**Homework - 1**

Contents of the folder -

1) Input files (*movies.dat, ratings.dat, users.dat*)

2) Scala code files (*task1.scala, task2.scala*)

3) JAR files (*SaiSree\_Kamineni\_task1.jar, SaiSree\_Kamineni\_task2.jar*)

4) Output Files (*SaiSree\_Kamineni\_result\_task1.txt/ part-00000, SaiSree\_Kamineni\_result\_task2.txt/ part-00000*)

Note - I haven't set up SPARK\_HOME variable. So I run the scripts and commands from inside "*spark-1.6.3-bin-hadoop2.4*" directory.

Place this UnZipped folder in *spark-1.6.3-bin-hadoop2.4* directory

Steps to run the jar files -

Task - 1

*./bin/spark-submit --class "HW1Task1" --master local[1] ./SaiSree\_Kamineni\_HW1/SaiSree\_Kamineni\_task1.jar "./SaiSree\_Kamineni\_HW1/ratings.dat" "./SaiSree\_Kamineni\_HW1/users.dat"*

Output will be stored in current directory that is *spark-1.6.3-bin-hadoop2.4* with name *SaiSree\_Kamineni\_result\_task1.txt*. Inside this directory *part-00000* has the output for this task.

Task - 2

*./bin/spark-submit --class "HW1Task2" --master local[1] ./SaiSree\_Kamineni\_HW1/SaiSree\_Kamineni\_task2.jar "./SaiSree\_Kamineni\_HW1/ratings.dat" "./SaiSree\_Kamineni\_HW1/users.dat" "./SaiSree\_Kamineni\_HW1/movies.dat"*

Output will be stored in current directory that is *spark-1.6.3-bin-hadoop2.4* with name *SaiSree\_Kamineni\_result\_task2.txt*. Inside this directory *part-00000* has the output for this task.

Output format

For both the tasks I casted the average type to float (As mentioned in discussion board).