Apache Hive 安裝手冊

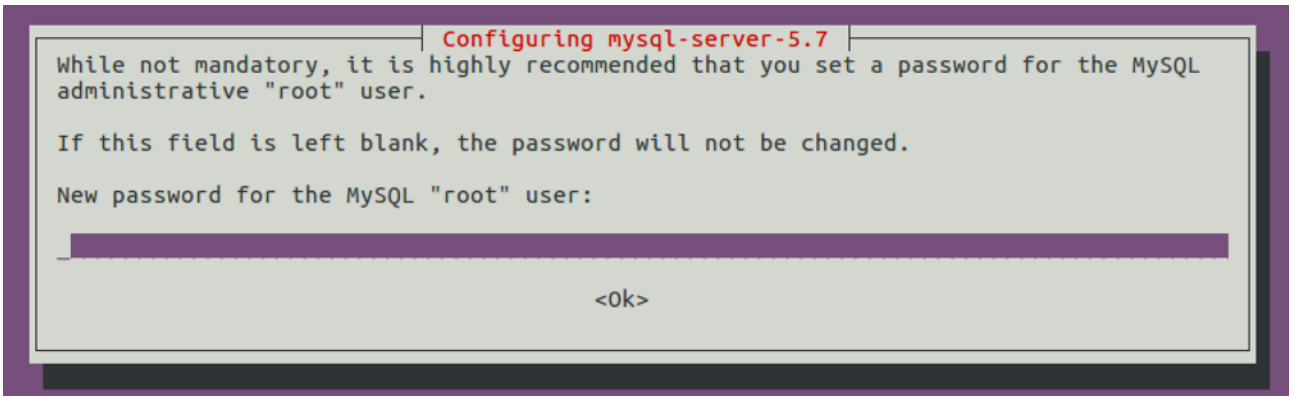
本文說明如何安裝Apache Hive、用Hive存資料與透過Java的JDBC存取Hive。

安裝MySQL

－在Terminal視窗中指令安裝：sudo apt-get install mysql-server

```
hduser@spark-single:~/Downloads$ sudo apt-get install mysql-server
[sudo] password for hduser:
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatically installed and are no longer required:
  linux-headers-4.4.0-31 linux-headers-4.4.0-31-generic linux-image-4.4.0-31-generic
  linux-image-extra-4.4.0-31-generic
Use 'sudo apt autoremove' to remove them.
The following additional packages will be installed:
  libevent-core-2.0-5 libhtml-template-perl mysql-client-5.7 mysql-client-core-5.7 mysql-common
  mysql-server-5.7 mysql-server-core-5.7
Suggested packages:
  libipc-sharedcache-perl mailx tinyca
The following NEW packages will be installed:
  libevent-core-2.0-5 libhtml-template-perl mysql-client-5.7 mysql-client-core-5.7 mysql-common
  mysql-server mysql-server-5.7 mysql-server-core-5.7
0 upgraded, 8 newly installed, 0 to remove and 195 not upgraded.
Need to get 18.3 MB of archives.
After this operation, 160 MB of additional disk space will be used.
Do you want to continue? [Y/n] Y
```

安裝過程中會詢問密碼，請輸入自訂密碼(不可空白或設為"root")



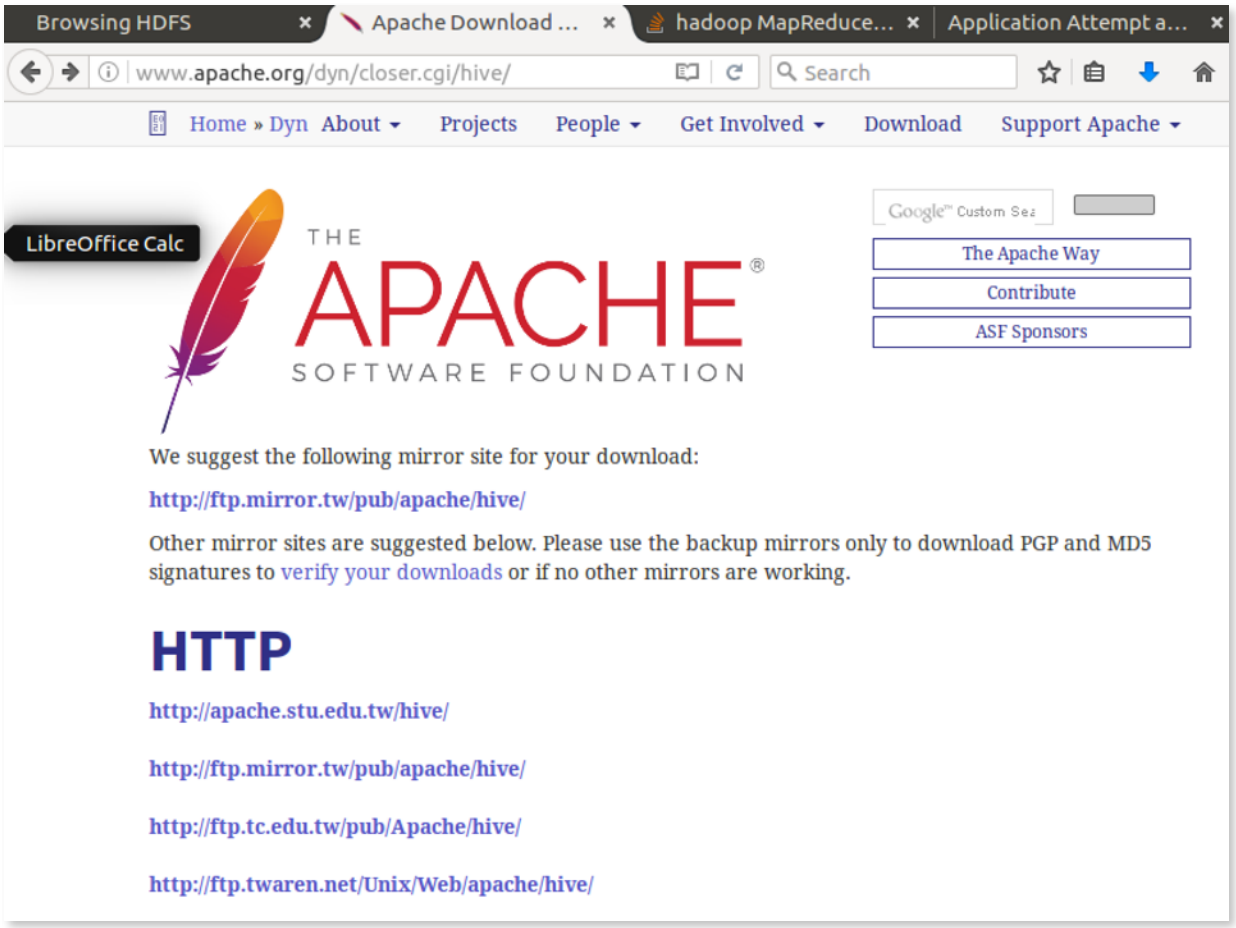
－安裝程序完成後，透過以下指令啟動及存取MySQL

- 啟動mysql服務 － sudo service mysql start
- 登入mysql － mysql -u root -p
- 顯示所有資料庫 － show databases;
- 退出mysql指令視窗 － quit

```
hduser@spark-single:~/Downloads$ sudo service mysql start
hduser@spark-single:~/Downloads$ mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 6
Server version: 5.7.16-0ubuntu0.16.04.1 (Ubuntu)

Copyright (c) 2000, 2016, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> show databases;
+--------------------+
| Database           |
+--------------------+
| information_schema |
| mysql              |
| performance_schema |
| sys                |
+--------------------+
4 rows in set (0.00 sec)

mysql>
```

Hive 安裝

－下載apache-hive-1.2.1-bin.tar.gz (http://ftp.tc.edu.tw/pub/Apache/hive/hive-1.2.1/)



– 完成hive下載後，解壓縮tar.gz檔，move到/usr/local下（指令如下：）

- tar zxvf apache-hive-1.2.1-bin.tar.gz
- sudo mv apache-hive-1.2.1-bin /usr/local/hive

```
hduser@spark-single:~/Downloads$
hduser@spark-single:~/Downloads$
hduser@spark-single:~/Downloads$ tar zxvf apache-hive-1.2.1-bin.tar.gz
```

```
hduser@spark-single:~/Downloads$
hduser@spark-single:~/Downloads$
hduser@spark-single:~/Downloads$ sudo mv apache-hive-2.1.1-bin /usr/local/hive
[sudo] password for hduser:
hduser@spark-single:~/Downloads$ ll /usr/local/hive
total 108
drwxrwxr-x  9 hduser hduser  4096 十二 11 11:40 ./
drwxr-xr-x 12 root   root    4096 十二 11 11:45 ../
drwxrwxr-x  3 hduser hduser  4096 十二 11 11:40 bin/
drwxrwxr-x  2 hduser hduser  4096 十二 11 11:40 conf/
drwxrwxr-x  4 hduser hduser  4096 十二 11 11:40 examples/
drwxrwxr-x  7 hduser hduser  4096 十二 11 11:40 hcatalog/
drwxrwxr-x  2 hduser hduser  4096 十二 11 11:40 jdbc/
drwxrwxr-x  4 hduser hduser 12288 十二 11 11:40 lib/
-rw-r--r--  1 hduser hduser 29003 十一 29 05:35 LICENSE
-rw-r--r--  1 hduser hduser   578 十一 29 22:09 NOTICE
-rw-r--r--  1 hduser hduser  4122 十一 29 05:35 README.txt
-rw-r--r--  1 hduser hduser 18501 十一 30 03:45 RELEASE_NOTES.txt
drwxrwxr-x  4 hduser hduser  4096 十二 11 11:40 scripts/
hduser@spark-single:~/Downloads$
hduser@spark-single:~/Downloads$
```

–將安裝路徑設定於bashrc

- sudo edit ~/.bashrc
- 將下面文字框內容加入.bashrc內容中
- source ~/.bashrc

```
#Hive Variables
export HIVE_HOME=/usr/local/hive
export PATH=$PATH:$HIVE_HOME/bin:$HIVE_HOME/conf
#Hive Variables
```

HIVE的設定

完成以上步驟，需再進行設定才能正常啟動hive；主要有以下幾個設定步驟：

－設定hive-site.xml

－設定hive-env.xml

－在HDSF建立hive使用目錄，並設定權限

－在mysql中建立儲存hive metastore之資料庫

－設定hive-site.xml，主要設定hive在HDSF上使用目錄的路徑及mysql連線資訊，操作步驟如下：

◦cd /usr/local/hive/conf

◦cp hive-default.xml.template hive-site.xml (一開始hive-site.xml不存在，由template建立)

◦若hive-site.xml已存在，則可略過上一步驟

◦sudo gedit hive-site.xml

◦確認hive-site.xml中以下設定：

－hive.metastore.warehouse.dir（應為 /user/hive/warehouse）

－hive.exec.scratchdir（應為 /tmp/hive）

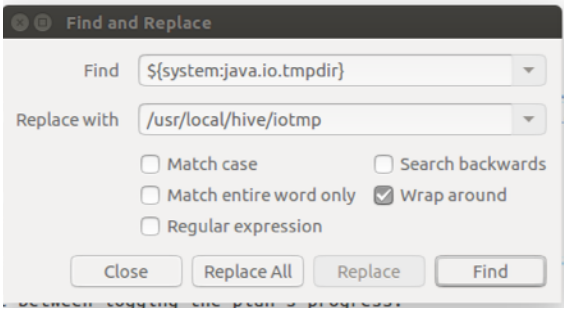
```
  <property>
    <name>hive.metastore.warehouse.dir</name>
    <value>/user/hive/warehouse</value>
    <description>location of default database for the warehouse</descriptio
n>
  </property>
```

```
  <property>
    <name>hive.exec.scratchdir</name>
    <value>/tmp/hive</value>
    <description>HDFS root scratch dir for Hive jobs which gets created wit
h write all (733) permission. For each connecting user, an HDFS scratch dir
: ${hive.exec.scratchdir}/&lt;username&gt; is created, with ${hive.scratch.
dir.permission}.</description>
  </property>
```

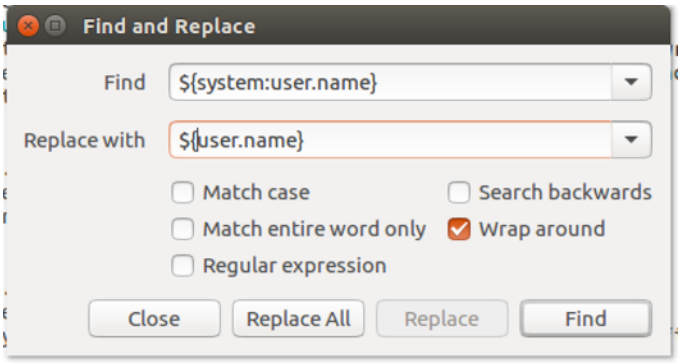
◦在hive-site.xml設定mysql連線資訊（將原使用Derby的部份改為mysql）

◦將hive-site.xml中的${system:java.io.tmpdir}取代為/usr/local/hive/iotmp（應取代4處），否則在執行hive時會出現java.net.URISyntaxExcepton

```
<property>
   <name>javax.jdo.option.ConnectionURL</name>
   <value>jdbc:mysql://localhost:3306/hive_metadata?
createDatabaseIfNotExist=true</value>
</property>

<property>
   <name>javax.jdo.option.ConnectionDriverName</name>
   <value>com.mysql.jdbc.Driver</value>
</property>

<property>
   <name>javax.jdo.option.ConnectionUserName</name>
   <value>hive</value>
</property>

<property>
   <name>javax.jdo.option.ConnectionPassword</name>
   <value>hive</value>
</property>
```



◦將hive-site.xml中的${system:user.name}取代為${user.name}（應取代3處），否則在執行hive時會出現java.net.URISyntaxExcepton



－設定hive-env.xml，主要設定HADOOP安裝目錄及hive設定檔位置，操作步驟如下：

◦cd /usr/local/hive/conf

◦cp hive-env.sh.template hive-env.sh (一開始hive-env.sh不存在，由template建立)

◦若hive-env.sh已存在，則可略過上一步驟

◦sudo gedit hive-env.sh

◦調整hive-env.sh中以下設定：

－HADOOP_HOME=/usr/local/hadoop

－export HIVE_CONF_DIR=/usr/local/hive/conf

－在HDSF建立hive使用目錄，並設定權限，步驟如下：

◦hadoop fs -mkdir -p /user/hive/warehouse（hive.metastore.warehouse.dir的設定）

◦hadoop fs -mkdir -p /tmp/hive（hive.exec.scratchdir的設定）

```
# Set HADOOP_HOME to point to a specific hadoop install directory
 HADOOP_HOME=/usr/local/hadoop

# Hive Configuration Directory can be controlled by:
 export HIVE_CONF_DIR=/usr/local/hive/conf
```

◦hadoop fs -chmod 777 /user/hive/warehouse

◦hadoop fs -chmod 777 /tmp/hive

◦透過 hadoop fs -ls -R / 指令查看HDFS目錄是否建立及權限設定

```
drwxrwxrwx   - hduser supergroup          0 2016-12-11 17:55 /tmp/hive
drwxr-xr-x   - hduser supergroup          0 2016-12-11 17:59 /user
drwxr-xr-x   - hduser supergroup          0 2016-12-11 17:54 /user/hive
drwxrwxrwx   - hduser supergroup          0 2016-12-11 17:54 /user/hive/warehouse
```

－取得mysql connector jar檔、並在mysql中建立儲存hive metastore之資料庫

◦為讓hive能順利存取mysql，需將mysql connector jar檔放入hive安裝目錄之lib目錄下

- 至https://dev.mysql.com/downloads/connector/j/ 下載最新之connector jar檔（目前為5.1.40，下載檔名為mysql-connector-java-5.1.40.tar.gz）
- 解壓縮tar.gz檔，並將解壓縮目錄中之jar檔copy到/usr/local/hive/lib中，指令如下：
  - tar zxvf mysql-connector-java-5.1.40.tar.gz
  - cp mysql-connector-java-5.1.40/mysql-connector-java-5.1.40-bin.jar /usr/local/hive/lib/

◦接著在MySQL shell中，建立一個專屬Hive的帳號並給予授權，指令步驟如下：

- mysql -u root -p

```
mysql> create database hive_metadata; #建立一個hive的database
Query OK, 1 row affected (0.00 sec)

mysql> grant all on *.* to'hive'@'%' identified by 'hive';  #建立一個MySQL使用者
Query OK, 0 rows affected, 1 warning (0.00 sec)

mysql> flush privileges;
Query OK, 0 rows affected (0.00 sec)

mysql> show databases;
+--------------------+
| Database           |
+--------------------+
| information_schema |
| hive_metadata      |
| mysql              |
| performance_schema |
| sys                |
+--------------------+
5 rows in set (0.00 sec)
```

  - mysql > create database hive_metadata; #建立一個hive的database
  - mysql> grant all on *.* to 'hive'@'%' identified by 'hive'; #建立一個MySQL使用者，帳號跟密碼都是hive，且用%代表在任何hostname都可登入
  - mysql> flush privileges; #更新User清單
  - mysql> select host, user from mysql.user; #查看所有帳號狀態
  - mysql> exit; #結束mysql

- ［註1］測試hive與metadata資料庫連線設定指令：schematool -initSchema -dbType mysql
- ［註2］測試hive帳號能否登入mysql：mysql -h localhost -u hive -p hive_metadata

啟動HIVE

完成上述的安裝及設定步驟，即可進行hive相關測試

- 在Terminal中輸入hive —version指令可檢視目前hive的版本

```
hduser@spark-single:~$ hive --version
Hive 2.1.1
Subversion git://jcamachorodriguez-rMBP.local/Users/jcamachorodriguez/src/worksp
aces/hive/HIVE-release2/hive -r 1af77bbf8356e86cabbed92cfa8cc2e1470a1d5c
Compiled by jcamachorodriguez on Tue Nov 29 19:46:12 GMT 2016
From source with checksum 569ad6d6e5b71df3cb04303183948d90
hduser@spark-single:~$
```

- 在Terminal中輸入hive指令即可啟動hive shell（看到hive>才表示正常啟動）

```
hduser@spark-single:~$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hive/lib/log4j-slf4j-impl-2.4.1.jar
!/org/slf4j/impl/StaticLoggerBinder.class]
```

```
hive>
```

- 在hive shell中輸入以下指令，建立及檢視hive table：
  - create table test(id String);
  - show tables;
  - desc test;

```
hive> show tables;
OK
Time taken: 1.012 seconds
hive> create table test(id String);
OK
Time taken: 0.45 seconds
hive> show tables;
OK
test
Time taken: 0.052 seconds, Fetched: 1 row(s)
hive> desc test;
OK
id                      string
Time taken: 0.148 seconds, Fetched: 1 row(s)
hive>
```

  - ctrl + c 退出hive shell，在Terminal中輸入hadoop fs -ls -R /user/hive指令，可看到在HDSF的/user/hive/warehouse底下有test目錄被建立

```
hduser@spark-single:~$ hadoop fs -ls -R /user/hive
drwxrwxrwx   - hduser supergroup          0 2016-12-12 01:21 /user/hive/warehouse
drwxrwxrwx   - hduser supergroup          0 2016-12-12 01:21 /user/hive/warehouse/test
hduser@spark-single:~$
```