

# 빅데이터

Big data Framework

# 프레임워크



02

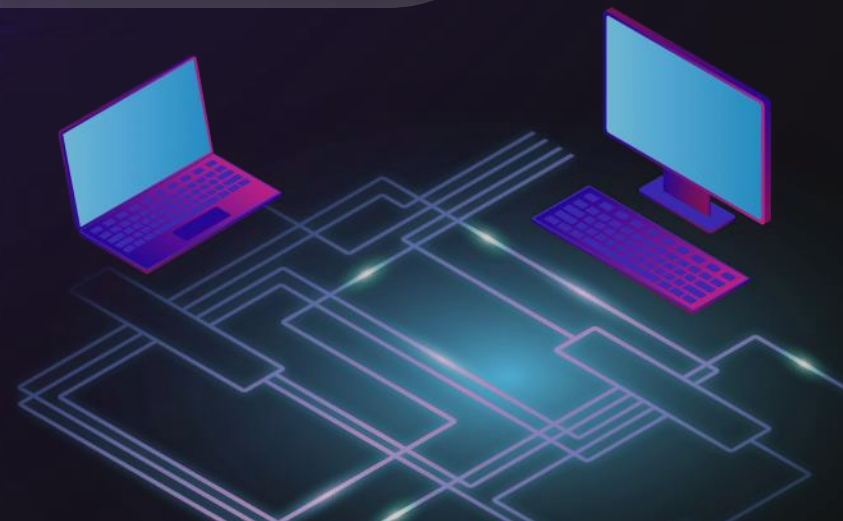
Big data Framework

# 빅데이터 기반 기술 - OS 및 가상화 기술



01

# 분산 시스템과 리눅스 운영체제



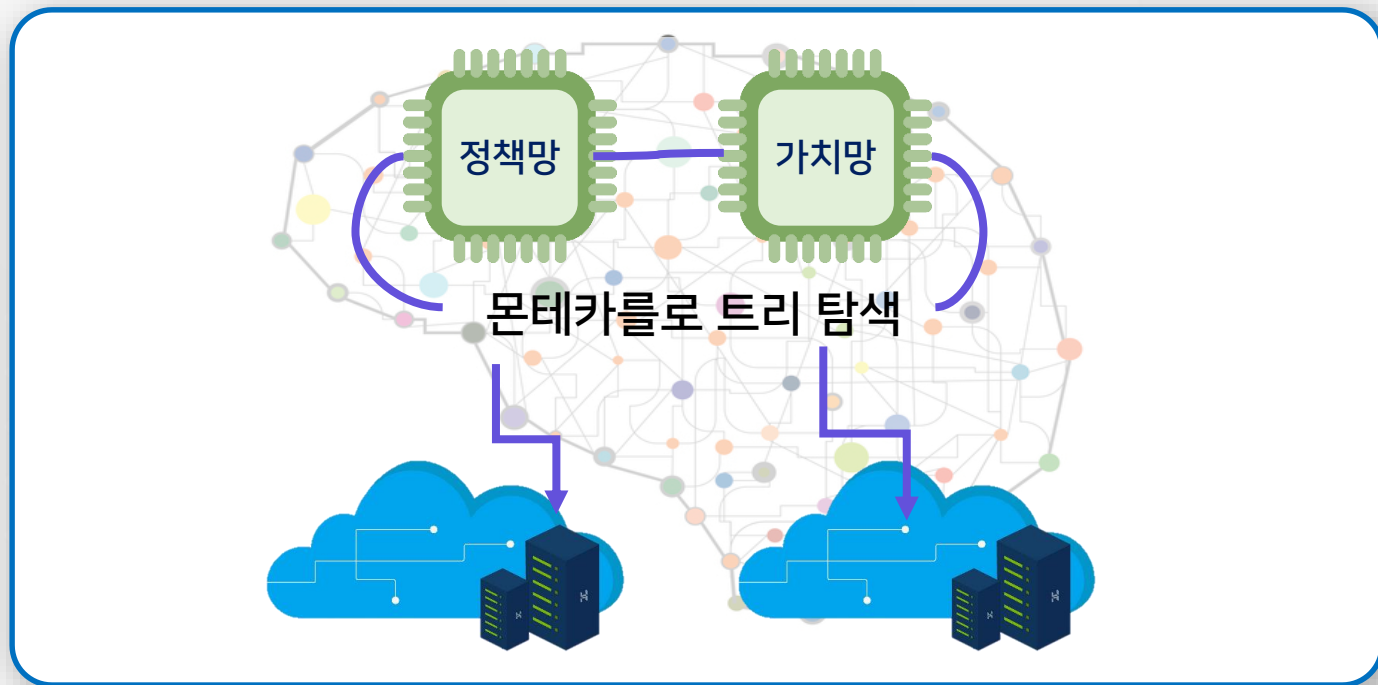
# 01 | 분산 시스템과 리눅스 운영체제

- ❏ 분산 시스템이 어떻게 활용되는지 설명할 수 있다.
- ❏ 분산 시스템의 원리와 구조를 설명할 수 있다.



### 1 분산 시스템은 어떻게 활용되는가?

- 알파고의 경우 총 1202개의 CPU와 176개의 GPU를 사용하여 천문학적인 계산을 수행하였음
- 슈퍼컴퓨팅 급의 고성능 계산 자원 활용을 위해 많은 컴퓨팅 자원을 네트워크로 연결함



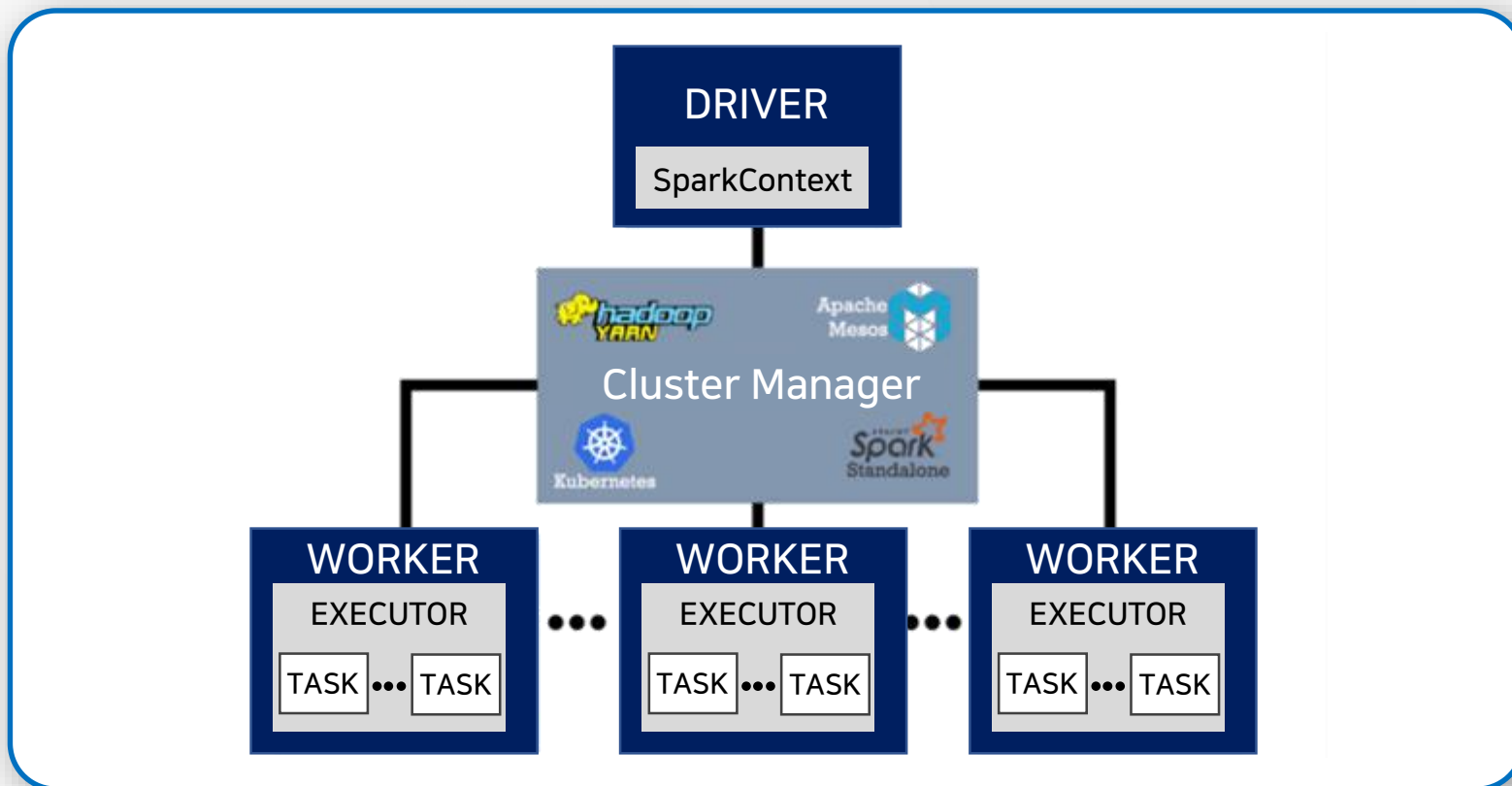
## 2 분산 시스템의 기본적 구조

- ❖ 어플리케이션이 네트워크로 연결된 여러 노드에서 수행됨
- ❖ 컴퓨팅 자원 관리자와 실제 데이터를 처리하는 워커 노드로 존재함



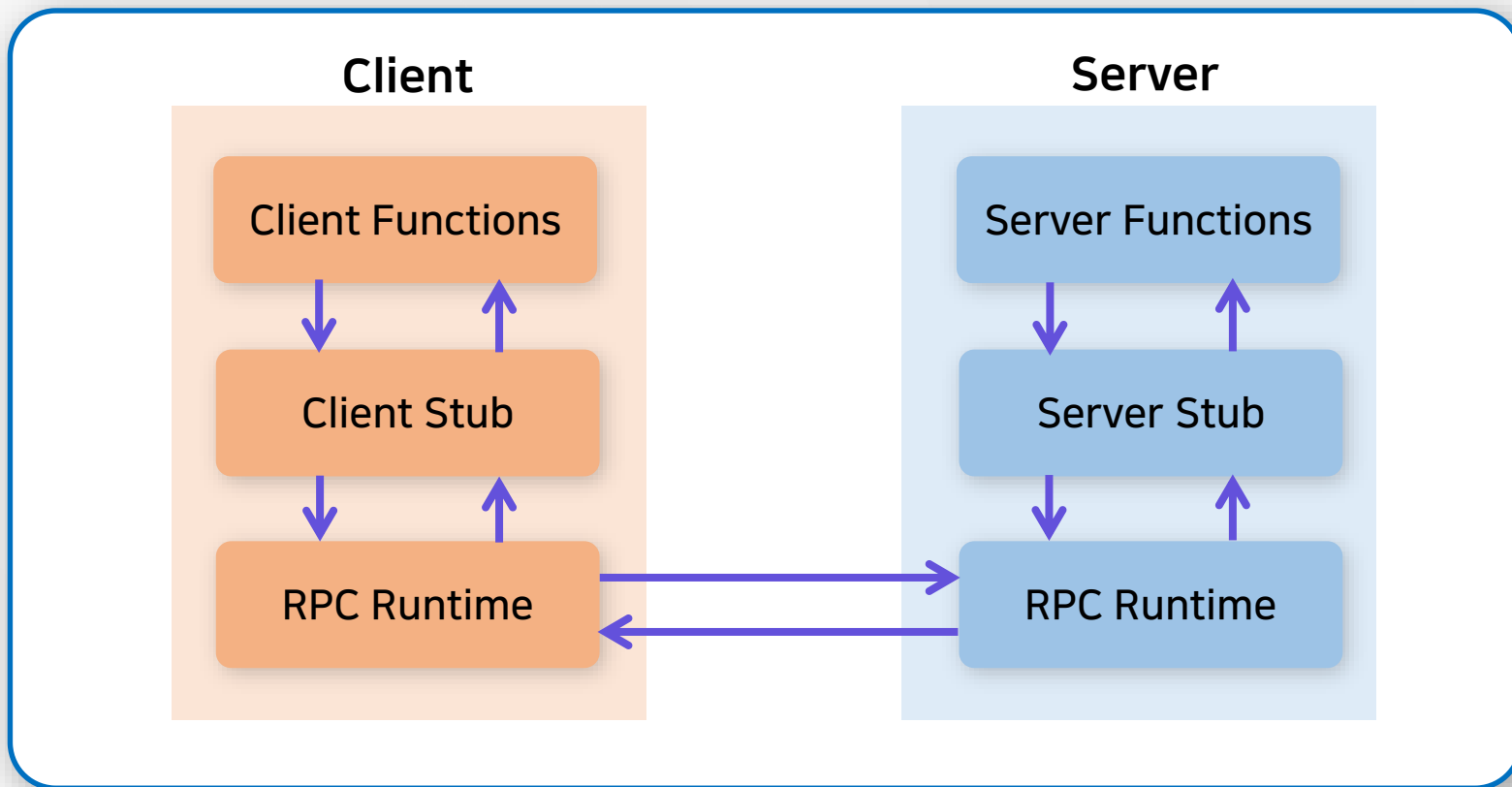
## 1 분산 시스템의 기본적 구조

- ❖ 전체 작업(job)을 Task 단위로 쪼개 여러 Worker에 분배
- ❖ Worker는 자신에게 할당된 Task를 수행 후 결과 전송



### 1 분산 시스템의 기본적 구조

- 🏗️ RPC(Remote Procedure Call)을 통한 노드 간 통신
- 🏗️ Server/Client 간 원격 함수 호출





### 1 Linux 특징



#### 다중 사용자, 다중 프로세스

- 여러 사용자가  
동시에  
사용 가능



#### 셸 명령 기반 동작



#### 우수한 네트워킹

- 리눅스에서부터  
네트워킹 시작
- FTP, Telnet,  
WWW 등



#### 분산 시스템에 활용 용이

## 2 Linux 명령어



### Linux 셸 명령어 구조



**\$ls -al**

↑      ↑  
명령어    옵션

```
ubuntu@DESKTOP-MK9G127:~$ ls -al
total 32
drwxr-xr-x 4 ubuntu ubuntu 4096 Jul 11 17:51 .
drwxr-xr-x 3 root  root  4096 Jul 11 17:34 ..
lrwxrwxrwx 1 ubuntu ubuntu  27 Jul 11 17:51 .aws -> /mnt/c/Users/StarcomPC/.aws
lrwxrwxrwx 1 ubuntu ubuntu  29 Jul 11 17:51 .azure -> /mnt/c/Users/StarcomPC/.azure
-rw----- 1 ubuntu ubuntu   5 Jul 11 17:37 .bash_history
```

## 3 Linux 주요 명령어 정리

### 1 현재 경로에 존재하는 파일 보여줌(ls)

- `ls -al`
  - 디렉토리 내 모든 파일 및 폴더 리스팅
- `ls -alh`
  - 디렉토리 내 모든 파일 및 폴더 리스팅과 더불어 파일 사이즈를 눈에 보기 쉽게 해 줌
- `ls -ltr`
  - 디렉토리 내 모든 파일 및 폴더의 수정시간 기준 정렬

### 2 디렉토리 옮기기(cd)

- `change directory`
- `cd [이동할 디렉토리]`

## 3 Linux 주요 명령어 정리

### 3 현재 작업중인 디렉토리 출력

- pwd

### 4 파일 복사 (cp)

- cp [arg] [복사할 파일] [복사될 파일]
  - cp -r (폴더까지 복사)
  - cp -f (같은 파일이 있다면 강제로 덮어쓰움)

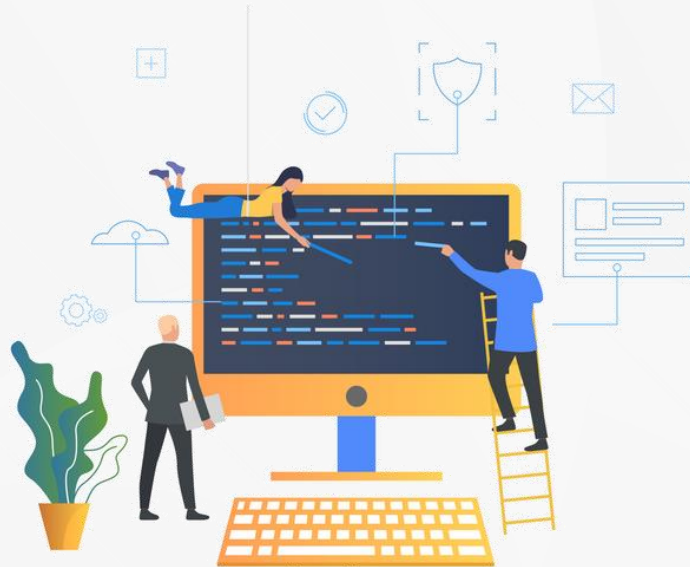
### 5 파일 이동 (mv)

- mv [arg] [이동할 파일명] [이동할 위치]

## 3 Linux 주요 명령어 정리

### 6 파일 지우기 (rm)

- `rm -r [디렉토리]` 디렉토리를 포함해서 삭제
- `rm -v [파일]` 삭제한 내용에 대해 표시
- `rm -f [파일]` 강제로 삭제



## 1 분산 시스템의 활용

- 고성능의 계산 자원 활용을 위해  
많은 컴퓨팅 자원을 네트워크로 연결

## 2 분산 시스템의 원리

- 전체 일(job)을 Task단위로 쪼개  
여러 Worker에 분배하여 분산처리



## 3 리눅스 운영체제의 특징

- 다중 사용자, 다중 프로세스
- 셸 기반 명령 기반 동작
- 우수한 네트워킹
- 분산 시스템에 활용하기에 용이

✎ 빅데이터 컴퓨팅 기술, 2014, 박두순, 한빛미디어