



# 데이터 분석 기반 AI 시스템 개발자 양성 과정

01. 빅데이터와 인공지능의 이해

조창제

강의자료

2025.08.

# 목차

## CONTENTS



### I

정보기술 개념과 구성요소 및 그 관계

## 정보기술

### II

빅데이터와 Data Base 개념

## 빅데이터

### III

인공지능 및 통계의 개념과 구분

## 인공지능





### 1.정보기술(Information Technology, IT)

1) 정의: 컴퓨터 및 통신 기술을 포함한 모든 정보와 관련된 기술을 의미

2) 배경:

- ① 현대 사회의 발전에 따라 **정보가 중요한 자원으로 부상**함에 따라, 이를 **관리하고 처리하는 기술**이 요구됨
- ② 산업화로 인해 효율적 생산 및 관리가 중요해짐에 따라, **자동화 및 최적화** 수요가 증가

3) 필요성:

- ① 원활한 **정보 교환**을 위한 기술과 **정보를 보호**하고 관리할 수 있는 기술에 대한 필요성이 대두





### 1.정보기술(Information Technology, IT)

#### 1) 구성:

- ① 하드웨어, 소프트웨어, 네트워크, 보안, 소프트웨어 개발 및 유지보수로 구분

#### 2) 기대효과:

- ① 정보를 효율적으로 관리, 비즈니스 프로세스 개선, 의사결정 지원, 기술혁신 및 경쟁력 확보

#### 3) 한계점:

- ① 비즈니스 효율성 향상, 경쟁력 강화, 비용 절감 등 **실질적 이점**을 제공하지 못하면 불필요할 수 있음





### 1.하드웨어

#### 1) 입력장치

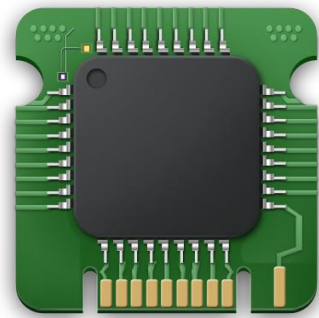
- ① 키보드, 마우스, 터치패드, 스캐너, 카메라, 마이크 등

#### 2) 출력장치

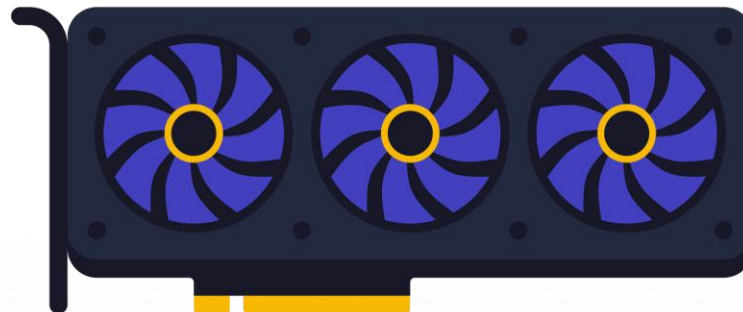
- ① 모니터, 프린터, 스피커

#### 3) 연산장치

- ① 중앙처리장치(Central Processing Unit, CPU)
- ② 그래픽처리장치(Graphics Processing Unit, GPU)



CPU



GPU





### 1.하드웨어

#### 4) 기억장치

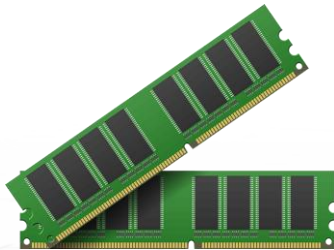
- ① 주기억장치(Main Memory)
  - RAM(Random Access Memory): 휘발성
  - ROM(Read Only Memory): 비휘발성
- ② 보조기억장치(Auxiliary Memory)

#### 5) 제어장치

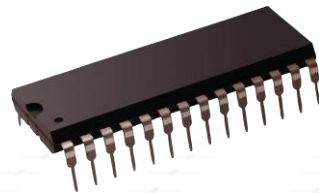
- ① 메인보드

#### 6) 전원장치

- ① 무정전전원장치(Uninterruptible Power Supply, UPS)



RAM



ROM



보조기억장치



메인보드



### 1.하드웨어

#### 7) 처리 과정



출근  
(부팅-ROM)



공장  
(주기억장치-RAM)



이동  
(파일 읽기)



창고  
(보조기억장치)



관리자  
(전사적 관리-CPU)



생산라인  
(그래픽 처리-GPU)



## 2.소프트웨어

### 1) 구분

#### ① 시스템 소프트웨어

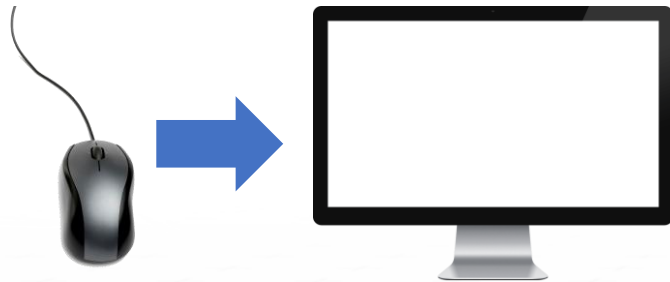
- OS나 **Driver**와 같이 하드웨어와 상호작용하여 시스템을 관리하는 소프트웨어

#### ② 미들웨어

- OS와 응용 소프트웨어 중간에서 조정 및 **중개의 역할**을 수행하는 소프트웨어

#### ③ 응용소프트웨어

- 워드, 스프레드시트, 디자인 프로그램 등과 같이 **특정 작업을 수행**하기 위해서 사용하는 소프트웨어



시스템 소프트웨어



Session 관리



응용 소프트웨어





## 2.소프트웨어

### 2) 구성

- ① 플랫폼: 여러 이해관계자들이 모여 가치를 창출하고 교환할 수 있는 **디지털 생태계**를 의미
  - 예시 - 유튜브, 배달의민족, 쿠팡, 멜론, 야놀자, AWS, Android, Azure, GCP, Windows 등
- ② 프레임워크: 어떠한 일을 처리하기 위해 **기본 구조와 규칙을 제공**하는 소프트웨어
  - 예시 - Django, Flask, FastAPI, Tensorflow, Torch, PyQt



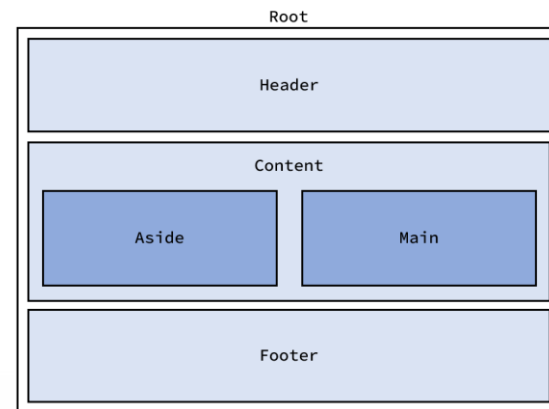
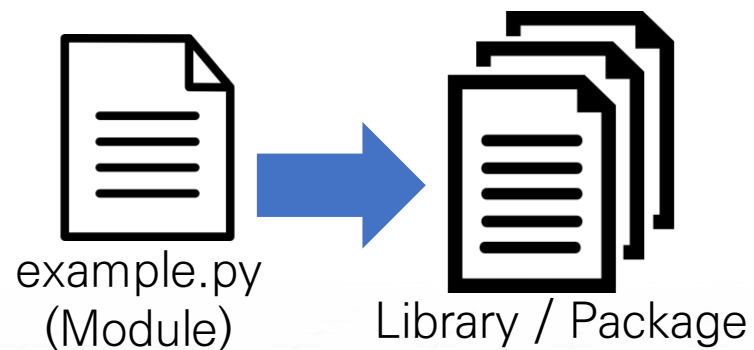
디지털 생태계 예시



## 2.소프트웨어

### 2) 구성

- ③ 모듈: 특정 기능을 가진 코드 단위
  - 예시 - example.py
- ④ 라이브러리: 특정 작업을 수행하기 위해 제공되는 **재사용 가능한 코드 집합**(함수, 클래스, 모듈)
- ⑤ 패키지: 관련된 여러 모듈과 하위 라이브러리들의 집합
- ⑥ 컴포넌트: 버튼, 폼과 같이 독립적으로 설계된 소프트웨어 구성요소(UI 요소를 포함)





### 2.소프트웨어

#### 3) 개발과정

##### ① 요구사항분석

- 사용자 요구 사항 분석, **필요한 기능 및 성능** 요구사항을 명확히 정의, 프로젝트 목적성 파악

##### ② 기획

- 프로젝트 및 제품 **일정** 및 예산, 리소스 계획, 개발 방법론, **프로젝트 방식, 팀 구성** 등을 결정

##### ③ 설계

- 시스템의 구조, 아키텍처, **사용자 인터페이스**, 데이터베이스, 알고리즘 등을 설계



요구사항 정의



프로젝트 기획



소프트웨어 설계



### 2.소프트웨어

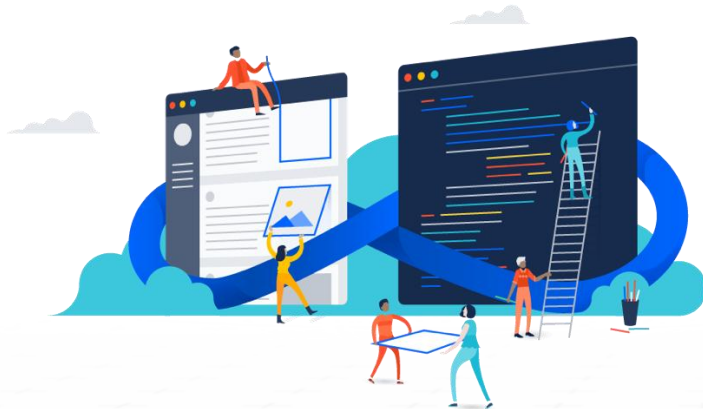
#### 3) 개발과정

##### ④ 개발

- 프로그래밍을 통한 제품 및 **소프트웨어 개발** 작업 수행

##### ⑤ 테스트

- 개발된 소프트웨어 및 제품의 **요구사항을 충족**하고 예상대로 동작하는지 확인
- 단위, 통합, 시스템, 성능 테스트 등의 다양한 단계를 포함



소프트웨어 개발



테스트 수행



## 2.소프트웨어

### 3) 개발과정

#### ⑥ 배포

- 테스트 완료 제품 및 소프트웨어를 **실제 환경에 배포**
- 사용자 교육 및 문서 작성 등의 작업을 수행

#### ⑦ 유지보수

- 배포된 제품 및 소프트웨어의 **관리 및 유지보수**
- 버그수정, 기능추가, 보안 업데이트 등



소프트웨어 배포



소프트웨어 업데이트





### 3.네트워크

#### 1) 정의

- ① 송신자의 메시지를 수신자에게 전달하는 과정
- ② 한 지점에서 원하는 다른 지점까지 의미 있는 정보를 **빠르고 정확하게** 상대방이 이해할 수 있도록 전달하는 것을 의미
- ③ 프로토콜(Protocol)
  - 통신망 내, 통신을 원하는 양측 시스템에서 데이터를 주고 받기 위해 미리 약속된 **운영상 규약**



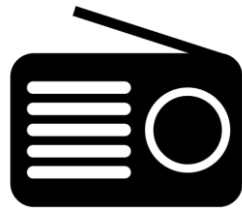




### 3.네트워크

#### 2) 구분

- ① 단방향통신(Simplex)
  - 데이터를 전송만 할 수 있고 수신할 수는 없음
- ② 반이중통신(Half Duplex)
  - 데이터를 송/수신할 수 있으나, 동시에 송/수신할 수 없음
- ③ 전이중통신(Full Duplex)
  - 데이터를 송/수신할 수 있으며, 동시에 송/수신이 가능



라디오  
(단방향통신)



무전기  
(반이중통신)



전화기  
(전이중통신)



### 4.보안

#### 1) 구분

##### ① 물리적보안

- 건물, 장치, 자산 등을 물리적으로 보호하는 것

##### ② 네트워크보안

- 네트워크상 정보를 보호하고 네트워크를 안전하게 유지하는 것

##### ③ 데이터보안

- 데이터의 기밀성, 무결성, 가용성을 보호하는 것

##### ④ 암호화

- 데이터를 안전하게 전송하거나 저장하기 위해 정보를 암호화 하는 것



물리적보안  
(출입통제, 내부망 등)



네트워크보안  
(방화벽)



데이터/파일 암호화



### 5.데이터베이스

#### 1) 정의

- ① 체계적이고 구조화된 방식으로 데이터를 저장, 관리, 검색할 수 있도록 하는 시스템
- ② 데이터 무결성
  - 데이터의 범위를 제한하거나 중복 발생이 불가하게 설정하여 데이터 무결성 확보 가능
- ③ 중복 최소화(일원화)
  - 데이터를 DB 서버 한 곳에만 저장함으로써, 같은 데이터를 각자 관리하는 문제가 없어짐
- ④ 다중 사용자 지원



파일 시스템



데이터베이스 관리 시스템  
(DBMS)



다중 사용자 지원



### 5.데이터베이스

#### 1) 정의

##### ⑤ 보안

- 사용자별로 접근 가능한 요소를 따로 관리할 수 있음
- 무결성으로 인해 데이터가 변조되지 않음

##### ⑥ 데이터 독립성

- 물리적, 논리적으로 잘 설계된 DB의 경우, 파일 구조 변경시에도 응용프로그램들이 영향을 받지 않거나 유지보수가 쉬움



데이터접근  
권한관리



데이터 독립성 및  
의존성



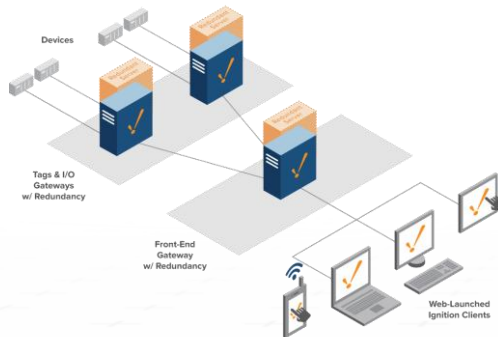
### 6.자원관리 및 유지보수

#### 1) 이중화기술

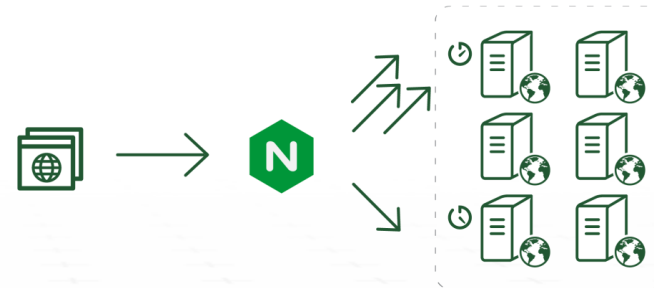
- ① 정의: **시스템의 안정성**을 높이기 위해 동일한 구성요소를 중복하여 배치하는 기술
  - 서버를 여러 대 구동해두고, 메인 서버에 장애가 발생하면 대기중인 서버가 동작하게 하는 기술
  - DB의 경우 이러한 기술을 리플리케이션(Replication)이라 부름

#### 2) 로드밸런싱

- ① 정의: 여러 장비에 트래픽을 나누어 **각 장비에 부하를 분산**해주는 기술
  - DB의 경우 이러한 기술을 샤딩, 파티셔닝이라 부름
  - 하나의 DB에서 테이블 분할은 파티셔닝(Partitioning) 여러 DB에 테이블을 분할하면 샤딩(Sharding)이라 부름



서버 이중화



트래픽 분산



### 6.자원관리 및 유지보수

#### 3) 버전관리

- ① 정의: 소프트웨어의 버전을 관리해주는 기술
- ② 목적: 변경 이력을 체계적으로 관리하는 시스템, 협업 지원
  - 특정 시점의 상태로 되돌릴 수 있음
  - 여러 개발자가 동시에 다른 작업을 할 수 있도록 브랜치(Branch) 및 병합(Merge) 기능 제공



이중화 기술  
(Kubernetes, k8s)



로드 밸런서  
(Nginx)



버전관리  
(Git)





### 7.관계

#### 1) 직업구분

- ① 마케팅 및 기획
- ② 개발자(웹, 데스크톱, 모바일, 펌웨어, 임베디드)
- ③ AI 개발자(Researcher, Engineer), 프롬프트 엔지니어, 레이블링
- ④ 데이터 관리자(DBA)
- ⑤ 서버 및 시스템 관리자(유지보수)
- ⑥ 네트워크/클라우드 전문가
- ⑦ 보안 전문가
- ⑧ 프로젝트 관리자
- ⑨ 디자이너(UI/UX, 웹, 그래픽)
- ⑩ 제품 판매(CS) 등

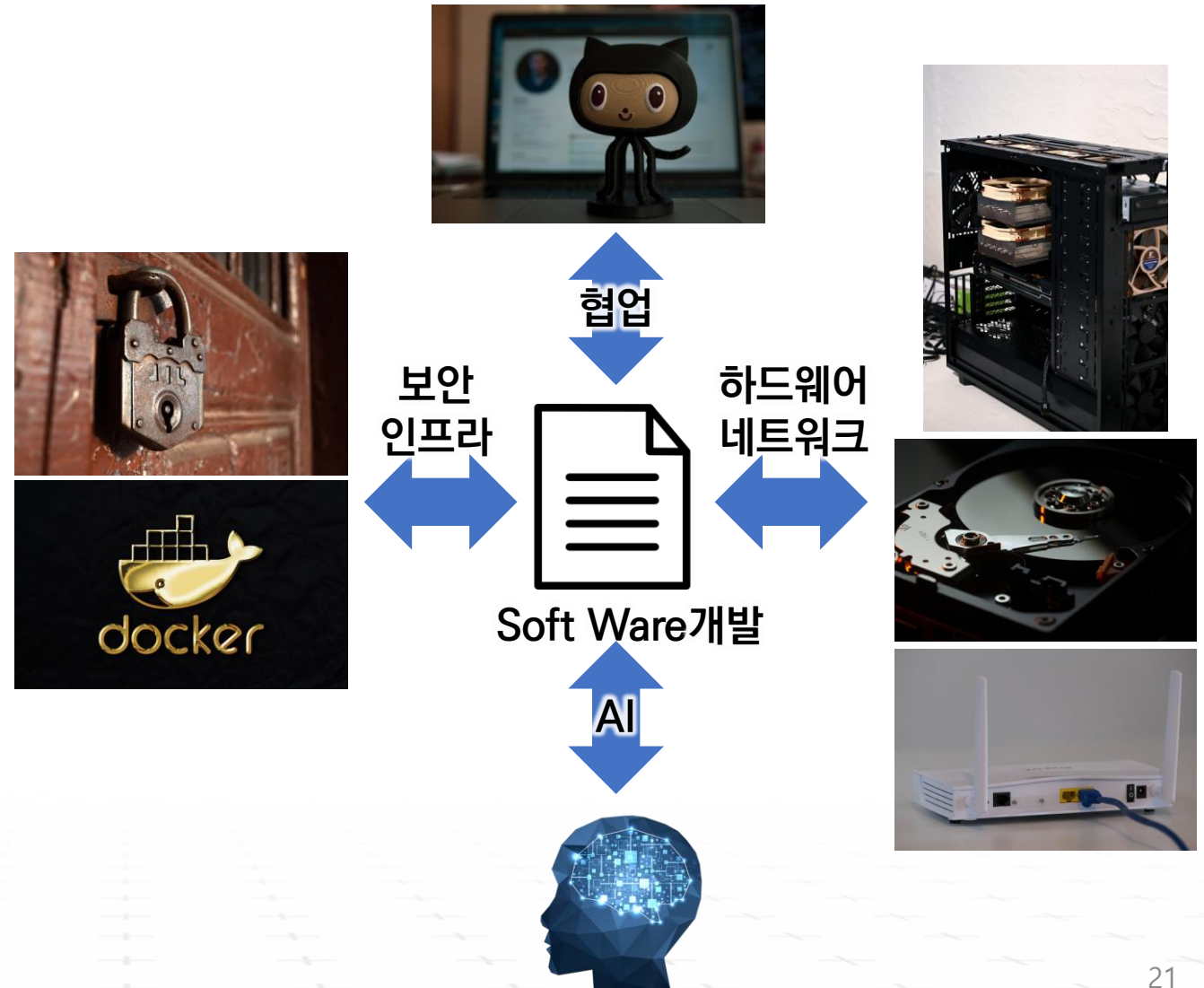




### 7.관계

#### 1) 소프트웨어 개발 특징

- ① 여러 개발자와 함께 소프트웨어를 개발
- ② 경우에 따라 하드웨어, 네트워크, AI 지식이 요구
- ③ 보안 및 유지보수 확장성을 고려해야 함





### 7.관계

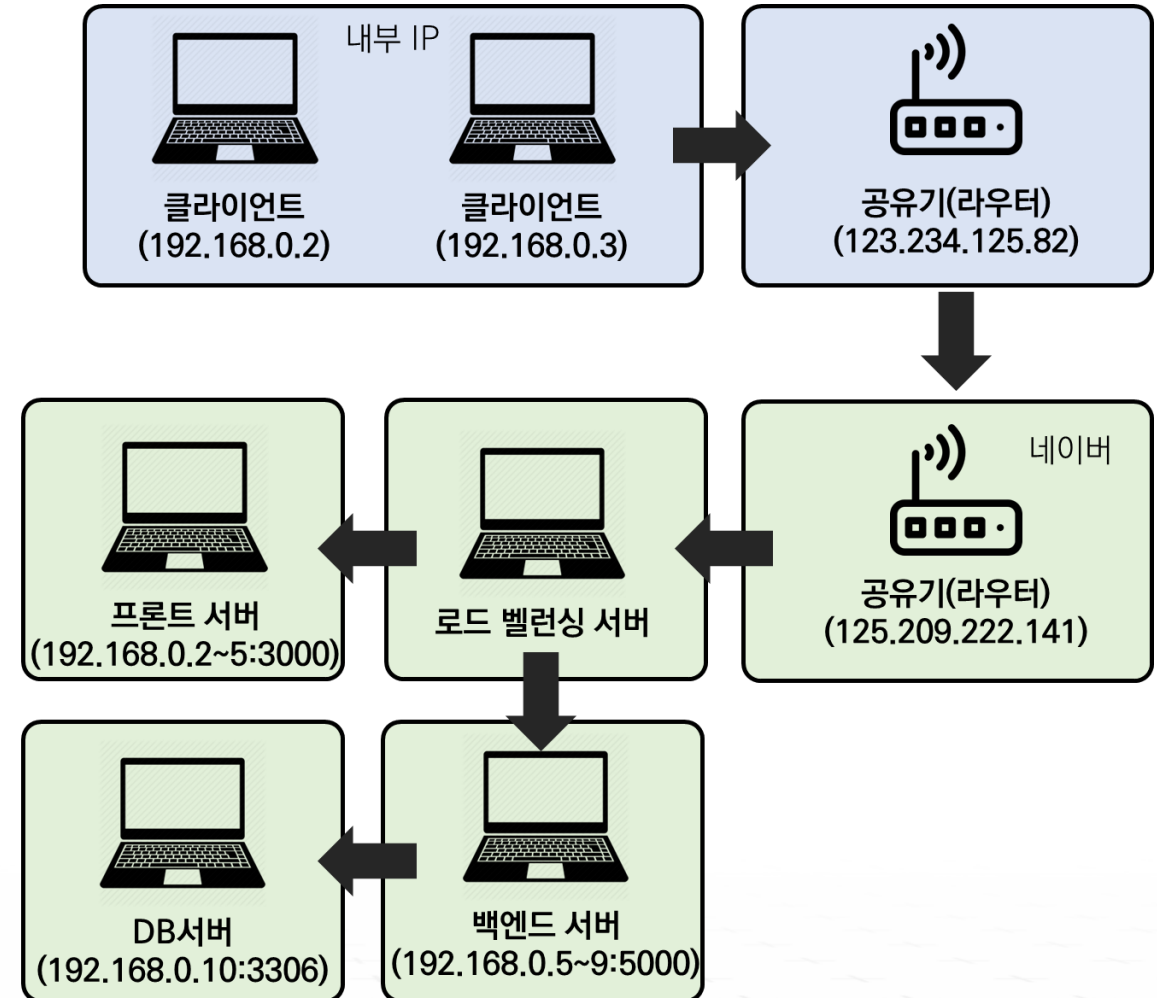
#### 2) 프론트 사례

##### ① 사용자가 서버 접속(192.168.0.2)

- 123.234.125.82 -> 125.209.222.141:80 접속
- 125.209.222.141:443으로 리다이렉트
- 192.168.0.2:3000에서 HTML/XML 반환
- 네이버 라우터 통해 192.168.0.2로 반환

##### ② 로그인 요청

- 123.234.125.82 -> 125.209.222.141:80 접속
- 192.168.0.2:3000에서 로드밸런싱 서버에게 백엔드 요청
- 192.168.0.5:5000서버 DB에 확인 요청
- 확인 처리 후 로드밸런싱 서버 거쳐서 프론트 서버에 반환





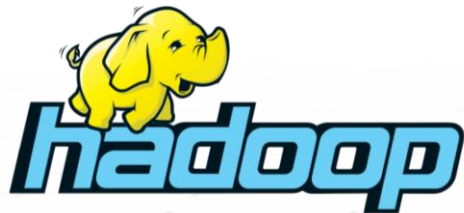
### 8. 빅데이터

#### 1) 정의

- ① 속도, 용량, 연산량 등의 문제로 하나의 컴퓨터로 서비스 되지 못할 정도로 방대한 데이터를 의미
- ② 주로 DB와 Cluster의 융합 기술을 의미. 즉, 여러 컴퓨터를 하나의 DB로 사용하는 기술
  - 예시 - 하둡, 스파크, 카프카 등

#### 2) 핵심

- ① 규모(Volume): 트랜잭션, 용량, 레코드
- ② 다양성(Variety): 정형, 비정형, 반정형
- ③ 속도(Velocity): 일반 처리, 실시간 처리, 스트림
- ④ 정확성(Veracity): 데이터 신뢰성
- ⑤ 가치(Value): 기업에 가치 있는 데이터





### 1. 정의

#### 1) 지능

- ① 한 개인이 문제에 대해 합리적으로 사고하고 해결하는 **인지적인 능력과 학습능력을 포함하는 총체적인 능력**
- ② 인지 또는 인식은 어떠한 대상을 분별하여 판별하는 것을 의미
- ③ 학습은 연습이나 경험의 결과로 일어나는 행동의 지속적 변화를 의미
- ④ 즉, 지능은 어떠한 **대상을 분별하여 판별하는 능력과 경험에 따른 학습능력**

#### 2) 인공지능

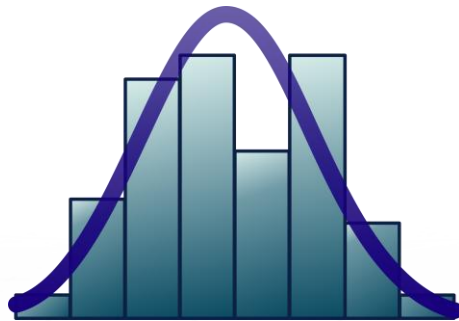
- ① 인간의 학습, 추론, 지각능력을 인공적으로 구현하려는 컴퓨터 과학분야
- ② 컴퓨터비전(Computer Vision, CV): 사물을 인지하는 능력을 모방한 분야
- ③ 자연어처리(Natural Language Processing, NLP): 언어 능력을 모방한 분야
- ④ 이 밖에도 다양한 분야가 존재



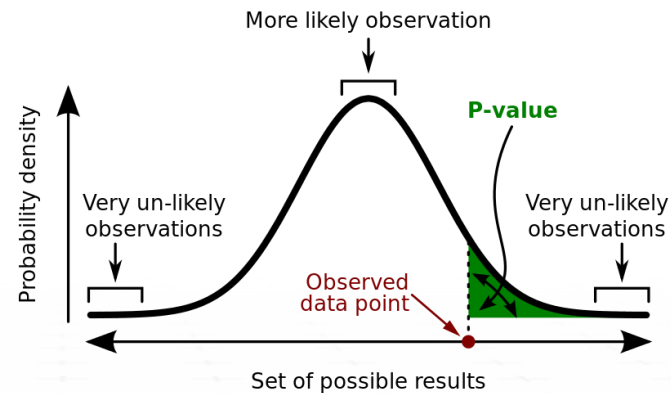
### 1. 정의

#### 1) 통계학

- ① 데이터의 속성을 정량적으로 기술하거나 데이터를 바탕으로 가설을 추론하는 분야
- ② 기술통계
  - 자료를 수치로 요약하여 정보를 제공
  - 예시 - 평균, 분산, 히스토그램
- ③ 추론통계
  - 샘플을 통해 모집단의 특성을 추론
  - 예시 - 선거추론,  $H_0/H_1$



기술 통계



가설검정





### 1. 정의

#### 1) 데이터마이닝

- ① 데이터 속에 숨겨져 있던 정보를 찾아내는 것을 의미

#### 2) 머신러닝

- ① 데이터를 바탕으로 훈련을 통해 추정/예측

#### 3) 딥러닝

- ① 인공신경망을 사용하여 학습하는 기술

#### 4) 딥러닝과 머신러닝의 차이

- ① 과거에는 SVM, Random Forest 같은 **전통 알고리즘**을 머신러닝, **다층 신경망**을 딥러닝이라 구분
- ② 현대에는 머신러닝은 패턴을 학습하고 추론하는 분야로 정의, 딥러닝을 **머신러닝의 한 분야**로 제시



데이터 마이닝



딥러닝

## 1. 자료 형식에 따른 구분

- 1) 표형식 자료(Tabular)에 대한 연구
- 2) 이미지 자료에 대한 연구
- 3) 언어 자료에 대한

## 2. 학습 방식에 따른 구분

- 1) 지도학습: 목표변수  $y$ 에 대해 학습하는 방식
- 2) 비지도학습: 목표변수 없이  $X$ 의 유사성을 바탕으로 군집하는 방식
- 3) 강화학습: 액션에 대한 이익과 패널티를 바탕으로 스스로 학습하는 방식



지도학습



비지도학습



강화학습



# Thank You

Email: [qkdrk777777@naver.com](mailto:qkdrk777777@naver.com)