

先將老師的 image pull 使用老師給的 github 網站 ,到 終端 將以下全部的指令執行一次。

The screenshot shows a web browser window with the URL `github.com/sdwangntu/hadoop-cluster`. The page content includes instructions for pulling Docker images, creating an overlay network, launching the cluster, and launching a development container. Below the browser window, a terminal window is open, showing the execution of a Hadoop job. The terminal output displays a series of timestamps and counts, followed by the command to run the Hadoop job.

Pull image from dockerhub

- `docker pull sdwangntu/hive-metastore-db`
- `docker pull sdwangntu/hadoop3hbase-spark-hive`

Create an attachable overlay network called for example "my-attachable-network"

- `docker network create -d overlay --attachable my-attachable-network`

Launch cluster: (note the "hadoop-master" hostname is required as it is, the "hadoop-worker" hostname can be renamed to others; the "mysql" hostname is used in the hive setting - hive-site.xml)

- `docker run --hostname=mysql --name mysql --network my-attachable-network -d sdwangntu/hive-metastore-db`
- `docker run --hostname=hadoop-master --name hadoop-master --network my-attachable-network -d sdwangntu/hadoop3hbase-spark-hive`
- `docker run --hostname=hadoop-worker --name hadoop-worker --network my-attachable-network -d sdwangntu/hadoop3hbase-spark-hive`

Note the container execution order. And also check the startup routine "start-hadoop.sh" to see the roles played by each container.

Launch a development container (this is a hadoop development container and is used for development and job submission) the "hadoop-dev" hostname is required as it is

- `docker run --hostname=hadoop-dev --name hadoop-dev -v $(pwd):/home --network my-attachable-network -d sdwangntu/hadoop3hbase-spark-hive`

Terminal output:

```
11/Mar/2004:00 6
11/Mar/2004:01 1
11/Mar/2004:02 1
11/Mar/2004:03 6
11/Mar/2004:06 19
11/Mar/2004:07 6
11/Mar/2004:08 8
11/Mar/2004:10 3
11/Mar/2004:11 19
11/Mar/2004:12 17
11/Mar/2004:13 46
11/Mar/2004:14 15
11/Mar/2004:15 27
11/Mar/2004:16 6
11/Mar/2004:18 4
11/Mar/2004:20 14
11/Mar/2004:23 1
12/Mar/2004:01 1
12/Mar/2004:02 2
12/Mar/2004:03 1
12/Mar/2004:04 5
12/Mar/2004:05 11
12/Mar/2004:08 1
12/Mar/2004:09 2
12/Mar/2004:11 25
12/Mar/2004:12 22
12/Mar/2004:13 3
root@hadoop-worker:~/Cloud-Computing-hw3# hadoop jar /opt/hadoop/share/hadoop/tools/lib/hadoop-streaming-3.1.2.jar -D mapred.map.tasks=2 -mapper "python3 $PWD/python/mapper.py" -reducer "python3 $PWD/python/reducer.py" -input "/tmp/visitCount" -output "/tmp/visitCount_outdir"
```

用 hadoop 執行 mapping 與 reducer

```
root@hadoop-worker: ~/Cloud
11/Mar/2004:02 1
11/Mar/2004:03 6
11/Mar/2004:06 19
11/Mar/2004:07 6
11/Mar/2004:08 8
11/Mar/2004:10 3
11/Mar/2004:11 19
11/Mar/2004:12 17
11/Mar/2004:13 46
11/Mar/2004:14 15
11/Mar/2004:15 27
11/Mar/2004:16 6
11/Mar/2004:18 4
11/Mar/2004:20 14
11/Mar/2004:23 1
12/Mar/2004:01 1
12/Mar/2004:02 2
12/Mar/2004:03 1
12/Mar/2004:04 5
12/Mar/2004:05 11
12/Mar/2004:08 1
12/Mar/2004:09 2
12/Mar/2004:11 25
12/Mar/2004:12 22
12/Mar/2004:13 3
root@hadoop-worker:~/Cloud-Computing-hw3# hdfs dfs -ls /tmp/visitCount_outdir
Found 2 items
-rw-r--r-- 1 root supergroup 0 2020-11-04 07:22 /tmp/visitCount_outdir/_SUCCESS
-rw-r--r-- 1 root supergroup 1736 2020-11-04 07:22 /tmp/visitCount_outdir/part-00000
root@hadoop-worker:~/Cloud-Computing-hw3#
```

執行成功，觀看其結果

Java 版本:

建立環境變數

```
export HADOOP_CLASSPATH=$JAVA_HOME/lib/tools.jar
```

compile java 程式:

```
hadoop com.sun.tools.javac.Main visitCount.java
```

```
jar cf vc.jar visitCount*.class
```

執行 hadoop:

```
hadoop jar vc.jar visitCount /tmp/visit_count/access_log
```

```
/tmp/visitCountJava_outdir
```

結果圖:

```
07/Mar/2004:23 22
08/Mar/2004:00 21
08/Mar/2004:01 21
08/Mar/2004:02 27
08/Mar/2004:03 22
08/Mar/2004:04 26
08/Mar/2004:05 37
08/Mar/2004:06 17
08/Mar/2004:07 31
08/Mar/2004:08 44
08/Mar/2004:09 63
08/Mar/2004:10 39
08/Mar/2004:11 34
08/Mar/2004:12 45
08/Mar/2004:13 37
08/Mar/2004:14 23
08/Mar/2004:15 9
08/Mar/2004:16 2
08/Mar/2004:17 2
08/Mar/2004:18 9
08/Mar/2004:19 6
08/Mar/2004:20 23
08/Mar/2004:22 20
08/Mar/2004:23 1
09/Mar/2004:01 12
09/Mar/2004:02 15
09/Mar/2004:03 1
09/Mar/2004:05 24
09/Mar/2004:06 29
```

