**Instructions of Setting up and Running Spark Project**

**Preparing the Environment**

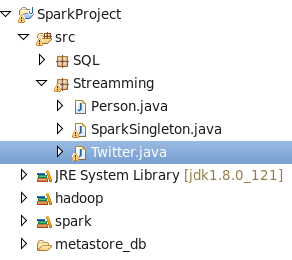
**Step 1:** Unzip “SparkProjectFiles.zip” file. Import the project “SparkProject” in Eclipse.

**Step 2:** Configure the build path by adding the entire required libraries mentioned below.

* Spark Core, Spark SQL, Spark Stream
* Java JDK 8, JRE
* Hive

**Running Project Part 1 – Spark Stream**

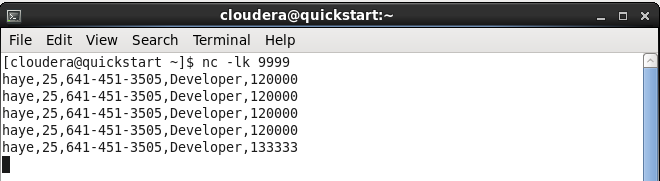
**Step 1:** Run the Twitter.java file as *Java application* from the project that is under package “Streamming”.



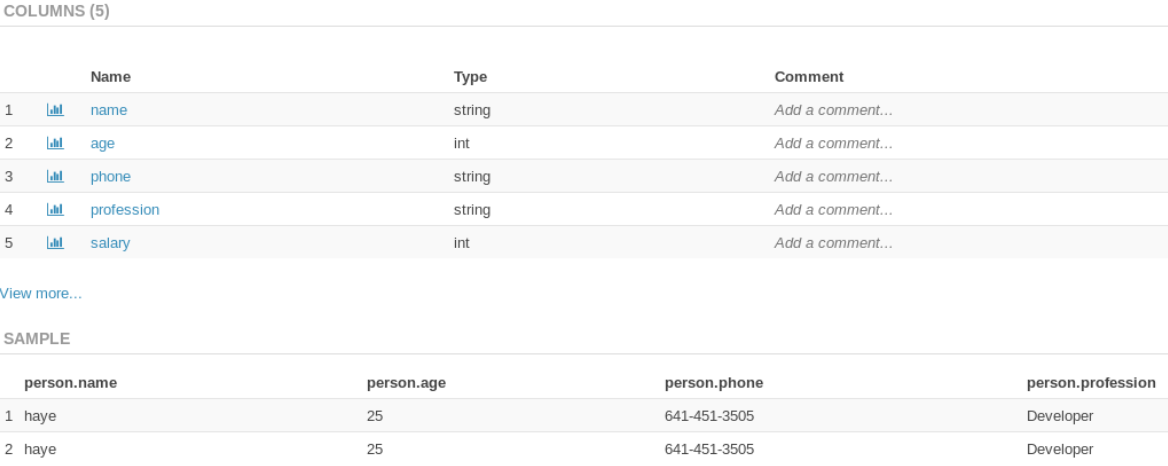
**Step 2:** Run the server using shell by typing following command.



**Step 3:** Write sample data in the shell in the format of *name, age, phone number, job, salary*.

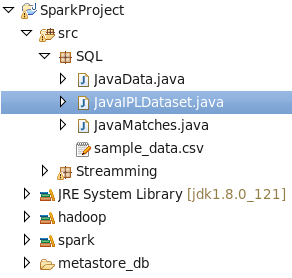


**Step 4:** To view the saved data go to Hive using data exploration tool like Hue. The data will be saved as shown below.

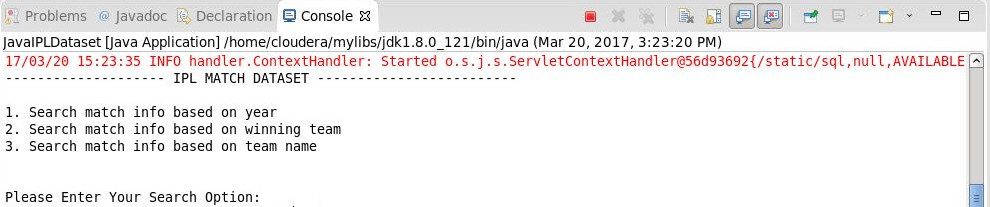


**Running Project Part 2 – Spark SQL and Dataset**

**Step 1:** From project, run the file name “JavaIPLDataset.java” as *Java Application* that is under package named “SQL”.

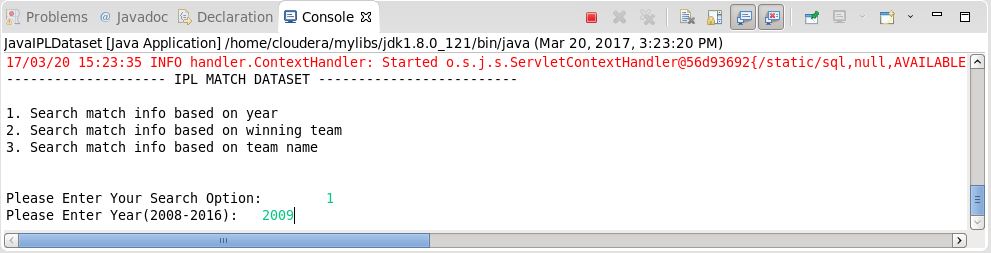


**Step 2:** You will be prompted to enter the filter option as shown in image below.



**Step 3:** Enter the required option in number along with the *parameter* to filter the data further. Below shows all the three options selection condition, with picture.

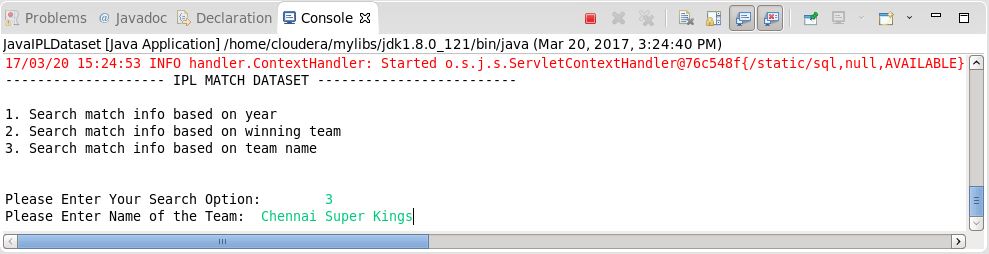
* If option 1 is selected, enter the year between 2008 and 2016 to filter the records further.



* If option 2 is chosen, enter the name of the winning team to filter the data further.

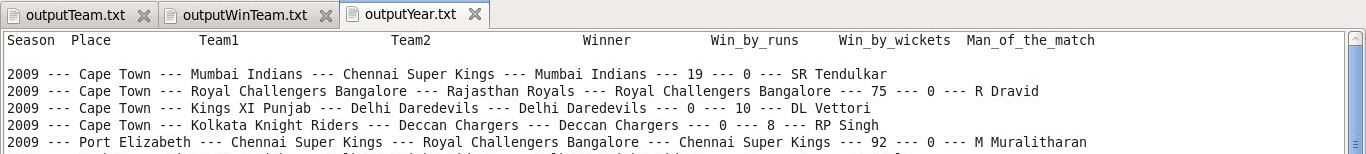


* If option 3 is opted, enter the team name to filter the data further.

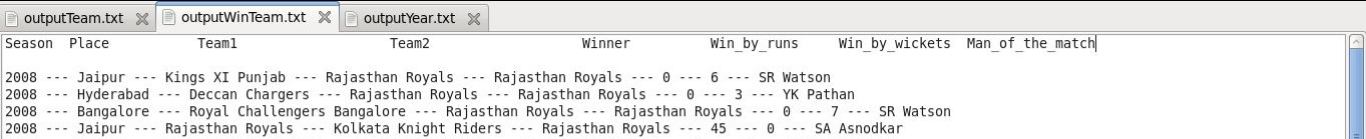


**Step 4:** To view the output containing records filtered according to the user input see the following information.

* For output of filter option 1 open the file named outputYear.txt that is generated inside the project in Eclipse IDE.



* For output of filter option 1 open the file named outputWinTeam.txt that is generated inside the project in Eclipse IDE.



* For output of filter option 1 open the file named outputTeam.txt that is generated inside the project in Eclipse IDE.

