

빅데이터 이해와 활용사례

2015.09

조완섭

충북대학교 경영정보학과
대학원 비즈니스데이터융합학과
wscho@chungbuk.ac.kr

043-261-3258

010-2487-3691



McKinsey Global Institute



June 2011

Big data: The next frontier
for innovation, competition,
and productivity

충북대학교 빅데이터 소개



2012.03

비즈니스데이터융합학과 설립
미래부 빅데이터 석사과정 사업
(4개학과, 교수 7, 석박사 50여명)

2013.09

빅데이터생활형서비스센터 설립
(미래부 대학 IT연구센터사업)

한국빅데이터서비스학회
(학술대회 10여회)

2015.05

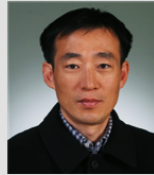
- 청주시 교통 빅데이터분석 사업 (청주시청)
- 제조빅데이터 사업 (산자부)
- 다양한 SNS 분석 활용 사업 (충북도청 등)



충북대학교 빅데이터 소개

조완섭 (曹完燮)

소속 : 충북대학교 경영대학 경영정보학과
 대학원 비즈니스데이터융합학과 (학과장)
 연락처: 010-2487-3691, 043-261-3258, wscho@chungbuk.ac.kr
 학력 : 1996.2 한국과학기술원 (전산학, 공학박사)
 경력 :
 2012.3~현재 충북대 대학원 비즈니스데이터융합학과 학과장
 (미래부 지원 빅데이터 인력양성 석사과정)
 2015~2017 미래창조과학부 전문위원(미래부)
 수상 : 2015.4 국무총리표창 (정보화 - 빅데이터 확산)



관심분야

빅데이터, 빅데이터 거버넌스, 비즈니스 인텔리전스, 기업정보화, DB

연구실적 및 활동

(1) 저서 및 역서

- 빅데이터 이해와 활용, 충북대학교출판부, 2015년 2월
- 빅데이터 거버넌스 (빅데이터 활용의 전제조건), 홍릉과학출판사, 2014.04
- 데이터베이스 시스템, 피어슨 코리아, 2010 (Database Systems, Elmasri 원저)

(2) 강연

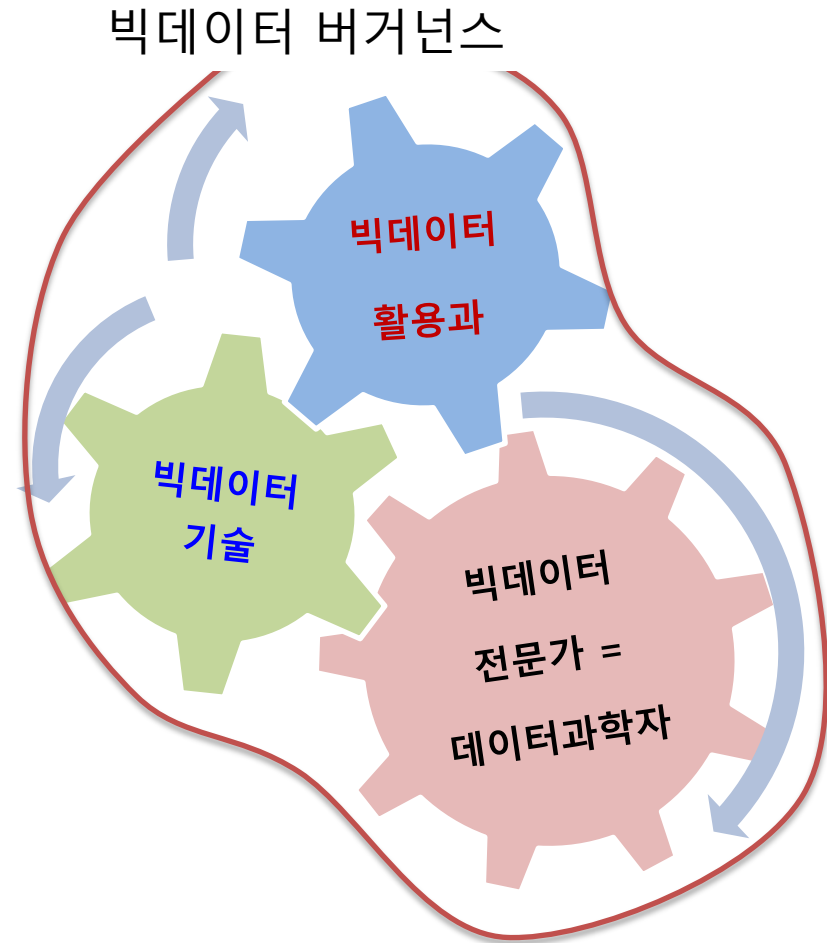
- 경기도청, 경기도공무원 빅데이터 심화활용 교육, 2015.06.01~(3일간)
- 경기도청, 빅데이터 이해와 활용, 세미나, 2015.04.28
- 질병관리본부, 빅데이터와 의생명분야 활용방안, 2015.03.
- 한국건설기술연구원, 빅데이터 이해와 활용 (Hive 실습), 2014.08.27~28
- 통계청, 빅데이터 활용사례와 거버넌스, 2014.06 (2일)
- 유한양행, 빅데이터 활용사례와 거버넌스, 2014.5.22.
- 현대자동차 (남양연구소), 빅데이터와 자동차 산업, 2014.05.14
- 한국교육학술정보원(KERIS), 빅데이터 활용과 거버넌스, 2014.03.21.
- 청주시청, 빅데이터 활용과 거버넌스, 2013.10.4.
- LG 전자, Big Data 활용과 비즈니스 혁신, 2012.7.

(3) 언론보도

- KBS 9시 뉴스(중양), 아동성범죄 위험구역에 CCTV가 없다, 2015.08.31
- 연합뉴스, 정보통합전산센터, 빅데이터 연구기관과 협의체 구성, 2015.08.21
- KBS 청주방송, 빅데이터 청주시 교통혼잡지도, 2014.4
- KBS 청주방송, 청주시 교통카드 3천만건 분석, 환승 어디가 많나?, 2014.3
- KBS 청주방송, 빅데이터가 바꾸는 세상, 2014.2.
- KBS 청주방송, 지자체도 빅데이터 시대, 2013.11.20.
- 디지털타임스, 빅데이터, 거버넌스에도 관심을 가져야, 2013.11.19.
- 블로터닷넷, 빅데이터학과에선 뭘 배울까, 2012.5.13.
- 디지털타임스, 국내 첫 빅데이터 학과, 2012.12.17.
- 중앙일보, 충북대 전국최초 빅데이터 석사과정 신설, 2012.2.29
- 매일경제, 충북대 빅데이터SW 석사과정, 2012.02.28.

목차

- 개요
- 빅데이터 활용
- 데이터과학자
- 빅데이터 기술
 - 오픈소스 플랫폼
- 국내동향
- 결론



개요

● 빅데이터 시대의 도래

전세계 데이터는
매년 40% 증가

세계는 2010년 zettabyte 시대에 돌입
1 Zettabyte : 美의회도서관 정보(235TB, 11/4 기준)
의 4백만 배, 16GB iPad로 축구장 넓이로 쌓아도
대기권 높이 두배

빅데이터 활용으로 연간
미국 의료분야에서 \$3,300억 가치생산
유럽 공공분야에서 €2,500억 절감효과

2018년까지 미국에서만 연간
분석전문가 14~19 만명, 데이터 기반 관리자 150 만명 추가 수요

(출처) McKinsey (2011,05), "Big Data: The next frontier for innovation, competition, and productivity"

빅 데이터 생성 속도

- 하루 250경 바이트 비정형 정보
- 매달 10억 여 개 트윗
- 매달 300억 여 개 페이스북 메시지
- 1조 대 이상 모바일 기기로 가속화

Overload

Global information created and available storage (Exabytes)



Source: IDC (2011)

개요

● 빅데이터란 ?

- 기존 컴퓨팅 기술로는 저장, 관리, 분석이 불가능할 정도로 큰 데이터 집합과 관련 기술, 인력 등을 포괄하는 의미
- IT 기술에서 출발했으나 사회, 문화, 정치 등 삶 전체의 이슈, 혁신 패러다임으로 부각 (Economist, Gartner, McKinsey, NYT)

● 빅데이터의 특징 : 3Vs

	기존 데이터	빅데이터
크기(volume)	MB, GB	TB, PB 시대에 따라 다른 기준
다양성(variety)	정형 데이터 위주 (매출, 재고, 인사 생산, 고객 데이터)	비정형 데이터 (동영상, SNS, 음악, Stream data 등)
속도(velocity) (생성, 유통, 활용주기)	수시간~수주	수초, 수분 (실시간 분석 요구)

개요

• 빅데이터는 생성 원천은 ?

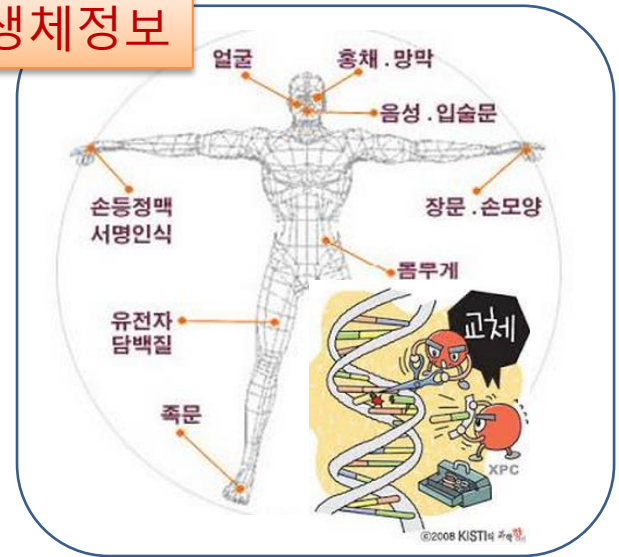
비즈니스 데이터



Social Data



생체정보



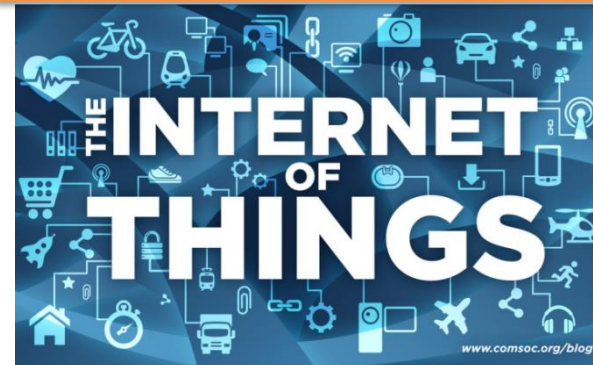
동영상

YouTube, CC 카메라, 의료장비 등



사물인터넷(IoT) / M2M

2014년 500억개 무선단말기(Ericsson)



빅데이터 - 효과와 전망

• 빅데이터 효과

- 인류가 직면한 **의료, 환경, 식량, 에너지** 등의 분야에서 해결방안 제시



Economist
(2010.05)

- SNS와 M2M 센서 등을 통해 도처에 존재하는 데이터의 효과적 분석으로 전세계가 직면한 환경, 에너지, 식량, 의료문제에 대한 해결책을 제시



Gartner
(2011.03)

- 데이터는 21세기의 원유이며 데이터가 미래 경쟁 우위를 좌우
- 기업들은 다가온 데이터 경제시대를 이해하고 정보 공유를 늘려 **Information silo**를 극복해야 함



Big data: The next frontier for innovation, competition, and productivity

McKinsey
(2011.05)

- 빅데이터의 활용에 따라 기업/공공 분야의 경쟁력 확보와 생산성 개선, 사업혁신/신규사업 발굴
- 특히 의료, 공공행정 등 5대 분야에서 **6천억불 이상의 가치 창출** 예상

빅데이터 - 효과와 전망

• 빅데이터의 효과 - 산업분야별 경제적 효과



의료 건강(美)

- 매년 \$3,300억 가치
- 年 ~0.7%의 생산성 증가

스페인 의료비의 2배



공공, 행정 부문(유럽)

- 매년 €2,500억 가치
- 年 ~0.5%의 생산성 증가

그리스 GDP 규모



개인 위치 정보(글로벌)

- 서비스 공급자 매출은 \$1,000억 이상
- 사용자 혜택은 \$7,000억



소매업(美)

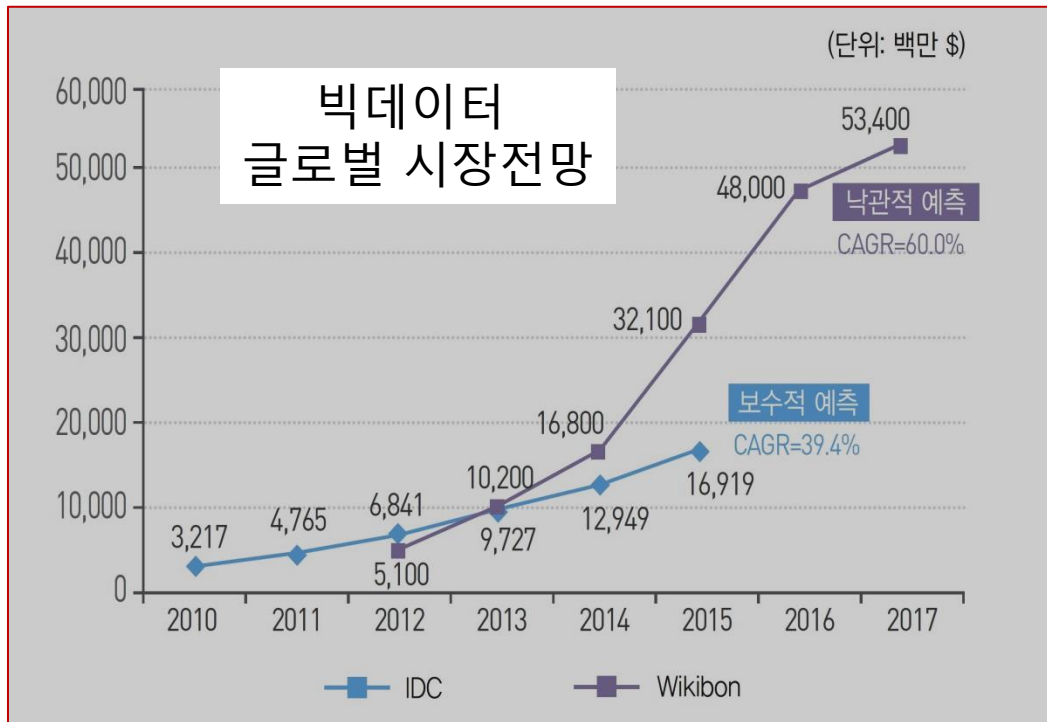
- 이윤 60% 증가 가능
- 年 0.5~1% 생산성 증가



제조업

- 제품개발비 50% 감소
- 운전자본 7% 절감 가능

빅데이터 시장



세부 분야별 시장전망

(단위 : 백만달러, %)

구분	2010	2011	2012	2013	2014	2015	비중
하드웨어	919	1,371	2,269	3,368	4,474	5,756	34
SW	1,062	1,415	1,851	2,476	3,376	4,625	27
서비스	1,236	1,979	2,721	3,883	5,099	6,538	39
합계	3,217	4,765	6,841	9,727	12,949	16,919	100

빅데이터 - 활용사례

- **KBS 보도자료 “빅데이터, 세상을 바꾸다”**

<http://news.kbs.co.kr/economic/2012/02/01/2428163.html>

[관련 동영상](#)



- 병원 미숙아실의 각종 의료장비에서 생성되는 스트림 빅데이터
 - 초당 10,000건의 데이터 발생
- 실시간 통합, 분석을 통하여 응급상황 조기 예측 (**2시간전=>24시간 전**)
- 미숙아 사망률의 획기적인 감소, 의사와 간호사 노동 감소

빅데이터 - 활용사례

구글 무인자동차 제원

광선 레이더

천장에 탑재된 레이저 거리-이동 감지장치. 회전하는 센서가 지붕에 달려 있다.

비디오카메라

실내 뒷거울에 장착된 센서가 교통신호기 변화 등 차 주변의 작은 움직임까지 측정해 지도상 정확한 위치를 표시하도록 돕는다.

컴퓨터

운행에 관련된 정보를 모으고 구글 서버와 교신하면서 조종을 지시.

수동 조작 전환 장치

비상시 빨간색 버튼을 누르면 수동 조종이 가능.

위치 파악기

차의 운동 상태를 감지하고 운전 조작 정보를 처리.

전파 탐지기

차로부터 사물이 떨어져 있는 거리와 주위 장애물의 이동 감지.



빅데이터 - 활용사례

2014.11.03

‘구글·애플 때문에’…고민 깊어가는 현대차



이성규

현대자동차의 고민이 깊어지고 있다. BMW, GM, 벤츠 등 경쟁 자동차 업체 때문만은 아니다. 애플, 구글과 같은 IT 거인들의 자동차 시장을 향한 거침없는 행보가 시름을 배가시키고 있다.

불과 10여년 전까지만 해도 검색업체가 자동차 업체와 경쟁 구도를 형성할 것이라곤 상상하기 쉽지 않았다. 간혹 ‘미디어의 경쟁사는 나이키가 될 수 있다’는 미래학자의 전망이 제기되곤 했지만 긴박한 현실로 인식되기엔 논리적 빈틈이 적지 않았다.

자동차 운영체제는 ? 안드로이드

하지만 지금은 당면한 고민거리가 됐다. 특히 IT 소프트웨어 분야에서 경쟁 열위 상황인 자동차 업계는 자칫 시장의 주도권을 잃어버릴 수 있을 것이라는 한숨까지 나오고 있다.

현대차와 부딪히지 않을 것만 같았던 구글과 애플. 이제 현대차는 하드웨어 생산 기업이 될 것인지 테슬라 처럼 소프트웨어와 하드웨어를 완벽하게 통합한 혁신적인 자동차 기업이 될 것인지 선택해야 할 시점에 직면해 있다.

빅데이터 - 활용사례

- Google Glass

- http://www.youtube.com/watch?v=T2uwki9jk_8
- http://www.youtube.com/watch?v=cHix5yWJ_0Y

- 구글 글라스 체험단 모집 (미국)

- 엄청난 데이터가 생성될 것임 (lifelog data)
- 지금으로서는 응용 분야를 가늠하기조차 어려운 상황
- 스마트폰, 디카 대체 ?
- 사람들의 생활 모습은 ?

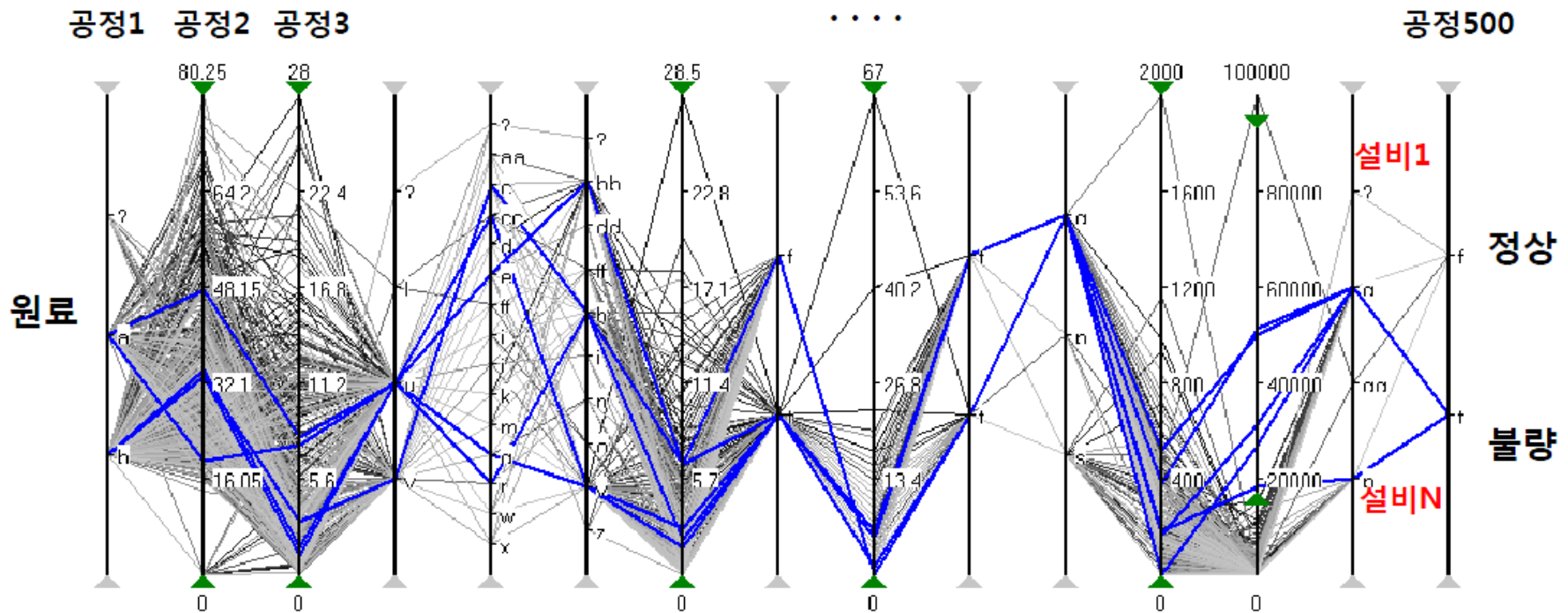
CIO 구글 글라스의 기업 침투...
10가지 뚜렷한 신호



빅데이터 기반 제조업 첨단화 - 반도체

- **이슈사항:** 반도체 전체공정(500여개)에서 수율에 영향을 미치는 인자 및 설비 발견
- **데이터량:** 센서정보(FDC data: Fault Detection & Classification data)는 연간 약 **10TB**정도
- **분석방법:** Tracing 분석 → 저수율 생산경로 추적 → 저수율 공정/설비 탐색 → 개선 → **수율 향상**

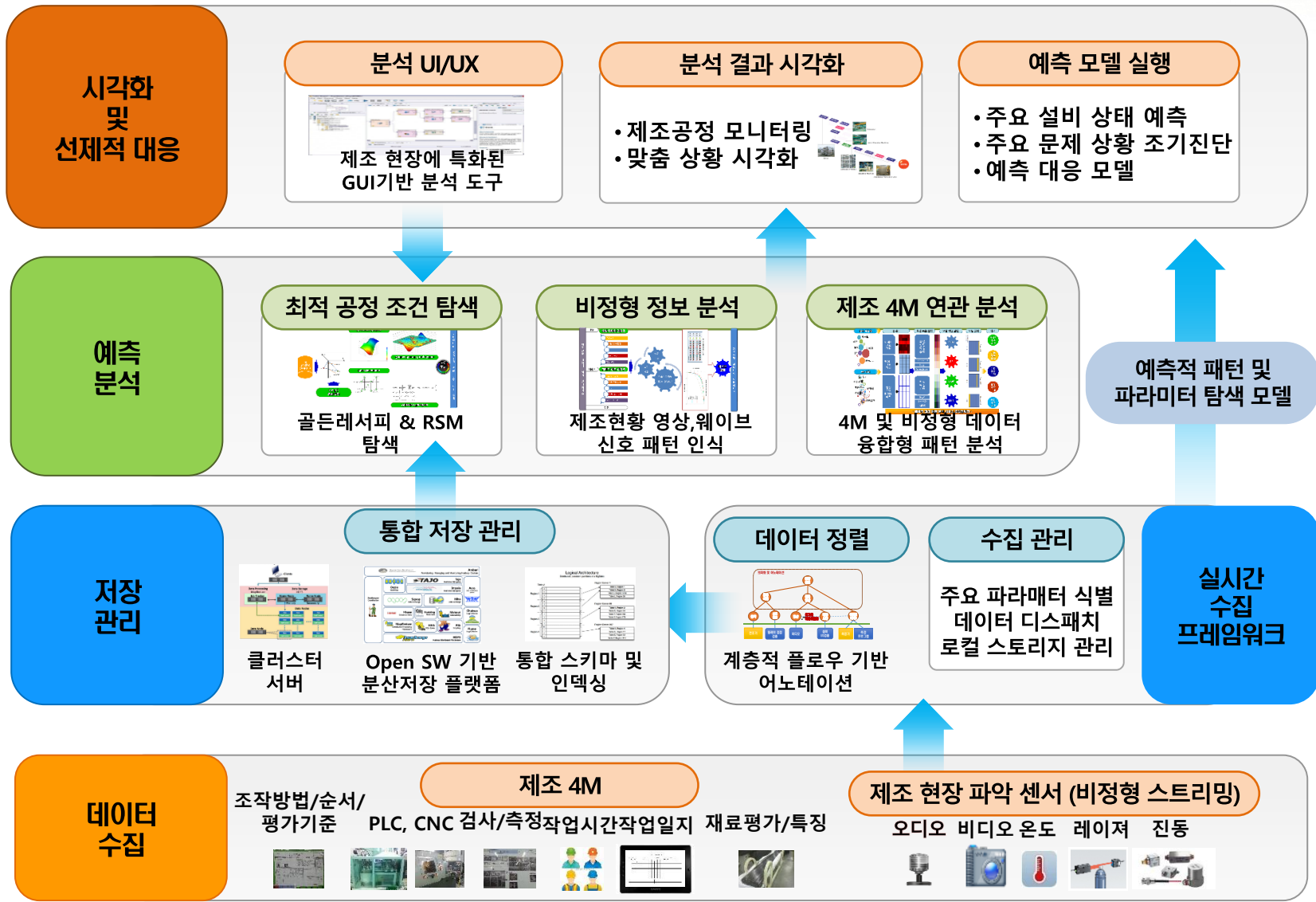
▪ 500여 공정 별/설비 별 경로



(출처) 빅데이터와 제조부문 적용사례 (이씨마이너) 발표자료, 2012 : S 반도체 사례

빅데이터 기반 제조업 첨단화 - 자동차부품

연구 목표의
명확성



빅데이터 활용사례 - 금융산업



현금=>통장=>카드=>스마트폰 (클릭 몇 번으로)
은행지점=>ATM기기=>스마트폰
결제 => 송금, 외환거래, 대출, 투자, 보험, 증권 ...

뉴스 증권 부동산 **정책·금융** 기업 Weekly Biz 오피니언

정책·금융 ▾

"연봉 반토막도 좋다" 금융맨들 짐싸서 핀테크로

이신영 기자 ▾

기사

100자평(1)

핀테크 성공은 '빅데이터'에 달렸다

2015빅데이터리더스포럼 9일 열려

상품개발, 부정방지, 신용평가, 마케팅 등 4대 금융 업무가 빅데이터 활용으로 빠르고 정확하게 처리 가능



입력 : 2015.02.09 03:04

["잘있기 이제 사람들은 이 과정보다 불편하다고 여긴다.페이팔 서비스나 알리페이, 애플페이 등은 은행 계좌나 카드 등을 한번만 연결해놓으면 그 다음부터는 결제나 송금 등을 바로 이용할 수 있다. 결제할 때 필요한 중간 과정은 알아서 대행해준다. 금융 서비스의 핵심인 본인인증을 대신해주며, 카드 정보 입력이나 계좌번호 입력, ARS 인증이나 전화 인증, 보안 코드 입력 등과 같은 다소 번거로운 과정도 생략해준다(엄밀하게 말하면 이 역시 대행해준다). 즉, 계속 반복해서 입력해야 했던 과거의 방식에서 벗어나 한번 입력하면 그 다음부터는 보다 간편하고, 쉽고, 빠르게 해당 서비스를 이용할 수 있다.

“ 앱 하나로 관제없이 더 빠르고 정확하게 택시 호출 ”

기존 콜 택시



관제센터



1분 마다 택시 위치정보
전송으로 탑승위치 부정확



단골택시



택시 오는 모습이
지도에 보인다

승객호출

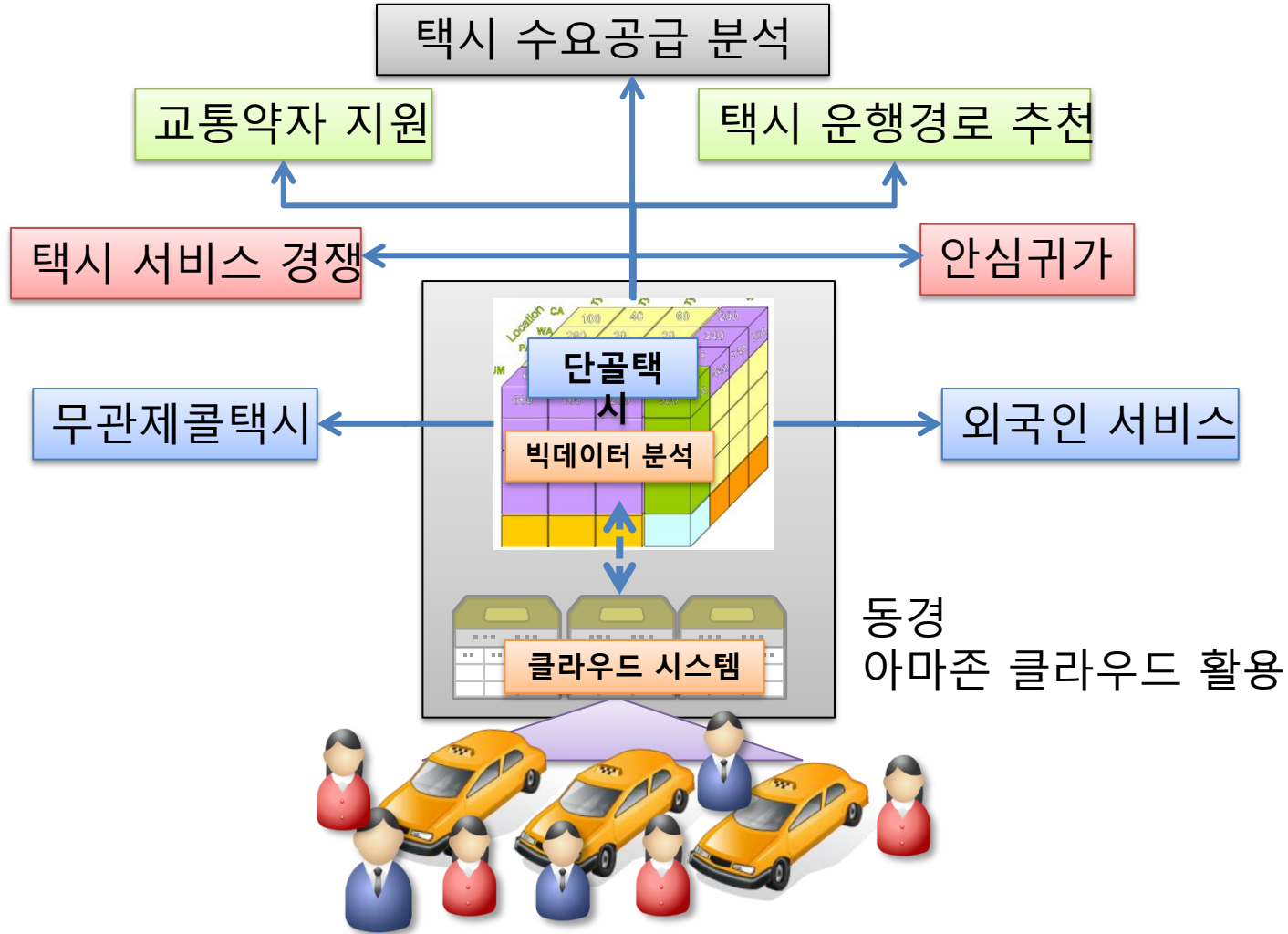


기사수락

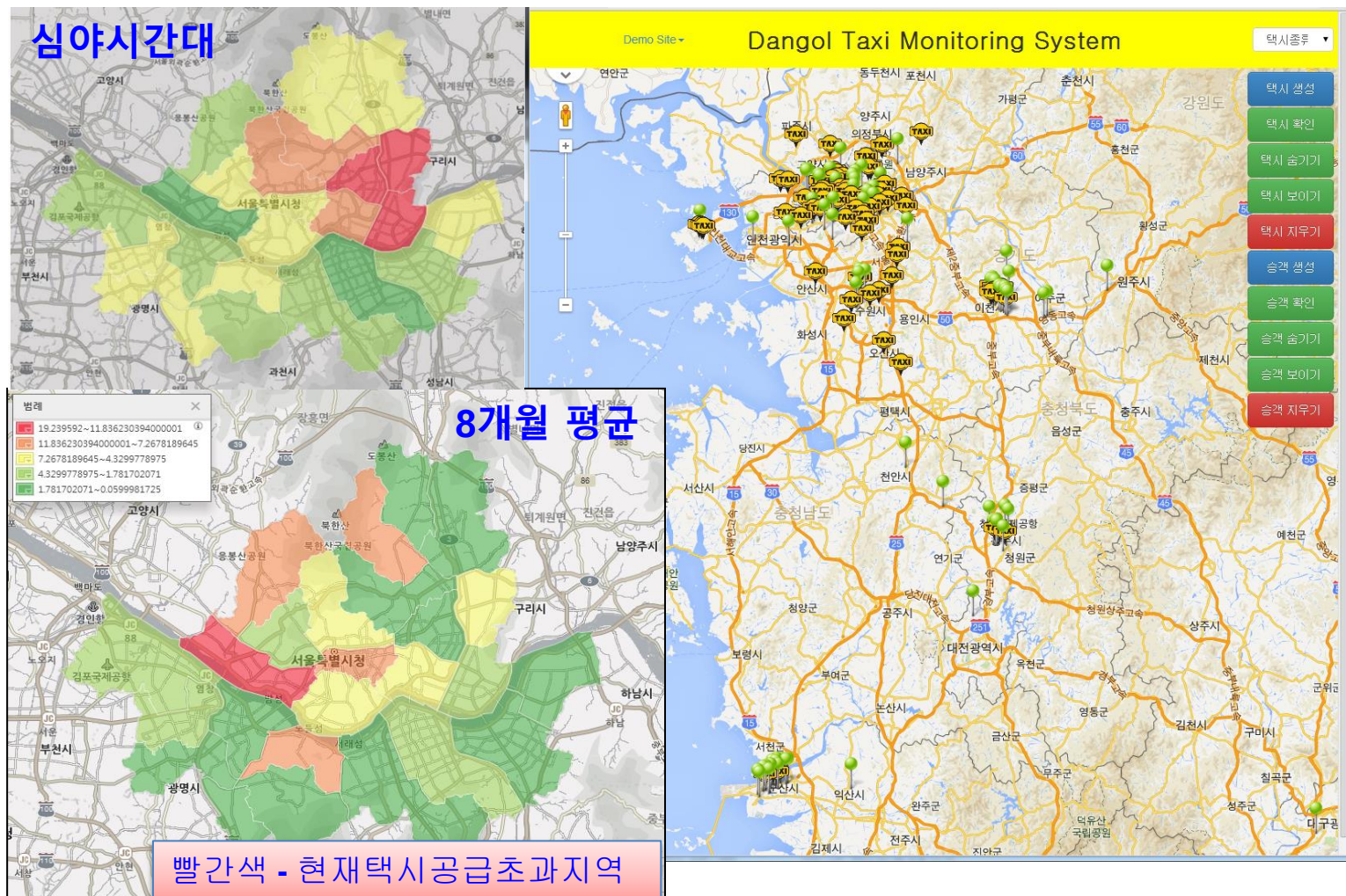


5초 마다 위치정보 전송함으로써
정확도 개선

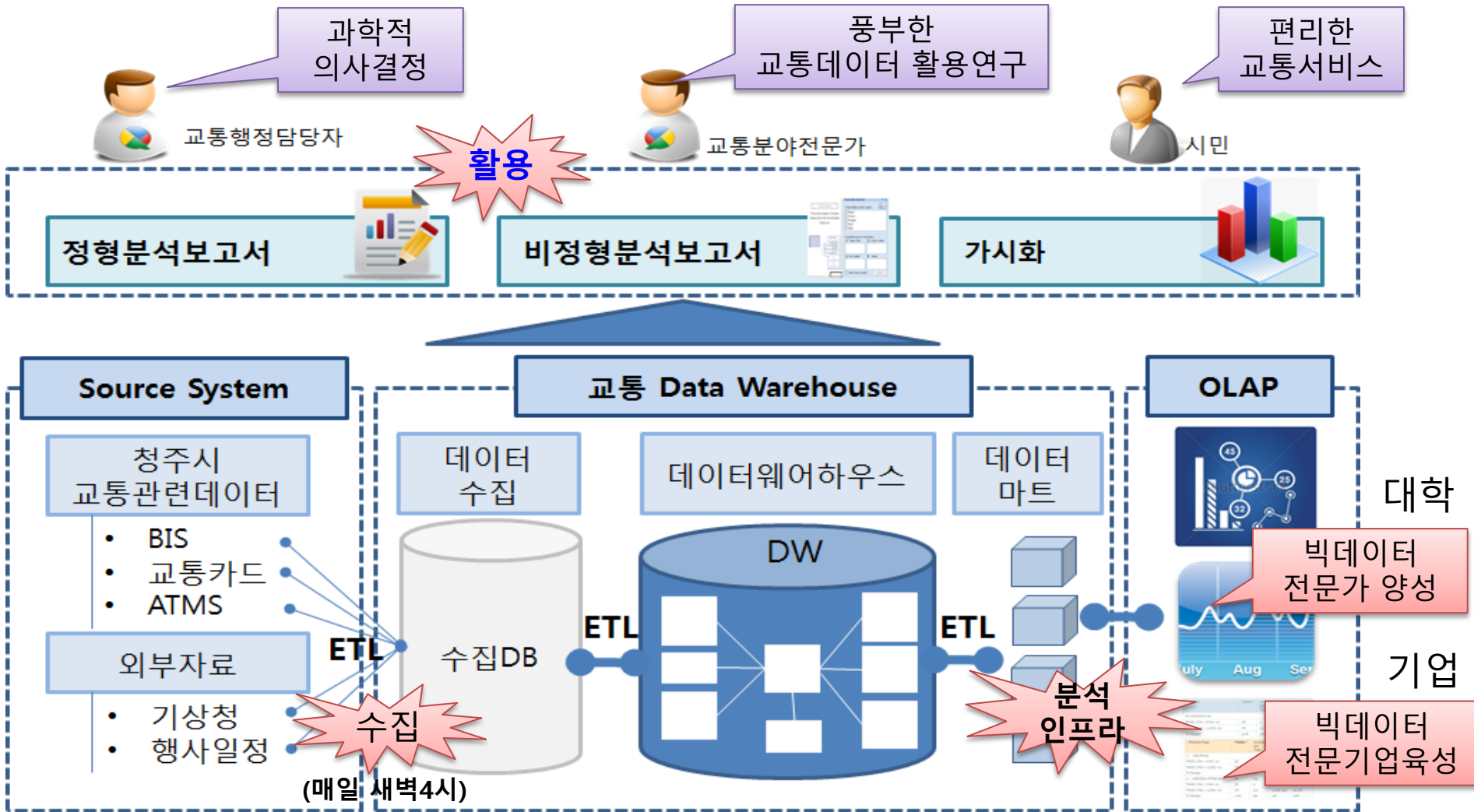
빅데이터 활용사례 - 교통



빅데이터 활용사례 - 교통 (택시)



빅데이터 활용사례 - 교통 (대중교통, 청주시)



빅데이터 활용사례 - 교통 (대중교통, 청주시)

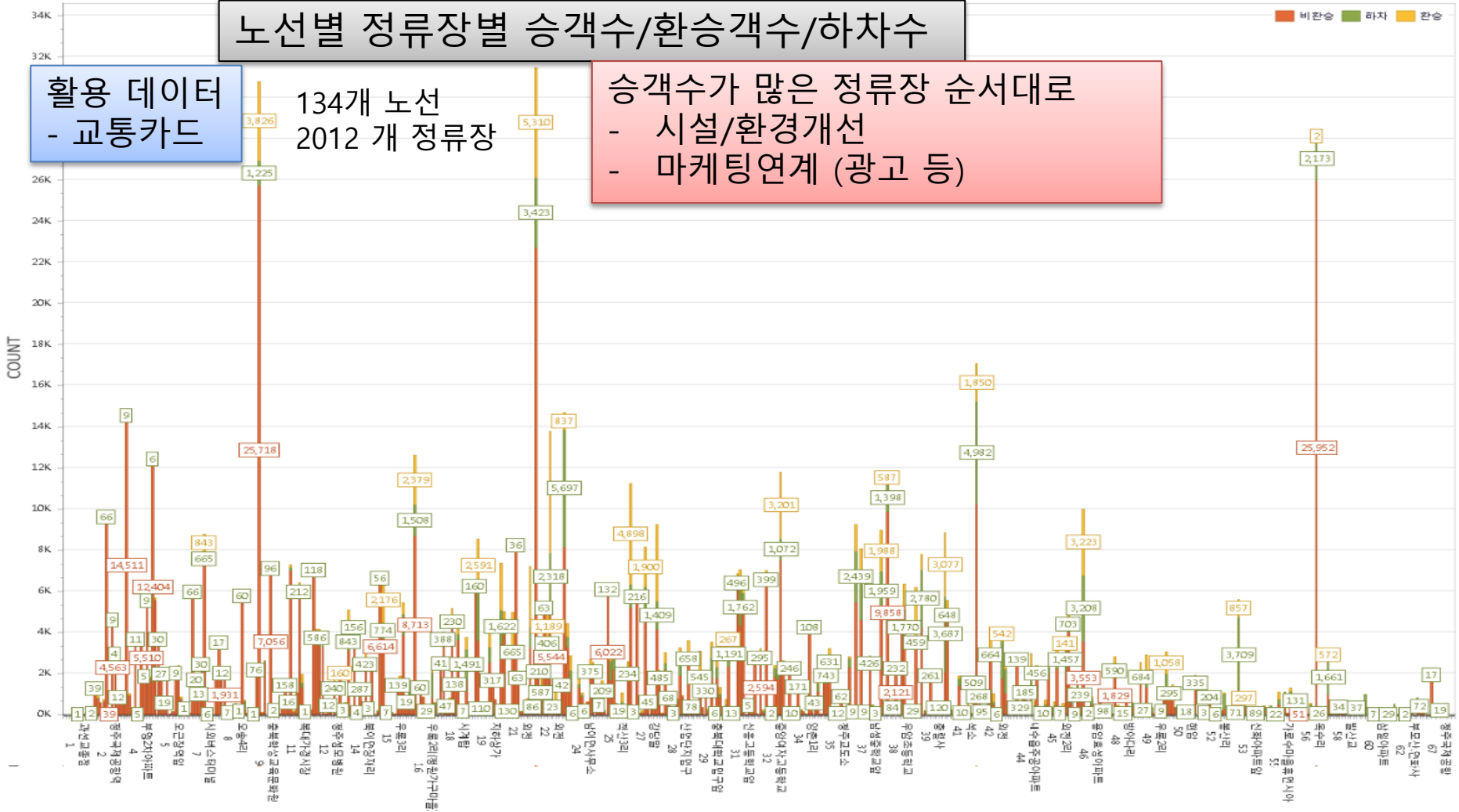
노선별 정류장별 승객수/환승객수/하차수 [단위: 명]

노선별 정류장별 승객수/환승객수/하차수

활용 데이터
- 교통카드

134개 노선
2012 개 정류장

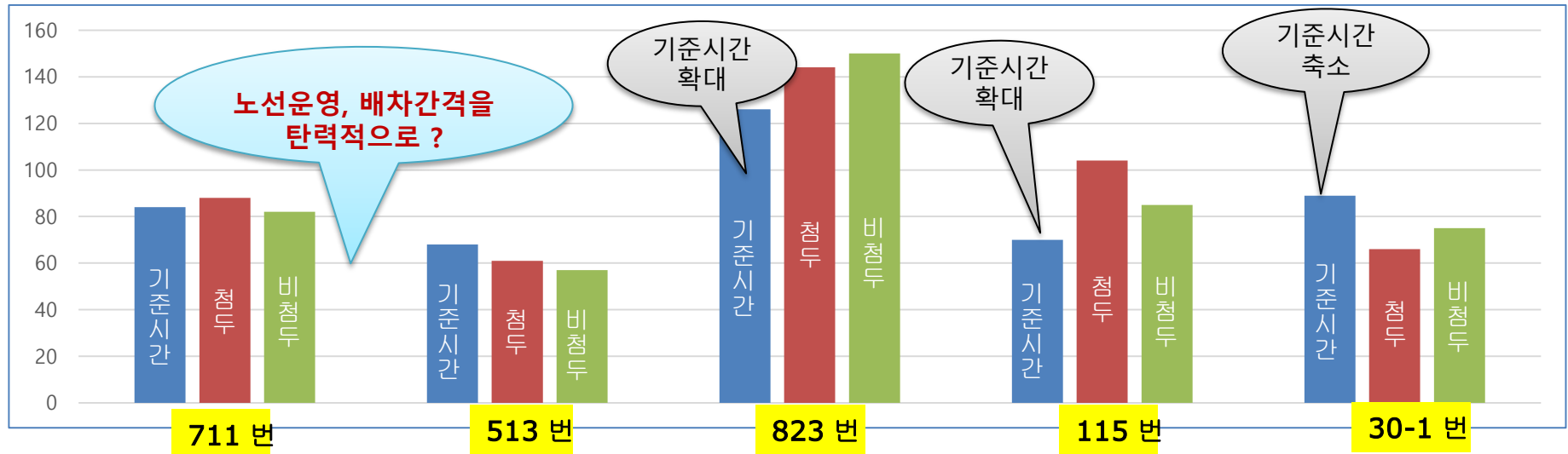
승객수가 많은 정류장 순서대로
- 시설/환경개선
- 마케팅연계 (광고 등)



빅데이터 활용사례 - 교통 (대중교통, 청주시)

노선별 평균운행시간 분석을 통한 기준운행시간 조정

	노선명	기준운행시간	첨두 실제운행시간	비첨두 실제운행시간
OK!	711	84분	88분	82분
배차간격조정 제안	513	68분	61분	57분
	823	126분	144분	150분
탄력적 배차 제안	115	70분	104분	85분
	30-1	89분	66분	75분



빅데이터 활용사례 - 교통 (대중교통, 청주시)

요일별/시간대별/날씨별/이벤트별 최대승객 환승 정류장 n개는?

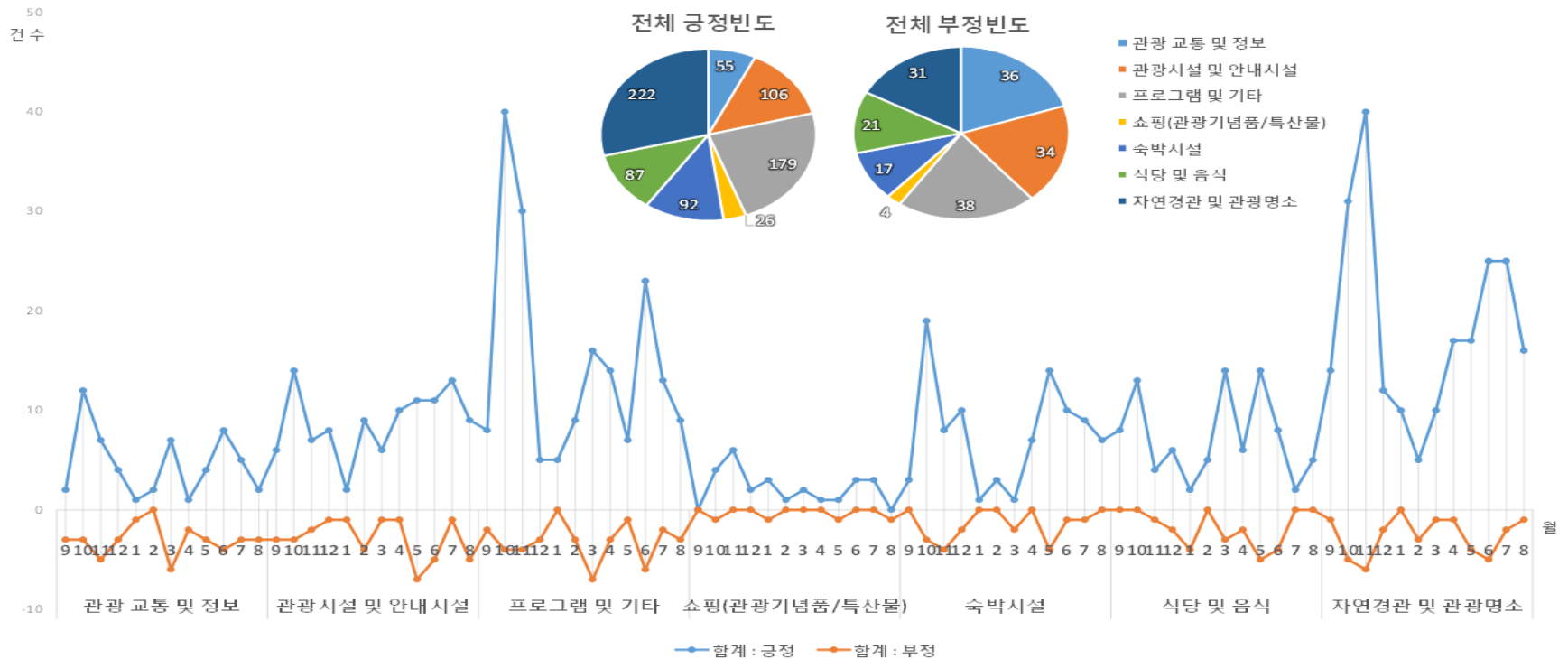
환승보조금을 최소화하도록 노선개편
월별/요일별/시간대별 탄력적 노선운영



충북 관광분석 - 350개 관광지별 이슈, 연관, 감성 분석

- 청남대는 자연경관, 산책, 트래킹 등 자연경관에 대한 긍정적인 의견이 높았으며, 전반적으로 시설 관리는 잘 되어 있으나, 매표소 안내에 대한 불만과 여름에는 햇볕을 피할 곳이 없다는 부정적인 의견이 나타났음.
- 관광 교통 및 정보에 관한 부정적인 의견이 많았으며, 특히 사전 예약제 시행으로 불편하다는 의견이 있었음.
- 매표소에서 청남대까지 이동 수단(셔틀버스)과 주차장 확충이 필요한 것으로 보임.

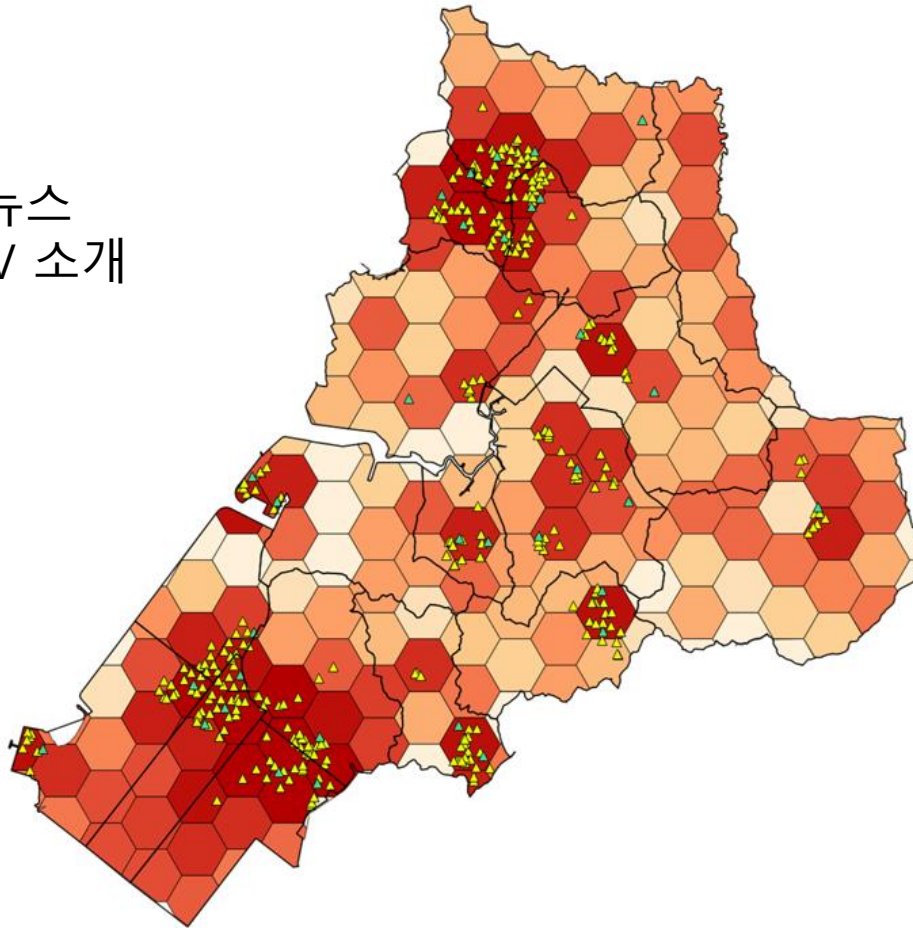
청원 청남대



스마트 치안

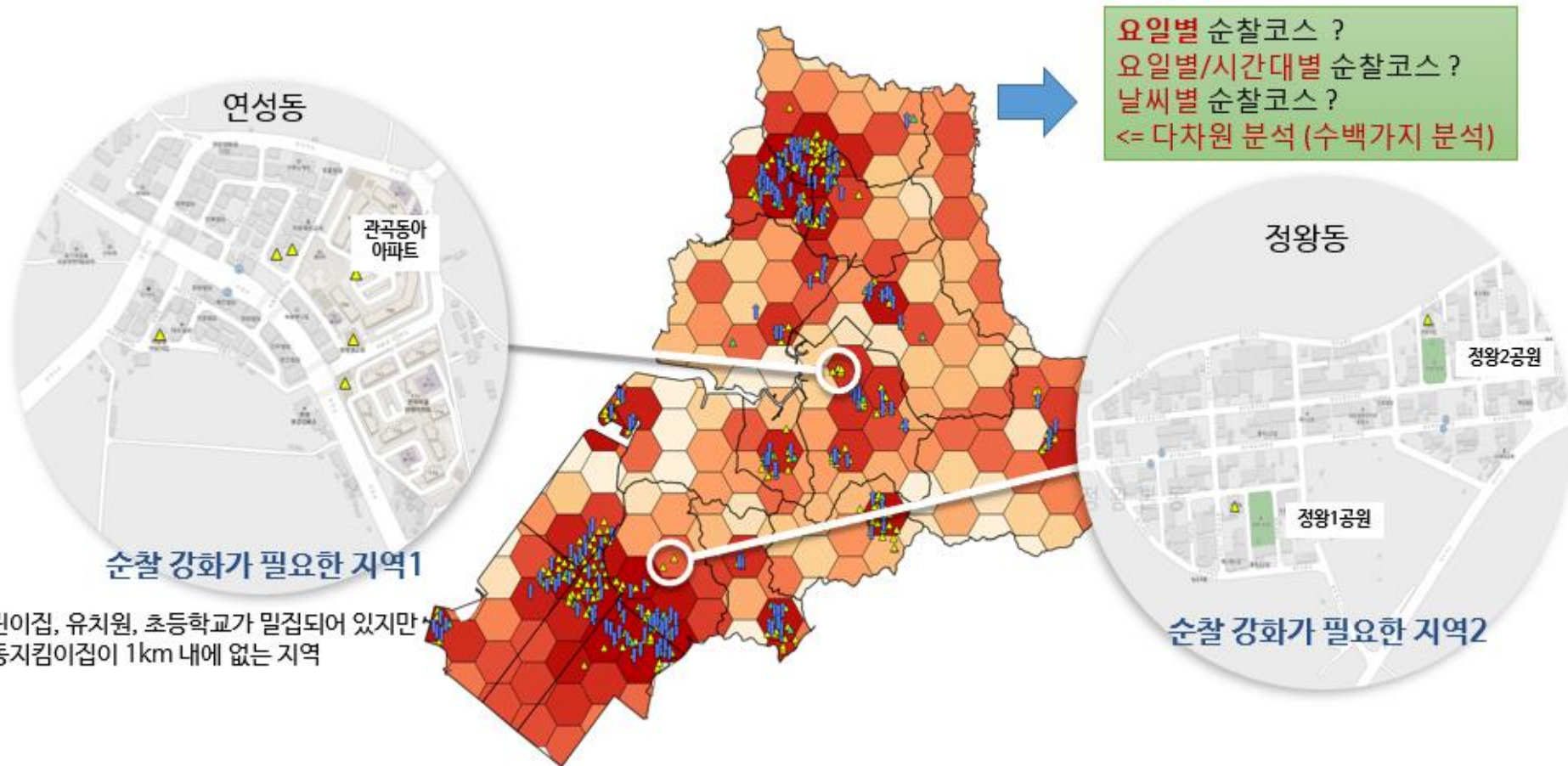
- 시흥시 어린이 시설현황, 범죄발생 지역 분석

KBS 9시 뉴스
중국 CCTV 소개



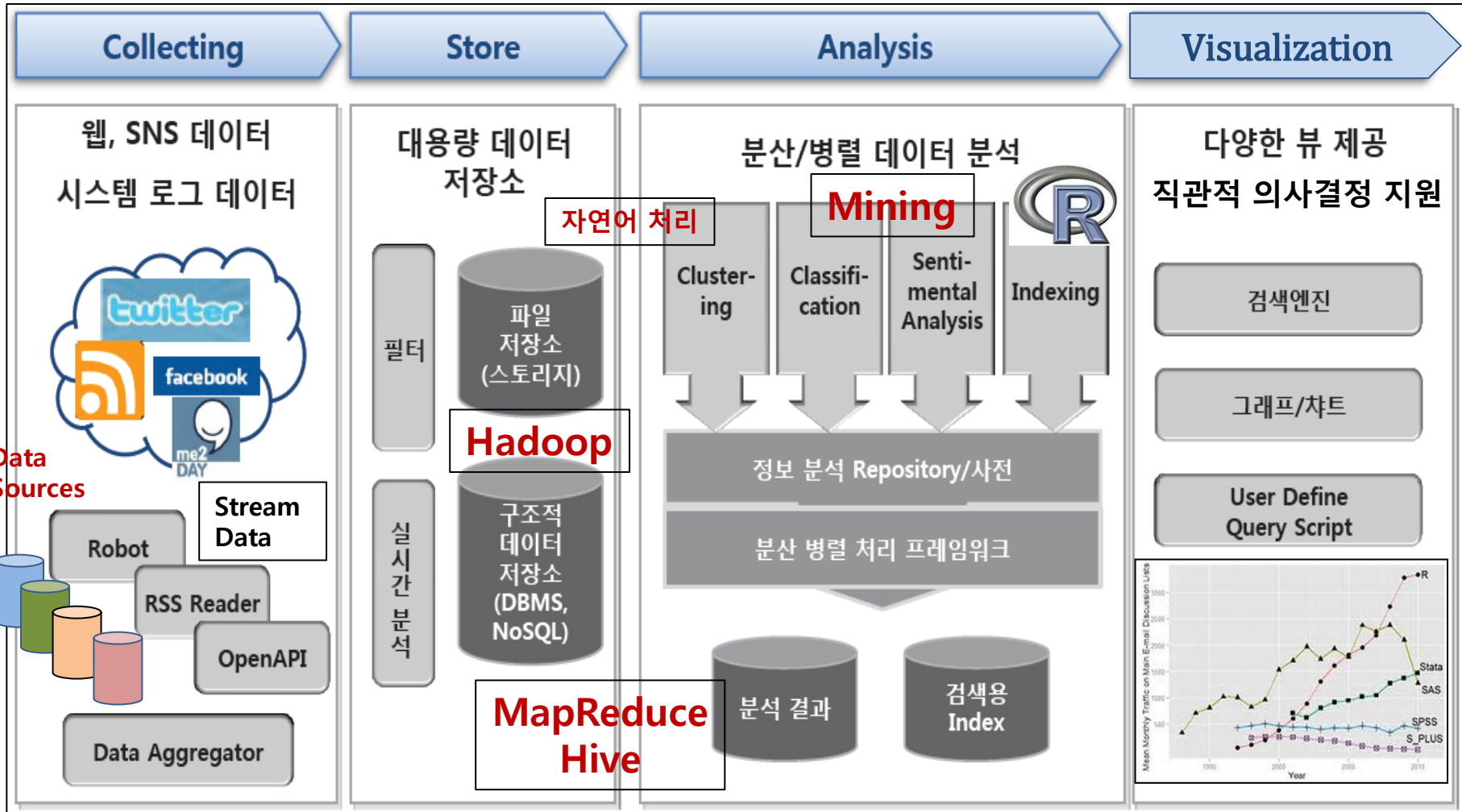
스마트 치안

- 순찰지역 탄력적인 추천, CCTV 사각지대 분석



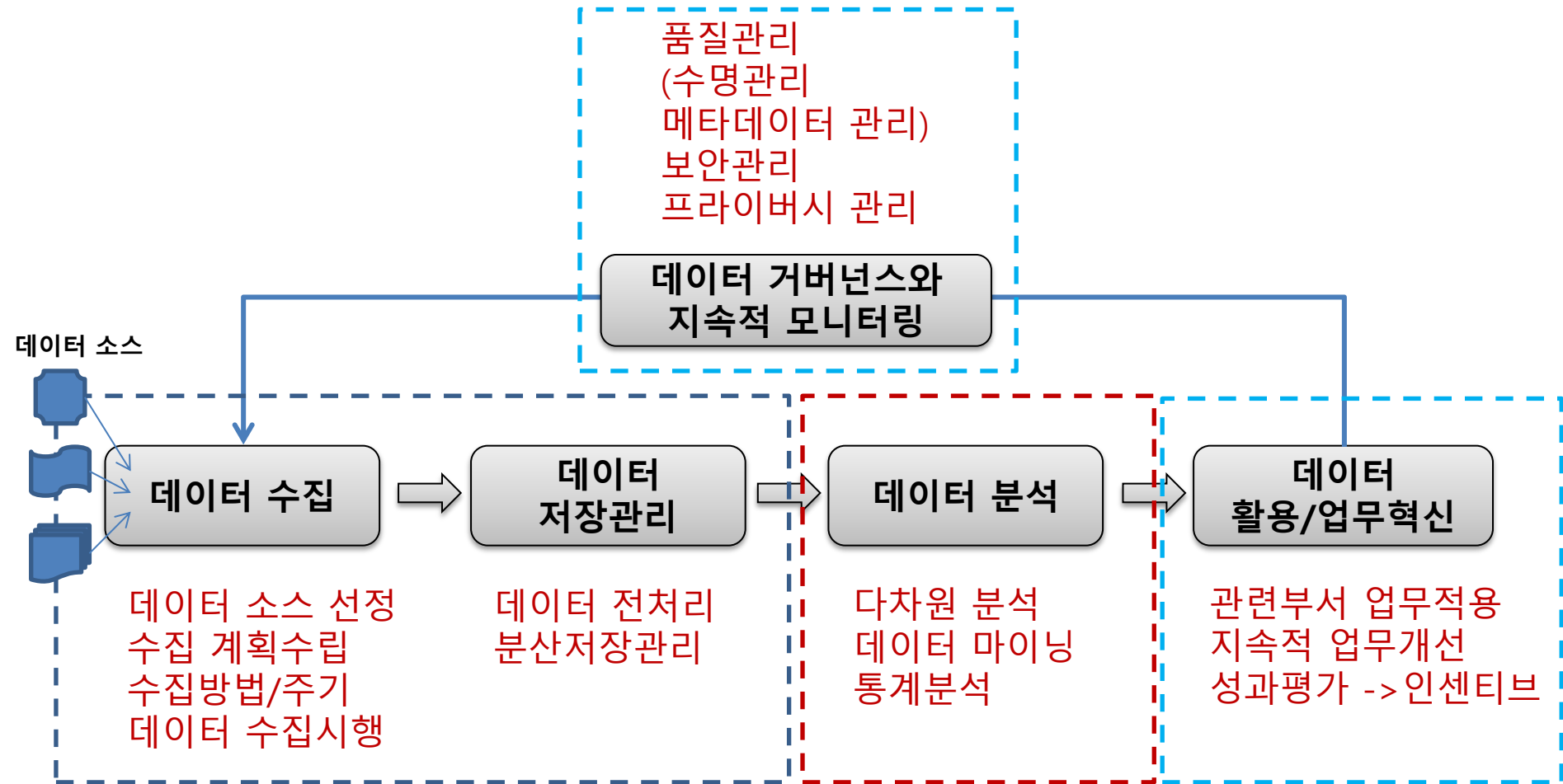
빅데이터 Project

- Data Lifecycle - 처리과정



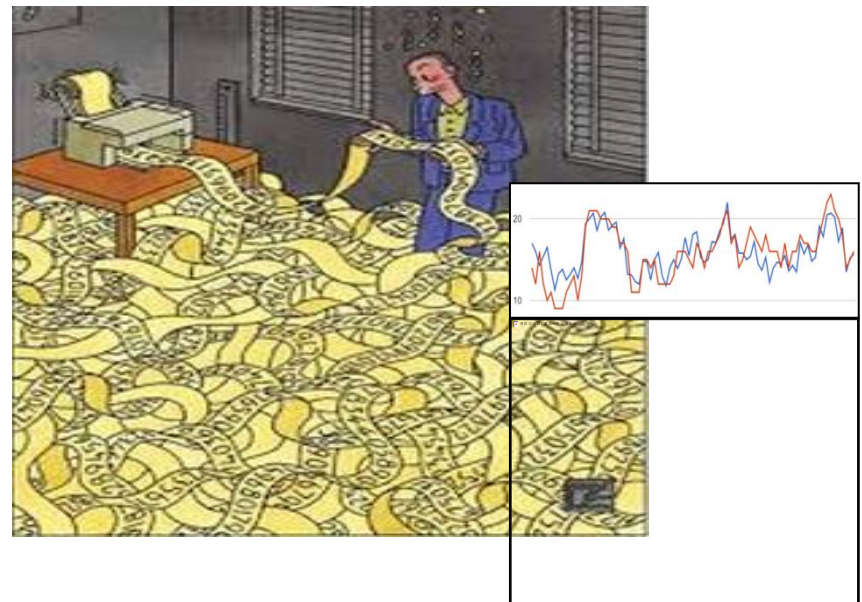
빅데이터 Project

- 빅데이터 사업수행과 활용 절차



데이터과학자

- **데이터 분석을 통해 가치(data product)를 창출하는 전문가**
 - 데이터의 다양한 소스를 찾아 데이터를 추출하고,
 - 크고 복잡한 데이터를 구조화/단순화하여 분석이 가능하게 만들며,
 - 이상한 데이터를 실시간으로 탐지해 내고,
 - 적절한 통계적 모델링을 통하여 예측을 수행하며,
 - 인사이트를 발견한 후 시각화하여 비즈니스의 방향을 제시함



데이터 과학자

- 데이터 과학자에 대한 언론 전망

2012년 최고의 신규 유망직종 (CNN)



Data Scientist: The Sexiest Job of the 21st Century

데이터과학자는 그들은 구하기도 힘들고,
고용하려면 비싸고, 과학-IT-분석능력을
겸비한 사람은 찾기 어렵다 (HBR, 2012)



최근 기업들은
머리는 수학/통계 지식으로,
손은 컴퓨터 해커 수준으로,
눈은 예술적 안목을 가진 사람 (data scientist)을
찾는데 애를 먹고 있음 (NPR, November 30, 2011)

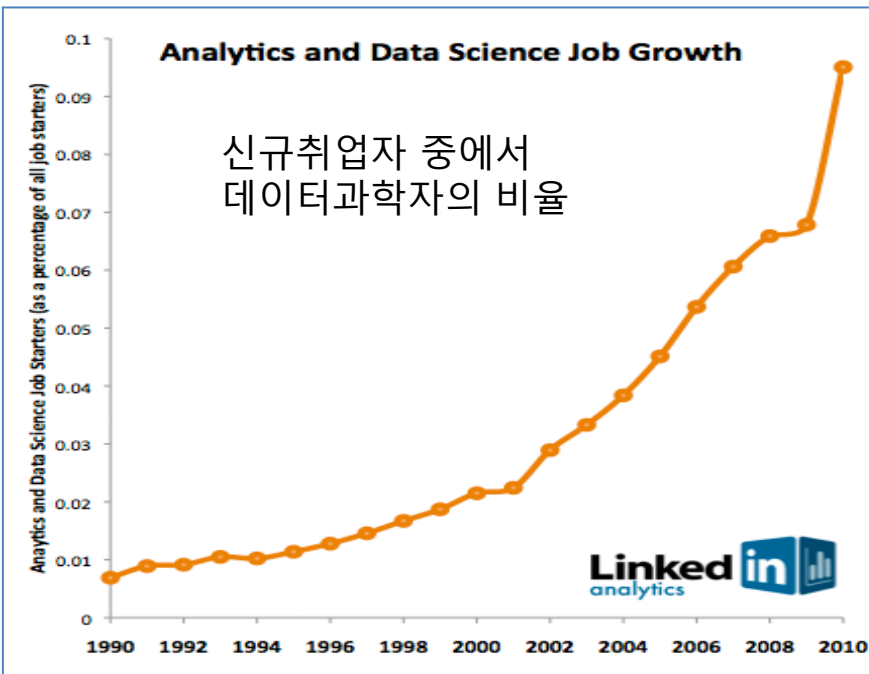


데이터과학자

● 데이터과학자 수요

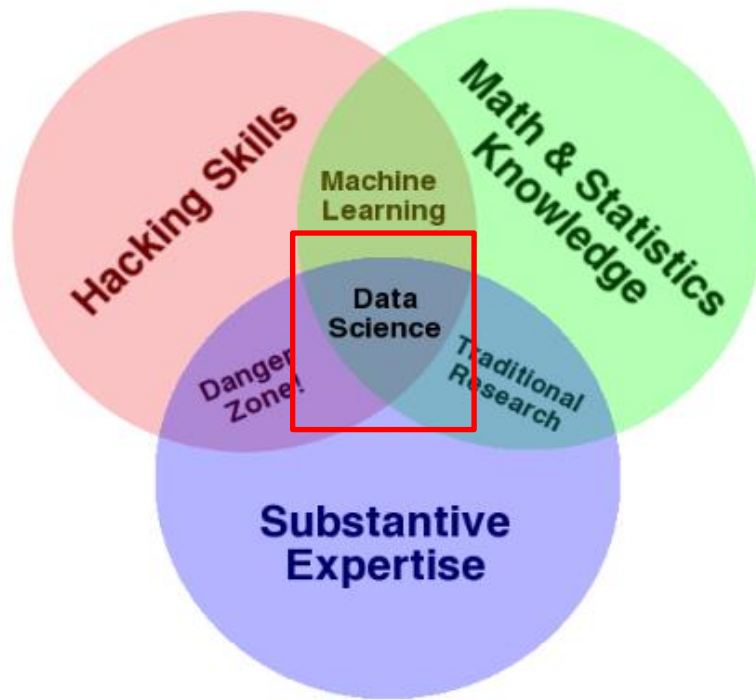
- 향후 3-5년 이내에

- 미국은 18만명의 분석전문가, 150만명의 데이터기반 관리자 필요할 것으로 예상 (맥킨지 2011년 보고서)
- 영국은 50,000여명이 필요할 것으로 예측
- 우리나라는 2017년까지 약 11,000명의 빅데이터 전문인력이 필요할 것으로 예상 (KISDI Premium Report「(14-10) 빅데이터 2.0시대, 주요 이슈와 정책적 시사점



데이터 과학자 - 역량

- **Data Scientist – Venn Diagram**



데이터과학자는
 IT 기술 (hacking skills),
 수학 및 통계 지식
 현장 업무 지식 (business, science 등)을
 모두 갖추어야 한다.

두 가지만 갖추면
 기존 과학자나 기계학습 혹은
 위험한 사람(danger zone)이 된다.

The skills needed for data scientists

출처 : <http://www.dataists.com/2010/09/the-data-science-venn-diagram/>

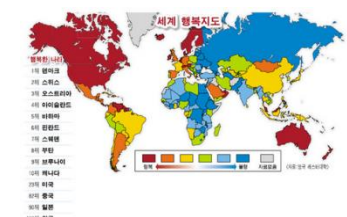
데이터과학자 - 기업의 분석지능

• 분석지능 (AQ • Analysis Quotient)

- 빅 데이터를 수집, 저장하고 이로부터 통찰력을 얻으며, 미래 상황까지 예측하는 능력
- 기업의 AQ의 4단계
 - 첫째 단계의 기업은 이제 막 **데이터 분석에 관심을 갖기 시작한 회사**
 - 둘째 단계는 어느 정도 데이터를 활용하지만 **비즈니스와 연계하지는 못하는 기업들**
 - 셋째 단계는 과거 데이터를 분석할 수는 있어도 **미래를 예측하지는 못하는 기업들**
 - 네째 단계 기업들은 과거와 현재, 미래를 예측할 수 있는 데이터를 모두 분석하고 활용해 **성과를 창출하는 회사**
- AQ가 높은 기업일수록 리스크 관리와 혁신 능력도 뛰어남
- 조직에서 데이터과학자 육성, 채용 필요

데이터과학자 - 기업의 분석지능

• 직관을 능가하는 분석 (MIT, CEO 3000명 설문결과)



[MIT Sloan Management Review] Big Data, Analytics and the Path From Insights to Value

데이터 과학자 - 교육프로그램

- 데이터 과학자 양성 프로그램(미국 50여개 대학교)

School	Program	Offline	Online	Degrees
Stevens Institute of Technology	Business Intelligence & Analytics http://www.stevens.edu/sit/graduate/master-of-science-business-intelligence-and-analytics.cfm	Yes	Yes	M.S.
CMU	Data Science	Yes	?	Certificate
North Carolina State University	Analytics http://analytics.ncsu.edu/?page_id=1799			
Northwestern University	The Master of Science in Analytics http://www.analytics.northwestern.edu/			
Stanford University	Data Mining http://scpd.stanford.edu/			Certificate
University of Cincinnati	Business Analytics http://business.uc.edu/departments/obais/oba.html			
Oxford University	Data and Systems Analysis	?	Yes	Undergrad Adv. Diploma
University of California San Diego	Data Mining http://extension.ucsd.edu/programs/index.cfm?vaction=certdetail&vcertificateid=128&vstudyareaid=14	No	Yes	Grad Certificate
University of Washington	Data Science http://www.pce.uw.edu/certificates/data-science.html	Yes	Yes	Certificate
University of San Francisco	Analytics http://www.usfca.edu/management/analytics/program/	Yes	No	M.S.
RUTGERS, The State University of New Jersey	Professional Science Master's degree in Analytics http://www.business.rutgers.edu/msis/academic-programs/ug-bait	Yes	?	M.S.
Syracuse University	Certificate of Advanced Studies in Data Science http://ischool.syr.edu/academics/graduate/datascience/index.aspx/index.aspx	Yes	?	Grad Certificate
Johns Hopkins University	Institute for Data Intensive Engineering and Science			

2012년 3월
충북대학교
비즈니스데이터융합학과
(미래부지원 데이터과학자 양성
석사과정)

오픈소스 기반의 빅데이터 기술



목차

- 개요
- 빅데이터 기술
- 클라우드 컴퓨팅
- Hadoop & Databases
- 데이터 분석기술
 - 다차원 분석
 - 통계 분석 : R
 - 데이터 마이닝
- 빅데이터 시각화 기술

개요

• 빅데이터 - 새로운 IT기술과 분석기술이 요구됨

기존방식으로 처리하기엔 데이터 규모가 크고 컴퓨팅 파워가 부족하기 때문

기존 Data 처리 방식

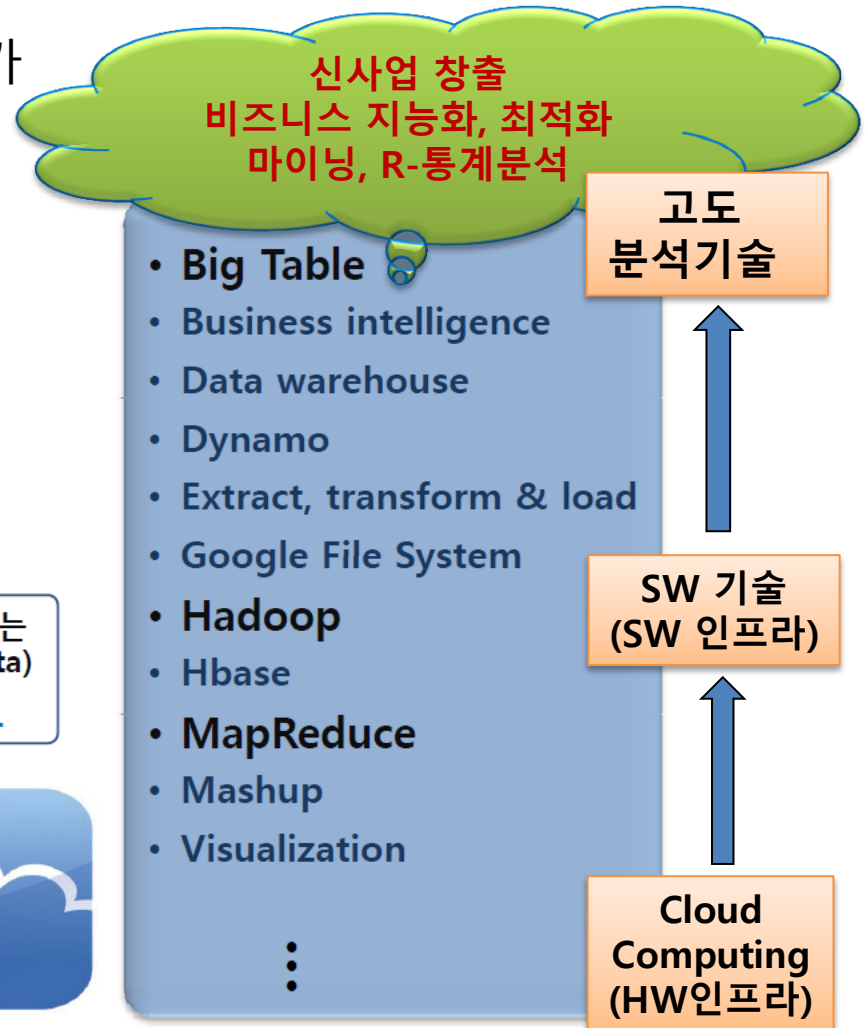
- Cluster Analysis
- Data Mining
- Network Analysis
- Pattern Recognition
- Regression
- Simulation
- Cluster Analysis
- Crowd-sourcing
- Predictive Modeling
- ⋮



기존 방식으로 다룰 수 없는 큰 규모의 데이터 (Big data)
→ 새로운 솔루션 필요

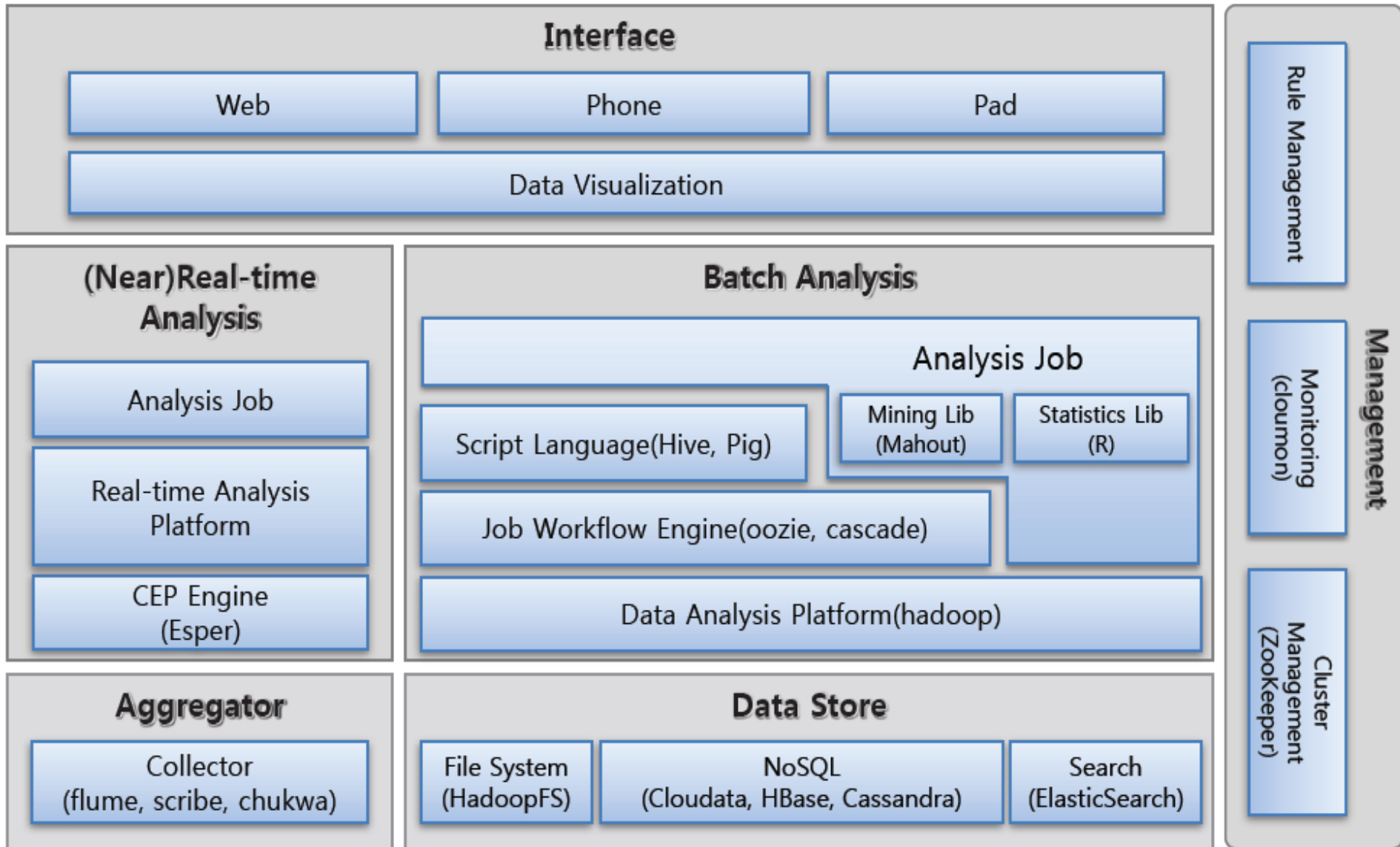


클라우드 기반 솔루션으로 전환

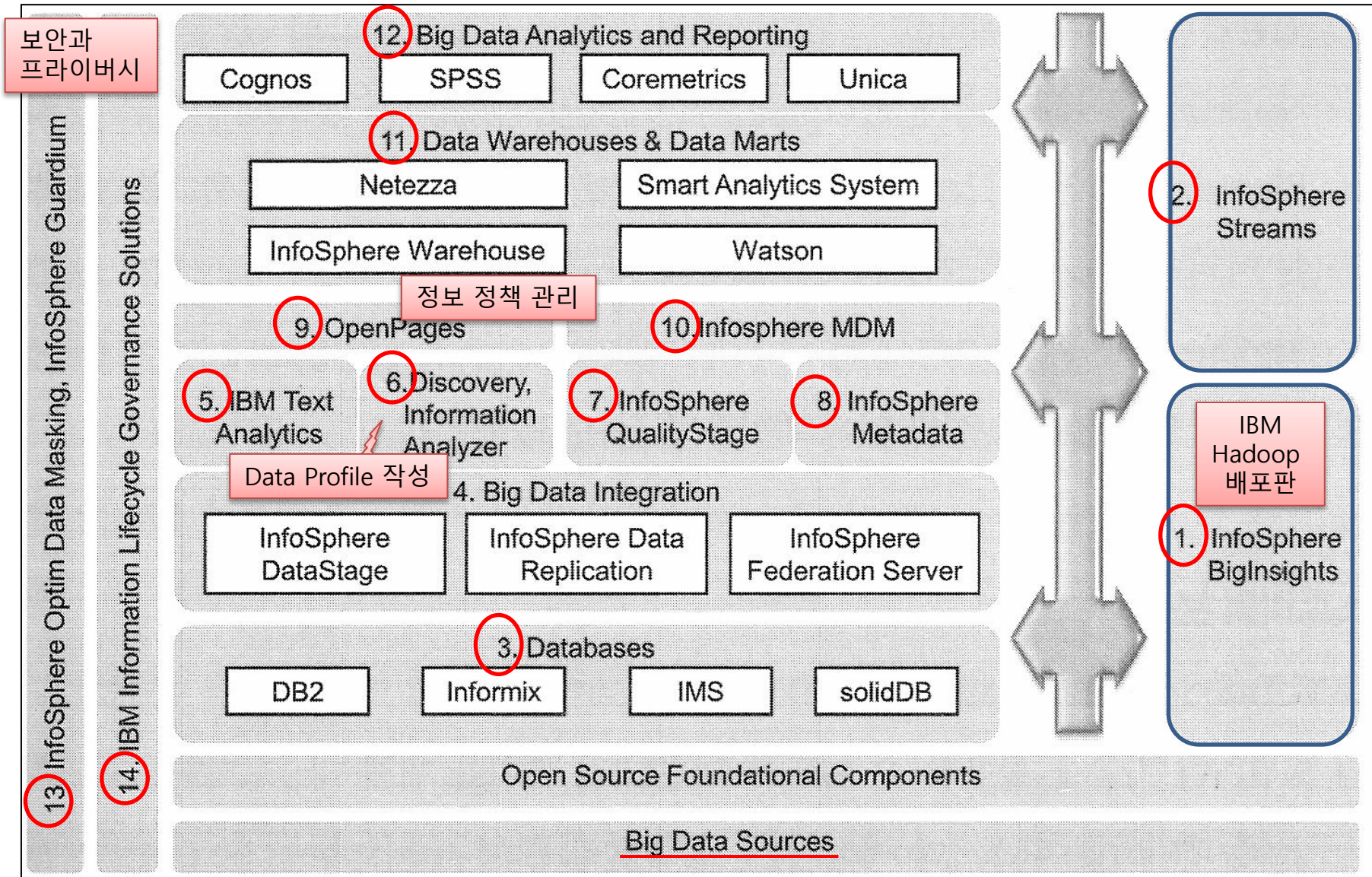


Big Data 처리 기술들

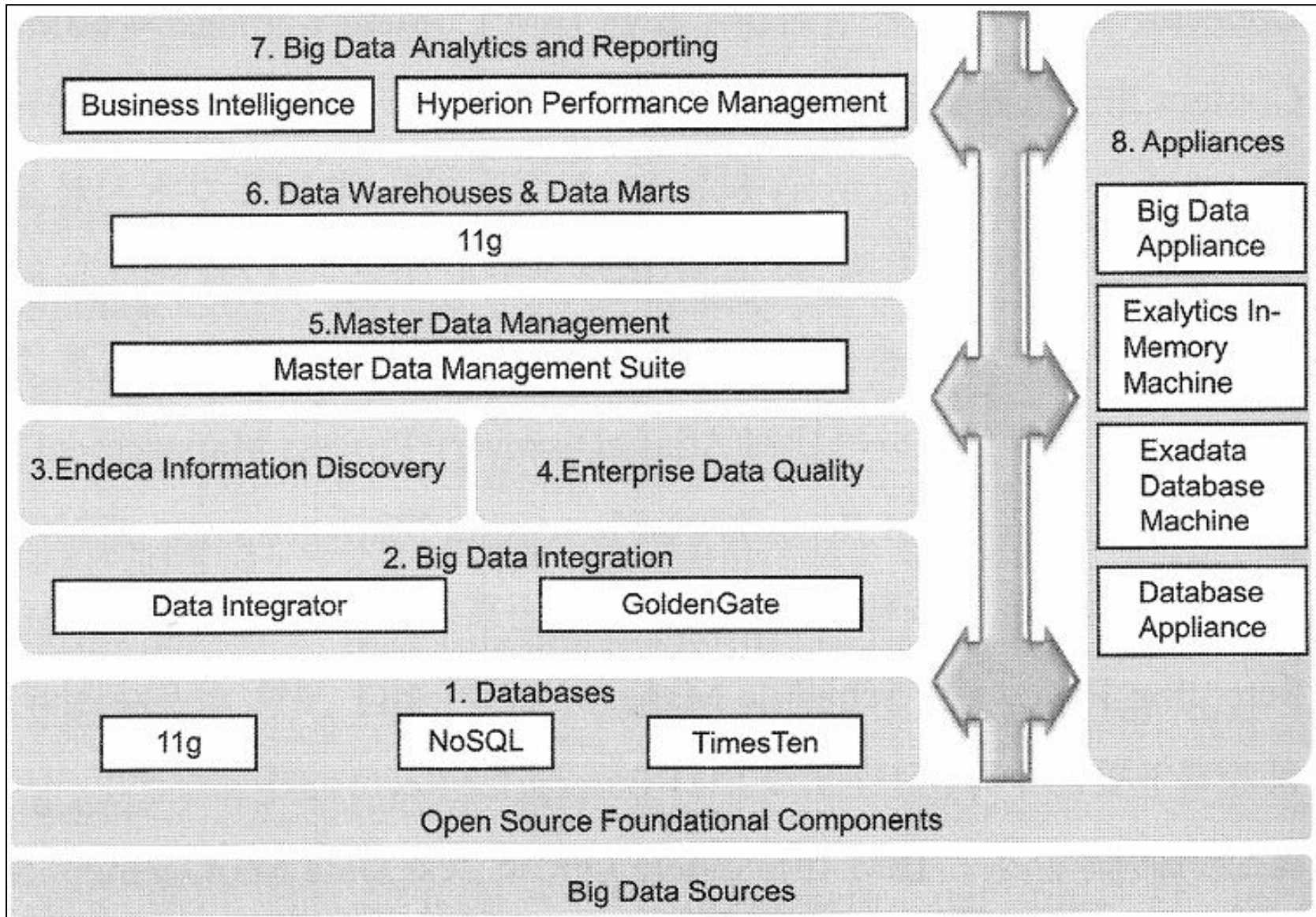
Big Data 관련기술 (요약)



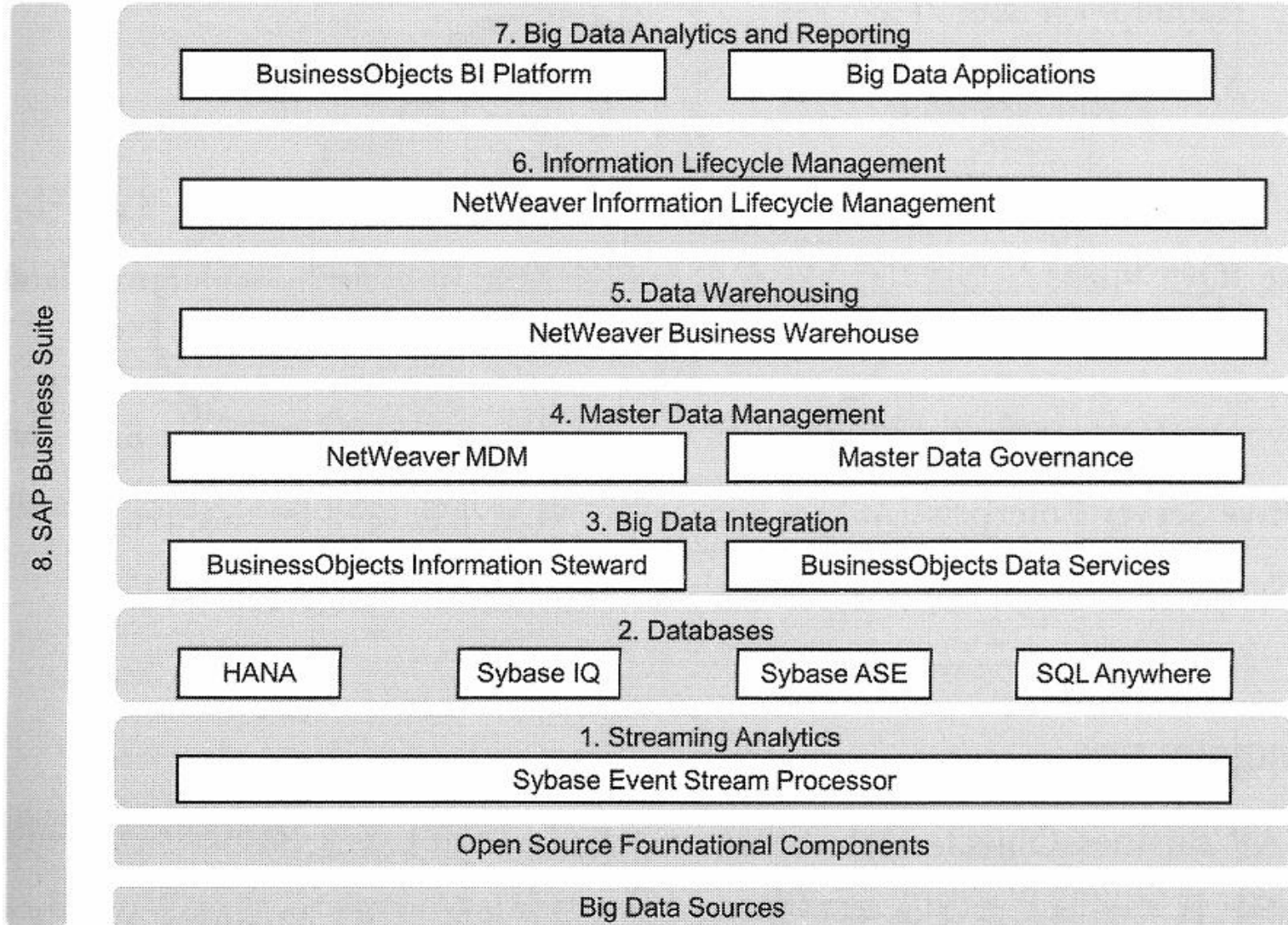
Big Data Platform - IBM



Big Data Platform – Oracle

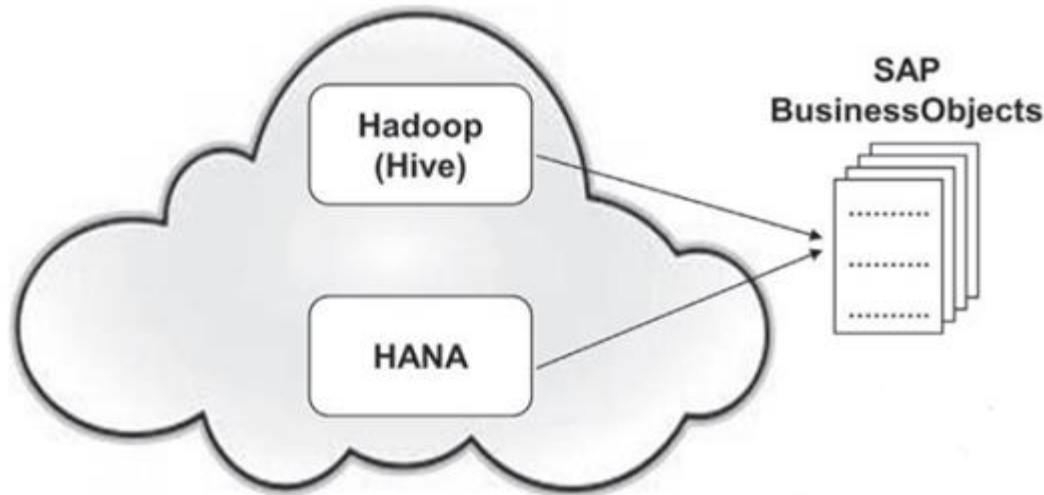


Big Data Platform – SAP



Big Data Platform – SAP

- SAP BusinessObjects
 - 클라우드 상의 Hadoop과 HANA 인스턴스에 대해 연합 쿼리를 작성할 수 있음
- SAP Smart Meter Analytics
 - 엄청난 양의 공공 요금 M2M 데이터를 분석하고 HANA에서 실행함



Big Data Platform

- MS 사

- MS SQL : DBMS
- Microsoft SQL Server Integration Services (SSIS®) : ETL 툴
- Microsoft SQL Server Parallel Data Warehouse 용 쌍방향 Hadoop 커넥터 발표
- Microsoft SQL Server Reporting Services (SSRS) : 분석 툴
- Microsoft의 Hive Open Database Connectivity (ODBC) 드라이버는 Microsoft SQL Server Analysis Services (SSAS), PowerPivot, Power View의 사용자들이 Hadoop 데이터와 상호작용할 수 있도록 지원함
- 또한 Microsoft의 엑셀용 Hive 추가 기능은 사용자들이 스프레드시트 환경에서 Hadoop과 상호작용할 수 있게 지원함

=> **Power BI**

빅데이터 기술

요소기술	설 명	해당 기술
빅데이터 수집	조직내부와 외부의 분산된 여러 데이터 소스로부터 필요로 하는 데이터를 검색하여 수동 또는 자동 으로 수집하는 과정과 관련된 기술로 단순 데이터 확보가 아닌 검색/수집 /변환을 통해 정제된 데이터를 확보하는 기술	<ul style="list-style-type: none"> ■ ETL ■ 크롤링 엔진 ■ 로그 수집기 ■ 센싱 ■ RSS, Open API 등
빅데이터 공유	서로 다른 시스템간의 데이터 공유	<ul style="list-style-type: none"> ■ 멀티 테넌트 데이터 공유 ■ 협업 필터링 등
빅데이터 저장	작은 데이터라도 모두 저장하여 실시간으로 저렴하게 데이터를 처리하고, 처리된 데이터를 더 빠르고 쉽게 분석 하도록 하여, 이를 비즈니스 의사 결정에 바로 이용하는 기술	<ul style="list-style-type: none"> ■ 병렬 DBMS ■ 하둡(Hadoop) ■ NoSQL 등
빅데이터 처리	엄청난 양의 데이터의 저장·수집·관리·유통·분석을 처리 하는 일련의 기술	<ul style="list-style-type: none"> ■ 실시간 처리 ■ 분산 병렬 처리 ■ 인-메모리 처리 ■ 인-데이터베이스 처리

빅데이터 기술

요소기술	설 명	해당 기술
빅데이터 분석	데이터를 효율적으로 정확하게 분석하여 비즈니스 등의 영역에 적용하기 위한 기술로 이미 여러 영역에서 활용해온 기술임	<ul style="list-style-type: none"> ■ 통계 분석 ■ 데이터 마이닝 ■ 텍스트 마이닝 ■ 예측 분석 ■ 최적화 ■ 평판 분석 ■ 소셜 네트워크 분석 등
빅데이터 시각화	자료를 시각적으로 묘사하는 학문으로 빅데이터는 기존의 단순 선형적 구조의 방식으로 표현하기 힘들기 때문에 빅데이터 시각화 기술이 필수적임	<ul style="list-style-type: none"> ■ 편집 기술 ■ 정보 시각화 기술 ■ 시각화 도구

빅데이터
표준

시스템
관리기술

빅데이터
전송기술

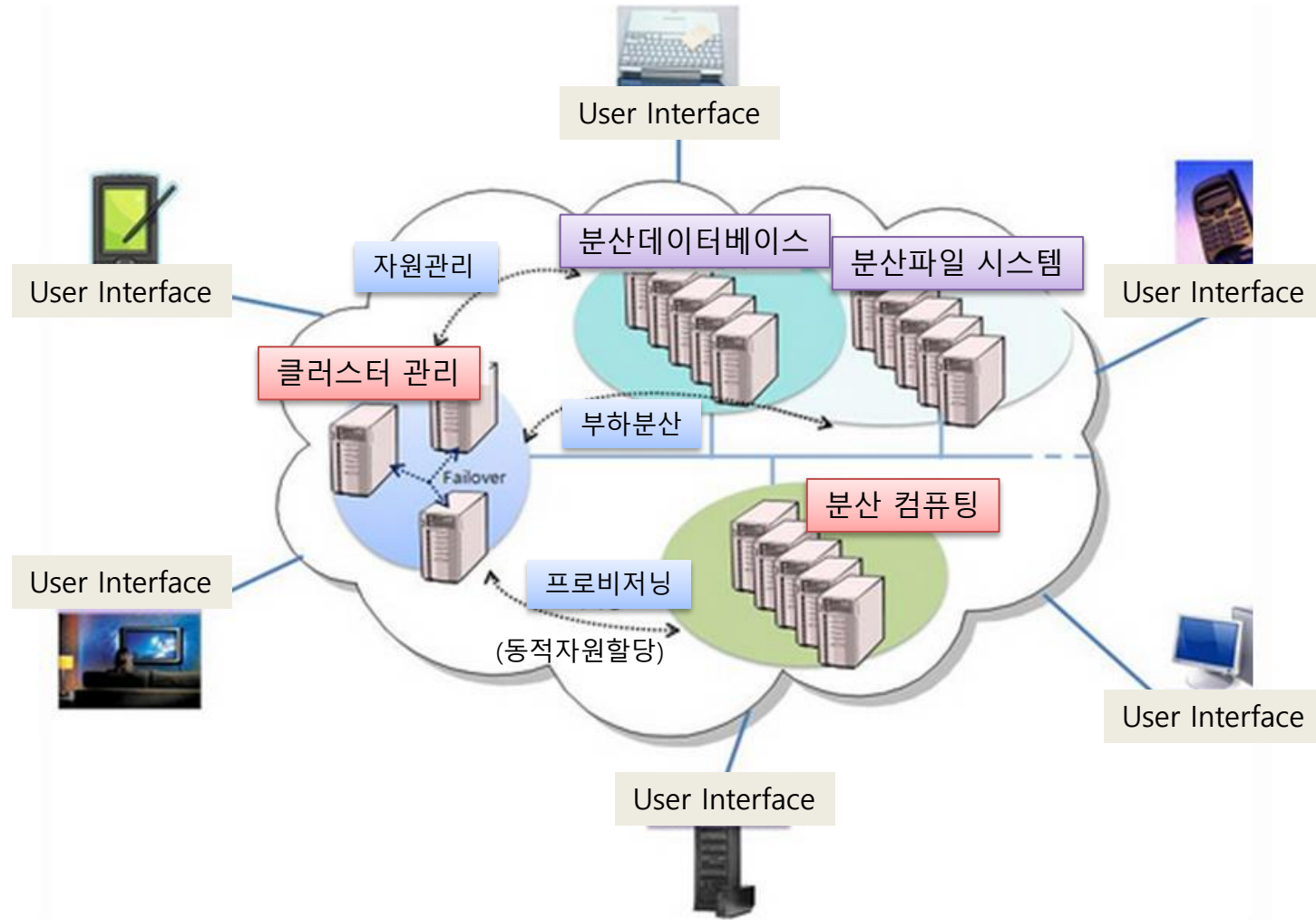
클라우드
핵심 SW(open)

Big Data 관련기술 (요약)

구분	설명	오픈 소스
데이터 수집	데이터 발생원으로부터 안정적인 저장소로 저장하는 기능 수행	Flume, Scribe, Chukwa
원본 데이터 저장	수집된 데이터를 안정적으로 저장하는 저장소 비구조적 데이터 저장소로 주로 대용량 파일 저장소	Hadoop FileSystem MogileFS
트랜잭션 데이터 저장	원본 데이터를 실시간으로 저장, 조회 처리를 하기 위한 저장소 구조적 저장소 또는 검색 엔진 기술을 활용	NoSQL(Cloudata, HBase, Cassandra) Katta, ElasticSearch
실시간 분석 플랫폼	데이터 수집과 동시에 분석을 수행 복잡한 분석보다 count, sum 등 단순한 aggregation 연산 정도 수행	S4, Storm
배치 분석 플랫폼	전체 또는 부분 데이터에 대해 복잡하고 다양한 분석 수행 대용량 처리를 위해 분산, 병렬처리가 필요 단순 텍스트 분석부터 그래프 분석까지 다양한 분석 모델 지원	Hadoop MapReduce(Hive, Pig) Giraph, GoldenOrb
데이터 마이닝/통계 도구	Cluster, Classification 등과 같이 데이터 마이닝을 위한 기본 알고리즘 라이브러리 및 도구	Mahout, R
클러스터 관리 및 모니터링	대부분 분산 시스템으로 구성되기 때문에 전체 클러스터에 대한 관제 및 모니터링도 복잡	ZooKeeper, HUE, Cloumon
데이터 Serialization	이기종 플랫폼 및 다양한 종류의 솔루션을 사용하기 때문에 데이터 전송 및 처리에 대한 표준 프레임워크 필요	Thrift, Avro, ProtoBuf

클라우드 컴퓨팅

- Cloud Computing Platform



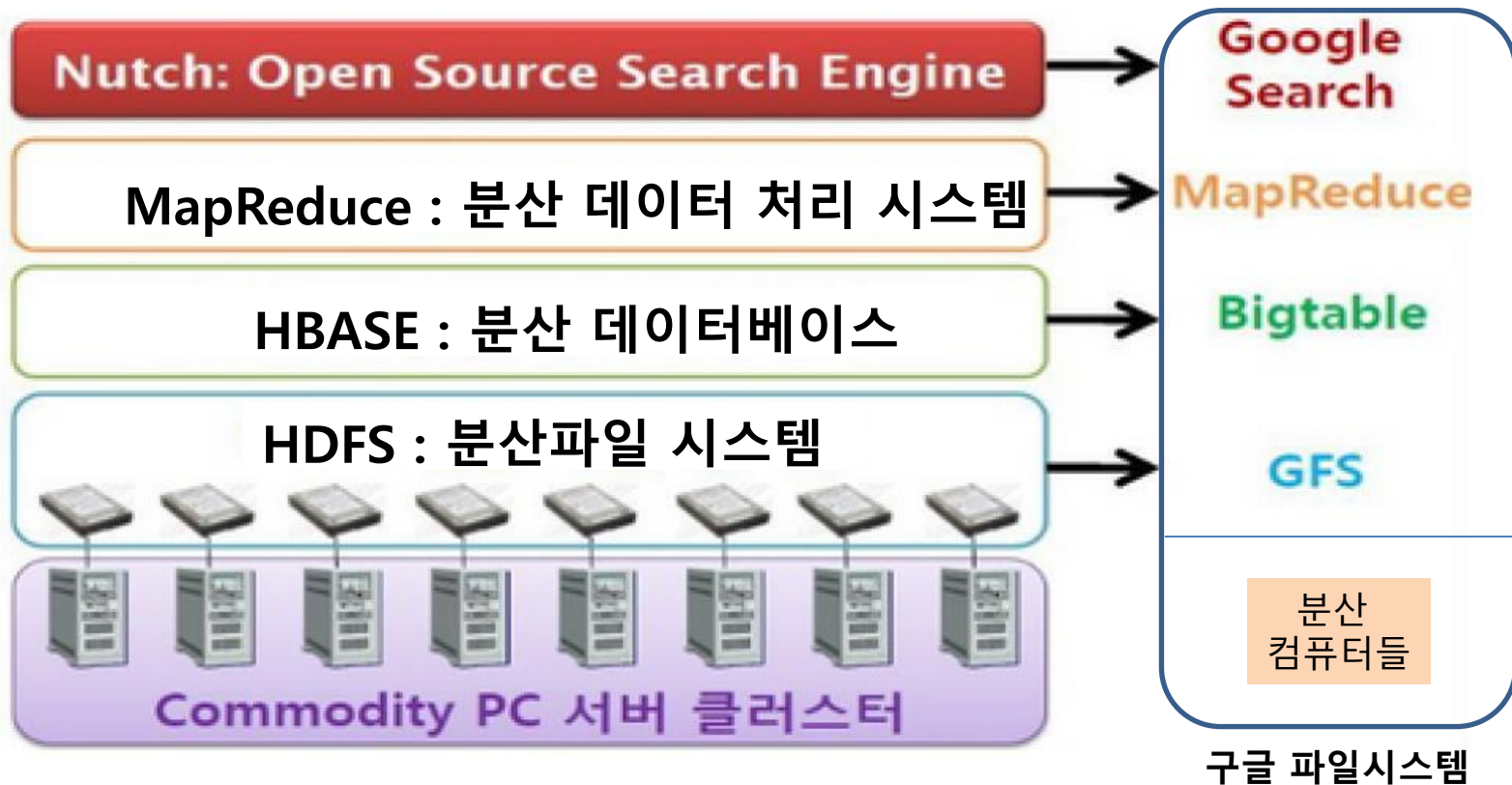
Hadoop

- 연결된 여러 컴퓨터에 데이터를 분산하여 저장, 처리한 후, 통합하여 최종 결과를 생성하는 고성능 컴퓨팅 기술
 - Hadoop 분산파일 시스템(HDFS : Hadoop Distributed File System)
 - 파일을 적당한 크기로 나눠서 각 노드에 분산 저장함
 - 데이터 유실이나 부하 분산을 위해 각 블록의 복사본 (Replication)을 유지함
 - MapReduce 프레임워크
 - Map함수는 데이터를 여러 개의 데이터 조각으로 나눠서 여러 대의 컴퓨터에서 분산 처리함
 - Reduce함수 각 노드가 계산한 결과를 하나로 통합함



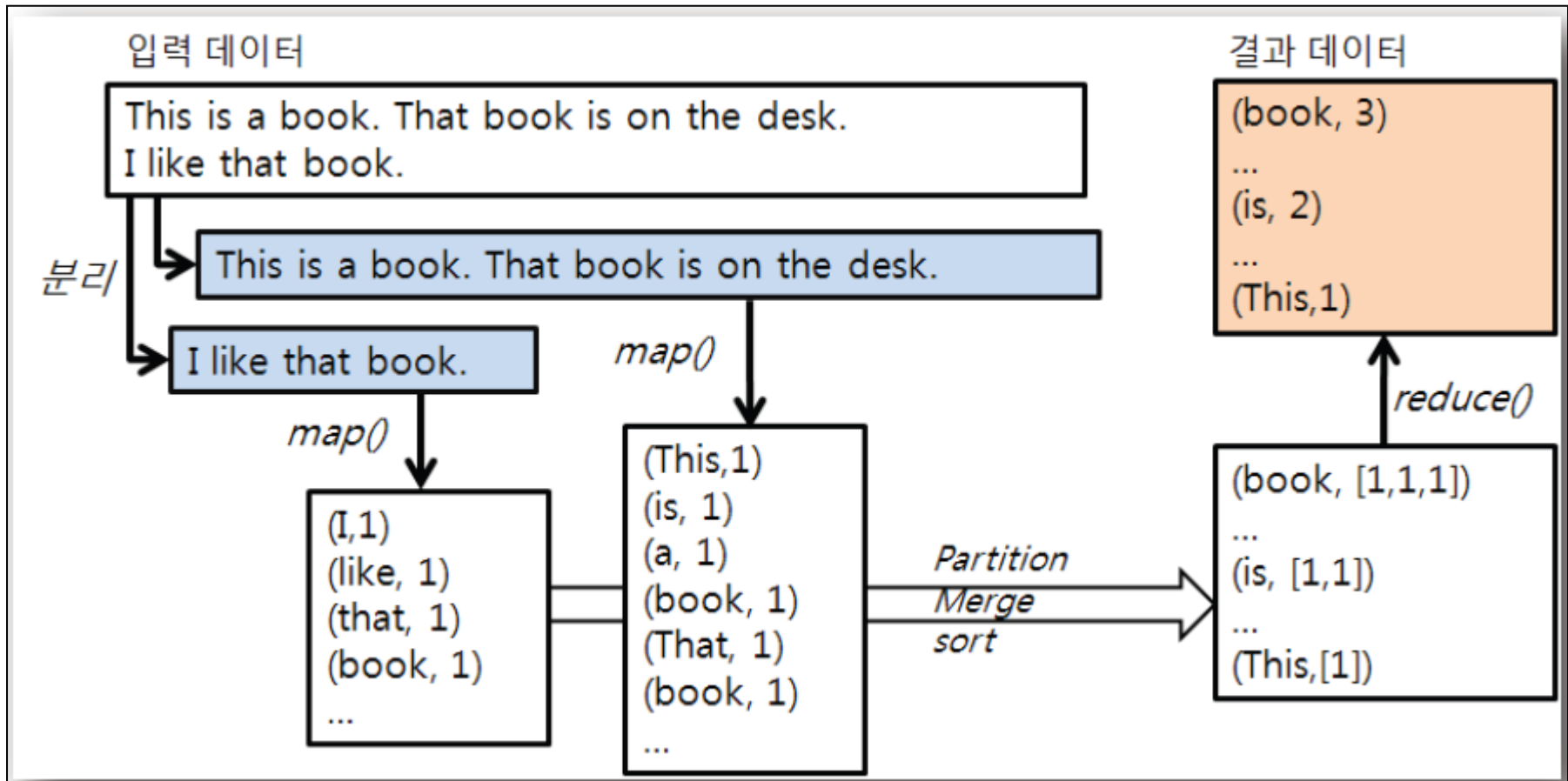
Hadoop

- Hadoop 과 구글 파일 시스템



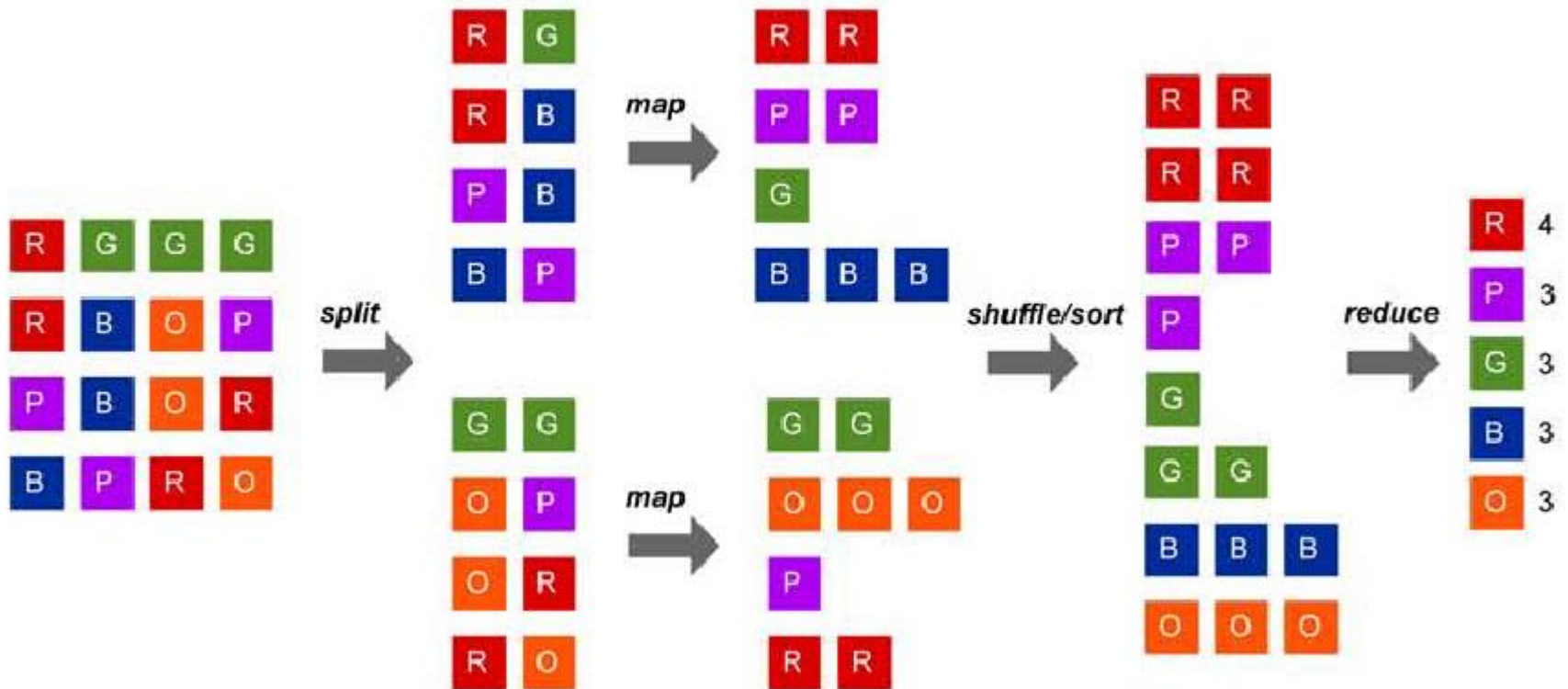
Hadoop

- MapReduce를 활용한 단어 출현횟수 counting



Hadoop

- **MapReduce 기술** – 컬러 사각형 개수 counting



ETL 도구

- ETL (Extraction, Transformation, and Loading)
 - 다양한 데이터 소스로부터 데이터를 추출하여 정제 (표준화)하여 통합하고 데이터베이스에 로딩하는 도구

< 주요 데이터 수집 기술 >

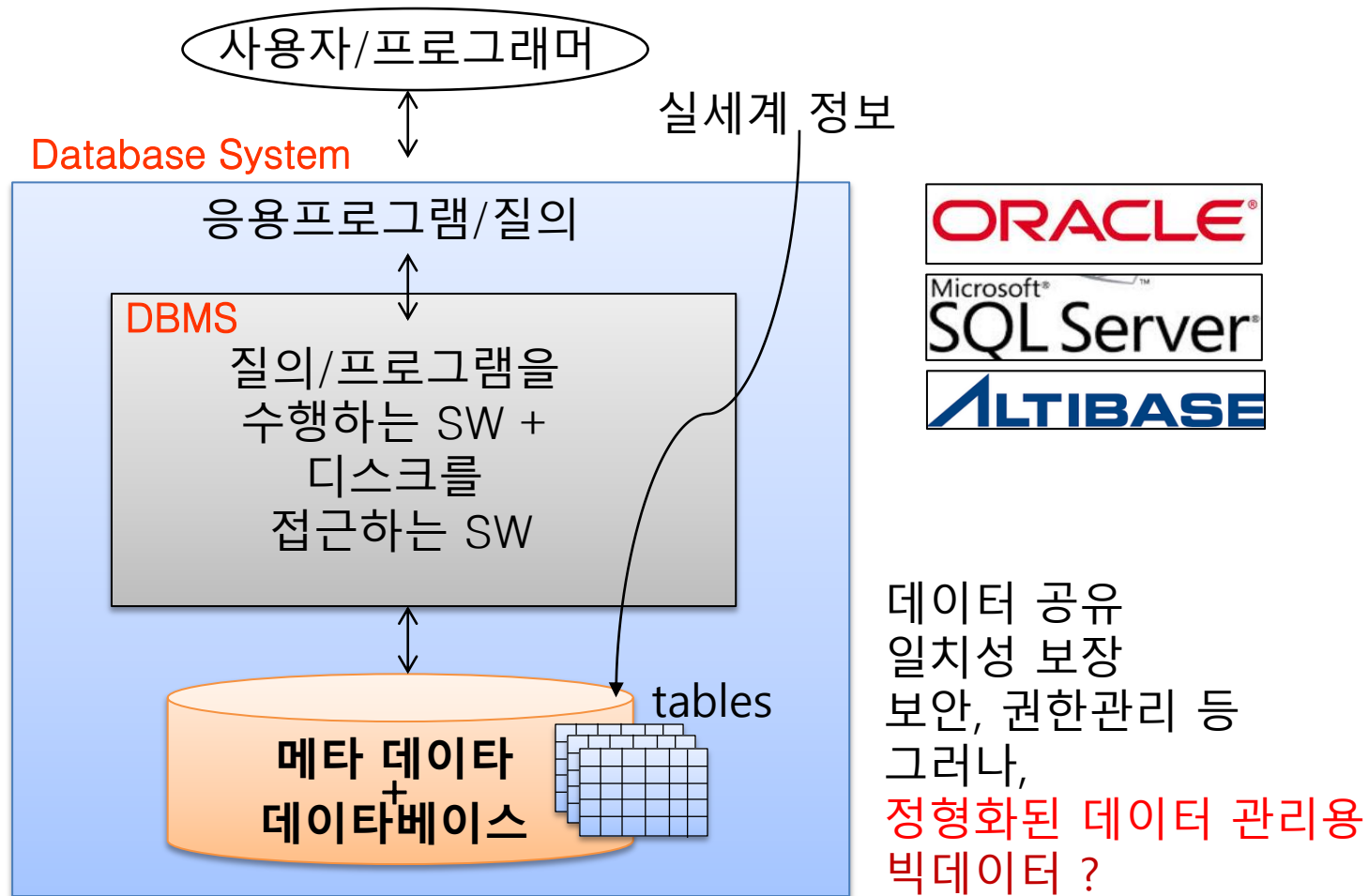
구 분	특징	비고
Crawling	- SNS, 뉴스, 웹 정보 등 인터넷상에서 제공되는 웹문서·정보 수집	웹문서 수집
FTP	- TCP/IP 프로토콜을 활용하는 인터넷 서버로부터 각종 파일들을 송수신 - 보안을 강화하기 위해 SFTP 사용 고려 - 서버간 연동시에는 전용 네트워크 구축 고려	File 수집
Open API	- 서비스, 정보, 데이터 등을 어디서나 쉽게 이용할 수 있도록 개방된 API로 데이터 수집방식 제공 - 다양한 어플리케이션을 개발할 수 있도록 개발자와 사용자에게 공개	실시간 데이터 수집

ETL 도구

RSS	<ul style="list-style-type: none"> - RSS(Really Simple Syndication)는 Web기반 최신의 정보를 공유하기 위한 XML 기반 콘텐츠 배급 프로토콜 	콘텐츠 수집
Streaming	<ul style="list-style-type: none"> - 인터넷에서 음성, 오디오, 비디오 데이터를 실시간으로 수집할 수 있는 기술 	실시간 데이터 수집
Log Aggregator	<ul style="list-style-type: none"> - 웹서버 로그, 웹 로그, 트랜잭션 로그, 클릭 로그, DB의 로그 등 각종 로그 데이터를 수집하는 오픈 소스 기술 - 종류 : Chukwa, Flume, Scribe 등 	로그수집
RDB Aggregator	<ul style="list-style-type: none"> - 관계형 데이터베이스에서 정형 데이터를 수집하여 HDFS(하둡 분산파일시스템)이나 HBase와 같은 NoSQL에 저장하는 오픈 소스 기술 - 종류 : Sqoop, Direct JDBC/ODBC 등 	RDB 기반 데이터 수집

데이터베이스 시스템

- 기존 database systems



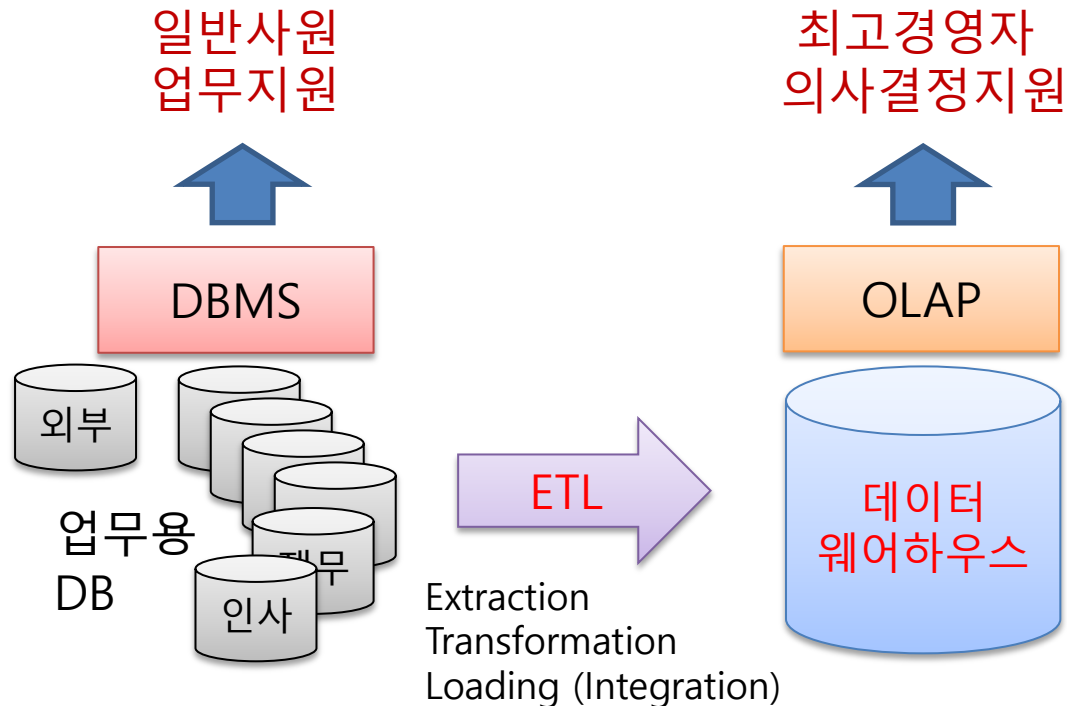
Big Data 용 DBMS

- **NoSQL** (Not Only SQL)
 - 기존의 RDBMS는 빅데이터 처리에 적합하지 않음
 - NoSQL은 빅데이터 처리에는 유용하지만 RDBMS에서 제공하는 데이터의 일관성과 유효성은 보장하지 않음 (Read only, 분석용으로 적합)
 - 20개 이상의 NoSQL 제품이 보급됨
 - **BigTable**, Dynamo, **Cassandra**, CouchDB, **MongoDB**, Hbase, Riak, Voldemort 등
 - Oracle NoSQL DB2.0 출시 (2012.01)

다차원 분석

- **Data Warehouse(DW), OLAP**

- DW : 의사결정용 데이터베이스
- OLAP (On-line Analytical Processing)은 data warehouse 상에서 온라인 다차원 분석처리를 지원하는 도구

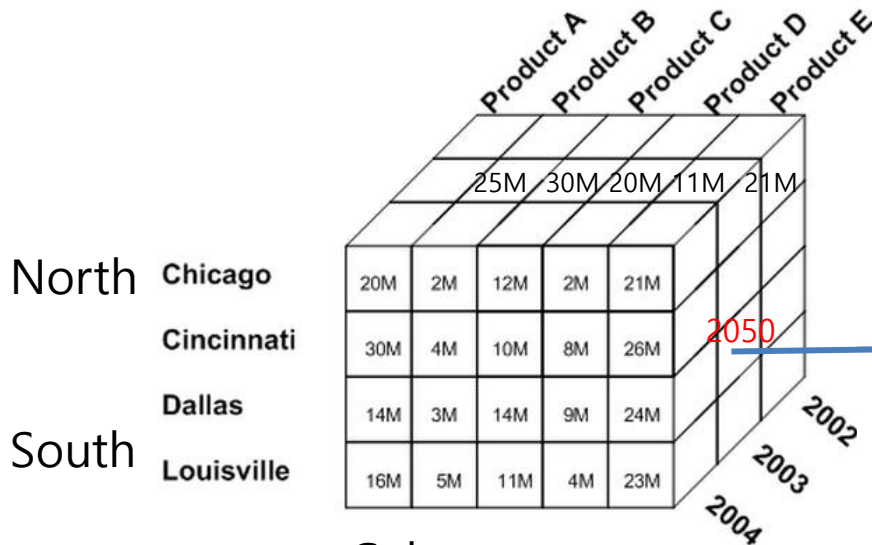


다차원 분석 - 사례

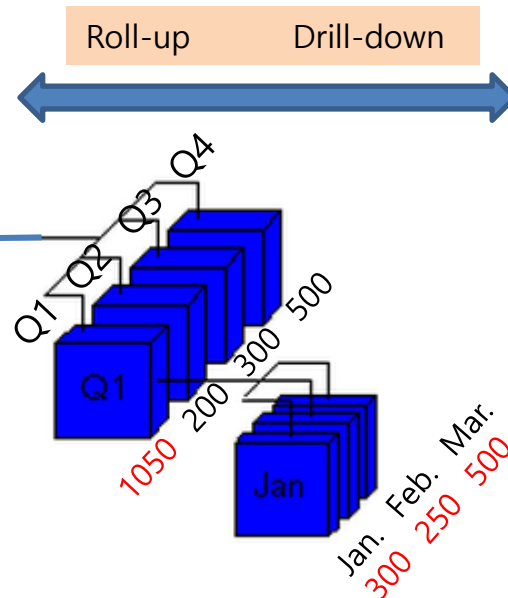
매출액
\$500억



년도별 매출액은 ?
 2003년 분기별 매출액은 ?
 2003년 1분기 월별 매출액은 ?
 도시별 매출액은 ?
 North/South 지역별 매출액은 ?
Dallas 지역의 Product E에 대한 2003년 분기별 매출액은 ?
 도시별로 2003년 월별 매출액은 ?
 제품별, 년도별 매출액은 ?
 ... (수많은 분석)



Cube
수백개 cubes



다차원 분석

- **Hive**(<http://hadoop.apache.org/hive/>)
 - Data warehouse system으로 Hadoop compatible file systems에 저장된 large datasets에 대하여 간편한 data summarization, ad-hoc queries, analysis 기능을 제공함 (OLAP)
 - SQL-like 언어로 하둡상에서 맵리듀스 처리를 간단하게 작성, 실행할 수 있는 툴
 - Open source (Apache License), ANSI SQL 지원
 - Facebook의 Main Data Warehousing System으로 활용됨
- **Pig**(<http://hadoop.apache.org/pig/>)
 - 스크립트 언어 형태로 MapReduce 프로그램을 개발할 수 있는 플랫폼으로 현재 Hadoop의 서브 프로젝트
 - Cascading과 유사하게 고수준의 처리를 위한 쉬운 문법을 제공하며 스크립트 형태이기 때문에 빠른 프로그래밍 및 피드백이 가능함 (Yahoo에서 주도적으로 개발)

다차원 분석

- 배치 분석 : Script language



Hive

```
Hive> CREATE TABLE invites (foo INT, bar STRING) PARTITIONED BY (ds STRING);
hive> LOAD DATA LOCAL INPATH './examples/files/kv1.txt' OVERWRITE INTO TABLE invites;
hive> SELECT a.foo FROM invites a WHERE a.ds='2008-08-15';
hive> FROM pokes t1 JOIN invites t2 ON (t1.bar = t2.bar) INSERT OVERWRITE TABLE events SEL
ECT t1.bar, t1.foo, t2.foo;
```



Pig

```
Visits = load '/data/visits' as (user, url, time);
Visits = foreach Visits generate user, Canonicalize(url), time;
Pages = load '/data/pages' as (url, pagerank);
VP = join Visits by url, Pages by url;
UserVisits = group VP by user;
UserPageranks = foreach UserVisits generate user,
AVG(VP.pagerank) as avgpr;
GoodUsers = filter UserPageranks by avgpr > '0.5';
store GoodUsers into '/data/good_users';
```

다차원 분석 - 실시간

- **Streaming data processing**
 - ETL, Storing, Processing, Mining, Alarming
 - In-Memory database systems
 - ESPER
 - 하드 디스크에 저장하지 않고 바로 메모리상에서 분석함으로써 실시간 분석결과를 도출함
- **Spark**
 - In-Memory based multidimensional analysis system
 - Hive 보다 빠른 속도로 빅데이터 다차원 분석을 지원함

R 언어

- **R 언어 – SAS, SPSS의 대안으로 급성장**
 - 데이터를 다루고 분석하는데 특화된 언어
 - 통계분석 (마이닝) 및 시각화를 위한 언어 혹은 개발환경으로 오픈 소스임
 - 수천가지 라이브러리를 사용하여 기본적인 통계분석 기법부터 모델링, 최신 데이터 마이닝 기법까지 구현/개선이 가능
 - 다양한 분야에 적용 가능한 마이닝 기법들을 제공함
 - 마케팅, 금융, CRM(고객관리), 생명공학 및 의학, GIS, Stream Data 등
 - Java, C/C++, Python 등 다른 프로그래밍 언어와의 연결 용이
 - 빅데이터 분석이 필요한 조직에서 대용량 데이터 통계분석과 데이터 마이닝을 위한 솔루션으로 정착되고 있음
 - 예: 약 30 라인으로 SNS (Twitter) 분석에서 시각화까지 구현할 수 있음

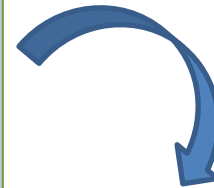
R 언어

Example : R을 사용한 트위터 팔로워 분석과 결과 : 단 20여 line으로 결과생성 !

```

library(twitteR)
library(KoNLP)
library(wordcloud)
library(tm)
gogamza <- getUser("gogamza")
gogamza.followers <- gogamza$getFollowers()

#팔로워들의 자기 소개를 벡터에 적재 한다.
followerDesc <- c()
for(i in gogamza.followers){
  followerDesc <- append(followerDesc, i$description)
}
#쓸모없는 문자들을 제거한다.
followerDesc <- gsub("\n","", followerDesc)
followerDesc <- gsub("\r", "", followerDesc)
nouns <- Map(extractNoun, followerDesc)
wordsvect <- unlist(nouns, use.name=F)
#쓸모없는 문자들을 제거한다. 특히 영문자의 경우 tm의 stopwords를 활용한다.
wordsvect <- wordsvect[-which(wordsvect %in% stopwords("english"))]
wordsvect <- gsub("[[:punct:]]","", wordsvect)
wordsvect <- Filter(function(x){nchar(x)>=2}, wordsvect)
wordcount <- table(wordsvect)
pal <- brewer.pal(8,"Dark2")
wordcloud(names(wordcount),freq=wordcount,scale=c(4,0.5),min.freq=10,
  random.order=T,rot.per=.1,colors=pal)
  
```



RHive

- RHive = R + Hive



- ◆ Language : R or ANSI-SQL
- ◆ R-Hive Bridge
- ◆ R Export
- ◆ R 기반 분산 처리 Framework

- ◆ 가장 널리 사용하는 Analytic Tool
- ◆ CRAN : 4,000+ Rich R library Set
- ◆ 용이한 Library/Procedure 제작
- ◆ 다양한 Visualization, IDE 도구

- ◆ Hadoop 기반 분산 병렬 처리
- ◆ ANSI SQL : Low Learning Cost
- ◆ 용이한 기능 확장 : UDF, UAF

KRUG (Korean R Users Group)

GNU의 철학에 입각하여, R을 한국어 사용자가 올바르게 쉽게 사용 될 수 있도록 문서를 번역하고 지식과 기술을 공유하는 사용자 모임; 2007년 1월부터 공식적으로 활동

빅데이터 - 마이닝

● 데이터 마이닝

- 데이터로부터 체계적이고 자동적으로 (통계학에서 패턴 인식에 이르는 다양한 계량 기법을 사용하여) 규칙이나 패턴을 찾아 내는 기술
- 데이터 마이닝 기법은 통계학에서 발전한 탐색적자료분석, 가설 검증, 다변량 분석, 시계열 분석, 일반선형모형 등의 방법론과 데이터베이스 측면에서 발전한 OLAP (온라인 분석 처리: On-Line Analytic Processing), 인공지능 진영에서 발전한 SOM(Self Organizing Maps), 신경망, 전문가 시스템 등의 기술적인 방법론이 사용됨

● 빅데이터 마이닝

- 빅데이터의 특징인 “대용량, 비정형, 실시간” 마이닝이 가능하도록 기존의 기법들을 확장함
- IoT/M2M 등의 보편화로 Stream Data Mining 기술의 필요성 증대

빅데이터 - 마이닝

● 데이터 마이닝 기술의 종류

– 분류(Classification)

- 일정한 데이터 집단에 대한 특성 정의를 기준으로 분류함
- 예: 경쟁자에게로 이탈한 고객들을 분류함

– 군집화(Clustering)

- 어떤 특성을 공유하는 데이터 그룹을 찾음.
- 군집화는 미리 정의된 특성에 대한 정보를 갖지 않는다는 점에서 분류와 구분됨 (예 : 유사 행동 집단의 구분)

– 연관관계(Association)

- 관련이 있는 (동시에 발생하는) 데이터들을 찾아냄
- 예: 장바구니에서 함께 구매되는 상품들

빅데이터 - 마이닝

- **데이터 마이닝 기술의 종류 (계속)**

- **순차패턴(Sequencing)**

- 순차적으로 발생하는 데이터들을 찾아냄
 - 예: 비디오 대여 순서정보

- **예측(Forecasting)**

- 대용량 데이터 집합내의 패턴을 기반으로 미래를 예측
 - 예: 수요예측

- **텍스트 마이닝**

- 자연어 처리 기술을 기반으로 텍스트의 의미를 자동으로 파악함
 - 예: SNS 기반의 고객감성 분석

텍스트 마이닝

● 정의

- 텍스트 기반의 데이터(문서)로부터 새로운 정보를 발견하는 기술
- 다음과 같은 자연어 처리 기술도 필요함
 - 형태소 분석 기술 (morpheme processing technique)
 - 구문 분석 기술 (syntactic processing technique)
 - 문맥 처리 기술 (context processing technique)
 - 의미 처리 기술 (semantic processing technique)
 - 문장 합성 기술 (sentence generation technique)

● 관련분야

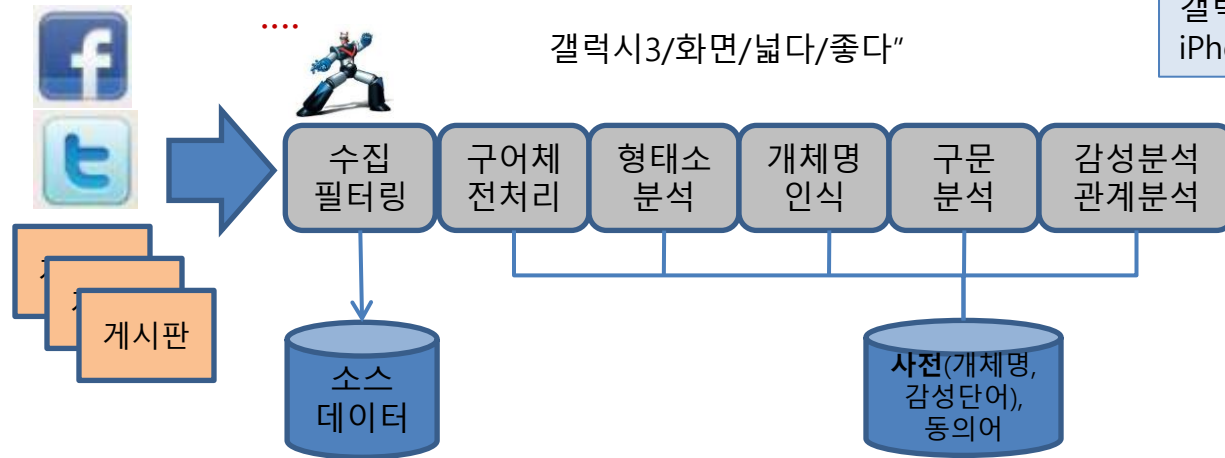
- Data Mining : 데이터로부터 새로운 패턴이나 지식 발견
- Web Mining : 웹 상의 데이터를 분석하여 지식을 발견하는 기술
- Statistics : 통계학
- Information Retrieval : 정보 검색
- Computational Linguistic & NLP : 자연 언어 분석 및 활용 학문

텍스트 마이닝

• 응용분야

- SNS 분석

"갤럭시 3 화면이 넓어 짱이다"
 "iPhone4 무게가 가벼워 좋아"



화면이 크다	45%
디자인이 예쁘다	25%
크기가 작다	15%
화질이 좋다	15%

	긍정	부정
갤럭시3	30,599	1,500
iPhone4	56,456	3,500

- 테러, 범죄, 재난
 조기예측에 응용됨

갤럭시3 iPhone4 ... 갤3 아폰4	화면 бат데리 무게 화질 ...	좋다 넓다 크다 작다 짱이다
제품T	속성T	감성단어T

Dictionary

빅데이터 시각화 (visualization)

- 시각화란 ?
 - 방대한 양의 자료를 분석해서 한눈에 볼 수 있도록 도표나 차트 등으로 정리하는 것
 - 시각화 자체가 분석의 한
- 시각화의 효과
 - 데이터로부터 정보를 습득하는 시간의 절감으로 즉각적인 상황판단이 가능해짐
 - 자료를 습득하는 사람의 흥미유발과 빠른 확산 촉진
 - 자료를 기억하는데 기여함
- 뉴욕 타임스 (유권자 분석)
 - http://www.nytimes.com/interactive/2012/11/11/sunday-review/counties-moving.html?_r=0

시각화의 예

Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dessiné par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en lettres des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Léguir, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 23 Octobre. Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout, qui avaient été détachés sur Minsk et Mielow et qui rejoignent vers Orescha et Witebsk, avaient toujours marché avec l'armée.

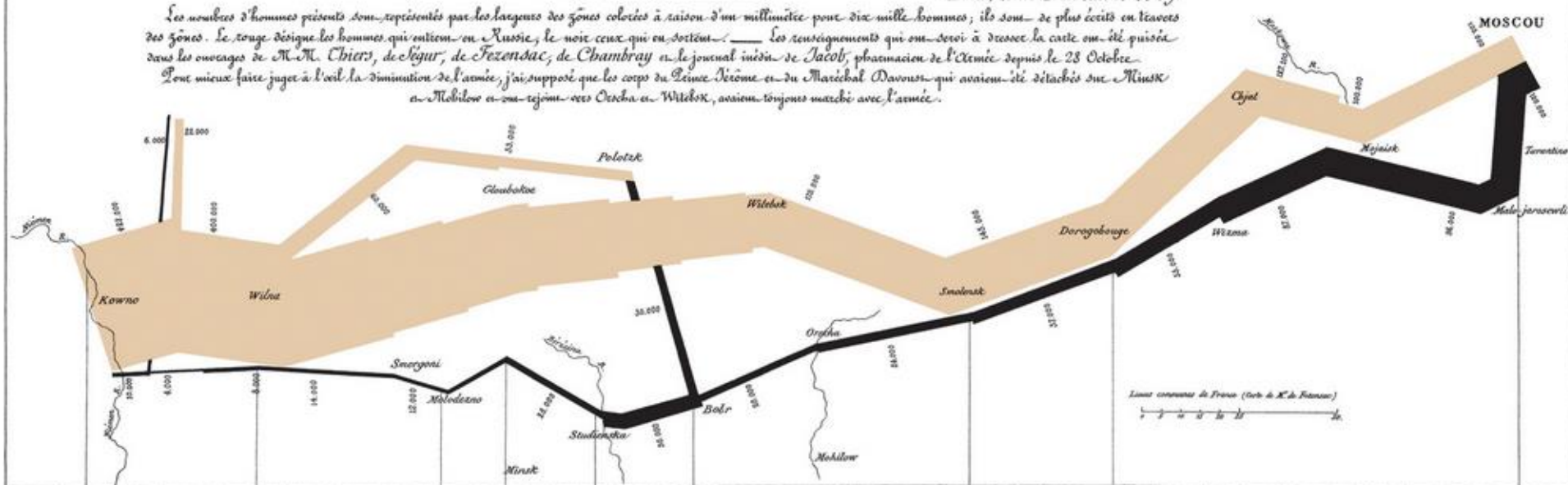
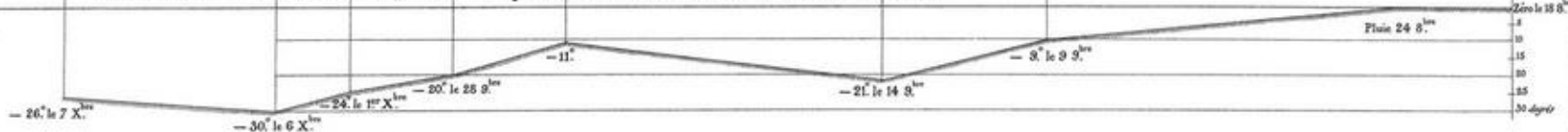


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.



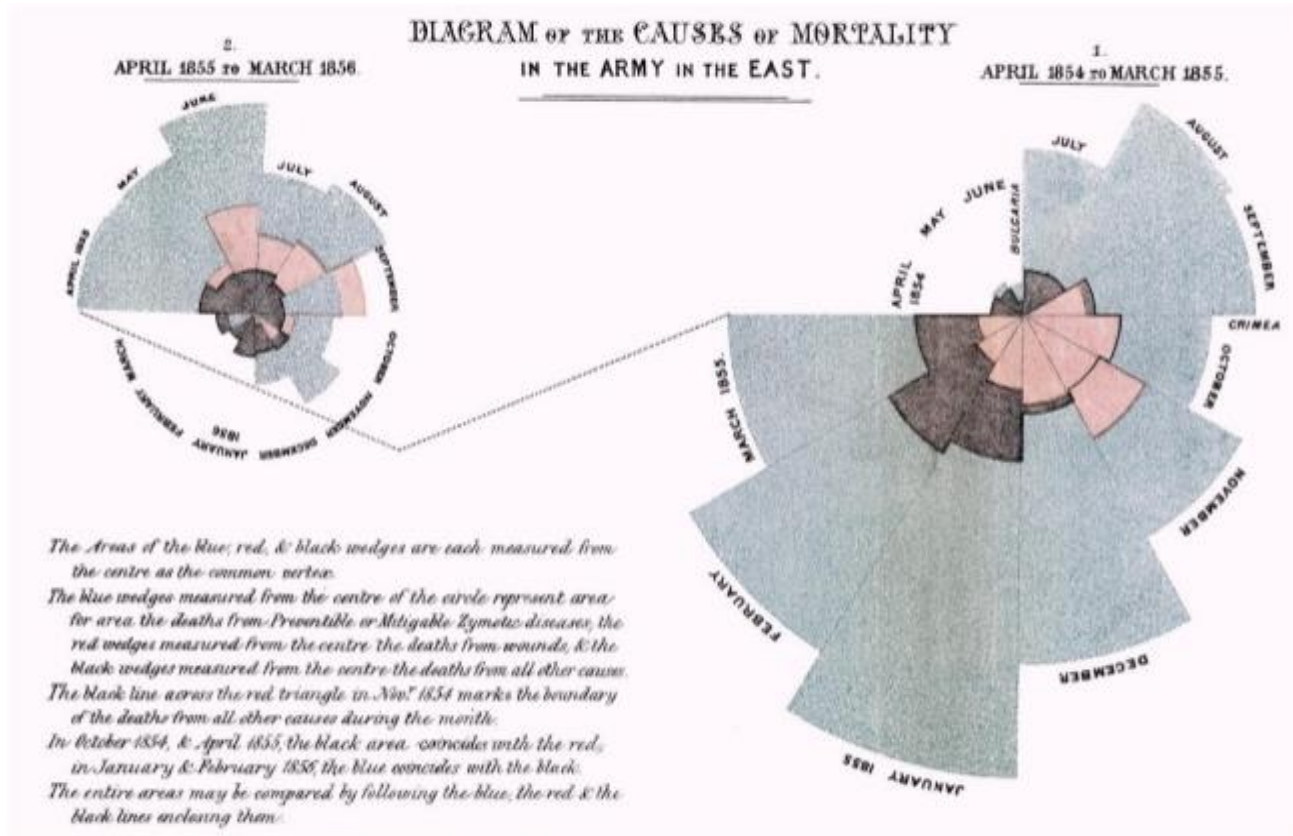
Les Cosaques passent au galop le Nilman gelé.

Dessiné par Régisier, l. Par. 5^{ème} Mars 51 0^{ème} à Paris.

Imp. Lit. Régisier et Deso. dat.

나폴레옹군이 러시아 원정에서 완패하고 모스크바로부터 프랑스로 복귀하는 이동경로 별로 그들이 겪은 추위와 병사감소를 한눈에 시각화 (샤를 미나르, 1861 제작)

시각화의 예



<http://en.wikipedia.org/wiki/File:Nightingale-mortality.jpg>

■ 나이팅게일의 폴라 지역 다이어그램

- 18세기에 전쟁중인 영국의 간호사로 있던 나이팅게일이 전쟁터에서 사망하는 병사보다 열악한 위생 환경 때문에 사망하는 병사가 더 많다는 사실을 깨닫고 이 폴라그래프를 고안해 정부를 설득하는 데 활용
 - * 출처: 빅데이터를 위한 데이터 시각화 (이지선)

시각화 도구

○ 구글 차트

- 속도가 빠르며, 모바일에서도 잘 작동함
- 동적인 데이터인 경우에는 구글 스프레드시트에 데이터를 저장한 후 Visualization Query로 읽어와 차트를 만들 수도 있음
- 잘 모르는 초보자를 위한 위자드 툴도 제공
- 스프레드시트와 차트 API를 함께 씀

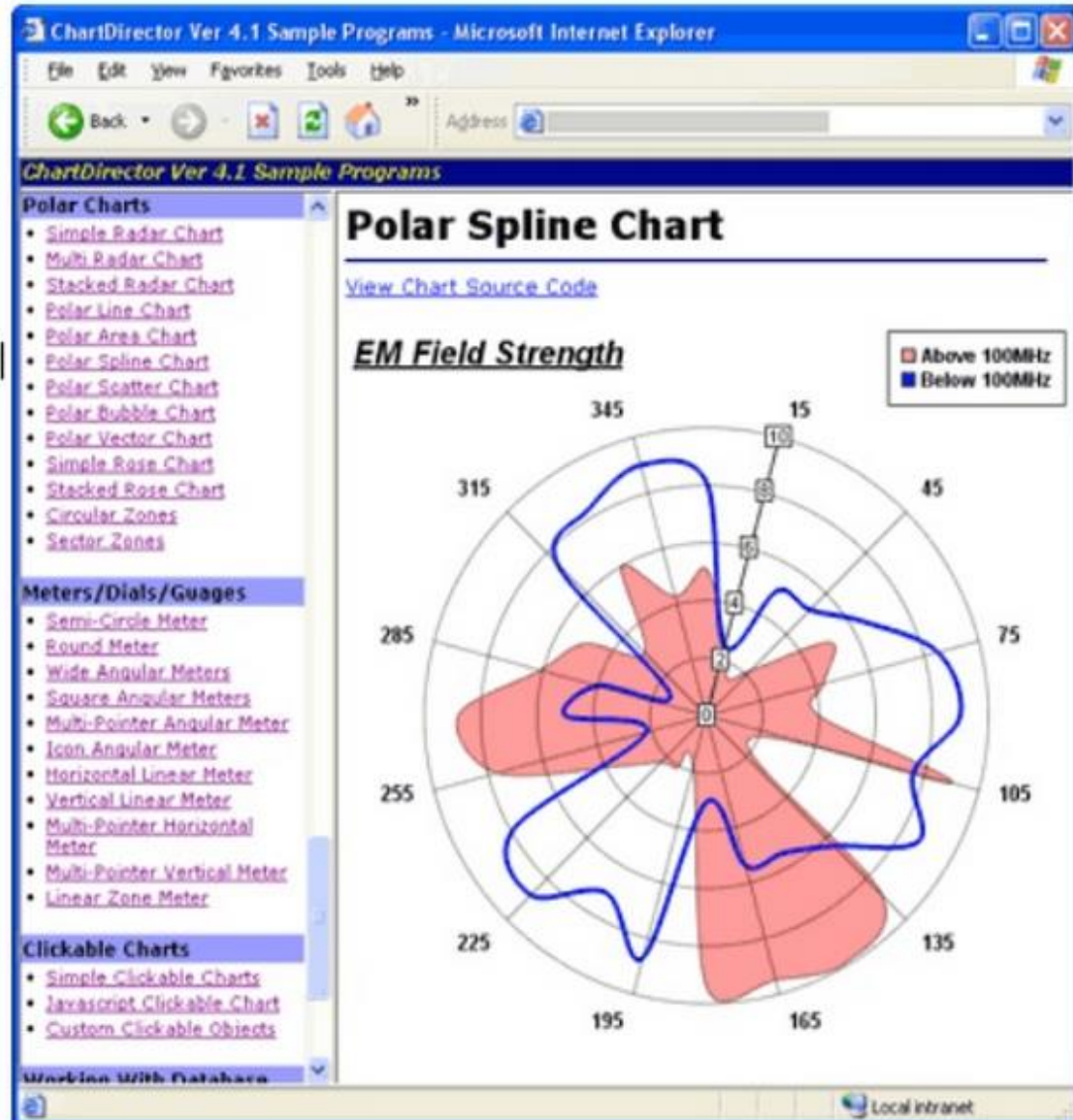


* 출처: 빅데이터를 위한 데이터 시각화 (이지선)

시각화 도구

파이썬 <http://python.org>

- 대규모 데이터를 무리없이 다룸 큰 규모의 계산에 쓰임
- 적은 코드만으로 많은 기능을 만듦
- 다양한 라이브러리를 이용할수 있음
- 미학적으로는 훌륭하지 않더라도 데이터 탐색 단계라면 훌륭한 툴



* 출처: 빅데이터를 위한 데이터 시각화 (이지선)

시각화 도구

- 시각화 도구들
 - 엑셀, CVS/JSON, 구글차트 API, Flot, Raphael, D3 (Data-Driven Documents), Visual.ly, NodeBo, R, Weka, Gephi
- 맵핑 (지도) 도구
 - Modest Maps, Leaflet, Polymaps, OpenLayers, Kartograph, CartoDB

국내동향

dongA.com 뉴스 | 오피니언 | 스포츠 | 연예 | 스튜디오 | 동아오토 | 부동산

“비료 어디에 더 좋까”... 드론이 척척 찾아내

김유영기자, 이상연기자 | 입력 2015-05-29 03:00:00 | 수정 2015-05-29 03:00:00

댓글보기 { 1 } | 폰트 + - | 뉴스듣기 { 여성 남성 } | 

ICT 접목한 첨단농업기술 확산



26일 충남 서천군 서천읍의 '금화친환경재배단지'에서 농민들이 드론을 날리고 있다. 특수

경기도, CCTV 위치선정에 빅데이터 분석결과 활용

경기도와 수원시, 경기지방경찰청이 CCTV 설치지역 선정에 빅데이터 분석결과를 반영하고 범죄 취약지역에 범죄예방 환경디자인을 적용하는 등 안전시범도시 구축사업을 추진한다.



국내동향

“강원을 빅데이터 산업의 광맥으로 육성”

기사입력 2015-05-12 03:00:00 | 기사수정 2015-05-12 03:00:00

[f](#) [좋아요](#) 0 |
 [T](#) [Tweet](#) 0 |
 [@](#) [f](#) [t](#) |
 [기사보내기](#)

경기도 '빅데이터 전문가' 250명 양성한다

최종수정 2015.05.14 08:38 기사입력 2015.05.14 08:38 경쟁률 7.1 : 1

朴대통령 “경제도약 이끌 새 자원”… 네이버, 도와 1050억원 펀드 조성



춘천 강원대에 강원창조경제혁신센터 출범 박근혜 대통령의 강원창조경제혁신센터를 방문해 관계자들을 하고 있다. 박 대통령 왼쪽은 한중호 창조혁신센터 상현 네이버 대표, 춘천=뉴시스

충청북도 => 의생명 빅데이터
공공행정, 과학기술
빅데이터로 특화

경상남도 => 제조빅데이터
(독일 Industry 4.0)

연합뉴스

2015.06.05 (금) ☁️ 광주 15.6 °C

통계청장 "인구총조사, 빅데이터 이용해 1천400억 절약"

송고시간 | 2015/05/28 06:00

[f](#) [T](#) [g+](#) [@](#)

보건복지 정책에 빅데이터 활용 추진

정부, 기본 계획 수립... 정책 방향성·우선순위 결정에 도움 기대

이형근 기자 bass007@dt.co.kr | 입력: 2015-01-08 19:09
[2015년 01월 09일자 8면 기사]

관광

관광산업 일자리 미스매치, '빅데이터'로 분석한다
제주도, 빅데이터 시스템 도정현안 적용

정부가 연내 보건복지 정책에 빅데이터를 접목해 효율성을 높인다. 빅데이터를 활용하면 보건의료분야 정책 방향성과 우선순위를 결정하는데 도움이 될 것으로 전망된다.

마윈 최장의
아시아리더십퍼런스
말말말



젊은이들에게 보내는 메시지

"청년들에게 오늘은 힘든 날이다. 내일은 더 힘들다. 하지만 오레는 아름다울 것이다. 오레에는 많은 사람들이 사라질 것이기 때문이다. 꿈을 가져라"

성공의 비결을 설명하면서

"고래잡이를 하면서 돈 버는 사람은 거의 없다. 새우는 고래보다 훨씬 적다. 하지만 새우잡이의 꿈을 10년 동안 지키면 돈을 벌 수 있다"

기업 내 여성의 역할을 강조하여

"10명의 남자가 회의하면 2시간동안 서로 싸운다. 반면 여성들이 토론하면 매우 논리적이고 편안하게 대화가 이뤄진다. 남을 존중하고 경청하는 여성의 특징이 기업의 성공을 이끈다"

IT업계 변화를 예견하면서

"앞으로 30년간 DT(데이터 기술) 혁명에 기반한 중소기업의 시대가 열릴 것이다. 중소기업이 창의력과 아이디어를 발휘하도록 대기업과 정부가 지원해야 한다"

자선 사업의 이유를 밝히며

"돈을 버는 것보다 제대로 쓰는 게 더 어렵다. 빌 게이츠가 자선사업을 하면서 한 머리가 늘지 않았다. (356억달러 재산은) 사회적 책임의 일환이다. 형운은 남과 나뉘어"



마윈 "IT시대 저물고 향후 30년간은 DT시대 온다"

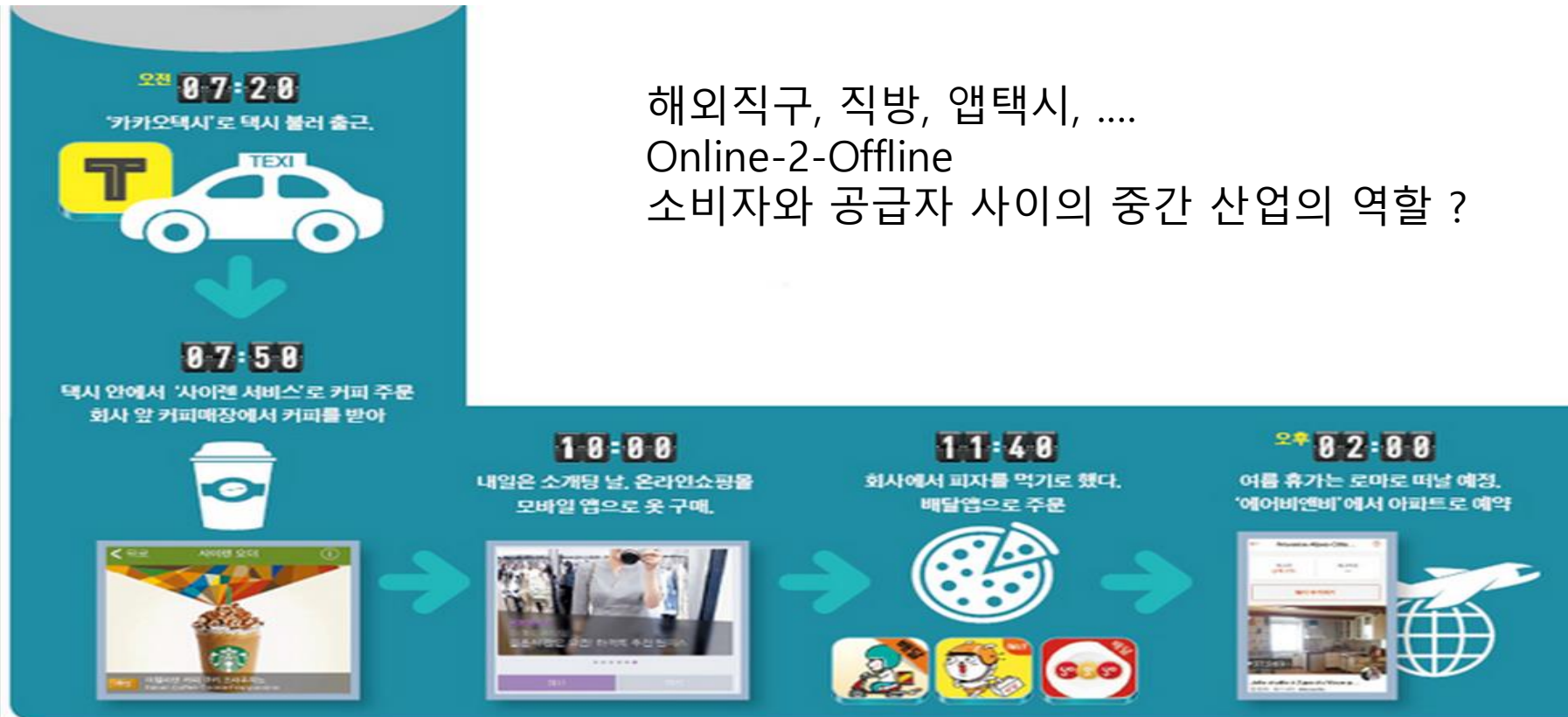
숫자로 보는 마윈과 알리바바

356 억달러	8500 만원	2314 억달러	1억명	34000 명
자산(약 30조원)	창업비용	시장가치(약 242조원)	하루 이용자	직원(간접고용 140만명)

※2014년 기준

국내동향

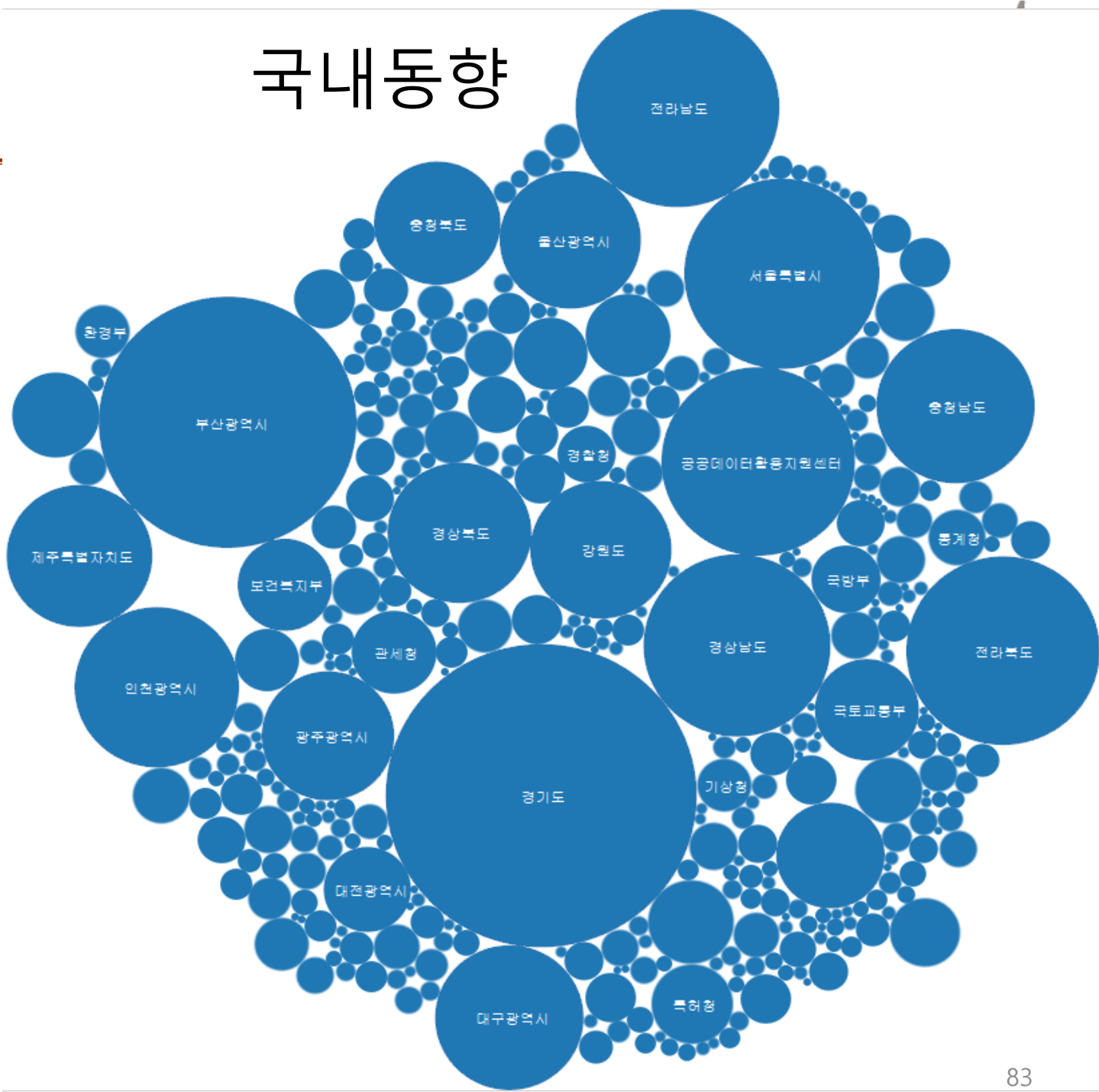
해외직구, 직방, 앱택시, ...
Online-2-Offline
소비자와 공급자 사이의 중간 산업의 역할 ?



기관 및 지역별 공개데이터 수

	orgal	갯수
1	경기도	1515
2	부산광역시	1041
3	전라남도	651
4	서울특별시	597
5	전라북도	587
6	공공데이터활용지원센터	581
7	경상남도	548
8	인천광역시	424
9	충청남도	393
10	대구광역시	346
11	제주특별자치도	333
12	경상북도	325
13	울산광역시	313
14	강원도	310
15	광주광역시	275
16	충청북도	251
17	산업통상자원부	185
18	국토교통부	170
19	보건복지부	140
20	행정자치부	119
21	대전광역시	119
22	농림축산식품부	115
23	과학기술정보통신부	113

국내 동향



평가결과 특·광역시 중에서는 부산광역시가 최우수에 올랐고 인천광역시와 광주광역시가 우수 기관을 기록했다. 도(道)에서는 경상남도, 제주도, 경상북도 순이었고 시(기초)는 시흥시(경기), 청주시(충북), 창원시(경남), 순천시(전남), 구미시(경북), 아산시(충남) 순이었다.

경남 창원, 기계·ICT융합 '메카트로닉스' 허브로

2015-04-09 17:44:44



경남도는 홍준표 도지사 취임 이후 경남발 혁신사례로 전격 추진된 '공공임대아파트 분양전환 가격 개선', '맞춤형 기업트랙 설치를 통한 청년일자리 제공', '거가대로 운영 재구조화 등을 통한 재정건전화', '빅데이터 분석을 통한 축제 활성화' 등이 우수한 평가를 받았다.

정부와 두산그룹이 기계산업과 첨단 ICT(정보통신기술)를 융합해 경남지역을 '메카트로닉스(기계+전자)' 허브로 육성한다. 전통적 제조업인 기계산업에 빅데이터, 사물인터넷(IoT), 로봇기술 등을 결합해 제조업 혁신을 선도하겠다는 구상이다. 또 물산업을 신성장동력으로 육성하고 지역 특산 천연물을 활용한 항노화 바이오산업 전진기지로 조성한다.

국내동향

홈 > 교육

충북대, 전국최초 빅데이터 분야 학석·박사과정 개설

조아라 기자 |

빅데이터로 대국민 서비스 개선

충청북도 특화 관광모델 개발
 기상청은 기후정보 융합 추진
 서울시는 교통분야 추가 활용

이형근 기자 bass007@dt.co.kr | 입력: 2014-06-29 19:00
 [2014년 06월 30일자 9면 기사]

[인터뷰] “지방자치도 빅데이터 시대”

이경희 충북대학교 비즈니스데이터융합학과 빅데이터센터 선임연구원

충북, 교량·터널에 ICT 적용한 첨단재난안전시스템 구축

홈 > 지역 > 충북

청주시, 교통정보 빅데이터 분석시스템 구축

-21일 착수보고회, 과학적 교통정책 수립 기대-

시흥경찰서 스마트 치안 시스템 활용

빅데이터의 효과

미래사회 특성		빅 데이터의 역할	
불확실성	→	통찰력	<ul style="list-style-type: none"> • 사회현상, 현실세계의 데이터를 기반으로 한 패턴분석과 미래전망 • 여러 가지 가능성에 대한 시나리오 시뮬레이션 • 다각적인 상황이 고려된 통찰력을 제시 • 다수의 시나리오로 상황 변화에 유연하게 대처
리스크	→	대응력	<ul style="list-style-type: none"> • 환경, 소셜, 모니터링 정보의 패턴 분석을 통한 위협징후, 이상 신호 포착 • 이슈를 사전에 인지·분석하고, 빠른 의사결정과 실시간 대응 지원 • 기업과 국가 경영의 투명성 제고 및 낭비요소 절감
스마트	→	경쟁력	<ul style="list-style-type: none"> • 대규모 데이터 분석을 통한 상황인지, 인공지능 서비스 등 가능 • 개인화, 지능화 서비스 제공 확대 • 소셜(니즈)분석, 평가, 신용, 평판 분석을 통해 최적의 선택 지원 • 트렌드 변화 분석을 통한 제품 경쟁력 확보
융합	→	창조력	<ul style="list-style-type: none"> • 타분야와의 결합을 통한 새로운 가치창출(의료 정보, 자동차정보, 건물정보, 환경정보 등) • 인과관계, 상관관계가 복잡한 컨버전스 분야의 데이터 분석으로 안전성 향상, 시행착오 최소화 • 방대한 데이터 활용을 통한 새로운 융합시장 창출

결론

- **빅데이터가 성공하려면 ?**
 - **데이터 기반 조직문화 형성**
 - 직관보다 데이터 기반의 **과학적 의사결정 중시 문화**가 필요
 - **CEO 의지**가 가장 중요한 관건임
 - **조직의 분석지능 제고**
 - 빅데이터 시대에 분석지능이 높은 조직(기업, 국가)일수록 **혁신 능력**과 **리스크 관리 능력**이 우수함(MIT Report) - 데이터과학자
 - 작은 업무에서라도 빅데이터 분석을 시작! (가장 확실한 교육)
 - **분석결과가 업무혁신으로 연결되어야 함**
 - 작은 데이터부터 빅데이터까지 모두 고려하여 가치 창출
 - 업무혁신으로 연결되지 않는 빅데이터는 의미가 없음

결론

- **지속가능한 빅데이터가 되려면 ?**

- 빅데이터의 활용과 함께 데이터 거버넌스가 구축되어야 !

빅데이터 거버넌스는

데이터의 품질보장, 프라이버시 보호, 데이터 수명관리, 전담조직과 규정정립, 데이터 소유권과 관리권 명확화 등을 통하여 빅데이터가 적시에 필요한 사람에게 제공되도록 체계를 확립하는 것

빅데이터 거버넌스가 확립되지 못하면

- 품질이 낮은 데이터를 중요한 의사결정에 사용함으로써 심각한 문제를 야기
- 개인 프라이버시 관련 데이터로 인해 빅브라더의 우려가 현실화
- 폭증하는 데이터에 대한 메타 데이터 & 수명 관리 소홀로 인한 IT 비용 급증
- 데이터 관련 문제를 전담하고 책임지는 조직과 인력이 없다면 빅데이터 효과는 일회성에 그칠 것임