



# 데이터 사이언스

# 오늘의 학습내용

---

- 데이터와 데이터 사이언스
- 데이터 분석과 데이터 관리 기술
- 데이터 사이언스 요소 기술

## ◆ 데이터와 데이터 사이언스

### ■ 4차 산업혁명과 데이터

- ❖ 4차 산업혁명은 초연결성(Hyper-Connected)과 초지능화(Hyper-Intelligent)의 특성.
- ❖ 사물인터넷(IoT) 및 클라우드 등 정보통신기술(ICT)을 통해 인간과 인간, 사물과 사물, 인간과 사물이 상호 연결.
- ❖ 빅데이터와 인공지능 등으로 보다 지능화된 사회로의 변화 예측.

### 4차 산업혁명이란?

인공지능, 빅데이터, 초 연결 등으로 촉발되는 지능화 혁명, 그리고 그 이상

기술력 + 범위 증가

점점 빨라지는 산업혁명 주기

증기기관 18세기  
1차 산업혁명

전기에너지 19~20세기  
2차 산업혁명

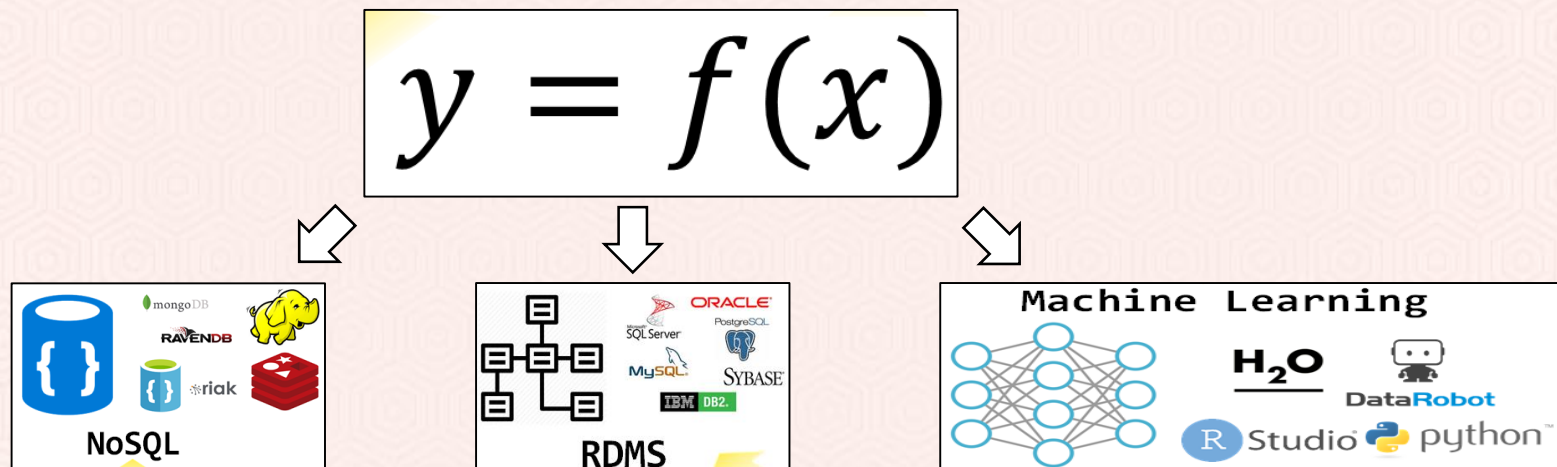
컴퓨터, 인터넷  
> 전통 IT [논리]  
20세기 후반  
3차 산업혁명

인공지능 > [직관]  
빅데이터  
초 연결  
Data 반복학습  
4차 산업혁명 21세기 초반

## ▲ 데이터와 데이터 사이언스

### ■ 데이터와 가치

- ❖ 빅데이터 시대의 데이터 가치
- ❖ 중요한 정보를 관계형 데이터베이스 기반 체계적인 관리에서 빅데이터를 저장하고 처리하고 활용할 수 있는 기술로 발전
- ❖ 최근에는 기계학습과 딥러닝 기법을 적용하여 혁신을 통해 가치를 창출



## ▶ 데이터와 데이터 사이언스

### ■ 데이터와 빅데이터

- ❖ 빅데이터란 인사이트와 의사결정, 자동화 과정을 강화할 수 있는 정보처리 과정의 혁신적인 형태
- ❖ 높은 비용 효율을 요구하고 방대하며, 빠르고 다양한 형태의 정보 자산(assets)

#### DATA SCIENCE

- 비정형 혹은 정형 데이터를 정제, 준비, 분석하는 활동

#### BIG DATA

- 엄청난 양의 데이터를 말하며, 다양한 형태의 정보 자산

#### DATA ANALYTICS

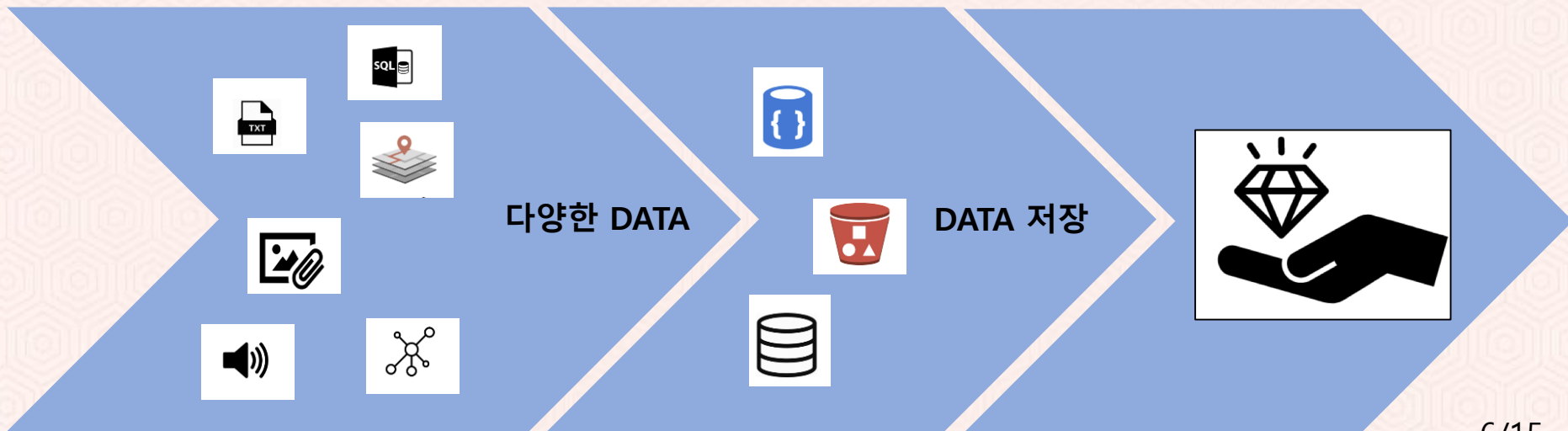
- 알고리즘과 수학적 처리 과정을 적용하여 데이터를 다루는 활동



# ▶ 데이터와 데이터 사이언스

## ■ 데이터 사이언스

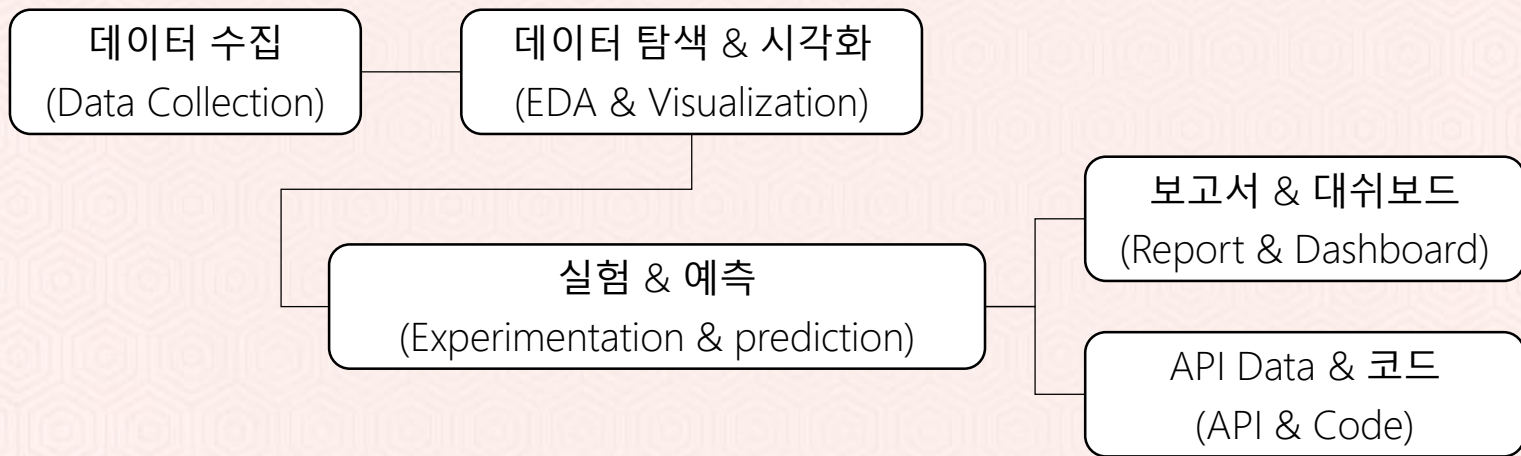
- ❖ 데이터 사이언스(data science)란, 데이터 마이닝(Data Mining)과 유사
- ❖ 정형 및 비정형 형태를 포함한 다양한 데이터로부터 지식과 인사이트를 추출하는데 과학적 방법론, 프로세스, 알고리즘, 시스템을 동원하는 융합분야



## ▲ 데이터와 데이터 사이언스

### ■ 데이터 사이언스 작업 흐름도

- ❖ 데이터를 수집하고 이를 이해하는 과정과 추상화하는 과정을 거쳐 최종 산출물
- ❖ 사람이 최종 소비자가 되는 보고서/대쉬 보드 기반과 기계가 최종 소비자가 되는 API/코드 형태로 최종 작업이 이루어짐



## ▶ 데이터 분석과 데이터 관리 기술

### ■ 데이터 분석

- ❖ 데이터 분석은 분석의 대상에 대한 문제점을 찾고 해결할 데이터를 수집 분석하여 해결방안을 찾아 이를 업무와 시스템에 도입하는 것을 의미.
- ❖ 데이터 분석은 문제의 발견, 데이터 수집 및 가공, 데이터 분석 및 액션의 과정을 통하여 이루어짐.
- ❖ 데이터 분석이란 데이터로부터 의사결정을 위해서 데이터를 모으고 변환하며, 분류하고, 그래서 특징을 찾아내어 한눈에 파악할 수 있도록 시각적인 표현을 해 주는 모든 행위를 의미.



## ▶ 데이터 분석과 데이터 관리 기술

### ■ 데이터 분석

- ❖ 데이터 분석은 인사이트를 도출하기 위해서 알고리즘과 수학적 처리과정을 적용하여 해당 정보에 대한 결론을 도출하고 패턴을 찾기 위한 목적으로 원시 데이터를 다루는 과학

복잡한 패턴을 다루는  
딥러닝 알고리즘

+

대규모 연산이 가능한  
병렬연산 기술

+

다양한 분야에서 측정  
방대한 데이터

다양한 분야에서 활용



자율주행 무인 차



가사도우미 로봇



인공지능 비서

## ▲ 데이터 분석과 데이터 관리 기술

### ■ 데이터 관리

❖ 빅데이터 수집, 저장, 분석, 가시화 관리 메카니즘

Data 소스	수집	저장	처리	분석	표현
내부 데이터	수동	정형, 비정형	일괄처리	전 처리	차트
외부 데이터	자동	데이터 등급	실시간 & 배치	분석방법	도표
미디어		장치(CD 등)	분산 병렬 처리		분포도
					디스플레이

# ▶ 데이터 분석과 데이터 관리 기술

## ■ 데이터 관리

### ❖ 데이터 관리 체계 구조

#### 빅데이터 운용

##### 빅데이터 수집 및 정제

- 데이터 추출
- 데이터 변환
- 데이터 로딩

##### 빅데이터 저장 및 관리

- NoSQL
- 확장성 RDBMS
- 분산파일시스템

##### 빅데이터 수집 및 정제

- 분산 컴퓨팅
- MAP REDUCE

##### 빅데이터 수집 및 정제

- 탐구 요인 분석
- 확인 요인 분석

##### 빅데이터 수집 및 정제

- 표
- 그래프

사용자 측면의 빅데이터 플랫폼 요소

제공자 측면의 빅데이터 플랫폼 요소

문제해결 환경

사용자 인터페이스

성능, 비용, 보안, 자원, 품질 관리

## ▶ 데이터 분석과 데이터 관리 기술

### ■ 데이터 관리

#### ❖ 데이터를 통한 지능 자동화를 목표

##### 지능 자동화

지능 시스템을 기반으로 한 전문가 시스템 및 의사결정 지원 시스템 개발 연구

딥 러닝을 포함한 우수한 성능을 가지는 Machine 러닝 기법 관련 연구

다양한 Machine 러닝 기법 간의 융합을 통한 장점 극대화 및 지능형 자동화 시스템 구축연구

Machine – Human Cooperation 기반으로 한 대규모 통합 시스템 관련 연구

##### 데이터 기반 지능 자동화

###### 인공지능

분류, 회귀, 군집화 등의 기법을 이용한 지능 시스템 기반 연구

###### 고장진단 및 예측

인공지능 기법을 이용한 건전도 평가, 고장진단 및 예측

###### 로보틱스

로봇의 기구학적, 동역학적 해석을 통한 기반 기술 구현

###### 물류 자동화

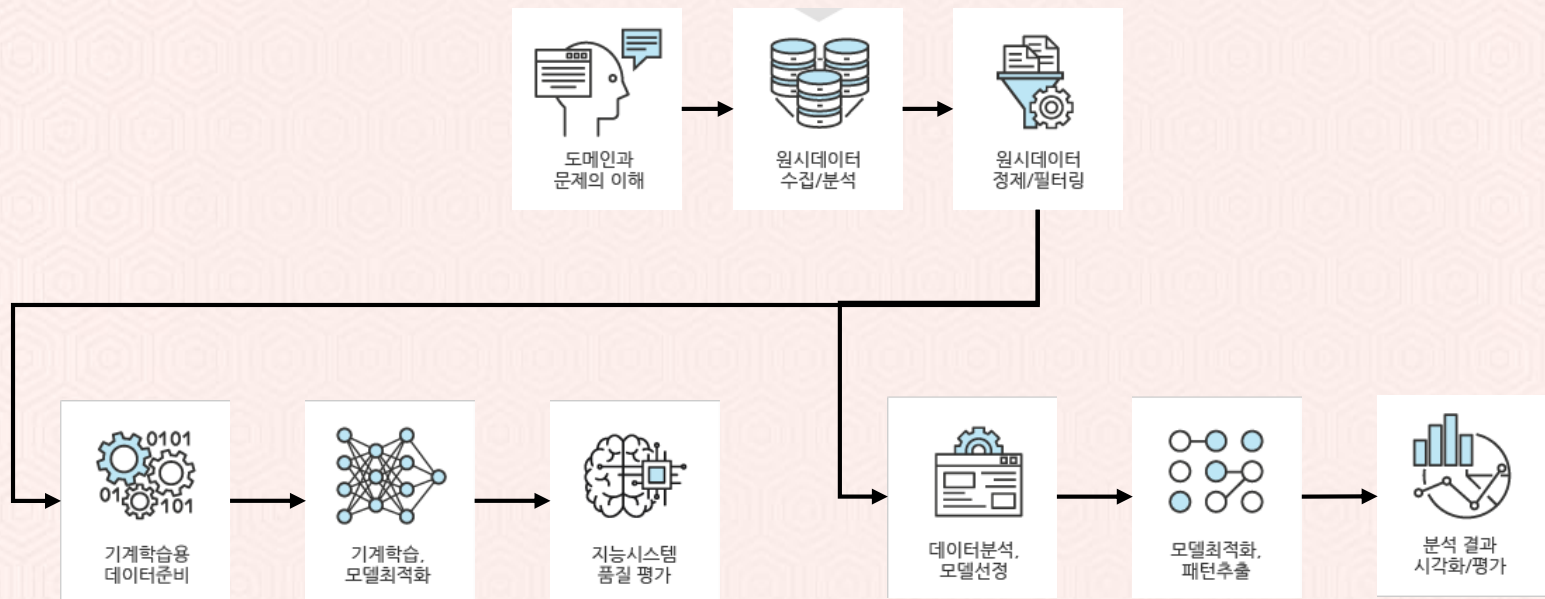
AGV, UAV 등 첨단 기기를 이용한 다양한 형태의 물류 관련 응용기술 연구

###### Internet of Things (IoT)

센서 네트워크 기반 실시간 데이터 전송 기술 및 환경 구축에 대한 연구

## ▶ 데이터 분석과 데이터 관리 기술

### ■ 데이터 과학, 빅데이터, 데이터 분석 관계





## ▲ 데이터 사이언스 요소 기술

### ■ 빅데이터 요소 기술 구성 및 분류

요소기술	설명	해당 기술
빅데이터 수집	조직내부와 외부의 분산된 여러 데이터 소스로부터 필요로 하는 데이터를 검색하여 수동 또는 자동으로 수집하는 과정과 관련된 기술로 단순 데이터 확보가 아닌 검색/수집/변환을 통해 정제된 데이터를 확보하는 기술	ETL / 크롤링엔진 / 로그 수집기 / 센싱 / RSS, Open API 등
빅데이터 공유	서로 다른 시스템간의 데이터 공유	멀티 테넌트 데이터 공유 / 협업 필터링 등
빅데이터 저장	작은 데이터라도 모두 저장하여 실시간으로 저렴하게 데이터를 처리하고, 처리된 데이터를 더 빠르고 쉽게 분석하여 이를 비즈니스 의사 결정에 이용하는 기술	병렬 DBMS/하둡(Hadoop) /NoSQL 등
빅데이터 처리	엄청난 양의 데이터의 저장·수집·관리·유통·분석을 처리하는 일련의 기술	실시간 처리 / 분산 병렬 처리 / 인-메모리 처리 / 인-데이터베이스 처리
빅데이터 분석	데이터를 효율적으로 정확하게 분석하여 비즈니스 등의 영역에 적용하기 위한 기술로 이미 여러 영역에서 활용해 온 기술	통계 분석 / 데이터 마이닝 / 텍스트 마이닝 / 예측 분석 / 최적화 / 평판분석 / 소셜네트워크 분석 등
빅데이터 시각화	자료를 시각적으로 묘사하는 학문으로 빅데이터는 기존의 단순 선형적 구조의 방식으로 표현하기 힘들기 때문에 빅데이터 시각화 기술이 필수적임	시간시각화 / 분포시각화 / 관계시각화 / 비교시각화 / 공간시각화 / 인포그래픽

## ▲ 데이터 사이언스 요소 기술

### ■ 빅데이터 요소 기술 구성 및 분류

- ❖ 데이터 엔지니어 (Data Engineer): 데이터를 저장하고 관리, SQL/자바/스칼라/파이썬
- ❖ 데이터 분석가 (Data Analyst): 탐색적 데이터 분석, 스프레드시트/SQL/상용 대쉬보드 도구
- ❖ 데이터 과학자 (Data Scientist): 통계학 기반 실험과 통계 모형 개발, 자료 추출 SQL, 실험과 모형 R/파이썬
- ❖ 기계학습 데이터 과학자 (Machine Learning Scientist): 기계학습 모형 개발, R/파이썬