

Create Google Cloud Project พวัก hadoop-project

Compute Engine==➡ VM instances

Firewall rule

sudo su – ชื่อจะกลายเป็น root@hadoop-docker(จะอยู่ใน docker)

apt-get update ให้กลายเป็นล่าสุด

apt-get install docker.io ติดตั้ง docker

apt-get install docker.io เอา cloudera ใส่ docker

docker images ดูว่าที่เราติดตั้งมีรีปาว

docker run -v /root:/mnt --hostname=quickstart.cloudera --privileged=true -t -i -p

8888:8888 -p 8880:8880 -p 9092:9092 -p 2181:2181 -p 11122:11122

cloudera/quickstart /usr/bin/docker-quickstart คำสั่งรันตัว cloudera

Login to Hue: <http://<<external-ip-address>>:8888>



- |    | Step   | Conclusion |
|----|--|------------|
| 1. | cloud.google.com -> Try free   |            |
| 2. | Create Project   |            |
| 3. | Create VM Instance   |            |
| 4. | Add firewall rule  |            |
|    | <ul style="list-style-type: none"><li>• Target : All instances in the network</li><li>• Source IP ranges: 0.0.0.0/0</li><li>• Allowed protocols and ports : Allow all</li></ul>  |            |
| 5. | Connect to the instance (via SSH)  |            |
|    | <ul style="list-style-type: none"><li>• Update OS : apt-get update</li><li>• Install Docker : apt-get install docker.io</li><li>• Install Cloudera in Docker : docker pull cloudera/quickstart:latest</li><li>• Run Docker : docker run -v /root:/mnt --hostname=quickstart.cloudera --privileged=true -t -i -p 8888:8888 -p 8880:8880 -p 9092:9092 -p 2181:2181 -p 11122:11122 cloudera/quickstart /usr/bin/docker-quickstart</li></ul> | 33         |





`docker ps -a -> [container id]` รู้ container id ว่าอะไรรันอยู่บ้าง

`docker exec -it [container id] bash` เข้า root@quickstart เพื่อเข้า cloudera คนละ session

`docker rm [container id]` ลบอันที่ไม่ใช่ออก

cluster แบบ UI

dataproc--> clusters--> create--> ตั้งชื่อตามจารย์

<input type="checkbox"/> Name ^	Zone	Recommendation	In use by	Internal IP	External IP	Connect
<input type="checkbox"/>  arisa-m	us-east1-c			10.142.0.4 (nic0)	34.73.207.92	SSH ▾ ⋮
<input type="checkbox"/>  arisa-w-0	us-east1-c			10.142.0.3 (nic0)	34.73.197.194	SSH ▾ ⋮
<input type="checkbox"/>  arisa-w-1	us-east1-c			10.142.0.5 (nic0)	34.73.144.39	SSH ▾ ⋮
<input type="checkbox"/>  hadoop-docker	us-east1-c			10.142.0.2 (nic0)	34.73.5.146 ↗	SSH ▾ ⋮

เลือก arisa-m เพื่อประมวลผล กด ssh

แบบใช้ google cloud shell





`#gcloud dataproc clusters create thanisa-dataproc \`

`--zone=us-central1-a \`

`--master-machine-type=n1-standard-2 \`

`--worker-machine-type=n1-standard-2` สร้าง dataproc create cluster

สร้าง instance ชื่อ cloudera หลายๆอัน แต่ยังไม่เป็น cluster เพราะยังไม่ได้เชื่อมกัน เหมือนแยกกันอยู่

<input type="checkbox"/> Name ^	Zone	Recommendation	In use by	Internal IP	External IP	Connect
<input type="checkbox"/>  cloudera1	us-east1-b			10.142.0.2 (nic0)	35.190.189.7 <a href="#">↗</a>	SSH ▾ ⋮
<input type="checkbox"/>  cloudera2	us-east1-b			10.142.0.3 (nic0)	34.73.48.232 <a href="#">↗</a>	SSH ▾ ⋮
<input type="checkbox"/>  cloudera3	us-east1-b			10.142.0.4 (nic0)	35.196.1.76 <a href="#">↗</a>	SSH ▾ ⋮
<input type="checkbox"/>  cloudera4	us-east1-b			10.142.0.5 (nic0)	34.73.208.42 <a href="#">↗</a>	SSH ▾ ⋮

ต้องset ให้อยู่อันเดียวกัน root พาสเวิร์ดเดียวกัน แต่ละตัวต้องเข้าถึงกันได้  
ไปที่ ssh แต่ละตัว

ต้องเข้า 1. `sudo -i` เพื่อเข้าถึง `root@cloudera1`

2. `#passwd` เปลี่ยนพาส

3. ทำการ config บางอย่าง โดยเปลี่ยนค่า 3 ตัว เป็น yes

- ใช้คำสั่ง `# nano /etc/ssh/sshd_config` เพื่อแก้ไข ไฟล์

In the file, change the following 3 configurations

**PermitRootLogin** yes

..

**ChallengeResponseAuthentication** yes

...

**PasswordAuthentication** yes

เปลี่ยนเสร็จ ก็กด `ctrl x` แล้วกด `y`

4. Generate SSH key ขึ้นมาใหม่

```
# service ssh restart
```

```
# ssh-keygen
```

```
# service ssh restart
```

5. ให้ทุกตัวเชื่อมต่อกันไปหากันได้ (การทำให้เป็น cluster เพื่อให้ทุกตัวประมวลผล) กันได้

ของ cloudera1

```
# ssh-copy-id root@cloudera2
```

```
# ssh-copy-id root@cloudera3
```

```
# ssh-copy-id root@cloudera4
```

ของ cloudera2

```
# ssh-copy-id root@cloudera1
```

```
# ssh-copy-id root@cloudera3
```

```
# ssh-copy-id root@cloudera4
```

6. `#sysctl -w vm.swappiness=0` ใส่ทำไม??????????

7. Hands-on ติดตั้ง Cloudera

- wget <http://archive.cloudera.com/cm5/installer/latest/cloudera-manager-installer.bin>

-chmod u+x cloudera-manager-installer.bin

-sudo ./cloudera-manager-installer.bin

-----รอการติดตั้ง-----

8. เปิด <http://<<external-ip>>:7180> 7180 ไม่เคยสร้าง สร้างfirewallใหม่เลย

## Module 3 HDFS กับ GCS

Importing File from Local to HDFS แบบ UI ทำการติดต่อ hdfs

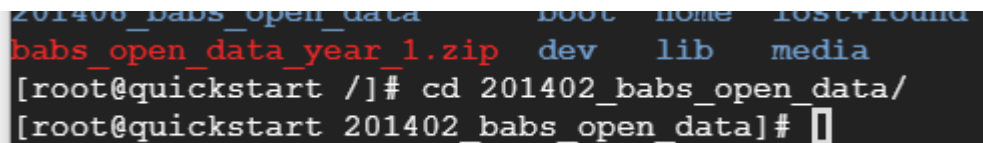
เข้า Hue :8888 คลิก file browser manage hdfs

สร้าง folder คลิก new ตั้งชื่อ input เป็นการ upload ไฟล์ขึ้น cloud เอาไฟล์อะไรก็ได้ขึ้น

Manage Files/Directories in HDFS Hadoop เพื่อติดต่อ hdfs แบบ command

ดาวน์โหลดข้อมูลจากเว็บ ไป ใส่ hdfs

1. yum install wget ติดตั้ง wget
2. wget "web"
3. yum install unzip
4. #ls คือดูว่า ไฟล์ที่โหลดมา อยู่ไหม
5. ทำการ unzip "ชื่อไฟล์"
6. ls
7. cd 201402\_babs\_open\_data/ เข้าไปในโฟลเดอร์ที่โหลดมา



```
201402_babs_open_data  boot  home  lost+found
babs_open_data_year_1.zip  dev  lib  media
[root@quickstart /]# cd 201402_babs_open_data/
[root@quickstart 201402_babs_open_data]#
```

8. copy file ไป Hdfs ไปโฟลเดอร์ที่ต้องการ

Ex ไปที่ user cloudera input #hdfs dfs -put 201402\_trip\_data.csv/user/cloudera

ตรวจสอบ ว่าเข้าป่าว หรืออาจจะ ตรวจสอบด้วยการ ls

สรุป wget ละ put

## คำสั่ง

#hdfs dfs -ls /user/cloudera ดูว่ามีอะไรอยู่ในโฟลเดอร์บ้าง

#hdfs dfs -mkdir /user/cloudera/rawdata สร้างโฟลเดอร์ที่ชื่อว่า rawdata อยู่ใน  
user/cloudera

#hdfs dfs -rm /user/cloudera/rawdata/201402\_trip\_data.csv

ลบไฟล์ 201402 ออกจาก rawdata

#hdfs dfs -mv /user/cloudera/201402\_trip\_data.csv /user/cloudera/rawdata  
ย้ายไฟล์

## ถ้าใช้ google storage

ไปที่ storage--> เลือก storage -> Browser -> สร้าง Bucket กรอกข้อมูลลงไป

เลือกcloud storage -> สร้าง folder

ใช้ cloud shell ไม่ต้อง yum install wget ใช้ได้เลยจ้า

#gsutil cp pg2600.txt gs://<ชื่อ bucket ที่ตั้งอะ>/data (เอาไปใส่ไว้ใน data)

เป็นคำสั่งให้เราเข้าถึง google cloud storage ได้

## HBASE

ต้องอยู่ใน root@quick #hbase shell

List ดูว่ามีตารางอะไรบ้าง?

create 'emp', 'personal', 'professional' สร้างตารางชื่อว่า emp มีคอลัมน์  
Familyอยู่ข้างใน ('personal', 'professional')

put 'emp', '1', 'personal:name', 'raju' ใส่ข้อมูลลงไป ระบุชื่อไป สร้าง  
key ชื่อว่าname

#scan 'name tale' ดูว่า อะไรอยู่ในตารางบ้าง พวกvalue

```
hbase(main):003:0> scan 'emp'
ROW          COLUMN+CELL
 1          column=personal:name, timestamp=1552559555310, value=Arisa
1 row(s) in 0.0710 seconds
hbase(main):004:0> █
```

#get 'emp', '1' เอาข้อมูลออกมาดูอ่ะ ตามที่กำหนด

สามารถเข้าไปดูใน HUE ได้ กด Data Browsers เลือก HBASE

ลบตาราง

```
exists 'emp'
•disable 'emp'
•drop 'emp'
•exists 'emp'
```

ลบเฉพาะคอลัมน์ในตาราง

**delete 'tableName', 'rowId', 'colFamily:colName', timestamp**

ลบแถวในตาราง **deleteall 'ชื่อตาราง', 'rowid'**