

4차 산업혁명과 사물인터넷

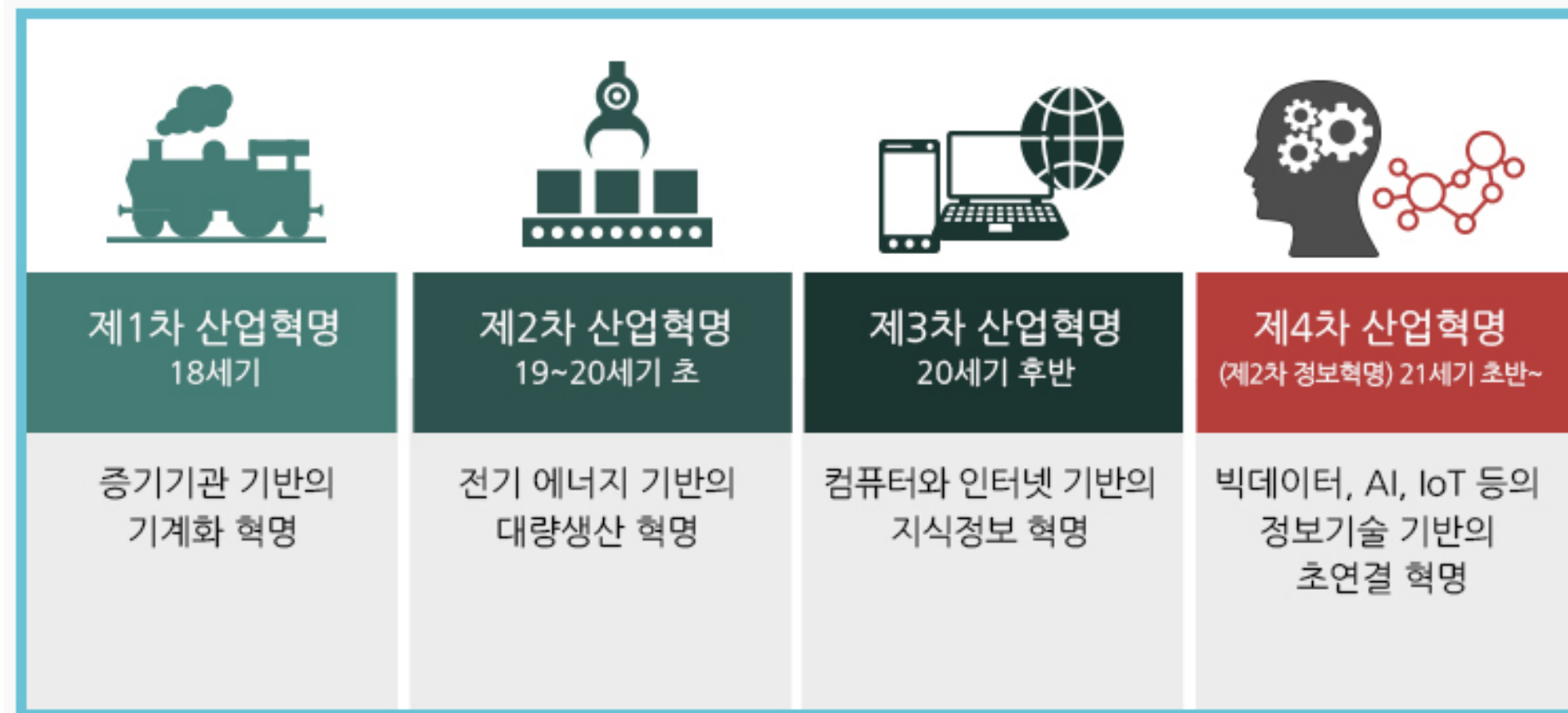
(빅데이터 개념 및 주요 사례)

제6강

우 송 대 학 교
철도 전기시스템학과
최 상 성

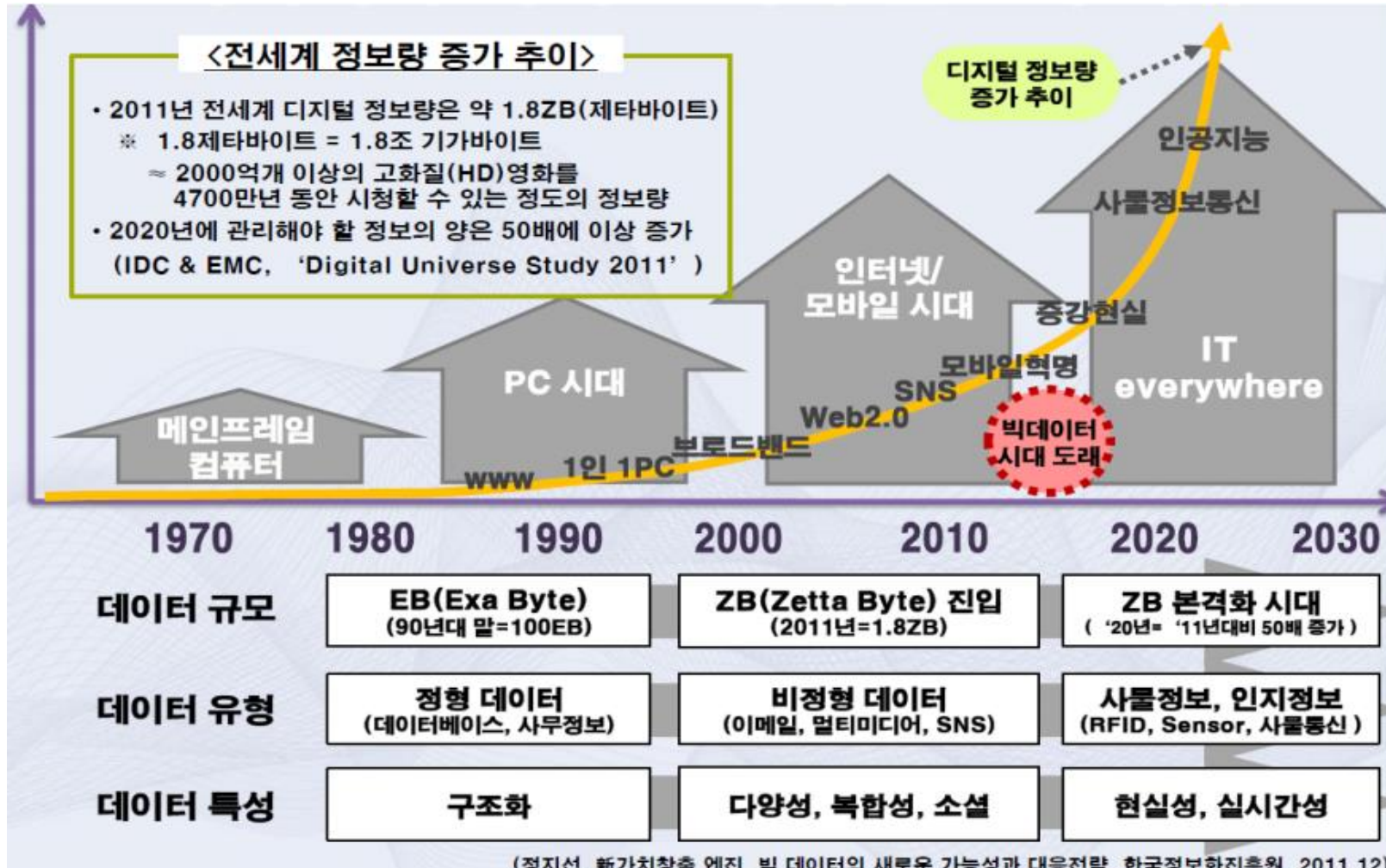
제4차 산업혁명

- 사물인터넷과 **빅데이터**, 인공지능 등의 정보기술 기반의 **초연결 혁명**



빅데이터의 출현(1/3)

빅데이터의 출현 환경



빅데이터의 출현(2/3)

빅데이터 출현 배경

- Web의 활동 증가

→ 구글, SNS 등 인터넷 서비스 활용 정보, 페이스북에 매달 20억 건 이상의 사진이 등록

- 모바일 기기 사용의 증가

→ 통화 및 문자 메시지 증가. 전자기기에 스마트 센서가 추가되고 이로부터 로그가 수집

- 새로운 IT 기구의 출현

→ RFID 리더기, GPS 네비게이션, 지진 측정기 등)이 출현은 데이터 생산을 더욱 확대

- 생명공학, 비디오 등 다양한 생산 원천

→ 게놈 서열(genomic sequence), MRI 사진과 같은 다양한 의학 정보가 증가

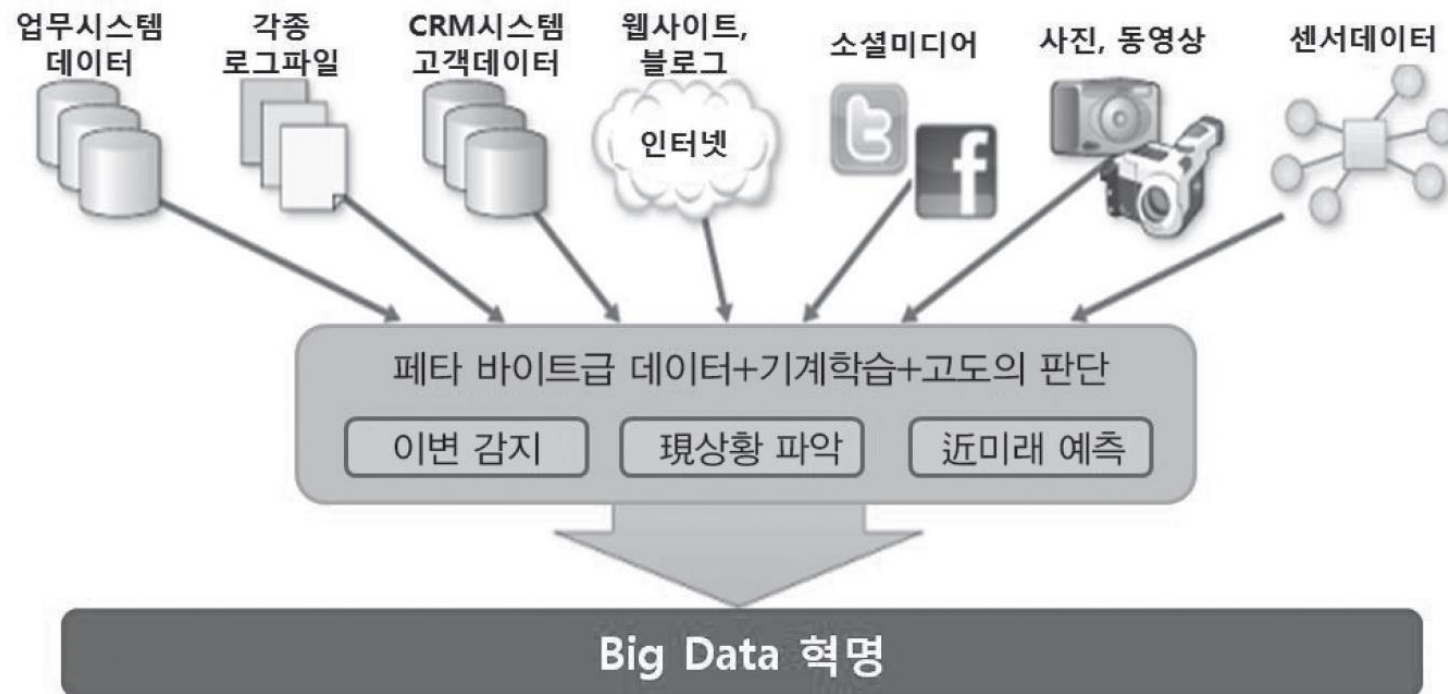


[1분 동안 인터넷에서 생성되는 데이터 양]

빅데이터의 출현(3/3)

▪ 빅데이터의 정의

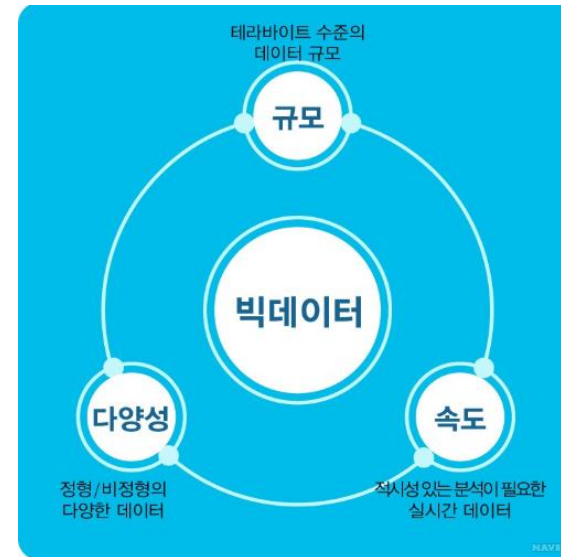
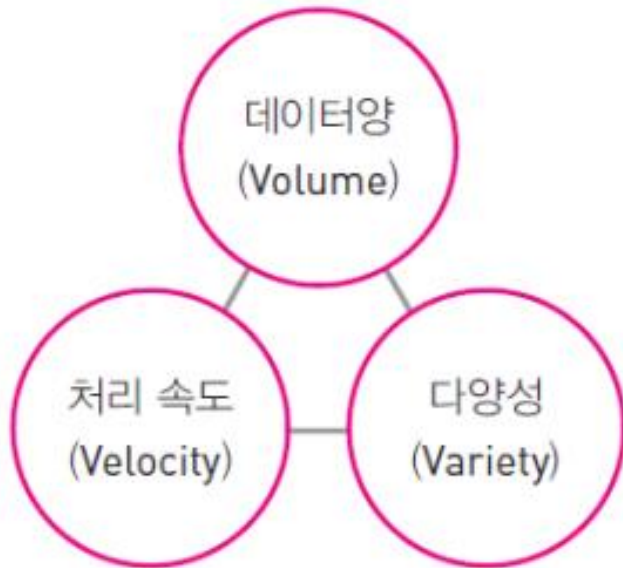
- 빅데이터란 단순히 대용량의 데이터만을 의미하는 것이 아니라 **대용량의 데이터를 활용·분석**하여 가치 있는 **정보를 추출**하는 정보화 기술을 총칭
- 초기에 데이터 규모와 기술적인 측면에서 출발했으나, 빅데이터의 가치와 활용효과 측면으로 의미가 확대되는 추세



빅데이터의 특성(1/2)

▪ 빅데이터 특성 - 3Vs

- **데이터 양(Volume)**: 큰 데이터 규모.(대용량의 레코드 혹은 컬럼)
- **처리 속도(Velocity)**: 데이터량의 빠른 증가 또는 적시성 있는 응대를 위한 빠른 데이터 처리 속도
- **다양성(Variety)**: 다양한 데이터 소스로 부터의 다양하고 복잡한 구조



빅데이터의 특성(2/2)

▪ 빅데이터 특성 - 분석 방법 측면

- 정확성(Veracity)

- 빅데이터는 현황 분석보다는 미래예측 또는 판별 결과의 정확성에 초점
- 결과의 정확성 기준이 주어지게 되며 이는 검증을 통해 측정될 수 있으므로, 검증이 함께 고려된다는 점

- 즉시성(Instancy)

- 빅데이터 분석은 데이터 처리 속도에 있어 즉시성 있게 필요한 결과를 도출
(기존 DW를 통한 현황 통계 분석은 지난달의 결과를 몇 주 후에 산출 하기도 함)
- 빅데이터는 즉시성 있게 필요한 결과를 도출
(경우에 따라 빠른 처리 속도로 실시간에 가까운 응답을 요구)

빅데이터 처리 과정



빅데이터 소스(1/6)

- 구조화 정도에 따른 **데이터 분류**
 - 비구조적 데이터(Unstructured data)
 - 구조화가 가능한 데이터 (Quasi-structured data)
 - 준구조화된 데이터(Semi-structured data)
 - 구조화된 데이터 (Structured data)



빅데이터 소스(2/6)

- **비구조적 데이터(Unstructured data)**
 - 데이터 항목에 대해 사전에 정의된 설명이 없거나 구조화가 어려운 데이터
 - 대략 전 세계 데이터의 80% 이상은 비구조적 데이터



아이폰5, 11월2일 출시 유력...26일부터 예약판매

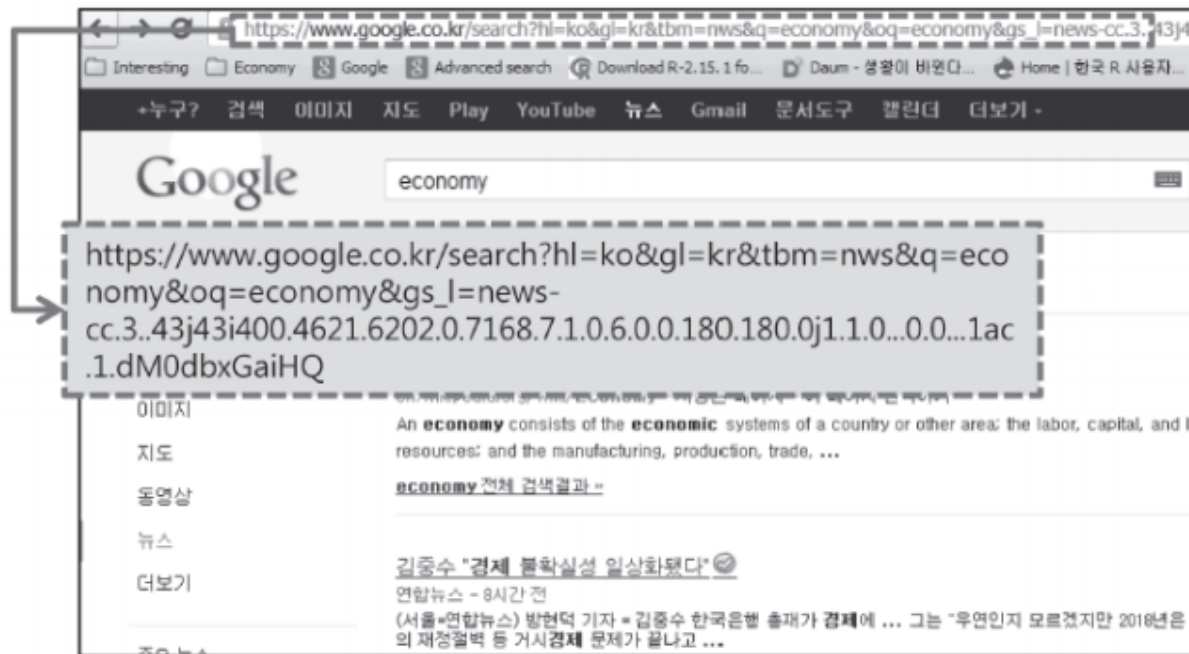
애플의 아이폰5가 결국 10월을 넘기고 11월 초에 출시될 것으로 보인다. 19일 업계에 따르면 애플의 아이폰5가 다음 달 2일 0시부터 국내 출시될 전망이다. 예약 판매는 26일 0시부터 진...

2012-10-19 16:08

[비구조적 데이터: 신문 뉴스 내용]

빅데이터 소스 (3/6)

- **구조화가 가능한 데이터(Quasi-structured data)**
 - 시간과 노력을 들이면 구조화 가 가능 → 데이터 구조에 대한 충분한 지식이 요구됨
 - 사용자가 방문한 URL(Uniform resource locator)은 “디렉터리명”, “화면 번호” 등 많은 정보를 포함.



[구조화가 가능한 데이터: 구글 검색 URL 주소에 담긴 키워드 데이터]

빅데이터 소스(4/6)

- **준구조화된 데이터(Semi-structured data)**

- 준구조화된 데이터는 정보 항목의 의미를 설명하는 태그(tags)나 기호(markers)를 가짐
- 마크업 언어(Markup Language), XML(eXtensible Markup Language), email, EDI(Electronic Data Interchange), HTML(HyperText Markup Language) 등 데이터는 정보 항목을 설명하는 내용을 자체에 포함

```
<?xml version="1.0"?>
<note>
  <to>Tove</to>
  <from>Jani</from>
  <heading>Reminder</heading>
  <body>Don't forget me this weekend!</body>
</note>
```

[준구조화된 데이터: XML 소스 코드]

빅데이터 소스(5/6)

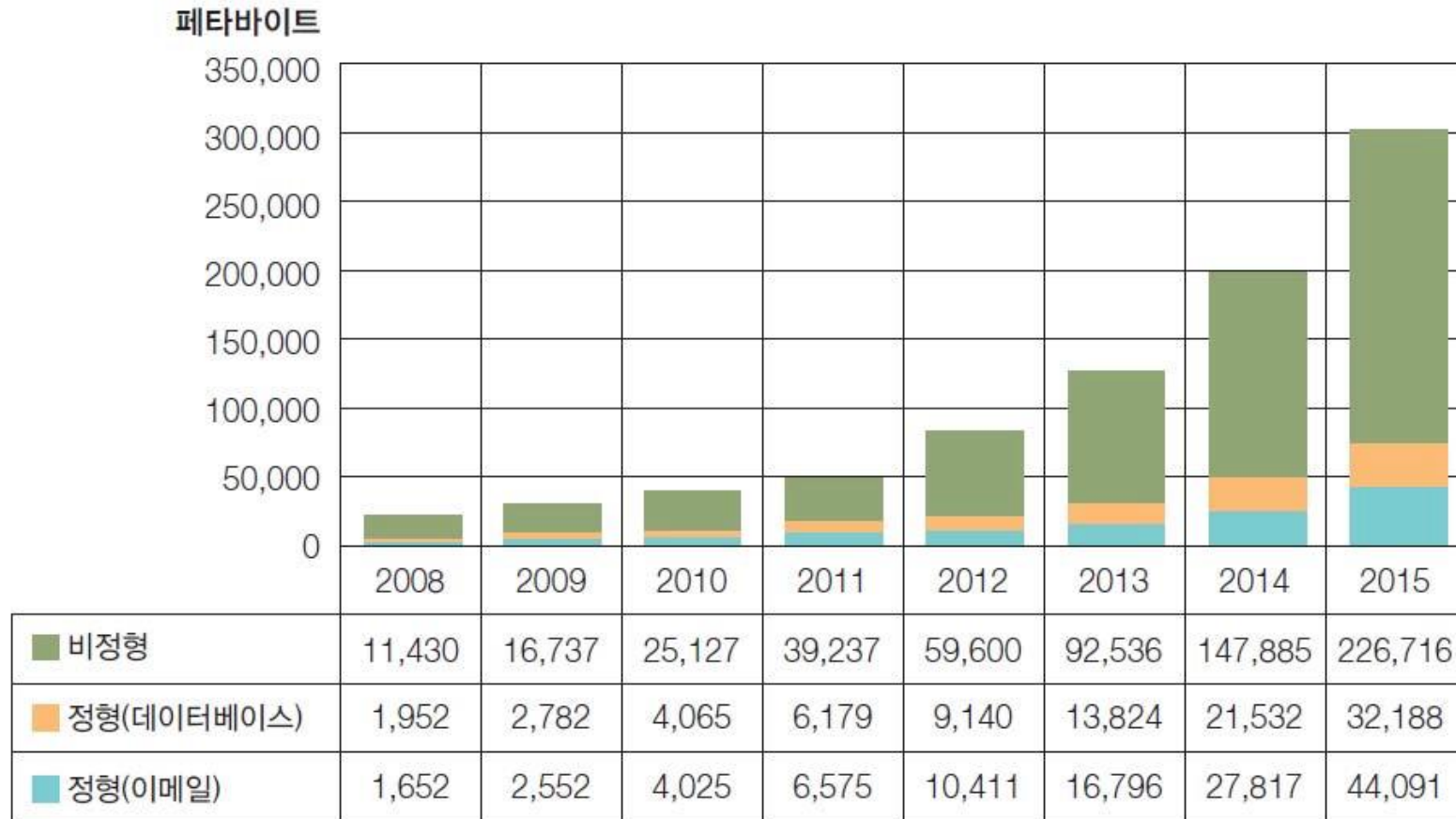
- **구조화된 데이터(Structured data)**
 - 데이터가 정형화된 속성으로 구조화가 가능한 경우
 - 스프레드시트(Excel)나 관계형 데이터베이스 (Relational Data Base Management System)

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

[구조화된 데이터: 차량 종류별 성능 비교표]

빅데이터 소스(6/6)

정형 vs 비정형 데이터



빅데이터 수집

■ 내부 데이터 수집

- 자체적으로 보유한 내부 파일 시스템, 데이터베이스, 관리 시스템, 센서 등에서 정형 데이터를 수집

■ 외부 데이터 수집

- 인터넷으로 연결된 외부에서 비정형 데이터를 수집

방법	설명
로그 수집기	내부에 있는 웹 서버의 로그를 수집, 즉 웹 로그, 트랜잭션 로그, 클릭 로그, 유의 로그 데이터 등 수집
크롤링	주로 웹 로봇으로 거미줄처럼 얽혀 있는 인터넷 링크를 따라 다니며 방문한 웹 사이트의 페이지라든지 소셜 데이터 등 인터넷에 공개되어 있는 데이터 수집
센싱	각종 센서로 데이터 수집
RSS 리더/오픈 API	데이터 생산 공유 참여 환경인 웹 2.0을 구현하는 기술로 필요한 데이터를 프로그래밍으로 수집
ETL (Extraction Transformation and Loading)	데이터의 추출, 변환, 적재의 약자로, 다양한 소스 데이터를 취합해 데이터를 추출하고 하나의 공통된 형식으로 변환하여 데이터 웨어하우스에 적재하는 과정 지원

빅데이터 저장

- 의미 있는 **정보 추출**을 위해서는 **효율적으로 데이터를 저장 관리**
- **대용량, 비정형, 실시간성**을 수용할 수 있는 **저장 방식**

접근 방식	설명	제품
분산 파일 시스템	컴퓨터 네트워크로 공유하는 여러 호스트 컴퓨터 파일에 접근할 수 있는 파일 시스템	GFS(Google File System) HDFS(Hadoop Distributed File System) Amazon S3 File System
NoSQL	데이터 모델을 단순화해서 관계형 데이터 모델과 SQL을 사용하지 않는 모든 DBMS 또는 데이터 저장 장치	Cloudata, Hbase, Cassandra
병렬 DBMS	다수의 마이크로프로세서를 사용하여 여러 디스크의 질의, 갱신, 입출력 등 데이터 베이스 처리를 동시에 수행하는 데이터 베이스 시스템	VoltDB, SAP HANA, Vertica, Greenplum, Netezza
네트워크 구성 저장 시스템	서로 다른 종류의 데이터 저장 장치를 하나의 데이터 서버에 연결하여 총괄적으로 데이터를 저장 및 관리	SAN(Storage Area Network) NAS(Network Attached Storage)

빅데이터 처리

- 방대한 양의 데이터와 데이터 생성 속도, 데이터 종류의 다양성을 통합적으로 고려하여 처리
- 실시간 분석의 중요성이 증가되면서 **실시간 스트림 데이터** 처리를 위한 기술이 개발

용어	설명
빅데이터 일괄 처리 기술	<ul style="list-style-type: none"> - 빅데이터를 여러 서버로 분산하여 각 서버에서 나누어 처리하고, 이를 다시 모아서 결과를 정리하는 분산 병렬 기술 방식 - 구글 맵리듀스(구글에서 분산 컴퓨팅을 지원할 목적으로 제작 발표한 소프트웨어 프레임워크, 함수형 프로그래밍에서 일반적으로 사용되는 Map과 Reduce 함수를 기반으로 주로 구성)하둡 맵리듀스, 마이크로소프트 Dryad 등이 있음
빅데이터 실시간 처리 기술	스트림 처리 기술로 강화된 스트림 컴퓨팅을 지원하는 IBM의 InfoSphere Streams, 분산 환경에서 스트리밍 데이터를 분석할 수 있게 해주는 트위터의 Storm
빅데이터 처리 프로그래밍 지원 기술	분산 데이터를 처리하는 프로그래밍 언어인 구글의 Sawzall과 명렬 처리를 하는 고성능 데이터-플로우 언어와 실행 프레임워크인 하둡 Pig

빅데이터 분석(1/2)

- 통계학, 전산학, 기계 학습과 데이터 마이닝에서 사용하던 기술을 활용
- 대규모 데이터를 처리할 수 있도록 처리 방식으로 개선하거나 보완

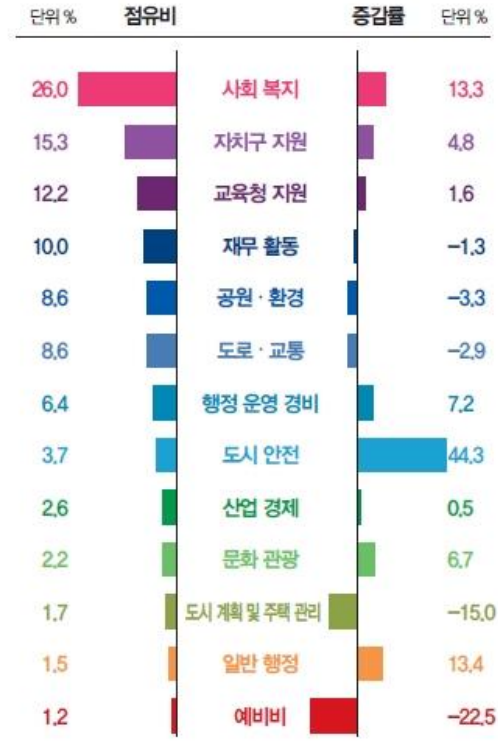
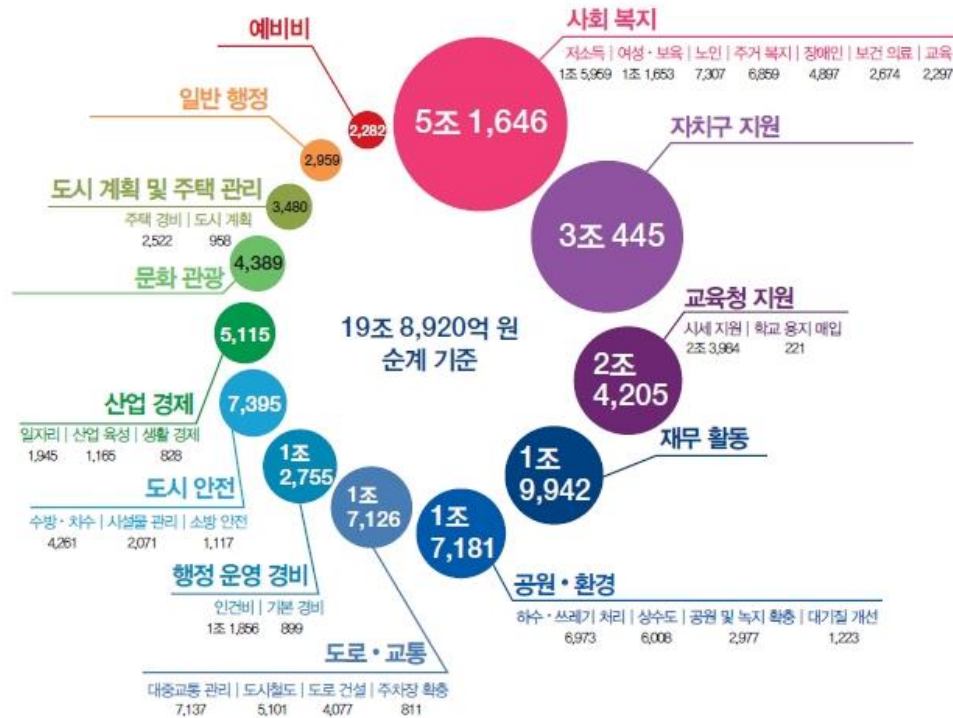
용어	설명
텍스트 마이닝 (Text Mining)	- 자연어 처리 기술을 사용해 인간의 언어로 쓰인 비정형 텍스트에서 유용한 정보를 추출하거나 다른 데이터와 연계성을 파악하여, 분류나 군집화 등 빅데이터에 숨겨진 의미 있는 정보를 발견하는 것
웹 마이닝 (Webb Mining)	- 인터넷에서 수집한 데이터를 데이터 마이닝 기법으로 분석하는 것
오피니언 마이닝 (Opinion Mining)	- 다양한 온라인 뉴스와 소셜 미디어 코멘트, 사용자가 만든 콘텐츠에서 표현된 의견을 추출 분류 이해하고 자산화 하는 컴퓨팅 기술 - 텍스트 속의 감성과 감동, 여러가지 감정 상태를 식별하려고 감성분석 사용 - 마케팅에서는 Buzz(입소문) 분석이라고도 함
리얼리티 마이닝 (Reality Mining)	- 휴대폰 등 기기를 사용하여 인간 관계와 행동 양태 등을 추론 하는 것 - 통화량, 통화 위치, 통화 상태, 대상, 내용 등을 분석하여 사용자의 인간관계, 행동 특성 등 정보를 찾아냄
소셜 네트워크 분석 (Social Network Analysis)	- 수학의 그래프 이론을 바탕으로 소셜 네트워크 서비스에서 소셜 네트워크 연결 구조와 연결 강도를 분석하여 사용자의 명성 및 영향력을 측정하는 것

빅데이터 분석(2/2)

용어	설명
분류 (Classification)	<ul style="list-style-type: none"> - 미리 알려진 클래스들로 구분되는 훈련 데이터군을 학습시켜 새로 추가 되는 데이터가 속할 만한 데이터군을 찾는 지도 학습(Supervised Learning) 방법 - 가장 대표적인 방법으로 KNN(K-Nearest Neighbor)이 있음
군집화 (clustering)	<ul style="list-style-type: none"> - 특성이 비슷한 데이터를 합쳐 군(group)으로 분류하는 학습 방법 - 분류와 달리 훈련 데이터군을 이용하지 않기 때문에 비지도 학습(unsupervised Learning) 방법 - 트위터에서 주로 사진/카메라를 논의하는 사용자군과 게임에 관심있는 사용자군 등 관심사나 취미에 따라 분류
기계학습 (machine Learning)	<ul style="list-style-type: none"> - 인공지능 분야에서 인간의 학습을 모델링 한것 - 컴퓨터가 학습할 수 있도록 하는 알고리즘과 기술을 개발하여 수신한 이메일의 스팸 여부를 판단할 수 있도록 훈련 - 결정 트리(decision Tree) 등 기호적 학습, 신경망이나 유전자 알고리즘 등 비기호적 학습, 베이지안(bayesian)이나 마코프(hidden Markov) 등 확률적 학습 등 다양한 기법이 있음
감성 분석 (sentiment Analysis)	<ul style="list-style-type: none"> - 문장의 의미를 파악하여 글의 내용에 긍정/부정, 좋음/나쁨을 분류하거나 만족/불만족 강도를 지수화, 그런 다음 이 지수를 이용하여 고객의 감성 트렌드를 시계열적으로 분석하고 고객 감성 변화에 기업의 신속한 대응 및 부정적인 의견의 확산을 방지하는데 활용

빅데이터 수집

- 특정 기준에 따라 분석한 데이터의 특징이나 분석 결과를 분석가와 사용자들이 쉽게 이해할 수 있도록 **그림**이나 **그래프** 등으로 **표현**
- 빅데이터 분석에서는 표현 단계가 가장 중요한 기술 분야로 취급



순계 : 총예산에서 회계 간 중복분 제외한 실질적인 예산

빅데이터 수집

제품/기술	최초 개발	최초 공개	주요 기능 및 특징
Tag Cloud	D. Coupland	1995년	태그의 연관성에 따른 빈발도 및 관계 분석
Gephi	M. Bastian	2008년	데이터를 네트워크 형태로 생성 후 표현
GraphViz	AT&T	1988년	흐름도나 트라 다이어그램 생성 표현 툴
Processing	MIT Media Lab	2001년	그래픽 디자인을 위한 프로그래밍 언어
Fusion Tables	Google	2009년	대용량 데이터를 표현해 주는 온라인 서비스
Tableau	P. Hanrahan	2003년	데이터의 시각적 분석과 리포팅 도구 제공
TinkerPro	A. Averbuch	2009년	그래프를 처리하는 통합 서비스형 시스템
Clustergram	M. Schonlau	2002년	계층적 군집화에 적합한 데이터 표현 툴
Spatial Information Row	-	-	특정 정보를 기준으로 데이터 흐름 표현

빅데이터 응용 사례: 한국관광공사 외국인 관광산업 지원

➤ 추진 배경

- 중국인 관광객 10년 전 대비 500% 증가, 외래 관광객중 가장 높은 비중 과 소비 규모, 관광 제도 개선으로 인한 자유여행객의 증가 예상
- 외국인 관광객 1000만시대, 중국인 관광객은 전체 관광객의 33%로 1위, 1인당 경비 지출이 중국인 관광객이 가장 높음
- 중국 단체 여행객의 감소, 개별 여행객 수가 증가하고 있음
- 중국인 관광객의 획일화된 관광지 및 관광 패턴, 관광 일정으로 인해 재방문률의 정체와 재방문 의향이 낮은 상황
- 국내 입국 관광객의 모바일 인프라를 활용하여 보다 다양한 국내 관광 정보의 제공을 통한 관광 만족도를 높일 필요성

빅데이터 응용 사례: 한국관광공사 외국인 관광산업 지원

➤ 추진 목적

- 중국인 관광객의 여행 패턴 및 소비 패턴 분석을 통해 중국인 관광객 대상 정책 활용 및 사업 환경 개선에 활용
- 중국인 관광 트렌드를 분석하여 개인 맞춤형 관광 정보 제공, 추가 관광지 개발, 관광지 추천 정보 제공

빅데이터 응용 사례: 한국관광공사 외국인 관광산업 지원

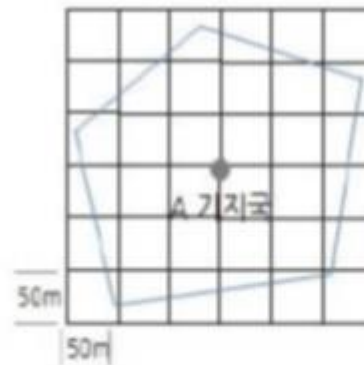
➤ 추진 내용(1)

▪ 중국인 유동 인구 산출

- KT 로밍 데이터를 활용한 일단위 서울시 지역에 대한 법정 동 단위 유동 인구 추출
- 통화 발생 기지국 커버리지를 포함하는 50m x 50m 셀 단위의 유동 인구 정보 분석



중국인 유동인구 데이터 기반
서울시 법정동 리스트 선정



일자별/시간대별/법정동별
상위 15개 기지국 및 셀 격자 추출

- 중국인 관광객 및 내국인 유동인구 데이터 분석
- 시간대별 내/외국인 자주 가는 지역 패턴 비교 분석 활용
- 결제 데이터와의 데이터 융합을 통한 데이터 시너지 확보

빅데이터 응용 사례: 한국관광공사 외국인 관광산업 지원

➤ 추진 내용(2)

- BC 카드 데이터를 활용한 상권별 중국인 매출액 분석
 - 고객 유형 정보 분석: 블록단위(거주지주소기준) 성, 연령 고객 유형
 - 블록 단위 카드 거래의 성별/연령대별, 시간대별/요일별 패턴 내국인 거래 패턴 분석
 - 블록 단위 카드 거래 실적 데이터를 통해 중국인 거래 패턴 분석



빅데이터 응용 사례: 한국관광공사 외국인 관광산업 지원

➤ 추진 내용(3)

▪ 여행 정보 안내

- 중국인 대상 모바일앱 서비스에 데이터 분석 결과를 활용
- 지도기반의 여행지 안내, 여행 코스 추천, 인기 가맹점 소개 등 정보 제공



빅데이터 응용 사례: 존스킨화장품 신제품 발굴

➤ 추진 배경

- 존스킨한의원이 2008년에 런칭한 존스킨화장품은 한의원의 연구 자료를 바탕으로 한방화장품을 제조 및 판매하는 업체
- 전문 한방 코스메슈티컬 브랜드로 시작되어 2009년부터 고객확대를 위해 유통을 확대
- 존스킨화장품의 고객들은 주로 한의원을 방문하는 고객들과 치료/개선 효과를 필요로 하는 고객
- 지속적인 성장과 한방화장품의 시장 확대를 위하여 대중들에게 쉽게 다가갈 수 있는 제품 개발이 필요



빅데이터 응용 사례: 존스킨화장품 신제품 발굴

➤ 추진 목적

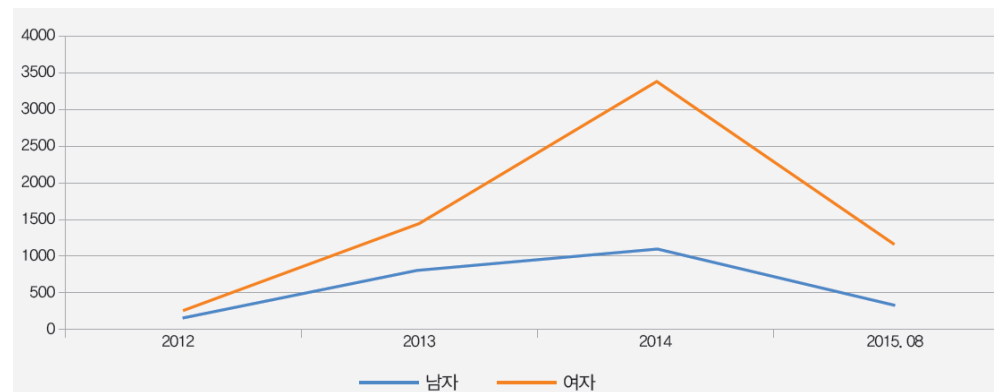
- 대중적이면서도 한방화장품의 특성을 살린 신제품 출시가 필요
- 신제품 출시를 위한 Target 고객과 이들이 원하는 요구를 분석
 - 어떤 고객층을 대상으로 해야 할까?
 - 고객의 욕구는 무엇일까?

빅데이터 응용 사례: 존스킨화장품 신제품 발굴

➤ 추진 내용(1)

▪ 고객 분석

- SNS(Social Network Service) 데이터를 분석
- 고객 트렌드 분석결과 중 존스킨화장품의 눈길을 끈 것은 우선 남성고객에 대한 언급량이 지속적으로 증가하고 있는 추세
- 남성들도 외모를 잘 가꾸어야 성공할 수 있다는 사고가 사회전반에 형성되면서, 패션과 미용에 아낌없이 투자



빅데이터 응용 사례: 존스킨화장품 신제품 발굴

➤ 추진 내용(2)

▪ 구매 목적 분석

- 고객들이 화장품을 살 때 어떤 목적을 가지고 구매하는지를 분석
- 구매목적과 관련된 키워드를 분석한 결과, ‘어머니’와 ‘선물’에 대한 언급이 가장 많았지만, ‘아버지, 남편, 남친, 부모님’도 주요 키워드로 분석

Taxonomy	키워드	총합계
목적	어머니	986265
	선물	808973
	친구	616401
	가족	48427
	아버지	28454
	남편	23431
	남친	21352
	부모님	13075
	선물용	8139
	여친	7096
	여자친구	4551
	선물하다	3274
	남자친구	1230
	선물상품	1134

빅데이터 응용 사례: 존스킨화장품 신제품 발굴

➤ 추진 내용(3)

▪ 제품 기능 도출

- 남성으로 신제품의 Target을 결정한 존스킨화장품은 구체적인 제품을 기획하고 화장품 구입의 주 목적 분석
- 피부, 여드름, 피부관리, 트러블키트 등이 주요하게 언급되는 키워드로 ‘피부트러블’에 대한 수요가 많은 것으로 파악
- ‘피부트러블’을 관리하기 위한 것이라는 빅데이터 분석결과를 가지고 ‘피부트러블’을 주제로 한 남성전용화장품을 기획

연관어	빈도
치료	34349
효과	14553
차	14382
건강	13332
피부	11867
추천	10957
어머니	9977
전문	8620
피부과	6390
여드름	6192
성분	4732
천연	4024
크림	3189
샴푸	2979
향	2905
시술	2757
한의원	2429
할인	2353
탈모	1764
주름개선	1293
원료	1280
아토피	930

빅데이터 응용 사례: 신한카드 고객 맞춤형 타겟 마케팅

➤ 추진 배경

- 이용패턴 분석으로 최적화된 맞춤형 마케팅 전략 수립과 효과 제고
- 카드 사용 고객의 관점에서 생각하는 소비 생활 서비스 및 상품 제공으로 매출 증가 기대
- 특정 고객을 타겟으로 설정하는 마케팅 보다 현재 보유하고 있는 고객의 성향을 분석하여 영업 전략 도출

➤ 추진 목적

- 카드 비즈니스의 본원적인 경쟁력 강화와 고객만족경영을 위해 공익성을 담은 빅데이터 기반 사업의 강화
- 고객의 다양한 요구와 최신의 니즈를 명확하게 반영하여 상품에 반영
- 브랜드 가치 및 고객만족도 제고, 사업 포트폴리오 강화, 차별화된 금융상품 및 서비스 제공

빅데이터 응용 사례: 신한카드 고객 맞춤형 타겟 마케팅

➤ 추진 내용(1)

▪ 트렌드 분석

- 고객의 다양한 요구와 최신의 요구를 명확하게 파악하고 상품에 반영
- 카드 결제 내역을 분석해서 기존 고객의 이용 특성을 분석하여 현 상황 파악 및 향후 고객 유치 과정에서 알맞은 카드 서비스를 제공할 수 있는 전략 수립

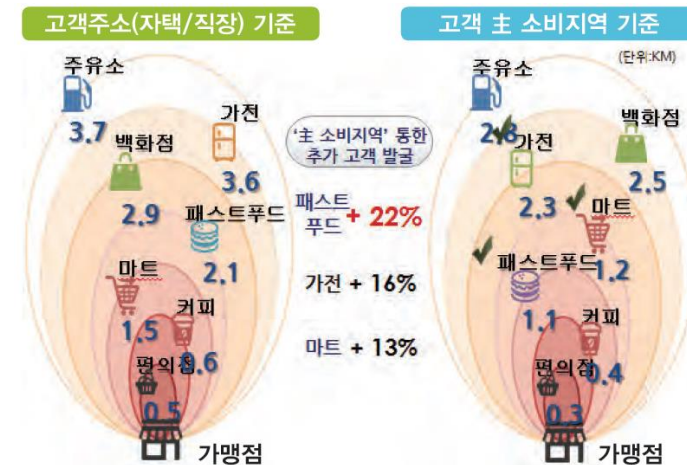
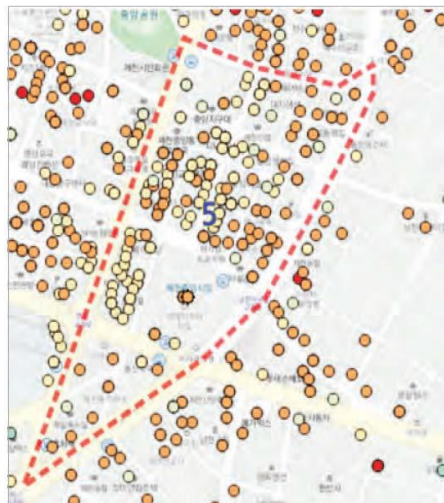


빅데이터 응용 사례: 신한카드 고객 맞춤형 타겟 마케팅

➤ 추진 내용(2)

▪ 카드 사용자 분석

- 특정 지역의 업종을 음식, 관광/레저, 편의점/쇼핑으로 구분하여 소비자가 결제한 상점에 대해 분석한 것으로 고객의 소비 성향과 카드 사용 집중 지역을 다룬 정리된 데이터 확보 가능
- 가맹점에서 파악하기 어려운 고객의 주 소비지역 거리 데이터를 통해 정확도나 이용도가 높은 타겟 발굴



빅데이터 응용 사례: 파riba게트 날씨 마케팅

➤ 추진 배경

- 국내에서는 최근 케이웨더, 웨더뉴스 등 기상정보 제공업체를 통해 고정적으로 날씨 정보를 회원들에게 제공하는 업체 증가
- 기업에서는 기상 정보를 분석하여 이를 적극적으로 마케팅에 활용하고 있으며 날씨에 따른 전략상품으로 차별화
- 기상 특징이 서로 다른 지역은 지역의 기후에 적합한 제품군을 선별해 전략상품으로 활용

➤ 추진 목적

- 전국 기상 관측 자료와 점포별 판매 데이터 분석
- 날씨에 따라 제품 선호도가 바뀌는 점을 활용하여 재고 관리 및 마케팅에 활용

빅데이터 응용 사례: 파리바게트 날씨 마케팅

➤ 추진 내용(1)

▪ 날씨판매 지수

- 날씨 판매지수는 날씨에 따른 판매율을 나타내는 지수
- 최근 5년간 전국 169개 지점의 일별 매출과 기상자료를 통계기법으로 지수화한 지수 개발
- 가맹점에서는 판매량을 예측하고 주문량을 조절할 수 있는 찬스로스를 방지하고 재고 부담 감소
- 기상관측 자료와 10억 건 이상의 점포별 상품 판매 데이터를 분석하여 실시간으로 전국 3,100여개 파리바게트 점포의 단말기에 제공

빅데이터 응용 사례: 파리바게트 날씨 마케팅

➤ 추진 내용(2)

- 날씨 예측을 통해 판매 수요 예측 및 생산 관리 - 날씨별로 판매가 높은 빵을 파악하여, 점포 단말기로 주문량 권장 가능
- 파리바게트 매장의 계산대 단말기 화면에는 ‘일별 날씨 판매지수 최대 변동’ 이라는 항목에 제품 이름이 표기
- 날씨 지수를 도입한지 한달만에 조리빵 매출 30% 증가



빅데이터 시장(1/2)

- 한국 빅데이터 시장, **연평균 24% 이상 성장 전망**

세계 빅데이터 시장규모 전망



빅데이터 주요 세그먼트별 국내 시장규모 전망



빅데이터 시장(2/3)

➤ 글로벌 IT 기업의 빅데이터 추진 현황

기업명	빅데이터 사업 추진 현황
EMC	<ul style="list-style-type: none"> - 데이터 저장부터 관리, 분석까지 빅데이터와 관련한 모든 것을 제공하기 위해 그린 플럼, 아이실론 등 빅데이터 솔루션 업체 및 데이터 관련 다수 업체 인수 - 빅데이터 스토리지 솔루션(아이실론, 아트모스), 콘텐츠 관리 솔루션(다큐멘텀)
HP	<ul style="list-style-type: none"> - BI 솔루션 업체 버티카, 기업용 검색 엔진 업체 오토노미 인스 - 버티카와 오토노미를 결합하여 빅 데이터 시장에 진입 - 기업경영 의사 결정, 경영정보 분석 등 경영 지원 전략 수립 서비스 제공
IBM	<ul style="list-style-type: none"> - 분석용 데이터 저장 관리 업체(네티자), 데이터 통합업체(에센셜), 분석 솔루션 업체(코그너스) emd 비즈니스 분석 관련 업체 인수 - 빅데이터 솔루션: infosphere Biginsight(Hadoop), InforSphere Streams
Oracle	<ul style="list-style-type: none"> - 세계적인 DB 업체 하이페리온 인수로 분석 기술 확보 - 오라클 빅데이터 어플라이언스 제품 출시
SAS	<ul style="list-style-type: none"> - 고급 분석(advanced Analytics)을 위한 HAP(high Performance Analytics) 기반의 SEMMA 방법론 제시 - IT+분석+비즈니스 통합 플랫폼구현SAS 빅데이터 분석 플랫폼 Solution MAP)
Teradata	<ul style="list-style-type: none"> - 데이터웨어하우징 및 비즈니스 인텔리전스(BI) 전문 업체 - 비정형 데이터의 고급분석, 관리 솔루션 업체 인수(애스터데이터) - 애스터 맵리듀스 플랫폼 제시

빅데이터 시장(4)

➤ 국내 IT 기업의 빅데이터 추진 현황

기업명	빅데이터 사업 추진 현황
그루터	<ul style="list-style-type: none"> - 빅데이터 관련 플랫폼, 기술, 솔루션 전문 기업 - 빅데이터 플랫폼 구축 및 컨설팅 서비스, 빅데이터 분석 및 데이터 제공 서비스, 빅데이터 분석 플랫폼 제공 서비스 구축
넥스알	<ul style="list-style-type: none"> - 넥스알 빅데이터 분석 플랫폼(NDAP: Next Data Analytics Platform). 넥스알 빅데이터 분석 솔루션(RHve) 구축
다음소프트	<ul style="list-style-type: none"> - SNS 정보 기반 여론 진단 서비스, 소셜미디어 트위터, 블로그 트렌드 분석 - 소셜미디어 상의 데이터들에서 의미 있는 정보를 찾고, 조직화함으로써 정보간의 관계나 패턴, 트렌드 등을 분석하는 서비스 제공
사이람	<ul style="list-style-type: none"> - 소셜네트워크 분석 소프트웨어 넷마이너(netminer) 개발: 대규모 소셜 네트워크 및 데이터 간의 관계를 계량적으로 분석해 패턴을 파악하고 시각화 하는 기능을 제공 - 소셜 네트워크 분석 응용솔루션 및 컨설팅 제공
솔트룩스	<ul style="list-style-type: none"> - 비정형 빅데이터 분석 및 시멘틱 기술 전문 기업 - 비정형 빅데이터 분석 플랫폼(Trueestory), 클라우드 기반 시멘틱 검색 플랫폼(in2), 시멘틱 기반 빅데이터 추론 플랫폼(STORM), 빅데이터 분석 서비스 플랫폼(O2)서비스

해외 빅데이터 활용 서비스

구분	내용
미국 국세청, 탈세방지시스템	<ul style="list-style-type: none"> 방대한 자료로부터 이상 징후를 찾아내고 예측 모델링을 통해 과거의 행동 정보를 분석하여 사기 패턴과 유사한 행동 검출 페이스북이나 트위터를 통해 범죄자와 관련된 소셜 네트워크를 분석하여 범죄자 집단에 대한 감시 시스템 마련
뉴욕주 시라큐스시, 스마트 시티	<ul style="list-style-type: none"> 데이터 분석을 통해 낙후된 지역에서 공통적으로 나타나는 현상을 파악 수집된 정보를 분석·통합하여 낙후지역의 특성에 맞춘 새로운 도시 개발 모델 제시
덴마크, 베스타스 윈드 시스템	<ul style="list-style-type: none"> 바람의 방향 및 높이에 따른 변화요소, 날씨, 조수 간만의 차, 위성 이미지, 지리 데이터, 날씨 모델링 조사 등의 데이터 이용 동력을 얻기 위한 터빈의 관리 및 배치를 위해 데이터 분석 시스템을 이용하여 에너지 효율성을 높일 수 있는 방안 마련
엘마트, 일마트랩	<ul style="list-style-type: none"> 일마트랩을 이용한 소비자 소비 패턴 조사를 통해 점포 운영에 반영 소셜 미디어를 통해 대규모 데이터를 수집하여 리얼타임으로 해석된 추출된 정보를 이용하여 상품판매를 촉진하는 기법
FBI, 범인 검거 체계	<ul style="list-style-type: none"> 미제 사건 용의자 및 실증자에 대한 DNA 정보를 포함한 12만 명의 범죄자 DNA 정보 저장 1시간 내에 범인 DNA 분석을 위한 주정부 데이터베이스 연계 및 빅데이터 실시간 분석
코카콜라, 마케팅 전략	<ul style="list-style-type: none"> 트위터, 페이스북에서 발생하는 코카콜라 관련 데이터를 수집 분석 코카콜라에 비우호적인 정보가 증가하는 국가나 지역을 대상으로 홍보를 강화하는 등 실시간 대응 가능
할리우드, 홍행 수익 예측	<ul style="list-style-type: none"> 트위터 등 소셜 네트워크서비스(SNS)가 미국 할리우드에서 영화의 흥행 여부를 미리 판단 영화사들은 개봉 전 SNS를 면밀히 검토해 영화에 대한 고객들의 관심 정도를 파악하고 홍보 예산 책정
미국 국립보건원, 유전자 분석	<ul style="list-style-type: none"> 1,700명의 유전자 정보를 아마존 클라우드에 저장하여 누구나 데이터 이용 가능 난치병 및 불치병과 관련된 유전자 정보를 공유하고 분석함으로써 새로운 치료제 개발 가능성 제시

국내 빅데이터 활용 서비스

구분	내용
국민권익위원회, 민원정보분석 시스템	<ul style="list-style-type: none"> 홈페이지 민원, 제안, 콜센터 상담 등을 통해 축적된 민원 데이터를 종합적·체계적으로 분석하여 정책에 환류할 수 있도록 지원 과거 민원 발생 현황 등을 분석한 월별, 지역별 주요 민원 캘린더를 제작하여 사회적 이슈를 민원 지도 형태로 제공
한국도로공사, 고객 목소리 분석 시스템	<ul style="list-style-type: none"> 콜 상담서비스, 민원관리 시스템, 채팅 상담 시스템을 고도화된 언어처리 기법으로 분석하여 고객만족활동에 도움이 될 수 있는 지표와 이슈 도출 주요 이슈 사항이나 불만을 사전에 파악하여 대응할 수 있는 기반 마련하고 서비스 전략 수립이나 정책 수립을 위한 의사결정 지원
한국수자원공사, 스마트 워터 그리드	<ul style="list-style-type: none"> 수도관 중간에 유량, 수질, 유수율(물공급량과 수도요금의 비율) 등을 관측하는 센서가 설치되어 있어 수도관의 정보를 실시간으로 관제 컴퓨터에 전송 수도관 파손으로 누수가 발생하면 수도관에 설치된 센서에서 누수량 감지 및 누수 위치 파악
포스코, 원자재 구매 시스템	<ul style="list-style-type: none"> 철광석 가격에 영향을 미치는 남미, 호주 광산의 상황과 런던 금속거래소(LME)를 통해 수집한 광물 가격 데이터를 실시간으로 분석 고객사의 수요 데이터와 전 세계 철광산 및 현물 거래소의 가격 데이터를 조합한 후 비교하여 철광석 구매의 최적 타이밍과 가격대 결정
SK텔레콤, 스마트인사이트 시스템	<ul style="list-style-type: none"> 기업들이 원하는 키워드를 중심으로 온라인 여론을 분석하여 실시간으로 제공 기업의 평판을 실시간으로 모니터링하여 기업의 대응전략 마련을 지원
보건복지부, 수용자 중심의 복지서비스	<ul style="list-style-type: none"> 지방 자치단체에서 집행하는 약 120여 개의 복지급여 및 서비스 이력을 개인별, 가구별로 통합 관리 복지급여 지급과정에서 지급내역의 임의수정을 통한 부정 소지 차단 및 실명 확인 후 입금으로 제정의 투명성
분당 서울대병원, 임상의사결정지원 시스템	<ul style="list-style-type: none"> 현재 보유한 진료기록이 60TB 규모로 처리속도가 느려짐에 따라 빅데이터 분석으로 시간 단축 및 자연어처리 기능 강화 분당 서울대병원에서 부적절한 용량의 신독성 약물 처방률이 30.6%로 감소

감사합니다 !



for details about this presentation, please email to sschoi@wsu.ac.kr