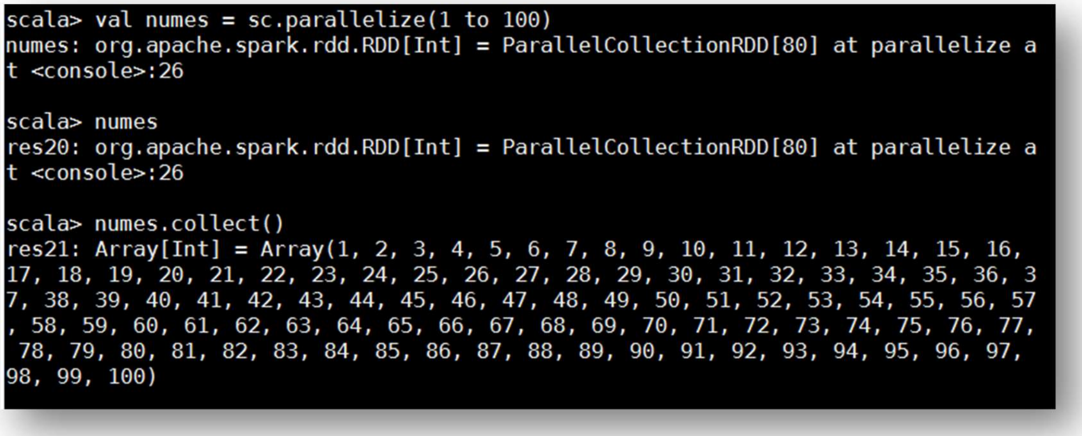
# Assignment 19.3

Create a dataframe with 1 to 100 and save as parquet file.

1. Created a RDD for numbers between 1 to 100

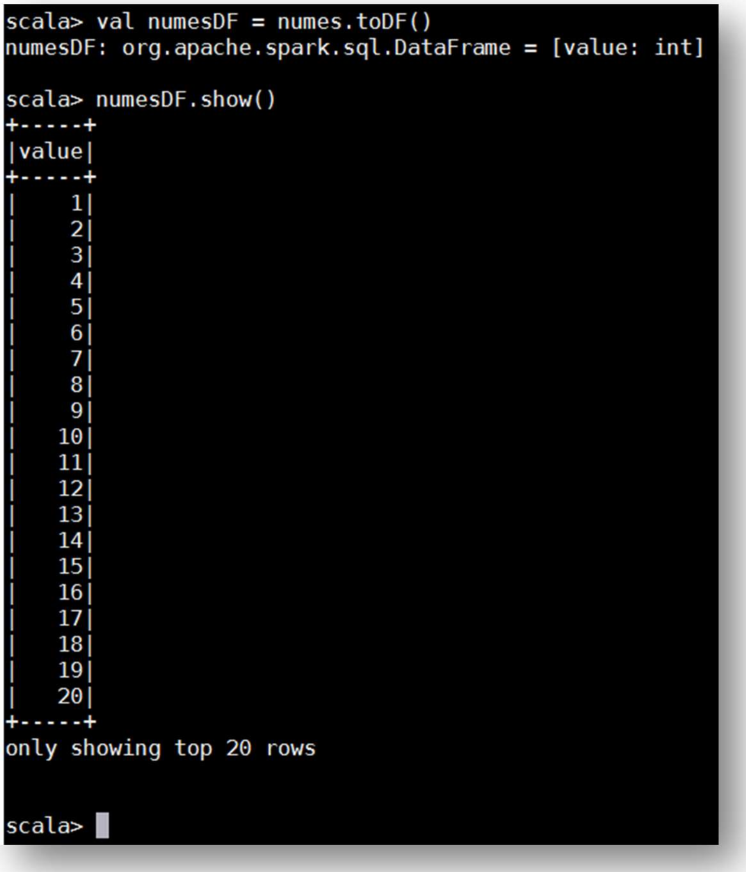
**val numes = sc.parallelize(1 to 100)**



```
scala> val numes = sc.parallelize(1 to 100)
numes: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[80] at parallelize a
t <console>:26

scala> numes
res20: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[80] at parallelize a
t <console>:26

scala> numes.collect()
res21: Array[Int] = Array(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,
17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 3
7, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57
, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77,
 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97,
98, 99, 100)
```

Now we will create a Dataframe for the above RDD

**val numesDF = numes.toDF()**



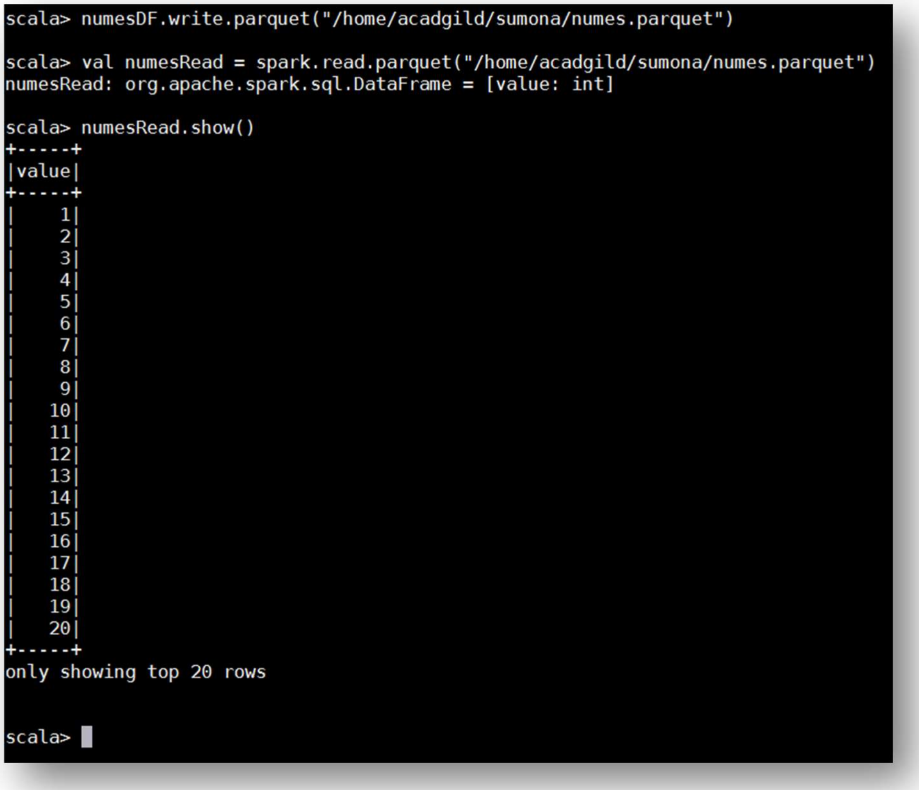
```
scala> val numesDF = numes.toDF()
numesDF: org.apache.spark.sql.DataFrame = [value: int]

scala> numesDF.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


scala> 
```

Now, we will save the dataframe as parquet file and then read it.

**numesDF.write.parquet("/home/acadgild/sumona/numes.parquet")**

**val numesRead = spark.read.parquet("/home/acadgild/sumona/numes.parquet")**

**numesRead.show()**



```
scala> numesDF.write.parquet("/home/acadgild/sumona/numes.parquet")

scala> val numesRead = spark.read.parquet("/home/acadgild/sumona/numes.parquet")
numesRead: org.apache.spark.sql.DataFrame = [value: int]

scala> numesRead.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


scala> 
```