

빅데이터의 이해

3주차 : 빅데이터와 미래기술

성공회대학교 열림공과대학
김종울

3. 빅데이터 기술에 대한 이해

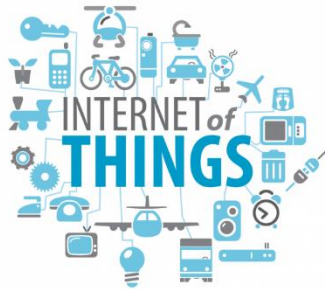
--- 강의 목표 ---

빅데이터 기술에 대한 이해

빅데이터 기술

빅데이터 기술

Related technology with big data



- Data mining
- Text mining



데이터 수집

웹, 일상 등 모든 분야에서 데이터를 수집.



STEP 1.
COLLECTING



STEP 3.
PROCESSING



STEP 5.
VISUALIZATION



STEP 2.
STORING



STEP 4.
ANALYSIS

데이터 저장

다양한 방식 다양한 환경에 효율적으로 저장.



데이터 분석

데이터로부터 유의미한 정보를 추출.



데이터 시각화

데이터에 의미를 부여하여 효율적으로 전달.



빅데이터 기술

빅데이터 기술



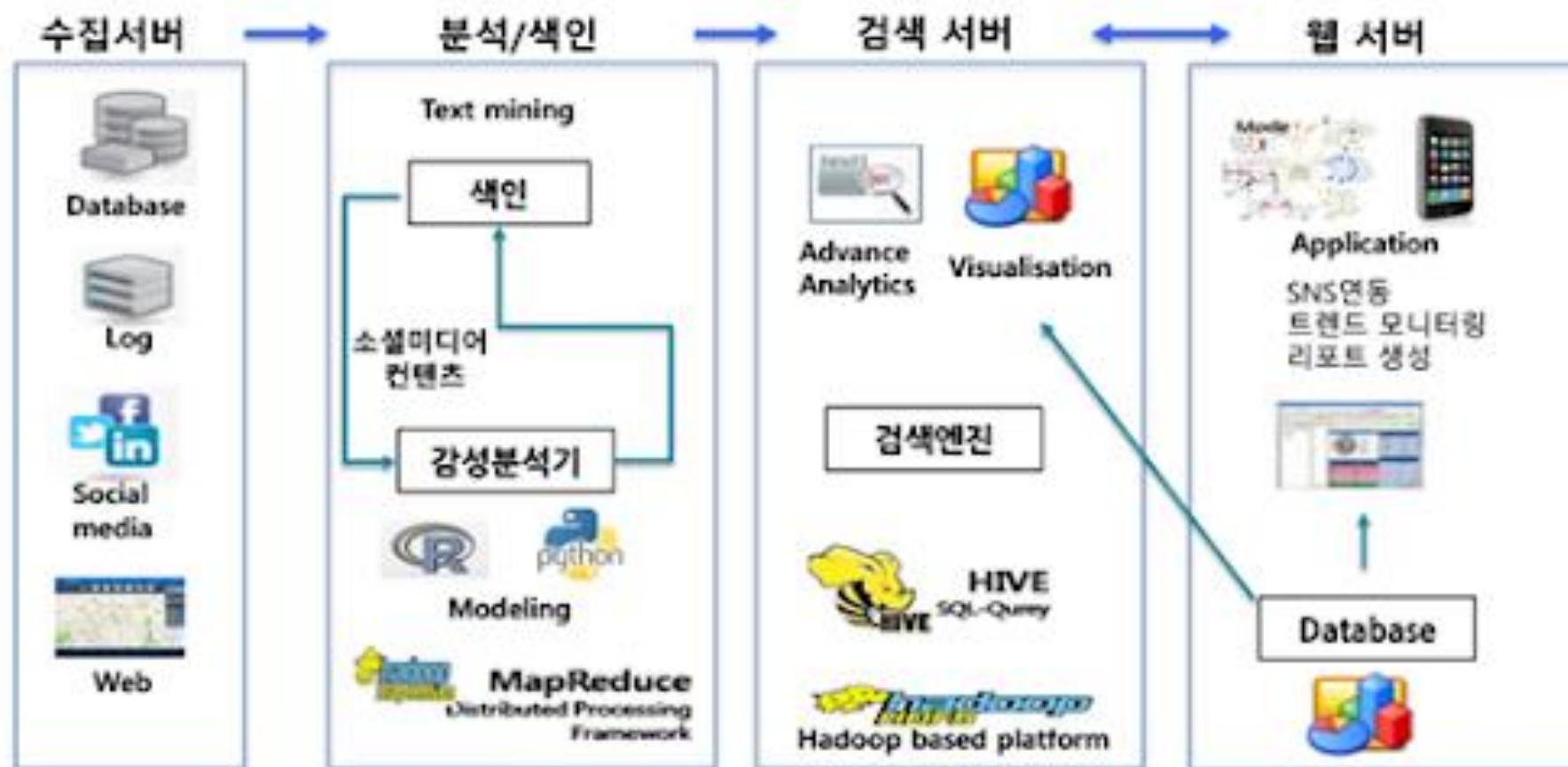
빅데이터 기술

빅데이터 기술



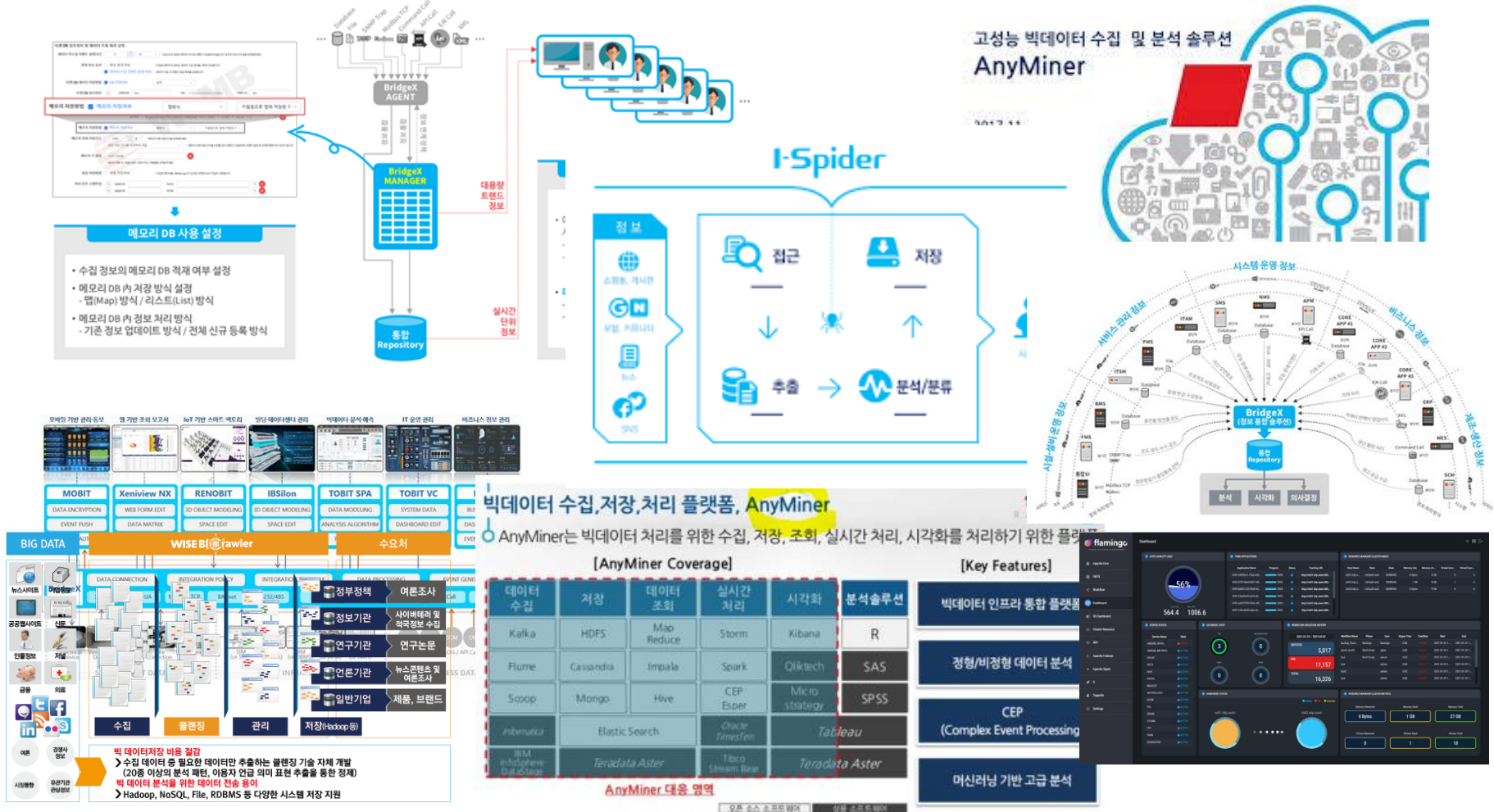
빅데이터 기술

빅데이터 기술



빅데이터 기술

데이터 수집(데이터 수집)



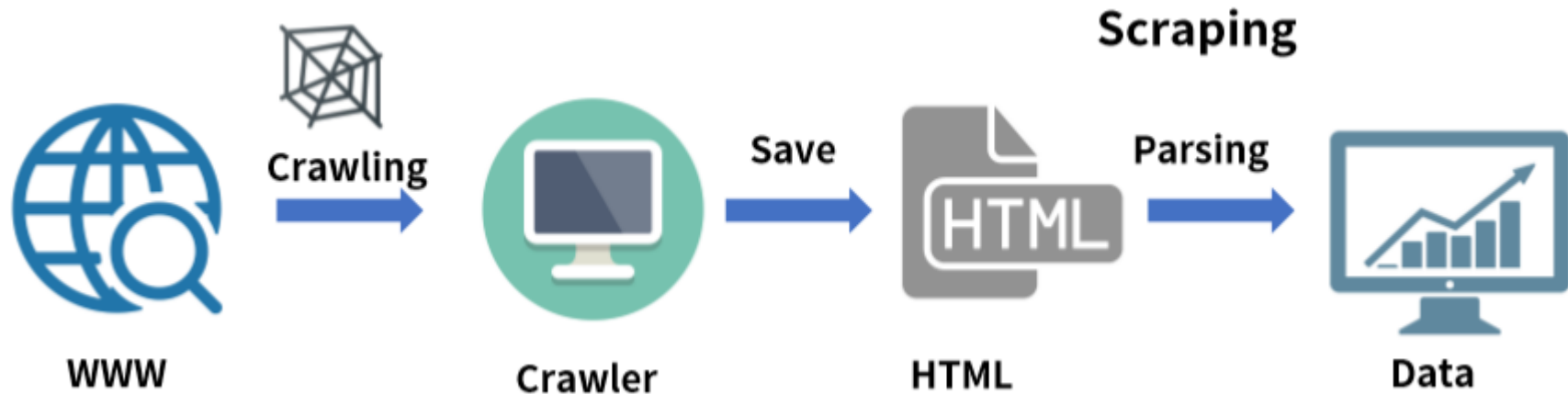
빅데이터 기술

데이터 수집(WEB 데이터 수집)

크롤링(Crawling)

다양한 정보를 활용하기 쉽도록 수집하는 행위

크롤러(Crawler): 크롤링을 하는 프로그램



빅데이터 기술

데이터 수집(WEB 데이터 수집)

Web crawling:

웹에서 링크를 타고다니며 웹페이지들을 수집하는 행위

예) Google bot이 전 세계의 웹사이트를 수집하여 검색 서비스 제공

Web scraping:

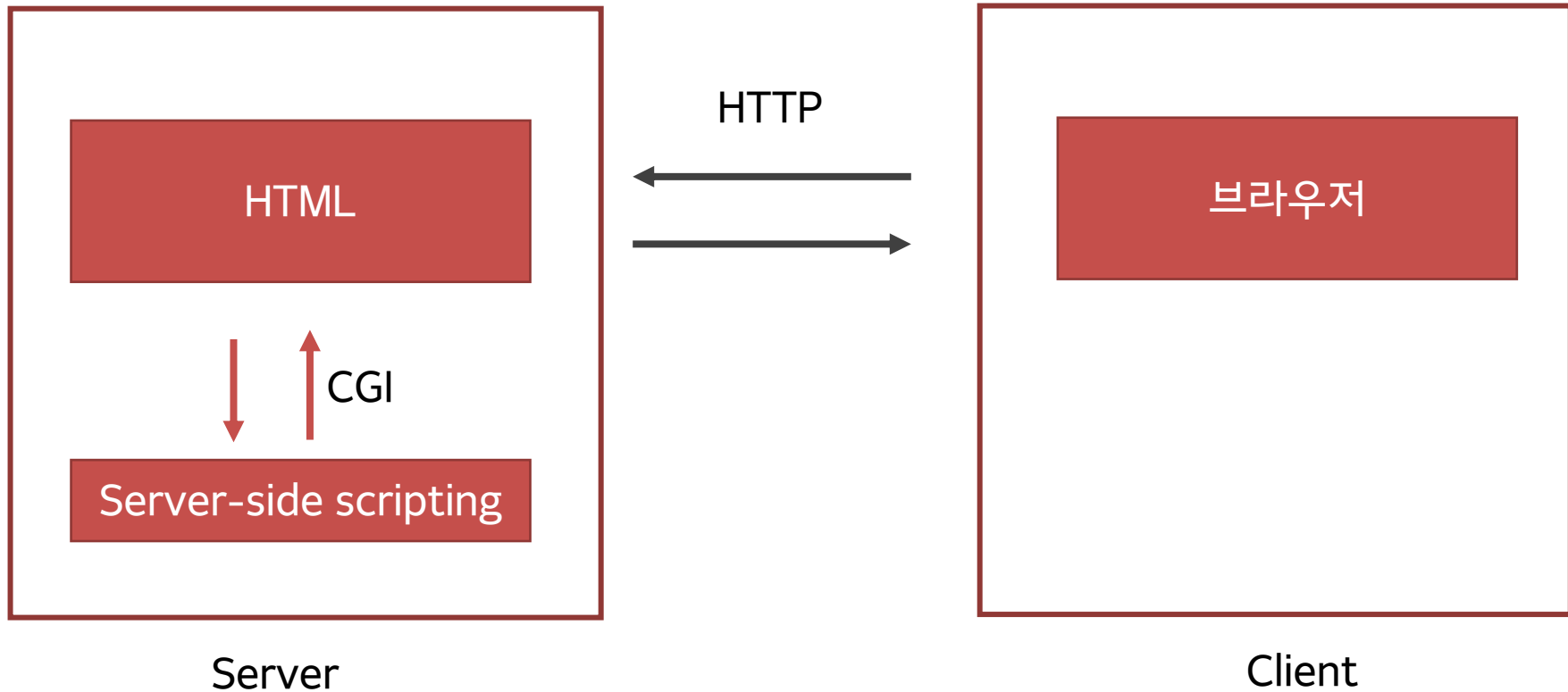
웹사이트에서 정보를 추출하는 행위

예) 쇼핑몰 가격비교 - 각 쇼핑몰 상품 페이지에서 상품 이름, 가격 등을 추출

빅데이터 기술

데이터 수집(WEB 데이터 수집)

웹서버의 구조와 크롤러 / 스크래핑에 대한 설명



빅데이터 기술

데이터 수집(데이터 수집)

대학안내

대학원안내

대학생활

학생지원

학사안내

열린강좌

총학생회

동아리

취업정보

연구/산학

WITH INU

서비스지킴이

부속기관 사이트

바로가기

GO

메뉴 빠르게 찾기

바로가기

GO

식당메뉴

오늘은 어디서 뭘 먹을까?

인천대학교의 학생식당 및 교직원식당 등의 주간 메뉴입니다.

학생1식당

생활원식당

교직원식당

카페테리아식당

사범대식당

일간

주간

29(월)

1코너	햄순두부찌개 / 고추장삼겹살볶음 / 훈제오리칠판구이(2인)	
2코너	등심돈까스&해쉬포테이토 / 떡갈비난자완스덮밥	
3코너	냉모밀&닭갈비덮밥	
4코너	삼겹살스테이크, 치즈함박스테이크, 고르곤졸라피자, 치즈오븐스파게티(토마토), 짜파스타, 도리아(불닭, 불고기), 불고기샐러드, 닭가슴살샐러드, 훈제연어샐러드, 옛날통닭	
5코너	신라면, 신라면정식, 만두라면, 줄면, 해물짬뽕라면, 폭탄라면, 떡라면, 치즈라면, 크림치즈김밥, 참치김밥, 즉석떡볶이, 짜파게티, 냄비우동	

30(화)

1코너	닭갈비볶음 / 설렁탕&소면 / 돈까스김치나베(2인)	
-----	------------------------------	--

Memory

Console

Elements

Sources

Network

Performance

1

X

</td><td>칼로라</td></tr> -->

<thead>...</thead>

<tbody>

<tr>

<th scope="row">1코너</th>

<td>햄순두부찌개 / 고추장삼겹살볶음 / 훈제오리칠판구이(2인)</td> == \$0

<td></td>

<td class="bdlrNone"></td>

</tr>

<tr>

<th scope="row">2코너</th>

<td>등심돈까스&해쉬포테이토 / 떡갈비난자완스덮밥</td>

<td></td>

<td class="bdlrNone"></td>

</tr>

<tr>

<th scope="row">3코너</th>

<td>냉모밀&닭갈비덮밥</td>

<td></td>

<td class="bdlrNone"></td>

</tr>

<tr>

<th scope="row">4코너</th>

<td>삼겹살스테이크, 치즈함박스테이크, 고르곤졸라피자, 치즈오븐스파게티(토마토), 짜파스타, 도리아(불닭, 불고기), 불고기샐러드, 닭가슴살샐러드, 훈제연어샐러드, 옛날통닭</td>

<td></td>

<td class="bdlrNone"></td>

... div #wrap_sub #Wrap_ContentBox #subContainer div #contents div div table tbody tr td

Styles

Event Listeners

DOM Breakpoints

Properties

Filter

:hov .cls +

element.style {

}

.tbTypeA tbody td {

border: 1px solid #d3d3d3;

padding: 7px 10px;

}

.tblList tbody td {

text-align: center;

border-bottom: 1px solid #eaeaea;

padding: 10px;

}

input, textarea, select, td, th {

font: 12px/150% '나눔고딕', 'Nanum Gothic', '돋움', Dotum, sans-

border

1

padding

7

10

169 x 36

10

7

1

Filter

Show all

border-bottom-col...

rgb(2...

빅데이터 기술

데이터 수집(데이터 수집)

```
node ㉸1 ㉸ bash ㉸2
kimjongmin:~/work/test $node test.js
29(월)
1코너 -> 햄순두부찌개 / 고추장삼겹살볶음 / 훈제오리철판구이 (2인)
2코너 -> 등심돈까스&해쉬포테이토 / 떡갈비난자완스덮밥
3코너 -> 냉모밀&닭갈비덮밥
4코너 -> 삼겹살스테이크, 치즈함박스테이크, 고르곤졸라피자, 치즈오븐스파게티 (토마토), 짜네파스타, 도리아 (
불닭, 불고기), 불고기샐러드, 닭가슴살샐러드, 훈제연어샐러드, 옛날통닭
5코너 -> 신라면, 신라면정식, 만두라면, 쫄면, 해물짬뽕라면, 폭탄라면, 떡라면, 치즈라면, 크림치즈김밥, 참치김밥
, 즉석떡볶이, 짜파게티, 냄비우동
30(화)
1코너 -> 닭갈비볶음 / 설렁탕&소면 / 돈까스김치나베 (2인)
2코너 -> 치킨치즈까스&쫄면 / 순두부비빔밥
3코너 -> 콘마요스팸김치덮밥
4코너 -> 삼겹살스테이크, 치즈함박스테이크, 고르곤졸라피자, 치즈오븐스파게티 (토마토), 짜네파스타, 도리아 (
불닭, 불고기), 불고기샐러드, 닭가슴살샐러드, 훈제연어샐러드, 옛날통닭
5코너 -> 신라면, 신라면정식, 만두라면, 쫄면, 해물짬뽕라면, 폭탄라면, 떡라면, 치즈라면, 크림치즈김밥, 참치김밥
, 즉석떡볶이, 짜파게티, 냄비우동
31(수)
1코너 -> 감자수제비 / 청양통돼지불백 / 삼겹살철판구이 (2인)
2코너 -> 카레함박스테이크&춘권 / 훈제연어회덮밥
3코너 -> 새우튀김우동&주먹밥
4코너 -> 삼겹살스테이크, 치즈함박스테이크, 고르곤졸라피자, 치즈오븐스파게티 (토마토), 짜네파스타, 도리아 (
불닭, 불고기), 불고기샐러드, 닭가슴살샐러드, 훈제연어샐러드, 옛날통닭
5코너 -> 신라면, 신라면정식, 만두라면, 쫄면, 해물짬뽕라면, 폭탄라면, 떡라면, 치즈라면, 크림치즈김밥, 참치김밥
, 즉석떡볶이, 짜파게티, 냄비우동
01(목)
1코너 -> 제육김치볶음 / 차돌박이된장찌개 / 순대철판볶음 (2인)
2코너 -> 불닭돈까스&감자튀김 / 잡채덮밥&짜장소스
3코너 -> 김치비빔국수&꼬마김밥
4코너 -> 삼겹살스테이크, 치즈함박스테이크, 고르곤졸라피자, 치즈오븐스파게티 (토마토), 짜네파스타, 도리아 (
불닭, 불고기), 불고기샐러드, 닭가슴살샐러드, 훈제연어샐러드, 옛날통닭
5코너 -> 신라면, 신라면정식, 만두라면, 쫄면, 해물짬뽕라면, 폭탄라면, 떡라면, 치즈라면, 크림치즈김밥, 참치김밥
, 즉석떡볶이, 짜파게티, 냄비우동
02(금)
1코너 -> 소고기미역국 / 스펀गी이정식 / 우거지갈비탕 (2인)
2코너 -> 치즈돈까스&맛탕 / 강된장보리비빔밥
3코너 -> 나시고랭&장각구이
4코너 -> 삼겹살스테이크, 치즈함박스테이크, 고르곤졸라피자, 치즈오븐스파게티 (토마토), 짜네파스타, 도리아 (
불닭, 불고기), 불고기샐러드, 닭가슴살샐러드, 훈제연어샐러드, 옛날통닭
5코너 -> 신라면, 신라면정식, 만두라면, 쫄면, 해물짬뽕라면, 폭탄라면, 떡라면, 치즈라면, 크림치즈김밥, 참치김밥
, 즉석떡볶이, 짜파게티, 냄비우동
03(토)
1코너 -> 오늘은 쉽니다
2코너 -> 오늘은 쉽니다
3코너 -> 오늘은 쉽니다
```

<https://jongmin92.github.io/2017/05/26/Emily/4-crawling/>

빅데이터 기술

데이터 수집(WEB 데이터 수집)

웹서버의 구조와 크롤러 / 스크래핑에 대한 예제

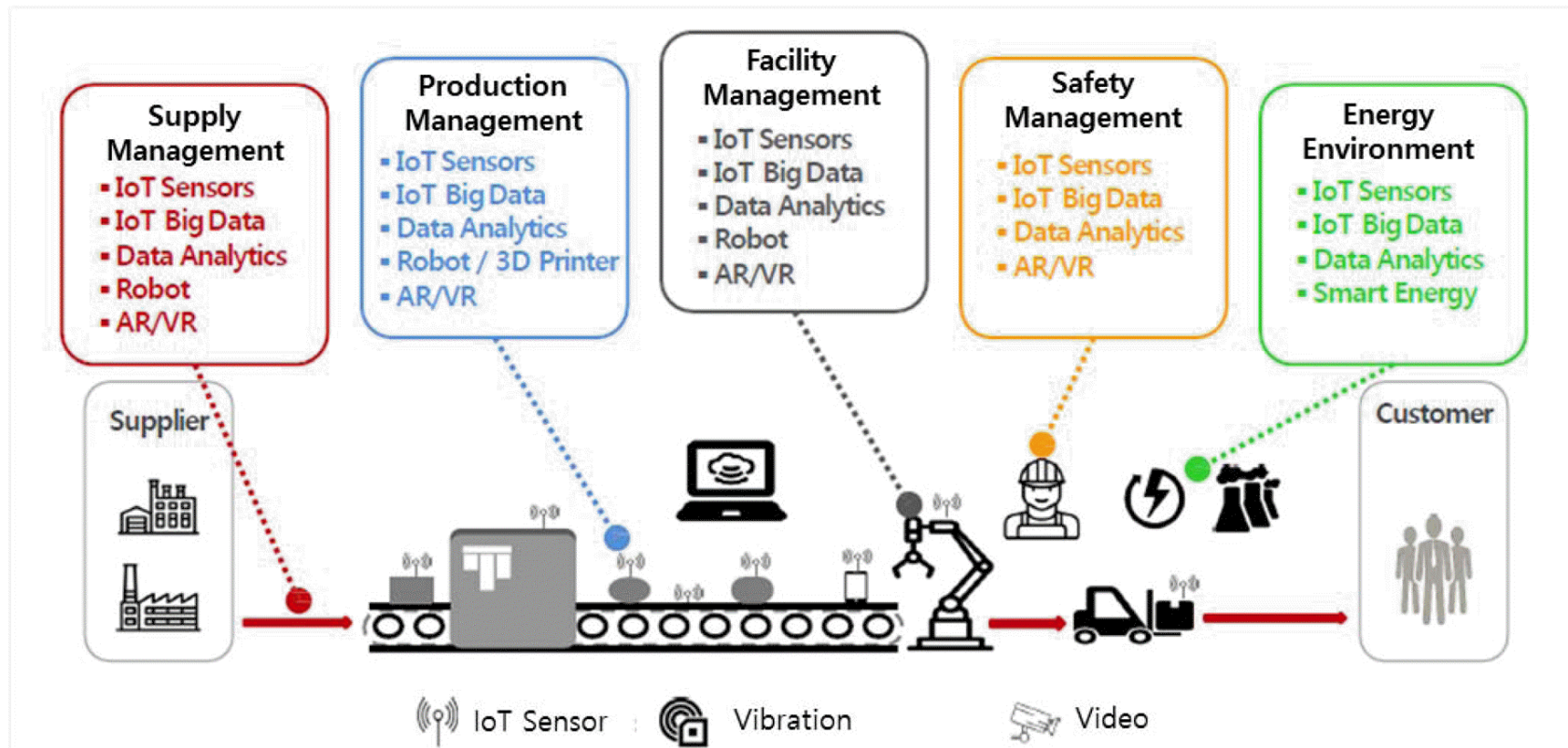
빅데이터 기술

IOT 센서에서의 데이터 수집

당신이 오늘 등교/출근 하면서 수집된 데이터는...?

빅데이터 기술

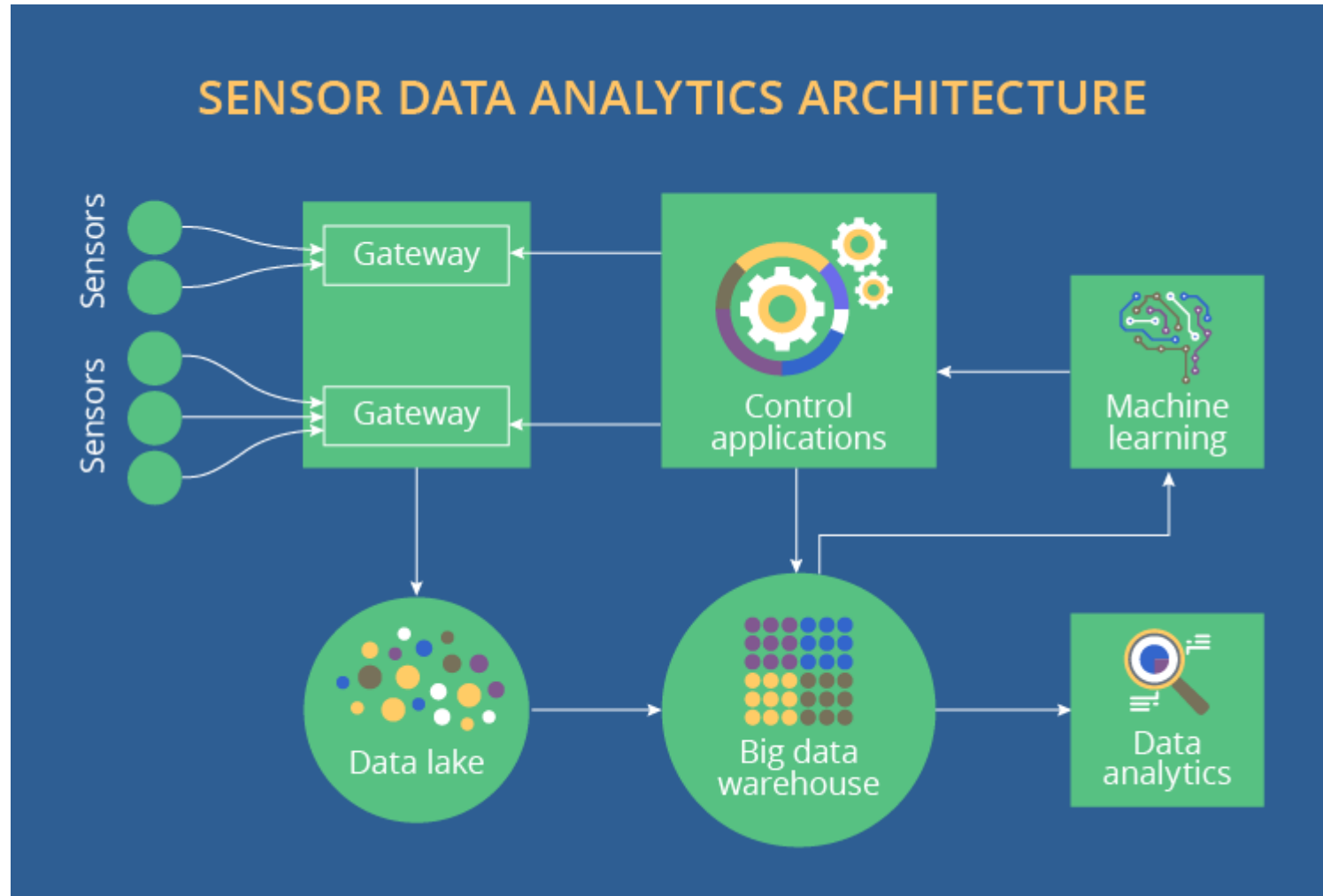
IOT 센서에서의 데이터 수집



※(cited from Korean Standards Association lectures by Min Hwa, Lee of Korea Creative Economy Research Network)





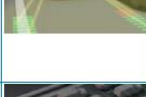




빅데이터 기술

데이터와 정보



빅데이터 기술

[표 1] 차량 제어를 위한 주요 응용기술의 예

구분	기술명	내용	예시	기반기술
ADAS (운전자 보조)	운전자 졸음 경고 (Driver Drowsiness Alert) / 운전자 상태 감시(Driver Status Monitoring)	운전자의 얼굴 모니터 링 및 피로도, 삼박수, 음주여부 등 현재의 상태 감시		영상인식, HM
	사각지대 감지 (BSM: Blind Spot Monitoring)	차량의 사각지대의 물체 감지		근거리 RADAR, 초음파센서
	주차 보조 장치 (IPAS: Intelligent Parking Assis System)	주차 공간 인식, 자동주차		초음파센서, 근거리 RADAR
	적응형 상향등 제어 (AHBC: Adaptive High Beam Control)	야간 주행 시 상황에 따른 자동 전조등 제어		카메라 영상인식, RADAR
	야간 시각 (NV: Night Vision)	야간 물체 인식		적외선 카메라
	차선이탈 경보장치 (LDW: Lane Departure Warning)	주행 중 차선이탈 시 운전자에게 경고 또는 조향제어		카메라 영상인식
운전자 보조	차로유지 지원장치 (LKAS: Lane Keeping Assist System)			
	차간거리 유지장치 (ACC: Adaptive Cruise Control)	차간거리 및 정속주행 유지		LiDAR, RADAR, 스테레오 카메라
	자동제동장치 (AEB: Autonomous Emergency Braking)	전방의 차량 및 장애물에 대한 인식, 사고발생 예측 및 제동 제어		LiDAR, RADAR, 초음파센서
	교통신호 인식 (TSR: Traffic Sign Recognition)	신호등 인식 및 표지판 인식		카메라 영상인식
	전후방 모니터링 (Front and Rear Vehicle Monitoring)	자동차 전방 및 후방의 상황감지		카메라 영상인식, 근거리 RADAR, 초음파센서
운전자 보조	보행자 감지 (PD: Pedestrian Detection)	보행자 감지		스테레오 카메라, 근거리 RADAR, LiDAR, 초음파센서, 적외선 카메라
	충돌방지(회피) 시스템 (Collision Avoidance System) / 교차로 충돌 회피 시스템(Collision Avoidance System)	차량의 충돌을 예측하고 제동 또는 회피하는 시스템		근거리 RADAR, 카메라 영상인식, LiDAR

※ 출처 : Autoweb.com, Continental, Euro NCAP, Scoda, Volvo, Mazda, GM, Toyota, Bosch

빅데이터 기술

[그림 2] 자율주행을 위한 자동차의 다양한 센서 예시



※ 괄호안은 탐지 거리(m) 및 범위(각도)

※ 출처 : Mercedes-benz (재편집)

빅데이터 기술

눈 셋 달린 자율주행차, 눈 하나는 사라진다?

https://biz.chosun.com/site/data/html_dir/2020/01/13/2020011303383.html

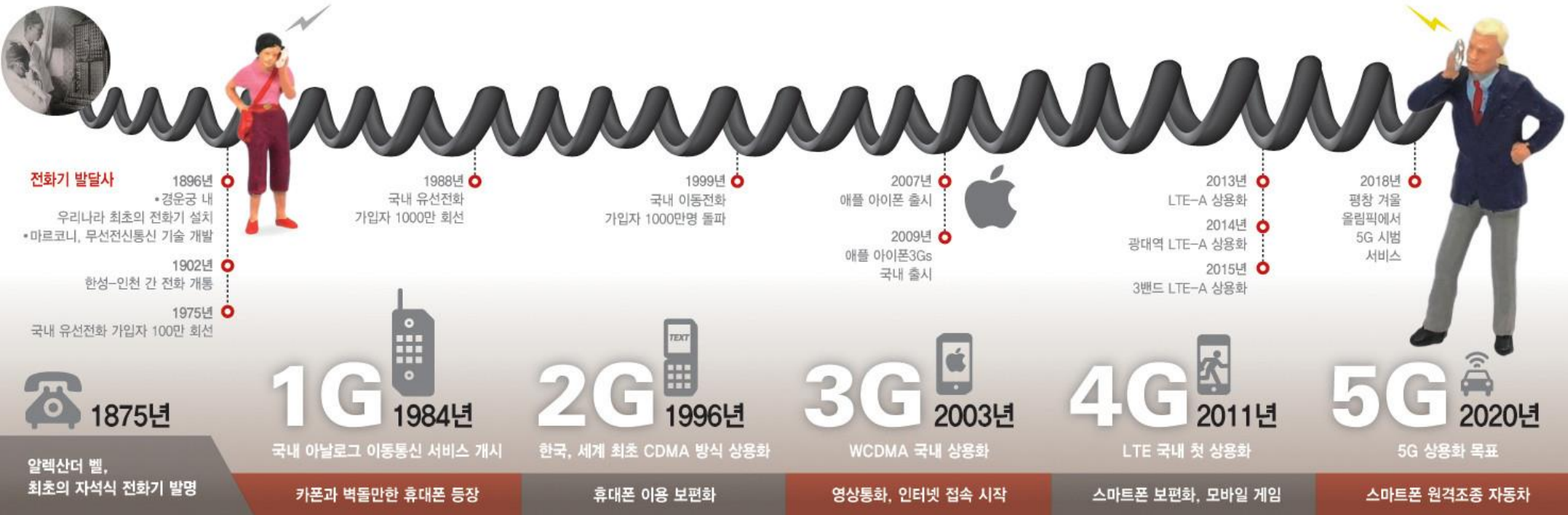
60년 만에 현실된 자율주행, 5G 기반 C-ITS로 완성될까?

https://www.e4ds.com/sub_view.asp?ch=11&t=0&idx=11017

라이다 수집 영상

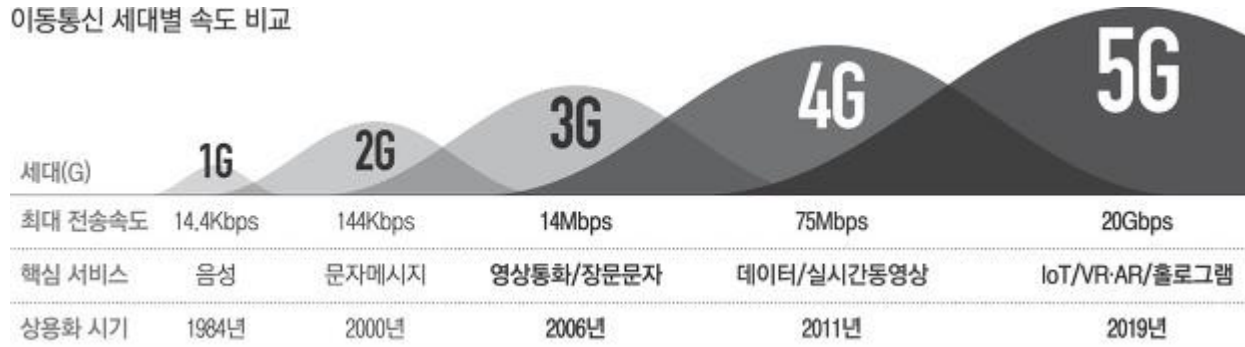
https://www.youtube.com/watch?v=yKOdrA4_-1g

빅데이터 기술



그래픽 이임정 기자 imjung@hani.co.kr

이동통신 세대별 속도 비교



빅데이터 기술

센서와 수집/네트워크/클라우드에 대한 이야기

빅데이터 기술

아직 수집 안끝남.

오픈데이터
공공데이터와 수집에 대해 알아보죠

[공공데이터포털 \(data.go.kr\)](http://data.go.kr)

빅데이터 기술

저장

가공 후 저장
저장 후 가공

빅데이터 기술

데이터베이스와 데이터베이스 관리시스템

DB (Data Base)	체계적으로 수집·축적하여 다양한 용도와 방법으로 이용할 수 있게 정리한 정보의 집합체
DBMS (Data Base Management System)	이용자가 쉽게 데이터베이스를 구축·유지할 수 있게 하는 관리 소프트웨어

빅데이터 기술

데이터베이스와 데이터베이스 관리시스템

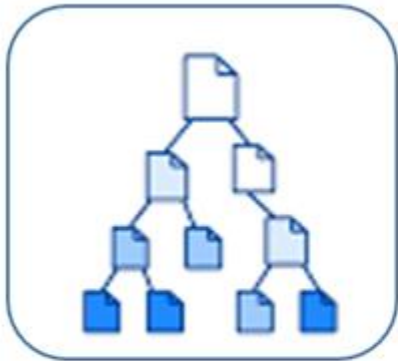
Users		
id	email	name
1	max@test.com	Maximilian Schwarzmüller
2	manu@test.com	Manuel Lorenz
3

Products			
id	title	price	description
1	Chair	120.99	A comfy chair
2	Book	10.99	Exciting book!
3

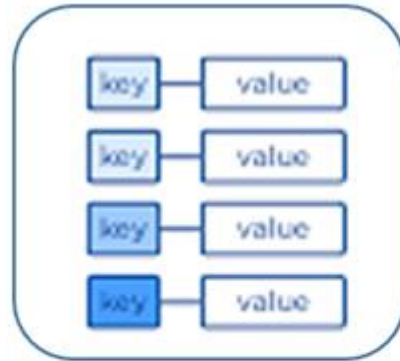
Orders		
id	user_id	product_id
1	2	1
2	1	1
3	2	2

RDBMS

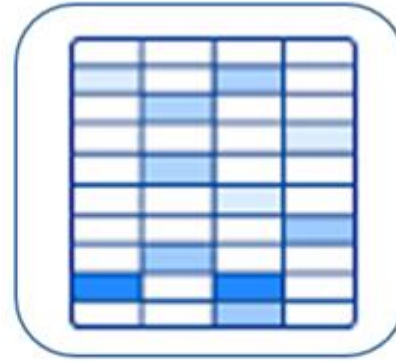
NoSQL



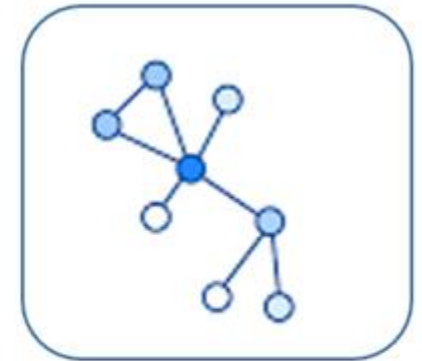
Document
Store



Key-Value
Store



Wide-Column
Store



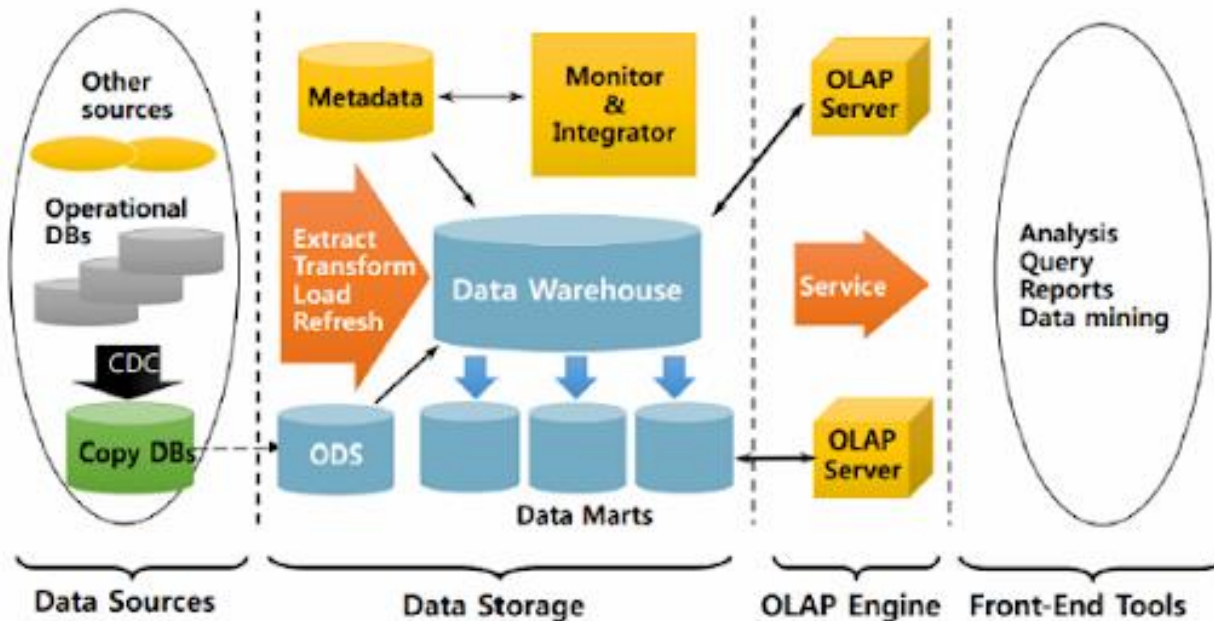
Graph
Store

빅데이터 기술

저장

데이터 웨어하우스(data warehouse)란 사용자의 의사 결정에 도움을 주기 위하여, 기간시스템의 데이터베이스에 축적된 데이터를 공통의 형식으로 변환해서 관리하는 데이터베이스를 말한다.

줄여서 DW로도 불린다



빅데이터 기술

저장

https://www.youtube.com/watch?time_continue=11&v=CMnJcZUjRyo&feature=emb_logo

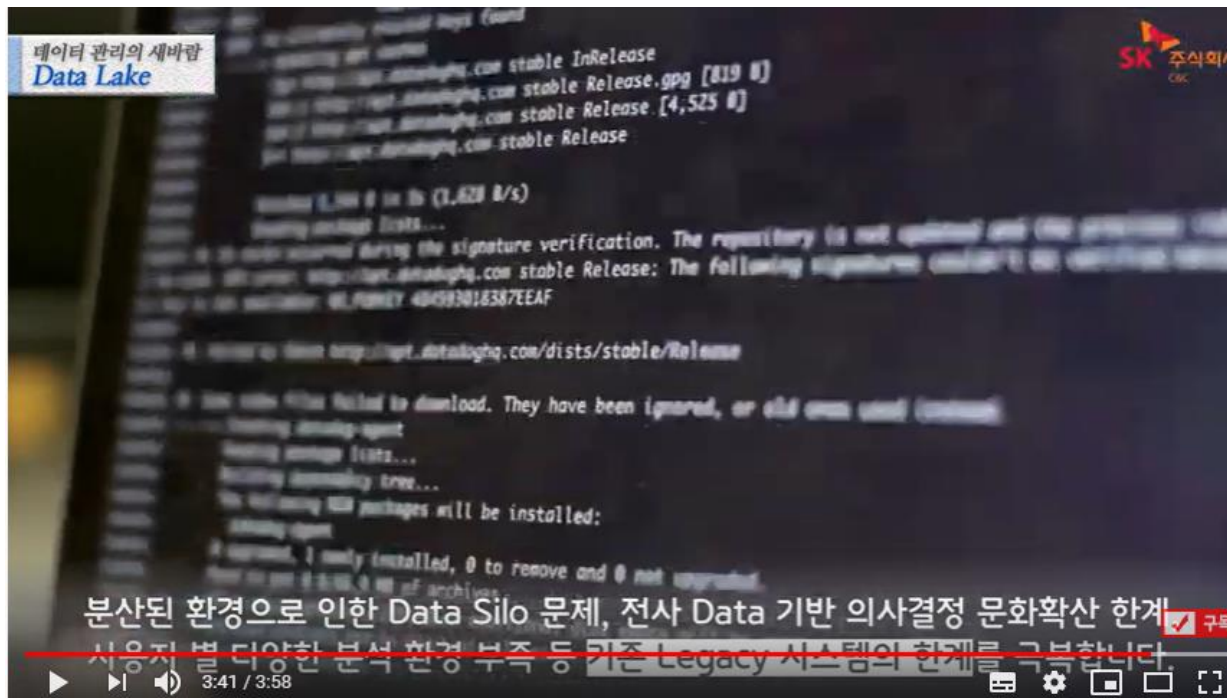
함께 꿈꾸는 더 큰 행복

함께해요 C&C

Biz & Tech

Storyteller

데이터 관리 패러다임의 새바람, 데이터 레이크(Data Lake)



빅데이터 기술

저장

데이터 레이크

https://www.youtube.com/watch?time_continue=11&v=CMnJcZUjRyo&feature=emb_logo



다음 주에 봐요