

Advanced Big Data Analytics

Course Code: UE20CS581

Submitted By: Nallani Mounika

Submitted By: Shilpa Hegade

Professor: Dr. K.V Subramaniam

SRN: PES1PG20CS023

SRN: PES1PG20052

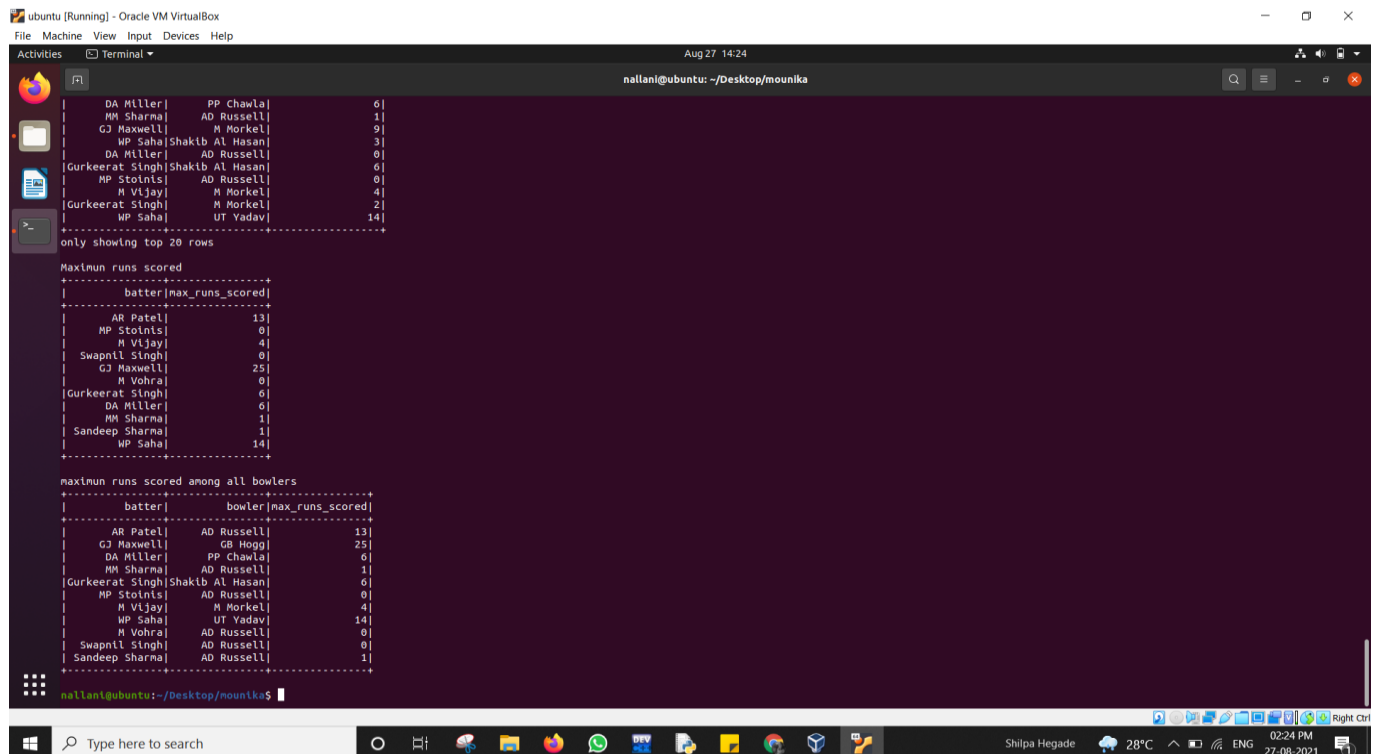
Assignment: IPL Analysis using Spark

Github Link: <https://github.com/MOUNIKA4567/Big-Data-IPL-DATA-analysis>

Learnings:

1. Installation of Apache Spark.
2. Working of map reduce program in spark .
3. Execution of Pyspark using sample file (<https://medium.com/expedia-group-tech/working-with-json-in-apache-spark-1ecf553c2a8c>)
4. How to write python program for large IPL data set querying.

Output: <https://github.com/MOUNIKA4567/Big-Data-IPL-DATA-analysis>



```
ubuntu [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Aug 27 14:24
nallani@ubuntu: ~/Desktop/mounika

DA Miller| PP Chawla| 6|
MM Sharma| AD Russell| 1|
GJ Maxwell| M Morkel| 9|
MP Sahaj| Shakib Al Hasan| 3|
DA Miller| AD Russell| 0|
Gurkeerat Singh| Shakib Al Hasan| 0|
MP Stoinis| AD Russell| 0|
M Vijay| M Morkel| 4|
Gurkeerat Singh| M Morkel| 2|
MP Sahaj| UT Yadav| 14|
-----
only showing top 20 rows

Maximum runs scored
-----
| batter|max_runs_scored|
-----
| AR Patel| 13|
| MP Stoinis| 0|
| M Vijay| 4|
| Swapnil Singh| 0|
| GJ Maxwell| 25|
| M Vohra| 0|
| Gurkeerat Singh| 0|
| DA Miller| 6|
| MM Sharma| 1|
| Sandeep Sharma| 1|
| MP Sahaj| 14|
-----

Maximum runs scored among all bowlers
-----
| batter| bowler|max_runs_scored|
-----
| AR Patel| AD Russell| 13|
| GJ Maxwell| GB Hogg| 25|
| DA Miller| PP Chawla| 6|
| MM Sharma| AD Russell| 1|
| Gurkeerat Singh| Shakib Al Hasan| 0|
| MP Stoinis| AD Russell| 0|
| M Vijay| M Morkel| 4|
| MP Sahaj| UT Yadav| 14|
| M Vohra| AD Russell| 0|
| Swapnil Singh| AD Russell| 0|
| Sandeep Sharma| AD Russell| 1|
-----

nallani@ubuntu: ~/Desktop/mounika$
```