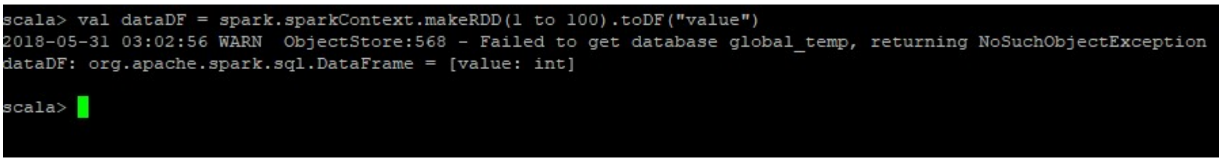
# Session 19 Assignment 3

Create a dataframe from 1 to 100 and write it to parquet file.

First, create a dataframe from 1 to 100

val dataDF = spark.sparkContext.makeRDD(1 to 100).toDF("value")

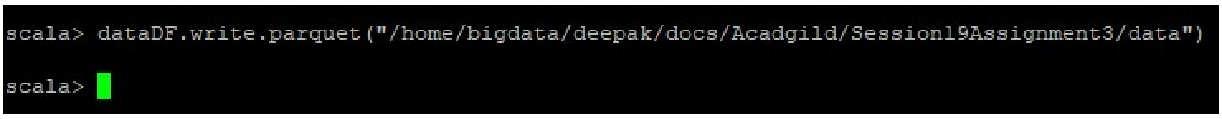


```
scala> val dataDF = spark.sparkContext.makeRDD(1 to 100).toDF("value")
2018-05-31 03:02:56 WARN  ObjectStore:568 - Failed to get database global_temp, returning NoSuchObjectException
dataDF: org.apache.spark.sql.DataFrame = [value: int]

scala>
```

Now, write it to parquet file

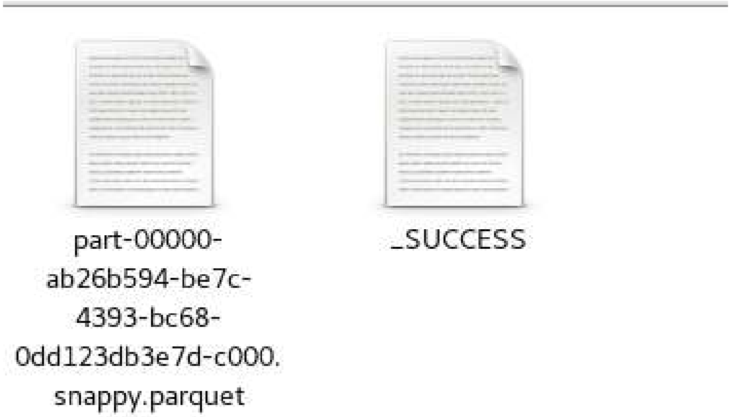
dataDF.write.parquet("/home/bigdata/deepak/docs/Acadgild/Session19Assignment3/data")



```
scala> dataDF.write.parquet("/home/bigdata/deepak/docs/Acadgild/Session19Assignment3/data")

scala>
```

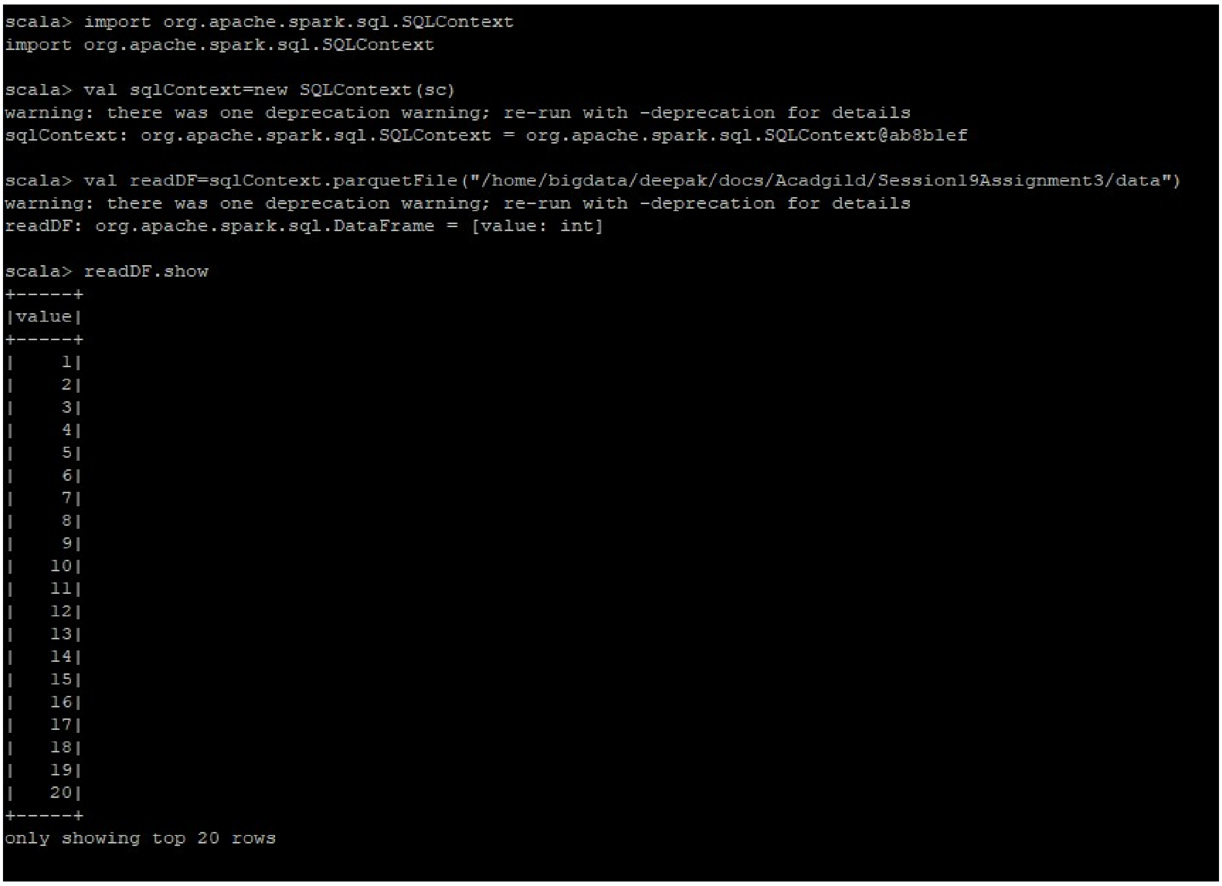
Once written, we are getting the following files



part-00000-ab26b594-be7c-4393-bc68-0dd123db3e7d-c000.snappy.parquet

_SUCCESS

Now, validate the parquet file by reading the contents back to dataframe



```
scala> import org.apache.spark.sql.SQLContext
import org.apache.spark.sql.SQLContext

scala> val sqlContext=new SQLContext(sc)
warning: there was one deprecation warning; re-run with -deprecation for details
sqlContext: org.apache.spark.sql.SQLContext = org.apache.spark.sql.SQLContext@ab8blef

scala> val readDF=sqlContext.parquetFile("/home/bigdata/deepak/docs/Acadgild/Session19Assignment3/data")
warning: there was one deprecation warning; re-run with -deprecation for details
readDF: org.apache.spark.sql.DataFrame = [value: int]

scala> readDF.show
+-----+
|value|
+-----+
|    1|
|    2|
|    3|
|    4|
|    5|
|    6|
|    7|
|    8|
|    9|
|   10|
|   11|
|   12|
|   13|
|   14|
|   15|
|   16|
|   17|
|   18|
|   19|
|   20|
+-----+
only showing top 20 rows
```