
빅데이터의 개념과 민간/공공의 활용사례

한국정보화진흥원 미래전략센터

이영주 수석연구원/정보시스템박사

CONTENTS

01 빅데이터 개요

02 공공부문 빅데이터 활용 사례

03 민간부문 빅데이터 활용 사례

04 빅데이터 시대 진로와 경력 관리

01 빅데이터 개요

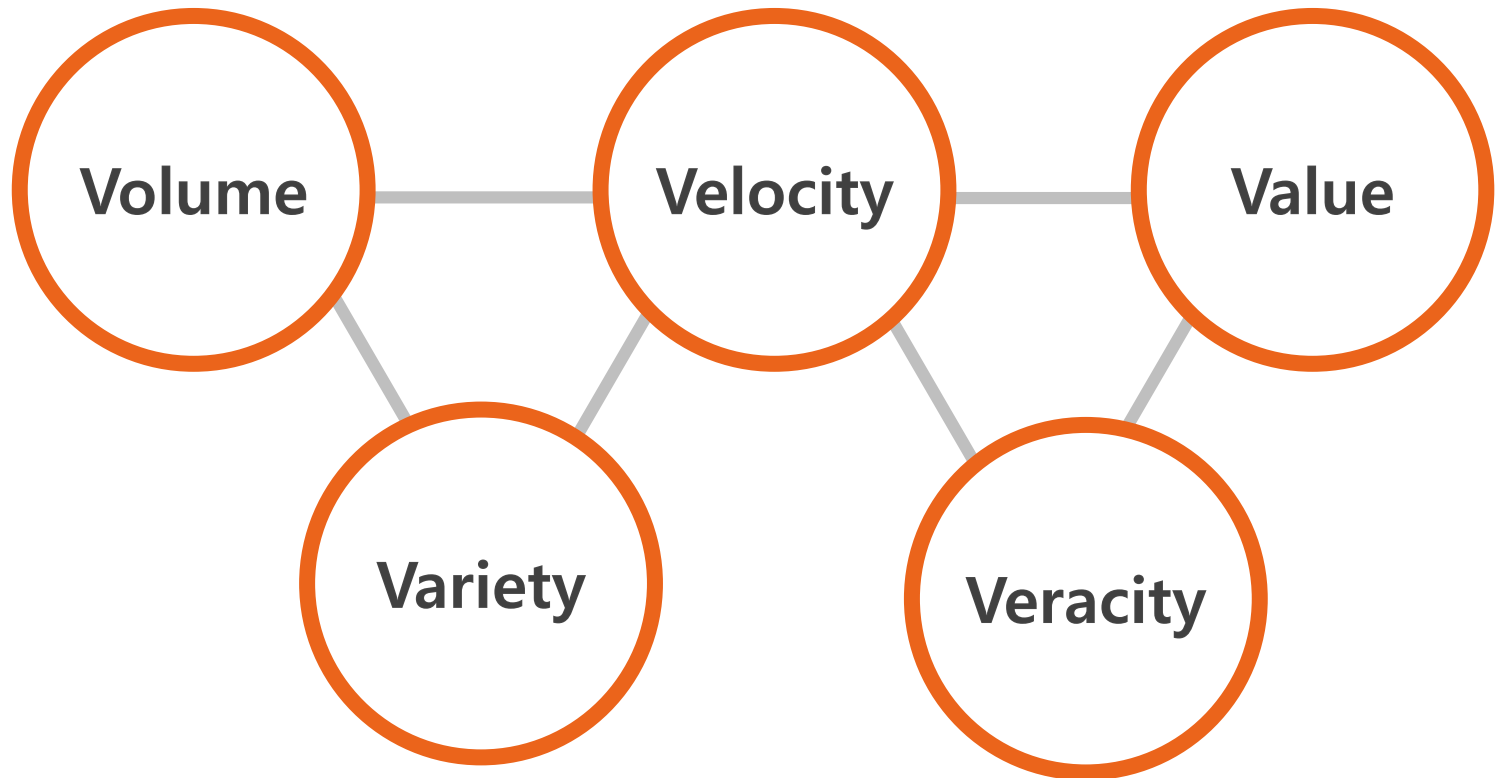


빅데이터란?

“기존 데이터베이스 관리도구의 데이터 수집, 저장, 관리, 분석
역량을 넘어서는 대량의 정형 또는 비정형 데이터 세트
및
이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술”
(맥킨지, 위키피디아 등)

“다양한 종류의 대규모 데이터로부터 저렴한 비용으로 가치를 추출하고,
데이터의 빠른 수집, 발굴, 분석을 지원하도록 고안된
차세대 기술 및 아키텍처”
(IDC)

빅데이터의 특징



빅데이터는 세상을 읽는 수단



※ 인터넷에서 생성되는 소셜미디어 데이터들은 이시간에도 1분에 약 1.3테라바이트(Terabite)이상 증가 중
(출처: <http://pennystocks.la/internet-in-real-time/>)

빅데이터의 등장 배경

✓ 다양한 비정형 데이터 생성

- Social Network Service(SNS)
- 인터넷 서비스의 확산과 웹 로그데이터
- API를 통한 데이터 제공

✓ 새로운 유형의 데이터 증가

- 센서를 통한 상황 · 활동데이터 발생 및 수집 가능
- 각종 전자 기기의 정보 수집 가능

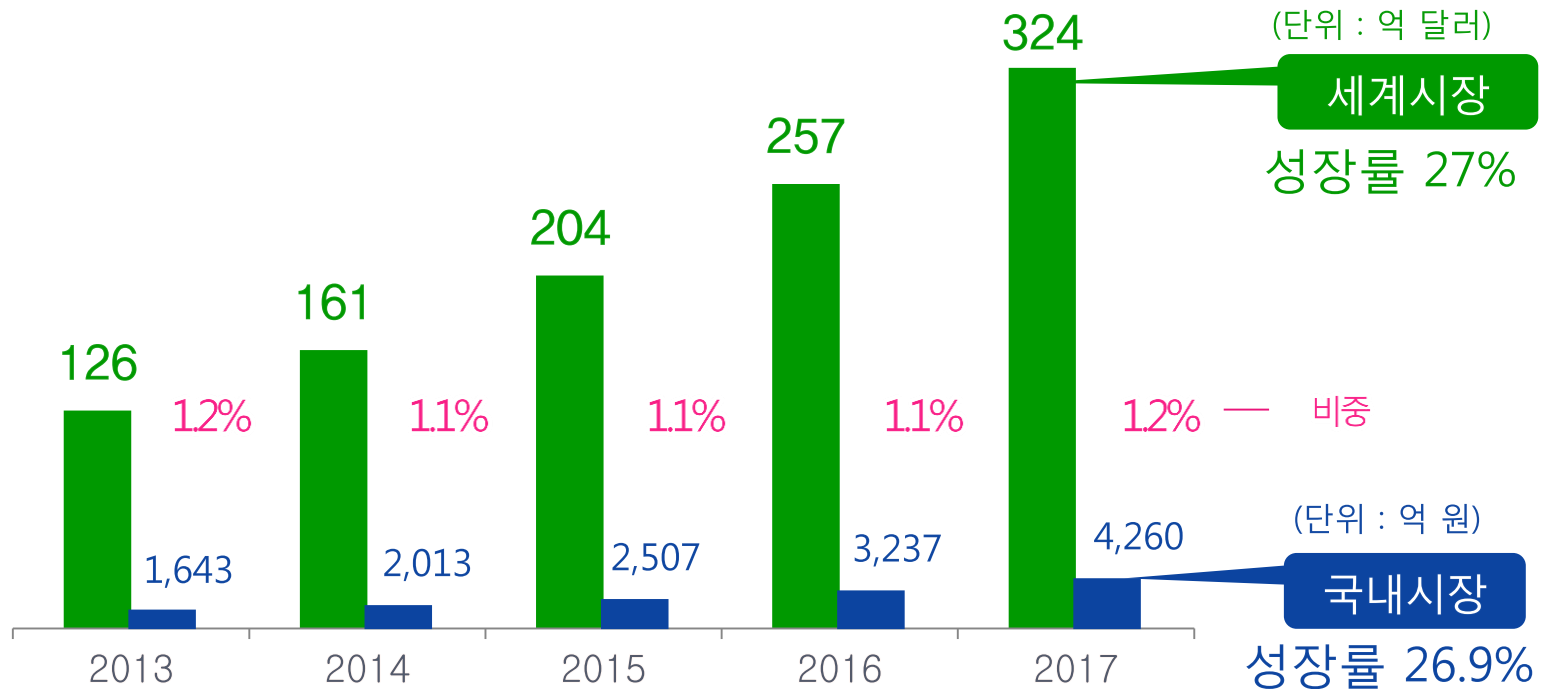
✓ 자동화된 데이터 수집과 처리, 분석 기술의 진화

- 컴퓨터 연산능력 증가, 데이터 저장 비용 감소

빅데이터에 대한 기대

Economist	✓ 데이터는 자본이나 노동력과 거의 동등한 레벨의 경제적 투입 자본, 비즈니스의 새로운 원자재 역할
Gartner	✓ 데이터는 21세기 원유, 데이터가 미래 경쟁 우위를 좌우 ✓ 기업은 다가올 '데이터 경제 시대'를 이해하고 정보 고립(Information Silo)을 경계해야 성공 가능
McKinsey	✓ 빅 데이터는 혁신, 경쟁력, 생산성의 핵심 요소 ✓ 의료, 공공행정 등 5대 분야에서 6천억불 이상 가치 창출
美 대통령 과학기술자문위	✓ 미국 정부기관들이 데이터를 지식으로, 지식을 행동으로 변환하는 전략에 집중해야 함을 주장
싱가포르	✓ 데이터를 기반으로 싱가포르를 위협하는 리스크에 대한 평가와 환경변화를 탐지

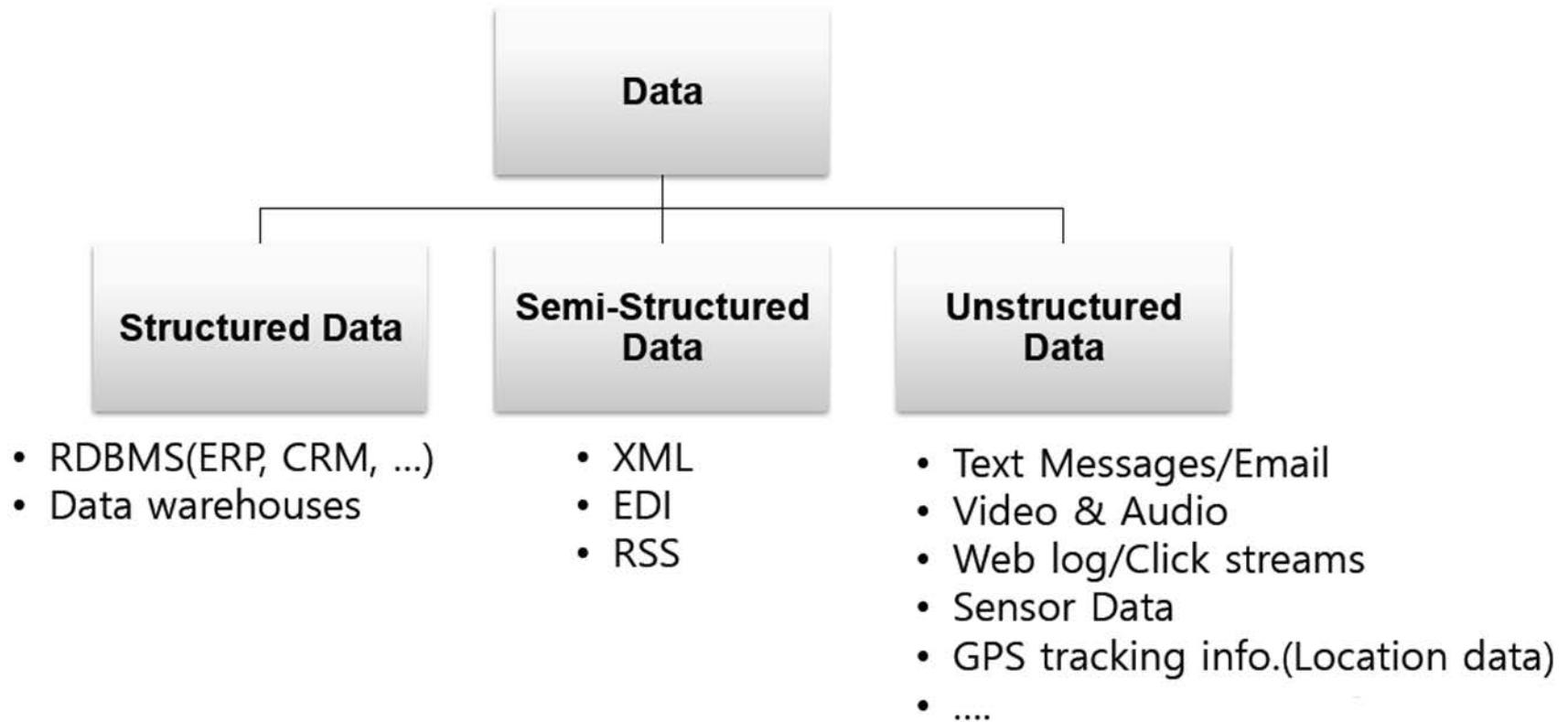
빅데이터 시장 전망



국내시장 출처 : KISTI Market Report 빅데이터 산업의 현황과 전망, April 2013, KISTI

세계시장 출처 : Worldwide Big Data Technology and Services 2013 – 2017 Forecast, December 2013, IDC

빅데이터의 유형



빅데이터의 가치

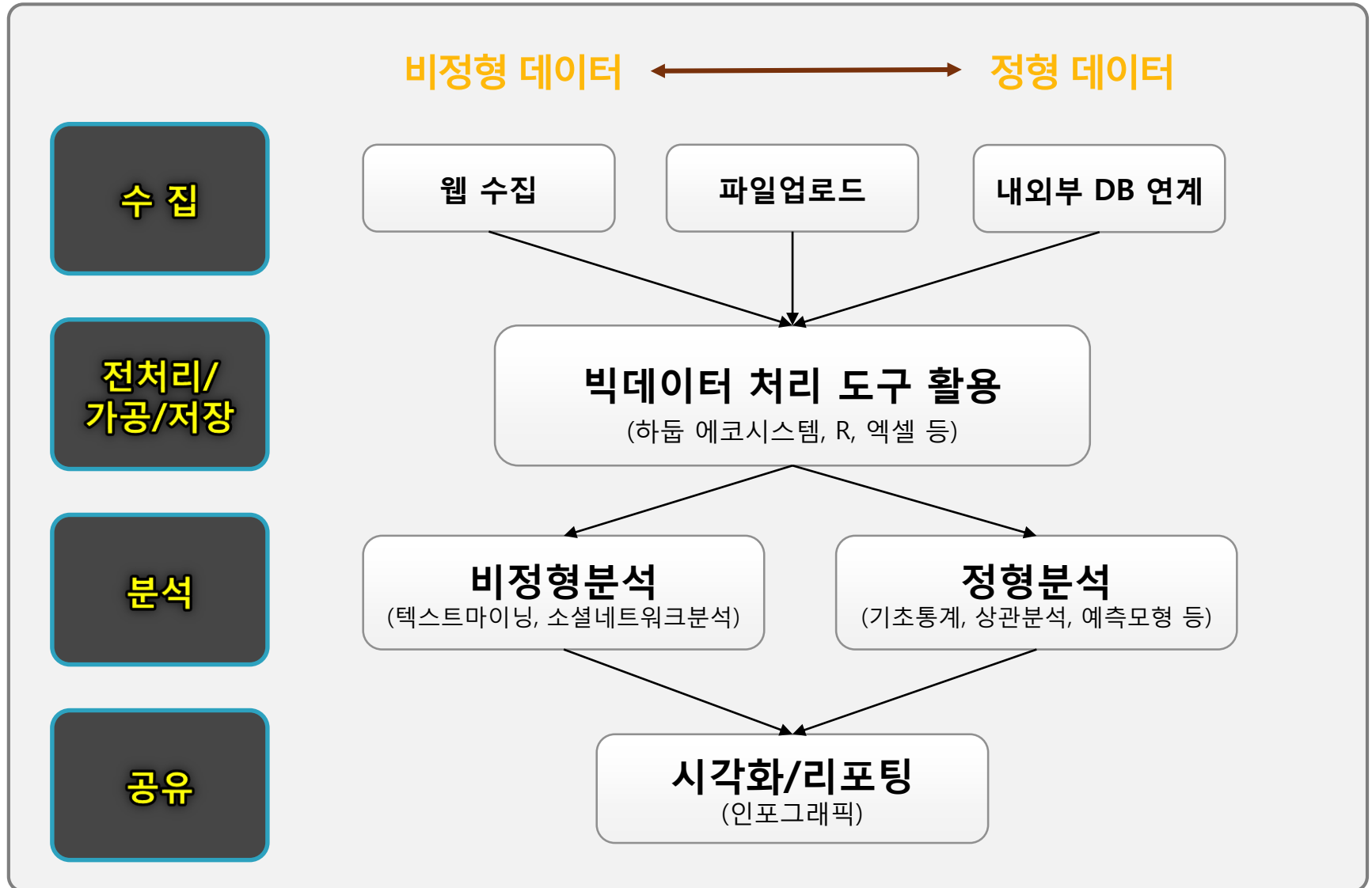
✓ 데이터 비즈니스로의 패러다임 전환

- 인터넷 기업 → 데이터 기업(페이스북, 구글, 트위터)
- 콘텐츠 무료제공/광고수익 → 사용자 데이터 획득, 새로운 비즈니스 창출

✓ 데이터분석 능력은 새로운 가치 창출의 기반

- 구체적이고 실제적인 분석을 통한 매출 증대, 비용 절감
- 데이터 증거에 기반한 객관적인 문제파악, 원인 분석
- 데이터 분석을 통한 실시간 의사결정, 투명한 의사결정
- 사회현안(범죄, 재난재해 등) 해결을 통한 무형의 가치 창출

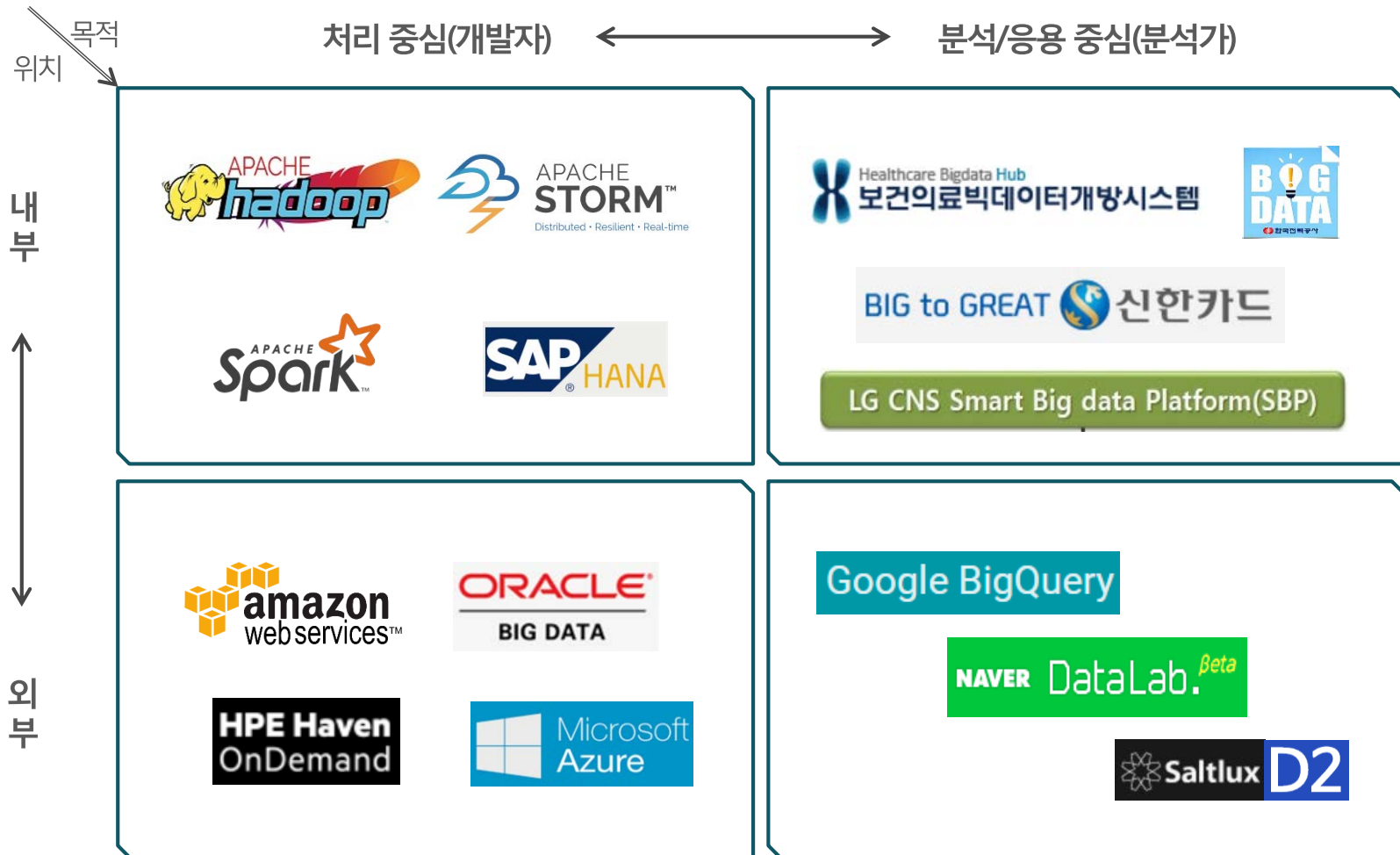
일반적인 빅데이터 처리 절차



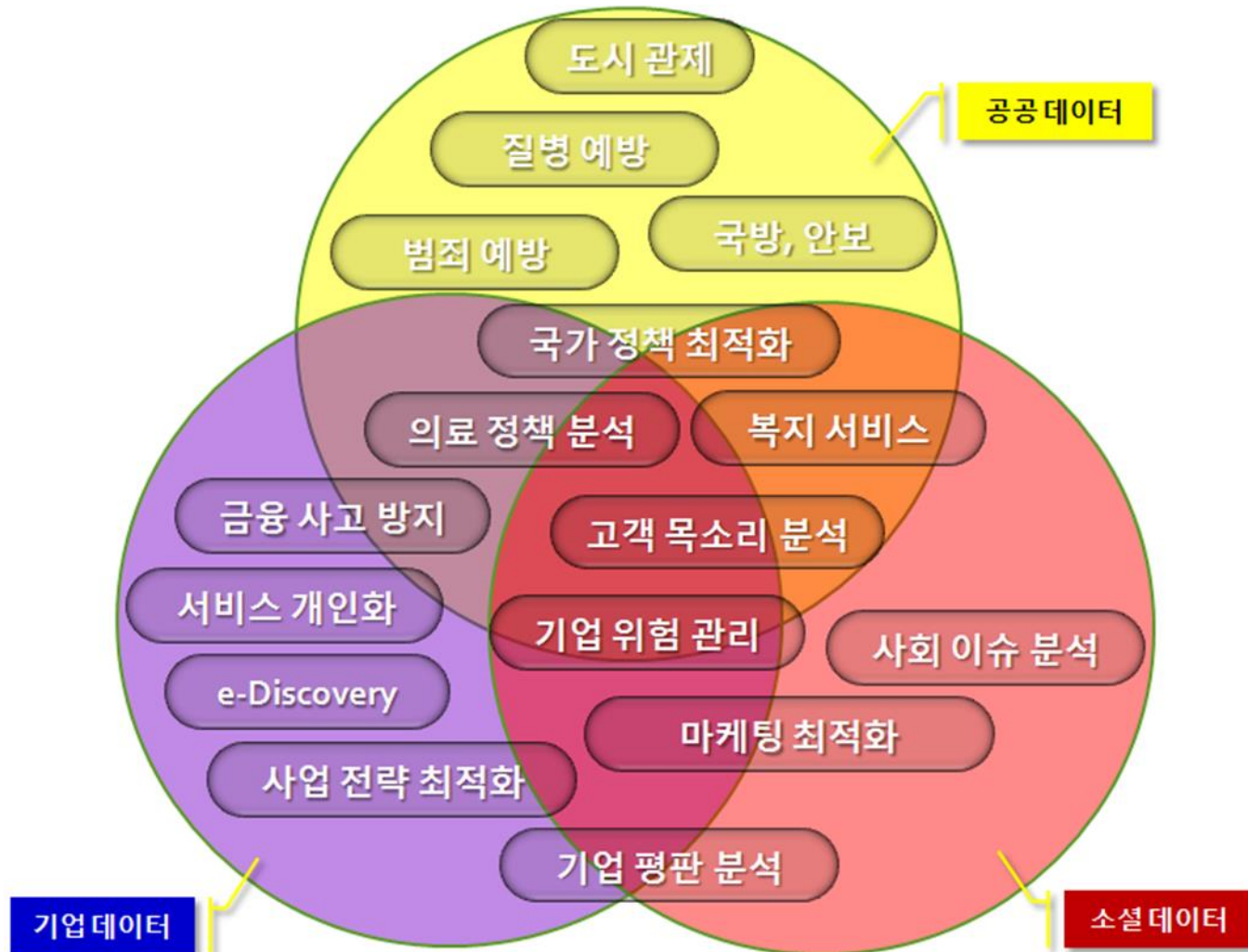
빅데이터 분석 인프라 구성도(예시)



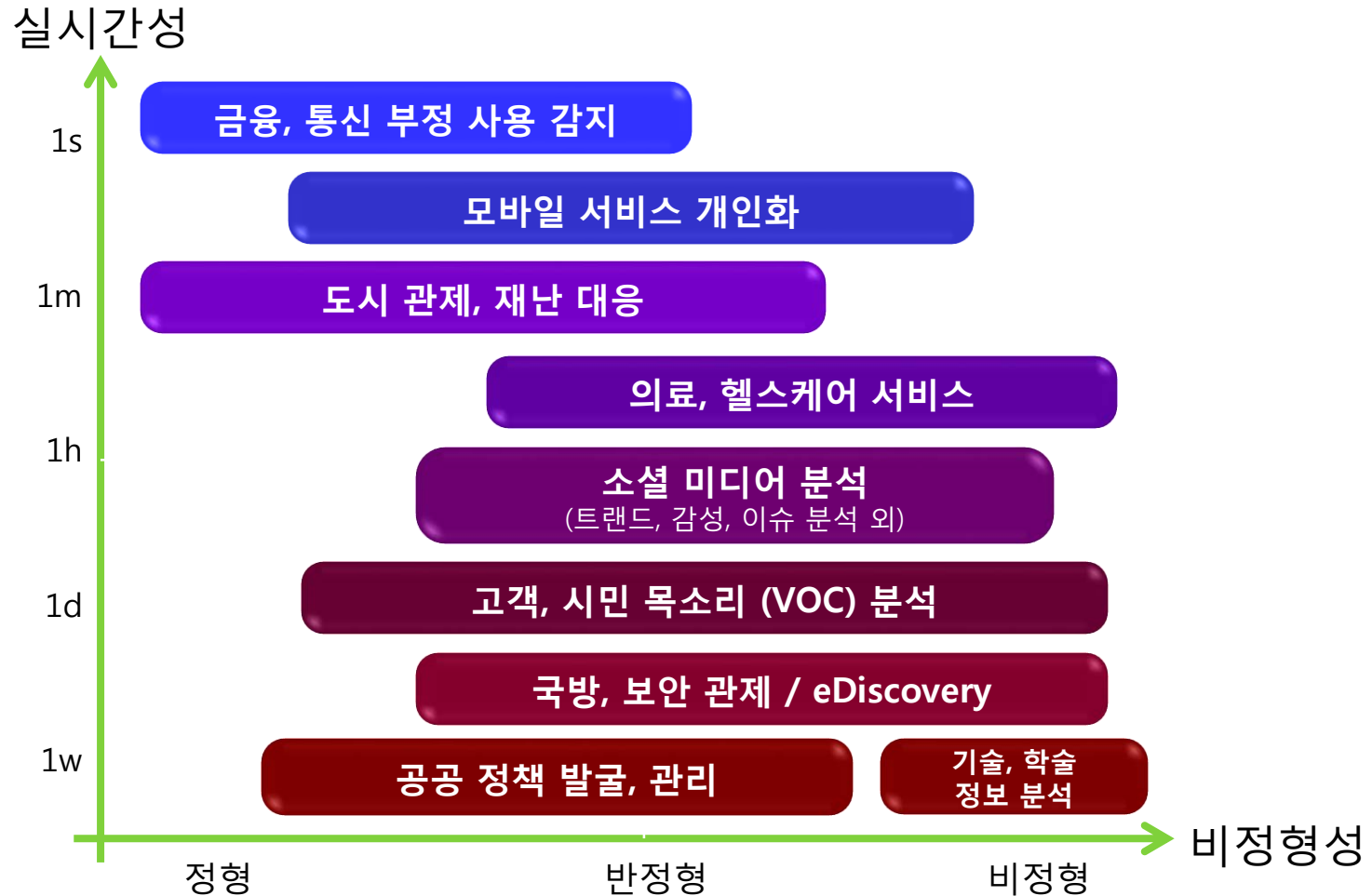
빅데이터 분석플랫폼의 유형



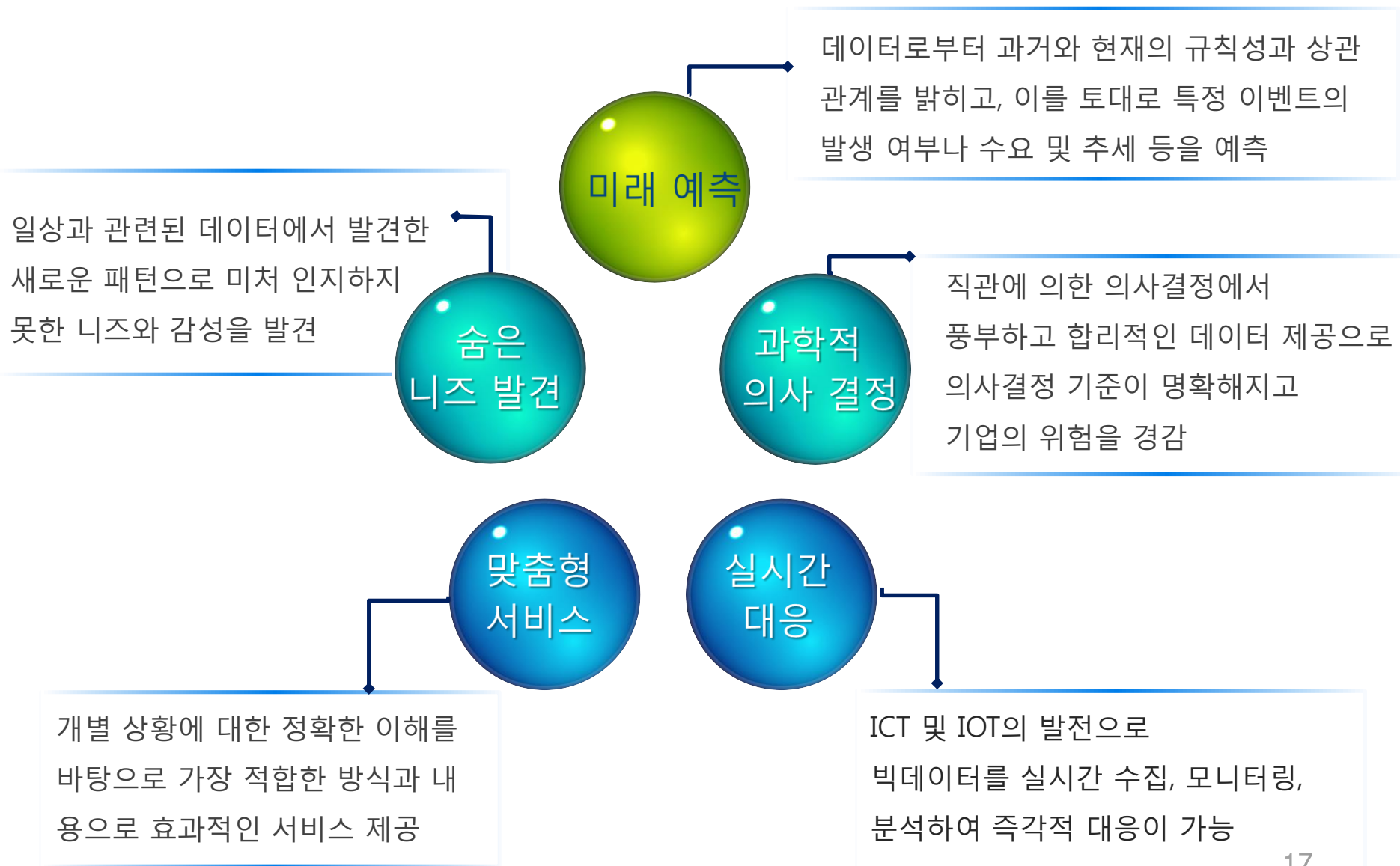
빅데이터 활용 분야 - 출처 관점



빅데이터 활용 분야 - 데이터 특성 관점



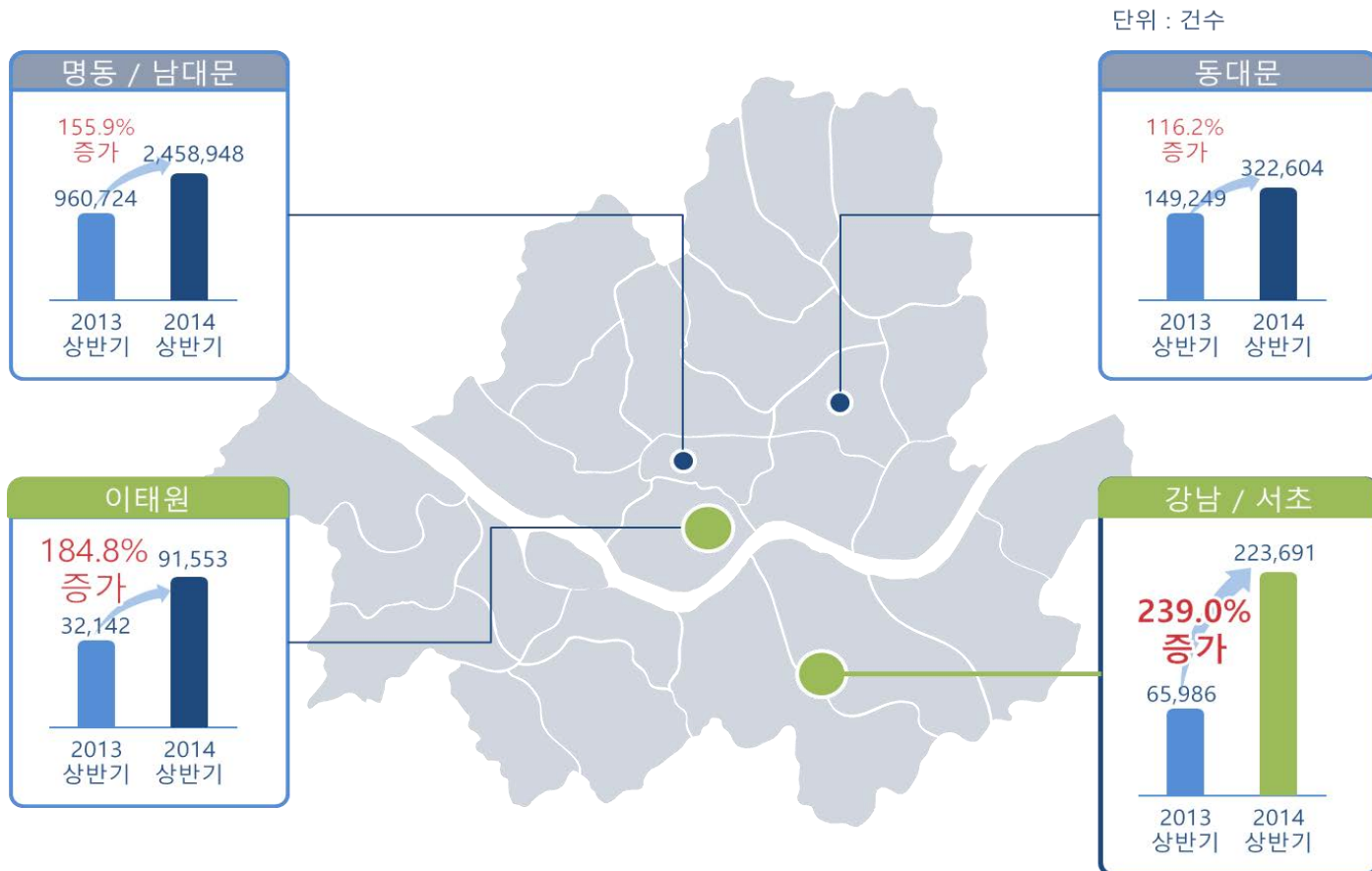
빅데이터 활용 분야 - 목적 관점



02 공공부문 빅데이터 활용 사례

[1] 정책 의사결정

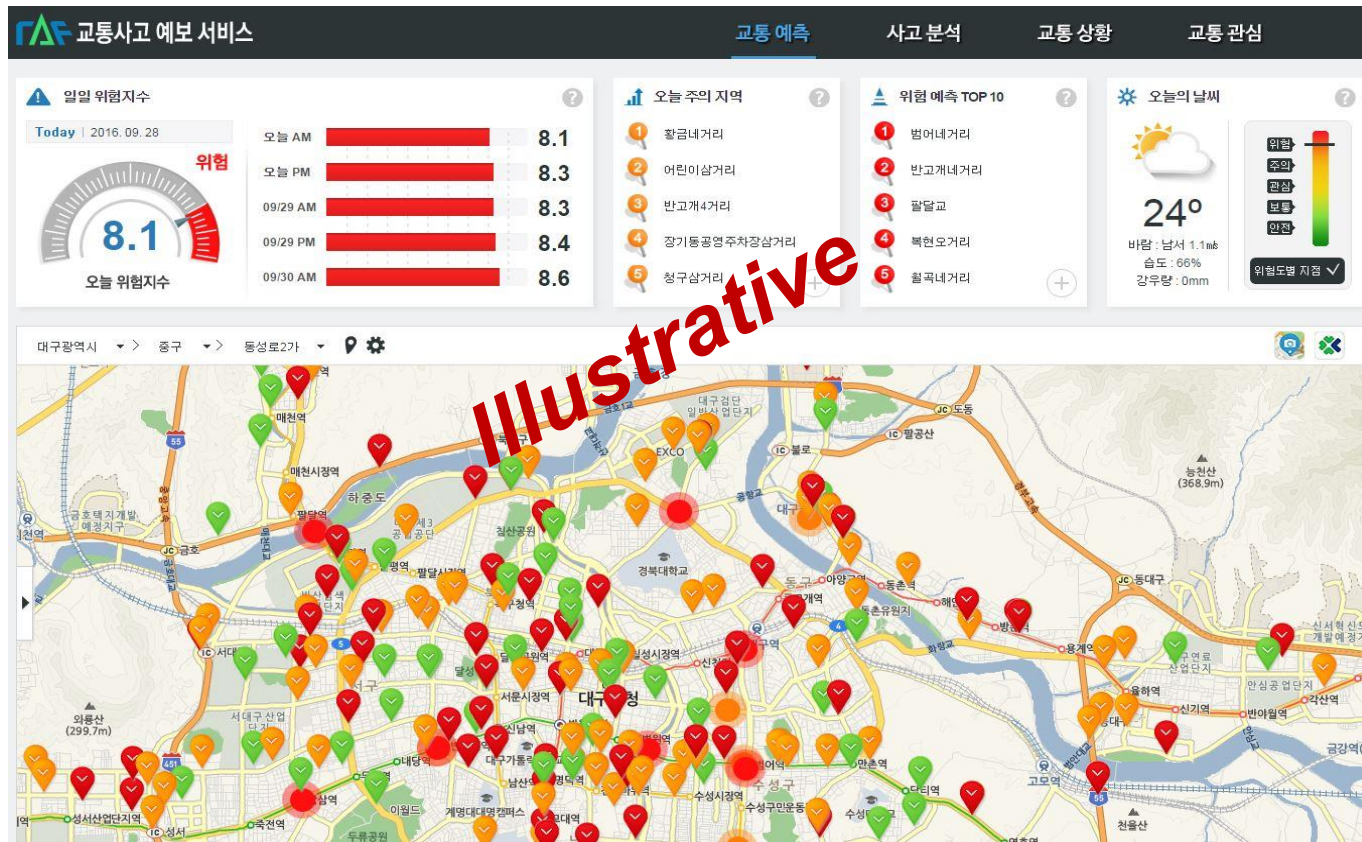
< 중국관광객 신용카드 거래 금액 변화 추이 >



* 자료 : NIA, 2014년 빅데이터 시범과제 추진 결과 중 발췌

[2] 위험 조기경보

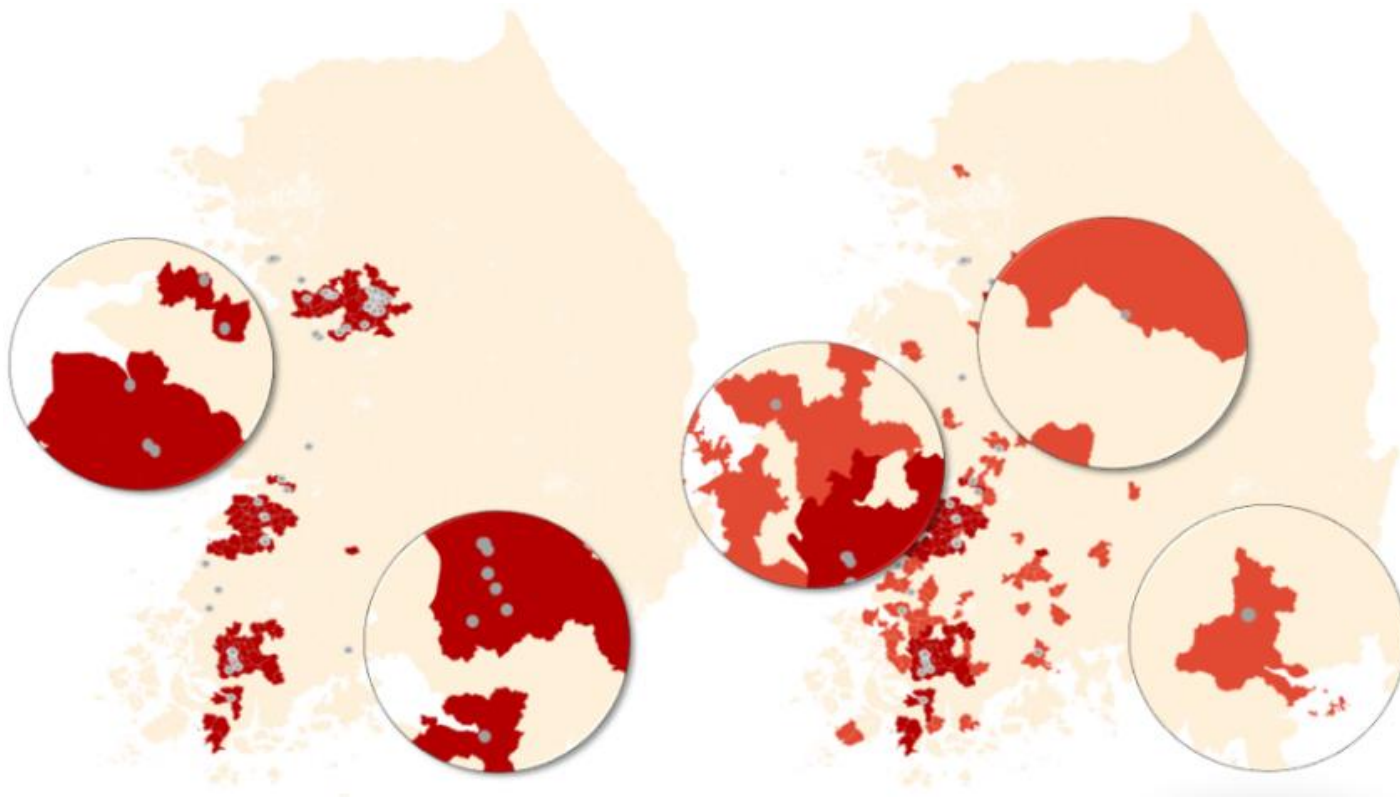
<교통 빅데이터를 결합한 사고위험 예보>



* 자료 : NIA, 2016년 데이터기반 미래전략 시범과제(진행중, 프로토타입)

[3] 재난재해 대응

<조류독감(AI) 발생지역 추적 및 확산 대응>



* 자료 : NIA, 2013년 빅데이터 시범과제 추진 결과 중 발췌

[4] 복지서비스 개선

<경력단절여성 구인구직 미스매치 현황 분석>

전체 경력단절여성

195만



새일센터 이용
적극적 구직자

20만

30대
55.3%



40대 이상
68.3%



전체 경력단절여성

✓ 경기도에 가장 많이
분포됨 29.8%

✓ 30대 경력단절 여성 비율이
가장 높음 55.3%

새일센터 이용 여성

✓ 경기도 새일센터 이용률
전국 최하위 수준임 6.1%

✓ 새일센터 이용자중 40대
이상 비율이 가장 높음
68.3%

✓ 30대 여성은 새일센터
이용률이 저조함 21.7%

vs.

취업지원

활성화 방안 1.

경기도 포함 수도권 거주 30대 고학력 계층을
새일센터 구직자로 유입 확대

- 새일센터 양질의 서비스 제공 (맞춤형 취업지원 교육)
- 고용 여건이 좋은 구인기업 지속적 발굴 -> 재취업 연계

새일센터의 구직자 특성을 고려한
차별화 전략 마련

- 40~50대 고연령층 대상 맞춤형 취업 지원 서비스

출처 : 2013 경력단절여성 통계, 통계청

* 자료 : NIA, 2014년 데이터기반 미래전략 컨설팅 추진 결과 중 발췌

[5] 실시간 환경분석

<신용카드 사용액 기반 소비동향 모니터링>

트렌드 분석

월간지수 | 주간지수 | 주간MMM지수

전체 | 지역별 | 업종별 | 소득분위별 | 가맹점규모별 | 연령대별

구간설정

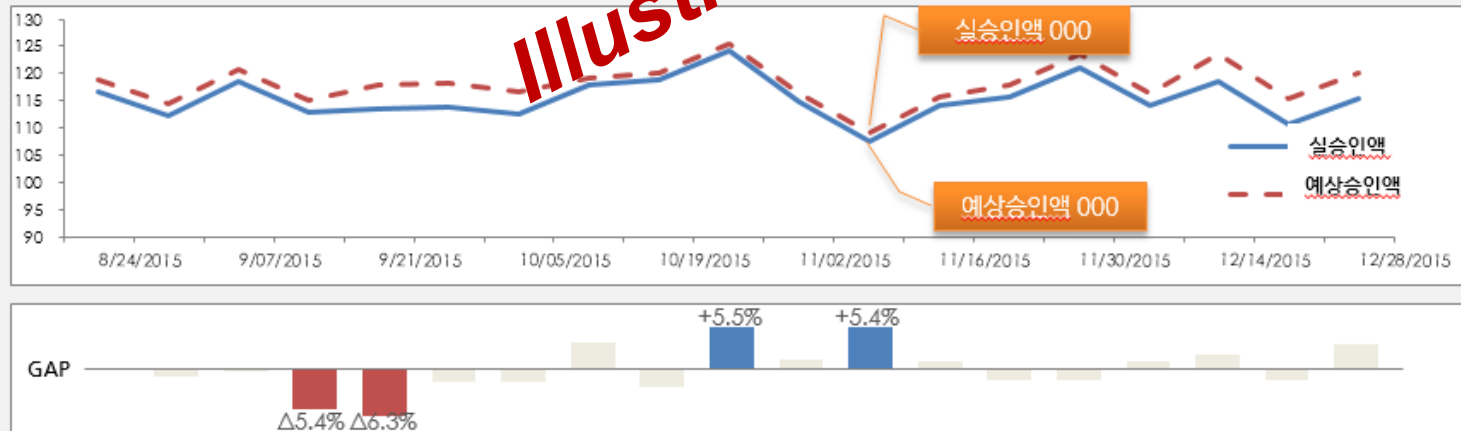
2015.02



2015.08



경상지수

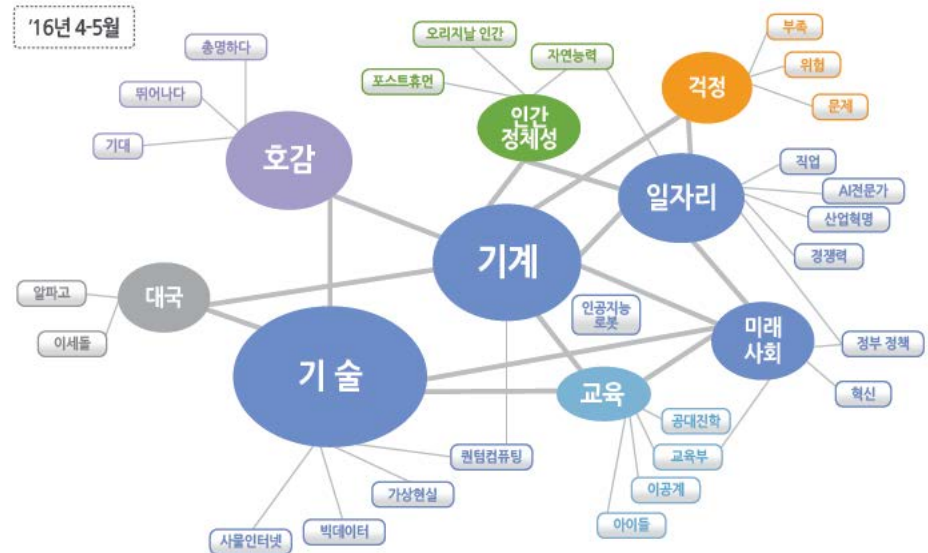
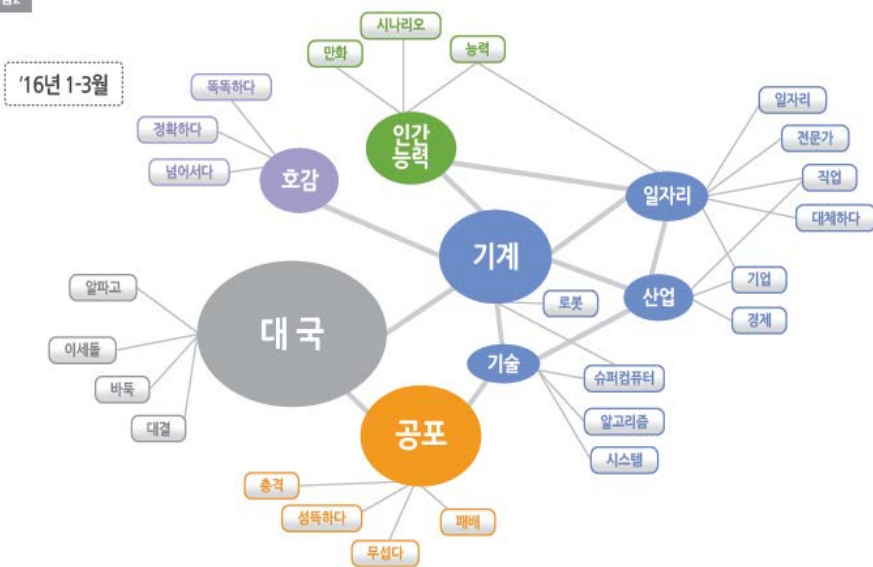


* 자료 : NIA, 2016년 데이터기반 미래전략 시범과제(진행중, 프로토타입)

[6] 트렌드 및 이슈분석

<알파고 쇼크로 본 인공지능과 미래사회 인식>

→전사



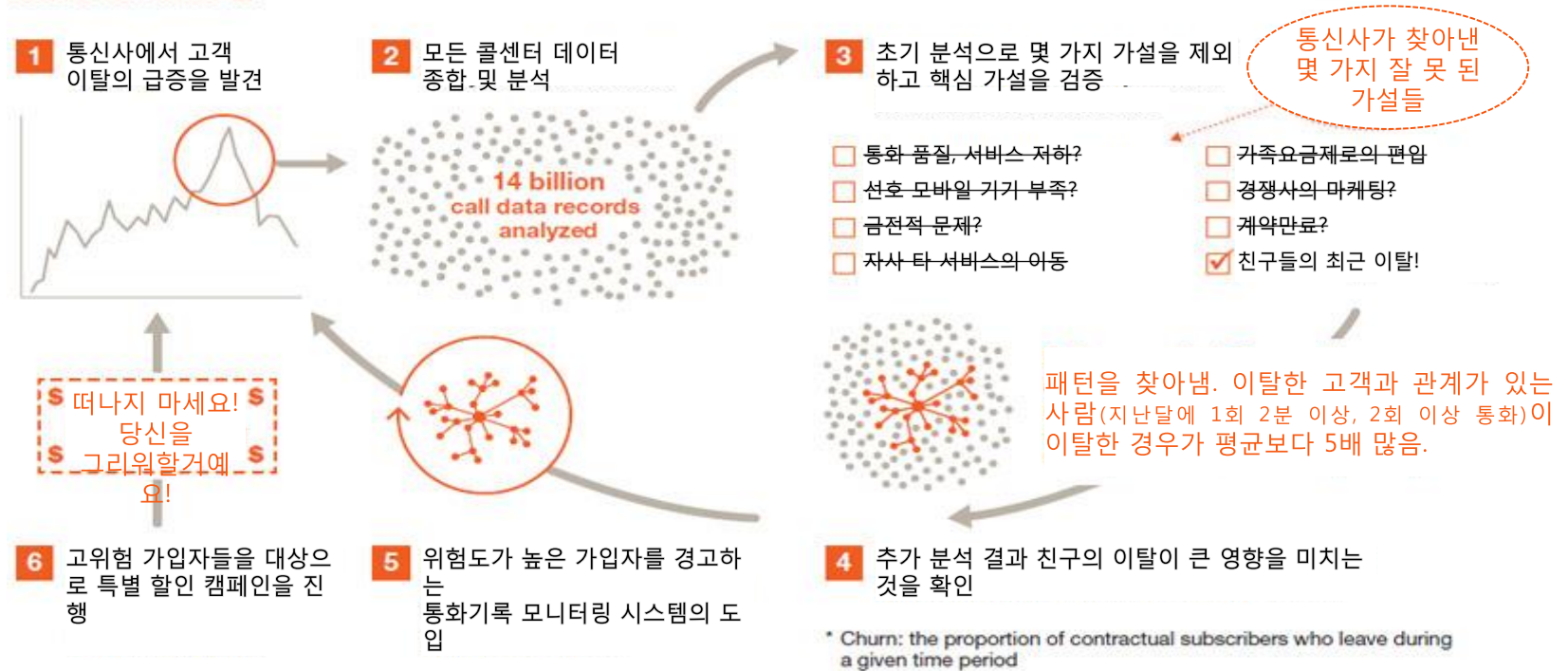
* 자료 : NIA, NEAR & Future Vol. 15, 소셜빅데이터 분석으로 살펴본 인공지능과 미래사회 인식 - 일자리를 중심으로

03 민간부문 빅데이터 활용 사례

[1] 고객 행위 분석

< 이동통신 고객의 번호이동 원인 분석 >

By analyzing billions of call records, carriers are able to obtain early warning of groups of subscribers likely to switch services. Here is how it works:



[2] 문제해결과 비용절감

< 항공기 유지보수를 위한 빅데이터 분석 >



한 대 당 14GB의
정보를 생산하는
1만 5천대 항공기에
대한
분석 진행



분석 플랫폼 가동
20분만에 제트엔진 세척 관련
해결책 생성
(일반적인 분석 방법으로는 한 달 여
소요)



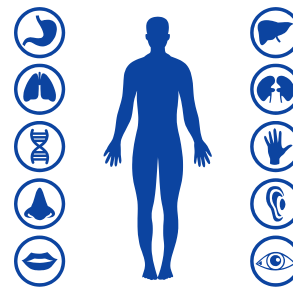
향후, 자사 제트 엔진을 적용한
1천만대 항공기로부터 1천500 TB에
해당하는 데이터를 수집 분석하여
새로운 문제 해결에 활용 계획



유지보수를 효율화를 위해 빅데이터 분석 기술을 도입하여
부품의 사용패턴 파악 및 마모 예상

[3] 전략 수립

< 빅데이터를 활용한 축구경기 전략/전술 수립 >



선수들이 훈련 시 몸에 부착한 센서에서 나오는 데이터는
분당 1만 2천여건
 (한 경기(90분)를 뛰는 동안 한 선수 당 432만여개의 데이터가 생성)

선수의 **호흡·맥박·활동량** 등의 데이터를 수집, **컨디션 관리·전술**에 활용



HANA플랫폼 기반의 매치 인사이트(Match Insight) 프로그램을 통해 2014 브라질 월드컵 우승에 기여, **12번째 선수**로 불림

[4] 수요 예측

< 내외부 빅데이터를 접목한 수요예측 >



최근 5년간 전국 169개 지점의 기상관측 자료와 10억 건 이상의점포 별 상품 판매 데이터를 분석

파리바게뜨 가맹점의 90%이상이
빅데이터를 활용한 날씨판매지수를
점포운영에 적극 활용해

매출 증대 및 영업이익 상승

(날씨지수 도입 후 한달 만에 조리빵 매출 30% 증가)



식품업계 최초 **날씨 판매지수** 제작,
매장 주변 날씨 예보, 요일 등의 요소를 종합하여 **판매 예측**

[5] 고객만족도 제고

< 빅데이터 기반 추천시스템 >



1 개인화된 도서 추천

- 인간 편집자 → 컴퓨터간 추천
- 개인화 추천 시스템 자신감



2 취급 상품 확장

- 책에서 모든 품목으로
- 구매 패턴반영 상품 추천



3 예측 배송 시스템

- 주문 이전에 물류센터 미리 도착
- 클릭시 즉시 배송, 배송시간 획기적 감소



사업초기부터 생성된 데이터를 축적하여 활용, 가능한 모든 업무 지능화

- 전세계 직원 수 13만명, 매출 740억 달러 (2013년 기준, 대한민국 GDP 6% 수준)

[6] 빅데이터 분석 기반 컨설팅

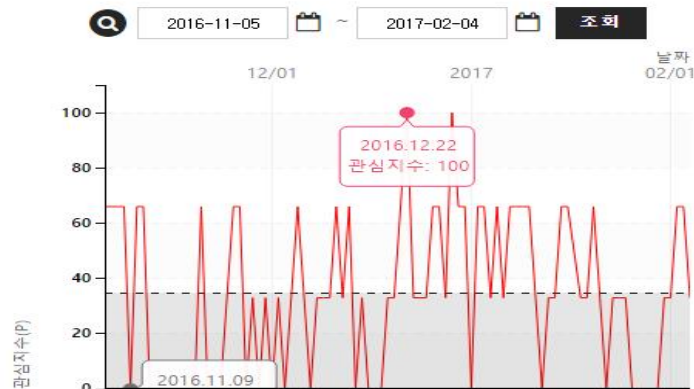
<의류시장 소비예측 및 마케팅 지원 >



*이미지 출처: WGSN, 삼성디자인넷, 스타일닷컴, PFin/한국그린트렌드마케팅연구소(GMTRI), 위클리트렌드

컨셉 관심도

컨셉 키워드의 정보량을 통해 해당 컨셉의 관심도를 확인할 수 있습니다.
최대 3개월까지 검색 기간 설정이 가능합니다.



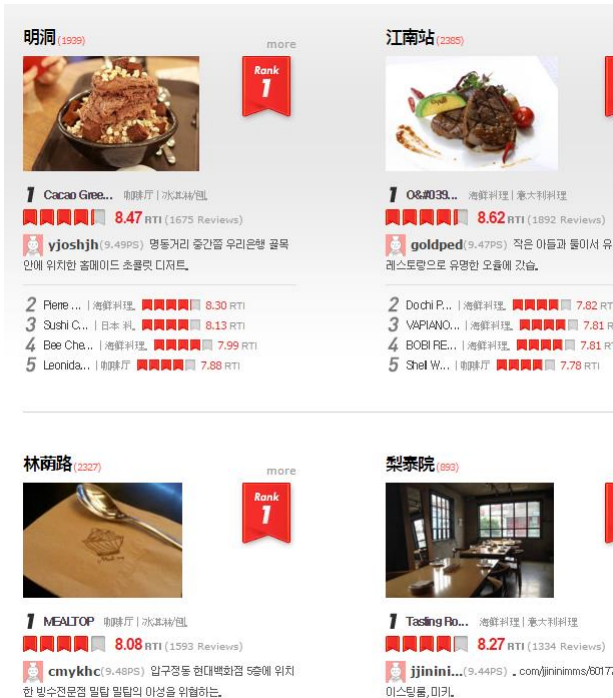
* '와이드' 컨셉에 대한 관심지수는 지난 2016. 4.23(토) 최초 최대치로 상승하였으며, 이후 계속해서 매우 큰 폭의 증감을 반복하고 있습니다. 이를 통해 사람들은 해당 컨셉에 대해 크진 않지만 지속적인 관심을 갖고 있는 것으로 보여집니다.

컨셉 연관어

해당 컨셉 키워드와 연관되는 상위 15개의 키워드를 보여줍니다.
거리가 가까울수록 밀접한 연관을 가집니다.



[7] 정보제공 서비스



➤ 기업명 : 레드테이블

➤ 인원 : 7명

➤ 주요 서비스

- 서울, 도쿄, 오사카, 베이징, 상하이 5개 도시의 레스토랑에 대한 데이터를 현지의 전문 웹사이트와 각종 SNS, 미디어 등을 통해 수집하고 있으며, 이를 서비스로 구현하여 4개 언어(한국어, 일본어, 중국어, 영어)로 제공
- 관광공사 Tour API, 서울시 상권분석 정보 연계 활용

*자료 : www.redtable.kr/zh/area



아시아 지역을 방문하는 관광객들에게 '현지인이 좋아하는 레스토랑 랭킹' 제공, 예약 및 결제까지 연계

[8] 앱 개발

- 기업명 : 케이에스비퓨처(주), ksbfuture Inc.
- 설립일 : 2013.1.21
- 대표자 : 박대건
- 주요제품 및 서비스
 - 도서추천서비스 센시북(SensiBook)(스마트폰 앱)
 - 비정형빅데이터분석 기반의 도서 추천 및 IOT책갈피를 활용한 독서 지도 서비스 센시북

*자료 : <http://sensibook.me/Demo>



센시북

도서를 추천 해주고, 추천된 도서를 완독할 수 있도록 도움을 주는 감성독서 도우미 센시북

[9] 신제품 개발

➤ 기업명 : (주)컨시더씨

➤ 주요 기능 및 서비스

- 미션기반 튜토리얼 - 레벨업/포인트 획득
- 테스트를 통한 체력 측정
- 멀티 플레이어 라이딩 제공
- 운동능력 변화 예측(개발 중)

➤ 관련 기술

- 모바일기기 <-> 장비 블루투스 통신
- GPS 데이터 모델링
- 스포츠과학 기반 활동데이터 분석

* 자료 : <http://www.rxcycle.me/#top>



가상현실 요소로 운동을 재미있게 시작하고, 바이오 데이터 피드백을 통해 사용자 체력을 분석하여 정확한 운동을 할 수 있도록 코칭

04 빅데이터 시대 진로와 경력 관리

유망직종, 데이터 과학자

SPOTLIGHT ON BIG DATA

Spotlight

ARTWORK Tamar Cohen, Andrew J Buboltz

2011, silk screen on a page from a high school yearbook, 8.5" x 12"

Data Scientist: *The Sexiest Job of the 21st Century*

**Meet the people who
can coax treasure out of
messy, unstructured data.**

by Thomas H. Davenport
and D.J. Patil

70 Harvard Business Review October 2012



데이터과학자의 정의

“Data Scientist는 복잡한 비즈니스 문제를 모델링하고, 인사이트를 도출하며, 통계학, 알고리즘, 데이터 마이닝 그리고 시각화 기법 등을 통해 그 속에서 기회를 찾아내는 사람이다.”

“그리고 이와 같은 고급분석(Advanced Analytics) 기술에 더해, 용량이 크고 다양한 유형의 Dataset을 다루는데 능숙하고, 특정한 목적 혹은 컴퓨팅 환경의 데이터베이스 아키텍처를 수립할 수 있으며, 분석 결과를 이해 관계자들과 커뮤니케이션 할 수 있어야 한다.”

“하지만 비즈니스 문제를 모델링하고 Data를 이해하는데 도움이 되는 특정 산업 영역에 대한 지식이 반드시 필요한 것은 아니다.”

데이터분석 프로젝트 구성



Collection

Size & speed
matters



Refinement

Get to answers, not
just ingredients



Delivery

For consumption
by end-users

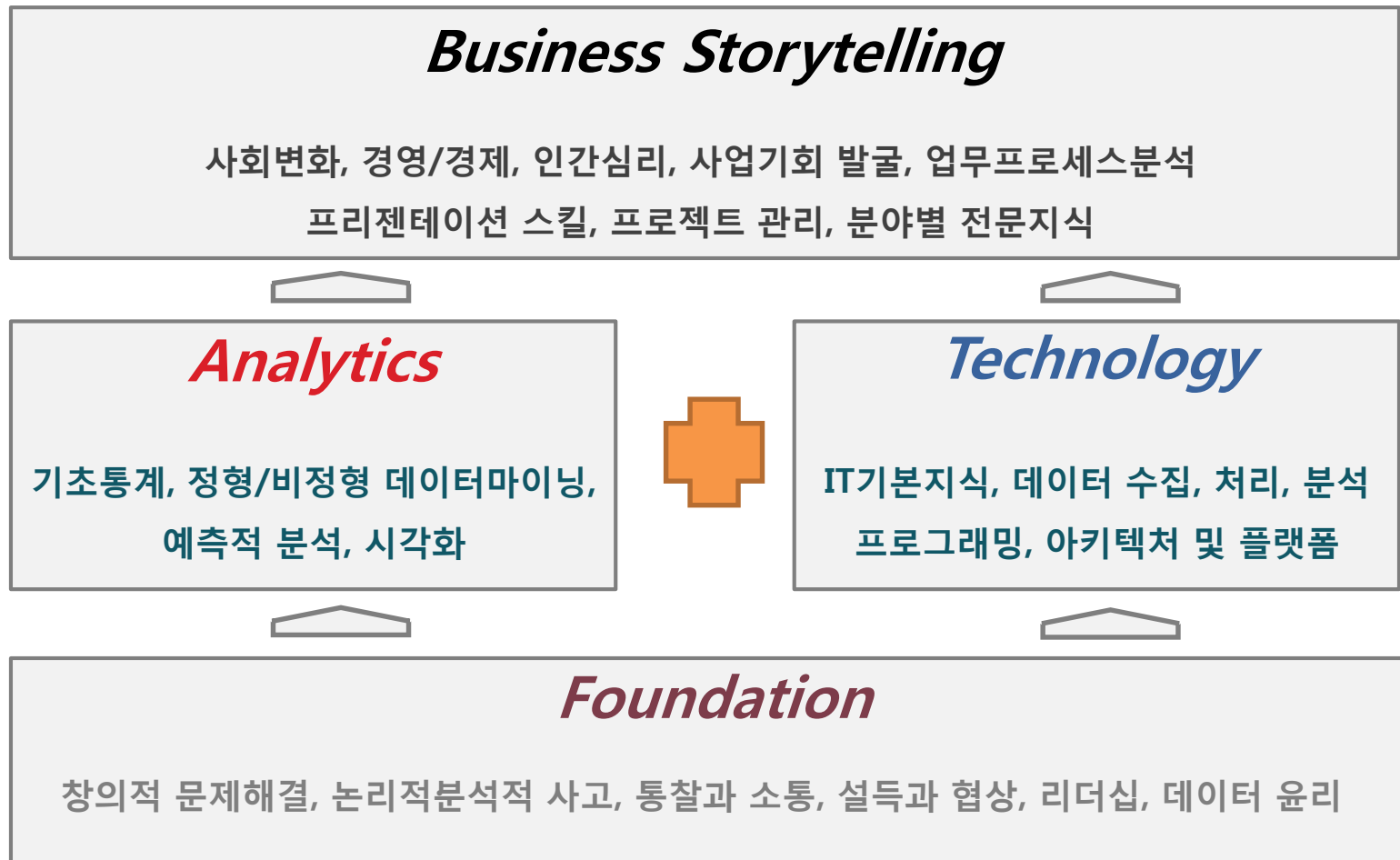
System Engineer

Data Engineer

Data Strategist

Data Scientist

데이터과학자의 역량체계



데이터과학자로의 첫걸음

<빅데이터 관련 자격증 현황>			
등록연도	자격명	등록	관리기관
2015년	산업데이터분석사	비공인	사단법인대한산업경영학회
2015년	광고데이터분석사	비공인	한국마케팅저널
2015년	공공빅데이터분석사	비공인	한국SW기술인협회
2015년	광고데이터분석사	비공인	한국평생교육진흥원
2015년	데이터분석사	비공인	주식회사 시소
2015년	경영빅데이터분석사	비공인	상상엠미디어 주식회사
2014년	기업실무형 데이터분석 마스터	비공인	한국표준협회
2014년	데이터마이닝 전문가	비공인	사단법인 한국고객관계관리협회
2014년	데이터 분석	비공인(내년 1월 국가공인 전환)	한국데이터베이스진흥원
2014년	데이터베이스 분석사	비공인	한국서비스산업진흥원
2014년	고객DB분석가	비공인	사단법인 한국고객관계관리협회

자료: 한국직업능력개발원

데이터과학자의 자세

- ✓ “Big” 말고 “Data”에만 집중하라
- ✓ 기술에 현혹되지 말고 문제해결에 주목하라
- ✓ 빅데이터가 모두 정답을 이야기해주지 않는다
- ✓ 데이터와 전문가의 직관을 결합하라
- ✓ ROI를 항상 염두에 두어라

Q&A
