

사전 학습

[K-Digital]

지능형 영상/이미지 처리를 위한 컴퓨터비전 소프트웨어 개발 과정

강의 주제 : 빅데이터 개요

빅데이터란?

■ 빅데이터

- 디지털 환경에서 생성되는 데이터로, 그 규모가 방대하고, 생성 주기도 짧고, 형태도 수치 데이터 뿐 아니라 문자, 영상 등의 다양한 종류의 데이터를 포함하는 대규모 데이터

■ 빅데이터 기술

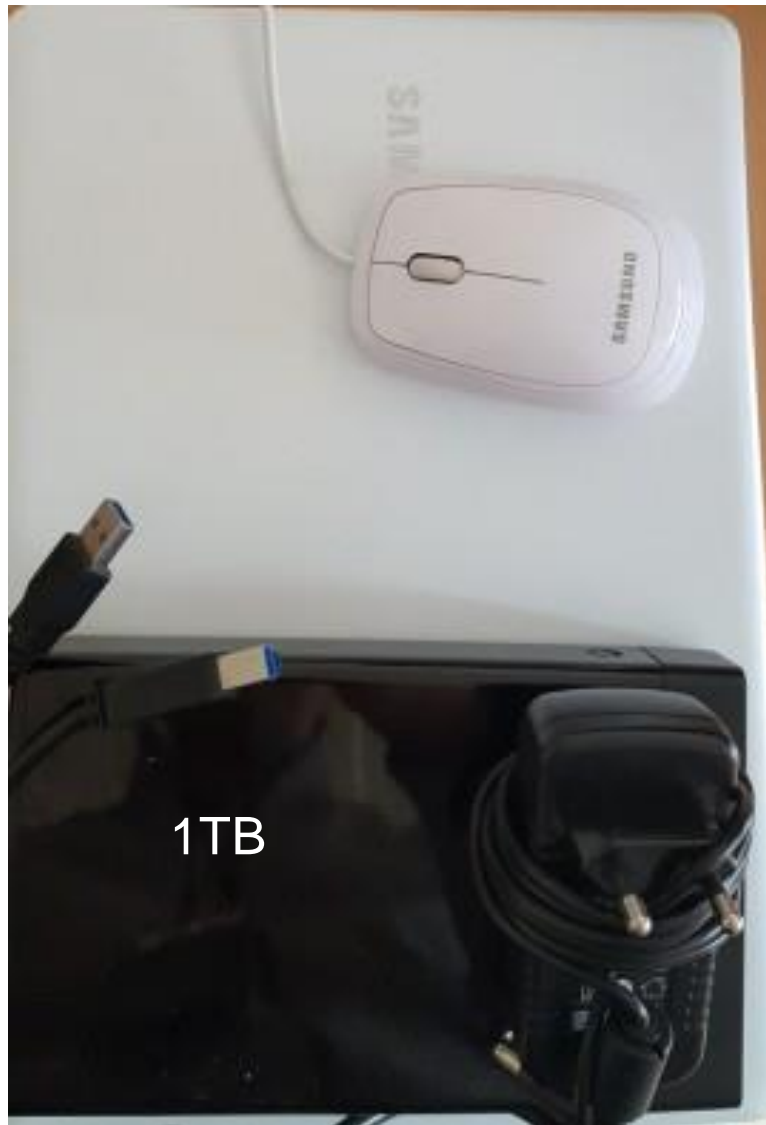
- 고객정보와 같은 정형화된 데이터 뿐 만 아니라 비정형의 대용량 데이터를 분석하여 의미 있는 정보를 추출하거나 미래를 예측하고 이를 시각화 할 수 있음

빅데이터의 시대

■ 출현 배경?

- 기업의 고객 데이터 수집 행위 증가
- 스마트 기기의 보급으로 영상/음성 데이터의 폭증
- 트위터나 페이스북 등 SNS의 급격한 확산으로 인한 데이터 폭증
- 사물인터넷 기기에 의한 센서 데이터 증가
- 데이터 저장매체의 비용 하락

빅데이터의 시대



2012

=



2019~

빅데이터의 시대

2019 *This Is What Happens In An Internet Minute*

그림 출처 : <https://www.visualcapitalist.com/what-happens-in-an-internet-minute-in-2019/>

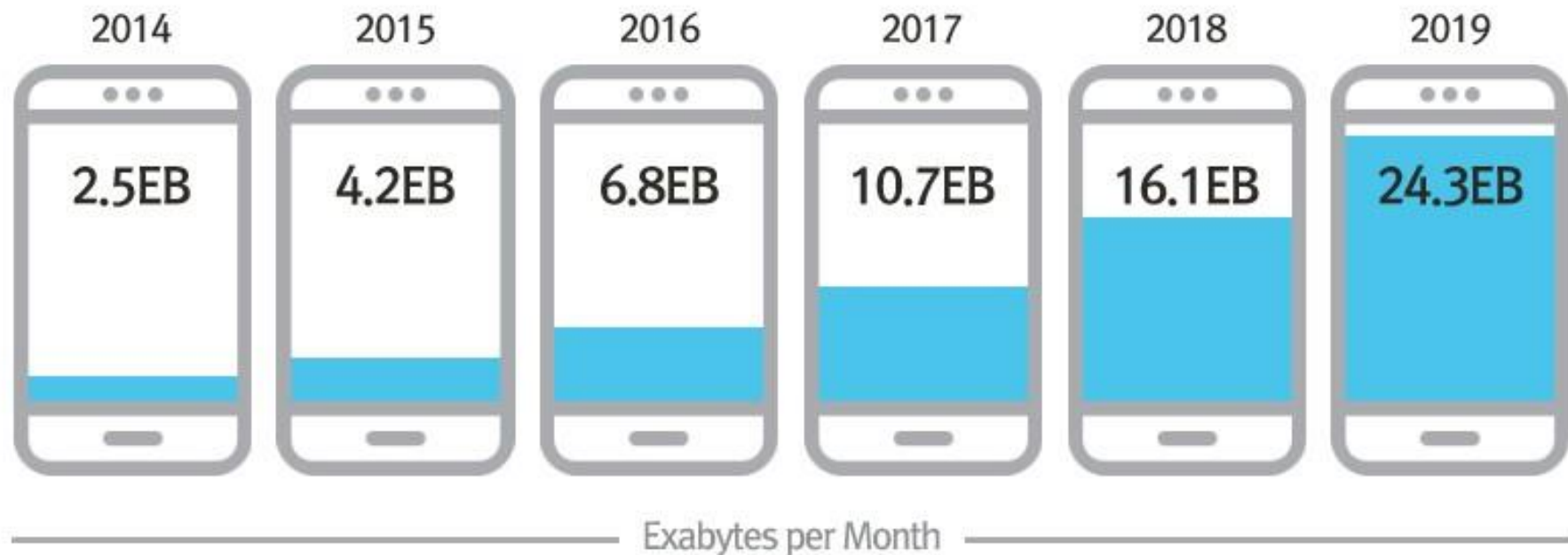


빅데이터의 시대

■ 모바일 데이터 증가 추세(월평균)

전 세계 모바일 데이터 트래픽 증가 추세

※ 출처: 시스코 비주얼 네트워킹 인덱스 보고서(2016. 01)



빅데이터의 시대

■ 현재와 앞으로의 전망

- 인류 문명이 시작된 이래 **2003년까지 축적된 데이터량은 5엑사 바이트(EB)**
 - ✓ EB는 십의 십팔 승을 뜻하는 단위로, 1기가바이트짜리 **영화 약 10억 편 분량**
- 그런데 2015년 한 달간 모바일에서만 생성된 데이터 양이 지난 수세기에 걸쳐 쌓아온 데이터 양과 비슷
- **2015년 월평균 발생한 모바일 데이터 트래픽은 4.2EB. 2020년이면 24.3EB로 증가할 것으로 예상**
 - ✓ 연간으로 환산하면 전 세계 인구가 1년간 약 81조 개의 이미지를 만드는 양

빅데이터의 시대

- 디지털 데이터의 용량 단위

1 테라바이트(TB) = 10^{12} bytes

1 페타바이트(PB) = 10^{15} bytes

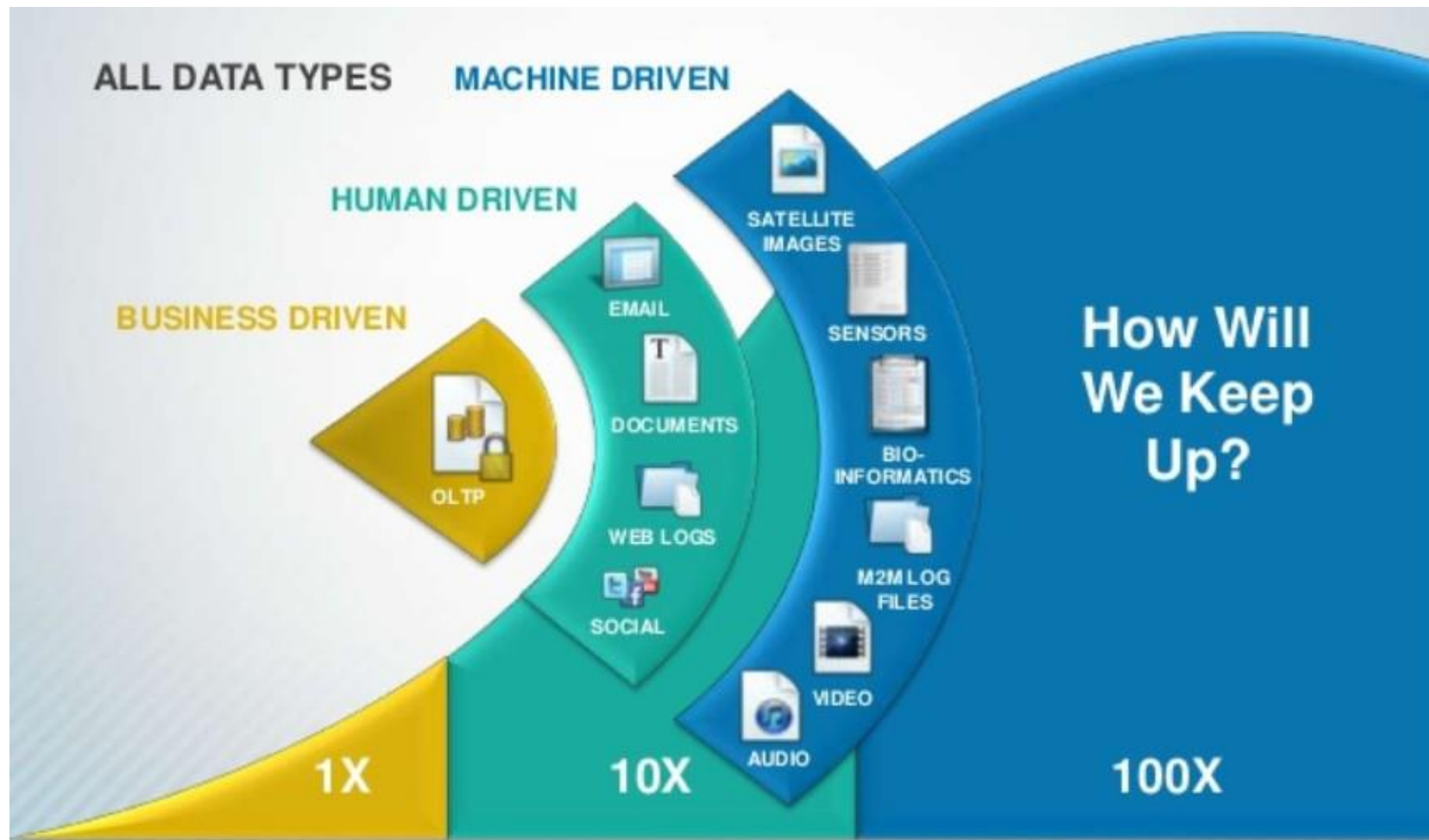
1 엑사바이트(EB) = 10^{18} bytes

1 제타바이트(ZB) = 10^{21} bytes

1 요타바이트(YB) = 10^{24} bytes

빅데이터의 시대

■ 빅데이터 생성의 주범?



빅데이터의 시대

1년 동안 미국 내 비행기 엔진에서 수집되는 데이터

Sensor data from a cross-country flight



빅데이터의 시대

■ 빅데이터 생성의 주범?

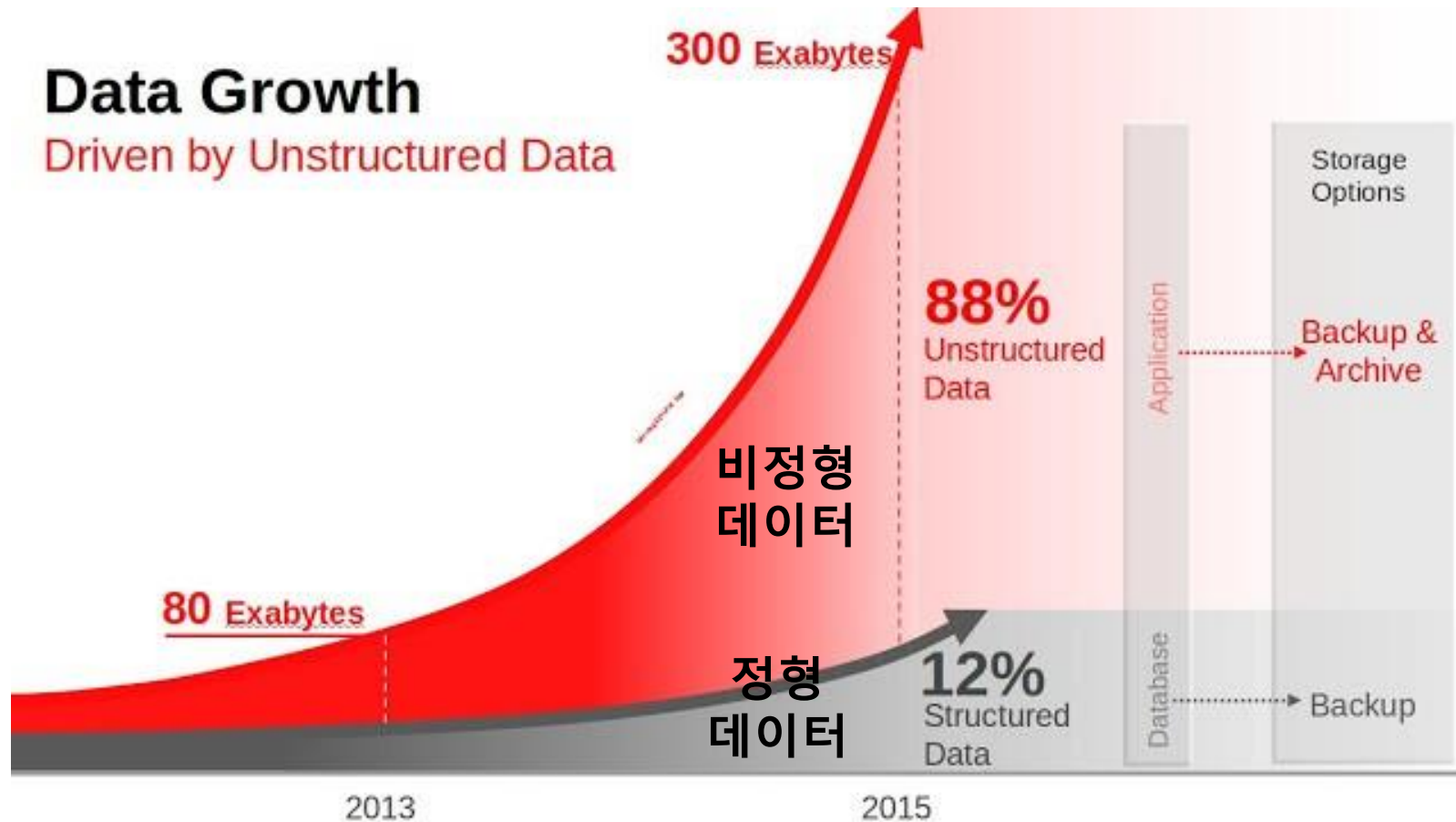


그림 출처 : <http://ddmix.blogspot.kr/2015/10/bigdata-1-intro.html>

빅데이터 특징

■ 3Vs

- **대용량** Volume : 데이터를 수집, 저장, 관리, 분석하는데 있어 기존의 데이터베이스 소프트웨어의 능력을 뛰어넘는 대용량의 데이터
- **속도** Velocity : 기존의 데이터베이스 소프트웨어의 능력을 뛰어넘는 빠른 속도로 생성되는 데이터, 소비자를 충족시키기 위한 빠른 분석, 처리 필요
- **다양성** Variety : 비정형화된 다양한 형태의 데이터
 - ✓ 정형 데이터 : 고정된 필드에 저장된 데이터
 - ✓ 비정형 데이터 : 형태와 구조가 복잡한 데이터(이미지, 오디오 등)



Volume
Data Quantity



Variety
Data Types



Velocity
Data Speed

빅데이터 형태별 분류

- **정형 데이터** | Structured Data
- **반정형 데이터** | Semi-structured Data
- **비정형 데이터** | Unstructured Data

빅데이터 형태별 분류

■ 정형 데이터

- 고정된 필드에 저장된 데이터
- 대표적인 예 : 엑셀에 항목 별로 입력된 데이터
- 예시
 - ✓ 고객 신상 데이터
 - ✓ 매출 데이터
 - ✓ 재고 데이터
 - ✓ 회계 데이터 등

선수	팀	팀연고지	포지션	등번호	생년월일	키	몸무게
최강조	일화천마	성남	MF	6	90/01/24	165	57
오춘식	대구FC	대구	MF	22	88/03/08	168	75
하리	아이파크	부산	FW	10	84/05/14	168	65
윤용구	드래곤즈	전남	MF	15	87/08/08	168	60
정도용	FC서울	서울	MF	40	86/05/28	168	68
전재호	일화천마	성남	MF	14	89/08/08	168	64
홍종하	제주유나이티드FC	제주	MF	32	88/12/21	169	74
오비나	시티즌	대전	MF	26	90/06/03	169	70
고창현	블루윙즈	수원	MF	8	93/09/15	170	64
이청용	블루윙즈	수원	MF	17	1988/07/02	180	69
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

빅데이터 형태별 분류

■ 반정형 데이터

- 고정된 필드에 저장되지는 않지만 메타데이터Metadata를 포함하는 데이터
- 대표적인 예 : HTML로 생성한 웹페이지
- 메타데이터 : 데이터에 대한 데이터
 - ✓ 데이터에 대한 추가적인 정보(예 : 저자명)를 추가하기 위한 데이터

HTML 문서에서 메타 데이터를 설명 :

```
<head>
<meta charset="UTF-8">
<meta name="description" content="Free Web tutorials">
<meta name="keywords" content="HTML,CSS,XML,JavaScript">
<meta name="author" content="Hege Refsnes">
</head>
```


빅데이터 형태별 분류

■ 비정형 데이터

- 고정된 필드에 저장되지 않고 형태와 구조가 복잡한 데이터
- 대표적인 예 : 이미지, 동영상, 음성과 같은 멀티미디어 데이터
 - ✓ 최근 CCTV, 블랙박스, 스마트폰, 드론 등으로 생성되는 이미지/동영상 데이터의 양이 급속히 증가
- 향후 10년 동안 비정형 데이터가 차지하는 비중은 약 90%에 달할 것으로 전망

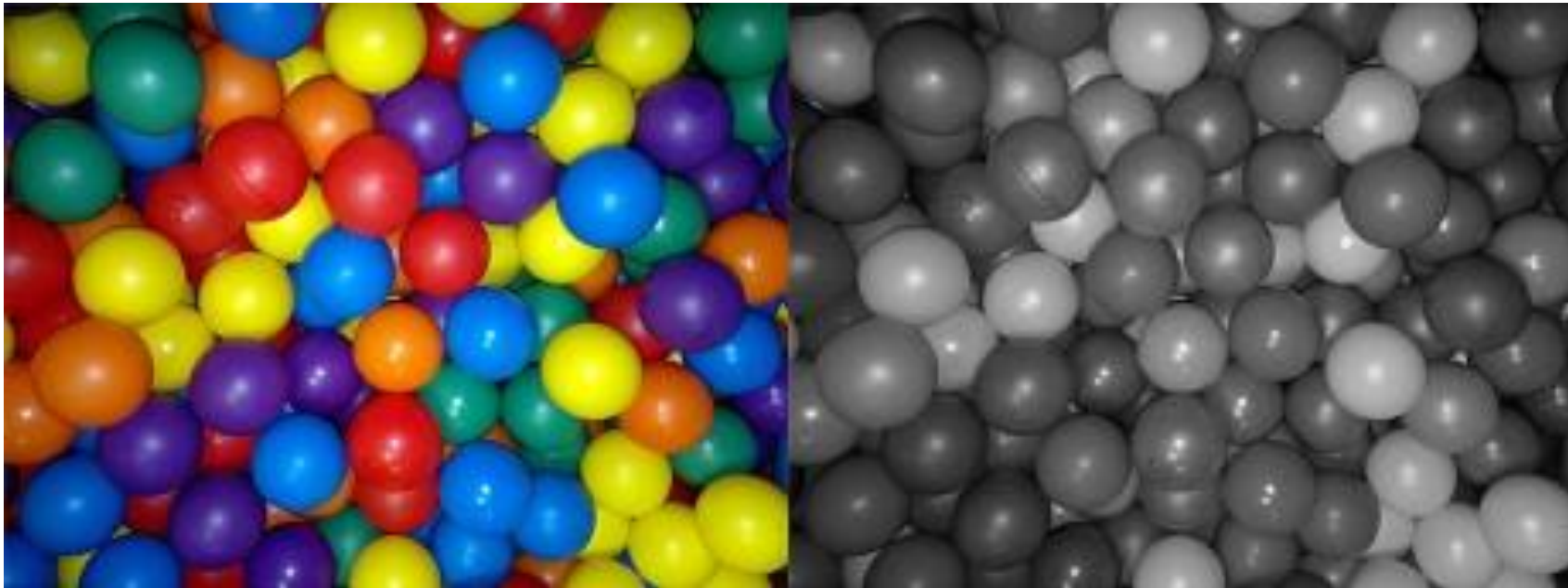
빅데이터 형태별 분류

- 이미지/동영상의 복잡성?
 - 고해상도 vs 저해상도



빅데이터 형태별 분류

- 이미지/동영상의 복잡성?
 - 컬러 vs 흑백



빅데이터 형태별 분류

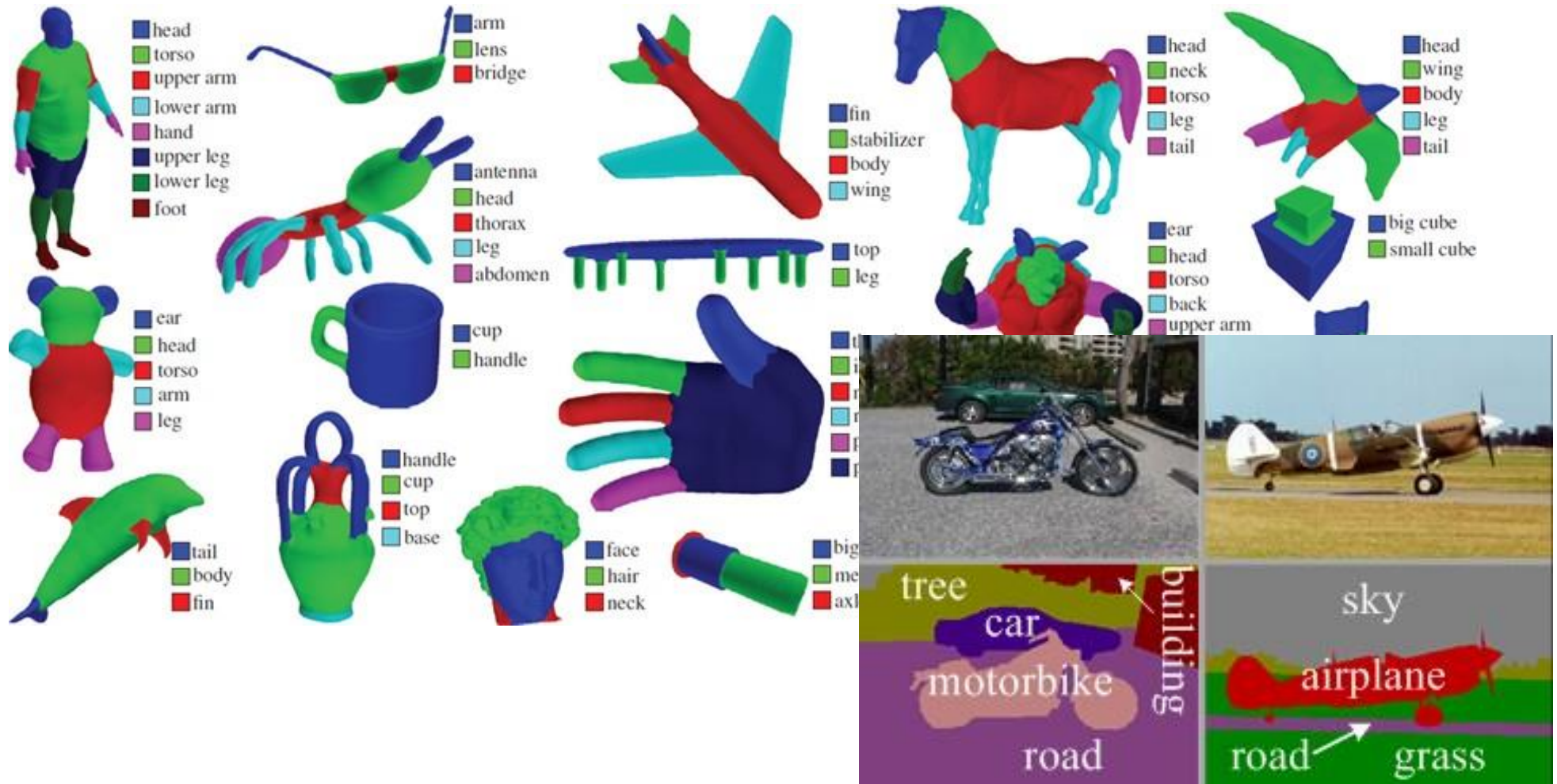
- 이미지/동영상의 복잡성?
 - 초당 프레임 수



fps : Frames per Second

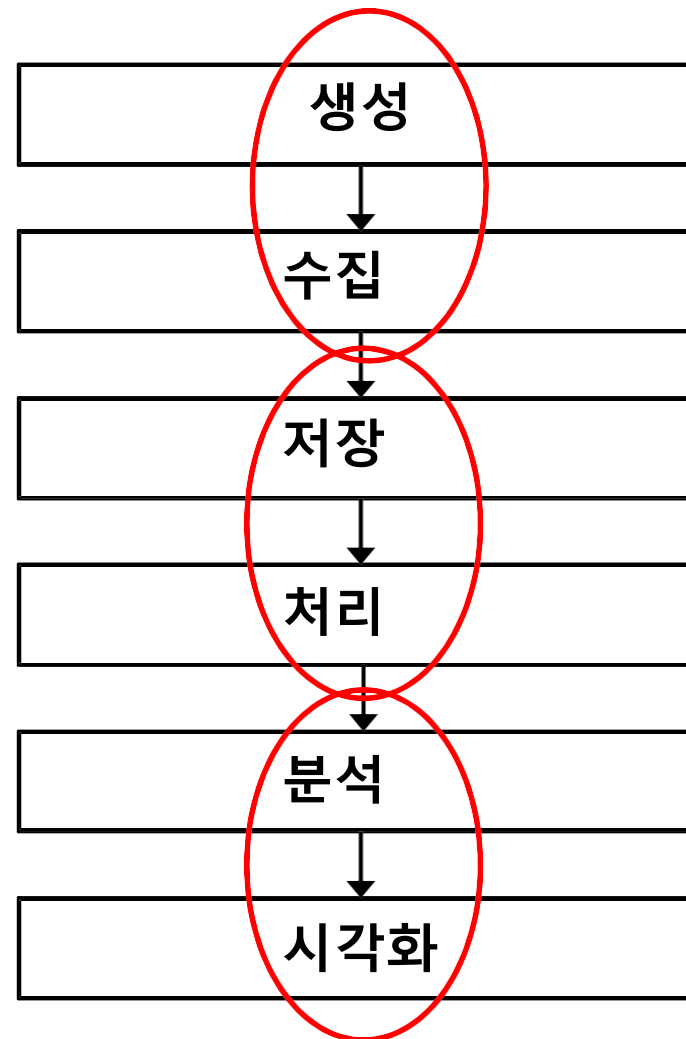
빅데이터 형태별 분류

- 이미지/동영상의 복잡성?
 - 데이터에 대한 설명(라벨) 제공의 여부



빅데이터 기술

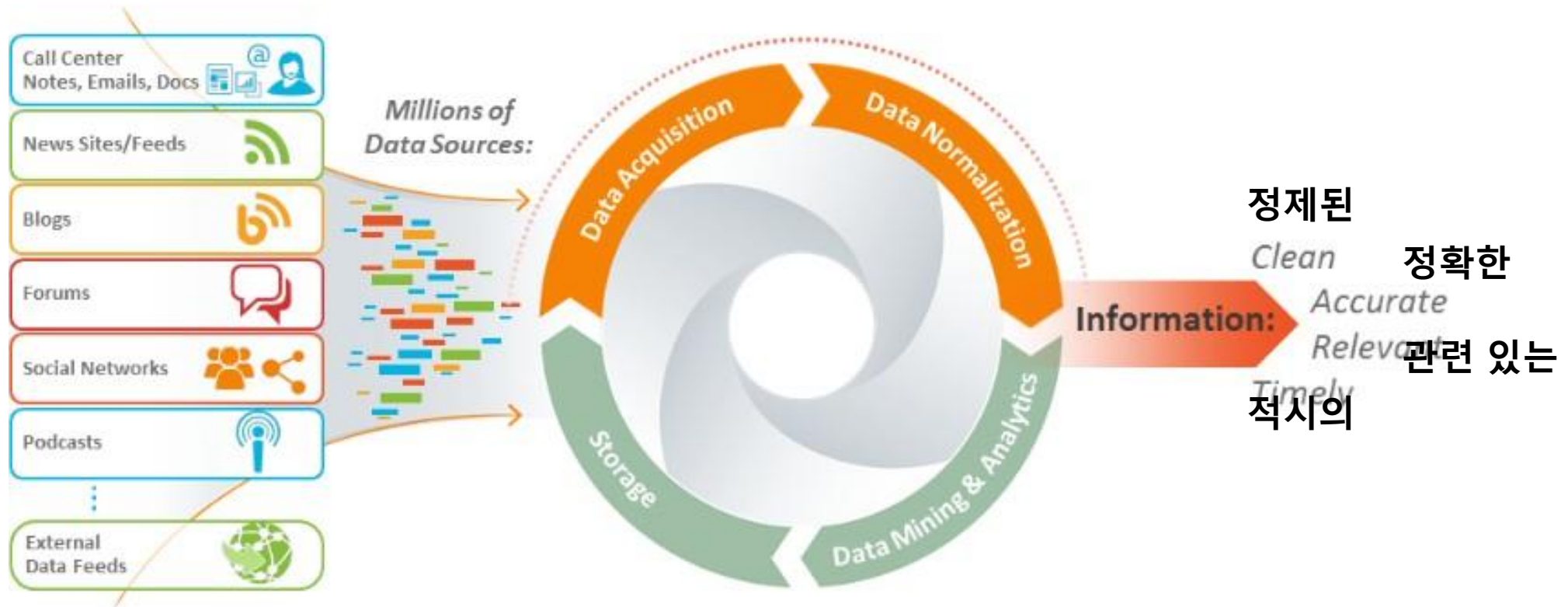
■ 빅데이터 기술 흐름도



빅데이터 기술

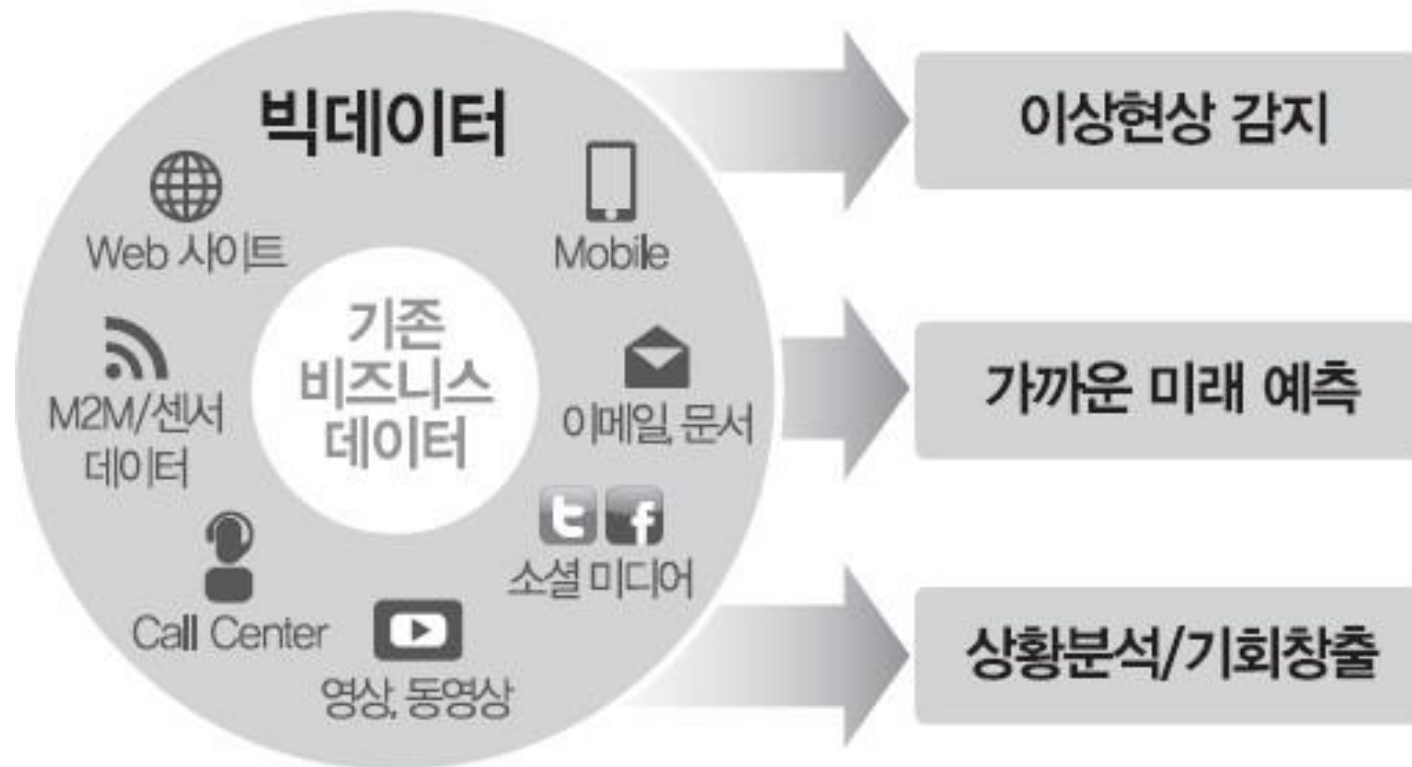
■ 빅데이터 기술의 목적

- 가공되지 않은 원본 데이터 Raw data로부터 정보 Information를 추출



빅데이터 활용

■ 활용범위와 역할



빅데이터 활용

■ 이상현상 감지

사람이 갑자기 쓰러진 상황에서 주변 사람의 행동 변화



정상시

쓰러진 사람 발생

주변 사람의 행동 변화

이상 인식 4개 유형

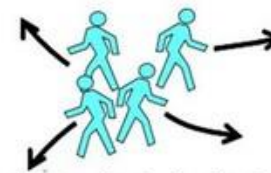
사람의 밀도나 흐름을 분석



군중변화를 감지

이상 혼잡

집단으로 도망 행동



밀도 높음

큰 흐름이 갑자기 발생

둘러싸기

집단 체류



흐름없이 고밀도

빅데이터 활용

- 이상현상 감지
 - 왜 빅데이터 인가?

기존 방식과 새로운 방식의 성능 차이

기존 방식



1명에서 소수명의 이상 행동만 감지

새로운 방식



수백명이 복잡하게 왕래하는
환경에서도 감시 가능

자료원: NEC보도 자료(<http://jpn.nec.com/press/201311/images/0702-01-01.pdf>)

■ 가까운 미래 예측 ‘인공지능’으로 15분 뒤 ‘도로 정체’ 예측한다

이동속도 4km/h 오차 범위로 가까운 미래 상황 시각화 가능

5~15분 뒤에 일어날 도로 상황을 예측해 시각적으로 보여주는 인공지능 기술이 개발됐다. “차량이 가다 서다를 반복하고 있습니다”라는 익숙한 교통안내가 “5분 뒤 시속 40km/h로 이동 가능합니다”처럼 구체적으로 바뀔 전망이다.

출처 : <http://ece.unist.ac.kr>

빅데이터 활용

- 상황분석 - CCTV 시각지대, 빅데이터로 찾는다 (동영상)



빅데이터 활용

■ 기회창출 - 빅데이터가 만든 대박상품 (동영상)



빅데이터 전문가

■ 빅데이터 전문가의 다양성

- **빅데이터 컨설턴트** : 어떤 목적으로 데이터를 수집해 어떤 결과를 도출하고 어떻게 활용할 것인지 설계
- **빅데이터 엔지니어** : 예측이나 가치 있는 정보를 도출하기 위해 시뮬레이션이 가능한 시스템 개발
- **빅데이터 분석가** : 데이터 패턴 분석 알고리즘 개발
- **빅데이터 시각화 전문가** : 데이터 추출 결과나 분석 결과를 가독성 있게 보여주는 역할

출처 :

https://www.software.kr/um/um02/um0201/um020104/um020104View.do?postId=23620&cs_sort=T

빅데이터 전문가

■ 빅데이터 전문가의 다양성

- **빅데이터 큐레이터** : 정보의 홍수 속에서 무엇을 먹을지, 볼지 결정하기 힘들어하는 현대인들에게 적합한 콘텐츠를 필터링하고 가공하여 제공
- **데이터 소거원** : 개인 사생활 침해와 관련된 정보를 삭제
- **데이터 장례사** : 고인의 디지털 흔적을 모두 삭제
- **기억 대리인** : 한 사람의 일생을 데이터베이스화해서 보관하고 의뢰인이 필요로 할 때 정보를 재생
- **스몰데이터 전문가** : 고객의 작은 행동이나 사소한 유형도 분석

출처 :

<https://www.software.kr/um/um02/um0201/um020104/um020104View.do?postId=23620&sort=T>

빅데이터 전문가

■ 빅데이터 전문가의 다양성



출처 :

https://www.software.kr/um/um02/um0201/um020104/um020104View.do?postId=23620&cs_sort=T