**INF 553 FALL 2018 ASSIGNMENT 1**

Name: Aditya Chavan

USC ID: 3416409224

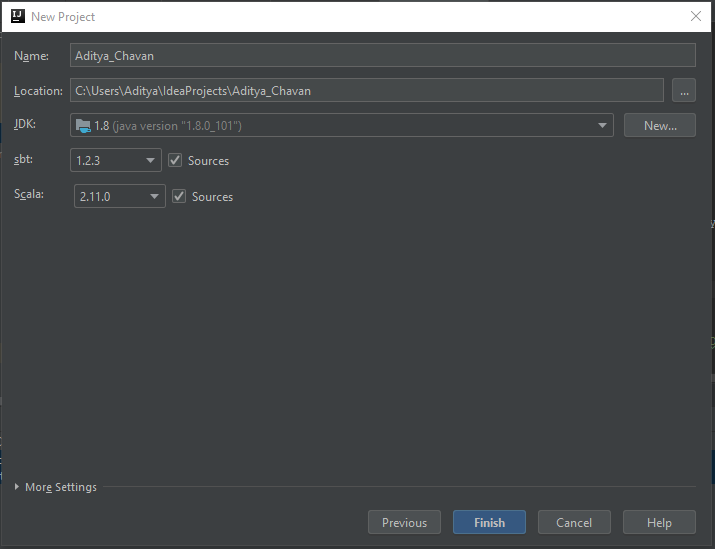
**I am Running the Scala Code, using the following IDE:**

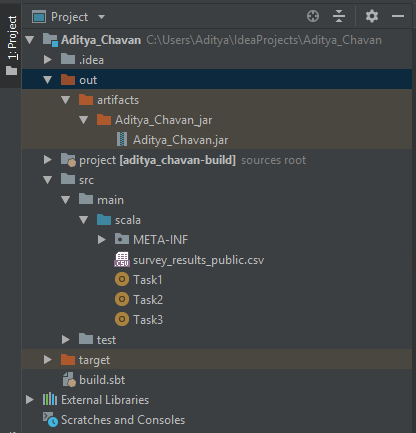
intellij IDEA

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

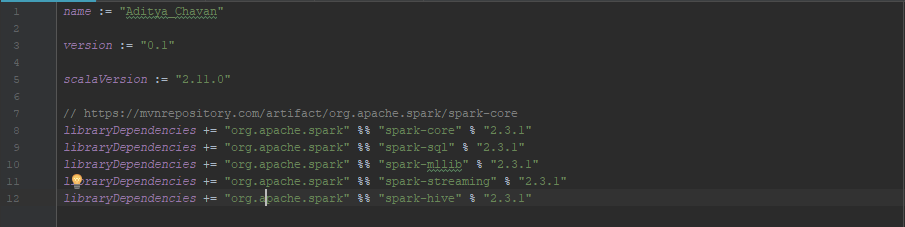
1. Java – 1.8.0\_181
2. sbt – 1.2.3
3. Hadoop – 2.7
4. Scala – 2.11.0
5. Spark – 2.3.1

**The project was made in intelliJ as given:**





**The build.sbt File looks as follows:**



**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>



“/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”

1. **To run Task2, execute the command:**

$SPARK\_HOME/bin/spark-submit --class Task2 Aditya\_Chavan.jar <input\_path> <output\_path>



“/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”

1. **To run Task3, execute the command:**

$SPARK\_HOME/bin/spark-submit --class Task3 Aditya\_Chavan.jar <input\_path> <output\_path>



“/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”