Spark 2.0

Uses SparkSession, Row Objects.

Horton Work included Spark 1 and Spark 2.

If you want to use Spark 2. You need to set the environment variable

Export SPARK\_MAJOR\_VERSION=2. This tells HortonWorks that you want to use Spark 2

Then spark-submit LowestRatedMovieDataFrame.py

To run with machine learning file,

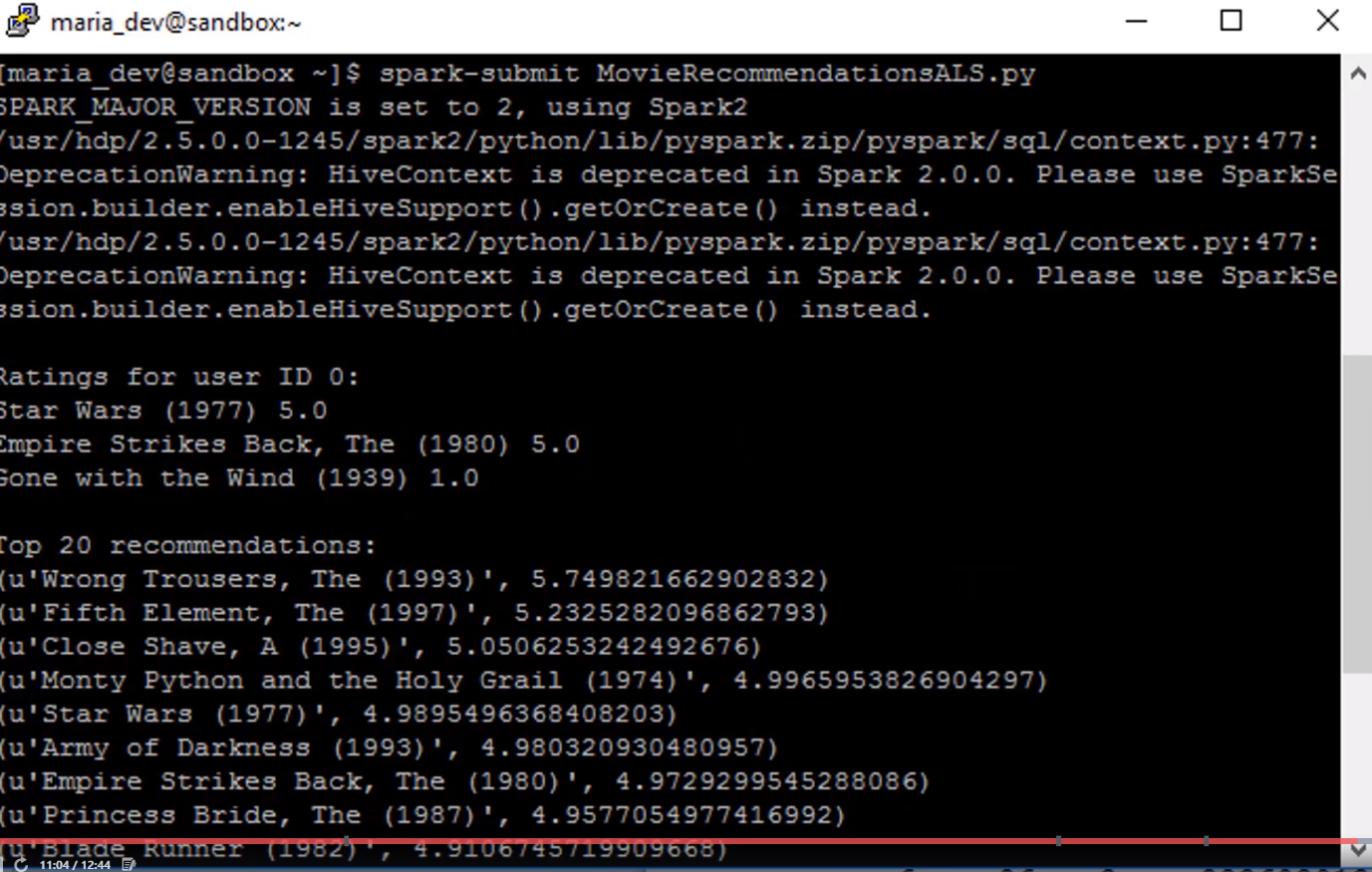
Install numpy – sudo pip install numpy==1.16

Upload 3 lines of data into u.data file. (see lecture)

Less <python file> to look at the contents of the python file

Export SPARK\_MAJOR\_VERSION = 2

And spark-submit MovieRecommendationsALS.py.

It has random element, so the result might not be the same