Table of Contents

[Project: Part 1 – Spark Stream 1](#_Toc477808432)

[Project: Part 2 – Spark SQL and Datasets 3](#_Toc477808433)

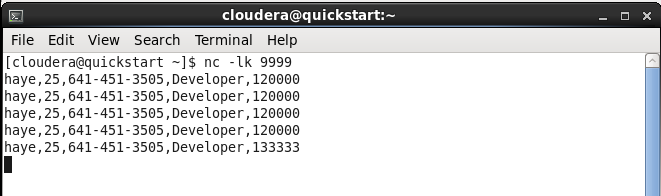
# Project: Part 1 – Spark Stream

**About Project**

The project works on Spark Stream technology that reads the streaming data from the UNIX Netcat channel and stores it inside Hive table. The data can be further accessed from the Hive and analyzed further.

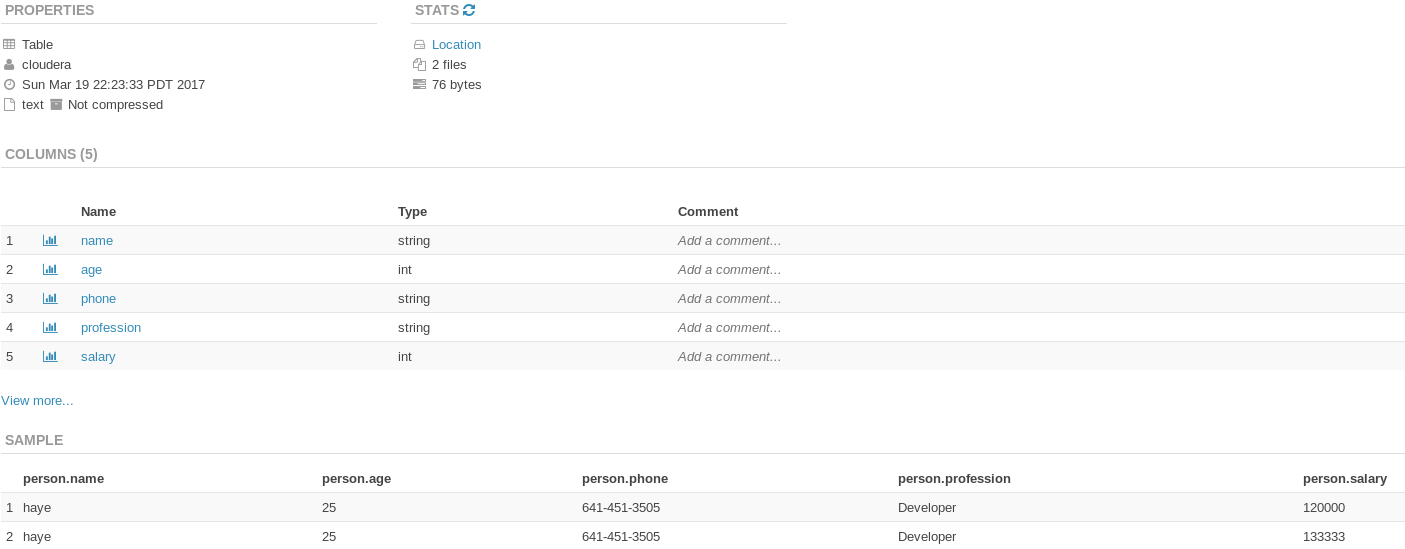
**Input**

Following image depicts how we can run the UNIX Netcat server with the command “*nc –lk 9999*” in shell. After the server is running the data that are streamed into it will be saved into Hive. The image shows the sample data that is in the structure of *name, age, phone number, job and salary*. This structure should be maintained in order to save it in Hive table, or else it cannot read and won’t save the data.



**Output**

The below image depicts how the data streamed into the server is saved to the Hive table by the project system. It also shows the table schema that is present in Hive table.



# Project: Part 2 – Spark SQL and Datasets

**About Project**

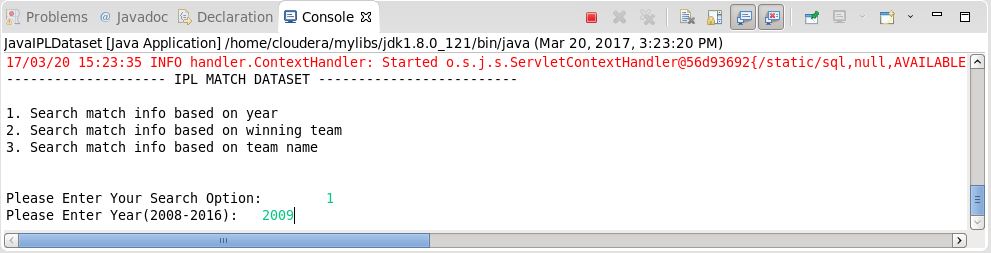
The project provides the service of showing IPL Cricket Match information based on the request given by users. The project loads the csv format big data information of IPL cricket matches from 2008 to 2016. The user can filter the records based on three parameters.

1. Year: displays all IPL matches played on that year.
2. Winning Team: displays all match information won by that team so far.
3. Team Name: displays all matches the team has played so far.

The system uses Spark SQL and Datasets to load the big data and filters it based on user request and generates the output in a text file. If filtered by year *outputYear.txt* file is generated. For other filter option *outputWinTeam.txt* and *outputTeam.txt* file is generated, respectively.

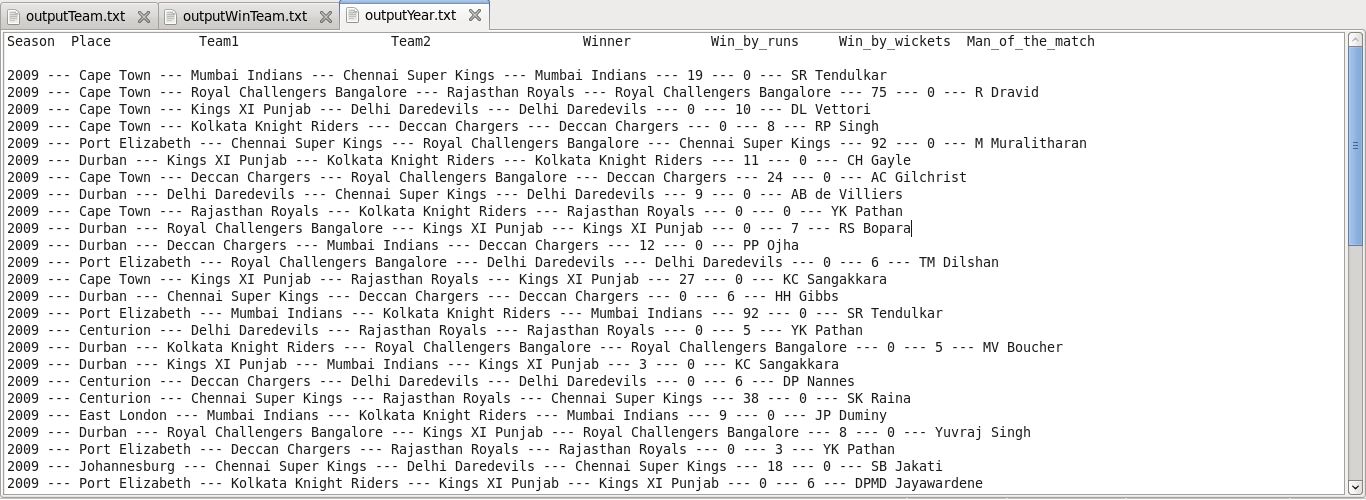
**Input – option 1**

The image shows the condition where search option 1 is selected and 2009 is entered as parameter to filter the records.



**Output – option 1**

The output shows information about all the matches played on that year. The information contains data like season, place where match is played, team names, winner team name, win by runs and win by wicket information and man of the match.



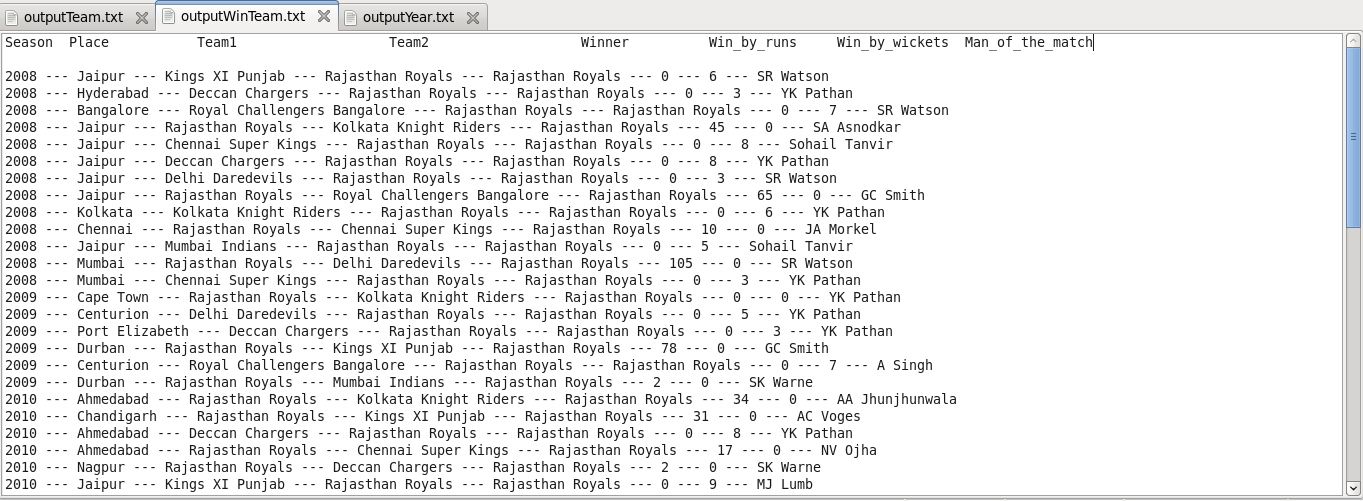
**Input – option 2**

The picture depicts the condition where filter option 2 was selected and winner team name “Rajasthan Royals” is given to filter the records.



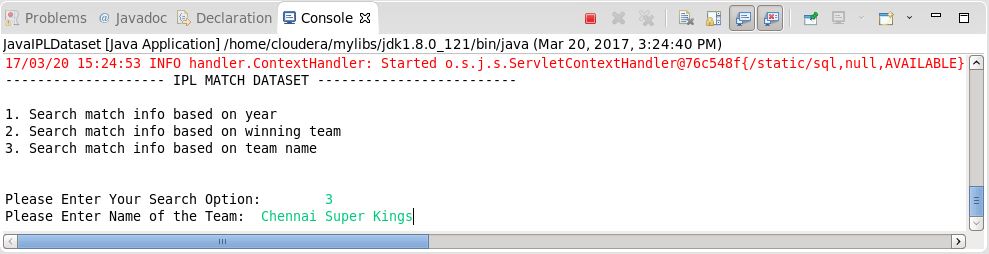
**Output – option 2**

The output displays information about all the matches till 2016 in which “Rajasthan Royals” has won.



**Input – option 3**

The picture shows the condition where filter option 3 was selected and “Chennai Super Kings” is entered as filter parameter.



**Output – option 3**

The output shows information about all the matches played by “Chennai Super Kings” in the IPL, whether the team won or lose, till 2016.

