

AI 융합전문가
(기초과정 3차시)

데이터분석

2023년 9월 17일



강사 소개_ 홍용기 컨설팅학 박사



| 소속 | 주요 업무 |
|----------------|--|
| (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, 데이터바우처 사업수행 |

● 자격증

1. 경영지도사 (인적자원, 2016)
2. 창업보육매니저 (BI협회, 2018)
3. 기업·기술가치평가사 (기업·기술가치평가협회, 2018)
4. 기업재난관리사 실무과정 (행정안전부, 2019)
5. 데이터분석 준전문가 ADsP (데이터산업진흥원 K-Data, 2021)
6. 빅데이터 분석기사 (과학기술정보통신부 · 통계청, 2021)
7. 국제공인컨설턴트 CMC (ICMCI, 국제경영컨설팅협의회, 2023)

강사 소개_ 홍용기 컨설팅학 박사



● 학력

- 79.03 ~ 82.02 덕수상업고등학교 졸업
- 02.03 ~ 06.02 한국방송통신대학교 법학과 졸업(법학사)
- 17.03 ~ 19.02 한성대학교 지식서비스컨설팅학과 졸업(컨설팅학 석사)
- 19.03 ~ 22.02 한성대학교 스마트융합컨설팅학과 졸업(컨설팅학 박사)

● 주요 경력

- 1985 ~ 1995 삼풍백화점, 동아씨티백화점, 태평백화점 등 신규개점 프로젝트 외
- 1996 ~ 2000 프라임그룹 강변테크노마트 신규개점 프로젝트 외
- 2008 ~ 2010 코스닥상장기업 관리이사(CFO)
- 2012 ~ 2015 건설회사 관리담당 상무
- 2017 ~ 현재 컨설턴트

컨설팅 시장 진입이 안되고 어려웠던 시절

한국경제

Hankyung.com
기업을 찾는 생각

알고 보면 재미있는 경영이야기 - 기업 경영과 자본

▲기업형 출판기 전문위원

누구나 알고 있는 것처럼 기업을 경영하면서 어렵거나 자본이 필요하다. 그게 사실로는 대체로 맞지만, 자본과 함께 자본으로 나눌 수 있는 자본의 가장 대표적인 주체가 바로 자본이다.

자본은 우리가 흔히 부정하고 부정하는 민간자본과 자가자본으로 나눌 수 있다. 그러면 기업은 입장에서 민간자본은 이를 자금화하려고 또 다른 통로를 찾고, 개인에게는 개인으로 찾고, 그리고 그에 따른 통로를 찾고자 한다. 그래서 기업은 이익이 날으면 투자를 하고 그 결과로 많은 돈을 벌거나, 나아가 성사 회사가 될 것을 염두하고 상장을 목표로 한다. 다른 투자를 확장하고자 노력하는 경향에 대해서 있다면 그에 대한 통로를 찾구려는 가설이 있다.

따라서 기업의 경영 성장을 나타내는 재무제표의 표기방식에 있어서도 이런 특징과 적용된다. 기업의 재무상황을 나타내는 '자본구조'에서도 부정과 자본이 표시되는 우측 한의 표시 속에서 외부자본과 내부자본에 중요시 한다는 의미로 자본보다 부정과 위험에 표시된다.

일례 기간의 경영실력을 나타내는 '순이자율'의 경우에도 마찬가지다. '복리이자보상비율'이다.

마음에서 이익을 감출하는 순서에 있어서도 경영이익에서 민간자본에 대한 이자비용을 먼저 처리한 이후에 회동적으로 당기손이익을 감출한다.

그런데 중소기업을 경영하는 대표 혹은 경영주는 민간자본의 부정에 대해서는 유연하게 이를 저금화하지만, 자기자본에 대해서는 그 비용이 발생하기 어렵기 때문에 찾을 때는 본인이 의도한 바와는 정반대로 저기자본에 대해서는 저기자본에 대해서도 비용이 발생하고 그에 대한 비용 겝수 등을 수립해야 안전적인 경영을 할 수 있다.

입장에 따라서 자본금을 충자하는 주주, 투자가 믿었는데 생각에서도, 회사에 대해서 민간자본처럼 영향력을 행사할 수 있는 이자율을 받을 수 있다. 예전에는 경영실적이 좋지 않아 나쁜 회사에 투자를 하거나 투자하지 않으면 회사를 본사를 찾을 경우 투자 원금과 회수액에 대한 보상이 되기 때문에 투자가 본인에 투자하는 이유로는 민간자본보다 높다고 할 수 있다.

위험률이 높다고 하는 것은 위험이 높을 때 요구하는 수익률이 높거나 아니이고, 투자자 회사의 자본으로 투자하는 투자자는 회사보다 더 높은 수익률을 요구하는 것이 자국의 민족관이다. 그에 따라 자기자본은 비록 민간자본에 비해보다 더 높을 것이다.

자본비용은 민간자본 비용과 자기자본 비용을 가중평균하는 방식으로 산출하게 된다. 가중평균자본비용(WACC)을 간단히 설명하면, 경영자들은 민간자본 비용이 7%이고, 민간자본 비용과 자기자본 비용이 각각 5%, 10%라고 가정할 때,

$$\text{WACC} = 0.4 \times 5\% + 0.6 \times 10\% = 8\%$$

WACC은 자본비용과는 산출하는 데나 경영자는 활용도 한다. 다만 WACC이 가중평균 자본비용이라는 뜻은 이해하기 어렵다. 나아가 경영자는 예상 수익률을 계산하는 시점에서는 이를 것으로 기업 가치를 평가해서 주가적인 대출여부를 판단한다는 사실. 중소기업 대표님들은 명심하시길.

한국경제

한경 경영지원단

성장하는 기업의 5가지 조건

한치호·홍용기·이민석·최종렬 지음

한국경제신문

성장하는 기업의 5가지 조건

한치호·홍용기·
이민석·최종렬
지음

“어떻게 지속적으로 성장할 것인가?”

끌이 보이지 않는 저성장 시대의 위기를 돌파하고
무한 경쟁 시대에도 살아남는 전략을 알려준다!

한국경제신문

- 4 -



→ 컨설팅 1년 차

● 2017년

03 한성대학교 컨설팅대학원 석사 입학 (원우회장)

04 창업멘토링 (협성대)

05 특강 (여주대학교: 창업과 경영)

07 기술사업화 강의 (충남권 화학소재 중소기업 역량강화 세미나_보령 웨스토피아)

07 기술사업화 강의 (경제협력권 산업육성 R&D 3차년도 세미나_대전 한화리조트)

08 석사 2학기

08~10 팀티칭 8주 출강(여주대학교: 보안안전경영과/세무회계정보과)

09 기술사업화 강의 (제품단위 포트폴리오 컨소시엄 제3차 연구회_제주 리젠틴마린)

10 경영지원 멘토링 (창업보육센터_을지대학교 성남캠퍼스)

10 사업계획고도화 강의 (창조경제타운_상암DMC)

11 기술사업화 창업교육 (KRISO_대전 선박해양플랜트연구소)

11 비즈니스모델수립 멘토링 (4차 산업 융합프로그램_성결대학교)

12~18.03 초정밀 접목로봇 보급사업 경제성 분석 (농업기술실용화재단)



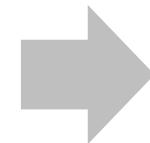
지구파트너스 설립



→ 컨설팅 2년 차

● 2018년

- 01~02 NCS 일학습병행제 훈련교재개발 (남서울대학교)
- 03 석사 총원우회장 (7개 특수대학원 전체, 석사 1,000여 명)
- 04 기술사업화 지원사업 신청업체 서면평가/대면평가 (대전경제통상진흥원)
- 05 공공기관 외부면접관 (LH공사 채용 및 정규직 전환_집단면접 및 개별면접)
- 05 TMC사업 우수기업집중지원 전문가그룹컨설팅 (주)대경기술지주)
- 06 공공기관 외부면접관 (한국가스공사 신규채용_개별면접)
- 06~08 중소기업 컨설팅지원사업 (중소벤처기업부 쿠폰제 컨설팅)
- 06~08 사회적기업 경영컨설팅 (사회적기업진흥원)
- 08 석사 4학기
- 09 통계 프로그램 SPSS, AMOS 등 습득
- 11 스마트팩토리 관련 시뮬레이션 프로그램 강의 (국방전직교육원, 6일)
- 11~12 베트남 창업진출 전략보고서 (한국옐로우페이지)



R 프로그램 영접



→ 컨설팅 3~4년 차

● 2019년

- 01~04 해상풍력발전사업 사업타당성 및 경제성분석 (현대그룹)
- 03 한성대학교 박사 입학 (원우회장)
- 05 공공기관 외부면접관 (LH공사 신규채용)
- 05 창업캠프 "창업 해보젠했나?" (제주관광대학교, 2박3일)
- 05~12 축산 ICT 융복합확산사업 및 빅데이터 컨설팅 (농림수산식품교육문화정보원)
- 08~09 글로벌 액셀러레이터 MAP 제작 연구용역 (창업진흥원)
- 09~20.03 안산시 창업생태계 조성사업 (안산시 & 행정안전부)

● 2020년

- 05~08 ISP 컨설팅 (창업진흥원: 창업기업 확인시스템 구축)
- 08 청주시 창업보육센터 활성화사업 집중멘토링 (충북대 창업지원단)
- 10 창업프로젝트 경제성확인 동영상 강의 (강원창조경제혁신센터 G-아카데미, 비대면)
- 11 소상공인·소기업을 위한 인사조직 강의 (강동구 사회적경제지원센터, 비대면)
- 10~11 (주)SMT랩 기업가치평가 (중소기업 제조혁신바우처사업)
- 10~12 ISP 컨설팅 (한국연구재단: 중장기 정보화전략계획)



ISP 컨설팅 시작



→ 컨설팅 5~6년 차

● 2021년

- 01~04 ISP 컨설팅 (소상공인시장진흥공단: 소상공인지원사업 디지털전환)
- 06~12 강의 (농림수산식품교육문화정보원: 청년후계농)
- 07~09 연구용역 (경기도: 자동차부품 공급 및 수요현황조사)
- 11~22.06 ISP 컨설팅 (서울시: 차세대지방세입정보시스템 통합구축 변화관리)

● 2022년

- 01 데이터바우처지원사업 공급기업 지정평가 (K-Data)
- 02 **한성대학교 졸업 (컨설팅학 박사)**
- 03 AI학습용데이터 구축사업(1차) 평가 (NIA: 08, 09, 43, 48 데이터)
- 04~11 **강의 (농림수산식품교육문화정보원: 청년후계농 대상)**
- 04~12 **산업전문인력 AI역량강화지원사업 (정보통신산업진흥원 NIPA)**
- 04 AI학습용데이터 구축사업(2차) 평가 (NIA: 74, 75 데이터)
- 05 AI학습용데이터 구축사업(3차) 평가 (NIA: 06 데이터)
- 06~10 **데이터바우처지원사업 수행 (K-DATA)**
- 07~12 **연구용역 (광양만권경제자유구역청: EV 관련 사업모델 구축)**
- 08 빅데이터 이론 및 R코딩 강의 (경영기술지도사회, 4일 32시간)
- 08 빅데이터센터 추가선정 평가 (NIA: 스마트팜 &문화 부문)
- 08~12 ISP 컨설팅 (경찰대학: 국민체감형 치안안심플랫폼 구축)



NIA 전문위원 등록



데이터바우처 PM



→ 컨설팅 7년 차

● 2023년

- 01 (주)제타데이터 설립
- 03 강의 (경영기술지도사회: 빅데이터 이론 및 R코딩)
- 03~10 강의 (농림수산식품교육문화정보원: 청년후계농)
- 05 강의 (서울여성능력개발원 강동센터: 데이터분석 전문가 양성과정 중 Python 심화)
- 06~11 데이터바우처지원사업 수행 (K-DATA)
- 07 강의 (서울여성능력개발원 용산센터: 데이터분석 전문가 양성과정 중 Python 심화)
- 08~ 요르단 빅데이터 기반 치안정보 통합관리 시스템 구축사업 → ODA 사업 영접
- 09 데이터 가치평가사업 수행 (중소벤처기업진흥공단)



데이터 관련 비즈니스 수행 실적



발급번호:00KIH83K-W6YQ-OA64-CG1Z

● 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
23.08~24.08. 요르단 경찰청 ODA

● 2022년 AI학습용데이터 구축사업 평가

- 1차 08 방송콘텐츠 대화체 음성인식 데이터
09 방송콘텐츠 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~ 중소벤처기업진흥공단

소프트웨어기술자 경력증명서

| 성명 | 홍용기 | | 생년월일 | 1964.08.25 | | |
|-------|-------------------|----------------------------------|-------------------------|-----------------------------------|--------|--------------------|
| 현 근무처 | 회사명 | | 사업자등록번호 | | | |
| | 전화번호 | | 업종 | | | |
| | 소재지 | | | | | |
| 근무경력 | 확인여부 | 근무기간 | 회사명 | 담당업무 | | |
| 기술자격 | | | | 부서/직위 | | |
| | 종목 및 등급 | 등록번호 | 취득일 | 발급기관 | | |
| | 빅데이터분석기사 | BAE-002000023 | 2021.07.16 | 한국데이터산업진흥원 | | |
| 학력 | ADsP(데이터 분석 준전문가) | ADsP-028000961 | 2021.04.09 | 한국데이터산업진흥원 | | |
| | 학교명 | 학과(전공) | 수학기간 | 학위 | | |
| | 기간 | 과정 | 수료번호 | 교육기관 | | |
| 교육 | 수여일 | 종류 | 상훈기관 | 근거 | | |
| 상훈 | 확인여부 | 참여사업명 | 참여기간 | 발주자 | | |
| | 확인 | 국민 체감형 치안 안심 플랫폼 구축 정보화전략계획 사업 | 2022.08.04 ~ 2022.12.31 | 경찰대학 (주)메타로 직컨설팅 | 프리랜서 | IT컨설팅 > 정보기술기획 |
| | 확인 | 서울시 차세대 지방세 입정보시스템 통합구축·변화관리 컨설팅 | 2021.11.01 ~ 2022.06.30 | 서울특별시 (주)메타로 직컨설팅 | 프리랜서 | IT컨설팅 > 정보기술컨설팅 |
| | 확인 | 소상공인지원사업 디지털전환 정보화전략 계획(ISP) 용역 | 2021.01.04 ~ 2021.05.03 | 소상공인 시장진흥 공단 (주)메타로 직컨설팅 | 수석컨설턴트 | IT컨설팅 > 정보기술컨설팅 |
| | 확인 | 한국연구재단 중장기 정보화전략계획(ISP) 수립 | 2020.10.05 ~ 2021.01.04 | 한국연구 재단 (주)메타로 직컨설팅 | 수석컨설턴트 | IT컨설팅 > 정보기술컨설팅 |
| | 확인 | 창업기업평가시스템 구축을 위한 정보화전략계획(ISP) 수립 | 2020.05.18 ~ 2020.08.17 | 창업진흥 원 (주)메타로 직컨설팅 | 수석컨설턴트 | IT컨설팅 > 정보기술컨설팅 |
| | | | | | | |

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

2023년 01월 25일

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.



What to do?



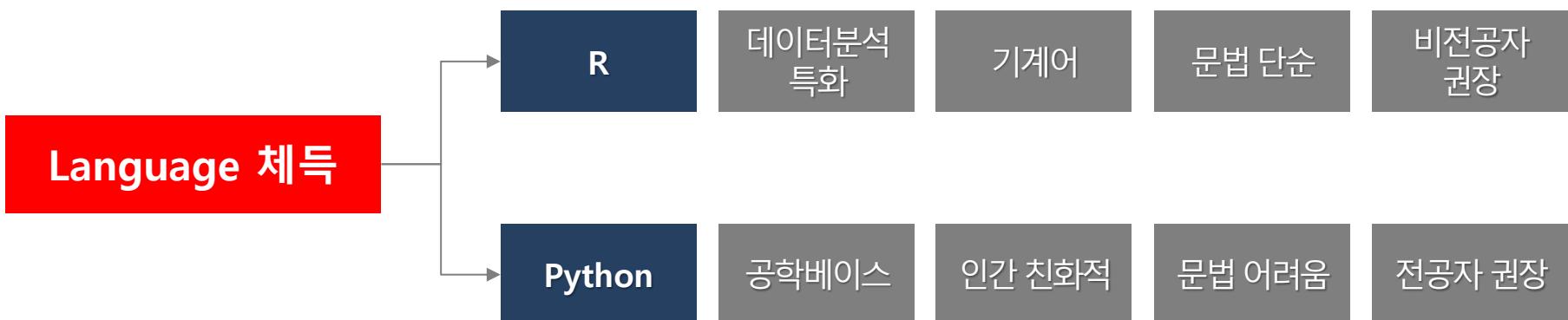
혹자는
인생 뭐 별거 있냐고 할테지만...

인생은 어쩌면
수많은 연결고리들의 집합일지도...

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



How to do?



2023년 데이터 분석 자격증 시험일정

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



| 전체 | 빅데이터분석기사 | 데이터분석 | SQL | 데이터아키텍처 | 접수가능 시험 |
|----|----------|-------|-----|---------|---------|
|----|----------|-------|-----|---------|---------|

| 구분 | 회차 | | 접수기간 | 수험표발급 | 시험일 | 사전점수발표 | 결과발표 | 증빙서류 제출기간 |
|--------------|-----|----|------------|-------|---------|--------|-------|-------------|
| 빅데이터 분석기사 | 제6회 | 필기 | 3.6~10 | 3.24 | 4.8(토) | 4.21 | 4.28 | 5.2~11 |
| | | 실기 | 5.22~26 | 6.9 | 6.24(토) | 7.7 | 7.14 | - |
| | 제7회 | 필기 | 8.21~25 | 9.8 | 9.23(토) | 10.6 | 10.13 | 10.16~10.26 |
| | | 실기 | 10.30~11.3 | 11.17 | 12.2(토) | 12.15 | 12.22 | - |

| 전체 | 빅데이터분석기사 | 데이터분석 | SQL | 데이터아키텍처 | 접수가능 시험 |
|----|----------|-------|-----|---------|---------|
|----|----------|-------|-----|---------|---------|

| 구분 | 회차 | | 접수기간 | 수험표발급 | 시험일 | 사전점수발표 | 결과발표 | 증빙서류 제출기간 |
|-------------------------------|------|---|---------|-------|----------|--------|-------|-----------|
| 데이터분석 준전문가 AISPA | 제36회 | - | 1.25~31 | 2.10 | 2.26(일) | 3.17 | 3.24 | - |
| | 제37회 | - | 4.17~21 | 5.4 | 5.20(토) | 6.9 | 6.16 | - |
| | 제38회 | - | 7.17~21 | 8.4 | 8.19(토) | 9.8 | 9.15 | - |
| | 제39회 | - | 9.18~22 | 10.6 | 10.21(토) | 11.10 | 11.17 | - |

Artificial Intelligence

Machine Learning

Deep Learning

Generative AI



Artificial Intelligence

인간의 지능을 복제하거나 능가할 수 있는 지능형 기계를 만들고자 하는 컴퓨터 과학 분야



Machine Learning

기계가 기존 데이터에서 학습하고 해당 데이터를 개선하여 의사 결정 또는 예측을 할 수 있도록 하는 AI의 하위 집합



Deep Learning

신경망 레이어를 사용하여 데이터를 처리하고 결정을 내리는 기계 학습 기술



Generative AI

프롬프트나 기존 데이터를 기반으로 새로운 문서, 시각, 청각 콘텐츠를 생성하는 기술



기술의 가치

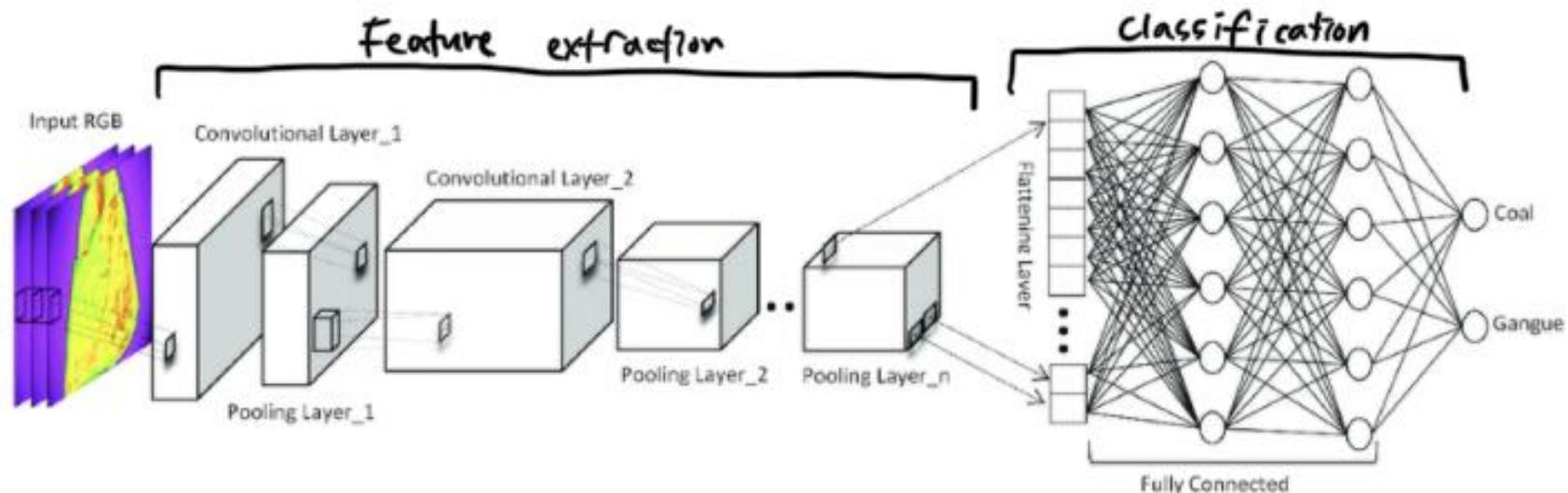
“가치를 아는 사람에게 기술이 가야 빛을 발한다.”

WWW



Deep learning – Convolutional Neural Network; CNN

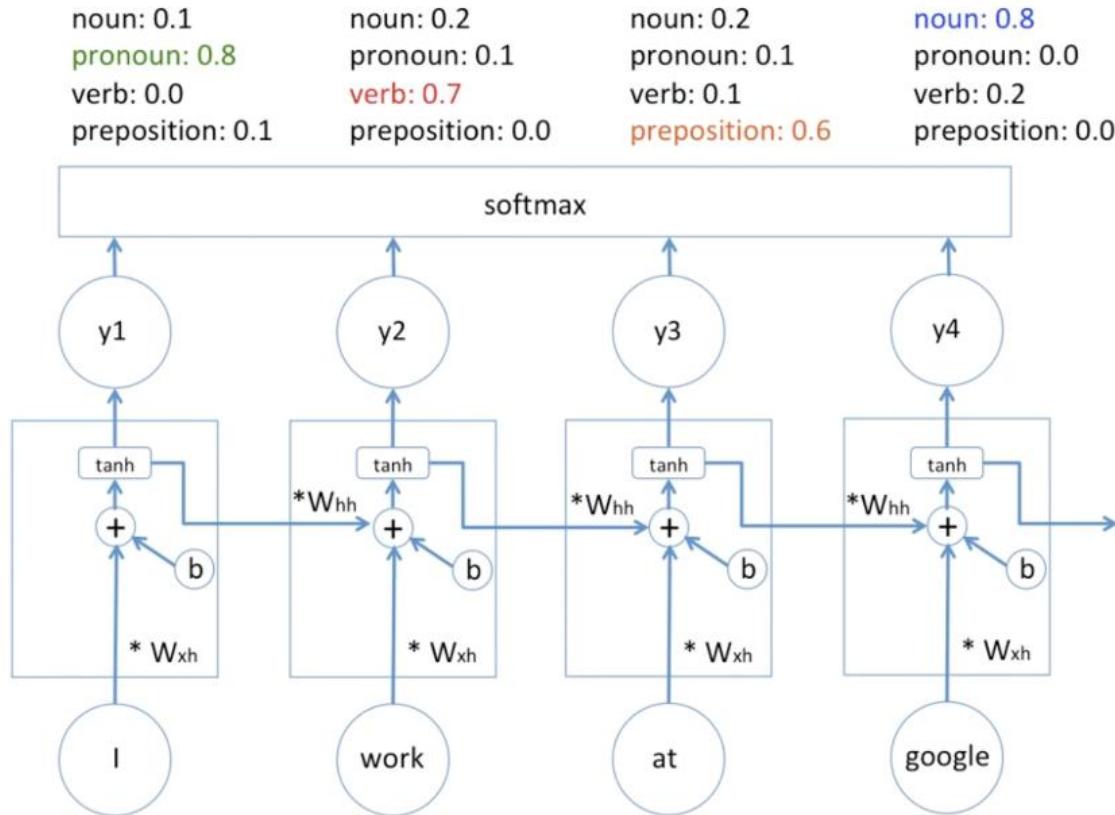
CNN 구조



AlexNet, GoogleNet, ResNet, DenseNet, Region Based CNNs, CNNs for NLP ...

※출처: 블로그 CNN 주요 모델들 (17.10, <https://ratsgo.github.io/deep%20learning/2017/10/09/CNNs/>)

Deep learning – Recurrent Neural Network; RNN



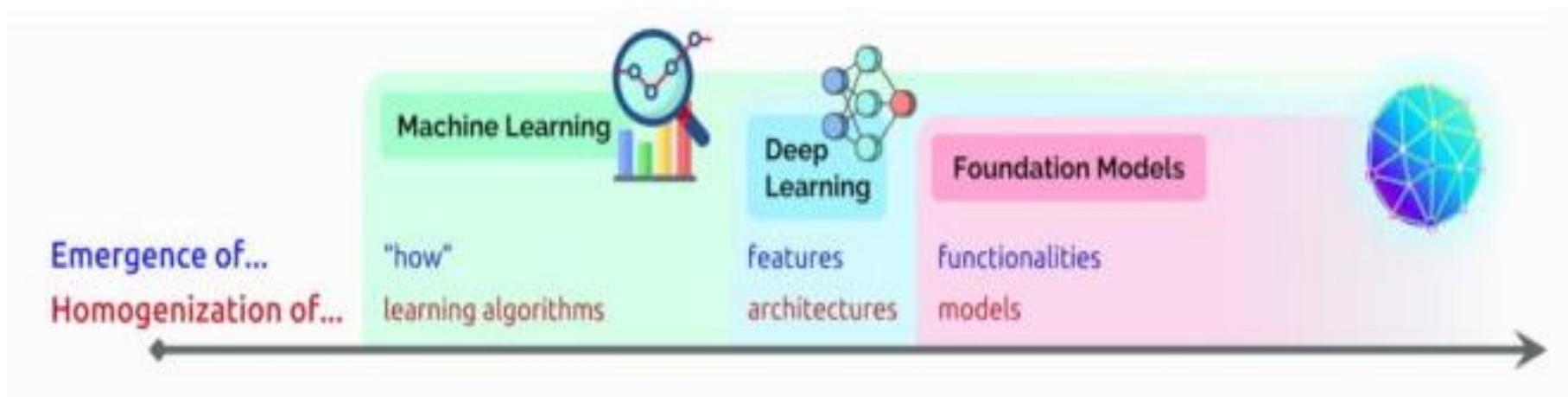
멀리 떨어져 있는
시간 단계와의 의존성이
예측에 중요한 특징이 될
때가 많음에 따라
LSTM*, GRU와 같이**
RNN 변형구조 설계

(참조) * Long Short-Term Memory (장단기 메모리)** Gate Recurrent Unit (게이트 순환장치)

※출처: 블로그 [딥러닝] RNN이 뭘까? (22.01, <https://velog.io/@hhong/RNN>)

Foundation Model

- 2021년 여름, 100명이 넘는 스탠포드 연구자로 구성된 팀이 214페이지 분량의 논문 발표



이 논문에서 연구자들은 트랜스포머(transformer) 모델과 대규모 언어 모델(LLM), 구축이 계속되고 있는 기타 신경망들이 일명 **파운데이션 모델(foundation model)**이라는 새롭고 중요한 범주를 구성한다고 분석

History of the foundation model

- 2017년 NIPS*(인공지능 세계최고권위 학회)를 통해 트랜스포머(Transformer) 발표
- 그전까지 NLP** 분야에서 사용되어 오던 인공신경망 구조(주로 RNN)가 아닌 "Attention"이라는 메커니즘을 활용해 번역부분 SOTA***를 차지하며 화려하게 데뷔
- 그 후, 분야를 넓혀 이미지 분류나 이미지 인식 등 NLP를 넘어 컴퓨터 비전분야까지, 당시 인공지능계 핫이슈로 부각

(참조)

* Neural Information Processing Systems

** Natural Language Processing (자연어 처리)

*** State-of-the-art (현재 존재하는 최고 수준의 결과)



※출처: 블로그 Transformer는 이렇게 말했다, "Attention is all you need."(21.03, <https://blog.promedius.ai/transformer/>)

History of the foundation model

- 트랜스포머는 BERT* 같은 대규모 언어 모델에 영감 주었고, AI 관련 논문의 표현을 빌리자면 '2018년을 자연어 처리의 분수령'으로 만듦
- 구글이 BERT를 오픈 소스 소프트웨어로 출시했고, 이는 후속 제품군의 탄생으로 이어졌으며, 더 크고 강력한 언어 모델을 구축하기 위한 경쟁 시작
- 2020년 오픈AI(OpenAI)의 연구진이 또 하나의 기념비적인 트랜스포머 모델인 GPT-3를 발표
- GPT-3는 약 1조 개에 가까운 단어를 포함한 데이터세트로 훈련. 파라미터의 수가 1,750억 개에 달함 ↪ GPT-4는 170조 개

(참조)

* Bidirectional Encoder

Representations from Transformers

구글이 공개한 사전 훈련된 AI
언어모델로 자연어를 이해하기
위한 양방향 학습모델 모두
지원하는 알고리즘



History of the foundation model

- 가장 최근의 ChatGPT는 엔비디아 GPU 1만 개로 훈련했으며, 두 달만에 1억 명 이상의 사용자 확보
- 이 모델의 출시는 관련 테크놀로지의 활용법을 많은 이들에게 선보였다는 점에서 '아이폰 모멘트 (iPhone moment)'라고도 불림

텍스트에서 이미지로

ChatGPT가 데뷔하던 즈음, 또 다른 차원의 신경망인 확산 모델(diffusion model)이 큰 인기를 끌었습니다. 텍스트로 된 설명을 예술적인 이미지로 바꾸는 확산 모델은 놀라운 이미지들로 소셜 미디어에서 입소문을 타면서 일반 사용자와도 친숙해졌습니다.

확산 모델을 다룬 최초의 논문은 2015년에 조용히 발표됐습니다. 하지만 트랜스포머 모델들이 그랬듯 이 기법도 이내 열광적인 관심에 휩싸였죠.

옥스퍼드 대학에서 AI를 연구하는 제임스 손턴(James Thornton)에 따르면 작년 한 해에만 200개가 넘는 확산 모델 관련 연구가 발표됐습니다.

미드저니(Midjourney)의 데이비드 홀즈(David Holz) CEO는 트위터에 게시한 글에서 자사의 확산 모델 기반 텍스트 이미지 변환 서비스가 440만 명 이상의 사용자를 보유하고 있다고 밝혔습니다. 그는 한 인터뷰(로그인 필요)에서 이 서비스의 AI 추론에 1만 개 이상의 NVIDIA GPU가 활용된다고 말한 바 있습니다.

Theatre D`opera Spatial (스페이스 오페라 극장)

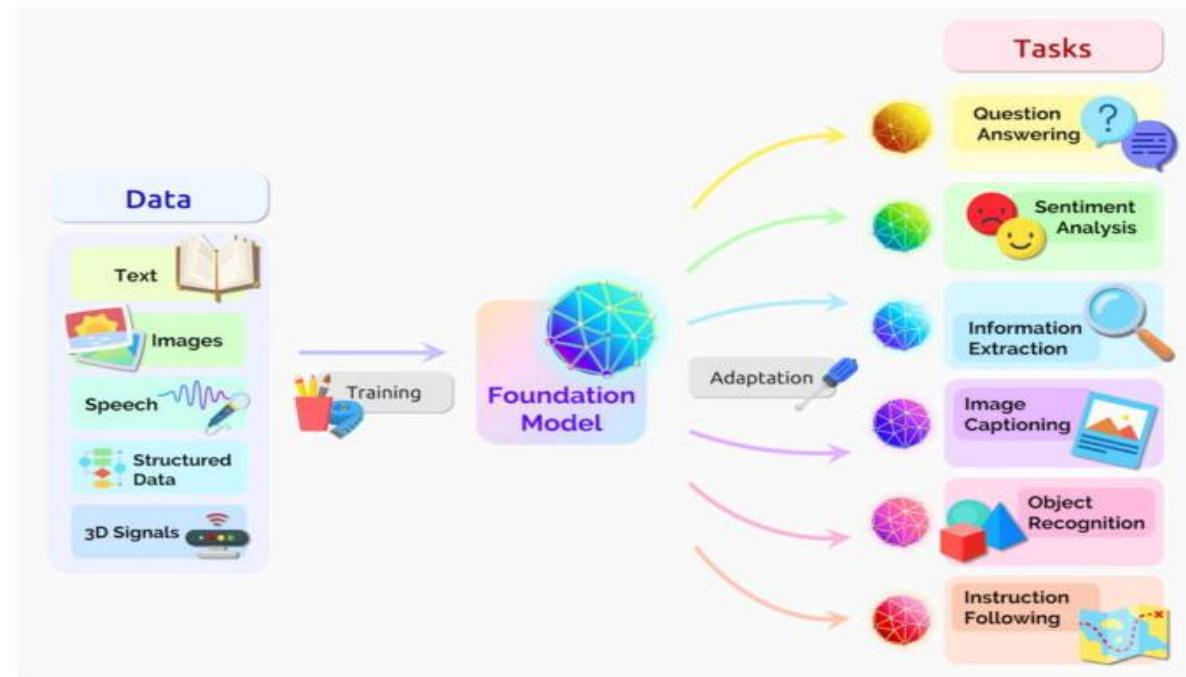


Definition of foundation model

- 논문에 따르면 파운데이션 모델은 산더미 같은 원시 데이터에서 대개 비지도학습(unsupervised learning)을 통해 훈련된 AI 신경망으로 광범위한 작업에 응용이 가능

파운데이션 모델을 정의할 때는
'더 쉬운 데이터 수집'과
'지평선 만큼 광활한 가능성'이라는
두 가지 개념으로 이해

파운데이션 모델은 일반적으로
레이블(label)이 없는 데이터 세트로
학습하므로 레이블링에 필요한
시간과 비용을 절감



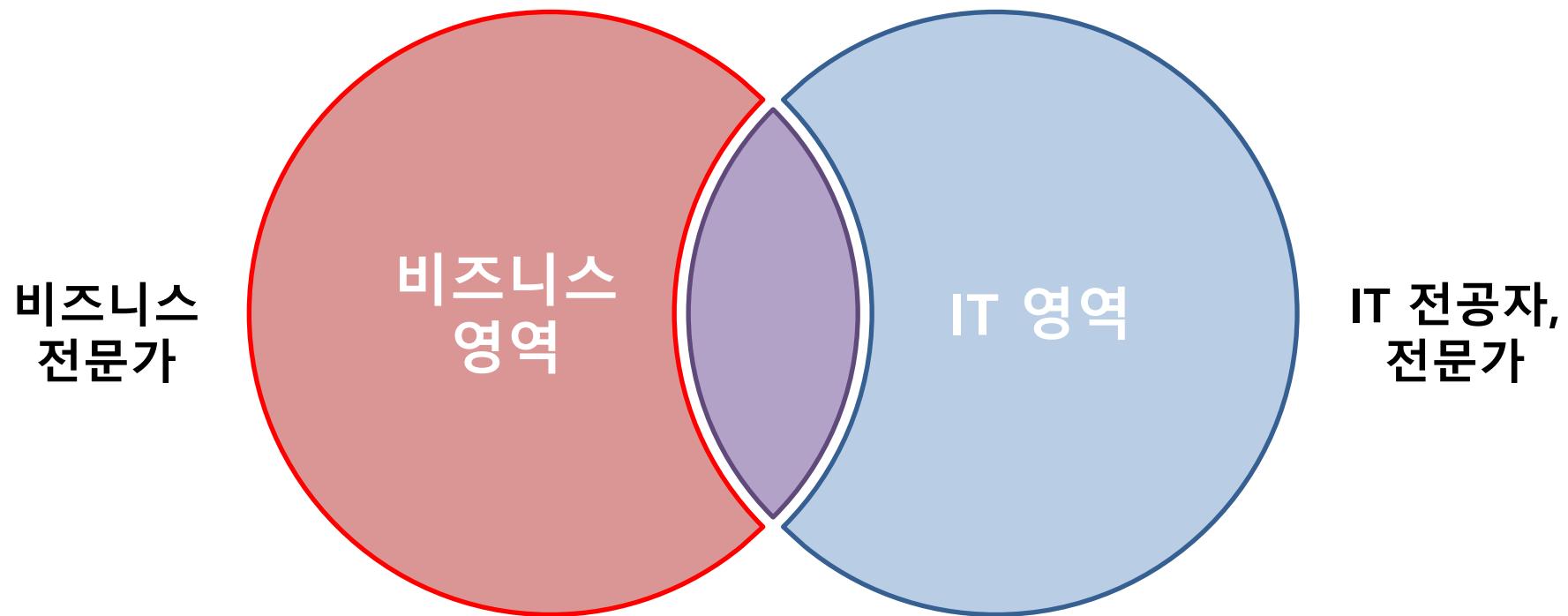
2021년 논문에서 연구자들은 파운데이션 모델이 다양한 용도로 사용되고 있다고 보고했습니다.



DATA

‘추론과 추정의 근거를 이루는 사실’

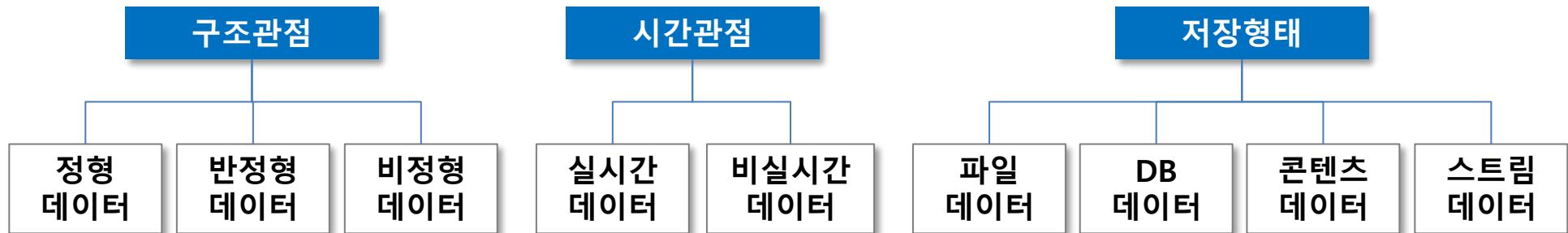
(A thing given or granted; something known or assumed as fact and made the basis of reasoning or calculation; an assumption of premiss from which inferences are drawn. OED, Vol. IV 264)



“컨설턴트는 어느 영역까지 cover 해야 할까요?”

‘데이터분석’이라고 하면
어떠한 데이터를 어느 범위까지
분석하는 것일까요?

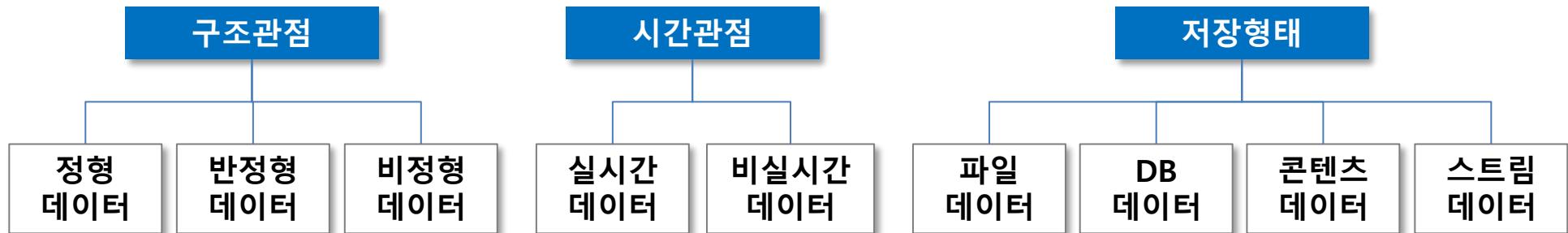
Data type



| 구조관점의 데이터 유형 |
|-----------------|
|-----------------|

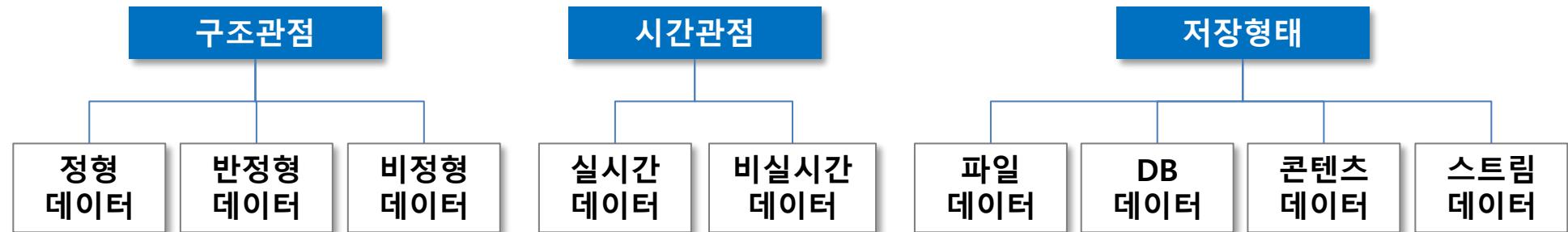
| 유형 | 설명 | 종류 |
|---------|--|-------------------------------|
| 정형 데이터 | 정형화된 스키마(형태) 구조 기반으로 행과 열로 구성 | 관계형 데이터베이스, 스프레드 시트 |
| 반정형 데이터 | 스키마 구조 기반이나 값과 형식에서 일관성을 가지지 않는 데이터 | XML, HTML, 로그데이터, JSON, 센서데이터 |
| 비정형 데이터 | 스키마 구조 형태를 가지지 않으며 고정된 필드에 저장되지 않는 데이터 | SNS, 게시판, 텍스트, 이미지, 오디오/비디오 |

Data type



| 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 |

Data type



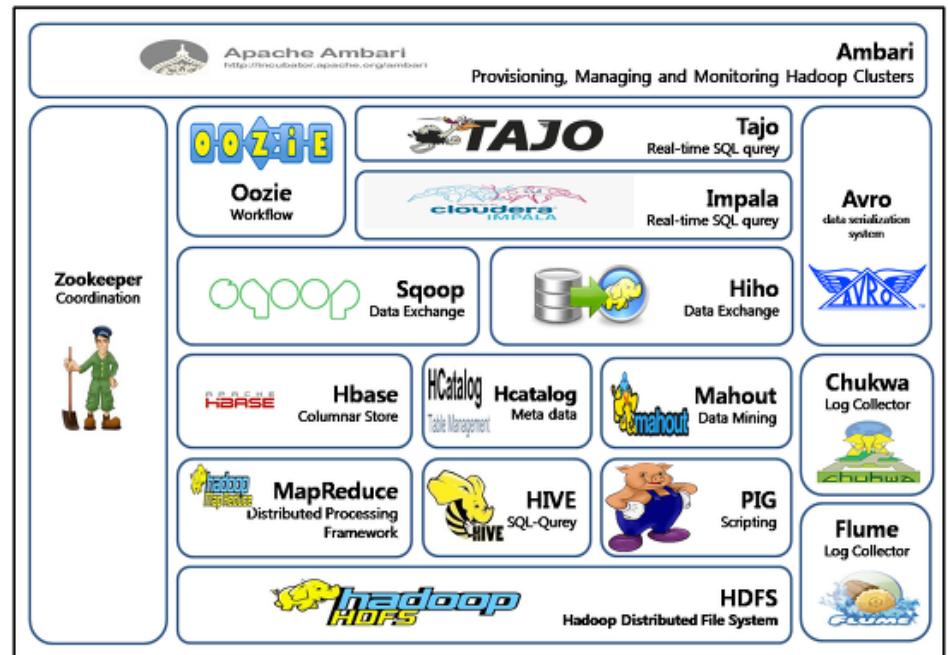
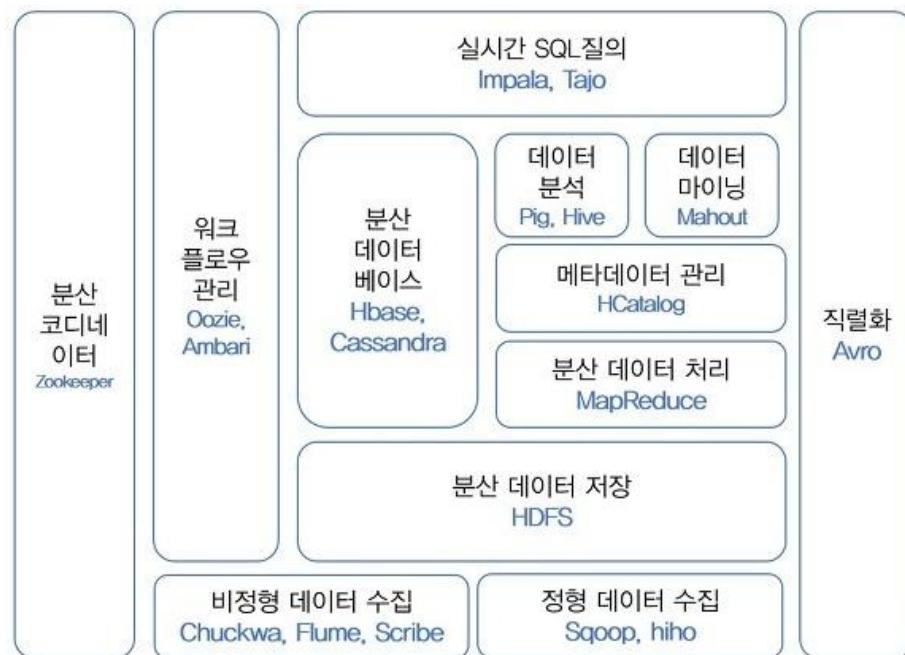
| 저장형태 관점의 데이터 유형 |
|-----------------------|
|-----------------------|

| 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 techniques

데이터 수집 방식 및 기술

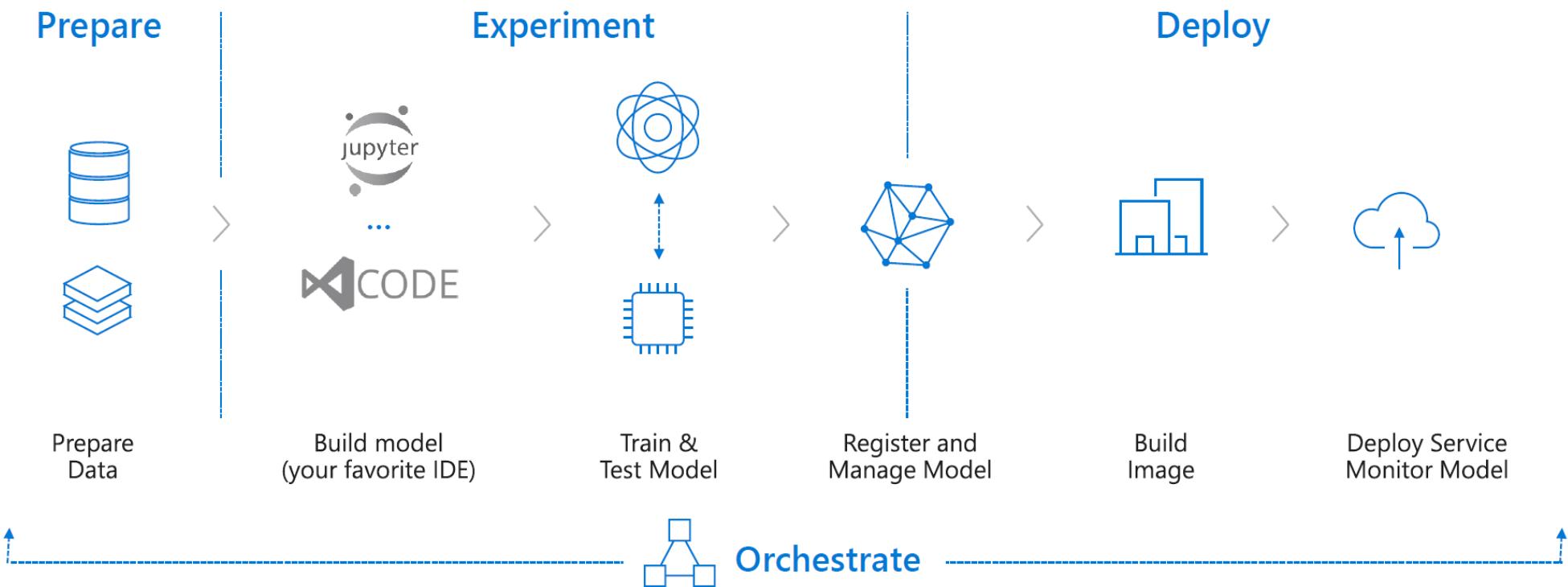
| | | | |
|--|---|---|--|
| <ul style="list-style-type: none"> ETL(Extract, Transform, Load) FTP(File Transfer Protocol) 스쿱(Sqoop) 스크래파이(Scrapy) | <ul style="list-style-type: none"> 아파치 카프카(Apache Kafka) 플럼(Flume) 스크라이브(Scribe) 척와(Chukwa) | <ul style="list-style-type: none"> CEP(Complex Event Processing) 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) |
|--|---|---|--|



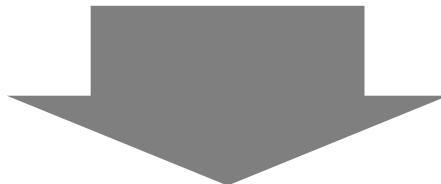
※출처: 하둡 에코시스템(Hadoop-Ecosystem)이란 <https://butter-shower.tistory.com/73>

Machine Learning

Typical E2E Process



데이터분석 = 데이터수집 + 실질적인 데이터 분석



“어느 것이 더 어려운가?”

어느 것이 더 많은 노력과 시간이 들어가나?

Python in Excel

Python in Excel.xlsx Public • Saved

File Home Insert **Formulas** Data Review View Comments Share Catch up

Insert Function Sum AutoSum Logical Text Date & Time Python Reset Diagnostics Defined Names Trace Precedents Trace Dependents Remove Arrows Watch Window Calculation Options Calculation

D3 PY #Announcing Python in Excel!
DataFrame=xl("A1:B10", headers=True)
DataFrame.groupby('Category').agg('mean')

| | A | B | C | D | E | F | G | H | I | J |
|----|-----------------|------|---|------------------|---|---|--------------|---|---|---|
| 1 | Category | \$ | | DataFrame | | | Image | | | |
| 2 | Components | \$20 | | | | | | | | |
| 3 | Bikes | \$17 | | | | | | | | |
| 4 | Accessories | \$ 9 | | | | | | | | |
| 5 | Bikes | \$ 9 | | | | | | | | |
| 6 | Clothing | \$ 8 | | | | | | | | |
| 7 | Accessories | \$ 4 | | | | | | | | |
| 8 | Clothing | \$ 4 | | | | | | | | |
| 9 | Components | \$ 3 | | | | | | | | |
| 10 | Components | \$ 1 | | | | | | | | |
| 11 | | | | | | | | | | |
| 12 | | | | | | | | | | |
| 13 | | | | | | | | | | |

Category

| Category | \$ |
|-------------|----|
| Accessories | 7 |
| Bikes | 13 |
| Clothing | 6 |
| Components | 8 |

Components Clothing Bikes Accessories

size
447, 274

ANACONDA.

Python in Excel

Display Settings

Seamlessly aggregate and visualize your data with Python in Excel.

Understanding the structure of Python code

● 웹크롤링 코드 구조 예시

```
# 모듈, 라이브러리, 클래스 import
import time
import pandas as pd
import numpy as np
from datetime import datetime as dt
from bs4 import BeautifulSoup
from selenium import webdriver

# 변수 할당
a = 100
a, b = 100, 200
bs = BeautifulSoup()

# 사용자 정의 함수
def 함수명(매개변수):
    :
    return X

# 조건문
if amount == count:
    print('크롤링 완료')
```

```
# 반복문
for i in range(last_page):
    :
    data.append()
    print(data)

# 반복문 + 조건문 + 반복문
def abc(driver=Chrome, last_page=None):
    for i in range(last_page):
        if last_page == 0:
            print('리뷰글 없음')
            break
        for j in range(1, 11):
            count += 1
    return musinsa

# 데이터 프레임으로 만들기
df = pd.DataFrame(data, columns=['가격'])

# 저장
df.to_csv('C:/data.csv')
```

Understanding the structure of Python code

● 데이터분석 코드 구조 예시

```
# 모듈, 라이브러리, 클래스 import
import pandas as pd
import numpy as np
import os

# 파일 불러오기
df = pd.read_excel()

# 파일 구조, 크기, 형식 등 확인
df.shape / df.head() / df.tail()
df.info() / df.describe()

# 변수 할당
temperature = re_edg['온도']
humidity = re_edg['습도']
fine_dust = re_edg['미세먼지']

# 사용자 정의 함수
def 함수명(매개변수):
    :
    return X
```

```
# 반복문
for i in range(last_page):
    :
    data.append()
    print(data)

# 반복문 + 조건문 + 반복문
for i in range(제품리스트):
    if last_page == 0:
        print('마지막 제품입니다')

# 그래프
import matplotlib.pyplot as plt
from matplotlib import font_manager, rc
plt.figure(figsize=(14, 8))
plt.plot(x, y, color='k', label='불량률')
plt.title('년-월별 불량률', fontsize=20)
plt.xlabel('년-분기', fontsize=14)
plt.ylabel('COEF', fontsize=14)

plt.savefig('불량률 현황.jpg')
```

Python

Octoparse

octoparse.com

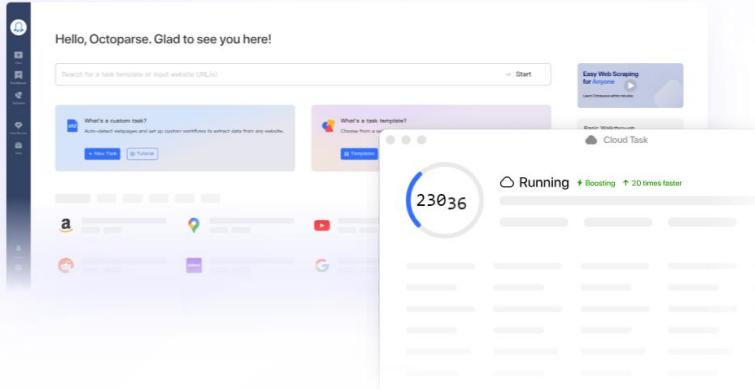
Ready, Set, Save! Get Ready for the Octoparse Mega Sale - Up to 40% OFF!

Octoparse Solutions Download Pricing Resources Login Start a free trial EN

Easy Web Scraping for Anyone

Octoparse is your no-coding solution for web scraping to turn pages into structured data within clicks.

Start a free trial Watch a demo



Hello, Octoparse. Glad to see you here!

Search for a task template or input matching URLs

Start

Easy Web Scraping for Anyone

What's a custom task? What's a task template?

New Task Tutorial

Cloud Task

Running 23036 Boosting 20 times faster

Join millions of data-driven organizations who benefit from Octoparse every day!

P&G SONY AUDI JCB pwc accenture

Read Customer Stories →

🔍

Google extension : Tab Save

Chrome | chrome://extensions

확장 프로그램

확장 프로그램 검색

개발자 모드

| | | |
|---|--|---|
| Absolute Enable Right Click & Copy Force Enable Right Click & Copy | AIPRM for ChatGPT AIPRM adds a list of curated prompt templates for SEO, SaaS and more to ChatGPT. Try today. | DeepL Translate: Reading & writing translator Translate while you read and write with DeepL Translate, the world's most accurate translator. |
| Enhanced ChatGPT Enhanced ChatGPT extension adds new features to ChatGPT, including chat export and prompt templates. A must-have for ChatGPT users. | Google Docs 오프라인 인터넷에 연결하지 않고도 문서, 스프레드시트 및 프레젠테이션을 설정, 생성, 확인합니다. | Language Reactor 당신이 공부하는 언어로 영화와 드라마를 보면서 자신의 언어실력을 즐겁고 효과적인 방법으로 향상 시키세요. |
| ShareGPT: Share your ChatGPT conversations ShareGPT is a Chrome extension that lets you your wildest ChatGPT conversations with one click. | Tab Save One-click download for files from open tabs, lists of URLs, or scholarly article PDFs from their associated DOI. | WebChatGPT: 인터넷 액세스가 가능한 ChatGPT 웹 브라우징을 통해 관련 웹 검색 결과로 ChatGPT 플랫폼을 강화하세요. |
| YouTube & Article Summary powered by ChatG... Summarize YouTube videos and web articles and save time with this extension, powered by ChatGPT. | 프롬프트 지니: ChatGPT 자동 번역기 ChatGPT 쓸 때 질문을 영어로 번역해 주고, 답변도 한글로 번역해 주는 번역기입니다. | |

Python

ChatGPT – Advanced Data Analysis
(→ Code interpreter)

⚡ GPT-3.5

🤖 GPT-4

ChatGPT PLUS

Come up with concepts
for a retro-style arcade game

Show me a code snippet
of a website's sticky header

↳ Share

Create a charter
to start a film club

Make a content strategy
for a newsletter featuring free local weekend e...

⊕ Send a message



Q & A