



# 大數據資料處理 – Hadoop MapReduce Programming

交通部高速公路闡道資料  
(<http://tisvcloud.freeway.gov.tw/>)  
+ 網路資料自動下載程式

車輛(VehicleType) +24小時統計  
闡道(GantryID\_On) +24小時統計  
特定闡道(GantryID\_On)+24小時統計

授課教授：王經篤 (Jing-Doo Wang)

Email: [jdwang@asia.edu.tw](mailto:jdwang@asia.edu.tw)

亞洲大學 資訊工程系 (CSIE, Asia University)

- TDCS資料擷取範例程式:

- [http://dns2.asia.edu.tw/~jdwang/TeachingCourses/Hadoop/TDCS\\_GIDSequence\\_MapReduceParser\\_24Hour\\_jdwang2017\\_10\\_13.zip](http://dns2.asia.edu.tw/~jdwang/TeachingCourses/Hadoop/TDCS_GIDSequence_MapReduceParser_24Hour_jdwang2017_10_13.zip)

- 注意：需如匯入WordCount\_jdwang 專案

- 重新載入 Hadoop external Jar

- 設定Project環境變數 ( HADOOP\_HOME, PATH )

WordCount\_jdwang 的內容

輸入過濾器文字

- > 資源
- > 建置器
- > 執行/除錯設定
- > 專案參照
- Java 建置路徑**
- > Java 程式撰寫風格
- > Java 編輯器
- > Java 編譯器
- > Javadoc 位置

遺漏 94 建置路徑項目。

程式碼(S) 專案(P) 程式庫(L) 次序與匯出(O)

建置路徑上的 JAR 與類別資料夾(T) :

- > activation-1.1.jar - H:\windoop\hadoop\share\hadoop\cor
- > apacheds-i18n-2.0.0-M15.jar - H:\windoop\hadoop\share
- > apacheds-kerberos-codec-2.0.0-M15.jar - H:\windoop\had
- > api-asn1-api-1.0.0-M20.jar - H:\windoop\hadoop\share\h
- > api-util-1.0.0-M20.jar - H:\windoop\hadoop\share\hadoop
- > asm-3.2.jar - H:\windoop\hadoop\share\hadoop\commor
- > avro-1.7.4.jar - H:\windoop\hadoop\share\hadoop\comm
- > commons-beanutils-1.7.0.jar - H:\windoop\hadoop\share\
- > commons-beanutils-core-1.8.0.jar - H:\windoop\hadoop\s
- > commons-cli-1.2.jar - H:\windoop\hadoop\share\hadoop\
- > commons-codec-1.4.jar - H:\windoop\hadoop\share\hadc
- > commons-collections-3.2.1.jar - H:\windoop\hadoop\shar
- > commons-compress-1.4.1.jar - H:\windoop\hadoop\share
- > commons-configuration-1.6.jar - H:\windoop\hadoop\sha
- > commons-digester-1.8.jar - H:\windoop\hadoop\share\ha

新增 JAR(J)...  
新增外部 JAR(X)...  
新增變數(V)...  
Add Library...  
新增類別資料夾(C)...  
Add External Class Folder...  
編輯(E)...  
移除(R)  
移轉 JAR 檔...

套用(A)

確定 取消

先點選  
Java  
建置路徑

會發現所有的jar檔都有紅色的XX

```
BAD_ID=0  
CONNECTION=0  
IO_ERROR=0  
WRONG_LENGTH=0  
WRONG_MAP=0  
WRONG_REDUCE=0  
File Input Format Counters  
Bytes Read=9000000  
File Output Format Counters  
Bytes Written=28
```

Job Done!

編輯配置

編輯啟動配置內容

執行 Java 應用程式

※請在名稱輸入1.值也輸入1.  
按下確認，才能在新增 2

名稱(N) : WordCount\_jdwang

主要 引數 Java 執行時期環境(JRE) 類別路徑 來源 環境 共用(C)

要設定的環境變數(S) :

分別輸入1. HADOOP\_HOME  
2. PATH

變數

新建環境變數

名稱(N) :

值(V) :

變數(B)...

確定

取消

新建(E)...

選取(L)...

編輯(D)...

移除(O)

附加環境到原生環境(A)

以指定的環境取代原生環境(P)

1. \${eclipse\_home}\..\Hadoop

2. %PATH%;\${eclipse\_home}\..\hadoop\bin

回復(V)

套用(Y)



確定

取消