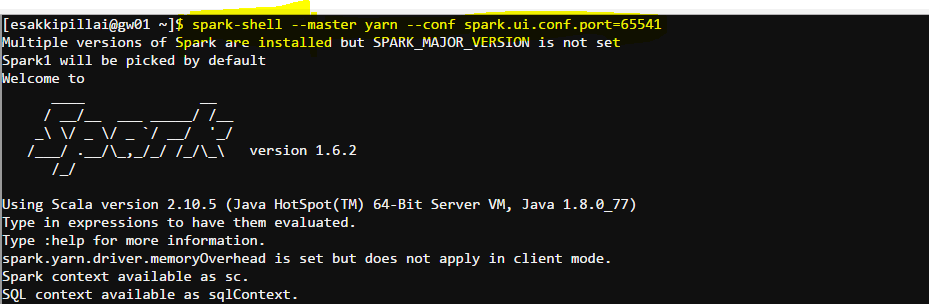
Spark SQL

* SPARK SQL is the Component on top of the spark core
* SPARKSQL introduces a new data abstraction called Schema RDD which provides support for the structured and Semi-structured data
* SPARKSQL can also be used to execute SQL queries written using either a basic SQL OR HIVEQL
* SPARKSQL can also used to read the data from the existing hive installation
* Structured data is the one which has the schema – that is the known set of fields for each record , when we have these type of data SPARKSQL makes it easier to load and make efficient to query
* SPARKSQL Provides a Programming abstraction called data frame and act as distributed SQL Query Engine.

To open the Spark Shell



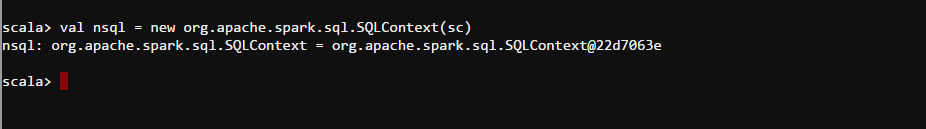
Creating the SQL Context

Val sc = sparkCommon.SparkContext

Val sqlcontext = new org.apache.spark.sqlContext(sc)

In Lab default sqlContext is HiveContext

So creating a new SQLCOntext



To Create Multiple Directories in HDFS as a Stretch



