

# GCP의 VM끼리 서로 통신하게 클러스터 구축하기

## About

하둡은 여러개의 클러스터 위에서 구동되는 분산환경 시스템이다. 물론, machine 1개로도 구동이 되지만 이는 분산 환경 시스템의 의미에 맞지 않다. 여러 개의 machine들이 서로 통신하며 분산환경 시스템이 구축되는 것이 하둡을 십 분 활용하는 것이다. 이를 위해 GCP에서 여러 대의 VM을 만들텐데, 이들이 서로 통신하게 설정을 해주어야 한다. 이번 글에서는 바로 이 부분에 대해서 살펴본다.

## Prerequisite

- 이미 GCP에 여러 대의 VM을 만들었다고 가정한다. 여기서는 4개의 VM을 만들었는데, 각 VM의 이름을 `hadoop-master-01`, `hadoop-worker-01`, `hadoop-worker-02`, `hadoop-worker-03` 으로 설정 하였다.

## Let's Start!

먼저 sshkey를 생성한다.

```
$ ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
```

`~/.ssh/` 라는 디렉토리에 `id_rsa` 라는 키를 만들라는 명령어이다. 이 명령어를 수행하면 `id_rsa`, `id_rsa.pub` 이 생기는 것을 확인할 수 있다.

```
$ cd ~/.ssh
$ ls -al
```

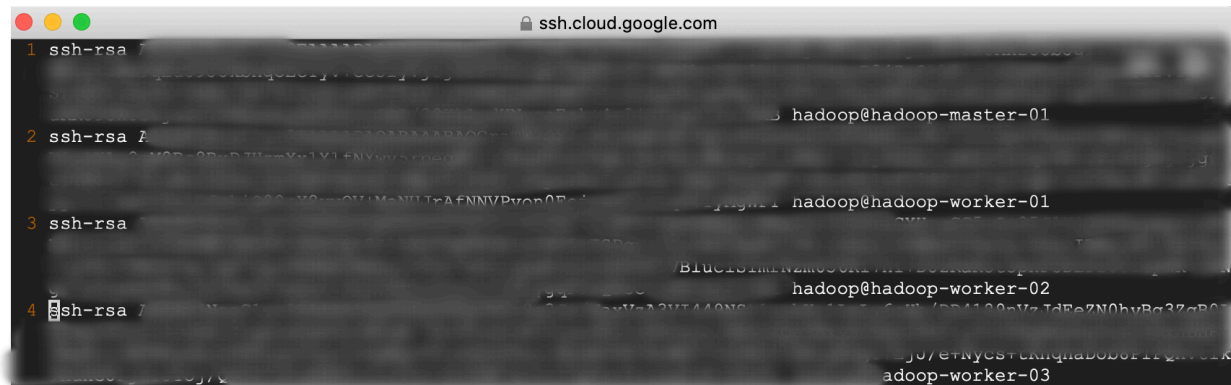
쉽게 말해서, `id_rsa` 는 나만 볼 비밀키이고 `id_rsa.pub` 는 남들에게 공개될 공개키이다. 다른 VM과 통신할 때에는 `id_rsa.pub` 을 사용한다. 그 내용을 보고 싶다면 아래의 명령어를 사용해보자.

```
$ cat id_rsa
$ cat id_rsa.pub
```

뭔가 `id_rsa` 가 더 긴 것을 확인할 수 있다. 이제 `id_rsa.pub` 의 내용을 복사하여 `authorized_keys` 에 저장 해보자. 이 파일에는 본인 클러스터의 공개키 뿐만 아니라 서로 송수신할 클러스터의 공개키를 담아둔다.

```
$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
$ vi ~/.ssh/authorized_keys
```

`vi` 을 통해서 에디터 모드로 들어가고 `:set nu` 을 통해서 한 줄로 되어있나 확인해보자. 이를 확인했으면, 다른 클러스터에 있는 `id_rsa.pub` 도 모두 복사하여 한 곳으로 모은다. 이 작업을 각 클러스터에서 반복하자. 나는 네 개의 VM을 만들었으므로 최종 아래의 형태가 나왔다.



```
ssh.cloud.google.com
1 ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAQDA... hadoop@hadoop-master-01
2 ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAQDA... hadoop@hadoop-worker-01
3 ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAQDA... hadoop@hadoop-worker-02
4 ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAQDA... hadoop@hadoop-worker-03
```

옆에 네 줄로 되어있는 것을 확인할 수 있다.

마지막으로 아래 설정을 통해 다른 계정이 읽고 쓸수 없게 만들어준다.

```
$ chmod 600 ~/.ssh/authorized_keys
$ cd ~/.ssh
$ ls -al
```

최종적으로 아래와 같이 `authorized_keys`에 `-rw-----` 가 나와야 한다.

```
-rw-----. 1 hadoop hadoop 1620 Oct 14 16:29 authorized_keys
```

이제 아래 명령어로 다른 클러스터에 로그인할 수 있다.

```
$ ssh hadoop-worker-01
```