

SENG 480A: Summer 2017

Assignment 3 output

Yaxi Yu (Chloe)

1. Find the average rating for each user and movie

Files: MovielensAvgReducer.java

MovielensAvgMapper.java

MovielensAvgCombiner.java

Output sample:

```
C:\Users\Yaxi\Desktop\assign3_hadoop_spark\hadoop_examples_post>more output_movielens\part-r-00000
u1      3.8724696356276303
u10     3.4508196721311477
u100    3.4285714285714284
u100017 3.0
u100032 2.0
u100034 4.0
u100083 1.5
u100106 4.5
u100159 3.5
u100163 2.25
u100226 1.5
u100272 4.0
u1003    3.0
u100304 3.0
u100306 2.5
u100326 2.5
u100365 2.5
u100383 3.5
u100390 1.5
u1004    2.5
u100450 3.5
u100487 2.25
u100498 1.875
u1005    2.5
u100517 3.0
u100527 4.0
u100553 5.0
u100556 4.125
u100581 2.3333333333333335
u1006    3.0
u1007    2.727272727272727
u100714 3.8333333333333335
u100745 2.25
u1008    3.5
u100843 4.5
u1009    3.25
u101     3.8823529411764706
u1010   3.0294117647058822
-- Movie (98) --
```

2. Find the average rating for each genre

Files: MovielensGenreAvgReducer.java

MovielensGenreAvgMapper.java

MovielensGenreAvgReducer.java

MovielensGenreAvg.java (change directory)

Output sample:

```
C:\Users\Yaxi\Desktop\assign3_hadoop_spark\hadoop_examples_post>more output_movielens\part-r-00000
(no genres listed)      3.7777777777777777
Action 3.4456128030751034
Adventure 3.5203933324249443
Animation 3.636061588330632
Children 3.4661866359447004
Comedy 3.4460369221059275
Crime 3.679638509774991
Documentary 3.813299232736573
Drama 3.681779585269932
Fantasy 3.518063244815614
Film-Noir 3.9557017543859647
Horror 3.3152430044182624
IMAX 3.5711343472750317
Musical 3.5987928843710293
Mystery 3.6795409836065573
Romance 3.556164666942491
Sci-Fi 3.4604295476732836
Thriller 3.5185023771790807
War 3.8172139303482586
Western 3.5664225941422596
```

3. Repeat 1 using Spark

Files: SparkMovielensAvg.java

Output sample:

```
C:\Users\Yaxi\Desktop\assign3_hadoop_spark\SparkJava8Examples_post>more output_m
ovielens\part-00000
(m4427,4,27777777777777778)
(m1088,3,358490566037736)
(m25920,3,5)
(m5791,3,9615384615384617)
(m64614,3,3333333333333335)
(u539,4,0)
(m8477,4,25)
(m3516,3,0625)
(m1889,3,0)
(m33592,2,5)
(m133419,2,9166666666666665)
(m4280,3,75)
(m114601,2,75)
(m3656,5,0)
(m100106,4,5)
(u123,3,757575757575758)
(m26547,4,0)
(u481,4,137614678899083)
(m103810,3,0)
(m5785,3,3181818181818183)
(m3969,2,9565217391304346)
(m8917,3,425)
(m8641,3,514705882352941)
(m26302,2,5)
(m118696,3,1)
(m5483,3,6)
(m3181,3,8333333333333335)
(m4035,4,0)
-- More (0%) --
```

4. Repeat 2 using Spark
Files: SparkMovielensGenresAvg.java
Output sample:

```
C:\Users\Yaxi\Desktop\assign3_hadoop_spark\SparkJava8Examples_post>more output_
ovielens\part-00000
(War,3,8172139303482586)
(Fantasy,3,518063244815614)
(Western,3,5664225941422596)
(Musical,3,5987928843710293)
(Horror,3,3152430044182624)
(IMAX,3,5711343472750317)
(Crime,3,679638509774991)
(Animation,3,636061588330632)
(Children,3,4661866359447004)
(Sci-Fi,3,4604295476732836)
(Comedy,3,4460369221059275)
(Documentary,3,813299232736573)
(Mystery,3,6795409836065573)
(Thriller,3,5185023771790807)
((no genres listed),3,7777777777777777)
(Adventure,3,5203933324249443)
(Romance,3,556164666942491)
(Drama,3,681779585269932)
(Film-Noir,3,9557017543859647)
(Action,3,4456128030751034)
```