

KCA연구 2018

# 빅데이터를 이용한 효율적인 전파감시 방안 연구

(최종보고서)

2019. 02.

한국방송통신전파진흥원

연구수행기관 : (사)한국전자통신학회

이 보고서는 한국방송통신전파진흥원의 출연에 의한  
재정지원으로 이루어졌습니다.

# 제출물

한국방송통신전파진흥원장 귀하

본 보고서를 『빅데이터를 이용한 효율적인 전파감시 방안 연구』의 결과보고서로 제출합니다.


2019년 02월

연구기관 : (사)한국전자통신학회  
                  신현식 교수  
총괄책임자 : 석경휴 교수  
참여연구원 : 윤홍주 교수  
                  최세일 교수  
                  홍완표 교수  
                  김규철 교수  
                  정유선 박사

## 요 약 문

1. 제목 : 빅데이터를 이용한 효율적인 전파감시 방안 연구

2. 연구내용 및 범위

연구 주제	연구 내용
전파감시데이터 활용 절차	<ul style="list-style-type: none"> <li>빅데이터 사업 추진 단계별 점검 절차               <ul style="list-style-type: none"> <li>① 요건 정의, ② 데이터 수집, ③ 데이터 전처리, ④ 분석 모델링, ⑤ 모델 구현, ⑥ 시각화, ⑦ 검증 및 안정화</li> </ul> </li> <li>기관의 데이터 품질관리               <p>데이터 진단을 통해 도출된 이슈 및 시사점에 따라 기관이 보유한 데이터의 활용성 향상을 위해서는 품질관리목표에 맞춰 계획, 구축, 운영, 활용 단계별 품질활동이 필요하다.</p> <ul style="list-style-type: none"> <li>중앙전파관리소 데이터 품질관리(Plan-Do-Check-Act)                    </li> </ul> </li> <li>기관의 데이터 품질 진단 및 개선절차               <p>데이터 품질진단 및 개선은 총 4단계의 절차를 거쳐 진행하도록 한다.</p> <ul style="list-style-type: none"> <li>① 진단 대상 선정, ②진단 방법 정의, ③ 진단 수행, ④ 진단 결과 분석 및 관리</li> </ul> </li> </ul>
전파감시 데이터 분석 및 과제선정	<ul style="list-style-type: none"> <li>분석 과제 선정 및 필요 데이터               <ul style="list-style-type: none"> <li>① 전파품질측정 데이터 및</li> <li>② 고조파측정 데이터</li> <li>③ 주파수 점유율 조사 데이터 등</li> </ul> </li> <li>분석 과제 “전파 혼·간섭 예방” 선정               <p>무선국의 출력증강, 고조파 발생 등에 의한 전파 혼·간섭 예방을 위해 전파감시 빅데이터의 분석을 통한 선제적인 대응이 필요하다.</p> </li> </ul>

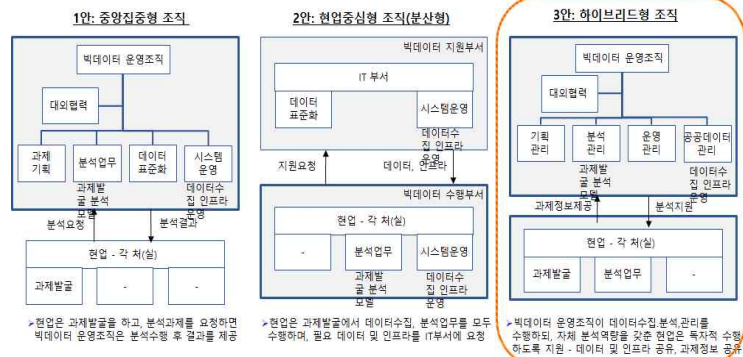
	<ul style="list-style-type: none"> <li>분석과제 추진에 따른 기대효과 무선국 출력증가 및 고조파로 인한 전파 혼·간섭 사전과악 및 예방조치가 가능하다.</li> <li>- 주파수 이용현황(점유율) 신뢰도 제고</li> <li>- 무선국의 허가지역 외 불법 운용여부 파악</li> <li>- 시스템 장애 발생 전 예방조치 가능</li> </ul> <p>무선국의 변칙운용 및 고조파를 분석하여 전파 혼·간섭을 예방하고, 주파수이용현황 조사결과의 신뢰도 제고 및 불법전파 파악, 시스템 장애 발생 전 예방조치가 가능하다.</p>
<p>분석모델 시각화 제안</p>	<ul style="list-style-type: none"> <li>정보 시각화 절차           <div> <div>정보 조직화 단계</div> <ul style="list-style-type: none"> <li>• 사용자의 정보 인지에 관여</li> <li>• 수집된 데이터를 분류하고 배열하여 조직화하고 질서를 부여</li> </ul> <div>↓</div> <div>정보 시각화 단계</div> <ul style="list-style-type: none"> <li>• 사용자의 정보 지각에 관여</li> <li>• 효율적인 정보 전달을 위해 시각, 청각 등의 감각 기관에 최적의 자극의 제시</li> </ul> <div>↓</div> <div>상호작용 단계</div> <ul style="list-style-type: none"> <li>• 정보와 사용자 간의 상호작용 측면의 사용자 경험을 디자인</li> <li>• 정보의 인지적 요인과 지각적 요인을 함께 활용</li> </ul> </div> </li> <li>빅데이터 시각화 라이브러리 구성 방안           <div> <div>플랫폼 통합 관리</div> <div>각 분석과제 개별 활용</div> </div> </li> <li>빅데이터 시각화 표현 유형           <p>① 네트워크 시각화 유형, ② 계층 시각화 유형, ③ 통계적 분석 시각화 유형, ④ 빈도 시각화 유형, ⑤ 2차원 비교 시각화 유형, ⑥ 입출력 시각화 유형, ⑦ 다측정 시각화 유형, ⑧ 상관분석 및 비율 시각화 유형, ⑨ 지도 시각화 유형, ⑩ 시간 시각화 유형</p> </li> </ul>
데이터 분석가 양성	<ul style="list-style-type: none"> <li>빅데이터분석가 자격분류</li> <li>① 데이터분석전문가, ② 데이터분석준전문가</li> </ul>

필요성 및  
교육 방안

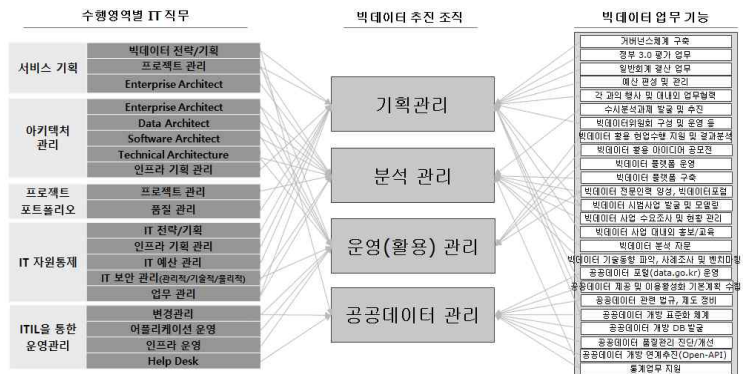
## ○ 국내의 데이터 직무 구분

DA	데이터 개발자		데이터 엔지니어		데이터 분석가	DBA	데이터 과학자	데이터 컨설턴트	데이터 마케터
	DB 개발자	빅데이터 개발자	DB 엔지니어	빅데이터 엔지니어					

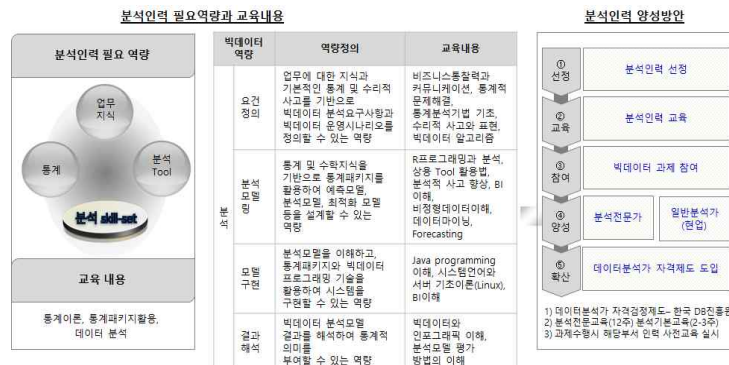
## ○ 빅데이터 조직 대안 설정



## ○ 빅데이터 추진 조직 부문 도출




## ○ 빅데이터 분석 인력 양성 방안




## ○ 빅데이터 운영 인력 양성 방안

### 운영인력 필요역량과 교육내용

운영인력 필요 역량	빅데이터 역량	역량정의	교육내용
	데이터 수집	빅데이터 플랫폼에 대한 지식과 빅데이터 수집유형, 수집방법, 수집기술을 이해하고, 사용할 수 있는 역량	데이터 플랫폼 이론, 빅데이터의 수집
	데이터 저장	빅데이터 플랫폼에 대한 지식과 빅데이터 저장유형, 저장방법, 저장기술을 이해하고, 사용할 수 있는 역량	데이터 플랫폼 이론, 빅데이터의 저장 및 관리, 빅데이터 파티션 및 분석, 빅데이터 플랫폼 구축 및 활용
	데이터 보안	빅데이터 보안과 개인정보보호 등에 대한 마인드를 갖고 있으며, 데이터보안기술 지식을 기반으로 데이터보안을 실현할 수 있는 역량	빅데이터 보안분석, 빅데이터 환경의 개인정보보호 대책
	시스템 운영	빅데이터 플랫폼을 이해하고, 보안 및 복구기술 등을 적용하여 시스템의 안정적 운영을 유지할 수 있는 역량	빅데이터 플랫폼과 각종 용량성장기술을 이해하고, 빅데이터 환경의 설치와 운용, 빅데이터 아키텍처 설계방법론
	데이터 활용 관리	빅데이터의 활용을 예측하고, 용량성장기술을 기반으로 데이터저장장치 및 저장방법을 선택하고 적용할 수 있는 역량	-

### 운영인력 양성방안



1) 기술전문가 교육-한국 DB진흥원 (10주)  
2) 과제 수행시 해당 인력 사전교육 실시

## ○ 빅데이터 플랫폼 구축방안

<b>수집 플랫폼</b>	<ul style="list-style-type: none"> <li>단계별 구축전략에 따른 빅데이터 플랫폼 설계</li> <li>내/외부데이터 수집(ETL, Crawling, API등) 설계</li> <li>데이터 수집분야 아키텍처 설계</li> </ul>
<b>저장 플랫폼</b>	<ul style="list-style-type: none"> <li>데이터 적재·보관·처리·보안·관리 분야 아키텍처 설계</li> <li>데이터 백업구조 설계</li> </ul>
<b>분석 플랫폼</b>	<ul style="list-style-type: none"> <li>위험정보 분석을 위한 표준 프레임워크 정의 및 구축방안 수립</li> <li>사업 추진과제별 분석모형(서술, 진단, 예측, 처방, 최적화) 설계</li> </ul>
<b>활용 플랫폼</b>	<ul style="list-style-type: none"> <li>데이터 시각화 Tool 정의</li> <li>텍스트, 이미지, 동영상 등 다양한 데이터를 분석 제공하는 활용 플랫폼</li> </ul>

전과감시

빅데이터

시스템 구축

방안

## ○ 연차별 사업 추진 방안

중앙전과관리소의 빅데이터 분석 플랫폼 구축은 빅데이터 플랫폼 구축 연구과제 수행(정보화 전략 계획 수립), 전과 혼·간섭 예방 빅데이터 분석 시범사업 및 빅데이터 분석 플랫폼 구축, 빅데이터 분석 플랫폼 고도화에 따른 연차별 사업 추진 전략을 수립하고, 각 단계별 목표는 아래와 같다.

- 1차 년도 : 연구과제 결과를 바탕으로 일부 과제 검증을 위한 시범사업
- 2차 년도 : 시범사업을 통해 검증된 빅데이터 분석 플랫폼 구축 및 활용방안 수립
- 3차 년도 : 빅데이터 플랫폼 서비스 고도화 및 확대



### 3. 연구결과

#### ◇ 전파감시데이터 활용 절차

- 전파 분야 빅데이터 분석과제의 추진 담당자가 추진절차를 실제 업무에 적용하기 위해서 점검해야 할 “빅데이터 사업 추진 단계별 점검 절차”를 제시하였다.

#### ◇ 전파감시데이터 현황분석



- 전파감시 분석과제 선정을 위해 전파감시업무 현황을 파악하고, 전파감시 데이터 현황 및 업무에 따른 DB현황 상세 테이블을 분석하였다.
- 또한, 전파감시 데이터 활용방안을 조사하고, 선정된 분석과제의 연관 데이터를 정리하여 활용 가능한 데이터 목록을 도출하고 내부데이터 현황을 진단하였다.

#### ◇ 분석모델 시각화 제안

- 실시간 저장되는 대량의 데이터를 모두 해석하기에는 많은 제약이 따르기 때문에 데이터를 선별적으로 선택해서 짧은 시간에, 적은 양의 자료를, 해석하기 쉬운 형태로 시각화하는 것이 가장 적합한 방법이라고 할 수 있다. 이에 시각화 도구 및 기술선정 능력을 고려하여 데이터 시각화 절차에 맞춰서 전파감시 데이터를 시각화하는 방법을 고려해야 한다.
- 분석과제 시각화는 우선적으로 빅데이터 플랫폼의 다차원 분석 솔루션에서 제공하는 시각화 유형을 사용하고, JavaScript 기반의 무료 차트 라이브러리를 활용하되 추가적 요건이 있는 경우 D3 등의 개발 도구를 활용할 수 있다.
- 과제별 시각화 도구 적용은 빅데이터 플랫폼의 다차원 분석 솔루션에서 제공하는 시각적 표현 기법을 우선 사용하고, Google Chart, HIGHCHARTS, Fusion Chart, Keylines 등 JavaScript 기반의 시각화 오픈 소스를 활용할 수 있다. 기존 솔루션에서 제공되지 않는 사용자에게 특화된 표현이 필요한 경우 D3을 활용하여 구현할 수도 있으며,
- 전파 혼·간섭 예방 분석모델의 시각화 방법은 지도 시각화 유형에 통계적 분석 시각화, 빈도 시각화, 시간 시각화 방안이 가장 효과적일 것으로 판단된다.

#### ◇ 데이터 분석가 양성 필요성 및 교육 방안

- 중앙전파관리소의 업무에 빅데이터를 적용 시, 각 분야별로 적용하는 경우에는 자칫 예산만 낭비하고 실패를 낼 가능성이 높다. 그러므로, 우선 빅데이터 적용이 필요한 분야를 도출하여 이를 통합한 빅데이터 플랫폼의 구축과, 분석 목적에 해당하는 데이터에 관한 정비작업과 통합된 빅데이터 관리 및 분석을 할 수 있는 능력을 갖춘 독자적인 조직(빅데이터센터(가칭))이 필요하다. 그러나 현재 여건상 이러한 독자적인 조직을 구축하기에는 예산상의 한계가 있을 수도 있다.
- 따라서 단기적으로는 빅데이터 전문가를 포함한 전담팀 혹은 전담인력을 포함하는 조직의 구성이 필요하고, 장기적으로 해당 조직을 정비, 보완해 나아가야 할 것이다.
- 빅데이터 추진 조직은 내부조직 간의 경계가 모호한 업무 또는 긴밀한 협력체계가 요구되는 업무를 고려하여 기획관리, 분석관리, 운영관리, 공공데이터 관리의 4개 기능을 수행하는 新 빅데이터 사업팀으로 구성하고, 빅데이터 조직 및 인력의 역량을 감안하여 데이터의 통합관리에 적합한 하이브리드형 조직으로 구성하는 방안이 효과적이다.
- 조직 구성 방안의 소요인력 도출에 있어서는 소요인력 산정방법에 따라 업무 기능별 적정 인력 추정치를 도출하였고, 향후 빅데이터 분석시스템의 규모에 따라 소요인력은 변동될 수 있다.
- 빅데이터 분석 조직의 주요업무는 내부 및 외부데이터 조사, 빅데이터 분석시스템 구축 및 운영관리, 빅데이터 분석 및 활용서비스 개발, 데이터 개방 및 활성화이며, 주요업무에 맞춰 적정 소요 인력을 산정하였고, 중앙전파관리소의 부서별 인사이동을 대비한 데이터 인력 양성방안으로는 분석인력과 운영인력 양성방안으로 구분해서 양성하는 방안을 제안하였다.

#### ◇ 전파감시 빅데이터 시스템 구축 방안

- 빅데이터 인프라 구축은 빅데이터 분석과제를 수행하기 위한 플랫폼으로 필요한 하드웨어 및 소프트웨어를 설치하고 구성해야 한다. 빅데이터 플랫폼은 데이터 수집, 데이터 저장 및 처리, 데이터 분석, 데이터활용, 데이터 보안 및 관리 영역으로 구성되고, 수행 기간은 주관부서 주도로 5-6개월 예상된다.
- 빅데이터 인프라 구축 추진 주용 내용은 ① 빅데이터 플랫폼 인프라(하드웨어 및 소프트웨어) 도입, ② 인프라 설치 및 구성, ③ 구성 테스트 수행이 주요 내용으로 소요예산은 내·외부 분석에 따른 컨설팅 절차가 필요하다.
- 빅데이터 플랫폼 운영을 위해서는 빅데이터 및 인프라 운영을 위한 각각의 담당자 지정이 필요하다. 빅데이터 운영 담당자는 데이터 수집 및 DBA 역할을 하고, 인프라 운영 담당자는 플랫폼내의 시스템 운영을 담당하도록 지정되어야 하고, 데이터 보안 및 작업관리와 데이터 마트 백업 방안도 신중히 고려해야 한다.
- 빅데이터 분석 플랫폼은 다양한 유형의 데이터를 활용할 수 있고, 데이터 활용 이전에 요구되던 시간을 줄이고 실시간으로 발생하는 데이터에 대해서도 분석할 수 있도록 구성되어야 한다.
- 전파감시 데이터의 특성 상 데이터 내 패턴 및 연관성을 분석하기 위해 인터랙티브 방식으로 탐색하는 디스커버리 분석 기술, 통계 및 마이닝 기법을 통한 모델링 작업, 패턴 분류 및 예측을 위한 예측 분석 기술을 권장한다.
- 향후에는 전파감시 과정에서 생성된 음성데이터를 텍스트로 변환 후 자연어 처리 기술을 적용하여 비정형 텍스트 데이터에서 특성을 추출하고 정형화하여 텍스트에 담긴 콘텐츠를 분석하기 위한 텍스트 분석 기술의 적용도 필요하다.

#### 4. 기대효과

- ◇ 전파감시 빅데이터 시스템 구축은 전파관리 정보서비스의 향상으로 이어져 전파관리정보의 고부가가치 창출 및 국민의 생명과 재산을 보호, 전파관리 산업의 발전을 견인할 수 있다.
- ◇ 기술(정책)적 측면에서는, 새로운 사회, 새로운 산업혁명의 도래로 전파환경이 급속도로 변화하는 상황에서 효율적인 전파이용을 위한 변화상 분석 및 대응방안 마련에 필요한 자료를 제공하고, 전파서비스의 지속적인 성장에 필요한 데이터를 창출하여 전파이용환경 개선에 필요한 기반을 마련할 수 있으며,
- ◇ 산업·경제적으로는 전파 이용 산업의 발전과 활성화에 기여할 수 있고, 전파 인지도 확산 및 전파서비스의 이용 확대로 이어져 경제적 상승효과를 기대할 수 있다.

# 목 차

제 1장 서론 .....	1
제1절 연구개요 .....	3
1. 연구 목적 .....	3
2. 연구 필요성 .....	3
제2절 빅데이터 분석 및 활용 .....	4
1. 빅데이터 개념 .....	4
2. 빅데이터 활용목적 .....	5
3. 빅데이터 활용분야 .....	6
4. 빅데이터 활용 기대효과 .....	7
제3절 빅데이터 사업 추진 단계별 점검 절차 .....	9
1. 요건 정의 .....	9
2. 데이터 수집 .....	15
3. 데이터 전처리 .....	18
4. 분석 모델링 .....	21
5. 모델 구현 .....	23
6. 시각화 .....	25
7. 검증 및 안정화 .....	29
제4절 내부 데이터 품질 확보방안 .....	32
1. 데이터 품질관리 .....	32

2. 데이터 품질관리 계획 단계 .....	33
3. 데이터 품질관리 구축 단계 .....	34
4. 데이터 품질관리 운영 단계 .....	36
5. 데이터 품질관리 활용 단계 .....	38
6. 데이터 품질진단 및 개선절차 .....	39
 제 2장 전파감시 빅데이터 현황 및 활용 .....	40
제1절 전파감시 데이터 현황분석 .....	40
1. 전파감시업무 개요 .....	40
2. 전파감시 데이터 현황 .....	41
3. 전파감시 데이터 분석 .....	42
 제2절 전파감시 빅데이터 활용 .....	43
1. 전파감시 데이터별 활용방안 .....	43
2. 전파감시 빅데이터 분석과제 선정 .....	45
3. 전파감시 빅데이터 분석과제 연관 데이터 .....	45
4. 전파감시 빅데이터 분석과제 연관 데이터 상세분석 .....	48
5. 전파감시 빅데이터 분석과제 연관 데이터 필드 연결 .....	53
 제 3장 분석모델 시각화 .....	56
제1절 시각화 사전 점검 사항 .....	56
1. 텍스트 기반 빅데이터 시각화 활용 .....	56

2. 빅데이터 시각화 TOOL .....	62
3. 빅데이터 시각화 TOOL 비교 .....	68
4. 빅데이터 시각화 라이브러리 구성 방안 .....	69
5. 빅데이터 시각화 표현 유형 .....	70
 제2절 『전파 혼·간섭 예방』분석모델 시각화 방안 .....	79
1. 전파감시 빅데이터 시각화 초기화면 .....	84
2. 지역별 전파감시 및 전파혼신 실적 .....	85
3. 불법·지정 외 의심 주파수 .....	88
 제 4장 빅데이터 분석가 양성 필요성 및 교육 방안 .....	90
제1절 빅데이터 분석 및 자격 .....	90
1. 빅데이터 분석기법 .....	90
2. 빅데이터 처리기술 .....	92
3. 빅데이터 분석가 자격분류 .....	94
4. 빅데이터 분석가가 되기 위해 필요한 기술 .....	96
5. 빅데이터 전문가가 되기 위해 필요한 기술 .....	96
 제2절 빅데이터 교육과정 운영 현황 .....	97
1. 데이터 산업 인력 현황 .....	97
2. 국외 교육과정 운영 현황 .....	99
3. 국내 빅데이터 전문인력 운영 현황 및 정책 .....	105

4. 빅데이터 전문인력 직무 분석 .....	115
5. 중앙전파관리소의 빅데이터 조직 구성 방안 .....	129
 제 5장 전파감시 빅데이터 시스템 구축 방안 .....	140
제1절 빅데이터 플랫폼 구축 방안 .....	140
1. 빅데이터 분석 시범사업 .....	140
2. 빅데이터 플랫폼 구축 .....	141
3. 연차별 사업 추진 방안 .....	145
 제2절 빅데이터 플랫폼 설계 .....	148
1. 빅데이터 플랫폼 설계 목적 및 범위 .....	148
2. 빅데이터 플랫폼 설계 절차 .....	148
3. 빅데이터 플랫폼 설계 접근 방안 .....	149
4. 빅데이터 레퍼런스 아키텍처 .....	150
5. 빅데이터 플랫폼 구성요소 정의 .....	152
 제3절 빅데이터 수집 플랫폼 구성 .....	153
1. 빅데이터 수집 플랫폼 개요 .....	153
2. 빅데이터 수집 플랫폼 구성 및 기능 .....	154
3. 빅데이터 수집 방안 .....	158
 제4절 빅데이터 저장 플랫폼 구성 .....	161



1. 빅데이터 저장 플랫폼 개요 .....	161
2. 빅데이터 저장 플랫폼 구성 및 기능 .....	162
3. 데이터 보안 및 작업관리 .....	164
4. 데이터 보안 및 작업관리 구성 방안 .....	167
5. ETL 및 데이터 정제방안 .....	173
 제5절 빅데이터 분석 플랫폼 구성 .....	176
1. 빅데이터 분석 프레임워크 .....	176
2. 빅데이터 분석 기술 .....	178
3. 빅데이터 분석 유형 .....	179
4. 빅데이터 분석영역의 특징 .....	186
5. 빅데이터 플랫폼 서비스 .....	189
6. 빅데이터 분석 솔루션 도입 .....	191
7. 빅데이터 분석 영역별 주요 솔루션 .....	192
 제 6장 결론 및 제언 .....	193
 참고자료 .....	197

## 표 목 차

표 1-1 빅데이터 기술과 DW/BI 기술의 차이점 .....	5
표 1-2 결과공유 항목 .....	28
표 1-3 품질관리목표 .....	32
표 1-4 품질관리 계획을 수립함에 있어 고려해야 할 주요 내용 .....	33
표 1-5 기관의 데이터 산출물 관리 .....	35
표 1-6 데이터 산출물 점검 및 최신성 확보 .....	37
표 1-7 데이터 활용성과 평가 .....	38
표 1-8 데이터 품질 진단 및 개선절차 .....	39
표 2-1 전파감시 업무 설명자료 .....	40
표 2-2 전파감시 데이터 설명자료 .....	41
표 2-3 전파감시 데이터 분석자료 .....	42
표 2-4 전파감시 데이터 활용방안 .....	43
표 2-5 “전파 혼·간섭 예방” 분석과제 연관 데이터 .....	45
표 3-1 빅데이터 시각화 도구 비교 .....	68
표 3-2 빅데이터 시각화 라이브러리 구성 방안 .....	69
표 3-3 네트워크 시각화 유형 .....	70
표 3-4 계층 시각화 유형 .....	71
표 3-5 통계적 분석 시각화 유형 .....	72
표 3-6 빈도 시각화 유형 .....	73
표 3-7 2차원 비교 시각화 유형 .....	73
표 3-8 입출력 시각화 유형 .....	74

표 3-9 다측정 시각화 유형 .....	75
표 3-10 상관분석 및 비율 시각화 유형 .....	76
표 3-11 지도 시각화 유형 .....	77
표 3-12 시간 시각화 유형 .....	77
표 4-1 데이터솔루션 산업 분류에 따른 정의 .....	94
표 4-2 EU의 인력양성 프로그램 .....	99
표 4-3 영국의 인력양성 프로그램 .....	99
표 4-4 독일의 인력양성 프로그램 .....	100
표 4-5 미국의 정부 주도 인력양성 프로그램 비교 .....	101
표 4-7 일본의 정부 주도 인력양성 프로그램 비교 .....	104
표 4-8 싱가포르의 인력양성 프로그램 비교 .....	104
표 4-9 빅데이터 관련 인력현황(2017년 기준) .....	106
표 4-10 정부주도 빅데이터 전문인력 양성 추진과제 .....	107
표 4-11 국내 데이터 관련 학부과정 개설 현황(일부) .....	109
표 4-12 국내 대학원 데이터 전문인력 양성 과정 현황(일부) .....	109
표 4-13 연세대학교 빅데이터학과의 재직자 대상 계약학과 교육과정 ....	110
표 4-14 국내 데이터 관련 학부 및 대학원 과정 개설 현황 .....	111
표 4-15 삼성 SDS 멀티캠퍼스의 빅데이터 아카데미 과정 .....	112
표 4-16 교육과정 세부내용 .....	113
표 4-17 데이터 직무 분류 .....	116
표 4-18 국가직무능력표준(NCS)의 빅데이터 관련 직무 .....	117
표 4-19 빅데이터 직무 능력 모형 .....	118
표 4-20 네덜란드 DATAFLOQ의 데이터 직무 구분 .....	119

표 4-21 국내외 데이터 직무 구분 비교표 .....	122
표 4-22 데이터 직무구분 및 필요 역량 .....	124
표 4-23 빅데이터 역량 정의 .....	126
표 4-24 빅데이터 조직 대안 비교 .....	129
표 5-1 빅데이터 플랫폼 구축 기본적인 H/W 사양 .....	143
표 5-2 빅데이터 플랫폼 구축 기본적인 S/W 사양 .....	143
표 5-3 TeraONE 확장성과 유연성을 보강한 기능 .....	151
표 5-4 빅데이터 수집 플랫폼 구성 및 기능 .....	154
표 5-5 리소스 관리 및 데이터 관리 .....	162
표 5-6 데이터 저장 및 처리 노드 .....	163
표 5-7 데이터 보완 및 작업관리 기능 설명 .....	165
표 5-8 데이터 마트 백업 방안 .....	170
표 5-9 네임노드 백업 방안-하둡 플랫폼 백업 .....	171
표 5-10 빅데이터 플랫폼 운영 인력 산정 .....	172
표 5-11 실시간 분석 기능 설명 .....	180
표 5-12 예측 분석 기능 설명 .....	181
표 5-13 의사결정 분석 기능 설명 .....	182
표 5-14 다차원 분석 기능 설명 .....	183
표 5-15 디스커버리 분석 기능 설명 .....	184
표 5-16 텍스트 분석 기능 설명 .....	184
표 5-17 빅데이터 분석 영역의 특징 .....	186
표 5-18 빅데이터 분석 상용 솔루션 특징 .....	191
표 5-19 빅데이터 분석 영역별 주요 솔루션 특징 .....	192

## 그 립 목 차

그림 1-1 빅데이터의 속성 3V+1V .....	4
그림 1-2 요건 정의 단계의 주요 활동 .....	9
그림 1-3 데이터 수집 단계의 주요 활동 .....	15
그림 1-4 데이터 전처리 단계의 주요 활동 .....	18
그림 1-5 분석모델링 단계의 주요 활동 .....	22
그림 1-6 모델구현 단계의 주요활동 .....	23
그림 1-7 시각화 단계의 주요 활동 .....	25
그림 1-9 검증 단계의 주요 활동 .....	29
그림 1-10 통합 테스트 절차 예시 .....	30
그림 1-11 데이터 품질관리 단계 .....	33
그림 2-1 쿼리 작업을 위한 공통 필드 .....	53
그림 2-2 쿼리 작업을 위한 내부 필드 .....	54
그림 2-3 가상 DB 필드 연결 .....	55
그림 3-1 빅데이터 시각화의 예 .....	57
그림 3-2 정보 시각화 절차 .....	58
그림 3-3 인포그래픽의 예 .....	59
그림 3-4 정보 디자인에서 빅데이터 시각화의 영역 .....	59
그림 3-5 빅데이터 시각화 단계 .....	60
그림 3-6 풍속, 풍향 데이터 시각화 예제 .....	61
그림 3-7 우리나라 시정 데이터 시각화 .....	61
그림 3-8 Candela Tools .....	62

그림 3-9 Datawrapper Tools .....	63
그림 3-10 Google Data Studio Tools .....	63
그림 3-11 Google Charts Tools .....	64
그림 3-12 MyHeatMap Tools .....	65
그림 3-13 RawGraphs Tools .....	66
그림 3-14 Dygraphs Tools .....	67
그림 3-15 시각화 라이브러리 구성 방안 .....	69
그림 3-16 네트워크 시각화 유형 .....	71
그림 3-17 계층 시각화 유형 .....	72
그림 3-18 통계적 분석 시각화 유형 .....	73
그림 3-19 빈도 시각화 유형 .....	73
그림 3-20 2차원 비교 시각화 유형 .....	74
그림 3-21 입출력 시각화 유형 .....	75
그림 3-22 다측정 시각화 유형 .....	76
그림 3-23 상관분석 및 비율 시각화 유형 .....	76
그림 3-24 지도 시각화 유형 .....	77
그림 3-25 시간 시각화 유형 .....	78
그림 3-26 전파품질측정 시각화 .....	80
그림 3-27 전파감시실적 및 위반실적 시각화(Ⅰ) .....	80
그림 3-28 전파감시실적 및 위반실적향 시각화(Ⅱ) .....	81
그림 3-29 전파품질측정 레이어 별 시각화 .....	81
그림 3-30 실시간 전파품질 및 전파혼신 상황 모니터링 시각화 .....	82

그림 3-31 데이터 변환(시트작업) .....	82
그림 3-32 데이터 변환(필드작업) .....	83
그림 3-33 데이터 변환(필드 시각화 적용 작업) .....	83
그림 4-1 데이터 과학자의 수요 .....	97
그림 4-2 데이터 산업 인력 현황 .....	98
그림 4-3 데이터산업의 데이터 직무 인력 환경 및 부족률 .....	98
그림 4-4 직무별 빅데이터 전문인력 현황 및 전망(전체산업) .....	106
그림 4-5 KMR 빅데이터 분석 전문가 양성과정 .....	112
그림 4-6 빅데이터 조직 설계 방향 .....	129
그림 4-7 빅데이터 조직 대안 설정 .....	131
그림 4-8 서울시 빅데이터 업무 운영 조직 .....	132
그림 4-9 경기도 빅데이터 업무 운영 조직 .....	132
그림 4-10 한국도로공사 빅데이터 업무 운영 조직 .....	133
그림 4-11 의사 결정 필요 주요 역할 .....	134
그림 4-12 빅데이터 추진 조직 부문 도출 .....	134
그림 4-13 빅데이터 운영 조직 직무구분 .....	135
그림 4-14 빅데이터 운영 조직 업무 .....	135
그림 4-15 소요인력 산정방법 .....	136
그림 4-16 소요 적정 인력 산정 .....	137
그림 4-17 빅데이터 운영 조직 구성 안 .....	138
그림 4-18 빅데이터 분석 인력 양성 방안 .....	138
그림 4-19 빅데이터 운영 인력 양성 방안 .....	139

그림 5-1 빅데이터 시스템 구축안(HW+SW) .....	142
그림 5-2 연차별 사업 추진 계획 .....	145
그림 5-3 1차 년도 예상 목표 구성도 .....	146
그림 5-4 2차 년도 예상 목표 구성도 .....	147
그림 5-6 플랫폼 4단계 구축방안 .....	148
그림 5-5 빅데이터 플랫폼 설계 절차 .....	148
그림 5-7 빅데이터 플랫폼 설계 접근 방안 .....	149
그림 5-8 TeraONE 엔터프라이즈 플랫폼 .....	150
그림 5-9 기관 빅데이터 플랫폼의 방향성 .....	151
그림 5-10 빅데이터 플랫폼 구성요소 정의 .....	152
그림 5-11 빅데이터 수집 플랫폼 구성 고려사항 .....	153
그림 5-12 빅데이터 수집 플랫폼 컴포넌트 .....	153
그림 5-13 DBMS 데이터 수집 방안 .....	158
그림 5-14 내부 데이터 API 수집 방안 .....	158
그림 5-15 공공 데이터 수집 방안 .....	159
그림 5-16 크롤링을 이용한 외부 데이터 수집 방안 .....	159
그림 5-17 외부 데이터 수집 방안 .....	160
그림 5-18 빅데이터 저장 플랫폼 구성 시 고려사항 .....	161
그림 5-19 공통 데이터 저장 및 관리 컴포넌트 .....	161
그림 5-20 빅데이터 저장 플랫폼 구성방안 .....	164
그림 5-21 데이터 보안 및 작업관리 컴포넌트 .....	164
그림 5-22 빅데이터 플랫폼 관리 고려 사항 .....	167



그림 5-23 빅데이터 플랫폼 보안 구성 고려 사항 .....	167
그림 5-24 빅데이터 플랫폼 보안 적용 방안 .....	168
그림 5-25 빅데이터 플랫폼 백업 대상 .....	168
그림 5-26 OS/어플리케이션 백업 방안 .....	169
그림 5-27 데이터 백업 수행 예시 .....	170
그림 5-28 DR 구축 방안 .....	172
그림 5-29 ETL 고려사항 및 구성 .....	173
그림 5-30 ETL 처리 방안 .....	174
그림 5-31 ETL 흐름도 .....	174
그림 5-32 6가지 품질 지표 .....	175
그림 5-33 데이터 오류 유형 별 정제 방안 .....	175
그림 5-34 분석 시스템 구성 .....	176
그림 5-35 빅데이터 접근 방법 .....	176
그림 5-36 빅데이터 분석 단계 .....	177
그림 5-37 빅데이터 분석 기술 .....	178
그림 5-38 빅데이터 분석 컴포넌트 .....	179
그림 5-39 데이터 및 분석 흐름도 .....	188
그림 5-40 빅데이터 플랫폼 서비스 영역 .....	189
그림 5-41 빅데이터 플랫폼 업무절차 .....	189

## 제 1장 서론

빅데이터 분석은 몇 년 전부터 클라우드 컴퓨팅 등과 함께 4차 산업혁명을 주도하는 핵심 기술로 주목을 받으며 등장하였으며 이제는 데이터를 활용한 분석을 대신하는 통상적인 표현으로 자리 잡았다. 그러나 빅데이터 분석이라는 표현이 빈번하게 사용되고 큰 기대를 받고 있음에도 불구하고 최근 실시된 한국정보화진흥원(:NIA)의 조사결과 등에 따르면 아직까지 국내에서의 빅데이터 활용 수준은 기대에 크게 미치지 못 하고 있는 상황이다. 이는 빅데이터 분석을 활용해 고객만족도를 크게 향상시키고 내부 업무의 운영 효율성을 극대화하고 있는 해외 기업들의 사례와 비교해 보면 4차 산업혁명 시대의 도래를 앞둔 시점에서 시급히 개선되어야 할 부분으로 판단된다. 이에 본 연구에서는 국내 빅데이터 산업 활성화에 필요한 인식과 문화, 제도 및 인프라 측면에서의 개선을 촉진하기 위한 방편으로, 4차 산업혁명 시대 핵심자원인 전과 데이터를 활용하여 효율적인 전과감시 방안에 대해 연구해 보고자 한다.

선진국의 애플, 아마존, 구글, 페이스북 등과 같은 국제적 선도 기업에서는 일찍이 데이터의 가치를 인식하고 데이터를 기업의 중요 자산으로 생각하여 데이터의 수집, 관리, 보관, 활용 등에 혈안이 되어 있다. 이러한 현상은 바이두, 텐센트, 알리바바 등이 데이터 가치화 경쟁에 뛰어 들면서 전통적 기업들과 비즈니스 경쟁력에서 점점 그 격차가 벌어지고 있다. 그러나 미래의 비즈니스 기반을 이루는 데이터의 가치에 대한 우리의 인식은 낙후되어 있어 향후 국내 기업의 경쟁력 및 나아가 국가 경쟁력이 상당히 염려되고 있는 실정이다. 이러한 상황에서 우리는 이미 확보하고 있는 중요한 데이터에 대해서도 그 가치를 인식하지 못함으로 인해 중요 자산을 사용하지 못하고 방치하고 있는 현실에 처해 있으며, 전과관련 데이터도 그 중요성에 비추어 마찬가지로 사용되지 못하고 있다.

전과감시를 통하여 이미 획득한 데이터의 양과 질은 이미 그 가치를 판단하기 어려운 만큼 중요하고, 이러한 전과 데이터의 양은

유비쿼터스, 사물인터넷, 무선자동차, 드론기술, 위성통신기술 등과 같은 무선통신기술의 발전과 더불어 앞으로 급속히 확대될 것으로 전망된다. 이와 같이 미래의 빅데이터 시대에 있어서 전파자원 데이터의 중요성은 더더욱 급속히 증대될 것으로 예상됨에 따라 전파데이터 관련 기관은 데이터 가치 인식을 제고하고 빅데이터 활용 방법을 습득하여야 할 필요가 있다고 생각된다. 따라서 본 연구과제에서는 이러한 필요성에 비추어 전파관련 기관에서 실제로 필요한 문제 중의 하나인 효율적인 전파감시 방안을 해결하기 위하여 빅데이터를 활용하는 방법으로 “빅데이터 사업 추진 단계별 점검 절차”를 제안한다.

1장에서는 빅데이터의 본질을 파악하여 우리 사회에 적합한 빅데이터 활용 영역과 활용방안을 모색하고, 전파 분야 빅데이터 분석과제의 추진 담당자가 추진절차를 실제 업무에 적용하기 위해서 점검해야 할 “빅데이터 사업 추진 단계별 점검 절차”를 제시하였다.

2장에서는 전파감시 분석과제의 선정을 위해 전파감시업무 현황을 파악하고, 전파감시 데이터 현황 및 업무에 따른 DB현황 상세 테이블을 분석하였다.

3장에서는 실시간 저장되는 대량의 데이터를 모두 해석하기에는 많은 제약이 따르기 때문에 데이터를 선별적으로 선택해서 짧은 시간에 해석하기 쉬운 형태로 시각화하는 데 가장 적합한 시각화 방법을 제안하였다.

4장에서는 중앙전파관리소의 업무에 빅데이터를 적용할 경우, 빅데이터 교육 방안과 운영인력 양성방안을 제안하였다.

5장에서는 전파감시의 빅데이터를 공통으로 활용할 수 있는 빅데이터 플랫폼을 설계하고, 활용 가능한 전파감시 빅데이터 시스템 구축 방안을 제안하였다

## 제1절 연구개요

### 1. 연구 목적

전과 분야 빅데이터의 다양한 가치에 대한 인식을 제고하고 효율적인 빅데이터 분석과 활용으로 국가, 지자체 및 공공기관 업무의 효과를 극대화하고, 빅데이터 분석을 통한 공공정책 의제발굴 · 정책결정 및 추진전략 수립의 방향을 이해하고 올바르게 적용할 수 있는 가이드를 제공하고자 한다.

전과 분야 빅데이터 분석과제의 추진 담당자가 추진절차를 실제 업무에 적용하기 위해서 점검해야 할 단계별 항목과 체크리스트를 제시하고, 실제 업무에서 담당자가 쉽게 활용할 수 있는 체계를 제시하고자 한다.

### 2. 연구 필요성

국가 안전 및 치안, 의료, 교육, 복지, 환경 등 사회 전반에 걸쳐서 빅데이터의 활용 가능성이 높아지고 실제 수요도 크게 증가하고 있다. 그러나 이를 수행하는데 참고할 지침이 마련되어 있지 않아 실무 담당자들이 실제 추진과정에서 불필요한 시행착오를 거쳐 추진될 우려가 있다.

전과 분야 빅데이터를 업무에 효과적으로 적용하기 위한 절차 및 방법 등에 대해 체계적으로 설명이 필요하고, 실무 담당자를 위하여 과제 목표와 방향 설정, 과제 추진 및 관리를 위한 체계적인 절차와 점검 항목 등을 제시할 필요가 있다.

## 제2절 빅데이터 분석 및 활용

### 1. 빅데이터 개념

빅데이터는 기존의 방법으로 처리가 어려운 대용량 혹은 다양한 유형의 실시간 생성되는 데이터의 집합을 의미한다. 빅데이터 분석은 시간, 공간 그 외 조건들의 변화에 따른 데이터의 분포와 데이터 간의 상호관계를 다양한 관점으로 조망하므로 의미 있는 패턴을 발견하고 세상을 이해해 가는 과정이다.

초창기 『일반적인 DB SW가 저장·관리·분석하는 범위를 초과하는 데이터』 등 규모를 강조하던 관점에서(McKinsey, 2011), 최근에는 『다양한 종류의 데이터로부터 저렴한 비용으로 가치를 추출하고 초고속 분석을 지원하는 기술』 등으로 가치창출 및 활용으로 초점이 이동되고 있다(IDC, 2011).



그림 1-1 빅데이터의 속성 3V+1V

빅데이터는 경제, 사회 전반에서 혁신을 주도하는 일종의 플랫폼 기술(GPT: General Purpose Technology)을 의미하며 효과적으로 활용시 정부·기업·개인 모두에 막대한 효용을 수반한다.

## 2. 빅데이터 활용목적

빅데이터를 분석하는 목적은 과거에 어떤 일이 일어났고, 그 원인이 무엇이며, 앞으로 어떤 일이 일어날 것인지를 예측하기 위함이다. 미래를 예측할 수 있다면 합리적인 의사결정을 통하여 기관 및 국민에게 엄청난 이득을 가져다줄 수 있으며, 과학적으로 데이터에 근거해서 의사결정을 하게 되면 문제와 결론이 명확한 합리적 의사결정을 할 수 있다.

기존 DW(: Data Warehouse), BI(: Business Intelligence), 통계 분석 등은 비싼 하드웨어, 소프트웨어에서 실행되며, 이 또한 데이터가 많아지면, 고가의 하드웨어를 추가해도 실행할 수 없다. 그래서 나온 것이 빅데이터 기술이며, 우선 빅데이터는 PC와 같이 상대적으로 저렴한 하드웨어와 공개 소프트웨어에서 운영되고, 많은 데이터를 처리할 경우에도 하드웨어만 계속해서 늘려 나가면 아무 문제없이 컴퓨팅 성능을 계속해서 올릴 수 있다. 결국 기존 RDBMS 기술의 고질적인 문제점을 빅데이터 소프트웨어가 해결했는데 이것이 바로 하둡(Hadoop), NoSQL DB이다.

표 1-1 빅데이터 기술과 DW/BI 기술의 차이점

	빅데이터 기술	DW/BI 기술
데이터	정형과 비정형 데이터 (SNS, 사진, 동영상, 위치 정보, 지리 정보, 기타)	정형 데이터
기본 기술	Hadoop, No SQL, CEP 등 대단히 다양	RDB에 근간
특징	수십, 수백억 건 이상도 처리 가능, HW만 늘려 가면 퍼포먼스가 증가함(Scale-out)	데이터가 수십억 건 이상 되면 HW를 더 많이 사용해도 퍼포먼스가 증가하지 않음
서버	저가의 일반 x86 서버 사용	주로 Unix(x86에 비해 고가)
소프트웨어	대체로 공개 소프트웨어 (물론 상용 소프트웨어도 있음)	모두 고가 상용 소프트웨어

데이터 처리	3개 노드에 데이터 중복해서 카피함, 1~2개의 노드에 문제가 동시에 있어도 자동 복구	데이터 중복을 안 하는 대신 시스템은 중복, 고비용
데이터 처리 방식	배치 처리(Hadoop), 실시간 데이터 처리(CEP 등), 다양한 분석 방식(NoSQL 등)	주로 배치 처리
참고사항	공개 소프트웨어이므로 기술 지원 문제, 기존 RDB와는 완전히 다른 기술	고가이고, 데이터 처리 양이 늘어난다면 한계에 이르며, 비정형 데이터는 처리하기 어려움
향후 트렌드	기존의 DW/BI에 빅데이터 기술을 도입하는 방향, 비정형 데이터를 처리해야 하는 분야에 부분적으로 도입 가능, DW/BI에 빅데이터 기술을 융합하는 형태	

### 3. 빅데이터 활용분야

#### 1) 맞춤형 서비스 분야

- (시민목소리 분석) 특정주제에 대한 시민의 목소리를 이해하고 그 추이를 분석함. 민원 접수 데이터와 소셜 데이터에 기반 할 수 있으며, 정책의제 발굴과 전략 방향 설정 등에 활용할 수 있다.
- (사회이슈 분석) 사회이슈의 자동 감지와 연관 주제의 동향 분석을 통해 잠재 정책 수요를 발굴함. 지역별 이슈를 도출하고 지역별 맞춤형 대국민 서비스 전략 수립이 가능함. 주요 일간지, 소셜 데이터, 민원 접수 데이터 등이 분석 대상이다.
- (맞춤형 민원서비스) 지역별, 기관별 민원인의 행동 및 요구사항 분석을 통해 민원 행정서비스의 개인 맞춤형 서비스를 구현함. 지역별 기관별 서비스 사용로그, 게시판 및 민원 접수 데이터, 포털게시판과 질의응답 데이터 등이 주요 분석 대상이다.

#### 2) 국민 안전 분야

- (범죄예방과 대응) 지역별, 시간별, 이벤트별, 유형별, 범죄 패턴 분석과 이를 활용한 지역별, 시기별 맞춤형 범죄예방 전략 도출

이 가능함. 범죄 및 경찰업무기록, 보고서, 뉴스 및 소셜 데이터 등이 주요 분석 대상이다.

- (도시관제 및 재난대응) 도심 내의 문제를 조기 파악하거나 재난을 조기에 감지하여 대응할 수 있도록 하며, 도시민과 재난지역의 시민 목소리를 바르게 이해하여 응급활동에 반영할 수 있는 정책의제 발굴에 활용한다. 센서 데이터, CCTV, 소셜 데이터 등이 주요 분석 대상이다.

#### 4. 빅데이터 활용 기대효과

공공 부문에서 축적된 빅데이터를 분석하고 활용하면 다양한 편익과 기대효과를 창출하게 되며, 특히 대국민 서비스 향상, 재난 관리, 안보, 범죄 예방, 부패 방지, 공공관리 등 공공적 가치 증진에 기여할 수 있다.

- (대국민 서비스 만족도) 개방과 협력을 통해 지역별 맞춤형 민원 서비스, 교육 환경 개선, 교통 환경 최적화 등에서 빅데이터 분석 결과가 활용 되므로 대국민 서비스의 질을 높이고, 서비스 만족도를 제고한다.
- (예산 절감) 정책 의제 발굴을 위한 비용과 시간을 절약할 수 있으며, 의료 복지 서비스, 국가 R&D 정책 입안, 교육 정책과 범죄 예방 등 거의 모든 부분에서 예산 절감 효과가 있음. 예를 들어, 조세 분야에서 빅데이터 분석 결과를 활용할 경우 탈루된 세금을 회수하는 등 상당한 예산절감 효과를 기대할 수 있다.
- (정책운영 효과성 향상) 빅데이터 분석을 통한 가시적인 정책효과 예측으로 국민의 수요에 부응하는 공공서비스를 제공함으로써 국가 정책 운영의 효과를 향상시킨다.
- (사회비용 절감) 환경, 교통 등의 분야에서 빅데이터를 활용하여 국민의 개인적인 비용 절감은 물론 사회 전체가 공동으로 부담해



야 하는 사회 비용을 절감할 수 있다.

- (합리적 정책 수립) 고부가가치 정보 활용으로 글로벌 환경변화에 대응하고 합리적 국가정책을 수립하고 집행한다.
- (위기대응 능력향상) 재난안전 분야에서 빅데이터의 활용은 정책권자의 신속한 의사결정을 지원할 뿐만 아니라, 국민의 위기 대응능력을 향상 시킬 수 있다.
- (업무효율성 제고) 빅데이터 분석을 통해 미래예측에 근거한 선제적 행정서비스를 제공함으로써 공공의 업무 효율성을 제고한다.
- (공공 기관과 국가 투명성 제고) 공공데이터 공개와 빅데이터 기반의 관리, 투명하고 근거 있는 내부감사 등은 기관과 국가의 투명성을 향상시키며 국제 경쟁력 확보와 지속적 성장을 견인한다.
- (정보 비대칭 해소) 공공 빅데이터 분석결과를 공공데이터포털, 빅데이터 포털 등을 통해 공개·공유함으로써 공공-민간, 중앙정부-지자체간 정보 비대칭 문제를 해소한다.
- (산업 활성화) 민간·공공의 융합지식을 민간에 개방·공유하여 다양한 비즈니스 창출 및 산업 활성화에 기여한다.

<b>※ 참고 - 범정부 빅데이터 추진체계</b>
범정부 빅데이터 활용 추진체계 및 인프라 업무는 크게 행정안전부(공공데이터정책과)와 과학기술정보통신부에서 담당한다. 행정안전부는 정부차원의 빅데이터 활용에 관한 전반적인 업무와, 대용량의 정형 또는 비정형의 데이터 셋에 관한 제도·정책 및 기본계획의 수립·시행, 빅데이터 기반 구축·운영 및 활용에 관한 업무를 총괄 한다. 과학기술정보통신부는 대학 및 민간차원의 빅데이터 활용에 관한 업무를 담당한다.

### 제3절 빅데이터 사업 추진 단계별 점검 절차

#### 1. 요건 정의

요건 정의 단계는 빅데이터 분석과 활용을 통해 해결하고자 하는 현안에 대해 명확히 이해하고 정의하며, 사업 추진을 위해 필요한 환경을 분석하고 사업의 추진 계획을 수립하는 과정을 포함한다.

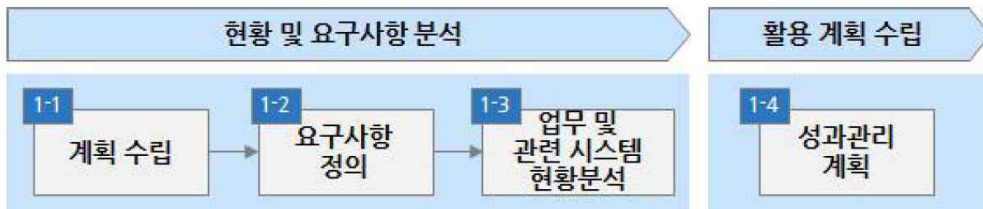


그림 1-2 요건 정의 단계의 주요 활동

##### 1) 계획 수립

계획 수립 단계에서는 사업의 전체 수행 일정계획을 수립하고, 요구사항을 정의하기 위한 실무자 및 관계자와의 인터뷰 계획을 수립한다.

#### ☞ 점검 사항

##### ◦ 일정 계획 수립

- 사업단이 제시한 일정계획(WBS)이 제안요청서 및 제안서에 기술된 내용과 일치하는가?
- 사업 추진 일정이 제안서의 내용과 불일치하는 경우, 기술협상 단계나 사업수행계획서 제출 이전에 사전 협의를 거쳤는가?
- 사업 수행 일정의 각 단계별로 투입하고자 하는 기술자의 등급과 투입 공수는 사업수행계획서 상의 투입 계획과 일치하는가?

##### ◦ 인터뷰 계획 수립

- 인터뷰 수행 대상은 사업목적 달성에 적합한 인원으로 도출되었나?
- 인터뷰 수행 대상에 업무 담당자와 IT 담당자가 고르게 분포되어 있나?
- 인터뷰 수행 일정은 사전에 인터뷰 대상자와 협의하여 확정하였는가?
- 인터뷰 질의서는 사업자가 문제의 현안을 파악하고 환경을 분석하는데 적합하게 도출되었나?

## 2) 요구사항 정의

요구사항 정의 단계에서는 사업의 기본적인 이해를 바탕으로 현업 및 IT 담당자와 인터뷰를 수행하고 현안을 분석하며 관련 자료를 수집, 정리하여 요구사항을 정의한다.

### ☞ 점검 사항

#### ◦ 요구사항 정의

- 제안요청서의 요구사항 내용을 모두 포함하는가?  
(사업계획서, 제안요청서 등 계획단계에서 제시된 목표와 추진 전략을 잘 이해하고 있는지 파악)  
(제안요청서의 요구사항들이 요구사항 정의서의 상세설명과 맞는지 파악)
- 사업 관련자의 추가 요구사항이 잘 반영되었나?  
(사업 관련자와의 인터뷰를 통한 추가 요구사항 등이 잘 반영되었는지 파악)
- 요구사항이 구체적으로 잘 정의되었는가?  
(요구사항들의 설명이 상세하게 되어 있는지 파악)  
(요구사항을 구현하는 방법이 적합한지 파악)  
(요구사항의 전제조건 및 제약사항이 있는지 파악)

(요구사항과 관련 산출물이 맞는지 파악)

- 요구사항 추적표는 향후 검수를 위한 항목들이 잘 관리되고 있나?

(요구사항정의서와 요구사항추적표의 내용이 일치하는지 파악)

(제안요청서 요구사항 ID 및 요구사항명 그리고 산출물과 구현내용이 일치하는지 파악)

#### ◦ 인터뷰 수행

- 인터뷰 계획서 또는 질의서의 내용을 모두 포함하여 인터뷰가 진행되었나?

- 예정된 인터뷰 대상자들과 이슈 없이 인터뷰가 진행되었나?

- 추가 인터뷰 일정이 필요한가?

(수행된 인터뷰 외에 추가 인터뷰가 필요한지 파악)

- 인터뷰 결과서는 사업의 이해와 요구사항 분석을 위해서 잘 작성되었나?

(수행된 인터뷰로 사업단이 사업의 전반적인 이해를 하였는지 파악)

(인터뷰가 요구사항들을 파악하기에 충분하였는지 파악)

(요구사항의 추가 및 변경 사항이 잘 반영되었는지 파악)

(요구사항의 세부항목들이 구체화 되었는지 파악)

#### ◦ 선행사업 분석 및 자료조사

- 과제 수행 시 필요한 자료는 잘 도출되고 활용되고 있나?

- 국내외 선행사례들을 연구하고 시사점을 도출하여 사업에 적용하였나?

(유사한 국내·외 선행 사례를 조사하고 필요한 내용이 잘 정리가 되었는지 파악)

(선행 사례 조사를 바탕으로 수행 사업에서 필요한 데이터, 보유기관, 분석모델링 방법이 정리되었는지 파악)

### 3) 업무 및 관련 시스템 현황 분석

시스템 현황 분석 단계에서는 앞으로 사업을 수행하기 위해 활용될 빅데이터 포털, 수집대상 데이터, 데이터를 수집할 환경의 현황을 분석하여 사업 준비를 시작하는 단계이다.

#### ☞ 점검 사항

##### ◦ 빅데이터 포털 분석

- 빅데이터 포털을 사용하여 사업을 수행한다면 그 범위와 대상은 명확히 구분 하였나?

(빅데이터 포털을 활용할 목적이 명확한지 파악)

(활용 범위가 구체적으로 어디까지인지 파악)

- 빅데이터 포털의 H/W와 S/W의 현황을 파악하였는가?

- 사전 준비 사항을 점검하였나?

(빅데이터 포털의 활용 범위에 따라 작업내용을 구체화 하였는지 파악)

(작업내용별 예상 소요 일정을 계획 하였는지 파악)

(빅데이터 포털의 개발환경 사용을 행정안전부와 협의하였는지 파악)

##### ◦ 수집 데이터 현황분석

- 분석 목표 달성을 위해 필요한 수집 대상 데이터 목록을 정의하였나?

(선행사례 조사 및 인터뷰를 통해 필요한 데이터 목록을 정리 하였는지 파악)

(수집 대상 데이터의 외부 데이터 보유기관을 조사하였는지 파악)

- 분석 목표 달성을 위한 분석 항목을 정의하였나?

(요구되는 분석 항목을 정의하고 각 항목별 데이터 현황을 파악하였는지 확인)

(다른 유관기관이 보유하고 있는 데이터와 병합할 방법을 정

의하였는지 파악

◦ 수집환경 분석

- 내부망 또는 외부망에서 데이터 수집이 가능한지 여부는 조사하였나?
- 외부망을 활용한다면 데이터 수집 루트를 정하였는가?
- 데이터 수집을 위한 개발환경 계획이 잘 되었는가?
- 데이터 수집을 위한 환경 분석이 완료되었는가?

4) 성과관리 계획

성과관리 계획 단계에서는 사업 수행 결과를 업무에 적용하여 평가할 기준 및 계획을 세우고, 결과를 실제 업무에 적용하기 위한 단계적 활용방안과 타 지역(또는 전국) 확산을 위한 방안을 계획한다.

☞ 점검 사항

◦ 성과관리 계획

- 사업목적에 적합한 성과목표를 세웠는가?  
(성과목표가 분석과제에 합당한지 검토)  
(분석과제에 대하여 성과목표가 현실성이 있는지 파악)
- 성과목표에 대하여 대상, 기간, 측정 항목들이 명확하고 구체적으로 작성되었는가?
- 성과 측정시기별 계획안이 마련되었는가?  
(측정 시기 간격이 성과를 측정하기에 적절한지 파악)  
(성과측정이 단기, 장기로 계획안이 마련되었는지 파악)

◦ 성과관리 지표 정의

- 성과관리 지표가 분석과제를 평가하기 위해 타당한가?  
(지표의 측정주체가 정의되어 있는지 파악)

(지표 측정 산식의 근거가 타당한지 검토)

(측정시기별 목표치 설정근거가 타당한지 검토)

- 성과관리 운영방안을 수립하였는가?

(성과점검의 기관별 역할이 명확한지 파악)

(기관들 간의 성과평가 범위를 정의 하였는지 파악)

(성과점검 및 성과평가의 절차가 세워졌는지 파악)

#### ◦ 성과관리 활용계획

- 성과지표의 단기/중기/장기 활용계획을 수립하였나?

- 과제 수행 결과 활용 계획을 합리적으로 세웠는가?

(활용계획의 목표가 현실성이 있는지 파악)

(활용계획 설정 근거가 타당한지 점검)

- 분석모형의 일반화, 확산 및 운영방안 수립 등의 활용방안을 제시하였는가?

(단기 및 중·장기 활용계획의 적용대상 업무 정의)

(과제 확산으로 얻을 수 있는 기대효과 명시)

(과제 확산 계획의 타당성 검토)

## 2. 데이터 수집

데이터 수집 단계에서는 분석 데이터를 정의하고 수집 환경을 구성하여 데이터를 수집한다.

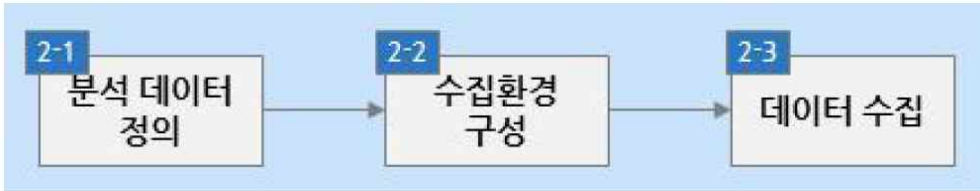


그림 1-3 데이터 수집 단계의 주요 활동

### 1) 분석 데이터 정의

분석 데이터 정의 단계에서는 분석 범위에 맞는 데이터 수집 대상 목록을 정리하고 보유기관 및 담당자를 조사한다.

#### ☞ 점검 사항

##### ◦ 필요 데이터 정의

- 분석 대상 범위 정의가 완료되었는가?  
(분석 대상 범위 정의가 요구사항을 만족 시키는지 점검)
- 데이터 수집 대상 목록을 작성하였는가?  
(사전 조사 및 인터뷰를 바탕으로 수집 대상 목록 작성)  
(추가 수집 대상 데이터가 없는지 파악)  
(데이터 종류, 양, 보관 방식, 수집주기와 분석주기, 백업과 이중화 등 원시(raw) 데이터에 대한 요구사항 도출 및 분석)
- 수집 대상 데이터 보유기관 및 담당자를 조사하였는가?
- 보유기관이 데이터를 제공하는 형식을 조사하였는가?
- 보유기관으로부터 데이터 메타정보(칼럼, 건수, 크기, 길이)를 입수하였는가?



◦ **데이터 확보 방안 수립 및 점검**

- 보유기관으로부터 데이터 확보 및 수집 방안을 세웠는가?
- 조사한 데이터는 사용 가능한 데이터인가?
- 조사한 데이터는 수집 가능한가?
- 데이터 수집 항목 중 민간 기업에서 수집되는 경우가 있다면, 데이터 수집 비용은 적절한가?

◦ **개인 정보에 대한 점검**

- 수집 예정 데이터에 개인정보가 포함되어 있는가?  
(개인정보 침해 가능성 검토를 하였는지 확인)
- 수집 예정 데이터에 개인정보가 포함되어 있다면, 유출될 문제에 대하여 방안을 세웠는가?  
(개인정보 데이터 제공 범위 협의)  
(개인정보 보안 방안 협의)

2) 수집 환경 구성

수집 환경 구성 단계에서는 기관이 보유한 수집 환경이나, 빅데이터 포털 인프라를 활용하여 데이터를 수집하고 이를 위한 환경을 구성한다.

☞ **점검 사항**

◦ **필요 데이터 정의**

- 분석 대상 범위 정의가 완료되었는가?  
(분석 대상 범위 정의가 요구사항을 만족 시키는지 점검)
- 데이터 수집 대상 목록을 작성하였는가?  
(사전 조사 및 인터뷰를 바탕으로 수집 대상 목록 작성)  
(추가 수집 대상 데이터가 없는지 파악)

(데이터 종류, 양, 보관 방식, 수집주기와 분석주기, 백업과 이  
중화 등 원시(Rawf) 데이터에 대한 요구사항 도출 및 분석)

- 수집 대상 데이터 보유기관 및 담당자를 조사하였는가?
- 보유기관이 데이터를 제공하는 형식을 조사하였는가?
- 보유기관으로부터 데이터 메타정보(칼럼, 건수, 크기, 길이)  
를 입수하였는가?

#### ◦ 데이터 확보 방안 수립 및 점검

- 보유기관으로부터 데이터 확보 및 수집 방안을 세웠는가?
- 조사한 데이터는 사용 가능한 데이터인가?
- 조사한 데이터는 수집 가능한가?

### 3) 데이터 수집

데이터 수집 단계에서는 수집 대상을 명확하게 정의하여 보유기  
관에 요청하여 수집한다.

#### ☞ 점검 사항

##### ◦ 데이터 수집

- 데이터를 수집하기 위한 보유기관과의 협의 일정을 세웠는가?
- 데이터 수집을 위해 보유기관에 보낼 협조 공문을 준비하였는가?
- 보유기관에 요청할 데이터 리스트는 구체적으로 작성하였는가?

(대상 데이터의 시간, 공간의 범위가 명확한지 파악)

(모든 보유기관에 요청 시 대상 데이터의 시간, 공간의 범위  
동일한지 파악)

(요청 데이터 목록을 정의하였는지 파악)

- 기관 내·외부 데이터 수집을 완료하였는가?
- 데이터 수집 시 주의사항

(데이터의 품질, 수집기술, 데이터 보안 및 개인정보보호 문제

등 고려할부분이 많으므로 전문가의 조언이 필요)  
(데이터 수집활동은 분석 결과의 품질을 좌우하는 중요한 과정이므로 분석에 필요한 데이터 항목들이 반드시 포함될 수 있도록 사전 점검 필요)

### 3. 데이터 전처리

데이터 전처리 단계는 수집된 데이터를 파악하여 데이터의 오류를 점검하고, 데이터를 정제함과 동시에 매칭 key 값을 기준으로 여러 데이터를 융합하고 기초 통계 분석을 통해 데이터의 고유 패턴을 파악한다.

분석 대상 업무가 1회성이 아닌, 지속적으로 서비스를 수행해야 하는 경우 빅데이터 포털이나 기관 자체 인프라를 활용하여 데이터 수집, 정제, 표준화, 융합이 자동화 되도록 시스템 연계를 추진하고, 모듈을 설계하고 구현한다.



그림 1-4 데이터 전처리 단계의 주요 활동

#### 1) 데이터 정제

데이터 정제 단계에서는 다양한 매체로부터 수집된 데이터를 기초 분석하여 오류가 있는지 파악하고, 원하는 형태로 변환하여 빅데이터화 한 후 저장하고 데이터를 활용할 수 있도록 품질을 확인하고 관리한다.

#### ☞ 점검 사항

##### ◦ 데이터 정제

- 데이터 기초 분석을 하였는가?
- 오류데이터 제거, 캐릭터 셋 통일, 길이 체크, 날짜 포맷 통일 등 Data 검증 및 정제 과정을 거쳤는가?
- 코드 데이터가 정의된 코드 값에 맞게 변환되었나?
- 데이터에 포함된 개인정보는 모두 비식별화 하였는가?
- 사용하는 모든 데이터가 개인정보보호법에 따른 제약사항을 준수하고 있나?

## 2) 데이터 저장

데이터 저장 단계에서는 데이터 구조를 표준화하고, 모든 데이터(테이블)를 융합(매칭)할 주요 key 값을 찾는다.

데이터를 융합하여 데이터의 정보 및 패턴을 파악하기 위해 기초 통계 분석을 하며, 빅데이터 포털이나 기관 자체 인프라를 활용하여 데이터 관리가 자동화되도록 한다.

#### ☞ 점검 사항

##### ◦ 데이터 표준화 및 융합

- 수집된 데이터의 속성(칼럼명, 유형(문자형, 숫자형 등), 크기 및 길이, 필수 여부 등) 정의가 되어 있는가?
- 분석에 필요한 데이터 목록과 유형이 정의 되고 변환 기준이 정의되었는가?
- 수집된 데이터는 구조를 통일하고 병합하는 표준화 과정을 거쳤는가?

(데이터 품질 측정)

◦ **데이터 관리 및 수집 자동화**

- 데이터를 추출-변환-탑재(ETL, Extract - Transform- Load) 하기 위한 변환 시스템은 정상 작동하나?
- 데이터 보유기관으로부터 데이터를 주기적으로 수집이 가능하도록 연계가 이루어졌는가?

◦ **데이터 기초 통계 분석**

- 정제된 데이터를 파악하기 위해 기초 통계(평균, 분산, 최소값, 최대값, 사분위수 등)를 추출하였는가?  
(추출된 기초 통계가 데이터 전체를 파악하기에 적합한지 검토)
- 데이터 현황 분석을 하였는가?  
(데이터를 시간별, 공간별 등 다른 시각으로 접근하여 분석하였는지 검토)  
(데이터의 고유한 패턴을 파악하였는지 검토)

3) 시스템 연계

시스템 연계 단계에서는 빅데이터 포털이나 기관 자체 인프라를 활용하여 데이터 정제 및 융합이 자동화 될 수 있도록 데이터 연계 개발을 수행한다.

☞ **점검 사항**

◦ **빅데이터 포털을 활용하는 경우**

- 빅데이터 포털의 활용 범위를 정하였는가?
- 주요 연계 데이터를 정하였는가?
- 빅데이터 포털에서 활용할 데이터 저장방식을 정하였는가?
- 빅데이터 포털의 전문분석 도구(SBP SRA, Tableau)를 활용하여 연계하는가?

- 데이터 연계 방식을 설계하였는가?
- 데이터 보완 관리에 대하여 고려하였는가?
- (데이터 접근 제어가 가능한지 점검)
- (사용자 인증 절차 기능이 있는지 점검)
- (분석 프로세스 단계별로 나오는 결과들이 외부의 불법적인 침입으로 유출 되지 않도록 조치가 취하여졌는지 확인)

#### 4) 모의 설계 구현

모듈 설계 및 구현 단계에서는 표준화 및 융합, 정제를 위한 환경을 분석하여 모듈을 설계하고 구현한다.

##### ☞ 점검 사항

- **표준화 모듈 설계 및 구현**
  - 표준화 모듈 설계를 하였는가?
  - 표준화 모듈 구현이 잘 되었는가?
- **데이터 정제 모듈 설계 및 구현**
  - 데이터 정제 모듈 설계를 하였는가?
  - 데이터 모듈 구현이 잘 되었는가?

#### 4. 분석 모델링

분석 모델링 단계에서는 분석 대상 및 범위를 정하여 분석 목적을 구현하기 위한 분석 방법론을 설계하는 단계이고, 분석모델 설계가 타당한지 자문위원회 회의를 거쳐 논의하고, 내용을 추가 반영하여 분석모델 설계를 완성 한다.

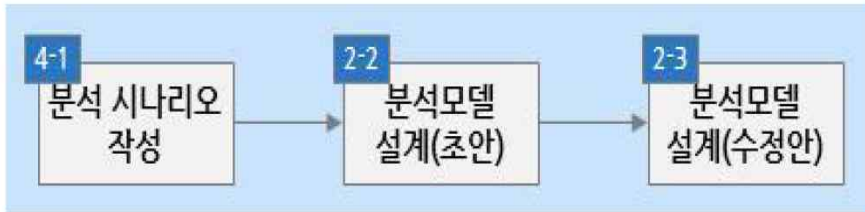


그림 1-5 분석모델링 단계의 주요 활동

#### ☞ 점검 사항

##### ◦ 분석시나리오 작성

- 분석 대상 및 범위가 요구사항에 맞게 정해졌는가?
- 분석에 필요한 데이터가 정의 되었는가?
- 분석 목표별 구현 모델(ex. 분류모델, 지수모델, 예측모델 등)을 정하였는가?

##### ◦ 분석모델 설계(초안)

- 필요한 데이터 항목이 정해졌는가?
- 데이터 항목별 표준화 방법을 정하였는가?  
(데이터의 단위를 고려하여 표준화하였는지 파악)
- 데이터 수집 가능 항목에 따라, 단계별로 모델이 설계 되었는가?
- 분석 검증 통계기법을 정하였는가?  
(검증방법이 타당한지 판단)

##### ◦ 분석모델 설계(수정안)

- 현업, 빅데이터 전문가, 실무자들로 구성된 자문위원단과 분석 모델 설계를 검토하였는가?
- 분석모델 설계(수정안)가 타당한가?

## 5. 모델 구현

모델 구현 단계에서는 분석모델링 설계를 알고리즘으로 구현하기 위해 알고리즘 모델링을 설계하고 모듈을 개발하여 분석 알고리즘을 구현한다.



그림 1-6 모델구현 단계의 주요활동

### 1) 분석 알고리즘 모델링

분석 알고리즘 모델링 단계에서는 분석모델링 설계에 따라 알고리즘을 설계하며 만약 분석모델링 설계에 ‘텍스트 마이닝’ 분석 기법을 활용 한다면 단어 사전 구축이 매우 중요한 단계로써 사전구축에 대하여 점검할 필요가 있다.

#### ☞ 점검 사항

##### ◦ 사전 구축

- 분석 대상의 언어(ex. 한글) 사전을 구축하였는가?
- 분석 목표가 되는 키워드(ex. 갈등 인지 단어)를 정의하였는가?

(키워드 사전으로 분석 목표를 달성할 수 있는지 검토)

- 분석 목적과 거리가 먼 명사인 불용어 사전을 구축하였는가?
- (불용어 사전 구축 전과 후의 분석 결과 비교)

##### ◦ 분석 알고리즘 모델링

- 분석모델에 대하여 조사하였는가?
- (분석모델에서 사용한 분석기법이 알고리즘화가 가능한지 점검)



- 분석 알고리즘 모델을 설계하였는가?  
(분석모델을 기반으로 알고리즘을 설계하였는지 확인)

## 2) 모듈 개발

모듈 개발 단계에서는 분석 알고리즘을 적용할 모듈을 정의하고 설계하며, 분석 알고리즘을 적용하여 동작에 문제가 없는지 확인한다.

### ☞ 점검 사항

#### ◦ 모듈 기능 정의 및 설계

- 모듈에 대한 기능을 정의하였는가?  
(기존에 존재하는 언어 사전이 분석에 활용이 가능한지 확인)  
(누락된 알고리즘이 없는지 점검)  
(중복되거나 누락된 모듈 기능이 없는지 검토)
- 모듈 설계를 실시하였는가?  
(H/W와 S/W등의 제반 상황을 고려하여 모듈 구축가능성을 검토)  
(분석모델 프로세스와 모듈설계안이 일치하는지 검토)

#### ◦ 모듈 개발

- 모듈 개발 결과물이 모델 설계와 일치하는가?
- 구축한 모듈이 정상적으로 동작하는가?

## 6. 시각화

시각화 단계에서는 분석 결과가 단순 수치가 아닌, 분석가 이외의 사람들이 이해하기 쉽도록 그림 및 도표를 활용하여 표현하며 수집한 데이터 및 분석 결과를 담당자 및 관련 기관들과 공유하기 위한 목적으로 필요하다면 Open-API 형태로 제공하도록 개발한다.

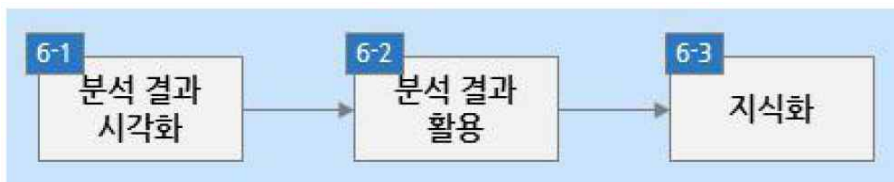


그림 1-7 시각화 단계의 주요 활동

### 1) 분석 결과 시각화

분석 결과 시각화 단계에서는 분석 결과를 다양한 관점에서 고찰하고 패턴을 쉽게 발견할 수 있도록 시각화를 설계하고 개발한다.

#### ☞ 점검 사항

##### ◦ 시각화 설계

- 시각화 설계 목차로 메뉴 구조도를 작성하였는가?
- 목차마다 화면이 설계되어 있는가?
- 화면 설계에 구체적인 설명이 되어 있는가?  
(화면 구성이 이해하기 쉽고 한 눈에 들어오는지 확인)
- 화면을 스토리 보드에 맞게 구성하였는가?  
(시각화에 서론, 본론, 결론과 같이 스토리가 담겨 있나 검토)
- 분석 결과가 잘 반영되었는가?  
(분석 결과 의미가 잘 표현되었는지 검토)  
(분석 대상의 기본정보를 제공하여 분석의 이해를 돕고 있는지 검토)  
(다양한 관점의 시각화 방식이 구현 가능한 설계인지 검토)

### ◦ 시각화 개발

- 분석 결과별로 시각화를 개발하였는가?  
(분석 결과를 통해 데이터를 바로 확인할 수 있도록 시각화되었는지 검토)  
(모든 분석 결과의 의미가 시각화로 쉽게 이해가 되는지 검토)
- 시각화 화면 설계 시 실무자의 의견이 반영되었는가?
- 데이터 분석 결과를 추출, 시각화 작동되는 단계까지 작동에 오류가 없는가?

## 2) 분석 결과 활용

분석 결과 활용 단계에서는 수집 데이터 및 분석 결과를 담당자 및 관련 기관들과 공유할 수 있도록 필요하다면 Open-API를 개발한다.

### ☞ 점검 사항

#### ◦ 활용 계획 수립

- 빅데이터 분석 및 활용에 대한 계획을 수립할 때 전체 조직 구성원들의 아이디어와 의견이 반영되었나?
- 분석 결과를 활용하고, 적용하고, 활성화 계획을 갖고 있나?  
(분석 결과의 단기 활용계획과 중장기 활용 계획 수립 및 상세화)  
(빅데이터 활용과 관련하여 단계적이고 구체적인 계획 수립)
- 분석종료 후 결과 활용을 위한 관련부서 협조와 예산확보는 가능한가?  
(수립한 빅데이터 분석 및 활용 방안은 기술적, 경제적으로 충분히 실현가능해야 함)

#### ◦ 분석 결과의 배포 및 추적

- 분석 완료된 결과를 리포트 형태로 구성하고 이를 유통, 배

포했는가?

◦ **Open-API**

- Open-API의 데이터 정의 및 규격을 설정하였는가?
- 데이터 수요자와 개방 방식을 정하였는가?

(데이터 수요자의 범위를 실무 담당자, 관련기관, 일반 사람들 등 접근 권한을 조정할 수 있는 기능이 존재 하는지 확인

(Open-API의 개방 방식이 수요자에게 편리한 방식인지 검토)

- 데이터의 공공성을 판단하여 공공기관과 대국민에게 개방할 데이터를 구분하였는가?

◦ **테스트**

- 시각화 및 Open-API 테스트 시나리오를 수립하고 테스트를 실시하였는가?

3) 분석 결과 공유

결과공유 단계에서는 향후 타기관이 유사한 분석과제를 수행할 때, 빅데이터 포털을 활용할 수 있도록 분석 과정과 결과를 공유한다.

☞ **점검 사항**

◦ **결과공유**

- 결과공유 대상과 범위(분석 과정, 시각화 결과 등)를 정하였는가?
- 빅데이터 포털을 활용하여 결과공유를 하였는가?
- 결과공유에 분석 수행 과제의 주제 및 목적이 잘 드러났는가?
- 분석과제의 데이터 수집>정제>표준화>융합>분석>검증 단계에 맞춰 설명이 잘 되어 있는가?
- 원시데이터와 분석대상 데이터의 샘플 및 속성 정보를 제공하였는가?

- 빅데이터 포털을 활용하여 분석 프로세스 절차대로 수행하였는가?
- 분석 과제의 결과공유를 위해 빅데이터 포털에 과제를 등록할 때 다음과 같은 항목들을 입력하여 빅데이터 담당자들이 참조하고 활용할 수 있도록 한다.

표 1-2 결과공유 항목

빅데이터 포털에 과제 등록 시 입력 내용			
입력 항목	<ul style="list-style-type: none"> <li>• 과제명, 과제 필요성 및 목표, 과제 설명</li> <li>• 전제조건 및 제약사항</li> <li>• 활용 데이터</li> <li>• 분석 기법 및 시각화 결과</li> <li>• 분석결과 설명 및 시사점</li> <li>• 활용 계획</li> </ul>	업로드 파일	<ul style="list-style-type: none"> <li>• 분석결과</li> <li>• 시각화 샘플</li> <li>• 결과 보고서</li> <li>• 샘플데이터</li> </ul>

## 7. 검증 및 안정화

검증 및 안정화 단계에서는 사업 전반에 대한 결과물(시각화, 분석 프로세스 자동화 등)을 통합 테스트하여 검증하고, 사용자 테스트 후 검수를 거쳐 분석 결과를 사용할 실무자를 위하여 교육을 실시하고 인수인계를 한다.

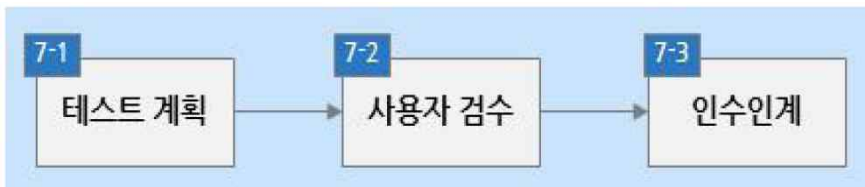


그림 1-9 검증 단계의 주요 활동

### 1) 테스트 계획

테스트 계획 단계에서는 프로젝트 추진과 관련된 테스트 활동 중 통합 테스트 계획을 세워 사업의 통합테스트 요구사항과 목표, 테스트 범위, 절차, 요구 되는 자원, 일정 등에 대한 계획을 수립한다.

#### ☞ 점검 사항

##### ◦ 테스트 계획 수립

- 테스트 대상 및 범위를 선정하였는가?
- 테스트 대상 및 범위가 사업 결과물(자동화 시스템 및 시각화 등)을 판단하는 데 적합한가?
- 테스트 유형별로 점검항목을 구체화하였는가?
- 테스트 항목을 단계적으로 정하였는가?
- 테스트 예상 결과를 도출하였는가?
- 테스트 수행 주체 및 참여자를 선정하였는가?

## 2) 사용자 검수

사용자 검수 단계에서는 테스트 계획에 따라 통합테스트를 실시하여 프로젝트를 점검하고 검수한다.

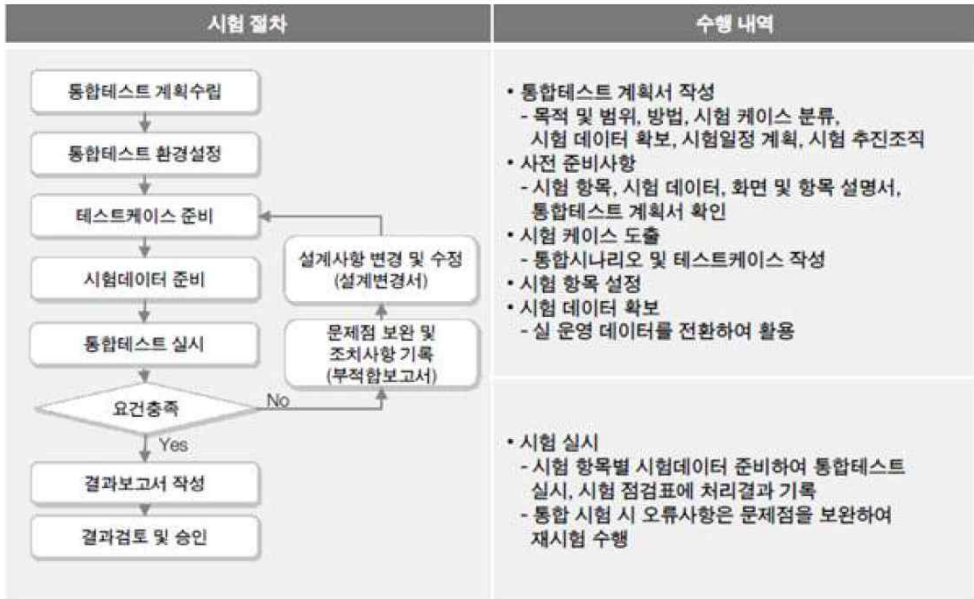


그림 1-10 통합 테스트 절차 예시

### ☞ 점검 사항

#### ◦ 통합 테스트

- 테스트 계획에 따라 진행하였는가?
- 테스트 결과가 예상 결과와 동일한가, 다르다면 그 원인을 찾고 해결하였는가?
- 분석결과 검증을 위해 관련 전문가와 동행한 현장 실사를 진행하였나?

(예를 들어, CCTV 사각지대의 경우, 분석 결과 사각지대로 파악된 지점을 직접 찾아가 실사하고 분석 결과와 실재가 일치하는지 검증 필요)

◦ 사용자 승인 테스트

- 사용자가 통합 테스트에 참석하였는가?
- 혹은 사용자를 위한 시연을 실시하였는가?

3) 인수인계

인수인계 단계에서는 실제 업무에서 분석 결과를 활용하게 될 실무자를 위한 교육을 실시한다.

☞ 점검 사항

◦ 인수인계

- 사용자를 위한 시스템(분석 자동화 및 시각화 화면) 사용자 지침서를 작성하였는가?

(분석 프로세스는 단계별 설명이 충분한지 검토)

(알고리즘 단계별 주석의 설명이 이해하기 쉬운지 검토)

(시각화 화면의 구성에 대한 설명이 모두 되어 있는지 검토)

(시스템 작동 단계별 매뉴얼이 구성되어 있는지 검토)

- 사용자 및 운영자를 위한 교육을 계획하고 실시하였는가?

(교육 대상자가 적합한지 검토)

(인수인계를 위해 교육 내용이 충분한지 검토)

(교육 대상자의 참석률이 높았는지 확인)



## 제4절 내부 데이터 품질 확보방안

### 1. 데이터 품질관리

데이터 진단을 통해 도출된 이슈 및 시사점에 따라 기관이 보유한 데이터의 활용도 향상을 위해서는 기관 데이터의 품질향상이 필요하며 품질 확보를 위한 방안으로 표 5-1의 품질관리목표에 맞춰 계획, 구축, 운영, 활용 단계별 품질활동이 필요하다.

표 1-3 품질관리목표

구분	품질요소	품질기준	검사내용
데이터 품질	완전성	개별 완전성	필수컬럼의 값 누락여부 확인
	유효성	범위 유효성	컬럼 값이 주어진 범위 내에 존재하는지 확인
	일관성	형식 표준화	날짜, 시간, 등의 정보가 동일 형식으로 표준화 되었는지 확인
	보오나성	오류 조치율	오류 및 누락 데이터의 조치 및 처리 여부 확인

#### 1) 데이터 품질관리 단계

- ① **(품질관리 계획수립)** 데이터를 구축하거나, 보유 데이터를 운영 및 활용하는 단계별 주요 품질관리 활동을 반영한 품질관리 계획을 수립한다.
- ② **(구축단계 품질관리)** 데이터를 구축 시 준수해야 할 데이터 표준의 적용·점검, 데이터 구조의 일관성 확보, 데이터관리 산출물의 품질확보 등을 수행 한다.
- ③ **(운영단계 품질관리)** 보유하고 있는 데이터를 운영하는 단계에서 준수해야 할 품질진단·개선, 연계데이터 정합성 유지, 주요 산출물의 현행화 점검 등을 수행한다.
- ④ **(활용단계 품질관리)** 데이터를 활용하면서 발생하는 품질오류 신고 및 차기 품질계획에 반영하기 위한 데이터 활용성과 등을 평가한다.

• 중앙전파관리소 데이터 품질관리(Plan-Do-Check-Act)



그림 1-11 데이터 품질관리 단계

## 2. 데이터 품질관리 계획 단계

매년 정보화 계획을 수립하는 단계에서 고려해야 할 데이터 품질 측면의 계획을 수립하는 활동을 의미한다.

### 1) 데이터 품질관리

표 1-4 품질관리 계획을 수립함에 있어 고려해야 할 주요 내용

구분	주요 내용
중앙전파관리소 차원의 데이터 품질관리 목표 수립	<ul style="list-style-type: none"> <li>국가 및 기관의 품질관리 정책을 일관성 있게 추진하기 위한 기관 차원의 중장기 품질관리 목표와 로드맵 수립</li> <li>품질 목표는 기관의 중장기 정보화 계획 등에 반영하여 지속적으로 관리</li> </ul>
중점 품질관리 대상 데이터베이스 선정	<ul style="list-style-type: none"> <li>기관 차원의 품질관리 목표와 부합되는 중점 품질관리 대상을 선정하여 품질관리 계획에 반영</li> <li>3가지 선정 기준(기관의 미션 수행관점, 서비스 제공 관점, 품질 이슈 관점)을 고려하여 선정하고 지속적으로 유지, 관리</li> </ul>
품질 진단 및 개선 계획 수립	<ul style="list-style-type: none"> <li>중점 품질관리 대상 데이터베이스 중에서 당해 연도에 추진할 품질 진단 및 개선 대상 선정</li> <li>품질 진단 계획과 품질 개선 계획으로 구분하여 품질관리 계획에 반영</li> </ul>
표준화 적용 계획 수립	<ul style="list-style-type: none"> <li>중점 품질관리 대상 이외에 신규 구축되는 데이터베이스를 포함하여 표준화 적용 대상을 선정 후 품질 관리 계획</li> </ul>

	<p>에 반영</p> <ul style="list-style-type: none"> <li>표준화 적용 계획 수립을 위해서는 중앙전파관리소 차원의 데이터 표준화 정책 및 방향, 표준화 범위 등에 대한 정립 및 표준화 수준에 대한 현황 분석이 선행되어야 함</li> </ul>
연계 데이터 품질 확보 계획 수립	<ul style="list-style-type: none"> <li>연계 데이터 품질확보를 위해 기관 차원의 연계 데이터 현황 및 품질 이슈에 대한 분석이 선행되어야 함</li> <li>중앙전파관리소가 보유, 운영 중인 데이터베이스 이외에 신규 구축 데이터베이스도 포함해서 계획에 반영 필요</li> <li>데이터 연계 기관간 사전 협의를 통한 연계 타당성 및 연계 데이터의 품질에 대한 검토 필요</li> </ul>
품질관리 산출물 점검 및 최신성 확보 계획 수립	<ul style="list-style-type: none"> <li>중점 품질관리 대상 데이터베이스 중 산출물 관리(산출물 점검 및 최신성 확보) 대상을 선정하여 품질관리 계획에 반영</li> <li>대상 선정 시 데이터 변경 요인에 대한 식별을 통해 점검 및 최신성 확보 대상 데이터베이스를 선정하고, 점검에 따른 현재까지의 진행사항이 반영된 산출물을 이해관계자에게 통보하는 것이 중요</li> <li>이를 위해서는, 중앙전파관리소 차원의 품질관리 점검 대상 산출물 목록이 사전에 정의되어 있어야 함</li> </ul>

### 3. 데이터 품질관리 구축 단계

정보시스템 신규 도입 또는 고도화 등을 추진하기 위한 구축사업 단계에서 고려해야 할 데이터 품질관리 활동을 의미한다.

#### 1) 데이터 품질관리(데이터 표준화)

데이터 품질을 보장하기 위한 기본 활동으로써, 기관이 보유하고 있는 정보시스템에 구축되어 있는 데이터 정보 요소의 명칭, 정의, 형식, 규칙에 대한 원칙을 수립하여 이를 전사적으로 적용하고, 지속적으로 관리하는 제반 활동을 의미한다.

기관차원에서 사용하는 용어의 의미와 형식에 규칙을 정해서 의사소통의 혼란을 방지하고, 데이터의 정확성과 일관성을 유지하여 궁극적으로 고품질의 데이터를 확보하는데 목적이 있다.

## 2) 데이터 품질관리(데이터 산출물 관리)

정보시스템의 신규 도입 및 시스템 고도화 등의 구축단계에서 생성되는 산출물들 중 데이터 품질과 관련된 산출물을 생성하고 검증하며, 현재까지의 진행사항을 반영하는 일련의 활동을 의미한다.

구축단계에서 관리해야 할 품질관리 대상 산출물 목록을 기본으로 하고, 기관의 여건에 맞는 품질관리 대상 산출물을 정의하여 관리할 수 있다.

표 1-5 기관의 데이터 산출물 관리

산출물명		필수	권장	기관 전체	개별 DB
데이터 표준	데이터사전 정의서	○		○	○
	코드 정의서	○		○	○
	도메인 정의서	○		○	○
논리 데이터 요소	엔티티 정의서	○			○
	속성 정의서	○			○
	논리데이터 모델 다이어그램	○			○
물리 데이터 요소	테이블 정의서	○			○
	컬럼 정의서	○			○
	물리 데이터모델 다이어그램	○			○
	데이터베이스 정의서	○			○
연계 데이터 목록		○		○	○
오너쉽 정의서			○		○
업무 규칙 정의서			○		○
테이블 대 응용 프로그램 상관도			○		○

#### 4. 데이터 품질관리 운영 단계

기관이 생성·보유·활용하고 있는 데이터를 운영함에 있어 데이터의 품질수준을 향상시키기 위한 제반 활동을 의미한다.

##### 1) 기관의 데이터 품질관리(연계 데이터의 품질관리)

데이터 보유기관과 이를 활용하는 기관과의 연계데이터에 대한 표준화된 관리 및 최신성 확보를 통해 연계 데이터의 정합성을 확보하고, 기관 내 연계 정보 관리(기관 내 시스템간의 연계정보에 대한 관리)와 기관 간 연계 정보 관리(보유기관과 활용기관 사이의 연계 정보에 대한 관리)를 의미한다.

##### 2) 기관의 데이터 품질관리(데이터 품질 진단 및 개선)

기관 내부적으로 보유하고 있는 데이터와 데이터를 관리하는 체계에 대한 전반적인 진단을 통해 원인분석 및 개선사항 도출하고 이를 토대로 실제 개선활동을 수행한다.

##### 3) 기관의 데이터 품질관리(데이터 산출물 점검 및 최신성 확보)

운영단계에서 실시되는 데이터의 품질 개선 및 데이터 변경 등에 따른 관련 산출물 등 문서의 최신성 확보 및 관련 이해관계자들에게 변경사항을 통지하는 것을 포함하고, 정기적 비정기적 데이터 산출물의 최신성 확보 수준을 점검하여 수정 보완하는 일련의 작업까지도 포함한다.

표 1-6 데이터 산출물 점검 및 최신성 확보

단계	정의
진단대상 정의	품질 이슈에 대한 수요 및 현황으로 조사하여 품질 진단 대상 데이터베이스를 선정하고, 진단 방향성을 정의
품질진단 실시	품질 진단 대상에 대한 상세 수준의 품질 진단 계획 수립 후 품질 진단 영역별 진단을 실시
진단결과 분석	오류 원인 분석, 업무 영향도 분석을 통해 개선과제를 정의
개선계획 수립	품질 개선 관제별 개선 방향 정의 및 개선 추진을 위한 추진 계획을 수립
개선 수행	상세 품질개선 계획 수립 및 개선 영역별 품질 개선 실시
품질 통제	목표 대비 결과 분석, 평가를 통한 품질관리 목표 재설정 및 지속적 품질 통제 수행

## 5. 데이터 품질관리 활용 단계

“활용단계 품질관리”란 기관이 생성·보유하고 있는 데이터를 활용함에 있어 발생하는 품질 이슈를 인지하여 개선 활동으로 연계하는 품질오류 신고관리와 품질 개선 결과를 포함하는 데이터의 활용 성과를 평가하여, 계획단계의 품질 목표 재조정 등에 반영하는 일련의 활동으로 구성한다.

### 1) 기관의 데이터 품질관리

기관에서 제공하는 데이터 및 정보에 대한 품질을 유지하기 위해 정보를 활용하는 사용자들이 정보를 활용하는 과정에서 발견하는 다양한 오류를 기관에서 알 수 있도록 기록, 관리 기관의 대외 서비스 신뢰도 향상 및 유지한다.

### 2) 기관의 데이터 품질관리(데이터 활용성과 평가)

내·외부 사용자의 활용 목적을 고려하여 데이터의 활용성과를 평가한다.

표 1-7 데이터 활용성과 평가

구분	데이터 활용성과 평가
제공정보의 품질 수준을 평가	데이터 품질 진단 킴 개선에 사용되는 품질지표 등을 활용하여 데이터 정확성, 유효성, 일관성 등을 평가
사용자 만족도 평가	성과 평가의 대상이 되는 데이터의 주요 사용자를 식별하여, 전반적인 데이터 활용도에 대한 서비스 만족도를 조사하여 평가
보유정보 연계 대상 기관의 수 등을 평가	기관이 보유하고 있는 데이터가 연계활용이 많은 데이터 베이스인 경우, 보유 정보를 연계해서 활용하는 기관의 수 또는 증가율 등을 이용하여 활용성과를 평가
공공정보 공개 대상 목록 및 제공, 활용건수 등	활용을 위해서는 공공정보의 개방 및 공유가 적극적으로 이루어져야 하므로 개방, 공유 관점에서 기관이 보유하고 있는 데이터 얼마나 많이 공개되고 활용될 수 있는지를 평가

## 6. 데이터 품질진단 및 개선절차

데이터 품질진단 및 개선은 총 4단계의 절차를 거쳐 진행하도록 한다.

표 1-8 데이터 품질 진단 및 개선절차

진단 대상 선정	진단 방법 정의	진단 수행	진단 결과 분석 및 관리
<ul style="list-style-type: none"> <li>· 데이터품질 관리를 수행을 위한 데이터품질 관리 수행 계획 수립</li> <li>· 품질 진단 대상에 대한 기준 정의</li> <li>· 데이터품질 관리를 위한 계획 수립 및 품질관리 대상 데이터 범위 확정</li> </ul>	<ul style="list-style-type: none"> <li>· 정의된 업무규칙 등을 이용하여 품질 진단 계획 수립</li> <li>· 기 정의된 품질 진단 기준에 따른 데이터 품질 수준 측정</li> <li>· 측정계획 수립, 품질 측정, 오류 데이터 리포팅</li> </ul>	<ul style="list-style-type: none"> <li>· 오류의 근본원인 파악</li> <li>· 측정 결과에 대한 현상검토 및 오류유형 검토를 기반으로 주요 품질 대상 선정 및 데이터 구조 분석</li> <li>· 오류원인 분석 및 조치계획 수립</li> </ul>	<ul style="list-style-type: none"> <li>· 품질 개선 대상 선정</li> <li>· 개선 방안 수립</li> <li>· 데이터 정제 및 프로세스 개선 수행</li> <li>· 개선안 도출</li> <li>· 개선 위험요소 및 영향도 평가</li> <li>· 개선대상 선정 및 품질개선 수행</li> </ul>
기타	<ul style="list-style-type: none"> <li>· 진단대상: 보유(수집 및 가공, 제공)하고 있는 모든 데이터</li> <li>· 진단주기: 정기적인 검사와 담당자 요청 또는 데이터 오류 발생 시 수행하는 특별검사(비정기적)</li> <li>· 진단 대상 선정 및 진단 방법 정의 후 구체적인 데이터 품질 확보방안을 제시하도록 함</li> </ul>		



## 제 2장 전파감시 빅데이터 현황 및 활용

### 제1절 전파감시 데이터 현황분석

#### 1. 전파감시업무 개요

주파수 9kHz~275GHz 범위내의 전파감시가 가능한 주파수를 대상으로 하여 전파관련 법령의 준수여부 및 각종 전파이용에 관한 자료조사 업무 등을 수행하고 있다.

표 2-1 전파감시 업무 설명자료

구 분	주요 내용
전파품질감시	<ul style="list-style-type: none"> <li>· On-air 상의 전파품질을 측정하여 기술기준 위반 여부를 판단</li> <li>- 무선국에서 발사되는 전파에 대한 주파수의 허용편차, 점유 주파수 대역폭, 스퓨리어스 발사강도 등을 측정</li> </ul>
불법전파 탐사	<ul style="list-style-type: none"> <li>· 불법주파수 및 불법무선국 등 불법전파 탐사를 수행하며, 불법무선국 조사단속 업무와 연계하여 처리, 불법무선국의 방향 탐지와 무선국 확인을 위해 필요한 경우 운용감시(교신 내용 청취) 등을 병행하여 수행</li> </ul>
혼신 조사	<ul style="list-style-type: none"> <li>· 전파 혼신 민원 접수 시, 혼신 신호의 측정 및 신호원의 방향 탐지를 통해 전파 혼신의 발생 원인을 규명 및 제거</li> </ul>
불법무선국 및 변칙운용 무선국 단속	<ul style="list-style-type: none"> <li>· 현장 조사단속을 통해 불법무선국 및 변칙운용 무선국 색출</li> </ul>
전파의 방향 탐지	<ul style="list-style-type: none"> <li>· 조난통신, 긴급통신 등 중요통신, 불법전파, 방해전파, 혼신 전파를 대상으로 전파 송신원의 방향을 탐지하여 위치 추적 및 제거</li> </ul>
이동전파감시	<ul style="list-style-type: none"> <li>· 고정감시가 미치지 못하는 지역을 대상으로 전파이용질서 확립 및 각종 전파 자료 조사</li> </ul>

## 2. 전파감시 데이터 현황

표 2-2 전파감시 데이터 설명자료

데이터 구분 (중)	데이터 구분 (소)	설명
국내전파 감시	전파품질측정(편이, 편차, 점유대역폭 등)	허가·신고무선국에 대한 기술기준 위반여부 측정
	불법전파탐색(주파수)	허가대역에서 허가·신고 없이 불법으로 사용하는 무선국과 허가대역 외에서 출현하는 불법주파수를 감시
	주파수이용량조사(주파수별 이용효율%)	매년 조사대상 주파수를 정하여 매월 2주간 이용률을 조사
	고속스펙트럼측정(광대역 전파스펙트럼)	고속으로 넓은 주파수 대역의 전파스펙트럼을 측정
	스푸리어스(불요파, 고조파)	측정·특정 주파수의 불요파 및 제2,3,4,5차 고조파를 측정
	전파 잡음분석	측정된 전파스펙트럼을 정해진 잡음분석용 주파수로 분석해 전파 잡음값 추출
국제단파 감시	국제단파측정	ITU의 IFIC 주파수 및 HFBC 등록 주파수에 대해 국내출현여부 및 위반여부 확인을 위해 측정
	스펙트럼점유율측정	일정한 간격(step)으로 국제단파감시 주파수 대역에 대한 hit여부 및 점유율을 측정
해상·항공 감시	전파품질측정, 불법전파탐색, 고속스펙트럼, 주파수이용률, 고조파, 마스크, 잡음조사	해상·항공 중요주파수에 대해 혼신발생여부 위주로 감시
GNSS 감시	GPS(미국), GLONASS(러시아), BeiDou(중국), Galileo(유럽) 위성신호 감시 데이터	주로 GPS 대역의 전파 혼신여부를 감시
위성전파 감시	위성주파수의 대역폭, 레벨, PFD, EIRP 위반여부 등	ITU에 등록된 국내외 정지위성, 비정지위성에 대한 위반여부 및 우리나라 위성의 혼신영향 여부 등 감시

### 3. 전파감시 데이터 분석

표 2-3 전파감시 데이터 분석자료

업무구분	부서별 서비스 운영 내용	데이터 구분	생성 데이터	데이터 저장소	데이터 저장형태
전파품질측정 (편이, 편차, 점유대역폭 등)	허가·신고무선국에 대한 기술기준 위반 여부 측정	측정자료	전파품질 DB 참고	지소 DB 및 통합 DB	DB
		스펙트럼	파일(확장자 : ref)	로컬PC	파일(.ref)
		스냅샷	파일(확장자 : jpg)	로컬PC	이미지 파일(.jpg)
		음성	파일(확장자 : wav)	음성서버	음성 파일(.wav)
불법전파탐색 (주파수)	허가대역에서 허가·신고 없이 불법으로 사용하는 무선국과 허가대역 외에서 출현하는 불법주파수를 감시	측정자료	불법탐사 DB 참고	지소 DB 및 통합 DB	DB
		스펙트럼	파일(확장자 : bin)	로컬PC	파일(.bin)
		스냅샷	파일(확장자 : jpg)	로컬PC	이미지 파일(.jpg)
		음성	파일(확장자 : wav)	음성서버	음성 파일(.wav)
주파수이용량조사 (주파수별 이용효율%)	매년 조사대상 주파수를 정하여 매월 2주간 이용률을 조사	측정자료	주파수이용률 DB 참고	지소 DB 및 통합 DB	DB
		스냅샷	파일(확장자 : jpg)	로컬PC	이미지 파일(.jpg)
고속스펙트럼측정 (광대역 전파스펙트럼)	고속으로 넓은 주파수 대역의 전파스펙트럼을 측정	측정자료	고속스펙트럼 DB 참고	지소 DB 및 통합 DB	DB
		스펙트럼	파일(확장자 : ref)	로컬PC	파일(.ref)
		스냅샷	파일(확장자 : jpg)	로컬PC	이미지 파일(.jpg)
스퓨리어스 (불요파, 고조파)	측정·특정 주파수의 불요파 및 제 2,3,4,5차 고조파를 측정	측정자료	고조파 DB 참고	지소 DB 및 통합 DB	DB
		스냅샷	파일(확장자 : jpg)	로컬PC	이미지 파일(.jpg)
전파 잡음분석	측정된 전파스펙트럼을 정해진 잡음분석용 주파수로 분석해 전파 잡음값 추출	측정자료	잡음조사 DB 참고	지소 DB 및 통합 DB	DB

## 제2절 전파감시 빅데이터 활용

### 1. 전파감시 데이터별 활용방안

표 2-4 전파감시 데이터 활용방안

구 분	데이터 출처	데이터명	전파감시 데이터 활용	
			기 준	활 용(안)
① 측 정 데 이 터	1세대, 2세대 전파감시시스 템,  국내단파 감시시스템	· 전파품질측정	· 전파품질위반 확인	· 무선국 검사 데이터와 비교·분석하여 고정감시국소의 On-Air 측정 시 장비 환경설정 기준 및 측정 데이터 신뢰도 확보 · 정상 신호의 누적스펙트럼 파형을 저장·분석하여 이상신호(혼신) 발생 즉시 파악 ※ 정상적인 신호의 샘플을 저장하여 머신러닝(ML) 분석을 통한 이상신호 자동탐지
		· 불법전파	· 불법전파여부 확인	· 불법전파의 출현시간, 신호레벨, 출현방향 등의 데이터를 누적 분석하여 불법전파 이용패턴 파악
		· 주파수이용현황	· 주파수 회수·재배치	· 지역 내 허가주파수에 대한 점 유율, 신호레벨 분석정보를 무선국 허가 신청자에게 제공
		· 고속스펙트럼	· 불법주파수 확인	· 스펙트럼 데이터를 장기간 누적하여 전파환경 변화 파악 · 지역 내 허가주파수에 대한 점 유율, 신호레벨 분석정보를 무 선국 허가 신청자에게 제공
		· 스푸리어스	· 불요파 발생여부 확인	· 지역 내 허가주파수에 대한 스 푸리어스 데이터를 분석하여 중요통신 주파수에 대한 영향 여부와 시간경과에 따른 변화 파악
		· 전파 잡음	· 주파수이용현황조사 의 기준레벨 및 전 파관리맵의 잡음지 도에 활용	· 장기간 누적·분석하여 측정지 점에 대한 전파환경변화 파악
		· 음성	· 불법전파 사용자 확인	· 음성->텍스트 변환 후 데이터를 분석하여 불법전파의 사용자와

				사용 장소를 찾아내고, 이동시간 및 이동경로 예측
	고정방탐 시스템	· 전파방향탐지	· 불법전파 출현방향 확인	<ul style="list-style-type: none"> <li>· 불법전파 측정데이터와 연계하여 불법전파 이동경로 파악</li> <li>※ 감시국소 주변의 모든 이동경로 상에서 모의신호를 발사하여 신호세기과 방탐각도 데이터를 저장하여 머신러닝(AI)으로 분석</li> </ul>
	국제단파 감시시스템	· 국제단파감시	· 국제단파 위반 확인	· 국내 단파 허가 주파수의 방송 시간과 국제 유입전파의 방송 시간을 분석하여 국내 방송 주파수에 대한 혼신 영향여부 확인
		· 음성	· 국제단파 사용언어 확인	· 음성->번역->텍스트 변환 후 데이터를 분석하여 방송내용 확인
	해상·항공 감시시스템	· 전파품질측정	· 해상·항공 중요 주파수의 혼신 발생여부 확인	<ul style="list-style-type: none"> <li>· 정상 신호파형을 저장·분석하여 이상신호(혼신) 발생 즉시 파악</li> <li>※ 정상적인 신호의 샘플을 저장하여 머신러닝(AI) 분석을 통한 이상신호 자동탐지</li> <li>· 주파수의 측정데이터를 관련 기관에 주기적으로 제공</li> </ul>
② 실 적 데 이 터	방송통신 통합정보 시스템	<ul style="list-style-type: none"> <li>· 기술기준위반</li> <li>· 허가사항위반</li> <li>· 불법무선국</li> <li>· 전파혼신실적</li> <li>· 조난통신실적</li> <li>· 전파환경조사</li> </ul>	· 실적 통계작성	· 위반 유형별로 패턴을 분석하여 시간경과에 따른 위반패턴의 변화추이를 시각화하여 제공
③ 업 무	인접국 유입전파 조사	· 유입현황(주파수, 전계강도)	· 국제 전파혼신 대응	<ul style="list-style-type: none"> <li>· 인접국 유입전파의 유입현황과 기상정보(날씨, 강우, 습도, 온도), 해양정보(수온, 파고, 염분 등), 전리층 상태정보 등을 종합 분석하여 인접국 전파가 유입되는 환경조건 분석</li> <li>· 국내 방송 주파수와 중복되는 유입전파의 측정 데이터를 국내 방송국에 제공</li> </ul>

## 2. 전파감시 빅데이터 분석과제 선정

무선국의 출력증강, 고조파 발생 등에 의한 전파 혼·간섭의 예방을 위해서는 전파감시 빅데이터 분석을 통한 선제적인 대응이 필요할 것으로 판단됨에 따라 본 연구과제에서는 “전파 혼·간섭 예방”을 빅데이터 분석과제로 선정하였다.

### 1) 분석과제 선정

- 전파감시 빅데이터 분석과제명 : “전파 혼·간섭 예방”

### 2) 분석과제 선정에 따른 기대효과

- 무선국 출력증가 및 고조파, 지정 외 주파수 운용 등으로 인한 전파 혼·간섭을 사전 파악하여 예방조치가 가능하다.
- 주파수 이용현황(점유율) 신뢰도 제고
- 무선국의 허가지역 외 불법 운용여부 파악
- 시스템 장애 발생 전 예방조치 가능

## 3. 전파감시 빅데이터 분석과제 연관 데이터

표 2-5 “전파 혼·간섭 예방” 분석과제 연관 데이터

연관 데이터	데이터 속성	비 고
전파품질측정 데이터	주파수, 채널파워, 점 유대역폭, 편이, 편차	<ul style="list-style-type: none"> <li>· 무선국의 송신출력 조정여부 파악</li> <li>· 제2~제5 고조파의 점유대역폭 확인</li> <li>· 전파품질 체크로 시스템 고장여부 파악</li> <li>· 주파수 이용효율조사 채널을 측정하여 최적의 주파수이용효율조사 임계레벨 파악</li> </ul>
고조파측정 데이터	주파수, 기본파 채널파워, 제2~제5고조파 채널파워	<ul style="list-style-type: none"> <li>· 제2~제5 고조파의 채널파워 확인</li> </ul>

주파수이용 효율 데이터	주파수, 대역폭, 측정 시간	· 전파품질측정 데이터와 비교
무선국 허가DB	주파수, 허가번호, 대 역폭, 설치장소	· 고조파와 중첩되는 허가 무선국 확인 · KCA 검사 무선국 설치장소 확인
무선국 검사이력	허가번호, 시설자명, 판정, 검사종류, 검사 일자, 차기검사일자, 설치장소	· 무선국의 검사일자, 검사결과 등 확인
KCA 검사결과	허가번호, 시설자명, 검사일자, 검사종류, 검사결과	
전파 감시실적	주파수, 시설자 주소, 적발일자, 조치일자, 위규명, 적발시작시간	· 전파품질(전계강도, 편이, 편차, 대역폭) 변화원인 분석
전파 혼신실적	주파수, 혼신·피혼신원 주소, 접수일자, 접수 시간, 종료시간, 혼신 원인, 혼신 시설자명, 피혼신 시설자명, 제 거일자	
장비 고장· 수리내역	고장일시, 복구일시, 고장이력, 조치내용	· 장비 고장·수리 시점의 전파품질 상태를 분석하여 장비 이상여부를 예측
시스템 자기진단 Log	장비 접속상태 및 장 에이력	· 시스템 자기진단 Log 및 고장·수리내역, 전파품질측정값과의 연관성을 파악하여 장비 이상여부를 예측

### 1) 전파품질측정데이터

지역별 허가 주파수, 주파수이용효율조사 주파수, 20W 이상의 무선국(방송국, 기지국, 고정국)의 제2~제5 고조파 주파수를 대상으로 주간 2시간, 야간 2시간 이상 전파품질(전계강도, 점유대역폭, 편이, 편차) 측정해서 무선국의 전계강도 변화를 파악하여 방송국, 고정국의 검사 전·후 출력 조정여부 파악(평균 전계강도가 5%를 초과 또는 감소할 경우 운용자에게 통보)하고, 20W 이상의 무선국(방송국,

기지국, 고정국)의 제2~제5 고조파의 점유대역폭을 측정하여 점유대역폭 내에 허가된 주파수가 있는지 파악한다.

전파품질 측정결과의 변화와 고장·수리내역, 자기진단 Log와 연계하여 장비이상 시 전파품질의 패턴변화를 파악하고, 신호가 없을 때 전파품질모드에서 측정된 잡음레벨을 주파수이용효율조사 시, 기준레벨로 활용하고자 한다.

## 2) 고조파측정 데이터

20W 이상의 무선국(방송국, 기지국, 고정국)의 제2~제5 고조파를 대상으로 주간 2시간, 야간 2시간 이상 채널파워 측정하고, 고조파의 채널파워가  $20\text{dB}\mu\text{V}$  이상 측정된 경우, 전파품질측정결과에서 점유대역폭을 확인하여 점유대역폭 내에 허가된 주파수가 있는지 확인하고 허가주파수가 있는 경우 운용자에게 통보한다.

## 3) 주파수이용효율 데이터

주파수이용효율 조사결과, 허가되지 않는 지역에서 주파수 이용효율이 조사되는 경우 불법전파로 보고 운용자에게 통보한다.

## 4) 무선국 허가DB

$20\text{dB}\mu\text{V}$  이상의 제2~제5 고조파가 허가 주파수와 중첩되는지 파악하고, KCA 무선국 검사결과의 허가번호와 연계하여 설치장소를 파악한다.

## 5) 무선국 검사이력, KCA 검사결과

무선국의 검사일자를 기준으로 송신출력 조정이 이루어지는지 파악한다.

## 6) 장비 고장·수리내역, 시스템 자기진단 Log

장비 고장·수리내역을 시스템에 등록하고, 시스템이 기본적으로 갖추고 있는 자기진단을 통해 쌓인 LOG 데이터를 전파품질측정 데이터(전계강도, 편이, 편차 등) 변화와 비교하여 연관성 파악한다.



#### 4. 전파감시 빅데이터 분석과제 연관 데이터 상세분석

◇ 허가데이터	분석 과제 필요 데이터	◇ 전파감시실적 데이터	분석 과제 필요 데이터
<ul style="list-style-type: none"> <li>- 무선국 허가DB</li> <li>- 검사일정관리</li> <li>- 검사결과이력</li> <li>- 유선방송검색</li> <li>- 전파응용설비</li> <li>- 진흥원검사결과</li> <li>- 무선기기조회</li> </ul>	<ul style="list-style-type: none"> <li>○</li> <li>×</li> <li>○</li> <li>×</li> <li>×</li> <li>×</li> <li>×</li> </ul>	<ul style="list-style-type: none"> <li>- 전파기술기준위반</li> <li>- 허가사항위반</li> <li>- 무선국품질관리</li> <li>- 불법무선국</li> <li>- 허가신고제외무선국위반</li> <li>- 전파혼신실적</li> <li>- 전파환경조사처리</li> <li>- 누설전자파조사</li> <li>- TV방송수신장애</li> <li>- 불법방송통신기자재</li> <li>- 불법감청설비</li> <li>- 불법스팸</li> <li>- 통신보안준수지도</li> <li>- 통신보안교육지도</li> </ul>	<ul style="list-style-type: none"> <li>○</li> <li>○</li> <li>×</li> <li>○</li> <li>○</li> <li>○</li> <li>×</li> <li>×</li> <li>×</li> <li>×</li> <li>×</li> <li>×</li> <li>×</li> <li>×</li> </ul>
◇ 장애(고장)데이터	분석 과제 필요 데이터		
<ul style="list-style-type: none"> <li>- 시스템 장애(고장) 이력</li> </ul>	○		

DB 연결 가능

DB 연결 가능

DB 연결 가능

◇ 인증데이터	분석 과제 필요 데이터	◇ 측정데이터	분석 과제 필요 데이터
<ul style="list-style-type: none"> <li>- 인증현황</li> <li>- 세관장확인물품현황</li> </ul>	<ul style="list-style-type: none"> <li>×</li> <li>×</li> </ul>	<ul style="list-style-type: none"> <li>- 전파품질 환경 설정 : TRFJ_MRENV_MST</li> <li>- 전파품질 결과 : TRFJ_MRRSLT_DTL</li> <li>- 불법전파 환경설정 : TRFJ_MIENV_MST</li> <li>- 불법전파 결과 : TRFJ_MIRSLT_DTL</li> <li>- 고속스펙트럼 환경 설정 : TRFJ_MEENV_MST</li> <li>- 고속스펙트럼 결과 : TRFJ_MERSLT_DTL</li> <li>- 음성파일 환경 : TRFJ_MFENV_MST</li> <li>- 음성파일 결과 : TRFJ_MFRSLT_DTL</li> <li>- 주파수 이용률 측정 상위 정보 : TRFJ_MLENV_MST</li> <li>- 주파수이용률 주기별 결과 : TRFJ_MLRSLT_DTL</li> <li>- 고조파 환경 설정 : TRFJ_MHENV_MST</li> <li>- 고조파 측정 결과 : TRFJ_MHRSLT_DTL</li> <li>- 마스크 측정 환경 설정 : TRFJ_MSENV_MST</li> <li>- 마스크측정 인접채널 정보 : TRFJ_MSENV_DTL</li> <li>- 마스크측정 측정 결과 : TRFJ_MSRSLT_DTL</li> </ul>	<ul style="list-style-type: none"> <li>×</li> <li>○</li> <li>×</li> <li>×</li> <li>×</li> <li>×</li> <li>×</li> <li>×</li> <li>×</li> <li>○</li> <li>×</li> <li>×</li> <li>×</li> <li>×</li> <li>×</li> <li>×</li> <li>×</li> </ul>

DB 연결 가능

DB 연결 가능

※ “장애(고장) 데이터”(추가 기능개선 요청 필요) 및 “전파잡음 측정결과”는 현재 구축 중인 3세대 전파감시시스템에서 확보

# 1) 허가데이터

## - 무선국 허가 DB조회

허가번호	무선국구분	허가상태	무선(방송) 국명	국종	주파수	단위	주파수대역 별	변조방식 1
[연관Data] Text	Text	Text	Text	[연관Data] Text	[연관Data] Text	[연관Data] Text	Number	Text
기기명칭 (형식/명칭)	장치내역	통신방식	통신상대방	제조사	기기일련번 호(제조번 호)	지정용도	공중선명	편파
Text	Text	Text	Text	Text	Text	Text	Text	Text
이득	해발고	지상고(안 테나설치높 이)	전력	단위	전파형식 (영상)	전파형식 (음성)	출력용도	통신사항
Number	Number	Number	Number	Text	Text	Text	Text	Text
목적사항	주민/사업 자등록번호	시설자명	시설자우편 번호	시설자주소	시설자도로 명주소	시설자전화 번호	시설자FAX	운영순위
Text	Text	Text	Text	[연관Data] Text	Text	Text	Text	Text
호출명칭	호출부호	최초허가일 자	허가일자	유효기간	준공기한	허가변경일 자	관리관서	허가관서
Text	Text	Date	[연관Data] Date	Text	Date	Date	Text	Text
검사기관	허가담당자	이동채(선 박)길이	선박구분	항해구역	주정박항	선박톤수	이동범위	방송종류
Text	Text	Text	Text	Text	Text	Text	Text	Text
시설종류	방송구역	연허세통보	처리일자	부관사항	고지주소우 편번호	고지주소	폐지일자	휴지기간
Text	Text	Text	Text	Text	Number	Text	[연관Data] Date	Date
여객정원	무선국명							
Text	[연관Data] Text							

## 2) 검사데이터

### - 검사결과(이력) 조회

허가번호	시설자명	호출부호	호출명칭	검사일련번호	검사일자	검사주기	차기검사일자	무선국명
[연관Data] Text	[연관Data] Text	Text	Text	[연관Data] Number	[연관Data] Date	Number	[연관Data] Date	[연관Data] Text
식별번호	무선국종별	설치장소	검사종류	무선국상태	판정	종합평가	납입방법	검사수수료
Text	Text	[연관Data] Text	[연관Data] Text	Text	[연관Data] Text	Text	Text	Number
검사관	관리관서							
Text	[연관Data] Text							

## 3) 전파감시실적 데이터

### - 전파기술기준위반

상태	적발관서	관리번호	조치구분	적발일자	허가번호	시설자명	무선국명	시설자코드
Text	[연관Data] Text	[연관Data] Text	Text	[연관Data] Text	Text	Text	Text	Text
시설자주소 (지번)	시설자주소 (도로명)	시설자구분	국종명	위규명	조치결과	조치일자	출력여부	주파수
Text	Text	Text	Text	[연관Data] Text	Text	[연관Data] Date	Text	[연관Data] Text
적발시작시간	적발종료시간	적발자번호	적발자명	적발구분	적발좌석	적발방법	적발지역	업종
[연관Data] Number	Number	Text	Text	Text	Text	Text	[연관Data] Text	Text
문서제목	교신내용	위규여부	의뢰수신관 서명	최초등록자	최초등록자 명	입력일자	마지막수정 자	마지막수정 자명
Text	Text	Text	Text	Text	Text	Date	Text	Text
마지막수정 일자	SMS송부							
Date	Text							

- 허가사항위반

상태	적발관서	관리번호	조치구분	적발일자	허가번호	시설자명	호출명칭	시설자주소 (지번)
Text	[연관Data] Text	[연관Data] Text	Text	[연관Data] Date	Text	Text	Text	Text
시설자주소 (도로명)	시설자구분	국종명	위규명	조치결과	과태료	조치일자	주파수	적발지역
Text	Text	Text	[연관Data] Text	Text	Text	[연관Data] Date	[연관Data] Text	Text
적발시작시 간	적발종료시 간	적발자번호	적발자명	적발좌석	적발방법	적발지역	업종	제작회사
[연관Data] Number	Number	Text	Text	Text	Text	[연관Data] Text	Text	Text
기기모델명	문서제목	교신내용	가중위규여 부	의회수신관 서명	최초등록자	최초등록자 명	입력일자	마지막수정 자
Text	Text	Text	Text	Text	Text	Text	Date	Text
최초등록자 명	마지막수정 일자	SMS송부						
Text	Date	Text						

- 불법무선국

상태	적발관서	관리번호	조치구분	적발일자	시설자명	시설자주소 (지번)	시설자주소 (도로명)	주인/사업 자번호
Text	[연관Data] Text	[연관Data] Text	Text	[연관Data] Text	Text	Text	Text	Text
국종	주파수	불법구분	적발무선국 수	조치무선국 수	조치결과	적발좌석	적발방법	조치일자
Text	[연관Data] Text	Text	Number	Number	Text	Text	Text	[연관Data] Date
입력일자	적발시작시 간	적발종료시 간	적발자번호	적발자명	전파형식	처리기간	전파법재범 여부	통신사실확 인자료요청 유무
Date	[연관Data] Number	Number	Text	Text	Text	Number	Text	Text
위규명	적발구분	시설자구분	업종구분	사용용도	사용동기	적발장소	제작회사	기기모델명
[연관Data] Text	Text	Text	Text	Text	Text	Text	Text	Text
기기명칭	관련문서	과태료	벌금	교신내용	의회수신관 서	최초등록자	최초등록자 명	마지막수정 자
Text	Text	Text	Text	Text	Text	Text	Text	Text

마지막수정 자명	마지막수정 일자	적발지역
Text	Date	[연관Data] Text

- 허가신고제외무선국위반

상태	적발관서	관리번호	조치구분	적발일자	시설자명	시설자주소 (지번)	시설자주소 (도로명)	주민/사업 자번호
Text	[연관Data] Text	[연관Data] Text	Text	[연관Data] Date	Text	[연관Data] Text	[연관Data] Text	Text
주파수	국종	위규분류	위규명	조치결과	적발좌석	적발방법	조치일자	입력일자
[연관Data] Text	Text	Text	[연관Data] Text	Text	Text	Text	[연관Data] Date	Date
적발시작시 간	적발종료시 간	적발자번호	적발자명	업종구분	시설처구분	적발구분	업무구분	관련문서
[연관Data] Number	Number	Text	Text	Text	Text	Text	Text	Text
참고사항	제조회사	기기모델명	최초등록자	최초등록자 명	마지막수정 자	마지막수정 자명	마지막수정 일자	
Text	Text	Text	Text	Text	Text	Text	Date	

※ 허가신고제외무선국위반에는 “적발지역” 데이터가 없어 “시설자주소”로 대체

- 전파혼신실적

상태	접수번호	접수관서	접수일자	접수시간	시설자명 (피혼신)	시설자명 (혼신)	혼신원인	혼신등급
Text	[연관Data] Text	[연관Data] Text	[연관Data] Date	[연관Data] Number	[연관Data] Text	[연관Data] Text	[연관Data] Text	Text
조치결과	주파수	출동일시	종료일시	민원처리시 간	접수방법	출동구분	조사자	민원처리답 변
Text	[연관Data] Number	Date	[연관Data] Date	Number	Text	Text	Text	Text
사용장비명	모델명	혼신제거일 자	제거방안	기타민원	비고	민원요지	혼신원(혼 신원 근거)	처리결과 (관계법령)
Text	Text	Number	Text	Text	Text	Text	Text	Text
피시설자번 호	피무선국명	피전파형식	피전화번호	피우편번호	피주소1	피주소2	피허가일자	피시설자분 류
Text	Text	Text	Text	Text	[연관Data] Text	[연관Data] Text	Date	Text
혼신유형	최초등록자	최초등록자 명	입력일자	마지막수정 자	마지막수정 자명	마지막수정 일자	혼신국 설 치장소	
Text	Text	Text	Date	Text	Text	Date	[연관Data] Text	

## 1) 쿼리 작업을 위한 공통 필드 연결

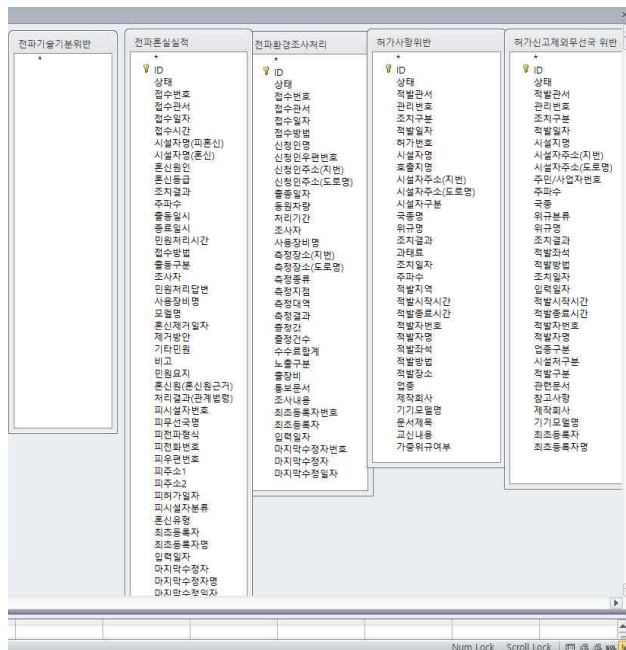
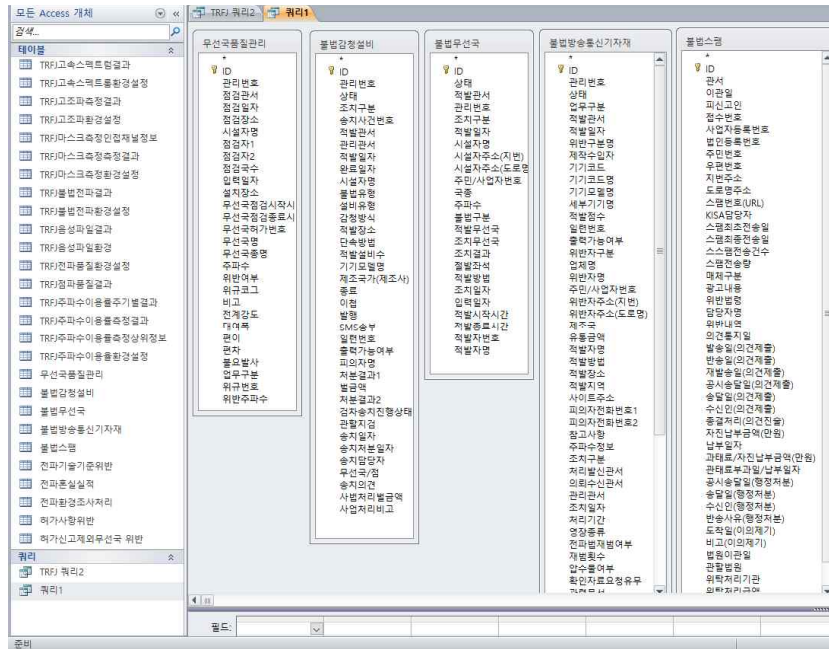


그림 2-1 쿼리 작업을 위한 공통 필드

## 2) 쿼리 작업을 위한 내부 필드 연결

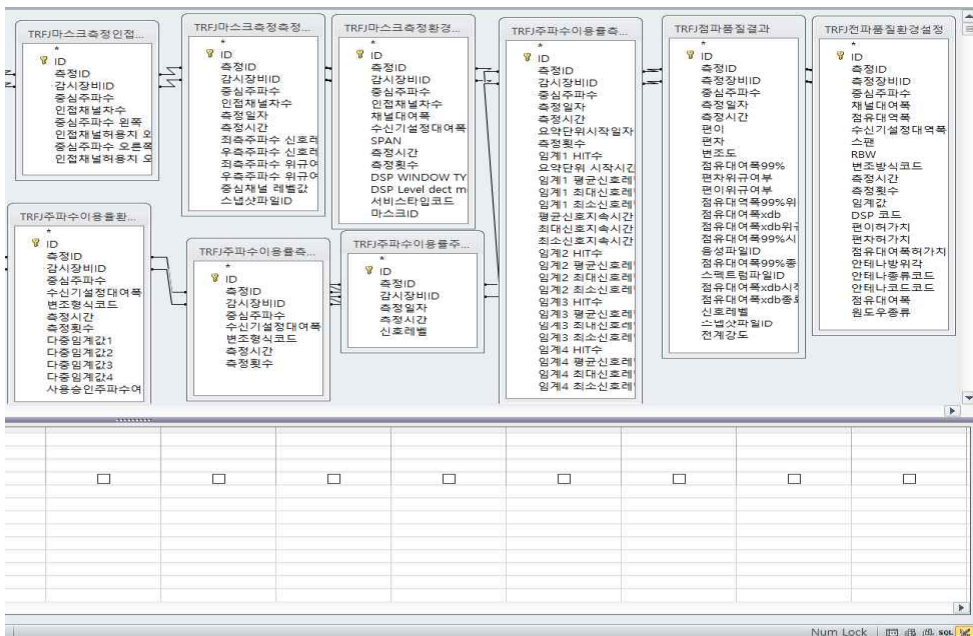
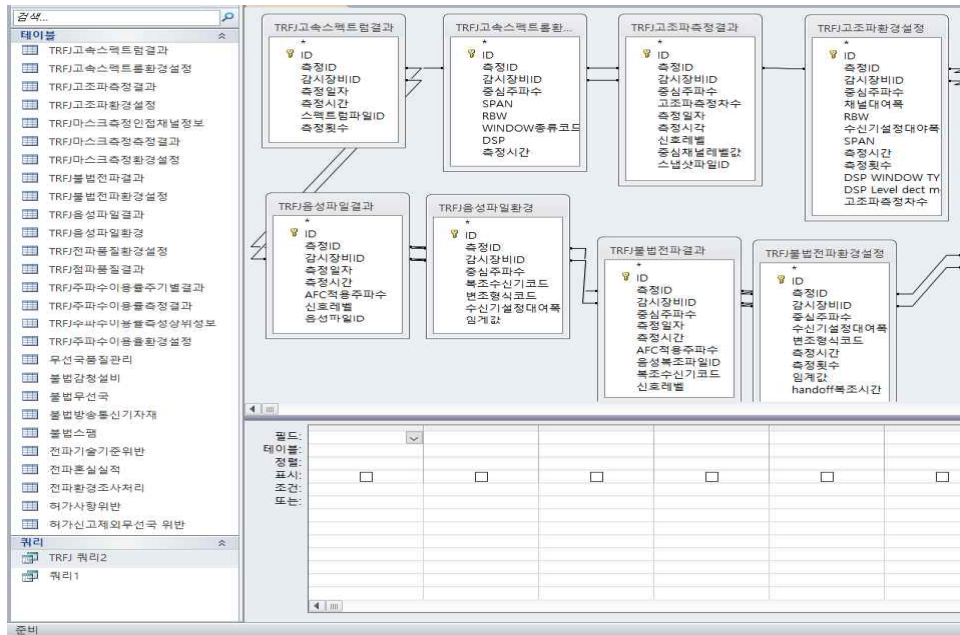


그림 2-2 쿼리 작업을 위한 내부 필드

### 3) 가상 DB 필드 연결



그림 2-3 가상 DB 필드 연결



## 제 3장 분석모델 시각화

### 제1절 시각화 사전 점검 사항

#### 1. 텍스트 기반 빅데이터 시각화 활용

빅데이터 분석에서는 분석 기법뿐 아니라 그에 대한 표현 기술 또는 시각화(Visualization) 방법도 매우 중요한 요소이다. 데이터를 어떻게 표현하느냐에 따라 새로운 트렌드나 패턴을 찾아낼 수도, 찾아내지 못할 수도 있기 때문이다. 또한 분석 결과에 대한 적절한 시각화는 타인과 의사소통할 때 효과적이다. 따라서 데이터 시각화는 정보를 분명하고 효과적으로 전달하는 것을 목적으로 한다.

시각화 단계에서는 분석 결과가 단순 수치가 아닌, 분석가 이외의 사람들이 이해하기 쉽도록 그림 및 도표를 활용하여 표현하며 수집한 데이터 및 분석 결과를 담당자 및 관련 기관들과 공유하기 위한 목적으로 필요하다면 Open-API 형태로 제공하도록 개발한다.

##### 1) 시각화 도구 및 기술 선정 능력

- 시각화 화면 구성요소별 특징 분석 능력
- 시각화 스토리보드 기반 시각화 구성 요소 도출 능력
- 시각화 플랫폼 구현 기술
- 시각화 표현 방안 고안 능력
- 시각화 기법 고안 능력
- 시각화 구성요소 배치, 조정 능력
- 다양한 그래픽 디자인 라이브러리 활용 능력
- 시각화 입력 데이터 요건 정의 능력
- 시각화 입력 데이터 식별 및 정의 능력
- 데이터와 시각화 요소 매핑 요건 정의 능력
- 데이터와 시각화 요소 매핑 능력

- 시각화 도구 및 기술을 활용을 통해 시각화 구현 프로그래밍 능력

데이터 시각화는 빅데이터 분석 결과를 쉽게 이해할 수 있도록 시각적 수단을 활용하여 나타내는 것을 말한다. 도표, 이미지, 통계 그래프 등을 통해 빅데이터 안에 있는 정보를 전달한다. 수많은 데이터를 한두 장의 그림으로 표현하기 때문에 해당 비즈니스를 잘 아는 전문가가 만들어야 한다고 생각하지만, 비즈니스 전문가들은 원하는 결과를 미리 유추하면서 만드는 경향이 있을 수 있어 객관적인 결과를 만들지 못하는 경우도 종종 나타날 수 있기 때문에 가급적이면 객관화된 시각화 도구를 활용하는 것이 좋다. 그림 3-1은 “한눈에 보는 2012년 서울시 예산”을 나타내고 있다. 일반적으로, 이러한 데이터를 보여줄 때는 그림 좌측의 엑셀과 같이 나열식으로 보여주는 경우가 많았지만 이것을 그림의 우측과 같이 데이터 시각화로 변형할 경우 한눈에 알아보기 쉽게 표현된다.

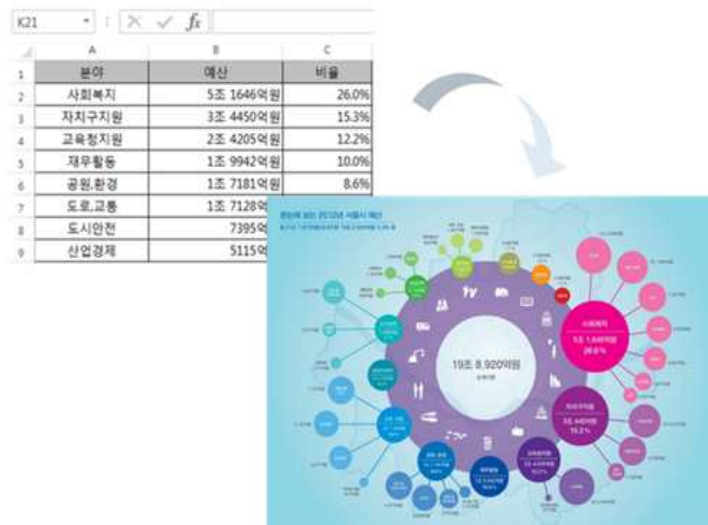


그림 3-1 빅데이터 시각화의 예

## 2) 정보 시각화 절차

정보를 보는 사람은 너무나 많은 정보로 인해 해석에 어려움을 느끼고 보는 것만으로도 피로움을 느낄 정도로 데이터가 쏟아지고 있다. 짧은 시간에, 적은 양의 자료를, 해석하기 쉬운 형태로 보여주는 데이터 시각화는 이러한 점을 해소하기 위해 가장 적합한 방법이라고 할 수 있다.

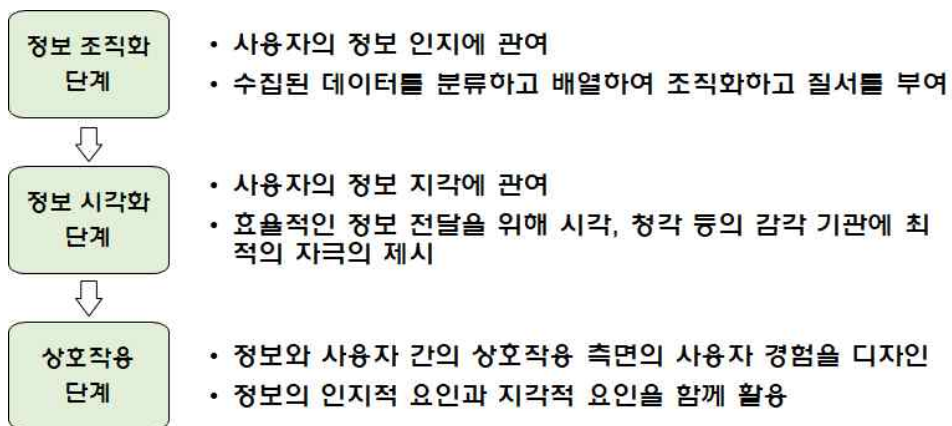


그림 3-2 정보 시각화 절차

### ① 정보 시각화(Information Visualization)

데이터 시각화를 위해서는 수집된 정보를 분류하고 배열하여 조직화한 후, 효율적인 정보 전달을 위해 시각, 청각 등의 감각 기관에 최적의 자극을 제시한다. 마지막으로, 정보를 사용자가 잘 인지할 수 있도록 사용자 상호작용(User Interaction or User experience)을 제공한다.

### ② 인포그래픽(Infographic)

정보 그래픽이라고도 하는 인포그래픽은 정보와 데이터, 지식을 시

각적으로 표현하는데 복잡한 정보를 빠르고 명확하게 전달하는 것이 핵심이다. 교통 표지판이나 복잡한 지하철 노선도 등이 대표적이다. 하나의 도표나 이미지에 수많은 정보를 요약해서 전달하기 때문에 신문이나 방송 같은 미디어나 교육 분야에서 활용이 많이 되고 있다.



그림 3-3 인포그래픽의 예  
- 출처: 연합뉴스 데이터



그림 3-4 정보 디자인에서 빅데이터 시각화의 영역  
- 출처: 숙명여대 빅데이터 분석

### ③ 빅데이터 시각화

앞에서 언급된 데이터 시각화, 정보 시각화, 인포그래픽은 빅데이터를 시각화하는 방법이라고 할 수 있다. 빅데이터 시각화는 보는

사람의 흥미를 유발 시키고, 정보를 습득하는 시간을 단축 시켜 빠른 상황 판단을 할 수 있도록 한다. 또한, 정보를 빠르게 확산할 수 있고, 한번 보고도 자료를 오래 기억할 수도 있다. 무엇보다 정보에 대해 효과적으로 커뮤니케이션 할 수 있다. 빅데이터 시각화는 크게 2가지로 분류된다. 하나는 객관적인 정보를 전달하는 “정보형”, 또 다른 하나는 주장하는 것을 알리는 “설득형” 이다. 빅데이터 시각화는 설득정보다는 정보형에 약간 더 가깝기 때문에 객관적인 데이터에 기반한 정보를 제공한다.

전문 도구를 사용해 그래픽으로 나타냈다고 무조건 좋은 것은 아니다. 빅데이터 시각화 도구는 사용자에게 겉으로 보이는 모습만 나타내는 것이고, 안으로는 분석된 데이터를 체계적으로 분류하고 분류된 정보가 오류 없이 잘 활용되도록 시각화나 상호작용, 프로그래밍까지 잘 되어야 한다는 것을 기억해야 한다.

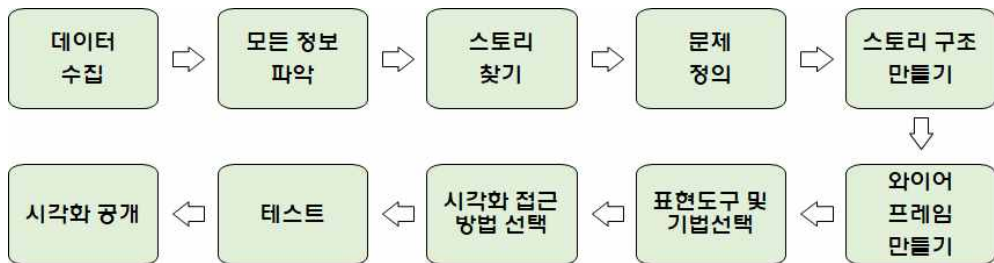


그림 3-5 빅데이터 시각화 단계

국내 자동 기상 관측망인 AWS/ASOS는 빅데이터 시각화를 사용하여 전국 573개 지점, 수평 해상도 13km, 1분 단위의 데이터 수집을 통해 기상 데이터를 분석하여, 홈페이지를 통해 실황 자료, 초단기 예보, 단기 예보가 제공되고 있다. 풍속, 풍향에 대한 예제를 보여주고 있으며, R의 `coord_polar()` 함수를 통해 시각화하고 있다.

과거에는 풍속과 풍향에 대해 글로 표현되어 제공되는 경우가 대부분이었다. 하지만, 그림 3-6, 3-7과 같이 시각화된 자료가 제공됨으로써 풍속, 풍향 데이터를 한눈에 살펴볼 수 있다.

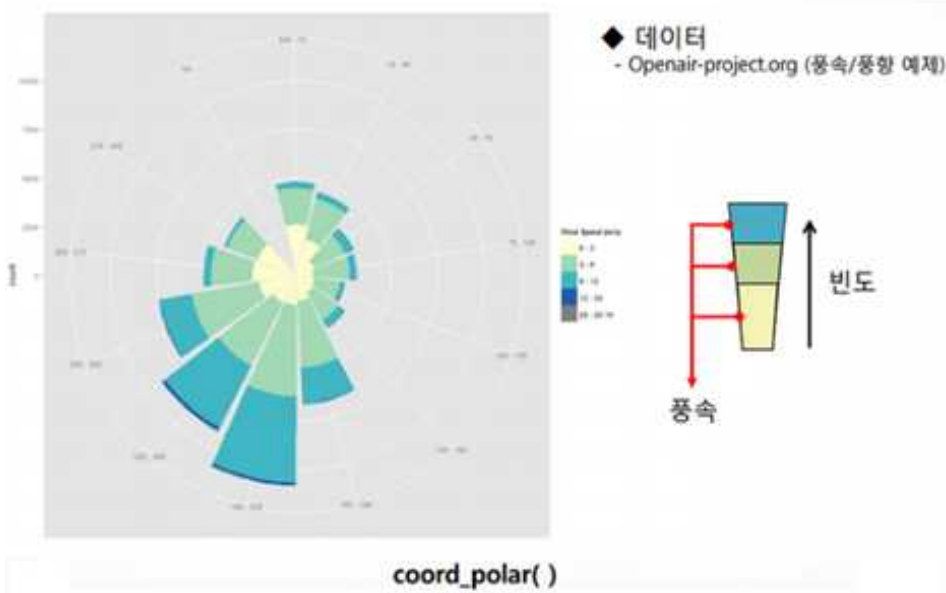


그림 3-6 풍속, 풍향 데이터 시각화 예제

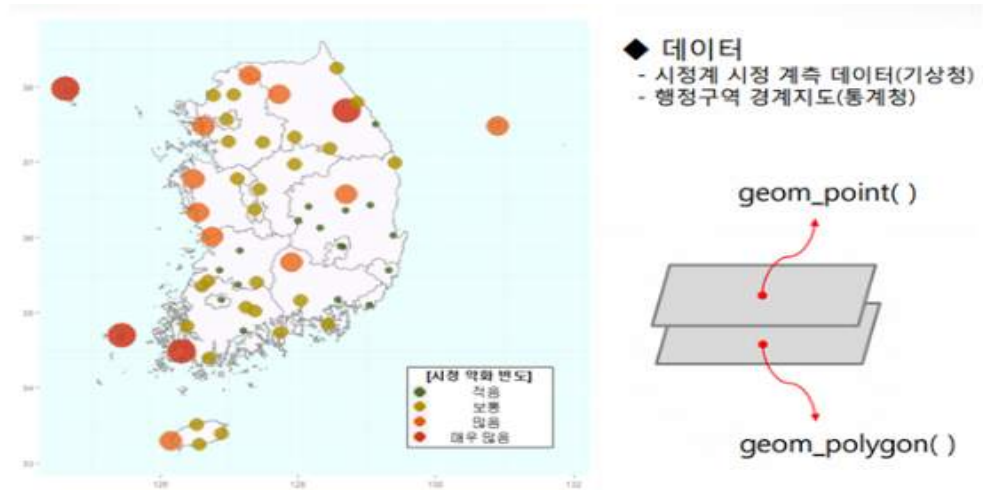


그림 3-7 우리나라 시정 데이터 시각화

## 2. 빅데이터 시각화 TOOL

### 1) Candela

JavaScript를 알고 있다면 오픈 소스 도구를 사용하여 풍부한 데이터 시각화를 만들 수 있다. Candela는 상호 운용 가능한 웹 시각화 구성 요소의 오픈 소스이다.

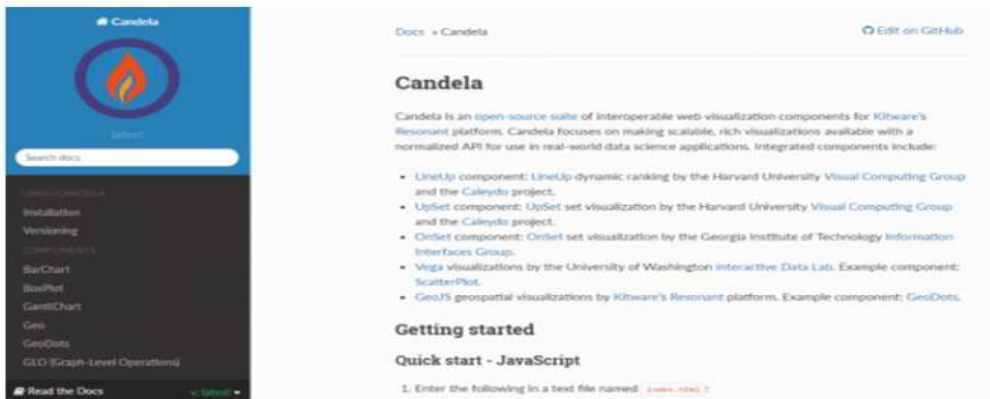


그림 3-8 Candela Tools

### 2) Charted

Charted는 무료 데이터 시각화 도구로서 CSV 파일과 Google 스프레드 시트에서 선 그래프 또는 막대 차트를 만들 수 있다. 여기에는 LineUp 구성 요소, UpSet 구성 요소, OnSet 구성 요소, Vega 시각화 및 GeoJS 지형 공간 시각화를 포함한 통합 구성 요소가 포함된다. 도구는 데이터를 저장하거나 조작하지 않는다. 순전히 시각화에 중점을 두었으며, 레이블 및 메모가 있는 선 또는 누적형 차트를 만드는 기본 기능이 제공된다.

### 3) Datawrapper

Datawrapper는 모바일 친화적인 데이터 시각화 도구로서 몇 초 안에 차트 및 보고서를 작성이 가능하다. 단일 사용자를 위한 무료 버전의, 이 도구는 10,000개의 월간 차트 뷰를 지원한다. 가로 막대형 차트, 분할형 차

트, 누적형 차트, 도트 플롯, 화살표 플롯, 영역형 차트, 산점도, 심볼맵 및 choropleth 맵과 같은 다양한 유형의 시각화를 만들 수 있으며 코딩이나 설계 기술이 필요하지 않다.



그림 3-9 Datawrapper Tools

#### 4) Google Data Studio

Google의 데이터 시각화 도구인 Google Data Studio는 Gmail 계정이 있으면 무료이며 쉽게 설정할 수 있다. 또한 Google 애드워즈, Google 애널리틱스, YouTube 애널리틱스 및 Google 스프레드 시트와 같은 Google 제품으로 손쉽게 연결이 가능하다.

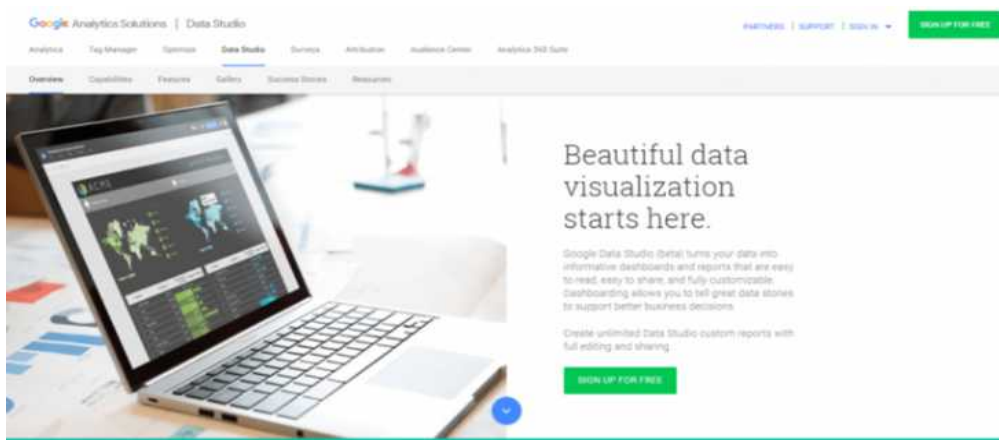


그림 3-10 Google Data Studio Tools



## 5) Google Charts

Google의 또 다른 간단하고 무료 데이터 시각화 도구는 Google Charts 도구이다. 이 도구에는 시각화를 위한 대화식 차트 및 데이터 도구가 함께 제공된다.

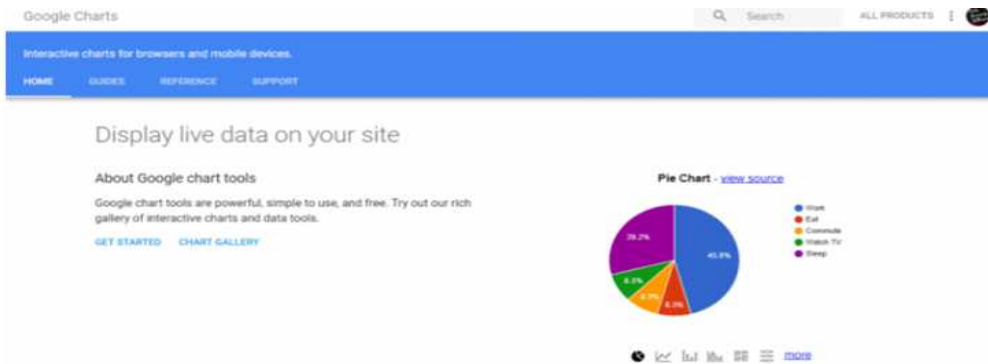


그림 3-11 Google Charts Tools

## 6) Leaflet

Leaflet는 모바일 친화적 대화형 지도를 만들 수 있는 오픈 소스 자바스크립트 라이브러리이다. 이 틀은 기능을 추가하기 위한 많은 플러그인이 있으며 다양한 데스크톱 및 모바일 플랫폼의 호환성이 좋으며 R에서도 많이 사용된다.

## 7) MyHeatMap

MyHeatMap은 대화형으로 지도 데이터를 볼 수 있는 무료 도구이다. 이 도구의 무료 버전은 공개 지도만 제공하며 무료 지도 각각에 대해 20 개의 데이터 포인트를 추가 할 수 있다. 색으로 구분 된 히트 맵을 사용하여 데이터를 쉽게 이해할 수 있으며 같은 지도에서 데이터 세트 간에 전환 기능이 제공된다.

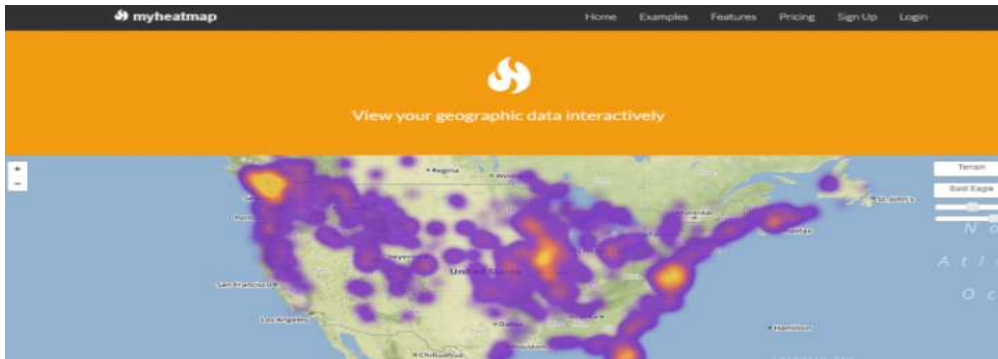


그림 3-12 MyHeatMap Tools

## 8) Openheatmap

Openheatmap을 사용하면 스프레드시트를 이용한 지도 생성이 가능하며 CSV 파일이나 Google 시트를 업로드 하여 몇 초 안에 대화 형 온라인 지도를 만들 수 있다. 이 도구는 우편번호 별 고객 인구 통계와 같은 데이터를 설명하는 데 사용할 수 있다.

## 9) Palladio

Palladio는 복잡한 과거 데이터를 시각화하기 위한 도구이다. 지도보기, 그래프보기, 목록보기 및 갤러리 보기와 같은 기능이 제공되며 CSV, TAB 또는 TSV파일의 데이터 시각화가 가능하다. 그래프보기를 사용하면 데이터 차원 간의 관계를 시각화 할 수 있으며 데이터는 선으로 연결된 노드로 표시가 된다. 반면에 목록보기에서는 사용자 지정 목록을 만들기 위해 데이터를 정렬 기능과 도구에는 그리드 내에서 데이터를 표시하는 갤러리 기능이 있다.

## 10) RawGraphs

RawGraphs는 TSV, CSV, DSV 또는 JSON 데이터를 시각화 할 수 있는 오픈 소스 플랫폼으로 사용이 간편하며 데이터를 차트로 변환할 수 있다.

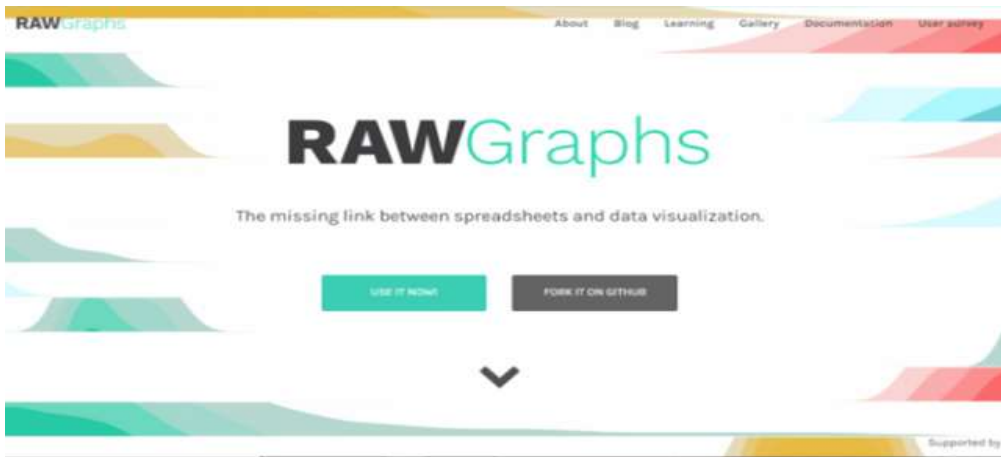


그림 3-13 RawGraphs Tools

#### 11) Timeline

Timeline은 보고서의 타임 라인을 만들 수 있으며 제공된 템플릿을 사용하여 Google 드라이브 계정을 연결하여 Google 스프레드시트에서 타임 라인 생성이 가능하다. 또한 JSON을 사용하여 사용자 정의 설치를 작성할 수 있다.

#### 12) Chartist.js

Chartist.js는 반응형 차트를 빠르고 쉽게 만들 수 있는 데이터 시각화 도구로서 이 도구는 큰 유연성 제공과 사용자 정의가 가능하다.

#### 13) Dygraphs

Dygraphs는 JavaScript를 기반으로 하는 오픈 소스 차트 라이브러리이다. 밀집된 데이터세트에서 분석 기능을 제공하고 모든 브라우저에서 사용자 정의가 가능하다. 이 도구는 오차 막대 / 신뢰 구간에 대한 강력한 지원을 제공한다.

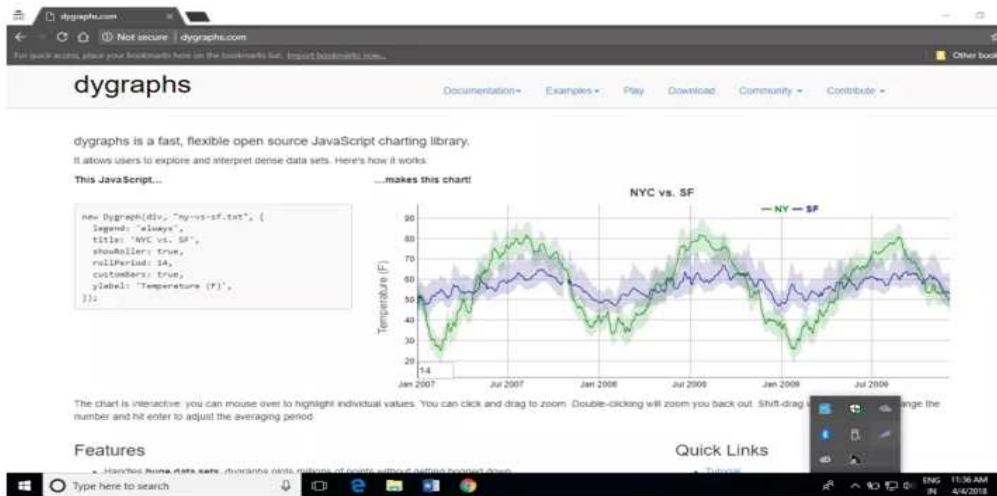


그림 3-14 Dygraphs Tools

그 외에 ColorBrewer, D3.js, Plotly, Polymaps, Weave 등의 다양한 무료 도구가 제공되고 있다.

### 3. 빅데이터 시각화 TOOL 비교

차트 라이브러리를 제공하는 주요 시각화 Tool로 Google Charts, HIGHCHARTS, Fusion Chart 그리고 Keylines가 있으며, 사용자가 원하는 형태의 차트를 직접 구현할 수 있는 D3에 대해 아래 표3-1에서 비교 하였다.

표 3-1 빅데이터 시각화 도구 비교

	Google Charts	HIGHCHAR TS	Fusion Chart	Keylines	D3.js
개요	Google 에서 제공하는 JavaScript 기반 무료 차트 라이브러리	HighSoft 사에서 제공하는 JavaScript 기반 차트 라이브러리	Sencha Labs 에서 제공하는 JavaScript 기반 무료 차트 라이브러리	그래프 시각화 응용 프로그램을 작성하는 데 도움이 되는 JavaScript 라이브러리	데이터를 기반으로 도큐먼트를 조작하기 위한 JavaScript 라이브러리
사용 방법	Google Chart 라이브러리를 개발 어플리케이션에 추가하고 필요에 따라 커스터마이징 하여 사용	Java Script 라이브러리를 다운로드하여 사용 CDN(: Content Delivery Network)서비스를 활용하여 라이브러리 직접 접근	표현하고자 하는 chart 의 sample code 및 JavaScript 라이브러리를 다운로드하여 사용	최신 버전을 다운로드 받아서 사용하거나 어플리케이션에서 직접 접근해서 사용	최신 버전을 다운로드 받아서 사용하거나 D3 site에 어플리케이션에서 직접 접근해서 사용
라이 센스 정책	Creative Common Attribution 3.0 License, Sample Code 는 Apache 2.0 License 적용 받음으로 무료로 사용 가능	개발자 수에 따라 구매 비용 있으며 유지보수 및 지원 서비스 포함	복제, 수정, 배포 등 소프트웨어 활용 전반에 걸쳐 무료로 사용 가능	개발자 수에 따라 구매 비용 있으며 유지보수 및 지원 서비스 포함	BSD License 적용을 받으며 무료로 사용 가능
주요 특징	HTML5 및 SVG 기술을 기반으로 차트 생성을 함으로 표준 기반 브라우저 및 모바일 디바이스에서 표현 가능	다양한 차트 유형을 제공하며 커스터마이징 이 용이함. 차트 출력 등 부가적 기능 모바일 및 데스크탑 표준	Google Group 에서 차트 개발자와 질의응답이 가능  API 문서를 통하여 상세한	Neo4j 그래프 데이터베이스의 데이터를 탐색하여 연결을 형성하고 숨겨진 추세를 찾아 낼 수 있는 강력한	표준기반 브라우저 지원 대량의 데이터 처리 및 인터랙티브 표현 용이 새로운 유형의 차트

		브라우저 지원다수 고객 reference 확보	사용가이드 제공	그래프 시각화 또는 링크 분석	구현의 유연성 높음
--	--	---------------------------------------	-------------	------------------------	---------------

#### 4. 빅데이터 시각화 라이브러리 구성 방안

시각화 라이브러리는 JavaScript 기반의 오픈 소스 라이브러리를 활용하되, 플랫폼에서 통합 관리하는 방안과, 각 분석과제에서 개별적으로 활용하는 방안이 있다.

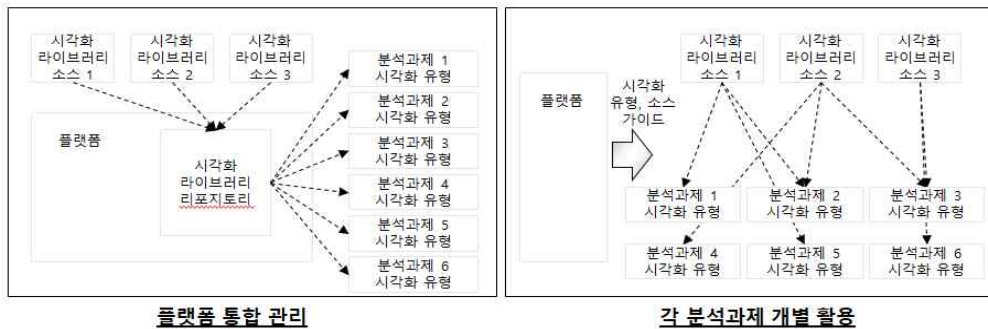


그림 3-15 시각화 라이브러리 구성 방안

표 3-2 빅데이터 시각화 라이브러리 구성 방안

	플랫폼 통합 관리	각 분석과제 개별 활용
설명	시각화 오픈 소스 제공 사이트에서 시각화 유형별 라이브러리를 다운받아 리포지토리에 저장하고 메뉴 구조를 통해 각 분석과제에서 필요로 하는 라이브러리 제공	플랫폼에서는 시각화 유형 및 관련 라이브러리 소스를 제공 등 가이드만 제공하고, 각 분석과제에서 필요한 라이브러리를 소스 사이트에 직접 접속하여 활용함
장점	각 분석과제에서는 플랫폼에서 제공하는 라이브러리 직접 활용 가능 분석과제별 유사한 시각화 표현이 가능함	각 분석과제의 요건에 맞게 최신 라이브러리 적용 가능 라이브러리 관리 위한 시스템 및 인력 비용이 절감
단점	시각화 라이브러리 관리 인력 및 추가 시스템 도입이 필요 최신 라이브러리 지속적 파악 및 업데이트 필요	분석과제별 독자적인 시각화 표현으로 통일성 결여

분석과제 시각화는 우선적으로 빅데이터 플랫폼의 다차원 분석 솔루션에서 제공하는 시각화 유형을 사용하고, JavaScript 기반의 무료 차트 라이브러리를 활용하되 추가적 요건이 있는 경우 D3 등 개발 도구를 활용할 수 있다.

과제별 시각화 도구 적용은 빅데이터 플랫폼의 다차원 분석 솔루션에서 제공하는 시각적 표현 기법을 우선 사용하고, Google Chart, HIGHCHARTS, Fusion Chart, Keylines 등 JavaScript 기반 시각화 오픈 소스를 활용한다. 기존 솔루션에서 제공되지 않는 사용자 특화된 표현이 필요한 경우 D3을 활용하여 구현할 수 도 있다.

## 5. 빅데이터 시각화 표현 유형

사용자들에게 시각화를 통하여 정보를 효과적으로 전달하기 위하여 여러 유형의 표현 방식들이 있으며 “전파 혼·간섭 예방 분석모델”에 적용할 만한 표현 유형들은 다음과 같다.

### 1) 네트워크 시각화 유형

표 3-3 네트워크 시각화 유형

구분	시각화 유형	설명	주요활용 형태
네트워크 (Network)	매트릭스 도표 (matrix diagram)	두 개 이상의 정보 그룹들 간의 연관성을 매트릭스 형식으로 표현한 도표	두 개의 데이터 항목들 간의 논리적 연관성 표현
	노드 링크 다이어그램 (node link diagram)	노드는 원 또는 박스로 표시하고, 연결은 선으로 기하 평면에 표현하는 그래프	개체들 간의 연결 관계를 표현
	아크 다이어그램 (arc diagram)	노드는 하나의 선 위에 위치시키고, 연결은 만원으로 표현하는 그래프	데이터 항목들의 분류 및 연결점을 표현

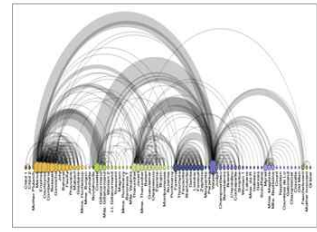
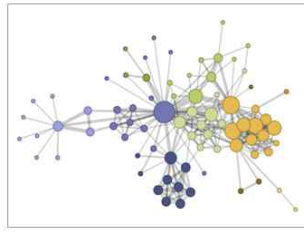
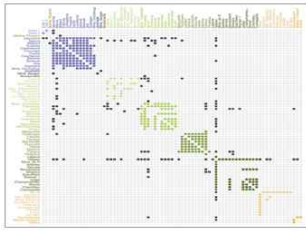


그림 3-16 네트워크 시각화 유형

## 2) 계층 시각화 유형

표 3-4 계층 시각화 유형

구분	시각화 유형	설명	주요활용 형태
계층 (Tree/ Hierar chy)	트리 맵 (tree map)	양적 변수에 의해 크기와 순서가 정해지는 사각형으로, 상위 수준은 다른 사각형을 포함하는 사각형으로 표현하는 그래프	많은 계층 구조를 가지는 정량적 데이터의 표현
	썬버스트 (sunburst)	햇살 모양을 기반으로 계층을 구성하는 항목의 크기를 면적으로 표현하는 그래프	정량적 데이터의 계층 구조 표현
	아이씨클 (icicle)	고드름 모양을 연상시키는 표현으로 위에서 아래 방향으로 계층을 표현하는 그래프	정량적 데이터의 계층 구조 표현
	방사형 트리 (radial tree)	방사형 형태로 외부로 향하여 하위 수준을 트리 구조로 표현한 그래프	다수의 구성요소를 가지는 데이터의 계층 구조 표현
	일반 트리 차트 (general tree visualization)	일반적으로 계층구조 차트, 조직 차트로 불리며 계층 구조를 박스 형태로 표현한 그래프	상대적으로 단순한 구조의 계층 데이터 표현



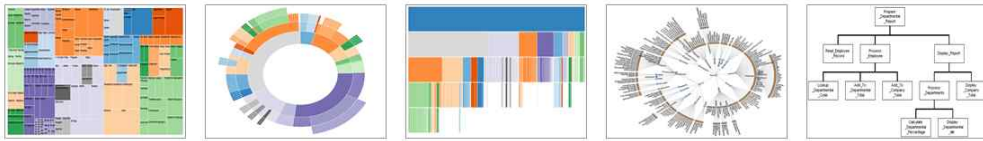


그림 3-17 계층 시각화 유형

### 3) 통계적 분석 시각화 유형

표 3-5 통계적 분석 시각화 유형

구분	시각화 유형	설명	주요활용 형태
통계분석 (Analytics)	박스 플롯 (box and whisker plot)	최대값, 최소값, 중앙값, 사분편차를 사용하여 추정값들의 분포와 이상치 값을 표현하는 그래프	계측 데이터의 통계 값 표현
	평행 좌표 (parallel coordinates plot)	여러 변수들에 대하여 각각의 Y 축을 가지며, 연결선을 통해 패턴을 파악할 수 있도록 표현하는 그래프	다변량 분석을 위한 주요 인자들을 표현
	페어스플롯 (pairs plot)	데이터 셋에 있는 변수를 대상으로 서로 쌍으로 상관관계를 분석할 수 있도록 매트릭스 형태로 표현한 그래프	여러 변수들 간의 상관관계 표현

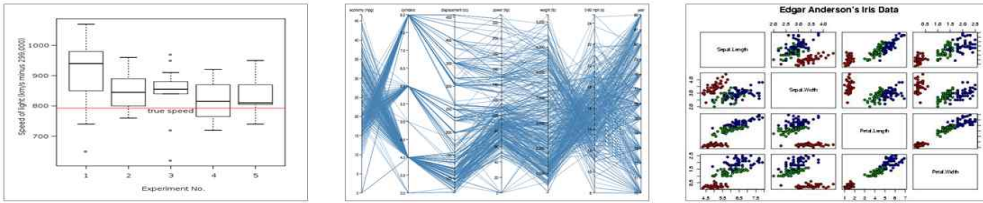


그림 3-18 통계적 분석 시각화 유형

#### 4) 빈도 시각화 유형

표 3-6 빈도 시각화 유형

구분	시각화 유형	설명	주요활용 형태
빈도 (Frequency)	워드 클라우드 (Wordle, tag cloud)	단어의 크기를 통하여 핵심 단어를 직관적 으로 파악할 수 있 도록 표현하는 기법	단어의 출현 빈도를 표현
	불규칙 버블차 트 (unordered bubble chart)	일반적인 버블 차 트에서 축과 관계없이 값의 크기를 나타내기 위한 그래프	여러 분류 항목의 빈도 또는 비율을 표현

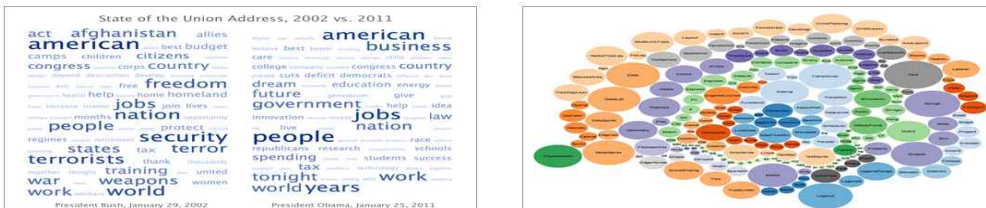


그림 3-19 빈도 시각화 유형

#### 5) 2차원 비교 시각화 유형

표 3-7 2차원 비교 시각화 유형

구분	시각화 유형	설명	주요활용 형태
2차원 비교 (2D)	히트맵 (heat map)	매트릭스 내 개별 값들의 크기를 색상 으로 표현하는 그래프	목표 값 대비 두 번 수들의 패턴 표현

Comparison)	모자이크 차트 (mosaic display/marimekko chart)	일종의 바 차트로서 동일한 높이를 가지며 각각의 바는 다른 변수에 의하여 분할된 형태로 표현하는 그래프	2차원 변수들에 대한 비율을 표현
	방사형 그래프 (radar chart)	중심에서 방사형 형태의 위치까지의 거리를 통해 수치 변수들의 항목을 비교하는 그래프	2차원 변수들에 대한 크기 비교

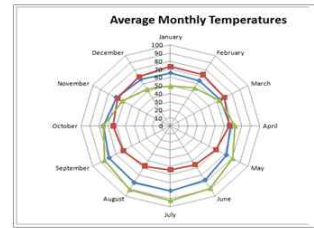
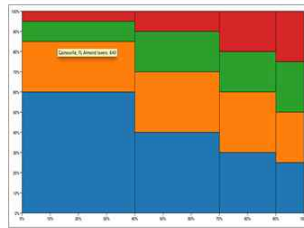
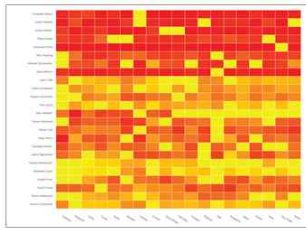


그림 3-20 2차원 비교 시각화 유형

## 6) 입출력 시각화 유형

표 3-8 입출력 시각화 유형

구분	시각화 유형	설명	주요활용 형태
입출력 (In/Output)	샌키 다이어그램 (sankey diagram)	복잡한 네트워크 상의 흐름을 표현하기 위한 그래프	두 변수의 크기에 따른 분석 대상 값의 크기 표현 시간 또는 기준 별 다양한 변수들의 크기 비교
	폭포 차트 (waterfall chart)	플러스 및 마이너스 측면에서 항목들을 순차적으로 구성하여 값들의 누적 영향을 표현하는 그래프	시간 또는 기준 별 다양한 변수들의 크기 비교

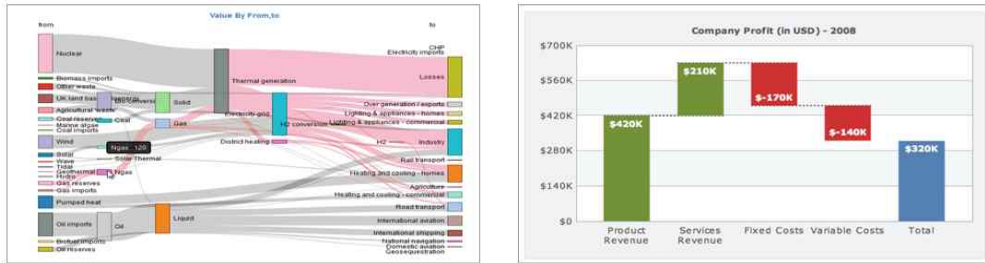


그림 3-21 입출력 시각화 유형

## 7) 다측정 시각화 유형

표 3-9 다측정 시각화 유형

구분	시각화 유형	설명	주요활용 형태
다측정 (multi value)	버블 차트 (bubble chart)	3차원의 데이터에 대하여 XY 축을 기준으로 한 원의 크기로 3번째 변수 값을 나타내는 표현	두 변수의 크기에 따른 분석 대상 값의 크기 표현
	선 도표 (line chart)	통계 수치를 곡선이나 꺾은선으로 나타내는 도표	시간 또는 기준 별 다양한 변수들의 크기 비교
	막대 그래프 (bar chart)	각 변수의 값을 사각형 바의 길이 값을 수직 또는 수평을 기준으로 표현하는 그래프	시간 또는 기준 별 다양한 변수들의 크기 비교

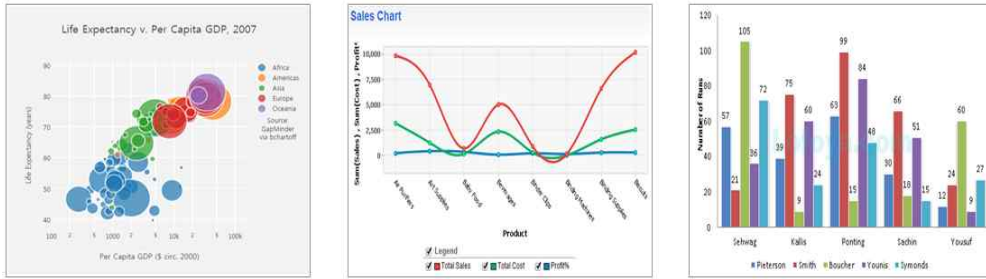


그림 3-22 다측정 시각화 유형

## 8) 상관분석 및 비율 시각화 유형

표 3-10 상관분석 및 비율 시각화 유형

구분	시각화 유형	설명	주요활용 형태
상관분석 (correlation)	산점도 (scatter plot)	직교 좌표계를 이용하여 두 개 변수 간의 상관관계를 표현하는 그래프	두 변수간의 상관관계 표현
비율 (proportion)	파이 차트 (pie chart)	개별 영역의 크기의 상대적인 비중을 원형으로 표현하기 위한 그래프	항목간의 상대적인 크기의 비율을 표현
	히스토그램 (histogram)	도수분포표를 자료로 하여, 계급 구간을 밑면으로 하고 도수를 높이로 하여 표현하는 그래프	구간 별 크기 및 상대적인 차이를 표현

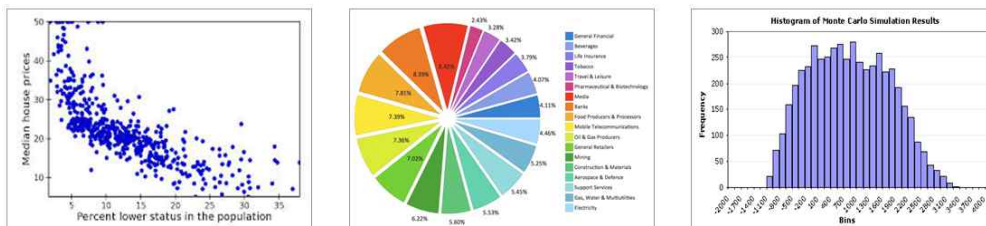


그림 3-23 상관분석 및 비율 시각화 유형

## 9) 지도 시각화 유형

표 3-11 지도 시각화 유형

구분	시각화 유형	설명	주요활용 형태
지도 (MAP)	등치지역도 (choropleth)	변수 값 크기구분에 따라 지역 색상을 구분하여 표현하는 지도	정량 데이터의 지역별 구분
	통계지도 (cartogram)	변수 값의 크기에 따라 바탕이 되는 지역의 크기를 왜곡하여 표현하는 지도	정량 데이터의 지역별 구분
	점분포도 (dot distribution map)	통계 내용을 점의 크기와 개수를 이용하여 표현하는 지도	이벤트 데이터의 지역별 구분

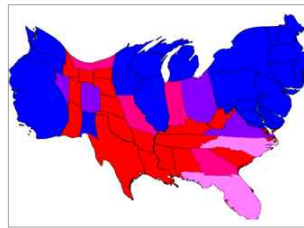
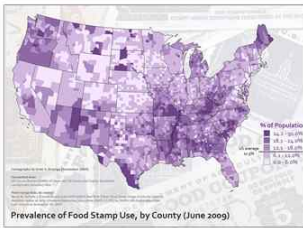


그림 3-24 지도 시각화 유형

## 10) 시간 시각화 유형

표 3-12 시간 시각화 유형

구분	시각화 유형	설명	주요활용 형태
시간 (Temporal)	연대표 (timeline)	이벤트를 연대 순서로 나열하여 표현하는 그래프	이벤트 정보의 시간 기준 표현
	시계열 그래프 (time series)	시간 경과에 따른 값의 변화를 선도표를 사용하여 표현하는 그래프	계측 및 계량 데이터의 시계열 표현

	연결 산점도 (connected scatter plot)	시간의 흐름에 따른 두 변수에 대한 변화를 나타내기 위한 그래프	계측 및 계량 데이 터와 주요 요인과의 관계 표현
--	---------------------------------------	---	-----------------------------------

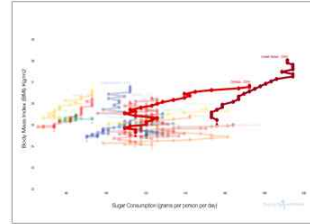
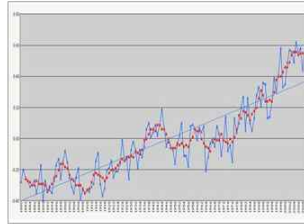
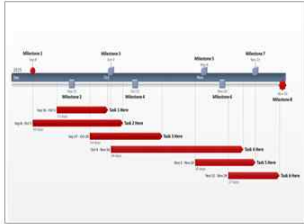


그림 3-25 시간 시각화 유형

## 제2절 『전파 혼·간섭 예방』 분석모델 시각화 방안

- 무선국 송신출력 변화 및 장비고장 예측 분석이 가능하도록 기본적으로 주파수를 먼저 선택하여 전파품질 측정값을 추세선 그래프로 확인 가능하도록 시각화
- 전체 또는 일부지역에 대한 측정값을 확인할 수 있도록 시각화
- 전파지정기준의 “지정용도별”로 측정값을 확인할 수 있도록 시각화
- 전파법 위반사항, 전파 혼신내역, 무선국 검사내역 등의 내역 일자별 리스트화 및 지역별 MAP 시각화
- 검출 내역을 지역별 검색 또는 체크리스트를 통하여 업종·기관별 통계 차트 실시간 시각화 적용
- 검출 내역의 일자별 지역 누적 현황 시각화와 지역별 비교 시각화
- 무선 국종에 따른 비교분석 차트 및 리스트 시각화
- 추세선 그래프 아래에 열별로 주파수이용효율 데이터, 전파감시·혼신실적, 무선국 검사일자, 장비고장·수리내역, 시스템 Log 데이터를 시간축(가로)상에 나열하여 원인을 분석할 수 있도록 시각화
- GIS를 통해 감시국소를 중심으로 선택한 주파수의 허가 무선국을 도식화
- 무선국의 전계강도 변화(5% 초과), 고조파의 허가주파수(20dB $\mu$ V 이상)의 운용자 알람(팝업)을 통한 전파품질 추세선 그래프 확인
- 시각화 화면에 무선국의 전계강도 변화, 고조파의 허가주파수 운용자 통보 기준조건을 조정할 수 있도록 입력창 필요
- 허가 무선국의 국종, 지정용도, 허가장소, 이동범위, 이동체 등록번호 등 상세 내역을 확인하여 불법 지정 외 의심주파수 분석하도록 시각화
- 측정 및 검출 데이터는 과거일자, 기간조회, 분기별, 전월, 최근 일자 설정 등의 기능으로 기간별·일자별 조회 및 시각화



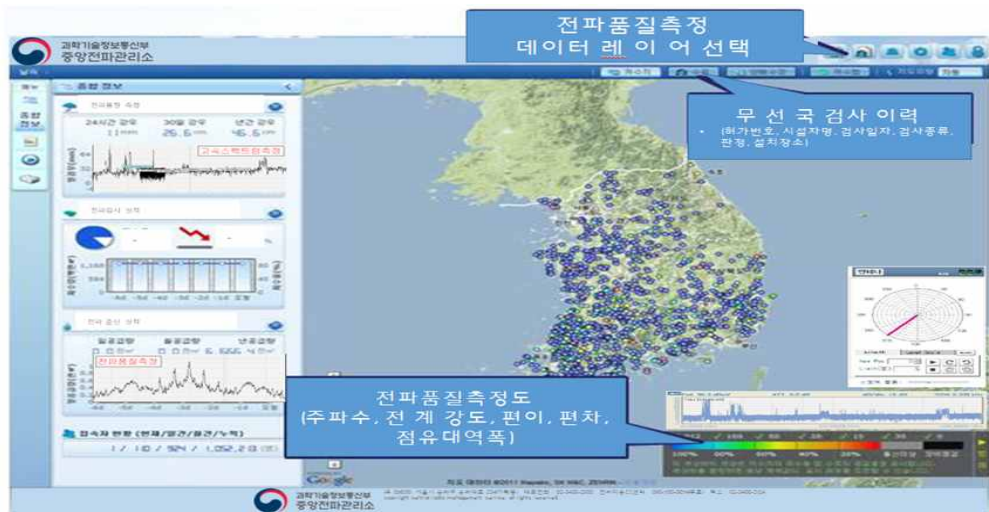


그림 3-26 전파품질측정 시각화

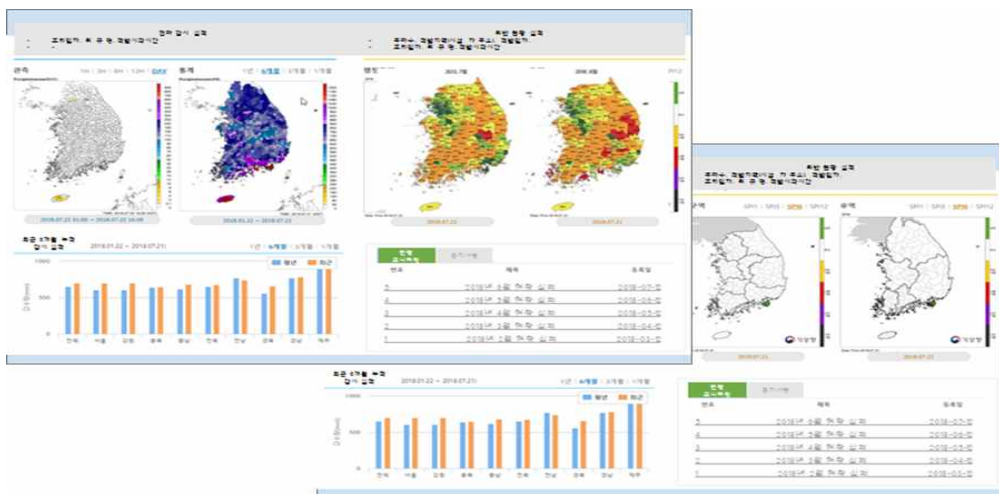


그림 3-27 전파감시실적 및 위반실적 시각화(I)

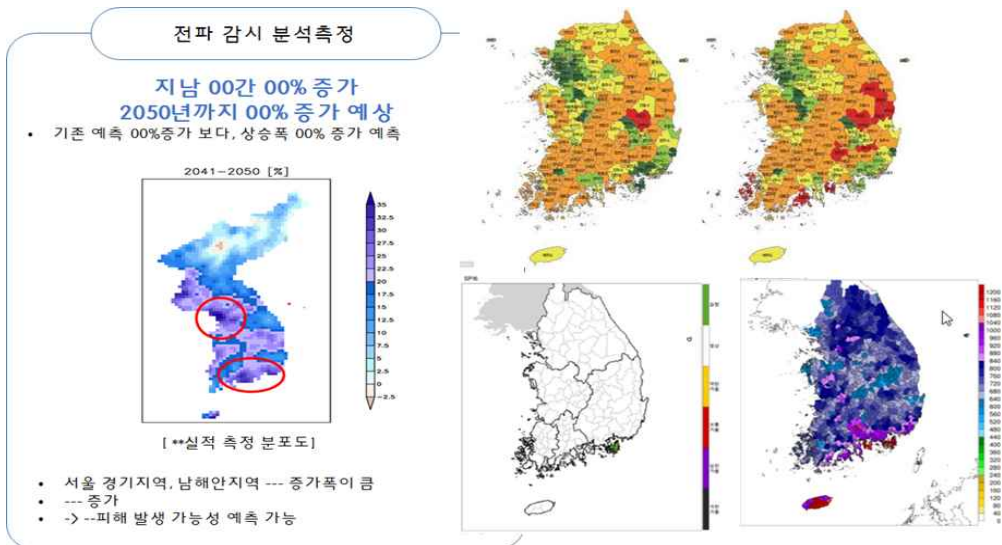


그림 3-28 전파감시실적 및 위반실적항 시각화 (Ⅱ)



그림 3-29 전파품질측정 레이어 별 시각화



그림 3-30 실시간 전파품질 및 전파혼신 상황 모니터링 시각화

본 연구과제에서는 “Google Data Studio”와 “Google Charts Tools”를 이용, 중앙전파관리소에서 제공한 엑셀 데이터를 가지고 임의 설계 구현 하였습니다.

엑셀로 작성된 data파일을 구글 환경에 맞도록 변환하여 적재시킨다. 이때 데이터 필드와 데이터 속성들은 데이터 정제와 검증이 완료된 상태로 적용해야만 한다. 그렇지 않을 경우 정확하지 않은 결과나 실행 오류가 발생한다.

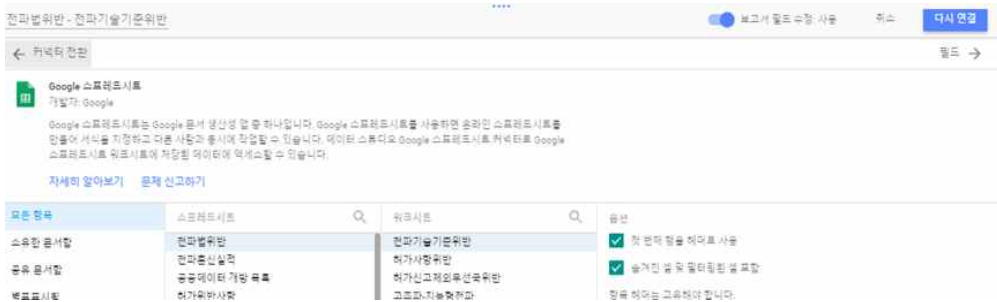


그림 3-31 데이터 변환(시트작업)

데이터 스튜디오에서 원하는 DB를 선택 후 시각화한다. 이때 시각화 용도에 맞도록 수식을 산출하여 필드를 생성함으로써 데이터 소스로 활용한다. 새로이 생성된 필드와 각 개체에 연동되어 데이터의 추가, 갱신, 삭제 데이터를 시각화에 적용할 수 있다.

전파법위반-전파기술기술훈위반

필드 추가

번호	필드	유형	집계	설명
1	상태	RBC 텍스트	▼	업종
2	전파법위반	RBC 텍스트	▼	업종
3	관리번호	RBC 텍스트	▼	업종
4	조치구분	RBC 텍스트	▼	업종
5	전파법일자	RBC 날짜(YYYYMMDD)	▼	업종
6	허가번호	RBC 텍스트	▼	업종
7	시설자명	RBC 텍스트	▼	업종
8	시설자주소(지번)	RBC 텍스트	▼	업종
9	시설자주소(도로명)	RBC 텍스트	▼	업종
10	시설자구분	RBC 텍스트	▼	업종
11	업종명	RBC 텍스트	▼	업종

필드 새로고침 38개 중 38개 필드

그림 3-32 데이터 변환(필드작업)

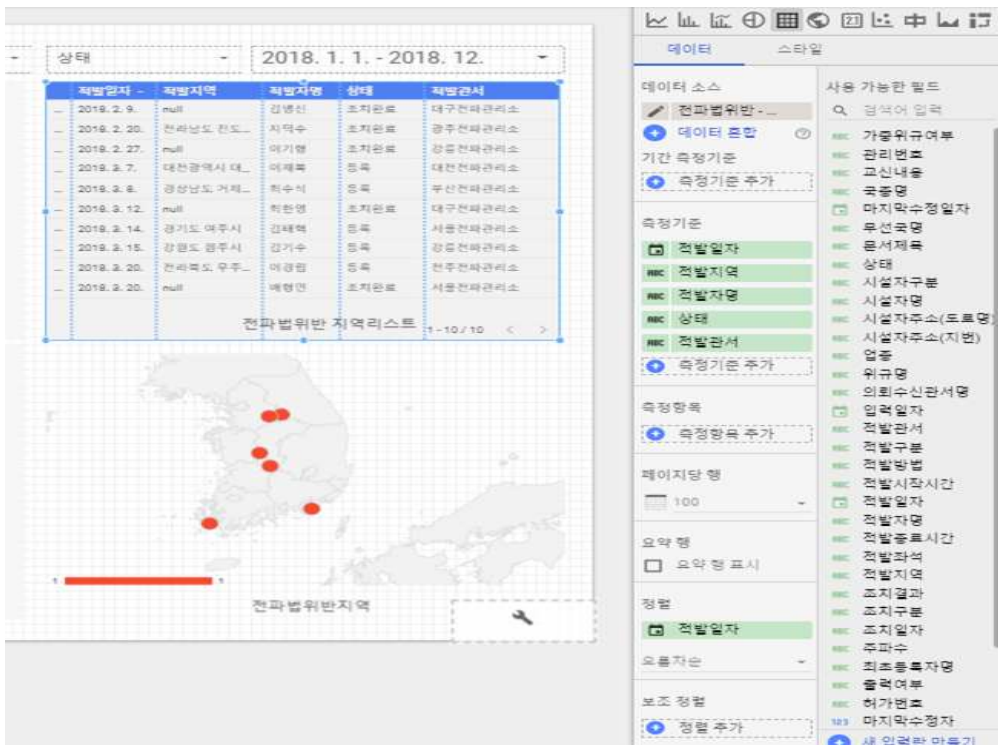


그림 3-33 데이터 변환(필드 시각화 적용 작업)

## 1. 전파감시 빅데이터 시각화 초기화면

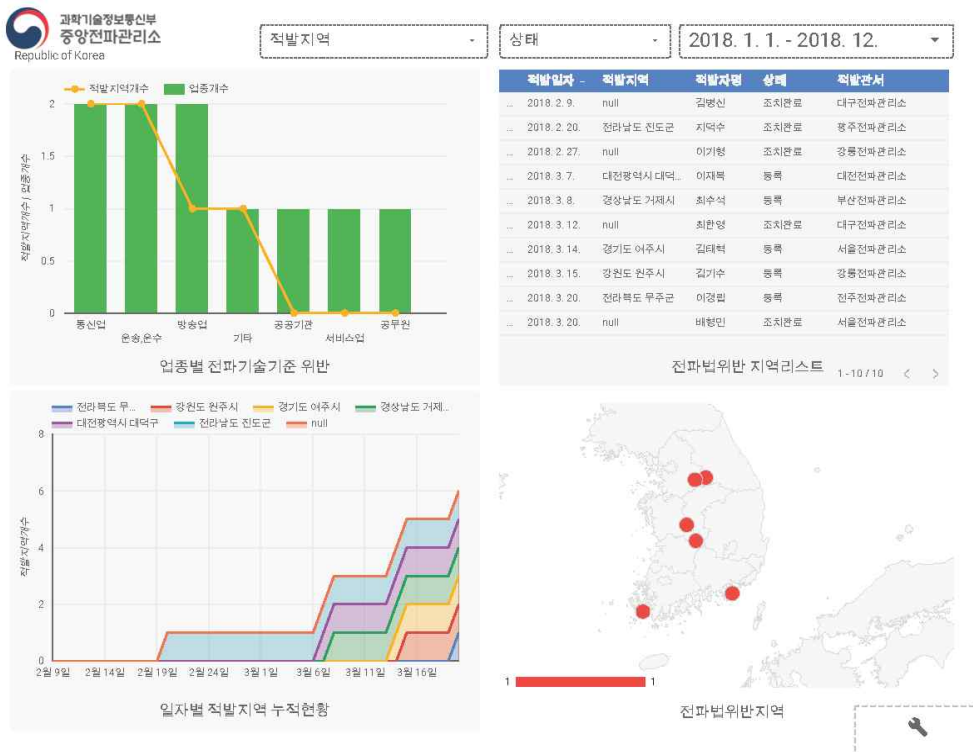


<https://datastudio.google.com/reporting/1DhoMeJVxkVQJInLPFJo80NfDnJLucqPs/page/2FJa>

그림 3-34 전파감시 빅데이터 초기화면

분석하고자 하는 주파수 대역 및 주파수 선택 시 신호세기 및 주파수 편차, 최대 주파수 편이, 점유 주파수 대역폭을 체크리스트 또는 검색어 입력 기능을 통해 화면에 표시되도록 설정한다. 이에 따른 전파 혼신, 전파법 위반, 장비고장 수리 내역, 무선 허가내역 등 검출 내용을 하나 또는 일부를 중복 선택하여 하나의 화면에 map과 신호형태로 동시 확인할 수 있다. 전파품질 분석을 위해 최대 주파수, 주파수 편차, 실시간 감시로 분리하여 체크리스트 형태로 보여준다. 고조파분석 및 주파수 이용률 또한 동일한 형태로 이루어진다. map형태의 시각화는 위 선택사항에 따라 지역이 도시되며 해당지역의 검출 수를 카운트하여 준다. 또한 다중 표시가 될 경우 원하는 지역을 검색하여 쉽게 검색되도록 제공한다. 이때 내역을 링크를 통해 자세히 볼 수 있도록 제공한다.

## 2. 지역별 전파감시 및 전파혼신 실적



<https://datastudio.google.com/reporting/1DhoMeJVxkVQJInLPFJo80NfDnJLucqPs/page/kKrB>

그림 3-35 지역별 전파 감시 실적 및 위반 현황 실적 화면01

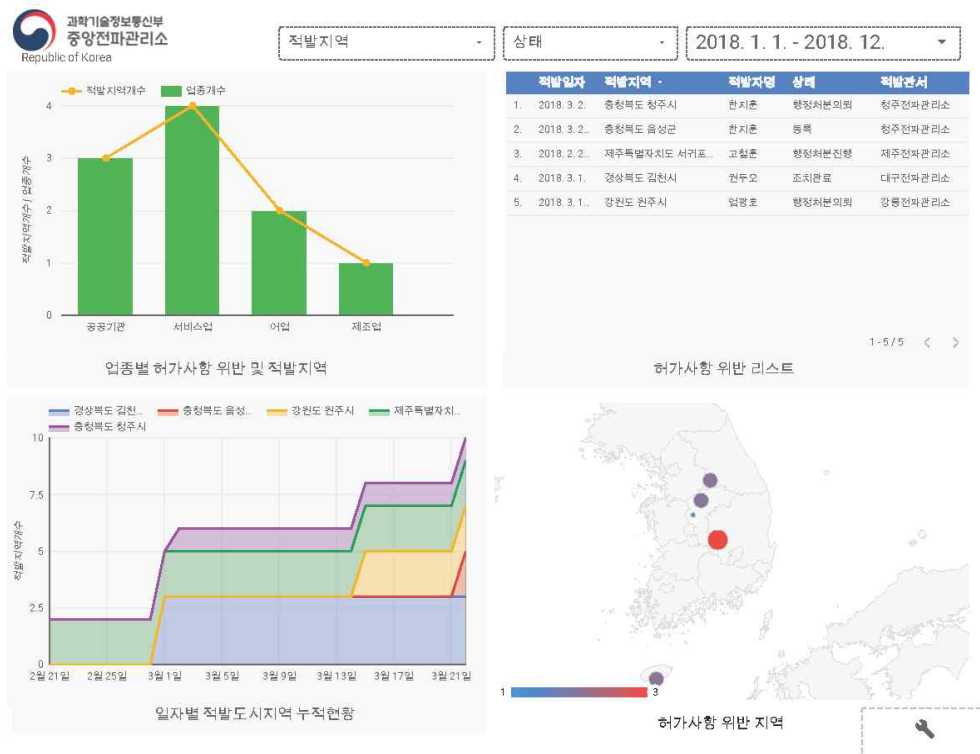
검출된 정보는 지역과 검출된 지역의 등록상태에 따라 map, 그래프 및 테이블 형태로 화면에 도시된다. 지역은 검색어 입력과 체크리스트 형태로 사용자가 쉽게 다중 또는 단일 선택이 가능하다.

등록여부 및 조치 상태는 다중 및 단일선택이 가능하며 그 결과는 직관적으로 쉽게 확인하여 볼 수 있다. 선택 및 검색에 따라 즉시 원하는 정보가 map, 막대그래프, 누적그래프, 테이블 형태로 제공된다. 일자별, 월별, 년별, 분기별 설정이 가능하며 과거기록 및 현재기록을 연동시키면 모든 내용을 화면에 도시할 수 있으며, 검색하고자 하는 특정기간이나 일자를 체크하여 시각화 할 수 있다.

선택 상태 및 “or” 선택 상태에 따라 검출된 적발일자, 적발자명,



상태, 적발관서, 적발방법 등의 내역은 텍스트 형태로 적용된다. 또한, 검출 내역은 업종과 지역별 이중 차트로 비교하여 그래프 형태로 제공되어 분석이 가능하다. 검출 내역을 누적그래프 형태로 시각화시 일자별로 누적현황을 제공한다. map형태의 시각화는 위 선택사항에 따라 지역이 도시되며 해당지역의 검출 수를 카운트하여 주며 쉽게 검출지역을 확인하여 볼 수 있다.



<https://datastudio.google.com/reporting/1DhoMeJVxkVQJInLPFJo80NfDnJLucqPs/page/RNtZ>

그림 3-36 지역별 전파 감시 실적 및 위반 현황 실적 화면02

적발지역

상태

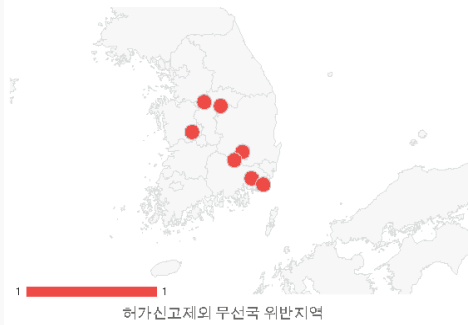
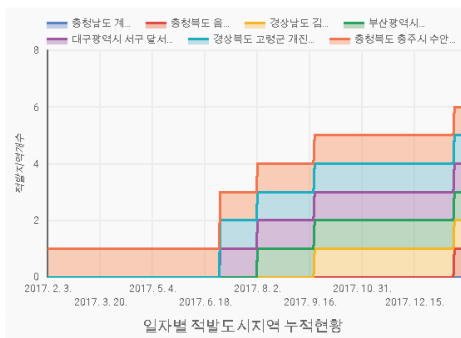
2018. 1. 1. - 2018. 12.



적발일자	적발자명	상태	적발관서	적발방법
1. 2017. 7. 1.	홍보자기	조치완료	대구전파관리소	이동감시
2. 2017. 8. 2.	채광용	조치완료	부산전파관리소	고정감시
3. 2017. 9. 20.	채광용	조치완료	부산전파관리소	고정감시
4. 2017. 2. 3.	주령	조치완료	청주전파관리소	혼신조사
5. 2018. 1. 26.	이만두	조치완료	대전전파관리소	고정감시
6. 2018. 1. 18.	신재창	조치완료	대전전파관리소	고정감시
7. 2017. 7. 1.	김치국	조치완료	대구전파관리소	이동감시

허가신고제외 무선국 위반 리스트

1-7/7 < >



<https://datastudio.google.com/reporting/1DhoMeJVxkVQJInLPFJo80NfDnJLucqPs/page/fVtZ>

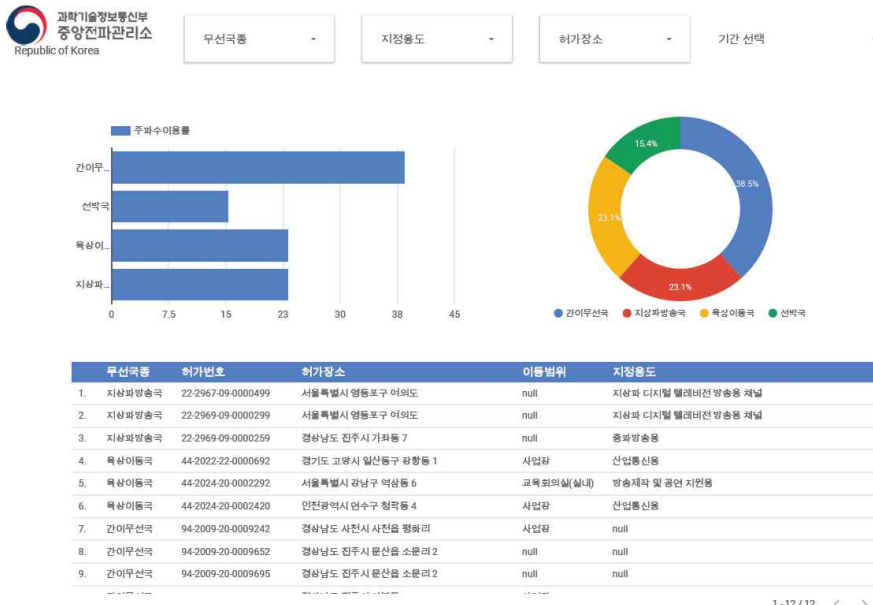
그림 3-37 지역별 전파 감시 실적 및 위반 현황 실적 화면03

허가사항 위반 검출 내역은 위와 동일한 형태로 제공되나 상태는 행정처분진행, 행정처분의뢰, 조치 상태 및 등록 여부 등의 체크리스트로 단일 또는 다중 선택이 가능하며 해당 내용은 사용자가 원하는 정보를 화면에 동시에 도시하여 준다.

허가 신고제외 무선국 검출 내역은 위와 동일한 형태로 제공되나 상태대신 적발 방법으로 혼신조사, 이동감시, 고정감시 등의 형태로 체크리스트 형태로 제공된다. 이때 단일 및 다중 선택이 가능하며 해당 내용은 화면에 도시된다.



### 3. 불법·지정 외 의심 주파수



<https://datastudio.google.com/reporting/1DhoMeJVxkVQJInLPFJo80NfDnJLucqPs/page/ZQJa>

그림 3-38 불법·지정 외 의심 주파수 화면

불법 지정 외 의심주파수는 검출된 내용에 따라 화면에 도시된다. 이때 무선국종, 지정용도, 허가 장소, 기간 선택이 가능하다.

지상파방송국, 육상이동국, 선박국, 간이 무선국 등으로 나누어 선택가능하며 단일 및 다중선택을 할 수 있으며 검색어 입력을 통하여 국종 검색이 가능하다.

지상파, 중파방송용, 산업통신용, 방송제작 및 공연지원용등 범주별 단일 및 다중 선택과 검색어 입력을 통하여 원하는 검출 내역에 대한 주파수 이용률과 해당 자세한 내역을 차트와 Table형태로 제공하여 시각화한다.

허가장소는 입력된 주소지에 따른 것으로 단일·다중 선택이 가능하며 원하는 지역 주소지를 도시 단위로 검색어 입력이 되도록 하였다.

검출된 내역은 주파수 이용률 및 국종에 따라 막대 및 원형 차트

형태로 시각화된다. 국종별 주파수 이용률은 막대그래프 형태로 수치화되어 도시되며 원형 차트는 해당 국종의 허가 비중을 시각화한다. 일자별, 월별, 년별, 분기별 설정이 가능하며 과거 기록과 현재 기록을 연동 시 모두 화면에 도시할 수 있으며, 검색하고자 하는 특정기간이나 일자를 별도로 체크하여 시각화 할 수 있다.

## 제 4장 빅데이터 분석가 양성 필요성 및 교육 방안

### 제1절 빅데이터 분석 및 자격

#### 1. 빅데이터 분석기법

글로벌 컨설팅 그룹 맥킨지&컴퍼니 산하 경제연구소인 맥킨지 글로벌 인스티튜트는 빅데이터 기술을 빅데이터 분석을 위한 ‘분석 기법’과 데이터를 수집, 조작, 관리, 분석하기 위한 ‘처리기술’로 구분했다. 분석기법에서 대표적인 방법은 마이닝 기법이다.

광산에서 광물을 캐낸다는 의미를 가진 마이닝(Mining) 기법은 다량의 데이터에 숨겨진 패턴과 관계 등을 파악해 미래를 전망할 수 있는 정보를 추출해 낸다. 현재 기업의 의사결정, 마케팅, 고객 관리에 적용될 뿐 아니라 금융이나 교육 등 다양한 영역으로 활용되는 중이다. 최근엔 그림과 영상 그리고 문서처럼 형태와 구조가 복잡해 정형화하기 어려운 비정형 데이터의 증가로 인해 다양한 분석기법이 주목받고 있다.

##### 1) 텍스트 마이닝

텍스트 마이닝(Text Mining)이란 대규모의 문서(Text)에서 의미 있는 정보를 추출하는 것을 말한다. 분석대상이 비구조적인 문서정보라는 점에서 데이터 마이닝과 차이가 있다. 텍스트 마이닝은 텍스트 분석(Text Analytics), 텍스트 데이터베이스로부터 지식 발견, 문서 마이닝(Document Mining) 등으로 불리기도 한다.

대용량의 텍스트 데이터에서 의미 있는 정보를 추출하고, 해당 정보와 연계된 정보를 파악하는 등 텍스트가 가진 단순한 정보 이상의 의미 있는 결과를 얻어 낼 수 있다는 장점을 가지고 있다.

##### 2) 오피니언 마이닝

인물, 이슈 등에 대한 대중들의 의견이나 평가, 감정 등을 분석하

는 기법을 뜻한다. 이를 통해 신상품 시장 규모를 예측하거나 소비자 반응에 대해 사전 파악이 가능하다. 소셜 미디어에 게재된 정형·비정형 텍스트가 전달하려는 의도가 긍정적인지 부정적인지 판별할 수 있기 때문이다. 사실과 의견을 구분하고, 의견을 추출하여, 긍정과 부정으로 나누고 그 강도까지 측정한다.

### 3) 웹 마이닝

웹 마이닝(Web Mining)은 인터넷을 이용하는 과정에서 생성되는 웹 로그(Web Log) 정보나 검색어로부터 유용한 정보를 추출하는 웹 대상의 데이터 마이닝이다. 웹 마이닝에선 웹 데이터의 속성이 반정형 혹은 비정형이고, 링크 구조를 형성하고 있기 때문에 별도의 분석기법이 필요하다.

웹 마이닝은 분석 대상에 따라 웹 구조 마이닝(Web Structure Mining)과 웹 유지지 마이닝(Web Usage Mining), 그리고 웹 콘텐츠 마이닝(Web Contents Mining)으로 구분된다. 이 중 웹 콘텐츠 마이닝은 웹 페이지에 저장된 콘텐츠로부터 웹 사용자가 원하는 정보를 빠르게 찾는 기법으로 검색엔진에 많이 사용된다.

## 2. 빅데이터 처리기술

빅데이터 시대를 맞이하면서 데이터의 폭발적 증가와 비정형 데이터 증가로 데이터의 저장 및 처리에 어려움이 발생하고 있다.

대응을 위해선 스토리지 장비를 추가하는 것이 가장 쉬운 해결책이지만, 엄청난 속도로 증가하는 데이터양을 따라잡기에는 비용 면에서 한계가 있다. 이에 대안으로 현재 대기업에서 사용하는 MPP(Massive Parallel Processing) 방식 역시 고비용 구조로 인해 거의 한계 상황에 다다르고 있다. MPP란 고도병렬처리로 프로그램을 여러 부분으로 나눠 여러 프로세서가 각 부분을 동시 수행하는 것을 말한다.

### 1) 하둡

빅데이터 처리 기술을 이야기할 때 빠지지 않고 등장하는 것이 바로 ‘하둡(Hadoop)’이다. 저가 서버와 하드디스크를 이용해 빅데이터를 상대적으로 쉽게 활용해 처리할 수 있는 분산파일 시스템인 하둡은 야후의 지원으로 개발됐다. 현재는 아파치 소프트웨어의 프로젝트로 관리되고 있다.

하둡은 빅데이터 활용을 가능하게 만든 빅데이터 플랫폼의 핵심기술이자 사실상의 표준이다. 하둡은 빅데이터를 저장하는 분산파일 시스템(Hadoop Distributed File System, HDFS)과 분산 병렬 처리하는 맵리듀스(MapReduce)로 구성되어 있다. 저가의 장비와 스토리지를 활용하여 저비용으로 방대한 양의 데이터 저장 및 처리가 가능하다. HDFS를 통해 다양한 형태의 초대용량 데이터를 분산 저장하며 맵리듀스로 빅데이터를 초고속으로 처리할 수 있다. 하둡은 크게 네 가지 용도를 가지고 있다. 첫째는 검색엔진 색인저장소(Indexing), 둘째는 데이터 분석 또는 통계분석, 셋째는 데이터의 전처리(Table Precomputation and Rollup), 마지막은 정형 데이터의 저장소(Structured Data Storage)로서의 용도다.

## 2) R

오픈 소스 프로젝트 ‘R’은 통계 계산 및 시각화를 위한 언어 및 개발환경을 제공한다. 또 R 언어와 개발환경을 통해 기본적인 통계 기법부터 모델링, 최신 데이터 마이닝 기법까지 구현하며 개선이 가능하다. 이렇게 구현한 결과는 그래프 등으로 시각화할 수 있으며, Java나 C, Python 등의 다른 프로그래밍 언어와 연결도 용이하다. Mac OS, 리눅스/유닉스, 윈도우 등 대부분의 컴퓨팅 환경을 지원하는 것도 장점이다.

## 3) NoSQL

NoSQL은 Not-Only SQL, 혹은 No SQL을 의미하며, 전통적인 관계형 데이터베이스(RDBMS : Relational Database Management System)와 다르게 설계된 비관계형 데이터베이스를 의미한다. 대표적인 NoSQL 솔루션으로는 Cassandra, Hbase, MongoDB 등이 존재한다. NoSQL은 테이블 스키마(Table Schema)가 고정되지 않고, 테이블 간 조인(Join) 연산을 지원하지 않으며, 수평적 확장(Horizontal Scalability)이 용이하다는 특징을 가진다.

### 3. 빅데이터 분석가 자격분류

#### 1) 데이터분석 전문가

데이터분석전문가(ADP : Advanced Data Analytics Professional)는 데이터 이해 및 처리 기술에 대한 기본 지식을 바탕으로 데이터 분석 기획, 데이터 분석, 데이터 시각화 업무를 수행하고 이를 통해 프로세스 혁신 및 마케팅 전략 결정 등의 과학적 의사결정을 지원하는 직무를 수행하는 전문가를 의미한다.

#### 2) 데이터분석 준전문가

데이터분석준전문가(ADsP : Advanced Data Analytics Semi-Professional)는 데이터 이해에 대한 기본 지식을 바탕으로 데이터 분석 기획 및 데이터 분석 등의 직무를 수행하는 실무자를 의미한다.

표 4-1 데이터솔루션 산업 분류에 따른 정의

데이터 종류	정의	비고
정형 데이터 분석	<ul style="list-style-type: none"> <li>· DBMS의 데이터를 사용자의 이용목적에 맞도록 간편하게 보고서를 생성하여 활용 할 수 있도록 지원하는 솔루션(데이터리포팅)</li> <li>· 미리 정의된 관점에 따라 사용자가 기준을 유연하게 변경하면서 데이터를 분석할 수 있도록 지원하는 솔루션(OLAP) 데이터 집합을 통계적으로 분석하여 의미 있는 인사이트를 찾아내는 과정을 지원하는 솔루션 등(데이터마이닝)</li> </ul>	Reporting 솔루션, OLAP(Online Analytical Processing), 통계분석패키지
비정형 데이터 분석	<ul style="list-style-type: none"> <li>· 웹사이트 데이터를 추출 및 분석하여 기술트렌드 및 기업 평판 등을 분석하기 위해 적용하는 솔루션(웹크롤링)</li> <li>· 주로 문서 및 웹사이트의 데이터에서 키워드를 찾아내고 빈도 및 연관관계 등을 분석하는 솔루션 등(텍스트마이닝)</li> </ul>	웹 크롤링(Web Crawling), 텍스트 마이닝, 비디오데이터 가공 솔루션, 음성 데이터

	<ul style="list-style-type: none"> <li>· 신문기사, 논문, 특허 등의 동향정보로부터 주요 개체 및 개체와 관련된 사실정보를 정형화된 형태로 추출하여 분석에 활용하는 솔루션(정보추출)</li> <li>· 음성 및 영상 등의 데이터로부터 의미 있는 데이터를 추출하여 분석하는 솔루션</li> <li>· 소셜미디어 내의 인맥정보를 이용하여 사회관계망을 분석하고, 포스팅 되는 내용을 통해 최신 이슈와 트렌드를 분석하는 솔루션</li> </ul>	가공 솔루션, 트위터, 페이스북 등의 API 연계 솔루션(Social Analytics 솔루션)
실시간 데이터 분석	<ul style="list-style-type: none"> <li>· 데이터 분석을 통해 위험 신호 또는 징후를 감지하여 위기나 재난에 대처할 수 있도록 예측하는 솔루션</li> </ul>	CEP(Complex Event Processing), BAM(Business Activity Monitoring)
데이터 시각화 분석	<ul style="list-style-type: none"> <li>· 데이터 세트의 의미를 시각적으로 표현하거나 탐색하는 과정을 지원하는 솔루션</li> </ul>	시각화 및 비주얼 분석솔루션(Visual Analytics 솔루션)



#### 4. 빅데이터 분석가가 되기 위해 필요한 기술

- 프로그래밍 기술 : R과 파이썬 등 프로그래밍 언어의 기초 필수
- 통계적 기술과 수학 : 추론적 통계 및 실험 설계
- 머신러닝기술
- 데이터 wrangling 기술 : 데이터를 매핑하고 데이터를 보다 편리하게 사용할 수 있도록 다른 형식으로 변환하는 기능
- 의사소통 방법과 데이터 시각화 기술

#### 5. 빅데이터 전문가가 되기 위해 필요한 기술

- 분석 능력 : 막대한 양의 데이터들을 적절하게 감지할 수 있는 능력. 이는 분석적인 문제해결 능력으로 어떤 데이터가 목적에 적합한 해결법과 가장 밀접한지 결정하는 것
- 창의성 : 데이터를 조합하고 해석하며 분석하는 새로운 데이터 전략을 창조해낼 능력
- 수학과 통계적 기술 : 수치 처리(number crunching)능력
- 컴퓨터 과학 : 컴퓨터는 모든 데이터 전략의 베이스로서 프로그래머는 데이터를 통찰력으로 처리하는 알고리즘을 고안할 수 있는 능력 필요
- 비즈니스 스킬 : 빅데이터 전문가는 비즈니스 성장뿐만 아니라 이익을 창출하는 기본 프로세스와 함께 비즈니스의 목표를 이해할 수 있는 능력 필요

## 제2절 빅데이터 교육과정 운영 현황

### 1. 데이터 산업 인력 현황

맥킨지 보고서에서는 2018년까지 미국에서만 140,000~190,000명의 데이터 분석 전문가가 추가로 필요할 것이고, 1,500,000명의 데이터 분석 기반의 관리자가 필요할 것이라고 분석하고 있다. 이와 같이 데이터 과학자에 대해 언급하는 이유는 쏟아지는 방대한 데이터 속에서 의미를 발굴하고 그것을 비즈니스 가치로 연결하는 사람들이 바로 데이터 과학자이기 때문이다.

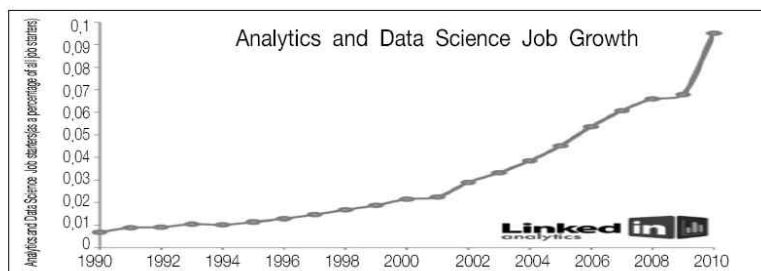


그림 4-1 데이터 과학자의 수요

그림 4-1은 데이터 과학자의 수요를 조사한 LinkedIn사의 보고서이다. 2010년에 신규취업자 1,000명 중에서 1명(0.1%)이 데이터 과학자임을 보여주고 있다. 2010년을 전후로 하여 미국의 많은 대학들에서 앞을 다투어 데이터 과학자 양성 프로그램을 만들고 있으며, 국내에서도 이 같은 현상은 마찬가지이다. 국내에서 처음으로 (2012년 3월) 과학기술정보통신부의 지원으로 대학원 과정에 빅데이터 전문가 양성 석사과정 학과를 만들어 운영하고 있다.

데이터 과학자는 호기심, 창의성, 객관성, 논리적/구조적 사고, 인내심, 상식, 그리고 세부 사항에 대한 이해력 등의 자질을 갖추어야 한다고 주장하고 있다. 쓰레기더미 같은 방대한 데이터로부터 정제, 통합, 분석 과정을 거쳐 의미 있는 결과를 만들어 내는 일은 마치 과학 수사관이 사건현장의 혈흔이나 발자국 지문 등을 근거로 사건을 풀어 나가는 것과 유사하다는 측면에서 중요한 자질로 인식되고 있다.

2017년 데이터산업 인력: 총 294,753명

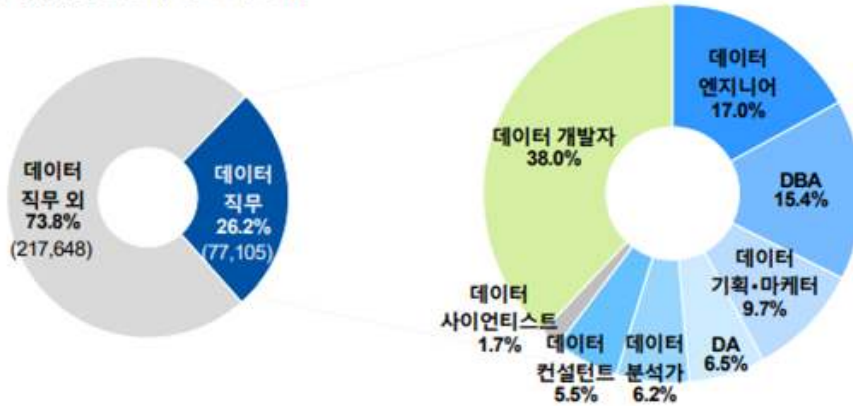
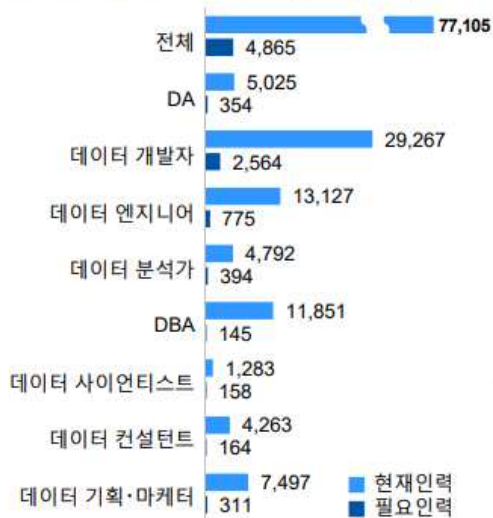


그림 4-2 데이터 산업 인력 현황

※ 2017년 데이터산업의 데이터직무 인력: 총 77,105명



※ 2017년 데이터산업의 데이터직무별 인력 부족률

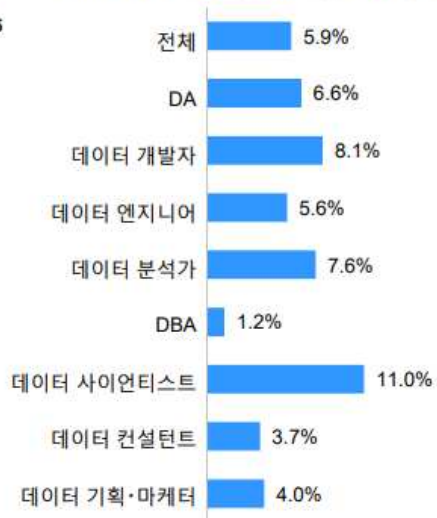


그림 4-3 데이터산업의 데이터 직무 인력 현황 및 부족률  
(2017년 데이터 산업 현황조사 -한국데이터진흥원)

## 2. 국외 교육과정 운영 현황

### 1) EU

EC(: European Commission)의 예산으로 범 유럽 포털을 통해 유럽 내 데이터교육을 위한 교재·커리큘럼·교육방법 등 관련 표준을 제시하고, 교육기관 정보를 제공함으로써 국가별 교육주체들이 이를 기반으로 전문인력을 양성하도록 돕고 있다.

표 4-2 EU의 인력양성 프로그램

프로그램	전담기관	교육기관	지원형태	교육방법	역량	수준	학습지원
유럽 데이터사이언스 아카데미 (EDSA)	유럽연합 위원회 (EC)	EDSA	<ul style="list-style-type: none"> <li>교육 제공</li> <li>인프라 지원</li> <li>연구 지원</li> </ul>	<ul style="list-style-type: none"> <li>온라인 교육</li> <li>이론/실습 교육</li> <li>셀프학습</li> </ul>	<ul style="list-style-type: none"> <li>기반</li> <li>기술</li> <li>분석</li> <li>사업</li> </ul>	<ul style="list-style-type: none"> <li>초급</li> <li>중급</li> <li>고급</li> </ul>	<ul style="list-style-type: none"> <li>여름학교</li> <li>워크숍</li> <li>컨퍼런스</li> </ul>

### 2) 영국

영국은 정부나 공공기관이 중심이 되고 대학과 협력체계를 구축하여 공공기관 임직원이나 기업체 재직자를 대상으로 실무역량을 강화하는데 중점을 두고 있다.

표 4-3 영국의 인력양성 프로그램

유형	프로그램	전담기관	교육기관	대상	역량	수준	학습지원	협력기관
정부 및 공공기관	훈련캠프 (bootcamp)	데이터센터연합(DCA)	동런던 대학교	구직자	<ul style="list-style-type: none"> <li>기반</li> <li>기술</li> </ul>	초급	고용 지원	<ul style="list-style-type: none"> <li>C-Net(훈련기업), Telecity 및 Telehouse (데이터센터 회사)</li> </ul>
	데이터과학촉진 프로그램	내각사무국 (Cabinet Office)	정부 디지털서비스센터 (GDS)	재직자	<ul style="list-style-type: none"> <li>분석</li> </ul>	고급 고급	<ul style="list-style-type: none"> <li>노트북 지급</li> <li>멘토링</li> </ul>	<ul style="list-style-type: none"> <li>국가통계처, 과학문야 정부 기관 및 분석 전문가</li> </ul>
대학원	기계학습 석사 프로그램	유니버시티 칼리지 런던	컴퓨터공학과	대학원생	<ul style="list-style-type: none"> <li>분석</li> </ul>	고급	-	-
	빅데이터 텍스트분석 석사과정	엑식스 대학교	컴퓨터공학과/전자공학과	대학원생	<ul style="list-style-type: none"> <li>분석</li> </ul>	고급	-	-

### 3) 독일

독일의 대학들은 데이터 관련 학과를 별도로 만들기보다는 보통 정보학이나 통계학, 경영학과 내에 데이터 관리 및 개발 관련 세부 영역을 구성하여 지원하고 있다. 몇몇 대학들은 데이터산업에 종사하고 있는 사람들을 위해 별도의 자격증 과정을 운영하고 있는데, 대표적으로 뮌스터 대학의 ‘데이터사이언스 인증과정’을 들 수 있다. 독일 인력양성 프로그램의 특징 중 하나는 인증과정이 많다는 점이다. 교육을 통해 지급 받는 자격증 또는 수료증이 실제 취업 시 자격증명으로 활용되고 있으며, 그만큼 수여조건이 까다롭다.

표 4-4 독일의 인력양성 프로그램

유형	교육기관	과정명	교육대상	역할	수준	학습자원	신학점력
정부 공공 기관	베를린 빅데이터센터	데이터 과학 석사	대학생 대학원생	분석	고급	• 논문 지원	-
	드레스덴/라이프치히 데이터 서비스센터	빅데이터 훈련과정	대학생 대학원생	기반 기술 분석	중급	• 논문 지원	-
	프라운호퍼 빅데이터 연구연령	데이터 과학자 기초인증	재직자	기반 기술 분석	중급 중급	• 자격증 발급	-
	프라운호퍼 빅데이터 연구연령	데이터 과학자 심화	재직자	기술 분석	고급 중급	-	-
	프라운호퍼 빅데이터 연구연령	데이터 과학자 산업별 특화	재직자	사업	고급	-	-
	브란덴부르크공과대학교	데이터 과학 인증	재직자 대학생	분석	고급	-	o
대학원	도르트문트공과대학교 컴퓨터통계학과	데이터 사이언스 - 빅데이터 인증	재직자 대학생	기술 분석	중급 고급	• 수료증 발급	-
	알브레히트 지그마링겐대학교	데이터 사이언스	대학원생	기반 기술 분석	초급 중급 고급	• 네트워크 지원 • 고용지원	o
	위네부르크 요아피나대학교	메니지먼트 데이터 사이언스	대학원생	사업	초급 중급 고급	-	-
	메니 프리드리히스라대학교	컴퓨터 데이터 사이언스	대학원생	사업	중급 고급	-	-
	헝니츠공과대학교 비즈니스	비즈니스 인텔리전스 - 분석	대학원생	분석 사업	고급	• 실용성 • 논문지원	o
민간	비트콜 아카데미	데이터 과학자 프랙티셔널 자격증	재직자	기반 기술 분석 사업	중급 고급	• 자격증 발급 • 교육시간 인정 • 필요요금 할인	o
	예술경영연합회 기술아카데미	데이터 메니지먼트 프로그램	재직자	분석	고급	-	-
	DWH Academy	데이터 웨어하우스	재직자	기술	중급 고급	• 자격증 발급 • 교육시간 인정	-
		비즈니스 인텔리전스	재직자	분석	고급	• 자격증 발급 • 교육시간 인정	-
	SanData IT트레이닝센터	SanData 빅데이터 교육 프로그램	재직자	기술	고급	• 자격증 발급 • 교육시간 인정	-
	ORDIX 교육센터	ORDIX 빅데이터 교육 프로그램	재직자	기술	중급	-	-

#### 4) 미국

미국의 데이터 인력양성 프로그램은 정부, 대학, 민간에서 활발하게 진행되고 있다. 정부에서 주도하는 인력양성 프로그램은 주로 국립과학재단과 백악관 과학기술정책실이 주관하고 있으며, 빅데이터 연구인력 양성프로그램과 빅데이터 분야의 인력수요를 충족하기 위한 고용촉진 프로그램으로 구분할 수 있다.

표 4-5 미국의 정부 주도 인력양성 프로그램 비교

전담 기관	프로그램	교육 기관	과정명	지원형태	교육 대상	교육방법	역량	수준
국립과학재단(NSF)	IGERT 프로그램	• 대학 및 연구기관	CIF22 Track	• 교육 제공	재직자	• 이론 · 실습 교육	분석	중급
	컴퓨팅 탐험대	• 대학		• 연구 지원	재직자	• 연구 지원	사업	고급
	DIDC	• 국가인문학기금		• 연구 지원	대학생 대학원생	• 연구 지원	기술	고급
	아이디어랩	• 국립과학재단(NSF)		• 교육 제공	대학생 대학원생	• 워크숍	사업	고급
	빅데이터 지역혁신 허브	• 콜롬비아대학교	북동부 빅데이터 허브	• 인프라 지원 • 교육 제공	재직자	• 인턴십 • 워크숍 • 지식교류	기반 기술 분석	고급
		• 조지아공과대학교 • 노스캐롤라이나대학교	남부 빅데이터 허브	• 인프라 지원 • 교육 제공	재직자	• 이론 · 실습 교육	분석	고급
		• 일리노이 얼버나 삼패인대학교	중서부 빅데이터 허브	• 인프라 지원 • 교육 제공	재직자	• 워크숍 • 회의	기반	고급
		• 워싱턴대학교 • UC 버클리 • UC 샌디에고 슈퍼컴퓨터센터	서부 빅데이터 허브	• 인프라 지원 • 교육 제공	재직자	• 워크숍 • 프로젝트 • 전시회	사업	고급
과학기술정책실(OS TP)	IT 훈련 · 인증 파트너십	• 마이크로소프트	마이크로소프트의 훈련 · 인증 프로그램	• 교육 제공	구직자	• 이론 · 실습 교육	기술	초급
		• 오라클 아카데미 (오라클대학)	오라클의 훈련 · 인증 프로그램	• 교육 제공	구직자	• 온라인 교육 • 이론 · 실습 교육	기술	초급
		• HP	HP의 훈련 · 인증 프로그램	• 교육 제공	구직자 재직자	• 이론 · 실습 교육	기술	초급
	TechHire 프로그램	• Opportunity@Work	데이터 분석교육	• 교육 제공	구직자 대학생	• 이론 · 실습 교육	분석	고급
		• 뉴욕 Tech Talent Pipeline initiative	데이터 분석교육	• 교육 제공	일반인	• 이론 · 실습 교육	분석	초급

표 4-6 미국 대학 빅데이터 학과의 교육과정

1. IT & 데이터 관리 분야	
Advanced Data Mining and Web Analytics(1)	Information Management(5)
Analytics for Big Data(4)	Integrating IT Architecture(8)
Analytics for Social Networks(10)	Introduction to Data Management(1)
Business Intelligence Capstone(1)	Introduction to Data Warehousing and Workflow Management(4)
Computation for Analytics(10)	Introduction to Programming in SAS(10)
Computer Information Systems(5)	Knowledge Discovery in Databases(8)
Data acquisition(10)	Machine Learning(10)
Data Analytics Programming(1)	Programming Macros/sql(3)
Data and Information Visualization(10)	SAS Programming(7)
Data Driven Decision Making(2)	
Data Management and Information Processing(4)	Social and Digital Media Analytics(2)
Data Visualization(2, 4, 9)	Social Media Analytics(1)
Data Warehousing and Business Intelligence(8)	Social Network Analytics(8)
Database Management(7)	Strategic Data Management(8)
Dealing with Big Data(2)	System Simulation(5)
Distributed Computing(10)	Technologies(6)
Distributed Databases(10)	Text Mining(3, 10)
Geospatial Analytics(3)	Web Analytics(3, 8, 10)
2. 경영, 비즈니스 관련 분야	
Introduction to Business Data Analytics(1)	Managing for Quality(2)
Advanced Decision Models(2)	Market Modeling(2)
Advanced Modeling(3)	Marketing Analytics(10)
Analysis of Markets(1)	Marketing Management(5)
Analytical Consulting Project Leadership(4)	Marketing Metrics(1)
Analytics for Competitive Advantage(4)	Marketing Online(8)
Business Communications for Analytics(10)	Operations Analytics(2)
Business Strategies for Big Data(10)	Operations Management(5)
Computational Finance(1)	Opt. Analysis(9)
Customer Analytics(3)	Optimization and Decision Analysis(1)
Decision Analysis(4, 5)	Optimization and Heuristics(4)
Decision Models(2)	Optimization and Simulation(3)
Decision-making(6)	Optimization(9)
Digital Analytics and Strategy: An Introduction(2)	Organization Behavior(5)
Electronic Commerce(7)	Pricing and Revenue Optimization(1)
Ethical, Legal, & Social Responsibilities of Business(6)	Probabilistic Models for Finance(2)
Financial Analytics(3)	Process Analytics and Optimization(8)
Financial Decision Making(8)	Revenue Management & Pricing(2)
Financial Enterprise Risk Engineering(8)	Risk Analytics(3)
Financial Management(1, 6)	Seminar in Advanced Business Problems(7)
Information Technology in Supply Chain Management(5)	Strategic Leadership & Management(6)
Introduction to Data-Driven Business Strategies(10)	Strategic Marketing(6)
Investment and Capital Markets(8)	Strategy, Change and Analytics(2)
Management Science(5)	Supply Chain Analytics(1)
Managerial Economic Analysis(6)	Supply Chain Management(7)
Managerial Effectiveness(6)	Turning Data into Revenue(5)

표 4-6은 미국의 10개 대학에서 빅데이터 분야로 개설한 학위과정의 교육과정을 편의상 ‘IT 및 데이터 관리 분야’, ‘경영, 비즈니스 관련 분야’로 구분하여 분류한 것이다(해당 내용은 각 대학의 홈페이지를 조사하여 정리한 것임). 일부 중복된 과목도 있지만 중복된 만큼 많은 대학에서 개설한다는 의미가 있다. 조사 대상으로 한 대학교는 다음과 같으며, 학위과정이 아닌 재직자 교육이나 자격증 과정은 제외하였다.

- ① 텍사스 주립대학 맥컴스 경영대학원(McCombs School of Business)
- ② 뉴욕 대학(New York University) 산하 스텐 경영대학원(Stern School of Business)
- ③ 노스 캐롤라이나 주립대학(North Carolina State University)
- ④ 노스웨스턴 대학(Northwestern University)
- ⑤ 미시간 주립대학(University of Michigan) 산하 디어본 경영 대학(Dearbon College of Business)
- ⑥ 아이오와 두부쿠의 로라스 칼리지(Loras College)
- ⑦ 루이지애나 주립 대학(Louisiana State University)
- ⑧ 스티븐스 공과대학교(Stevens Institute of Technology)
- ⑨ 신시내티 대학(University of Cincinnati)
- ⑩ 샌프란시스코 대학(University of San Francisco)



## 5) 일본

일본은 데이터 사이언스 스쿨에서 종합적인 교육 서비스를 제공하고 있다. 데이터 사이언스 스쿨은 입문과정과 실천과정으로 구분하여 프로그램을 제공하고 있는데, 교육생들은 이곳에서 데이터 분석의 기본 지식부터 비즈니스 현장에서 사용되는 실천적인 데이터 분석(통계 분석) 기법까지 습득할 수 있다.

표 4-7 일본의 정부 주도 인력양성 프로그램 비교

프로그램	과정명	전담기관	교육기관	교육방법	역량	수준
데이터 사이언스 스쿨	사회인을 위한 데이터 사이언스 입문	총무성 통계국	gacco	<ul style="list-style-type: none"> <li>• 온라인 교육</li> <li>• 이론·실습 교육</li> </ul>	• 기반	• 초급
	사회인을 위한 데이터 사이언스 훈련			<ul style="list-style-type: none"> <li>• 온라인 교육</li> <li>• 이론·실습 교육</li> </ul>	• 분석	<ul style="list-style-type: none"> <li>• 중급</li> <li>• 고급</li> <li>• 고급</li> </ul>

## 6) 싱가포르

싱가포르는 정보통신부 산하 정보통신개발청에서 데이터 전문인력 양성을 선도하고 있다. 대표적인 온라인 교육 프로그램은 MOOC은 10개의 모듈로 구분하여 교육을 진행하고 있는데, 통과 여부에 따라 참가자에 대한 평가와 구직능력을 분석하는 실무형 교육을 진행하는 것이 특징이다. CITREP 프로그램은 장비 업체 전문인력을 대상으로 기술변화 대응 훈련과 정보통신 융합과 신기술 연구를 지원하고 있다. (표를 앞장으로 당김)

표 4-8 싱가포르의 인력양성 프로그램 비교

프로그램	교육기관	교육대상	교육방법	역량	수준	학습지원	협력기관
MOOC 프로그램	코세라, 콘스텀킨스대 학교	일반인	<ul style="list-style-type: none"> <li>• 온라인 교육</li> <li>• 프로젝트</li> </ul>	• 분석	<ul style="list-style-type: none"> <li>• 중급</li> <li>• 고급</li> </ul>	<ul style="list-style-type: none"> <li>• 자격증 발급</li> <li>• 멘토링</li> <li>• 장학금</li> </ul>	—
CITREP	대학 및 교육기관	재직자	• 프로젝트	• 기술	• 초급	—	—
CLT	Google, IBM, SAS 등	구직자 재직자	<ul style="list-style-type: none"> <li>• 이론·실습 교육</li> </ul>	<ul style="list-style-type: none"> <li>• 기술</li> <li>• 분석</li> </ul>	<ul style="list-style-type: none"> <li>• 초급</li> <li>• 중급</li> </ul>	<ul style="list-style-type: none"> <li>• 멘토링</li> <li>• 장학금 지급</li> </ul>	• IDA
COAs 훈련 프로그램	인력양성센터(COAs)	재직자	<ul style="list-style-type: none"> <li>• 이론·실습 교육</li> </ul>	• 분석	<ul style="list-style-type: none"> <li>• 중급</li> <li>• 고급</li> <li>• 고급</li> </ul>	<ul style="list-style-type: none"> <li>• 급여 보조</li> </ul>	<ul style="list-style-type: none"> <li>• EMC, Microsoft 등</li> </ul>

### 3. 국내 빅데이터 전문인력 운영 현황 및 정책

빅데이터 프로그램의 교육과정은 데이터 과학자 양성 프로그램의 교육과정과 같이 데이터 과학자가 가져야 할 역량을 배양하는 데 초점을 맞추어야 한다. 즉, IT 기술과 수학 및 통계지식 그리고 현업 지식을 갖추기 위한 교과목들로 구성된다. 물론 기존의 IT 분야 세부 기술 관련 전체 과목보다는 데이터 관련 과목들(예: 데이터베이스, 데이터 마이닝, 데이터베이스 프로그래밍 등)에 포커스를 맞추어 공부하는 것이 중요하다.

#### 1) 빅데이터 인력현황

일반 산업의 빅데이터 인력은 전체 3,631명 중 빅데이터 개발자가 1,042명(28.7%)의 비중으로 가장 높은 분포를 보였으며 빅데이터 분석가가 727명(20.0%), 빅데이터 엔지니어 598명 (16.5%) 순으로 인력 비중이 높은 것으로 나타난 반면, 빅데이터 컨설턴트는 전체 인력의 10% 미만인 344명 수준으로 가장 낮은 인력 분포를 보이고 있다.

2017년 국내 빅데이터 업무를 담당하는 전문 인력은 총 9,955명으로 조사됐으며, 3년 후인 2020년도에는 현재 인력에서 60.4% 증가한 15,963명의 인력이 필요한 것으로 나타났다. 이에 일반 산업을 중심으로 빅데이터에 대한 활용이 크게 확대되며 관련 전문인력의 수요가 크게 증가할 것으로 예상된다.

2017년도 데이터 산업과 일반 산업을 포함한 전체 산업에서의 빅데이터 전문 인력을 직무별로 살펴보면 빅데이터 개발자가 2,908명으로 전체 29.2% 수준으로 가장 높게 나타났으며, 데이터 사이언티스트 1,803명(18.1%), 빅데이터 컨설턴트 1,724명(17.3%) 순으로 인력이 분포되어 있는 것으로 조사됐다. 2020년도에는 빅데이터 인프라 구축이 확대되면서 빅데이터 분석에 대한 수요가 상승해 해당 인력에 대한 수요가 크게 증가할 것으로 전망되기에 관련 직무에 대한 전문인력 양성 및 수급을 위한 대책 마련이 필요하다.

표 4-9 빅데이터 관련 인력현황(2017년 기준)

현재 인력 직무 구분		데이터산업		일반산업	
		인력 수	%	인력 수	%
전체 DB 인력	① DA	5,025	6.5	5,046	15.7
	② 데이터 개발자	29,267	38.0	11,987	37.2
	③ 데이터 엔지니어	13,127	17.0	3,507	10.9
	④ 데이터 분석가	4,792	6.2	3,606	11.2
	⑤ DBA	11,851	15.4	6,012	18.6
	⑥ 데이터 사이언티스트	1,283	1.7	520	1.6
	⑦ 데이터 컨설턴트	4,263	5.5	741	2.3
	⑧ 데이터 기획/마케터	7,497	9.7	796	2.5
	합계	77,105	100.0	32,215	100.0
빅데이터 인력	⑨ 빅데이터 개발자	1,866	29.5	1,042	28.7
	⑩ 빅데이터 엔지니어(하둡/NoSQL)	1,081	17.1	598	16.5
	⑪ 빅데이터 분석가	381	6.0	727	20.0
	⑫ 데이터 사이언티스트	1,283	20.3	520	14.3
	⑬ 빅데이터 컨설턴트	1,369	21.6	355	9.8
	⑭ 빅데이터 기획/마케터	344	5.5	389	10.7
	합계	6,324	100.0	3,631	100.0

자료 : 2017년 데이터산업 현황조사, 한국데이터진흥원

구분	2017년		2020년		'17년 대비 '20년 필요인력 증가	
	인력 수	비중(%)	인력 수	비중(%)	인력 수	성장률(%)
빅데이터 개발자	2,908	29.2	5,258	32.9	2,350	80.8
빅데이터 엔지니어(하둡/NoSQL)	1,679	16.9	2,586	16.2	907	54.0
빅데이터 분석가	1,108	11.1	2,439	15.3	1,331	120.1
데이터 사이언티스트	1,803	18.1	2,299	14.4	496	27.5
빅데이터 컨설턴트	1,724	17.3	2,233	14.0	509	29.5
빅데이터 기획/마케터	733	7.4	1,148	7.2	415	56.6
Total	9,955	100.0	15,963	100.0	6,008	60.4

자료 : 2017년 데이터산업 현황조사, 한국데이터진흥원

그림 4-4 직무별 빅데이터 전문인력 현황 및 전망(전체산업)

## 2) 빅데이터 전문인력 양성 주요 정책

우리나라 데이터 인력양성은 정부기관 주도의 추진체계를 갖추고 있다. 데이터 인력양성은 과학기술정보통신부를 주무부처로 한국데이터진흥원을 전담기관으로 하여 기업체 재직자를 주요 대상으로 전문인력 양성 교육을 추진하고 있으며, 고용노동부는 구직자 대상으로 직업능력개발을 통한 취업 목적의 교육을 담당하고 있다.

### ① 과학기술정보통신부의 빅데이터 전문인력 양성

정부는 2012년 12월, 교육과학기술부, 행정안전부, 지식경제부, 방송통신위원회, 국가과학기술위원회 공동으로 '스마트 국가 구현을 위한 빅데이터 마스터 플랜'을 발표했다. 마스터 플랜에서는 빅데이터 활용기반 조성을 위해 빅데이터 공유·활용 인프라 구축, 기술개발지원과 더불어 전문인력 양성을 체계적으로 추진할 것이라고 밝힌 바 있다. 이후 설립된 과학기술정보통신부의 주도하에 데이터 전문인력 양성이 이루어지고 있다.

표 4-10 정부주도 빅데이터 전문인력 양성 추진과제

추진과제	세부내용
데이터 과학자급 전문인력 양성	<ul style="list-style-type: none"> <li>• 대학원의 ITRC(IT Research Center)를 통해 빅데이터 전문 ITRC 센터를 지정하여 운영·지원</li> <li>• ITRC는 IT 분야의 고급 석·박사 양성을 통해 국가가 추진하는 신성장 동력 산업 발전에 필요한 전문인력을 집중 육성하고 핵심 기술에 대한 연구 역량을 강화</li> </ul>
현업 실무 전문가 중심의 빅데이터 전문인력 양성	<ul style="list-style-type: none"> <li>• 빅데이터 전문가를 양성하기 위한 빅데이터 아카데미 운영 및 실무 현업 사례와 유사한 빅데이터 파일럿 프로젝트 기획·개발</li> <li>• CEO-CIO의 인식개선을 통한 데이터 활용 제고를 위한 공개강좌</li> <li>• 여건상 교육 참석이 불가능한 기업 대상 현장방문 교육 지원</li> <li>• 교육 연수 중 실시한 우수 프로젝트 사례를 발굴하여 향후 신규 비즈니스 창출 방향 제시를 위한 사례 발표회 개최</li> </ul>
대학대학원의 빅데이터 잠재인력양성 지원	<ul style="list-style-type: none"> <li>• 빅데이터전력센터(K-ICT 빅데이터 센터)는 대학대학원 등에서 빅데이터 교육 지원을 위해 커리큘럼 보급 및 실습 인프라 제공</li> <li>• 다양한 빅데이터 사례를 분석해 볼 수 있도록 실습에 필요한 데이터 세트, 분석 시나리오와 기법 등을 개발, 각 대학 등에 보급</li> <li>• 빅데이터 교육 과정을 운영하는 대학대학원 등과 협의체를 구성하여 관련 정보 등 공유</li> </ul>

‘빅데이터 산업 발전 전략’ 중 빅데이터 전문인력 양성 추진과제(2013)

대학·대학원의 빅데이터 잠재인력 양성 지원을 위해 데이터 전략센터(K-ICT 빅데이터 센터)는 빅데이터 교육 지원을 위한 커리큘럼 보급, 교육에 필요한 실습 인프라를 제공하는 것과 함께 다양한 빅데이터 사례를 분석해 볼 수 있도록 실습에 필요한 데이터 세트, 분석 시나리오와 기법 등을 개발하여 대학·대학원에 보급하고, 빅데이터 교육과정을 운영하는 대학·대학원 등과 협의체를 구성하여 관련 정보 등을 공유하고 있다.

## ② 고용노동부의 직업능력 개발 지원

고용노동부는 졸업예정자, 실업자, 재직자 등 대상별 직업능력 개발을 지원하는 다양한 사업들을 추진하고 있는데, 최근 들어 사업별로 데이터 관련 교육과정이 많이 개설되고 있는 추세이다. 특히, '청년취업 아카데미'사업의 일환으로 (사)한국정보과학진흥협회, (사)한국IT비즈니스진흥협회, 한국소프트웨어산업협회, 한국ICT융합협동조합, 한국생산성본부, 단국대학교 등 산업별협회나 협동조합 또는 교육기관을 참여기관으로 하여 다양한 데이터 관련교육과정을 운영하고 있다. 이밖에 기업교육기관과 각종교육센터 및 훈련기관이 주관하는 '국가기간·전략산업 직종훈련', 민간 교육기관과 협업으로 추진하는 '내일배움카드', 표준협회 등과 함께 추진하는 '중소기업 핵심직무능력향상 지원사업' 등을 추진하고 있다.

## 3) 대학주도 교육과정

### ① 학부과정

최근 국민대, 충북대, 성균관대 등에서 데이터 관련 학부과정을 개설하였으며, 대부분 기존 학과 내 세부 전공 또는 연계 전공으로 운영 중이다. 소속 단과대는 △경영대학 △IT대학 △문과대학 등 다양하며, 학과 및 학위명도 학교별로 차이를 보이고 있다.

표 4-11 국내 데이터 관련 학부과정 개설 현황(일부)

학교명	단과대학	학과명	학위명	설립년도
국민대학교	경영대학	경영학부 빅데이터 경영통계전공	경영학사	2013
대구의원대학교	글로벌경영대학	데이터경영학과	경영학사	2015
동의대학교	자연과학대학	데이터정보학과	통계학사	2006
서울여자대학교	단과대 구분 없음	데이터과학전공	공학사	2016
충북대학교	경영대학	경영정보학과 빅데이터 전공	경영학사	2015
성균관대학교	문과대학	데이터사이언스학과	미정	2015

※ 자료 : 대학알리미 홈페이지(<http://www.academyinfo.go.kr/>)

학부의 정규 교과과정은 산업에서 요구하는 빅데이터 전문인력 역량에 맞춰 융합형·실무형 인재육성을 지향하고 있다. 최근에는 개설한 강의의 50% 이상이 기업체 실무 전문가가 참여하여 수업을 진행하는 방식으로 변모하고 있으며, 특히 학생들이 대학 졸업 후에도 실무경험 부족으로 취업의 어려움을 겪는 점을 감안해 과거와 달리 인턴십 및 교육을 확대해 나가고 있는 추세이다.

## ② 대학원 과정

2012년 설립된 충북대학교 비즈니스 데이터융합학과를 비롯해 10개 이상의 대학원에서 데이터 관련 교육과정을 운영하고 있으며 현재도 지속적으로 개설되고 있다.

표 4-12 국내 대학원 데이터 전문인력 양성 과정 현황(일부)

학교명	단과대학	학과명	학위명	설립년도
충북대학교	단과대 구분 없음	비즈니스 데이터융합학과	공학석사	2012
	일반대학원	빅데이터학과	공학석·박사	2015
국민대학교	경영대학원	빅데이터경영MBA	경영학석사	2013
	일반대학원 인문사회계열	데이터사이언스학과	경영학석·박사	2014
서울과학기술대학교	일반대학원 공학계열	데이터사이언스학과	공학석사	2014
연세대학교	정보대학원	빅데이터학과 (SMB)	정보시스템학석사	2014
이화여대	일반대학원	빅데이터분석학과	데이터분석석·박사	2014
고려대학교	정보대학원	빅데이터응용 및 보안학과	공학석사	2015
단국대학교	일반대학원 공학계	데이터사이언스학과	데이터사이언스학석사	2015
과학기술연합 대학원 대학교	단과대 구분 없음	빅데이터과학	석·박사	2014

※ 자료 : 대학알리미 홈페이지(<http://www.academyinfo.go.kr/>)

학부과정과 마찬가지로 소속이 다양하지만 주로 정보대학원이나 일반대학원 내에 있으며 경영대학원 내 MBA 과정으로 개설된 곳도 있다. 교육과정은 학교별로 조금씩 다르지만, 대부분 20명 이하의 소수 인원을 대상으로 진행되고 있다. 졸업 시 보통 공학이나 경영학 학위가 수여되지만, 일부 대학원에서는 데이터 분석학 또는 데이터 사이언스학 학위를 수여하기도 한다.

연세대학교 빅데이터학과는 기업과 연계해 학비를 지원해 주며, 기업 맞춤형 인재교육에 초점을 둔 재직자 대상의 '계약학과' 형태로 운영되고 있다. 교육은 상시근로자 5인 이상 사업장에 재직 중이며 기업에서 50% 이상 등록금 지원이 가능한 자를 대상으로 하며, 오프라인 강의·사례연구·실습·프로젝트가 진행된다.

표 4-13 연세대학교 빅데이터학과의 재직자 대상 계약학과 교육과정

교육과정	교육내용	
디지털비즈니스	△소셜미디어 분석론 △산업별 빅데이터 응용 △빅데이터 비즈니스 전략 △빅데이터 관련 법과 제도	△비즈니스환경과 소셜네트워크 △IT 비즈니스 다이내믹스 △프로젝트 관리론 △빅데이터 관리 체계
컴퓨터 공학	△빅데이터 컴퓨팅 △대용량 멀티미디어 자료 처리 △디지털 포렌식 △빅데이터 처리 표준 및 제도	△하둡 플랫폼 운용 △데이터 프로그래밍 실무 △네트워크 보안
응용 통계학	△비즈니스 환경의 예측적 분석론 △소셜애널리틱스 △데이터 최적화와 시뮬레이션	△R-데이터마이닝 △빅데이터 시각화 기법 △빅데이터 분석 및 응용
경영학	△마케팅 엔지니어링	△빅데이터의 경영 적용

※ 자료 : 연세대학교 빅데이터학과 홈페이지

국내 대학 및 대학원에서 운영하는 데이터 관련 정규 교과과정은 최근 산업에서 원하는 인재상에 걸맞게 실무형·학문 융합형으로 변모하고 있다. 기업의 데이터 관련 업무는 팀 단위로 이루어지며, 모든 직무역량을 갖춘 인력보다 세분화된 전문역량을 갖춘 인재를 선호하기 때문에 대학과 대학원의 교과과정도 세분화되고 실무 중심으로 편성되어야 한다.



표 4-14 국내 데이터 관련 학부 및 대학원 과정 개설 현황

학교		단과대 및 소속	정원	학위	설립년도
학부	국민대학교	경영대학 경영학부 빅데이터 경영통계전공	50명	경영학사	2013
	강원대학교	IT대학 컴퓨터정보통신공학과 데이터 분석 전공	60명	공학사	2014
	충북대학교	경영정보학과 빅데이터 전공	미정	경영학사	2015
	성균관대학교	문과대학 데이터사이언스학과	미정	미정	2015
대학원	충북대학교	비즈니스 데이터융합학과	11명	공학석사	2012
	국민대학교	경영대학원 빅데이터경영MBA	15명	경영학석사	2013
	국민대학교	일반대학원 인문사회계열 데이터사이언스학과	5명	경영학석·박사	2014
	서울과학기술대학교	일반대학원 공학계열 데이터사이언스학과	10명	공학석사	2014
	연세대학교	정보대학원 빅데이터학과 (SMBA)	15명	정보시스템학석사	2014
	이화여대	일반대학원 빅데이터분석학과	5명	데이터분석학석·박사	2014
	고려대학교	정보대학원 빅데이터응용 및 보안학과	20명 내외	공학석사	2015
	단국대학교	일반대학원 공학계 데이터사이언스학과	유동적	데이터사이언스학 석사	2015



#### 4) 민간주도 교육과정

국내 민간 기업에서 운영 중인 데이터 전문인력 양성과정은 주로 IT분야의 대기업 혹은 서비스 공급기업에서 기업부설 교육센터나 교육 전문기관을 통해서 운영되고 있다. 대부분의 민간 교육기관은 정부지원 교육사업을 위탁 운영하고 있으며, 단기간 내 업무수행에 필요한 직무역량 및 전문성을 습득시키기 위한 프로그램 중심으로 교육을 하고 있다.

표 4-15 삼성 SDS 멀티캠퍼스의 빅데이터 아카데미 과정

구분	입문	핵심	고급
기술 전문가	NoSQL 데이터 모델링	실시간 빅데이터 분석	—
		MongoDB! 개발부터 관리까지	
		클라우드 컴퓨팅을 위한 하둡(Hadoop)	
분석 전문가	빅데이터 분석과 시각화 입문	R을 이용한 데이터 분석	R을 이용한 예측분석 모델링
		디지털영상 데이터 분석(VA)	Spark를 활용한 빅데이터 분석
	이정훈의 빅데이터 분석 기본	이정훈의 Excel Based Machine Learning	오픈소스를 활용한 머신러닝과 딥러닝
		빅데이터를 위한 파이썬	
공통 역량	빅데이터 분석과 시각화 입문	이정훈의 Excel Based Machine Learning	—
	이정훈의 빅데이터 분석 기본	핵심! 빅데이터 분석 전문가	

※ 자료 : 삼성 SDS 멀티캠퍼스 홈페이지(www.multicampus.co.kr)

#### ① KMR 빅데이터 분석 전문가 양성과정(그림 넓이 조정하고 앞으로 당김)



그림 4-5 KMR 빅데이터 분석 전문가 양성과정

표 4-16 교육과정 세부내용

교육모듈	교육일정	세부내용
빅데이터 분석을 위한 R프로그래밍	10월 6일	<ul style="list-style-type: none"> <li>• 데이터의 이해</li> <li>• 데이터 수집, 가공 그리고 활용</li> <li>• 빅데이터 분석 환경 구성</li> </ul>
R을 활용한 대용량 데이터 처리 및 조작	10월 13일	<ul style="list-style-type: none"> <li>• 외부데이터 불러오기</li> <li>• 데이터 핸들링</li> </ul>
R과 SQL DB 연동 및 조작	10월 27일	<ul style="list-style-type: none"> <li>• Mysql</li> <li>• AWS 예시</li> </ul>
빅데이터 패턴 탐색 및 시각화	11월 3일	<ul style="list-style-type: none"> <li>• 패턴인식의 이해</li> <li>• 데이터 시각화 기본개념과 기법</li> </ul>
데이터 마이닝 개념과 분석기법	11월 10일	<ul style="list-style-type: none"> <li>• 패턴인식과 사후분석의 이해</li> <li>• 회귀분석과 분산분석</li> </ul>
데이터 분석의 심화와 리포팅	11월 17일	<ul style="list-style-type: none"> <li>• 모델평가 기법의 이해</li> <li>• 리포팅의 이해와 활용</li> </ul>

우리나라 정부·공공기관 주도 프로그램을 기반·기술·분석·사업 역량을 기준으로 구분하였을 경우, 대부분 기술·분석 역량 중심으로 교육이 이뤄지고 있으며, 기반역량이나 사업역량을 강화하기 위한 과목은 다소 부족하다. 반면 해외의 경우 우리나라에 비해 4개 역량별 과목이 비교적 골고루 분포되어 있다. 따라서 향후 데이터 과학자 수요에 대응하기 위해서는 기반역량과 사업역량을 강화하기 위한 인력양성 프로그램이 다수 개발되어야 한다.

향후 교육수준의 향상과 기업체의 다양한 요구 수준을 감안했을 때 고급인력에 대한 수요는 급격히 증가할 것으로 예상되므로, 추후 '기반·사업역량 강화를 위한 고급프로그램 확대'가 필요하다.

다양한 주제를 중심으로 교육기회를 확대하고, 데이터 전문 인력을 양성하는 교육주체를 보다 다양화할 필요가 있다. 해외 사례에서 보듯이 일반 국민과 학생 및 시민을 대상으로 데이터의 중요성과 활용방법을 교육하는 것은 데이터 산업의 저변을 확대하고 궁극적으로 데이터 산업의 발전을 이끌어 내기 위한 것이다. 따라서 정

부, 공공기관 뿐만 아니라 다양한 주체에 의한 데이터 교육의 확대가 매우 중요하다.

또한 해외 전문 인력 유입환경을 통해 데이터산업의 수준을 업그레이드 할 수 있으므로 국내 전문 인력의 양성과 함께 해외 전문 인력을 유입할 수 있는 정책이 필요하다. 해외 전문 인력의 국내 유입을 확대하고, 국내 전문 인력의 해외 유출은 최소화하는 것이 데이터산업 육성의 가장 큰 과제라고 할 수 있다

빅데이터 R&D와 인력양성 융합을 위해 단순한 강좌나 실습에서 한걸음 더 나아가 산업체가 당면한 애로기술을 빅데이터를 이용해서 풀어 나가는 R&D프로젝트를 통한 데이터 고급인력 양성 프로그램을 운영하는 것도 필요하다.

온라인 데이터 인력 플랫폼을 구축하고 운영함으로써 전국 규모의 데이터 인력 정보 공유를 통한, 전문 인력과 기업이 서로 소통할 수 있는 방안을 모색해야 한다.

#### 4. 빅데이터 전문인력 직무 분석

##### 1) 데이터 산업 및 전문인력 정의

IT시대에서 DT(: Data Technology)시대로 접어들면서 데이터의 분석과 활용의 중요성이 점차 높아지고 있다. 이에 데이터산업은 데이터의 축적과 데이터베이스 구축에 초점을 둔 전통적인 데이터베이스산업의 테두리에서 벗어나 데이터를 기반으로 한 새로운 비즈니스 영역까지를 포함하는 개념으로 인식되고 있다. 즉, 데이터산업은 전통적인 데이터베이스산업과 더불어 빅데이터, 클라우드 및 사물인터넷 등 최근 떠오르는 신기술 시장 영역을 포함하여, '데이터의 생산·수집·처리·관리·유통·분석·활용 등을 지원하거나 이와 관련된 제품과 서비스를 제공하는 산업'으로 정의할 수 있다.

기존 데이터베이스산업에서 데이터산업으로 의미가 확장·진화해감에 따라, 데이터 전문 인력의 직무도 데이터의 저장과 자료의 축적 중심에서 데이터의 분석 및 활용 중심으로 변화하여, 최근에는 데이터 분석가, 데이터 과학자와 같은 직무 비중이 점차 늘어나고 있다.

한국데이터진흥원에서 발간한 '2015년 데이터산업 현황조사'에서는 데이터 전문 인력을 다음과 같이 정의하고 있다. “기업에서 필요한 데이터베이스를 기획하거나 생성된 데이터나 데이터베이스를 관리하고, 운영하는 인력을 말하며, 또한 기업에서 활용하는 데이터 및 데이터베이스의 보안·성능·품질 향상을 위한 인력까지 포함한다. 기존에 포함하였던 '데이터 입력자'와 같은 단순 업무 종사자는 제외한다.”

데이터산업의 정의에 기반하여 데이터 전문 인력을 정의하면, '데이터의 생산·수집·처리·관리·유통·분석·활용 등을 지원하거나 이와 관련된 제품과 서비스를 제공하는 모든 인력 중 '데이터 입력자'와 같은 단순 업무자를 제외한 모든 전문인력'이라고 할 수 있다. 한국데이터진흥원은 데이터 전문인력에는 △데이터 아키텍트(DA) △데이터 개발자 △데이터 엔지니어 △데이터 분석가 △데이터베이스 아키텍트(DBA) △데이터 과학자 △데이터 컨설턴트 △데이터 마케터 등이 포함되어 있다고 정의한다.

## 2) 데이터 전문인력 직무 및 역량

### ① 국내의 데이터 직무 구분

한국데이터진흥원이 발표한 '2015년 데이터산업 현황 조사 결과 보고서'에 따르면 DA(: Data Architect), DB개발자, 빅데이터 개발자, DB엔지니어, DBA(: Database Administrator) 등 10가지로 구분하고 있다.

표 4-17 데이터 직무 분류

DA	데이터 개발자		데이터 엔지니어		데이터 분석가	DBA	데이터 과학자	데이터 컨설턴트	데이터 마케터
	DB 개발자	빅데이터 개발자	DB 엔지니어	빅데이터 엔지니어					

한국데이터진흥원은 △DBA △DB개발자 △데이터 컨설턴트 등 전통적인 직무를 비롯해 최근 수요가 급증하고 있는 △데이터 과학자 △빅데이터 개발자 △빅데이터 엔지니어 등 빅데이터 관련 직무를 데이터 직무에 포함하고 있다.

한국정보화진흥원에서 수행한 '2015년 빅데이터 시장현황 조사' 자료에 의하면, 데이터 전문 인력을 △하둡/NoSQL/MapReduce엔지니어 △빅데이터 분석가 △빅데이터 기술영업/마케터 △빅데이터 컨설팅 △빅데이터 기타인력 등 5가지 유형으로 분류하고 있다.

빅데이터 개발자는 하둡/NoSQL/MapReduce엔지니어 직무에 포함된 것으로 보이며, 데이터 과학자는 별도의 직무로 분류하지 않았으나 상당수 빅데이터 분석가로 분류한 것으로 추측된다. 그 밖에 빅데이터 기술영업/마케터, 컨설팅 등 새로운 직무가 세분화되고 있으나, 기존 직무와 중복되는 업무영역이 있는 등 분류기준이 아직 명확하게 정립되지 않은 상태라는 것을 알 수 있다.

고용노동부 국가직무능력표준(NCS : National Competency Standards)의 데이터 관련 직무는 아직까지 '빅데이터 분석', 'DB 엔지니어링' 직무로만 구분하고 있다.

‘빅데이터 분석’ 직무는 능력단위를 △빅데이터 분석 기획 △빅데이터 수집 △빅데이터 저장 △빅데이터 처리 △분석용 데이터 탐색 △통계 기반 데이터 분석 △머신러닝 기반 데이터 분석 △텍스트마이닝 기반 데이터 분석 △빅데이터 분석 결과 시각화 등 9개 항목으로 구분하고 있다. ‘DB엔지니어링’ 직무는 능력단위를 △데이터베이스 요구사항 분석 △개념 데이터 모델링 △논리 데이터베이스 설계 △물리 데이터베이스 설계 △데이터베이스 구현 △데이터 품질관리 △데이터 전환 설계 △데이터 전환 △데이터베이스 성능확보 △데이터 표준화 △SQL활용 △SQL응용 등 12개 항목으로 구분하고 있다. 그 외 △데이터 개발 △데이터 기획 △데이터 기술지원 △데이터 기술영업 △데이터 관리 등 다른 데이터 관련 직무는 아직 별도로 분류하지 않고, 일반적인 IT분야 직무나 사무 행정 직무에 포함하고 있어, 데이터 분야의 성장에 따른 직무 및 능력단위에 대한 세분화가 과제로 남아 있다.

표 4-18 국가직무능력표준(NCS)의 빅데이터 관련 직무

빅데이터 관련 직무	능력단위	능력요소
빅데이터 분석	빅데이터 분석 기획	도메인 이슈 도출, 분석목표 수립, 프로젝트 계획, 보유데이터 자산 확인
	빅데이터 수집	데이터 수집 계획, 빅데이터 수집 시스템 구성, 내부 데이터 수집, 외부 데이터 수집, 데 이터 변환, 수집 데이터 검증
	빅데이터 저장	빅데이터 저장 계획 수립, 빅데이터 저장 모 델 설계, 빅데이터 저장 관리 시스템 구성, 빅 데이터 적재, 빅데이터 운영
	빅데이터 처리	빅데이터 처리 계획, 빅데이터 처리 시스템 구성, 분산병렬 수행, 실시간 수행, 이벤트 처 리 수행
	분석용 데이터 탐색	기본 통계 확인, 데이터 분포 분석, 변수간 관 계 확인, 데이터 정제, 데이터 변화 적재, 분 석용 데이터 검증

	통계 기반 데이터 분석	가설 설정, 빅데이터 모델 개발, 빅데이터 모델 평가 검증, 빅데이터 모델 운영방안 마련
	머신러닝 기반 데이터 분석	머신러닝 수행 방법 계획, 데이터 셋 분할, 지도학습 모델 적용, 자율학습 모델 적용, 모델 성능 평가, 학습결과 작용
	텍스트마이닝 기반 데이터 분석	텍스트 마이닝 기반 텍스트 분석 실행방법 계획, 텍스트 변환, 단어 사전 구축, 텍스트 분류 결과 분석, 정형 데이터 결합 분석
	빅데이터 분석 결과 시각화	분석 결과 스토리텔링, 분석 정보 시각화, 분석 정보 시각표현

학계에서도 데이터 직무를 체계화하려는 시도들이 계속되고 있다. 특히 빅데이터 활용 증가에 따라 빅데이터 직무 및 직무역량을 구분하려는 연구가 많이 시도되고 있다. 그 중 '빅데이터 직무능력 참조모형에 관한 융합적 연구'에서는 작업자 중심으로 직무를 구분하여 빅데이터 직무 유형을 크게 △빅데이터 기획 △빅데이터 분석 △빅데이터 관리 및 기술(지원) △빅데이터 활용 등 4개 분야로 나누고, 빅데이터 관련 직업능력을 지식, 기술, 태도로 나누어 설명하고 있다.

표 4-19 빅데이터 직무 능력 모형

빅데이터 직무	직무정의
빅데이터 기획	빅데이터 분석을 통해 의사결정과 실행과정에 필요한 정보와 통찰력을 도출하기 위해 분석기획 발굴, 필요한 분석과 데이터 정의 및 분석체계의 구체화를 계획하는 직무
빅데이터 분석	의사결정과 실행과정에 필요한 정보와 통찰력을 도출하기 위해 빅데이터 기획으로부터 정의된 분석 기회를 다양한 분석도구를 활용하여 데이터를 실제 분석하는 직무
빅데이터 관리 및 기술(지원)	빅데이터 분석을 통해 의사결정과 실행과정에 필요한 정보와 통찰력을 도출하기 위해 분석에 필요한 데이터를 수집 저장 관리 및 처리하고 이와 관련되는 H/W, S/W, 인프라 등을 지원하는 직무
빅데이터 활용	빅데이터 분석을 통해 의사결정과 실행 과정에 필요한 정보와 통찰력을 실제 업무에 활용하거나 사업에 적용하는 등의 수요측면에서의 직무

## ② 빅데이터 전문 인력 직무 구분

표 4-20 네덜란드 DATAFLOQ의 데이터 직무 구분

빅데이터 직무	필요 역량
데이터 총괄책임자 (Chief Data Officer)	<ul style="list-style-type: none"> <li>· 조직 내에서 빅데이터 전략에 대한 총괄 및 CEO에게 직접 보고 책임 조직에서 활용되는 빅데이터가 전략에 부합하도록 업무 추진·감독</li> <li>· 조직이 전략적으로 활용할 빅데이터의 우선순위 정의</li> <li>· 강한 리더십과 고위경영진 및 임원과의 소통 능력</li> <li>· 핵심 사업 영역에서의 정보 관리 리딩 경험</li> <li>· Hadoop, MapReduce 등 빅데이터 솔루션에 대한 전문성</li> <li>· 데이터 거버넌스 및 데이터 품질관리 경험</li> <li>· 시장에서 통용되는 주요 빅데이터 솔루션과 제품에 대한 이해도</li> <li>· 우수 사례와 방법론을 도입하여 추진할 수 있는 전문성</li> </ul>
빅데이터 과학자 (Big Data Scientist)	<ul style="list-style-type: none"> <li>· 한 분야에 국한되지 않는 광범위한 스킬</li> <li>· 비즈니스 전략뿐 아니라 알고리즘을 짜기 위한 통계학, 수학, 예측 모델링 능력</li> <li>· 분석을 통해 발견한 사항을 말과 글로써 효과적으로 소통하는 능력</li> <li>· 다양한 종류의 프로그래밍 언어 사용 능력</li> <li>· Hadoop, Hive 등에 대한 이해</li> <li>· 복수의 시스템과 데이터셋을 통합하는 방법에 대한 이해</li> <li>· 자연어 처리(Natural Language Processing)에 대한 이해</li> <li>· 알고리즘을 개발하기 위한 컴퓨터 사용 능력</li> <li>· 모델을 공유하기 위한 개념적 모델링(Conceptual modelling)</li> <li>· 통계적 분석 능력</li> </ul>
빅데이터 분석가 (Big Data Analyst)	<ul style="list-style-type: none"> <li>· 빅데이터 과학자를 보조하며 주어진 시스템 안에서 데이터를 다루며 분석 수행</li> <li>· 실시간 분석 및 Tableau S와 같은 비즈니스 인텔리전스 플랫폼에 대한 이해</li> <li>· SQL database, 프로그래밍 언어, R이나 SPSS와 같은 통계 소프트웨어 역량</li> </ul>



	<ul style="list-style-type: none"> <li>· 하둡과 맵리듀스에 대한 최소한의 기초 지식</li> <li>· 가용 데이터로부터 새로운 시사점(insight)을 개발하는 능력</li> </ul>
빅데이터 비주얼라이저 (Big Data Visualizer)	<ul style="list-style-type: none"> <li>· 빅데이터를 시각화하여 조직 또는 상사가 이해하기 쉽게 변환하는 역량</li> <li>· 유저 인터페이스, 타이포그래피에 대한 이해</li> <li>· 인터페이스 디자인, 사용자경험(UX) 디자인, 시각디자인에 대한 이해</li> <li>· 새로운 패턴과 시사점(insight)을 찾기 위한 창조적 사고방식</li> <li>· 빅데이터 과학자의 의도에 맞추기 위한 빅데이터 분석 과정에 대한 이해</li> </ul>
빅데이터 매니저 (Big Data Manager)	<ul style="list-style-type: none"> <li>· 조직 내 기술부서와 전략 부서를 연결하기 위한 양측 모두의 업무에 대한 이해</li> <li>· 빅데이터 분석의 결과를 조직의 상급 관리자에게 설명하는 능력</li> <li>· 조직의 전략과 비전을 빅데이터 부서에 이해시키기 위한 소통 능력</li> <li>· Python, R, Java, Ruby, SQL 등의 프로그래밍 언어에 대해 충분한 지식</li> <li>· Hadoop, Hive 또는 MapReduce 에 대한 최소한의 기초지식</li> <li>· 자연어처리, 기계학습, 개념 모델링, 통계 분석, 예측 모델링 경험</li> </ul>
빅데이터 솔루션 아키텍트 (Big Data Solution Architect)	<ul style="list-style-type: none"> <li>· 빅데이터 솔루션의 구조에 대한 이해</li> <li>· Hadoop과 같은 기술을 이용하여 빅데이터 솔루션을 제공하는 방법 이해 <ul style="list-style-type: none"> <li>• Hadoop 애플리케이션 실제 활용 경험</li> </ul> </li> <li>· Hadoop 솔루션의 수명주기 관리</li> <li>· 적절한 아키텍처 설계에 필요한 그 밖의 다양한 기술</li> </ul>
빅데이터 엔지니어 (Big Data Engineer)	<ul style="list-style-type: none"> <li>· 빅데이터 솔루션 아키텍트가 설계한 것을 구축하는 역량</li> <li>· 빅데이터 솔루션의 개발, 유지관리, 테스트, 평가 역량</li> <li>· 일간 기준으로 페타바이트 또는 엑사바이트 크기의 데이터를 다루는 역량</li> <li>· Java, Linux, C++, Ruby, Python 또는 R 등 다양한 프로그래밍 언어 지식</li> <li>· 일반적으로 큰 규모의 데이터를 수집, 파싱, 관리, 분석, 시각화하고 복수의 플랫폼을 사용하여 정보를 시사점(insight)으로 전환시키는 역량</li> </ul>

빅데이터 연구원 (Big Data Researcher)	<ul style="list-style-type: none"> <li>· 빅데이터의 활용 방안에 대한 심층 연구 수행 역량</li> <li>· 빅데이터 기술로 다양한 분야에서 새로운 비즈니스 또는 제품 개발 기회 발굴</li> <li>· 다양한 빅데이터 기술과 데이터베이스, 프로그래밍 언어에 대한 견고한 이해</li> <li>· Research 프로젝트를 통해 텍스트 마이닝, 군집분석, recommendation analysis, 극단값 확인, 예측 분석 또는 이와 유사한 기법 수행</li> <li>· 예산과 기획을 포함하는 프로젝트 관리 능력</li> </ul>
빅데이터 컨설턴트 (Big Data Consultant)	<ul style="list-style-type: none"> <li>· 빅데이터 및 빅데이터 전략에 대해 조언</li> <li>· 비즈니스 경험과 함께 빅데이터 툴에 대한 광범위한 기술적 지식</li> <li>· 빅데이터 전략을 설계하고 빅데이터를 수집, 저장, 분석 시각화하는 능력</li> <li>· Hadoop, Hive, HBase, MongoDB 또는 MapReduce 등에 대한 이해</li> </ul>

### 3) 국내외 데이터 전문인력 직무 구분 비교

국내외 데이터 직무를 구분하고 있는 주요 기관들을 살펴본 결과, 직무 개수와 각각의 직무 명칭이 다소 상이한 것으로 나타났다. 또한 분류된 개별 직무가 하나의 업무 영역 혹은 업무 프로세스를 담당하기도 하지만, 경우에 따라서는 여러 업무 영역 혹은 업무 프로세스에 걸쳐 발생하며, 그 기준도 기관별로 다양한 형태로 나타나고 있다.

국내외 데이터 직무를 단순 비교해 보면, 국내는 데이터 분석 관련 직무를 '데이터 분석가'(Data Analyst) 또는 '데이터 과학자'(Data Scientist)로 단순화시키고 있지만, 해외에서는 여기에 더하여 '데이터 비주얼라이저'(Data Visualizer), '데이터 디자이너'(Data Designer), '데이터 연구원'(Data Researcher) 등으로 보다 세부적으로 구분하고 있다. 데이터 분석 관련 직무가 늘어나고 있

고, 종합적인 데이터 설계와 시각화의 중요성도 높아지고 있어 국내에서도 보다 세분화된 데이터 분석 관련 직무 분류와 직무별 맞춤형 교육과정 개발이 필요하다.

국내에는 일반적으로 관리업무가 대부분 별도의 직무로 구분되고 있지 않은 반면, 해외에는 △데이터 총괄 책임자(Chief Data Officer)를 비롯해 △데이터 관리자(Data Administrator) △빅데이터 매니저(Big Data Manager) △빅데이터 프로젝트 매니저(Big Data Project Manager)등으로 세분화 되어 있다. 향후 데이터 활용 확대에 따라 기획·관리업무의 중요성이 더욱 커질 것은 자명하므로, 관련 역량을 키우기 위한 직무도 세분화할 필요가 있다.

표 4-21 국내외 데이터 직무 구분 비교표

구분		한국 데이터진흥원	한국 정보화 진흥원	고용 노동부	학계	DATA FLOQ	SAS
데이터설계		데이터 아키텍트	하둡 /NoSQL/ Map Reduce 엔지니어	DB 엔지니어 링	빅데이 터 관리 및 기술(지 원	빅데이터 솔루션 아키텍트	빅데이터 아키텍트
데이터 개발		데이터 개발자				빅데이터 엔지니어	빅데이터 개발자
데이터 운영관리		데이터 엔지니어					
데이터 활용	데이 터분 석	데이터분석가/ 데이터 과학자	빅데이터 분석가	빅데이터 분석	빅데이 터 분석	빅데이터 분석가/ 빅데이터 과학자	빅데이터 분석가/ 데이터 과학자
	데이 터시 각화					빅데이터 비주얼라이 저	빅데이터 디자이너
	데이 터연 구					빅데이터 연구원	데이터 과학자

데이터 관리	데이터 과학자/ DBA	-		빅데이 터 활용/ 빅데이 터 관리 및 기술(지 원	총괄책임자 / 빅데이터 매니저/ 빅데이터 과학자	프로젝트 매니저/ 데이터 과학자/ 빅데이터 관리자
데이터 판매	데이터 마케터	빅데이터 기술영업/ 마케터	-	-	-	-
데이터 컨설팅	데이터 컨설턴트	빅데이터 컨설턴트	-	빅데이 터 기획	빅데이터 컨설턴트	-

#### 4) 데이터 전문 인력 직무 및 필요 역량

데이터산업에 있어 기술과 활용의 형태가 매우 빠르게 변화함에 따라, 직무 분류가 아직 하나의 큰 줄기로 수렴되지 못하고 기관의 기능과 특성에 따라 다양하게 나타나고 있다. 여러 기관이 분류한 직무를 하나의 기준에 따라 재분류하는 것은 각 분류체계의 목적을 간과하여 본질을 왜곡할 수 있고, 다양성이 강한 데이터 산업의 특성에 부합하지 않을 수 있기 때문에 주의가 필요하다. 데이터 산업의 중요성이 날로 부각되고 있고, 다양한 분야에서 인력 수요가 부족한 실정에서 현장에 곧바로 투입될 수 있는 인력을 적시에 양성하기 위해서는 현재의 직무분류체계 보다 더욱 세분화된 직무분류체계와 직무별 필요역량을 구체적으로 정의할 필요가 있다.

표 4-22 데이터 직무구분 및 필요 역량

직무구분	기술적 역량	비즈니스 역량
데이터 아키텍처	<ul style="list-style-type: none"> <li>· Oracle, Java, SQL, Hadoop, SQL Server에 대한 이해</li> <li>· 데이터 모델링, ETL, EA, 오픈 소스 애널리틱스 활용 역량</li> <li>· Hadoop 애플리케이션 실제 활용 경험과 Hadoop 솔루션의 수명주기 관리 역량</li> </ul>	<ul style="list-style-type: none"> <li>· 빅데이터 솔루션의 구조에 대한 이해를 바탕으로 Hadoop과 같은 기술을 이용하여 빅데이터 솔루션을 제공하는 역량</li> <li>· 적절한 아키텍처 설계에 필요한 그 밖의 다양한 역량</li> </ul>
데이터 엔지니어	<ul style="list-style-type: none"> <li>· DB를 이용하여 프로그래밍을 하는 직무로 JavaScript, SQL, Linux, Oracle, Hadoop, C++, Ruby, Phyton 등 다양한 프로그래밍 언어를 활용하여 빅데이터 처리, LOD 구축 등 데이터 개발 역량</li> <li>· 일간 기준 페타바이트 또는 엑사바이트 크기 데이터 취급 역량</li> </ul>	<ul style="list-style-type: none"> <li>· 빅데이터 솔루션 아키텍처가 설계한 것을 구축하고 빅데이터 솔루션의 개발, 유지관리, 테스트, 평가 역량</li> <li>· 큰 규모의 데이터를 수집, 저장, 관리, 분석하고 복수의 플랫폼을 사용하여 정보를 시사점(insight)으로 전환시키며 이와 관련되는 H/W, S/W, 인프라를 지원하는 역량</li> </ul>
데이터	· Oracle, SQL, Java 프로그램에	· 데이터 과학자를 보조하며 주

분석가	<p>대한 이해를 바탕으로 통계/머신러닝/텍스트마이닝 기반 데이터 분석 역량</p> <ul style="list-style-type: none"> <li>· 실시간 분석 및 Tableau SW와 같은 비즈니스 인텔리전스 플랫폼 이해</li> <li>· SQL database, 프로그래밍 언어, R이나 SPSS와 같은 통계 소프트웨어 활용 역량</li> </ul>	<p>어진 시스템 안에서 다양한 분석도구를 활용하여 데이터를 실제 분석하는 역량</p> <ul style="list-style-type: none"> <li>· 하둡과 맵리듀스에 대한 최소한의 기초 지식을 가지고 가용 데이터로부터 새로운 시사점(insight)을 개발하는 능력</li> </ul>
데이터 비주얼라이저	<ul style="list-style-type: none"> <li>· 분석 결과 스토리텔링, 분석 정보 시각화, 분석 정보 시각표현</li> <li>· 유저 인터페이스, 타이포그래피에 대한 이해</li> <li>· 인터페이스 디자인, 사용자경험(UX) 디자인, 시각디자인에 대한 이해</li> </ul>	<ul style="list-style-type: none"> <li>· 데이터를 시각화하여 조직 또는 상사가 이해하기 쉽게 변환하는 역량</li> <li>· 새로운 패턴과 시사점(insight)을 찾기 위한 창조적 사고방식을 갖추고 데이터 과학자의 의도에 맞추기 위한 데이터 분석 과정 이해</li> </ul>
데이터 과학자	<ul style="list-style-type: none"> <li>· Hadoop, Java, NoSQL, C++, 파이썬, R 등 다양한 프로그래밍언어 사용 능력</li> <li>· 통계학, 분석학(Analytics), 수학 등 한 분야에 국한되지 않는 지식</li> <li>· 알고리즘을 개발하기 위한 컴퓨터 사용 및 모델을 공유하기 위한 개념적 모델링 능력</li> <li>· Research 프로젝트를 통해 텍스트 마이닝, 군집분석, recommendation analysis, 극단값 확인, 예측 분석 또는 이와 유사한 기법 수행 능력</li> <li>· 복수의 시스템과 데이터 셋을 통합하는 방법에 대한 이해</li> </ul>	<ul style="list-style-type: none"> <li>· 다양한 내외부 소스로부터 데이터를 추출하고 가장 시급한 과제에 대한 데이터 기반 솔루션 고안</li> <li>· 조직내부 및 외부 데이터를 관리/활용/분석하는 체계를 만들고, 프로세스 혁신 및 신제품 개발, 마케팅 전략 결정 등 의사결정 유도</li> <li>· 효과적인 데이터 시각화 및 보고를 통해 예측과 결과를 경영진과 IT 부서에 통보</li> <li>· 데이터 기술로 다양한 분야에서 새로운 비즈니스 또는 제품개발 기회 발굴 및 예산과 기획을 포함하는 프로젝트 관리 및 자문 능력</li> </ul>

프로젝트 매니저 (표배경 색 통일)	<ul style="list-style-type: none"> <li>· Oracle, Netezza, Business Objects and Hyperion</li> <li>· ETL, Agile Software Development</li> <li>· Python, R, Java, Ruby, SQL 등의 프로그래밍 언어에 대한 지식</li> <li>· Hadoop, Hive 또는 MapReduce에 대한 기초지식</li> <li>· 자연어처리, 기계학습, 개념 모델링, 통계 분석, 예측 모델링 경험</li> </ul>	<ul style="list-style-type: none"> <li>· 필요한 분석과 데이터 정의 및 분석체계의 구체화를 계획하는 직무</li> <li>· 조직 내 기술부와 전략 부서의 업무에 대해 이해와 소통능력</li> <li>· 데이터 분석의 결과를 조직의 상급 관리자에게 설명하는 능력</li> </ul>
DBA (표배경 색 통일)	<ul style="list-style-type: none"> <li>· Linux, MySQL, Puppet, Hadoop, Oracle</li> <li>· Disaster Recovery, Configuration Management, Clustering, ETL 관련 역량</li> <li>· Hadoop, MapReduce 등 빅데이터 솔루션에 대한 이해를 바탕으로 한 데이터 거버넌스 구축 및 데이터 품질관리 역량</li> </ul>	<ul style="list-style-type: none"> <li>· 조직 내 데이터 전략 총괄 및 데이터 업무 추진 감독</li> <li>· 강한 리더십을 바탕으로 한 고위경영진 및 임원과의 소통능력</li> <li>· 데이터의 우선순위 결정 및 핵심사업에서 정보 관리 리딩 경험</li> <li>· 시장에서 통용되는 주요 데이터 솔루션과 제품에 대한 이해</li> </ul>

표 4-23 빅데이터 역량 정의

빅데이터 역량그룹-역량요소		역량정의	참조모델: 빅데이터 역량모델(NIA: BCM)
기획	환경 분석	빅데이터 정책 및 적용사례와 기술 환경을 이해하고, 빅데이터의 비즈니스 활용분야와 빅데이터 기술 동향에 대한 지식을 기반으로 빅데이터 아키텍처 설계 및 향후 발전 방향에 대한 전략을 도출할 수 있는 역량	빅데이터 비즈니스 이해, 빅데이터 트렌드, 빅데이터 플랫폼이론, 빅데이터 아키텍처, 전략적 사고방식, 빅데이터정책

	과제정의	業에 대한 지식과 기본적인 통계 및 수리적 사고를 기반으로 빅데이터 활용 아이디어를 도출하고, 과제화 할 수 있는 역량	통찰과 소통, 기초 통계이론, 수리적/정량적 사고방식, 업종특화지식, 업무 프로세스지식, 빅데이터 최적화모델링
	협업	빅데이터 사업기회를 찾아내고, 정보 공유 및 활용방안을 제시하며 대내외적으로 이해관계자와 협력하고 협상할 수 있는 역량	설득과 협상, 빅데이터 리더십
	성과관리	빅데이터 분석모형에 대한 이해와 빅데이터 적용결과에 대한 성과평가를 하고, 경영진 및 이해관계자에게 보고할 수 있는 역량	분석모형 및 성과평가, 의사결정 및 성과관리
처리	데이터수집	빅데이터 플랫폼에 대한 지식과 빅데이터 수집유형, 수집방법, 수집기술을 이해하고, 사용할 수 있는 역량	빅데이터 플랫폼 이론, 빅데이터 수집
	데이터저장	빅데이터 플랫폼에 대한 지식과 빅데이터 저장유형, 저장방법, 저장기술을 이해하고, 사용할 수 있는 역량	빅데이터 플랫폼 이론, 빅데이터저장 및 관리, 빅데이터 처리 및 분석, 빅데이터플랫폼 구축 및 활용
	데이터표준화	메타데이터관리와 마스터데이터에 대한 지식과 관련 솔루션을 기반으로 데이터품질 개선을 위한 데이터표준화 활동을 수행할 수 있는 역량	-
	데이터통합	데이터모델링 및 데이터웨어하우스에 대한 지식과 데이터아키텍처를 설계하고, 구현 및 운영을 총괄할 수 있는 역량	빅데이터 아키텍처
분석	요건정의	業에 대한 지식과 기본적인 통계 및 수리적 사고를 기반으로 빅데이터 분석요구사항과 빅데이터 운영시나리오를 정의할 수 있는 역량	통찰과 소통, 기초 통계이론, 수리적/정량적 사고방식, 업종특화지식, 업무 프로세스지식



	분석모델링	통계 및 수학지식을 기반으로 통계패키지를 활용하여 예측모델, 분석모델, 최적화모델 등을 설계할 수 있는 역량	통계패키지 활용, 분석적 마인드, BI, 데이터마이닝, 비정형데이터마이닝, 빅데이터 예측분석, 업종특화지식, Business Analytics
	모델구현	분석모델을 이해하고, 통계패키지와 빅데이터 프로그래밍 기술을 활용하여 시스템을 구현할 수 있는 역량	빅데이터 프로그래밍, BI, 데이터마이닝, 비정형데이터마이닝
	결과해석	빅데이터 분석모델 결과를 해석하여 통계적 의미를 부여할 수 있는 역량	논리적 자기표현, 분석모형 및 성과평가
시각화	데이터 표현	빅데이터 분석모델 결과를 해석하고, 논리적으로 표현할 수 있는 역량	논리적 자기표현, 기초통계이론
	시각화 구현	시각화 도구를 이용하여 분석결과를 화면으로 구현할 수 있는 역량	빅데이터 프로그래밍
운영	시스템 운영	빅데이터 플랫폼을 이해하고, 백업 및 복구기술 등을 적용하여 시스템의 안정적 운영을 유지할 수 있는 역량	빅데이터 플랫폼이론, 빅데이터 플랫폼 구축 및 활용, 빅데이터 아키텍처
	데이터 용량관리	빅데이터의 용량을 예측하고, 용량산정 결과를 기반으로 데이터 저장장치 및 저장방법을 선택하고 적용할 수 있는 역량	-
	데이터 보안	빅데이터 보안과 개인정보보호 등에 대한 마인드를 갖고 있으며, 데이터보안기술 지식을 기반으로 데이터 보안을 실현할 수 있는 역량	빅데이터 윤리의식

## 5. 중앙전파관리소의 빅데이터 조직 구성 방안

### 1) 빅데이터 조직 설계 방향

빅데이터 조직은 비즈니스가치 창출에 기여하고, 현업 요구에 신속 대응하며, 분산된 역량을 집중화 할 수 있도록 명확한 역할 정의와 협업체계 및 빅데이터 전담조직으로서의 효과적인 역할 수행을 위한 조직을 설계하여야 한다.

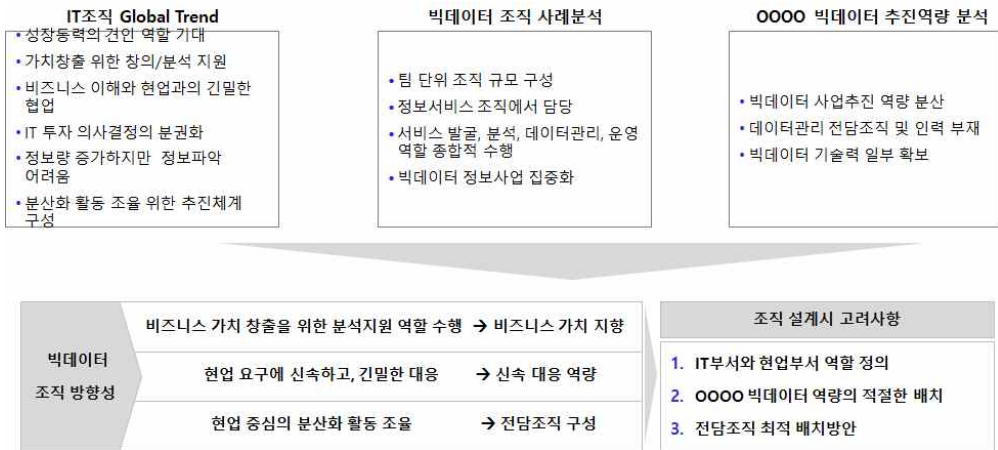


그림 4-6 빅데이터 조직 설계 방향

빅데이터 조직구성을 위해 세 가지 조직 대안의 특징 및 장단점을 비교하여 최적의 조직 구성안을 비교하면 표 4-24와 같다.

표 4-24 빅데이터 조직 대안 비교

항목	1안: 중앙집중형	2안: 현업중심형(분산형)	3안: 하이브리드형
설명	<ul style="list-style-type: none"> <li>빅데이터 과제기획, 분석, 데이터관리, 데이터수집, 제공을 위한 대외협력 등</li> <li>빅데이터 운영조직에서 모든 업무 수행</li> </ul>	<ul style="list-style-type: none"> <li>빅데이터 활용이 필요한 각 처(실)에서 자율적으로 사업을 수행</li> </ul>	<ul style="list-style-type: none"> <li>빅데이터 과제기획, 데이터관리, 데이터수집, 제공을 위한 대외협력 등 모든 업무를 빅데이터 운영조직에서 수행</li> </ul>

			<ul style="list-style-type: none"> <li>· 단, 분석업무의 경우 현업이 주도적으로 수행하고, 빅데이터 운영조직은 분석지원 업무를 수행</li> </ul>
특징	<ul style="list-style-type: none"> <li>· 빅데이터 기술, 자원, 인력을 빅데이터 운영 조직에 집중</li> </ul>	<ul style="list-style-type: none"> <li>· 빅데이터 기술, 자원, 인력이 각 부서 분산</li> </ul>	<ul style="list-style-type: none"> <li>· 효율적 수행을 위해 역량 집중 가능</li> <li>· 분석역량은 현업부서에 위치하고, 기술 및 관리역량은 IT부서에 위치함</li> </ul>
장점	<ul style="list-style-type: none"> <li>· 체계적인 빅데이터 과제 수행 및 자원관리 용이</li> <li>· 전사적 전략과의 일치가 쉬운 수준</li> <li>· 사업의 자율성과 신속성 확보 가능</li> </ul>	<ul style="list-style-type: none"> <li>· 현업의 필요에 따라 빅데이터 사업을 자율적으로 신속하게 진행</li> <li>· 데이터 및 인프라 자원의 효율적 활용 가능</li> <li>· 현업 맞춤형 빅데이터 사업 진행</li> </ul>	<ul style="list-style-type: none"> <li>· 부서의 전문성 유지와 효율적 재분배로 인적자원 관리용이</li> <li>· 데이터 및 인프라 자원의 효율적 활용 가능</li> <li>· 전사적 전략과의 일치가 보통 수준</li> <li>· 사업의 자율성과 신속성 확보 부분적 가능</li> </ul>
단점	<ul style="list-style-type: none"> <li>· 빅데이터 운영조직 규모가비대화</li> <li>· 비즈니스 전문성을 갖춘 현업인력 확보 어려움 (인사발령, 배치 문제 등)</li> <li>· 데이터 및 인프라 자원이 IT부서와 개별적 관리로 효율성 저하</li> </ul>	<ul style="list-style-type: none"> <li>· 전사 차원에서 자원과 인력 중복 투자에 의한 자원 낭비</li> <li>· 전사차원의 빅데이터 사업방향성 상실 우려</li> <li>· 전사적 전략과의 일치가 어려운 수준</li> </ul>	<ul style="list-style-type: none"> <li>· 빅데이터 운영조직과 현업간 사업추진 균형 유지관리 필요</li> <li>· 조직 내 거버넌스(직원평가 등)가 제대로 갖추어 지지 않은 경우 오히려 사업추진의 혼란을 야기</li> </ul>
유사 사례	기상청, 통계청, 통합전산센터		신한카드, 신한은행, 우리은행, KB국민은행

빅데이터 조직 및 인력 역량을 감안하고 데이터의 통합관리가 필요한 빅데이터 사업 특성상 (3안)하이브리드형 조직으로 구성하는 것이 효과적이다.

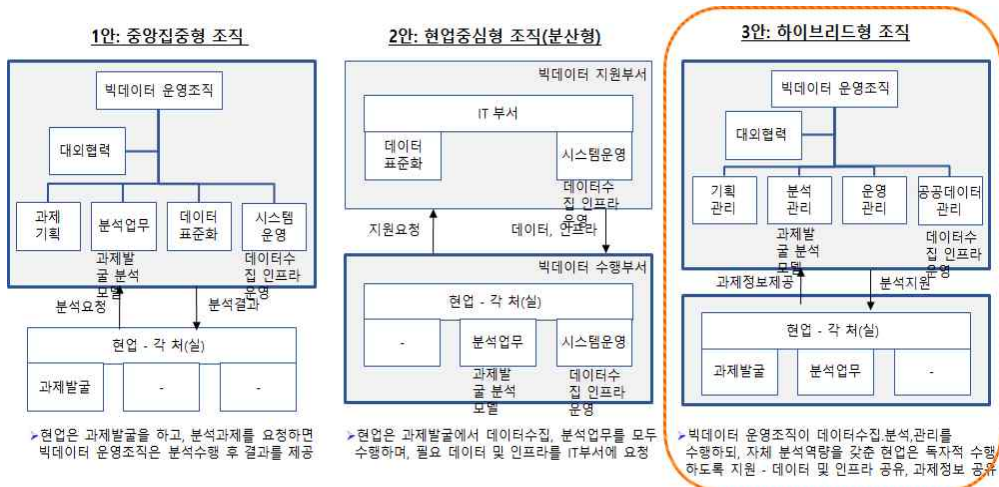


그림 4-7 빅데이터 조직 대안 설정

## 2) 타기관 운영 사례

타 기관의 빅데이터 전담조직 운영 현황을 살펴보면 교통관련 정보를 다루는 공공기관 중 빅데이터 전담팀을 운영하는 곳은 서울시와 경기도이다.

실제적인 업무가 원활히 이루어지는 것은 서울시이고, 경기도는 빅데이터 활용 기반 구축단계이다.

서울시에서는 빅데이터 전략팀의 주요 업무는 기관 내·외부 빅데이터를 수집하여 활용전략을 구축하고 이를 추진하는 것으로 수행업무를 크게 4가지로 구분할 수 있다.

- 내외부 빅데이터 수집
- 빅데이터 현황조사를 통한 사업발굴
- 빅데이터 공유, 활용 플랫폼 구축 및 고도화
- 운영관리

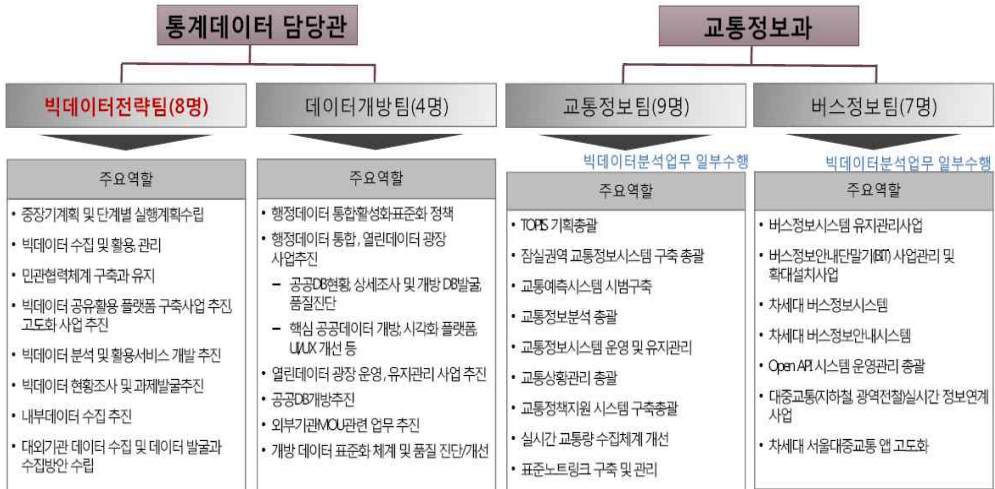


그림 4-8 서울시 빅데이터 업무 운영 조직

경기도의 빅데이터 담당관은 주로 빅데이터를 활용한 사업을 개발하고 추진하기 위한 기반을 다지는 단계로 그와 관련된 교육과 사업 개발 등이 주 업무이다.

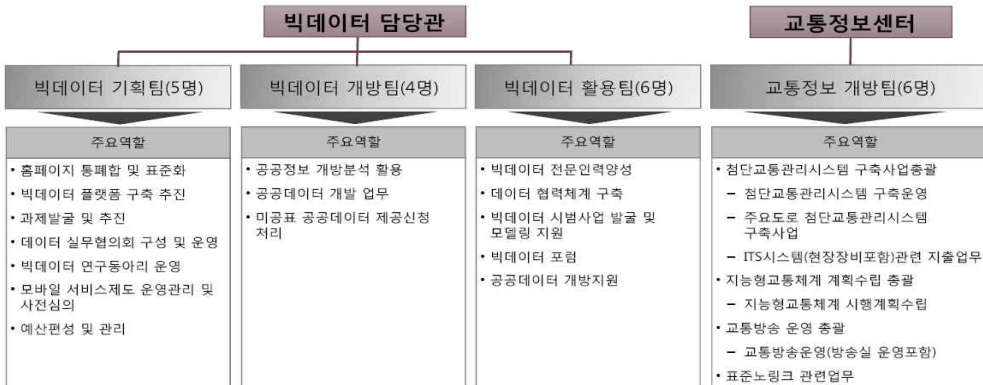


그림 4-9 경기도 빅데이터 업무 운영 조직

한국도로공사에서는 기획본부의 정보처에서 빅데이터 시스템을 구축하는 인원은 1인이며, ICT 센터의 공공정보팀에서 빅데이터를 활용한 업무개발과 몇 가지 빅데이터 시범과제를 추진 중에 있다.

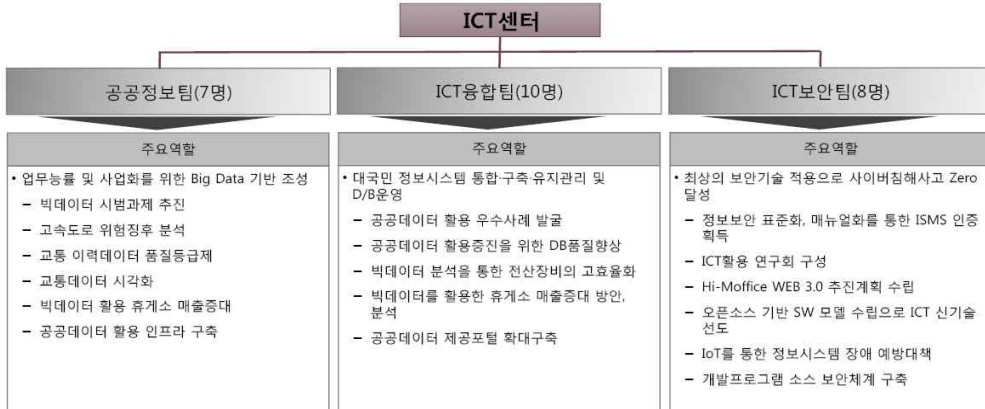


그림 4-10 한국도로공사 빅데이터 업무 운영 조직

### 3) 중앙전파관리소의 빅데이터 운영조직 업무 도출

빅데이터 운영조직에 대한 사례 분석결과, 신규과제 발굴, 공유·활용 플랫폼 구축사업, 플랫폼 운영 및 유지관리의 업무가 주로 수행되고 있고, 빅데이터 분석 조직의 주요업무는 내부 및 외부데이터 조사, 빅데이터 분석시스템 구축 및 운영관리, 빅데이터 분석 및 활용서비스 개발, 데이터 개방 및 활성화로 구성된다.

중앙전파관리소에는 내부 조직사이에서 경계가 모호한 업무나 긴밀한 협력이 요구되는 업무를 위한 내부 의사결정 조직이 필요하다.

- ① 개별조직의 데이터 간 긴밀한 연계 체계가 요구
- ② 빅데이터 활용시스템 구축 및 고도화를 위한 개별 시스템의 데이터 품질 향상 요구
- ③ 빅데이터 분석 및 활용시스템 운영에 따른 데이터 개방 범위(타 조직의 데이터가 포함된)와 이에 따른 소요인력 투입에 대한 업무 등

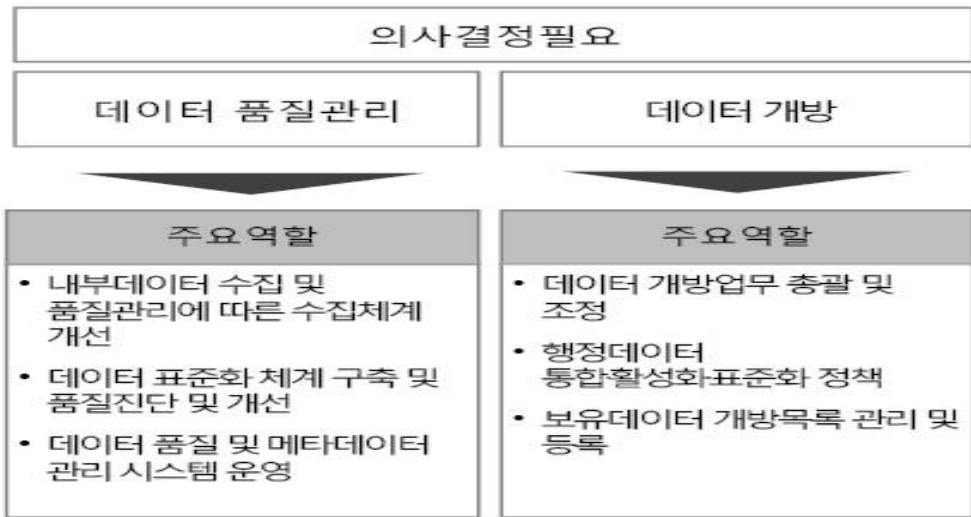


그림 4-11 의사 결정 필요 주요 역할

IT Governance 기준으로 도출된 수행 영역별 IT직무를 관련 분야별 분류 및 빅데이터 시스템 조직형태를 반영한 결과 정보화 운영 조직은 4개 부문으로 구성할 것을 권고한다.

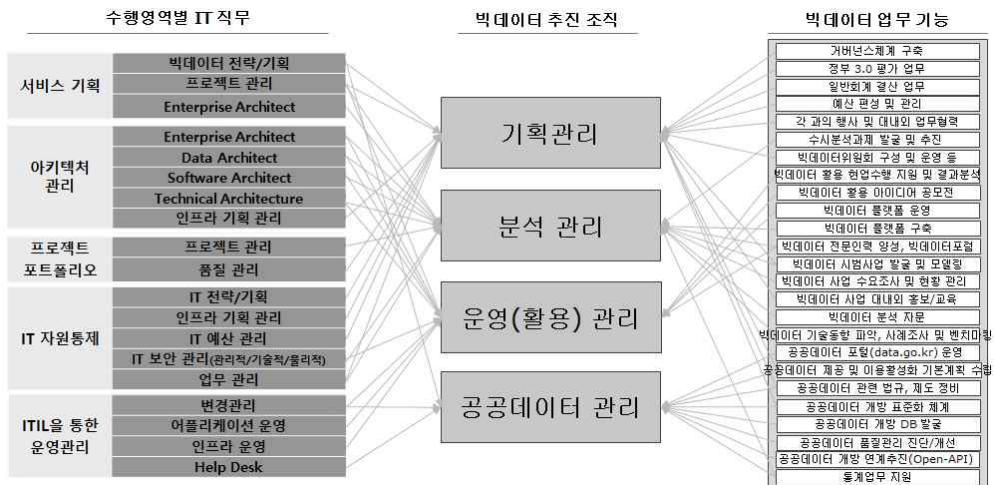


그림 4-12 빅데이터 추진 조직 부문 도출



빅데이터 추진 조직은 기획관리, 분석관리, 이용(활용)관리 및 공공데이터 관리의 4개 기능을 수행하는 新 빅데이터 사업팀으로 구성방안을 수립한다.

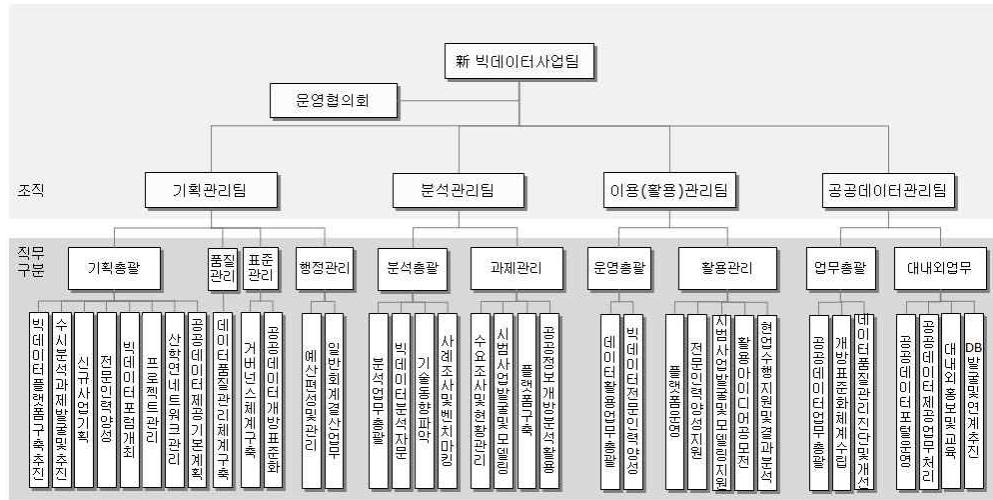


그림 4-13 빅데이터 운영 조직 직무구분

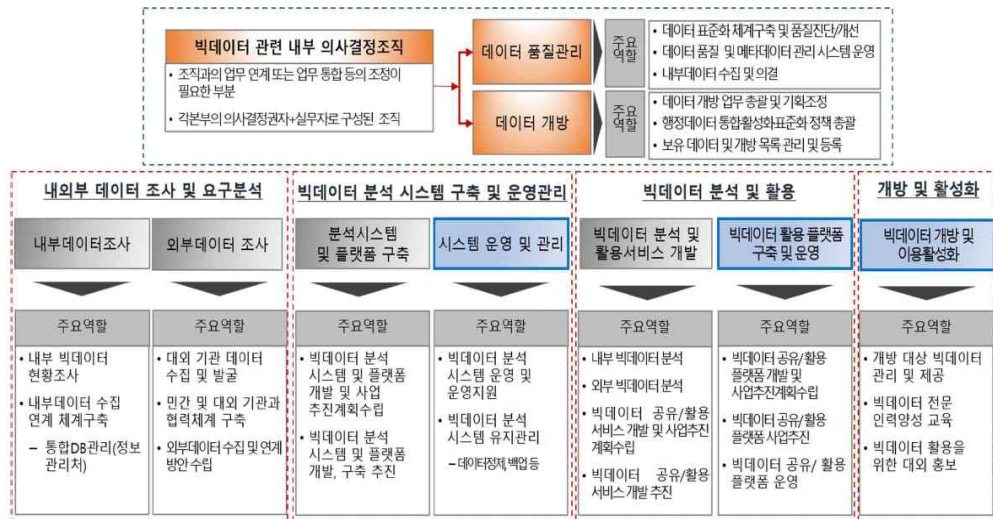


그림 4-14 빅데이터 운영 조직 업무

■ : 1차 조직, ■ : 2차 조직(시스템 구축 이후), ■ : 조직간 업무 조정 및 협의 필요



4) 중앙전파관리소의 빅데이터 운영조직 소요인력 도출

조직 구성 방안에서 소요인력 도출에 있어서, 소요인력 산정방법으로는 Benchmarking과 인력Needs 분석방법을 토대로 업무 기능에 따른 적정 인력 추정치를 도출한다.

[illegible]

그림 4-15 소요인력 산정방법

데이터수집 1명, 빅데이터 분석시스템 및 플랫폼 운영관리 2명  
빅데이터 분석 및 활용 3명, 데이터 개방 및 이용 활성화에 1명의  
인력이 필요할 것으로 판단되나, 향후 빅데이터 분석 및 활용시스  
템 규모에 따라 소요인력은 변동될 수 있다.

분야	세부업무	기 능	FTE	적정인력
데이터 수집	내부데이터	내부 빅데이터 현황조사	0.3	1
		내부 부처간 데이터 연계체계 구축	0.1	
	외부데이터	외부데이터 수집 및 연계 방안 수립	0.1	
		대외기관 데이터 수집 및 발굴	0.1	
		민간 및 대외기관과 협력체계 구축	0.2	
빅데이터 분석시스템 및 플랫폼 운영 관리	분석시스템 및 플랫폼 구축	빅데이터 분석 시스템 개발 및 사업 추진 계획수립	0.5	2
		빅데이터 분석 시스템 개발 추진	0.2	
	시스템 운영 및 관리	빅데이터 분석시스템 운영 및 운영지원	1	
		빅데이터 분석시스템 유지보수	0.2	
소계			2.7	3

분야	세부업무	기 능	FTE	적정인력
빅데이터 분석 및 활용	빅데이터 분석	내부 빅데이터 분석	0.5	3
		외부 빅데이터 분석	0.3	
	빅데이터 활용 서비스개발	빅데이터 공유, 활용 서비스 개발 및 사업 추진 계획수립	0.5	
		빅데이터 공유, 활용 서비스 개발 추진	0.5	
	빅데이터 공유, 활용 플랫폼구축 및 운영	빅데이터 공유, 활용 플랫폼 개발 및 사업 추진 계획 수립	0.5	
		빅데이터 공유, 활용 플랫폼 개발 추진	0.5	
		빅데이터 공유, 활용 플랫폼 운영	0.3	
빅데이터 개방 및 이용활성화	데이터 개방	공공데이터 개방	0.1	1
	빅데이터 이용 활성화	전문인력 양성을 위한 교육	0.2	
		빅데이터 활용을 위한 대외 홍보	0.2	
소계			3.6	4

<sup>1)</sup> FTE (Full Time Equivalent) : 1인이 일 년 동안 full time으로 일하는 경우와 동일한 측정치

<sup>2)</sup> 적정인력=ROUND(직무별FTE합계)

그림 4-16 소요 적정 인력 산정

## 5) 중앙전과관리소의 빅데이터 운영조직 구성 안

빅데이터 분석 조직의 주요업무는 내부 및 외부데이터 조사, 빅데이터 분석시스템 구축 및 운영관리, 빅데이터 분석 및 활용서비스 개발, 데이터 개방 및 활성화로 구성하고, 내부데이터의 수집 및 요구분석에 따른 데이터 품질관리와 개방 관련 의사결정을 수행하기 위한 별도의 조직구성이 필요하다.



### 운영인력 필요역량과 교육내용



빅데이터 역량	역량정의	교육내용
데이터 수집	빅데이터 플랫폼에 대한 지식과 빅데이터 수집유형, 수집방법, 수집기술을 이해하고, 사용할 수 있는 역량	데이터 플랫폼 이론, 빅데이터 수집
데이터 저장	빅데이터 플랫폼에 대한 지식과 빅데이터 저장유형, 저장방법, 저장기술을 이해하고, 사용할 수 있는 역량	데이터 플랫폼 이론, 빅데이터 저장 및 관리, 빅데이터 처리 및 분석, 빅데이터 플랫폼 구축 및 활용
데이터 보안	빅데이터 보안과 개인정보보호 등에 대한 마인드를 갖고 있으며, 데이터보안기술 지식을 기반으로 데이터보안을 실현할 수 있는 역량	빅데이터 보안분석, 빅데이터 환경의 개인정보보호 대책
시스템 운영	빅데이터 플랫폼을 이해하고, 백업 및 복구기술 등을 적용하여 시스템의 안정적 운영을 유지할 수 있는 역량	빅데이터 플랫폼과 하둡 이해, 하둡에코시스템 설치와 활용, 빅데이터 아키텍처 설계방법론
데이터 용량 관리	빅데이터의 용량을 예측하고, 용량산정결과를 기반으로 데이터저장장치 및 저장방법을 선택하고 적용할 수 있는 역량	-

### 운영인력 양성방안

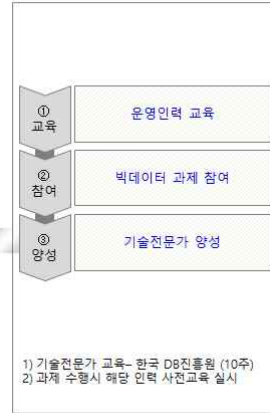


그림 4-19 빅데이터 운영 인력 양성 방안

## 제 5장 전파감시 빅데이터 시스템 구축 방안

### 제1절 빅데이터 플랫폼 구축 방안

중앙전파관리소의 빅데이터 플랫폼은 전파감시 빅데이터와 외부 연관 데이터를 모두 수집·저장하여 그 연관성을 분석할 수 있는 플랫폼으로 사업의 안정적인 추진을 위해 ①빅데이터 분석 시범사업, ②플랫폼 구축, ③고도화 순으로 단계적인 사업 추진이 필요하다.

빅데이터 분석 시범사업은 전파감시 빅데이터 분석모델 개발을 위한 분석 플랫폼 구성 요소중 최소한의 하드웨어 및 소프트웨어 등 빅데이터 분석에 필요한 기반을 구축하고, 전파감시 빅데이터 시범 과제 수행을 의미하며, 빅데이터 플랫폼 구축 시에는 빅데이터 수집, 데이터 저장 및 처리, 데이터 분석, 데이터 활용, 데이터 보안 및 관리 영역으로 구성하여야 하며, 단계별 사업기간은 약 5~6개월 정도가 소요될 것으로 예상된다.

#### 1. 빅데이터 분석 시범사업

빅데이터 플랫폼은 기존 정보시스템 구축 방법, 대상 및 활용 방법에 차이가 있다. 더욱이 빅데이터 플랫폼을 통해 얻으려고 하는 것은 기존 정보시스템에서 제공하지 않거나 하지 못하는 부분을 모형화하여 플랫폼에 탑재하는 과정을 거친다. 빅데이터를 수집, 분석 및 의미 있는 활용을 위해서는 빅데이터 플랫폼에서 얻어지는 결과가 업무에 도움이 될 수 있어야 한다. 그러기 위해서는 다양한 업무적인 요구사항을 수집, 분석과제로 정의하고 장기적으로 활용할 수 있는 플랫폼을 구축하기 위해서 시범사업을 통해 필요성을 검증하고 장기적인 활용 전략을 수립하는 과정을 거쳐야 한다.

빅데이터 분석 플랫폼의 구축은 어떤 제품을 사용하느냐에 따라 구성이 달라 질 수 있다. 예를 들어 하둡 플랫폼과 엘라스틱서치[

Elasticsearch]는 제품 구성이 완전히 다르므로, 시범사업을 위한 시스템 구성을 위해서는 우선적으로 수요기관의 의사결정이 필요하며, 시범사업을 통해서 빅데이터 분석의 효과를 검증하고 실증된 시스템의 효과를 바탕으로 장기적인 빅데이터 분석 플랫폼을 구축하여야 한다.

#### <시범사업 시 추진사항>

- 시범과제 정의 : 연구과제를 통해 도출된 과제 중 시범과제를 선정하고 빅데이터 분석을 위한 표준화 및 상세 설계
- 데이터 수집 : 내부 데이터 수집, 외부 데이터 수집
- 데이터 저장/처리 : 수집된 데이터 저장, 저장 데이터 분석을 위한 모델 개발 및 데이터 분석
- 데이터 보안 및 관리 : 보안 클러스터 설치
- 데이터 분석 : 개발 모델을 통한 데이터 분석
- 시각화 : 분석 결과 시각화, 다차원분석, 예측분석, 의사결정 분석, 텍스트분석, 디스커버리분석)

## 2. 빅데이터 플랫폼 구축

중앙전파관리소 전파감시 데이터의 다양성, 향후 전파관리정보시스템의 고도화 사업 등을 고려하여 H/W는 가용성 및 확장성(CPU, Memory, Disk)을 제공하는 최신의 시스템인 DL380 Generation 10 제품, 스토리지로 설계하고, SW는 안정성을 바탕으로 웹 브라우저만으로 다차원 통계 분석을 할 수 있는 인메모리 기반의 통계분석 및 시각화 솔루션, 다양한 전자문서의 DATA와 비정형 DATA의 수집이 가능한 문서추출 솔루션 등 전 세계적으로 가장 레퍼런스가 많은 MapR 플랫폼을 활용한 구축안을 그림 5-1과 같이 제안한다.

전파감시 빅데이터 플랫폼의 구축을 위해서는 표 5-1, 표 5-2과 같이 기본적인 H/W와 S/W가 필요하다.

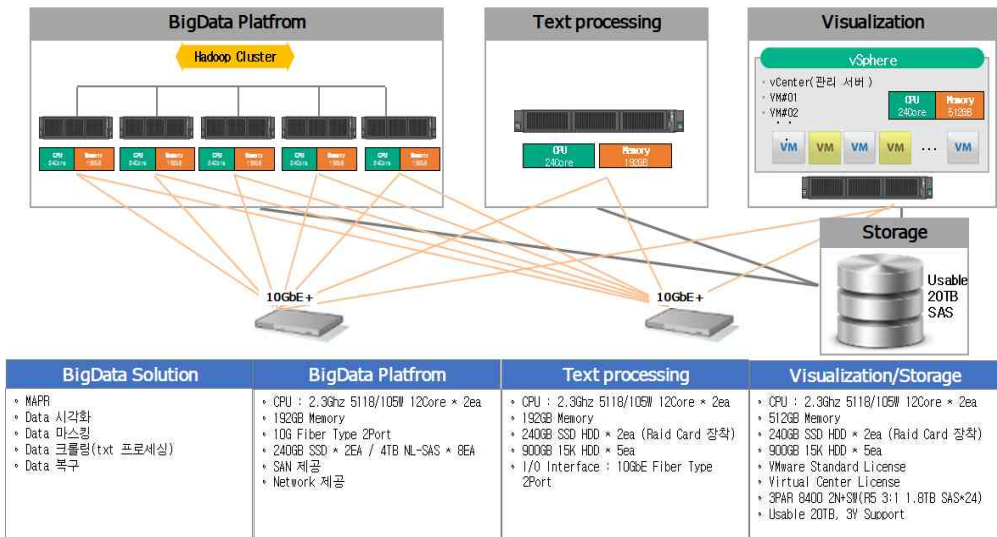


그림 5-1 빅데이터 시스템 구축안(HW+SW)

빅데이터 플랫폼은 오픈소스를 기반으로 x86서버 H/W를 활용하여 구축하는 것이 일반적이지만, 이는 빅데이터 구축 방법에 따라 달라질 수 있다.

### 1) 하드웨어 및 기능

- 수집서버 : 주기적인 데이터 수집, 정제, 변환 및 로딩의 역할 수행
- 네임노드(Name Node) : 빅데이터를 저장하기 위한 HDFS 파일 시스템을 유지 관리하기 위한 메타데이터를 관리하고, 데이터를 저장하는 데이터 노드를 모니터링 및 클라이언트에서 데이터 파일에 접근하려는 요청을 접수 및 관리
- 데이터노드(Data Node) : HDFS 형태로 데이터를 저장하는 역할 수행
- 분석 서버 : 하둡분산파일시스템(HDFS)에 저장된 데이터를 분석
- 시각화 서버 : 빅데이터 분석 결과를 다양한 통계분석 및 시각화 기법을 통해 사용자 분석 환경 제공

표 5-1 빅데이터 플랫폼 구축 기본적인 H/W 사양

구분	CPU	메모리	디스크	OS	비고
수집서버	8Core	256G	Usable 10TB	Linux	이중화 추천
네임노드	8Core	128G	Usable 5TB/ 노드	Linux	이중화
데이터노드	12Core	256G	Usable 15TB/노드	Linux	3개노드(기본)
분석서버	12Core	256G	Usable 10TB	Linux	
시각화 서버	8Core	256G	Usable 5TB	Linux/Windows	

## 2) 소프트웨어

빅데이터 분석 플랫폼을 구축하기 위해서는 다양한 S/W를 활용해야 하고, 일반적으로 오픈소스가 활성화 되어 있는 플랫폼 관련 S/W인 하둡(Hadoop)을 활용하여 구성한다. 하둡은 빅데이터 저장 및 관리 환경에 가장 널리 활용되고 있는 오픈소스로 장기적인 관점에서 접근할 수 있는 S/W이다. 반면에 데이터 수집을 위한 S/W 및 시각화 S/W는 오픈소스 보다는 보다 안정적이고 관리가 용이한 상용 S/W를 추천한다.

표 5-2 빅데이터 플랫폼 구축 S/W 기본사양

구분	오픈소스 구분	OS	비고
수집 S/W	상용	Linux	
빅데이터 플랫폼	오픈소스	Linux	
시각화 S/W	상용	Windows/Linux	
암호화 S/W	상용	Linux	
DBMS	오픈소스	Linux/Windows	

## 3) 빅데이터 플랫폼 구축을 위한 방법론

시범사업을 거쳐 구축되는 빅데이터 플랫폼은 다양한 데이터를



수집하고, 분석하는 과정을 거쳐 활용할 수 있는 장기적인 시스템 구조로 구축되어야 한다. 시범사업에서 도출된 이슈 등을 개선하고 보다 안정적으로 활용할 수 있는 시스템을 구축하여야 한다.

(1) 시범사업 이슈 및 개선 방안 도출

: 시범사업을 통해서 구축한 과제에 대한 평가와 활용 및 개선 방안을 도출하고 빅데이터 플랫폼 구축에 적용할 수 있도록 함.

(2) 빅데이터 플랫폼 아키텍처 정의

: 빅데이터 수집에서 저장, 분석 및 시각화를 위한 상세 아키텍처를 수립

(3) 빅데이터 분석 과제 상세화

: 빅데이터 플랫폼을 통해서 서비스할 과제를 상세화하고, 대상 데이터의 현황 파악 및 대상 데이터 선정을 통해 서비스 과제 상세화

(4) 빅데이터 플랫폼 구축

① 데이터 수집 프로세스 설계 및 구축

② 빅데이터 플랫폼 설치 및 테스트

③ 빅데이터 분석 모델 설계 및 개발

④ 빅데이터 적재 및 분석

⑤ 시각화 설계 및 구축

(5) 빅데이터 활용 방안 수립

: 빅데이터 플랫폼을 통해 구축된 서비스 과제의 실 업무 활용 방안을 수립하고 플랫폼의 확장 방안 등을 수립

(6) 빅데이터 교육 방안 수립

: 플랫폼의 활용을 극대화 하고, 빅데이터 분석 및 업무 활용을 위해 사용자들이 적극적으로 참여할 수 있도록 빅데이터 플랫폼 활용 교육을 주기적으로 수행

### 3. 연차별 사업 추진 방안

중앙전파관리소의 빅데이터 분석 플랫폼 구축은 빅데이터 플랫폼 구축 연구과제 수행(정보화 전략계획 수립), 전파 혼·간섭 예방 빅데이터 분석 시범사업 및 빅데이터 분석 플랫폼 구축, 빅데이터 분석 플랫폼 고도화에 따른 연차별 사업 추진 전략을 수립하고, 각 단계별 목표는 아래와 같다.



그림 5-2 연차별 사업 추진 계획

- ◆ 1차 년도 : 연구과제 결과를 바탕으로 시범과제 검증(시범사업)
- ◆ 2차 년도 : 시범사업을 통해 검증된 빅데이터 분석 플랫폼을 구축하고, 빅데이터 활용방안 수립
- ◆ 3차 년도 : 빅데이터 플랫폼 서비스 고도화 및 분석업무 확대

#### 1) 1차 년도 상세

- 수행 과제 : 전파 혼·간섭 예방을 위한 빅데이터 시범 분석 환경 구현
- 구현 목표 : 연구과제 결과 도출된 빅데이터 과제 시범 구현

- 소요 기간 : 5 개월 ~ 6개월
- H/W 도입 장비 : 수집서버, 분석 및 시각화 서버
- S/W 도입 장비 : 수집도구(ETL), 시각화 도구, DBMS(오픈소스)

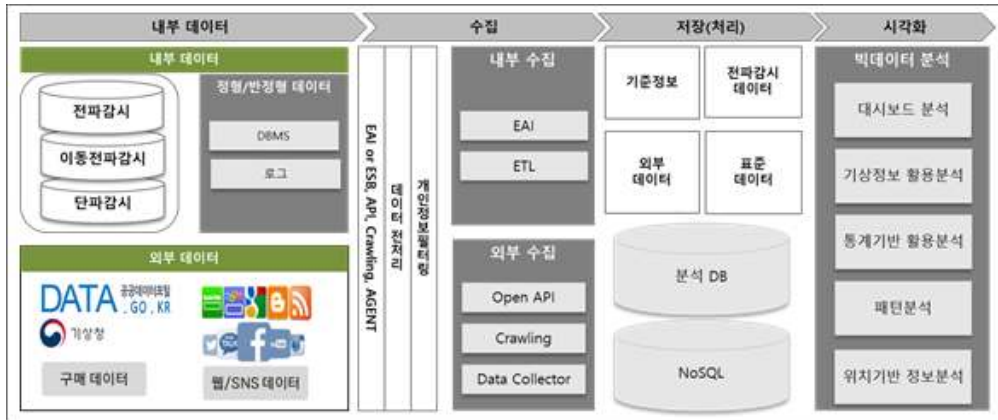


그림 5-3 1차 년도 예상 목표 구성도

1차 년도 목표 구성도는 상세 시범사업 계획에 따라 변경될 수는 있으나, 최소한 내부데이터를 활용한 빅데이터 분석 가치에 대한 검증할 수 있는 환경을 구성해야 하며, 2차 년도 사업을 위한 충분한 기반으로 활용되어야 한다.

## 2) 2차 년도 상세

- 수집 대상 : 내부데이터, 외부데이터(기상정보 등, 외부데이터의 경우 필요성에 따라 활용 방안 검토)
- 구현 목표 : 시범구현으로 검증된 빅데이터 분석 플랫폼 구축
- 소요 기간 : 6개월
- H/W 도입 장비 : 네임노드(2대), 데이터노드(3대), 분석서버
- S/W 도입 장비 : 시각화 라이선스 확대, 암호화 도구, 하둡 및 오픈소스

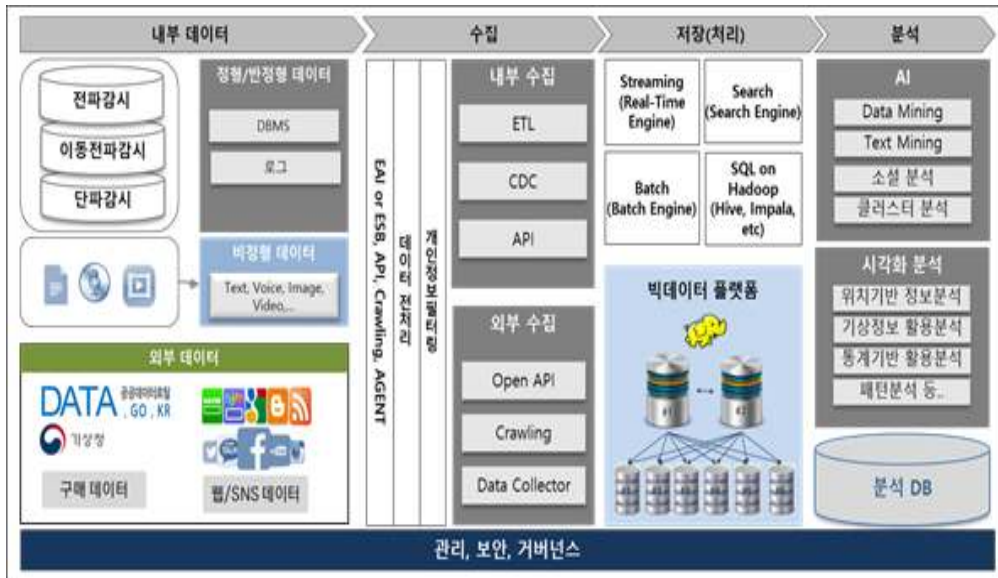


그림 5-4 2차 년도 예상 목표 구성도

2차 년도 사업은 장기적인 관점에서 안정적이고 확장성 있는 빅데이터를 분석하고 서비스할 수 있는 체계로 구축해야 한다. 전파감시 데이터 및 외부 연계 데이터의 확장에 따라 데이터 노드가 확장이 되어야 하며, 다양한 오픈소스를 적용할 수 있어야 한다.

### 3) 3차 년도 상세

- 수집 대상 : 확대 과제와 관련된 내부 및 외부 데이터
- 구현 목표 : 서비스 과제 고도화 및 신규 과제 확대
- 소요 기간 : 3개월 ~ 4개월
- H/W 도입 장비 : 2차년도 구축 플랫폼 활용
- S/W 도입 장비 : 기존 제품 활용 및 시각화 라이선스 확대

## 제2절 빅데이터 플랫폼 설계

### 1. 빅데이터 플랫폼 설계 목적 및 범위

중앙전파관리소의 빅데이터를 공통으로 활용할 수 있는 빅데이터 플랫폼을 설계하고, 플랫폼에 대한 세부 구축방안을 수립하기 위해서는 활용 가능한 최적의 빅데이터 플랫폼 아키텍처를 설계하고 위협정보 데이터의 수집 및 저장, 분석, 활용의 최적화 방안을 도출해야 한다.

플랫폼에 대한 세부 구축방안은 수집, 저장, 분석, 활용의 4단계 구축방안으로 구성된다.

<b>수집 플랫폼</b>	<ul style="list-style-type: none"> <li>• 단계별 구축전략에 따른 빅데이터 플랫폼 설계</li> <li>• 내/외부데이터 수집(ETL, Crawling, API등) 설계</li> <li>• 데이터 수집분야 아키텍처 설계</li> </ul>
<b>저장 플랫폼</b>	<ul style="list-style-type: none"> <li>• 데이터 적재·보관·처리·보안·관리 분야 아키텍처 설계</li> <li>• 데이터 백업구조 설계</li> </ul>
<b>분석 플랫폼</b>	<ul style="list-style-type: none"> <li>• 위협정보 분석을 위한 표준 프레임워크 정의 및 구축방안 수립</li> <li>• 사업 추진과제별 분석모형(서술, 진단, 예측, 처방, 최적화) 설계</li> </ul>
<b>활용 플랫폼</b>	<ul style="list-style-type: none"> <li>• 데이터 시각화 Tool 정의</li> <li>• 텍스트, 이미지, 동영상 등 다양한 데이터를 분석 제공하는 활용 플랫폼 구축</li> </ul>

그림 5-6 플랫폼 4단계 구축방안

### 2. 빅데이터 플랫폼 설계 절차

중앙전파관리소의 빅데이터 플랫폼 설계 절차는 빅데이터 플랫폼 설계, 빅데이터 인프라 구성, 빅데이터 거버넌스 체계 수립, 빅데이터 분석센터 기획, 빅데이터 분석 프로세스 정립으로 구분 할 수 있다.

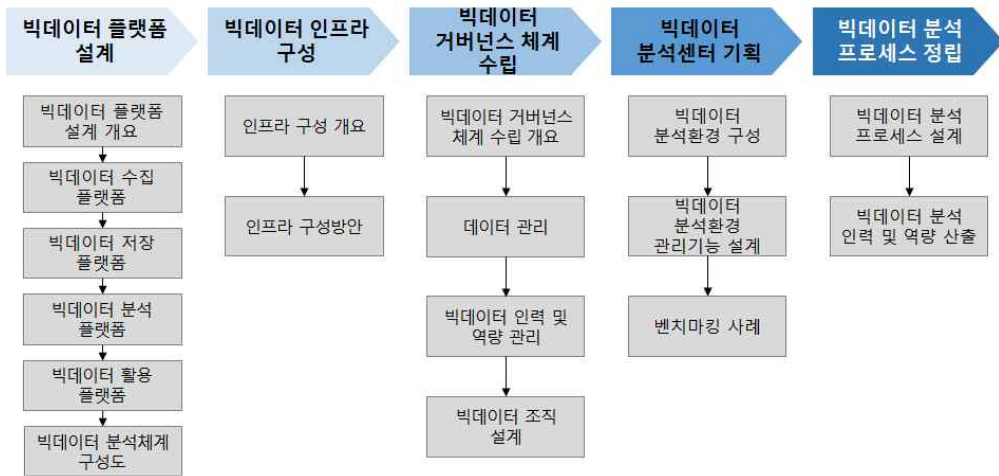


그림 5-5 빅데이터 플랫폼 설계 절차

### 3. 빅데이터 플랫폼 설계 접근 방안

빅데이터 플랫폼 설계를 위해 빅데이터 레퍼런스와 관련기술을 분석하고 기관의 인프라 환경 분석 및 구축 전략, 빅데이터 과제 발굴 및 선정과정에서 도출된 내용을 바탕으로 빅데이터 플랫폼을 설계한다.

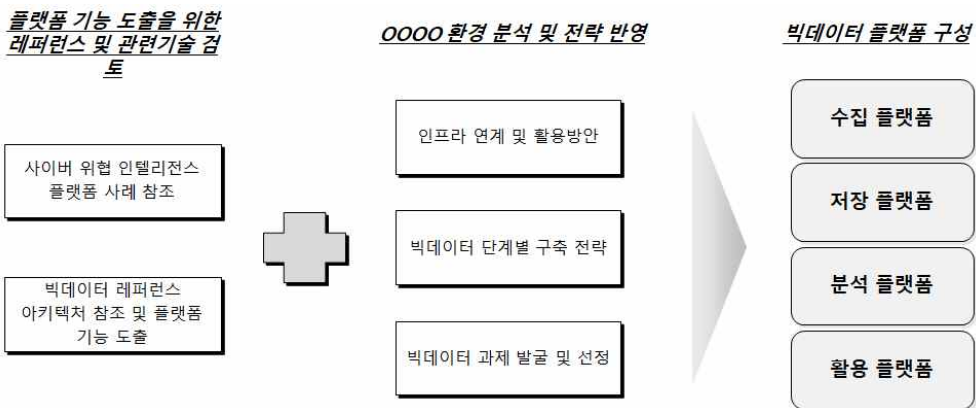


그림 5-7 빅데이터 플랫폼 설계 접근 방안

#### 4. 빅데이터 레퍼런스 아키텍처

TeraONE은 고객이 효율적으로 데이터를 관리할 수 있는 DW 환경을 최적화한 아키텍처를 기반으로 진화해 새로운 엔터프라이즈 플랫폼을 제시한다.

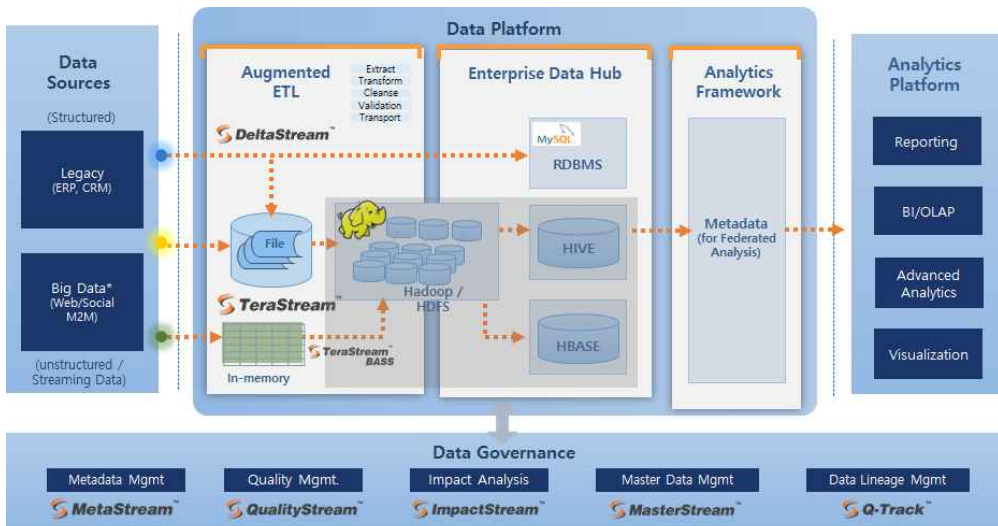


그림 5-8 TeraONE 엔터프라이즈 플랫폼

TeraONE은 데이터 통합 및 거버넌스 적용 솔루션에 오픈소스의 확장성과 유연성을 보강한 Hybrid 플랫폼으로 다음과 같은 기능을 제공한다.

표 5-3 TeraONE 확장성과 유연성을 보강한 기능

영역	기능
Augmented ETL	<ul style="list-style-type: none"> <li>다양한 데이터 유형 지원 : 표준 연계 모듈을 통하여 프로그램 개발의 편의성 및 재사용성 제공</li> <li>데이터 수집 현황을 GUI 환경으로 제공</li> </ul>
Enterprise Data Hub	<ul style="list-style-type: none"> <li>데이터 처리/저장/보안/품질관리 등 데이터를 안전하게 활용하기 위한 일련의 프로세스를 지원</li> <li>다양한 데이터 형식 변환, 다양한 형식 변환 기능 제공</li> <li>데이터 저장을 위해 RDBMS, NoSQL, 분산파일시스템 등 다양한 방식 제공</li> </ul>

Analytics Framework / Data Insight	<ul style="list-style-type: none"> <li>데이터 사이언티스트에게 데이터에 대한 제약을 최소화한 유연한 분석 환경을 제공하기 위한 Data Federation기능 제공</li> <li>다양한 분석 및 시각화 기능 기본 지원(R, HIVE 등 오픈소스 지원)</li> <li>필요시 타 회사와의 파트너십을 통해 기능 제공(용이한 연동 환경 제공)</li> </ul>
Data Governance	<ul style="list-style-type: none"> <li>빅데이터 환경 운영을 위한 정책/규정 및 각종 기준, Role, Data Source 속성 등을 통합적으로 관리</li> <li>개인정보 관리 기능을 통한 운영 기능, 비즈니스 메타데이터와 테크니컬 메타데이터 간의 링크로 현업과 IT 부서간 융합 지원 기능</li> </ul>
Admin	<ul style="list-style-type: none"> <li>대시보드 형태의 모니터링을 통하여 name node 및 data node 시스템 정보 제공</li> <li>Map/Reduce 작업 개수 및 프로그램 성공/실패 정보 제공</li> <li>Ecosystem 진행 상황 모니터링, 오퍼레이션 리포트, 로그 이벤트 추적 등</li> <li>플랫폼 전반에 걸쳐 워크플로우 기능을 통한 사용자 편의 기능</li> </ul>

레퍼런스 아키텍처를 바탕으로 빅데이터 아키텍처에 필요한 Use Cases를 통해 중앙전과관리소 빅데이터 플랫폼의 방향성을 도출하였다.

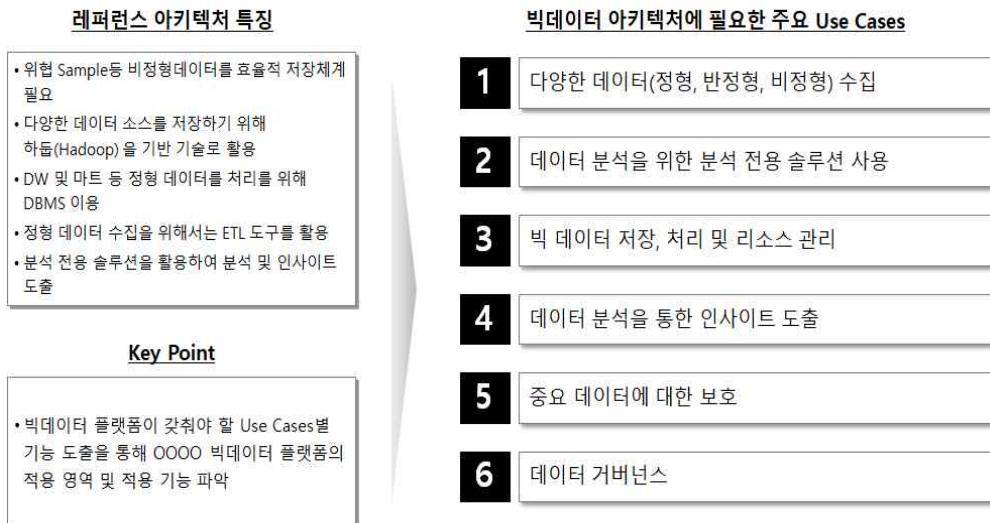


그림 5-9 기관 빅데이터 플랫폼의 방향성



## 5. 빅데이터 플랫폼 구성요소 정의

빅데이터 레퍼런스 아키텍처와 적용 가능한 기능 도출을 통해 빅데이터 플랫폼의 기능영역을 크게 6가지 영역으로 설정하고 각 기능영역에 필요한 구성요소(컴포넌트)를 정의한다.



그림 5-10 빅데이터 플랫폼 구성요소 정의

### 제3절 빅데이터 수집 플랫폼 구성

#### 1. 빅데이터 수집 플랫폼 개요

빅데이터 수집 플랫폼은 데이터 연계 및 수집, 처리를 위한 인프라 구성을 통해 중앙전파관리소에 다양한 분석을 수행할 수 있는 기반을 제공할 수 있다.

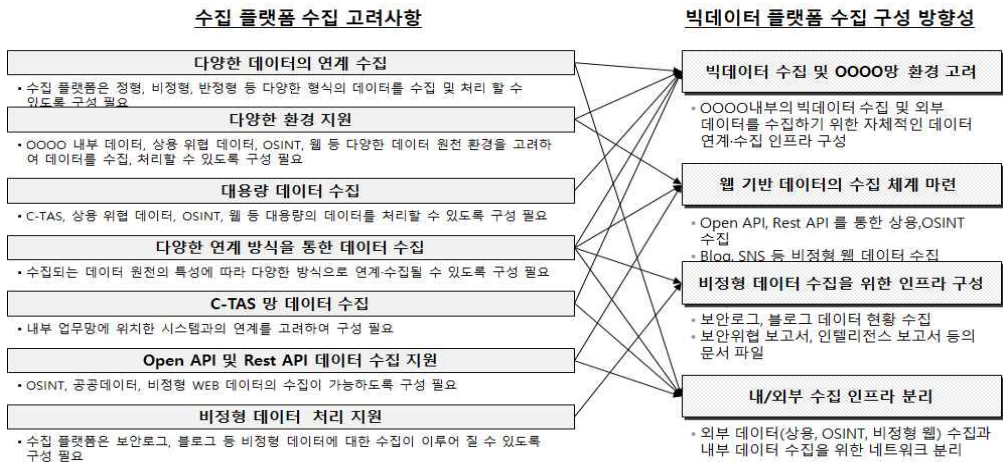


그림 5-11 빅데이터 수집 플랫폼 구성 고려사항

빅데이터 플랫폼에서 내부 및 외부의 원천 데이터를 수집하기 위한 아키텍처는 그림 5-12와 같다.



그림 5-12 빅데이터 수집 플랫폼 컴포넌트

## 2. 빅데이터 수집 플랫폼 구성 및 기능

내부의 데이터 수집을 위한 부분으로는 시스템에 직접 연결하는 방식이 아닌 경우, 수집서버와 중계서버를 이용한 내부 시스템으로부터의 데이터 수집, 기관 내부에 존재하는 DBMS로부터 데이터를 수집하기 위한 DBMS 수집, API 수집, 로그데이터 수집, Open API 수집, 웹 수집(Crawler), API 수집 컴포넌트 구성요소이다.

표 5-4 빅데이터 수집 플랫폼 구성 및 기능

영역	구성요소	기능 설명	고려사항
Agent 연동 수집	라우팅 및 연동 컴포넌트	<ul style="list-style-type: none"> <li>서비스 어플리케이션들의 연결을 통한 서비스 제공을 위한 라우팅 기능</li> <li>서비스 라우팅(Static Routing, Dynamic Routing, Complex Routing) 기능</li> <li>수집 대상 레거시 시스템과의 연동을 위한 Pluggable 어댑터 기능</li> </ul>	
	메세지 변환	<ul style="list-style-type: none"> <li>원천 시스템으로부터 전달되는 데이터에 대한 변환 기능 (XML to C-Text, JSON to C-Text, JSON to CSV, XML to C-Text등)</li> <li>데이터 포맷과 형태 등을 자유롭게 변환하고 가공하는 기능</li> </ul>	
	프로토콜 변환	<ul style="list-style-type: none"> <li>이기종 분산 환경과 다양한 프로토콜 간에 메시지를 주고 받도록 통신 프로토콜에 대한 변환 기능 (JSON/HTTP to JMS 등)</li> </ul>	
	모니터링	<ul style="list-style-type: none"> <li>통합적인 모니터링이 가능해야 하며 진행되고 있는 프로세스의 상태 및 데이터에 대한 조회가 기능</li> <li>장애 발생 시 장애 원인이 문제 파악 가능한 관리 기능</li> </ul>	
	관련 솔루션	<ul style="list-style-type: none"> <li>오픈소스 : Apache Kafka</li> <li>상용 : TibcoActiveMatrixBusinessWork</li> </ul>	
DBMS 수집	수집정보 관리	<ul style="list-style-type: none"> <li>수집 대상 정보(DBMS, ODBC, JDBC 등)와 연계 방식 설정 기능</li> <li>수집 정보 및 수집정보 전달 대상에 대한관리(추</li> </ul>	C-TAS 데이터 수집

		<ul style="list-style-type: none"> <li>가/변경/삭제) 기능</li> <li>수집 주기(시간, 분, 초)를 설정하는 기능</li> </ul>	
	수집 에이전트	<ul style="list-style-type: none"> <li>에이전트 관리 및 운영(시작, 중지 등) 기능</li> <li>DBMS 메타정보를 통해 수집대상 테이블 및 컬럼을 선택하는 기능</li> <li>수집대상 데이터의 범위를 설정하는 기능</li> <li>원천 시스템의 부하를 최소화하면서 데이터를 수집하는 기능</li> </ul>	
	통계 및 모니터링	<ul style="list-style-type: none"> <li>ETL 작업에 대한 자세한 통계 정보 제공 기능</li> <li>ETL Job 실행 오류, 이상동작에 대한 모니터링 및 복구기능</li> </ul>	
	관련 솔루션	<ul style="list-style-type: none"> <li>오픈소스 : Apache Sqoop</li> <li>상용 : DataStreams TeraStream/DeltaStream, IBM InfosphereDataStage, InformaticaBigData Management, Oracle ODI</li> </ul>	
API 수집	수집정보 관리	<ul style="list-style-type: none"> <li>수집 대상 정보(URI)에 대한 연계방식 설정 기능</li> <li>수집 정보 및 수집정보 전달 대상에 대한관리 (추가/변경/삭제) 기능</li> <li>수집 주기(시간, 분, 초)를 설정하는 기능</li> </ul>	
	수집 에이전트	<ul style="list-style-type: none"> <li>에이전트 관리 및 운영(시작, 중지 등) 기능</li> <li>RESTFul방식의 API 기능</li> <li>다양한 형태의 수집정보(JSON, XML, CSV, C-Text) 기능</li> <li>수집된 정보(XML, JSON, CSV 등)와 C-Text와의 매핑 기능</li> </ul>	
	통계 및 모니터링	<ul style="list-style-type: none"> <li>각 수집 활동에 대한 자세한 통계 정보 제공 기능</li> <li>에이전트 오류, 이상동작에 대한 모니터링 및 복구기능</li> </ul>	
로그 데이터 수집	수집정보 관리	<ul style="list-style-type: none"> <li>수집 대상 정보(디렉토리, 파일, 확장자 등)와 연계방식 설정 기능</li> <li>수집 정보 및 수집정보 전달 대상에 대한관리 (추가/변경/삭제) 기능</li> <li>수집 주기(시간, 분, 초)를 설정하는 기능</li> </ul>	
	수집 에이전트	<ul style="list-style-type: none"> <li>에이전트 관리 및 운영(시작, 중지 등) 기능</li> <li>파일 수집시 설정된 저장단위로 파일을 수집하거나 로그 데이터를 토큰 (블록) 단위로 나누어 전송하는 기능</li> </ul>	

		<ul style="list-style-type: none"> <li>· 파일을 압축하여 전송하는 기능</li> <li>· 수집 대상 파일의 변경(추가, 수정, 삭제)을 감지하고 통보하는 기능</li> </ul>	
	통계 및 모니터링	<ul style="list-style-type: none"> <li>· 각 활동(로그파일 전송 등)에 대한 자세한 통계 정보 제공 기능</li> <li>· 에이전트 오류, 이상동작에 대한 모니터링 및 복구기능</li> </ul>	
	관련 솔루션	<ul style="list-style-type: none"> <li>· 오픈소스 : Apache Flume, fluentd, Graylog, Logstash</li> <li>· 상용 : Splunk, TIBCO LogLogic, VMWare vRealize Log Insight</li> </ul>	
Open API 수집	수집정보 관리	<ul style="list-style-type: none"> <li>· 수집 대상 정보(URI)에 대한 연계방식 설정 기능</li> <li>· 수집 정보 및 수집정보 전달 대상에 대한관리 (추가/변경/삭제) 기능</li> <li>· 수집 주기(시간, 분, 초)를 설정하는 기능</li> </ul>	
	수집 에이전트	<ul style="list-style-type: none"> <li>· 에이전트 관리 및 운영(시작, 중지 등) 기능</li> <li>· Open API 및 RESTFul 방식의 Open API 기능</li> <li>· 다양한 형태의 수집정보(JSON, XML, RSS, CSV) 기능</li> <li>· 수집된 정보(XML, JSON 등)와 DBMS와의 매핑 기능</li> </ul>	
	통계 및 모니터링	<ul style="list-style-type: none"> <li>· 각 수집 활동에 대한 자세한 통계 정보 제공 기능</li> <li>· 에이전트 오류, 이상동작에 대한 모니터링 및 복구기능</li> </ul>	
웹 수집 (Crawler)	수집정보 관리	<ul style="list-style-type: none"> <li>· 수집 정보 및 수집정보 전달 대상에 대한관리 (추가/변경/삭제) 기능</li> <li>· 수집 주기(시간, 분, 초)를 설정하는 기능</li> <li>· 구성관련 정보를 자동으로 반영하거나 자유롭게 입력 할 수 있는 관리 기능</li> </ul>	웹 수집 데이터를 플랫폼으로 연계하기 위해서는 OOOO 자료연계 서버 이용
	수집 에이전트	<ul style="list-style-type: none"> <li>· 수집대상리스트에서 웹문서 등 콘텐츠를 수집하는 기능</li> <li>· 규칙, 확률 또는 학습 기반으로 문서를 분류하는 기능</li> <li>· 데이터 수집 시 불필요한 수집이 일어나지 않도록 사전에 대상을 등록하여 선별적 수집이 되도록 하는 기능</li> <li>· 웹 사이트에 정보 수집을 방지하기 위한 로봇배제</li> </ul>	

		표준(robots.txt) 및 메타태그 준수	
	통계 및 모니터링	<ul style="list-style-type: none"> <li>· 각 활동(크롤링 구문 분석과 인덱싱 등)에 대한 자세한 통계 정보 제공 기능</li> <li>· 에이전트 오류, 이상동작에 대한 모니터링 및 복구기능</li> </ul>	
	관련 솔루션	<ul style="list-style-type: none"> <li>· 오픈소스 : Apache Nutch, Heritex, Scrapy</li> <li>· 상용 :와이즈넷 WISE BICrawler, 솔트룩스 Tornado,KONAN Social Crawler, 80 legs, Mozenda</li> </ul>	
API 수집	수집정보 관리	<ul style="list-style-type: none"> <li>· 수집 대상 정보(URI)에 대한 연계방식 설정 기능</li> <li>· 수집 정보 및 수집정보 전달 대상에 대한관리 (추가/변경/삭제) 기능</li> <li>· 수집 주기(시간, 분, 초)를 설정하는 기능</li> </ul>	
	수집 에이전트	<ul style="list-style-type: none"> <li>· 에이전트 관리 및 운영(시작, 중지 등) 기능</li> <li>· RESTFul방식의 API 기능</li> <li>· 다양한 형태의 수집정보(JSON, XML, CSV, C-TeX) 기능</li> <li>· 수집된 정보(XML, JSON, CSV 등)와 C-TeX와의 매핑 기능</li> </ul>	
	통계 및 모니터링	<ul style="list-style-type: none"> <li>· 각 수집 활동에 대한 자세한 통계 정보 제공 기능</li> <li>· 에이전트 오류, 이상동작에 대한 모니터링 및 복구기능</li> </ul>	

### 3. 빅데이터 수집 방안

중앙전파관리소 내부 망에 존재하는 데이터를 수집하는 경우에는 DBMS 수집 방식을 통해 대상 데이터를 수집한다.

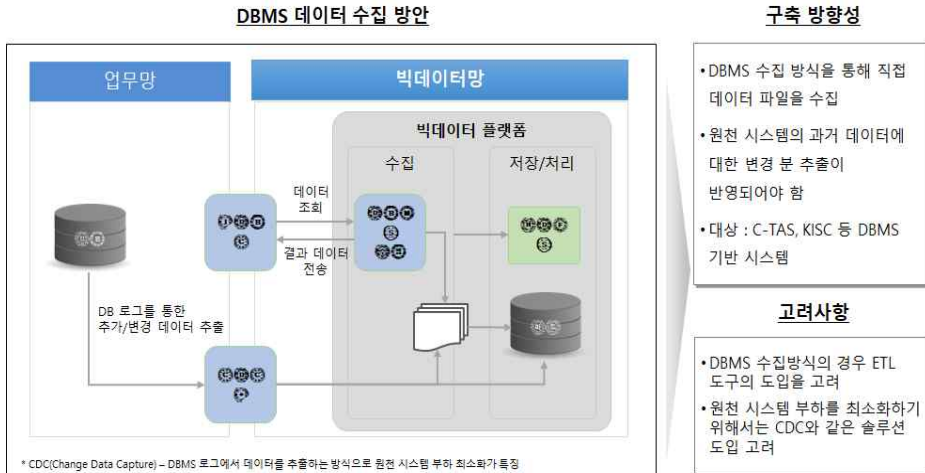


그림 5-13 DBMS 데이터 수집 방안

REST API를 이용하는 내부시스템의 경우에는 수집서버가 직접 전송된 데이터를 받아 처리하고, REST API를 제공하는 내부 시스템의 경우에는 수집서버가 주기적으로 접근해 데이터를 수집한다.

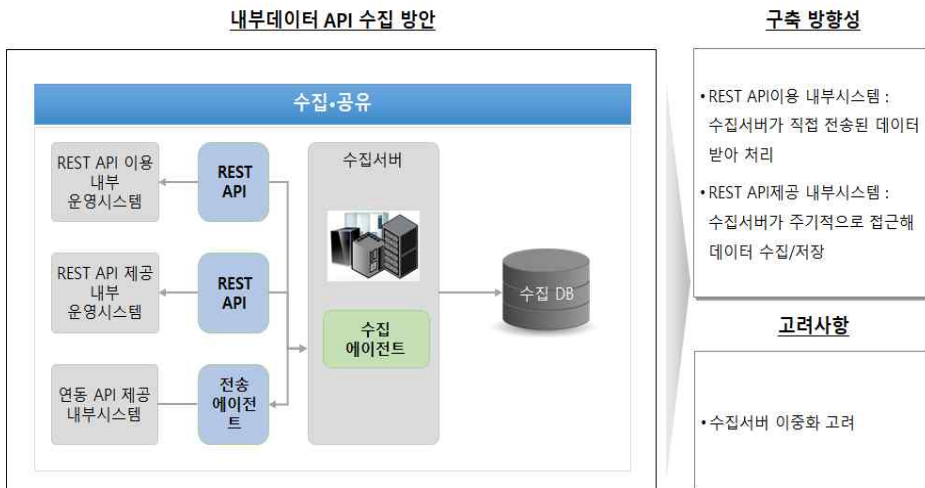


그림 5-14 내부 데이터 API 수집 방안

오픈데이터의 수집대상 데이터가 선정되면 API를 검색해 활용 신청하여 인증키를 부여받아 데이터를 수집한다.

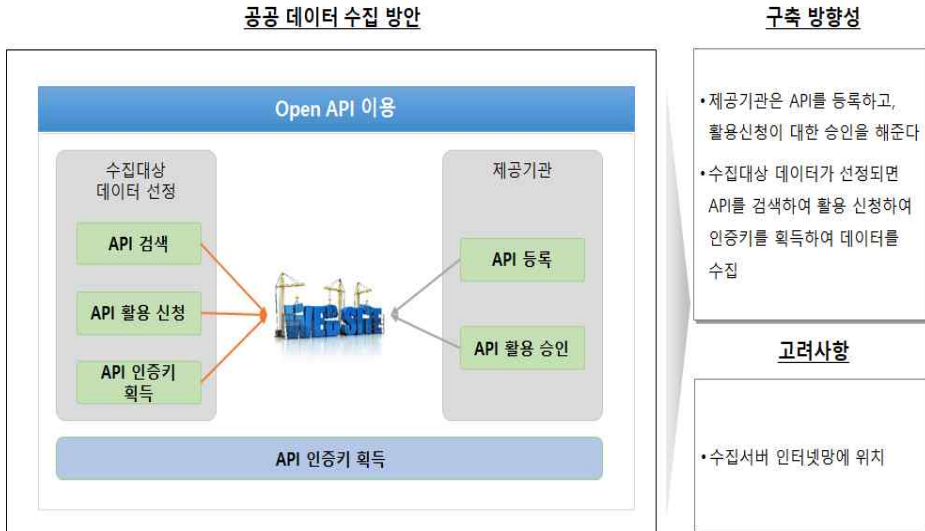


그림 5-15 공공 데이터 수집 방안

국내외 주요 인텔리전스 웹사이트의 데이터는 웹사이트 크롤링과 자체 API, REST API를 이용하여 수집한다.

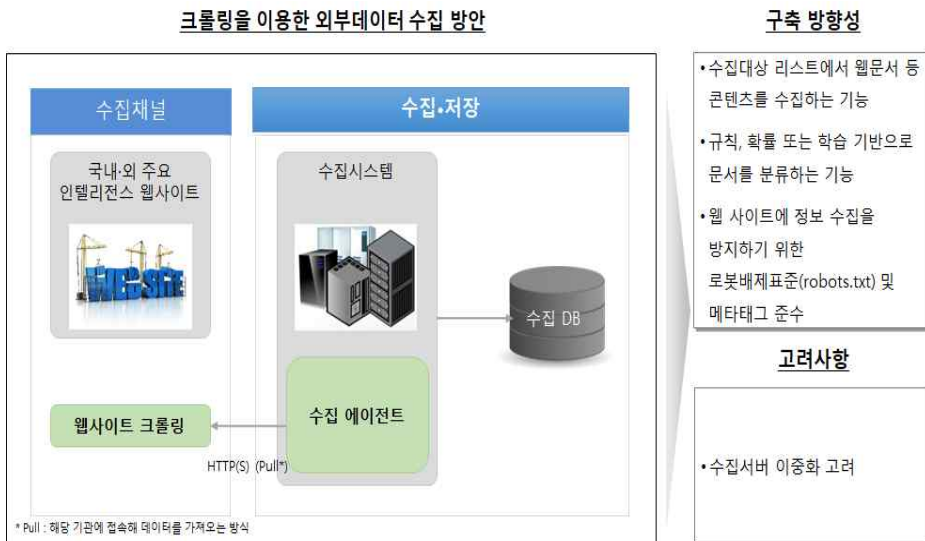
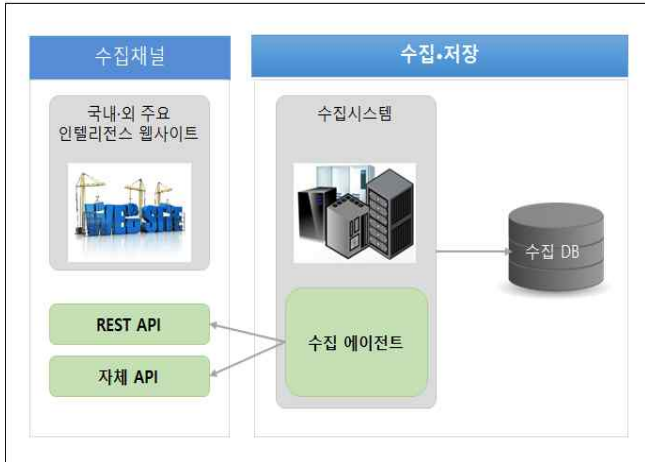


그림 5-16 크롤링을 이용한 외부 데이터 수집 방안



### 외부데이터 수집 방안



### 구축 방향성

- CISCO Threat Grid, Umbrella와 같은 외부 상용 인텔리전스 데이터는 제공하는 API를 이용해서 수집
- RESTful API를 이용해 수집
- 수집에이전트에서 정상적인 데이터수신을 체크한 뒤 수집DB에 저장

### 고려사항

- 수집서버 이중화 고려

그림 5-17 외부 데이터 수집 방안

## 제4절 빅데이터 저장 플랫폼 구성

### 1. 빅데이터 저장 플랫폼 개요

빅데이터 저장 플랫폼은 수집 데이터 저장 및 처리를 위한 인프라 구성을 통해 기관의 내부 빅데이터 및 외부 데이터를 저장하여 사용자들에게 다양한 분석을 수행할 수 있는 기반을 제공한다.

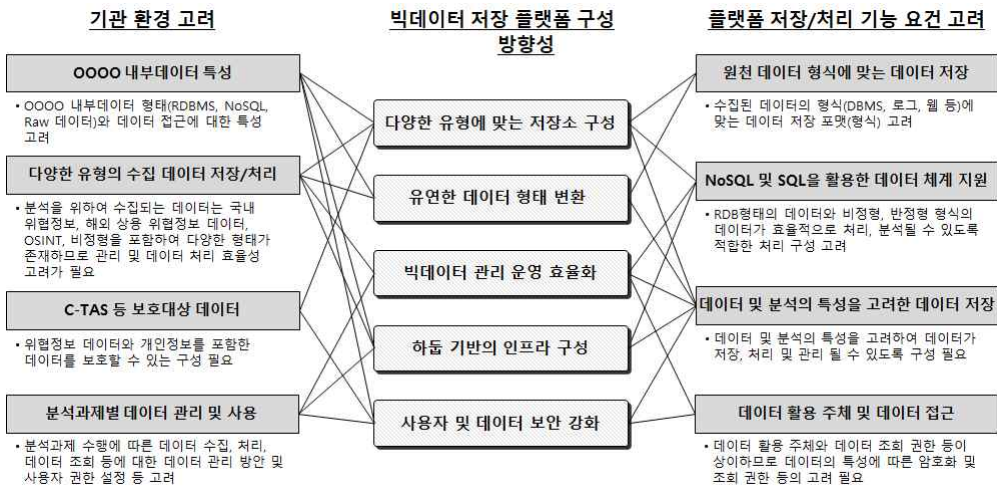


그림 5-18 빅데이터 저장 플랫폼 구성 시 고려사항

빅데이터 플랫폼에서 데이터를 저장 및 관리하기 위한 공통 데이터 저장 및 관리 아키텍처는 그림 5-19와 같다.



그림 5-19 공통 데이터 저장 및 관리 컴포넌트

## 2. 빅데이터 저장 플랫폼 구성 및 기능

빅데이터 플랫폼 내의 리소스와 데이터 작업을 수행하기 위한 리소스 및 데이터 관리 컴포넌트의 구성요소는 표 5-5와 같고, 빅데이터 플랫폼 내의 실제 빅데이터가 저장되는 데이터 분산파일 시스템인 하둡을 구성하기 위해 데이터 저장 및 처리를 위한 컴포넌트의 구성요소는 표 5-6과 같다.

표 5-5 리소스 관리 및 데이터 관리

영역	구성요소	기능 설명	고려사항
리소스 관리	리소스 매니저	<ul style="list-style-type: none"> <li>· 빅데이터 클러스터 내의 자원 상태를 관리하는 기능</li> <li>· 노드 매니저와 통신하면서 컨테이너 실행 및 모니터링</li> </ul>	
	타임라인 서버	<ul style="list-style-type: none"> <li>· 어플리케이션 수행 이력 저장 및 관리</li> </ul>	
데이터 관리	네임 노드	<ul style="list-style-type: none"> <li>· 네임노드와 secondary 네임노드로 구성</li> <li>· 파일시스템의 디렉터리, 파일명, 파일블록 등 네임 스페이스를 관리하는 일종의 마스터 역할 기능</li> <li>· 정의된 값에 따라 주기적으로 네임 노드의 파일 시스템 이미지를 스냅샷해 생성하는 기능</li> <li>· 클라이언트에서 작업 요청 시 메모리에 저장된 메타정보를 이용하여 파일 입출력을 수행할 수 있도록 관리</li> </ul>	리소스 관리 (Resource Manager) 및 네임노드는 이중화 구성 필요
	Secondary 네임노드 (또는 Standby 네임노드)	<ul style="list-style-type: none"> <li>· Secondary 네임노드는 정의된 값에 따라 주기적으로 네임 노드의 파일 시스템의 이미지를 스냅샷해 생성 (이중화 구성 시 Standby 네임노드가 수행)</li> </ul>	

표 5-6 데이터 저장 및 처리 노드

영역	구성요소	기능 설명	고려사항
데이터 저장 및 처리 노드	데이터 노드	<ul style="list-style-type: none"> <li>· 파일에 대한 실질적인 데이터 입출력에 대한 처리 기능</li> <li>· 데이터 유실을 방지하기 위해 동일한 파일을 여러 개의 블록으로 나누어 저장하며 3벌로 복제하여 분산 저장</li> </ul>	<ul style="list-style-type: none"> <li>· 데이터 3벌 복제 시 Rack Awareness*를 고려하여 2벌은 동일 랙 (Rack)에 위치하고 1벌은 물리적으로 다른 랙에 위치하도록 구성</li> </ul>
	노드 매니저	<ul style="list-style-type: none"> <li>· 전체 데이터 처리 노드에서 수행되는 프로세스</li> <li>· 리소스 매니저와 통신하면서 어플리케이션을 수행하기 위한 컨테이너 실행 및 관리</li> </ul>	
	컨테이너	<ul style="list-style-type: none"> <li>· CPU, 디스크, 메모리 등과 같은 속성으로 정의</li> <li>· 하나의 작업은 여러 Task로 분리되어 실행되며, 분리된 각 Task들은 하나의 컨테이너 내에서 실행</li> </ul>	
	어플리케이션 마스터	<ul style="list-style-type: none"> <li>· 하나의 프로그램에 대한 마스터 역할 수행</li> <li>· 수행되는 프로그램에 대한 실행상태를 모니터링 및 관리</li> <li>· 필요한 자원에 대해 Resource Manager에게 요청</li> </ul>	

빅데이터 플랫폼에서 데이터를 저장하기 위해서는 수집된 원천데이터의 유형을 구분하고 압축률, 쓰기성능, 읽기 성능을 고려하여 사용목적에 따라 데이터 저장 파일 형식을 선택하여 저장해야 한다.

## 원천 데이터 형태



원천 데이터  
성격에 따른  
파일 변환

## 하둡(Hadoop) 저장 방안

원천	파일 포맷	압축 률	쓰기 성능	읽기 성능	호환 성	내 용
DBMS	Parquet	상	하	상	중	<ul style="list-style-type: none"> <li>컬럼 기반 바이너리 형태의 파일로 DBMS 데이터 형식을 저장할 때 사용</li> <li>컬럼 별 기본 통계를 보유하고 있고 특정 컬럼 위주로 조회하는 경우 효율적으로 I/O 처리</li> </ul>
	Orc	상	하	상	하	<ul style="list-style-type: none"> <li>컬럼 기반 바이너리 형태의 파일로 DBMS 데이터 형식을 저장할 때 사용 (Hive 저장 포맷으로 개발됨)</li> <li>컬럼 별 기본 통계 보유 (min, max, sum, count)</li> </ul>
CSV JSON	TEXT	하	상	하	상	<ul style="list-style-type: none"> <li>csv, json 형식 등의 파일</li> <li>다양한 어플리케이션 간에 호환성이 좋음</li> <li>별도의 변환 없이 사람이 인식할 수 있는 형태</li> </ul>
Binary	Sequence	하	상	중	중	<ul style="list-style-type: none"> <li>바이너리 형식의 key/value 형식</li> <li>Map-Reduce 작업 수행 시 map 과 reduce 간에 데이터 전달을 위해 사용됨</li> </ul>
JSON	Avro	상	중	하	중	<ul style="list-style-type: none"> <li>주로 JSON 포맷을 직렬화(Serialization) 하여 다양한 어플리케이션 (MapReduce, Pig, Kafka 등) 에서 활용</li> </ul>

그림 5-20 빅데이터 저장 플랫폼 구성방안

## 3. 데이터 보안 및 작업관리

빅데이터 플랫폼에서 데이터 보안 및 작업관리를 위한 데이터 보안 및 작업관리 컴포넌트 아키텍처이다.



그림 5-21 데이터 보안 및 작업관리 컴포넌트

빅데이터 공통기반내의 보안 컴포넌트 구성 요소로서 사용자 인증, 접근 제어, 데이터 암호화, 개인정보 비식별화, 보안 모니터링으로 구성되고 빅데이터 플랫폼의 전체 클러스터를 관리하기 위한 구성 요소로서 설치/업그레이드, 통합 관리, 구성관리, 모니터링 등으로 구성한다.

표 5-7 데이터 보완 및 작업관리 기능 설명

영역	구성요소	기능 설명	고려사항
보안 관리	사용자 인증	· OOOO SSO 와의 사용자 인증 연동 기능 · 사용자가 플랫폼에 접근할 경우 데이터에 어떤 조치를 취할 수 있는가를 미리 정할 수 있어야 함	· 암호화는 하둡의 T D E (Transparent Data Encryption) 기능 사용가능 · 사용자 인증 및 개인정보 비식별화에 대해서는 별도 솔루션 필요
	접근 제어	· 사용자가 플랫폼 및 데이터 마트 데이터에 대한 조회권한을 정의하는 기능 · 데이터 조회에 대한 권한을 가지고 있는지를 체크하고 통제하는 기능	
	암호화	· 개인정보 보호를 위해 개인정보보호법을 준수하는 개인정보 비식별화 기술을 지원해야 함 · 검증된 암호화 알고리즘을 지원해야 함 · 암호화 키는 별도의 안전한 서버에서 관리되어야 함 · 인가된 사용자만이 검증된 알고리즘을 사용하여 암호/복호화 하여야 하며, 암호/복호화 정보 기록과 감사 정보가 보관되어야 함	
	개인정보 비식별화	· 저장되는 데이터에 개인정보 포함여부를 사전에 체크하고 통제하는 기능 · 개인정보 데이터를 비식별하고 관리하는 기능	
클러스터 관리	보안 모니터링	· 빅데이터 플랫폼 사용자에 대한 내부 감사/통제 기능 · 플랫폼 접근 내역에 대해 권한 기반 접근제어/마스킹 등의 접근 통제 보안 정책을 지원하는 기능	
	통합 관리	· 빅데이터 클러스터에 대한 전체 통합 관리 기능 · 관리자 UI를 통한 개별 구성요소에 대한 관리 기능 · 빅데이터 클러스터내의 각 모듈들을 시작, 중지, 재 시작 할 수 있는 기능 · 사용자 및 그룹에 대한 생성, 변경, 삭제 기능 · 사용자 및 그룹별 화면 조회 및 변경 권한 부여 기능	
	설치/업그레이드	· 빅데이터 플랫폼 내의 빅데이터 클러스터에 설치되는 각 모듈들을 설치 및 패치, 버전 업그레이드 할 수 있는 기능	

		<ul style="list-style-type: none"> <li>· 데이터 증가에 의한 노드 추가 등으로 확장 (Scale-out)시에 관리할 수 있는 기능</li> </ul>	
	구성관리	<ul style="list-style-type: none"> <li>· 각 구성요소에 대한 설정을 변경 및 적용할 수 있는 기능</li> <li>· 성능 구성과 관련된 값(메모리, 버퍼 등)을 설정하고 조정할 수 있는 기능</li> </ul>	
워크플로우 관리	작업관리 설정	<ul style="list-style-type: none"> <li>· 작업에 대한 정의 및 작업에 대한 구성 관리 기능</li> <li>· 전체 빅데이터 플랫폼 클러스터 내에서의 작업 컨트롤 기능</li> </ul>	
	작업스케줄 관리	<ul style="list-style-type: none"> <li>· 작업 등록 및 예상스케줄 관리를 통한 작업 예약 기능</li> <li>· 작업 선 후행 관리 기능</li> </ul>	
	작업 실행	<ul style="list-style-type: none"> <li>· 작업 직접 수행 및 중지</li> </ul>	
	작업 모니터링	<ul style="list-style-type: none"> <li>· 작업 성공 및 실패에 대한 로그 관리 기능</li> <li>· 작업 수행 상태 및 전체 클러스터내의 작업 모니터링</li> <li>· 작업 관련 보고서 기능</li> </ul>	
	관련 솔루션	<ul style="list-style-type: none"> <li>· TeraOne, Apache Oozie, Azkaban (by Linkedin)</li> </ul>	
백업	OS 영역 백업	<ul style="list-style-type: none"> <li>· OS 및 OS 영역 내에 존재하는 어플리케이션에 대한 백업</li> </ul>	<ul style="list-style-type: none"> <li>· 데이터 백업을 위해서는 별도의 백업 솔루션 필요</li> <li>· 어플리케이션 버전 관리를 위해서는 버전 관리 컨트롤 소프트웨어 활용</li> </ul>
	어플리케이션 버전관리	<ul style="list-style-type: none"> <li>· 버전관리 컨트롤을 통한 어플리케이션 버전 관리 기능</li> </ul>	
	빅데이터 플랫폼 백업	<ul style="list-style-type: none"> <li>· 원본데이터 백업 수행하는 기능</li> <li>· 빅데이터 플랫폼 내의 하둡 데이터를 백업 수행하는 기능</li> <li>· 빅데이터 플랫폼 내의 NoSQL 데이터를 백업 수행하는 기능</li> </ul>	
	데이터 마트 백업	<ul style="list-style-type: none"> <li>· 데이터 마트 DBMS에 대한 전체 백업(Full Backup) 수행 기능</li> <li>· 데이터 증가분에 대해서만 백업(Incremental Backup)할 수 있는 기능</li> </ul>	

또한, 빅데이터 플랫폼 내의 워크플로우 관리 컴포넌트의 구성요소로서 데이터 처리를 위한 작업 관리, 작업 스케줄 관리, 모니터링을 포함하여 구성하고, 빅데이터 플랫폼 내의 백업 컴포넌트 구성요소로서 OS 및 어플리케이션 백업, 데이터 백업, DBMS 기반의 데이터 마트 백업으로 구성한다.

#### 4. 데이터 보안 및 작업관리 구성 방안

빅데이터 플랫폼은 다수의 서버와 다양한 기술이 혼재하여 구성됨으로써 일원화된 관리가 쉽지 않으므로 빅데이터 환경에 맞는 체계화된 관리가 필요하다.

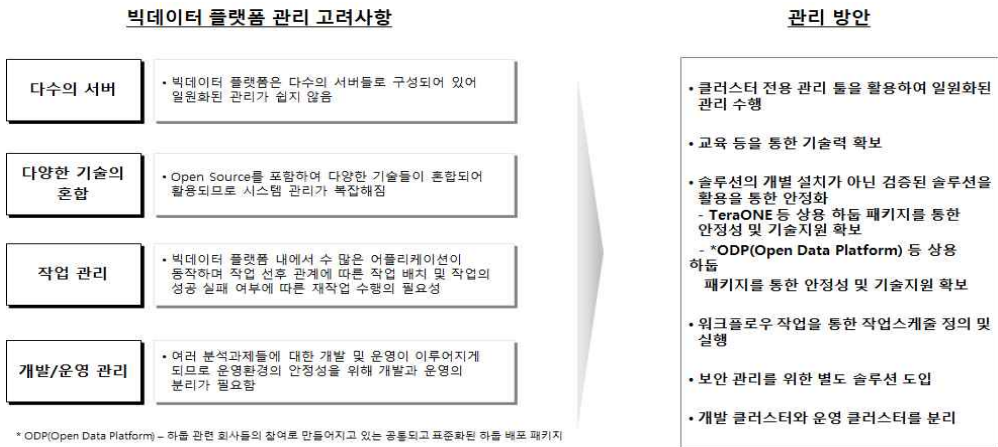


그림 5-22 빅데이터 플랫폼 관리 고려 사항

빅데이터 플랫폼에서의 보안은 데이터 수집, 저장, 분석과 데이터를 사용하기 위해 접근하는 사용자 접근 제어 및 모니터링, 데이터 암호화를 통한 정보보호 측면을 고려하여 구성하여야 한다.

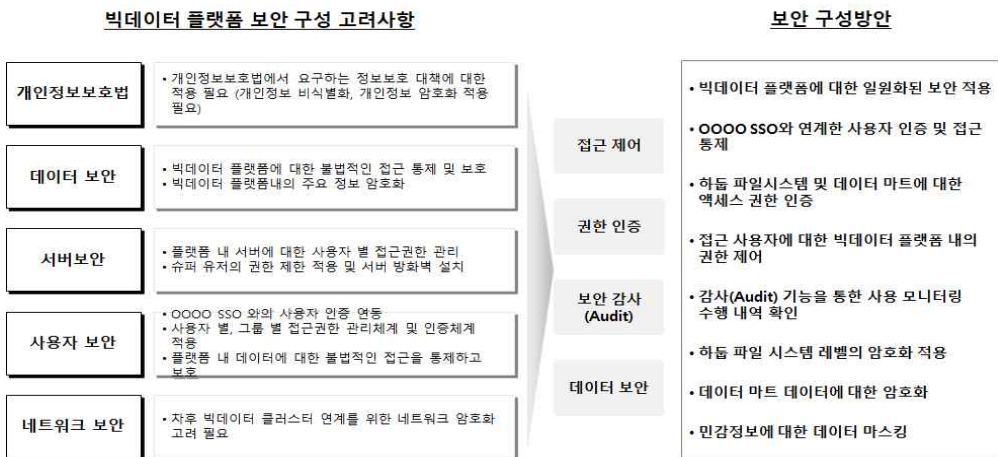


그림 5-23 빅데이터 플랫폼 보안 구성 고려 사항



빅데이터 플랫폼에 저장된 중요정보의 유출을 방지하기 위해 사용자 인증, 접근통제 및 권한 할당, 사용자 작업에 대한 모니터링을 수행하고 민감 데이터의 경우 암호화 기능을 적용하여 데이터를 보호해야 한다.

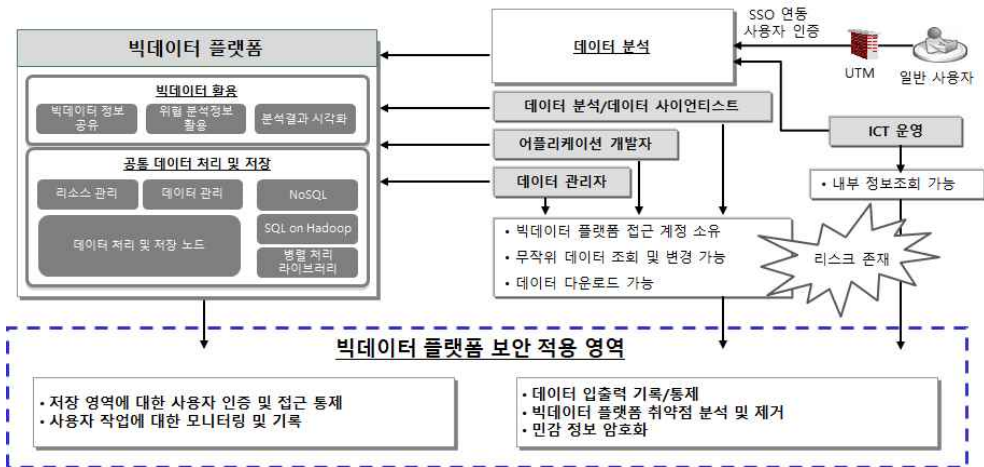


그림 5-24 빅데이터 플랫폼 보안 적용 방안

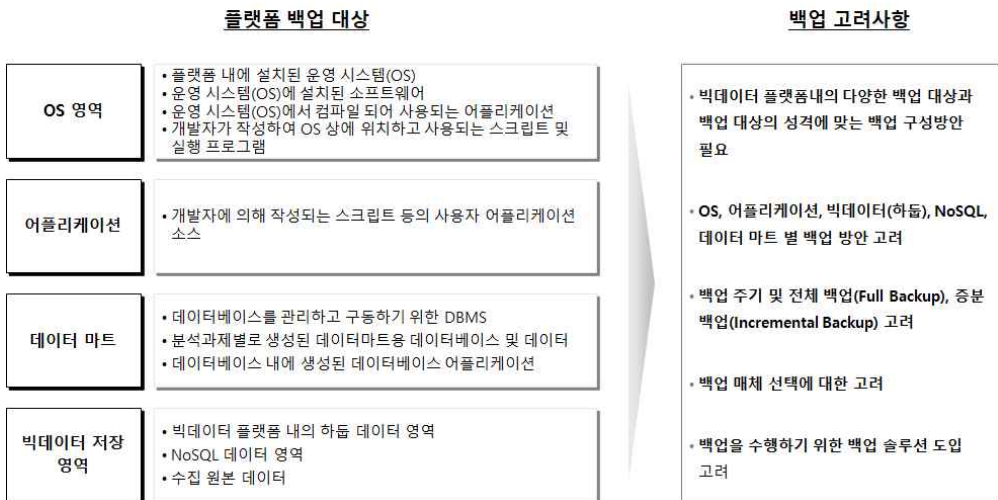


그림 5-25 빅데이터 플랫폼 백업 대상

빅데이터 플랫폼을 운영하기 위해 기본이 되는 부분으로 운영 시스템(OS), 설치된 소프트웨어 및 구성정보(Configuration) 등을 포함하므로 디스크 미러링, 파일 시스템, 이미지 백업의 3가지 백업 구성을 권고한다.


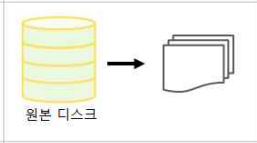
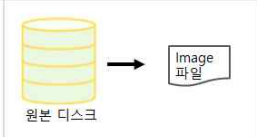
백업 방안	설명	비고
<div data-bbox="193 533 303 701"> <p>디스크 미러링 구성</p>  </div>	<ul style="list-style-type: none"> <li>• 추가 디스크 구성을 통해 OS 영역의 디스크를 미러링 함으로써 디스크 장애 발생시 미러링 구성된 디스크로 대체</li> <li>• 장점 : 추가 디스크 구성으로 간단히 구현 가능</li> <li>• 단점 : 잘못된 정보까지 미러링 됨으로써 원본 디스크 이미지의 손상이 발생할 경우 미러링 디스크로 손상되기 때문에 정상복구가 불가능</li> </ul>	<ul style="list-style-type: none"> <li>• 미러링 구성 권고</li> </ul>
<div data-bbox="193 701 303 842"> <p>파일시스템 백업</p>  </div>	<ul style="list-style-type: none"> <li>• 백업 받고자 하는 특정 파일시스템을 선별하여 백업 수행</li> <li>• 장점 : 백업이 필요한 파일 시스템을 선별하여 백업 가능</li> <li>• 단점 : 복원 시에는 운영체제(OS)를 미리 설치하고 백업 수행 필요</li> </ul>	<ul style="list-style-type: none"> <li>• 중요 파일 시스템 백업 권고</li> </ul>
<div data-bbox="193 842 303 977"> <p>이미지 백업</p>  </div>	<ul style="list-style-type: none"> <li>• 원본 디스크의 이미지 파일을 생성하여 백업</li> <li>• 장점 : 원본과 동일한 이미지를 생성하므로 동일한 백업 생성 가능</li> <li>• 단점 : 복원 시 디스크의 파티션이 원본 이미지와 동일한 형태로 구성 필요</li> </ul>	<ul style="list-style-type: none"> <li>• 이미지 백업 권고</li> </ul>

그림 5-26 OS/어플리케이션 백업 방안

데이터 마트는 데이터베이스의 백업 방안을 사용하며 백업 정책에 의해 전체 백업, 증분 백업, 차등 백업을 조합하여 백업 수행하는 것을 권고한다.

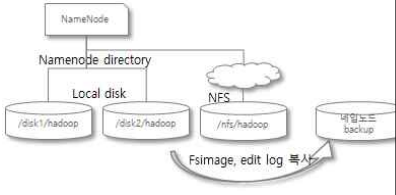

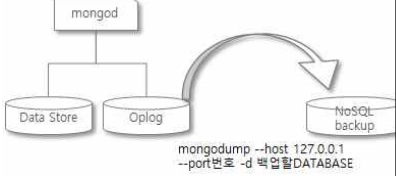
표 5-8 데이터 마트 백업 방안

백업 모드	내 용
Full	· 데이터베이스에 존재하는 전체 데이터를 백업
Schema	· 데이터는 백업하지 않고, 데이터베이스에서 사용하는 스키마 (테이블, 뷰, 프로시저, 권한 등) 만을 백업
Incremental (증분 백업)	· 이전 백업(Full, Differential, Cumulative) 이후에 데이터베이스에 변경된 내용을 백업
Differential (차등 백업)	· 이전 Full 백업 이후에 데이터베이스에 변경된 내용을 백업. 두 가지 백업본 (Full, Differential)만 가지고 데이터 복구



그림 5-27 데이터 백업 수행 예시

표 5-9 네임노드 백업 방안-하둡 플랫폼 백업

백업 방안	내 용	백업 작업 예시	비고
네임노드 디렉토리 복사	<ul style="list-style-type: none"> <li>· 네임노드 디렉토리는 여러 경로에 설정 가능 <ul style="list-style-type: none"> <li>- 예시: 네임노드 파일시스템 2개, NFS 1개</li> </ul> </li> <li>· 네임노드 백업을 위해 네임노드 디렉토리에 있는 fsimage 및 edit 파일 등을 백업 스토리지에 복사</li> <li>· 네임노드 저장공간 만큼의 백업 공간 필요(디렉토리 및 block 이 네임노드 메모리에서 150 bytes 정도를 사용하므로 3복제 적용시 대략 100만개 파일당 1GB 메모리 필요)</li> </ul>		백업 권고
하둡 스냅샷 (snapshot)	<ul style="list-style-type: none"> <li>· 하둡스냅샷 기능을 이용하여 특정 시점의 파일시스템 정보 저장</li> <li>· 데이터 block 을 추가로 복사하지 않음</li> </ul>		백업 권고
MongoDB 백업	<ul style="list-style-type: none"> <li>· MongoDB는 기본적으로 복제기능을 제공</li> <li>· 서버와 네트워크 장애 발생에 대비해 다른 서버에 백업하고 해당 서버가 동작이 불가능할 경우 다른 보조 서버에서 서버의 역할을 담당함</li> <li>· 마스터 - 슬레이브 방식으로 복제</li> <li>· 복제 기능과 별도로 mongodump 명령으로 백업</li> <li>· Dump 폴더에 백업할 데이터</li> <li>· 베이스명이 기록되어 있으면 백업완료</li> </ul>		백업 권고

향후 DR 구축을 위해서는 동일한 저장 용량의 하둡 백업 클러스터를 구성해야 하며, 주기적으로 distcp를 통해 데이터를 복사하거나 수집 단계에서 두 개의 클러스터에 모두 데이터를 적재하는 방식으로 백업을 수행해야 한다.

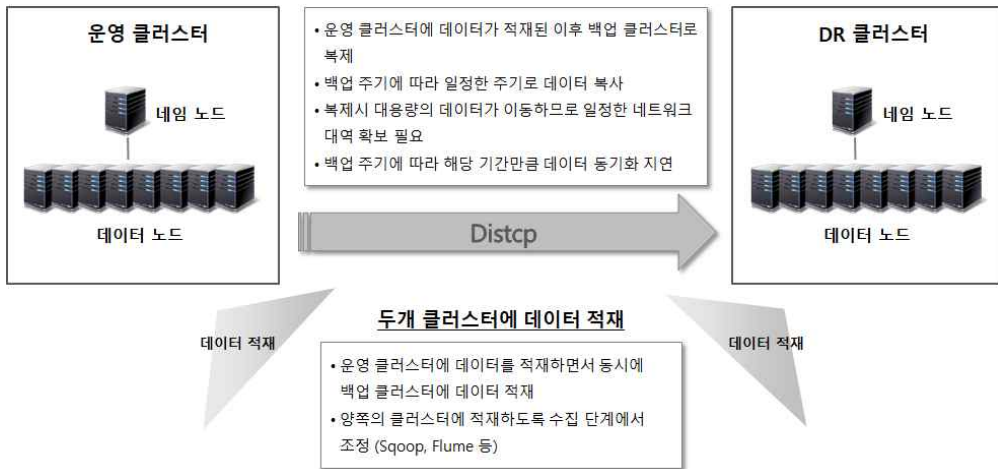


그림 5-28 DR 구축 방안

빅데이터 플랫폼 운영을 위해서는 빅데이터 관리 및 인프라 운영을 위한 별도의 담당자 지정이 필요하며, 빅데이터 관리는 데이터 수집 및 DBA 역할, 인프라 운영은 플랫폼내의 시스템 운영을 담당하도록 지정되어야 한다.

표 5-10 빅데이터 플랫폼 운영 인력 산정

역할 구분	역할 정의	세부업무 정의	기능 구분	예상인원
빅데이터 운영	Data수집, 적재 DBA	<ul style="list-style-type: none"> <li>· 원천 시스템으로부터 데이터 수집, 변환, 적재 업무</li> <li>· 데이터초기 및 변경 적재</li> <li>· 데이터 정합성 검증</li> <li>· 하둡 데이터 관리</li> <li>· NoSQL 관리</li> <li>· 데이터 마트 DBMS 관리</li> </ul>	빅데이터 플랫폼 데이터 관리	1명

인프라 운영	빅데이터 플랫폼 시스템 운영	<ul style="list-style-type: none"> <li>· 인프라 관리 및 모니터링</li> <li>· 빅데이터 플랫폼 하드웨어 관리</li> <li>· 빅데이터 플랫폼 소프트웨어 관리</li> <li>· 사용자 지원 및 교육</li> </ul>	빅데이터 플랫폼 시스템 관리	1명
--------	-----------------	---	-----------------	----

## 5. ETL 및 데이터 정제방안

빅데이터 플랫폼의 ETL은 데이터 원천 시스템의 부하를 최소화하도록 고려되어야 하며 ETL 성능과 작업 프로세스가 자동화 되도록 구성되어야 한다.

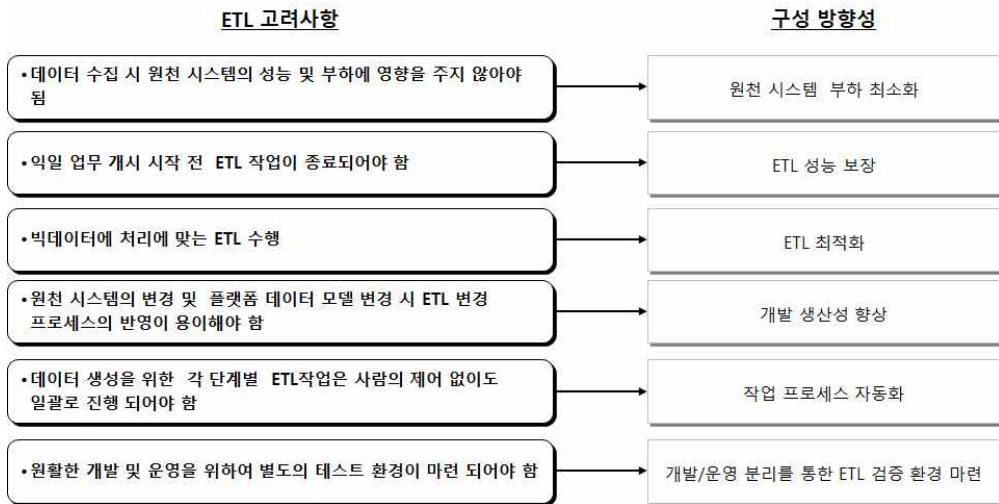


그림 5-29 ETL 고려사항 및 구성

ETL 처리를 위해서는 원천데이터로부터 데이터 추출, 정제, 처리, 플랫폼으로의 데이터 적재가 이루어져야 한다.

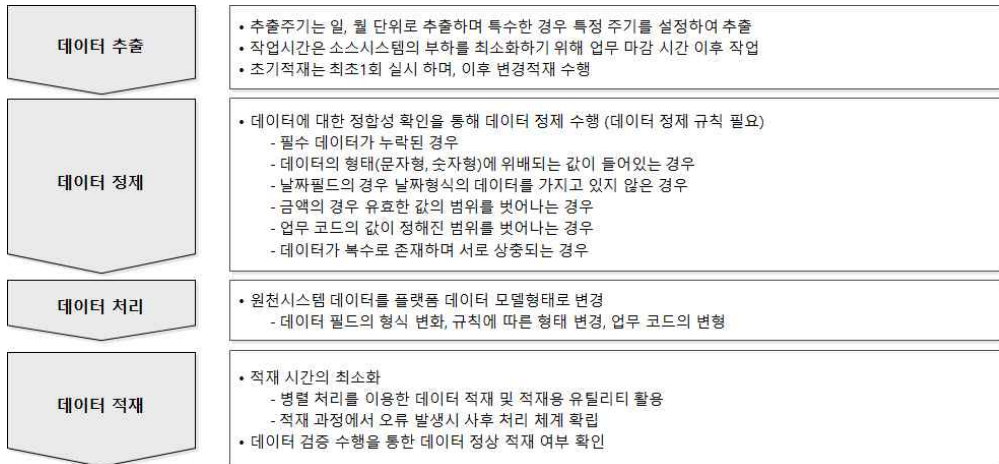


그림 5-30 ETL 처리 방안

빅데이터 플랫폼은 원천 시스템에서 수집된 데이터의 성격에 따른 분리작업과 Mapping 작업을 통해 하둡 파일 시스템(HDFS)에 디렉토리별로 데이터가 저장되도록 구성되어야 한다.

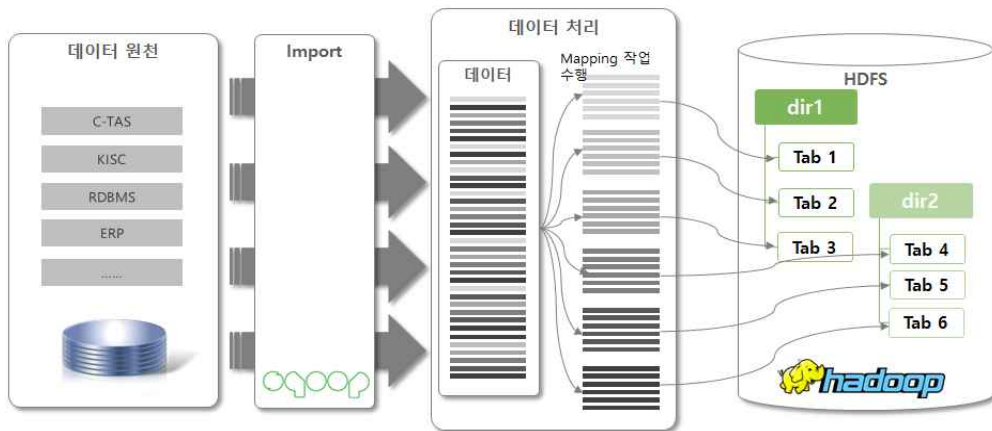


그림 5-31 ETL 흐름도

수집된 데이터에 대해 6가지 품질지표 (DQI : Data Quality Indicator) 의 유형별 측정기준을 수립하여 데이터 정제 방안을 수립하고 데이터 품질을 관리해야 한다.

품질지표 유형	품질지표 유형 설명	품질 측정 예시
충실성 (Completeness)	• 데이터에 값이 채워져 있는 비율	• 데이터 필드에 값이 채워져 있는가?
유효성 (Validity)	• 데이터가 Business Rule 또는 일반적 데이터 구성 기준에 부합하는 비율	• 숫자 필드에 숫자가 아닌 값이 존재하는가?
일치성 (Consistency)	• 연관성을 갖는 항목의 데이터 값들이 서로 논리적으로 모순이 되지 않는 비율	• 주소의 지역과 전화번호의 지역이 서로 일치하는가?
명료성 (Comprehensiveness)	• 입력된 데이터 값의 구분이 모호하지 않은 비율	• 의미가 정의되지 않은 코드값이 존재하는가?
정확성 (Accuracy)	• 데이터의 값이 실제 상황을 확실히 반영하는 비율	• 보유중인 데이터가 최근의 정보인가?
유일성 (Uniqueness)	• 입력된 데이터가 유일하게 존재하는 비율	• 고객번호가 유일하게 존재하는가?

그림 5-32 6가지 품질 지표

데이터의 정제는 원천 데이터 오류, 수행 데이터 오류, 최종 데이터 오류 등 3가지 요인에 의해 발생할 수 있으며 발생한 데이터의 오류는 데이터 또는 어플리케이션 수정을 통해 해결해야 한다.

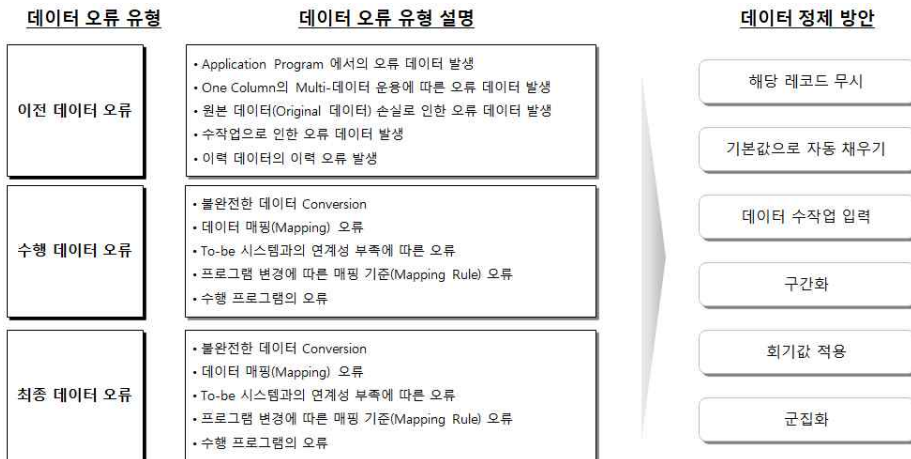


그림 5-33 데이터 오류 유형 별 정제 방안





분석을 통하여 얻고자 하는 해답의 목적 및 의사결정을 위한 자동화 정도에 따라 분석 단계가 구분되며, 분석 플랫폼에서는 각각의 분석이 이루어질 수 있어야 한다.

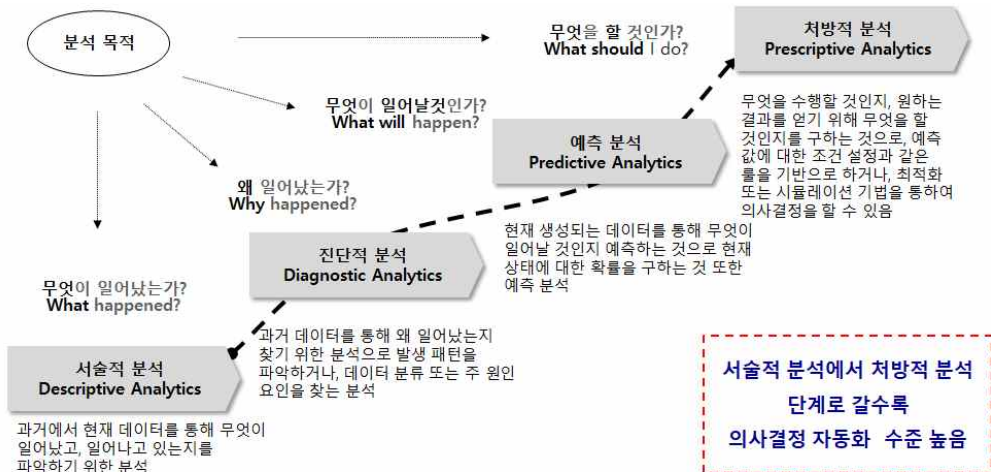


그림 5-36 빅데이터 분석 단계

## 2. 빅데이터 분석 기술



그림 5-37 빅데이터 분석 기술

빅데이터 분석기술 중 하나인 실시간 분석은 스트리밍 데이터를 이동 상태에서 실시간으로 분석하는 기술이다.

의사결정 및 최적화 기술은 비즈니스 룰 또는 최적화 기법을 기반으로 의사 결정을 위해 최적의 대안을 도출하는 기술이다.

디스커버리 분석은 데이터 내 패턴 및 연관성을 발견하기 위해 인터랙티브 방식으로 탐색하는 기술이다.

예측 분석은 통계 및 마이닝 기법을 통해 모델링 작업을 수행하고, 패턴 분류 및 예측을 하기 위한 기술이다.

다차원 분석은 사용자가 원하는 형식의 비정형 질의 및 다차원 분석을 수행하기 위한 기술이다.

텍스트 분석은 자연어 처리 기술을 적용하여 비정형 텍스트 데이터에서 특성을 추출하고 정형화하여 텍스트에 담긴 콘텐츠를 분석하기 위한 기술이다.

### 3. 빅데이터 분석 유형

그림 5-38은 빅데이터 및 사이버 위협정보 분석을 위해 실시간 데이터 분석, 대용량 데이터 기반 분석 모델 생성, 사용자 점점 분석 영역으로 구분되고 6가지 분석 유형으로 구성된 분석 컴포넌트를 나타낸다.



그림 5-38 빅데이터 분석 컴포넌트

표 5-11는 저장되기 이전의 스트리밍 데이터를 가공, 분석을 수행하는 실시간 분석의 구성요소이다.

표 5-11 실시간 분석 기능 설명

영역	구성 요소	기능 설명	고려사항
실시간 분석	수집 및 확인	<ul style="list-style-type: none"> <li>· 실시간 데이터 값을 입력 받고, 각 데이터에 해당되는 기준 정보를 파악함</li> <li>· 데이터 제공 원천과 사전 규약에 의하여 정의된 데이터 항목을 기준 정보로 메모리에 로딩</li> </ul>	<ul style="list-style-type: none"> <li>· 예측 분석에서 생성한 분석 모델을 PMML (: Predictive Model Markup Language) 형태로 실시간 분석 솔루션에 불러들여 실시간 발생하는 데이터에 적용</li> </ul>
	분류 및 계산	<ul style="list-style-type: none"> <li>· 수집되는 원천 데이터를 분석 목적에 맞게 처리하는 기능</li> <li>· 타임 윈도우 구간 내 평균, 최대, 최소 값들 계산하거나, 다른 데이터 항목들과의 산술식에 의한 계산</li> <li>· 필요 시, 데이터 정제 룰에 기반한 데이터 정제</li> </ul>	
	스코어링 및 결정	<ul style="list-style-type: none"> <li>· 수집 데이터에 대하여 분석을 수행하고 분석 결과를 제공하는 기능</li> <li>· 예측 및 패턴 분석 수행을 통한 입력 데이터 값에 기반한 분석 결과 스코어를 산출하는 기능</li> <li>· 입력 데이터 값의 변화에 대하여 정의된 임계치를 기준으로 임계치 내에 존재하는 지 벗어났는지에 대한 판단</li> </ul>	
	이벤트 실행	<ul style="list-style-type: none"> <li>· 분석 수행 결과를 연계 어플리케이션에 전달하거나 데이터베이스에 저장하는 기능</li> <li>· 예측 및 패턴 분석 수행 결과 또는 임계치를 벗어난 경우에 대한 로그를 데이터베이스에 저장</li> <li>· 실시간 발생 데이터 및 분석 결과를 어플리케이션 화면에 실시간으로 제공</li> </ul>	

표 5-12는 통계 및 기계학습 기법을 통해 패턴 분류 및 예측 분석 모델링 작업을 수행하는 예측 분석의 구성요소이다.

표 5-12 예측 분석 기능 설명

영역	구성요소	기능 설명	고려사항
예측 분석	데이터 수집 및 변환 처리	<ul style="list-style-type: none"> <li>· 분석 모델의 입력 데이터를 원천 시스템에 직접 접속하여 데이터 추출, 병합, 함수 적용을 통한 데이터 변환 수행</li> <li>· 분석의 정합성을 높이기 위한 신규 파생 데이터 생성 기능</li> </ul>	<ul style="list-style-type: none"> <li>· 실시간 분석 솔루션 또는 DB에서 분석 결과를 반영할 수 있도록 분석 모델을 PMML 형식으로 추출</li> <li>· 생성된 분석 모델을 전사적 차원에서 체계적으로 관리하기 위한 관리 도구 검토 필요</li> </ul>
	분석 모델링	<ul style="list-style-type: none"> <li>· 통계 모형 기반 및 기계 학습 기반의 분석 모델을 생성할 수 있는 기능</li> <li>· 통계 분석, 연관성 분석, 군집 분석, 분류 분석, 시계열 분석 등 다양한 유형의 분석을 수행할 수 있는 알고리즘 제공</li> </ul>	
	분석 결과 저장 및 출력	<ul style="list-style-type: none"> <li>· 분석 모델 적용 결과 값을 입력 값과 함께 데이터베이스 또는 파일로 저장하는 기능</li> <li>· 생성된 분석 모델을 XML 형식으로 출력하여 데이터베이스 분석 또는 실시간 분석 솔루션에 import 수행</li> </ul>	
	분석 결과 시각화	<ul style="list-style-type: none"> <li>· 버블 차트, 히트 맵, 산점도 등 다양한 시각화 차트를 제공함으로 분석가가 입력 데이터 및 결과 데이터를 효과적으로 분석</li> <li>· 분석 그래프 결과를 이미지 파일 형식으로 저장</li> </ul>	

표 5-13은 비즈니스 룰 또는 최적화 기법을 기반으로 최적의 대안을 도출하는 의사결정 분석의 구성요소이다.

표 5-13 의사결정 분석 기능 설명

영역	구성요소	기능 설명	고려사항
의사 결정 분석	비즈니스 룰 정의	<ul style="list-style-type: none"> <li>· 의사결정을 위해 필요한 업무 데이터를 수집하고 필터링 등의 조건을 통해 선택 기능</li> <li>· 업무 데이터 변수들과 모델링의 스코어링 결과를 대상으로 다양하게 조합하여 비즈니스 룰을 생성</li> </ul>	<ul style="list-style-type: none"> <li>· 비즈니스를 기반으로 시뮬레이션을 통한 최적화 방식과 수리적 모델을 통한 최적화 방식으로 구분됨</li> <li>· 비즈니스를 정의 기반 솔루션은 예측 분석 솔루션과 통합되는 것이 일반적</li> <li>· 비즈니스를 정의 모델을 운영 환경에서 실행하기 위한 모델 관리 기능 필요, 솔루션에 함께 통합되어 있기도 함</li> </ul>
	시물 레이션	<ul style="list-style-type: none"> <li>· 비즈니스 룰 정의에 따른 예상 결과 값을 모니터링</li> <li>· 조건식을 변화시켜 가면서 결과 값을 비교할 수 있는 시뮬레이션 기능</li> </ul>	
	최적화 모델링	<ul style="list-style-type: none"> <li>· 선형 계획법(Linear Programming), 혼합정수 프로그래밍(Mixed Integer Linear Programming), 이차 방정식 프로그래밍(Quadratic Programming) 등 최적화 문제를 풀기 위한 모델링 기능</li> </ul>	
	최적화 데이터 처리	<ul style="list-style-type: none"> <li>· 통합 개발 환경을 통해 최적화 수행을 위해 필요한 데이터 세트를 읽어 들이고 데이터 변환 및 결과 값을 데이터베이스 또는 파일로 저장</li> </ul>	

표 5-14는 사용자가 원하는 형식의 비정형 질의 및 다차원 분석을 수행하는 다차원 분석의 구성요소이다.

표 5-14 다차원 분석 기능 설명

영역	구성요소	기능 설명	고려사항
다차원 분석	정형 보고서	<ul style="list-style-type: none"> <li>· 일반 사용자가 조회하기 위한 정형 보고서 생성 및 조회</li> <li>· 필터링, 정렬, 컬럼 이동 등 기본적인 분석을 위한 기능</li> <li>· 차트, 그리드, 교차표 등 다양한 형식을 단일 보고서에서 조회</li> </ul>	
	비정형 질의	<ul style="list-style-type: none"> <li>· 사용자가 메타 정보를 이용하여 추출하고자 하는 데이터를 선택하고 조회할 수 있는 기능</li> <li>· 쿼리한 결과를 파일로 저장하는 기능</li> </ul>	
	대시보드	<ul style="list-style-type: none"> <li>· 시계열, 지도 등 다양한 유형의 시각화 표현을 통하여 전반적인 현황에 대하여 종합적으로 조회 및 분석</li> <li>· 사용자별로 개인화된 대시보드 구성</li> </ul>	
	분석 기능	<ul style="list-style-type: none"> <li>· 사용자가 다양한 관점에서 데이터를 조회할 수 있는 기능</li> <li>· 관점 변경, 드릴 업, 드릴 다운 등 인터랙티브한 분석 기능</li> <li>· 다양한 차트를 통하여 시각적으로 데이터를 표현하는 기능</li> </ul>	



표 5-15은 다양한 종류의 데이터 내 패턴 및 연관성을 발견하기 위해 인터랙티브 방식으로 탐색하는 디스커버리 분석의 구성요소입니다.

표 5-15 디스커버리 분석 기능 설명

영역	구성요소	기능 설명	고려사항
디스커버리 분석	패턴 매칭	<ul style="list-style-type: none"> <li>패턴 발견을 위하여 다양한 원천으로부터 데이터를 불러와서 조건에 맞게 조회할 수 있는 기능</li> <li>텍스트 데이터, 차트, 문서, 정형 데이터 등 다양한 특정 조건을 기준으로 조회할 수 있는 기능</li> </ul>	텍스트 분석 솔루션과 함께 연계되는 경우 효과적으로 구현 될 수 있음
	검색	<ul style="list-style-type: none"> <li>데이터 저장영역에서 원하는 결과를 추출하기 위한 검색 기능</li> <li>데이터 포맷, 위치에 관계없이 데이터 소스에 접근하여 검색하고 결과를 제공하는 기능</li> </ul>	

표 5-16은 자연어 처리 기술을 적용하여 텍스트 데이터 내 패턴 및 연관 분석 등을 수행하기 위한 텍스트 분석의 구성요소이다.

표 5-16 텍스트 분석 기능 설명

영역	구성요소	기능 설명	고려사항
텍스트 분석	자연어 처리	<ul style="list-style-type: none"> <li>사람이 사용하고 이해하는 언어를 컴퓨터가 이해할 수 있는 형태로 구조화하고 분류하는 기능</li> <li>형태소 분석을 기본으로 구문분석, 의미 분석을 수행하는 기능</li> </ul>	빅데이터 플랫폼으로 다양한 과제 수행을 위해서는 형태소 분석, 구문 분석, 의미 분석과 같은 자연어 처리 엔진을 기반으로 하는 텍스트 분석이 이루어져야 함
	인덱싱	<ul style="list-style-type: none"> <li>자연어 처리를 기반으로 텍스트 데이터를 분석이 이루어 질 수 있도록 단어들에 대하여 색인을 작성하는 기능</li> </ul>	

		<ul style="list-style-type: none"> <li>· 비구조화된 콘텐츠로부터 개념, 단어, 구분, 분류, 개체명을 추출할 수 있도록 어노테이션을 추가하는 기능</li> <li>· 텍스트 문서에 대한 분류, 작성자, 일자, 제목 등 메타 데이터를 추출하는 기능</li> </ul>	
	텍스트 마이닝	<ul style="list-style-type: none"> <li>· 텍스트 데이터에서 자연어 처리를 통하여 처리된 데이터를 대상으로 감성 분석, 연관 분석 등의 텍스트 기반의 분석을 수행하는 기능</li> <li>· 특정 키워드 기반의 문서를 검색할 수 있는 기능</li> <li>· 단어들 간의 빈도, 연관 분석 및 추세 분석 기능</li> <li>· 감성 분석 기능</li> </ul>	

#### 4. 빅데이터 분석영역의 특징

데이터 관점, 분석 관점, 비즈니스 사용자 관점, 분석 및 개발 관점에서 각각의 분석 영역에서의 특징은 표 5-17과 같다.

표 5-17 빅데이터 분석 영역의 특징

구분	데이터 관점	분석 관점	비즈니스 사용자 관점	분석 및 개발 관점
실시간 수집 데이터 분석	<ul style="list-style-type: none"> <li>정형스트리밍 데이터 입력 (IoT, 센서 등 지속적 발생 데이터)</li> <li>이력분석 위해 필요한 데이터를 저장 (파일 또는 DB 저장)</li> </ul>	<ul style="list-style-type: none"> <li>솔루션 내 로직 구현을 통하여 분석 실행</li> <li>예측 분석의 분석 모델을 import 하여 운영 환경에서 분석 실행</li> </ul>	<ul style="list-style-type: none"> <li>일반 사용자가 솔루션에 직접 접속하지는 않고 실시간 분석 결과 값을 과제별 어플리케이션을 통해 조회</li> </ul>	<ul style="list-style-type: none"> <li>실시간 분석 솔루션의 개발 환경에 대한 기술을 가지고 있는 개발자에 의해 개발</li> <li>분석 로직은 비즈니스 사용자가 제공</li> </ul>
대용량 데이터 기반 분석 모델 생성	<ul style="list-style-type: none"> <li>분석대상 유형은 정형 데이터</li> <li>예측 분석의 경우 하둡 기반의 플랫폼 저장/처리 영역 내 모델 실행 가능</li> </ul>	<ul style="list-style-type: none"> <li>분석 솔루션 클라이언트를 통해 데이터 분석 및 모델 생성 과정을 수행함</li> <li>생성된 모델을 운영 시점에 실행할 수 있도록 deploy 필요</li> <li>예측 분석 모델의 경우 실시간 솔루션 및 인 데이터베이스 실행 가능</li> </ul>	<ul style="list-style-type: none"> <li>일반 사용자는 분석 솔루션을 사용하지 않고 분석 결과를 어플리케이션을 통해 조회</li> <li>분석 기술이 있는 현업 사용자는 예측 분석 및 비즈니스를 관리 솔루션 사용</li> </ul>	<ul style="list-style-type: none"> <li>예측 분석, 최적화 분석에 대한 전문 지식을 가지는 분석가가 분석 모델을 생성</li> <li>룰 기반 의사결정의 경우 현업 사용자가 직접 개발 가능</li> </ul>
다차원 분석 및 텍스트 분석	<ul style="list-style-type: none"> <li>다차원 분석은 기본적으로 정형 데이터를 대상으로 분석</li> <li>사용자가 수행하는 텍스트 분석은 자연어 처리를 통하여 정형 데이터 형식으로 변환된 데이터들</li> </ul>	<ul style="list-style-type: none"> <li>솔루션 개발 단계에서 화면 구성 및 데이터 접근에 대한 구성은 이루어짐</li> <li>사용자는 데이터 선택 및 조회를 통하여 분석 실행</li> </ul>	<ul style="list-style-type: none"> <li>주 분석 주체는 일반 사용자</li> <li>일반 사용자가 분석 솔루션을 통해 개발된 화면에 접근하여 분석</li> </ul>	<ul style="list-style-type: none"> <li>일반 사용자가 분석을 수행하기 위해서는 다차원 분석 및 텍스트 분석 각각의 솔루션에 대한 기술을 가지고 있는 개발자에 의해</li> </ul>

	대상으로 분석, 필요 시 텍스트 원본을 함께 조회		석을 수행함	하여 사전 개발이 필요
디스커버리 분석	<ul style="list-style-type: none"> <li>· 분석대상 유형은 정형 데이터 및 비정형 데이터(텍스트 데이터)가 주 대상</li> <li>· 유의미한 데이터 관계가 발견되는 경우, 추가 분석을 위해 데이터 변환 등을 통해 분석 결과를 요약 영역에 저장</li> </ul>	<ul style="list-style-type: none"> <li>· 디스커버리 분석 솔루션에서 제공하는 기능을 활용하여 개발된 화면을 기반으로 사용자가 분석하고자 하는 데이터를 인터랙티브하게 조회하면서 분석 수행</li> </ul>	<ul style="list-style-type: none"> <li>· 일반 사용자가 분석 솔루션을 통해 개발된 화면에 직접 접속해서 분석 수행</li> </ul>	<ul style="list-style-type: none"> <li>· 디스커버리 분석 솔루션의 개발 환경에 대한 스킬을 가진 개발자가 개발</li> <li>· 비즈니스 사용자는 분석에 필요한 데이터에 대한 정보 제공</li> </ul>
텍스트 분석 (자연어 처리)	<ul style="list-style-type: none"> <li>· 분석대상 유형은 비정형 텍스트 데이터 (텍스트 파일, 웹 데이터 등)</li> <li>· 자연어 처리를 통하여 텍스트 비정형 데이터로 분석 가능한 정형 데이터를 생성하여 저장</li> </ul>	<ul style="list-style-type: none"> <li>· 텍스트 분석 솔루션 자체 자연어 처리 엔진을 통하여 비정형 데이터의 분석이 이루어짐</li> </ul>	<ul style="list-style-type: none"> <li>· 자연어 처리 단계에서는 비즈니스 사용자가 솔루션에 접근할 필요는 없음</li> </ul>	<ul style="list-style-type: none"> <li>· 자연어 처리를 위해 필요한 용어 사전, 처리 로직 등을 텍스트 분석 솔루션에 대한 스킬을 가진 개발자가 개발</li> <li>· 비즈니스 사용자는 필요 시 용어 및 로직에 대한 정보 제공</li> </ul>

그림 5-39는 데이터 흐름, 분석 결과 저장 중심의 데이터 및 분석 흐름도이다.

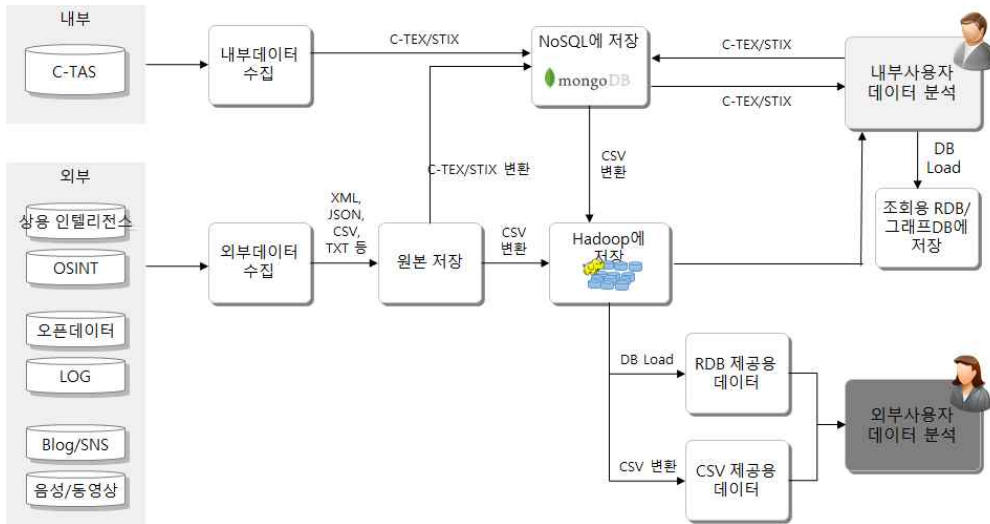


그림 5-39 데이터 및 분석 흐름도

## 5. 빅데이터 플랫폼 서비스

빅데이터 플랫폼은 각 분석과제에서 필요로 하는 데이터의 수집, 저장, 관리 기능과 분석 기능을 제공하고 활용플랫폼은 분석결과를 표현하여 정보사용자에게 전달한다.

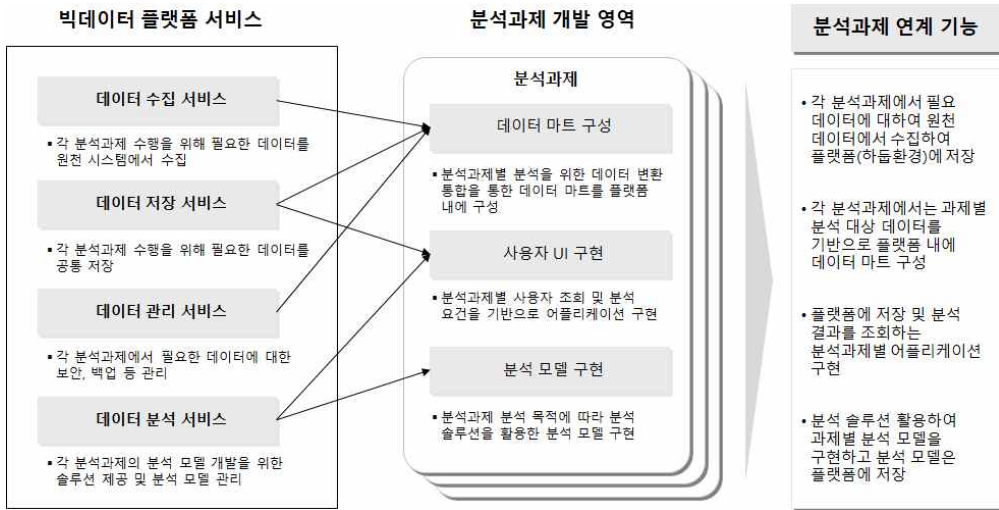


그림 5-40 빅데이터 플랫폼 서비스 영역

빅데이터 플랫폼 제공 서비스를 활용하여 각 분석과제에서는 자체 어플리케이션을 구현하여 사용자에게 분석 데이터를 제공할 수 있다.

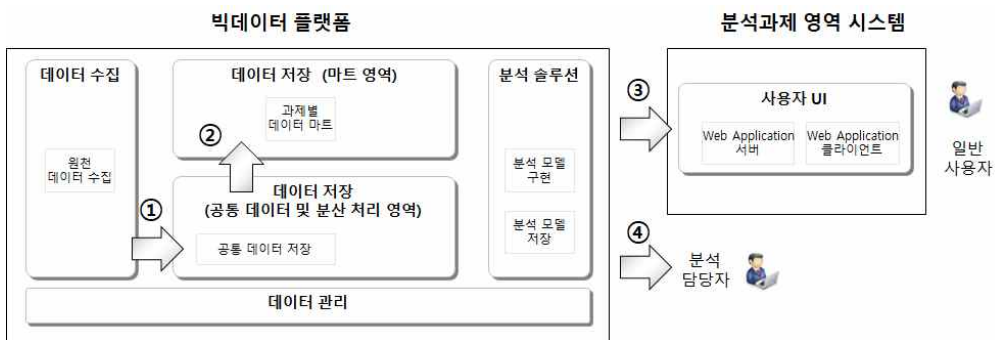


그림 5-41 빅데이터 플랫폼 업무절차

그림 5-37에서 빅데이터 플랫폼과 분석과제 영역 시스템에서의 업무를 확인해보면 아래와 같다.

- ① 각 분석과제에서 필요 데이터에 대하여 원천 데이터에서 수집하여 플랫폼(하둡환경)에 저장
- ② 각 분석과제에서는 과제별 분석 대상 데이터를 기반으로 플랫폼 내에 데이터 마트 구성
- ③ 플랫폼에 저장 및 분석 결과를 조회하는 분석과제별 어플리케이션 구현
- ④ 분석 솔루션 활용하여 과제별 분석 모델을 구현하고 분석 모델은 플랫폼에 저장

## 6. 빅데이터 분석 솔루션 도입

빅데이터 분석 솔루션 도입을 위해서는 분석 편의성, 빅데이터 처리 역량, 전사 차원의 기능, 에코 시스템 협업 체계, 기술지원 및 시스템 유지보수에 대한 고려가 필요하다.

표 5-18 빅데이터 분석 상용 솔루션 특징

구분	상용 솔루션	오픈소스
분석 편의성	<ul style="list-style-type: none"> <li>· 일반 사용자도 분석 쉽게 접근할 수 있는 GUI 기반의 사용 환경 제공</li> <li>· 고급 사용자를 위한 코딩 및 스크립트 방식 지원</li> <li>· 솔루션 내에 다양하고 수려한 시각화 지원</li> </ul>	<ul style="list-style-type: none"> <li>· 일반 분석 사용자 보다는 개발 기술이 있는 사용자들을 대상으로 개발되어 주로 Code Programming 방식임</li> <li>· 분석/업무 지식 이외에 추가적으로 데이터 처리 메커니즘에 대한 이해가 요구</li> <li>· 수려한 시각화를 위해서는 별도 개발 필요</li> </ul>
빅데이터 처리 역량	<ul style="list-style-type: none"> <li>· Hadoop pushback 등의 기능을 통해 빅데이터 처리 가능하며 다양한 프로젝트 사례를 통하여 빅데이터 처리 역량이 검증되어 있음</li> </ul>	<ul style="list-style-type: none"> <li>· 데이터를 내장메모리로 이동하여 분석하는 구조로 빅데이터 처리가 불가능하며 샘플링 기반으로 접근</li> </ul>
엔터프라이즈 레벨 특화 기능	<ul style="list-style-type: none"> <li>· 분석 결과의 자산화(협업/공유/배포/버전관리 등)</li> <li>· 분석 결과 성능 평가, 사용자 권한 관리 및 보안</li> <li>· 예측 모델의 경우 모니터링 및 자동 예측 성능 보정</li> </ul>	<ul style="list-style-type: none"> <li>· 단일 사용자 또는 연구자를 위한 분석 환경 제공을 목적으로 개발되어 엔터프라이즈 레벨을 위한 관리 기능 없음</li> </ul>
에코 시스템 협업 체계	<ul style="list-style-type: none"> <li>· 벤더에서 요구하는 자격요건을 충족하는 검증된 협력업체를 통해 서비스 제공하여 안정적 프로젝트 수행</li> </ul>	<ul style="list-style-type: none"> <li>· 전문 취급 업체가 없거나, 있더라도 엔지니어들의 수준차가 많이 나 안정적인 프로젝트 수행이 어려움</li> </ul>
교육 훈련 및 기술 지원	<ul style="list-style-type: none"> <li>· 일반적으로 벤더에서 운영하는 공식 지원 채널 존재</li> <li>· 상시 대응 가능한 채널 확보 및 문제 발생 시 즉시 대응</li> </ul>	<ul style="list-style-type: none"> <li>· 오픈소스 관련 커뮤니티에 질의-응답을 통해서만 문제 해결</li> <li>· 체계적 지원에 대한 보장이 불가능함</li> <li>· 시스템 장애 발생 시 책임 소재가 불분명함</li> </ul>
시스템 유지보수	<ul style="list-style-type: none"> <li>· 자동모델 관리 기능을 제공하여 투입 인력을 최소화하여 모델 성능 관리 가능</li> <li>· 업그레이드나 픽스 적용 시 벤더에서 하위 호환성 보장, 별도 영향도 분석 없이 최소화된 노력으로 시스템 유지보수 가능</li> </ul>	<ul style="list-style-type: none"> <li>· 모델 유지 보수 위해 지속적으로 고급 모델링 인력 투입 필요</li> <li>· 업그레이드나 픽스 적용 시 영향도 분석 및 적용 후 안정화에 지속적인 인력 투입 필요</li> </ul>



## 7. 빅데이터 분석 영역별 주요 솔루션

빅데이터 분석을 위한 각 분석 영역별 오픈소스 및 주요 상용 솔루션은 표 5-19와 같다.

표 5-19 빅데이터 분석 영역별 주요 솔루션 특징

영역	구분	솔루션	벤더 또는 조직
다차원 분석 (Analysis & Reporting)	오픈 소스	Pentaho Business Analytics	Pentaho
	상용 솔루션	Business Objects BI	SAP
	상용 솔루션	Cognos BI	IBM
	상용 솔루션	SAS Visual Analytics	SAS
	상용 솔루션	Tableau Desktop	Tableau
예측 분석 (Predictive Analytics)	오픈 소스	Mahout	APACHE
	오픈 소스	R	R Foundation
	오픈 소스	SPARK Mlib	APACHE
	상용 솔루션	SAS Enterprise Miner	SAS
	상용 솔루션	SPSS Modeler	IBM
의사결정 분석 (Decision Management)	상용 솔루션	Oracle Real Time Decision	Oracle
	상용 솔루션	SAS Decision Manager	SAS
	상용 솔루션	SPSS Analytical Decision Management	IBM
의사결정 분석 - 최적화 전문 (Optimization Analytics)	오픈 소스	GNU Linear Programming Kit	GNU
	오픈 소스	LP Solve	GNU
	상용 솔루션	ILOG Optimization	IBM
	상용 솔루션	Xpress Optimization Suite	FICO
실시간 분석 (Real Time Analytics)	오픈 소스	SPARK Streaming	Apache
	오픈 소스	Storm	Apache
	상용 솔루션	DeltaStream	Datastreams
	상용 솔루션	InfoSphere Streams	IBM
디스커버리 분석 (Discovery Analytics)	오픈 소스	ElasticSearch	Elastic
	오픈 소스	QlikView	QlickTech
	상용 솔루션	Endeca Information Discovery	Oracle
	상용 솔루션	Watson Explorer	IBM
텍스트 분석 (Text Analytics)	상용 솔루션	KONAN Text Analytics	코난
	상용 솔루션	SAS Text Miner	SAS
	상용 솔루션	Watson Content Analytics	IBM
	상용 솔루션	WISE TEA	와이즈넷
플랫폼 (Platform)	오픈 소스	KIME Analytics Platform	KNIME
	상용 솔루션	IBM PMQ	PMQ

## 제 6장 결론 및 제언

본 연구과제에서는 전파감시 빅데이터에 대한 다양한 가치를 인식하고, 효율적인 전파감시 빅데이터의 분석을 통해 향후 전파정책의 추진방향을 올바르게 이해하고 적용할 수 있도록 가이드를 제시하였다.

현재까지 중앙전파관리소에서 많은 양의 전파감시 데이터가 생성되고는 있으나 주파수이용현황 조사 데이터를 제외하면 제대로 활용되지 못하는 실정으로, 향후 빅데이터 분석시스템 구축을 통해 전파감시 빅데이터에 대한 활용도 제고가 필요할 것으로 판단된다.

중앙전파관리소가 본 연구과제를 바탕으로 빅데이터 전문가를 양성하고, 빅데이터 분석 플랫폼을 구축하여 적극 활용한다면, 4차 산업혁명시대 전파환경이 급속도로 변화하는 상황에서 주도적인 역할을 해 나아갈 수 있을 것이다.

### ◇ 전파감시데이터 활용 절차

- 전파 분야 빅데이터 분석과제의 추진 담당자가 추진절차를 실제 업무에 적용하기 위해서 점검해야 할 “빅데이터 사업 추진 단계별 점검 절차”를 제시하였다.

### ◇ 전파감시데이터 현황분석

- 전파감시 분석과제 선정을 위해 전파감시업무 현황을 파악하고, 전파감시 데이터 현황 및 업무에 따른 DB현황 상세 테이블을 분석하였다.
- 또한, 전파감시 데이터 활용방안을 조사하고, 선정된 분석과제의 연관 데이터를 정리하여 활용 가능한 데이터 목록을 도출하고 내부데이터 현황을 진단하였다.

### ◇ 분석모델 시각화 제안

- 실시간 저장되는 대량의 데이터를 모두 해석하기에는 많은 제약이 따르기 때문에 데이터를 선별적으로 선택해서 짧은 시간에, 적은 양의 자료를, 해석하기 쉬운 형태로 시각화하는 것이 가장 적합한 방법이라고 할 수 있다. 이에 시각화 도구 및 기술선정 능력을 고려하여 데이터 시각화 절차에 맞춰서 전과감시 데이터를 시각화하는 방법을 고려해야 한다.
- 분석과제 시각화는 우선적으로 빅데이터 플랫폼의 다차원 분석 솔루션에서 제공하는 시각화 유형을 사용하고, JavaScript 기반의 무료 차트 라이브러리를 활용하되 추가적 요건이 있는 경우 D3 등의 개발 도구를 활용할 수 있다.
- 과제별 시각화 도구 적용은 빅데이터 플랫폼의 다차원 분석 솔루션에서 제공하는 시각적 표현 기법을 우선 사용하고, Google Chart, HIGHCHARTS, Fusion Chart, Keylines 등 JavaScript 기반의 시각화 오픈 소스를 활용할 수 있다. 기존 솔루션에서 제공되지 않는 사용자에게 특화된 표현이 필요한 경우 D3을 활용하여 구현할 수도 있으며,
- 전과 혼·간섭 예방 분석모델의 시각화 방법은 지도 시각화 유형에 통계적 분석 시각화, 빈도 시각화, 시간 시각화 방안이 가장 효과적일 것으로 판단된다.

#### ◇ 데이터 분석가 양성 필요성 및 교육 방안

- 중앙전과관리소의 업무에 빅데이터를 적용 시, 각 분야별로 적용하는 경우에는 자칫 예산만 낭비하고 실패를 낼 가능성이 높다. 그러므로, 우선 빅데이터 적용이 필요한 분야를 도출하여 이를 통합한 빅데이터 플랫폼의 구축과, 분석 목적에 해당하는 데이터에 관한 정비작업과 통합된 빅데이터 관리 및 분석을 할 수 있는 능력을 갖춘 독자적인 조직(빅데이터센터(가칭))이 필요하다. 그러나 현재 여건상 이러한 독자적인 조직을 구축하기

에는 예산상의 한계가 있을 수도 있다.

- 따라서 단기적으로는 빅데이터 전문가를 포함한 전담팀 혹은 전담인력을 포함하는 조직의 구성이 필요하고, 장기적으로 해당 조직을 정비, 보완해 나아가야 할 것이다.
- 빅데이터 추진 조직은 내부조직 간의 경계가 모호한 업무 또는 긴밀한 협력체계가 요구되는 업무를 고려하여 기획관리, 분석관리, 운영관리, 공공데이터 관리의 4개 기능을 수행하는 新 빅데이터 사업팀으로 구성하고, 빅데이터 조직 및 인력의 역량을 감안하여 데이터의 통합관리에 적합한 하이브리드형 조직으로 구성하는 방안이 효과적이다.
- 조직 구성 방안의 소요인력 도출에 있어서는 소요인력 산정방법에 따라 업무 기능별 적정 인력 추정치를 도출하였고, 향후 빅데이터 분석시스템의 규모에 따라 소요인력은 변동될 수 있다.
- 빅데이터 분석 조직의 주요업무는 내부 및 외부데이터 조사, 빅데이터 분석시스템 구축 및 운영관리, 빅데이터 분석 및 활용서비스 개발, 데이터 개방 및 활성화이며, 주요업무에 맞춰 적정 소요 인력을 산정하였고, 중앙전파관리소의 부서별 인사이동을 대비한 데이터 인력 양성방안으로는 분석인력과 운영인력 양성방안으로 구분해서 양성하는 방안을 제안하였다.

#### ◇ 전파감시 빅데이터 시스템 구축 방안

- 빅데이터 인프라 구축은 빅데이터 분석과제를 수행하기 위한 플랫폼으로 필요한 하드웨어 및 소프트웨어를 설치하고 구성해야 한다. 빅데이터 플랫폼은 데이터 수집, 데이터 저장 및 처리, 데이터 분석, 데이터활용, 데이터 보안 및 관리 영역으로 구성되고, 수행 기간은 주관부서 주도로 5-6개월 예상된다.
- 빅데이터 인프라 구축 추진 주용 내용은 ① 빅데이터 플랫폼 인프라(하드웨어 및 소프트웨어) 도입, ② 인프라 설치 및 구

성, ③ 구성 테스트 수행이 주요 내용으로 소요예산은 내·외부 분석에 따른 컨설팅 절차가 필요하다.

- 빅데이터 플랫폼 운영을 위해서는 빅데이터 및 인프라 운영을 위한 각각의 담당자 지정이 필요하다. 빅데이터 운영 담당자는 데이터 수집 및 DBA 역할을 하고, 인프라 운영 담당자는 플랫폼내의 시스템 운영을 담당하도록 지정되어야 하고, 데이터 보안 및 작업관리와 데이터 마트 백업 방안도 신중히 고려해야 한다.
- 빅데이터 분석 플랫폼은 다양한 유형의 데이터를 활용할 수 있고, 데이터 활용 이전에 요구되던 시간을 줄이고 실시간으로 발생하는 데이터에 대해서도 분석할 수 있도록 구성되어야 한다.
- 전파감시 데이터의 특성 상 데이터 내 패턴 및 연관성을 분석하기 위해 인터랙티브 방식으로 탐색하는 디스커버리 분석 기술, 통계 및 마이닝 기법을 통한 모델링 작업, 패턴 분류 및 예측을 위한 예측 분석 기술을 권장한다.
- 향후에는 전파감시 과정에서 생성된 음성데이터를 텍스트로 변환 후 자연어 처리 기술을 적용하여 비정형 텍스트 데이터에서 특성을 추출하고 정형화하여 텍스트에 담긴 콘텐츠를 분석하기 위한 텍스트 분석 기술의 적용도 필요하다.

## 참고자료

- 서진완. (2004). 전자지방정부 구현을 위한 모델개발과 적용. 교수 논총.
- 양순애. (2013). 지역정보화의 효율적인 추진을 위한 IT 거버넌스 발전방안. 한국지역정보화학회지. 16(1) : 77-96.
- 양혜영. (2012). 빅데이터를 활용한 기술기획 방법론. Korea Institute of Science & Technology Evaluation and Planning.
- 윤상오. (2013). 빅데이터의 위험유형 분류에 관한 연구. 한국지역정보화학회지. 16(2) : 93-122.
- 정보통신산업진흥원. (2011). 빅데이터 확산에 따른 도전과 기회. 최신 IT동향.
- 조영임. (2012). 최신인공지능시스템. 홍릉과학출판사.
- 한국정보화진흥원. (2010). 모바일 시대를 넘어 AI 시대로. IT&Future Strategy.
- 한국정보화진흥원. (2011). 신가치창출 엔진, 빅데이터의 새로운 가능성과 대응전략. IT&Future Strategy.
- 한국정보화진흥원. (2012a). 빅데이터로 진화하는 세상. 빅데이터 연구센터. 2012.
- 한국정보화진흥원. (2012b). 빅데이터 시대 : 효과적인 공공정보 개방을 위한 데이터 플랫폼 구축
- 방향. IT&Future Strategy. 2012.
- 한국정보화진흥원. (2012c). 빅데이터 시대. AI의 새로운 의미와 가치. IT&Future Strategy. 2012.
- 한국정보화진흥원. (2012d). 빅데이터 시대의 인재, 데이터 사이언티스트의 역할과 가능성. IT&Future Strategy. 2012.
- 한국정보화진흥원. (2012e). 2012 국가정보화백서, 빅데이터 시대의 국가발전 전략. 2012.
- 한국정보화진흥원. (2013). 새로운 미래를 여는 빅데이터 시대.

- 행정안전부. (2012). 주요정책 과제 추진실적 보고서.
- Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze, (2009), "An Introduction to Information Retrieval", Cambridge University Press, England.
- 서울시청(2013. 9. 4.), “심야전용 ‘올빼미버스’ 9개 노선출발!”(<http://traffic.seoul.go.kr/archives/13102>).
- 조완섭(2012. 7. 11.), 「Big Data 시대 - 데이터 분석인력의 가치」, 13th Conference on SW Quality Insight.
- 한국정보화진흥원(2012. 8. 8.), “빅데이터 시대의 인재, 데이터 사이언티스트의 역할과 가능성”, 「IT & Future Strategy」, 제8호.
- 김경원, 빅데이터를 활용한 경찰의 범죄예측 활성화 방안, 동국대학교, 2015.2.
- 김성태, 성공적인 빅데이터 활용을 위한 3대 요소 : 자원, 기술, 인력, 한국정보화진흥원, 2012.4.12.
- 김성태, 빅데이터 시대의 데이터 자원 확보와 품질 관리 방안, 한국정보화진흥원, 2012.5.7.
- 노규성, 사례분석을 통한 지방행정의 빅데이터 활용 전략, Journal of Digital Convergence, 12(1): 88-97, 2014.
- 박대현 · 송동현, 비정형 데이터 활성화의 정치, 경제, 문화적 함의, INTERNET & SECURITY FOCUS, 2014.2.
- 박봉원, 21세기 연금술사 ‘데이터 사이언티스트’ 그리고 강원도, 강원발전연구원, 2014.5.6.
- 박봉원, 공공데이터 개방과 강원도, 강원발전연구원, 2015.5.8.
- 박시룡 · 안순권 · 정선화 · 이효정, 빅데이터를 활용한 과학적·효율적 행정 구현 방안, 서울경제 백상경제연구원, 2013.12.
- 변화수 · 양석모 · 배현아, 빅데이터 분야 해외 선진사례 연구, 안전행정부 정부통합전산센터, 2013.11.
- 백성진, 미국의 IT 산업 전망 및 정책 동향, 한국산업기술진흥원, 2014.3.14.

- 배동민 · 박현수 · 오기환, 빅데이터 동향 및 정책 시사점, 정보통신연구원, 2013.6.1.
- 빅데이터 국가전략 포럼, 2013 빅데이터 국내 사례집, 2013.
- 신영진, 공공분야의 빅데이터 추진과 개인정보보호에 관한 연구: 개인정보 보호법의 주요 내용과 개선과제를 중심으로, Internet and Information Security, 제3권, 제2호, pp.90-105, 2012.
- 신신애 · 김성현 · 송경빈 · 류승희 · 정규진 · 솔리라, 창조 경제 실현을 위한 빅데이터국내 사례집, 한국정보문화진흥원 빅데이터 분석활용센터, 2014.03.
- 엄희경 · 최두진 · 박성찬 · 장휴정, 도시브랜딩을 위한 빅데이터 활용에 관한 연구, 한국브랜드디자인학회, 13(3):193-206, 2015.
- 유영성, 지자체의 공공 빅데이터 정책 사례 연구, 경기개발연구원, 2014.12.
- 이유택, 빅데이터 시대의 개인 데이터 보호와 활용, IT & Future Strategy, 2013.6.21.
- 이경원 · 이한일 · 김근형, 입법과정에서의 빅데이터 분석의 필요성 연구, 법제처, 2013.
- 이정호, 빅데이터를 활용한 대한지적공사 사업효율 증대 및 이미지 제고 방안, 인천광역시, 2013.11.
- 이재학 · 김병곤 · 박정구 · 이민호 · 이정훈 · 정용규 · 조중환 · 최재섭 · 허 준, 빅데이터 환경하의 ICT 데이터 유통체계 진단 및 생태계 조성방안, 남서울대학교 산학협력단, 2013.11.
- 이이삭, 빅데이터 지식재산권 문제에 관한 연구, 성균관대학교 석사 학위논문, 2015.
- 임상규, 빅데이터를 활용한 스마트 재난관리 전략, 2014
- 이원규, 빅데이터(Big Data)를 활용한 부산시 도시관리 전략, 부산발전연구원, 2014.4
- 정지선, 성공적인 빅데이터 활용을 위한 3대 요소 : 자원, 기술, 인력, IT & Future Strategy, 2012.4.12.



- 장병열 · 김영돈, 빅데이터 기반 융합 서비스산업 창출 방안, 과학기술정책연구원, 2013.6.14
- 정지선, 새로운 미래를 여는 빅데이터 시대, 한국정보화진흥원, 2011.12.30
- 조완섭, 빅데이터시대 과학기술정책방향, 한국과학기술정책연구원, 2013
- 정광렬, 문화예술 분야 빅데이터 활용을 위한 기초 연구, 한국문화관광연구원, 2014
- 최창우 · 윤종인 · 오병철 · 조희진 · 전표진 · 신동민, 빅데이터 활용 사례 분석을 통한 민 · 관 융합 빅데이터 활용, 한국지역정보개발원, 2014.12.30.
- 한국IT학회 컨소시엄,(2015), 공단 빅데이터 활용방안 연구 용역.
- 홍릉과학출판사(2014년 초 번역출판예정), 「빅데이터 거버넌스」 (원서: Big Data Governance - An Emerging Imperative, MC Press, 2012).
- CIO(2012. 12. 3.), “데이터과학자 육성, 채용 ‘5가지 조건’”(http://www.ciokorea.com/news/14901).
- IDG Tech Report(2012. 8. 23.), “데이터 속에서 가치를 창출하는 데이터과학자”(http://www.itworld.co.kr/techlibrary/77427).
- IT World(2013. 4. 12.), “데이터과학자가 되고 싶다면”(http://www.itworld.co.kr/news/81316).
- CNN Money(2012. 11), Best new jobs in America(http://money.cnn.com/gallery/pf/2012/11/01/best-new-jobs-in-america/3.html).
- Conway, D.(2010), The data science venn diagram, Dataists(http://www.dataists.com/2010/09/the-data-science-venn-diagram/database).
- Datanami Staff(2012. 4. 10.), Six Big Name Schools with Big Data

Programdatanami([http://www.datanami.com/datanami/2012-04-10/six\\_big\\_name\\_schools\\_with\\_big\\_data\\_programs.html?page=2](http://www.datanami.com/datanami/2012-04-10/six_big_name_schools_with_big_data_programs.html?page=2)).

- EMC, Data Science and Big Data Analytics(<http://education.EMC.com/DataScience>).
- McKinsey(2011. 5), Big Data: The next frontier for innovation, competition, and productivity.
- Michael Rappa(2012. 6), Master of Science in Analytics, Goals, Learning, Outcomes, North Carolina State University, Institute for Advanced Analytics Report.
- Mike Loukides(2010. 6), What is Data Science?, An O'Reilly Radar Report.
- NASCIO(2008. 4), "Data Governance - Managing Information As An Enterprise Asset: Part I - An Introduction", NASCIO Governance Series([www.nascio.org](http://www.nascio.org)).
- Steve ALaValle, et al.(2011), "Big Data, Analytics and the Path From Insights to Value", MIT Sloan Management Review.
- Thomas H. Davenport and D.J. Patil(2012. 11), "Data Scientist: The Sexiest Job of the 21<sup>st</sup> Century", Harvard Business Review.
- Tom Groenfeldt(2013. 5. 18.), Big Data and Data Scientists - It's An Issue Of Degree(s), Forbes.
- William Finzer, The Data Science Education Dilemma, Report(supported by National Science