



AI 융합전문가
(7기 입문 4차시)

데이터분석

2024.05.25.





현재 하는 일과 보유 자격증



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

◎ 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)



강사 홍용기 컨설팅학 박사



데이터분석 관련 비즈니스

◎ 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
23.08~24.07 ODA (요르단 경찰청 PSD)

발급번호:00KH83K-W6YQ-OA64-CG1Z

소프트웨어기술자 경력증명서

성명	홍용기		생년월일	1964.08.25	
현 근무처	회사명		사업자등록번호		
	전화번호		업종		
	소재지				
기술자격	확인여부	근무기간	회사명	담당업무	
				부서/직위	
	증목 및 등급	등록번호	취득일	발급기관	
학력	빅데이터분석기사	BAE-002000023	2021.07.16	한국데이터산업진흥원	
	ADsP(데이터 분석 전문가)	ADsP-028000961	2021.04.09	한국데이터산업진흥원	
교육	학교명	학과(전공)	수학기간	학위	
	기간	과정	수료번호	교육기관	
	수여일	종류	상훈기관	근거	
기술경력	확인여부	참여사업명	참여기간	발주자	
	확인	국민 체감형 치안 안심 플랫폼 구축 정보화전략계획 사업	2022.08.04 ~ 2022.12.31	경찰대학교 (주)메타로직컨설팅	프리랜서 IT컨설팅 > 정보기술기획
	확인	서울시 차세대 지방세 임정보시스템 통합구축 변화관리 컨설팅	2021.11.01 ~ 2022.06.30	서울특별시 (주)메타로직컨설팅	프리랜서 IT컨설팅 > 정보기술컨설팅
	확인	소상공인지원사업 디지털전환 정보화전략 계획(ISP) 용역	2021.01.04 ~ 2021.05.03	(주)메타로직컨설팅	수석컨설턴트 IT컨설팅 > 정보기술컨설팅
	확인	한국연구재단 중장기 정보화전략계획(ISP) 수립	2020.10.05 ~ 2021.01.04	한국연구재단 (주)메타로직컨설팅	수석컨설턴트 IT컨설팅 > 정보기술컨설팅
	확인	창업기업확인시스템 구축을 위한 정보화전략계획(ISP) 수립	2020.05.18 ~ 2020.08.17	창업진흥원 (주)메타로직컨설팅	수석컨설턴트 IT컨설팅 > 정보기술컨설팅

◎ 2022년 AI학습용데이터 구축사업 평가

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

◎ 데이터 가치평가 컨설팅

- 23.09~23.11 중소벤처기업진흥공단

「소프트웨어 진흥법」제24조제3항 및 같은 법 시행규칙 제13조제3항에 따라 소프트웨어기술자의 경력
사항을 증명합니다.

2023년 01월 25일





데이터분석 관련 강의

- 데이터분석 및 실전 R코딩 (경영기술지도사회, 빅데이터 분석기사 자격증 취득 과정)
 - 데이터분석 Python 심화과정 (서울 여성능력개발원 강동 여성인력개발센터 / 용산 여성인력개발센터)
 - No-coding 크롤링과 데이터분석 (경영지도사회 중부지회)
 - AI & ChatGPT 활용 및 데이터분석 컨설팅 방법론 (경영기술지도사회, 국제공인컨설턴트 CMC 양성과정)

HOME > City&Global

홍용기 컨설팅학박사, 한국경영기술지도사회 '빅데이터 분석 실무 과정' 성료

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

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



중요기사

- 외대교수 시국선언? "나들 암가짐"
- 한일 정상회담 전하는 KBS 라디오
- 민경욱 대표, "중앙선관위를 직무
- 김기현, 가족들을 청탁을 만나 '맞'
- 언론단체들, "대통령은 기소된 한

인기뉴스

- "北해경 관련 산관위에 대한 의
- [최재용 칼럼] 언공지능(AI) 시대, 위협과 기...
- [평양언의 손편지] (359) 도독이 제 발 차...
- [전정희 작가 칼럼] 가정의 말 5월말을 맞이하여
- 속도로 서울시의원, 전기차 충전구역 확장안

성서로일하기센터의 '빅데

(주제) 데이터 흥용기 대표, 용산여성인력개발센터의 '빅데이터분석을 위한 파이썬(Python) 심화 과정' 성료

윤성일 | ① 출판 2023.08.10 19:51 | ② 댓글 0

중요기사

- 이동근 아들 친구, '난 학폭 피해자 아닌, 당시 ...'
- '대한민국 외교복 종별화' 관련 부서 청진 차관...
- 이와함께 '선택적 길사' 받겠다는 선관위에 작별...

중요기사

- '북한증증' 영웅에게 40억국민증...,'박진식 보훈장...'
- 새CBS공동취재, 'KBS의 자본적인 자사 관점 평창...
- 김정경 전 mbc사장, '국가 폭력이란?'
- 법사위 실전 최강호 '만죽거리지 말리 VS 한동...'
- 김의겸, 고(故) 재모 송영 사건 '우사 기록 찾고 ...'



디자인융합교육원 & **FUN TODAY**
제1회 빅GPT 활용
객원기자 양성 과정

모집인원
모집인원 모집인원 모집인원 모집인원 모집인원 모집인원 모집인원 모집인원 모집인원

모집기간

강사 홍용기 컨설팅학 박사



책 쓰기

KYOBOTextbook

통합검색 대입을 결정하는 초등 영어 공부법

대학교재 어린이 베스트 신상품 이벤트 사은품 PICKS CASTing 교보ONLY

국내도서 경제/경영 경영전략 경영전략일반

성장하는 기업의 5가지 조건

한치호, 홍용기, 하현식, 최종철 저자(글)
한국경제신문 2018년 01월 29일

10.0
(1개의 리뷰)
도움돼요
(100%의 구매자)

“
도움돼요
(100%의 구매자)

성장하는
기업의
5가지
조건

한치호·홍용기·
하현식·최종철
저자

“어떻게 지속적으로 성장할 것인가?”
글이 보이지 않는 저성장 시대의 위기를 돌파하고
무한 경쟁 시대에도 살아남는 전략을 알려준다!

한국경제신문

KYOBObib

eBook <돈의 심리학> 모건 하우저 신작!

챗GPT AI로 데이터분석 마스터하기

초보자도 가능한 노코딩 AI 데이터분석

eBook 6,300원 sam 구독

챗GPT AI로 데이터분석 마스터하기

초보자도 가능한 노코딩 AI 데이터분석

챗GPT AI로 데이터분석 마스터하기

홍용기 이현구 윤성임 저자

디즈비즈북스

2023년 05월 17일 출간

10
(1개의 리뷰)
도움돼요
(100%의 구매자)

eBook 상품 정보

- 파일 정보: pdf (11.43MB)
- ISBN: 9791169982573
- 듣기(TTS): 가능 ①

디즈비즈북스
메타버스교육센터





책 쓰기



AI융합비즈니스포럼
<https://www.aimaster.kr>



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?



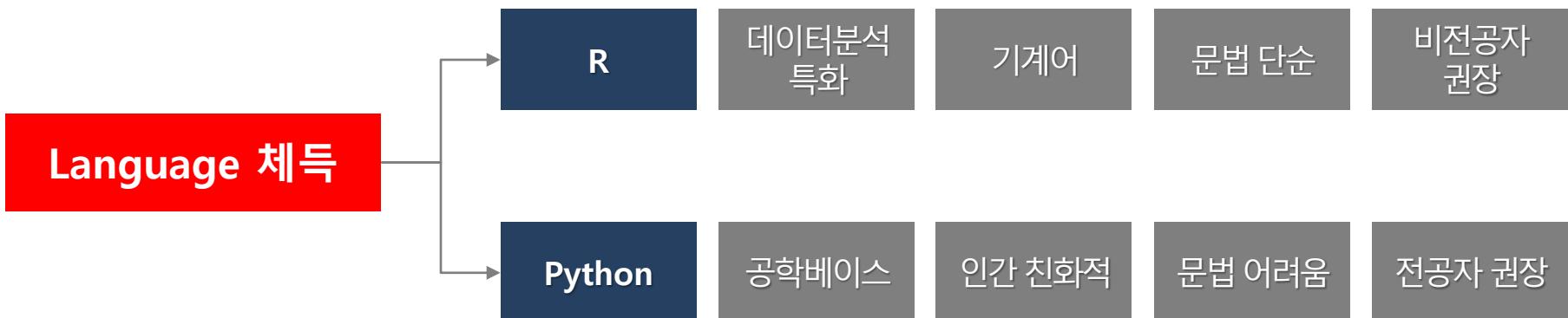
혹자는
인생 뭐 별거 있냐고 할테지만...

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?



2024년 데이터분석 자격증 시험일정

→ K-data 데이터자격검정(<https://www.dataq.or.kr/www/accept/schedule.do>)



구분	회차	접수기간	수험표발급	시험일	사전점수공개 및 재검토 접수	결과발표	증빙서류 제출기간	
빅데이터 분석기사	제8회	필기	3.4~8	3.22	4.6(토)	4.19~23	4.26	4.29~5.9
		실기	5.20~24	6.7	6.22(토)	7.5~9	7.12	-
	제9회	필기	8.5~9	8.23	9.7(토)	9.20~24	9.27	9.30~10.10
		실기	10.28~11.1	11.15	11.30(토)	12.13~17	12.20	-
데이터분석 전문가 	제32회	필기	1.22~26	2.8	2.24(토)	3.15~19	3.22	-
		실기	3.22~29	4.12	4.27(토)	5.17~21	5.24	5.24~31
	제33회	필기	7.1~5	7.26	8.10(토)	8.30~9.3	9.6	-
		실기	9.9~13	9.27	10.12(토)	11.1~5	11.8	11.8~15
데이터분석 준전문가 	제40회	-	1.22~26	2.8	2.24(토)	3.15~19	3.22	-
	제41회	-	4.8~12	4.26	5.11(토)	5.31~6.4	6.7	-
	제42회	-	7.1~5	7.26	8.10(토)	8.30~9.3	9.6	-
	제43회	-	9.30~10.4	10.18	11.3(일)	11.22~26	11.29	-



경영정보 시각화능력(BI specialist) 소개

빅데이터시대 취업 필수 국가기술자격증

경영정보시각화능력 BI Specialist

대한상공회의소

경영정보시각화능력(Business Intelligence Specialist)

기업의 내외부 정보를 시각화 요소들을 사용하여 효과적으로 표현하고 전달하는 역량을 평가하는 국가기술자격입니다.

경영정보시각화능력(BI Specialist)을 취득해야하는 이유

- 취업 활용도 높은 자격**
취업에 도움되는 자격
- 전세계적으로 기업, 공공기관 BI 도입으로 인력수요 증가
- 빅데이터시대 필수역량**
데이터분석, 보고서 작성의 핵심역량 - 시각화능력
- 미래 소프트웨어**
엑셀, 파워포인트에 이은 미래 필수 사무 소프트웨어
- 국가기술자격증**
국가에서 시각화분야 전문성을 인정하는 자격증
- 어느 직무에나 필요**
컴퓨터활용능력과 같이 취업 및 사무에 요구되는 필수 자격
- 비즈니스 인사이트**
BI 툴을 사용하여 직접 비즈니스 인사이트 도출

컴활에 이은 사무 필수 자격증

**Business
Intelligence
Specialist**



AI융합비즈니스포럼

매출 분석 보고서



매출금액	매출금액, CalendarYear	이익률, CategoryName	매출금액, CalendarYear 및 ChannelName
1,657,001,820	2017 29% 2018 32% 2019 39%	Accessories: 51.62% Clothing: 49.20% Components: 43.84% Bikes: 40.33%	Channel... Catalog: 2019 Online: 2018 Reseller: 2017 Store: 2019
매출건수			
61,473			
매출원가	매출금액 및 이익률, ChannelName 및 CalendarYear	매출금액, EnglishMonth 및 CalendarYear	
953,471,477	CalendarYear: 2017, 2018, 2019 이익률: 42.40%, 42.45%, 42.50%, 42.55%	CalendarYear: 2017, 2018, 2019 EnglishMonth: Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec	
매출이익	Online: 0.6십억 Reseller: 0.4십억 Store: 0.2십억 Catalog: 0.0십억	Jan: 55.0만 Feb: 48.0만 Mar: 52.0만 Apr: 54.0만 May: 58.0만 Jun: 48.0만 Jul: 55.0만 Aug: 48.0만 Sep: 45.0만 Oct: 58.0만 Nov: 45.0만 Dec: 55.0만	
703,530,343			
이익률			
42.5%			

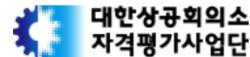
Copyright © 2021  Power BI all right reserved

Country: 모두 선택, Australia, Canada, France, Germany, United Kingdom, United States



경영정보 시각화능력(BI specialist) 2024 시험일정

로그인 · 회원가입 · 기관소개



종목소개

개별접수

단체접수

마이페이지

고객센터

활용현황



경영정보시각화능력 국가기술자격

종목소개

시험안내

시험문제

시험일정

관련자료

FAQ

시험일정

사무정보

컴퓨터활용능력

경영정보시각화능력

워드프로세서

비서

엑셀플러스

파워포인트플러스

한글플러스

IT Plus

유통/마케팅

유통관리사

전자상거래관리사

전자상거래운용사

경영정보시각화능력 시험일정

- 인터넷접수 여건이 안되는 수험자께서는 인터넷접수 기간중 해당상의를 방문해 주시면 접수에 필요한 시스템(사진 스캔 등)을 제공해 드립니다.
- 접수기간 중이라도 수험자가 많을 경우 시험장은 조기 마감될 수 있습니다.
- 정기 검정 원서접수 마지막 날의 접수마감은 18:00 까지 입니다.(수험료 결제까지 완료하고 접수증이 확인되어야 원서접수가 된 것입니다.)

정기시험 일정안내

2024

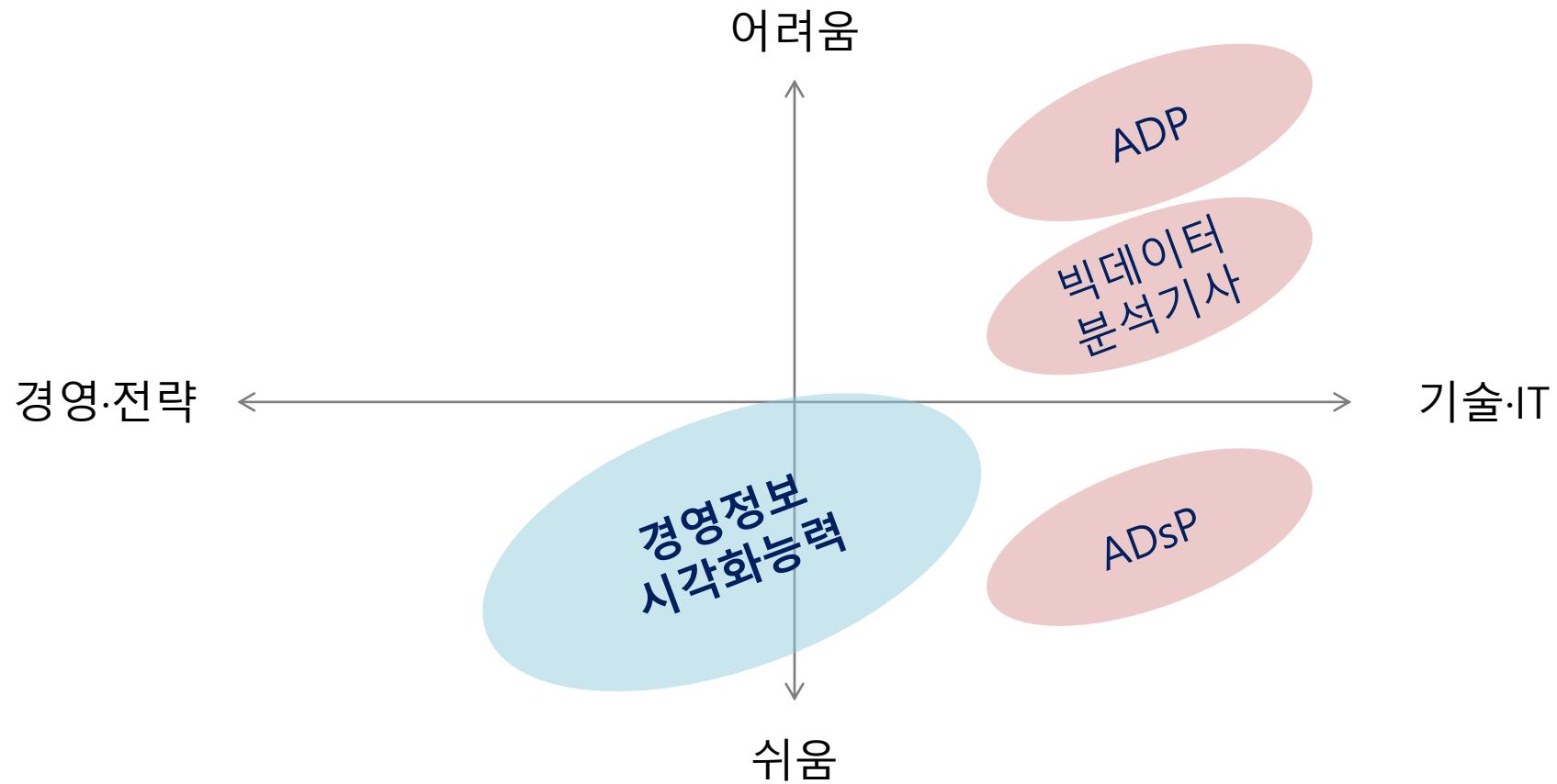
조회

종목	회별	구분	등급	인터넷접수	시험일자	발표일자
경영정보시각화능력	1	필기	단일등급	(1차) 2024.03.18 ~ 2024.03.24 (2차) 2024.04.17 ~ 2024.04.23	2024.05.18	2024.06.18
경영정보시각화능력	1	실기	단일등급	2024.08.28 ~ 2024.09.03	2024.09.28	2024.11.18
경영정보시각화능력	2	필기	단일등급	(1차) 2024.09.30 ~ 2024.10.06 (2차) 2024.10.30 ~ 2024.11.05	2024.11.30	2024.12.31



경영정보 시각화능력(BI specialist) 국가기술자격증

→ <https://license.korchaam.net/co/examguide.do?cd=0108&mm=28>



다같이 생각해 봅시다...



기술의 가치

“가치를 아는 사람에게 기술이 가야 빛을 발한다.”

WWW



사례: 빅카인즈 활용 <https://www.bigkinds.or.kr>



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 &
Consulting**

II .

Database &
Schema

III .

Internet &
Web



DATA

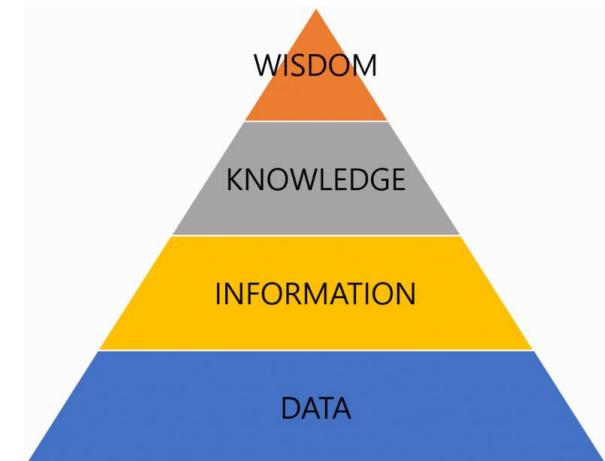
‘추론과 추정의 근거를 이루는 사실’

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



- ✓ 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

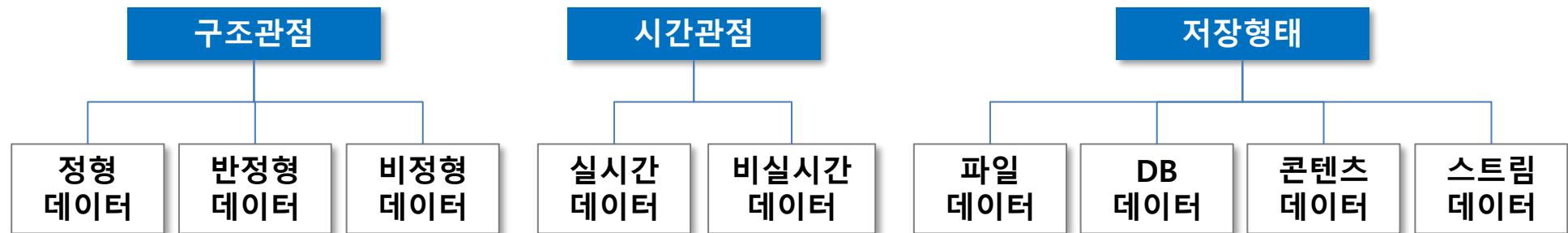


‘데이터분석’이라고 하면…

어떠한 데이터를 어느 범위까지
분석해야 하는 것인가?



Type of Data

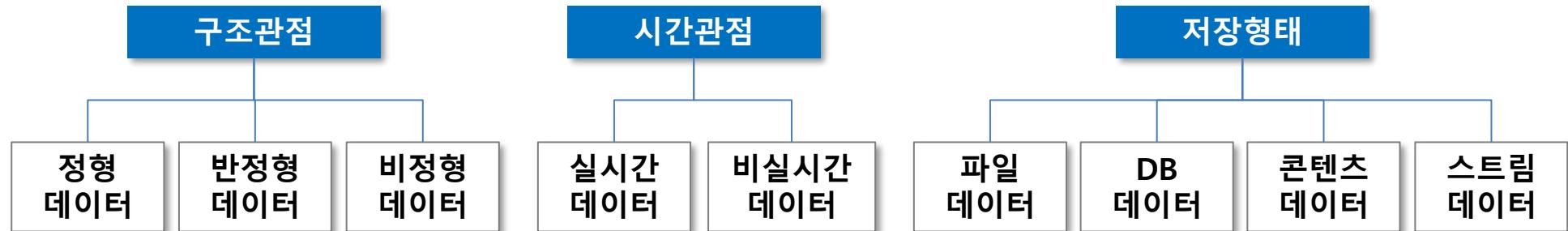


구조관점의 데이터 유형

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



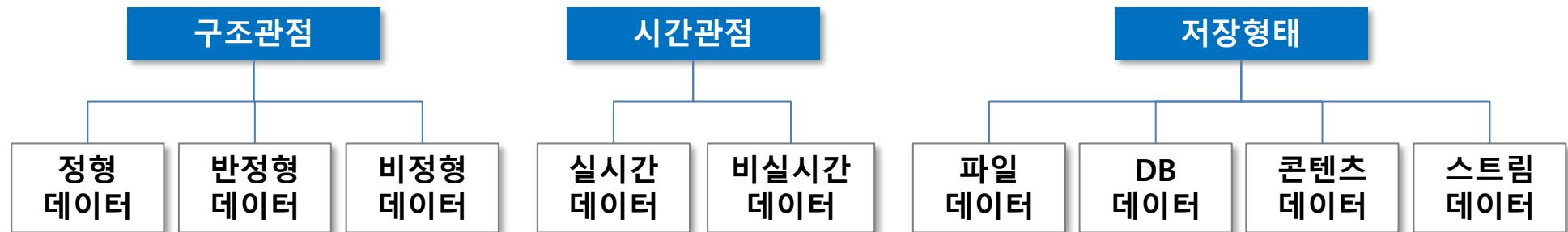
Type of Data



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



Type of Data



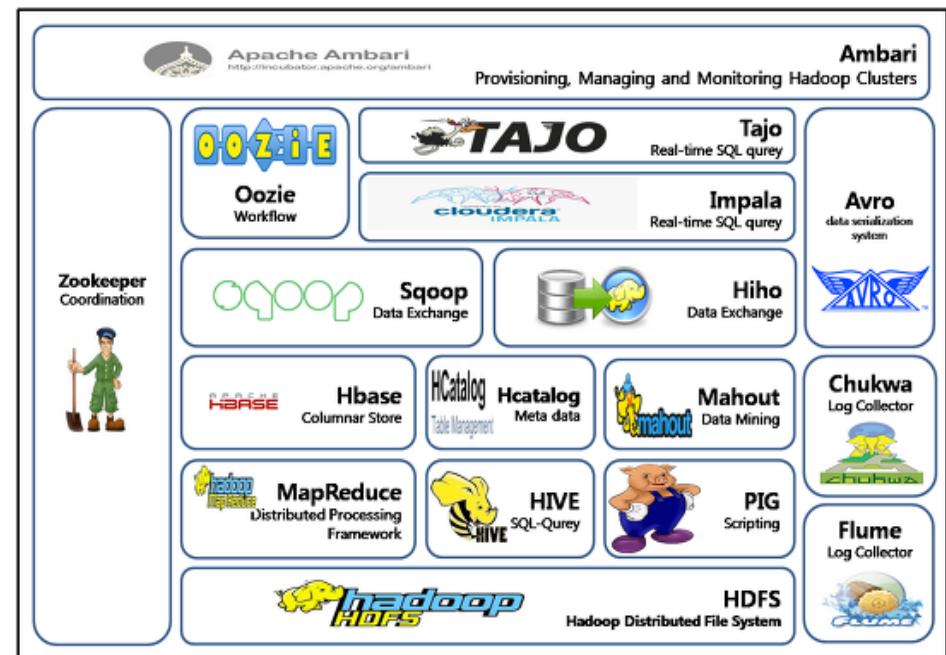
저장형태 관점의 데이터 유형

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



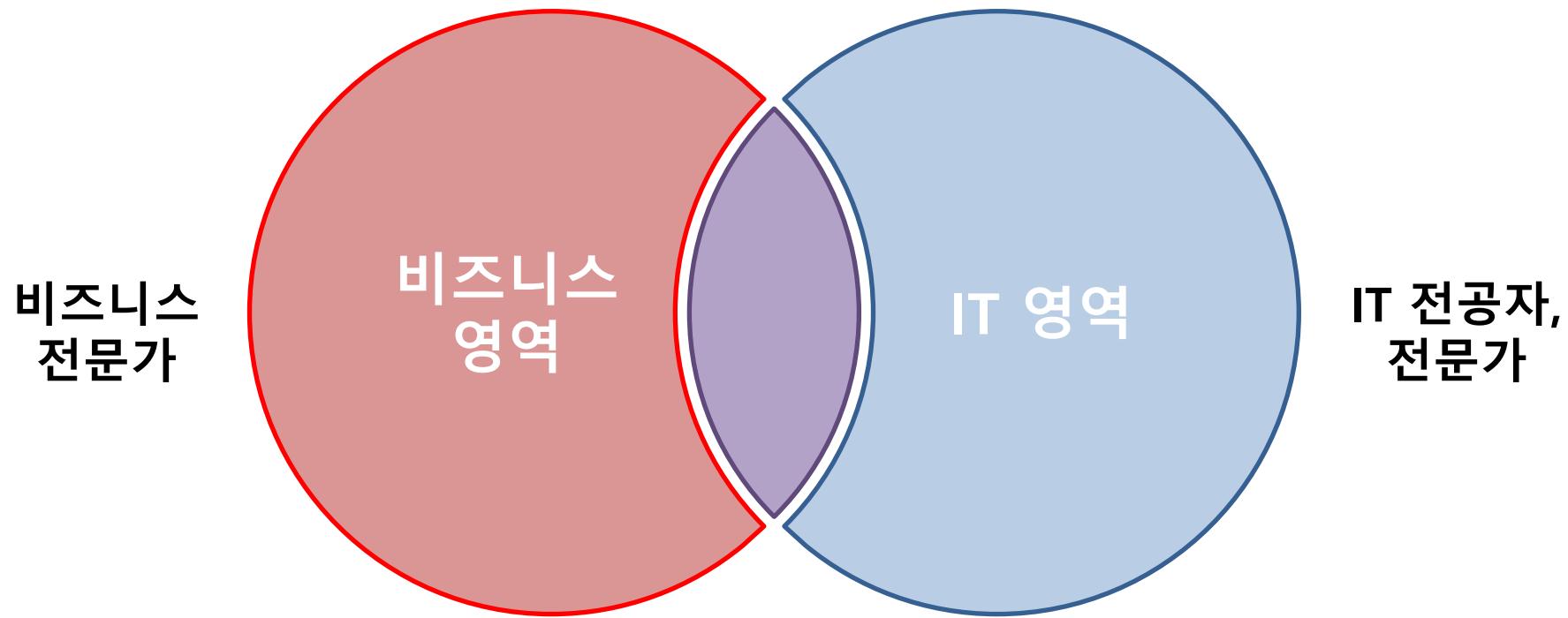
Data Collection Methods and Technologies

<ul style="list-style-type: none"> ETL(Extract, Transform, Load) FTP(File Transfer Protocol) 스쿱(Sqoop) 스크래파이(Scrapy) 	<ul style="list-style-type: none"> 아파チ 카프카(Apache Kafka) 플럼(Flume) 스크라이브(Scribe) 척와(Chukwa) 	<ul style="list-style-type: none"> CEP(Complex Event Processing) EAI(Enterprise Application Integration) CDC(Change Data Capture) ODS(Operational Data Store) 	<ul style="list-style-type: none"> 크롤링(Crawling) RSS(Rich Site Summary) Open API 스트리밍(Streaming)
--	---	---	--



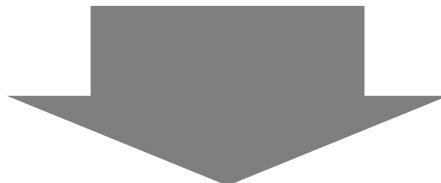
※출처: 하둡 에코시스템(Hadoop-Ecosystem)이란 <https://butter-shower.tistory.com/73>





“컨설턴트는 어느 영역까지 cover해야 할 것인가?”

강의/컨설팅 차원에서의 데이터 분석



데이터를 통한 문제 해결 vs 문제 해결을 위한 데이터

“업체가 보유한 데이터로도 충분하다.”



Data Analysis Framework

데이터 수집

전처리

탐색

모델링

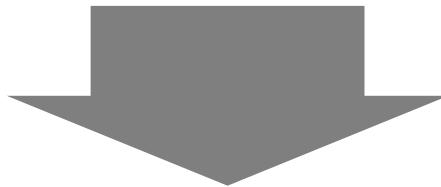
보고/적용

- 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
- 시각화자료첨부

“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 &
Consulting

II .

**Database &
Schema**

III .

Internet &
Web



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





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	¶ 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)

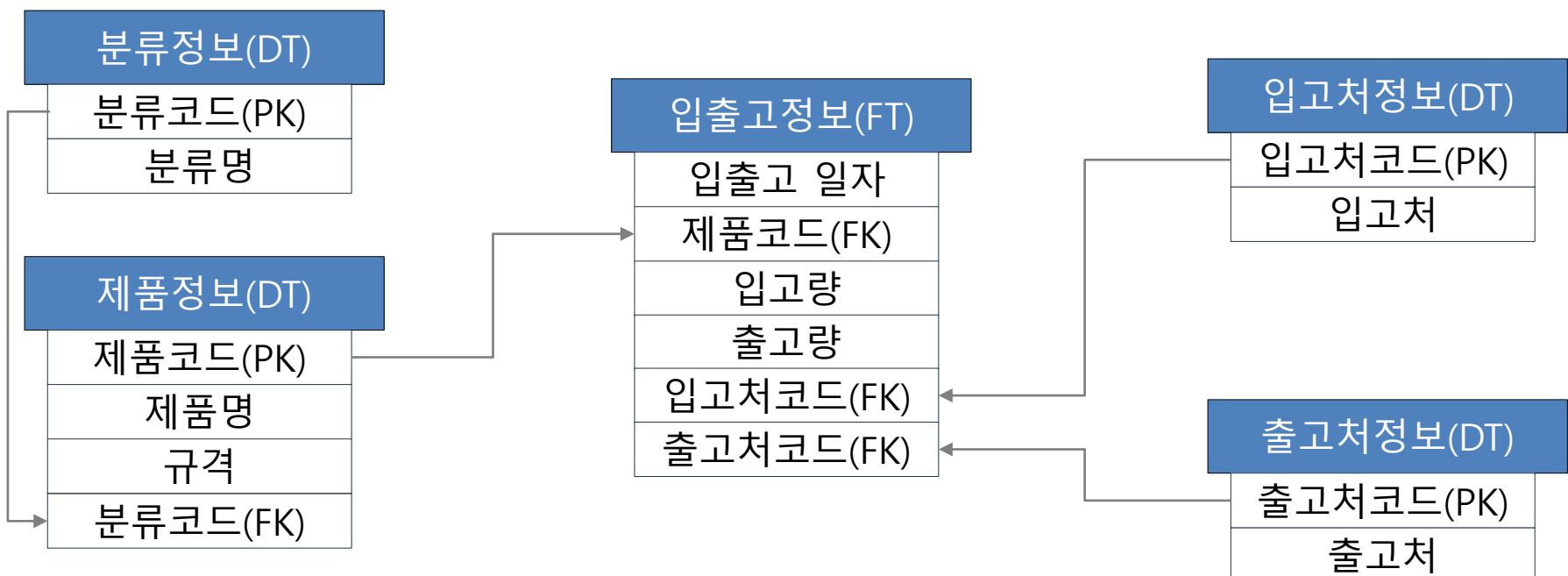
- ✓ Unique한 값으로 중복이 허용되지 않음 → 테이블 생성 시 반드시 하나 이상의 필드/열을 지정
- ✓ Null값 포함이 불가 (Non-null)
- ✓ 가장 세분성(Granularity)이 높아야 함





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

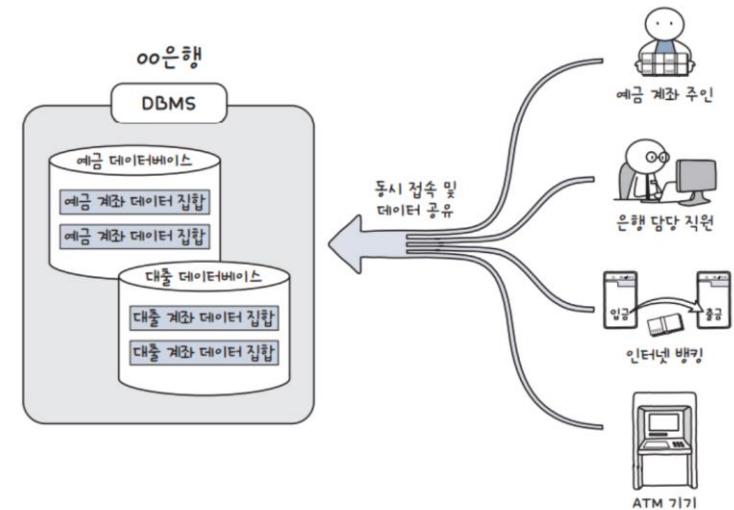
※ 참고: Star 스키마와 Snowflake 스키마 <https://www.integrate.io/ko/blog/snowflake-schemas-vs-star-schemas-what-are-they-and-how-are-they-different-ko/>





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



※그림출처 혼공리들의스터디공간 <https://hangongrihablog.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 &
Consulting

II .

Database &
Schema

III .

**Internet &
Web**



Internet & Web

Internet = Web

VS

Internet ≠ Web

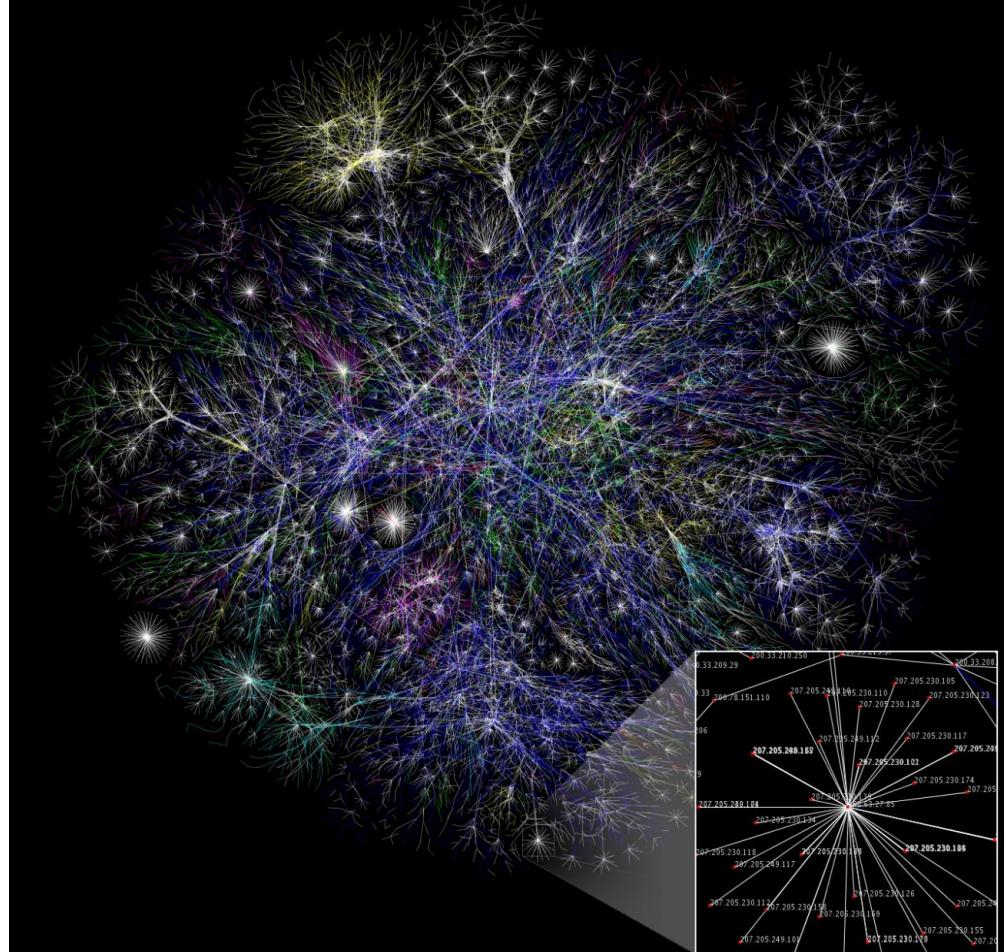


Internet

인터넷은 인간이 발명해 놓고도 이해하지 못하는 최초의 발명품이며, 역사상 최대 규모의 무정부주의에 대한 실험이다.

The Internet is the first thing that humanity has built that humanity doesn't understand, the largest experiment in anarchy that we have ever had.

- Eric Emerson Schmidt



※라우터를 통해 연결된 인터넷을 시각화한 그림(위키백과)



Internet

인터넷(Internet)은
인터넷 프로토콜 스위트(TCP/IP)를 기반으로 하여 전 세계적으로
연결되어 있는 컴퓨터 네트워크 통신망을 일컫는 말이다.
그야말로 인류의 역사상 전례 없는 거대한 정보의 바다인 셈이다.

흔히 웹(WEB)이라고 줄여 부르는
월드 와이드 웹(World Wide Web; WWW)만 생각하기 쉽지만
인터넷은 월드 와이드 웹, 전자 메일, 파일 공유(토렌트, eMule 등),
웹캠, 동영상 스트리밍, 온라인 게임, VoIP, 모바일 앱 등
다양한 서비스들을 포함한다.

※출처: 나무위키(<https://namu.wiki/인터넷>)





WWW의 탄생

1989년 3월, CERN(유럽 입자 물리 연구소)의
소프트웨어 공학자 팀 버너스리는 CERN에서 인사 재배치 등으로
기존에 수행했던 실험 결과를 비롯한 각종 문서들이 유실되는 비율이
높은 것을 보고 이를 줄이기 위해
Information System: A Proposal을 제안하였다.

또한 여러 연구기관에 흩어져 있는 문서들을 체계화하여
전 세계의 대학 및 연구소들끼리 정보를 신속하게 교환할 수 있도록
해야 한다고 판단하여 문서 뿐만 아니라 소리, 동영상 등을 망라하는
데이터베이스를 구축하고 이를 전문 열람 소프트웨어로 열람하는
방식을 생각해 냈다.



인터넷과 WWW

워낙 WWW가 대세이기에 WWW를 인터넷으로 착각하는 경우가 많지만, 웹은 TCP/IP 기반 물리적 통신망인 인터넷을 활용한 서비스로 인터넷의 하위 개념으로 볼 수 있다. 위키백과에 따르면 WWW은 다음 세 가지의 기능으로 요약할 수 있음

첫 번째, 통일된 웹 자원의 위치 지정 방법

→ 예를 들면 URL(Uniform Resource Locator)

두 번째, 웹의 자원 이름에 접근하는 프로토콜(protocol) → 예를 들면 HTTP

세 번째, 자원들 사이를 쉽게 항해할 수 있는 언어 → 예를 들면 HTML

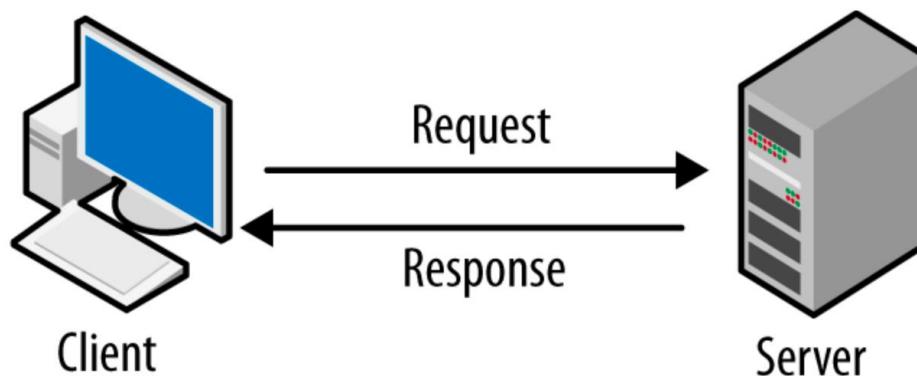




HTTP

Hyper Text Transfer Protocol은 W3 상에서 정보를 주고받을 수 있는 프로토콜(규약)

- HTTP는 클라이언트와 서버 사이에 이루어지는 요청/응답 (request/response) 프로토콜(규약)
- 클라이언트인 웹 브라우저가 HTTP를 통하여 서버로부터 웹페이지(HTML)나 그림 정보를 요청하면, 서버는 이 요청에 응답하여 필요한 정보를 해당 사용자에게 전달
- 이 정보가 모니터와 같은 출력 장치를 통해 사용자에게 나타나는 것서 흔히 볼 수 있는 htm이나 html 확장자가 바로 이 언어로 작성된 문서



※그림 출처: <https://velog.io/@seosu2000/Client-Server란 무엇인가>





HTML

< > ... </>

<https://namu.wiki/w/HTML>





HTML

웹사이트의 모습을 기술하기 위한 마크업 언어

- 프로그래밍 언어가 아니라 마크업 정보를 표현하는 마크업 언어로 문서의 내용 이외의 문서의 구조나 서식 같은 것을 포함
- HTML의 ML이 마크업 언어라는 뜻으로 웹사이트에서 흔히 볼 수 있는 .htm이나 .html 확장자가 바로 이 언어로 작성된 문서

```
<!DOCTYPE html>
<html>
  <head>
    <meta charset="utf-8">
  </head>
  <body>
    Hello, world!
  </body>
</html>
```





HTML은 ‘정보전달’이 주목적

디자인 요소

UI



HTML은 ‘정보전달’에 충실

디자인 요소

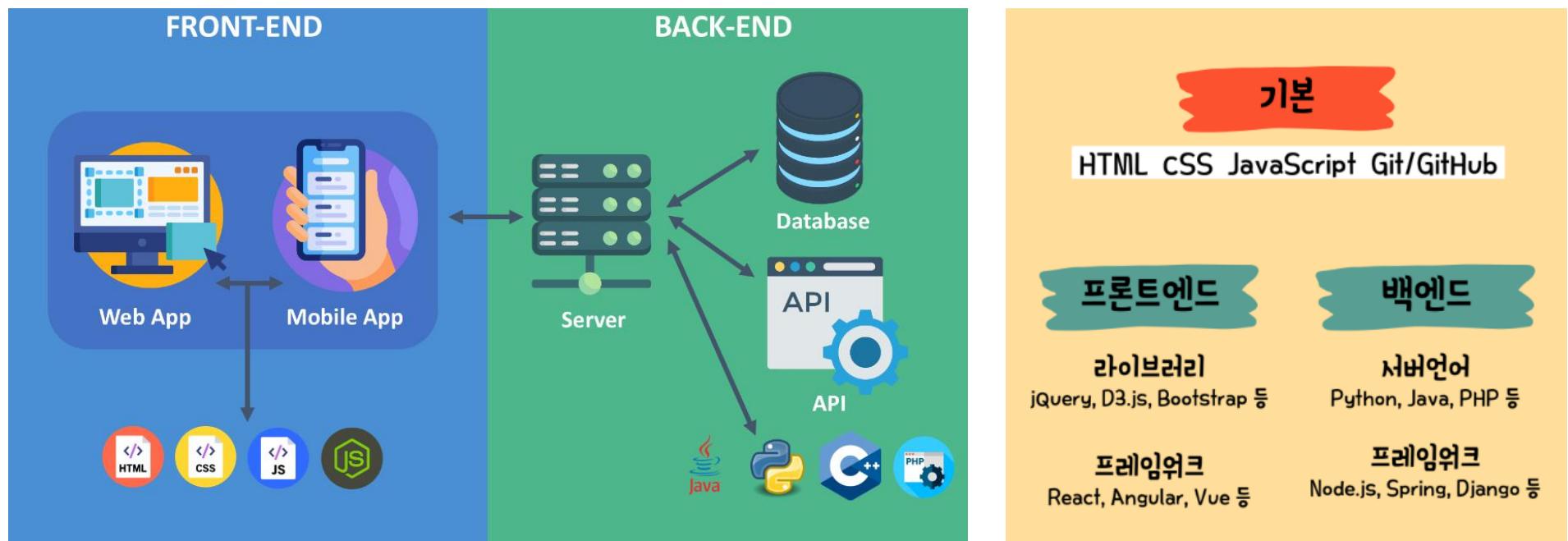
CSS (Cascading Style Sheet)

UI

JavaScript



웹개발은 크게 기본, 프론트엔드, 백엔드 등 3가지 영역으로 구분



※그림 출처: <https://velog.io/@xenxxn/01>, <https://1.pearlvely.com/5>



검색 엔진 최적화

검색엔진최적화(Search Engine Optimization)는 검색엔진으로부터 웹사이트나 웹페이지에 대한 트래픽의 품질과 양을 개선하는 과정

- 웹 페이지 검색엔진이 자료를 수집하고 순위를 매기는 방식에 맞게 웹 페이지를 구성해서 검색 결과의 상위에 나올 수 있게 함
- 웹 페이지와 관련된 검색어로 검색한 검색 결과 상위에 나오게 된다면 방문 트래픽이 늘어나기 때문에 효과적인 인터넷 마케팅 방법 중의 하나이며 비용처리 없는 마케팅이라고 할 수 있음
- 기본적인 작업 방식은 특정한 검색어를 웹 페이지에 적절하게 배치하고 다른 웹 페이지에서 링크가 많이 연결되도록 하는 것
- 구글 등장 이후 검색 엔진들이 콘텐츠의 신뢰도를 파악하는 기초적인 지표로 다른 웹사이트에 얼마나 인용되었나를 사용하기 때문에 타 사이트에 인용되는 횟수를 늘리는 방향으로 최적화함





웹 데이터 기반

단 3줄로 가능한 쿠팡 상품 댓글 크롤링

1. Python 설치 (<https://dotiromooook.tistory.com/32>)
2. 앱축 폴더 앱축 해제
3. 폴더 경로 복사
4. 명령 프롬프트 또는 “**cmd**” 실행 → DOS 화면 뜸
5. > cd 폴더 경로 붙여 넣고 → Enter → 해당 폴더로 이동
6. 해당 폴더 > pip install -r ./requirements.txt → Enter
7. 크롤링 할 쿠팡 상품 댓글 페이지 url → Copy
8. > python main.py → Enter
9. 복사된 url 주소 붙여넣기 → Enter → 페이지 수 지정 → Enter
10. 생성된 엑셀파일 열어서 확인





“꾸준한 연습과 반복을 권합니다.”

1월
22일

암기만 하는 것의 문제는
스스로 생각하지
않는데 있다.



爲學日益
位學日益

古之爲文者字字而辨其旨則使中其理後世誦成句而勲其全文莫猶古也
고지위문자자자이변기지즉사중기리후세송성구이표기전문학유고야

옛날 사람들은 글자마다 뜻을 헤아려 이치에 맞게 썼다.
후세에는 만들어진 글을 외워 그대로 쓴다. 그래서 글이 옛날만 못하다.
〈여유당전서〉

1월
23일

좋은 문장은 좋은 글들을
무수하게 흉내 내다가
결국에는 자신만의 글에
도달했을 때 나온다.



位學日錄

人之有文章 猶草木之有榮華耳
榮華不可以襲取之也 文章不可以襲取之也

인지유문장 유초폭지유영화이
영화불가이습취지야 문장불가이습취지야

사람이 문장을 지님은 초목에 꽂이 피는 것과 같다.
꽃은 갑작스레 얻을 수 없다. 문장도 갑작스레 얻을 수 없다.
(여우단정선)