INF 553 FALL 2018 ASSIGNMENT 2

Name: Aditya Chavan
USC ID: 3416409224
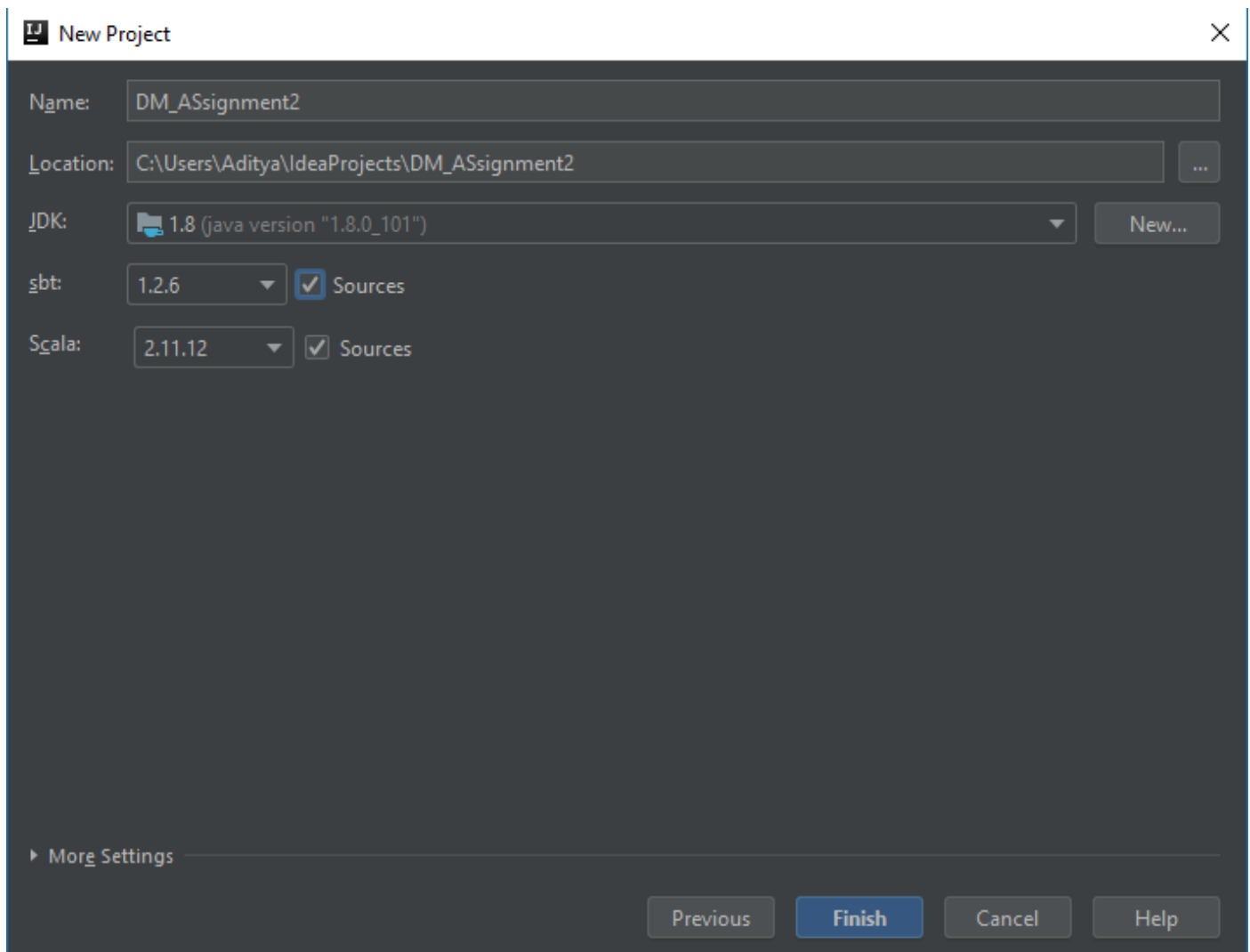
I am Running the Scala Code, using the following IDE:
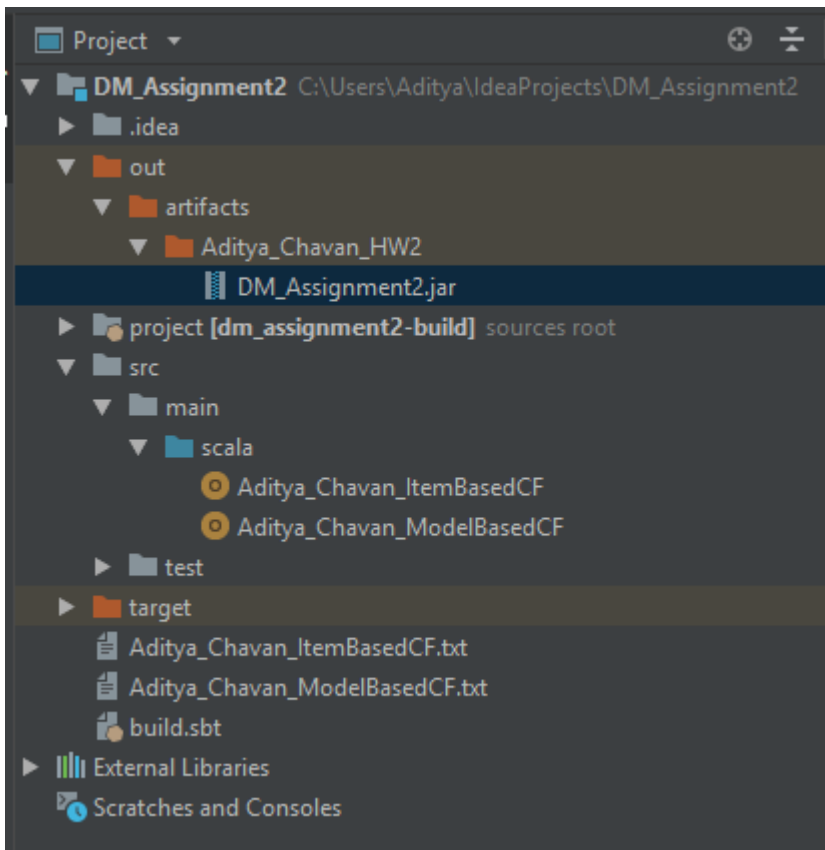
intellij IDEA

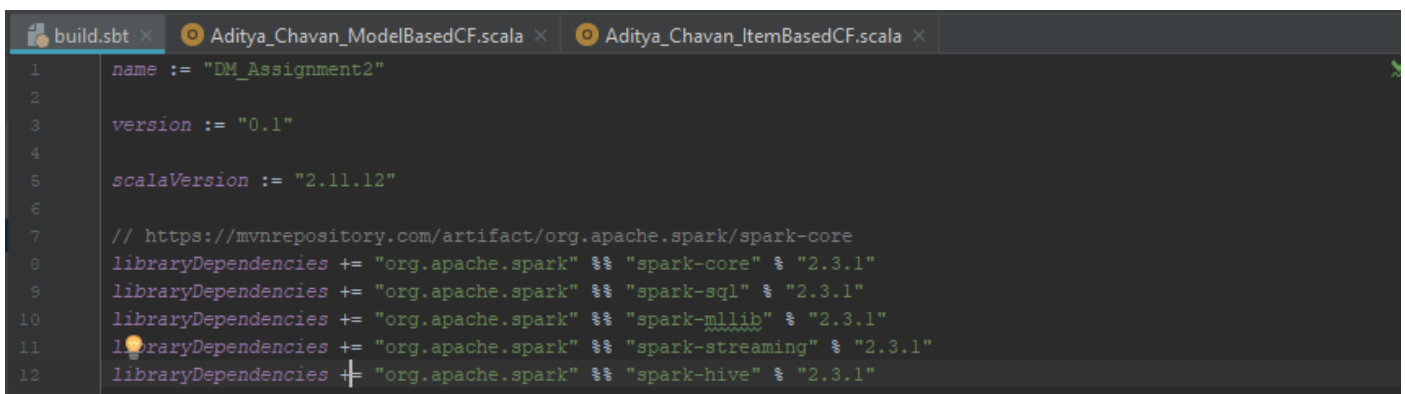**As per instructions, the environment was set as follows:**

1) Java – 1.8.0_181
2) sbt – 1.2.3
3) Hadoop – 2.7
4) Scala – 2.11.0
5) Spark – 2.3.1

The project was made in intelliJ as given:

The build.sbt File looks as follows:


```
build.sbt ×    Aditya_Chavan_ModelBasedCF.scala ×    Aditya_Chavan_ItemBasedCF.scala ×
1      name := "DM_Assignment2"
2
3      version := "0.1"
4
5      scalaVersion := "2.11.12"
6
7      // https://mvnrepository.com/artifact/org.apache.spark/spark-core
8      libraryDependencies += "org.apache.spark" %% "spark-core" % "2.3.1"
9      libraryDependencies += "org.apache.spark" %% "spark-sql" % "2.3.1"
10     libraryDependencies += "org.apache.spark" %% "spark-mllib" % "2.3.1"
11     libraryDependencies += "org.apache.spark" %% "spark-streaming" % "2.3.1"
12     libraryDependencies += "org.apache.spark" %% "spark-hive" % "2.3.1"
```

Steps to Run the code using command line

1. Change to directory from the folder that contains *Aditya_Chavan_hw2.jar*

2. **To run Task1(Model Based CF), execute the command:**
$SPARK_HOME/bin/spark-submit --class ModelBasedCF Aditya_Chavan_hw2.jar <training_file_path> <testing_file_path>

```
apples-MacBook-Pro-2:Mahima_Gupta_hw2 Mahima$ /Users/mahima/Desktop/spark-2.3.1-bin-hadoop2.7/bin/spark-submit --class ModelBasedCF Aditya_Chavan_hw2.jar /Users/mahima/De
sktop/train_review.csv /Users/mahima/Desktop/test_review.csv
```

3. **To run Task2 (Item Based CF), execute the command:**
$SPARK_HOME/bin/spark-submit --class ItemBasedCF Aditya_Chavan_hw2.jar <training_file_path> <testing_file_path>

```
apples-MacBook-Pro-2:Mahima_Gupta_hw2 Mahima$ /Users/mahima/Desktop/spark-2.3.1-bin-hadoop2.7/bin/spark-submit --class ItemBasedCF Aditya_Chavan_hw2.jar /Users/mahima/De
ktop/train_review.csv /Users/mahima/Desktop/test_review.csv
```

Collaboration Filtering Speed & Accuracy

Ranges	Task 1 (ModelBasedCF)	Task 2 (ItemBasedCF)
>=0 and <1	35797	26037
>=1 and <2	5790	16132
>=2 and <3	2070	2700
>=3 and <4	985	346
>=4	594	21
RMSE	**1.06869268803**	**1.09802373279**
Time	29 sec	197 sec

Task 1

```
>=0 and <1: 35797
>=1 and <2: 5790
>=2 and <3: 2070
>=3 and <4: 985
>=4: 594
RMSE: 1.0686926880305345
Time: 29 sec.
```

Task 2

```
>=0 and <1: 26037
>=1 and <2: 16132
>=2 and <3: 2700
>=3 and <4: 346
>=4: 21
RMSE: 1.0980203732797909
Time: 197 sec.
```