



Hadoop Install on Ubuntu



Outline

- Before install
 - Install Hadoop(single node)
 - Launch & Test
- 



Outline

- Before install
- Install Hadoop(single node)
- Launch & Test



Before Install

- 在這篇tutorial 所有指令都是以一般user身分下完成 而非sudoer(root)
- 安裝ssh和vim

```
sudo apt-get install ssh vim
```

Install sun java JDK

- 通常Linux 預設的java版本是openjdk ,但只有Sun JDK 支援Hadoop ,所以這邊要把JDK 改成SUN的版本
- 這篇投影片是以ubuntu 64-bit 環境下作的 注意若在32-bit 要裝不同版本的jdk
- 下載 jdk-7u40-linux-x64.gz (<http://tinyurl.com/cnafy3t>)
- 下載完後,解壓縮並安裝

```
tar -xvf jdk-7u40-linux-x64.gz
sudo mkdir /usr/lib/jvm
sudo mv jdk1.7.0_40/ /usr/lib/jvm/
sudo update-alternatives --install "/usr/bin/java" "java" "/usr/lib/jvm/jdk1.7.0_40/bin/java" 1
sudo update-alternatives --install "/usr/bin/javac" "javac" "/usr/lib/jvm/jdk1.7.0_40/bin/javac" 1
sudo update-alternatives --install "/usr/bin/javaws" "javaws" "/usr/lib/jvm/jdk1.7.0_40/bin/javaws" 1
sudo update-alternatives --install "/usr/bin/jar" "jar" "/usr/lib/jvm/jdk1.7.0_40/bin/jar" 1
sudo update-alternatives --config java
sudo update-alternatives --config jar
sudo update-alternatives --config javac
sudo update-alternatives --config javaws
java -version
```

Install sun java JDK


- 若出現此畫面 表是修改成功

```
java version "1.7.0_40"  
Java(TM) SE Runtime Environment (build 1.7.0_40-b43)  
Java HotSpot(TM) 64-Bit Server VM (build 24.0-b56, mixed mode)
```



Outline

- Before install
 - Install Hadoop(single node)
 - Launch & Test
- 



Install Hadoop(Single node)

- 下載並解壓縮 hadoop-1.2.1

```
wget http://apache.cdpa.nsysu.edu.tw/hadoop/common/hadoop-1.2.1/hadoop-1.2.1.tar.gz  
tar -xvf hadoop-1.2.1.tar.gz
```




Install Hadoop(Singlenode):Configuration

- 在hadoop-1.2.1資料夾下
- Edit conf/hadoop_env.sh:
在其中加一行

```
export JAVA_HOME=/usr/lib/jvm/jdk1.7.0_40/
```

Install Hadoop(Single node):Configuration

- 編輯 conf/core-site.xml:

```
<configuration>
  <property>
    <name>fs.default.name</name>
    <value>hdfs://hostname:9000</value>
  </property>
</configuration>
```

Note: **hostname**是指電腦的名稱,若不知hostname,可下指令 **hostname** 查詢



Install Hadoop(Single node):Configuration

- 編輯 conf/hdfs-site.xml:

```
<configuration>
```

```
  <property>
```

```
    <name>dfs.replication</name>
```

```
    <value>1</value>
```

```
  </property>
```

```
</configuration>
```



Install Hadoop(Single node):Configuration

- 編輯 conf/mapred-site.xml:

```
<configuration>
```

```
  <property>
```

```
    <name>mapred.job.tracker</name>
```

```
    <value>hostname:9001</value>
```

```
  </property>
```

```
</configuration>
```

Note: **hostname**是指電腦的名稱

Setup Passphraseless ssh & Set PATH

- 使之後ssh登入不需要密碼

```
ssh-keygen -t dsa -P "" -f ~/.ssh/id_dsa  
cat ~/.ssh/id_dsa.pub >> ~/.ssh/authorized_keys
```

- 編輯~/.bashrc

```
vim ~/.bashrc
```

在最後加一行

```
export PATH=$PATH:"hadoop路徑"/bin
```

```
export PATH=$PATH:"hadoop路徑"/bin
```

Note:hadoop路徑指的是hadoop-1.2.1資料夾的絕對路徑



Outline

- Before install
 - Install Hadoop(single node)
 - Launch & Test
- 



Execution

- 格式化HDFS

```
hadoop namenode -format
```

- 運行hadoop

```
start-all.sh
```

- 檢查 hadoop 是否運行正常

```
hadoop dfsadmin -report
```

- 當要關掉hadoop時

```
stop-all.sh
```

Execution

結果:若有如下畫面 表示HDFS成功運行 HDFS底下有一個datanode

```
ubuntu@ip-10-176-226-70:~/hadoop-1.2.1$ hadoop dfsadmin -report
```

```
Configured Capacity: 8455118848 (7.87 GB)
```

```
Present Capacity: 6385065999 (5.95 GB)
```

```
DFS Remaining: 6385037312 (5.95 GB)
```

```
DFS Used: 28687 (28.01 KB)
```

```
DFS Used%: 0%
```

```
Under replicated blocks: 0
```

```
Blocks with corrupt replicas: 0
```

```
Missing blocks: 0
```

```
-----  
Datanodes available: 1 (1 total, 0 dead)
```

```
Name: 10.176.226.70:50010
```

```
Decommission Status : Normal
```

```
Configured Capacity: 8455118848 (7.87 GB)
```

```
DFS Used: 28687 (28.01 KB)
```

```
Non DFS Used: 2070052849 (1.93 GB)
```

```
DFS Remaining: 6385037312(5.95 GB)
```

```
DFS Used%: 0%
```


```
DFS Remaining%: 75.52%
```

```
Last contact: Sat Sep 21 15:35:48 UTC 2013
```




Execution

- http://hadoop.apache.org/docs/r0.18.3/hdfs_shell.html
- Example:
 - `hadoop fs -ls /` 查詢HDFS內的檔案
 - `hadoop fs -put input /` 將本機的input檔案送到HDFS中



Browse the web interface for the NameNode and the JobTracker

- NameNode - <http://localhost:50070/>
 - JobTracker - <http://localhost:50030/>
- 

Test: Word Count

- 在hadoop-1.2.1資料夾底下
 - `vim input` 在本機端建立input檔案,檔案內容一篇文字檔即可
 - `hadoop fs -put input /` 將input 檔案放入HDFS
 - `hadoop jar hadoop-examples-1.2.1.jar wordcount /input /output` 執行wordcount
 - `hadoop fs -cat /output/*` 看執行結果
- 執行結果:

```
road, 1
robbery 1
said 3
say 1
scene 2
sending 1
senior 1
seven 1
shooting 1
shootout 1
shoppers 1
shoppers. 1
shot. 1
someone 1
still 1
store, 1
story 2
strike 1
surrounded 1
taken 1
terrorist 1
that 1
```



Question

- 這份文件只教single node安裝,若想玩玩看cluster,可以上hadoop 網站 (http://hadoop.apache.org/docs/stable/cluster_setup.html),或是也可以上討論區發問
- 