Name – Spursh Ujjawal

SJSUID - 010471071

Ques – 22

Link - http://jmcauley.ucsd.edu/data/amazon/links.html

From the Dashboard select the Data & Analytics.

Click on Analytics and create New Instance and New Notebook.

I'm using Scala for my Notebook.

Drag and drop the data file to Data Source.

Run the following queries:

- import org.apache.spark.sql.functions._
- val sqlContext = new org.apache.spark.sql.SQLContext(sc)
- val df = sqlContext.read.json("swift://notebooks.spark/reviewsMusic.json.gz2")
- df.sort(desc("overall")).show(10)

My project aims at selecting the top 10 Music Records from the given Amazon data set.
The data set is provided by the owner by dropping him a mail.
Login to Bluemix and create a new instance in Data and Analytics. Create a new notebook in Scala, as I'm using Scala for my project. Write the above queries. Upload the required data in the data source. Run the above queries and we get the top 10 Music records from the data set.