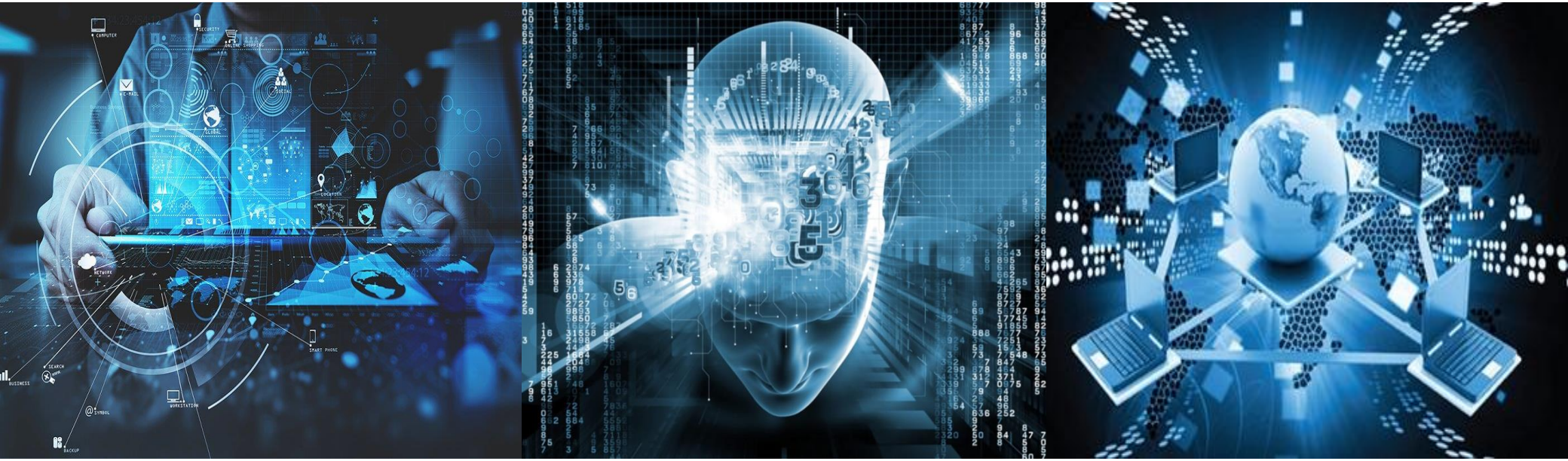
JB Financial Group

비파이 크래프트
Find Information to design valuable society



# 데이터분석(DS) 과정 개요

## 교육 과정 소개

IT Competency Improvement Training

Kim Jin Soo

# Profile

경북대학교 전자공학과 (시스템공학) 졸업
경북대학교 전자공학과 (정보통신) 석사 졸업
고려사이버대 디지털경영학과

## 경력 및 자격

前 **대우정보시스템** 기술연구소 선임연구원
前 **SBS 미디어넷** 기술개발CP 개발팀장
前 E-Biz 기업솔루션 전문벤처기업 기술연구소장
前 기업통합보안솔루션 벤처기업 CTO
前 산업보안전문가포럼 회장
前 경기창조경제혁신센터 **경기콘텐츠진흥원** CSP

**빅데이터기술전문가(Bigdata Technical Expert)**
**산업보안전문가(Industrial Security Professional)**

빅파이 크래프트
Find Information to design valuable society

# 김 진 수
## CEO / DataActionist

現 서울T직업전문학교 빅데이터 전문강사
現 멀티캠퍼스 IT기술교육강사 & 수석멘토
現 한국경제신문 빅데이터센터 전문위원
現 한국SW기술진흥협회 기술위원
現 경기문화창조허브 스타트업플래너

現 INNOTURN 파트너즈 대표컨설턴트
現 중기부 중소기업지원단 현장클리닉위원
現 고용노동부 일터혁신컨설팅 컨설턴트
現 산인공 NCS/일학습병행제 컨설턴트
現 서울시 정보통신분과 청년취업 멘토

# 생각해 봅시다!!

- 정보화 시대, 데이터와 정보의 차이?
- 컴퓨터가 바라본 데이터의 조건?
- 빅데이터의 서막, 오픈소스진영의 반란?
- 국내 빅데이터 시장의 현황
- 기술을 알면 보이는 것도 달라진다
- 소프트웨어를 공부해야만 하는 이유

# 빅데이터 - Word Cloud

# 데이터의 이해

## 정보화 시대 → 빅데이터 시대 !!

- 우리가 다루는 것은?  데이터는 무엇인가?
- 도대체 무엇을 수집하고, 어떻게 표현하고,
  관리는 왜 해야하고, 활용은 어떤 방식으로 해야할까?

# 데이터의 이해

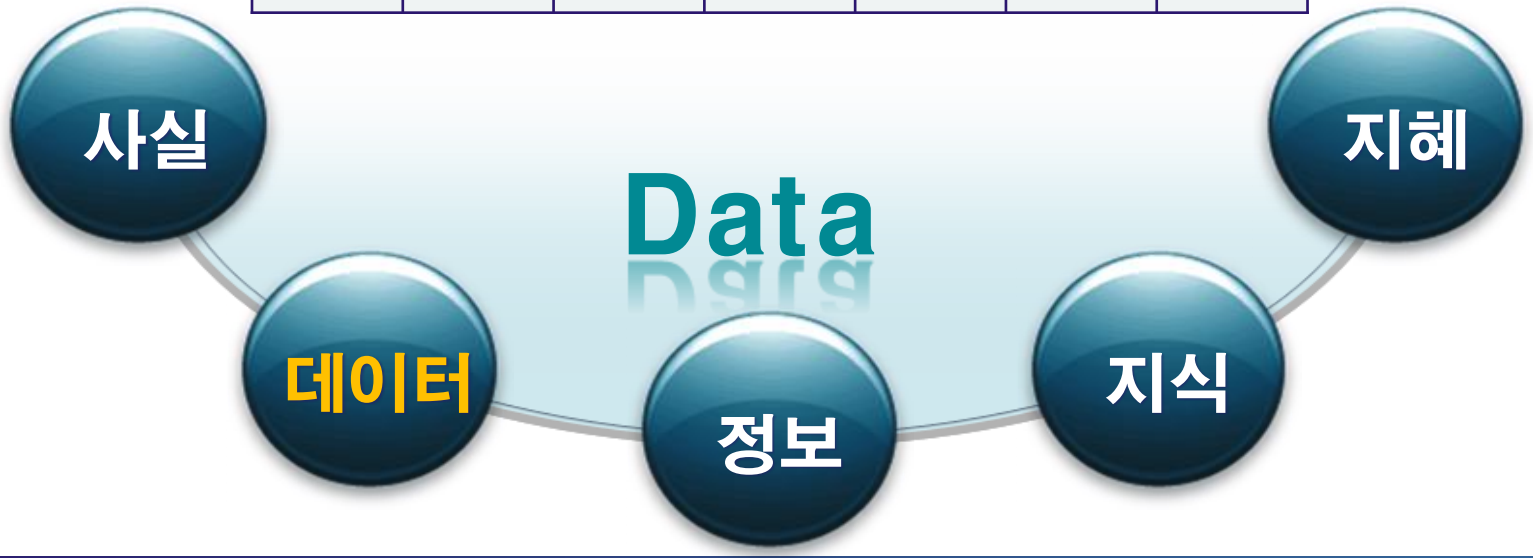
## 사실 : 객관적 실재

- 전현무가 롯데리아에서 빅불세트를 먹었다.
- 전지현은 순살치킨팩을 시켜 먹었다.

## 데이터: 객관적 실재의 체계화

- 사실을 특정한 방식을 통해 체계화

| 고객명 | 일자 | 주문 | 음료 | 사이드 | 가격 | 비고 |
|---|---|---|---|---|---|---|
| 전현무 | 22.04.01 | 빅불세트 | 콜라 | 양념감자 | 7,500원 | 리필1번 |
| 전지현 | 22.04.02 | 순살치킨 | 커피 | 치즈스틱 | 6,800원 | - |
| - | - | - | - | - | - | - |



# 데이터의 이해

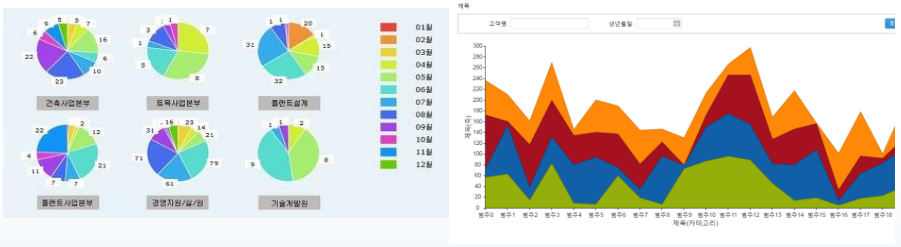
## 정보 : 데이터 + 의미

- 데이터를 특정 목적과 문제 해결에 도움이 되도록 가공한 것
- 전현무는 4월에 3번, 전지현은 4월에 처음 방문

# 데이터의 이해

## 지식 : 정보 + 가치

- 정보를 집적하고 체계화하여 장래의 일반적 사용에 대한 보편성 확보
- 전년도 8월 매출이 최고, 11월 매출이 최저...



# 데이터의 이해

## 지혜 : 지식 + 추론

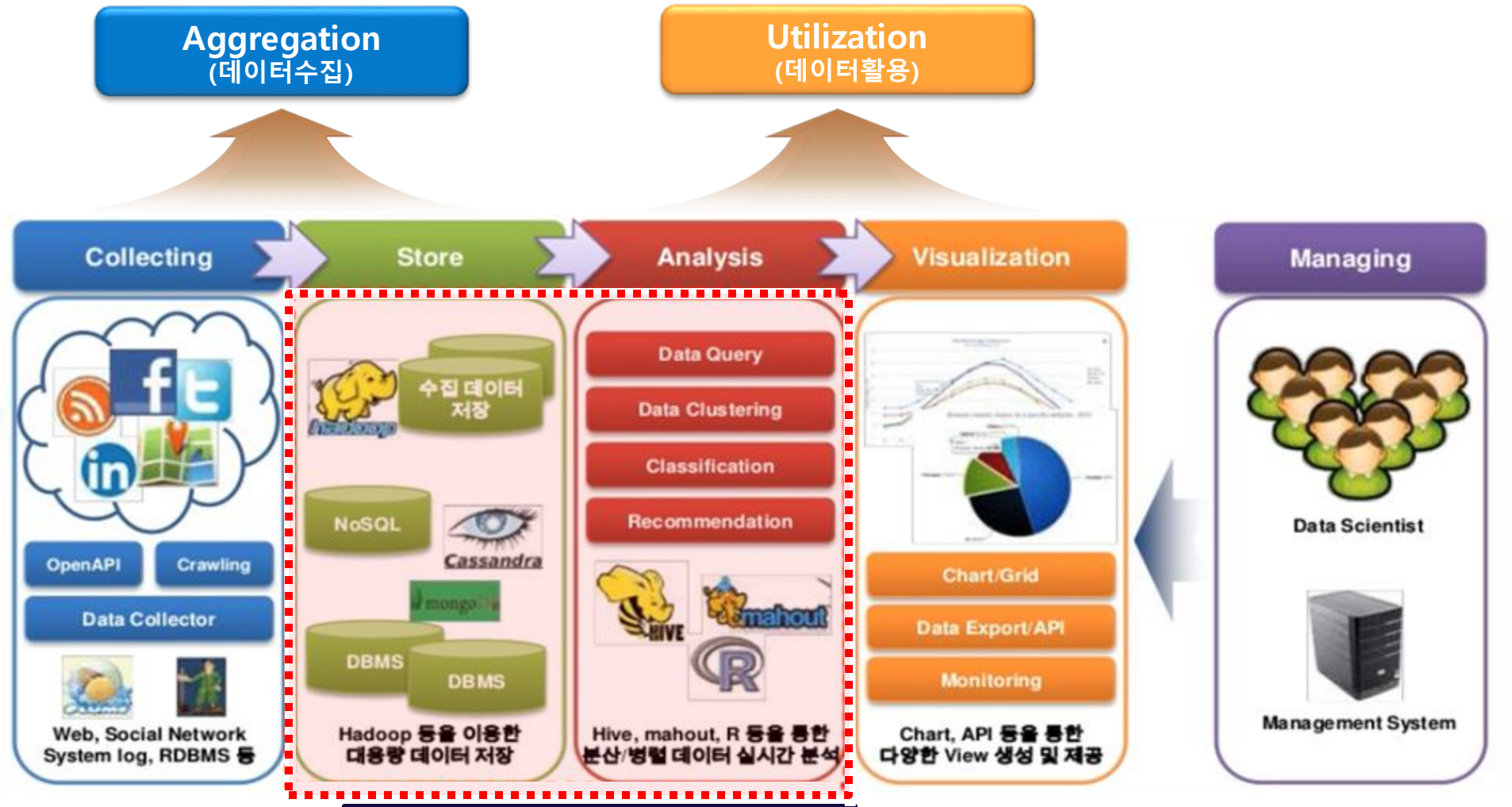
- 추론 및 문제해결을 위한 지적 능력
- 이달의 추천메뉴, 모바일 주문 서비스 이벤트



# 컴퓨터가 바라본 데이터는 조건

- **정보화시대에서 데이터의 요건?**
  - 통합, 저장, 운영, 공유
- **효율적 관리가 중요**
  - 성능과 비용적 측면
- **IT분야를 주도하는 글로벌 기업의 특징**
  - OS와 DBMS

# 빅데이터 처리 4단계




빅데이터 요소기술이 투입됨

BigData Solution의 기능 및 처리 흐름과 관리구조

# 빅데이터의 서막, 오픈소스진영의 반란

- 2005년 이전의 일반적인 데이터 관리, 대용량데이터
- 고가가 아닌 저가 시스템, 중앙집중방식이 아닌 분산처리방식
- Hadoop 기술의 등장
- 빅데이터 생태계(EcoSystem)를 이루는 다양한 기술들
- CAP이론

  → Pick Two : Consistency, Availability, Partition Tolerance

# BigData Landscape

# 데이터과학, Data Science

WIKIPEDIA
The Free Encyclopedia

## Data science

From Wikipedia, the free encyclopedia

*Not to be confused with information science.*

**Data science**, also known as **data-driven science**, is an interdisciplinary field about scientific processes and systems to extract knowledge or insights from data in various forms, either structured or unstructured,[1][2] which is a continuation of some of the data analysis fields such as statistics, machine learning, data mining, and predictive analytics,[3] similar to Knowledge Discovery in Databases (KDD).

Turing award winner Jim Gray imagined data science as a "fourth paradigm" of science (empirical, theoretical, computational and now data-driven) and asserted that "everything about science is changing because of the impact of information technology" and the data deluge.[4][5]

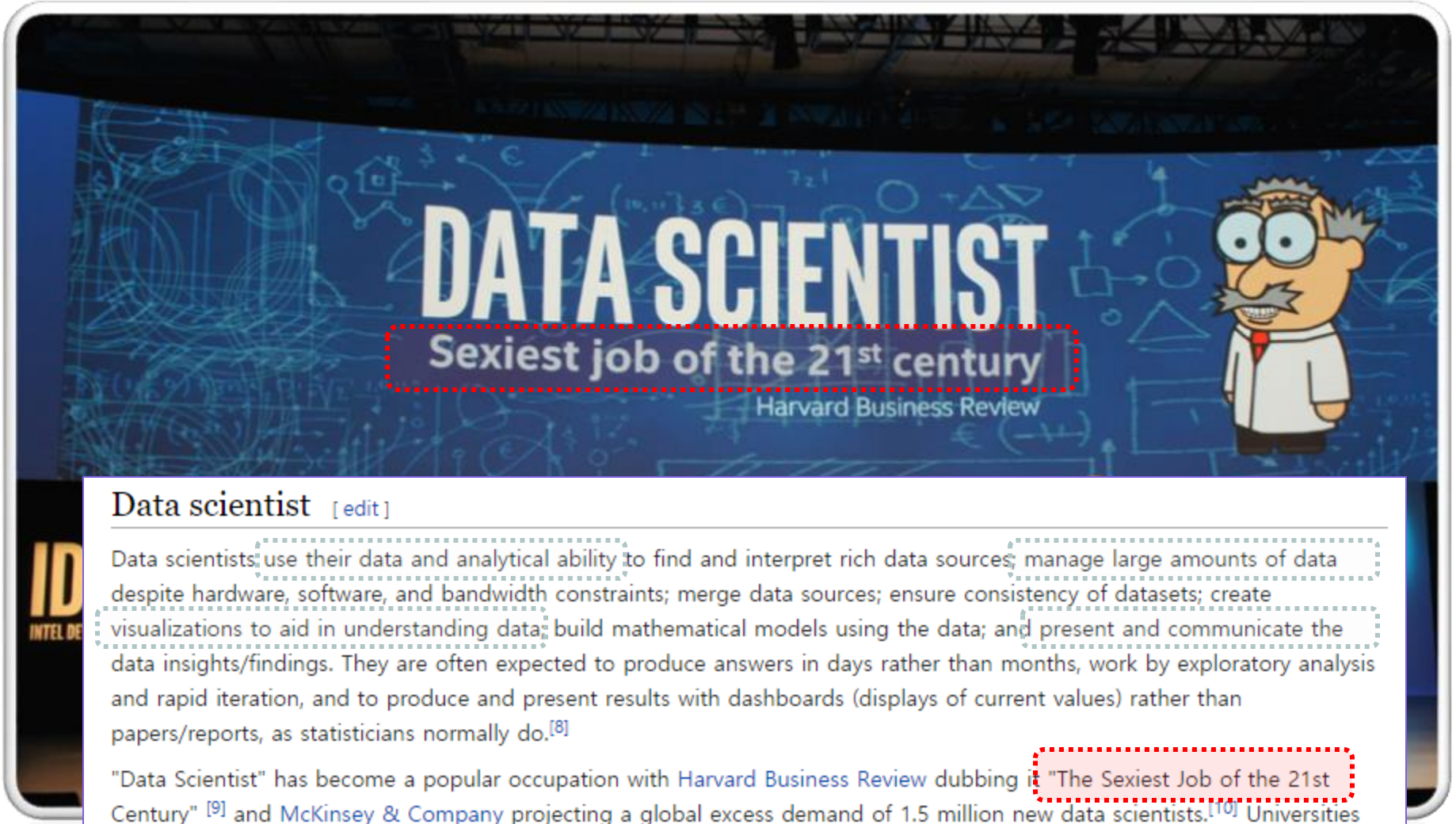
데이터 사이언스(data science)이란, 데이터와 관련된 연구를 하는 학문이다.

데이터의 구체적인 내용이 아닌 서로 다른 성질의 내용이나 형식의 데이터에 공통으로 존재하는 성질, 또는 그것들을 다루기 위한 기술의 개발에 착안점을 둔다는 특징을 가진다.

사용되는 기술은 여러분야에 걸쳐있으며 수학, 통계학, 계산기과학, 정보공학, 패턴인식, 기계학습, 데이터마이닝, 데이터베이스 등과 관련이 있다.

데이터 사이언스는 생물학, 의학, 공학, 사회학, 인문과학 등의 여러 분야에 응용되고 있다.

# 빅데이터 전문가, Data Scientist

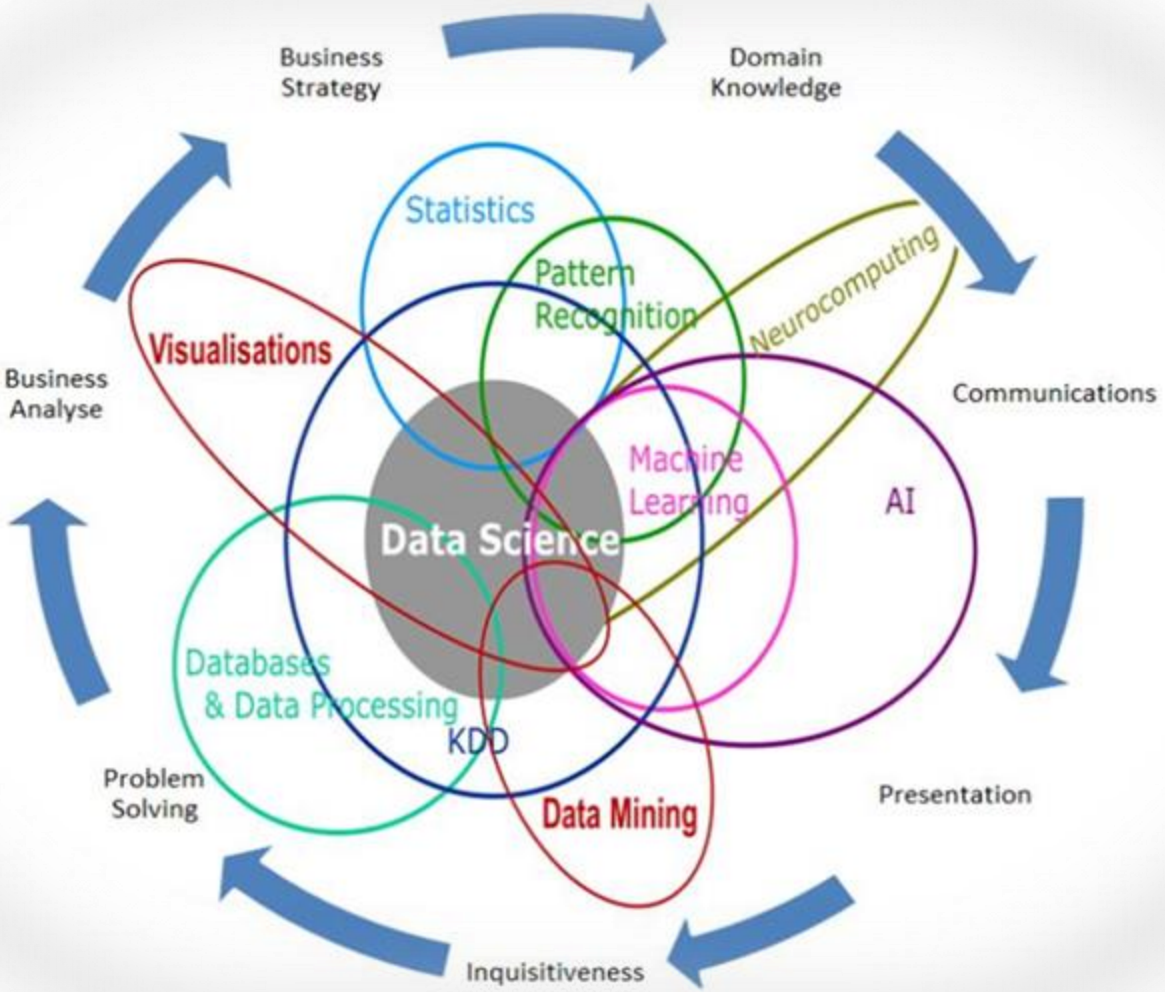



## Data scientist [ edit ]

Data scientists use their data and analytical ability to find and interpret rich data sources; manage large amounts of data despite hardware, software, and bandwidth constraints; merge data sources; ensure consistency of datasets; create visualizations to aid in understanding data; build mathematical models using the data; and present and communicate the data insights/findings. They are often expected to produce answers in days rather than months, work by exploratory analysis and rapid iteration, and to produce and present results with dashboards (displays of current values) rather than papers/reports, as statisticians normally do.[8]

"Data Scientist" has become a popular occupation with Harvard Business Review dubbing it "The Sexiest Job of the 21st Century" [9] and McKinsey & Company projecting a global excess demand of 1.5 million new data scientists.[10] Universities are offering masters courses in data science.[11] Shorter private bootcamps are also offering data science certificates including student-paid programs like General Assembly to employer-paid programs like The Data Incubator.[12]

# 데이터과학의 관심분야


Business Strategy
Domain Knowledge
Communications
Presentation
Inquisitiveness
Problem Solving
Business Analyse
Statistics
Pattern Recognition
Neurocomputing
Visualisations
Machine Learning
AI
Data Science
Databases & Data Processing
KDD
Data Mining

# 국내 빅데이터 시장의 현황

❖ **국가 R&D 과제를 보면 짐작할 수 있다.**

❖ **국가 정책의 이해**

- 기술과제선정,
- 사업화,
- 활성화(비즈니스 확산),
- 표준화,
- 제도적 보완

❖ **국가/학계/기업 합작 성공 사례**

- WiBro

# 기술을 알면 보이는 것도 달라진다

❖ 통신시장의 주도권 싸움

❖ 3G, 4G, 5G 의미

❖ 4차산업혁명시대 트랜드기술

- IoT, 사물통신
- Cloud, 클라우드
- BigData, 빅데이터
- Mobile, 모바일
- Social, 소셜서비스

❖ 플랫폼을 주도하는 회사가 앞서간다

# 소프트웨어를 공부해야만 하는 5가지 이유

- ❖ **데이터를 내 마음대로 다룰 줄 알아야 한다.**
  - 데이터를 핸들링 하는 사람이 핵심.
- ❖ **SW교육 의무화 시대, 코딩은 점점 일상 속으로~**
- ❖ **하드웨어와 통합개발환경의 개선, 데이터 접근 용이**
- ❖ **전문가 영역 vs 분석가 영역**
- ❖ **알면 소통할 수가 있다.**

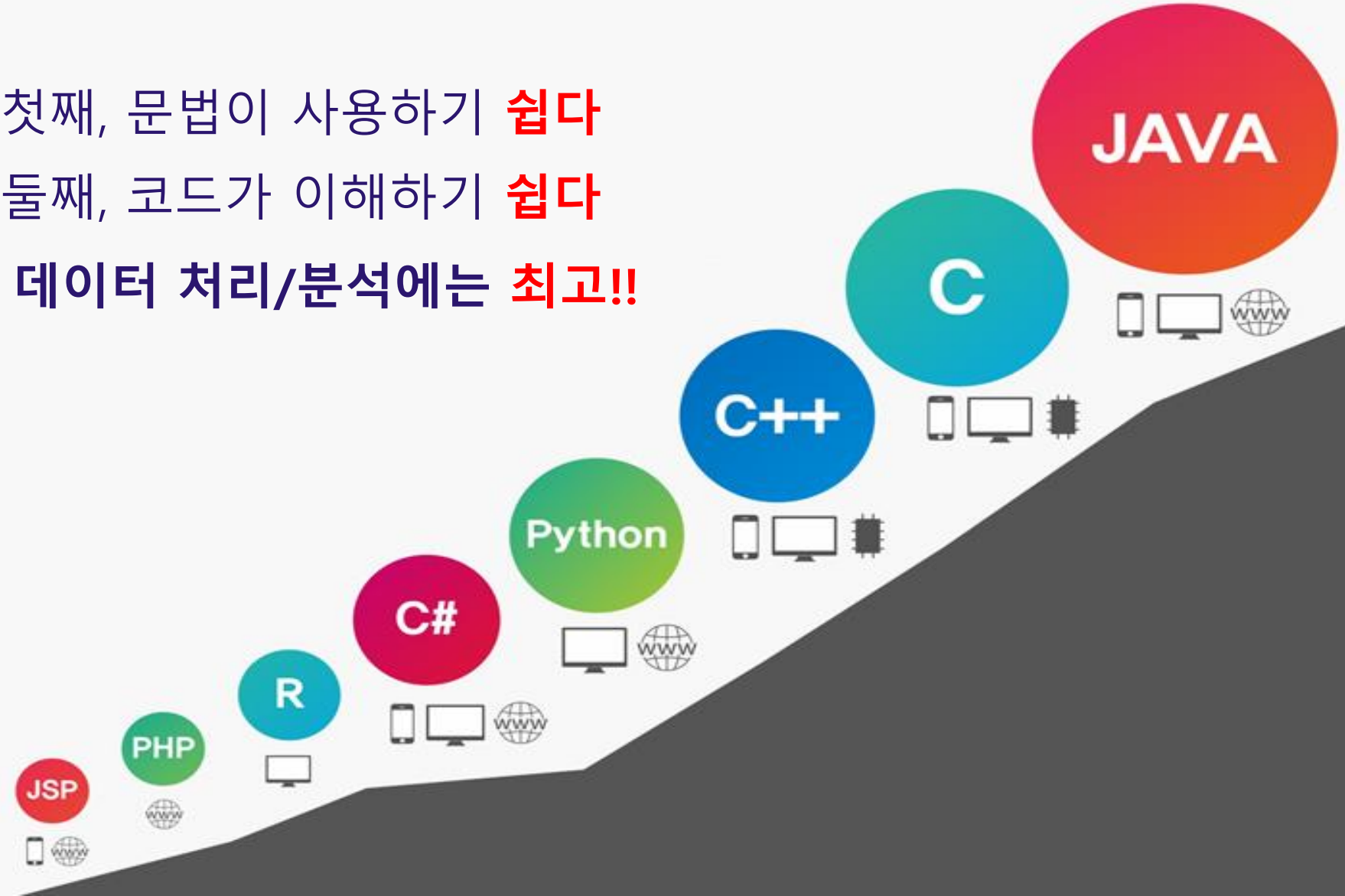
# BigData ♥ Software

# 수많은 소프트웨어, 왜 파이썬을...

첫째, 문법이 사용하기 **쉽다**

둘째, 코드가 이해하기 **쉽다**

∵ **데이터 처리/분석에는 최고!!**



# 파이썬의 특징

## 파이썬의 철학

- Beautiful is better than ugly
- Explicit is better than implicit
- Simple is better than complex

## 오픈소스

- Battery included. 이미 만들어진 라이브러리 이용
- 많은 Library 지원, 많은 Platform 지원

## 높은 생산성

- Life is too short, You need Python
- Prototype → Product

# 파이썬 특강 커리큘럼

- **파이썬 입문** **PyCharm (Integrated Development Environment, IDE)**
  - 코딩, 디버그, 컴파일, 배포 등 프로그램 개발에 관련된 모든 작업
- **파이썬 고급** **Jupyter Notebook(Powerful Interactive Computing )**
  - 대화형 컴퓨팅, 데이터 처리 및 분석 관련 라이브러리 활용

| Session 1 | Session 2 |
|---|---|
| 1 파이썬 개요 : SW이해, 환경설정 | 1 파이썬 리뷰 : Data, File I/O |
| 2 데이터 이해 : Data 종류 및 구조체 | 2 객체지향PG : Function, Class |
| 3 제어문 학습 : 조건, 반복, 제어문 | 3 통계기반PG : Numpy, Pandas |
| 4 함수의 활용 : 함수의 생성 및 활용 | 4 시각화 모듈 : Matplotlib, Seaborn |
| 5 객체와 모듈 : OOP, 클래스 이해 | 5 데이터 분석 : Kaggle, Data Portal |

# 효율적인 학습을 위한 퀵가이드

## Be Data Actionist !!

### 1단계 익숙해지기

- SW/PG
- IDE/Tool
  - ✓ Coding
  - ✓ Debugging
  - ✓ Compile
  - ✓ Deploy

### 2단계 데이터 이해

- 데이터 종류
- 데이터 구조체
  - ✓ LIST
  - ✓ TUPLE
  - ✓ SET
  - ✓ DICT

### 3단계 데이터 처리

- 라이브러리
- 데이터 전처리
  - ✓ Indexing
  - ✓ Slicing
  - ✓ Filtering
  - ✓ Sorting

### 4단계 데이터 활용

- 데이터 크롤링
- Open Data
  - ✓ Public
  - ✓ Social
  - ✓ Script
  - ✓ Web/App

# 파이썬 특강 – 심화과정 (관심분야)

## ❖ 웹프로그래밍 기본

- 장고(Django) : 파이썬으로 작성된 오픈 소스 웹프레임워크
- 모델-뷰-컨트롤러(MVC) 패턴기반의 웹 애플리케이션 제작

## ❖ 인공지능 구현을 위한 프로그래밍

- TensorFlow : AI 프로그래밍을 위한 오픈소스 소프트웨어 라이브러리
- 머신러닝의 메커니즘 이해 및 최적화 기법
- 딥러닝 모델 학습 및 응용

## ❖ 금융데이터 분석 및 활용

- Quantitative Analysis : 기술적분석, 재무재표분석, 모멘텀
- 주식자동매매 프로그래밍 기법

## ❖ 사물인터넷(IoT) 프로그래밍

- 아두이노 및 라즈베리파이(raspberry pi) 학습
- 입출력 부품(센서)을 활용한 설계 및 코딩

# 참고도서


데이터 과학을 위한
파이썬
프로그래밍
Python
Programming
지진 데이터 시각화, 선거와 이름통계 분석 등 실사례 사용
파이썬 라이브러리를 활용한 데이터 분석
Python for
Data Analysis
O'REILLY
파이썬을 이용한
웹크롤링과
데이터 분석
INFINITYBOOKS
matplotlib을 이용한
데이터 시각화 프로그래밍
파이썬으로 구현하는 데이터 분석용 2D & 3D 그래프
파이썬
데이터 분석
실무 테크닉
100
데이터의
보이지 않는 손
휴먼 빅데이터로 밝혀낸 인간·조직·사회의 법칙
1위
행복, 운, 불임까지 데이터로 컨트롤할 수 있다!



김 진 수
CEO, Data Actionist

100-791 서울특별시 중구 청파로 463번지 3F BigData R&D Center
**CP.** 010-5670-3847 **Tel.** 02-360-4047 **Fax.** 02-360-4899
**E-mail.** bigpycraft@gmail.com
http://www.bigpycraft.com

# 감사합니다!