



大數據資料處理 – Hadoop MapReduce Programming

Hadoop File System (HDFS) 系統操作

授課教授：王經篤 (Jing-Doo Wang)

Email: jdwang@asia.edu.tw

亞洲大學 資訊工程系 (CSIE, Asia University)

學習目標

- Windoop 之 HDFS 檔案與目錄之上傳與下載

首先呢我們先依照上個單元的執行方式，先執行eclipse，讓DFS Location出現input跟output

```
package org.windoop;

import java.io.BufferedWriter;
import java.io.OutputStreamWriter;
import java.util.Random;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.*;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.lib.input.*;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class presidentElection {

    public static void main(String[] args) throws Exception {
        String[] candidate = {"A", "B", "C"};
        String[] location = {"Taichung", "Taipei", "Kaohsiung"};
        Random r;
        boolean useResourceManager = true;
        //boolean useJobTracker = true; <-- for hadoop 1.0

        //Step 0. Set configuration
```

快速存取 | Java | Map/Reduce

概要

- org.windoop
 - presidentElection
 - main(String[]) : void

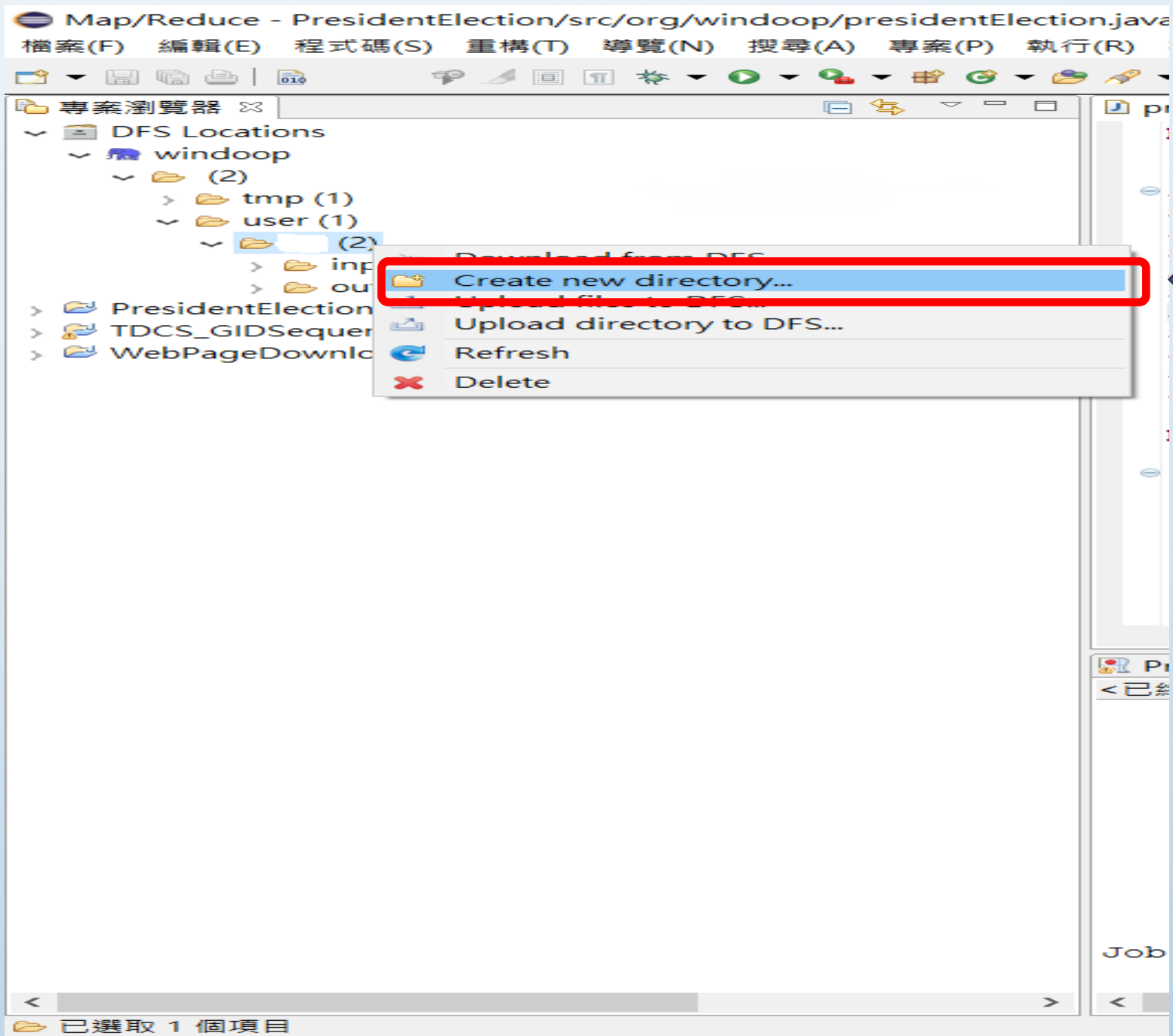
Problems | Tasks | Javadoc | Map/Reduce Locations | 主控台

<已終止> presidentElection [Java 應用程式] C:\Program Files\Java\jre1.8.0_121\bin\javaw.exe (2017年7月5日 上午9:40:17 直接存取)

```
Total committed heap usage (bytes)=735576064
Shuffle Errors
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!
```

切記執行結果一定要顯示job done!

已選取 1 個項目

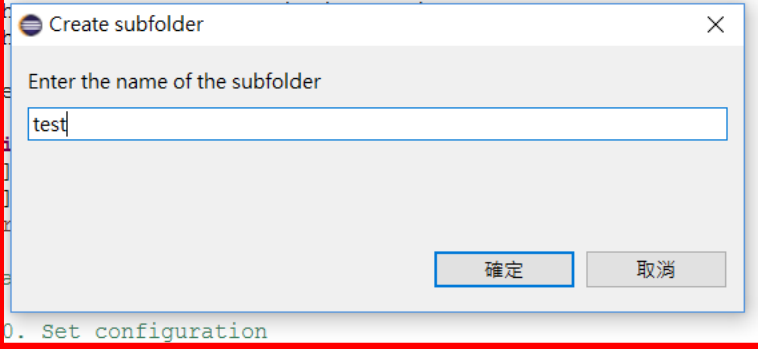


點選user底下資料夾右鍵，在使用者底下新增一個資料夾

DFS Locations

- windoop
 - (2)
 - tmp (1)
 - user (1)
 - (2)
 - input (3)
 - output (2)
 - PresidentElection
 - TDCS_GIDSequence_MRE
 - WebPageDownload_jdwang

```
package org.windoop;  
  
import java.io.BufferedWriter;  
import java.io.OutputStreamWriter;  
import java.util.Random;  
import org.apache.hadoop.conf.Configuration;  
import org.apache.hadoop.fs.*;  
import org.apache.hadoop.io.IntWritable;  
import org.apache.hadoop.io.Text;  
import org.apache.hadoop.mapreduce.lib.input.*;  
import org.apache.hadoop.mapreduce.lib.output.*;  
import org.apache.hadoop.mapreduce.Mapper;  
import org.apache.hadoop.mapreduce.Reducer;  
  
public class presidentElection {  
  
    public static void main(String[] args) {  
        String inputPath = args[0];  
        String outputPath = args[1];  
        Random r = new Random();  
        boolean isTest = false;  
        //boolean isTest = true;  
  
        //Step 0. Set configuration
```



輸入完想要的資料夾
名稱按下確認

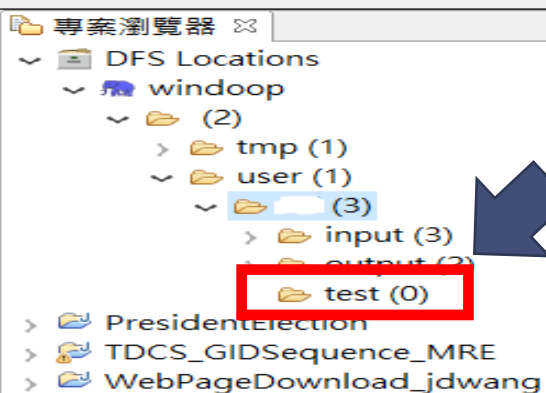
Problems Tasks Javadoc Map/Reduce Locations 主控台

<已終止> presidentElection [Java 應用程式] C:\Program Files\Java\jre1.8.0_121\bin\javaw.exe (2017年7月5日 上午9:40:17 直接存取)

```
Total committed heap usage (bytes)=735576064  
Shuffle Errors  
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!





重新整理完後，
就會看到自己新增的資料夾

```
package org.windoop;

import java.io.BufferedWriter;
import java.io.OutputStreamWriter;
import java.util.Random;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.*;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.lib.input.*;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class presidentElection {

    public static void main(String[] args) throws Exception {
        String[] candidate = {"A","B","C"};
        String[] location = {"Taichung", "Taipei", "Kaohsiung"};
        Random r;
        boolean useResourceManager = true;
        //boolean useJobTracker = true; <-- for hadoop 1.0

        //Step 0. Set configuration
```

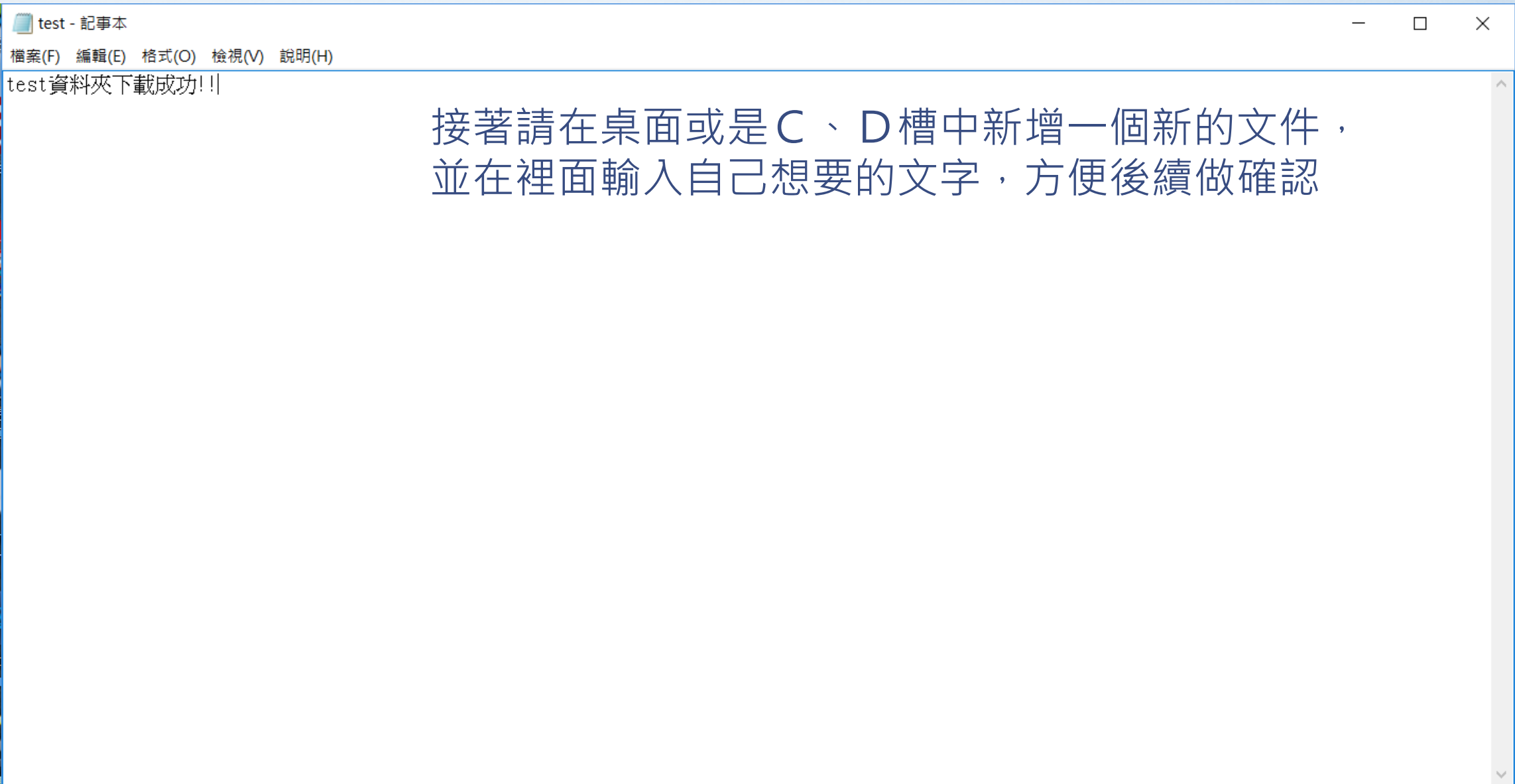
```
<已終止> presidentElection [Java 應用程式] C:\Program Files\Java\jre1.8.0_121\bin\javaw.exe (
    Total committed heap usage (bytes)=735576064

    Shuffle Errors
        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!



Map/Reduce - PresidentElection/src/org/windoop/presidentElection.java - Eclipse平台

檔案(F) 編輯(E) 程式碼(S) 重構(T) 導覽(N) 搜尋(A) 專案(P) 執行(R) 視窗(W) 說明(H)

專案瀏覽器

- DFS Locations
 - windoop
 - (2)
 - tmp (1)
 - user (1)
 - (3)
 - input (3)
 - output (2)
 - test (0)
 - PresidentElection
 - TDCS_GIDSequenc
 - WebPageDownloa

Download from DFS...
Create new directory...
Upload files to DFS...
Upload directory to DFS...
Refresh
Delete

回到eclipse，選取剛剛建立的資料夾，點選右鍵，選取Upload to DFS....

```
package org.windoop;

import java.io.BufferedWriter;
import java.io.OutputStreamWriter;
import java.util.Random;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.*;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.lib.input.*;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class presidentElection {

    public static void main(String[] args) throws Exception {
        String[] candidate = {"A", "B", "C"};
        String[] location = {"Taichung", "Taipei", "Kaohsiung"};
        Random r;
        boolean useResourceManager = true;
        //boolean useJobTracker = true; <-- for hadoop

        //Step 0. Set configuration
    }
}
```

Problems Tasks @ Javadoc Map/Reduce Locations 主控台

<已終止> presidentElection [Java 應用程式] C:\Program Files\Java\jre1.8.0_121\bin\java.exe

```
Total committed heap usage (bytes)=735576
Shuffle Errors
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!
```

已選取 1 個項目



<已終止> presidentElection [Java 應用程式] C:\Program Files\Java\jre1.8.0_121\bin\javaw.exe (2017年7月5日 上午9:40:17 直接存取)

```
Total committed heap usage (bytes)=735576064
Shuffle Errors
BAD_ID=0
CONNECTION=0
```

Map/Reduce - PresidentElection/src/org/windoop/presidentElection.java - Eclipse平台
檔案(F) 編輯(E) 程式碼(S) 重構(T) 導覽(N) 搜尋(A) 專案(P) 執行(R) 視窗(W) 說明(H)

專案瀏覽器

- DFS Locations
 - windoop
 - (2)
 - tmp (1)
 - user (1)
 - (3)
 - input (3)
 - output (2)
 - test (0)**
 - PresidentElection
 - TDCS_GIDSequence
 - WebPageDownload

Download from DFS...
Create new directory...
Upload files to DFS...
Upload directory to DFS...
Refresh
Delete

對剛剛建立的“test”資料夾按下 Refresh 重新整理“test”資料夾

```
package org.windoop;

import java.io.BufferedWriter;
import java.io.OutputStreamWriter;
import java.util.Random;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.*;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.lib.input.*;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class presidentElection {

    public static void main(String[] args) throws Exception {
        String[] candidate = {"A", "B", "C"};
        String[] location = {"Taichung", "Taipei", "Kaohsiung"};
        Random r;
        boolean useResourceManager = true;
        //boolean useJobTracker = true; <-- for hadoop 1.0

        //Step 0. Set configuration
```

Problems Tasks @ Javadoc Map/Reduce Locations 主控台

<已終止> presidentElection [Java 應用程式] C:\Program Files\Java\jre1.8.0_121\bin\javaw.exe (201

```
Total committed heap usage (bytes)=735576064
Shuffle Errors
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!
```

已選取 1 個項目

DFS Locations

windoop

(2)

tmp (1)

user (1)

(3)

input (3)

output (2)

test (1)

test.txt (20.0 b, r3)

PresidentElection

TDCS_GIDSequence_MRE

WebPageDownload_jdwang

上傳成功會顯示出剛
剛選的文件

presidentEle... PresidentEle... PresidentEle... h

package org.windoop;

```
import java.io.BufferedWriter;
import java.io.OutputStreamWriter;
import java.util.Random;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.*;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.lib.input;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input;
import org.apache.hadoop.mapreduce.lib.outp
```

public class presidentElection {

```
public static void main(String[] args) {
    String[] candidate = {"A", "B", "C"};
    String[] location = {"Taichung", "T
    Random r;
    boolean useResourceManager = true;
    //boolean useJobTracker = true; <--

    //Step 0. Set configuration
```

Problems Tasks Javadoc Map/Reduce Locations 主

<已終止> presidentElection [Java 應用程式] C:\Program Files\Java\j

```
Total committed heap usage (by
Shuffle Errors
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!



```
package org.windowoop;
```

```
import java.io.BufferedWriter;  
import java.io.OutputStreamWriter;  
import java.util.Random;  
import org.apache.hadoop.conf.Configuration;  
import org.apache.hadoop.fs.*;  
import org.apache.h...  
import org.apache.h...  
import org.apache.h...  
import org.apache.h...  
import org.apache.h...
```

```
public class presid
```

```
    public static v
```

```
        String[] ca
```

```
        String[] lo
```

```
        Random r;
```

```
        boolean use
```

```
        //boolean u
```

```
        //Step 0. S
```

problems Tasks @ Java

终止> presidentElection [Java

```
    Total
```

```
    Shuffle Errors
```

```
    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
```

Done!

Copy to local directory

Copy the selected files and directories from the distributed filesystem to a local directory

桌面
> OneDrive
> 本機
> 媒體櫃
> 網路
> 家用群組
>

資料夾(F): 桌面

建立新資料夾(M)

確定

取消

在此選取想要放
至下載資料夾的
位置

7月5日 上午9:40:17 直接存取)

檔案常用共用檢視

test

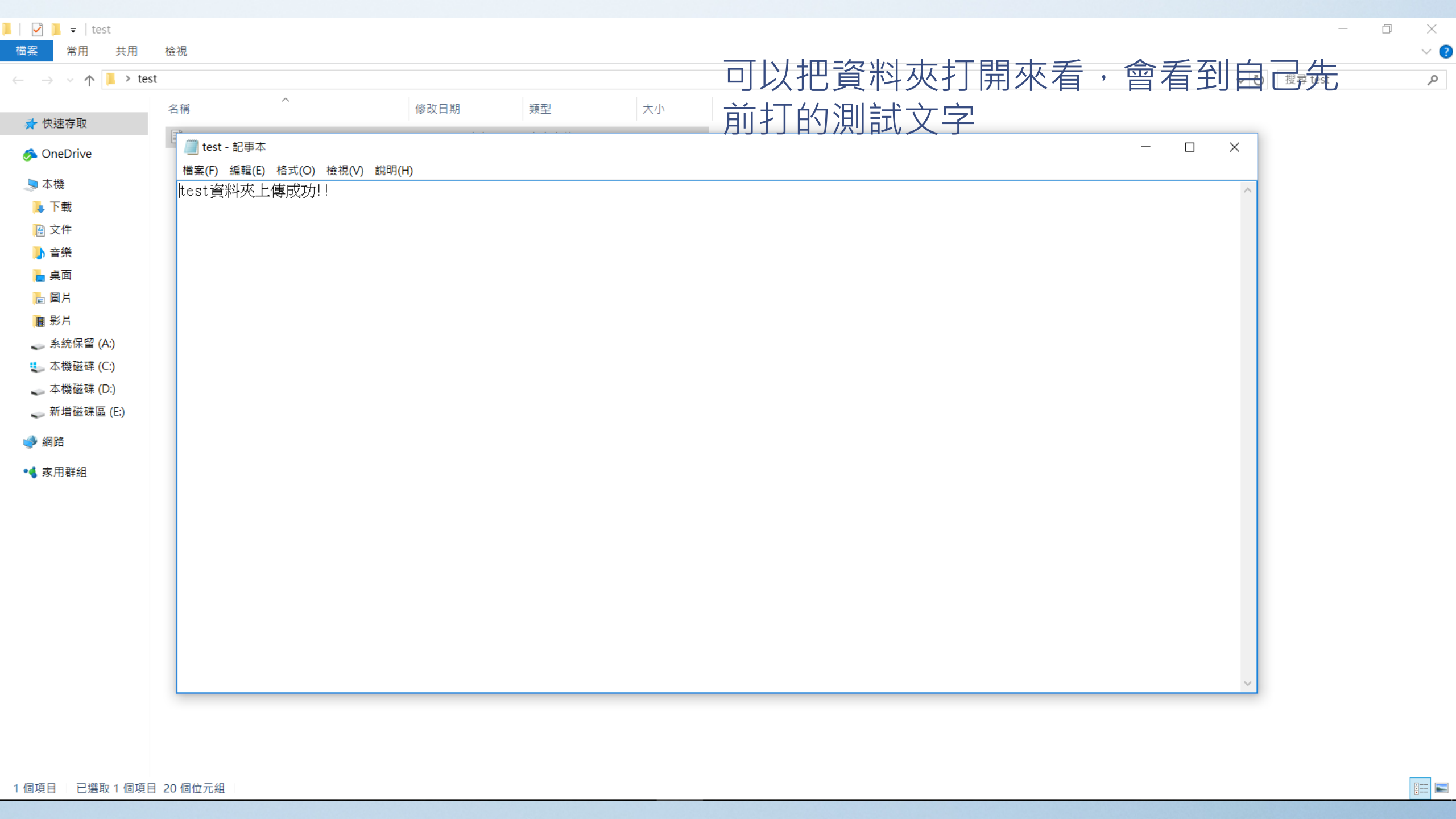
快速存取OneDrive本機下載文件音樂桌面圖片影片系統保留 (A:)本機磁碟 (C:)本機磁碟 (D:)新增磁碟區 (E:)網路家用群組

名稱	修改日期	類型	大小
test	2017/7/5 上午 10...	文字文件	1 KB

1 個項目

↑

下載成功後，請到剛剛選的位置，找尋自己下載的資料夾，點選進去可看到裡面有先前放的文件



可以把資料夾打開來看，會看到自己先前打的測試文字

test - 記事本

檔案(F) 編輯(E) 格式(O) 檢視(V) 說明(H)

test資料夾上傳成功!!