連線H204 Linux PC cluster主機使用範例

****

**電腦規格**

CPU：8核心 3.40GHz \*8

RAM：32 GB

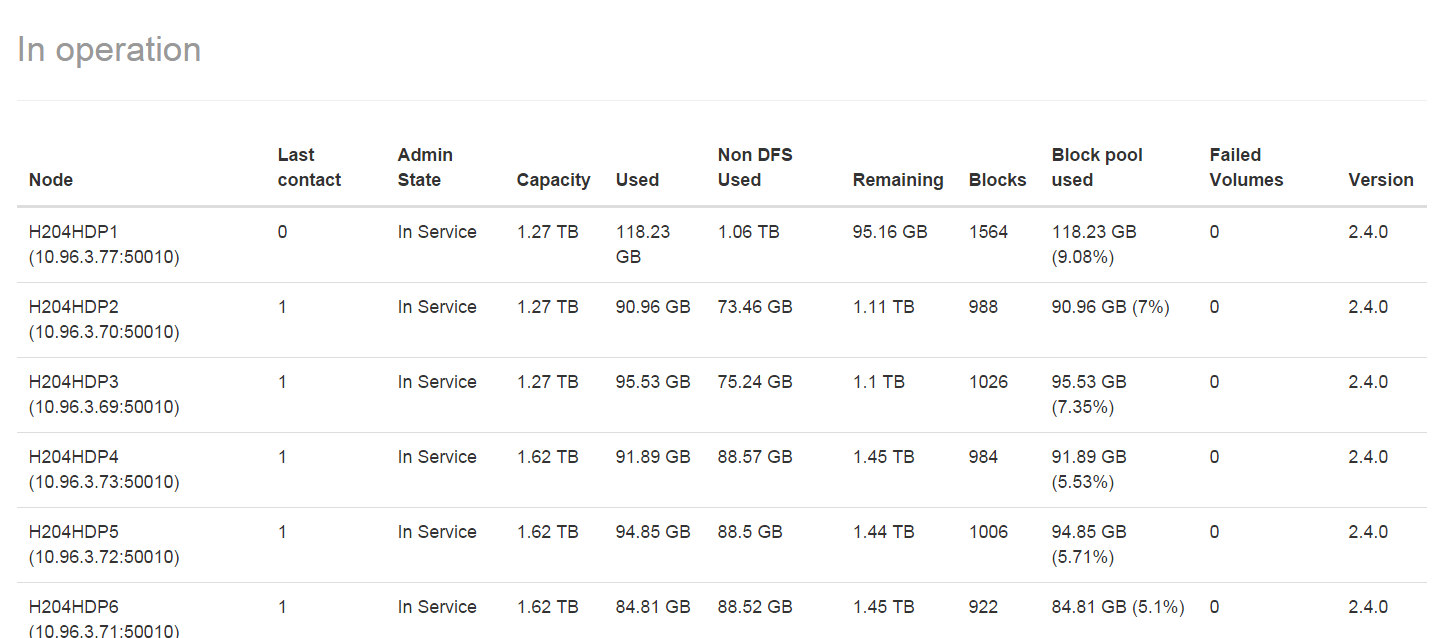
HD ：2 TB

OS ：ubuntu 14.04 64bit

**Hadoop Cluster**

Hadoop 2.4.0 Version

6 nodes hadoop cluster



1. putty

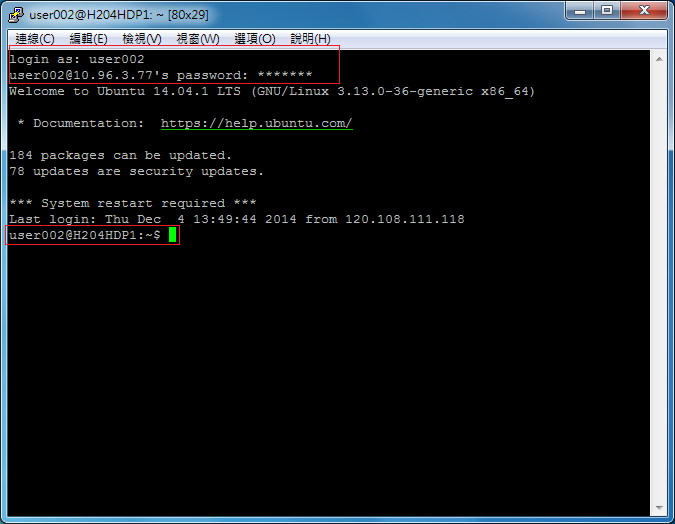
輸入IP : 10.96.3.77



**ListofHadoopUsers 帳密**

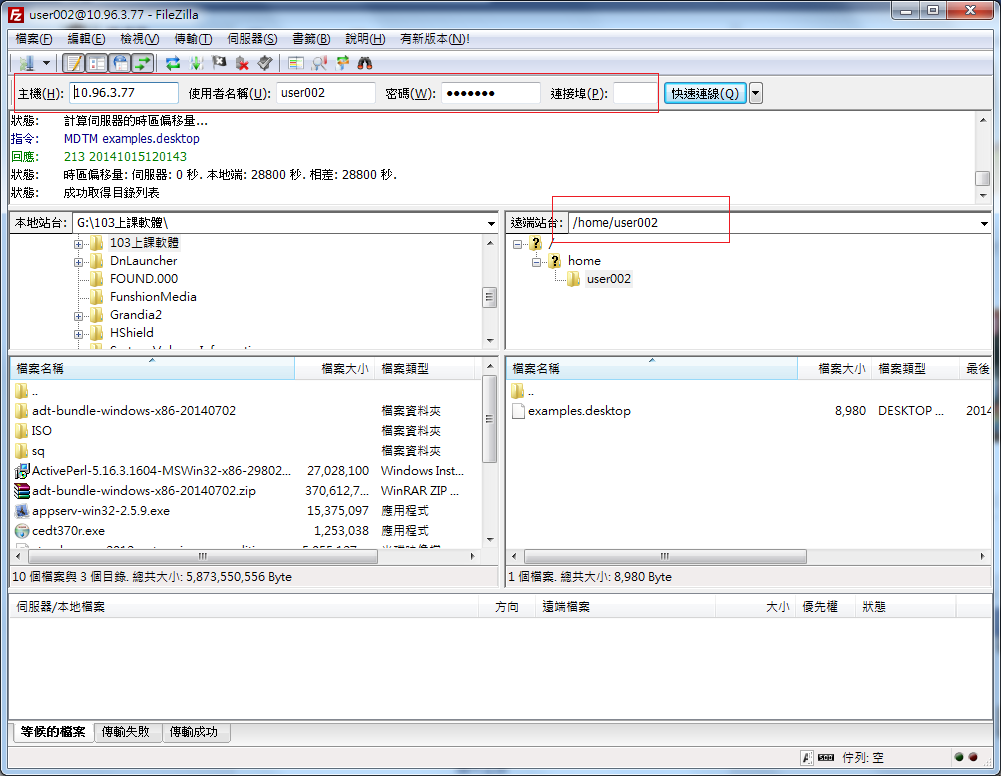
(課堂確定)

輸入帳密後，「顯示userXXX@H204HDP1:~$　」，表是連線成功。



1. FTP （程式+資料）

使用FileZilla FTP軟體，連線至Hadoop。



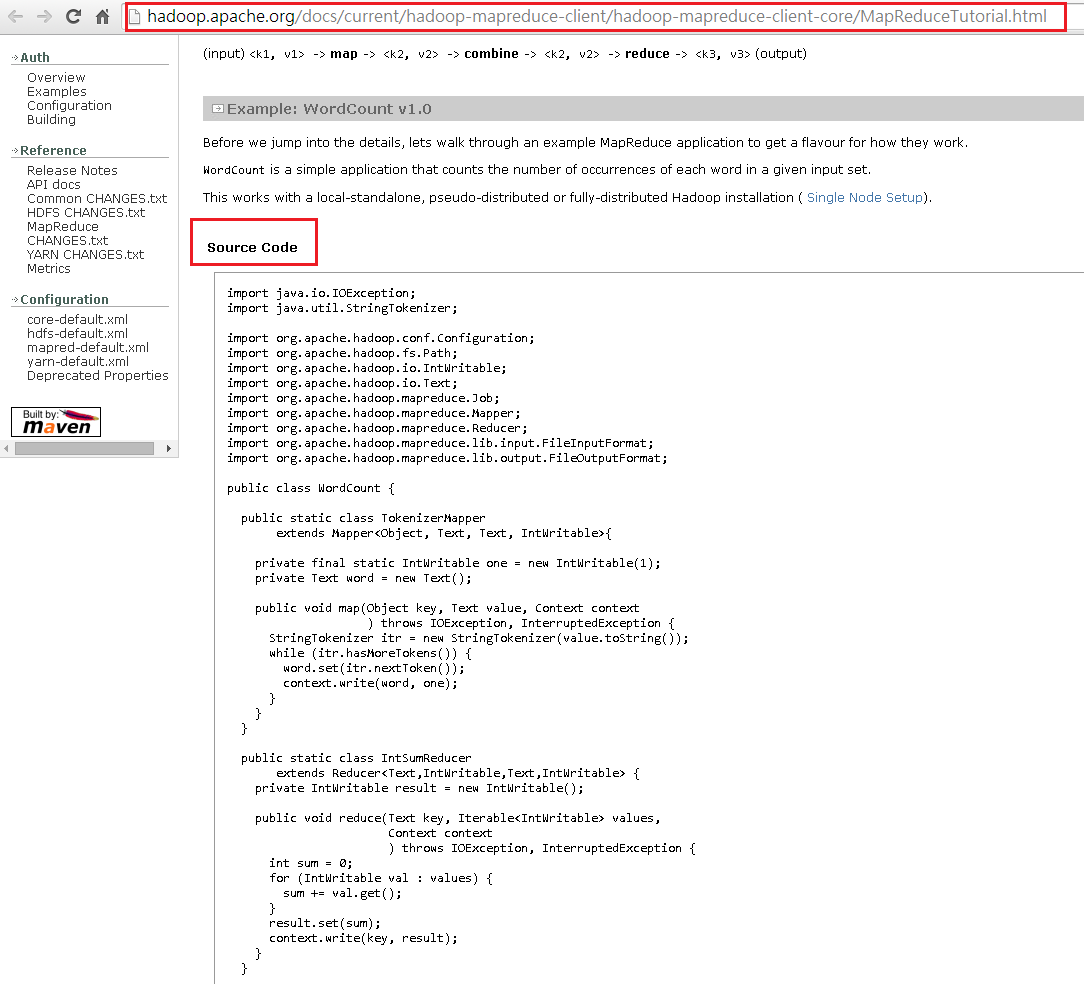
上傳檔案至hadoop



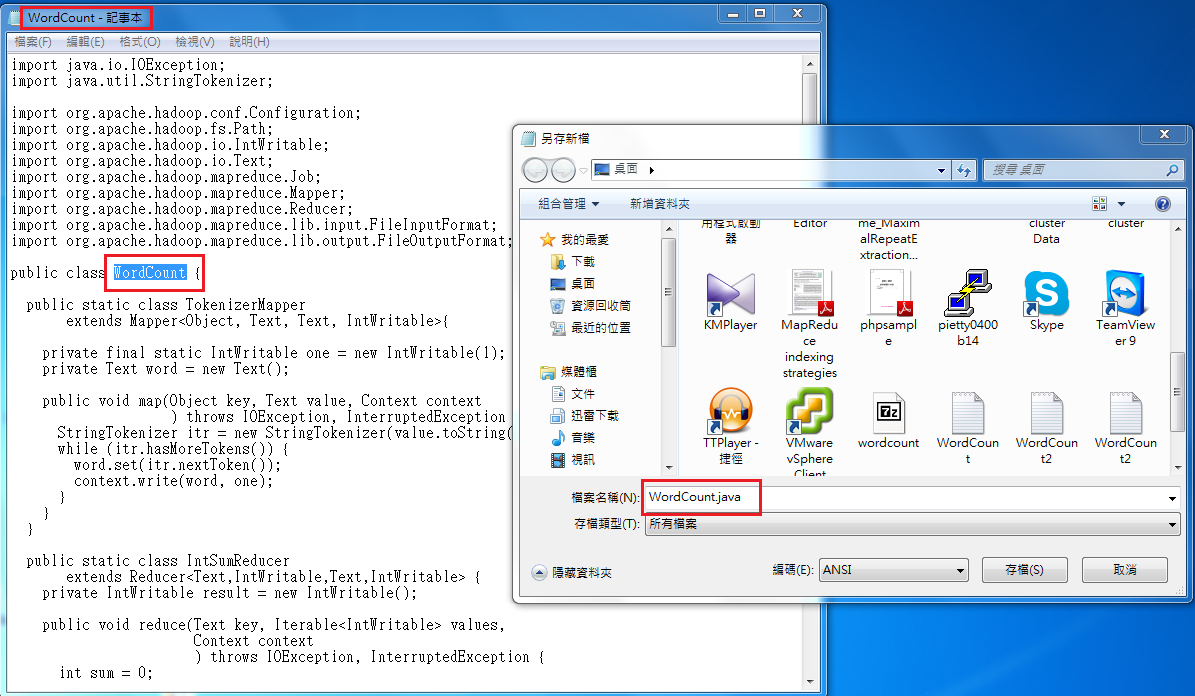
1. Wordcount Map Reduce Example

http://hadoop.apache.org/docs/current/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html

首先到這網站下載wordcount Source Code



複製程式碼至記事本，另存新檔：WordCount.java



把WordCount.java檔上傳至hadoop

請在路徑: /home/user002/ 底下新增一個資料夾: wordcount

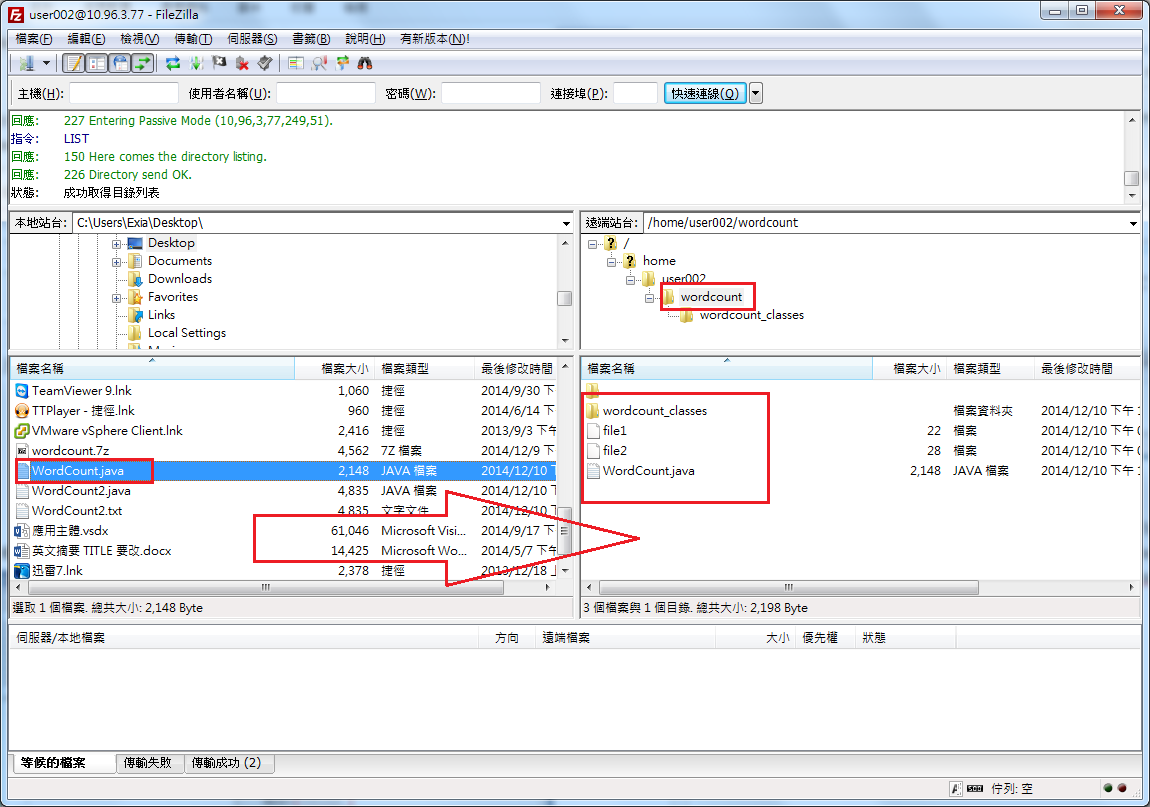
**Wordcount資料夾**裡要有三個檔案、一個資料夾

wordcount\_classes (資料夾)

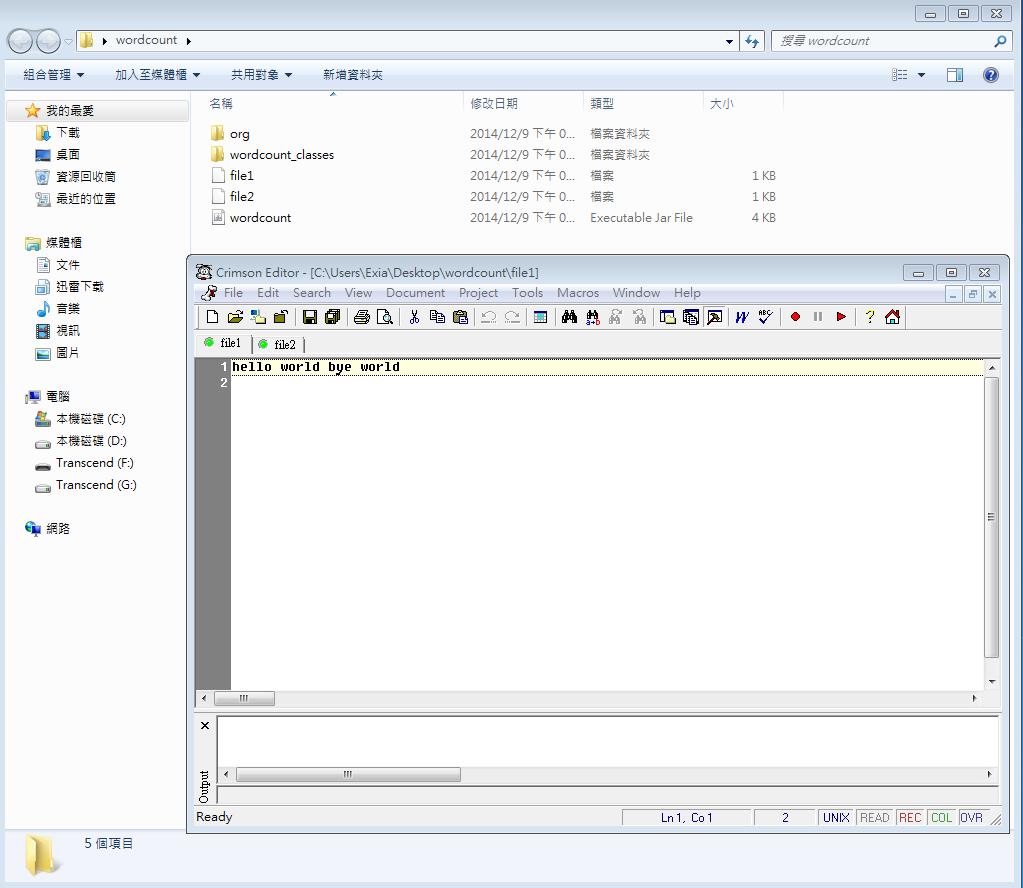
WordCount.java (程式碼)

File1.text (input)

File2.text (input)

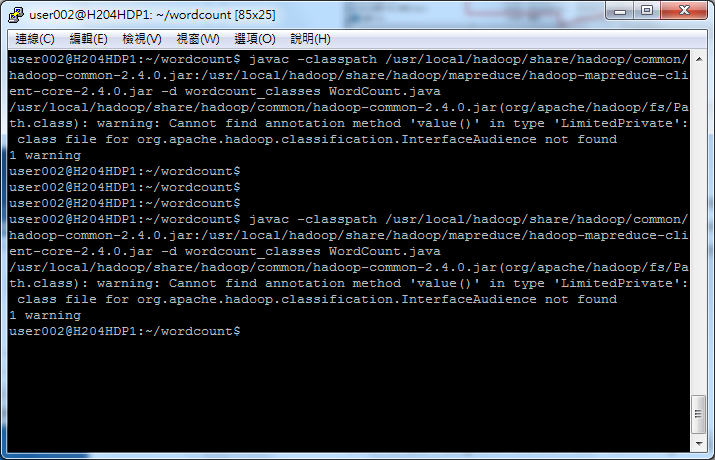


File1.txt input檔:內容(wordcount 要查詢的內容)



1. Compile Code

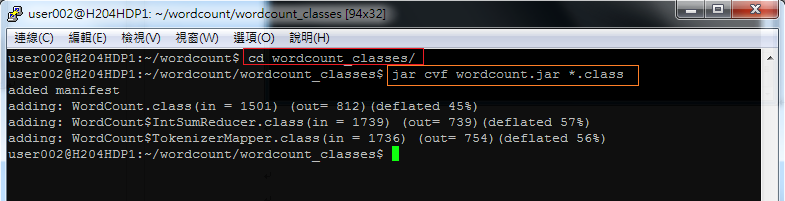
javac -classpath /usr/local/hadoop/share/hadoop/common/hadoop-common-2.4.0.jar:/usr/local/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-client-core-2.4.0.jar -d wordcount\_classes WordCount.java



1. Make Jar File

cd wordcount\_classes/

jar cvf wordcount.jar \*.class

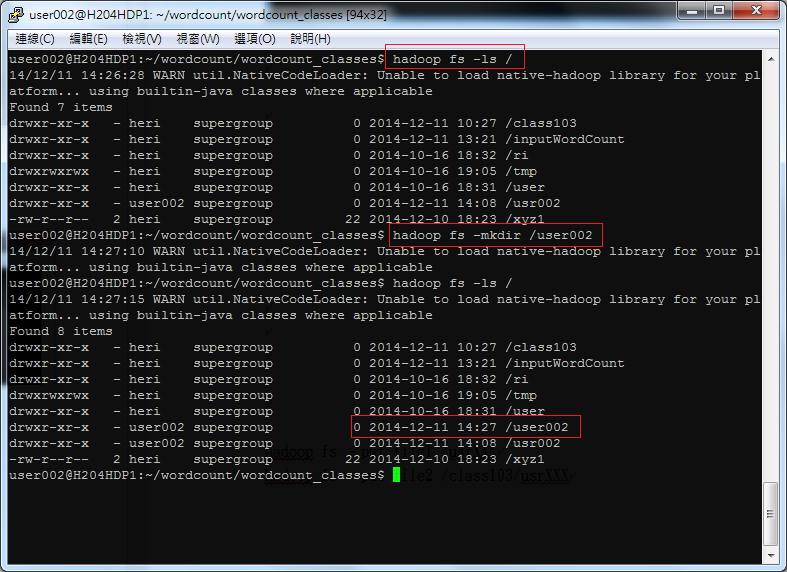


1. Put input file to Hadoop File System

從local端傳送檔案至HDFS (Hadoop File System)

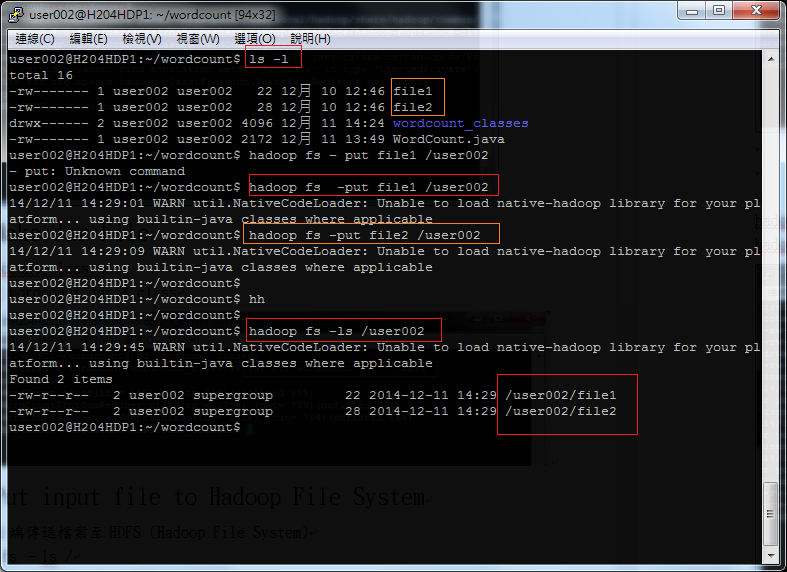
hadoop fs -ls / (查詢HDFS上的使用者)

hadoop fs -mkdir /userXXX (在HDFS上新增userXXX資料夾)



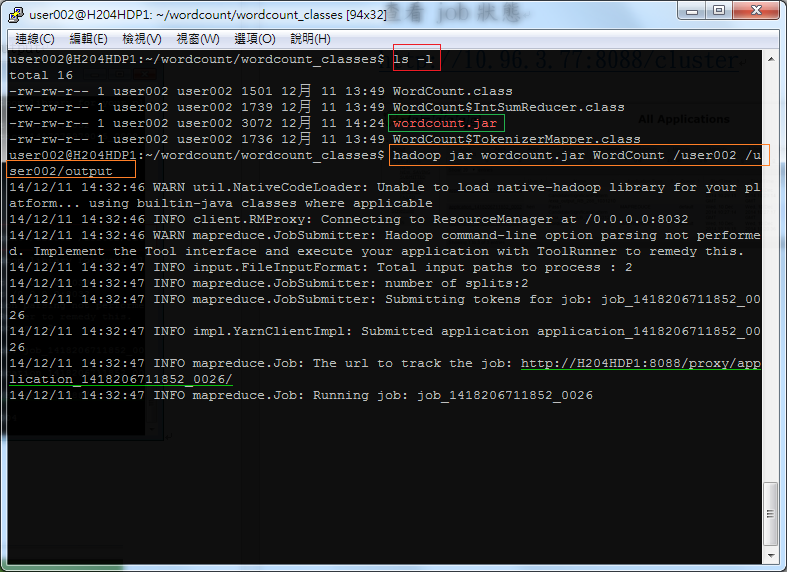
hadoop fs -put file1 /userXXX (把file1檔上傳至userXXX資料夾裡)

hadoop fs -put file2 /userXXX (把file2檔上傳至userXXX資料夾裡)



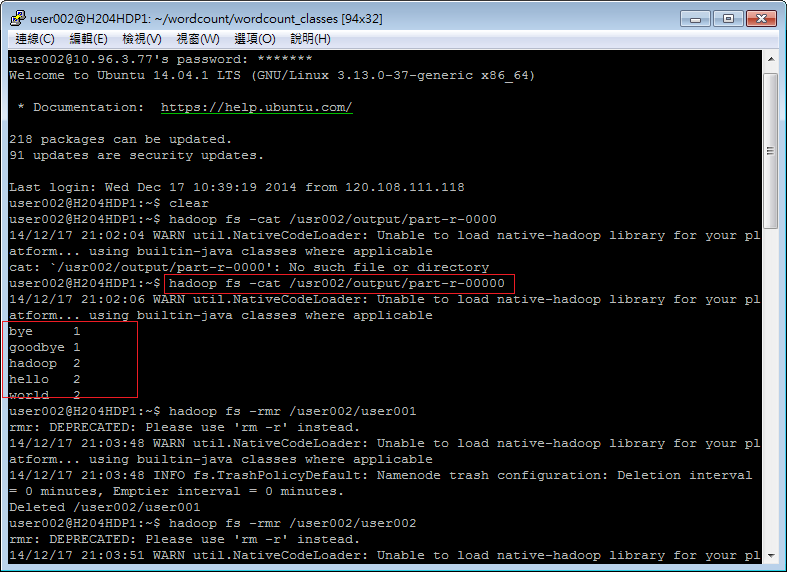
1. Running WordCount

hadoop jar wordcount.jar WordCount /userXXX /userxxx/output



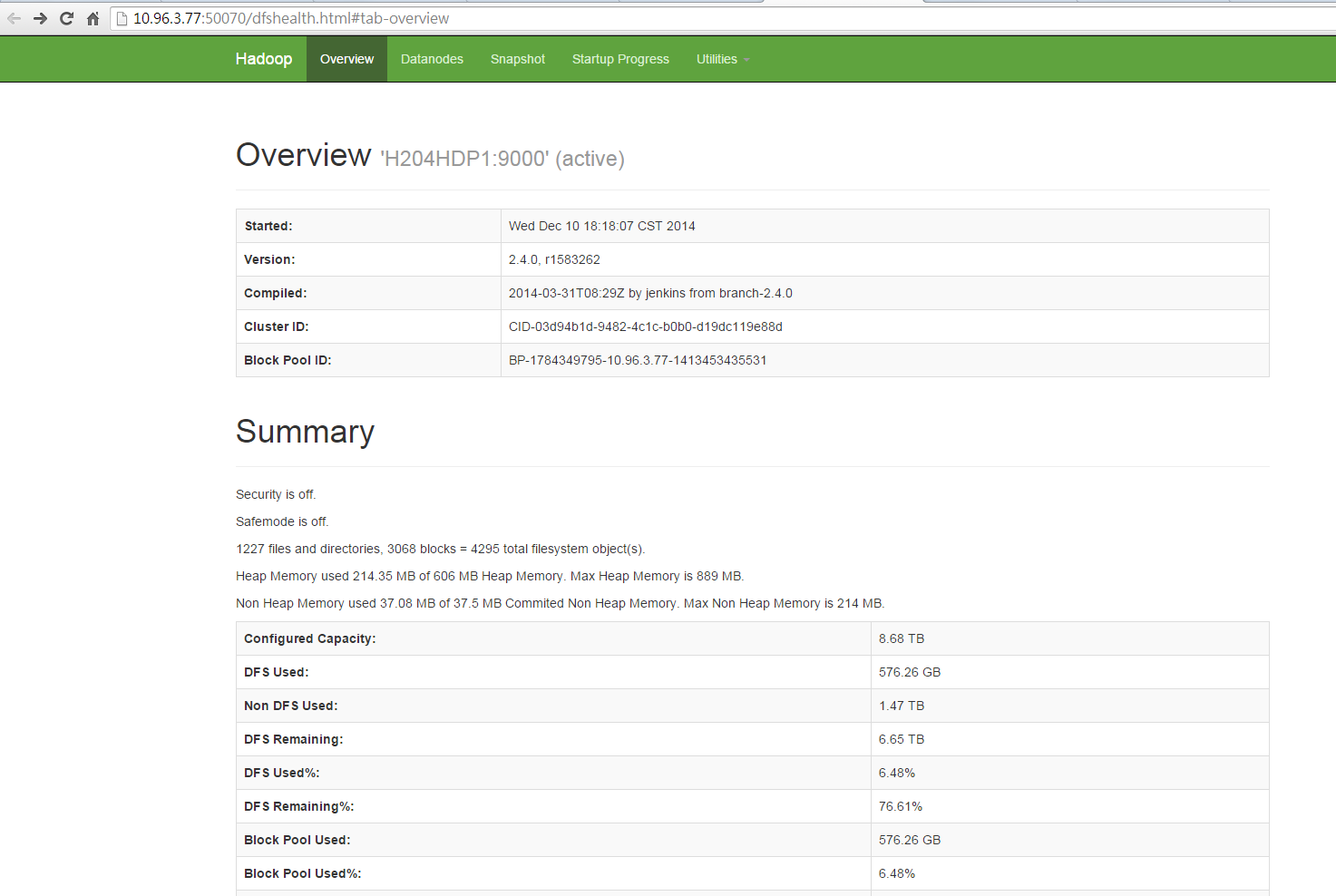
查看結果

Hadoop fs -cat /user002/output/part-r-00000 (output檔目的地)

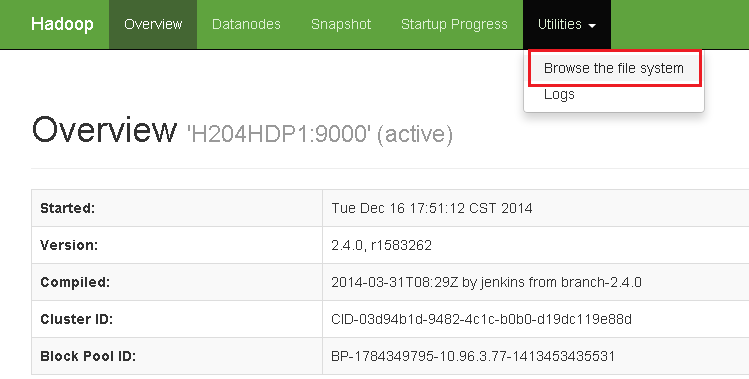


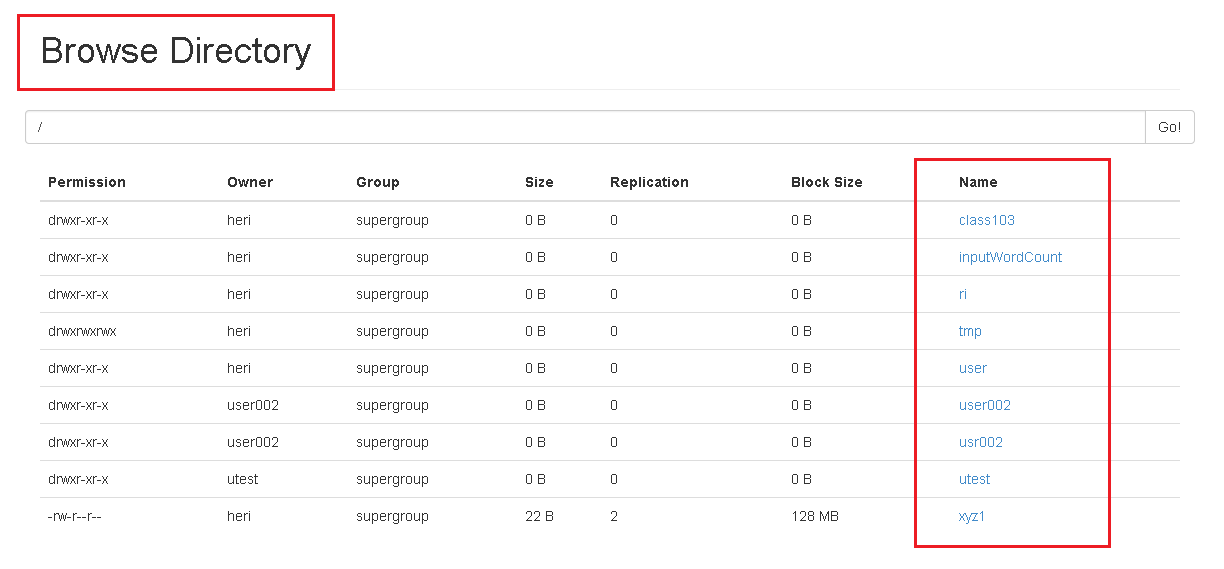
可查看hadoop 狀態

<http://10.96.3.77:50070/>



點選Browse the file system查看HDFS上的檔案





查看job狀態

<http://10.96.3.77:8088/cluster>

