

Report

學號: F74012227

使用軟體: IntelliJ IDEA 2016.1.2

作業系統: Windows 7

檔案:

- 程式碼
 - ScalaProject
- 輸出結果
 - output/one
 - output/two

結果:

1.

- 題目:
統計頻率大於等於 K 的單詞
- 指令:

```
[root@cdc-course-05 ~]# spark-submit --class ProblemOne --master spark://cdc-course-05:10005 ScalaProject.jar hdfs://cdc-course-05:8020/spark/input/movies_small.dat hdfs://cdc-course-05:8020/spark/output/one 2
```

- 輸出:
見檔案
輸出格式 : (單詞, 頻率)

2.

- 題目:
給定 2 個文件 ratings.dat 和 movies.dat，通過這 2 個文件找出平均得分大於等於 4.0 的電影
- 指令:

```
[root@cdc-course-05 ~]# spark-submit --class ProblemTwo --master spark://cdc-course-05:10005 ScalaProject.jar hdfs://cdc-course-05:8020/spark/input/movies.dat hdfs://cdc-course-05:8020/spark/input/ratings.dat hdfs://cdc-course-05:8020/spark/output/two 4.0
```

- 輸出:

見檔案

輸出格式 : (電影識別符,(平均得分,電影名稱))