**Session 19 Assignment 3**

**// Parallelizing data**

scala> val data1To100 = sc.parallelize(1 to 100)

data1To100: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[11] at parallelize at <console>:27

**// Creating a dataframe**

scala> val dataFrame = data1To100.toDF()

dataFrame: org.apache.spark.sql.DataFrame = [\_1: int]

**// Writing data into parquet file**

scala> dataFrame.write.parquet("One\_To\_Hundread.parquet")

**// Reading data from parquet file and assigning to rdd**

scala> val readOneToHundread = sqlContext.read.parquet("One\_To\_Hundread.parquet")

readOneToHundread: org.apache.spark.sql.DataFrame = [\_1: int]

**// Displaying the data – Only 20 row max**

scala> readOneToHundread.show()

+---+

| \_1|

+---+

| 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

**// Displaying the data – All Rows**

scala> readOneToHundread.collect.foreach(println)

[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]

[37]

[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]

scala>