**ReadMe**

#Requirement:

* Python 3.\*
* Java Runtime Environment 1.8.\*
* Scala
* PySpark 2.1.\*

#Setup:

1. Install “findspark”, a helper module to load PySpark API if you use any IDE

>> **pip install findspark**

1. Copy this project folder “Movie-rec” into a convenient directory.
2. Copy the “movie-rec-data” folder into the user’s home directory (“~/”)
3. Run main\_rethink.py

NB: To run on the larger data set, replace the contents of “movie-rec-data” with the larger set