

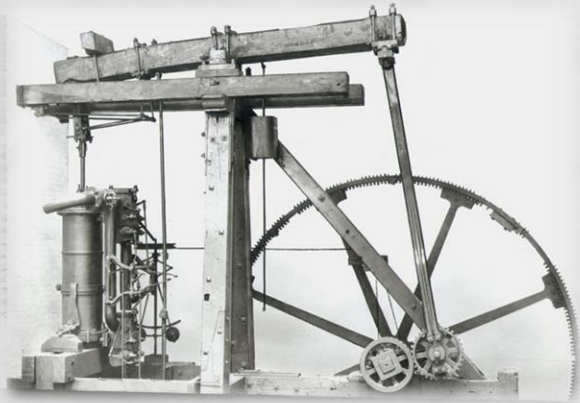
I . Overview

1. The Fourth Industrial Revolution

The Fourth Industrial Revolution

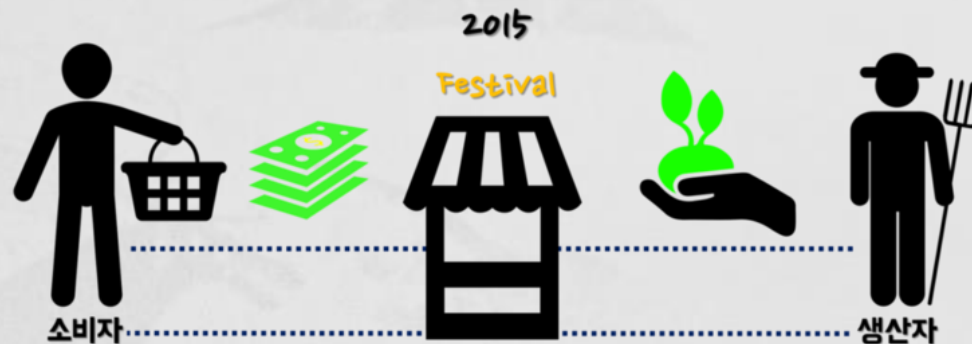
Definition

- 1차 산업혁명(1780) : 증기, 석탄



Definition

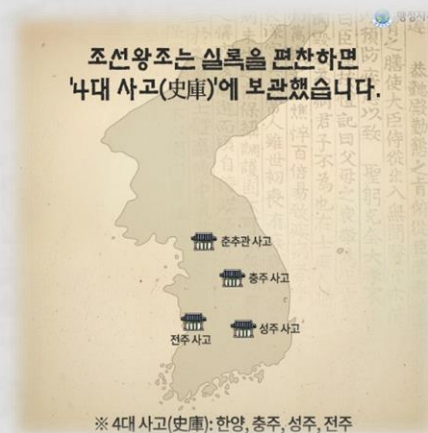
-



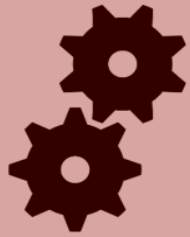
The Fourth Industrial Revolution

Definition

- 3차 산업혁명(1970) : Digital, Computer, Cyber World



The Fourth Industrial Revolution



제 1차 산업혁명

18세기

증기기관 기반의
기계화 혁명

증기기관을 활용하여
영국의 섬유공업이
거대산업화



제 2차 산업혁명

19세기~20세기 초

전기에너지 기반의
대량생산 혁명

공장에 전력이 보급
되어 벨트 컨베이어를
사용한 대량 생산보급

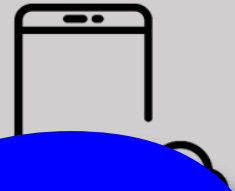


제 3차 산업혁명

20세기 후반

컴퓨터와 인터넷 기반의
지식정보 혁명

인터넷과 스마트
혁명으로 미국주도의
글로벌 IT기업 부상



제 4차 산업혁명

2017년

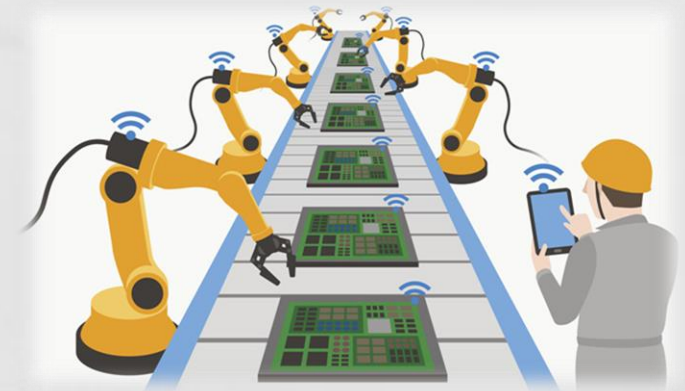
IOT/빅데이터/인공지능
의
만물 초지능 혁명

사람과 공간
초연결 초지능화
하여 산업구조
사회 시스템 혁신

The Fourth Industrial Revolution

Definition

- 4차 산업혁명(2016) : 자동화와 연결성 극대화, **Cyber Physical System(CPS)**



The Fourth Industrial Revolution



Big Data

Data Analytics?

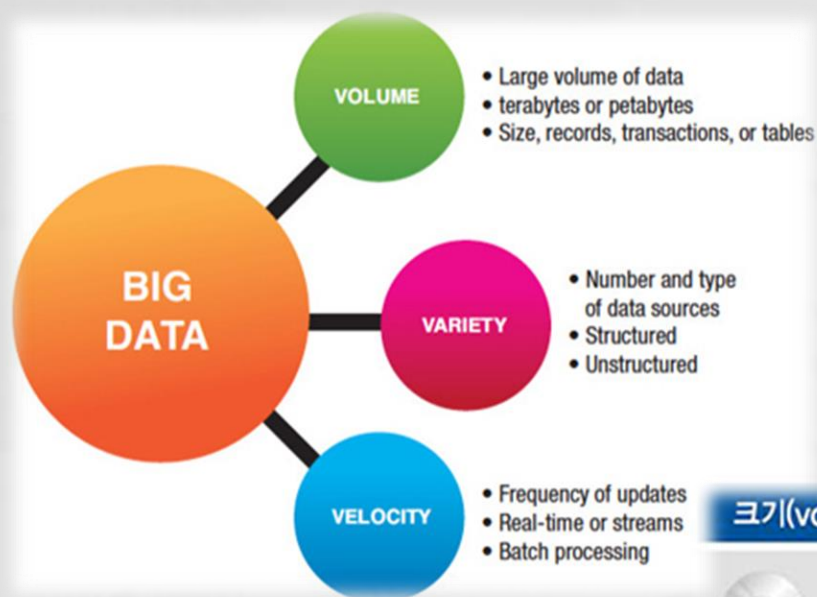
What is Data Analytics?



Definition of Data

- 데이터 : **추론과 추정의 근거를 이루는 사실**(옥스퍼드 대사전)
 - 데이터(Data)라는 용어는 1646년 영국 문헌에 처음 등장
 - 라틴어인 Dare(주다)의 과거분사형으로 **주어진 것**이란 의미로 사용
- 기술(記述)적(Descriptive) : **객관적 사실**을 기록하는 것
 - 경영학, 통계학 등에서 **기술적이고 사실적인 의미**로 발전
 - 다른 객체와의 상호관계 속에서 가치를 가지는 것으로 설명
- 자료(Data)를 분류하여 정보(Information)를 생성 및 활용
 - 자료(Data) -> 정보(Information) -> 지식(Knowledge) -> 지혜(Wisdom)

Big Data 3V

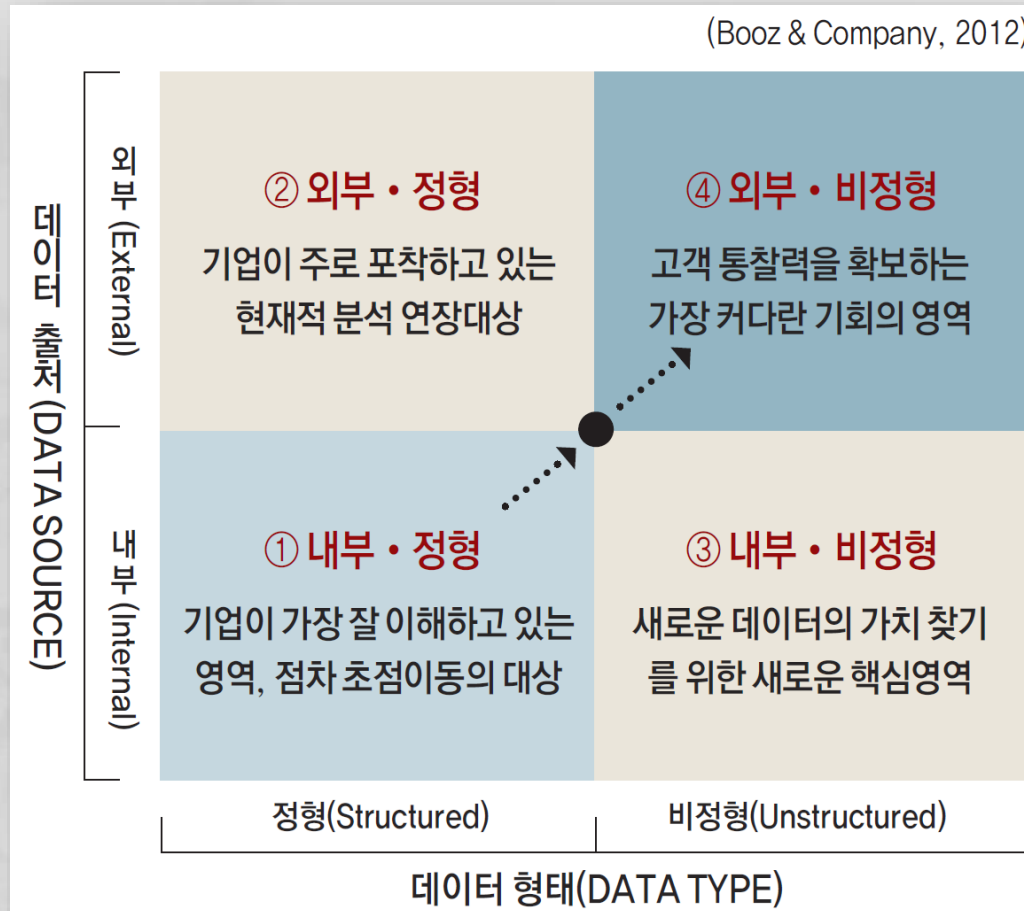


Big Data 6V

- 기존 3V 모델 : Volume, Variety, Velocity
- 6V 모델로 확장
 - Volume(크기) : TB 또는 PB 이상의 방대한 크기
 - Variety(다양성) : 정형(RDB) 및 비정형(음성, 영상, 사진, SNS)
 - Velocity(속도) : 실시간 생성 및 빠른 속도로 데이터 분석
 - Veracity(신뢰성) : 의사결정을 위한 데이터의 품질과 신뢰성 확보
 - Value(가치) : 올바른 분석으로 비즈니스 의사결정을 지원
 - Visualization(시각화) : 복잡한 대량의 데이터를 시각적으로 표현

Data Type & Source

데이터 형태별 가치



Data Science

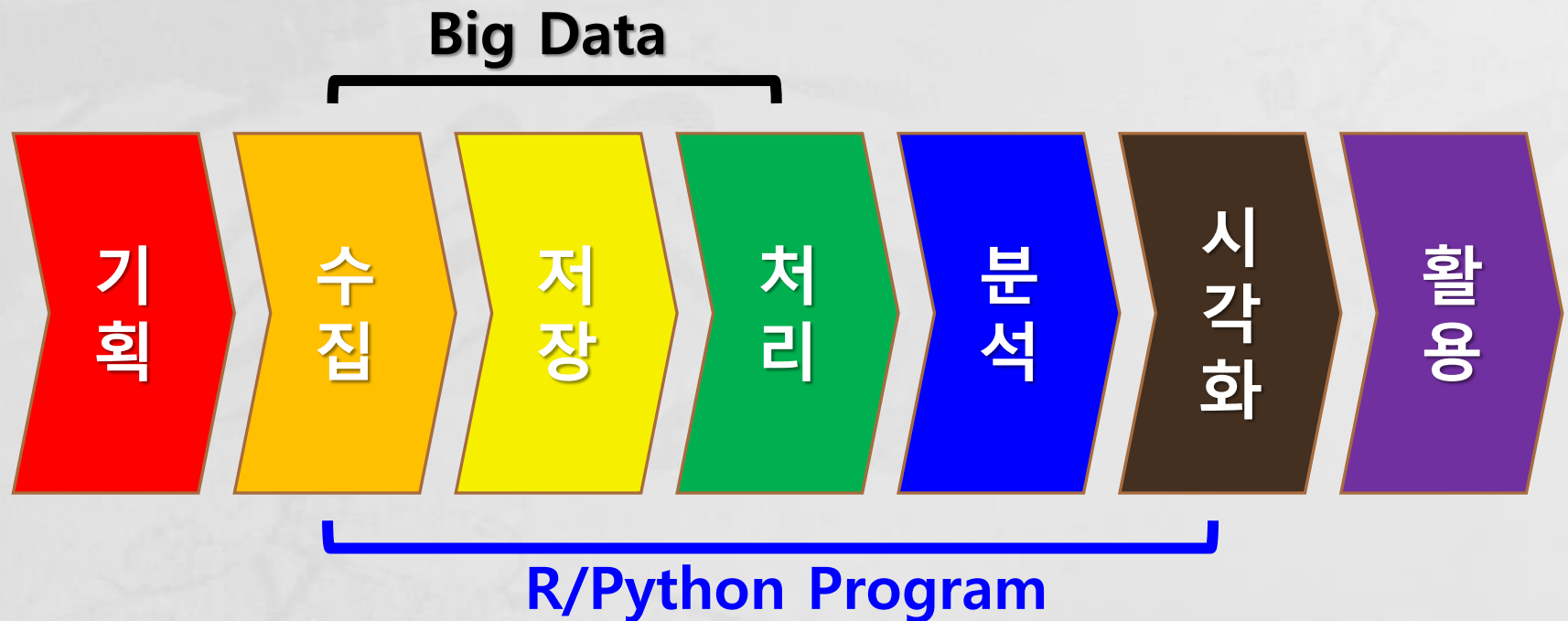
Data Scientist



LG그룹 블로그 | www.lgblog.co.kr

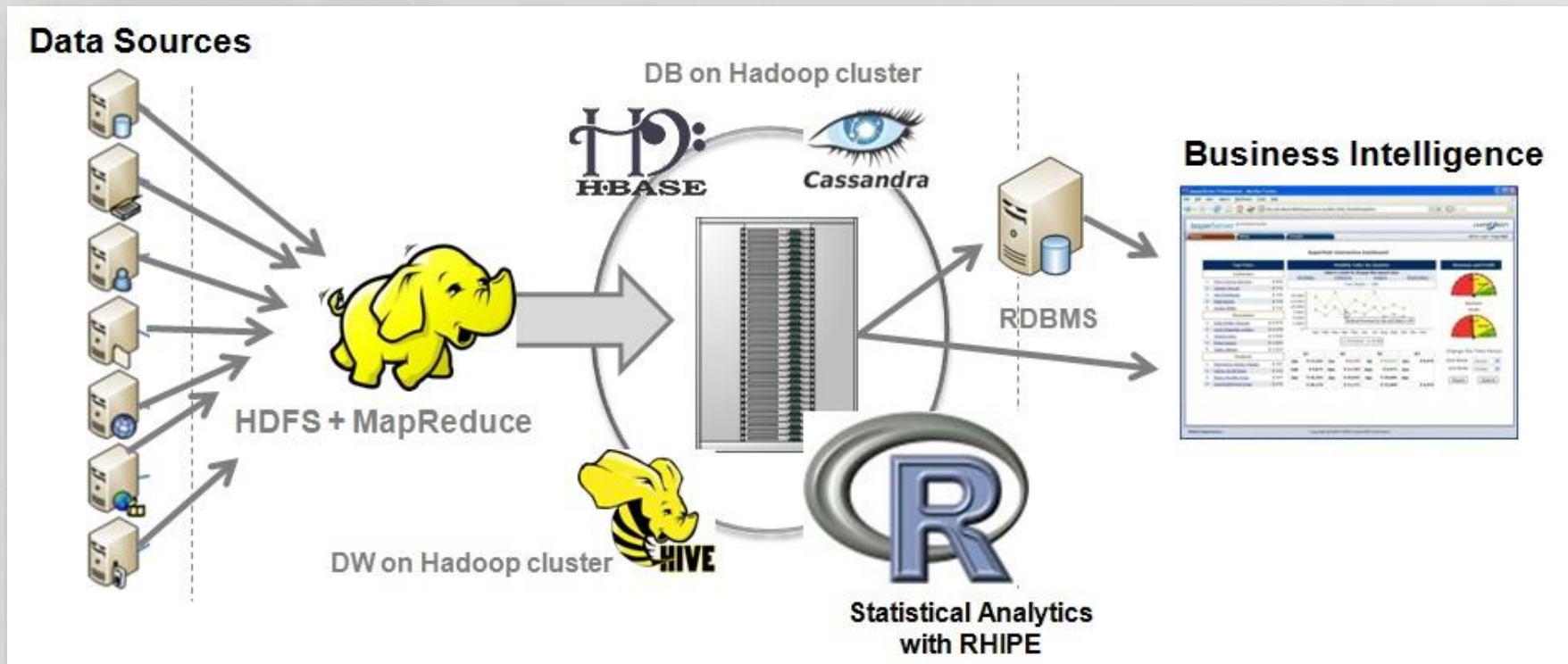
Data vs Big Data

Big Data 단계



Big Data

Data Engineering vs. Data Analytics



Big Data

공공 데이터 포털(<https://www.data.go.kr>)

The screenshot shows the Korea Open Data Portal homepage. At the top, there's a navigation bar with the logo 'DATA . GO . KR' and links for '로그인' (Login), '회원가입' (Sign Up), '사이트맵' (Site Map), 'ENGLISH', and social media icons. Below this is a secondary navigation bar with '데이터셋' (Data Set), '제공신청' (Request for Provision), '활용사례' (Use Cases), '정보공유' (Information Sharing), and '이용안내' (Usage Guide).

The main content area features a search bar with the placeholder text '데이터를 검색해보세요!' (Search for data!). To the right of the search bar is a dropdown menu set to '2.미세먼지' (2. PM2.5). Below the search bar, a message reads: '여러분이 원하는 데이터, 데이터 1번가 로 신청해주세요!' (Data you want, please apply to Data 1st Place!).

Below this message is a grid of data categories under the heading '국가중점데이터' (National Key Data). The categories include: 건축정보 (Building Information), 교통사고정보 (Traffic Accident Information), 국민건강정보 (National Health Information), 상권정보 (Commercial District Information), 수산정보 (Fisheries Information), 실시간 수도정보 (Real-time Water Supply Information), 농수축산가격정보 (Agriculture, Forestry, and Fisheries Price Information), 등산로 정보 (Hiking Trail Information), 부동산종합정보 (Real Estate Comprehensive Information), 통합재정정보 (Integrated Financial Information), 지방행정정보 (Local Government Information), 부동산거래정보 (Real Estate Transaction Information), 식의약품종합정보 (Food and Drug Comprehensive Information), 지방재정정보 (Local Finance Information), 법령정보 (Regulatory Information), and a '+ 더보기' (More) button.

At the bottom, there are three main promotional banners. The left banner is for '데이터1번가' (Data 1st Place), encouraging users to apply for public data. The middle banner is for '오픈데이터 포럼' (Open Data Forum), mentioning a policy workshop and a forum for data utilization. The right banner is for '품질관리 및 진단 기본교육' (Quality Management and Diagnosis Basic Education), a course for public officials.

Below the banners, there are two more sections: '공공데이터 활용사례' (Public Data Use Cases) and '인기 데이터' (Popular Data), which includes links for '파일데이터' (File Data) and '오픈 API' (Open API).

Big Data

Kaggle(<https://www.Kaggle.com/competitions>)

kaggle

Search

Q

Competitions

Datasets

Kernels

Discussion

Learn

...

Competitions

Documentation

InClass

General

InClass

Sort by


Grouped

All Categories

Search competitions


Q

18 Active Competitions


TWO SIGMA


Two Sigma: Using News to Predict Stock Movements
Use news analytics to predict stock price performance
Featured · 6 months to go · news agencies, time series, finance, money

\$100,000
2,902 teams


Los Alamos
NATIONAL LABORATORY
EST. 1943

LANL Earthquake Prediction
Can you predict upcoming laboratory earthquakes?
Research · 5 months to go · earth sciences, physics, signal processing

\$50,000
319 teams

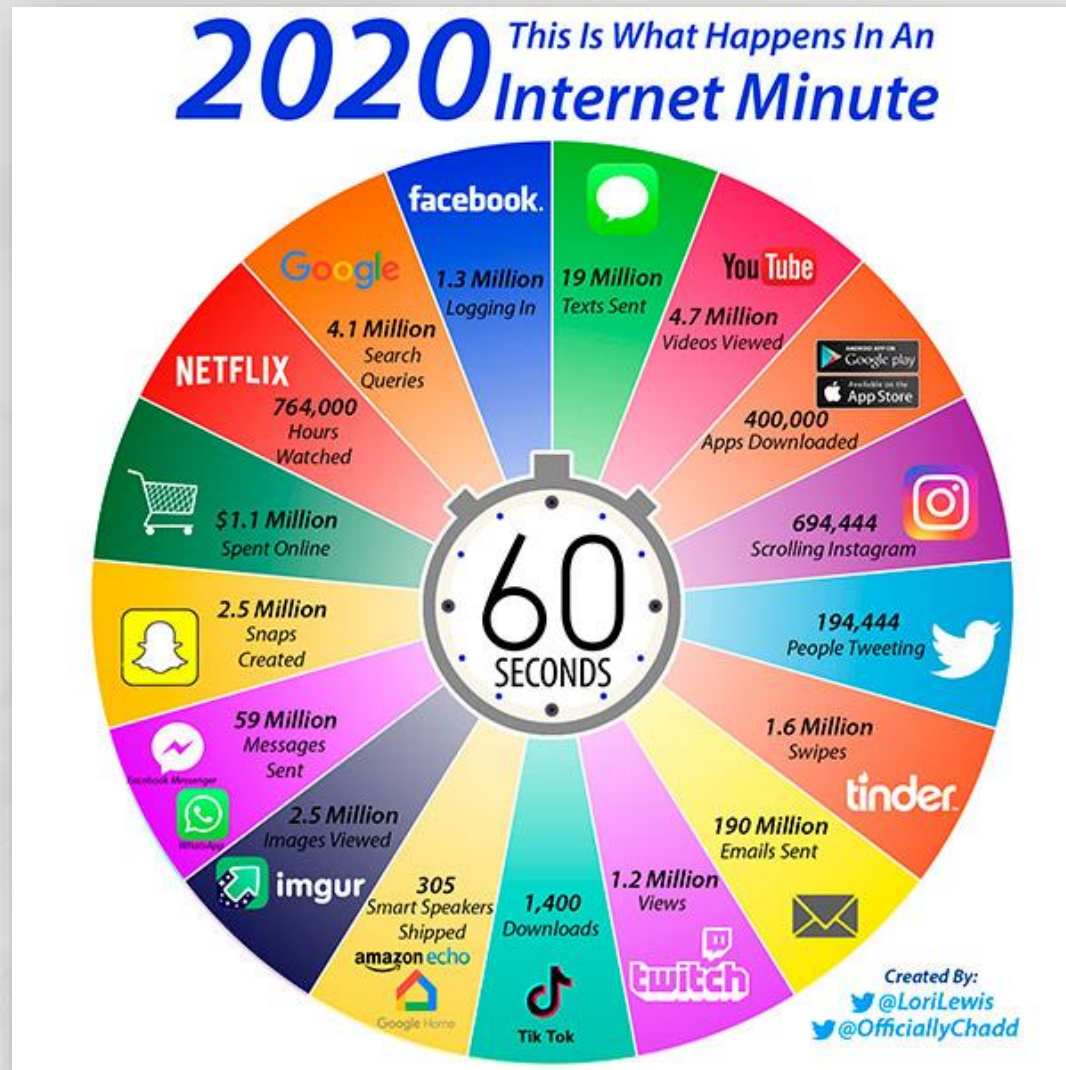

ELO MERCHANT

Elo Merchant Category Recommendation
Help understand customer loyalty
Featured · a month to go · banking, tabular data, regression

\$50,000
2,630 teams

Big Data

Internet Minute 2020



Principles of Cloud Computing

Before we start



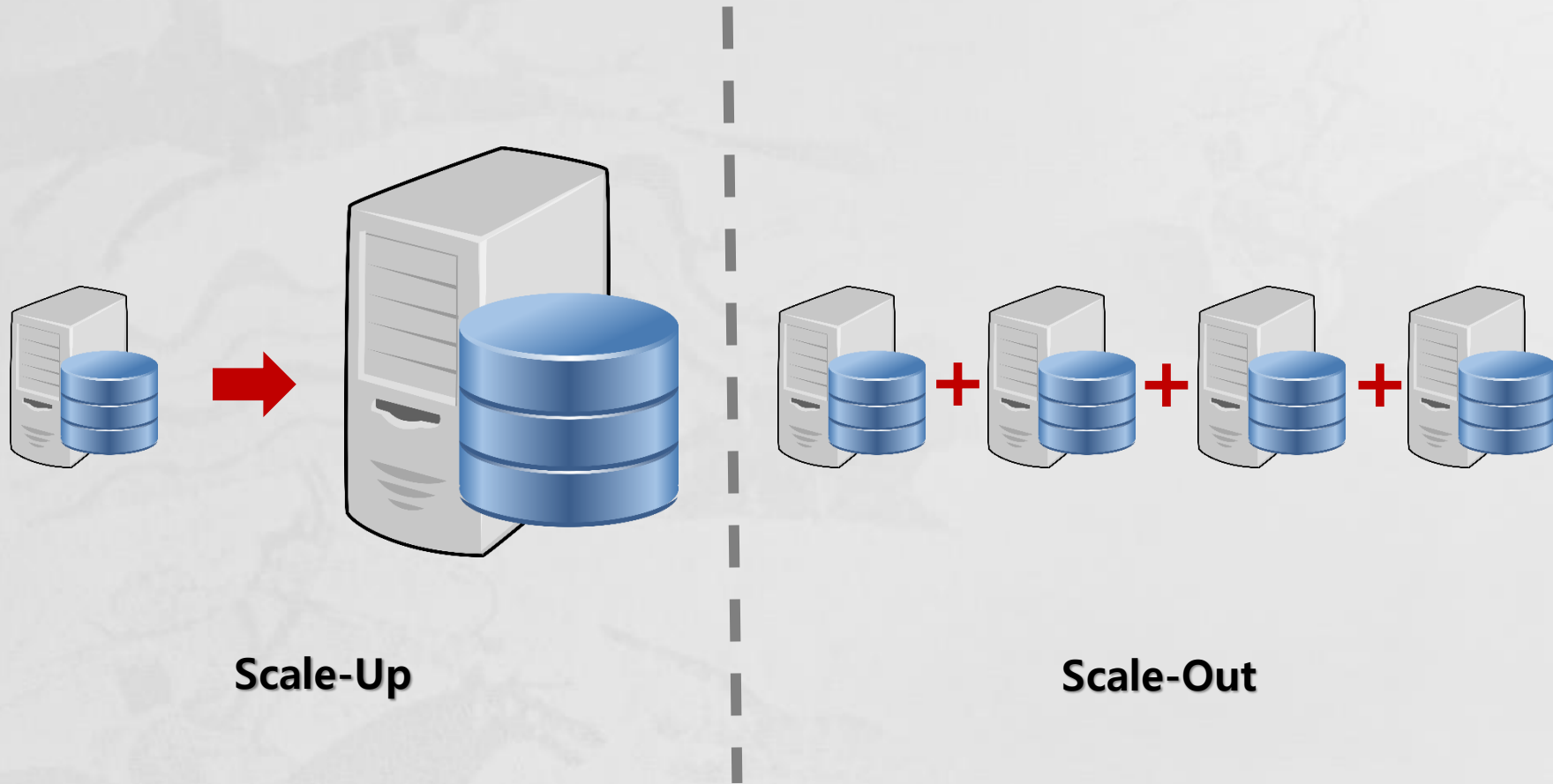
Principles of Cloud Computing

Definition

- 클라우드 컴퓨팅은 인터넷(**Network**) 기술을 이용하여 내-외부 고객들에게 확장성(**Scalable**) 있고 탄력적(**Elastic**)인 IT 서비스가 제공되는 방식
- Cloud is a Style of Computing** where **Scalable** and **Elastic** IT-related capabilities are provided **as a service** to customer using **Internet technologies**
- 확장성(범위성)은 늘어나는 것만이 아니라 줄어드는 것도 포함
- 컴퓨팅 자원이 필요한 만큼 늘어나거나 적정 수준으로 줄어들 수 있음
- 이러한 작업(요청)이 수분에서 수십분 이내로 가능
- 확장성은 범위, 탄력성은 시간

Principles of Cloud Computing

Scale-Up vs. Scale-Out



Principles of Cloud Computing

Cloud Computing Benefits

- Reduced Cost(Pay-Per-Use, Economics of Scale)
- Automated(Updates, Security Patches, Backups ...)
- On demand(Flexibility + Scalability = Elasticity)
- Mobility(Accessible from any **web enabled** device)
- Shared Resources(Multi-Tenancy)
- Agility and Scalability
- Back to core business
- More IT functionality for a lower price

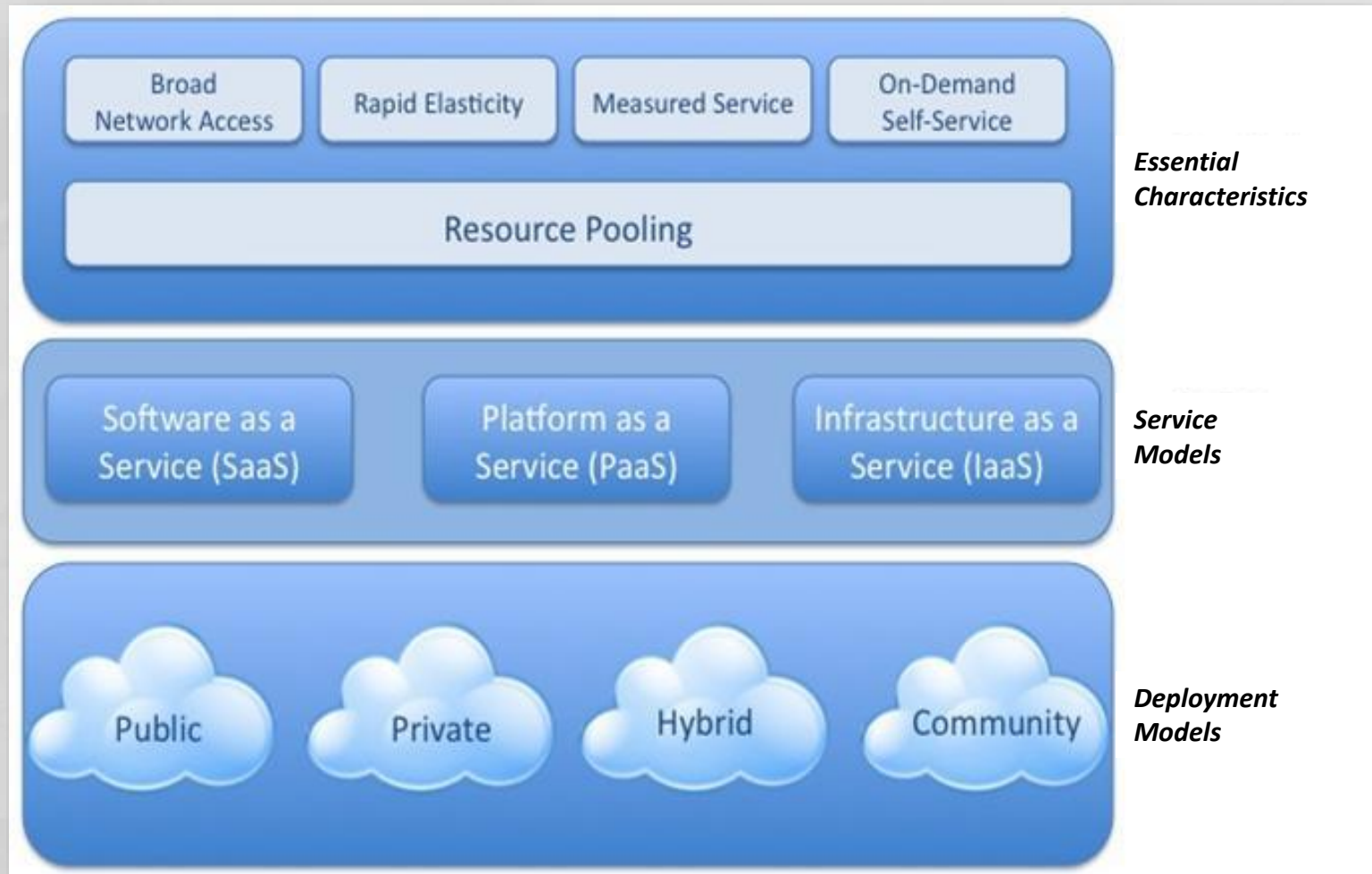
Principles of Cloud Computing

Cloud Computing Limitations

- Internet Access(No Internet = No Cloud)
- Security(How do you know?)
- Privacy(What **legislation** or regulations?)
- Vendor Lock-in(Application **migration** may be impossible)

Principles of Cloud Computing

Structure - NIST



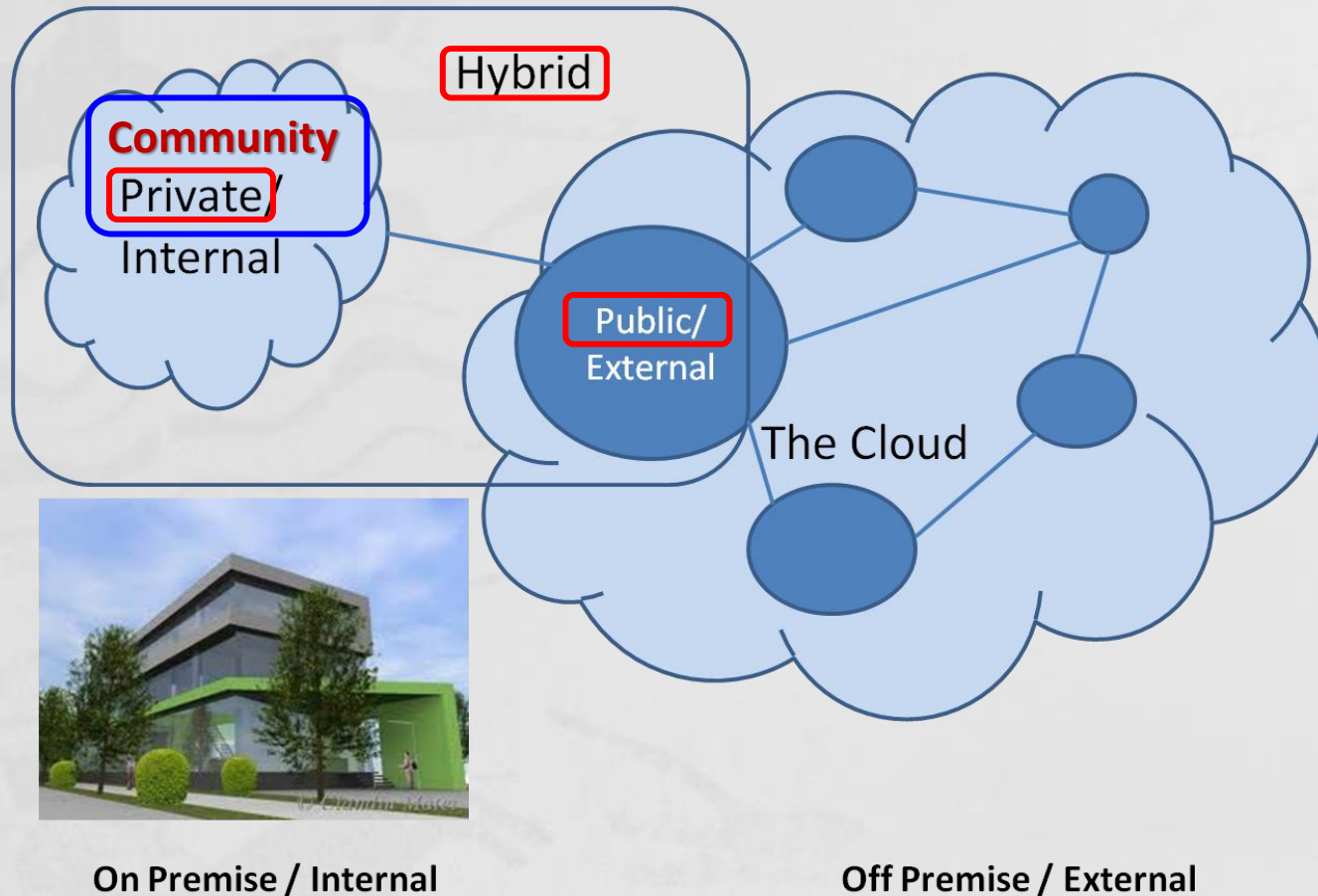
Principles of Cloud Computing

Five Characteristics

- 주문형 셀프서비스(On-Demand Self-Service)
 - 사용자가 별도의 기술습득 없이 필요할 때 온라인으로 즉시 사용
- 광대역망 액세스(Broad Network Access)
 - 네트워크를 통한 컴퓨팅 자원 접근(Any time, Any place, Any device)
- 자원 공동관리(Resource Pooling)
 - 다중 임대 모델을 통한 자원 할당(Multi-Tenancy)
- 빠른 요구탄력성(Rapid Elasticity)
 - 비즈니스 상황에 따른 컴퓨팅자원의 탄력적 사용(Flexibility, Scalability)
- 도수제(Measured Service)
 - 서비스를 사용한 만큼 비용 지불(Pay-Per-Use, Pay as you go)

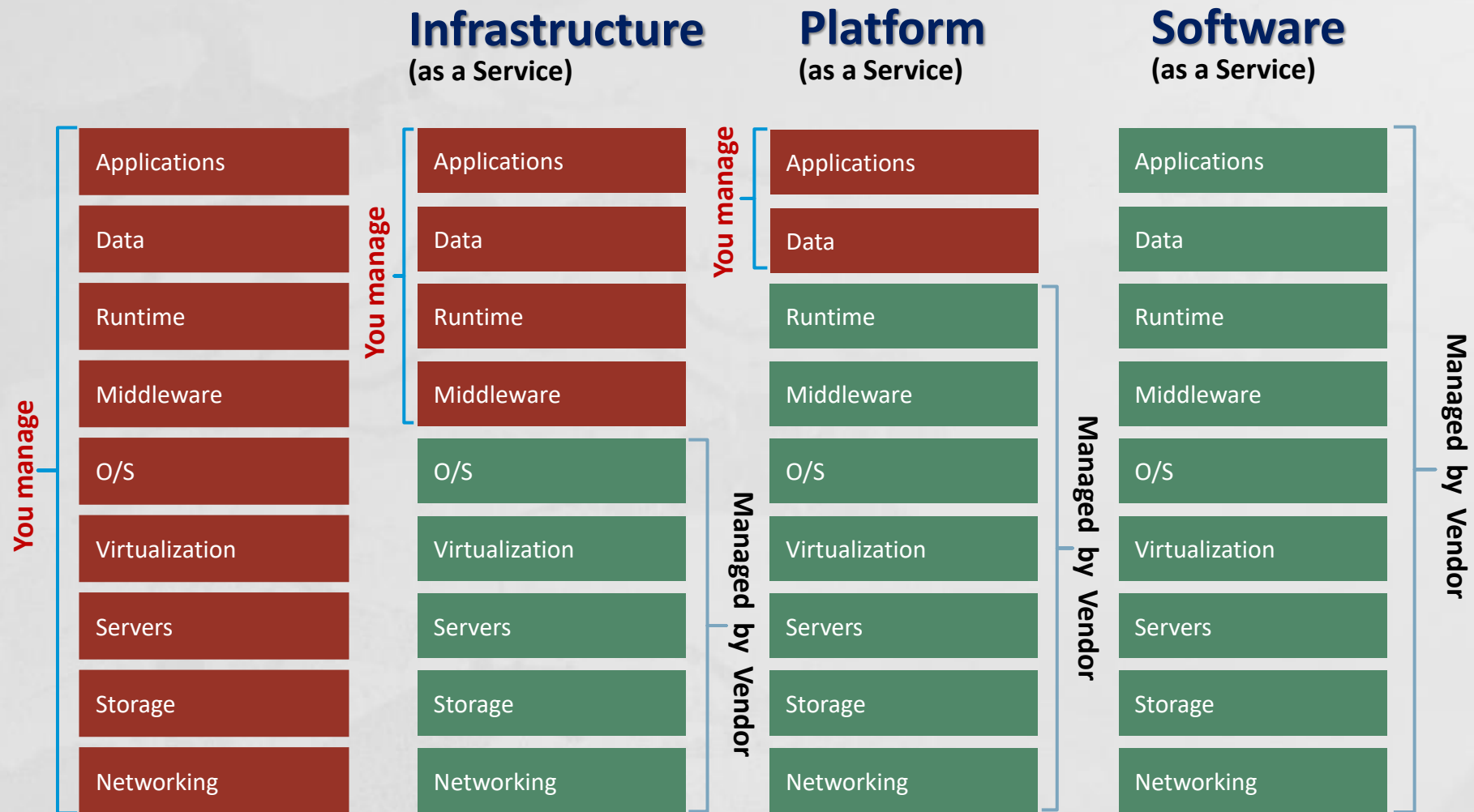
Cloud Computing Deployment Models

Four Deployment Models



Cloud Computing Service Models

Tree Service Models



Cloud Computing Service Models

Software as a Service(SaaS)

- 소프트웨어를 네트워크를 통하여 서비스 받음
- 설치 형태가 아닌 서비스 형태로 제공한다는 의미(as a Service)
- 사용자가 네트워크로 접속하여 메일, CRM, ERP, Office 프로그램 사용
- 사용자는 실행을 위해 Software를 추가로 설치할 필요가 없음
- Software가 실행되는 컴퓨터나 저장장치에 대해 알지 못함
- Software의 실행 환경은 서비스 제공자에 의해 관리
- 무료로 서비스가 제공되거나 이용한 만큼 비용 지불
- Google Apps, salesforce.com, MS Office 365, Naver Office

Cloud Computing Service Models

Platform as a Service(PaaS)

- 플랫폼이란 Software 개발환경이나 실행하기 위한 기본적인 모듈군
- 애플리케이션이나 서비스가 실행(운영)되는 환경을 제공
- 서비스 개발 및 운영 환경에 따라서 다양한 미들웨어가 필요
- 일반적으로 회사 내 서버에 직접 구축하는 방식으로 진행
- 사용자가 제공 받을 수 있는 실행 환경이 제한적인 단점을 극복
- 실행 환경 표준화에 따른 애플리케이션의 개발이 가능
- 네트워크를 통해 서비스 받고 이용한 만큼 비용 지불
- Google Cloud Platform, force.com, Windows Azure

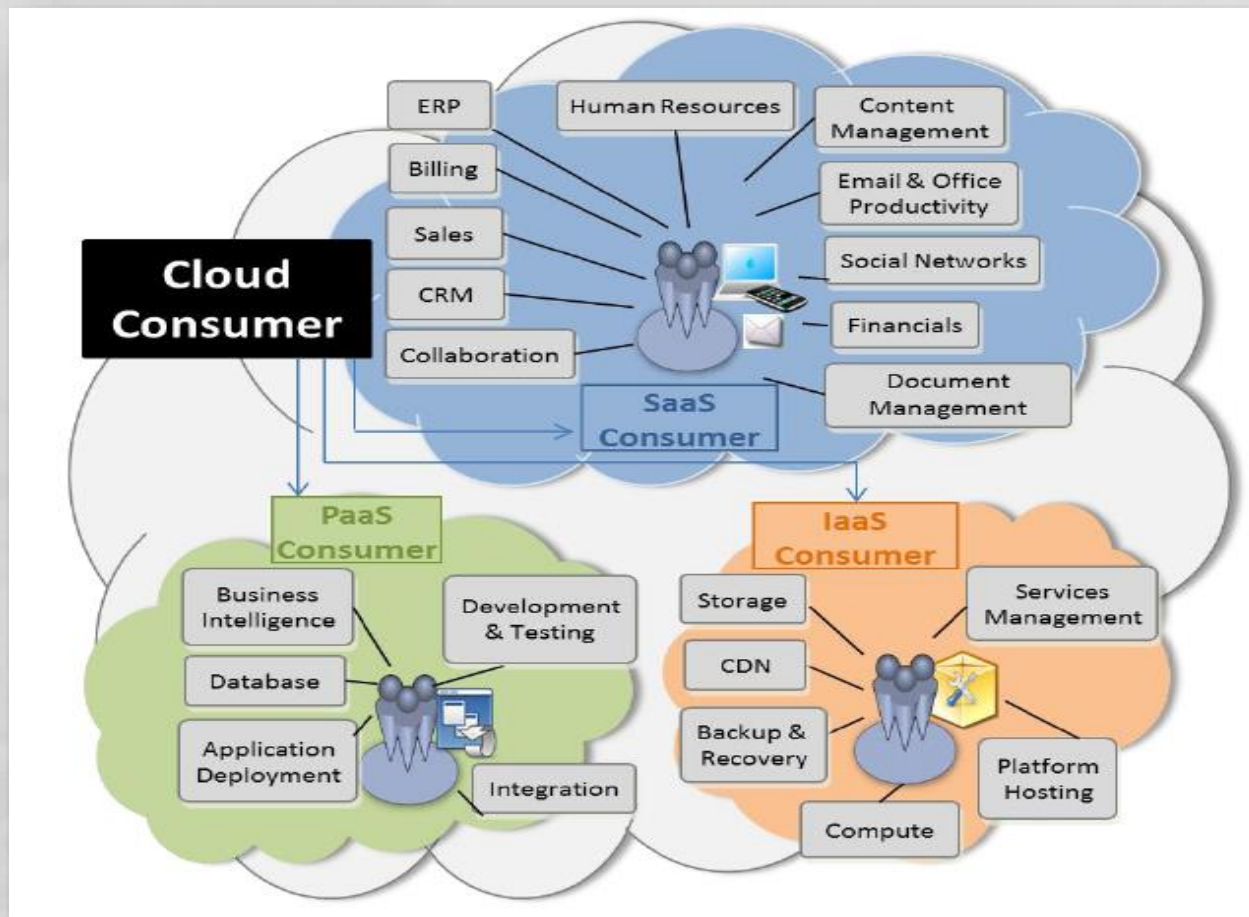
Cloud Computing Service Models

Infrastructure as a Service(IaaS)

- 인프라(Infrastructure)는 서버, 스토리지, 네트워크 등의 하드웨어를 의미
- 서버, 스토리지, 데이터베이스 등과 같은 물리적인 IT 자원
- 업무시스템을 구축할 수 있도록 자원을 할당해 주는 서비스
- 물리적 자원을 제공하기 위하여 가상화 기술을 활용
- 하드웨어 리소스나 저장장치를 인터넷을 경유해 이용하는 서비스
- 자원에 대해 사용한 만큼만 비용을 지불하는 경제성
- 데이터 센터를 직접 구축하거나 퍼블릭 클라우드를 이용
- Amazon Web Service(AWS), Windows Azure

Cloud Computing Service Models

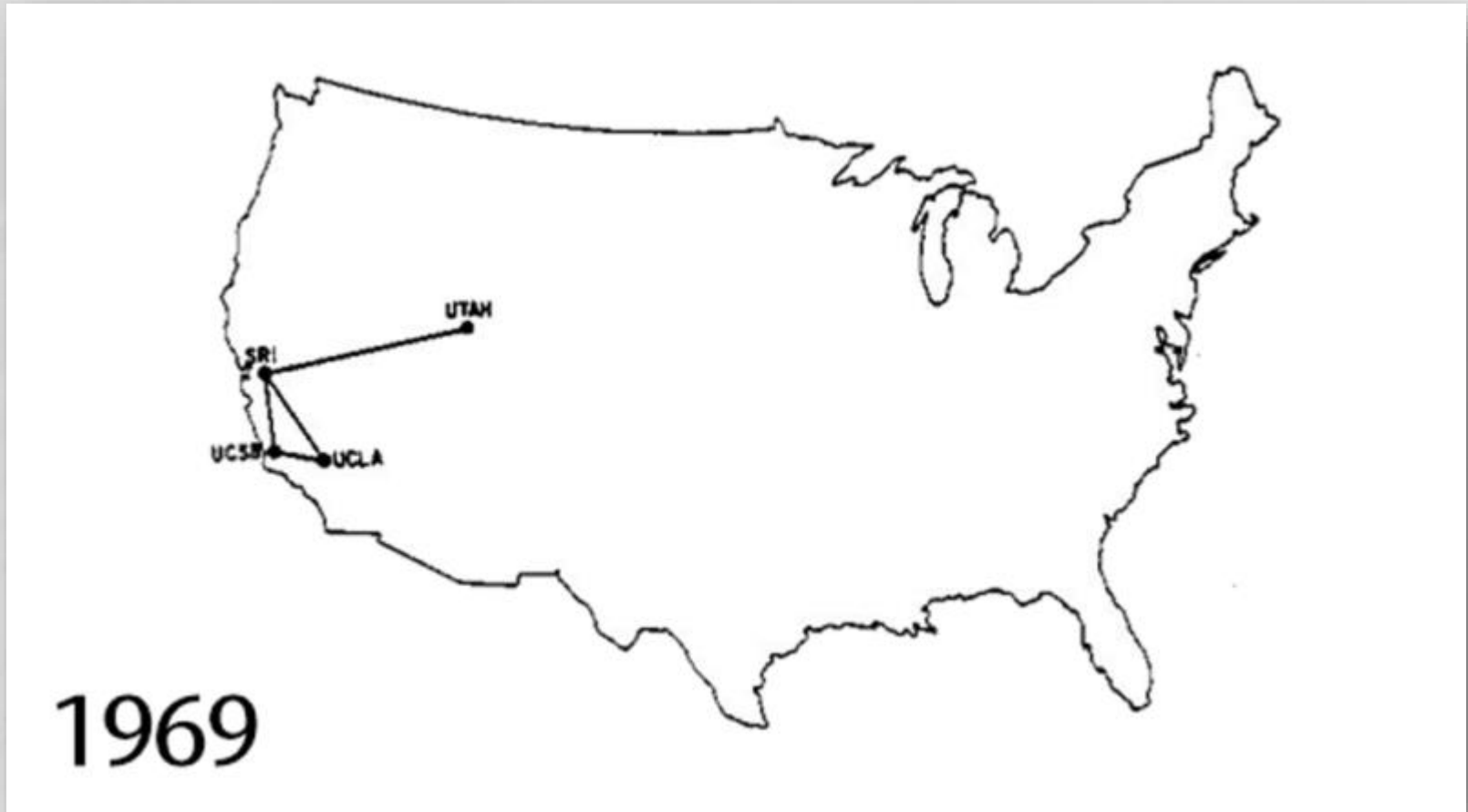
Examples of Services



Cloud Computing Technologies

Cloud Computing Technologies

Network



Cloud Computing Technologies

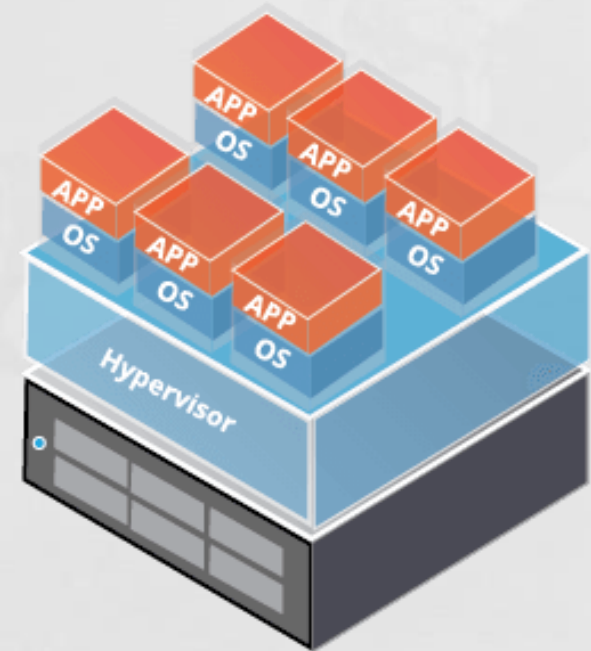
Virtualization

- 가상화는 “실제로 존재하지 않는 것을 존재하는 것처럼 보여줌” “실제로는 복수로 존재하고 있는 것을 논리적으로 통합하여 다른 새로운 것으로 보여줌과 같은 의미
- 구체적으로는 시스템 이나 네트워크를 구성하는 각종 물리 자원을 논리적으로 마치 다른 자원인 것처럼 인식시키는 기술이 나 방법을 의미
- 가상화 기술을 적용함에 따라서 자원을 유효하게 활용할 수 있으며 확장성 · 가용성 향상 운영관리 개선 등을 기대

Cloud Computing Technologies

Virtualization

- Exists since the 1970s in mainframe environments
- IBM VM/370(1972)



Cloud Computing Technologies

Virtualization

- 각종 물리 자원을 논리적으로 마치 다른 자원인 것처럼 인식시키는 기술이나 방법
- 대상 : 서버, 데스크톱, 스토리지, 네트워크, 소프트웨어 등 모든 IT 자원
- 구현 방법
 - 자원 가상화
 - 물리적 자원을 다른 특성을 지닌 별도의 논리자원으로 보여 줌
 - 가상메모리 = 주기억장치(물리적 메모리) + 보조기억장치(하드디스크)
 - 자원 분할
 - 단일 물리 자원을 논리적으로 분할하여 별도의 물리 자원으로 인식, 동작
 - 서버 가상화 : 한대의 서버 위에서 여러 개의 OS를 동시에 운영
 - 자원 통합
 - 복수의 물리 자원을 논리적으로 통합하여 단일한 물리 자원처럼 동작
 - 서버 클러스터링 : 복수의 서버를 통합하여 높은 처리 능력을 가진 하나의 단일 서버로 동작

Cloud Computing Technologies

Virtualization

◦ 서버 분할

- 단일 물리서버를 분할하여 복수의 논리서버 환경을 구축하여 여러 시스템을 운영
- 일반적인 서버 가상화는 대부분 서버 분할 방식
- 구현방법
 - 파티셔닝 : 물리 하드웨어 자원을 분할하여 사용
 - 가상머신 : 서버 가상화 소프트웨어에 의해 하드웨어 자원을 배분하여 사용

◦ 서버 통합

- 복수의 물리서버를 연계하여 마치 한대의 서버가 동작하는 것처럼 동작
- 구현방법
 - HA 클러스터링 : 복수의 서버를 통합하여 단일 서버 이상의 높은 가용성 지원
 - 부하분산 클러스터링 : 복수의 서버를 통합하고 병렬처리하여 고성능 처리

Cloud Computing Technologies

Virtualization

가상화 기술 대상 지원	가상화 기술 내용
서버	<ul style="list-style-type: none">• 한 대의 물리서버를 분할하여 여러 대의 각 회사의 논리서버 환경으로 구축• 여러 대의 서버를 논리적으로 통합하여 높 각 회사의 서버 클러스터링 제품은 처리능력을 가진 한 대의 서버로 구동
데스크톱	<ul style="list-style-type: none">• 한대의 물리서버 속에서 개별 유저가 사용하는 PC 환경 구축
스토리지	<ul style="list-style-type: none">• 복수의 물리 하드디스크를 통합하여 한대의 하드디스크로 인식• 한대의 물리 하드디스크를 분할하여 복수의 하드디스크로 인식
네트워크	<ul style="list-style-type: none">• 물리적인 LAN 접속 구성과는 다른 논리적인 LAN을 구축• 다른 사용자와 공유하는 인터넷 등의 네트워크상에 가상 전용선을 구축• 복수의 네트워크 장비를 한 대로 인식시켜 운영관리 작업을 용이하게 함• 한대의 네트워크 장비를 여러 대로 인식시켜 사용자와 서버에 따라 기능과 설정을 변경

Cloud Computing Technologies

Virtualization

• 비용 감소

- 용도나 부하에 따라 서버 등 자원을 분할 구성 하면 서버 수를 줄일 수 있음
- 하드웨어에 대한 투자비 용을 감소시키는 것과 함께 설치 공간이나 소비 전력 의 낭비도 줄일 수 있음

• 자원(리소스)의 유효 활용

- CPU나 메모리 등의 자원을 보다 효율적으로 이용
- 여러 서버의 각 CPU 사용률이 매우 낮은 Idle 상태(대기상태)가 많다면, 이를 한 대의 가상서버에 집약하여 CPU 처리 능력을 낭비 없이 활용

• 높은 확장성

- 서버 등을 분할 구성하면 분할된 각 서버에서 CPU나 메모리 등의 자원을 동적으로 추가·삭제할 수 있음
- 서버 등의 통합 구성을 통해서도 통합된 자원을 동적으로 추가·삭제하는 것이 가능

Cloud Computing Technologies

Virtualization

• 가용성 향상

- 서버 등을 통합 구성하면 통합된 자원에 장애가 발생하는 경우에도 나머지 자원에서 연속적인 처리를 할 수 있음
- 처리 능력은 떨어지지만 통합 구성 이 아닌 경우에 비하여 향상된 가용성을 가질 수 있음

• 운영관리 개선

- 서버 등을 분할/통합 구성함으로써 인터페이스의 통합 운영관리가 가능해지므로 관리 대상인 물리 자원을 줄일 수 있음