



# AI 융합전문가 4차시 데이터분석 I

2024.09.29..





## 현재 하는 일



소속	주요 업무
(주)제타데이터 대표이사	데이터 분석, 전략 및 컨설팅, 데이터 가치평가 ODA 컨설팅 (Official Development Assistance)
(주)지구파트너스 감사	창업보육, 투자, 기업·기술가치평가, 사업타당성 분석
(주)메타로직 컨설팅 수석	ISP 컨설팅 (Information Strategy Planning) ISMP 컨설팅 (Information System Master Plan)

### ◎ 자격증

- 경영지도사31기 (인적자원, 2016)
- 창업보육매니저 (BI협회, 2018)
- 기업·기술가치평가사 (기업·기술가치평가협회, 2018)
- 기업재난관리사 실무과정 (행정안전부, 2019)
- 데이터분석 준전문가 ADsP (데이터산업진흥원 K-Data, 2021)
- 빅데이터 분석기사 (과학기술정보통신부 · 통계청, 2021)
- 국제공인컨설턴트 CMC (ICMCI, 2023)
- 인공지능(AI) 활용마스터1급 (뉴미디어교육연구소, 2024)





## 데이터 관련 비즈니스

### 한국경제

Hankyung.com  
가정을壇내는 생각

**알고 보면 재미있는 경영이야기 - 기업 경영과 자본**

▲기업원 홍용기 전문위원

누구나 알고 있는 것처럼 기업을 관리하면서 예전 가지는 자본이 필요하다. 그리고 사업으로부터 벌어들인 자본으로 나눌 수 있고, 그들의 차등으로 차등으로 나눌 수 있다.

자본은 우리가 충분히 투자하고 부른 가 분을 모아놓아서 상황으로 부른다. 다른 모든 부분을 종상하고도 남은 자리에 자신이 있다면 그에 대한 뿐만 아니라 가기 속도를加快하는 이유를 찾고, 그리고 빠르게 상황의 변화를 감지하고, 그에 따른 대처방법을 찾을 수 있다.

따라서 기업의 운영 상황을 나타내는 재무회계의 표기방식에 있어서도 이런 논거가 적용된다. 기업의 재무상표를 나타내는 '재무상표'에서도 부채로 순서에 있어서 타인 자본을 더 놓았지만, 의미로 자본보다 부채가 위에서 표시된다.

결국 기간의 경영실력을 나타내는 '순익제산부'의 경우에도 마찬가지다.

마음에서 이익을 산출하는 순서에 대해서는 영업이익에서 타인 자본에 대한 이자비용을 먼저 차감한 이후에 차액으로 순기반 이익을 산출한다.

그럼에 순기반 이익을 계약하는 대표가 혹은 권역주는 타인 자본의 부채에 대해서는 국익의 이자를 지금까지는 차기자본에 대해서 그 비용이 발생하지 않는 것으로 생각에서는 부채는 의의로 많다. 하지만 그렇지도 않다. 기업을 경영하는 관점에서 차기자본에 대해서는 비용이 발생하고 그에 대한 부채 계수 등을 수립해야 민영적인 경영을 할 수 있다.

인장을 바꿔서 차관을 출자하는 주주, 즉 투자자 입장에서 생각해도 출자한 금액에 대해서 타인 자본처럼 경영실과 무관하게 이자와 벌을 수 있고, 경영성이 있어 벌을 수 있는 범위도 확장하게 되어 나아가 회사 분을 모을 경우 투자 벌을 줄수록 더 많은 보상을 받지 않으면서 차기 자본에 투자하는 위험률은 타인 자본에 높다고 할 수 있다.

위험률이 높다고 하는 것은 위험이 높을 때를 요구하는 수익률이 높거나 아니고, 실제로 회사의 자본으로 투자를 했을 때 차관은 것이다. 차관은 차관으로 나누어 주면 회사에서 차관을 수령할 수 있다. 예컨대 영업이익이 3.2%라면 3.2%를 1.6로 나누면 2%를 수령하고, 영업 차입금이 5%라면 2%로 나누어 40억을 수령해 나가게 된다.

또한 영업이익 3.2억을 WACC이 0.08로 나누면 위에서 산출한 범위로 같은 수준인 3.2억을 수령할 수 있는데 영업이익이 영구한 만큼 영업이익에 대한 차익률은 영업의 차익률로 나누어 주고, 일부 신용평가 기관에서는 이런 방식으로 산출되는 차익률을 그 기관의 차익률로 평가기도 한다.

뿐만 아니라, 영업이익 3.2억을 WACC이 0.08로 나누면 위에서 산출한 범위로 같은 수준인 3.2억을 수령할 수 있는데 영업이익이 영구한 만큼 영업의 차익률은 영업의 차익률로 나누어 주고, 일부 신용평가 기관에서는 이런 방식으로 산출되는 차익률을 그 기관의 차익률로 평가된다.

WACC은 타인 자본 비율과 차기 자본 비율을 가중평균하는 방법으로 산출하게 된다. 가중평균기준비율 WACC을 각각의 실증하면, 경영자본은 타인 자본과 차기자본 비율이 각각 4/10은 6/10이고, 타인 자본 비율과 차기자본 비율이 각각 5%, 10%라고 생각할 때,

\* 가중평균 자본비율 WACC  

$$= 0.4 \times 5\% + 0.6 \times 10\% = 8\%$$

WACC은 재무管理水平가 산출하면 됨으로 경영자는 활용도 된다. 다만 가중평균기준비율은 몇몇 이해하면 된다. 나아가 이해하기 어렵고 활용도 어렵다. 그것이 차익률을 계산하는데 있어서 좋았지만, 그것과는 별도로 일부 신용평가 기관에서는 이런 식으로 기관차익률을 평가하는데 사실, 순기반 차익률을 명식화시킨다.

한경 경영지원단  
<http://clean.hankyung.com>

## ◎ 정보화전략계획수립(ISP) 컨설팅 수행

- |              |                   |
|--------------|-------------------|
| 20.05~20.08. | 창업진흥원             |
| 20.10~20.12. | 한국연구재단            |
| 21.01~21.04. | 소상공인시장진흥공단        |
| 21.11~22.06. | 서울특별시             |
| 22.08~22.12. | 경찰대학교             |
| 23.08~24.05  | ODA (요르단 경찰청 PSD) |

## ◎ 2022년 AI학습용데이터 구축사업 평가

- |    |  |
|----|--|
| 1차 | 08 방송 콘텐츠 대화체 음성인식 데이터<br>09 방송 콘텐츠 한국어·영어 통번역 데이터<br>43 갑각류 종자생산 데이터<br>48 식생 탄소 포집량 식별 데이터 |
| 2차 | 74 축산 기자재(소, 돼지) 3D 데이터<br>75 소(한우, 젖소) 및 돼지 발정행동 데이터  |
| 3차 | 06 인공지능 신기술 선도(자유 공모)  |

## ◎ 데이터 가치평가 컨설팅

- 23.09~23.11 중소벤처기업진흥공단





## 데이터 관련 강의

- 데이터분석 및 실전 R코딩 (경영기술지도사회, 빅데이터 분석기사 자격증 취득 과정)
- 데이터분석 Python 심화과정 (서울 여성능력개발원 강동 여성인력개발센터 / 용산 여성인력개발센터)
- 파이썬 코딩을 통한 크롤링 자동화 인텐시브 과정 (경영지도사 및 컨설턴트)
- AI & ChatGPT 활용 및 데이터분석 컨설팅 방법론 (경영기술지도사회, 국제공인컨설턴트 CMC 양성과정)
- AI & 데이터분석 (매경아카데미, 동북아 ICT 포럼)

HOME > City&Global

### 홍용기 컨설팅학박사, 한국경영기술지도사회 '빅데이터 분석 실무 과정' 성료

김현진 기자 | ○ 승인 2023.03.27 16:54 | ○ 댓글 0

경영컨설팅 과정에서 데이터 분석을 결합한 객관적인 사업계획수립, 마케팅 전략 제시 역량을 갖추고자



**주제타데이터 홍용기 대표, 강동여성인력개발센터/강동여성새로일하기센터의 '빅데이터분석 파이썬(Python) 심화 과정' 성료**

윤성임 | ○ 승인 2023.06.05 11:53 | ○ 댓글 0

『2023 경력단절여성대상 직업교육훈련 빅데이터분석 전문가 양성과정』의 한 부분으로 빅데이터분석기사(BAE) 자격증 취득 의지 고취 및 이어지는 RVA/업무자동화 과정에 대한 기대감 높아

(주)제타데이터 홍용기 대표이사(컨설팅학박사/경영지도사)는 지난 5월 23일부터 6월 1일까지 서울시 강동여성인력개발센터/강동여성새로일하기센터에서 진행한 '빅데이터분석에 필요한 파이썬(Python) 심화' 과정을 진행하였다.



**주제타데이터 홍용기 대표, 용산여성인력개발센터의 '빅데이터분석을 위한 파이썬(Python) 심화 과정' 성료**

윤성임 | ○ 승인 2023.08.10 19:51 | ○ 댓글 0

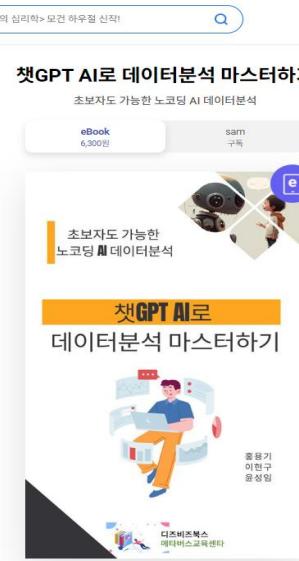
(주)제타데이터 홍용기 대표는 지난 7월 13일부터 7월 26일까지 용산 전자랜드에서 빅데이터분석을 위한 파이썬(Python) 심화과정 강의를 진행했다.



# 강사 홍용기 컨설팅학 박사



# 책 쓰기 프로젝트



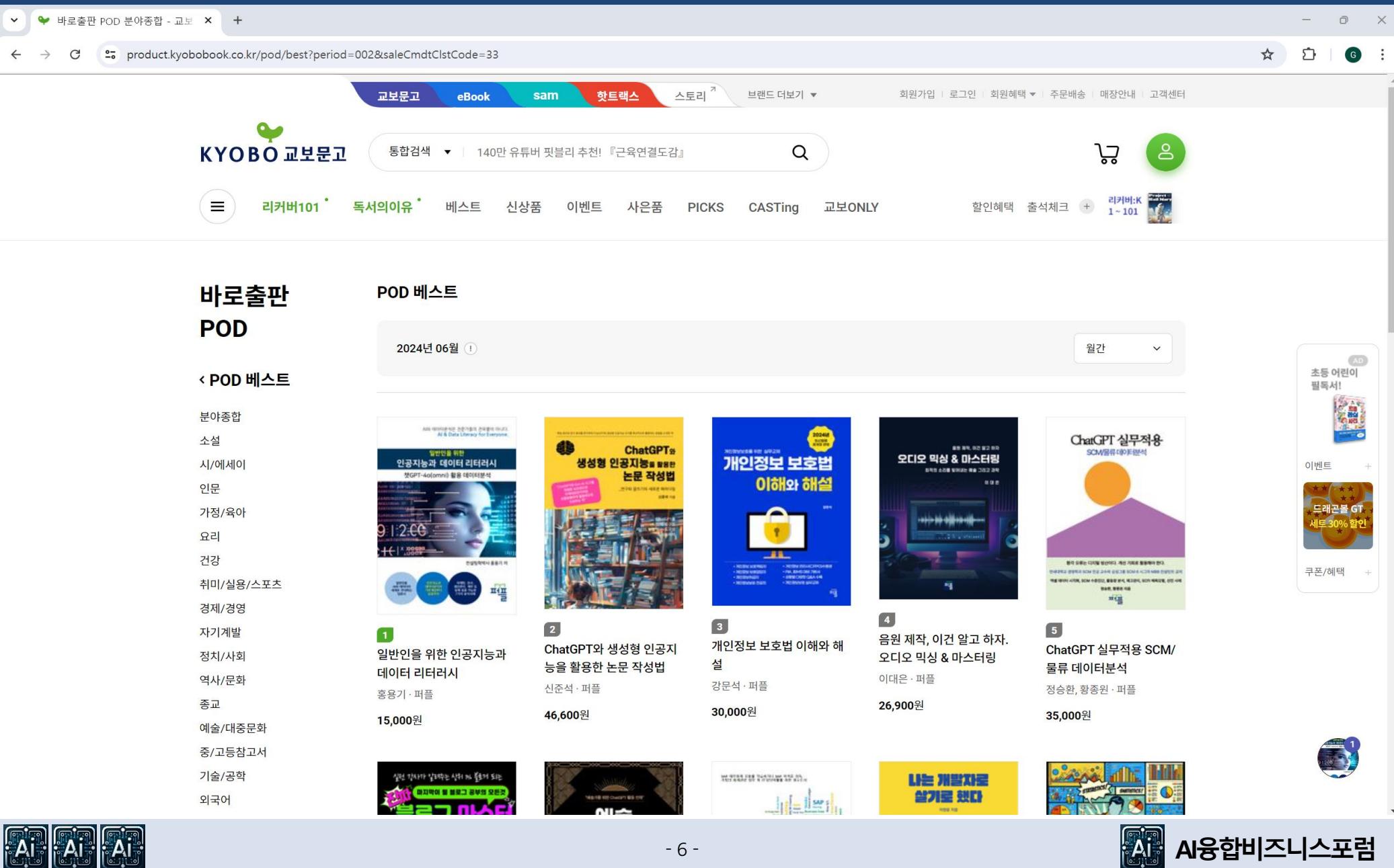
2023년 5월



2024년 4월



## 강사 홍용기 컨설팅학 박사



혹자는  
인생 뭐 별거 있냐고 할테지만...

인생은 어쩌면  
수많은 연결고리들의 집합일지도...

현재를 살면서 긴장을 늦출 수 없는 이유는  
이 연결고리가 앞으로 어떻게 연결될지 지금은 알 수 없기 때문이다.

대체 불가능한 사람으로 살아남기 위해 해야 할 것

# What to do?



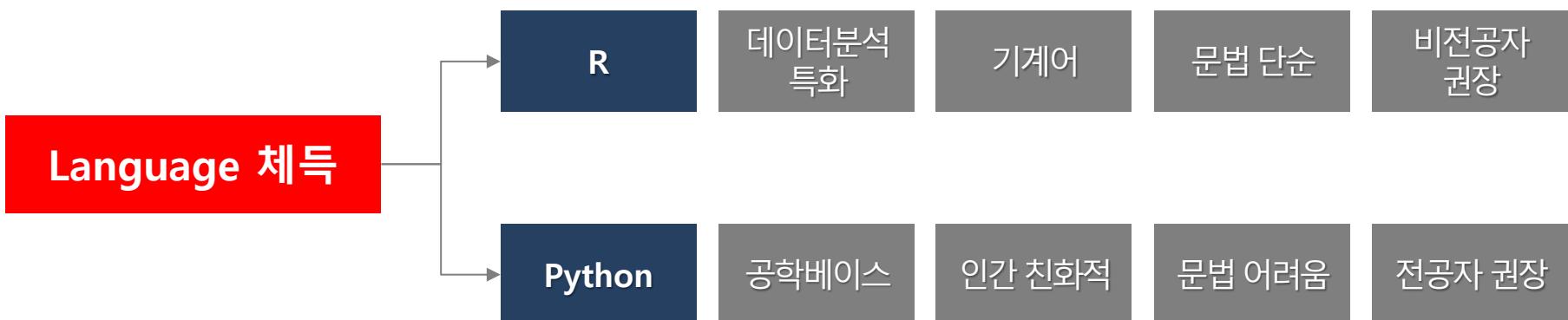
혹자는  
인생 뭐 별거 있냐고 할테지만...

인생은 어쩌면  
수많은 연결고리들의 집합일지도...

현재를 살면서 긴장을 늦출 수 없는 이유는  
이 연결고리가 앞으로 어떻게 연결될지 지금은 알 수 없기 때문이다.

대체 불가능한 사람으로 살아남기 위해 해야 할 것

# How to do?



생각해 봅시다...



## 기술의 가치

“가치를 아는 사람에게 기술이 가야 빛을 발한다.”

www



가족끼리 암호를 정해 두셨나요?

KBS  
NEWS

KBS  
NEWS

# 진짜같은 가짜, 딥페이크 직접 제작해보니





*Like every great presentation, I've divided my talk into three subjects. Steve Jobs -*

I .

---

**Data &  
Information**

II .

---

Database &  
Schema

III .

---

Data  
Gathering



# DATA

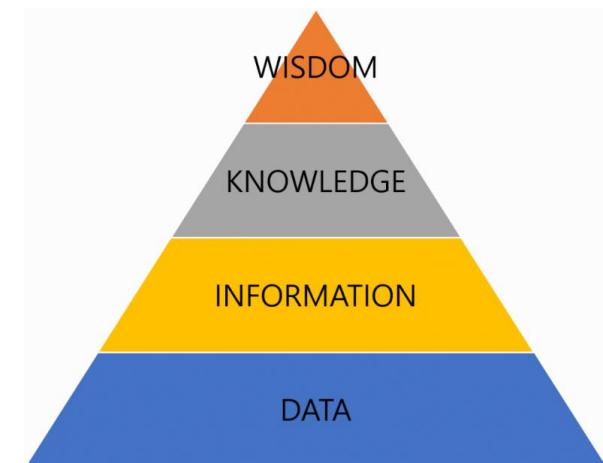
**‘추론과 추정의 근거를 이루는 사실’**

(A thing given or granted; something known or assumed as fact and made the basis of reasoning or calculation; an assumption of premiss from which inferences are drawn. OED, Vol. IV 264)



# Data vs Information

피라미드 요소	설명
지혜 (wisdom)	<ul style="list-style-type: none"><li>● 근본 원리에 대한 깊은 이해를 바탕으로 도출되는 창의적 아이디어</li><li>● 상황이나 맥락에 맞게 규칙을 적용하는 요소</li></ul>
지식 (knowledge)	<ul style="list-style-type: none"><li>● 다양한 정보를 구조화하여 유의미한 정보로 분류하고 일반화 시킨 결과물</li><li>● 정보에 기반해 찾아진 규칙</li></ul>
정보 (information)	<ul style="list-style-type: none"><li>● 여러가지 데이터 중에 사용자에게 필요한 데이터</li><li>● 사용자의 필요에 따라 정제되거나 가공된 데이터를 정보라고 부름</li></ul>
데이터 (data)	<ul style="list-style-type: none"><li>● 객관적 사실로서 가공하기 전의 순수한 자료</li><li>● 가공되지 않은 데이터는 그 자체로는 의미를 지니기 어려움</li></ul>



- ✓ 전국의 지역별로 조사한 매 시간별 기온 측정자료 → 데이터
  - ✓ 어떤 목적에 의해 이 데이터를 기반으로 월별 또는 계절별 평균 기온을 산출 → 정보

ex. 어떤 지역의 온도에 따라 혹은 일교차에 따라 어떤 지역에서 어떤 과일을 재배하겠다는 등의 의사결정에 활용

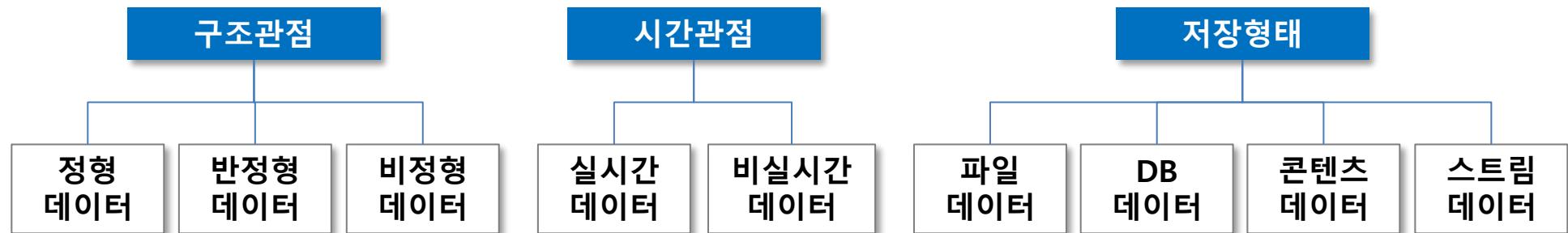


# ‘데이터분석’이라고 하면…

어떠한 데이터를 어느 범위까지  
분석해야 하는 것인가?



# Type of Data

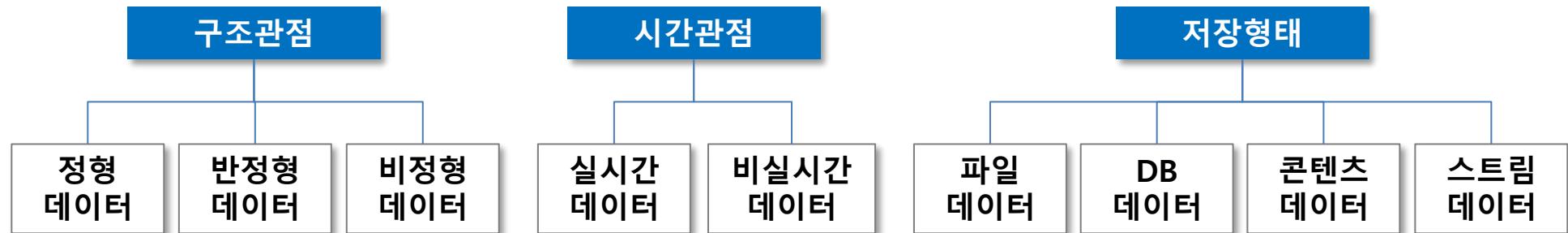


구조관점의 데이터 유형
-----------------

유형	설명	종류
정형 데이터	정형화된 스키마(형태) 구조 기반으로 행과 열로 구성	관계형 데이터베이스, 스프레드 시트
반정형 데이터	스키마 구조 기반이나 값과 형식에서 일관성을 가지지 않는 데이터	XML, HTML, 로그데이터, JSON, 센서데이터
비정형 데이터	스키마 구조 형태를 가지지 않으며 고정된 필드에 저장되지 않는 데이터	SNS, 게시판, 텍스트, 이미지, 오디오/비디오



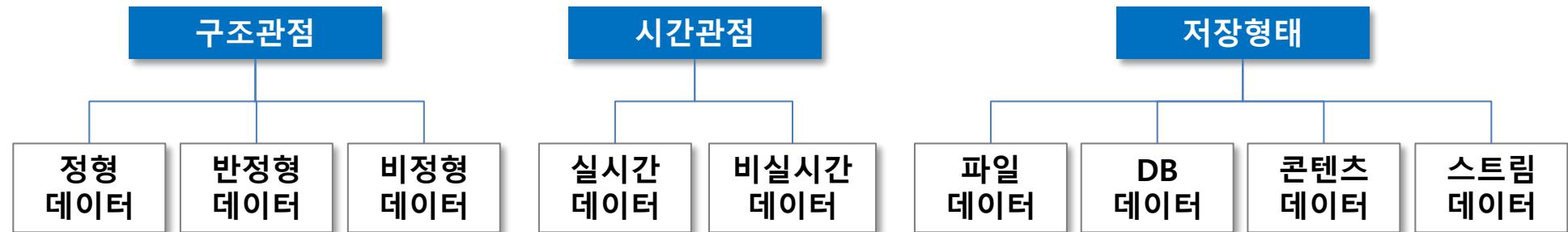
# Type of Data



유형	설명	종류
실시간 데이터	생성된 이후 수 초 ~ 수 분 이내에 처리되어야 의미가 있는 현재 데이터	센서 데이터, 시스템 로그, 보안 장비 로그, 알람, 네트워크 장비 로그
비실시간 데이터	이후에 처리되어야 의미가 있는 과거 데이터	통계, 웹 로그, 서비스 로그, 구매정보, 디지털 헬스케어 정보



# Type of Data



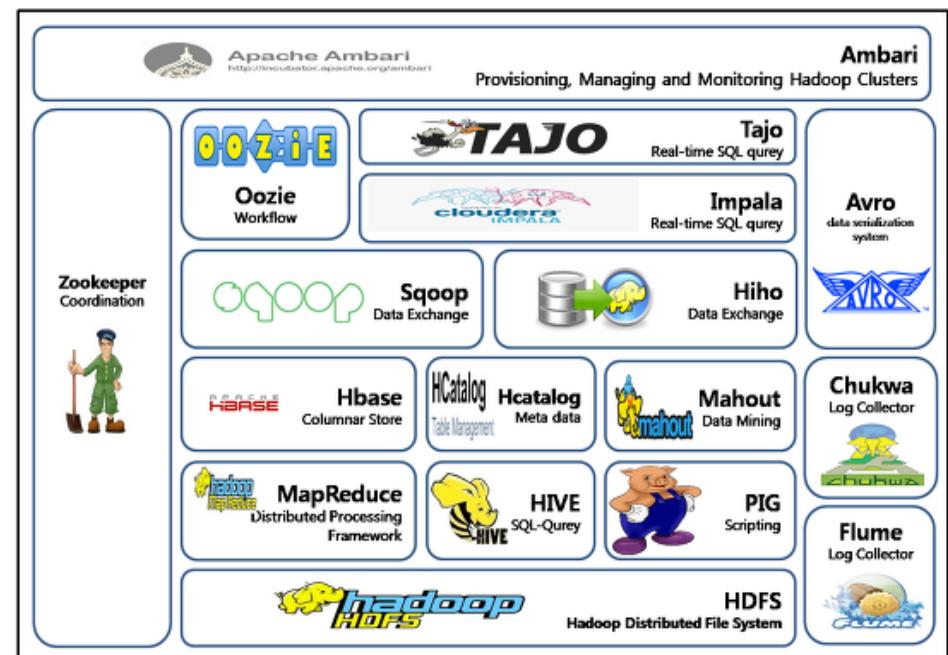
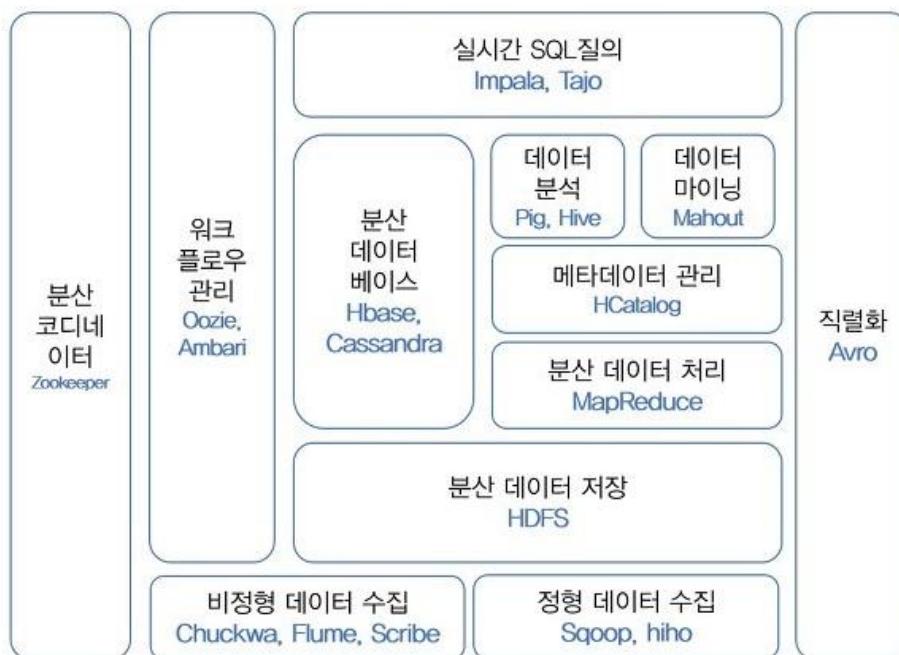
저장형태 관점의 데이터 유형
-----------------------

유형	설명
파일 데이터	시스템 로그, 텍스트, 스프레드 시트 등과 같이 파일 형식으로 파일 시스템에 저장되는 데이터
데이터베이스 데이터	관계형 데이터베이스, NoSQL, 인메모리 데이터베이스 등에 저장된 데이터
콘텐츠 데이터	텍스트, 이미지, 오디오, 비디오 등과 같이 개별적으로 데이터 객체로 구별 가능한 미디어 데이터
스트림 데이터	센서 데이터, HTTP 트랜잭션, 알람 등과 같이 네트워크를 통해 실시간으로 전송되는 데이터



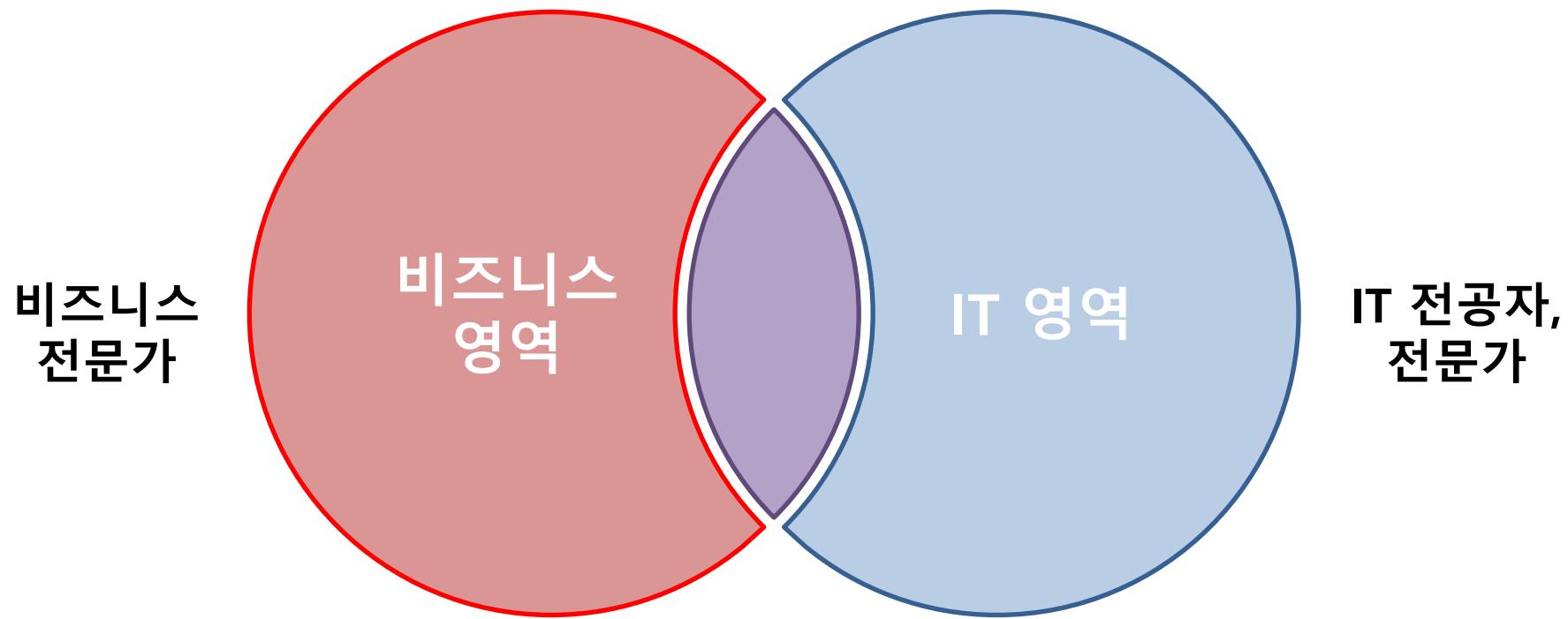
# Data Collection Methods and Technologies

▪ ETL(Extract, Transform, Load)	▪ 아파チ 카프카(Apache Kafka) ▪ 플럼(Flume)	▪ CEP(Complex Event Processing) ▪ EAI(Enterprise Application Integration)	▪ 크롤링(Crawling) ▪ RSS(Rich Site Summary)
▪ FTP(File Transfer Protocol)	▪ 스크라이브(Scribe)	▪ CDC(Change Data Capture) ▪ ODS(Operational Data Store)	▪ Open API
▪ 스쿱(Sqoop)	▪ 척와(Chukwa)		▪ 스트리밍(Streaming)



※출처: 하둡 에코시스템(Hadoop-Ecosystem)이란 <https://butter-shower.tistory.com/73>





“강사/컨설턴트는 어느 영역까지 cover해야 할 것인가?”

# Data Analysis Framework

## 데이터 수집

## 전처리

## 탐색

## 모델링

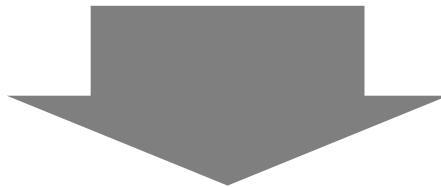
## 보고/적용

- 수십 기가바이트에 이르는 대용량 데이터 기반
- 데이터 마이닝 수행에 필요한 데이터 선별하여 샘플링 가능
- 데이터 분할 및 전처리
- 일관성이 없고 불완전한 데이터 정리
- 데이터를 분석하기 좋은 형태로 가공
- 데이터수치화 및 변수 간의 관계 관찰
- 시각화활용
- 분석 모형을 활용한 모델 구축
- 모델의 적합성 평가
- 분석결과 보고
- 시각화 자료첨부

“데이터분석 → 데이터수집 + 실질적인 데이터 분석”



**데이터분석 = 데이터수집 + 실질적인 데이터 분석**



# “어느 것이 더 어려운가?”

# 어느 것이 더 많은 노력과 시간이 들어가나?





*Like every great presentation, I've divided my talk into three subjects. Steve Jobs -*

I .

---

Data &  
Information

II .

---

**Database &  
Schema**

III .

---

Data  
Gathering



## 일반적으로 엑셀 프로그램을 사용하는 방식

	날짜	분류	품명	규격	Code	이전재고량	입고량	출고량	유실량	입고처	출고처	특기사항
149	2016-07-15	형강	ANGLE	45x45x3tx6m	T016-02	0	5			현대제철		
150	2016-07-15	형강	ANGLE	65x65x6tx10m	T016-03	92	1	55		현대제철	제1공장	
151	2016-07-20	가공식품	간장	샘표	F001-10	52		50			조립반	
152	2016-07-20	스위치	리미트	SD-34	E500-24	142		120			조립반	
153	2016-07-20	철자재	철판(EGI)	3.2*1219*2438	T016-08	39		33			조립반	
154	2016-07-20	형강	환봉	Φ 22*8m	T016-01	4		2			조립반	
155	2016-07-24	가공식품	고추장	순창	F001-11	0	3			샘표간장		
156	2016-07-24	가공식품	고추장	순창	F001-11	3	5			샘표간장		
157	2016-07-24	가공식품	고추장	순창	F001-11	8		5			제2공장	
158	2016-07-24	센서	말굽	MMG-RE	E300-04	2		3			제2공장	
159	2016-07-24	식품	김치	담근김치	F222-01	13	8			청정원		
160	2016-07-31	스위치	푸쉬버튼	SD-37	E500-27	79		5			금형반	
161	2016-07-31	철자재	파이프	20*40*1.4*8m	S103-01	38	2			포스코		
162	2016-07-31	철자재	파이프	30*30*1.4*6m	S103-02	43		11			제1공장	
163	2016-08-26	센서	온도	DHS-23	E300-05	34	5			포스코		
164	2016-08-28	가공식품	간장	샘표	F001-10	2		1			제3공장	
165	2016-08-29	기타	페인트	암적색	T001-01	76		3			제3공장	
166												
167												

출처: [https://www.kanghanet/bbs/board.php?bo\\_table=excel01&wr\\_id=3750&sca=%EC%9E%AC%EA%B3%A0%EA%B4%80%EB%A6%AC](https://www.kanghanet/bbs/board.php?bo_table=excel01&wr_id=3750&sca=%EC%9E%AC%EA%B3%A0%EA%B4%80%EB%A6%AC)





## Primary key, Foreign key

재고관리엑셀프로그램V44.xls - Microsoft Excel

	날짜	분류	품명	규격	Code	이전재고량	입고량	출고량	유실량	입고처	출고처	특기사항
149	2016-07-15	형강	ANGLE	45x45x3tx6m	T016-02	0	5			현대제철		
150	2016-07-15	형강	ANGLE	65x65x6tx10m	T016-03	92	1	55		현대제철	제1공장	
151	2016-07-20	가공식품	간장	샘표	F001-10	52		50			조립반	
152	2016-07-20	스위치	리미트	SD-34	E500-24	142		120			조립반	
153	2016-07-20	철자재	철판(EGI)	3.2*1219*2438	T016-08	39		33			조립반	
154	2016-07-20	형강	환봉	Φ 22*8m	T016-01	4		2			조립반	
155	2016-07-24	가공식품	고추장	순창	F001-11	0	3			샘표간장		
156	2016-07-24	가공식품	고추장	순창	F001-11	3	5			샘표간장		

기본 키(Primary key)는 특정 행을 유일하게 식별할 수 있는 속성 (ex. 주민등록번호)

- ✓ Unique한 값으로 중복이 허용되지 않음 → 테이블 생성 시 반드시 하나 이상의 필드/열을 지정
- ✓ Null값 포함이 불가 (Non-null)
- ✓ 가장 세분성(Granularity)이 높아야 함





## 데이터 모델링

Dimension 테이블과 Fact 테이블은  
특정한 기본키(primary key)와 외래키(foreign key)를 기준으로 상호 연결되며  
DT:FT는 일대다(1:n)의 관계(cardinality, 선택도)를 갖게 됨



# Data Analysis Approach



# 데이터 모델링

분류코드	분류명
1	형강
2	가공식품
3	스위치
4	철자재
5	센서
6	식품
7	기타

제품코드	제품코드2	제품명	규격	분류코드
1	T016-02	ANGLE	45	1
2	T016-03	ANGLE	65	1
3	F001-10	간장		2
4	E500-24	리미트		3
5	T016-08	철판		4
6	T016-01	환봉		1
7	F001-11	고추장		2
8	E300-04	말굽		5
9	F222-01	김치		6
10	E500-27	푸쉬버튼		3
11	S103-01	파이프	20	4
12	S103-02	파이프	40	4
13	E300-05	온도		5
14	T001-01	페인트		7

입출고일자	제품코드	입고량	출고량	입고처코드	출고처코드
2024-07-15	1	5		1	
2024-07-16	2	1		1	
2024-07-16	2		55		1
2024-07-17	3		50		2
2024-07-18	4		120		2
2024-07-19	5		33		2
2024-07-20	6		2		2
2024-07-21	7	3		2	
2024-07-22	7	5		2	
2024-07-23	7		5		3
2024-07-24	8		3		3
2024-07-25	9	8		3	
2024-07-26	10		5		4
2024-07-27	11	2		4	
2024-07-28	12		11		1
2024-07-29	13	5		4	
2024-07-30	3		1		5
2024-07-31	14		3		5

입고처코드	입고처
1	현대제철
2	샘표간장
3	청정원
4	포스코

출고처코드	출고처
1	제1공장
2	조립반
3	제2공장
4	금형반
5	제3공장

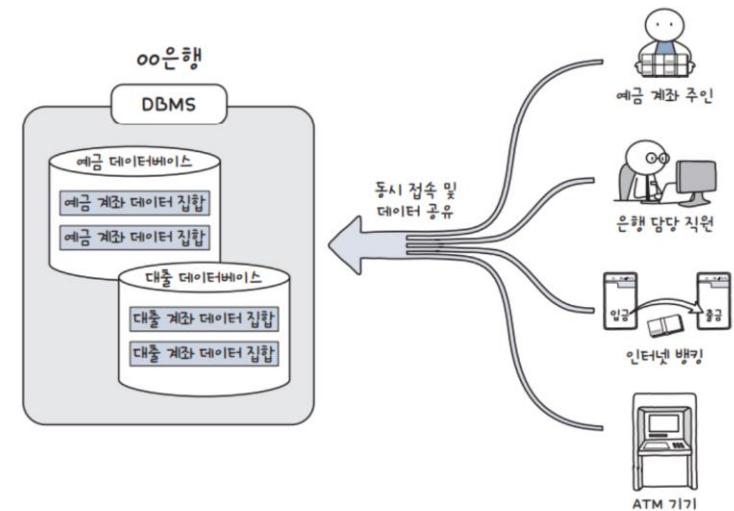
※ 참고: Star 스키마와 Snowflake 스키마 <https://www.integrate.io/ko/blog/snowflake-schemas-vs-star-schemas-what-are-they-and-how-are-they-different-ko/>





## 데이터베이스

- **데이터베이스**(database, DB): 여러 사람이 공유하여 사용할 목적으로 체계화해 통합, 관리하는 데이터의 집합이다. (위키백과)
- **DBMS(Database Management System)**: 데이터베이스를 운영하고 관리하는 소프트웨어(예: MySQL, Oracle, SQL 서버, PostgreSQL, MariaDB)
- **SQL(Structured Query Language)**: 구조화된 질의 언어라는 뜻으로 관계형 데이터베이스에서 사용되는 언어
- 데이터베이스는 1개 이상의 **테이블**로 이루어지며, 테이블은 **fact 테이블**과 **dimension 테이블**로 구분됨



※그림 출처: 혼공리들의 스터디 공간 <https://hangongrihabook.tistory.com/50>



# Data Analysis Approach



## 데이터베이스

```
Microsoft Windows [Version 10.0.22631.3155]
(c) Microsoft Corporation. All rights reserved.

C:\Users\821033669010>cd "C:\program files\MariaDB 10.8\bin"

C:\Program Files\MariaDB 10.8\bin>mysql.exe -u root -p
Enter password: *****
Welcome to the MariaDB monitor. Commands end with ; or \g.
Your MariaDB connection id is 6
Server version: 10.8.5-MariaDB mariadb.org binary distribution

Copyright (c) 2000, 2018, Oracle, MariaDB Corporation Ab and others.

Type 'help;' or '\h' for help. Type '\c' to clear the current input
statement.

MariaDB [(none)]> SHOW databases;
+-----+
| Database |
+-----+
| financial_terms
| information_schema
| investar
| kotra
| mysql
| opentutorials
| performance_schema
| scraping
| sys
| wikipedia
| workbench
+-----+
11 rows in set (0.002 sec)
```

```
MariaDB [(none)]> USE opentutorials;
Database changed
MariaDB [opentutorials]> SHOW tables;
+-----+
| Tables_in_opentutorials |
+-----+
| author
| topic
| topic_backup
+-----+
3 rows in set (0.010 sec)

MariaDB [opentutorials]> DESC topic;
+-----+-----+-----+-----+-----+-----+
| Field | Type      | Null | Key | Default | Extra   |
+-----+-----+-----+-----+-----+-----+
| id    | int(11)   | NO  | PRI | NULL    | auto_increment |
| title | varchar(100)| NO  |     | NULL    |                 |
| description | text | YES |     | NULL    |                 |
| created | datetime | NO  |     | NULL    |                 |
| author_id | int(11) | NO  |     | NULL    |                 |
+-----+-----+-----+-----+-----+-----+
5 rows in set (0.018 sec)

MariaDB [opentutorials]> SELECT * FROM topic;
+-----+-----+-----+-----+-----+
| id | title          | description          | created          | author_id |
+-----+-----+-----+-----+-----+
| 1  | MySQL Database | MySQL is Relational ... | 2023-12-19 10:03:32 | 1 |
| 2  | Oracle          | Oracle is ...        | 2023-12-19 10:09:53 | 1 |
| 5  | MongoDB         | MongoDB is ...       | 2023-12-19 11:29:44 | 1 |
+-----+-----+-----+-----+-----+
3 rows in set (0.011 sec)
```





*Like every great presentation, I've divided my talk into three subjects. Steve Jobs -*

I .

---

Data &  
Information

II .

---

Database &  
Schema

III .

---

**Data  
Gathering**



# Disk Operation System

로컬 디스크 (C:)

내 PC > 로컬 디스크 (C:) >

로컬 디스크 (C:) 검색

새로 만들기 +

파일 폴더

이름 수정한 날짜 유형 크기

이름	수정한 날짜	유형	크기
17.10 Jigoo Partners	2024-07-25 오후 2:19	파일 폴더	
19.09 CNKG	2024-07-19 오후 7:13	파일 폴더	
22.06 Inner People	2024-03-28 오후 8:59	파일 폴더	
23.01 Zeta Data	2024-01-17 오전 10:39	파일 폴더	
24.04_위너아이_valuation	2024-06-12 오전 10:39	파일 폴더	
Coupaing-Review-Crawling	2024-04-27 오후 11:07	파일 폴더	
CREON	2022-11-21 오후 5:31	파일 폴더	
daishin	2022-11-21 오후 10:07	파일 폴더	
Data Science	2024-03-05 오후 11:35	파일 폴더	
Hadoop	2023-04-01 오후 1:21	파일 폴더	
INNORIX_Agent	2024-07-14 오후 3:17	파일 폴더	
INZISOFT	2023-01-11 오후 1:54	파일 폴더	
KED	2024-04-24 오후 3:27	파일 폴더	
KMPlayer	2022-05-30 오후 6:41	파일 폴더	
KWIC	2024-04-24 오후 3:28	파일 폴더	
Langchain	2023-05-09 오후 9:28	파일 폴더	
Logs	2024-07-14 오후 2:25	파일 폴더	
musinsa	2023-09-13 오후 2:54	파일 폴더	
myPackage	2023-07-07 오후 5:24	파일 폴더	
mySite	2022-11-20 오후 4:04	파일 폴더	
PerfLogs	2022-05-07 오후 2:24	파일 폴더	
Program Files	2024-04-25 오전 12:31	파일 폴더	
Program Files (x86)	2024-07-04 오전 1:06	파일 폴더	
Python311	2024-01-15 오후 12:12	파일 폴더	
Risk	2022-12-07 오후 6:49	파일 폴더	
SCAN	2023-01-12 오전 10:53	파일 폴더	
spark-2.4.5-bin-hadoop2.7	2023-04-01 오후 1:41	파일 폴더	
Temp	2023-03-07 오후 4:57	파일 폴더	
VirtualEnv	2022-10-24 오전 10:29	파일 폴더	
Windows	2024-07-27 오전 12:25	파일 폴더	
사용자	2023-08-16 오후 4:40	파일 폴더	
agentlog	2023-12-27 오후 4:24	텍스트 문서	1KB
DumpStack	2024-02-11 오후 11:06	텍스트 문서	12KB
ServiceTest	2024-07-12 오전 10:13	텍스트 문서	1KB

34개 항목

명령 프롬프트

```
C:\>dir
C 드라이브의 볼륨에는 이름이 없습니다.
볼륨 일련 번호 : 7C9C-4BF6
```

C:\ 디렉터리

```
2024-07-25 오후 02:19 <DIR> 17.10 Jigoo Partners
2024-07-19 오후 07:13 <DIR> 19.09 CNKG
2024-03-28 오후 08:59 <DIR> 22.06 Inner People
2024-01-17 오전 10:39 <DIR> 23.01 Zeta Data
2024-06-12 오전 10:39 <DIR> 24.04_위너아이_valuation
2023-12-27 오후 04:24 <DIR> 663 agentlog.txt
2024-04-27 오후 11:07 <DIR> Coupang-Review-Crawling
2022-11-21 오후 05:31 <DIR> CREON
2022-11-21 오후 10:07 <DIR> daishin
2024-03-05 오후 11:35 <DIR> Data Science
2024-02-11 오후 11:06 <DIR> 12,288 DumpStack.log
2023-04-01 오후 01:21 <DIR> Hadoop
2024-07-14 오후 03:17 <DIR> INNORIX_Agent
2023-01-11 오후 01:54 <DIR> INZISOFT
2024-04-24 오후 03:27 <DIR> KED
2022-05-30 오후 06:41 <DIR> KMPlayer
2024-04-24 오후 03:28 <DIR> KWIC
2023-05-09 오후 09:28 <DIR> Langchain
2024-07-14 오후 02:25 <DIR> Logs
2023-09-13 오후 02:54 <DIR> musinsa
2023-07-07 오후 05:24 <DIR> myPackage
2022-11-20 오후 04:04 <DIR> mySite
2022-05-07 오후 02:24 <DIR> PerfLogs
2024-04-25 오전 12:31 <DIR> Program Files
2024-07-04 오전 01:06 <DIR> Program Files (x86)
2024-01-15 오후 12:12 <DIR> Python311
2022-12-07 오후 06:49 <DIR> Risk
2023-01-12 오전 10:53 <DIR> SCAN
2024-07-12 오전 10:13 <DIR> ServiceTest.txt
2023-04-01 오후 01:41 <DIR> spark-2.4.5-bin-hadoop2.7
2023-03-07 오후 04:57 <DIR> Temp
2023-08-16 오후 04:40 <DIR> Users
2022-10-24 오전 10:29 <DIR> VirtualEnv
2024-07-27 오전 12:25 <DIR> Windows
3개 파일 13,541 바이트
31개 디렉터리 35,911,946,240 바이트 남음
```

C:\>



# 웹 데이터 가져오기

단 3줄로 가능한 쿠팡 상품 댓글 크롤링

1. 인용 사이트 : <https://github.com/JaehyoJJAng/Coupa...>
  2. Git 설치 (<https://git-scm.com/downloads/win>)  
→ 참조 사이트 : <https://m.blog.naver.com/insideimagix/223325204776>
  3. Python 설치  
→ (<https://dotiromoook.tistory.com/32>)
  4. 명령 프롬프트 또는 “cmd” 실행 → DOS 화면 뜸
  5. \$ git clone https://github.com/JaehyoJJAng/Coupa...git
  6. \$ cd Coupa...Review-Crawling
  7. \$ pip install -r ./requirements.txt

**반드시  
참조**

실행1 \$ python main.py

실행2 복사된 url 주소 붙여넣기 → Enter → 페이지 수 지정 → Enter





## MS Excel vs Google Sheets



Google Sheets