19NSIP-B081011-06

국토공간정보의 빅데이터관리, 분석 및 서비스 플랫폼 기술개발

최종보고서

2020

국토교통부

국토교통과학기술진흥원

보안 과제( ), 일반 과제(○) / 공개(○), 비공개( ) 발간등록번호(11-B552989-000217-01)

국토공간정보연구사업 제6차연도 최종보고서

R&D / 19NSIP-B081011-06

국토공간정보의 빅데이터 관리, 분석 및 서비스 플랫폼 기술개발 (연구단)

2020. 02.

주관연구기관/안양대학교 산학협력단
협동연구기관/한국전자통신연구원
협동연구기관/한국과학기술원
공동연구기관/아이씨티웨이(주)
공동연구기관/(주)포스웨이브
공동연구기관/서울대학교 산학협력단
공동연구기관/(주)시터스
공동연구기관/연세대학교 산학협력단
공동연구기관/(주)가이온
공동연구기관/(주)유엔진솔루션즈
공동연구기관/(주)공간인소프트
공동연구기관/(주)포도

국 토 교 통 부
국토교통과학기술진흥원

제 출 문

국토교통과학기술진흥원장 귀하

본 보고서를 공간빅데이터 연구단 '국토공간정보의 빅데이터 관리, 분석 및 서비스 플랫폼 기술개발'(연구개발 기간 : 2014.02.14. ~ 2019.12.31.)의 최종보고서로 제출합니다.

2020. 02.

주관연구기관 : 안양대학교 산학협력단 (대표자) 권 원 현 (인)

협동연구기관 : 한국전자통신연구원 (대표자) 김 명 준 (인)

협동연구기관 : 한국과학기술원 (대표자) 신 성 철 (인)

공동연구기관 : 아이씨티웨이 (대표자) 조 유 복 (인)

공동연구기관 : 포스웨이브 (대표자) 변 재 학 (인)

공동연구기관 : 서울대학교 산학협력단 (대표자) 윤 의 준 (인)

공동연구기관 : 시터스 (대표자) 이 준 표 (인)

공동연구기관 : 연세대학교 산학협력단 (대표자) 이 충 용 (인)

공동연구기관 : 가이온 (대표자) 강 현 섭 (인)

공동연구기관 : 유엔진솔루션즈 (대표자) 장 진 영 (인)

공동연구기관 : 긍간인소프트 (대표자) 김 원 균 (인)

공동연구기관 : 포도 (대표자) 권 일 룡 (인)

주관연구기관책임자(안양대학교 산학협력단) : 신 동 빈
협동연구기관책임자(한국전자통신연구원) : 이 강 우
협동연구기관책임자(한국과학기술원) : 이 재 길
공동연구기관책임자(아이씨티웨이) : 김 상 수
공동연구기관책임자(포스웨이브) : 황 규 문
공동연구기관책임자(서울대학교 산학협력단) : 유 기 윤
공동연구기관책임자(시터스) : 이 종 오
공동연구기관책임자(연세대학교 산학협력단) : 허 준
공동연구기관책임자(가이온) : 이 헌 규
공동연구기관책임자(유엔진솔루션즈) : 장 진 영
공동연구기관책임자(긍간인소프트) : 김 원 균
공동연구기관책임자(포도) : 이 효 균

국토교통부소관 연구개발사업 운영규정 제29조에 따라 최종보고서 열람에 동의합니다.

보고서 요약서

과제 고유 번호	19NSIP-B08 1011-06	해당 단계 연구 기간	2019.01.01.~ 2019.12.31.	단계구분	6/6
연구사업명	중사업명	국토공간정보연구사업			
	세부사업명	국토공간정보의 빅데이터 관리, 분석 및 서비스 플랫폼 기술개발			
연구과제명	대과제명	국토공간정보의 빅데이터 관리, 분석 및 서비스 플랫폼 기술개발			
	세부과제명	공간 빅데이터 저장관리 인프라 기술 개발(1세부) 공간 빅데이터 분석 및 가시화 기술 개발(2세부) 공간 빅데이터 서비스 활성화 및 실증 연구(3세부)			
연구책임자	신동빈	해당단계 참여연구원 수	총: 명 내부: 명 외부: 명	해당단계 연구비	정부: 4,527,000천원 민간: 1,541,423천원 계: 6,068,423천원
		총 연구기간 참여연구원 수	총: 명 내부: 명 외부: 명	총 연구비	정부: 15,729,000천원 민간: 5,329,050천원 계: 21,058,050천원
연구기관명 및 소속 부서명	안양대학교 산학협력단 스마트도시공간연구소	참여기업명: 한국전자통신연구원, 한국과학기술원, 아이씨티웨이, 포스웨이브, 서울대학교 산학협력단, 시터스, 연세대학교 산학협력단, 가이온, 유엔진솔루션즈, 공간인소프트, 포도			
국제공동연구	상대국명:			상대국 연구기관명:	
위탁연구	연구기관명:			연구책임자:	
※ 국내·외의 기술개발 현황은 연구개발계획서에 기재한 내용으로 갈음					
연구개발성과의 보안등급 및 사유	일반				

9대 성과 등록·기탁번호

구분	논문	특허	보고서 원문	연구시설·장비	기술요약 정보	소프트웨어	화합물	생명자원		신품종	
								생명정보	생물자원	정보	실물
등록·기탁 번호	Y	Y	Y	Y	Y	Y	N	N	N	N	N

국가과학기술종합정보시스템에 등록한 연구시설·장비 현황

구입기관	연구시설·장비명	규격 (모델명)	수량	구입연월일	구입가격 (천원)	구입처 (전화)	비고 (설치장소)	NTIS 등록번호

요약

- 본 연구는 “저비용, 고성능, 고확장성의 공간 빅데이터(Geospatial Big Data)의 저장, 관리, 분석 및 서비스 제공을 위한 핵심기술과 이를 구현한 공간 빅데이터 관리, 분석 및 서비스 플랫폼 개발”을 최종 목표로 하고 있음
  - ▪ 본 연구에서의 공간 빅데이터는 기존 공간 정보시스템에서 다루던 대용량 공간 데이터(수치지도, 지적도, 지형도 등) 뿐만 아니라 공간 정보를 포함한 빅데이터(위치기반 SNS, GeoSensor 정보, 부동산 거래 정보, 버스 카드 리더기 정보 등)로 정의함
- 공간 빅데이터의 수집, 저장, 관리 및 질의 기능 제공을 위한 핵심기술과 이를 구현한 공간 빅데이터 저장관리 시스템을 연구 개발하여 4개의 핵심성과를 개발
- 소셜 미디어 데이터, 기존 공간 데이터, 이동 객체 데이터, 센서 데이터 등에 내재된 공간 빅데이터를 위한 다양한 최첨단 분석 알고리즘 및 가시화 기술 개발하여 4개의 핵심성과를 개발
- 공간 빅데이터 플랫폼 활용 시스템 구축을 위해 구체적인 구현방안을 제시하고, 핵심서비스 모델과 실증 서비스 기술을 개발하여 공간 빅데이터 서비스 시스템 구축 및 6개의 핵심성과를 개발

보고서 면수

476

요약문

<table>
<tr><td>연구의
목적 및 내용</td><td colspan="5">본 연구는 “저비용, 고성능, 고확장성의 공간 빅데이터(Geospatial Big Data)의 저장, 관리, 분석 및 서비스 제공을 위한 핵심기술과 이를 구현한 공간 빅데이터 관리, 분석 및 서비스 플랫폼 개발”을 최종 목표로 하며, 본 연구를 통해 개발된 플랫폼은 다음 성능 조건을 만족시킴
▪ 기존 방식 대비 데이터 처리 속도가 최소 30% 이상 향상된 페타바이트 이상의 공간 빅데이터 저장, 관리, 처리를 위한 Spatial Hadoop & Analytics 기술 개발
▪ 기존 공간 빅데이터 시스템 구축비용 대비 최소 50% 이상 저렴한 공간 빅데이터 SW (Spatial Hadoop & Analytics) 기술 개발
본 연구는 최종 목표 달성을 위해 다음과 같은 세부 연구 내용으로 구성함
▪ 공간 빅데이터 저장 관리기술, 처리 언어 시스템, 질의 처리 기술, 시맨틱 기반 ETL 도구 등 공간 빅데이터 저장 관리 인프라 기술 개발(TRL 8 : 시제품 설치 및 성능시험 단계)
- 공간 빅데이터 저장관리 인프라 기술 개발
- 인메모리 기반 공간 빅데이터 저장관리 프레임워크 개발
- 공간 빅데이터 웨어하우스 핵심기술 개발
▪ 공간 빅데이터 분석 프레임워크, Batch 분석, 이동객체 Batch 분석, Interactive 분석 기술 등 공간 빅데이터 분석 핵심기술과 컴포넌트 개발(TRL 8 : 시제품 설치 및 성능시험 단계)
- 공간 빅데이터 분석 프레임워크 기술 개발
- 공간 빅데이터 일괄(Batch) 분석 컴포넌트 기술 개발
- 공간 빅데이터 Interactive 분석 컴포넌트 기술 개발
▪ 공간 빅데이터 활용 서비스 프레임워크, UI/UX 및 Open API, 실증 서비스 시스템 구축 등 서비스 활성화 및 실증(TRL 8 : 시제품 설치 및 성능시험 단계)
- 공간 빅데이터 서비스 활성화 방안 연구
- 공간 빅데이터 서비스 실용화 연구
- 공간 빅데이터 서비스 실증 연구</td></tr>
<tr><td>연구개발성과</td><td colspan="5">▪ 공간 빅데이터의 수집, 저장, 관리 및 질의 기능을 제공하는 고성능, 고확장성의 공간 빅데이터 저장 관리 시스템
- Hadoop기반 공간 빅데이터 저장 프레임워크(SW), 공간 빅데이터 질의처리 엔진(SW), 인메모리기반 공간 빅데이터 저장관리 프레임워크(SW), 스트림 기반 분산 공간 ETL 도구(SW) 등
▪ 소셜 및 실시간 취득 공간정보 빅데이터 분석 알고리즘 및 SW
- 소셜 미디어 정보를 이용한 공간 빅데이터 분석 플랫폼, 공간 통계/비통계 기반 빅데이터 분석 알고리즘 및 시스템, 실시간 공간 빅데이터 분석 알고리즘 등
▪ 공간 빅데이터 플랫폼 활용을 위한 서비스 활성화 및 실증
- 공간 빅데이터 활용 서비스 모델 및 시나리오, 공간 빅데이터 서비스 제공을 위한 UI/UX, OpenAPI, 공간 빅데이터 분석 결과 연계 및 가시화 모듈(SW) 등</td></tr>
<tr><td>연구개발성과의
활용계획
(기대효과)</td><td colspan="5">기술적 기대성과
▪ 최신 IT 인프라 기반 공간 빅데이터 플랫폼 구축으로 미래 지식기반 사회를 지향하는 선진국 대비 기술 경쟁 우위 확보 기대
- 대규모 공간 빅데이터에 대한 효율적인 저장·관리·검색 기능과 일괄/실시간 분석을 제공할 수 있는 공간 빅데이터 원천 기술 확보 기대
- 기존 유무선 통신, IoT, 스마트폰 등 세계적인 경쟁력을 보유한 IT 인프라와의 공간 빅데이터 처리기술 융합을 통한 독자적인 공간정보 신 시장 창출 및 경쟁력 확보 기대
사회/경제적 파급효과
▪ 국내 공간정보 SW의 고도화를 통한 외산 소프트웨어 중심의 기존 공간정보 시장 재편 및 해외 시장 진출 기대
- 국내 공간정보 시장의 해외의존도 저감('12년 80%이상→'20년 50% 이하), 공간정보 지식서비스 세계시장 점유율 제고 ('20년 10%이상)
▪ 국내 공간 빅데이터의 효율적 저장 관리와 대국민 참여/공유/확산 서비스를 통한 정부 3.0 실현 기대
- 정부 3.0 : 개방형 공간 빅데이터 플랫폼 서비스를 통한 상생경제 실현 (국가 공간정보/공공정보의 중소기업 제공 확대)
활용방안
▪ 국가의 공간 빅데이터 체계 구축을 위한 핵심기술로 활용 기대
- 공간 빅데이터 체계 구축을 위한 플랫폼 구축에 본 과제에서 개발된 공간 빅데이터 저장 관리 및 분석기술 적용 기대
▪ 공간정보 제공을 위한 공간 빅데이터 연계 서비스 제공에 활용
- 실시간 이동경로 분석, 대규모 교통 모니터링, 재난재해 대응 등과 같은 다양한 공간 빅데이터 실증 서비스 제공을 위한 핵심기술로 공간 빅데이터 저장 관리, 분석 및 서비스 기술 활용 기대</td></tr>
<tr><td>국문핵심어
(5개 이내)</td><td>공간 빅데이터</td><td>Spatial 하둡</td><td>공간 빅데이터 분석</td><td>공간 빅데이터 서비스</td><td>개방형 API</td></tr>
<tr><td>영문핵심어
(5개 이내)</td><td>Geo-Spatial Big Data</td><td>Spatial Hadoop</td><td>Geo-Spatial Big Data Analytics</td><td>Geo-Spatial Big Data Service</td><td>Open API</td></tr>
</table>

< 목 차 >

1. 연구개발과제의 개요 ····· 1

가. 추진 배경 및 목적 ····· 1

(1) 추진 배경 ····· 1

(2) 추진 목적 및 연구의 필요성 ····· 2

나. 연구개발 목표 ····· 5

(1) 최종 목표 ····· 5

(2) 중점 추진 내용 ····· 5

다. 연구개발의 내용 및 범위 ····· 7

(1) 공간 빅데이터 저장관리 인프라기술 개발 ····· 7

(2) 공간 빅데이터 분석 및 가시화 기술 개발 ····· 14

(3) 공간 빅데이터 서비스 활성화 및 실증 연구 ····· 21

(4) 연구내용 변동사항 ····· 28

2. 연구수행내용 및 성과 ····· 41

가. 연구수행내용 ····· 42

(1) 공간 빅데이터 저장관리 인프라 기술 개발 ····· 42

(2) 공간 빅데이터 분석 및 가시화 기술 개발 ····· 131

(3) 공간 빅데이터 서비스 활성화 및 실증 연구 ····· 295

나. 연구개발 성과 ····· 379

(1) 최종 목표에 따른 연구개발 성과 ····· 379

(2) 공간 빅데이터 저장관리 인프라 기술 개발 ····· 382

(3) 공간 빅데이터 분석 및 가시화 기술 개발 ····· 389

(4) 공간 빅데이터 서비스 활성화 및 실증 연구 ····· 391

3. 목표 달성도 및 관련 분야 기여도 ····398

가. 연차별 연구목표 달성도 ····398

(1) 공간빅데이터 저장 관리 인프라 기술 개발 ····398

(2) 공간 빅데이터 분석 및 가시화 기술 개발 ····403

(3) 공간 빅데이터 서비스 활성화 및 실증 연구 ····413

나. 연구의 한계 및 시사점 ····419

다. 관련 분야 기여도 ····437

(1) 공간빅데이터 저장 관리 인프라 기술 개발 ····437

(2) 공간 빅데이터 분석 및 가시화 기술 개발 ····440

(3) 공간 빅데이터 서비스 활성화 및 실증 연구 ····442

4. 연구개발성과의 활용 계획 ····444

가. 활용계획 ····444

(1) 공간빅데이터 저장 관리 인프라 기술 개발 ····444

(2) 공간 빅데이터 분석 및 가시화 기술 개발 ····455

(3) 공간 빅데이터 서비스 활성화 및 실증 연구 ····460

나. 추가 연구 및 사업화 방안 ····467

(1) 공간 빅데이터 저장관리 인프라기술 개발 ····467

(2) 공간 빅데이터 분석 및 가시화 기술 개발 ····468

(3) 공간 빅데이터 서비스 활성화 및 실증 연구 ····469

붙임1. 참고 문헌 ····470

붙임2. 공간 빅데이터 관리, 분석 및 서비스 플랫폼의 세부별 모듈 구성도(물리적 관점) ··474

붙임3. 공간 빅데이터 관리, 분석 및 서비스 플랫폼의 세부 모듈 간 연계 구성도 ····475

< 그림 목차 >

<그림 1-1> 연구의 추진배경 ····· 1

<그림 1-2> 빅데이터와 공간정보 기술의 융합 필요성 ····· 2

<그림 1-3> 연구개발 최종목표 ····· 5

<그림 1-4> 중점추진 내용 ····· 6

<그림 1-5> 공간 빅데이터 저장관리 인프라기술 개발의 주요연구내용 및 구조 ····· 7

<그림 1-6> 1세부의 세세부별 연구수행 내용 및 구조 ····· 8

<그림 1-7> 연차별 연구내용 ····· 9

<그림 1-8> 공간 빅데이터 분석 및 가시화 기술 개발의 주요연구내용 및 구조 ····· 14

<그림 1-9> 연차별 연구내용 ····· 15

<그림 1-10> 공간 빅데이터 서비스 활성화 및 실증 연구의 주요연구내용 및 구조 ····· 21

<그림 1-11> 연차별 연구내용 ····· 22

<그림 2-1> Hadoop 기반 공간 빅데이터 저장 프레임워크(핵심성과 A) ····· 43

<그림 2-2> 공간 빅데이터 저장 관리 시스템 구조 ····· 45

<그림 2-3> 공간 빅데이터 처리 시스템 내부 표현(예시) ····· 48

<그림 2-4> Record 형식 데이터 저장 및 접근 ····· 49

<그림 2-5> 공간 빅데이터 연산의 예: ExecSlope ····· 49

<그림 2-6> 연산자 기반 공간 빅데이터 처리 구조 ····· 50

<그림 2-7> 공간 빅데이터 처리 연산자 ····· 50

<그림 2-8> 복합 연산의 예: 지하철역 50m 이내 지적 검색 ····· 50

<그림 2-9> MR 작업 최적화 ····· 51

<그림 2-10> Hadoop을 이용한 R-Tree 색인 구축 ····· 52

<그림 2-11> 색인 공간의 크기 비교 ····· 53

<그림 2-12> 겹침 영역의 크기 비교 ····· 53

<그림 2-13> 공간 연산자 및 색인 생성 및 활용 성능평가 결과 ····· 54

<그림 2-14> 인메모리 Quad-Tree를 활용한 공간 분할의 예 ····· 54

<그림 2-15> 서울지역의 공간 색인 구성의 예 ·····55
<그림 2-16> 공간 색인 생성 관련 연산자 수행 과정 ·····55
<그림 2-17> Map/Reduce 기반 공간 색인 생성 과정 ·····56
<그림 2-18> 공간 빅데이터 언어 규격 정의(예시) ·····56
<그림 2-19> 공간분석 plan 저작 언어 프로토타입 ·····57
<그림 2-20> 기존의 데이터 적재 과정 및 데이터 분석 과정 ·····57
<그림 2-21> 기존 공간 빅데이터 플랫폼에서의 데이터 적재와 분석 과정 ·····58
<그림 2-22> 공간 빅데이터 플랫폼에서 고속 데이터 적재와 분석 과정 ·····58
<그림 2-23> 데이터 적재 방식에 따른 성능 비교 ·····59
<그림 2-24> 기존 중첩 루프 공간 조인 방법 ·····60
<그림 2-25> 개발된 중첩 루프 공간 조인 방법 ·····60
<그림 2-26> 공간 조인 성능 평가 ·····61
<그림 2-27> ‘10분 정책지도 테스트’ 공간 분석 응용 ·····62
<그림 2-28> 10분 정책지도 테스트 분석 결과 ·····62
<그림 2-29> 연구단-사업단 플랫폼 성능 평가 결과 ·····63
<그림 2-30> 대용량 DTG 데이터를 활용한 차량통행량 분석 ·····63
<그림 2-31> C#용 클라이언트 모듈 활용 응용 ·····65
<그림 2-32> GeoServer 플러그인 활용 ·····65
<그림 2-33> GUI 기반 공간 빅데이터 응용 개발 도구 ·····66
<그림 2-34> 스크립트 공간 분석(10분 정책지도 중 일부) ·····67
<그림 2-35> 워크플로우 가시화 도구 (Composite Analysis 보기) ·····68
<그림 2-36> 워크플로우 가시화 도구 (Module Analysis 정의 화면) ·····69
<그림 2-37> 워크플로우 가시화 도구 (Plan Analysis 정의 화면) ·····69
<그림 2-38> 사업단 실증 과정 ·····70
<그림 2-39> 공간분석 워크플로우 저작 도구 ·····72
<그림 2-40> 공간빅데이터 탐색기 ·····72
<그림 2-41> 대용량 공간 빅데이터 원격 접근 가시화 ·····74

<그림 2-42> GeoServer 연동을 통한 레이어 계시 기능 화면 ·····75
<그림 2-43> 공간색인 파티션 기반 클라이언트 캐싱 ·····75
<그림 2-44> 추가된 공간 연산자 ·····77
<그림 2-45> 공간 빅데이터 질의처리 엔진(핵심성과 B) ·····78
<그림 2-46> 공간 빅데이터 질의처리 엔진 구조 ·····81
<그림 2-47> 공간 빅데이터 질의언어 정의 ·····82
<그림 2-48> 공간 빅데이터 DBMS 인터페이스 시스템 ·····83
<그림 2-49> 공간 빅데이터 DBMS 인터페이스 기본기능 모듈 ·····84
<그림 2-50> 공간 빅데이터 DBMS 인터페이스 기본기능 테스트 ·····84
<그림 2-51> 질의실행계획 생성기와 질의실행기 ·····85
<그림 2-52> XML로 표현된 질의 플랜(일부) ·····85
<그림 2-53> 질의언어 파서와 검증기 ·····86
<그림 2-54> 웹 기반 클라이언트 인터페이스 ·····87
<그림 2-55> 공간 빅데이터 질의처리 엔진 실행 흐름 ·····88
<그림 2-56> 공간 빅데이터 JDBC 드라이버를 활용한 클라이언트 ·····89
<그림 2-57> 기존 공간분석 처리 방식 ·····90
<그림 2-58> 스트림 기반 분석 처리 방식 ·····90
<그림 2-59> 이벤트 처리 시스템 구조 ·····91
<그림 2-60> 공간 빅데이터 인터페이스 프레임워크 전체 구조 ·····92
<그림 2-61> 공간 빅데이터 클라이언트 ·····93
<그림 2-62> 건물 영역의 중첩 조사에 의한 잠재적 오류 식별 ·····94
<그림 2-63> 공간 빅데이터 질의처리 시스템 구조 및 외부 연동 ·····94
<그림 2-64> 공간 빅데이터 질의처리 엔진 테스트 구조 ·····95
<그림 2-65> 공간 빅데이터 클라이언트 프로그램에 의한 기능 테스트 ·····95
<그림 2-66> 스트림 기반 분산 공간 ETL 도구(핵심성과D) ·····96
<그림 2-67> 실시간 고속 ETL 기술의 필요성 및 요구사항 ·····97
<그림 2-68> 공간 빅데이터 웨어하우징 주요 기능 ·····98

<그림 2-69> 워크플로우 기반 ETL 기능 프레임워크 ····· 99
<그림 2-70> ETL 시스템 구조 및 처리 플로우 ····· 99
<그림 2-71> ETL 규칙(예시) ····· 100
<그림 2-72> ETL 수집 알고리즘 ····· 101
<그림 2-73> 실시간 분산 병렬 ETL 구조 ····· 102
<그림 2-74> 다중 공간 데이터 수집 모듈 설계도 ····· 103
<그림 2-75> Spatial ETL ToolKit 기본 화면 ····· 103
<그림 2-76> 공간 스트림 빅데이터 수집 및 변환 모듈 순서도 ····· 104
<그림 2-77> 공간 스트림 빅데이터 적재 모듈 연구개발 결과 ····· 104
<그림 2-78> 병렬 및 분산처리 모듈 구현도 ····· 105
<그림 2-79> 스트림 기반 공간 ETL 처리 엔진 ····· 106
<그림 2-80> 스트림 기반 공간 ETL 처리 엔진 연동 API 흐름도 ····· 107
<그림 2-81> 제공 API 리스트 Lite 버전 ····· 108
<그림 2-82> 스트림 기반 공간 ETL 규칙 처리 엔진 구성도 ····· 109
<그림 2-83> 공간 센서 스트림 데이터 수집 및 분산처리 모듈 설계안 ····· 110
<그림 2-84> 스트림 기반 공간 ETL 규칙처리 엔진 구성도 ····· 111
<그림 2-85> 스트림 기반 공간 ETL 처리 프로그램에 직관적 표현 방식 ····· 112
<그림 2-86> CSV to Shapefile 변환 (QGIS) ····· 112
<그림 2-87> Shapefile 내 공간 좌표계를 EPSG:4326으로 변환하고 CSV file 포맷으로 HDFS에 저장한 결과 ····· 113
<그림 2-88> 부가 기능: 레코드의 수를 출력하는 기능 ····· 113
<그림 2-89> 부가 기능: 일부 저장된 공간 레코드 정보를 출력하는 기능 ····· 113
<그림 2-90> (좌)Ambari Metrics 메인화면, (우)공간 스트림 ETL 관리도구 메인화면 ····· 114
<그림 2-91> 공간 ETL 워크플로우 관리 주기 ····· 114
<그림 2-92> 공간 ETL 워크플로우 관리 화면 ····· 115
<그림 2-93> 공간 ETL 파일 Browsing 화면 ····· 115
<그림 2-94> Apache NiFi 메인화면 및 10종의 API 라이브러리 목록 ····· 116

<그림 2-95> Shapefile을 CSV, GeoJson 포맷으로 변환 후 로컬에 저장 ·····116
<그림 2-96> (좌)HUE를 이용한 결과 파일 생성 확인, (우) QGIS를 이용한 결과 파일 가시화 ····117
<그림 2-97> QGIS를 이용한 컬럼 정보 변환 확인 ·····117
<그림 2-98> 인메모리 기반 공간빅데이터 저장관리 프레임워크 개념구조(핵심성과 0) ·····118
<그림 2-99> 실시간 공간빅데이터 처리시스템안의 데이터 흐름 ·····119
<그림 2-100> 실시간 공간 빅데이터 처리시스템 개념 구성도 ·····119
<그림 2-101> 실시간 공간 빅데이터 플랫폼 아키텍쳐 ·····120
<그림 2-102> 인메모리 기반 공간색인 모듈 ·····121
<그림 2-103> 색인 생성 성능 비교 ·····122
<그림 2-104> 색인 조회 성능 비교 ·····122
<그림 2-105> 공간연산자 성능 비교(10회 평균) ·····123
<그림 2-106> 이동객체 빅데이터 저장 플랫폼 개념설계 ·····123
<그림 2-107> 이동객체 저장 구조 설계안 ·····124
<그림 2-108> DTG 자료 저장 내용 ·····124
<그림 2-109> 자료량에 따른 초당 처리 속도 ·····125
<그림 2-110> 고유 속성수에 따른 처리시간 ·····125
<그림 2-111> 이동경로 최적화 예시 ·····126
<그림 2-112> 시공간 색인 구조 개념도 ·····126
<그림 2-113> 시간축 분할 예 ·····127
<그림 2-114> 시공간 색인을 통한 조회 ·····127
<그림 2-115> 시공간 질의 프로세스 ·····128
<그림 2-116> 공간 질의 프로세스 ·····128
<그림 2-117> SparkSQL 기반 Spatial SQL 목록 ·····129
<그림 2-118> OpenSource(GeoSpark) Vs 연구 성과물 ·····130
<그림 2-119> 공간 빅데이터 소셜 분석 기술 및 공간 빅데이터 분석 및 가시화 컴포넌트 구현 (핵심성과 E) ·····133
<그림 2-120> 소셜 데이터에 대한 공간 빅데이터 분석 모델 기법 ·····134

<그림 2-121> 공간 빅데이터 분석 모델 설계도 ····135
<그림 2-122> 연구흐름도 ····135
<그림 2-123> 트윗(point), 주택가격지수(polygon) 및 격자망 데이터(grid) ····138
<그림 2-124> 트윗 격자 데이터의 핫스팟 분석 결과 ((a) 500m grid, (b) 1,000m grid, (c) 2,000 m grid) ····138
<그림 2-125> 주택매매가격지수 격자 데이터의 핫스팟 분석 결과 ····139
<그림 2-126> 주택전세가격지수 격자 데이터의 핫스팟 분석 결과 ····139
<그림 2-127> 소셜 미디어 데이터의 공간자료화 프로세스 ····141
<그림 2-128> 중복 데이터 제거 프로세스 ····142
<그림 2-129> 제안된 변환 기법을 적용한 사례 ····144
<그림 2-130> 소셜 미디어 데이터 분석 서비스 필터링 결과 화면 ····145
<그림 2-131> 캐싱 Work-flow ····145
<그림 2-132> 버퍼링 프로세스 Work-flow ····146
<그림 2-133> 감성 분석기반 장소선호도 분석 분포 및 차트 서비스 사례 ····153
<그림 2-134> GPS 정보 추정 분포 및 차트 서비스 ····154
<그림 2-135> 장소명 대체어 분석 분포 및 차트 서비스 사례 ····156
<그림 2-136> LDA 문서생성 과정 (D: 말뭉치 전체 문서 개수, K: 전체 토픽수, N: d번째 문서의 단어 수) ····156
<그림 2-137> 공공시설 매력도 및 접근성 분석 차트 및 분포서비스 ····157
<그림 2-138> 시간정보를 표시하는 방법의 예: (a) color-coded road segments, (b) free-from color-coded regions, (c) Distance Cartograms ····158
<그림 2-139> (a) 시간공간의 불연속성으로 인한 공간겹침의 예, (b) S-GAP 을 통한 평면상에 공간겹침 없이 DC를 제작하는 예시 ····159
<그림 2-140> 장소 감성어 사전 분포 및 차트 서비스 사례 ····160
<그림 2-141> 장소 테마 분석 차트 서비스 ····162
<그림 2-142> 머신러닝 기반 문서분류 서비스 ····163
<그림 2-143> 머신러닝 기반 문장분류 서비스 ····164

<그림 2-144> key player 탐지 차트 및 분포 서비스 ····· 165
<그림 2-145> 유사 장소 검색 서비스 화면 예시> ····· 167
<그림 2-146> 토픽 기반 장소 검색 서비스 화면 예시> ····· 167
<그림 2-147> 임베딩 알고리즘 ····· 168
<그림 2-148> 사용자 맞춤 추천 연구흐름도 ····· 169
<그림 2-149> 사용자 맞춤 POI 추천 ····· 169
<그림 2-150> 단어 임베딩을 활용한 사용자 관심 이벤트 추출 서비스 화면 예시 ····· 170
<그림 2-151> 단어 임베딩을 활용한 사용자 관심 이벤트 추출 서비스 화면 예시 ····· 171
<그림 2-152> 공공자전거 수요예측 기술 전체 흐름도 ····· 172
<그림 2-153> 통행목적을 고려한 공공자전거 수요예측 가시화 서비스 ····· 173
<그림 2-154> 02 Overall Architecture ····· 175
<그림 2-155> SAP HANA Geo-Spatial Features ····· 177
<그림 2-156> SAP HANA 아키텍처 ····· 178
<그림 2-157> GIS Tools for Hadoop ····· 179
<그림 2-158> GIS Tools for Hadoop ····· 182
<그림 2-159> 소셜 미디어 공간 빅데이터 수집 흐름 ····· 183
<그림 2-160> SNA(Social Network Analysis) 분석용 실행파일 구성 ····· 183
<그림 2-161> 환경설정 파일내용 ····· 184
<그림 2-162> 수집 모듈 실행 화면 ····· 184
<그림 2-163> 수집 정보 결과 ····· 184
<그림 2-164> 인스타그램(Instagram) 실행파일 구성 ····· 185
<그림 2-165> 인스타그램(Instagram) 환경설정 파일 예시 ····· 185
<그림 2-166> 인스타그램(Instagram) 수집 모듈 실행 화면 ····· 185
<그림 2-167> 인스타그램(Instagram) 수집 모듈 인증 화면 ····· 186
<그림 2-168> 인스타그램(Instagram) 수집 화면 ····· 186
<그림 2-169> 유투브(Youtube) 실행파일 구성 ····· 186
<그림 2-170> 유투브(Youtube) 환경설정 파일 예시 ····· 187

<그림 2-171> 유투브(Youtube) 수집 모듈 실행 화면 ·····187
<그림 2-172> 유투브(Youtube) 수집 화면 ·····187
<그림 2-173> 수집 데이터 DB 적재 흐름 ·····188
<그림 2-174> 파일변환(JSON->CSV) 흐름 ·····188
<그림 2-175> V-World의 Geocoder API 기능 ·····189
<그림 2-176> 필터링 설정 화면 ·····190
<그림 2-177> 행정검색 설정 화면 ·····190
<그림 2-178> 검색어 입력에 의한 위치검색 설정 화면 ·····191
<그림 2-179> 캐시 모듈 구조 ·····191
<그림 2-180> 버퍼링 모듈 구조 ·····192
<그림 2-181> 프리패칭 모듈 구조 ·····193
<그림 2-182> 정합 모듈 분석 흐름도 ·····193
<그림 2-183> 주제도 모듈 분석 흐름도 ·····195
<그림 2-184> OpenAPI 처리 흐름도 ·····198
<그림 2-185> OpenAPI 연동 규격 시스템 구조 ·····198
<그림 2-186> OpenAPI 응답 결과 화면 ·····198
<그림 2-187> 공간 빅데이터 분포 분석 모듈의 시스템 구조도 ·····199
<그림 2-188> 관광 POI 사전의 데이터 현황 ·····201
<그림 2-189> 행정동 단위의 데이터 현황 ·····202
<그림 2-190> 시군구 단위의 데이터 현황 ·····202
<그림 2-191> 분포 표출용 정보의 데이터 현황 ·····203
<그림 2-192> 시도 레벨 ·····203
<그림 2-193> 시군구 레벨 ·····203
<그림 2-194> 읍면동 레벨 ·····204
<그림 2-195> 시설물 레벨 ·····204
<그림 2-196> 분포 데이터 현황 및 상세 정보 ·····204
<그림 2-197> 공간 빅데이터 차트 분석 모듈의 시스템 구조도 ·····205

<그림 2-198> 차트 표출용 정보의 데이터 현황 ·····206
<그림 2-199> 차트 데이터 현황 및 상세 정보 ·····206
<그림 2-200> 예측 지수 Z 산정을 위한 연구 흐름도 ·····207
<그림 2-201> 유동인구 예측용 가공 도구 ·····207
<그림 2-202> 유동인구 예측 가시화 ·····208
<그림 2-203> json 형식 결과 ·····208
<그림 2-204> xml 형식 결과 ·····208
<그림 2-205> 이벤트 탐지를 위한 분석 흐름도 ·····209
<그림 2-206> 시계열 분석을 통한 이상치 탐지 ·····209
<그림 2-207> 단어 임베딩 및 이상 시점의 핫키워드들을 워드클라우드 표출 ·····210
<그림 2-208> json형태의 장소 정보 추출 결과 화면 ·····210
<그림 2-209> 통행 목적을 고려한 공유자전거 수요예측을 위한 분석 흐름도 ·····211
<그림 2-210> 바플롯(barplot) 시계열 분석 ·····211
<그림 2-211> 박스플롯(boxplot) 시계열 분석 ·····212
<그림 2-212> 산점도(regplot) 시계열 분석 ·····212
<그림 2-213> 히트맵(heatmap) 시계열 분석 ·····213
<그림 2-214> 예측 모델 및 수요 예측의 시계열 분석 ·····213
<그림 2-215> json형태의 공공자전거 수요 예측 결과 화면 ·····213
<그림 2-216> 1세부 연계 구조도 ·····214
<그림 2-217> 데이터 저장 ·····214
<그림 2-218> 데이터 목록 확인 ·····215
<그림 2-219> 데이터 스키마 확인 ·····215
<그림 2-220> 데이터 내용 확인 ·····215
<그림 2-221> 데이터 조회를 위한 시스템 구조 ·····216
<그림 2-222> 사업단 및 연구단 구성 개념도 ·····217
<그림 2-223> R&D 성과물의 사업단 적용 후 구성 개념도 ·····218
<그림 2-224> R&D 성과물의 사업단 실증을 위한 H/W 구성 ·····218

<그림 2-225> R&D 성과물의 사업단 실증을 위한 S/W 구성 ·····219
<그림 2-226> 머신러닝 기반 문서분류 모듈 처리 흐름도 ·····220
<그림 2-227> 형태소 분석 (도구1) ·····220
<그림 2-228> 공공시설 관련 문서분류 (도구2) ·····220
<그림 2-229> 토픽분석 (도구3) ·····221
<그림 2-230> 공간접근성 분석 (도구4) ·····221
<그림 2-231> 학습 데이터 증강 도구 (도구1) ·····221
<그림 2-232> 문장 추출 모델 학습 도구 (도구2) ·····222
<그림 2-233> 특정 주제 포함 문장 추출 도구 (도구3) ·····222
<그림 2-234> 분석모듈 설치위치 ·····223
<그림 2-235> 머신러닝 기반 문서분류 모듈의 크롬브라우저 사용 ·····223
<그림 2-236> 머신러닝 기반 문서분류 모듈의 가시화 도구 ·····224
<그림 2-237> 가시화 결과 ·····224
<그림 2-238> 공공시설 매력도 측정 및 접근성 분석의 크롬브라우저 사용 ·····224
<그림 2-239> 공공시설 매력도 측정 및 접근성 분석의 가시화 도구 ·····225
<그림 2-240> 가시화 결과 ·····225
<그림 2-241> 머신러닝 기반 특정주제 문장분류 모듈의 크롬브라우저 사용 ·····225
<그림 2-242> 머신러닝 기반 특정주제 문장분류 모듈의 가시화 도구 ·····226
<그림 2-243> 가시화 결과 ·····226
<그림 2-244> 주요연구내용 ·····226
<그림 2-245> 연간 공간 빅데이터 관련 문헌 빈도 ·····228
<그림 2-246> 공간 빅데이터 관련 문헌의 저자가 소속된 상위 10개 기관 ·····228
<그림 2-247> 국가별 공간 빅데이터 문헌 빈도 ·····228
<그림 2-248> 공간 빅데이터 연구 분야 ·····229
<그림 2-249> 미들웨어 간 인터페이스 관계도 ·····232
<그림 2-250> 공간 빅데이터 Batch 분석 컴포넌트 아키텍쳐 ·····232
<그림 2-251> 분석 요청 처리에 대한 시퀀스 다이어그램 ·····233

<그림 2-252> 오전 8시 최종 대중접근성 지수 ·····235
<그림 2-253> 오후 3시 최종 대중접근성 지수 ·····235
<그림 2-254> Parking Lot Footprints 결과 ·····235
<그림 2-255> 공간 빅데이터 분석 컴포넌트 프로토타이핑 ·····236
<그림 2-256> 대중교통 접근성 분석 진행 UI ·····236
<그림 2-257> 그리드 인덱싱 과정 Pseudo-Code ·····237
<그림 2-258> 강남구 상업지역의 통합 공시지가 이상지역 분석 결과 ·····238
<그림 2-259> 강남역-역삼역 사이의 통합 공시지가 이상지역 분석 결과 ·····238
<그림 2-260> 의사결정나무 기능 Framework ·····239
<그림 2-261> 의사결정나무 기능 분석 결과 ·····239
<그림 2-262> AWS와 Spark를 활용한 공간 빅데이터 분석 시스템 ·····240
<그림 2-263> AWS에 구성된 분석 클러스터 ·····240
<그림 2-264> 웹 콘솔의 Analysis Module 탭과 Component 탭 ·····241
<그림 2-265> DBSCAN에 대한 설명 팝업 화면 ·····241
<그림 2-266> 외부 전문가 소프트웨어 성능 평가 확인서 ·····242
<그림 2-267> 서울시 행정동별 보행자 교통사고 이상치 지점 탐색 결과 ·····242
<그림 2-268> 18대 대선 결과 및 중요 지역인자 검출 ·····243
<그림 2-269> 인구 유동패턴 기반 도시 군집 결과 ·····243
<그림 2-270> 빈집 발생 영향 요인 탐색 Flow Chart ·····244
<그림 2-271> 신규주택 비율(좌) 및 노후주택 비율(우)에 대한 빈집발생 영향요인 분석 결과 ·····245
<그림 2-272> 시간대별 생활인구 이상치 탐색 ·····245
<그림 2-273> Random Forest 분산처리 모니터링 화면 ·····245
<그림 2-274> 딥러닝을 활용한 대기오염물질 농도 예측 ·····246
<그림 2-275> 서울시 6개구 일산화탄소 예측 결과 (단위:ppm) ·····247
<그림 2-276> Recurrent Neural Network (RNN) ·····247
<그림 2-277> 따릉이 자전거 대여소별 부족/과다량 도시화 ·····247
<그림 2-278> 행정중심복합도시 내 유동인구 패턴 분석 결과 ·····249

<그림 2-279> LISA 알고리즘 분산처리 성능 테스트 결과 ·····251
<그림 2-280> 서울시 공시지가 이상지역 분석 결과 ·····251
<그림 2-281> 공간제약 클러스터링의 Workflow ·····251
<그림 2-282> 2차년도 공간 군집화 결과와, 공간 제약 클러스터링의 결과 비교 ·····252
<그림 2-283> 특정 지역, 특정 시점의 인구 유동량 예측 기능의 Workflow ·····252
<그림 2-284> 예측 유동인구의 분포와 실제 유동인구 분포의 시각화 ·····253
<그림 2-285> 유동인구 주요 패턴도출 기능의 Workflow ·····253
<그림 2-286> 시간대-월간 인구 유동 특징 분석 결과 ·····254
<그림 2-287> 1번 컴포넌트와 4번 컴포넌트의 공간적 분포 특징 매핑 결과 ·····255
<그림 2-288> 공간 가중치를 고려한 유병율 회귀분석 Local R2 매핑 결과 ·····256
<그림 2-289> 공간 빅데이터 분석 Batch 솔루션 ·····257
<그림 2-290> Drag-and-drop 및 저장 / 편집 기능 ·····257
<그림 2-291> E2SFCA 분석컴포넌트 결과 ·····258
<그림 2-292> Decision Tree 분석 컴포넌트 결과 ·····258
<그림 2-293> 주요연구내용 ·····258
<그림 2-294> 연간 이동객체데이터 관련 문헌 빈도 ·····260
<그림 2-295> 이동객체 데이터 관련 문헌의 저자가 소속된 상위 10개 기관 ·····260
<그림 2-296> 국가별 이동객체 데이터 문헌 빈도 ·····260
<그림 2-297> 이동객체 데이터 연구 분야 ·····260
<그림 2-298> 1, 2세부 간 컴포넌트 인터페이스 관계 다이어그램 ·····266
<그림 2-299> 이동객체 Batch 분석 컴포넌트 아키텍처 ·····266
<그림 2-300> Workflow 등록 시퀀스 다이어그램 ·····267
<그림 2-301> 서울시 00시~01시 택시 승·하차 point 데이터 ·····268
<그림 2-302> 서울시 집계구 단위 공간 군집화 결과 ·····269
<그림 2-303> 이동객체데이터 비통계기반 분석 컴포넌트 분석결과 ·····269
<그림 2-304> 이동객체데이터 분석 컴포넌트 연산자 동작 정확도 실험 결과 ·····270
<그림 2-305> 공간 군집화 분석 시작 UI ·····270

<그림 2-306> 공간 군집화 분석 진행 UI ·····270
<그림 2-307> 택시 서비스 필요지역 분석 결과 ·····271
<그림 2-308> 부산시 노인 복지시설 입지선정 분석 결과 ·····272
<그림 2-309> 택시 공차 지역 분석 결과 ·····272
<그림 2-310> AWS와 Spark를 활용한 이동객체데이터 분석 시스템 ·····273
<그림 2-311> 웹 콘솔의 Analysis Module 탭과 Component 탭 ·····273
<그림 2-312> 공간 빅데이터 및 이동객체 데이터 일괄 분석 기능 TTA 인증 ·····274
<그림 2-313> Space Time Scan Statistics 기반의 시공간 택시 수요 탐지 결과 ·····274
<그림 2-314> Map Matching 결과와 이에 따른 Network 단위의 택시 유동분석 ·····275
<그림 2-315> 택시 토픽 기반 LDA 클러스터링 ·····276
<그림 2-316> 오전 8-9시 강남지역 단거리 택시 운행 밀집 지역 도출 결과 ·····277
<그림 2-317> 소방시설 서비스 가능지역 Point 결과 도출 (전체시간) ·····277
<그림 2-318> 09~12시 Prefix-Span 군집화 결과 ·····278
<그림 2-319> 개발기능을 통해 제안한 유동인구 데이터를 이용한 활동지도 ·····279
<그림 2-320> 돌아가는 택시 도출 결과 ·····281
<그림 2-321> NW 알고리즘을 통해 얻어진 유사도 matrix ·····281
<그림 2-322> 비 정상적인 택시 이동 경로 도출 결과 ·····282
<그림 2-323> OD를 이용한 node 기반 분석 결과 (서울, 부산, 강원도) ·····282
<그림 2-324> OD를 이용한 link 기반 분석 (서울의 주요 인구 유동 도출) ·····283
<그림 2-325> 서울 5대 생활권 및 버스 권역과의 결과 비교 ·····283
<그림 2-326> 공간 빅데이터 분석 Batch 솔루션 ·····284
<그림 2-327> Drag-and-drop 및 저장 / 편집 기능 ·····284
<그림 2-328> Map Matching 및 Percentile Rank 분석 컴포넌트 결과 ·····285
<그림 2-329> 핵심성과 H의 대표 이미지 ·····286
<그림 2-330> UseCase ·····293
<그림 2-331> CEP 엔진 사업단 시범서비스 실행 화면 ·····294
<그림 2-332> OLAP 엔진 사업단 시범서비스 실행 화면 ·····294

<그림 2-333> 공간 빅데이터 플랫폼의 융복합 활용방안 개발(핵심성과 I) ·····296
<그림 2-334> V-World(공간정보 오픈플랫폼) 서비스 프레임워크 ·····297
<그림 2-335> 공간 빅데이터 플랫폼의 서비스 프레임워크 ·····299
<그림 2-336> 공간 빅데이터 서비스 활성화 지원을 위한 기본방향 및 대응과제 ·····300
<그림 2-337> 공간 빅데이터 실증서비스 시스템의 아키텍처 구성(안) ·····302
<그림 2-338> 사업단 시스템 지원을 위한 연구단 성과의 단계별 적용 방안 ·····303
<그림 2-339> 공간 빅데이터 플랫폼 메인화면 구성 ·····305
<그림 2-340> 실증시각화 화면 구성 ·····305
<그림 2-341> 공간분석프레임 화면 구성 ·····305
<그림 2-342> 관리서비스 화면 구성 ·····306
<그림 2-343> OpenAPI서비스 화면 구성 ·····306
<그림 2-344> 토지건물정보 기반 서비스 시나리오 적용예시(워크플로우) ·····307
<그림 2-345> 비정형텍스트정보 기반 서비스 시나리오 적용예시(워크플로우) ·····307
<그림 2-346> 네트워크정보 기반 서비스 시나리오 적용예시(워크플로우) ·····307
<그림 2-347> 공간 빅데이터의 공공/민간부문 핵심서비스 모델 도출 및 제공방안 개발(핵심성과 J) ·····308
<그림 2-348> 공간 빅데이터 서비스 차별화 방안 및 활용전략 ·····312
<그림 2-349> 공간 빅데이터 서비스 분야 우선순위 도출 ·····313
<그림 2-350> 서비스 유형화에 따른 핵심서비스모델 개발 절차 ·····313
<그림 2-351> 공간 빅데이터 비즈니스 모델 개발 절차 ·····315
<그림 2-352> 공간 빅데이터 서비스 유형 구분 ·····323
<그림 2-353> 토지건물정보 기반 서비스 시나리오 ·····324
<그림 2-354> 전국단위 격자분석 결과 ·····325
<그림 2-355> 공시지가 이상지역 및 에너지 사용량 군집패턴 분석 결과 ·····326
<그림 2-356> 비정형 텍스트정보 기반 서비스 시나리오 ·····327
<그림 2-357> 도봉구 공원의 공간 감성 분석 결과 ·····328
<그림 2-358> 공원 감성분석맵의 격자분석 결과 ·····328
<그림 2-359> 네트워크정보 기반 서비스 시나리오 ·····329

<그림 2-360> 전국 도로별 통행량 및 읍면동별 교통사고건수 분석 결과(2016년 9월) ·····330
<그림 2-361> 전국 도로별 통행량 및 읍면동별 교통사고건수 분석 결과(2016년 9월) ·····331
<그림 2-362> 공간 빅데이터 서비스 플랫폼의 시스템 아키텍처 구성 ·····332
<그림 2-363> 공간 빅데이터 서비스 플랫폼의 저장관리 Layer 구성 ·····333
<그림 2-364> 공간 빅데이터 서비스 플랫폼의 분석Layer 구성 ·····334
<그림 2-365> 공간 빅데이터 서비스 플랫폼의 표출 Layer 구성 ·····335
<그림 2-366> 공간 빅데이터 서비스 플랫폼의 분석Layer 구성 ·····336
<그림 2-367> 공간 빅데이터 서비스 플랫폼의 분석Layer 구성 ·····337
<그림 2-368> 공간 빅데이터 플랫폼 핵심서비스별 주요기능 ·····339
<그림 2-369> 인구이동 시각화 서비스 구성 ·····341
<그림 2-370> 유동인구 서비스 구성 ·····341
<그림 2-371> Open API 구성 ·····342
<그림 2-372> Open API 설계 과정 ·····342
<그림 2-373> Open API 개발환경 구성 및 테스트 데이터셋 구축 ·····343
<그림 2-374> Open API 설계 ·····344
<그림 2-375> 웹페이지를 통한 Open API 레퍼런스 정보 제공 ·····344
<그림 2-376> 공간빅데이터 Open API 서비스 사이트 개발 프로세스 ·····345
<그림 2-377> Open API 서비스 사이트 특징 ·····345
<그림 2-378> Open API 서비스 사이트 소개 페이지 ·····346
<그림 2-379> Open API 서비스 사이트 레퍼런스 지도서비스 페이지 ·····347
<그림 2-380> Open API 서비스 사이트 레퍼런스 데이터서비스 페이지 ·····347
<그림 2-381> 성능 테스트 프로세스 ·····348
<그림 2-382> 성능검증 스크립트 항목별 설명 ·····348
<그림 2-383> Nathan Shedroff의 정보 이해도 ·····350
<그림 2-384> 정보전달 목적 사용자 환경 우선순위 조사 결과 ·····350
<그림 2-385> 의사결정 활용 목적 사용자 환경 우선순위 조사 결과 ·····350
<그림 2-386> 행동변화 유발 목적 사용자 환경 우선순위 조사 결과 ·····351

<그림 2-387> 데이터 파악 목적 사용자 환경 우선순위 조사 결과 ·····351
<그림 2-388> 논리모델 설계 ·····352
<그림 2-389> 물리모델 설계 ·····352
<그림 2-390> 시스템 기능 구조 ·····353
<그림 2-391> 공간 빅데이터 서비스 제공 User Interface 구성 ·····353
<그림 2-392> 메인 페이지 UI/UX 적용 ·····354
<그림 2-393> 시각화 서비스 메뉴 구조 단순화 ·····354
<그림 2-394> 공간분석 프레임 메뉴 UI/UX 개선 ·····355
<그림 2-395> 공공 데이터 수집 스케줄러 Open API 서비스 ·····355
<그림 2-396> 사용성 검증 체크리스트 전체 결과 ·····356
<그림 2-397> 사용성 검증 체크리스트 상세 결과 ·····356
<그림 2-398> 범죄 지도(Crime Map) ·····357
<그림 2-399> 서울시 심야버스노선수립 지원 시스템 (출처 : 한국ESRI) ·····358
<그림 2-400> V World 오픈플랫폼 서비스 활용도 ·····358
<그림 2-401> 공간 빅데이터 실증서비스 서비스 아키텍처 ·····359
<그림 2-402> 공간빅데이터 프로세스 ·····359
<그림 2-403> 공간빅데이터 서비스 컨테이너 ·····360
<그림 2-404> 관리 서비스 구성 ·····360
<그림 2-405> 공간 빅데이터 서비스 통합 기술 구성도 ·····361
<그림 2-406> 공간 빅데이터 공통 데이터베이스 ·····361
<그림 2-407> 공간 빅데이터 서비스 연계 구성 ·····363
<그림 2-408> 공간 빅데이터 실증 시각화 서비스 구성 ·····363
<그림 2-409> 공간 빅데이터 관리 서비스 구성 ·····364
<그림 2-410> 공간 빅데이터 관리 패키지 ·····364
<그림 2-411> 저장, 관리, 분석, 가시화 등 개별 서비스 기술 구성도 ·····365
<그림 2-412> 1세부 연계 구성도 ·····365
<그림 2-413> 2세부 연계 구성도 ·····366

<그림 2-414> 통합 모듈 구성도 ·····366
<그림 2-415> 통합 연계 흐름도 ·····367
<그림 2-416> 통합 연계 기능 구성도 ·····367
<그림 2-417> 공간빅데이터 통합 서비스 최종목표 ·····368
<그림 2-418> 공간빅데이터 서비스 실증 시스템 ·····369
<그림 2-419> 공간 분석 프레임 프로세스 ·····369
<그림 2-420> 공간 연산 기능 구성 화면 ·····370
<그림 2-421> VWORLD 융합 기반 빅데이터 서비스 시스템 개요 ·····371
<그림 2-422> LOD 영상 병렬처리 개념도 ·····371
<그림 2-423> 공공 공간 빅데이터 설계 개념도 ·····372
<그림 2-424> 프로토 타입 모니터링/관리 화면설계 ·····372
<그림 2-425> 시스템 메뉴구조 설계도 ·····373
<그림 2-426> 공공 공간 빅데이터 서비스 및 확장 기술 프로세스 ·····374
<그림 2-427> 공공 공간 빅데이터 서비스 연계 및 확장 기술 개발 ·····374
<그림 2-428> 공공 공간 빅데이터 및 서비스, 분석 기법 ·····375
<그림 2-429> 공공 공간 빅데이터 서비스 확장 ·····375
<그림 2-430> 공공 공간 빅데이터 빅데이터 생성 및 공간 연산 수행 ·····376
<그림 2-431> 공공 공간 빅데이터 서비스 연계 및 확장 ·····376
<그림 2-432> 일원화된 형식의 데이터를 표준화된 결과데이터 생성 ·····377
<그림 2-433> 공용 포맷 데이터 생성 ·····377
<그림 2-434> 공간 빅데이터 서비스 시스템 확장 구축 ·····378
<그림 2-435> 서비스 연계 실증 시각화 ·····378
<그림 2-436> 공간 빅데이터 저장관리 시스템 성능평가(2017년) ·····384
<그림 2-437> 1세부 학술적 성과 ·····386
<그림 2-438> 3세부 언론, 방송 등 홍보성과 ·····394
<그림 2-439> 3세부 전시활동 참가 홍보성과 ·····394
<그림 2-440> 3세부 연구단 홈페이지 구성 ·····395

<그림 2-441> 3세부 웹진 발간 현황 ·····395
<그림 4-1> 성과물 테스트 및 통합활용을 위한 테스트베드 ·····447
<그림 4-2> 사업단 공간분석 라이브러리 ·····447
<그림 4-3> 성과물 배포 및 기술지원 ·····448
<그림 4-4> 택시 승하차 로그 및 유동인구 분석을 통한 지하철역 추천 분석의 연산 플로우 ·····449
<그림 4-5> 택시 승하차 로그 및 유동인구 분석을 통한 지하철역 추천 분석 ·····449
<그림 4-6> 대도시 상업지구내 지역활성화 정도 분석 ·····450
<그림 4-7> 도시쇠퇴현황과 도시재생사업현황 분석 ·····450
<그림 4-8> 공간빅데이터 사업단 플랫폼 적용 개요 ·····451
<그림 4-9> API를 활용한 개발 컴포넌트 제공 예시 ·····456
<그림 4-10> Batch 분석 솔루션 2차년도 초기 모델 ·····457
<그림 4-11> OLAP 엔진과의 Marmot 연계 ·····457
<그림 4-12> '독립 플랫폼 구축형' 사업유형 예시 ·····462
<그림 4-13> '핵심요소기술 활용형' 사업유형 예시 ·····462
<그림 4-14> '공간 빅데이터 서비스 연계형' 사업유형 예시 ·····462

< 표 목차 >

<표 1-1> 연구개발 내용 및 범위 ····· 10

<표 1-2> 연구개발 내용 및 범위 ····· 16

<표 1-3> 연구개발 내용 및 범위 ····· 23

<표 1-4> 연차별 연구개발계획서 대비 연구내용 변동사항(1세부) ····· 28

<표 1-5> 연차별 연구개발계획서 대비 연구내용 변동사항(2세부) ····· 33

<표 1-6> 연차별 연구개발계획서 대비 연구내용 변동사항(3세부) ····· 37

<표 2-1> 핵심연구성과 ····· 41

<표 2-2> 1세부 연구내용 요약 ····· 42

<표 2-3> 공간 빅데이터 저장관리 사용자 요구사항 ····· 44

<표 2-4> 공간 빅데이터 저장관리 시스템 요구사항 ····· 44

<표 2-5> 공간 빅데이터 CRUD 인터페이스 ····· 47

<표 2-6> 공간빅데이터 시스템 성능평가 활용 데이터 ····· 62

<표 2-7> 스크립트 기반 공간 빅데이터 처리 언어 시험 항목 리스트 ····· 68

<표 2-8> MapReduce 기반 워크플로우 가시화도구 시험 항목 리스트 ····· 69

<표 2-9> 공간 빅데이터 저장관리 시스템 실증용 시험 항목 리스트 ····· 71

<표 2-10> 공간 빅데이터 처리 언어용 내부 코드 실행기 시험 항목 리스트 ····· 73

<표 2-11> 대용량 공간 빅데이터 원격 접근 기능 시험 항목 리스트 ····· 74

<표 2-12> 대용량 공간 빅데이터 클라이언트 캐싱 모듈 시험 항목 리스트 ····· 76

<표 2-13> GUI기반 공간빅데이터 응용 개발 도구 시험 항목 리스트 ····· 77

<표 2-14> 공간 빅데이터 질의처리 엔진 사용자 요구사항 ····· 79

<표 2-15> 공간 빅데이터 질의처리 엔진 시스템 요구사항 ····· 80

<표 2-16> 공간 빅데이터 질의언어 문법(일부 예시) ····· 82

<표 2-17> 공간 빅데이터 질의언어(예시) ····· 82

<표 2-18> 인메모리 데이터 처리용 기본 공간 연산자 ····· 122

<표 2-19> 인메모리 기반 공간분석 라이브러리 ····· 128

<표 2-20> 실시간 인메모리 기반 저장 플랫폼 실증 항목 ·····130
<표 2-21> 2세부 연구내용 요약 ·····131
<표 2-22> SCC values between tweet data and housing price indices for each grid size ····140
<표 2-23> 공간분석을 위해 선택된 속성필드 ·····140
<표 2-24> 대상 소셜 미디어 데이터 수집 ·····142
<표 2-25> 대상 소셜 미디어 정보에서 제공되는 위치정보 분석결과 ·····142
<표 2-26> 소셜 미디어 정보에서의 위치정보 변환 기법 ·····143
<표 2-27> 소셜 미디어 정보의 필터링 기준 ·····144
<표 2-28> 정의된 사용자 생성 콘텐츠 ·····146
<표 2-29> 사용자 생성 콘텐츠의 구분 ·····147
<표 2-30> 사용자 생성 콘텐츠에서 추출된 위치 정보의 정확도 분석 결과 ·····147
<표 2-31> 사용자 생성 콘텐츠로부터 정합 기술로 추출된 위치 정보 ·····148
<표 2-32> 공간 감성어 사전 구축 결과 ·····148
<표 2-33> 감성지수 정의 ·····149
<표 2-34> 문서 분류 결과 ·····150
<표 2-35> 추가 감성어 DB 구축 결과 ·····151
<표 2-36> 감성어와 이모티콘이 함께 사용된 경우 감성 판단 ·····151
<표 2-37> 감성지수벡터 및 감성값 정의 ·····152
<표 2-38> 데이터 처리 결과 ·····153
<표 2-39> 탐지된 위치에 대한 좌표정보 추출 결과 일부(뉴스데이터) ·····155
<표 2-40> 장소명 대체어 분석 결과 (서울특별시, 부산광역시 사례) ·····155
<표 2-41> 공공도서관 매력도지수 결과 ·····157
<표 2-42> 정적 방법으로 구축한 장소 감성어 사전 구축 결과(일부) ·····160
<표 2-43> 동적 방법으로 구축한 장소 감성어 사전 구축 결과(일부) ·····160
<표 2-44> 장소 감성어 사전에 대한 정확도 검증 결과 ·····160
<표 2-45> 장소 테마 분석 결과(일부) ·····161
<표 2-46> 일부 reference data에 대한 SVMs와 RVM 성능비교 결과 ·····162

<표 2-47> 텍스트 분류 결과(일부) ·····164
<표 2-48> '관광'과 관련된 뉴스기사의 LDA 결과(일부) ·····165
<표 2-49> 브런치 카테고리 장소 임베딩 결과 코사인 유사도가 가장 높은 장소 상위 5개 ·····166
<표 2-50> POI, 사용자 데이터 수집 ·····168
<표 2-51> 02 플랫폼 레이어 ·····174
<표 2-52> 02 플랫폼의 물리 구성 ·····174
<표 2-53> 02 플랫폼 세부기술 ·····176
<표 2-54> Geometry API for Java ·····180
<표 2-55> 평가지표 ·····180
<표 2-56> 공간분석을 위한 대상 필드 ·····188
<표 2-57> 필터링 기능 정의 ·····189
<표 2-58> R 코딩을 이용한 데이터 필터링 기술 분석 ·····194
<표 2-59> R 코딩을 이용한 위치 추출 기술 분석 (POI 포함) ·····194
<표 2-60> R 코딩을 이용한 위치 추출 기술 분석 (POI 제외) ·····194
<표 2-61> R 코딩을 이용한 좌표 추출 기술 분석 ·····195
<표 2-62> R 코딩을 이용한 장소를 포함한 소셜 미디어의 텍스트 추출 기술 분석 ·····195
<표 2-63> R 코딩을 이용한 감성분석의 기술 분석 ·····196
<표 2-64> R 코딩을 이용한 장소별 언급빈도수 산출의 기술 분석 ·····197
<표 2-65> R 코딩을 이용한 장소별 선호도 산출의 기술 분석 ·····197
<표 2-66> OpenAPI 연동 규격서 ·····198
<표 2-67> 분포 분석 모듈의 테이블 목록 ·····199
<표 2-68> 관광 POI 사전의 테이블 상세 목록 ·····200
<표 2-69> 행정동 단위의 테이블 상세 목록 ·····200
<표 2-70> 시군구 단위의 테이블 상세 목록 ·····200
<표 2-71> 분포 표출용 테이블 상세 목록 ·····201
<표 2-72> 행정동 단위별 중심좌표 기준의 레벨관리 ·····203
<표 2-73> 차트 분석 모듈의 테이블 목록 ·····205

<표 2-74> 관광 POI 사전의 테이블 상세 목록 ·····205
<표 2-75> R&D기술 적용 핵심 현안 ·····216
<표 2-76> 프로그램 설치 목록 및 버전 ·····219
<표 2-77> R 기반 공간빅데이터 분석기능 검토 ·····230
<표 2-78> 공간 빅데이터 비통계 분석 기능 검토 ·····230
<표 2-79> 구현 환경 ·····231
<표 2-80> KOPSS 공간정보 분석 기능 ·····233
<표 2-81> Score model 속도 측정 결과 ·····244
<표 2-82> DBSCAN 컴포넌트 분산처리 속도 측정 결과 ·····246
<표 2-83> SK 지오비전 유동인구 데이터 명세 ·····248
<표 2-84> 15개 공간 빅데이터 Batch 분석 컴포넌트 ·····250
<표 2-85> CNN-LSTM 컴포넌트 분산처리 속도 측정 결과 ·····256
<표 2-86> R 기반 이동객체데이터 분석기능 검토 ·····261
<표 2-87> 이동객체 데이터 비통계 연구 검토 ·····262
<표 2-88> 구현환경 ·····266
<표 2-89> STSS 컴포넌트 분산처리 속도 측정 결과 ·····275
<표 2-90> MapMatching 컴포넌트 분산처리 속도 측정 결과 ·····276
<표 2-91> IFEZ에서 제공하는 데이터 현황 ·····278
<표 2-92> 15개 이동객체데이터 Batch 분석 컴포넌트 ·····280
<표 2-93> 공간 빅데이터 OLAP UI/UX 요구사항 정의 ·····292
<표 2-94> 3세부 연구내용 요약 ·····295
<표 2-95> 공간 빅데이터 실증서비스 시스템과 사업단 시스템의 차별성 ·····302
<표 2-96> 사업단 시스템 지원을 위한 연구단 성과의 단계별 적용 방안 ·····303
<표 2-97> 공간 빅데이터 플랫폼(사업단) 실증 대상 기술 ·····304
<표 2-98> 국내외 빅데이터 및 공간 빅데이터 서비스 사례 분석결과 ·····309
<표 2-99> 미래사회 특성에 따른 공간 빅데이터의 역할 ·····311
<표 2-100> 공간 빅데이터 분석기능 우선순위 도출 ·····313

<표 2-101> 공공/민간부문 비즈니스 모델 개발 ····316
<표 2-102> 토지건물정보 기반 서비스 사용자 분석 ····324
<표 2-103> 토지건물정보 기반 서비스 시나리오 활용데이터 ····325
<표 2-104> 비정형 텍스트정보 기반 서비스 시나리오 활용데이터 ····327
<표 2-105> 네트워크정보 기반 서비스 사용자 분석 ····329
<표 2-106> 네트워크정보 기반 서비스 시나리오 활용데이터 ····330
<표 2-107> 공간 빅데이터 플랫폼 핵심서비스별 주요기능 ····338
<표 2-108> 공간데이터 Open API 적용 사례 ····340
<표 2-109> 서비스별 UI/UX 기술 및 적용사례 ····349
<표 2-110> UI/UX 요구사항 목록 ····351
<표 2-111> 공간 빅데이터 관리 서비스 구조 ····362
<표 2-112> 공간 빅데이터 실증 시각화 서비스 구조 ····362
<표 2-113> 세부 서비스 연계 정의 ····373
<표 2-114> 핵심성과별 최종목표와 연구개발내용 ····379
<표 2-115> 연구개발 성과 종합 ····388
<표 2-116> 연구개발 성과 종합 ····390
<표 2-117> 3세부 기타성과 ····396
<표 2-118> 연구개발 성과 종합 ····397
<표 3-1> 연차별 연구목표 달성도 ····398
<표 3-2> 연차별 연구목표 달성도 ····403
<표 3-3> 연차별 연구목표 달성도 ····413
<표 3-4> 연구의 한계 및 시사점 ····421
<표 3-5> 연구의 한계 및 시사점 ····426
<표 3-6> 연구의 한계 및 시사점 ····433
<표 3-7> 관련 연구분야 기여도 ····438
<표 3-8> 관련 연구분야 기여도 ····441
<표 3-9> 관련 연구분야 기여도 ····443

<표 4-1> 핵심연구성과의 활용계획 ·····451
<표 4-2> 공간 빅데이터 소셜 분석 및 가시화 컴포넌트 적용을 위한 필요사양 ·····456
<표 4-3> 핵심연구성과의 활용계획 ·····458
<표 4-4> 공간 빅데이터 전문인력 양성 및 지속적 홍보 ·····460
<표 4-5> 민간부문을 대상으로 한 지속적인 홍보 방안 수립 ·····461
<표 4-6> 공간 빅데이터 관련 기업 및 지자체 등과 전략적 교육 협력체계 ·····461
<표 4-7> 공간 빅데이터 실증 시스템 활용 사업 유형 및 설명 ·····463
<표 4-8> 핵심연구성과의 활용계획 ·····465
<표 4-9> 추가 연구 및 사업화 방안 ·····467
<표 4-10> 추가 연구 및 사업화 방안 ·····468
<표 4-11> 추가 연구 및 사업화 방안 ·····469

공간 빅데이터 관리, 분석 및 서비스 플랫폼 개요

❑ 공간 빅데이터 관리, 분석 및 서비스 플랫폼

○ 공간 빅데이터 서비스 플랫폼은 공간 빅데이터 저장관리 및 질의 시스템(1세부), 공간 빅데이터 분석 시스템(2세부), 공간 빅데이터 서비스 플랫폼(3세부)을 통하여 원시데이터의 입력/가공/저장/관리/질의/분석을 지원하고, 사용자에게 다양한 맞춤형 공간 빅데이터 서비스를 제공



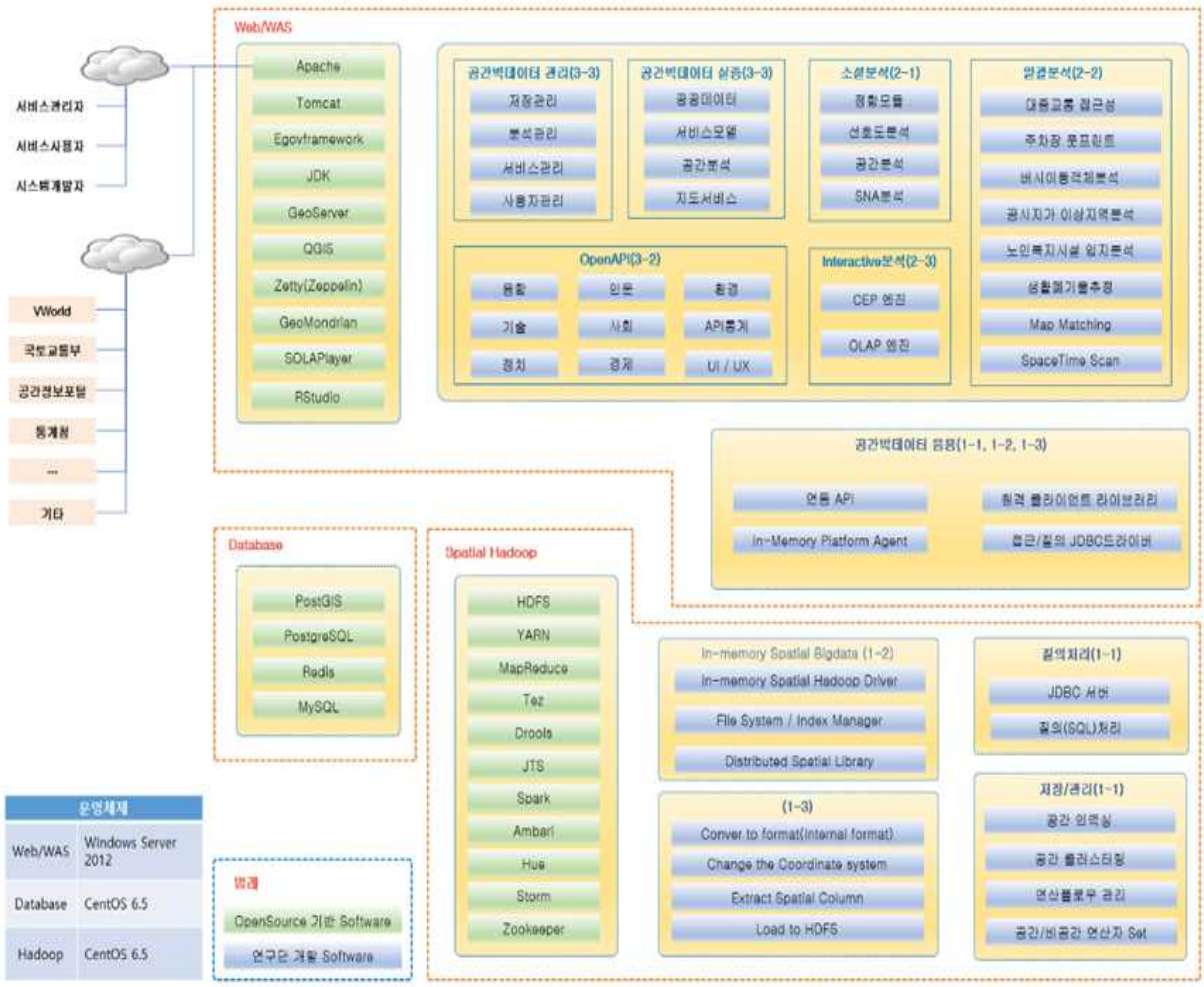
< 공간 빅데이터 관리, 분석 및 서비스 플랫폼의 목표 이미지 >

○ (**기능 구성**) 공간 빅데이터 서비스 플랫폼의 물리적 기능 구성은 크게 DB부문, Spatial Hadoop부문, Web/WAS부문으로 구성, 각 부문의 구현에 활용된 오픈소스 기반 소프트웨어, 연구단 개발 성과 소프트웨어, 운영체제 기반은 다음과 같음

< 공간 빅데이터 서비스 플랫폼의 물리적 기능 구분 및 소프트웨어 명세 >

물리적 구분	연구단 개발 소프트웨어		오픈소스 기반 소프트웨어	운영 체제
	세부구분	모듈 구분		
DataBase	-	-	PostGIS, PostgreSQL, Redis, MySQL	CentOS 6.5
Spatial Hadoop	질의처리(1-1)	JDBC 서버, 질의(SQL)처리	HDFS, YARN, MapReduce, Tez, Drools, JTS, Spark, Ambari, Hue, Storm, Zookeeper	CentOS 6.5
	배치기반 저장/관리(1-1)	공간 인덱싱, 공간 클러스터링, 연산플로우 관리, 공간/비공간 연산자 Set		
	인메모리 기반 저장/관리 (1-2)	In-memory Spatial Hadoop Driver, File System / Index Manager, Distributed Spatial Library		

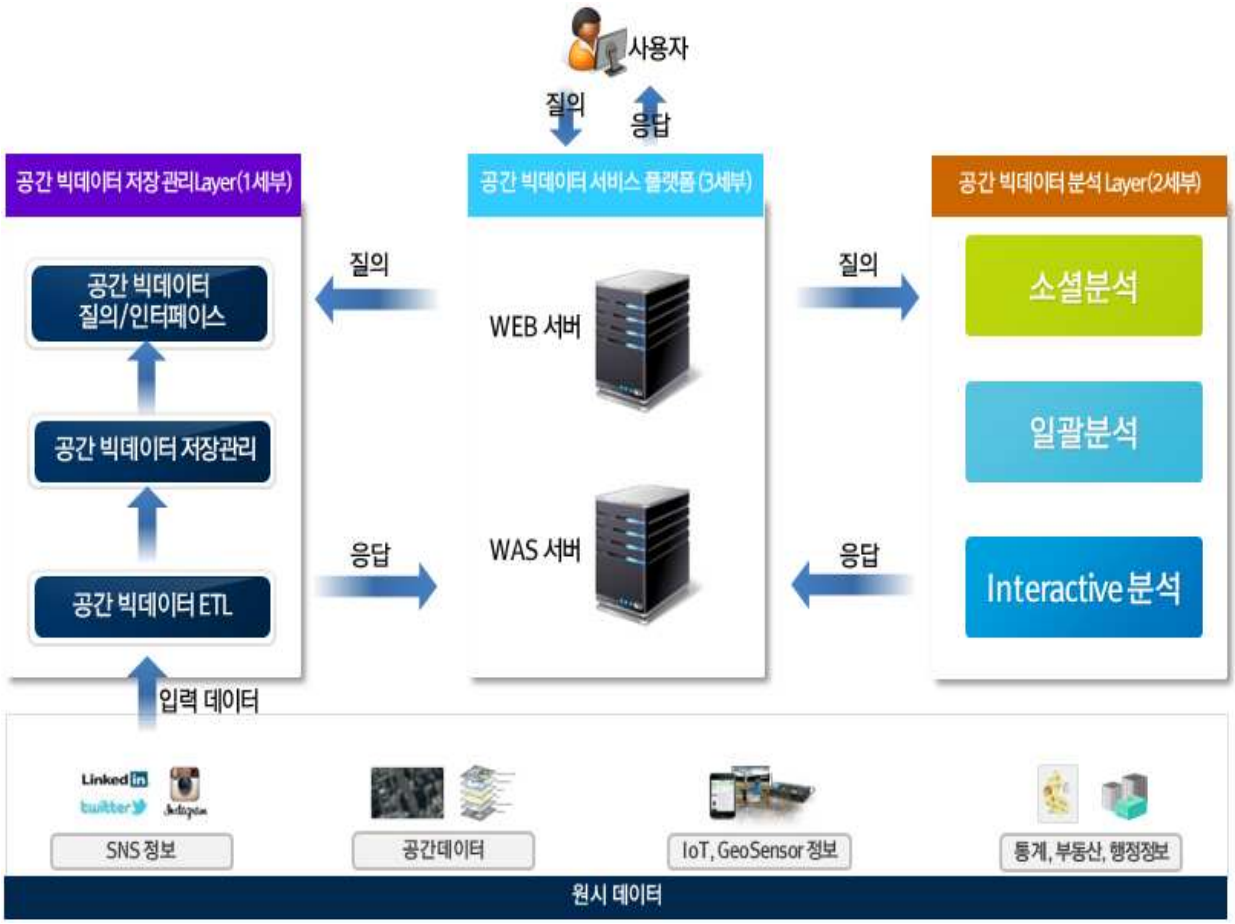
물리적 구분	연구단 개발 소프트웨어		오픈소스 기반 소프트웨어	운영 체제
	세부구분	모듈 구분		
	공간 빅데이터 웨어하우스(1-3)	Convert to format(Internal format), Change the Coordinate system, Extract Spatial Column, Load to HDFS		
Web/WAS	소셜분석(2-1)	정합모듈, 선호도분석, 공간분석, SNA분석	Apache, Tomcat, Egovframework, JDK, GeoServer, QGIS, Zetty(Zeppelin), GeoMondrian, SOLAPlayer, RStudio	Windows Server 2012
	일괄분석(2-2)	대중교통 접근성, 주차장 풋프린트, 이동객체분석, 공시지가 이상지역 분석, 노인복지시설 입지분석, 생활폐기물 추정, Map Matching, SpaceTime Scan 등		
	Interactive분석(2-3)	CEP 엔진, OLAP엔진		
	OpenAPI(3-2)	융합, 인문, 환경, 기술, 사회, API통계, 정치, 경제, UI/UX		
	공간빅데이터 관리(3-3)	저장관리, 분석관리, 서비스관리, 사용자관리		
	공간빅데이터 실증(3-3)	공공데이터, 서비스모델, 공간분석, 지도서비스		



< 공간 빅데이터 관리, 분석 및 서비스 플랫폼의 세부별 모듈 구성도 >

❍ (**세부 간 연계**) 공간 빅데이터 서비스 플랫폼(3세부)은 세부 간 공통 인터페이스를 통하여 공간 빅데이터 저장관리 및 질의 시스템(1세부)과 공간 빅데이터 분석 시스템(2세부)의 개별 모듈들을 호출하며, 분석결과를 받아 서비스 발행, OpenAPI와 UI/UX를 통해 서비스 사용자에게 최종 전달됨

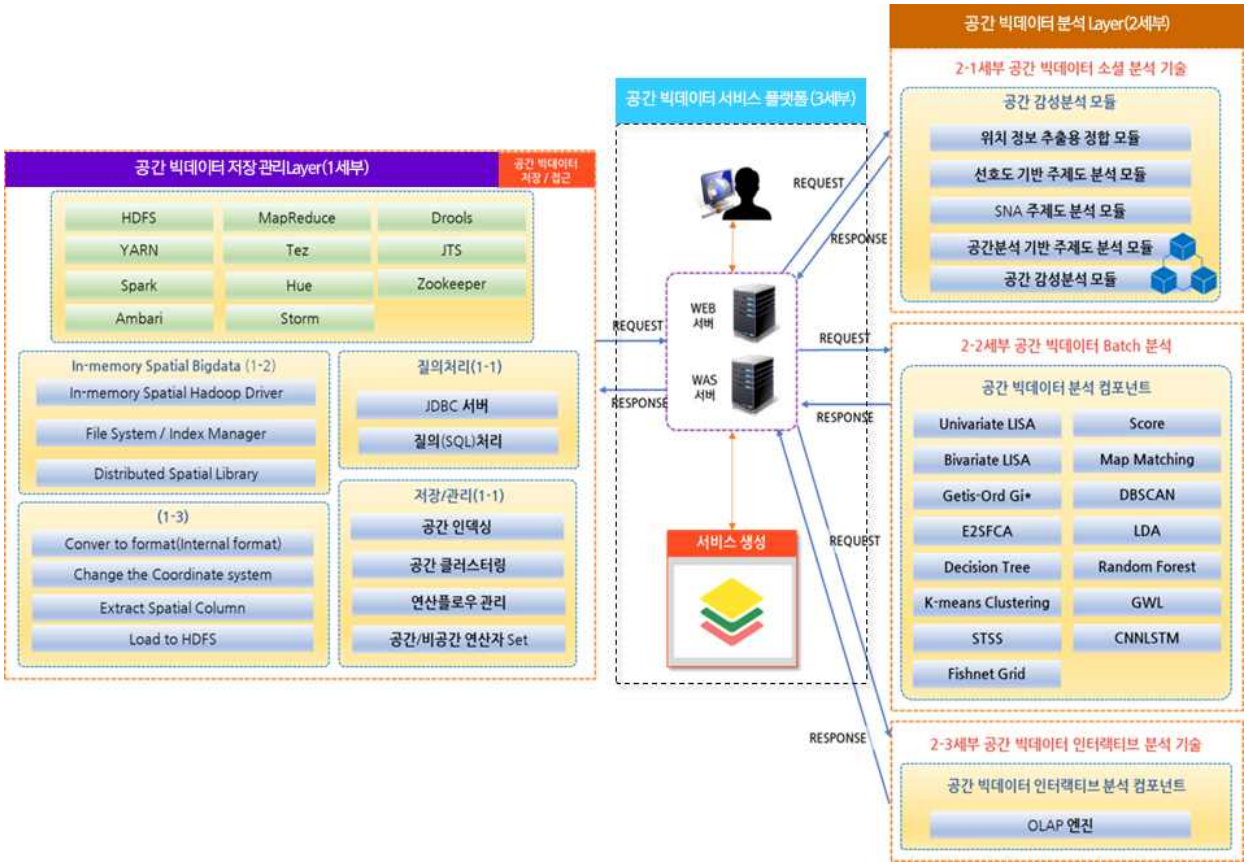
- (**공간 빅데이터 저장관리 및 질의 시스템**) 공간 빅데이터 원시데이터를 공간 하둡 에코시스템에 입력 데이터로 적재, 적재된 데이터의 전처리 가공/저장/관리, 데이터들을 질의/인터페이스를 통하여 공간 빅데이터 서비스 플랫폼에 전달하는 기능을 제공
- (**공간 빅데이터 분석 시스템**) 공간 빅데이터 저장관리 및 질의 시스템으로부터 질의를 통하여 응답받은 결과데이터를 입력데이터로 활용하여 소셜분석, 일괄분석, Interactive 분석을 수행하고 그 결과 데이터를 공간 빅데이터 서비스 플랫폼에 제공
- (**공간 빅데이터 서비스 플랫폼**) 공간 빅데이터 저장관리 및 질의 시스템과 공간 빅데이터 분석 시스템에서 질의 및 응답받은 결과데이터들을 융·복합하여 다양한 사용자의 수요를 충족시키는 맞춤형 서비스를 제공



< 공간 빅데이터 관리, 분석 및 서비스 플랫폼의 세부 간 연계 >

○ (**세부 모듈 간 연계**) 공간 빅데이터 저장관리 및 질의 시스템(1세부)의 공간 빅데이터 저장/접근, 공간 연산 기능과 소셜분석(2-1세부), 일괄분석(2-2세부), Interactive 분석(2-3세부) 기술을 통합적으로 연계

- 모듈단위로 탑재한 세부연계 기술을 효율적으로 사용하고 분석할 수 있는 공간 분석프레임 설계
- 2-1세세부의 공간 분석 모듈을 이용한 특정 지역의 감성 분석 처리 후, 1세부 공간 연산을 통해 격자 형식의 결과를 도출하는 등 일련의 시나리오별 처리 프로세스 기술 개발



< 공간 빅데이터 관리, 분석 및 서비스 플랫폼의 세부 모듈 간 연계 구성도 >

1. 연구개발과제의 개요

가. 추진 배경 및 목적

(1) 추진 배경

○ 4차 산업혁명 등 산업 환경의 변화에 따라 모든 것이 연결되어 데이터 수집, 분석, 처리가 고도화되고, 기계가 지능을 가지고 스스로 학습하면서 새로운 가치를 창출하는 방향으로 전환되고 있음

- 4차 산업혁명에서 빅데이터 기술개발의 중요성이 지속적으로 강조되고 있으며, 빅데이터 중에서도 데이터 형식의 복잡성과 대용량 처리를 위한 기술수요가 급증할 것으로 예상되는 공간 빅데이터(Geospatial Big Data) 처리 기술에 대한 관심이 급증하고 있음



<그림 1-1> 연구의 추진배경

❍ IBM, Google, Oracle, HP 등 글로벌 기업들은 막대한 투자를 기반으로 공간 빅데이터 시장을 독점하고 있으며, 이로 인하여 글로벌 기업중심의 빅데이터 생태계가 지속적으로 확장되고 있음

- GeoWeb, SensorWeb, GeoBI 등과 같이 공간 정보와 다른 정보들이 융합된 빅데이터 시장이 급성장하고 있으며, 외산 GIS Tool을 중심으로 기술시장이 형성되고 있음

(2) 추진 목적 및 연구의 필요성

❍ 공간 빅데이터는 국가 공공정보의 근간이 되는 핵심정보로서, 다양한 분야와의 융합 활용이 가능한 기본 빅데이터이며, 기존 방식으로 저장, 관리, 분석할 수 있는 범위를 초과하는 대규모의 데이터를 구성하고 있음

❍ 현재 국토교통부는 4차 산업혁명의 산업 환경변화에 대응하기 위하여 공간정보를 7대 신산업 육성 분야 중 하나로 선정

❍ 특별히 공간 빅데이터의 중요성을 인식하고 공공 및 민간부문의 효율적인 공간 빅데이터 활용을 지원하여 국내 공간 빅데이터 기술 개발과 관련 산업의 활성화를 촉진할 필요가 있음



<그림 1-2> 빅데이터와 공간정보 기술의 융합 필요성

❍ 공간 빅데이터의 생산과 활용이 사회 전반에 걸쳐 빠르게 증가하고 있는 가운데 볼륨, 복잡도 및 다양성이 높은 공간 데이터의 효율적인 저장관리 방법의 필요성이 대두되고 있으며 다양한 레거시 공간 데이터와 실시간 빅데이터의 연동 및 처리 방식이 필요하게 됨

- 본 연구는 이를 해결하기 위하여 빅데이터 플랫폼에서 공간 빅데이터(Geospatial Big Data)를 효율적으로 저장, 관리, 질의, 분석하고 서비스를 제공하기 위한 핵심기술과 이를 구현한 공간 빅데이터 저장관리 플랫폼을 개발함

공간 빅데이터

본 연구에서의 공간 빅데이터는 기존 공간 정보시스템에서 다루던 대용량 공간 데이터(수치지도, 지적도, 지형도 등) 뿐만 아니라 공간 정보를 포함한 빅데이터(위치기반 SNS, GeoSensor 정보, 부동산 거래 정보, 버스 카드 리더기 정보 등)로 정의함

❍ 빅데이터와 공간정보를 융합한 공간 빅데이터 관련 기술 필요

- 빅데이터 플랫폼 시장의 글로벌 기업 시장선점으로 인해 외산제품에 대한 의존도가 높아지고 있으며, 빅데이터와 위치정보를 기반으로 한 공간정보 빅데이터 융합 시장이 급성장하고 있음

※ 빅데이터 활용 시 IBM, Google, Oracle, HP, EMC, SAP, Teradata 등의 글로벌 기업의 외산 소프트웨어가 80%이상을 차지

- 기존 빅데이터와 공간정보를 융합한 공간 빅데이터와 관련한 기술개발은 초기단계임

❍ Hadoop 프레임워크를 확장한 “Spatial Hadoop” 프레임워크가 요구됨

- Hadoop 프레임워크는 빅데이터의 병렬분산처리를 지원하고 있는 범용 프레임워크이나, 기본적으로 텍스트 정보 중심의 정보처리에 적합한 구조와 기능을 제공함

※ 하둡(Hadoop): 방대한 양의 자료를 여러 대의 컴퓨터로 된 클러스터에서 처리할 수 있는 분산 응용프로그램을 지원하는 소프트웨어 프레임워크로, 최근 빅데이터 처리 방법으로 각광받고 있음

- 반면 2차원 이상의 좌표로 된 vector 공간정보로 된 빅데이터 처리에 있어서는 공간데이터의 특성을 고려하고 있지 않아 적합하지 않으며, 비공간 데이터를 처리하는 것과 같은 방법으로 처리할 경우 성능이 저하됨

- 기존 Hadoop 솔루션들로는 처리할 수 없거나 비효율적으로 처리하는 공간정보 처리에 특화된 적재, 공간색인, 공간 연산 및 분석 기능 등 기능을 추가하여 기존 Hadoop 시스템을 확장함으로써 공간정보 처리 성능 향상을 위하여 최적화된 “Spatial Hadoop” 프레임워크의 필요성이 제시되고 있음

- 이러한 기능 확장은 단순히 Hadoop 시스템에 공간정보를 처리하는 응용 수준의 layer를 올리는 방식으로는 효과적인 구현이 불가능하며, 기존 Hadoop 시스템의 전 layer에 걸쳐 공간정보 처리 기능을 수정 및 확장하는 기술이 필요함

❍ 공간 빅데이터 관련 국내 기술 경쟁력의 신속한 확보 필요

- IoT(사물인터넷)와 센서기술의 발전으로 인해 실시간으로 입력되는 빅데이터가 증가추세이며 이중 80% 이상이 공간정보를 포함하고 있음
- 이에 따라 공간데이터와 스트림 데이터를 모두 처리할 수 있는 공간 빅데이터 처리기술에 대한 수요가 급증하고 있으며 핵심기술 확보가 시급함
- 현재 빅데이터 관련 원천기술은 국외 글로벌 기업들 주도로 빠르게 발전하는 반면 국내 공간 빅데이터 관련 기술력은 매우 미약한 실정이므로 기술 경쟁력 확보를 위하여 기술개발에 대한 투자가 시급함
- 공간 빅데이터와 관련된 적극적인 기술개발을 통해 해외 의존도 저감과 새로운 공간 빅데이터 시장에서의 기술적 우위를 통한 시장 점유율 제고가 필요함

❍ 국토교통부 중심으로 확대되고 있는 공공 및 민간부문 공간 빅데이터 활용 사업 지원

- 대부분 공간정보를 포함하는 공공의 통계, 행정정보, 민간의 교통, 유동인구정보 등을 융합하여 새로운 서비스와 가치를 창출할 수 있음
- 국토·도시정책 수립과 집행에 필수적인 공간빅데이터 통합분석활용체계의 기술적 핵심임

나. 연구개발 목표

(1) 최종 목표

○ 본 연구는 “저비용, 고성능, 고확장성의 공간 빅데이터(Geospatial Big Data)의 저장, 관리, 분석 및 서비스 제공을 위한 핵심기술과 이를 구현한 공간 빅데이터 관리, 분석 및 서비스 플랫폼 개발” 을 최종 목표로 하고 있음

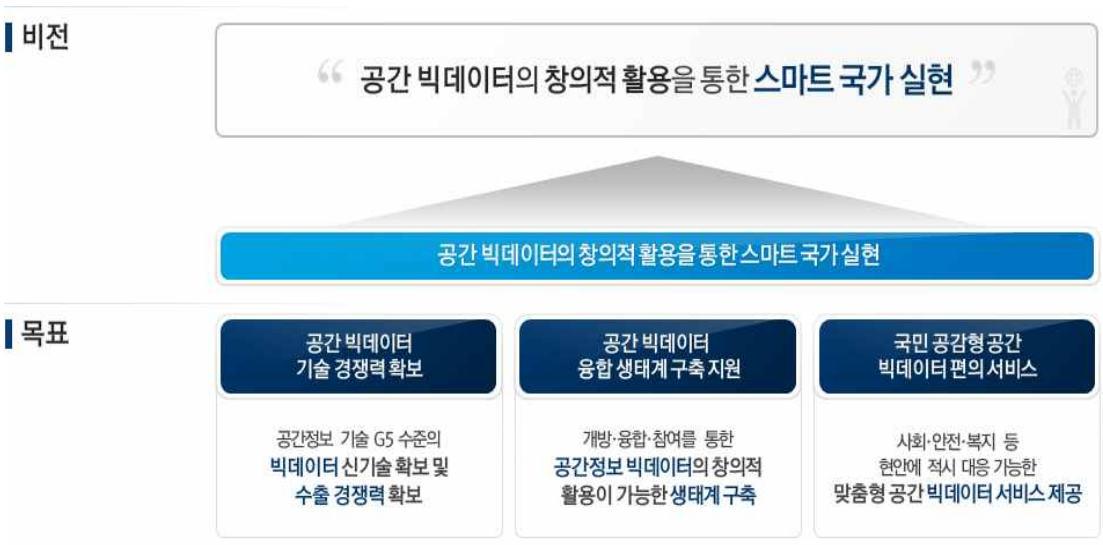
- 본 연구에서의 공간 빅데이터는 기존 공간 정보시스템에서 다루던 대용량 공간 데이터(수치지도, 지적도, 지형도 등) 뿐만 아니라 공간 정보를 포함한 빅데이터(위치 기반 SNS, GeoSensor 정보, 부동산 거래 정보, 버스 카드 리더기 정보 등)로 정의함

(2) 중점 추진 내용

○ 본 연구의 중점 추진 내용은 공간 빅데이터 저장관리 인프라 기술 개발, 공간 빅데이터 분석 및 가시화 기술 개발, 공간 빅데이터 서비스 활성화 및 실증 연구 3부문으로 구성

○ 공간 빅데이터 저장관리 인프라 기술 개발

- 공간 빅데이터 저장관리 인프라 기술 개발
- 공간 빅데이터 웨어하우스 핵심기술 개발
- 인메모리 기반 공간 빅데이터 저장관리 프레임워크 개발



<그림 1-3> 연구개발 최종목표

❍ 공간 빅데이터 분석 및 가시화 기술 개발

- 공간 빅데이터 소셜 분석 및 가시화 컴포넌트 기술 개발
- 공간 빅데이터 일괄분석 및 컴포넌트 기술 개발
- 이동객체 데이터 일괄분석 및 컴포넌트 기술 개발
- 공간 빅데이터 CEP 및 OLAP 엔진 개발

❍ 공간 빅데이터 서비스 활성화 및 실증 연구

- 공간 빅데이터 플랫폼의 융복합 활용방안 개발
- 공간 빅데이터의 공공/민간부문 핵심 서비스 모델 도출 및 제공방안 개발
- 공간 빅데이터 확장 OpenAPI 서비스 및 UI/UX 개발
- 공공시스템 기반 국토교통분야 및 타부처 공간 빅데이터 실증 서비스 시스템 개발

중점추진 내용



<그림 1-4> 중점추진 내용

다. 연구개발의 내용 및 범위

(1) 공간 빅데이터 저장관리 인프라기술 개발

(가) 주요연구내용 및 구조

○ **공간 빅데이터의 수집, 저장, 관리 및 질의 기능 제공을 위한 핵심기술과 이를 구현한 고성능, 고확장성의 공간 빅데이터 저장관리 시스템 개발**

○ **(1-1 세세부)** 공간 빅데이터 저장관리 인프라 기술 개발

- Hadoop 환경에서 공간 빅데이터를 HDFS에 저장관리하는 기술, 공간/비공간 데이터를 처리하는 핵심 연산, 공간데이터의 빠른 접근을 위한 공간색인, 질의언어 처리 기술 및 이들 기능을 외부에서 활용할 수 있도록 하는 응용 인터페이스 개발

○ **(1-2 세세부)** 인메모리 기반 공간 빅데이터 저장관리 프레임워크 개발

- 공간 빅데이터를 분산환경에 효율적으로 저장하고 메모리 기반으로 고속 접근, 처리 및 분석할 수 있는 핵심기술 및 이를 구현한 실시간 공간 빅데이터 플랫폼 개발

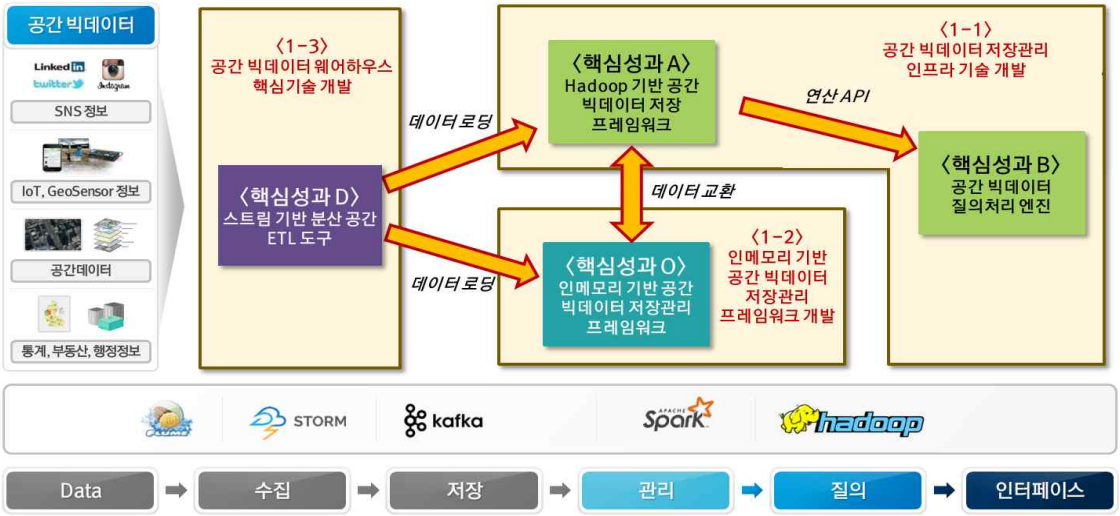
○ **(1-3 세세부)** 공간 빅데이터 웨어하우스 핵심기술 개발

- 다양한 형태를 가지는 공간데이터의 신속, 정확한 추출 및 변환, 적재 기능을 제공하는 공간 스트림 빅데이터의 병렬분산처리 모듈 및 공간 ETL 도구 개발



<그림 1-5> 공간 빅데이터 저장관리 인프라기술 개발의 주요연구내용 및 구조

※ ETL(Extract, Transformation, Load) : 시스템간 데이터를 이동·처리하는 형태로, 동일 또는 타 형태의 데이터 소스로부터 데이터를 추출하고, 조회 또는 분석을 목적으로 적절한 포맷이나 구조로 데이터를 저장하기 위해 데이터를 변환하고, 최종 대상(저장소)으로 변환 데이터를 적재하는 것을 말함



<그림 1-6> 1세부의 세세부별 연구수행 내용 및 구조

❍ 성과물의 활용

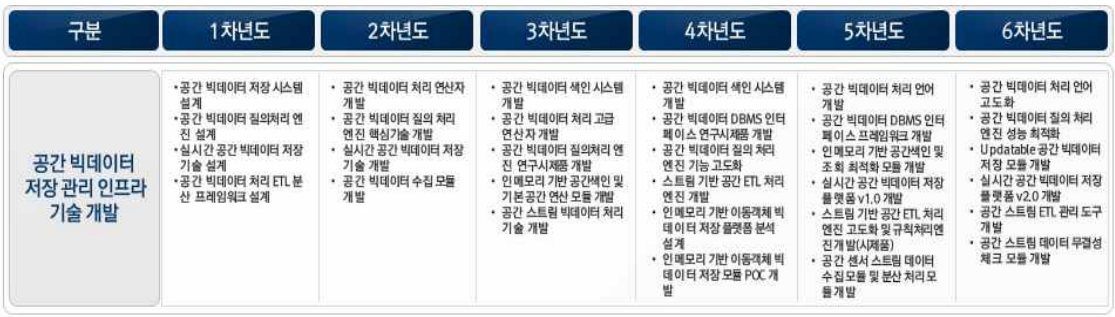
- 1세부 성과물의 사용자는 주로 “공간분석 응용 개발자”로서 본 시스템에서 제공하는 API 및 인터페이스를 통하여 데이터를 사용하여 상위의 분석 및 서비스를 개발함

※ 본 시스템은 전체 구조에서 DBMS와 유사한 역할을 함

- 성과물의 활용은 특정 분야의 응용으로 범위가 한정되지는 않으며 공간 빅데이터를 필요로 하는 또는 활용하고자 하는 응용개발자는 모두 성과물을 사용할 수 있음

- 성과물을 활용하는 사용자는 다음과 같이 구분할 수 있음
  - 공간빅데이터 플랫폼을 사용하는 일반 사용자
  - 공간분석 시스템을 고도화하려는 개발자
  - 정부 및 공공기관, 민간분야의 공간정보 비전문가

(나) 연차별 연구내용



구분	1차년도	2차년도	3차년도	4차년도	5차년도	6차년도
공간 빅데이터 저장 관리 인프라 기술 개발	• 공간 빅데이터 저장 시스템 설계 • 공간 빅데이터 질의처리 엔진 설계 • 실시간 공간 빅데이터 저장 기술 설계 • 공간 빅데이터 처리 ETL 분산 프레임워크 설계	• 공간 빅데이터 처리 연산자 개발 • 공간 빅데이터 질의 처리 엔진 핵심기술 개발 • 실시간 공간 빅데이터 저장 기술 개발 • 공간 빅데이터 수집 모듈 개발	• 공간 빅데이터 색인 시스템 개발 • 공간 빅데이터 처리 고급 연산자 개발 • 공간 빅데이터 질의처리 엔진 연구시제품 개발 • 인메모리 기반 공간색인 및 기본공간 연산 모듈 개발 • 공간 스트림 빅데이터 처리 기술 개발	• 공간 빅데이터 색인 시스템 개발 • 공간 빅데이터 DBMS 인터페이스 연구시제품 개발 • 공간 빅데이터 질의 처리 엔진 기능 고도화 • 스트림 기반 공간 ETL 처리 엔진 개발 • 인메모리 기반 이동객체 빅데이터 저장 플랫폼 분석 설계 • 인메모리 기반 이동객체 빅데이터 저장 모듈 POC 개발	• 공간 빅데이터 처리 언어 개발 • 공간 빅데이터 DBMS 인터페이스 프레임워크 개발 • 인메모리 기반 공간색인 및 조회 최적화 모듈 개발 • 실시간 공간 빅데이터 저장 플랫폼 v1.0 개발 • 스트림 기반 공간 ETL 처리 엔진 고도화 및 규칙처리엔진 개발(시제품) • 공간 센서 스트림 데이터 수집모듈 및 분산 처리모듈 개발	• 공간 빅데이터 처리 언어 고도화 • 공간 빅데이터 질의 처리 엔진 성능 최적화 • Updatable 공간 빅데이터 저장 모듈 개발 • 실시간 공간 빅데이터 저장 플랫폼 v2.0 개발 • 공간 스트림 ETL 관리 도구 개발 • 공간 스트림 데이터 무결성 체크 모듈 개발

<그림 1-7> 연차별 연구내용

○ 1차년도에는 공간 빅데이터 저장관리 인프라 기술 개발에 필요한 관련 연구 조사 및 요구사항과 주요 핵심 기능 도출, 시스템 전체의 구조에 대한 설계를 수행함

○ 2차년도에는 1차년도에 도출된 설계안을 기반으로 공간 빅데이터 저장관리 인프라의 주요 핵심 요소 기술을 개발함

○ 3차년도에는 2차년도에 개발된 핵심 요소 기술을 적용하여 공간 빅데이터 저장관리 인프라의 프로토타입을 개발함

○ 4차년도에는 핵심 모듈의 대부분을 개발하고 결과물 중 일부를 타 세부과제에 공유하여 사용할 수 있도록 하고 ‘공간빅데이터체계구축사업’에 활용 검증용으로 1차 시범적용하여 결과물에 대한 사용 가능성을 확인함

○ 5차년도는 개발된 결과물을 실제 현장에 적용하기 위하여 ‘공간빅데이터체계구축사업’에서 운영 중인 기존 플랫폼과 연동 작업을 통하여 결과물의 현장 적용을 지원하고, 운영 플랫폼 환경에서의 연구 결과물의 최적화, 안정화 작업을 수행함

○ 6차년도에는 연구 결과물의 최종 실증 및 적용을 위해 타 세부 과제물들과의 통합 및 테스트를 실시하고, 이를 현장에 적용하여 이에 따른 피드백을 바탕으로 시스템을 안정화 및 고도화함

○ 연차별 세부 연구내용 및 범위는 <표 1-1>과 같음

<표 1-1> 연구개발 내용 및 범위

구분		세부과제명	세부연구 목표	연구개발 수행내용	수행기관
연도	세세부				
1차 년도	1-1	공간 빅데이터 저장관리 인프라 기술 개발	• 공간 빅데이터 저장 시스템 요구사항 분석 및 설계	• 공간 빅데이터 저장 시스템 아키텍처 도출 • 공간 빅데이터 모델 정의 • 공간 빅데이터 CRUD(Create, Read, Update, Delete) 인터페이스 개발 • 공간 빅데이터 카탈로그 관리 인터페이스 개발 • 공간 빅데이터 수집 인터페이스 개발	한국전자통신연구원
			• 공간 빅데이터 저장 시스템 구축 (Proof-of-Concept 버전)	• 공간 빅데이터 저장 및 단순 탐색 기능 개발	
			• 공간 빅데이터 질의 처리 엔진 개발 (Proof-of-Concept 버전)	• 공간 빅데이터 질의 처리 엔진 아키텍처 설계 • 공간 빅데이터 기본 연산자 정의 • 공간 빅데이터 질의 언어 정의 • 공간 빅데이터 질의 언어 파싱 및 기본검색 기능 개발	
			• 공간 빅데이터 DBMS 인터페이스 설계 및 개발(Proof-of-Concept 버전)	• 공간 빅데이터 DBMS ODBC 인터페이스 설계 • 공간 빅데이터 DBMS JDBC 인터페이스 설계 • 공간 빅데이터 Map/Reduce 인터페이스 설계 • 공간 빅데이터 DBMS 인터페이스 기본 기능 개발	
			• 공간 빅데이터 웨어하우징 기술 분석/설계	• 수집대상 데이터 및 수집방법 분석 • 공간 빅데이터 웨어하우스 관리 기술 분석/설계	
			• 정형/반정형/비정형 시공간 데이터 및 서비스 연계기술 분석/설계	• 공간/소셜 빅데이터 웨어하우스 사용자 인터페이스 분석/설계	
	1-2	인메모리 기반 공간 빅데이터 저장관리 프레임워크 개발	• 실시간 공간 빅데이터 저장 플랫폼 분석 및 설계	• 실시간 공간 빅데이터 저장 기술 설계 • 실시간 공간 빅데이터 처리 플랫폼 아키텍처 설계	포스웨이브
	1-3	공간 빅데이터 웨어하우스 핵심기술 개발	• 공간 빅데이터 DBMS 인터페이스 설계 및 개발(Proof- Of-Concept 버전)	• 공간 빅데이터 DBMS ODBC 인터페이스 설계 • 공간 빅데이터 DBMS JDBC 인터페이스 설계 • 공간 빅데이터 Map/Reduce 인터페이스 설계 • 공간 빅데이터 DBMS 인터페이스 기본 기능 개발	아이씨티웨이

구분		세부과제명	세부연구 목표	연구개발 수행내용	수행기관
연도	세세부				
2차 년도	1-1	공간 빅데이터 저장관리 인프라 기술 개발	• 공간 빅데이터 처리 연산자 개발	• 기본 공간 연산자(basic spatial operators) 개발 • 비공간 연산자(non-spatial operators) 개발	한국전자통신연구원
			• 공간 빅데이터 질의 처리 엔진 핵심기술 개발	• 공간 빅데이터 질의 플랜 생성 모듈 개발 • 스트림 공간데이터 질의 언어 파싱 및 의미검증 모듈 개발 • 스트림 공간데이터 질의처리 모듈(기본기능) 개발	
	1-2	인메모리 기반 공간 빅데이터 저장관리 프레임워크 개발	• 실시간 공간 빅데이터 저장 연구시제품 개발	• 실시간 데이터 처리를 위한 인메모리 데이터 처리 모듈 개발	포스웨이브
	1-3	공간 빅데이터 웨어하우스 핵심기술 개발	• 공간 빅데이터 수집 모듈 개발	• 다중 공간 빅데이터 수집 모듈 개발 • 다중 공간 빅데이터 적재 모듈 개발	아이씨티웨이
			• 공간 스트림 빅데이터 수집 모듈 개발	• 공간 스트림 빅데이터 수집 모듈 개발(단일 노드) • 공간 스트림 빅데이터 적재 모듈 개발(단일 노드)	
3차 년도	1-1	공간 빅데이터 저장관리 인프라 기술 개발	• 공간 빅데이터 색인 시스템 개발	• 공간 빅데이터 색인 개발(1차)	한국전자통신연구원
			• 공간 빅데이터 처리 고급 연산자 개발	• 고급 공간 연산자(advanced spatial operators) 개발(2차) • 공간 조인 연산자(spatial join operators) 개발	
			• 공간 빅데이터 질의 처리 엔진 연구시제품 개발	• 공간 빅데이터 입력 스트림 관리 모듈 구현 • 스트림 공간 빅데이터 분산 질의처리 모듈 개발	
	1-2	인메모리 기반 공간 빅데이터 저장관리 프레임워크 개발	• 인메모리 기반 공간색인 및 기본 공간 연산 모듈 개발	• 인메모리 기반 공간색인 모듈 개발 • 인메모리 데이터 처리를 위한 기본 공간 연산자 개발	포스웨이브
	1-3	공간 빅데이터 웨어하우스 핵심기술 개발	• 공간 스트림 빅데이터 처리 기술 개발	• 공간 스트림 빅데이터의 병렬/분산 처리 모듈 개발	아이씨티웨이
4차 년도	1-1	공간 빅데이터 저장관리 인프라 기술 개발	• 공간 빅데이터 색인 시스템 개발	• 공간 빅데이터 색인 개발(2차) • MapReduce를 활용한 고속 공간 빅데이터 색인 구축 기술 개발	한국전자통신연구원
			• 공간 빅데이터 DBMS 인터페이스 연구시제품 개발	• 공간 빅데이터 DBMS ODBC 클라이언트/서버 모듈 개발 • 공간 빅데이터 DBMS JDBC 클라이언트/서버 모듈 개발	
			• 공간 빅데이터 질의 처리 엔진 기능 고도화	• 스트림 공간 빅데이터 질의 스케줄링 모듈 개발 • 스트림 공간 빅데이터 이벤트 처리 모듈 개발	

구분		세부과제명	세부연구 목표	연구개발 수행내용	수행기관
연도	세세부				
	1-2	인메모리 기반 공간 빅데이터 저장관리 프레임워크 개발	• 인메모리 기반 이동객체 빅데이터 저장 플랫폼 설계 및 개발 및 POC 개발	• 인메모리 기반 이동객체 빅데이터 저장 플랫폼 분석 설계 • 인메모리 기반 이동객체 빅데이터 저장 모듈 POC 개발	포스웨이브
	1-3	공간 빅데이터 웨어하우스 핵심기술 개발	• 스트림 기반 공간 ETL도구 핵심 처리 엔진 기술 개발	• 스트림 기반 공간 ETL 처리 엔진 개발 • 스트림 기반 공간 ETL 처리 엔진 연동 POC 개발	아이씨티웨이
5차년도	1-1	공간 빅데이터 저장관리 인프라 기술 개발	• 공간 빅데이터 처리 언어 개발	• 공간 빅데이터 처리 언어 문법 설계 • 공간 빅데이터 저장 시스템 고도화	한국전자통신연구원
			• 공간 빅데이터 저장 시스템 고도화	• 색인 기반 고속 공간 조인 연산자 개발 • 고속 데이터 입력을 위한 지연 색인 갱신 기술 개발	
			• 원격 공간 빅데이터 활용 기술	• 클라이언트-서버 기반 원격 공간 빅데이터 접근 모듈 개발	
			• 공간 빅데이터 저장 관리 인프라 시범 서비스 적용 지원(1)	• 공간 빅데이터 저장관리 프레임워크 통합	
			• 공간 빅데이터 분석 응용 개발 도구(1)	• GUI 기반 공간 빅데이터 응용 개발 도구 개발	
			• 공간 빅데이터 DBMS 인터페이스 프레임워크 개발	• 스트림 공간 빅데이터 인터페이스 클라이언트/서버 모듈 개발 • 스트림 공간 빅데이터 인터페이스 통합처리 서버 프레임워크 개발	
	1-2	인메모리 기반 공간 빅데이터 저장관리 프레임워크 개발	• 인메모리 기반 이동객체 빅데이터 저장 플랫폼 개발	• 이동객체 빅데이터 저장 모듈 개발 • 이동객체 빅데이터 공간색인 모듈 개발	포스웨이브
			• 실시간 공간 빅데이터 저장 플랫폼 고도화 (1차)	• 인메모리기반 공간색인 및 조회 최적화 • 실시간 공간 빅데이터 저장 플랫폼 v1.0	
			• 실시간 공간 빅데이터 저장 플랫폼 고도화 (2차)	• 실시간 공간빅데이터 저장 플랫폼 V2.0(1)	
	1-3	공간 빅데이터 웨어하우스 핵심기술 개발	• 공간 스트림 데이터 처리 기술 고도화	• 공간 센서 스트림 데이터 수집 및 분산 처리 모듈 개발	아이씨티웨이
			• 스트림 기반 공간 ETL 처리 엔진 고도화	• 스트림 기반 공간 ETL 처리 엔진 연동 API 개발 • 스트림 기반 공간 ETL 규칙 처리 엔진 개발(POC 수준) • 스트림 기반 공간 ETL 규칙 처리 엔진 개발(2차배포판)	

구분		세부과제명	세부연구 목표	연구개발 수행내용	수행 기관
연도	세세부				
6차 년도	1-1	공간 빅데이터 저장관리 인프라 기술 개발	• 공간 빅데이터 처리 언어 개발(2)	• 내부 코드 실행기 개발	한국 전자 통신 연구 원
			• 공간 빅데이터 저장관리 시스템 고도화	• Map/Reduce 기반 워크플로우 가시화 도구 개발 • 대용량 공간 빅데이터 원격 접근 기능 개발 • 대용량 공간 빅데이터 클라이언트 캐싱 모듈 개발	
			• 공간 빅데이터 저장관리 인프라 시범 서비스 실증 지원(2)	• 공간 빅데이터 사업단 저장 시스템 연동 모듈 개발 • 시범 서비스 공간 빅데이터 연산자 및 어댑터 개발 • 시범 서비스 실증 통합 테스트	
			• 공간 빅데이터 분석 응용 개발 도구(2)	• GUI 기반 공간 빅데이터 응용 개발 도구 고도화	
			• 공간 빅데이터 질의 처리 엔진 성능 고도화	• 스트림 공간 빅데이터 분산 질의처리 최적화 모듈 개발	
	1-2	인메모리 기반 공간 빅데이터 저장관리 프레임워크 개발	• 실시간 공간 빅데이터 저장 플랫폼 고도화 (2차)	• 실시간 공간 빅데이터 저장 플랫폼 v2.0 (2) • 실시간 인메모리 기반 저장 플랫폼 실증	포스 웨이 브
	1-3	공간 빅데이터 웨어하우스 핵심기술 개발	• 스트림 기반 공간 ETL 처리 엔진 고도화	• 스트림 기반 공간 ETL 처리 엔진 고도화 및 연구 시제품 개발	아이 씨티 웨이
			• 스트림 기반 공간 ETL 처리 엔진 관리 도구 개발	• 공간 스트림 ETL 관리 도구 개발 • 공간 연산 워크플로우 분석도구 개발	

(2) 공간 빅데이터 분석 및 가시화 기술 개발

(가) 주요연구내용 및 구조

○ 공간 빅데이터 분석 프레임워크, Batch 분석, 이동객체 Batch 분석, Interactive 분석 기술 등 공간 빅데이터 분석 핵심기술과 컴포넌트 개발

○ **(2-1세세부)** 공간 빅데이터 소셜 분석 및 가시화 컴포넌트 기술 개발

- 서울대학교 산학협력단은 사용자가 생성한 비정형 데이터(트위터, 네이버블로그, 다음 아고라, 뉴스기사, 블로그, 트립어드바이저, 망고플레이트, 구글리뷰, Yelp 등)를 이용하여 공간 빅데이터 정합, 분포, 차트, 가시화 및 주제도 분석을 위한 알고리즘을 개발함

○ **(2-2세세부)** 공간 빅데이터 일괄분석 및 컴포넌트 기술 개발

- 연세대학교 산학협력단은 공간 빅데이터 및 이동객체데이터의 일괄(Batch) 분석 컴포넌트 개발을 목표로, 1~6차년도에 걸쳐 공간 빅데이터 요구조건 검토 / 분석 컴포넌트 설계 및 개발 / 일괄(Batch) 분석 솔루션 설계 및 개발 / 분석 컴포넌트 고도화 및 최적화 등을 수행하였음

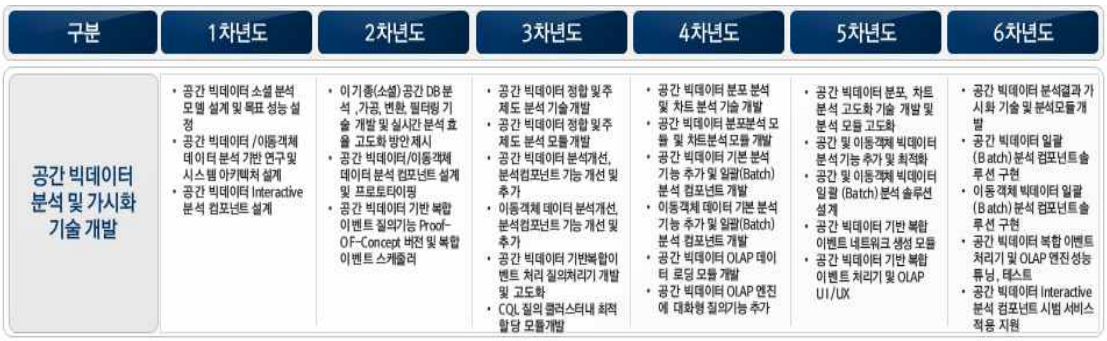
○ **(2-3세세부)** 공간 빅데이터 CEP 및 OLAP 엔진 개발

- 한국과학기술원 및 유엔진은 실시간 입력되는 공간 빅데이터를 대화형으로 신속 정확하게 분석할 수 있는 기술을 개발하였으며 크게 (1) 공간 빅데이터 분석 기반 복합 이벤트 처리 기술과 (2) 공간 빅데이터 기반 대화형 분석 기술로 구성



<그림 1-8> 공간 빅데이터 분석 및 가시화 기술 개발의 주요연구내용 및 구조

(나) 연차별 연구내용



구분	1차년도	2차년도	3차년도	4차년도	5차년도	6차년도
공간 빅데이터 분석 및 가시화 기술 개발	• 공간 빅데이터 소셜 분석 모델 설계 및 목표 성능 설정 • 공간 빅데이터 /이동객체 데이터 분석 기반 연구및 시스템 아키텍처 설계 • 공간 빅데이터 Interactive 분석 컴포넌트 설계	• 이기종(소셜)공간 DB 분석 ,가공, 변환, 필터링 기술 개발 및 실시간 분석 효율 고도화 방안 제시 • 공간 빅데이터/이동객체 데이터 분석 컴포넌트 설계 및 프로토타이핑 • 공간 빅데이터 기반 복합 이벤트 질의기능 Proof-OF-Concept 버전 및 복합 이벤트 스케줄러	• 공간 빅데이터 정합 및주제도 분석 기술개발 • 공간 빅데이터 정합 및주제도 분석 모듈개발 • 공간 빅데이터 분석개선, 분석컴포넌트 기능 개선 및 추가 • 이동객체 데이터 분석개선, 분석컴포넌트 기능 개선 및 추가 • 공간 빅데이터 기반복합이벤트 처리 질의처리기 개발 및 고도화 • CQL 질의 클러스터내 최적 할당 모듈개발	• 공간 빅데이터 분포 분석 및 차트 분석 기술 개발 • 공간 빅데이터 분포분석 모듈 및 차트분석모듈 개발 • 공간 빅데이터 기본 분석 기능 추가 및 일괄(Batch) 분석 컴포넌트 개발 • 이동객체 데이터 기본 분석 기능 추가 및 일괄(Batch) 분석 컴포넌트 개발 • 공간 빅데이터 OLAP 데이터 로딩 모듈 개발 • 공간 빅데이터 OLAP 엔진에 대화형 질의기능 추가	• 공간 빅데이터 분포, 차트 분석 고도화 기술 개발 및 분석 모듈 고도화 • 공간 및 이동객체 빅데이터 분석기능 추가 및 최적화 • 공간 및 이동객체 빅데이터 일괄 (Batch) 분석 솔루션 설계 • 공간 빅데이터 기반 복합 이벤트 네트워크 생성 모듈 • 공간 빅데이터 기반 복합 이벤트 처리기 및 OLAP UI/UX	• 공간 빅데이터 분석결과 가시화 기술 및 분석모듈개발 • 공간 빅데이터 일괄 (Batch) 분석 컴포넌트 솔루션 구현 • 이동객체 빅데이터 일괄 (Batch) 분석 컴포넌트 솔루션 구현 • 공간 빅데이터 복합 이벤트 처리기 및 OLAP 엔진 성능 튜닝, 테스트 • 공간 빅데이터 Interactive 분석 컴포넌트 시범 서비스 적용 지원

<그림 1-9> 연차별 연구내용

○ 1차년도에는 비정형 데이터의 공간 빅데이터 수집 및 가공, 공간 빅데이터 및 이동객체데이터 분석 요구조건을 검토하기 위하여 R을 이용한 통계기반 분석기능과 Machine Learning 등을 이용한 비통계기반 분석기능을 검토하고, 분석 컴포넌트 아키텍처 및 인터페이스를 설계, 공간 빅데이터 Interactive 분석 컴포넌트 요구사항 정의서 및 설계서를 작성하였음

○ 2차년도에는 저장 모듈 개발과 공간 빅데이터 분석 시스템의 모듈 개발로 위치 정합 모듈개발, 공간 빅데이터 및 이동객체데이터의 기본 분석기능을 개발하고, 분석 컴포넌트 알고리즘 개발 및 프로토타입을 구현, 공간 빅데이터 분석 기반 복합 이벤트 처리(CEP) 시스템을 개발함으로써 공간정보와 연관된 스트림 데이터(예 : 교통 데이터)로부터 복합적인 이벤트를 효율적으로 검색하여 실시간 분석 기능을 지원하고 상황 인식을 통해 적절하게 특정 이벤트에 대응기술 개발

○ 3차년도에는 주제도 분석 모듈개발, 2차년도에 개발한 공간 빅데이터 및 이동객체데이터의 기본 분석기능의 성능을 개선하고, 분석 컴포넌트의 성능 향상 및 사용성을 확보, 공간 빅데이터 기반 복합 이벤트 스케줄러 모듈, 질의 처리기 모듈, 질의 최적 할당 모듈, 복합 이벤트 네트워크 생성 모듈, UI/UX를 개발

○ 4차년도에는 분포/차트 분석 모듈개발, 공간 빅데이터 및 이동객체데이터의 기본 분석기능을 추가 개발하고, 통계기반 / 비통계기반 분석 성능을 향상시켰으며 일괄(Batch) 분석을 위한 컴포넌트 추가 개발, 공간 빅데이터 기반 대

화형 분석 지원(OLAP) 시스템을 개발함으로써 공간정보와 연관된 여러 통계 값 및 정보(예 : 부동산, 지적 데이터)들을 공간축 혹은 시간축을 따라 큰 단위 혹은 작은 단위로 반복적으로 재탐색하여 대화형으로 검색 및 계산하는 분석 수행

○ 5차년도는 고도화 모듈개발, 공간 빅데이터 및 이동객체데이터의 기본 분석 기능 추가 및 기 개발된 분석 기능을 고도화 및 최적화하였고, 이를 일괄 분석 컴포넌트 상에 반영, 공간 빅데이터 OLAP 데이터 로딩 모듈, 드릴다운/롤업 모듈, 대화형 분석기 모듈, UI/UX를 개발

○ 6차년도에는 분석결과 가시화 모듈개발, 공공 / 민간 수요 기반 일괄(Batch) 분석 시나리오 설계 및 공간 빅데이터 및 이동객체데이터의 분석기능을 추가하였고, 공간 빅데이터 및 이동객체데이터 일괄(Batch) 분석 솔루션을 구현, 공간 빅데이터 Interactive 분석 컴포넌트 테스트 및 튜닝을 수행하고, 시범서비스 적용을 지원

○ 연차별 세부 연구내용 및 범위는 <표 1-2>과 같음

<표 1-2> 연구개발 내용 및 범위

구분		세부과제명	세부연구 목표	연구개발 수행내용	수행기관
연도	세세부				
1차 년도	2-1	공간 빅데이터 소셜 분석 기술 개발	• 공간 빅데이터 분석을 위한 모델 설계 및 목표 성능 설정	• 공간 빅데이터 분석모델 현황분석 및 요구사항 정의 • 공간 빅데이터 분석 목표모델 정의	서울대학교 산학협력단
		공간 빅데이터 분석 및 가시화 컴포넌트 구현	• 공간 빅데이터 분석을 위한 모델 설계 및 목표 성능 설정	• 데이터 특성별 분석모델 설계	시터스
	2-2	공간 빅데이터 분석 요구조건 분석	• 공간 빅데이터 및 이동객체데이터 통계기반 분석 요구조건 분석(R의 공간분석기능 검토) • 공간 빅데이터 및 이동객체데이터 비통계기반 분석 요구조건 분석(Machine Learning 기법 등)	• 공간 빅데이터 및 이동객체데이터 R 기반 분석기능 각 12종, 9종 검토 • 공간 빅데이터 및 이동객체데이터 비통계기반 분석 관련 각 21편, 24편 논문 비교분석	연세대학교 산학협력단
		공간 빅데이터 분석 컴포넌트 아키텍처 및 인터페이스 설계	• 공간 빅데이터 및 이동객체데이터 분석 컴포넌트 아키텍처 설계 • 공간 빅데이터 및 이동객체데이터 비통계기반 분석 컴포넌트 인터페이스 설계	• 사례 분석을 통하여 공간 빅데이터 Batch 분석 컴포넌트와 각 컴포넌트를 관리해야 할 컴포넌트의 필요성 확인 • Batch 분석 컴포넌트 아키텍처 및 인터페이스 설계	엘에이치

구분		세부과제명	세부연구 목표	연구개발 수행내용	수행기관
연도	세세부				
	2-3	공간 빅데이터 Interactive 분석 컴포넌트	• 공간 빅데이터 Interactive 분석 컴포넌트 요구사항 정의서	• 공간 빅데이터 Interactive 분석 컴포넌트 요구사항 정의서 작성	한국과학기술원
			• 공간 빅데이터 Interactive 분석 컴포넌트 설계서	• 공간 빅데이터 Interactive 분석 컴포넌트 설계서 작성	
2차 년도	2-1	공간 빅데이터 소셜 분석 기술 개발	• 공간 빅데이터 분석을 위한 데이터 가공, 변환, 필터링 및 분석효율 고도화	• 이기종 공간DB 특성 분석	서울대학교 산학협력단
				• 이기종 DB간 데이터 가공, 변환, 필터링 기술 개발	
				• 다중 레이어를 위한 캐싱/프리패칭/버퍼링 기술 개발	
		공간 빅데이터 분석 및 가시화 컴포넌트 구현	• 공간 빅데이터 분석을 위한 데이터 가공, 변환, 필터링 및 분석효율 고도화	• 이기종 DB간 데이터 가공, 변환, 필터링 모듈 개발	시터스
				• 다중 레이어를 위한 캐싱/프리패칭/버퍼링 모듈 개발	
	2-2	공간 빅데이터 및 이동객체데이터 분석 컴포넌트 설계	• 공간 빅데이터 및 이동객체데이터 기반 기본 분석기능 선정 및 신규 개발 기능 결정(Predictive 기능 및 Descriptive 기능) • 통계기반 분석 컴포넌트 기초 설계 • 비통계기반 분석 컴포넌트 기초 설계	• 공간 및 이동객체 빅데이터 기본 분석 컴포넌트 기능 선정 및 개발 • 통계/비통계 기반 분석 컴포넌트 기초 개념 설계 및 개발	연세대학교 산학협력단
		공간 빅데이터 및 이동객체데이터 분석 컴포넌트 프로토타이핑	• Batch 분석을 위한 middleware 프로토타이핑 • 컴포넌트간 내부 인터페이스 프로토타이핑 • 컴포넌트 관리 프로토타이핑 • Batch 분석을 위한 기본 컴포넌트 알고리즘 기술 개발 및 프로토타이핑 작성	• 선정된 개발 기능의 컴포넌트화 과정 수행 • 각 컴포넌트 별 기초 설계 수행 • 기초 설계된 컴포넌트를 활용한 개발 기능 프로토타입 구현	엘에이치
	2-3	공간 빅데이터 Interactive 분석 컴포넌트	• CQL 질의 메모리 상주 색인 및 저장 모듈	• CQL 질의 메모리 상주 색인 및 저장 모듈 개발	한국과학기술원
			• 공간 빅데이터 기반 복합 이벤트 스케줄러 모듈	• 공간 빅데이터 기반 복합 이벤트 스케줄러 모듈 개발	유누스
3차 년도	2-1	공간 빅데이터 소셜 분석 기술 개발	• 공간 빅데이터 정합 및 주제도 분석 기술	• 공간 빅데이터 정합 및 주제도 분석 기술 개발	서울대학교 산학협력단
		공간 빅데이터 분석 및 가시화 컴포넌트 구현	• 공간 빅데이터 정합 및 주제도 분석 기술	• 공간 빅데이터 정합 및 주제도 분석 모듈 개발	시터스

구분		세부과제명	세부연구 목표	연구개발 수행내용	수행기관
연도	세세부				
	2-2	공간 빅데이터 및 이동객체데이터 분석 개선	• 공간 빅데이터 및 이동객체 데이터 기본 분석기능 개선 • 통계기반 분석 성능 향상 • 비통계기반 분석 성능 향상	• 기존 제작된 공간 빅데이터 및 이동객체데이터 분석 알고리즘의 성능 향상을 위하여 MBR기법 등 추가적인 분석기법을 도입함. • 통계기반 알고리즘인 Getis Ord Gi*를 이용한 분석의 확장성 확보 • 비통계기반 알고리즘인 LISA를 이용한 분석의 확장성 확보	연세대학교 산학협력단
		공간 빅데이터 및 이동객체데이터 분석 컴포넌트 기능 개선 및 추가	• 컴포넌트간 내부 기능 개선 • 컴포넌트 관리 기능 개선 • Batch 분석을 위한 기본 컴포넌트 알고리즘 기술 • 추가 개발 및 기능 개선	• 공간 빅데이터 및 이동객체 데이터 분석 모듈 4개 사항 추가 • Spark System을 이용하여 성능 향상 • Web-platform 에서 UI/UX 기반으로 구현되어 사용성 확보	키위 컴퍼니
	2-3	공간 빅데이터 Interactive 분석 컴포넌트	• CQL 질의 클러스터 내 최적 할당 모듈	• CQL 질의 클러스터 내 최적 할당 모듈 개발	유누스
			• 공간 빅데이터 기반 질의 처리기 모듈	• 공간 빅데이터 기반 질의 처리기 모듈 개발	한국과학기술원
4차 년도	2-1	공간 빅데이터 소셜 분석 기술 개발	• 공간 빅데이터 분포 및 차트 분석 기술	• 공간 빅데이터 분포 분석 기술 개발 • 공간 빅데이터 차트 분석 기술 개발	서울대학교 산학협력단
		공간 빅데이터 분석 및 가시화 컴포넌트 구현	• 공간 빅데이터 분포 및 차트 분석 모듈	• 공간 빅데이터 분포 분석 모듈 개발 • 공간 빅데이터 차트 분석 모듈 개발	시터스
	2-2	공간 빅데이터 일괄(Batch) 분석 컴포넌트 기술 개발	• 공간 빅데이터 기본 분석 기능 추가 • 통계기반 분석 성능 최적화 • 비통계기반 분석 성능 최적화	• 공간 빅데이터 기본 분석기능 개선 • 통계기반 분석 성능 향상 • 비통계기반 분석 성능 향상 • 컴포넌트간 내부 기능 추가 • 컴포넌트 관리 기능 개선	연세대학교 산학협력단
		이동객체 데이터 Batch 분석 컴포넌트	• 이동객체데이터 기본 분석 기능 추가 • 통계기반 분석 성능 최적화 • 비통계기반 분석 성능 최적화	• 이동객체 데이터 기본 분석기능 개선 • 통계기반 분석 성능 향상 • 비통계기반 분석 성능 향상 • Batch 분석을 위한 기본 컴포넌트 알고리즘 기술 추가 개발 및 기능 개선	가이온

구분		세부과제명	세부연구 목표	연구개발 수행내용	수행기관
연도	세세부				
	2-3	공간 빅데이터 Interactive 분석 컴포넌트	• 공간 빅데이터 OLAP 데이터 로딩 모듈	• 공간 빅데이터 OLAP 데이터 로딩 모듈 개발	유누스/유엔진솔루션즈
			• 공간 빅데이터 OLAP 드릴다운/롤업 모듈	• 공간 빅데이터 OLAP 드릴다운/롤업 모듈 개발	한국과학기술원
5차 년도	2-1	공간 빅데이터 소셜 분석 기술 개발	• 공간 빅데이터 분포 및 차트 분석 고도화 기술	• 공간 빅데이터 분포 분석 고도화 기술 개발	서울대학교 산학협력단
				• 공간 빅데이터 차트 분석 고도화 기술 개발	
			• 공간 빅데이터 분석 결과 가시화 기술 (정적 가시화 1건)	• 공간 빅데이터 분석결과 가시화 기술(정적 가시화) 개발	
		공간 빅데이터 분석 및 가시화 컴포넌트 구현	• 공간 빅데이터 분포 및 차트 분석 고도화 모듈	• 공간 빅데이터 분포 분석 고도화 모듈 개발	시터스
				• 공간 빅데이터 차트 분석 고도화 모듈 개발	
			• 공간 빅데이터 분석 결과 가시화 모듈 (정적 가시화 1건)	• 공간 빅데이터 분석결과 가시화 모듈 개발 (정적 가시화)	
	2-2	공간 빅데이터 일괄(Batch) 분석 컴포넌트 기술 개발	• 공간 빅데이터 • 기본 분석 기능 추가 및 최적화 • 이동객체 데이터 기본 분석 기능 추가	• 공간 빅데이터 기본 분석기능 추가 • 통계기반 분석 성능 최적화 • 비통계기반 분석 성능 최적화 • 이동객체 기본 분석기능 추가 • 통계기반 분석 성능 최적화 • 비통계기반 분석 성능 최적화	연세대학교 산학협력단
		공간 빅데이터 일괄(Batch) 분석 컴포넌트 시스템 구현	• 공간 빅데이터 일괄(Batch) 분석 컴포넌트 개발 • 이동객체 데이터 일괄(Batch) 분석 컴포넌트 개발	• 공간 빅데이터 신규 분석기능 컴포넌트 개발 (3종) • 공간 빅데이터 분석 컴포넌트 기능 고도화 • 이동객체 데이터 신규 분석기능 컴포넌트 개발 (2종) • 이동객체 데이터 분석 컴포넌트 기능 고도화	가이온

구분		세부과제명	세부연구 목표	연구개발 수행내용	수행기관
연도	세세부				
	2-3	공간 빅데이터 Interactive 분석 컴포넌트	• 공간 빅데이터 OLAP 대화형 분석기 모듈	• 공간 빅데이터 OLAP 대화형 분석기 모듈 개발	한국과학기술원
			• 공간 빅데이터 기반 복합 이벤트 네트워크 생성 모듈	• 공간 빅데이터 기반 복합 이벤트 네트워크 생성 모듈 개발	
			• 공간 빅데이터 기반 복합 이벤트 처리기 UI/UX	• 공간 빅데이터 기반 복합 이벤트 처리기 UI/UX 개발	유엔진솔루션즈
			• 공간 빅데이터 OLAP UI/UX	• 공간 빅데이터 OLAP UI/UX 개발	
6차 년도	2-1	공간 빅데이터 소셜 분석 기술 개발	• 공간 빅데이터 분석결과 가시화 기술	• 공간 빅데이터 분석결과 가시화 기술 고도화 개발	서울대학교 산학협력단
		공간 빅데이터 분석 및 가시화 컴포넌트 구현	• 공간 빅데이터 분석결과 가시화 모듈	• 공간 빅데이터 분석결과 가시화 모듈 개발 (동적 가시화)	시터스
			• 공간 빅데이터 사업단 실증	• 공간 빅데이터 분석모듈 사업단 적용 및 실증	
	2-2	공간 빅데이터 및 이동객체데이터 일괄(Batch) 분석 설계	• 공간 빅데이터 및 이동객체데이터 일괄(Batch) 분석 설계	• 공공/민간 수요기반 Batch 분석 시나리오 설계 • 공간 빅데이터 및 이동객체데이터 Batch 솔루션 설계	연세대학교 산학협력단
		공간 빅데이터 및 이동객체데이터 일괄(Batch) 분석 솔루션 구현 및 실증	• 공간 빅데이터 및 이동객체데이터 일괄(Batch) 분석 솔루션 구현 및 실증	• 공간 빅데이터 및 이동객체데이터 분석 Batch 솔루션 구현 여부 • 공간 빅데이터 및 이동객체데이터 분석 솔루션 사업단 실증 (2건)	가이온
	2-3	공간 빅데이터 Interactive 분석 컴포넌트	• 공간 빅데이터 Interactive 분석 컴포넌트 테스팅 및 튜닝	• 공간 빅데이터 Interactive 분석 컴포넌트 테스팅 및 튜닝	한국과학기술원
			• 공간 빅데이터 Interactive 분석 컴포넌트 시범서비스 적용 지원	• 공간 빅데이터 Interactive 분석 컴포넌트 시범서비스 적용 지원	유엔진솔루션즈

(3) 공간 빅데이터 서비스 활성화 및 실증 연구

(가) 주요연구내용 및 구조

○ **공간 빅데이터 플랫폼 활용을 지원하는 서비스의 활성화 및 실증을 위하여 공간 빅데이터 활용 서비스 프레임워크, 공간 빅데이터 서비스 제공을 위한 UI/UX, OpenAPI, 공간 빅데이터 분석 결과 연계 및 가시화 모듈(SW) 등을 개발 및 실증**

○ **(3-1 세세부)** 빅데이터 서비스 활성화 방안 연구

- 공간 빅데이터 서비스 플랫폼의 구체적인 구현방안과 활용 시스템 실증을 위한 서비스 모델 개발

○ **(3-2 세세부)** 공간 빅데이터 서비스 실용화 연구

- 공공 빅데이터를 표준화된 OpenAPI 기반으로 개방·공유·활용 가능하게 하여 사용자 활용성을 높이고, 사용자 편의를 고려한 상호작용 기술과 사용자의 이용 경험을 바탕으로 공간 빅데이터 서비스 제공을 위한 UI/UX 개발

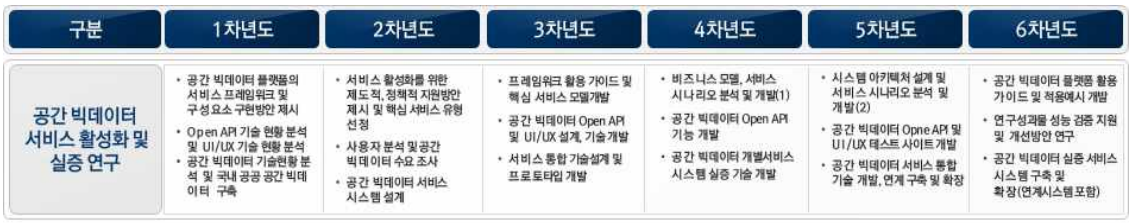
○ **(3-3 세세부)** 공간 빅데이터 서비스 실증 연구

- 공간 빅데이터를 관리, 분석 및 활용할 수 있는 통합 기술과 연계 및 확장 서비스 기술 개발 및 실증



<그림 1-10> 공간 빅데이터 서비스 활성화 및 실증 연구의 주요연구내용 및 구조

(나) 연차별 연구내용



구분	1차년도	2차년도	3차년도	4차년도	5차년도	6차년도
공간 빅데이터 서비스 활성화 및 실증 연구	• 공간 빅데이터 플랫폼의 서비스 프레임워크 및 구성요소 구현방안 제시 • Open API 기술 현황 분석 및 UI/UX 기술 현황 분석 • 공간 빅데이터 기술현황 분석 및 국내 공공 공간 빅데이터 구축	• 서비스 활성화를 위한 제도적, 정책적 지원방안 제시 및 핵심 서비스 유형 선정 • 사용자 분석 및 공간 빅데이터 수요 조사 • 공간 빅데이터 서비스 시스템 설계	• 프레임워크 활용 가이드 및 핵심 서비스 모델개발 • 공간 빅데이터 Open API 및 UI/UX 설계, 기술개발 • 서비스 통합 기술설계 및 프로토타입 개발	• 비즈니스 모델, 서비스 시나리오 분석 및 개발(1) • 공간 빅데이터 Open API 기능 개발 • 공간 빅데이터 개별서비스 시스템 실증 기술 개발	• 시스템 아키텍처 설계 및 서비스 시나리오 분석 및 개발(2) • 공간 빅데이터 Opne API 및 UI/UX 테스트 사이트 개발 • 공간 빅데이터 서비스 통합 기술 개발, 연계 구축 및 확장	• 공간 빅데이터 플랫폼 활용 가이드 및 적용예시 개발 • 연구성과물 성능 검증 지원 및 개선방안 연구 • 공간 빅데이터 실증 서비스 시스템 구축 및 확장(연계시스템포함)

<그림 1-11> 연차별 연구내용

- 1차년도에는 공간 빅데이터 플랫폼의 서비스 프레임워크 및 구성요소 구현방안 제시, OpenAPI 기술 현황 분석 및 UI/UX 기술 현황 분석, 공간 빅데이터 기술현황 분석 및 국내 공공 공간 빅데이터 구축
- 2차년도에는 서비스 활성화를 위한 제도적, 정책적 지원방안 제시 및 핵심서비스 유형 선정, 사용자 분석 및 공간 빅데이터 수요조사, 공간 빅데이터 서비스 시스템 설계
- 3차년도에는 프레임워크 활용 가이드 및 핵심 서비스 모델 개발, 공간 빅데이터 Open API 및 UI/UX 설계 및 관련 기술 개발, 서비스 통합 기술 설계 및 프로토타입 개발
- 4차년도에는 비즈니스모델, 서비스 시나리오 분석 및 개발(1), 공간 빅데이터 OpenAPI 기능 개발, 공간 빅데이터 개별 서비스 시스템 실증 기술 개발
- 5차년도에는 시스템 아키텍처 설계 및 서비스 시나리오 분석 및 개발(2), 공간 빅데이터 OpenAPI 및 UI/UX 테스트 사이트 개발, 공간 빅데이터 서비스 통합 기술 개발, 연계 구축 및 확장
- 6차년도에는 공간 빅데이터 플랫폼 활용 가이드 및 적용예시개발, 연구성과물 성능 검증 지원 및 개선방안 연구, 공간 빅데이터 실증 서비스 시스템 구축 및 확장(연계시스템 포함)

○ 연차별 세부 연구내용 및 범위는 <표 1-3>과 같음

<표 1-3> 연구개발 내용 및 범위

구분		세부과제명	세부연구 목표	연구개발 수행내용	수행기관
연도	세세부				
1차 년도	3-1	공간 빅데이터 플랫폼의 융복합 활용방안 개발	• 공간 빅데이터 융합 서비스의 서비스 프레임워크 및 구성요소 구현방안 제시	• V-World 융합기반 • 공간 빅데이터 플랫폼의 • 서비스 프레임워크 도출	안양대학교 산학협력단
				• 서비스 프레임워크 구성요소 개발방안 제시	
		공간 빅데이터의 공공/민간 부문 핵심 서비스 모델 도출 및 제공 방안 개발	• 핵심 서비스 모델 도출을 위한 서비스 현황 분석 및 시사점 도출	• 공공/민간 부문, 서비스 유형별 공간 빅데이터 서비스 현황 수집 및 분석	
				• 국내외 공간 빅데이터 활용사례 분석을 통한 기존 서비스와 공간 빅데이터 서비스와의 차별화 방안 도출	
	3-2	공간 빅데이터 서비스 실용화 연구	• 공간 빅데이터 확장 Open API 서비스 개발	• V-World에서 제공하는 Open API 사이트를 통한 현황파악	공간인소프트
				• Open API 국내외 적용 사례 조사	
			• 공간 빅데이터 서비스 제공 UI/UX 개발	• 서비스별 UI/UX 기술 및 적용 사례 분석	
				• 공간 빅데이터 서비스 제공을 위한 UI/UX 개발목표 및 범위 정의	
	3-3	공간 빅데이터 및 다양한 공공 시스템 기술현황 분석	• 공간 빅데이터 현황 및서비스 구축 사례 분석	• 공간 빅데이터 서비스 적용 국외/국내 사례를 분석	이지스
			• 공간정보 오픈플랫폼 실증 기술 현황 분석	• 공간정보 오픈플랫폼 적용사례를 분석 및 기반 기술 현황 분석	
		국내 공공 공간 빅데이터 구축 및 시각화 기술 현황 분석	• 국내 공공 공간 빅데이터 구축 및 시각화 기술 현황 분석	• 공공 공간 빅데이터 연계서비스를 위한 국내 빅데이터 구축 및 시각화 기술(서비스) 현황 분석	

구분		세부과제명	세부연구 목표	연구개발 수행내용	수행기관
연도	세세부				
2차 년도	3-1	공간 빅데이터 플랫폼의 융복합 활용방안 개발	• 공간 빅데이터 융합 서비스의 프레임워크 매뉴얼 작성	• 공간 빅데이터 서비스 활성화를 위한 • 정책적, 제도적 • 지원방안 제시(1)	안양대학교 산학협력단
		공간 빅데이터의 공공/민간 부문 핵심 서비스 모델 도출 및 제공 방안 개발	• 공간 빅데이터 서비스 유형화 및 핵심 서비스 모델 개발	• 핵심 서비스 유형 및 • 서비스 모델 개발(1)	
	3-2	공간 빅데이터 서비스 실용화 연구	• 공간 빅데이터 확장 Open API 서비스 개발	• 공공시스템 기반 공간 빅데이터 Open API 서비스를 위한 공간 빅데이터 콘텐츠 구성 및 인터페이스 설계	공간인소프트
				• 공간정보와의 연계 범위 및 API 종류 설계	
			• 공간 빅데이터 서비스 제공 UI/UX 개발	• 사용자 분석 및 공간 빅데이터 니즈 조사	
	3-3	공간 빅데이터 서비스 시스템 부문별 서비스 및 실증 기술 설계	• 빅데이터 서비스 아키텍처 및 프레임워크 설계	• 각 개별서비스에 대한 프로세스 과정을 확인하고 통합적으로 활용할 수 있는 구조의 실증형 빅데이터 서비스 아키텍처 및 프레임워크 설계	이지스
			• 공간 빅데이터 서비스 시스템 실증 기술 설계	• 서비스모델을 분석 기술로 융합DB를 생성 등의 서비스 시스템 실증 기술 설계	
			• 저장, 관리, 분석, 가시화 등 개별 서비스 설계	• 프로토타입 화면설계 등의 개별 서비스 설계, 관리서비스 스트럭처 설계	
		공공 공간 빅데이터 서비스 연계/확장 기술 설계(1) 및 프로토타입 개발	• 공공 공간 빅데이터 서비스 연계 및 확장 기술 설계(1)	• 국토교통분야에 대한 서비스를 개발하기 위한 사용자, 필요데이터 정의 등 기본 설계	
			• 공공 공간 빅데이터 서비스 연계 및 확장 기술 프로토타입 개발	• 유즈케이스, DB설계, 화면설계 등을 통하여 기본적인 프로토타입을 구성	

구분		세부과제명	세부연구 목표	연구개발 수행내용	수행기관
연도	세세부				
3차년도	3-1	공간 빅데이터 플랫폼의 융복합 활용방안 개발	• 공간 빅데이터 융합 서비스의 프레임워크 매뉴얼 작성	• 공간 빅데이터 서비스 활성화를 위한 • 정책적, 제도적 • 지원방안 제시(2)	안양대학교 산학협력단
				• 서비스 프레임워크 • 활용 가이드라인 개발	
		공간 빅데이터의 공공/민간 부문 핵심 서비스 모델 도출 및 제공 방안 개발	• 공간 빅데이터 서비스 유형화 및 핵심 서비스 모델 개발	• 핵심 서비스 유형 및 • 서비스 모델 개발(2)	
	3-2	공간 빅데이터 서비스 실용화 연구	• 공간 빅데이터 확장 Open API 서비스 개발	• 공간 빅데이터 확장 서비스를 위한 Open API 기능 개발	공간인소프트
			• 공간 빅데이터 서비스 제공 UI/UX 개발	• 공간 빅데이터 서비스 제공을 위한 UI/UX 상세설계	
	3-3	공간 빅데이터 서비스 통합 기술 설계 및 프로토타입 개발	• 공간 빅데이터 서비스 통합 기술 설계	• 국토교통분야에 대한 서비스를 개발하기 위한 사용자와 필요데이터의 정의, 주요 핵심 DB에 대한 설계 등 상세 기술 설계	주식회사 포도
			• 공간 빅데이터 서비스 통합 기술 프로토타입 개발	• 공간 빅데이터 서비스 통합기술 설계를 기반한 프로토타입 개발	
		공공 공간 빅데이터 서비스 연계/확장 기술 설계(2)	• 공공 공간 빅데이터 서비스 연계 및 확장 기술 설계(2)	• 시나리오에 유연하게 대처 할 수 있는 서비스 모델에 대한 기초 설계 및 서비스 모델 기반의 공간 빅데이터 서비스 연계 수행	
4차년도	3-1	공간 빅데이터의 공공/민간 부문 핵심 서비스 모델 도출 및 제공 방안 개발	• 핵심 • 서비스의 비즈니스 모델 개발	• 공공/민간부문 공간 빅데이터 활용을 위한 비즈니스 모델 분석 및 개발	안양대학교 산학협력단
				• 공공/민간부문 공간 빅데이터 활용을 위한 서비스 시나리오 개발(1)	
	3-2	공간 빅데이터 서비스 실용화 연구	• 공간 빅데이터 확장 Open API 서비스 개발	• 공간 빅데이터 확장 서비스를 위한 Open API 기능 개발	공간인소프트
				• 공간 빅데이터 Open API 서비스 사이트 개발	

구분		세부과제명	세부연구 목표	연구개발 수행내용	수행기관
연도	세세부				
	3-3	공간 빅데이터 서비스 시스템 실증 및 연계 기술 개발	• 공간 빅데이터 서비스 시스템 실증 기술 개발	• Hadoop 프레임워크 및 2세부의 분석 모듈 등 데이터의 시각화 및 정량화 가능한 실증 서비스 시스템 기술을 개발하여 체계적이고 효율적인 통합관리 서비스 제공	주식회사 포도
			• 저장, 관리, 분석, 가시화 등 개별 서비스 기술 개발	• 연구과제 수행으로 도출되는 빅데이터 저장, 관리, 분석 및 가시화 등의 빅데이터 기반 기술에 대한 시각화 및 정량화 등 개별서비스에 대한 실증 서비스 기술을 개발	
		공공 공간 빅데이터 서비스 연계 및 확장 기술 개발	• 공공 공간 빅데이터 서비스 연계 및 확장 기술 개발	• 기관,형태,포맷별로 규격화된 분류체계 확립과 이를 토대로 한 공공 공간 빅데이터 정제기법 연구	
5차 년도	3-1	공간 빅데이터의 공공/민간 부문 핵심 서비스 모델 도출 및 제공 방안 개발	• 핵심 • 서비스의 비즈니스 모델 개발	• 공공/민간부문 공간 빅데이터 활용을 위한 서비스 시나리오 개발(2)	안양대학교 산학협력단
			• 핵심서비스 제공방안 개발	• 서비스 구현을 위한 시스템 아키텍처 설계	
				• 핵심서비스의 주요 기능 도출(1)	
	3-2	공간 빅데이터 서비스 실용화 연구	• 공간 빅데이터 확장 Open API 서비스 개발	• 공공시스템 기반 공간 빅데이터 Open API 서비스 테스트 사이트 개발	공간인소프트
			• 공간 빅데이터 서비스 제공 UI/UX 개발	• 공간 빅데이터 서비스 제공을 위한 UI/UX 테스트 시스템 개발	
	3-3	공간 빅데이터 서비스 시스템 실증 및 연계 기술 개발	• 저장, 관리, 분석, 가시화 등 개별 서비스 기술 개발	• 관리서비스 개발을 통해 하둡시스템 데이터 모니터링, 분석 컴포넌트 연계를 통한 분석관리, 서비스 모델 연계를 통한 서비스관리 및 사용자, 통계관리 등의 개별 서비스 기술 개발 수행 • 공간 연산, 공간 분석 데이터의 시각화를 통한 공간 빅데이터 가시화 기술 개발 수행	주식회사 포도

구분		세부과제명	세부연구 목표	연구개발 수행내용	수행기관
연도	세세부				
			• 공간 빅데이터 서비스 통합 기술 개발	• 공간 빅데이터 연산 및 저장 연계, 분석 기능 연계 등 각 세부 기술 연계 수행	
		공공 공간 빅데이터 서비스 연계 및 확장 기술 개발	• 공공 공간 빅데이터 서비스 연계 및 확장 기술 개발	• 기관,형태,포맷별로 규격화된 분류체계 확립과 이를 토대로 한 공공 공간 빅데이터 정제기법 개발	
6차 년도	3-1	공간 빅데이터 플랫폼의 융복합 활용방안 개발	• 공간 빅데이터 활용 가이드 개발	• 공간 빅데이터 플랫폼 활용 가이드 개발	안양대학교 산학협력단
				• 핵심 서비스모델 기반 활용 가이드 적용예시	
		공간 빅데이터의 공공/민간 부문 핵심 서비스 모델 도출 및 제공 방안 개발	• 핵심서비스 제공방안 개발	• 핵심서비스의 주요 기능 도출(2)	
	3-2	공간 빅데이터 서비스 실용화 연구	• 공간 빅데이터 확장 Open API 서비스 개발	• 연구성과물 성능 검증 지원 및 개선방안 연구	공간인소프트
			• 공간 빅데이터 서비스 제공 UI/UX 개발	• 공공시스템 기반 공간 빅데이터 Open API 서비스를 적용한 UI/UX 테스트 사이트 개발	
				• 연구성과물 성능 검증 지원 및 개선방안 연구	
	3-3	공간 빅데이터 서비스 시스템 실증 및 연계 기술 개발	• 공간 빅데이터 서비스 통합 기술 개발	• 1세부의 공간 빅데이터 저장/접근, 공간 연산 기능과 2-1세부의 공간 감성분석 모듈, 2-2세부의 공간 빅데이터 배치분석, 2-3세부의 공간 빅데이터 인터랙티브 분석 기술을 통합적으로 연계	주식회사 포도
		공간 빅데이터 서비스 실증 시스템 구축 (연계서비스 포함)	• 공간 빅데이터 서비스 시스템 구축	• 공간 연산, 공간 분석 및 분석 진행 상태 등의 기능을 이용할 수 있는 공간 분석 프레임을 통해 분석 완료된 것을 실증 시각화 서비스에서 확인 가능한 시스템 구축	
		공간 빅데이터 서비스 실증 시스템 확장 구축	• 공공 시스템 기반 공간 빅데이터 서비스 시스템 확장 구축	• 공공 공간 빅데이터의 표준화, 정제, 분석을 통해 표준화된 데이터를 생성하고 이를 공간연산에 활용	

(4) 연구내용 변동사항

(가) 공간 빅데이터 저장관리 인프라 기술 개발 (1세부)

○ 연구기간 중 최신 기술 동향 및 연구 환경의 변화 등 변동되는 상황에 따라 <표 1-4>와 같이 연구 내용을 변경하여 연구개발을 추진함

○ 인메모리 공간 빅데이터 처리기술 연구내용을 구체화하여 3차년도 이후 계획에 반영함(2016.02)

- Spark 등 인메모리 빅데이터 처리기술에 대한 관심이 증가함에 따라 인메모리 기반 공간 빅데이터 처리기술을 독립된 핵심성과(O)로 정의하고 연구내용을 구체화하여 이후 연차의 연구내용에 반영함
- 이후 연구내용에 대한 우선순위를 도출하여 시스템의 기반기술이라고 할 수 있는 공간 빅데이터 처리 연산자 연구 수행을 3차년도에 조기 시작하고 공간빅데이터 처리 언어 개발의 일부는 4차년도에 추진하는 등 일정을 조정하였음

○ 성과물의 사업단 플랫폼 탑재 및 연계활용 계획 및 일정에 대응하기 위하여 연구내용을 조정하고 일정을 앞당겨 조기 수행하도록 함(2018.01)

- 연구 성과물을 '공간 빅데이터 체계구축사업' 에 활용하기 위하여 필요한 성능 비교 및 연계 작업 등 수행을 위하여 '공간 빅데이터 저장관리 인프라 시범서비스 적용 지원' 의 일부를 1년 앞당겨 5차년도(2018)에 조기 수행함

○ 공간빅데이터 사업단 적용 및 타 세부 연계를 고려하여 5차년도 이후에는 성과물의 활용성을 높일 수 있는 연구내용을 추가 및 조정하여 추진함(2018.01)

- 공간분석 응용개발자에게 손쉽고 직관적인 공간 빅데이터 분석 응용 개발을 가능하게 하여 성과물의 실용화 및 보급/확산에 큰 기여를 할 것으로 예상되는 '원격 공간 빅데이터 활용 기술' 과 'GUI 기반 공간 빅데이터 분석 응용 개발 도구' 를 연구내용에 추가함
- '공간 NoSQL DBMS 개발' 을 '원격 공간 빅데이터 활용 기술' 과 '공간 빅데이터 분석 응용 개발 도구' 로 대체하여 연구개발을 수행함
- 실시간 공간 빅데이터 플랫폼의 활용성을 높이기 위해 플랫폼 아키텍처를 구성하는 핵심 기능인 '인메모리 기반 공간질의 인터페이스와의 통합' 을 선행하여 연구개발을 수행함
- 타 세부 연구 결과물 및 공간 ETL 처리 엔진과 연동되어 공간 정보 추출/변환/적재 시 동작되어 공간 연산 중 불필요한 연산을 줄일 수 있는 '스트림 기반 공간 ETL 규칙처리 엔진 개발' 의 일정을 앞당겨 5차년도에 POC(Proof-Of-Concept) 수준으로 연구를 수행함

<표 1-4> 연차별 연구개발계획서 대비 연구내용 변동사항(1세부)

기존 변경
삭제 신규

구분	핵심연구성과 (level1)		단위 성과 (level2)		연차					
					1차 년도	2차 년도	3차 년도	4차 년도	5차 년도	6차 년도
1세부	A	Hadoop 기반 공간 빅데이터 저장 프레임워크	A-1	공간 빅데이터 저장 시스템 요구사항 정의서	기존					
			A-2	공간 빅데이터 저장 시스템 개념 설계서	기존					
			A-3	공간 빅데이터 저장 시스템 인터페이스 정의서	기존					
			A-4	공간 빅데이터 저장시스템(Proof-of-Concept) 버전) 모듈	기존					
			A-5	실시간 공간 빅데이터 저장 플랫폼 설계서	삭제					
			A-6	공간 빅데이터 처리 연산자 v1.0 개발		기존				
			A-7	스트림 공간 빅데이터 언어 규격 정의		기존			변경	
			A-8	공간 빅데이터 처리 언어 파서 모듈		삭제				
			A-9	공간 빅데이터 처리 언어 내부 코드 규격		삭제				
			A-10	공간 빅데이터 색인 모듈 구현(1)		기존	변경			
			A-11	공간 빅데이터 색인 모듈 구현(2)		기존		변경		
			A-12	실시간 공간 빅데이터 저장 연구시제품			기존			
			A-13	공간 빅데이터 처리 연산자 v2.0 개발			기존			
			A-14	공간 빅데이터 처리 언어 컴파일러 모듈 v1.0			기존			
			A-15	공간 빅데이터 처리 언어용 내부 코드 실행기 구현			기존			변경
			A-16	공간 빅데이터 처리 언어 규격 v2.0(BNF 형식)			기존			
			A-17	공간 빅데이터 고속 색인 생성기 구현			기존		변경	
			A-18	실시간 공간 빅데이터 저장 플랫폼 v1.0			기존			
			A-19	이동객체 빅데이터 처리 플랫폼 설계서			기존			
			A-20	색인 기반 고속 공간 조인 연산자 구현				기존	변경	
			A-21	MapReduce 기반 워크플로우 가시화도구 개발				기존		변경
			A-22	MapReduce 기반 공간 빅데이터 처리 최적화 모듈				삭제		
			A-23	공간빅데이터 저장 관리 인프라 시범 서비스 적용 지원(1)				기존	변경	

기존 변경
삭제 신규

구분	핵심연구성과 (level1)		단위 성과 (level2)		연차					
					1차 년도	2차 년도	3차 년도	4차 년도	5차 년도	6차 년도
			A-24	클라이언트-서버 기반 원격 공간 빅데이터 접근 모듈 개발				기존	변경	
			A-25	공간빅데이터 저장 관리 인프라 시범 서비스 적용 지원(2)				기존		변경
			A-26	GUI 기반 공간 빅데이터 응용 도구 개발						신규
			A-27	GUI 기반 공간 빅데이터 응용 도구 개발(2)						신규
			A-28	대용량 공간 빅데이터 원격 접근 기능 개발						신규
			A-29	대용량 공간 빅데이터 클라이언트 캐싱 모듈 개발						신규
			A-30	2세부 공간 연산자 연동 모듈 개발						신규
	B	Hadoop 기반 공간 빅데이터 질의 처리 엔진	B-1	공간 빅데이터 질의 처리 엔진 요구사항정의서	기존					
			B-2	공간 빅데이터 질의 처리 엔진 설계서	기존					
			B-3	공간 빅데이터 질의 언어 정의서	기존					
			B-4	공간 빅데이터 ODBC/JDBC 인터페이스 설계	기존					
			B-5	공간 빅데이터 DBMS 인터페이스 기본기능 모듈 구현	기존					
			B-6	공간 빅데이터 질의 플랜 생성 모듈 구현		기존				
			B-7	스트림 공간 빅데이터 질의언어 파싱 및 의미 검증 모듈 구현		기존				
			B-8	스트림 공간 빅데이터 질의처리 모듈 구현		기존				
			B-9	스트림 공간 빅데이터 분산 질의처리모듈 구현		기존	변경			
			B-10	공간 빅데이터 DBMS/ODBC/JDBC 클라이언트/서버 모듈 구현			기존	변경		
			B-11	스트림 공간 빅데이터 질의 스케줄링모듈 구현				기존		
			B-12	스트림 공간 빅데이터 이벤트 처리 모듈 구현				기존		
			B-13	스트림 공간 빅데이터 인터페이스 통합처리 서버 프레임워크 구현				기존	변경	
			B-14	스트림 공간 빅데이터 분산 질의처리 최적화 모듈 구현				기존		변경

기존 변경
삭제 신규

구분	핵심연구성과 (level1)		단위 성과 (level2)		연차					
					1차 년도	2차 년도	3차 년도	4차 년도	5차 년도	6차 년도
	C	공간 NoSQL DBMS	C-1	공간 NoSQL DBMS 요구사항 정의서			삭제			
			C-2	공간 NoSQL DBMS 시스템구조 설계서			삭제			
			C-3	공간 NoSQL DBMS 확장 플러그인 모듈			삭제			
			C-4	공간 NoSQL DBMS 확장 색인 모듈			삭제			
			C-5	공간 NoSQL DBMS 확장 연산 모듈			삭제			
			C-6	공간 NoSQL DBMS 질의 처리 통합모듈				삭제		
			C-7	이동객체 모델 정의서				삭제		
			C-8	NoSQL DBMS 이동객체 확장 색인 모듈				삭제		
			C-9	NoSQL DBMS 이동객체 확장 연산 모듈				삭제		
	D	워크플로우기반 공간 ETL 도구개발	D-1	공간 빅데이터 웨어하우스 사용자/시스템/기술 요구사항 분석	기존					
			D-2	공간 빅데이터 웨어하우스 사용자/서비스/플랫폼 연동 및 ETL 처리 핵심 API 설계	기존					
			D-3	ETL 워크플로우 생성을 위한 ETL Rule 설계	기존					
			D-4	ETL워크플로우 자동 생성 모듈 설계	기존					
			D-5	워크플로우기반 ETL 데이터 처리 엔진 설계	기존					
			D-6	다중 공간 데이터 수집 모듈 개발		기존				
			D-7	공간 스트림 데이터 수집 모듈 개발		기존				
			D-8	공간 스트림 데이터 병렬/분산 처리모듈 개발		기존	변경			
			D-9	스트림 기반 공간 ETL 처리 엔진 개발		기존		변경		
			D-10	스트림 기반 공간 ETL 처리 엔진 연동 API 개발		기존			변경	
			D-11	스트림 기반 공간 ETL 규칙 처리 엔진 개발		기존			변경	
			D-12	공간 센서 스트림 데이터 수집 및 분산처리 모듈 개발				기존	변경	
			D-13	스트림 기반 공간 ETL 규칙 처리 엔진 개발				기존		변경
			D-14	공간 스트림 ETL 관리 도구 개발				기존		변경

기존 변경
삭제 신규

구분	핵심연구성과 (level1)		단위 성과 (level2)		연차					
					1차 년도	2차 년도	3차 년도	4차 년도	5차 년도	6차 년도
			D-16	스트림 기반 공간 ETL 처리 엔진 연동 API 개발(PoC 수준)				신규		
			D-17	스트림 기반 공간 ETL 규칙 처리 엔진 개발(2차 배포판)					신규	
			D-18	공간 연산 워크플로우 분석도구 개발						신규
	O	인메모리 기반 공간 빅데이터 저장 관리 프레임워크 개발 * 핵심성과 O의 연구내용은 2차년도이후 핵심성과A에서 이동한 것임	O-1	실시간 공간 빅데이터 저장 플랫폼 설계서	신규					
			O-2	실시간 공간 빅데이터 저장 연구시제품		신규				
			O-3	인메모리 기반 공간색인 모듈			신규			
			O-4	인메모리 데이터 처리를 위한 기본 공간 연산자			신규			
			O-5	인메모리 기반 이동객체 빅데이터 저장 플랫폼 설계				신규		
			O-6	인메모리 기반 이동객체 빅데이터 저장 모듈					신규	
			O-7	인메모리 기반 공간색인 및 조회 최적화 모듈					신규	
			O-8	실시간 공간 빅데이터 저장 플랫폼 v1.0					신규	
			O-10	실시간 공간 빅데이터 저장 플랫폼 v2.0(1)					신규	
			O-11	인메모리 기반 이동객체 빅데이터 저장 모듈(PoC 수준)				신규		
			O-12	실시간 공간 빅데이터 저장 플랫폼 v2.0(2)						신규
			O-13	실시간 인메모리 기반 저장 플랫폼 실증						신규

(나) 공간 빅데이터 분석 및 가시화 기술 개발 (2세부)

○ 다음의 최신 기술 동향, 연구 환경의 변화에 따라 <표 1-5>와 같이 연구 내용 변경 추진

○ 공간 빅데이터의 공간적 특성을 고려한 시각화 연구 초기단계

- 빅데이터를 분석하고 가시화하기 위한 사용 소프트웨어가 이미 널리 사용 중이나 위치정보를 포함하고 있는 공간 빅데이터를 가시화하기 위한 연구는 아직 초기 단계
- 공간 빅데이터를 기하학적, 토폴로지 및 의미론적 정보에 따라 평균을 기준으로 구분하여 공간적 다양성과 이질성을 분석하는 연구 진행 중
- 공간 빅데이터의 공간적 특성을 고려하여 머신러닝과 딥러닝 기술을 적용하는 분석 알고리즘을 개발함으로써 가시화 측면에서 그 특성이 두드러질 수 있는 가시화 알고리즘의 개발 및 이를 제공하는 모듈을 추가로 개발

<표 1-5> 연차별 연구개발계획서 대비 연구내용 변동사항(2세부)

기존 변경 삭제 신규

구분	핵심연구성과 (level1)		단위 성과 (level2)		연차					
					1차 년도	2차 년도	3차 년도	4차 년도	5차 년도	6차 년도
2세부	E	공간 빅데이터 소셜 분석 기술 개발	E-1	공간 빅데이터 분석모델 현황분석 및 요구사항 정의	기존					
			E-2	데이터 특성별 분석모델 설계서	삭제					
			E-3	공간 빅데이터 분석 목표모델 정의	기존					
			E-4	이기종 공간DB 특성 분석		기존				
			E-5	이기종 DB간 데이터 가공, 변환, 필터링 기술 개발		기존				
			E-6	다중 레이어를 위한 캐싱/프리패칭/버퍼링 기술 개발		기존				
			E-7	공간 빅데이터 정합 및 주제도 분석 기술 개발			기존			
			E-8	공간 빅데이터 분포 분석 기술 개발			기존	변경		
			E-9	공간 빅데이터 차트 분석 기술 개발				기존		
			E-10	공간 빅데이터 분포 분석 고도화 기술 개발				기존	변경	
			E-11	공간 빅데이터 차트 분석 고도화 기술 개발					신규	
			E-12	공간 빅데이터 분석결					신규	

기존 변경
삭제 신규

구분	핵심연구성과 (level1)		단위 성과 (level2)		연차					
					1차 년도	2차 년도	3차 년도	4차 년도	5차 년도	6차 년도
				과 가시화 기술 개발						
			E-13	공간 빅데이터 분석결과 가시화 기술개발(정적 가시화)					신규	
			E-14	공간 빅데이터 정합 및 주제도 분석 모듈 개발			신규			
			E-15	공간 빅데이터 분포 분석 모듈 개발				신규		
			E-16	공간 빅데이터 차트 분석 모듈 개발				신규		
			E-17	공간 빅데이터 분포 분석 모듈 고도화					신규	
			E-18	공간 빅데이터 차트 분석 모듈 고도화					신규	
			E-19	공간 빅데이터 분석결과 가시화 모듈 개발(정적 가시화)					신규	
			E-20	공간 빅데이터 분석결과 가시화 모듈 개발						신규
			E-21	공간 빅데이터 분석 모듈 사업단 적용 및 실증						신규
	F	공간 빅데이터 Batch 분석 컴포넌트 기술개발	F-1	공간 빅데이터 통계기반 요구조건 분석	기존					
			F-2	공간 빅데이터 비통계기반 요구조건 분석	기존					
			F-3	공간 빅데이터 분석 컴포넌트 인터페이스 설계	기존					
			F-4	공간 빅데이터 분석 컴포넌트 아키텍처 설계	기존					
			F-5	공간 빅테이더 기본분석 기능 선정 및 개발 기능 결정		기존				
			F-6	통계기반 분석 컴포넌트 설계		기존				
			F-7	비통계기반 분석 컴포넌트 설계		기존				
			F-8	공간 빅데이터 분석 컴포넌트 프로토타이핑		기존				
			F-9	공간빅데이터기본분석기능개선			기존			
			F-10	통계기반 분석 성능 향상			기존			
			F-11	비통계기반 분석 성능 향상			기존			
			F-12	공간 빅데이터 분석 컴포넌트 기능 고도화			기존	변경	변경	
			F-13	공간 빅데이터 기본 분석 기능 추가				기존		
			F-14	통계기반 분석 성능 최적화				기존		
			F-15	비통계기반 분석 성능 최적화				기존		

기존 변경
삭제 신규

구분	핵심연구성과 (level1)		단위 성과 (level2)		연차					
					1차 년도	2차 년도	3차 년도	4차 년도	5차 년도	6차 년도
			F-16	공공/민간수요기반Batch 분석시나리오설계					신규	
			F-17	공간 빅데이터 분석 Batch 솔루션 설계					신규	
			F-18	공간 빅데이터 분석 Batch 솔루션 구현				신규	신규	
			F-19	공간 빅데이터 분석 솔루션 사업단 실증					신규	
	G	이동객체데이터 Batch 분석 컴포넌트	G-1	이동객체데이터 통계기반 요구조건 분석	기존					
			G-2	이동객체데이터비통계기반요구조건분석	기존					
			G-3	이동객체데이터분석컴포넌트인터페이스설계	기존					
			G-4	이동객체데이터분석컴포넌트아키텍처설계	기존					
			G-5	공간빅테이더기본분석기능선정및개발기능결정		기존				
			G-6	통계기반분석컴포넌트설계		기존				
			G-7	비통계기반분석컴포넌트설계		기존				
			G-8	이동객체데이터 분석 컴포넌트 프로토타이핑		기존				
			G-9	이동객체데이터기본분석기능개선			기존			
			G-10	통계기반 분석 성능 향상			기존			
			G-11	비통계기반 분석 성능 향상			기존			
			G-12	이동객체데이터 분석 컴포넌트 기능 추가 및 최적화			기존			
			G-12	이동객체데이터 분석 컴포넌트 기능 고도화				신규	신규	
			G-13	이동객체데이터 기본 분석 기능 추가				기존		
			G-14	통계기반 분석 성능 최적화				변경	변경	
			G-15	비통계기반 분석 성능 최적화				변경	변경	
			G-16	공공/민간수요기반Batch 분석시나리오설계						신규
			G-17	이동객체데이터 Batch 솔루션 설계						신규
			G-18	이동객체데이터 분석 Batch 솔루션 구현					신규	신규
			G-19	이동객체 데이터 분석 솔루션 사업단 실증						신규
	H	공간 빅데이터 CEP	H-1	공간 빅데이터 Interactive 분석 컴포넌	기존					

기존 변경
삭제 신규

구분	핵심연구성과 (level1)		단위 성과 (level2)		연차					
					1차 년도	2차 년도	3차 년도	4차 년도	5차 년도	6차 년도
				트 요구사항 정의서						
			H-2	공간 빅데이터 Interactive 분석 컴포넌트 설계서	기존					
			H-3	CQL 질의 메모리 상주 색인 및 저장 모듈		기존				
			H-4	공간 빅데이터 기반 복합 이벤트 스케줄러 모듈		기존				
			H-5	CQL 질의 클러스터 내 최적 할당 모듈			기존			
			H-6	공간 빅데이터 기반 질의 처리기 모듈			기존			
			H-7	공간 빅데이터 OLAP 데이터 로딩 모듈			기존	변경		
		및 OLAP 엔진 개발	H-8	공간 빅데이터 OLAP 드릴다운/롤업 모듈				기존		
			H-9	공간 빅데이터 OLAP 대화형 분석기 모듈				기존	변경	
			H-10	공간 빅데이터 기반 복합 이벤트 네트워크 생성 모듈				기존	변경	
			H-11	공간 빅데이터 기반 복합 이벤트 처리기 UI/UX					신규	신규
			H-12	공간 빅데이터 OLAP UI/UX					신규	
			H-13	공간 빅데이터 Interactive 분석 콤포넌트 테스팅 및 튜닝						신규
			H-14	공간 빅데이터 Interative 분석 콤포넌트 시범 서비스 적용 지원						신규

(다) 공간 빅데이터 서비스 활성화 및 실증 연구 (3세부)

○ 다음의 최신 기술 동향, 연구 환경의 변화에 따라 <표 1-6>와 같이 연구 내용 변경 추진

○ 공간 빅데이터 서비스의 활용을 저해하는 기술적 장벽을 낮추고, 연구성과의 시의성을 높일 수 있는 계기 마련

- 공간 빅데이터를 보유한 관계기관들은 관련기술에 대한 전문지식과 데이터 활용에 대한 기술적 장벽이 높아 공간 빅데이터가 갖는 잠재적 가치를 충분히 활용하지 못하고 있는 실정임
- 공간 빅데이터 서비스를 범용적으로 활용할 수 있는 기술적 지원방안을 구체적으로 제시하여 관련기술의 이해와 활용을 저해하는 기술적 장벽을 낮춤
- 공간 빅데이터 서비스 기술을 필요로 하는 관계기관과의 협의를 통하여 구체적인 기술적 적용방안을 마련, 연구 성과의 시의성을 높일 수 있는 계기 마련

○ 공간 빅데이터 서비스의 범용성 확보를 위한 수요분석과 구체적인 구현방안 마련 필요

- 공간 빅데이터 서비스의 잠재적인 공급처와 수요처인 관계기관과의 협의를 통하여 공간 빅데이터 서비스를 범용적으로 지원하는 공통모듈의 지원범위 및 주요기능 도출
- 공통모듈의 주요기능을 현재 구현중인 공간 빅데이터 서비스 실증 시스템에 적용하는 구체적인 방안 제시

<표 1-6> 연차별 연구개발계획서 대비 연구내용 변동사항(3세부)

기존 변경 삭제 신규

구분	핵심연구성과 (level1)		단위 성과 (level2)		연차					
					1차 년도	2차 년도	3차 년도	4차 년도	5차 년도	6차 년도
3세부	I	공간 빅데이터 플랫폼 활용 시스템 구현 방안	I-1	V-World 융합기반 공간 빅데이터 플랫폼의 서비스 프레임워크 도출	기존					
			I-2	서비스 프레임워크 구성요소 개발방안	기존					
			I-3	서비스 프레임워크 운영을 위한 정책적, 제도적 방안 제시		변경	변경			
			I-4	서비스 프레임워크 활용 가이드라인 개발		기존	변경			
			I-5	공간 빅데이터 플랫폼 활용 가이드 개발			기존			변경
			I-6	핵심 서비스모델 기반			기존			변경

기존 변경
삭제 신규

구분	핵심연구성과 (level1)		단위 성과 (level2)		연차					
					1차 년도	2차 년도	3차 년도	4차 년도	5차 년도	6차 년도
				활용 가이드 적용 예시						
			I-7	공간 빅데이터 플랫폼 활용을 위한 활용가이드 제시				삭제		
			I-8	핵심 서비스 모델을 기반으로 한 활용가이드 적용예시 제공				삭제		
	J	공간 빅데이터 핵심 서비스 모델	J-1	공공/민간 부문, 서비스 유형별 공간 빅데이터 서비스 현황 수집 및 분석	기존					
			J-2	국내외 공간 빅데이터 활용사례 분석을 통한 기존 서비스와 공간 빅데이터 서비스와의 차별화 방안 도출	기존					
			J-3	핵심 서비스 유형 및 서비스 모델 개발(1)		기존				
			J-4	핵심 서비스 유형 및 서비스 모델 개발(2)		기존	변경			
			J-5	공공/민간부문 공간 빅데이터 활용을 위한 비즈니스 모델 분석 및 개발			기존	변경		
			J-6	공공/민간부문 공간 빅데이터 활용을 위한 서비스 시나리오 개발			기존	변경	변경	
			J-7	서비스 구현을 위한 시스템 아키텍처 설계				기존	변경	
			J-8	핵심서비스의 주요기능 도출				기존	변경	변경
	K	공간 빅데이터 확장 Open API 서비스	K-1	V-World에서제공하는 OpenAPI사이트를통한현황파악	기존					
			K-2	Open API 국내외 적용 사례 조사	기존					
			K-3	공간 빅데이터 Open API 서비스를 위한 공간 빅데이터 콘텐츠 구성 및 인터페이스 설계		기존				
			K-4	공간정보와의 연계범위 및 API 종류 설계		기존				
			K-5	기능성과 사용성을 고려한 공간 빅데이터 확장 서비스를 위한 Open API 기능 개발			변경	변경		
			K-6	공공시스템 기반 공간 빅데이터 Open API 서비스 테스트 사이트 개발				변경	변경	
			K-7	연구성과물 성능 검증 지원 및 개선방안 연구				기존		변경
	L	공간 빅데	L-1	서비스별 UI/UX 기술	기존					

기존 변경
삭제 신규

구분	핵심연구성과 (level1)		단위 성과 (level2)		연차					
					1차 년도	2차 년도	3차 년도	4차 년도	5차 년도	6차 년도
				및 적용 사례 분석						
		이터 서비스 제공 UI/UX	L-2	공간 빅데이터 서비스 제공을 위한 UI/UX 개발 목표 및 범위 정의	기존					
			L-3	사용자 분석 및 공간 빅데이터 니즈 조사		기존				
			L-4	공간 빅데이터 서비스 제공을 위한 UI/UX 상세 설계		기존	변경			
			L-5	공간 빅데이터 서비스 제공을 위한 UI/UX 테스트 시스템 개발			기존		변경	
			L-6	V-World 기반 공간 빅데이터 Open API 서비스를 적용한 UI/UX 테스트 사이트 개발				기존		변경
			L-7	연구성과물 성능 검증 지원 및 개선방안 연구				기존		변경
	M	국토교통 분야 공간빅데이터 실증 서비스 시스템	M-1	공간 빅데이터 현황 및 서비스 구축 사례 분석	기존					
			M-2	공간정보 오픈플랫폼 실증 기술 현황 분석	기존					
			M-3	빅데이터 서비스 아키텍처 및 프레임워크 설계		기존				
			M-4	공간 빅데이터 서비스 시스템 실증 기술 설계		기존				
			M-5	저장, 관리, 분석, 가시화 등 개별 서비스 설계		기존				
			M-6	공간 빅데이터 서비스 통합 및 연계 기술 설계		기존	변경			
			M-7	공간 빅데이터 서비스 통합 기술 프로토타입 개발			기존			
			M-8	공간 빅데이터 서비스 시스템 실증 기술 개발			기존	변경		
			M-9	저장, 관리, 분석, 가시화 등 개별 서비스 기술 개발			기존		변경	
			M-10	공간 빅데이터 서비스 통합 기술 개발				기존	변경	변경
			M-11	공간 빅데이터 서비스 시스템 구축						신규
	N	V-World 융합 기반 행정, 통계 등 타 부처 공간빅데이터 실증 서비스 시스템	N-1	국내 공공 공간 빅데이터 구축 및 시각화 기술 현황 분석	기존					
			N-2	공공 공간 빅데이터 서비스 연계 및 확장 기술 설계(1)		기존				
			N-3	공공 공간 빅데이터 서비스 연계 및 확장 기술 프로토타입 개발		변경	기존			

기존 변경
삭제 신규

구분	핵심연구성과 (level1)		단위 성과 (level2)		연차					
					1차 년도	2차 년도	3차 년도	4차 년도	5차 년도	6차 년도
			N-4	공공 공간 빅데이터 서비스 연계 및 확장 기술 설계(2)			변경	기존		
			N-5	공공 공간 빅데이터 서비스 연계 및 확장 기술 개발				신규	신규	
			N-7	공공 시스템 기반 공간 빅데이터 서비스 시스템 확장 구축						신규

2. 연구수행내용 및 성과

❍ 연구개발 목표를 수행하여 개발한 핵심 연구성과는 <표 2-1>과 같으며, 자세한 연구수행내용은 다음 '가. 연구수행내용' 에서 제시함

<표 2-1> 핵심연구성과

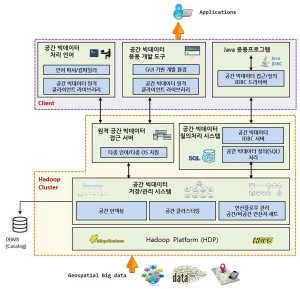
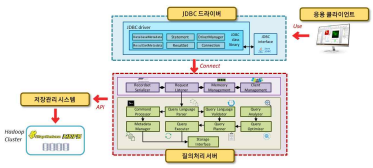
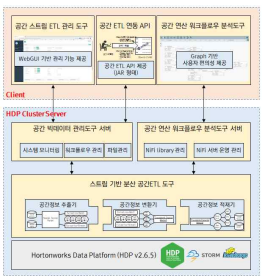
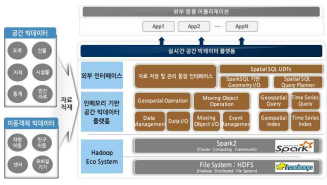
과제명		핵심성과물(level1)
1세부	공간빅데이터 저장관리 인프라 기술 개발	• (핵심성과 A) Hadoop 기반 공간 빅데이터 저장 프레임워크
		• (핵심성과 B) 공간 빅데이터 질의처리 엔진
		• (핵심성과 D) 스트림 기반 분산 공간 ETL 도구
		• (핵심성과 O) 인메모리 기반 공간빅데이터 저장관리 프레임워크
2세부	공간 빅데이터 분석 및 가시화 기술 개발	• (핵심성과 E) 공간 빅데이터 소셜 분석 기술, 공간 빅데이터 분석 및 가시화 컴포넌트 구현
		• (핵심성과 F) 공간 빅테이터 Batch 분석
		• (핵심성과 G) 이동객체테이터 Batch 분석
		• (핵심성과 H) 공간 빅데이터 Interactive 분석 기술 개발
3세부	공간 빅데이터 서비스 활성화 및 실증 연구	• (핵심성과 I) 공간 빅데이터 플랫폼의 융복합 활용방안 개발
		• (핵심성과 J) 공간 빅데이터의 공공/민간 부문 핵심 서비스 모델 도출 및 제공 방안 개발
		• (핵심성과 K) 공간 빅데이터 확장 Open API 서비스
		• (핵심성과 L) 공간 빅데이터 서비스 제공 UI/UX
		• (핵심성과 M) 공공시스템기반 국토교통분야(부동산, 지적, 교통 등) 공간 빅데이터 실증 서비스 시스템
		• (핵심성과 N) 공공시스템기반 타 부처 공간 빅데이터 실증 서비스 시스템

가. 연구수행내용

(1) 공간 빅데이터 저장관리 인프라 기술 개발

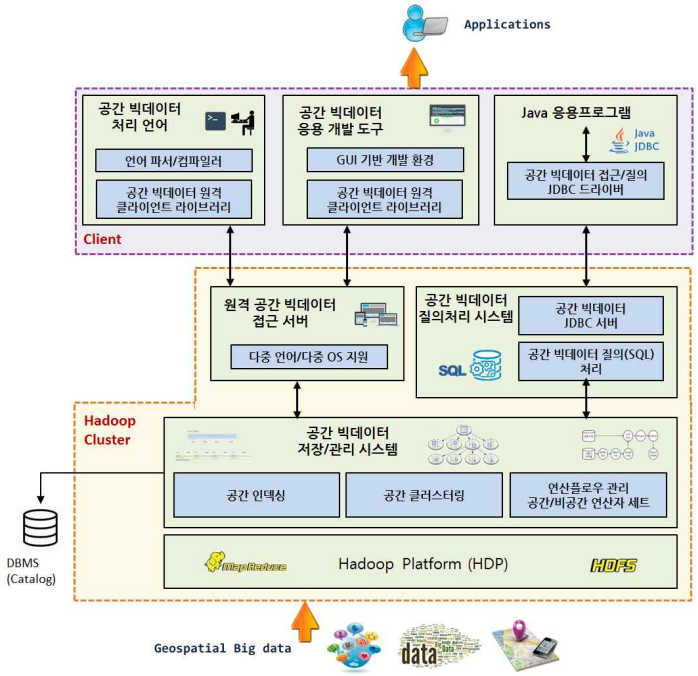
○ 1세부의 연구내용을 수행하여 다음과 같은 4개의 핵심성과를 도출함

<표 2-2> 1세부 연구내용 요약

연구 핵심 성과		수행 기관	연차	연구수행 내용 및 성과
A	Hadoop 기반 공간 빅데이터 저장 프레임워크	한국전자통신연구원	1~6	Hadoop 플랫폼에서 공간 빅데이터의 분석 및 응용 서비스를 개발할 수 있게 지원하는 공간 빅데이터 저장/탐색/검색 API와 저장관리 엔진 및 도구들을 개발함 
B	공간 빅데이터 질의처리 엔진	한국전자통신연구원	1~6	공간 빅데이터 저장 프레임워크에 저장된 공간 빅데이터에 대한 질의처리를 지원하는 질의처리 엔진 및 이를 응용프로그램에서 접근하여 사용할 수 있는 응용프로그램 인터페이스를 개발함 
D	스트림 기반 분산 공간 ETL 도구	아이씨티웨이	1~6	배치 및 스트림 형태로 제공되는 공간 빅데이터를 신속, 정확하게 Hadoop 플랫폼에 적재하기 위해 공간 빅데이터 추출/변환/적재 API 및 사용자 편의성을 위해 공간 ETL 관리 도구들을 개발함 
O	인메모리 기반 공간빅데이터 저장관리 프레임워크	포스웨이브	1~6	Hadoop 및 Spark 기반으로 공간 빅데이터를 인메모리상에서 효율적으로 저장 관리하고, 고속 처리 및 분석이 가능한 엔진 및 도구들을 개발함 

(가) Hadoop 기반 공간 빅데이터 저장 프레임워크(핵심성과 A)

○ Hadoop 빅데이터 플랫폼에서 고속의 스트림 방식으로 제공되는 공간 빅데이터를 활용한 분석 및 응용 서비스를 개발하기 위해 제공되는 공간 빅데이터 저장, 탐색 및 검색을 위한 API와 이를 구현한 엔진, 도구들의 집합



<그림 2-1> Hadoop 기반 공간 빅데이터 저장 프레임워크(핵심성과 A)

○ 특징

- 공간 빅데이터 처리를 위하여 공통적으로 사용되는 여러 접근 패턴을 구현한 **단위 연산자(operator) 기반으로 개발되어 활용성과 확장성이 높음**
- 데이터를 스트림 기반 방법으로 처리하며 데이터 및 연산의 특성에 따라 공간 빅데이터 연산이 맵리듀스 방법 또는 인메모리 방법으로 선택적으로 실행되어 **공간 빅데이터 처리 속도가 우수함**
- 저장관리 기능을 사용하기 위한 API 외에도 원격 접근 인터페이스, 응용개발 도구, 스크립트언어를 지원하여 **공간 빅데이터 분석 응용 개발 편의성이 높음**

○ 시스템 구조

- 저장관리 시스템 엔진은 HDFS에 공간 데이터 및 공간 인덱스를 저장하며, 공간/비공간 연산자 및 이들의 조합을 Hadoop 프레임워크 환경에서 실행함
- 원격 공간 빅데이터 접근 서버는 다양한 언어와 운영체제에서 공간 분석 응용을 개발할 수 있는 원격 인터페이스를 제공함
- 공간 빅데이터 응용개발 도구는 GUI 기반으로 연산자를 조합하여 공간 분석 응용의 로직을 생성하고 실행할 수 있는 인터페이스를 제공함

① [단위성과 A-1] 공간 빅데이터 저장 시스템 요구사항 정의 (1차연도)

❍ 공간 빅데이터 저장관리 시스템의 요구사항을 정의함

- 공간 빅데이터 저장관리 시스템은 기존의 공간정보 처리 시스템과 최근에 각광을 받은 빅데이터 처리 기술을 접목하여 대규모의 공간 빅데이터를 효율적으로 수집, 저장, 관리할 수 있는 기능을 제공하는 것을 목표로 함
- 공간 빅데이터 저장관리 시스템의 연구목표에 기반하여 사용자 일반 요구사항과 사용자 서비스 요구사항을 포함하는 사용자 요구사항을 정의함
- 사용자 및 환경 이해를 위한 공간 빅데이터 저장 및 탐색 기능, 공간 빅데이터 색인 관리, MR 기반 공간 빅데이터 처리 기능 및 공간 빅데이터 처리 언어 등을 포함하는 시스템 요구사항을 정의함

❍ 공간 빅데이터 저장관리 시스템의 시스템 개념분석 및 목표 시나리오 설정 등을 통하여 사용자 요구사항을 정의함

<표 2-3> 공간 빅데이터 저장관리 사용자 요구사항

요구사항 ID	내용
UFR-001	1페타 바이트 이상의 데이터를 처리할 수 있어야 한다.
UFR-002	기존의 다양한 공간 정보 포맷을 사용할 수 있어야 한다.
UFR-003	스크립트 언어를 이용하여 저장 시스템에 저장된 데이터를 처리할 수 있어야 한다.
UFR-004	응용 개발용 인터페이스를 제공하여 공간 분석 응용을 사용할 수 있어야 한다.
UFR-005	기초적인 공간 분석 기능을 제공한다.
UFR-006	여러 컴퓨터를 이용하여 분산하여 데이터를 처리할 수 있어야 한다.
UFR-007	이미 구축된 데이터를 적재할 수 있어야 한다.
UFR-008	원격에서 저장 시스템에 접속하여 공간 분석을 수행할 수 있어야 한다.
UFR-009	배치 분석 뿐만 아니라 실시간 분석 형식도 지원한다.

❍ 공간 빅데이터 저장관리 시스템의 사용자 요구사항을 기반으로 개발자 관점에서의 시스템 요구사항을 정의함

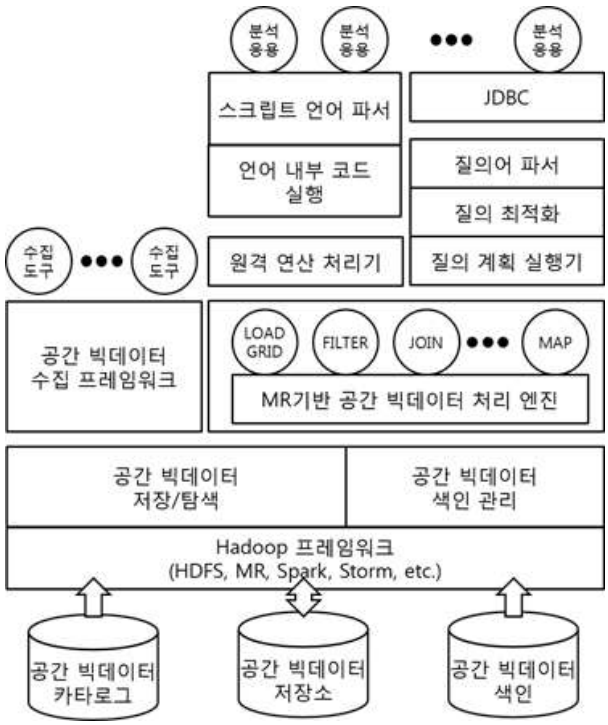
<표 2-4> 공간 빅데이터 저장관리 시스템 요구사항

요구사항 ID	내용
SFR-001	Hadoop 프레임워크 상에서 구동되어야 한다.
SFR-002	공간 정보를 HDFS를 통해 저장되고 입출력 될 수 있다.
SFR-003	JSON 형식으로 표현된 공간 정보를 인식할 수 있어야 한다.
SFR-004	WKB 형식으로 표현된 공간 정보를 인식할 수 있어야 한다.
SFR-005	WKT 형식으로 표현된 공간 정보를 인식할 수 있어야 한다.
SFR-006	GeoJSON 형식으로 표현된 공간 정보를 인식할 수 있어야 한다.
SFR-007	기존 공간 정보를 import하거나 export 할 수 있어야 한다.
SFR-008	공간 빅데이터 처리용 절차형 언어를 제공한다.
SFR-009	공간 빅데이터 언어는 컴파일되어 내부 수행 폼으로 변환되어, 나중에 바로 실행될 수 있어야 한다.
SFR-010	내부 수행 폼은 원격에서 실행될 수 있어야 한다.

요구사항 ID	내용
SFR-011	분석 응용을 위한 응용 개발용 API를 제공한다.
SFR-012	원격에서 응용 개발 API를 호출할 수 있어야 한다.
SFR-013	맵리듀스를 통한 공간 분석 작업을 수행할 수 있어야 한다.
SFR-014	공간 정보 접근용 색인 기능을 제공한다.
SFR-015	제공되는 공간 분석에 OGC 표준의 Simple Feature Access에 기술된 연산을 지원한다.
SFR-016	실시간 공간 분석을 위한 기능을 제공한다.
SFR-017	순차적인 공간 정보 접근뿐만 아니라 임의의 위치에 공간 정보를 추가 삭제가 가능해야 한다.
SFR-018	그리드 형식의 데이터를 저장할 수 있는 기능을 제공한다.
SFR-019	벡터 형식의 공간 데이터를 저장할 수 있는 처리 기능을 제공한다.
SFR-020	그리드 형식의 공간 데이터를 처리할 수 있는 기능을 제공한다.
SFR-021	벡터 형식의 공간 데이터를 처리할 수 있는 기능을 제공한다.

② [단위성과 A-2] 공간 빅데이터 저장 시스템 개념 설계 (1차연도)

❍ 대용량의 공간 빅데이터를 Hadoop을 활용하여 생성, 저장, 접근 및 처리하는 기능을 하는 공간 빅데이터 저장 시스템의 구조를 도출함



<그림 2-2> 공간 빅데이터 저장 관리 시스템 구조

❍ 전체 시스템은 Hadoop 프레임워크를 기반으로 하여 구축되며 Hadoop 프레임워크 상에 공간 빅데이터의 저장, 탐색을 맡는 “공간 빅데이터 저장/탐색” 기능, 신속한 공간 빅데이터 접근을 지원하기 위한 “공간 빅데이터 색인 관리” 기능, 공간 빅데이터를 접근하고 조작하기 위한 “MR 기반 공간 빅데이터 처리 연산” 기능이 위치함

※ MR(Map/Reduce, 맵리듀스)은 대용량 데이터 처리를 분산 병렬 컴퓨팅에서 처리하기 위한 목적으로 구글에서 2004년 발표한 소프트웨어 프레임워크이자 프로그래밍 모델임

❍ 공간 빅데이터 저장/탐색

- HDFS에 저장된 공간 빅데이터의 저장 포맷과 이에 따른 데이터 입출력을 담당하는 서브시스템임
- 공간 빅데이터의 주요 데이터 형태인 레코드 형식과 그리드 형식의 데이터에 대한 포맷과 해당 데이터를 HDFS에 입출력하는 부분을 담당함

❍ 공간 빅데이터 색인 관리

- HDFS는 순차적인 입출력에 특화된 파일 시스템이지만 공간 데이터를 다루는 상황에서는 특정값을 갖는 공간정보를 신속히 접근하는 방법은 제공하지 않기 때문에, HDFS 기반의 색인 관리 및 색인 기반 공간정보 접근 기능이 필요함

❍ 공간 빅데이터 연산자

- 공간 빅데이터 연산자는 공간 빅데이터 저장관리 시스템에서 제공하는 데이터를 손쉽게 접근 및 처리할 수 있도록 공통적으로 사용되는 여러 접근 패턴을 구현한 모듈이며 사용자/응용 들의 기본 인터페이스임
- 모듈화된 공간 빅데이터의 저장 및 관리를 제공하기 위해 “공간 빅데이터 처리 연산자” 개념과 연산자를 수행하는 “연산자 처리기” 개념을 도입함

❍ 공간 빅데이터 원격 연산 처리

- 공간 빅데이터 처리 연산자는 Hadoop이 설치된 곳에서 수행되기 때문에, 원격의 응용프로그램이 이 연산을 사용하기 위해서는 연산자를 원격에서 호출할 수 있도록 하는 기능이 필요함
- 이 기능은 “원격 연산 처리기” 부분에서 원격 함수 호출 (RPC: Remote Procedure Call) 기법을 사용하여 담당함

❍ 공간 빅데이터 처리 언어

- 응용 개발자가 API 기반의 인터페이스인 공간 빅데이터 연산자를 직접적으로 호출하여 응용을 개발할 수도 있지만, 응용 개발의 편의 및 생산성 향상을 위해 스크립트 언어 기반의 공간 빅데이터 처리 언어를 제공함
- 공간 빅데이터 처리 언어는 이러한 문제를 해결하기 위한 서브시스템으로 스크립트 언어를 제공하여 이 언어를 이용하여 연산자를 직접 호출하지 않고도 간단한 스크립트 명령을 통해 공간 빅데이터의 접근, 처리, 분석을 수행할 수 있도록 함

③ [단위성과 A-3] 공간 빅데이터 저장 시스템 인터페이스 정의 (1차연도)

❍ 공간 빅데이터를 다루는 기본 연산인 CRUD(Create, Read, Update, Delete) 인터페이스를 정의함

<표 2-5> 공간 빅데이터 CRUD 인터페이스

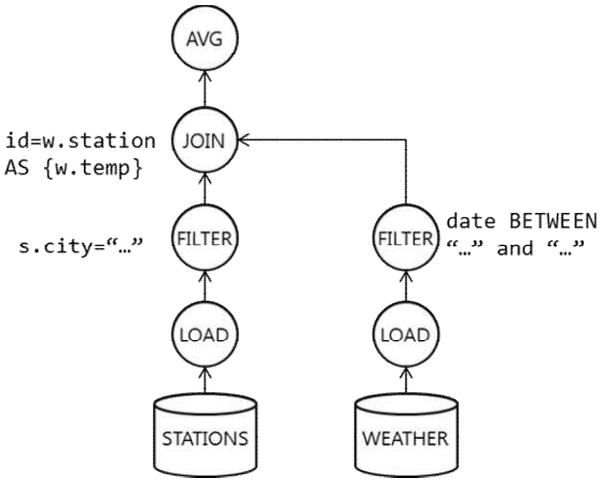
함수 Signature	주요내용
Object getAttribute(int idx)	주어진 번호(Idx)에 해당하는 번째의 필드 값을 반환한다.
Object getAttribute(String name)	주어진 이름에 해당하는 필드 값을 반환한다.
int getAttributeCount()	하나의 레코드에 포함된 필드의 개수를 반환한다.
List getAttributes()	레코드에 포함된 모든 필드 값의 리스트를 반환한다.
Object getDefaultGeometry()	레코드에 포함 Geometry 정보를 반환한다.
SimpleFeatureType getFeaturetype()	레코드의 메타 정보를 반환한다.
void setAttribute(int idx, Object value)	주어진 번호(idx)에 해당하는 필드 값을 설정한다.
void setAttribute(String name, Object value)	주어진 필드명에 해당하는 필드 값을 설정한다.
void setAttributes(List values)	주어진 필드 값 리스트를 차례대로 레코드의 필드 값으로 설정한다.
void setAttribute(Object[] values)	주어진 필드 값 배열을 차례대로 레코드의 필드 값으로 설정한다.
void setDefaultGeometry(Object geometry)	레코드의 geometry 정보를 설정한다.

❍ 공간 빅데이터에 대한 주요 공간 연산자 인터페이스를 설계함

- 공간 빅데이터 저장관리 시스템에서 제공하는 공간데이터를 손쉽게 접근 및 처리할 수 있는 기본 연산자를 설계함
- 공간 빅데이터 연산자는 공간 데이터 관련 OGC의 심플 피처 모델을 지원할 뿐만 아니라, 공간 빅데이터의 적재 및 변경에 관련된 기능도 제공함
- 주요 연산으로는 HDFS에 저장된 데이터를 적재하는 Load, 조건을 만족하는 공간 데이터만을 뽑아내는 Filter, 두 개 이상의 공간 데이터를 서로 조합하여 새로운 공간 데이터를 생성하는 Join 등을 포함함
- 연산자가 제공하는 인터페이스만 이용하여 공간 빅데이터 연산자를 활용할 수 있도록 하여 구현 투명성을 제공함

❍ 공간 빅데이터 처리 언어 시스템의 인터페이스를 설계함

- 응용 개발자가 연산자를 직접 호출하지 않고도 간단한 스크립트 명령을 통해 복잡한 빅데이터 분석을 수행할 수 있도록 함
- <그림2-3>은 설계된 언어를 통해 작성된 공간 빅데이터 처리 프로그램의 예이며, 처리언어는 공간 연산자의 트리로 구성되어 반복적으로 호출되어 실행되게 됨



<그림 2-3> 공간 빅데이터 처리 시스템 내부 표현(예시)

④ [단위성과 A-4] 공간 빅데이터 저장시스템(POC 버전) 모듈 (1차연도)

❍ 공간 빅데이터 저장 시스템의 설계를 검증하기 위한 POC(Proof-Of-Concept) 버전의 시스템을 개발함

❍ 일반 파일시스템에 저장된 공간정보를 공간 빅데이터 저장시스템의 관리하에 있는 HDFS에 저장하는 모듈을 개발하고 데이터를 화면에 출력하여 확인함

- Hadoop 환경에서 Map/Reduce 기법을 활용하여 다수의 클러스터에 작업을 분배하여 병렬적으로 작업을 처리하기 위해서는 작업에 필요한 공간데이터가 HDFS에 저장되어야 함

- Record 형식의 공간 데이터를 저장하는 방법은 각 레코드를 차례대로 기록하는 방법을 사용하며, Geometry 데이터는 WKB 포맷을 사용하여 저장함

※ WKB(Well-Known Binary)는 OGC 표준에서 정의한 공간데이터의 교환 포맷으로, 미리 정의된 바이너리 데이터 포맷으로 공간객체를 표기하는 방법임

- Shape 포맷과는 달리 Geometry 정보와 기타 부가 정보를 별도로 저장하지 않고, 한 레코드에 포함된 모든 데이터를 함께 저장하는 방법을 사용함

- <그림 2-4>는 대한민국 구 행정구역 정보를 기록한 Record 형식의 데이터를 HDFS에서 저장한 뒤 읽어들여 화면에 출력한 예시를 보여줌

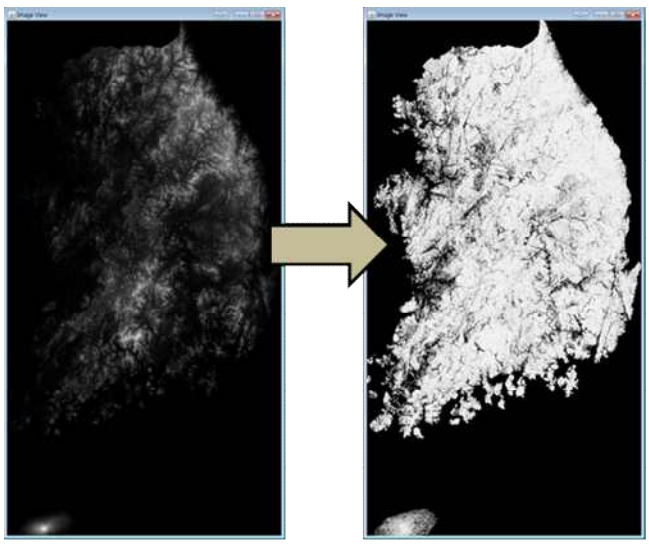
❍ 설계된 공간 빅데이터 처리 연산 모듈 중 시스템 설계 검증에 활용될 연산들을 구현하여 확인함

- 각 연산자는 Map/Reduce 작업으로 구현되었음

- <그림 2-5>는 개발된 공간 빅데이터 연산 중 하나인 "ExecSlope" 연산, 즉 고도(elevation) 정보의 Grid 데이터를 입력으로 받아 경사도(slope)를 계산하는 연산의 수행 결과를 보여줌



<그림 2-4> Record 형식 데이터 저장 및 접근



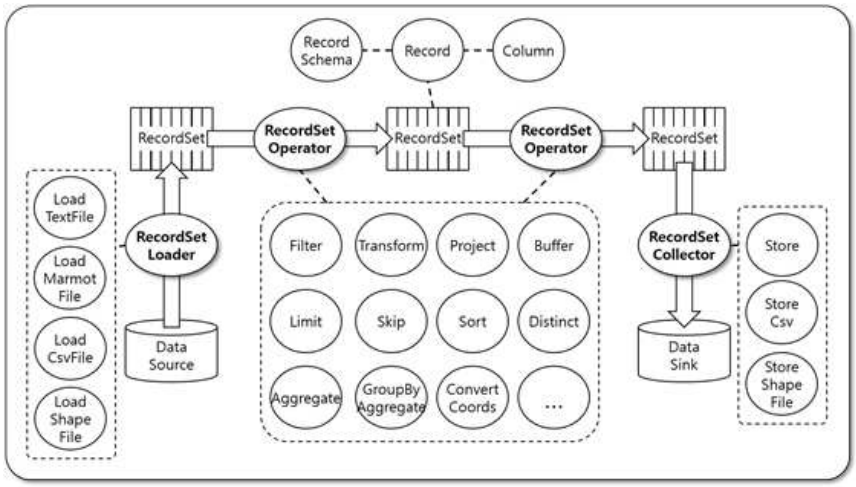
<그림 2-5> 공간 빅데이터 연산의 예: ExecSlope

⑤ [단위성과 A-6] 공간 빅데이터 처리 연산자 구현 모듈 v1.0 (2차연도)

❍ 공간 빅데이터의 저장 및 관리 기능을 제공하기 위한 레코드 연산자 기반 공간 빅데이터 처리 구조를 설계하고, 이를 구현하는 기본 공간 연산자와 비공간 연산자를 개발함

❍ 기본 공간 연산자(basic spatial operators) 및 비공간 연산자(non-spatial operators)를 개발함

- 실시간 공간 레코드 스트림 및 연산자 기반 공간 빅데이터 처리 구조를 개발함



<그림 2-6> 연산자 기반 공간 빅데이터 처리 구조

- 기본 공간 연산자 9종 및 비공간 연산자 14종을 개발함

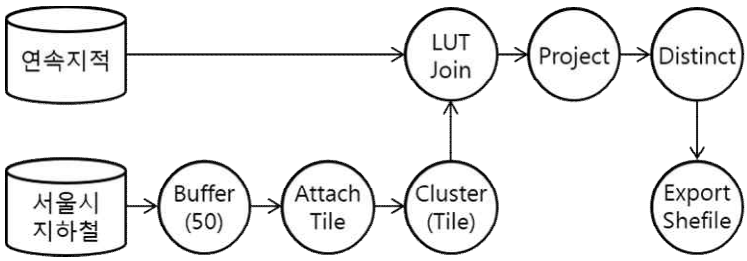
비공간 연산자 (14종)	Unary 연산 (10종)	Aggregate, Filter, Project, GroupByAggregate, Distinct, Skip, Limit, NoOp, Shard, Transform
	Loader 연산 (2종)	LoadMarmotFile, LoadTextFile
	Collector 연산 (2종)	Store, StoreAsCsv
공간 연산자 (10종)	Unary 연산 (8종)	AttachMapTile, Buffer, ConvertCoordinates, AggrUnion, Bbox, AggrConvexHull, Union, Intersection,
	Loader 연산 (1종)	LoadShapefile
	Collector 연산 (1종)	StoreShapefile

<그림 2-7> 공간 빅데이터 처리 연산자

- 개발된 연산자에 대한 동작 정확도 확인을 위해 JUnit을 활용하여 연산자 동작 정확도를 검증함

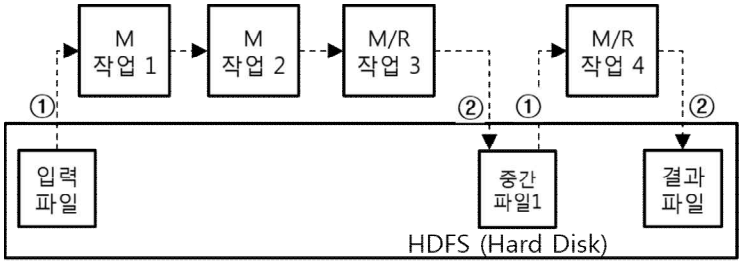
○ 공간 빅데이터 연산 플로우 (복합연산) 실행기 개발

- 하둡에서 수행하기 위한 복합 연산 플로우를 Map/Reduce 작업들로 자동으로 변환하는 기능을 구현함



<그림 2-8> 복합 연산의 예: 지하철역 50m 이내 지적 검색

- 최적화된 Map/Reduce 변환 방식으로 기존 Map/Reduce 처리방법 대비 처리 성능을 향상시킴



<그림 2-9> MR 작업 최적화

○ 공간 빅데이터 연산 성능 평가

- 개발된 공간 빅데이터 연산 및 연산 실행기의 성능을 주요 유사 시스템들과 비교 평가함
- 평가 환경은 4대의 데이터 노드, 1대의 네임 노드, 그리고 1대의 Yarn Resource Manager로 구성된 하둡 시스템을 구축하여 실시함
- 성능 평가 대상으로는 '2014년도 국토교통부 공간빅데이터체계구축사업단'에 납품된 공간 빅데이터 엔진, 미네소타 대학에서 개발한 SpatialHadoop v.2.3, PostGIS 2.1, PIG 및 HIve를 선정하여 평가함
- 성능 검증용 데이터로는 연속지적도를 포함한 주요 국토부 보유 공간 DB와 VWorld 로그 데이터, 그리고 Twitter와 같은 SNS 로그 데이터를 활용함
- 성능 평가 결과는 대규모(128MB 이상) 공간 데이터에 대한 분석의 경우 개발된 시스템은 비교 대상인 시스템들에 비해 우수한 성능을 보임
- 소규모(128MB 이하)의 데이터 분석의 경우는 전통적인 시스템인 PostGIS 보다는 처리 성능은 떨어지나, 하둡 기반의 비교 대상 시스템들에 비해서는 월등한 성능을 보임
- 공간 빅데이터외에 일반 빅데이터에 대한 처리도 기존 PIG와 HIVE와 같은 기존 시스템들과 유사한 성능을 보여, 일반 빅데이터 분석용으로도 사용 가능함을 알 수 있음

⑥ [단위성과 A-10] 공간 빅데이터 색인 모듈(1) (3차연도)

○ 공간 빅데이터에 대한 R-Tree 색인을 구축함에 있어 색인을 신속하게 구축하고 색인 내부의 자료구조를 효율적으로 생성하여 검색 성능을 높임

- 샘플링을 통해 공간 빅데이터의 분포를 파악하고 이를 바탕으로 효율적인 구조의 R-Tree 색인을 구축함
- 이 과정에서 Hadoop을 이용하여 공간 빅데이터를 병렬적으로 처리하고 색인을 동시에 구축함으로써 색인 구축에 따르는 소요시간을 줄임

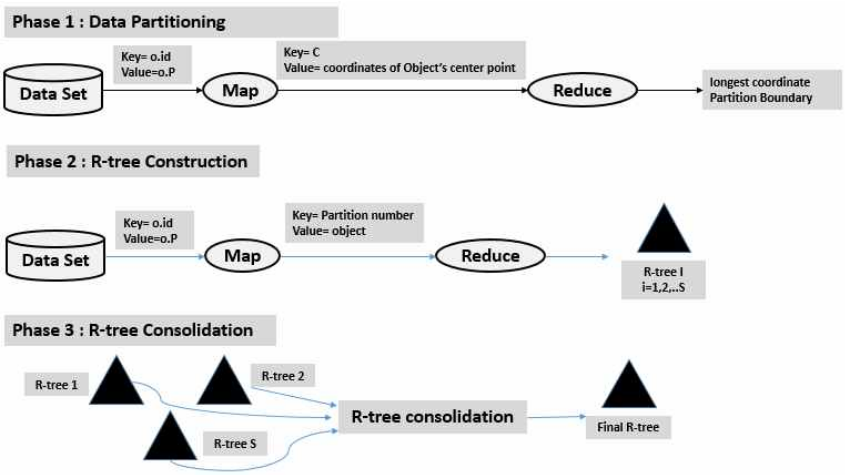
○ 색인 구축은 다음 3단계로 이루어짐

- [샘플링을 통한 파티션] 단계
  - 공간 빅데이터 집합에 대해 샘플링을 통해 데이터들의 공간적 분포를

파악하고 이것을 균등한 개수의 데이터 파티션으로 분할함

- 파티션 방식은 Z-Order 곡선 또는 STR 기법을 적용함
- Hadoop의 Map/Reduce를 활용하여 샘플링 데이터에 대해 병렬적으로 수행함

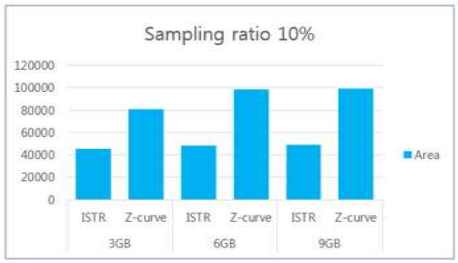
- [지역 R-Tree 색인 구축] 단계
  - 빅데이터 전체 집합을 앞 단계에서 분할한 파티션 단위로 읽고 각각의 파티션에 대해 지역 R-Tree 색인을 구축함
  - Hadoop의 Map/Reduce를 사용하여 전체 데이터 집합에 대해 병렬적으로 수행함
- [전역 R-Tree 색인 구축] 단계
  - 이전 단계에서 구축된 지역 R-Tree들을 병합하여 전역 R-Tree 색인을 구축하는 단계이며 지역 R-Tree들의 최상위 노드들을 자식 노드로 하는 최상위 노드를 새로 구축함



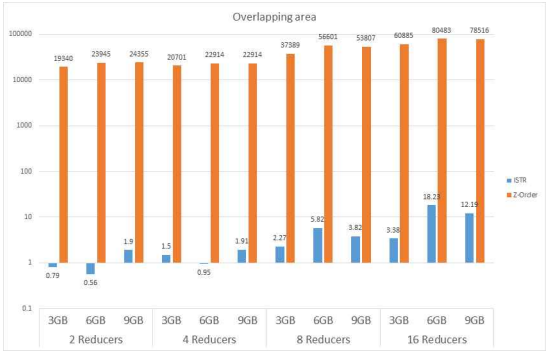
<그림 2-10> Hadoop을 이용한 R-Tree 색인 구축

○ 성능 평가

- 성능평가는 8대의 노드로 구성된 Hadoop 클러스터를 이용하여 수행함
- 기존 기법에서는 파티션을 위해 Z-Order 곡선을 사용하였으나 제안한 기법에서는 STR 기법을 개선한 기법(ISTR)을 적용함
- 전체적으로 기존 기법에 비해 R-Tree에서 검색에 사용되는 색인 공간의 크기는 50% 이상 줄어들었으며, 특히 최상위 노드의 MBR들 간의 겹침은 90% 이상 줄일 수 있었음
- 이것은 R-Tree 검색 과정에서 방문해야 하는 노드들의 개수를 크게 줄일 수 있음을 의미하고 따라서 검색 성능이 향상됨을 나타냄



<그림 2-11> 색인 공간의 크기 비교



<그림 2-12> 겹침 영역의 크기 비교

⑦ [단위성과 A-13] 공간 빅데이터 처리 연산자 구현 모듈 v2.0 (3차연도)

❍ 공간 색인을 활용하는 공간 질의를 포함한 고급 공간 연산자 인터페이스 및 모듈을 개발함

- 색인 파일 읽기 및 쓰기 모듈을 개발함
- Map/Reduce 환경에서 색인을 활용한 공간 검색 모듈을 개발함

❍ 공간 색인을 활용하여 고속으로 수행되는 다양한 공간 조인 연산자 (클립 연산, 거리기반 연산, 교차반전 연산 등)들의 인터페이스를 설계하고 동작 모듈을 구현함

- 중첩 루프기반 공간 조인 연산자
- 클립 연산자
- 거리 기반 연산자
- 교차 반전 연산자
- 반경제한 KNN(K-Nearest Neighbors) 연산자

❍ 카탈로그(Catalog)를 개발하여 레이어 개념을 구현함

- DBMS 기반 카탈로그 저장관리 모듈 개발
- 레이어 관리 모듈 개발
- 레이어 파일 읽기 및 쓰기 모듈 개발

❍ 공간 연산자 및 색인 생성 및 활용 성능평가

- 공간 연산자의 동작 정확도 및 성능을 평가하였으며 모두 목표치를 만족함 (그림 2-13)
- 외부 기관(TTA)에 의한 시험을 통하여 신뢰성을 확보함

시험 결과

ID	시험 항목	목표치	측정치
TC1	동작 정확도 - 공간인덱스 생성	100%	100%
TC2	동작 정확도 - 공간인덱스 기반 공간 검색	100%	100%
TC3	속도 향상 - 거리기반 선택	30% 이상	68.30%
TC4	속도 향상 - 클립 분석	30% 이상	58.40%
TC5	속도 향상 - 교차 중첩	30% 이상	93.43%
TC6	속도 향상 - 공간조인 중첩	30% 이상	68.71%
TC7	질의처리 동작 정확도 - 비공간 검색 및 연산	100%	100%
TC8	질의처리 동작 정확도 - 공간 검색 및 연산	100%	100%
TC9	질의처리 동작 정확도 - 데이터셋 정확도	100%	100%

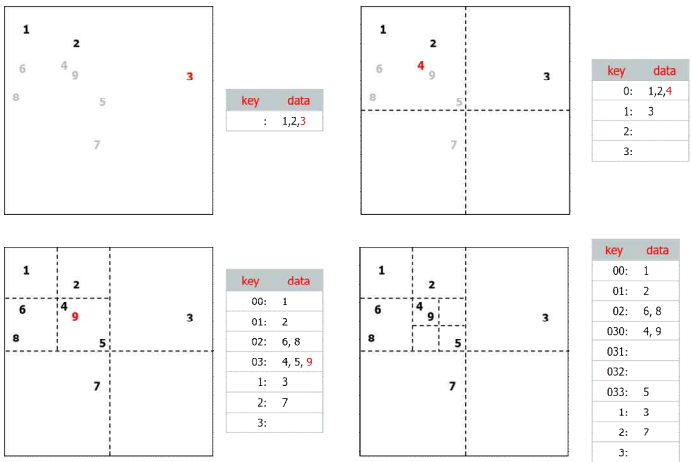
<그림 2-13> 공간 연산자 및 색인 생성 및 활용 성능평가 결과

⑧ [단위성과 A-11] 공간 빅데이터 색인 모듈(2) (4차연도)

❍ 전년도(3차년도)에 개발된 공간 색인 시스템을 Hadoop Map/Reduce 환경에서 고속으로 생성되고 처리될 수 있도록 고도화함

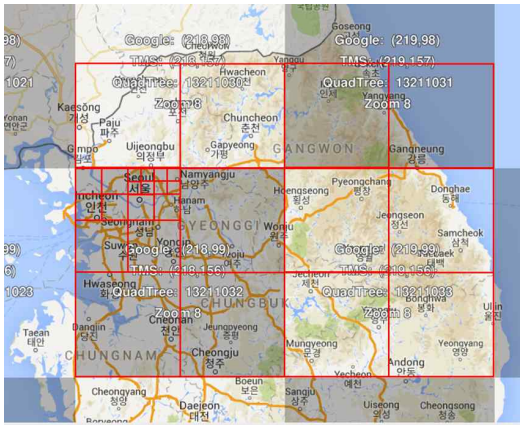
❍ 공간 빅데이터 색인 개발(2차)

- 인메모리 기반 Quad-Tree 공간색인 모듈을 개발함
- 인메모리에 특화된 경량화된 공간 색인 모듈을 개발함



<그림 2-14> 인메모리 Quad-Tree를 활용한 공간 분할의 예

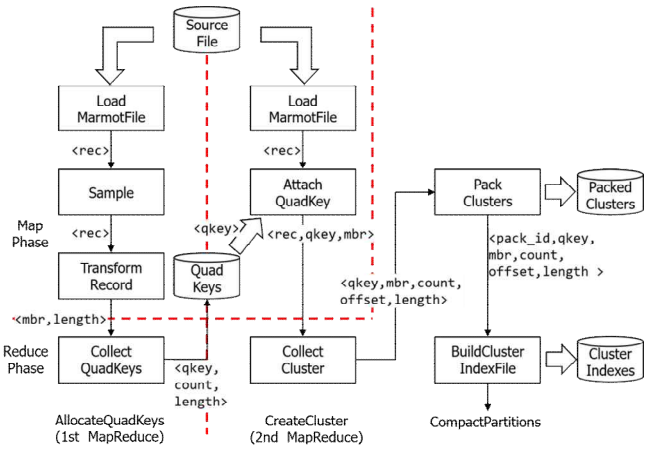
- 구글 맵 타일 방식의 공간 분할 방식을 활용하여 공간 조인 성능을 향상함



<그림 2-15> 서울지역의 공간 색인 구성의 예

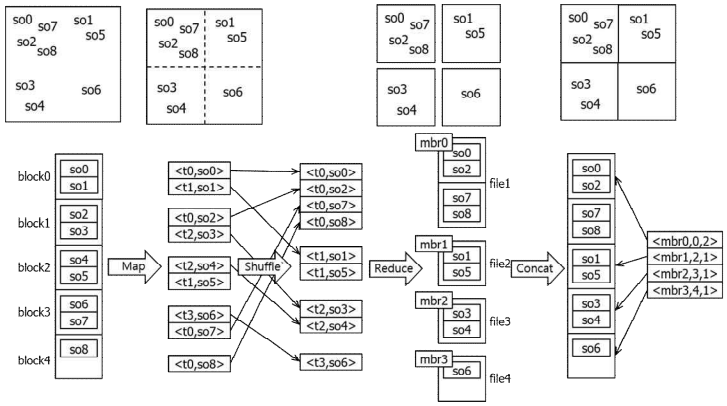
❍ Map/Reduce를 활용한 고속 공간 빅데이터 색인 구축 기술 개발

- Map/Reduce 기반 공간 인덱스 생성 기술을 개발함
- 2단계 Map/Reduce 공간 인덱스를 생성함
  - 1단계에서는 데이터중 일부를 샘플링하여 생성될 Quad-Tree Bucket 크기를 추정함
  - 2단계에서는 추정된 bucket을 기준으로 공간 인덱스를 생성함



<그림 2-16> 공간 색인 생성 관련 연산자 수행 과정

- 2단계 인덱스 생성 방법을 통하여 클러스터 인덱스(Global)와 클러스터 내 인덱스 (Local)를 생성함
- Cluster-packing을 통하여 HDFS 파일 개수를 최소화함



<그림 2-17> Map/Reduce 기반 공간 색인 생성 과정

⑨ [단위성과 A-7] 스트림 공간 빅데이터 언어 규격 정의 (5차연도)

❍ 공간 빅데이터 처리 언어의 문법 규격을 정의하여 공간 빅데이터 언어 규격 정의서를 작성함

- 공간 빅데이터 처리 언어는 공간 빅데이터 시스템에서 제공하는 API를 활용하여 공간 분석 응용 개발이 익숙하지 않은 개발자들에게 스크립트 언어를 이용하여 공간 분석 응용을 작성할 수 있도록 하는 기능을 제공함
- 공간 빅데이터 처리 언어는 공간 빅데이터 분석 로직을 구성하는 여러 공간/비공간 연산자의 모음으로 구성됨
- 공간 빅데이터 처리 언어는 스크립트 기반의 언어로서 StringTemplate 라이브러리를 활용하여 분석 로직을 구성하는 각각의 공간/비공간 연산자의 구성 템플릿을 정의하는 방법으로 규격을 정의함
- <그림 2-18>은 StringTemplate을 이용하여 ‘parse_csv’라는 비공간 연산의 규격을 정의한 예를 보여줌

```
parse_csv(header, delim=",", quote=[]) ::= <<
{ "parseCsv": {
    "delimiter": "<delim>",
    "options": {
      "headerColumn": "<header>"<if(quote)>,
      "quote": "<quote>"
      <endif>
    }
  }
}
>>
```

<그림 2-18> 공간 빅데이터 언어 규격 정의(예시)

❍ GUI 기반 공간분석 Plan 저작 언어 prototype을 개발함

- 공간 빅데이터를 처리하는 단위 연산을 순차적으로 나열한 분석 로직인 plan을 생성할 수 있음
- Operator 블록을 드래그 & 드랍하면 원하는 위치에 블록이 쌓이고, 블록을 더블클릭하면 가장 마지막에 블록이 쌓임

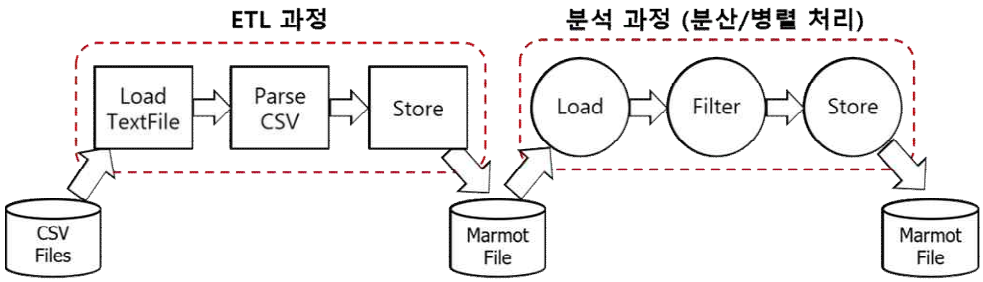


<그림 2-19> 공간분석 plan 저작 언어 프로토타입

⑩ [단위성과 A-17] 공간 빅데이터 고속 색인 생성기 모듈 (5차연도)

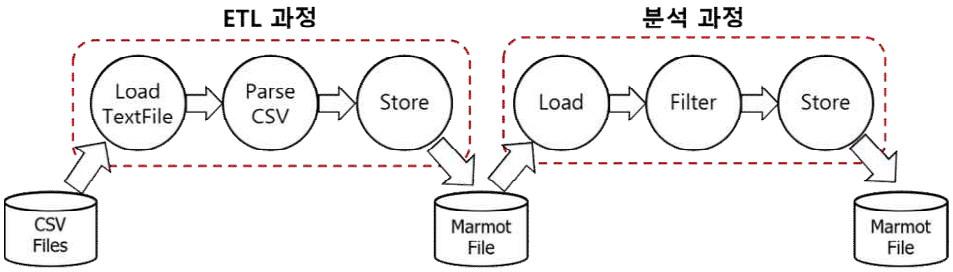
○ 데이터 적재 과정의 일부를 분석 과정에 통합하여 적재 부하를 최소화함으로써 고속으로 추가되는 대용량 로그 데이터의 적재 및 활용 방식을 개선함

- 기존 데이터 적재는 별도의 ETL 시스템을 통해 사전에 수행하는 것이 일반적인 방법이었음
  - 응용개발자의 분석에 앞서 필요한 데이터를 적재하는 방식을 사용함
  - 일반적으로 데이터 적재 및 전처리 과정은 분석 과정과 유사하게 워크플로우 방식으로 수행됨
  - 일반적으로 단일 노드에서 수행되어 scalability 문제가 발생될 여지가 크며 ETL 과정이 고속으로 처리되기에 한계를 지님
  - 특히 다수(다량)의 센서/로그 데이터를 지속적으로 적재하는 경우 시스템 bottleneck이 될 수 있음



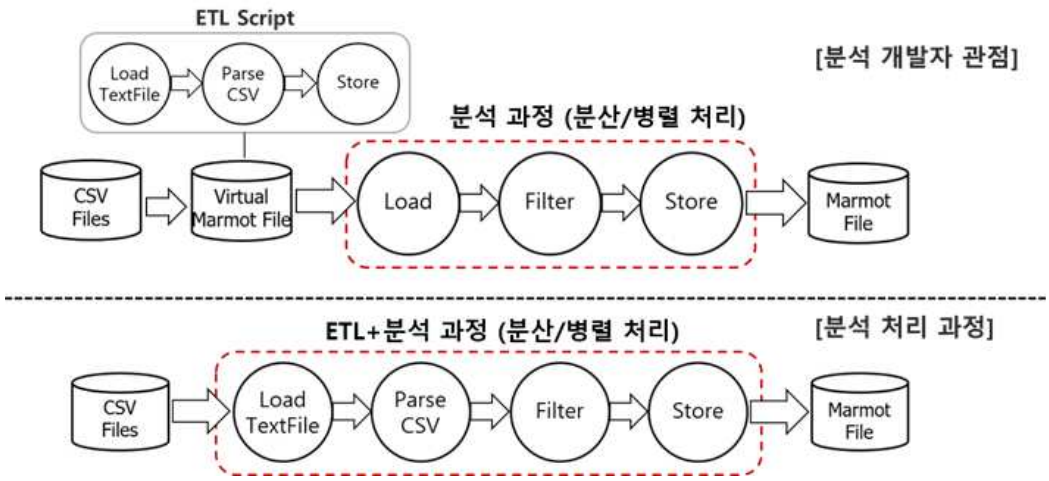
<그림 2-20> 기존의 데이터 적재 과정 및 데이터 분석 과정

- 전년도(4차년도)까지의 공간 빅데이터 플랫폼에서는 분석에서 사용하는 공간/비공간 연산과 워크플로우 수행 방식을 활용하여 ETL 과정 수행을 수행함
  - ETL 과정과 분석 과정을 동일한 방법으로 기술할 수 있는 장점을 가지므로 응용개발자가 단일 방식으로 적재 시스템과 분석 시스템을 수행할 수 있음
  - 적재 과정에서 분석 과정 중에 사용할 수 있는 다양한 분석 연산을 활용할 수 있어 복잡한 ETL 과정 수행 가능한 장점을 가짐
  - 그러나 이 방법은 적재 과정이 단일 노드에서 수행되기 때문에 다량의 데이터를 적재해야 하는 경우 시스템의 bottleneck을 유발하게 됨



<그림 2-21> 기존 공간 빅데이터 플랫폼에서의 데이터 적재와 분석 과정

- 당해연도 연구에서는 고속으로 추가되는 대용량 로그 데이터의 적재 및 활용 방식을 개선하기 위해 데이터 정제 작업을 데이터 적재시 수행하지 않고 데이터 적재 과정의 일부를 분석 과정에 통합하여 적재 부하를 최소화함
  - 데이터 적재 과정에서는 데이터 전처리 없이 단순한 원시 데이터 복사와 전처리 과정을 수행할 ETL 스크립트를 적재하는 방법으로 전처리 과정을 마침
  - 데이터 분석 과정에서 전처리 연산을 함께 분산/병렬 방식으로 수행함
  - 데이터 적재 과정에 제공되는 ETL 스크립트를 통해 응용 개발자에게는 해당 데이터의 전처리가 이미 완료된 것처럼 보이게 하여 기존 운용 중이던 분석 응용의 수정 없이 개발된 방식을 사용할 수 있도록 함



<그림 2-22> 공간 빅데이터 플랫폼에서 고속 데이터 적재와 분석 과정

○ 개발된 방법은 ETL 작업에서 데이터 전처리 과정을 생략하고 원시 데이터의 적재(복제)만을 통해 적재 작업이 신속히 완료될 수 있는 장점을 가짐

- ETL 전처리 과정과 분석 과정이 통합되어 하둡 플랫폼에서 분산/병렬 방식으로 실행하여 대용량 센서/로그 데이터 전처리 소요시간이 감소됨
- 개발된 방법의 성능 평가를 실시하였음. 성능 평가 환경은 14대의 데이터 노드를 구성된 하둡 플랫폼에서 실시되었음
- (분석 기능 선정 기준) 공간 빅데이터 적재 과정에서 가장 많이 사용하는 공간 데이터 형식인 CSV 파일을 성능 실험에 사용함. 개발 기술은 대용량 데이터의 효과적인 적재를 목표로 하기 때문에, 연구단에서 보유하고 있는 가장 큰 CSV

데이터인 ‘도로교통안전공단’에서 제공한 DTG 데이터 31.3GB를 대상으로 실험을 실시하였음. 또한 개발된 방법이 갖는 추가 오버헤드를 측정하기 위해 분석 모듈의 복잡도에 따라 ‘count’, ‘버퍼 추정’, 그리고 ‘공간 조인’을 활용한 분석 시간도 함께 측정함

- <그림 2-23>에서 보면 약 31GB 크기의 CSV 파일을 적재하는 경우 기존의 방법에서는 약 24분 47초의 시간이 소요된 반면 제안된 방법에서는 약 5분 27초가 소요되어 약 4.5배의 성능 차이를 보임

데이터: DTG (31.3GB)

작업	기존 ETL	지연 ETL	비율
적재	24:47	05.27	0.22 (4.5x)
count	03.06	04.43	1.52
버퍼 추정	06.17	07.36	1.21
공간조인	12.31	14.29	1.16

<그림 2-23> 데이터 적재 방식에 따른 성능 비교

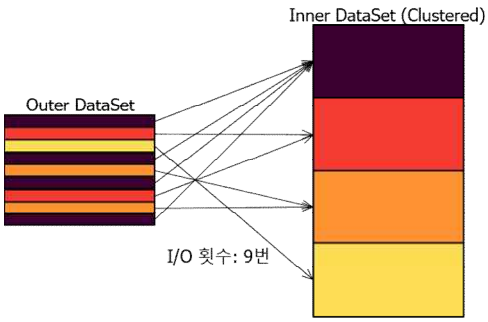
- (결과 의미) 본 성능 실험을 통해 개발된 기술을 활용하면 대규모 공간 데이터를 적재할 때, 그 소요 시간이 기존의 방법보다 월등히 줄어드는 것(4배 이상)을 알 수 있음. 이는 적재 과정 중 수행되는 데이터 정제 작업을 분석 과정과 함께 Hadoop에서 분산/병렬로 처리하여 얻는 결과인 것을 알 수 있음
- (결과의미) 개발된 방법은 데이터 분석 과정에 추가 부하를 유발하기 때문에, 분석 작업 수행 성능이 일부 저하되는 것을 확인할 수 있었음 (소요 시간 기준 1.16배 ~ 1.52배 증가). 수행하는 분석 작업의 복잡도가 증가 (count -> 공간조인)함에 따라 성능 저하 효과는 점차 줄어드는 것을 볼 수 있었음. 일반적인 공간 분석은 ‘공간조인’ 연산 정도의 복잡도를 갖는 다수의 연산으로 구성되기 때문에 개발된 기술로 유발되는 추가 부하는 그리 크지 않을 것으로 예상됨

⑪ [단위성과 A-20] 색인 기반 고속 공간 조인 연산자 모듈 (5차연도)

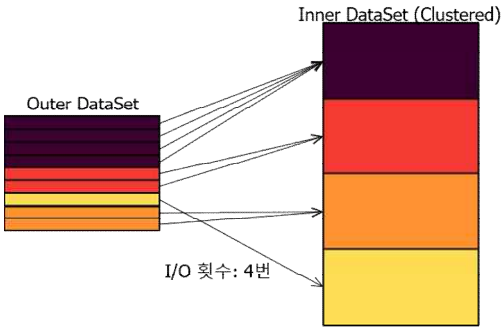
❍ Hadoop 환경에서 Map/Reduce 방식을 활용하는 중첩 루프(Nested-Loop) 기반 공간 조인의 성능을 개선함

- 공간 빅데이터 플랫폼에서는 공간 분석의 많은 부분을 차지하는 대용량의 데이터 사이의 공간 조인의 성능이 매우 중요한 부분을 차지하며 특히 대부분의 공간 조인의 기반 알고리즘인 중첩 루프 방식의 성능은 전체 공간 빅데이터 플랫폼의 성능에 매우 중요함
- 중첩 루프 기반 공간 조인 과정에서 outer 데이터셋의 각 HDFS 파일 블록을 전체 inner 데이터셋과 조인할 때 outer 데이터 블록의 레코드의 순서를 재구성하는 방법을 통해 공간 조인시 발생되는 HDFS I/O 횟수를 줄임

- <그림 2-24>는 기존의 중첩 루프 방식의 공간 조인 과정에서 하나의 outer 데이터 블록에 대해 inner 데이터셋 전체와 조인하는 과정이며 해당 outer 블록의 공간 조인과정에서는 총 9번의 I/O 횟수가 발생됨
- <그림 2-25>는 개선된 중첩 루프 방식의 공간 조인 방식이며 기존의 공간 조인 방법과 달리 총 4번의 HDFS 파일 블록 I/O가 발생됨



<그림 2-24> 기존 중첩 루프 공간 조인 방법



<그림 2-25> 개발된 중첩 루프 공간 조인 방법

- 개발된 방식에서는 outer 블록에 포함된 각 레코드에 대해 공간 조인을 수행하기 앞서서 각 레코드가 어떤 inner 블록과 매치되는가에 따라 동일 inner 블록과 조인된 레코드들을 서로 grouping하는 사전 작업을 수행함
- 그림에서 왼쪽 outer 블록에 포함된 레코드들이 조인에 앞서 조인될 inner 블록에 따라 group된 것을 볼 수 있음
- 이를 통해 한 그룹에 속한 레코드들은 공간 조인 과정에서 모두 동일 inner 블록만을 참조하기 때문에 한 번의 HDFS 블록 I/O를 통해 공간 조인이 가능함

❍ HDFS I/O 횟수를 줄임으로써 공간 조인의 시간 성능을 기존 방법에 비해 약 2.9배 개선함

- <그림 2-26>은 기존 공간 조인 방식과 연구 개발된 공간 조인 방식의 성능 실험 결과를 보여줌
- 그림에서 “재배열” 방식은 본 연구를 통해 개발된 공간 조인 방식을 의미함
- 공간 조인에 사용하는 내부 클러스터 캐쉬의 크기를 64mb와 128mb 두 가지 경우로 하여 성능을 측정함.
- (분석 기능 선정 기준) 성능 평가 환경은 14대의 데이터 노드로 구성된 하둡 플랫폼에서 실시되었음. 본 기술은 공간 조인의 성능을 개선하는 것이므로, 공간 분석에서 자주 사용하는 공간 데이터 중에서 비교적 크기가 큰 데이터인 ‘전국 건물 주소’ 데이터세트 (약 927mb)와 실험으로 위해 임의로 생성한 데이터세트를 활용하여 공간 조인 수행 속도를 측정하였음. 개발된 기술은 공간 빅데이터 시스템에서 사용하는 공간 조인용 내부 캐쉬에 따라 성능 차이가 발생할 수 있기 때문에, 64mb 크기와 128mb 크기의 내부 캐쉬를 사용하는 경우 각각에 대해 성능을 측정하였음.

Cluster Cache 크기	조인 방식	클러스터 수	I/O 횟수	소요시간
64mb	기존	12	495	17:08
	재배열	12	49	5:59 **(2.9x)**
128mb	기존	12	495	17:08
	재배열	12	12	3:53

<그림 2-26> 공간 조인 성능 평가

- 실험 결과는 내부 캐쉬의 크기에 따라 성능 차이가 많이 나는 것을 알 수 있음
- 내부 캐쉬의 크기가 64mb인 경우는 개발된 공간 조인 방식이 기존의 방법에 비해 약 2.9배의 공간 조인 성능 차이를 보임
  - 기존의 방식에서는 총 495회의 I/O가 발생되는 반면 개선된 방식에서는 49회의 I/O만 발생하는 것을 볼 수 있으며 성능의 차이는 I/O 횟수의 차이에 기인한다고 볼 수 있음
  - 캐슁 효과에 의해 기존 방법에서도 발생되는 I/O 횟수가 495회에서 14회로 급격히 줄어든 것을 볼 수 있음
- 내부 캐쉬의 크기가 128mb로 커지는 경우에는 공간 조인 성능의 차이는 그리 크지 않을 뿐만 아니라 기존 방식이 소요시간이 더 작은 것을 볼 수 있음
  - 발생되는 I/O 횟수가 개선된 방식에서 49회에서 12회로 줄어들지만 상대적으로 비율은 적음
  - 개발된 방식은 기존의 방식과 달리 공간 조인 과정에서 각 outer 블록마다 동일 위치의 레코드를 group하는 오버헤드를 가지기 때문인 것으로 판단됨
- (결과 의미) 개발된 방법은 내부 캐쉬의 크기가 작은 경우 기존의 방법에 비해 월등한 성능 개선 효과를 보이며, 내부 캐쉬가 커짐에 따라 개발된 방법의 개선 효과가 떨어지는 것을 알 수 있음. 그러나 개발된 방법은 기존의 방법과 달리 캐쉬 크기에 따른 성능 낙폭이 작아 다양한 시스템 설정 환경에서 안정적으로 운영될 수 있는 것을 알 수 있음

⑫ [단위성과 A-23] 공간 빅데이터 저장관리 인프라 시범서비스 적용 지원(1) (5차연도)

○ 공간 빅데이터 저장관리 인프라 시스템의 사업단 활용 가능성을 확인하기 위해 사업단에서 운용 중인 시스템과의 성능을 비교평가함

- 성능평가는 두 시스템을 동일한 하드웨어 환경에서 동일한 공간 분석 응용을 수행한 후 그 처리 소요시간을 측정하는 방법을 사용함
- (분석 선정 기준) 성능평가 활용 대상 공간 분석은 평가 결과의 왜곡을 방지하고자 사업단에서 실제로 운영하는 공간 분석 응용 중에서 사업단에서 직접 선정하여 제공한 '10분 정책지도 테스트'의 분석 모델을 활용함. 활용 분석 과정은 총 7단계의 분석 작업으로 구성되며(그림 2-27) 실험 구동 환경은

연구단과 사업단 시스템 동일하게 설정함

- 위 성능 평가는 동일하게 사업단에서 제공한 5대의 데이터 노드로 구성된 하둡 플랫폼에서 수행하였음

- 공간 빅데이터 시스템의 성능 평가에 활용된 데이터는 <표 2-6>과 같음

※ 10분 정책지도 : 2017년 공간 빅데이터 활용 수요분석과제에서 빅데이터 체계의 활용도를 높이기 위해 도출된 4개 사례 중 하나로, 주민이 어린이집, 노인복지시설, 공원 등 사회복지시설을 도보로 10분 이내에 접근하여 이용 가능한지를 분석하는 것임

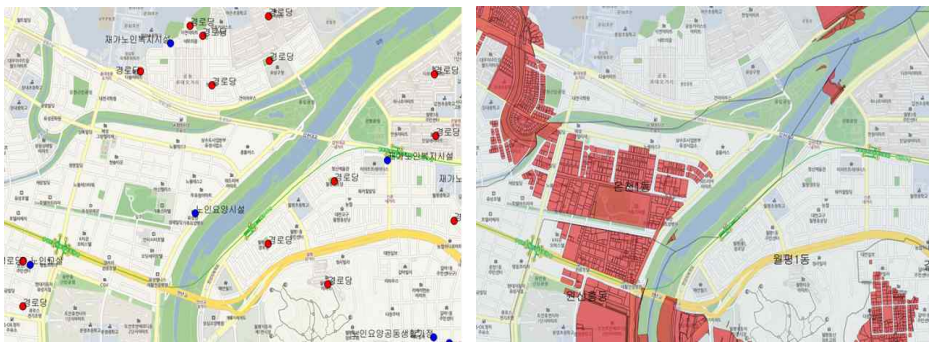
<표 2-6> 공간빅데이터 시스템 성능평가 활용 데이터

데이터셋 이름	레코드 건수	데이터 용량
연속지적도	37,648,906건	14.1Gb
전국노인복지시설	80,069건	34.8Mb
전국 인구밀도	102,219건	171Mb
행정동 코드	3,495건	37.5Mb



<그림 2-27> '10분 정책지도 테스트' 공간 분석 응용

❍ '10분 정책지도 테스트'를 활용한 노인 복지시설 추가 설립 대상지 추천 분석을 수행하고 분석 결과를 가시화함



<그림 2-28> 10분 정책지도 테스트 분석 결과

❍ '10분 정책지도 테스트'의 수행 시간을 비교함

- <그림 2-29>는 앞의 공간분석 응용에 대하여 7단계의 성능 분석 작업과 1단계의 추가 분석 작업의 소요시간을 측정한 결과를 보여줌

- 전체 '10분 정책지도 테스트' 분석 응용 수행시간은 사업단 시스템과 연구단 시스템(본 시스템)에 대하여 각각 38분 39초와 20분 01초가 소요되어 사업단 엔진 (개선 버전) 대비 84%의 성능 향상율을 보였음

- (분석 결과 의미) 본 실험은 사업단에서 자주 활용하는 분석 라이브러리를 활용한 분석 모델을 통해 평가한 결과로서, 연구단에서 개발한 공간 빅데이터 플랫폼을 사업단에 적용할 경우 의미 있는 성능 개선 효과를 볼 수 있을 것으로

예상함

- (분석 결과 의미) 실험 결과를 볼 때 공간 조인을 사용하는 분석 모듈들(교차분석, 교차반전)이 일반 공간 분석 모듈들 (영역분석, 버퍼추정, 영역분석)보다 연구단 플랫폼의 사업단 플랫폼 대비 성능 향상율이 더 월등한 것으로 나타나는 것을 알 수 있음. 이를 통해 연구단에서 개발한 공간 조인 기술의 성능 향상 효과를 확인할 수 있었음

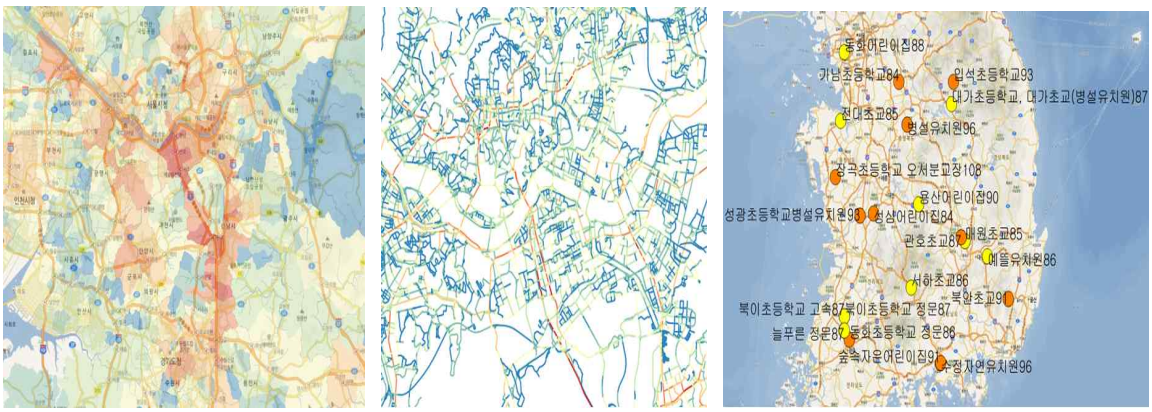
분석내용	사업단	연구단	성능향상율
1.영역분석	00:15	00:10	50%
2.버퍼추정	00:20	00:13	54%
4.중심점추출	00:21	00:30	-30%
5.영역분석	00:13	00:07	86%
6.교차분석	00:21	00:06	250%
3.교차반전	17:56	07:48	130%
7.클립분석	08:38	08:04	7%
8.추가분석	10:35	04:03	161%
합계	38:39	20:01	84%

<그림 2-29> 연구단-사업단 플랫폼 성능 평가 결과

○ 대용량 데이터인 도로교통공단의 DTG 데이터를 활용하여 차량 통행량을 분석함

- 레코드가 약 82.7억건이고 크기가 821GB에 이르는 대용량의 DTG 데이터(1개월 분량)를 활용하여 읍면동, 도로링크별 차량통행량과 어린이 보호구역 주변 차량 운행속도를 분석함

- 각 DTG 레코드에 공간 정보를 부가(읍면동/도로/격자 식별자 부여)하여 분석에 사용하였으며 다양한 교통 관련 분석에 활용 가능함



<그림 2-30> 대용량 DTG 데이터를 활용한 차량통행량 분석

⑬ [단위성과 A-24] 클라이언트-서버 기반 원격 공간 빅데이터 접근 모듈 개발 (5차연도)

❍ Hadoop 기반 플랫폼에서 응용을 구동하기에 필수적인 복잡한 라이브러리 활용 없이도 원격에서 최소한의 라이브러리를 통해 공간 분석 서비스를 활용할 수 있는 gRPC를 이용한 원격 인터페이스 접근 모듈을 개발함

※ gRPC(Remote Procedure Calls): 구글이 최초로 개발한 오픈소스 원격 프로시저 호출 시스템이며, 전송을 위해 HTTP/2를, 인터페이스 정의 언어로 프로토콜 버퍼를 사용함. 많은 언어를 대상으로 크로스 플랫폼 클라이언트 및 서버 바인딩을 생성할 수 있음

- Hadoop API 의존성을 배제할 수 있어 많은 공간 분석 개발자들이 Hadoop 환경에 익숙하지 않는 것을 고려할 때 응용 개발 편의성을 높일 수 있음
- 특정 언어나 특정 OS에 국한되지 않고, 다양한 언어와 운영체제에서 공간 분석 응용을 개발할 수 있는 환경을 제공함
- Java, Python, C# 언어를 지원하며, Windows, Linux 등 복수의 OS를 지원함

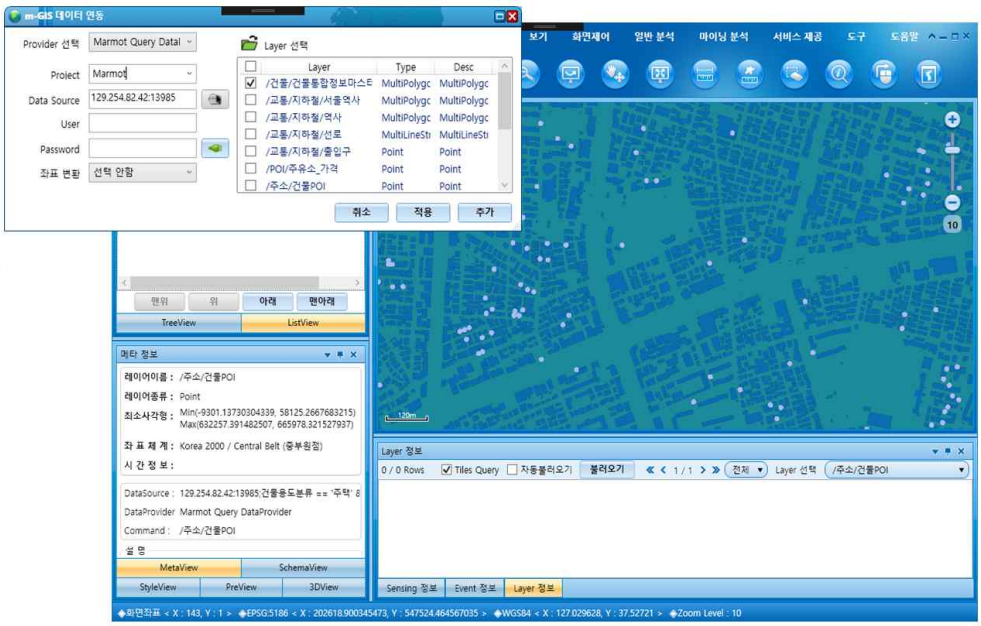
❍ 공간 빅데이터 플랫폼을 클라이언트/서버 구조로 확장하여 서버 모듈을 Hadoop 클러스터 상에서 구동시키고, 클라이언트 모듈을 네트워크를 통해 원격에서 접속할 수 있도록 함

❍ 공간 빅데이터 플랫폼에서 제공하는 기능을 원격에서 네트워크를 통해 활용하기 위해 Google에서 개발한 통신 프레임워크인 gRPC와 데이터 직렬화 도구인 protocol-buffer를 이용한 원격 접근 인터페이스를 개발함

※ 프로토콜 버퍼(Protocol Buffers): 구조화된 데이터를 직렬화하는 방식으로, 유선이나 데이터 저장을 목적으로 서로 통신할 프로그램을 개발할 때 유용함

❍ Java, Python, 그리고 C# 언어용 원격 접근 모듈을 개발하여 응용 개발자가 복수의 프로그래밍 언어에서 공간 빅데이터 시스템의 기능을 활용할 수 있도록 함

- C# 언어의 경우 개발된 원격 접근 모듈의 동작 확인을 위해 공간 빅데이터를 접근하여 가시화하는 데이터 제공자 (data provider)를 개발하였으며 이를 통해 C# 언어용 원격 접근 모듈의 동작을 확인함(그림 2-31)
- Java 언어용 원격 접근 모듈의 경우에는 GeoServer의 가시화 플러그인을 개발하여, 원격 접근 모듈의 동작을 확인함(그림 2-32)



<그림 2-31> C#용 클라이언트 모듈 활용 응용



<그림 2-32> GeoServer 플러그인 활용

- GeoServer 플러그인의 경우에는 대용량 공간 데이터의 가시화에 필연적으로 발생할 수 있는 메모리 부족 현상을 회피하기 위해 샘플링 기법을 개발함
- 데이터 샘플링은 가시화 범위에 포함될 공간 데이터 개수의 추정을 통해, 샘플된 데이터가 일정 부분에 쏠리지 않고 대상 구역에 적절히 분포될 수 있도록 개발함
- 인덱스 파티션 기반의 데이터 접근 방식과 클라이언트 캐싱 기법을 활용한 대용량 데이터의 가시화시에도 고속으로 데이터 가시화가 가능하도록 하는 기술을 개발함

⑭ [단위성과 A-26] GUI 기반 공간 빅데이터 응용 개발 도구 개발(1) (5차연도)

❍ 공간 빅데이터 저장관리 시스템의 기능을 활용하는 분석 응용을 개발할 때 시스템 API를 활용한 프로그래밍에 익숙하지 않은 응용개발자를 돕기 위하여 개발작업을 도와주는 GUI 기반의 도구를 개발함

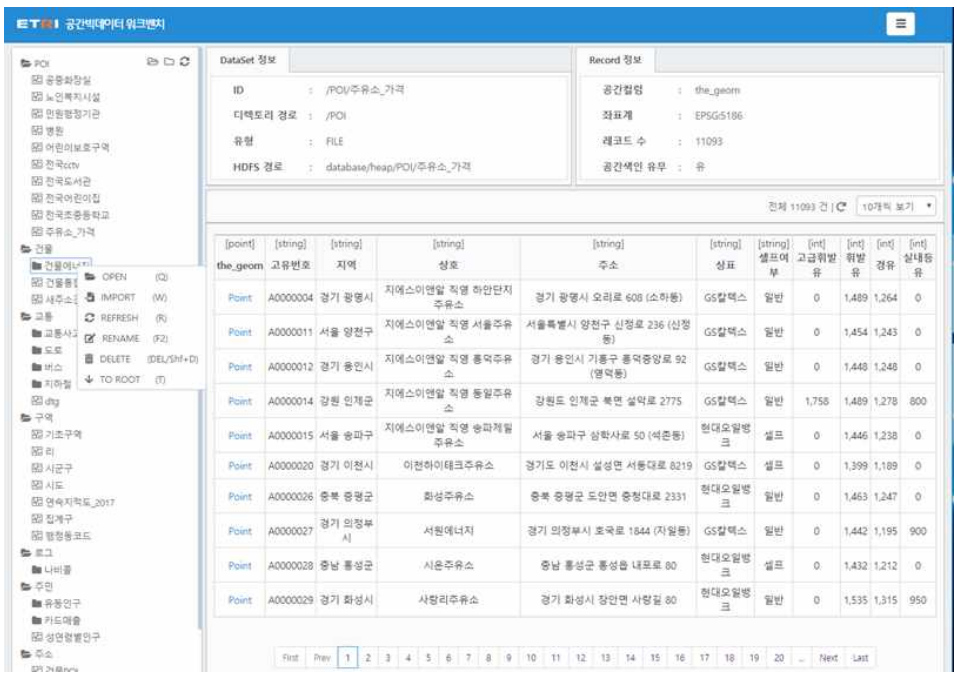
- 응용개발자의 응용 개발 편의를 도모하고 공간 빅데이터 저장관리 시스템의 기능을 사용하는 공간 분석의 생산성을 향상시킬 수 있음
- 공간 빅데이터 분석 응용 개발 도구는 웹 서비스 형태로 개발되었으며, 사용자는 웹브라우저를 이용하여 서비스를 이용할 수 있음

❍ 데이터셋 탐색 기능

- 공간 빅데이터 저장시스템에서 제공하는 API를 활용하여 공간 빅데이터에 저장된 데이터셋을 탐색하고, 그 내용을 browse할 수 있는 기능을 개발함
- 공간 빅데이터 저장시스템에 저장된 데이터셋들의 이름, 위치 등을 변경할 수 있는 기능 및 관련 GUI를 개발함

❍ 공간 빅데이터 Import/Export 기능

- 공간 빅데이터 저장시스템에 저장된 데이터셋을 다양한 형식(shapefile, csv, geojson 형식)의 파일로 export할 수 있는 기능 및 관련 GUI를 개발함
- 다양한 형식(shapefile, csv, geojson, line 기반 텍스트 형식)의 외부 파일을 공간 빅데이터 저장소에 import할 수 있는 기능 및 관련 GUI를 개발함



<그림 2-33> GUI 기반 공간 빅데이터 응용 개발 도구

⑮ [단위성과 A-15] 공간 빅데이터 처리 언어용 내부 코드 실행기 모듈 (6차연도)

❍ 스크립트 기반 공간 빅데이터 처리 언어 구동 엔진 개발

- Groovy 언어 기반 DSL 개발
- 스크립트 언어 장점과 기존 Java, Groovy 언어를 활용한 확장 가능
- 공간 분석에 활용되는 주요 기능을 스크립트 언어로 활용 가능
- 지금까지 개발된 주요 공간분석 응용을 개발된 스크립트 언어로 재 구현함

```
// '노인복지시설_경로당_추출_버퍼' 추출
createDataSet "tmp/노인복지시설_버퍼", {
  geometry 'the_geom(EPSG:5186)'
  force = true
  from plan('대도시지역 추출') {
    load "POI/노인복지시설"
    filter "induty_nm == '경로당'"
    project 'the_geom'
    buffer "the_geom", 400
  }
}
clusterDataSet "tmp/노인복지시설_버퍼“

// '인구밀도_2017_중심점추출_10000이상' 추출
createDataSet "tmp/인구밀도_중심점", {
  geometry GC_INFO
  force = true
  from plan('인구밀도_2017_중심점추출_10000이상') {
    load "주민/인구밀도_2000"
    centroid 'the_geom'
    filter "value >= 10000"
    project 'the_geom'
  }
}
clusterDataSet "tmp/인구밀도_중심점“

deleteDataSet "tmp/노인복지시설_버퍼“
exportDataSetToShapefile("tmp/인구밀도_중심점“,\
                         “tmp/result", { })
```

<그림 2-34> 스크립트 공간 분석(10분 정책지도 중 일부)

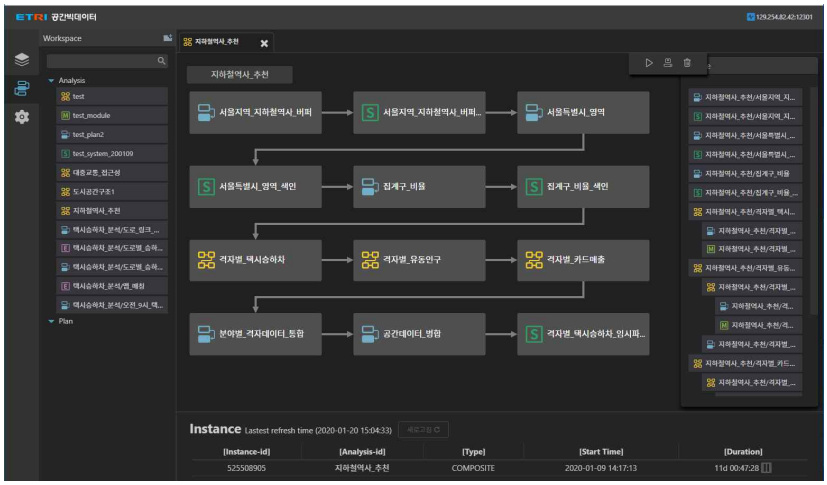
❍ 스크립트 기반 공간 빅데이터 처리 언어 시험 계획서 및 결과서 작성

<표 2-7> 스크립트 기반 공간 빅데이터 처리 언어 시험 항목 리스트

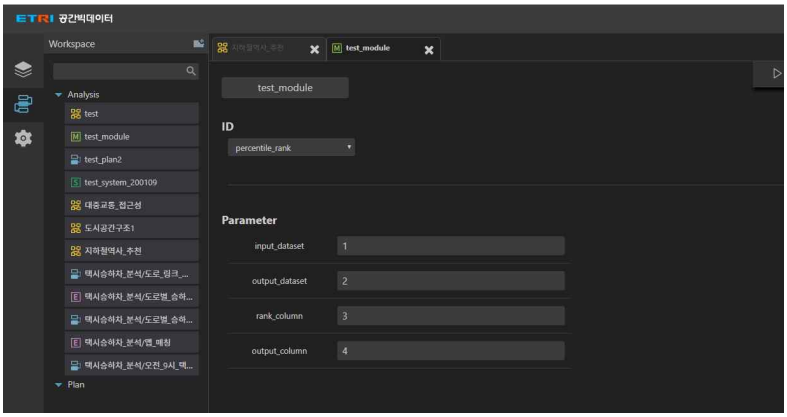
시험항목 ID	시험항목 이름	시험 내용
TC-01	데이터세트 생성	주어진 스키마를 갖는 데이터세트를 생성 동작을 확인한다.
TC-02	데이터세트 삭제	주어진 이름의 데이터세트를 Marmot에서 삭제한다.
TC-03	공간 클러스터 생성	Marmot script를 통해 데이터세트에 공간 색인을 생성하는 것을 확인한다.
TC-04	데이터세트 이동	주어진 데이터세트의 경로를 변경시킨다.
TC-05	플랜 실행	플랜 수행의 정확도를 시험한다.
TC-06	세이프파일 Import	주어진 세이프파일을 import하여 데이터세트를 생성한다.
TC-07	세이프파일 Export	데이터세트를 세이프파일 형태로 export 시킨다.
TC-08	CSV 파일 Import	주어진 CSV 파일을 import하여 데이터세트를 생성한다
TC-09	CSV 파일 Export	데이터세트를 CSV 파일 형태로 export 시킨다

⑯ [단위성과 A-21] MapReduce 기반 워크플로우 가시화도구 개발 (6차연도)

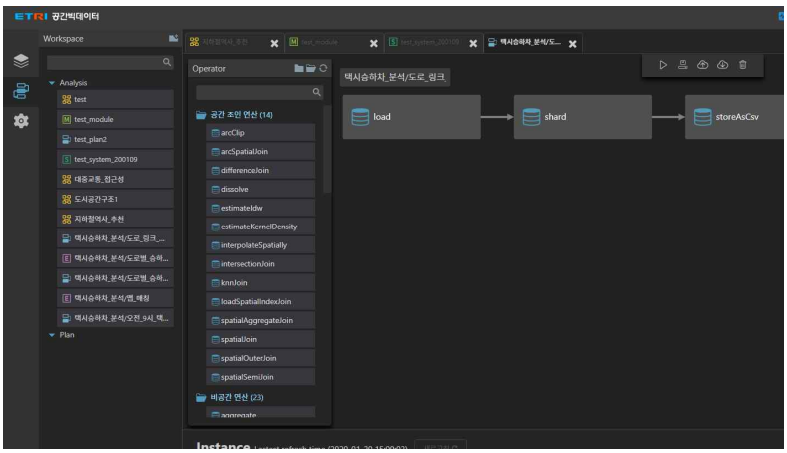
- ❍ 수행 중인 분석 응용의 강제 중지 기능을 구현하였음
- ❍ 작성된 공간분석 플랜의 관리 기능을 구현하였음
- ❍ 5가지 형식의 공간 분석 (Analysis) 타입 및 관련 모듈을 개발하였음
  - Plan Analysis: 공간 분석 플랜으로 구성된 분석 모듈
  - Module Analysis: 공간 분석 플랫폼 내부 프로세스로 구현된 분석 모듈
  - System Analysis: 공간 분석 플랫폼 제공 내부 도구로 구현된 분석 모듈
  - Extern Analysis: 외부 레거시 시스템으로 구현된 분석 모듈
  - Composite Analysis: 긴존 공간 분석 모듈의 플로우로 구현된 분석 모듈
- ❍ 분석 플랜 워크플로우 가시화를 위한 웹기반 GUI 도구 개발



<그림 2-35> 워크플로우 가시화 도구 (Composite Analysis 보기)



<그림 2-36> 워크플로우 가시화 도구 (Module Analysis 정의 화면)



<그림 2-37> 워크플로우 가시화 도구 (Plan Analysis 정의 화면)

❍ MapReduce 기반 워크플로우 가시화 도구 시험 계획서 및 결과서 작성

<표 2-8> MapReduce 기반 워크플로우 가시화도구 시험 항목 리스트

시험항목 ID	시험항목 이름	시험 내용
TC-01	Plan Analysis 생성 및 등록	Plan Analysis 생성 및 등록 기능을 확인한다.
TC-02	Plan Analysis 실행	Plan Analysis 실행 기능을 확인한다.
TC-03	System Analysis 생성 및 등록	System Analysis 생성 및 등록 기능을 확인한다.
TC-04	System Analysis 실행	System Analysis 실행 기능을 확인한다.
TC-05	Module Analysis 생성 및 등록	Module Analysis 생성 및 등록 기능을 확인한다.
TC-06	Module Analysis 실행	Module Analysis 실행 기능을 확인한다.
TC-07	Extern Analysis 생성 및 등록	Extern Analysis 생성 및 등록 기능을 확인한다.
TC-08	Extern Analysis 실행	Extern Analysis 실행 기능을 확인한다.
TC-09	Composite Analysis 생성 및 등록	Composite Analysis 생성 및 등록 기능을 확인한다.
TC-10	Composite Analysis 실행	Composite Analysis 실행 기능을 확인한다.
TC-11	Analysis 목록 및 등록정보 확인	등록 analysis 목록 확인 기능을 확인한다.
TC-12	등록된 analysis 제거	Analysis 목록 제거 기능을 확인한다.

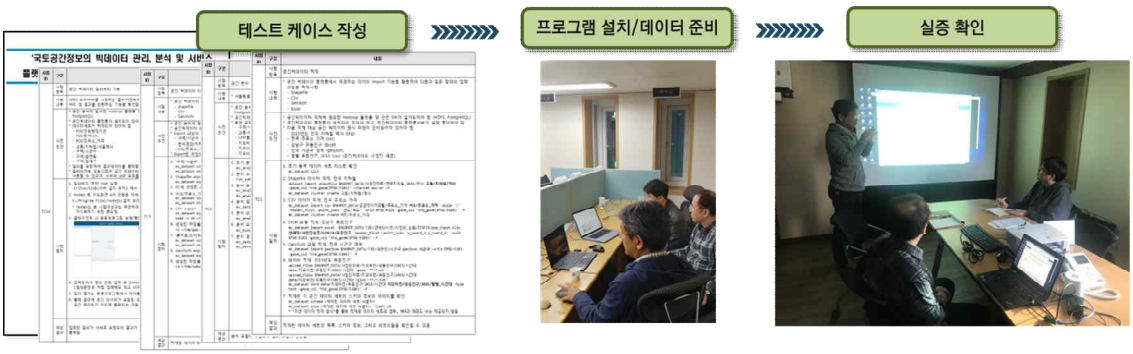
⑰ [단위성과 A-25] 공간 빅데이터 저장관리 인프라 시범서비스 적용 지원(2) (6차연도)

❍ 사업단 시스템에 공간 빅데이터 저장관리 시스템의 성과물을 설치하고 시범 운용을 통해 실증

- 사업단에 방문하여 성과물을 설치하고 성과물 기능을 테스트 실증
  - 장소 : 공간빅데이터사업단 사무실(서울 방배동)
- 2019.11.05.~06 에 실증 대상 기능을 11개의 TC(테스트케이스)로 정의하고 실증 확인
  - 1세부 : 공간분석 플로우처리 기능, 공간데이터 import/export 기능
  - 2세부 연동 : 고급 공간분석 컴포넌트(2-2세세부), OLAP 분석 시스템(2-3세세부)
- 2019.11.19. 에 실증 대상 기능에 대한 4개의 TC를 추가적으로 실증
  - 공간 빅데이터 가시화 지원 기능, 공간 빅데이터 스크립트 언어, 공간 빅데이터 질의처리 기능
  - 공간 ETL 관리도구

❍ 공간 빅데이터 저장관리 시스템 성과물 실증 과정은 다음과 같은 단계로 수행함

- 테스트케이스 수립
- 사업단 시스템에 프로그램 설치 및 데이터 준비
- 프로그램 실행 및 기능 확인



<그림 2-38> 사업단 실증 과정

❍ 실증 시험에 필요한 공간 빅데이터 저장관리 시스템의 시험 항목(테스트케이스)은 <표2-9>와 같으며 세부 내용을 시험 절차서로 작성함

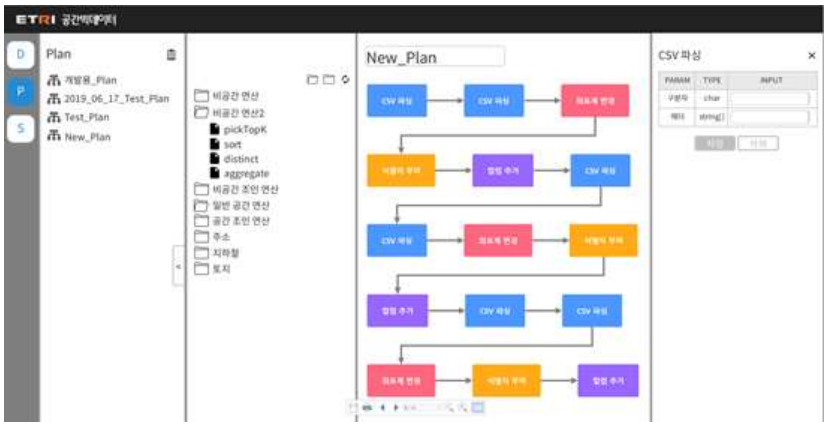
<표 2-9> 공간 빅데이터 저장관리 시스템 실증용 시험 항목 리스트

시험항목 ID	시험항목 이름	시험 내용
TC-01	공간 빅데이터 Import 기능	공간 빅데이터 플랫폼에서 제공하는 데이터 import 기능을 활용하여 다음과 같은 형태의 입력 파일을 적재시킴 - Shapefile - CSV - GeoJson - Excel
TC-02	공간 분석 플로우 처리	'서울특별시 지하철 후보지 추천' 분석 모델을 수행시켜 공간 분석 플로우 처리 기능을 확인함
TC-03	공간 빅데이터 Export 기능	공간 빅데이터 저장 플랫폼에 저장된 데이터 세트를 다음과 같은 형태의 파일로 export 시킴 - shapefile - CSV - GeoJSoN
TC-07	공간 빅데이터 저장 시스템간 연동	- 1-1 세세부 플랫폼과 1-2 세세부 플랫폼 기능 연동 테스트 - 1-1 세세부 플랫폼을 이용하여 원시 자료 정제 및 가공 수행 - 1-2 세세부 플랫폼을 이용하여 1-1세세부에서 가공한 자료를 이용하여 Cluster And Outlier 분석 수행후 분석결과 저장 - 1-1 세세부 플랫폼에서 1-2세세부 분석결과 자료 로딩 및 확인
TC-08	'E2SFCA' 분석 컴포넌트	'강남구 대중교통 접근성 분석 모델'을 활용하여 2-2세세부를 통해 개발된 'E2SFCA' 분석 컴포넌트의 기능과 1-1세세부 성과물 연동성을 실증함
TC-09	'MapMatching'분석 컴포넌트, 'PercentileRank' 컴포넌트	'도로별 택시 운행 분석 모델'을 활용하여 2-2세세부를 통해 개발된 'MapMatching' 분석 컴포넌트와 'PercentileRank' 분석 컴포넌트의 기능을 실증하고, 1-1세세부 성과물 연동성을 실증함
TC-10	'Decision Tree' 컴포넌트	'시군구별 폐기물 요인 분석 모델'을 활용하여 2-2세세부를 통해 개발된 'Decision Tree' 컴포넌트의 기능과 1-1세세부 성과물 연동성을 실증함
TC-11	공간 빅데이터 OLAP	2-3세세부 개발 성과물과 1-1세세부 성과물 연동성을 실증함
TC-12	공간빅데이터 스크립트 언어	'서울특별시 지하철 후보지 추천' 분석 모델의 일부를 공간빅데이터 스크립트 언어로 작성하여 수행시켜 기능을 확인함
TC-13	공간빅데이터 가시화 지원 기능	대용량 공간 빅데이터 (2018년도 건물통합정보마스터)를 GeoServer을 통해 가시화하여 공간 빅데이타 가시화 기능을 확인함
TC-14	공간 빅데이터 질의처리 기능	공간 빅데이터 플랫폼에 저장된 데이터를 SQL기반 질의언어를 통하여 질의하여 결과 데이터를 가져오는 기능을 확인

⑱ [단위성과 A-27] GUI 기반 공간 빅데이터 응용 개발 도구 개발(2) (6차연도)

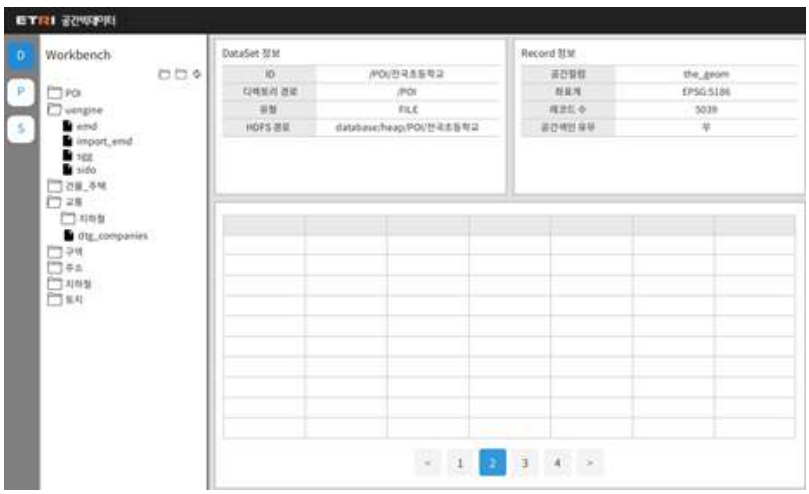
❍ 공간 빅데이터 저장관리 시스템의 시스템 API를 활용한 프로그래밍에 익숙하지 않은 응용개발자의 개발작업을 도와주는 GUI 기반 응용 개발 도구의 기능을 확장함

- 공간 빅데이터 데이터셋 및 속성 정보를 조회하는 탐색기의 기능을 개선함
- 공간 빅데이터 단위 연산의 연속적 수행을 분석 플랜으로 저작하는 도구를 개발함



<그림 2-39> 공간분석 워크플로우 저작 도구

❍ 기 개발된 공간 데이터세트 탐색기 모듈과 연동 작업 진행 중



<그림 2-40> 공간빅데이터 탐색기

❍ GUI기반 공간 빅데이터 응용 개발 도구 시험 계획서 및 결과서 작성

<표 2-10> 공간 빅데이터 처리 언어용 내부 코드 실행기 시험 항목 리스트

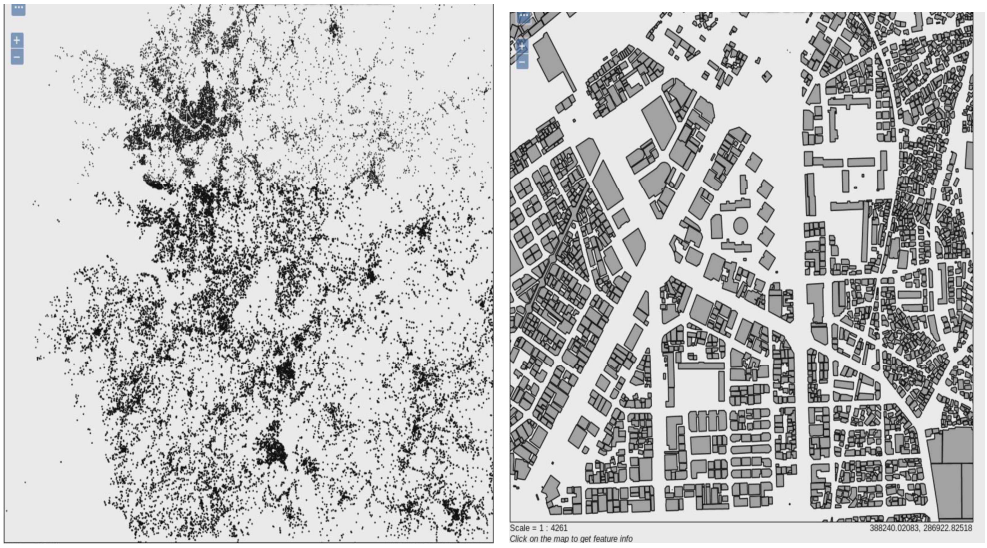
시험항목 ID	시험항목 이름	시험 내용
TC-01	데이터세트 열람	공간 분석 개발 도구의 데이터세트 열람 기능을 확인한다.
TC-02	데이터세트 이름 변경	공간 분석 개발 도구의 데이터세트 이름 변경 기능을 확인한다.
TC-03	데이터세트 이동	공간 분석 개발 도구의 데이터세트 이동 기능을 확인한다.
TC-04	데이터세트 삭제	공간 분석 개발 도구의 데이터세트 삭제 기능을 확인한다.
TC-05	Shapefile Import 기능	공간 분석 개발 도구의 Shapefile Import 기능을 확인한다.
TC-06	CSV 파일 Import 기능	공간 분석 개발 도구의 CSV 파일 Import 기능을 확인한다.
TC-07	Shapefile Export 기능	공간 분석 개발 도구의 Shapefile Export 기능을 확인한다.
TC-08	CSV 파일 Export 기능	공간 분석 개발 도구의 CSV 파일 Export 기능을 확인한다.
TC-09	분석 플랜 저작 기능	공간 분석 개발 도구의 분석 플랜 저작 기능을 확인한다.
TC-10	분석 플랜 download 기능	공간 분석 개발 도구의 분석 플랜 download 기능을 확인한다.
TC-11	분석 플랜 실행 기능	공간 분석 개발 도구의 분석 플랜 실행 기능을 확인한다.
TC-12	Project 연산	Project 연산 모듈 기능을 확인한다.

⑲ [단위성과 A-28] 대용량 공간 빅데이터 원격 접근 기능 개발 (6차연도)

❍ Hadoop 기반 플랫폼에서 복잡한 라이브러리의 활용 없이도 원격에서 최소한의 라이브러리를 통해 공간 빅데이터를 사용하여 분석 서비스를 활용할 수 있는 원격 인터페이스 접근 기능을 개선함

❍ 썸네일(thumbnail) 기반 고속 샘플링 기법을 통하여 가시화 대상 공간데이터를 선별함으로써 가시화 작업의 부담을 낮춤

- 서버로부터 대용량 데이터를 가져와서 클라이언트에서 가시화할 때 작업의 부담을 낮춤
- 클라이언트에서 대용량 데이터를 한번에 가져올 때 발생하는 메모리 부족 현상을 방지할 수 있음
- 클라이언트에서 화면을 축소/확대/이동하여 서버로부터 공간데이터를 다시 가져와야 할 때 속도 저하를 방지할 수 있음



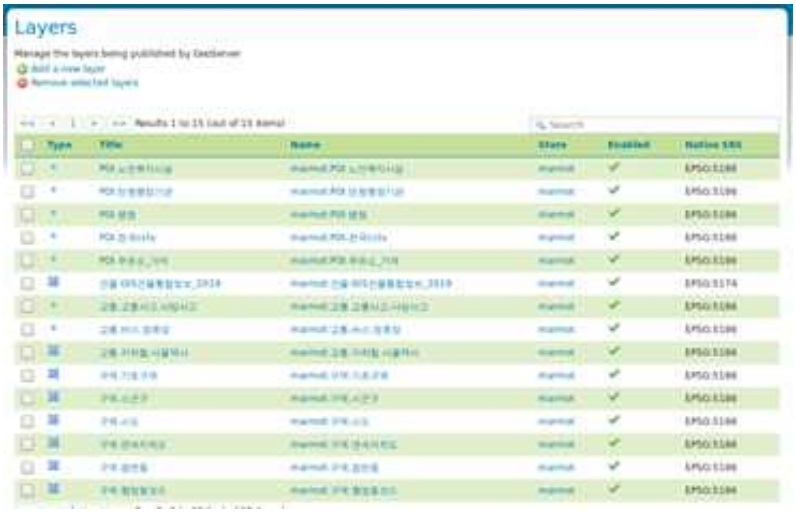
<그림 2-41> 대용량 공간 빅데이터 원격 접근 가시화

❍ 대용량 공간 빅데이터 원격 접근 기능 시험 계획서 및 결과서 작성

<표 2-11> 대용량 공간 빅데이터 원격 접근 기능 시험 항목 리스트

시험항목 ID	시험항목 이름	시험 내용
TC-01	Thumbnail 파일 생성	주어진 데이터세트의 thumbnail 파일이 생성되는 것을 확인한다.
TC-02	Thumbnail 파일 삭제	기존 thumbnail 파일이 삭제되는 것을 확인한다.
TC-03	데이터세트 삭제시 thumbnail 파일 삭제	Thumbnail이 존재하는 데이터세트를 삭제할 때 thumbnail 파일이 같이 삭제되는 것을 확인한다.
TC-04	Thumbnail 파일을 활용한 원격 데이터세트 접근	원격 데이터세트 접근시 thumbnail이 존재하는 경우 이를 활용하는 것을 확인한다.
TC-05	Thumbnail 파일을 활용하지 않는 원격 데이터세트 접근	원격 데이터세트 접근시 thumbnail이 존재하지 않는 경우 이를 활용하지 않은 상태에서도 동작하는 것을 확인한다.
TC-06	공간 색인을 활용한 원격 데이터세트 접근	원격 데이터세트 접근시 thumbnail을 사용할 수 없는 경우, 공간 색인을 사용하여 데이터를 접근하는 기능을 확인한다.
TC-07	로컬 캐쉬를 활용한 원격 데이터세트 접근	데이터세트의 극히 일부의 데이터를 접근하는 경우, 로컬 캐쉬를 사용하여 접근하는 것을 확인한다.

❍ 대용량 공간 빅데이터 원격 접근 기능 확인을 위한 GeoServer 연동 플러그인 모듈 개발

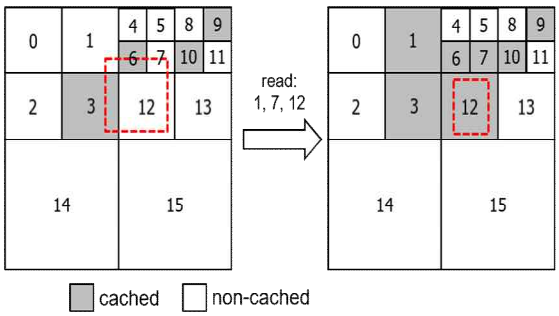


<그림 2-42> GeoServer 연동을 통한 레이어 계시 기능 화면

⑳ [단위성과 A-29] 대용량 공간 빅데이터 클라이언트 캐싱 모듈 개발 (6차연도)

❍ 클라이언트에서 가시화 대상 공간데이터를 가져올 때 인덱스 기반 클라이언트 캐싱(caching) 기능을 제공함

- 공간 빅데이터 저장관리 서버에서 데이터를 공간적으로 분할하여 관리하는 단위인 파티션(partition) 단위의 클라이언트 캐싱 기능을 제공함
- 가시화 Quad-Tree 기반 인덱스 파티션 단위로 데이터 전달
- 공간 색인 기반 데이터 파티션 정보를 활용하여 신속한 질의 결과 레코드 수 추정을 활용함



	캐쉬 여부	전체 레코드수	Overlap 영역 범위 비율	대상 레코드수
파티션 1	X	4500	0.05	225
파티션 3	O	2300	0.2	460
파티션 6	O	4000	0.1	400
파티션 7	X	1500	0.07	105
파티션 12	X	1000	0.9	900
			추정치	2090

샘플 개수: 500 (설정값)
샘플 비율: 500 / 2090
= 0.24

⇒ 1, 3, 6, 7, 12 파티션에서 질의 영역과 각각 겹치는 레코드들 중 0.24 확률로 레코드 선택

<그림 2-43> 공간색인 파티션 기반 클라이언트 캐싱

❍ 클라이언트 캐싱 성능 향상을 위한 파티션 prefetching 기능 구현

- 서버기반 질의 처리 및 샘플링 작업을 수행하는 과정을 대기하는 중에 다중 쓰레드 기법을 활용하여 다음 질의에 사용할 가능성이 있는 영역의 데이터를 미리 캐싱함
- 사용자 설정을 통해 prefetching 기능을 사용할지의 여부를 결정할 수 있는 기능 제공함

❍ 대용량 공간 빅데이터 클라이언트 캐싱 모듈 및 결과서 작성

<표 2-12> 대용량 공간 빅데이터 클라이언트 캐싱 모듈 시험 항목 리스트

시험항목 ID	시험항목 이름	시험 내용
TC-01	지역 데이터세트 캐쉬 저장소 생성	지역 캐쉬가 활성화됨에 따라 지역 캐쉬 저장소가 생성되는 것을 확인한다.
TC-02	데이터세트 접근에 따른 캐싱 기능 확인	데이터세트의 파티션 접근에 따라 로컬 캐쉬에 없는 경우, 로컬 캐쉬에 추가되는 기능 확인한다.
TC-03	시간 경과에 따른 파티션 무효화 기능 확인	지역 캐쉬에 저장된 파티션이 일정기간 동안 접근되지 않는 경우 무효화되는 기능 확인한다.

㉑ [단위성과 A-30] 2세부 공간 연산자 연동 모듈 개발(6차연도)

❍ 공간 빅데이터 저장관리 시스템 고도화의 일환으로 주요 공간 연산자를 추가적으로 구현함

- 추가할 공간 연산자의 선정은 연구단과 타 세부에서 제시한 것을 우선적으로 함
- 공간 연산자의 용어와 의미는 ArcGIS에서 사용되는 공간 연산자를 기준으로 하여 정의함
- 연구단에서 제공한 입력/출력 데이터를 활용하여 동작을 확인

공간연산	동작 내용	비고
Clip	Input 파일을 Clip 파일 모양으로 잘라내기	연산
Split	Input 파일을 Split 파일 모양으로 분할하여 잘라내기	연산, 프로세스
Intersect	두 개의 Input 파일 간 겹치는 부분(교집합)의 파일 생성하기	연산
Spatial Join	Target 파일 각 feature의 공간적 위치에 해당하는 Join Feature의 속성 붙이기	연산
Union	두 개의 Input 파일 합치기	프로세스
Buffer	Input 파일을 기준으로 일정거리만큼 면적이 증가된 폴리곤 데이터 생성	프로세스
Feature to Point	점, 선, 면 형태의 Input 파일을 기반으로 Point 형태의 Output 파일 만들기	연산
Multipart To Singlepart	여러 개 Feature를 마치 한 개 인 것처럼 묶인 상태의 Input 파일을 여러 개의 Feature로 분할	연산
Merge	인접한 여러 Input 파일을 하나의 파일로 합치기	프로세스
Select by location	두 개 이상의 데이터 간 위치정보에 기반하여 특정 Feature 선택하기	연산
Define Projection	Input 파일의 공간 정보에 대한 좌표 체계를 정의한다.	API
Project	Input 파일의 공간 정보의 좌표 체계를 변환시킨다.	연산

<그림 2-44> 추가된 공간 연산자

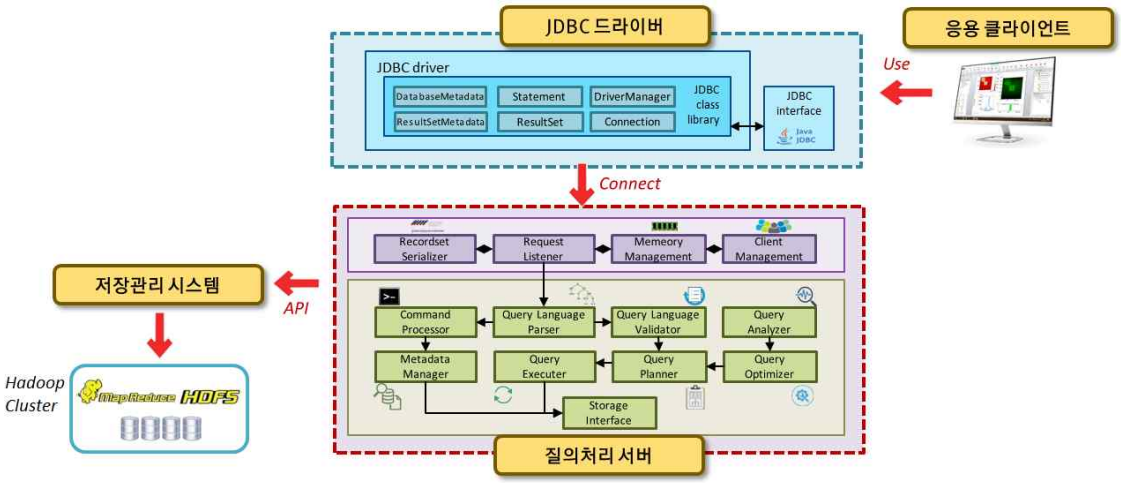
❍ 2세부 공간 연산자 연동 모듈 시험 계획서 및 결과서 작성

<표 2-13> GUI기반 공간빅데이터 응용 개발 도구 시험 항목 리스트

시험항목 ID	시험항목 이름	시험 내용
TC-01	Clip 연산	Clip 연산 모듈 기능을 확인한다.
TC-02	Split 연산	Split 연산 모듈 기능을 확인한다.
TC-03	Intersect 연산	Intersect 연산 모듈 기능을 확인한다.
TC-04	Union 연산	Union 연산 모듈 기능을 확인한다.
TC-05	Buffer 연산	Buffer 연산 모듈 기능을 확인한다.
TC-06	Feature to Point 연산	Feature to Point 연산 모듈 기능을 확인한다.
TC-07	Multipart to Singlepart 연산	Multipart to Singlepart 연산 모듈 기능을 확인한다.
TC-08	Merge 연산	Merge 연산 모듈 기능을 확인한다.
TC-09	SpatialJoin 연산	SpatialJoin 연산 모듈 기능을 확인한다.
TC-10	Select by location 연산	Select by Location 연산 모듈 기능을 확인한다.
TC-11	Define Projection 연산	Define Projection 연산 모듈 기능을 확인한다.
TC-12	Project 연산	Project 연산 모듈 기능을 확인한다

(나) 공간 빅데이터 질의처리 엔진(핵심성과 B)

○ 공간 빅데이터 저장 프레임워크에 저장된 공간 빅데이터에 대한 질의(query) 처리를 지원하는 질의처리 엔진 및 이를 응용프로그램에서 접근하여 사용할 수 있는 응용프로그램 인터페이스



<그림 2-45> 공간 빅데이터 질의처리 엔진(핵심성과 B)

○ 특징

- SQL 기반 질의언어를 사용하여 공간 빅데이터를 접근, 조회, 조작하고 개발자가 응용프로그램을 편리하게 개발할 수 있게 해주는 표준 인터페이스를 제공함
- Hadoop 환경에 설치된 공간빅데이터 저장관리 프레임워크 위에서 **SQL기반 질의언어 및 표준 인터페이스를 사용**할 수 있게 해줌
- SQL을 지원하는 GIS나 공간 DBMS를 사용하던 기존 legacy 시스템이 **공간 빅데이터 저장관리시스템을 사용하도록 전환하거나 통합하기 용이**하게 해줌

○ 시스템 구조

- 외부 클라이언트의 요청에 대하여 데이터를 전달하는 서버 형태로 구동됨
- 시스템 구조상 공간 빅데이터 저장관리 시스템과 응용 클라이언트의 사이에 위치하며 저장관리 시스템과는 API를 통하여 연동됨
- SQL 기반 질의언어를 파싱하고 분석하여 공간 빅데이터 저장관리 시스템에서 제공하는 연산 및 플랜으로 변환하여 실행함
- Java 기반 클라이언트에서 표준적인 방법으로 데이터를 접근하고 조회할 수 있는 JDBC 드라이버를 제공하여 이를 통하여 질의처리엔진과 클라이언트가 연결되어 데이터를 주고받음

※ JDBC(Java DataBase Connectivity) : Java언어로 된 프로그램에서 DBMS 또는 데이터 저장소를 접근하고 데이터를 입출력하는 표준 인터페이스

① [단위성과 B-1] 공간 빅데이터 질의처리 엔진 기술분석, 요구사항 정의 (1차연도)

❍ 공간 빅데이터 질의처리 엔진의 사용자 요구사항 및 개발자 관점에서의 시스템 요구사항, 주요 기능을 정의함

- 공간 빅데이터 저장 프레임워크의 공간 데이터에 대한 SQL 기반의 공간 질의를 처리하고 이와 연동하여 Hadoop 플랫폼 기반의 응용 프로그램에서 접근하여 사용할 수 있는 응용프로그램 인터페이스를 제공하는 것이 목표임

❍ 공간 빅데이터 질의처리 엔진의 사용자 관점의 요구사항을 분석하여 시스템 요구사항을 정의함

<표 2-14> 공간 빅데이터 질의처리 엔진 사용자 요구사항

요구사항 ID	내용
USR-201	SBQE는 테이블을 생성하는 SQL 기반의 문장을 입력받을 수 있어야 한다.
USR-202	SBQE는 속성질의가 포함된 SQL 기반의 질의 문장을 입력받을 수 있어야 한다.
USR-203	SBQE는 OGC Simple Feature 모델의 공간데이터 연산함수가 포함된 SQL 기반의 질의 문장을 입력받을 수 있어야 한다.
USR-204	SBQE는 속성과 공간 질의에 대하여 질의 실행 순서를 계획할 수 있어야 한다.
USR-205	SBQE는 속성과 공간 질의에 대한 연산 기능을 호출할 수 있어야 한다.
USR-206	SBQE는 C/C++ 개발 환경에서 ODBC 인터페이스를 제공해야 한다.
USR-207	SBQE는 Java 개발 환경에서 JDBC 인터페이스를 제공해야 한다.
USR-208	SBQE는 메타데이터 등 자체 필요한 데이터의 관리를 위한 디스크 기반의 저장소를 가져야 한다.
USR-209	SBQE는 공간 빅데이터 저장관리 인프라의 API와 연동할 수 있어야 한다.
USR-210	SBQE는 하부 API가 확장될 때 다수의 API 셋을 호출할 수 있게 확장될 수 있어야 한다.
USR-211	SBQE는 하둡 분산환경에서 master-slave 구조로 운용 가능해야 한다.

❍ 공간 빅데이터 질의처리 엔진의 주요 기능을 정의함

- 질의 입력 수신: 외부로부터 공간 빅데이터에 대한 질의를 입력받을 수 있게 대기함
- 질의언어 파싱(parsing): 공간 데이터 타입과 공간 연산 조건(predicate)이 포함된 SQL 기반의 공간데이터 질의언어를 파싱하고 스키마에 맞는지 검증함
- 질의 계획수립(planning): 공간 질의언어 파싱 결과로부터 질의 실행 계획을 트리 형태로 생성함
- 질의 최적화: 공간 데이터와 비공간 데이터의 특성 및 데이터 분산을 고려하여

질의 전략을 수립함

- 질의 실행: 질의 실행계획을 공간 빅데이터 연산 API의 조합으로 생성하여 호출 및 실행하여 질의 결과 데이터를 생성함
- API 호출: 하부의 공간빅데이터 저장관리 인프라의 API 를 호출하고 결과 데이터를 얻어옴

❍ 사용자 요구사항과 주요 기능목록으로부터 개발자 관점에서 사용자 요구사항을 만족시키기 위한 공간 빅데이터 질의처리 엔진의 시스템 요구사항을 정의함

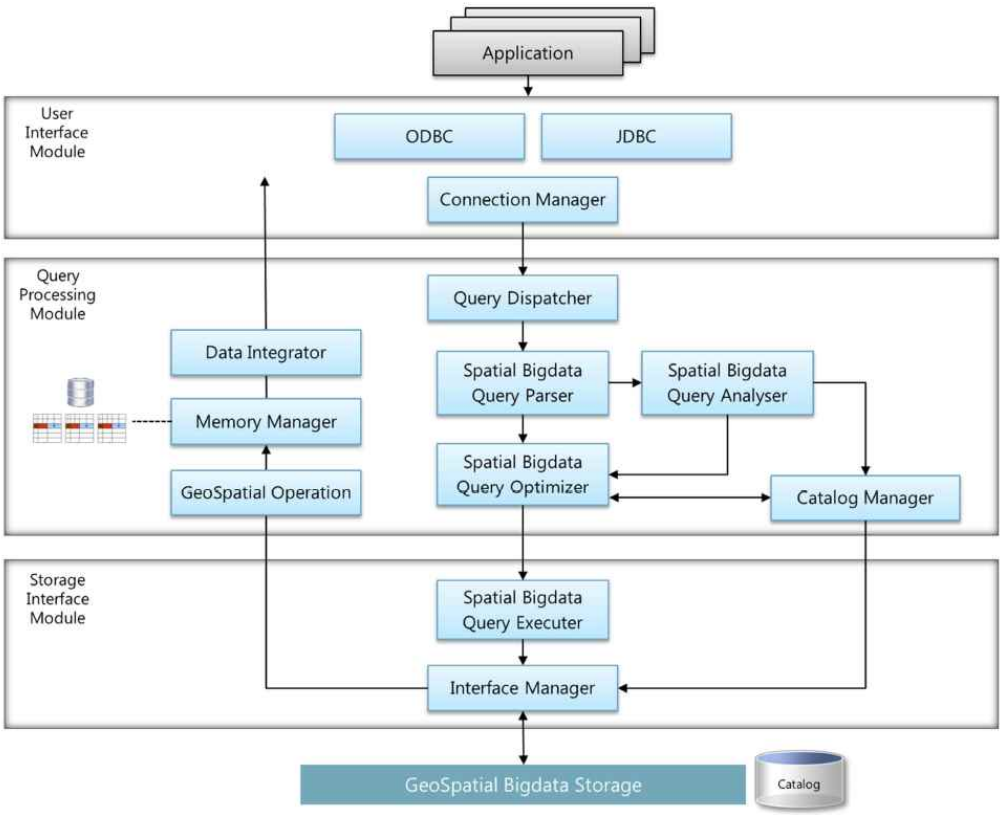
<표 2-15> 공간 빅데이터 질의처리 엔진 시스템 요구사항

요구사항 ID	내용
SFR-201	SBQE는 외부 SQL 호출을 수신하는 서버를 실행할 수 있어야 한다.
SFR-202	SBQE는 외부 SQL 호출을 수신하는 서버를 종료할 할 수 있어야 한다.
SFR-203	SBQE의 서버는 실행 후 대기상태로 외부 SQL 요청을 수신할 수 있어야 한다.
SFR-204	SBQE의 서버는 외부 SQL 요청을 내부 처리 모듈로 전달할 수 있어야 한다.
SFR-205	SBQE의 서버는 내부 처리 모듈의 처리 결과를 외부로 전송할 수 있어야 한다.
SFR-206	SBQE는 공간 질의언어 중 DML을 파싱할 수 있어야 한다.
SFR-207	SBQE는 공간 질의언어 중 SELECT문을 파싱할 수 있어야 한다.
SFR-208	SBQE는 공간 질의 실행계획을 생성할 수 있어야 한다.
SFR-209	SBQE는 속성/공간 데이터의 특성에 따라 질의 실행계획을 최적화할 수 있어야 한다.
SFR-210	SBQE는 외부의 공간 질의 실행 모듈을 호출할 수 있어야 한다.
SFR-211	SBQE는 HDFS의 API를 호출할 수 있어야 한다.
SFR-212	SBQE는 공간 질의 실행계획을 MR 연산으로 변환할 수 있어야 한다.
SFR-213	SBQE는 공간 질의 실행의 결과로 수신된 데이터를 디스크기반의 내부 저장소에 저장할 수 있어야 한다.
SFR-214	SBQE는 디스크기반의 내부 저장소에 저장된 데이터를 검색하여 읽을 수 있어야 한다.
SFR-215	SBQE는 공간 질의 실행의 결과로 수신된 데이터를 메인메모리에 저장할 수 있어야 한다.
SFR-216	SBQE는 메인메모리에 저장된 데이터를 검색하여 읽을 수 있어야 한다.
SFR-217	SBQE는 ODBC 인터페이스 호출을 입력받을 수 있어야 한다
SFR-218	SBQE는 ODBC 인터페이스 호출의 결과를 전송할 수 있어야 한다.
SFR-219	SBQE는 JDBC 인터페이스 호출을 입력받을 수 있어야 한다.
SFR-220	SBQE는 JDBC 인터페이스 호출의 결과를 전송할 수 있어야 한다.
SFR-221	SBQE는 질의 언어 입력을 위한 사용자 인터페이스를 제공해야 한다.
SFR-222	SBQE는 master-slave 노드간에 메시지 및 데이터를 전송할 수 있어야 한다.
SFR-223	SBQE는 공간 빅데이터 저장관리 인프라와 연동가능한 인터페이스를 제공해야 한다.

② [단위성과 B-2] 공간 빅데이터 질의 처리 엔진 기능분석, 설계 (1차연도)

❍ 공간 빅데이터 저장 프레임워크의 공간 데이터에 대하여 SQL-like 공간 질의를 처리하는 질의처리 엔진의 기본적인 기능과 구조를 설계함

- 공간 빅데이터 질의처리 엔진은 공간 빅데이터 저장관리 시스템의 상부에 API를 통하여 결합되어 빅데이터 플랫폼 기반으로 공간 DBMS와 같은 구조를 구성함
- 공간 빅데이터 질의처리 엔진의 구조는 <그림 2-46>과 같음



<그림 2-46> 공간 빅데이터 질의처리 엔진 구조

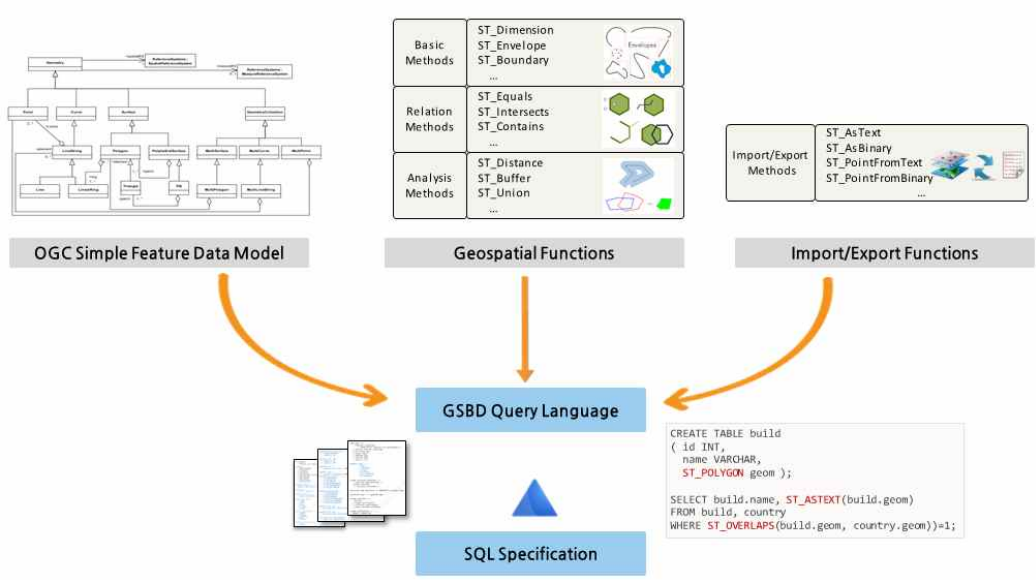
❍ 공간 빅데이터 질의처리 엔진의 주요 기능을 다음과 같이 설계함

- SQL 기반의 질의언어를 파싱하여 내부 형태의 파스 트리(parse tree)를 생성
- 질의언어의 의미(semantics)가 잘못되었는지 검증
- 공간 빅데이터 저장관리 시스템의 메타데이터를 조회하여 질의계획을 최적화
- 질의계획으로부터 하부 저장관리시스템의 API에 의한 질의실행을 수행
- 질의실행의 결과로 수신되는 데이터를 통합 및 처리하여 질의결과 데이터를 생성
- 질의결과 데이터를 요청한 클라이언트에게 표준적인 인터페이스로 전송

③ [단위성과 B-3] 공간 빅데이터 질의 언어 정의 (1차연도)

❍ SQL 기반으로 공간 데이터 지원을 위하여 OGC Simple Features의 공간데이터 타입과 메서드를 확장하여 공간 빅데이터 질의언어를 정의함

- 공간 빅데이터를 질의할 수 있는 SQL 기반의 질의언어를 정의함
- 공간데이터를 처리하기 위하여 OGC의 Simple Feature 데이터 모델 기반 데이터 타입, 공간 연산 함수, 공간 데이터 입출력 함수를 정의함



<그림 2-47> 공간 빅데이터 질의언어 정의

- 데이터 정의(DDL) 및 데이터 조작(DML) 중 기본적인 것들을 우선적으로 정의함(표 2-16에서 질의언어 문법을, 표 2-17에서 질의언어를 예시함)

<표 2-16> 공간 빅데이터 질의언어 문법(일부 예시)

```
<query specification> ::=
    SELECT [ <set quantifier> ] <select list> <table expression>
<select list> ::=
      <asterisk>
    | <select sublist> [ { <comma> <select sublist> }... ]
<select sublist> ::=
      <derived column>
    | <qualifier> <period> <asterisk>
<table expression> ::=
    <from clause>
    [ <where clause> ]
    [ <group by clause> ]
    [ <having clause> ]
```

<표 2-17> 공간 빅데이터 질의언어(예시)

```
SELECT geom
FROM building
WHERE ST_INTERSECTS (ST_PolyFromText('POLYGON( ( 67 13, 67 18, 59 18, 59 13, 67 13) )',1))

SELECT geom
FROM building
WHERE ST_CONTAINS (geom, ST_POINTFROMTEXT('POINT(36 128)',0))=1;
```

④ **[단위성과 B-4] 공간 빅데이터 ODBC/JDBC 인터페이스 설계 (1차연도)**

❍ 기존 관계형 데이터베이스와의 데이터 호환성 및 공간 빅데이터 DBMS의 원활한 사용을 위해 ODBC/JDBC/MapReduce 인터페이스를 설계함

❍ 응용프로그램에서 공간질의를 실행하기 위해 사용하는 공간 빅데이터 DBMS ODBC 인터페이스 설계

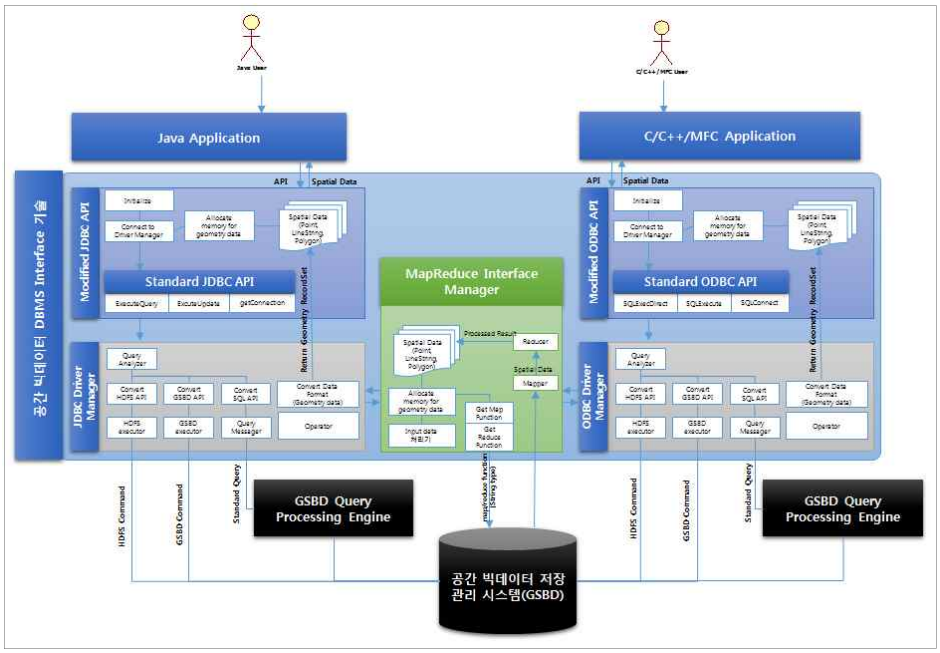
- 공간 정보의 저장을 위한 데이터 포맷 정의 및 변환기 설계
- 표준 ODBC API 인식을 위한 질의분석기 설계
- ODBC 사용자를 위한 HDFS/GSBD/SQL API 변환기 설계

❍ 응용프로그램에서 공간질의를 실행하기 위해 사용하는 공간 빅데이터 DBMS JDBC 인터페이스 설계

- 공간 정보의 저장을 위한 데이터 포맷 정의 및 변환기 설계
- 표준 JDBC API 인식을 위한 질의분석기 설계
- JDBC 사용자를 위한 HDFS/GSBD/SQL API 변환기 설계

❍ 응용프로그램에서 공간질의를 실행하기 위해 사용하는 Map/Reduce 인터페이스 설계

- Map/Reduce 입/출력 데이터 포맷 설정
- Map Function 생성 모듈
- Reduce Function 생성 모듈



<그림 2-48> 공간 빅데이터 DBMS 인터페이스 시스템

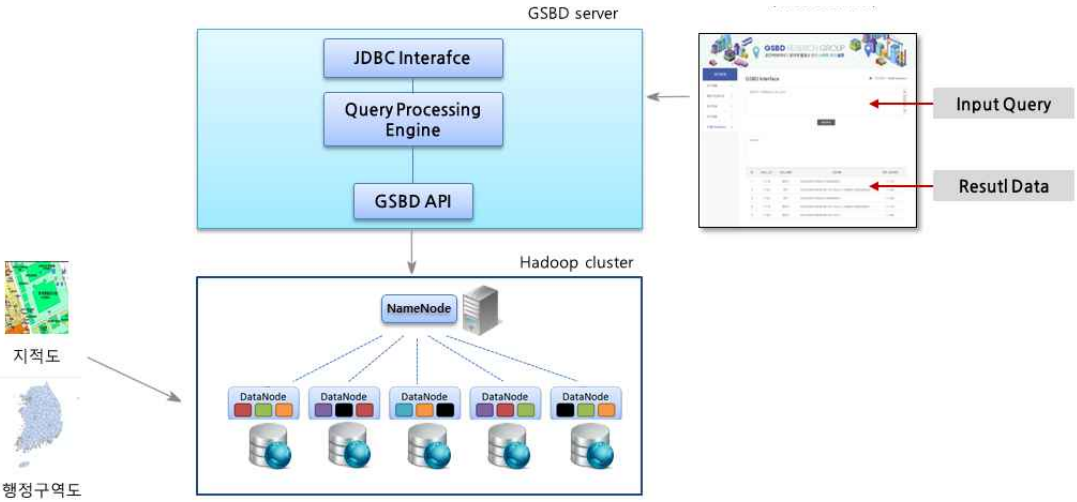
⑤ [단위성과 B-5] 공간 빅데이터 DBMS 인터페이스 기본기능 모듈 구현 (1차연도)

❍ 공간 빅데이터 질의처리 엔진의 SQL 질의 및 JDBC 인터페이스 기본 기능을 개발하여 공간 빅데이터 DBMS 인터페이스 기본 기능을 구현하고 테스트함

- 질의처리 엔진 접속 기능
- 콘솔에 의한 SELECT 문장 실행 요청 기능
- JDBC 클라이언트에 의한 SELECT 문장 실행 요청 기능
- 결과 데이터 수신 기능

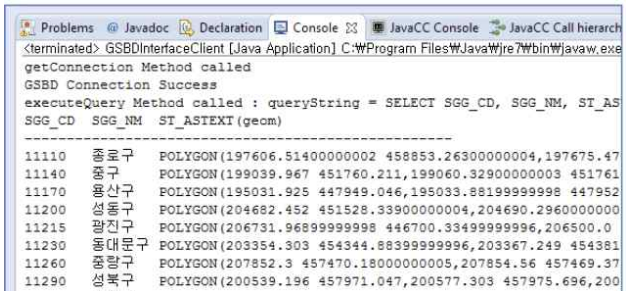
❍ 공간 빅데이터 DBMS 인터페이스 기본 기능 개발

- 질의언어 정의 및 파서 설계에 기반하여, 공간데이터를 포함한 SELECT 문에 대한 기본 파싱기능을 개발함
- 기본 검색을 위하여 공간 빅데이터 저장관리시스템에 저장된 데이터에 대한 SELECT * 및 기본적인 WHERE 조건에 의한 질의를 실행함



<그림 2-49> 공간 빅데이터 DBMS 인터페이스 기본기능 모듈

❍ 콘솔 클라이언트를 통하여 질의처리 엔진에 접속하고 인터페이스를 호출하여 메시지와 데이터가 교환되는 것을 테스트하여 확인함

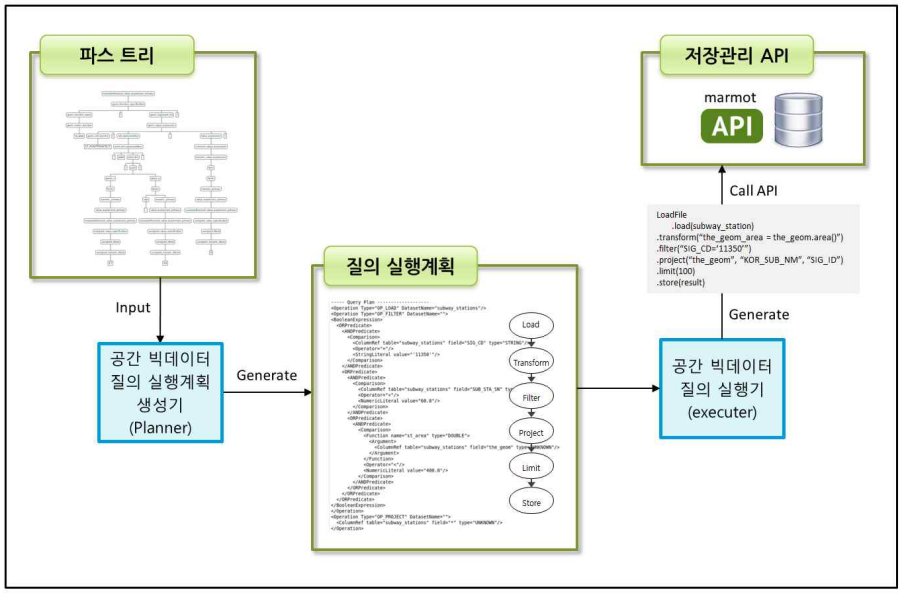


<그림 2-50> 공간 빅데이터 DBMS 인터페이스 기본기능 테스트

⑥ [단위성과 B-6] 공간 빅데이터 질의 플랜 생성 모듈 구현 (2차연도)

❍ 공간 질의언어에 대한 질의처리의 논리적 실행순서를 생성하는 공간 빅데이터 질의 플랜 생성 모듈을 개발함

- 파싱 및 의미검증 과정을 거친 파스 트리를 분석하여 데이터 접근을 최소화하기 위한 질의처리의 단위연산(operation)의 논리적 실행순서를 질의 플랜(query plan)으로 생성함
- 질의 플랜 생성 모듈은 파스 트리로부터 기본 질의 실행계획을 수립하여 연산 트리(operation tree)를 생성함
- 질의 플랜을 사용자 또는 개발자에게 출력할 경우나 모듈간에 교환하는 경우 표준적인 방법으로 표현하기 위하여 XML 기반의 표기 방식을 사용함



<그림 2-51> 질의실행계획 생성기와 질의실행기

```
----- Query Plan ------------------
<Operation Type="OP_LOAD" DatasetName="subway_stations"/>
<Operation Type="OP_FILTER" DatasetName="">
<BooleanExpression>
  <ORPredicate>
    <ANDPredicate>
      <Comparison>
        <ColumnRef table="subway_stations" field="SIG_CD" type="STRING"/>
        <Operator="="/>
        <StringLiteral value="'11350'"/>
      </Comparison>
    </ANDPredicate>
    <ORPredicate>
      <ANDPredicate>
        <Comparison>
          <ColumnRef table="subway_stations" field="SUB_STA_SN" type="LONG"/>
          <Operator="<"/>
          <NumericLiteral value="60.0"/>
        </Comparison>
      </ANDPredicate>
```

<그림 2-52> XML로 표현된 질의 플랜(일부)

⑦ [단위성과 B-7] 스트림 공간 빅데이터 질의언어 파싱 및 의미검증 모듈 구현 (2차연도)

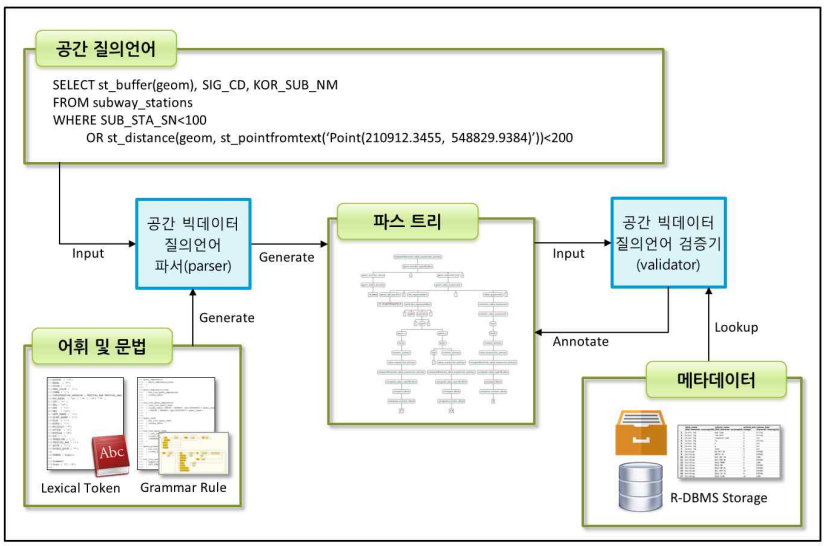
❍ 공간 빅데이터 질의언어를 처리하여 질의수행을 위한 준비를 하는 파서(parser) 및 의미검증기(validator) 모듈을 개발함

❍ 공간 질의언어 파서를 개발함

- 공간 질의언어를 파싱(parsing)하여 주어진 SQL 문장이 적합한 구문인지를 검사하는 구문(syntax) 분석 및 파싱을 수행하고 파스 트리(parse tree)라는 트리 형태의 데이터구조를 출력함
- SQL 문장을 파싱하는 단계에서 공간데이터 타입을 지원하기 위하여 데이터 타입과 조건함수(predicate)에 대하여 인식이 가능하도록 확장된 문법을 사용함
- 파싱 모듈은 파서 생성기(parser generator)인 ANTLR[1] 4.5 기반으로 정의된 입력 문법으로부터 생성함

❍ 공간 빅데이터 질의 의미검증 모듈을 개발함

- 테이블명, 데이터타입, 함수명, 파라미터타입 등 질의언어 파스 트리 각 요소의 의미를 검증(validation)함
- 질의언어의 SELECT절 및 WHERE절의 조건부에 공간데이터 필드나 함수들이 올 수 있으며 이러한 함수들은 정의된 공간 함수 명세에 대하여 검증함
- 공간빅데이터 저장소에 입력되어 있는 데이터셋의 테이블명, 칼럼명 등 메타데이터[2]를 조회하여 질의언어 요소의 의미를 검증하며, 이를 위하여 카탈로그 인터페이스 모듈을 개발함



<그림 2-53> 질의언어 파서와 검증기

1) 주어진 문법을 입력으로 하여 parser의 기본 프레임워크를 생성하는 프로그램으로, Java로 된 오픈소스 소프트웨어이다. UNIX나 Linux에서의 Yacc과 같은 역할을 한다.
2) 본 시스템에서 메타데이터는 카탈로그라고 부르며 메타데이터의 저장관리는 HDFS가 아닌 RDBMS를 사용한다.

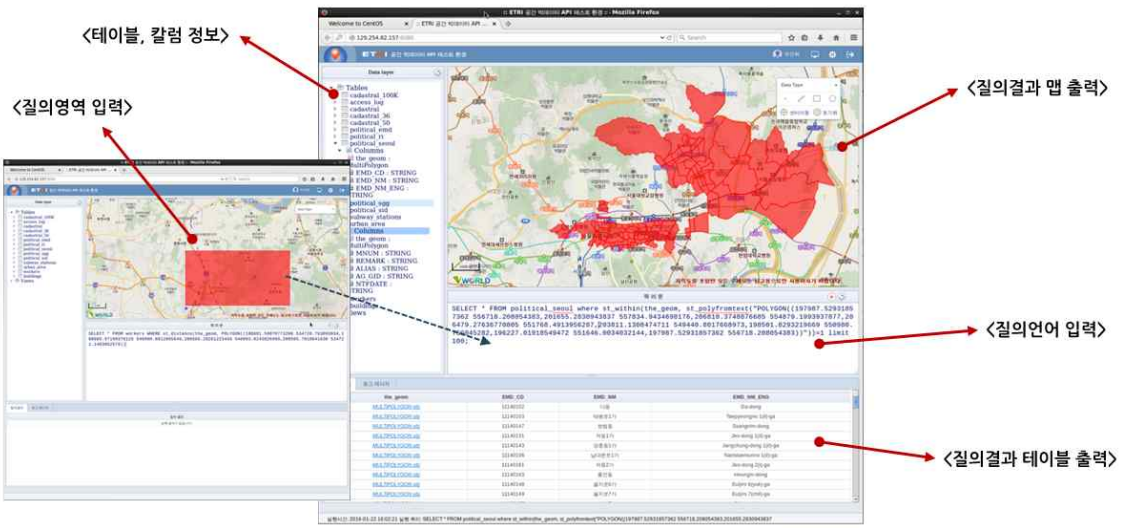
⑧ [단위성과 B-8] 스트림 공간 빅데이터 질의처리 모듈(기본기능) 구현 (2차연도)

❍ 공간빅데이터 질의언어에서 질의언어 파싱, 의미검증, 질의플랜 생성 단계를 거친 질의플랜에 따라 실제 저장관리시스템에서 데이터를 접근하여 가져올 수 있는 질의실행 모듈을 개발함

- 공간빅데이터 저장관리 시스템으로부터 실제 데이터를 가져오기 위하여 질의플랜을 공간 빅데이터 저장관리 인프라의 단위연산(operation) API로 변환하고 실행함
- 질의실행은 스트림 처리 방식으로 이루어지며 한번에 데이터를 가져오지 않고 데이터를 하나씩 fetch하기 위한 함수(저장 시스템에서 제공되는 API)를 반복하여 호출하여 데이터를 가져옴
- 데이터 로딩, 필터, 프로젝션, 저장 등 질의언어의 기본 기능으로 지원되는 기능에 대하여 실행 기능이 지원됨
- SQL 형태로 된 공간질의언어는 저장시스템 API 메서드들의 조합으로 변환되어 호출하는 형태로 공간 질의를 실행할 수 있음

❍ 공간 빅데이터 질의를 위한 웹 기반 클라이언트 인터페이스를 개발함

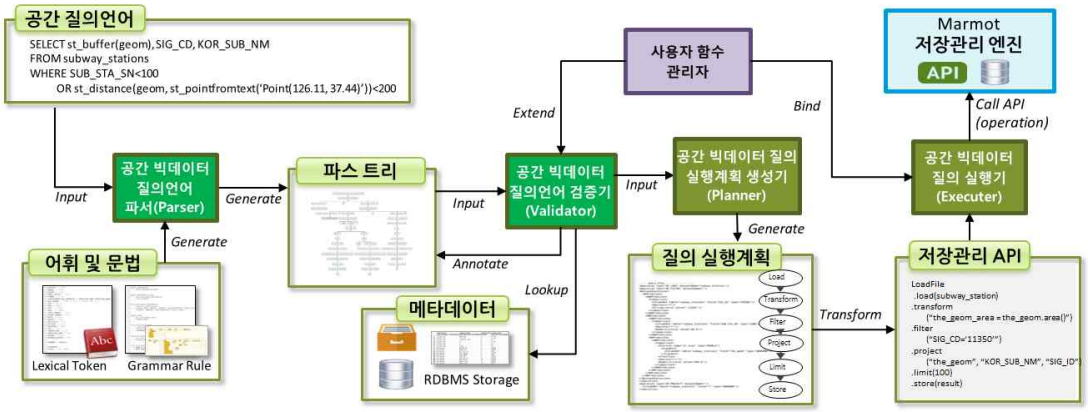
- 공간 빅데이터 질의처리 서버에 연결하여 공간질의 입출력 테스트에 활용할 웹 기반의 클라이언트 인터페이스를 개발함(그림 2-54)
- 지도에서 입력한 질의영역 공간객체를 좌표 텍스트로 변환하고 공간질의 결과를 화면에 출력하는 기능을 개발함
- 공간빅데이터 질의처리 엔진과 웹 클라이언트 인터페이스 사이에 JDBC 서버/클라이언트 모듈의 프로토타입을 개발하여 연동함



<그림 2-54> 웹 기반 클라이언트 인터페이스

⑨ [단위성과 B-9] 스트림 공간빅데이터 분산 질의처리 모듈 구현 (3차연도)

❍ 공간 빅데이터 저장관리 프레임워크에 저장된 공간데이터에 대한 SQL 기반 질의언어에 의하여 공간 질의를 처리하는 질의처리 엔진 모듈을 개발함



<그림 2-55> 공간 빅데이터 질의처리 엔진 실행 흐름

❍ 공간 빅데이터 질의언어 파싱 모듈 및 의미검증 모듈을 확장 개발함

- 데이터 정의(DDL) 및 데이터 조작(DML) 기능들에 추가하여 공간 조인을 포함한 공간 빅데이터 질의언어 문법을 확장 정의함
- 데이터셋의 메타정보를 저장하는 카탈로그의 API를 활용하여 질의언어 내의 테이블명, 칼럼명, 칼럼 데이터타입을 검증하는 의미검증 기능을 확장 개발함
- 함수의 입력 파라미터 개수, 데이터 타입, 반환 타입에 대한 함수 명세를 관리하고 질의언어 내의 함수표현식을 검증하는 기능을 개발함

❍ 분산 저장 시스템 및 공간색인 연동 질의처리 모듈을 개발함

- 하부의 공간 빅데이터 저장관리 시스템이 사용하는 저장관리 인프라 API(Application Program Interface) 연동 기능을 개발함
- 공간 인덱스 생성 질의언어 지원 기능을 개발함
- 저장관리 시스템의 처리방법의 특성에 따라 질의의 성격에 적합한 API를 호출함

❍ 스트림 기반 공간 빅데이터 질의실행 및 처리 기능을 개발함

- 질의 실행계획을 공간 빅데이터 저장관리 시스템의 레코드 스트림 연산자 기반 단위연산(operation) API로 변환하고 실행함
- 대용량 데이터를 처리하는 스트림 기반 처리 방법에 의하여 공간빅데이터 저장관리 시스템 및 공간색인 시스템과 연동하고 질의결과를 교환함

❍ 공간 조인 질의언어 지원 및 처리기능을 개발함

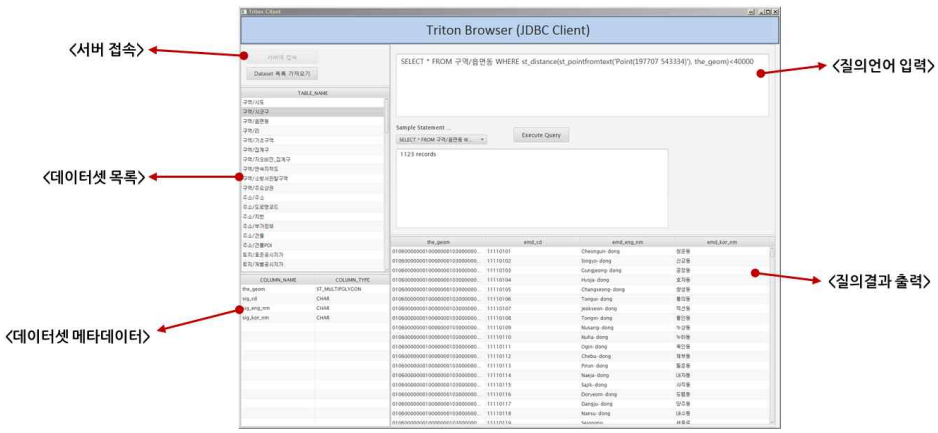
- 질의언어 수준에서 공간 조인을 표현하고 수행할 수 있도록 JOIN...ON 키워드를 사용하여 공간 관계연산 및 거리기반 조건에 의한 공간 조인 기능을 지원함

⑩ [단위성과 B-10] 공간 빅데이터 DBMS ODBC/JDBC 클라이언트/서버 모듈 구현 (4차연도)

❍ SQL 기반 질의언어 및 표준 인터페이스를 사용하여 공간 빅데이터를 접근, 조회하고 개발자가 응용프로그램을 편리하게 개발할 수 있게 해주는 공간 빅데이터 DBMS JDBC 클라이언트/서버 모듈을 개발함

❍ 공간 빅데이터 질의언어 처리 기능 위에서 동작하고 외부의 데이터 요청에 대한 처리를 하는 JDBC(Java DataBase Connectivity) 서버를 개발함

- JDBC 인터페이스를 구현하여 Java 언어를 사용하는 외부 클라이언트가 사용할 수 있는 JDBC 드라이버를 개발함
- 외부 클라이언트는 JDBC 드라이버를 등록하여 사용할 수 있음
- 공간 빅데이터 저장소의 고유 API를 사용하는 대신 다른 공간 DBMS를 사용할 때와 동일한 방식으로 SQL 언어 기반으로 공간 빅데이터 저장 시스템의 데이터를 접근 및 조작할 수 있도록 지원함
- Connection, Statement, ResultSet, ResultSetMetaData 등 JDBC 객체(Java의 interface) 및 이들의 표준 API 함수를 구현함
- JDBC 드라이버를 사용하는 Java 프로그램의 테스트와 검증을 위하여 Java로 작성된 클라이언트 프로그램을 개발하였으며 저장시스템 서버 접속, 메타데이터 조회, 질의언어 요청 및 결과(레코드) 수신 등의 기능을 구현함



<그림 2-56> 공간 빅데이터 JDBC 드라이버를 활용한 클라이언트

❍ 공간 빅데이터 DBMS ODBC 인터페이스를 개발함

- C++ 개발 환경에서 공간 빅데이터 저장관리 및 질의처리 엔진의 기능을 표준적인 방법으로 사용할 수 있도록 하는 ODBC(CLI) 인터페이스를 개발함
- JNI handler를 활용하여 C++을 통해 공간 빅데이터 저장관리 및 질의처리 엔진 서버의 JDBC API 함수를 연동할 수 있도록 ODBC(CLI) 표준 함수를 인터페이스로 구현함

⑪ [단위성과 B-11] 스트림 공간 빅데이터 질의 스케줄링 모듈 구현 (4차연도)

❍ 스트림 형태의 공간 빅데이터를 처리하기 위해 파이프라인 방식으로 데이터를 처리하는 하둡 기반의 시스템을 설계하고 구현함

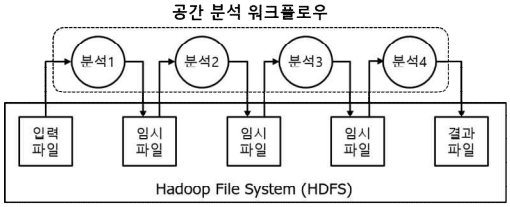
- 공간 연산들의 체인으로 분석 응용을 기술하고, 입력 공간 빅데이터에 크기에 따라 Map/Reduce 작업 또는 In-Memory 작업을 동적으로 선택하여 수행함으로써 다양한 형태 및 크기의 공간 데이터를 고속으로 처리할 수 있음
- Map/Reduce 방식으로 처리하는 경우 사용되는 Map/Reduce 횟수를 최소화하는 기능을 통해 고속의 공간 분석 응용의 수행 성능을 보여줌
- 스트림 기반의 파이프라인(pipeline) 방식의 데이터 처리 기능을 제공하여 데이터의 병렬 처리를 통해 한정된 메모리 환경에서 기존의 공간 데이터 처리 방법이 가지는 한계를 극복하고 높은 확장성을 제공함

❍ 데이터 병렬 작업, 컨트롤 병렬 작업, 로컬 메모리 사용을 통해 기존의 빅데이터 처리 시스템에 비해 차별성을 가짐

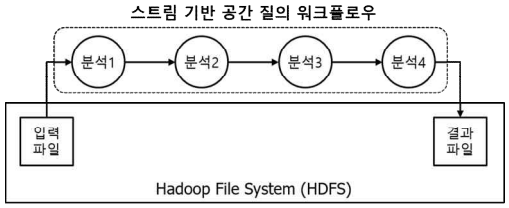
- 사용자는 빅데이터에 관련된 전문 지식의 필요 없이 개발하고자 하는 공간 분석 알고리즘만 파악된다면 다수의 공간/비공간 데이터 처리 연산자를 조합하여 개발하고자 하는 공간분석 작업을 구현할 수 있음
- 특정한 데이터의 특성에 종속되지 않도록 플러그인 방식으로 구현되어 높은 확장성을 제공함
- 스트림 기반의 파이프라인 방식의 데이터 처리 기능을 제공하여 제한된 메모리상에서 데이터의 병렬 처리를 통해 성능이 향상되도록 구현함

❍ 스트림 기반의 처리방법으로 공간분석을 스케줄링하여 파일 I/O를 감소시켜 성능 향상을 가져옴

- 여러 분석들을 묶어 한번에 스트림 형식으로 처리하여 중간 파일 생성으로 인한 불필요한 파일 I/O를 제거시킴
- 분석에 사용하는 공간 빅데이터 파일 크기 증가에 따른 I/O 부담 증가가 한정됨
- 전체 분석응용을 구성하는 연산 모듈 개수가 증가하여도 이에 따른 I/O 부담 증가가 한정됨



<그림 2-57> 기존 공간분석 처리 방식

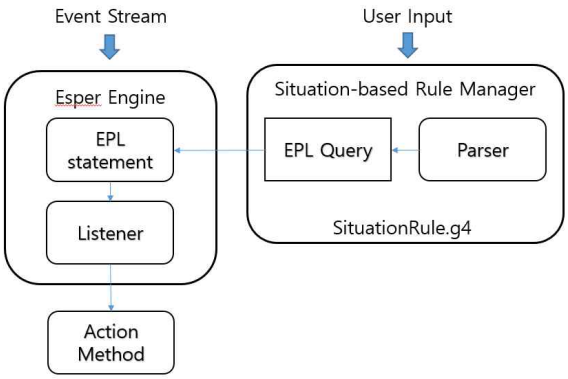


<그림 2-58> 스트림 기반 분석 처리 방식

⑫ [단위성과 B-12] 스트림 공간 빅데이터 이벤트 처리 모듈 구현 (4차연도)

○ ECA 규칙 기반 컴퓨팅에 의한 이벤트 처리를 위해 CEP를 이용하는 이벤트 처리 도구를 개발하여 이벤트 처리 성능을 향상시킴

- ECA(Event-Condition-Action) 규칙은 이벤트들의 조건을 고려하여 이벤트에 관한 동작이 수행되는 이벤트 중심의 컴퓨팅 방법임
- 모든 기능은 유입되는 이벤트에 의해서 동작하고, 이벤트가 규칙 조건의 논리 테스트를 만족하게 되면 규칙에 정의한 대응 동작을 취하는 방식으로 운영됨
- 복합적인 예외에 대한 설명의 부담을 줄이기 위해 예외 상황에 대한 '상황 패턴'을 정의하고, 해당하는 예외상황이 발생할 경우 대응할 동작을 정의함
- 공간 빅데이터 이벤트 처리 모듈은 정규식 기반의 구성 규칙정의를 통해 이벤트 스트림을 처리하며 <그림 2-59>와 같이 규칙과 엔진이 계층구조로 이루어져 있음

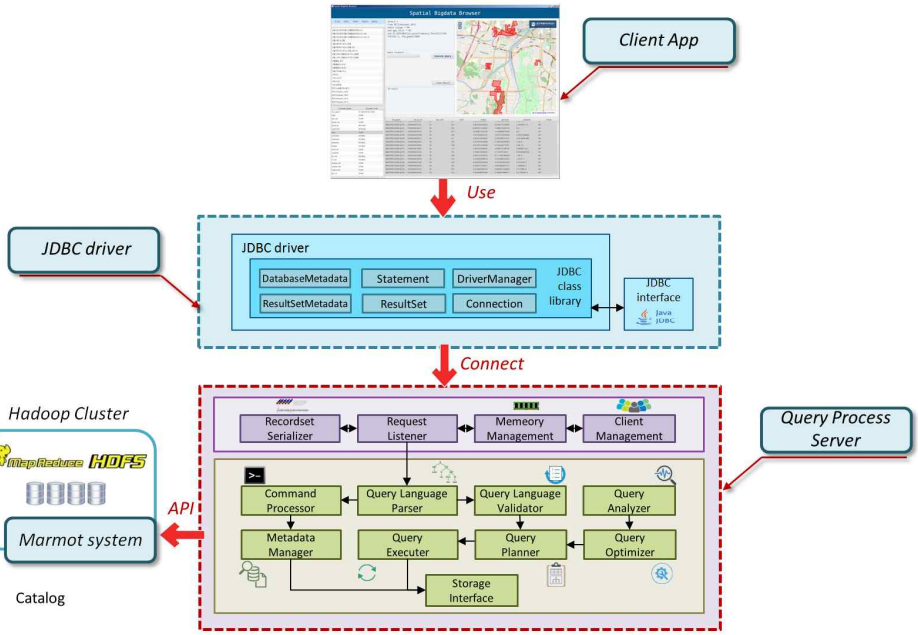


<그림 2-59> 이벤트 처리 시스템 구조

- 사용자가 입력하는 규칙은 상황규칙 문법을 통해서 어휘 및 구문 분석이 이루어짐
- 사용자가 상황규칙 문법을 통하여 정의한 상황 및 동작 규칙은 파서에서 Condition과 Action 부분으로 나누어, 각각 EPL(Event Processing Language) Query와 Action Method를 생성함
- Esper 엔진은 이 EPL 문장을 통해서 지속적인 이벤트 스트림의 질의를 처리함
- 기존의 ECA 기반의 컴퓨팅 중 Event와 Condition에 해당하는 부분을 EPL 문장을 통해서 수행할 수 있음
- Esper는 EPL 질의문을 통해 지속적으로 들어오는 이벤트 스트림을 받아 실시간으로 스트림 안의 상황 패턴을 찾아냄
- 이벤트들은 EPL 내부의 조건 절 및 패턴을 통해 예외상황으로 판별되고, 판별 결과는 리스너로 전달되어 리스너 콜백을 이용하여 Action을 수행함
- Action의 내용은 사용자가 기술하는 자바코드로 정의되며, 이 동작은 실시간으로 들어오는 데이터에 대해 지속적이고 빠르게 수행됨

⑬ [단위성과 B-13] 스트림 공간 빅데이터 인터페이스 통합처리 서버 프레임워크 구현 (5차연도)

❍ 공간 빅데이터 질의처리 서버, JDBC 드라이버, Java 응용프로그램 (UI 기반 클라이언트)을 확장개발하고 이들을 연계통합한 프레임워크를 구현함



<그림 2-60> 공간 빅데이터 인터페이스 프레임워크 전체 구조

❍ 공간 빅데이터 질의처리 서버

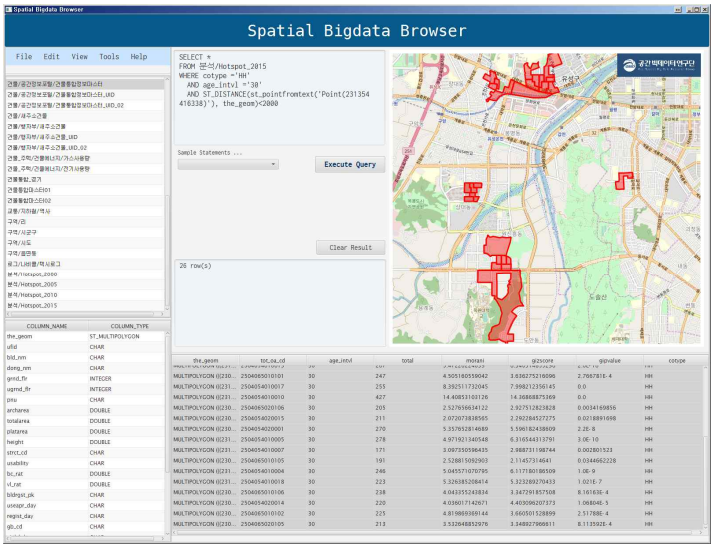
- 저장관리엔진에 연결하는 커넥터, 질의언어를 분석하여 질의실행계획을 생성하는 질의엔진, 외부 요청을 수신하여 레코드를 전달하는 서버로 구성됨
- ALTER TABLE 등 DDL 문장을 지원하도록 질의처리 서버의 기능을 확장함

❍ 공간 빅데이터 클라이언트(UI 클라이언트 응용프로그램)

- 클라이언트의 기본 기능에 추가하여 질의언어 입력 도구, 지도 데이터 출력 등의 기능을 개발함
- JDBC 드라이버 라이브러리를 참조하여 서버의 시스템과 통신하고 수행됨
- 화면 출력은 OpenLayers API를 활용한 Web Control을 응용프로그램에 포함하여 구현함

❍ 공간 빅데이터 인터페이스 통합처리 서버 프레임워크 개발

- 공간 빅데이터 인터페이스는 JDBC 드라이버의 서버 인터페이스 호출과 데이터 송수신, 객체 관리 등의 기능을 통하여 통합됨
- 공간 빅데이터 질의처리 엔진 구조를 기반으로, 공간 빅데이터 질의처리 서버와 JDBC 드라이버, UI 기반의 공간 빅데이터 클라이언트를 개발하여 연계함



<그림 2-61> 공간 빅데이터 클라이언트

❍ 공간 빅데이터 저장관리 인프라 기능 및 질의처리 인터페이스를 대용량 공간 데이터의 데이터 오류 검사에 활용

- (선정 기준) 현재 '국가공간정보포털'에서 제공하는 지도 데이터 중에서 자주 활용되면서 대용량 데이터인 '건물통합정보마스터' 데이터셋[3] (데이터 크기 약 1,439만 건, 약 11.1Gbyte)에서 공간 및 비공간 데이터의 오류 검출
- 비공간 데이터의 경우 UFID 값의 포맷 비정상(누락, 특수문자 포함, 자리수 불일치 등)을 탐지함
- 공간 데이터의 경우 건물의 영역(geometry)이 서로 겹치는지 판단에 의하여 오류 가능성을 탐지하며, 이 연산은 두 건물 영역(Polygon)의 겹침(Overlaps) 공간연산에 대한 데이터셋의 자체조인(self join)에 의하여 처리됨
- 두 개의 서로 다른 건물의 영역(좌표)은 서로 겹치지 않는다는 기본적 가정하에, 서로 겹치는 건물은 중복 입력된 것이거나 서로 다른 버전이라는 잠재적 오류 가능성으로 판단함[4]
- (결과 의미) 이러한 대용량 데이터 처리는 데이터 용량 문제로 기존 GIS로는 직접 수행할 수 없으며 본 공간 빅데이터 저장관리 시스템의 기능을 활용하여 수행 가능함
- (결과 의미) '건물통합정보마스터' 데이터셋에서 UFID 값의 포맷 비정상 오류를 포함한 심각한 오류를 발견할 수 있었으며, 연속지적도를 포함한 데이터셋에서도 오류를 발견하였음
- (결과 의미) 다른 대용량 공간데이터에도 존재할 수 있는 잠재적 오류를 탐지하는데 본 시스템을 사용할 수 있을 것으로 판단됨

3) 데이터셋의 출처는 국가공간정보포털이며 shp 파일을 본 시스템에 import한 뒤 연산 처리를 수행하였음
4) 겹치는 것으로 판단된 건물은 확실한 오류라는 의미가 아니라 오류 '가능성'이 있는 후보라고 보아야 하며, 실제 오류 여부는 원본 데이터의 관리 정책에 기반하여 데이터 생성 또는 관리 주체가 판단하게 됨

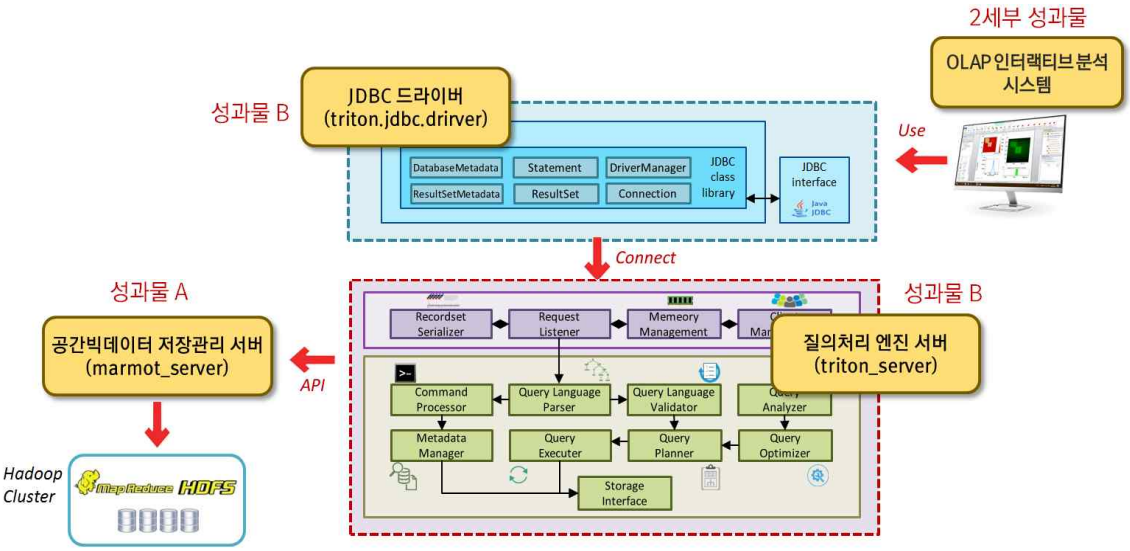


<그림 2-62> 건물 영역의 중첩 조사에 의한 잠재적 오류 식별

⑭ [단위성과 B-14] 스트림 공간 빅데이터 분산 질의처리 최적화 모듈 구현 (6차연도)

❍ 전년도까지 개발한 공간 빅데이터 질의처리 엔진과 JDBC 드라이버를 테스트하고 안정화 및 기능 고도화를 수행함

- JDBC 드라이버의 배포에 의하여 질의처리 서버를 2-3세부 OLAP 인터랙티브 분석 시스템과 연동함
- 외부 시스템 연동에 의하여 오류를 수정하고 기능을 보완함
- 메타데이터와 스키마 관리 일부 기능을 추가함

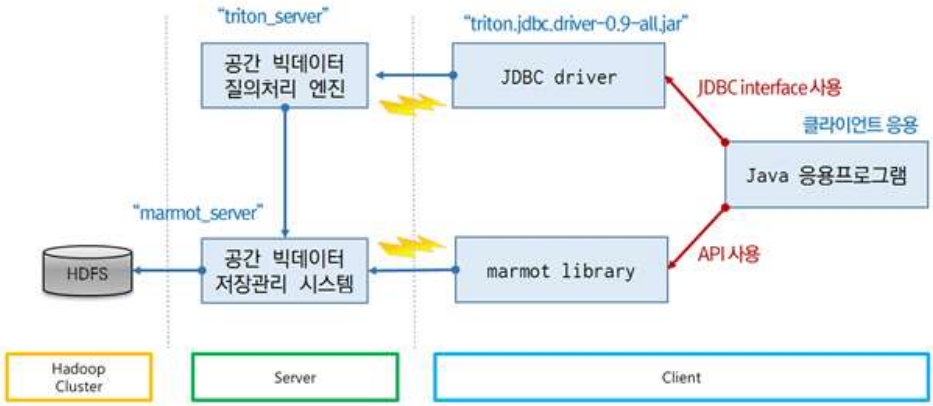


<그림 2-63> 공간 빅데이터 질의처리 시스템 구조 및 외부 연동

❍ 스트림 기반으로 데이터를 처리할 수 있는 질의는 중간 파일을 생성하지 않고 스트림 기반으로 데이터를 전송하는 구조로 모듈을 구현

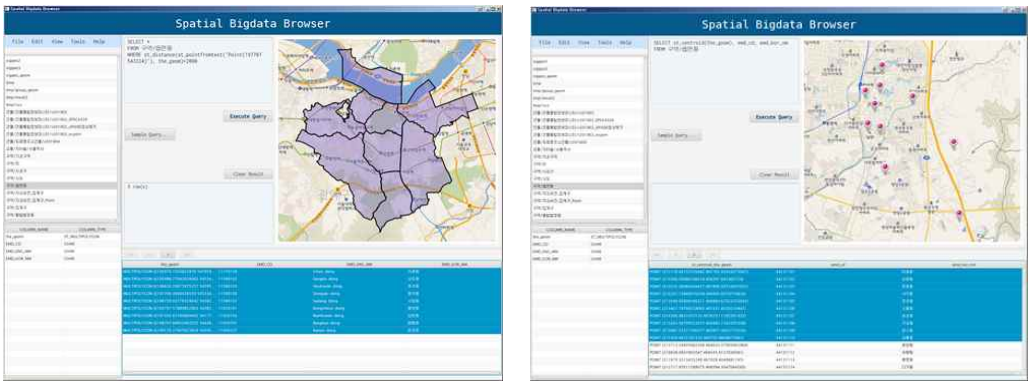
- 전체 선택(SELECT ALL) 및 조건 검색(WHERE), 함수 질의의 일부에 대하여 스트림 기반 데이터 처리를 하여 클라이언트 프로그램의 최초응답시간 개선

❍ 공간 빅데이터 질의처리 엔진의 기능에 대하여 질의 유형별로 13개의 범주로 구분된 테스트케이스를 정의하여 기능 테스트를 수행함



<그림 2-64> 공간 빅데이터 질의처리 엔진 테스트 구조

- 클라이언트 응용프로그램이 JDBC 드라이버를 사용하여 서버에 접속하고 UI를 통하여 공간 빅데이터 질의처리 엔진(서버)의 기능을 테스트함
- 클라이언트 프로그램에서 지도 화면은 공간정보오픈플랫폼(VWorld)의 API를 활용하여 출력함



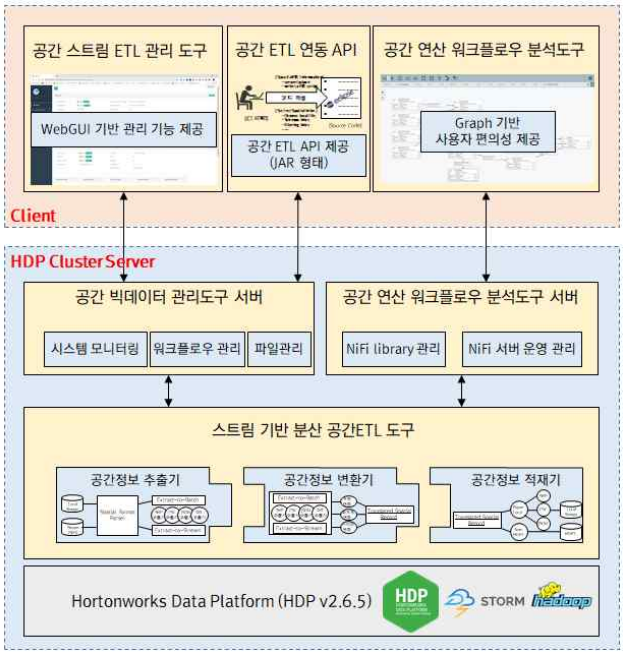
<그림 2-65> 공간 빅데이터 클라이언트 프로그램에 의한 기능 테스트

❍ 공간 빅데이터 사업단 시스템에 성과물(공간 빅데이터 질의처리 엔진)을 설치하고 실행 테스트하여(테스트케이스14) 동작을 실증함

- 공간 빅데이터 플랫폼에 저장된 데이터를 SQL기반 질의언어를 통하여 질의하여 결과 데이터를 가져오는 기능을 확인함
- 2-3세세부 성과물 ‘공간 빅데이터 OLAP’과 연계한 형태로 테스트하여 타 세부 성과물에서 JDBC 드라이버를 통하여 공간 데이터를 질의하는 기능을 확인함

(다) 스트림 기반 분산 공간 ETL 도구(핵심성과 D)

○ 공간 빅데이터 저장 프레임워크에 외부 공간 빅데이터를 저장하기 위한 ETL 규칙을 정의하고, 분산 환경에서 동작 가능한 ETL 워크플로우를 생성하여 배치/스트림 기반 병렬 및 분산 형태로 구동시키는 엔진 및 이를 통해 ETL 작업을 수행하는 시스템



<그림 2-66> 스트림 기반 분산 공간 ETL 도구(핵심성과D)

○ 특징

- 공간 빅데이터를 효율적으로 관리하기 위한 웹 기반의 관리도구와 공간 ETL 핵심 기능의 활용도를 높이고자 Apache NiFi 라이브러리 형태로 제공되며 일반 Java 개발자를 위해 API 형태로도 제공됨
- 공간 빅데이터 유형에 따라 배치/스트림 공간 정보 ETL 서비스를 제공함
- 공간 ETL 작업 중 사용자의 반복적인 추출/변환/적재 명령 입력 횟수를 줄이기 위해 워크플로우 관리기능을 제공함

○ 시스템 구조

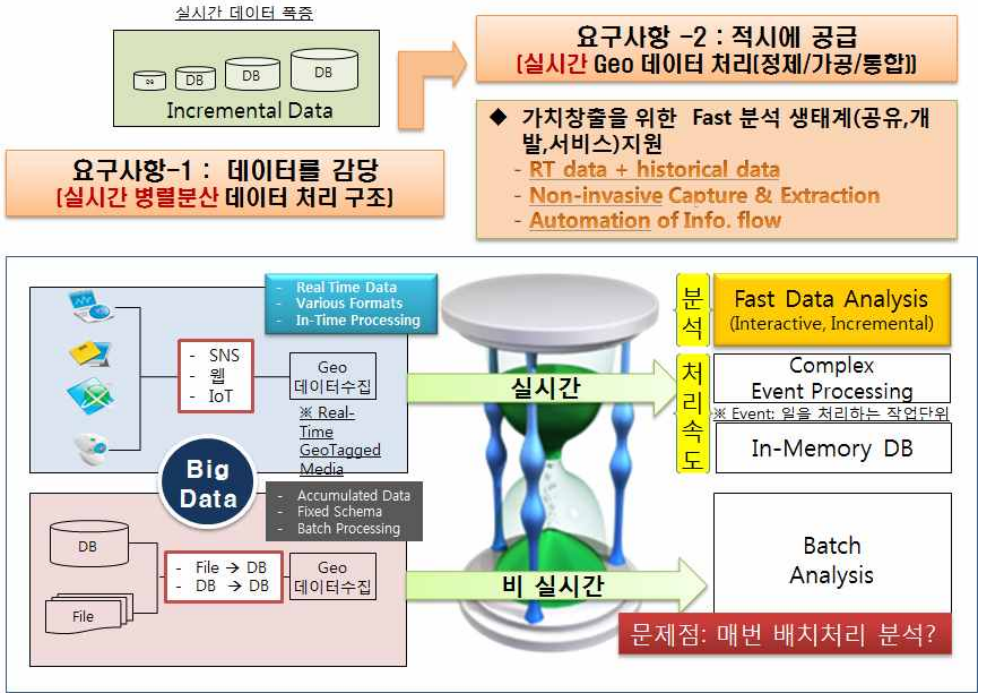
- 스트림 기반 분산 공간 ETL 도구는 JAR 형태의 라이브러리를 통해 핵심 기능을 제공함
- 공간 빅데이터 관리도구 서버는 웹 형식으로 전달되는 요구사항을 처리하는 서버로 결과값을 웹으로 전달하여 서비스함
- NiFi 사용자를 위해 스트림 기반 분산 공간 ETL 서버 내에 공간 연산 워크플로우 분석도구 서버를 구동하고 공간 ETL 핵심 기능을 라이브러리로 제공함으로써 그래프 기반의 워크플로우 생성/수정 서비스를 제공함

① [단위성과 D-1] 공간 빅데이터 웨어하우스 사용자/시스템/기술 요구사항 분석 (1차연도)

❍ 공간 빅데이터에 대한 실시간 병렬 분산 기술 기반 동적 데이터 처리 기술 및 요구사항을 분석하고 주요 기능을 제시함

❍ 분산/병렬 공간 빅데이터 처리 기술에 대한 연구 필요성 및 요구가 증가하고 있음

- 다중 사일로, 융합 DB 등 데이터웨어하우스의 확장과 관련한 구조가 필요함
- 빈번히 발생하는 트랜잭션 처리에 대한 최적화 요구로 in-memory, 캐시 서버 등과 같이 적시에 대응하기 위한 기술에 대한 요구가 증가하고 있음
- 소셜 네트워크, 사물인터넷의 등장으로 실시간 스트리밍 빅데이터에 대한 처리 필요성이 증가하고 있음
- 가용성과 적시성을 확보하기 위해 데이터웨어하우스 기술도 기존의 DW(Data Warehouse) 기술에서 한층 진화된 실시간 처리에 대한 요구 및 기존 ETL, 쿼리 등에서 해결하지 못한 데이터 통합과 관련한 다양한 요구사항이 존재함



<그림 2-67> 실시간 고속 ETL 기술의 필요성 및 요구사항

❍ ETL 기술 및 주요 기업 동향을 조사함

- 국내 ETL 시장은 데이터스트림즈, IBM, 오러클, 인포매티카 4개 사가 경쟁하고 있으며 최근에는 Hadoop 기반 솔루션으로 진화하고 있음
- 기존의 'ETL' 기능에서 'Integration'의 개념으로 확장되고 있음

❍ Hadoop 환경에서 사용되는 데이터웨어하우스 기술을 조사하고 분석함

- Hive는 하둡의 상위에 위치한 데이터 웨어하우스 구성 요소로써 SQL 쿼리를

맵리듀스하는 작업을 위해 컴파일하고 클러스터에서 실행되게 함

- Sqoop은 RDBMS에서 HDFS 파일시스템으로 데이터를 이동시키거나 HDFS에서 RDBMS로 데이터를 이동시키기 위한 하나의 유틸리티임
- Flume은 분산환경에서 대용량의 로그 데이터를 효과적으로 수집하여 합친 후 다른 곳으로 전송할 수 있는 Streaming Data Flow 기반 서비스를 제공함

○ 공간정보 빅데이터와 관련하여 현재 산업계에서 활용되거나 연구되고 있는 데이터 웨어하우징 관련 ETL 기술의 특징과 의미를 파악하여 공간정보 빅데이터 ETL 도구의 개발방향 및 고려사항을 정립함

○ 실시간 병렬 분산 기술로 고속 데이터 처리를 지원하기 위한 병렬분산 공간빅데이터 처리기술의 주요 기능을 제시함

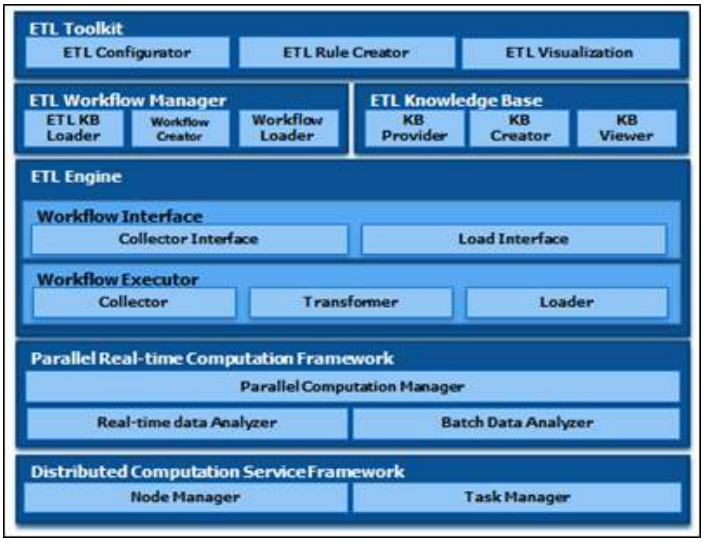


<그림 2-68> 공간 빅데이터 웨어하우징 주요 기능

② [단위성과 D-2] 공간 빅데이터 웨어하우스 사용자/서비스 /플랫폼 연동 및 ETL 처리 API 설계서 (1차연도)

○ ETL 시스템의 데이터 처리를 위한 전체적인 시스템 구성 및 기능과 주요 API를 정의하고 설계함

○ ETL 도구의 실시간 데이터 처리를 위한 기능 프레임워크는 분산/병렬 컴퓨팅 프레임워크 기반하에 ETL 툴킷과 워크플로우 매니저, 엔진으로 구성됨

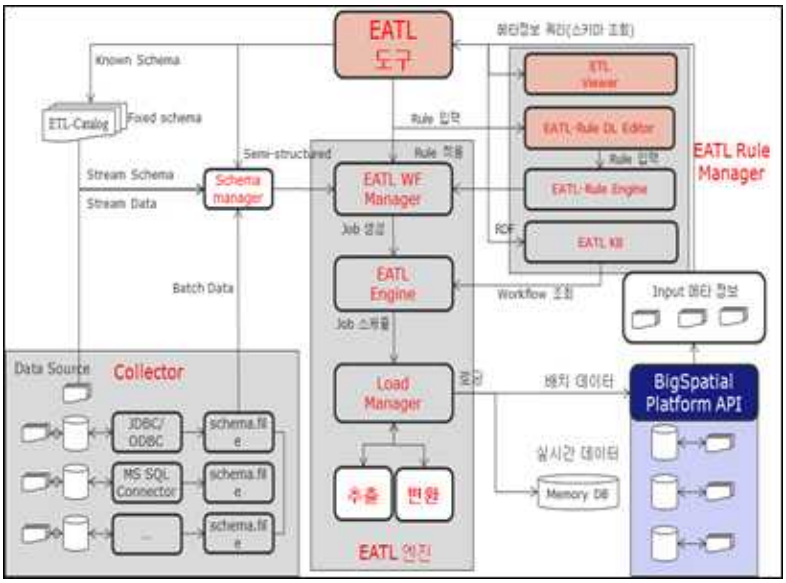


<그림 2-69> 워크플로우 기반 ETL 기능 프레임워크

❍ ETL 시스템의 구조는 다음과 같은 서브시스템으로 구성됨

- ETL 도구: ETL의 다양한 설정정보를 사용자에게 쉽게 사용할 수 있도록 제공하고, 실제적인 작업의 결과를 제공하기 위한 관리 툴
- ETL 컬렉터: Manager가 질의 처리, 자원 관리, 세션관리, 데이터 관리 및 플랫폼 관리 관련된 주요 기능을 수행함
- ETL 엔진: 각 데이터 노드에서 로딩된 정보를 이용하거나, 수집된 정보의 정제, 가공을 실질적으로 수행하는 ETL 작업 엔진
- ETL Rule 매니저: ETL 도구에 의해 설정되는 규칙을 관리하는 규칙 생성 및 설정엔진

❍ 워크플로우 기반 ETL 동작을 지원하기 위한 ETL 처리 플로우 및 인터페이스를 설계함(그림 2-69)



<그림 2-70> ETL 시스템 구조 및 처리 플로우

③ [단위성과 D-3] 워크플로우 생성을 위한 ETL Rule 설계 (1차연도)

○ 다양한 로직으로 이루어진 워크플로우에 기반하여 ETL 엔진이 분산 플랫폼에서 처리할 수 있는 룰(Rule)의 명세를 위해 요구되는 규칙 설정 메커니즘을 정의함

○ 입력되는 데이터의 포맷과 ETL 대상이 되는 ETL 명세언어 정보를 이용하여 센서데이터의 시맨틱 ETL을 위한 방법인 ETL 규칙을 정의하고 기술함

- 공간데이터 ETL의 ETL 규칙 포맷을 새롭게 정의하고, 이 ETL 규칙을 프로세싱하기 위한 메커니즘을 제공함
  - 시맨틱 ETL을 위한 프로세스의 형태를 정의하고 ETL 명세언어 어휘를 기준으로 시맨틱 어노테이션(Semantic Annotation)의 타입을 정의함
  - ETL 규칙은 JSON 포맷과 호환이 가능한 <키, 값> 구조를 가짐
  - ETL 규칙을 BNF(Backus-Naur Form) 포맷으로 정의하고 JSON 형태의 입력데이터에 대하여 ETL 규칙을 작성함(그림 2-70 예시)
  - ETL 규칙에서 입력 센서데이터의 키는 ETL 규칙의 키와 동일하며 입력데이터 각각의 키에 해당되는 ETL 규칙을 정의함
- ETL 규칙은 ETL 명세언어의 각 어휘(Vocabulary)와 입력 센서데이터의 <키, 값>을 이용하여, 수집•ETL•저장의 작업을 정의하고, 이를 실시간 분산 병렬 프레임워크에서 실행함

TRANSLATION RULE EXAMPLE

```
tdID= [
        {"jobtype":1, "subject":resource@tdID$",
         "object":"resource@transducerCategory$"},
        {"jobtype":2,"subject":resource@tdID$",
         "predicate":"resource@produces",
         "object":"resource@transducerCategory$"}
      ]
manufacturer= [
        {"jobtype":3,"subject":resource@tdID$",
         "predicate":"resource@manufacturer",
         "object":"manufacturer$", "datatype":"string"}
      ]
productNo= [
        {"jobtype":3,"subject":resource@tdID$",
         "predicate":"resource@productNo",
         "object":"productNo$", "datatype":"string"}
      ]
min= [
        {"jobtype":3,"subject":resource@tdID$",
         "predicate":"resource@min",
         "object":"min$", "datatype":"float"}
      ]
max= [
        {"jobtype":3,"subject":resource@tdID$",
         "predicate":"resource@max",
         "object":"max$", "datatype":"float"}
      ]
```

<그림 2-71> ETL 규칙(예시)

④ [단위성과 D-4] ETL 워크플로우 자동 생성 모듈 설계 (1차연도)

○ 분산 플랫폼에서 다양한 로직으로 처리할 수 있는 룰에 기반하여 ETL 기능을 순차적으로 처리할 수 있는 워크플로우의 생성 메커니즘을 정의하고 워크폴로우 생성모듈을 설계함

○ 각 노드에 분산된 분산노드의 ETL 엔진은 사용자가 수집하도록 명세한 데이터 또는 시스템에서 저장하고 있던 공간 데이터를 원하는 분석에 사용할 수 있도록 변환, 정제하기 위한 데이터 처리를 수행함

- 워크플로우는 효율적이고 명확한 로직에 기반하여 데이터를 처리할 수 있는 실행코드를 의미함
- 생성된 워크플로우는 분산노드에 분배되어지고 각 엔진들은 실제적인 데이터 처리를 위해 수령한 워크플로우를 실행하는 구조를 가짐

○ ETL 워크플로우 생성기

- ETL 워크플로우는 분산노드에서 실행할 하나의 실행가능한 코드를 만드는 과정이며 EATL 토폴로지, EATL 수집기, EATL 변환기로 구성됨
- ETL 토플로지는 시맨틱 ETL 타스크를 수집(Collect), 변환(Translate), 저장(Add)의 3 개의 과정으로 정의하고, 각 타스크를 병렬로 처리하는 실시간 분산 병렬 프레임워크인 아파치 스톰을 기반으로 처리하는 역할을 수행함
- ETL 수집기는 아파치 스톰의 스파우트를 상속하여 구현하며, JSON 포맷의 센서데이터를 개별 <키, 값>의 튜플로 나누어, 변환기에 전송함
- ETL 변환기는 아파치 스톰의 볼트를 상속하여 구현하며, 수집기로부터 전송받은 튜플과 ETL 블랙보드와 ETL 지식베이스를 이용하여 변환작업을 수행함

```
Algorithm 1 Collect Sensor Data
define class Collector extends Spout :
DataQueue Q
open() :
    JsonToken[] tokenList = parseJSON(input_sensor_data_stream)
    reportID = findReportID(tokenList)
    if reportID is not NULL then
        createDataMap(reportID , tokenList);
    end if

    for each JsonToken t in tokenList  do
        Q.add(t)
    end for


nextTuple() :
    while Q is not EMPTY
        JsonToken t = Q.poll()
        if t is not NULL then
            createDataMap(reportID , tokenList);
            createTuple(<reportID.t.getKey(), t.getValue()>);
        end if
    end while
```

<그림 2-72> ETL 수집 알고리즘

⑤ [단위성과 D-5] 워크플로우 기반 ETL 데이터 처리 엔진 설계 (1차연도)

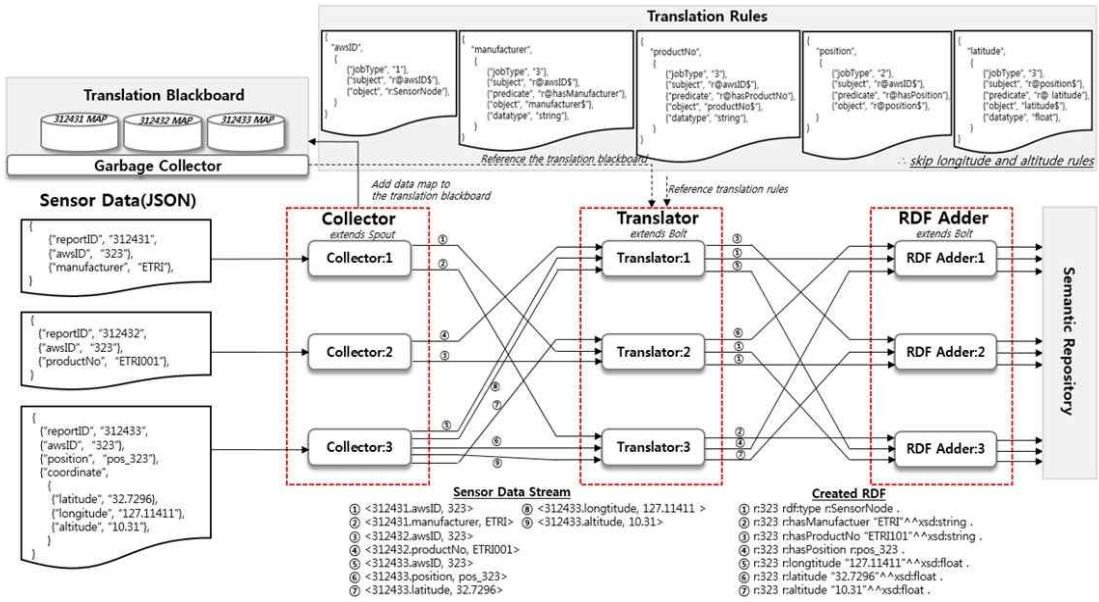
❍ ETL 워크플로우 생성기에서 생성된 실행가능 코드를 분산노드에서 병렬적으로 처리하는 ETL 데이터 처리 엔진의 기능들을 정의하고 기본적 설계를 수행함

❍ ETL 엔진

- ETL 엔진은 공간빅데이터 분산 플랫폼의 데이터 노드에서 동작하는 공간 데이터 처리 엔진임
- 공간 기능의 처리 및 공간 분석을 위해 요구되는 데이터의 정제, 가공, 분석 기능을 실제적으로 제공함
- ETL 워크플로우를 수신하여 실행하는 엔진으로써 워크플로우에 포함된 사용자의 설정에 기반한 실행코드를 수행함

❍ ETL 엔진을 통한 분산/병렬 처리 개념

- ETL 엔진은 맵리듀스를 이용한 병렬 분석으로 수행됨
  - 분석기는 분석규칙으로부터 생성된 Rete 네트워크를 클러스터링으로 분산된 색인기(패턴/조인) 로컬 메모리에 저장하고 트리플 형태의 대량의 입력 Fact에 대한 색인작업을 병렬로 수행함
  - 색인 작업은 Rete 네트워크의 패턴 네트워크를 이용하여 패턴매칭 테스트를 수행하는 패턴색인 작업과 조인네트워크를 이용하여 조인매칭 테스트를 수행하는 조인색인 작업으로 나눌 수 있음
- <그림 2-72>는 실시간 분산 병렬 ETL 예를 나타냄

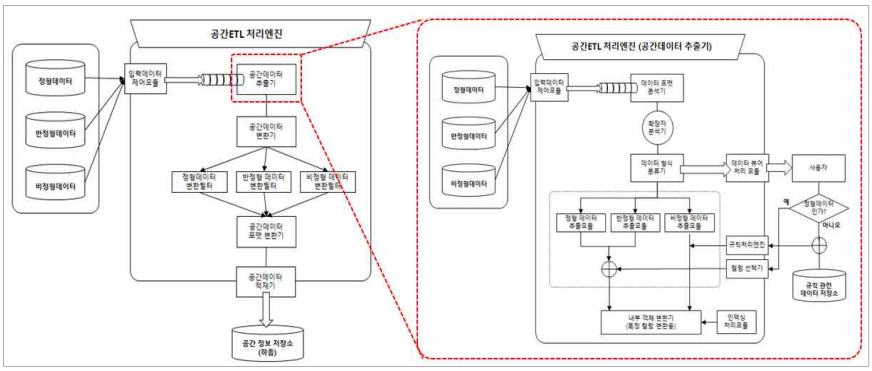


<그림 2-73> 실시간 분산 병렬 ETL 구조

⑥ [단위성과 D-6] 다중 공간 데이터 수집 모듈 개발 (2차연도)

❍ 다양한 공간정보 유형(정형, 반-정형, 비정형)에 매칭되는 공간 데이터 수집 모듈을 개발함으로 배치 기반의 공간정보들을 내부 공간포맷으로 변환하여 공간정보 활용성을 증가시키는 모듈임

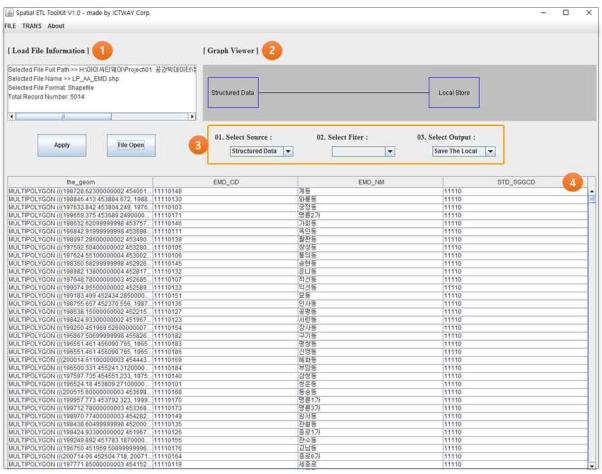
- 다양한 공공 데이터에서 공간 데이터 수집을 위한 ETL 핵심 기능 중 하나이고 여러 포맷으로 존재하는 공간 데이터 레코드들을 수집하기 위한 모듈임
- 공간 데이터 수집 모듈은 입력 데이터 제어 모듈, 데이터 포맷 분석기, 데이터 형식 분류기, 내부 객체 변환기, 규칙 처리 엔진으로 구성됨



<그림 2-74> 다중 공간 데이터 수집 모듈 설계도

❍ Spatial ETL ToolKit GUI는 Spatial ETL ToolKit 기본 라이브러리를 기반으로 구현되었으며 Spatial ETL ToolKit 기본 라이브러리는 1-1 세세부의 marmot(공간 빅데이터 저장관리시스템) 라이브러리를 이용하여 구현함

- “Load File Information”(①) : 로컬 저장소에서 로딩할 파일 선택 시 해당 파일의 절대경로와 선택된 파일 이름 표시; “확장자 분석기”를 통한 공간 데이터 유형 표시; 그리고 파일 내에 존재하는 레코드 수를 표시
- “Graph Viewer” (②) : ③에서 선택된 정보를 기반으로 화면에 정보를 표시



<그림 2-75> Spatial ETL ToolKit 기본 화면

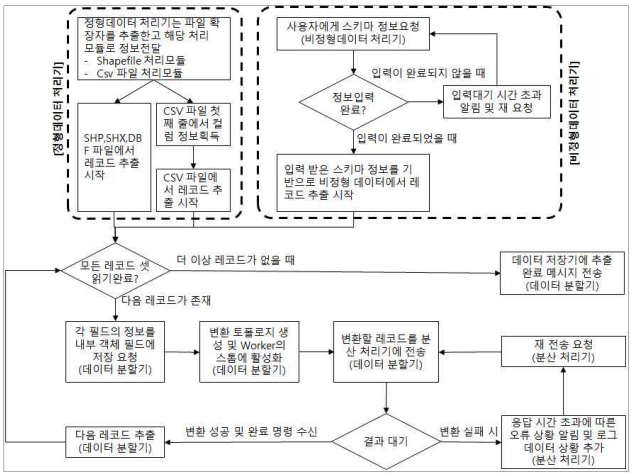
⑦ [단위성과 D-7] 공간 스트림 데이터 수집 모듈 개발 (단일 노드)(2차연도)

○ 공간정보 중 스트림 형태의 공간정보를 실시간 수집하기 위한 모듈로 공간 센서, 공간 스트림 데이터, 트위터, 페이스북 등과 같은 스트림 기반 공간정보 수집 기능을 제공함

○ 공간 스트림 데이터 수집 모듈 설계

- 공간 스트림 빅데이터 수집 모듈은 스트림 형태로 웹서버, 공간 센서 등과 연결 상태를 유지하여 수집하는 모듈이며 아파치 카프카와 스톰을 이용하여 스트림 이벤트 기반으로 동작함

- 스트림 형태의 데이터는 배치기반 데이터 처리 방식처럼 정형/반정형/비정형 데이터로 구분하지 않고 비정형 데이터로 구분하여 처리해야 함

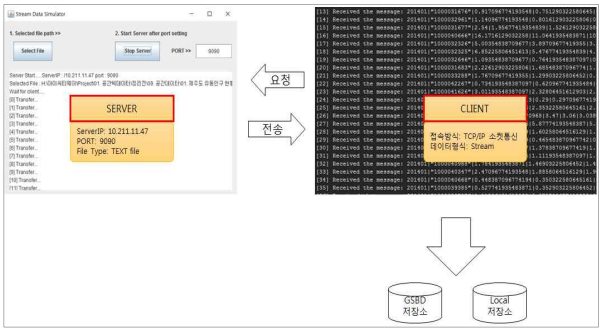


<그림 2-76> 공간 스트림 빅데이터 수집 및 변환 모듈 순서도

○ 공간 스트림 데이터 수집 모듈 구현

- 배치기반 다중 공간 데이터 적재 모듈과 내부 방식은 같으며 이를 처리하기 위해서는 공간 스트림 빅데이터 수집 모듈 개발 시 개발된 “Stream Data Simulator” 를 이용해야 함

- 적재 방식은 로컬 저장소를 이용하는 방법과 1-1 세세부의 분산 저장소를 이용하는 방법이 있으며 아파치 스톰을 이용한 적재 토폴로지 등록 및 실행 시 관리자 권한이 필요함



<그림 2-77> 공간 스트림 빅데이터 적재 모듈 연구개발 결과

⑧ [단위성과 D-8] 공간 스트림 데이터의 병렬/분산 처리 모듈 개발 (3차연도)

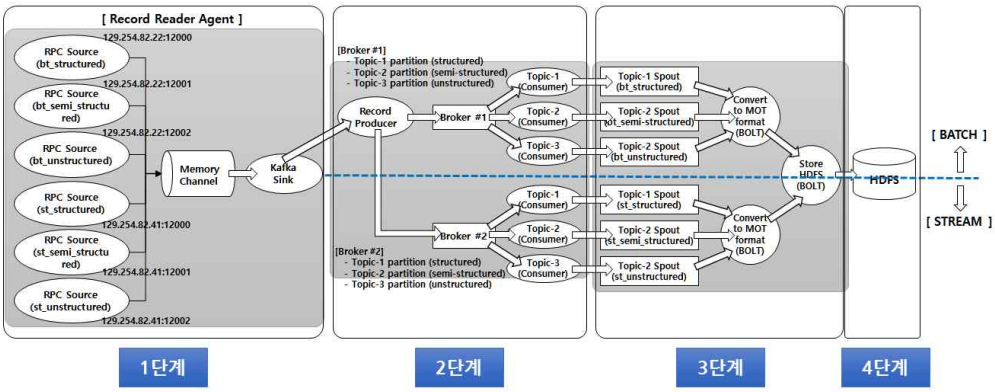
❍ 배치/스트림 기반 공간정보를 신속, 정확하게 ETL 처리하기 위한 모듈로 공간정보를 병렬 및 분산 처리하여 ETL 수행 성능을 향상시키기 위한 모듈임

❍ 공간 스트림 데이터의 병렬/분산 처리 모듈 설계 및 구현

- 배치 또는 스트림 기반 공간데이터 수집 처리기는 모든 공간 데이터 유형을 포함하여 처리가 이뤄져야 하며, 수집 방법에 대해서는 전송 데이터 포맷에 따라 다르게 설계되어야 하고 전송 시 활용되는 데이터 포맷은 일반적으로 많이 쓰이는 바이너리 또는 문자열 포맷으로 전송되어야 함
- 배치기반 공간데이터 전송 모듈은 정형데이터를 다루는 정형데이터 추출 모듈, 반-정형데이터를 다루는 반-정형데이터 추출 모듈, 비정형데이터를 다루는 비정형데이터 추출 모듈로 나뉨
- 각 추출 모듈은 병렬 및 분산 처리 모듈에 추출한 공간 레코드 혹은 공간 정보 전송을 위해 “RPC Clinet” 모듈을 사용함

❍ 공간 스트림 데이터의 병렬/분산 처리 모듈 기능과 성능은 외부기관에서 동작 속도 및 정확도를 검증하였음

- TC1: 배치기반 공간 정보 수집 및 적재 모듈 처리 속도 측정
- TC2: 스트림 기반 공간 정보 수집 및 적재 모듈 처리 속도 측정
- TC3: Shape 파일 수집 및 적재 모듈 정확도 측정
- TC4: CSV 파일 수집 및 적재 모듈 정확도 측정
- TC5: 아두이노 기반 센서 정보 수집 및 적재 모듈 정확도 측정
- TC6: 라즈베리파이 기반 센서 정보 수집 및 적재 모듈 정확도 측정



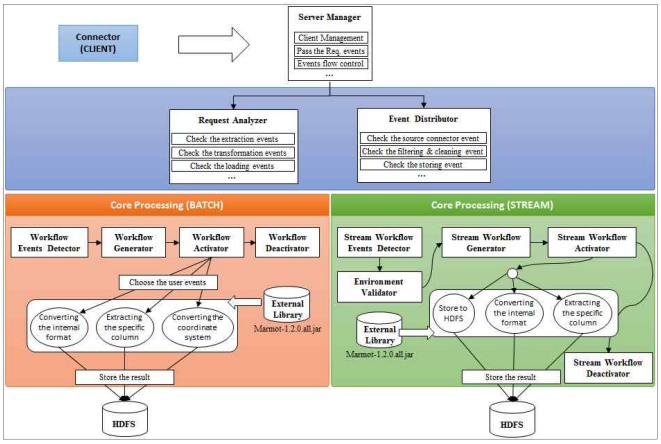
<그림 2-78> 병렬 및 분산처리 모듈 구현도

⑨ [단위성과 D-9] 스트림 기반 공간 ETL 처리 엔진 개발 (4차연도)

❍ 스트림 기반 공간 ETL 처리 엔진은 서버-클라이언트 모델로 병렬 및 분산 처리 방식으로 공간 ETL을 처리하며 배치/스트림 기반의 공간정보를 ETL 처리하는 엔진임

❍ 스트림 기반 공간 ETL 처리 엔진 (서버)

- 스트림 기반 공간정보 처리 시 스트림 성격상 공간정보가 연속적으로 발생하기 때문에 공간정보를 수집, 변환, 적재하기 위해서는 Apache Flume, Apache Kafka, Apache Storm 조합으로 공간정보를 처리하고 최종적으로 공간 레코드만 추출 또는 내부 포맷으로 변환된 공간정보를 파일 형태로 HDFS에 적재함
- 클라이언트에서는 배치 기반일 경우 파일의 확장자를 이용하여 자동으로 파일 포맷을 구분할 수 있으나 공간 ETL 연동 API를 원활하게 사용하기 위해 옵션 관리자를 두어 파일 포맷을 미리 입력받아 처리함



<그림 2-79> 스트림 기반 공간 ETL 처리 엔진

❍ 스트림 기반 공간 ETL 처리 엔진 (클라이언트)

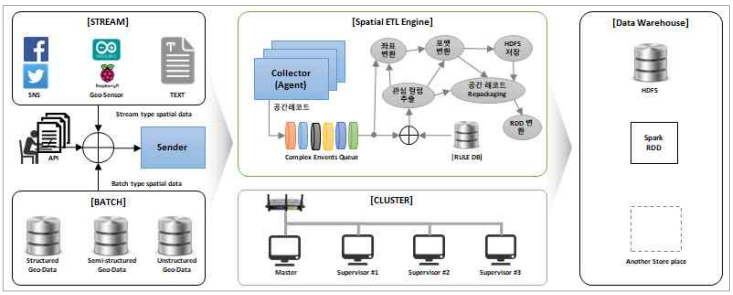
- 스트림 기반 공간 ETL 처리 엔진(클라이언트)는 JAVA로 구현되었으며 사용자(개발자)가 사용하기 편리하도록 API 형태의 라이브러리로 제공됨
- Source Connector는 배치 또는 스트림 기반의 다양한 공간 정보의 추출을 위해 지원하는 표준 포맷 수가 많아야 하므로 확장성 있는 구현 방식을 적용함
- 구현된 클라이언트의 Package 구조는 “SpatialETL.Client.Server”, “SpatialETL.Client.SourceConnector”, “SpatialETL.Client.Record”, “SpatialETL.Client.Table”, “SpatialETL.Client.Storing” 으로 구성됨
- SpatialETL.Client.Server” 는 서버 관리기를 구현한 것이며 “SpatialETL.Client.SourceConnector” 는 Source Connector를 구현한 것이고 “SpatialETL.Client.Record”, “SpatialETL.Client.Table” 은 레코드 및 테이블 생성기를 구현한 것임
- 또한, “SpatialETL.Client.Storing” 은 적재기를 구현한 Package임

⑩ [단위성과 D-16] 스트림 기반 공간 ETL 처리 엔진 연동 API (POC 수준) (4차연도)

❍ 스트림 기반 공간 ETL 처리 엔진을 Java 라이브러리 형태로 서비스하기 위해 JAR 형태로 추출, 변환, 적재 기능을 POC 수준으로 제공하는 API임

❍ 스트림 기반 공간 ETL 처리 엔진 연동 API

- 공간 ETL 처리 엔진 연동 API는 두 가지 모드로 설계 및 구현되었으며 첫 번째는 사용자(개발자)가 코드 상에서 참조할 수 있도록 라이브러리 형태로 제공되도록 설계 및 구현되었고, 두 번째는 간단한 공간 ETL 기능 테스트를 위해 Command Line Interface에서 동작 가능하도록 스크립트 형태로 제공되도록 설계 및 구현되었음



<그림 2-80> 스트림 기반 공간 ETL 처리 엔진 연동 API 흐름도

❍ 연동 API 추출 기능

- 배치 기반으로 존재하는 공간 정보를 추출하기 위해 접두사가 “LoadTo-”로 시작하는 API를 호출하면 되고, 스트림 기반의 공간 정보를 추출 및 생성하기 위해서는 접두사가 “ConnectTo-”로 시작하는 API를 호출하면 됨

- 스트림 기반 공간 ETL 처리 엔진 연동 API는 1-1세세부에서 개발한 라이브러리(marmot-1.2.0)에서 추출기능을 사용하며 추후 질의처리 엔진 및 저장소와의 원활한 연동을 위해 외부 라이브러리에서 내부 포맷 변환 API를 사용함

❍ 연동 API 변환 기능

- 공간정보 변환 시 요구되는 기본적인 정보 이외에도 스트림 형태로 유입되는 공간정보에 대한 공간 레코드 생성 및 공간 테이블 생성 기능을 지원하며 기존 공간 레코드 내에 특정 컬럼만 따로 새로운 공간 테이블에 쓰는 기능이 구현됨

- 변환 기능은 추출한 공간정보의 스키마 정보, 테이블 정보, 연동 센서 서버 정보, URL 정보를 생성 및 수정하는 API임

❍ 연동 API 적재 기능

- 적재 기능은 연결된 HDFS에 공간정보를 적재하는 기능이며 공간 테이블 적재경로, 파일 이름, 파일 확장자 정보는 사용자에게 파라미터로 전달받은 후 구동되는 시스템 내 사용자가 파라미터로 전달한 특정 위치에 적재하게 됨

⑪ [단위성과 D-10] 스트림 기반 공간 ETL 처리 엔진 연동 API 개발 (5차연도)

- ❍ 스트림 기반 공간 ETL 처리 엔진을 Java 라이브러리 형태로 서비스하기 위해 JAR 형태로 추출, 변환, 적재 기능을 배포 수준으로 제공하는 API임
- ❍ 스트림 기반 공간 ETL 처리 엔진 연동 API 중 수집 기능은 “SensorDataCollector”와 “SNSDataCollector”이고 사용횟수가 빈번한 SNS 데이터 수집 기능인 “TwitterCollector”는 별도로 제공되며 수집 API 접근은 “SpatialETL.Client.SourceConnector.*”로 가능하며 배치 형태 뿐만 아니라 스트림 형태의 공간정보 수집을 위해 API 서비스를 제공함
- ❍ 스트림 기반 공간 ETL 처리 엔진 연동 API 중 변환 기능은 기존 배치 형태의 공간정보 처리 시 워크플로우 생성 모듈을 통한 ETL 워크플로우 이벤트 생성 기능 API 추가됨
- ❍ ETL 워크플로우 이벤트는 추출/변환/적재 이벤트를 XML 형태로 로컬에 저장한 후 “DataFlowManager”를 통해 서버로 전송하여 처리하는 기능임
  - ETL 워크플로우 이벤트의 추출 규칙에는 Source 기본 정보가 포함되며 총 레코드 수, 컬럼 수, 스키마 보유 유/무, 배치 또는 스트림 모드 정보 등이 있음
  - 또한 변환 규칙에는 새롭게 생성할 공간정보 테이블에 몇 개의 컬럼 정보를 추출해서 별도로 생성할지 결정하는 부분과 생성할 공간정보 테이블의 좌표계 정보등이 표시되어 있음
  - ETL 워크플로우 이벤트 XML 파일은 “SpatialETL.Client.DataFlowManager”에 이벤트 전송 기능을 이용하여 서버로 전송되는데 서버는 전송받은 XML 파일을 내부 로컬 저장소에 같은 XML 타입으로 저장하고 이벤트 정보를 로드하여 ETL 워크플로우 작업을 수행하게 됨

	Function Name	Description	Used external library
Source Connector	SpatialETL.Client.ServerManager.Server(String, int)	Server IP, Port정보를 이용하여 서버 인스턴스 생성	Marmot-5.3.0.6.all.jar
	SpatialETL.Client.ServerManager.Server(initClientEvn)	initClientEnv Class를 이용하여 서버 인스턴스 생성	Marmot-5.3.0.6.all.jar
	SpatialETL.Client.ServerManager.connect()	설정된 서버 정보를 토대로 접속 요청	Marmot-5.3.0.6.all.jar
	SpatialETL.Client.ServerManager.execute(EventsManager, RecordSet)	설정된 이벤트 워크플로우를 이용하여 서버에 Raw 공간 정보를 전송하고 규칙 정보를 전달하는 함수	Marmot-5.3.0.6.all.jar
	SpatialETL.Client.ServerManager.isConnected()	서버 접속 유무를 알 수 있는 플래그 값	Marmot-5.3.0.6.all.jar
	SpatialETL.Client.ServerManager.sendMessage(String)	연결된 서버에 직접적으로 문자열값을 전송하는 함수	Marmot-5.3.0.6.all.jar
	SpatialETL.Client.ServerManager.close()	접속된 서버를 닫는 함수	Marmot-5.3.0.6.all.jar
	SpatialETL.Client.SourceConnector.URICollector	RESTful interface, geoJSON, WKT, 등과같이 스트림 형태로 제공되는 URI에서 공간 정보를 수집하는 함수	Marmot-5.3.0.6.all.jar
	SpatialETL.Client.DataFlowManager.sendRule(String)	사용자 정의 규칙을 토대로 만들어진 내부 워크플로우 규칙 값을 서버로 전송하는 함수	Marmot-5.3.0.6.all.jar
	SpatialETL.Client.DataFlowManager.sendEventCommand(String)	연결된 서버에 이벤트 명령어를 보내는 함수	Marmot-5.3.0.6.all.jar
	SpatialETL.Client.EventsManager	사용자 정의 규칙 정보를 토대로 이벤트 워크플로우를 생성하는 함수	Marmot-5.3.0.6.all.jar
	Function Name	**Description**	**Used external library**
Filtering & Cleaning	SpatialETL.Client.Filter.extractColumn()	모든 Column 정보를 추출하여 새로운 레코드 생성	Marmot-5.3.0.6.all.jar
	SpatialETL.Client.Filter.extractColumn(String)	하나의 Column 만 추출하여 새로운 레코드 생성	Marmot-5.3.0.6.all.jar
	SpatialETL.Client.Filter.extractColumn(int, String[])	지정된 Column 정보를 이용하여 새로운 레코드 생성	Marmot-5.3.0.6.all.jar
	SpatialETL.Client.Filter.extractJSONColumn(String, int, String[])	JSON 정보 내 지정된 Column 정보를 토대로 새로운 레코드 생성	Marmot-5.3.0.6.all.jar
	SpatialETL.Client.Filter.changeCoordinate()	기본으로 설정된 EPSG:5186으로 좌표계를 변환하는 함수	Marmot-5.3.0.6.all.jar
	SpatialETL.Client.Filter.changeCoordinate(String)	설정된 좌표계 값을 토대로 입력 레코드의 공간 좌표값을 변환하는 함수	Marmot-5.3.0.6.all.jar
	Function Name	**Description**	**Used external library**
Storer	SpatialETL.StoreToHDFS()	현재 연결된 HDFS에 결과물을 저장하는 함수	Marmot-5.3.0.6.all.jar
	SpatialETL.StoreToLocal()	Local 저장소에 결과물을 저장하는 함수	Marmot-5.3.0.6.all.jar

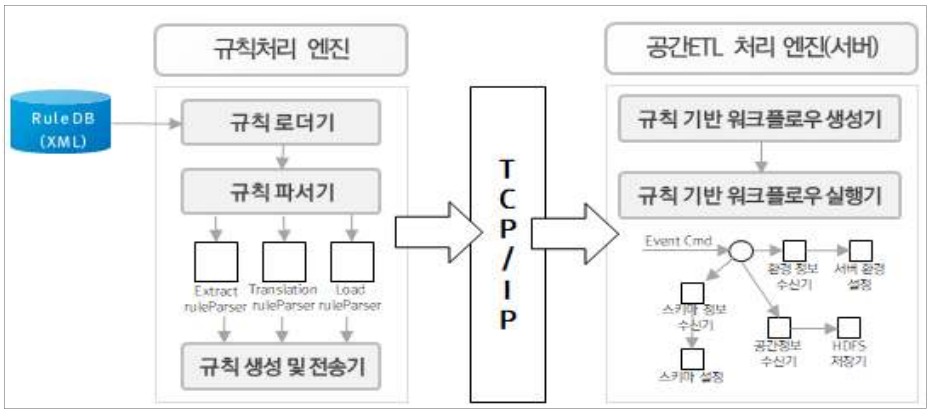
<그림 2-81> 제공 API 리스트 Lite 버전

⑫ [단위성과 D-11] 스트림 기반 공간 ETL 규칙 처리 엔진 개발(POC 수준) (5차연도)

❍ POC 수준의 스트림 기반 공간 ETL 규칙 처리 엔진은 독립적으로 구동이 가능하며 사용자가 작성한 Drools 규칙을 파싱하여 XML 형태의 규칙 정보로 변환 서비스를 제공함

❍ 공간 ETL 규칙 처리 엔진(POC)

- 내부 워크플로우는 추출, 변환, 적재가 순서대로 이뤄지며 크게 배치 형태의 공간정보 처리와 스트림 형태의 공간정보 처리로 나뉘게 됨
- 배치 형태의 공간정보 처리는 내부 워크플로우 규칙에 따라 우선 Source 파일 절대경로 값을 읽어오고 관심 컬럼 정보, 좌표계 설정 값, 내부 워크플로우 번호 순서로 구동됨
- 스트림 형태의 공간정보 처리 또한 내부 워크플로우 규칙에 따라 우선 서비스되는 스트림 타입 정보, 수집주기 정보, 좌표계 설정 값, 내부 워크플로우 번호 순서로 구동됨
- 배치 형태의 공간 정보를 처리하기 위해서는 Shapefile과 같은 정형화된 공간정보 이외 CSV, JSON 파일처럼 응용 분야가 많으면서 반-정형화된 공간 정보는 추가적으로 스키마 정보를 사용자에게 입력받아 처리함
- 또한, 스트림 형태의 공간정보를 처리하기 위해서는 스트리밍되는 공간 정보의 타입 정보와 수집주기 정보, 관심 컬럼 정보 및 스키마 정보를 사용자에게 입력받아 새로운 공간레코드 테이블로 만들어 제공함
- 배치 형태의 공간정보 처리는 Shapefile, CSV file, JSON file에 대한 ETL 워크플로우가 제공되며 스트림 형태의 공간정보 처리는 GeoJSON, WKT, WKB, Twitter에 대한 ETL 워크플로우가 제공됨
- 규칙에 대한 정의는 Drools 규칙을 사용하며 Shapefile, CSV file, JSON file에 공간 정보 추출 규칙은 POC 버전과 배포용 버전이 동일하게 테스트되어짐



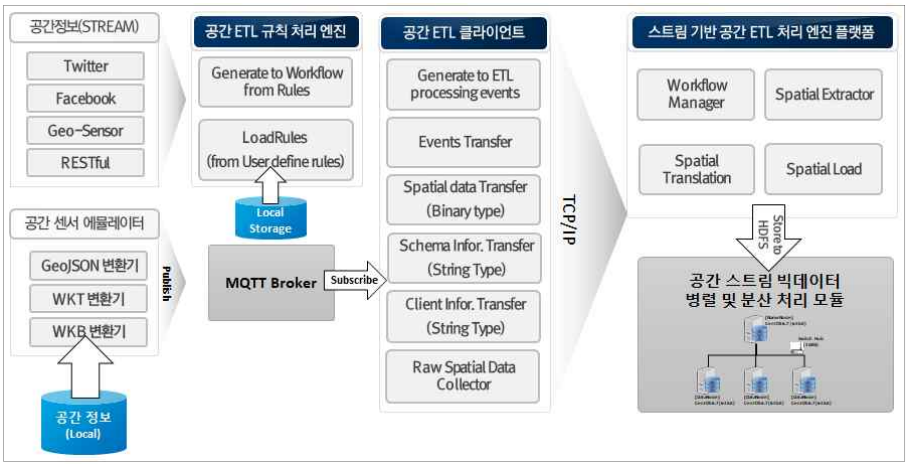
<그림 2-82> 스트림 기반 공간 ETL 규칙 처리 엔진 구성도

⑬ [단위성과 D-12] 공간 센서 스트림 데이터 수집 및 분산 처리 모듈 개발 (5차연도)

❍ 공간 센서 스트림 데이터 수집 모듈은 다양한 스트림 형태의 공간정보를 실시간으로 ETL 처리하기 위해서는 제공되는 공간 스트림 타입정보, 수집주기, 추출하고자 하는 컬럼정보 등을 사용자에게 제공받고 공간 스트림 정보를 서비스하는 MQTT Broker의 IP주소와 Port 정보 그리고 Publish되는 Topic 이름을 제공받아 수집 및 분산 처리하여 HDFS에 적재하는 모듈임

❍ 공간 센서 스트림 데이터 수집 및 분산 처리 모듈 설계

- 구현된 모듈은 공간 센서 에뮬레이터, 공간 스트림 수집 모듈, 규칙 기반 공간 스트림 수집 모듈이며 공간 센서 에뮬레이터는 Console에서 실행하고 규칙 기반 공간 스트림 수집 모듈의 규칙은 Drools를 활용하였음
- “SpatialETL.Client” 수집모듈 실행에 앞서 항상 “SpatialETL.Server”가 구동중이어야 하며 “SpatialETL.Client” 환경설정 파일인 XML 파일에 Server IP, Port 정보가 기재되어 있어야 하고 marmot 서버도 구동되어 있어야 함
- 공간 센서 에뮬레이터는 Shapefile을 입력받아 스트림 형태로 변환하며 Source 파일(*.shp) 경로와 변환할 스트림 모드 정보를 필수적으로 입력받음
- 다양한 공간 센서들의 스트리밍 속도를 고려하여 ms 단위로 전송 속도 조절이 가능하고 정확한 메시지 전달을 위해 Console 화면 출력 기능을 제공함
- “SpatialETL.Client.SourceConnector.SensorDataCollector”는 GeoJSON, WKT, WKB 수집 모듈에서 필요한 정보를 API 형태로 사용자에게 입력받음
- “SensorDataCollector” 클래스 생성 시 수집 주기, 스트림 타입, MQTT Broker IP, Port 정보, ROI 컬럼 정보등을 매개변수로 주어 인스턴스를 생성해야 함
- Sensor Data Collector 샘플 코드는 MQTT broker 연결 정보와 HDFS 연결정보를 바탕으로 수집되는 공간정보를 연결된 HDFS에 적재하는 모듈임

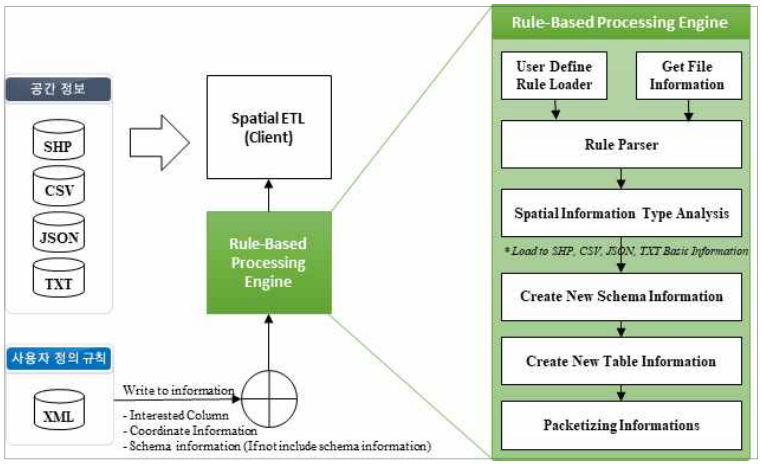


<그림 2-83> 공간 센서 스트림 데이터 수집 및 분산처리 모듈 설계안

⑭ [단위성과 D-17] 스트림 기반 공간 ETL 규칙 처리 엔진 개발 (5차연도)

○ 스트림 기반 공간 ETL 규칙 처리 엔진은 기존 공간 ETL 처리 엔진(클라이언트)에서 선택된 공간정보에 대한 정보와 설정한 관심 컬럼 명, 새로운 레코드 정보를 토대로 ETL 규칙을 정의하고 이를 서버로 전송하여 워크플로우 동작 이벤트(공간데이터 파일 정보, 스키마 존재 유/무, 새로운 스키마 정보)를 생성하는 엔진임

- 스트림 기반 공간 ETL 규칙 처리 엔진 POC 버전은 독립적인 프로그램으로 실행이 가능하며 스트림 기반 공간 ETL 처리 엔진에 직접적인 구동이 불가능함
- 공간 ETL 규칙 처리 엔진 POC 버전과 공간 ETL 처리 엔진을 연동하기 위해서는 먼저 공간 ETL 규칙 처리 엔진에서 사용자 정의 또는 미리 정의된 규칙 정보를 기반으로 ETL 워크플로우 정보를 로컬 저장소에 저장한 후 해당 파일정보를 이용하여 공간 ETL 처리 엔진에 적용할 수 있음
- 스트림 기반 공간 ETL 규칙 처리 엔진 2차 배포용은 기존 공간 ETL 처리 엔진에 Built-In되어 API 호출 형식으로 운용이 가능한 라이브러리임
- 패키지 타입의 JAR 파일을 제공하여 내부적으로 스트림 규칙이 필요한 경우나 스트림 규칙을 실행 시 사용할 수 있도록 함
- 스트림 기반 공간 ETL 규칙 처리 엔진 2차 배포용을 사용하는 연동 API 기능을 추가하고 공간 스트림 수집 모듈 연동 API 기능 또한 추가함
- JSON file을 제외한 Shapefile과 CSV file은 두 가지 규칙이 존재하는데 하나의 파일만 선택하여 ETL 작업을 수행하는 경우와 동일한 여러 파일을 ETL 작업 수행 요청을 하는 경우임
- Shapefile인 경우 파일명 혹은 폴더명만을 지정하면 “shp_single” 혹은 “shp_multiple” 규칙이 생성되어 내부 ETL 워크플로우를 선택하고 이벤트를 생성하여 서버에 관련 정보를 전송함

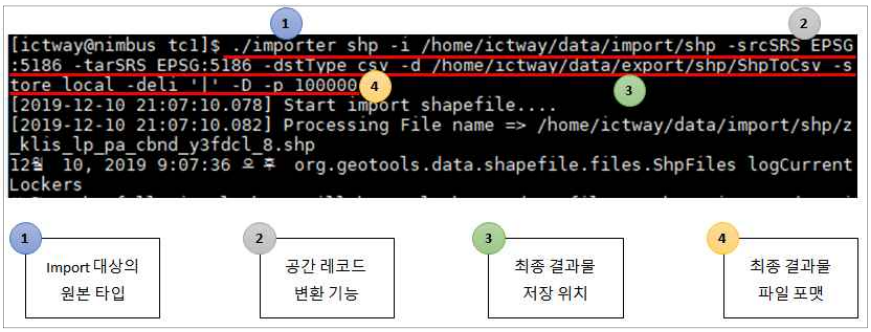


<그림 2-84> 스트림 기반 공간 ETL 규칙처리 엔진 구성도

⑮ [단위성과 D-13] 스트림 기반 공간 ETL 처리엔진 고도화 및 연구시제품 개발 (6차연도)

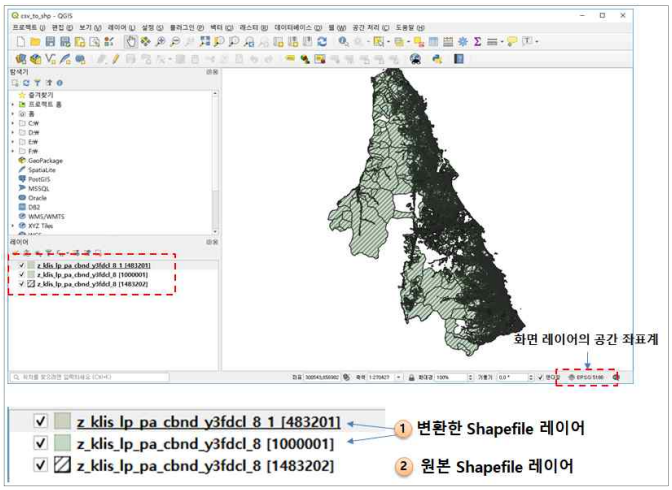
❍ 스트림 기반 공간 ETL 처리 엔진은 로컬 저장소에 있는 다양한 공간 정보를 추출하고, 공간 분석이 용이하게 변환하여 분석 플랫폼에 적재하는 엔진으로 ESRI 사의 Shapefile, CSV file, GeoJson file에서 공간 정보를 추출하는 기능, 공간 포맷 변환 및 좌표계를 변환하는 기능, 그리고 로컬, HDFS, Marmot 저장소에 적재하는 기능을 제공함

- 공간 ETL 처리 엔진의 사용자 접근성을 높이기 위해 독립적 구동이 가능한 Java의 JAR 라이브러리 형태로 제공됨
- 또한, 사용 편의성 및 접근성을 용이하기 위해 도움말 기능이 내포되어 있으며 직관적 표현 방식으로 공간 ETL 명령어가 제공됨



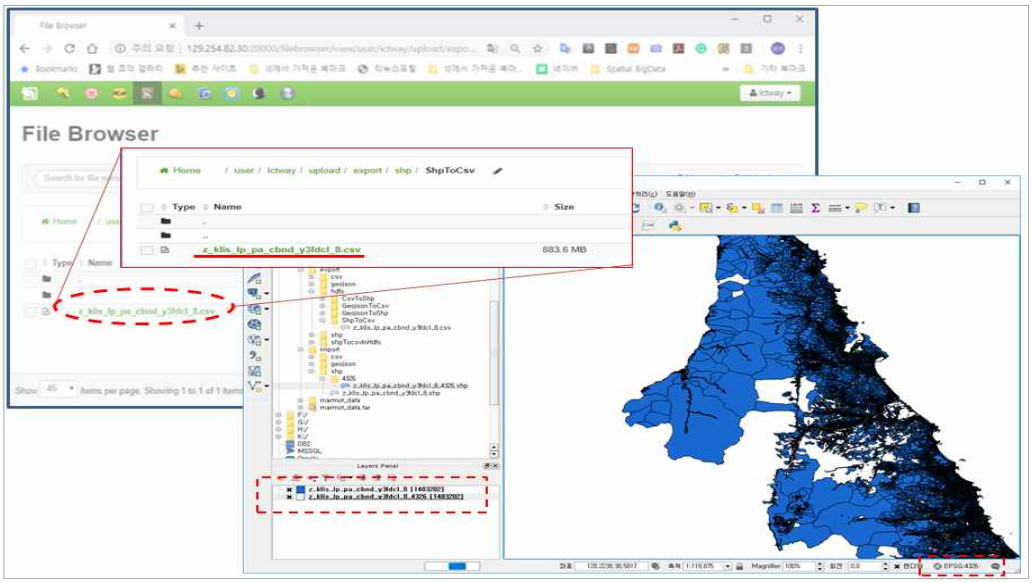
<그림 2-85> 스트림 기반 공간 ETL 처리 프로그램에 직관적 표현 방식

- 추출 기능에는 ESRI사의 Shapefile 내 공간 정보 추출 기능, CSV file 내 공간 정보 추출 기능, GeoJson file 내 공간 정보 추출 기능이 제공됨
- 변환 기능에는 좌표계 변환 기능, 새로운 공간 레코드 생성 기능, 공간 레코드의 컬럼 재배열 기능, ESRI 사의 Shapefile 변환 기능, CSV 타입으로 변환 기능, GeoJson 타입으로 변환 기능이 제공됨



<그림 2-86> CSV to Shapefile 변환 (QGIS)

- 적재 기능에는 로컬 저장소에 적재 기능, HDFS에 적재 기능, Marmot에 적재 기능이 제공됨



<그림 2-87> Shapefile 내 공간 좌표계를 EPSG:4326으로 변환하고 CSV file 포맷으로 HDFS에 저장한 결과

- 부가 기능에는 공간 정보 형태가 CSV와 GeoJson 타입으로 구성된 파일인 경우 공간 레코드 수를 세어주는 카운팅 기능이 제공됨

```
[ictway@nimbus tcl]$ ./importer count -i /home/ictway/data/export/hdfs/shp/ShpToGeojson/z_klis_lp_pa_cbnd_y3fdcl_8.json -t ge
ojson -m
[2019-12-10 02:52:35.321] Start Record Counting....
[2019-12-10 02:53:00.123] File name: /home/ictway/data/export/hdfs/shp/ShpToGeojson/z_klis_lp_pa_cbnd_y3fdcl_8.json
[2019-12-10 02:53:00.123] Reading total count: 1483202
[ictway@nimbus tcl]$
```

<그림 2-88> 부가 기능: 레코드의 수를 출력하는 기능

- 또한, 변환되어 로컬 저장소에 저장된 파일의 용량이 큰 경우 편집기로 그 내용을 확인하기 어려운 경우가 존재하는데 일부 내용만을 화면에 출력해 주는 라인 뷰어 기능이 제공됨
- 외부 기관(TTA) 검증을 통해 동작 정확도 100%를 확인함

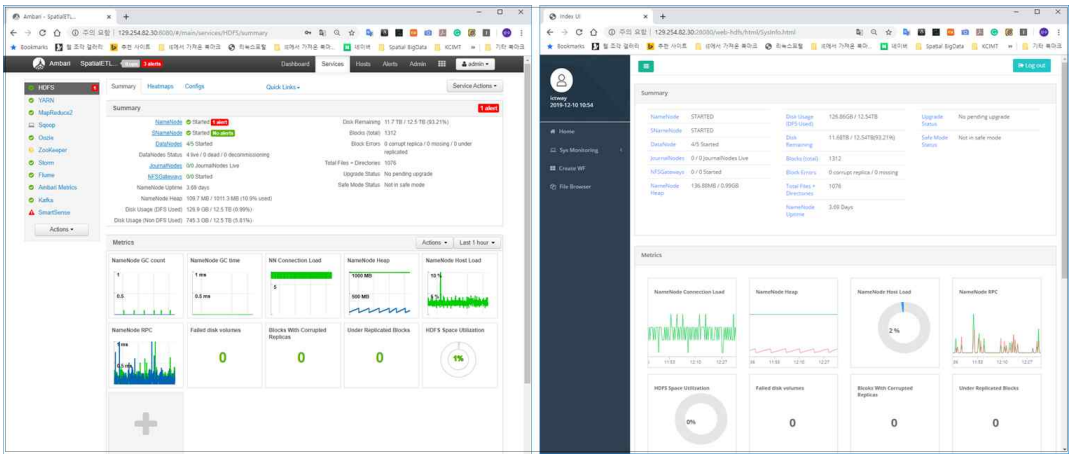
```
[ictway@nimbus tcl]$ ./lineViewer -i /home/ictway/data/export/hdfs/shp/ShpToGeojson/z_klis_lp_pa_cbnd_y3fdcl_8.json -
cnt 10
{"SRS":"EPSG:4326", feature:[{"type":"Feature","geometry":{"type":"MultiPolygon","coordinates":[[[[128.4354,38.4566],
[128.4354,38.4566],[128.4356,38.4566],[128.4357,38.4566],[128.4358,38.4566],[128.4359,38.4566],[128.4359,38.4566],[12
8.436,38.4565],[128.436,38.4565],[128.4361,38.4564],[128.4361,38.4565],[128.4361,38.4564],[128.436,38.4563],[128.4359
,38.4563],[128.4359,38.4562],[128.4359,38.4563],[128.4358,38.4563],[128.4357,38.4563],[128.4355,38.4563],[128.4354,38
.4563],[128.4354,38.4562],[128.4353,38.4562],[128.4353,38.4562],[128.4353,38.4563],[128.4354,38.4563],[128.4354,38.45
65],[128.4354,38.4566]]]]},"properties":{"big_sq":"2383348","big_fx":1.0,"pnu":"4282025322104080001","jibun":"408-1
","bchk":"1","sgg_oid":213594,"col_adm_se":"42820"},"id":"z_klis_lp_pa_cbnd_y3fdcl_8.1"}]}
{"SRS":"EPSG:4326", feature:[{"type":"Feature","geometry":{"type":"MultiPolygon","coordinates":[[[[128.4369,38.4566],
[128.437,38.4566],[128.437,38.4566],[128.4372,38.4565],[128.4372,38.4565],[128.4372,38.4565],[128.4373,38.4564],[128.
4373,38.4562],[128.4373,38.4562],[128.4372,38.4561],[128.4372,38.456],[128.4372,38.4559],[128.4373,38.4558],[128.4373
,38.4558],[128.4373,38.4557],[128.4373,38.4555],[128.4374,38.4553],[128.4374,38.4552],[128.4374,38.455],[128.4374,38.
4549],[128.4375,38.4549],[128.4375,38.4548],[128.4374,38.4549],[128.4374,38.455],[128.4374,38.4552],[128.4373,38.4553
],[128.4373,38.4555],[128.4373,38.4557],[128.4372,38.4558],[128.4372,38.4558],[128.4372,38.4559],[128.4372,38.456],[1
28.4372,38.456],[128.4372,38.4561],[128.4373,38.4561],[128.4373,38.4562],[128.4373,38.4563],[128.4372,38.4564],[128.4
372,38.4564],[128.4372,38.4565],[128.4372,38.4565],[128.4372,38.4565],[128.4372,38.4565],[128.4369,38.4566]]]]},"prop
erties":{"big_sq":"2383349","big_fx":1.0,"pnu":"4282025322104140001","jibun":"414-1","bchk":"1","sgg_oid":213595,"c
ol_adm_se":"42820"},"id":"z_klis_lp_pa_cbnd_y3fdcl_8.2"}]}
{"SRS":"EPSG:4326", feature:[{"type":"Feature","geometry":{"type":"MultiPolygon","coordinates":[[[[128.4368,38.4794],
[128.4369,38.4794],[128.4368,38.479],[128.4368,38.4794]]]]},"properties":{"big_sq":"2383350","big_fx":1.0,"pnu":"4282
025322105160004","jibun":"516-4","bchk":"1","sgg_oid":216028,"col_adm_se":"42820"},"id":"z_klis_lp_pa_cbnd_y3fdcl_8
.3"}]}
```

<그림 2-89> 부가 기능: 일부 저장된 공간 레코드 정보를 출력하는 기능

⑯ [단위성과 D-14] 공간 스트림 ETL 관리 도구 개발 (6차연도)

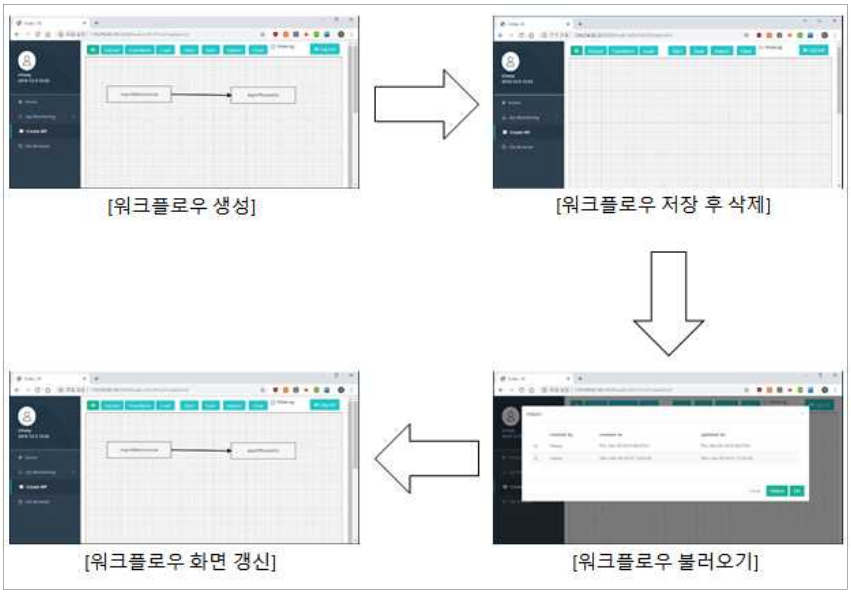
○ 공간 스트림 ETL 관리 도구는 웹 인터페이스를 통해 스트림 기반 공간 ETL 처리 엔진의 사용 편의성을 제공하며 공간 ETL 워크플로우 관리 기능, 기 설치된 HDP(Hortonworks Data Platform) 자원 사용량 모니터링 기능, HDFS File Browsing 기능을 제공함

- 웹 인터페이스 방식으로 스트림 기반 공간 ETL 처리 엔진을 효율적으로 관리 또는 운영하며 HDFS와 Storm의 시스템 자원 사용량 모니터링 기능, 그래프 기반 공간 ETL 워크플로우 생성 및 구동 기능, HDFS 파일 관리 기능을 제공함



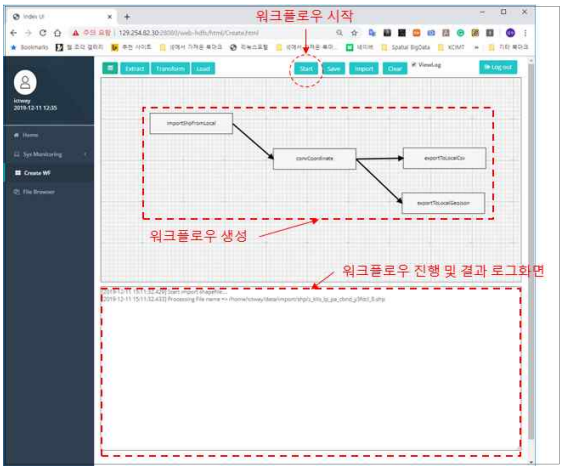
<그림 2-90> (좌)Ambari Metrics 메인화면, (우)공간 스트림 ETL 관리도구 메인화면

- 자체 DB를 통해 로그인 기능, 공간 ETL 워크플로우 저장 및 불러오기 기능이 제공되며 HDFS 파일 관리 기능을 통한 결과 확인이 실시간 가능하며 로컬로 다운로드 기능이 제공됨



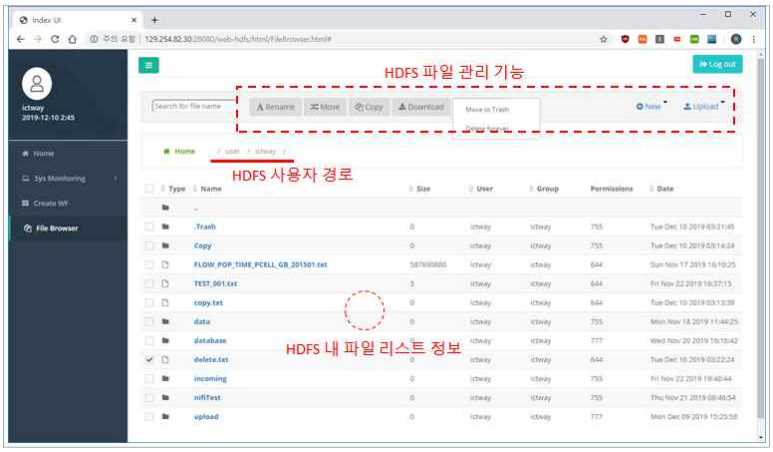
<그림 2-91> 공간 ETL 워크플로우 관리 주기

- 공간 ETL 워크플로우 관리 기능에서는 추출 API 리스트, 변환 API 리스트, 적재 API 리스트가 드래그앤드롭 방식으로 서비스됨
- 추출 API 리스트에는 로컬에 존재하는 ESRI사의 Shapefile에서 공간 정보를 추출하는 기능, 로컬에 존재하는 CSV file에서 공간 정보를 추출하는 기능, 로컬에 존재하는 GeoJson file에서 공간 정보를 추출하는 기능이 서비스됨
- 변환 API 리스트에는 좌표계를 변환하는 기능, 공간 레코드 내 컬럼을 변환하는 기능이 서비스됨
- 적재 API 리스트에는 ESRI 사의 Shapefile 포맷으로 변환하여 로컬, HDFS에 적재하는 기능, CSV file 포맷으로 변환하여 로컬, HDFS에 적재하는 기능, GeoJson file 포맷으로 변환하여 로컬, HDFS에 적재하는 기능, Marmot에 결과물을 적재하는 기능이 서비스됨



<그림 2-92> 공간 ETL 워크플로우 관리 화면

- HDFS 파일 관리 기능에는 파일 또는 디렉토리의 Rename 기능, Move 기능, Copy 기능, Delete 기능, New 기능 등이 서비스됨
- 외부 기관(TTA) 검증을 통해 동작 정확도 100%를 확인하고 사용자 만족도 설문조사를 통해 사용자 만족도 80점 이상을 달성함

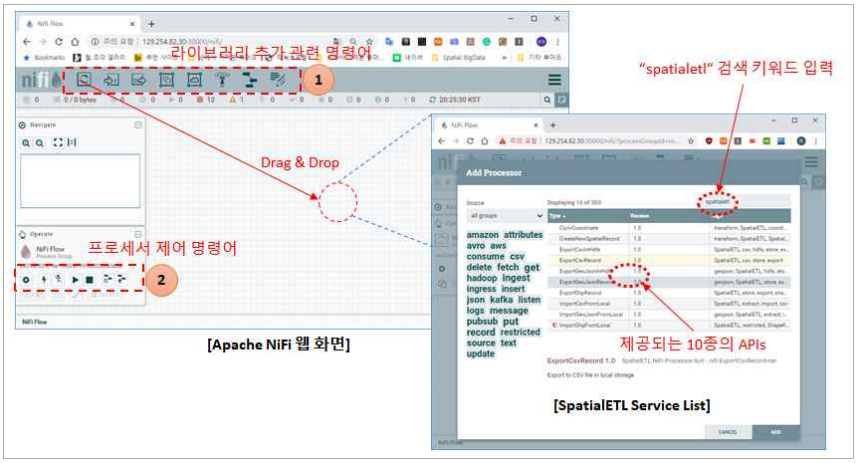


<그림 2-93> 공간 ETL 파일 Browsing 화면

⑰ [단위성과 D-18] 공간 연산 워크플로우 분석도구 개발 (6차연도)

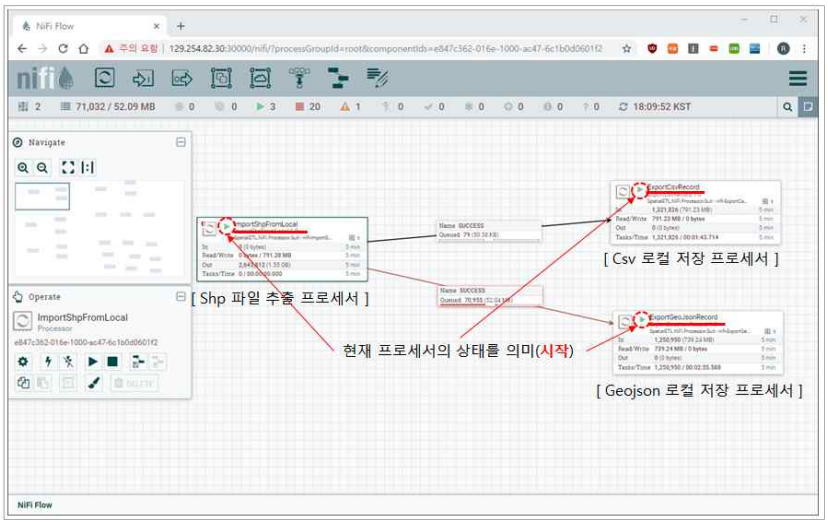
○ 공간 연산 워크플로우 분석 도구는 Apache NiFi Platform에서 동작 가능한 사용자 정의 프로세서들로 총 10종이 서비스되며 제공 형태는 라이브러리 형태인 NAR 파일로 제공함

- 10종의 사용자 정의 프로세서들은 기존 NiFi에서는 제공되지 않는 프로세서들로 공간 ETL하기 위한 기능들임
- Apache NiFi Platform 환경 구성 시 클러스터 환경 모드로 설치가 이뤄져야 하며 각 노드에 구동 프로세서를 지정하여 구동해야 함



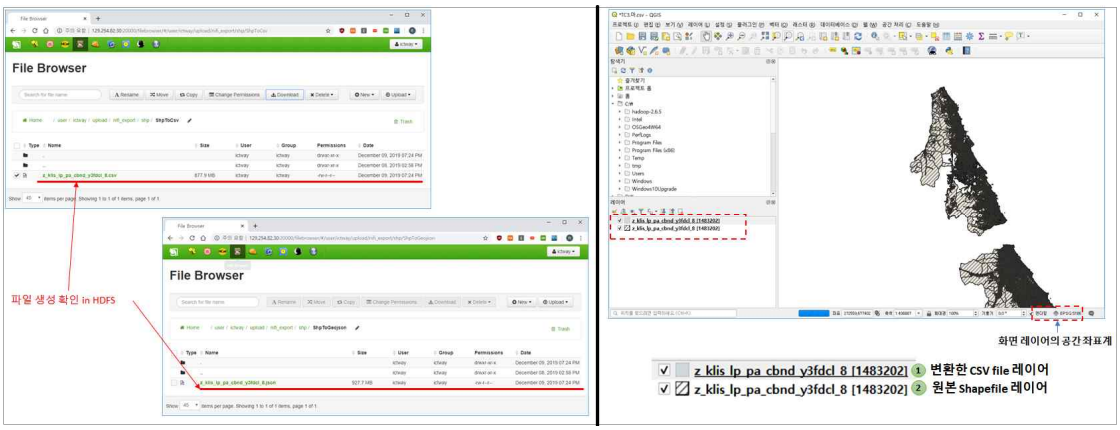
<그림 2-94> Apache NiFi 메인화면 및 10종의 API 라이브러리 목록

- 또한 제공되는 NAR 파일을 설치된 Apache NiFi의 lib 경로에 복사한 후 NiFi를 재시작하여 서비스 동작 여부를 확인해야 함
- Apache NiFi는 lib 디렉토리에 존재하는 서비스 라이브러리들을 자동 압축 해제하여 서비스 프로세서로 자동 변환함

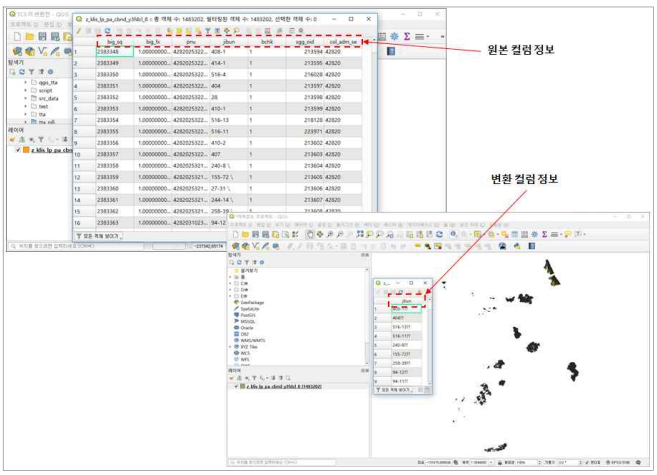


<그림 2-95> Shapefile을 CSV, GeoJson 포맷으로 변환 후 로컬에 저장

- 공간 ETL 워크플로우 생성은 제공되는 공간 ETL 프로세서들을 활용하여 공간 ETL 워크플로우를 구성해야 하고 각 프로세서들의 속성 값을 설정해야 하며 연결선에 대한 관계를 설정한 뒤 구동해야 함
- 추출 서비스에는 로컬에 존재하는 ESRI사의 Shapefile에서 공간 정보를 추출하여 전송하는 프로세서, 로컬에 존재하는 CSV file에서 공간 정보를 추출하여 전송하는 프로세서, GeoJson file에서 공간 정보를 추출하여 전송하는 프로세서가 서비스됨
- 변환 서비스에는 원본 좌표계를 변환 좌표계로 변환하여 재전송하는 프로세서, 공간 레코드 컬럼을 재배열하여 재전송하는 프로세서가 서비스됨
- 적재 서비스에는 로컬에 ESRI 사의 Shapefile, CSV file, GeoJson file 형태로 적재하는 프로세서, HDFS에 CSV file, GeoJson file 형태로 적재하는 프로세서가 서비스됨
- 외부 기관(TTA) 검증을 통해 동작 정확도 100%를 확인함



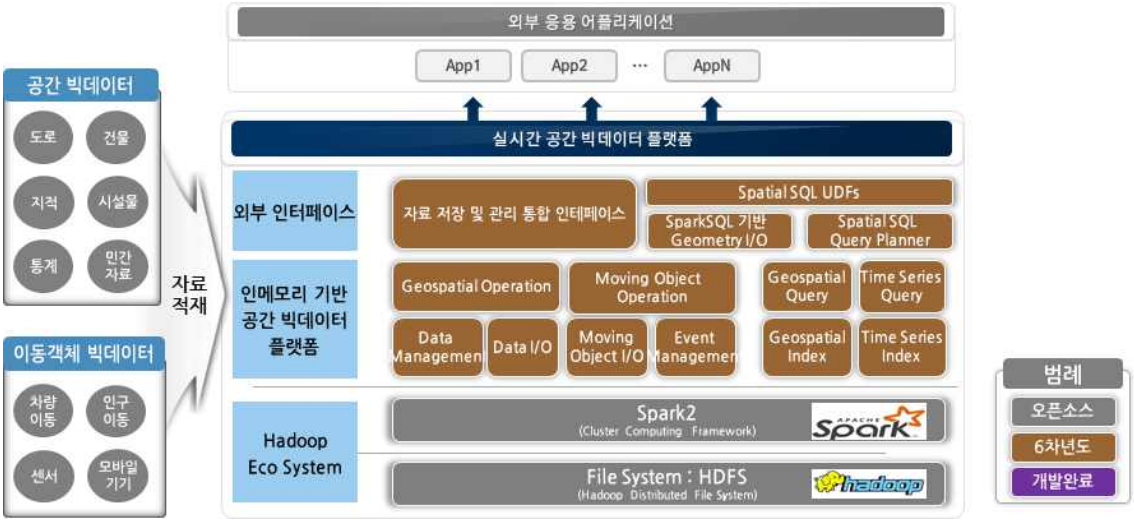
<그림 2-96> (좌)HUE를 이용한 결과 파일 생성 확인, (우) QGIS를 이용한 결과 파일 가시화



<그림 2-97> QGIS를 이용한 컬럼 정보 변환 확인

(라) 인메모리 기반 공간빅데이터 저장관리 프레임워크(핵심성과 O)

❍ Hadoop 및 Spark 기반에서 분산 메모리를 활용하여 공간 빅데이터 및 이동객체를 저장하고 분석하기 위한 엔진과 응용서비스를 개발하기 위한 API 및 도구의 집합



<그림 2-98> 인메모리 기반 공간빅데이터 저장관리 프레임워크 개념구조(핵심성과 O)

❍ 특징

- 공간 빅데이터 및 이동객체를 인메모리 기반 오픈소스인 Apach Spark을 활용하여 분산 메모리 기반 병렬처리를 통해 고속의 자료 저장/가공/분석 기능들을 제공하는 플랫폼임
- Hadoop에 저장된 공간 빅데이터를 인메모리 기반 기술을 이용하여 고속의 공간색인 생성 및 조회 기능 및 분석 기능을 제공함
- 인메모리 기반 기본 공간 분석 라이브러리 제공을 통해 기본적인 공간 분석 모델 설계 및 구현을 할 수 있게 해줌
- SparkSQL 기반의 SpatialSQL을 지원을 통해 공간분석 기능과 SQL을 이용한 공간분석 연계 분석을 용이하게 해줌

❍ 시스템 구조

- 공간 빅데이터는 HDFS에 적재된 이후 플랫폼의 분석 기능을 활용하여 분석시 분산 메모리 시스템을 이용하여 고속 분석이 실행됨
- Spark에서 일반적으로 사용되는 Scala 및 Python 기반 Script 기반의 분석과 동일한 방식으로 플랫폼의 분석 기능을 호출하여 사용이 가능함
- Client에서 공간분석 수행을 요청하면, Hadoop cluster에 저장된 공간자료를 분산 메모리기반에서 분석을 수행하고, 분석 결과는 분산메모리상에 존재하며, 필요시 분석 결과를 다시 Hadoop Cluster에 저장이 가능함

① [단위성과 O-1] 실시간 공간 빅데이터 저장 플랫폼 설계서 (1차연도)

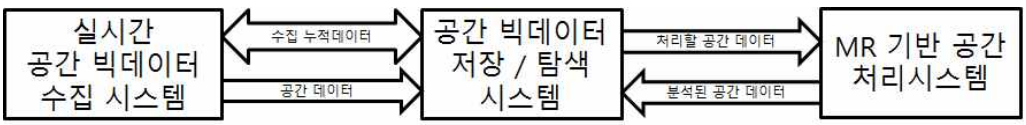
○ 인메모리 기반 공간빅데이터 저장 및 관리 플랫폼의 요구사항 정의를 통해 플랫폼의 기능 및 개념설계를 수행함

- 인메모리 기반 플랫폼 설계시 분산메모리의 특성상 각 노드의 메모리 용량에 대한 최적화와 노드간 데이터 교환에 의한 네트웍 병목현상에 대한 고려가 필요하며, 메모리에 저장된 자료를 HDFS에 저장 및 관리하기 위한 인터페이스 설계가 필요함
- 실시간성으로 수집되는 데이터의 경우 수집된 자료의 전처리 및 가공이 어려우므로 특정공간단위의 고정공간키를 정의하여 수집되는 자료의 종류 및 형태가 바뀌더라도 수용할 수 있는 비정형 자료구조를 기반으로 설계가 필요함
- 고정 공간키 기반의 자료 수집 방안 예시

```
고정 공간 키 - 표준화된 공간영역을 지정하고 고정 키 값을 부여하여 관리

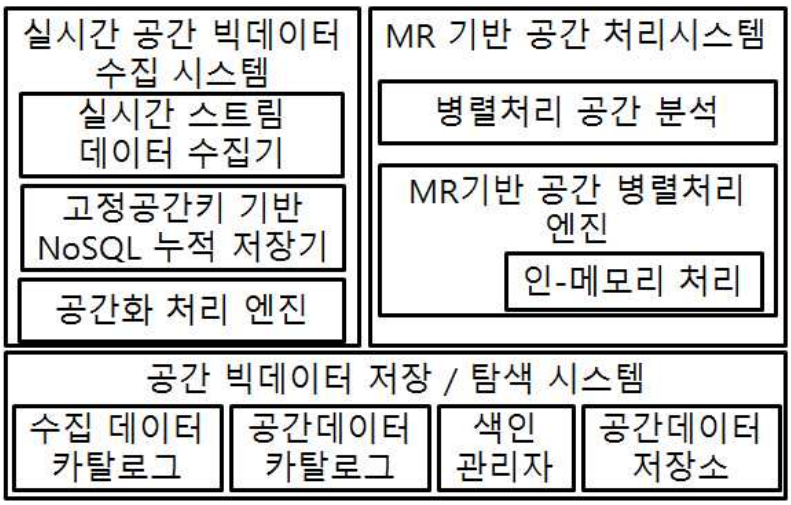
수집된 데이터의 공간의 표현은 고정키를 기준으로 관리
고정키 기반으로 수집된 데이터 -
{ 공간키=고정키1, 대중교통=지하철, 10층이상의고층건물=3, 용도=상업지구, Date=20141231 }
{ 공간키=고정키2, 용도=농업지구, 농수로인접=3, Date=20141231 }
{ 공간키=고정키1, 대중교통=지하철;버스, Date=20150101 }
```

- 실시간 공간빅데이터 처리시스템 안의 데이터 흐름은 <그림 2-98>과 같음



<그림 2-99> 실시간 공간빅데이터 처리시스템안의 데이터 흐름

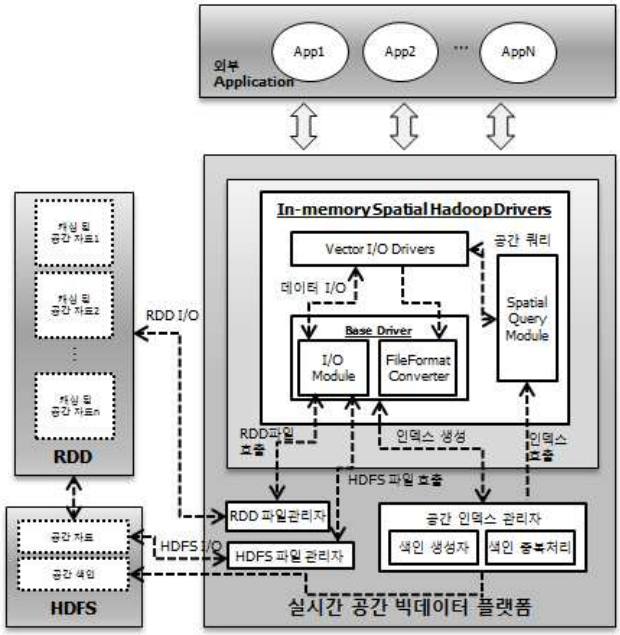
- 실시간 공간빅데이터 플랫폼의 주요 기능에 기반하여 플랫폼을 구성하는 각 시스템의 중요 컴포넌트 정의를 통해 플랫폼의 개념 설계안을 작성함(그림 2-99)



<그림 2-100> 실시간 공간 빅데이터 처리시스템 개념 구성도

② **[단위성과 O-2] 실시간 공간 빅데이터 저장 연구시제품 (2차연도)**

○ 인메모리 환경에서 공간 빅데이터의 저장 및 처리 기능 제공을 위한 핵심기술과 이를 구현한 실시간 공간 빅데이터 저장관리 시스템 개발을 목표로 시스템의 구조 및 인터페이스를 설계하고 공간색인 모듈과 연구시제품을 개발함



<그림 2-101> 실시간 공간 빅데이터 플랫폼 아키텍쳐

○ 실시간 공간 빅데이터 저장 연구시제품 개발

- 실시간 공간 빅데이터 플랫폼으로 공간빅데이터를 적재하고 공간 메타를 관리하는 HDFS 파일 관리자를 개발함
- HDFS에 저장된 공간 빅데이터의 인메모리 기반 파일 시스템인 RDD에 적재하고 관리하는 RDD파일 관리자를 개발함
- 적재된 공간 데이터의 색인을 생성 및 관리하고, 색인을 통한 공간 자료 조회를 지원하는 공간 인덱스 관리자를 개발함

○ 공간 빅데이터 기반 공간색인 생성 및 조회모듈 개발

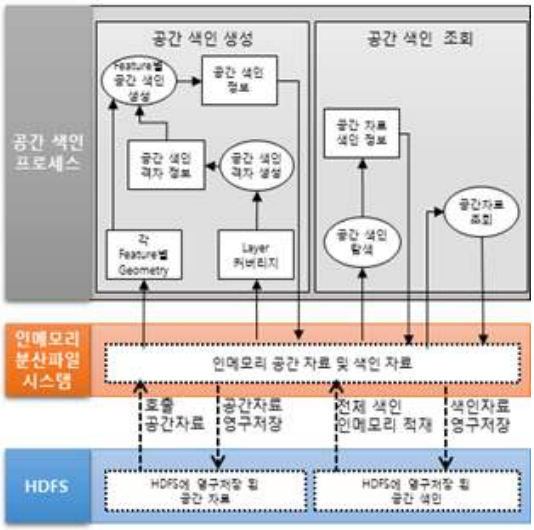
- 공간 빅데이터 시스템에 분산 저장된 공간 빅데이터 수준의 대용량 데이터에 대하여 1초 미만 또는 수 시간 이내의 공간 검색 및 조회 기능을 제공할 수 있도록 개발함
- 공간 빅데이터의 분산 저장 후 공간 인덱스 격자를 바탕으로 Feature별로 색인을 생성 후 격자 단위로 저장함
- 공간 영역 질의를 통한 공간 데이터 호출 및 호출 결과에 대한 중복자료 제거를 수행함

③ [단위성과 O-3] 인메모리 기반 공간색인 모듈 (3차연도)

○ 2차년도에 개발한 공간 색인 모듈을 인메모리 기반 설계 및 구현을 통해 고속 공간 색인 생성 모듈 및 인메모리 분산 처리를 이용한 공간 색인 조회 모듈을 구현함

○ 인메모리 기반 색인 생성 및 조회 모듈 개발

- 공간 색인 생성 및 조회 프로세스



<그림 2-102> 인메모리 기반 공간색인 모듈

- 공간색인 자료 생성 대상 자료는 초기 RDD상의 메모리에 적재된 후, 색인 생성 프로세스를 통해 그리드 격자 체계에 의해 색인을 생성하고, 생성된 색인 정보는 HDFS상에 저장하여 관리함

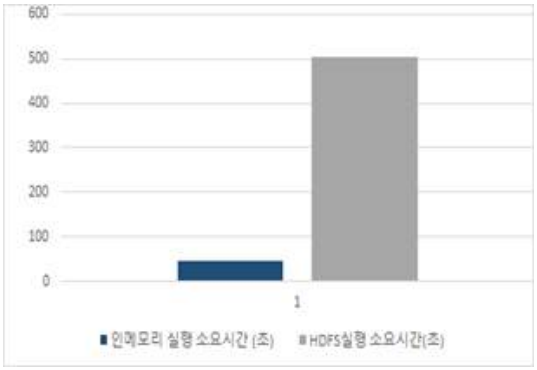
- HDFS에 저장된 공간 색인 자료를 RDD상의 메모리에 적재 후, 검색영역을 기준으로 각 색인 데이터 검색 및 중복자료 제거 수행을 통해 HDFS에 저장된 공간자료에 대한 색인정보를 추출함

○ 색인 생성 및 조회 모듈 성능 비교 실험

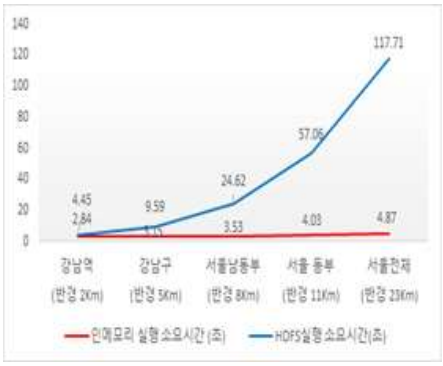
- 평가 환경은 8대의 데이터 노드, 1대의 네임 노드, 그리고 1대의 Yarn Resource Manager로 구성된 하둡 시스템을 구축하여 실시하였으며, 시험자료는 전국 연속지적도(약3600만건)를 활용하였음

- HDFS 기반의 색인 모듈 대비 인메모리기반 색인 모듈의 비교 테스트결과 색인 생성 시간은 1/10수준으로 향상되었고, 색인 검색 시간은 HDFS 기반의 경우 검색영역이 커질수록 검색속도가 저하되는 반면 인메모리 기반의 경우 검색속도가 거의 차이가 나지 않음을 확인함

- (결과 의미) 성능 비교 실험을 통해 인메모리 기반의 색인 모듈이 HDFS 기반의 색인 모듈보다 색인생성 시간을 단축시키고, 색인 검색시 작은 영역의 경우는 큰 성능차이가 없으나 큰 영역에 대한 대용량 검색시에는 속도 향상을 기대할 수 있음



<그림 2-103> 색인 생성 성능 비교



<그림 2-104> 색인 조회 성능 비교

** 성능 테스트용 시험 자료 : 전국 연속지적도 (약 3800만건)

④ [단위성과 O-4] 인메모리 데이터 처리를 위한 기본 공간 연산자 (3차연도)

❍ 인메모리 기반 데이터 처리 기본 공간 연산자 7종을 개발하였으며, 객관적인 평가를 위하여 TTA를 통해 성능 및 정확도 검증을 수행하였음

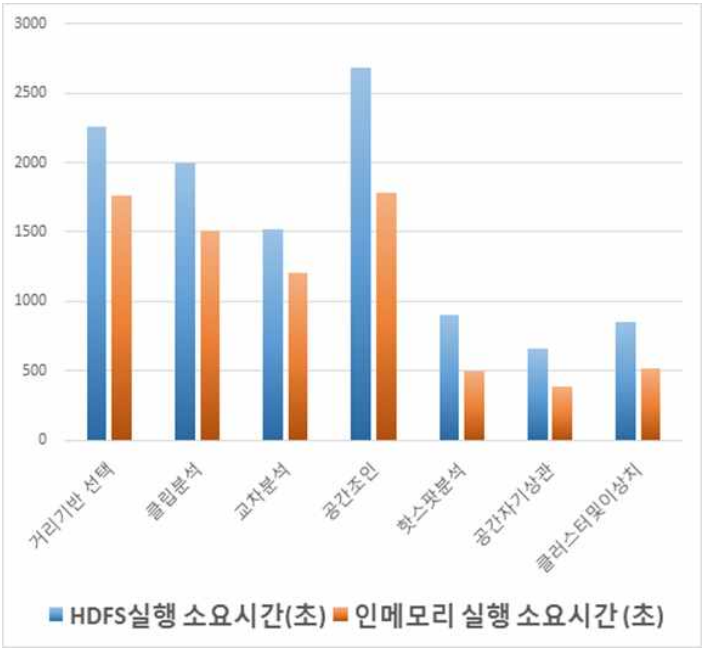
<표 2-18> 인메모리 데이터 처리용 기본 공간 연산자

공간 연산자 명	활용 공간 연산자	연산 특징
거리기반 선택	Intersects, Buffer	레이어간 연산, 공간색인 사용
클립 분석	Intersects, Intersection	레이어간 연산, 공간색인 사용
교차 중첩	Intersects	레이어간 연산 공간색인 사용
공간조인 중첩	Intersects, Within	레이어간 연산, 공간색인 사용
핫스팟 분석	Intersects	공간통계 분석, 공간색인 사용
공간 자기상관	Intersects	공간통계 분석, 공간색인 사용
클러스터 및 이상치	Intersects	공간통계 분석, 공간색인 사용

❍ 공간 색인 및 연산자 성능 평가

- (분석 기능 선정) 공간 색인 생성 모듈 및 연산자 7종에 대하여 2014년도 국토부 공간 빅데이터 사업단에 납품된 공간 빅데이터 엔진과 기본 공간 분석 라이브러리를 동일한 환경에서 평가를 진행함
- 평가 환경은 8대의 데이터 노드, 1대의 네임 노드, 그리고 1대의 Yarn Resource Manager로 구성된 하둡 시스템을 구축하여 실시함
- 성능 시험은 동일한 연산에 대하여 10회 반복 수행후 평균값에 대하여 비교함
- 속도향상율(%) = (HDFS기반 - 인메모리기반) / HDFS기반 * 100
- 인메모리 공간색인 생성은 공간 하둡 대비 91.18% 성능 향상이 있었으며, 공간색인 조회 모듈은 반경 5km 이상의 큰 영역에 의한 공간 쿼리시에도 속도저하가 발생하지 않음
- (결과 의미) 성능 실험에 사용한 공간분석 라이브러리 7종은 공간 분석 또는 공간자료 정제/융합시에 매우 자주 사용되는 기능들이며, 공간분석의 특성상

단일 연산이 아닌 복합연산으로 이루어지기 때문에 인메모리 기반 공간분석 라이브러리의 활용을 통해 공간자료 분석 및 가공시간을 단축할 수 있음

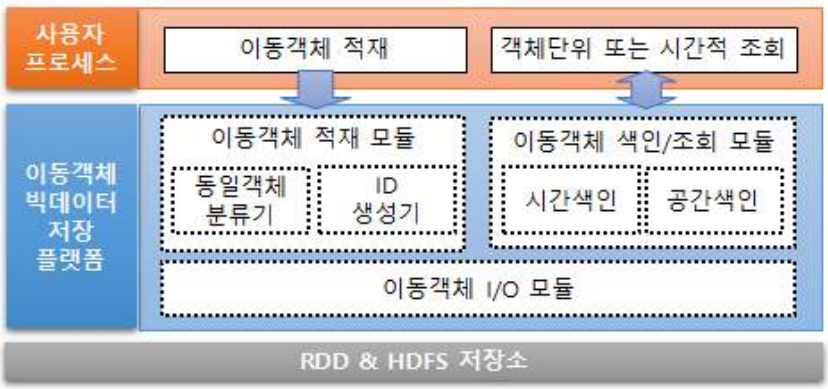


<그림 2-105> 공간연산자 성능 비교(10회 평균)

⑤ [단위성과 O-5] 인메모리 기반 이동객체 빅데이터 저장 플랫폼 설계서 (4차연도)

○ 인메모리 기반 공간 빅데이터 저장 플랫폼 개념 설계를 수행하였음

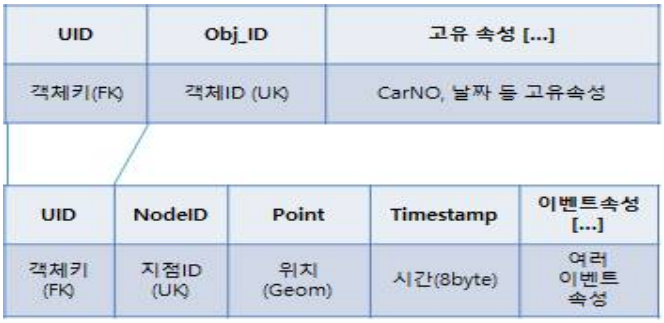
- 이동객체 공간 빅데이터 저장플랫폼의 기능정의 및 개념설계를 위한 국내·외 이동객체 기반 활용 사례 및 연구 사례 조사 및 분석 수행
- 이동객체 저장 모듈은 이동객체가 입력되면 적재 모듈을 통해 동일 객체인지 분류하고, 적재된 이동 지점 데이터마다 고유 ID를 부여함
- 이동객체 색인 및 조회 모듈은 지속적 또는 고속으로 적재되는 이동객체 데이터를 다루게 되므로 색인의 생성 및 관리는 프로세스 오버헤드를 최소화하는 방향으로 설계가 필요함



<그림 2-106> 이동객체 빅데이터 저장 플랫폼 개념설계

❍ 인메모리 기반 공간 빅데이터 저장 구조 설계

- 이동 객체 정보는 이동객체의 고유 정보를 저장하는 비공간 테이블과 각 객체의 이동 지점 정보를 저장하는 테이블로 구성됨
- 이동 지점 정보 테이블은 각 이동객체의 고유 정보를 참조하는 객체키(UID)와 이동 지점의 고유정보 값을 나타내는 지점ID(Node_ID), 해당 지점의 위치 정보를 저장하는 Point 컬럼, 해당 지점이 적재된 시간을 나타내는 Timestamp와 이동객체의 지점 적재시점의 이벤트 정보 컬럼으로 구성됨



<그림 2-107> 이동객체 저장 구조 설계안

⑥ [단위성과 O-11] 인메모리 기반 이동객체 빅데이터 저장 (POC 수준) (4차연도)

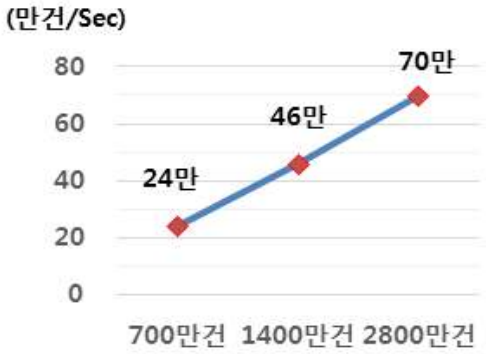
❍ 인메모리 기반 공간 빅데이터 저장 모듈 POC 수행을 통한 설계안 검증을 수행함

❍ 교통안전공단 운행기록 분석 시스템에서 수집되고 있는 DTG(Digital Tacho Graph) 자료에 대한 분석을 통해 이동객체 자료의 특징 및 구조에 대한 연구를 수행함

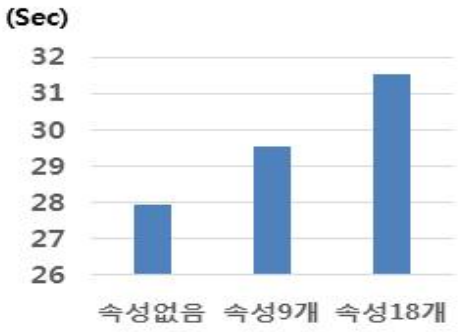


<그림 2-108> DTG 자료 저장 내용

- 운행정보를 담는 텍스트 파일의 첫 줄은 운행기록장치 모델명, 차대번호, 자동차유형, 자동차등록번호, 운송사업자등록번호 등으로 최초 사용 시 등록하는 정보이며, 운전자코드는 운송사업자가 설정하는 정보임



<그림 2-109> 자료량에 따른 초당 처리 속도



<그림 2-110> 고유 속성수에 따른 처리시간

- 교통안전공단에서 수집한 DTG 데이터의 적재 실험을 통해 클러스터의 메모리 용량 한도내에서는 처리하고자 하는 이동객체 자료의 건수는 처리 성능에 영향이 없으며 이동객체가 가지고 있는 고유 속성필드의 수에 따라 처리 성능에 영향이 있음을 확인함

⑦ [단위성과 O-6] 인메모리 기반 이동객체 빅데이터 저장 모듈 (5차연도)

❍ 공간빅데이터 수준의 이동객체 데이터를 저장 및 관리하고 실시간 수준의 데이터를 처리하기 위한 인메모리 기반 이동객체 빅데이터 저장 플랫폼을 개발함

❍ 이동객체 이동경로 최적화 기술 개발

- 이동객체의 경우 일반적으로 시간에 따라 지속적으로 위치정보가 변경이 발생되는 정보이어야 하나, 실제 수집되는 이동객체 정보의 경우 위치가 시간에 따라 변하지 않는 경우가 발생하며, 이로 인해 동일한 위치에 장시간 여러개의 이동정보가 누적되어 생성되는 경우가 많이 발생하게 됨
- 정지 상태에서 발생하는 이동객체 정보의 경우 실제 이동경로를 생성하거나 조회할시 불필요한 Node정보가 밀집한 지역에 다수 생성되어, 저장 공간의 낭비 및 불필요한 연산 Overhead를 발생시키게 됨
- 이동객체 정지상태를 체크하여 정지상태에서 지속적으로 수집되는 정보를 최적화를 통해 적재 용량 및 시공간 검색의 효율 및 속도를 개선할수 있음
- 이동객체가 동일한 Time Slot 구간 내에서 정지 임계치(ex. 30m) 범위내에서 이동할 경우, 해당 이동객체를 정지상태로 판단하고 임계치 이내에서 수집되는 이동객체 정보는 수집 및 적재하지 않음
- 정지상태로 판단되는 이동객체의 경우, 정지 지속 상태 지속시간을 duration 컬럼에 기록하여, 시공간 색인 검색시 활용함

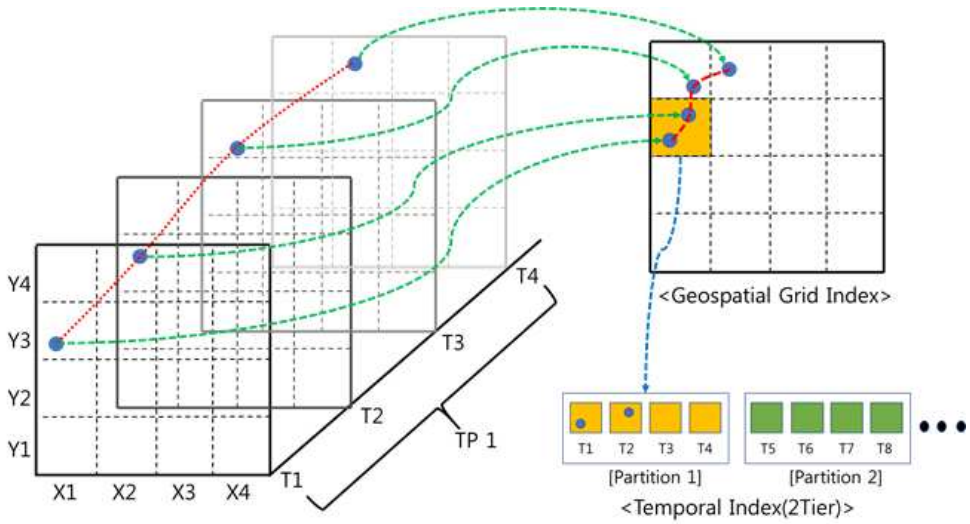


<그림 2-111> 이동경로 최적화 예시

⑧ [단위성과 O-7] 인메모리 기반 공간색인 및 조회 최적화 모듈 (5차연도)

❍ 이동객체 자료에 대한 인메모리 기반 시공간 색인 모듈을 개발함

❍ 시공간 색인 구조 설계안



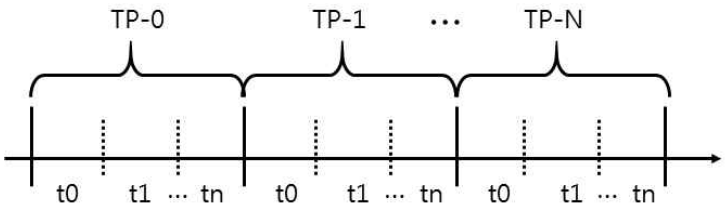
<그림 2-112> 시공간 색인 구조 개념도

❍ 공간 영역의 색인 생성 방법

- 레이어 전체 영역을 균일한 격자로 나누어 격자에 포함된 공간 정보를 그룹핑하여 색인 정보로 사용함
- 공간 자료가 여러 격자에 걸치는 경우 모든 격자에 내용을 저장하고 조회 시 중복 제거하여 사용함

❍ 시간축에 대한 색인 생성 방법

- 시간 축을 Time Partition으로 일정한 간격으로 분할하고, 다시 Time Partition을 여러개의 Time Slot으로 나누어 2Tier의 시간 분할을 함
- 수집되는 이동객체의 위치 데이터의 수집 주기에 따라 색인에 대한 Time Partition 및 Time Slot 구간을 조정함

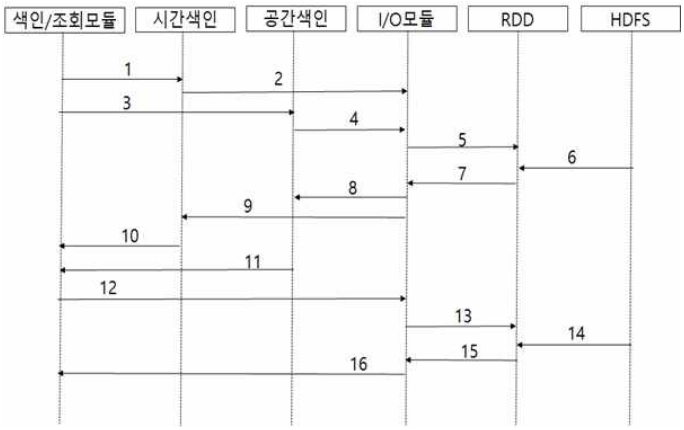


<그림 2-113> 시간축 분할 예

⑨ [단위성과 O-8] 실시간 공간 빅데이터 저장 플랫폼 v1.0 (5차연도)

❍ 인메모리 기반 공간 빅데이터 플랫폼의 저장 관리 인터페이스 및 적재된 자료 접근, 조회하기 위한 기능 및 공간 분석 라이브러리 개발을 수행함

❍ 시공간 색인을 이용한 이동객제 조회 모듈



<그림 2-114> 시공간 색인을 통한 조회

- 색인/조회 모듈에서 시공간조회에 대한 요청을 받아들이면 과정 1,3에 의해 시간색인 공간색인 순으로 색인키를 생성하여 요청하고, 과정 2,4에서는 해당 시공간키에 해당하는 색인 자료를 I/O모듈을 통해 확인함
- RDD는 RDD상에 데이터가 있는 경우 과정 6을 생략하고 없는 경우에는 과정 6에 의해 HDFS로부터 색인 정보를 가져옴
- 과정 7에 의해 전달된 색인 정보는 과정 8.7에 의해 각 색인 모듈로 전달됨
- 과정 10,11에 의해 색인에 의해 요청될 실제 데이터를 정리하여 색인/조회 모듈로 전달함
- 색인/조회 모듈은 실제 데이터 정보를 바탕으로 과정 12에서 I/O모듈을 통하여 데이터를 요청하며 과정 13에서 RDD로부터 데이터를 로드함

❍ 인메모리 기반 공간 분석라이브러리 13종을 개발함

<표 2-19> 인메모리 기반 공간분석 라이브러리

구분	공간 연산자 명
밀도 분석 (2종)	Point Density, Kernel Density
공간 통계 분석 (4종)	Optimized Hotspot, Incremental AutoCorrelation, Group Analysis, High Low Clustering,
기본 분석 (7종)	Centroid, Difference, Diff-Intersects Crosses, WithIn, Overlaps, Touch

⑩ [단위성과 O-10] 실시간 공간 빅데이터 저장 플랫폼 v2.0 (1) (5차연도)

❍ 인메모리 기반 공간 빅데이터 플랫폼에 저장된 공간 자료에 대한 공간질의를 하기 위한 통합 인터페이스 개발을 수행함

❍ 시공간 질의 수행 프로세스

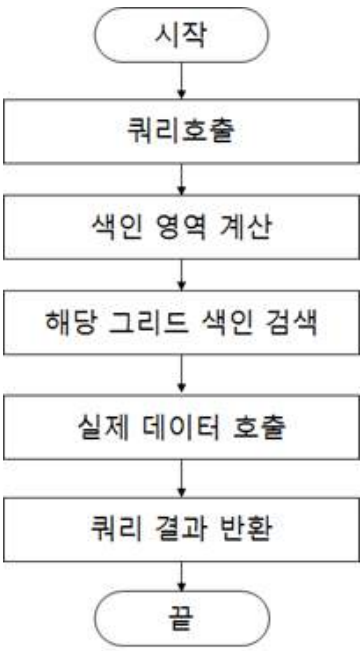
- 시공간 쿼리 영역이 호출되면 시간, 공간 순으로 색인 정보를 탐색함
- 시간색인 및 공간색인에 대하여 각각 탐색된 색인 정보 중 동일한 색인 정보를 검출하여 교집합을 생성함
- 색인 정보에 대하여 실제 데이터를 호출하고 쿼리 결과를 반환함

❍ 공간 질의 수행 프로세스

- 입력받은 조회 영역으로 색인 영역을 계산하고 해당 그리드의 색인을 검색함
- 검색된 색인 결과를 가지고 실제 데이터를 호출함
- 해당 결과를 반환하면 공간색인을 통한 조회가 완료됨



<그림 2-115> 시공간 질의 프로세스



<그림 2-116> 공간 질의 프로세스

⑪ [단위성과 O-12] 실시간 공간 빅데이터 저장 플랫폼 v2.0 (2) (6차연도)

❍ 인메모리 기반 공간 빅데이터 플랫폼의 인터페이스 개선을 통해 플랫폼의 활용성을 높이고, SparkSQL기반의 공간질의 처리 모듈 기능을 개발하였음

❍ 플랫폼 인터페이스 개선

- 기존에 Spark 1.x 구현된 자료 입출력 인터페이스를 Spark 2.3 기반의 자료 입출력 인터페이스로 재구현 수행
- 기존에 구현된 RDD 기반의 색인 모듈을 DataSet 기반의 색인 모듈로 개선을 통해 성능 및 메모리 최적화함

❍ SparkSQL 기반 공간질의 처리 모듈 개발

- SparkSQL 기반에서 활용할수 있는 Geometry Column Type 개발
- Geometry Type에 대한 WKT와 WKB 자료 유형간 자동 변환 기능 개발
- SparkSQL 기반으로 공간질의 연산자 37종 개발 및 동작 정확도 100% 확인(TTA 시험 실시)

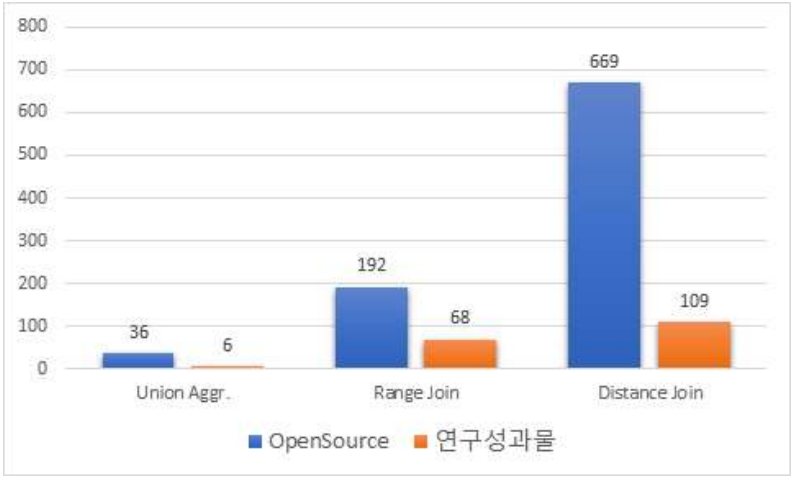
구분	Function 명
Constructor	ST_GeomFromGeoJSON
	ST_GeomFromText
	ST_GeomFromWKB
	ST_MakeLine
	ST_MakePoint
	ST_MakePolygon
Output	ST_AsBinary
	ST_AsGeoJSON
	ST_AsText
Accessor	ST_Dimension
	ST_Envelope
	ST_GeometryN
	ST_GeometryType
	ST_IsClosed
	ST_IsEmpty
	ST_IsSimple
	ST_IsValid
	ST_SRID

구분	Function 명
Relationships and Measurements	ST_Area
	ST_Centroid
	ST_Contains
	ST_Covers
	ST_Crosses
	ST_Distance
	ST_Equals
	ST_Intersects
	ST_Length
	ST_Touches
	ST_Within
	ST_SetSRID
	ST_Transform
Processing	ST_Buffer
	ST_ConvexHull
	ST_Difference
	ST_Intersection
	ST_SymDifference
	ST_Union

<그림 2-117> SparkSQL 기반 Spatial SQL 목록

❍ SparkSQL 기반 공간질의 성능 실험

- (분석 기능 선정 기준) 평가 환경은 8대의 데이터 노드, 1대의 네임 노드, 그리고 1대의 Yarn Resource Manager로 구성된 하둡 시스템을 구축하여 실시하였으며, 시험용 자료로 서울지역 연속지적도와 서울지역 노인복지시설 자료를 활용하였음
- 성능시험 항목의 경우 GeoSpark이 제공하고 있는 SpatialSQL 기준으로 성능 시험 항목을 선정하여 성능 비교를 실시함
- 인메모리 기반 공간질의 기능을 가지고 있는 OpenSource(GeoSpark)과 연구 성과물과의 성능 비교 결과 평균 399% 성능 향상 확인 (TTA 시험 실시)



<그림 2-118> OpenSource(GeoSpark) Vs 연구 성과물

- 성능 비교 결과 모두 연구 성과물이 더 빠른 성능을 보였으며, 특히 Distance Join의 경우 GeoSpark 대비 6배 이상 빠른 성능을 나타내었음
- (결과 의미) SparkSQL의 경우 Spark기반에서 자료 분석시 매우 많이 활용되나, GeoSpark이 제공하는 SpatialSQL 10여종 정도로 실무에 활용되지는 못하고 있음
- (결과 의미) 본 연구를 통해 개발된 SpatialSQL 37종은 가장 많이 활용되는 기능들을 선정하여 개발하였으며, GeoSpark보다 빠른 성능 확보를 통해 대용량 자료에 대한 공간 분석시에 활용될 수 있음

⑫ [단위성과 O-13] 실시간 인메모리 기반 저장 플랫폼 실증 (6차연도)

○ 메모리 기반 공간 빅데이터 플랫폼을 국토부 공간 빅데이터 사업단에 실증을 통해 플랫폼의 호환성을 검토하고, 개발된 플랫폼의 성능 및 기능에 대한 실증을 수행함

○ 실증 수행 내용

- 사업단 플랫폼과의 호환성을 위한 별도의 자료 포맷 변환 모듈 개발 및 연계
- 1~6차년도 주요 연구 성과물을 대상으로 실증 수행
- 2차에 걸친 핵심 연구성과 실증 테스트 진행 (1차 : '19.11.06, 2차 : '19.11.19)

<표 2-20> 실시간 인메모리 기반 저장 플랫폼 실증 항목

구분	주요 연구 성과물	비고
실시간 공간 빅데이터 저장 플랫폼	인모메리 기반 공간 빅데이터 저장 기능	
	인모메리 기반 공간 색인/검색/조회 기능	
	인메모리 기반 공간라이브러리 23종	
	SparkSQL기반 질의 처리 기능	
인메모리 기반 이동객체 빅데이터 저장 플랫폼	인메모리 기반 이동객체 빅데이터 저장 및 관리 기능	
	인메모리 기반 이동객체 빅데이터 색인/검색/조회 기능	

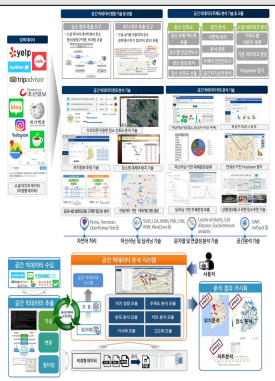
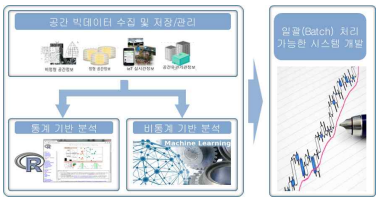
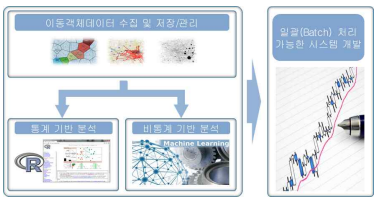
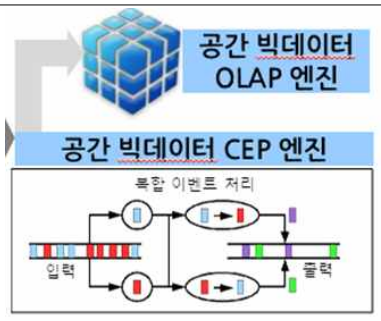
(2) 공간 빅데이터 분석 및 가시화 기술 개발

○ 2세부의 연구내용을 수행하여 다음과 같이 4개의 핵심성과를 도출함

○ 1세부의 공간 빅데이터 저장 및 관리 기술을 통해 수집된 공간 빅데이터를, 그 종류 및 특성에 따라 각 세세부에서 분석하고자 함

○ 서울대/시터스에서는 공간 정보가 포함된 소셜미디어 데이터(비정형데이터)를 분석하는 솔루션을, 연세대/가이온에서는 레거시 공간빅데이터 및 이동객체데이터를 분석하는 솔루션을, 한국과학기술원/유엔진솔루션에서는 센서로부터기 수집되거나 실시간으로 수집되는 공간 데이터를 분석하는 솔루션을 개발함

<표 2-21> 2세부 연구내용 요약

연구 핵심 성과		수행 기관	연차	연구수행 내용 및 성과
E	공간 빅데이터 소셜 분석 기술, 공간 빅데이터 분석 및 가시화 컴포넌트 구현	서울대/시터스	1~6	사용자가 생성한 비정형 데이터(트위터, 네이버블로그, 다음 아고라, 뉴스기사, 블로그, 트립어드바이저, 망고플레이트, 구글리뷰, Yelp 등)를 이용하여 공간 빅데이터 정합, 분포, 차트, 가시화 및 주제도 분석을 위한 알고리즘 및 이를 기반으로 공간 빅데이터 분석 및 가시화 컴포넌트를 개발함 
F	공간 빅테이터 Batch 분석	연세대/가이온	1~6	공간 빅데이터를 분석하기 위한 분석 컴포넌트 설계, 최적화 및 고도화를 수행하고 Web platform에서 drag and drop 방식을 통해 사용자가 원하는 분석 시나리오를 생성할 수 있는 분석 솔루션을 구현함 
G	이동객체 데이터 Batch 분석	연세대/가이온	1~6	이동객체데이터를 분석하기 위한 분석 컴포넌트 설계, 최적화 및 고도화를 수행하고 Web platform에서 drag and drop 방식을 통해 사용자가 원하는 분석 시나리오를 생성할 수 있는 분석 솔루션을 구현함 
H	공간 빅데이터 Interactive 분석 기술 개발	한국과학기술원/유엔진솔루션즈	1~6	실시간 입력되는 공간 빅데이터를 대화형으로 신속 정확하게 분석할 수 있는 기술을 개발하였으며 크게 공간 빅데이터 분석 기반 복합 이벤트 처리 기술과 공간 빅데이터 기반 대화형 분석 기술로 구성되어 있음 

❍ 2세부는 분석대상 데이터의 성격에 따라 연구주제가 결정되어 있으므로, 이질적인 공간 빅데이터를 동시에 요구하는 응용에서 적절하게 연계될 수 있음

❍ 따라서 2세부의 각 세세부들의 연구성과의 통합 연계성을 제시함

- 공간 빅데이터 소셜분석(서울대/시터스)과 공간 빅데이터/이동객체 데이터 배치 분석(연세대/가이온) 연계: 소셜 데이터에 포함된 공간정보와 공간 빅데이터/이동객체 데이터에 포함된 공간정보는 서로 다른 형태의 사용자 모빌리티 정보를 보유하고 있으므로 다양한 데이터를 분석에 활용하여 다각적인 분석이 가능함
  - 예: 소셜 데이터와 공간 빅데이터로 공공자전거 수요예측을 수행하여 각기 다른 입력데이터로 다각적인 분석이 가능함을 보임 (E-13, F-15 참조)
- 공간 빅데이터/이동객체 데이터 배치 분석(연세대/가이온)과 공간 빅데이터 Interactive 분석(한국과학기술원/유엔진솔루션즈) 연계: 공간 빅데이터/이동객체 데이터 배치 분석(연세대/가이온)의 분석결과를 실시간 분석 혹은 대화형 분석의 입력 데이터로 활용해서 다차원적인 분석이 가능함
  - 예: 공간 빅데이터인 Origin-Destination(O/D) 자료를 통해 유동인구 패턴을 산출하고, 이를 공간 빅데이터 OLAP 엔진의 입력으로 사용해 다양한 레벨의 지역 및 기간에서의 패턴을 대화형으로 탐색할 수 있음 (G-17, H-14 참조)
  - 예: 이동객체 데이터인 택시운행 데이터로 택시 승하차 Hot & Cold 지역(혹은 교통사고 위험지역)을 산출하고, 이를 공간 빅데이터 CEP 엔진의 규칙 입력으로 사용해 해당 지역에 택시가 진입할 시 빠르게 알람을 줄 수 있음 (G-13, H-14 참조)

(가) 공간 빅데이터 소셜 분석 기술, 공간 빅데이터 분석 및 가시화 컴포넌트 구현 (핵심성과 E)



<그림 2-119> 공간 빅데이터 소셜 분석 기술 및 공간 빅데이터 분석 및 가시화 컴포넌트 구현 (핵심성과 E)

○ 사용자가 생성한 비정형 데이터(트위터, 네이버블로그, 다음 아고라, 뉴스기사, 블로그, 트립어드바이저, 망고플레이트, 구글리뷰, Yelp 등)를 이용하여 공간 빅데이터 정합, 분포, 차트, 가시화 및 주제도 분석을 위한 알고리즘 및 이를 기반으로 공간 빅데이터 분석 및 가시화 컴포넌트를 개발함

① 공간 빅데이터 분석을 위한 모델 설계 및 목표 성능 설정

㉮ [단위성과 E-1] 공간 빅데이터 분석모델 현황분석 및 요구사항 정의(1차연도)

○ 소셜 데이터는 분포 분석 분야에서 Spatial Autocorrelation, Spatial Association rules, Classification/Clustering, Hotspot 분석 기법들이 주로 적용될 수 있음

○ 변화 분석인 Anomaly Detection, 연결성 분석인 Information Diffusion, Influence Estimation, Community Diagnostics 분석 기법이 주로 적용될 수 있음

○ 소셜 데이터에 대한 텍스트 마이닝을 통해 Unigram/Bigram, Recursive Neural Model, Sentiment Treebank 등의 감성 분석 기법이 적용될 수 있음

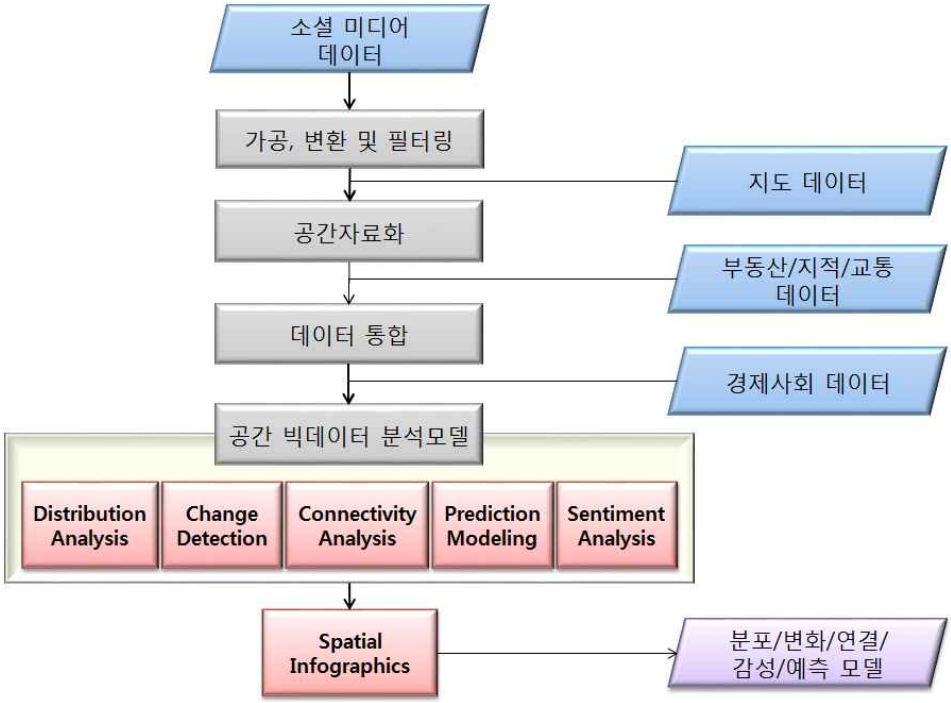
- 기본 공간정보와의 중첩 표현을 통해 Map-based Infographics, Network Visualization 등의 시각화 기법이 적용될 수 있음



<그림 2-120> 소셜 데이터에 대한 공간 빅데이터 분석 모델 기법

㉯ [단위성과 E-3] 공간 빅데이터 분석 목표모델 정의(1차연도)

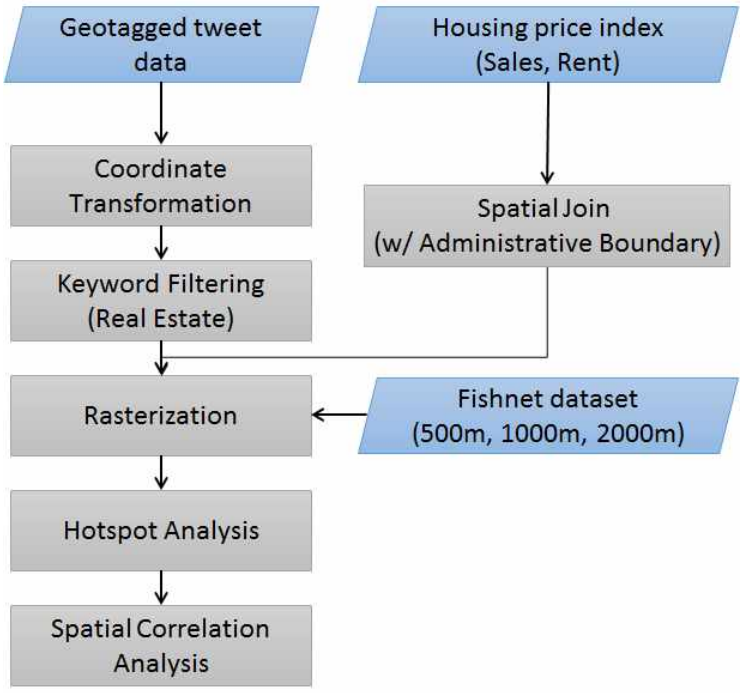
○ 공간 빅데이터 분석 모델은 소셜 미디어 데이터를 주요 대상으로 하고 있으며, 이를 공간자료화하기 위한 가공, 변환 및 필터링, 공간자료화 과정을 수행해야 함



<그림 2-121> 공간 빅데이터 분석 모델 설계도

❍ 목표모델 적용 사례

- 반정형/비정형 데이터인 위치기반 소셜 미디어 정보와 부동산 가격 정보에 대한 공간적 분포를 비교분석 함으로써 두 데이터 간의 공간적 상관성을 분석하기 위한 격자 기반의 분석 방법론을 제안하였음



<그림 2-122> 연구흐름도

❍ 대상데이터

- 본 연구의 대상 데이터는 소셜 미디어 데이터와 부동산 가격지수 데이터임
- 소셜 미디어 데이터로는 지오태그 정보를 포함하고 있는 트윗(tweet) 데이터를 주요 대상 데이터로 하였음
- 트윗 데이터는 2013년 8월 한달간 수집된 자료로 한정하였고 공간적 범위는 서울특별시를 중심으로 하여 일산, 분당, 과천 등 주요 신도시들을 포함하는 수도권으로 하였음
- 트윗 데이터의 위치값에 대하여 사각형 형태의 공간질의(Spatial Query)를 적용하였는데, 범위는 좌상단 (176778.7, 472112.8), 우하단 (218778.7, 432112.8)으로 하였음
- 부동산 가격 정보로는 한국감정원에서 구축, 제공하고 있는 행정구역(시군구) 단위의 주택가격지수(매매지수, 전세지수)를 보조 대상 데이터로 하였으며, 시간적 범위는 2013년 8월로 한정하고 공간적 범위는 트윗 데이터의 수집범위에 맞게 서울특별시(25개 구), 인천광역시(계양구, 부평구, 남동구의 3개 구), 경기도(고양시, 부천시, 안양시, 성남시 등) 일부지역을 포함하여 총 51개 시군구에 대한 주택가격지수(2013년 8월 주택매매지수와 주택전세지수)를 수집하였음

❍ 격자기반 공간분석 방법론

- 데이터 공간자료화
  - 위치 태그를 가진 트윗 데이터를 수집하기 위해, 트위터 공개 API[5]를 활용하여 위치 태그를 가진 트윗 데이터 중 대상 범위 내에 포함되는 트윗 데이터만을 자동으로 추출하였음
  - 키워드 필터링을 적용하였고, 부동산 관련 키워드(부동산, 주택가격, 집값, 주택매매, 아파트 시세, 전세, 월세 등)를 포함한 트윗 데이터들을 필터링 하였음
  - 총 193,831개 트윗 데이터 중 키워드 필터링을 통해 85개 데이터가 추출되었음
  - 트윗 데이터 지오태그 상의 위치정보는 경위도 좌표로 되어 있어 좌표변환을 통해 경위도 좌표를 TM(Transverse Mercator) 좌표체계로 변환하였으며, 이를 위해 국토지리정보원의 NGIpro(ver. 2.53) SW를 활용하였음
  - 주택가격지수(매매지수, 전세지수) 데이터는 행정경계(시군구) 폴리곤 데이터를 이용하여 공간조인을 통해 공간자료화 하였음
- 격자 데이터화(Rasterization)
  - 트윗 데이터와 주택가격지수 데이터는 각각 포인트 데이터와 폴리곤 데이터 형태이기 때문에 데이터의 형식에도 차이가 있지만 공간적 세밀도에서도 차이를 보임
  - 이에 격자 데이터셋 형태로 두 데이터를 변환시킴으로써 공간적 분포의 비교분석과 공간적 연관성 분석을 보다 용이하게 하고자 하였음
  - 격자형 데이터셋의 구축을 위해 ArcMap toolbox 중 ‘create fishnet’ 기능을 적용하여

5) 트위터 공개 API: 소셜 네트워크 서비스 업체인 ‘트위터(Tweeter)’에서 자사의 트윗 데이터의 활용을 지원하기 위해 개발자들에게 공개한 데이터 제공 인터페이스. 트윗, 사용자, 위치정보 등의 정보를 수집할 수 있음. https://dev.twitter.com/overview/api (2014년 12월 28일 방문)

격자망 데이터를 생성하였음

- 격자망 데이터셋의 크기는 트윗 데이터를 추출한 범위와 동일하게 설정하였으며, 격자망 크기는 500m, 1000m, 2000m 순으로 적용하였음
- 각 크기의 격자망 데이터셋에 트윗 데이터를 중첩시켜 각 격자 내에 포함된 트윗 데이터의 개수를 세어 격자의 속성정보로 입력하였음
- 주택가격지수 역시 격자망 데이터셋과 중첩하여 각 격자에 대한 가격지수를 재계산 하여 입력하였음. 이때 한 격자 내에 다수의 행정경계 폴리곤이 만날 경우에는 각 폴리곤의 격자 내 포함비율을 계산하여 면적 비율을 가중치로 한 주택가격지수의 가중평균합을 구하였음

- Hotspot 분석

- 소셜 미디어인 트윗 데이터와 주택가격지수 데이터의 공간적 분포를 시각적으로 비교하기 위해 Hotspot 분석을 실시하였음
- ArcMap toolbox 중, ‘Hotspot analysis’ 도구를 이용하여 격자화된 트윗 데이터와 주택가격지수 데이터에 대한 국지적 공간 클러스터링을 탐색하였음

- 공간적 상관성 분석

- 공간적 상관성 분석은 트윗 데이터와 주택가격지수 데이터의 공간적인 분포 패턴이 어느 정도의 유사성 또는 상관성을 가지는지를 수치적으로 파악하기 위해 이변량 공간 상관성 측도인 Spatial Lag Pearson’s r을 적용하였음
- Spatial Lag Pearson’s r이란 일반적인 상관분석에서 쓰이는 Pearson’s r에 spatial lag(공간시차)를 적용한 것으로, 여기서 공간시차란 공간가중치 행렬(Spatial Weight Matrix)에 의해 정의된 이웃들의 가중평균합으로서 계산함

$$\widetilde{x_i} = \sum_j w_{ij} x_j$$

- 여기서 w_{ij}는 i와 j가 인접하면 1, 그렇지 않으면 0의 값을 갖는 이진행렬로 행표준화가 행해진 값이며, x_j는 j번째 변수를 의미함
- Spatial Lag Pearson’s r은 다음 식과 같이 공간시차가 적용된 변수 X와 Y의 공분산을 각 표준편차의 곱으로 나눔으로써 계산할 수 있음

$$r_{\widetilde{X},\widetilde{Y}} = \frac{\sum_i (\widetilde{x_i} - \overline{\widetilde{x}})(\widetilde{y_i} - \overline{\widetilde{y}})}{\sqrt{\sum_i (\widetilde{x_i} - \overline{\widetilde{x}})^2} \sqrt{\sum_i (\widetilde{y_i} - \overline{\widetilde{y}})^2}}$$

- 여기서 $\widetilde{x_i}$, $\widetilde{y_i}$는 i 위치에서 두 변수 X와 Y의 공간시차에 의한 값이며, $\overline{\widetilde{x}}$, $\overline{\widetilde{y}}$는 X와 Y의 공간시차에 의한 이웃 값들의 평균임
- 트윗 데이터와 주택가격지수의 격자 데이터에 대하여 공간적 상관계수(Spatial Correlation Coefficient, SCC)를 계산하였음

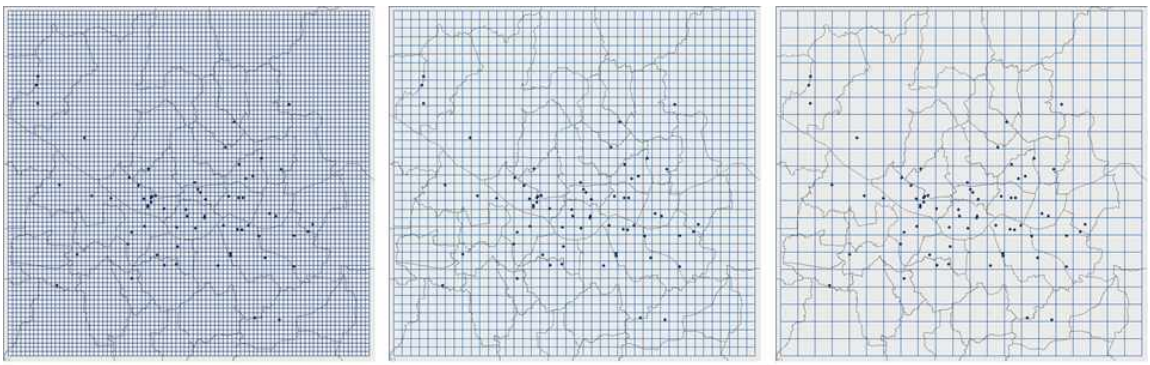
$$SCC_{AB} = \frac{\sum_{i}^{n} (\widetilde{A}_i - \overline{\widetilde{A}})(\widetilde{B}_i - \overline{\widetilde{B}})}{\sqrt{\sum_{i}^{n} (\widetilde{A}_i - \overline{\widetilde{A}})^2} \sqrt{\sum_{i}^{n} (\widetilde{B}_i - \overline{\widetilde{B}})^2}}$$

- 여기서 SCC_{AB}는 A, B 데이터 간의 공간적 상관계수, i는 각 격자, n은 격자의 개수, $\widetilde{A}_i$는 i번째 격자에서 A 데이터의 공간시차에 의한 격자값, $\widetilde{B}_i$는 i번째 격자에서 B 데이터의 공간시차에 의한 값을 의미함
- 이에 대한 결과값은 0과 1사이의 범위를 갖게 되며, 값의 해석은 일반적으로 0.3 이상이면 뚜렷한 공간적 상관성이 있다고 보고, 0.7 이상이면 강한 공간적 상관성을 보인다고 할 수 있음

○ 적용 및 결과

- 데이터 격자화 및 통합

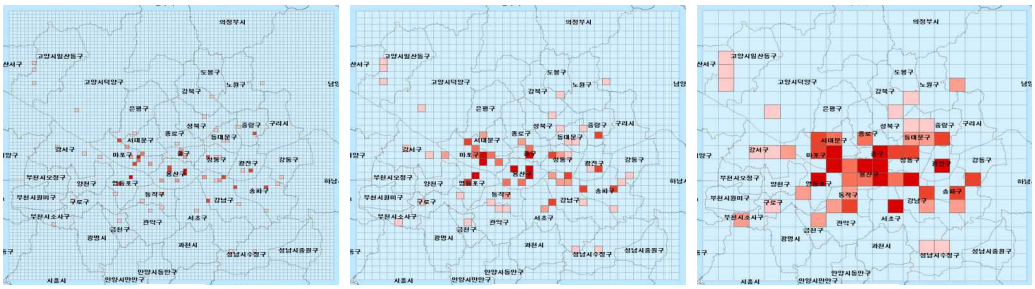
- 트윗 데이터와 주택가격지수 데이터 각각의 공간적 세밀도를 고려하였을 때, 500m, 1,000m의 크기로 격자를 생성한 경우에는 격자간 거리가 너무 좁아 트윗 데이터를 포함하는 격자의 갯수가 전체 격자의 1% 이내가 되어 격자값의 전체적인 분포특성을 파악하기에 어려움이 있을 것으로 예상되며, 2,000m의 격자크기가 두 데이터의 공간적 세밀도에 적절하다고 판단됨



(a) (b) (c)

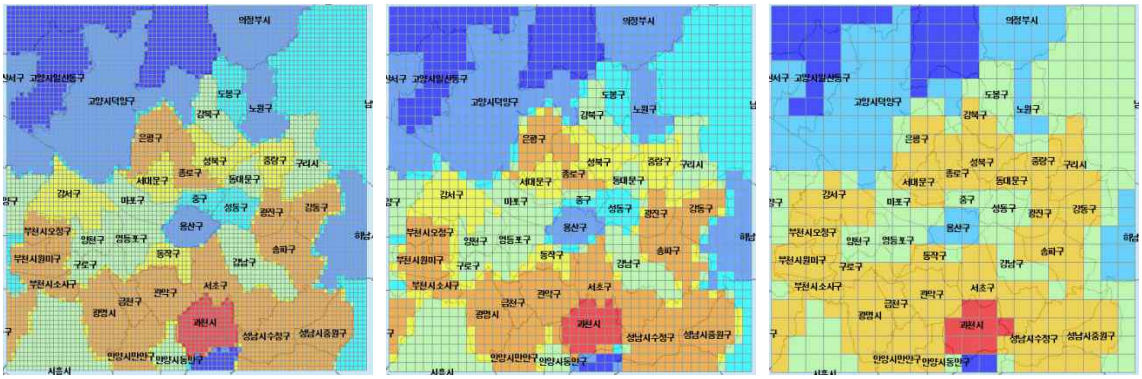
<그림 2-123> 트윗(point), 주택가격지수(polygon) 및 격자망 데이터(grid)
((a): 500m, (b): 1000m, (c): 2000m)

- Hotspot 분석 결과



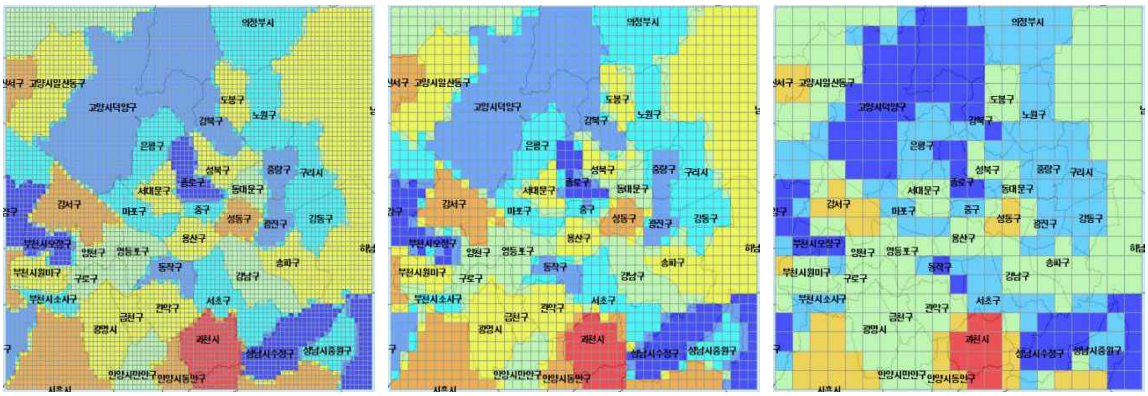
(a) (b) (c)

<그림 2-124> 트윗 격자 데이터의 핫스팟 분석 결과 ((a) 500m grid, (b) 1,000m grid, (c) 2,000m grid)



(a) (b) (c)

<그림 2-125> 주택매매가격지수 격자 데이터의 핫스팟 분석 결과
((a) 500m grid, (b) 1,000m grid, (c) 2,000m grid)



(a) (b) (c)

<그림 2-126> 주택전세가격지수 격자 데이터의 핫스팟 분석 결과
((a) 500m grid, (b) 1,000m grid, (c) 2,000m grid)

- 트윗 데이터에 대한 Hotspot 지역은 500m 격자 데이터와 1,000m 격자 데이터에서는 주로 용산구와 중구, 마포구와 영등포구에서 탐색되었으며, 2,000m 격자 데이터에서는 추가적으로 서초구와 광진구, 강남구, 송파구에서도 탐색되었음
- 주택매매가격지수 데이터에 대한 Hotspot 지역은 과천시에서 가장 두드러지게 나타났으며, 그 외에 서울 서남부 지역과 강동 지역, 부천시와 성남시에서 약한 Hotspot 지역이 탐색되었음. 고양시와 용산구, 안양시 등은 Coldspot 지역으로 탐색되었다. 이러한 분포 패턴은 격자망의 크기가 달라짐에 따라 큰 차이가 없었음
- 주택전세가격지수 데이터에 대한 Hotspot 지역은 과천시가 강한 Hotspot 지역으로 나타났으며, 그 외에 성동구, 강서구, 시흥시, 고양시 일산서구도 Hotspot 지역으로 탐색되었음. 반면에 종로구, 인천시 계양구, 부천시 오정구, 성남시 수정구와 광주시는 Coldspot 지역으로 탐색되었음
- 격자망의 크기에 따른 분포 패턴의 차이는 역시 크지 않은 것으로 나타났음
- 결과적으로, 트윗 데이터의 Hotspot 지역과 주택가격지수 데이터의 Hotspot 지역은 공간적 연관성이 높지 않은 것으로 나타났음. 이는 트윗 데이터가 많이 발생하는 지역이라고 해서 주택가격이 높지 않을 수 있다는 것을 의미함

- 공간적 상관성 분석 결과

- 트윗 데이터와 주택가격지수 데이터의 공간적 상관성을 분석하기 위해 격자화된 트윗

데이터와 매매지수, 전세지수 데이터 간의 SCC를 측정하였음

<표 2-22> SCC values between tweet data and housing price indices for each grid size

SCC	tweet data(500m)	tweet data(1,000m)	tweet data(2,000m)
House sales price index	0.0277	0.0666	0.1734
House lease price index	-0.0072	-0.0130	0.2411

- 500m와 1,000m 격자 크기에서의 트윗 데이터와 주택가격지수 데이터 간의 상관성은 매우 낮거나 오히려 음의 관계를 나타내고 있으나 2,000m 격자 크기에서는 아주 약간 상관관계를 보이고 있음
- 이는 부동산과 관련된 트윗 데이터의 양이 많다고 해서 꼭 주택가격이 높은 것은 아니라는 것을 의미함
- 격자 크기에 따라 공간적 상관성의 분석 결과가 달리 나타날 수 있음
- 이는 이종의 공간 데이터 간의 통합을 위해 격자 데이터화 방법론을 적용하는데 있어서 적절한 격자망의 크기를 선정하는 것이 매우 중요한 부분임을 보여주고 있음

② [단위성과 E-4] 이기종 공간DB 특성 분석(2차연도)

○ 소셜 미디어 데이터 중에서 트위터, 인스타그램, 유튜브, 구글 트렌드를 대상으로, 위치(공간)과 관련된 필드, 사용자가 입력한 텍스트나 사진 등의 멀티미디어 필드, 사회적 네트워크 관련 필드 등이 간 빅데이터 분석을 위한 중요한 필드로 분석됨

<표 2-23> 공간분석을 위해 선택된 속성필드

구분	오브젝트	대상 필드		데이터 타입
트위터	Users	id		integer
	Tweets	created_at		string
		id		integer
		text		string
		retweet_count		integer
		favorite_count		integer
		coordinates	coordinates	float
	Entities	hashtags	text	string
	Places	id		string
		country		string
		country_code		string
		place_type		string
		name		string
		full_name		string
		url		string
인스타그램	/media/search	id		string
		tags		array of string
		location		string
		created_time		integer
		images	url	string
		caption	text	string

구분	오브젝트	대상 필드	데이터 타입
유튜브	video	snippet.publishedAt	datetime
		snippet.description	string
		contentDetails.duration	string
		statistics.viewCount	unsigned long
		statistics.likeCount	unsinged long
		statistics.dislikeCount	unsigned long
		recordingDetails.locationDescription	string
		recordingDetails.recordingDate	datetime
		fileDetails.creationTime	string
		suggestions.tagSuggestions[].tag	string

③ [단위성과 E-5] 이기종 DB간 데이터 가공, 변환, 필터링 기술(2차연도)

㉮ 가공 (Processing)

○ 공간자료화 프로세스

- 소셜 미디어 데이터를 분석하여 공간분석을 위한 필드를 분류하고, 소셜 미디어 데이터를 수집과 축적함. 이후 축적된 파일을 JSON에서 CSV파일로 변환한 후 일자별 취득된 데이터를 통합함. 이때 중복되는 데이터를 제거하는 정제과정을 거침



<그림 2-127> 소셜 미디어 데이터의 공간자료화 프로세스

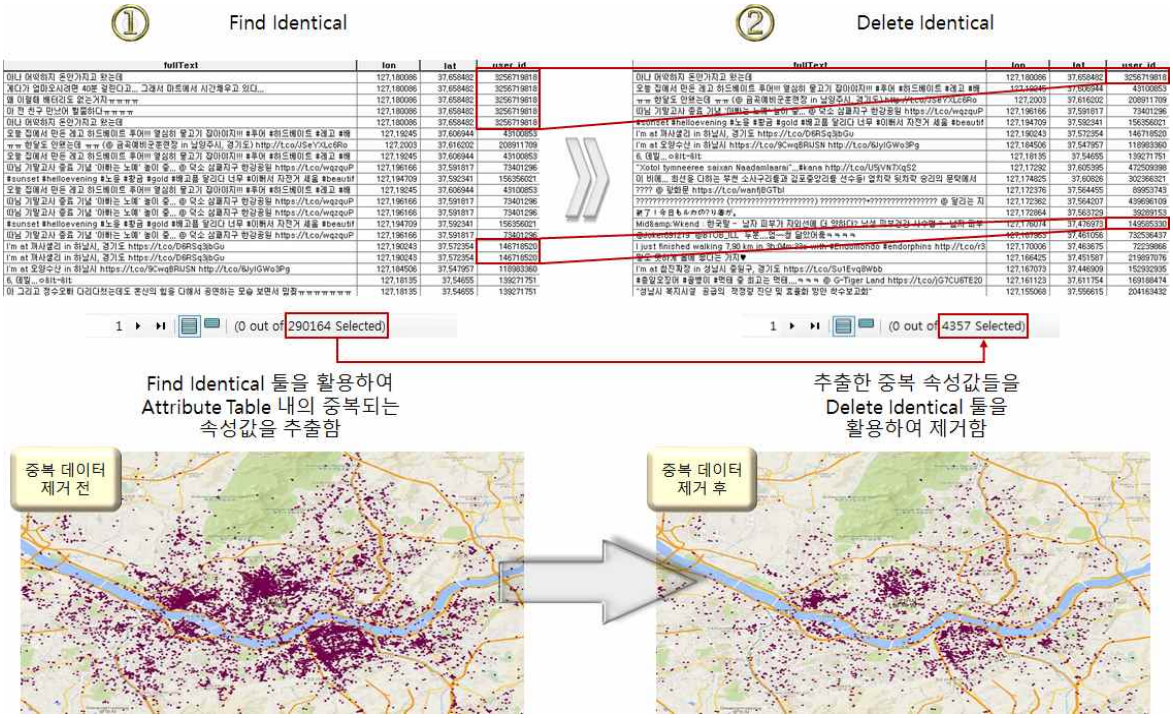
○ 실험 및 결과

- 소셜 미디어 데이터를 정해진 기간 동안 지오태깅된 데이터만 수집한 후, 수집된 데이터를 일자별로 병합함

<표 2-24> 대상 소셜 미디어 데이터 수집

소셜 미디어	수집 기간	획득 데이터
트위터	2015.06.22 ~ 2015.09.30	255,728
인스타그램	2015.07.01 ~ 2015.10.02	11,715
유튜브	2015.06.01 ~ 2015.09.17	22,721

- ArcMap에서 제공하는 ‘Find Identical(Data Management)’과 ‘Delete Identical’ 툴을 사용하여 속성정보 중 중복되는 정보를 제거함



<그림 2-128> 중복 데이터 제거 프로세스

㉯ 변환 (Conversion)

❍ 소셜 미디어 정보에서의 위치(공간)정보 분석

- 트위터, 인스타그램, 유튜브와 같은 소셜 미디어 정보에서 제공하는 위치정보는 크게 ① 경위도 좌표와 ② 사용자 입력 행정경계나 건물이나 상점의 명칭과 같은 POI(Point Of Interest) 명칭을 제공하는 필드로 구분됨
- 사용자 입력 행정경계 명칭은 행정경계 명칭 사전을 바탕으로 유사한 행정경계 명칭이 정형화된 행정경계 명칭으로 미리 정의되어야 하므로, 행정경계 명칭 사전을 구축함. 이후 V-World에서 제공하는 오픈API 중에서 지도검색 API의 Geocoder API를 이용해서 경위도 좌표로 지오코딩(Geocoding)함
- 이에 소셜 미디어 정보의 ① 경위도(WGS84 타원체) 좌표와 ② 사용자 입력 행정경계 명칭필드(입력값)를 WGS84타원체의 경위도 좌표로 변환하는 과정은 경위도 좌표 -> 사용자 입력 행정경계 명칭 순으로 진행됨

<표 2-25> 대상 소셜 미디어 정보에서 제공되는 위치정보 분석결과

SNS	위치정보		
	구분	필드 이름	필드 설명
트위터	프로필	user.location	사용자 입력 행정구역/POI 명칭
	트윗	coordinates.coordinates	경위도 좌표
		place.country_code	
		place.country	국가
		palce.full_name	
인스타그램	사진	place.place_type	
		data__location__longitude	위도 좌표
		data__location__latitude	경도 좌표
		data__location__name	사용자 입력 행정구역/POI 명칭
유튜브	video	items__recordingDetails__location__latitude	경도 좌표
		items__recordingDetails__location__longitude	위도 좌표

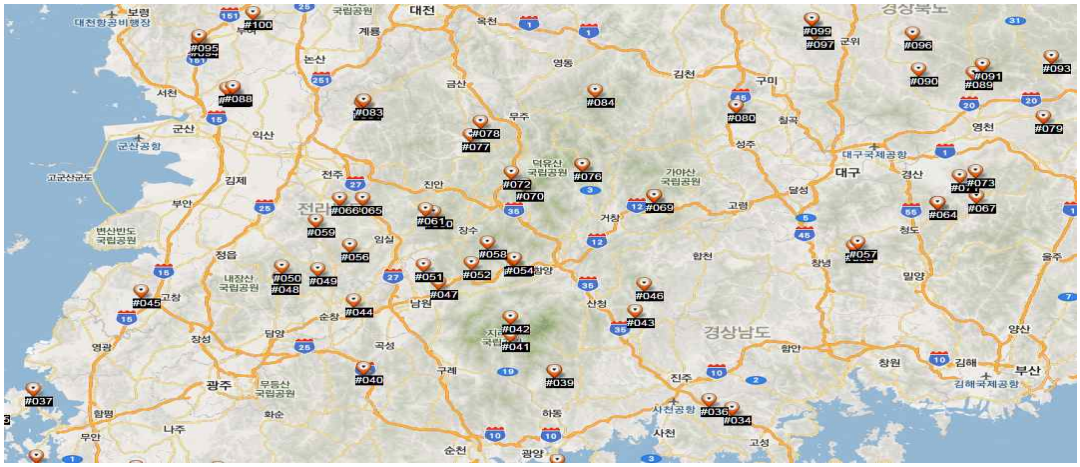
<표 2-26> 소셜 미디어 정보에서의 위치정보 변환 기법

① 경위도(WGS84 타원체) 좌표	
입력 값	트위터 / 인스타그램 / 유튜브 경위도 좌표 필드
1. 입력값을 (lon, lat)필드에 복사	
출력 값	경위도(WGS84 타원체) 좌표

② 사용자 입력 행정경계 명칭	
입력 값	트위터 / 인스타그램 / 유튜브 사용자 입력 행정경계 필드
1. 입력값과 행정경계 명칭 사전과 유사도 구하기 2. 유사도가 100%, 즉 행정경계 명칭 사전과 일치하는 경우만 입력 값을 사전에서 정의된 행정경계 명칭으로 변경 3. 정의된 행정경계 명칭을 V-World Geocorder API의 파라미터로 이용하여 경위도(WGS84 타원체) 좌표로 변환	
출력 값	경위도(WGS84 타원체) 좌표

❍ 실험 및 결과

- 2015년 7월 11일부터 2016년 1월 11일까지 수집된 트위터, 인스타그램, 유튜브 939,233건에서 대해서 제안된 변환 기법을 적용하여 V-World에서 제공하는 지도에 표출하는 실험을 진행함
- 그 결과,트위터, 인스타그램, 유튜브 데이터가 지오코딩되어 행정경계 명칭에 해당하는 경위도 좌표로 변환되어 표출되는 것을 시각적으로 확인할 수 있었음



<그림 2-129> 제안된 변환 기법을 적용한 사례

㉰ 필터링 (Filtering)

❍ 데이터 필터링(Data Filtering) 정의

- 오류 발견, 보정, 삭제 및 중복성 확인 등의 과정을 통해 데이터 품질을 향상 시키는 기술로, 가공되고, 변환되어진 소셜 미디어 정보를 얼마나 효과적으로 정제하여 공간 정보와 병합하는지, 즉 지도위에 얼마나 효과적으로 보여줄 수 있는지가 연구의 초점

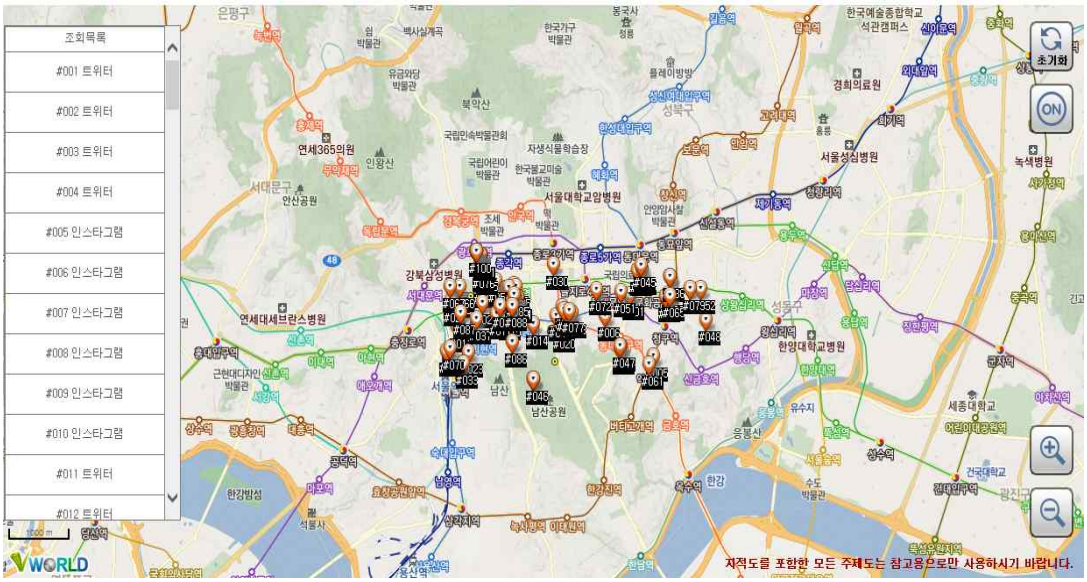
❍ 데이터 필터링 기준 설정 및 적용

- 소셜 미디어 정보의 필터링 기준을 공간적인 특성에 맞게 정의하고 설정 할 수 있는 기능을 제공해야 함. 따라서 공간적 범위, 시간적 범위, 데이터 내용의 구분을 통하여 사용자가 공간 소셜 데이터 분석이 편리하도록 설정

<표 2-27> 소셜 미디어 정보의 필터링 기준

구분	특징
검색범위	전체
	행정구역
	위치 / 검색어 기반(5~100km) 검색
검색기간	오늘 / 1개월 / 3개월 / 6개월
데이터 선택(출처)	필터링 소셜 미디어 데이터 선택
	트위터 / 인스타그램 / 유튜브
멀티미디어 유형(데이터)	텍스트 / 동영상 / 사진

- 검색범위를 행정구역(서울특별시 중구)으로 하고, 검색범위를 2015년 12월 12일부터 2016년 1월 12일로 설정하고, 데이터 선택(출처)을 모든 소셜 미디어 데이터로 설정한 후 필터링된 결과는 아래와 같음

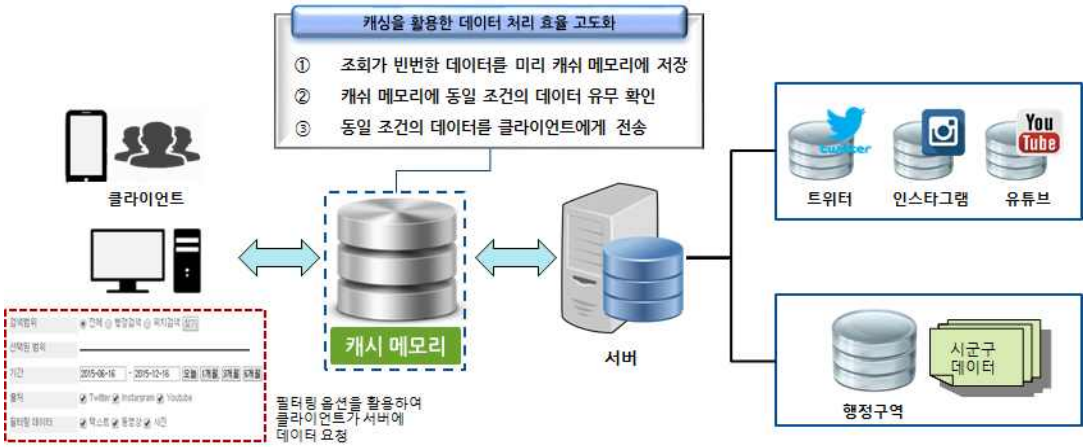


<그림 2-130> 소셜 미디어 데이터 분석 서비스 필터링 결과 화면

④ [단위성과 E-6] 다중 레이어를 위한 캐싱/프리패칭/버퍼링 기술(2차연도)

㉮ 캐싱 (Caching)

○ 사용자(클라이언트)측 조회패턴을 미리 예측하여 평균적으로 조회량이 많은 데이터를 중간 저장소인 캐시 메모리에 미리 저장하고, 사용자가 요청한 데이터를 캐시 메모리에서 직접 전달해 효율성을 높임



<그림 2-131> 캐싱 Work-flow

㉯ 버퍼링 (Buffering)

○ 버퍼를 데이터 중간저장소의 기능으로 활용하여 전체 소셜 네트워크 데이터를 한곳에 통합 보관하고 필요한 상황에 따라 요청 데이터를 분배함



<그림 2-132> 버퍼링 프로세스 Work-flow

㉰ 프리패칭(Pre-fetching)

○ 한꺼번에 많은 량의 데이터를 지도위에 표출하는 것은 서비스의 속도나 사용자의 접근성 하락을 유발하므로, 한 번에 지도위에 구현되는 소셜 미디어 정보를 100개로 제한하여 서버에서 클라이언트로 전송함

⑤ [단위성과 E-7-①] 공간 빅데이터 정합 기술(3차연도)

○ 소셜 미디어 데이터와 비정형 데이터에서 장소 정보(행정구역명, POI(Point of Interests)명)와 위치 정보(위 · 경도)를 추출하는 정합 기술 개발

㉮ 사용자 생성 콘텐츠 정의 및 정확도 분석

○ 사용자 생성 콘텐츠: 트위터, 유투브, 웹 문서(네이버, 뉴스 등)에서 사용자가 직접 입력한 생성된 텍스트

<표 2-28> 정의된 사용자 생성 콘텐츠

대상 데이터	필드명	필드 설명
트위터	text	UTF-8 표준 언어코드로 된 트윗 데이터의 텍스트 정보
유투브	snippet	동영상의 제목, 설명, 카테고리 등 동영상에 대한 기본 세부정보
	snippet.description	동영상에 대한 설명
웹 문서	-	제목, 글 내용

○ 이들 텍스트는 다시 유형화된 텍스트와 비유형화된 텍스트로 구분 가능

<표 2-29> 사용자 생성 콘텐츠의 구분

텍스트 구분	설명	사례
유형화된 텍스트	'@', 'at', 'in'을 포함한 텍스트	I'm at 원조 질정소곰탕 https://t.co/ixfnTQkmhV http://t.co/ts1abctOIL Graduation Day ??? (with neysa Rimas and 6 others at 성균관대학교 (Sungkyunkwan Univ.)) ? https:// I'm at 다음카카오 (daumkakao) in 성남시 경기도 https://t.co/MJK6WNscbe http://t.co/6clZgGNQAV
비유형화된 텍스트	그 외 텍스트	습기많고 무더운 수요일 아일리아헤어 인덕원점도 기분좋은 하루를 시작합니다~^^ 아일리아헤어는 고객의 건강한 아름다움을 위해 정성을... https://t.co/TQS6HH822d

○ 사용자 생성 콘텐츠에 대한 정확도 분석

- 2015-07-22에 수집된 지오태깅된 트윗 데이터 797건에 대해 수동으로 200개의 장소 정보(행정구역명, POI명)를 추출한 후 구글 Maps API를 이용하여 위치 정보(위 · 경도)를 추출하여 지오태깅된 위치 정보와 장소 정보로 추출된 위치 정보의 거리차이로 평가
- 154개의 장소 정보에 대해서 위치 정보를 취득하였으며, 지오태깅된 데이터와의 거리차이에 대한 상자도표를 그려보면 평균(62.45m)이하에 편중되어 분포됨
- 이들 위치 정보에서 최대값을 임계치로 추정. 즉, 지오태깅된 위치 정보와 장소 정보에서 추출된 위치 정보의 거리차이가 645m보다 큰 경우는 이상치로 판별함
- 그 결과, 전체 데이터 중에서 154개 중에서 45개가 이상치로 분석되어 약 71%의 데이터가 지오태깅된 위치 정보와 사용자 생성 콘텐츠에서 분석된 위치 정보가 유의미한 거리차이로 판별되어, 사용자 생성 콘텐츠에 입력된 장소 정보를 이용하여 위치 정보를 추출하는 것이 유의미한 것임을 확인함

<표 2-30> 사용자 생성 콘텐츠에서 추출된 위치 정보의 정확도 분석 결과

평균	거리차이 최대값	이상치 (거리차이 > 695m)
62.450m	695m	45개(29%)

㈏ 정합 기술

○ 텍스트에 지오태깅된 데이터가 있는지 확인한 후 없는 경우 텍스트 처리를 진행함

○ 텍스트 처리 시 해당 텍스트가 유형화 되어있는지 체크한 후 유형화된 경우 POI를 추출하고 그렇지 않은 경우 행정구역 DB를 이용하여 해당 텍스트와 시맨틱 매칭을 진행하고 행정구역으로 이루어진 장소 정보를 추출함

○ 최종적으로 얻어낸 정보를 동시에 이용하여 장소 정보를 병합하고 정규화 함

- 유형화된 텍스트 예시) 소화 잘 되는 햄에그치아바타 (@ 김포국제공항 국내선 (GMP - Domestic Terminal) - @kacgimpo in 서울특별시) https://t.co/WRn432vQ91
- 유형화된 텍스트처리를 통해 얻을 수 있는 장소정보: 김포국제공항 국내선

- 시맨틱 매칭을 통해 얻을 수 있는 장소 정보 : 서울특별시
- 두 정보를 융합하여 최종적으로 ‘서울특별시 김포국제공항 국내선’으로 정규화

○ 정규화된 장소 정보를 키워드로 구글 Maps API를 이용하여 위치 정보(위 · 경도)를 추출

- 응답이 2개 이상인 경우 해당 텍스트에 대한 위치 정보라고 보기 어렵기 때문에 다중반환 항목을 만들어 ‘1’로 표시하고 구글API에서 제공하는 모든 위치 정보를 저장하여 사용자에 제공함

<표 2-31> 사용자 생성 콘텐츠로부터 정합 기술로 추출된 위치 정보

장소정보	도로명주소	위치 정보	다중반환
충청남도 공주시	대한민국 충청남도 공주시	36.446535, 127.119153	0
대전광역시 충남대학교	대한민국 대전광역시 유성구 대학로 99	36.369987, 127.345955	0
RABBBT LOUNGE	대한민국 서울특별시 마포구 연남동 567-37	37.560609, 126.922597	1
	대한민국 서울특별시 서초구 반포동 105-8	37.498190, 126.997552	

⑥ [단위성과 E-7-②] 공간 빅데이터 주제도 분석 기술(3차연도)

○ 소셜 미디어 데이터와 비정형 데이터에서 장소 선호도 기반, 공간 분석(spatial analysis) 기반, 소셜 네트워크 분석(social network analysis, SNA) 기반 주제도 생성을 위한 기술 개발

㉮ 장소별 선호도 기반 주제도 분석

○ 공간 감성어 사전을 이용하여 장소에 대한 감성을 분석하고, 장소에 대한 언급빈도수를 통합하여 장소 선호도 기반의 주제도 분석 기술

[STEP 1] 공간 감성어 사전 구축

○ 네이버 검색 API에서 ‘서울시 테마공원’으로 검색했을 때 도출되는 POI에 대해 구글 맵에서 작성된 장소 리뷰글을 이용하여 장소에 대한 감성을 표현하는 어휘를 제공하는 공간 감성어 사전을 구축함

○ 구축된 공간 감성어 사전(SSLex_single, SSLex_combi)은 SSLex_single 테이블은 총 112개 단어로 구성되며 SSLex_combi 테이블은 104개 조합으로 구성됨

- SSLex_single 테이블은 독립적으로 하나의 단어를 구성하는 형용사, 동사, 명사로 구성되며, SSLex_combi 테이블은 공간 감성어 조합({명사, ‘많다’/‘적다’}, {명사, 형용사}, {명사, 동사})으로 구성됨

<표 2-32> 공간 감성어 사전 구축 결과

테이블명	품사	긍정	부정	중립	합계
SSLex_single	형용사	24	13	8	45

테이블명	품사	긍정	부정	중립	합계
	동사	6	1	6	13
	명사	36	10	8	54
소계		66	24	22	112
SSLex_combi	명사+'많다', '적다'	13	13	22	48
	명사+형용사	44	4	3	51
	명사+동사	3	0	2	5
소계		60	17	27	104
합계		126	41	49	216

[STEP 2] 장소별 선호도 분석

○ SNS 데이터로부터 장소를 포함한 텍스트 추출, 각 장소에 대한 언급빈도수 산출, 공간 감성어 사전을 이용한 감성분석, 장소 선호도 산출로 이루어짐

○ 공간 감성어 사전을 이용한 감성분석

- 텍스트별로 추출된 해당 감성어의 감성 극성과 확률을 곱한 값을 긍정과 부정으로 구분하여 누적평균하고, 이를 해당 텍스트의 긍정값(PT)과 부정값(NT)으로 정의

<표 2-33> 감성지수 정의

	장소 POI에 대한 감성지수(S)	텍스트의 감성값(T)	비고
긍정 (Positive)	$S_P = \sum_{i=1}^{F_k} PT_{(i)}$	$PT_{(i)} = \frac{\sum_{j=1}^{n}\{match(i,j) \times Pb_{(j)}\}}{\sum_{j=1}^{n} match(i,j)}$	$match(i,j) = \begin{cases} 1 & (\text{텍스트 } i\text{에 포함된 } j\text{(형용사,동사,명사)가 공간감성어사전에 존재하고 '장소 POI/공간주제어'와 3어절 이내일 경우}) \\ 0 & (\text{그 외의 경우}) \end{cases}$
부정 (Negative)	$S_N = \sum_{i=1}^{F_k} NT_{(i)}$	$NT_{(i)} = -\frac{\sum_{j=1}^{n}\{match(i,j) \times Pb_{(j)}\}}{\sum_{j=1}^{n} match(i,j)}$	

○ 장소에 대한 언급빈도수와 감성분석 결과를 고려한 장소 선호도 산출

$$\Pr eference_{(k)} = SD \times (1 + |S_P + S_N|) \times F_k$$

- 여기서, k는 장소 POI이며, SD(Sentiment Direction)는 감성방향으로 긍정($S_P > |S_N|$)일 경우 +1, 부정($S_P < |S_N|$)일 경우 -1을 가짐. S_P (Sentiment index of Positive)는 장소 POI에 대한 긍정지수, S_N (Sentiment index of Negative)은 장소 POI에 대한 부정지수, F_k (Frequency of k)는 장소 POI에 대한 언급빈도수를 의미

㉯ 공간 분석 기반 주제도 분석

○ 웹 문서 중 네이버 카페글에서 주제어를 선정하여, 해당 주제어의 언급빈도수를 이용하여 공간분석 중 공간자기상관 분석을 통해 주제도를 분석하는 기술

○ 자연어 처리 및 위치 정보 추출

- RL 제거, 영문 제거, 공백 제거 등의 자연어 처리를 수행한 후, 앞에서 개발된 정합 기술을 적용하여 행정구역 시군구 단위의 위치 정보를 추출

○ 문서 분류 과정

- 형태소 분석 및 품사 부착 등의 전처리 과정, 분석에 이용될 자질을 선정하는 자질 선정(feature selection), 자질 가중치로 자질을 표현하는 문서 표현(indexing), 데이터를 훈련 데이터(training data)와 테스트 데이터(test data)로 나누고, 훈련 데이터를 통해 SVMs(Support Vector Machines)을 트레이닝 하는 과정, 트레이닝 된 SVMs를 통해 문서를 분류하는 과정으로 이루어짐

- SVMs의 매개변수(parameter)로 튜닝을 통해 cost는 6, gamma는 0.1로 설정함

○ 공간 분석

- 시군구 단위로 주제어에 대한 언급빈도가 구한 후 공간자기상관 분석을 통해 주제도를 생성함. 공간자기상관 분석으로는 Moran’s I, Local Moran’s I, Getis-ord Gi*가 있음

<표 2-34> 문서 분류 결과

	카페이름	글제목	날짜	글 내용	자연어처리 결과	코드	시군구명	경도	위도	SVMs
1	·˙°³o 경주아이맘·˙°³o	헐~~미쳤나봐요ㅠ	2013-12-06	재워야하는데ㅠ 진짜 개념을 난지도에 갖다버린 집구석들이 와이리 많은지ㅠ 그런 인간들은 아파트에 못 살도록 했음 좋겠네요ㅠ 우찌 그래 지밖에 모르는지.... 에휴...층간소음없는 곳으로 가고싶네요~~~~ㅠ	재워야하는데 진짜 개념을 난지도에 갖다버린 집구석들이 와이리 많은지 그런 인간들은 아파트에 못 살도록 했음 좋겠네요 우찌 그래 지밖에 모르는지 에휴 층간소음없는 곳으로 가고싶네요	47130	경상북도 경주시	129.2359	35.826669	1
2	별내클로버 (별내,구리 갈매국민임...	층간소음 해결방안...!!!	2015-12-02	층간소음 해결방안에 대하여.. 별내 클로버 회원님들 잘 지내고 계시지요.보다나은 생활을 위해 층간소음의 해결책에 대하여 방법을 제시하오니 참고하시기를 바랍니다.별...	층간소음 해결방안에 대하여 별내 클로버 회원님들 잘 지내고 계시지요 보다나은 생활을 위해 층간소음의 해결책에 대하여 방법을 제시하오니 참고하시기를 바랍니다 별	41310	경기도 구리시	127.13112	37.5992 3	-1

㈐ SNA 기반 주제도 분석

○ SNS 데이터에서 키워드별 사용자(ID) 분류, 노드와 링크의 강도를 충분히 반영하는 가중 네트워크 생성, key player를 탐지함

○ 가중 네트워크 생성

$$WKC(A,B) = \frac{\in vest(B,A)}{\sum \in vest(B,X)} \times \frac{n(cokeywords(A,B))}{n(keywords(B))} \times n(followers(A))$$

- 여기서, $\in vest(B,A)$ = A와 B의 연결정도(B가 A에게 투자한 정도)
- $\sum \in vest(B,X)$ = B와 다른 사용자들(X) 간의 연결정도
- $n(cokeywords(A,B))$ = A와 B가 함께 언급한 키워드 수
- $n(keywords(B))$ = B가 언급한 키워드 수
- $n(followers(A))$ = A의 follower 수

❍ 가중 연결정도 중심성, 가중 근접 중심성, 가중 매개 중심성을 분석하여 그 값이 큰 사용자를 특정 키워드에 대한 key player로 탐지함

⑦ [단위성과 E-8, E-10] 공간 빅데이터 분포 분석 기술(4,5차연도)

❍ 이모티콘 기반 장소 선호도 분석 기술

- 긍정 및 부정 이모티콘 리스트 생성하고, 구글 장소 리뷰를 이용하여 해당 이모티콘과 관련된 긍정 및 부정 감성어를 추가하여 추가 감성어DB 구축

<표 2-35> 추가 감성어 DB 구축 결과

구분	긍정	부정
이모티콘	9	9
추가 감성어	369	336

- 추가 감성어 DB로 장소의 선호도를 분석하기 위하여 이모티콘과 추가 감성어를 이용하여 4가지 CASE로 구분하여 감성값 계산

<표 2-36> 감성어와 이모티콘이 함께 사용된 경우 감성 판단

CASE	감성어(w)	이모티콘(e)	최종 감성
1	긍정	긍정	긍정
2	긍정	부정	긍정
3	부정	긍정	부정
4	부정	부정	부정

<표 2-37> 감성지수벡터 및 감성값 정의

	장소 POI(k)에 대한 감성지수벡터($\overrightarrow{S}$)	텍스트(i)의 감성값(T)
긍정 (Positive)	$\overrightarrow{S_P} = \sum_{i=1}^{F_k} PT(i)$	$PT(i) = \sum_{j=1}^{n} \{match(i,j) \times Pb(j)\} + \sum_{e=1}^{m} match(i,e,w)$
부정 (Negative)	$\overrightarrow{S_N} = \sum_{i=1}^{F_k} NT(i)$	$NT(i) = -\sum_{j=1}^{n} \{match(i,j) \times Pb(j)\} - \sum_{e=1}^{m} match(i,e,w)$

- 여기서, i는 트위터 텍스트, k는 POI, F_k는 k의 언급빈도수, $Pb(j)$는 감성어 j의 감성확률, n은 i에 존재하는 j의 개수를 의미. $match(i,j)$는 텍스트 i에 포함된 j가 공간 감성어 사전에 존재하고, POI, 지시대명사, 공간 주제어와의 거리가 3어절 이내일 경우 1, 그 외의 경우 0을 갖음
- e는 긍정 및 부정 이모티콘, m은 i에 존재하는 e의 개수, w는 공간 감성어 사전 및 추가 감성어 DB에 저장된 감성어를 의미. $match(i,e,w)$는 텍스트 i에 포함된 e가 w와 함께 사용되었을 경우 1, 그 외의 경우 0을 갖음
- 각 장소 POI에 대한 언급빈도수와 감성분석 결과를 모두 고려하여 장소 선호도를 계량화하기 위하여 장소 선호도 산출식 제안. 1차 선호도를 min-max로 정규화한 값이 장소 선호도

$$P(k)_{1st} = SO \times SS \times W(F_k)$$

- 여기서, 감성방향(SO)은 POI k의 긍정지수벡터($\overrightarrow{S_P}$)와 부정지수벡터($\overrightarrow{S_N}$)의 크기비교를 통해 긍정(+1) 또는 부정(-1)으로 결정

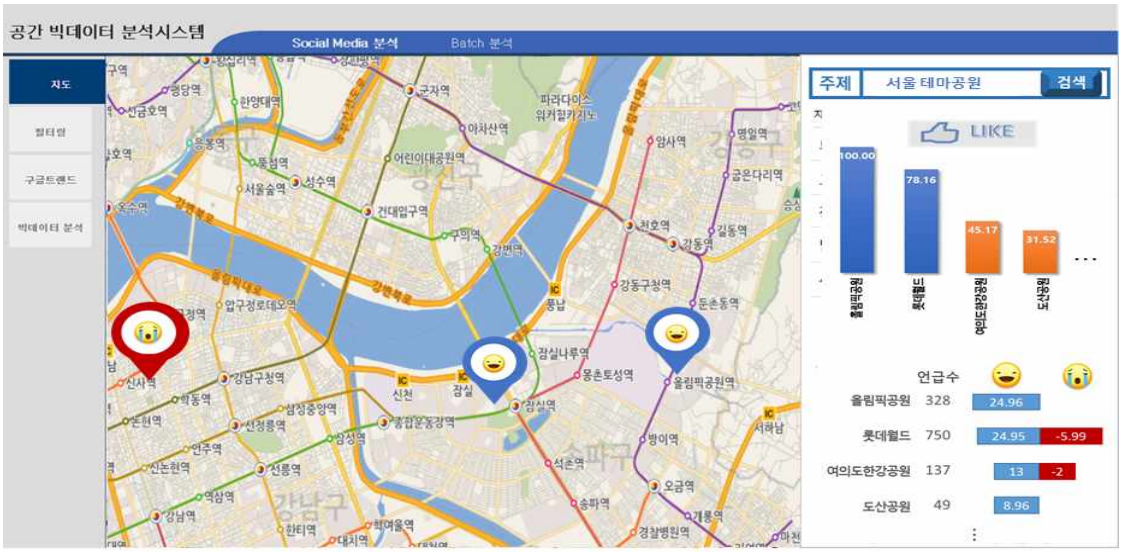
$$SO = \begin{cases} +1, \| \overrightarrow{S_P} \| \geq \| \overrightarrow{S_N} \| \\ -1, \| \overrightarrow{S_P} \| < \| \overrightarrow{S_N} \| \end{cases}$$

- 감성점수(SS)는 POI k의 긍정지수벡터와 부정지수벡터의 합벡터의 크기에 1을 더한 값

$$SS = 1 + \| \overrightarrow{S_P} + \overrightarrow{S_N} \|$$

- 각 POI k의 언급빈도수에 따른 가중치 $W(F_k)$

$$W(F_k) = 2 - 2^{-(F_k/wf)}$$



<그림 2-133> 감성 분석기반 장소선호도 분석 분포 및 차트 서비스 사례

❍ 트위터 데이터의 GPS 정보 추정

- GPS 정보를 가지고 있지 않은 트위터들을 공간분석에 활용하기 위해 사용자 주거정보, 텍스트 정보, 시군구 수준의 행정구역 정보를 이용하여 트위터의 GPS 정보를 추정하는 방법을 제시
- GPS 정보를 가지고 있는 트위터는 전체의 1% 미만이지만 시군구 수준의 행정구역 정보를 모두 포함하면 위치정보를 가지고 있는 트위터는 전체의 약 3% 정도이며 시군구 수준의 행정구역 정보를 GPS 정보로 변환
- GPS 정보를 가진 트위터를 시군구 수준의 행정구역을 기준으로 분류한 뒤 해당 행정구역 내에 발생한 트위터들을 클러스터링 하고 가장 밀집도가 높은 클러스터의 중심점을 해당 행정구역의 대표 좌표로 결정

<표 2-38> 데이터 처리 결과

위도	경도	시군구	텍스트	ID	time	class	POI
37.44934	126.45082	Ongjin-gun	... 인천국제공항 ...	6.95E+17	Mon Feb 01 12:53:48 +0000 2016	GPS	NA
35.97848	126.71499	Gunsan-si	... @ZerialLim: RT@Cloveraid백제에서는 ...	6.95E+17	Wed Feb 03 23:57:12 +0000 2016	sigungu	NA
NA	NA	Gwangjin-gu	... @kimkasu 원준씨가 여기에 ...	6.95E+17	Wed Feb 03 20:09:49 +0000 2016	home to sigungu	NA

위도	경도	시군구	텍스트	ID	time	class	POI
NA	NA	NA	... I'm at SK하이닉스(SK hynix) in 이천시 경기도 ...	6.95E+17	Fri Feb 05 15:13:03 +0000 2016	poi	경기도 이천시 SK하이닉스

- 트위터 데이터를 class에 따라 분류하여 GPS는 클러스터링에 사용하여 시군구 수준의 행정구역마다 대표좌표를 추정하였고, sigungu에는 클러스터링 결과 도출한 대표좌표들을 각각 부여. home to sigungu는 주거정보에서 추출한 시군구 정보이며 sigungu로 분류된 트위터들과 마찬가지고 대표좌표들을 부여. POI 추출결과는 따로 이용하지 않았으며 API를 통해 좌표만 부여함



<그림 2-134> GPS 정보 추정 분포 및 차트 서비스

❍ 개체명인식을 활용한 비정형 텍스트 데이터 위치탐지 기술

- 비정형 텍스트 데이터가 묘사하는 위치를 탐지하여 소속 행정구역을 읍면동 단위까지 추출하여 좌표정보를 부여함
- 분석 데이터는 2017년 1월부터 2018년 1월에 수집된 3,241개의 화재관련 기사문장들과 2014년 7월부터 2017년 11월까지 수집된 86,133건의 도봉구 민원데이터를 대상으로 함
- 장소정보를 추출하기 위해 한국어 문장 또는 문서를 입력받아 분석하는 Exobrain 언어분석 API를 활용하여 개체명을 추출함
- 개체명 중 장소정보를 활용하기 위해 행정구역을 탐지할 “축약어를 포함한 법정동 Reference”를 구축하였고, 도로명을 탐지하기 위해 “도로명 Reference”를 구축하였음
- 장소정보 간 위계를 고려하여 정리한 뒤, 지오코더를 활용하여 화재관련 뉴스데이터와 도봉구 민원데이터를 입력하여 각 텍스트에 대한 위치를 탐지할 수 있었음

<표 2-39> 탐지된 위치에 대한 좌표정보 추출 결과 일부(뉴스데이터)

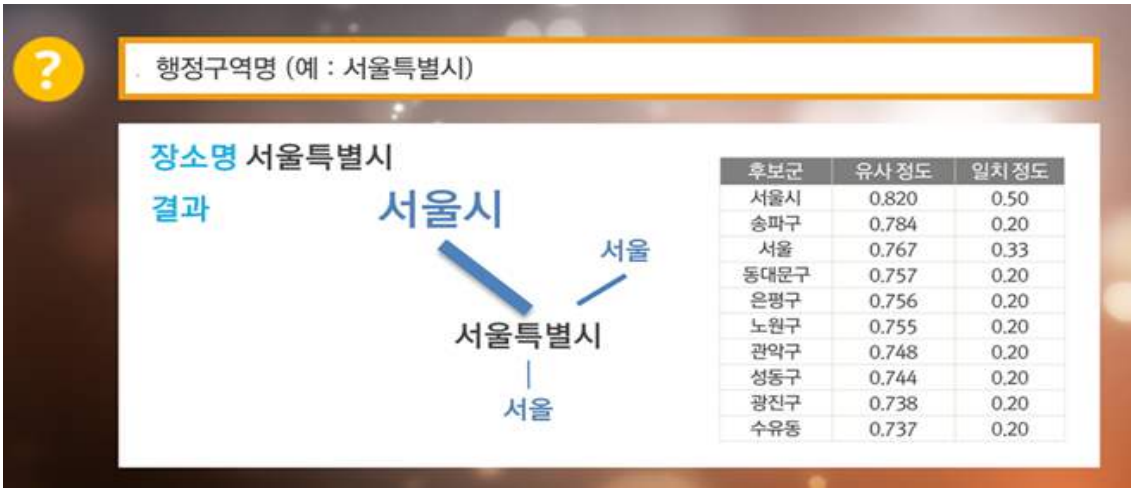
	text	탐지된 위치정보	좌표정보	
			위도(Lat)	경도(Lng)
1	밀양 세종병원 화재 뉴스를 접한 최승동(54)씨는 요즘 불안하다	경상남도 밀양시 세종병원	35.4761	128.7661
2	서울 은평구 불광동 미성아파트 화재로 의식을 잃은 채 병원에 실려갔던 일가족 3명이 모두 사망했다	서울특별시 은평구 불광동 미성아파트	37.6147	126.9281
3	"다중이용 화재취약시설 전수조사"문재인 대통령은 29일 밀양 세종병원 화재사고에 대한 정치권의 책임공방과 관련 "이 책임에서 자유로운 사람은 아무도 없을 것"이라고 말했다	경상남도 밀양시 세종병원	35.4761	128.7661

❍ 딥러닝 기반 장소명 대체어 분석 기술

- Word2vec을 통해 워드 임베딩 공간에서 비슷한 문맥을 가진 단어들이 서로 가까운 공간 분포를 가지게 되므로 특정 장소명의 대체어를 추출
- 1GB정도의 한국어 위키피디아를 Word2vec을 통해 학습하고, 특정 장소명과의 코사인 유사도(Cosine Similarity, CS)를 계산하여 상위 10개의 단어를 대체어 후보군으로 추출
- 추출된 대체어 후보군과 특정 장소명과의 편집거리(Edit Distance, ED)를 구하여 그 값이 임계값 이하인 경우 특정 장소명의 대체어로 탐지

<표 2-40> 장소명 대체어 분석 결과 (서울특별시, 부산광역시 사례)

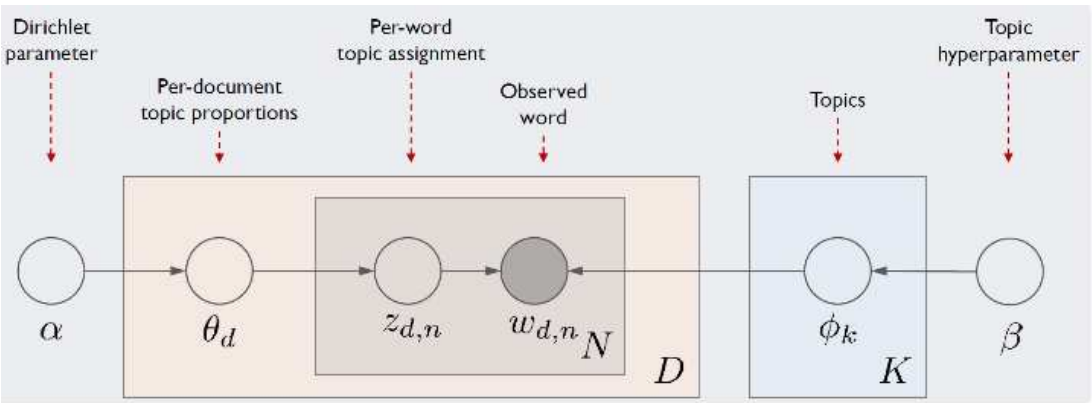
서울특별시			부산광역시		
후보군	CS	ED	후보군	CS	ED
서울시	0.820	2	부산진구	0.781	3
송파구	0.784	5	해운대구	0.750	5
서울	0.767	3	대구광역시	0.742	2
동대문구	0.757	5	사하구	0.741	5
은평구	0.756	5	부산	0.739	3
노원구	0.755	5	수영구	0.735	5
관악구	0.748	5	부산시	0.733	2
성동구	0.744	5	동래구	0.729	5
광진구	0.738	5	인천광역시	0.726	2
수유동	0.737	5	연산동	0.722	4



<그림 2-135> 장소명 대체어 분석 분포 및 차트 서비스 사례

❍ 공공시설 매력도 측정 및 접근성 분석 기술

- 공공시설의 매력도를 서비스를 이용하는 이용자 관점에서 측정하기 위해 SNS 데이터인 블로그글을 수집하여 공공시설 선호요인을 도출하고, 이를 지수화시켜 공공시설 매력도를 산출하였음
- 공공시설 방문관련 네이버블로글을 수집하여 전처리를 위해 NLP을 사용하였으며, 대용량 문서인 블로그의 특징을 고려하여 방문관련 적합문서를 분류하기 위해 SVM을 적용하였음, 추출된 문서를 바탕으로 토픽모델링인 LDA(Latent Dirichlet Allocation) 를 적용하여 공공시설 선호요인을 도출하였음



<그림 2-136> LDA 문서생성 과정 (D: 말뭉치 전체 문서 개수, K: 전체 토픽수, N: d번째 문서의 단어 수)

- 공공시설 매력도지수 산출을 위해 각 선호요인별 지표를 구축하여 표준화 및 가중치부여를 실시하였으며, 이때 관련 통계자료가 없는 경우를 고려하여 비정형 데이터를 통한 지표구축방법을 추가로 제시하였음
- 최종적으로 공공시설 매력도지수를 공급요인으로 고려한 공간접근성을 측정하여 이용자입장에서 지역별로 어떻게 나타나는지를 분석하였음
- 공간접근성 분석 모델 중 2SFCA모델에 공공시설 매력도지수를 공급요인으로 적용하였으며, 실세계 반영을 위해 도로네트워크를 사용하여 임계거리를 설정하였음



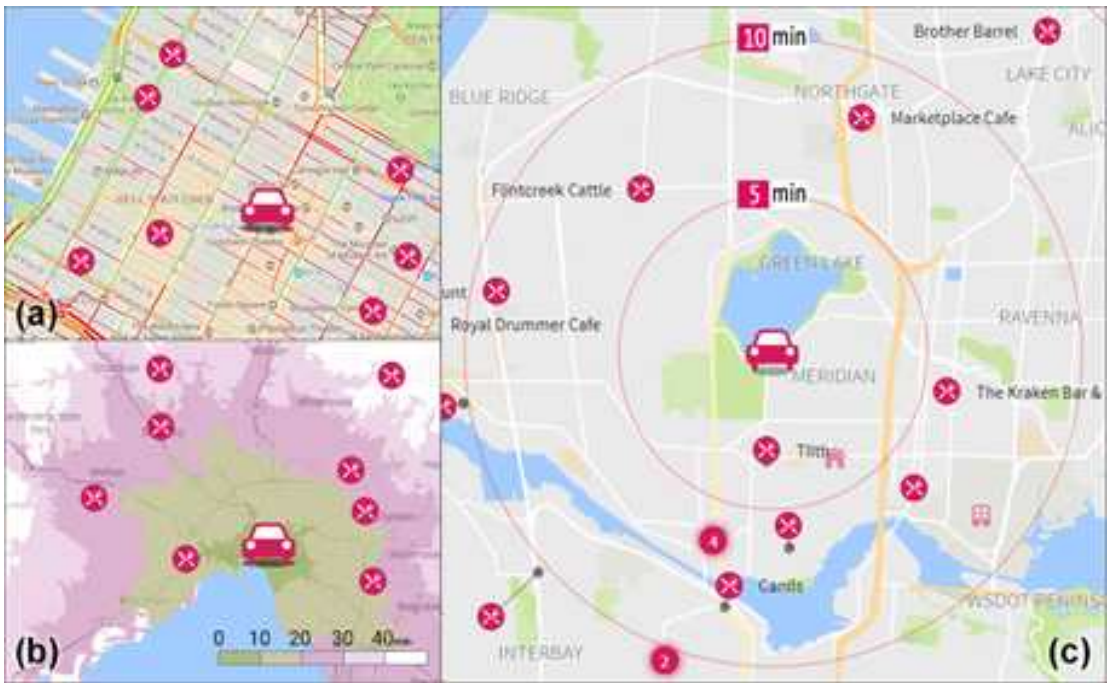
<그림 2-137> 공공시설 매력도 및 접근성 분석 차트 및 분포서비스

<표 2-41> 공공도서관 매력도지수 결과

도서관명	주변 환경	이용편리성	문화 프로그램	육아	장서 및 자료현황	매력도 지수
송파	0.099	0.182	0.132	0.056	0.146	0.614
중랑구립	0.000	0.000	0.035	0.026	0.000	0.061
양천	0.048	0.108	0.205	0.025	0.093	0.478
마포평생	0.043	0.114	0.083	0.160	0.070	0.470
정독	0.212	0.119	0.116	0.047	0.091	0.585
성동	0.025	0.120	0.028	0.069	0.179	0.422
노원	0.046	0.030	0.049	0.013	0.096	0.234
강북	0.037	0.015	0.022	0.036	0.156	0.266
노원평생	0.034	0.022	0.078	0.184	0.021	0.338
광진	0.048	0.008	0.094	0.065	0.190	0.405
반포	0.035	0.154	0.010	0.037	0.101	0.337
남산	0.107	0.109	0.153	0.039	0.065	0.472
고척	0.039	0.062	0.046	0.027	0.185	0.359
종로	0.055	0.068	0.019	0.016	0.217	0.377
서울	0.082	0.096	0.000	0.000	0.180	0.359

❍ 사용자 중점 기반 시간지도 생성 기술

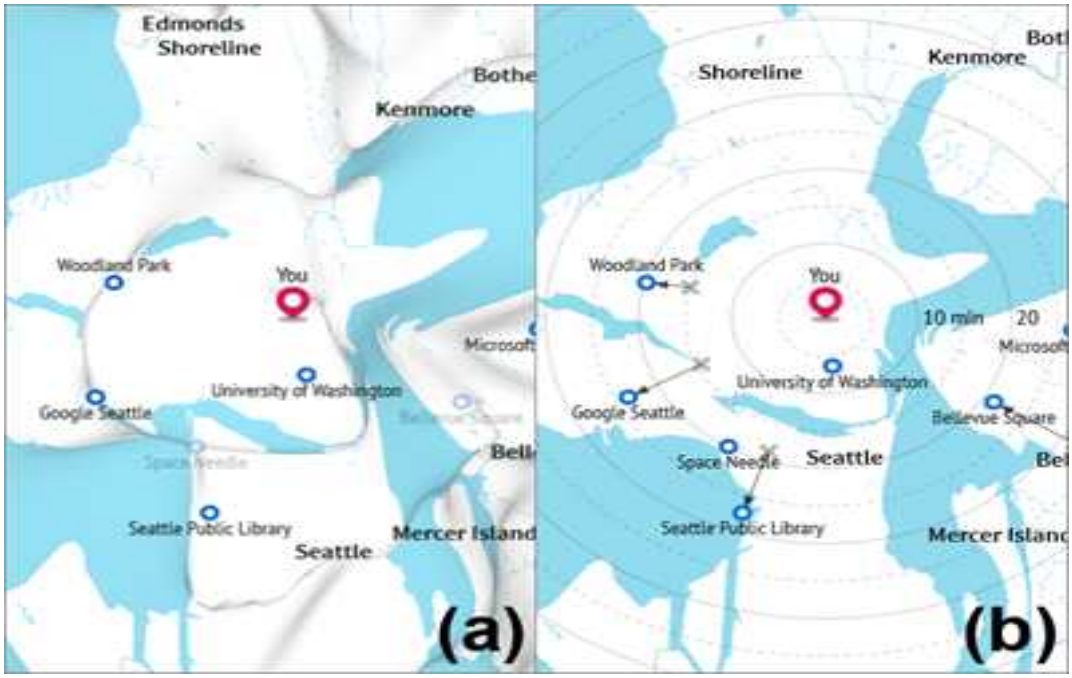
- 사용자 중점기반 시간지도 (Distance Cartogram, 이하 DC로 표기함) 는 사용자의 현재 지점을 중점으로, 현재 지점과 그 지점 이외의 지도상에 표시된 위치들까지의 이동시간을 기반으로 지도상의 지점들을 중점과 가깝거나 멀게 위치를 변위시켜줌으로서 거리 (position)라는 시각채널을 활용하여 양적 정보인 시간정보를 표시함
- 이러한 DC의 거리기반 시간정보 표시는 기존의 색상을 이용한 시간정보 표시보다 (그림 (a),(b) 참조) 현저히 높은 정밀도로 사용자가 정보를 인지할 수 있는 장점이 있음



<그림 2-138> 시간정보를 표시하는 방법의 예: (a) color-coded road segments, (b) free-from color-coded regions, (c) Distance Cartograms

- Quadtree Timespace Partitioning (QTP)는 쿼드트리 공간분할이라는 기술을 활용하여 중점에서부터 지도상에 존재하는 모든 점 까지의 최소이동시간을 O(nlogn) 복잡도로 실제 트레픽 데이터를 기반으로 연산하게 함으로서 DC를1초 안에 사용자에게 제공할 수 있음. QTP를 활용하여 인터랙티브한 DC의 제공이 가능함
- 시간지도를 생성하는 과정에서 제작되어야 하는 중점과 지도상의 모든 포인트들까지의 시간거리가 매핑되어 있는 “시간공간”은 불연속적임. 예를 들면 지도상에 물리적으로 두 점이 서로 가까이 있다고 해도 그 두 점으로 이동하는데 소요되는 이동시간은 두 점간의 물리적인 근접성에도 불구하고 크게 달라질 수 있음
- 이러한 불연속성은 2D 공간에 “공간적인 겹침”없이 지도를 사영하는 것을 불가능하게 하고, 한 편으로 인하여 물리적으로 가까운 두 위치도 DC에서는 멀게 보일 수 있게 하는 이른바 “위상침해” 를 통한 사용자적 관점에서의 인지적 난관을 제공함. 또한 위상침해 이외에도 일정 수준 이상으로 지도상의 지리적 윤곽들이 왜곡되는 경우 사용자가 원래 윤곽을 이해할 수 없게 될 때 DC의 사용성은 크게 떨어짐

- Shape-retaining Geo-Contextual Anchoring Projection (S-GAP)은 이러한 두 인지적 문제를 해결하기 위한 기술로, 위상침해가 없이 DC를 제작하게 하고, 다른 한 편으로 일정 수준 이상의 윤곽 왜곡이 일어나지 않게 함
- 본 연구를 통해 사용자들은 DC를 활용함으로써 목적지까지 걸리는 시간을 빠른 시간 안에 정확하게 인지할 수 있는 장점이 있음



<그림 2-139> (a) 시간공간의 불연속성으로 인한 공간겹침의 예, (b) S-GAP 을 통한 평면상에 공간겹침 없이 DC를 제작하는 예시

⑧ [단위성과E-9, E-11] 공간 빅데이터 차트 분석 기술(4,5차연도)

❍ 머신러닝 기반 장소 감성어 사전구축

- 사용자가 영어로 작성한 장소 리뷰 데이터에 대해 머신러닝 기법(Word2Vec, 연관규칙 마이닝, bag-of-words 로지스틱 회귀분석)을 활용하여 장소의 특성과 이를 서술하는 감성어휘를 추출하고, 장소 감성어 조합을 생성하여 이에 대한 감성점수를 계산, 장소 감성어 사전을 구축하는 방법 제시
- 장소 감성어 사전을 구축하는 방법은 크게 2가지로 나뉨: 1) 사전에 정해진 장소의 특성을 이용하는 정적인(static) 방법, 2) 데이터로부터 장소의 특성을 새롭게 추출하는 동적인(dynamic) 방법
- 다양한 장소 카테고리 중 'Italian Restaurant'에 대해 총 317개의 장소 감성어 조합을 생성, 각각에 대한 감성점수를 계산함. 이 중에서 정적 방법으로 구축한 장소 감성어 조합이 156개, 동적 방법으로 구축한 조합이 191개임

<표 2-42> 정적 방법으로 구축한 장소 감성어 사전 구축 결과(일부)

ID	PF	PSW	SC
Static_1	ambiance_NN	romantic_JJ	0.964626
Static_20	decor_NN	gorgeous_JJ	0.966204
Static_37	food_NN	delicious_JJ fresh_JJ	0.992520
Static_104	food_NN	horrible_JJ	0.216304
...	...	...	...

<표 2-43> 동적 방법으로 구축한 장소 감성어 사전 구축 결과(일부)

ID	PF	PSW	SC
Dynamic_5	anything_NN	special_JJ	0.869838
Dynamic_27	bread_NN	soft_JJ	0.932412
Dynamic_31	burger_NN	good_JJ	0.923305
Dynamic_43	crust_NN	crispy_JJ	0.948817
...	...	...	...

- 정확도 검증 결과, 정밀도(precision), 재현율(recall) accuracy, F1 값 모두 정적 방법으로 구축한 장소 감성어 사전에서 더욱 높게 나타남

<표 2-44> 장소 감성어 사전에 대한 정확도 검증 결과

구분	정적 방법	동적 방법
precision	0.85	0.52
recall	0.92	0.63
accuracy	0.84	0.51
F1	0.88	0.57



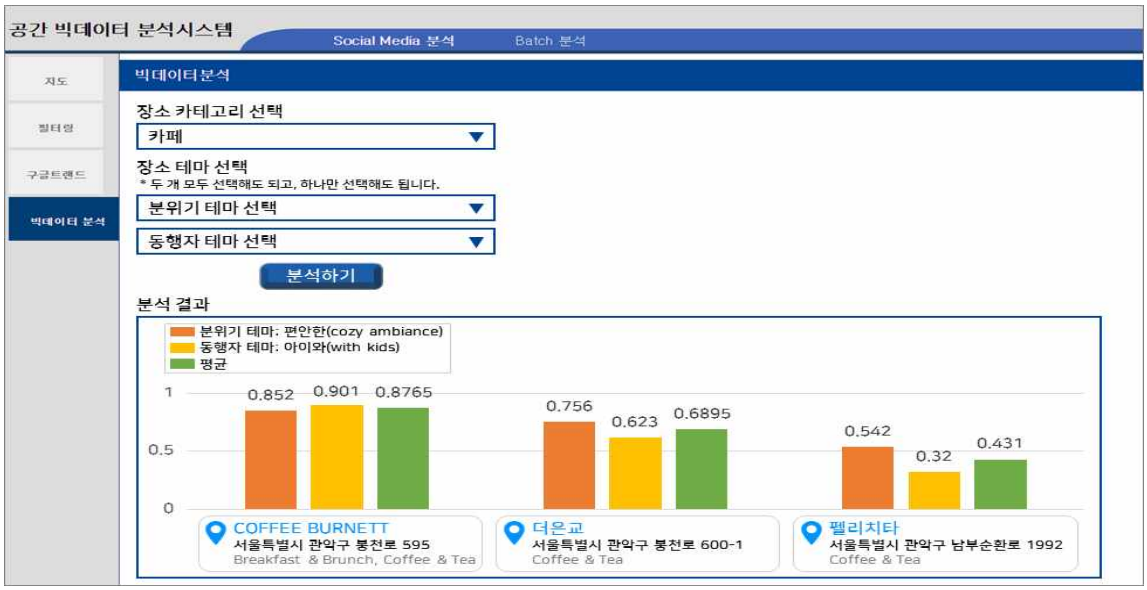
<그림 2-140> 장소 감성어 사전 분포 및 차트 서비스 사례

❍ 영문 장소리뷰로부터 장소 테마 분석 기술

- 임베딩 기술인 Word2Vec은 유사한 문맥의 단어들을 벡터 공간에서 가까운 거리에 위치시키므로 이를 이용해 단어 간 연관도 계산 가능
- 즉, 장소별 리뷰 데이터를 각각 임베딩한 후 장소별로, 장소 테마를 나타내는 주어(테마 단어)와 이를 설명하는 서술어(테마 표현 단어) 간 코사인 유사도를 계산함으로써 장소별 세부 테마 연관도를 분석
- Yelp 데이터셋에서 카페 관련 카테고리 장소 중 리뷰 개수가 많은 상위 10개 장소를 실험 대상으로 하여, 각 장소별로 작성된 영문 리뷰 텍스트에 Word2Vec을 적용해 장소별 임베딩을 수행
- 장소별 임베딩 결과에서, 분위기(ambiance), 동행자(companion), 시간(time) 테마별로 테마 단어(theme word, tw) 및 테마 표현 단어(theme representation word, trw)와 매칭되는 단어를 추출, 이에 대한 연관도를 계산
- 연관도 값이 계산되지 않은 경우는 그 장소에서 해당 테마 조합(테마 단어 + 테마 표현 단어)에 대한 사용자 관심도가 떨어진다는 것을 의미하고, 연관도 값이 클수록 그 장소가 해당 테마 조합에 상대적으로 가장 적합한 장소라는 것을 의미함

<표 2-45> 장소 테마 분석 결과(일부)

theme (tw)	trw	Lux Central	Amélie's French Bakery & Café	Baguette Cafe	Liberty Market	Jean Philippe Patisserie	Kung Fu Tea	Grouchy John's Coffee	Sunrise Coffee	Snooze An AM Eatery	Max Brenner
ambiance (ambiance/ ambience/ atmospher/ environment/ decor/vibe/ cor/interior/ deco/setting/ layout/setup)	clean	0.4806	0.4278	0.5686	0.5765	0.3024	0.6008	0.5071	0.5935	0.5191	0.3741
	comfortable	0.3753	0.4535	_	0.4273	_	_	0.3953	0.4839	0.4665	_
	cozy	0.5198	0.6653	0.3280	_	_	_	0.4669	0.4865	_	_
	dark	0.3538	0.2236	_	_	0.1983	_	0.1892	0.2150	_	0.3276
	pleasant	0.3371	0.3685	0.5112	0.4128	_	_	0.4217	0.2763	0.2888	0.4465
	simple	0.4168	0.3176	0.3840	0.3806	0.1964	_	0.1268	0.3891	_	0.3888
	strange/wired	0.2951	0.2539	0.2610	0.2349	_	0.1171	_	_	_	0.2767
	trendy/hip	0.5289	0.4373	_	0.5911	_	_	_	_	0.6415	_
	unique	0.4700	0.4421	0.2866	0.4114	0.3233	0.1950	0.3179	0.3287	0.4789	0.4030
companion (with)	friend	0.4261	0.3363	0.2430	0.4301	0.2329	0.2044	0.3533	0.3785	0.3646	0.2454
	girlfriend	0.3074	0.1016	0.2593	0.2262	_	_	_	0.0431	0.2221	-0.0320
	boyfriend	0.0993	0.1675	0.1508	0.1200	0.1657	0.2143	_	0.0787	0.1846	0.1133
time (time)	morning	0.3507	0.2676	0.2485	0.3105	0.5214	_	0.3763	0.5160	0.4334	0.3654
	afternoon	0.4034	0.2057	0.1199	0.1502	0.1836	_	0.2009	0.3034	_	0.2600
	night	0.4785	0.3053	_	0.2034	0.3450	0.2374	0.1089	0.3748	0.3924	0.3614



<그림 2-141> 장소 테마 분석 차트 서비스

❍ 머신러닝 기반 문서분류

- 사용자가 한글로 작성한 네이버 블로그 글에 대해 Relevance Vector Machine(RVM)을 통해 문서 분류를 진행함. 이를 위해, 자연어 처리 된 한글 문서에 대해, 형태소 분석을 진행하고 그 중 동사, 명사, 형용사를 추출한다. 그리고 RVM을 통해 문서를 분류
- RVM의 문서 분류 성능 평가를 위해 같은 데이터에 대해 선형 SVMs를 적용하여 그 결과를 비교함. 이를 위해 두 알고리즘에 동일한 방식으로 Document Term Matrix를 생성하여 자질을 선정하고 자질 가중치를 부여함
- 실험을 통해 결과를 검증한 후, 검증된 데이터를 RVM reference dataset으로 구축함. RVM reference dataset은 이후 임의의 문서 집단이 입력되었을 때, 각각의 문서가 층간소음 관련 글인지 아닌지 판별하는데 이용됨

<표 2-46> 일부 reference data에 대한 SVMs와 RVM 성능비교 결과

구분	SVMs	RVM
precision	0.82	0.92
recall	0.87	0.74
F1	0.84	0.82

- 정확도 검증 결과, 정밀도(precision)는 RVM에서 더 높았으며, 재현율(recall)은 SVMs에서 조금 더 높은 값을 보임. F1 값은 RVM에서 조금 낮게 나왔지만 비슷한 결과를 보임. RVM은 SVMs와 달리 결과가 probabilistic하며, parameter C가 필요 없다는 점 등 다양한 장점을 가짐



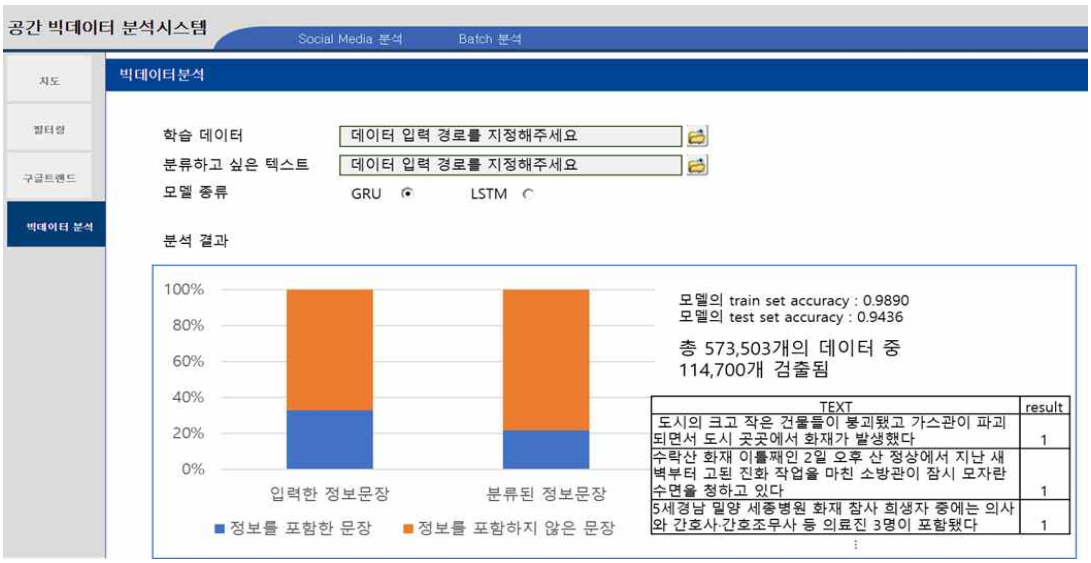
<그림 2-142> 머신러닝 기반 문서분류 서비스

❍ 머신러닝 기반 문장분류 기술

- 소셜미디어 데이터를 이용한 딥러닝 모델에 데이터 증강 기법을 적용하여 재난 정보 중 특히 화재 재난을 포함한 문장을 텍스트 데이터에서 탐지하는 정확도를 향상
- 데이터 증강 기법(Data Augmentation)이란 딥러닝 및 기계학습 분야에서 인위적으로 데이터의 양을 증가시키는 기술임. 텍스트 데이터에서는 문장 속 단어의 순서나 특정 표현이 빠짐으로써 문장의 의미가 달라질 수 있으므로 이미지와 같은 증강 기법을 사용할 수 없음
- 학습된 Word2Vec 모델을 이용하여 뉴스 문장 데이터를 동의어로 치환. 이때, 모든 형태소에 대해 치환이 진행하면 조사와 어미, 접미사까지 바뀌어버려 이해할 수 없는 문장이 생성되는 문제가 발생. 따라서 명사에 한해서 동의어로 치환
- 뉴스 문장 5,931개를 두 배로 증강시킨 데이터를 이용하여 재난 문장 분류 모델을 학습한 결과 정확도 0.9456, F1-score는 0.6540, AUC(The Area of Under a ROC Curve)의 값은 0.8371임
- 기존의 사전 기반의 문장탐지 방식은 사용자가 많은 양의 사전을 직접 지속해서 관리 및 업데이트를 해주어야 하지만 딥러닝에서는 새로운 라벨링 데이터를 추가만 하면 모델이 스스로 학습하여 파라미터를 갱신함

<표 2-47> 텍스트 분류 결과(일부)

TEXT	Y	NEW_PREDICTED
임직원들은 이 장비로 추락과 화재 등 현장에서 발생할 수 있는 각종 사고 상황에 대한 대처 방안을 배울 수 있다	0	1
도시의 크고 작은 건물들이 붕괴됐고 가스관이 파괴되면서 도시 곳곳에서 화재가 발생했다	1	1
불이 난 사실을 모른 채 잠든 다른 주민들을 깨우기 위해 서였다	1	1
그는 지난 26일 밀양 세종병원 화재 참사 현장을 방문해 "이 직전의 이곳(경남)의 행정의 최고 책임자가 누구였는지 한 번 봐야겠다"며 홍 대표의 책임론을 제기한 바 있다	0	1
...	...	...



<그림 2-143> 머신러닝 기반 문장분류 서비스

❍ Key player 탐지 기술

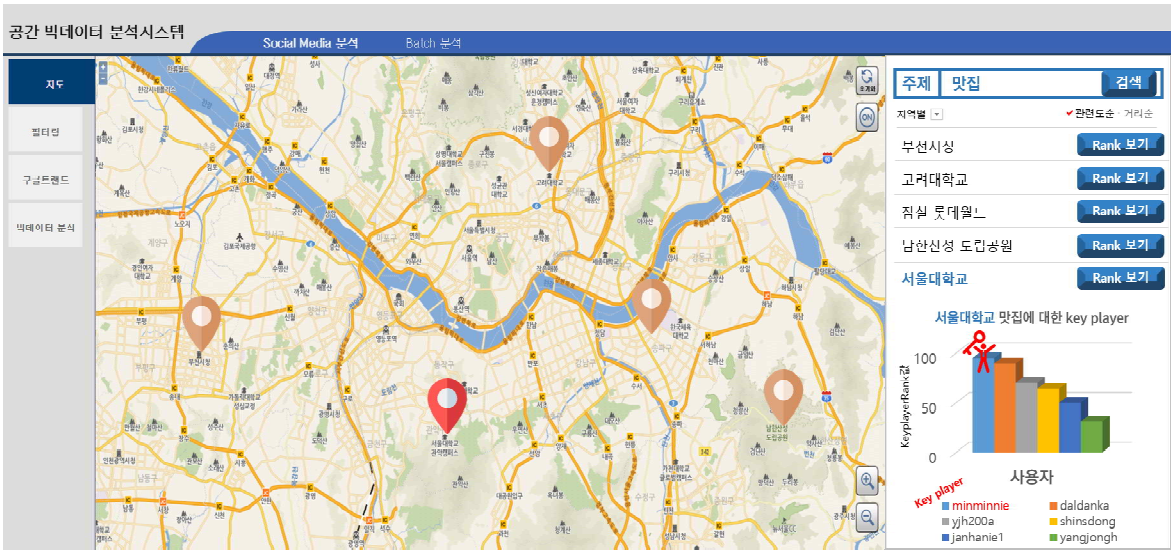
- 뉴스기사에서 LDA를 통해 주제를 추출하고, 트위터 데이터에서 각 주제별 키워드를 언급하는 사용자들을 선별하여 사용자들 간의 소통도와 주제 유사도가 가장 높은 key player를 KeyplayerRank를 통해 탐지함
- 데이터 전처리를 위해 일정기간 동안 수집된 뉴스기사를 문장 단위로 분리한 후 단어별로 tokenize함. tokenize된 결과에서 한자, 영문자, 숫자와 같은 불용어를 제거하고, 형태소분석을 실시한 뒤 명사(NNP, NNG)만을 추출함
- 추출된 명사들을 가지고 주제를 추출하는데, 주제를 추출하는 방법에 있어 토픽 모델링 기법 중 가장 대표적인 LDA(Latent Dirichlet Allocation)를 이용하여 각 주제 추출 및 주제별 키워드 선정

<표 2-48> '관광'과 관련된 뉴스기사의 LDA 결과(일부)

구분	해당 키워드	주제
Topic 1	0.019*"와인" + 0.007*"패키지" + 0.006*"체험" + 0.006*"양조장" + 0.006*"이용" + 0.006*"대표" + 0.005*"가족" + 0.005*"만원" + 0.005*"포도" + 0.005*"진행"	와인축제
Topic 2	0.027*"축제" + 0.019*"행사" + 0.017*"진행" + 0.016*"문화" + 0.016*"체험" + 0.014*"이벤트" + 0.014*"개최" + 0.013*"프로그램" + 0.013*"한국" + 0.011*"공연"	문화
Topic 3	0.024*"스키" + 0.013*"스키장" + 0.010*"슬로프" + 0.008*"시즌" + 0.008*"이용" + 0.007*"시간" + 0.007*"리프트" + 0.006*"운영" + 0.006*"전주" + 0.005*"안전"	스키

- 뉴스기사와 동일한 기간의 트위터 데이터를 Open API를 통해 수집하고, LDA를 이용하여 추출된 키워드들을 언급한 트위터 이용자들을 추출하여 각 주제별 연결 관계를 형성함
- 형성된 연결 관계에 코사인 주제 유사도(C_t)를 가중치로 곱하여 각 주제별 가중 네트워크를 생성함
- 기존의 PageRank 알고리즘에서 변형된 KeyplayerRank를 이용하여 key player 탐지. 이때, KeyplayerRank는 사용자들 간의 소통도와 주제 유사도를 고려하여 KeyplayerRank값이 가장 높은 key player를 탐지하는 알고리즘임
- 여기서 소통도는 트위터 이용자의 retweet(RT)된 횟수, reply된 횟수 그리고 favorite수를 고려한 지수임
- p_i의 KeyplayerRank를 구할 때, N은 p_i와 연결된 모든 트위터 이용자 수이고, T_i는 p_i가 쓴 전체 트윗 수이며, $L(p_i)$는 p_i에서 나가는 외부 연결을 의미함

$$KeyplayerRank(p_i) = \frac{1-d}{N} \times \frac{RT수(i) + reply수(i) + favorite수(i)}{T_i} + d\sum C_t \times \frac{KeyplayerRank(p_j)}{L(p_j)}$$



<그림 2-144> key player 탐지 차트 및 분포 서비스

⑨ [단위성과 E-12, E-13] 공간 빅데이터 분석결과 가시화 기술(정적 가시

화, 5,6차년도)

㉮ 토픽 기반 장소 임베딩 기술

○ 장소 리뷰 데이터에 토픽별 워드 임베딩 기술을 적용해 토픽 기반으로 장소를 임베딩하는 기술 개발

- 수집된 한국어 장소 리뷰 텍스트 데이터로부터 자연어 처리(형태소 분석)를 거쳐 LDA 기반 토픽 모델링 과정에서 토픽 추출, 단어별 토픽 할당 후 토픽별 워드 임베딩을 통해 장소를 벡터로 임베딩

○ 망고플레이트를 대상으로 서울 내 POI 리뷰 데이터 크롤링 후 일반 명사(NNG), 고유 명사(NNP), 동사(VV), 형용사(VA), 어근(XR) 품사에 해당하는 단어를 포함하는 리뷰 데이터를 추출한 결과, 209,406개 리뷰가 추출됨

○ 추출된 리뷰 데이터에 대해 LDA 기반 토픽 모델링을 적용해 4개 토픽(서비스, 가성비, 외부 분위기, 내부 분위기)을 선정했으며, 해당 결과가 토픽별 워드 임베딩(topical word embedding)의 인풋으로 활용됨

○ 단어와 토픽을 기반으로 학습하는 토픽별 워드 임베딩 과정에서 추출된 단어 벡터와 토픽 벡터의 연쇄 결합(concatenate)을 통해 토픽별 단어 임베딩이 이루어짐

○ 각 장소별로 작성된 리뷰에 대해, 토픽 모델링 과정에서 선정된 4개 토픽에 해당하는 단어-토픽 조합만을 추출해, 조건부 확률로 가중합을 계산, 해당 장소의 벡터로 확정함으로써 결과적으로 총 20,553개의 장소를 벡터로 임베딩함

○ 모든 장소에 대해 장소 임베딩 결과인 장소 벡터를 확정함으로써 코사인 유사도를 통해 장소 간 유사도를 계산할 수 있음

- 아래 표는 20,553개의 장소 중 일부 장소에 대해 해당 장소와 코사인 유사도가 가장 높은 장소 상위 5개를 추출한 결과임

- 기준이 되는 ‘10 꼬르소꼬모’의 카테고리인 ‘브런치/버거/샌드위치’와 1~5위로 추출된 장소의 카테고리인 ‘카페/디저트’와 유사함

<표 2-49> 브런치 카테고리 장소 임베딩 결과 코사인 유사도가 가장 높은 장소 상위 5개

순위	POI 명	주소	카테고리	유사도
기준	10 꼬르소꼬모	강남구 청담동 79	브런치 / 버거 / 샌드위치	-
1	달콤한거짓말	마포구 상수동 317-11	카페 / 디저트	0.98533210
2	카페디올	강남구 청담동 98-3	카페 / 디저트	0.97685640
3	코코브루니	마포구 동교동 165-3	카페 / 디저트	0.97519923
4	마망갸또	강남구 신사동 524-27	카페 / 디저트	0.97460386
5	글래머러스펭귄	용산구 한남동 743-41	카페 / 디저트	0.97385797

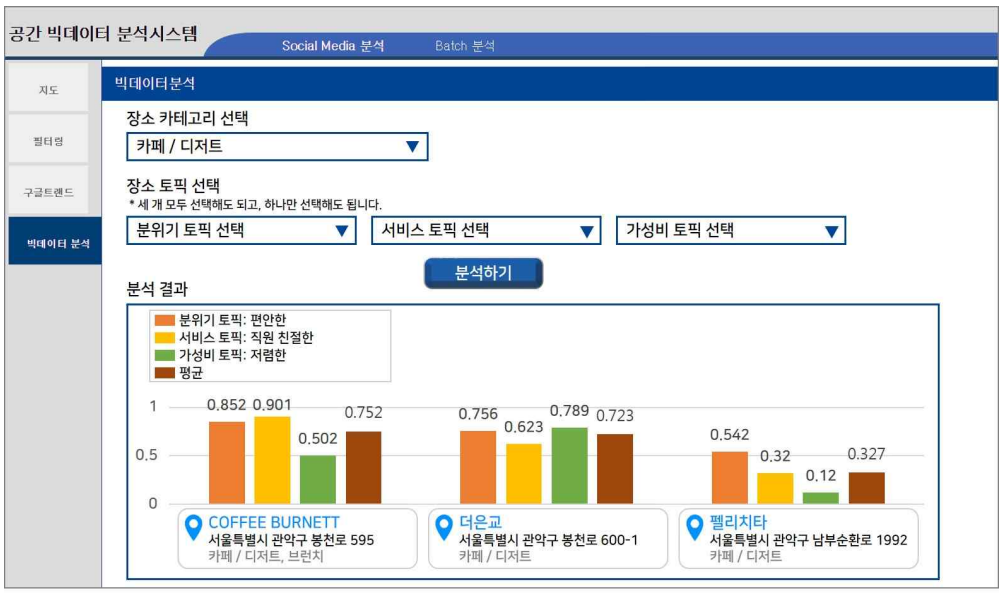
○ 아래 두 그림은 장소 검색 서비스 기술 예시로, 상단 그림은 유사 장소 검색 서비스 예시 화면으로, 사용자가 좋아하는 특정 장소를 검색하면 그와 유사한

장소가 유사도가 높은 순서대로 표출되는 서비스임



<그림 2-145> 유사 장소 검색 서비스 화면 예시>

○ 하단 그림은 토픽 기반 장소 검색 서비스 예시로, 사용자가 선택한 카테고리에 대해 분위기, 서비스, 가성비 토픽을 모두 선택했을 경우에 대한 분석 결과 예시인데, 사용자가 선택한 토픽별 연관도를 계산해 평균값이 높은 순서대로 시각화됨



<그림 2-146> 토픽 기반 장소 검색 서비스 화면 예시>

○ 토픽 기반 장소 임베딩 기술을 기반으로 한 이러한 서비스들은 기존 서비스에 비해 사용자에게 보다 다양한 장소 검색 환경을 제공할 수 있다는 차별점이 있음

㉯ 인공 신경망을 이용한 사용자 맞춤 추천

❍ Yelp, Tripadvisor, 망고플레이트 등 사용자 참여형 리뷰 데이터를 이용한 사용자 맞춤 추천 기술 개발

- 사용자 정보, POI 정보를 취합한 후 collaborative filtering 기술 기반의 추천

❍ Yelp 데이터를 이용하여 41,488개의 poi와 74,254명의 사용자 정보, 2,011,110개의 리뷰로 학습시킴

❍ 추출된 리뷰 데이터에서 아래 표와 같이 POI는 평점, 가격대, category(종류), 총 리뷰수를 사용자 데이터는 평균평점, 총 리뷰수를 이용하였음

<표 2-50> POI, 사용자 데이터 수집

POI	Feature type	User	Feature type
평점	Quantitative	평균 평점	Quantitative
가격대	Categorical	총 리뷰 수	Quantitative
종류	Categorical		
총 리뷰 수	Quantitative		

❍ 수집된 데이터는 quantitative data는 Min-max normalization [0,1]로, categorical data는 One-hot Encoding으로 전처리함

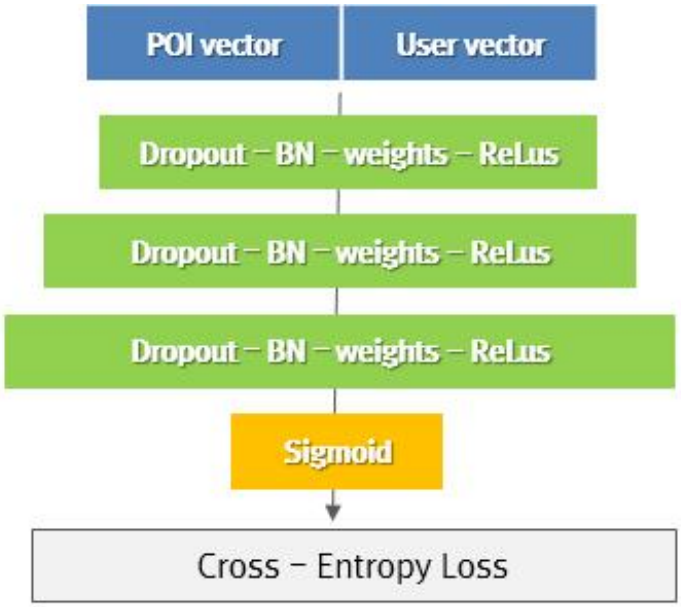
❍ 전 처리된 데이터는 아래 알고리즘으로 임베딩을 함

for POI_i
Categorical $P_{i,C}$: $\tanh(W_C POI_{i,C} + b_C)$
Numeric $P_{i,N}$: $\tanh(W_N POI_{i,N} + b_N)$
POI embedding $P_i = concat(P_{i,C}, P_{i,N})$
$P_{i,N}$
for $User_i$
Categorical U_C: $\tanh(W_C User_{i,C} + b_C)$
Numeric U_N: $\tanh(W_N User_{i,N} + b_N)$
User embedding $U_i = concat(U_{i,C}, U_{i,N})$

<그림 2-147> 임베딩 알고리즘

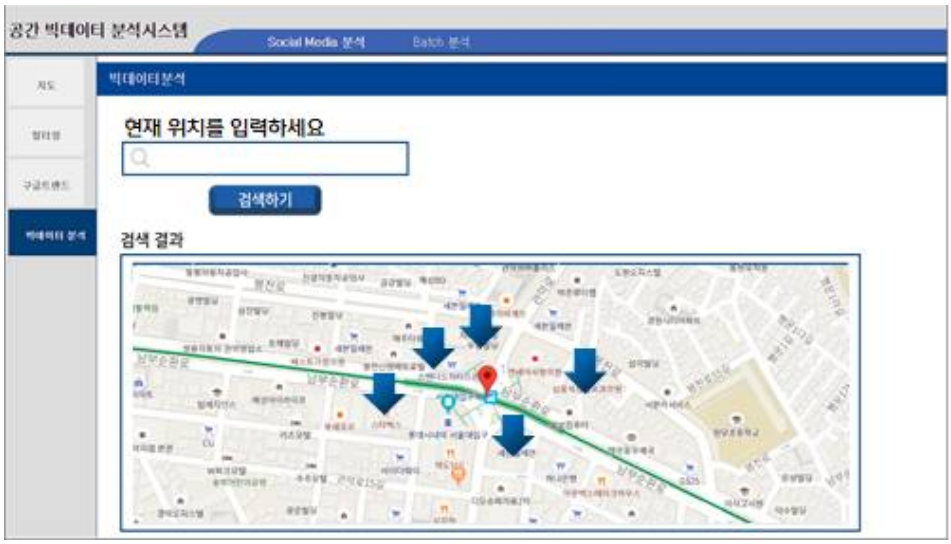
❍ 임베딩된 백터는 concatenate한 후 multi-layer perceptron 모델에 input이 됨

- Output은 user가 평가하는 그 POI의 점수를 나타냄. 아래 그림은 model을 표현한 그림임
- Dropout과 batch normalization으로 일반화를 높임
- Cross-Entropy Loss로 back propagation으로 Neural network에 parameter를 학습시킴



<그림 2-148> 사용자 맞춤 추천 연구흐름도

○ 아래 그림은 서비스 기술 예시로, 사용자 위치를 받으면 사용자 근처 POI중에 높은 순서로 K(K=5)개의 POI를 추천해주는 예시임



<그림 2-149> 사용자 맞춤 POI 추천

⑩ [단위성과 E-13] 공간 빅데이터 분석결과 가시화 기술(동적 가시화, 6차연도)

㉮ 단어임베딩을 활용한 사용자관심 이벤트 탐지

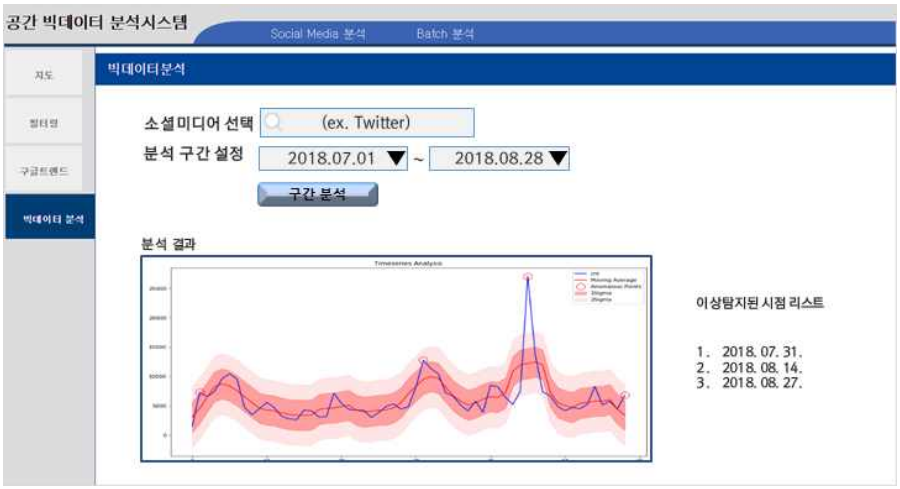
○ 소셜 미디어 데이터에 단어 임베딩 기술을 적용해 사용자 관심 이벤트를 탐지하여 관련 텍스트를 제공

- 수집된 한국어 소셜 미디어 텍스트 데이터로부터 자연어 처리(형태소 분석)를 거쳐 사전에 위키피디아(Wikipedia) 데이터로 학습시킨 단어임베딩 모델을 활용하여 이상탐지된 이벤트 중 사용자 관심 이벤트에 관련된 텍스트를 제공함

❍ 트위터 데이터를 대상으로 수집 가능한 텍스트 데이터를 수집하여 378,072개의 포스팅(posting)을 확보하고, 해당 텍스트의 발행일자와 본문내용을 추출하여 활용함

❍ 발행일자를 기준으로 일자별로 카운팅(counting) 하여, 시계열 분석을 통해 이동평균보다 1-시그마 이상 벗어난 시점을 이상시점으로 탐지하고 시각화를 통해 제공함

❍ 아래 그림은 소셜 미디어 이상시점 탐지 기술 예시로, 분석을 원하는 소셜 미디어와 분석 구간을 설정하면, 해당 기간 중 트윗 발행 수가 급증한 날을 알려줌



<그림 2-150> 단어 임베딩을 활용한 사용자 관심 이벤트 추출 서비스 화면 예시

❍ 한국어 위키피디아(2019년 4월 기준) 데이터를 활용하여 약 416만 문장으로 Word2Vec Skip-gram 300차원 단어 임베딩 모델을 구축하였음

❍ 단어 임베딩 모델을 구축하게 되면, 모델 내 등록단어에 포함된 단어들에 대해서 300차원의 벡터로 표현할 수 있기 때문에 단어 간 거리, 유사도 등을 측정할 수 있게됨

❍ 모델 내 등록단어(vocabulary) 간 거리를 계산한 거리행렬(distance matrix)를 구축함

❍ 이상시점 중 사용자가 특정 시점을 선택하면, 해당 시점 내 많이 언급되는 키워드들이 워드클라우드(wordcloud) 형식으로 보여지고, 사용자가 이벤트 단어를 설정하면, 관련 텍스트를 추출하게됨

❍ 워드클라우드는 텍스트 내 명사들만을 추출하여 빈도에 따라 글자의 크기를 달르게 하여 주요 이벤트 키워드들을 쉽게 보이게 함

○ 아래 그림은 사용자 관심 이벤트 관련 텍스트 추출 예시로, 탐지한 이상시점 중 관심있는 날짜를 선택 후 이벤트탐지를 누르면, 해당 시점의 다양한 키워드들이 나오는데, 이 중 사용자가 관심 이벤트를 설정하면 관련 텍스트들을 추출할 수 있음



<그림 2-151> 단어 임베딩을 활용한 사용자 관심 이벤트 추출 서비스 화면 예시

○ 단어 임베딩을 활용한 사용자관심 이벤트 탐지 기술을 기반으로 한 위 서비스는 특정 키워드를 포함해야 하는 기존 서비스에 비해 특정 키워드를 포함하지 않아도 사용자가 선택한 이벤트 단어와의 ‘거리’ 또는 ‘코사인 유사도(cosine similarity)’를 통해 관련 단어까지도 포함하는 텍스트를 추출하여, 소셜미디어 플랫폼에서 특정 시점에 특정 이벤트에 대한 텍스트들을 확인할 수 있음

○ 추후 텍스트 감성분석을 통해 해당 키워드 또는 이슈에 대한 여론의 감성을 파악할 수 있으며, 이는 선거에서의 지지율, 신규 출시 품목에 대한 호감도 등을 파악할 수 있을 것으로 기대됨

㉯ 통행목적을 고려한 공공자전거 수요예측 기술

○ 대중교통통합모델인 MaaS(Mobility as a Service)가 하나의 트렌드로 떠오르면서 공유자전거에 대한 위상이 높아지고 있음

- 본 기술은 이용자들에게 친숙한 대중교통수단인 지하철역과 연계하여 이용할 수 있는 공공자전거의 수요를 예측하고자 함
- 정확한 수요예측을 위해서는 다양한 시공간데이터가 필요한데, 기존 연구의 경우 데이터 구득의 문제로 대부분 시간 혹은 공간에 치우쳐 데이터를 구축하였음
- 시간에 초점을 맞춘 경우는 날씨 데이터를 주로 활용하여 예측하였으며, 공간데이터를 고려한 경우는 정적상태의 영향력만을 고려하였음
- 본 기술을 같은 공간이라도 시간대에 따라 활용도가 달라질 수 있기 때문에 공

간의 동적 상태를 반영한 데이터를 구축하여 정확한 수요예측에 기여하고자 함

- 특히 공공자전거의 통행목적을 고려하여 이에 대한 수요예측모델 기술을 개발함

❍ 공공자전거 수요에 미치는 요인을 크게 날씨요인, 대중교통요인, 토지이용요인으로 나누어서 구축하였음

- 날씨는 다른 교통수단과 달리 자전거 이용에 영향을 주는 요인이므로, 기온, 습도, 풍속, 강수, 적설량 등을 구축하여 비교함

- 대중교통요인의 경우 자전거와 연계하여 이용할 수 있는 점을 고려하여 자전거 대여소 근처의 지하철역의 실시간 하차인원에 대한 데이터를 스마트카드를 이용하여 구축하였음

❍ 토지이용요인은 시간대별로 공간의 동적 상태를 반영하기 위해 자전거 통행목적을 구분한 후, 각각의 대여수요에 해당하는 데이터를 구축하였음



<그림 2-152> 공공자전거 수요예측 기술 전체 흐름도

- 자전거 통행목적은 크게 통근, 통학, 귀가, 여가로 구분하였으며, 해당 데이터에 시간대별 분포를 고려하여 추출하였음

- 통근, 통학, 귀가는 가구통행조사에서 실시한 통행목적별 출발시간 분포표를 사용하였음

- 여가의 경우는 대략적인 흐름이 있는 통근통학과 달리 POI별로 상이하므로 구글 장소에서 제공하는 체크인 데이터를 활용하여 시간대별 분포를 파악하였음, 이때 체크인데이터의 경우 실세계를 반영하는 정확한 데이터는 아니므로 표준화작업이 필요함

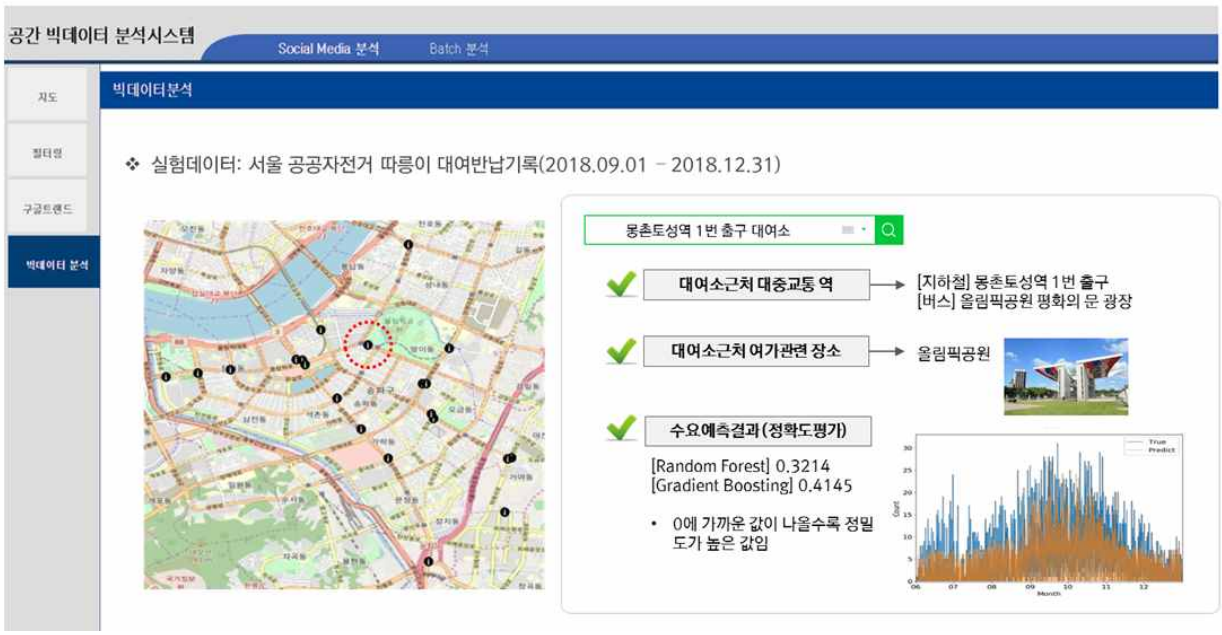
❍ 수요를 예측하는 머신러닝 기법으로 다양한 방법론들이 있는데, 본 연구에서는 분류와 회귀 문제의 다양한 데이터셋에서 효과적이라고 입증된 랜덤포레

스트와 그레디언트 부스팅을 사용하여 자전거 수요를 예측하였음

- 둘 다 모델을 구성하는 기본요소로 결정트리를 사용하였으며, 여러개의 결정트리를 묶어 강력한 모델을 만드는 앙상블 기법임
- 모델의 정확도를 평가하기 위해서는 RMSLE(Root Mean Squared Logarithmic Error)를 사용하였는데, 과대평가된 항목보다는 과소평가 된 항목에 패널티를 주는 방식으로 0에 가까운 값이 나올수록 정밀도가 높은 값임

$$\text{RMSLE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\log(p_i+1)-\log(a_i+1))^2}$$

- 이때 p_i가 예측값이고, a_i가 실제값으로 계산됨
- 또한 일반화 성능을 측정하기 위해 데이터를 여러번 반복해서 나누고 여러 모델을 학습하는 k-fold cross validation을 실시하였음



<그림 2-153> 통행목적을 고려한 공공자전거 수요예측 가시화 서비스

⑪ [단위성과 E-2] 데이터 특성별 분석모델 설계서(1차연도)

㉮ 공간 빅데이터 분석 프레임워크 사례분석

○ O2 플랫폼

- O2 플랫폼은 솔트룩스에서 개발 운영 중인 소셜 빅데이터 분석 플랫폼으로 실시간으로 대용량의 소셜데이터를 수집하고, 수집된 대용량의 데이터(빅데이터)를 기반으로 분석 기능을 제공
- 수집된 빅데이터는 실시간으로 분산 색인이 되어 빠른 검색이 가능하며 검색 및 분석 기능을 Open API로 제공하여 외부 데이터나 분석 도구와 연계가 가능함

- 플랫폼의 안정성을 모니터링하기 위한 자동 모니터링 기능을 제공
- O2 플랫폼은 크게 수집 레이어, 검색과 색인 및 분석 레이어, 서비스 인터페이스 레이어로 구성되어 원시 데이터의 저장부터 분석 결과를 통한 의사 결정까지의 단계를 지원
- 의사결정에 이르는 모든 단계에서의 분석을 위해 O2 플랫폼은 NoSQL과 하둡 에코 시스템에 기반 하여 구성되어 있음
- O2 플랫폼은 3개 레이어로 이루어져 있으며, 각 레이어는 기능적 모듈들을 포함

<표 2-51> O2 플랫폼 레이어

레이어	모듈	기능
서비스 인터페이스 레이어	서비스	• 실시간 주요 토픽 추출 • R연동 인터페이스 • 연관 토픽 추출 • 시각화 • 토픽 트렌드 추출 • 데이터 검색 및 정제
	모니터링	• 대시보드 • 모듈 오류 검출 • 수집량 통계 • 오류 알람 • 색인 통계 • 인스턴스 감시
검색, 색인 분석 레이어	색인검색	• 분산색인 • 분산 쿼리 병합 • 쿼리 연산 최적화
	분석	• 형태소 분석 • 토픽 연관성 분석 • 문서 분류 • 이슈 토픽 분석 • 개체명 인식 • 토픽 트랜드 분석
수집 레이어	수집	• 온디멘드 수집 • RSS 기반 수집 • 웹기반 수집 • 분산 수집 • 소셜 콘텐츠 수집 • 이미지 썸네일링

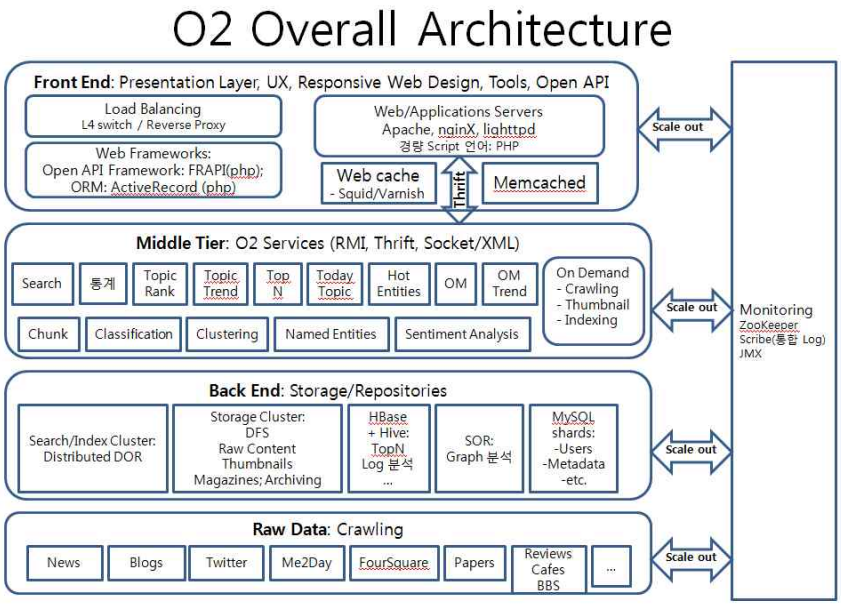
- O2 플랫폼의 물리 구성

<표 2-52> O2 플랫폼의 물리 구성

레이어	스펙
서버 하드웨어	• 197개 서버 • 1630개 CPU 코어 • 10TB 이상의 메인 메모리 • 500TB 이상의 하드디스크
DBMS & 파일 시스템	• NoSQL • DFS • 데이터 3중화

- O2 플랫폼 아키텍처

- O2 플랫폼은 대용량 비정형 데이터를 수집 및 분석을 지원하는 플랫폼으로, 플랫폼은 4개의 단으로 구성
- Raw Data 단은 뉴스, 블로그, 트위터, 미투데이 등의 소셜 미디어 데이터 및 대용량 이미지 파일, 웹 텍스트, 논문, 카페, 리뷰글 등의 데이터 수집을 담당
- Back End 단은 수집된 데이터를 저장하고 관리하는 MySQL, MongoDB, DFS등의 저장소로 이루어짐
- Middle 단은 비정형 데이터를 분석하고 서비스를 제공하는 검색, 텍스트 마이닝 및 자연어 처리 분석 모듈로 이루어짐
- 텍스트 마이닝 기능으로 자동분류, 자동군집, 연관단어 추출, 특성 추출, 트랜드 분석, 자연어 처리 기능으로 개체명 인식, 감성어 분석, 여론 분석 등을 포함



<그림 2-154> O2 Overall Architecture

- O2 세부 기술

- 전반적으로 O2 플랫폼의 하부 기술들은 오픈소스에 의존
- 프론트엔드 단은 주로 외부 인터페이스를 위한 레이어로 오픈소스 또는 상용 솔루션으로 구성
- 백엔드 단 역시 하부 기술은 오픈소스로 이루어져 있음
- 예컨대, MySQL이나 Hbase, Hive, 분산파일시스템과 같은 기본적인 스토리지/리파지토리 솔루션은 오픈소스임
- 미들티어 단은 주로 분석, 분류, 색인하는 역할을 담당하는 레이어로 <O2> 플랫폼의 핵심 기술임
- 백엔드 단은 뉴스, 블로그, 트위터, 미투데이, 포스퀘어, 논문, 카페, BBS 등등 크로링한 Raw Data를 저장하고 클러스터링하고 색인화함
- 이 결과를 가지고 미들티어 단에서 자동분류를 하고, 자동 클러스터링화하며, 문서를

요약하며 추출된 특성으로부터 그래프화하고 트렌드를 분석

<표 2-53> O2 플랫폼 세부기술

기술명	설명
텍스트 추출	• 이메일, 오피스 문서등의 분석 대상이 되는 문서의 내부 텍스트를 추출하는 기술
형태소 분석	• 텍스트에서 문장, 단어, 그리고 형태소 단위로 분석하는 단계
특성 추출	• 각 문서의 중요키워드에 해당하는 특성을 추출하고, 특성들의 형태와 중요도에 따른 특성벡터를 구성하는 단계
정보추출 (개체, 속성, 사실, 사건, etc)	• - Entity (개체): 사람, 제품, 조직과 같은 관심 대상이 되는 개체 • – Attribute (속성): 이름, 별명, 종류, 설명등과 같은 개체에 대한 속성 • – Fact (사실): 두 개 혹은 여러 개체간의 관계 • – Event (사건): 여러 개의 개체가 관련된 활동
자동분류	• 사용자에 의해 미리 정의된 분류 체계에 대해 기계가 자동 학습 또는 규칙을 통해 자동으로 문서를 분류
자동군집	• 추출된 특성값을 바탕으로 형태와 중요도에 따른 벡터를 구성하고 타 문서와의 유사도 및 관련성을 계산
문서요약	• 문서의 기본적인 내용을 유지하면서 문서의 중요 내용을 추출하여 자동으로 복잡도를 줄여 요약하는 작업
Top N	• 추출된 특성들을 분석하여 일정 기간 동안의 토픽 순위변화를 나타내어 급격하게 떠오르고 지는 이슈를 표현하는 기능
Topic Lank	• 문서를 분석하고 문서 내에 발생하는 Topic 간의 자동 군집을 통해 시멘틱 네트워크 관계를 생성하는 기술
Topic Trend	• 일정 기간 동안의 토픽 별 동향 분석을 나타내는 기술
Opinion Mining	• 상품 리뷰, SNS 텍스트 및 블로그 댓글등을 기반으로 사용자의 의견을 분석하여 자동으로 평가를 도출해 내는 기술

- O2 플랫폼 구조 분석

• O2 플랫폼은 기본적으로 소셜 분석을 위한 플랫폼으로 실시간 분석이 가능하다는 점에서 주목할 만한 기술을 탑재한 것으로 보임

• 그러나, O2 플랫폼은 공간정보에 대한 활용을 그다지 전면에 내세우고 있지 않음

• 즉, 공간정보를 다루는 기술모듈이 존재하고 있지 않음

• Raw Data에서 크롤링으로 트위터나 포스퀘어 데이터를 수집하기는 하나, 실제 분석 모듈에서 공간정보를 활용하고 있지 않음

• 결론적으로 O2 플랫폼은 실시간 소셜 분석 플랫폼이긴하나, 플랫폼의 기반이 시맨틱 검색에서 출발했기 때문에 공간정보에 대한 처리가 부족함

❍ SAP HANA Geospatial

- 원래 SAP HANA는 High-Performace Analytic Appliance의 약자로 인메모리, 컬럼-지향적(column-oriented)인 관계형 데이터베이스 관리시스템임

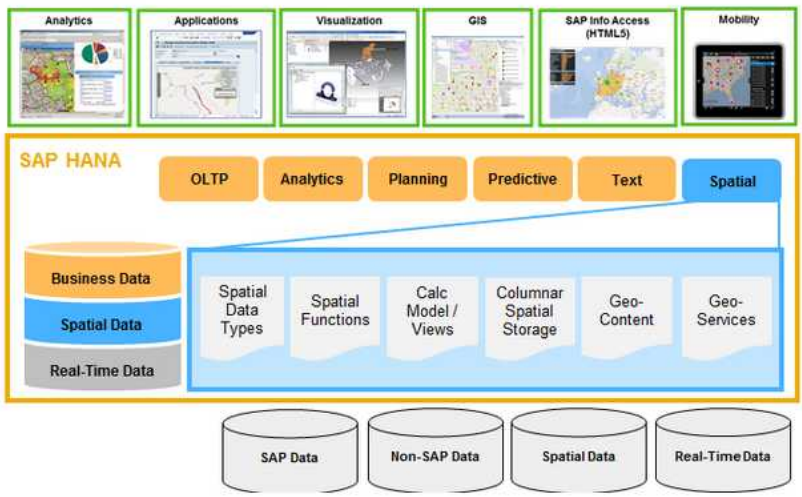
- HANA는 같은 플랫폼에서 복잡한 질의 처리와 고성능 트랜잭션 처리를 동시에

다룰 수 있도록 설계됨

- SAP HANA Geospatial 기능은 SAP HANA SP6에서 처음 도입되었으며 SAP HANA Geospatial Engine과 다른 엔진들을 결합함으로써 SAP은 강력한 공간 빅 데이타 분석 플랫폼을 확보

- SAP HANA Spatial

  - HANA Spatial은 ST_POINT, ST_GEOMETRY 등과 같은 공간 데이터 타입을 저장하고 처리할 수 있는 기능을 제공
  - 이러한 데이터 타입들은 애플리케이션 개발자들이 공간 정보와 그 속성 데이터를 연결할 수 있게 함
  - 예컨대, 기업 테이블에서 각 기업의 위치를 포인트로 저장하고, 배송 영역을 폴리곤으로 저장한다고 하면 이들은 다음과 같이 SQL 문으로 표현할 수 있음
  - SELECT * FROM Locations WHERE DeliveryArea.ST_Contains (new ST_POINT(1,1)) = 1h
  - SAP HANA는 예측, 계획 그리고 텍스트 검색과 같은 다양한 엔진을 제공함
  - 기존의 엔진과 Geospatial 엔진을 조합하여 복잡한 알고리즘을 적용하여 복합 데이터셋과 예측추론에 적용이 가능
  - 특히 Nokia의 HERE map 서비스를 HANA Spatial 기능으로 내장함으로써 HANA XS 엔진을 사용해 앱을 생성하는 것도 가능
  - 공간 데이터가 HANA로 로딩되면, 사용자는 HANA 모델(분석, 계산 뷰)을 생성하거나 확장하며 이때 HANA Studio를 사용하는데, 사용자는 SAP Lumira 또는 커스텀 HTML5 애플리케이션을 이용해 모델들의 분석과 시각화가 가능함
  - SAP HANA SQL 스크립트는 공간 데이터에 접근할 수 있도록 Spatial SQL MM 표준을 지원하도록 확장되었음



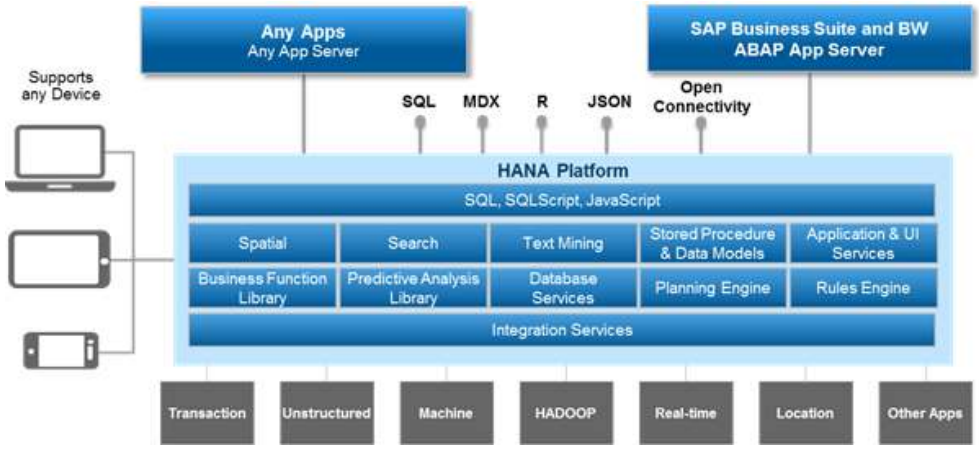
<그림 2-155> SAP HANA Geo-Spatial Features

- SAP HANA 플랫폼 구조

  - SAP HANA의 플랫폼 아키텍처는 통합 서비스 레이어와 각종 처리 모듈의 미들티어 레

이어, 그리고 타 언어와 바인딩하며 각종 애프리케이션과 서버와 통신할수 있게 하는 인터페이스 레이어, 이렇게 3개 레이어로 구성

- 아래 그림에서 볼 수 있듯이 SAP HANA는 플랫폼, 데이터베이스, 그리고 데이터 프로세싱 기능을 제공하며 이와 동시에 예측, 계획, 텍스트 처리, 공간정보, 그리고 비즈니스 분석 라이브러리를 제공
- 기능 서비스는 부수적인 기본 기능으로 이 기능을 이용해 사용자의 애플리케이션을 SAP HANA에서 바로 동작할 수 있도록함
- 예컨대, 대표적인 기능 서비스의 예가 텍스트 검색이며 이 서비스를 이용해 사용자는 빠르게 사용자의 앱을 개발해 사용자가 다른 종류의 비정형 데이터에서 컨텐트를 검색할 수 있도록 함



<그림 2-156> SAP HANA 아키텍처

- SAP HANA 플랫폼 구조 분석

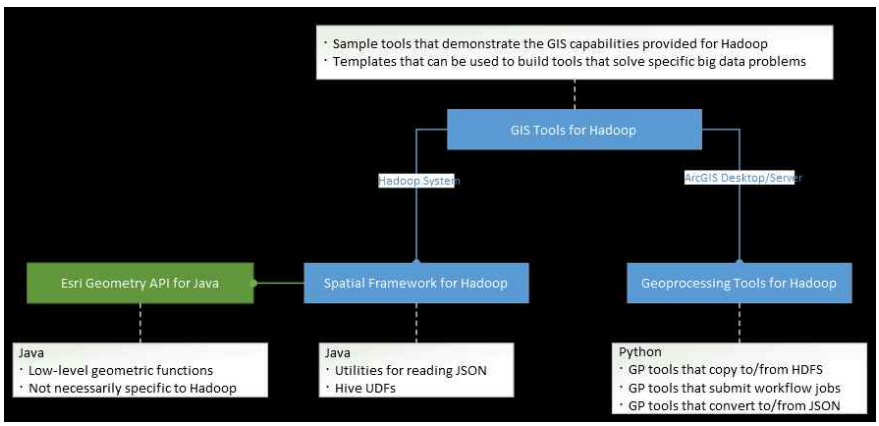
- SAP HANA는 원래 인메모리 데이터베이스로 여기에 다양한 모듈들이 부가되면서, SAP HANA는 오랜기간 비즈니스 분석 플랫폼으로서 역할을 수행함
- 2013년 SP6가 발표되면서 빅데이타와 공간정보를 다룰 수 있도록 확장되었으며 비즈니스 분야에서 강력한 공간 빅데이타 분석 플랫폼으로 역할을 수행함
- SAP HANA는 엄밀히 얘기해서 공간정보 빅데이타 플랫폼 분석 플랫폼은 아니지만 다양한 라이브러리를 제공하고 있어 비즈니스 분야에서 공간정보 분석 플랫폼으로 충분히 활용할 수 있도록 확장되어 있음
- 다만, SAP HANA 플랫폼에는 전문적인 공간정보 빅데이타 분석 모델을 위한 도구를 제공하고 있지 않기 때문에 이러한 도구들은 서비스 인터페이스 레이어를 통해, 사용자가 직접 개발해 활용해야 함
- 예컨대, 공간정보를 다루는 기능은 제공하지만, 정작 공간정보를 정확히 분석하기 위해서는 ESRI의 ArcGIS와 연동하는 방식을 취해야 함
- 아울러 각종 지적도와 같은 기본지도의 경우에도 노키아의 HERE맵과 연동이 필요함
- 이렇듯 아키텍처 상으로는 매우 훌륭한 공간정보 빅데이타 분석 플랫폼이지만, 공간정보 빅데이타를 다루기 위해서는 인터페이스를 통해 다른 솔루션과의 연동이 필요함

❍ ESRI GIS Tools for Hadhoop

- 하둡 사용자들은 종종 공간정보가 포함된 데이터를 다루게 되는데, 이러한 데이

터를 다루기위한 선택이 매우 제한적임

- ESRI에서는 하둡 애플리케이션에서 공간 데이터를 다룰 수 있도록 프레임워크를 오픈소스로 발표하였음
- GIS Tools for Hadoop은 GIS 툴의 집합으로 공간정보 빅데이터를 분석하기 위해 Hadoop을 이용하는 경우 사용자의 편의성을 증대해 주기 위한 하둡을 위한 공간 프레임워크임
- 이 툴들은 하둡 툴박스를 위한 지오프로세싱 툴을 사용하도록 되어있는데, ArcGIS 지오프로세싱 환경에서 하둡 시스템에 접근할 수 있도록 해줌
- 이 프레임워크가 유용한 이유는 ArcGIS 지오프로세싱 환경에서 기존 하둡 애플리케이션과 연동이 가능하기 때문임
- ESRI GIS Tools for Hadoop 특징
  - 샘플 툴들이 제공되어 하둡을 사용해서 GIS 문제들의 솔루션을 찾기위한 모든 리소스의 풀스택 구현물을 사용해 볼 수 있음
  - 또한 특정 문제의 솔루션을 위한 커스텀 툴을 구축하기 위한 템플릿도 제공
  - 커스텀 툴 구축을 위한 리소드는 아래와 같음
  - Spatial Framework for Hadoop
    Java helper utilities for Hadoop developers
    Hive spatial user-defined functions
  - Esri Geometry API Java - Java geometry library for spatial data processing
  - Geoprocessing Tools - ArcGIS Geoprocessing tools for Hadoop
  - Hadoop 기반에서 빅 데이터 공간정보 애플리케이션 작성을 위한 오프소스 프레임워크 & API 툴킷
  - Hadoop 애플리케이션과 ArcGIS 지오프로세싱 환경을 결합
  - 단순 피처(Point, Line, Polygon) 처리와 기본적인 분석 기능을 지원
- ESRI GIS Tools for Hadoop 아키텍처
  - ESRI GIS tools for Hadoop은 깃허브에 오픈소스로 공개
  - https://github.com/Esri/gis-tools-for-hadoop



<그림 2-157> GIS Tools for Hadoop

- 위 그림에서 보듯이 GIS Tools for Hadoop은 Hadoop 시스템과 ArcGIS Desktop/Server와의 징검다리 역할을 수행
- 하둡 파트의 공간 프레임워크는 자바로 인터페이스가 제공되며 Esri Geometry API를 이용해 공간정보를 다룰 수 있게 됨
- 관련 컴포넌트는 아래와 같음
- Geodatabase 피처 클래스와 JSON 형식파일 (Hadoop 파일 시스템, HDFS)간 변환 도구
- ArcGIS와 Hadoop간 데이터 복사 도구
- Hadoop의 워크플로우 실행 및 처리상태 점검 도구
- 아울러 Esri Geometry API는 자바로 제공되며 패키지명과 분석 및 연산기능은 아래와 같음
- 패키지명 : com.esri.core.geometry.*

<표 2-54> Geometry API for Java

Relationship Analysis	Operations
• equals • disjoint • touches • crosses • within • contains • overlaps •	• buffer • cut • clip • convexHull • intersect • union • difference

- ESRI GIS Tools for Hadoop 구조분석
  - GIS Tools for Hadoop은 ESRI의 GIS 기술과 하둡의 분산 빅데이타 처리 기술을 접속하려는 ESRI의 노력으로 이루어진 결과물임
  - ESRI가 공간정보 빅데이타를 다루기 위해 강하게 두 개의 시스템을 통합한 것이 아니라 느슨하게 시스템을 통합함으로써 각자의 사용자가 이질감을 느끼지 않고 서로의 시스템을 도입하게 하는 것이 가능해짐
  - 그러나, 이러한 접근방법은 통합되고 일관된 사용자 인터페이스와 사용자 운용환경을 제공해 주지 못하므로 확장 및 사용성에는 한계가 존재
  - 게다가 분산 노드의 확장에 있어 ArcGIS의 라이센스를 고려해야 하므로 확장 비용도 무시못할 고려 사항 중 하나임

㉯ 분석 프레임워크 평가 지표

❍ 공간정보 빅데이터 분석 프레임워크는 전 세계적으로 초기수준이며, 이제 발전하고 있는 분야이기 때문에 너무 구체적인 정량적 평가지표를 미리 설정하는 것은 합리적이지 않을 수 있음

❍ 왜냐하면 과제 종료될 시점에는 다른 요인들에 의해 경쟁 프레임워크의 성능이 기하급수적으로 증가할 것이기 때문임

❍ 따라서 분석 프레임워크의 평가지표는 입력대비 출력 성능지표보다는 확장성에 초점을 맞추어 평가지표를 삼는 것이 바람직하다고 간주됨

<표 2-55> 평가지표

평가지표			설명
일반 특성	사용자 인터페이스	시각적 표현 가능	• 사용자가 분석 프레임워크를 다루는데 있어 시각적 표현이 가능한지를 평가하는 지표. • 가능/불가능으로 평가함.
		사용의 편리함	• 사용자 인터페이스는 사용이 용이해야 함. • 정성적인 방법으로 1~5단계로 측정함.
	사용자 접근 용이성		• 접근용이성에 대해 용이함/안용이함으로 평가.
	최초 설치 및 사용환경 친숙도		• 초기 설치 방법 및 사용환경의 친숙도를 1~5 단계로 측정함. • 특지 자바 가상머신 사전설치 여부, 한글화된 설명서 등을 검사함.
추출 특성	공간정보 추출여부		• Raw Data에서 공간정보를 추출하는지 여부를 평가 • 예/아니오로 평가.
	추출 수준		• 개념과 관계를 추출하는지를 검사. 체크 여부로 평가
	자동화 여부		• Raw Data로부터 자동으로 탁사노미/개념/관계를 추출하는지를 평가. • 1~5단계로 평가함.
	알고리즘 선택 및 확장 가능성		• 알고리즘 선택 및 확장 가능한지를 평가.
	신뢰성		• 정성적인 평가로 신뢰성있게 추출하는지를 평가. • 1~5단계로 평가함.
	확장성		• 추출부의 확장이 가능한지를 평가. • 1~5단계로 평가함.
저장소 특성	비정형 데이타		• 비정형 데이터를 저장할 수 있는지 평가. • 예/아니오로 평가.
	리전 분리성		• 테이블의 부분집합을 관리할 수 있는지 평가 • 예/아니오로 평가
	데이터 이중화		• 데이터 이중화를 지원하는지 그리고 어느 정도로 지원하는지를 평가 • 1~5단계로 평가.
분석 모델 특성	분석 모델 수용성		• 다양한 분석모델을 수용하도록 구현되었는지를 평가. • 1~5단계로 평가한다.
	기존 공간정보 지원여부		• 분석모델에 기존 공간정보를 지원해 줄 수 있는지 여부를 평가. 예/아니오로 평가한다.
비용 특성	장비 비용		• 분산확장을 위해 노드를 추가했을 때의 평균 비용으로 평가. • 노드당 비용을 달러로 평가한다.
인터페이스 특성	언어 바인딩		• 타 언어로 바인딩이 가능하지 여부로 평가. • 1~5단계로 평가.
	웹서비스 여부		• RESTfull API를 제공하는지 여부로 평가. • 예/아니오로 평가.

㉮ 분석 프레임워크 개념적 설계

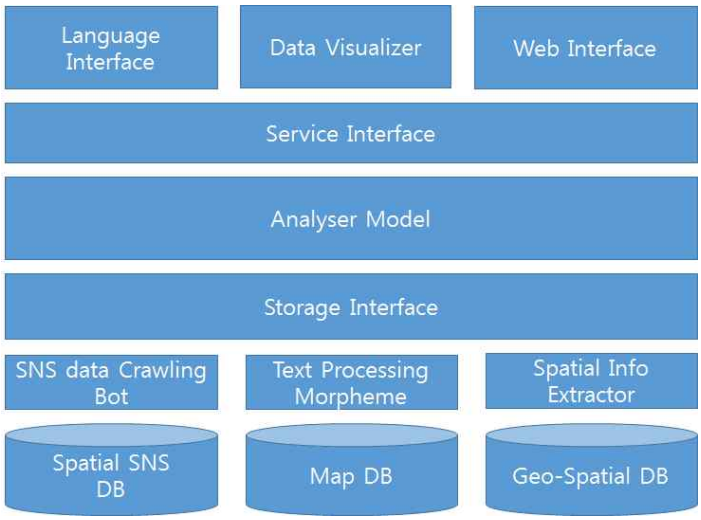
○ 프레임워크 설계 기준

- 오픈소스 솔루션 위주로 세부 컴포넌트로 활용

- 대용량의 데이터를 다루는 만큼, 하드웨어 비용을 최소화할 수 있도록 아키텍처를 구성
- 다양한 분석 모델을 적용할 수 있도록 모델 인터페이스를 규격화하고 컨트롤러를 구성
- 기존의 다양한 공간정보 데이터베이스를 분석 모델에 활용할 수 있도록 규격화

❍ 프레임워크 설계

- 공간정보 빅데이터 분석 프레임워크는 크게 세 가지 레이어로 설계함



<그림 2-158> GIS Tools for Hadoop

- 첫째는 데이터 수집 레이어로 각종 SNS 등에서 데이터를 수집하고 공간정보를 추출해 이를 데이터 리파지토리에 저장
- 둘째는 데이터 리파지토리 레이어다. 이 레이어는 분석 프레임워크의 성능과 직결되는 레이어로 검증된 기술을 이용해 레이어를 구성하도록 설계
- 셋째는 데이터 분석 레이어로 서울대에서 설계한 분석 모델을 컨트롤 할 수 있도록 설계

- 데이터 수집

- 다양하고 수많은 소스로부터 데이터를 추출하기 위한 레이어로 소스가 다양하기 때문에 파이프라인 방식으로 수집기와 전 처리기를 사용할 수 있도록 설계함
- 아울러 다양한 연산으로부터 데이터를 탐색하고 사용자와 인터액션하며, 정형 또는 비정형 데이터에 관계없이 데이터를 수집

- 데이터 리파지토리

- Hadoop, NoSQL 데이터베이스 응용툴들을 이용해 막대한 양의 데이터를 최소한의 비용으로 저장할 수 있도록 설계
- 소셜 미디어를 비롯해 기존에 구축되어 있는 다양한 형태의 공간정보등을 재사용할 수 있도록 데이터베이스 연결이 가능토록 설계
- 여기에 볼륨, 다양성, 속도의 특징을 가지는 데이터를 동적으로 분산 처리하고 수집할 수 있도록 함

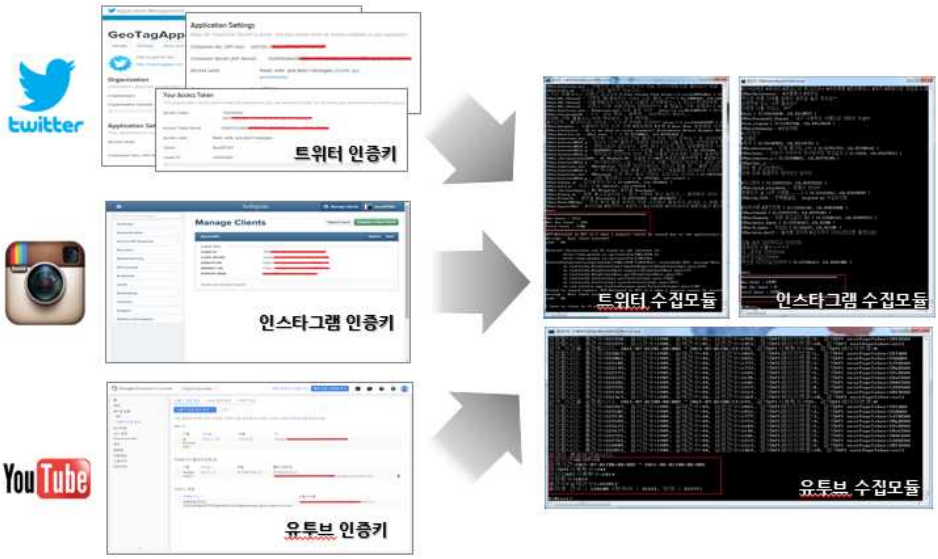
- 데이터 분석

  • 서울대에서 설계할 분석 모델을 수용하기 위한 레이어로 분포 분석, 변화 탐지, 연결성 분석, 예측 모델링, 감성 분석, 공간 인포그래픽스 등등의 방법들을 대응할 수 있도록 설계

⑫ [단위성과 E-5] 이기종 DB간 데이터 가공, 변환, 필터링 모듈 개발(2차연도)

㉮ 소셜 미디어 공간 데이터 수집 모듈 개발

○ 공간 빅데이터 분석 프레임워크 기술 개발을 위한 데이터 확보를 위해 소셜 미디어의 공간 빅데이터 특성을 분석하고, 수집 대상을 정의하여 트위터, 인스타그램, 유투브의 수집 모듈을 개발



<그림 2-159> 소셜 미디어 공간 빅데이터 수집 흐름

○ 트위터(Twitter) 수집 모듈

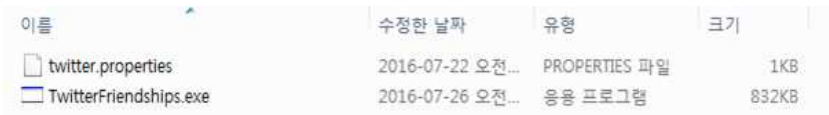
- Twitter Friendships REST APIs를 이용하여 트위터에 트윗한 사용자 및 팔로우한 사용자의 정보를 수집

- 실행파일 구성

  • JAVA 기반의 개발 도구를 이용하여 실행파일 생성

  • 실행파일 : TwitterFriendships.exe

  • 환경설정파일 : twitter.properties



<그림 2-160> SNA(Social Network Analysis) 분석용 실행파일 구성

- 환경설정 파일내용

  • Twitter Application Mangement 사이트에서 앱등록 및 인증키 발급

  • 참고사이트 : https://apps.twitter.com/

```
1  # Replace this with an API key available at https://apps.twitter.com/
2  twitter.consumerKey=zoD[redacted]
3  twitter.consumerSecret=OoWh[redacted]
4  twitter.accessToken=1830[redacted]
5  twitter.accessTokenSecret=53i90Y[redacted]
```

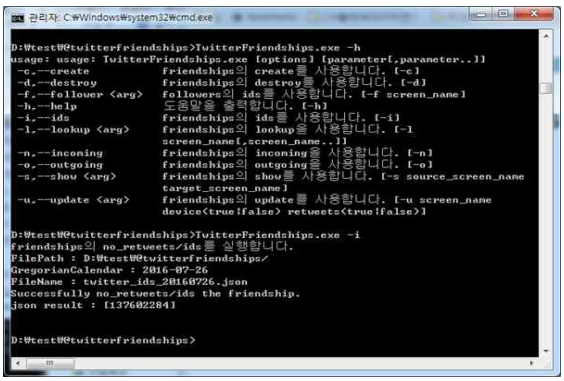
<그림 2-161> 환경설정 파일내용

- 실행

  • 실행에 필요한 파일들을 동일한 Directory에 위치

  • 도스모드 창에서 실행

  • 정상 동작시 동일 Directory에 json 파일 생성



<그림 2-162> 수집 모듈 실행 화면

- 데이터 수집

  • 수집 대상은 특정시간에 이슈화된 트렌드 목록을 기준

  • 수집 항목 : 트윗 사용자, 계정정보(screen_name, user_id), follower 여부

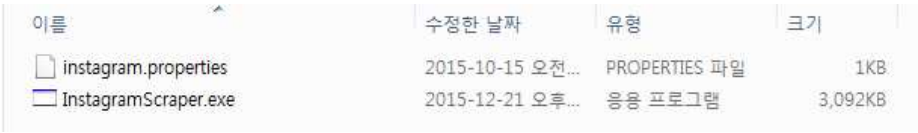
#	사용자	screen_name	user_id	followering	Trends
163	AndCat	@cat_cons	1521215323	N	북한 쪽지
164	KBS 뉴스	@KBSnews	164509454	N	지하철 1호선
165	선린	@goodboySR	1973977849	N	대본리딩
166	이제민	@love2poem	116169345	N	최순실 게이트
167	뉴스모아보기	@digestnewskr	755636719559598081	N	출입문 고장
168	Dae San Hwang	@daesanhwang	1159312776	N	주력 산업 경쟁력
169	경향신문	@kyunghyang	118720645	Y	김정일 4시간
170	Ｊａｃｋ Ｙｕｎ	@u21216	2949686616	N	열차 고장
171	ACCESS ent	@ACCESS_ent	142197815	N	메탈리카
172	대한민국 국회의원 표창원	@DrPyo	74048201	N	최순실 모녀
173	박범계	@bkfire1004	201168492	N	문재인 색깔론
174	?argaux	@Chateau_Margaux	1061444353	N	원성 1호기 수명 연장
175	산들바람	@Roangram	137244200	N	외교장관
176	내가 친노다	@gomirae	190595309	N	작년 고작 26억
177	이 도	@tpwhd28	119946531	N	남북정상회담 추진
178	Jee-Hyun Ha	@jhnha	118640231	N	권석천의 시시각각
179	까칠한 도담파파	@csswook	123430989	N	그림마당
180	위키트리 WIKITREE	@wikitree	47856818	Y	비속어 리포트 확점
181	❀월넛 하잇데루요❀	@kimdotori_KOP	775543586556411904	N	페미니즘
182	계간홀로~연애하지 않을 자유~	@quarterlyalone	1137531284	N	피해자분
183	skycat	@bleuchatte	47551286	N	하나의 자은택
184	굴희[플레티넘]	@femiforsave	735120002664136706	N	미지의 세계

<그림 2-163> 수집 정보 결과

❍ 인스타그램(Instagram) 수집 모듈

- 실행파일 구성

- JAVA 기반의 개발 도구를 이용하여 실행파일 생성
- 실행파일 : InstagramScraper.exe
- 환경설정파일 : instagram.properies



<그림 2-164> 인스타그램(Instagram) 실행파일 구성

- 환경설정 파일내용

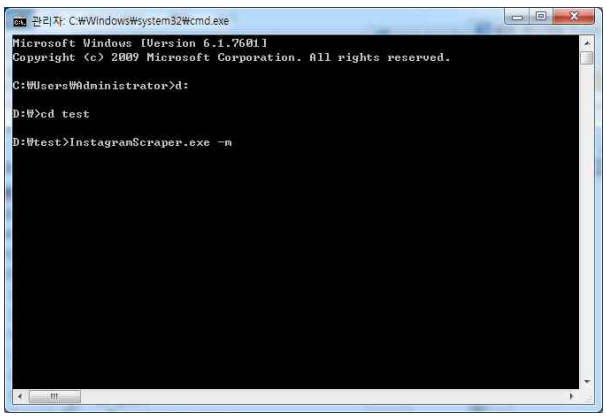
- 사용자 인증을 위한 key와 secret
- 조회 조건 입력

```
# Replace this with an API key available at https://instagram.com/developer/clients/manage/
instagram.clientid=7fa6515f[redacted]
instagram.clientsecret=aadb921db[redacted]
instagram.redirecturi=https://ins[redacted]
# 조회반경(단위m, 최대 5km임)
instagram.locationRadius=1000000
# 검색 기간 설정
# 검색 월, maxtimestamp에서 앞으로
instagram.months=1
# 검색 기준일, maxtimestamp에서 입력한 months 값 만큼 앞의 달 수를 mintimestamp로 정하여 정보를 조회
instagram.maxtimestamp=2015-10-15
```

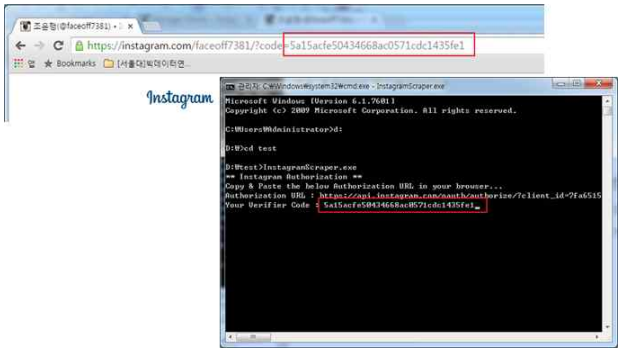
<그림 2-165> 인스타그램(Instagram) 환경설정 파일 예시

- 실행

- 실행에 필요한 파일들을 동일한 Directory에 위치
- 도스모드 창에서 실행
- 사용 가능여부를 확인하기 위한 인증코드 획득의 추가 단계 수행



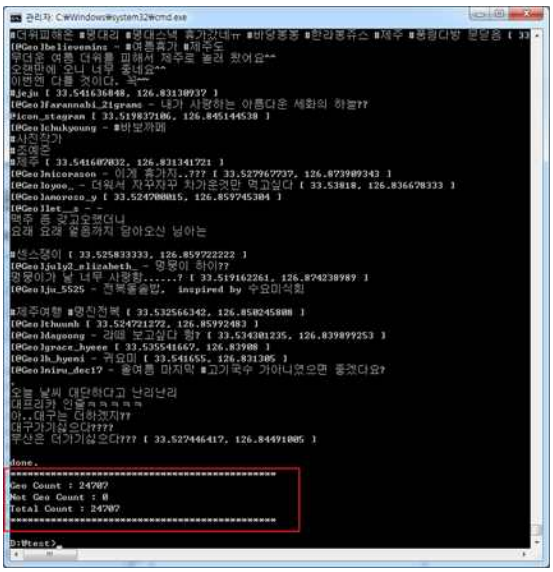
<그림 2-166> 인스타그램(Instagram) 수집 모듈 실행 화면



<그림 2-167> 인스타그램(Instagram) 수집 모듈 인증 화면

- 데이터 수집

- 사용자 이름, 등록내용 및 공간정보 유무와 수집된 전체 개수를 결과로 표시
- 수집되는 데이터는 JSON 형태의 파일로 생성

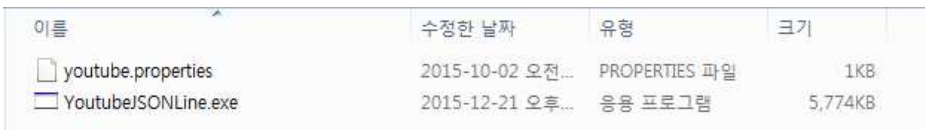


<그림 2-168> 인스타그램(Instagram) 수집 화면

❍ 유투브(Youtube) 수집 모듈

- 실행파일 구성

- JAVA 기반의 개발 도구를 이용하여 실행파일 생성
- 실행파일 : YoutubeJSONLine.exe
- 환경설정파일 : youbube.properies



<그림 2-169> 유투브(Youtube) 실행파일 구성

- 환경설정 파일내용

- 사용자 인증을 위한 API key

• 조회 조건 입력

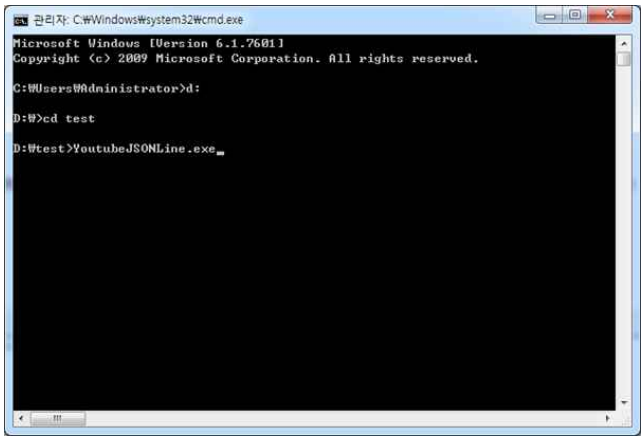
```
# Replace this with an API key available at code.google.com/apis/console
youtube.apikey=AIzaSyA4[redacted]
# 조회위치 : 서울특별시청 좌표
youtube.location=37.566663,126.978422
# 반경 (최대 1000km임)
youtube.locationRadius=1000km
# 검색 단위는 초(second), 60 * 60 = 3600(1시간)
youtube.term=3600
# timeAfter(포함)와 timeBefore(미포함) 해당 기간 동안 term 단위로 list search api 접속을 시도.
youtube.timeAfter=2015-10-01T00:00:00Z
youtube.timeBefore=2015-11-01T00:00:00Z
# nextPageToken 값의 사용 유무.(0:미사용, 1:사용)
youtube.nextPageTokenNotUse=1
```

<그림 2-170> 유투브(Youtube) 환경설정 파일 예시

- 실행

• 실행에 필요한 두 개의 파일을 동일한 경로에 위치

• 도스모드 창에서 실행

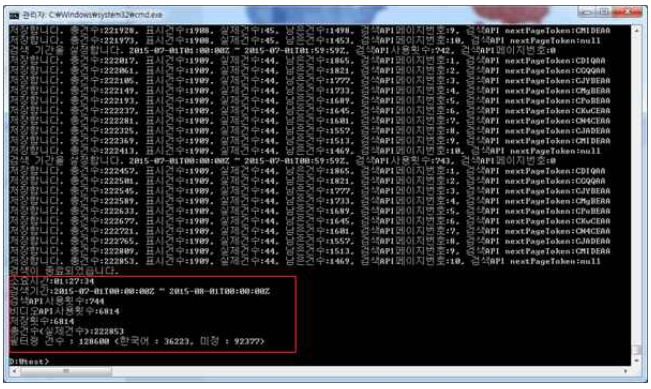


<그림 2-171> 유투브(Youtube) 수집 모듈 실행 화면

- 데이터 수집

• 소요시간, 검색기간, 수집개수 등을 결과로 표시

• 수집되는 데이터는 JSON 형태의 파일로 생성



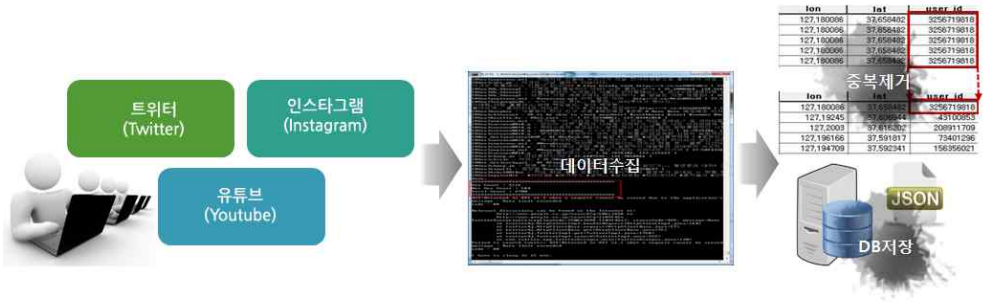
<그림 2-172> 유투브(Youtube) 수집 화면

○ 수집 데이터 DB 적재 및 중복 제거

- 수집된 소셜 미디어 데이터를 DB 서버에 저장하여 사용자 클라이언트에서 필터

링을 통한 공간 정보 서비스를 제공하고, 수집원의 계속적인 데이터 수집에 의한 데이터 증가로 인해 DB 적재용 스케쥴링을 구축하여 반자동화 시스템을 운영할 수 있도록 제공

- 공간 빅데이터 적재를 위한 데이터베이스 설계
- DB 업데이트를 위한 스케쥴링 관리
- 소셜 미디어 데이터의 동일 id에 대한 중복 데이터 제거



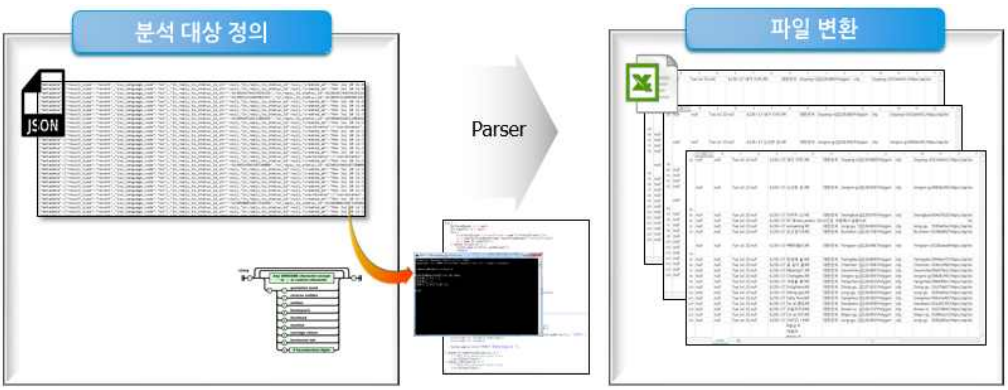
<그림 2-173> 수집 데이터 DB 적재 흐름

㉯ 데이터 변환 모듈 개발

❍ 데이터 분석을 위한 CSV 파일 변환

- 수집된 소셜 미디어 데이터 중 공간분석을 위한 대상 필드를 선정하여 CSV 형태의 파일로 변환할 수 있도록 제공

- 공간분석을 위한 대상 필드 선정
- JSON Parser 기술 적용
- 추출된 공간 정보를 CSV 형태의 파일로 변환



<그림 2-174> 파일변환(JSON->CSV) 흐름

<표 2-56> 공간분석을 위한 대상 필드

구분	위치정보	기타정보
트위터	• created_at • coordinates	• id • hashtags

	• place • withheld_in_countries	• text • retweet_count • favorite_count
인스타그램	• location_id • longitude • latitude • location_name	• type • users_in_photo • comments • created_time • likes
유투브	• regionRestriction • locationDescription • location • recordingLocation	• id • title • description • viewCount • creationTime • publishedAt

❍ V-World의 Geocoder API 사용

- VWorld는 지도검색 API 레퍼런스에서 좌표로 주소를 변환할 수 있도록 Geocoder API를 제공하며, 일일 지오코딩 요청건수는 최대 3만건으로 제한되어 있음

• 수집된 소셜 미디어 공간 빅데이터를 관리하기 위해 DB에 적재할 경우 좌표 데이터를 이용하여 주소로 변환할 수 있도록 사용

• 공간 빅데이터 분석 시스템에서 필터링의 위치 검색시 지도를 선택하면 선택된 위치에 주소를 표시할 수 있도록 사용



<그림 2-175> V-World의 Geocoder API 기능

㉮ 데이터 필터링 모듈 개발

❍ 공간 정보 필터링을 위한 기능 정의

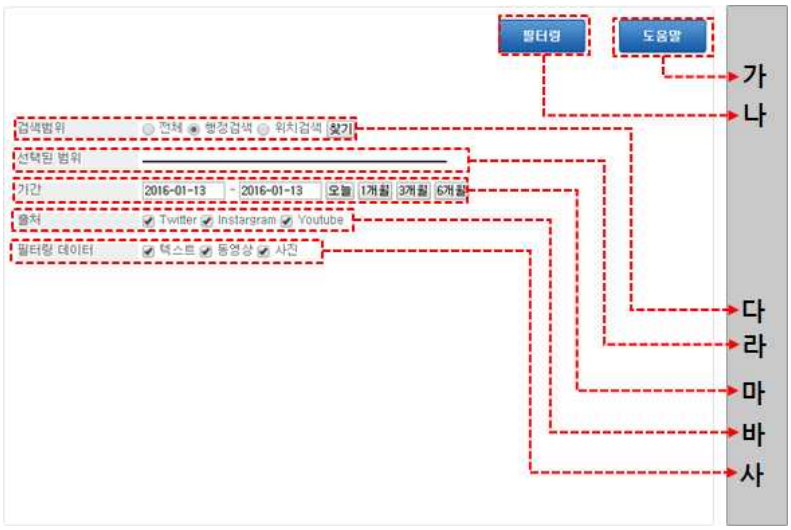
<표 2-57> 필터링 기능 정의

구분	적용부분	지오코딩
검색범위	• 검색된 위치의 선택 범위 지정	전체/행정검색/위치검색
선택된 범위	• 행정검색 및 위치검색의 선택된 위치	목록 표시
기간	• 검색 기간 설정	오늘/1개월/3개월/6개월

출처	• 소셜 미디어 데이터 선택	트위터/인스타그램/유투브
필터링 데이터	• 멀티미디어 유형	텍스트/동영상/사진

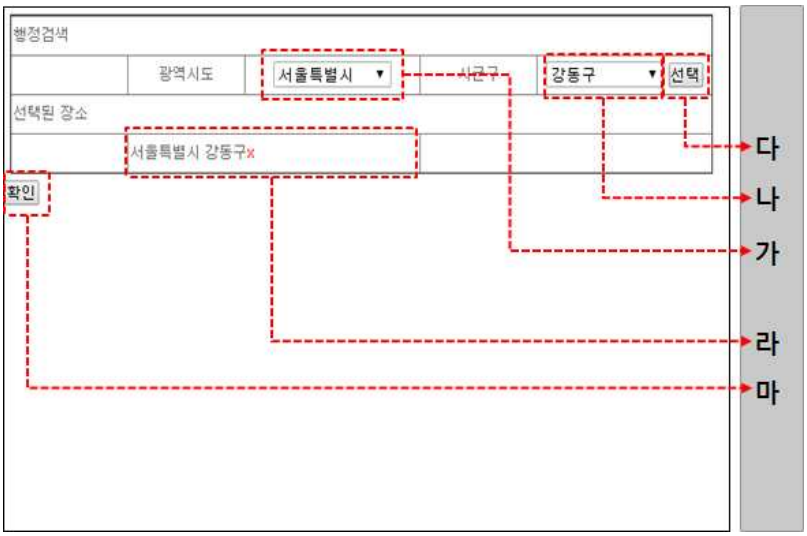
❍ 필터링 UI 및 시스템 구축

- 필터링 설정



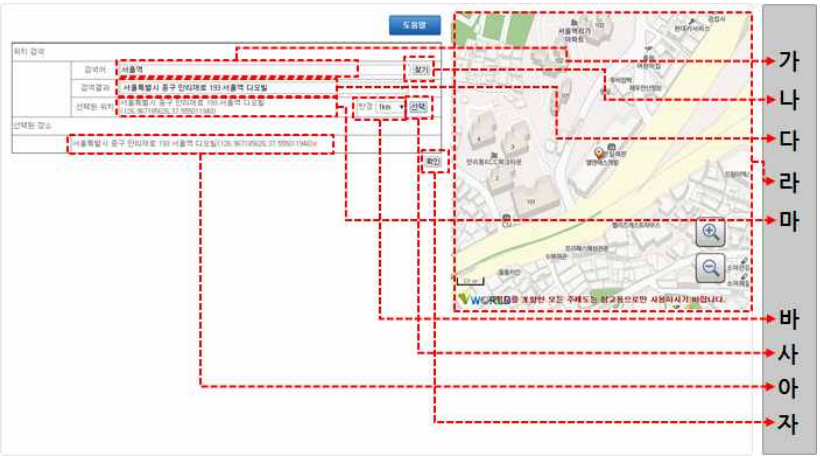
<그림 2-176> 필터링 설정 화면

- 행정검색화면



<그림 2-177> 행정검색 설정 화면

- 위치검색화면



<그림 2-178> 검색어 입력에 의한 위치검색 설정 화면

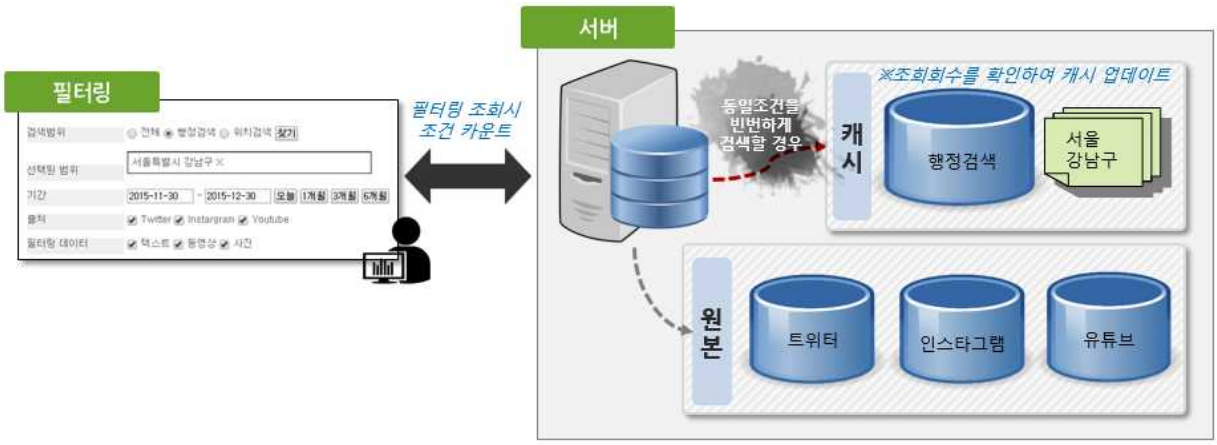
⑬ [단위성과 E-6] 다중 레이어를 위한 캐싱/프로패칭/버퍼링 모듈 개발(2차 연도)

㉮ 다중 레이어를 위한 캐싱 모듈 개발

○ 공간 빅데이터 분석 시스템 구축시 사용자 클라이언트에서 필터링에 의한 소셜 미디어 공간 빅데이터의 조회 결과가 많아질 경우 수행 속도가 늦어지며, 동일한 조건을 빈번하게 조회하는 행위 또한, 수행 속도는 반복적으로 떨어지게 됨. 이에, 서버에서는 클라이언트에 빠른 정보를 제공하기 위해 캐시 모듈을 탑재하여 효과적인 실시간 분석을 할 수 있도록 고도화 함

○ 캐싱 모듈 개발에 의한 주요 내용은 아래와 같음

- 동일한 조건을 빈번하게 필터링할 경우 수행속도를 줄이기 위해 제공
- DBMS의 메모리 테이블을 이용하여 별도의 데이터 저장
- 캐시의 업데이트를 위한 스케쥴링 관리



<그림 2-179> 캐시 모듈 구조

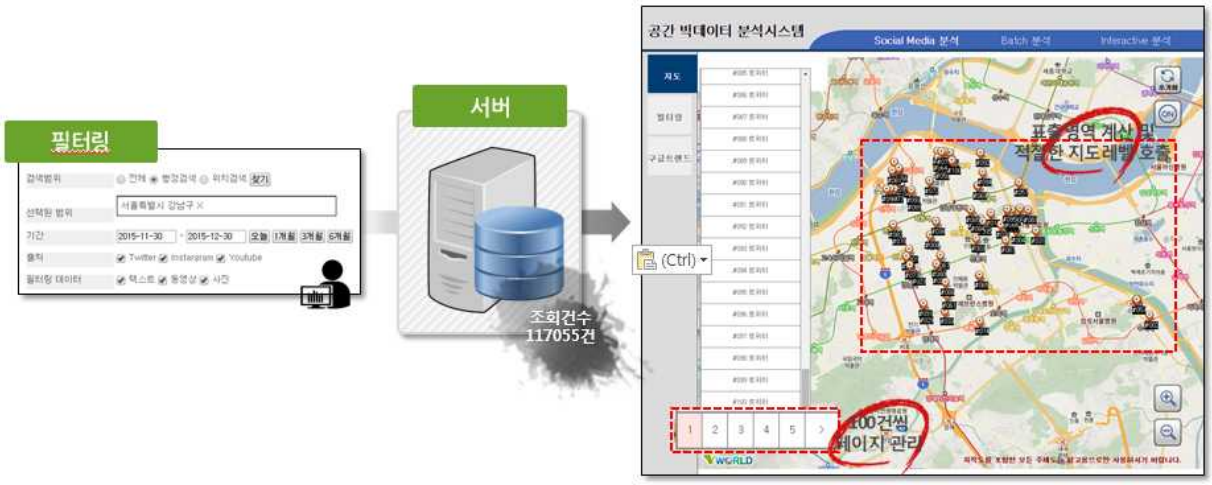
㉯ 다중 레이어를 위한 버퍼링 모듈 개발

○ 사용자 클라이언트가 필터링에 의한 소셜 미디어 공간 빅데이터의 조회 결과

전체를 받아 공간 정보를 지도에 표시할 경우 데이터 양이 많을수록 표출하기 위한 데이터의 시인성 및 수행 속도의 문제가 발생할 수 있음. 이를 위해 클라이언트는 페이지당 표출 정보를 관리하여 지도 표출에 대한 효과적인 실시간 분석을 할 수 있도록 고도화 함

○ 버퍼링 모듈 개발에 의한 주요 내용은 아래와 같음

- 표출을 위한 관리 페이지에서 현재 선택된 페이지만 지도 표출을 위해 제공
- 표출할 정보들의 영역을 계산하여 적절한 지도 레벨 표시



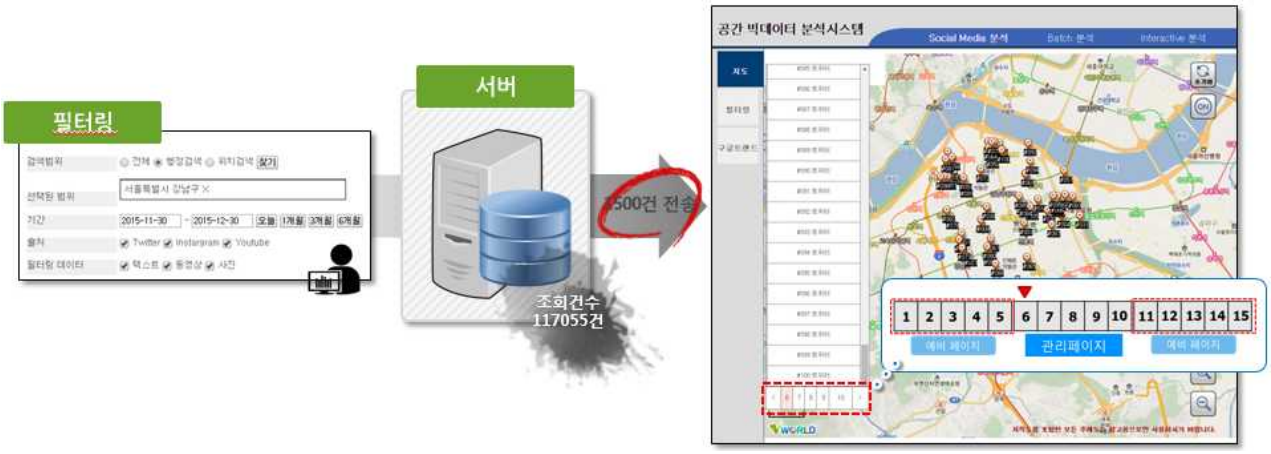
<그림 2-180> 버퍼링 모듈 구조

㉮ 다중 레이어를 위한 프리패칭 모듈 개발

○ 사용자 클라이언트는 버퍼 모듈에 의해 관리 페이지 만큼의 데이터를 갖고 있기 때문에 관리 페이지 외에 다른 페이지를 선택시 서버에 추가 요청을 해야 함. 때문에, 이를 반복적으로 수행하게 되면 속도 저하의 문제가 발생하여 표출 되고 있는 관리 페이지 외에 예비 페이지가 필요하게 됨. 프리패칭 모듈은 이를 해결하기 위해 예비 페이지를 추가로 전송하여 효과적인 실시간 분석을 할 수 있도록 고도화 함

○ 프리패칭 모듈 개발에 의한 주요 내용은 아래와 같음

- 수시로 서버에 정보 요청시 발생되는 속도 저하를 해결하기 위해 예비 페이지 관리
- 표출을 위한 관리 페이지를 기준으로 전후의 예비 페이지 관리



<그림 2-181> 프리패칭 모듈 구조

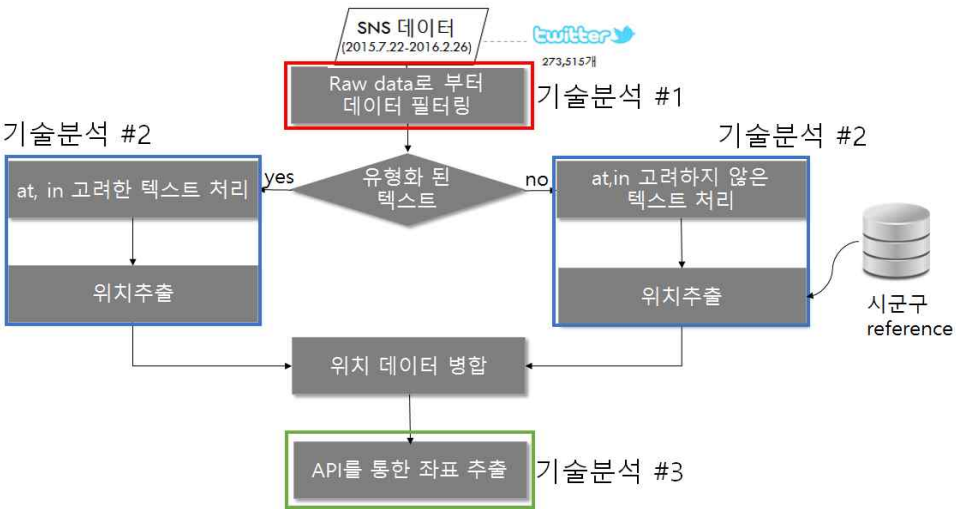
⑭[단위성과 E-14] 공간 빅데이터 정합 및 주제도 분석 모듈 개발(3차연도)

○ 공간 빅데이터 정합 및 주제도 분석 기술

- 소셜 미디어 데이터를 이용한 사용자 관심도, 이슈, 사회현상 등의 주제별 테마를 분석하기 위한 대표적인 정보는 위치좌표 및 텍스트 분석으로 현재까지 수집된 대부분의 공간 정보 빅데이터는 80%이상이 위치정보를 포함하지 않고 있으며, 텍스트 분석을 위한 방법이 필요하게 됨
- 이를 위한 대안으로 다음과 같은 주요 기술 요소를 개발
  - 소셜 미디어 데이터에서 위치정보를 추출하여 이종 DB와 정합하는 기술에 따른 모듈 개발
  - 소셜 미디어의 자연어 처리 및 웹크롤링에 의한 다양한 주제의 데이터 분석을 위한 알고리즘 기술에 따른 모듈 개발

㉮ 정합 모듈의 기술 분석

○ 소셜 미디어 데이터에서 위치정보를 추출하기 위한 기술 분석은 다음과 같음



<그림 2-182> 정합 모듈 분석 흐름도

❍ 데이터 필터링 기술 (기술분석 #1)

- 수집된 소셜 미디어 데이터를 읽어온 후, 중복 정보를 제거하고 위치정보로 판단되는 지오태킹된 데이터와 그렇지 않은 데이터를 구분

<표 2-58> R 코딩을 이용한 데이터 필터링 기술 분석

```
rawdata <- data.frame(text = rawdata$text[!duplicated(rawdata$text)] ,coordinates = rawdata$coordinates.coordinates[!duplicated(rawdata$text)] , stringsAsFactors = FALSE)
data <- cbind(text = rawdata$text[rawdata$coordinates == " "] )
```

❍ 위치 추출 기술 (기술분석 #2)

- 소셜 미디어 데이터의 텍스트에 포함되어 있는 유형화된 위치정보 추출

<표 2-59> R 코딩을 이용한 위치 추출 기술 분석 (POI 포함)

```
mat_place <- matrix(nrow = nrow(data), ncol =1)

remove_at <- function(tempstring){
  tempstring <- unlist(strsplit(tempstring, " " ))
  i <- 1
  while(i <= length(tempstring)){
    if(length(grep("@+" , tempstring[i])) !=0 & length(grep("@$",tempstring[i])) == 0) {
      tempstring <- tempstring[-i]
      i <- i - 1
    }
    i <- i + 1
  }
  tempstring <- paste0(tempstring , collapse = " ")
  return(tempstring)
}
```

<표 2-60> R 코딩을 이용한 위치 추출 기술 분석 (POI 제외)

```
# 위치 추출 레퍼런스로 사용되는 전국 시군구 코드.csv 파일
code[4] <- substr(code[[3]],1,nchar(code[[3]])-1)
# 군구명2 를 만드는 과정으로, 군구명2는 군구명에서 마지막 글자 한 글자를 뺀 것임.
code[5] <- paste(code[[2]], code[[3]], sep = " ")
df_count <- data.frame(matrix(nrow=length(df3[[1]]),ncol=length(code[[1]])))
# 행의 개수가 위치 추출할 데이터의 개수이고, 열의 개수가 시군구코드의 개수인 데이터프레임을 생성
code[[6]][is.na(code[[6]])] <- 0     # 중복여부
code[[7]][is.na(code[[7]])] <- 0     # 두글자 이하
# 1이 입력되지 않은 셀을 0으로 채워줌.

mat_place2 <- data.frame(df3 , sigunguname , sigungu , stringsAsFactors = FALSE) # 시군구만 선택했을시(POI 체크박스 선택하지 않았을 때)
```

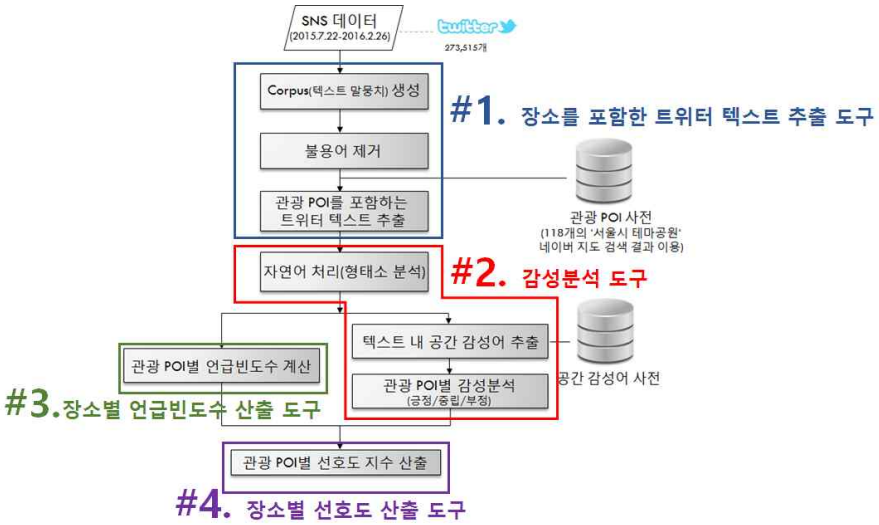
❍ 좌표 추출 기술 (기술분석 #3)

<표 2-61> R 코딩을 이용한 좌표 추출 기술 분석

```
geocode_result = c()
for(loc in loc_list){
  if(!is.na(loc)){
    # 검색어를 제외한 url
    url = "http://maps.googleapis.com/maps/api/geocode/xml?sensor=false&language=ko&address="
    geocode_url = paste0(url, loc, "%27" )
                        geocode_xml = read_xml(geocode_url, encoding='UTF-8')
                        geocode_lat = geocode_xml %>% xmlNode('lat') %>% xml()
                        geocode_lon = geocode_xml %>% xmlNode('lng') %>% xml_text()
                        # 지명, 위도, 경도를 1행짜리 데이터 프레임으로 구성한다
                        # 이 때 위도,경도값을 숫자로 변환한다
                        geocode_data = data.frame(lat = as.numeric(geocode_lat), lon =
as.numeric(geocode_lon))}
}
final_place <- cbind(mat_place2, geocode_result$lat , geocode_result$lon)
```

㉯ 주제도 분석 모듈의 기술 분석

○ 소셜 미디어 데이터에서 자연어 처리 및 주제도 분석을 위한 기술 분석은 다음과 같음



<그림 2-183> 주제도 모듈 분석 흐름도

○ 장소를 포함한 소셜 미디어의 텍스트 추출 기술

- 수집된 소셜 미디어 데이터를 읽어온 후, Semantic matching을 통해 장소를 포함하는 텍스트 추출

<표 2-62> R 코딩을 이용한 장소를 포함한 소셜 미디어의 텍스트 추출 기술 분석

```
myM<-matrix(ncol = 3) # 100% 매칭되어 얻어진 텍스트 모음
myMM<-matrix(ncol = 3) # 100% 매칭되어 얻어진 개별 텍스트
myFM<-matrix(ncol = 3) # 레빈티쉬 편집거리 적용하여 얻어진 개별 텍스트
mySM<- matrix(ncol = 3) # 레빈티쉬 편집거리 적용하여 얻어진 텍스트 모음
colnames(myM)<-c('POI','Text','Morph')
colnames(myMM)<-c('POI','Text','Morph')
```

```
colnames(myFM)<-c('POI','Text','Morph')
colnames(mySM)<-c('POI','Text','Morph')
for(i in seq(1,length(poi_park))){
  for(j in seq(1,length(ntext))){
    if(str_detect(ntext[j],poi_park[i])){ # 테마공원 poi와 100%매칭되는 텍스트 추출하기
      myMM[,1]<- poi_park[i]
      myMM[,2]<- ntext[j]
      myMM[,3]<- toString(SimplePos22(ntext[j]))
      myM<-rbind(myM,myMM)
      print(myMM) }
  }
}

myM<- myM[-1,]
write.csv(myM,file = 'Twitter_including테마공원.csv')
```

❍ 감성분석 기술

- 공간 감성어 사전을 활용하여 장소를 포함한 소셜 미디어 텍스트 추출 결과에서 형태소별로 세분화하여 공간 감성어를 추출하고, 긍정, 중립, 부정의 감성방향과 감성확률을 부여 및 감성점수를 산출

<표 2-63> R 코딩을 이용한 감성분석의 기술 분석

```
#감성어 추출하기
myM<- read.csv('Twitter_including테마공원.csv',header = T,stringsAsFactors = F)
myM<- myM[,-1]

#1. 3가지 문장 유형에 따라 감성어 추출을 위한 table 만들기
testMM<- cbind(myM,matrix(nrow= nrow(myM),ncol = 13))
colnames(testMM)<-c('POI','Text','Morph','1번 유형_주어(POI/DPN)','1번 유형_부정어(Ms)','1번 유형_감성어(SP)','2번 유형_주어(SS)','2번 유형_부정어(Ms)','2번 유형_감성어(SP)','3번 유형_주어(SS)','3번 유형_부정어(Ms)','3번 유형_감성어(많/적다)','PT','CP','NT','CN')
# PT: Positive index of Text, CP: COUNT of PT , NT: Negative index of Text, CN :COUNT of NT를 말함
class(testMM)
testMM <- as.data.frame(testMM)
testM<- testMM # 복사본 만들기

#5. 개별 텍스트의 긍정(PT) 및 부정지수(NT) 구하기
for(i in seq(1,nrow(testM))){
  if(testM$PT[i] != 0){
    testM$PT[i]<- testM$PT[i]/testM$CP[i]
  }
  if(testM$NT[i] != 0){
    testM$NT[i]<- testM$NT[i]/testM$CN[i]
  }
}

write.csv(testM,file='트위터에서 감성어 추출한 결과.csv')

#6.테마공원 poi별 PT,NT 합산하여 SP(Sentiment index of Positive), SN(Sentiment index of Negative)구하기
testM$POI<- as.factor(testM$POI)
SPresult<- tapply(testM$PT,testM$POI,sum)
```

```
SPresult<- as.data.frame(SPresult)
SNresult<- tapply(testM$NT,testM$POI,sum)
SNresult<- as.data.frame(SNresult)

SPresult
SNresult
nres<-nrow(SPresult)

fina<- cbind(SPresult,SNresult)
fina<- cbind(matrix(rownames(SPresult)),fina)
fina<- cbind(fina,matrix(ncol = 2,nrow = nres))
rownames(fina)<- c(seq(1,nres))
colnames(fina)<- c('POI','SP','SN','언급빈도수','선호도')
fina<- as.data.frame(fina)
class(fina)
```

❍ 장소별 언급빈도수 산출의 기술 분석

<표 2-64> R 코딩을 이용한 장소별 언급빈도수 산출의 기술 분석

```
testM$POI<- as.character(testM$POI)
fina$POI<- as.character(fina$POI)
for(i in seq(1,nrow(fina))){
  fina$언급빈도수[i]<- sum(str_count(testM$POI,fina$POI[i]))
}
```

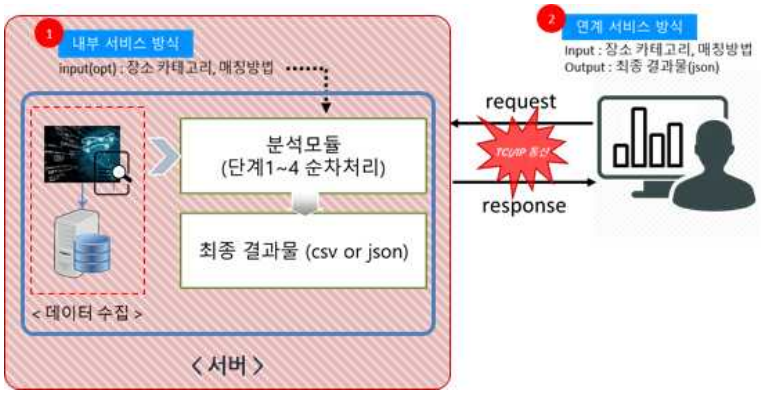
❍ 장소별 선호도 산출의 기술 분석

<표 2-65> R 코딩을 이용한 장소별 선호도 산출의 기술 분석

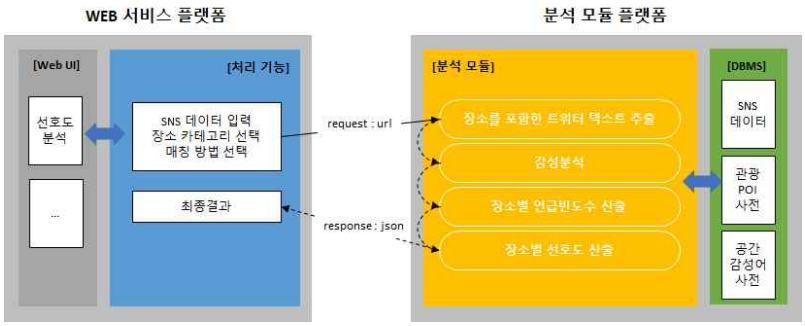
```
for(i in seq(1,nrow(fina))){
  if(fina$SP[i]==abs(fina$SN[i])){ #중립
    fina$선호도[i]<- 1
  }else{
    if(fina$SP[i]>abs(fina$SN[i])){    #긍정
      SD<- 1
    }else if(fina$SP[i]<abs(fina$SN[i])){  #부정
      SD<- -1
    }
    fina$선호도[i]<- SD*(1+abs(fina$SP[i]+fina$SN[i]))*fina$언급빈도수[i]
  }
}
```

㉰ 정합 및 주제도 분석 모듈의 OpenAPI 서비스

❍ 앞선 기술 분석에 의한 소셜 미디어 데이터 정합 및 주제도 분석 모듈을 개발하고 이에 따른 분석모듈 플랫폼과 서비스 플랫폼 간의 연동을 위한 OpenAPI 구축 및 서비스 제공



<그림 2-184> OpenAPI 처리 흐름도



<그림 2-185> OpenAPI 연동 규격 시스템 구조

<표 2-66> OpenAPI 연동 규격서

(가) 요청 목록

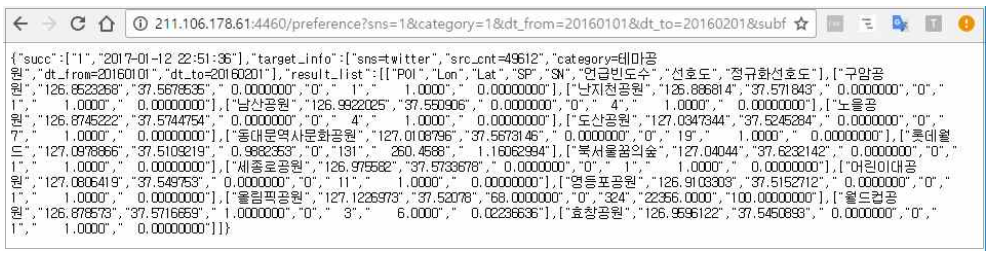
파라미터	데이터타입	설명	예
sns_cd	INTEGER	수집된 SNS 데이터를 선택 1:twitter, 2:Instagram, 3:youtube	sns_cd=1
date_from	STRING	조회 시작 날짜 (YYYYMMDD)	
date_to	STRING	조회 종료 날짜 (YYYYMMDD)	
category_cd	INTEGER	장소 카테고리 선택 1:테마공원, 2:문화재, 3:음식점, 4:카페, 5:숙박	category_cd=1
match_cd	INTEGER	매칭 방법 선택 1:100%일치, 2:100%일치 및 편집거리 적용	match_cd=1

(나) 응답 목록

필드	데이터타입	설명
succ	STRING	성공여부 ('Y' or '1':성공, 'N' or '2':실패)
result_id	STRING	출력 ID (YYYMMDDHHMMSS)
result_list.poi_nm	STRING	장소명 (테마공원 POI)
result_list.xcrd	STRING	경도
result_list.ycrd	STRING	위도
result_list.scale	STRING	감성점수
result_list.freq	STRING	언급빈도수
result_list.rate	STRING	선호도

○ 시험 테스트

- http://localhost:4460/preference?sns=1&category=1&dt_from=20160101 &dt_to=20160201&subfn=4



<그림 2-186> OpenAPI 응답 결과 화면

⑮ [단위성과 E-15, E-16, E-17, E-18] 공간 빅데이터 분포 및 차트 분석 모듈 개발과 고도화 (4, 5차연도)

㉮ 공간 빅데이터 분포 및 차트 서비스 모듈 개발

- ○ 이모티콘 기반 장소 선호도 분석 모듈
- ○ 트위터 데이터의 GPS 정보 추정 모듈
- ○ 개체명 인식을 활용한 비정형 텍스트 데이터 위치탐지 모듈
- ○ 딥러닝 기반 징소명 대체어 분석 모듈
- ○ 공공시설 매력도 측정 및 접근성 분석 모듈
- ○ 사용자 중점 기반 시간지도 생성 모듈
- ○ 머신러닝 기반 장소 감성어 사전 구축 모듈
- ○ 영어 장소리뷰로부터 장소 테마 분석 모듈
- ○ 머신러닝 기반 문서 분류 모듈
- ○ 머신러닝 기반 문장 분류 모듈 및 고도화
- ○ Key Player 탐지 모듈

㉯ 공간 빅데이터 분포 분석 및 고도화 모듈 개발

- ○ 소셜 미디어 데이터를 이용한 공간 빅데이터 분석 서비스 모듈을 통해 분석된 결과를 분포 분석에 사용하기 위하여 모듈을 구현
- ○ 이를 위하여 다음과 같은 주요 기술 요소를 개발
  - 분석 서비스 모듈을 이용하여 분포 표출용 데이터 생성 및 DBMS 관리
  - 행정동 단위의 공간적 분포 분석 기술 제공
  - AI 기술이 적용된 다양한 분포 분석 기술 제공



<그림 2-187> 공간 빅데이터 분포 분석 모듈의 시스템 구조도

- ○ 분포 분석 모듈의 데이터베이스
  - 테이블 목록

<표 2-67> 분포 분석 모듈의 테이블 목록

순번	테이블ID	테이블명
1	tb_lcode_theme_poi	관광 POI 사전 (118개 서울시 테마공원)
2	tb_lcode_addr	행정동 단위의 지역코드 및 경위도 중심좌표
3	tb_lcode_sgg	시군구 단위의 지역코드 및 경위도 중심좌표
4	tb_result_prefer_disp	분포 표출용 정보 (서비스 분석 모듈 결과 및 시군구별 중심좌표)

- 테이블 상세 목록

• 관광 POI 사전 (118개 서울시 테마공원)

<표 2-68> 관광 POI 사전의 테이블 상세 목록

TABLE ID		tb_lcode_theme_poi	테이블명	관광 POI 사전		
NO	영문명	한글명	TYPE	KEY	NULL	비고
1	poi	관광 POI	VARCHAR(50)		NOT NULL	
2	lat	위치좌표(위도)	VARCHAR(50)			
3	lon	위치좌표(경도)	VARCHAR(50)			
4	lcode	지역코드	VARCHAR(20)			
5	addr	주소	VARCHAR(50)			

• 행정동 단위의 지역코드 및 경위도 중심좌표

<표 2-69> 행정동 단위의 테이블 상세 목록

TABLE ID		tb_lcode_addr	테이블명	행정동 지역코드		
NO	영문명	한글명	TYPE	KEY	NULL	비고
1	sido	시도명	VARCHAR(20)		NOT NULL	
2	sigungu	시군구명	VARCHAR(20)			
3	ldong	행정동명	VARCHAR(20)			
4	lcode	지역코드	VARCHAR(20)			행정동 기준
5	cenx	중심좌표(경도)	VARCHAR(20)			
6	ceny	중심좌표(위도)	VARCHAR(20)			

• 시군구 단위의 지역코드 및 경위도 중심좌표

<표 2-70> 시군구 단위의 테이블 상세 목록

TABLE ID		tb_lcode_sgg	테이블명	시군구 지역코드		
NO	영문명	한글명	TYPE	KEY	NULL	비고
1	sido	시도명	VARCHAR(20)		NOT NULL	
2	sigungu	시군구명	VARCHAR(20)			
3	ldong	행정동명	VARCHAR(20)			

4	lcode	지역코드	VARCHAR(20)			시도 시군구 행정동 기준
5	cenx	중심좌표(경도)	VARCHAR(20)			
6	ceny	중심좌표(위도)	VARCHAR(20)			

• 분포 표출용 정보 (서비스 분석 모듈 결과 및 시군구별 중심좌표)

<표 2-71> 분포 표출용 테이블 상세 목록

TABLE ID		tb_result_prefer_disp	테이블명	분포 표출용 정보		
NO	영문명	한글명	TYPE	KEY	NULL	비고
1	seq_id	일련번호	BIGINT	PK	NOT NULL	
2	dt_from	조회시작날짜	VARCHAR(20)			
3	dt_to	조회종료날짜	VARCHAR(20)			
4	disp_type	분포유형	VARCHAR(20)			
5	disp_name	지역 및 장소명	VARCHAR(20)			
6	disp_cnt	분포지수	VARCHAR(20)			
7	sp	긍정지수	VARCHAR(10)			
8	sn	부정지수	VARCHAR(10)			
9	cenx	중심좌표(경도)	VARCHAR(20)			
10	ceny	중심좌표(위도)	VARCHAR(20)			
11	reg_dt	기록날짜	DATE			

- 데이터 현황

• 관광 POI 사전 (118개 서울시 테마공원)

poi	lat	lon	lcode	addr
도산공원	37.52452837	127.0347344	1168010700	서울특별시 강남구 신사동 649-9
월드컵공원	37.57166587	126.878573	1144012700	서울특별시 마포구 상암동 481-6
서대문독립공원	37.57369591	126.9564513	1141010900	서울특별시 서대문구 현저동 101
허브천문공원	37.53712688	127.153134	1174010600	서울특별시 강동구 둔촌동 산94
효창공원	37.5450893	126.9596122	1117011900	서울특별시 용산구 효창동
파리공원	37.53482381	126.8760578	1147010200	서울특별시 양천구 목동 906
서서울호수공원	37.52769006	126.8301276	1147010300	서울특별시 양천구 신월동 산68-3
탑골공원	37.57086993	126.9879816	1111013800	서울특별시 종로구 종로2가 38-1
노을공원	37.57447542	126.8745222	1144012700	서울특별시 마포구 상암동
남산공원	37.55090601	126.9922025	1114014200	서울특별시 중구 예장동 산5-85 남산제1별관
와룡공원	37.59066071	126.9898657	1111017300	서울특별시 종로구 명륜3가 산2-14
동대문역사문화공원	37.56731457	127.0108796	1114016200	서울특별시 중구 신당동
몽마르뜨공원	37.4948491	127.0033973	1165010700	서울특별시 서초구 반포동
남산골 전통정원	37.55785009	126.9937279	1114013800	서울특별시 중구 필동2가 84-1
학동공원	37.51589086	127.025601	1168010800	서울특별시 강남구 논현동
오목공원	37.52764686	126.8729515	1147010200	서울특별시 양천구 목동
천호공원	37.54432959	127.1260777	1174010900	서울특별시 강동구 천호동 360
낙성대공원	37.47121184	126.9599812	1162010100	서울특별시 관악구 봉천동 228
장충단공원	37.55794731	127.0041312	1114014400	서울특별시 중구 장충동2가 196
서초문화예술공원	37.46749186	127.0302213	1165010200	서울특별시 서초구 양재동 200
용마폭포공원	37.57306379	127.0903092	1126010100	서울특별시 중랑구 면목동 산1-4
앙카라공원	37.51757779	126.929963	1156011000	서울특별시 영등포구 여의도동 56
대현산배수지공원	37.55709179	127.0214645	1120010900	서울특별시 성동구 금호동1가 산1-72
영등포공원	37.5152712	126.9103303	1156010100	서울특별시 영등포구 영등포동
서울창포원	37.68948201	127.0476507	1132010800	서울특별시 도봉동 4

<그림 2-188> 관광 POI 사전의 데이터 현황

• 행정동 단위의 지역코드 및 경위도 중심좌표

sido	sigungu	ldong	lcode	cenx	ceny
서울특별시	종로구	청운동	1111010100	126.96932	37.58921
서울특별시	종로구	신교동	1111010200	126.96794	37.58451
서울특별시	종로구	궁정동	1111010300	126.9727	37.5847
서울특별시	종로구	효자동	1111010400	126.97195	37.58267
서울특별시	종로구	창성동	1111010500	126.97247	37.58056
서울특별시	종로구	통의동	1111010600	126.97287	37.57807
서울특별시	종로구	적선동	1111010700	126.9738	37.57532
서울특별시	종로구	통인동	1111010800	126.97022	37.58016
서울특별시	종로구	누상동	1111010900	126.96273	37.5804
서울특별시	종로구	누하동	1111011000	126.96828	37.57929
서울특별시	종로구	옥인동	1111011100	126.96341	37.58354
서울특별시	종로구	체부동	1111011200	126.97087	37.57759
서울특별시	종로구	필운동	1111011300	126.96814	37.5773
서울특별시	종로구	내자동	1111011400	126.97174	37.57532
서울특별시	종로구	사직동	1111011500	126.96596	37.57541
서울특별시	종로구	도렴동	1111011600	126.97475	37.57313
서울특별시	종로구	당주동	1111011700	126.97401	37.57159
서울특별시	종로구	내수동	1111011800	126.97197	37.57333
서울특별시	종로구	세종로	1111011900	126.9768	37.58023
서울특별시	종로구	신문로1가	1111012000	126.97392	37.57019
서울특별시	종로구	신문로2가	1111012100	126.96923	37.57087
서울특별시	종로구	청진동	1111012200	126.98018	37.57145
서울특별시	종로구	서린동	1111012300	126.98042	37.56952
서울특별시	종로구	수송동	1111012400	126.98065	37.5737
서울특별시	종로구	중학동	1111012500	126.97968	37.57532
서울특별시	종로구	종로1가	1111012600	126.98017	37.57027
서울특별시	종로구	공평동	1111012700	126.98318	37.57131
서울특별시	종로구	관훈동	1111012800	126.98468	37.57416

<그림 2-189> 행정동 단위의 데이터 현황

• 시군구 단위의 지역코드 및 경위도 중심좌표

sido	sigungu	ldong	lcode	cenx	ceny
서울특별시			1100000000	126.97936	37.56428
부산광역시			2600000000	129.07806	35.17575
대구광역시			2700000000	128.60289	35.86833
인천광역시			2800000000	126.70717	37.45208
광주광역시			2900000000	126.91725	35.15961
대전광역시			3000000000	127.38667	36.34833
울산광역시			3100000000	129.31481	35.53606
세종특별자치시			3600000000	127.29575	36.58956
경기도			4100000000	127.01167	37.27056
강원도			4200000000	127.73172	37.88186
충청북도			4300000000	127.49261	36.63269
충청남도			4400000000	126.67311	36.65364
전라북도			4500000000	127.10803	35.81608
전라남도			4600000000	126.46217	34.81358
경상북도			4700000000	128.60136	35.88758
경상남도			4800000000	128.69403	35.23425
제주도			5000000000	126.49939	33.48647
서울특별시	종로구		1111000000	126.98139	37.56967
서울특별시	중구		1114000000	127.00011	37.56086
서울특별시	용산구		1117000000	126.96786	37.53536
서울특별시	성동구		1120000000	127.03883	37.55928
서울특별시	광진구		1121500000	127.08547	37.53553
서울특별시	동대문구		1123000000	127.04206	37.57086
서울특별시	중랑구		1126000000	127.09483	37.60306
서울특별시	성북구		1129000000	127.01981	37.58700
서울특별시	강북구		1130500000	127.02822	37.63686
서울특별시	도봉구		1132000000	127.04836	37.66656
서울특별시	노원구		1135000000	127.05956	37.65131
서울특별시	은평구		1138000000	126.93150	37.59933
서울특별시	서대문구		1141000000	126.93825	37.57647

<그림 2-190> 시군구 단위의 데이터 현황

• 분포 표출용 정보 (서비스 분석 모듈 결과 및 시군구별 중심좌표)

seq_id	dt_from	dt_to	disp_type	disp_name	disp_cnt	sp	sn	cenx	ceny	reg_dt
1	20170901	20170929	0	노을공원	1	0	0	126.8745222	37.5744754	2017-09-29
2	20170901	20170929	0	동대문역사문화공원	1	0	0	127.0108796	37.5673146	2017-09-29
3	20170901	20170929	0	롯데월드	1	1	-1	127.0978866	37.5109219	2017-09-29
4	20170901	20170929	0	북서울꿈의숲	1	0	0	127.04044	37.6232142	2017-09-29
5	20170901	20170929	0	어린이대공원	1	0	0	127.0806419	37.549753	2017-09-29
6	20170901	20170929	0	올림픽공원	1	0	0	127.1226973	37.52078	2017-09-29
7	20170901	20170929	0	용마랜드	1	1	0	127.105646	37.594497	2017-09-29
8	20170901	20170929	0	월드컵공원	1	0	0	126.878573	37.5716659	2017-09-29
9	20170901	20170929	0	효창공원	1	0	0	126.9596122	37.5450893	2017-09-29
10	20170901	20170929	1	서울특별시 중구 신당동	1	0	0	127.01367	37.55792	2017-09-29
11	20170901	20170929	1	서울특별시 송파구 잠실동	1	1	-1	127.08339	37.51197	2017-09-29
12	20170901	20170929	1	서울특별시 강북구 번동	1	0	0	127.03675	37.62904	2017-09-29
13	20170901	20170929	1	서울특별시 광진구 능동	1	0	0	127.08173	37.55058	2017-09-29
14	20170901	20170929	1	서울특별시 송파구 방이동	1	0	0	127.1254	37.51568	2017-09-29
15	20170901	20170929	1	서울특별시 중랑구 망우동	1	1	0	127.10633	37.59959	2017-09-29
16	20170901	20170929	1	서울특별시 마포구 상암동	2	0	0	126.87999	37.57072	2017-09-29
17	20170901	20170929	1	서울특별시 용산구 효창동	1	0	0	126.96091	37.54281	2017-09-29
18	20170901	20170929	2	서울특별시 중구	1	0	0	127.00011	37.56086	2017-09-29
19	20170901	20170929	2	서울특별시 강북구	1	0	0	127.02822	37.63686	2017-09-29
20	20170901	20170929	2	서울특별시 광진구	1	0	0	127.08547	37.53553	2017-09-29
21	20170901	20170929	2	서울특별시 송파구	2	1	-1	127.10881	37.51267	2017-09-29
22	20170901	20170929	2	서울특별시 중랑구	1	1	0	127.09483	37.60306	2017-09-29
23	20170901	20170929	2	서울특별시 마포구	1	0	0	126.90356	37.56264	2017-09-29
24	20170901	20170929	2	서울특별시 용산구	1	0	0	126.96786	37.53536	2017-09-29
25	20170901	20170929	3	서울특별시	2	2	-1	126.97936	37.56428	2017-09-29
26	20170901	20170920	0	노을공원	1	0	0	126.8745222	37.5744754	2017-09-29
27	20170901	20170920	0	동대문역사문화공원	1	0	0	127.0108796	37.5673146	2017-09-29
28	20170901	20170920	0	롯데월드	1	1	-1	127.0978866	37.5109219	2017-09-29
29	20170901	20170920	0	북서울꿈의숲	1	0	0	127.04044	37.6232142	2017-09-29
30	20170901	20170920	0	어린이대공원	1	0	0	127.0806419	37.549753	2017-09-29

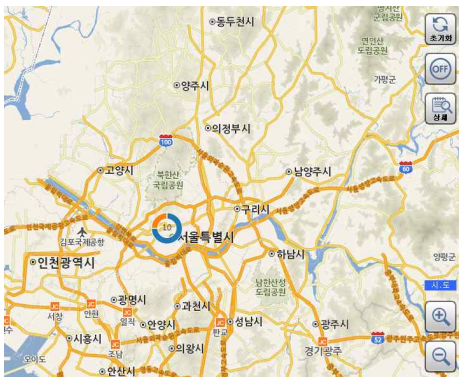
<그림 2-191> 분포 표출용 정보의 데이터 현황

❍ 분포 분석 모듈의 서비스 화면

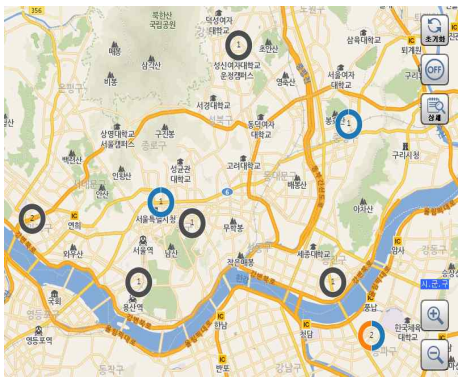
- 행정동 단위별 중심좌표 기준의 레벨관리

• 행정동 및 시군구 단위의 지역코드와 위치좌표를 데이터베이스화하고 분석 모듈 서비스 결과의 시설물 위치에 매칭하여 레벨관리

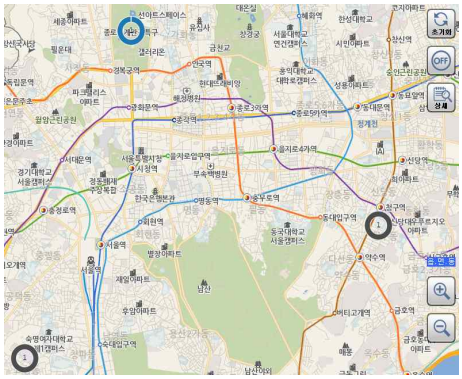
<표 2-72> 행정동 단위별 중심좌표 기준의 레벨관리



<그림 2-192> 시도 레벨



<그림 2-193> 시군구 레벨



<그림 2-194> 읍면동 레벨



<그림 2-195> 시설물 레벨

- 분포 분석 모듈에 의한 데이터 현황 및 상세 정보



분포유형	지역 및 시설명	분포지수	긍정지수	부정지수	위도	경도
0	노을공원	1	0	0	37.5744754	126.8745222
0	동대문역사문화공원	1	0	0	37.5673146	127.0108796
0	롯데월드	1	1	-1	37.5109219	127.0978866
0	북서물꿈의숲	1	0	0	37.6232142	127.04044
0	세종로공원	1	1	0	37.5733678	126.975582
0	어린이대공원	1	0	0	37.549753	127.0806419
0	올림픽공원	1	0	0	37.52078	127.1226973

<그림 2-196> 분포 데이터 현황 및 상세 정보

㉮ 공간 빅데이터 차트 분석 및 고도화 모듈 개발

○ 소셜 미디어 데이터를 이용한 공간 빅데이터 분석 서비스 모듈을 통해 분석된 결과를 차트 분석에 사용하기 위하여 모듈을 구현

○ 이를 위하여 다음과 같은 주요 기술 요소를 개발

- 분석 서비스 모듈을 이용하여 차트 표출용 데이터 생성 및 DBMS 관리
- 시계열 및 연결성 등의 다양한 분석 기술 제공
- AI 기술이 적용된 다양한 차트 분석 기술 제공



<그림 2-197> 공간 빅데이터 차트 분석 모듈의 시스템 구조도

❍ 차트 분석 모듈의 데이터베이스

- 테이블 목록

<표 2-73> 차트 분석 모듈의 테이블 목록

순번	테이블ID	테이블명
1	tb_result_prefer_chart	차트 표출용 정보 (서비스 분석 모듈 결과)

- 테이블 상세 목록

• 차트 표출용 정보 (서비스 분석 모듈 결과)

<표 2-74> 관광 POI 사전의 테이블 상세 목록

TABLE ID		tb_result_prefer_chart	테이블명	차트 표출용 정보		
NO	영문명	한글명	TYPE	KEY	NULL	비고
1	seq_id	일련번호	BIGINT	PK	NOT NULL	
2	sns_type	SNS종류	VARCHAR(10)			
3	src_cnt	정보추출 건수	VARCHAR(50)			
4	category	장소종류	INT			
5	matching	매칭방법	INT			
6	dt_from	조회시작날짜	VARCHAR(10)			
7	dt_to	조회종료날짜	VARCHAR(10)			
8	poi	장소명	VARCHAR(50)			
9	lon	위치좌표(경도)	VARCHAR(50)			
10	lat	위치좌표(위도)	VARCHAR(50)			
11	sp	긍정지수	VARCHAR(10)			
12	sn	부정지수	VARCHAR(10)			
13	ment_cnt	언급빈도수	VARCHAR(50)			
14	prefer_cnt	선호도 지수	VARCHAR(50)			
15	prefer_normal	정규화 선호도 지수	VARCHAR(50)			
16	reg_dt	기록날짜	DATE			

- 데이터 현황

• 차트 표출용 정보 (서비스 분석 모듈 결과)

seq_id	sns_type	src_cnt	category	matching	dt_from	dt_to	poi	lon	lat	sp	sn	ment_cnt	prefer_cnt	prefer_normal	reg_dt
1	1	562420	1	1	20170901	20170929	노을공원	126.8745222	37.5744754	0	0	1	1	0	2017-09-29
2	1	562420	1	1	20170901	20170929	동대문역사문화공원	127.0108796	37.5673146	0	0	3	1	0	2017-09-29
3	1	562420	1	1	20170901	20170929	롯데월드	127.0978866	37.5109219	1	-1	53	1	0	2017-09-29
4	1	562420	1	1	20170901	20170929	북서울꿈의숲	127.04044	37.6232142	0	0	1	1	0	2017-09-29
5	1	562420	1	1	20170901	20170929	어린이대공원	127.0806419	37.549753	0	0	2	1	0	2017-09-29
6	1	562420	1	1	20170901	20170929	홀림픽공원	127.1226973	37.52078	0	0	14	1	0	2017-09-29
7	1	562420	1	1	20170901	20170929	용마랜드	127.105646	37.594497	1	0	3	6	100	2017-09-29
8	1	562420	1	1	20170901	20170929	월드컵공원	126.878573	37.5716659	0	0	1	1	0	2017-09-29
9	1	562420	1	1	20170901	20170929	효창공원	126.9596122	37.5450893	0	0	1	1	0	2017-09-29
10	1	358887	1	1	20170901	20170920	노을공원	126.8745222	37.5744754	0	0	1	1	0	2017-09-29
11	1	358887	1	1	20170901	20170920	동대문역사문화공원	127.0108796	37.5673146	0	0	1	1	0	2017-09-29
12	1	358887	1	1	20170901	20170920	롯데월드	127.0978866	37.5109219	1	-1	46	1	0	2017-09-29
13	1	358887	1	1	20170901	20170920	북서울꿈의숲	127.04044	37.6232142	0	0	1	1	0	2017-09-29
14	1	358887	1	1	20170901	20170920	어린이대공원	127.0806419	37.549753	0	0	2	1	0	2017-09-29
15	1	358887	1	1	20170901	20170920	홀림픽공원	127.1226973	37.52078	0	0	12	1	0	2017-09-29
16	1	358887	1	1	20170901	20170920	용마랜드	127.105646	37.594497	0	0	1	1	0	2017-09-29
17	1	358887	1	1	20170901	20170920	월드컵공원	126.878573	37.5716659	0	0	1	1	0	2017-09-29
18	1	76110	1	1	20170901	20170905	롯데월드	127.0978866	37.5109219	0	-1	13	-26	-100	2017-09-29
19	1	76110	1	1	20170901	20170905	홀림픽공원	127.1226973	37.52078	0	0	3	1	0	2017-09-29
20	1	186318	1	1	20170901	20170910	동대문역사문화공원	127.0108796	37.5673146	0	0	1	1	0	2017-09-29
21	1	186318	1	1	20170901	20170910	롯데월드	127.0978866	37.5109219	0	-1	18	-36	-100	2017-09-29
22	1	186318	1	1	20170901	20170910	북서울꿈의숲	127.04044	37.6232142	0	0	1	1	0	2017-09-29
23	1	186318	1	1	20170901	20170910	어린이대공원	127.0806419	37.549753	0	0	2	1	0	2017-09-29
24	1	186318	1	1	20170901	20170910	홀림픽공원	127.1226973	37.52078	0	0	9	1	0	2017-09-29
25	1	277044	1	1	20170901	20170915	동대문역사문화공원	127.0108796	37.5673146	0	0	1	1	0	2017-09-29
26	1	277044	1	1	20170901	20170915	롯데월드	127.0978866	37.5109219	0	-1	29	-58	-100	2017-09-29
27	1	277044	1	1	20170901	20170915	북서울꿈의숲	127.04044	37.6232142	0	0	1	1	0	2017-09-29
28	1	277044	1	1	20170901	20170915	어린이대공원	127.0806419	37.549753	0	0	2	1	0	2017-09-29
29	1	277044	1	1	20170901	20170915	홀림픽공원	127.1226973	37.52078	0	0	10	1	0	2017-09-29
30	1	887812	1	1	20170801	20170929	노을공원	126.8745222	37.5744754	0	0	2	1	0.000000	2017-09-29

<그림 2-198> 차트 표출용 정보의 데이터 현황

○ 차트 분석 모듈의 서비스 화면

- 차트 분석 모듈에 의한 데이터 현황 및 상세 정보



<그림 2-199> 차트 데이터 현황 및 상세 정보

⑯ [단위성과 E-19] 공간 빅데이터 분석결과 가시화 모듈 개발(정적가시화)(5차연도)

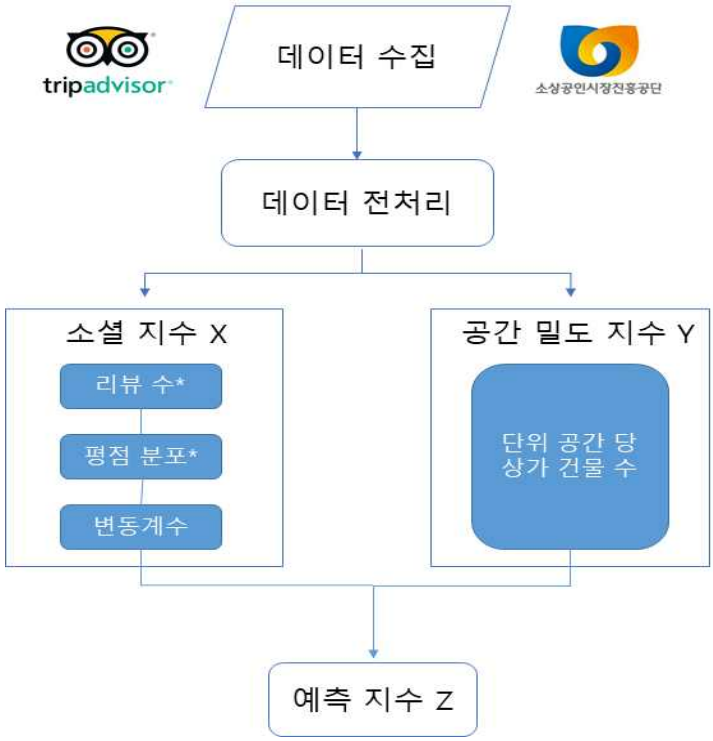
㉮ 리뷰 데이터 및 공간밀도 지수를 이용한 유동 인구 모듈

○ 빅데이터는 Long tail 분포를 주로 사용하여 기존 통계량으로 표현하는데 한

계가 있어 Head/tail breaks(H/T breaks) 기법을 통해 가시화 성능 향상 분석

❍ 장소 리뷰가 포함된 비정형 데이터 분석 및 공간특성(가로특성)을 고려하여 유동인구 지수 개발 후 H/T breaks를 적용하여 가시화 분석

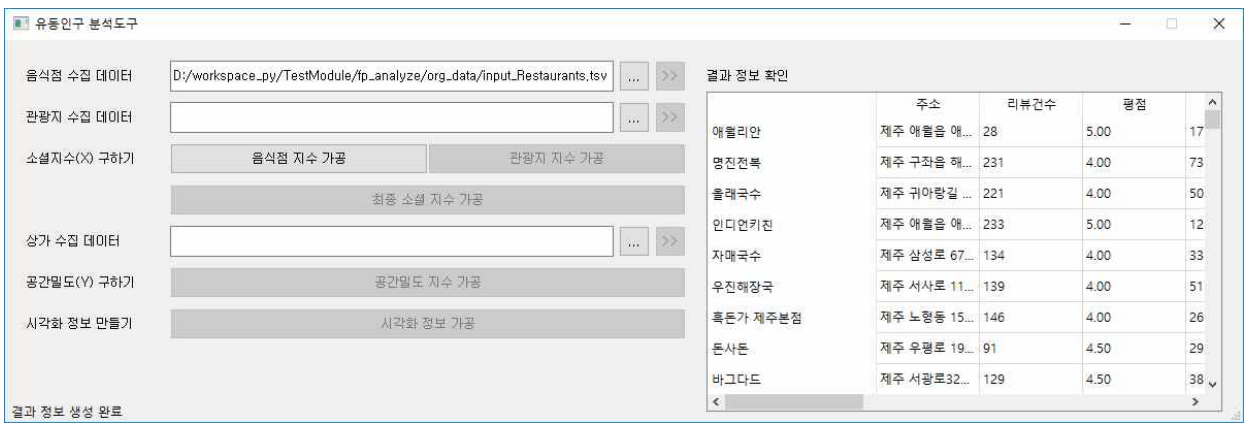
❍ 웹크롤를 통해 수집한 트립어드바이저의 소셜지수(X)를 구하고, 공공데이터포털에서 제공하는 상가업소 자료를 이용하여 공간밀도(Y)를 구해 유동인구 예측지수(Z)를 도출



<그림 2-200> 예측 지수 Z 산정을 위한 연구 흐름도

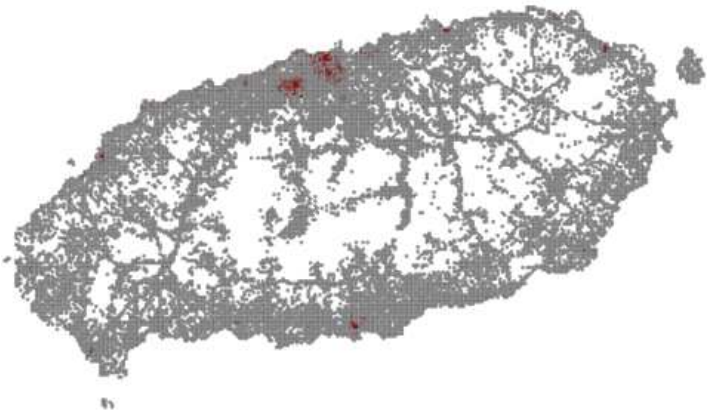
❍ 분석 결과

- 분석용 가공 도구



<그림 2-201> 유동인구 예측용 가공 도구

- 분석 결과 가시화



<그림 2-202> 유동인구 예측 가시화

- OpenAPI 서비스 결과



<그림 2-203> json 형식 결과



<그림 2-204> xml 형식 결과

⑰ [단위성과 E-20] 공간 빅데이터 분석결과 가시화 모듈 개발(동적가시화)(6차연도)

㉮ 소셜 미디어 이벤트 탐지 모듈

- ○ 사용자가 입력한 분석 조회 기간에 해당하는 수집된 비정형 데이터를 읽어온 후, 작성기간(Time or Created_at)을 활용하여 급등한 이상시점을 탐지
- ○ 이상 탐지된 시점의 텍스트를 가져와서 단어 임베딩을 분석하는 텍스트 기반의 이벤트 탐지
- ○ ETRI[6]에서 제공하는 언어 분석 API로 우리나라 대표 언어처리 분야의 인공지능 엑소브레인(Exobrain)의 개체명인식[7] 기술을 활용하여 비정형 데이터에서 장소 정보 추출

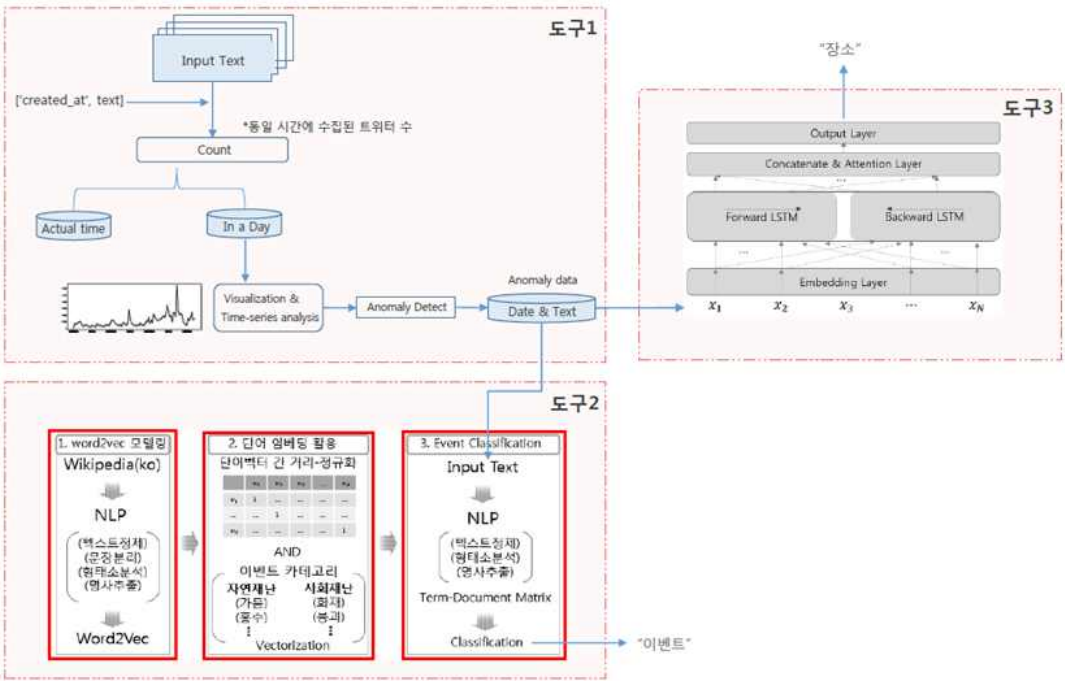
---

6) ETRI(Electronics and Telecommunications Research Institute, 한국전자통신연구원)은 정보, 통신, 전자, 방송 및 성과 관련 융·복합기술 분야의 산업원천 기술 개발 및 성과 확산을 통해 국가경제·사회 발전에 기여한다는 목적으로 설립된 과학기술정보통신부 산하 과학기술분야 정부출연 연구기관

7) 개체명 인식 기술은 인명(PS), 지명(LC), 기관명(OG) 등과 같은 개체명을 인식하는 기술로 특정 개체를 표현하는 단어에 대한 의미 정보를 제공. 개체명 태그셋은 15개 대분류 및 146개 세분류로 구성된 TTA 표준 개체명 태그셋(TTAK.KO-10.0852)을 사용

❍ 추출된 장소 개체명을 장소 대체어(축약어, 오타 등)를 고려하여 정규화된 행정구역 장소정보 추출

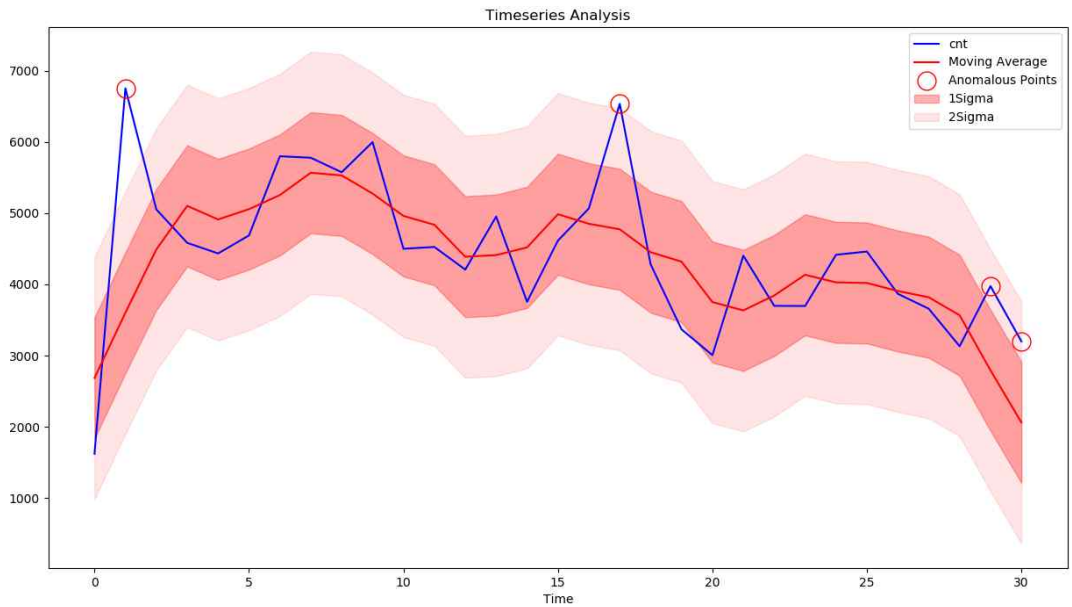
❍ 장소 개체명 중 기관명(OG)이 포함된 텍스트를 장소후보군으로 선정 후 행정구역 장소정보와 융합하여 POI 수준의 경위도 좌표로 된 위치정보 추출



<그림 2-205> 이벤트 탐지를 위한 분석 흐름도

❍ 분석 결과

- 분석 결과 가시화



<그림 2-206> 시계열 분석을 통한 이상치 탐지



<그림 2-207> 단어 임베딩 및 이상 시점의 핫키워드들을 워드클라우드 표출

- OpenAPI 서비스 결과

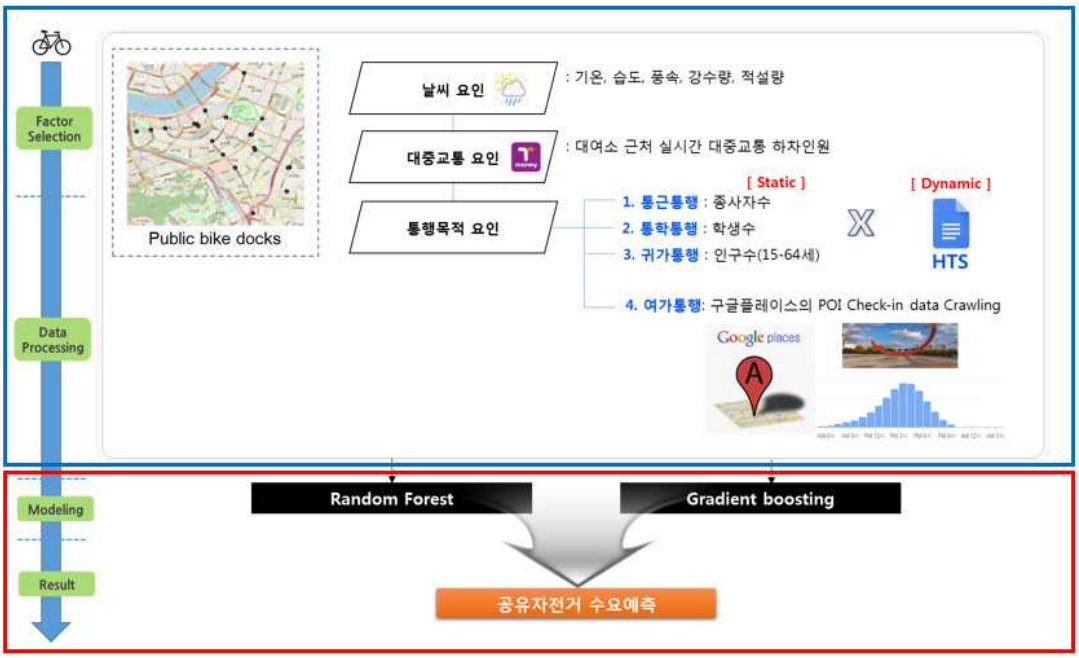


<그림 2-208> json형태의 장소 정보 추출 결과 화면

㉯ 통행 목적을 위한 공공자전거 수요예측 모듈

❍ 사용자가 알고 싶어 하는 공공자전거역에 대한 기본사항 및 훈련 데이터에 대한 시각화를 제공하는 기술

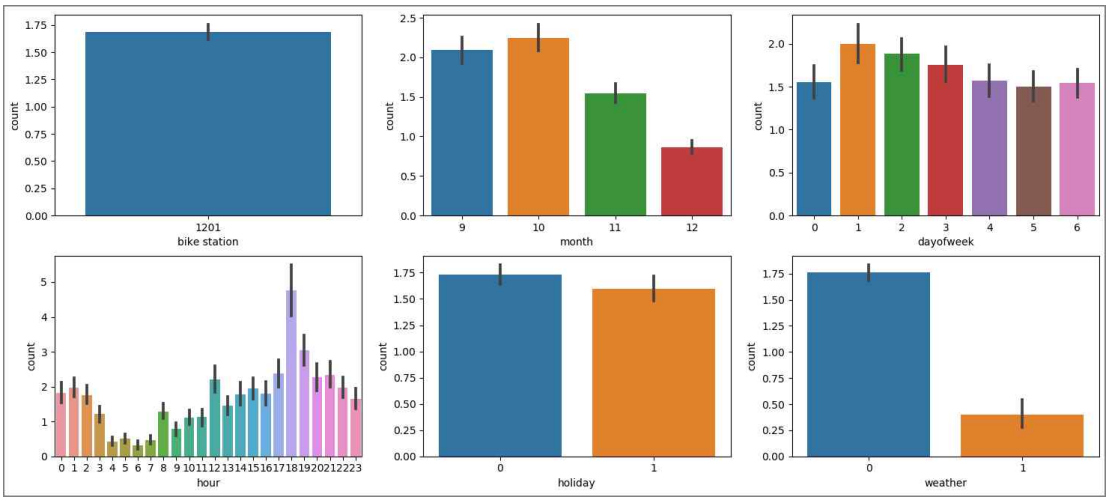
❍ 공공자전거 대여관련 전체 대이터를 훈련 데이터와 테스트 데이터롤 나눈 후 다양한 머신러닝 기법을 적용하여 수요를 예측하는 기술



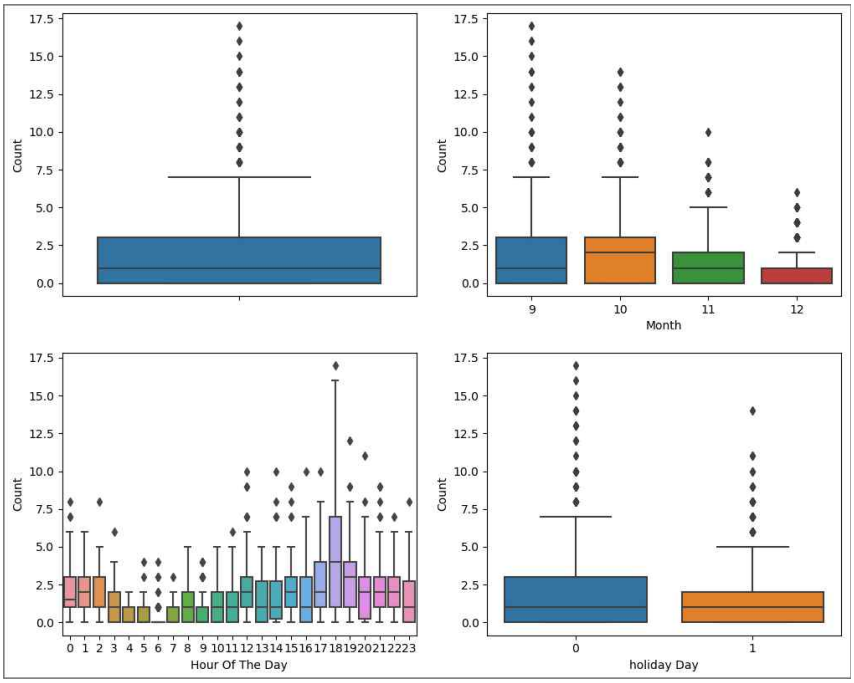
<그림 2-209> 통행 목적을 고려한 공유자전거 수요예측을 위한 분석 흐름도

❍ 분석 결과

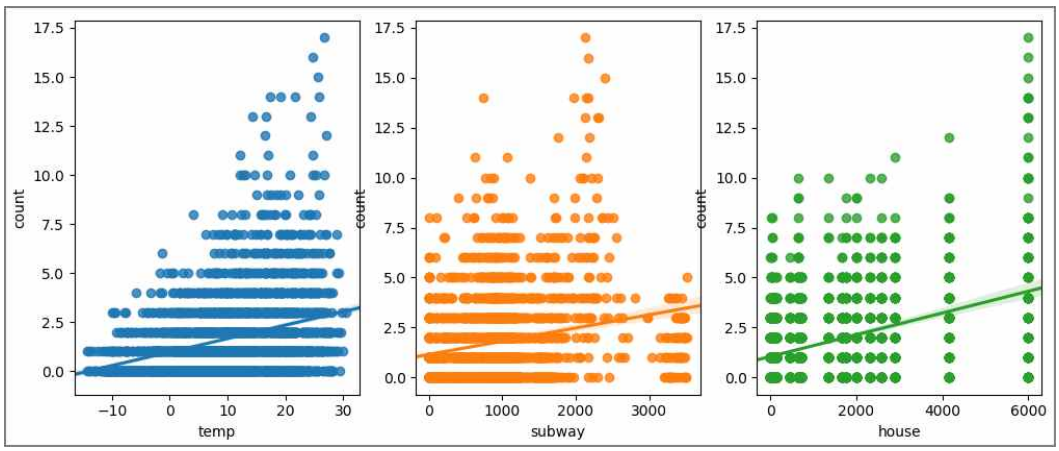
- 분석 결과 가시화



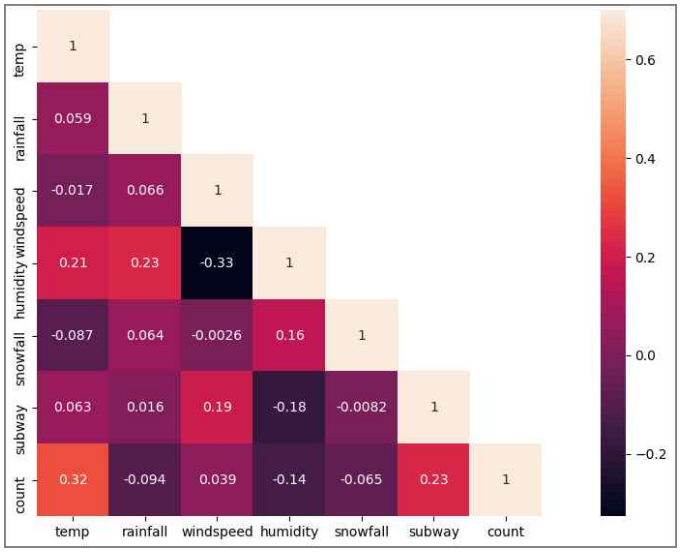
<그림 2-210> 바플롯(barplot) 시계열 분석



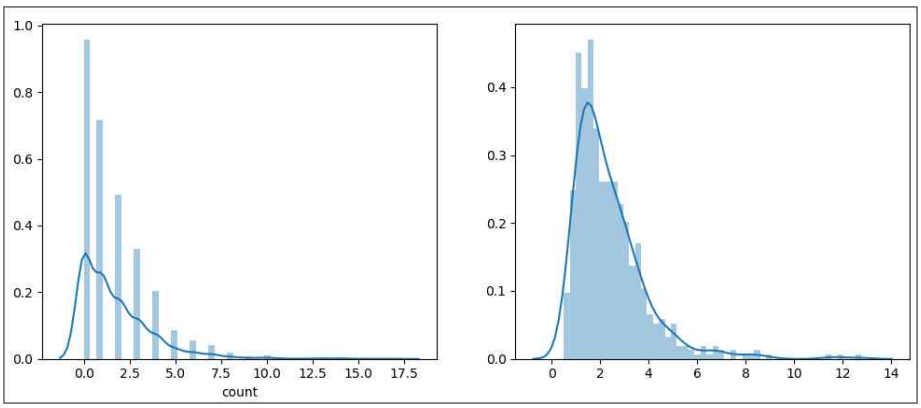
<그림 2-211> 박스플롯(boxplot) 시계열 분석



<그림 2-212> 산점도(regplot) 시계열 분석

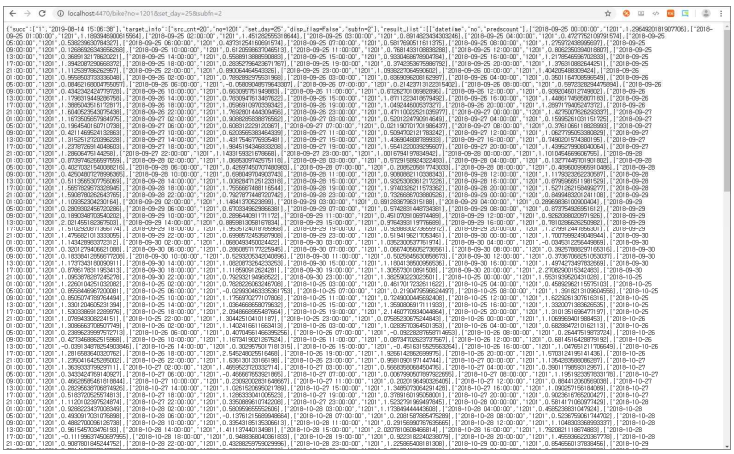


<그림 2-213> 히트맵(heatmap) 시계열 분석



<그림 2-214> 예측 모델 및 수요 예측의 시계열 분석

- OpenAPI 서비스 결과



<그림 2-215> json형태의 공공자전거 수요 예측 결과 화면

㉰ 공간 빅데이터 저장 및 사용을 위한 Hadoop 연계

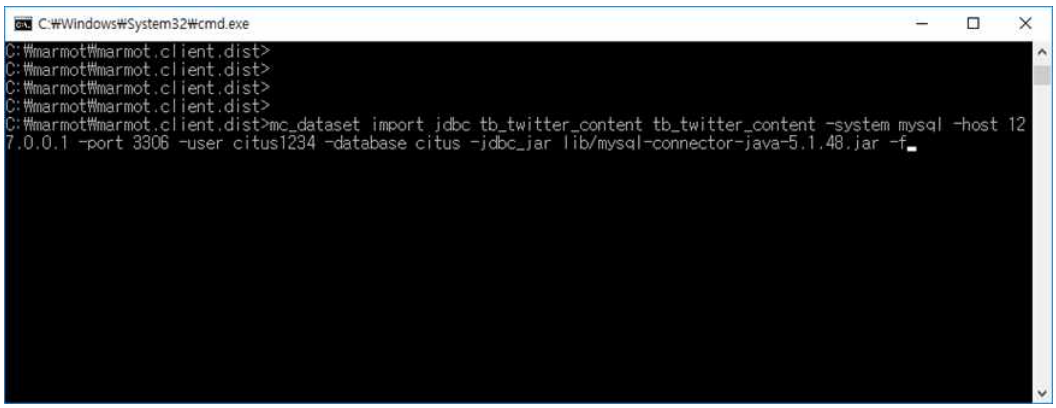
1세부의 공간 빅데이터 저장/접근 기능을 탑재한 marmot 시스템과 연계 기능 개발



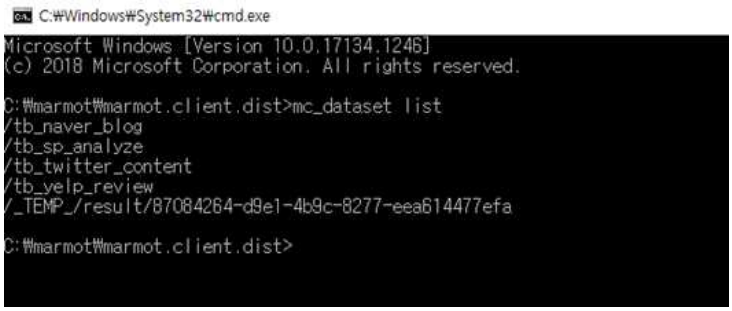
<그림 2-216> 1세부 연계 구조도

❍ 공간 빅데이터 저장을 위한 Marmot 시스템 사용

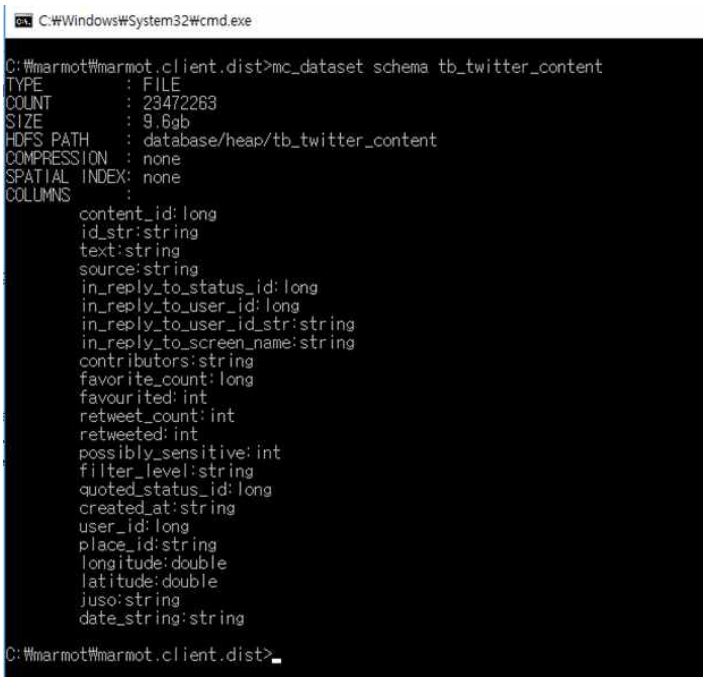
- 1세부 연구성과물과 2세부 연구성과물의 통합 가능성을 검증하기 위해 Marmot Command Scripts를 통한 Hadoop 기반 NoSQL Database 형태로 데이터 저장
  - Marmot과의 연결을 위한 JDBC import 모듈 설치 및 환경설정
  - Marmot Command Scripts를 사용하여 서버 접속 여부 확인 및 데이터 저장



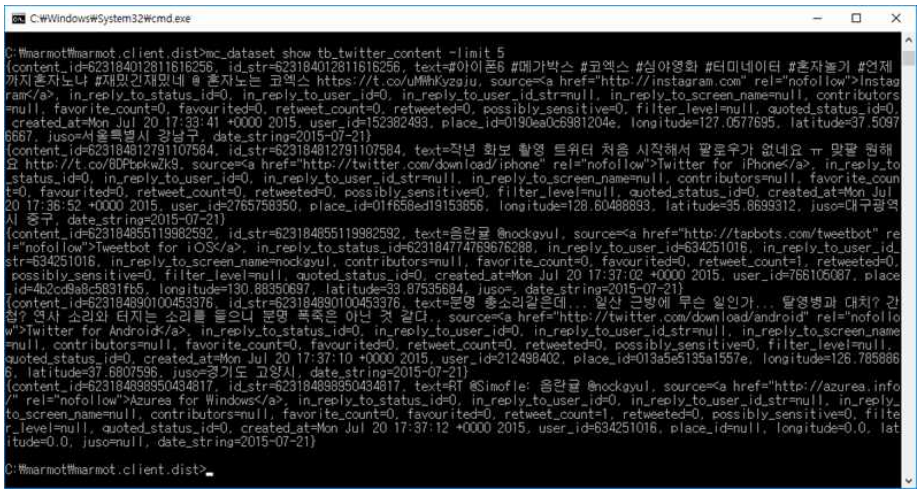
<그림 2-217> 데이터 저장



<그림 2-218> 데이터 목록 확인



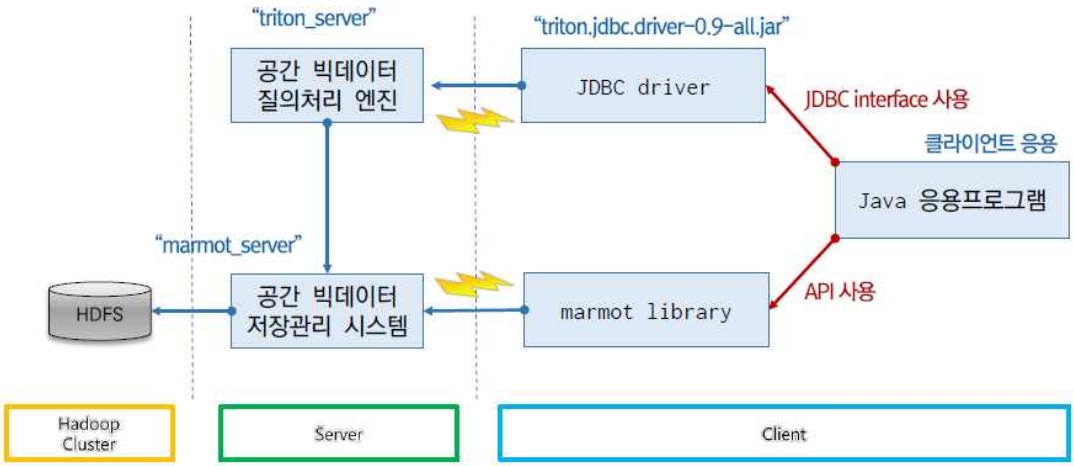
<그림 2-219> 데이터 스키마 확인



<그림 2-220> 데이터 내용 확인

❍ 공간 빅데이터 사용을 위한 JDBC Driver 모듈 사용

- 1세부 연구성과물인 Marmot에 내장된 데이터 질의처리 엔진과 JDBC Driver 모듈을 이용하여 저장된 데이터 조회
  - 공간 빅데이터 질의처리 엔진(triton)은 1세부가 개발한 공간 빅데이터 저장관리 엔진(marmot) 위에서 동작하는 서버
  - 질의처리 서버에 접속하고 데이터를 교환하는 것은 클라이언트 사이드의 JDBC Driver를 통해 가능
  - 분석모듈에서 JDBC Driver(jar 파일)의 라이브러리를 참조하여 JDBC 인터페이스를 통해 서버에 질의를 요청함으로써 결과 데이터를 수신



<그림 2-221> 데이터 조회를 위한 시스템 구조

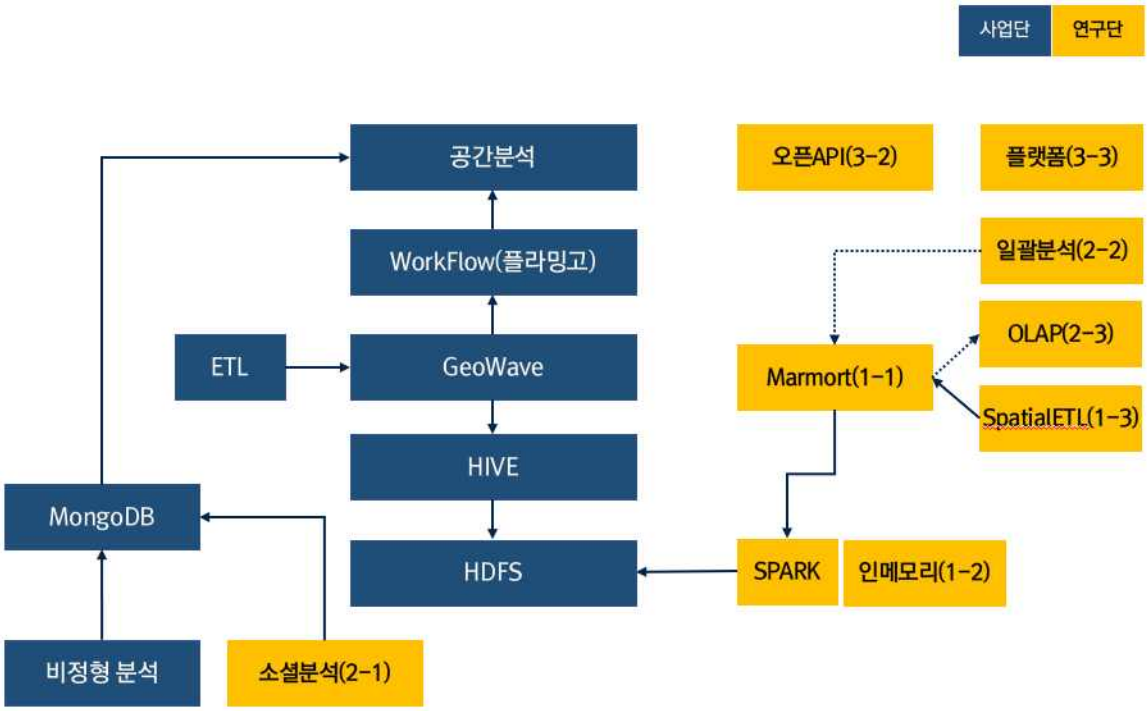
⑱ [단위성과 E-21] 공간 빅데이터 분석 모듈 사업단 적용 및 실증(6차연도)

㉮ 공간 빅데이터 R&D 기술의 적용성 검토 의견

- ❍ 윈도우 환경의 분석모듈을 리눅스 환경으로 변경 작업
- ❍ 사업단 수집 데이터의 MongoDB 연계 기능 구현
- ❍ 데이터 조회관련 분석 모듈 소스 변경
- ❍ 개발 서버에 단계별 분석모듈 적용 및 일정 기간 데이터로 안정화 유지

<표 2-75> R&D기술 적용 핵심 현안

현안	해결방안	사업단 검토
- 분석모듈 검증용으로 비정형 데이터 자체 수집 - 수집 및 가공 데이터 활용을 위한 별도 RDBMS 관리 (MySQL)	- 사업단 비정형 데이터 활용 - 수집 및 가공 데이터 활용을 위한 RDBMS 정의 및 변경 - 호환 연계 어댑터 개발	- 비정형 데이터는 MongoDB에 저장 중



<그림 2-222> 사업단 및 연구단 구성 개념도

❍ 연구단 검토 의견

- 이슈 및 제약사항

  • 윈도우 개발환경의 분석모듈을 리눅스 개발환경으로 변경 필요
  • 사업단 수집 데이터에 맞게 분석 모듈 변경 필요
  • 사업단 수집 서버의 MongoDB 연계 기능 필요
  • 분석모듈용 개발언어 설치 및 라이브러리 사용에 따른 버전 확인

- 해결 방안

  • 리눅스 개발환경의 적용을 위한 개발 기간 산정 및 문제점 검토
  • 사업단 수집 데이터에 맞게 분석 모듈 소스 변경
  • 사업단 수집 서버의 MongoDB 연계 기능 구현
  • 버전에 맞는 분석모듈용 개발언어 설치 및 라이브러리 설치
  • 개발 서버에 단계별 적용 및 일정 기간 데스트로 안정화 유지

❍ 사업단 검토 의견

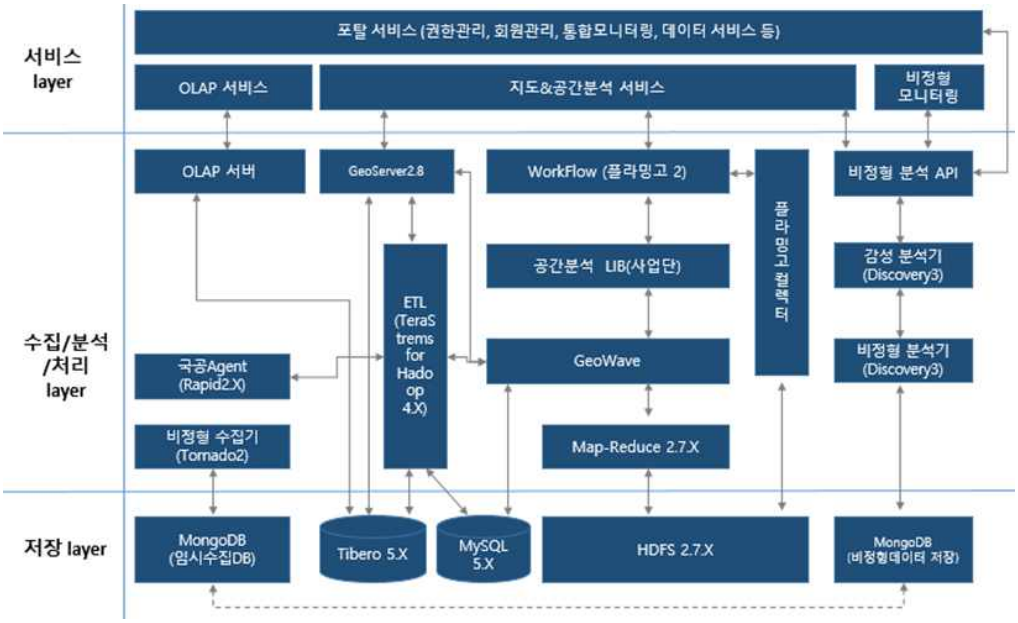
- 이슈 및 제약사항

  • 데이터 분석을 위한 비정형 데이터 포맷의 상이함
  • R&D 성과물 활용을 위한 서비스 부재

- 해결 방안

  • 비정형 데이터의 공동활용을 위한 데이터 변환 기능 개발
  • 연구단 성과물에서 제공하는 API를 활용하여 서비스할 수 있는 기능(화면) 개발

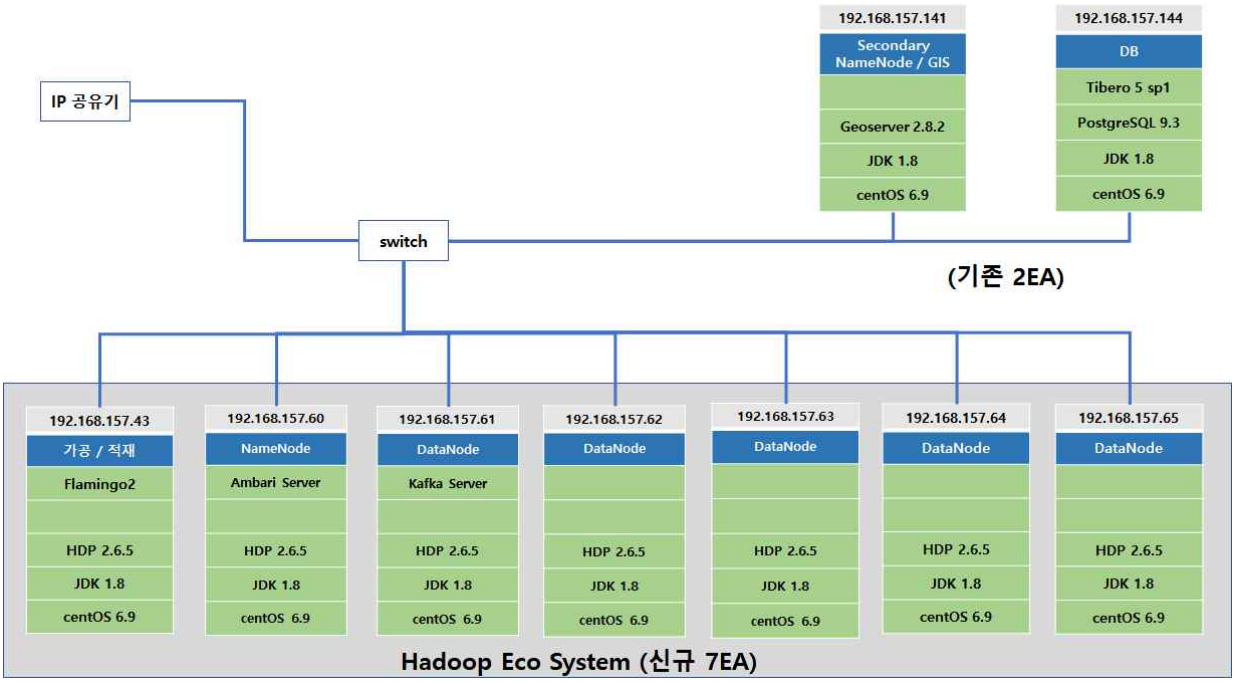
㉯ 공간 빅데이터 분석을 위한 개발환경 구축



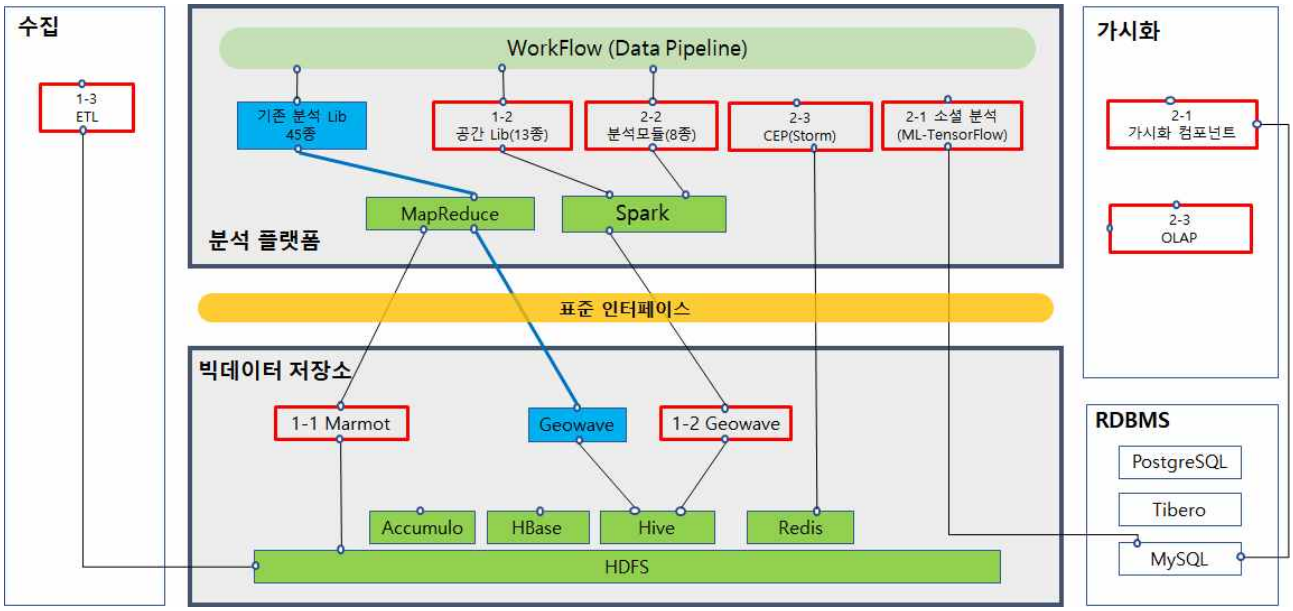
<그림 2-223> R&D 성과물의 사업단 적용 후 구성 개념도

○ 시스템 구조

- 사업단은 R&D 연구단 성과를 실증하기 위한 개발 환경 제공



<그림 2-224> R&D 성과물의 사업단 실증을 위한 H/W 구성



<그림 2-225> R&D 성과물의 사업단 실증을 위한 S/W 구성

○ 개발 도구 설치

- 리눅스 환경의 OS 및 기본 프로그램만 제공

<표 2-76> 프로그램 설치 목록 및 버전

	항목	버전
1	CentOS	v6.9
2	JDK	v1.8
3	HDP - Hadoop - Zookeeper - Spark - Storm - Kafka - Flume - Zeppelin - Hive	v2.6.5 v2.7.3 v3.4.6 v2.3.0 v1.1.0 v1.0.0 v1.5.2 v0.7.3 v1.2.1

㉰ 분석 모듈 적용

○ 분석모듈 개요

- 머신러닝 기반 문서분류 모듈 (층간 소음 분석)

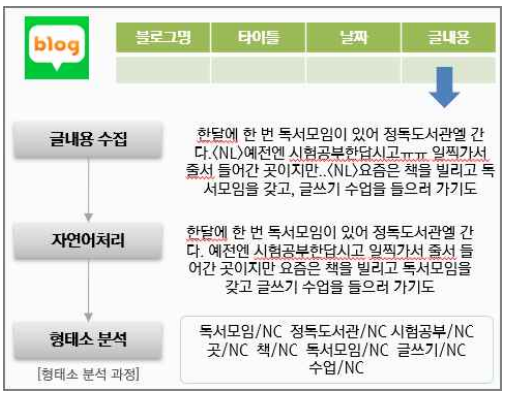
• 네이버 카페의 층간소음 관련 키워드 검색으로 웹크롤을 이용한 데이터 수집 및 분석



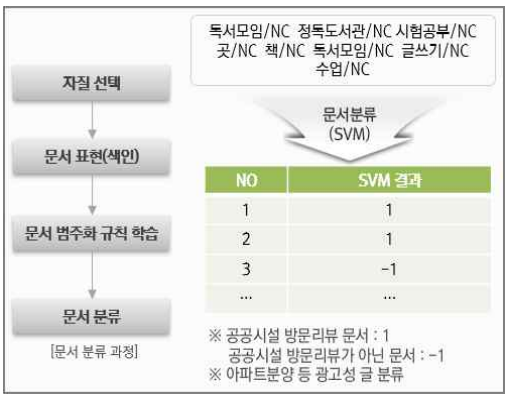
<그림 2-226> 머신러닝 기반 문서분류 모듈 처리 흐름도

- 공공시설 매력도 측정 및 접근성 분석 모듈

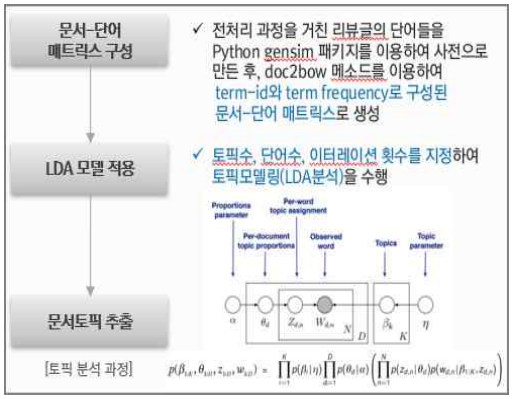
• 공공시설 방문리뷰 데이터를 수집하여 형태소 분석을 수행한 후 공공시설 관련 문서를 분류



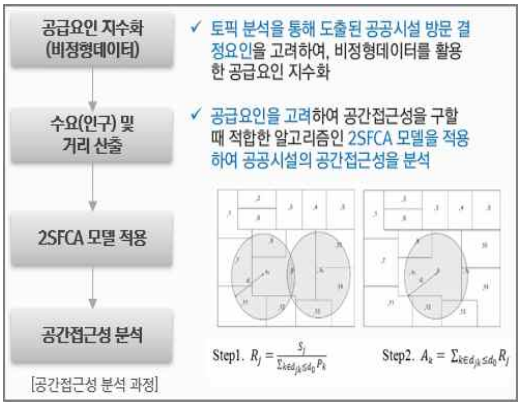
<그림 2-227> 형태소 분석 (도구1)



<그림 2-228> 공공시설 관련 문서분류 (도구2)



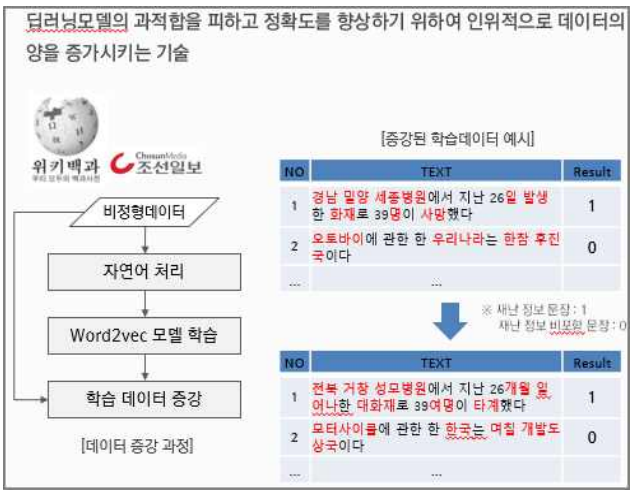
<그림 2-229> 토픽분석 (도구3)



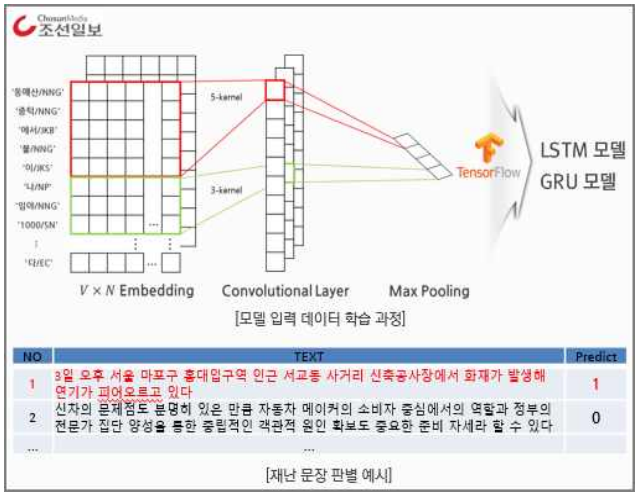
<그림 2-230> 공간접근성 분석 (도구4)

- 머신러닝 기반 특정주제 문장 분류 모듈

  - 딥러닝이나 기계학습 분야에서 인위적으로 데이터 양을 증가시키는 데이터 증강 기법 적용
  - 순환 신경망 기술의 RNN(Recurrent Neural Network) 중 LSTM 모델과 GRU 모델로 학습
  - 학습된 모델을 적용하여 특정주제가 포함된 문장을 추출 (트위터에 재난 정보가 포함된 트윗 추출)



<그림 2-231> 학습 데이터 증강 도구 (도구1)



<그림 2-232> 문장 추출 모델 학습 도구 (도구2)



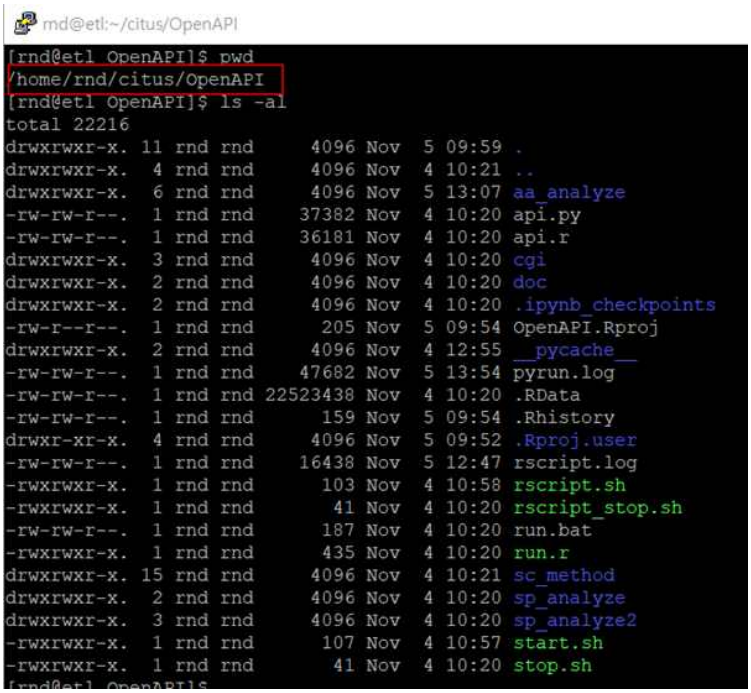
<그림 2-233> 특정 주제 포함 문장 추출 도구 (도구3)

❍ 분석모듈 설치

- 프로그램 설치 목록 및 버전

  • Python v3.5.6, R v3.6.2, rstudio-server v1.2.5019

- 분석모듈 설치위치



<그림 2-234> 분석모듈 설치위치

❍ 분석모듈 실행결과

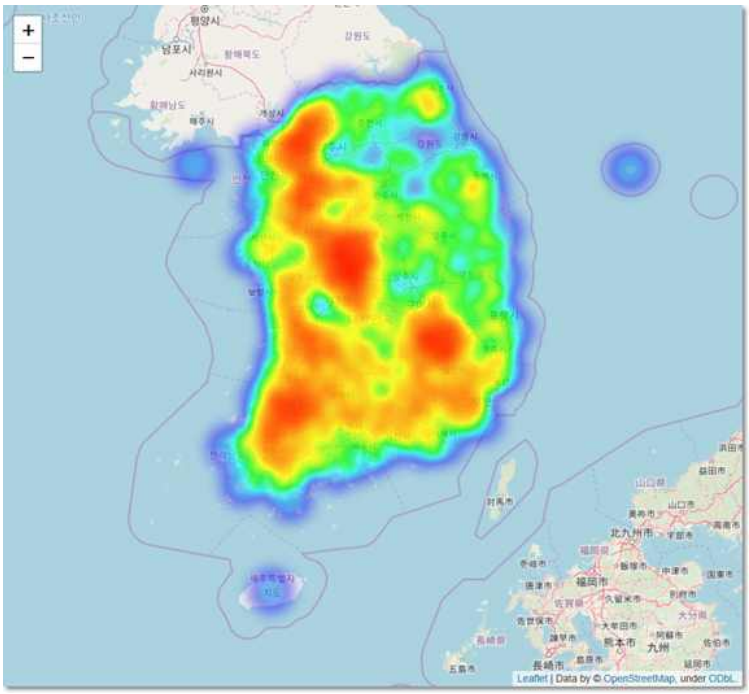
- 머신러닝 기반 문서분류 모듈 (층간 소음 분석)



<그림 2-235> 머신러닝 기반 문서분류 모듈의 크롬브라우저 사용



<그림 2-236> 머신러닝 기반 문서분류 모듈의 가시화 도구



<그림 2-237> 가시화 결과

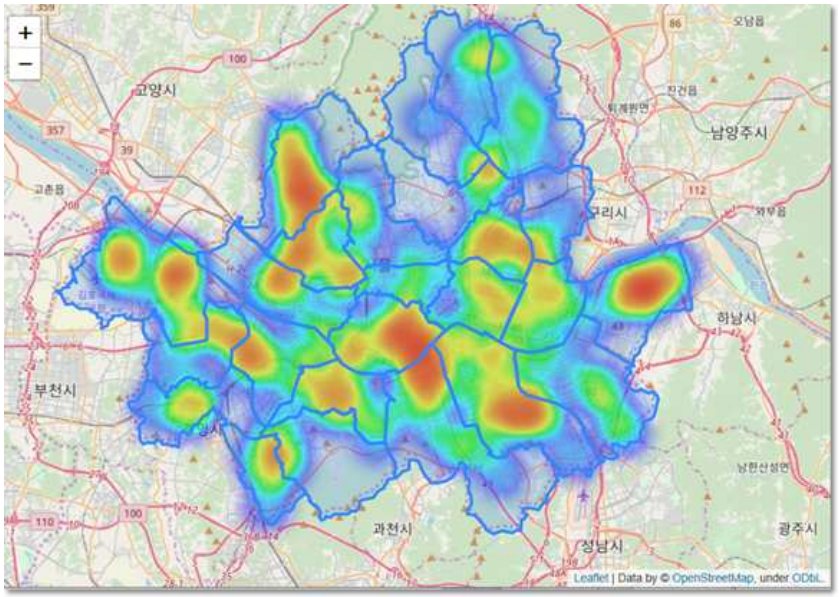
- 공공시설 매력도 측정 및 접근성 분석 모듈



<그림 2-238> 공공시설 매력도 측정 및 접근성 분석의 크롬브라우저 사용



<그림 2-239> 공공시설 매력도 측정 및 접근성 분석의 가시화 도구



<그림 2-240> 가시화 결과

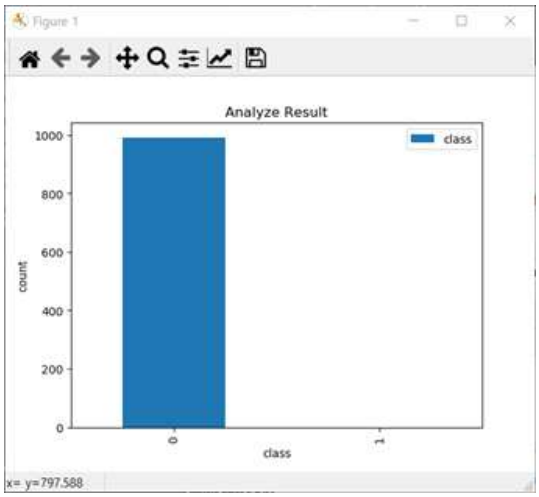
- 머신러닝 기반 특정주제 문장 분류 모듈



<그림 2-241> 머신러닝 기반 특정주제 문장분류 모듈의 크롬브라우저 사용

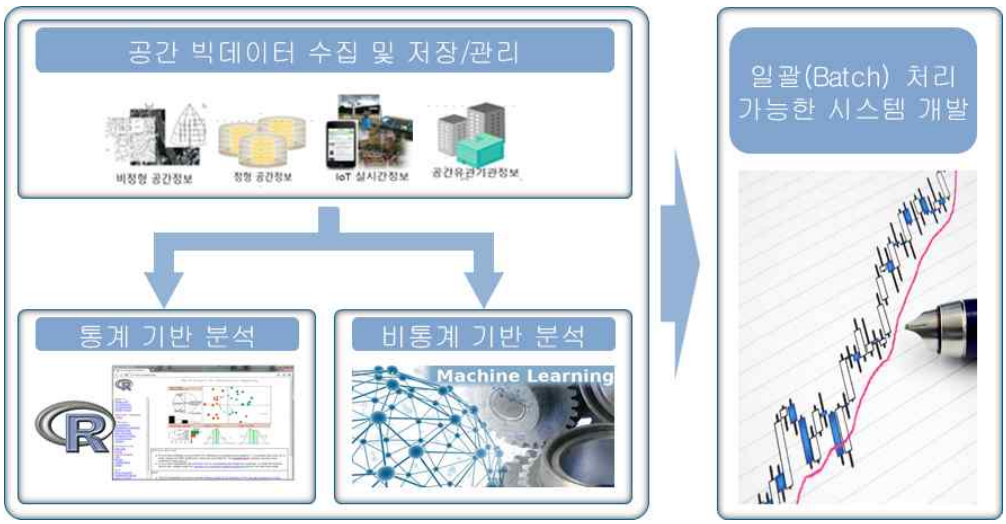


<그림 2-242> 머신러닝 기반 특정주제 문장분류 모듈의 가시화 도구



<그림 2-243> 가시화 결과

(카) 공간 빅데이터 Batch 분석 컴포넌트



<그림 2-244> 주요연구내용

❍ 공간 빅데이터 Batch 분석 컴포넌트는 1세부과제에서 구축된 빅데이터 저장 관리 플랫폼을 통해 전달된 공간 빅데이터에 통계기반·비통계기반의 분석기

법 도구들을 일괄(Batch)처리로 적용 가능한 시스템을 의미함

❍ 통계적·비통계적 분석기법 도구를 사용해 Predictive 또는 Descriptive 분석 기능들의 일련 과정을 일괄처리 함으로서 수집된 공간 빅데이터에서 정해진 Pattern을 지속적으로 분석해 내는 시스템임

❍ 분석 컴포넌트들은 분산처리 시스템인 Spark 시스템을 기반으로 구성되어, Machine Learning 등을 이용한 통계·비통계 기반 공간 빅데이터 분석 플랫폼을 구축하는 기반이 됨

❍ 개발된 분석 컴포넌트들은 Web-platform 기반의 Batch 분석 솔루션으로 구현되었음

❍ **연세대학교 산학협력단**

- 분석 기능 선정을 위한 통계기반 분석 요구조건 및 비통계기반 공간 빅데이터 분석 요구조건을 검토하였음
- 검토한 내용을 기반으로 개발 기능을 결정하고, 통계기반 및 비통계기반 공간 빅데이터 분석 컴포넌트를 설계하였음
- 이후, 새로운 공간 빅데이터 분석 기능 추가 및 기본 분석 기능의 성능 향상, 기 개발 기능의 고도화 / 최적화를 달성하였음
- 최종적으로, 개발된 공간 빅데이터 분석 컴포넌트들을 기반으로 공공 / 민간 수요기반 공간 빅데이터 Batch 분석 시나리오 및 솔루션을 설계하였음

❍ **가이온**

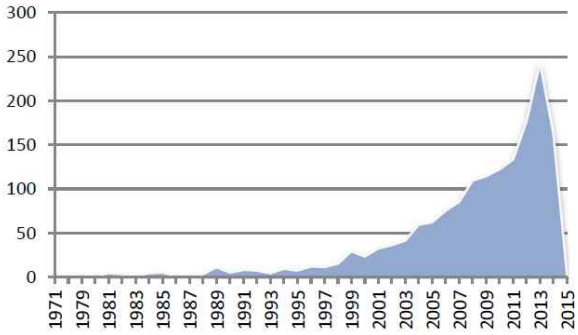
- 공간 빅데이터 분석 컴포넌트의 사용자 인터페이스 및 아키텍처를 설계하였고, 이를 기반으로 분석 컴포넌트 프로토타입을 구현하였음
- 공간 빅데이터 분석 컴포넌트를 추가하고 연산자 동작 정확도 및 처리 속도를 향상시켰음
- 공간 빅데이터 분석 컴포넌트의 기능을 고도화하였으며, 분산처리 속도 향상을 위하여 Spark를 통해 분석 기능을 구현하였음
- 최종적으로, 공간 빅데이터 Batch 분석 솔루션을 구현하였으며 공인인증 및 소프트웨어 등록을 통해 구현된 솔루션의 성능을 확인하였음

① [단위성과 F-1] 공간 빅데이터 통계기반 요구조건 분석 - R의 공간분석기능 검토

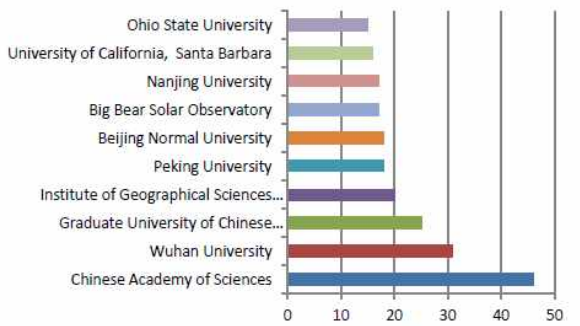
❍ SCOPUS DB에서 “Spatial Big Data” 라는 키워드로 검색된 전체 문헌의 초록을 텍스트마이닝과 LDA를 사용하여 분석하여 공간 빅데이터의 글로벌 연

구동향을 파악함

❍ 그 결과, 공간 빅데이터 연구가 급증하고 있으며, 대학 및 연구기관에서 연구가 주로 이뤄지고 있는 것으로 나타남

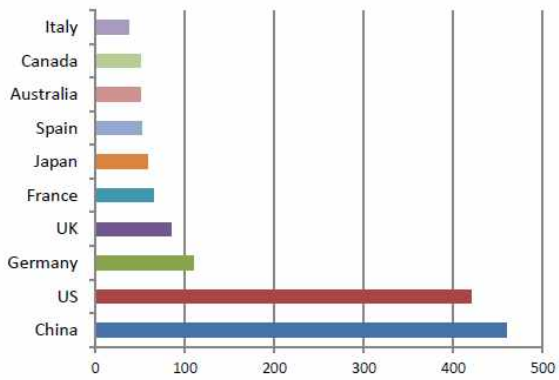


<그림 2-245> 연간 공간 빅데이터 관련 문헌 빈도

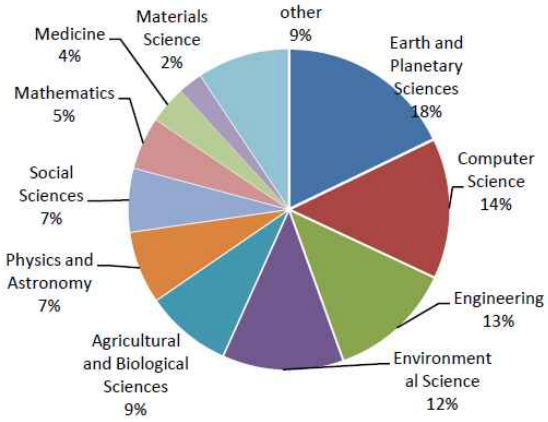


<그림 2-246> 공간 빅데이터 관련 문헌의 저자가 소속된 상위 10개 기관

❍ 중국, 미국, 독일, 영국 등이 공간 빅데이터 연구를 주도하고 있으며, Optics, Astronomy, Computer Science 분야 등 다양한 분야에서 공간 빅데이터에 대한 접근이 이뤄지는 것을 확인함



<그림 2-247> 국가별 공간 빅데이터 문헌 빈도



<그림 2-248> 공간 빅데이터 연구 분야

❍ 공간 빅데이터 관련 문헌을 대상으로 텍스트마이닝을 통해 도출된 문서-용어 행렬에 토픽모델로써 LDA(Latent Dirichlet Allocation)을 적용하여 40 여개의 주제를 발견하고, 각 주제간의 관계를 파악하기 위해 주제 네크워크를 분석하여 다음과 같은 그룹을 확인하였음

- 그룹 1 : 자연환경을 바탕으로 한 산업/경제
- 그룹 1-1 : 자연환경과 대기, 기상 관측, 바이오매스
- 그룹 1-2 : 도시에 대한 지리 연구, 공해와 경제, 그리고 농작물의 가격과 산업/경제
- 그룹 2 : 공간 빅데이터 분석을 위한 방법론과 이미지 처리기술 및 알고리즘
- 그룹 3 : 도시에서의 공간데이터 처리를 위한 센서 기술과 알고리즘, 사회적 커뮤니티와 주거, 그리고 GPS와 IT 인프라를 바탕으로 한 도시에서의 교통과 오존 등에 대한 주제

❍ 또한, 국토공간계획지원체계(KOPSS) 에서 제공하는 분석모형 및 세부기능 현황을 조사하여, 세부기능에 대응되는 공간 빅데이터 연구사례를 조사하고, 이로부터 공공 및 민간 분야에서 공간 빅데이터의 통계기반 요구조건을 도출함

- 지역이 : 지역산업분석 -> 산업특화 분석 (특화지수, LQ)
- 국토리 : 개발잠재력 분석 -> 가중치 분석
- 재생이 : 정비대상후보지 분석 -> 정비계획 수립대상지 분석 -> 공간 군집화
- 재생이 : 재정비촉진지구 분석 -> 광역 노후 불량률 분석
- 시설이 : 공급적정성 평가 -> 공급적정성 평가지수 산정
- 시설이 : 공급적정성 평가 -> 수요적정성 평가

❍ 공간 빅데이터 통계기반 요구조건을 분석하기 위하여 오픈 소스 통계 패키지인 R의 공간 분석 기능을 검토함

<표 2-77> R 기반 공간빅데이터 분석기능 검토

	분석방법	R 패키지
공간 빅데이터	Geostatistics	gstat
		geoR
		geoRglm
		vardiag
		spBayes
	Spatial Regression	nlme
		sphet
		McSpatial
		spatialprobit
	Ecological analysis	ade4
		adehabitat
		pastecs
		gstudio

② [단위성과 F-2] 공간 빅데이터 비통계기반 요구조건 분석 - Machine Learning 검토

❍ 8편의 국내외 공간정보 분석 관련 논문을 비교 분석하여 연구 동향을 파악하고 향후 연구 방향을 계획함

- Clustering 기반 분석은 다양한 공간 데이터의 위치 정보를 고려하여 지리적, 위치적 특성에 따라 관련성이 높은 공간 데이터들을 같은 그룹으로 분류하고, 특정 속성에 가중치를 부여하는 방식으로 클러스터링을 수행함
- Association Rule 기반 분석은 공간 데이터와 공간 데이터, 공간 데이터와 비공간 데이터, 비공간 데이터와 비공간 데이터 사이에 하나 혹은 그 이상의 공간 객체들이 서로 연관되어있는 규칙을 분석함
- Prediction and Evaluation은 Clustering 혹은 Association rule 분석에 속하지 않는 분석들로, 방대한 양의 공간 빅데이터 모니터링을 기반으로 traffic velocity 혹은 vacant taxi를 예측하는 연구를 수행한 사례 등이 조사됨

<표 2-78> 공간 빅데이터 비통계 분석 기능 검토

분석 기법	연구제목	데이터	연구목적	알고리즘	특징
Clustering	• iTaxi: Context-Aware Taxi Demand Hotspots Prediction • UsingOntologyandDataMiningApproaches	• GPS data	• predicts taxi demand distribution based on • time,weathercondition,andlocation	• Context-based filtering • Clustering(K-means,Hierarchical,DBSCAN)	• 4-step approach is proposed to solve the taxi demand analysis problem different clustering methods applied
	• 교통카드자료를 이용한 환승정류장의 유형별 입지특성에 관한 연구	• 교통카드 데이터	• 환승통행 특성 분석	• k-means clustering	• 차량별 버스정류장 도착시간을 수집하여 교통카드자료와의 matching 시도

분석 기법	연구제목	데이터	연구목적	알고리즘	특징
	-대구시를 중심으로-				
	• 교통카드자료를 이용한 통행패턴분석과 정책활용방안 연구 -경기도를 중심으로-	• 교통카드데이터	• 통행패턴분석/환승정류소 유형구분 및 활용방안 도출	• k-means clustering/CHAID 분석	• 3가지의 시나리오를 설정하고, 그에 따른 환승거점 정류장 선정
	• 아침 첨두시간대 지하철 이용수요의 결정요인에 관한 연구	• 지하철 승하차 데이터/토지관리 정보	• 통행량이 집중되는 첨두시간대의 지하철 통행량과 토지이용간의 관계 분석	• Linear/Semi-log/Log Model	• 역세권 개발이나 대중교통지향형 개발 등에 도움이 될 것으로 기대
	• 특허 데이터에 기초한 지식창출활동의 공간분석	• 한국공개용 특허공보 데이터	• 지식창출활동의 지역별 성장과 공간분포 및 공간적 불평등과 공간적 연관성 분석	• Moran's I/Geary C/LISA(Local Indicator of Spatial Association)	• 특허 원시자료를 DB로 구축하여 공간통계와 지리정보체계를 활용하여 지식창출활동 패턴이 공간적 연관성 분석
Association Rule	• 대용량 공간 데이터로 부터 빈발 패턴 마이닝	• FIMI Repository T10I4D100K or Chess	• 압축된 패트리샤 빈발 패턴 트리(이하 CPFP-Tree라 칭함)라는 새로운 자료구조와 이를 사용한 빈발 패턴 마이닝 알고리즘 제안	• Apriori, FP-Growth 알고리즘	• CPFP-Tree를 이용한 빈발 패턴 마이닝 알고리즘이 기존의 알고리즘들 보다 더 나은 성능을 보인다는 것을 실험 평가를 통하여 확인함.
Prediction and Evaluation	• Taxi-Aware Map: Identifying and Predicting Vacant Taxis in the City	• GPS data	• indentify and predict vacant taxis in the city	• Error-based learning algorithm and method for detecting adequacy of historical data using mutual information	• present a predictive model for the number of vacant taxis in a given area • likelihoodofeachpossiblenumberofvacanttaxisgiventimeofthedy,dayoftheweek,and weathercondition계산
	• Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment	• GPS data	• prove the concept for a traffic monitoring system based on GPS-enabled mobile phones	• 특정 알고리즘을 사용하지 않고 실험을 설계	• Results suggest that a 2-3% penetration of cell phones in the driver population is enough to provide accurate measurements of the velocity of the traffic flow

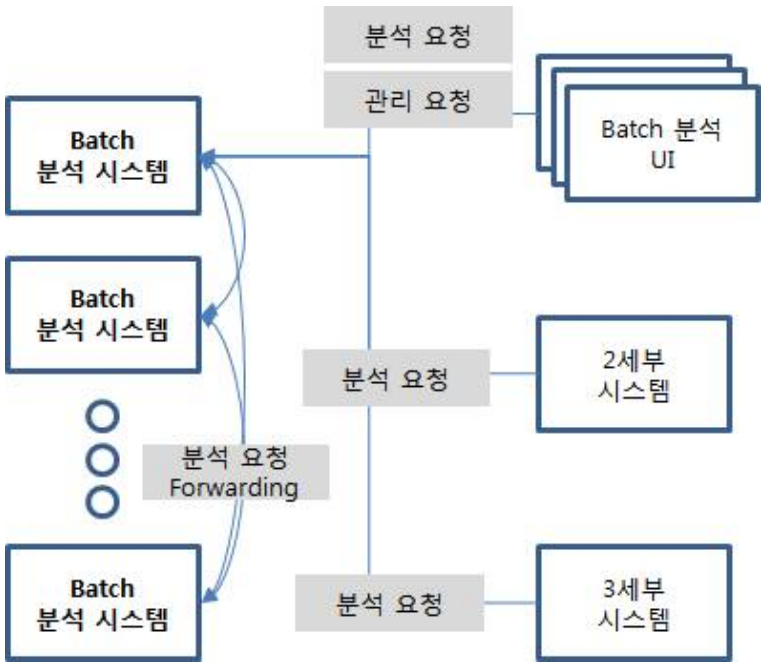
③ [단위성과 F-3] 공간 빅데이터 분석 컴포넌트 인터페이스 설계

❍ 공간 빅데이터 분석 컴포넌트 인터페이스 설계를 위한 구현 환경은 다음과 같음

<표 2-79> 구현 환경

구현 환경	Hadoop 1.2.1	MongoDB 2.6.5
	Hbase 0.98.3	Hive 0.13.1
	Flume 1.5.0.1	R studio Server 0.98.1074
	JAVA 1.6.0_45	

❍ 공간 빅데이터 분석 컴포넌트 아키텍처에서 각 컴포넌트가 유기적으로 동작하기 위해 세부 모듈 컴포넌트 구성 인터페이스 설계와 세부 모듈 컴포넌트 간 인터페이스 설계 작업이 요구됨

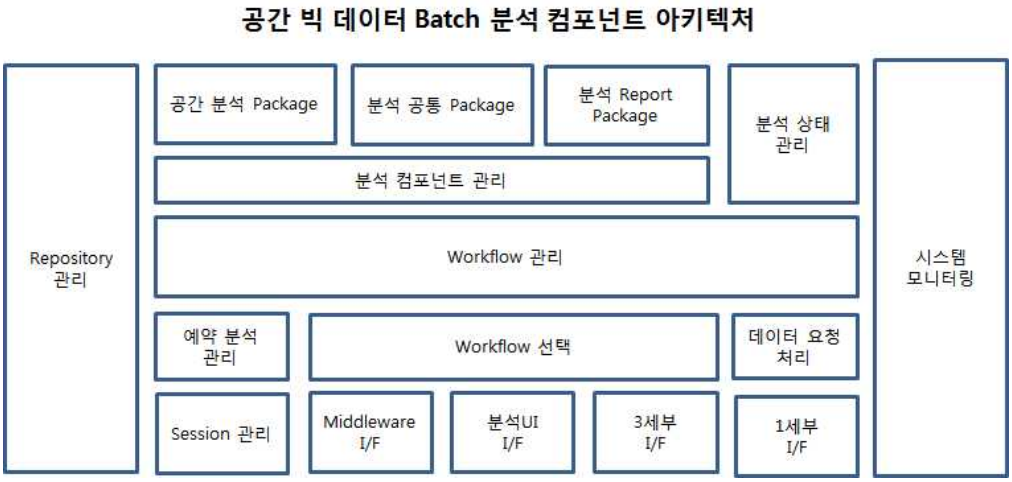


<그림 2-249> 미들웨어 간 인터페이스 관계도

○ 각 모듈의 인터페이스 설계를 통하여, 로그인, 사용자 관리, 세션 관리, I/f 관리, 분석 UI 관리 및 Work Flow 관리, 분석 컴포넌트 관리 등의 컴포넌트가 정의됨

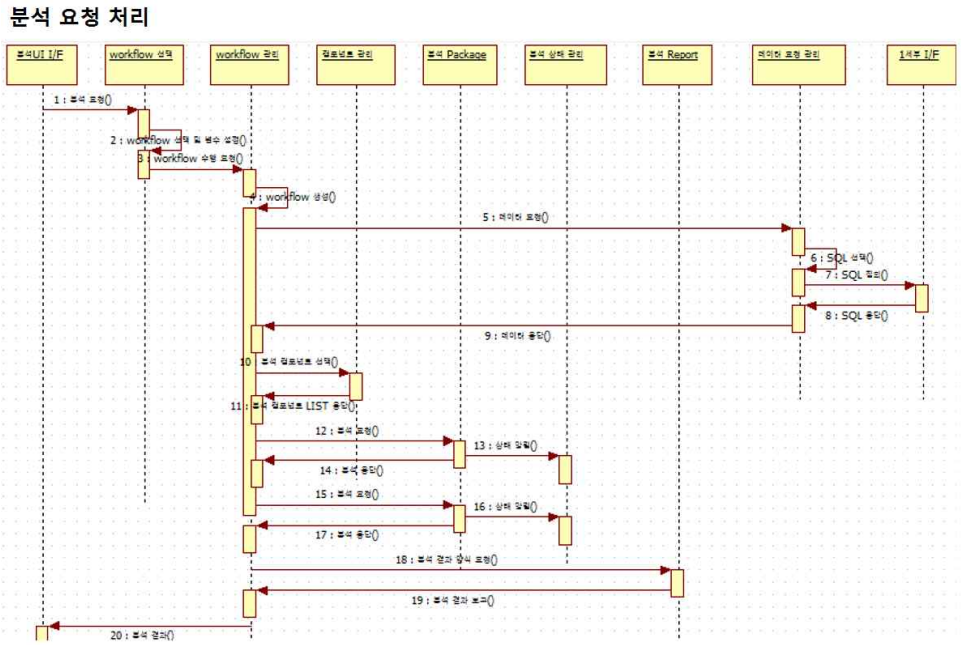
④ [단위성과 F-4] 공간 빅데이터 분석 컴포넌트 아키텍처 설계

○ 공간 빅데이터 분석 컴포넌트 아키텍처 설계를 위하여 필요한 기능을 List-up 하고, User case 다이어그램을 통해 분석 컴포넌트의 workflow를 정의함



<그림 2-250> 공간 빅데이터 Batch 분석 컴포넌트 아키텍쳐

❍ 또한 서비스 시나리오 분석을 통해, 예상되는 분석에 요구되는 컴포넌트들이 어떻게 동작하는지에 대한 시퀀스 다이어그램을 도출하였음



<그림 2-251> 분석 요청 처리에 대한 시퀀스 다이어그램

⑤ [단위성과 F-5] 공간 빅데이터 기본분석 기능 선정 및 개발 기능 결정

❍ 국토교통부에서 서비스 중인 국토공간계획지원체계(KOPSS, Korea Planning Support Systems)의 공간정보 분석기능을 분석하여 공간 빅데이터의 활용이 가능한 세부기능들을 확인함

<표 2-80> KOPSS 공간정보 분석 기능

분야	Level-1	Level-2		
지역이	지역지표분석	지역지표	교차분석	기초통계분석
		공간자기상관분석	공간군집도분석	공간분화도분석
	지역구조분석	권역별 현황보기	인접시군 현황보기	
	지역산업분석	산업분포현황보기	산업특화분석	산업잠재력 분석
	지역개발사업분석	현황검색		
국토리	토지수요산정	인구변화추정	용지규모산정	
	입지여건분석	토지이용규제	인문환경분석	자연환경분석
		토지이용 현황분석		
	개발잠재력분석	입지요인평가	가중치 분석	종합잠재력
	대안비교분석	수용토지현황	토지보상비	토공량/형상분석
	용도별잠재력분석	주변여건분석	입지별 잠재력 분석	
	토지용도입지배분	용도별 입지배분	토지이용 지표분석	토지이용 비교분석
	재정비	광역노후불량률	광역도로점유율분석	개략토지수용비분석

분야	Level-1	Level-2		
재생이	촉진지구분석	분석		
		용도지역별 부지특성		
	개발잠재력분석	인구현황	지가분석	도로접근성분석
		지하철접근성분석		
	지역별 노후불량 현황	건물노후도분석	과소필지비율분석	
	정비계획 수립대상지 분석	노후불량건축물분석	필지별호수 밀도분석	과소필지분석
		도로점유율분석	공간군집화	후보지 추출
	관심지역 정비기준 분석	정비기준 현황분석	토지 및 건축물조사	노후불량률 시뮬레이션
시설이	기반시설공급적 정성 평가	분석시설 선택	분석지역 선택	시설수요분석
		공급적정성평가		
	기반시설 입지분석	입지분석	후보지 분석	접근석분석

❍ 기존 KOPSS에서 도로접근성과 지하철 접근성 기능을 제공하나, 버스접근성에 대한 기능은 제공하지 않고 있기에, E2SFCA 방법을 적용하여 지하철뿐만 아니라, 버스도 통합한 통합 대중교통 접근성을 제공하는 기능을 선정함

❍ 또한 주거 지역의 거주자 주차 수요를 추정하는 모델을 개발하고 이를 바탕으로 주차 수급 상태를 나타내는 주차 Footprint Index를 도출하는 기능을 선정하였는데, 주차문제는 국민체감 교통불편 1위, 지자체 민원 1위로서 국민안전을 위협함은 물론 지역상권 활성화의 어려움을 주기 때문에 본 기능이 필요하다고 판단되어 선정함

⑥ [단위성과 F-6] 공간 빅데이터 통계기반 분석 컴포넌트 기초 설계

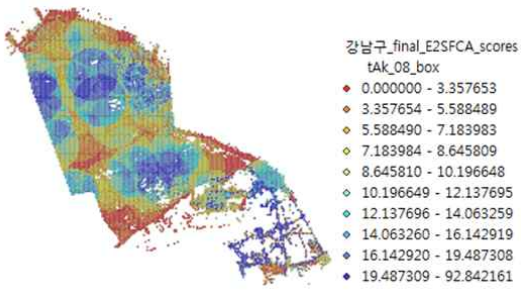
❍ 공간 빅데이터 통계기반 분석 컴포넌트로 대중교통 접근성 분석을 수행함

❍ 기존의 지하철 접근성의 한계를 보완하기 위해, 버스 정류장/노선 운행과 지하철역/노선 운행 정보를 반영한 대중교통 접근성 서비스를 제공하고자 함

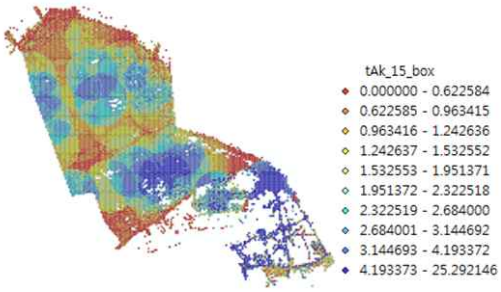
❍ 접근성을 구하는 방법으로 수요와 공급을 모두 고려한 Enhanced two-step Floating Catchment Area Method (E2SFCA)를 본 대중교통 접근성 서비스 구현 계산에 적용하였으며, 정류장별 대중교통 수요 대비 공급량(Rj) 값을 계산하고, 유동인구 조사지점별 대중교통 접근성 지수(Ak) 값을 계산하여 E2SFCA score를 계산함

❍ 데이터의 전처리를 통해, 버스 데이터, 지하철 데이터, 유동인구 데이터를 E2SFCA Score에 계산에 적합하게 변환하고, E2SFCA Score 계산을 통해서 각 지역별 ‘유동인구 조사지점별 대중교통 접근성 지수’ 도출한 뒤 QGIS를 통해 E2SFCA Score를 시각화함

❍ 아래는 E2SFCA를 통하여 구한 9월 출근 시간대 (오전 8시)와 오후 시간대 (오후 3시)의 최종 대중교통 접근성 지수로 값이 클수록 접근성이 좋음을 의미함



<그림 2-252> 오전 8시 최종 대중접근성 지수



<그림 2-253> 오후 3시 최종 대중접근성 지수

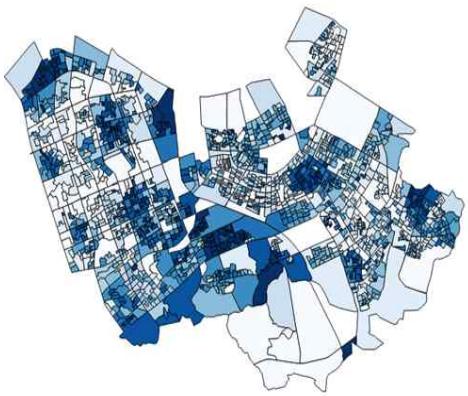
⑦ [단위성과 F-7] 공간 빅데이터 비통계기반 분석 컴포넌트 기초 설계

❍ 공간 빅데이터 통계기반 분석 컴포넌트로 Parking Lot Footprints를 개발함

❍ 본 기능은 주거 지역의 거주자 주차 수요를 추정하는 모델을 개발하고 이를 바탕으로 주차 수급 상태를 나타내는 주차 Footprint Index를 도출함

❍ 자동차 등록대수 데이터, 다세대-다가구 연립주택 데이터 그리고 유동인구 데이터 통계기반 분석컴포넌트 기초설계를 위하여 Parking Lot Footprints에 대한 컴포넌트 기초설계를 수행함

❍ 실 거주자를 통계청의 거주민 등록 수가 아닌 공간 빅데이터로 산정하여 Parking lot footprints를 도출하였으며, 집계구별 다세대/다가구 밀집 특별 주차 조사구역 여부, 거주자 주차 수급 만족 비율, 거주자 주차 거주지 수급 만족 비율, 권장 주차장 확충 비율 등을 보여줌



집계구ID	1123064050003
동, 구	강남구, 역삼1동
다세대/다가구 밀집 특별 주차 조사 구역	해당
거주자 주차 수급 만족 비율	67.07 %
거주자 주차 거주지 수급 만족 비율	50 %
권장 주차장 확충 비율 (現 대비)	49.08 %

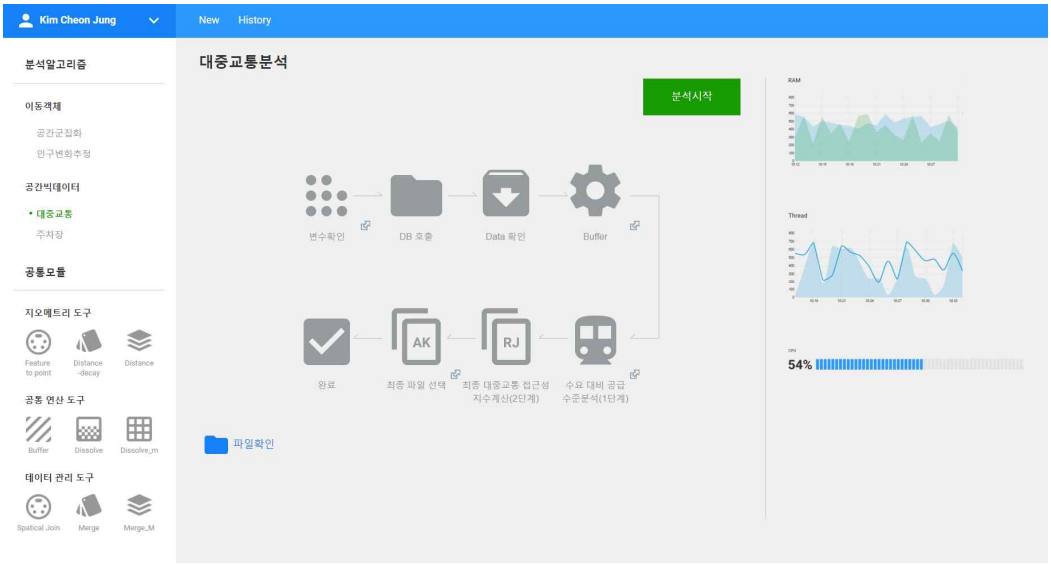
<그림 2-254> Parking Lot Footprints 결과

⑧ [단위성과 F-8] 공간 빅데이터 분석 컴포넌트 프로토타이핑

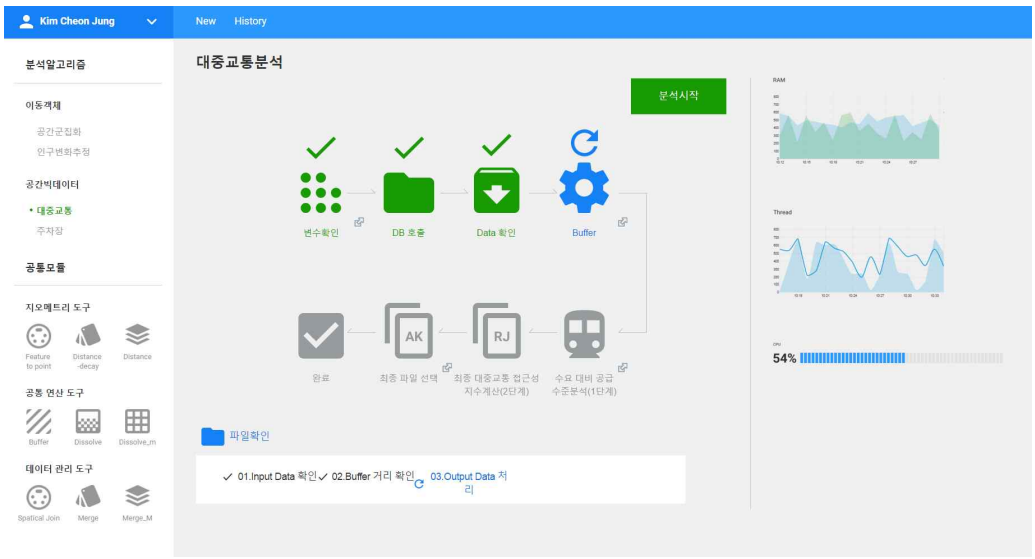
❍ 공간 빅데이터 분석 컴포넌트 프로토타이핑을 수행하고, 연산자 동작 정확도

및 기준대비 처리속도를 측정함

❍ 데이터 질의 / 응답 - 분석 - 결과보고의 과정을 통해 사용자에게 결과를 보여주는 프로토타입을 제작함



<그림 2-255> 공간 빅데이터 분석 컴포넌트 프로토타이핑



<그림 2-256> 대중교통 접근성 분석 진행 UI

⑨ [단위성과 F-9] 공간 빅데이터 기본 분석 기능 개선

❍ 공간 빅데이터 기본 분석기능 추가 및 성능 향상, 기존 분석기능의 분석 개선 작업을 수행하였음

❍ 정의된 거리를 통하여 공간 가중행렬을 연산하는 Fixed Distance Band Matrix의 분산처리가 가능한 형태로의 연산 프로세스를 개발하여, 공간적 자기상관을 검증하는 알고리즘인 LISA 분석, Getis-Ord Gi* 등 공간가중행렬 연산이 요구되는 분석 컴포넌트들의 분산처리를 가능하게 함

```
Hori_index[0] = Minimum(Point_X)
Hori_index[1] = Minimun(Point_X) + 1(Maximum(Point_X) - Minimum(Point_X)/N)
      ⋮
Hori_index[N] = Minimun(Point_X) + N(Maximum(Point_X) - Minimum(Point_X)/N)

Verti_index[0] = Minimum(Point_Y)
Verti_index[1] = Minimun(Point_Y) + 1(Maximum(Point_Y) - Minimum(Point_Y)/N)
      ⋮
Verti_index[N] = Minimun(Point_Y) + N(Maximum(Point_Y) - Minimum(Point_Y)/N)

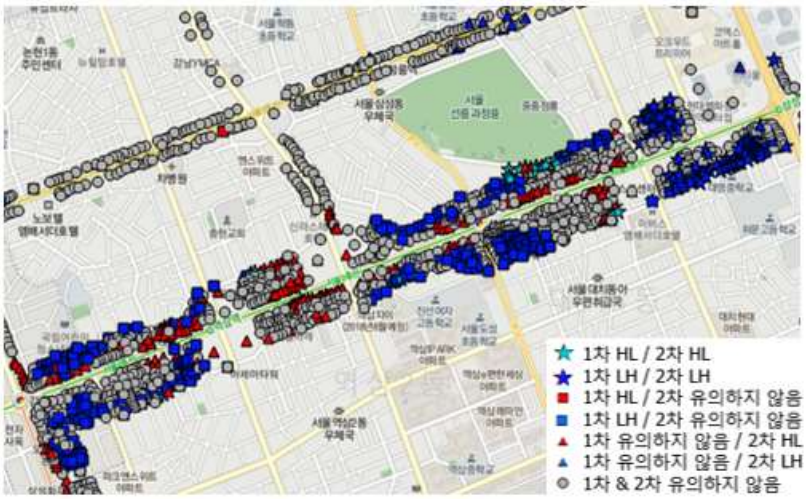
Sort Data with Point_X
For 1:SizeofData
 if (Hori_index[0] <= Point_X < Hori_index[1]) -> temp indexing  1
 else if (Hori_index[1] <= Point_X < Hori_index[2]) -> temp indexing  2
   ⋮
 else temp indexing  N

Sort Data with Point_Y
For 1:SizeofData
 if (Verti_index[0] <= Point_Y < Verti_index[1]) and temp index ==  A  -> indexing  A+(1-1)N
 else if (Verti_index[1] <= Point_Y < Verti_index[2]) and temp index ==  A  ->  indexing  A+(2-1)N
   ⋮
 else temp index ==  A  -> indexing  A + (N-1)N
```

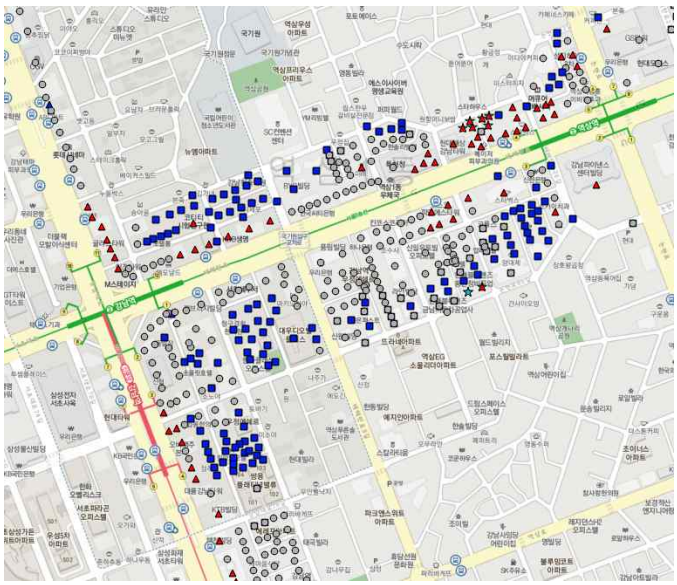
<그림 2-257> 그리드 인덱싱 과정 Pseudo-Code

⑩ [단위성과 F-10] 통계기반 분석 성능 향상

- ❍ 공간 빅데이터 일괄 분석 플랫폼의 통계기반 분석 성능을 향상시키기 위하여, 통계 기반의 신규 컴포넌트 기능을 개발함
- ❍ 공시지가 이상지점 발견 기능은 LISA(Local Indicator of Spatial Association) 방법을 통하여 주변 공시지가 및 유동인구에 대비한 공시지가 이상치를 발견하는 기능으로서, 공시지가의 형평성과 공평성을 위한 통계적 근거를 제시함
  - 주변 공시지가에 비해 공시지가가 높은 1차 HL outlier의 경우 공시지가를 낮추어 세금 감면을 고려해 볼 필요가 있는 지역으로, LH outlier의 경우 공시지가를 높여 세금 인상을 고려해 볼 필요가 있는 지역으로 파악됨
  - 주변 유동인구에 비해 공시지가가 높은 2차 HL outlier의 경우 공시지가를 낮추어 세금 감면을 고려해 볼 필요가 있는 지역으로, LH outlier의 경우 공시지가를 높여 세금 인상을 고려해 볼 필요가 있는 지역으로 파악됨



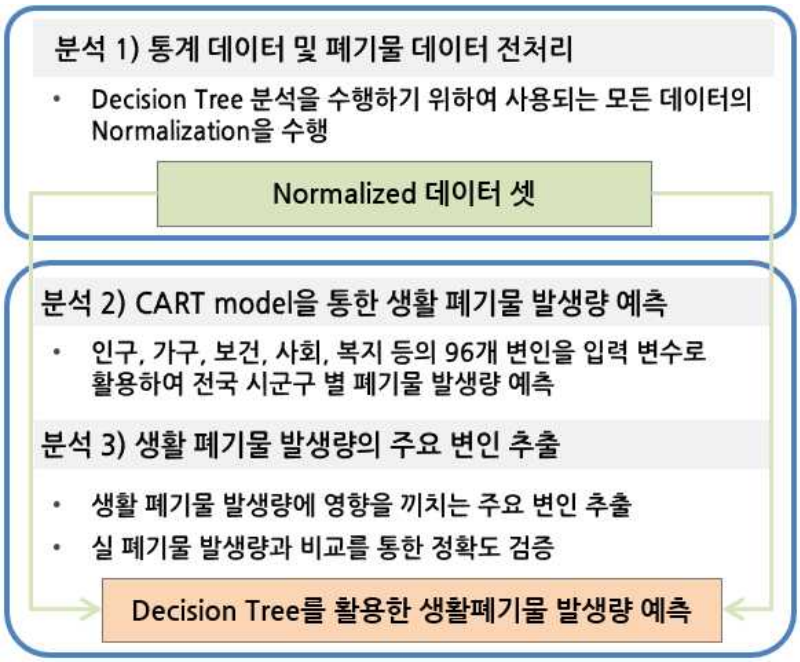
<그림 2-258> 강남구 상업지역의 통합 공시지가 이상지역 분석 결과



<그림 2-259> 강남역-역삼역 사이의 통합 공시지가
이상지역 분석 결과

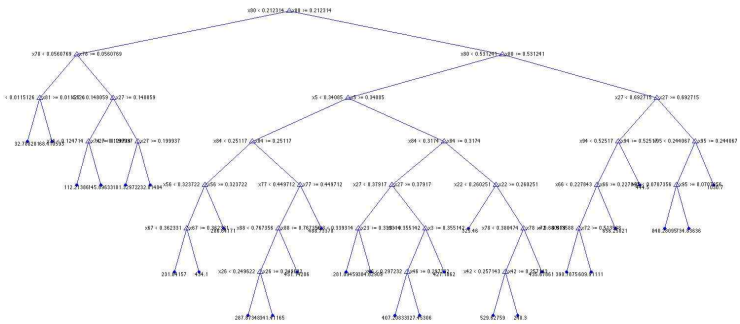
⑪ [단위성과 F-11] 비통계기반 분석 성능 향상

- ❍ 공간 빅데이터 일괄 분석 플랫폼의 통계기반 분석 성능의 향상을 위해, 비통계 기반의 신규 컴포넌트 기능인 의사결정나무를 이용한 생활 폐기물 추정 기능을 개발함
- ❍ 의사결정나무를 이용한 생활 폐기물 추정은 비통계 기반의 의사결정나무 방식을 이용하여, 각 시군구의 인구, 사회, 경제 통계량을 입력변수로 하여 폐기물 발생량을 예측하는 기능임



<그림 2-260> 의사결정나무 기능 Framework

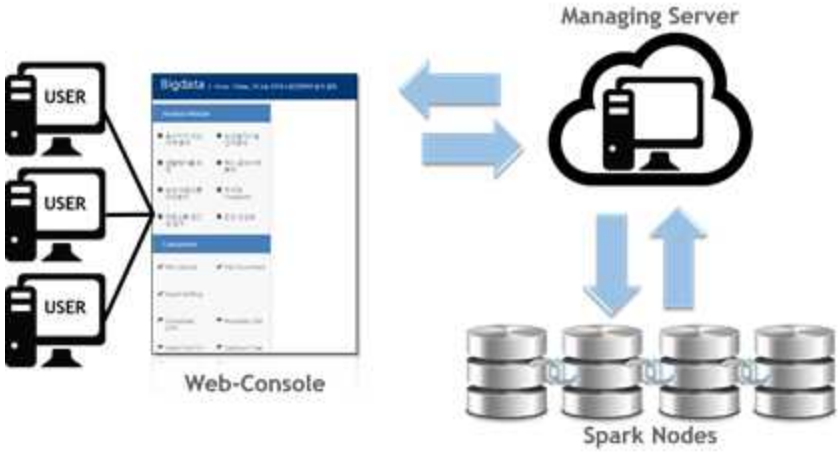
○ 의사결정나무 분석 결과, 생활 폐기물 추정에 있어 가장 연관된 상위 변수들은 모두 거주인구가 높은 시군구의 특성과 일치하였으며, 이를 통하여 지역의 거주 인구가 생활 폐기물 발생량과 높은 연관성을 가진다는것을 것을 확인함



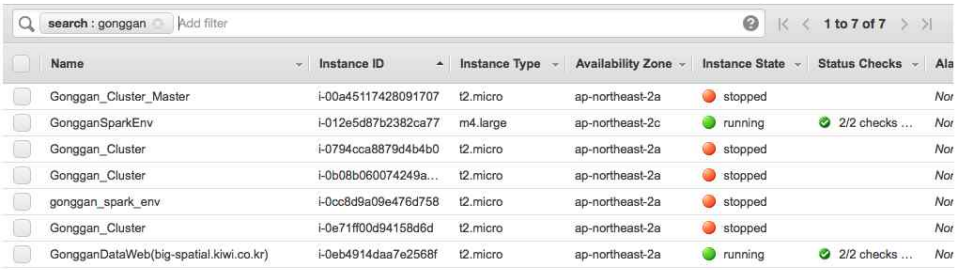
<그림 2-261> 의사결정나무 기능 분석 결과

⑫ [단위성과 F-12] 공간 빅데이터 분석 컴포넌트 기능 고도화

○ Apache Spark를 통한 In-Memory 방식의 분석과 분산처리 환경을 구성하였으며, AWS (Amazon Web Services)의 가상 인스턴스를 활용하여 서버를 구성하였음. 또한 Scala 언어를 통한 분석 scripting 작업으로 분석 모듈 실행환경을 구성하여 재사용성을 강화함



<그림 2-262> AWS와 Spark를 활용한 공간 빅데이터 분석 시스템



search : gonggan | Add filter | 1 to 7 of 7

Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks	Ala
Gonggan_Cluster_Master	i-00a45117428091707	t2.micro	ap-northeast-2a	stopped		Nor
GongganSparkEnv	i-012e5d87b2382ca77	m4.large	ap-northeast-2c	running	2/2 checks ...	Nor
Gonggan_Cluster	i-0794cca8879d4b4b0	t2.micro	ap-northeast-2a	stopped		Nor
Gonggan_Cluster	i-0b08b060074249a...	t2.micro	ap-northeast-2a	stopped		Nor
gonggan_spark_env	i-0cc8d9a09e476d758	t2.micro	ap-northeast-2a	stopped		Nor
Gonggan_Cluster	i-0e71ff00d94158d6d	t2.micro	ap-northeast-2a	stopped		Nor
GongganDataWeb(big-spatial.kiwi.co.kr)	i-0eb4914daa7e2568f	t2.micro	ap-northeast-2a	running	2/2 checks ...	Nor

<그림 2-263> AWS에 구성된 분석 클러스터

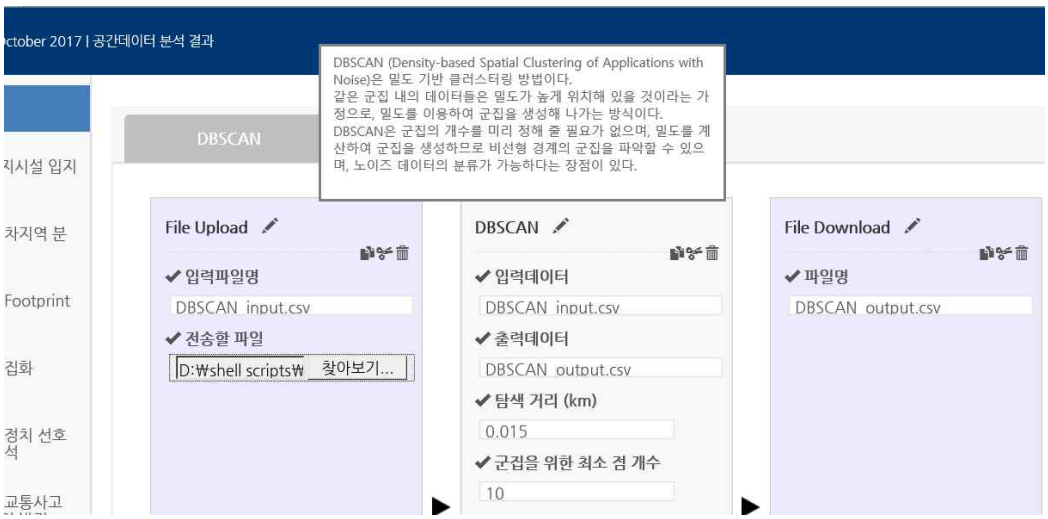
❍ 또한 웹 콘솔의 UI & UX 개선으로 직관적인 인터페이스를 조성함. 개발 기능 분석을 수행할 수 있는 Analysis Module 탭과 공간 빅데이터 분석 컴포넌트를 활용할 수 있는 Component 탭으로 구분하였으며, 3차년도 및 4차년도 개발 기능들을 모두 모듈 탭과 컴포넌트 탭에 탑재함



<그림 2-264> 웹 콘솔의 Analysis Module 탭과 Component 탭

❍ 분석기능의 모듈화를 통해 각 컴포넌트의 재사용이 가능하도록 Spark의 Scala로 개발하고, input setting을 통해 다양한 분석에 활용할 수 있도록 확정성을 높였으며, 재사용성을 강화한 Scala 어플리케이션을 쉘스크립트를 통한 편리한 실행환경을 구축하여 사용성을 강화함

❍ Web UI를 개선하여 분석 종료 후 결과를 다운로드 할 수 있도록 하였으며, 각 컴포넌트의 핵심 알고리즘에 대한 도움말 기능과 요구되는 변수에 따라 다양한 값을 받을 수 있도록 개발 컴포넌트를 수정함



<그림 2-265> DBSCAN에 대한 설명 팝업 화면

❍ 개선된 공간 빅데이터 분석 플랫폼의 성능을 평가하기 위하여, AWS의 분석 노드 개수를 늘려가며 기능의 분산처리 성능 테스트를 수행하고, 이를 외부

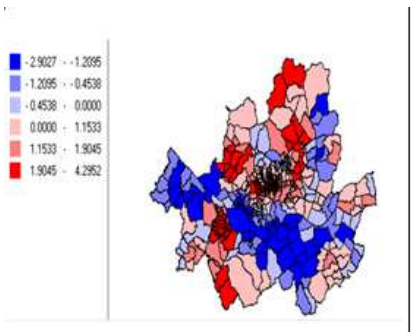
전문가로부터 평가받음

기관명	충북대학교 소프트웨어학과			
평가 확인자명	박 명 호 (서명)		소속/직책	초빙교수
전화번호	043-249-1239		E-Mail	bluemhp@chungbuk.ac.kr
홈페이지				
주 소	충청북도 청주시 충북대학교 S1-4동 317호			
성능평가 확인	Score - Model	합격	STSS	합격
	Random Forest	합격	Map - Matching	합격
	DBSCAN	합격		

<그림 2-266> 외부 전문가 소프트웨어 성능 평가 확인서

⑬ [단위성과 F-13] 공간 빅데이터 기본 분석 기능 추가

- 공간 빅데이터 기본 분석 기능을 추가 및 개선함
- 유동인구 대비 보행자 교통사고 건수가 다른 지역들에 비해 많은 지역을 탐색하는 보행자 교통사고 이상치 지점 탐색 기능을 개발함
- 기대 사고건수(= 전체 지역의 보행자 교통사고수 / 전체 지역의 인구수 * 해당 지역의 인구수)를 기반으로 실제 발생건수와 발생 기대건수 사이의 통계적 검정을 수행하는 Score model을 적용하여 각 행정동별 유동인구와 보행자 사고건수 사이에서 기대 사고보다 더 많은 사고가 발생하는 이상지역을 도출하는 기능임



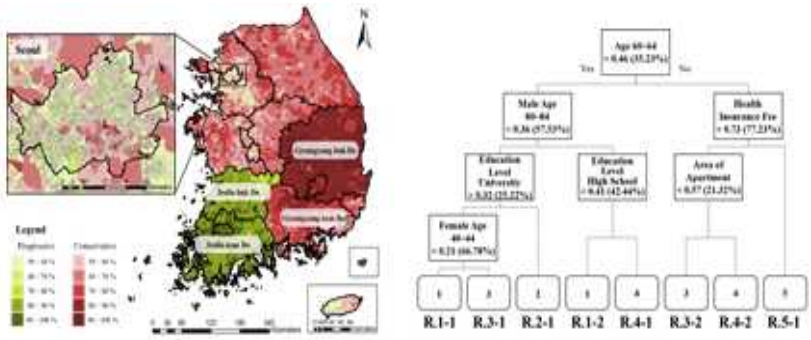
ID(1)	ID(2)	U(i)	Z[U(i)]	Obs	Exp
seoul	Beon-dong	2.405	2.415	3.000	1.880
seoul	Suyu-dong	0.262	0.262	2.000	2.461
seoul	Ui-dong	2.942	2.945	2.000	0.333
seoul	Ssangmun-dong	0.685	0.687	3.000	2.423
seoul	Banghak-dong	1.148	1.152	3.000	2.239
seoul	Chang-dong	0.124	0.125	2.000	2.330
seoul	Dobong-dong	2.938	2.944	5.000	1.498
seoul	Wolgye-dong	1.213	1.218	3.000	2.730
seoul	Gongneung-dong	-0.952	-0.955	2.000	3.212
seoul	Hagye-dong	-0.964	-0.968	1.000	1.547
seoul	Sanggye-dong	0.904	0.906	5.000	3.190
seoul	Junggye-dong	-1.595	-1.599	0.000	2.434
seoul	Susaek-dong	2.766	2.773	1.000	0.501
seoul	Nokbeon-dong	1.393	1.398	1.000	0.694
seoul	Bulgwang-dong	0.032	0.033	1.000	1.800
seoul	Galhyeon-dong	0.292	0.293	0.000	1.349
seoul	Gusan-dong	1.866	1.871	2.000	0.612
seoul	Yangjae-dong	-2.072	-2.080	1.000	3.676

<그림 2-267> 서울시 행정동별 보행자 교통사고 이상치 지점 탐색 결과

- 중요 지역 인자 검출을 통한 지역별 정치적 선호요인 분석은 표출된 결과로부터 원인 요인을 탐색하는 Reverse Engineering 관점의 탐색임. 18대 대선 결과에 영향을 미친 중요 지역인자들을 Random Forest 알고리즘을 통하여

도출함

❍ 전국 단위 투표구별 479개의 지역 인자 데이터 (인구통계, 경제 및 산업, 지역사회 건강 데이터 등)와 18대 대선 투표결과를 이용하여 특정 정당을 선호하게 유도하는 주요 지역인자들을 산출함



<그림 2-268> 18대 대선 결과 및 중요 지역인자 검출

❍ 인구 유동 패턴 기반 도시 군집화 기능은 DBSCAN 방법론을 통하여 도시 내 교통 관련 유동 정보로 도시 군집화를 찾는 기능임. 서울특별시의 교통기능을 군집화하여 새로운 도시 군집을 밝혀냄

❍ 대중교통 빅데이터를 활용해 인구 유동 패턴을 추출하고 지역적 기능 탐색을 통하여 새로운 도시 군집을 발견할 수 있도록 도움



<그림 2-269> 인구 유동패턴 기반 도시 군집 결과

⑭ [단위성과 F-14] 공간 빅데이터 통계기반 분석 성능 최적화 설계

❍ 공간 빅데이터 분석 기능 성능 최적화를 위해 Spark의 RDD(Resilient Distributed DataSet) 데이터 구조를 활용하여 분산처리가 가능한 개발을 진행

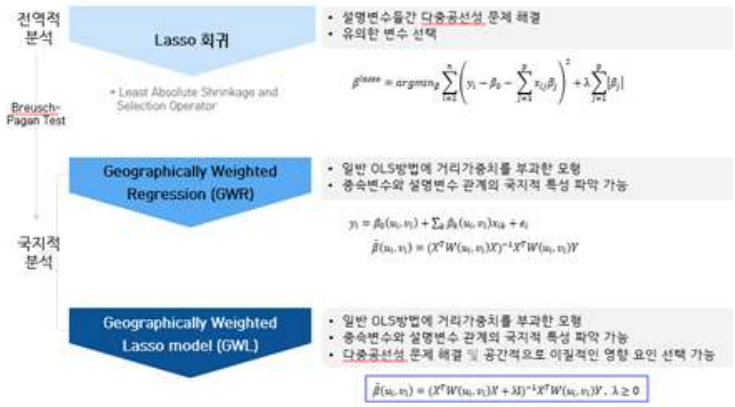
하였으며 연차별 연구 목표에 따라 신규 기능을 개발함

❍ 기 개발된 Score model 컴포넌트에 RDD를 기반으로 Map-Reduce를 적용하였으며, 분산처리 성능을 검증하기 위해 Amazon EMR (Elastic MapReduce) 서비스 상에서 실험 데이터를 사용한 처리속도 측정을 통해 결과를 평가함

<표 2-81> Score model 속도 측정 결과

사용 노드 수	소요 시간(분)
1	2.5
2	1.9
4	1.5
8	1.5
16	1.6

❍ 빠른 속도로 증가하고 있는 빈집의 발생 영향 요인을 탐색하기 위하여 Geographically Weighted Lasso (GWL) Model을 빈집 발생 영향 요인들로 알려진 데이터와 유동인구 데이터에 적용하여, 빈집 발생 영향 요인 탐색 기능을 개발함



<그림 2-270> 빈집 발생 영향 요인 탐색 Flow Chart

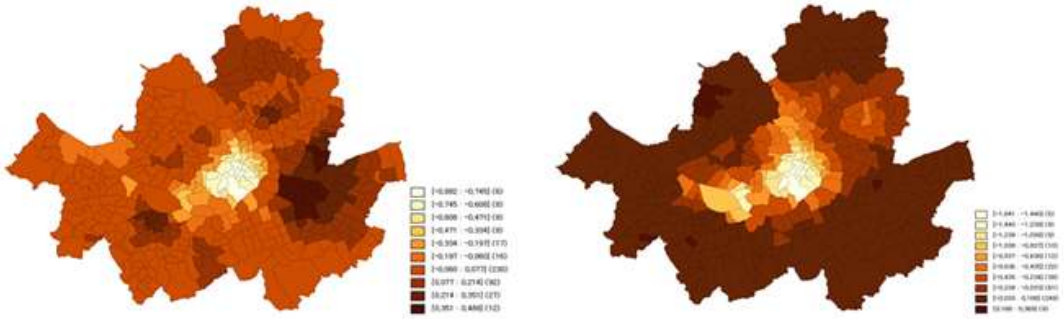
❍ 분석을 통해, 신규 주택 비율과 빈집발생 간의 상관관계, 노후주책 비율과 빈집발생의 상관관계 등을 파악하고, 기존의 미국, 일본의 연구결과와 비교함

❍ 서울시에서 거주하는 인구와 실제 생활하는 인구는 그 수치에 있어서 차이가 크며, 서울에서 생활하는 인구 데이터를 이용하여 시간대별 생활인구 이상치 (폭발적 증가 / 감소)를 파악하는 기능인 시간대별 생활인구 이상치 탐색 기능을 개발함

❍ 시간대별 생활인구 이상치 탐색 기능은 유동인구 및 이동객체 데이터의 지역별 증감 현상을 이용해 Volcano / Swarm을 찾아내고, 이를 통하여 Event Detection을 함. SVDD (Support Vector Data Description)는 데이터의 확률분

포가정이 필요 없이 kernel function을 이용하여 주어진 데이터에 맞게 학습할 수 있음

변수	선행연구 결과	GWL (Φ = 712.5488, λ = 0.001)				
		Min	Mean	Max	POS	NEG
5년 미만 신규주택 비율	신규주택비율: 양의 영향	-0.882	0.009	0.488	285	139
15년 이상 노후주택 비율	노후주택비율: 양의 영향	-1.641	-0.170	0.369	203	221



<그림 2-271> 신규주택 비율(좌) 및 노후주택 비율(우)에 대한 빈집발생 영향요인 분석 결과



<그림 2-272> 시간대별 생활인구 이상치 탐색

⑮ [단위성과 F-15] 공간 빅데이터 비통계기반 분석 성능 최적화 설계

- ❍ 공간 빅데이터 분석 기능 성능 최적화를 위해 Spark의 RDD(Resilient Distributed DataSet) 데이터 구조를 활용하여 분산처리가 가능한 개발을 진행하였으며 연차별 연구 목표에 따라 신규 기능을 개발함
- ❍ 기 개발된 Random Forest과 DBSCAN 컴포넌트에 RDD를 기반으로 MapReduce를 적용하였으며, 분산처리 성능을 검증하기 위해 Amazon EMR 서비스 상에서 실험 데이터를 사용한 처리속도 측정을 통해 결과를 평가함



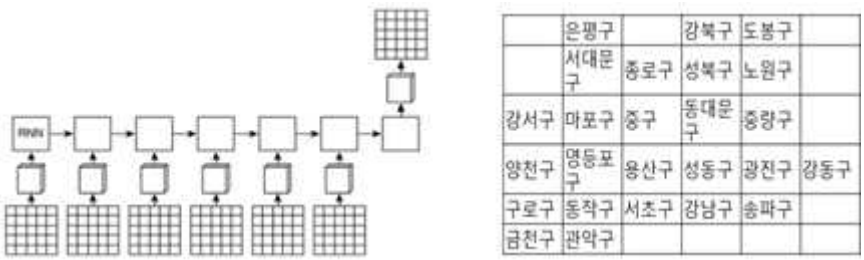
<그림 2-273> Random Forest 분산처리 모니터링 화면

<표 2-82> DBSCAN 컴포넌트 분산처리 속도 측정 결과

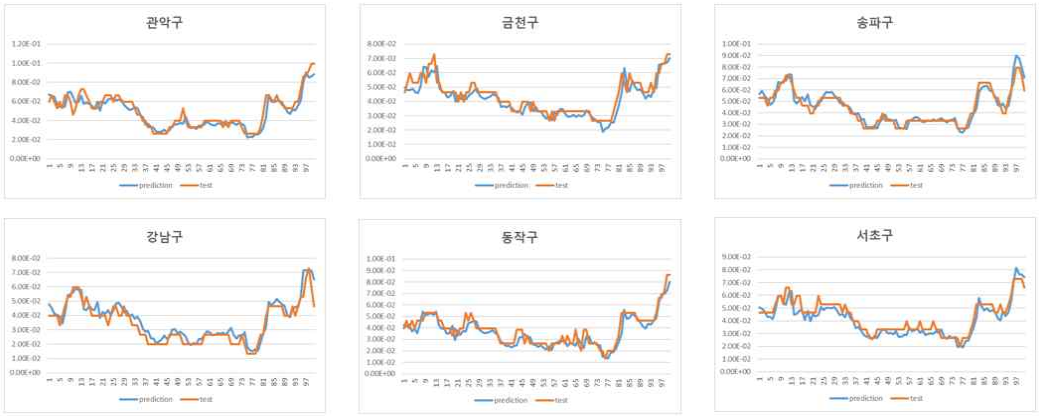
사용 노드 수	소요 시간(분)
1	6.0
2	4.5
4	3.2
8	2.4
16	2.3

❍ 공간적 영향 및 시간적 영향을 동시에 고려하는 모델인 시공간 분석 기능을 개발하기 위하여 시계열 분석 신경망 방법론 중 가장 대표적인 LSTM (Long-Short Term Memory)를 적용하였으며, 이를 서울시 대기 시계열 빅데이터에 적용하여 대기오염물질 농도를 예측하는데 활용함

❍ 대기 데이터 같은 경우 주위 지역과의 연관성이 있기 때문에 CNN을 적용해 근처 지역과의 연관성을 반영한 Feature map을 도출한 뒤 LSTM 모델에 학습을 진행함



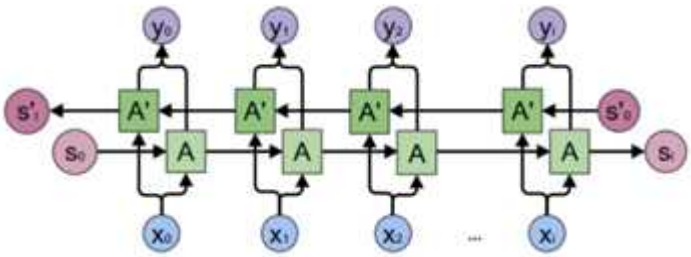
<그림 2-274> 딥러닝을 활용한 대기오염물질 농도 예측



<그림 2-275> 서울시 6개구 일산화탄소 예측 결과 (단위:ppm)

❍ 공공자전거 운영관리 비용의 큰 부분을 차지하는 공공자전거의 재배치 문제의 해결을 목표로 대용량 시계열 대여소 데이터의 학습을 통해 대여소 상태를 예측하는 공공 자전거 최적 재배치 기능을 개발함

❍ RNN을 기반으로 현재까지 축적된 데이터와 실시간으로 취득되는 대여소의 상황을 입력 자료로 활용하여 임의의 시점의 대여소의 수요 및 공급량을 예측함



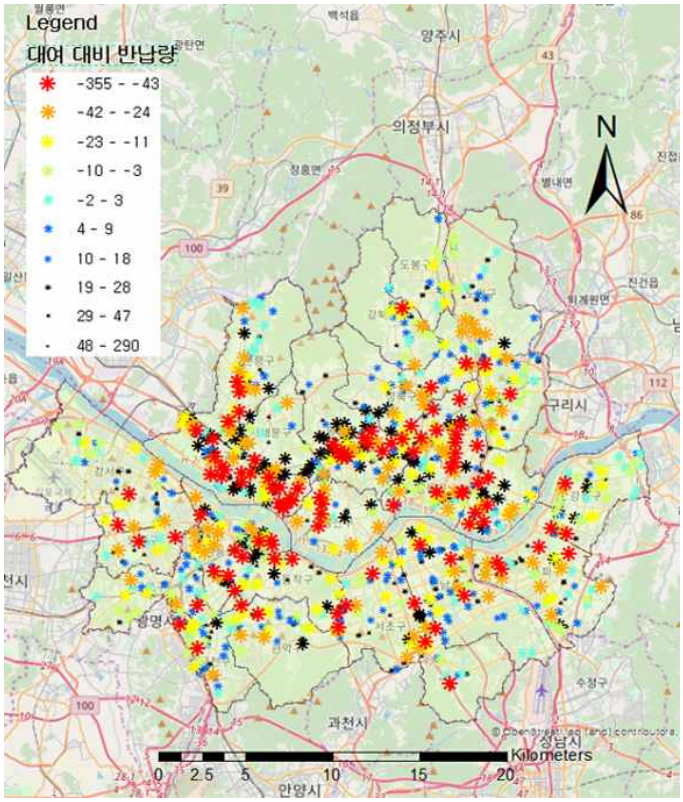
<그림 2-276> Recurrent Neural Network (RNN)

❍ 기존 재배치 차량의 효율적인 운영을 위하여 도로 네트워크와 대여소 위치 데이터, 수요 예측 결과를 이용해 재배치 시간 및 수량의 최적화 방안 도출할 수 있음

⑯ [단위성과 F-16] 공공/민간 수요기반 Batch 분석 시나리오 설계

❍ 개발된 기능의 활용에 대한 검증을 위해, 세종시 건축도시공간연구소(AURI)의 행정중심복합도시 내 유동인구 패턴을 연세대학교가 개발한 기능으로 분석함

❍ 분석 시나리오에 따라, 시간단위, 나이대 단위, 성별로 구분된 유동인구 데이터를 분석하였으며, 인구 유출입양이 많은 지역을 도출함



<그림 2-277> 따릉이 자전거 대여소별 부족/과다량 도시화

<표 2-83> SK 지오비전 유동인구 데이터 명세

Data Category	Variable	Description or Category
Floating population	Date	Data generation date, "YearMonth"
	X coordinate	UTM-K coordinate system
	Y coordinate	UTM-K coordinate system
	Age group	1=10~19, 2= 20~29, 3=30~39, 4=40~49, 5=50~59, 6=60~ (years)
	Time period	from k hour to k+1 hour (k=00, ..., 23)
	Gender	0 = Male, 1 = Female
	Weekday	Monday ~ Sunday
	Floating population	Monthly Average Floating Population
Official land price	Name of district at level 1	The name of Si / Do
	Name of district at level 2	The name of Si / Gun / Gu
	Name of district at level 3	The name of Dong
	Land code	The official number of the land
	Official land price	The official land price
	Code of district at level 1	The code of Si / Gun / Gu
	Code of district at level 2	The code of Dong
	Code of district at level 3	The code of the plot
	Plot name	The name of the plot
	Main serial land number	Main serial number
	Sub serial land number	Sub serial number
	Announced year	Announced year

❍ AURI와의 현장 검증을 통해, 정책연구기관의 분석 연구에 있어, 연세대학교에서 개발한 분석 기능들을 실제로 활용하여 분석 결과를 도출하였으며, 지역의 유동인구 패턴 도출에 활용 가능함을 확인 함



<그림 2-278> 행정중심복합도시 내 유동인구 패턴 분석 결과

❍ 세종시 내 행정복합도시의 인구 분포를 분석하여 미세먼지 관련 정책 연구에 사용함. 건축도시공간연구소에서 주관하여 작성한 “거주자 행태를 고려한 생활밀착형 미세먼지 대응방안 연구 Reducing Daily Exposures to Particulate Matter through Urban Environment” 보고서에 관련 결과가 담김

❍ AURI와의 현장 검증을 통해, 연세대학교에서 개발한 분석 기능들의 활용 가능성을 확인하였으며, 실제 공공/연구기관에서 본 기능을 활용해 정책 연구를

수행할 수 있음을 입증함

❍ Hourly Comparison을 위한 세 가지 시나리오, Age Comparison을 위한 세 가지 시나리오, Gender Comparison을 위한 세 가지 시나리오를 만들어 유동인구 패턴을 분석하여 시각화 함

⑰ [단위성과 F-17] 공간 빅데이터 Batch 솔루션 설계

❍ 4개 공간 빅데이터 분석 기능 추가로 총 15개 공간 빅데이터 분석 기능 개발 완료 및 Batch 솔루션 설계

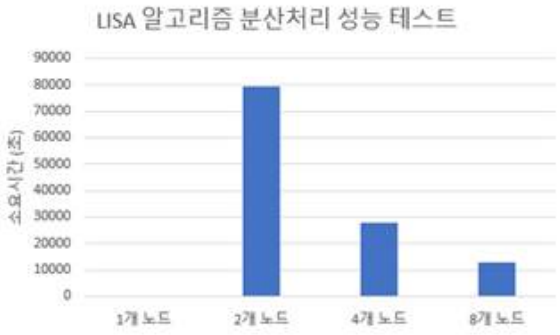
❍ 1차년도부터 6차년도까지 개발한 공간 빅데이터 Batch 분석 솔루션을 이루는 분석 컴포넌트는 총 15개이며, 이를 기반으로 분석 솔루션을 설계함

<표 2-84> 15개 공간 빅데이터 Batch 분석 컴포넌트

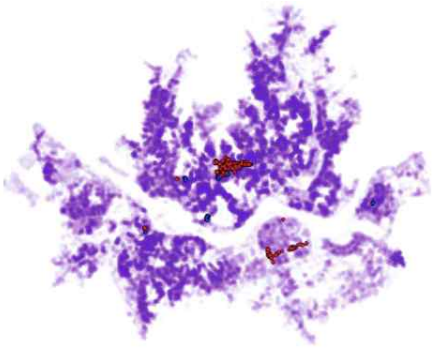
	공간 빅데이터 Batch 분석 컴포넌트 명
1	대중교통 접근성 분석
2	주차장 사용성 풋프린트 (footprint) 분석
3	공시지가 이상지역 분석
4	생활폐기물 추정
5	지역별 정치 선호도 주요 요인 도출
6	보행자 교통사고 이상치 지점 탐색
7	인구유동패턴 기반 공간 군집화
8	공공 자전거 최적 재배치 시스템
9	딥러닝 기반 시공간 분석 기능
10	시간대별 생활인구 이상치 탐색
11	빈집 발생 영향요인 탐색
12	유동인구 및 인구통계 기반 공간 군집화
13	특정 지역, 특정 시점의 인구 유동량 예측
14	유동인구 데이터를 이용한 주요 패턴 도출
15	공간 가중치를 고려한 유병율 회귀분석

❍ 지속적으로 연구 및 개발을 진행하며, 기 개발된 분석 컴포넌트들에 대한 성능 개선 작업 지속 진행함

❍ 공간 빅데이터 연산 효율을 높이기 위해 Locality Sensitive Hashing 방식을 적용하였으며 이를 활용해 Getis Ord-Gi*, Score Model, LISA 등의 알고리즘에서 분석 속도 및 분석 효율이 높아짐을 확인함



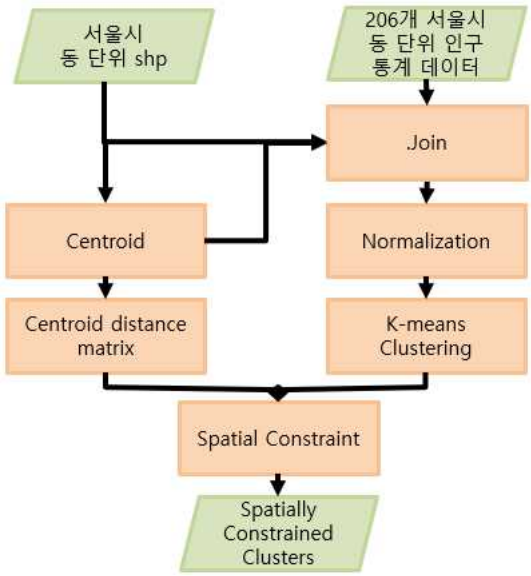
<그림 2-279> LISA 알고리즘 분산처리 성능 테스트 결과
(1개 노드: 메모리 부족으로 처리 불가)



<그림 2-280> 서울시 공시지가 이상지역 분석 결과

❍ 공간 빅데이터 분석 기능 추가 - 유동인구 및 인구통계 기반 공간 군집화

❍ 유동인구 및 인구통계 기반 공간 군집화 기능은, 실제 근접하지 않은 지역이 동일 군집으로 분류되는 기존 공간 군집화의 한계점을 극복하여, 공간적 일관성을 가지는 공간 군집화를 할 수 있어 더 고도화된 지역별 공간 군집을 도출할 수 있음

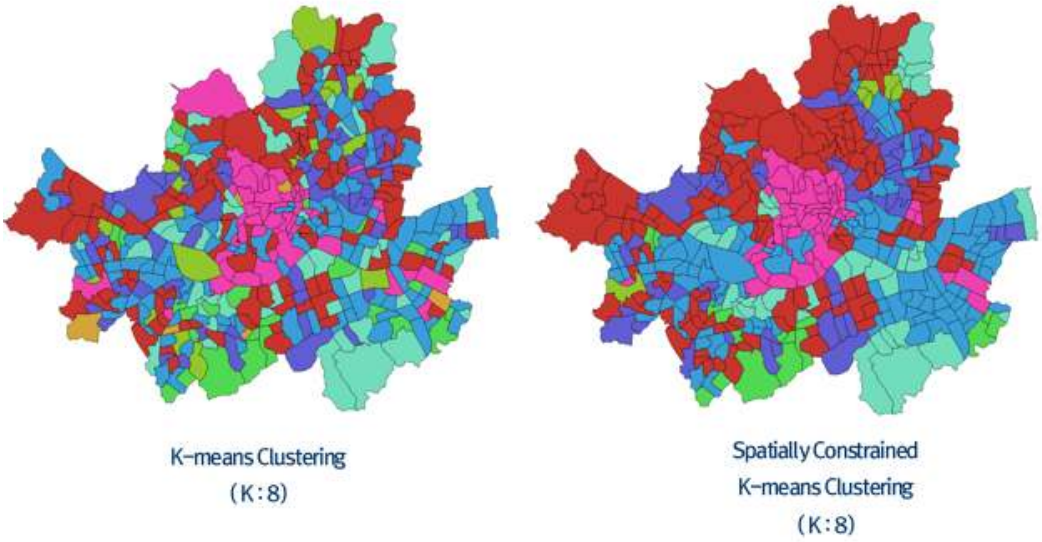


<그림 2-281> 공간제약 클러스터링의 Workflow

❍ 공간제약 클러스터링을 구현하기 위해 서울시 206개 동단위 인구통계 데이터를 활용하였으며, 공간 연산자를 활용하여 해당 데이터를 정제함

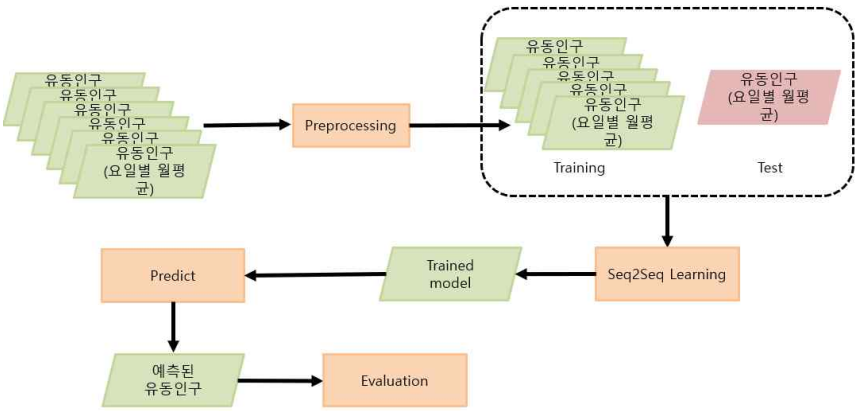
❍ 계산된 거리 메트릭스와 공간 군집을 이용하여, 최인접 중점 좌표 (Nearest Neighbor)를 탐색하고 공간 군집이 상이한 경우, 같은 공간 군집으로 변환함

❍ 이를 전체 포인트에 반복하여 최종적으로 공간적 제약조건을 가진 공간 군집을 도출함



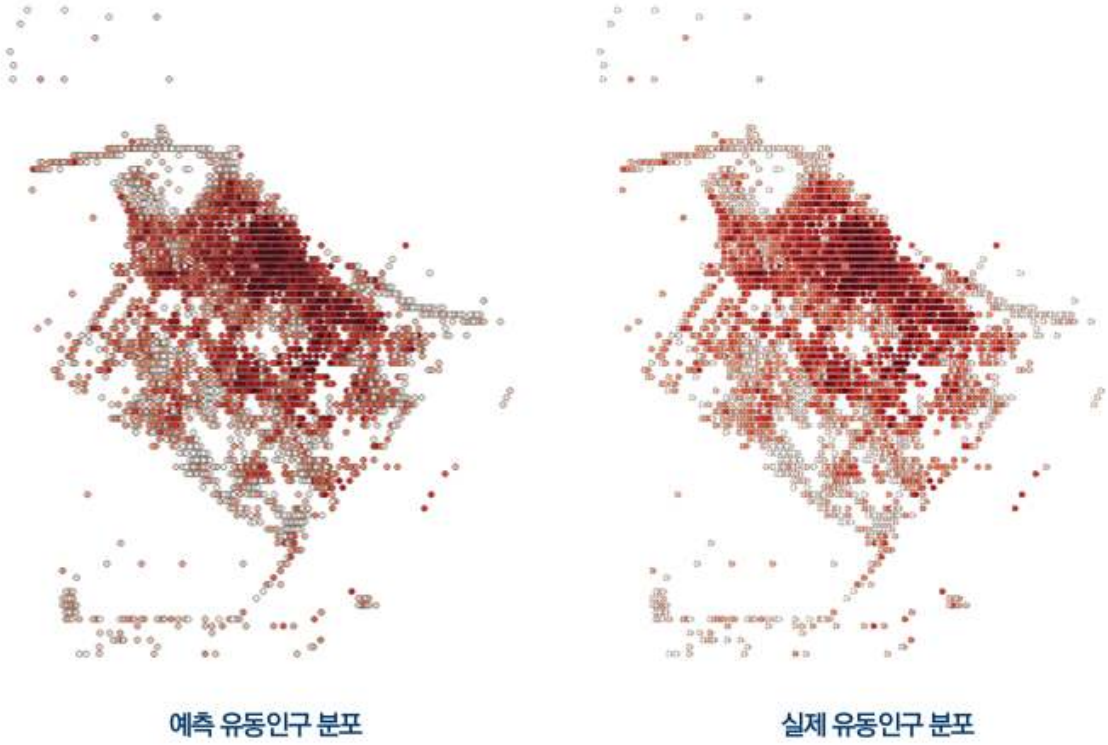
<그림 2-282> 2차년도 공간 군집화 결과와, 공간 제약 클러스터링의 결과 비교

○ 클러스터링 결과를 기 개발된 K-menas Clustering 방식과 비교하였을 시 공간적으로 근접한 지역이 같은 군집으로 분류됨을 확인할 수 있음
○ 공간 빅데이터 분석 기능 추가 - 특정 지역/시점 인구 유동량 예측 기능
○ 특정 지역 / 시점 인구 유동량 예측 기능에서는 인구 유동에는 시간 흐름에 따른 변화 패턴이 존재한다는 점을 이용하여, 유동인구 데이터에 시간적인 상관관계를 고려한 Long-Short Term Memory (LSTM) 기반의 Sequence to sequence를 적용함



<그림 2-283> 특정 지역, 특정 시점의 인구 유동량 예측 기능의 Workflow

○ 기능 구현을 위해 SK 지오비전에서 수급한 유동인구 데이터를 활용하였으며, 딥러닝 모델에 활용하기 위해 유동인구의 정규화를 진행한 뒤 Seq2Seq Learning 알고리즘에 적용함

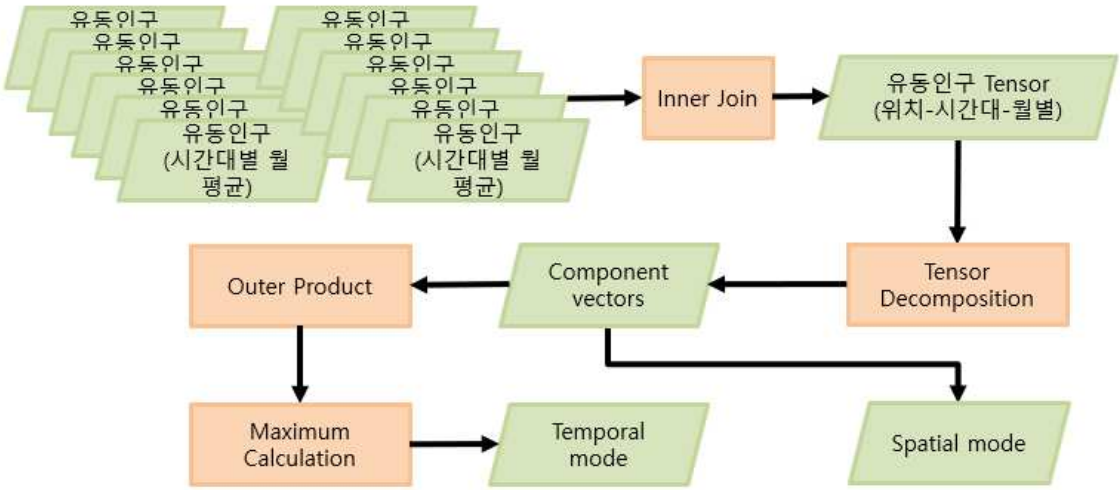


<그림 2-284> 예측 유동인구의 분포와 실제 유동인구 분포의 시각화

❍ 알고리즘에 의해 분석된 결과와 실제 유동인구 분포를 활용해 구축된 모델의 정확도를 판별하였으며, RMSE 8.3명, %RMSE 10.6%로 도출 됨

❍ 공간 빅데이터 분석 기능 추가 - 유동인구 데이터를 이용한 주요 패턴 도출

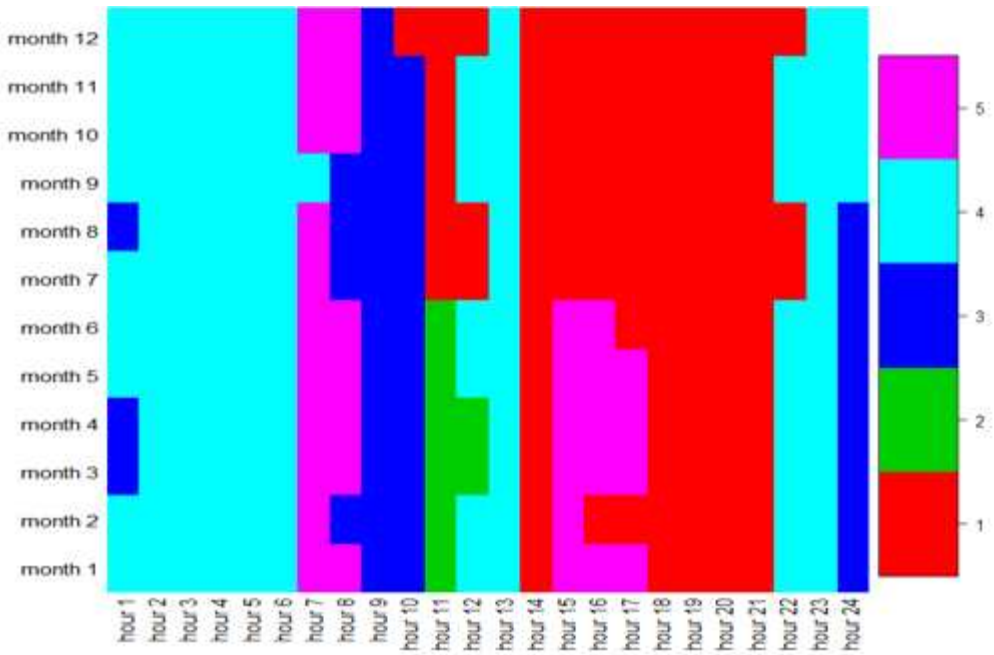
❍ 유동인구 데이터는 성별 / 연령별 / 요일별의 3개 축을 갖는 tensor라고 할 수 있음. 다차원 빅데이터에서 패턴과 비정상 특성을 찾는데 활용되는 Tensor decomposition을 이용하여 성별 / 연령별 / 요일별 주요 이상 패턴을 도출함



<그림 2-285> 유동인구 주요 패턴도출 기능의 Workflow

❍ 각 축에 대하여 도출된 5개의 컴포넌트 벡터는 각각 유동인구의 위치 특징, 시간대 특징, 월별 특징을 담고 있으며, 이들의 외적을 통하여 두 특징을 비교할 수 있음

❍ 시간 축을 담당하는 두 컴포넌트 벡터의 외적을 통해서 기존 Tensor가 담고 있는 특징에 대한 시각화가 수행되며, 월별 - 시간대 별 인구 유동 특징을 표현하는 매트릭스가 계산되고, 총 5개 매트릭스 중 최대값을 가지는 셀을 시각화하여 패턴을 구분함



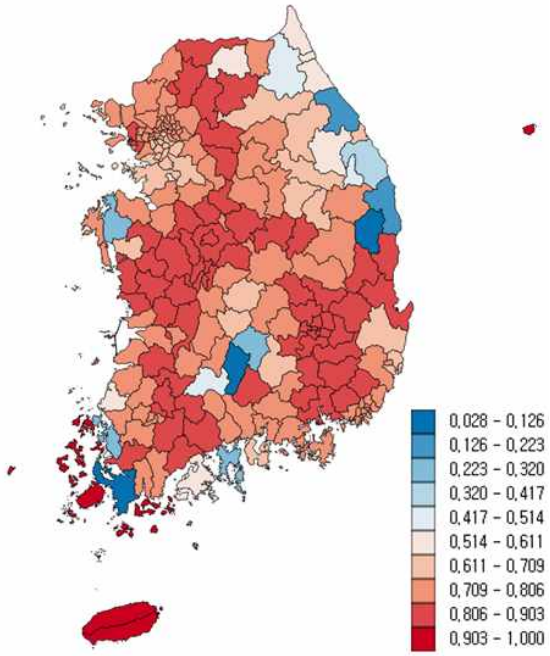
<그림 2-286> 시간대-월간 인구 유동 특징 분석 결과

❍ 분석 결과, 오후 2시~9시 사이의 군집 (1번)과 밤 10시~아침6시 사이 (4번)의 두 큰 패턴과 나머지 시간대의 작은 패턴으로 구분되는 것을 확인하였음



<그림 2-287> 1번 컴포넌트와 4번 컴포넌트의 공간적 분포 특징 매핑 결과

- 1번과 4번 군집간 공간 분포 분석 결과, 주거 지역과 상권 지역에 대한 큰 분포의 차이가 발생하는 것으로 확인되었으며, 이는 송도 지역의 인구유동의 패턴이 가장 큰 차이를 보이는 공간적 요인이 주거 지역과 상업 지역에 의한 구분이라는 것을 의미함
- 공간 빅데이터 분석 기능 추가 - 공간 가중치를 고려한 유병율 회귀분석
- 유병율은 지역적 특성에 따라 관련성의 크기 및 방향이 지역적으로 차이가 있으므로, 해당 기능은 지역간 특성의 차이를 반영할 수 있는 지리적 가중회귀분석 (Geographically weigthted regression, GWR)을 통해 유병율의 지역적 특성을 분석함
- 지역 간의 특성의 차이를 반영할 수 있는 지리적 가중회귀분석 모델을 사용하여, 고혈압 유병율과 연관성이 높다고 알려진 당뇨, 이상지질혈증, 뇌졸중에 대한 유병율을 종속변수로 사용하여, 독립변수인 고혈압과의 지역적 상관관계를 파악함



<그림 2-288> 공간 가중치를 고려한 유병율 회귀분석 Local R2 매핑 결과

○ 독립변수로 쓰인 세 질병의 유병률이 내륙 지역에서는 고혈압의 유병률과 높은 상관성이 있으나, 일부 해안 지역에서는 낮은 상관성을 보이는 것으로 나타남

⑱ [단위성과 F-18] 공간 빅데이터 Batch 솔루션 구현

○ 개발된 공간 빅데이터 분석 기능을 분산처리 기술을 적용하여 구현하고, 분산처리 성능을 확인함

○ GWL 컴포넌트, CNN-LSTM 컴포넌트, SCC 컴포넌트, Seq2Seq 컴포넌트, Tensor Decomposition 컴포넌트에 대한 구현을 수행하였으며, 연산 노드를 늘려가며 분산처리 처리성능을 확인함

<표 2-85> CNN-LSTM 컴포넌트 분산처리 속도 측정 결과

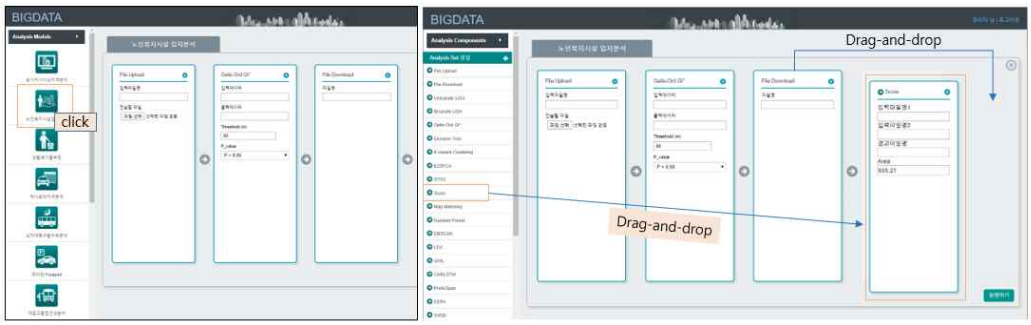
사용 노드 수	소요 시간(분)
1	40.18
2	22.25
4	12.57
8	9.17
16	7.52

❍ 총 15개 공간 빅데이터 분석 기능을 포함한 Batch 솔루션 구현
❍ UI / UX 측면에서 각 분석 컴포넌트별 독립성을 보장하여 확장성과 사용성을 높이고, Web-based platform 상에 drap-and-drop 기능을 구현함



<그림 2-289> 공간 빅데이터 분석 Batch 솔루션

❍ 분석 컴포넌트를 조합하여 분석 시나리오를 자유롭게 설정할 수 있으며 저장 및 편집이 가능함



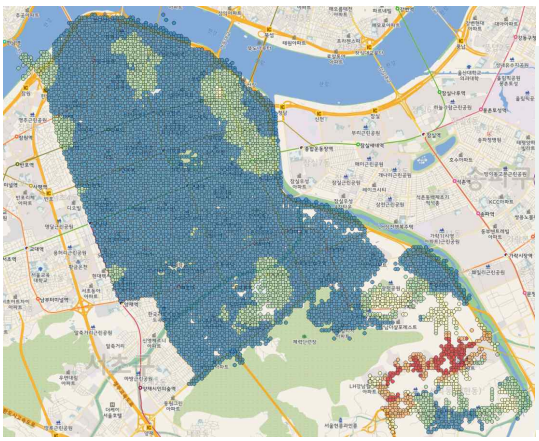
<그림 2-290> Drag-and-drop 및 저장 / 편집 기능

⑲ [단위성과 F-19] 공간 빅데이터 분석 솔루션 사업단 실증

❍ 국토공간정보의 빅데이터 관리, 분석 및 서비스 플랫폼 기술개발 연구단의 R&D 성과물이, 공간빅데이터 분석 플랫폼 (사업단 플랫폼)에 적용이 가능한지 실증 테스트를 수행하는 시험을 진행하였음
❍ 1세부 연구성과물과의 통합연계를 보이기 위해서 1-1세세부, 1-3세세부의 개발 기능의 결과와 2-2세부(연세대학교, 가이온)의 공간 빅데이터 분석 기능 2종 (E2SFCA, Decision Tree)을 공간빅데이터 분석 플랫폼 내에 설치하

여 기능 연동을 수행하고, 그 결과를 확인하였음

- ❍ E2SFCA 모델은 공간 빅데이터 분석 과제 2차년도에 개발한 대중교통 접근성 분석 모델의 분석 컴포넌트로, 강남구를 대상으로 강남구 버스 데이터, 강남구 지하철 데이터 및 강남구 유동인구 데이터를 통해 강남구의 대중교통에 대한 접근성을 평가하는 분석 기능임
- ❍ Decision Tree 컴포넌트는 공간 빅데이터 분석 과제 3차년도에 개발한 생활 폐기물 요인 분석 모델의 분석 컴포넌트로, 시군구별 다양한 인구, 사회, 경제 변수를 통해 시군구의 생활 폐기물을 발생량을 추정하고 어떤 변수가 생활 폐기물 발생에 큰 영향을 미쳤는지 확인하는 기능임
- ❍ 역시 분석 모델 수행을 위한 데이터는 1세부의 성과물을 활용하여 가져왔으며, 이를 2-2세부(연세대학교, 가이온)에서 개발한 분석 컴포넌트에 적용함
- ❍ 분석 결과가 파일로 생성되고 화면에 가시화됨을 확인함

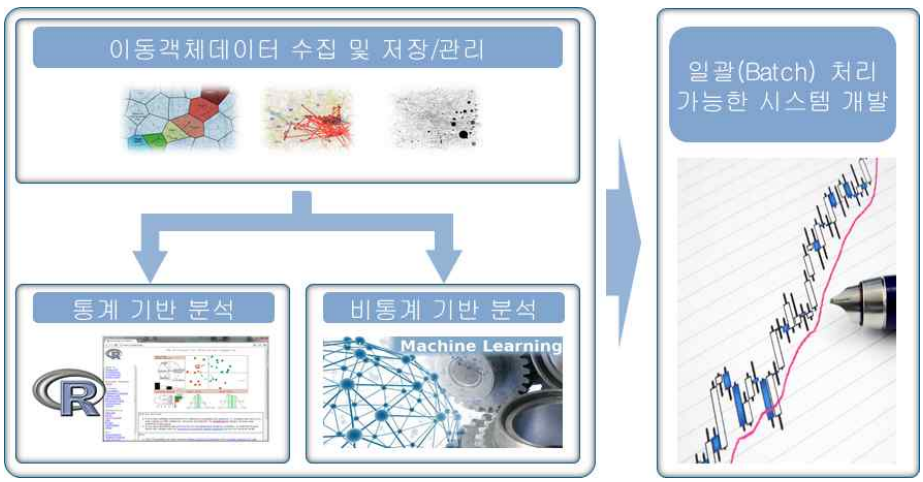


<그림 2-291> E2SFCA 분석컴포넌트 결과



<그림 2-292> Decision Tree 분석 컴포넌트 결과

(타) 이동객체데이터 Batch 분석 컴포넌트



<그림 2-293> 주요연구내용

○ 이동객체 데이터 Batch 분석 컴포넌트는 1세부과제에서 구축된 빅데이터 저장 관리 플랫폼을 통해 전달된 이동객체 데이터에 통계기반·비통계기반의 분석기법 도구들을 일괄(Batch)처리로 적용 가능한 시스템을 의미함

○ 통계적·비통계적 분석기법 도구를 사용해 Predictive 또는 Descriptive 분석 기능들의 일련 과정을 일괄처리 함으로서 수집된 이동객체데이터에서 정해진 Pattern을 지속적으로 분석해 내는 시스템임

○ 분석 기능들은 R을 이용한 통계기반 이동객체데이터 분석과 Machine Learning 등을 이용한 비통계기반 이동객체데이터 분석으로 이루어져 있음

○ 개발된 분석 컴포넌트들은 Web-platform 기반의 Batch 분석 솔루션으로 구현되었음

○ 연세대학교 산학협력단

- 분석 기능 선정을 위한 통계기반 분석 요구조건 및 비통계기반 이동객체데이터 분석 요구조건을 검토하였음
- 검토한 내용을 기반으로 개발 기능을 결정하고, 통계기반 및 비통계기반 이동객체데이터 분석 컴포넌트를 설계하였음
- 이후, 새로운 이동객체데이터 분석 기능 추가 및 기본 분석 기능의 성능 향상, 기 개발 기능의 고도화 / 최적화를 달성하였음
- 최종적으로, 개발된 이동객체데이터 분석 컴포넌트들을 기반으로 공공 / 민간 수요기반 이동객체데이터 Batch 분석 시나리오 및 솔루션을 설계하였음

○ 가이온

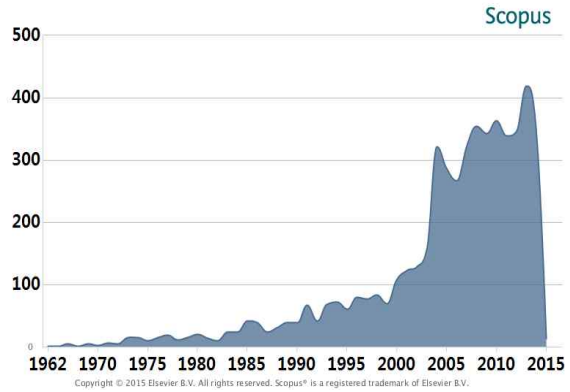
- 이동객체데이터 분석 컴포넌트의 사용자 인터페이스 및 아키텍처를 설계하였고, 이를 기반으로 분석 컴포넌트 프로토타입을 구현하였음
- 이동객체데이터 분석 컴포넌트를 추가하고 연산자 동작 정확도 및 처리 속도를 향상시켰음
- 이동객체 데이터 분석 컴포넌트의 기능을 고도화하였으며, 분산처리 속도 향상을 위하여 Spark를 통해 분석 기능을 구현하였음
- 최종적으로, 이동객체 데이터 Batch 분석 솔루션을 구현하였으며 공인인증 및 소프트웨어 등록을 통해 구현된 솔루션의 성능을 확인하였음

① [단위성과 G-1] 이동객체데이터 통계기반 요구조건 분석 - R의 공간분석 기능 검토

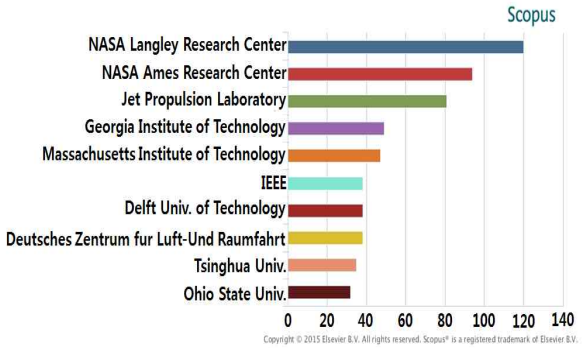
○ 이동객체 데이터에 대한 분석의 글로벌 연구동향을 파악하고자 하였으며, 이

를 위해 SCOPUS DB에서 “Trajectory Data Analysis” 라는 키워드로 검색된 전체 문헌의 초록을 텍스트마이닝과 LDA를 사용하여 분석함

❍ 그 결과, 최근 이동객체 궤적 분석 연구에 대한 관심이 급증하고 있으며, 대학 및 연구기관에서의 연구가 주로 이뤄지고 있는 것으로 나타남

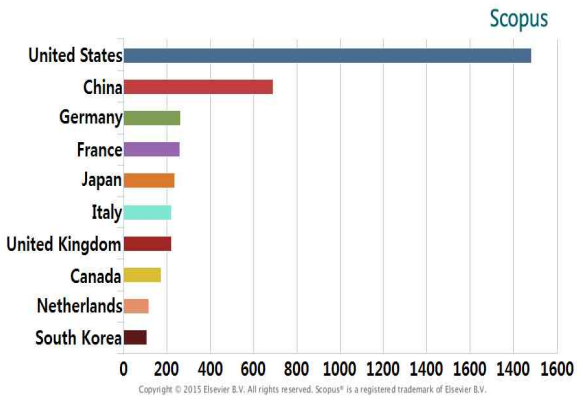


<그림 2-294> 연간 이동객체데이터 관련 문헌 빈도

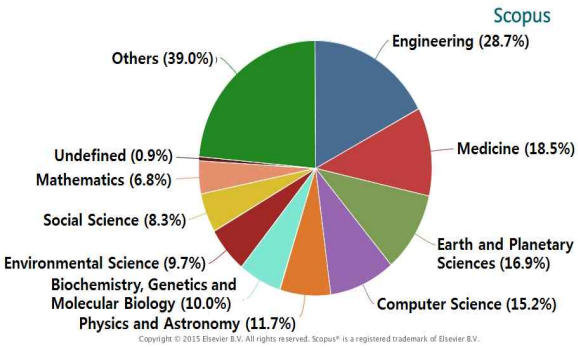


<그림 2-295> 이동객체 데이터 관련 문헌의 저자가 소속된 상위 10개 기관

❍ 국가 중에서는 미국, 중국, 독일, 프랑스 등이 주도하고 있는 것으로 나타났으며, Engineering, Earth and Planetary Sciences, Computer Science 분야 등 다양한 분야에서 이동객체 궤적 분석에 대한 접근이 이뤄짐을 확인함



<그림 2-296> 국가별 이동객체 데이터 문헌 빈도



<그림 2-297> 이동객체 데이터 연구 분야

❍ 이동객체 데이터 통계기반 분석으로 서울시 심야버스 노선 최적화 사례의 경우 KT 통화량 빅데이터를 이용한 통계 분석을 통해 최적화 노선을 선정한 것으로 보고되었으며, dynamic population mapping을 위해 회귀분석이 이용되었던 사례가 있음

❍ 대표적인 오픈 소스 통계 패키지인 R의 이동객체데이터 분석 패키지를 조사한 바, 이동객체데이터 분석에 적합한 패키지는 다음과 같음

<표 2-86> R 기반 이동객체데이터 분석기능 검토

	분석방법	R 패키지
이동 객체 데이터	Statistical Analysis	adehabitatHR
		adehabitatHS
		adehabitatLT
		InfDim
		OLScurve
		SGP
		trajectories
		traj
		move

② [단위성과 G-2] 이동객체데이터 비통계기반 요구조건 분석 - Machine Learning 검토

❍ 총 24편의 이동객체 궤적(trajectory) 데이터 분석 관련 논문을 비교 분석하여 현재까지 진행되어온 이동객체 궤적 데이터 연구를 분석함

❍ 이동객체 궤적 데이터 마이닝 기법을 크게 세 카테고리로 구분해, 연구에 사용한 데이터, 알고리즘, 그리고 기타 특징들을 조사하여 비교 분석하였음

❍ 이동객체 궤적 데이터를 분석하는 세 가지 방법은, 클러스터링(clustering)을 기반으로 하는 방법, 데이터 분류(classification)를 기반으로 하는 방법, 마지막으로 각 데이터가 가지는 관계를 기반으로 분석하는 방법이 있음

<표 2-87> 이동객체 데이터 비통계 연구 검토

분석 기법	연구제목	데이터	연구목적	알고리즘	특징
Clustering	• Unveiling the complexity of human mobility by querying and mining massive trajectory data	• GPS data • Mobile phone	• Predict traffic jams and congestion	• M-altas, the querying and mining	• Powerful basis for understanding mobiliry pattern getting semantic meaning
	• Mining long, sharable patterns in trajectories of moving objects	• GPS data	• Traffic condition 파악	• Naive approach to LSP mining • projection based LSP mining • SQL based implementation	• Present method for efficiently extracting long, sharable frequent routs • (x,y,t)형태의 trajectory data를 region based &road network based framework로 변환
	• A graph-based approach to vehicle trajectory analysis	• GPS data	• Facilitating the understanding of spatial and temporal patterns in trajectories	• converts trajectory data to a graph -based representation spatially constrained graph partitioning method	• proposes a graph-based approach that converts trajectory data to a graph-based representation and treats them as a complex network • Can search and visualize trajectory clusters that existing methods cannot find
	• A personal route prediction system based on trajectory data mining	• GPS data	• 과거의 정보를 통해 한 개인의 route를 예측 • 광고, 교통량 정보 반영에 활용 가능	• A novel mining • CRPM algorithm	• The proposed system has many potential applications. For example, appropriate notifications to users of their predicted routes, and advertisers can send targeted sales messages to shoppers.
	• Trajectory Clustering: A Partition-and-Group Framework	• GPS data • Radio -telemetry location	• trajectory database에서 Sub-trajectory를 파악 • 사용자들에게 새로운 insight를 제공	• minimum description length (MDL) principle • density-based line-segment clustering algorithm • TRA-CLUS algorithm	• 기존에는 전체 trajectory 들 중 clustering을 통하여 common trajectory들을 발견하는 것이 목적이었기에 common sub- trajectory를 놓치기 쉬움 • 하지만 common sub-trajectory도 유용한 자료가 될 수 있기에 이를 파악하는 프레임워크를 개발
	• An Incremental DPMM-Based Met	• 비디오 영상물	• Surveillance by in	• PCA-Based 자체 Algorithm, String M	• DPMM(Dirichlet Process Mixture Model)을 활용한 모델을 사용

분석 기법	연구제목	데이터	연구목적	알고리즘	특징
	hod for Trajectory Clustering, Modeling and Retrieval	내의 움직이는 물체	dexing of motion events	atch, Polynomial Fitting Algorithm, Euclidean Distance-based Algorithm	• DFT(Discrete Fourier Transform) Coefficient를 활용하면 다차원 Trajectory에 대하여 적용 가능
	• Clustering and Visualization of bankruptcy trajectory using self-organizing map	• 기업의 재무·재정 상태	• 기업의 재무·재정 상태를 2D Space로 옮겨 시각화하여 Trajectory 이용 재정 상태 분석 및 예측	• 프랑스 재무 비율 분석 기준 Self-Organizing map Algorithm	• 재무 상태를 2D Space로 옮겨 지도로 만든 뒤 Trajectory를 이용하여 기업의 재정 상태 분석 및 예측
	• 도로 네트워크에서 이동 객체 궤적 분석을 통한 도로 혼잡 구간 판별 기법	• 도로 정보, 궤적 데이터 • 시간당 객체 이동 정보 • (simulation data)	• 교통 혼잡 구역을 클러스터링 하여 방향별 혼잡 구간을 파악	• DBSCAN (Density-based spatial clustering of applications with noise) 클러스터링 알고리즘	• 도로 내 차량 및 도로 환경을 고려하여 도로 혼잡 구간을 판별하는 기법을 사용 • 도로 길이와 폭 기반의 도로 혼잡도를 사용하여 혼잡 구역을 발견하였다. 방향 별 클러스터링 수행
Classification	• TraClass: Trajectory Classification Using Hierarchical Region-Based and Trajectory-Based Clustering	• GPS data • satellite image	• vessel classification for fishery, pollution, border, off-shore activities, search and rescue control	• Region-Based Clustering • Trajectory-Based Clustering	• High classification accuracy owing to the collaboration between the two types of clustering
	• Mining Discriminative Patterns for Classifying Trajectories on Road Networks	• GPS data • trajectories patterns	• Traffic planning	• Frequent pattern-based classification	• Partial sequential pattern을 활용한 Classification이 Closed sequential pattern을 활용한 경우보다 시간,정확도, 손실에서 우위
	• Trajectory outlier detrctio	• radio-telemetr	• Trajectory outlier	• Partition-and-detect framewo	• temporal information 또한 고려한 연구가 필요.

분석 기법	연구제목	데이터	연구목적	알고리즘	특징
	n: a partition and detect framework	y data • satellite data	detection	rk • TRAOD algorithm	• 기존에 부족했던 outlier detection for trajectory data 에 대한 선행 연구로서의 의의
	• Object Trajectory-Based Activity Classification and Recognition Using Hidden Markov Models	• 2D projections of the object's centroid location at each instant of time	• Security	• Gaussian mixture models (GMM) • hidden Markov model (HMM)	• Object motion-based analysis and recognition • A novel framework for motion trajectory-based statistical modeling and classification • temporal modeling in the formulation, robust time-invariant modeling of the PDF
	• Goal-based trajectory analysis for unusual behaviour detection in intelligent surveillance	• Car part dataset	• Surveillance	• Particle filtering algorithm	• unusual behavior detection higher-level features of intentionality를 이용하여 unusual behavior를 detection 함
Clustering& Classification	• TripVista: Triple Perspective Visual Trajectory Analytics and Its Application on Microscopic Traffic Data at a Road Intersection	• Laser scanning data	• traffic pattern 파악	• Triple Perspective Visual Trajectory Analytics (TripVista)	• exploring and analyzing complex traffic trajectory data • visualizing microscopic traffic trajectory data at a road intersection • Investigating trajectories from different angles, including spatial, temporal and multi-dimensional perspectives.
	• Policies based on trajectory libraries	• simulation data	• 장애물·미로와 같은 상황에서 Trajectory 활용 보다 나은 경로 찾기	• A* Algorithm	• 보다 복잡한 미로 혹은 실세계에서의 한계 존재
	• Biped walking control using a trajectory library	• Simulation data	• Biped 이동하는 조건 하에, Trajectory 활용하여 최적의 경로 찾기	• Linear Inverted Pendulum Model Differential Dynamic Programming	• 경로 찾기에 있어 Biped 이동이 전제되어 발과 지면에 의한 제약이 존재

분석 기법	연구제목	데이터	연구목적	알고리즘	특징
Association	• A system for destination and future route prediction based on trajectory mining	• Mobile GPS data	• 예상 목적지 및 예상 경로 결정을통한 광고 메시지 전달	• forward-Backward Matching(Clustering based) • Continuous Route Pattern Mining	• FBM알고리즘을 통해 Time-based Mining의 불연속적 시간 관측으로부터 발생하는 문제 최소화 • 사전 지식이 없는 환경에서도 예측 가능
	• PutMode : prediction of uncertain trajectories in moving objects databases	• Street identifier, speed, direction • Moving objects databases	• criminal tracking analysis	• Continuous Time Bayesian Networks(CTBNs) • PutMode (Prediction of uncertain trajectories in Moving objects databases • TCTBNs(trajectory CTBNs)	• predicting the motion behaviors of moving objects in order to obtain the possible trajectories based on TCTBNs • he trajectory inside an electronic map has numerous distribution rules due to several factors, e.g., city planning, traffic planning, population distribution, etc.
	• Trajectory Similarity Analysis in Movement Parameter Space	• 운동하는 물체의 다양한 Parameters	• Hurricane movement pattern analysis	• n차원에서의 Euclidean Distance	• 비슷한 운동 특성을 가진 물체의 Trajectory 간 유사성 및 상대성 분석 • 서로 다른 Parameter로 이루어진 다차원 공간에서의 움직이는 물체의 Trajectory간 유사성 추출
etc.	• Vehicle Trajectory Optimization for application in Eco-Driving	-	• To reduce fuel consumption of road vehicles by trajectory optimization	• Eco-driving algorithm	• Given a drivecycle a so called ’eco-drive cycle’ is identified in which a vehicleperforms the same distance with the same stops in equivalent • time, while consuming less fuel.
	• Novel Approaches to the indexing of moving obejct trajectories	-	• trajectory-based query 시스템 개발	• TB-tree, queries	• This paper presents a set of pure • spatiotemporal queries, the so called trajectory-based(topological and navigational) queries, as well as combined queries
	• Mining frequent trajectory patterns in spatial-temporal databases	• a spatial-temporal database	• mining thefrequent trajectory patterns in a spatial-temporal data	• GBM(graph-based mining) algorithm	• It utilizes the adjacency property to reduce the search space. • it outperforms A priori-based and PrefixSpan-based methods by more than one order of magnitude.

분석 기법	연구제목	데이터	연구목적	알고리즘	특징
			base		
	• Modeling, storing and mining moving object database	• GPS data • MOD	• traffic condition, optimal routes and prediction of traffic jam	• Map matching algorithm	• Realizing and organizing a traffic management system

○ 관련 연구 사례를 조사한 결과, clustering과 classification 방식이 16건으로 주를 이루었고, 연관 관계를 기반의 분석이 3건, 기타 5건이 조사됨

○ 연구 목적은 교통량 데이터 및 GPS 데이터를 활용한 교통 패턴 파악 및 위치 예측이 총 7건으로 가장 많았으며, MOD(Moving Object Database)를 활용한 범죄자 위치 분석 등의 방범 관련 연구가 4건, 광고 활용 2건 순으로 나타남

③ [단위성과 G-3] 이동객체데이터 분석 컴포넌트 인터페이스 설계

○ 이동객체데이터 분석 컴포넌트 인터페이스 설계를 위한 구현 환경은 다음과 같음

<표 2-88> 구현환경

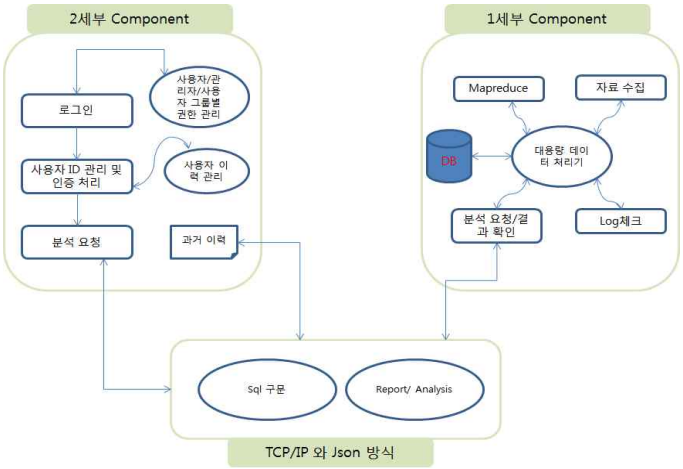
구현 환경	Hadoop 1.2.1	MongoDB 2.6.5
	Hbase 0.98.3	Hive 0.13.1
	Flume 1.5.0.1	R studio Server 0.98.1074
	JAVA 1.6.0_45	

○ 사례 분석을 통한 이동객체데이터 분석 컴포넌트 및 각 컴포넌트들을 관리해야 할 컴포넌트의 필요성 확인

○ 또한, 1세부 및 3세부 컴포넌트와의 인터페이스 관계를 정의하여 통합에 대한 고려를 수행함

④ [단위성과 G-4] 이동객체데이터 분석 컴포넌트 아키텍처 설계

○ 공간 빅데이터 분석 컴포넌트 아키텍처 설계를 위하여 필요한 기능을 List-up하고, User case 다이어그램을 통해 분석 컴포넌트의 workflow를 정의함

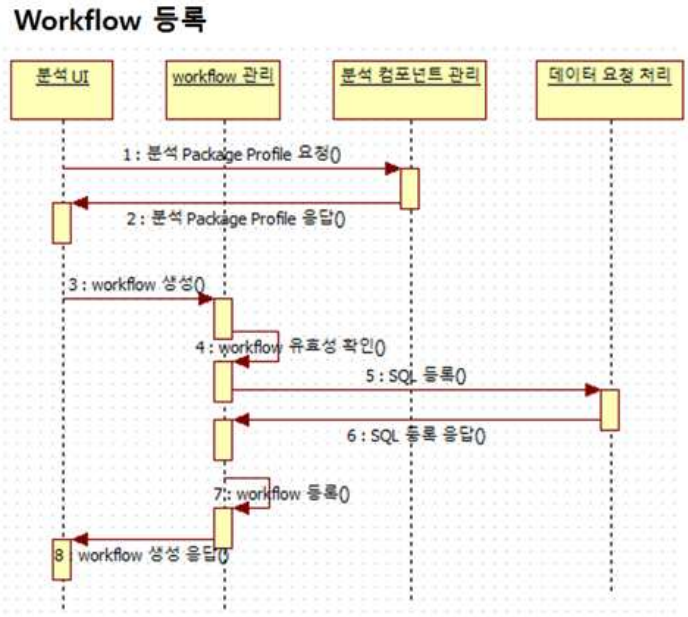


<그림 2-298> 1, 2세부 간 컴포넌트 인터페이스 관계 다이어그램



<그림 2-299> 이동객체 Batch 분석 컴포넌트 아키텍처

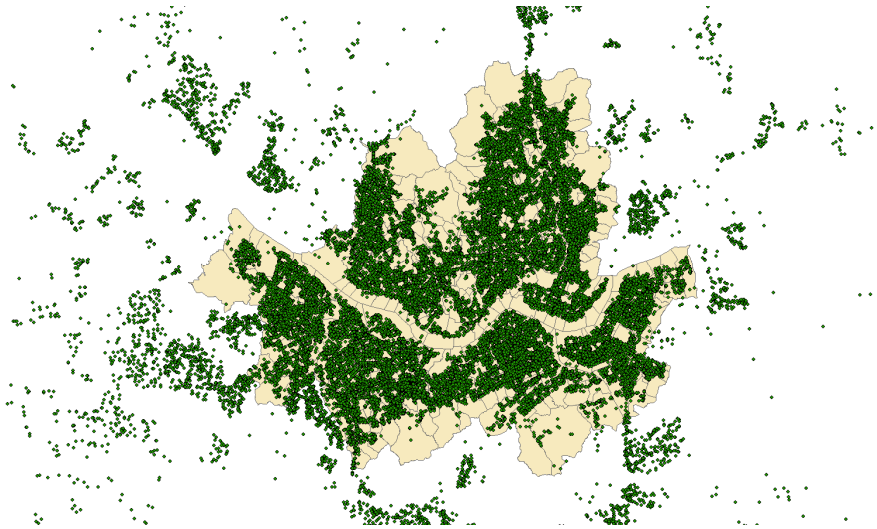
❍ 또한 서비스 시나리오 분석을 통해, 예상되는 분석에 요구되는 컴포넌트들이 어떻게 동작하는지에 대한 시퀀스 다이어그램을 도출하였음



<그림 2-300> Workflow 등록 시퀀스 다이어그램

⑤ **[단위성과 G-5] 이동객체데이터 기본분석 기능 선정 및 개발 기능 결정**

- ❍ 국토교통부에서 서비스 중인 국토공간계획지원체계(KOPSS, Korea Planning Support Systems)의 공간정보 분석기능을 분석하여 이동객체 데이터의 활용이 가능한 세부기능들을 확인함
- ❍ 이동객체데이터 기본분석 기능선정 및 신규개발 기능결정 보고서에서는 1차년도 수요조사 연구 결과를 활용하여 이동객체 데이터 기본분석기능 선정 및 신규개발기능 결정에 기준이 되는 논문을 검색하였으며, 3가지 기준을 통해 논문을 선정함
- ❍ 선정된 논문에 대하여 Input Data, Theory, Procedure, Results를 분석하였으며, 당해 연도의 연구 목표인 기본분석기능 선정 및 신규개발기능 결정에 활용함
- ❍ 이동객체데이터를 활용하였거나 혹은 활용이 가능할 것으로 예상되는 논문을 조사하였으며, 6가지 논문 및 한 가지 사례를 선정하여 분석함. 선정된 논문에 대한 분석결과를 바탕으로, 이동객체데이터 기본분석기능 및 신규개발기능을 정하였으며, 공간군집화 및 야간 대중교통 수요분석, 2가지 응용을 선정함

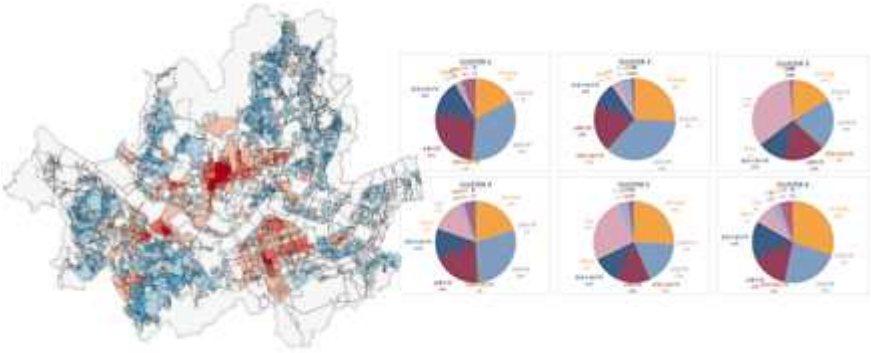


<그림 2-301> 서울시 00시~01시 택시 승·하차 point 데이터

⑥ **[단위성과 G-6] 이동객체데이터 통계기반 분석 컴포넌트 기초 설계**

- ❍ 이동객체데이터 통계기반 분석 컴포넌트로 공간 군집화 기능을 개발함
- ❍ 공간상에서 비슷한 패턴을 가지고 있는 지역을 하나의 클러스터로 묶는 공간군집화를 구현함
- ❍ K-means clustering을 활용하여 기능이 개발되었으며, 구현된 공간군집화의

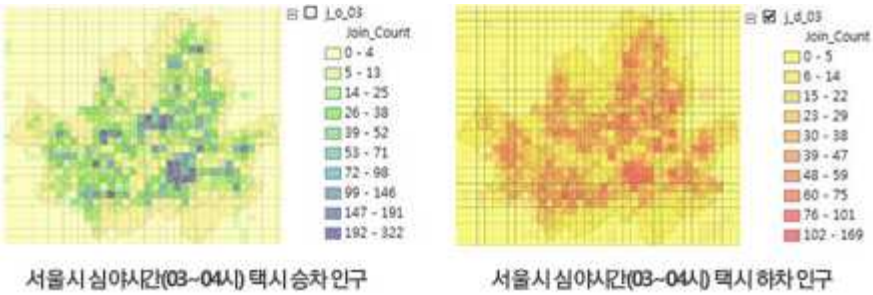
응용에는 스마트카드 데이터와 버스정류장 데이터가 Input data로 활용되어, 그 결과, 서울시 내부의 유사한 지역이 군집화되어 표시됨



<그림 2-302> 서울시 집계구 단위 공간 군집화 결과

⑦ [단위성과 G-7] 이동객체데이터 비통계기반 분석 컴포넌트 기초 설계

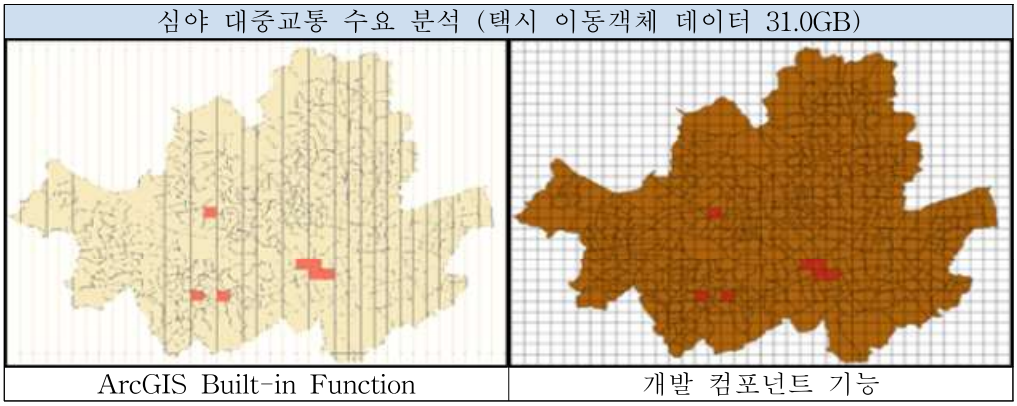
- ❍ 이동객체데이터 통계기반 분석 컴포넌트로 심야 대중교통 수요분석 기능을 개발함
- ❍ 심야 시간에 운행되는 택시의 승차 및 하차 정보를 통하여 심야 시간대의 인구 이동경로를 파악하고, 택시 이용인구가 밀집되어있는 지역이 심야 대중교통 추가 수요지로 선정됨



<그림 2-303> 이동객체데이터 비통계기반 분석 컴포넌트 분석결과

⑧ [단위성과 G-8] 이동객체데이터 분석 컴포넌트 프로토타이핑

- ❍ 이동객체데이터 분석 컴포넌트 프로토타이핑을 수행하고, 연산자 동작 정확도를 측정함
- ❍ 택시 이동객체 데이터를 이용한 심야 대중교통 수요분석을 각각 상용 소프트웨어인 ArcGIS의 Toolbox와 분석 컴포넌트 이용하여 얻은 결과, 연산자 동작 정확도 100%로 두 결과가 일치하는 모습을 보임

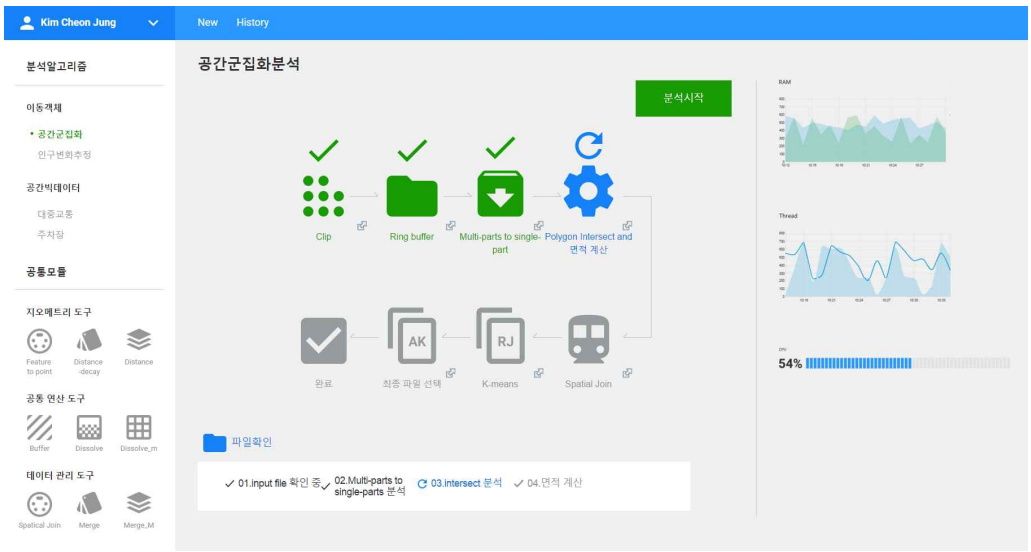


<그림 2-304> 이동객체데이터 분석 컴포넌트 연산자 동작 정확도 실험 결과

❍ 데이터 질의 / 응답 - 분석 - 결과보고의 과정을 통해 사용자에게 결과를 보여주는 프로토타입을 제작함



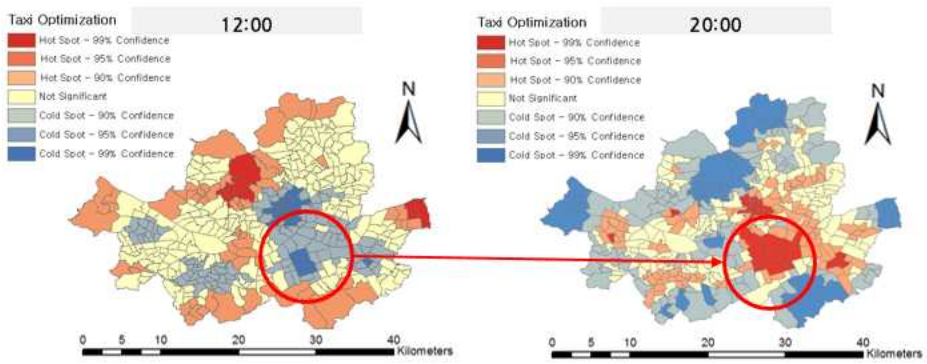
<그림 2-305> 공간 군집화 분석 시작 UI



<그림 2-306> 공간 군집화 분석 진행 UI

⑨ [단위성과 G-9] 이동객체데이터 기본 분석 기능 개선

- ❍ 기존 분석기능의 분석 개선 작업으로 Getis-Ord Gi* 방법론과 유동인구 데이터를 통하여 2차년도 분석 기능인 심야 대중교통 수요 분석의 분석 결과를 개선함
- ❍ 시간대별 택시 공차지역과 유동인구를 통하여 시간대별 택시 서비스의 요구지역 도출 시 수요와 공급을 모두 고려하였음

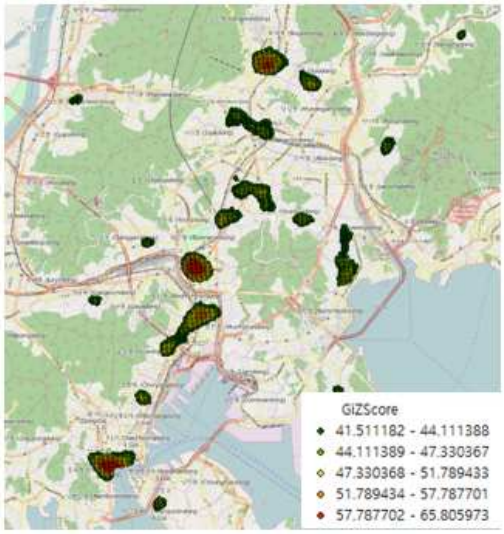


<그림 2-307> 택시 서비스 필요지역 분석 결과

- ❍ 시간대별/요일별 서울시 택시 서비스 필요도를 표현한 지도는, 택시 서비스의 최적화를 목표로 실제 택시 서비스 수요자 및 공급자가 택시 서비스를 효율적으로 이용하도록 의사결정 기반자료로 사용될 수 있음

⑩ [단위성과 G-10] 통계기반 분석 성능 향상

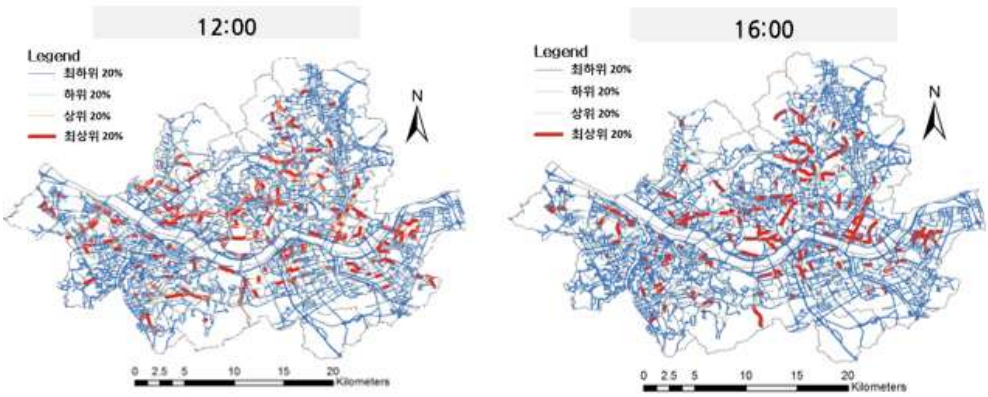
- ❍ 이동객체 데이터 일괄 분석 플랫폼의 통계기반 분석 성능을 향상시키기 위하여 통계 기반의 신규 컴포넌트 기능을 개발함
- ❍ 노인 복지시설 입지선정 기능은 노인 유동인구와 거주인구를 통하여 실제 노인들이 거주하고, 활동하고, 모이는 지역에 기반해 입지를 선정하는 방식을 제시함
- ❍ 노인 유동인구 및 노인 거주인구 자료를 이용해 통합 노인인구를 추정하고, Getis-Ord Gi*를 이용해 노인 잠재수요의 최대 군집 지역을 도출함



<그림 2-308> 부산시 노인 복지시설
입지선정 분석 결과

⑪ [단위성과 G-11] 비통계기반 분석 성능 향상

- ❍ 이동객체 데이터 일괄 분석 플랫폼의 비통계 기반 분석 성능을 향상시키기 위하여 비통계 기반의 신규 컴포넌트 기능을 개발함
- ❍ 대표적인 이동객체 데이터인 택시의 승하차 위치정보를 이용하여 택시의 공차지역을 분석하는 기능을 개발함
- ❍ 시간대별 택시 공차분포를 분석하여 시간대 별 택시 공차 분포를 확인하고, 택시 서비스 공급 불균형 지역을 파악할 수 있음

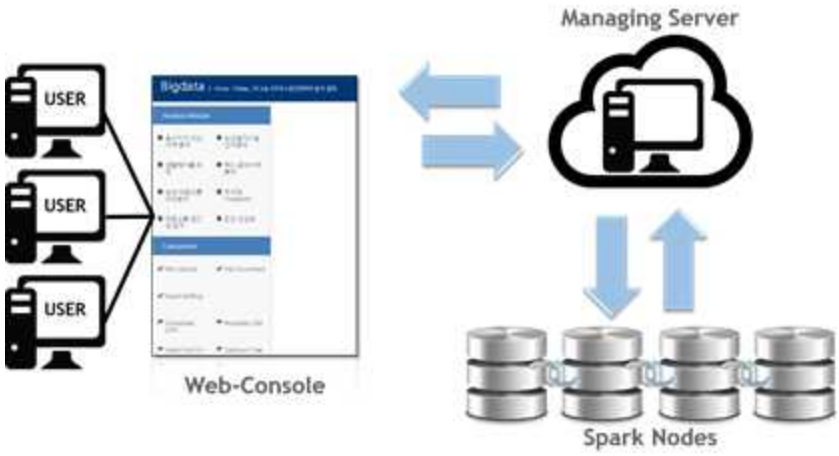


<그림 2-309> 택시 공차 지역 분석 결과

⑫ [단위성과 G-12] 공간 빅데이터 분석 컴포넌트 기능 고도화

- ❍ Apache Spark를 통한 In-Memory 방식의 분석과 분산처리 환경을 구성하였으며, AWS (Amazon Web Services)의 가상 인스턴스를 활용하여 서버를 구성하였음. 또한 scala 언어를 통한 분석 scripting 작업으로 분석 모듈 실행환경을

구성하여 재사용성을 강화하였음



<그림 2-310> AWS와 Spark를 활용한 이동객체데이터 분석 시스템

❍ 또한 웹 콘솔의 UI & UX 개선으로 직관적인 인터페이스를 조성함. 개발 기능 분석을 수행할 수 있는 Analysis Module 탭과, 이동객체데이터 분석 컴포넌트를 활용할 수 있는 Component 탭으로 구분하였으며, 3차년도 및 4차년도 개발 기능들을 모두 모듈 탭과 컴포넌트 탭에 탑재하였음



<그림 2-311> 웹 콘솔의 Analysis Module 탭과 Component 탭

❍ 분석기능의 모듈화를 통해 컴포넌트의 재사용이 가능하도록 spark의 scala로 구현하고 성능을 증빙하기 위해 샘플 데이터들을 활용하여 최종 데이터를 산출함

○ 구현된 기능의 성능을 증빙하기 위해 공인인증기관(TTA)의 공인인증을 받음



시험 결과서

한국정보통신기술협회
소프트웨어시험인증연구소
주소: 서울시 마포구 월드컵북로 396
전화: 02-2132-4040, Fax: 02-2132-4099

결과서번호: BT-B-16-0312

TTA

1. 의 뢰 자
. 회사(기관)명 : ㈜케이위컴퍼니
. 주 소 : (04326) 서울특별시 용산구 소월로 2 나길 33 (후암동)
. 계약일자 : 2016 년 12 월 23 일
2. 제 품 명 : 공간 빅데이터 일괄(Batch) 분석 컴포넌트 v0.4
3. 시험기간 : 2017. 1. 31. ~ 2017. 2. 3.

확 인	작성자	승인자
	성 명 : 이 공 선 (서명)	직 위 : 삼양소프트웨어시험센터장 성 명 : 김 경 훈 (서명)

2017. 2. 3.

한국정보통신기술협회 회장

TTA-SE1-16-A(R)

페이지 (1) / (총 23)

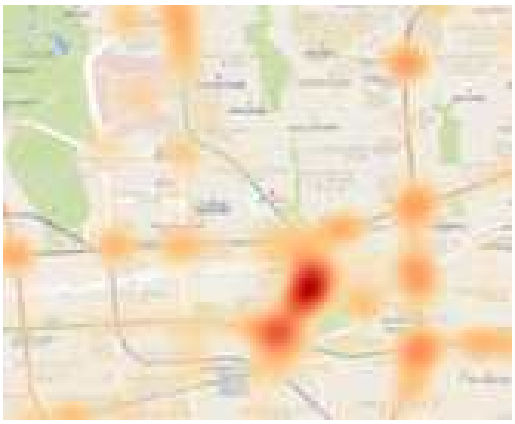
<그림 2-312> 공간 빅데이터 및 이동객체
데이터 일괄 분석 기능 TTA 인증

⑬ [단위성과 G-13] 이동객체데이터 기본 분석 기능 추가

○ 이동객체데이터 기본 분석 기능을 추가 및 개선함

○ 택시 승하차 시공간 클러스터 탐색 기능은 Space-Time Scan Statistics를 이용해 시간대 별 택시 승/ 하차 Hot & Cold 지역을 탐색할 수 있도록 개발됨. 시공간 탐색 알고리즘은 이동객체 데이터 뿐 아니라 다른 빅데이터 분석에도 활용 가능함

○ 그 결과, 특정 시간대의 특정 지역의 택시 수요 공급 불균형을 판단 할 수 있고, 효과적인 택시 운행노선을 선정할 수 있음



<그림 2-313> Space Time Scan Statistics
기반의 시공간 택시 수요 탐지 결과

○ 맵매칭을 통한 택시 운행 탐지 및 최적화 기능에서는 공간 택시 운행 최적화

의 효율적인 분석을 위하여 Map-Matching 기법을 도입함

❍ 이동객체 데이터인 운행 택시의 이동경로 데이터를 Network 기반의 도로망과 연결하여 네트워크 분석이 가능하도록 만들어줌

❍ 실제 Signal 오차가 존재하는 택시 GPS데이터의 경우 택시 위치 좌표가 도로상에 존재하지 않는 오류가 많아 Network 분석시 분석 결과가 실황과 다를 가능성이 크므로, GPS의 오차를 최소화 하기 위한 map-matching process를 수행하고 이를 기능으로 개발함



<그림 2-314> Map Matching 결과와 이에 따른 Network 단위의 택시 유동분석

⑭ [단위성과 G-14] 통계기반 분석 성능 최적화

❍ 이동객체 데이터 분석 기능 성능 최적화를 위해 Spark의 RDD(Resilient Distributed DataSet) 데이터 구조를 활용하여 분산처리가 가능한 개발을 진행하였으며 연차별 연구 목표에 따라 신규 기능을 개발함

❍ 기 개발된 STSS 컴포넌트에 RDD를 적용하여, MapReduce를 적용하였으며, 분산처리 성능을 검증하기 위해 Amazon EMR 서비스 상에서 실험 데이터를 사용한 처리속도 측정을 통해 결과를 평가함

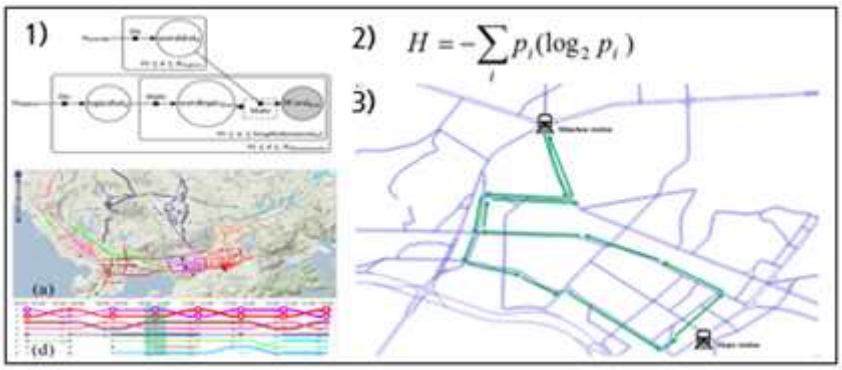
<표 2-89> STSS 컴포넌트 분산처리 속도 측정 결과

사용 노드 수	소요 시간(분)
1	58
2	33
4	19
8	11
16	8.2

❍ 공간 클러스터링 및 환승지점 탐색 기능 개발

❍ Trajectory 데이터에 LDA 모형을 적용하여 이동 토픽들을 탐색하고, 각 도로

별 이동 토픽들에 속할 확률을 이용하여 각 도로들의 entropy를 계산해, 다양한 이동 토픽을 공유하는 도로 (지역)을 탐색함



<그림 2-315> 택시 토픽 기반 LDA 클러스터링

⑮ [단위성과 G-15] 비통계기반 분석 성능 최적화

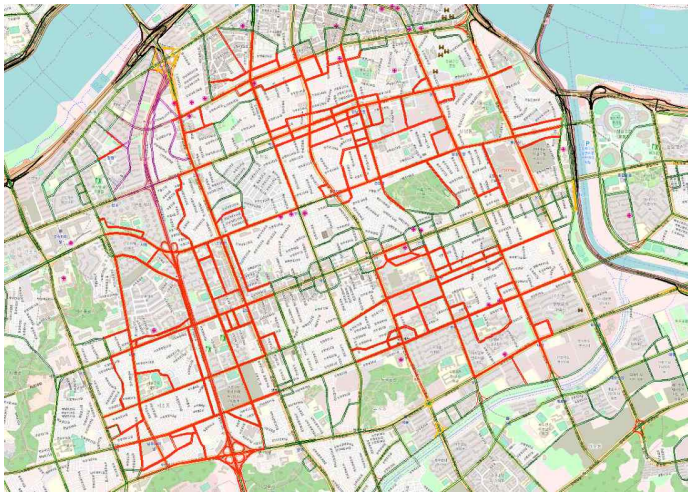
- ❍ 이동객체 데이터 분석 기능 성능 최적화를 위해 Spark의 RDD(Resilient Distributed DataSet) 데이터 구조를 활용하여 분산처리가 가능한 개발을 진행하였으며 연차별 연구 목표에 따라 신규 기능을 개발함
- ❍ 기 개발된 Map Matching 컴포넌트에 RDD 기반 MapReduce를 적용하였으며, 분산처리 성능을 검증하기 위해 Amazon EMR 서비스 상에서 실험 데이터를 사용한 처리속도 측정을 통해 결과를 평가함

<표 2-90> MapMatching 컴포넌트 분산처리 속도 측정 결과

사용 노드 수	소요 시간(분)
1	60
2	30.1
4	17
8	9.8
16	9.1

- ❍ 택시 단거리 운행 패턴 도출 / 데이터 기반 공공시설 서비스가능지역 분석 / 이동객체 데이터를 활용한 운행패턴도출 - 3가지 신규 기능 개발
- ❍ one-one shortest path finding을 통해 단거리 운행 택시 경로를 추출하는 기능을 개발함. 단거리 택시 운행은 대중교통이 취약하거나 이동이 불편한 지역에 집중되므로, 시간대 별 단거리 택시 운행 밀집 지역을 도출하여 해당지

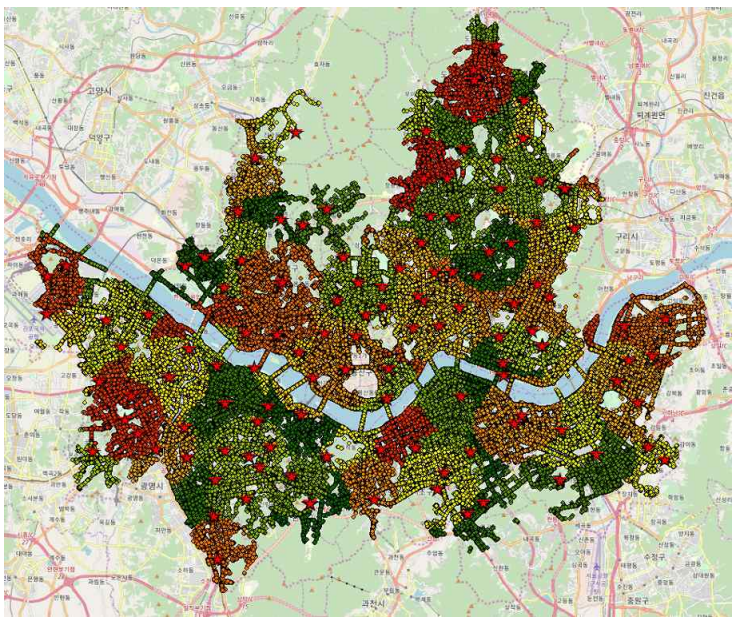
역 내 대중교통 관련 정책 결정에 기여할 수 있음



<그림 2-316> 오전 8-9시 강남지역 단거리 택시 운행 밀집 지역 도출 결과

❍ 기존 Network 기반 Service Area Analysis가 도로망 데이터에 대한 의존도가 크다는 한계를 극복하기 위하여, Grid 기반으로 Service Area Analysis를 수행하여 시설 접근성 분석을 수행하는 기능을 개발함

❍ 분석을 통해 서울 내부에 위치하는 소방서의 5분과 10분 이내 서비스 가능지역을 도출함

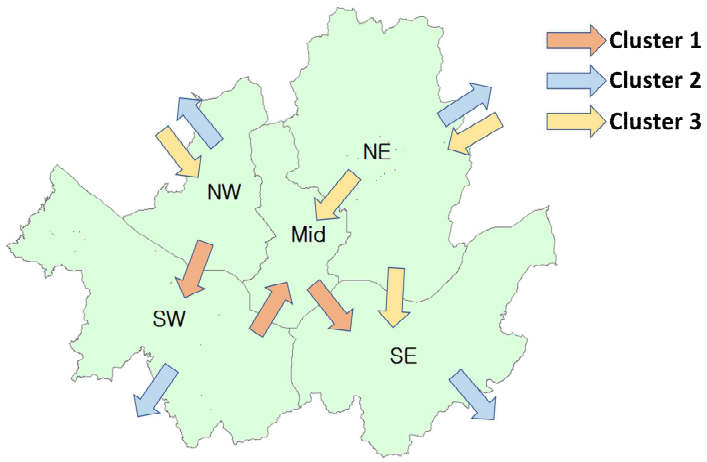


<그림 2-317> 소방시설 서비스 가능지역 Point 결과 도출 (전체시간)

❍ Map Matching을 통하여 DTG 경로를 매핑한 뒤, 택시가 운행한 지역적 Sequential prefix를 도출함. 해당 Sequence prefix의 연계성 / 연속성을 고려

해 Sequential Pattern을 도출할 수 있는 Prefix Span 알고리즘을 이용해 DTG 상의 이동객체 운행 패턴을 연산함

❍ Trajectory 패턴화를 위한 택시 운행지역의 ID 부여하여 서울 지역을 5대 권역(도심, 서북. 서남, 동남, 동북) 및 서울 외의 지역으로 구분하여 지역 ID를 설정하고, 이를 재구성하여 택시 Trajectory를 생성, Prefix-span Algorithm 을 적용하여 자주 나타난 지역이동 패턴을 도출함



<그림 2-318> 09~12시 Prefix-Span 군집화 결과

⑯ [단위성과 G-16] 공공/민간 수요기반 Batch 분석 시나리오 설계

❍ 개발된 기능의 활용에 대한 검증을 위하여, 인천경제자유구역(ifez) 스마트시티 내 개방형 데이터 허브 모델과 관련하여 연세대학교에서 개발한 공간 및 이동객체 빅데이터 분석 기능들을 적용하여 다양한 데이터 활용 방안을 제안함

<표 2-91> IFEZ에서 제공하는 데이터 현황

데이터종류	데이터명	
교통	실시간 교통정보	구간소통정보
		돌발상황정보
		CCTV정보
	대중교통	버스, 지하철역, 항공, 선박 정보
환경	대기오염정보	미세먼지, 초미세먼지, 오존 농도
	도로환경정보	노면상태, 마찰계수 수집값
	기상정보	시간별 기온, 습도, 날씨
	생활환경정보	식중독지수, 체감온도, 발열지수, 불쾌지수, 자외선지수, 동파지수
재난	재난상황, 행동요령, 대피소 위치	
안전	자전거 사고다발지역, 어린이 사고다발지역, 무단횡단 사고다발지역, 스쿨존 어린이 사고다발지역	
생활	병의원, 약국, 자전거, 공원, 문화/행사, 관광/맛집, 무료 WiFi, 주차장, 공공기관, 교육기관, 상가 정보	

❍ 또한, 민간 데이터와의 연계를 통한 데이터 활용 방안을 제안하고, 제안된 데이터 중 활용 가능한 데이터를 통해 실제 기 개발된 분석 기능을 활용하여 분석 결과를 도출 함

❍ 연계 방안으로써, 대기오염, 교통사고, 유동인구, 재난, CCTV, 에너지 데이터에 기반한 스마티시티 데이터 허브 기반 데이터 활용을 제안하였으며, 특히 유동인구에 기반한 생활지도 활용에서 연세대학교의 분석기능들을 활용한 예상 결과를 도출하여 현장 검증을 수행함

❍ IFEZ와의 현장 검증을 통해, 정책연구기관의 분석 연구에 있어, 연세대학교에서 개발한 분석 기능들을 실제로 활용하여 분석 결과를 도출하였으며, 지역의 유동인구 패턴 도출에 활용 가능함을 확인하였음



<그림 2-319> 개발기능을 통해 제안한 유동인구 데이터를 이용한 활동지도

❍ 본 현장 검증을 통하여, 공공기관에서 보유하고 있는 다양한 빅데이터에 기반 한 분석 기능들이 활용 될 수 있음을 확인하였으며, 이러한 빅데이터를 분석할 수 있는 기존 개발기능들 활용하는 분석 시나리오를 설계함

- 첫 번째 시나리오로, “보행자 교통사고 위험지역 및 이상지역 도출 기능”을 통하여, 유동인구 기반 보행자 교통사고가 많이 발생하는 지역을 도출하는 시나리오를 제안
- 두 번째 시나리오로, “노인복지관 최적 입지선정 기능”을 통하여, 지역의 유동인구의 연령별 속성 정보를 통한 공공 시설 확중 서비스 지역 도출을 제안
- 마지막 시나리오로, “딥 러닝 기반 시공간 분석 기능”을 통해, 도시의 대기오염물질 수준을 예측하는 시나리오를 제안

⑰ [단위성과 G-17] 이동객체데이터 Batch 솔루션 설계

- 1차년도부터 6차년도까지 개발한 공간 빅데이터 Batch 분석 솔루션을 이루는 분석 컴포넌트는 총 15개이며, 이를 기반으로 분석 솔루션을 설계함
- 특히, 기존에 사용했던 이동객체데이터인 택시 경로 데이터와 더불어 휴대전화 위치정보 기반으로 구축된 Origin-destination (O/D) 자료를 사용하여 그래프 분석 기능을 적용함

<표 2-92> 15개 이동객체데이터 Batch 분석 컴포넌트

	이동객체데이터 Batch 분석 컴포넌트
1	심야대중교통 수요분석
2	공간 군집화
3	노인복지시설 입지선정
4	택시 공차 분석
5	택시 승하차 시공간 클러스터 탐색
6	맵 매칭 기반 택시 운행 최적화
7	택시 단거리 운행 패턴 분석
8	데이터 기반 공공 서비스 접근성 분석
9	이동 토필 기반 클러스터 탐색
10	화물차량 DTG 기반 운행패턴 분석
11	인구 유동 기반 Community 구역 도출
12	돌아가는 택시 찾기
13	계층 군집분석을 이용한 비정상적인 택시 경로 탐지
14	Graph 기반 지역별 인구 유입-유출 패턴 도출
15	Graph 기반 인구 유동 패턴 도출

- 지속적으로 연구 및 개발을 진행하며, 기 개발된 분석 컴포넌트들에 대한 성능 개선 작업 지속 진행함
- 공간 빅데이터 연산 효율을 높이기 위해 Locality Sensitive Hashing 방식을 적용하였으며 이를 활용해 Getis Ord-Gi*, Score Model, LISA 등의 알고리즘에서 분석 속도 및 분석 효율이 높아짐을 확인함
- 이동객체 데이터 분석 기능 추가 - 돌아가는 택시 찾기
- 택시데이터에 분산처리가 적용된 DTW 알고리즘을 수행하고, 이를 통하여 택시 이동객체 데이터를 Hierarchical Clustering 하여 돌아가는 택시를 도출하는 분석을 진행함



<그림 2-320> 돌아가는 택시 도출 결과

○ 해당 분석기능을 통해 다수의 택시가 활용하는 구간이 아닌 우회 구간을 통해 도착지역으로 이동하는 택시를 도출할 수 있음

○ 결과는 위 그림과 같이, 같은 지역에서 출발하고 도착지역이 동일한 택시 중 다른 택시가 주로 활용하는 경로가 아닌 타 경로 (우회경로)를 사용하는 택시를 도출 할 수 있음

○ 해당 분석의 경우 DTW(Dynamic Time Wariping) 알고리즘을 활용하여 경로의 절대 길이가 서로 다른 이동객체(택시)들을 표준화 시키는 방식으로 진행함

○ 이동객체 데이터 분석 기능 추가 - 계층 군집분석을 이용한 비정상적인 택시 경로 감지

○ DTW 방식과 유사한 NW(Needleman-Wunsch) 알고리즘을 적용하여 이동객체 경로의 이상행동을 판단하는 분석기능도 개발되었음

○ 기존 DTW 기반 방식은 이동경로의 표준화된 '길이' 를 기반으로 이상경로를 도출하였다면, NW 알고리즘의 경우 이동객체가 이동한 Node / Link의 순서 및 배열을 활용하여 경로간의 유사도를 구축하고 이를 통해 이상경로를 도출함

	서울31자5914	서울33바6436	서울33자2618	서울33자6727	서울34사1668	서울34사1890	서울34아9758
서울31자5914	NA	21	12	16	17	21	21
서울33바6436	21	NA	15	11	24	4	24
서울33자2618	12	15	NA	18	15	15	15
서울33자6727	16	11	18	NA	15	11	15
서울34사1668	17	24	15	15	NA	24	12
서울34사1890	21	4	15	11	24	NA	24
서울34아9758	21	24	15	15	12	24	NA

<그림 2-321> NW 알고리즘을 통해 얻어진 유사도 matrix

○ 모든 택시 이동궤적 시퀀스에 대하여 NW 알고리즘을 적용하면, 위 그림과

같은 유사도 matrix를 얻을 수 있으며 유사도 매트릭스는 행과 열의 길이가 이동객체의 수와 같고, 주대각선 성분이 NA인 대칭 행렬임.

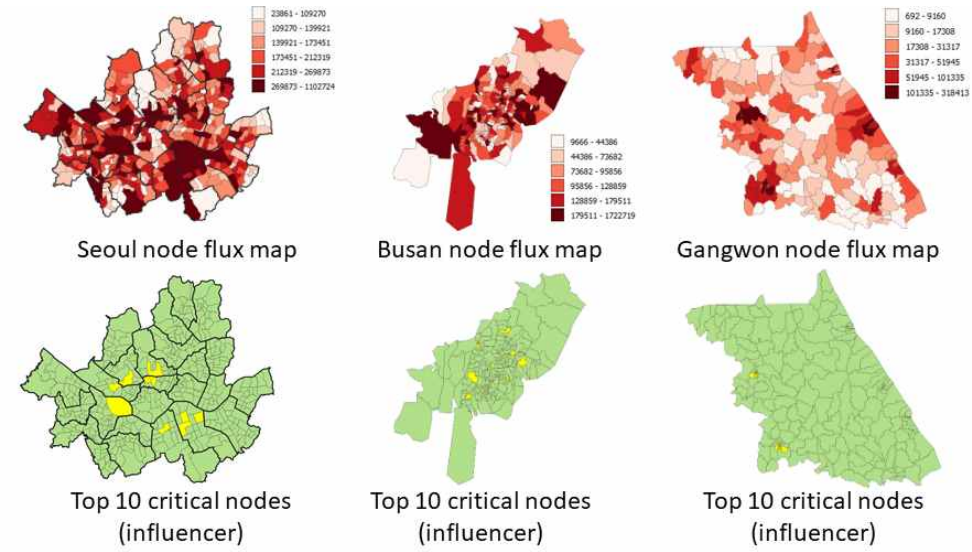
❍ 행렬의 각 성분은 그 성분의 행과 열에 해당하는 이동궤적 간의 유사도를 나타내며, 동일 시작점과 도착점을 지니는 다수의 경로를 NW알고리즘에 적용하면 유사하지 않은 '이상' 경로를 도출할 수 있음



<그림 2-322> 비 정상적인 택시 이동 경로 도출 결과

❍ 이동객체 데이터 분석 기능 추가 - Graph 기반 지역별 인구 유입-유출 패턴 도출

❍ 유동인구 OD 자료는 방향이 있는 가중치 그래프의 일종으로, 특정 지역을 node로 표현할 수 있는 장점이 있음. 이를 이용해 지역과 지역 사이의 인구 유입-유출 패턴을 파악할 수 있는 기능을 개발함

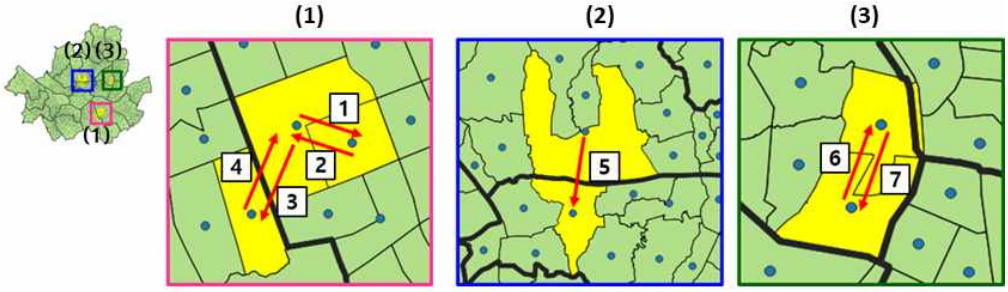


<그림 2-323> OD를 이용한 node 기반 분석 결과 (서울, 부산, 강원도)

❍ 이동객체 데이터 분석 기능 추가 - Graph 기반 인구 유동 패턴 도출

❍ 지역 간 인구 유동을 그래프의 link, 유동 인구수는 link weight으로 보고 지역별 내부 연결성 및 인구 유동의 특성을 파악할 수 있는 기능을 개발함

❍ 특히 행정 구역별 전체의 연결 밀도를 도출하는 Global clustering coefficient를 통해 각 지역을 구분할 수 있는 장점을 제공함



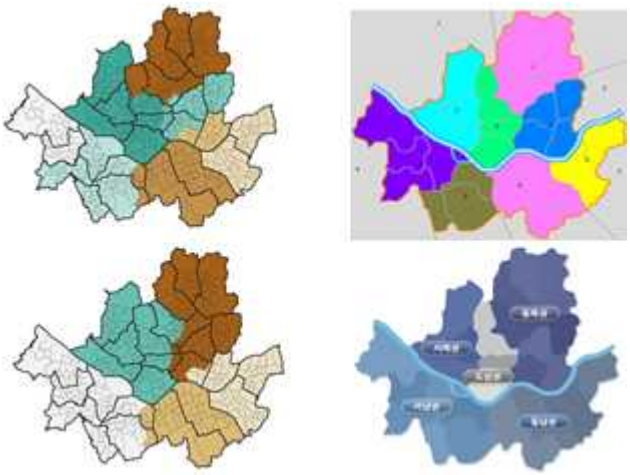
	Origin	Destination	Link weight
1	역삼1동	역삼2동	73,936
2	역삼2동	역삼1동	72,113
3	역삼1동	서초2동	63,239
4	서초2동	역삼1동	61,512
5	명동	종로1,2,3,4가동	58,025
6	장안1동	장안2동	53,558
7	장안2동	장안1동	52,682

<그림 2-324> OD를 이용한 link 기반 분석 (서울의 주요 인구 유동 도출)

❍ 이동객체 데이터 분석 기능 추가 - 인구 유동 기반 커뮤니티 구역 도출

❍ node와 link로 표현된 인구 유동 자료를 기반으로 그래프 군집을 도출하는 Louvain algorithm을 이용하여 유동인구 기반 생활권을 도출하는 기능을 개발함

❍ 도출된 생활권을 기존 행정구역과 비교 분석하여 도시 계획, 공공 시설 입지 선정, 도시 기능 최적화 등 의사 결정에 도움을 줄 수 있음

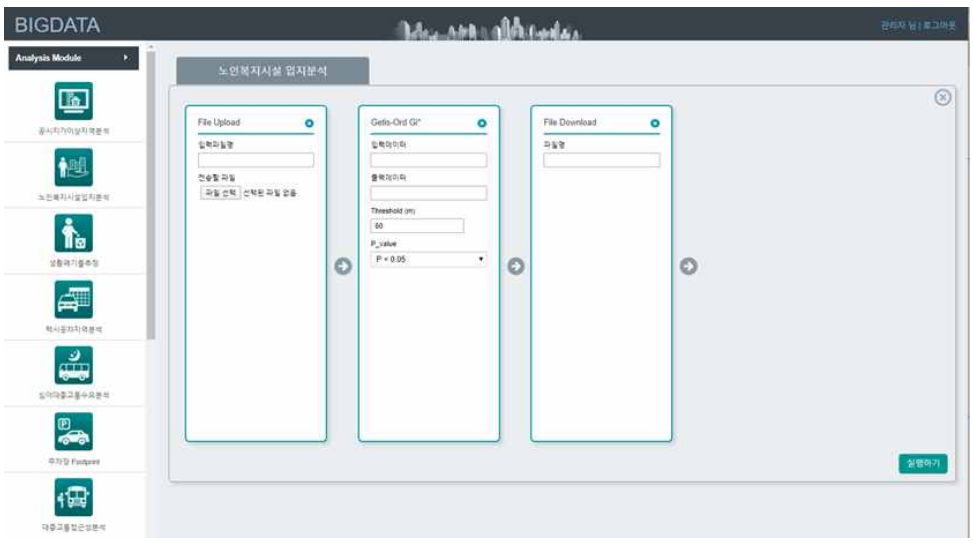


<그림 2-325> 서울 5대 생활권 및 버스 권역과의 결과 비교

⑱ [단위성과 G-18] 이동객체데이터 Batch 솔루션 구현

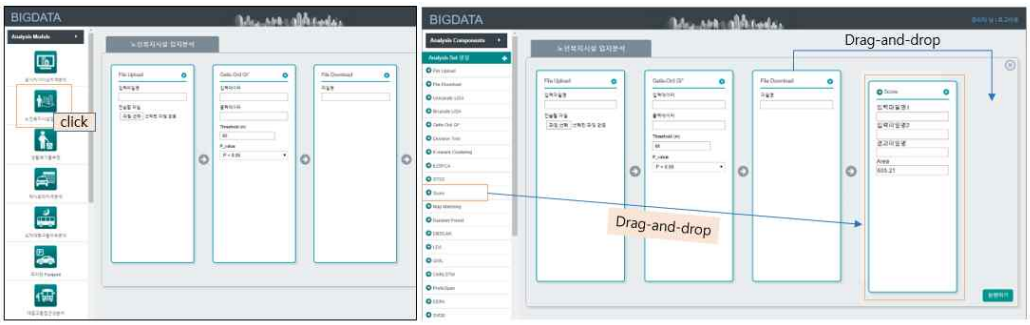
❍ 총 15개 이동객체데이터 분석 기능을 포함하는 Batch 솔루션 구현

❍ UI / UX 측면에서, 각 분석 컴포넌트별 독립성을 보장하여 확장성과 사용성을 높이고, Web-based platform 상에 drap-and-drop 기능을 구현함



<그림 2-326> 공간 빅데이터 분석 Batch 솔루션

❍ 분석 컴포넌트를 조합하여 분석 시나리오를 자유롭게 설정할 수 있으며 저장 및 편집이 가능함



<그림 2-327> Drag-and-drop 및 저장 / 편집 기능

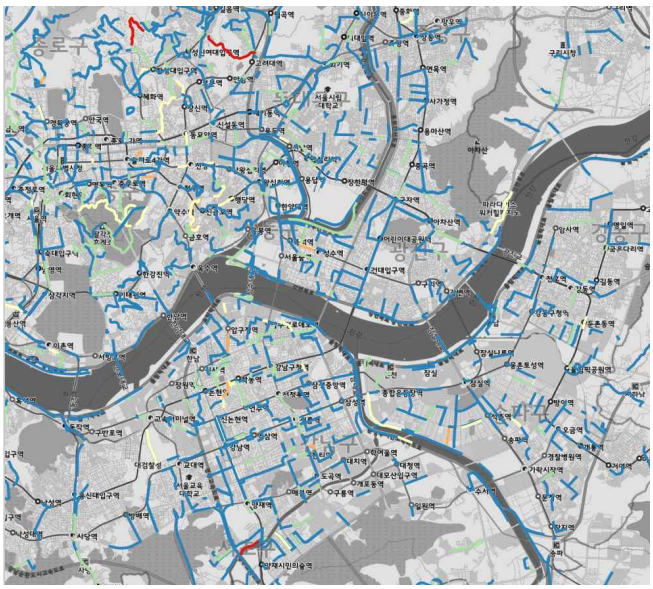
⑲ [단위성과 G-19] 이동객체데이터 분석 솔루션 사업단 실증

❍ 국토공간정보의 빅데이터 관리, 분석 및 서비스 플랫폼 기술개발 연구단의 R&D 성과물이, 공간빅데이터 분석 플랫폼 (사업단 플랫폼)에 적용이 가능한지 실증 테스트를 수행하는 시험을 진행하였음

❍ 1세부 연구성과물과의 통합연계를 보이기 위해서 1-1세세부, 1-3세세부의 개

발 기능의 결과와 2-2세부(연세대학교, 가이온)의 이동객체 데이터 분석 기능 2종 (Map Matching, Percentile Rank)을 공간 빅데이터 분석 플랫폼 내에 설치하여 기능 연동을 수행하고, 그 결과를 확인하였음

- ❍ 시험은 2-2세부가 기 개발한 이동객체 데이터 Batch 분석 솔루션의 'Map Matching' 분석 컴포넌트와 'Percentile Rank' 컴포넌트에 대하여 사업단의 플랫폼 내에서 1세부의 성과물과의 연동성을 실증하였음
- ❍ 역시 분석 모델 수행을 위한 데이터는 1세부의 성과물을 활용하여 가져왔으며, 이를 2-2세부(연세대학교, 가이온)에서 개발한 분석 컴포넌트에 적용함
- ❍ 정상적으로 분석 결과가 파일로 생성되고 화면에 가시화됨을 확인함



<그림 2-328> Map Matching 및 Percentile Rank 분석 컴포넌트 결과

(파) 공간 빅데이터 Interactive 분석 컴포넌트 (핵심 성과 H)

- ❍ 실시간 입력되는 공간 빅데이터를 대화형으로 신속 정확하게 분석할 수 있는 기술로서, 크게 공간 빅데이터 분석 기반 복합 이벤트 처리 기술과 공간 빅데이터 기반 대화형 분석 기술로 구성. 공간 빅데이터 분석 기반 복합 이벤트 처리 기술을 이용하여 고속 고용량으로 입력되는 이벤트 데이터 스트림에 대해서 사용자가 관심을 가지는 이벤트 패턴을 신속하게 탐지하는 작업을 수행할 수 있음. 공간 빅데이터 기반 대화형 분석 기술을 이용하여 대용량의 공간 데이터셋을 다차원으로 나누어 분석할 수 있음. 특히 사용자가 전문 지식이 없는 경우에도 최소한의 정보만으로 쉽고 간편하게 관련 기술을 서버에 배포하여 사용할 수 있는 기능을 함께 제공함



<그림 2-329> 핵심성과 H의 대표 이미지

① [단위성과 H-1][단위성과 H-2] 공간 빅데이터 Interactive 분석 컴포넌트 요구사항 정의서 / 컴포넌트 설계서

❍ 공간 빅데이터 CEP 요구사항 정의

- 6개의 공간 데이터 타입과 7개의 공간 연산자, 8개의 공간 수치 연산자를 정의

❍ 공간 빅데이터 OLAP의 시각화(UI/UX) 설계 요구 사항

- 데이터 차원(공간, 시간, 주제)과 시각화 멤버 Measure, 시각화 차트 제시

❍ 공간 빅데이터 OLAP의 시각화(UI/UX) 설계 구체화

- 지도와 함께 통계값, 테이블, 다이어그램, 시간변화를 여러 개 동시에 표시하고 롤업 및 드릴다운이 가능

❍ 공간 빅데이터 CEP 설계

- DE-9IM(Dimensionally Extended Nine-Intersection Model)을 이용해 공간 위상 연산자를 설계
- Drools Fusion에 지금까지 정의한 공간 데이터 타입, 위상 연산자, 수치 연산자를 구현하고, 공간 CEP 시스템의 아키텍처의 설계도와 UML 다이어그램을 제작

❍ 공간 빅데이터 OLAP 설계

- OLAP 엔진 시스템 아키텍처를 보이고 각 구성요소를 설명

② [단위성과 H-3] CQL 질의 메모리 상주 색인 및 저장 모듈

❍ 복합 이벤트 처리

- 복합 이벤트 처리를 소개하고 이를 수행할 시스템 아키텍처와 구성요소를 설명
- 복합 이벤트를 Rule 기반으로 처리, 저장, 색인하는 Rete 알고리즘을 설명

- Rete 알고리즘을 이용하여 CEP 엔진 역할을 수행하는 소프트웨어인 Drools 설명
- 실시간으로 스트림 데이터를 분산해 처리하는 Apache Storm 및 Kafka 설명

❍ 공간 빅데이터 기반 복합 이벤트 입출력 설계 및 개발

- 공간정보의 입력으로 GeoJSON 포맷을 설명하고 주어진 GeoJSON 스키마에 따라 자동으로 일반화된 객체를 만들어주는 객체 일반화 코드를 구현
- 공간데이터의 MGRS 좌표로 되어있는 위치정보를 이용한 공간 데이터를 색인 및 그룹화 구현

❍ 공간 빅데이터 기반 복합 이벤트 질의 기능 Proof-Of-Concept 버전 개발

- JTS 라이브러리와 Drools를 사용해 공간 데이터 타입, 공간 연산자를 구현
- 화물차 복합 이벤트 처리 시스템을 개발해 공간 빅데이터 기반 복합 이벤트 처리 시스템의 실제 응용을 보이고 성능을 측정함

③ [단위성과 H-4] 공간 빅데이터 기반 복합 이벤트 스케줄러 모듈

❍ CI를 이용한 자동 배포 시나리오 소개

- Git에서 소스코드의 버전을 관리하고, Jenkins는 이 소스코드를 빌드하고 bashscript를 실행하여 공간 빅데이터 복합 이벤트 처리 시스템을 구동함
- 자동으로 Git에서 소스코드를 받아와 빌드하고 이를 배포하는 소프트웨어인 Jenkins 소개 및 사용법 설명

❍ 배포된 Storm topology의 관리를 위한 Storm Scheduler 소개 및 테스팅

- Storm scheduler의 아키텍처에 포함된 Storm web UI, Storm CLI Tool, REST API를 설명
- 이를 이용해 실행된 topology의 Thread 수와 지연시간을 바꿔가며 테스팅
- 테스트를 위한 아키텍처와 그 결과를 그래프로 나타냄

❍ 복합 이벤트 처리의 최적화를 위한 Rebalancing 프로세스 제안 및 구현

- Rebalancing은 Storm의 기능 중 하나로 Worker의 수와 Executer 수를 실행 도중 변경하여 실시간으로 데이터를 최적화해서 처리하는 기능임
- 여러 Rebalancing 구현 방안을 계획하고 각각의 구현 아키텍처를 설명함

❍ 최종 복합 이벤트 스케줄러 아키텍처 구현

- Kafka를 이용해 데이터를 Storm에 입력시키고, Jenkins를 이용해 소스코드를 빌드하고 배포함
- Storm Scheduler를 이용해 배포된 topology를 모니터링하고, rebalancing 기준에 도달하는지 검사하고, 도달하면 rebalancing을 진행함

④ **[단위성과 H-5] CQL 질의 클러스터 내 최적 할당 모듈**

❍ Rule 할당 시나리오 개요

- 각 rule의 처리 시간을 체크하고 재분배가 필요한 rule을 파악하여 topology에 재분배를 진행하고 다시 가동함
- 구현을 위해 Kafka, Storm Rest API, Bashscript, Ganglia를 사용함

❍ Rule 할당 구현

- Storm UI Rest API의 기능과 사용방법을 설명하고 예시 코드를 작성
- Supervisor에서 storm ui API를 사용해 로그를 조회하고 실행 시간을 마스터 서버인 Nimbus에 전송
- Nimbus에서 다른 Bashscript를 실행해 rule을 재분배한 후, bolt를 재기동함

❍ Rule 할당 테스트

- Storm 로그를 모니터링해 실행 시간이 임계값을 넘으면 rule을 재배치함
- Bashscript를 사용해 데이터를 Storm에 전송하고, 전송된 데이터를 탐지하는 rule에 지연이 발생하면 실제로 rule이 이동하는 것을 로그를 통해 확인함

⑤ **[단위성과 H-6] 공간 빅데이터 기반 질의 처리기 모듈**

❍ 토폴로지 일반화 및 최적화 시나리오 작성

- 사용자가 제공하는 스트림 데이터 스키마와 드룰스 룰 파일을 입력으로 받아 자동으로 최적화된 Storm topology를 제안함
- 토폴로지 최적화를 위해 데이터 기반 병렬화나 작업 기반 병렬화를 사용하여 사용자가 제공하는 스키마와 룰 파일로부터 최적의 토폴로지 구성을 JSON 형태로 출력함

❍ 토폴로지 최적화를 위한 병렬화 개선

- 대표적인 두가지 병렬화 방안인 데이터 기반 병렬화와 작업 기반 병렬화 설명
- 공간 데이터 스트림의 키를 정의하고 동적으로 분배해 데이터 기반으로 병렬화하는 적응적 공간 키 그룹핑 알고리즘을 개발하고 구현하고 실험을 통해 각 서버의 부하 불균형을 줄임
- Rete 네트워크를 분할하는 알고리즘을 개발하고 구현해 공간 데이터 스트림 처리를 작업 기반 병렬화로 최적화하고, 최적화한 토폴로지의 성능을 실험을 통해 검증함

⑥ [단위성과 H-7] 공간 빅데이터 OLAP 데이터 로딩 모듈

❍ SQL to Marmot Plan 변환기

- 공간 빅데이터의 대화형 분석을 위한 공간 빅데이터 OLAP에서 발생한 다차원분석 쿼리인 SQL 문을 받아들여 1세부에서 개발한 Hadoop 기반 NoSQL Database인 Marmot에 연계하기 위한 Plan (Marmot 데이터베이스의 Client API를 통한 질의 방식)을 생성함
- OLAP 서버로 선정된 GeoMondrian 도구에 의하여 사용자가 대화형으로 다양한 국가/시/군/구 단위의 드릴다운, 필터링, 필터링을 통하여 생성되는 SQL의 유형을 도출하여 패턴을 파악한 후, 변환함
- Mondrian에서 발생한 SQL 유형들은 주로 Join Query와 where filter 로 구성되는 복잡한 쿼리의 유형임. 이를 단순한 Marmot Plan 으로 직접 일일이 구현하여 결과 데이터가 동일하게 나타나는 샘플 코드 집합을 구현함
- 직접 구현한 코드들을 제네릭하게 프로그램에서 생성할 수 있도록, SQL to Marmot Plan 변환기를 개발하여 1세부 연구성과물과의 연동을 원활하게 함

❍ Marmot 데이터베이스 JDBC Driver

- 1세부에서 개발한 Marmot의 데이터베이스가 GeoMondrian OLAP 플랫폼에 탑재될 수 있도록 상기의 전환모듈을 JDBC Driver 형식의 (Strategy Pattern) 어댑터로 작용하여 몬드리안에 피버팅 결과의 데이터를 제공함
- OLAP 서버는 일반적인 RDBMS를 데이터소스로 하여 분석 SQL을 생성하여 사용자가 원하는 데이터를 표시하게 되기 때문에, 공간 빅데이터에 최적화된 데이터베이스 질의로 전환 전송할 수 있는 쿼리로의 전환과 드라이버를 개발함
- JDBC를 인터페이스로 사용하는 GeoMondrian을 위하여 Marmot Plan 컨버터를 JDBC Driver 형태로 포장하여 개발함

⑦ [단위성과 H-8] 공간 빅데이터 OLAP 드릴다운/롤업 모듈

❍ 다양한 종류의 기본 공간 연산자 개발

- 일반적인 MDX에 공간 연산 기능이 추가된 Spatial MDX를 설명하고, Spatial MDX에서 사용되는 기본 공간 연산자들을 설명함
- 둘 이상의 공간 데이터 사이의 위상에 대한 연산자들인 Topological Predicates에 대해 설명함
- 공간 데이터의 특성을 수치화하는 값에 대한 연산자들인 Scalar Attributes에 대해 설명함

❍ 시간/공간 차원에 대한 드릴다운/롤업 기능 연구 개발

- 일반적인 OLAP에 대해 정의하고 OLAP의 주요 특징, OLAP과 RDBMS의 비교, OLAP의 데이터 구조 및 OLAP의 주요 연산자들에 대해 설명함

- GeoMondrian의 특징, GeoMondrian의 Cube 스키마 및 하위 구조, GeoMondrian의 공간 데이터 처리과정, GeopMondrian의 소스코드 분석 결과에 대해 설명함
- GeoMondrian의 Cube 스키마를 통해 시간/공간 차원에 대한 드릴다운/롤업 기능 구현 방법에 대해 설명함

❍ Dimensional Manager를 통한 차원 기능 개발

- 공간 빅데이터 OLAP 아키텍처를 제시하고 아키텍처 상의 각 부분에 대해 목적과 흐름을 설명함
- 공간 빅데이터 OLAP의 한 부분인 Dimensional Manager의 주요 기능과 다른 부분들과의 관계에 대해 설명함
- Dimensional Manager에서 사용되는 주요 요소들(member, hierarchy, level, dimension)에 대해 설명함
- Dimension의 간단한 구현 예제와 Dimension을 사용해서 시간 차원을 구현한 예제를 보이고, 각 예제 소스코드에 대해서 분석함

❍ 시각화 방안 제시 및 개선

- 시각화를 위한 기본 플랫폼인 Spatialytics에 대해 설명하고 이를 빌드하고 배포하는 방법에 대해 보다 상세하게 설명함
- 시각화를 보다 개선하기 위해 전반적인 레이아웃, 예제 스타일 추가, 속성값 표시, 예제 MDX 질의 추가 등의 사항을 개선함
- 실제 부동산 실거래가 데이터를 사용해서 공간 빅데이터 OLAP 적재를 위한 스타 스키마 설계, 테이블 설계, 데이터 입력, Cube 스키마 설계, 공간 OLAP 실행 과정을 설명함

❍ SQL Generaor 개발

- 공간 빅데이터 OLAP의 한 부분인 SQL Generator의 주요 기능에 대해 설명하고, 공간 빅데이터 OLAP의 엔진으로 Marmot을 사용하기 위해서는 SQL Generator가 Marmot을 지원해야하는 것을 설명함
- 공간 빅데이터 OLAP에서 현재 지원하고 있는 PostGIS가 사용하는 SQL과 Marmot이 사용하고 있는 SQL을 비교하여 PostGIS SQL에서 ETRI SQL로의 변환 가능성을 제시함
- SQL Generator에서 공간 객체 관련 함수와 SELECT 질의 관련 구현을 수정하여 공간 빅데이터 OLAP에서 Marmot을 지원하도록 수정함

⑧ [단위성과 H-9] 공간 빅데이터 OLAP 대화형 분석기 모듈

❍ 공간 데이터 타입 및 공간 연산자의 밀결합 구현 필요성 제시

- 일반적인 OLAP과 공간 빅데이터 OLAP의 차이를 설명하고 이를 통해 공간 데이터 타입과 공간 연산자들의 효율적인 구현의 필요성을 제시함

- 공간 데이터 타입과 공간 연산자들을 구현할 때 상위 레벨 인터페이스를 사용한 구현 방식인 소결합 기법과 엔진 레벨에서 통합한 구현 방식인 밀결합 기법을 비교함
- 소결합 기법 구현을 밀결합 기법 구현으로 변경하였을 때의 일반적인 장점을 제시함

❍ 공간 데이터 타입 및 공간 연산자들의 밀결합 구현을 위한 분석

- GeoMondrian의 아키텍처 및 질의 처리 과정을 분석해서 구현이 필요한 부분과 구현을 수정할 부분을 확인함
- GeoMondrian에서 공간 데이터 타입 및 공간 연산자들이 모두 소결합 기법으로 구현되어 있는 것을 확인함
- GeoMondrian에서 공간 데이터 타입 및 공간 연산자들을 밀결합 기법으로 구현하였을 때 기대되는 장점을 분석함

❍ 공간 데이터 타입의 밀결합 구현 및 공간 빅데이터 OLAP 구현 수정

- 공간 빅데이터 OLAP에서 공간 데이터 타입을 저장하거나 불러올 때, 공간 빅데이터 OLAP에 구현된 공간 데이터 타입 대신 PostGIS나 Marmot과 같은 엔진들에 구현된 공간 데이터 타입을 사용하도록 구현함
- GeoMondrian에서 질의를 파싱할 때 사용하는 JAVA API에서 공간 데이터 타입을 지원하지 않기 때문에 파싱 결과에 공간 데이터 타입이 포함되어 있지 않는 문제를 확인함
- 공간 빅데이터 OLAP의 파서 모듈이 공간 데이터 타입을 지원하도록 수정하여 위 문제를 해결함

❍ 공간 연산자들의 밀결합 구현 및 공간 빅데이터 OLAP 구현 수정

- 공간 빅데이터 OLAP에서 공간 연산을 수행할 때, 공간 빅데이터 OLAP에 구현된 API 기반의 공간 연산자를 사용하는 대신 PostGIS나 Marmot과 같은 엔진들에 구현된 공간 연산자를 사용하도록 구현함
- 공간 빅데이터 OLAP에서 질의의 다른 부분들을 모두 처리한 후 마지막에 공간 연산자를 처리하도록 공간 빅데이터 OLAP의 사용자 정의 함수 형태로 지원하고 있는 것을 확인함
- 공간 빅데이터 OLAP의 파서 모듈에서 공간 연산자를 파싱하여 엔진에서 처리한 뒤 그 결과를 받아오도록 수정하여 위 문제를 해결함

⑨ [단위성과 H-10][단위성과 H-11] 공간 빅데이터 기반 복합 이벤트 네트워크 생성 모듈/UI/UX

❍ 분산 복합 이벤트 처리 토폴로지 자동 생성 설명 및 구현

- 분산 복합 이벤트 처리 토폴로지를 정의하고 필요 이유와 구성 요소를 설명함
- Storm을 이용한 토폴로지 생성 과정과 준비 과정에 필요한 각종 용어를 정의하고 생성 예시를 보이고 수동 생성의 한계를 설명
- 토폴로지 자동 생성 모듈의 장점을 설명하고, Redis를 이용해 데이터를 입력받고 어떻게 처리되는지 그 흐름도를 단계별로 설명함
- 각 단계별 구현 코드와 방식, 설정파일 명세를 설명하고 예시를 보임

❍ 분산 복합 이벤트 처리 최적 자동 이벤트 및 룰 분배 모듈 구현

- 모든 서버의 부하의 합을 최소화하는 최적 이벤트 및 룰 분배를 정의하고, 서버 부하를 모델링함
- 사용자에게서 받은 룰 파일과 스트림 데이터 스키마로부터 모델링한 부하를 탐욕적으로 최소화하는 휴리스틱을 개발함
- 개발한 휴리스틱으로 합성 데이터를 받는 복합 이벤트 처리 시스템을 최적화하는 실험을 진행해 서버의 리소스 사용량과 데이터 복제량을 줄임

⑩ [단위성과 H-12] 공간 빅데이터 OLAP UI/UX

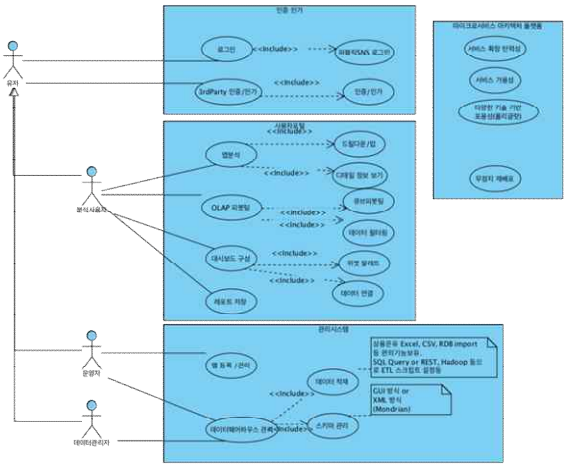
❍ 공간 빅데이터 OLAP UI/UX 요구사항 정의

- Actor 정의

<표 2-93> 공간 빅데이터 OLAP UI/UX 요구사항 정의

구분	설명
유저	시스템 전반적 사용자로, 인터넷에서 SaaS 형태로 언제든 접근하여 가능한 별도의 가입절차 없이 시스템을 사용할 수 있는 유저
분석 사용자	주된 사용자로, 데이터 분석을 위한 피보팅, 가공, 차트, 대시보드를 생성하는 작업을 하고 그 결과를 공유하거나 파일로 저장하여 업무에 활용하는 유저
운영자	시스템 운영자로 기반 데이터 적재, 스키마 관리, 맵 관리 등 시스템 기반 자료와 기능을 관리함
데이터 관리자	데이터 적재와 관련된 관리를 주로 하는 시스템 관리자

- UseCase 도출



<그림 2-330> UseCase

❍ 공간 빅데이터 OLAP UI/UX를 위한 기본 플랫폼 개발

- Rest API 방식의 MDX Client, Rest Mondrian 개발
- 사용자가 요청한 MDX 쿼리는 백엔드에 위치한 GeoMondrian 기반의 공간 OLAP 라이브러리에 전달됨
- 공간 OLAP 라이브러리를 통해 해석된 MDX 쿼리는 SQL로 변환되어 공간 데이터베이스에 데이터를 요청함
- 반환된 데이터는 공간 OLAP상에서 Cube 단위의 데이터 구조로 관리되며 질의된 MDX에 대한 결과는 다시 클라이언트 단의 웹 브라우저에 전달됨
- MVVM패턴의 VueJS를 사용한 UI 개발
- OLAP Vue에서는 위의 Rest Mondrian에 요청한 결과의 공간 정보를 이용하여 지도상에 시각적으로 표시함

⑪ [단위성과 H-13] 공간 빅데이터 Interactive 분석 컴포넌트 테스팅 및 튜닝

❍ 공간 빅데이터 CEP 테스팅 및 튜닝

- 총 7종의 공간 연산자 (Intersect, Disjoint, Overlap, Equal, Contain, Touch, Within)의 처리에 있어서 싱글 머신과 분산 머신에서의 결과를 비교 테스트하여 정확도를 확인함
- 공간 빅데이터 CEP와 실행에 필요한 다른 프로그램들을 패키징한 Docker 이미지를 제작, 배포하여 사용자의 환경에 관계없이 공간 빅데이터 CEP를 용이하게 사용할 수 있게 개선함

❍ 공간 빅데이터 OLAP 테스팅 및 튜닝

- 2세부에서 개발한 OLAP엔진과 1세부에서 개발한 Marmot을 연동하여 Marmot JDBC에 접속 및 SQL Query가 정확히 실행됨을 테스트함

- 특히 1세부에서 개발한 Marmot과의 밀결합을 통해 공간 빅데이터 OLAP 엔진에서 공간 연산을 처리할 때 데이터베이스 수준에서 더 효율적으로 연산을 수행할 수 있도록 지원
- OLAP엔진과 UI/UX 모듈을 연동하여 REST API를 통해 MDX Query가 OLAP 엔진에서 정확히 수행됨을 테스트함
- 공간 빅데이터 OLAP의 UI/UX 접속을 테스팅하여 지도와 다양한 차트를 포함한 시각화 요소들이 정상적으로 생성되는 것을 보이고, 시간/공간 차원에 대한 드릴다운/롤업 기능이 정상적으로 작동하는 것을 확인함

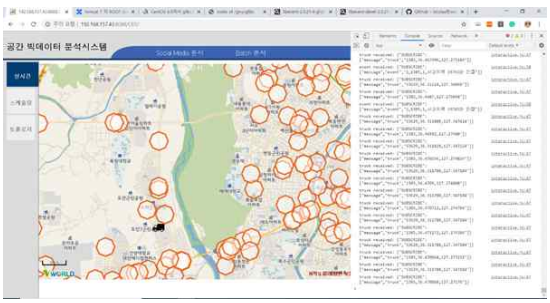
⑫ [단위성과 H-14] 공간 빅데이터 Interactive 분석 컴포넌트 시범 서비스 적용 지원

❍ CEP엔진 사업단 시범서비스 적용 지원

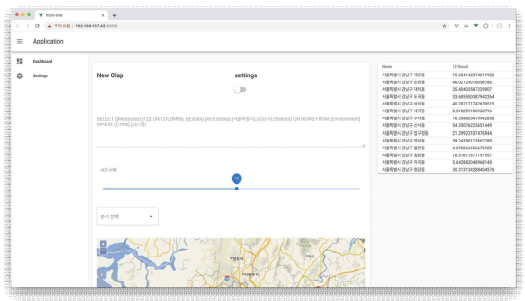
- 한국교통안전공단의 차량 센서 데이터(DTG)와 공간 연산자를 활용해 서울/대전 2700여개 사고지역 근처에 접근하는 차량을 실시간 탐지

❍ OLAP엔진 사업단 시범서비스 적용 지원

- 총 5종의 데이터(유동인구데이터, 시도 데이터, 시군구 데이터, 읍면동 데이터, 블록 데이터)를 1세부의 Marmot을 데이터베이스에 저장하여, Marmot JDBC 모듈을 통해 OLAP 엔진으로 시도별 시간대별 평균 유동인구 처리



<그림 2-331> CEP 엔진 사업단 시범서비스 실행 화면



<그림 2-332> OLAP 엔진 사업단 시범서비스 실행 화면

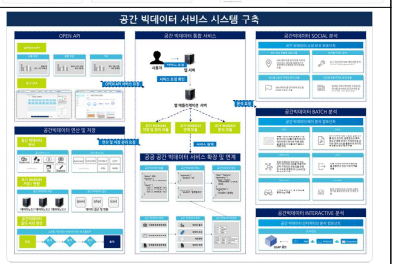
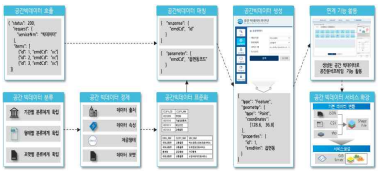
❍ CEP/OLAP 엔진 TTA 인증

- CEP 엔진의 규칙 분배 모듈에 대한 성능 향상 및 정확도 시험을 통해 무작위 분배 방식 대비 제안 방법의 처리량 300% 향상 및 정확도 100% 달성을 확인함
- OLAP 엔진의 MDX 질의 정확도 및 응답 시간 시험을 통해 기존 PostGIS와 연동한 OLAP 엔진 대비 Marmot을 연동한 OLAP엔진이 정확한 결과를 49% 더 빠르게 반환함을 확인함

(3) 공간 빅데이터 서비스 활성화 및 실증 연구

○ 3세부의 연구내용을 수행하여 다음과 같이 6개의 핵심성과를 도출함

<표 2-94> 3세부 연구내용 요약

연구 핵심 성과		수행 기관	연차	연구수행 내용 및 성과
I	공간 빅데이터 플랫폼의 융복합 활용방안 개발	안양대학교 산학협력단	1~6	• 연구단 기술을 통해 개발된 공간 빅데이터 플랫폼의 활용을 활성화하기 위해 관련 정책 및 활용사례를 검토하여 공간 빅데이터 플랫폼이 지향해야 할 방향을 설정하고 정책 및 제도적, 기술적 지원방안 마련 
J	공간 빅데이터의 공공/민간 부문 핵심 서비스 모델 도출 및 제공 방안 개발	안양대학교 산학협력단	1~6	• 국내외 공간 빅데이터 사례를 분석하여 공간 빅데이터 플랫폼에서 제공하는 서비스의 차별화 방안을 마련하고, 비즈니스 모델, 서비스 시나리오 등 핵심서비스 개발 및 서비스 제공을 위한 방안 마련 
K	공간 빅데이터 확장 Open API 서비스	공간인	1~6	• 공간 빅데이터를 표준화된 Open API 기반으로 상호 개방·공유·활용 가능하게 하고 실제 Business에 손쉽게 적용할 수 있도록 하는 기술을 개발함 
L	공간 빅데이터 서비스 제공 UI/UX	공간인	1~6	• 사용자 접근성을 고려한 상호작용 기술과 사용자의 경험을 바탕으로 효율적인 공간 빅데이터 서비스 제공을 위한 UI/UX 방안을 개발함 
M	공공시스템기반 국토교통분야(부동산, 지적, 교통 등) 공간 빅데이터 실증 서비스 시스템	포도	1~6	• 1세부의 공간 빅데이터 저장/접근, 공간연산 기능과 2세부의 공간 분석 기술, 3세부의 활용방안 및 UI/UX를 통합적으로 연계한 공간 빅데이터 실증 서비스 시스템 구축 
N	공공시스템기반 타 부처 공간 빅데이터 실증 서비스 시스템	포도	1~6	• 공공 공간 빅데이터 서비스를 연계하고 공공 공간 빅데이터 정제기법 연구를 통해 생성된 표준화된 공공 공간 빅데이터를 활용하여 유연한 시나리오 생성이 가능한 공간 빅데이터 실증 서비스 시스템 확장 구축 

(가) 공간 빅데이터 플랫폼의 융복합 활용방안 개발(핵심성과 I)

○ 연구단 기술을 통해 개발된 공간 빅데이터 플랫폼의 활용을 활성화하기 위해 기존의 공간 빅데이터와 관련한 정책 및 활용사례 등을 검토하여 공간 빅데이터 플랫폼이 지향해야할 방향을 설정하고, 정책 및 제도적, 기술적 지원방안 마련을 주요 목표로 함



<그림 2-333> 공간 빅데이터 플랫폼의 융복합 활용방안 개발(핵심성과 I)

○ 기술적 지원방안 부문

- 기술적 지원방안 부문에서는 공간 빅데이터 플랫폼이 지향해야할 방향을 검토하여 공간 빅데이터 플랫폼의 사용자 및 제공 서비스 등을 고려한 시스템 아키텍처를 설계하고, 최종 개발된 공간 빅데이터 플랫폼의 사용자 활용가이드를 개발하는 것이 주요 내용임

○ 정책 및 제도적 지원방안 부문

- 정책 및 제도적 지원방안 부문에서는 공간정보 및 빅데이터와 관련된 정책 현황, 법제도 현황, 해외사례 등을 검토하여 정책 및 제도적 측면, 법제마련 측면, 사업추진 측면, 해외사례 측면으로 구분하여 각 시사점을 도출하고, 공간 빅데이터 플랫폼 활성화를 위한 대응방안과 중앙부처 및 유관기관의 역할을 설정하는 것이 주요 내용임

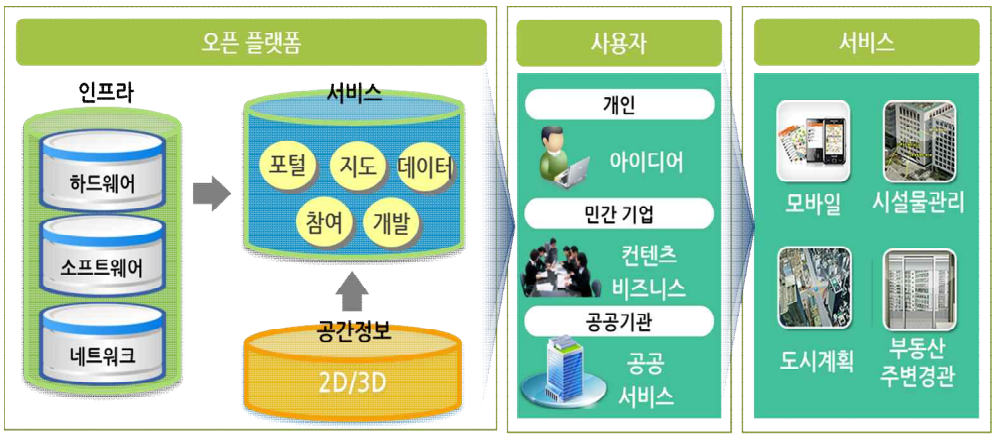
① [단위성과 I-1] V-World 융합기반 공간 빅데이터 플랫폼의 서비스 프레임워크 도출(1차년도)

○ V-World는 공간정보를 활용한 산업활성화 및 국가 경쟁력 제고를 목적으로 구축되었으며, 일반국민 누구나 쉽게 접근하고 활용할 수 있도록 인터넷과 스마트폰으로 지도기반의 다양한 정보를 제공

- 토지, 환경, 도시계획 등 국토관리 분야의 다양한 정보를 2D/3D 형태의 지도기반 형태로 제공

○ V-World의 기능은 크게 인프라, 공간정보, 서비스로 구성되며, 서비스는 포털서비스, 지도서비스, 데이터서비스, 참여서비스, 개발자 서비스로 구분됨

- 포털서비스는 V-World의 모든 서비스를 한 눈에 확인할 수 있도록 메인화면에 콘텐츠별로 구성되어 링크를 통해 관련 세부정보 제공
- 지도서비스는 사용자의 편의성을 고려하여 웹, 스마트폰, 테블릿PC 등 다양한 매체를 기반으로 국가공간정보, 3차원 건물, 영상지도, 행정경계, 교통시설 등의 공간정보를 제공
- 데이터서비스는 사용자가 다양한 공간정보를 손쉽게 검색할 수 있도록 메타데이터 기반의 검색을 지원하고, 그 외 데이터 상세정보 및 미리보기 지원, 데이터 조회수에 따른 인기데이터 정보 등을 함께 제공
- 참여서비스는 사용자가 주변지역의 맛집, 등산로, 데이트 코스 등 관심분야에 대한 콘텐츠를 플랫폼 상에 직접 작성함으로써 다양한 사용자와 정보를 공유할 수 있도록 지원하는 커뮤니티 서비스 제공
- 개발자서비스는 V-World 플랫폼에서 제공하는 공간정보를 활용하여 사용자가 직접 애플리케이션을 개발할 수 있도록 OpenAPI 기능 제공



<그림 2-334> V-World(공간정보 오픈플랫폼) 서비스 프레임워크

○ V-World의 제공 서비스를 검토한 결과, V-World 융합기반의 공간 빅데이터 플랫폼의 서비스 프레임워크 도출을 위한 방안을 다음과 같이 정리함

- 수집・저장측면 : V-World는 공공데이터 위주의 데이터 수요 발생에 따라 데이

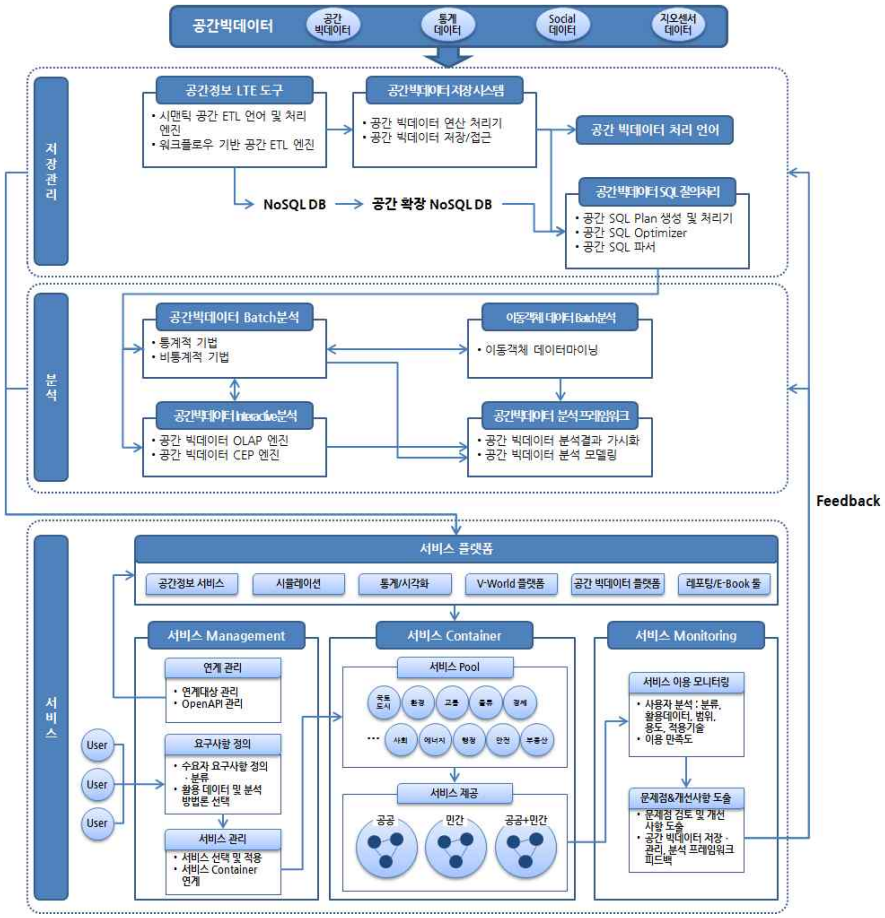
터 보유기관과 연계를 통해 데이터를 확보하고 있으며, 그에 따라 공간 빅데이터 플랫폼에서는 V-World와 연계 및 융합을 위해 데이터 호환성 확보에 필요한 표준체계를 수립하고, 다양하게 수집되는 정형, 비정형 및 실시간 데이터를 저장하고 갱신 관리하는 시스템 구축 및 데이터 확보방안 마련 필요

- 처리・분석측면 : V-World는 타 기관이 보유한 공공데이터를 수집하여 단순 편집・가공・변환을 통해 자체 웹기반의 2D/3D 엔진으로 정보를 제공하며, OpenAPI 서비스를 통해 보유하고 있는 데이터를 제공하고 있음. 공간 빅데이터 플랫폼과의 연계・융합 측면에서 살펴보면, 다양한 유형의 빅데이터의 처리・분석 지원을 위해 우선적으로 시스템의 성능개선 및 질의처리 엔진 개발이 필요하며, 데이터 특성에 맞는 다양한 분석 방법이 요구됨
- 표현・서비스측면 : V-World는 자체 웹기반 엔진을 통해 사용자 체감을 높이는 2D/3D 시각화 기능으로 국토공간정보를 제공하나, 공간정보와 통계 및 행정 정보와의 융복합을 통한 새로운 분석 서비스 제공측면이 부족함에 따라 공간 빅데이터 플랫폼에서는 공간정보 기반의 정보 제공 뿐 아니라 다양한 형태의 융복합 분석을 통한 새로운 서비스 도출이 요구됨

② [단위성과 I-2] 서비스 프레임워크 구성요소 개발 방안(1차년도)

○ V-World 오픈플랫폼, 공간 빅데이터 체계 구축사업 시스템, 교통 빅데이터 시스템, 공간정보 시맨틱 웹서비스 시스템 등 공간정보를 기반으로 한 유사 시스템들의 서비스 프레임워크를 분석하여 공간 빅데이터 플랫폼의 서비스 프레임워크 구성요소를 크게 저장・관리, 분석, 서비스의 3개 항목으로 구분하고, 각 구성요소별 개발 방향을 다음과 같이 설정함

- 저장・관리 : 공간정보 기반의 데이터를 포함한 다양한 유형의 대용량 데이터를 효율적으로 저장하고, 빠른 속도로 처리할 수 있는 시스템 개발
- 분석 : 정형데이터, 공간데이터 외 SNS 정보, 센서를 통한 실시간 정보 등 다양한 유형의 데이터를 분석할 수 있는 분석기술 개발
- 서비스 : 공공뿐 아니라 민간에서도 플랫폼을 활용할 수 있도록 활용주체별 서비스 모델을 개발하고, 다양한 분석결과 등을 OpenAPI를 통해 제공할 수 있는 서비스 개발



<그림 2-335> 공간 빅데이터 플랫폼의 서비스 프레임워크

③ [단위성과 I-3] 공간 빅데이터 서비스 활성화를 위한 정책적, 제도적 지원방안 (2・3차년도)

❍ 공간 빅데이터와 관련한 국내외 정책 및 계획, 추진사업, 해외사례 등을 검토하여 시사점을 도출하였으며, 그를 기반으로 공간 빅데이터 서비스 활성화 지원을 위한 기본방향과 대응과제를 다음과 같이 설정함

- 공간 빅데이터 관련 법제 정비

• 현재 우리나라는 공간 빅데이터와 관련한 법제적 기반이 부재하여 공간 빅데이터에 대한 명확한 정의, 범위, 데이터 구축 및 관련 조직 등이 마련되지 못함에 따라 공간 빅데이터의 활성화가 이루어지지 못하고 있어 공간 빅데이터에 대한 새로운 법안 제정이 요구됨

• 또한, 현재 각 부처별로 빅데이터 관련 업무가 산재되어 있어 기관 간의 협력과 체계적인 정책 수립 및 추진에 어려움이 따르므로 각 부처별 협력을 위한 상위의 전담기구를 신설((가칭) 국가 공간 빅데이터 위원회)하여 국가정책 및 계획 수립의 추진이 필요

- 공간 빅데이터 활성화를 위한 협력체계 구축

• 체계적인 공간 빅데이터의 구축 및 활용을 위해 중앙정부, 지자체, 민간의 명확한 역할 정립과 이들 간의 협력적인 거버넌스 및 추진체계 마련이 요구됨

• 협력적인 거버넌스 및 추진체계 구성을 통해 공공·민간부문의 고품질 정보를 구축, 공유하고 이를 기반으로 새로운 산업의 활성화를 기대할 수 있음

- 개인정보 활용 및 보호를 위한 기준 마련

• 공간 빅데이터를 활용하여 의미 있는 정보를 도출하기 위해서는 각종 개인정보를 포함한 다양한 정보가 요구되나, 개인정보 침해가 우려됨에 따라 「개인정보보호법」에 규정된 개인정보 비식별화의 범위를 확대하고 의무화함으로써 다양한 공간 빅데이터를 구축·활용할 수 있는 환경 조성이 요구됨

• 신용카드사, 이동통신사 등 민간 기업에서 구축하는 데이터는 활용도가 높으나 데이터의 공개 및 공유를 강요할 수 있는 법적 근거가 없으며, 또한 일정 비용을 지불해야 사용 가능함에 따라 민간기업의 데이터 공개를 유도할 수 있는 대책 마련이 요구됨

- 공간 빅데이터 표준화

• 중앙 및 지자체에서 구축하는 공공데이터와 민간에서 구축하는 다양한 민간데이터 간의 포맷이나 형식이 달라 이들 간의 융합을 통해 새로운 정보를 분석하는데 한계가 따르므로 데이터 결합을 위한 공통 Key 생성 등 데이터 표준화의 선행이 요구됨

• 빅데이터 분석은 정형화된 데이터 뿐 아니라 SNS, 블로그, 뉴스기사 등 텍스트 형태의 비정형화 데이터를 포함하기 때문에 비정형데이터의 공간화 및 공간정보와의 융합 분석 등에 대한 다양한 분석기법 개발 연구가 필요함



<그림 2-336> 공간 빅데이터 서비스 활성화 지원을 위한 기본방향 및 대응과제

④ [단위성과 I-4] 서비스 프레임워크 활용 가이드라인 개발(3차년도)

○ 공간 빅데이터 연구와 공간 빅데이터 체계 구축 사업단에서 추진하는 사업은 국토교통부(국토교통과학기술진흥원) 주관으로 추진되며 “공간 빅데이터 활용을 위한 기술 및 서비스 개발”이라는 동일한 목적을 가짐

○ 그에 따라 연구단과 사업단은 공간 빅데이터 연구 시작시점부터 종료시점(2014~2019년)까지 지속적인 업무협의를 거쳐 2가지 방안을 제시함

- 연구단과 사업단의 개발이 거의 동일한 시점에 시작하므로 지속적인 업무협의회

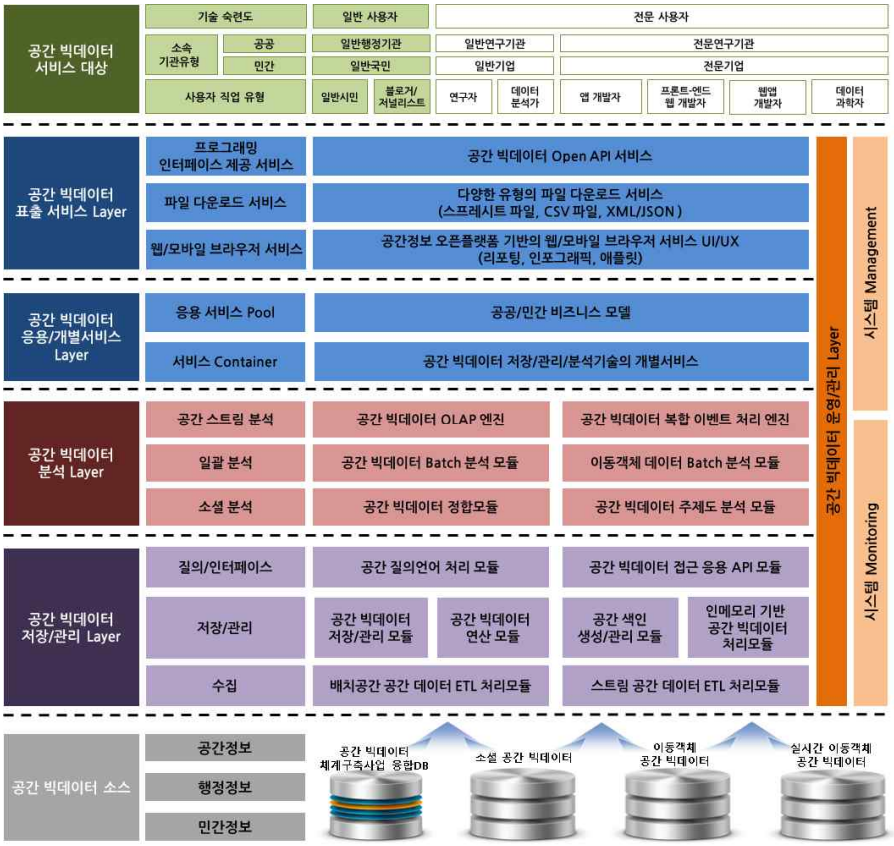
를 개최하여 최대한 중복되는 부분이 없도록 업무를 조정하는 것을 전제로 함

- (1안) 연구단 개발성과를 단독 시스템에 개발(공간 빅데이터 실증서비스 시스템 구축)
- (2안) 시기적으로 사업단이 시스템 구축을 먼저 완료함에 따라 향후 연구단의 개발 성과 중 일부를 사업단 시스템에 적용하는 방안(실용화 방안)

❍ 본 가이드라인에서는 2개 방안을 고려하여 연구단의 개발 성과를 종합한 실증서비스 시스템 아키텍처 설계(안)을 마련하고, 또한 사업단 기능과의 차별성 검토를 통해 연구단 개발 기술을 사업단에 적용하기 위한 방안을 마련함

❍ (1안) 공간 빅데이터 실증서비스 시스템 아키텍처 구성(안)

- 앞서 공간 빅데이터 서비스 프레임워크 구상(안)을 통해 구성요소를 저장·관리, 분석, 서비스의 3개 항목으로 구분하였으며, 이를 좀 더 구체화하여 시스템 아키텍처 구성(안)을 작성함
  - 공간 빅데이터 소스 : 통계데이터, 행정정보데이터, 공간정보데이터, 소셜데이터, 이동객체 데이터, 실시간 이동객체 데이터 등이 공간 빅데이터 플랫폼의 주요 데이터 소스로 활용됨
  - 공간 빅데이터 저장/관리 Layer : Hadoop 환경에서 공간 빅데이터를 고속으로 저장, 관리, 접근하는 기능을 제공하는 공간 빅데이터 저장관리 시스템으로 SQL 기반의 공간 질의 언어를 처리하는 질의처리 엔진, 응용 프로그램에서 공간 빅데이터를 접근하여 사용할 수 있는 응용프로그램 인터페이스 등을 지원
  - 공간 빅데이터 분석 Layer : Hadoop 환경에서 다양한 유형의 데이터를 분석할 수 있는 분석 기술을 제공하며, 소셜분석, 일괄분석, 실시간 분석 등을 지원
  - 공간 빅데이터 응용/개별서비스 Layer : 공간 빅데이터 저장·관리 및 분석모듈을 통해 산출되는 개별 서비스와 공공/민간 비즈니스 모델을 통해 산출된 융복합 분석 결과인 응용서비스를 제공하며, 사용자가 직관적으로 이해하기 쉽도록 UI와 UX를 지원
  - 공간 빅데이터 표출서비스 Layer : 서비스 사용자의 목적과 기술숙련도, 요청사항 등에 따라 웹/모바일 브라우저 서비스, 파일 다운로드 서비스, 프로그래밍 인터페이스 서비스 등 유형을 구분하여 데이터 및 정보를 제공
  - 공간 빅데이터 서비스 대상 : 공간 빅데이터 플랫폼을 사용하는 이용자를 직업, 소속, 기술적 숙련도 등에 따라 유형을 구분하고, 사용자 유형에 따른 응용/개별서비스와 표출서비스 등의 제공 범위를 고려



<그림 2-337> 공간 빅데이터 실증서비스 시스템의 아키텍처 구성(안)

❍ (2안) 공간 빅데이터 실용화 방안

- 연구단과 사업단의 시스템 구성과 기능별 차이를 구분하기 위해 두 시스템간의 아키텍처를 비교하고 기능별 차별성을 검토하여, 연구단의 개발 기술을 사업단에 적용하기 위한 방안을 제시

<표 2-95> 공간 빅데이터 실증서비스 시스템과 사업단 시스템의 차별성

(연구단) 실증서비스 시스템 아키텍처 구성요소	사업단 시스템과의 차별성
공간 빅데이터 소스	• 사업단은 공간정보, 행정정보, 민간정보 및 이전 사업을 통해 생성된 융합DB를 주요 데이터 소스로 활용하며, 연구단은 공간정보, 행정정보, 민간정보를 비롯한 이동객체 데이터와 실시간 이동객체 데이터 등을 활용하므로 사업단에 비해 좀 더 다양한 형태의 데이터를 활용함
공간 빅데이터 저장/관리 Layer	• 연구단에서 개발한 공간 빅데이터 저장관리 시스템은 공간 빅데이터에 대한 공간 인덱스를 Hadoop 분산환경에 맞게 지원하고, 엔진수준의 공간연산자와 특화된 질의처리 기능을 지원하여 공간 빅데이터 저장, 접근, 연산, 질의처리 성능이 모두 향상됨 • 또한 인메모리 기반의 분산병렬처리 기술을 통해 실시간 빅데이터의 처리 성능이 향상됨
공간 빅데이터 분석 Layer	• SNS 등 비정형 텍스트에서 위치정보 추출 기술 및 장소에 대한 감성분석 기술을 개발함으로써 소셜데이터를 활용한 다양화된 분석이 가능 • 이동객체 데이터의 일괄분석, 스트림데이터 분석 등은 사업단에서 지원하지 않는 기술로서 사업단 분석모듈과 차별성이 있음
공간 빅데이터 응용/개별서비스 Layer	• 사업단은 서비스 대상을 한정하여 공공분야에 해당하는 응용서비스를 개발함으로써 정책지원 및 의사결정에 활용하나, 연구단은 사용자 범위를 확대하여 공공과 민간에서 활용할 수 있는 서비스를 구분하여 개발

(연구단) 실증서비스 시스템 아키텍처 구성요소	사업단 시스템과의 차별성
공간 빅데이터 표출서비스 Layer	• 사업단은 사용자에게 최적화된 분석도구를 이용하여 지도기반의 효과적인 시각화를 통한 대화형 분석도구를 제공하며, 연구단은 수요조사를 통해 도출된 서비스에 따라 적합한 정보제공 유형을 설정하여 제공함
공간 빅데이터 서비스 대상	• 사업단 시스템의 주사용자는 공공기관에서 공간 빅데이터를 사용한 경험이 있는 사람을 대상으로 하며, 연구단 시스템의 사용자는 공공을 비롯한 민간의 전문가, 일반국민을 포함하므로 서비스 대상범위가 좀 더 넓음

❍ 연구단 개발성과를 사업단 시스템에 적용하기 위한 단계적 방안을 다음과 같이 계획함



<그림 2-338> 사업단 시스템 지원을 위한 연구단 성과의 단계별 적용 방안

<표 2-96> 사업단 시스템 지원을 위한 연구단 성과의 단계별 적용 방안

단계구분	연구추진일정	기술지원전략	주요내용
1단계 ('14~'16)	1차년도	개별모듈의 구현	• 공간 빅데이터 연구단과 사업단의 지속적인 협의를 통해 연구단과 사업단의 차별화되는 핵심요소기술 선정 • 선정된 핵심요소기술을 지원하는 공간 빅데이터 저장·관리·분석·서비스 모듈 구현
	2차년도		
	3차년도		
2단계 ('17~'18)	4차년도	적용가능성 테스트 및 플랫폼 탑재모듈 선정 (1차 기술지원)	• 3차년도까지 개발한 연구단의 성과를 사업단 플랫폼에 적용하여 융합분석, 시계열분석 중심의 플랫폼으로 확장될 수 있도록 1차 기술지원 • 구현된 개별모듈을 사업단의 테스트용 서버에 이전하여 적용가능성을 테스트하고, 테스트 결과를 바탕으로 최종적으로 사업단 플랫폼에 탑재할 모듈 선정
	5차년도		
3단계 ('19)	6차년도	플랫폼 탑재모듈 개선 및 고도화 (2차 기술지원)	• 5차년도까지 개발한 연구단의 성과를 사업단 플랫폼에 적용하여 실시간 데이터 수집 및 처리, Spatial Hadoop을 이용한 대용량 처리 및 분석, 대용량/실시간 처리가 가능한 플랫폼으로 확장될

단계구분	연구추진일정	기술지원전략	주요내용
			수 있도록 2차 기술지원 • 1차 기술지원을 통하여 선정된 탑재모듈의 추가 개선

❍ 6차년도까지의 연구단 연구개발 성과를 최종 검토한 결과 사업단에 실증하는 대상 기술을 다음과 같이 설정함(연구단과 사업단 협의 결과)

<표 2-97> 공간 빅데이터 플랫폼(사업단) 실증 대상 기술

구분	연구기관	실증 대상 기술	주요내용
1세부	한국전자통신연구원	• 공간빅데이터 저장 관리 플랫폼 (저장구조, 검색 인터페이스, 분석 워크플로우 제어 기능, 비공간 연산 분석 라이브러리)	• 스트림 기반의 공간 분석 플로우 수행 기능을 제공하는 공간빅데이터 저장 관리 플랫폼을 사업단 플랫폼에 이식하고, 기존 시스템의 인터페이스와 연계하여 실증. 특히 속성 기반 분석과 여러 개의 단위 분석을 한번의 Map/Reduce 작업으로 처리하는 복합 분석 모듈을 실증함
	포스웨이브	• 실시간 인메모리 기반 저장 플랫폼 v1.0 (저장구조, 색인/검색 인터페이스, 공간분석라이브러리 23종)	• Spark 기반의 실시간 공간빅데이터 저장 플랫폼 및 공간분석라이브러리(23종)을 사업단 플랫폼에 이식하고, 기존 시스템의 인터페이스와 연계하여 실증
2세부	시터스	• 공간빅데이터 분석 및 가시화 기술 (층간소음 분석 기술, 공공시설 매력도 측정 및 접근성 분석 기술(예: 공공도서관), 머신러닝 기반 특정주제 문장 분류 기술(예: 재난재해 문장))	• 사업단과 중복되지 않는 기술들로 실증 대상 기술을 선정함 • 기존 윈도우 개발 환경의 분석 모듈을 사업단에서 사용하고 있는 리눅스 개발 환경에 맞게 변경 및 적용 • 1세부 시스템의 연동을 통한 분석 모듈 결과 제공
	㈜가이온	• 공간 빅데이터 분석모듈 : E2SFCA 분석 모듈, Decision tree 분석 모듈 • 이동객체 빅데이터 분석모듈 : Fishnet Algorithm 모듈, Map-Matching 모듈	• 사업단과 중복되지 않는 기술들로 실증 대상 기술을 선정함 • E2SFCA는 대중교통 접근성 분석, Decision tree는 생활폐기물 배출추정, Fishnet Algorithm은 심야 대중교통 수요분석, Map-Matching은 택시운영 최적화를 통해 실증함 • 1세부 시스템의 연동을 통한 분석 모듈 결과 제공
	㈜유엔진솔루션즈	• 공간 빅데이터 OLAP 엔진 기술 (지도상에서의 드릴 다운 및 롤업 인터페이스, 관련 차트 가시화 기술)	• 1세부 Marmot과의 연동을 통해 사업단에 실증하는 대신 이미 사업단에서 사용중인 Hive와의 연동을 빠르게 시도함

⑤ [단위성과 I-5] 공간 빅데이터 플랫폼 활용가이드 개발(6차년도)

❍ 공간 빅데이터 플랫폼 활용가이드는 사용자가 공간 빅데이터 플랫폼에서 제공하는 서비스를 쉽게 이해하고, 활용할 수 있도록 지원하기 위한 목적으로 개발

❍ 공간 빅데이터 플랫폼 활용가이드(매뉴얼)는 화면 구성, 실증시각화 서비스, 공간분석프레임, 관리서비스, OpenAPI순으로 서비스 항목을 구분하여 작성

❍ 메인화면



<그림 2-339> 공간 빅데이터 플랫폼 메인화면 구성

○ 실증시각화

- 공공데이터, 공간분석 데이터 및 서비스모델 등을 가시화함
- 각 화면항목별로 리스트 내용, 검색 결과 등 순으로 매뉴얼을 작성



<그림 2-340> 실증시각화 화면 구성

○ 공간분석프레임

- 세부별 기능을 모듈단위로 연계하여 공간연산 및 공간분석을 수행하고, 작업목록에서 분석 완료된 기능은 실증시각화 서비스에서 표출
- 각 연산별로 파일업로드에서 분석결과 도출 및 실증시각화 서비스에서 표출까지 일련의 과정을 매뉴얼로 작성



<그림 2-341> 공간분석프레임 화면 구성

❍ 관리서비스

- 공간 빅데이터 서비스 시스템을 전반적으로 관리할 수 있는 서비스로 하둡 리소스를 확인하고 공간분석 리스트 및 통계 확인이 가능
- 관리항목별 제공내용의 상세내역을 확인하는 방법으로 매뉴얼 작성



<그림 2-342> 관리서비스 화면 구성

❍ OpenAPI서비스

- 공간 빅데이터 핵심서비스 및 서비스 시나리오에 활용된 원시데이터와 분석결과 데이터를 제공하고, 공공데이터 포털에서 OpenAPI 형태로 제공하는 공공데이터를 수집·관리하는 서비스
- 레퍼런스 항목에 부문별 제공 원시데이터와 분석결과데이터를 확인하고 OpenAPI로 가져오는 일련의 순서에 따라 매뉴얼 작성
- 스케쥴러 목록 및 해당항목의 정보 및 이력주기를 확인하는 방법으로 매뉴얼 작성



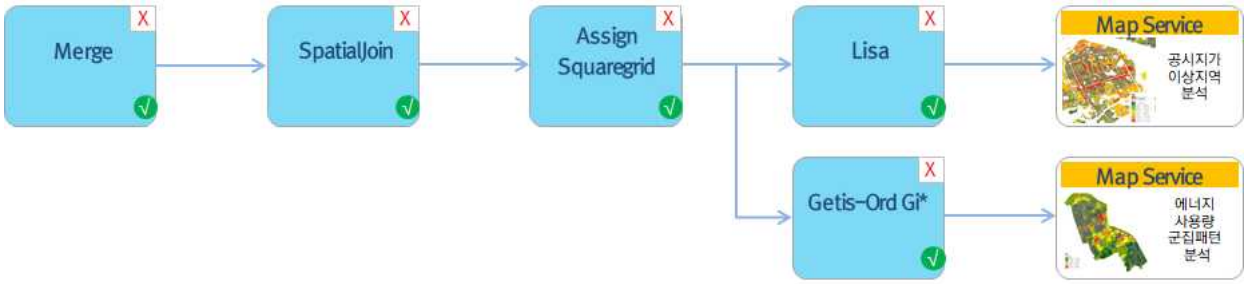
<그림 2-343> OpenAPI서비스 화면 구성

⑤ [단위성과 I-6] 핵심서비스 모델 기반 활용가이드 적용예시(6차년도)

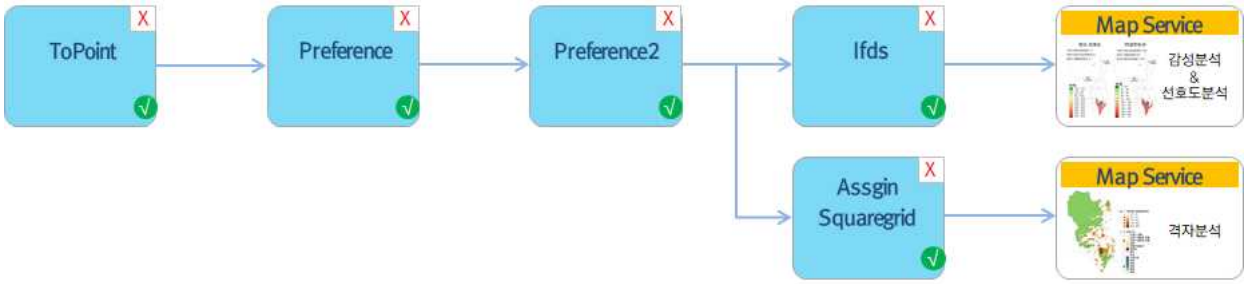
❍ 핵심서비스 모델을 기반으로 한 활용가이드 적용예시 개발

- 서비스 유형에 따라 구분한 3개의 서비스 시나리오를 기반으로 분석모듈 적용

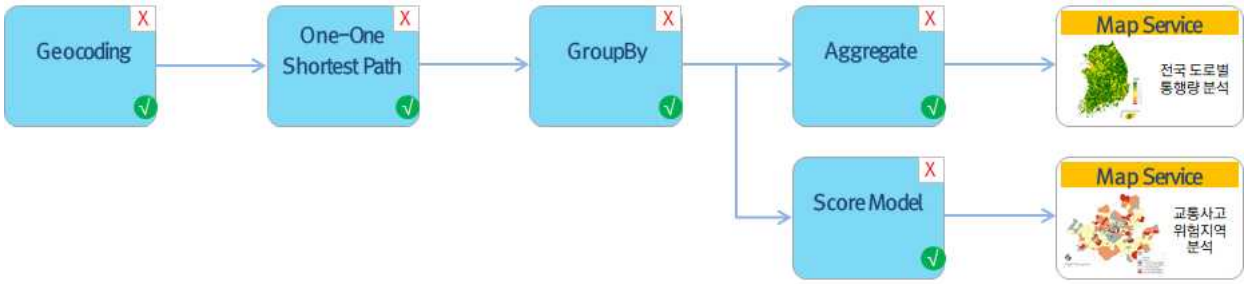
순서에 맞춰 해당모듈을 워크플로우 상에 배치하고, 파일 업로드에서 실증시각화, 분석결과의 OpenAPI 서비스까지 일련의 과정을 매뉴얼로 작성하여 사용자가 워크플로우 형식의 분석모듈 활용을 지원할 수 있도록 매뉴얼 작성



<그림 2-344> 토지건물정보 기반 서비스 시나리오 적용예시(워크플로우)



<그림 2-345> 비정형텍스트정보 기반 서비스 시나리오 적용예시(워크플로우)



<그림 2-346> 네트워크정보 기반 서비스 시나리오 적용예시(워크플로우)

(나) 공간 빅데이터의 공공/민간부문 핵심서비스 모델 도출 및 제공방안 개발(핵심성과 J)

○ 공간 빅데이터의 공공/민간부문 핵심서비스 모델 도출 및 제공방안 개발은 국내외 빅데이터와 공간 빅데이터 사례를 분석하여 공간 빅데이터 플랫폼에서 제공하는 서비스의 차별화 방안을 도출하고, 비즈니스 모델, 서비스 시나리오 개발 등 핵심 서비스 개발과 서비스 제공을 위한 방안 마련을 주요 목표로 함



<그림 2-347> 공간 빅데이터의 공공/민간부문 핵심서비스 모델 도출 및 제공방안 개발(핵심성과 J)

○ 공간 빅데이터 서비스 차별화 방안 도출 부문

- 국내외 빅데이터 및 공간 빅데이터 사례 총 40건을 분석하여 빅데이터와 공간 빅데이터 간의 데이터 특성, 활용분야, 활용주체에 따른 각 시사점을 분석하여 공간 빅데이터 서비스의 범위와 역할을 설정하고, 공간 빅데이터 서비스만의 차별화 방안을 도출

○ 공간 빅데이터 비즈니스 모델 및 서비스 시나리오 개발 부문

- 수요조사를 수행하여 공간 빅데이터 분석기능과 서비스 분야별 우선순위를 도출하고, 분야별 핵심서비스 유형을 개발함. 이를 토대로 핵심서비스 분야별 메가트렌드와 외부환경 분석을 수행하고, 연구단 개발 기술 및 공공데이터 활용 가능여부 등을 검토하여 비즈니스 모델 및 서비스 시나리오를 개발

○ 공간 빅데이터 시스템 아키텍처 설계 및 주요기능 도출 부문

- 공간 빅데이터 플랫폼에서 효과적인 서비스 제공을 위해 최적화된 형태의 시스템 아키텍처를 설계하고, 플랫폼 제공 서비스에 따른 주요기능을 도출하여 사용자로 하여금 공간 빅데이터 플랫폼 서비스를 손쉽게 활용할 수 있도록 지원

① [단위성과 J-1] 공공/민간부문 서비스 유형별 공간 빅데이터 서비스 현황 수집 및 분석(1차년도)

○ 공간 빅데이터 서비스 사례를 분석하기 위해 국내외 빅데이터 및 공간 빅데이터와 관련한 총 40건의 서비스 사례를 검토하였으며, 이를 토대로 빅데이터 및 빅데이터의 특성, 활용분야, 활용주체를 구분하여 사례분석 결과를 도출함

<표 2-98> 국내외 빅데이터 및 공간 빅데이터 서비스 사례 분석결과

구분		시사점
데이터 특성	빅데이터 (3V)	• 정적인(Static) 형태의 데이터 분석 사례가 대부분이며, 실시간 데이터 수집을 통한 분석사례는 거의 없는 것으로 나타남 • 서비스 측면보다는 분석측면의 성향이 강하게 나타남 • 분석결과를 효과적으로 전달할 수 있는 시각화 지원 기술 부족 • 타시스템과 연계를 통한 융복합 활용지원 기술 부족 • 빅데이터 관련 서비스 사례 대부분이 대용량(Volume)>신뢰성(Veracity)>가치(Value) 순으로 데이터 특성이 적용된 것으로 나타남
	공간 빅데이터 (7V)	• 정적인(Static) 형태의 공간데이터를 활용한 분석 사례가 대부분이며, 실시간 데이터 수집을 통한 분석 사례가 적은 것으로 나타남 • 공간정보 데이터의 수집을 통한 분석보다는 빅데이터 분석을 가시화하는 과정에서 시각화 자료로서 공간정보 데이터가 활용되는 편임 • 공간 빅데이터 관련 서비스 사례에서는 빅데이터에 분석에 비해 7V의 특성이 비교적 잘 나타났으며, 가치(Value)>시각성(Visualization)>대용량(Volume) 순으로 데이터 특성이 적용된 것으로 나타남
활용 분야	인문·사회	• 의료, 범죄, 민원 등 공공기관의 행정 및 정책 지원을 위해 빅데이터 및 공간 빅데이터 분석이 주로 활용됨
	산업·경제	• 카드사, 백화점 등에서 고객정보 분석 통해 고객의 성향 및 수요 등을 파악하여 기업의 수익 증대 및 마케팅에 활용하는 사례가 많은 것으로 나타남
	환경	• 기후, 산사태, 폭우, 지진 등 수십년간 축적된 상데이터를 종합 분석하여 미래를 예측하고 대응하기 위한 전략 마련에 주로 활용됨
	기반시설	• 지역 내 교통혼잡 해결, 버스노선 결정, 내비게이션 개발 등 교통과 관련한 문제해결을 위해 실시간 교통정보를 수집하여 분석하고, 시각화하는 서비스가 주로 활용됨
	기타	• 테니스, 배구, 농구, 야구 등 스포츠 종목별 선수 개인기록이나 경기별 기록을 분석하여 상대 선수 및 팀별 경기전략 등에 활용됨
활용 주체	공공	• 공공기관에서는 의료, 범죄, 민원을 비롯한 기상예보, 교통문제 해결 등 사회 전반적인 문제 해결을 위해 빅데이터 또는 공간 빅데이터 분석 및 서비스를 주로 활용하는 것으로 나타남
	민간	• 민간기업에서는 수년간 축적된 고객정보를 기반으로 고객의 수요 및 성향을 파악하여 그에 따른 점포입지 분석, 마케팅 등 기업의 매출 증대를 위해 주로 활용하는 것으로 나타남

○ 구분 항목별 서비스 사례를 분석한 결과, 공간 빅데이터 서비스가 지향해야 할 종합시사점을 도출하면 다음과 같음

- 공간 빅데이터는 빅데이터에 비해 7개의 데이터 특성 가운데 시각성(Visualization)이 두드러지게 나타나는 것을 알 수 있음
- 공간 빅데이터는 일반적인 대용량의 빅데이터 분석을 통해 얻어진 결과를 공간적으로 가시화하여 보여주는 것에 그치지 않고, 공간정보 데이터를 수집하여 다양한 공간분석 기법을 통해 새로운 가치를 생성할 수 있도록 데이터의 수집과

분석과정에 있어서 공간정보 데이터의 활용이 요구됨

- 활용분야별로는 서비스 주 사용자인 활용주체가 그 목적에 따라 공공과 민간으로 명확하게 구분되는 경향을 나타냄. 즉, 인문・사회, 환경, 기반시설 분야에서는 주로 공공에서 사회문제 해결을 위해 빅데이터 및 공간 빅데이터를 활용하는 것으로 나타났으며, 산업・경제, 기타 분야에서는 주로 민간기업의 마케팅이나 매출증대를 위한 목적으로 빅데이터 및 공간 빅데이터를 활용하는 것으로 나타남

② [단위성과 J-2] 공간 빅데이터 서비스의 차별화 방안 도출(1차년도)

❍ 앞서 공간 빅데이터 사례분석을 통해 얻어진 시사점을 토대로 공간 빅데이터 서비스의 차별화 방안을 도출하기 위해 공간 빅데이터 서비스 분야를 재설정하고, 서비스 활용주체를 구분하여 공간 빅데이터 특성이 잘 드러날 수 있도록 공간 빅데이터의 역할 및 범위를 설정함

- 우선, 공간 빅데이터 서비스 분야 재설정을 위해 미래사회 이슈를 검토하여 분야별 메가트렌드를 정리하고, 분야별 메가트렌드에 대응하기 위한 공간 빅데이터 역할을 정의함

❍ 국토연구원(2014)의 연구에 따르면, 미래에는 인구, 사회, 기술, 기후변화, 경제, 정치 등 6개 분야에서 트렌드 변화가 있을 것으로 예측하였으며, 이를 기반으로 미래사회 특성을 크게 4가지로 정리하고, 그에 대응하는 공간 빅데이터의 역할을 검토함

- 첫째, 미래사회에서 발생할 수 있는 일을 정확하게 예측하는 것은 어려운 일임에 따라 불확실성은 미래사회 특성 중 하나에 해당되며, 인문, 사회분야의 다양한 정보를 활용하여 시계열적 분석을 통해 미래를 예측함으로써 미래사회 변화에 유연하게 대처가 가능함

- 둘째, 환경오염, 자원부족, 경제 양극화, 정치적 이슈 등 환경, 정치, 경제 각 분야에서는 리스크의 특성을 지니고 있으며, 공간 빅데이터는 이러한 위험요소의 분석을 통해 이상 징후를 미리 발견하고, 문제 해결을 위한 의사결정을 지원이 가능함

- 세 번째는 지속적인 기술의 발전으로 인해 미래사회는 좀 더 스마트한 형태의 도시로 변모할 것이며, 공간 빅데이터는 축적된 데이터를 분석하여 다양한 정보를 제공함으로써 인공지능(AI) 등의 기술과 함께 스마트한 도시를 형성하는 도구가 될 것임

- 네 번째 미래사회 특성은 융합으로, 미래사회는 다양한 분야의 융・복합을 통해 새로운 시장이 창출될 것으로 예상되며, 공간 빅데이터는 타 분야간의 정보결합을 통해 새로운 가치를 생성하는 도구로서 중요한 역할을 할 것임

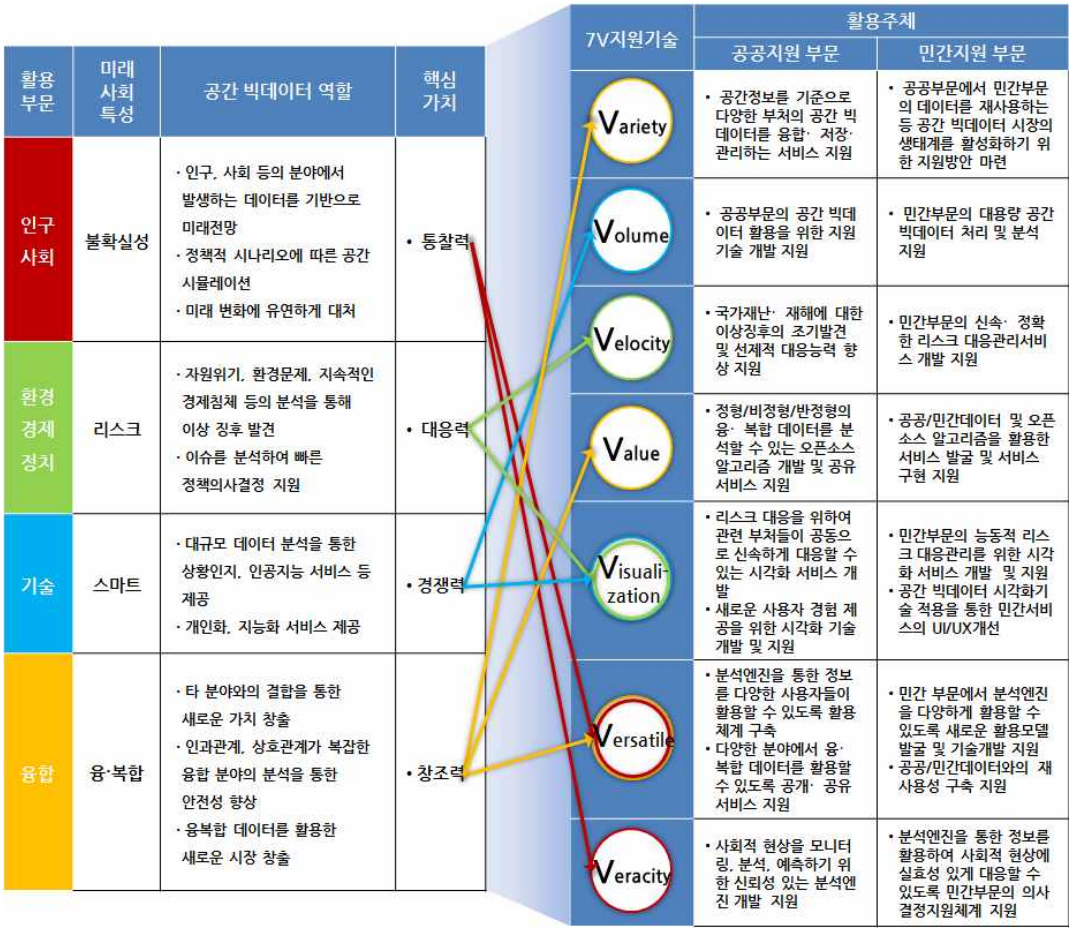
<표 2-99> 미래사회 특성에 따른 공간 빅데이터의 역할

분야	트렌드 변화	미래사회 특성	공간 빅데이터의 역할	
인구	• 저출산, 고령화 • 1인 가구, 다문화가구 등 가족 형태 변화	불확실성	• 인구, 사회분야에서 발생하는 각종 데이터를 기반으로 미래전망 • 정책적 시나리오에 따른 공간 시뮬레이션 • 미래 변화에 유연하게 대처	통찰력
사회	• 개인의 삶의 질 중시 • 건강, 여가, 문화수요 증대 • 사회관계 단절			
기술	• 과학기술의 융복합화 확대 • 신산업 및 기술경쟁 가속화	스마트	• 대규모 데이터 분석을 통한 상황 인지, 인공지능 서비스 등 제공 • 개인화, 지능화 서비스 제공	경쟁력
기후변화	• 기후변화, 재해, 질병위험 증가 • 자원부족 심화	리스크	• 자원위기, 환경문제, 지속적인 경제 침체 등의 분석을 통해 이상 징후 발견 • 이슈를 분석하여 빠른 정책의사결정 지원	대응력
경제	• 경제저성장 추세 및 양극화 심화 • 글로벌 교류 확대			
정치	• 분권화와 참여, 소통방식 다양화 • 남북간 교류 및 경제협력 증대			
융합			• 타분야와의 결합을 통한 새로운 가치 창출 • 인과관계, 상호관계가 복잡한 융합 분야 분석을 통한 안전성 향상 • 융복합 데이터를 활용한 새로운 시장 창출	창조력

❍ 이처럼 공간 빅데이터는 다양한 분야에서 미래에 대응할 수 있는 기술이라고 할 수 있으며, 공간 빅데이터의 정의에 따라 활용 가능한 범위는 크게 관리, 분석, 서비스로 구분할 수 있음

- 관리(Resource) : 복잡한 대용량의 공간 빅데이터를 수집·저장하여 다양한 정보를 관리하는데 활용 가능
- 분석(Analysis) : 공간 빅데이터 분석 기술을 통해 다양한 형태의 대용량 데이터를 분석하여 새로운 가치를 생성하는데 활용 가능
- 서비스(Service) : 공간 빅데이터 관리 및 분석을 통해 생성된 결과를 활용하여 다양한 분야에서 의사결정을 지원하는데 활용 가능

❍ 지금까지 살펴 본 바와 같이 미래사회 특성을 통해 공간 빅데이터의 역할과 범위를 살펴보았으며, 공간 빅데이터가 지원할 수 있는 서비스를 다음과 같이 공공과 민간으로 구분하여 차별화 방안을 도출함



<그림 2-348> 공간 빅데이터 서비스 차별화 방안 및 활용전략

③ [단위성과 J-3/J-4] 공간 빅데이터 핵심 서비스 유형 및 서비스 모델 개발 (2 · 3차년도)

㉮ 수요조사를 통한 서비스 유형화

○ 공간 빅데이터 서비스 유형화 및 핵심서비스 모델을 개발하기 위해 수요조사를 수행함

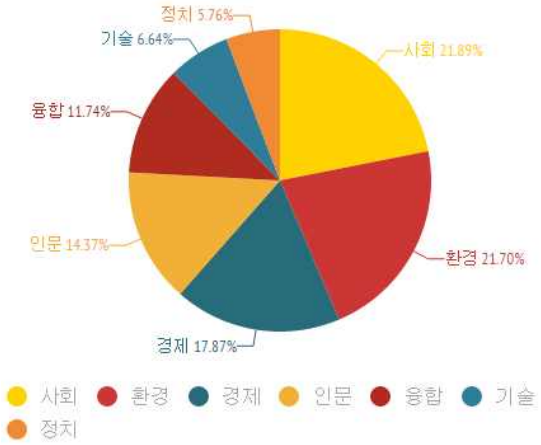
- 수요조사 수행일시 : 2015년 5월 15일 ~2015년 6월 8일, 총 140부 회수
- 수요조사 대상자 : 공간정보 및 빅데이터와 관련이 있는 학생, 교수, 공공기관, 일반기업, 연구직 등
- 수요조사 범위 : 공간 빅데이터 서비스 분야, 분석기능, 활용데이터, 데이터 제공유형, 서비스 제공방법 등

○ 또한, 공간정보 및 빅데이터 관련 직종에 10년 이상의 경력을 가진 전문가를 대상으로 공간 빅데이터 서비스 분야와 분석기능에 대한 평가를 진행하였으며, 수요조사와 전문가 평가 결과를 토대로 QFD(Quality Function Deployment) 기법을 적용하여 공간 빅데이터 서비스 분야와 분석 기능에 대한 관련성을

통해 상대적 중요도와 우선순위를 도출함

○ 사회부문이 21.89%로 가장 중요한 서비스 유형으로 파악되었으며, 그 다음으로 환경(21.70%), 경제(17.87%), 인문(14.37%), 융합(11.74%), 기술(6.64%), 정치(5.76%) 순위로 분석됨

○ 공간 빅데이터에서 요구되는 분석기능의 우선순위로는 핫스팟 분석(12.15%)이 가장 중요한 기능으로 나타났으며, 그 다음으로는 패턴분석(11.54%), 군집분석(9.42%) 순으로 나타남



<그림 2-349> 공간 빅데이터 서비스 분야 우선순위 도출

<표 2-100> 공간 빅데이터 분석기능 우선순위 도출

분석기능	중요도(%)
핫스팟분석	12.15
패턴분석	11.54
군집분석	9.42
연결성분석	8.85
근접분석	8.28
빈도분석	7.23
선호도분석	6.87
입지적합성분석	6.63
토픽연관성분석	6.57
키워드분석	6.36
시계열분석	6.34
공간가중회귀분석	5.33
교차분석	2.93
로지스틱회귀분석	1.51

○ 서비스 부문별로 요구되는 분석기능이 다소 차이를 보이는 것으로 나타났으며, 각 부문별 효율적인 서비스 제공을 위해서는 서비스모델 적용분야 도출, 분석기능의 조합, 프로세스 모델링, 데이터의 연계방안 마련 등이 요구됨

㈏ 서비스 유형화에 따른 핵심 서비스 모델 개발

○ 앞서 도출한 서비스 우선순위를 기반으로 세분화된 서비스 분야를 도출하고, 1세부 및 2세부 개발 분석기술과 공공시스템 데이터의 연계방안을 검토하여 핵심서비스 모델을 개발



<그림 2-350> 서비스 유형화에 따른 핵심서비스모델 개발 절차

○ 일련의 절차를 거쳐 부문별 핵심서비스 모델을 개발한 결과는 다음과 같음

- 사회부문 : 환자의료이용지도(Health Map) 서비스

  • 목적: 환자의 연령, 질환종류, 소득수준과 이동행태, 거주지 내 의료기관 분포 현황, 진료비 등을 종합적으로 분석하여 의료자원 분포와 의료이용간의 상관관계를 분석하고, 의료이용 패턴을 분석함으로써 정부의 의료보건 정책 시 의사결정 지원이 가능

- 환경부문 : 온실가스 배출량 분석 및 모니터링 서비스

  • 목적: 산업부문별, 지역별 온실가스 배출량을 시계열 자료로 구축하여 지속적인 온실가스 배출량 모니터링을 수행함으로써 정책도입에 따른 온실가스 감축효과를 판단할 수 있는 시뮬레이션 시스템을 구축하고, 의사결정 지원 가능

- 경제부문 : 공간 빅데이터 활용 상권분석 서비스

  • 목적: 공공의 행정정보, 통계정보, 공간정보 및 민간의 카드사용 정보, 유동인구 정보 등 다양한 정보를 공간단위로 융합 분석하여 공공 및 개인사업자를 위한 상권분석을 수행함으로써 지역경제 및 상권 활성화 지원

- 인문부문 : 인구이동데이터의 시・공간 가시화 서비스

  • 목적: 공공부문에서 통계자료 형태로 관리・제공하는 인구이동데이터를 시・공간정보 기반으로 가시화하여 노령인구의 증가율, 생산가능 인구의 감소율, 지역별 인구이동변화 패턴 등을 신속하게 파악하여 관련정책 수립 및 모니터링을 위한 주요 지원정보로 활용

- 융합부문 : 사회취약계층 중심 데이터기반 맞춤형 공간정책 지원 서비스

  • 목적: 에너지, 주거, 의료, 교통, 교육 등 분야별 취약성 지표를 도출하고, 기준을 설정한 후 이를 기반으로 분야별 가중치를 적용하여 공간분석을 수행함으로써 종합적인 관점의 사회취약지역을 도출함으로써 복지정책의 의사결정을 지원

- 기술부문 : 공간정보 기반의 기술네트워크 분석 서비스

  • 목적: 특정 기술분야에 대한 기술생태계의 모습을 시공간 및 특정 주제별로 가시화하여 기술현황을 직관적으로 이해할 수 있는 주제도를 구축하고, 사용자가 시공간 범위 및 특정 주제를 설정하여 맞춤형 기술 주제도를 작성할 수 있도록 지원

- 정치부문 : 공간 빅데이터를 활용한 선거전략 지원 서비스

  • 목적: 페이스북, 트위터 등 SNS정보 및 역대선거정보, 인구주택 센서스 정보 등을 활용하여 정당별, 지역별 공간적 영향력을 종합적으로 분석함으로써 선거 후보자의 공약에 따른 선거결과의 사전예측 정보를 제공

④ [단위성과 J-5] 공공/민간부문 공간 빅데이터 활용을 위한 비즈니스 모델 분석 및 개발(4차년도)

○ 앞서 핵심서비스 모델 개발을 통해 공간 빅데이터의 활용분야 및 활용방안에 대한 방법론과 방향을 살펴보았으며, 이를 토대로 연구단에서 개발한 공간 빅데이터 기술과 플랫폼의 효율적인 활용을 위해 공공과 민간으로 구분하여 공간 빅데이터 비즈니스 모델을 개발함

○ 공간 빅데이터 비즈니스 모델 개발을 위한 절차는 다음과 같음

- 미래사회와 관련한 국내외 다수의 자료를 검토하여 공간 빅데이터에서 최종적으로 제공할 6개의 서비스 분야를 도출하고, 분야별 메가트렌드 및 외부환경(Opportunity & Threat) 분석을 수행함
- 외부환경(OT) 분석은 메가트렌드에 따른 시장 환경을 분석하여 공간 빅데이터 플랫폼에서 제공할 비즈니스 모델의 방향을 설정하기 위함이며, 서비스 분야별로 공공과 민간으로 구분하여 기회요인과 위협요인을 분석함
- 다음 과정으로, 공공과 민간으로 구분한 외부환경(OT)분석 결과에 정책과 기술개발 동향을 검토하였으며, 공공부문의 경우, 국토교통부의 부서별 주요 추진사업 및 공간정보 업무 등을 검토하고, 국토교통부와 과학기술정보통신부에서 수행 중인 R&D사업을 검토함
- 민간부문의 경우, 외부환경(OT) 분석결과와 관련하여 현재 진행되고 있는 사업이나 기술 등을 검토함
- 마지막으로, 연구단에서 개발 중인 저장, 관리기술(1세부 개발 기술) 및 분석기술(2세부 개발기술)과 공공시스템에서 제공하는 데이터의 활용가능 여부를 검토하여 공간 빅데이터 비즈니스 모델 영역을 설정하고, 영역별 비즈니스 모델 구상(안)을 도출함
- 공공과 민간으로 구분하여 도출된 비즈니스 모델 구상(안)에 따라 공간 빅데이터 플랫폼에서 제공 가능한 서비스 내용, 서비스 사용 대상자, 서비스 방법 및 이유 등을 정립하여 총 30개의 비즈니스 모델을 개발함



<그림 2-351> 공간 빅데이터 비즈니스 모델 개발 절차

○ 6개 서비스 분야별 공공과 민간으로 구분되어 개발된 30개의 공간 빅데이터 비즈니스 모델은 다음과 같음

<표 2-101> 공공/민간부문 비즈니스 모델 개발

분야	부문	비즈니스 모델 영역	비즈니스 모델 개발				
			WHAT	WHO	WHEN	HOW	WHY
인문	공공	부동산 통합정보 제공	• 토지이용계획, 지적도, 개별주택가격, 개별공시지가 등 부동산정보를 통합하여 공간정보를 기반으로 정보를 제공	• 부동산 정책업무 관련 공무원 및 공인중개사, 일반시민	• 공시지가는 분기별 또는 연단위별로 정보 제공 • 실거래가 등의 정보는 거래가 발생하는 경우, 수시 갱신 적용	• 공시지가 및 실거래가 비율 등의 분석을 통해 이상지역을 조기에 발견하여 정책에 활용	• 현재 제공하고 있는 각종 부동산 정보는 필지단위 정보로서 대용량의 공간 빅데이터 기술을 활용하여 기존 정보를 보다 빠르게 처리하고, 이상지역을 자동으로 검색하여 제공
		국토변화 모니터링	• 토지피복도, 수치지형도, 용도지역지구도 등 공공시스템에서 제공하고 있는 전국단위의 대용량 공간정보의 분석을 통해 변화 정보를 모니터링하여 신뢰성 있는 정보를 제공	• 국토변화 관련 정책 및 계획 수립 공무원, 연구기관, 국토 및 도시계획 관련 엔지니어링 업체, GIS및 SI 업체에서 활용 가능	• 모니터링의 목표 및 대상에 따라 분석 단위를 설정하여 분석 할 수 있으며, 보통 전국단위에서는 1년 단위의 시계열 분석 시행	• 공간 빅데이터 연산 및 인메모리 기반 처리모듈을 통해 대용량의 전국단위 국토정보에 대한 접근, 처리, 연산을 지원	• 기존 GIS 기술로 처리가 불가능 하였던 대용량 공간정보를 공간 빅데이터 기술로 처리, 분석함으로써 관련 정책 및 계획 수립에 활용
	민간	환자의료 이용지도	• 국민건강보험 공단의 빅데이터와 각종 공간정보와 통계 DB를 활용하여 의료기관 이용정보를 분류하여 각 주제별 환자 의료이용지도를 구축하여 제공	• 공공부문의 의료보건 계획 수립을 위한 기초자료로 활용하고, 각종 정책의 모니터링에 활용, 민간 부문에서는 의료보건산업 인프라 기반을 구축, 의료부문의 시설입지 컨설팅에 활용	• 일반적인 의료이용지도의 경우 1년 단위로 주제도를 구축하여 제공하며, 이용행태나 특성에 따라 월/분기별로 분석하여 제공	• 대상상병, 의료이용 형태, 환자특성 등의 정보를 구축하여 가공하고, 공간정보와의 매칭 작업을 통해 공간 빅데이터 기술을 활용한 공간분석 및 가시화된 의료이용지도 구축	• 급격한 인구구조 변화 등에 대응한 지속가능한 보건의료 정책을 통한 맞춤형 복지 구현을 위해 공간 빅데이터를 활용한 환자 의료이용지도를 구축하여 의료보건 정책 활용
		IoT기반 의료지도	• 건강보험공단에서 구축한 의료 빅데이터와 공간정보와 각종 웨어러블 IoT에서 구축된 건강관련 정보를 융합하여 분석함으로써 이용자의 특성에 따라 의료시설의 위치 및 시설 등의 각종 정보를 실시간으로 제공하고 필요에 따라 응급이송 등에 활용 가능	• 일반 웨어러블 IoT 디바이스 및 실시간 대응을 요하는 중증환자 등을 대상으로 디바이스를 제공하의료 서비스 제공 가능	• IoT 기반 실시간 정보를 처리하여 기존 구축된 의료정보와 결합하여 이용자의 형태에 따라 각종 의료관련 정보를 제공	• IoT 기기에서 전송되는 실시간 정보를 공간 빅데이터 Interactive 분석 기술을 활용, 복합 이벤트를 처리하여 환자 및 의료시설 이용을 위한 이송 등에 활용	• IoT 기기의 확대보급에 따라 실시간 대응을 요하는 증증환자 및 일반환자들에게 신속한 의료서비스의 제공

분야	부문	비즈니스 모델 영역	비즈니스 모델 개발				
			WHAT	WHO	WHEN	HOW	WHY
		노인여가 복지시설 분석	• 노인 여가복지시설 분포 및 선호활동 분석 지도 구축	• 지역 내 복지센터 관리 및 노인복지 관련 업무 담당자	• 지역 내 거주노인 및 복지센터 분포도는 주기적으로 업데이트하여 상시 제공. 노인복지관 신설 등 필요에 따라 선호여가활동을 조사하여 반영	• 지역 내 거주노인 분포 및 노인대상 복지센터의 분포를 지도화하여 표시. • 빅데이터 분석을 통한 노인 선호여가 정보 제공	• 지역 내 노인의 정서적 건강관리 및 복지지원을 목적으로 시행
		감염병 예측 및 경보를 위한 시스템 구축	• 바이러스, 유행성 질병 등의 이력을 지도화 하고, 감염병 확산 알고리즘을 통한 예측 시스템 구축	• 대한민국 국민 누구나	• 누적정보의 통계분석을 상시제공 • 유행성 질병 및 바이러스 유행 시, 알고리즘을 확산 분석 결과 제공	• 유행성질병 및 신종 바이러스의 이력을 분석하여 결과정보를 제공. 바이러스 확산알고리즘을 통한 질병확산 예측 및 경보 정보 제공	• 국민의 질병 예방 및 전염병 확산 방지를 위한 목적으로 시행
		AI확산 예측 및 예방관리 지도	• 기존의 AI 발생 누적정보 및 철새도래지, 습지 등을 지도에 가시화하여 표현	• 농림축산식품부의 AI 관련 업무부서 등	• 새로 업데이트 되는 정보를 누적정보와 통계하여 주기적으로 업데이트	• 농림축산식품부에서 주기적으로 관리하고 있는 AI관련 누적데이터를 통계분석하여 지도에 가시화 표현	• AI 발생시기에 대비하여 AI가 자주 발생하는 지역의 사전 예방을 목적으로 시행
		스마트 시티와 연계한 스마트홈 서비스	• 각종 정보(센서를 통한 에너지, 물, 생활패턴 등)를 활용하여 스마트 시티와 연계하여 생활에 필요한 각종 기설 등에 대한 체계적인 공급이 가능	• 스마트 시티를 운용하는 공무원 및 기업에서 활용 가능하며, 이를 통해 일반 가구(시민)에서 각종 생활 패턴에 대한 다양한 정보를 모니터링 할 수 있음	• 스마트홈 및 스마트 시티에서 생성되는 정보를 실시간으로 수집하여 처리・분석하여 직접 대응	• 스마트홈에서 제공되는 각종 정보와 스마트 시티(통합관제센터)와 연계하고, 공간 빅데이터의 Batch 및 Interactive 분석기술을 적용하여 적정 시설의 밀도 및 사용량 등을 실시간으로 관측하여 공급 가능	• 기존 공공에서 일방적으로 제공되었던 각종 시설 및 공공서비스에서 벗어나 스마트홈을 통해 제공 되는 다양한 정보를 기반으로 맞춤형 공공시설 및 서비스 제공함으로써 비용 및 인원감축 가능
사회 가치	공공	스마트 사회 안전망 구축	• 국토지리정보원이 구축 및 제공하는 격자데이터를 기반으로, 모바일, 웹의 소셜미디어를 통하여 생성되는 데이터를 분석, 재난/재해및 응급상황에 신속하게 대처하기 위한 네트워크 구축	• 지자체의 소방, 방범, 치안을 책임지는 지방경찰청, 지구대 , 소방서등의 모니터링 관계자	• 지속적으로 안전 취약지역 분석을 위한 데이터를 수집 및 모니터링, 상황발생시 관련정보를 대응시설에 신속하게 전달 및 공유	• 범죄/치안 등이 빈번하게 발생하는 지역을 분석하여 재발방지를 위한 대응을 하는데 있어 대응 우선지역을 선별하기 위한 기준으로 분석정보를 활용, 상황발생시 대응시설에 신속하게 정보를 전달 및 정보공유체계 구축	• 안전취약지역과 관련된 정보를 수집 분석하여 미래에 발생할 위험을 선제적으로 방지 및 대응하기 위한 정보 수집, 분석, 대응체계 마련이 필요
		교통 안전망	• 교통안전부문, 교통수요·공급부문, 교통융합부문 3부	• 교통시설의 운영 및 관리를 통하여 서비스 시나리	• 디지털 운행기록장치(DTG) 데이터와 교통사고데이터	• 도로네트워크를 기준으로 디지털 운행기록장치	• 교통사고 위험지역 분석에 있어서 운전자, 운행기

분야	부문	비즈니스 모델 영역	비즈니스 모델 개발				
			WHAT	WHO	WHEN	HOW	WHY
		구축	문으로 구분하여 교통 빅데이터 융·복합 활용을 위한 대표적인 서비스 시나리오 제시	오와 연관된 교통 빅데이터를 구축 및 관리하는 기관을 공급처로 설정, 교통 빅데이터의 분석을 통하여 교통시설 또는 관련 서비스의 운영 및 관리의 효율화를 지원하거나 분석결과의 융·복합 활용을 지원하는 공공/민간부문을 수요처로 설정	간의 상관관계분석 및 판별분석을 통하여 교통사고의 공간적 패턴(대중교통/화물차/보행자 교통사고 위험지역)과 시간적 패턴(교통사고 발생 시간/요일 등)을 분석, 분석결과를 토대로 교통사고 선제대응을 위한 최적의 단속지역과 시간을 제시	(DTG) 데이터와 교통사고 데이터를 저장·관리·분석하고, 분석결과와 서비스 사용자의 운행기록 데이터를 융·복합하여 운전자에게 사용자 맞춤형 안전운전지도를 제공하는 플랫폼 서비스 제시	록, 교통 사고 데이터 간의 정확한 Matching에 대한 기술적 애로사항이 존재하여 이를 공간 빅데이터 기술구현성과로 해결
		화물안전 운송지도	• 국토부는'13년부터 화물운송실적관리시스템을 통해 화물자동차 운수사업자의 화물운송실적을 관리하고 있음 • 위험물 안전운송을 위하여 공간 빅데이터의 이동객체, 실시간 이동객체 분석 모듈을 활용하기 위한 방안 제시	• 도로교통안전공단, 경찰청 화물운송안전지도 관계자	• 위험물 운송차량의 사고정보를 주기적으로 수집, 분석하여 사고방지를 선제적으로 대응하기 위한 모델 개발	• 복합 이벤트 처리분석을 통한 위험물 운송관리(안전한 위험물 운송을 위해 SNS, 도로정보, 센서정보 등을 활용한 복합 이벤트 처리분석 지원)	• 위험물 운송차량의 사고는 대형사고로 이어지며, 사고인근지역에 피해를 줄 수 있기 때문에, 사고를 미연에 방지하거나 최소화하기 위한 방안이 필요함
		대중교통 취약지역 지원을 위한 버스노선/택시운행 최적화 분석 서비스	• 운행기록분석시스템(eTAS)의 디지털 운행기록(DTG) 데이터를 대상으로 버스운행과 택시운행에 대한 시·공간적 승하차 패턴을 분석하여 대중교통 취약지역을 분석, 버스노선 연장 또는 추가 신설과 택시운행 최적화 분석을 수행	• 대중교통시설 운영·관리기관, 버스/택시 운수업체	• 버스운행과 택시운행에 대한 시·공간적 승하차 패턴을 분석하여 대중교통 취약지역을 분석, 최적화된 버스노선 경로와 운행시간을 제시, 택시 공차시 버스노선 서비스 지역과 중첩되지 않으면서 택시운행을 최적화할 수 있는 택시 서비스 지역 및 최적 경로 제시	• 교통카드 정산소의 정보수집을 위한 교통 빅데이터를 활용하여 대중교통 이용 인구밀도 Map 작성, 상권분석을 위한 유동인구Map으로 활용, 시간대별 대중교통 Accessibility Map을 구축하여 대중교통 취약지구를 추출, 취약인구를 분석하는데 활용	• 교통 빅데이터를 활용하여 대중교통수요와 공급을 매칭시켜, 대중교통 취약지역에 거주하는 시민의 불편 해소
		대한민국 관광지도	• 전국의 여행정보를 제공하고, 여행지에 대한 여러 사람의 감성분석 결과를 함께 보여줌으로써 여행지에 대한 다양한 정보 제공	• 대한민국 국민 누구나	• 관광지에 대한 정보는 상시 제공하고, 블로그나 SNS를 통한 감성분석 결과는 주기적으로 업데이트하여 제공	• 전국지도에 관광지역을 표시하고, 사용자 선택 시 사진, 여행정보, 감성분석 결과 등의 정보를 메타데이터로 함께 제공	• 우리나라의 관광지역에 대한 다양한 정보를 제공하고, 관광산업의 활성화를 목적으로 시행
		온라인 모바일	• 온라인(포털사이트, 커뮤니티), SNS(페이스북, 트위	• 온라인 및 보바일 기기를 이요하는 일반시민들이	• 온라인에서 생성되는 각종 정보는 필요에 따라 실	• 온라인 및 SNS 등에서 언급되는 온라인 범죄관련	• 정보통신 기술의 발전 및 이용량 증가에 따른 온라

분야	부문	비즈니스 모델 영역	비즈니스 모델 개발				
			WHAT	WHO	WHEN	HOW	WHY
		범죄예방을 위한 정보제공	터, 인스타그램) 등에서 언급되는 온라인 범죄가 테깅되어 있는 정보를 수집하여 이를 범죄별로 유형화하고, 빈도수 등을 도출하여 지도에 표출하여 관련 정보와 함께 제공	직접적인 사용자로서 온라인 범죄를 사전에 예방하고, 공공부문인 수사기관에서는 이를 바탕으로 인력배치 정보 추적 등이 가능	시간으로 수집하여 처리·분석하며, 대용량의 데이터는 일/월/분기/년 단위로 분석하여 지도화	정보를 수집하고, 이를 유형별(보이스피싱, 온라인 사기 등)로 구분하여 빈도를 분석하여 지도화하여 제공	인 범죄증가로 피해건수 및 액수가 급증가하고 있어 사전 예방을 위한 정보제공과 수사를 위한 기초자료로서 공간 빅데이터 기술을 활용하여 제공
경제	공공	공간정보 사업체 지도	• 기초자치별, 광역자치별 공간정보산업의 사업체수, 종사자수, 품목별 수출입 통계를 분석하여 산업동향분석 결과를 Map과 시계열 그래프로 시각화 서비스 제공	• 공간정보관련 취업생, 공간정보산업진흥관련 업무담당자, 공간정보산업정책 개발자	• 공간정보와 관련된 사업체정보 또는 일자리 정보를 지속적으로 모니터링하여 정보 수요자에게 신속하게 전달 및 공유	• 공간정보 산업통계지도: 공간정보기술사전을 기반으로 공간정보기술과 연관된 표준산업분류코드를 도출, 관련 분류코드를 기준으로 시계열별로 지역별 사업체수와 종사자수 통계를 지도화하여 제공	• 국내 공간정보 산업의 동향을 파악하고, 관련 산업진흥을 위한 정책수립에 공간정보 사업체지도가 필요함
	민간	공간정보 산업/기술/무역/직업 동향정보 제공	• 국내 공간정보 산업의 진흥을 위하여 관련된 정보를 키워드를 기준으로 구축, 공간정보 기술/산업/무역/직업에 대한 정보접근성 향상 지원	• 공간정보관련 취업생, 공간정보산업진흥관련 업무담당자, 공간정보산업정책 개발	• 새롭게 등장하는 공간정보와 관련된 신기술의 키워드를 지속적으로 모니터링 및 수집하고, 관련 산업, 기술, 무역, 직업코드와 연계하여 정보 수요자에게 신속하게 전달 및 공유	• 공간정보기술 용어사전을 구축하고, 이를 기준으로 산업분류, 기술분류, 무역품목분류, 직업분류 항목들과의 연계성 분석을 통해 공간정보와 연관된 분류코드를 도출, 도출된 분류코드에 해당되는 통계정보를 공간정보를 기반으로 시각화 하여 제공	• 새롭게 등장하는 공간정보관련 신기술에 대한 정보 접근성을 향상시키고, 관련 산업/기술/무역/직업 간의 통합된 이해를 제공하여 공간정보산업진흥에 기여
환경	공공	홍수위험 지도	• 홍수에 대비하여 과거 침수피해지역을 지도상에 표현하고, 강수량 및 하천수위에 따라 3차원 형식으로 침수위험지역을 시뮬레이션. IoT기술을 통한 실시간 위험감지정보 제공	• 과거 침수피해 경험이 있거나 침수피해가 예상되는 지역(지자체 단위)의 재해관련 업무 담당자	• 태풍, 집중호우 시기에 대비하여 수위에 따른 3D 침수위험지역 정보 및 IoT 기술을 통한 실시간정보를 상시적으로 제공	• 하천단위로 강수량, 하천수위, 침수피해 실적 등의 정보를 통계 분석하고, 수위별 침수지역을 3D 형태로 시뮬레이션하여 직관적인 침수위험지역 정보 제공	• 태풍, 집중호우에 대비하여 침수피해지역의 사전 피해예방을 목적으로 시행
		건물 에너지 통합 모니터링	• 전국 690만여개의 건축물 정보와 에너지원별 사용정보를 수집하여 건물에너지 통합 DB를 구축하여	• 국민, 실무자, 공무원 등 공공·민간 모든 부문에서 활용가능하며, 감축목표수립 및 달성을 위한 정	• 월 단위의 에너지 사용량 정보와 건축물 정보를 연계하여 관련 정보를 제공할 수 있으며, 건축물 정	• 건축물정보, 공간정보(건물, 도로, 공통)와 에너지 사용정보(전기, 도시가스, 지역난방 : UFID 활용)을	• 건축물 및 에너지 정보는 공간속성을 갖는 빅데이터로서 공간 빅데이터 기술을 활용하여 저장·관

분야	부문	비즈니스 모델 영역	비즈니스 모델 개발				
			WHAT	WHO	WHEN	HOW	WHY
			건물부문에서의 온실가스 감축을 위한 정보 제공	책수립 지원, 통계분석, 대민포털 등으로 제공 가능	보의 주기적인 갱신을 통한 적용이 요구됨	활용하여 건물에너지 통합 DB를 구축하여 건물부문에서의 에너지소비증명, 정보공개, 계획 및 목표관리제 지원, 인증 등에 활용	리 할 수 있으며, 이를 통해 효과적인 온실가스 감축이 이루어질 수 있음
		기존 건축물의 에너지 사용량 및 녹색 건축물 보급지도	• 전국 690만여개의 건축물 정보와 에너지원별 사용정보를 수집하여 건물에너지 통합 DB를 구축하여 건물부문에서의 온실가스 감축을 위한 정보 제공	• 국민, 실무자, 공무원 등 공공·민간 모든 부문에서 활용가능하며, 감축목표수립 및 달성을 위한 정책수립 지원, 통계분석, 대민포털 등으로 제공 가능	• 월 단위의 에너지 사용량 정보와 건축물 정보를 연계하여 관련 정보를 제공할 수 있으며, 건축물 정보의 주기적인 갱신을 통한 적용이 요구됨	• 건축물정보, 공간정보(건물, 도로, 공통)와 에너지사용정보(전기, 도시가스, 지역난방 : UFID 활용)을 활용하여 건물에너지 통합 DB를 구축하여 건물부문에서의 에너지소비증명, 정보공개, 계획 및 목표관리제 지원, 인증 등에 활용	• 건축물 및 에너지 정보는 공간속성을 갖는 빅데이터로서 공간 빅데이터 기술을 활용하여 저장·관리 할 수 있으며, 이를 통해 효과적인 온실가스 감축이 이루어질 수 있음
		토지이용 및 토지이용 변화와 산림 (LULUCF) 지도	• LULUCF(산림, 농경지, 초지, 습지, 정주지, 기타토지) 각 부문의 변화량을 전국단위에서 추적, 분석하여 제공함으로써 국가 온실가스 인벤토리 통계 구축 지원	• 국무총리실(국가온실가스 종합정보센터) 및 각 카테고리별 관장기관 및 산정기관에서 면적 및 전용정보를 추적하는데 활용 가능	• 매년 단위 각 카테고리별 유지되는 면적 및 전용되는 면적의 추적이 가능하여야 하며, 2017년 현재의 통계 인정을 위해서는 과거 20년 부터의 변화정보가 추적되어야 함	• 전국단위의 토지이용 변화 및 전용정보를 분석하기 위해서는 기존 GIS 소프트웨어 및 기술에 한계가 있으며, 위성영상 및 토지피복도 등의 공간정보를 대상으로 공간빅데이터의 분산처리 기술을 적용하여 변화정보를 추적하여 제공	• 현재 국가 온실가스 인벤토리 중 LULUCF 부문의 변화량을 추적하지 못하여 국제사회에서의 공식 통계로 인정받지 못하고 있어, 이러한 토지이용 변화정보를 체계적으로 구축하여 제공함으로써 국제사회에서의 인정 가능
	민간	신재생 에너지 확보를 위한 기반지도	• 과학기술정보통신부에서 제공하는 신재생에너지 자원지도는 전국을 대상으로 태양, 풍력, 수력, 바이오매스, 지열로 구분하여 100*100m로 제공되며, 이를 기반으로 주제도 및 시계열분석 및 공간분석을 통한 자원량 평가에 활용	• 신재생에너지의 기초정보로 연구개발에 활용, 신재생 관련 기업의 사업평가 및 타당성 조사 비용 절감	• 신재생에너지지도는 에너지원의 특성에 따라 분기별로 적용하여 주제도화하고, 시계열별 변화분석 결과를 제공	• 각 에너지별로 인공위성, 기상 및 해상, 지형, 토지피복도, 바이오매스별 기타통계, 강수량 관측자료 등 대용량의 다양한 정보를 대상으로 공간정보와 융합된 분석을 위해 공간빅데이터의 수집저장 기술을 활용하여 가시화	• 전지구적 기후변화에 대응하기 위해 기존 화석연료 사용량의 감축과 신재생에너지 확대를 위해 각 지역별 활용가능한 신재생에너지 자원지도 구축을 통해 효율적인 신재생에너지 활용이 가능
		가정 상업 GHG	• 가정, 상업 건축물 정보와 에너지원별 사용정보를 수집하여 통합 DB를 구축	• 국민, 실무자, 공무원 등 공공·민간 모든 부문에서 활용가능하며, 감축목	• 월 단위 에너지 사용량 정보와 가정, 상업 건축물 정보를 연계하여 관련 정	• 가정, 상업 건축물 공간정보(건물, 도로, 공통)와 에너지사용정보(전기, 도시	• 가정, 상업 건축물 및 에너지 정보는 공간속성을 갖는 빅데이터로서 공간

분야	부문	비즈니스 모델 영역	비즈니스 모델 개발				
			WHAT	WHO	WHEN	HOW	WHY
		배출지도	하여 가정, 상업 건축물에서의 GHG 배출량 지도 구축	표수립 및 달성을 위한 정책수립 지원, 통계분석, 대민포털 등으로 제공 가능	보를 제공할 수 있으며, 가정, 상업 건축물 정보의 주기적인 갱신을 통한 적용이 요구됨	가스, 지역난방 : UFID 활용)을 활용하여 가정, 상업 건물에너지 통합 DB를 구축하여 에너지소비증명, 정보공개, 계획 및 목표관리제 지원, 인증 등에 활용	빅데이터 기술을 활용하여 저장·관리 할 수 있으며, 이를 통해 GHG 배출량 지도 구축 가능
		소음 발생지도	• 생활주변에서 발생하는 소음원의 위치와 그에 따른 피해지역의 현황지도 구축	• 지자체의 생활민원 및 주민복지 관련 업무담당자	• 과거 지역 내 소음으로 인한 피해지역정보는 상시 제공, 새로운 소음피해 민원 발생시 또는 IoT 기술을 통한 소음원 확보시 분석시행, 결과정보 제공	• IoT 기술을 통한 소음발생 지역 위치 정보 획득 및 지도에 표시, SNS를 통해 피해정보 유추하여 지도작성하여 피해확산 등을 분석	• 소음으로 인한 주민생활의 피해를 줄이고, 주민의 삶의 질 향상을 위한 목적으로 시행
		원자력 안전관리 지도	• 현재 제공 중인 원전정보 및 실시간 방사능 감지정보에 사고고장 누적정보, 지진발생 정보 등을 추가하여 원자력안전관리지도 구축	• 대한민국 국민 누구나	• 현재 전국 방사능 감지 정보는 상시 제공 중이며, 그 와 더불어 위험지역분석 정보도 주기적으로 업데이트하여 상시 제공	• 전국 원전의 위치 및 개별 세부정보 제공. 지진발생 위치의 누적정보와 도심지 및 거주민 거리 정보 등과 중첩분석을 통해 위험분석 및 결과 정보 제공	• 잦은 지진의 발생, 노후한 원자력 시설로 인한 원자력시설물의 붕괴 위험을 예방하고자 하는 목적으로 시행
		미세먼지 취약지도	• 현재 제공 중인 실시간 미세먼지 농도정보에 원인별, 시기별 미세먼지 누적정보를 추가하여 미세먼지취약지도 구축	• 대한민국 국민 누구나	• 연도별, 계절별, 월별, 시간대별 미세먼지 농도 누적정보 및 병원 환자수와의 상관관계, 중국발 미세먼지 영향분석 결과 등 상시공개	• 16년간의 누적결과를 통계분석하여 시기별, 시간별 미세먼지 취약지역을 지도화하고, 지역별 병원환자수와의 상관관계, 중국발 미세먼지 영향분석 결과를 제공	• 국민의 건강관리 및 쾌적한 생활 환경조성을 목적으로 시행
		폐기물 정보 시스템	• 지역의 에너지 관리 및 온실가스 저감을 위해 폐기물 발생량, 재활용 비율 등의 정보 구축	• 지자체의 폐기물관리 업무담당자	• 주간단위로 폐기물 발생량 및 수거량 등의 정보를 업데이트하여 제공	• 지역 내 폐기물 발생량을 위치별로 제공하고, 폐기물 용량의 지역별 원인별 분석, 재활용 비율 분석, 에너지화 방안 등의 결과정보 제공	• 지자체 내의 지역별 폐기물 발생량 파악, 지역의 재활용 및 에너지 관리 목적으로 시행
		환경변화 입체관측	• UAV, 드론등을 활용하여 원격으로 수집된 공간정보를 신속하게 분석하여 환경변화분석 및 대응하	• 발전환경입지계획원, 농업환경생태연구원, 생태계복원관리연구원, 지역환경정책연구원	• UAV, 드론 등을 통하여 인편으로 접근하기 어려운 지역의 환경오염을 원격으로 모니터링	• UAV, 드론에 탑재된 센서에서 송출되는 모니터링 정보를 이동통신망을 활용하여 원격으로 수신받	• 인편으로 접근하기 어려운 지역과 특정지점의 환경오염을 모니터링하여 환경오염에 신속하게 대

분야	부문	비즈니스 모델 영역	비즈니스 모델 개발				
			WHAT	WHO	WHEN	HOW	WHY
			는 기술개발			아 환경오염의 정도를 공간정보 기반으로 분석하여 모니터링	응하기 위한 목적으로 수행
정치	공공	북한의 공간변화 모니터링	• 기존에 구축된 각종 북한 지역 공간정보를 기반으로 용도지역 등 북한의 공간적 변화를 파악하는 모니터링 자료 구축	• 통일부, 국방부, 국토부 등 관련부처의 북한지역 정보관리 업무 담당자	• 위성영상 등을 통해 북한 지역의 공간변화를 주기적으로 모니터링	• 기존에 구축되어있는 공간정보 자료를 활용하여 기존현황을 파악하고, 위성영상을 통해 변화 상황 파악	• 향후 통일에 대비하여 북한지역의 공간변화를 파악하고자 하는 목적으로 시행
기술	공공	IoT기반 공간상황 인식 및 사물정보 기술	• 텍스트, 센서를 포함한 다양한 공간정보 기반의 스트림데이터를 격자단위로 수집하여 특정한 이벤트가 발생할 때, 신속하게 관련지역의 격자를 파악하고 관련상황에 대응	• 지자체의 소방, 방범, 치안을 책임지는 지방경찰청, 지구대 , 소방서등의 모니터링 관계자	• 특별한 응급상황이 발생하거나, 지역적 사건이 발생했을 때,관련 소셜데이터, 센서데이터를 격자를 기반으로 수집 및 분석하여 응급 상황의 공간적 패턴을 신속하게 파악및 대응	• 교통카드이용현황, CCTV 영상정보 등 센싱정보 수집체계를 기반으로 교통혼잡도 분석, 실시간 재난대응체계, 대중교통 구간획정 등 지능형 공간정보 분석모델에 적용할 수 있는 공간 빅데이터 분석기능 도출	• 재난/재해 및 응급사태가 발생했을 때 발생지점을 격자를 기반으로 찾고,그 변화의 패턴을 신속하게 파악하는 것이 필요
기술	민간	장래 교통망 변화에 따른 지역변화 분석	• 도로망 GIS DB와 교통 분석용 네트워크를 활용하여 장래년도의 교통망 변화 시나리오를 설정, 디지털 운행기록(DTG) 데이터를 기준으로 장래년도의 교통망 변화 시나리오에 적용했을 때 각 교통존별 접근성(네트워크 거리의 OD Matrix) 변화를 비교분석하여 영향력의 지역적 변화를 시·공간적으로 시각화하여 제공	• 지방국토관리청, 한국도로공사, 지자체, 공공연구기관, 지역 부동산 컨설팅 업체	• 장래 도로망확충이 확정된 지역에 거주하는 시민, 지역 공무원이 도로망 확충시 변화하는 접근성 등에 대한 정보를 제공하여 관련 사업 및 정책추진을 위한 이해를 제공	• 도로망 확충에 따른 접근성의 변화를 교통존을 기준으로 시·공간 시각화 및 비교분석 실증, 디지털 운행기록장치의 데이터를 활용하여 시계열별 도로망 확충에 따른 통행경로의 변화를 시뮬레이션 및 비교분석 실증	• 도로망 확충을 위한 다양한 사업들이 추진되고 있으나, 사업완료 후 교통접근성 향상 등과 같이 지역적 변화에 따른 정보를 시민들이 알기 쉽게 전달하기 위한 수단이 부족함

⑤ [단위성과 J-6] 공공/민간부문 공간 빅데이터 활용을 위한 서비스 시나리오 개발(4·5차년도)

○ 서비스 시나리오는 연구단 개발 기술과 공간 빅데이터 플랫폼에서 연계 활용 및 분석이 가능한 데이터 유형을 검토하여 어떤 비즈니스 모델을 적용하더라도 범용적으로 활용 가능한 서비스를 도출할 수 있도록 서비스 유형을 재구분함

- 토지건물정보 기반 서비스 : 국토공간정보의 최소단위이자 기본이 되는 단위인 필지 및 건물정보를 활용하여 공간 빅데이터 분석을 통해 플랫폼에서 제공 가능한 서비스 유형
- 비정형 텍스트정보 기반 서비스 : SNS, 뉴스, 블로그 등 비정형 텍스트 정보를 활용하여 공간 빅데이터 분석을 통해 플랫폼에서 제공 가능한 서비스 유형
- 네트워크정보 기반 서비스 : 지하철, 버스, 택시, 화물차 등 각종 교통수단 및 도로정보를 활용하여 공간 빅데이터 분석을 통해 플랫폼에서 제공 가능한 서비스 유형



<그림 2-352> 공간 빅데이터 서비스 유형 구분

㉮ 토지건물정보 기반 서비스

○ 서비스 목적

- 정부는 2013년부터 공공데이터 개방 정책을 추진해오고 있으며, 이에 따라 다양한 형태의 공공정보가 개방되어 다양한 부문에서 이를 활용한 가치 창출이 이루어지고 있음
- 개방되는 공공데이터는 종류가 점점 다양해지고 있으며, 공간해상도가 높아지고 있어 이를 활용·처리하기 위한 기술개발이 요구됨
- 이에 따라 공공데이터의 기본적인 정형데이터 형식을 지니며 국토공간정보의 최소단위이자 기본이 되는 필지 및 건물 데이터를 활용하여 연구단 개발 기술 적용함으로써 범용적으로 활용이 가능한 서비스 시나리오를 개발하고자 함
- 본 연구에서는 필지기반의 공시지가와 에너지사용량 데이터를 활용하여 연구단에

서 개발한 분석모듈을 적용하여 결과 도출까지의 과정을 시나리오 주제로 선정함

○ 서비스 사용자 분석

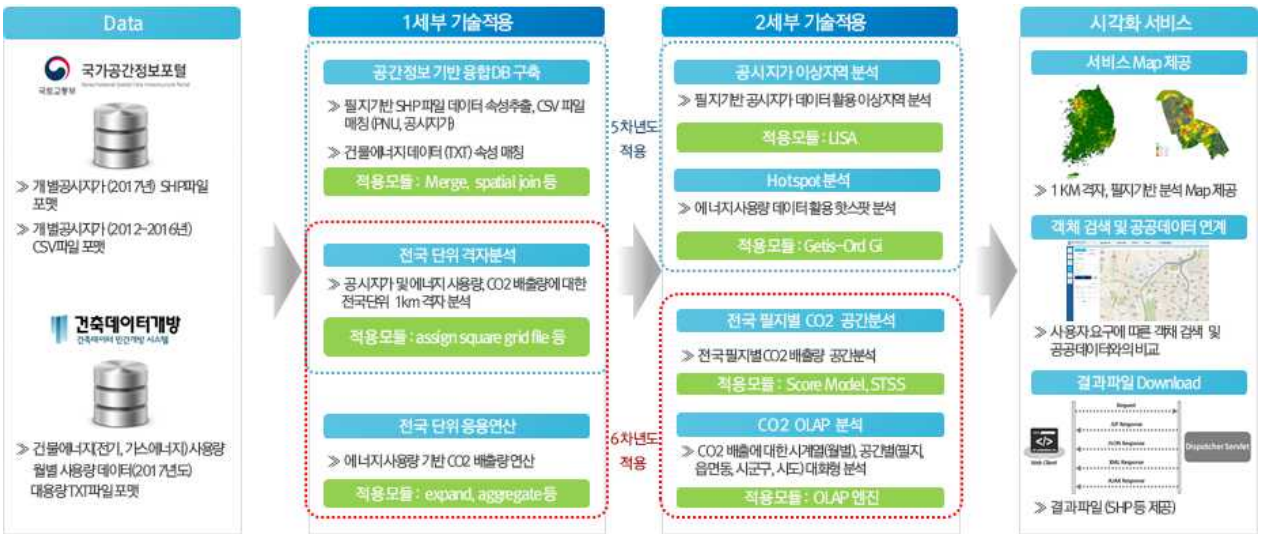
- 토지건물정보 기반 서비스를 활용하는 사용자를 기술숙련도에 따라 일반사용자와 전문사용자로 구분하고, 소속 및 주체에 따라 이를 다시 공공과 민간으로 구분하여 서비스 사용자를 도출함

<표 2-102> 토지건물정보 기반 서비스 사용자 분석

기술 숙련도	소속기관		서비스 사용자	서비스 사용분야
일반 사용자	공공	일반 행정기관	행정기관 공무원	• 국토·도시계획, 부동산, 건물에너지 관련각종 계획수립 및 운용 담당 일반 공무원 • 부동산, 건물에너지 등과 관련된 업무 담당 일반 공무원
		민간	일반국민	• 일반시민
전문 사용자	공공	일반 연구기관	기본정보활용 연구원	• 인문, 사회, 환경 등 다양한 연구 분야에서 공시지가, 건물에너지 정보를 확인 및 활용하고자 하는 일반 연구원
		전문 연구기관	상세정보활용 연구원	• 부동산 및 에너지관련 계획 및 정책수립 관련 전문기관의 연구원
	민간	일반기업	일반기업	• 일반 엔지니어링 업체, 부동산 관련 업체 등
		전문기업	국토변화 관련사업 기업	• 국토, 도시계획 관련 엔지니어링 업체, GIS 및 SI 관련 기업

○ 서비스 시나리오 개요

- 정부에서 개방하고 있는 대용량의 공공정보인 개별공시지가 및 에너지사용량 데이터를 활용하여 1세부와 2세부에서 개발한 분석 기술을 적용하여 다양한 분석결과를 도출함



<그림 2-353> 토지건물정보 기반 서비스 시나리오

○ 활용데이터

- 국토교통부 국가공간정보포털에서 제공하고 있는 개별공시지가 데이터와 건축데이터 민간개방시스템에서 제공하는 에너지사용량 데이터 활용

- 융합DB 구축을 위해 활용되는 개별공시지가 데이터는 2017년 SHP파일 포맷 개별공시지가 데이터와 2011~2016년 CSV파일 포맷 형태의 개별공시지가 데이터임
- 건물에너지 사용량 데이터는 2017년 월별 전기, 가스 사용량에 대한 TXT파일 포맷 데이터임

<표 2-103> 토지건물정보 기반 서비스 시나리오 활용데이터

항목	내용	파일 포맷	데이터 용량	출처
Cadastral Official Price (개별공시지가)	17개 시도별 공간정보 데이터 (2017년도 기준)	SHP	31.7GB	국토교통부 국가공간정보포털
Cadastral Official Price (개별공시지가)	17개 시도별 개별공시지가 (2011~2016년)	CSV	17.1GB	
Gas Usage (가스사용량)	2017년 월별 가스사용량	TXT	444MB	건축데이터 민간개방시스템
Electricity Usage (전기사용량)	2017년 월별 전기사용량	TXT	1.92GB	

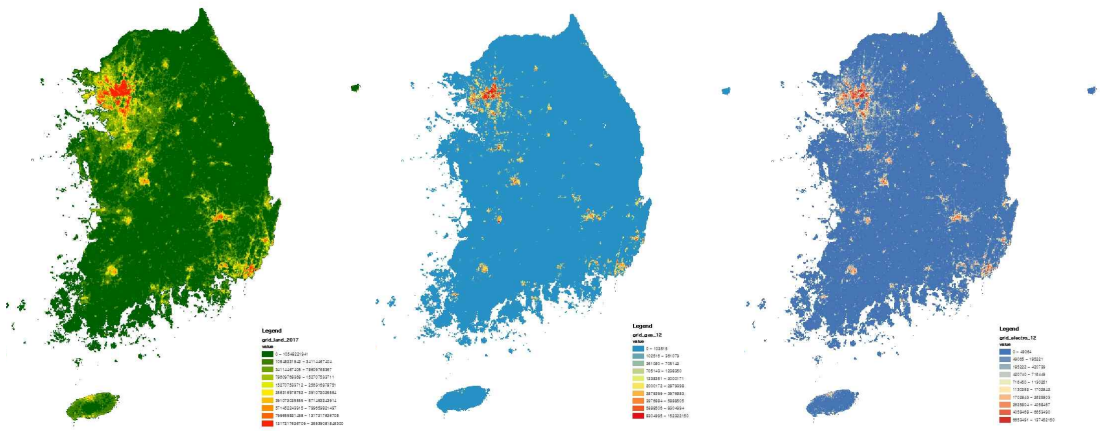
○ 1세부 개발 기술 적용

- 공간정보 기반 융합DB 구축

- 1단계 : 1세부 개발 기술 중 Merge 기능을 활용하여 시도단위로 나뉘어 있는 SHP 파일 형태의 2017년 개별공시지가 데이터를 전국단위의 한 파일로 생성
- 2단계 : 1세부 개발 기술 중 Spatial Join 기능을 활용하여 SHP파일 형태의 개별공시지가 데이터에 포함된 필지번호(PNU)를 기준으로 CSV파일 형태의 연도별(2011~2016) 개별공시지가 데이터와 TXT파일 형태의 가스·전기에너지 사용량 데이터를 가져와 필지단위 공간정보 기반의 융합DB를 생성
- 최종 생성된 데이터는 SHP파일 형태이며, 필지단위로 2011~2017년 개별공시지가 정보와 2017년 월별 전기사용량과 가스사용량 정보를 포함하고 있음

- 전국단위 격자분석

- 공간정보 기반 융합DB 구축으로 최종 생성된 파일은 대용량 데이터로 한 번에 화면에 띄어 확인하기가 어려우므로 전국을 1km 격자로 나누어 해당되는 값을 격자로 표현
- 1세부 개발 기술 중 Assign squaregrid 모듈 적용



<a> 개별공시지가(2017년) <b> 가스사용량(2017년 12월) <c> 전기사용량(2017년 12월)

<그림 2-354> 전국단위 격자분석 결과

○ 2세부 개발 기술 적용

- 공시지가 이상지역 분석
  - 융합DB의 개별공시지가 정보와 서울시 강남구의 유동인구 데이터를 활용하여 서울시 강남구의 공시지가 이상지역을 도출함
  - 2세부 개발모듈 중 LISA 분석 모듈을 적용
- 에너지사용량 군집패턴 분석
  - 융합DB의 가스 및 전기에너지 사용량에 대한 군집 패턴을 분석하여 에너지 사용량이 많은 Hot Spot 지역 도출
  - 2세부 개발모듈 중 Getis-Ord Gi* 분석 모듈 적용



<a> 공시지가 이상지역 분석 결과 <b> 가스에너지사용량 군집 분석 결과 <c> 전기에너지사용량 군집 분석 결과

<그림 2-355> 공시지가 이상지역 및 에너지 사용량 군집패턴 분석 결과

㉯ 비정형 텍스트정보 기반 서비스

○ 서비스 목적

- 공공과 민간에서 매일 생성되는 많은 양의 비정형 텍스트 데이터는 다양한 사람의 의견을 포함하고 있어 빅데이터 분석에 유용한 정보를 제공함
- 따라서 비정형 텍스트 데이터를 활용하여 공간정보와 융·복합을 통해 새로운 가치를 발굴하고 다양한 업무에 활용할 수 있는 기술지원 체계가 요구됨
- 이에 따라 민원이나 SNS, 뉴스, 블로그 등 비정형 형태의 텍스트 데이터를 활용하여 연구단의 개발 기술을 적용함으로써 범용적으로 활용이 가능한 서비스 시나리오를 개발하고자 함
- 본 연구에서는 민원데이터와 뉴스데이터를 활용하여 서울시 도봉구 내 소재하는 공원에 대한 선호도 분석 및 감성분석 결과를 도출하는 일련의 과정을 시나리오 주제로 선정함

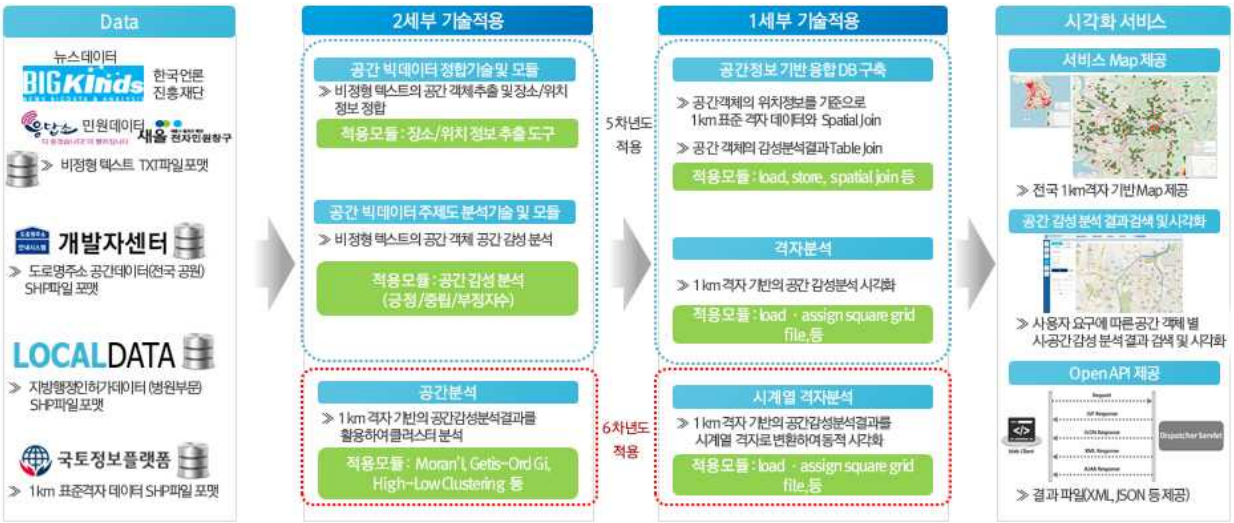
○ 서비스 사용자 분석

- 서울시 도봉구 내 소재하는 공원에 대한 선호도 분석 결과를 알고 싶은 사람 누구나 서비스 사용자가 될 수 있음

○ 서비스 시나리오 개요

- 민원데이터와 뉴스데이터 등 비정형 텍스트 정보와 정부에서 개방하고 있는 공간

데이터를 활용하여 융합 DB를 구축하고, 1세부와 2세부에서 개발한 분석 기술을 적용하여 분석결과를 도출함



<그림 2-356> 비정형 텍스트정보 기반 서비스 시나리오

❍ 활용데이터

- 서울시 및 도봉구의 민원게시판에 게시된 민원데이터와 국토교통부 국가공간정보포털 및 국토정보플랫폼에서 제공하고 있는 공간데이터 활용

<표 2-104> 비정형 텍스트정보 기반 서비스 시나리오 활용데이터

항목	내용	파일포맷	데이터용량	출처
비정형 텍스트 데이터	도봉구 민원데이터 공간 빅데이터 사업단 뉴스데이터 빅카인즈 뉴스데이터	TXT CSV	1.74GB	도봉구홈페이지 국토부 공간 빅데이터 사업단 한국언론진흥재단 빅카인즈
전국 공원 데이터	공원데이터	SHP	54.7MB	도로명주소 개발자 센터
전국 1km 표준격자 데이터	1km 표준격자 데이터	SHP	110MB	국토정보지리원 국토정보플랫폼

❍ 2세부 개발 기술 적용

- 공간 빅데이터 정합기술을 통한 장소 위치 정보 추출

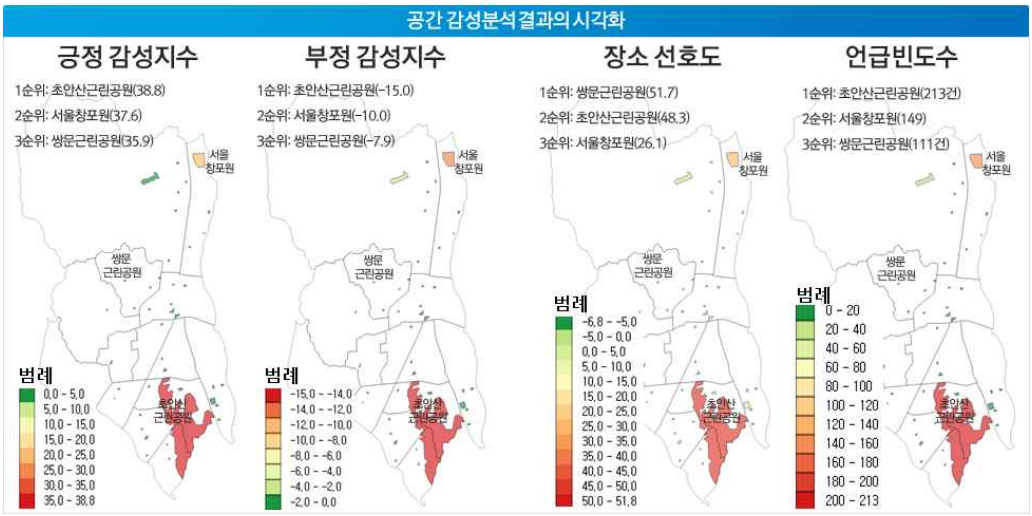
  • 1단계 : 민원데이터 리스트에서 서울시 도봉구 내 소재하고 있는 공원데이터의 위치정보와 공원명칭을 이용하여 도봉구 내 공원관련 민원 추출

  • 2단계 : 추출한 공원관련 민원 중에서 민원 내용에 포함된 공원의 명칭 또는 주소정보, 좌표정보를 활용하여 해당공원의 위치와 민원정보를 매칭

- 공간 감성 분석

  • 해당공원의 위치와 매칭된 민원정보를 2세부에서 개발한 감성분석 모듈에 적용하여 공간 감성 분석을 수행하여 공원별 긍정/중립/부정 지수를 지도상에 표현

  • 도봉구 내 공원에 대한 감성분석 결과와 도봉구 공원 현황도 간의 위치정보를 이용하여 Table Join을 수행하여 공원 감성 분석맵 도출



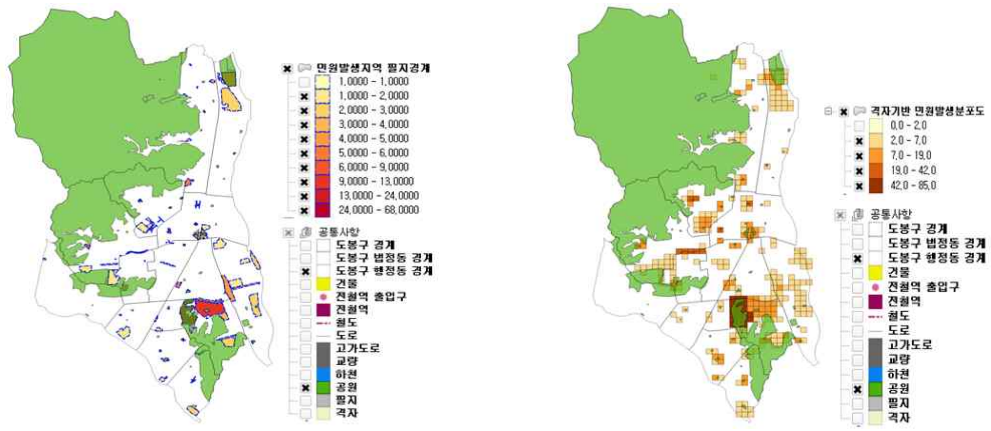
<그림 2-357> 도봉구 공원의 공간 감성 분석 결과

○ 1세부 개발 기술 적용

- 격자분석

• SHP파일 형태의 도봉구 공원 감성분석 결과맵과 전국 표준격자 데이터를 중첩하여 도봉구의 공원 감성분석 맵을 격자로 표현

• 1세부 개발 기술 중 Assign Squaregrid 모듈 적용



<a> 필지기반의 민원발생 분포도 시각화 <b> 격자기반의 민원발생 분포도 시각화

<그림 2-358> 공원 감성분석맵의 격자분석 결과

㉰ 네트워크정보 기반 서비스

○ 서비스 목적

- 4차산업에 따른 환경변화와 정부의 공공데이터 개방 정책 등으로 ICT 기술융합 신산업이 유망사업으로 떠오르고 있으며, 특히 교통 분야의 경우, 수집·축적되는 대용량의 데이터를 활용한 교통안전 향상, 신산업 육성 등의 정부사업이 활발하게 추진되고 있음

- 교통카드, 하이패스, 차량 내비게이션을 비롯한 교통사고 정보, 사업용차량의 운행기록정보 등 다양한 정보가 수집·축적되고 있으며, 이렇게 수집·축적되는 데

이터를 활용하여 새로운 가치를 창출할 수 있는 서비스 체계 마련이 요구됨

- 따라서 수집·축적되는 대용량의 교통 빅데이터를 연구단 개발 기술을 적용하여 범용적으로 활용 가능한 서비스 시나리오를 개발하고자 함
- 본 연구에서는 사업용 차량의 운행기록데이터와 교통사고 데이터를 활용하여 연구단에서 개발한 분석모듈을 적용한 결과 도출까지의 과정을 시나리오 주제로 선정함

○ 서비스 사용자 분석

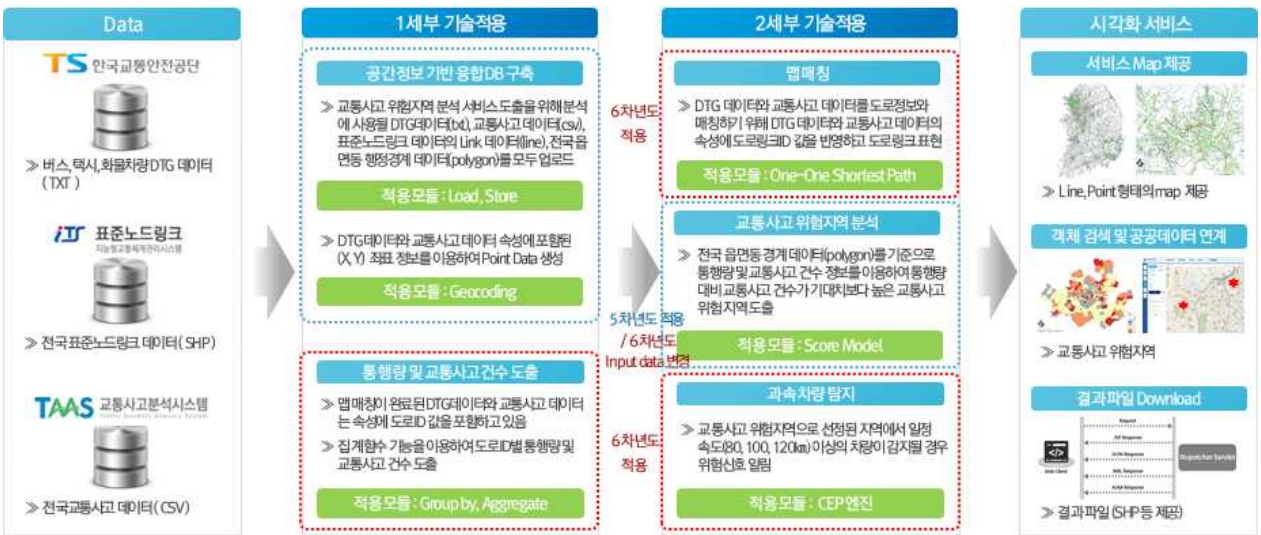
- 서비스 사용자는 크게 공공과 민간으로 구분되며, 공공의 경우 중앙정부 및 광역시도와 지자체로, 민간의 경우 전문 연구기관, 일반기업, 일반시민으로 세분화할 수 있음

<표 2-105> 네트워크정보 기반 서비스 사용자 분석

구분		서비스 사용자	서비스 사용분야
공공	중앙정부 및 광역시도	한국도로공사 및 지역별 도로관리사업소	• 고속도로 대상 주요 사고위험지역 선정 • 주요 교통사고 위험과 관련한 고속도로 상의 교통시설 및 제도 등 검토 시 활용 • 고속도로 내 운전자 쉼터 등 휴식처 마련 시 활용
	지자체	지자체 및 관할 경찰서	• 관할지역 내 교통수단, 교통시설 또는 교통사업자와 관련된 산업, 제도 등과 관련하여 교통사고를 줄일 수 있는 방안 마련 시 활용
민간	전문 연구기관	교통 및 관련분야 연구기관	• 교통사고 원인분석 등 관련 연구 수행 시 활용 • 대중교통 이용패턴, 유동인구 등을 활용한 대중교통 취약지역 분석 등 관련 연구 수행 시 활용
	기업	운수업체, 내비게이션업체, SI업체	• 대중교통 대상 주요 사고지점을 파악하여 운전자에게 위험지역 정보 제공 및 안전인식 고취 • 운수업체의 경우 시내버스, 마을버스 등 지역 내 운영노선에 대한 적정성 검토 및 노선변경, 신설 시 최적 노선 선정에 활용
	시민	일반시민, 웹/앱 개발자	• 교통사고 다발지점에 대한 정보 획득 • 교통사고정보 및 차량운행패턴정보를 활용하여 새로운 아이디어가 부합된 웹/앱 개발

○ 서비스 시나리오 개요

- 사업용 차량의 운행기록 정보와 교통사고 정보를 활용하여 1세부와 2세부에서 개발한 분석 기술을 적용하여 교통사고 위험지역을 도출함



<그림 2-359> 네트워크정보 기반 서비스 시나리오

○ 활용데이터

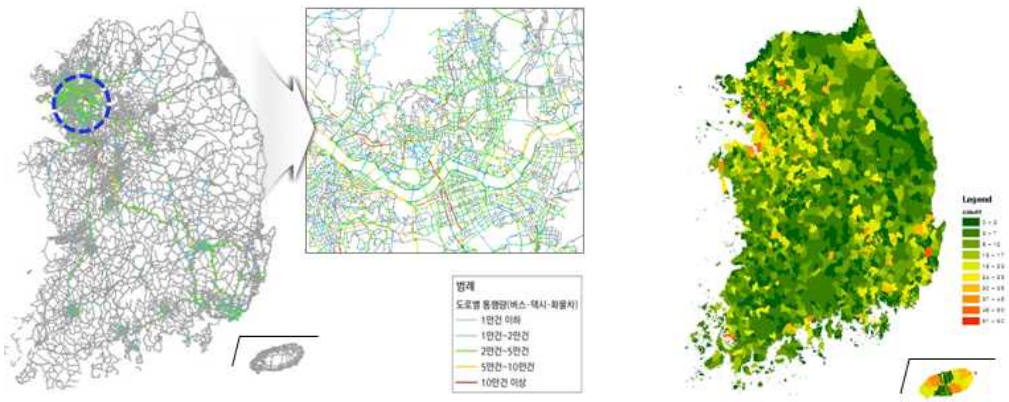
- 한국교통안전공단에서 관리하고 있는 차량 운행기록 정보(DTG데이터)와 국가교통정보센터에서 제공하는 전국표준노드링크 데이터, 도로교통공단에서 제공하는 전국 교통사고 데이터 활용

<표 2-106> 네트워크정보 기반 서비스 시나리오 활용데이터

항목	내용	파일포맷	데이터용량	출처
운행기록데이터	택시, 화물차, 버스 등 사업용 차량의 운행기록데이터	TXT	821GB	국토교통부 TS교통안전공단
도로데이터	전국표준노드링크	SHP	346MB	국토교통부 국가교통정보센터
교통사고 데이터	전국 교통사고 데이터	CSV	65KB	행정안전부 도로교통공단

○ 1세부 개발 기술 적용

- 공간정보 기반 융합DB 구축
  - 운행기록데이터(TXT)와 교통사고 데이터(CSV)의 속성에 포함된 좌표정보(x, y)를 이용하여 각각의 Point data 생성
  - 1세부 개발 기술 중 Geocoding 모듈 적용
- 통행량 및 교통사고 건수 도출
  - 맵 매칭한 결과 데이터인 운행기록데이터와 교통사고 데이터에는 도로링크ID 값이 포함되어 있으며, 집계함수를 통해 도로링크ID별 통행량 및 교통사고 건수 도출
  - 1세부 개발 기술 중 Group by & Aggregate 모듈 적용



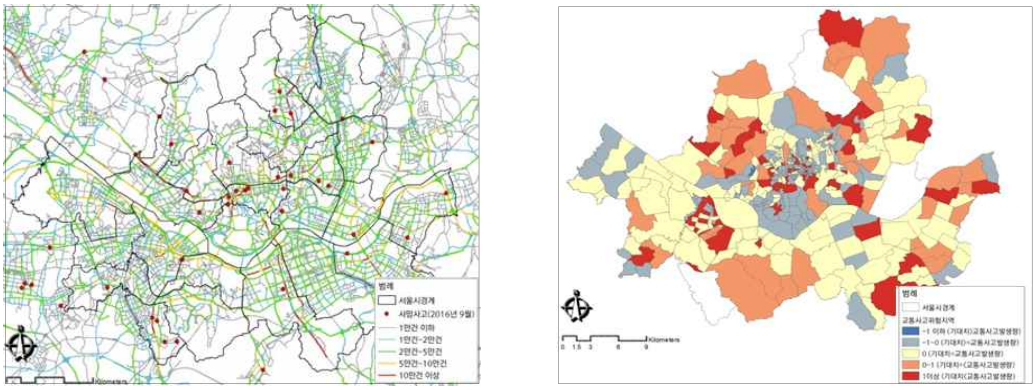
<a> 전국 도로별 통행량 분석 결과 <b> 전국 읍면동별 교통사고건수 분석 결과

<그림 2-360> 전국 도로별 통행량 및 읍면동별 교통사고건수 분석 결과(2016년 9월)

○ 2세부 개발 기술 적용

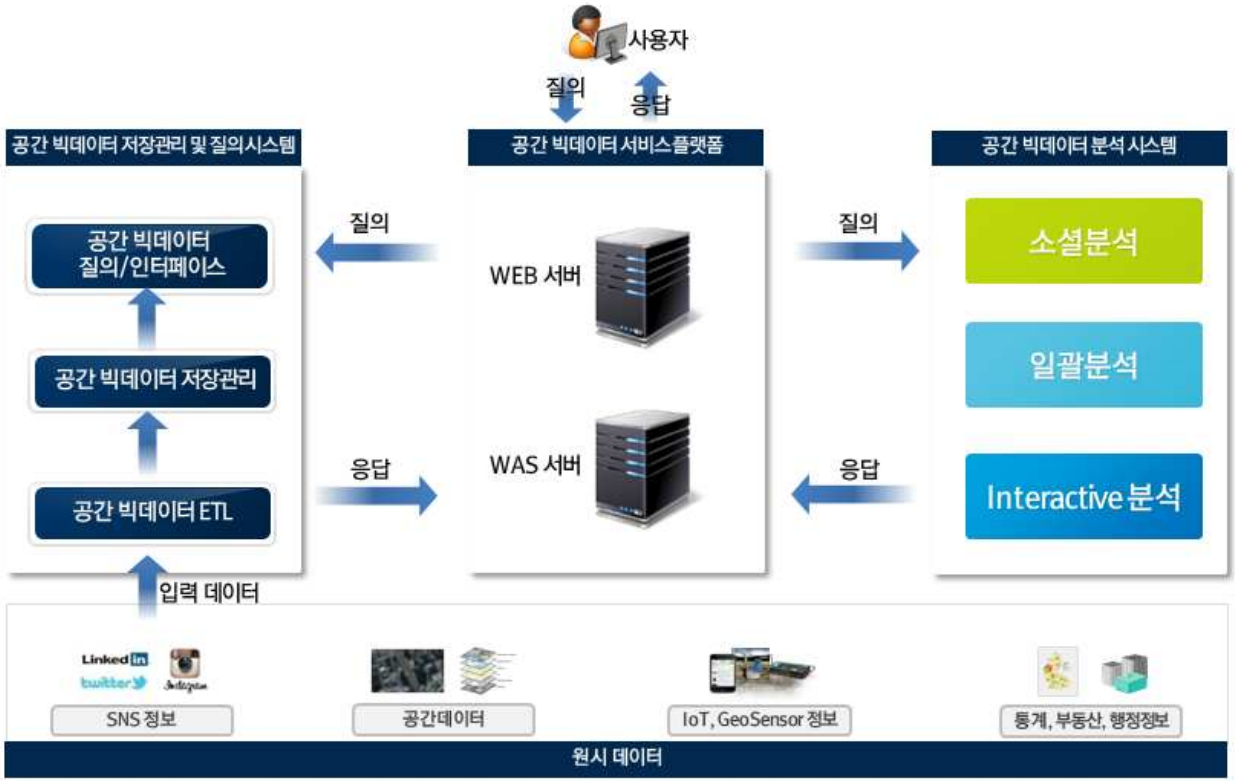
- 맵 매칭(Map Matching)
  - 운행기록데이터와 교통사고 데이터를 도로데이터와 각각 매칭(Matching)하여 운행기록데이터 및 교통사고 데이터의 속성에 해당 지점의 도로링크ID 값을 부여
  - 2세부 개발 기술 중 One-One Shortest Path 모듈 적용
- 교통사고 위험지역 분석

- 전국 읍면동 행정경계 데이터를 기반으로 통행량 정보와 교통사고 건수를 이용하여 교통사고 위험지역 도출
- 2세부 개발 기술 중 Score Model 모듈 적용



<그림 2-361> 전국 도로별 통행량 및 읍면동별 교통사고건수 분석 결과(2016년 9월)

⑥ [단위성과 J-7] 서비스 구현을 위한 시스템 아키텍처 설계(5차년도)



<그림 2-362> 공간 빅데이터 서비스 플랫폼의 시스템 아키텍처 구성

○ 공간 빅데이터 서비스 플랫폼의 시스템 아키텍처는 공간 빅데이터 저장관리 및 질의 시스템, 공간 빅데이터 분석 시스템, 공간 빅데이터 서비스 플랫폼을 통하여 공간 빅데이터 원시데이터의 입력/가공/저장/관리/질의/분석을 통해 사용자에게 다양한 맞춤형 공간 빅데이터 서비스를 제공하는 것을 주요 목표로 함

○ 공간 빅데이터 저장관리 및 질의 시스템은 공간 빅데이터 원시데이터를 공간 하둡 에코시스템에 입력 데이터로 적재, 적재된 데이터의 전처리 가공/저장/관리, 데이터들을 질의/인터페이스를 통하여 공간 빅데이터 서비스 플랫폼에 전달하는 기능을 제공

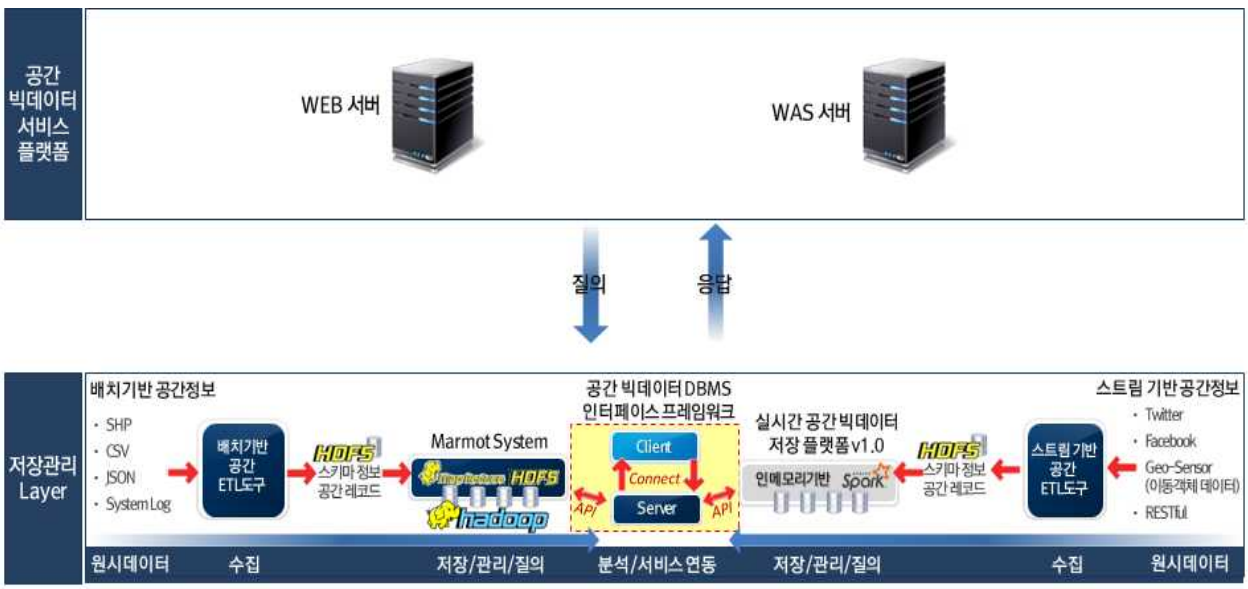
○ 공간 빅데이터 분석 시스템은 공간 빅데이터 저장관리 및 질의 시스템으로부터 질의를 통하여 응답받은 결과데이터를 입력데이터로 활용하여 소셜분석, 일괄분석, Interactive 분석을 수행하고 그 결과 데이터를 공간 빅데이터 서비스 플랫폼에 제공

○ 공간 빅데이터 서비스 플랫폼은 공간 빅데이터 저장관리 및 질의 시스템과 공

간 빅데이터 분석 시스템에서 질의 및 응답받은 결과데이터들을 융·복합하여 다양한 사용자의 수요를 충족시키는 맞춤형 서비스를 제공

○ 저장관리 Layer

- Hadoop 환경에서 공간 빅데이터를 고속으로 저장, 관리, 접근하는 기능을 제공하는 공간 빅데이터 저장관리 시스템, SQL 기반의 공간 질의 언어를 처리하는 질의 처리 엔진, 응용 프로그램에서 공간 빅데이터를 접근하여 사용할 수 있는 응용프로그램 인터페이스 등을 지원
- 기존 텍스트 형식의 빅데이터 처리기술(Hadoop)과 공간정보(Spatial)가 융합된 대용량 공간 빅데이터의 수집, 저장, 분석 기술 지원
- 배치 기반의 공간정보 수집/저장관리/질의/분석서비스 연동
  - SHP, CSV, JSON, System Log 등 대용량의 배치기반 정형, 반정형, 비정형 공간 원시데이터를 배치기반의 공간 ETL도구를 통하여 수집하여 HDFS기반의 스키마 정보와 공간레코드 데이터로 저장, 공간 하둡기반의 Marmot System에서 이 데이터를 가공/저장/관리, 공간 빅데이터 DBMS 인터페이스 프레임워크를 활용하여 원격의 클라이언트가 서버를 통해 Marmot System의 데이터 가공/저장/관리를 질의 및 결과 값을 전달받는 기능 제공
- 스트림 기반의 공간정보 수집/저장관리/질의/분석서비스 연동
  - Twitter, Facebook, Geo-Sensor(이동객체 데이터), Restfull 데이터 등 대용량의 스트림 기반 공간 원시데이터를 스트림 기반의 공간 ETL도구를 통하여 수집하여 HDFS기반의 스키마 정보와 공간레코드 데이터로 저장, 실시간 공간 빅데이터 저장 플랫폼에서 이 데이터를 가공/저장/관리, 공간 빅데이터 DBMS 인터페이스 프레임워크를 활용하여 원격의 클라이언트가 서버를 통해 실시간 공간 빅데이터 저장 플랫폼의 데이터 가공/저장/관리를 질의 및 결과 값을 전달받는 기능 제공



<그림 2-363> 공간 빅데이터 서비스 플랫폼의 저장관리 Layer 구성

○ 분석 Layer

- Hadoop 환경에서 다양한 형식의 공간 빅데이터를 고속으로 분산 분석하는 기능을 제공하는 재사용 가능한 분석모듈지원
- 특별히, 소셜 빅데이터를 기반으로 공간 빅데이터를 분석하는 소셜분석, 이동객체 공간 빅데이터의 자동화된 분석을 지원하는 일괄분석, 실시간 스트리밍 공간 빅데이터를 지원하는 분석엔진지원
- 소셜분석
  - (공간 빅데이터 정합모듈) 공간 빅데이터의 다양한 분석을 위한 기본 프레임에 해당하는 분석모델을 설계함으로써 공간 빅데이터 분석의 설계와 자동 프로세싱 구현 모듈
  - (공간 빅데이터 주제도 분석 모듈) 주제도, 분포도, 차트 분석 등의 공간 빅데이터 분석 기술과 분석결과 시각화를 통해 사용자의 요구에 맞는 분석 결과를 실시간 제공 모듈
- 일괄분석
  - (공간 빅데이터 Batch 분석 모듈) 통계기반, 비통계 기반의 분석 기법 도구들을 이용하여 공간 빅데이터의 일괄(Batch) 처리가 가능한 공간 빅데이터 분석 기능 지원하는 모듈
  - (이동객체 데이터 Batch 분석 모듈) 공간 빅데이터 일괄 분석 기능 개발과 마찬가지로, Trajectory 정보가 포함된 이동객체 데이터를 이용하여 통계기반, 비통계 기반의 분석 기법을 도입하여 일괄(Batch) 처리가 가능한 이동객체 분석 시스템을 지원하는 모듈
- Interactive 분석
  - (공간 빅데이터 OLAP 엔진) 공간 OnLine Analytical Processing(OLAP) 엔진: 공간 정보와 연관된 통계값 및 정보(예: 부동산, 지적 데이터)들을 대화형으로 검색 및 분석을 수행하는 모듈
  - (공간 빅데이터 복합 이벤트 처리(CEP) 엔진) 공간 Complex Event Processing(CEP) 엔진: 스트림 공간 빅데이터(예 : 교통 데이터)로부터 복합적인 이벤트를 검색하여 실시간 분석을 수행하는 모듈



<그림 2-364> 공간 빅데이터 서비스 플랫폼의 분석Layer 구성

○ 표출 Layer

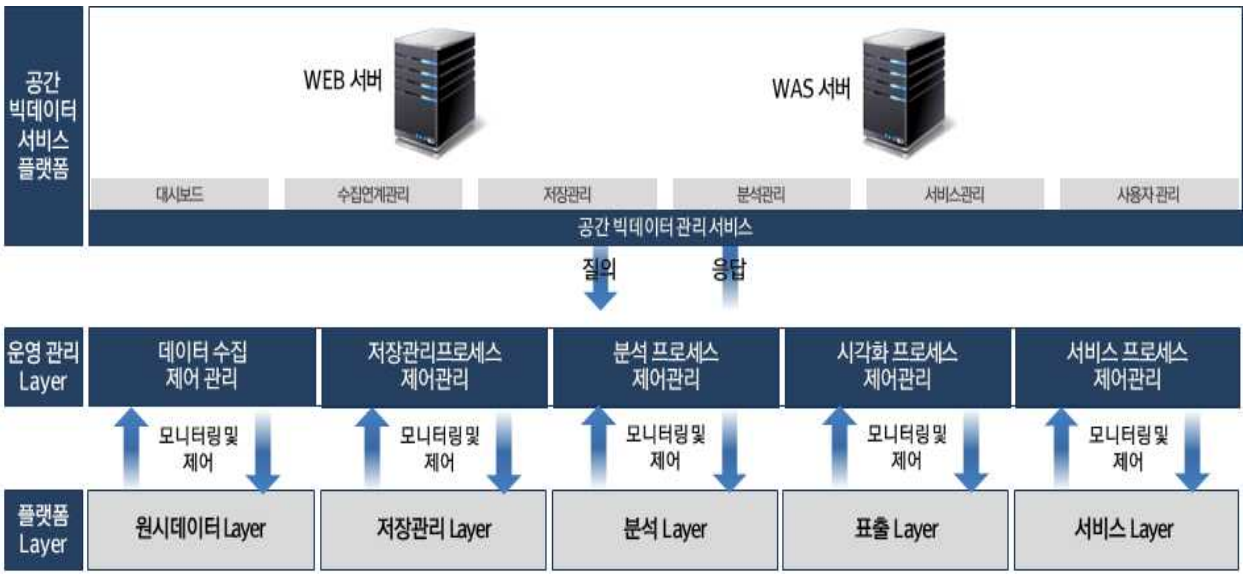
- 공간 빅데이터 저장·관리 및 분석을 통하여 산출된 결과를 기반으로 사용자에게 알기 쉽게 전달하기 위하여 공간 빅데이터 개별서비스와 응용서비스의 사용자 경험(UI) 및 사용자 인터페이스(UX)지원
- 공간 빅데이터 저장·관리 및 분석을 통하여 산출된 결과를 RESTful WEB API방식으로 사용자에게 요청받은 정보를 XML, JSON 포맷으로 제공하는 서비스 모듈 지원
- 공간 빅데이터 서비스 기술은 서비스 사용자의 사용목적과 기술적 숙련도에 따라 웹/모바일 브라우저 서비스, 파일 다운로드 서비스, 프로그래밍 인터페이스 제공 서비스와 같이 3가지 서비스 기술유형으로 구분하고, 각 유형별로 맞춤형 서비스 지원
  - (공간정보 오픈플랫폼 기반의 웹/모바일 브라우저 서비스 UI/UX) 웹/모바일 브라우저 서비스 유형은 공간 빅데이터 저장·관리와 분석모듈을 통하여 산출된 결과를 공간정보 오픈플랫폼 기반의 웹/모바일 브라우저를 통하여 사용자에게 알기 쉽게 전달하기 위하여 리포팅, 인포그래픽, 애플릿 형식의 시각화정보를 제공
  - (파일 다운로드 서비스) 파일 다운로드 서비스 유형은 공간 빅데이터 저장·관리와 분석모듈을 통하여 산출된 결과를 서비스 사용자가 추가활용하기 위한 목적으로 스프레드시트 파일, CSV 파일, XML/JSON 파일 등 다양한 파일형식으로 제공
  - (프로그래밍 인터페이스 제공 서비스) 프로그래밍 인터페이스 제공 서비스 유형은 공간 빅데이터 저장·관리와 분석모듈을 통하여 산출된 결과 또는 개별 모듈의 기능을 웹서비스 형태로 제공하여 서비스 사용자가 다양한 웹/모바일 응용서비스를 구현. 웹서비스는 공간 빅데이터 저장·관리와 분석모듈을 통하여 산출된 결과 데이터 또는 서비스의 성숙도에 따라 JSON, XML, OpenAPI 형식으로 제공



<그림 2-365> 공간 빅데이터 서비스 플랫폼의 표출 Layer 구성

○ 운영관리 Layer

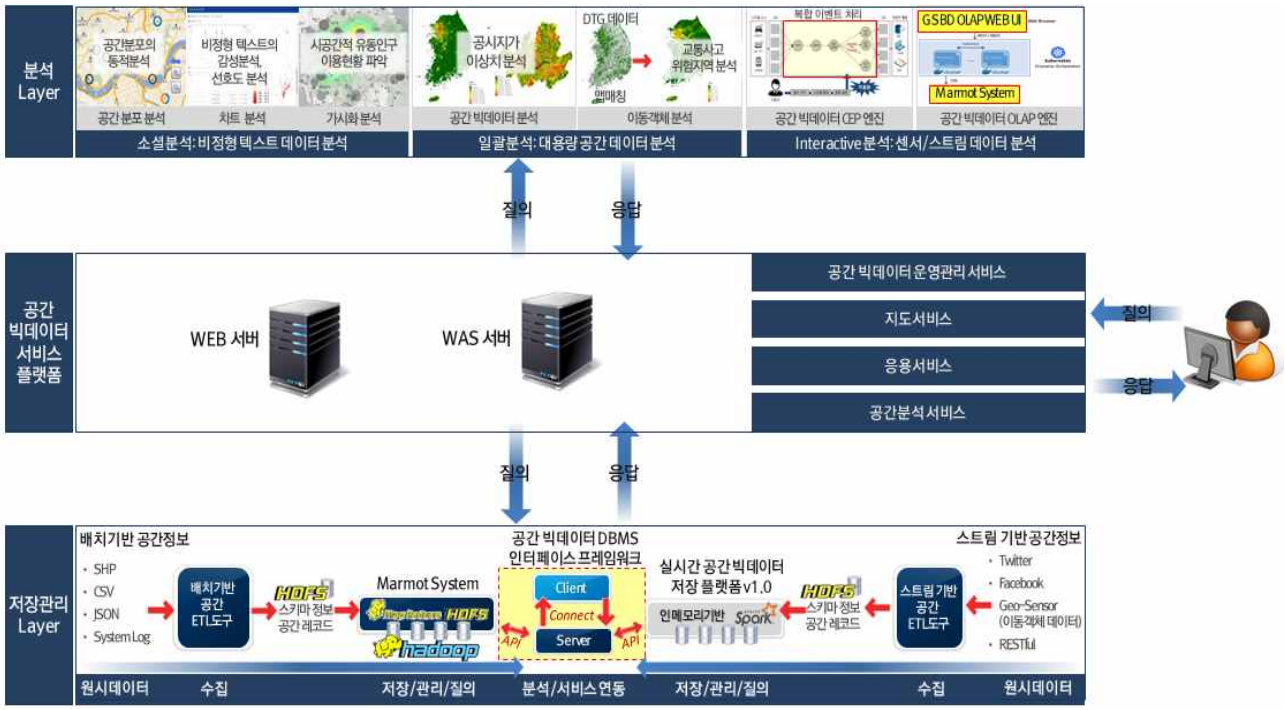
- 플랫폼 내부적으로 서비스 Container에 있는 개별 서비스와 응용서비스 Pool에 있는 응용서비스를 통하여 산출된 결과를 통합 관리할 수 있는 기능 제공플랫폼 외부적으로 타 플랫폼과의 연계활용을 위한 서비스 인터페이스 관리기능 제공
- 서비스 Container와 응용서비스 Pool에 있는 다양한 공간 빅데이터 서비스 이용현황 모니터링 기능 제공
  - (데이터 수집 제어관리) 수집/연계를 위한 데이터 관리, 오류관리, 스케줄링 관리, 수집/연계 데이터의 통계치 표출 기능
  - (저장 관리 프로세스 제어관리) 저장된 데이터 리스트, 데이터 분류별 통계치 표출 기능
  - (분석 프로세스 제어관리) 분석 데이터 리스트, 분석 기능별 통계치를 표출하고 분석기능 중 스케쥴 기능
  - (시각화 프로세스 제어관리) 시각화 데이터의 리스트, 저장/관리/분석 실증 시각화 서비스의 시각화 프로세스 제어, OpenAPI의 시각화 프로세스 제어기능
  - (서비스 프로세스 제어관리) 서비스 데이터 리스트, 서비스 모델 관리(GIS 서버 정보, 워크스페이스 리스트, 데이터 스토어 리스트, 레이어 리스트 등), 서비스 통계치 등의 표출 기능



<그림 2-366> 공간 빅데이터 서비스 플랫폼의 분석Layer 구성

○ 공간 빅데이터 서비스 플랫폼

- 공간 빅데이터 저장·관리·분석모듈이 제공하는 기능을 기반으로 범용적인 개별 서비스 제공
- 공간정보의 빅데이터 관리, 분석 및 서비스 플랫폼 실증을 위한 공공과 민간부문의 비즈니스모델 제시
  - (공간 빅데이터 운영관리 서비스) 공간 빅데이터의 수집연계/저장/분석/서비스/사용자관리 및 모니터링을 위한 대시보드 서비스 제공
  - (지도 서비스) 특정 공간 데이터의 레이어와 특정 속성을 검색하는 서비스 제공 및 공공데이터를 Web Map Service 및 Web Feature Service를 통하여 공간 시각화하는 지도서비스 제공
  - (공간 분석 서비스) 저장관리 Layer, 분석 Layer, 표출 Layer에서 제공하는 개별적인 기능들을 사용자에게 제공
  - (응용 서비스) 저장관리 Layer, 분석 Layer, 표출 Layer에서 제공하는 개별적인 기능들을 연계하여 사용자에게 다양한 맞춤형 공간 빅데이터 서비스를 제공



<그림 2-367> 공간 빅데이터 서비스 플랫폼의 분석Layer 구성

⑦ [단위성과 J-8] 핵심서비스의 주요기능 도출(5・6차년도)

○ 공간 빅데이터 플랫폼은 기능에 따라 4개의 서비스 항목으로 구분되며, 기능별 정의는 다음과 같음

- 공간분석 서비스 : 1세부 및 2세부에서 개발한 공간분석 기능을 제공하는 서비스
- 시각화 서비스 : 공간분석 결과 및 타부처 주제도 매쉬업 등의 결과를 시각화하여 보여주는 서비스
- OpenAPI 서비스 : 공간 빅데이터 플랫폼 상에서 분석한 결과데이터 및 원시데이터를 제공하는 서비스
- 플랫폼관리 서비스 : 공간 빅데이터 플랫폼 운영관리자를 위한 서비스로 데이터 저장, 분석, 서비스 및 사용자 관리 등을 지원

<표 2-107> 공간 빅데이터 플랫폼 핵심서비스별 주요기능

서비스 구분	주요기능		
공간분석 서비스	1세부	레이어관리	• 레이어 검색, 레이어 목록, 작업 목록
		공간 빅데이터 ETL	• 배치기반 공간 데이터 ETL 처리엔진, 스트림기반 공간 빅데이터 ETL 처리엔진
		공간 빅데이터 저장프레임워크	• 공간 빅데이터 색인, 색인조회, 공간연산자(31종)
		인메모리기반 공간 빅데이터 저자프레임워크	• 인메모리 기반 공간색인, 공간색인 조회, 이동객체 저장 및 색인, Updatable 공간 빅데이터 저장, 인메모리 데이터 처리 공간연산자(23종)
		공간 빅데이터 질의처리	• 공간 빅데이터 JDBC서버, 공간 빅데이터 질의(SQL) 처리, 공간 빅데이터 처리 언어
		원격 공간 빅데이터 활용	• 클라이언트 서버 기반 원격 공간 빅데이터 접근, 공간 빅데이터 응용개발 도구
	2세부	Social분석	• 데이터 필터링, 장소추출, 좌표추출, 공간감성분석, 장소별 언급빈도수 산출, 장소별 선호도 빈도수 산출
		Batch분석	• Clustering&Regression 분석(6종), Factor분석(2종), Anomaly Detection(2종), Score Model, SVDD, Network분석(3종), Hotspot분석(2종), Accessibility분석(2종)
		Interactive분석	• 공간 빅데이터 CEP엔진, 공간 빅데이터 OLAP엔진
시각화 서비스	검색결과 시각화		• 전체검색, 명칭검색, 지번검색, 도로명 검색
	공공데이터 시각화		• 브이월드 주제도 매쉬업, 통계청
	서비스모델 시각화		• 토지건물정보기반 서비스, 비정형 텍스트정보 기반 서비스, 네트워크정보 기반 서비스
OpenAPI 서비스			• Map Service, Json/XML, OpenAPI 카탈로그, OpenAPI 검색, OpenAPI 레퍼런스, OpenAPI 서비스 게시판, OpenAPI 이용현황정보, OpenAPI 관리정보
플랫폼관리 서비스	수집데이터 제어관리		• 수집목록관리, 오류관리, 수집통계관리, 스케쥴관리
	가공·저장·프로세스 제어관리		• 데이터관리, 분류관리, 저장/변환통계관리
	분석프로세스 제어관리		• 분석관리, 분석통계관리
	서비스프로세스 제어관리		• 융합DB목록관리, 서비스모델관리, 통계관리
	사용자관리		• 사용자 목록관리, 권한관리, 통계관리



<그림 2-368> 공간 빅데이터 플랫폼 핵심서비스별 주요기능

(다) 공간 빅데이터 확장 Open API 서비스(핵심성과 K)

① [단위성과 K-1] 공간정보 오픈플랫폼 Open API현황 분석 (1차연도)

○ V-WORLD는 다양한 수요에 대응할 수 있도록 단순 조회에서부터 원시 데이터 직접 제공까지의 서비스 제공

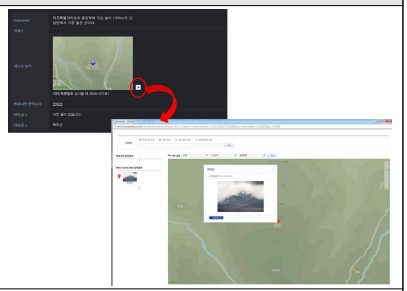
○ 포털 서비스, 통합지도서비스, 개발자 센터, 데이터센터, 참여서비스, 3D데스크톱 서비스, 3D Map(영문지도서비스) 총 7개의 서비스 제공

○ 2D지도 API, 3D지도 API, 배경지도 API, WMS/WFS API, 데이터 API, 지도검색 API, Static Map API 총 7개로 구성

② [단위성과 K-2] Open API 기술의 국내외 적용사례조사 (1차연도)

○ 주요 공간데이터 Open API 적용 사례

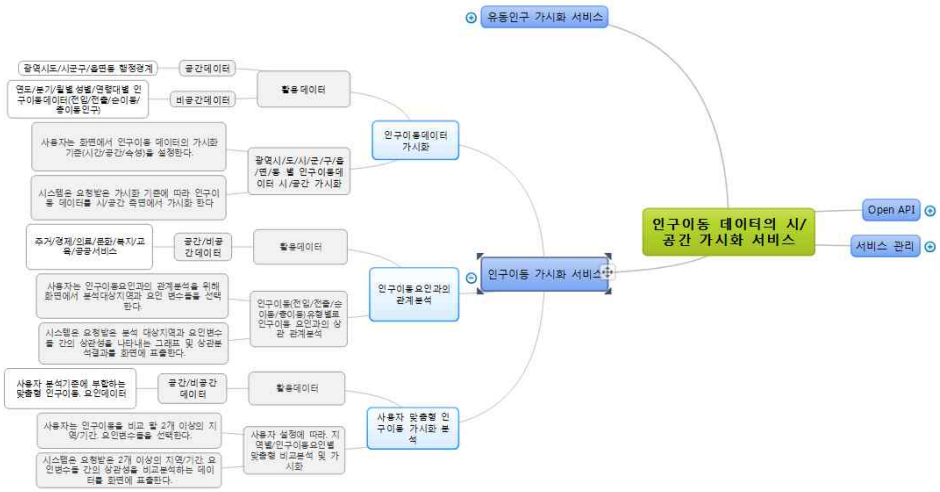
<표 2-108> 공간데이터 Open API 적용 사례

사이트	설명	비고
두산백과사전 Encyber 지도	• 네이버 지도 API • 지도 위에 백과사전 항목을 표시해주는 서비스 제공	
한국 석유공사 오피넷	• 올레맵 API • 웹, 모바일 기반의 주유소/LPG 충전소 가격정보를 제공	
Leerstandsmelder 서비스	• GoogleMaps • 부동산 거래를 위해 주소정보, 사진, 그리고 더 상세한 정보를 Leers사이트를 통해 공유	
Haiti 재난 구조	• OpenStreetMap • 참여형 지도서비스 • 시민 및 단체들이 직접 참여하여 지도 제작	

③ [단위성과 K-3] 공간 빅데이터 Open API 서비스 콘텐츠 구성 및 인터페이스설계 (2차연도)

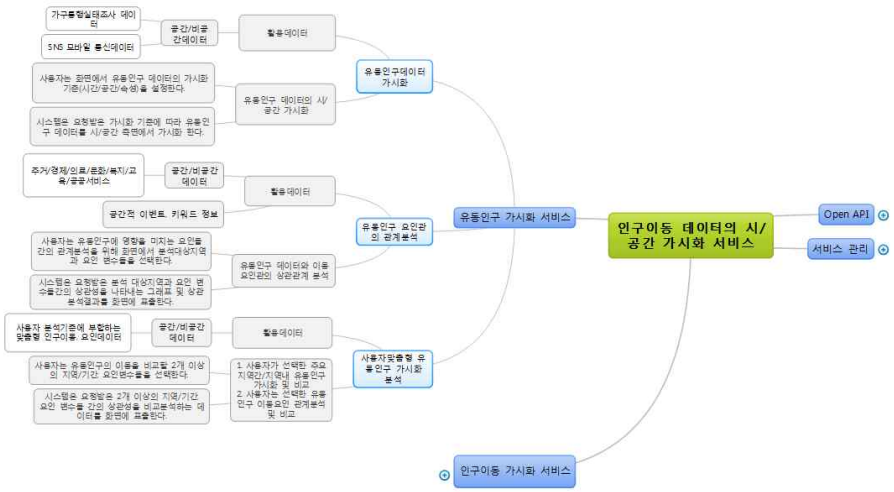
○ 공간빅데이터 Open API 서비스 컨텐츠는 3세부의 핵심 서비스 유형 및 모델 개발의 결과물인 교통부문의 인구이동·유동인구 데이터의 시·공간 가시화 서비스를 대상으로 함

○ 인구이동은 거주를 목적으로 인구가 현재 거주하고 있는 장소에서 거주하고자 하는 장소로 옮겨가는 현상으로 사회적 조건, 이동 동기, 이동자의 사회와 문화적 수준, 생활태도, 가치관 등이 관련되어 나타나는 매우 복잡한 현상임



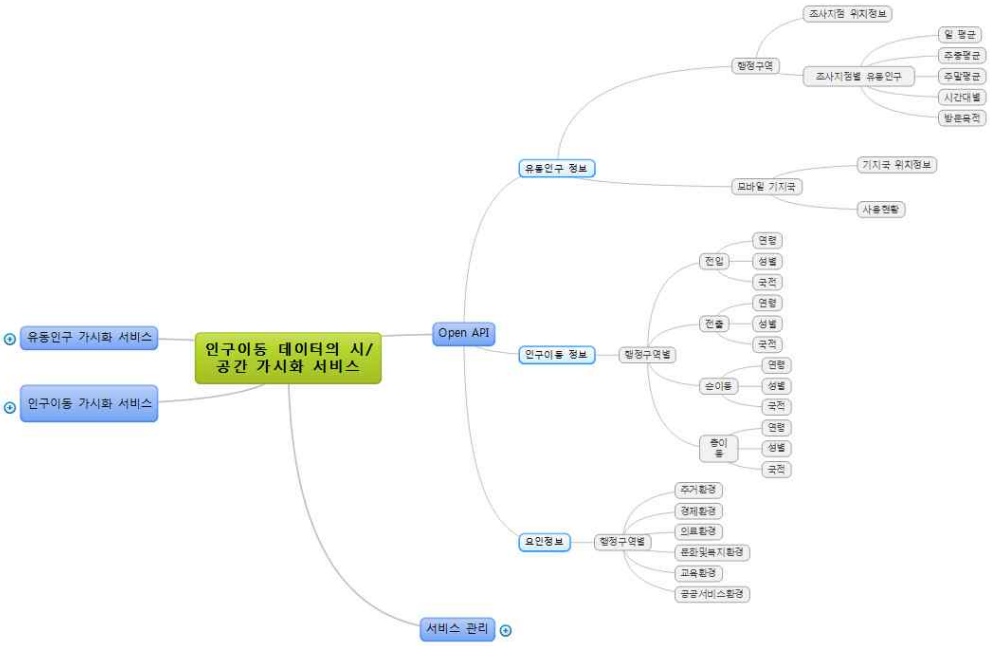
<그림 2-369> 인구이동 시각화 서비스 구성

○ 유동인구는 특정지점에서 일정시간을 기준으로 유입, 유출, 이동한 총 인구수로 건물 유·출입 인구와 가로 보행인구 합으로 볼 수 있음



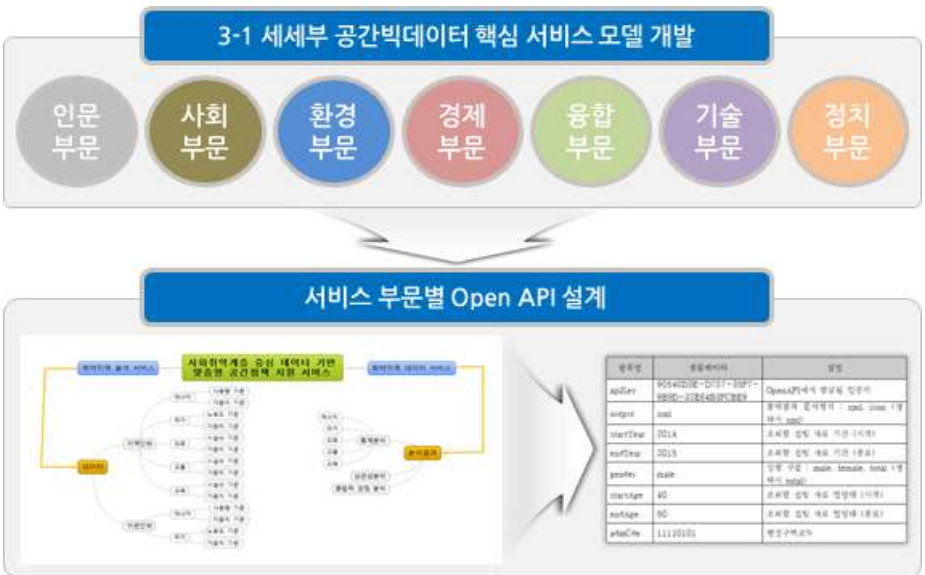
<그림 2-370> 유동인구 서비스 구성

○ 인구이동 및 유동인구 가시화 서비스의 Open API는 사용자에게 인구이동 또는 유동인구 현상의 데이터와 영향요인에 대한 정보를 제공하고 새로운 서비스를 만들 수 있는 기반을 지원함



<그림 2-371> Open API 구성

④ [단위성과 K-4] 공간 빅데이터 공간정보와의 연계범위 및 API 종류 설계 (2차연도)



<그림 2-372> Open API 설계 과정

○ 공간빅데이터 Open API 서비스 콘텐츠는 3세부의 핵심 서비스 유형 및 모델 개발의 결과물인 교통부문 서비스인 인구이동·유동인구 데이터의 시·공간 가시화 서비스를 대상으로 함

○ 국·내외 인구이동 및 유동인구 데이터의 가시화 사례를 조사하고 가시화 서비스의 주요 서비스 기능을 정리함

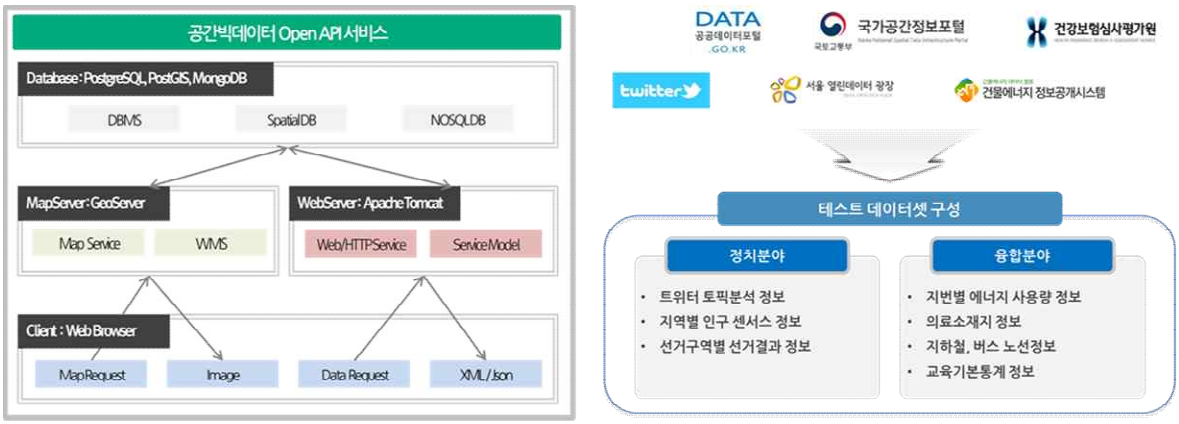
⑤ [단위성과 K-5] 공간 빅데이터 Open API 기능 개발 (3~4차연도)

○ 빅데이터 및 공간 빅데이터 서비스 사례 현황, 연구동향 및 사회현안 등을 고려하여 구체적인 서비스모델 개발을 위한 핵심서비스모델 세부적용분야를 도출함

○ Postgresql, PostGIS를 이용하여 SNS, 공간정보, 통계 데이터를 저장, 관리, 검색할 수 있는 환경을 구성함

○ MapServer를 이용하여 공간데이터의 지도 서비스 제공 WebServer를 이용하여 사용자 요구에 따른 검색 결과를 제공함

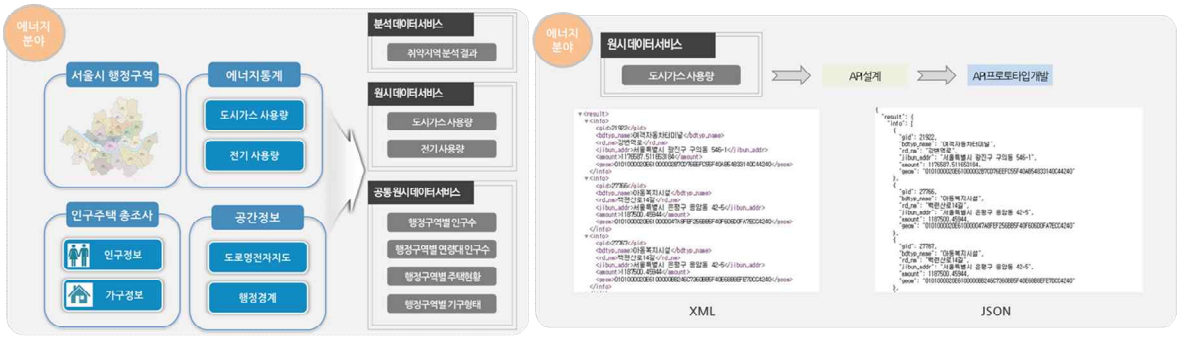
○ SNS, 선거정보, 에너지 사용량, 건물종합정보, 의료기관 정보, 교육기본통계, 대중교통 시설물 정보 등 수집한 데이터를 이용하여 융합, 정치, 기술 부문 핵심 서비스 모델을 위한 테스트 데이터셋을 구축함



<그림 2-373> Open API 개발환경 구성 및 테스트 데이터셋 구축

○ 핵심 서비스 모델을 바탕으로 Open API 서비스를 설계하기 위해 융합, 정치, 기술 부문의 핵심 서비스 모델에서 요구사항을 정의하고 요구사항에 따른 Open API 서비스 목록을 도출함

○ 서비스 내용에 따라 분석 데이터 서비스, 원시데이터 서비스, 공통원시데이터 서비스로 구분하고 Open API 서비스별 요청변수, 출력결과 등 상세 설계 수행함



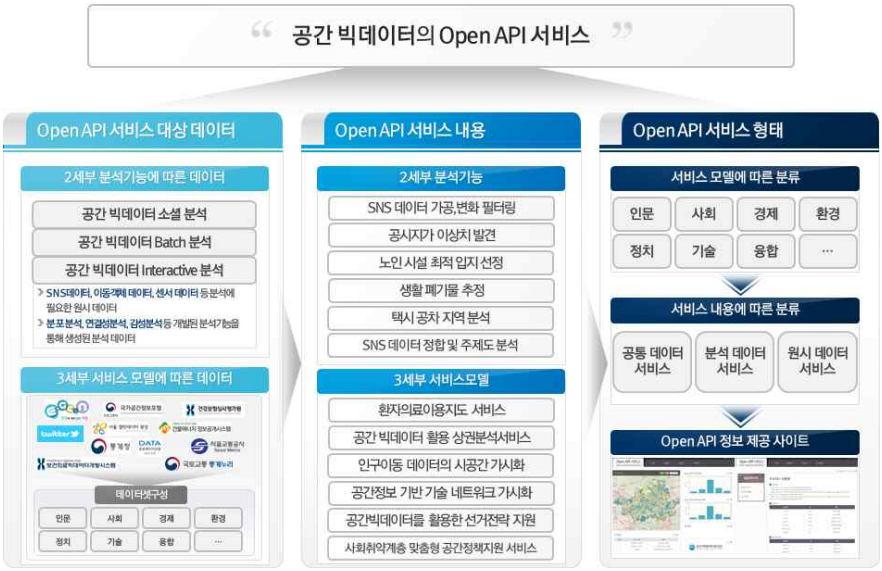
<그림 2-374> Open API 설계

○ Open API 서비스별 요청변수, 출력결과 등 상세 설계를 바탕으로 기능을 개발하고 Open API 서비스를 웹페이지를 통해 레퍼런스 정보 제공함



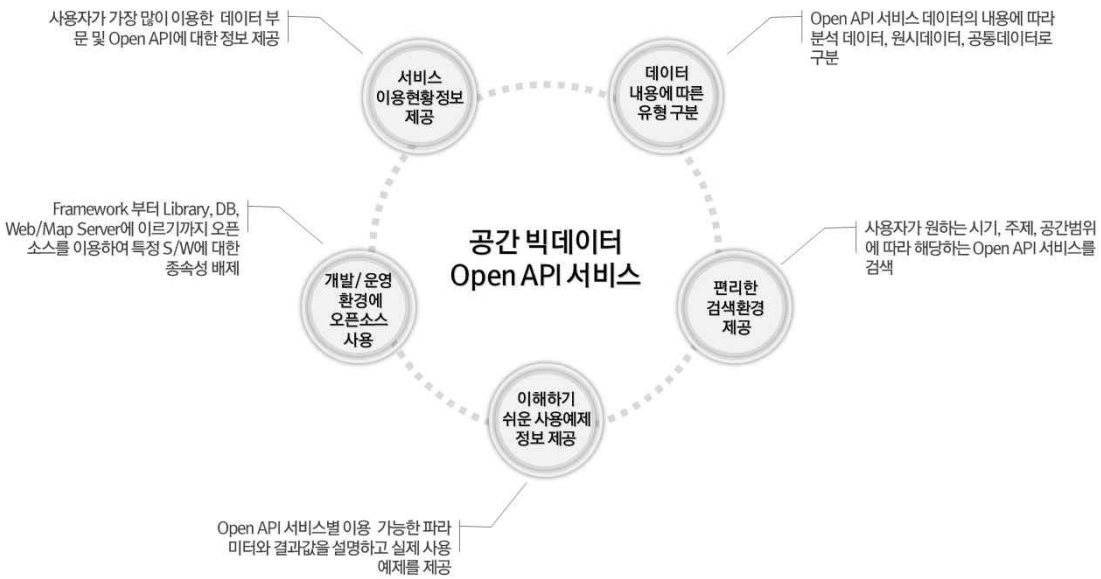
<그림 2-375> 웹페이지를 통한 Open API 레퍼런스 정보 제공

⑥ [단위성과 K-6] 공간 빅데이터 Open API 서비스 사이트 개발



<그림 2-376> 공간빅데이터 Open API 서비스 사이트 개발 프로세스

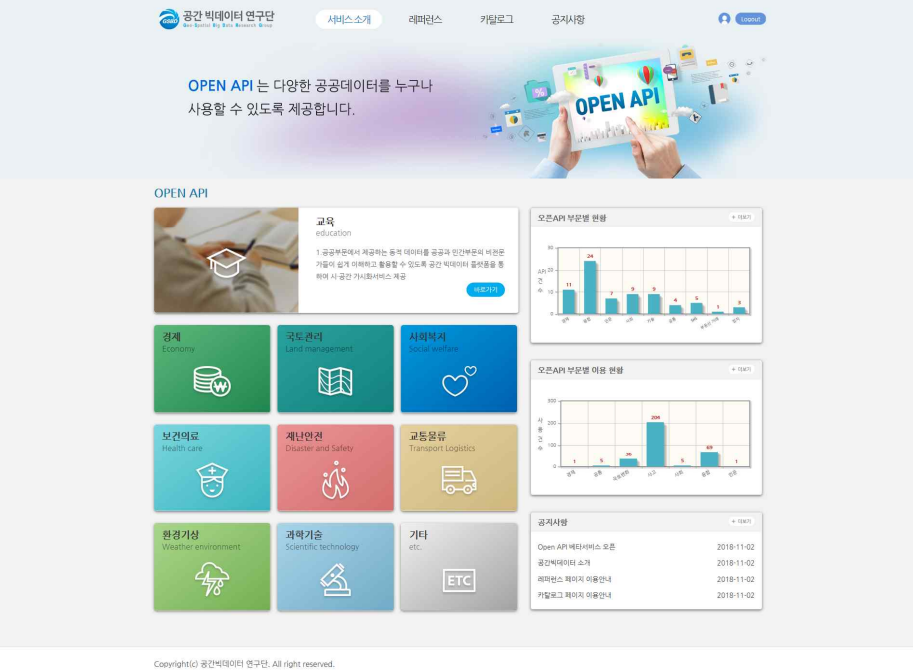
- ❍ 사용자가 공간 빅데이터 Open API 서비스를 이용할 수 있도록 Open API별 서비스 이용 정보를 제공하는 웹 사이트를 개발
- ❍ 공간 빅데이터 Open API 서비스 사이트는 데이터 내용에 따라 Open API 서비스를 구분하고 검색 환경을 제공하여 Open API 이용 정보와 사용예제를 제공함



<그림 2-377> Open API 서비스 사이트 특징

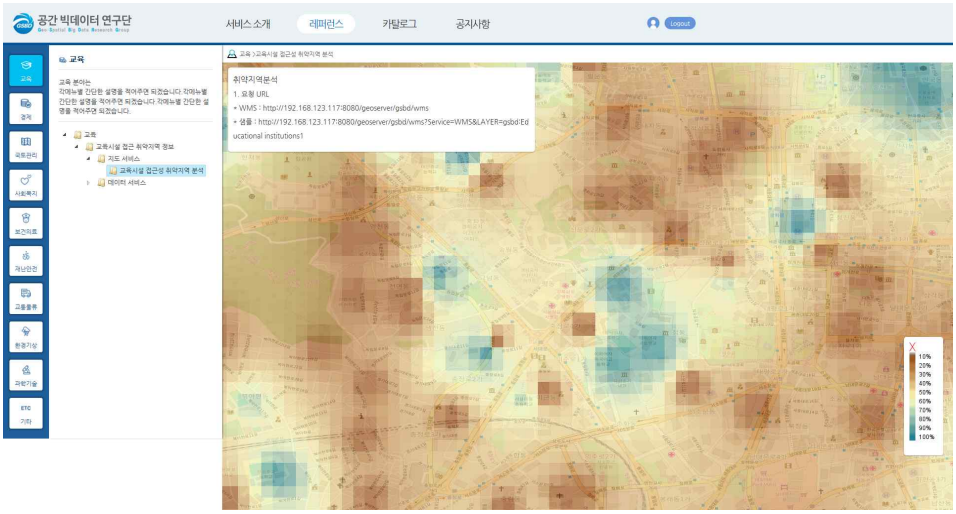
○ 핵심 서비스 모델의 시나리오에 따라 분석 결과, 분석에 활용한 데이터, 분석된 현상을 이해하기 위한 데이터들을 Open API 서비스의 내용에 따라 분석 데이터, 원시데이터, 공통데이터로 구분하고 서비스 함

○ 각각의 Open API 서비스는 사용자가 이용하기 쉽도록 입력 가능한 파라미터와 예시 값을 설명하고 사용자가 응답받을 수 있는 결과값과 내용을 예시를 들어 설명함, 또한 가장 일반적인 형태의 Json/xml 요청사항을 예제로 정리하여 실제 반환되는 값을 화면상에 제공 함



<그림 2-378> Open API 서비스 사이트 소개 페이지

○ 공간 빅데이터 Open API 서비스 사이트 소개 페이지는 구성에 따라 자주 사용하는 Open API 목록, Open API 부문 별 현황 차트, 부문별 이용현황 차트를 표시하도록 개발함, 각 서비스 부문의 목록은 클릭시 해당 레퍼런스 페이지로 이동하도록 개발됨



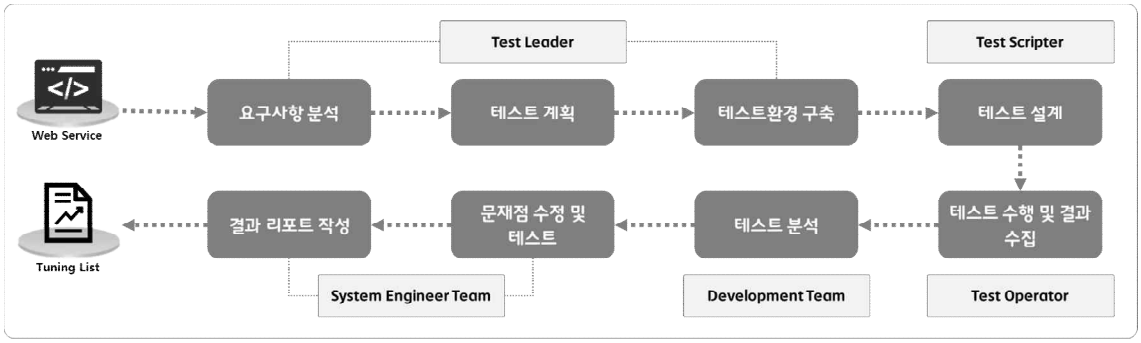
<그림 2-379> Open API 서비스 사이트 레퍼런스 지도서비스 페이지



<그림 2-380> Open API 서비스 사이트 레퍼런스 데이터서비스 페이지

○ 레퍼런스 페이지는 좌측의 서비스 부문을 선택하면 해당 부문에서 서비스하고 있는 Open API 서비스가 트리로 나타나고 해당 항목을 선택하면 오른쪽 컨텐츠 화면에서 해당 정보를 조회할 수 있도록 개발함

⑦ [단위성과 K-7] 연구성과물 성능 검증 지원 및 개선방안 연구



<그림 2-381> 성능 테스트 프로세스

○ 공간 빅데이터 Open API 서비스 성능 테스트는 서비스 및 서비스 시스템의 성능을 확인하기 위해서 실제 사용환경과 비슷한 환경에서 테스트를 진행함



<그림 2-382> 성능검증 스크립트 항목별 설명

○ 웹페이지 디스플레이 응답시간

- 동시 접속자 수가 100명인 경우 10개 부문 104개 데이터 모두 목표치인 3초 이내로 나타나 목표치 이내로 검증됨.
- 동시 접속자 수가 200명인 경우 10개 부문 104개 데이터 모두 목표치인 3초 이내로 나타나 목표치 이내로 검증됨.

○ 성능검증 시스템사용률-CPU사용률

- 10개 부문 104개 데이터 검증시 CPU 평균 사용률은 모두 목표치인 평균 90%이하로 나타나 목표치 이내로 검증됨.

○ 성능검증 시스템사용률-메모리사용률

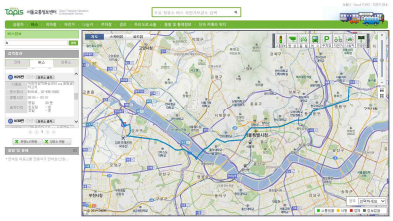
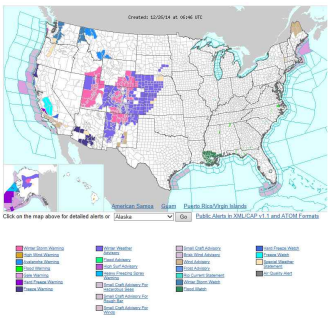
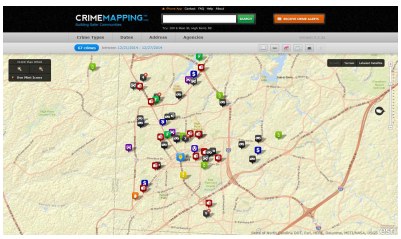
- 10개 부문 104개 데이터 검증시 메모리 평균 사용률은 모두 목표치인 평균 75% 이하로 나타나 목표치 이내로 검증됨.

(라) 공간 빅데이터 서비스 제공 UI/UX(핵심성과 L)

① [단위성과 L-1] 서비스별 UI/UX 기술 및 적용사례 분석 (1차연도)

❍ 서비스별 UI/UX 기술 및 적용사례

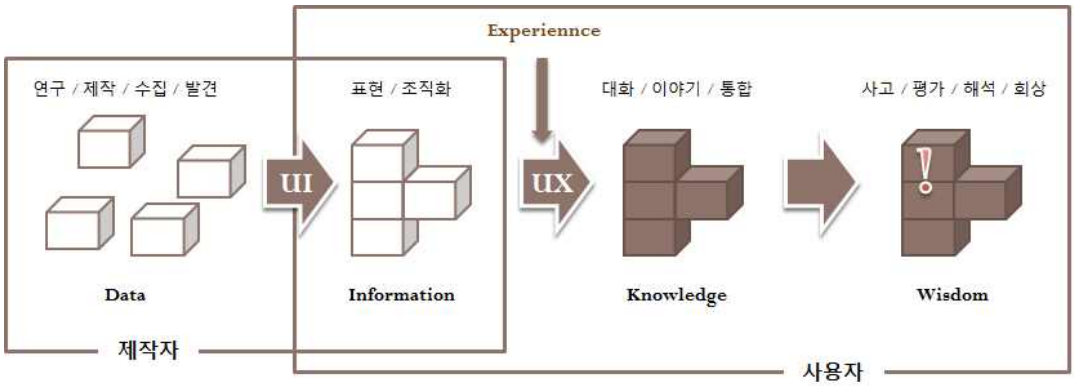
<표 2-109> 서비스별 UI/UX 기술 및 적용사례

사이트	설명	비고
T map	SK 텔레콤 스마트폰 기반의 대표적인 길 안내 시스템	
서울시 심야버스 노선	서울시 제공 심야시간대의 유동인구와 목적지 통행량 등을 분석 노선 및 배차 간격설정에 활용	
National Weather Service	미국 여름철 폭염 특보, 고온건강•경보, 눈보라, 스톰 및 돌풍, 홍수, 화재 경보와 같은 기상경보를 분석	
CrimeMapping	미국 미국 전역에 대한 범죄 정보를 제공 후속 범죄의 가능성을 예측하여 제공	

② [단위성과 L-2] 공간 빅데이터 서비스 제공을 위한 UI/UX 목표정의 (2차연도)

❍ UI(User Interface)/UX(User Experience) 역할

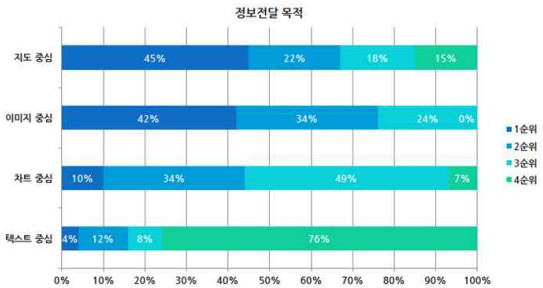
- 일반적으로 공간 빅데이터의 연구와 관심은 새로운 코어 기술의 개발 및 발견 그리고 데이터 분석 과정에 어떻게 적용 할지에 집중되고 있음
- 데이터의 구조가 더 복잡하고 더 많은 단계의 분석을 필요로 할 수 록 사용자가 공간 빅데이터의 분석 결과로부터 정보와 지식을 취득하여 의사결정 과정에 도달하기 위해서는 분석결과의 표현을 명확하고 직관적으로 해야 함



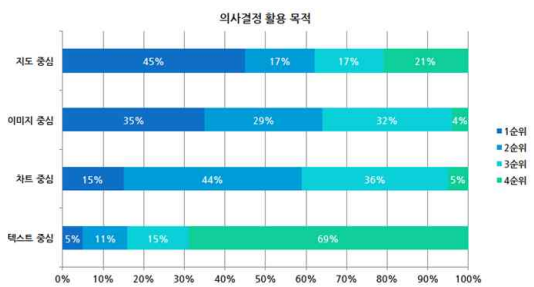
<그림 2-383> Nathan Shedroff의 정보 이해도

③ [단위성과 L-3] 공간 빅데이터 서비스 UI/UX 니즈 조사 (2차연도)

- ❍ 공간 빅데이터는 다양한 활용 분야에서 수요가 크게 증가할 것으로 예상되며 이러한 수요에 따라 핵심 서비스에 따른 사용자 기능 요구사항 도출이 필요함
- ❍ 공간 빅데이터를 주로 사용하는 기업, 대학, 연구소의 사용자를 대상으로 핵심 서비스 도출을 위한 문항과 사용자 기능 요구사항 도출을 위한 문항으로 구분하여 설문조사를 수행함
- ❍ 설문조사의 내용은 통계처리를 위한 문항, 공간빅데이터에 대한 이해도를 파악하기 위한 문항, 우선순위 및 서비스 활용 분야를 파악하기 위한 문항, 활용 우선순위를 파악하기 위한 문항, 서비스 활용 영역에 대한 문항, 사용 목적에 따른 시각화 요구사항 파악을 위한 문항으로 구성 되어 있음
- ❍ 공간 빅데이터 사용 목적에 따른 시각화 요구사항을 정의하기 위하여 4가지 시각화 목적을 제시하고 가장 적합한 인터페이스를 조사 함



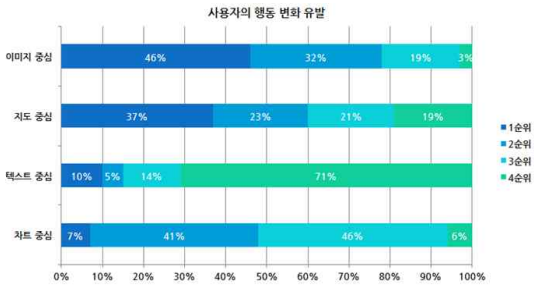
<그림 2-384> 정보전달 목적 사용자 환경 우선순위 조사 결과



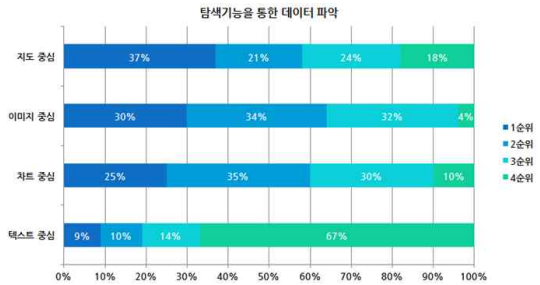
<그림 2-385> 의사결정 활용 목적 사용자 환경 우선순위 조사 결과

- ❍ 정보전달 목적은 지도중심(1순위:45%, 2순위:22%)과 이미지 중심(1순위:42%, 2순위:34%)의 사용자 환경을 1순위, 2순위로 응답한 비율이 높게 나타남
- ❍ 의사결정 과정에 활용은 지도중심(1순위:45%, 2순위:17%), 이미지 중심(1순

위:35%, 2순위:29%), 차트 중심(1순위:15%, 2순위:44%)을 사용자 환경의 1순위, 2순위로 응답한 비율이 높게 나타남



<그림 2-386> 행동변화 유발 목적 사용자 환경 우선순위 조사 결과



<그림 2-387> 데이터 파악 목적 사용자 환경 우선순위 조사 결과

○ 사용자의 행동변화를 유발하기 위한 경우는 이미지 중심(1순위:46%, 2순위:32%), 지도 중심(1순위:37%, 2순위:23%), 차트중심(1순위:7%, 2순위:41%)을 사용자 환경의 1순위, 2순위로 응답한 비율이 높게 나타남

○ 탐색기능을 통하여 데이터를 파악하기 위한 경우는 지도 중심(1순위:37%, 2순위:21%), 이미지 중심(1순위:30%, 2순위:34%), 차트중심(1순위:25%, 2순위:35%)을 사용자 환경의 1순위, 2순위로 응답한 비율이 높게 나타남

○ 공간빅데이터 서비스를 위한 UI/UX의 요구사항을 사용자 편의성, 직관성, 다양성, 일관성, 탐색의 효율성, 작업의 효율성 측면으로 분류하고 각 요구사항에 따른 상세 내용을 정리함

<표 2-110> UI/UX 요구사항 목록

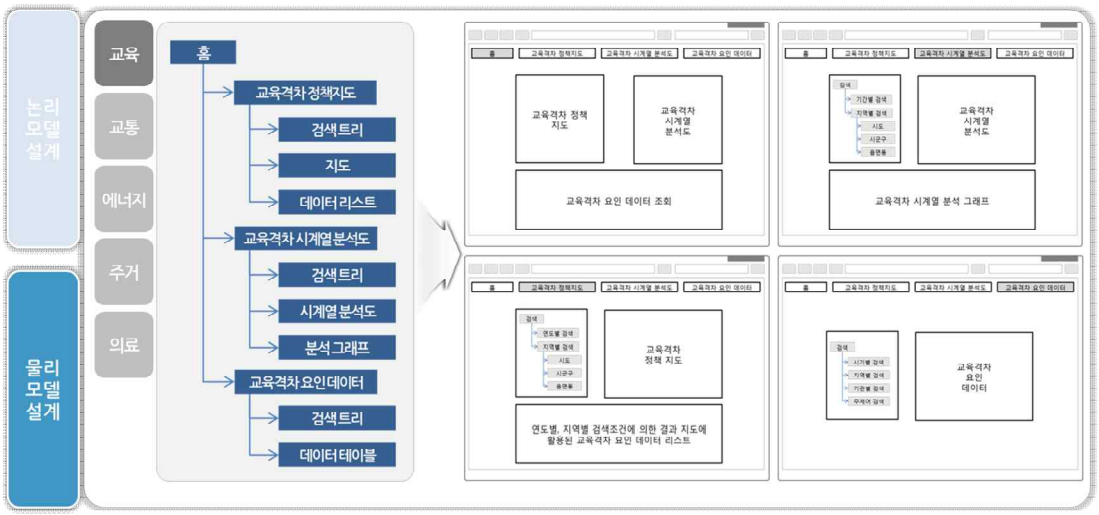
요구사항 분류	요구사항 명칭
사용자 편의성	사용자 편의 기능 제공
	사용자 편의를 고려한 적절한 콤포넌트 추천
직관성	입력 선택 영역의 명료화
	버튼 특성을 고려한 배치
	기능 목적에 부합하는 용어 사용
	정보 특성을 고려한 디자인
다양성	업무 특성별 특화 기능(분석된 그래프 정보 등) 제공
	다양한 사용자를 고려한 일반사용자/관리자 기능 제공
일관성	업무 및 소프트웨어 표준을 고려한 정보 제공
탐색 효율성	입력 정보 특성을 고려한 컨텐츠 배치
작업 효율성	입력 기능의 복잡한 단계 최소화
	작업 순서를 고려한 정보 배치
	사용자 편의를 위한 단축키 제공

④ **[단위성과 L-4] 공간 빅데이터 서비스 제공을 위한 UI/UX 상세설계 (3차 연도)**

- ○ 공간 빅데이터 서비스 UI/UX 상세설계 단계를 논리모델 설계와 물리모델 설계로 구분하여 진행함
- ○ UI/UX 논리모델 설계를 위해 사용대상이 되는 사용자 특징을 정의하고 시나리오를 작성함
- ○ UI/UX 물리모델 설계는 사용자에게 전달하려는 정보에 따라 화면 구조를 다르게 하고 이를 와이어프레임으로 작성함
- ○ 논리모델과 물리모델을 취합하여 스토리보드를 작성함



<그림 2-388> 논리모델 설계



<그림 2-389> 물리모델 설계

⑤ [단위성과 L-5] 공간 빅데이터 서비스 제공을 위한 UI/UX 테스트 시스템 개발 (5차연도)

○ 공간 빅데이터 서비스 실증시스템은 1세부와 2세부의 모듈을 연계하고, 3세부의 서비스 모델에 대하여 시각화, 분석, 저장, 관리 할 수 있는 실증 기술을 구현하는 것을 목적으로 함

○ 각각의 실증 기술은 시스템 상 실증시각화 서비스, 공간분석 프레임, 관리서비스, OpenAPI의 메인 메뉴로 구성함



<그림 2-390> 시스템 기능 구조

○ 앞선 시스템 기능 구조와 기능 스토리모델을 바탕으로 정보 유형별 User Interface를 구성하고 서비스 제공을 위한 시스템 개발함



<그림 2-391> 공간 빅데이터 서비스 제공 User Interface 구성

⑥ [단위성과 L-6] 공간 빅데이터 Open API 서비스를 적용한 UI/UX 테스트 시스템 개발 (6차연도)

○ 공간 빅데이터 서비스 메인페이지, 실증시각화, 공간분석프레임, 시스템 관리, OpenAPI의 단위 기능에 대한 User Interface를 개선하고, User Experience 요소를 적용함



<그림 2-392> 메인 페이지 UI/UX 적용

○ 기존 서비스 모델 중심의 메인페이지를 공간 빅데이터 서비스 실증 시스템 구성 내용의 메인페이지로 개선하고 Infinite Scroll을 이용한 User Experience 요소 적용함



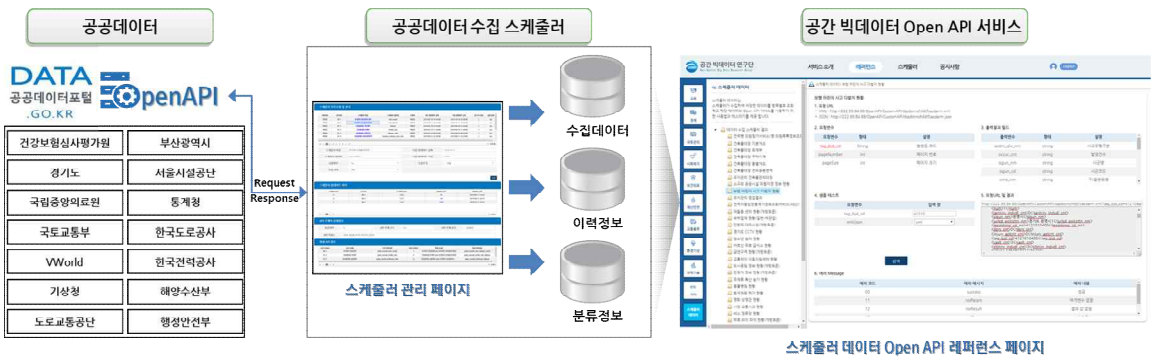
<그림 2-393> 시각화 서비스 메뉴 구조 단순화

○ 공간연산, 소셜분석, 기타분석 시각화 메뉴를 삭제하고 분석기능 자체에서 결과를 조회 하도록 시각화 메뉴를 단순화함



<그림 2-394> 공간분석 프레임 메뉴 UI/UX 개선

○ 공간 연산 및 분석 모듈 사용성 개선을 위해 서비스 시나리오를 통한 가이드 제시 및 연산자를 통합 배치함

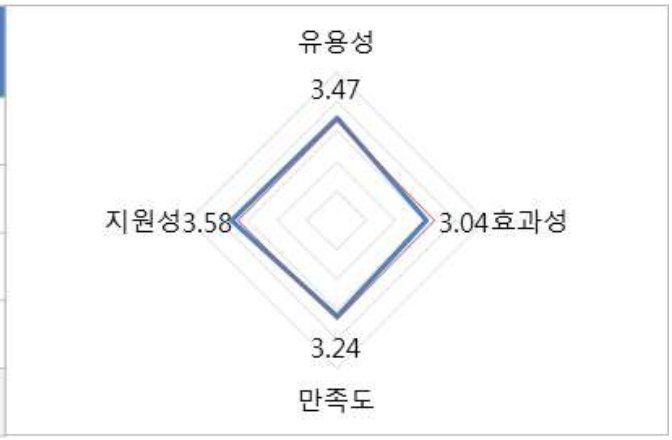


<그림 2-395> 공공 데이터 수집 스케줄러 Open API 서비스

○ 공공데이터 수집 스케줄러는 공간 빅데이터 서비스 실증 시스템에서 구현한 확장 서비스의 공공데이터를 대상으로 데이터를 주기적으로 수집하여 이력정보 및 분류 정보를 생성하여 저장하고 이를 Open API 서비스로 제공함

⑦ [단위성과 L-7] 연구성과물 성능검증 지원 및 개선방안 연구 (6차연도)

카테고리	점수 (5점 척도)
유용성	3.47
효과성	3.04
만족도	3.24
지원성	3.58
전체 평균	3.33



<그림 2-396> 사용성 검증 체크리스트 전체 결과

❍ 사용성 검증 체크리스크 분석결과 지원성에 대한 만족도가 3.58로 가장 높게 나타났으며, 효과성에 대한 만족도가 3.04로 가장 낮게 나타남.



<그림 2-397> 사용성 검증 체크리스트 상세 결과

❍ 사용성 검증 결과 업무에 사용의향이 없는 사유로 검색 및 분석 기능에 대한 이해와 설명이 필요한 부분이 많아 활용에 어려움이 있거나 필요한 데이터가 존재하지 않다는 의견이 존재하였음

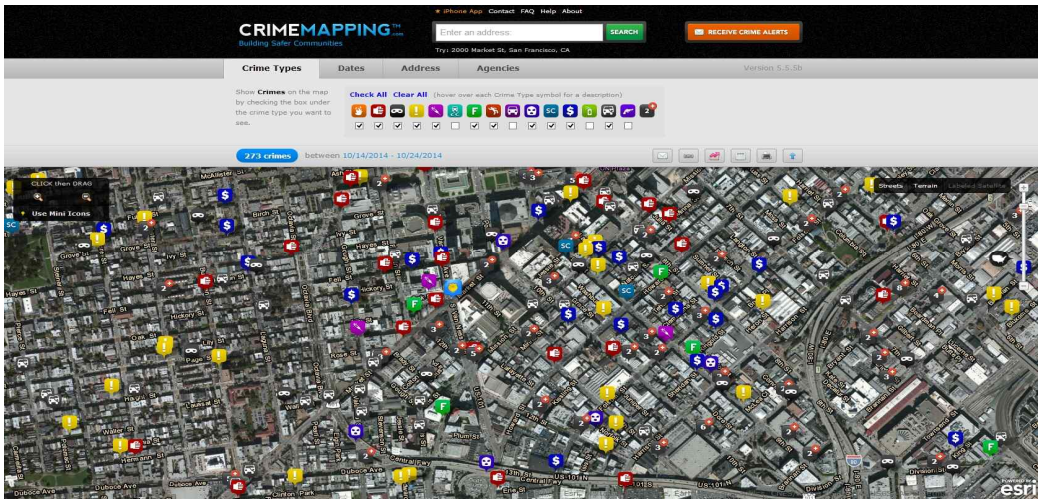
❍ 개선사항으로는 공간분석 프레임의 기능에 대한 설명창 제공, 입력기능 도우미와 같은 안내 기능 추가, 오류 및 에러 사항에 대한 피드백 인터페이스 추가 등이 도출됨

(마) 공공시스템기반 국토교통분야 (부동산, 지적, 교통 등) 공간 빅데이터 실증 서비스 시스템(핵심성과 M)

① [단위성과 M-1] 공간 빅데이터 현황 및 서비스 구축 사례 분석

○ 수치지도, 지적도, 지형도, DEM 등의 기본적인 공간정보 이외에도, 공공 및 민간 부문의 공간정보에 대한 요구가 지속적으로 증가

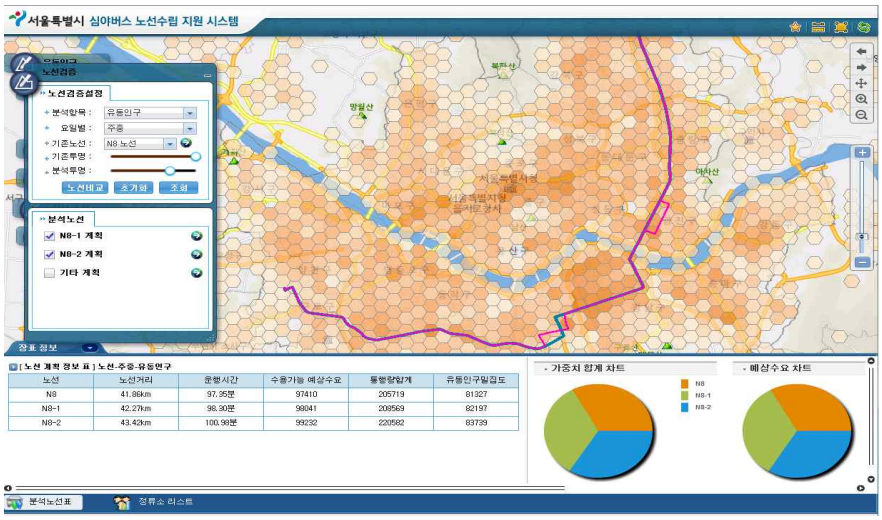
○ 공간 빅데이터 서비스 적용 국외 사례 분석



<그림 2-398> 범죄 지도(Crime Map)
(출처: http://www.crimemapping.com/map/ca/sanfrancisco)

- 샌프란시스코는 범죄 관련 데이터를 공간정보와 융합하여 범죄 지도를 작성하고 활용을 통해 범죄를 예방하고, 효율적인 경찰 인력 배치 및 순찰 시스템을 구축
- 범죄발생 현황을 범죄 유형별로 공간정보 상에 표현하고 있음.

○ 공간 빅데이터 서비스 적용 국내 사례 분석



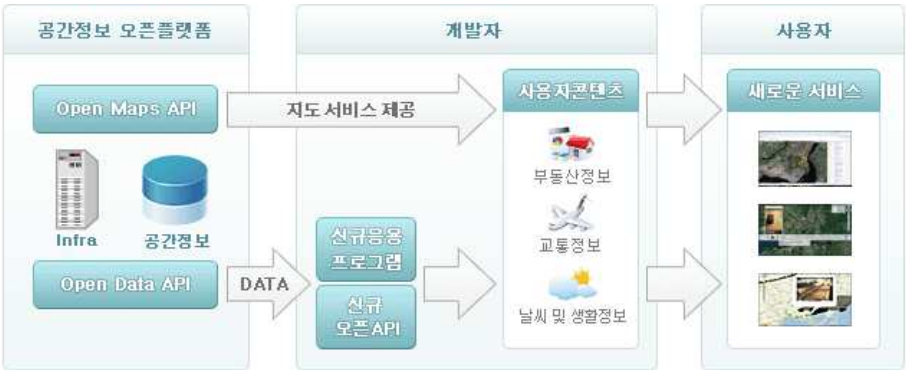
<그림 2-399> 서울시 심야버스노선수립 지원 시스템 (출처 : 한국ESRI)

- 서울시의 교통데이터와 KT의 통화량 약 30억 건의 빅데이터에 대한 유동인구 밀집도 분석을 실시하고, 요일별/시간대별 유동인구에 대한 공간패턴 분석

② [단위성과 M-2] 공간정보 오픈플랫폼 실증 기술 현황 분석

○ 공공 공간 데이터의 수요가 급증하고 이를 활용한 서비스가 확산됨에 따라 공공기관이 보유하고 있는 데이터의 개방 요구가 증가하였음.

○ 오픈플랫폼 적용 사례 분석



<그림 2-400> V World 오픈플랫폼 서비스 활용도
(출처 : V World 포털서비스)

- 일반 국민, 학교·연구소, 기업, Open API 개발자 등으로 서비스 대상자가 제한되어 있지 않으며, Cross Browsing을 지원하고 국내, 북한, 백두산, 전세계, 해저별 영상 해상도 차별 구축
- 기본 지도정보뿐만 아니라, 교통정보도, 택지개발지구도, 개발제한구역, 3D 산사태위험지도 등의 주제도 제공
- 오픈 API 개발자를 위한 3D GIS IE 플러그인 방식 오픈 제공

③ [단위성과 M-3] 빅데이터 서비스 아키텍처 및 프레임워크 설계

○ 공간빅데이터의 효율적인 구성을 위해 사용자의 요구사항을 논리적으로 분해한 것을 기반으로 각 개별 서비스에 대한 프로세스 과정을 확인하고 API를 통해 통합적으로 활용할 수 있는 구조의 아키텍처를 설계 구현

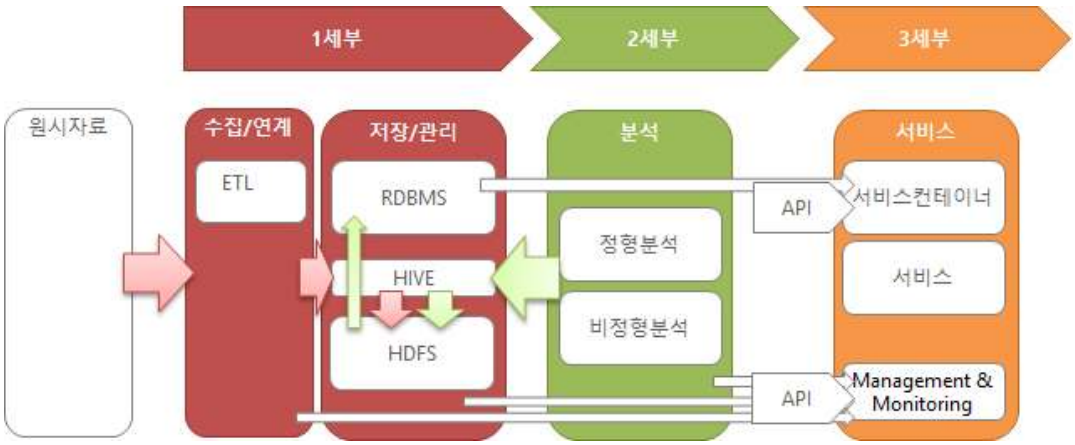
○ 서비스 아키텍처 및 프레임워크



<그림 2-401> 공간 빅데이터 실증서비스 서비스 아키텍처

- 서비스 아키텍처는 각각의 기술들이 공간빅데이터 Management를 통하여 개별기술들이 관리되고 활용되어 공간플랫폼기반위의 서비스 컨테이너를 통해 사용자들에게 지속적인 활용이 가능함
- 협업화된 형태로 지원프로그램, 컴파일러, 코드 라이브러리, 도구세트, API등과 같이 프로젝트나 솔루션 개발을 가능하게 하는 여러 가지 서로 다른 컴포넌트를 포함

○ 공간빅데이터 프로세스



<그림 2-402> 공간빅데이터 프로세스

- 공간빅데이터의 프로세스는 원시자료를 1세부 수집/연계, 저장/관리, 2세부의 분석, 3세부의 서비스의 프로세스로 구성함

- 각 세부로부터 수집/연계, 저장/관리, 분석 과정 및 결과 내용을 API를 통하여 Management & Monitoring에 정보와 LIST등을 관리자 화면에 표시함

④ [단위성과 M-4] 공간 빅데이터 서비스 시스템 실증 기술 설계

○ 서비스 시스템 실증 기술 설계

- 분석기능을 통한 융합DB화된 정보는 서비스 컨테이너에 탑재되어 사용자가 필요에 따라 사용할 수 있음

○ 서비스 컨테이너

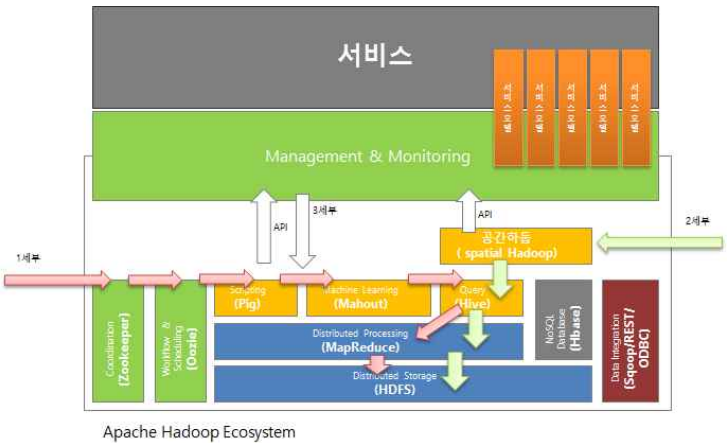


<그림 2-403> 공간빅데이터 서비스 컨테이너

- 서비스 컨테이너는 서비스모델을 분석기술로 융합DB를 생성하고 관리자가 필요에 따라 어떤 분야에 활용할 지 결정

⑤ [단위성과 M-5] 저장, 관리, 분석 가시화 등 개별 서비스 설계

○ 저장, 관리,분석 가시화 등 개별 서비스 설계



<그림 2-404> 관리 서비스 구성

- 프로토타입 화면설계 등의 개별 기술설계, 관리서비스 구조 설계

- 유즈케이스 및 기본 하둡구조를 기반으로 관리자화면 설계

⑥ [단위성과 M-6] 공간 빅데이터 서비스 통합 기술 설계

❍ 시스템 통합 기술 설계



<그림 2-405> 공간 빅데이터 서비스 통합 기술 구성도

- 공간 빅데이터의 효율적인 구성을 위해 서비스를 논리적으로 분해한 것을 기반으로 저장, 분석 데이터 및 서비스 데이터 등 각 개별 서비스에 대한 프로세스 과정을 확인할 수 있는 통합 관리 및 가시화 서비스 설계

❍ 데이터베이스 설계

테이블 ID	USERS		테이블명	사용자정보			
테이블개요	공간 빅데이터 사용자 정보 관리						
No	컬럼ID	컬럼명	타입	길이	PK	NULL	비고
1	OID	고유번호	NUMERIC		PK	NOT NULL	
2	USER_ID	아이디	VARCHAR	256		NOT NULL	
3	USER_NAME	이름	VARCHAR	256		NOT NULL	
4	USER_PSWD	비밀번호	VARCHAR	256		NOT NULL	
5	AUTHORITY	사용자 권한	NUMERIC	1		NOT NULL	
6	CREATED	생성일	DATE				
7	MODIFIED	수정일	DATE				

테이블 ID	LAYERS		테이블명	하둡 저장 레이어			
테이블개요	공간 빅데이터 하둡 시스템 저장 레이어 정보 관리						
No	컬럼ID	컬럼명	타입	길이	PK	NULL	비고
1	NAME	레이어명	VARCHAR		PK	NOT NULL	
2	FOLDER	레이어폴더	VARCHAR			NOT NULL	
3	TYPE	레이어타입	VARCHAR			NOT NULL	
4	SRID	레이어좌표계	VARCHAR			NOT NULL	
5	GEOM_COLUMN	레이어좌표	VARCHAR			NOT NULL	
6	ENVELOPE	레이어지표	VARCHAR			NOT NULL	
7	COUNT	레이어카운트	NUMERIC			NOT NULL	
8	SCHEMA	레이어스키마	VARCHAR			NOT NULL	
9	HDFS_PATH	파일시스템경로	VARCHAR			NOT NULL	

테이블 ID	METRO_CODE		테이블명	광역시도 코드			
테이블개요	광역시도 행정구역 코드						
No	컬럼ID	컬럼명	타입	길이	PK	NULL	비고
1	METRO_CD	광역시도 코드	INT		PK	NOT NULL	
2	METRO_NM	광역시도 명	VARCHAR	256		NOT NULL	

테이블 ID	SGG_CODE		테이블명	시군구 코드			
테이블개요	시군구 행정구역 코드						
No	컬럼ID	컬럼명	타입	길이	PK	NULL	비고
1	SGG_CD	시군구 코드	INT		PK	NOT NULL	
2	SGG_NM	시군구 명	VARCHAR	256		NOT NULL	
3	METRO_CD	광역시도 코드	INT			NOT NULL	
4	METRO_NM	광역시도 명	VARCHAR	256		NOT NULL	

테이블 ID	읍면동_CODE		테이블명	읍면동 코드			
테이블개요	읍면동 행정구역 코드						
No	컬럼ID	컬럼명	타입	길이	PK	NULL	비고
1	EMD_CD	읍면동 코드	INT		PK	NOT NULL	
2	EMD_NM	읍면동 명	VARCHAR	256		NOT NULL	
3	SGG_CD	시군구 코드	INT			NOT NULL	
4	SGG_NM	시군구 명	VARCHAR	256		NOT NULL	
5	METRO_CD	광역시도 코드	INT			NOT NULL	
6	METRO_NM	광역시도 명	VARCHAR	256		NOT NULL	

테이블ID	VWORLDS		테이블명	공공데이터 주제도			
테이블개요	공공데이터 주제도 리스트						
No	컬럼ID	컬럼명	타입	길이	PK	NULL	비고
1	OID	고유번호	INT		PK	NOT NULL	
2	CLASSIFY	구분	VARCHAR	256		NOT NULL	
3	NAME	한글명칭	VARCHAR	256		NOT NULL	
4	WMS_NAME	WMS 명	VARCHAR	256		NOT NULL	
5	WFS_NAME	WFS 명	VARCHAR	256		NOT NULL	
6	STYLE	스타일 명	VARCHAR	256		NOT NULL	
7	REMARK	비고	VARCHAR	10		NOT NULL	

<그림 2-406> 공간 빅데이터 공통 데이터베이스

- 공간 빅데이터 통합 기술 시스템에 요구되는 기능 및 데이터의 효율적인 관리를 위하여 무결성, 일관성, 효율성 및 확장성 등을 고려하여 설계
- 공간 빅데이터의 효율적인 관리를 위하여 Geo Spatial 기능을 지원하고 있는 RDBMS 계열의 PostgreSQL의 사용

⑦ [단위성과 M-7] 공간 빅데이터 서비스 통합 기술 프토토타입 개발

○ 시스템 구조

<표 2-111> 공간 빅데이터 관리 서비스 구조

세부 서비스명	사용자	관리자
	세부 서비스 내용	
Dashboard	하둡 리소스 관리 및 CPU 등의 리소스 관리를 통합적으로 표출	
수집/연계 관리	수집/연계를 위한 데이터 관리, 오류 관리, 스케쥴링 관리, 수집/연계 데이터의 통계치 표출	
저장 데이터 관리	저장된 데이터 리스트, 데이터 분류별 통계치 표출	
분석 관리	분석 데이터 리스트, 분석 기능별 통계치 표출하고 추후 분석 기능의 스케쥴 기능의 가능성 검토	
서비스 연계 관리	서비스 데이터 리스트, 서비스 모델 관리(Gis 서버 정보, 워크스페이스 리스트, 데이터스토어 리스트, 레이어리스트 등), 서비스 통계치 등의 표출	
사용자 관리	사용자 리스트, 권한 관리 등의 사용자 관리 수행	

- 정부, 지자체, 공공기관 등에서 제공하는 공공데이터의 서비스 연계 방안 및 데이터 처리 기술 설계
- 공간 빅데이터의 프로세스는 원시자료를 1세부 수집/연계, 저장/관리, 2세부의 분석, 3세부의 서비스의 프로세스로 구성함

<표 2-112> 공간 빅데이터 실증 시각화 서비스 구조

세부 서비스명	사용자	관리자
	세부 서비스 내용	
지도 기본 서비스	전체 검색, 명칭 검색, 지번 검색, 도로명 검색 등의 기본 검색 기능	
공공데이터	146종의 공공데이터 매쉬업하여 주제도 표출	
공간 연산 기능	Distance, Intersection, Dissolve, Merge, Convexhull, SpatialJoin, Clip, Buffer 등의 기본적인 공간 연산 기능 연계를 통한 기본 공간 연산 기능 구현	
분석 모듈 연계	장소정보에서 위치정보 추출하는 공간 빅데이터 정합 모듈 연계,2세부에서 개발하는 공간 감성어 사전 구축 기술을 통한 장소 선호도 기반 주제도 분석 모듈, 장소 선호도 산출 모듈 등의 다양한 분석 모듈 연계하여 실증 시각화가 가능한 기능 구현	
서비스 모델	정치, 융합, 기술 등의 다양한 서비스 모델을 선정하여 센서스 및 국회의원선거 분석 서비스, 사회 취약 계층과 관련된 분석 서비스, 공간빅데이터 특허 관련 데이터 분석 서비스 등 서비스 모델의 실증을 수행하여 서비스 확장 기술 구현	

⑧ [단위성과 M-8] 공간 빅데이터 시스템 실증 기술 개발

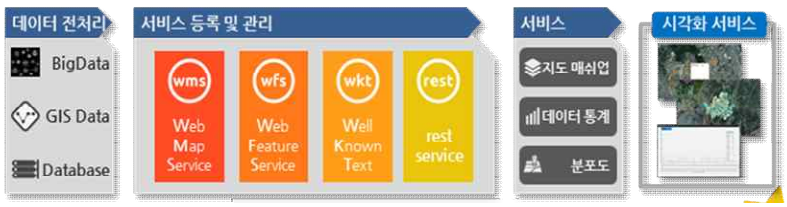
○ 공간 빅데이터 서비스 연계 구성



<그림 2-407> 공간 빅데이터 서비스 연계 구성

- 공통 개발 환경을 구성하고 컨테이너 방식의 세부별 서비스를 구성하여 연계시스템 기반 마련
- 하드웨어 구성, 소프트웨어 구성을 통한 시스템 통합 아키텍처를 설계하여 시스템 통합 및 서비스 연계를 위한 기반기술 마련

○ 공간 빅데이터 실증 시각화 서비스 구성



<그림 2-408> 공간 빅데이터 실증 시각화 서비스 구성

- 공간 빅데이터 실증 시각화 서비스를 구축하여 체계적으로 공간 빅데이터 연계 기술의 실증을 수행
- 공간 정보 검색, 분석 기능 연계 등 공간 빅데이터 실증 및 가시화 기술 개발

⑨ [단위성과 M-9] 저장,관리,분석,가시화 등 개별서비스 기술 개발

○ 공간 빅데이터 관리 서비스 구성



<그림 2-409> 공간 빅데이터 관리 서비스 구성

- 공간 빅데이터의 대시보드 기술 개발을 통해 리소스 모니터링 기능 제공
- 공간 빅데이터 저장 관리 기술 개발을 통해 저장된 데이터의 분류 및 통계 등의 관리 기능 제공
- 연계된 공간 연산 및 공간 분석 컴포넌트의 통계 모니터링 기능 제공
- 서비스 모델의 목록, GIS 서버 로그 등의 GIS 서비스 관리 기능 제공

○ 공간 빅데이터 관리 서비스 패키지

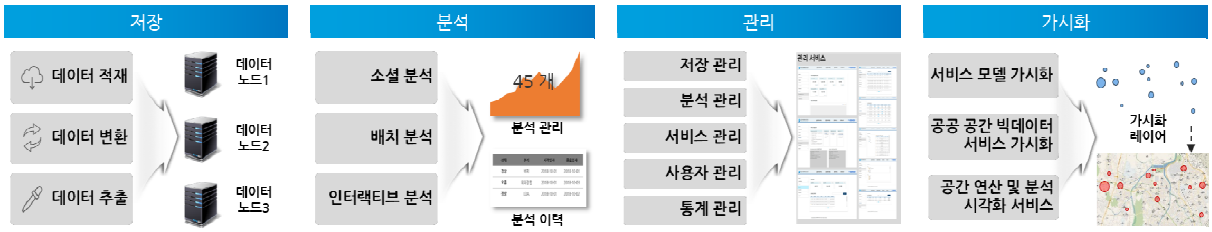
패키지명	클래스명	설명
geosbigdata.platform.mng.web	CollaboController	관리자 페이지 내 수집/연계 관리 페이지 Contoller
	StorageController	관리자 페이지 내 저장 관리 페이지 Contoller
	AnalysisController	관리자 페이지 내 분석 관리 페이지 Contoller
	ServiceController	관리자 페이지 내 서비스 관리 페이지 Contoller
	UserController	관리자 페이지 내 사용자 관리 페이지 Contoller
geosbigdata.platform.mng.util	App	연계된 하둡 Ambari 데이터를 사용가능한 Json 형식으로 변환
	Common	API를 활용한 하둡 Ambari 데이터 연계
	Timer	일정 시간 간격을 위해 현재 클라이언트 시간과 서버 시간의 시간 계산 수행
geosbigdata.platform.mng.service	DashVO	대시보드 페이지 내 리소스 관리를 위한 대시보드 VO
geosbigdata.platform.mng.service.coll	CollaboVO	수집 연계 관리를 위한 수집 연계 VO
	CollaboService	수집 연계 관리를 위한 수집 연계 Service 인터페이스
geosbigdata.platform.mng.service.coll.impl	CollaboServiceImpl	수집 연계 관리를 위한 수집 연계 Service
	CollaboServiceDAO	수집 연계 관리를 위해 데이터 조회, 수정 등 연계하는 수집 연계 Service DAO
geosbigdata.platform.mng.service.storage	StorageVO	저장 관리를 위한 저장 데이터 연계 VO
	StorageService	저장 데이터 관리를 위한 저장 관리 Service 인터페이스
geosbigdata.platform.mng.service.storage.impl	StorageServiceImpl	저장 데이터 관리를 위한 저장 관리 Service
	StorageServiceDAO	저장 데이터 관리를 위해 데이터 조회, 수정 등 연계하는 저장 관리 Service DAO

패키지명	클래스명	설명
geosbigdata.platform.mng.service.analysis	AnalysisVO	분석 관리를 위한 분석 데이터 연계 VO
	AnalysisService	분석 데이터 관리를 위한 분석 관리 Service 인터페이스
geosbigdata.platform.mng.service.analysis.impl	AnalysisServiceImpl	분석 데이터 관리를 위한 분석 관리 Service
	AnalysisServiceDAO	분석 관리를 위해 분석 데이터 조회, 수정 등 연계하는 분석 관리 Service DAO
geosbigdata.platform.mng.service.service	ServiceVO	서비스 관리를 위한 서비스 데이터 연계 VO
	ServiceService	서비스 데이터 관리를 위한 서비스 관리 Service 인터페이스
geosbigdata.platform.mng.service.service.impl	ServiceServiceImpl	서비스 데이터 관리를 위한 서비스 관리 Service
	ServiceServiceDAO	서비스 관리를 위해 서비스 데이터 조회, 수정 등 연계하는 서비스 관리 Service DAO
geosbigdata.platform.mng.service.user	UserVO	사용자 관리를 위한 사용자 데이터 VO
	LoginVO	로그인 관리를 위한 로그인 데이터 VO
	UserService	사용자 관리를 위한 사용자 Service 인터페이스
	LoginService	로그인 관리를 위한 로그인 Service 인터페이스
	CustomAuthenticationSuccessHandler	로그인 성공 시 권한별 관리를 위한 Handler
	CustomAuthenticationFilureHandler	로그인 실패 시 권한별 관리를 위한 Handler
geosbigdata.platform.mng.service.user.impl	UserServiceImpl	사용자 관리를 위한 사용자 관리 Service
	LoginServiceImpl	로그인 관리를 위한 로그인 관리 Service
	UserServiceDAO	사용자 관리를 위해 사용자 데이터 조회, 수정 등 연계하는 사용자 관리 Service DAO
	LoginServiceDAO	로그인 관리를 위해 로그인 데이터 조회, 수정 등 연계하는 로그인 관리 Service DAO

<그림 2-410> 공간 빅데이터 관리 패키지

- 저장, 관리, 분석 가시화 등의 개별 서비스 개발을 위해 패키지를 정의 및 구축
- 저장관리 컨트롤러, 대시보드 컨트롤러, 분석 컨트롤러 등 각 개별 서비스에 컨트롤러 등록

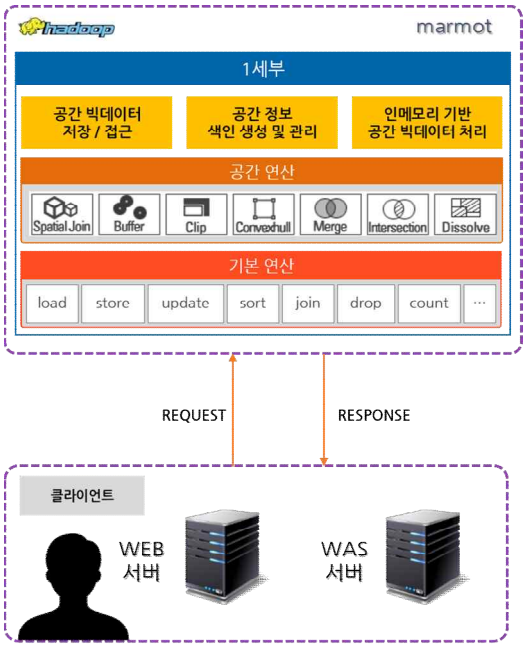
○ 저장, 관리, 분석 가시화 등 개별 서비스 기술 구성



<그림 2-411> 저장, 관리, 분석, 가시화 등 개별 서비스 기술 구성도

- 관리서비스 개발을 통해 하둡시스템 데이터 모니터링, 분석 컴포넌트 연계를 통한 분석 관리, 서비스 모델 연계를 통한 서비스 관리 및 사용자, 통계 관리 등의 개별 서비스 기술 개발 수행
- 타세부에서 개발한 공간연산, 공간분석 기술을 연계하여 개별 서비스로 발행하고 공간 연산, 공간 분석 결과정보를 지도상에 시각화하여 공간 빅데이터 분석기능의 가시화 기술 개발 수행

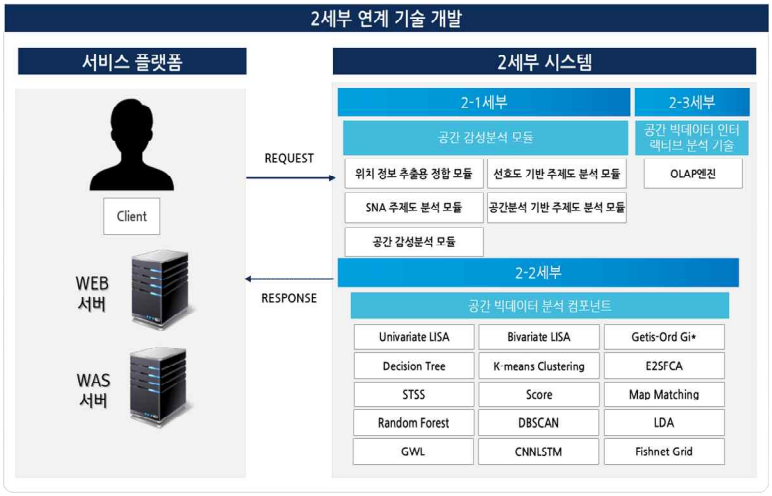
○ 1 세부 연계 기능 개발



<그림 2-412> 1세부 연계 구성도

- 1세부의 공간 빅데이터 저장/접근 기능, 공간 정보 색인 생성 및 관리 기능, 인메모리 기반 공간 빅데이터 처리 등과 공간 연산, 기본 연산 기능을 탑재한 marmot 시스템과의 연계 기능 개발

○ 2 세부 연계 기능 개발

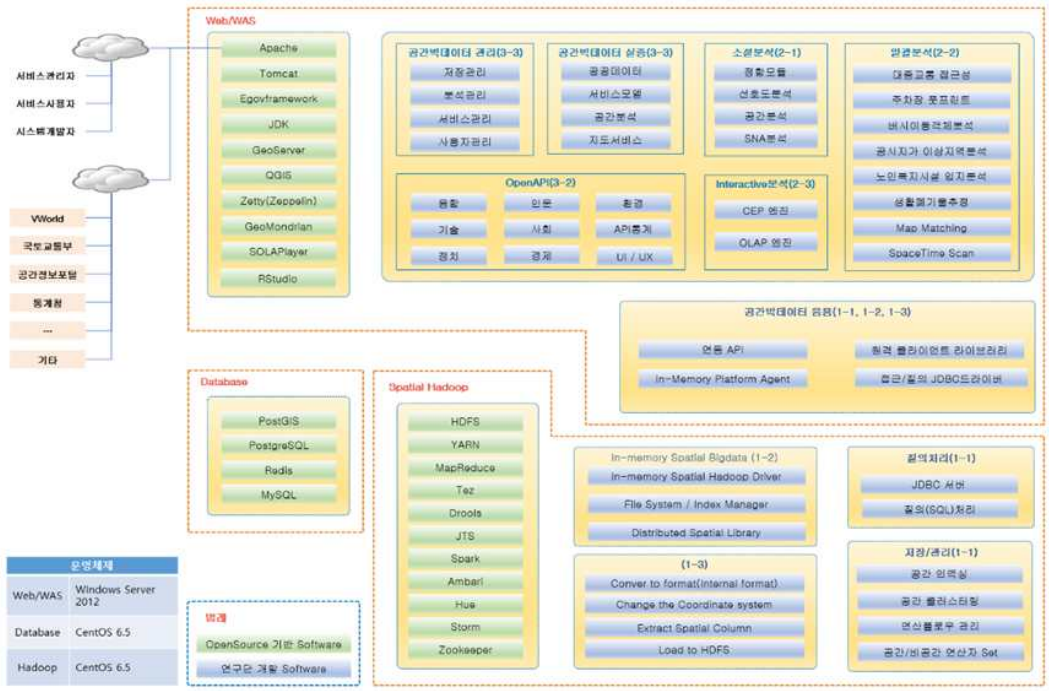


<그림 2-413> 2세부 연계 구성도

- 2-1세부의 공간 감성분석 모듈, 2-2세부의 공간 빅데이터 분석 컴포넌트, 2-3세부의 OLAP엔진 연계 기능 개발

⑩ [단위성과 M-10] 공간 빅데이터 서비스 통합 기술 개발

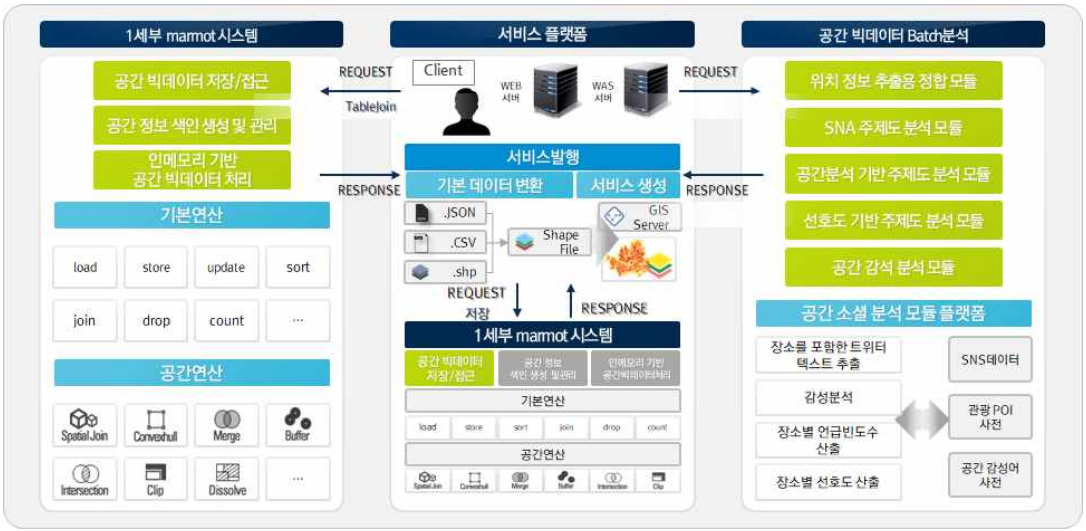
❍ 통합 모듈 구성도



<그림 2-414> 통합 모듈 구성도

- WEB/WAS : 웹서버를 통해 서비스하는 개별 서비스 모듈
- Spatial Hadoop : HSFS 기반 소프트웨어 모듈
- Database : 통합 연계 DB 구성

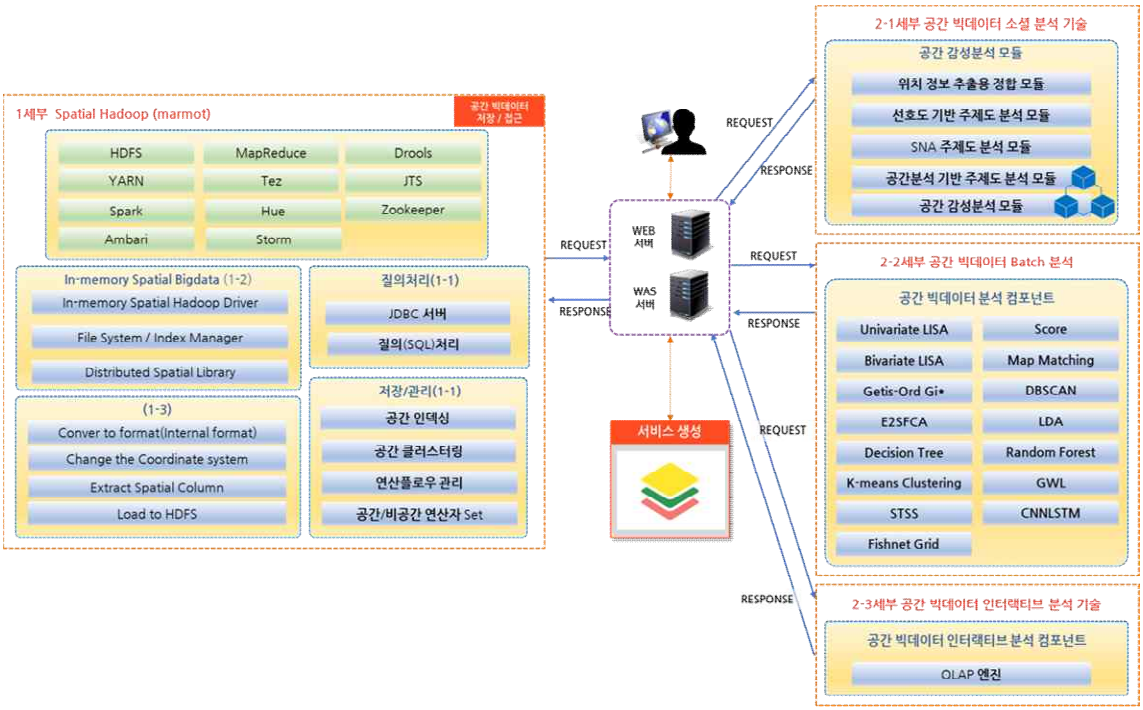
❍ 통합 연계 흐름도



<그림 2-415> 통합 연계 흐름도

- 1세부 Mamot 시스템 : 분석기능 요청 및 HDFS 연계 모듈 제어
- 2세부 Spack 분석 : 분석기능 요청 및 Spark 연계 모듈 제어
- 서비스 플랫폼 : User Interface 구성, 세부별 모듈 호출 및 분석결과 가시화

○ 통합 연계 기능 개발



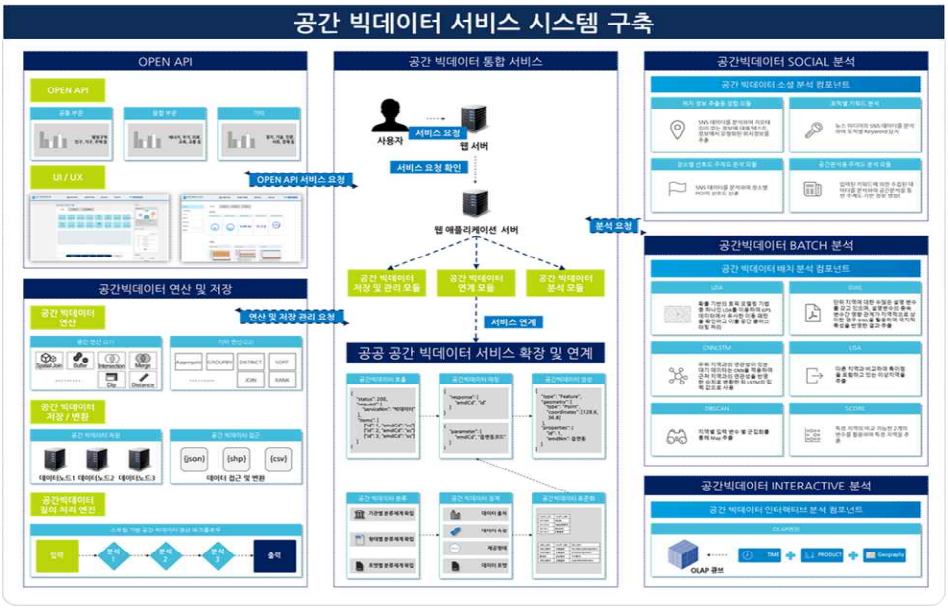
<그림 2-416> 통합 연계 기능 구성도

- 1세부의 공간 빅데이터 저장/접근, 공간 연산 기능과 2-1세부의 공간 감성분석 모듈, 2-2세부의 공간 빅데이터 배치 분석, 2-3세부의 공간 빅데이터 인터랙티브 분석 기술을 통합적으로 연계
- 2-1세세부의 공간 분석 모듈을 이용한 특정 지역의 감성 분석 처리 후 1세부 공

간 연산을 통해 격자 형식의 결과를 도출하는 등 일련의 시나리오별 처리 프로세스 기술 개발

- 모듈단위로 탑재한 세부연계 기술을 효율적으로 사용하고 분석할 수 있는 공간분석프레임 설계

○ 공간 빅데이터 통합 서비스



<그림 2-417> 공간빅데이터 통합 서비스 최종목표

- 1세부의 공간 및 비공간 연산, 공간 조인 등의 기능을 연계하여 공간 분석 프레임의 공간 연산 기능 개발

- 2-1세세부의 위치정합기술, 주제도분석 기술 등의 기능을 연계하여 공간 분석 프레임의 소셜분석 기능 개발

- 2-2세세부의 통계 및 비통계 등의 배치 분석 등의 기능을 연계하여 공간 분석 프레임의 배치분석 기능 개발

- 타세부에서 도출되는 공간 빅데이터 저장, 관리, 분석 및 가시화 등의 개별 서비스를 연계 및 통합하여 하나의 통합관리 시스템에서 관리하기 위한 통합기술을 개발하고 공간 빅데이터 서비스 시스템 구축을 위한 기반 마련

⑪ 단위성과 M-11] 공간 빅데이터 서비스 시스템 구축

❍ 공간 빅데이터 시스템 구축



<그림 2-418> 공간빅데이터 서비스 실증 시스템

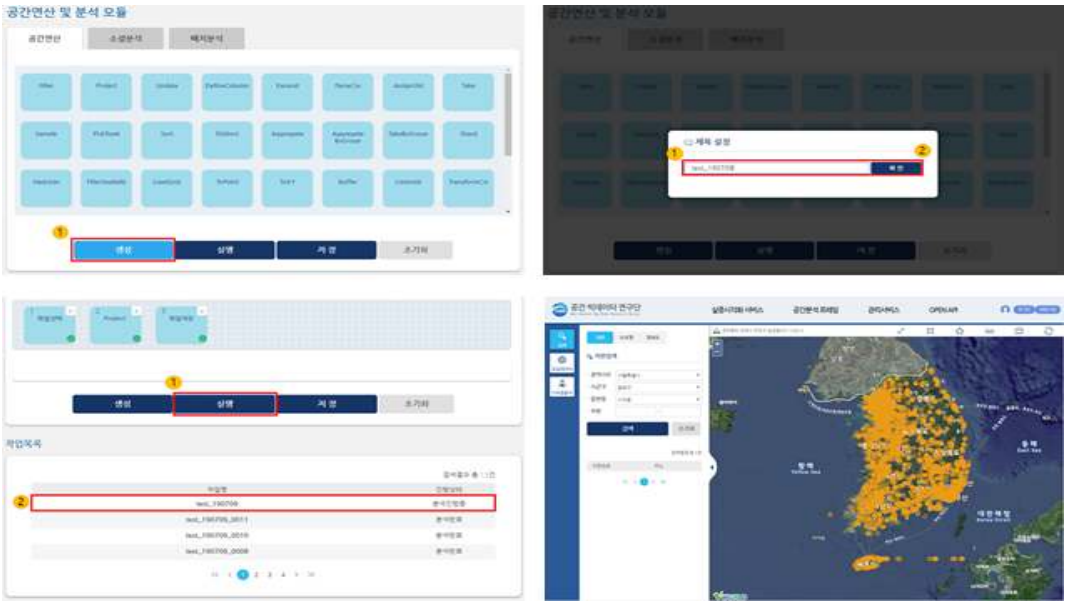
- 통합된 공간 연산 및 분석 기능과 공공 데이터 또는 로컬 데이터를 활용하여 공간 정보를 분석, 관리 할 수 있는 공간 빅데이터 실증 시스템 구축
- 품질 인증(GS) 수행 과정을 거쳐 안정성 높은 시스템 제공
- 공간 정보 기반 공간 빅데이터 서비스 실증 시스템을 이용하여 보다 쉽게 공간 정보를 활용하고 가공할 수 있는 직관적인 인터페이스 제공으로 공공 및 민간에서 활용 가능하도록 통합시스템 구축

❍ 공간 분석 프레임



<그림 2-419> 공간 분석 프레임 프로세스

- 공간 빅데이터의 연산 및 분석, 저장 데이터와 분석 상태 등의 기능을 이용할 수 있는 공간 분석 프레임 구축



<그림 2-420> 공간 연산 기능 구성 화면

- 공간 분석 프레임은 저장된 데이터셋을 호출하여 입력 데이터로 사용할 수 있으며 또한 사용자로부터 파일을 입력받아 처리 할 수 있음.
- 입·출력 파일의 포맷은 SHP, CSV, JSON 등 정형데이터 및 비정형 데이터로 설정 가능하며, 각각의 연산, 분석 모듈 별로 필요한 설정 값들을 함께 입력 받음
- 사용자로부터 입력받은 데이터를 공간 분석 프레임에서 분석하여 해당 결과를 실증 시각화 서비스에서 가시화 시키는 '입력-분석-저장-출력' 일련의 프로세스를 구축하였으며, 사용자는 이러한 공간 분석 프레임을 사용하여 다양한 공간 연산 기능 및 공간 분석 기능 활용하여 자신만의 서비스 모델을 개발할 수 있음

(바) 공공시스템 기반 타 부처 공간 빅데이터 실증 서비스 시스템(핵심성과 N)

① 단위성과 N-1] 국내 공공 공간 빅데이터 구축 및 시각화 기술 현황 분석

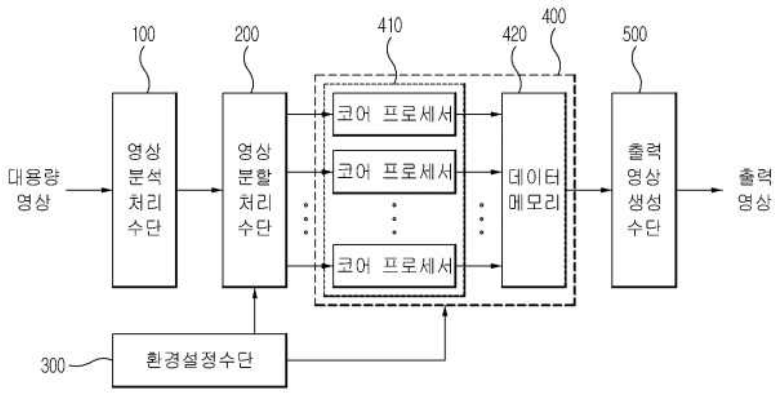
❍ 정부에서 생성하는 공간정보 기반의 행정, 통계 등의 빅데이터를 공간빅데이터 시스템에서 활용가능하도록 서비스 시스템 확장 구축



<그림 2-421> VWORLD 융합 기반 빅데이터 서비스 시스템 개요

- 정부에서 공간정보와 연계해 생성한 공공 빅데이터를 활용 할 수 있는 연계 기술 개발을 통해 빅데이터 시스템의 활용성과 공공성 향상
- 공간 빅데이터 실증 서비스 시스템에서 제공하는 동일한 수준의 서비스를 타 부처의 빅데이터를 활용하여 제공 할 수 있도록 시스템 확장 구축

❍ 시각화 기술 현황 분석 - LOD 영상 병렬 처리



<그림 2-422> LOD 영상 병렬처리 개념도

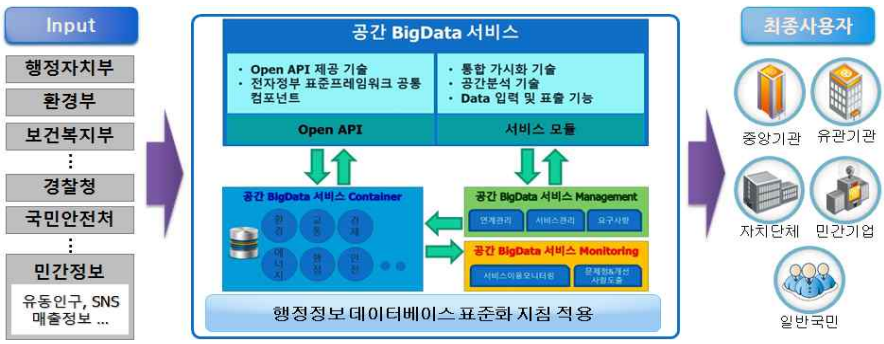
- 대용량 영상을 다수의 코어 프로세서를 통해 병렬방식으로 LOD 영상 생성
- 자원의 효율적인 이용으로 신속한 LOD 영상 생성처리 가능

○ GeoWeb, GeoCloud

- 개별 기기에서 프로그램을 설치하고 저장하던 방식에서 진화하여 네트워크를 통해 공간자원을 할당하는 기술
- 대량의 공간 빅데이터의 수집·분석 결과를 제공하기 위한 기반 기술

② [단위성과 N-2] 공공 공간 빅데이터 서비스 연계 및 확장 기술 설계(1)

○ 국토교통분야에 대한 서비스를 개발하기 위한 프로토타입 BASE 구성

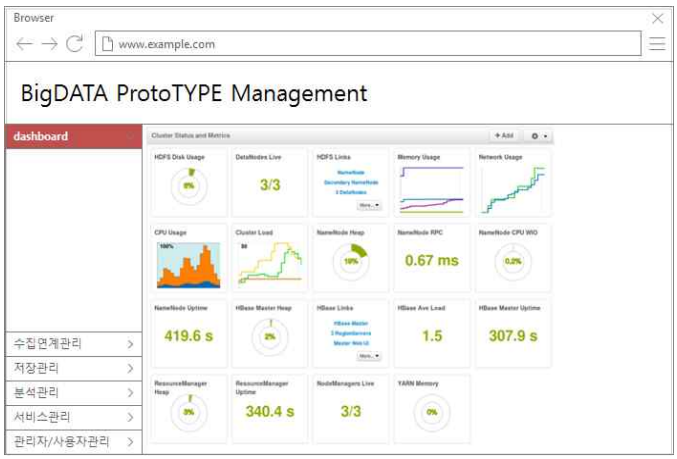


<그림 2-423> 공공 공간 빅데이터 설계 개념도

- 국토교통분야에 대한 서비스를 개발하기 위한 사용자와 필요데이터에 대한 정의, 주요 핵심 DB에 대한 설계, 테이블에 대한 기본 설계
- 설계 연계 및 확장을 위하여 서비스 컨테이너의 꾸준한 추가를 위하여 분석된 자료와 충분한 확인 과정 수행
- 이를 위하여 관리자의 확인 과정과 메타데이터 및 분석된 내용의 정확성 및 연계성 확인

③ [단위성과 N-3] 공공 공간 빅데이터 서비스 연계 및 확장 기술 프로토타입 개발

○ 공공 공간 빅데이터의 가시화화 구현



<그림 2-424> 프로토 타입 모니터링/관리 화면설계

- 공간빅데이터 서비스 연계를 위해 기본적인 프로토타입을 구성하기 위해 유즈케이스, DB설계, 테이블설계, 화면설계 등을 통하여 기본적인 프로토타입을 구성하는 일련의 과정 설계
- 임의로 분석영역 및 분석도구를 설정하여 빅데이터 결과를 한정 할 수 있도록 분석 툴 제공

④ [단위성과 N-4] 공공 공간 빅데이터 서비스 연계 및 확장 기술 설계(2)

○ 공공 공간 빅데이터의 서비스 구조 설계



<그림 2-425> 시스템 메뉴구조 설계도

- 통합적으로 활용 가능한 가시화 모듈을 설계 및 구현하여 공공 공간 빅데이터의 연계 및 확장 서비스의 데이터 시각화를 효과적으로 할 수 있도록 구조 설계

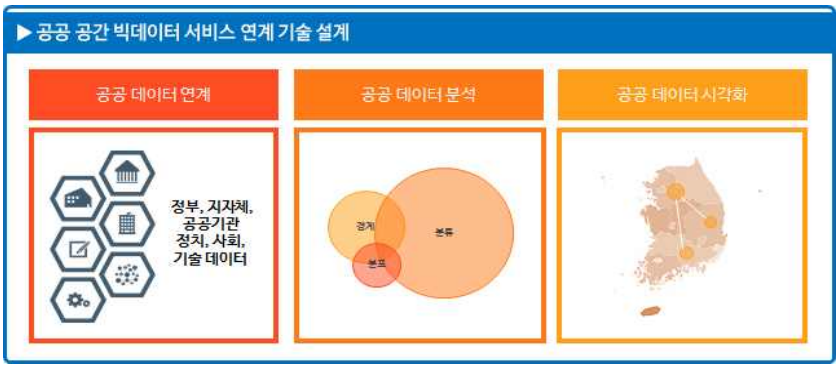
○ 공공 공간 빅데이터 서비스 연계

<표 2-113> 세부 서비스 연계 정의

구분	내용
정치 부문 서비스	국회의원선거, 가구, 주택, 인구, 선거구, SNS 등 정부 정책 및 정치와 관련된 데이터 및 요인변수들 간의 관계 분석, 시각화 서비스 제공
융합 부문 서비스	에너지, 의료, 교육, 주거, 교통 등 사회 취약 계층과 관련된 다양한 분야의 데이터를 융합하여 비교분석할 수 있도록 하는 시각화 서비스 제공
기술 부문 서비스	공간 빅데이터 기반 특허 정보를 공간/비공간 측면에서 비교 분석하는 서비스를 제공함으로써, 인용 특허 기관별 연관성을 파악
서비스 관리	서비스 모델과 데이터 관리 및 서비스 발행과 같은 기능을 지원함으로써 데이터 시각화 및 분석 처리에 최적화된 서비스 제공

- 지자체, 공공기관 등에서 제공하는 공공데이터의 서비스 연계 방안 및 데이터 처리 기술 설계
- 교육, 행정, 재정 등의 공공 데이터 분석을 통해 공공 공간 빅데이터와 접목할 수 있는 서비스 연계 방안 채택 및 연계 기술 설계

○ 공공 공간 빅데이터 서비스 연계 기술 설계



<그림 2-426> 공공 공간 빅데이터 서비스 및 확장 기술 프로세스

- WMS, WFS, WKT와 같은 OGC에서 정의한 표준 GIS 포맷 등 자바스크립트 상에서 지형지물 표현에 최적화된 개방형 표준 형식의 연계기술 설계
- 시나리오별 적용가능한 서비스 모델에 대한 기초 설계
- 공공 데이터와의 연계를 통해 향후 공공기관 데이터 뿐 아니라 민간 데이터와도 연계 할 수 있는 개방형 기술을 선택

⑤ [단위성과 N-5] 공공 공간 빅데이터 서비스 연계 및 확장 기술 개발

- 공공 공간 빅데이터 서비스 연계 및 확장 기술 개발



<그림 2-427> 공공 공간 빅데이터 서비스 연계 및 확장 기술 개발

- 공공 공간 데이터의 분석 및 시각화 처리를 위한 GIS기반의 데이터 처리 기술 개발

- 통계청과 행정안전부가 제공 중인 공간 정보 및 일반 통계 정보를 API 서비스를 통해 연계하고 이를 분석데이터로 활용할 뿐만 아니라 공간정보로 서비스 할 수 있도록 공용 기술 개발 수행

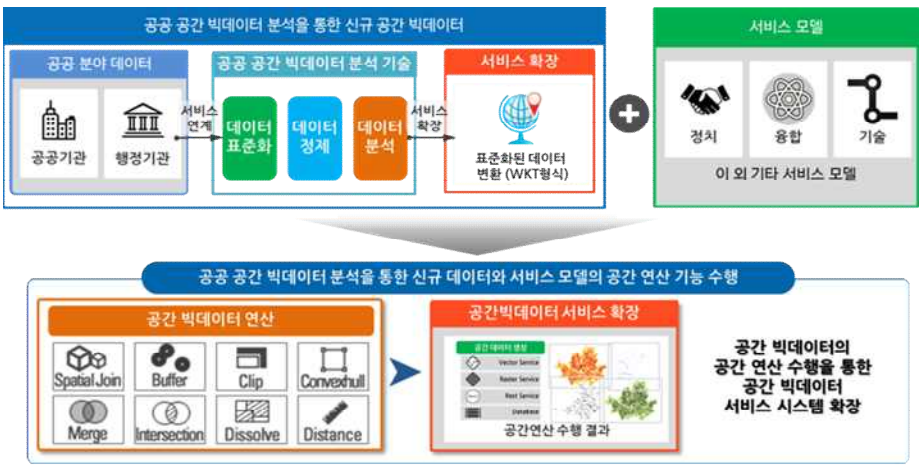
○ 공공 공간 빅데이터 및 서비스 분석 기법



<그림 2-428> 공공 공간 빅데이터 및 서비스 분석 기법

- 기관, 형태, 포맷 별로 규격화된 분류체계 확립과 이를 토대로 한 공공 공간 빅데이터의 정제 기법 연구
- 정제된 데이터에 통계, 마이닝, 네트워크, 클러스터 등 다양한 분석 기법을 적용하여 표준화된 공간 빅데이터 도출
- 데이터 변환, 분석 활용, 서비스 활용을 위한 공공 공간 빅데이터 서비스 확장 기술 개발 수행

○ 공공 공간 빅데이터 서비스 확장



<그림 2-429> 공공 공간 빅데이터 서비스 확장

- 공공 공간 빅데이터의 표준화, 정제, 분석을 통해 표준화된 데이터(WKT 형식)를 생성하고, 이를 통한 Spatial Join, Buffer, Clip 등의 공간 연산 기능을 수행함으로써 공공 공간 빅데이터의 서비스 확장 수행

⑥ [단위성과 N-5] 공공 공간 빅데이터 서비스 연계 및 확장 기술 개발

○ 공공 공간 빅데이터 생성 및 공간 연산 수행



<그림 2-430> 공공 공간 빅데이터 빅데이터 생성 및 공간 연산 수행

- 표준화된 공공 공간 빅데이터를 공간 빅데이터 시스템에 연계하여 각종 공간연산 및 분석기술에서 입력자료로 활용할 수 있도록 표준화된 공공 공간 빅데이터 생성기술 개발

○ 공공 공간 빅데이터 서비스 연계 및 확장



<그림 2-431> 공공 공간 빅데이터 서비스 연계 및 확장

- 기존에 연계한 국토교통부, 행정안전부 이외에도 기상청, 한국전력공사, 한국도로공사, 국립중앙의료원 등 공간연산 및 분석기법에 다양하게 활용 할 수 있는 공간정보를 제공하는 다양한 부처의 공공기관 데이터 연계

○ 다양한 형식의 공간 빅데이터 분류체계 일원화



<그림 2-432> 일원화된 형식의 데이터를 표준화된 결과데이터 생성

- 공공 시스템 기반 공간 빅데이터 서비스 시스템 확장 구축의 기반 마련
- 기존 개방 데이터의 중복, 누락 및 공간정보 속성이 상이하여 서비스 활용이 어려운 문제점에 대한 표준화 가능방안 설계
- 다양한 형식의 공공 공간 빅데이터 분류 체계를 일원화하여 서비스 기술 개발 수행

○ 표준화된 공용 데이터포맷(GeoJson) 생성 및 표출

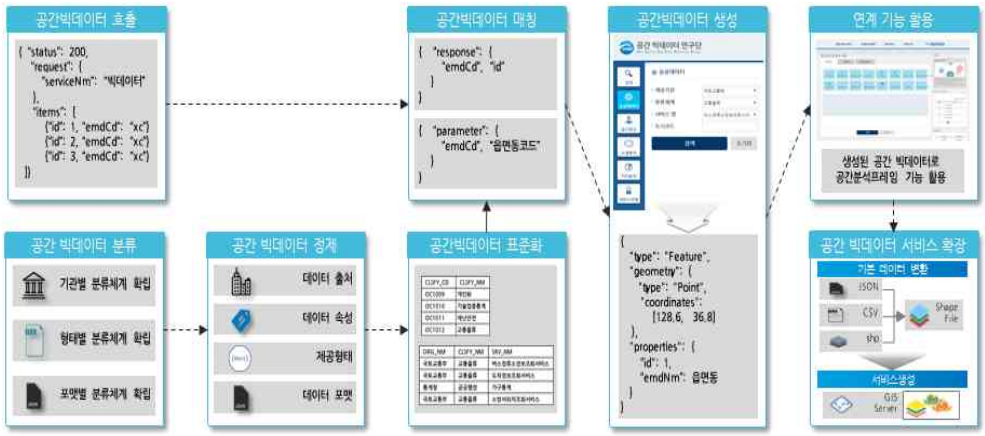


<그림 2-433> 공용 포맷 데이터 생성

- Point, LineString, Polygon, Geometry Collection 등 다양한 Geometry 유형을 지원하는 공용 포맷 데이터 생성

⑦ [단위성과 N-7] 공공 시스템 기반 공간 빅데이터 서비스 시스템 확장 구축

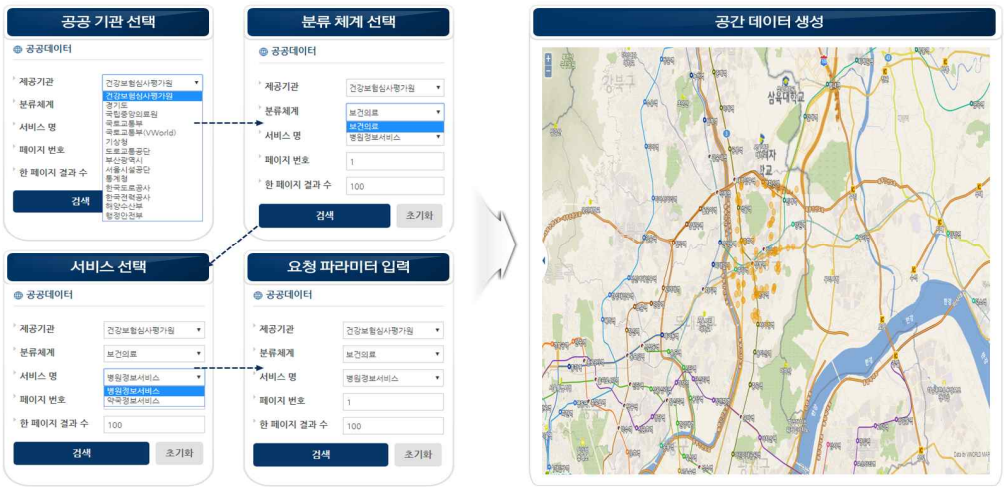
○ 공공 공간 빅데이터 서비스 연계 실증 시각화



<그림 2-434> 공간 빅데이터 서비스 시스템 확장 구축

- 기존 연계한 데이터를 기반으로 공공 공간 빅데이터 구축
- 사용자가 선택한 공공 공간 빅데이터를 표준화, 정제, 분석을 통해 공용포맷 형태로 데이터를 구축하고 이를 활용하여 서비스 연계가 가능한 시스템으로 확장

○ 공공 공간 빅데이터 서비스 연계 데이터 생성



<그림 2-435> 서비스 연계 실증 시각화

- 공공 공간 빅데이터를 사용자가 분류체계, 서비스 선택, 요청파라미터 입력 등의 작업을 통해 공공 공간데이터 형태로 표출
- 데이터의 처리 속도가 빠르고, 데이터 파싱이 간단하여 공간분석프레임의 공간연산, 공간 분석 기능 수행 시 광범위한 활용 가능
- 향후 다른 공공 공간 빅데이터와의 확장이 용이하고 기존에 제공하지 못했던 다양한 종류의 공간 빅데이터 서비스를 제공하는 기반 구축

나. 연구개발 성과

(1) 최종 목표에 따른 연구개발 성과

○ 본 연구는 “저비용, 고성능, 고확장성의 공간 빅데이터(Geospatial Big Data)의 저장, 관리, 분석 및 서비스 제공을 위한 핵심기술과 이를 구현한 공간 빅데이터 관리, 분석 및 서비스 플랫폼 개발”을 최종 목표로 설정

○ 최종 목표인 “저비용, 고성능, 고확장성”에 대한 정량적 목표치로 1) 기존방식대비 데이터 처리속도가 최소 30% 이상 향상된 페타바이트 이상의 공간 빅데이터 저장, 관리, 처리를 위한 Spatial Hadoop & Analytics 기술 개발 2) 기존 공간 빅데이터 시스템 구축비용 대비 최소 50% 이상 저렴한 공간 빅데이터 SW (Spatial Hadoop & Analytics) 기술 개발로 설정

○ 공간 빅데이터 저장관리 인프라 기술 개발, 공간 빅데이터 분석 및 가시화 기술 개발, 공간 빅데이터 서비스 활성화 및 실증 연구를 통하여 정량적 목표치에 상응하는 저비용, 고성능, 고확장성을 위한 공간 빅데이터 서비스 시스템을 구축 및 실증함

○ 세부별 핵심성과에 따라 최종목표에 대응하는 연구개발성과는 <표 2-114>와 같음

<표 2-114> 핵심성과별 최종목표와 연구개발내용

구분	핵심성과	최종목표와의 관계	연구개발내용
1세부	• Hadoop 기반 공간 빅데이터 저장 프레임워크	저비용, 고성능	• Hadoop 플랫폼에서 공간 빅데이터의 분석 및 응용 서비스를 개발할 수 있게 지원하는 공간 빅데이터 저장/탐색/검색 API와 저장관리 엔진 및 도구들을 개발 • 대용량 공간 빅데이터를 기존 방식 대비 30% 이상 향상된 성능으로 저장 관리할 수 있는 기술을 개발
	• 공간 빅데이터 질의처리 엔진	고확장성	• 공간 빅데이터 저장 프레임워크에 저장된 공간 빅데이터에 대한 질의 처리를 지원하는 질의처리 엔진 및 이를 응용프로그램에서 접근하여 사용할 수 있는 응용프로그램 인터페이스를 개발 • gRPC를 이용한 원격 인터페이스를 지원하여 Java, Python, C# 등 여러 프로그래밍 언어로 공간 빅데이터 저장시스템의 기능을 사용 지원
	• 스트림 기반 분산 공간 ET	저비용, 고성능	• 배치 및 스트림 형태로 제공되는 공간 빅데이터를 신속, 정확하게 Hadoop 플랫폼에 적재하기 위해 공간 빅데이터

	L 도구		추출/변환/적재 API 및 사용자 편의성을 위해 공간 ETL 관리 도구들을 개발 • 고속으로 추가되는 대용량 로그 데이터의 적재 및 활용 방식을 개선하여 ETL 전처리과정을 분산/병렬 방식으로 실행하는 기술을 개발하여 적재 시간을 약 20%로 감소시킴
	• 인메모리 기반 공간 빅데이터 저장관리 프레임워크	고성능	• Hadoop 및 Spark 기반으로 공간 빅데이터를 인메모리상에서 효율적으로 저장 관리하고, 고속 처리 및 분석이 가능한 엔진 및 도구들을 개발 • 기존 Hadoop 기반 플랫폼에 비해 공간분석 연산시 210%, 공간색인 생성 및 조회시 각각 91%, 75%의 성능 향상
2세부	• 공간 빅데이터 소셜 분석 기술, 공간 빅데이터 분석 및 가시화 컴포넌트 구현	고성능	• 사용자가 생성한 비정형 데이터(트위터, 네이버블로그, 다음 아고라, 뉴스기사, 블로그, 트립어드바이저, 망고플레이트, 구글리뷰, Yelp 등)를 이용하여 공간 빅데이터 정합, 분포, 차트, 가시화 및 주제도 분석을 위한 알고리즘 및 이를 기반으로 공간 빅데이터 분석 및 가시화 컴포넌트를 개발 • 최신의 딥러닝 기술을 개발하여 제공함으로써 사용자들이 공간 빅데이터 분석 및 가시화를 위하여 쉽게 활용할 수 있도록 함
	• 공간 빅데이터 Batch 분석	고성능	• 공간 빅데이터를 분석하기 위한 분석 컴포넌트 설계, 최적화 및 고도화를 수행하고 Web platform에서 drag and drop 방식을 통해 사용자가 원하는 분석 시나리오를 생성할 수 있는 분석 솔루션을 구현 • 공간 빅데이터 비통계기반 분석기능의 경우 Oracle의 비통계적 공간분석기능 중 의사결정트리를 기준으로 25M records : 80Min, 100M records : 316Min 대비 130% 이상의 성능을 제공 • 연차 별 신규 분석기능 개발로 총 30개의 분석기능 제공
	• 이동객체 데이터 Batch 분석	저비용, 고성능	• 이동객체데이터를 분석하기 위한 분석 컴포넌트 설계, 최적화 및 고도화를 수행하고 Web platform에서 drag and drop 방식을 통해 사용자가 원하는 분석 시나리오를 생성할 수 있는 분석 솔루션을 구현 • 기존 공간 빅데이터 시스템 구축비용 대비 최소 50% 이상 저렴한 공간 빅데이터 SW (Spatial Analytics) 기술 개발
	• 공간 빅데이터 Interactive 분석 기술 개발	고성능, 고확장성	• 실시간 입력되는 공간 빅데이터를 대화형으로 신속 정확하게 분석할 수 있는 기술을 개발하였으며 크게 공간 빅데이터 분석 기반 복합 이벤트 처리 기술과 공간 빅데이터 기반 대화형 분석 기술로 구성되어 있음 • 공간 빅데이터에 특화된 복합 이벤트 처리 모듈 및 대화형 분석 모듈을 사용자 친화적으로 개발하여 비전문가도 손쉽게 분석 작업을 수행할 수 있도록 지원
3세부	• 공간 빅데이터 플랫폼의 융복합 활용방안 개발	고확장성	• 공간 빅데이터 관련 국내외 법제도, 정책, 추진사업 및 관련기술 검토를 통한 정책과제를 제안하여 산업 활성화 및 정책수립기반을 마련 • 공간 빅데이터 개념정립 및 서비스 시나리오 등을 주제로 SCI급 논문 2건, 일반 논문 4건 등의 게재를 통해 학술적 성과를 도출
	• 공간 빅데이터의 공공/민간 부문 핵심 서비스 모델 도출 및 제공 방안 개발	고확장성	• 공간 빅데이터 플랫폼 활용을 위한 사용자 매뉴얼 개발을 통해 공간 빅데이터 플랫폼의 활용성을 높임 • 공공/민간 다양한 데이터를 활용하여 공간 빅데이터 기술을 적용함으로써 새로운 서비스 발굴을 통한 新산업 창출 지원 가능
	• 공간 빅데이터	고확장성	• 공간 빅데이터를 표준화된 Open API 기반으로 상호 개방·공유·활용 가능하게 하고 실제 Business에 손쉽게

	확장 Open API 서비스		적용할 수 있도록 하는 기술을 개발
	• 공간 빅데이터 서비스 제공 UI/UX	고확장성	• 사용자 접근성을 고려한 상호작용 기술과 사용자의 경험을 바탕으로 효율적인 공간 빅데이터 서비스 제공을 위한 UI/UX 방안을 개발 • UI/UX 방안을 공간 빅데이터 실증 서비스 시스템에 적용하여 공간 빅데이터 서비스 사용자이 이용 편의성 증대
	• 공공시스템기반 국토교통분야(부동산, 지적, 교통 등) 공간 빅데이터 실증 서비스 시스템	저비용, 고확장성	• 1세부의 공간 빅데이터 저장/접근, 공간연산 기능, 2세부의 공간 분석 기술, 3세부의 활용방안 및 UI/UX를 통합적으로 연계한 공간 빅데이터 실증 서비스 시스템 구축을 통하여 하여 기존 공간 빅데이터 시스템 구축비용 대비 50% 이상 저렴한 고성능의 공간 빅데이터 SW (Spatial Hadoop & Analytics) 기술 개발 실증
	• 공공시스템기반 타 부처 공간 빅데이터 실증 서비스 시스템	저비용, 고확장성	• 공공 공간 빅데이터 서비스를 연계하고 공공 공간 빅데이터 정제기법 연구를 통해 생성된 표준화된 공공 공간 빅데이터를 활용하여 유연한 시나리오 생성이 가능한 공간 빅데이터 실증 서비스 시스템 확장 구축하여 기존 공간 빅데이터 시스템 구축비용 절감할 수 있는 고성능의 공간 빅데이터 SW (Spatial Hadoop & Analytics) 기술 개발 실증

(2) 공간 빅데이터 저장관리 인프라 기술 개발

(가) 공공적 성과

○ 국토교통부 재정사업 ‘공간빅데이터 체계구축사업’[8]에 직접 활용하여 공공사업에서 공간 빅데이터의 활용성을 높일 수 있는 공간 빅데이터 저장관리 기반 기술을 개발함

- 연구성과물을 공간빅데이터 사업단의 공간 빅데이터 분석 플랫폼에 직접 연동시키고 실증하여 개발한 기술이 공공사업에 활용되는 성과를 가져옴
- 연구 초기부터 기술의 적용대상 및 활용주체를 명확히 하고 단계별로 사업에 필요한 기술 위주로 요구사항을 직접 파악하여 성과물 개발을 진행하여 **사업과 긴밀하게 연계된 R&D 추진의 성공적 사례**가 됨
- 사업단 플랫폼에서 실제로 수행할 공간분석 모델에 따라 적용을 위한 기능을 비교검토하여 요구사항을 반영하고 시스템 구조를 검토하는 등 공간빅데이터 체계구축사업 추진단과 지속적인 협의를 하여왔으며 이를 통해 효과적인 성과물 적용을 가져옴

○ 공간 빅데이터를 활용하고자 하는 여러 분야의 국가/공공 사업에 기반기술로 활용할 수 있는 길을 열었음

- 공간 빅데이터를 다루는 기반기술이므로 공간빅데이터 체계구축사업 뿐만 아니라 교통, 부동산, 지적, 재난재해 등 공간 빅데이터를 활용할 필요가 있는 여러 분야의 국가/공공부문 사업에 기반 핵심기술로 적용이 가능함

○ 국토교통부를 포함한 국가 보유 데이터의 품질을 높일 수 있는 지원 기술을 개발함

- 국가공간정보포털 등을 통하여 공유되고 있는 대용량의 공간정보를 처리할 수 있는 기술을 개발하여 기존 GIS로는 수행하기 어려웠던 공간정보의 오류 탐지에 활용 가능함
- 공간 빅데이터 저장관리 시스템을 사용하여 국가중점개방데이터인 ‘건물통합정보마스터’(건수 약 1,400만여건, 크기 약 11.1Gb) 에 대한 UFID 오류, 건물영역 중첩 오류 등을 탐지할 수 있으며 이러한 분석은 대용량 데이터에 대한 공간/비공간 조인이 필요하므로 기존 GIS로는 수행하기 어려움

○ 제2차 공간정보산업진흥계획(2016~2020)의 전략적 추진과제에 활용할 수 있음

- 공간 빅데이터 저장관리 인프라는 공간 빅데이터를 다루는 기반 기술로, 제2차 공간정보산업진흥계획의 전략적 추진과제 중 ‘사물인터넷 기반 데이터 구축’ 및 ‘플랫폼 기반 공간정보 공유체계 통합 및 고도화’를 실현할 수 있음

---

8) 공간빅데이터 체계구축사업은 2019년부터 ‘공간 빅데이터 분석 플랫폼 유지관리 및 운영지원’ 으로 명칭이 변경되었으며 본 문서에서는 ‘공간 빅데이터 분석 플랫폼 사업’이라고 함

(나) 기술적 성과

○ 기존 빅데이터 저장관리 기술(Hadoop)을 고도화하여 공간정보에 특화된 공간정보 전용 빅데이터 처리 엔진을 개발하였으며 4개의 핵심성과를 개발하였음 (연구수행내용 및 성과 참조)

- [핵심성과 A] Hadoop 기반 공간 빅데이터 저장 프레임워크
- [핵심성과 B] 공간 빅데이터 질의처리 엔진
- [핵심성과 D] 스트림 기반 분산 공간 ETL 도구
- [핵심성과 O] 인메모리 기반 공간 빅데이터 저장관리 프레임워크

○ **대용량 공간 빅데이터를 기존 방식 대비 30% 이상 향상된 성능으로 저장 관리**할 수 있는 기술을 개발함

- 공간 빅데이터는 볼륨, 복잡도, 다양성이 높기 때문에 고성능, 저비용 저장관리 방법이 필수적이나, 관련 원천 기술은 글로벌 기업이 주도하고 있음
- Hadoop 기반 기술을 활용하여 공간 빅데이터의 고속 처리가 가능하고 확장성이 높은 빅데이터 플랫폼 기반 공간 빅데이터 저장·관리 기술을 개발하여 기존의 상용 공간 빅데이터 기술에 비해 처리속도를 향상하고 처리 비용을 감소시킴

○ 기존 GIS 시스템이나 공간 RDBMS로는 처리할 수 없을 정도의 **대용량 공간 데이터에 대한 분석이 가능**해짐

- 레코드 개수가 약 82.7억건이고 데이터 크기가 821Gb에 이르는 대용량데이터인 도로교통공단의 DTG 데이터(1개월 분량)를 활용하여 읍면동, 도로링크별 차량통행량과 어린이 보호구역 주변 차량 운행속도를 분석함
- 처리하는데 많은 시간과 저장공간이 필요한 공간 조인이 필요한 연산을 수행함
- 대용량의 연속지적도(14.1Gb)와 전국노인복지시설, 전국인구밀도, 행정동 데이터 사용도를 조인하여 노인 복지시설 추가설립 대상지 추천 분석을 수행함

○ 공간빅데이터 체계구축사업의 기존 플랫폼 대비 공간분석의 성능 향상을 가져옴

- (분석 대상 선정) 2017년도 성능 평가에서는 본 연구 성과물의 1차 활용처로 예상되는 공간 빅데이터 사업단에 사용되는 공간 빅데이터 엔진과, 공간 빅데이터 처리 시스템에서 가장 많이 언급되고 있는 미네소타 대학의 SpatialHadoop을 비교 대상으로 선정함
- (분석 기능 선정) 분석 기능은 공간 빅데이터 사업단에서 실제로 사용하는 공간분석 라이브러리 중 일부를 선정함. 미네소타 대학의 SpatialHadoop의 경우는 사업단의 공간 분석 라이브러리에 해당하는 공간 분석 모듈을 제공하지 않아 SpatialHadoop에서 제공하는 분석 모듈을 기준으로 성능 평가를 실시함



<그림 2-436> 공간 빅데이터 저장관리 시스템 성능평가(2017년)

- (결과 의미) 2017년의 성과물 성능평가에서 기존 **사업단 엔진 대비 공간 조인 연산 속도에서 최대 93.43% 성능 향상**을 가져왔으며 SpatialHadoop 대비 최대 300% 성능이 향상됨. 본 결과는 공간 빅데이터 사업단에서 실제로 사용하는 분석 모듈의 성능을 측정한 것이므로, 추후 연구단 성과물이 사업단 플랫폼에서 활용되는 경우 상당한 성능 향상 효과를 볼 수 있을 것으로 판단됨
- 2018년 사업단 플랫폼에서 실제로 운용중인 동일한 공간분석 모델 '10분 정책지도 테스트'를 동일한 데이터에 대하여 수행한 성능비교를 수행하였으며, 개발한 공간 빅데이터 플랫폼은 8가지의 분석 모듈에서 기존 **사업단 플랫폼 대비 약 84%의 성능 향상 효과**를 보임
- 위 성능 평가는 9대의 데이터 노드를 구성된 하둡 플랫폼에서 수행되었고, 운영체제는 CentOS 6.9 (64bit)이고, 각 노드는 32GB 메모리와 4TB 크기의 디스크가 장착되었음

○ 공간 빅데이터를 처리하기 위한 단위 기술을 하둡 환경에서 최적화하여 수행 성능이 우수한 기술을 개발함

- 하둡 환경에서 공간 빅데이터 검색 및 데이터 접근 효율화를 위한 공간 인덱스 및 분산 처리 기술을 개발하여 공간정보에 특화된 공간정보 전용 빅데이터 처리 엔진을 개발함
- 공간데이터를 비공간데이터와 동일한 방식으로 처리하는 기존 기술에 비하여 공간 인덱스 기반으로 공간데이터에 특화된 분산 데이터처리 방법을 사용하여 기존 시스템 대비 공간 빅데이터 질의처리 속도를 향상시킴
- 공간 조인 과정 중 발생하는 HDFS I/O 횟수를 최소화하여 Map/Reduce 기반 중첩 루프(Nested-Loop) 공간 조인의 성능을 개선하고 전체 공간 조인 성능을 약 2.9배 향상하는데 기여함

- 고속으로 추가되는 대용량 로그 데이터의 적재 및 활용 방식을 개선하여 ETL 전처리과정을 분산/병렬 방식으로 실행하는 기술을 개발하여 적재 시간을 약 20%로 감소시킴

○ 공간 분석 응용 개발자들이 편리하게 사용할 수 있는 지원도구를 개발하여 시스템의 활용성을 증대시킴

- gRPC를 이용한 원격 인터페이스를 지원하여 특정 언어나 특정 OS에 국한되지 않고, 다양한 언어와 운영체제에서 공간 분석 응용을 개발할 수 있는 환경을 제공하며, Java, Python, C# 등 여러 프로그래밍 언어로 공간 빅데이터 저장시스템의 기능을 사용할 수 있음
- 공간 빅데이터 분석 응용 도구를 개발하여 공간 분석 응용 개발 및 시스템 운영에 도움을 줄 수 있고 시스템의 활용성을 증대시킴
- 많은 공간 분석 개발자들이 Hadoop 환경에 익숙하지 않는 것을 고려할 때, 원격 공간 빅데이터 활용 기술은 응용 개발 편의성을 높일 수 있음

○ SQL 기반 질의언어와 표준 인터페이스를 제공하여 시스템의 기능을 쉽게 사용할 수 있는 방법을 제공함

- SQL 기반의 질의언어와 표준 JDBC 인터페이스를 지원하여 응용프로그램이 Hadoop 환경 및 저장소 사용법과 무관하게 공간 빅데이터 저장시스템의 기능을 표준적으로 쉽게 사용할 수 있는 활용성을 확보함

○ 인메모리 데이터 처리 기술을 사용한 인메모리 기반 공간 빅데이터 저장관리 프레임워크를 개발하여 공간 빅데이터 처리 성능을 개선함

- 빅데이터를 HDFS 기반으로 저장 및 처리하는 기존 기술과는 달리 빅데이터를 다수 서버의 메모리에 분산 저장 및 처리하는 시스템 구축을 통해 빈번한 I/O가 발생하여 디스크 I/O에 의존도가 높은 분석의 경우에도 성능 저하를 최소화함
- 기존 Hadoop 기반 플랫폼에 비해 공간분석 연산시 210%, 공간색인 생성 및 조회시 각각 91%, 75%의 성능 향상을 달성함
- 이동객체 자료에 대한 고속 색인 및 적재 기능을 통해 목표치인 초당 5000건을 상회하는 초당 20만건 이상의 적재 성능을 달성함

○ 모바일 데이터에 활용 가능한 이동객체 데이터 처리 플랫폼의 기반을 마련함

- 대용량의 이동객체 데이터를 처리할 수 있는 인메모리기반 공간 빅데이터 플랫폼 기술을 개발하여 공간 빅데이터 기반의 분석 시스템은 물론 향후 IoT/Mobile 분야에서 이동객체 데이터를 처리하는 분석시스템의 기반 플랫폼으로 활용할 수 있음

○ 공간분석에서 많이 사용되는 분석 기능을 하둡 환경에서도 사용할 수 있는 공간분석 라이브러리를 지원함

- 거리기반 분석, 클립분석, 핫스팟분석 등 공간분석에서 많이 사용되는 12종의 공간분석 라이브러리를 지원함

(다) 학술적 성과

○ 공간 빅데이터 저장관리 시스템을 개발하는데 필요한 핵심기술을 연구 및 개발하여 학술논문 게재/발표 및 지적재산권 확보 성과를 달성함

○ 공간 빅데이터 저장관리 시스템 핵심기술에 대한 SCI 논문 2건을 포함하여 **학술논문 8건을 게재**하여 해당 분야의 학술 발전에 기여함

- 대표적으로 하둡 환경에서 공간빅데이터의 처리를 단위연산자의 연결로 구성하고 스트림 기반으로 효율적으로 수행하는 방법에 대한 "High-Performance Geospatial Big Data Processing System Based on MapReduce" 논문을 작성하여 ISPRS International Journal of Geo-Information(SCI 학술지) 에 2018년 게재함

○ 실시간 공간 빅데이터 저장관리 핵심기술 관련하여 **24건의 학술발표논문을 발표**하여 해당 분야의 학술 교류 및 기술 확산에 기여함

- 대표적으로 하둡 및 Spark 환경에서 핫스팟 분석의 분산처리 방법 및 성능향상 효과를 제시한 "공간 통계 핫스팟 분석의 분산처리 고속화 설계" 논문을 2017년 한국컴퓨터종합학술대회에 발표하여 최우수논문상을 수상함

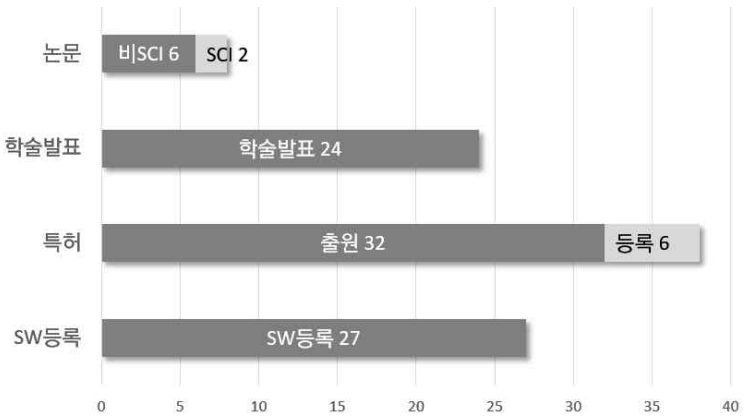
○ 공간 빅데이터 저장관리의 핵심적인 기술을 특허로 작성하여 32**건을 출원**하였으며 그중 **6건이 등록**되어 핵심 기술에 대한 지적재산권을 확보함

- 대표적으로 공간 빅데이터 처리를 기본 단위인 연산 기반으로 구성하는 핵심 방법을 제시한 "연산 체인으로 구성된 프로그램을 이용하여 맵리듀스 작업을 생성하는 방법 및 장치" 를 2018년 출원함

- 대표적으로 분산환경에서 연속적으로 입력되는 공간데이터를 시스템에 적재하는 기술을 특허로 출원한 "스트림 기반 공간 데이터를 추출, 변환 및 적재하기 위한 시스템 및 방법" 이 2018년 등록됨

○ 공간 빅데이터 저장관리 기능을 구현한 모듈 또는 시스템 **27건을 소프트웨어로 등록**하여 프로그램 지적재산권을 확보함

- 대표적으로 공간 빅데이터의 질의처리 엔진을 구현한 java 언어 프로그램인 "하둡기반 공간빅데이터 질의처리 프로그램" 을 2016년 한국저작권위원회에 등록함



<그림 2-437> 1세부 학술적 성과

(라) 경제적 성과

○ 국내 공간정보 기술 확대 및 공간정보 시장 개편을 기대함

- 세계 공간정보 시장에서 국내 기업의 시장 점유율이 상승하여 '20년에는 10%를 초과할 것으로 기대됨
- 국산 공간 ETL 핵심기술 확보를 통한 국내 공간정보 시장의 해외기술 의존도 감소를 기대함('12년 75% 이상 -> ' 25년 50% 이하)
- 전자지도 등 대용량 공간 빅데이터와 시간에 따라 빠르게 수집되는 실시간 공간 데이터를 융합한 관리 및 분석이 가능하여 빅데이터로부터 새로운 사회경제적 가치를 창출함
- 국내 공간정보 기술 확보로 외산 소프트웨어 중심의 기본 공간정보 시장을 개편하고 해외시장 진출을 기대함
- 공간정보 기반의 실시간 빅데이터 저장관리 기술 확보를 통하여 IoT 등과 연계한 빅데이터 활용 및 서비스 다양화가 예상됨
- 실시간 빅데이터의 효과적인 처리와 관리를 지원하기 위한 핵심기술로서 빅데이터 기술 선도 및 관련 산업 활성화가 기대됨

○ 미래 산업발전과 서비스 및 신시장 창출을 기대함

- 공간빅데이터 처리 기술을 기반으로 기존 정보와 융합을 통해 공간정보 분야에서 다양한 서비스 및 시장이 창출되고 새로운 일자리와 중소기업이 육성되는 산업발전 효과를 기대함
- 공간 빅데이터 플랫폼 구축으로 미래 지식기반 사회를 지향하는 선진국 대비 기술 경쟁 우위를 확보하고 공간정보 산업의 기술경쟁력을 확보할 것으로 기대함
- 무선통신, IoT, 스마트폰 등 세계적인 경쟁력을 보유한 IT 인프라와의 공간 빅데이터 처리기술 융합을 통한 독자적인 공간정보 신시장 창출 및 경쟁력 확보를 기대함
- 지도 등 대용량 공간 빅데이터와 스트림 형태의 빅데이터를 융합한 질의를 가능하게 하여 빅데이터로부터 새로운 가치 창출함

(마) 기타 성과

○ 각종 대용량 공간정보를 활용하는 공공기관의 업무 수행에 적용하여 분석기능을 지원함으로써 공공부문의 업무에 필요한 시간과 비용을 절약하여 업무 효율성을 증대시킬 수 있음

○ 타 R&D 과제에서 발생하는 공간 빅데이터의 기술적 이슈를 지원할 수 있는 기반기술을 개발하여 타 R&D 과제와 협업 성과를 가져옴

- 공간 빅데이터 저장관리 시스템을 사용하여 국가중점개방데이터인 '건물통합정보마스터'(건수 약 1,400만여건, 크기 약 11.1Gb) 에 대한 UFID 오류, 건물영역

중첩 오류 등을 탐지함으로써 데이터 전처리 및 오류수정에 효과적으로 활용하였음

- 공간정보를 다루는 최근 다른 R&D 과제들도 기존의 GIS로는 다루기 어려운 대용량의 데이터를 사용하는 추세이므로 타 R&D 과제에서 데이터 변환, 가공 및 전처리와 같은 기반 기술로서 활용할 수 있음

(바) 종합분석 및 결론

○ 공간 빅데이터 저장관리 시스템은 공간 빅데이터를 다루는 기존 시스템들이 다룰 수 없는 대용량의 데이터를 효과적으로 처리하고 분석하는 기술을 제시함으로써 공간빅데이터체계구축사업을 포함한 공간정보 관련 국가·공공 사업에서 대용량 공간데이터의 활용성을 크게 높였음

<표 2-115> 연구개발 성과 종합

<table>
<tr><th colspan="2">과제명</th><th>공공</th><th>기술</th><th>학술</th><th>경제</th><th>기타</th></tr>
<tr><td rowspan="3">1세부</td><td rowspan="3">공간빅데이터 저장관리 인프라 기술 개발</td><td>공간빅데이터 체계구축사업'을 포함한 공공 사업에서 기반기술로 활용하여 공간빅데이터의 활용성을 높임</td><td>기존 GIS 시스템이나 공간 RDBMS로는 처리할 수 없을 정도의 대용량 데이터에 대한 분석이 가능해짐</td><td>핵심기술을 연구 및 개발하여 학술논문 게재/발표 및 지적재산권 확보 성과를 달성함</td><td>공간정보 분야 해외기술의존도를 감소시키고 공간정보와 빅데이터 분야의 융합으로 서비스 시장 창출 및 산업 발전 효과가 기대됨</td><td>공간정보를 활용하는 공공기관의 업무 효율성을 증대시킬 수 있음</td></tr>
<tr><td colspan="5">종합</td></tr>
<tr><td colspan="5">공간 빅데이터를 다루는 기존 시스템들이 다룰 수 없는 대용량의 데이터를 효과적으로 처리하고 분석하는 기술을 제시함으로써 공간빅데이터체계구축사업을 포함한 공간정보 관련 국가·공공 사업에서 대용량 공간데이터의 활용성을 크게 높였음</td></tr>
</table>

(3) 공간 빅데이터 분석 및 가시화 기술 개발

(가) 공공적 성과

○ 민원과 같은 비정형 데이터 분석을 통하여 정책 결정에 대한 기본자료를 생성함으로써 데이터 기반의 과학적이고 합리적인 의사결정 지원 가능

○ 공간 빅데이터를 실시간 대화형으로 분석하여 시의적절하게 각종 분석 및 대응을 가능케 함

(나) 기술적 성과

○ 장소에 특화된 사전을 구축하기 위한 기술을 제시하고 이를 이용하여 장소 정보가 포함된 비정형 데이터로부터 장소에 대한 선호도를 분석하는 기술을 개발함

○ 최신의 딥러닝 기술을 개발하여 제공함으로써 사용자들이 공간 빅데이터 분석 및 가시화를 위하여 쉽게 활용할 수 있도록 함

- 누구나 활용할 수 있도록 개발된 성과물(모듈)을 API로 제공

○ 공간 빅데이터 및 이동객체데이터 Batch 분석 솔루션을 구현하였음

- 연차 별 신규 분석기능 개발로 총 30개의 분석기능을 제공함
- 공간 빅데이터 비통계기반 분석기능의 경우 Oracle의 비통계적 공간분석기능 중 의사결정트리를 기준으로 25M records : 80Min, 100M records : 316Min 대비 130% 이상의 성능을 제공함
- 기존 공간 빅데이터 시스템 구축비용 대비 최소 50% 이상 저렴한 공간 빅데이터 SW (Spatial Analytics) 기술 개발
- 4차년도와 5차년도에 각각 TTA인증과 KTC인증을 취득하였음

○ AWS (Amazon Web Service) 기반의 공간 빅데이터 일괄 분석 컴포넌트 시제품 2개를 제작하였음

○ 공간 빅데이터에 특화된 복합 이벤트 처리 모듈 및 대화형 분석 모듈을 사용자 친화적으로 개발하여 비전문가도 손쉽게 분석 작업을 수행할 수 있도록 함

(다) 학술적 성과

○ 특허 출원 14건, 등록 6건, 국내외 학술지 61건, 국내외 학술발표 81건을 달성함

(라) 경제적 성과

○ 사용자 개입이 적은 반자동화된 기술의 개발로 데이터 분석 비용 절감 및 유지비용 최소화

○ 딥러닝 기술이 적용된 훈련모델 생성을 위하여 어노테이션 데이터 생성을 위

한 고용창출

○ 비전문가가 최소한의 지식 및 기술로 실시간 데이터 분석을 가능하게 하여 추가적인 용역 및 자문 비용 절감

(마) 기타 성과

○ 공간 빅데이터 소셜 분석 기술 중 일부 기술(층간소음 분석, 딥러닝 기반 문장 분류 기술, 공공시설의 접근성 분석)은 연구단(3세부) 및 공간 빅데이터 체계 구축 사업(재정사업)에서 활용 및 적용

○ 이동객체데이터 비통계기반 분석 컴포넌트 개발 결과를 활용하여 택시 기사를 대상으로 한 Machine Learning 기반의 'Passenger finder' 라는 이름의 스마트폰 어플리케이션 프로토타입을 제작함. 2018년 12월부터 ㈜나비콜 및 개인 택시 기사를 대상으로 기술 자문 및 베타 테스트를 실시함

(바) 종합분석 및 결론

○ 공간 빅데이터 분석 및 가시화 기술 개발을 통해 소셜 데이터, 이동 객체 데이터, 실시간 데이터 등의 정형/비정형 공간 데이터를 사용자의 최소한의 지식으로 효과적으로 분석할 수 있는 기술을 개발하였음. 이를 기반으로 수십 건의 특허 및 논문 출판의 학술적 성과를 달성하였으며, 기업 대상으로 기술 자문 및 테스트를, 공공 정책 결정 및 수행 지원을 위해 공간 빅데이터 체계 구축 사업을 대상으로 자문 및 테스트를 진행하였음

<표 2-116> 연구개발 성과 종합

<table>
<tr><th colspan="2">과제명</th><th>공공</th><th>기술</th><th>학술</th><th>경제</th><th>기타</th></tr>
<tr><td rowspan="3">2세부</td><td rowspan="3">공간 빅데이터 분석 및 가시화 기술 개발</td><td>비정형 데이터 및 실시간 분석을 통해 정책 결정 및 수행을 지원</td><td>장소 선호도 기술, 딥러닝 기반 분석/가시화 기술, Batch 분석 솔루션, 대화형 분석 모듈 개발</td><td>특허 출원 14건, 등록 6건, SCI급 논문 19건, 일반 논문 42건, 국내외 학술발표 81건 달성</td><td>사용자 개입, 전문 지식 최소화로 분석 비용 절감 및 어노테이션 관련 고용 창출</td><td>공간 빅데이터 체계 구축 사업 및 기업 대상 기술 자문 및 베타 테스트</td></tr>
<tr><td colspan="5">종합</td></tr>
<tr><td colspan="5">공간 빅데이터 분석 및 가시화 기술 개발을 통해 소셜 데이터, 이동 객체 데이터, 실시간 데이터 등의 정형/비정형 공간 데이터를 사용자의 최소한의 지식으로 효과적으로 분석할 수 있는 기술을 개발하였음. 이를 기반으로 수십 건의 특허 및 논문 출판의 학술적 성과를 달성하였으며, 기업 대상으로 기술 자문 및 테스트를, 공공 정책 결정 및 수행 지원을 위해 공간 빅데이터 체계 구축 사업을 대상으로 자문 및 테스트를 진행하였음.</td></tr>
</table>

(4) 공간 빅데이터 서비스 활성화 및 실증 연구

(가) 공공적 성과

○ 공간빅데이터 활성화를 위해 국내외 관련 법제도, 정책 등에 대한 검토를 통한 정책 과제제안으로 효과적인 정책수립 기반 마련에 기여

- 국내외 빅데이터 서비스 제공내용 분석과 공간 빅데이터 서비스 차별화를 위한 공공/민간 부문의 활용방안의 제시를 통해 공간 빅데이터 활용도 향상을 위한 기반 마련
- 공간 빅데이터 추진체계의 구축, 관련 법제도 개선, 관련 기술개발, 지원사업의 추진, 공공DB 융복합 공유체계 마련 등 효과적인 공간 빅데이터 관련 정책 및 사업추진과 서비스 활성화를 위한 정책제안
- 공간 빅데이터 서비스 활성화를 위한 관련 법제 정비, 협력체계 구축, 개인정보 활용 및 보호 기준마련, 표준화 등의 정체과제 제안을 통해 체계적인 국가전략 및 계획수립 기반 마련

○ 다양한 공간 빅데이터 사업적용 기반마련과 공공시스템 공간 빅데이터 활용을 통한 확작성 증대

- 공간 빅데이터 기반 기술을 연계하여 실증하기 위한 통합 환경을 구축하고 이를 활용하여 추후 다양한 공간 빅데이터 사업에 적용 가능
- 다양한 공공 공간 빅데이터 서비스 연계를 수행하여 공공시스템 공간 빅데이터의 활용도를 높이고 서비스 모델 구현을 통한 공간 빅데이터 확장성 증대

(나) 기술적 성과

○ 세부별 공간 빅데이터 기술의 통합환경 구축

- 세부별 연구결과 기술을 통합하여 사용할 수 있는 환경을 구축하고 이를 활용해 공간빅데이터 분석 시스템 구축 시 기반 기술로 활용 가능

○ 공간빅데이터 서비스 시스템에 대해 GS인증을 취득하여 시스템 품질인증 및 시스템 안정성을 확보함으로 기반기술에 대한 품질인증 수행

○ 세부별로 구현된 공간연산 및 공간 분석 기능을 연계하여 분석가능한 공간 분석 프레임을 구축하여 이를 활용해 다양한 분석을 수행하며 수행된 분석결과를 가시화하여 사용자가 확인할 수 있는 실증시각화 서비스 개발

○ 공간 빅데이터 시스템의 자원을 전체적으로 관리할 수 있는 관리서비스 개발

○ 공공시스템 기반 공공 데이터 표준화 작업을 통해 이를 활용하여 공간연산, 공간 분석 기능과 연계하여 사용자가 원하는 서비스를 생성하고 이용할 수 있도록 시스템 확장 구축

○ 서비스 시나리오에 따른 데이터 분석 결과 가시화 사용자들에게 자료를 맞춤형 자료를 제공함으로써 사용자들이 공간 빅데이터를 위하여 쉽게 활용할 수 있도록 함

- 누구나 활용할 수 있도록 분석 자료를 OpenAPI로 제공 하며, 사용자 맞춤형 UI/UX로 가시화함

(다) 학술적 성과

○ 공간 빅데이터 활성화를 위한 공간빅데이터의 개념 및 활용방안, 다양한 서비스 제공방안 등에 대한 주제로 SCI급 논문 4건을 게재하여 학술적 기여

- 한국의 공간빅데이터 개념 및 활용방안을 주제로 개념 및 범위와 각 부문별 활용방안 등에 대한 SCI급 논문게재를 통해 명확한 공간빅데이터의 개념 및 활성화를 위한 학술적 기여
- 다양한 부문에서의 공간 빅데이터를 활용하여 연구개발중인 공간 빅데이터 저장관리 분석기술을 적용을 주제로 한 연구논문을 게재하여 공간 빅데이터의 기술적 활용성에 대한 학술적 기여

○ 공간 빅데이터 활용 및 실증과 관련한 비SCI급 논문 9건을 게재하여 공간정보 분야에서의 공간빅데이터 활용성과 홍보 및 학술적 가치 향상

- 공간 빅데이터 활용 및 실증과 관련한 서비스 프레임워크 개발, 실제 데이터를 활용한 서비스모델 개발, 실증에 필요한 요구기술 등에 대한 논문 게재를 통해 공간정보 분야에서의 공간빅데이터 활용성과 홍보 및 학술적 가치 향상에 기여

○ 공간 빅데이터 서비스 활성화 및 실증과 관련하여 국내외 학술발표 26.5건을 통해 연구개발내용에 대한 홍보 및 관련학계의 이해 증진

- 국내외 학술대회에서의 학술발표를 통해 공간 빅데이터의 이용 및 서비스, 통합실증과 관련한 연구개발 내용을 소개함으로써 관련학계에 대한 공간 빅데이터의 이해와 관심 증진

(라) 경제적 성과

○ 공간 빅데이터 실증 서비스 시스템 구축을 통해 공간 빅데이터 관련 사업의 진입 장벽을 낮춰 공간정보 시장의 활성화 기대

○ 외산 소프트웨어 중심의 기존 공간정보 시장을 재편하고 해외 시장에서의 경쟁력을 확보하여 국내 공간정보 관련 기업들의 해외진출에 기여

○ 공공 분야의 공간 빅데이터 분석을 통해 데이터를 표준화하고 표준화된 데이터를 공간 연산과 연계하여 새로운 서비스 구축

- 공간연산 기능 연계를 통해 다양한 서비스로 확장이 가능하고 이에 따라 새로운

비지니스 창출 기대

- 새로운 서비스를 구축을 통해 공간 빅데이터 서비스 확장을 도모하고 이를 통한 관련 산업의 활용도 및 확장성 증대

○ 공공 분야의 공간 빅데이터 분석을 통해 생성된 표준화된 데이터를 공간연산 기능 연계를 통해 다양한 서비스로 확장하고 이를 활용한 공간 정보 가치 확장 및 시장성 확보

○ 공공 분야의 공간 빅데이터 분석을 통해 데이터를 표준화하고 표준화된 데이터를 공간 연산과 연계하여 새로운 서비스를 구축할 수 있게끔 하여 서비스 확장을 도모하고 이를 통한 관련 산업의 활용도 및 확장성 증대

○ 공간 빅데이터 분석을 통해 도출된 경과를 다양한 계층의 사용자에게 제공함으로써 공간정보 산업 분야에서 새로운 서비스 제공이 가능해짐으로써, 자료의 활용도가 높아지고 공간정보 산업의 확장성이 용이해짐

(마) 기타 성과

○ 공간 빅데이터 통합 플랫폼과 관련한 활용 가이드라인 1건과 매뉴얼 1건 개발을 통해 향후 사용자들에 대한 활용성 증대

- 공간 빅데이터 통합 플랫폼에서 제공하는 프레임워크에 대한 설명 및 활용을 위한 가이드라인 개발

- 1~3세부 연구개발 기술을 연계한 통합플랫폼 활용을 위해 각 기능별 설명 및 예시 등을 포함한 매뉴얼 개발

○ 공간 빅데이터 관련 전문인력 양성 2인을 지원하여 관련업계에서 요구하는 전문인력 지원

- 공간 빅데이터 관련 석사과정 인력양성 지원을 통해 업계에서 요구하는 맞춤형 인력지원

○ 공간 빅데이터 연구개발 성과와 관련하여 언론, 홈페이지, 웹진발간 등을 통해 지속적인 홍보활동 추진

- 아시아투데이, YTN, 국토매일, 매일건설신문 등 다양한 멀티미디어를 통해 연구개발 내용에 대한 홍보성과 달성





<그림 2-438> 3세부 언론, 방송 등 홍보성과

- 국토교통기술대전, 스마트 국토엑스포, 스마트 건설·기술 안전 엑스포 등의 전시 활동 참여를 통해 연구개발성과 홍보





<그림 2-439> 3세부 전시활동 참가 홍보성과

- 연구단 홈페이지 개설 및 웹진발간 등을 통해 연구개발 성과 및 내용에 대한 지속적인 홍보활동
- 홈페이지는 공간빅데이터 소개, 연구단 소개, 연구분야, 정보마당, 홍보마당으로 구분되며, 공간빅데이터의 개념 및 역할, 기대효과, 연구단 조직도 및 연구진 소개, 연구분야별 목표와 내용, 연구 성과 및 웹진 등 다양한 세부 항목으로 구성됨



<그림 2-440> 3세부 연구단 홈페이지 구성

- 웹진은 연구단 홈페이지에서 다운로드 가능하며, 차년도별 1건 이상을 목표로 6차년도까지 14호를 발간하였으며, 국토공간정보의 빅데이터 관리, 분석 및 서비스 플랫폼 기술개발 연구 수행을 위한 기술동향, 연구과제별 주요 이슈와 행사 등을 기록하여 반영함



<그림 2-441> 3세부 웹진 발간 현황

<표 2-117> 3세부 기타성과

번호	연차	홍보내용	기관명	홍보일
1	1차	'국토공간정보연구사업' 연구수행기관 선정 (경기일보)	안양대학교	2014-08-13
2	1차	(아시아투데이) 국토공간정보의 빅데이터 서비스 플랫폼 개발 무엇이 달라질까	안양대학교	2014-10-07
3	1차	Geo-Spatial BIG DATA Vol.1	안양대학교	2014-11
4	2차	공간빅데이터 연구단 홈페이지 개설	안양대학교	2015-02
5	2차	Geo-Spatial BIG DATA Vol.2	안양대학교	2015-05
6	2차	Geo-Spatial BIG DATA Vol.3	안양대학교	2015-08
7	2차	Geo-Spatial BIG DATA Vol.4	안양대학교	2015-10
8	2차	Geo-Spatial BIG DATA Vol.5	안양대학교	2016-01
9	3차	2016 국토교통기술대전	안양대학교	2016-05-25~27
10	3차	Geo-Spatial BIG DATA Vol.6	안양대학교	2016-05
11	3차	도시, 사람을 품다 - 공간 빅데이터, 도시를 바꾸다	안양대학교	2016-07-08
12	3차	Geo-Spatial BIG DATA Vol.7	안양대학교	2016-08
13	3차	Geo-Spatial BIG DATA Vol.8	안양대학교	2016-11
14	4차	Geo-Spatial BIG DATA Vol.9	안양대학교	2017-02
15	4차	Geo-Spatial BIG DATA Vol.10	안양대학교	2017-06
16	4차	신동빈 안양대 도시정보공학과 교수 (전 한국공간정보학회 회장) '초연결 플랫폼'의 확산.. 공간정보 빅데이터 기술이 선도	안양대학교	2017-10-18
17	4차	Geo-Spatial BIG DATA Vol.11	안양대학교	2017-12
18	5차	Geo-Spatial BIG DATA Vol.12	안양대학교	2018-04
19	5차	2018년 스마트국토엑스포	안양대학교	2018-09-12~14
20	5차	[특별기고] 공간 빅데이터 플랫폼 활용성에 주목한다	안양대학교	2018-10-24
21	5차	Geo-Spatial BIG DATA Vol.13	안양대학교	2018-12
22	6차	Geo-Spatial BIG DATA Vol.14	안양대학교	2019-04
23	6차	2018 스마트 건설기술·안전 엑스포	안양대학교	2019.11.27~29

(바) 종합분석 및 결론

○ 공간 빅데이터 서비스 활성화 및 실증과 관련한 공공적 기반과 각 세부별 기술을 연계·통합하여 서비스 제공을 위한 기술적 기반을 마련하였음

- 공간 빅데이터 활성화를 위해 요구되는 정책·제도적 해결방안의 제시를 통한 정책수립의 기반을 제공
- 다양한 분야의 공간 빅데이터 분석을 위한 통합기술을 개발하고, 이를 통한 공간정보가치 확장 및 시장성 확보

○ 다양한 미디어를 통한 연구개발성과에 대한 지속적 홍보활동을 통해 공공 및 민간에서의 관심증대

- 신문, 방송, 온라인 웹, 박람회 등의 다양한 형태의 미디어를 통해 공간 빅데이터 연구개발성과에 대한 지속적인 홍보활동을 추진하였으며, 이에 따른 공공 및 민간에서의 관심증대와 공간 빅데이터 기술의 활용기반을 마련하였음

<표 2-118> 연구개발 성과 종합

<table>
<tr><th colspan="2">과제명</th><th>공공</th><th>기술</th><th>학술</th><th>경제</th><th>기타</th></tr>
<tr><td rowspan="3">3세부</td><td rowspan="3">공간 빅데이터 서비스 활성화 및 실증 연구</td><td>공간 빅데이터 활성화를 위해 요구되는 각종 정책과제를 발굴하여 제안함으로써 효과적인 정책수립 기반마련, 공간 빅데이터 사업 적용 및 활용성 증대</td><td>세부별 공간 빅데이터 기술을 통합하여 실증 가능한 통합 환경을 구축하여 공간 빅데이터를 활용한 공간정보 서비스 구축시 기반기술로 활용 가능. 공간 빅데이터의 제공을 통한 산업계 활성화 가능</td><td>공간빅데이터의 개념 및 활용범위 등을 차별화된 방안제시를 통한 학술적 기여와 공간빅데이터 분석결과 제시를 통해 다양한 부문에서의 활용성 증대</td><td>공간 빅데이터 실증 서비스 시스템 구축을 통해 관련산업 활성화와 표준화된 공간 빅데이터 분석을 통해 공간정보 가치 향상 및 시장성 확보</td><td>다양한 멀티미디어를 통한 연구개발 성과에 대한 지속적인 홍보를 통한 대국민 및 공공기관의 관심증대</td></tr>
<tr><td colspan="5">종합</td></tr>
<tr><td colspan="5">공간빅데이터 활성화를 위해 요구되는 정책 및 제도적 개선방안 제시를 통해 공공적 기반을 마련하고, 통합기술 개발을 통해 향후 다양한 공공부문에서의 공간 빅데이터 활용을 위한 기반을 마련하였음</td></tr>
</table>

3. 목표 달성도 및 관련 분야 기여도

가. 연차별 연구목표 달성도

(1) 공간빅데이터 저장 관리 인프라 기술 개발

○ 최종 연구목표는 공간 빅데이터의 수집, 저장, 관리 및 질의 기능 제공을 위한 핵심기술과 이를 구현한 고성능, 고확장성의 공간 빅데이터 저장관리 시스템을 개발하는 것이며 연차별 연구목표 대비 달성도는 다음 <표 3-1>과 같음

<표 3-1> 연차별 연구목표 달성도

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	달성내용	목표 달성 여부
1차 년도	공간 빅데이터 저장관리 인프라 기술 개발 (한국전자통신연구원)	Hadoop 기반 공간 빅데이터 저장 프레임워크 개발	• 공간 빅데이터 저장 시스템 기술 분석, 요구사항 분석	• 공간 빅데이터 요구사항 및 기능 정의	달성(100%)
			• 공간 빅데이터 저장 시스템 기능 분석, 개념 설계	• 공간 빅데이터 저장 관리자 구조 설계 • 공간 빅데이터 기반 분산 공간색인 관리자 및 공간연산자 설계	
			• 공간 빅데이터 저장 시스템 인터페이스 설계	• 공간 빅데이터 처리 언어 시스템 설계	
			• 공간 빅데이터 저장 시스템 개념 수준 구현	• 공간 빅데이터 저장 시스템 구축	
		공간 빅데이터 질의 처리 엔진 개발	• 공간 빅데이터 질의 처리 엔진 기술분석, 요구사항 정의	• 공간 빅데이터 질의처리 엔진 요구사항 정의	
			• 공간 빅데이터 질의 처리 엔진 기능분석, 설계	• 공간 빅데이터 질의 처리 엔진 구조 설계	
			• 공간 빅데이터 질의 언어 정의	• 공간 빅데이터 기본 연산자 및 질의 언어 정의	
			• 공간 빅데이터 ODBC/JDBC 인터페이스 설계	• 공간 빅데이터 DBMS ODBC / JDBC / Map/Reduce 인터페이스 설계 • 공간 빅데이터 DBMS 인터페이스 기본 기능 개발	

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	달성내용	목표 달성 여부
	인메모리 기반 공간 빅데이터 저장관리 프레임워크 개발 (포스웨이브)	인메모리 기반 공간빅데이터 저장관리 프레임워크 개발	• 실시간 공간 빅데이터 저장 플랫폼 분석 및 설계	• 실시간 공간 빅데이터 처리를 위한 인메모리 데이터 처리 시스템 구조 및 설계	달성(100%)
	공간 빅데이터 웨어하우스 핵심기술 개발 (아이씨티웨이)	스트림 기반 분산 공간 ETL 도구 개발	• 워크플로우기반 공간 ETL 도구 분석/설계	• EATL 기술 및 개발환경 분석 • 워크플로우 기반 공간 EATL 도구 설계 • 공간 EATL 시스템 설계	달성(100%)
2차 년도	공간 빅데이터 저장관리 인프라 기술 개발 (한국전자통신연구원)	Hadoop 기반 공간 빅데이터 저장 프레임워크 개발	• 공간 빅데이터 처리 연산자 개발	• 기본 공간 연산자(basic spatial operators) 개발 • 비공간 연산자(non-spatial operators) 개발	달성(100%)
		공간 빅데이터 질의 처리 엔진 개발	• 공간 빅데이터 질의 처리 엔진 핵심기술 개발	• 공간 빅데이터 질의 플랜 생성 모듈 개발 • 스트림 공간데이터 질의 언어 파싱 및 의미검증 모듈 개발 • 스트림 공간데이터 질의처리 모듈 (기본기능) 개발	
	인메모리 기반 공간 빅데이터 저장관리 프레임워크 개발 (포스웨이브)	인메모리 기반 공간빅데이터 저장관리 프레임워크 개발	• 실시간 공간 빅데이터 저장 연구시제품 개발	• 실시간 데이터 처리를 위한 인메모리 데이터 처리 모듈 개발	달성(100%)
	공간 빅데이터 웨어하우스 핵심기술 개발 (아이씨티웨이)	스트림 기반 분산 공간 ETL 도구 개발	• 공간 빅데이터 수집 모듈 개발	• 다중 공간 빅데이터 수집 모듈 개발 • 다중 공간 빅데이터 적재 모듈 개발	달성(100%)
			• 공간 스트림 빅데이터 수집 모듈 개발	• 공간 스트림 빅데이터 수집 모듈 개발 (단일 노드) • 공간 스트림 빅데이터 적재 모듈 개발 (단일 노드)	
3차 년도	공간 빅데이터 저장관리 인프라 기술 개발 (한국전자통신연구원)	Hadoop 기반 공간 빅데이터 저장 프레임워크 개발	• 공간 빅데이터 색인 시스템 개발	• 공간 빅데이터 색인 개발(1차)	달성(100%)
			• 공간 빅데이터 처리 고급 연산자 개발	• 고급 공간 연산자(advanced spatial operators) 개발(2차) • 공간 조인 연산자(spatial join operators) 개발	
		공간 빅데이터 질의 처리 엔진 개발	• 공간 빅데이터 질의처리 엔진 연구시제품 개발	• 공간 빅데이터 입력 스트림 관리 모듈 구현 • 스트림 공간 빅데이터 분산 질의처리 모듈 개발	
	인메모리	인메모리	• 인메모리 기반	• 인메모리 기반 공간색인 모듈	달성(100%)

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	달성내용	목표 달성 여부
	기반 공간 빅데이터 저장관리 프레임워크 개발 (포스웨이브)	기반 공간빅데이터 저장관리 프레임워크 개발	공간색인 및 기본공간 연산 모듈 개발	개발 • 인메모리 데이터 처리를 위한 기본 공간 연산자 개발	
	공간 빅데이터 웨어하우스 핵심기술 개발 (아이씨티웨이)	스트림 기반 분산 공간 ETL 도구 개발	• 공간 스트림 빅데이터 처리 기술 개발	• 공간 스트림 빅데이터의 병렬/분산 처리 모듈 개발	달성(100%)
4차 년도	공간 빅데이터 저장관리 인프라 기술 개발 (한국전자통신연구원)	Hadoop 기반 공간 빅데이터 저장 프레임워크 개발	• 공간 빅데이터 색인 시스템 개발	• 공간 빅데이터 색인 개발(2차) • MapReduce를 활용한 고속 공간 빅데이터 색인 구축 기술 개발	달성(100%)
		공간 빅데이터 질의 처리 엔진 개발	• 공간 빅데이터 DBMS 인터페이스 연구시제품 개발	• 공간 빅데이터 DBMS JDBC 클라이언트/서버 모듈 개발 • 공간 빅데이터 DBMS ODBC 클라이언트/서버 모듈 개발	
	인메모리 기반 공간 빅데이터 저장관리 프레임워크 개발 (포스웨이브)	인메모리 기반 공간빅데이터 저장관리 프레임워크 개발	• 인메모리 기반 이동객체 빅데이터 저장 플랫폼 분석 설계	• 인메모리 기반 이동객체 활용 사례 및 자료 분석을 통한 저장 플랫폼 설계	달성(100%)
			• 인메모리 기반 이동객체 빅데이터 저장 모듈 POC 수행	• 인메모리 기반 이동객체 빅데이터 저장 모듈 POC 개발을 통한 설계안 검증	
	공간 빅데이터 웨어하우스 핵심기술 개발 (아이씨티웨이)	스트림 기반 분산 공간 ETL 도구 개발	• 스트림 기반 공간 ETL 처리 엔진 개발	• 스트림 기반 공간 ETL 처리 엔진 개발 • 스트림 기반 공간 ETL 처리 엔진 연동 라이브러리(1차 POC 수준) 개발	달성(100%)
5차 년도	공간 빅데이터 저장관리 인프라 기술 개발 (한국전자통신연구원)	Hadoop 기반 공간 빅데이터 저장 프레임워크 개발	• 공간 빅데이터 처리 언어 개발	• 공간 빅데이터 처리 언어 문법 설계	달성(100%)
			• 공간 빅데이터 저장 시스템 고도화	• 색인 기반 고속 공간 조인 연산자 개발 • 고속 데이터 입력을 위한 지연 색인 갱신 기술 개발	
			• 원격 공간 빅데이터 활용 기술	• 클라이언트-서버 기반 원격 공간 빅데이터 접근 모듈 개발	
			• 공간 빅데이터 저장 관리 인프라 시범 서비스 적용 지원(1)	• 공간 빅데이터 저장 관리 프레임워크 통합	
			• 공간 빅데이터	• GUI 기반 공간 빅데이터 응용	

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	달성내용	목표 달성 여부
			분석 응용 개발 도구(1)	개발 도구 개발	
		공간 빅데이터 질의 처리 엔진 개발	• 공간 빅데이터 DBMS 인터페이스 프레임워크 개발	• 스트림 공간 빅데이터 인터페이스 클라이언트/서버 모듈 개발 • 스트림 공간 빅데이터 인터페이스 통합처리 서버 프레임워크 개발	
	인메모리 기반 공간 빅데이터 저장관리 프레임워크 개발 (포스웨이브)	인메모리 기반 공간빅데이터 저장관리 프레임워크 개발	• 인메모리 기반 이동객체 빅데이터 저장 플랫폼 개발	• 인메모리 기반 이동객체 저장 및 관리 모듈 개발 • 인메모리 기반 이동객체 빅데이터 시공간 색인 개발	달성(100%)
			• 실시간 공간 빅데이터 저장 플랫폼 고도화 (1차)	• 인메모리 기반 공간 분석 라이브러리 개발 (13종) • 인메모리 기반 공간 색인 및 저장 모듈 통합 및 고도화	
			• 실시간 공간 빅데이터 저장 플랫폼 고도화 (2차 일부)	• 공간질의 통합 인터페이스 설계 및 구현	
	공간 빅데이터 웨어하우스 핵심기술 개발 (아이씨티웨이)	스트림 기반 분산 공간 ETL 도구 개발	• 공간 스트림 빅데이터 처리 기술 개발	• 공간 센서 스트림 데이터 수집 및 분산 처리 모듈 개발	달성(100%)
			• 스트림 기반 공간 ETL 처리 엔진 고도화	• 스트림 기반 공간 ETL 규칙 처리 엔진 개발(POC 수준) • 스트림 기반 공간 ETL 규칙 처리 엔진 개발(배포용) • 스트림 기반 공간 ETL 처리 엔진 연동 API 개발	
6차 년도	공간 빅데이터 저장관리 인프라 기술 개발 (한국전자통신연구원)	Hadoop 기반 공간 빅데이터 저장 프레임워크 개발	• 공간 빅데이터 처리 언어 개발(2)	• Groovy기반 내부 코드 실행기 개발	달성(100%)
			• 공간 빅데이터 저장관리 시스템 고도화	• MapReduce 기반 분석 플랜 플로우 처리기 개발 • Web기반 공간 분석 워크플로우 저작 및 가시화 도구 개발 • 대용량 공간 빅데이터 원격 접근 기능 개발 • 대용량 공간 빅데이터 클라이언트 캐싱 모듈 개발 • 2세부 공간 연산자 연동 모듈 개발	
			• 공간 빅데이터 저장 관리 인프라 시범 서비스 실증 지원(2)	• 공간 빅데이터 사업단 저장 시스템 연동 모듈 개발 • 시범 서비스 공간 빅데이터 연산자 및 어댑터 개발 • 시범 서비스 실증 통합 테스트	
			• 공간 빅데이터 분석 응용 개발 도구(2)	• GUI 기반 공간 빅데이터 응용 개발 도구 고도화	

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	달성내용	목표 달성 여부
		공간 빅데이터 질의 처리 엔진 개발	• 공간 빅데이터 질의 처리 엔진 고도화	• 공간 빅데이터 질의처리 기능개선 및 최적화 모듈 개발 • 공간 빅데이터 질의처리 엔진 사업단 시스템 테스트 및 적용	
	인메모리 기반 공간 빅데이터 저장관리 프레임워크 개발 (포스웨이브)	인메모리 기반 공간빅데이터 저장관리 프레임워크 개발	• 실시간 공간 빅데이터 저장 플랫폼 V2 (2)	• SparkSQL 기반 공간 질의 처리 모듈 개발 • SparkSQL 기반 기본 공간연산자 모듈 개발 • 인메모리 기반 이동객체 및 공간자료 저장 및 질의 인터페이스 통합	달성(100%)
			• 실시간 인메모리 기반 저장 플랫폼 실증	• 자료 저장 관리 인터페이스 연계 모듈 개발	
	공간 빅데이터 웨어하우스 핵심기술 개발 (아이씨티웨이)	스트림 기반 분산 공간 ETL 도구 개발	• 스트림 기반 공간 ETL 처리 엔진 고도화	• 스트림 기반 공간 ETL 처리 엔진 고도화 • 스트림 기반 공간 ETL 처리 엔진 연구시제품	달성(100%)
			• 스트림 기반 공간 ETL 처리 엔진 관리 도구 개발	• 공간 스트림 ETL 관리 도구 개발 • 공간 연산 워크플로우 분석도구 개발	

(2) 공간 빅데이터 분석 및 가시화 기술 개발

○ 최종 연구목표는 공간 빅데이터 분석 프레임워크, Batch 분석, 이동객체 Batch 분석, Interactive 분석 기술 등 공간 빅데이터 분석 핵심기술과 컴포넌트 개발 하는 것이며 연차별 연구목표 대비 달성도는 다음 <표 3-2>과 같음

<표 3-2> 연차별 연구목표 달성도

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	달성내용	목표 달성 여부
1차년도	공간 빅데이터 소셜 분석 기술 개발 (서울대)	공간 빅데이터 분석을 위한 모델 설계 및 목표 성능 설정	• 공간 빅데이터 분석 모델 현황 분석 및 요구사항 정의	• 소셜 미디어 및 공간 빅데이터 분석을 위한 GISMO 모델링 분석, 소셜 미디어 및 공간 빅데이터를 위한 선험적 모델링 분석함으로써 목표한 바를 충실히 달성	달성(100%)
			• 공간 빅데이터 분석 목표 모델 정의	• 소셜 미디어 및 공간 빅데이터 특성과 정적/동적 분석 특성에 따른 목표모델 정의 및 성능 지표 설정, 소셜 미디어 및 공간 빅데이터 기반 분석 모델 설계, 데이터의 동적/정적 특성에 따른 분석 모델 설계함으로써 목표한 바를 충실히 달성	달성(100%)
	공간 빅데이터 분석 및 가시화 컴포넌트 구현 (시터스)	공간 빅데이터 분석을 위한 모델 설계 및 목표 성능 설정	• 데이터 특성별 분석 모델 설계	• O2 플랫폼, SAP HANA, ESRI GIS tools for Hadoop, Splunk 솔류션 분석 • 분석 프레임워크 개발 시 평가가 가능하도록 평가지표 설정 • 기존 솔루션 분석 후 평가지표를 달성하기 위한 분석 프레임워크의 개념적 설계 완료	달성(100%)
	공간빅테이터 Batch 분석 컴포넌트(연세대/키위 컴퍼니)	공간 빅데이터 일괄 분석 컴포넌트 기술 개발	• 공간 빅데이터 통계기반 분석 요구조건 분석 (R 의 공간분석기능 검토) • 공간 빅데이터 비통계기반 분석 요구조건 분석 (Machine Learning 기법 등) • 공간 빅데이터 분석 컴포넌트 아키텍처 설계 • 공간 빅데이터 분석	• 국토부에서 서비스 중인 국가공간계획지원체계(KOPSS) 분석을 통해 공공 및 민간 수요처의 요구조건을 분석 • 공간 빅데이터 통계/비통계 분석 기능 검토를 위해 R에서 제공하는 패키지와 literature review 조사를 병행 • 동적 계획법 (dynamic programing) 을 활용하여 간단한 분석 사례를 반복적으로 구현하여 분석 컴포넌트의 역할 및 인터페이스 특징을 도출	달성(100%)

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	달성내용	목표 달성 여부
			컴포넌트 인터페이스 설계		
	이동객체데이터 Batch 분석 컴포넌트(연세대/키위컴퍼니)	이동객체데이터 일괄 분석 컴포넌트 기술 개발	• 이동객체데이터 통계기반 분석 요구조건 분석(R의 공간분석기능 검토) • 이동객체데이터 비통계기반 분석 요구조건 분석(Machine Learning 기법 등) • 이동객체데이터 분석 컴포넌트 아키텍처 설계 • 이동객체데이터 분석 컴포넌트 인터페이스 설계	• 국토부에서 서비스 중인 국가공간계획지원체계(KOPSS) 분석을 통해 공공 및 민간 수요처의 요구조건을 분석 • 이동객체 데이터 통계/비통계 분석 기능 검토를 위해 R에서 제공하는 패키지와 literature review 조사를 병행 • 동적 계획법 (dynamic programing) 을 활용하여 간단한 분석 사례를 반복적으로 구현하여 분석 컴포넌트의 역할 및 인터페이스 특징을 도출	달성(100%)
	공간 빅데이터 Interactive 분석 컴포넌트(한국과학기술원/유누스)	공간 빅데이터 Interactive 분석 컴포넌트 요구사항 정의서	• 공간 빅데이터 Interactive 분석 컴포넌트 요구사항 정의서 작성	• 오픈소스 CEP 엔진 코드 분석 자료 작성 • 공간 CEP 시스템 아키텍처 설계도 및 UML 다이어그램 작성	달성(100%)
		공간 빅데이터 Interactive 분석 컴포넌트 설계서	• 공간 빅데이터 Interactive 분석 컴포넌트 설계서 작성	• 오픈소스 Spatial OLAP 엔진 코드 분석 자료 작성 • 공간 OLAP 시스템 아키텍처 설계도 및 UML 다이어그램 작성	달성(100%)
2차년도	공간 빅데이터 소셜 분석 기술 개발 (서울대)	공간 빅데이터 분석을 위한 데이터 가공, 변환, 필터링 및 분석효율 고도화	• 이기종 공간DB 특성 분석	• 소셜 미디어의 공간 빅데이터 수집을 위한 대상 정의, 버티컬 SNS로 사용자가 많은 인스타그램과 정보 신로도가 높은 유튜브, 정보 신뢰도가 높은 마이크로 블로깅 트위터, 검색결과를 제공하는 구글 트렌드, 수집기술은 OpenAPI, 수집범위는 대한민국 전역으로 분석함으로써 목표한 바를 충실히 달성	달성(100%)
			• 이기종 DB 간 데이터 가공, 변환, 필터링 기술 개발	• 소셜 미디어 공간 빅데이터 수집 모듈을 이용하여 데이터 수집 및 중복 필드 제거와 DB 저장 관리, 브이월드(Vworld) 지오코더 API를 이용한 좌표 및 주소의 부재 정보 취득, 검색범위, 검색기간,	달성(100%)

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	달성내용	목표 달성 여부
				데이터 출처, 데이터 유형 등을 이용한 필터링 모듈 제공함으로써 목표한 바를 충실히 달성	
			• 다중 레이어를 위한 캐싱/프리패칭/버퍼링 기술 개발	• 동일한 조건을 빈번하게 필터링할 경우 수행속도를 줄이기 위해 제공, DBMS의 메모리 테이블을 이용하여 별도의 데이터 저장, 표출할 정보의 영역을 계산하여 적절한 지도 레벨 표시, 표출을 위한 관리 페이지를 기준으로 전후의 예비 페이지 관리함으로써 목표한 바를 충실히 달성	달성(100%)
	공간 빅데이터 분석 및 가시화 컴포넌트 구현 (시터스)	공간 빅데이터 분석을 위한 데이터 가공, 변환, 필터링 및 분석 효율 고도화	• 이기종 DB 간 데이터 가공, 변환, 필터링 모듈 개발	• 소셜 미디어 공간 빅데이터 수집 모듈을 이용하여 데이터 수집 및 중복 필드 제거와 DB 저장 관리 개발 • 브이월드(V-World) 지오코더 API를 이용한 좌표 및 주소의 부재 정보 취득 기능 개발 • 검색범위, 검색시간, 데이터 출처, 데이터 유형 등을 이용한 필터링 모듈 개발	달성(100%)
			• 다중 레이어를 위한 캐싱/프리패칭/버퍼링 모듈 개발	• 동일한 조건을 빈번하게 필터링할 경우 수행속도를 줄이기 위해 모듈 개발 • DBMS의 메모리 테이블을 이용하여 별도의 데이터 저장 • 표출할 정보의 영역을 계산하여 적절한 지도 레벨 표시 • 표출을 위한 관리 페이지를 기준으로 전후의 예비 페이지 관리	달성(100%)
	공간빅데이터 Batch 분석 컴포넌트(연세대/키위 컴퍼니)	공간 빅데이터 일괄 분석 컴포넌트 기술 개발	• 공간 빅데이터 기본 분석 기능 선정 및 개발 기능 결정 여부 • 통계기반 분석 컴포넌트 기초 설계 여부 • 비통계기반 분석 컴포넌트 기초 설계 여부 • 공간 빅데이터 분석 컴포넌트 프로토타이핑	• 1차년도 연구수행 결과인 요구조건 분석을 활용하여, 문헌조사를 통한 공간 빅데이터에 필요한 기본 분석기능 선정 • 선정된 기본 분석기능 중, 새로운 형태의 데이터, 분석기법, 응용분야에 적용된 기능을 개발 기능으로 선정 • 각 컴포넌트 별 기초 설계를 수행하여 컴포넌트의 조합을 통한 해당 개발 기능 프로토타이핑 수행	달성(100%) • 2개 신규 기능 개발
	이동객체데이터 Batch 분석 컴포넌트(연세대/키위	이동객체데이터 일괄 분석 컴포넌트 기술 개발	• 이동객체데이터 기본 분석 기능 선정 및 개발 기능 결	• 1차년도 연구수행 결과인 요구조건 분석을 활용하여, 문헌조사를 통한 이동객체데이터에 필요한 기본 분석기능 선정	달성(100%) • 2개 신규 기능 개발

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	달성내용	목표 달성 여부
	컴퍼니)		정 여부 • 통 계 기 반 분석 컴포넌트 기초 설계 여부 • 비통계기반 분석 컴포넌트 기초 설계 여부 • 이동객체데이터 분석 컴 포 넌 트 프로토타이핑	• 선정된 기본 분석기능 중, 새로운 형태의 데이터, 분석기법, 응용분야에 적용된 기능을 개발 기능으로 선정 • 각 컴포넌트 별 기초 설계를 수행하여 컴포넌트의 조합을 통한 해당 개발 기능 프로토타이핑 수행	
	공간 빅데이터 Interactive 분석 컴포넌트(한국과학기술원/유누스)	CQL 질의 메모리 상주 색인 및 저장 모듈	• CQL 질의 메모리 상주 색인 및 저장 모듈 개발	• 이벤트 스트림의 이벤트 릴레이션 변환기 설계 • 이벤트 릴레이션의 이벤트 스트림 변환기 설계 • CQL 질의 저장 기술 개발 • CQL 질의 색인 기술 개발 • 고속 입력 이벤트를 위한 메모리 상주 처리 기술 연구	달성(100%)
		공간 빅데이터 기반 복합 이벤트 스케줄러 모듈	• 공간 빅데이터 기반 복합 이벤트 스케줄러 모듈 개발	• 고속 입력 이벤트의 샘플링 기능 개발 • 대기 이벤트를 위한 큐잉 모듈 개발	달성(100%)
3차년도	공간 빅데이터 소셜 분석 기술 개발 (서울대)	공간 빅데이터 정합 및 주제도 분석 기술	• 공간 빅데이터 정합 및 주제도 분석 기술 개발	• 소셜 미디어 데이터 및 비정형 데이터에서 행정구역(시군구) 및 POI까지 포함하는 장소명에 해당하는 장소정보를 추출하고, 이를 구글에서 제공하는 OpenAPI를 이용하여 위치정보를 취득하는 기술을 개발하였으며, 공간 감성어 사전을 구축하여 이를 이용한 장소별 선호도 기반 주제도 분석 기술, 특정 키워드에 대한 사용자 영향력을 고려한 key player 탐지를 위한 소셜 네트워크 분석 기반 주제도 분석 기술을 개발함으로써 목표한 바를 충실히 달성	달성(100%)
	공간 빅데이터 분석 및 가시화 컴포넌트 구현 (시터스)	공간 빅데이터 정합 및 주제도 분석 기술	• 공간 빅데이터 정합 및 주제도 분석 모듈 개발	• 소셜 미디어 데이터 및 비정형 데이터에서 행정구역(시군구) 및 POI까지 포함하는 장소명에 해당하는 장소정보를 추출하고, 이를 구글에서 제공하는 OpenAPI를 이용하여 위치정보를 취득하는 모듈 개발 • 공간 감성어 사전을 구축하여 이를 이용한 장소별 선호도 기반 주제도 분석 모듈, 특정 키워드에 대한 사용자 영향력을 고려한 key player 탐지를 위한 소셜 네트워크 분석 기반 주제도 분석 모듈,	달성(100%)

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	달성내용	목표 달성 여부
				비정형 데이터에서 공간분석 기반 층간소음용 주제어 분석 모듈 개발	
	공간빅테이터 Batch 분석 컴포넌트(연세대/키위컴퍼니)	공간 빅데이터 일괄 분석 컴포넌트 기술 개발	• 공간 빅데이터 기본 분석 기능 개선 및 향상 • 통계기반 분석 기능 추가 • 비통계기반 분석 기능 추가 • 공간 빅데이터 분석 컴포넌트 기능 추가 및 고도화	• 3차년도 개발 예정인 2개의 공간 빅데이터 분석 모듈의 개발이 완료되었고, 효과적인 공간 빅데이터 처리를 위해 MBR기법을 사용하여 성능 향상을 진행함에 있어 3차년도 연구 성과 목표를 달성	달성(100%) • 2개 신규 기능 개발
	이동객체데이터 Batch 분석 컴포넌트(연세대/키위컴퍼니)	이동객체데이터 일괄 분석 컴포넌트 기술 개발	• 이동객체데이터 기본 분석 기능 개선 및 향상 • 통계기반 분석 기능 추가 • 비통계기반 분석 기능 추가 • 이동객체데이터 분석 컴포넌트 기능 추가 및 고도화	• 3차년도 개발 예정인 2개의 이동객체 데이터 분석 모듈의 개발을 완료하고, 이를 이용하여 실제 사례에 대한 분석이 가능함에 있어 3차년도 연구 성과목표를 달성	달성(100%) • 2개 신규 기능 개발
	공간 빅데이터 Interactive 분석 컴포넌트(한국과학기술원/유누스)	공간 빅데이터 기반 질의 처리기 모듈	• 공간 빅데이터 기반 질의 처리기 모듈 개발	• 3차년도에 목표하였던 공간 빅데이터 기반 복합 이벤트 질의 처리기의 개발 및 고도화를 충실히 수행하였다. Storm의 최적 Topology를 자동으로 결정하기 위한 방법론을, 병렬처리 패러다임인 Data parallelism과 Task parallelism을 사용하여서 고안 • 값비싼 연산인 공간 조인 연산자들의 성능을 개선하기 위해서 공간 인덱싱 기법을 공간 CEP 엔진에 추가 • 공간 CEP 엔진의 사용성을 개선하여서, 2차년도에 API 수준이었던 것을 3차년도에는 SQL 수준까지 개선	달성(100%)
		CQL 질의 클러스터 내 최적 할당 모듈	• CQL 질의 클러스터 내 최적 할당 모듈 개발	• 3차년도 주요 업무인 'CQL질의 클러스트 내 최적 할당 모듈 개발'을 개발하였으며, 목표로 한 100개 이상의 데이터를 1초에 1.2회 이상 입력받아 100% 데이터 처리를 하	달성(100%)

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	달성내용	목표 달성 여부
				고, 임계값 초과 시 재기동 후 임계값 이하로 복구되도록 모듈을 구현하는 등 목표한 바를 충실히 달성	
4차년도	공간 빅데이터 소셜 분석 기술 개발 (서울대)	공간 빅데이터 분포 및 차트 분석 기술	• 공간 빅데이터 분포 분석 기술 개발	• 소셜 미디어 데이터에서 이모티콘을 고려하여 장소 선호도 분석 기술, 위치정보를 추정하는 기술, 딥러닝을 적용하여 장소명 대체어 분석 기술 개발 및 공공시설의 선호요인 분석 수행함으로써 목표한 바를 충실히 달성	달성(100%)
			• 공간 빅데이터 차트 분석 기술 개발	• 소셜 미디어 데이터에서 딥러닝을 적용하여 장소에 특화된 영어 감성어 사전 구축 기술, 특정 주제에 대한 문서분류 기술, 재난재해 관련 문장추출 기술, 토픽별 KeyPlayer 탐지 기술 개발함으로써 목표한 바를 충실히 달성	달성(100%)
	공간 빅데이터 분석 및 가시화 컴포넌트 구현 (시터스)	공간 빅데이터 분포 및 차트 분석 기술	• 공간 빅데이터 분포 분석 모듈 개발	• 소셜 미디어 데이터에서 이모티콘을 고려한 장소 선호도 분석 모듈, 위치정보를 추정하는 모듈, 딥러닝을 적용하여 장소명 대체어를 분석하는 모듈 및 공공시설의 선호요인 분석 모듈 개발	달성(100%)
			• 공간 빅데이터 차트 분석 모듈 개발	• 소셜 미디어 데이터에서 딥러닝을 적용하여 장소에 특화된 영어 감성어 사전 구축 모듈, 특정 주제에 대한 문서분류 모듈, 재난재해 관련 문장추출 모듈, 토픽별 Keyplayer 탐지 모듈 개발	달성(100%)
	공간빅테이터 Batch 분석 컴포넌트(연세대/가이온)	공간 빅데이터 일괄 분석 컴포넌트 기술 개발	• 공간 빅데이터 기본 분석 기능 개선 및 향상 • 통계기반 분석 기능 추가 • 비통계기반 분석 기능 추가 • 공간 빅데이터 분석 컴포넌트 기능 추가 및 고도화	• 4차년도 개발 예정인 2개의 공간 빅데이터 분석 모듈의 개발이 완료되었고, 효과적인 공간 빅데이터 처리를 위해 MBR기법을 사용하여 성능 향상을 진행함에 있어 3차년도 연구 성과 목표를 달성	달성(100%) • 5개 신규 기능 개발
	이동객체데이터 Batch 분석 컴포넌트(연세대/가이온)	이동객체데이터 일괄 분석 컴포넌트 기술 개발	• 이동객체데이터 기본 분석 기능 개선 및 향상 • 통계기반 분석 기능 추가 • 비통계기반	• 4차년도 개발 예정인 2개의 이동객체 데이터 분석 모듈의 개발을 완료하고, 이를 이용하여 실제 사례에 대한 분석이 가능함에 있어 3차년도 연구 성과목표를 달성	달성(100%) • 5개 신규 기능 개발

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	달성내용	목표 달성 여부
			분석 기능 추가 • 이동객체데이터 분석 컴포넌트 기능 추가 및 고도화		
	공간 빅데이터 Interactive 분석 컴포넌트(한국과학기술원/유누스/유엔진솔루션즈)	공간 빅데이터 OLAP 데이터 로딩 모듈	• 공간 빅데이터 OLAP 데이터 로딩 모듈 개발	• 4차년도에 목표하였던 공간 빅데이터 OLAP 엔진 개발 및 대화형 질의기능 추가에 대한 연구개발을 충실히 수행 • 1세부 개발물인 Marmot과 공간 OLAP인 GeoMondrian을 연동하여 공간 빅데이터 OLAP의 Proof-Of-Concept 버전을 완성	달성(100%)
		공간 빅데이터 OLAP 드릴다운/롤업 모듈	• 공간 빅데이터 OLAP 드릴다운/롤업 모듈 개발	• 공간과 시간에 대한 대화형 드릴다운/롤업 기능을 추가	달성(100%)
5차년도	공간 빅데이터 소셜 분석 기술 개발 (서울대)	공간 빅데이터 분포 및 차트 분석 기술	• 공간 빅데이터 분포 분석 고도화 기술 개발	• 소셜 미디어 데이터 및 비정형 데이터(다음 카페, 뉴스, 트립어드바이저 장소리뷰, 민원 데이터 등)에서 개체명인식을 활용한 비정형 텍스트 데이터 위치탐지 기술, 공공시설 매력도 측정 및 접근성 분석 기술, 사용자 중점 기반 시간지도 생성 기술을 개발함으로써 목표한 바를 충실히 달성	달성(100%)
			• 공간 빅데이터 차트 분석 고도화 기술 개발	• 소셜 미디어 데이터 및 비정형 데이터(다음 카페, 뉴스, Yelp 리뷰 등)에서 영문 장소리뷰 데이터로부터 장소 테마 분석 기술, 머신러닝 기반 문장 분류 기술 개발함으로써 목표한 바를 충실히 달성	달성(100%)
		공간 빅데이터 분석결과 가시화 기술	• 공간 빅데이터 분석결과 가시화 기술(정적 가시화) 개발	• 소셜 미디어 데이터 및 국문 장소리뷰 데이터와 공간밀도 지수를 이용한 유동인구 가시화 기술 개발을 개발함으로써 목표한 바를 충실히 달성	달성(100%)
	공간 빅데이터 분석 및 가시화 컴포넌트 구현 (시터스)	공간 빅데이터 분포 및 차트 분석 기술	• 공간 빅데이터 분포 분석 고도화 모듈 개발	• 소셜 미디어 데이터 및 비정형 데이터(다음 카페, 뉴스, 트립어드바이저 장소리뷰, 민원 데이터 등)에서 개체명인식을 활용한 비정형 텍스트 데이터 위치탐지 모듈, 공공시설 매력도 측정 및 접근성 분석 모듈, 사용자 중점 기반 시간지도 생성 모듈 개발	달성(100%)
			• 공간 빅데이터 차트 분석 고도화 모듈 개	• 소셜 미디어 데이터 및 비정형 데이터(다음 카페, 뉴스, Yelp 리뷰 등)에서 영문 장소리뷰 데이터로부터 장소테마	달성(100%)

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	달성내용	목표 달성 여부
			발	분석 모듈, 머신러닝 기반 문장 분류 모듈 개발	
		공간 빅데이터 분석 결과 가시화 기술	• 공간 빅데이터 분석 결과 가시화 모듈 개발 (정적 가시화)	• 소셜 미디어 데이터 및 국문 장소리뷰 데이터와 공간밀도 지수를 이용한 유동인구 가시화 모듈 개발	달성(100%)
	공간빅테이터 Batch 분석 컴포넌트(연세대/가이온)	공간 빅데이터 일괄 분석 컴포넌트 기술 개발	• 통계기반 분석 성능 최적화 • 비통계기반 분석 성능 최적화 • 공간 빅데이터 분석 컴포넌트 기능 추가 및 고도화 • 공간 빅데이터 분석 Batch 솔루션 구현	• 5차년도 개발 예정인 4개의 공간 빅데이터 분석 모듈의 개발이 완료되었고, 효과적인 공간 빅데이터 처리를 위해 Map-Reduce법을 사용하여 성능 향상을 진행함에 있어 5차년도 연구 성과 목표를 달성	달성(100%) • 4개 신규 기능 개발
	이동객체데이터 Batch 분석 컴포넌트(연세대/가이온)	이동객체데이터 일괄 분석 컴포넌트 기술 개발	• 통계기반 분석 성능 최적화 • 비통계기반 분석 성능 최적화 • 이동객체데이터 분석 컴포넌트 기능 추가 및 고도화 • 이동객체빅데이터 분석 Batch 솔루션 구현	• 5차년도 개발 예정인 4개의 이동객체 데이터 분석 모듈의 개발을 완료하고, 이를 이용하여 실제 사례에 대한 분석이 가능함에 있어 5차년도 연구 성과목표를 달성	달성(100%) • 4개 신규 기능 개발
	공간 빅데이터 Interactive 분석 컴포넌트(한국과학기술원/유엔진솔루션즈)	공간 빅데이터 OLAP 대화형 분석기 모듈	• 공간 빅데이터 OLAP 대화형 분석기 모듈 개발	• 공간 빅데이터 OLAP 과 공간 데이터베이스간의 밀결합을 구현하여 빅데이터 처리에 보다 효율적으로 수행	달성(100%)
		공간 빅데이터 기반 복합 이벤트 네트워크 생성 모듈	• 공간 빅데이터 기반 복합 이벤트 네트워크 생성 모듈 개발	• 공간 빅데이터 복합 이벤트 네트워크를 자동생성 모듈 및 이벤트 및 룰 최적 분배 모듈을 개발하여 전문 지식이 없는 사용자가 최소한의 정보로도 공간 빅데이터 복합 이벤트 네트워크를 생성	달성(100%)
		공간 빅데이터 기반 복합 이벤트 처리기 UI/UX	• 공간 빅데이터 기반 복합 이벤트 처리기 UI/UX 개발	• 공간 빅데이터 OLAP에서 2개 이상의 테이블에 총 1GB 이상의 데이터를 로딩, 두 가지 이상의 시각화 요소(예: 지도, 테이블)가 개발되었는지 확인	달성(100%)
		공간 빅데이터 OLAP UI/UX	• 공간 빅데이터 OLAP UI/UX 개발	• 지도상에서의 드릴다운/롤업이 올바르게 동작하도록 하는 등 목표한 바를 충실히 달성	달성(100%)

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	달성내용	목표 달성 여부
6차년도	공간 빅데이터 소셜 분석 기술 개발 (서울대)	공간 빅데이터 분석결과 가시화 기술	• 공간 빅데이터 분석결과 가시화 기술 개발	• 소셜 미디어 데이터 및 비정형데이터(트위터데이터, Yelp 리뷰), 체크인데이터, 국문 장소리뷰 데이터에서 정적 가시화(토픽기반 장소임베딩 기술, 사용자 맞춤 추천기술)와 동적 가시화(단어임베딩을 활용한 사용자관심 이벤트탐지기술, 통행목적을 고려한 공공자전거 수요예측 기술) 기술 개발을 개발함으로써 목표한 바를 충실히 달성	달성(100%)
	공간 빅데이터 분석 및 가시화 컴포넌트 구현 (시터스)	공간 빅데이터 분석결과 가시화 기술	• 공간 빅데이터 분석결과 가시화 모듈 개발	• 소셜 미디어 데이터를 이용하여 이상시점을 탐지 및 사용자 관심 이벤트 텍스트와 장소 정보를 추출하는 비정형 데이터 기반 이벤트 탐지 모듈 개발 • 공공자전거의 통행 목적을 고려하여 이에 대한 수요를 예측하는 분석 모듈 개발	달성(100%)
		공간 빅데이터 사업단 실증	• 공간 빅데이터 분석 모듈 사업단 적용 및 실증	• 사업단 적용성 검토를 통한 실증 대상 기술 정의 • 사업단 시스템에 실증 대상 기술을 적용하기 위한 주요 기술 검토 및 테스트	달성(100%)
	공간빅테이터 Batch 분석 컴포넌트(연세대/가이온)	공간 빅데이터 일괄 분석 컴포넌트 기술 개발	• 공공/민간 수요기반 Batch 분석 시나리오 설계 • 공간 빅데이터 Batch 솔루션 설계 • 공간 빅데이터 분석 Batch 솔루션 구현 여부 • 공간 빅데이터 분석 솔루션 사업단 실증 (2건)	• 6차년도 개발 예정인 5개의 공간 빅데이터 분석 모듈의 개발이 완료되었고, 특히 공공 / 민간 기반 시나리오 설계 시 새로운 Graph 형태의 OD 자료를 사용	달성(100%) • 5개 신규 기능 개발
	이동객체데이터 Batch 분석 컴포넌트(연세대/가이온)	이동객체데이터 일괄 분석 컴포넌트 기술 개발	• 공공/민간 수요기반 Batch 분석 시나리오 설계 • 이동객체데이터 Batch 솔루션 설계 • 이동객체데이터 분석 Batch 솔루	• 6차년도 개발 예정인 4개의 이동객체 데이터 분석 모듈의 개발을 완료하고, 이를 이용하여 실제 사례에 대한 분석이 가능함에 있어 6차년도 연구 성과목표를 달성	달성(100%) • 4개 신규 기능 개발

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	달성내용	목표 달성 여부
			션 구현 여부 • 이동객체데이터 분석 솔루션 사업단 실증(2건)		
	공간 빅데이터 Interactive 분석 컴포넌트(한국과학기술원/유엔진 솔루션즈)	공간 빅데이터 Interactive 분석 컴포넌트 테스팅 및 튜닝	• 공간 빅데이터 Interactive 분석 컴포넌트 테스팅 및 튜닝	• 실제 교통, 지적, 부동산 등의 데이터를 적용하여 테스트 • 기본 7종의 공간 연산자(intersect, contain, disjoint 등) 테스트 • 여러 데이터의 input을 조합한 다양한 rule 분산 및 처리 테스트 • 데이터베이스를 통한 실시간 데이터 입력 및 출력 여부 테스트 • 실제 교통, 지적, 부동산 등의 데이터를 적용하여 테스트 • 다양한 차원에 대한 Roll up & Drill down 기능 테스트 • 다양한 공간 데이터에 대해 지도상 시각화 기능 테스트 • 공간 데이터 저장소와의 밀결합 기능 테스트	달성(100%)
		공간 빅데이터 Interactive 분석 컴포넌트 시범서비스 적용 지원	• 공간 빅데이터 Interactive 분석 컴포넌트 시범서비스 적용 지원	• 공간 빅데이터 CEP/OLAP 엔진 UI/UX 사업단 시범 서비스 실증 완료 • 1, 2, 3세부 연동 서비스 개발 및 지원 • 공간 빅데이터 CEP/OLAP 엔진 패키징 및 문서화	달성(100%)

(3) 공간 빅데이터 서비스 활성화 및 실증 연구

○ 최종 연구목표는 공간 빅데이터 플랫폼 활용을 지원하는 서비스의 활성화 및 실증을 위하여 공간 빅데이터 활용 서비스 프레임워크, 공간 빅데이터 서비스 제공을 위한 UI/UX, OpenAPI, 공간 빅데이터 분석 결과 연계 및 가시화 모듈(SW) 등을 개발 및 실증하는 것이며 연차별 연구목표 대비 달성도는 다음 <표 3-3>과 같음

<표 3-3> 연차별 연구목표 달성도

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	달성내용	목표 달성 여부
1차 년도	공간 빅데이터 서비스 활성화 방안 연구 (안양대학교 산학협력단)	공간 빅데이터 융합 서비스의 서비스 프레임워크 및 구성요소 구현방안 제시	• V-World 융합 기반 공간 빅데이터 플랫폼의 서비스 프레임워크 도출	• 빅데이터 및 공간 빅데이터에 대한 선행연구 검토를 통한 공간 빅데이터의 개념 정립 • 공간정보 오픈플랫폼, 공간 빅데이터 체계구축사업, 기타 서비스프레임워크 선행연구 등 현황분석 • 플랫폼간 서비스 연계를 위해 수집·저장, 처리·분석, 표현·서비스 측면에서의 시사점 도출	달성(100%)
			• 서비스 프레임워크 구성요소 개발방안 제시	• 공간 빅데이터 서비스 프레임워크 구상(서비스 Management, Container, Monitoring, 플랫폼) 및 구성요소별 개발방안 제시	달성(100%)
		핵심 서비스 모델 도출을 위한 서비스 현황 분석 및 시사점 도출	• 공공/민간 부문, 서비스 유형별 공간 빅데이터 서비스 현황 수집 및 분석	• 빅데이터 및 공간 빅데이터 서비스 사례 40종에 대해 공간 빅데이터 특성에 따른 지원기술 적용 현황 분석, 인문·사회·, 산업·경제, 환경, 기반시설, 기타 및 공공/민간 부문별 기술적용 사례분석	달성(100%)
			• 국내외 공간 빅데이터 활용사례 분석을 통한 기존 서비스와 공간 빅데이터 서비스와의 차별화 방안 도출	• 공간 빅데이터 서비스의 역할 및 범위, 공공/민간 등 주체별 활용 전략, 서비스 차별화 방안 제시	달성(100%)
	공간 빅데이터 서비스 실용화 연구 (공간인소프트)	Open API 기술 현황 분석 UI/UX 기술 현황 분석	•	• Open API 기술의 국내외 적용사례조사, 공간정보 오픈플랫폼 Open API현황 분석, 서비스별 UI/UX 기술 및 적용사례 분석, 공간 빅데이터 서비스 제공을 위한 UI/UX 목표정의	달성(100%)
	공간 빅데이터 서비스 실증 연구(포도)	공간 빅데이터 현황 및 서비스 구축 사례 분석	• 현황 및 구축사례 분석	• 공간 빅데이터 사례현황 및 구축사례 분석	달성(100%)

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	달성내용	목표 달성 여부
		공간정보 오픈플랫폼 실증 기술 현황 분석	• 실증 기술 현황 분석	• 공간정보 오픈플랫폼 실증 기술 현황 분석	달성(100%)
		국내 공공 공간 빅데이터 구축 및 시각화 기술 현황 분석	• 시각화 기술 현황 분석	• 국내 공공 공간 빅데이터 구축 및 시각화 기술 현황 분석	달성(100%)
2차 년도	공간 빅데이터 서비스 활성화 방안 연구 (안양대학교 산학협력단)	공간 빅데이터 융합 서비스의 프레임워크 매뉴얼 작성	• 공간 빅데이터 서비스 활성화를 위한 • 정책적, 제도적 • 지원방안 제시 (1)	• 공간 빅데이터 서비스 활성화를 위한 정책적, 제도적 대응과제 도출	달성(100%)
		공간 빅데이터 서비스 유형화 및 핵심 서비스 모델 개발	• 핵심 서비스 유형 및 • 서비스 모델 개발(1)	• 공간 빅데이터 핵심 서비스 분야도출 • 도출된 핵심 서비스 분야에 2세부에서 개발중인 분석기법과 국내 공공기관 플랫폼의 연계성을 고려한 핵심 서비스모델 도출	달성(100%)
	공간 빅데이터 서비스 실용화 연구 (공간인소프트)	공간 빅데이터 Open API 설계 공간 빅데이터 서비스 제공을 위한 UI/UX 개발전략 수립	•	• 공공시스템 기반 공간 빅데이터 Open API 서비스 콘텐츠 구성 및 인터페이스설계 • 공간 빅데이터 서비스 UI/UX 니즈 조사	달성(100%)
	공간 빅데이터 서비스 실증 연구(포도)	빅데이터 서비스 아키텍처 및 프레임워크 설계	• 서비스 아키텍처 및 프레임워크 설계결과	• 서비스 아키텍처 및 프레임워크 설계	달성(100%)
		공간 빅데이터 서비스 시스템 실증 기술 설계	• 실증 기술 설계 결과	• 공간 빅데이터 서비스 시스템 실증 기술 설계	달성(100%)
		저장, 관리, 분석, 가시화 등 개별 서비스 설계	• 개별 서비스 설계결과	• 저장, 관리, 분석, 가시화 등 개별 서비스 설계	달성(100%)
		공공 공간 빅데이터 서비스 연계 및 확장 기술 설계(1)	• 공공 공간 빅데이터 서비스 연계 및 확장 기술 설계결과	• 공공 공간 빅데이터 서비스 연계 및 확장 기술 설계	달성(100%)
		공공 공간 빅데이터 서비스 연계 및 확장 기술 프로토타입 개발	• 공공 공간 빅데이터 서비스 연계 및 확장 기술 프로토타입 개발	• 공공 공간 빅데이터 서비스 연계 및 확장 기술 프로토타입 개발	달성(100%)
3차 년도	공간 빅데이터 서비스 활성화 방안 연구 (안양대학교 산학협력단)	공간 빅데이터 융합 서비스의 프레임워크 매뉴얼 작성	• 공간 빅데이터 서비스 활성화를 위한 • 정책적, 제도적 • 지원방안 제시	• 공간 빅데이터 서비스 활성화를 위한 정책적, 제도적 대응과제 도출	달성(100%)

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	달성내용	목표 달성 여부
			(2)		
			• 서비스 프레임워크 • 활용 가이드라인 개발	• 서비스 프레임워크 연계를 위한 활용 가이드라인 개발 • 공간 빅데이터 플랫폼 활용을 위한 1-2-3세부간의 연계 요소들에 대한 가이드라인 개발	달성(100%)
		공간 빅데이터 서비스 유형화 및 핵심 서비스 모델 개발	• 핵심 서비스 유형 및 • 서비스 모델 개발(2)	• 핵심 서비스 모델 개발 • 2차년도에 이어 융합, 기술, 기타(정치) 부문의 핵심 서비스 모델을 추가 개발	달성(100%)
	공간 빅데이터 서비스 실용화 연구 (공간인소프트)	공간 빅데이터 Open API 기능 개발 공간 빅데이터 서비스 제공을 위한 UI/UX 설계	•	• 핵심서비스 모델을 바탕으로 Open API 기능 개발 • 개발된 Open API 서비스 웹페이지를 S/W 등록 • 핵심서비스 모델에 대한 논리모델과 물리모델 정의 • 스토리보드 작성을 통해 UI/UX 설계 완료	달성(100%)
	공간 빅데이터 서비스 실증 연구(포도)	공간 빅데이터 서비스 통합 기술 설계	• 통합 기술 설계 결과	• 통합 기술 설계	달성(100%)
		공간 빅데이터 서비스 통합 기술 프로토타입 개발	• 공간 빅데이터 서비스 통합 기술 프로토타입 개발	• 공간 빅데이터 서비스 통합 기술 프로토타입 개발	달성(100%)
		공공 공간 빅데이터 서비스 연계 및 확장 기술 설계(2)	• 공공 공간 빅데이터 서비스 연계 및 확장 기술 설계 결과	• 공공 공간 빅데이터 서비스 연계 및 확장 기술 설계	달성(100%)
4차년도	공간 빅데이터 서비스 활성화 방안 연구 (안양대학교산학협력단)	핵심 서비스의 비즈니스 모델 개발	• 공공/민간부문 공간 빅데이터 활용을 위한 비즈니스 모델 분석 및 개발	• 공공/민간부문 공간 빅데이터 활용을 위한 비즈니스 모델 개발 • 미래사회 이슈 및 트렌드를 분석하여 공간 빅데이터의 경향을 파악하고, 정부정책 및 현재 이뤄지고 있는 주요 기술을 검토 • 연구단의 개발 분석기술 및 공공시스템 데이터의 활용가능 여부를 검토 • 국내외 공간 빅데이터 서비스 환경분석을 통한 서비스 전략을 수립 • 기존 핵심 서비스 모델과 1, 2세부 개발기술 분석을 통해 공공/민간 부문 공간 비데이터를 활용한 비즈니스 모델 개발	달성(100%)
			• 공공/민간부문 공간 빅데이터 활용을 위한 서비스 시나리오 개발(1)	• 공공/민간부문 공간 빅데이터 활용을 위한 서비스 시나리오 분석 및 개발 • 공공/민간부문 공간 빅데이터 활용을 위한 서비스 시나리오 개발 • 비즈니스 모델별로 공간 빅	달성(100%)

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	달성내용	목표 달성 여부
				데이터 플랫폼에서 서비스가 가능하도록 서비스 시나리오 개발 절차 및 시스템 아키텍처 적용방안 제시 • 4차년도에 개발된 비즈니스 모델을 기반으로 원시데이터의 유형 및 범용성, 연구단의 개발 기술 등을 고려하여 3개 유형의 서비스 시나리오를 개발 • 각 서비스 시나리오별로 원시데이터부터 저장·관리, 분석, 개별응용 서비스까지 서비스를 위한 분석단계별로 연구단의 개발모듈을 적용하여 최종서비스를 도출	
	공간 빅데이터 서비스 실용화 연구 (공간인소프트)	공간 빅데이터 Open API 기능 개발	•	• 핵심서비스 모델을 바탕으로 Open API 기능 개발 완료, 개발된 Open API 서비스 기능을 S/W 등록	달성(100%)
	공간 빅데이터 서비스 실증 연구(포도)	공간 빅데이터 서비스 시스템 실증 기술 개발	• 실증 기술 개발 결과	• 공간 빅데이터 서비스 시스템 실증 기술 개발	달성(100%)
		저장, 관리, 분석, 가시화 등 개별 서비스 기술 개발	• 개별서비스 개발 결과	• 개별서비스 개발 결과 보고서 및 시연	달성(100%)
		공공 공간 빅데이터 서비스 연계 및 확장 기술 개발(1)	• 공공 공간 빅데이터 서비스 연계 및 확장 기술 개발 결과	• 공공 공간 빅데이터 서비스 연계 및 확장 기술 개발 결과 보고서 및 시연	달성(100%)
5차 년도	공간 빅데이터 서비스 활성화 방안 연구 (안양대학교산학협력단)	핵심 서비스의 비즈니스 모델 개발	• 공공/민간부문 공간 빅데이터 활용을 위한 서비스 시나리오(2)	• 공공/민간부문 공간 빅데이터 활용을 위한 서비스 시나리오 분석 및 개발 • 공공/민간부문 공간 빅데이터 활용을 위한 서비스 시나리오 개발 • 비즈니스 모델별로 공간 빅데이터 플랫폼에서 서비스가 가능하도록 서비스 시나리오 개발 절차 및 시스템 아키텍처 적용방안 제시 • 3개 유형의 서비스 시나리오를 개발 및 각 서비스 시나리오별로 원시데이터부터 저장·관리, 분석 • 개별응용 서비스까지 서비스를 위한 분석단계별로 연구단의 개발모듈을 적용하여 최종서비스를 도출	달성(100%)
		핵심서비스 제공방안 개발	• 서비스 구현을 위한 시스템 아키텍처 설계	• 서비스 구현을 위한 시스템 아키텍처 설계 • 공간 빅데이터 실증서비스 시스템의 관리·운영을 위한 서비스 대상을 구분하고 공간 빅데이터 서비스 제공을	달성(100%)

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	달성내용	목표 달성 여부
				위한 일련의 과정에 따라 Layer별 기능을 정의하고 Layer간 관계분석을 수행	
			• 핵심서비스의 주요기능 도출 (1)	• 핵심서비스 주요기능 도출 • 공간 빅데이터 실증서비스 시스템 아키텍처에 따라 Layer별로 개발된 기능들을 정리하여 List-up하고, 공공/민간부문 서비스 시나리오에 따른 주요기능을 도출	달성(100%)
	공간 빅데이터 서비스 실용화 연구 (공간인소프트)	공간 빅데이터 Open API 서비스 사이트 개발 공간 빅데이터 서비스 제공을 위한 UI/UX 시스템 개발	•	• 개발된 Open API 서비스 이용정보를 사용자에게 제공하기 위한 웹 사이트 개발 완료 • 서비스의 단위 기능에 대한 User Interface를 구성하고 이를 시스템으로 개발 • 개발된 Open API 서비스 웹 사이트를 S/W 등록 완료 • 개발된 사용자 인터페이스를 S/W 등록 완료	달성(100%)
	공간 빅데이터 서비스 실증 연구(포도)	저장, 관리, 분석, 가시화 등 개별 서비스 기술 개발	• 개별 서비스 기술 개발	• 공간빅데이터 저장, 관리, 분석, 가시화 기능을 위한 개별 서비스 기술 개발 • 실증시각화 서비스 개발 • 관리서비스 개발	달성(100%)
		공간 빅데이터 서비스 통합 기술 개발	• 통합 기술 개발	• 세부별 기능을 통합하여 실증할 수 있는 공간분석프레임 기반기술 개발 • 자체적으로 마련한 연계 통합 기준으로 모듈형태의 기능을 연계하는 통합 기술 개발	달성(100%)
		공공 공간 빅데이터 서비스 연계 및 확장 기술 개발(1)	• 공공 공간 빅데이터 서비스 연계 및 확장 기술 개발	• 공공 공간 빅데이터 서비스 연계 및 확장 기술 개발 보고서 • 공공부문 공간데이터를 활용하여 공간분석프레임을 수행할 수 있는 확장 기술 개발	달성(100%)
6차 년도	공간 빅데이터 서비스 활성화 방안 연구 (안양대학교산학협력단)	공간 빅데이터 활용가이드 개발	• 활용 가이드 개발	• 플랫폼의 개념, 구성 및 기능별 개발 모듈 등 사용자가 공간 빅데이터 플랫폼을 활용하기 위한 가이드 개발 • 사업종료 이후 연구성과물의 활성화 및 사업화 방안 제시	달성(100%)
		핵심 서비스모델 기반 활용 가이드 적용예시	• 서비스 모델의 가이드 적용예시	• 기 개발한 서비스 시나리오를 기반으로 데이터 Load, 분석, 가시화, 저장 등 데이터의 입력부터 결과 도출까지 단계별 활용 가이드 적용예시 개발	달성(100%)
		핵심서비스 주요 기능 도출	• 주요기능 도출 (2)	• 공간 빅데이터 플랫폼의 주요 서비스 항목을 기능별로 구분하여 정의 • 5차년도에 도출한 공간 비데이터 플랫폼의 주요기능 List를 기준으로 추가 개발된 기능을 검토하여 최종 List-up 작성	달성(100%)

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	달성내용	목표 달성 여부
	공간 빅데이터 서비스 실용화 연구 (공간인소프트)	공간 빅데이터 Open API 서비스를 적용한 UI/UX 테스트 사이트 개발 연구성과물 성능 검증 지원 및 개선방안 연구	• 테스트 사이트 개발 및 연구성과물 성능 검증	• (개발보고서) 1건 작성 완료 • (성능검증 및 개선사항 보고서) 2건 작성 완료 • (S/W등록) 2건 달성(1건 추가달성)	달성(100%)
	공간 빅데이터 서비스 실증 연구(포도)	공간 빅데이터 서비스 통합 기술 개발	• 통합 기술 개발	• 세부별 기술을 통합하여 실증 가능한 공간분석프레임 개발	달성(100%)
		공간 빅데이터 서비스 시스템 구축	• 시스템 구축(테스트베드)	• 시스템 구축 보고서 • 안양대학교 테스트베드를 기반으로 시스템 구축 • 총2차에 걸쳐 31명에게 시연하고 GS인증을 통한 성능검증 • 사용자매뉴얼에 개별분석모듈에 대한 설명 및 분석결과 자료 작성 • 운용매뉴얼 • (GS인증) 1건 완료 • (S/W등록) 1건 완료(추가달성) • 학술발표 1건 완료(추가달성)	달성(100%)
		공공 시스템 기반 공간 빅데이터 서비스 시스템 확장 구축	• 시스템 구축(테스트베드)	• 시스템 확장 구축 보고서 • 공공데이터를 활용한 서비스 빅데이터 5종 이상 개발	달성(100%)

나. 연구의 한계 및 시사점

(1) 공간빅데이터 저장 관리 인프라 기술 개발

○ **(1-1 세세부)** 공간 빅데이터 저장관리 인프라 기술 개발

- 대용량의 공간 빅데이터를 다룰 수 있는 공간 빅데이터 처리 기술을 개발하였으며 기존의 시스템에 비해 성능을 향상시키고 대상 시스템인 ‘공간 빅데이터 분석 플랫폼’에 적용하여 활용 가능성을 실증함
- 시스템 성능 향상에 영향을 준 기술적 요소는 공간 색인 기법과 스트림 기반의 데이터 처리 방식으로 판단되며, 이는 향후 시스템 성능 향상을 위한 기술적 포커스를 시사함
- 기술의 실증 대상 시스템인 ‘공간 빅데이터 분석 플랫폼’의 개발 사업자의 요구와 필요성 제기에 따라 추가 연산자 개발이 필요할 것으로 예상됨
  - 본 시스템은 연산자 기능을 확장하기 용이하게 설계되었으므로 개발된 추가 연산자를 시스템에 쉽게 추가할 수 있을 것임
- 또한, 개발 사업자의 운영 경험에 따라 시스템 운영 시 공간분석 기능 외에 부수적인 기능의 필요성이 제시되었음
  - 예를 들면 대용량 공간 데이터의 가시화를 위한 지원과 최근 많은 활용되고 있는 다양한 프로그래밍 언어(Python, R, Scala 등)에 대한 지원을 들 수 있음
  - 이러한 기술들은 공간 분석과 직접 연관된 기술은 아니지만, 플랫폼의 활용 확산에 중요한 부분을 차지할 것으로 예상되어 이러한 부가 기능에 대한 추가적인 연구 및 개발이 필요함

○ **(1-2 세세부)** 인메모리 기반 공간 빅데이터 저장관리 프레임워크 개발

- 기본적인 연산 모듈 제공과 실제 운영 데이터 기반 검증등을 통해 인메모리 기반 공간 빅데이터 분석 플랫폼의 실용화 가능성을 높임.
- 인메모리 기반 이동객체 저장 관리 기능에 대한 연구 목표치는 달성하였으나, 향후 실용화를 위해서는 이동객체 자료에 대한 시공간 기반 분석 등의 차별화된 분석 기능들에 대한 추가적인 연구 및 개발이 필요함
  - 이동객체의 궤적 기반 패턴분석이나 향후 위치 예측기법과 같은 다양한 이동객체 분석기법에 대한 발굴 및 개발이 필요
  - 플랫폼과 타 시스템과의 연계 등을 통한 활용측면에서 GPX, NMEA, TCX와 같은 지원되는 입출력 포맷 확장 필요

○ **(1-3 세세부)** 공간 빅데이터 웨어하우스 핵심기술 개발

- 점차 증가하고 있는 방대한 공간정보와 다양한 포맷의 데이터들을 수집하고 변환, 적재하려는 요구가 증가하고 있지만 전통적인 ETL 방식으로 데이터 관리로 해결하는 데에는 한계가 있음
- 비정형적으로 발생하는 공간정보(소셜데이터, 센서데이터, 로그정보 등)를 실시간으로 수집, 정제, 통합하여 활용가능한 공간정보로 변환, 적재되는 새로운 방식에 관리 플랫폼이 필요함
- 국토, 지역, 도시 부문 정책에 빅데이터를 활용하기 위해서는 빅데이터를 지속적으로 발굴, 수집, 정제하는 체계를 구축하는 것이 필요함

<표 3-4> 연구의 한계 및 시사점

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	미달성 내용	연구의 한계 및 시사점
1세부	공간 빅데이터 저장관리 인프라 기술 개발 (한국전자통신연구원)	Hadoop 기반 공간 빅데이터 저장 프레임 워크 개발	• 공간 빅데이터 저장 시스템 기술 분석, 요구사항 분석	• 해당사항 없음	• 효율적인 실용화와 사업화를 위해 수요처의 요구사항들이 반영되었음
			• 공간 빅데이터 저장 시스템 기능 분석, 개념 설계	• 해당사항 없음	• 다년도 과제 수행으로 인해 지속적으로 발생하는 요구사항 변화에 대응할 필요가 있음
			• 공간 빅데이터 저장 시스템 인터페이스 설계	• 해당사항 없음	
			• 공간 빅데이터 저장 시스템 개념 수준 구현	• 해당사항 없음	• 이후 실제 시스템 개발에 많은 feedback을 제공하여 유용하였음
			• 공간 빅데이터 처리 연산자 개발	• 해당사항 없음	• 개발 초기에 확정된 공간 연사자 외에도 수요처 필요에 따른 공간 빅데이터 연산자를 개발할 필요가 발생함 • 이후에도 활용 범위 확장에 따른 추가 연산자 개발이 필요할 것으로 예상됨
			• 공간 빅데이터 처리 고급 연산자 개발		
			• 공간 빅데이터 색인 시스템 개발	• 해당사항 없음	• 결과적으로 공간 색인은 개발된 시스템의 성능 향상에 가장 큰 영향을 주었음 • 향후 연구개발 동향 변화에 따라 추가적인 색인 기법을 연구하는 것이 시스템 성능 향상에 많은 영향을 줄 것으로 예상됨
			• 공간 빅데이터 처리 언어 개발	• 해당사항 없음	• 공간 빅데이터 처리 언어는 공간 분석 응용의 생산성에 많은 기여를 한 것으로 판단됨 • 현재 개발된 언어에 보급을 위한 다양한 응용 라이브러리의 추가 개발이 필요함
			• 공간 빅데이터 저장 시스템 고도화	• 해당사항 없음	• 저장 시스템 개발 이후에 수요 대상처의 다양한 요구사항을 추가로 받아 고도화를 수행함
			• 공간 빅데이터 분석 응용 개발 도구	• 해당사항 없음	• 현 구현은 Web 기반 방식으로 구현한 상태이나 데이터 적재와 같은 기능은 client-server 방식이 더 효과적일 것으로 판단됨
			• 원격 공간 빅데이터 활용 기술	• 해당사항 없음	• 과제 제안 당시에는 없던 연구 내용이었으나, 현재 수요예상처의 애로사항을 해결하기 위해 추가된 연구 내용임
			• 공간 빅데이터 저장 관리 인프라 시범 서비스 적용 지원	• 해당사항 없음 •	• 실증 대상 시스템인 '공간 빅데이터 분석 플랫폼'의 개발 사업자 선정 지연으로 연구 수행 일정이 다소 지연됨

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	미달성 내용	연구의 한계 및 시사점
		공간 빅데이터 질의 처리 엔진 개발	• 공간 빅데이터 질의 처리 엔진 기술분석, 요구사항 정의	• 해당사항 없음	• 효율적인 실용화와 사업화를 위해 수요처의 요구사항들이 반영되었음
			• 공간 빅데이터 질의 처리 엔진 기능분석, 설계	• 해당사항 없음	• 현 연구목표는 배치기반의 질의 처리이며 추후 많은 수요가 발생할 이벤트 기반 실시간 질의 처리 기능의 추가 연구가 필요함
			• 공간 빅데이터 질의 언어 정의	• 해당사항 없음	
			• 공간 빅데이터 ODBC/JDBC 인터페이스 설계	• 해당사항 없음	• 최근 많은 각광을 받는 Python/R 언어를 위한 질의 인터페이스 추가 지원이 필요할 것으로 보임
			• 공간 빅데이터 DBMS 인터페이스 연구시제품 개발	• 해당사항 없음	
			• 공간 빅데이터 DBMS 인터페이스 프레임워크 개발	• 해당사항 없음	
			• 공간 빅데이터 질의 처리 엔진 핵심기술 개발	• 해당사항 없음	• 스트림 기반으로 한 데이터 처리 방식의 효율성이 높았음 • 추후 저장된 데이터 뿐만 아니라 여러 공간센서에서 제공된 이벤트에 reactive 반응할 수 있는 반응형 질의 처리에 관한 연구가 필요할 것으로 예상됨
			• 공간 빅데이터 질의처리 엔진 연구시제품 개발	• 해당사항 없음	• 시제품을 통해 현재 국가가 포털을 통해 제공하는 여러 공간 데이터에 오류가 다수 있음을 확인할 수 있었음
			• 공간 빅데이터 질의 처리 엔진 고도화	• 해당사항 없음	• 추후 저장된 데이터 뿐만 아니라 여러 공간센서에서 제공된 이벤트에 reactive 반응할 수 있는 반응형 질의 처리에 관한 연구가 필요할 것으로 예상
	인메모리 기반 공간 빅데이터 저장관리 프레임워크 개발 (포스웨이브)	인메모리 기반 공간빅데이터 저장관리 프레임워크 개발	• 실시간 공간 빅데이터 저장 플랫폼 분석 및 설계	• 해당사항 없음	• 효율적인 실용화와 사업화를 위해 수요처의 요구사항들이 반영되었음
			• 실시간 공간 빅데이터 저장 연구시제품 개발	• 해당사항 없음	• 완성도 및 현실성 향상을 위해 다양한 피드백을 통한 수정 및 보완 진행이 필요함
			• 인메모리 기반 공간색인 및 기본공간 연산 모듈 개발	• 해당사항 없음	• 실용화 및 사업화를 위해 공간빅데이터 기반의 차별화된 연산 모듈 연구 및 보완이 필요함
			• 인메모리 기반 이동객체 빅데이터 저장 플랫폼 분석 설계	• 해당사항 없음	• 실제 이동객체 데이터 분석을 기반으로 설계안의 적절성 및 실용화 가능성을 높임
			• 인메모리 기반 이동객체 빅데이터 저장	• 해당사항 없음	• 실제 이동객체 데이터를 기반으로 설계안의 저장 기능 및 성능의 적절성을 효율적으로 검증함

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	미달성 내용	연구의 한계 및 시사점
			모듈 POC 수행		
			• 인메모리 기반 이동객체 빅데이터 저장 플랫폼 개발	• 해당사항 없음	• 교통안전공단에서 수집하는 DTG 자료를 이용하여 구현된 시스템의 검증을 통해 실용화 가능성을 높임
			• 실시간 공간 빅데이터 저장 플랫폼 고도화 (1차)	• 해당사항 없음	• 다양한 공간연산 기능 추가와 인터페이스 개선을 통해 활용성을 높임
			• 실시간 공간 빅데이터 저장 플랫폼 고도화 (2차 일부)	• 해당사항 없음	• 일반 공간자료와 이동객체 자료에 대한 통합 인터페이스 제공을 통해 시스템의 범용성 및 활용성을 높임
			• 실시간 공간 빅데이터 저장 플랫폼 V2 (2)	• 해당사항 없음	• 향후 인메모리 기반 기술의 진화 및 Trend 변화에 맞춰 지속적인 보완 및 고도화가 필요함
			• 실시간 인메모리 기반 저장 플랫폼 실증	• 해당사항 없음	• 실운영 시스템과의 연계하여 플랫폼 안정성을 검증을 통해 향후 실용화 가능성을 높임
	공간 빅데이터 웨어하우스 핵심기술 개발 (아이씨티웨이)	스트림 기반 분산 공간 ETL 도구 개발	• 공간 빅데이터 ODBC/JDBC 인터페이스 설계	• 해당사항 없음	• ODBC/JDBC 인터페이스를 통한 공간 빅데이터 질의 처리 기능 접근성을 높임
			• 공간 빅데이터 DBMS 인터페이스 기본기능 모듈 구현	• 해당사항 없음	• 향후 단순질의 외 복합질의 및 공간질의에 대한 고도화가 필요함
			• 다중 공간 데이터 수집 모듈 개발	• 해당사항 없음	• 추후 다양한 공간정보 처리를 위해 2종 이상의 공간정보를 처리 가능하나 더 많은 공간정보 처리가 가능하게 보완이 필요함
			• 공간 스트림 데이터 수집 모듈 개발 (단일 노드)	• 해당사항 없음	• 향후 단일노드에서 뿐만 아니라 분산 모드에서 공간 스트림 데이터 수집이 가능하게 지속적인 보완이 필요함
			• 공간 스트림 데이터의 병렬/분산 처리 모듈 개발	• 해당사항 없음	• 분산 환경에서 공간정보 ETL 처리 기능 및 테스트 베드 제공으로 실용화 가능성을 높임
			• 스트림 기반 공간 ETL 처리 엔진 개발	• 해당사항 없음	• 분산 환경에서 배치/스트림 공간 ETL 처리 지원 서비스 엔진을 제공함으로 활용성을 높임
			• 스트림 기반 공간 ETL 처리 엔진 연동 API (POC 수준)	• 해당사항 없음	• 스트림 기반 공간 ETL 처리 엔진을 효율적으로 활용하기 위한 자바 라이브러리 형태의 API 서비스를 지원함으로 실용화 가능성을 높임
			• 스트림 기반 공간 ETL 처리 엔진 연동 API 개발	• 해당사항 없음	• 공간 ETL 처리 엔진 연동 API를 통해 사용자 접근이 용이하나 추후 지속적인 API 지원 목록 확장 및 보완이 필요함
			• 스트림 기반 공간 ETL 규칙 처리	• 해당사항 없음	• 사용자 요구사항을 반영하기 위한 스크립트 기반 규칙 처리 서비스를

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	미달성 내용	연구의 한계 및 시사점
			엔진 개발(POC 수준)		지원함으로 활용성을 높임 • 독립적 구동보다는 최종 성과물과 Built-In되어 구동되도록 보완이 필요함
			• 공간 센서 스트림 데이터 수집 및 분산 처리 모듈 개발	• 해당사항 없음	• 배치기반 공간정보 뿐만 아니라 스트림 기반으로 생성되는 공간정보에 대해 ETL 서비스를 지원함으로 활용성을 높임
			• 스트림 기반 공간 ETL 규칙 처리 엔진 개발(2차배포판)	• 해당사항 없음	• 스트림 기반 공간 ETL 처리엔진 내 Built-In되어 자동으로 규칙 생성되어 공간 ETL 서비스를 지원함으로 활용성을 높임
			• 스트림 기반 공간 ETL 처리엔진 고도화 및 연구시제품 개발	• 해당사항 없음	• 다양한 인터페이스 제공으로 공간 ETL 처리 엔진의 활용성을 높임
			• 공간 스트림 ETL 관리 도구 개발	• 해당사항 없음	• 웹 인터페이스를 통한 관리 도구 제공으로 활용성 및 실용화 가능성을 높임
			• 공간 연산 워크플로우 분석도구 개발	• 해당사항 없음	• 서로 다른 플랫폼에 공간ETL 서비스를 지원하기 위한 공간ETL NiFi 라이브러리 지원으로 활용성을 높임

(2) 공간 빅데이터 분석 및 가시화 기술 개발

○ **(2-1세세부)** 공간 빅데이터 소셜 분석 및 가시화 컴포넌트 기술 개발

- 비정형 데이터의 유료화 및 서비스 제한으로 다양한 데이터를 수집하는데 어려움이 있었음
- 장소나 위치정보가 포함된 장소 리뷰에 대한 한글 코퍼스가 부족하여 비정형데이터 분석을 위한 한글 비정형데이터를 이용한 훈련 모형 개발에 한계가 있었음
- 민간 포털사에서 제공하는 OpenAPI를 사용하기 위한 인증키 발급이 필요하며 최근 이들 API를 유료화하려는 움직임이 있어 향후 사용에 부담이 예상됨
- 현재 기개발된 기술들에 적용된 SNS데이터는 빅데이터이긴 하나, 분석의 어려움을 고려하여 데이터 수집기간 등의 제한을 두어서 소규모데이터로 실험을 진행하였음, 향후 대규모데이터 적용 등을 통한 실증분석이 필요함

○ **(2-2세세부)** 공간 빅데이터 일괄분석 및 컴포넌트 기술 개발

- 총 30가지의 기능 개발을 목표로 하여 빅데이터 분석 성능을 담보하는 인메모리 기반 플랫폼인 Spark와 AWS 클라우드 시스템을 통해 각 컴포넌트의 개발이 이루어졌으며, 사용자의 편의성과 분석의 확장성을 고려한 Web 기반 플랫폼을 구축하였음
- 컴포넌트의 인풋 데이터 형태를 독립적으로 지정할 수 있는 기능을 추가하는 등 통합 플랫폼의 연동 및 수요처의 다양한 형태의 데이터를 처리할 수 있도록 기능적 확장성을 높이는 개선 및 최적화 작업을 수행하였지만, 분석을 위하여 데이터를 전처리하여야 하는 한계점이 존재함
- 유동인구를 포함한 새로운 형태의 공간 빅데이터와 궤적 정보를 가지는 이동객체 빅데이터에 대한 여러 분석 기능을 지원하는 분석 플랫폼으로써, 공공 및 민간의 빅데이터 분석 및 활용에 많은 기여를 할 수 있을 것으로 판단됨

○ **(2-3세세부)** 공간 빅데이터 CEP 및 OLAP 엔진 개발

- 공간 빅데이터 Interactive 분석 컴포넌트 요구사항 및 설계서 정의를 통해 개발하고자 하는 CEP 및 OLAP 엔진의 기본 청사진을 설계함. 다년차 연구로 인해 시시각각 변화하는 기술 환경에 따른 수정 보완을 진행하였음.
- 복합 이벤트 처리 모듈 및 OLAP 모듈 개발을 위해 각 모듈을 세분화하여 연구개발을 진행함. 각 모듈별로 사용자의 개입 및 사전 지식을 최소화하기 위한 사용성을 높이는 작업을 성공적으로 수행함. 실제 환경에 맞추어 추후 시범 테스트 및 성능 고도화가 필요함.

<표 3-5> 연구의 한계 및 시사점

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	미달성 내용	연구의 한계 및 시사점
2세부	공간 빅데이터 소셜 분석 기술 개발 (서울대학교 산학협력단)	공간 빅데이터 분석을 위한 모델 설계 및 목표 성능 설정	공간 빅데이터 분석모델 현황분석 및 요구사항 정의	• 해당사항 없음	• 시간이 흐름에 따라 현황 및 요구사항이 변할 수 있어 지속적인 분석이 요구됨
			공간 빅데이터 분석 목표모델 정의	• 해당사항 없음	• 없음
		공간 빅데이터 분석을 위한 데이터 가공, 변환, 필터링 및 분석효율 고도화	이기종 공간DB 특성 분석	• 해당사항 없음	• 소셜 미디어 데이터별로 제공되는 데이터 및 형태가 지속적으로 변경되고 있어 이에 대한 지속적인 반영이 요구됨
			이기종 DB간 데이터 가공, 변환, 필터링 기술 개발	• 해당사항 없음	• 소셜 미디어 데이터별로 제공되는 데이터 및 형태가 지속적으로 변경되고 있어 이에 대한 지속적인 반영이 요구됨
			다중 레이어를 위한 캐싱/프리패칭/버퍼링 기술 개발	• 해당사항 없음	• 없음
		공간 빅데이터 정합 및 주제도 분석 기술	공간 빅데이터 정합 및 주제도 분석 기술 개발	• 해당사항 없음	• 트위터에 특화된 정합 기술이 개발되었음. • 그러나 이는 차년도 연구에서 범용적인 비정형데이터를 대상으로 한 기술로 고도화되었음 • 일부 기능들이 포털사 등에서 제공하는 OpenAPI를 적용하여 이에 대한 지속적인 모니터링이 요구됨
		공간 빅데이터 분포 및 차트 분석 기술	공간 빅데이터 분포 분석 기술 개발	• 해당사항 없음	• 장소나 위치정보가 포함된 장소 리뷰에 대한 한글 코퍼스 데이터가 부족하여 영문 리뷰 데이터로 일부 기술 개발 • 일부 기능들이 포털사 등에서 제공하는 OpenAPI를 적용하여 이에 대한 지속적인 모니터링이 요구됨
			공간 빅데이터 분포 분석 고도화 기술 개발	• 해당사항 없음	• 일부 기능들이 포털사 등에서 제공하는 OpenAPI를 적용하여 이에 대한 지속적인 모니터링이 요구됨
			공간 빅데이터 차트 분석 기술 개발	• 해당사항 없음	• 일부 기능들이 포털사 등에서 제공하는 OpenAPI를 적용하여 이에 대한 지속적인 모니터링이 요구됨
			공간 빅데이터 차트 분석 고도화 기술 개발	• 해당사항 없음	• 일부 기능들이 포털사 등에서 제공하는 OpenAPI를 적용하여 이에 대한 지속적인 모니터링이 요구됨
			공간 빅데이터 분석결과 가시화 기술 개발	• 해당사항 없음	• 장소나 위치정보가 포함된 장소 리뷰에 대한 한글 코퍼스 데이터가 부족하여 영문 리뷰 데이터로 일부 기술 개발 • 일부 기능들이 포털사 등에서 제공하는 OpenAPI를 적용하여 이에 대한 지속적인 모니터링이 요

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	미달성 내용	연구의 한계 및 시사점
					구됨
			공간 빅데이터 분석결과 가시화 기술 개발(정적 가시화)	• 해당사항 없음	• 다양한 오픈 소스를 활용한 기술 개발이 요구될 것으로 보여짐
	공간 빅데이터 분석 및 가시화 컴포넌트 구현 (시터스)	공간 빅데이터 분석을 위한 모델 설계 및 목표 성능 설정	데이터 특성별 분석모델 설계	• 해당사항 없음	• 분석 프레임워크의 평가지표는 입력대비 출력 성능지표보다는 확장성에 초점을 맞추어 정의할 필요가 있음
		공간 빅데이터 분석을 위한 데이터 가공, 변환, 필터링 및 분석효율 고도화	이기종 DB간 데이터 가공, 변환, 필터링 모듈 개발	• 해당사항 없음	• 비정형 데이터의 유료화 및 서비스 제한으로 다양한 데이터를 수집하는데 어려움이 있음
			다중 레이어를 위한 캐싱/프리패칭/버퍼링 모듈 개발	• 해당사항 없음	• 클라이언트/서버간 통신 및 하드웨어 장비 성능에 따라 영향이 크게 좌우함
		공간 빅데이터 정합 및 주제도 분석 기술	공간 빅데이터 정합 및 주제도 분석 모듈 개발	• 해당사항 없음	• 구글 API를 사용하기 위한 인증키 발급이 필요하며 최근 유료화가 되어 사용에 부담이 있음
		공간 빅데이터 분포 및 차트 분석 기술	공간 빅데이터 분포 분석 모듈 개발	• 해당사항 없음	• 다양한 분포 분석을 통해 시계열 및 가시화 정보 제공
			공간 빅데이터 분포 분석 모듈 고도화	• 해당사항 없음	• 다양한 분포 분석 고도화를 통해 시계열 및 가시화 정보 제공
			공간 빅데이터 차트 분석 모듈 개발	• 해당사항 없음	• 다양한 차트 분석을 통해 시계열 및 가시화 정보 제공
			공간 빅데이터 차트 분석 모듈 고도화	• 해당사항 없음	• 다양한 차트 분석 고도화를 통해 시계열 및 가시화 정보 제공
		공간 빅데이터 분석결과 가시화 기술	공간 빅데이터 분석결과 가시화 모듈 개발 (정적 가시화)	• 해당사항 없음	• 비정형 데이터의 리뷰와 공간밀도 지수를 이용한 유동 인구 예측 지수 제공
			공간 빅데이터 분석결과 가시화 모듈 개발	• 해당사항 없음	• 이상치 탐지를 위한 시계열 분석 및 단어 임베딩을 통한 핫키워드의 워드클라우드 표출 등 동적 가시화 모듈 제공
		공간 빅데이터 사업단 실증	공간 빅데이터 분석 모듈 사업단 적용 및 실증	• 해당사항 없음	• 사업단과 함께 공간 빅데이터의 R&D 기술의 적용성 검토 단계를 거쳐 개발환경 구축 및 분석모듈을 적용하여 실행 결과를 제시
	공간 빅데이터 일괄(Batch)	공간 빅데이터 일괄(Batc	공간 빅데이터 통계기반 요구조건 분석	• 해당사항 없음	• R 패키지와 Literature review 조사를 통해 다양한 알고리즘을 지원하는 라이브러리의 개발이 필

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	미달성 내용	연구의 한계 및 시사점
	분석 컴포넌트 기술 개발 (연세대학교 산학협력단)	h) 분석 컴포넌트 기술 개발	R의 공간분석기능 검토		요하다는 것을 확인함
			공간 빅데이터 비통계기반 요구조건 분석 Machine Learning 기법 검토	• 해당사항 없음	• 국가 공간계획지원체계(KOPSS) 분석을 통한 공공 및 민간 수요처의 요구조건들이 확인되었음
			공간 빅데이터 분석 컴포넌트 인터페이스 설계	• 해당사항 없음	• 유연한 인터페이스 설계를 위한 ESB (Enterprise Service Bus) 컨셉이 활용되었음
			공간 빅데이터 분석 컴포넌트 아키텍처 설계	• 해당사항 없음	• 타 세부와의 빅데이터 연계를 위한 데이터 송수신 방법의 수정이 요구됨
			공간 빅데이터 기본분석 기능 선정 및 신규 개발 기능 결정(Predictive 기능 및 Descriptive 기능)	• 해당사항 없음	• 향후 수요처의 수요 변화, 연구개발의 동향에 따른 유연한 형태의 기본 분석 기능이 필요할 것으로 예상됨
			통계기반 분석 컴포넌트 기초 설계	• 해당사항 없음	• 기존 계획을 수정하여 인메모리 기반의 Spark를 사용한 컴포넌트 설계를 통해 분석 기능 성능 향상에 큰 영향을 주었음
			비통계기반 분석 컴포넌트 기초 설계	• 해당사항 없음	• 기존 계획을 수정하여 인메모리 기반의 Spark를 사용한 컴포넌트 설계를 통해 분석 기능 성능 향상에 큰 영향을 주었음
			공간 빅데이터 분석 컴포넌트 프로토타이핑	• 해당사항 없음	• 컴포넌트의 조합을 통하여 기능이 동작하도록 프로토타이핑을 진행함
			공간 빅데이터 기본 분석 기능 개선	• 해당사항 없음	• 세부 간 기능 중첩 개발을 피하기 위하여, 기존 공간 연산자가 포함된 분석 기능을 개선함
			통계기반 분석 성능 향상	• 해당사항 없음	• 기존 Java 기반의 분석 컴포넌트를 Spark의 Scala 언어 기반으로 수정하여 분석 성능이 향상됨
			비통계기반 분석 성능 향상	• 해당사항 없음	• 기존 Java 기반의 분석 컴포넌트를 Spark의 Scala 언어 기반으로 수정하여 분석 성능이 향상됨
			공간 빅데이터 분석 컴포넌트 기능 추가 및 최적화	• 해당사항 없음	• 기본 연구 목표를 수정하여, 총 30가지의 분석 기능을 제공하는 것을 목표로 기능을 추가하였음
			공간 빅데이터 분석 컴포넌트 기능 고도화	• 해당사항 없음	• 입력 데이터의 배열과 독립적으로 구동할 수 있도록, 입력 부분에 대한 확장이 필요함
			공간 빅데이터 기본 분석 기능 추가	• 해당사항 없음	• 기본 연구 목표를 수정하여, 총 30가지의 분석 기능을 제공하는 것을 목표로 기능을 추가하였음
			통계기반 분석 성능 최적화	• 해당사항 없음	• 효율적인 분산처리 기능을 수행하기 위해 Spark RDD 형태의 데이터 구조로 변경을 통해 최적화를 수행함
			비통계기반 분석 성능 최적화	• 해당사항 없음	• 효율적인 분산처리 기능을 수행하기 위해 Spark RDD 형태의 데이터 구조로 변경을 통해 최적화를 수행함

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	미달성 내용	연구의 한계 및 시사점
			공공/민간 수요기반 Batch 분석 시나리오 설계	• 해당사항 없음	• 총 30가지의 기능을 여러 공공기관에의 적용을 통해 분석 플랫폼의 활용 가능성을 확인함
			공간 빅데이터 Batch 솔루션 설계	• 해당사항 없음	• 연구단 통합 플랫폼 연동 지원을 위하여, 컴포넌트 연동성을 강화하였음
			공간 빅데이터 분석 Batch 솔루션 구현	• 해당사항 없음	• 독립적 구동으로 개발되었으나, 최종 성과물과의 연동을 위한 보완 작업이 필요함
	이동객체 데이터 Batch 분석 컴포넌트 (연세대학교 산학협력단)	이동객체 데이터 Batch 분석 컴포넌트 기술 개발	이동객체데이터 통계기반 요구조건 분석 R의 공간분석기능 검토	• 해당사항 없음	• R 패키지와 Literature review 조사를 통해 다양한 알고리즘을 지원하는 라이브러리의 개발이 필요하다는 것을 확인함
			이동객체데이터 비통계기반 요구조건 분석 Machine Learning 기법 검토	• 해당사항 없음	• 국가 공간계획지원체계(KOPSS) 분석을 통한 공공 및 민간 수요처의 요구조건들이 확인되었음
			이동객체데이터 분석 컴포넌트 인터페이스 설계	• 해당사항 없음	• 유연한 인터페이스 설계를 위한 ESB (Enterprise Service Bus) 컨셉이 활용되었음
			이동객체데이터 분석 컴포넌트 아키텍처 설계	• 해당사항 없음	• 타 세부와의 빅데이터 연계를 위한 데이터 송수신 방법의 수정이 요구됨
			이동객체데이터 기본분석 기능 선정 및 신규 개발 기능 결정 (Predictive 기능 및 Descriptive 기능)	• 해당사항 없음	• 향후 수요처의 수요 변화, 연구개발의 동향에 따른 유연한 형태의 기본 분석 기능이 필요할 것으로 예상됨
			통계기반 분석 컴포넌트 기초 설계	• 해당사항 없음	• 기존 계획을 수정하여 인메모리 기반의 Spark를 사용한 컴포넌트 설계를 통해 분석 기능 성능 향상에 큰 영향을 주었음
			비통계기반 분석 컴포넌트 기초 설계	• 해당사항 없음	• 기존 계획을 수정하여 인메모리 기반의 Spark를 사용한 컴포넌트 설계를 통해 분석 기능 성능 향상에 큰 영향을 주었음
			이동객체데이터 분석 컴포넌트 프로토타이핑	• 해당사항 없음	• 컴포넌트의 조합을 통하여 기능이 동작하도록 프로토타이핑을 진행함
			이동객체데이터 기본 분석 기능 개선	• 해당사항 없음	• 세부 간 기능 중첩 개발을 피하기 위하여, 기존 공간 연산자가 포함된 분석 기능을 개선함
			통계기반 분석 성능 향상	• 해당사항 없음	• 기존 Java 기반의 분석 컴포넌트를 Spark의 Scala 언어 기반으로 수정하여 분석 성능이 향상됨
			비통계기반 분석 성능 향상	• 해당사항 없음	• 기존 Java 기반의 분석 컴포넌트를 Spark의 Scala 언어 기반으로 수정하여 분석 성능이 향상됨
			이동객체데이터 분석 컴포넌트 기능 추가 및 최적화	• 해당사항 없음	• 기본 연구 목표를 수정하여, 총 30가지의 분석 기능을 제공하는 것을 목표로 기능을 추가하였음
			이동객체데이터 분석 컴포넌트 기능	• 해당사항	• 입력 데이터의 배열과 독립적으로 구동할 수 있도록, 입력 부분

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	미달성 내용	연구의 한계 및 시사점
			고도화	없음	에 대한 확장이 필요함
			이동객체데이터 기본 분석 기능 추가	• 해당사항 없음	• 기본 연구 목표를 수정하여, 총 30가지의 분석 기능을 제공하는 것을 목표로 기능을 추가하였음
			통계기반 분석 성능 최적화	• 해당사항 없음	• 효율적인 분산처리 기능을 수행하기 위해 Spark RDD 형태의 데이터 구조로 변경을 통해 최적화를 수행함
			비통계기반 분석 성능 최적화	• 해당사항 없음	• 효율적인 분산처리 기능을 수행하기 위해 Spark RDD 형태의 데이터 구조로 변경을 통해 최적화를 수행함
			공공/민간 수요기반 Batch 분석 시나리오 설계	• 해당사항 없음	• 총 30가지의 기능을 여러 공공기관에의 적용을 통해 분석 플랫폼의 활용 가능성을 확인함
			이동객체데이터 Batch 솔루션 설계	• 해당사항 없음	• 연구단 통합 플랫폼 연동 지원을 위하여, 컴포넌트 연동성을 강화하였음
			이동객체데이터 분석 Batch 솔루션 구현	• 해당사항 없음	• 독립적 구동으로 개발되었으나, 최종 성과물과의 연동을 위한 보완 작업이 필요함
	공간 빅데이터 CEP 및 OLAP 엔진 개발 (한국과학기술원)	공간 빅데이터 CEP 및 OLAP 엔진 개발	공간 빅데이터 Interactive 분석 컴포넌트 요구사항 정의서	• 해당사항 없음	• 분석 컴포턴트 개발을 위한 및 요구사항을 잘 정의하였음. 추후 기술 발전에 따른 지속적인 수정 개선이 요구되었음.
			공간 빅데이터 Interactive 분석 컴포넌트 설계서	• 해당사항 없음	• 분석 컴포턴트 개발을 위한 및 설계서를 잘 정의하였음. 추후 기술 발전에 따른 지속적인 수정 개선이 요구되었음.
			CQL 질의 메모리 상주 색인 및 저장 모듈	• 해당사항 없음	• 해당 처리 모듈을 성공적으로 개발하였음. 색인 기법으로 인해 CQL의 질의 성능을 향상시킬 수 있었음.
			공간 빅데이터 기반 복합 이벤트 스케줄러 모듈	• 해당사항 없음	• 스케쥴러 모듈을 통해 시시각각 변화하는 이벤트 유입 및 리소스 변화에 효과적으로 대처할 수 있었음.
			CQL 질의 클러스터 내 최적 할당 모듈	• 해당사항 없음	• 할당 모듈을 통해 다양하게 변화하는 이벤트 및 규칙을 효과적으로 클러스터 내에 배분할 수 있었음
			공간 빅데이터 기반 질의 처리기 모듈	• 해당사항 없음	• 공간 빅데이터 질의 처리기 모듈을 실시간 환경에 맞게 개발 및 적용 하였음
			공간 빅데이터 OLAP 데이터 로딩 모듈	• 해당사항 없음	• 공간 빅데이터를 OLAP 시스템 분석을 위해 효과적으로 적재할 수 있었음. 1세부의 시스템과 연계하기 위한 추가 작업이 소요되었음.
			공간 빅데이터 OLAP 드릴다운/롤업 모듈	• 해당사항 없음	• 드릴다운/롤업 모듈을 통해 인터렉티브한 분석을 가능하게 하였음.
			공간 빅데이터 OLAP 대화형 분석기 모듈	• 해당사항 없음	• 대화형 분석기 모듈을 통해 드릴다운/롤업 모듈 등의 기본 분석 도구에 대한 사용성을 높였음.
			공간 빅데이터 기반 복합 이벤트	• 해당사항 없음	• 복합 이벤트 네트워크를 최소한의 지식으로 클러스터 내에 생성

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	미달성 내용	연구의 한계 및 시사점
			네트워크 생성 모듈		및 배포할 수 있는 기술을 개발하였음.
			공간 빅데이터 기반 복합 이벤트 처리기 UI/UX	• 해당사항 없음	• 복합 이벤트 처리기의 구성, 사용, 및 보고에 필요한 기본적인 데이터 흐름 및 클러스터 구성 로직을 사용자의 최소한의 지식만으로 구현 가능하게 되었음.
			공간 빅데이터 OLAP UI/UX	• 해당사항 없음	• OLAP UI/UX를 통해 공간 빅데이터의 다차원 분석을 보다 직관적으로 수행 가능하게 하였음.
			공간 빅데이터 Interactive 분석 컴포넌트 테스팅 및 튜닝	• 해당사항 없음	• OLAP/CEP 엔진 테스팅 및 튜닝을 통해 보다 다양한 실제 환경에서의 시스템 도입 가능성을 평가하였음.
			공간 빅데이터 Interactive 분석 컴포넌트 시범 서비스 적용 지원	• 해당사항 없음	• 시범 서비스 적용 지원을 위해 사업단 실증을 완료하였으며, TTA 인증을 통해 보다 현실적인 상황에서의 실제 시스템 성능을 평가하였음.

(3) 공간 빅데이터 서비스 활성화 및 실증 연구

○ **(3-1 세세부)** 공간 빅데이터 서비스 활성화 방안 연구

- 공간 빅데이터 서비스 플랫폼을 기반으로 부동산, 지적, 교통, 소셜 비정형 데이터 등 전국단위의 국토교통 공간 빅데이터를 융·복합하여 공공/민간 부문에서 활용할 수 있는 서비스 모델 개발 및 적용 가능성 실증
- 전국단위의 대용량 공간 빅데이터의 수집에 애로사항이 존재하여 다양한 형태의 공간 빅데이터를 연구 성과에 적용하는데 한계가 있음
- 다양한 유형의 대용량 공간 빅데이터를 확보하여 공간 빅데이터 서비스 플랫폼에 적용할 경우, 추가 보완되어야할 기술적 이슈의 검토와 기술개선이 요구됨

○ **(3-2 세세부)** 공간 빅데이터 서비스 실용화 연구

- 공간 빅데이터를 표준화된 Open API 기반으로 상호 개방•공유•활용 가능하게 하고 실제 Business에 손쉽게 적용할 수 있도록 하는 기술을 개발함
- 사용자 접근성을 고려한 상호작용 기술과 사용자의 경험을 바탕으로 효율적인 공간 빅데이터 서비스 제공을 위한 UI/UX 방안을 개발함

○ **(3-3 세세부)** 공간 빅데이터 서비스 실증 연구

- 세부별로 연구한 각각의 독립적인 환경을 통합하고 일련의 시나리오에서 공간 빅데이터 기능을 수행하고 실증할 수 있는 시스템 구축
- 타세부에서 개발한 결과를 통합하여 연계하는 과정에서 세부별로 다른 개발 정도에 따라 하나의 시스템으로 통합하기 위한 표준화 과정이 미비하여 통합 과정에서 많은 시행착오가 발생함
- 공간연산, 공간 분석 기능별로 상이한 분석시간을 가지고 있으므로 사용자 측면에서 분석시간 예측에 대한 어려움이 존재함.
- 공공기관에서 연계한 데이터 무결성 확보를 위한 주기적인 데이터 관리가 필요함

<표 3-6> 연구의 한계 및 시사점

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	미달성 내용	연구의 한계 및 시사점
3세부	공간 빅데이터 플랫폼의 융복합 활용방안 개발 (안양대학교 산학협력단)	공간 빅데이터 융합 서비스의 서비스 프레임워크 및 구성요소 구현방안 제시	V-World 융합기반 공간 빅데이터 플랫폼의 서비스 프레임워크 도출	• 해 당 사 항 없음	• 국가공간정보통합체계를 통해 구축 및 수집한 공간정보를 정부, 공공, 민간의 다양한 수요에 맞게 공간 빅데이터 기술을 적용하는 서비스 프레임워크 구축 필요
			서비스 프레임워크 구성요소 개발방안 제시	• 해 당 사 항 없음	• 서비스 프레임워크 구성요소 개발방안을 통하여 1-2-3세부간의 개별 구현 기능 및 기능 간 연계를 위한 통합적인 이해를 기반으로 하는 플랫폼 구축 필요
		공간 빅데이터 융합 서비스의 프레임워크 매뉴얼 작성	공간 빅데이터 서비스 활성화를 위한 정책적, 제도적 지원방안 제시	• 해 당 사 항 없음	• 공간 빅데이터 수집, 관리, 분석, 활용을 위한 정책적, 제도적 지원방안을 통해 정보서비스에 제한되는 각종 이슈해결 및 서비스 활용지침 마련에 활용 가능
			서비스 프레임워크 활용 가이드라인 개발	• 해 당 사 항 없음	• 공간 빅데이터 플랫폼 활용을 위한 1-2-3세부간의 연계요소들에 대한 가이드라인 개발을 통하여 공간 빅데이터 서비스 플랫폼 활용 전략 제시
		공간 빅데이터 활용가이드 개발	공간 빅데이터 플랫폼 활용 가이드 개발	• 해 당 사 항 없음	• 공간 빅데이터 플랫폼의 저장, 관리, 분석, 서비스 구현 기능에 대한 통합적인 이해를 제공하고, 관련 기술의 교육과 활용에 소요되는 비용과 시간을 절감하는데 기여
			핵심 서비스모델 기반 활용 가이드 적용예시	• 해 당 사 항 없음	• 적용예시를 통하여 기술 수요처 관계자가 현장적용 가능성과 기대효과를 미리 가늠해볼 수 있는 구체적인 근거를 제공
	공간 빅데이터의 공공/민간 부문 핵심 서비스 모델 도출 및 제공 방안 개발 (안양대학교 산학협력단)	핵심 서비스 모델 도출을 위한 서비스 현황 분석 및 시사점 도출	공공/민간 부문, 서비스 유형별 공간 빅데이터 서비스 현황 수집 및 분석	• 해 당 사 항 없음	• 공간 빅데이터 서비스 플랫폼 개발과정에서 다양한 공공시스템에 연계 및 적용과 민간부문의 신규 공간 빅데이터 서비스 적용을 고려하여 기술개발 추진 필요
			국내외 공간 빅데이터 활용사례 분석을 통한 기존 서비스와 공간 빅데이터 서비스와의 차별화 방안 도출	• 해 당 사 항 없음	• 기술 개발 결과물의 수요자 또는 시장 조사를 통해 요구되는 서비스를 발굴하여 연구개발 진행 추진 필요
		공간 빅데이터 서비스 유형화 및 핵심 서비스 모델 개발	핵심 서비스 유형 및 서비스 모델 개발(1)	• 해 당 사 항 없음	• 사용자 요구사항 및 개발 분석기술, 데이터 연계방안을 통하여 사회, 환경, 경제 등 다양한 분야의 수요와 요구기술을 적용하여 부문별 각종 서비스 모델 개발 지원 가능
			핵심 서비스 유형 및	• 해 당 사	• 융합, 기술, 기타(정치) 부문

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	미달성 내용	연구의 한계 및 시사점
			서비스 모델 개발(2)	항 없음	의 핵심 서비스 모델을 추가 개발하여 실증을 위한 계획을 구체화하고 사업단과 연계하여 실증할 수 있도록 지원 가능
		핵심 서비스의 비즈니스 모델 개발	공공/민간부문 공간 빅데이터 활용을 위한 비즈니스 모델 분석 및 개발	• 해 당 사 항 없음	• 미래사회 이슈 및 트렌드를 분석하여 공간 빅데이터의 경향을 파악하고, 정부정책 및 현재 이뤄지고 있는 주요기술을 검토하여 비즈니스모델을 개발함으로써 공간 빅데이터 서비스의 활용성 제고
			공공/민간부문 공간 빅데이터 활용을 위한 서비스 시나리오 개발	• 해 당 사 항 없음	• 전국단위의 부동산, 지적, 교통, 소셜네트워크 등 국토교통 공간 빅데이터와 민간데이터를 융복합하여 공공/민간 시스템에서 활용하는 서비스 모델 개발 및 실증 필요
		핵심서비스 제공방안 개발	서비스 구현을 위한 시스템 아키텍처 설계	• 해 당 사 항 없음	• 민간/공공 시스템의 서비스와 공간 빅데이터 서비스 플랫폼 간의 연계를 통하여 다양한 매시업 서비스를 생성할 수 있도록 지속적인 기술개선이 필요
			핵심서비스의 주요기능 도출(1)	• 해 당 사 항 없음	• 1-2-3세부 구현 성과 간의 보다 긴밀한 연계를 위한 지속적인 통합작업이 필요함
			핵심서비스의 주요기능 도출(2)	• 해 당 사 항 없음	• 기존의 공간 연산자를 개선하고, 새로운 공간 연산자를 개발하는 등 연구 성과를 지속적으로 개선하여 다양한 공간 빅데이터 분석 서비스를 지원해야함
	공간 빅데이터 확장 Open API 서비스 (공간인소프트)	Open API 기술 현황 분석	V-World에서 제공하는 Open API 사이트를 통한 현황파악	• 해 당 사 항 없음	• Open API 기술의 국내외 적용사례조사, 공간정보 오픈플랫폼 Open API현황 분석, 서비스별 UI/UX 기술 및 적용사례 분석, 공간 빅데이터 서비스 제공을 위한 UI/UX 목표정의
			Open API 국내외 적용 사례 조사	• 해 당 사 항 없음	
		공간 빅데이터 Open API 설계	공공시스템 기반 공간 빅데이터 Open API 서비스를 위한 공간 빅데이터 콘텐츠 구성 및 인터페이스 설계	• 해 당 사 항 없음	• 공공시스템 기반 공간 빅데이터 Open API 서비스 콘텐츠 구성 및 인터페이스설계
			공간정보와의 연계범위 및 API 종류 설계	• 해 당 사 항 없음	
		공간 빅데이터 Open API 기술 개발	기능성과 사용성을 고려한 공간 빅데이터 확장 서비스를 위한 Open API 기능 개발	• 해 당 사 항 없음	• 핵심서비스 모델을 바탕으로 Open API 기능 개발 • 개발된 Open API 서비스 웹 페이지를 S/W 등록
		공간 빅데이터 Open API	공공시스템 기반 공간 빅데이터 Open API 서비스 테스트 사이트	• 해 당 사 항 없음	• 핵심서비스 모델을 바탕으로 Open API 기능 개발 완료, 개발된 Open API 서비스 기능을 S/W 등록

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	미달성 내용	연구의 한계 및 시사점
		서비스 시스템 구축	개발		
			연구성과물 성능 검증 지원 및 개선방안 연구	• 해 당 사 항 없음	
	공간 빅데이터 서비스 제공 UI/UX (공간인소프 트)	UI/UX 기술 현황 분석	서비스별 UI/UX 기술 및 적용 사례 분석	• 해 당 사 항 없음	• 핵심서비스 모델에 대한 논리모델과 물리모델 정의 • 스토리보드 작성을 통해 UI/UX 설계 완료
			공간 빅데이터 서비스 제공을 위한 UI/UX 개발 목표 및 범위 정의	• 해 당 사 항 없음	
		공간 빅데이터 서비스 제공을 위한 UX 개발전략 수립	사용자 분석 및 공간 빅데이터 니즈 조사	• 해 당 사 항 없음	• 공간 빅데이터 서비스 UI/UX 니즈 조사 • 서비스의 단위 기능에 대한 User Interface를 구성하고 이를 시스템으로 개발
		공간 빅데이터 서비스 UI/UX 기술 개발	공간 빅데이터 서비스 제공을 위한 UI/UX 상세설계	• 해 당 사 항 없음	• 개발된 Open API 서비스 이용정보를 사용자에게 제공하기 위한 웹 사이트 개발 완료
			공간 빅데이터 서비스 제공을 위한 UI/UX 테스트 시스템 개발	• 해 당 사 항 없음	
		공간 빅데이터 UI/UX 서비스 시스템 구축	공공시스템 기반 공간 빅데이터 Open API 서비스를 적용한 UI/UX 테스트 사이트 개발	• 해 당 사 항 없음	• 개발된 Open API 서비스 웹 사이트를 S/W 등록 완료 • 개발된 사용자 인터페이스를 S/W 등록 완료
			연구성과물 성능 검증 지원 및 개선방안 연구	• 해 당 사 항 없음	
	공공시스템기반 국토교통분야 (부동산, 지적, 교통 등) 공간 빅데이터 실증 서비스 시스템 (포도)	공간 빅데이터 및 다양한 공공 시스템 기술현황 분석	공간 빅데이터 현황 및 서비스 구축 사례 분석	• 해 당 사 항 없음	• 공간 빅데이터 현황 및 서비스 구축 사례 분석
			공간정보 오픈플랫폼 실증 기술 현황 분석	• 해 당 사 항 없음	• 공공에서 제공하는 기본적인 공간정보 이외에도, 민간부문에서의 공간정보의 요구가 지속적으로 증가하는 추세
		공간 빅데이터 서비스 시스템 부문별 서비스 및 실증 기술 설계	빅데이터 서비스 아키텍처 및 프레임워크 설계	• 해 당 사 항 없음	• 선도적으로 공간 데이터를 개방했던 해외처럼 국내에서도 데이터 개방 요구가 증가
			공간 빅데이터 서비스 시스템 실증 기술 설계	• 해 당 사 항 없음	• 전자정부 프레임워크를 기반으로 컨테이너 형식의 프레임워크 설계 수행
			저장, 관리, 분석, 가시화 등 개별 서비스 설계	• 해 당 사 항 없음	• 각 세부별 연구성과 기술을 통합하여 관리 및 운용할 수 있는 시스템 통합기술을 설계
		공간 빅데이터 서비스 통합 기술 설계 및 프로토타입 개발	공간 빅데이터 서비스 통합 기술 설계	• 해 당 사 항 없음	• 세부별 연구기술을 개별적으로 실증 및 운용할 수 있도록 서비스 시스템 설계
			공간 빅데이터 서비스 통합 기술 프로토타입 개발	• 해 당 사 항 없음	• 세부별 기술들의 모니터링을 통해 개별 서비스에 대한 프로세스 확인

구분	과제명 (연구기관)	연구목표	연구목표별 Activity	미달성 내용	연구의 한계 및 시사점
		공간 빅데이터 서비스 시스템 실증 및 연계 기술 개발	공간 빅데이터 서비스 시스템 실증 기술 개발	• 해 당 사 항 없음	• 개별서비스를 통합하여 단일 서비스로 구성 • 세부별로 상이한 기술 개발 단계로 모든 기술에 대한 테스트 수행에 어려움
			저장, 관리, 분석, 가시화 등 개별 서비스 기술 개발	• 해 당 사 항 없음	• 하둡 프레임워크 및 2세부의 분석 모듈 등의 데이터 시각화 지원
			공간 빅데이터 서비스 통합 기술 개발	• 해 당 사 항 없음	• 세부별로 독립적인 개발환경을 취합하고 이를 통일된 환경으로 구축하기 위해 컨테이너 형식의 테스트배드 구축 및 지원
		공간 빅데이터 서비스 실증 시스템 구축 (연계서비스 포함)	공간정보 오픈플랫폼 기반 공간 빅데이터 서비스 시스템 구축	• 해 당 사 항 없음	• 세부별로 개발정도가 상이하여 정해진 시나리오만 제공가능하므로 통합적인 개발 일정 계획 수립 필요
	공공시스템기반 타 부처 공간 빅데이터 실증 서비스 시스템 (성과물 추가) (포도)	국내 공공 공간 빅데이터 구축 및 시각화 기술 현황 분석	국내 공공 공간 빅데이터 구축 및 시각화 기술 현황 분석	• 해 당 사 항 없음	• 상이한 분석 시간에 따른 사용자 측면의 어려움
		공공 공간 빅데이터 서비스 연계/확장 기술 설계(1) 및 프로토타입 개발	공공 공간 빅데이터 서비스 연계 및 확장 기술 설계(1)	• 해 당 사 항 없음	• 국내 공공 공간 빅데이터 구축현황의 예시로 VWorld 외에 활용가능한 공간정보 오픈플랫폼이 미비함
			공공 공간 빅데이터 서비스 연계 및 확장 기술 프로토타입 개발	• 해 당 사 항 없음	• 서비스 컨테이너의 꾸준한 추가를 위해 분석된 내용의 정확성 및 연계성 검증
		공공 공간 빅데이터 서비스 연계/확장 기술 설계(2)	공공 공간 빅데이터 서비스 연계 및 확장 기술 설계(2)	• 해 당 사 항 없음	• 서비스모델이 실제로 구현되고 분석되었다는 가정하에 설계를 추가함 • 프로토타입 개발을 통해 시각화와 서비스 모델이 실제로 구현되는지 확인
		공공 공간 빅데이터 서비스 연계 및 확장 기술 개발	공공 공간 빅데이터 서비스 연계 및 확장 기술 개발	• 해 당 사 항 없음	• 통계청, 국민안전처 등 공공 부문의 공간 빅데이터를 연계하고 공공 데이터 표준화 방안의 기반 마련
		공간 빅데이터 서비스 실증 시스템 확장 구축	공공 시스템 기반 공간 빅데이터 서비스 시스템 확장 구축	• 해 당 사 항 없음	• 공공 공간 빅데이터의 데이터 표준화 기술개발을 통해 공간빅데이터에서 사용가능한 데이터 형식으로 가공하고 이를 공간연산, 공간 분석 기능과 연계할 수 있는 확장 기능 지원

다. 관련 분야 기여도

(1) 공간빅데이터 저장 관리 인프라 기술 개발

○ **[핵심성과 A] Hadoop 기반 공간 빅데이터 저장 프레임워크**

- 공간 빅데이터의 저장 및 관리를 제공하기 위한 연산자를 제공함으로써 연산자를 조합하여 다양한 공간데이터 처리를 할 수 있는 프레임워크를 지원함
- 개발자가 MapReduce의 세부적인 데이터 처리 과정을 알지 못하여도 Hadoop 환경에서 공간 데이터의 효율적인 처리 기능을 쉽게 사용할 수 있게 함

○ **[핵심성과 B] 공간 빅데이터 질의처리 엔진**

- SQL 기반 질의언어에 의한 공간 빅데이터에 대한 질의를 지원하고 JDBC 표준 인터페이스를 지원함으로써 편리하게 공간 빅데이터 저장 프레임워크의 기능을 사용할 수 있게 함
- 외부의 응용프로그램이 Hadoop 환경에 저장된 공간 빅데이터를 이용한 분석 프로그램을 작성할 때 기존의 GIS 또는 공간 RDBMS의 데이터를 사용하던 방법과 동일하게 데이터를 접근하여 사용할 수 있음

○ **[핵심성과 C] 인메모리 기반 공간 빅데이터 저장관리 프레임워크**

- 공간 빅데이터 자료를 Hadoop에 적재후 인메모리 기반에서 저장 및 관리할수 있는 프레임워크를 지원하고, 이를 기반으로 공간분석 라이브러리 23종 개발을 통해 기본적인 공간분석 작업을 가능하게 함
- 인메모리 기반에서 이동객체 자료에 대한 고속 적재 및 시공간 색인 기능을 통해 이동객체 빅데이터 분석을 위한 기반 시스템을 마련함
- Spark에서 제공하는 SparkSQL 기반 문법에 Spatial SQL 37종을 개발하여 SQL 기반으로 다양한 공간분석 작업을 할 수 있음

○ **[핵심성과 D] 스트림 기반 분산 공간 ETL 도구**

- 전형적인 ETL 기능이외에 스트림 기반 공간 데이터 ETL 처리가 가능하며 RESTful 인터페이스 연동 기능 추가로 다양한 공간정보에 대한 ETL 처리를 지원함
- 공간 빅데이터 성능 향상을 위한 병렬 및 분산 처리 플랫폼을 구축하고 공간정보 ETL의 병렬/분산화를 통한 스트림 데이터 처리 속도 향상 서비스를 지원함
- 기존 ETL 처리에 반복적 수행을 자동화하기 위한 워크플로우 기반 처리 기능을 추가하고 동적인 워크플로우 동작을 위한 규칙 적용 기능을 지원함

<표 3-7> 관련 연구분야 기여도

구분	과제명 (연구기관)	핵심 연구성과명	관련 분야 기여도
1세부	공간 빅데이터 저장관리 인프라 기술 개발 (한국전자통 신연구원)	(A) Hadoop 기반 공간 빅데이터 저장 프레임워크	(연산자 기반 시스템) • 연산자와 레코드 스트림을 개념을 이용하여 다양한 데이터 처리 방법을 조합(복합 연산자)하여 개발할 수 있도록 하여 재활용성 및 확장성이 큰 성과물을 개발함 • 복합 연산자를 Hadoop 맵리듀스 방식으로 자동으로 매핑하는 기술과 매핑시 맵리듀스 작업을 최소화하는 방법을 개발함 (색인) • 생성된 R-Tree 간의 상호간섭, 즉 겹치는 영역 및 색인공간의 총합을 줄임으로써 R-Tree의 성능을 높임 • Hadoop의 병렬처리 기능을 사용하여 R-Tree 색인 구축 시간을 줄이고 개별 R-Tree의 색인공간을 적절하게 파티션함으로써 검색 성능을 높임 (공간 조인) • 분산색인 구조 기반의 공간 연산/분석 기능 지원 등 Hadoop 개선을 통한 공간 연산 성능을 향상시킴 • 다양한 공간 조인 연산에서 활용될 수 있는 Inner 블록의 grouping을 활용한 중첩 루프 조인 알고리즘을 개발함 • outer 데이터세트의 locality가 낮은 경우 성능 개선 효과가 높은 조인 알고리즘을 개발하였으며 최대 2.9배 조인 성능을 개선함 (공간 빅데이터 활용) • 원격 인터페이스와 분석응용 개발 도구 지원으로 활용성을 증대시킴 • 많은 공간 분석 개발자들이 Hadoop 환경에 익숙하지 않는 것을 고려할 때, 원격 공간 빅데이터 활용 기술은 응용 개발 편의성을 상당히 높일 수 있음 • 특정 언어나 특정 OS에 국한되지 않고, 다양한 언어와 운영체제에서 공간 분석 응용을 개발할 수 있는 환경을 제공함 • 공간 빅데이터 분석 응용 도구를 개발하여 공간 분석 응용 개발 및 시스템 운영에 도움을 줄 수 있음
		(B) 공간 빅데이터 질의처리 엔진	(질의언어에 의한 고급 질의) • SQL기반 질의언어를 통한 질의 기능을 지원하여 API 호출이 아닌 질의언어에 의하여 공간 빅데이터 저장관리 시스템의 기능을 사용할 수 있음 • OGC Simple Feature Specification에서 정의된 표준 공간데이터 타입과 함수를 사용할 수 있음 (표준 인터페이스) • 응용프로그램이 Java 코드에서 표준 인터페이스/클래스/메서드를 통하여 공간 빅데이터 저장관리시스템의 기능을 쉽고 편리하게 사용할 수 있게 함 • JDBC 표준 인터페이스를 제공함으로써 기존 다른 공간 DBMS나 GIS를 사용하던 응용프로그램이 공간 빅데이터 저장관리시스템으로 이전할 때 코드 변경을 최소화함 • 공간 빅데이터 저장관리시스템과 다른 공간 DBMS의 데이터를 동시에 사용하여 분석 알고리즘을 수행할 수 있는 유연성을

			제공함 • 서버의 변경사항이 클라이언트 응용프로그램의 인터페이스에 영향을 주지 않는 구조로 편의성과 확장성을 제공함
	인메모리 기반 공간 빅데이터 저장관리 프레임워크 개발 (포스웨이브)	(O) 인메모리 기반 공간빅데이터 저장관리 프레임워크	(공간분석 라이브러리) • 인메모리 기반 공간분석 라이브러리 23종 개발을 통해 플랫폼기반에서 기본적인 공간분석에 활용할수 있게 하였음 • 기존 공간통계 분석의 경우 분산 메모리 기반의 알고리즘 개발을 통해 고속의 분산처리를 통해 분석 성능을 향상 시켰음 (색인) • 그리드 기반의 색인 기법을 기반으로 일반적인 공간자료와 이동객체 공간자료에 대한 공간 및 시공간 색인을 인메모리 기반으로 고속 생성 및 조회가 가능하게 함 • 시공간 색인의 경우 공간그리드 및 시공간 그리드 기법을 고안하여 자료가 지속적으로 증가하더라도 색인 생성 및 조회시 성능 저하가 발생하지 않도록 하였음 (분석 인터페이스) • 인메모리 기반 분석시 사용하는 Scala와 Python 스크립트에서도 별도의 작업없이 동일한 인터페이스를 통해 플랫폼의 기능을 쉽게 활용할 수가 있음 • SparkSQL기반 37종의 Spatial SQL 개발을 통해 SQL기반으로 공간분석 작업을 수행할 수 있음
	공간 빅데이터 웨어하우스 핵심기술 개발 (아이씨티웨이)	(D) 스트림 기반 분산 공간 ETL 도구	(다양한 공간정보) • 공간정보에 대표적인 Shp, CSV, Geo-JSON 형태에 공간정보 ETL 기능을 제공하고 내부 포맷으로 변환하여 공간 빅데이터 저장관리시스템과 연동이 가능하게 함 • 소셜 네트워크 서비스에서 실시간 JSON 데이터를 취득하고 이를 활용하여 공간정보 ETL 처리 기능을 제공함 (병렬 및 분산 플랫폼) • 스트림 방식에 발생되는 공간 정보를 수집, 변환, 적재하는 공간 ETL 서버 엔진과 내부 클라이언트용 API를 개발하여 연동 실험함 • 다양한 스트림 형태의 공간정보를 생성하기 위한 공간 센서 에뮬레이터를 개발완료하고 이를 활용하여 Shp, CSV, TXT, Geo-JSON, WKT, WKB 수집 모듈을 고도화함 (공간 ETL 기능) • 다양한 형태의 공간 빅데이터와 batch와 interactive한 방식의 데이터 획득 방식을 지원하는 분산 병렬 공간 빅데이터 획득 기능이 구현됨 • 공간 ETL 규칙처리 엔진을 개발하여 사용자 정의 규칙 기반 ETL 워크플로우 서비스 실행 기능을 구현함

(2) 공간 빅데이터 분석 및 가시화 기술 개발

○ [핵심성과 E] 공간 빅데이터 소셜 분석 기술 개발

- 공간 데이터에서 위치정보(위도, 경도)를 추출하기 위한 기술을, 블로그, 트위터 구글 리뷰 등 장소에 기반한 주제도 분석을 위한 기술을 개발하였음
- 공간에 특화된 감성분석, 장소정보 분석, 문장 추출, 키플레이어 탐지 기술, 정적 및 동적 가시화 기술을 개발하였음을 개발하였음

○ [핵심성과 E] 공간 빅데이터 분석 및 가시화 컴포넌트 구현

- 다양한 비정형 데이터 수집에 따른 각종 공간 빅데이터 분석의 주제도 모듈 제공
- 소셜 미디어 데이터에서 공간 정보를 이용한 빅데이터 분석 서비스 제공
- 분석 모듈 가시화를 위한 분포 및 차트 모듈 제공
- 모든 분석 모듈의 분석 결과는 JSON 형태의 OpenAPI 서비스 제공

○ [핵심성과 F] 공간 빅데이터 일괄(Batch) 분석 컴포넌트 기술 개발

- 공간 빅데이터에 특화된 분석 솔루션은 사례를 거의 찾아볼 수 없으며, 탑재된 분석 컴포넌트를 사용자가 Drag & Drop 방식으로 목적에 맞게 사용할 수 있다는 점에서 사용자 편의성까지 갖추었음
- 웹플랫폼 기반으로 초기 서버 구축 이후에 공간 빅데이터 분석 결과 필요시 시공간에 구애받지 않고 제공받을 수 있으며, 대용량의 데이터 분석을 수행할 수 있음

○ [핵심성과 G] 이동객체 데이터 Batch 분석 컴포넌트

- 이동객체데이터에 특화된 분석 솔루션은 사례를 거의 찾아볼 수 없으며, 탑재된 분석 컴포넌트를 사용자가 Drag & Drop 방식으로 목적에 맞게 사용할 수 있다는 점에서 사용자 편의성까지 갖추었음
- 웹플랫폼 기반으로 초기 서버 구축 이후에 이동객체데이터 분석 결과 필요시 시공간에 구애받지 않고 제공받을 수 있으며, 대용량의 데이터 분석을 수행할 수 있음

○ [핵심성과 H] 공간 빅데이터 CEP 및 OLAP 엔진 개발

- 본 기술을 통해 많은 지식과 노하우가 요구되는 초고속 대용량의 공간 빅데이터 분

석(실시간 대화형 패턴 탐지 및 다차원 분석)의 진입 장벽을 낮춤. 특히 유저 친화적인 UI/UX 개발 및 클러스터 내 자동 배포 기능을 통해 활용성을 높임.

<표 3-8> 관련 연구분야 기여도

구분	과제명 (연구기관)	핵심 연구성과명	관련 분야 기여도
2세부	(E) 공간 빅데이터 소셜 분석 기술 개발 (서울대)	공간 빅데이터 분석을 위한 모델 설계 및 목표 성능 설정	• 공간 빅데이터를 정의하고 이를 분석하기 위한 모델 및 목표 성능을 설정하였다는데 의의가 있음
		공간 빅데이터 분석을 위한 데이터 가공, 변환, 필터링 및 분석효율 고도화	• 공간에 특화된 소셜 미디어 데이터 분석을 위하여 데이터 가공, 변환, 필터링 기술을 제시하였으며, 대용량인 지도 기반의 분석효율 고도화 기술을 제시하였음
		공간 빅데이터 정합 및 주제도 분석 기술	• 공간 데이터에서 위치정보(위도, 경도)를 추출하기 위한 기술을 개발하였으며, 블로그, 트위터 구글 리뷰 등 장소에 기반한 주제도 분석을 위한 기술을 개발하였음
		공간 빅데이터 분포 및 차트 분석 기술	• 공간에 특화된 감성분석, 장소정보 분석, 문장 추출, 키플레이어 탐지 기술을 개발하였음
		공간 빅데이터 분석결과 가시화 기술	• 공간에 특화된 정적 및 동적 가시화 기술을 개발하였음
	(E) 공간 빅데이터 분석 및 가시화 컴포넌트 구현 (시터스)	공간 빅데이터 분석을 위한 모델 설계 및 목표 성능 설정	• 기존 솔루션 분석 후 평가지표를 달성하기 위한 분석 프레임워크의 개념적 설계를 완료하였다는데 의미가 있음
		공간 빅데이터 분석을 위한 데이터 가공, 변환, 필터링 및 분석효율 고도화	• 공간에 특화된 소셜 미디어 데이터 분석을 위하여 데이터 가공, 변환, 필터링 모듈을 개발하였으며, 대용량인 지도 기반의 분석효율 고도화 모듈을 개발하였음
		공간 빅데이터 정합 및 주제도 분석 기술	• 공간 데이터에서 위치정보(위도, 경도)를 추출하기 위한 모듈을 개발하였으며, 블로그, 트위터 구글 리뷰 등 장소에 기반한 주제도 분석을 위한 모듈을 개발하였음
		공간 빅데이터 분포 및 차트 분석 기술	• 공간에 특화된 감성분석, 장소정보 분석, 문장 추출, 키플레이어 탐지 모듈을 개발하였음
		공간 빅데이터 분석결과 가시화 기술	• 공간에 특화된 정적 및 동적 가시화 모듈을 개발하였음
	(F) 공간 빅테이터 Batch 분석 (연세대/가이온)	공간 빅테이터 Batch 분석 컴포넌트	• 공간 빅데이터에 특화된 분석 솔루션은 사례를 거의 찾아볼 수 없음 • 탑재된 이동객체데이터 분석 컴포넌트를 사용자가 Drag & Drop 방식으로 목적에 맞게 사용할 수 있다는 점에서 사용자 편의성까지 갖추었음
	(G) 이동객체테이터 Batch 분석 (연세대/가이온)	이동객체테이터 Batch 분석 컴포넌트	• 이동객체데이터에 특화된 분석 솔루션은 사례를 거의 찾아볼 수 없음 • 탑재된 이동객체데이터 분석 컴포넌트를 사용자가 Drag & Drop 방식으로 목적에 맞게 사용할 수 있다는 점에서 사용자 편의성까지 갖추었음
	(H) 공간 빅데이터 Interactive 분석 기술 개발(한국과학기술원/유엔진솔루션즈)	공간 빅데이터 Interactive 분석 기술 개발	• 초고속 대용량의 공간 빅데이터로부터 실시간 대화형으로 다양한 패턴 탐지(CEP 모듈) 및 다차원 분석(OLAP 모듈)를 가능하게 함 • 유저 친화적인 UI/UX 개발을 통해 비전문가의 경우에도 쉽게 원하는 분석 작업을 수행할 수 있도록 지원

(3) 공간 빅데이터 서비스 활성화 및 실증 연구

○ [핵심성과 I] 공간 빅데이터 플랫폼의 융복합 활용방안 개발

- 공간 빅데이터 저장, 관리, 분석 핵심기술을 통합한 기존과 차별화 되는 새로운 유형의 서비스 개발을 통하여 관련 산업에서 각종 서비스를 창출하는데 기여
- 공공 분야의 공간 빅데이터 분석을 통해 생성된 표준화된 데이터를 공간연산 기능 연계를 통해 다양한 서비스로 확장하고 이를 활용한 공간 정보 가치 확장 및 시장성 확보

○ [핵심성과 J] 공간 빅데이터의 공공/민간 부문 핵심 서비스 모델 도출 및 제공 방안 개발

- 부동산, 지적, 교통, 비정형 텍스트 등 전국단위의 국토교통 공간 빅데이터와 민간데이터를 융·복합 활용하는 서비스모델 및 제공방안 제시를 통하여 관련분야의 인적, 물적, 자원, 절약가능
- 단순 공간정보 및 통계정보 뿐만 아니라 SNS 및 Web, 실시간 데이터 등을 제공함으로써 부문별 공간 빅데이터 수집에 필요한 자료수집/가공시간 감소효과 기대

○ [핵심성과 K] 공간 빅데이터 확장 Open API 서비스

- 공간 빅데이터를 표준화된 Open API 기반으로 상호 개방•공유•활용 가능하게 하고 실제 Business에 손쉽게 적용할 수 있도록 하는 기술을 개발함

○ [핵심성과 L] 공간 빅데이터 서비스 제공 UI/UX

- 사용자 접근성을 고려한 상호작용 기술과 사용자의 경험을 바탕으로 효율적인 공간 빅데이터 서비스 제공을 위한 UI/UX 방안을 개발함

○ [핵심성과 M] 공공시스템기반 국토교통분야(부동산, 지적, 교통 등) 공간 빅데이터 실증 서비스 시스템

- 공간 빅데이터 기반 기술을 연계하여 실증하기 위한 통합 환경을 구축함. 이를 통해 추후 관련 산업에서 공간 정보 서비스 시스템 구축 시 기반 기술로 활용 가능
- 공간정보의 시각화를 통해 직관적인 서비스가 가능하므로 다양한 사용자에게 쉬운 접근성 제공

○ 핵심성과 N] 공공시스템기반 타 부처 공간 빅데이터 실증 서비스 시스템

- 표준화된 형태의 공공데이터와 서비스를 매쉬업하여 새로운 서비스 이용이 가능하고 이를 표준화하여 새로운 비즈니스에 적용하고 이를 통한 다양한 지식정보 서비스 제공

<표 3-9> 관련 연구분야 기여도

구분	과제명 (연구기관)	핵심 연구성과명	관련 분야 기여도
3세부	공간 빅데이터 서비스 활성화 방안 연구 (안양대학교 산학협력단)	(I) 공간 빅데이터 플랫폼의 융복합 활용방안 개발	• 기존의 GIS 소프트웨어로 처리할 수 없는 공간 빅데이터의 저장, 관리, 분석, 서비스 기능에 대한 이해를 제공 • 공간 빅데이터 플랫폼 관련 기술의 교육과 활용에 소요되는 비용과 시간을 절감하는데 기여
			• 공간 빅데이터 플랫폼의 기술 수요처 관계자는 활용 가이드 적용 예시를 근거로 기술 적용 대상과 범위를 구체화 하여 현장적용 가능성과 기대효과를 가늠해 볼 수 있음 • 다양한 영역에서 공간 빅데이터를 활용할 수 있는 적용 예시를 제공하여 기술 수요처의 지속적인 발굴을 통한 활용영역 다변화와 기술적용 시장 확대를 통한 빅데이터 산업의 고용창출에 기여
		(J) 공간 빅데이터의 공공/민간 부문 핵심 서비스 모델 도출 및 제공 방안 개발	• 현장에서 활용되고 있는 대용량의 공간 빅데이터를 적용하여 현장적용 가능성이 실증된 서비스 시나리오를 제공 • 서비스 시나리오는 기술 수요처의 관계자가 공간 빅데이터 플랫폼의 기술적용 사례와 기대효과를 구체적으로 가늠해 볼 수 있는 근거 제공 • 다양한 영역에서 공간 빅데이터를 활용할 수 있는 서비스 시나리오를 개발하여 기술 수요처의 지속적인 발굴을 통한 활용영역 다변화와 관련 기술 적용 시장을 확대하여 궁극적으로 빅데이터 산업의 발전과 고용창출에 기여
	공간 빅데이터 서비스 실용화 연구 (공간인)	(K) 공간 빅데이터 확장 Open API 서비스	• 공간 빅데이터를 표준화된 Open API 기반으로 상호 개방·공유·활용 가능하게 하고 실제 Business에 손쉽게 적용할 수 있도록 하는 기술을 개발함
		(L) 공간 빅데이터 서비스 제공 UI/UX	• 사용자 접근성을 고려한 상호작용 기술과 사용자의 경험을 바탕으로 효율적인 공간 빅데이터 서비스 제공을 위한 UI/UX 방안을 개발함
	공간 빅데이터 서비스 실증 연구(포도)	(M) 공공시스템기반 국토교통분야(부동산, 지적, 교통 등) 공간 빅데이터 실증 서비스 시스템	• 개별 연구성과를 모듈화하여 통합시스템에 연계가능하도록 시스템을 설계하여 향후 추가적인 분석기법이 도입될 경우 확장이 용이한 기술로 활용가능
			• 공간 빅데이터 실증 서비스 시스템 구축을 통해 공간 빅데이터 관련 사업의 진입 장벽을 낮춰 공간정보 시장의 활성화 기대
		(N) 공공시스템기반 타 부처 공간 빅데이터 실증 서비스 시스템	• 공공데이터와 다양한 콘텐츠, 서비스를 매쉬업하여 새로운 유형의 서비스 이용이 가능함. 이를 통한 새로운 비즈니스 창출 기대
			• 새로운 비즈니스를 활용한 우수한 지식정보 서비스를 제공함으로써 민간 사업의 수익 창출 기대

4. 연구개발성과의 활용 계획

가. 활용계획

(1) 공간빅데이터 저장 관리 인프라 기술 개발

(가) 활용 방안

○ 본 연구의 성과물은 공간빅데이터체계구축사업, 공간정보 오픈플랫폼, 국가공간정보통합체계 등 공간정보 관련 국토교통부 중점 사업에 활용할 계획임

- 빠르게 수집되는 민간의 공간 빅데이터의 활용이 확산되는 추세에 따라 대용량/실시간 처리 기술을 개발하여 국토부의 공공 서비스와 융합된 형태의 서비스에 활용할 계획임
- 저비용의 컴퓨팅 하드웨어 인프라를 이용하여 기존의 공간정보 및 실시간 공간 빅데이터를 효율적으로 저장, 관리하고 검색하기 위한 기반 기술로 활용 가능함
- 공간 빅데이터를 수집하여 저장, 관리하고 검색하여 데이터 제공을 필요로 하는 공공 및 민간부문의 공간 빅데이터의 저장 관리 엔진으로서 활용이 가능함
- 해당 공간 빅데이터 플랫폼을 필요로 하는 사업에 적용할 때는 성과물을 1단계 시범 적용하고 관련 기술지원을 통해 검증하여 성과물 적용의 안정성을 확보함

○ 활용 주체 및 수요처

- 본 연구 성과물의 예상되는 활용주체로는 국토교통부, 국토지리정보원, 국가공간정보통합센터, 공간정보산업진흥원을 포함한 공간 빅데이터의 저장, 관리, 분석, 서비스가 필요한 공공기관이 있음
- 국토교통부의 공간빅데이터 체계구축사업은 공간 빅데이터 저장관리 인프라의 기능과 성과물을 활용할 대표적인 수요처임
- 국가 통합 NSDI 구축에서 공간정보의 저장, 관리, 검색 시스템으로 적용 가능함
- 공간정보산업진흥원의 V-World 서비스는 대용량의 공간데이터 관리를 필요로 하며 빅데이터를 다룰 수 있는 본 성과물과의 융합 활용을 추진함
- 그 외 공간 빅데이터 서비스를 제공하는 포털이나 서비스 제공 기업과의 협의를 통하여 활용 대상을 확대할 계획임
- 공간 빅데이터의 활용이 확산됨에 따라 정부/공공/민간을 포함한 ICT 융합 서비스에 폭넓게 활용할 수 있음

○ 사업자 기술이전을 통한 성과물 활용

- 개발한 공간 빅데이터 기반 핵심기술은 민간 사업자에게 기술이전하고 기술실시에 의하여 공공사업에 적용함으로써 기술 확산을 유도함
- 공간빅데이터 체계구축사업의 경우 기존 플랫폼 및 기술분석에 따라 대상 기술을 식별하고 기술실시에 의하여 성과물을 활용함

○ 세부 기술 특성에 따른 성과물 활용

- 전통적인 빅데이터 처리방식에 추가하여 인메모리 기반의 작업 처리 기술을 통합 적용하여 공간 빅데이터 저장 및 처리 기반 프레임워크로 활용함
- 표준적인 질의 언어 또는 인터페이스를 제공하여 공간 빅데이터를 다루는 기존의 타 분석시스템 또는 응용프로그램에 대하여 기존의 사용하던 GIS나 DBMS를 본 성과물로 대치하는 형태로 활용함
- 인메모리 기반 공간 빅데이터 저장관리 시스템은 실시간 공간 빅데이터 처리용 모듈로 적용 예정임
- 이동객체 처리 기술은 최근 IoT 및 이동객체 데이터 등의 증가에 따라 IoT 및 실시간 이동객체 데이터의 분석·처리를 위한 실시간 데이터 처리의 코어 기술로 활용 가능함
- 공간 빅데이터 저장소에 특화된 ETL 역할을 하는 공간 ETL 기술은 공간 데이터 수집/변환/적재 기능을 요하는 시스템에 활용하며, 특히 센서, IoT 등으로부터 빠른 주기로 계속 입력되는 공간 빅데이터를 처리해야 하는 분야에 적용하기에 적합함
- 공간 ETL 툴킷은 대용량의 공간 정보를 원하는 하둡 시스템에 신속, 정확하게 추출, 변환, 적재시 활용될 것으로 기대됨

(나) 공간빅데이터체계구축사업 적용

① 적용 개요

○ 1세부의 성과물은 공간 빅데이터를 사용하는 분석시스템이나 응용프로그램에서 빅데이터 저장관리를 위한 기반 시스템으로 활용될 수 있음

- 대표적으로 중점 추진하고 있는 적용 대상 시스템은 국토교통부의 재정사업인 ‘공간빅데이터 체계구축사업’ 임
- 연구과제 초기부터 국토교통부, KAIA, 연구단, 사업단 등이 참여하는 사용자협의체에서 결정 및 추진해온 내용에 기반하여 향후 적용을 추진함

○ 공간빅데이터 체계구축사업에 대한 핵심 기술 지원

- 공간 빅데이터 플랫폼에서 실제 제공하는 분석모델 및 서비스를 중심으로 저장·관리 핵심기술을 적용함
- 공간빅데이터 체계구축사업의 단계별 기술 수요에 따라 공간 빅데이터 기반 핵심

기술을 지원하고 피드백을 받음

- 실제 성과물의 적용은 사업자들에 대한 기술이전을 통하여 사업에 적용하도록 함
- 기술 검증 및 실용화 추진을 위하여 공간빅데이터 체계구축사업과의 협의에 의하여 시제품을 단계별로 적용하고 성능검증을 위한 테스트베드를 구축하여 활용하고 보유 또는 연계 가능한 데이터를 최대한 활용함

○ 연구단과 사업단의 긴밀한 협의를 통한 성과물 적용을 추진

- 공간빅데이터 사업단과 연구단간의 정기적인 협의를 통해 중복방지, 성과공유, R&D 성과 적용 등을 논의하기 위해 공간 빅데이터 사용자 협의체를 구성하고 주기적으로 회의를 개최함
- 국토교통부, KAIA, 사업단, 추진단, 연구단 등이 참여하는 사용자협의체에서 R&D에서 필요한 기능을 도출하고 성과물의 적용 방안을 정하며 각종 현안을 논의함
- 공간 빅데이터 플랫폼의 연구성과 테스트 및 검증에 사업단에서 구축 보유한 데이터 및 융합DB를 적극 활용함
- 사업단과 협의 하에 사업단 시스템에 연구 성과물을 설치하고 사업단과 합동으로 성과물 실증 시험을 수행함

② 성과물 연동 및 실증 시험 경과

○ 공간 빅데이터 협의체 개최

- 매년 공간빅데이터 체계구축사업의 사업자, 추진단, 국토교통부, KAIA와 연구단이 참석하는 공간 빅데이터 협의체 회의를 개최함
- R&D 성과물의 사업단 플랫폼 적용 방안, 적용 대상 기술, 사업단에서 요구하는 기능 등에 대한 논의를 진행함
- 논의 중 인지된 사업단의 기술적 애로사항을 해결하기 위해, 연구내용 중 일부를 변경하여 해당 문제를 해결할 수 있는 기술 개발을 수행함
  - 예를 들어 고속 가시화 지원 기술, 분석 플로우 처리 기술 등이 있음

○ 테스트베드 구축

- 연구단 차원의 테스트베드를 연구단(안양대학교)에 구축하여 하둡 클러스터를 구성하고 데이터를 적재하였으며, 1세부의 연구성과물을 설치하고 성과물 연계를 위한 환경을 설정함







<그림 4-1> 성과물 테스트 및 통합활용을 위한 테스트베드

❍ 적용 대상 성과물 분석

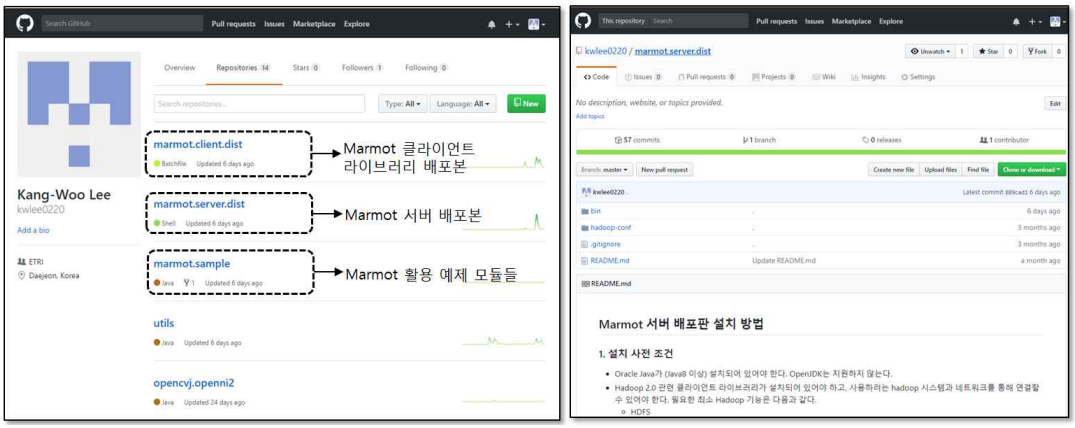
- 사업단 플랫폼에서 서비스되고 있는 34종의 공간분석 라이브러리와 15종 기초공간분석 컴포넌트를 분석
- 공간분석 라이브러리로부터 요구사항 및 적용 우선순위를 정리하고 성과물에서 연산을 지원할 수 있도록 개발을 진행함



<그림 4-2> 사업단 공간분석 라이브러리

○ 타 세부간 연계 활용 및 기술 지원

- 1세부과제의 성과물은 타 세부과제의 연구개발에 직접적으로 연계되어 사용될 수 있으므로 연구기간 동안 최종적으로 완성되기 이전이라도 타 세부 과제에 제공하여 사용 및 활용할 수 있도록 함
- 저장 시스템을 활용하기 위한 인터페이스와 관련된 연구 성과물을 2세부 및 3세부에 제공하여 이를 활용한 분석 응용 및 서비스를 개발할 수 있도록 지원함
- 개발된 성과물을 타 세부와 통합하기 위해 플랫폼 S/W를 2017년도부터 배포하고 활용 기술을 지원함
  - 공간 빅데이터 저장관리 플랫폼 서버(Binary), 클라이언트(Binary), 샘플(소스코드)를 GitHub에 배포하여 공유함[9)]
  - 공간 빅데이터 저장관리 시스템용으로 개발된 JDBC 드라이버를 2세부과제에 전달하여 JDBC 인터페이스에 의하여 ‘OLAP 인터랙티브 분석 시스템’과 연계하여 공간 빅데이터 저장관리 시스템의 기능과 데이터를 질의언어를 통하여 사용할 수 있게 지원함



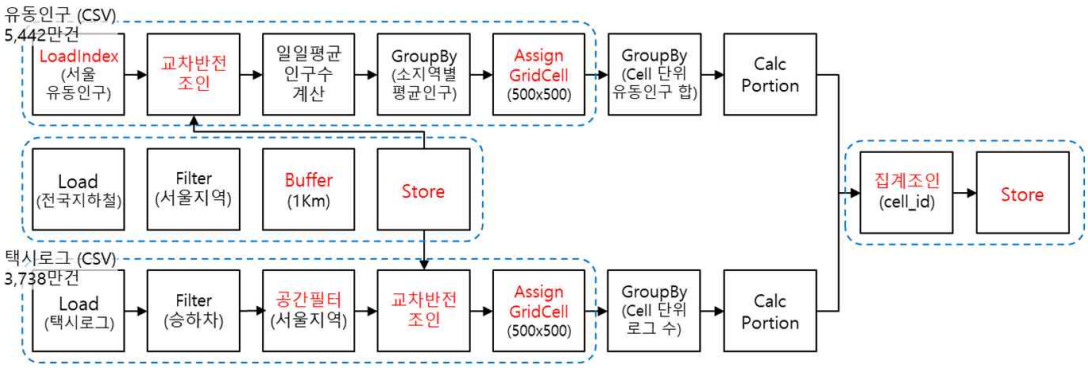
<그림 4-3> 성과물 배포 및 기술지원

○ 성과물을 활용하여 사업단 플랫폼에서 분석 수행

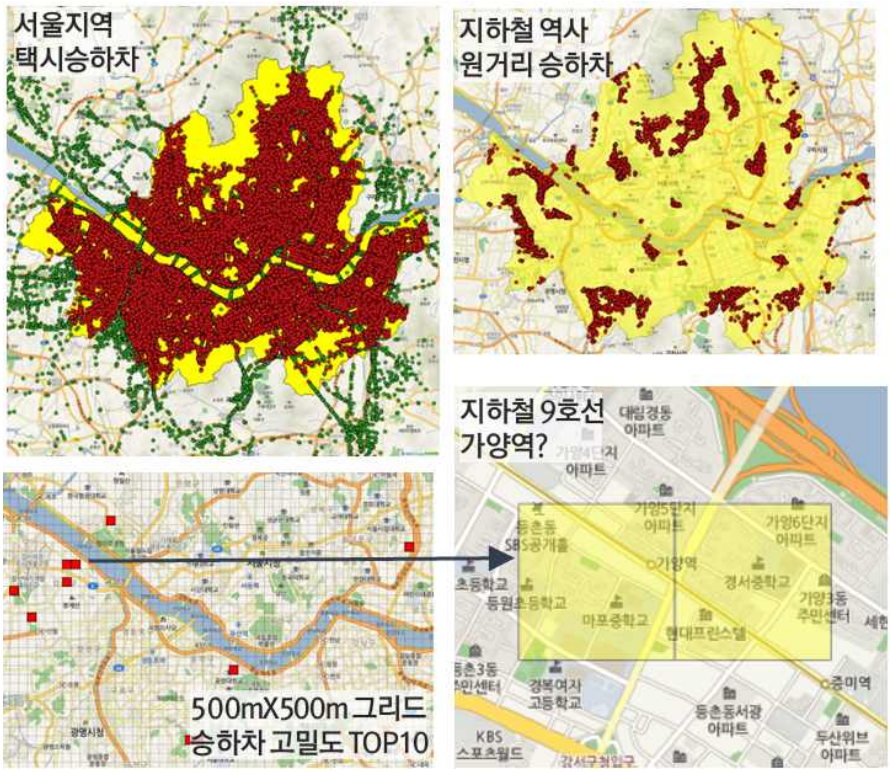
- 2017년에 ‘택시 승하차 로그/유동인구 분석을 통한 지하철역 추천’, ‘대도시 상업지구 내 지역활성화 정도 분석’과 ‘도시쇠퇴 현황과 도시재생사업 현황분석’ 사례를 통하여 성과물을 사업단 플랫폼에서 수행하고 사업단 플랫폼과의 연동 가능성을 확인함
  - 택시 승하차 로그/유동인구 분석을 통한 지하철역 분석은 대규모 공간 빅데이터를 활용하는 공간 분석이며, 개발한 주요 공간연산을 사용하여 분석 플로우를 구성함
  - 택시 로그(3,738만건, 3.7GB), 전국 유동인구(5,442만건, 8.39GB), 전국 지하철역사, 시도 행정구역 데이터를 사용함
  - 분석 시간은 약 1분 40초가 소요됨

9) GitHub 주소는 https://github.com/kwlee0220?tab=repositories

• 분석 프로세스를 공간연산의 플로우로 구성함(그림 4-4)



<그림 4-4> 택시 승하차 로그 및 유동인구 분석을 통한 지하철역 추천 분석의 연산 플로우



<그림 4-5> 택시 승하차 로그 및 유동인구 분석을 통한 지하철역 추천 분석

• 대도시 상업지역 내 건축물의 세부용도 분포를 파악하고 지역 경제 활성화 정도를 파악하는 분석

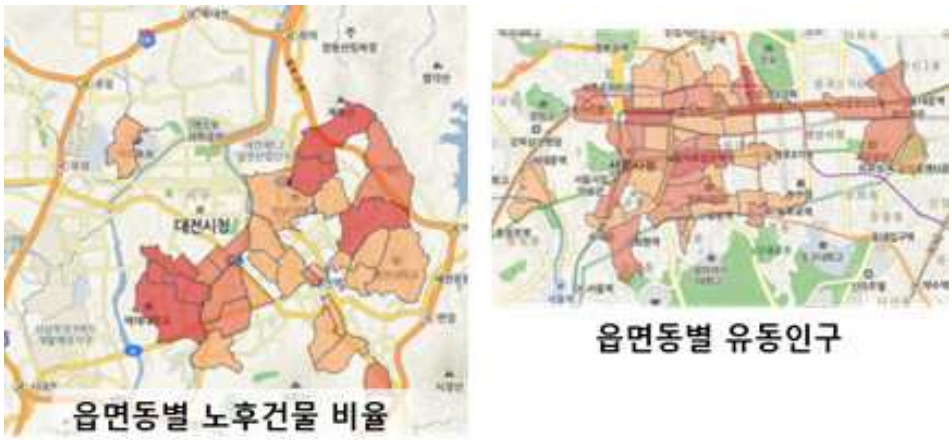
• 전국 유동인구 (5,442만건, 8.39GB), 전국 카드매출 (1억1천만건, 49.3GB), 건물통합정보(1,402만건, 18.8GB), 용도지역지구(6.4만건, 320MB) 데이터를 사용함

• 분석 시간은 11분 37초가 소요됨



<그림 4-6> 대도시 상업지구내 지역활성화 정도 분석

- 노후건축물 분포와 도시쇠퇴 현황을 비교하여 취약지역 상황을 파악하는 분석
- 전국 유동인구(5,442만건, 8.39GB), 전국 카드매출(1억1천만건, 49.3GB), 건물통합정보(1,402만건, 18.8GB) 데이터를 사용함[10)]
- 분석 시간은 22분 29초가 소요됨



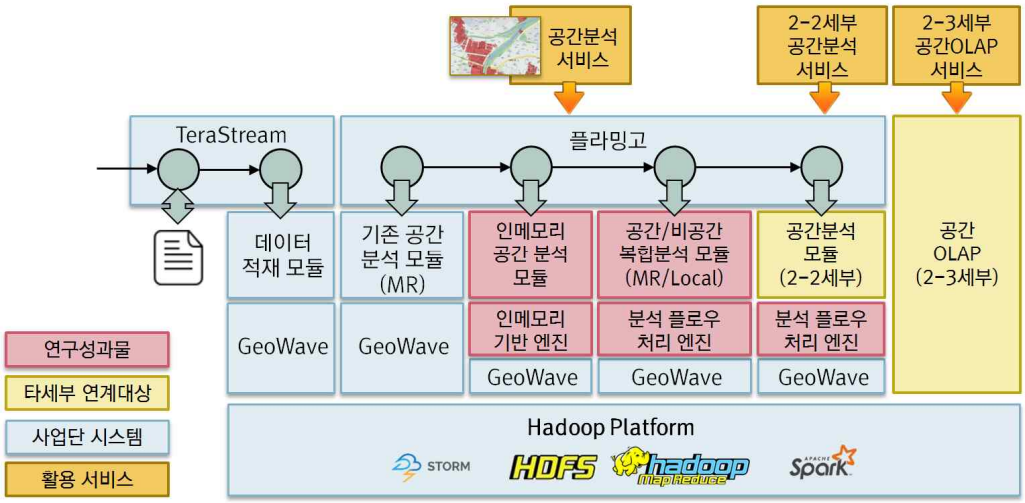
<그림 4-7> 도시쇠퇴현황과 도시재생사업현황 분석

③ 실증 시험 계획

❍ 공간빅데이터 사업단 플랫폼상의 실증 시험 방법

- 사업단 시스템과 연동하여 연구 성과물 실증
- 적용에 따른 기존 사업단 시스템의 변경 최소화
- 기존 GeoWave의 저장 포맷 및 색인 활용
- 기존 분석 워크플로우 엔진(플라밍고)과 연동
- 기존 운용 중이던 분석 서비스 구동 가능
- 실제 실증은 일부 분석 서비스를 통해 확인

10) 실제로는 분석에 필요한 데이터 부족으로 인하여 부분적인 과정만 수행함



<그림 4-8> 공간빅데이터 사업단 플랫폼 적용 개요

○ 공간빅데이터 사업단 플랫폼에 대한 세부 활용 계획

- 공간빅데이터 사업단 플랫폼에 대한 성과물 및 기능별 세부 활용 계획은 <표 4-1>과 같음

<표 4-1> 핵심연구성과의 활용계획

과제명(연구기관)		핵심성과	주요 활용계획	기대효과
1세부	공간 빅데이터 저장관리 인프라 기술 개발 (한국전자통신연구원)	(A) Hadoop 기반 공간 빅데이터 저장 프레임워크	(저장관리 모듈) • 일부 기능 (HDFS에 저장된 공간 빅데이터의 읽기/쓰기 기능)은 GeoWave와 연동하여 활용할 예정 • 그 외 공간데이터 활용 기술 및 Map/Reduce 작업 구동 기술을 적용함 (연산자) • 공간 분석 플로우 처리 기술을 적용할 수 있는 복합 공간/비공간 연산자 제공 • 속성 기반 라이브러리 위주로 적용하여, 현 사업단 플랫폼 구조 변경을 최소화하고, 1-2세세부 연동시 발생 가능한 문제를 최소화함	
		(B) 공간 빅데이터 질의 처리 엔진	• 2-3세세부에서 적용할 OLAP 엔진에서 SQL 기반 질의언어를 사용하여 연동하는 형태로 적용 • 현 공간 빅데이터 플랫폼에는 SQL을 활용하여 공간 분석을 수행하는 분석 서비스가 없어 사업단에 직접 적용할 수 없음 • 사업단 플랫폼과 직접 연동이 아닌 타 세부(2세부)와 연동하는 형태로 적용	
	인메모리 기반 공간 빅데이터 저장관리 프레임워크 개발 (포스웨이브)	(O) 인메모리 기반 공간빅데이터 저장관리 프레임워크	• 기존 개발된 인메모리 기반 공간 분석 라이브러리 23종은 국토부 공간 빅데이터 사업단 과 연계 실증을 통해 사업단 플랫폼 활용방안 검토중 • 인메모리 기반 공간색인 기술 및 공간 분석 라이브러리들은 자체 기술이전을 통해 자체 솔루션 개발 진행 예정	

	공간 빅데이터 웨어하우스 핵심기술 개발 (아이씨티웨이)	(D) 스트림 기반 분산 공간 ETL 도구	• 사업단 플랫폼의 기존 ETL 도구를 대체하지 않고 병행 구동이 가능한 형태로 적용 • 사업단 플랫폼 내에서 스트림 형태의 공간정보 처리 서비스가 제공되지 않아 직접 적용이 어려움 • 공간 연산 워크플로우 분석도구를 활용하여 공간 ETL 병행 처리하여 충돌을 최소화해야 함	• 추후 기존 플랫폼 내 공간 ETL 기능 수정 없이 실증이 가능할 것으로 예상됨

④ 공간빅데이터 사업단 플랫폼 적용 실증 시험

○ 공간빅데이터 사업단 시스템에 공간 빅데이터 저장관리 시스템의 성과물을 설치하고 시범 운용을 통해 실증

- 사업단에 방문하여 성과물을 설치하고 성과물 기능을 테스트 실증
  - 장소 : 공간빅데이터사업단 사무실(서울 방배동)
- 실증 대상 기능을 11개의 TC(테스트케이스)로 정의하고 실증 확인(2019.11.05.~06)
  - 1세부 성과물 기능: 공간분석 플로우처리 기능, 공간데이터 import/export 기능
  - 2세부 연동 기능: 고급 공간분석 컴포넌트(2-2세세부), OLAP 분석 시스템(2-3세세부)
- 실증 대상 기능에 대한 4개의 TC를 추가적으로 실증(2019.11.19.)
  - 공간 빅데이터 가시화 지원 기능, 공간 빅데이터 스크립트 언어, 공간 빅데이터 질의처리 기능
  - 공간 ETL 관리도구

○ 공간 빅데이터 저장관리 시스템 성과물 실증 과정은 다음과 같은 단계로 수행함

- 테스트케이스 수립
- 사업단 시스템에 프로그램 설치 및 데이터 준비
- 프로그램 실행 및 기능 확인

○ 실증 시험에 필요한 공간 빅데이터 저장관리 시스템의 시험 항목(테스트케이스)과 각 항목에 따른 시험 절차를 수립하고 이에 따라 실증 시험을 수행하였으며, 그 결과를 사업단과 함께 확인함

(다) 기대효과

<공공적 측면>

○ 공간 빅데이터를 활용한 공공서비스 확산 및 수준 제고

- 국토교통부 재정사업 ‘공간빅데이터 체계구축사업’에서의 기술적 장벽을 해결할 수 있는 핵심기술과 성과물 지원으로 사업의 성공적인 결과에 기여할 것임

- 공간 빅데이터 사업단 플랫폼의 기능 및 성능 고도화 등에 효과적으로 적용함으로써 공간 빅데이터를 활용하는 국내의 유사한 공간정보 서비스의 확산 및 활성화가 기대됨
- 공간 빅데이터의 생산과 활용이 사회 전반에 걸쳐 빠르게 증가하고 있는 것을 고려할 때 공간 빅데이터 처리 핵심기반기술을 제공함으로써 공공/민간의 공간 빅데이터의 활용성 증대 및 대민 서비스 확대 기반을 마련함

❍ 공간 빅데이터 대민서비스 확대

- 현재 서비스중인 공간정보 오픈플랫폼(V-World)에 데이터 저장관리 플랫폼으로서 적용할 경우 서비스 수준 향상과 대상 데이터의 확대에 기여할 것임
- 저비용 고품질의 공간 빅데이터 개방, 공유, 활용 생태계 구축의 기반을 제공함으로써 공간정보 기반 대국민 편의 서비스의 질을 향상시킴

〈기술적 측면〉

❍ 대규모 공간 빅데이터에 대한 효율적인 저장·관리·검색 기능과 분석을 제공할 수 있는 공간 빅데이터 원천 기술 확보를 기대함

❍ 공간 빅데이터를 다루는 기반이 되는 데이터 저장관리 기술을 하둡 환경에서 공간 빅데이터 처리가 필요한 시스템에 폭넓게 활용할 수 있음

- 공간 빅데이터 연산자를 통하여 공간 빅데이터 처리 기능을 제공함으로써 유연한 확장성을 제공하며 공간분석 응용프로그램에서 활용성이 높음
- 상황에 따른 동적 공간 연산자 수행 방법을 선택하도록 프로그램을 개발하여 공간 빅데이터 처리 성능을 향상시킬 수 있음
- 공간 빅데이터를 SQL 기반 질의언어에 의하여 질의할 수 있는 기능을 제공하여 응용프로그램에서의 사용 편의성을 제고하고 기존 legacy 시스템과 통합을 용이하게 함

❍ 최근 빅데이터 분야에서 주목받고 있는 인메모리 기술을 접목하여 인메모리 기반 고속 공간 빅데이터 처리 기능을 제공하여 기존 하둡기반 시스템 대비 공간 분석의 성능 향상을 가져옴

- 지속적 또는 고속으로 증가하는 이동객체 빅데이터 자료에 대한 효과적인 저장 방안 제공
- 최근 IOT 및 Mobile 기술 발달에 따라 수요가 증가하고 있는 이동객체 빅데이터 실시간 처리를 위한 저장관리 및 분석 플랫폼의 핵심기술로 활용될 수 있음

❍ 공간 스트림 빅데이터를 실시간으로 처리하며 데이터 특성에 의존적이지 않으며 규칙처리가 가능한 ETL 처리 엔진 제공

- 개발자 수준에서 ETL 툴킷을 이용한 유연한 ETL API 제공
- 스트림 기반 공간 빅데이터 분석을 위한 전처리 도구로써 활용

〈경제적 · 산업적 측면〉

○ 미래 산업의 기술경쟁력 확보

- 최신 IT 인프라 기반 공간 빅데이터 플랫폼 구축으로 미래 지식기반 사회를 지향하는 선진국 대비 기술 경쟁 우위를 확보하고 공간정보 산업의 기술경쟁력 확보 기대
- 기존 유무선 통신, IoT, 스마트폰 등 세계적인 경쟁력을 보유한 IT 인프라와의 공간 빅데이터 처리기술 융합을 통한 독자적인 공간정보 신시장 창출 및 경쟁력 확보 기대
- 지도 등 대용량 공간 빅데이터와 스트림 형태의 빅데이터를 융합한 질의를 가능하게 하여 빅데이터로부터 새로운 가치 창출

○ 국내 공간정보 기술 확대 및 공간정보 시장 개편

- 국산 공간 ETL 핵심기술 확보를 통한 국내 공간정보 시장의 SW 의존도 감소를 기대함(‘12년 75% 이상 -> ’ 25년 50% 이하)
- 전자지도 등 대용량 공간 빅데이터와 시간에 따라 빠르게 수집되는 실시간 공간데이터를 융합한 관리 및 분석이 가능하여 빅데이터로부터 새로운 사회경제적 가치 창출
- 국내 공간정보 기술 확보로 외산 소프트웨어 중심의 기본 공간정보 시장을 개편하고 해외시장 진출을 기대함
- 공간정보 기반의 실시간 빅데이터 저장관리 기술 확보를 통하여 IoT 등과 연계한 빅데이터 활용 및 서비스 다양화
- 실시간 빅데이터의 효과적인 처리와 관리를 지원하기 위한 핵심기술로서 빅데이터 기술 선도 및 관련 산업 활성화가 기대됨

〈사회적 측면〉

○ 공간 빅데이터와 공공정보를 시민들이 활용할 수 있게 될 것으로 예상

- ‘공간 빅데이터 플랫폼’을 활용하여 국가가 소유하고 있는 다양한 공간 빅데이터를 시민들이 적극 활용할 수 있는 통로를 제공하여 다양한 정보 선택권을 제공함
- 공간 빅데이터의 공개 및 활용을 용이하게 하여 유도하여 시민들의 정부활동에 대한 참여, 모니터링을 통해 국정 운영의 투명성을 제고하고 공공 정보에 대한 국민의 알권리를 확대시킴

(2) 공간 빅데이터 분석 및 가시화 기술 개발

(가) 활용 방안

○ 공간 빅데이터 소셜 분석 및 가시화 솔루션은 다음와 같은 공곤 기관에 활용 될 수 있음

- 본 연구 성과물의 예상되는 활용주체로는 국토교통부, 국토지리정보원, 한국국토정보공사, 국가공간정보통합센터, 공간정보산업진흥원, 국립재난안전연구원 등 비정형데이터에 대한 분석을 필요로하는 공공기관이 있음
- 또한 딥러닝 모델로 학습된 모형 및 이를 생산할 수 있는 기술이 개발되어 있으므로 민간 및 국민 등 비정형데이터를 이용하여 딥러닝 모델을 적용할 수 있을 것이라 예상됨

○ 공간 빅데이터 및 이동객체데이터 Batch 분석 솔루션은 이를 필요로 하는 공공기관 및 지방자치단체의 분석 툴박스로써 사용될 수 있음

- 공공기관 및 지방자치단체에서 보유하고 있는 공공데이터는 그 활용가치가 높지만 통일된 분석 솔루션이 마땅치 않은 상황임
- 공간 빅데이터 및 이동객체데이터 Batch 분석 솔루션은 공공기관 및 지방자치단체에서 보유한 대용량의 데이터를 시간과 장소에 구애받지 않고 분석할 수 있다는 장점을 지니며, 각 기관 및 지방자치단체가 필요로 하는 수요기반 유동인구 분석, 다차원 빅데이터 패턴 분석 등의 분석 시나리오를 설정하고 그 분석 결과를 의사 결정의 근거 자료로 사용 가능함
- 본 연구 성과물인 공간 빅데이터 및 이동객체데이터 Batch 분석 솔루션을 공공/연구기관인 건축도시공간연구소(AURI), 행정중심복합도시건설청의 연구에 현장 적용하였으며, 스마트시티 (IFEZ) 및 민간기관인 SK지오비전에 현장 적용하여 공공 및 민간의 빅데이터 분석 니즈에 맞추어 실제 다양한 기관에서의 본 연구 성과물이 활용될 수 있음을 확인함
- 공간 빅데이터 및 이동객체 데이터를 생성하는 스마트 시티 운영 지자체 (서울스마트시티센터, IFEZ, 고양 스마트시티 등), 공공 연구기관 (서울기술연구원, 행정중심복합도시건설청, 공간정보산업진흥원 등) 및 인구 유동 데이터를 생산하는 기업 (SKT, KT 등)이 본 연구 성과물의 주요 활용 주체로 예상됨

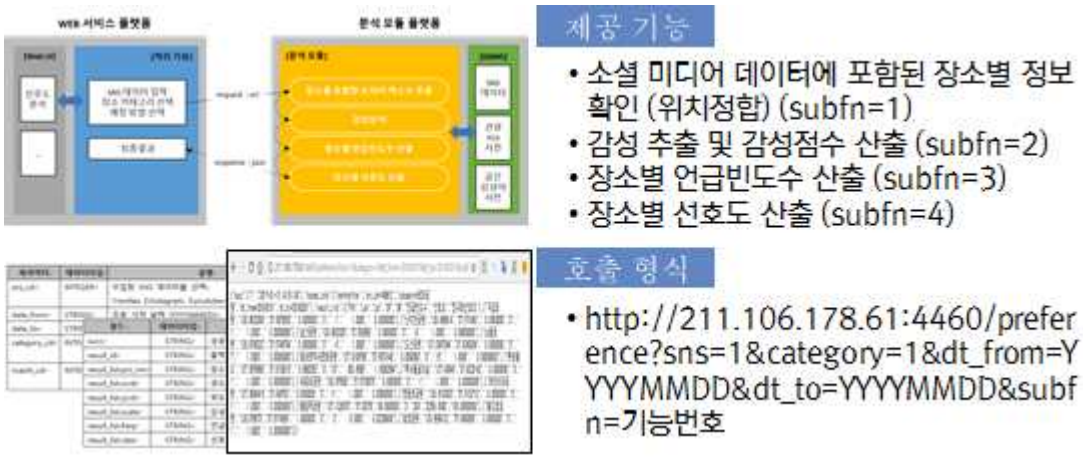
○ 공간 빅데이터 Interactive 분석 솔루션은 실시간 다차원 분석 솔루션으로 활용 가능함

- 공간 빅데이터에 특화된 실시간 다차원 분석이 가능한 솔루션은 널리 제공되고 있지 않음
- 향후 자율주행차 등의 대중화로 인해 실시간 공간 데이터가 고속 고용량으로 유입되는 환경에서, 사용자가 원하는 정보를 실시간으로 다차원 분석 가능함

- 특히 관련 전문 지식이 높지 않은 일반 사용자도 데이터 정보, 질의 정보 등의 최소한의 정보로 UI/UX를 활용하여 손쉽게 분석 작업을 수행할 수 있음

(나) 공간빅데이터체계구축사업 적용

○ 공간 빅데이터 정합 기술, 공간 빅데이터 분포 및 차트 분석 기술(딥러닝 기반 문장분류 기술, 공공시설의 접근성 분석)을 적용할 예정임



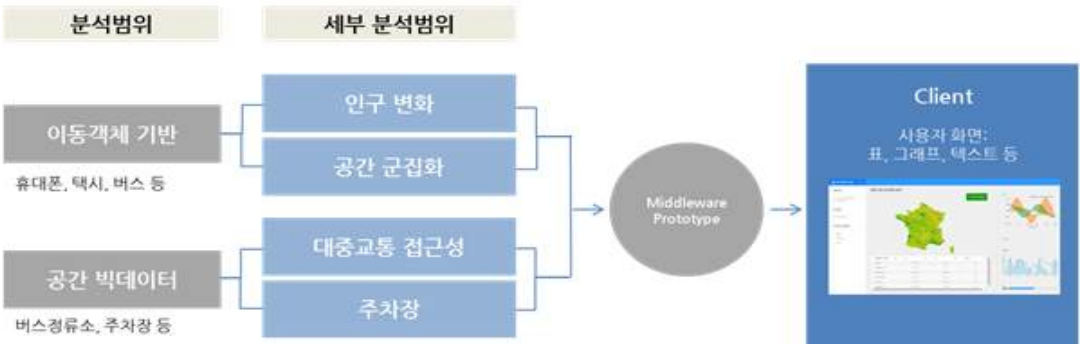
<그림 4-9> API를 활용한 개발 컴포넌트 제공 예시

<표 4-2> 공간 빅데이터 소셜 분석 및 가시화 컴포넌트 적용을 위한 필요사양

구분	품목(사양)	용도	사유
적용 시 하드웨어 사양	-CPU: 2.5 GHz -MEM:16GB이상 -HDD:100GB이상 -OS:CentOSrelease6.10(Final)64bit -그외:기본	분석모듈 서버	비정형 데이터 분석모듈 적용
적용 시 운영환경	자바 버전 1.7 이상	텍스트마이닝을위한라이브러리사용시필요	사업단의 자바 버전을 패치
	R 최신버전	통계도구와 연동하여 상관분석 등 고급통계 분석 활용	R설치후필요한패키지사전에설치 (DBI,e1071,KoNLP,maptools,rgdal,rgeos,spdep,stringi,tm,...)
	rstudio 최신버전	R 기반의 분석모듈 설치 및 디버깅 도구	
	python 2.x 최신버전	python 2.x용 분석모듈 설치 및 디버깅	python 설치후 패키지 사전에 설치
	python 3.x 최신버전	python 3.x용 분석모듈 설치 및 디버깅	python설치후패키지사전에설치 (numpy,scipy,scikit-learn,skrvm,pandas,konlpy,psutil,gensim,...)
	node.js 최신버전	분석모듈 결과의 openapi 제공	

○ 공간빅데이터체계구축사업에서 공간 분석 모듈을 구성함

- 1세부에서 저장 / 관리하는 공간 빅데이터를 다양한 분석 컴포넌트로 이루어진 Batch 분석 솔루션을 통해 분석 서비스를 제공할 수 있음
- 이미 2차년도부터 아래 그림과 같은 자체 분석 솔루션 사용자 인터페이스를 개발하여 제공하고 있음

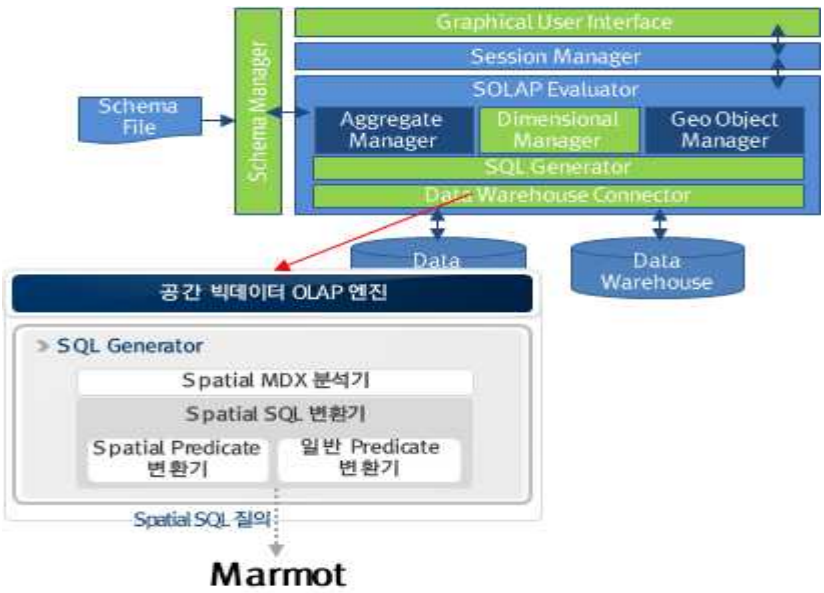


<그림 4-10> Batch 분석 솔루션 2차년도 초기 모델

○ 또한, 공간 빅데이터 및 이동객체데이터 분석 컴포넌트 30개에 대한 기술 지원이 가능함

○ 공간빅데이터체계구축사업에서 공간 빅데이터 Interactive 분석 컴포넌트를 시범 적용함

- 1세부에의 Marmot과 연계하여 공간 빅데이터를 저장하고, 이를 OLAP 모듈과 연동하여 공간 빅데이터 다차원 분석을 효율적으로 수행함



<그림 4-11> OLAP 엔진과의 Marmot 연계

(다) 기대효과

○ 공간 빅데이터 및 이동객체데이터 분석 전문 인력 양성

- 공간 빅데이터 및 이동객체데이터 Batch 분석 솔루션을 탑재하고 활용할 시 이를 전문적으로 다룰 수 있는 전문 인력을 양성할 수 있음
- 인력 양성으로 인한 향후 추가 결과 적용 및 분석 솔루션 탑재를 기대할 수 있음

○ 실시간 공간 빅데이터 분석을 통한 의사 결정 지원

- 실시간 공간 빅데이터의 Interactive 분석을 통해 공간 데이터를 활용한 정책 결정이나 공공 시스템 지원에 있어서 빠른 의사결정이 필요한 경우 실시간으로 관련 근거 및 판단 지표등을 도출할 수 있음

○ 외산 시스템 대체로 인한 비용 절감

- 공간 빅데이터 관련 솔루션이 많지 않은 상황에서, 외산 시스템을 대체하는 효과로 라이센스 비용 등 관련 예산을 절감할 수 있음

○ 공공적 성과

- 민원과 같은 비정형 데이터 분석을 통하여 정책 결정에 대한 기본자료를 생성함으로써 데이터 기반의 과학적이고 합리적인 의사결정 지원 가능

○ 기술적 성과

- 장소에 특화된 사전을 구축하기 위한 기술을 제시하고 이를 이용하여 장소 정보가 포함된 비정형 데이터로부터 장소에 대한 선호도를 분석하는 기술을 개발함
- 최신의 딥러닝 기술을 개발하여 제공함으로써 사용자들이 공간 빅데이터 분석 및 가시화를 위하여 쉽게 활용할 수 있도록 함

○ 경제적 성과

- 사용자 개입이 적은 반자동화된 기술의 개발로 데이터 분석 비용 절감 및 유지비용 최소화
- 딥러닝 기술이 적용된 훈련모델 생성을 위하여 어노테이션 데이터 생성을 위한 고용창출

<표 4-3> 핵심연구성과의 활용계획

과제명(연구기관)		핵심성과물(level1)	주요 활용계획	기대효과
2세부	(E) 공간 빅데이터 소셜 분석 기술 개발 (서울대/시터스)	공간 빅데이터 분석을 위한 모델 설계 및 목표 성능 설정	• 공간 빅데이터 분석을 위한 모델 및 목표 성능을 설정함으로 공간 빅데이터 분석 기술 개발 시 기초자료로 활용 가능	• 공간 빅데이터 분석을 위한 기반 마련
		공간 빅데이터 분석을 위한 데이터 가공, 변환, 필터링 및 분석효율 고도화	• 공간에 특화된 분석 및 대용량의 지도 정보를 처리하기 위한 기술 개발 시 활용 가능	• 공간에 특화된 빅데이터 분석을 위한 기반 마련
		공간 빅데이터 정합 및 주제도 분석 기술	• 정합기술은 비정형데이터에서 장소정보 추출 및 탐지 시 활용 가능하며, 주제도 분석 기술은 장소선호도기반, 공간분석기반,	• 비정형데이터에서 위치를 탐지하고 이를 통해 주제도로 시각화하는 기반 마련

과제명(연구기관)		핵심성과물(level1)	주요 활용계획	기대효과
			SNS기반 주제도 작성 시 활용 가능	
		공간 빅데이터 분포 및 차트 분석 기술	• 소셜미디어 데이터를 이용한알고리즘(7개)을 통해 분포 및 차트 서비스(14개 서비스)에 활용 가능	• 소셜미디어 데이터를 공간정보와 결합하여 이를 분포 및 차트화로 서비스하는 도구 마련
		공간 빅데이터 분석결과 가시화 기술	• 체크인데이터, 트위터데이터, Yelp데이터, 국내 장소리뷰데이터 등 다양한 데이터에 대한 정적 및 동적가시화 서비스 가능	• 공간빅데이터의 분석 기술을 정적 측면 및 동적 측면에서 가시화 서비스하는 도구 마련
		공간 빅데이터 분석을 위한 모델 설계 및 목표 성능 설정	• 공간 빅데이터 분석을 위한 모델 및 목표 성능을 설정함으로 공간 빅데이터 분석 기술 개발 시 기초자료로 활용 가능	• 공간 빅데이터 분석을 위한 기반 마련
		공간 빅데이터 분석을 위한 데이터 가공, 변환, 필터링 및 분석효율 고도화	• 공간에 특화된 분석 및 대용량의 지도 정보를 처리하기 위한 기술 개발 시 활용 가능	• 소셜 미디어 데이터 사용의 표준화 및 대용량 데이터 분석을 위한 기술 기반 마련
		공간 빅데이터 정합 및 주제도 분석 기술	• 공간 정합 및 주제도 분석의 추가 연구를 통해 기술이전 가능	• 공공기관에서 정보 모니터링 및 예측을 위한 분석 활용 기대
		공간 빅데이터 분포 및 차트 분석 기술	• 공간 빅데이터 사업단에 제공하여 국가 공간정보시스템에 사용할 수 있도록 추진	• 향후 실시간 빅데이터의 효과적인 처리와 관리를 지원하기 위한 핵심기술로서 빅데이터 기술 선도 및 관련 산업 활성화가 기대됨
		공간 빅데이터 분석결과 가시화 기술	• 공간 빅데이터 사업단에 제공하여 국가 공간정보시스템에 사용할 수 있도록 추진	• 향후 실시간 빅데이터의 효과적인 처리와 관리를 지원하기 위한 핵심기술로서 빅데이터 기술 선도 및 관련 산업 활성화가 기대됨
	(F) 공간 빅데이터 일괄(Batch) 분석 컴포넌트 기술 개발 (연세대/가이온)	공간빅테이터 Batch 분석 컴포넌트	• 공간 빅데이터 분석기술 및 결과를 필요로 하는 공공기관 및 지방자치단체에 탑재하고자 함	• 공간 빅데이터 분석 전문 인력 양성 • 관련 공공기관 및 산업 분야 추가 적용 가능
	(G) 이동객체데이터 Batch 분석 컴포넌트 (연세대/가이온)	이동객체데이터 Batch 분석 컴포넌트	• 이동객체데이터 분석기술 및 결과를 필요로 하는 공공기관 및 지방자치단체에 탑재하고자 함 • 승객 탑승 지점 추천 기능의 모바일 어플리케이션 시스템을 ㈜나비콜 시스템과 연계하고자 함	• 이동객체데이터 분석 전문 인력 양성 • 관련 공공기관 및 산업 분야 추가 적용 가능
	(H) 공간 빅데이터 CEP 및 OLAP 엔진 개발 (한국과학기술원/유엔진솔루션즈)	공간 빅데이터 Interactive 분석 컴포넌트	• 공간 빅데이터 사업단에 제공하여 국가 공간정보시스템에 사용할 수 있도록 추진	• 비전문가의 공간 빅데이터 Interactive 분석 활용도 증대 • 외산 소프트웨어 대체로 라이센스 비용 절감

(3) 공간 빅데이터 서비스 활성화 및 실증 연구

(가) 활용 방안

① 공간 빅데이터 전문인력 양성 및 지속적 홍보

○ 중앙부처, 지자체, 공공기관, 연구기관 등 업무대상에 맞는 적정한 활용모델과 가이드라인 개발·보급 및 교육 컨설팅 지원 필요

- 활용 모델 발굴 및 홍보, 동영상 교육콘텐츠 제공, 오프라인 교육 등 다차원적인 홍보와 교육계획 수립 추진
- 컨설팅, 교육, 커스터마이징, 관련 업무 연계, 제도 정책 등 지원체계 마련

<표 4-4> 공간 빅데이터 전문인력 양성 및 지속적 홍보

공간 빅데이터 전문인력 양성	교육과정 운영	교육프로그램 개발
- 공간 빅데이터 체계 운영·관리 전문가 양성 - 공간 빅데이터 체계 분석 전문가 양성	- 교육과정 수강생 모집 - 전문인력 양성계획 수립 및 교육실시	- 교육 콘텐츠 발굴 - 교육프로그램 기획 및 제작 - 교육프로그램 점검 및 보완

○ 사회현안 해결 및 수요창출 혁신 비즈니스 발굴을 위한 기업, 기관(지자체) 등과의 상호 협력 방안 마련

- 사회 현안 해결에 기여할 수 있는 사회 연결망, 공간분석, 인공지능 기술을 접목한 정책지원용 분석방법론 및 활용 모델 등 성공사례 발굴·전파
- 건물, 토지, 시설 등 국토·도시 분야뿐만 아니라 보건, 의료, 기후변화 등 사회적 중요도와 파급효과를 고려한 기업 및 기관 등과 상호 협력 · 교류 방안 마련

○ 공간 빅데이터 활성화를 위해 중앙정부-지자체-기업 간의 협력적 거버넌스체계를 구축하여 운용하도록 지원

- 중앙-지방-기업이 함께하여 정책 및 기타 문제 해결을 위한 다양한 이해관계자들을 합의 지향적으로 참여시킬 수 있도록 협력체계 구축 지원
- 공간 빅데이터 구축을 위해 중앙에서 장기전략을 수립하고, 이에 따라 공공부문에서의 데이터를 지방자치단체와의 협력을 통해 구축, 일반기업 및 시민과의 협력을 통해 민간부문 데이터 구축 · 활용

○ 공공기관 뿐만이 아닌 민간부문을 대상으로 한 지속적인 홍보 방안 수립을 통하여 공간 빅데이터 체계의 다차원적 서비스 활성화 도모

<표 4-5> 민간부문을 대상으로 한 지속적인 홍보 방안 수립

홍보전략 수립	세미나/공모전 개최	언론 홍보
- 홍보·활성화 대상자 정의 - 대상별 홍보매체 선정 - 대상별 맞춤형 홍보계획 수립	- 업무담당자 세미나 개최 - 서비스 개발 공모전 개최	- 보도자료 작성 및 배포 - 온라인 매체 활용

○ 단기적으로는 플랫폼 활용을 위한 컨설팅 조직 운영 또는 지자체, 공공기관, 연구기관의 전문 인력과 지속적인 교류 활성화 및 협력 네트워크 마련 필요

- 데이터 기반 정책업무 제도화로 지역개발계획 수립, 스마트도시계획 수립 및 사업효과 평가 등 관련 정책업무를 발굴하고 데이터 기반 업무모델 제시
- 정책실무자를 위한 데이터 서비스, 분석 및 시각화 시스템 등 수요자 맞춤형 정보, 컨설팅 서비스 제공
- 정책업무에 맞춤형으로 대응할 수 있는 분석컨설팅, 도메인 전문가의 해석지원, 교육 및 훈련 프로그램 등을 함께 제공하여 협업적 데이터 활용 지원
- 공간 빅데이터 관련 기업 및 지자체 등과 전략적 교육 협력체계 마련

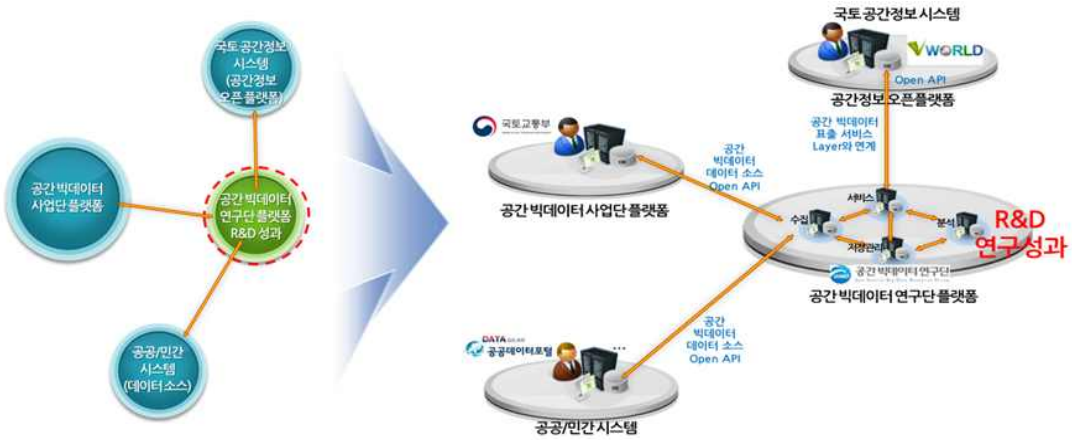
<표 4-6> 공간 빅데이터 관련 기업 및 지자체 등과 전략적 교육 협력체계

- 공간 빅데이터 생산·가공 및 분석 등에 대한 상호 기술 협력 - 공간 빅데이터 최신 기술, 트랜드 정보공유 등을 통해 협력체계 구축 - 공간 빅데이터 인재 양성을 위한 지원활동(강의, 현장학습, 아카데미 등) - 기타 우리나라 빅데이터 정책과 기술 발전을 위한 상호 협력적 공조체계 구축

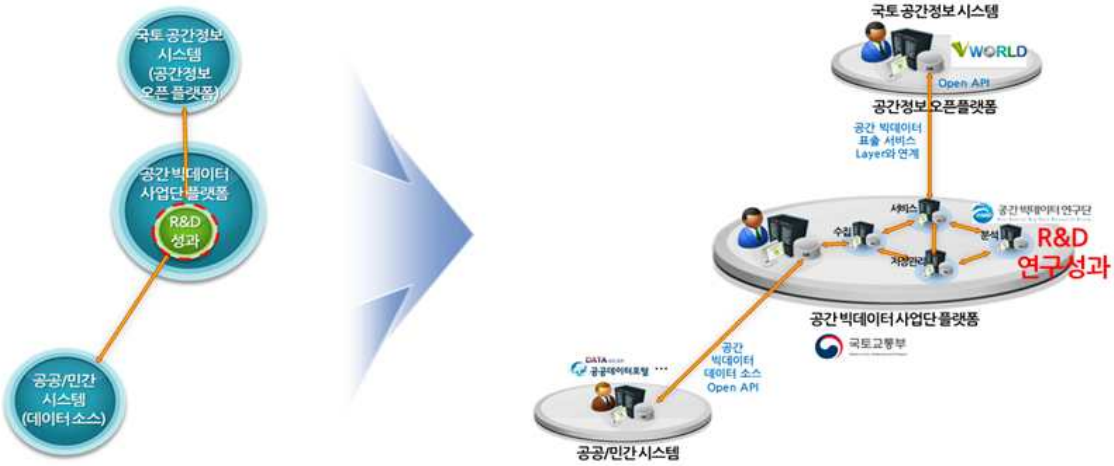
② 공간 빅데이터 실증 시스템 활용방안

○ 공간 빅데이터 실증 시스템을 공공·민간 시스템에 연계 활용하는 3가지 유형의 사업추진전략 제시

- ‘공간 빅데이터 플랫폼 구축형’ 사업은 공간 빅데이터 실증 시스템을 활용하여 독립적인 플랫폼과 운영환경을 구축하기 위한 사업유형이며, 독립플랫폼 축형과 핵심요소기술 활용형으로 세분화됨
- 공간빅데이터 체계구축사업에 R&D 성과를 성공적으로 적용하기 위해 기존에 운영해 오던 협의체(연구단-사업단-추진단)를 지속적으로 운영
  - 기관 내부 공간 빅데이터의 관리와 보안 문제로 외부 서비스를 활용하기 어렵거나, 기술적 제반 요구사항을 외부 서비스 기관에서 지원하지 못하는 경우, 혹은 자체 구축된 시스템을 통해 타 기관 및 대국민 서비스 수행을 목표로 할 때 타당한 유형



<그림 4-12> '독립 플랫폼 구축형' 사업유형 예시



<그림 4-13> '핵심요소기술 활용형' 사업유형 예시

- '공간 빅데이터 플랫폼 서비스 연계형' 사업은 독립적인 플랫폼을 자체 구축하지 않고, 외부 업체의 데이터와 시스템 환경을 활용하거나, 기관 내 시스템을 외부 업체의 시스템과 연계하여 활용하는 방식

• 외부 업체 시스템의 기능과 서비스 품질이 활용 목표 수준에 부합할 때, 매우 비용 효과적이며 조기 활용이 가능한 유형



<그림 4-14> '공간 빅데이터 서비스 연계형' 사업유형 예시

<표 4-7> 공간 빅데이터 실증 시스템 활용 사업 유형 및 설명

사업유형		설 명	용 례
공간 빅데이터 플랫폼 구축형	독립 플랫폼 구축형	• 독립된 운영·관리를 위한 공간 빅데이터 플랫폼의 구축을 전제로 하는 사업 유형 • 공간 빅데이터의 데이터 소스 제공을 목적으로 공공/민간 시스템과 공간 빅데이터 플랫폼의 수집 Layer와 연계 • 공간 빅데이터 플랫폼을 활용하여 공간 빅데이터의 저장·관리·분석·서비스를 이용 • 공간 빅데이터 사업단 플랫폼과 독립적인 플랫폼을 구축하고, 사업단의 공간 빅데이터 융합DB를 연계활용하는 방식이 이 유형에 속함	• 공간 빅데이터의 데이터 소스를 관리 및 운영하는 시스템을 보유하고 있으나, 공간 빅데이터의 저장·관리·분석·서비스를 독립적으로 구현하기 어려운 경우 • 공간 빅데이터의 데이터 소스를 제공하고, 데이터의 활용은 독립적으로 운영·관리되는 공간 빅데이터 플랫폼을 활용
	핵심 요소기술 활용형	• 공간 빅데이터 저장·관리·분석·서비스 구현을 위한 핵심요소기술을 적용하여 기존에 독립적으로 운영·관리되었던 빅데이터 플랫폼의 기능을 개선 또는 확장하기 위해 R&D 성과물을 활용하는 방식 • 공간 빅데이터 사업단 플랫폼의 테스트베드에 R&D성과를 적용하여 연계 활용하는 방식이 이 유형에 속함	• 내부 데이터의 관리정책과 보안문제로 외부 서비스를 활용하기 어려운 경우 • 공간 빅데이터 저장·관리·분석·서비스 기술을 외부 서비스 기관에서 지원하지 못하는 경우 • 자체 구축된 시스템을 통해 타 기관 및 대국민 서비스 수행을 목표로 할 경우
공간 빅데이터 플랫폼 서비스 연계형		• 독립적인 플랫폼과 운영환경을 구축하지 않고 공간 빅데이터 플랫폼에서 제공하는 OpenAPI서비스를 이용하여 공간 빅데이터 메쉬업 서비스를 구축하는 방식	• 외부 공공·민간 시스템의 기능과 서비스 품질이 활용목표 수준에 부합할 경우 • 서비스 구축비용이 효과적이며 조기 활용이 가능한 유형 • 독립적인 공간 빅데이터 플랫폼 서비스 구축 및 운영기관이 요구됨

(나) 기대효과

〈공공적 측면〉

○ 선제적 대응을 위한 예측기반 마련

- 국정 현안에 대한 사전대응 또는 미래예측과 관련하여 공간기반의 빅데이터 수집·분석체계를 활용함으로써 변화추이 및 위험징후를 빠르게 파악하여 선제적 공공서비스 제공

○ 맞춤형 정책지원 서비스 제공 가능

- 공간 기반의 빅데이터 수집·분석체계를 활용함으로써, 각 지역 및 개인에게 최적화된 맞춤형 도시계획 정책수립, 일자리 창출 등을 제공할 수 있음

○ 정부에서 개방하고 있는 공공데이터를 원시데이터로 활용하여 분석결과를 도출하도록 서비스 시나리오를 작성함에 따라 정부의 공공데이터 활용 활성화 정책에 기여

- 비즈니스 모델 및 핵심서비스 모델을 통해 제시된 보건, 복지, 안전, 환경 분야의 서비스 도출을 통해 사회적 이슈에 대한 방안 마련

〈기술적 측면〉

○ 공간 빅데이터 연구단 1, 2세부 개발기술의 유기적인 연계와 다분야 공공시스템의 데이터를 연계·활용함으로써 부문별 인적, 물적, 자원, 절약가능

- 단순 공간정보 및 통계정보 뿐만 아니라 SNS 및 Web, 실시간 데이터 등을 제공함으로써 부문별 공간 빅데이터 수집에 필요한 자료수집/가공시간 감소효과 기대

○ 공간정보 기반의 시·공간 분석, 시각화를 통해 직관적이고 신속한 의사 결정 및 편의성 지원

- 다양한 자료의 종류 및 방대한 양의 분석 결과 등 사용자가 요구하는 수요에 따른 Open API 제공과 사용자 위주의 UI/UX 개발을 통해 공간 빅데이터에 대한 접근성을 높이고 다양한 분야의 활용 가능성 기대

- 세부별 공간 빅데이터 기술 통합 환경구축을 통해 공간 빅데이터를 활용한 공간정보 서비스 구축 시 기반 기술로 활용 가능

- 공간 빅데이터 체계 구축을 위한 플랫폼 구축에 본 과제에서 개발된 공간 빅데이터 서비스 통합 기술을 접목시킴으로써 보다 완성도 높은 서비스 제공 및 개발 편의성 증대효과 기대

〈경제적·산업적 측면〉

○ 공간 빅데이터 관련 기술 확보를 통한 관련 산업 창출 및 활성화

- 공간 빅데이터 기반 새로운 기술 및 유형의 서비스 시나리오개발을 통해 관련 산업의 활성화기반을 마련하고 민간부문에서의 서비스 시나리오 개발을 위한 가이드 제공

- 공간 빅데이터 저장, 관리, 분석 핵심기술을 통합한 기존과 차별화되는 새로운 유형의 서비스 시나리오 개발을 통해 관련 산업에서 각종 서비스를 창출하는데 기여

- 대용량 데이터 분석 및 다차원분석기능을 제공함으로써 관련분야에서의 응용가능한 분석서비스로 활용가능

- Open API로 제공된 다양한 정보는 공공 및 민간분야에서 2차 가공을 통해 새로운 콘텐츠로 생성이 가능하며, 이를 통해 공간정보 및 관련 산업의 활성화가 기대되어 정부의 상업 활성화 정책에 기여할 것으로 기대됨

- 공간 빅데이터 실증 서비스 시스템 구축을 통해 공간 빅데이터 관련 사업의 진입장벽을 낮춰 공간정보 시장의 활성화 기대

○ 해외시장 선점을 위한 경쟁력을 확보

- 외산 소프트웨어 중심의 기존 공간정보 시장을 재편하고 해외 시장에서의 경쟁력

을 확보하여 국내 공간정보 관련 기업들의 해외진출에 기여

- 공공 분야의 공간 빅데이터 분석을 통해 생성된 표준화된 데이터를 공간연산 기능 연계를 통해 다양한 서비스로 확장하고 이를 활용한 공간 정보 가치 확장 및 시장성 확보

〈사회적 측면〉

○ 공간 빅데이터 서비스 제공을 위한 정책적, 제도적 지원방안을 제시함으로써 국내 4차 산업혁명에 대응한 기반마련 지원

○ 공간 빅데이터 분석을 통해 생성된 정보를 Open API로 제공하여 타 산업을 지원함으로써 데이터 공유체계 확산에 기여

- 국토교통분야 뿐만 아니라 타부처에서 활용할 수 있도록 공간 빅데이터 실증 서비스 시스템에 공공데이터 연계 등을 지원함으로써 국내 공간정보기반 사회 형성에 도움

○ 공간 빅데이터 서비스를 통한 국민 삶의 편의성 증대 기대

- 언제 어디서나 누구나 이용할 수 있는 공간정보, 통계정보, 행정정보, 공공DB 등이 융합된 빅데이터로부터 맞춤형 지식을 추출하여 서비스를 제공함으로써 국민 삶의 편의성 증대 기대

<표 4-8> 핵심연구성과의 활용계획

과제명(연구기관)		핵심성과물 (level1)	주요 활용계획	기대효과
3세부	빅데이터 서비스 활성화 방안 연구(안양대학교 산학협력단)	(I) 공간 빅데이터 플랫폼의 융복합 활용방안 개발	• 공간 빅데이터의 활용의 저변을 확대하기 위하여 전문 인력양성을 위한 교육 콘텐츠로 활용 • 공간 빅데이터 플랫폼의 수요처 발굴과 활용처의 사용자 직무교육을 위한 기술안내서 등의 콘텐츠로 활용	• 다양한 영역에서 공간 빅데이터를 활용할 수 있는 적용 예시를 제공하여 기술 수요처의 지속적인 발굴을 통한 활용영역 다변화와 기술 적용 시장 확대를 통한 빅데이터 산업의 고용 창출에 기여
		(J) 공간 빅데이터의 공공/민간 부문 핵심 서비스 모델 도출 및 제공 방안 개발	• 공간 빅데이터 플랫폼의 현장적용 가능성을 검증할 수 있는 실증적 예시 제공	• 공간 빅데이터 실증 서비스 시스템의 구현과 성과 활용을 위한 공간 빅데이터 활용가이드 개발 및 핵심서비스 모델의 활용가이드 적용 예시 개발을 통해 공간 빅데이터 실증 서비스 시스템의 활용도 향상
	공간 빅데이터 서비스 실용화 연구(공간인소프트)	(K) 공간 빅데이터 확장 Open API 서비스	• 공간 빅데이터를 표준화된 Open API 기반으로 상호 개방·공유·활용 가능하게 하고 실제 Business에 손쉽게 적용	• 공간 빅데이터를 표준화된 Open API 기반으로 상호 개방·공유·활용 기대

과제명(연구기관)		핵심성과물 (level1)	주요 활용계획	기대효과
		(L) 공간 빅데이터 서비스 제공 UI/UX	• 사용자 접근성을 고려한 상호작용 기술과 사용자의 경험을 바탕으로 효율적인 공간 빅데이터 서비스 제공	• 사용자 접근성을 고려한 상호작용 기술과 사용자의 경험 제공
	공간 빅데이터 서비스 실증 연구 (포도)	(M) 공공시스템기반 국토교통분야(부동산, 지적, 교통 등) 공간 빅데이터 실증 서비스 시스템	• 공간 저장, 공간연산, 공간 분석에 대한 가이드라인 제공 • 공간 빅데이터의 연산 및 분석 기능을 활용함으로써, 단체 및 개인에게 최적화된 공간 빅데이터 서비스 포털 구축 기반으로 활용	• 대용량 공간 빅데이터의 연산 및 분석 플랫폼 구축 기반 마련 • 교통, 사회, 경제 등 공공 공간 빅데이터의 활성화
		(N) 공공시스템기반 타 부처 공간 빅데이터 실증 서비스 시스템	• 표준화된 공간 데이터를 공간 연산과 연계하여 새로운 서비스를 구축할 수 있는 서비스 확장을 통한 서비스 포털 구축 기반으로 활용	• 새로운 서비스 확장을 통해 관련 산업의 활용도 및 확장성 증대 • 다양한 공공 공간 빅데이터 서비스 연계를 수행하여 공공시스템 공간 빅데이터의 활용도 증가 • 개방형 공간 빅데이터 서비스를 통한 맞춤형 민원 서비스 제공 및 관련 민간 사업의 사업 기획 기술지원

나. 추가 연구 및 사업화 방안

(1) 공간 빅데이터 저장관리 인프라기술 개발

○ 포털과 같은 중앙집중 서버 기반의 공간데이터 저장관리 기술 외에 경량의 공간 빅데이터 저장관리 플랫폼 개발이 필요함

- “공간빅데이터 체계구축사업”과 같은 포털 기반의 공간 빅데이터 분석 서비스는 사용자가 자신의 데이터를 포털에 적재해야 하므로 거부감이 있을 수 있음
- 경량의 공간 빅데이터 저장관리 플랫폼을 개발하여 배포하여 각 수요처가 자신의 Hadoop 플랫폼에 배포본을 직접 설치하여 사용하고 필요한 경우 적절한 기술지원을 받을 수 있도록 하는 형태의 기술 개발이 필요함

○ 관련 기술의 발전 트렌드에 따른 플랫폼의 고도화 기술 개발이 필요함

- Apache Tez를 활용하여 Map/Reduce 데이터 처리 방식의 한계를 극복하여 보다 고성능의 공간 빅데이터 처리 기능을 제공하는 기술이 필요함
- 스트림 기반 실시간 공간 빅데이터 분석 기능을 제공하는 공간 빅데이터 스트림 처리 기능 개발이 필요함
- 인메모리 기반 기술의 진화와 트렌드 변화에 따라 기 개발된 플랫폼의 고도화 및 수정 보완이 필요함
- 이동객체 자료의 활용을 위해 GPX, NMEA, TCX와 같은 이동객체 자료 포맷 지원 및 변환기능 개발이 필요함

<표 4-9> 추가 연구 및 사업화 방안

구분	과제명 (연구기관)	연구명 또는 사업명	주요내용
1세부	공간 빅데이터 저장관리 인프라 기술 개발 (한국전자통신연구원)	Hadoop 기반 공간 빅데이터 저장 프레임워크	• Apache Tez를 활용하여 Map/Reduce 데이터 처리 방식의 한계를 극복하여 보다 고성능의 공간 빅데이터 처리 기능을 제공할 기술이 필요함 • 현재 국토교통부에서 운영하고 있는 “공간빅데이터 체계구축사업”에서 제공하고 있는 포털 기반의 공간 빅데이터 분석 서비스는 사용자가 자신의 공간 데이터를 포털에 적재해야 하기 때문에 거부감이 있을 수 있음 • 이를 해결하기 위해서는 경량의 공간 빅데이터 저장관리 플랫폼을 개발하여 배포하여, 각 수요처가 자신의 Hadoop 플랫폼에 배포본을 직접 설치하여 사용하고 필요한 경우 적절한 기술지원을 받을 수 있도록 하는 형태의 기술 개발이 필요함
		공간 빅데이터 질의처리 엔진	• 지금까지 개발된 배치 기반의 공간 빅데이터 분석 뿐만 아니라 스트림 기반 실시간 공간

			빅데이터 분석 기능을 제공하는 공간 빅데이터 스트림 처리 기능 개발이 필요함
	인메모리 기반 공간 빅데이터 저장관리 프레임워크 개발 (포스웨이브)	인메모리 기반 공간빅데이터 저장관리 프레임워크	• 인메모리 기반 기술의 진화와 Trend 변화에 따라 기 개발된 플랫폼의 고도화 및 수정 보완이 필요함 • 이동객체 공간 자료에 대한 시공간기반의 분석 기능 및 모델 개발이 필요함 • GPX, NMEA, TCX와 같은 이동객체 정보 저장 포맷 improt/exprot기능 보완이 필요함
	공간 빅데이터 웨어하우스 핵심기술 개발 (아이씨티웨이)	스트림 기반 분산 공간 ETL 도구	• 다변하는 공간 정보 서비스 시장의 변화에 따라 연구개발된 공간 ETL 도구에 지속적인 고도화 및 새로운 공간 정보 분석을 위한 ETL 기능 개발이 필요함 • 다양하고 새롭게 등장하는 공간 센서 및 다양한 원천 데이터에서 공간 정보를 실시간 추출, 변환, 적재하는 공간 ETL 추가 연구가 필요함

(2) 공간 빅데이터 분석 및 가시화 기술 개발

○ 국가 차원의 공간 빅데이터 관련 코퍼스, 감성어 사전등을 구축하며 사용자 참여 플랫폼을 개발하고자 함. 또한 AWS (Amazon Web Service) 클라우드 시스템에 관련 기술을 탑재하여 추후 연구에 활용하며, 공간 빅데이터 연구단에 기술 적용 및 현장 검증을 통해 기술 고도화를 추진함.

<표 4-10> 추가 연구 및 사업화 방안

구분	과제명 (연구기관)	연구명 또는 사업명	주요내용
2세부	공간 빅데이터 소셜 분석 기술 개발 (서울대/시터스)	국가적 공간 빅데이터 분석 지원 연구	• 국가 차원의 공간 빅데이터(장소 리뷰 데이터) 코퍼스 구축 및 제공 기술 개발 • 국가 차원의 공간에 특화된 개체명 인식기 개발 및 제공 기술 개발 • 국가 차원의 공간 감성어 사전 구축 기술 및 사용자 참여 플랫폼 개발
	공간 빅데이터 일괄(Batch) 분석 컴포넌트 기술 개발 (연세대/가이온)	공간 빅테이터 Batch 분석 컴포넌트	• AWS(Amazon Web Service) 클라우드 시스템을 구축하여 공간 빅데이터 분석 컴포넌트를 연구실 내에 탑재하고 기능 고도화 및 향후 추가연구에 활용할 예정
	이동객체데이터 Batch 분석 컴포넌트 (연세대/가이온)	이동객체데이터 Batch 분석 컴포넌트	• AWS(Amazon Web Service) 클라우드 시스템을 구축하여 이동객체데이터 분석 컴포넌트를 연구실 내에 탑재하고 기능 고도화 및 향후 추가연구에 활용할 예정
	공간 빅데이터 CEP 및 OLAP 엔진 개발 (한국과학기술원/유엔진 솔루션즈)	공간 빅데이터 Interactive 분석 컴포넌트	• 공간 빅데이터 연구단에 기술 적용 및 현장 검증을 거쳐 기능 고도화

○ 공간 빅데이터 및 이동객체데이터 Batch 분석 플랫폼의 향후 활용을 위하여, 일반 및 고급 통계/비통계 기법을 활용한 신규 공간 빅데이터 분석 서비스 시나리오가 포함된 사용자 매뉴얼(총 30종의 분석 컴포넌트 기능)을 제시하여, 기본적인 데이터 분석에 대한 지식이 있는 학사 이상의 공공/민간 수요처의 사용자가 활용 가능하도록 제시하고자 함

(3) 공간 빅데이터 서비스 활성화 및 실증 연구

○ 공간 빅데이터 분석 플랫폼(사업단) 적용 및 안정화와 공간 빅데이터 전문인력 양성 및 지속적인 홍보활동 추진. 또한 공간 빅데이터를 표준화된 OpenAPI기반으로 상호 개방・공유・활용을 지원하는 등의 사업 추진

<표 4-11> 추가 연구 및 사업화 방안

구분	과제명 (연구기관)	연구명 또는 사업명	주요내용
3세부	빅데이터 서비스 활성화 방안 연구(안양대학교 산학협력단)	공간 빅데이터 분석 플랫폼(사업단) 적용 및 안정화 추진	• 공간 빅데이터 분석 플랫폼(사업단) 상에서 원활한 R&D 성과 실증 및 유지관리·활용을 위한 관련 분석모델·기술 등 지원방안 마련
		공간 빅데이터 전문인력 양성 및 지속적 홍보	• 활용모델과 가이드라인 개발·보급 및 교육 컨설팅 • 지속적인 정보와 노하우 공유 및 우수사례 확산
	공간 빅데이터 서비스 실용화 연구(공간인소프트)	공간 빅데이터를 표준화된 Open API 기반으로 상호 개방·공유·활용 필요	• 실제 Business에 손쉽게 적용할 수 있도록 지속적인 개술 개선 및 적용 확대 필요
			• 사용자 접근성을 고려한 상호작용 기술과 사용자의 경험을 바탕으로 효율적인 공간 빅데이터 서비스 제공 필요
	공간 빅데이터 서비스 실증 연구 (포도)	공간 빅데이터 서비스 실증 연구	• 분석 완료 예상 시간이 예측 불가하고 분석별로 완료시간이 상이하므로 사용자 측면에서 이를 해결하기 위한 알림 등을 전달할 수 있는 고도화 기능 필요
			• 공공 기관에서 연계한 데이터 무결성 확보를 위한 주기적인 데이터 관리 및 모니터링 시스템 구축 필요
			• 공공데이터 기반의 서비스 사례를 제공하고 커뮤니케이션이 가능한 소통채널에 대한 추가 연구 필요
			• SNS등 외부데이터 수집 및 연계관련 내용이 전 세부 통틀어 과제내용으로 존재하지 않기 때문에 향후 시스템 설치 및 운용 시 수집기능에 대한 추가적인 기술개발이 필요함

붙임1. 참고 문헌

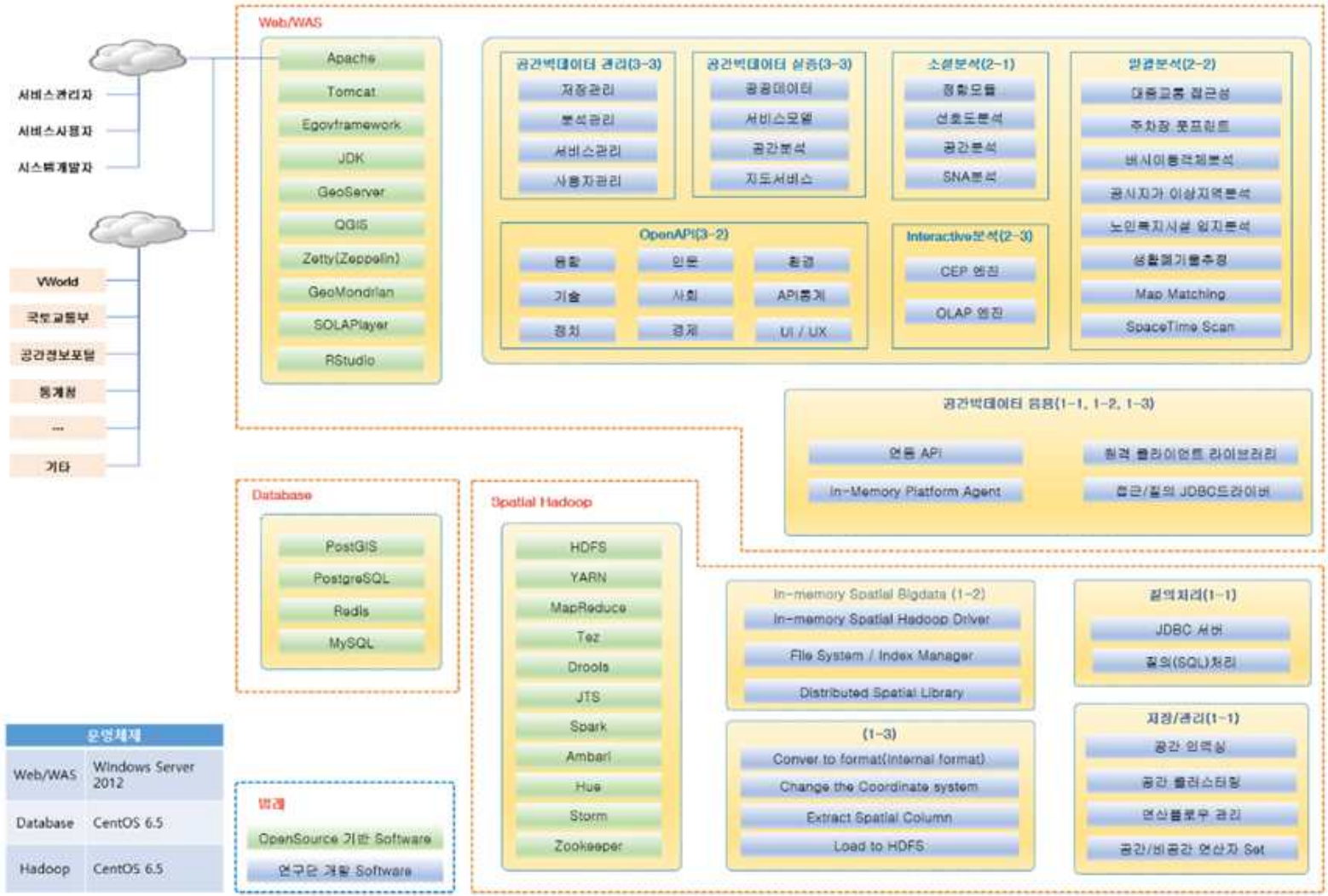
- Bum-Soo Kim, 2016, Boundary Image Matching Supporting Partial Denoising Using Time-Series Matching Techniques, Multimedia Tools and Applications 제76권 제6호.
- Jungeun Kim, 2015, Community Detection in Multi-Layer Graphs: A Survey, SIGMOD Record 제44권 제3호.
- Jungeun Kim, 2018, LinkBlackHole: Robust Overlapping Community Detection Using Link Embedding, IEEE Transactions on Knowledge and Data Engineering 제2018권.
- Minseo Kang, 2017, An experimental analysis of limitations of MapReduce for iterative algorithms on Spark, Cluster Computing 제20권 제4호.
- Nguyen Minh Hieu, 2017, A Benchmark Test of Spatial Big Data Processing Tools and a MapReduce Application, 한국측량학회지 제35권 제5호.
- Sangmin Kim, 2015, Development of a 3D Underground Cadastral System with Indoor Mapping for As-Built BIM: The Case Study of Gangnam Subway Station in Korea, Sensors 제15권 제12호.
- Sangmin Kim, 2017, Development of 3D Underground Cadastral Data Model in Korea : Based on Land Administration Domain Model, Land Use Policy 제60권.
- 강기연, 2016, Relationships Between Nutrient Patterns and Geodemographic Characteristics in Korea, Nutrition & Dietetics 제73권 제5호.
- 곽재하, 2017, Optimal location selection for the installation of urban green roofs considering honeybee habitats along with socio-economic and environmental effects, Journal of Environmental Management 제189권.
- 김경혜, 2018, 지리가중라소 모형을 활용한 서울시 빈집 발생 영향요인 탐색 연구, 대한산업공학회지 제44권.
- 김대욱, 2015, 3차원 공간플랫폼기반 제방관리시스템 설계 및 구축, 2015 KAGIS 춘계학술대회.
- 김대욱, 2015, V-World 프레임워크를 활용한 생활안전 서비스 기술개발, 2015 KAGIS 춘계학술대회.
- 김대욱, 2015, 공간정보 오픈플랫폼기반의 교통영향평가서비스 프레임워크 설계방안에 관한 연구, 2015 KAGIS 춘계학술대회.
- 김동하, 2016, 공간가중회귀 모형을 이용한 서울시 에너지 소비에 따른 이산화탄소 배출 분석, 대한 산업공학회지 제42권 제2호.
- 김민석, 2018, 장소 추천을 위한 방문 간격 보정, 데이타베이스 연구 제34권 제2호.
- 김민선, 2018, KeyplayerRank를 이용한 SNS상의 주제별 키플레이어 탐지기법, 한국지형공간정보학회지 제26권 제4호.
- 김소현, 2018, 실내 추적 데이터에서의 고객 재방문 패턴 탐색 연구, 데이타베이스 연구 제34권 제2호.
- 김정은, 2015, 다양한 실제 네트워크에서 클러스터링 계수와 클러스터 경향의 상관관계 분석, 데이타베이스연구 제31권 제2호.
- 김준성, 2016, 조건이 있는 k-최근접 객체 질의 처리 방법, 데이타베이스 연구 제32권 제2호.

- 김지영, 2018, A multi-criteria decision-making approach for geometric matching of areal objects, Transactions in GIS 제22권 제1호.
- 김창수, 2018, 하둡 및 Spark 기반 공간 통계 핫스팟 분석의 분산처리 방안 연구, 정보과학회 논문지 제45권 제2호.
- 김형진, 2014, 유전자 알고리즘을 이용한 서울시 군집화 최적 변수 선정, 한국지형공간정보학회 논문집 제22권 제4호.
- 박우진, 2015, 격자기반 분석을 통한 위치기반 소셜 미디어 데이터와 부동산 가격지수 간의 공간적 상관성 분석 연구, 한국지형공간정보학회지 제23권 제1호.
- 박정, 2017, An optimization approach for the placement of bicycle-sharing stations to reduce short car trips: An application to the city of Seoul, Transportation Research Part A 제105권.
- 박준민, 2015, 공간 빅데이터 서비스 활성화를 위한 정책과제 도출, 한국공간정보학회 제23권 제6호.
- 박준민, 2016, A study on policy and System Improvement Plan of Geo-Spatial Big Data Services on Korea, 한국측량학회 제34권 제6호.
- 방윤식, 2016, An Improved Map-Matching Technique Based on the Fréchet Distance Approach for Pedestrian Navigation Services, Sensors 제16권 제10호.
- 방윤식, 2016, 길찾기 과정의 도로명주소 체계 연계를 위한 선형 객체 매칭 방법, 한국지형공간정보학회지 제24권 제4호.
- 서양모, 2015, 공간빅데이터를 위한 정보 시각화 방법, 한국공간정보학회지 제23권 제6호.
- 양민선, 2018, A 3D simulation of river basin boundary based on the change of water-level using open government data in South Korea, 한국공간정보학회 제26권 제2호.
- 양종현, 2016, 교통사고 데이터와 GIS를 이용한 보행자사고 개선구역 선정 : 서울시를 대상으로, 한국측량학회지 제34권 제3호.
- 양종현, 2017, 비정형 데이터를 이용한 층간소음 탐지: 네이버 카페를 대상으로, 한국지형공간정보학회지 제25권 제3호.
- 양현석, 2016, 스마트폰 사용자의 중단가능성 예측 시 콜드 스타트 문제 해결 연구, 데이타베이스 연구 제32권 제3호.
- 양현석, 2016, 신규 스마트폰 사용자의 중단가능성 예측 정확도 향상 연구, 데이타베이스 연구 제32권 제2호.
- 어승원, 2016, Establishing the Process of Spatial Informatization Using the Data from Social Network, 한국측량학회지 제34권 제2호.
- 오원섭, 2016, 2012년 전국 지역사회 건강조사 자료를 이용한 시군구 단위 질병 유병률의 공간 자기상관도에 대한 연구, 한국측량학회지 제34권 제3호.
- 유선철, 2014, 공간 빅데이터의 개념 및 서비스 프레임워크 구상에 관한 연구, 한국공간정보학회 제22권 제6호.
- 유선철, 2015, A Study on Concepts and Utilization of Geo-Spatial Big Data in South Korea, 대한토목학회 제20권 제7호.
- 유선철, 2019, Analysis of Land Use, Land-Use Change, and Forestry (LULUCF) and Construction of Statistics in Korea, Sensors and Materials 제31권.

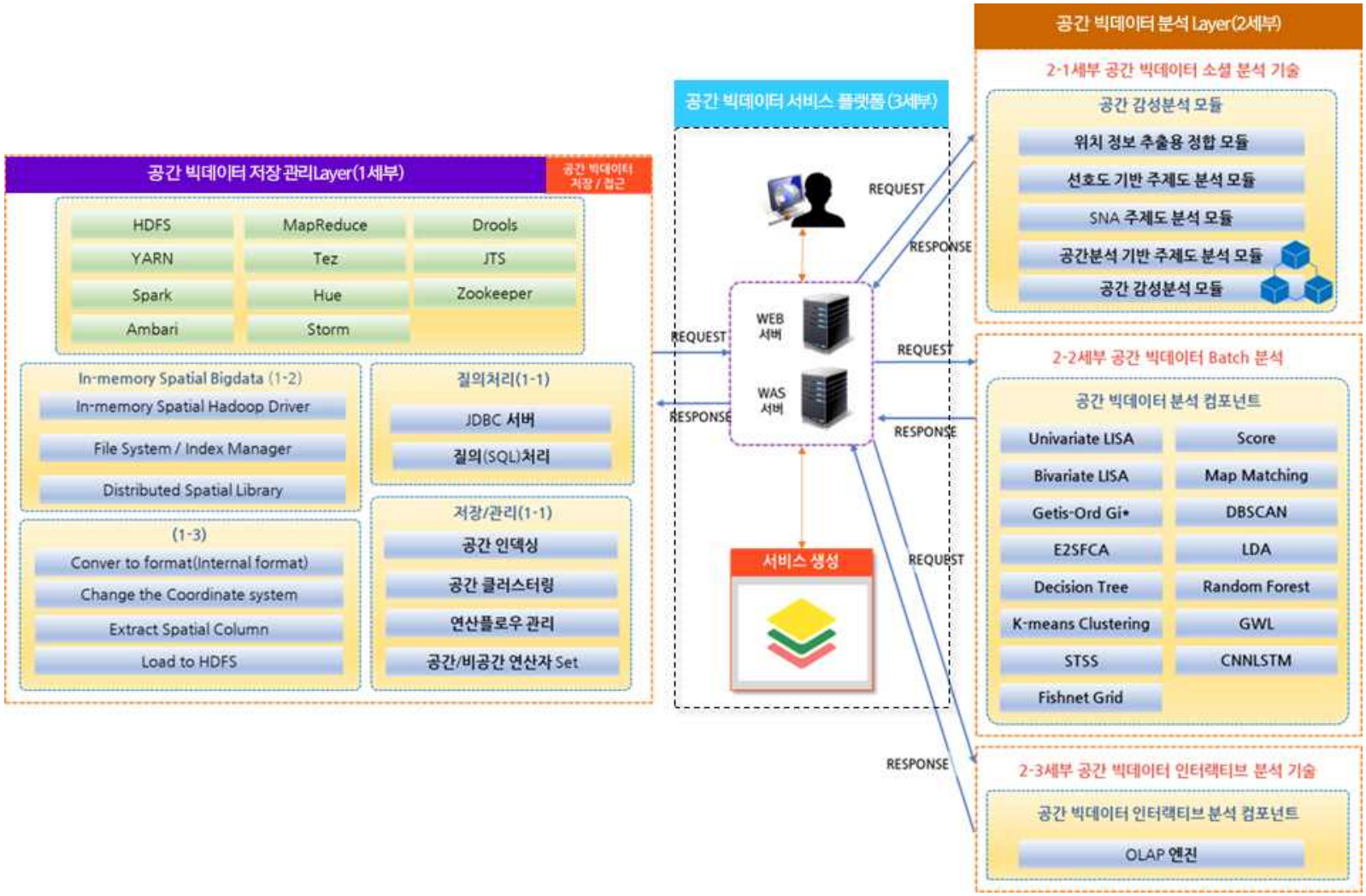
- 유선철, 2019, 공간 빅데이터를 활용한 서울시 건물부문 온실가스 배출량 공간 분석 연구, 대한공간정보학회지 제27권 제4호.
- 유수정, 2017, 한국 스마트 카드 데이터를 활용한 대중교통 승객의 이동 패턴 기반 위치 판별 기법, 데이타베이스 연구 제33권 제2호.
- 윤상현, 2018, Geographical variations and influential factors in prevalence of cardiometabolic diseases in South Korea, PLOS One 제13권.
- 윤성범, 2017, A combinatorial Optimization for Influential Factor Analysis : A Case Study of Political Preference in Korea, 한국측량학회지 제35권 제5호.
- 윤수식, 2015, 실시간 공간 빅데이터 분석을 위한 복합 이벤트 처리 시스템 설계 및 응용, 데이타베이스연구 제31권 제3호.
- 윤수식, 2017, 실시간 공간 빅데이터 스트림 분산 처리를 위한 부하 균형화 방법, 정보과학회논문지 제44권 제11호.
- 이명호, 2015, 공간 빅데이터 핵심서비스 선정에 관한 연구, 한국측량학회 제33권 제5호.
- 이명호, 2016, GIS 공간분석을 활용한 사회 취약지역의 분석에 관한 연구 : 서울특별시를 중심으로, 한국지형공간정보학회 제24권 제4호.
- 이영민, 2017, 한국어 장소 리뷰를 이용한 공간 정보 감성어 사전 구축 방법, 한국지형공간정보학회지 제25권 제2호.
- 이원경, 2017, Taxi vacancy duration: a regression analysis, Transportation Planning and Technology 제40권 제7호.
- 이원경, 2018, Utilizing mobile phone-based floating population data to measure the spatial accessibility to public transit, Applied Geography 제92권.
- 이원상, 2015, 공간 빅데이터 연구 동향 파악을 위한 토픽모형 분석, 대한 산업공학회지 제41권 제1호.
- 이재은, 2017, Machine Learning Classification of Buildings for Map Generalisation Purposes, ISPRS International Journal of Geo-Information 제6권 제10호.
- 이지원, 2019, SNS 데이터를 이용한 공공시설 매력도지수에 따른 접근성 분석기법, 한국측량학회지 제37권 제1호.
- 임장혁, 2018, 순환 신경망과 데이터 증강을 이용한 재난 정보 탐지 방법, 한국지형공간정보학회지 제26권 제4호.
- 조정희, 2018, High-Performance Geospatial Big Data Processing System Based on MapReduce, ISPRS International Journal of Geo-Information 제7권 제10호.
- 조정희, 2018, Marmot 공간 빅데이터 처리를 위한 하둡 기반의 고성능 데이터 저장관리 시스템, 한국지형공간정보학회지 제26권 제1호.
- 조준형, 2017, 공급적정성 및 서비스 효용성 평가를 기반 도보권 근린공원의 최적입지 선정, 대한국토도시계획학회지 제52권 제2호.
- 주인학, 2017, Hadoop에서 SQL 기반 질의언어를 지원하는 공간 빅데이터 질의처리 시스템, 한국정보전자통신기술학회 논문지 제10권 제1호.
- 차맹규, 2016, A Mobile Object Tracking Scheme by Wired/wireless Integrated Street Lights with RFID, 한국지형공간정보학회지 제24권 제2호.

- 채인영, 2017, 소셜 미디어 텍스트를 이용한 장소 선호도 분석 기법, 한국지형공간정보학회지 제25권 제4호.
- 최원욱, 2018, Study on the Development of Geo-Spatial Big Data Service System Based on 7V in Korea, 대한토목학회 제23권 제1호.
- 최원욱, 2018, Text mining geo-visualization of patent documents on geo-spatial big-data industry, Spatial Information Research 제27권 제1호.
- 최원욱, 2019, Development and Demonstration of Grid-Based SpatioTemporal Data Model for Monitoring Small Businesses, Sensors and Materials 제32권.
- 최장범, 2016, 공간정보 빅데이터 기반기술의 통합 시각화, 정량화 등의 실증서비스 기술 설계에 관한 연구, 2016 KAGIS 공동추계학술대회.
- 최장범, 2017, 공공 공간 빅데이터 서비스 확장 기술 개발을 위한 분석 기술 설계, 한국지형공간정보학회 추계학술대회.
- 최장범, 2018, 공공 공간 빅데이터 서비스 확장 기술 개발을 위한 분석 기술 개발, 2018 KAGIS 공동추계학술대회.
- 최중현, 2015, 복합 이벤트 스트림 질의의 성능 개선을 위한 질의 전처리 도구, 정보과학회 컴퓨팅의 실제 논문지 제21권 제8호.
- 최중현, 2018, Top-k 질의 처리를 위한 오픈소스 복합 이벤트 처리 시스템의 확장, 정보과학회 컴퓨팅의 실제 논문지 제24권 제4호.
- 최형주, 2017, 그래프 분할 기반 RETE 네트워크 분산 처리, 데이타베이스 연구 제33권 제2호.
- 최형환, 2015, 멀티 브라우저에서 작동 가능한 3D GIS 가시화 기술 개발, 한국공간정보학회.
- 최형환, 2015, 멀티브라우저 지원을 위한 3차원 공간정보체계 가시화 기술개발에 관한 연구, 한국공간정보학회.
- 최효신 , 2015, 지리적 표시제에 대한 관심이 농산물 가격변화 예측에 미치는 영향 연구 : 사과를 사례로, 대한 산업공학회지 제41권 제4호.
- 한은진, 2017, 외국인 증가와 주거이동과의 공간 상관관계 분석, 대한국토도시계획학회지 제52권 제4호.
- 한은진, 2018, Utilizing mobile phone-based floating population data to measure the spatial accessibility to public transit, Applied Geography 제92권.

붙임2. 공간 빅데이터 관리, 분석 및 서비스 플랫폼의 세부별 모듈 구성도(물리적 관점)



붙임3. 공간 빅데이터 관리, 분석 및 서비스 플랫폼의 세부 모듈 간 연계 구성도



주 의

1. 이 보고서는 국토교통부에서 시행한 국토공간정보연구사업의 연구보고서입니다.

2. 이 보고서 내용을 발표하는 때에는 반드시 국토교통부에서 시행한 국토공간정보연구사업의 연구결과임을 밝혀야 합니다.

3. 국가과학기술 기밀 유지에 필요한 내용은 대외적으로 발표 또는 공개하여서는 안 됩니다.