

Problem Statement

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

Solutions Question 1 Code:

```
val dataRDD = sc.parallelize(1 to 100)
```

```
val dataDF = dataRDD.toDF()
```

```
dataDF.write.parquet("data.parquet")
```

```
val newDataDF = spark.read.parquet("data.parquet")
```

```
newDataDF.show()
```

Screen-shot

```
scala> val dataRDD = sc.parallelize(1 to 100)
```

```
dataRDD: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[30] at parallelize at <console>:28
```

```
scala> val dataDF = dataRDD.toDF()
```

```
dataDF: org.apache.spark.sql.DataFrame = [value: int]
```

```
scala> dataDF.write.parquet("data.parquet")
```

Accadgild_Session_19_Assignment_19.3_Solutions

```
scala> val newDataDF = spark.read.parquet("data.parquet")
newDataDF: org.apache.spark.sql.DataFrame = [value: int]
```

```
scala>
```

```
scala> newDataDF.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> █
```

Activate Windows
Go to Settings to activate Windows.