

# 빅데이터의 이해

3주차 : 빅데이터의 개념 및 특징  
(2차시 복습 및 요약)

성공회대학교 열림공과대학  
김종울

# 1. 지난 시간의 복습

## 빅데이터의 개념 및 특징

--- 강의 목표 ---

데이터란 무엇인가?

데이터에 대한 이해

데이터, 정보, 지식, 지혜에 대한 이해

데이터베이스에 대한 이해

빅데이터에 대한 이해

빅데이터의 특징, 모델

# 데이터의 정의

## 데이터의 정의

‘데이터(Data)’라는 단어는 라틴어인 ‘Datum’의 복수형인 ‘Data’에서 그 유래를 찾을 수 있다. ‘datum’은 선물, 주어진 것, 차변(회계용어) 등의 의미를 담고 있다.

옥스포드 대사전에서는 데이터를 추론과 추정의 근거를 이루는 사실이라고 정의하고 있다.

위키피디아에서는 문자, 숫자, 소리, 그림, 형상, 단어 등의 형태로 된 의미 단어이며, 보통 연구나 조사들의 바탕이 되는 재료를 데이터라고 정의한다.

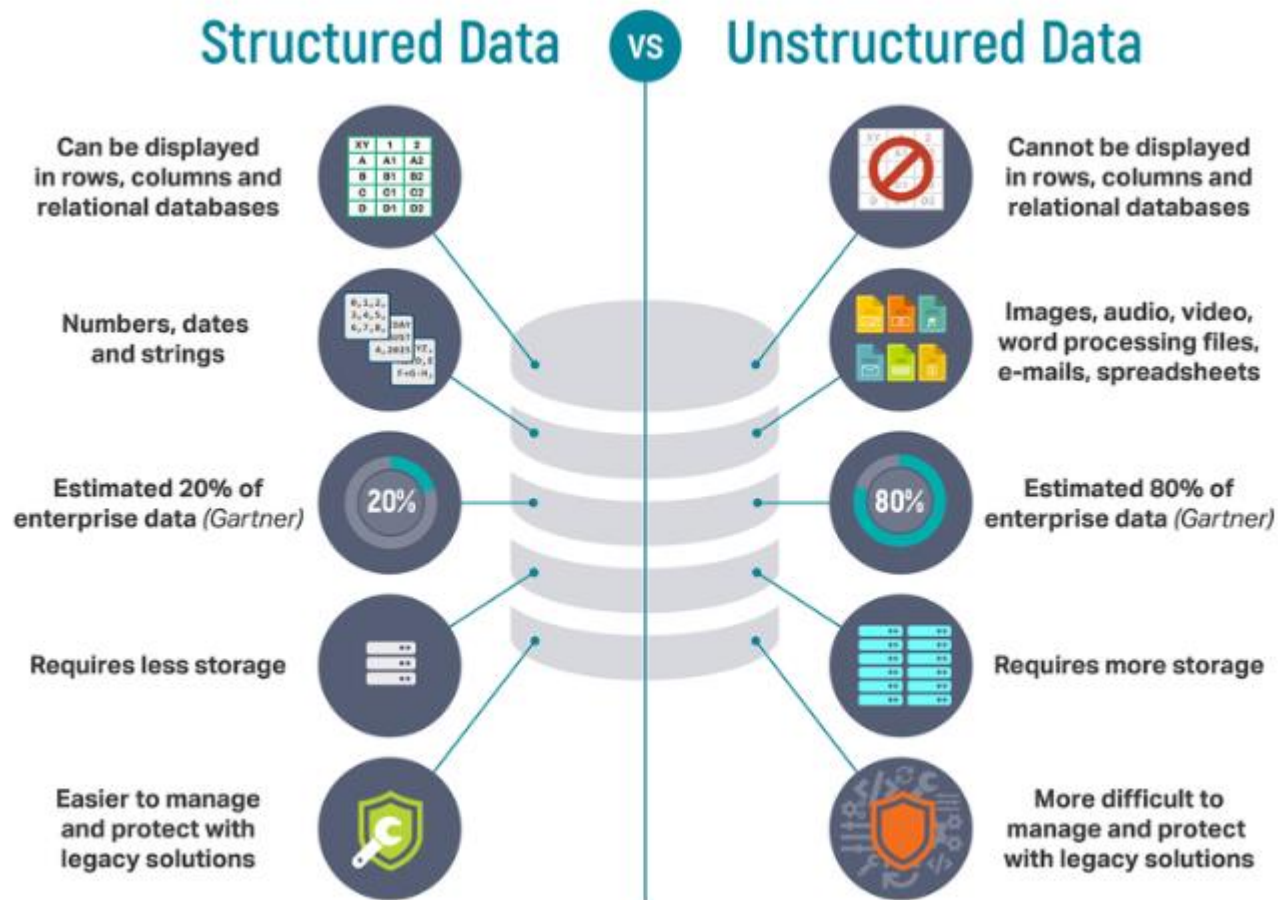
일반적으로 데이터란 이론을 세우는 데 기초가 되는 사실 또는 바탕이 되는 자료, 혹은 컴퓨터가 처리할 수 있는 문자, 숫자, 소리, 그림 따위의 형태로 된 자료 등을 의미한다.

여기서 중요한 것은 데이터는 정보가 아니라 자료라는 점이며 의미를 담고 있느냐 없느냐의 차이이다.

데이터가 의미를 담고 있으면 정보가 되지만, 의미가 없다면 그냥 자료로 남는다.

# 데이터의 정의

## 데이터의 유형



# 데이터의 정의

## 데이터의 유형

```
{
  "type": "champion",
  "format": "standAloneComplex",
  "version": "10.22.1",
  "data": {
    "Aatrox": {
      "version": "10.22.1",
      "id": "Aatrox",
      "key": "266",
      "name": "아트록스",
      "title": "다르킨의 검",
      "blurb": "한때는 공허에 맞서 싸웠던 슈리마의 명예로운 수호자 아트록스와 그의 종족은 결국 공허보다 위험한 존재가 되어 룬테라의 존속을 위협했지만, 교활한 필멸자의 마법에 속아넘어가 패배하게 되었다. 수백 년에 걸친 불인 끝에, 아트록스는 자신의 정기가 깃든 마법 무기를 휘두르는 머리색은 자들을 타락시키고 육신을 바꾸는 것으로 다시 한번 자유의 길을 찾아내었다. 이제 이전의 잔혹한 모습을 잃은 육체를 처지한 아트록스는 세상의 종말과 오랫동안 기다려온 복수를...",
      "info": {
        "attack": 8,
        "defense": 4,
        "magic": 3,
        "difficulty": 4
      },
      "image": {
        "full": "Aatrox.png",
        "sprite": "champion0.png",
        "group": "champion",
        "x": 0,
        "y": 0,
        "w": 48,
        "h": 48
      },
      "tags": [
        "Fighter",
        "Tank"
      ],
      "partype": "피의 샘",
      "stats": {
        "hp": 580,
        "hpperlevel": 90,
        "mp": 0,
        "mpperlevel": 0,
        "movespeed": 345,
        "armor": 38,
        "armorperlevel": 3.25,
```

# 데이터의 정의

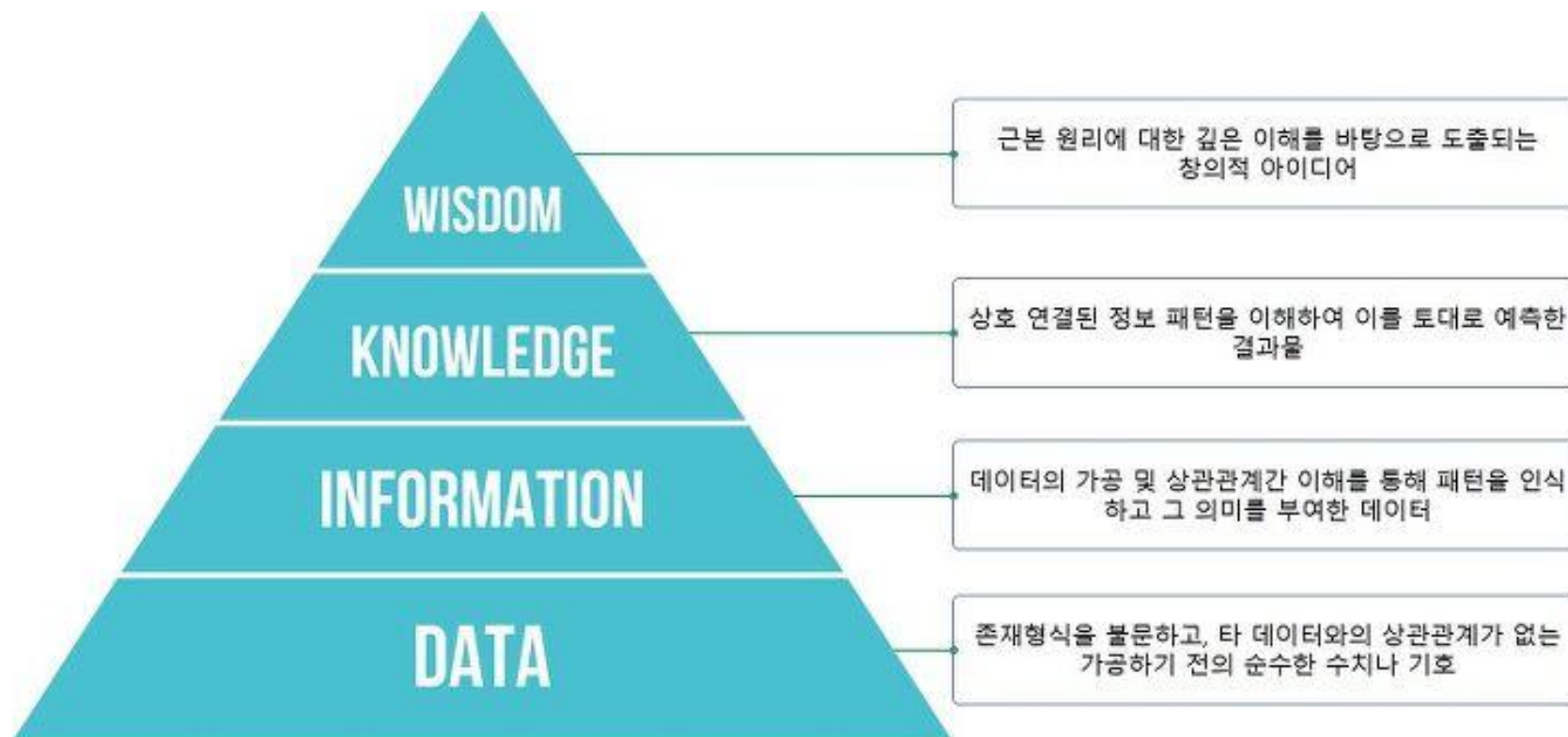
## 데이터와 정보



자료: Debra M. Amidon, 1997

# 데이터의 정의

## 데이터와 정보



# 데이터베이스

## 데이터베이스의 다양한 정의

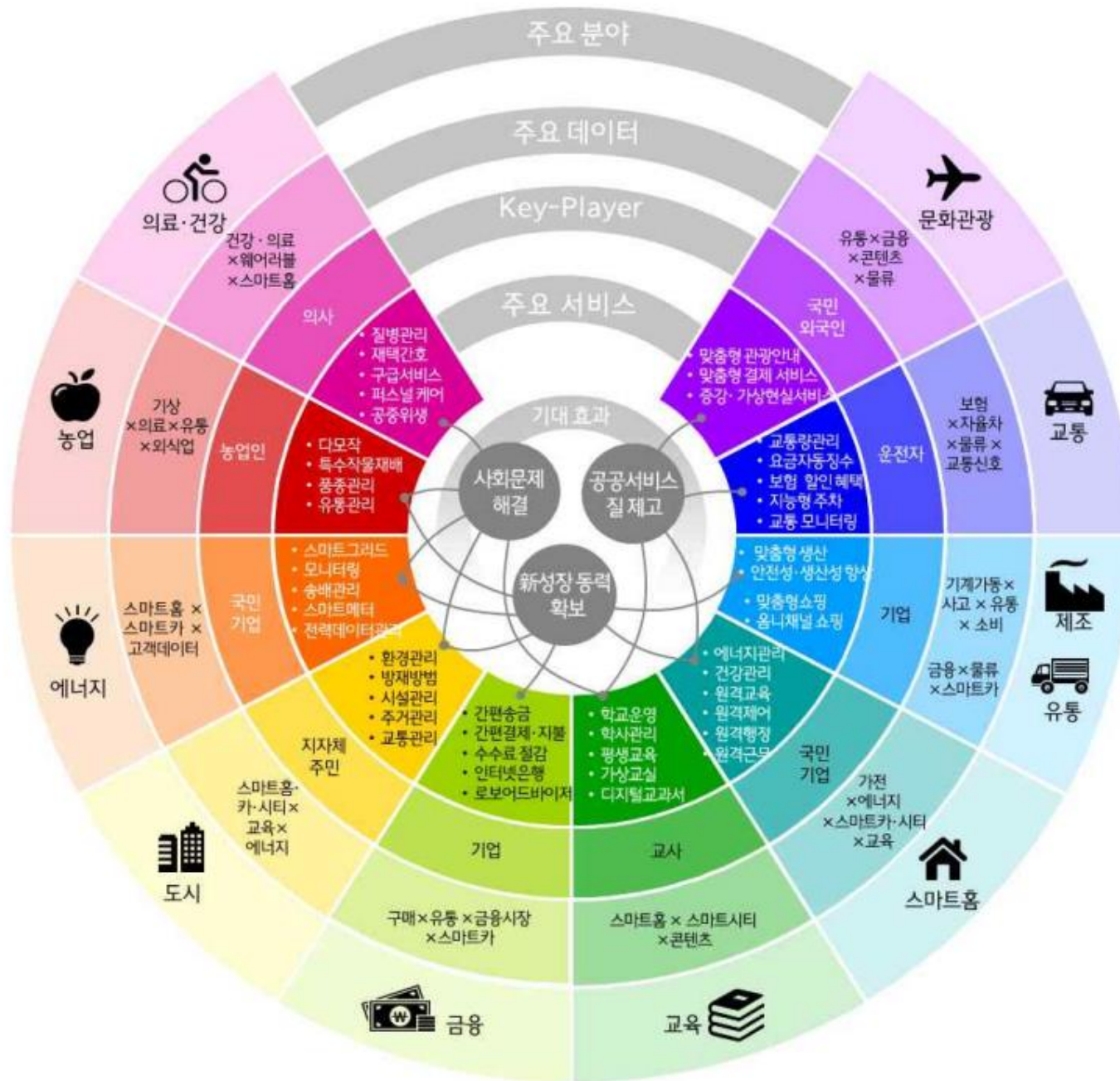
- **EU의 ‘데이터베이스의 법적 보호에 관한 지침’**: 데이터베이스란 체계적이거나 조직적으로 정리되고 전자식 또는 기타 수단으로 개별적으로 접근할 수 있는 독립된 저작물, 데이터 또는 기타 소재의 수집물이다.
- **우리나라의 ‘저작권법’**: 데이터베이스란 소재를 체계적으로 배열 또는 구성한 편집물로서 개별적으로 그 소재에 접근하거나 그 소재를 검색할 수 있게 한 것이다(법률적으로 기술 기반 저작물로 인정).
- **‘컴퓨터용어사전’, ‘정보통신용어사전(TTA)’**: 데이터베이스란 동시에 복수의 적용 업무를 지원할 수 있게 복수 이용자의 요구에 대응해서 데이터를 받아들이고 저장·공급하기 위해 일정한 구조에 따라 편성된 데이터의 집합이다.
- **‘위키피디아’**: 데이터베이스란 체계적으로 정렬된 데이터 집합을 의미한다. 데이터의 양과 이용이 늘어나면서 데이터베이스는 대용량의 데이터를 저장·관리·검색·이용할 수 있는 컴퓨터 기반의 데이터베이스로 진화했다.
- **‘한국데이터산업진흥원’**: 데이터베이스란 문자, 기호, 음성, 화상, 영상 등 상호 관련된 다수의 콘텐츠를 정보처리 및 정보통신기기에 의하여 체계적으로 수집·축적하여 다양한 용도와 방법으로 이용할 수 있게 정리한 정보의 집합체를 의미한다. 여기서 콘텐츠란 다양한 의미 전달 매체에 의해 표현된 데이터, 정보, 지식, 저작물 등의 인식 가능한 모든 자료를 의미한다.



# 데이터베이스

## 데이터베이스와 데이터베이스 관리시스템

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



자료 : 한국정보화진흥원(NIA) 미래전략센터

# 빅데이터의 이해

## 빅데이터의 정의

### 데이터의 규모에 초점을 맞춘 정의

- 기존 데이터베이스 관리도구의 데이터 수집, 저장, 관리, 분석 역량을 넘어서는 데이터 (맥킨지, 2011년)

### 업무 수행 방식에 초점을 맞춘 정의

- 다양한 종류의 대규모 데이터로부터 저렴한 비용으로 가치를 추출하고, 데이터의 빠른 수집, 발굴, 분석을 지원하도록 고안된 차세대 기술 및 아키텍처 (IDC, 2011년 6월)

## 빅 데이터

위키백과, 우리 모두의 백과사전.

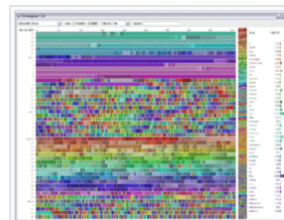
**빅 데이터**(영어: big data)란 기존 **데이터베이스** 관리도구의 능력을 넘어서는 대량(수십 **테라바이트**)의 정형 또는 심지어 데이터베이스 형태가 아닌 비정형의 데이터 집합조차 포함한<sup>[1]</sup> 데이터로부터 가치를 추출하고 결과를 분석하는 기술<sup>[2]</sup>이다.

다양한 종류의 대규모 데이터에 대한 생성, 수집, 분석, 표현을 그 특징으로 하는 빅 데이터 기술의 발전은 다변화된 현대 사회를 더욱 정확하게 예측하여 효율적으로 작동케 하고 개인화된 현대 사회 구성원마다 맞춤형 정보를 제공, 관리, 분석 가능케 하며 과거에는 불가능했던 기술을 실현시키기도 한다.

이같이 빅 데이터는 정치, 사회, 경제, 문화, 과학 기술 등 전 영역에 걸쳐서 사회와 인류에게 가치 있는 정보를 제공할 수 있는 가능성을 제시하며 그 중요성이 부각되고 있다.

하지만 빅데이터의 문제점은 바로 사생활 침해와 보안 측면에 자리하고 있다. 빅데이터는 수많은 개인들의 수많은 정보의 집합이다. 그렇기에 빅데이터를 수집, 분석할 때에 개인들의 사적인 정보까지 수집하여 관리하는 빅브라더의 모습이 될 수도 있는 것이다. 그리고 그렇게 모은 데이터가 보안 문제로 유출된다면, 이 역시 거의 모든 사람들의 정보가 유출되는 것이기에 큰 문제가 될 수 있다.

세계 경제 포럼은 2012년 떠오르는 10대 기술 중 그 첫 번째를 빅 데이터 기술로 선정<sup>[3]</sup> 했으며 대한민국 지식경제부 R&D 전략기획단은 IT 10대 핵심기술 가운데 하나로 빅 데이터를 선정<sup>[4]</sup> 하기도 했다.

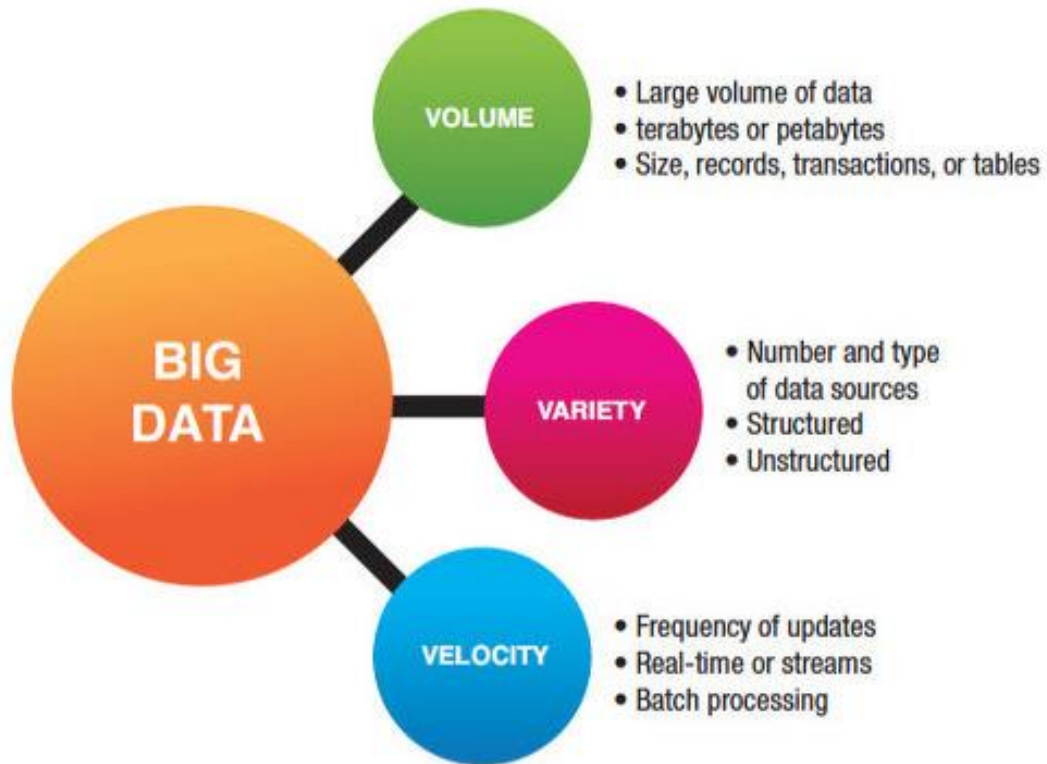


위키백과의 편집 현황의 시각화 자료(IBM 작성). 수 **테라바이트**의 용량을 지닌 위키백과의 텍스트 및 이미지 자료는 빅 데이터의 고전적 사례에 속한다.

# 빅데이터의 이해

## 빅데이터의 특징

### The Three Vs of Big Data



# 빅데이터의 이해

## 빅데이터의 특징

빅데이터의 특징은 **3V**로 요약할 수 있다.

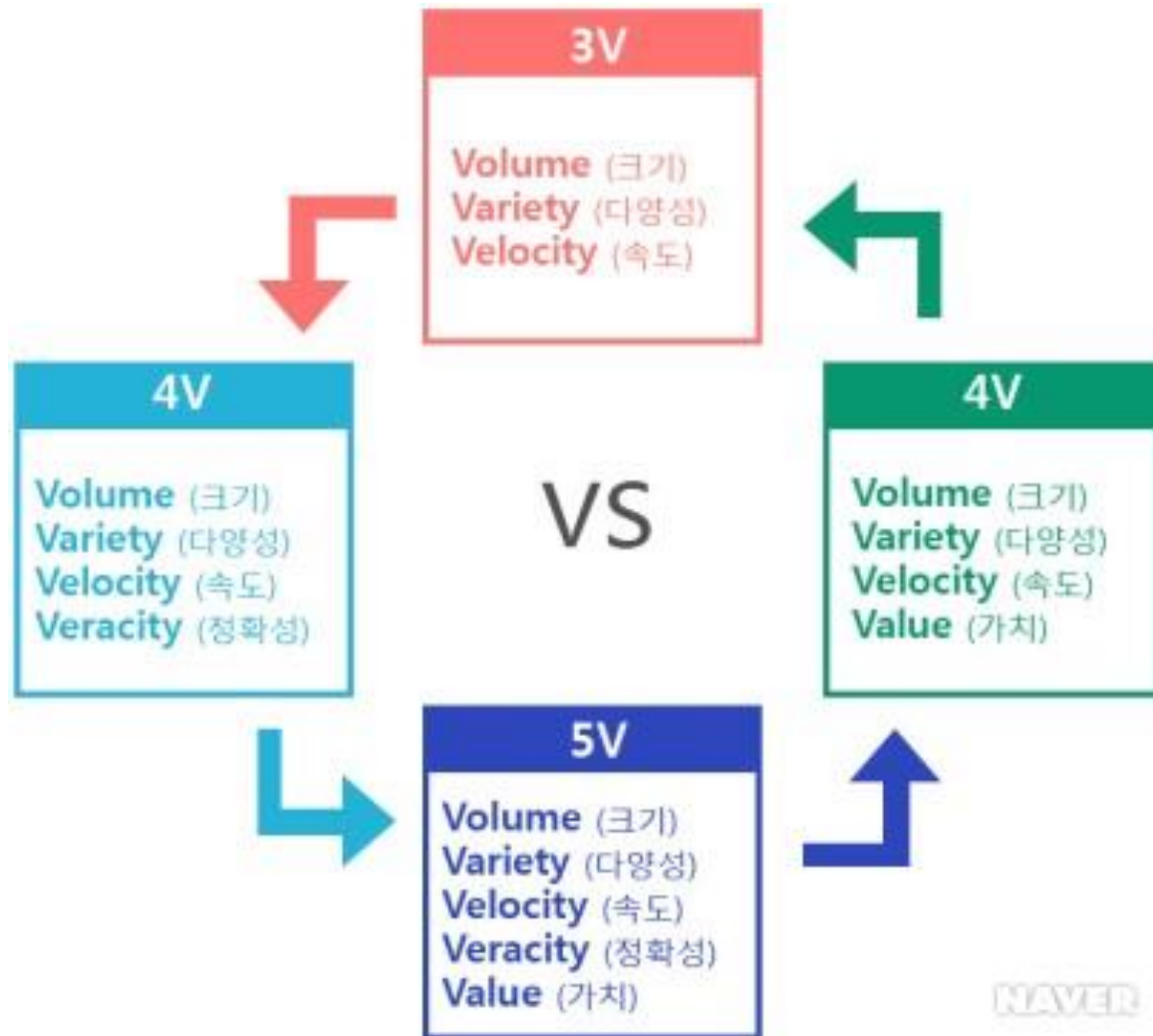
- Volume - 크기
  - 저장장치 가격의 하락, 네트워크 속도의 향상으로 수 페타바이트의 데이터가 매일 생성
- Variety - 다양성
  - 정형: DB 같이 고정형 필드에 저장되는 데이터
  - 반정형: XML, HTML 같이 메타데이터나 스키마등을 포함하는 데이터
  - 비정형: 동영상, SNS 메시지, 사진, 오디오 등 고정된 형태가 없는 데이터
- Velocity - 속도
  - 정보의 유통 속도
  - 데이터의 처리 속도

3V에 아래의 두가지를 추가해서 4V, 5V로 부르기도 한다.

- Value - 가치
  - 유의미한 가치를 가지는 지표
- Veracity - 정확성
  - 빅데이터를 이용하여 뽑아낸 데이터의 신뢰성

# 빅데이터의 이해

## 빅데이터의 특징

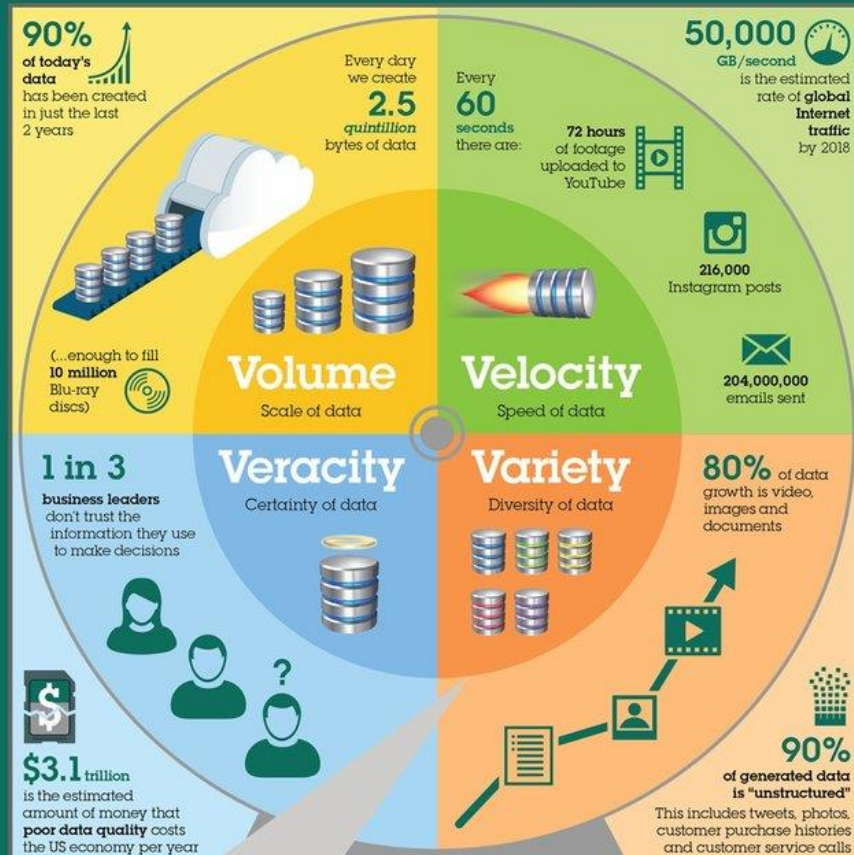




# 빅데이터의 이해

## 빅데이터의 특징

## Extracting business value from the 4 V's of big data



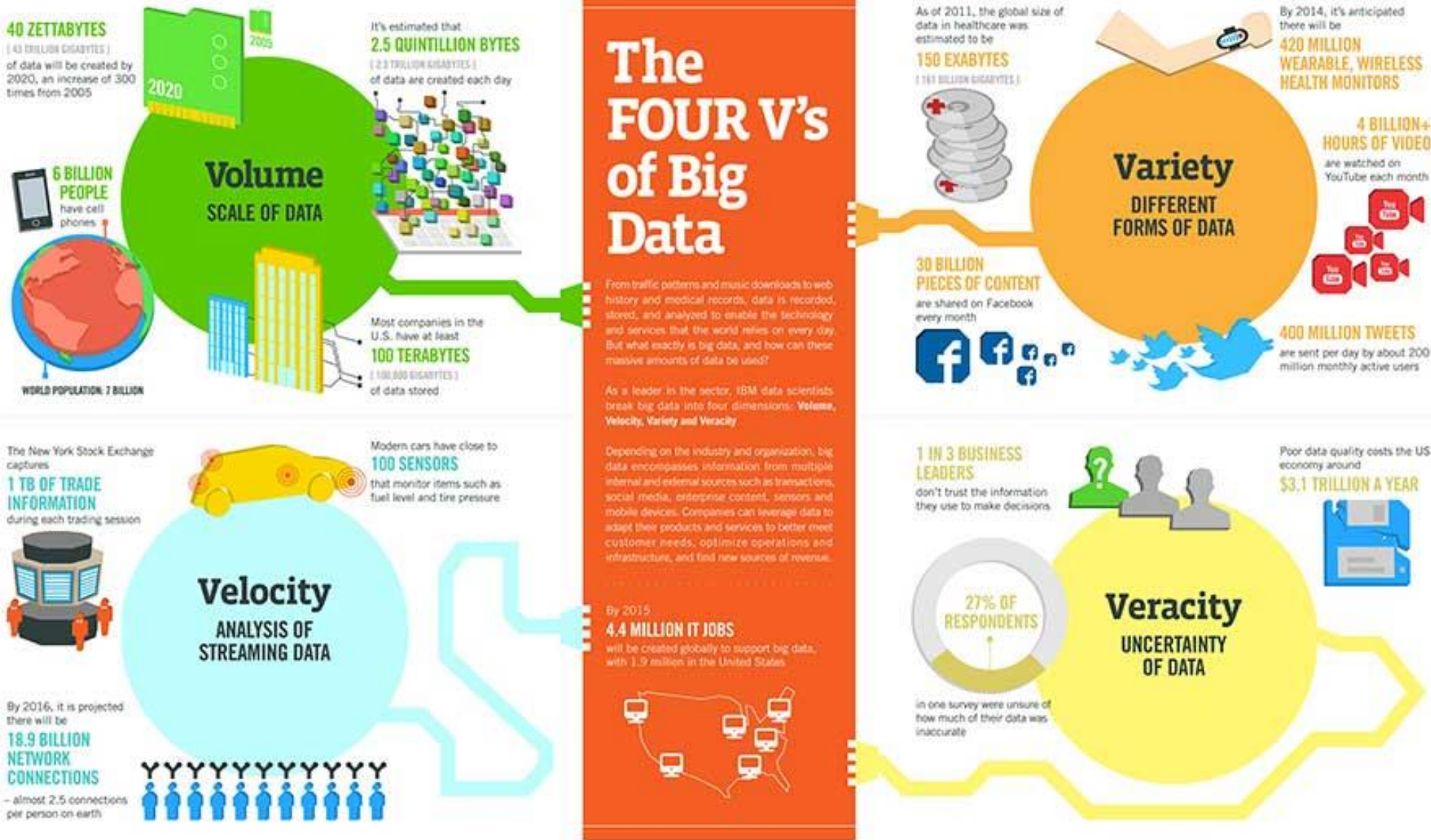
### The fifth "V"?

Big data = the ability to achieve greater **Value** through insights from superior analytics



**Case study:** A US-based aircraft engine manufacturer now uses analytics to predict engine events that lead to costly airline disruptions, with 97% accuracy. If this prediction capability had been available in the previous year, it would have saved \$63 million.

# 빅데이터의 이해





# 빅데이터의 이해

---

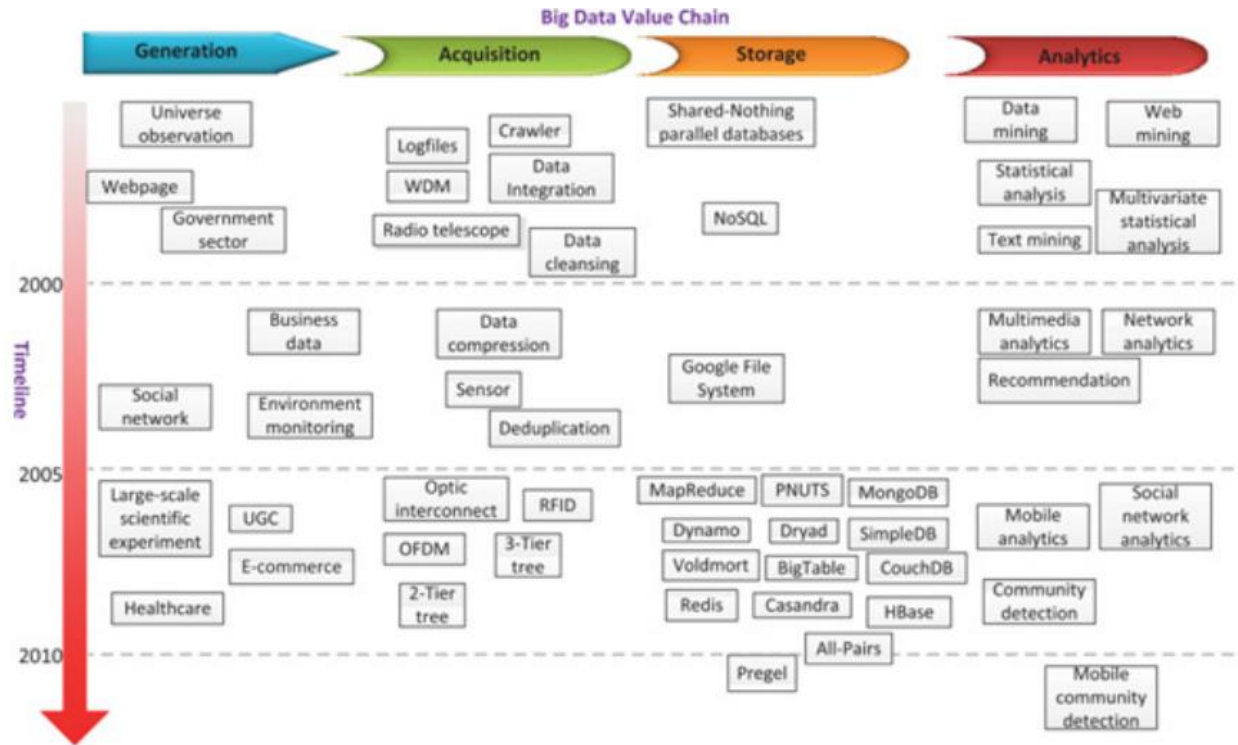
## 빅데이터의 출현배경

과학기술의 발달에 따라 빅데이터는 왜 늘어나게 되었을까요?

# 빅데이터의 이해

## 빅데이터의 출현 배경

2000년대 이후 플랫폼 및 분석과 관련된 다양한 기술이 개발되었고, 2010년 이후 빅데이터 기술이 주목을 받기 시작함



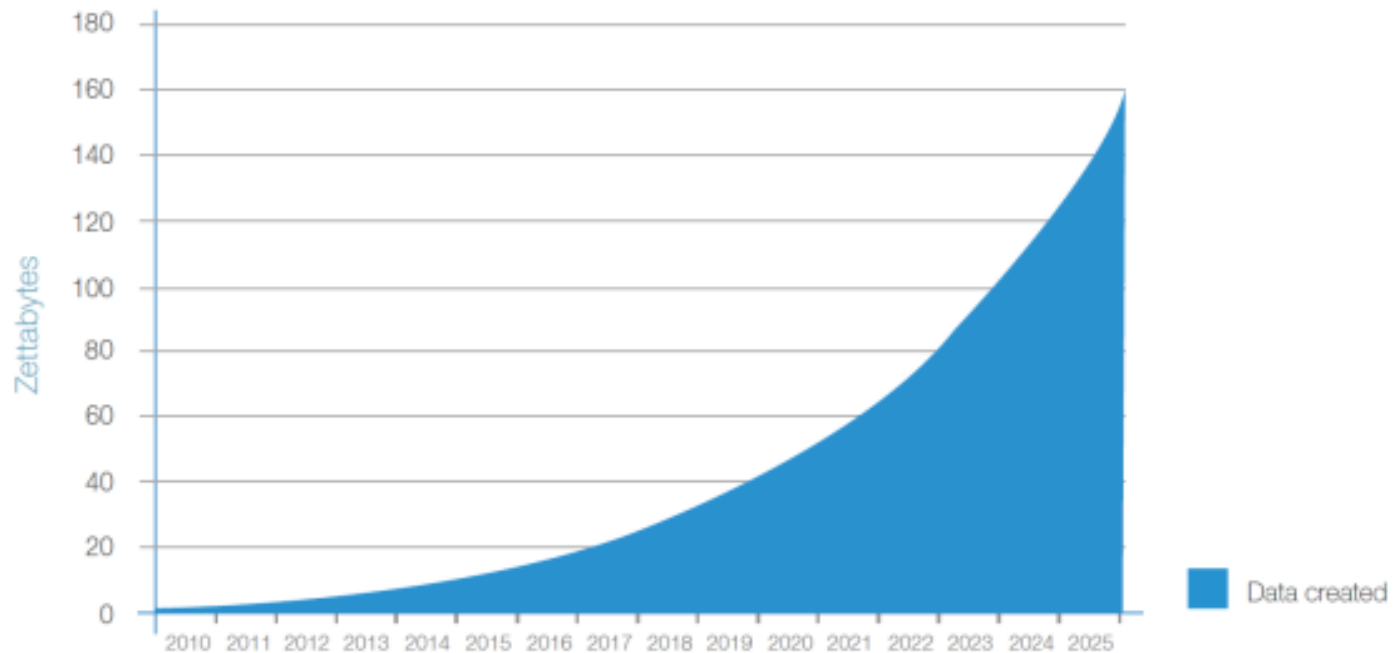
출처 : Hu et al.(2014)

[그림 2] 빅데이터 기술 발전 추세

# 빅데이터의 이해

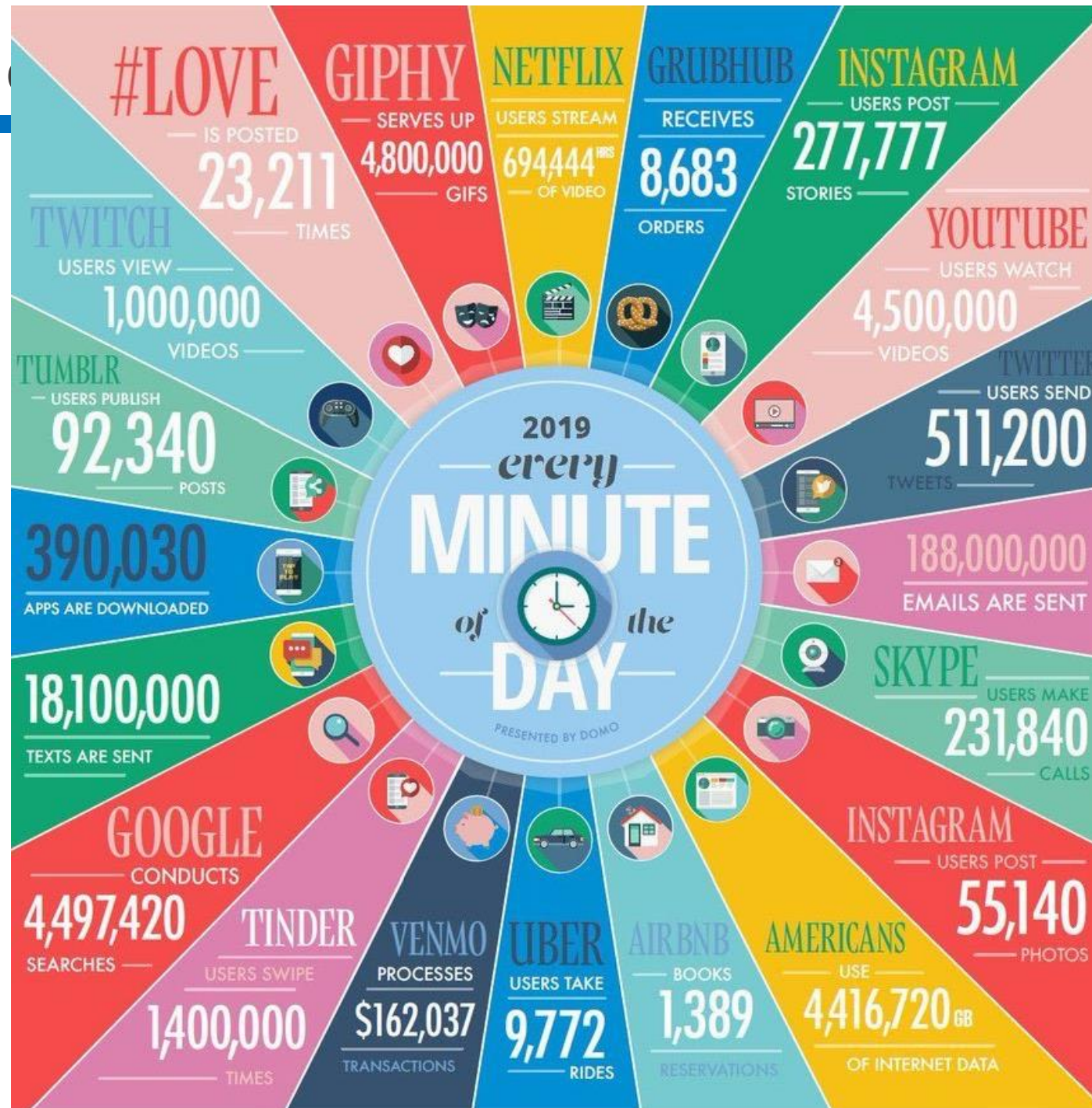
## 빅데이터의 출현 배경

모바일 기기의 확산 및 SNS의 활성화 등으로 전 세계 데이터량이 증가하기 시작하였고, 사물인터넷의 발달로 데이터량은 기하급수적으로 증가할 것으로 예상되고 있어 데이터로부터 새로운 가치와 혁신을 창출하기 위한 관련기술이 중요하게 부상하고 있음



출처 : Reinsel et al.(2017)

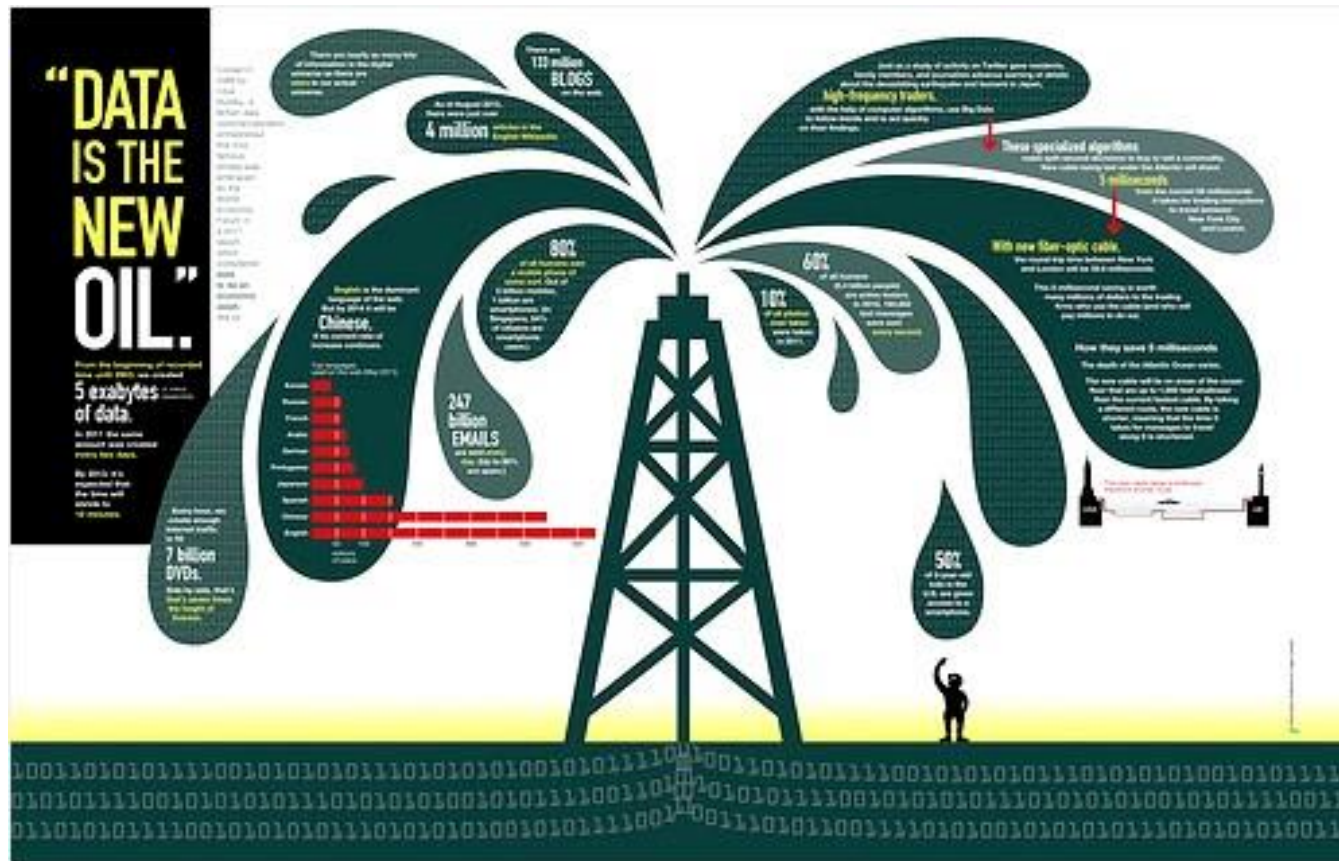
[그림 1] 연도별 세계 데이터 예측 규모





## 빅데이터에 거는 기대

2016년 다보스포럼에서 논의한 4차 산업혁명을 견인할 기술로 제시하고 있는 인공지능, 로봇 공학, 사물인터넷, 자율주행차 등은 빅데이터 기술과 밀접하게 연관되어 있으며, 가트너 (Gartner)는 빅데이터가 21세기 원유라고 발표하여 빅데이터의 중요성을 강조함



# 빅데이터의 이해

## 빅데이터에 거는 기대

빅데이터는 “산업혁명의 석탄·철”	제조업뿐만 아니라 서비스 분야의 생산성을 획기적으로 끌어올려 사회·경제·문화·생활 전반에 혁명적 변화를 가져올 것으로 기대
빅데이터는 “21세기 원유”	빅데이터도 원유처럼 각종 비즈니스, 공공기관 대 국민 서비스, 그리고 경제 성장에 필요한 정보를 제공하여 산업 전반의 생산성을 향상시키고 새로운 범주의 산업을 만들어낼 것으로 기대
빅데이터는 “렌즈”	렌즈는 통해 현미경이 생물학 발전에 끼쳤던 영향만큼, 빅데이터도 렌즈처럼 산업 발전에 큰 영향을 줄 것으로 기대
빅데이터는 “플랫폼”	플랫폼은 공동 활용의 목적으로 구축된 유무형의 구조물을 말한다. 빅데이터는 플랫폼으로서 다양한 서드파티 비즈니스에 활용될 것으로 기대된다.

# 빅데이터의 이해

---

## 빅데이터의 가치와 영향

빅데이터의 가치에 대해 생각해 봅시다.

# 빅데이터의 이해

## 빅데이터의 가치와 영향

현재 빅데이터 기술은 전 산업에 영향을 미치는 기반 기술로 자리매김하고 있으며, 지능형 분석을 통한 '변화 예측형 빅데이터' 기술로 발전하고 있음

최근 글로벌 주요 50개 기업 대상으로 설문조사를 한 결과, 95% 이상의 기업이 빅데이터 기술에 투자하고 있으며, 비용절감이나 혁신을 위한 새로운 방향 마련 등에서 성과를 거두고 있다고 응답함(NVP, 2017)

현재 데이터 간 연계활용, 실시간 분석 능력 향상, 지능·예측형 분석 기술 확보, 빅데이터 수집·처리·저장·분석에 최적의 성능을 얻을 수 있는 기술이 개발되고, 유의미한 정보 획득이라는 기존 빅데이터 기술의 목적이 변화 예측으로 바뀌고 있음(IITP, 2016a)

- 사물인터넷(IoT) 관련 기기로부터 시계열에 따라 생성되는 다양한 트랜잭션, 고객행동, 센서기록, 보안 위협 등이 2018년 이후에는 빅데이터 추세를 결정할 중요한 요소가 될 것으로 전망되어(Foote, 2017) 실시간으로 분석할 수 있는 기술들이 개발되고 있음
- - 빅데이터와 인공지능 기술이 결합하여 비즈니스 혁신이 이루어지고 있으며, 인공지능의 판단 및 학습능력을 가능하게하기 위해 빅데이터의 수집·분석·활용 기술은 필수적임

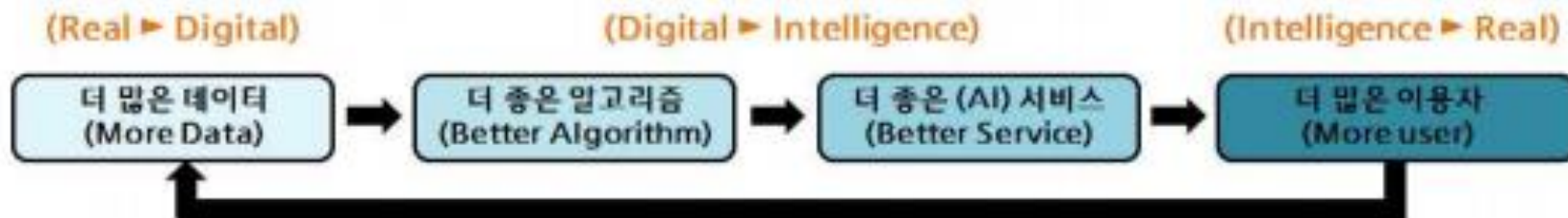


# 빅데이터의 이해

## 빅데이터의 가치와 영향

정보기술은 컴퓨터·인터넷이, 지능기술은 '데이터·AI·알고리즘'이 핵심

- IoT 기술을 통해 현실세계의 다양한 사물과 현상이 데이터화되고, 실시간으로 수집된 대량·다종의 데이터가 사이버세계로 수집·연계
- - 데이터를 AI 기술로 분석해서 맞춤형 서비스·마케팅, 비용·에너지 효율화, 안전·보안 예측이나 대응, 의사결정 등의 지능형 서비스로 연계
- - 그러나 지능기술·서비스를 근본적으로 차별화시키고 부가가치를 높이기 위해서는 동일 문제에 다양하게 접근할 수 있는 알고리즘이 중요



# 빅데이터의 이해

## 빅데이터의 가치와 영향

기업(국가) 경쟁력을 결정짓는 인프라 : 데이터(Data First)

- 지능화 시대에는 사물이 데이터를 생산하면서 세상의 모든 것을 데이터로 기록(Datafication)하고 의사결정에 필요한 데이터를 항상 활용

세상을 혁신하는 생활 속의 기술 : AI(AI Everywhere)

- 누가 더 좋은 알고리즘으로 어떤 분야에서 차별화된 AI 서비스를 제공하느냐가 경쟁우위를 확보하는 관건

팩트를 해석하는 중요한 관점 : 알고리즘(Data Storyteller)

- 데이터와 AI 기술을 결합해 차별화된 서비스나 최적의 결과를 찾아내는 지식 알고리즘이 경쟁력으로 작용

# 빅데이터의 이해

## 빅데이터의 위기 요인과 통제 방안

위기 요인	통제 방안
사생활 침해	동의에서 책임으로
책임 원칙 훼손	결과 기반 책임 원칙 고수
데이터 오용	알고리즘 접근 허용

# 빅데이터의 이해

---

빅데이터가 앞으로 삶을 어떻게 바꿀 것인가?



다음 주에 봐요