

INF 553 FALL 2018 ASSIGNMENT 1

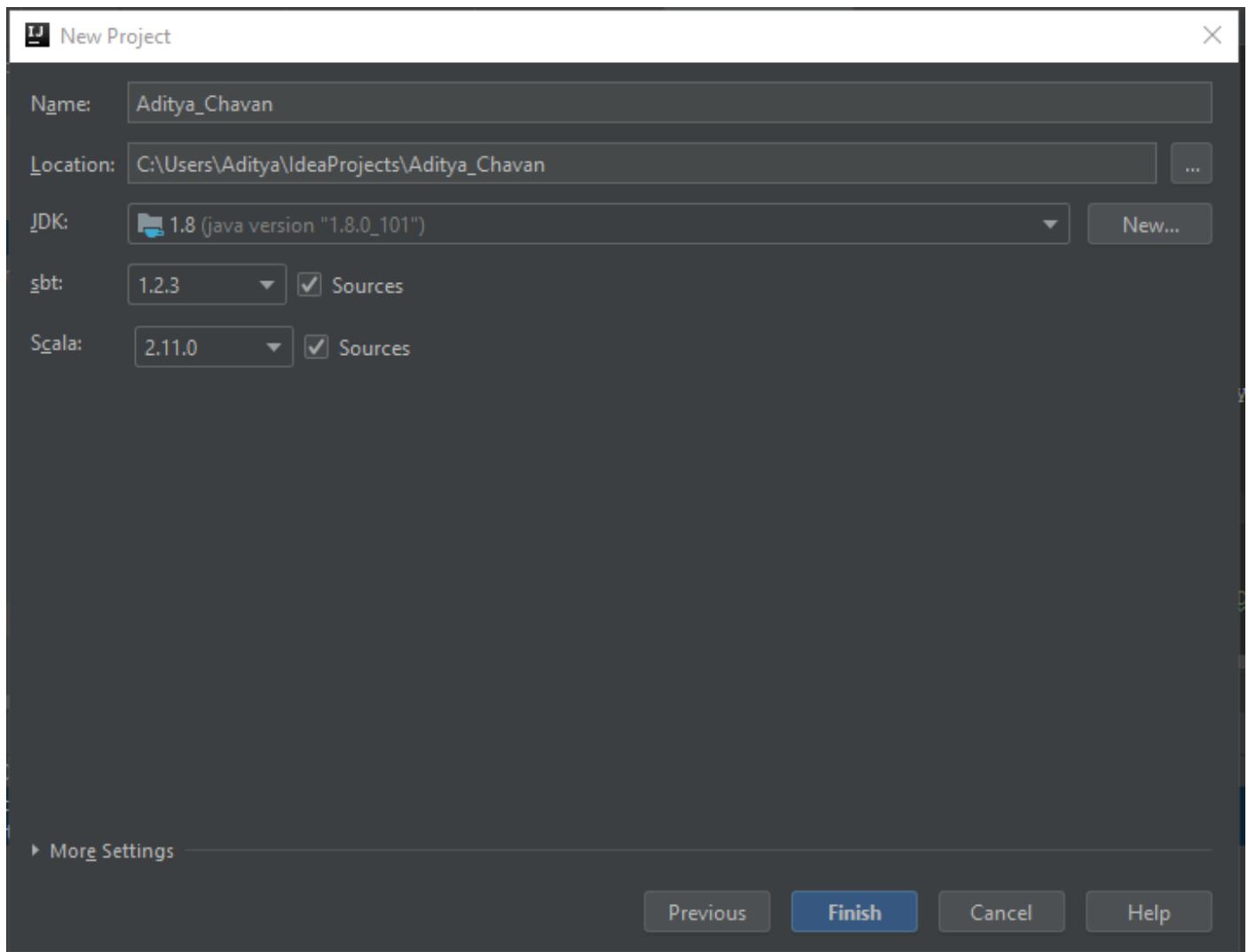
Name: Aditya Chavan
USC ID: 3416409224

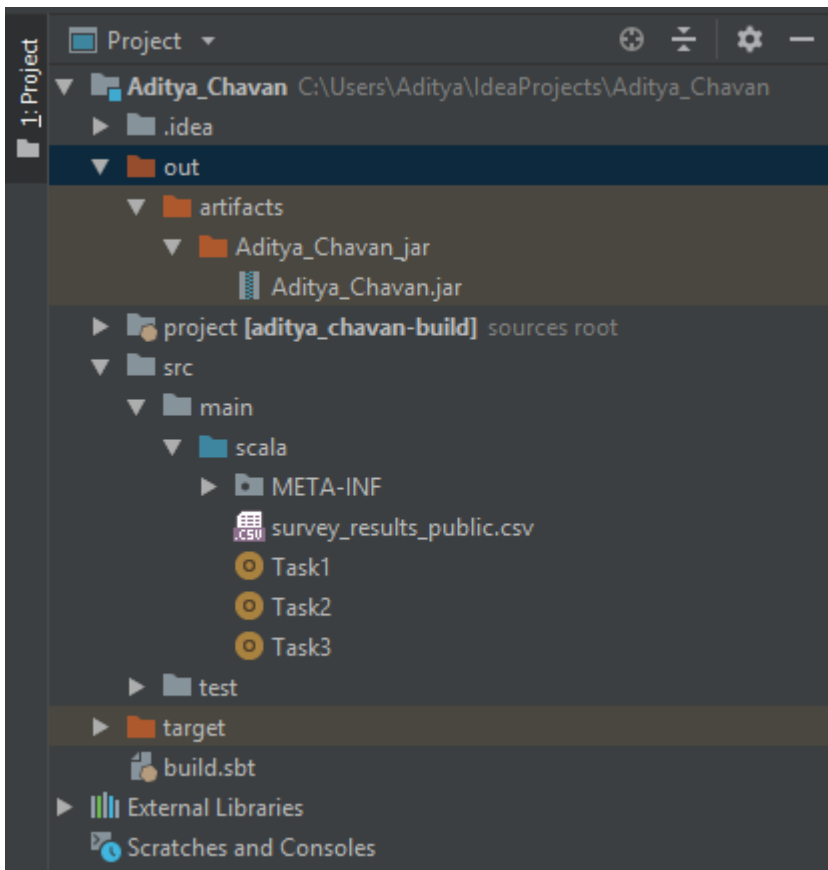
I am Running the Scala Code, using the following IDE:
intelliJ IDEA

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:

```

1  name := "Aditya_Chavan"
2
3  version := "0.1"
4
5  scalaVersion := "2.11.0"
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.jar*

2. To run Task1, execute the command:

```
$SPARK_HOME/bin/spark-submit --class Task1 Aditya_Chavan.jar
 <input_path> <output_path>
```

```

apples-MacBook-Pro-2:Aditya_Chavan Mahima$ /Users/mahima/Desktop/spark-2.3.1-bin-hadoop2.7/bin/spark-submit --class Task1 Aditya_Chavan.jar /Users/mahima/Desktop/stack-overflow-2018-developer-survey/survey_results_public.csv /Users/mahima/Desktop/Results.csv

```

```

"/Users/mahima/Desktop/spark-2.3.1-bin-hadoop2.7/bin/spark-submit --class Task1 Aditya_Chavan.jar /Users/mahima/Desktop/stack-overflow-2018-developer-survey/survey_results_public.csv /Users/mahima/Desktop/Results.csv"

```

3. To run Task2, execute the command:

```
$SPARK_HOME/bin/spark-submit --class Task2 Aditya_Chavan.jar  
<input_path> <output_path>
```

```
apples-MacBook-Pro-2:Aditya_Chavan Mahima$ /Users/mahima/Desktop/spark-2.3.1-bin-hadoop2.7/bin/spark-submit --class Task2 Aditya_Chavan.jar /Users/mahima/Desktop/stack-overflow-2018-developer-survey/survey_results_public.csv /Users/aditya/Desktop/Results.csv
```

```
"/Users/mahima/Desktop/spark-2.3.1-bin-hadoop2.7/bin/spark-submit --  
class Task2 Aditya_Chavan.jar /Users/mahima/Desktop/stack-overflow-2018-  
developer-survey/survey_results_public.csv  
/Users/aditya/Desktop/Results.csv"
```

4. To run Task3, execute the command:

```
$SPARK_HOME/bin/spark-submit --class Task3 Aditya_Chavan.jar  
<input_path> <output_path>
```

```
apples-MacBook-Pro-2:Aditya_Chavan Mahima$ /Users/mahima/Desktop/spark-2.3.1-bin-hadoop2.7/bin/spark-submit --class Task3 Aditya_Chavan.jar /Users/mahima/Desktop/stack-overflow-2018-developer-survey/survey_results_public.csv /Users/aditya/Desktop/Results.csv
```

```
"/Users/mahima/Desktop/spark-2.3.1-bin-hadoop2.7/bin/spark-submit --  
class Task3 Aditya_Chavan.jar /Users/mahima/Desktop/stack-overflow-2018-  
developer-survey/survey_results_public.csv  
/Users/aditya/Desktop/Results.csv"
```