



4차 산업혁명시대의 빅데이터 분석시스템 구축과 활용

2018. 12. 07 (금)

김정중
(주)이글로벌시스템

✓ 빅데이터분석시스템이 갖는 의미는?

데이터
과학

정형, 비정형 형태를 포함한 다양한 데이터로부터 지식과 인사이트를 추출하는데 과학적 방법론, 프로세스, 알고리즘, 시스템을 동원하는 융합 분야

빅데이터

대용량의 데이터로부터 가치를 추출하고 결과를 분석하는 기술

데이터
분석

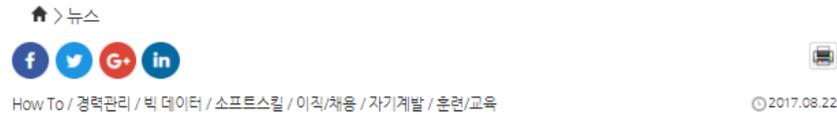
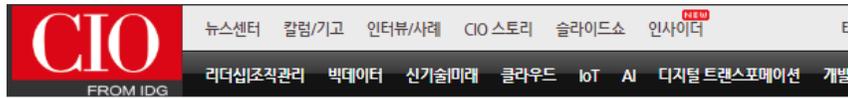
알고리즘과 수학적 처리 과정을 적용하여 데이터를 다루는 활동

각 구성요소들이 상호작용하거나 상호 의존하여 복잡하게 얽힌 하나의 집합체

시스템



✓ 데이터과학자에게 필요한 자질은 무엇인가?

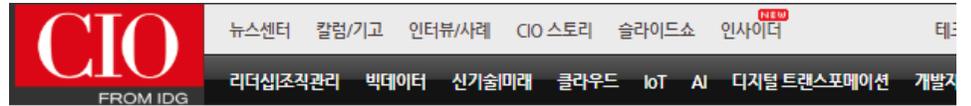


How To / 경력관리 / 빅 데이터 / 소프트스킬 / 이직/채용 / 자기계발 / 훈련/교육 © 2017.08.22

"최고의 직업 1위"...데이터 과학자란? 역할은? 되는 방법은?

Sarah K. White | CIO

데이터 과학자가 되는 경로는 업종에 따라 다르다. 그럼에도 불구하고 공통적으로 요구되는 역량과 경력, 학위 등이 있다. 여기 데이터 과학자 커리어를 시작하는데 필요한 정보를 정리했다.

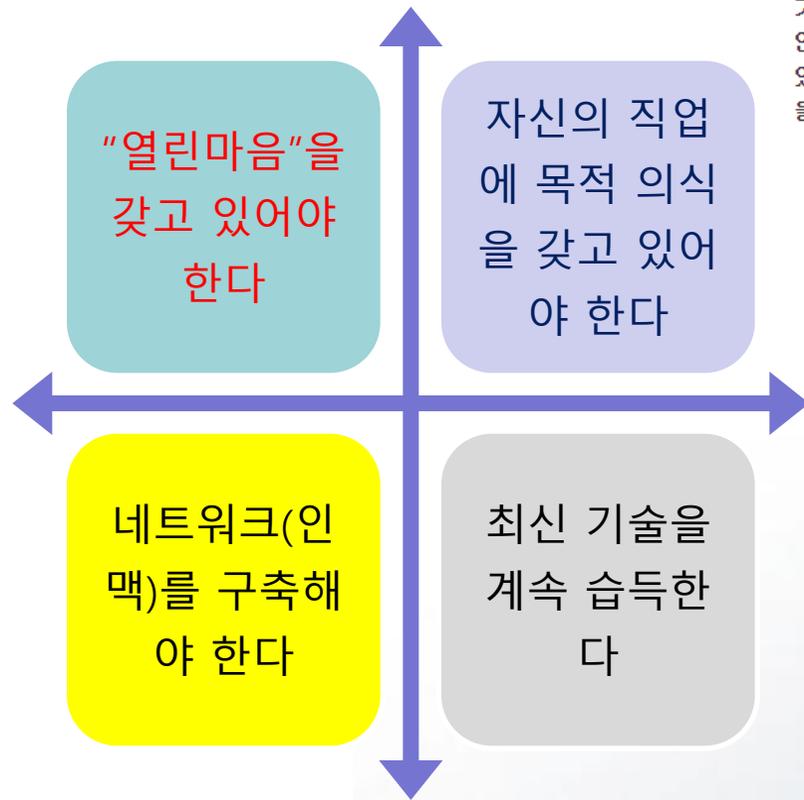


How To / 경력관리 / 빅 데이터 / 자기계발 © 2015.10.12

데이터 과학자에게 필요한 4가지 자질

Sarah K. White | CIO

기업들이 고객에 대해 온갖 정보를 무수히 수집하고 있다. 동시에 최근 발생했던 일련의 보안 사고에서 개인 정보가 대거 유출된 것이 알려지면서 개인 데이터에 대한 관심이 고조되고 있다. 기업들은 수 많은 정보를 수집할 수 있다. 그러나 이들 데이터를 보호하는 것에도 만전을 기해야 한다. 기업들은 또 '정보 수집'과 '분류 및 분석'은 별개라는 사실을 깨달고 있다.





목차

1. 빅데이터 및 데이터 분석 트렌드
2. 빅데이터 분석업무 이해
3. 빅데이터 분석시스템 사례
4. 분석업무를 위한 지식과 기술



1. 빅데이터 및 데이터 분석 트렌드

BIG DATA & AI LANDSCAPE 2018



빅데이터는 데이터만을 가리키지 않으며, 기술을 의미하지 않으며, 끊임없이 문제해결을 추구하는데 그 의의가 있다.

빅데이터 구성

인프라스트럭처

- 분석
- 데이터분석 플랫폼
 - 데이터 사이언스 플랫폼
 - BI 플랫폼
 - 시각화
 - 머신 러닝
 - 스피치 & NLP
 - 검색
 - 로그분석
 - 소셜분석
 - 웹/모바일/커머스 분석

애플리케이션

- Enterprise
- Industry

오픈소스

데이터 소스 및 API

데이터 리소스

Final 2018 version, updated 07/15/2018

© Matt Turck (@mattturck), Demi Obayomi (@demi_obayomi), & FirstMark (@firstmarkcap) mattturck.com/bigdata2018



> 1. 빅데이터 및 데이터 분석 트렌드

4차 산업혁명의 태동

구분	1차 산업혁명	2차 산업혁명	3차 산업혁명	4차 산업혁명
시기	18세기	19세기	20세기 후반	현재
특징	기계적 생산 및 증기기관	대량생산 및 전기에너지	컴퓨터와 인터넷 기반	인공지능, 빅데이터 분석, IoT 기반

- 2016년 다보스 포럼을 통해 전 세계 확산

- 2016년 1월 스위스 다보스에서 개최된 제 46차 세계경제포럼(WEF), "제 4차 산업혁명의 이해"라는 주제로 논의, 세계적으로 주목을 받기 시작
- 4차 산업혁명은 다양한 산업이 초연결성(Hyper-Connected), 초지능화(Hyper Intelligent) 기반 하에 상호 연결된 지능화 사회로 변화시킴
- **디지털, 바이오, 물리학 등 영역이 ICT 기술과 융·복합된 새로운 기술로 인류 삶의 방식 변혁 전망**

- 4차 산업혁명 시대, 4가지 융합 기술이 핵심

- 사물과 ICT의 융합: 사물인터넷 (IoT) ex) 센서기술, 스마트 홈 서비스, 공장자동화, 3D 프린팅
- 사람과 ICT의 융합: 사람과 장치를 연결하는 인터페이스 ex) 인터넷과 연결된 센서 부착 의류와 안경 등
- 운송수단과 ICT의 융합: 자율주행차 ex) 구글 자율주행차 등
- AI(인공지능)와 빅데이터의 융합: 딥러닝 ex) IBM "왓슨" 등, 빅데이터 정보를 AI가 스스로 분석해 스마트한 의사 결정

출처: 글로벌마켓레포트, "4차 산업혁명시대, 첨단제품 개발 트렌드와 시사점", 코트라, 2017

> 1. 빅데이터 및 데이터 분석 트렌드

빅데이터 특징과 활용

빅데이터 특징

1. 용량(Volume)

- 특성: 대량의 산재되어 있는 데이터
- 제약: 데이터 관점에서 관리 부재
- 방향: 분산되어있는 대량 데이터 분석 통한 가치 제공

2. 다양성(Variety)

- 특성: 분산된 데이터 중 비정형 80% 이상 차지
- 제약: 특정 분산 병렬 처리 방법 제공
- 방향: **다양한 비정형 데이터** 처리 방법 제공

3. 속도 (Velocity)

- 특성: 데이터 생성 후 유통에서 활용까지 수분 소요
- 제약: 일괄처리(실시간 처리 미지원)
- 방향: **실시간 데이터 조회와 처리**

4. 정확성 (Veracity)

- 특성: 흩어져 있는 방대한 데이터를 분석
- 제약: 분석의 정확도 향상에 노력
- 방향: 다양한 분석기법 및 시뮬레이션 처리

빅데이터 활용

활용분야	설명
미래예측	- 실시간 대용량 데이터 분석으로 정확도 높은 실시간 예측 - (예) 미국 포드사: 커넥티드 자동차 주행시스템에 운전자의 주행이력과 패턴을 분석하여 최적, 최단 주행경로 제안
숨은 필요 발견	- 대규모 데이터로부터 패턴을 발견해 숨은 필요사항 발견
리스크 경감	- 정량적 정보를 대폭 확보하여 위험의 감소와 정확도 증대 - (예) VISA: 카드 부정검지모델에서 부정패턴의 갱신을 매일 수회 실시하여, 정밀도를 향상
맞춤형 서비스	- 고객 개인별로 차별화해 유용한 정보 제공
실시간 대응	- 실시간 모니터링과 즉각적인 대응이 가능 - (예) 일본 Nishitetsu Store: 매일 단일상품별 원가율과 원가변동 추리를 분석하여, 이익률이 높은 상품에 대한 일자별 주력 마케팅 정책수립

가치
(Value)

의미 없는 데이터에 대해
문제 핵심과 가치를 정확히 분석하고,
새로운 관계를 발견 하도록 제공

> 1. 빅데이터 및 데이터 분석 트렌드

가트너 선정 10대 전략기술 트렌드 (Y 2018 ~ 2019)

지능

- ✓ 인공지능 강화 시스템 (AI Foundations)
- ✓ 지능형 앱 • 분석 (Intelligent Apps and Analytics)
- ✓ 지능형 사물 (Intelligent Things)

디지털

- ✓ 디지털 트윈 (Digital Twin)
- ✓ 클라우드에서 에지로 (Cloud to the Edge)
- ✓ 대화형 플랫폼 (Conversation Platforms)
- ✓ 몰입 경험 (Immersive Experience)

메시

- ✓ 블록체인 (Blockchain)
- ✓ 이벤트 기반 모델 (Event-Driven)
- ✓ 지속적이고 적응할 수 있는 리스크 및 신뢰평가 (CARTA) (Continuous Adaptive Risk and Trust)

가트너 선정
2018년 10대
전략기술 트렌드

✓ 지능형 (Intelligent)

✓ 분석 (Analytics)

용어에 주목하기

가트너 선정
2019년 10대
전략기술
트렌드

지능

- ✓ 자율주행 사물 (Autonomous Things)
- ✓ 확장된 분석기술 (Augmented Analytics)
- ✓ 인공지능 주도 개발 (AI-Driven Development)

디지털

- ✓ 디지털 쌍둥이 (Digital Twin)
- ✓ 자율권을 가진 엣지 (Empowered Edge)
- ✓ 몰입 경험 (Immersive Experience)

메시

- ✓ 블록체인 (Blockchain)
- ✓ 스마트 공간 (Smart Spaces)

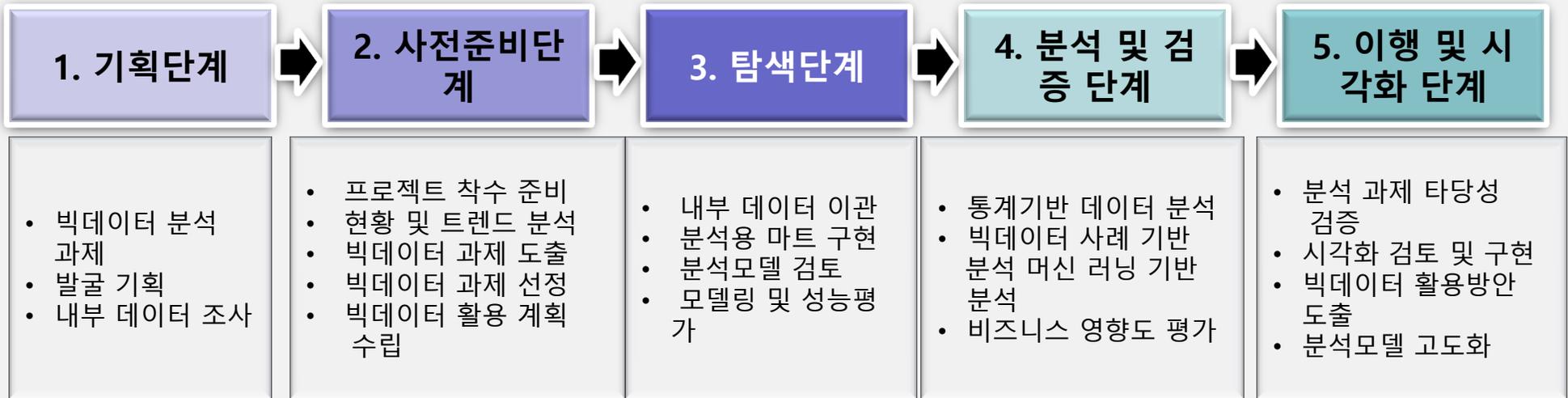
✓ 디지털 윤리와 개인정보보호 (Privacy and Ethics)

✓ 양자컴퓨팅 (Quantum Computing)

출처: Gartner(2016~19, "Gartner Top 10 Strategic Technology")

> 2. 빅데이터 분석업무 이해

빅데이터 분석업무 단계별 세부업무



빅데이터 과제 이행 시 역할

현업 부서 (업무담당자) 역할	정보화 부서 (IT 담당자) 역할
<ul style="list-style-type: none"> • 업무수요 제공 • 업무현안 및 문제점 제공 • 분석과제 발굴부터 현업부서 참여가 필수적 • 도출된 과제 후보들을 평가하여 빅데이터의 과제를 최종 결정 	<ul style="list-style-type: none"> • 아이디어 구현방안 제공 • 데이터 유무 확인 • 데이터 정보 제공 • 후보 과제를 평가하는 보조적 역할

> 2. 빅데이터 분석업무 이해

빅데이터 분석업무 극복 과제

기반 인프라 부족

- 수요, 공급자 간 데이터, 서비스 등을 공유하고 활용 할 수 있는 인프라 부족
- 선진국과 IT 솔루션, 데이터 분석 등 분야의 기술력 격차(2~5년)

전문인력 부족

- 빅데이터 가공·분석 전문 인력
- 개발자 에코시스템 부재 : 프로젝트를 통한 전문성 확보 기회

데이터 자원 확보 유통 어려움

- 데이터 개방 및 공유, 거래에 대한 거부감 (빅데이터 보유 업체·기관 간)
- 데이터 축적 > 업무활용 > 기술 성숙 > 데이터 축적의 선순환

기획력 부족

- 빅데이터를 활용한 서비스모델 기획력
 - 창의적인 서비스 아이디어 기획, 산업적 활용 등 사업화 연계

개인정보 보호

- 개인정보보호와 빅데이터의 안전한 활용

> 2. 빅데이터 분석업무 이해

분석업무 사례: 빅데이터를 통한 서비스 진화

과제명	빅데이터를 통한 서비스 진화		사업주체	SK플래닛
과제 필요성 및 목적	<ul style="list-style-type: none"> 기업과 판매 채널의 다양화로 공급이 증가하여 수요 분산을 야기 서비스를 통해 고객 행동을 이해하고 고객이 요구하는 쇼핑 경험 가치에 대한 통찰의 획득이 고객서비스의 주요 자원으로 부상 소비자의 탐색 노력을 줄이고 만족도를 높이는 추천 서비스로 매출 향상 목표 			
활용 데이터	내부 데이터	<ul style="list-style-type: none"> 고객 구매정보 데이터, 고객 로그 데이터 		
분석기법	<ul style="list-style-type: none"> 순차분석, 로그분석 / 통계 분석, 상관 분석, 패턴 분석, 군집 분석 			
추진 내용	<ul style="list-style-type: none"> 머신러닝 라이브러리, 아마존 웹 서비스 클라우드 서버를 사용하여 대용량 데이터 처리를 위한 플랫폼 구축 순차분석과 로그분석을 이용해 최근 3개월간 개인의 과거 사용 이력을 분석해 사용자의 활동 데이터를 활용하여 사용자의 관심 상품을 추천하여 목록 제공 데이터 베이스에 단순히 매칭하는 방식을 벗어난 새로운 알고리즘 개발 			
분석결과 활용계획	<ul style="list-style-type: none"> 빅데이터를 이용한 차별화된 O2O 서비스 계획 수립 			
결과 데이터	<ul style="list-style-type: none"> 사용자 행동 분석 데이터, 사용자 욕구에 최적화된 추천 데이터 			
기대효과	<ul style="list-style-type: none"> 서비스 사용자의 관심 증가로 인한 매출 및 만족도 향상. 			
시사점	<ul style="list-style-type: none"> 출처 : 2016년 글로벌 빅데이터 융합 사례집 대용량 수집분석을 통해 플랫폼을 구축하여 고객 만족도를 높이고 매출 증가로 연결되도록 함 			

> 2. 빅데이터 분석업무 이해

분석업무사례: 빅데이터 기반 고객 맞춤형 타겟 마케팅

과제명	빅데이터 기반 고객 맞춤형 타겟 마케팅	사업주체	신한카드
과제 필요성 및 목적	<ul style="list-style-type: none"> 카드 비즈니스의 본원적인 경쟁력 강화와 고객 만족 경영을 담은 빅데이터 기반의 사업의 강화 고객 소비 패턴 파악으로 카드 서비스와 혜택 제공으로 결제액 증가 목표 		
활용 데이터	내부 데이터	<ul style="list-style-type: none"> 결제 내역 데이터, 기본 인적 데이터 	
	외부 데이터	<ul style="list-style-type: none"> SNS 	
분석기법	<ul style="list-style-type: none"> 회귀분석, 텍스트 마이닝, 군집분석/ 통계 분석, 상관 분석, 패턴 분석, 예측 분석, 추세 분석, 군집 분석 		
추진 내용	<ul style="list-style-type: none"> 비정형 정보 분석 시스템 구축 및 활성화를 통한 비즈니스 성과창출 기반을 마련 로지스틱 회귀분석을 이용한 이탈 고객 예측 텍스트 마이닝을 통한 SNS 소비 트렌드 분석 소비자의 성향에 따라 군집분석으로 분류하여, 해당 군별로 마케팅 전략 수립을 지원하는 데이터 컨설팅 분석방법론 마련 		
분석결과 활용계획	<ul style="list-style-type: none"> 미래의 소비패턴 예측 및 잠재적인 소비 욕구에 맞춘 새로운 상품 체계 개발 		
결과 데이터	<ul style="list-style-type: none"> 주 소비 지역 거리 데이터, 소비 트렌드 분석 데이터 		
기대효과	<ul style="list-style-type: none"> 결제액 증가를 통한 매출 증대 		
시사점	<ul style="list-style-type: none"> 출처 : 2016년 글로벌 빅데이터 융합 사례집 정책수립관점으로, 빅데이터 기반 패턴 분석으로 마케팅 전략 수립 		

- 보건 의료(진료) 정보 분야
- 제조공정 분야
- 공공분야 (전파정보 빅데이터, 고속도로 주행속도)
- 보안 빅데이터 분야
- 영업정보시스템 및 고객지원센터

보건의료(진료) 정보 분야

사회적 이슈와 자동차 보험 사기

이슈

- ✓ 보험사기를 노린 의심환자로서 3년간 10회 이상 교통사고를 겪은 환자가 935명에 달하며, 이들이 찾은 의료기관도 특정한 장소에 몰려 있음
- ✓ 이로 인해 2015년 한해 보험사기로 적발된 금액이 6,459억에 달함

탐색 및 검증

- ✓ 교통사고 환자의 특성과 교통사고 환자를 진료한 의료기관의 특성을 분석
- ✓ 사고다발생 지표, 병상 가동률 지표, 비용 과다와 장기 입원 지표 등을 이상패턴탐지를 위한 지표로 활용

방안 제시

- ✓ 자동차보험 진료비 심사 자료를 활용하여, 보험사기 의심환자와 병원을 찾아내는 "보험사기 빅데이터 정보공유 체계"를 구축하는 방안을 추진

최근 3년 5회이상 사고 1만1460명... 환자 20%가 병원 6곳 몰려 정부 "보험사기 등 의심... 정보공유"

2013년 7월~2016년 6월 5회 이상 다발생 환자를 다수 진료한 병원

구분	소재지	환자수
A의원	대구	530명
B한방병원	서울	375명
C병원	인천	330명
D의원	대구	306명
E병원	인천	306명
F병원	대구	304명



자료: 건강보험심사평가원

- 심평원은 교통사고 진료비 심사 업무를 위탁받아 수행하고 있음.
 - 민간보험사들은 자기고객의 사고기록만 보유하지만, 심평원의 모든 보험사들의 고객 사고기록을 보관하고 있음.

- 심평원의 데이터를 분석하면 일정기간 동안 교통사고가 다발생 환자 파악이 가능.

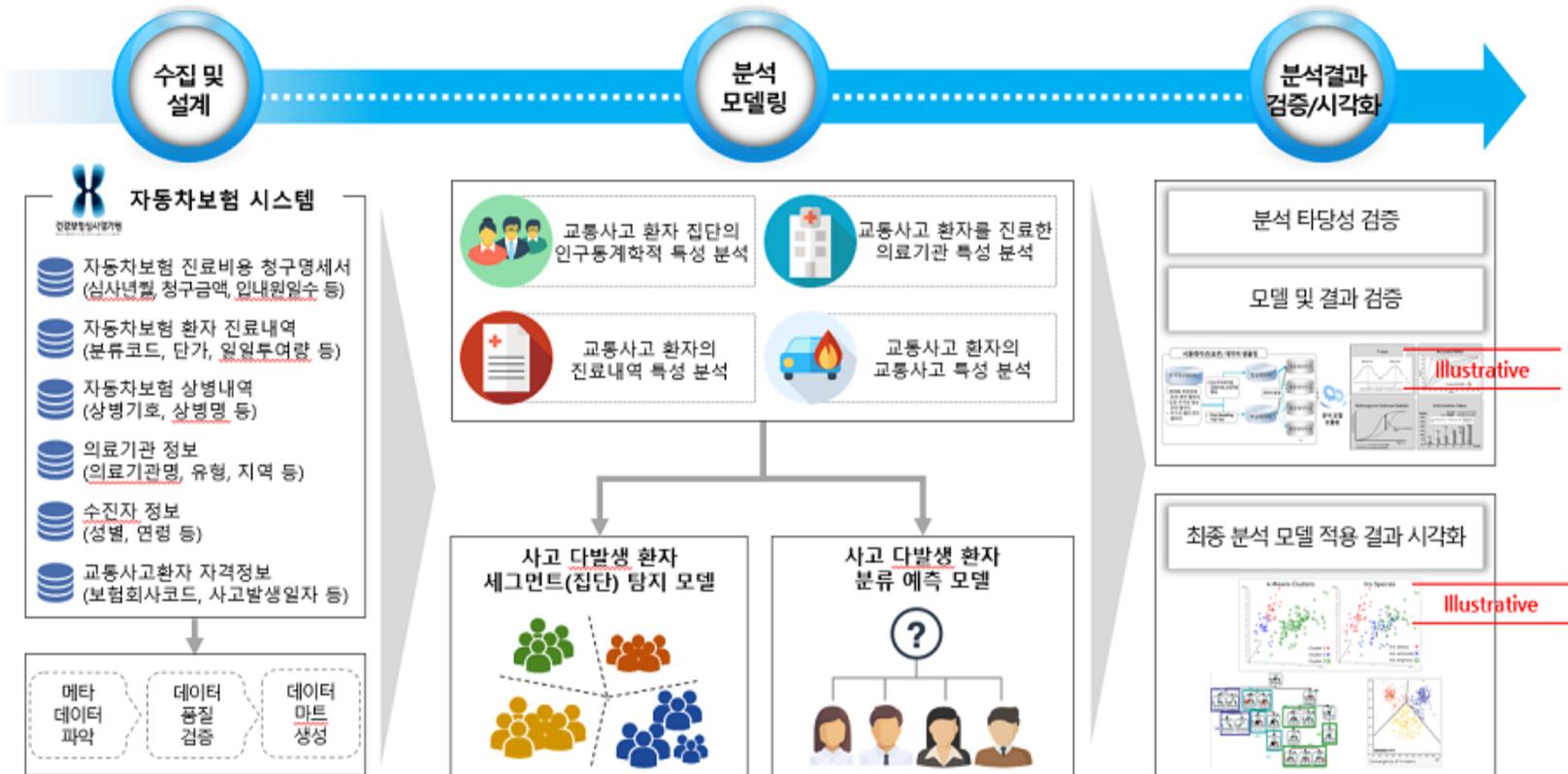
< 연도별 사고발생 횟수별 환자수 >

연도	계	1회	2회	3회 이상	5회 이상	10회 이상	20회 이상
'13.7.~'16.6.	4,360,492	3,781,572	471,014	96,446	10,525	857	78
2015	1,693,327	1,583,932	97,901	10,569	869	51	5
2014	1,650,723	1,545,998	93,949	10,014	701	55	6

출처: 건강보험심사평가원

자동차보험 누수방지를 위한 교통사고 다발생환자 탐지 및 예측

○ 분석 프로세스



자동차보험 누수방지를 위한 교통사고 다발생환자 탐지 및 예측



2차년도: 빅데이터 분석 및 서비스
 자동차보험 누수방지를 위한 교통사고 다발생
 환자 탐지 및 예측.pdf

교통사고 다발생 환자 및 의료기관 탐지 및 예측

군집분석

- 의료기관 및 사고환자 데이터로 K 평균 군집화 알고리즘 적용하여 각각 정상 집단 정의
- 데이터의 유사성에 따라 주어진 데이터를 K개의 군집으로 묶음

이상점수 산출

- 정상 집단의 중심점(평균)과 반경 R 산출
- 정상 집단의 중심점으로부터 거리 D 산출
- R과 D로 이상점수(Anomaly Score, AS) 산출
- 이상점수가 높을수록 이상 패턴일 확률 높음

예) ○○병원의 이상점수 = $\frac{D_i}{R_i} = \frac{10}{4} = 2.5$
 △△병원의 이상점수 = $\frac{D_i}{R_i} = \frac{44}{4} = 1.1$

➡ ○○병원이 △△병원 보다 이상 패턴일 확률 높음

이상패턴 탐지

- 모델을 생성할 때 이상점수 임계치
- 정상 집단으로부터 멀리 떨어져있었음을 경우 이상집단으로 판별
- 이상점수 임계치에 따라 이상집단



허위·과다 입원 등 연성보험사기 증가로 인한 사회적 보험료 상승을 방지하기 위해 관리체계 구축의 필요성이 증대되고 있습니다.

특히 자동차보험 사기 수법의 지능화·조직화 추세로 빅데이터와 정보기술을 활용한 보다 과학적인 자동차보험 청구 심사업무 구축이 필요합니다.

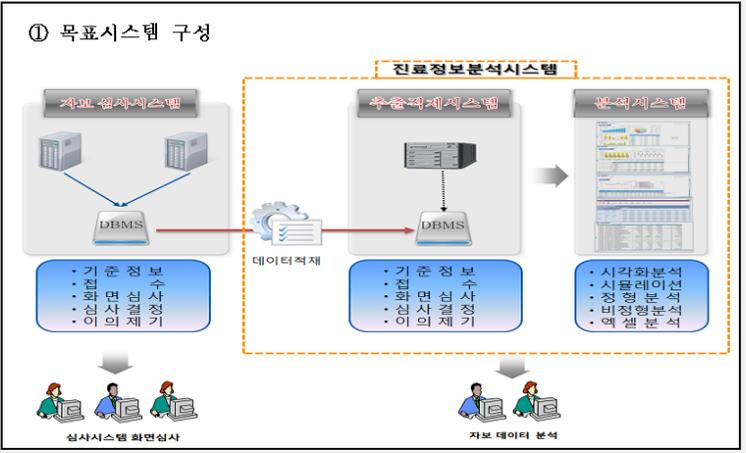


출처: 2017년 빅데이터 성과공유대회, 행정안전부, 한국정보화진흥원

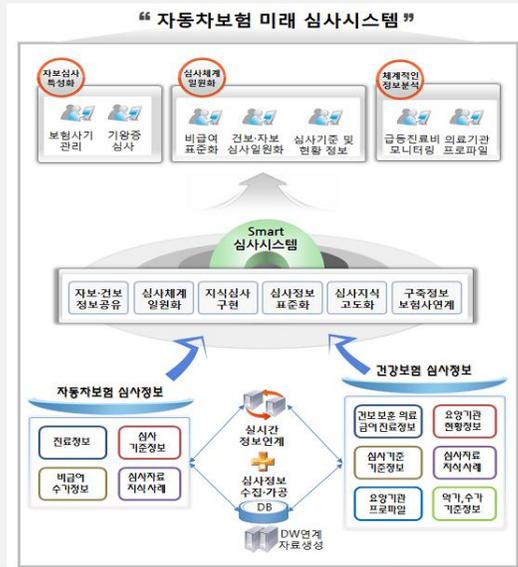
Case Study

빅데이터 분석시스템: 보건의료(진료)정보 분야

1차년도: 진료정보분석시스템



2차년도: 사고다발 발생 분석(차세대)



- 지역별 사고별 유형을 분석한 데이터를 시각화 하기
- 모니터링을 통한 문제 인식
- 임원 및 현업 담당자도 동일한 화면으로 같은 시점에 정보를 공유

- 주요지표를 통한 현황분석의 비교 및 검토
 - 상위 랭킹을 통한 전체 비율을 직관적으로 인식

- 관련된 다른 주요지표를 통한 현황분석의 비교 및 검토
 - 상위 랭킹의 변화를 통한 상대적 중요성을 인식

- 관련된 데이터의 화면 출력 및 레포팅
 - OLAP 기능을 통한 필요한 데이터를 생성

빅데이터 기반의 자동차보험 진료정보분석시스템

추진 목적

- 사용자가 직접 신속하고 다양한 분석으로 심사업무 효율성 제고
- 자동차 보험 진료정보에 근거한 과학적 심사 및 심사기준 개발
- 부당청구 개연성이 높은 의료기관 및 수진자 관리
- 자동차보험 진료비 통계정보 제공 등 정부 정책 지원
- 자동차보험 대내외정보 서비스를 위한 빅데이터 전략 수립

분석(Analytic)

- 빅데이터 기반의 분석전용 데이터
- 분석 및 시뮬레이션 실시간 결과 확인

업무연속성 확보

- 분석시스템을 활용한 내 외부 서비스 업무의 지속성 유지

사업범위

- 빅데이터 기반 자동차보험 분석(Analytic) 체계 구축
- 자동차보험 심사정보 분석시스템 인프라 구축
- 데이터의 분석 및 시뮬레이션을 통한 의사결정관리 프로세스 구축

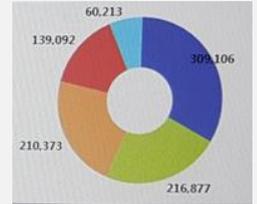
구축방법 및 전략

- 빅데이터 기반 분석(Analytic) 전문 심사 데이터의 확보
- 신속성과 효율성을 기반으로 한 데이터 분석 및 시뮬레이션 프로세스 제공
- 진료정보 분석 전용 장비 도입 및 구축
- 통계 데이터 생성, 진료 분석용 데이터 등 분석용 프로그램 개발

분석시스템 도입 효과

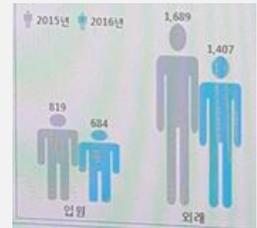
업무생산성 향상

- 단순 조회 업무(Reporting)에서 Ad-hoc분석, 통계, 시뮬레이션 업무까지 업무 개선 및 확대
- 2~4시간 소요된 업무 분석 시간을 분 이하로 감소함으로써, 업무 생산성 증대



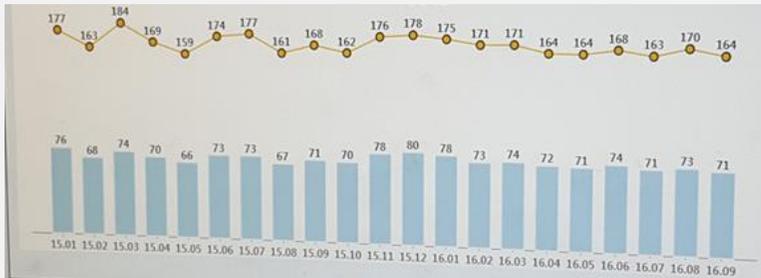
분석 데이터 정확성

- 원본 데이터를 분석환경에 그대로 사용함으로써, 다양한 뷰를 통한 분석기법 활성화
- 자료 분석 가능 기간이 1개월에서 3년으로 증가되어, 데이터 정확성 및 대표성 확보



빅데이터 현황분석

- 실시간 데이터를 심사운영시스템에 적용함으로써, 운영 데이터간 실시간 데이터 분석
- 실시간 Dashboard 구성을 통하여, 시각적이고 즉각적인 인사이트 발굴

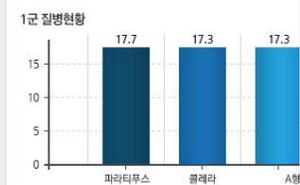


Case Study

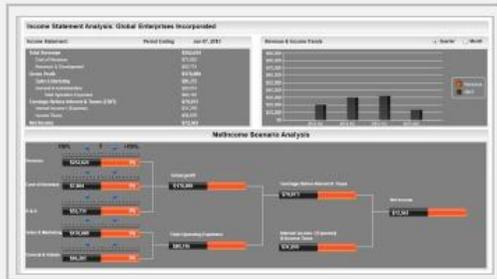
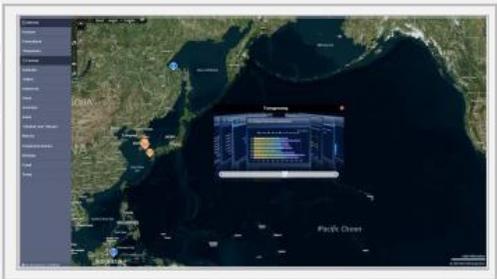
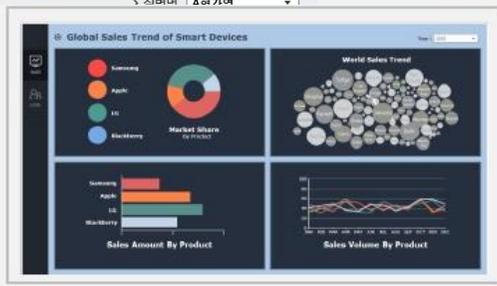
빅데이터 분석시스템: 보건의료(진료)정보 분야

다양한 대시보드의 시각적 역할

질병 발생 종합 현황 >년도 2013년 >군 3군 조회 상세

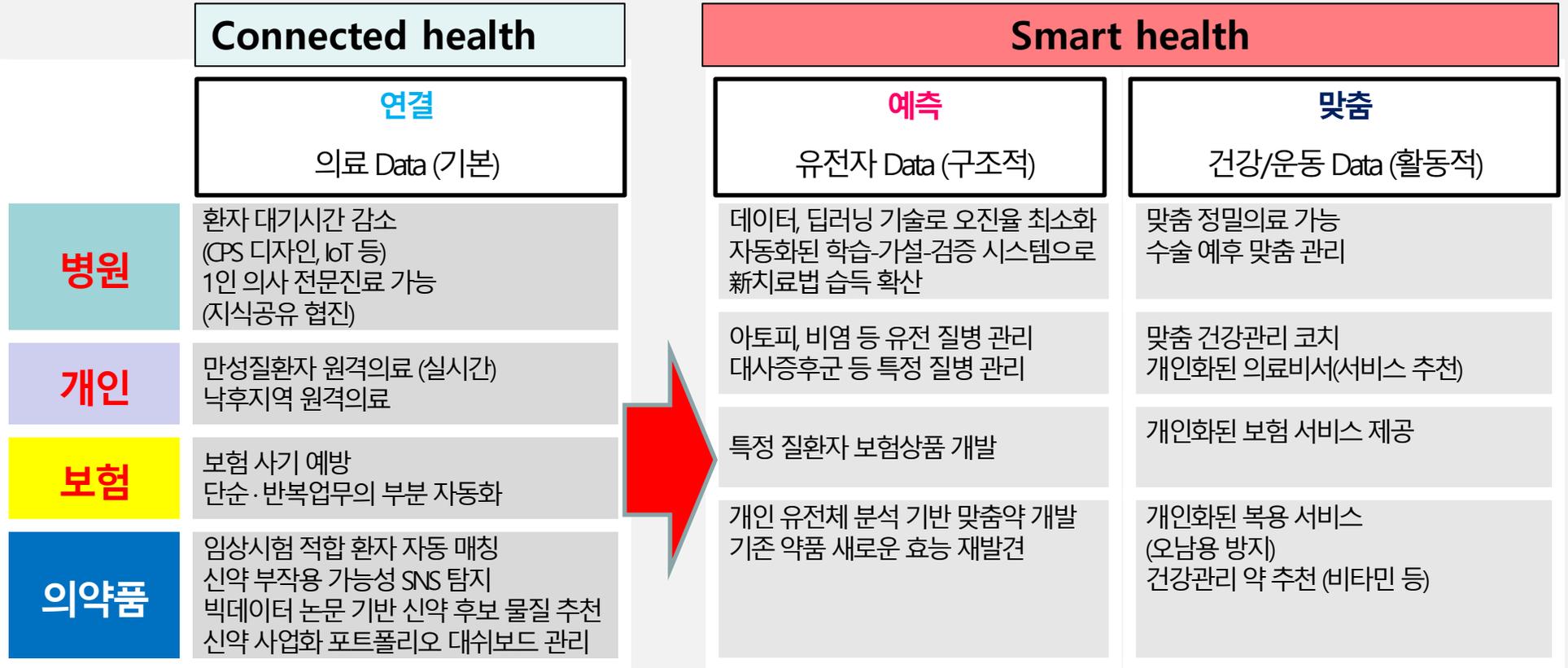


구분	2009년	2010년
1군		
A형간염	56	
세균성이질	51	
장출혈성대장균감염증	50	
장티푸스	47	
볼레라	67	
파라티푸스	42	
1군 소계	313	



By (주)비아이메트릭스

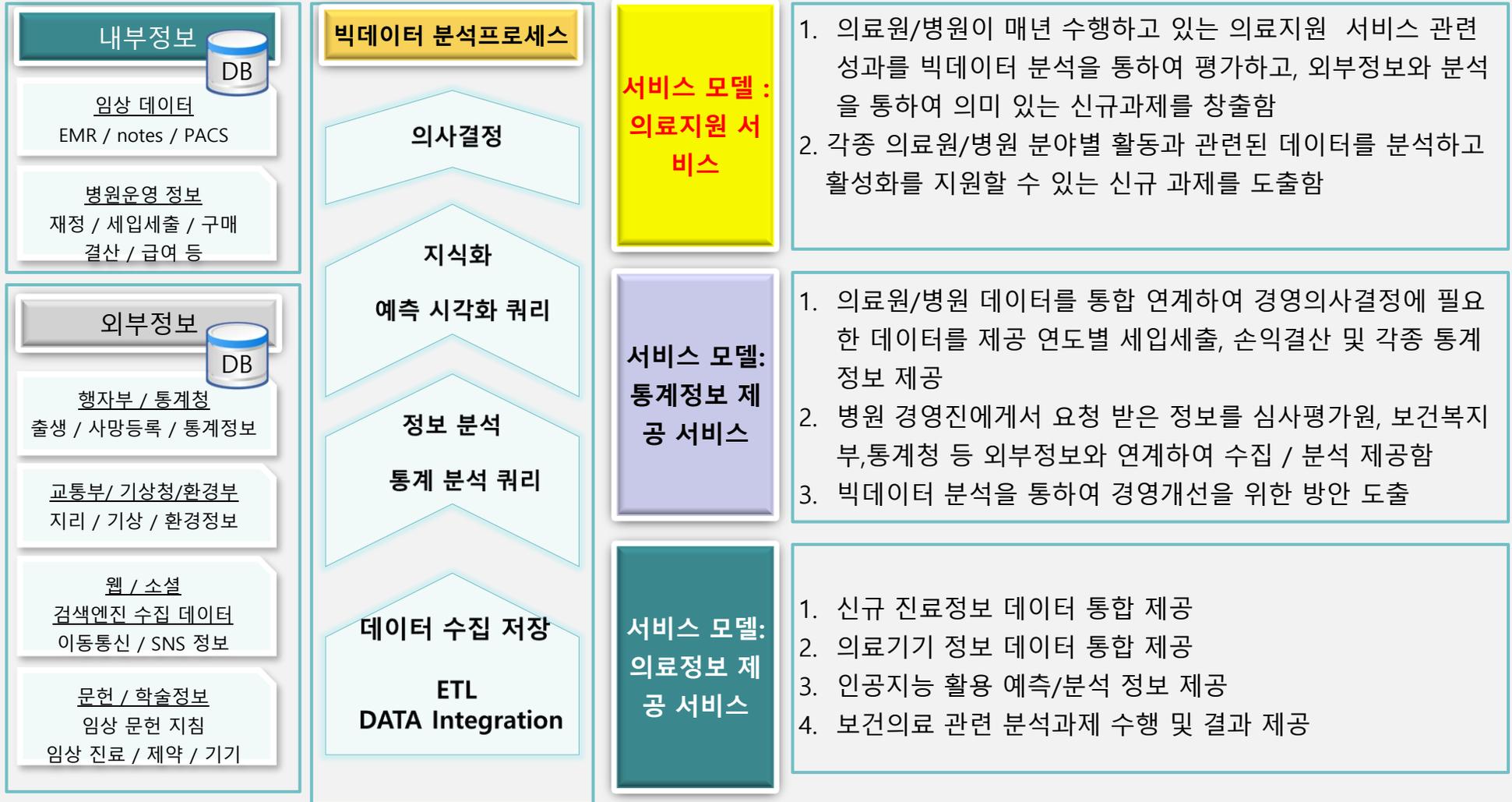
디지털 헬스케어의 미래



자료: SPRi. 인공지능과 헬스케어 산업 혁신 표 재가공

출처: 창조경제연구회 (KCERN)

의료원 빅데이터 분석센터 서비스 모델 개념도 (예시)



제조 공정 분야

Case Study

빅데이터 분석시스템: 제조공정분야

대용량 데이터 프로세싱 분석(Analytic)

4차 산업혁명 관점

- ✓ **인더스트리 4.0**
 - 사물인터넷, 빅데이터 등 IT와 물리 분야 기술을 활용한 생산방식의 전면적 개편
 - 사이버물리시스템에 의한 패러다임 변화

출처: 한국정보화진흥원, 2015

제조분야의 빅데이터

- ✓ **제조분야 기대수요**
 - IoT 인프라 확대에 따른 시장 창출 수요
 - 생산성 개선 프로젝트 수요 기대
 - 재고/생산량/불량률 예측 프로그램 개발
- ✓ **한국 빅데이터 "제조업에서 성공해야"**
 - 시장이 포화와 기술 평준화에 따른 한계에 부딪힌 제조업을 성장시킬 해결책 필요
 - 제조 공정의 개선과 관련하여 깊이 있는 통찰력을 제공하는 빅데이터 분석에 대한 높은 기대

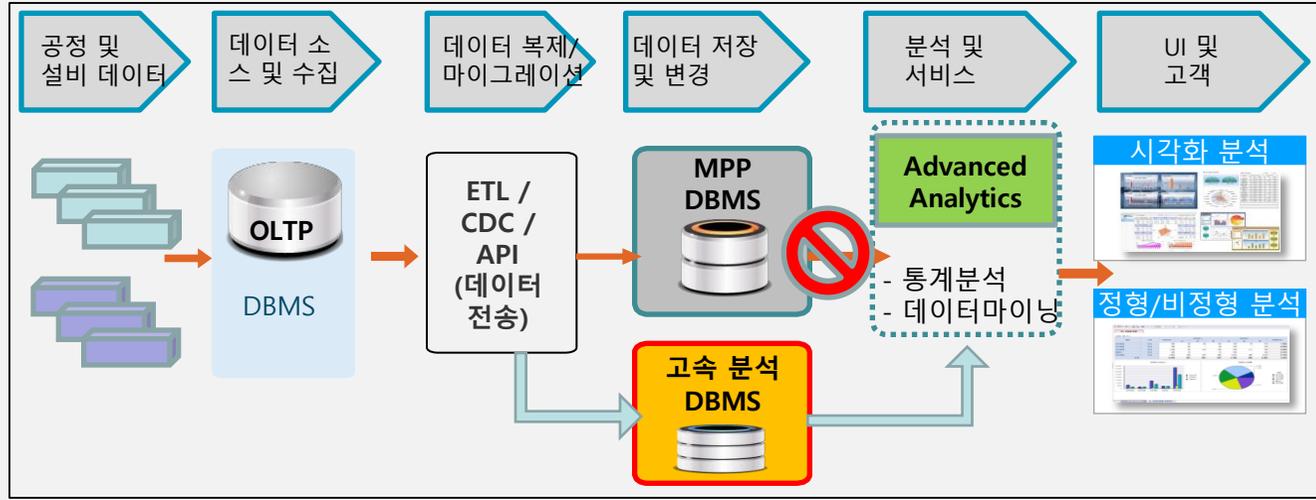
- ✓ **제조공정 품질불량 분석**
(테스팅 결과 분석에 따른 개선사항 발굴)

- ✓ **MES 도전과제**
 - 프로세스 제약 제어 (Process Limit Control)
 - 오류탐지와 분류 (Fault Detection)
 - 손실발생에 대한 예측 (Prediction)
 - Historical 데이터 분석
 - 실시간 작업 분배 (Real time Dispatching)

출처: MES 빅데이터 분석(활용)을 통한 생산성 및 품질 향상 사례, 김용한, 2013

- ✓ **MES 최적화 시뮬레이션 분석**
(설비관리의 상태분석, 공정재고의 최소화)

불량분석 및 이상점 Detection



- ✓ 공정 실시간 감시, 제품 불량 분석, 이상점 Detection 처리를 위한 설비 관련 데이터 분석
- ✓ 생산관리 및 6시그마 관점의 과학적 분석기법의 테스트 및 발굴
- ✓ 기존의 10억 건 정도의 표본추출에서 1,000억 건~5,000억 건 이상의 전수조사를 통한 획기적인 분석 방향 제시

제조업 혁신

Process :

R&D

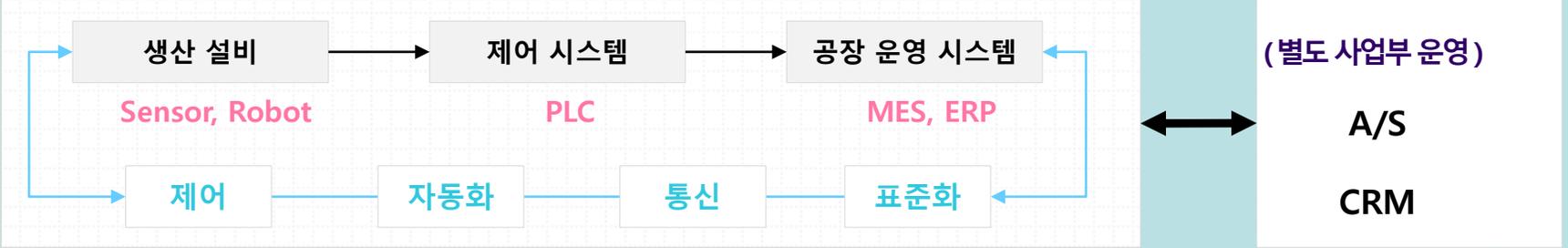
설계

부품

생산

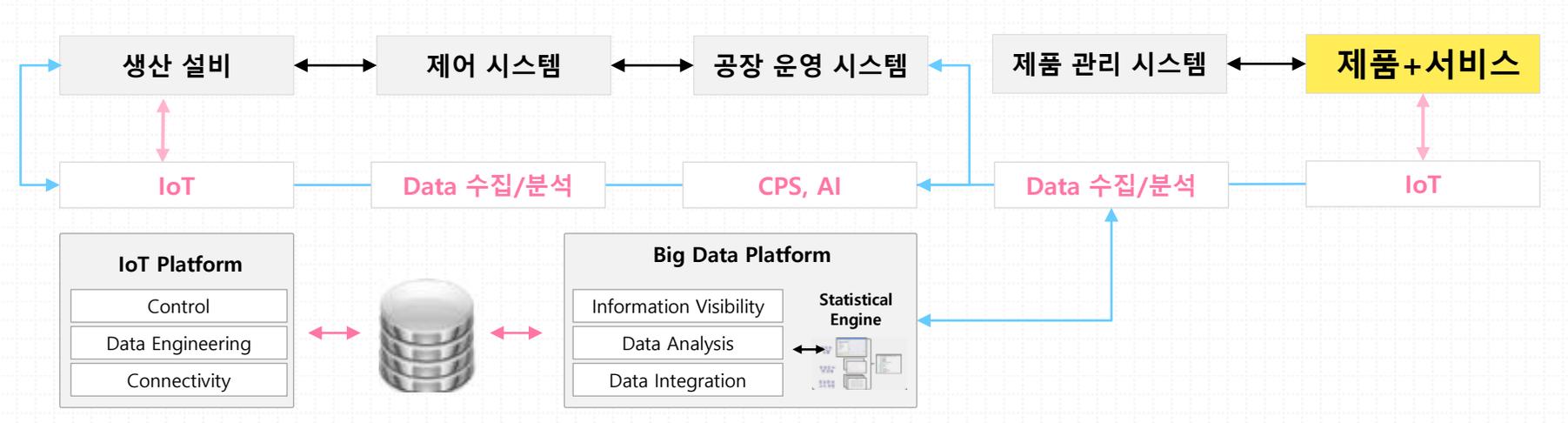
제품

자동화 공장



Smart Factory

연결화/ 지능화 /가상화



제품 판매 후 고객과 간헐적 관계(Buy-Sell)에서
제품-서비스와 고객과 상시연결(Relationship)으로

출처: 창조경제연구회 (KCERN)

스마트 산업 4 Step - 스마트팩토리

1 데이터 수집 IoT

각 설비에 센서 부착, IoT 소통 체계 구축
테스트 공정별 결과 데이터 수집

2 저장 & 분석 CLOUD/BIG DATA

조립 공정간 가동 정보, 생산 정보,
품질 정보 실시간 자동 분석

3 가치 창출 A.I

실시간 공장 운영 현황 분석 및 제어
롯데별, 공정별 시뮬레이션 분석결과를 공정에 반영

4 최적화 기술 융합

공장 자동화 → 불량률 0.001%,
에너지 비용 30% 감소 출처: 창조경제연구회 (KCERN)

AS - IS

- 제품 품질 측정 데이터 1일 10억 건의 규모로 분석 결과 도출
- 롯데별 공정별 데이터에서 일정한 패턴을 찾기까지 5일이 소요됨
- 분석결과가 부족하여 개선활동을 위한 업무범위가 한정되어, 개선활동에 일주일 이상이 소요됨
- 분석을 위한 원본 데이터를 여러 부서에서 여러 명의 담당자가 중복적으로 관리하여, 시스템/스토리지/과업 등에 비용이 지속적으로 발생함

개선사항

- 측정데이터 분석 단위를 3개월 90억 건 이상으로 확대하여 분석 건수의 지속적인 증가를 희망
- 롯데별, 공정별, 세부단위별, 부품별 등 일정한 패턴을 찾는 시간을 2시간 이내로 희망
- 단일화된 데이터 저장, 분석시스템 활용하여 설비의 최적화, 공정과 설비의 연관관계 분석을 자유롭게 하기를 희망

TO - BE

- 분석할 데이터 단위를 1년치 1천억 건으로 확대하여 지능형 분석이 가능
- 분석결과 도출 및 축적, 시뮬레이션에 필요한 빅데이터 분석시스템을 상시 활용하도록 제공
- 업무개선을 세부활동이 5일에서 4시간을 줄어들어 지속가능한 성과체계가 이루어짐

전파정보 빅데이터 활용시스템 구축

사업의 이해

전파정보
현황분석

- 전파정보 통합 운영
- 전파정보 공개 확대
- 전파정보 접근 제한
- 체계적 관리 미흡

정보기술
발전동향

- 개방성 제고
- 호환성 보장
- 재사용성 증대
- 표준화 구축

빅데이터
현황분석

- 새로운 가치 인식
- 공공/대기업 주도
- 선진 기술 격차
- 관련 산업 활성화

공유
융복합

사업의 목적

전파분야 신규 비즈니스 창출 및 대국민 전파정보 활용 촉진

「 유의미한 가치정보 발굴 」

- 〔 전파정보 빅데이터 플랫폼 구축 〕
- 〔 전파정보 활용 서비스 추가 개발 〕
- 〔 시스템 안정성 및 사용자 편의성 강화 〕

빅데이터 기반 확보

- 빅데이터 분석을 통한 선제적 정책 및 미래전략 수립 적극 활용
- 실제 주파수 이용을 고려한 분석 방안 검토
- 공공분야 빅데이터 연계

전파정보 정책 반영

- K-ICT 전략 (빅데이터 활용 지원. '15.0■
- 공공용 주파수 수급체계 ('16.06.24 시행)
- 주파수 분석·발굴·재(공동)사용·활용의 순환구조 정책 기반필요

활용기반 조성

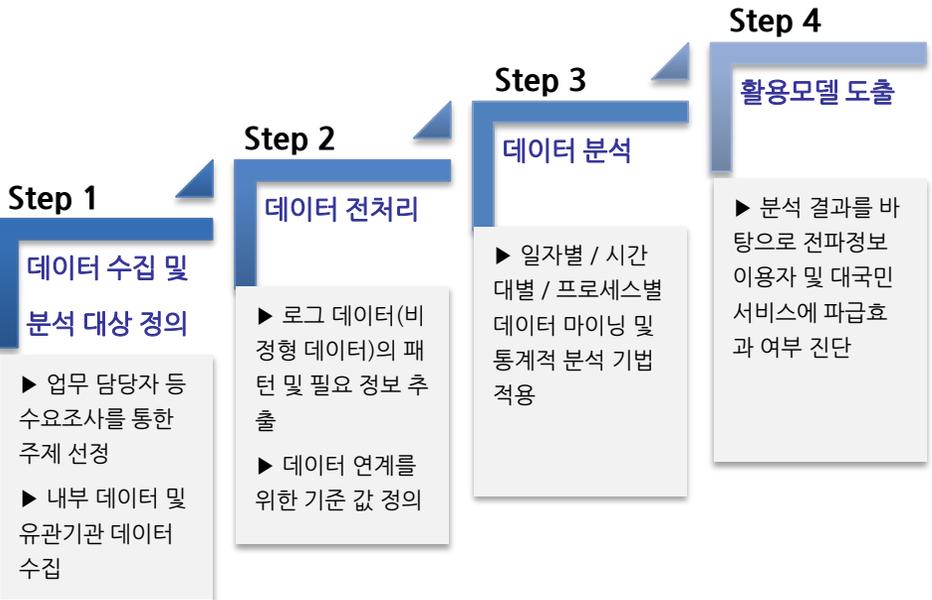
- 빅데이터를 통한 창의적 활용서비스 개발 및 신규 비즈니스 창출 도모
- 대국민 전파정보 활용 촉진
- 전파분야 선진 기술 적용에 따른 위상 제고

> Case Study

빅데이터 분석시스템: 공공분야(전파빅데이터)

전파정보 빅데이터 가치창출을 위한 활용모델

활용모델 도출 프로세스



활용 데이터



활용 모델 예시

『평창 동계올림픽 무선 트래픽 모니터링 서비스』

서비스 내용

- 국제적 이벤트인 평창 동계올림픽 및 유니버시아드 등 단기간 인구밀집이 예상되는 지역에 대하여 무선 트래픽 사용량 집중 모니터링

기대 효과

- 전파정보 이용자들에게 원활한 서비스를 제공을 통한 만족도 향상 기대
- 무선 트래픽 모니터링 및 패턴 분석을 통한 주파수 할당 등 전파정보 의사결정 근거 마련

서비스 제공



『해안 무선정보를 활용한 선박 사고 조기 대응 서비스』

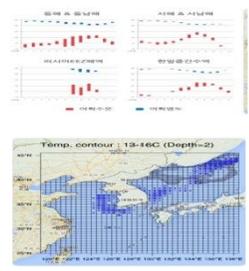
서비스 내용

- 해상 활동(어선, 유조선, 화물선 등)에서 발생하는 전파정보 모니터링 및 해도를 통한 시각화 제공 서비스

기대 효과

- 해상에서 발생하는 전파정보를 해도 기반을 통하여 위치정보 제공 서비스를 통한 사고 등 특이사항에 대하여 신속한 처리가 가능

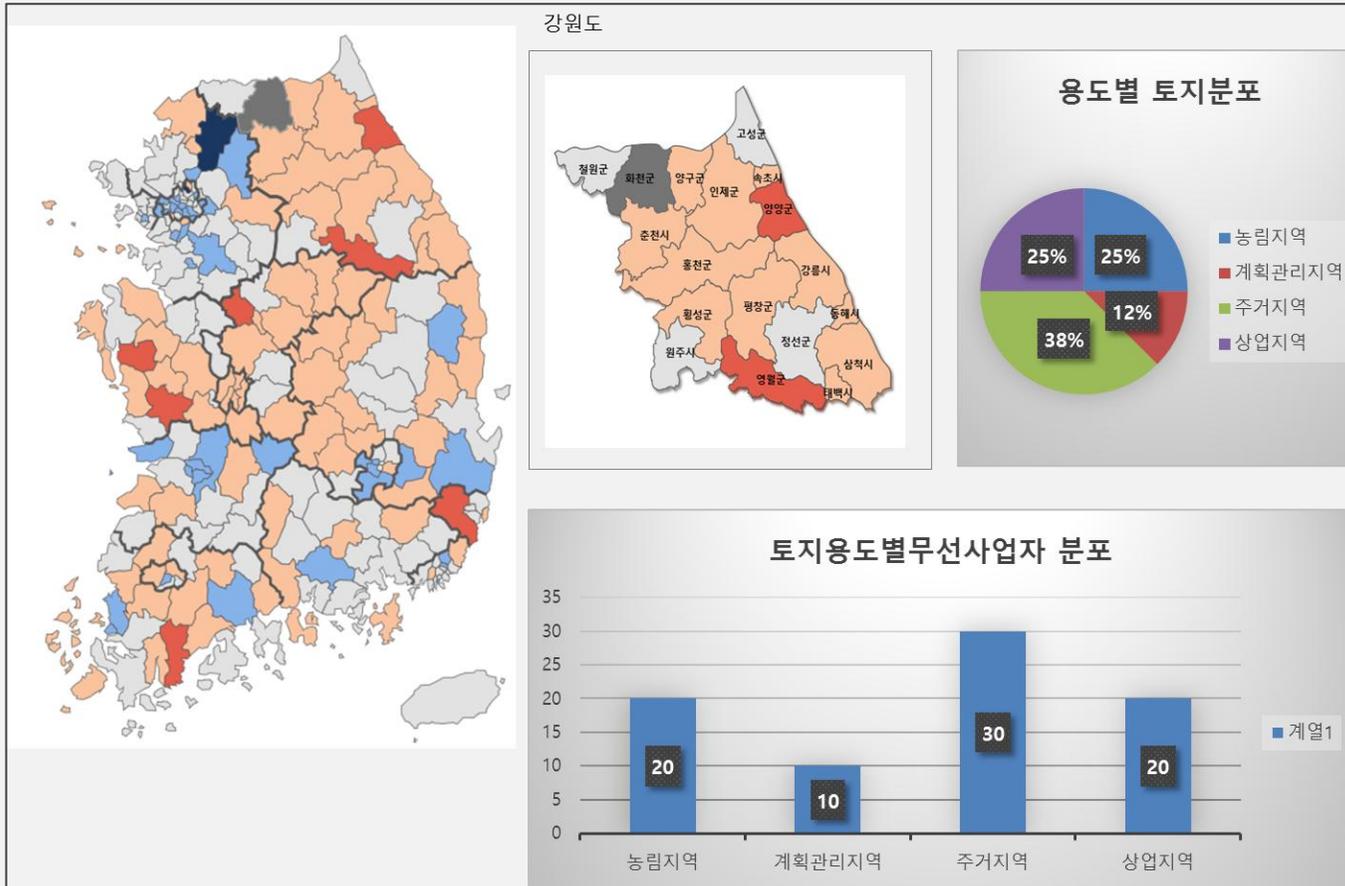
서비스 제공



> Case Study

빅데이터 분석시스템: 공공분야(전파빅데이터)

✓주제 상세: 토지용도별 무선국 분석 - cont.



구/성/방/안

화면 및 기능

- 지역선택 (도단위) -용도별 토지분포
- 지역선택(도단위) - 토지용도별 무선사업자 분포

조회방법

- 분석마트 쿼리 사용

> Case Study

■ 빅데이터 분석시스템: 공공분야(전파빅데이터)

✓주제 상세: 토지용도별 무선국 분석 - cont.

토지용도선택

주거지역

주거지역

강원도 홍천군 홍천읍 1

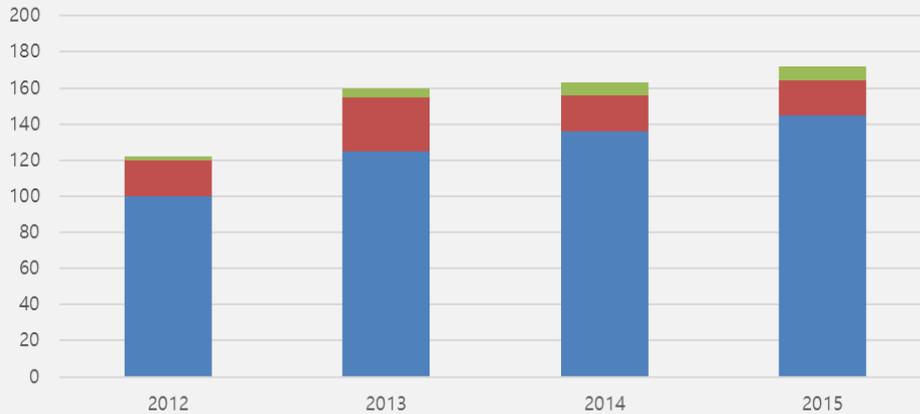
강원도 홍천군 홍천읍 2

강원도 홍천군 홍천읍 3

강원도 홍천군 남면

강원도 홍천군 북면

무선사업자 현황



토지용도변경

상업지역

선택

위치기반

도시기반

Analysis

용도변경 예측추이

연도	블루 선	적색 선	녹색 선
2012	100	20	0
2013	125	30	5
2014	135	25	5
2015	145	25	5

구/성/방/안

◆ 화면 및 기능

- 토지용도선택 (시군구단위) - 무선사업자 현황
- 지역선택(시군구/토지용도 단위) - 토지용도별 무선사업자 분포

◆ 분석 방법

- 토지용도변경(위치기반) - 가장 가까운 변경된 용도의 무선국 수 예측
- 토지용도변경(도시기반)- 선택한 도시의 토지용도로 변경될 경우의 무선국 수 예측

Case Study

빅데이터 분석시스템: 공공분야(전파빅데이터)

전파정보 빅데이터 서비스 예시

무선국 정보 - 이동통신 무선국 찾기

K-ICT 스펙트럼 Map

검색어를 입력하세요

전파라이브 **new** Spectrum/GIS Map 주파수 정보 무선국 정보 전파 활용 서비스 맞춤형 전파정보 LAB 자료실 커뮤니티 시스템 소개

무선국 정보 > 이동통신 무선국 찾기

현재위치 서울특별시 노원구 공릉동

범례 2G 3G 4G Wibro 2G(육내) 3G(육내) 4G(육내) Wibro(육내)

검색 서비스 전체

주소 서울특별시 노원구 화랑로 815, 삼육대학교 미래관(공릉동)

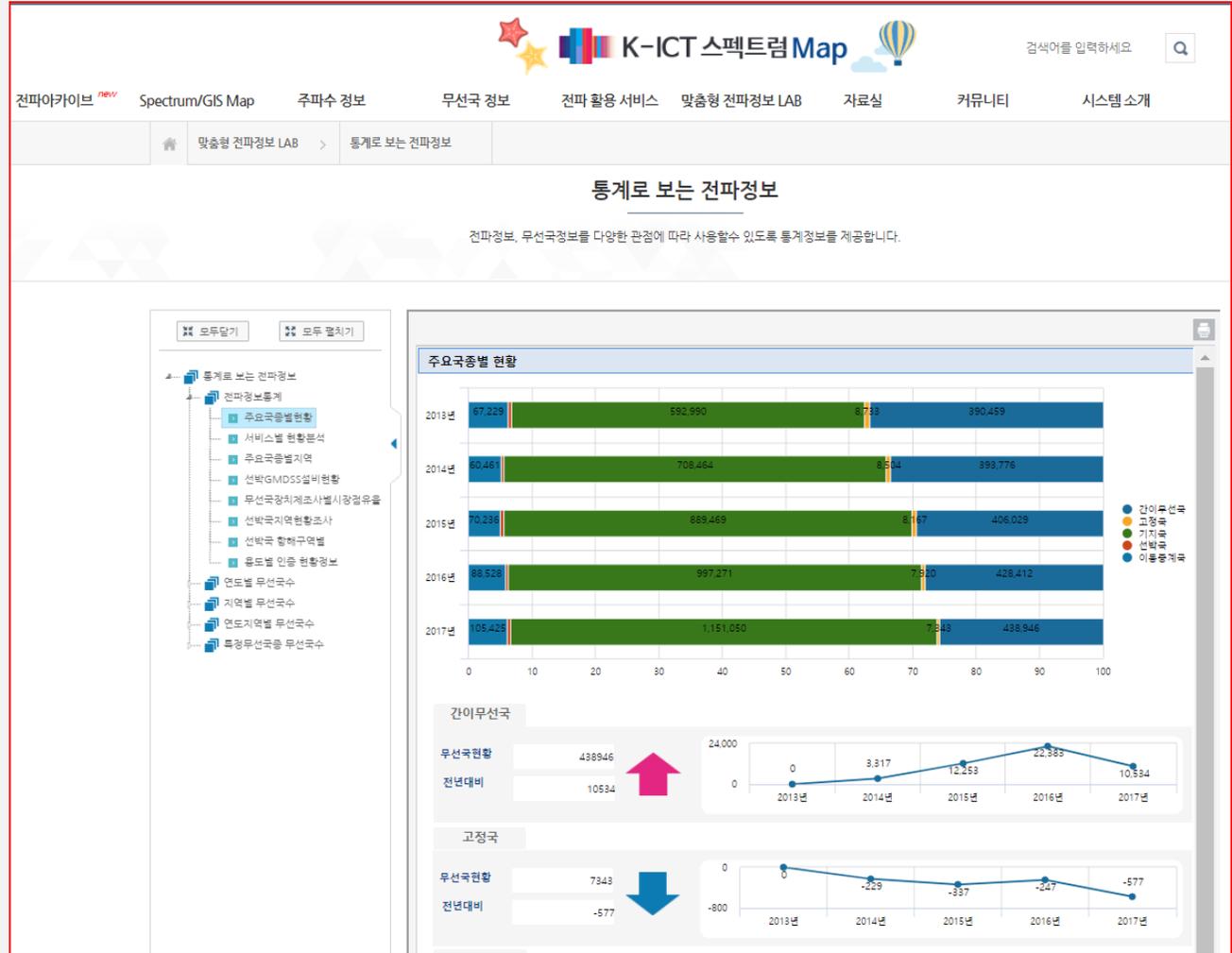
검색반경 1KM

이동통신 무선국 111 건 ※ 설치장소는 무선국 허가정보를 기반으로 제공되며, 지도상의 위치정보(GPS)는 실제위치와 차이가 있을 수 있으므로 참고하시기 바랍니다.

허가번호	설(상)지 장소	경도	위도	용도	인증번호(주장치)	일련번호	전세기수	주파수 (MHz)	출력 (W)	대표이득 (dB)	해발고 (m)	지상고 (m)	노출고 (m)	안테나 설치형태
322013110052724	서울특별시 노원구 공릉동 26-27번지	127° 06' 17.18"	37° 38' 04.46"	이동전화(LTE)서비스 제공용(지상_광기저국)	MSIP-CRM-3TC-RMBLROLO25	13A00113	2	1820	25	15	34	14	14	통신주(CP주)

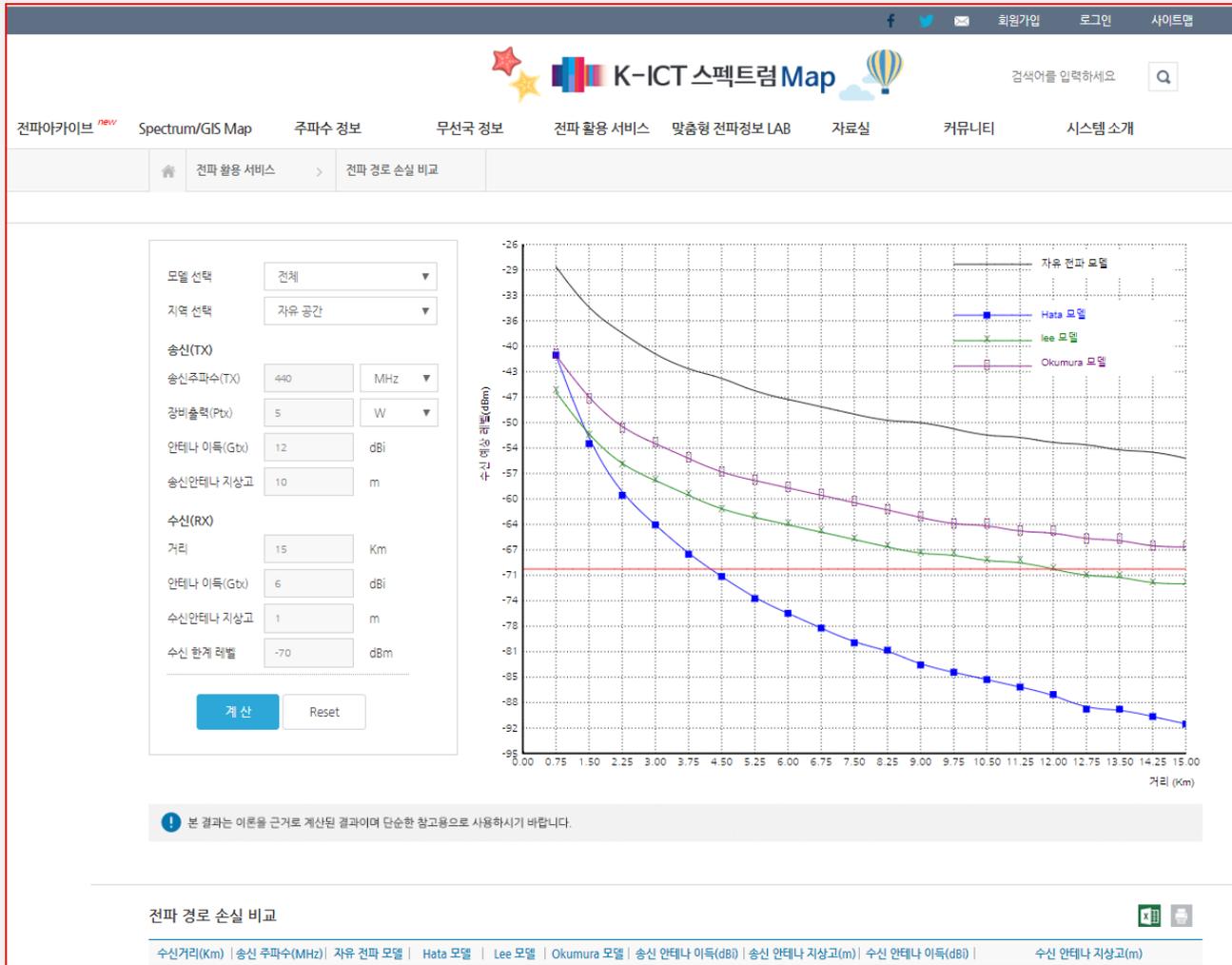
전파정보 빅데이터 서비스 예시 - cont.

맞춤형 전파정보 LAB - 통계로 보는 전파정보



전파정보 빅데이터 서비스 예시 - cont.

전파활용서비스 -
전파경로 손실비교

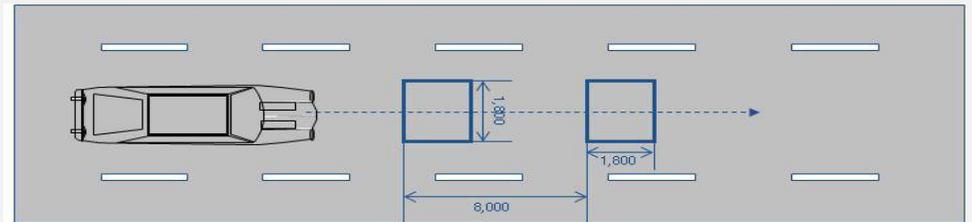


고속도로 주행속도 분석을 위한
주제 발굴 및 초기 탐색 과정

VDS (Vehicle Detect System)의 급감속 구간 탐색



아이디어 발굴을 구체화 하기 위한
분석 준비 단계에 해야 할 업무가 많음



LOOP SENSOR

▪ LOOP SENSOR 규격

구 분	규 격
측정가능 속도	10Km/h ~ 250Km/h
속도 오차율	± 5% 이하
검지 소요시간	5 msec 이하



VDS (Vehicle Detect System)의 급감속 구간 탐색

No	주제명
1	VDS 지점별 일단위 통행속도
2	VDS 지점별 인접 상류부 속도차이 산출

VDS (Vehicle Detect System)의 급감속 구간 탐색

중부내륙선 35일치 데이터: 약 5천 1백만 건 분석

Workflow (개선)

VDS_0450 (중부내륙선) 전체 데이터 대상 : 평균속도 (avg_speed) 0 이상인 경우만 포함, SQL 구성 조정으로 속도 향상 추진
 - 총 51,262,229건, 분석소요시간 15,608초 (약 4.3시간, 원격접속 기준)

Workflow Details:

- PreProcessing:** Includes nodes like RenameCols, FillVal, FillNull, DropCol, JoinData, AddDerCols, UnionData, FilterData, Melting, Casting, Normalizing, FillNulls, ReorderCol, SortData, FilterData2.
- Analytics:** Includes HistoGRAM-1, LinearReg-1, and KMeans-1.
- Visualization:** Includes Linear and Means nodes.
- Controls:** Includes a console window showing execution logs.

Mode	Node Name	Status	ElapsedTime	Node Name	Var Name	Var Value	Real Value
RunTime	KMeans-1	END	64.025 sec.				
RunTime	LinearReg-1	END	4.365 sec.				
RunTime	HISTOGRAM-1	END	6.496s				
RunTime	LinearReg-1	END	2.734 sec.				
RunTime	HISTOGRAM-1	END	6.496 sec.				
RunTime	RScript-5	END	360.073 sec.				
RunTime	VDS_0450	END	15124.416 sec.				

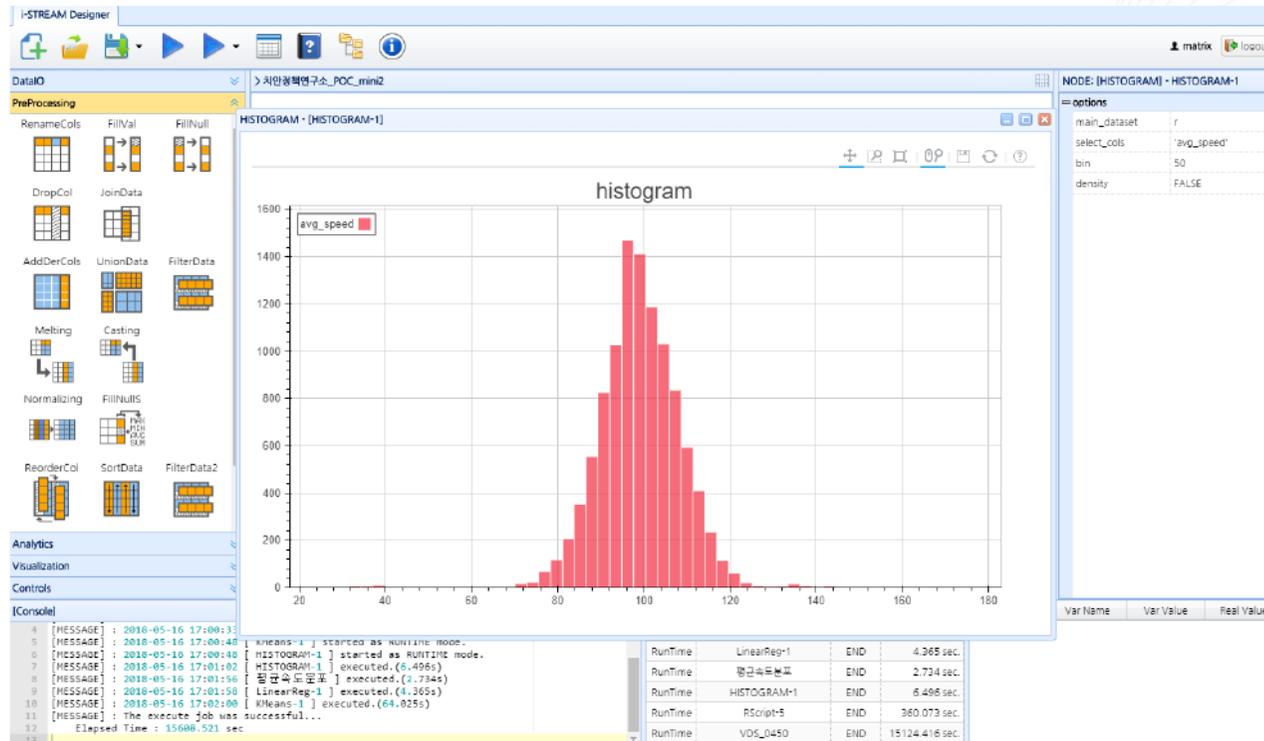
출처: 비아이매트릭스

VDS (Vehicle Detect System)의 급감속 구간 탐색

중부내륙선 35일치 데이터: 약 5천 1백만 건 분석

평균속도분포 (Histogram)

1일치 데이터와 크게 다르지 않은 분포이나 분산이 좀더 커짐 (평균속도가 좀더 넓은 구간에 분포됨)



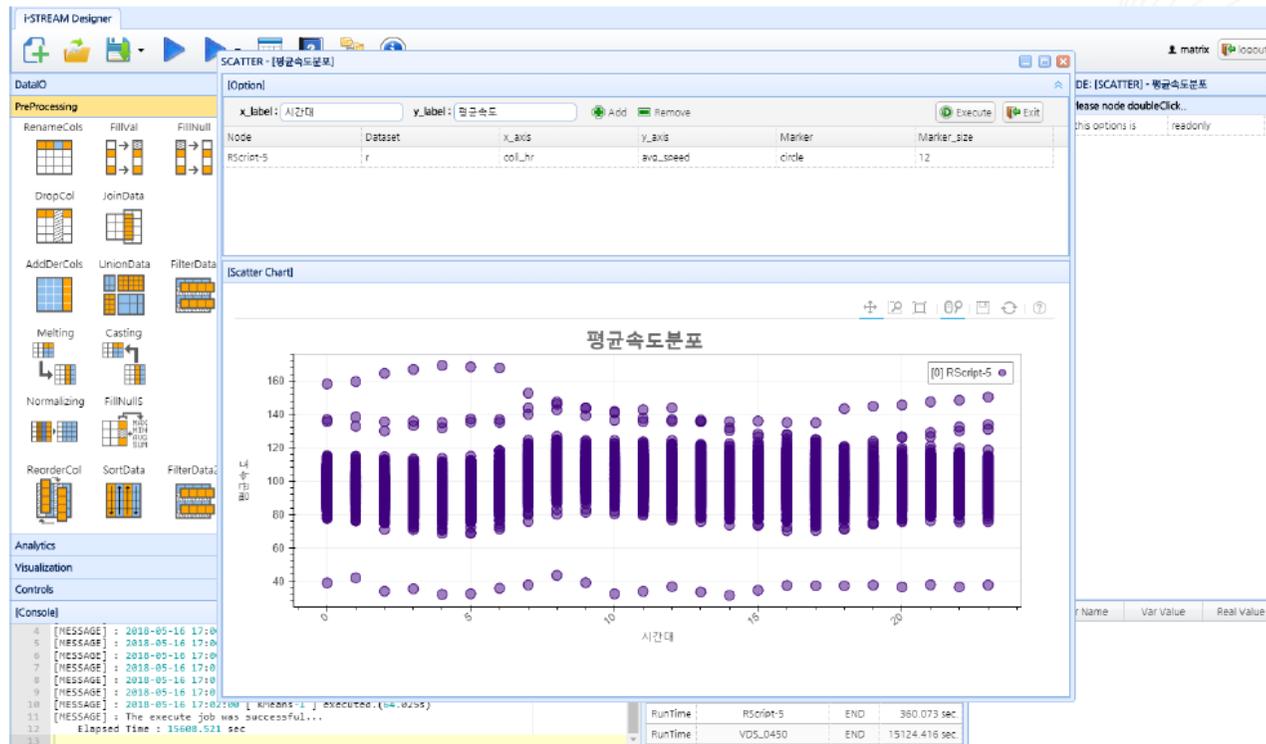
출처: 비아이매트릭스

VDS (Vehicle Detect System)의 급감속 구간 탐색

중부내륙선 35일치 데이터: 약 5천 1백만 건 분석

평균속도분포 (Scatterplot)

전형적으로 새벽시간대가 평균속도가 높고 저녁시간대가 높았던 1일치 분석결과는 보편적이지 않음을 관찰할 수 있음



출처: 비아이매트릭스

VDS (Vehicle Detect System)의 급감속 구간 탐색

중부내륙선 35일치 데이터: 약 5천 1백만 건 분석

회귀분석 (평균속도~시간대)

음의 상관관계이나 계수가 작아졌음 (-0.54 → -0.06) : 새벽시간대에 평균속도가 높으나 그 선형 관계성은 미미함
- 시간대별 분석이 필요함을 나타냄

Linear Regression - [LinearReg-1]

```

Stdout
1 Coll:
2 ln(formula = coll_hr ~ avg_speed, data = data)
3 Residuals:
4   Min       1Q   Median       3Q      Max
5 -14.8921  -5.8144  -0.1967   6.0846  15.0270
6 Coefficients:
7   (Intercept)  5.804085  0.688130  8.425  <2e-16
8   avg_speed   0.057458  0.006900  8.317  <2e-16 ***
9   ---
10  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
11 Residual standard error: 6.9 on 18654 degrees of freedom
12 Multiple R-squared:  0.00045, Adjusted R-squared:  0.000357
13 F-statistic: 69.17 on 1 and 18654 DF, p-value: < 2.2e-16
14
15 coll_hr ~ avg_speed
    
```

Data Result	Coll	Avg Speed	Predicted
0450VDE00100	0	89.38461538461539	10.93896635005014
0450VDE00200	0	93.11157455683004	11.154110789250277
0450VDE00300	0	100.65947403910991	11.587799686970786
0450VDE00400	0	98.19491094147583	11.4461902689222
0450VDE00500	0	98.82930863380747	11.482641035210380
0450VDE00600	0	94.49301429219073	11.233485862703140
0450VDE00700	0	97.50352230759035	11.406404300079960
0450VDE00800	0	101.18127448958548	11.017781414030434
0450VDE00900	0	98.00594300669456	11.438837008078891
0450VDE01000	0	98.24607603767004	11.449130124470399
0450VDE01200	0	99.86916085394209	11.54238969510035

출처: 비아이매트릭스

VDS (Vehicle Detect System)의 급감속 구간 탐색

중부내륙선 35일치 데이터: 약 5천 1백만 건 분석

K-Means 분류 결과 (K=5)

시간대와 평균속도 2개 축을 이용하여 5개의 중심으로 분류

- 새벽시간대 (3시, 5시경) / 오전 (11시경) 및 오후 (17시, 18시경)의 5개 클래스로 구분되며, 1일치 분석과 유사

Console Output:

```

1 k-means clustering with 5 clusters of sizes 2124, 2844, 2812, 1547, 1329
2 Cluster means:
3   col_hr avg_speed
4 1 11.748469 112.00111
5 2 5.018987 99.00991
6 3 17.703770 100.62176
7 4 18.276018 89.80198
8 5 3.966149 86.33693

```

Data Table:

vds_id	col_hr	avg_speed	CLUSTERED
0450VDE00100	0	89.3846153846	5
0450VDE00200	0	93.11157455680001	2
0450VDE00300	0	100.659474039	2

Options Panel:

```

options
main_dataset r
variables col_hr,avg_speed
k 5
method euclidean
n_start 5
max_iteration 100
cluster_variable_name CLUSTERED
class_variables []
parallel_degree auto

```

Centers Table:

centers.col_hr	centers.avg_speed	CLUSTERED
11.7434086629002	112.001106754172	1
5.01898734177215	99.0099102699929	2
17.7037695590527	100.621760654844	3
18.2760180995475	89.8019831068518	4
3.96613995485327	86.33693001354	5

Size Table:

size	CLUSTERED
2124	1
2844	2
2812	3
1547	4
1329	5

출처: 비아이매트릭스

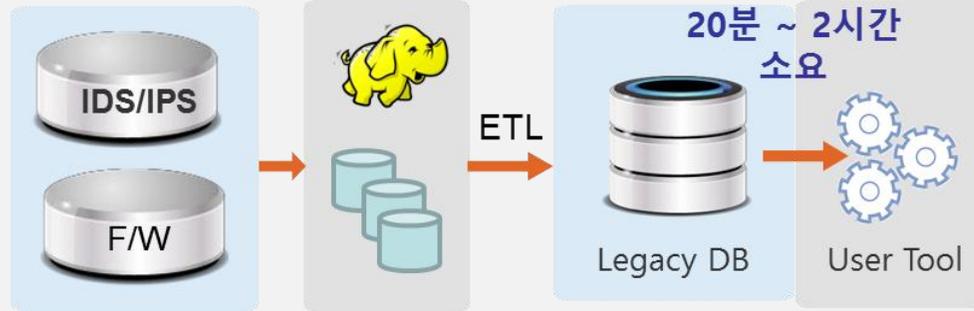
> Case Study

■ 빅데이터 분석시스템: 보안 빅데이터 분야

침해사고대응 보안관제

As-Is

- ▶ 그룹 내, 1,800 대 보안시스템 으로부터 보안 로그 수집
- ▶ 수집 에이전트와 하둡을 이용한 로그 수집 후 ETL 기능을 통한 분석 결과를 Legacy DB에 저장



사업명

침해사고 대응 관제
기술용 DBMS 공급 건

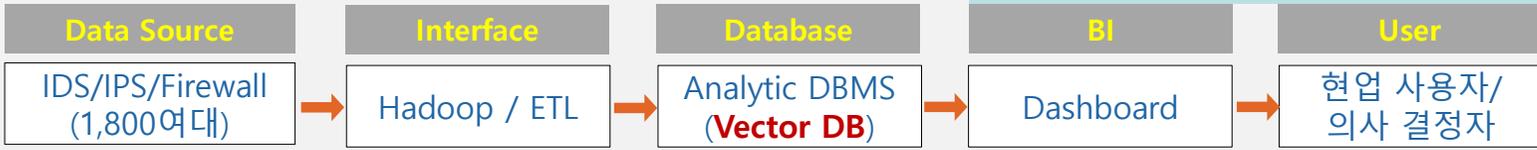
분석시스템 규모

- 동시 사용자 : 40여명
- 데이터 용량 : 3TB
- Big Table 자료 : 6.5억 건

프로젝트 기간

2014.12. ~ 2015.05. (6 개월)

To-Be



프로젝트 투입 인력

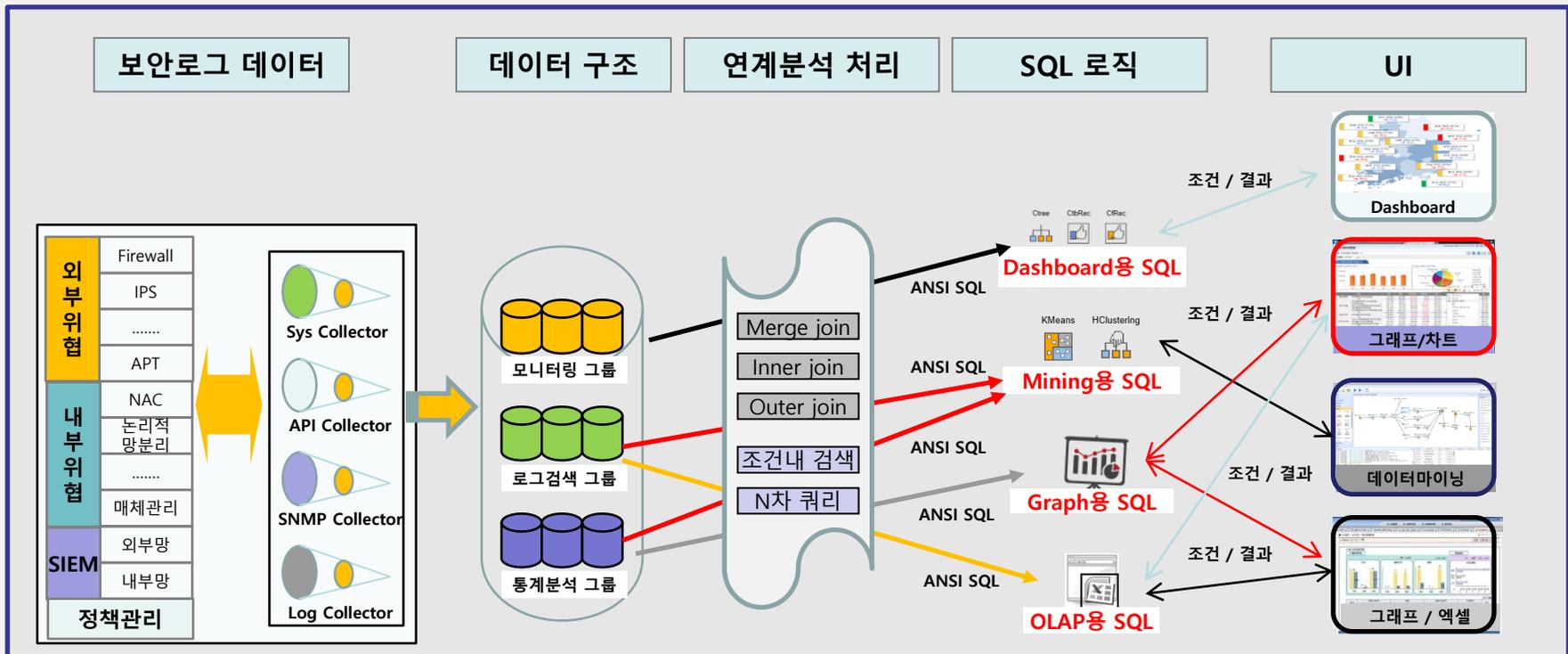
- Vector 지원 : 3MM
- Application 개발 : 6MM

솔루션 도입 효과

- 통합보안로그 분석 시, 빠른 검색 시간 제공 (Legacy DB사용 시 20분 이상 소요 → 초 단위로 결과 확인)
- 보안 문제 발생 시, 침입탐지 이상유무 레포팅 시간이 2시간 이상 소요되던 업무가 5분 이하로 유지
- 보안관제 담당자 40여명이 동시에 검색을 실시하는 경우에도, 쿼리 성능에 빠른 검색 시간 유지

보안로그 및 정책 데이터 연계분석 - 데이터 분석 구조를 명확히

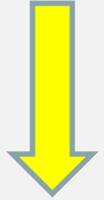
- ✓ 빅데이터 및 IoT 환경에서 진화하는 통합 보안 시스템 및 보안 정책 필요
- ✓ 보안 시스템 모니터링, Risk 감소, 실시가 대응의 중요성이 지속적으로 부각
- ✓ 저장용 데이터베이스에서 분석 전문 데이터베이스로의 진화
- ✓ 보안 및 정책로그의 보안 룰엔진(Rule Engine)과 시나리오의 자동화 및 대량 데이터에 대한 유연성 향상



빅데이터 분석 및 투자를 위한 시스템 규모 산정- 예시

동시사용 개수	사용한 CORE 수	CPU 사용률(%)	DISK IO (MB)	Processing Memory 사용량	수행시간 (초)	32 코아 대비 수행시간 증 가율	설명
1	8	800%	0	0.4GB	98초	350%	32코아 대비 3.5배 소요
	16	1,600%	0	0.4GB	52초	185%	상동
	32	3,200%	0	0.4GB	28초	0%	-
5	8	800%	0	0.55GB	491초	260%	2.6배
	16	1,600%	0	0.57GB	253초	136%	1.3배
	32	3,200%	0	0.57GB	186초	0%	-
10	8	800%	0	0.63GB	983초	358%	3.6배
	16	1,600%	0	0.65GB	502초	183%	1.8배
	32	3,200%	0	0.71GB	274초	0%	-

✓ 물리적 컴퓨터 파워가
2배 증가할 경우,
분석 소요시간이 50% 감소



✓ 데이터 저장장치의 속도가
2배 증가할 경우,
분석 소요시간이 20% 감소



1) CPU CORE 개수 별 SQL 수행 시간 비교

- 물리적 코어 수가 2배 증가할 때, SQL 수행 시간이 1/2로 단축되었으며 ((98-52)/98 = 0.46, 약 47%의 시간 감소), 물리적 코어 수가 4배 증가할 때, SQL 수행 시간이 1/4로 단축되었음 ((98-28)/98 = 0.71, 약 71%의 시간 감소)

2) DISK I/O별 SQL 수행 시간 비교

- DISK IO가 7.5배 증가할 때, SQL 수행시간은 413초에서 54초로 변화하므로, (413-54)/413 = 0.86, 약 86%의 시간 감소가 발생함
그러므로 DISK IO가 2배 증가할 때, 수행시간은 약 20%의 감소가 있을 것으로 추정함

> 4. 분석업무를 위한 지식과 기술

빅데이터 분석 업무 관련 학습 모듈

빅데이터 수집

빅데이터 저장

빅데이터 처리

데이터 탐색

통계기반 분석

ML 기반 정형분석

ML 기반 비정형분석

분석결과 시각화

빅데이터 수집	1. 데이터 수집 계획 수립하기	2. 빅데이터 수집 시스템 구성하기	3. 내부 데이터 수집하기	4. 외부 데이터 수집하기
	5. 데이터 변환하기	6. 수집 데이터 검증하기		
빅데이터 저장	1. 빅데이터 저장 계획 수립하기	2. 빅데이터 저장 모델 설계하기	3. 빅데이터 저장 관리시스템 구성하기	4. 빅데이터 적재하기
	5. 빅데이터 운영하기			
빅데이터 처리	1. 빅데이터 처리 계획 수립하기	2. 빅데이터 처리 시스템 구성하기	3. 분산병렬 수행하기	4. 실시간 수행하기
데이터 탐색	1. 기본통계 확인하기	2. 데이터 분포 분석하기	3. 변수간 관계 확인하기	4. 데이터 정제하기
	5. 데이터 변화 적재하기	6. 분석용 데이터 지표도출하기		
통계기반 데이터분석	1. 가설 설정하기	2. 빅데이터 모델 개발하기	3. 빅데이터 모델 평가 검증하기	4. 빅데이터 모델 운영방안 마련하기
머신러닝 기반 정형 데이터 분석	1. 머신러닝 수행방법 계획하기	2. 데이터셋 분할하기	3. 지도학습 모델적용하기	4. 자율학습 모델 적용하기
	5. 모델 성능 평가하기	6. 학습결과 적용하기		
머신러닝 기반 비정형 데이터 분석	1. 텍스트마이닝 기반 텍스트 분석 실행방법 계획하기	2. 텍스트 변환하기	3. 단어사전 구축하기	4. 텍스트 분류 결과 분석하기
	5. 정형 데이터 결합 분석 수행하기	6. 사회연결망 분석하기	7. 영상 인식하기	8. 음성 인식하기
빅데이터 분석 결과 시각화	1. 분석 결과 스토리텔링하기	2. 분석 정보 시각화하기	3. 분석 정보 시각표현하기	

출처: 국가직무능력표준, 2017

> 4. 분석업무를 위한 지식과 기술

빅데이터 수집 - 수집시스템 구성을 위한 지식과 기술 (예)



빅데이터 수집시스템 구성을 위한 필요 지식(예시)

- 빅데이터 수집 방침 및 목표
- 수집 대상 데이터 특성
- 데이터 유형별 수집 기술 종류
- 현행 개선 요구, 리스크, 이슈, 제약 사항 식별
- 개인정보보호 및 권한 관리
- 빅데이터 분석 대상 업무 도메인
- NoSQL (Not Only SQL) 논리적 구조
- 분산파일 시스템 논리적 구조
- 관계형 데이터베이스 논리적 구조
- 데이터 분석 및 설계 방법론
- 배치 처리 방법론

빅데이터 수집시스템 구성을 위한 필요 기술(예시)

- 데이터 유형별 데이터 수집 기술
- 데이터 수집 요건 분석 및 설계 능력
- 정보 시스템 및 데이터 인터페이스 설계 기술
- 고 가용성/부하 분산 방식의 데이터 전송 채널 구축 기술
- 수집 어댑터 및 플러그인 개발 능력
- 수집 시스템 구성 요소들의 구동/설정 배포/상태 모니터링 등 운영을 위한 관리 도구 구축 기술
- 데이터 필터링, 정제 기능 구축 능력
- 반정형 데이터 저장 NoSQL (Not Only SQL) 용량 산정 능력
- 정형 데이터 저장 관계형 데이터베이스 용량 산정 능력
- 관계형 데이터베이스 테이블 모델링 설계 및 구축 기술

출처: 국가직무능력표준, 2017

> 4. 분석업무를 위한 지식과 기술

빅데이터 저장 - 빅데이터 저장모델 설계하기를 위한 지식과 기술 (예)



빅데이터 저장모델 설계하기를 위한 필요 지식(예시)

- 관계형 데이터 모델링 기법
- 데이터웨어하우스 설계 기법
- 텍스트 저장모델 설계 기법
- 데이터 모델 표현 기법
- NoSQL (Not Only SQL) 설계 기법
- EDA (Exploratory Data Analysis) 기법
- 머신러닝 기법
- 텍스트 마이닝 기법

빅데이터 저장모델 설계하기를 위한 필요 기술(예시)

- 데이터 요구사항 분석능력
- 커뮤니케이션을 통한 요구사항 도출 능력
- 비즈니스 도메인 분석기술
- 데이터 요구사항과 기능 요구사항 연계 능력
- 데이터 저장모듈 설계도구 활용 능력

출처: 국가직무능력표준, 2017

> 4. 분석업무를 위한 지식과 기술

빅데이터 처리 - 빅데이터 처리시스템 구성 하기를 위한 지식과 기술 (예)



빅데이터 처리시스템 구성하기를 위한 필요 지식(예시)

- 빅데이터 처리 플랫폼
- 빅데이터 처리 인프라스트럭처
- 클라우드 IaaS 서비스와 클라우드 솔루션
- 클러스터 데이터베이스
- 빅데이터 처리 플랫폼 검증 방법
- 빅데이터 처리 개선 사항 도출

빅데이터 처리시스템 구성하기를 위한 필요 기술(예시)

- 빅데이터 처리 플랫폼 분석 기술
- 빅데이터 처리 플랫폼 설치 및 설정 능력
- 클라우드 인프라 구축 및 설정 능력
- 분산 병렬 컴퓨팅 솔루션 설치 및 설정 능력
- 오픈소스 소프트웨어 사용 능력

출처: 국가직무능력표준, 2017

> 4. 분석업무를 위한 지식과 기술

통계기반 데이터 분석 - 빅데이터 모델 운영방안 마련하기 위한 지식과 기술 (예)



빅데이터 모델 운영방안 마련하기 위한 필요 지식(예시)

- 빅데이터 분석 방법론
- 정보시스템 구축 방법론
- 정보시스템 생명주기
- 분석 모델 생명주기

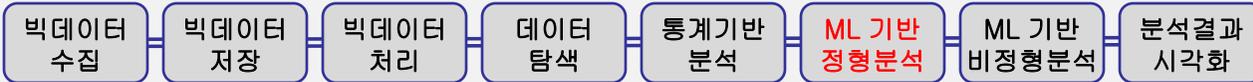
빅데이터 모델 운영방안 마련하기 위한 필요 기술(예시)

- 모델 알고리즘 설명서 작성 기술
- 시스템 설계 작성 기술
- UML(Unified Modeling Language) 기술
- SQL(Structured Query Language) 기술
- 정보시스템 개발언어 활용 기술

출처: 국가직무능력표준, 2017

> 4. 분석업무를 위한 지식과 기술

머신러닝 기반 데이터 분석 - 지도학습 모델 적용하기를 위한 지식과 기술 (예)



지도학습 모델 적용하기를 위한 필요 지식(예시)

- 현업 업무 및 수행 프로세스
- 통계학 기반의 분석방법론
- 선형대수 및 미적분, 최적화 등의 수학기론
- 분류 모델을 위한 지도학습 머신러닝 기법
- 예측 모델을 위한 지도학습 머신러닝 기법
- 최신 머신러닝 기법 발전 트렌드와 연구 동향

지도학습 모델 적용하기를 위한 필요 기술(예시)

- 문제 핵심 파악 및 이슈 도출 능력
- 프로젝트 기획 및 관리 능력
- 문제 및 이슈를 실제로 검증 가능한 일련의 명확한 가설로 만들어 낼 수 있는 능력
- 머신러닝 알고리즘을 실제로 구현해낼 수 있는 프로그래밍 기술
- 통계 및 데이터 분석에 특화된 언어를 다룰 수 있는 기술
- 기존의 알고리즘을 융합하거나 새로운 알고리즘을 설계하고 구현할 수 있는 능력
- 자동화 처리 및 시스템 구현 위한 개발 능력

> 4. 분석업무를 위한 지식과 기술

빅데이터 분석 결과 시각화 – 분석정보 시각화하기 위한 지식과 기술 (예)



분석정보 시각화하기 위한 필요 지식(예시)

- 시각화 기술 및 알고리즘
- 데이터 분석 대상 업무
- 시각화 정보 구조 및 구성 요소
- 시각화 구축의 목적과 사용 의도
- 시각화 시스템 작동 방식
- 화면구성 특징에 따른 시각화 기법
- 시각화 표현 기법
- UX(User Experience) 설계 방법론
- 효과적인 컬러 및 폰트 사용법
- 시각화 디자인 원칙

분석정보 시각화하기 위한 필요 기술(예시)

- 시각화 도구 및 기술 선정 능력
- 시각화 화면 구성요소 별 특징 분석 능력
- 시각화 스토리보드 기반 시각화 구성 요소 도출 능력
- 시각화 플랫폼 구현 기술
- 시각화 표현 방안/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

출처: 국가직무능력표준, 2017

> 4. 분석업무를 위한 지식과 기술

효과적인 분석시스템 구축을 위한 성공 요소

지식 습득에 대한 갈망

- 논리학, 심리학, 통계학, 경영학, 컴퓨터 공학 등
- 관련된 학문의 동향 주시하기
- 융합을 위한 필요한 지식을 탐구하기

폭넓은 기술 활용

- 컴퓨터 활용, 프로그래밍, 분석방법론 등
- 빅데이터 수집, 저장, 분석, 표현 기술을 습득
- 사전에 두 단계 깊이 있는 기술을 찾기

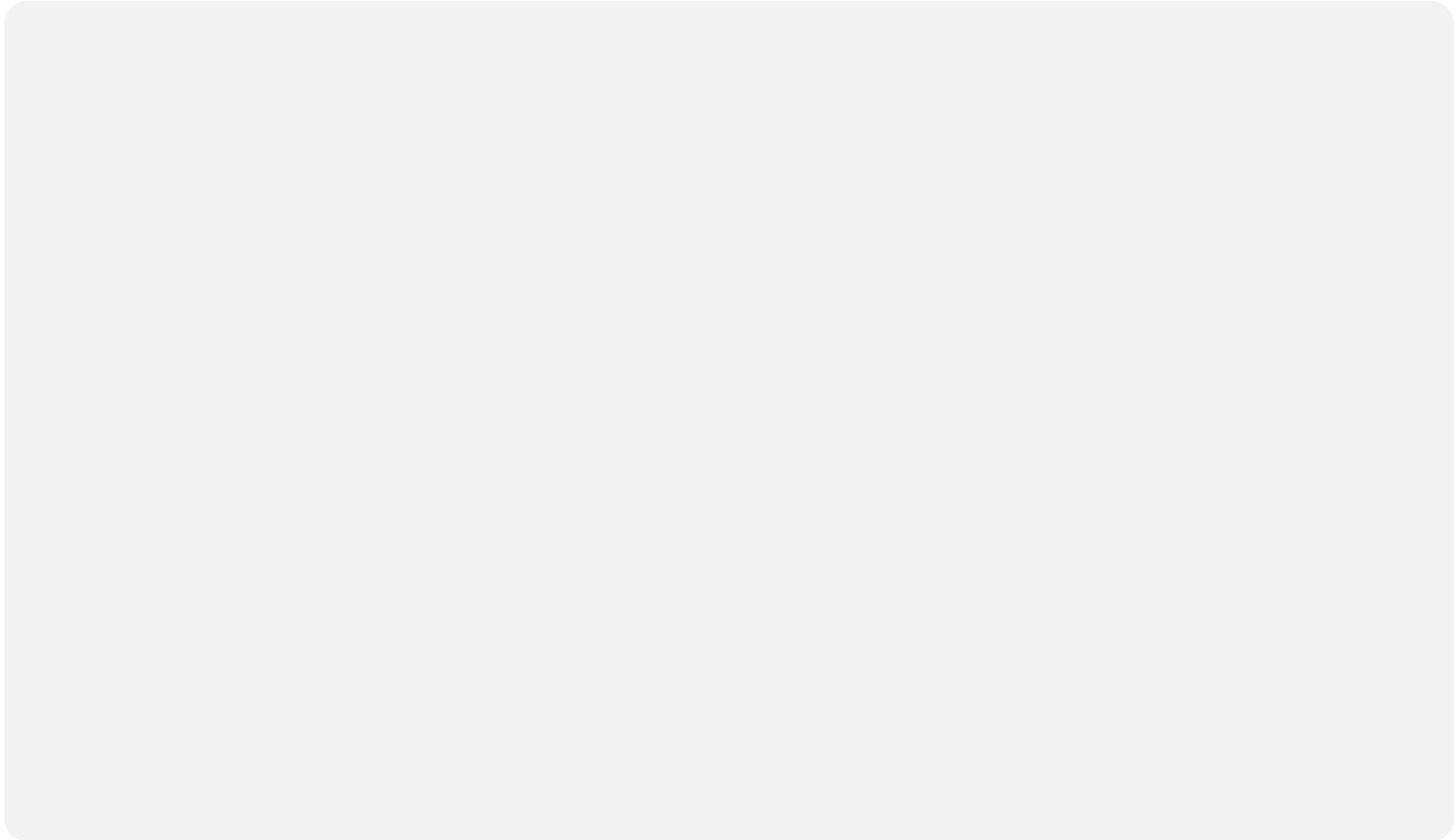
협업을 위한 자신의 태도

- 지식과 기술을 배우려는 자세가 필요함
- 함께 일하는 사람이 서로가 다름을 인정하기
- 어떤 일이든지 프로페셔널이 되기에 오래 걸림
(1만 시간의 법칙이 중요함)

도달 가능한 능력

- 무엇을 잘하는가? 정말 인정받을 수 있는가?
- 개인의 능력과 협업의 생산성 및 조화를 존중하기
- 멘토를 통한 자신의 서로 다른 능력 키우기
(다양한 방법을 통한 최고의 전문가 찾기)

> Q & A



감사합니다.



eGlobal
Global Technology, Make IT easy

서울시 강남구 선릉로 93길 54, 9층 (역삼동, 일환빌딩)

(주)이글로벌시스템

Tel : 82-2-6447-6988, Fax : 82-2-6447-6989,

E-Mail : sales@eglobalsys.co.kr (www.eglobalsys.co.kr)