AWS EC2

Category	
DATE	@2022/03/14
Complete	☐
etc	
열	

```
1. 제공기간:  ~ 공통 PJT 종료 시(종료 후 7일 이내 삭제 예정)
2. 서버 도메인: j6c207.p.ssafy.io (예: 서울1반 1팀 - i6a101.p.ssafy.io)
3. 접속 방법: 제공된 인증키(.pem)를 사용하여 ubuntu 계정으로 접속, 아래 방법 참조
https://blog.edit.kr/entry/AWS-EC2-%EC%97%90-putty%EB%A1%9C-%EC%97%B0%EA%B2%B0%ED%95%98%EA%B8%B0-PuTTYgen%EC%9C%BC%EB%A1%9C-pem%ED%8C%8C%EC%9D%BC%EC%9D%84-ppk%EB%A1%9C-%EB%B3%80%ED%99%98-%ED%95%84%EC%88%98

4. 방화벽은 없음, 관련 설정 불 필요
5. 아래 명령어들을 잘못 사용시, 시스템이 복구 불능 될 수 있음, PJT시 사용 할 일 없음
sudo ufw
sudo iptables
sudo chmod
sudo poweroff
sudo shutdown
sudo halt
sudo init
sudo rm -rf /
6. 재부팅 원하면 sudo reboot 만 사용
```

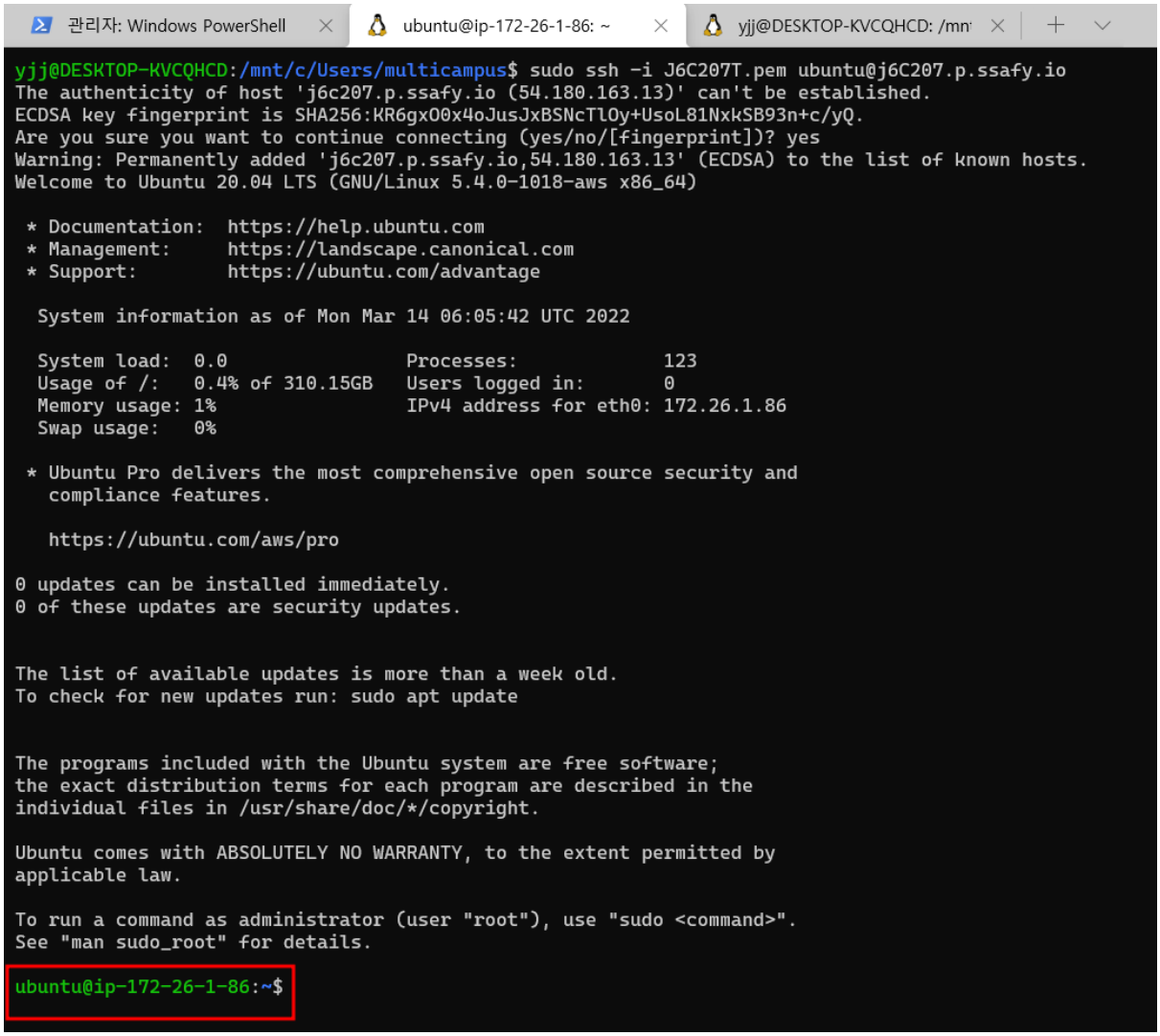
```
빅데이터(분산) 프로젝트의 대규모 분산 클러스터(하둡) 활용을 위해 제공

제공기간: 금일 ~ 특화 프로젝트 종료 시
서버 도메인: cluster.p.ssafy.io
접속 방법: 제공된 인증키(.pem)를 사용하여 j6<팀ID> 계정으로 접속

예: ssh -i <팀인증서> j6<팀ID>@cluster.p.ssafy.io # 본인 팀 인증서 사용
사용 안내: SSAFY GIT - Help - 매뉴얼 게시판의 "공용 하둡 클러스터 사용 안내" 참고
```

.pem으로 원격 리눅스 접속하기 (ssafy - AWS)

1. 로컬 우분투접속
2. .pem파일이 있는 위치로 이동
3. 다음명령어 실행

일반 서버

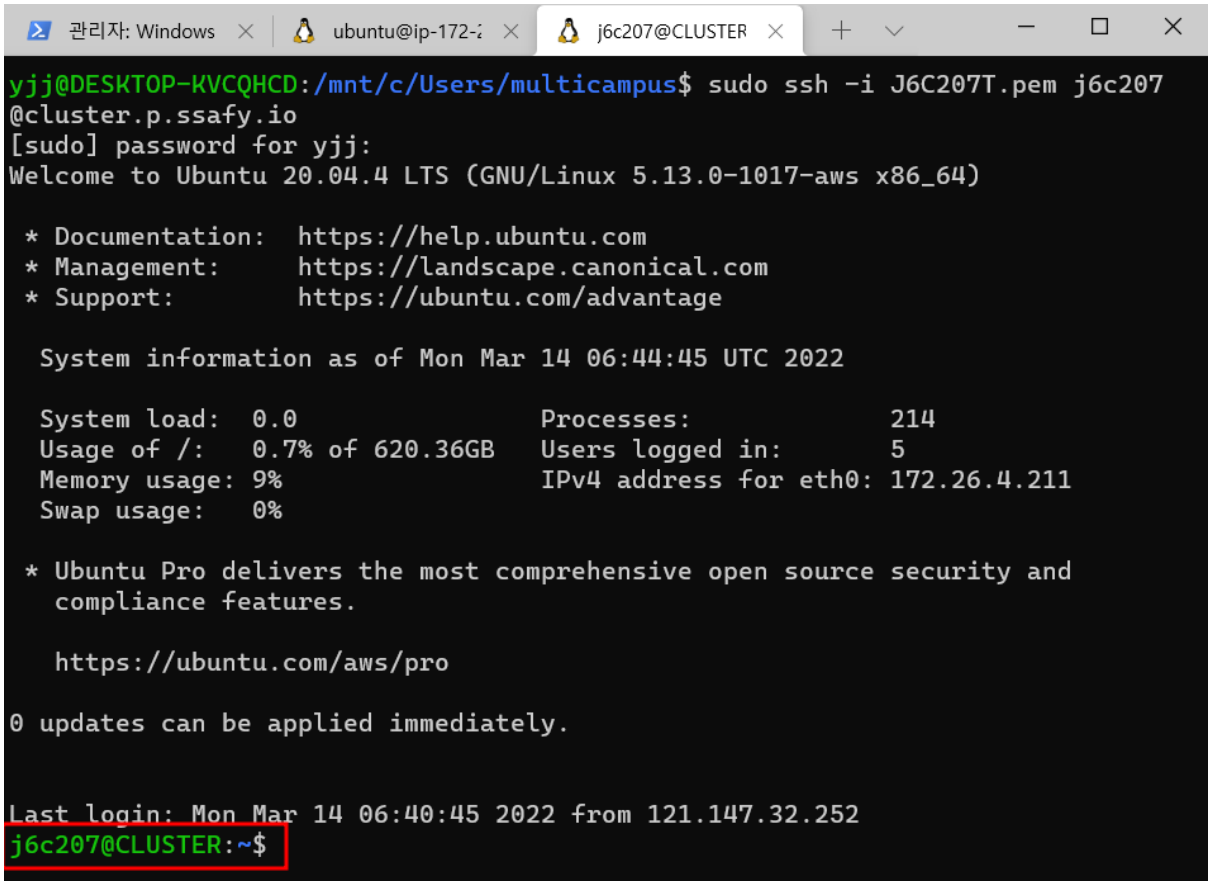
```
sudo ssh -i J6C207T.pem ubuntu@j6C207.p.ssafy.io
```



접속이 되었음을 확인

분산 서버

```
sudo ssh -i J6C207T.pem j6c207@cluster.p.ssafy.io
```



AWS EC2 IP 찾기

private IP address

```
curl http://169.254.169.254/latest/meta-data/local-ipv4
```

public IP address

```
curl http://169.254.169.254/latest/meta-data/public-ipv4
```

일반 서버

- private IP address : 172.26.1.86
- public IP address : 54.180.163.13

분산 서버

- private IP address : 172.26.4.211
- public IP address : 52.78.208.161

Docker 설치

https://docs.docker.com/engine/install/ubuntu/

1. 저장소 설정

```
sudo apt-get update

sudo apt-get install \
    ca-certificates \
    curl \
    gnupg \
    lsb-release
```

```
curl -fsSL https://download.docker.com/linux/ubuntu/gpg | sudo gpg --dearmor -o /usr/share/keyrings/docker-archive-keyring.gpg
```

```
echo \
  "deb [arch=$(dpkg --print-architecture) signed-by=/usr/share/keyrings/docker-archive-keyring.gpg] https://download.docker.com/linux/
  $(lsb_release -cs) stable" | sudo tee /etc/apt/sources.list.d/docker.list > /dev/null
```

2. 도커 엔진 설치

```
sudo apt-get update
 sudo apt-get install docker-ce docker-ce-cli containerd.io
```

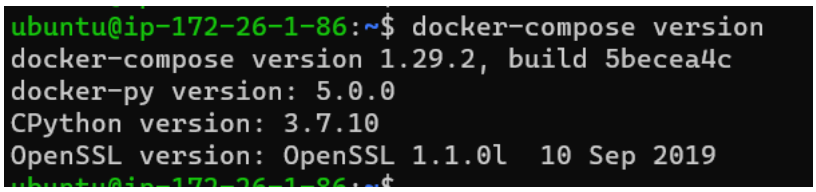
3. 도커 compose설치

https://darrengwon.tistory.com/793

```
sudo curl -L "https://github.com/docker/compose/releases/download/1.29.2/docker-compose-$(uname -s)-$(uname -m)" -o /usr/local/bin/doc
sudo chmod +x /usr/local/bin/docker-compose
```

docker compose 버전 확인

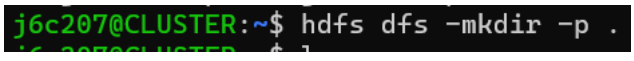
```
docker-compose version
```



하둡

사용자 홈 디렉토리 생성

```
hdfs dfs -mkdir -p .
```



텍스트 파일 생성

```
echo 'Hello SSAFY'>a.txt
```

```
j6c207@CLUSTER:~$ echo 'hello ssafy'>a.txt
j6c207@CLUSTER:~$ ls
a.txt
```

DFS 업로드

```
hdfs dfs -put a.txt
```

DFS 업로드 확인

```
hdfs dfs -cat a.txt
```

```
j6c207@CLUSTER:~$ hdfs dfs -put a.txt
j6c207@CLUSTER:~$ hdfs dfs -cat a.txt
hello ssafy
```

맵리듀스 예제 “GREP”실행

```
hadoop jar /usr/local/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.1.jar grep a.txt output 'SSAFY[a-z.!]+'
```

```
j6c207@CLUSTER:~$ hadoop jar /usr/local/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.1.jar grep a.txt output 'SSAFY[a-z.!]+'
2022-03-14 12:33:04,796 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2022-03-14 12:33:04,910 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2022-03-14 12:33:04,910 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2022-03-14 12:33:05,107 INFO input.FileInputFormat: Total input files to process : 1
2022-03-14 12:33:05,171 INFO mapreduce.JobSubmitter: number of splits:1
2022-03-14 12:33:05,276 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1638119158_0001
2022-03-14 12:33:05,276 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-03-14 12:33:05,376 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2022-03-14 12:33:05,376 INFO mapreduce.Job: Running job: job_local1638119158_0001
2022-03-14 12:33:05,377 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2022-03-14 12:33:05,383 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2022-03-14 12:33:05,383 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2022-03-14 12:33:05,384 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2022-03-14 12:33:05,418 INFO mapred.LocalJobRunner: Waiting for map tasks2022-03-14 12:33:05,418 INFO mapred.LocalJobRunner: Starting task: attempt_local1638119158_0001_m
_000000_0
2022-03-14 12:33:05,437 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2022-03-14 12:33:05,437 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2022-03-14 12:33:05,451 INFO mapred.Task:  Using ResourceCalculatorProcessTree : [ ]
2022-03-14 12:33:05,454 INFO mapred.MapTask: Processing split: hdfs://ip-172-26-4-211.ap-northeast-2.compute.internal:9000/user/j6c207/a.txt:0+12
2022-03-14 12:33:05,532 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
2022-03-14 12:33:05,532 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2022-03-14 12:33:05,532 INFO mapred.MapTask: soft limit at 83886080
2022-03-14 12:33:05,532 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
2022-03-14 12:33:05,532 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
2022-03-14 12:33:05,537 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2022-03-14 12:33:05,624 INFO mapred.LocalJobRunner:
2022-03-14 12:33:05,626 INFO mapred.MapTask: Starting flush of map output2022-03-14 12:33:05,640 INFO mapred.Task: Task:attempt_local1638119158_0001_m_000000_0 is done. A
nd is in the process of committing
2022-03-14 12:33:05,643 INFO mapred.LocalJobRunner: map
2022-03-14 12:33:05,643 INFO mapred.Task: Task 'attempt_local1638119158_0001_m_000000_0' done.
2022-03-14 12:33:05,650 INFO mapred.Task: Final Counters for attempt_local1638119158_0001_m_000000_0: Counters: 24
        File System Counters
                FILE: Number of bytes read=281318
                FILE: Number of bytes written=916500
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=12
                HDFS: Number of bytes written=0
```

맵리듀스 실행 결과 확인

```
hdfs dfs -cat output/*
```

```
j6c207@CLUSTER:~$ hdfs dfs -cat output/*
```

텍스트 파일 삭제

```
hdfs dfs -rm a.txt
hdfs dfs -rm -r output
rm a.txt
```

```
j6c207@CLUSTER:~$ hdfs dfs -rm a.txt
Deleted a.txt
```

```
j6c207@CLUSTER:~$ hdfs dfs -rm -r output
Deleted output
```

```
j6c207@CLUSTER:~$ rm a.txt
j6c207@CLUSTER:~$ ls
j6c207@CLUSTER:~$
```

a.txt 삭제됨

홈 디렉토리 삭제

```
hdfs dfs -rmdir .
```