

빅데이터 실무

강사 : 송재성 (zzyzigy@naver.com)

학습목표

- ▶ 제4차 산업혁명 부상과 의의
- ▶ 빅데이터 개론
- ▶ 빅데이터 처리과정과 관련기술
- ▶ 빅데이터 활용 예

제4차 산업혁명 부상

- ▶ '16년 **세계경제포럼**에서 4차 산업혁명을 의제로 제시하면서 세계적 관심 점화
- ▶ 4차 산업혁명 개념은 '12년 독일의 '**인더스트리 4.0**' 시작 당시 이미 사용되었으나, 세계경제포럼에서 거론된 후 재조명
- ▶ 국내에서는 이세돌과 알파고의 대국 이후 4차 산업혁명의 핵심기술 중 하나인 **인공지능**을 중심으로 이목 집중



제19대 대통령 대선후보



제4차 산업혁명 관련공약



세계 주요국가들이 일제히 4차 산업혁명에 대처하기 위한 경제정책 마련에 골머리를 앓고 있는 가운데 대통령 선거를 앞둔 우리나라에서도 주요 대선 후보들이 일제히 제조, 유통, 관광 등 기존 산업에 정보통신기술(ICT)을 융합해 신산업과 일자리를 만들어내겠다는 경제공약을 제시하고 나섰다.

제4차 산업혁명 관련공약



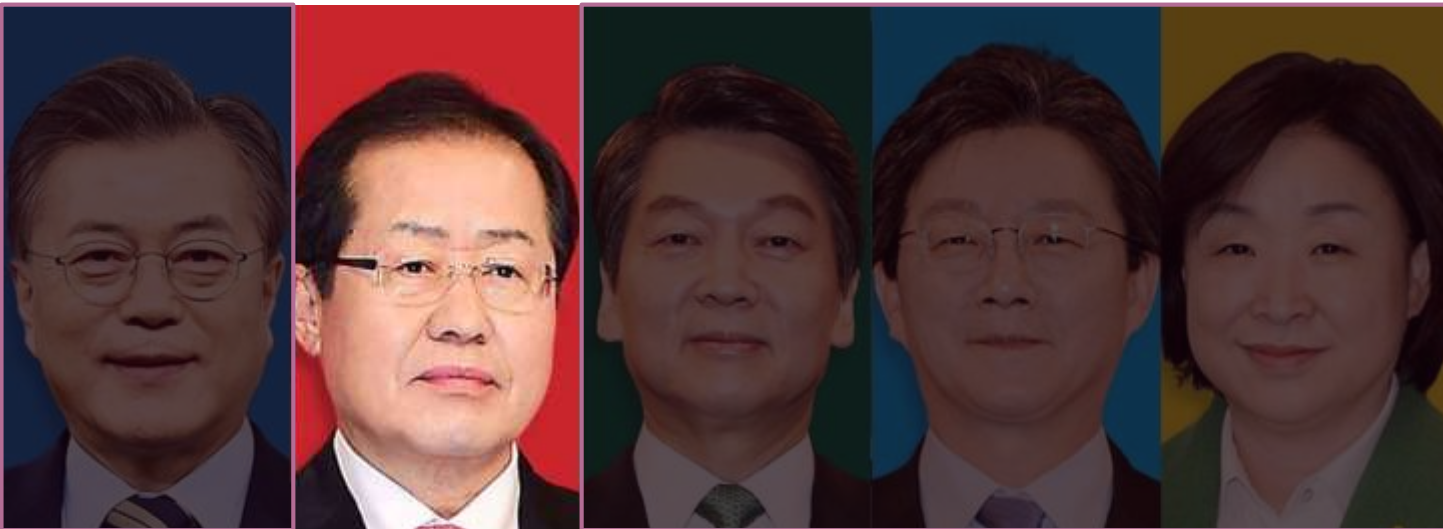
문재인 후보는 "정부가 4차 산업혁명이라는 황무지에 레일을 깔겠다"며 정부 주도의 5세대(5G) 이동통신망 구축, 혁신 벤처기업의 제품을 정부가 구매하고 마케팅까지 해주는 역할을 하겠다고 강조했다. 4차 산업혁명 준비에 모든 정부의 역량을 모으겠다는 것이다.

제4차 산업혁명 관련공약



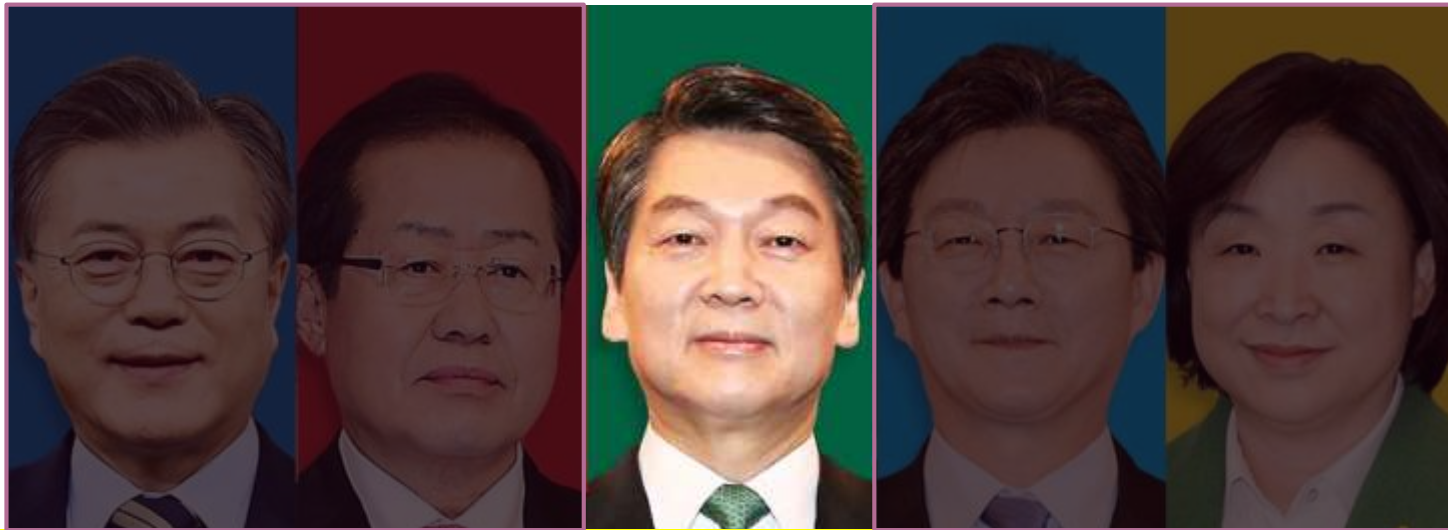
심상정 후보 역시 "정부주도 투자로 박정희정부가 경부고속도를 뚫은 것처럼 국가가 4차 산업혁명을 주도해야 한다"고 했다. 이렇게 해서 얻은 성과는 소수만 누리지 않도록 기본소득제를 도입하겠다는 것이다.

제4차 산업혁명 관련공약



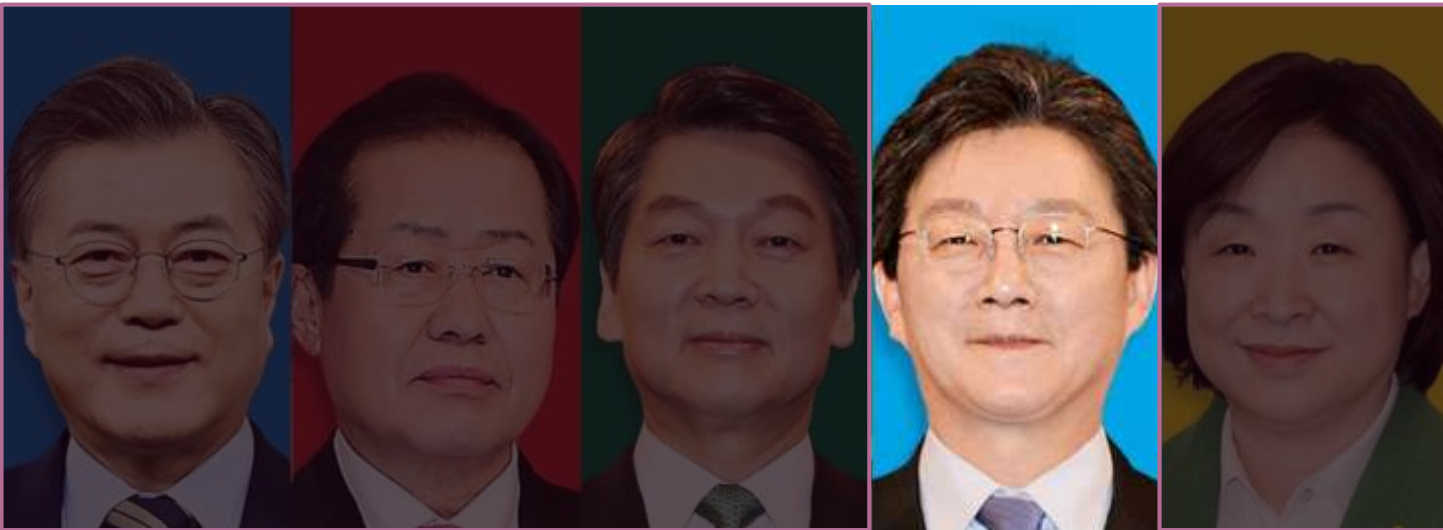
홍준표 후보는 정부와 민간이 함께 펀드를 마련해서 4차 산업혁명을 이끌어가는 창업기업과 중소기업들에 지원하겠다는 공약을 제시했다.

제4차 산업혁명 관련공약



안철수 후보는 "4차 산업혁명을 제대로 이끌기 위해서는 정부의 역할을 최소한으로 줄여야 한다"고 지적했다. "새로운 기술과 다양한 산업이 융합되기 때문에 이를 미리 예측해서 정부가 앞에서 끌고 가는 것은 사실상 불가능하다. 민간이 4차 산업혁명을 선도하면 정부는 뒤에서 민간을 밀어줘야 한다"고 했다.

제4차 산업혁명 관련공약



유승민 후보 역시 젊은이들이 창업에 뛰어들 수 있는 환경을 만들겠다고
창업 생태계를 활성화하는 것이 4차 산업혁명을 선도할 수 있는
방법이라고 강조했다.

산업혁명의 역사

			
제 1차 산업혁명	제 2차 산업혁명	제 3차 산업혁명	제 4차 산업혁명
18세기	19세기~20세기 초	20세기 후반	2015년~
증기기관 기반의 기계화 혁명	전기에너지 기반의 대량생산 혁명	컴퓨터와 인터넷 기반의 지식정보 혁명	IOT/CPS/인공지능 기반의 만물 초지능 혁명
증기기관을 활용하여 영국의 섬유공업이 거대산업화	공장에 전력이 보급 되어 벨트 컨베이어를 사용한 대량 생산보급	인터넷과 스마트 혁명으로 미국주도의 글로벌 IT기업 부상	사람, 사물, 공간을 초연결, 초지능화 하여 산업구조 사회 시스템 혁신

제4차 산업혁명

- ▶ 정보통신 기술이 다양한 산업들과 결합하여 지금까지 볼 수 없었던 새로운 형태의 제품과 서비스, 비즈니스를 창출하는 것을 말한다

제4차 산업혁명 대표기술

▶ **인공지능, 기계학습**, 로봇, 사물인터넷,
클라우드, 빅데이터, 3D 프린팅, 나노,
드론, 자율주행, 핀테크, 블록체인

외국 동향 - 미국

SMART
AMERICA



- ▶ 애플, 구글, 마이크로소프트, 페이스북, 아마존, IBM 등 세계를 주도하는 미국 소프트웨어 서비스 기업들은 인공지능(AI)과 자율주행차, 전기자동차, 에너지, 스마트공장, 드론(무인기), 재해대응 시스템 등 온갖 분야의 신사업을 모색 중이다
- ▶ 오바마 정부는 또 2013년 4월 뇌과학 연구를 지원하는 '브레인(BRAIN) 이니셔티브'를 출범시켰다. 2025년까지 50억 달러(5조 원)의 예산이 투입해 지능형 정보통신기술(ICT)을 개발할 것으로 전망된다.

외국 동향 - 독일



- ▶ 독일 정부는 2012년 10월에 도이체텔레콤, SAP, 지멘스 등 산업계가 제출한 '산업 4.0' (Industrie 4.0)이라는 보고서의 내용을 반영해 '하이테크 전략 2020'을 내놔다.
- ▶ 독일은 정부기관과 산업계, 학계 대표 뿐만 아니라 미국의 IIC 등 해외 기관까지 참여시키고 2억 유로(2천500억 원)의 예산을 투입해 인공지능(AI), 빅데이터, IoT, CPS, 디지털화, M2M(기계 대 기계), 로봇공학 등 기술을 활용해 다양한 ICT 기술이 융합된 '스마트 공장'을 구축하는 것이 목표다.

- 17

외국 동향 - 중국

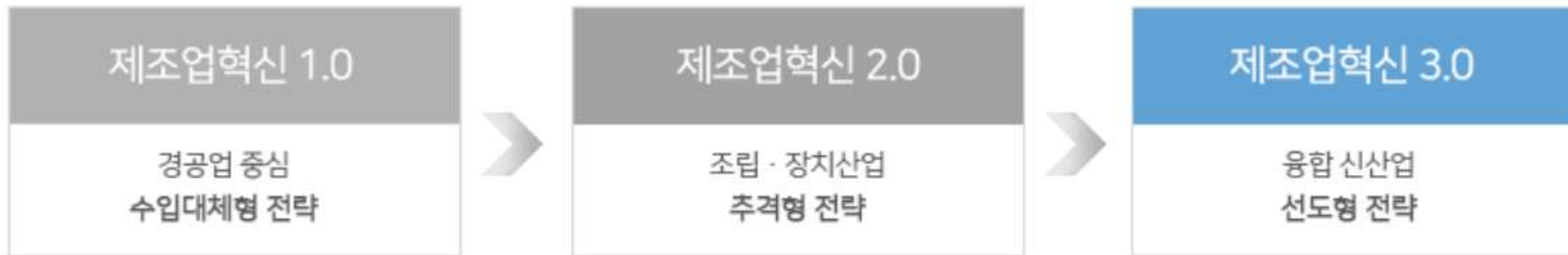


- ▶ 중국 정부가 2015년에 발표한 '중국 제조 2025' 정책에는 공업화와 정보화의 결합을 통해 혁신을 유도하고, ICT나 로봇뿐만 아니라 농업기계 장비, 소재, 의료 등에도 적용키로 했다.
- ▶ 중국 최대 포털업체 바이두는 이미 음성인식 등 AI 관련 연구의 일부 분야에서 미국 선도 업체들과 어깨를 나란히 하고 있으며, 5년 안에 자율주행차를 대량 생산한다는 목표를 세웠다. 또 게임업체 텐센트와 전자상거래업체 알리바바 등 다른 중국 ICT 기업들도 IoT, 빅데이터, AI, 클라우드 컴퓨팅 등 다방면의 연구개발에 매진하고 있다.

국내 동향 - 제조업혁신 3.0

- ▶ 제조업과 IT/SW, 서비스, 타 산업과의 융복합이 확산중이고, 3D프린팅, 스마트 공장 등 새로운 생산방식이 등장

[한국 제조업의 혁신 패러다임 변화]



웨인 크레츠키



- ▶ *I skate to where the puck is going to be, not where it has been.*
- ▶ 나는 펙이 지나간 자리가 아니라 나아갈 자리로 움직인다

빅데이터 등장배경

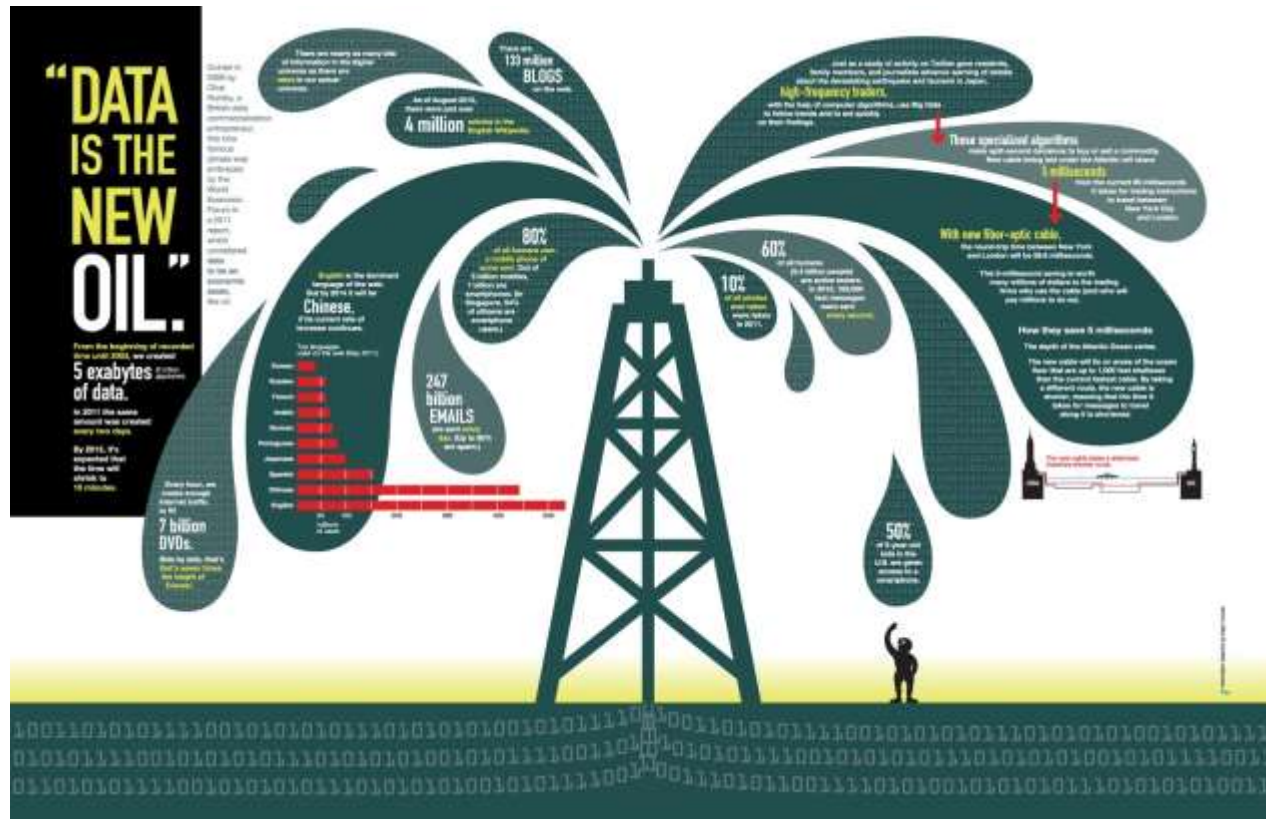


- ▶ 구글의 에릭 슈미츠에 의하면 2003년까지 인류가 쌓아올린 데이터의 양은 5 Exabyte 수준
- ▶ 2007년부터는 하루만에 그 정도의 분량을 쏟아내는 시대가 되었음

빅데이터 등장배경

- ▶ 디지털로 저장되는 데이터가 기하급수적으로 증가
- ▶ 서버로그, 기상/교통/공공 시스템의 센서 데이터, 보안카메라 이미지/동영상
- ▶ 블로그, 리뷰, 이메일, 사진, 동영상
- ▶ 엄청난 규모의 데이터 집합의 분석은 고성능 슈퍼 컴퓨터의 영역으로 국한
- ▶ 예측, 의사결정 지원을 목적으로 수집한 다양한 종류의 데이터 집합을 분석해서 활용하려는 산업 부문이 늘어나고 있음

빅데이터는 21세기 원유



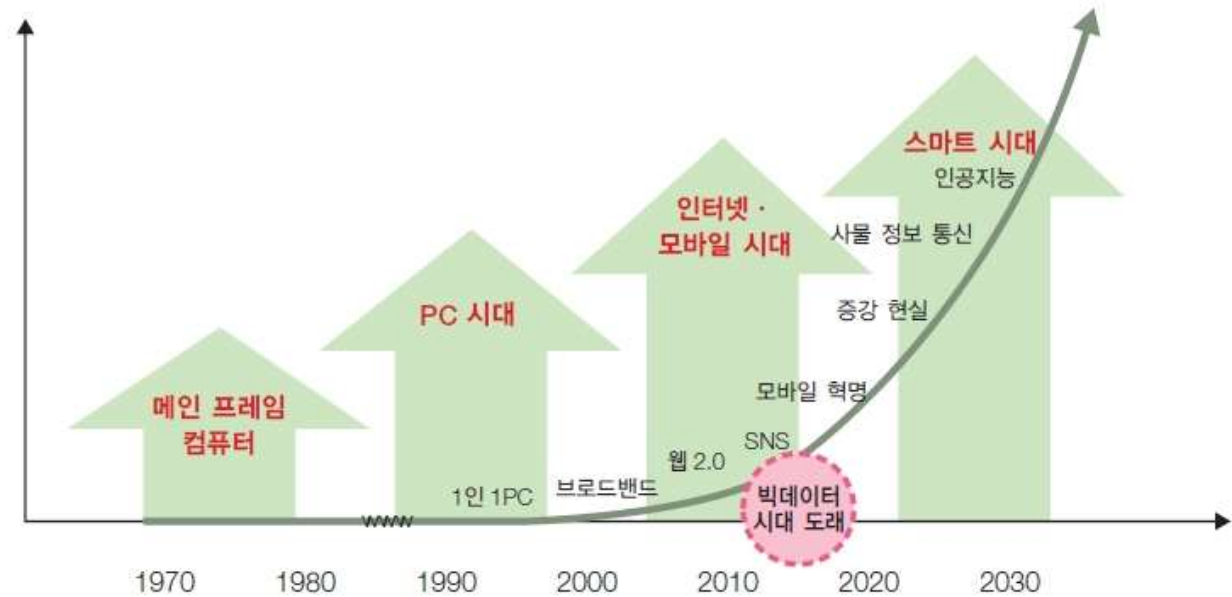
빅데이터는 21세기 원유

- ▶ 2011 기준, 웹에는 133만개의 블로그가 있다
- ▶ 2012.08 기준, 영어 위키피디아에는 400만개의 기사가 있다
- ▶ 2011 기준, 웹 공용어는 영어지만, 2014에는 중국어가 차지할 것이다
- ▶ 인류의 80%정도가 핸드폰을 소지하고 있다
- ▶ 매일 247만개의 이메일이 발송된다
- ▶ 매시간 생성되는 인터넷 트래픽은 7백만 개의 DVD를 채울 정도다
- ▶ 인류의 60% 정도가 문자메세지를 작성하고 전송한다
- ▶ 미국에서는 5세 아이의 50%정도가 스마트폰을 다룰 줄 안다고 한다

빅데이터는 21세기 원유

- ▶ 뉴욕 증권거래소에서는 하루에 4.5테라바이트의 데이터 발생
- ▶ 페이스북은 2,400억개의 사진을 보유하고 있고 매달 7 페타바이트 증가
- ▶ 계보조사 사이트인 앤서스트리는 약 10페타바이트의 데이터를 저장
- ▶ 인터넷 아카이브는 약 18.5페타바이트의 데이터를 저장
- ▶ 스위스 제네바에 있는 대형 하드론 입자가속기는 매년 30페타바이트의 데이터를 생산

빅데이터는 21세기 원유



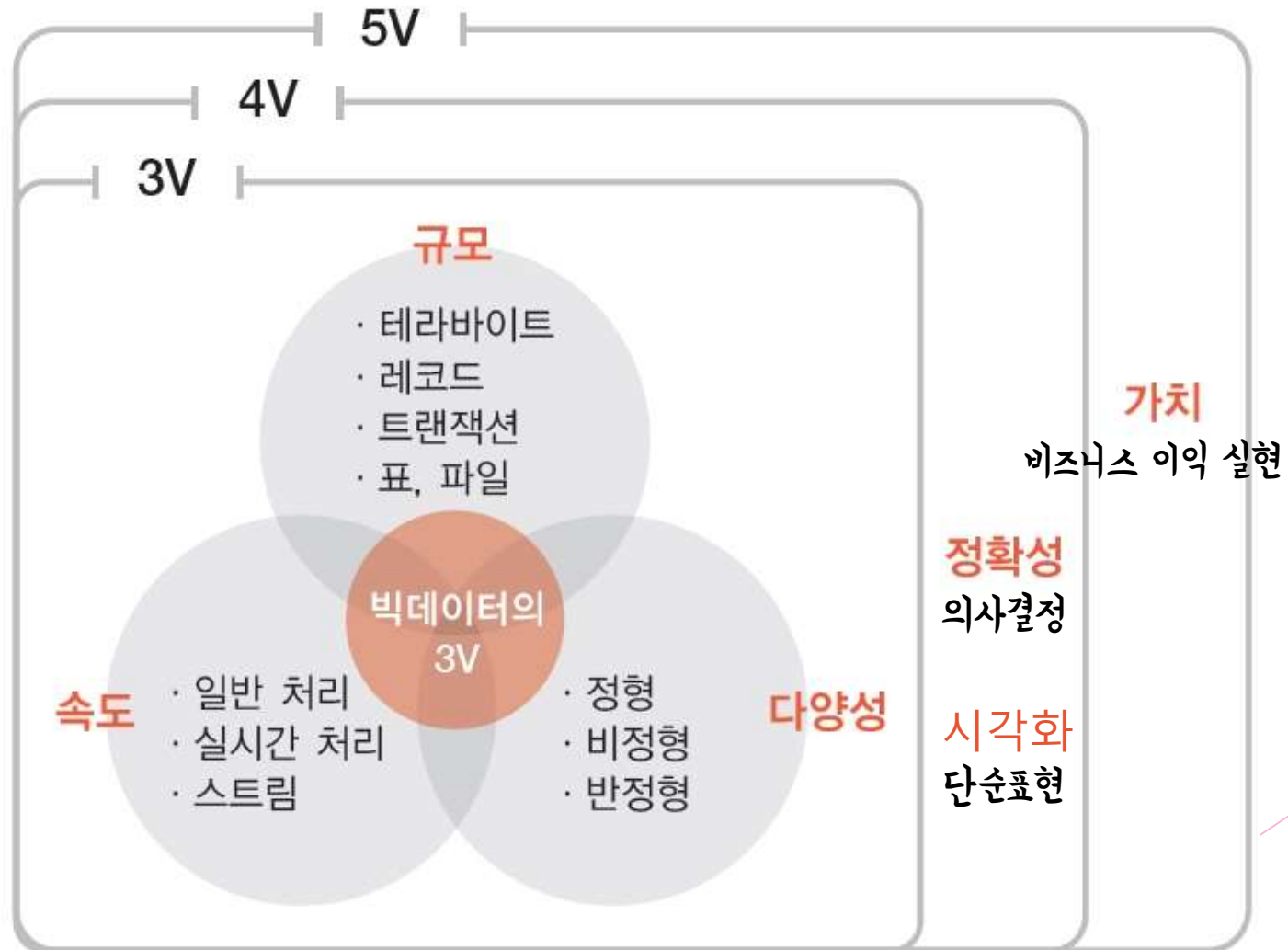
빅데이터란?

- ▶ 엄청난 용량의 데이터 집합
- ▶ 너무 커서 분석, 검색, 저장, 수정시
기존 어플리케이션으로 처리하기 힘들
- ▶ 즉, 데이터는 주변에 널려 있는데 그것을 저장하고
분석하는 일은 매우 어려운 상황에 놓임

빅데이터란?

- ▶ 하지만, 데이터의 크기에 중점을 두지 않고, 그것으로부터 가치를 추출하기 위해 사용되는 **아키텍처, 플랫폼 및 기술**을 총칭하는 의미

빅데이터 특성



빅데이터 종류

종류	설명
정형	고정된 필드에 저장된 데이터 예) 관계형 데이터베이스, 스프레드시트
반정형	고정된 필드에 저장되어 있지는 않지만, 메타데이터나 스키마 등을 포함하는 데이터 예) XML, HTML 텍스트
비정형	고정된 필드에 저장되어 있지 않은 데이터 예) 텍스트 분석이 가능한 텍스트 문서, 이미지 · 동영상 · 음성 데이터

전통적 데이터 vs 빅데이터

	전통적 데이터	빅데이터
생성주체	정부 및 기업등 조직	개인 및 시스템
데이터 유형	정형 데이터 조직 내부 데이터 주로 비공개 데이터	비정형 데이터 조직 외부 데이터 일부 공개 데이터
데이터 특징	데이터 증가량 관리 기능 신뢰성 높은 핵심 데이터	기하급수로 양적 증가 쓰레기 데이터 비중 높음
데이터 보유	정부, 기업등의 대부분 조직	인터넷 서비스 기업
처리 플랫폼	정형 데이터 생산/저장/분석/처리	비정형 데이터 생산/저장/분석/처리

전통적 데이터 vs 빅데이터

	전통적 데이터	빅데이터
하드웨어	고가의 저장장치 관계형 데이터베이스 서버 데이터웨어 하우스	저가의 리눅스 클라우드 환경(PAAS, IAAS)
하드웨어 확장	scale up(수직적 확장)	scale out(수평적 확장)
소프트웨어	관계형 데이터베이스 시스템 통계 패키지(SAS, SPSS) Data Mining	오픈소스 형태의 무료 패키지 Hadoop Ecosystem 무료 통계 시스템 (R)
통계이론	파레토법칙	롱테일법칙

빅데이터 기술 활용 수준

~2000년

수집, 적재 / 스토리지 인프라 : 대용량 저장소 구축



~2012년

분산 컴퓨팅 기반 분석 : 대용량 데이터 연산, 추출, 집계 분석



~2016년

데이터 마이닝 및 고급 분석 : 통찰력 기반 미래 예측

대용량 배치
분석

실시간
온라인 분석

기계학습

하둡



- ▶ 거대한 데이터를 분산시켜 저장할 수 있는 **HDFS**와 이렇게 저장된 데이터를 분석할 수 있는 **맵리듀스**로 구성된 빅데이터 기술
- ▶ 2005년, 더그 커딩이 오픈소스로 공개
- ▶ 하둡을 지원하는 오픈소스 덕에 기계학습과 실시간 분석이 가능하게 됨

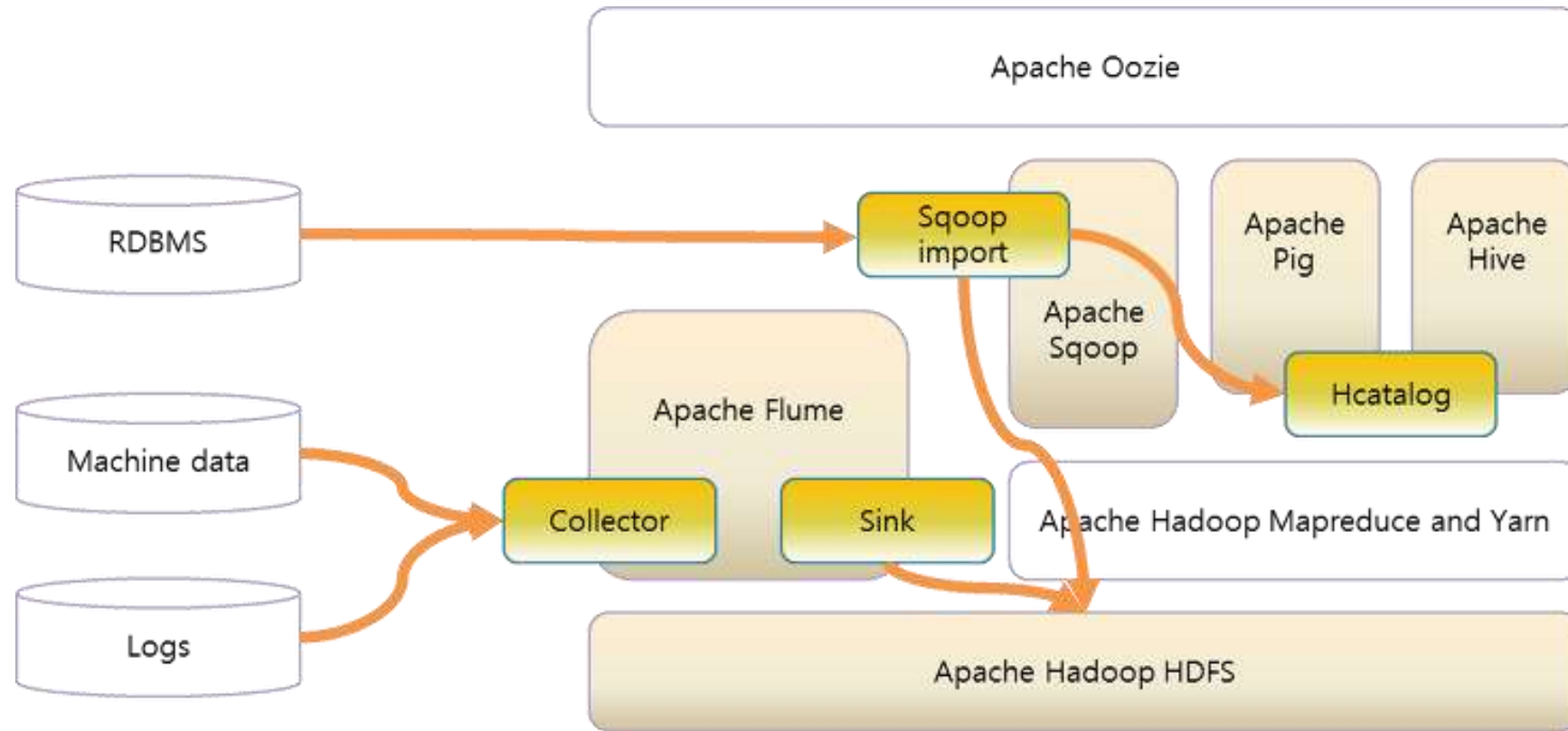
빅데이터 처리과정



빅데이터 수집

- ▶ 조직의 내/외부에 있는 다양한 시스템으로 부터
원천 데이터를 모으는 것
- ▶ 대규모 메시지, 대용량 데이터
- ▶ 정형/반정형/비정형 데이터
- ▶ 실시간 스트림 형태 데이터
- ▶ Flume, Chukwa, Strom, Esper

빅데이터 수집



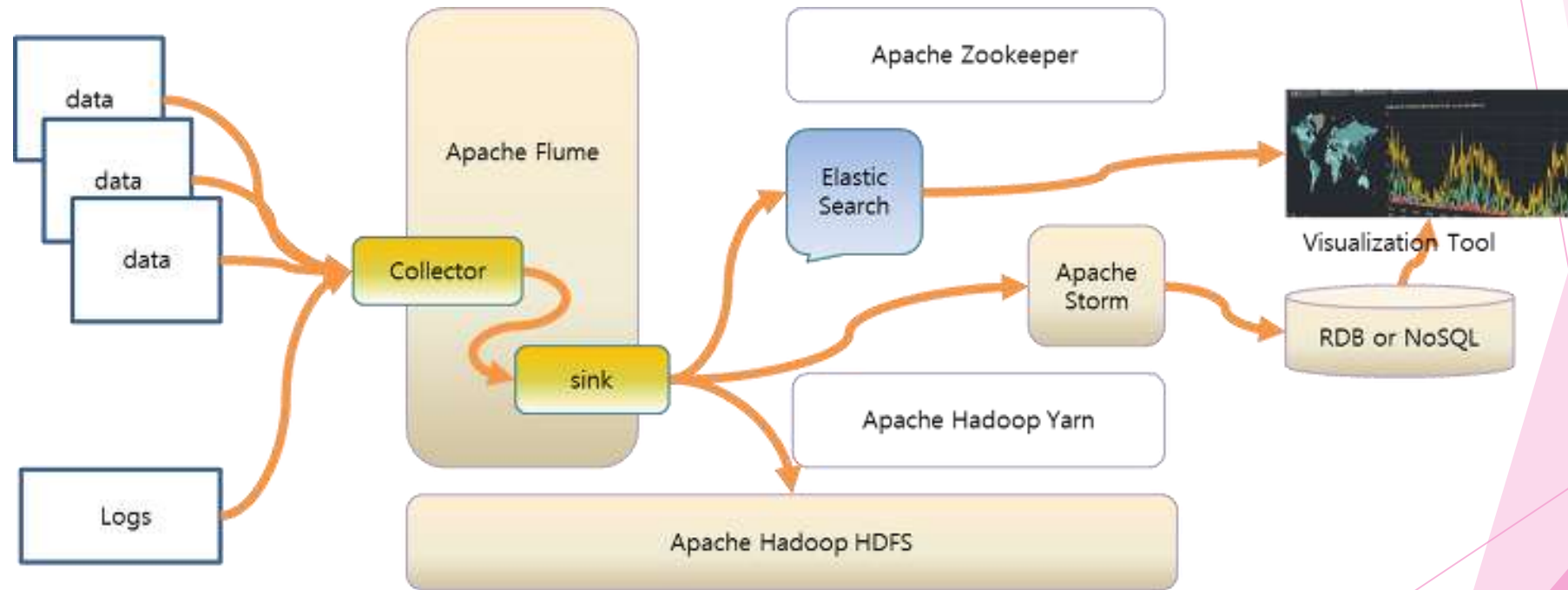
빅데이터 저장

- ▶ 수집한 데이터를 분산 저장소에 영구 또는 임시로 적재 하는 것
- ▶ 대용량 파일의 영구 저장 - HDFS
- ▶ 대규모 메시지를 영구 저장 NoSQL - HBase, MongoDB
- ▶ 대규모 메시지 처리 결과를 고속 저장 - 인메모리 캐시 Redis
- ▶ 대규모 메시지를 임시 저장 - Kafka, RabbitMQ

빅데이터 처리,탐색

- ▶ 대용량 저장소에 저장된 데이터를 분석에 활용하기 위해 데이터를 정형화 및 정규화 하는 것
- ▶ 데이터를 선택,변환,통합,축소 SQL on Hadoop - Hue, Hive, Spark SQL
- ▶ 정기적으로 발생하는 처리/탐색 작업 - oozie

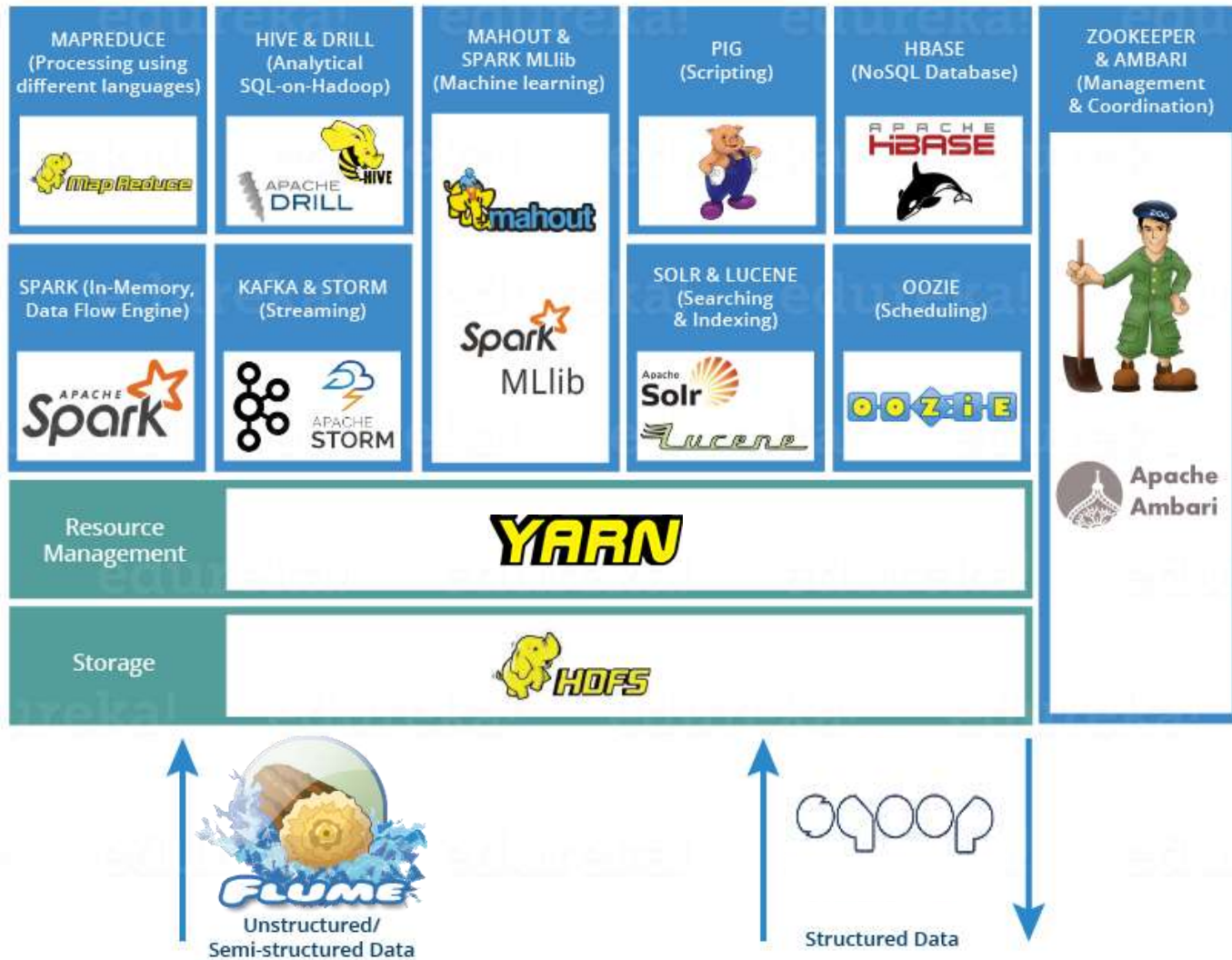
빅데이터 분석



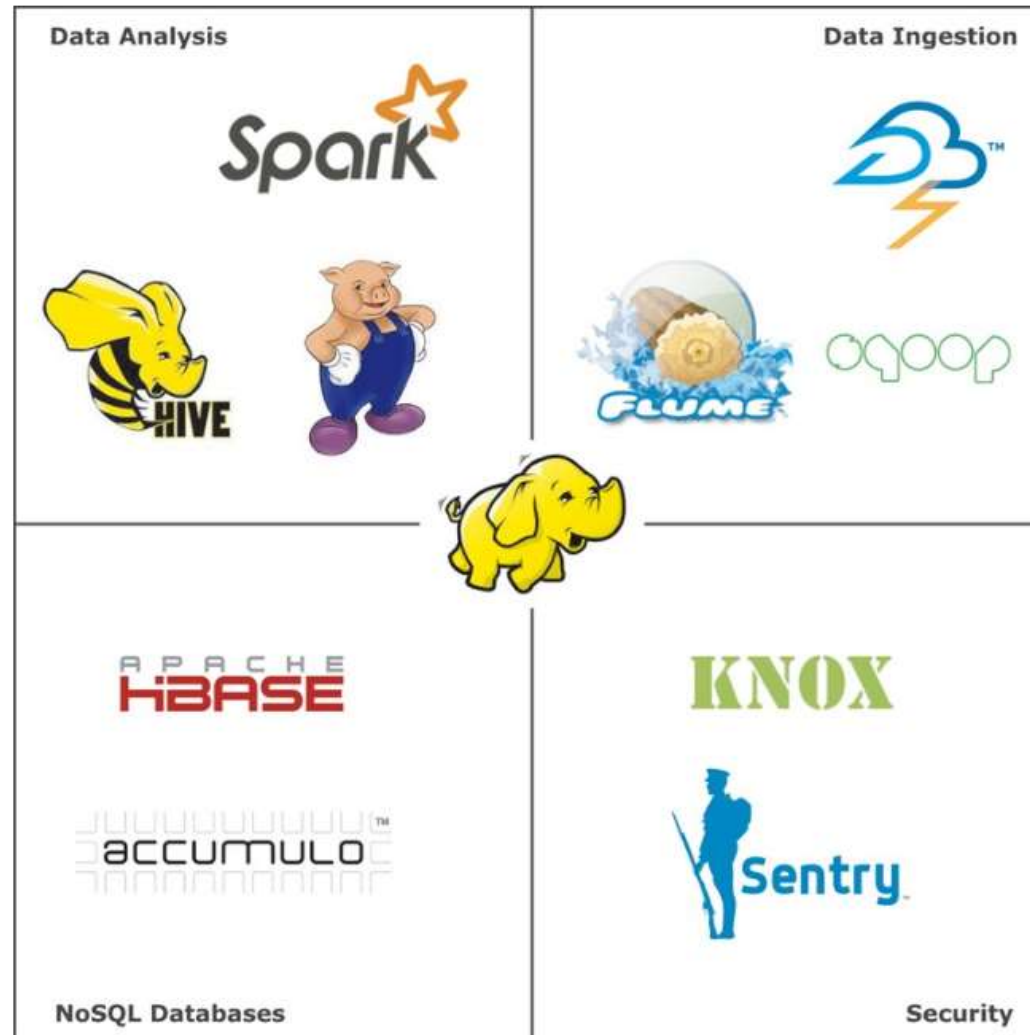
빅데이터 분석

- ▶ 대규모 데이터로부터 새로운 패턴을 찾고, 그 패턴을 해석해서 통찰력을 확보하는 것
- ▶ 통계, 데이터 마이닝, 텍스트 마이닝, 소셜 미디어 분석,
- ▶ 기계학습 활용 - 군집, 분류, 회귀, 추천등 고급분석 활용
- ▶ Impala, Zeppelin, Mahout, Sqoop

하둡 생태계 **hadoop**



하둡 생태계 **hadoop**



산업별 빅데이터 활용 - 포드

- ▶ 하이브리드(전기+내연) 자동차가 대세. 배터리 방전여부와 충전소 위치에 민감.
- ▶ 포드의 PHEV에 장착된 센서를 통해 배터리 잔량, 인근 충전소 찾기, 최저비용 충전알림등의 정보를 운전자에게 제공
- ▶ 시간당 25기가 정도의 센서 데이터 생성
이동통신망을 통해 포드의 클라우드 시스템으로 전송
운전자의 스마트 앱인 myford 모바일로 전송



산업별 빅데이터 활용 - 볼보

- ▶ 최근의 자동차는 전자기술이 집약되어 정교한 운전제어를 위해 많은 센서와 cpu가 내장
- ▶ 소비자의 자동차 운전과정 중 수집된 데이터를 본사 분석시스템에 자동 전송하여 데이터를 축적하고 이를 이용하여 제품 개선에 활용
- ▶ 종래에는 50만대나 팔린뒤에야 발견되었던 결함이 1800 - 2000대 정도만 출고한 후 발견하도록 있도록 개선



산업별 빅데이터 활용 - 금융

- ▶ **뱅크오브아메리카**

- ▶ 빅데이터 분석을 통해 신용리스크 모델을 고도화해 신용리스크에 대한 조기경보체계를 강화해왔으며, 비정형 데이터 분석을 통해 신용관리와 손실예측 처리시간도 단축시켰다.



- ▶ **시티은행(Citi bank)**

- ▶ IBM의 왓슨(Watson) 인공지능 시스템을 활용한 빅데이터 분석으로 기존 신용평가 모델을 보완해 연체여부, 연체기간, 거래내역, 패턴 분석 등 고객 계좌에 대한 보다 심층적인 분석을 실시해오고 있다



- ▶ **웰스파고(Wells Fargo)**

- ▶ 빅데이터 전문업체인 테라데이타(Teradata)와 협업으로 CSI(Customer Service Interaction) 3600이라는 프로젝트를 통해 광범위한 채널(콜센터, 지점, ATM, 창구직원, 이메일 등)의 데이터를 분석하여 고객이탈 확률과 추가상품 가입확률 등을 파악하게 되었으며, 이를 바탕으로 고객 맞춤형 마케팅을 추진해왔다.



산업별 빅데이터 활용 - 금융

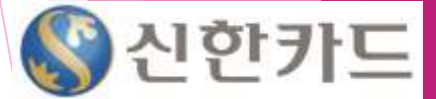
- ▶ 미국 아메리칸 익스프레스(American Express)
 - ▶ 위치 기반 SNS 회사인 포스퀘어(Foursquare)와 제휴하여 SNS 고객계정을 자사 카드와 연동시키고 고객의 거래 성향을 파악하는 데 도움이 되는 대량의 정보를 축적한 후 위치정보를 토대로 실시간 고객 맞춤형 프로모션을 실시해왔으며,
- ▶ 비자 (VISA)
 - ▶ 고객의 동의 하에 타겟 고객의 결제 위치, 결제시점, 구입품목 등을 실시간으로 파악하고 빅데이터 분석에 기반하여 할인쿠폰을 발송하는 실시간 문자 메시지 서비스를 실시해왔다.
- ▶ 마스터카드(Master Card)
 - ▶ 전 세계 200여 개국에서 19억 장의 카드와 32백만 개 카드가맹점 으로부터 매일 발생하는 방대한 결제데이터를 비식별화 처리를 한 후 기업체와 중소 상공인들에게 데이터를 판매해왔으며, IBM의 왓슨 인공지능 시스템을 활용해 카드 데이터를 분석하고 웰스파고 (Wells Fargo)와 여행중개업체 익스피디어(Expedia), 가방제조사 키플링 (Kipling) 등 많은 기업체와 중소 상공인들을 대상으로 사업 컨설팅 서비스도 추진해왔다.



산업별 빅데이터 활용 - 금융

▶ 신한카드

- ▶ 2013년 말 빅데이터 분석전담 조직인 빅데이터센터를 신설한 이후 고객 소비패턴과 선호 트렌드 분석, 고객군 유형화에 기반한 코드나인(Code 9) 카드 시리즈를 출시하였으며 각종 타깃 마케팅도 실시해왔다. 최근 들어서는 30여 개 이상의 기업과 기관들을 대상으로 컨설팅을 수행하며 빅데이터 컨설팅의 수익사업화를 추진하고 있다.



▶ 삼성카드

- ▶ 2016년 7월 가맹점 지원 통합 서비스 브랜드인 BMP(Big-data Marketing Partnership)를 카드업계 최초로 선보이며 가맹점 지원 사업을 적극 추진하고 있다. 동 사업은 가맹점 대상 맞춤형 빅데이터 분석 및 컨설팅, 가맹점 이용고객 의견의 실시간 수집 제공, 가맹점 혜택을 타깃 회원에게 자동 제공하는 서비스 등을 담고 있다.



▶ 국민카드

- ▶ 금융권 최초로 빅데이터 기반 실시간 마케팅 시스템을 도입했던 2015년 이후 콜센터 음성데이터를 텍스트로 변환해 분석함으로써 고객분석 정확도를 제고시켰으며, 최근에는 빅데이터 컨설팅 사업도 추진하고 있다. 2017년 들어서는 빅데이터 분석과 인공지능 기술 딥러닝에 기반한 부정사용방지시스템(FDS) 업그레이드를 추진 중에 있다.



산업별 빅데이터 활용 - 보험

- ▶ 미국 생명보험사 메트라이프(MetLife)
 - ▶ 빅데이터 기술을 활용하여 약 70개의 데이터베이스에 흩어져 있던 1억 명 이상의 고객 데이터를 하나의 시스템으로 통합하여 고객관리시스템인 The Wall 을 구축하고 이를 상품 판매 및 마케팅, 고객관리 등에 활용
- ▶ AIG와 AXA, Alliance 등 대해보험사
 - ▶ 운전자의 급가속, 급감속, 과속 등 운전습관 관련 데이터를 수집, 분석, 활용하여 안전한 운행을 한 고객들에 대해 보험료를 할인해 주는 방식의 운전습 관연계보험(UBI: Usage Based Insurance) 상품을 개발, 판매해왔다.
- ▶ 영국 대해보험사 아비바 (Aviva)
 - ▶ 고혈압, 당뇨, 우울증과 같은 질병에 걸릴 위험의 예측을 위해 기존의 의료진단정보(혈액 및 소변 샘플 분석 등)를 활용하는 방식 대신 신용평가보고서와 라이프스타일 데이터를 이용한 건강 위험 예측모형을 개발하여 언더라이팅(Underwriting) 단계에서 비용 절감을 추진하였다.



산업별 빅데이터 활용 - 유통

- ▶ 세계 최대 유통업체인 월마트(Walmart)

- ▶ 월마트 랩스(Walmart Labs)를 통해 소셜미디어에서 수집한 빅데이터를 이용해 소비자들의 심리와 행동양식을 파악, 상품구성 결정에 반영하고 있으며, 검색엔진인 폴라리스(Polaris)를 개발해 자사 온라인 쇼핑몰과 모바일 웹과 애플리케이션에 전면 적용하는 등 빅데이터로부터 소비자 패턴을 분석한 자료를 활용해 재고관리를 최적화하고 매장별로 소비자가 원하는 제품을 원활히 공급함으로써 고객만족도를 제고하고 있다.



- ▶ 미국의 온라인 쇼핑몰인 아마존닷컴(Amazon.com)

- ▶ 고객이 구입한 상품 정보를 분석해 구매 예상 상품을 추천하고 개인화된 쿠폰을 제공해 회사 매출의 약 35%가 빅데이터 기반 추천 시스템을 통해 발생하며 매년 이익의 10%를 추천 시스템 성능 향상에 투자하는 것으로 알려져 있다.



- ▶ 스페인 의류업체인 자라(ZARA)

- ▶ 전 세계 매장의 판매 및 재고 데이터를 실시간 분석해 최대 매출을 달성할 수 있는 재고 최적 분배 시스템을 개발해 최적화를 통한 실시간 고객의 니즈에 대응함으로써 현재 77개국에 2,000여 개의 매장을 가진 세계 최대 의류기업 중 하나로 부상하고 있다.



산업별 빅데이터 활용 - 물류

- ▶ 후지쯔(Fujitsu)의 농업용 빅데이터 분석 솔루션

- ▶ 후지쯔는 농지작업 실적과 작물 이미지 등 데이터를 분석해 수확량 증가와 품질을 향상시키는 클라우드 기반의 농업용 빅데이터 분석 솔루션을 2012년부터 제공하고 있다. 이는 기후와 토양환경 등에 대해 센서로부터 수집되는 데이터와 과거 수확 실적 등을 비롯한 빅데이터를 분석하여 최적의 파종, 농약 살포, 수확 시점을 제공하는 솔루션이다.



- ▶ 아마존의 예측 배송

- ▶ 빅데이터 분석을 이용한 고객 이해와 구매 추천의 선구자인 아마존은 예측 배송 이라는 또 다른 파격적 행보를 시도할 계획인데, 이를 위해 2013년 12월, 예측배송(anticipatory shipping) 서비스에 대한 특허를 취득한 바 있다. 예측 배송은 고객이 구매할지 여부가 불확실한 상황에서 고객 주소지 근처의 물류창고로 배송을 시작하는 것으로, 이는 기존 주문과 검색 내역, 위시 리스트와 쇼핑 카트에 담아놓은 상품, 반품 내역, 마우스 커서가 머무른 시간 등을 활용해 고객 자신보다 고객을 더 잘 이해하는 것으로 알려져 있는 아마존의 빅데이터 분석 역량에 바탕을 두고 있다.



산업별 빅데이터 활용 - 물류

▶ UPS(United Parcel Service)

- ▶ '2억5000만개 주소 데이터를 활용해 최적화된 배달 경로를 찾고 있다'며 '이를 통해 배송기사(운전자) 한 명당 하루 1마일을 덜 운행하게 됐다'고 말했다. UPS는 이 시스템을 통해 2017년까지 5만5000개의 배달 경로를 최적화할 계획으로, 이럴 경우 연간 5000만 달러에 이르는 연료비를 절감할 수 있을 것으로 전했다.

특히, 빅데이터는 예상치 못한 폭설, 홍수, 파업 등 자연재난 등이 발생할 경우 육,해,공 운송 채널 및 실시간 교통 상황을 분석해 대체운송경로를 찾는 데도 도움을 주고 있어 리스크(risk) 해결에도 큰 도움이 되고 있다



▶ DHL

- ▶ 2009년에 빅데이터를 도입해서 매일 기록되는 배송 도착지, 크기, 무게, 내용물 등 수 백건의 배송정보를 통해 소비자의 물류 서비스 이용 흐름을 파악하고 패턴을 분석해서 실시간 교통상황, 수신자 상황, 지리적, 환경적 요소를 고려해서 최적화된 배송 경로를 실시간으로 제공하여 배송 실패율 제로에 가깝게 만들고 있고, 불필요한 연비도 줄일 수 있다.

또한, 빅데이터를 장/단기 투자 규모 예측에도 활용하여 수요증가를 예측하여 허브 시설 확충, 물류센터 확장, 차량 증편에 대한 투자 규모를 결정해서 투자 위험 부담 최소화하도록 했다



산업별 빅데이터 활용 - 물류

- ▶ SK텔레콤의 상권 분석 및 타겟 마케팅 지원 서비스
- ▶ SK텔레콤은 국내 최대 이동통신 사업자로서 보유한 강점을 적극적으로 레버리지해 이동 인구와 지리 정보, 소비 업종과 상품 판매 현황을 종합적으로 분석하여 지오비전(Geovision)이라는 상권 분석 및 타겟 마케팅 지원 서비스를 개발하여 제공하고 있다.
- ▶ CJ대한통운
- ▶ 화물 차량에 태블릿PC와 디지털 운행기록계를 결합한 통합 단말기를 설치해 차량의 위치와 경로, 운송 중인 화물의 상태, 연료소모량, 속도 등을 실시간으로 파악하고 관제할 예정이다.
- ▶ (주)한진
- ▶ 최적의 운송경로를 찾아 택배 배송기사들에게 지급된 스마트폰과 태블릿 등에 네비게이션 기능을 통해 전달할 방침이다. 운전자가 화물을 내린 장소에서 가장 가까운 곳에 있는 다른 화물의 정보를 실시간으로 확인할 수 있어 공차율(화물칸이 빈 채로 운행하는 비율)을 낮출 수 있다는 게 회사 측 설명이다.



산업별 빅데이터 활용 - 의료

- ▶ 미국 보험사 웰포인트
- ▶ 방대한 규모의 최신 임상정보, 환자 보험기록 등 의료 빅데이터를 빠르게 분석하여 가장 연관성이 높은 맞춤형 정보를 추출하고 신뢰도를 기반으로 의료진에게 제공, 정확한 의사결정을 지원함
- ▶ 왓슨은 60만건의 의학적 근거와 200만장에 달하는 전공 서적, 2만 5000여 건의 사례 등을 분석 하고, 의료진에게 충분한 최신 의학 지식 제공이 가능하며 응급상황에서 빠른 판단을 지원하여 의료과실 (medication errors) 발생 확률을 낮추는 등 의료의 질을 제고하는 동시에 의료비를 절감할 수 있음



산업별 빅데이터 활용 - 의료

- ▶ 캐나다 온타리오 병원
- ▶ 임산부 14명 중 1명은 미숙아를 낳고, 미숙아 4명 중 3명은 사망을 한다고 한다. 의료기술이 발달하고 있긴 하지만 육안 진찰과 차트 정보에만 의존하는 방식으로는 이미 질병에 감염되고 난 뒤에 조치를 취하기 때문에 한발 늦을 수 밖에 없다. 더군다나 미숙아들은 말을 할 수 없기 때문에, 어디가 어떻게 아픈지 정확히 판단하기가 어렵다. 바이러스에 취약한 미숙아의 경우 더욱 위험하다.
- ▶ 캐나다 온타리오 병원에서는 미숙아 모니터링 장비를 통해 데이터를 수집해서 미숙아의 질병 감염 여부를 빠르게 판단하고 조치를 취해 더 많은 생명을 살릴 수 있었다고 한다
- ▶ 미숙아 모니터링 장비에서 생성되는 혈압, 체온, 심전도, 혈중산소 포화도 등의 생리학 데이터를 수집하고, 이 데이터들을 IBM에서 지원하는 스트리밍 데이터 분석기반 솔루션을 활용해 실시간으로 수집되는 데이터를 연속적으로 분석해서 패턴을 찾아냈다. 즉, 환자당 일 9000만건 이상의 생리학 데이터스트림을 실시간으로 분석했다고 한다.
- ▶ 수집한 실시간 생리학 데이터(혈압, 체온, 심전도, 혈중산소 포화도 등)와 질병 등의 이상징후 간의 연관성 분석을 했다. 즉, 생리학 데이터가 어떠한 현상을 보일 때 질병이 발견되는 지에 대한 연관성 패턴을 알아낸 것이다.

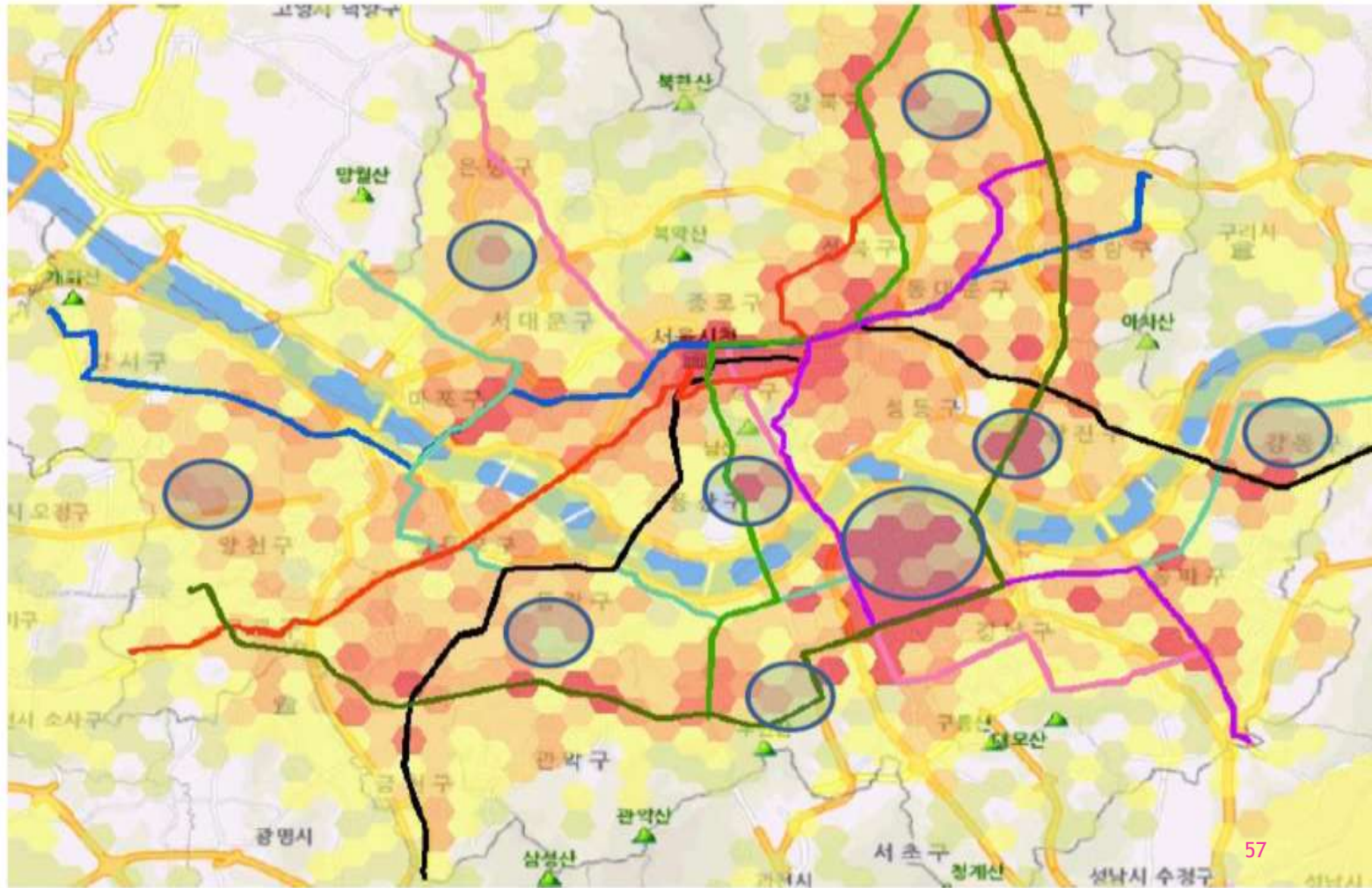


산업별 빅데이터 활용 - 공공

- ▶ 서울시 심야버스
- ▶ 2013년 4월 19일, 시범 운행된 심야버스 노선 최적화에 빅데이터를 활용했다. 기존에 전문가들이 모여 정했던 노선을 30억건의 콜데이터와 가입자 주소지(KT 제공) 정보, 택시 승하차 정보, 버스 정류장 및 교통량 정보등의 빅데이터를 활용해 분석하고, 시민들에게 좀더 편리한 방향으로 바꾼 것이다.
- ▶ 일단, 서울을 1km 반경의 1,250개 육각형 셀 단위로 나누고, 유동인구 교통수요량을 색상별로 표시하고, 지역별 유동인구와 이동통신가입자의 심야시간 통화량을 분석했다. 이를 통해 노선 최적화와 배차간격 조정 등의 판단 기반 인사이트를 얻었다. KT에서 제공된 정보는 유동인구 데이터로 가입자의 심야시간(0~5시) 통화 기지국 위치와 청구지 주소 데이터 통계치였다.
- ▶ 한달 간 데이터가 약 30억건에 달하는 양을 분석했다. 여기에 스마트카드를 통한 택시 승하차정보 1주일치 데이터도 분석했다. 이같은 데이터는 구역별 hexa본을 만들고 그에 유동인구 밀집도를 입력했다. 밀집도에 따라 색상을 시각화하고, 노선, 시간, 요일별 패턴을 분석했다. 여기에 노선 부근 유동인구 가중치를 계산해 노선을 최적화했다.
- ▶ 이를 통해 약 10%의 시민들이 더 심야버스를 이용할 수 있게 됐고, 하루 승객 이용자가 6000명에 달했고, 택시 승차거부가 9% 감소했고, 심야에 이동하는 여성이 11% 증가했다.

bus.go.kr / r BusMain.jsp

산업별 빅데이터 활용 - 공공



산업별 빅데이터 활용 - 공공



산업별 빅데이터 활용 - 공공

- ▶ 서울시 골목상권 서비스
- ▶ 2015년 12월, 서울시는 골목 상권에 창업하고자 하는 영세 소상공인을 대상으로 창업 위험도를 빅데이터 기반으로 예측해서 알려주는 서비스를 시행했다.
- ▶ 대형 유통시설이 들어서지 않은 큰 대로변 등의 뒷 골목과 같은 영세한 골목상권 1008개를 서울시 골목상권으로 규정하고, 신규 창업 또는 기존 자영업소 중 비중이 높은 지역내 외식업, 서비스업, 도소매업등 43개 생활밀착형 업종의 2000억개 빅데이터를 기반으로 서비스를 구축하였다.

산업별 빅데이터 활용 - 공공



산업별 빅데이터 활용 - 공공

- ▶ 건강보험심사평가원
- ▶ 심평원에서는 2007년부터 축적해 온 진료정보와 의약품, 지역별 요양기관 등의 의료분야 공공데이터 약 3,258억건을 보건 의료 빅데이터 개방시스템에 공개했다.
- ▶ 이 시스템을 통해 지역별 한자의 상태(질환 비중, 의료기관 방문 횟수, 의료비용 지출현황), 지역별 의료기관 현황(이용환자의 수, 각 진료과목별 내방횟수, 주변 의료기관과의 협력도)등을 얻을 수 있고, 이를 통해 각 대기업 및 벤처기업에서 각 회사 상황에 맞게 정보를 이용할 수 있다.

산업별 빅데이터 활용 - 공공

공공데이터

의료빅데이터

의료통계정보

고객지원

시스템소개



공개는 **넌리!** 제공은 **빨리!** 이용은 **편리!**

건강보험심사평가원에서
보유하고 있는 다양한 의료데이터를
국민에게 개방합니다.

데이터 서비스 현황

- 공공데이터 | 35 종
- Open API | 17 종
- 의료통계정보 |  133  62  3
- 원격분석시스템 | 총 150 계정(사용 146 계정)

의료통계
정보



국민관심질병



국민관심
진료행위



다빈도질병



질병(소분류)



진료행위



의료이용통계



의약품



의료자원



OECD 보건통계

산업별 빅데이터 활용 - 공공

- ▶ 국민건강보험공단
- ▶ 국민건강보험공단은 국토교통부와 협력하여 환자의료이용지도를 구축하였다. 기존의 건강보험공단이 보유한 환자상태 및 진료내역, 이용 의료기관등의 정보를 국토부가 보유한 공간정보기술과 결합하여 실제 의료이용 패턴을 분석했다.
- ▶ 이를 통해 환자 의료이용 지표를 산출하여 대국민 의료이용 실태, 의료자원 배치현황, 거점병원, 의료 취약지 도출등 의료이용 형태를 한눈에 알 수 있는 종합적인 공간정보 서비스이다

산업별 빅데이터 활용 - 공공

The screenshot displays the NHISS website interface. At the top, the NHISS logo and navigation links (로그인, 회원가입, 사이트맵, ENGLISH) are visible. Below the navigation bar, a horizontal menu lists services: 서비스이용안내, 데이터신청, 성과공유, 통계, 의료이용지표, 공공데이터, and 고객센터. The main content area, titled '공공데이터개방', features several service tiles:

- 건강보험자료 공유서비스**: A large tile on the left with a green background, featuring a hand holding colorful figures. It includes a '로그인' button and links for '회원가입' and '마이데이터발급신청하기'.
- 건강보험 빅데이터 표본연구 DB**: A white tile with a red heart icon, describing a sample research database for health insurance data.
- 건강보험 빅데이터 맞춤형연구 DB**: A yellow tile with a computer monitor icon, describing a customized research database.
- 원격연구포털**: A green tile with a lightbulb icon, describing a remote research portal.
- 의료이용지표**: A red tile with a bar chart icon, describing indicators for medical service utilization.
- 통계**: A grey tile with a bar chart icon, providing various statistical data.
- 의료이용지표순위**: A white tile with a list of medical service utilization indicators.

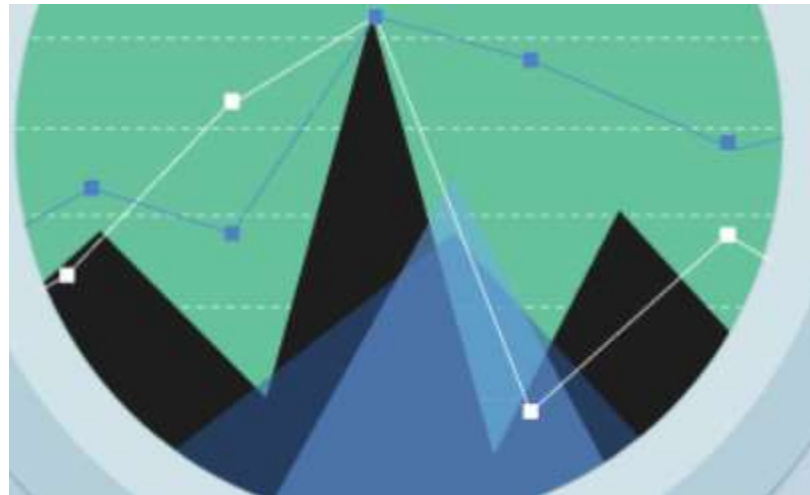
빅데이터 기획 전문가

- ▶ 빅데이터에 대한 다양한 시각과 기술, 분석 방법론, 활용 환경 등에 대한 이해를 바탕으로 빅데이터 활용 기회 발굴, 사업기획, 사업관리가 가능한 전문가



빅데이터 기술 전문가

- ▶ 빅데이터 수집, 저장, 처리, 운영관리 등 빅데이터 주요 기술에 대한 배경지식 및 고급기술 습득을 통한 실무형 빅데이터 전문가



빅데이터 분석 전문가

- ▶ 빅데이터 분석 기획, 분석방법 및 분석도구 활용 등에 대한 지식 함양을 기반으로 현업에서 빅데이터 분석을 통해 새로운 가치를 창출해낼 수 있는 빅데이터 전문가



빅데이터 센터

K-ICT 빅데이터센터
BIG DATA CENTER

데이터 큐브

분석활용

이용지원

지식공유

센터소개

검색어를 입력하세요



중소기업과 스타트업의
빅데이터 인프라를 지원합니다.

빅데이터 솔루션 매칭과 각종 교육 제공 등을 통해 창업 및 사업화를 돕습니다.



빅데이터센터
분석 인프라
활용교육

일정 : 2017.09.20~21

장소 : K-ICT 빅데이터센터 오픈랩
(스타트업캠퍼스 1동 6층)



인프라 이용



소셜분석



오픈랩



비식별조치



협력파트너

공지사항

더보기 +

17년도 9월 K-ICT 빅데이터센터 디러닝 교육

2017-09-14

17년도 9월 K-ICT 빅데이터센터 분석인프라 활용 교육

2017-09-12

<대회기간 연장> 글로벌 데이터톤(Datathon) 2017 행사 안내

2017-09-08

빅데이터 동향과 이슈

더보기 +



공공 데이터 포털

DATA 공공데이터포털
GO . KR

로그인 회원가입 사이트맵 ENGLISH

데이터셋 제공신청 활용사례 정보공유 이용안내

데이터를 검색해보세요!

3.기상청

여러분이 원하는 데이터, **데이터 1번가** 로 신청해주세요!

국가중점데이터

데이터 카테고리

건축정보 교통사고정보 국민건강정보 상권정보 수산정보 실시간 수도정보 농수축산가격정보 등산로 정보

부동산종합정보 통합재정정보 지방행정정보 부동산거래정보 식의약품종합정보 지방재정정보 법령정보 + 더보기

「데이터1번가」
여러분이 필요로 하는 공공데이터를 자유롭게
신청할 수 있는 데이터 소통창구

오픈데이터 포럼
민·관 협력을 통한 공공데이터 개방 정책 수립 및
공공데이터 활용성 증진을 위해 운영되는 단체

**공공기관 중복·유사
서비스 신고**
민간에서 개발한 공공데이터 활용 서비스가
공공기관에서 제공하거나 개발예정인 서비스와 기능이
중복되거나 유사한 경우, 이를 법적절차에 따라
조사·검토 및 협의를 통하여 공공기관의 서비스를 개선합니다.
(민원번호)

감사합니다