**Assignment 3**

**How to run the code:**

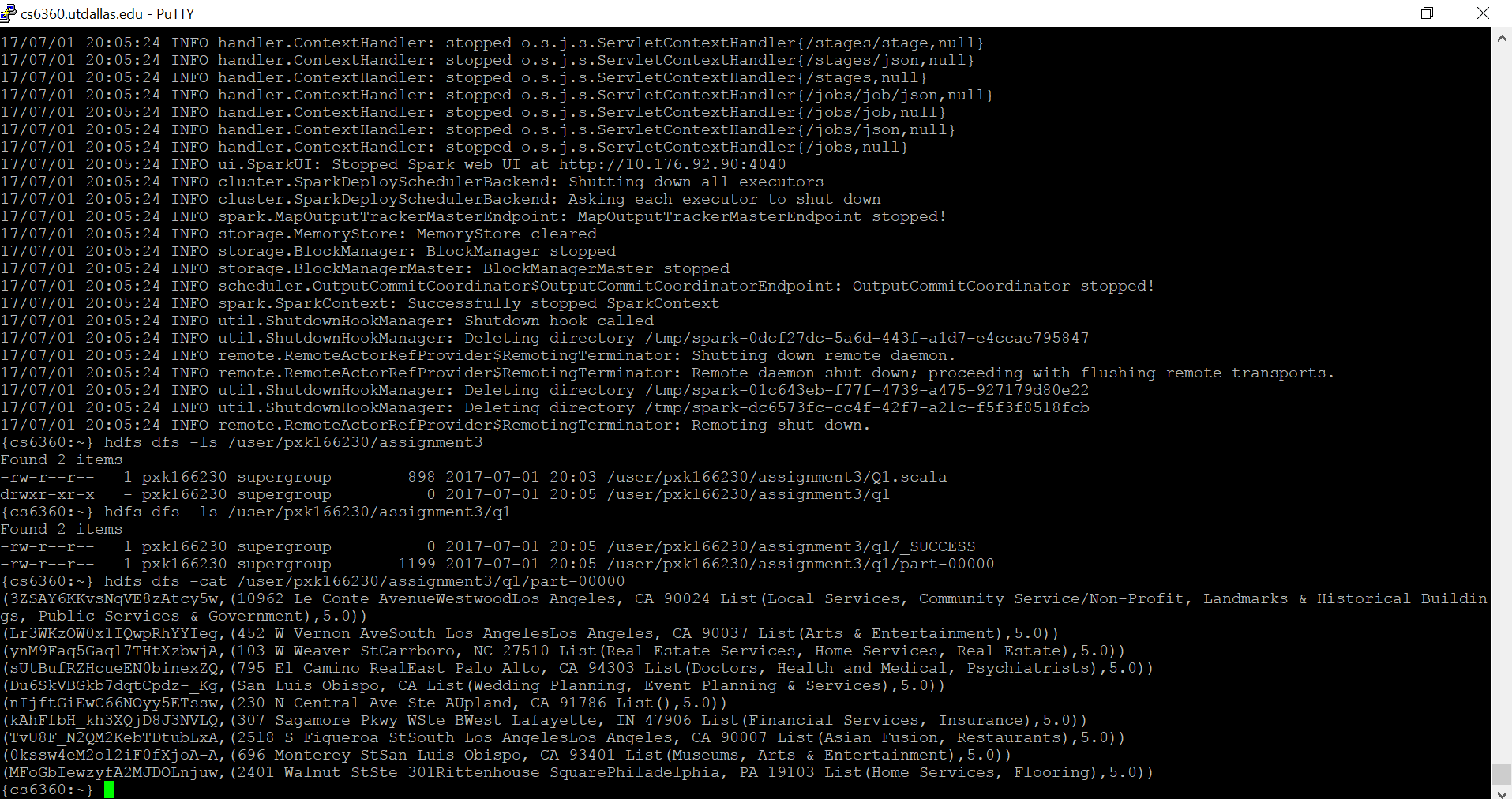
Copy the scala files into HDFS and use the following commands to run:

Q1: spark-shell -i Q1.scala

Output is stored in folder q1. Retrieve it by using the command hdfs dfs –cat <output-folder>. If the folder already exists, delete it by using hdfs dfs -rm -r <output-folder>

hdfs dfs -cat /user/pxk166230/assignment3/q1/part-00000

Screenshot of the output:

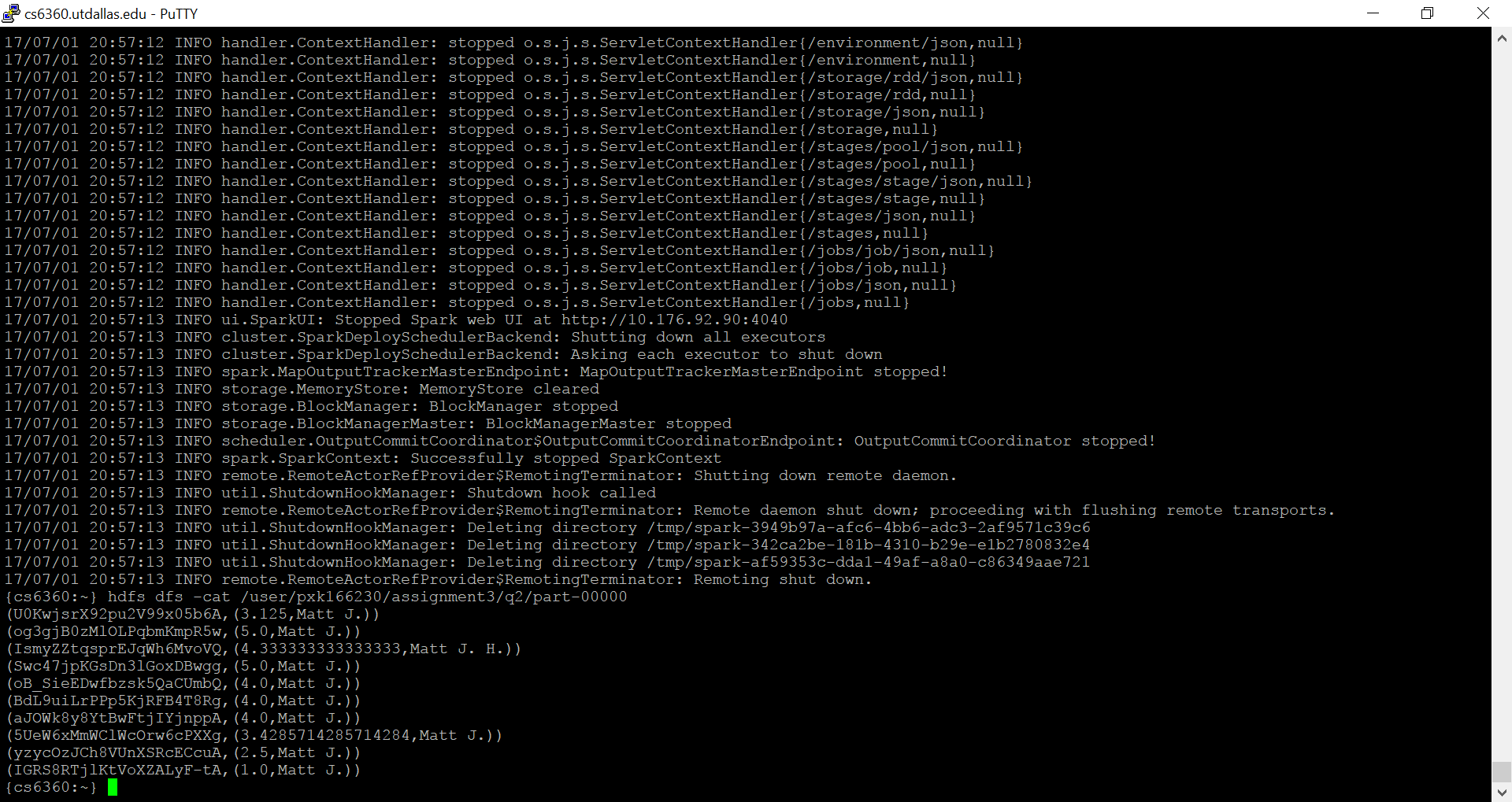


Q2: spark-shell -i Q2.scala

Output is stored in folder q2. Retrieve it by using the command hdfs dfs –cat <output-folder>

hdfs dfs -cat /user/pxk166230/assignment3/q2/part-00000

Screenshot of the output:

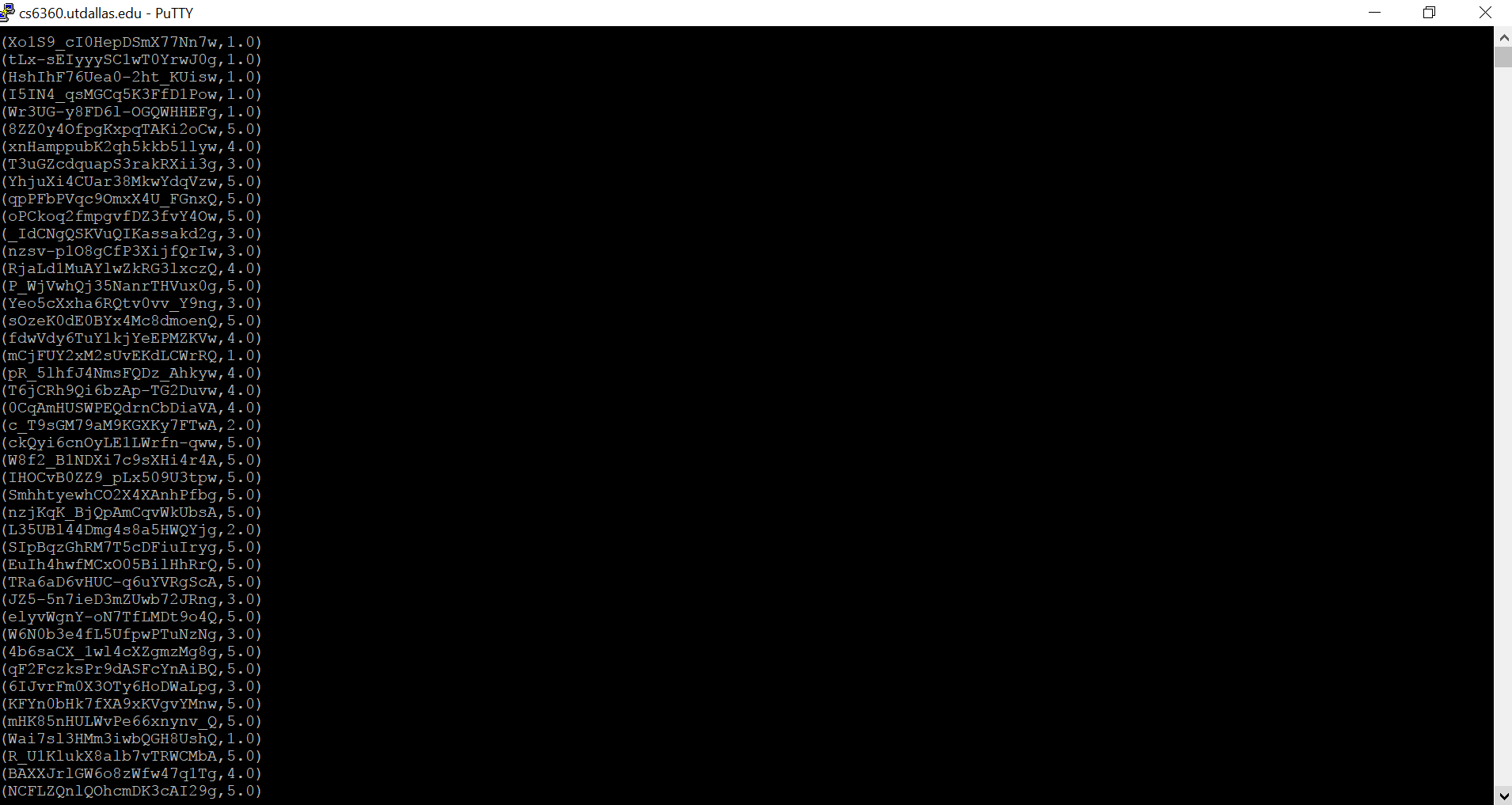
****

Q3: spark-shell -i Q3.scala

Output is stored in folder q3. Retrieve it by using the command hdfs dfs –cat <output-folder>

hdfs dfs -cat /user/pxk166230/assignment3/q3/part-00000

Screenshot of the output:

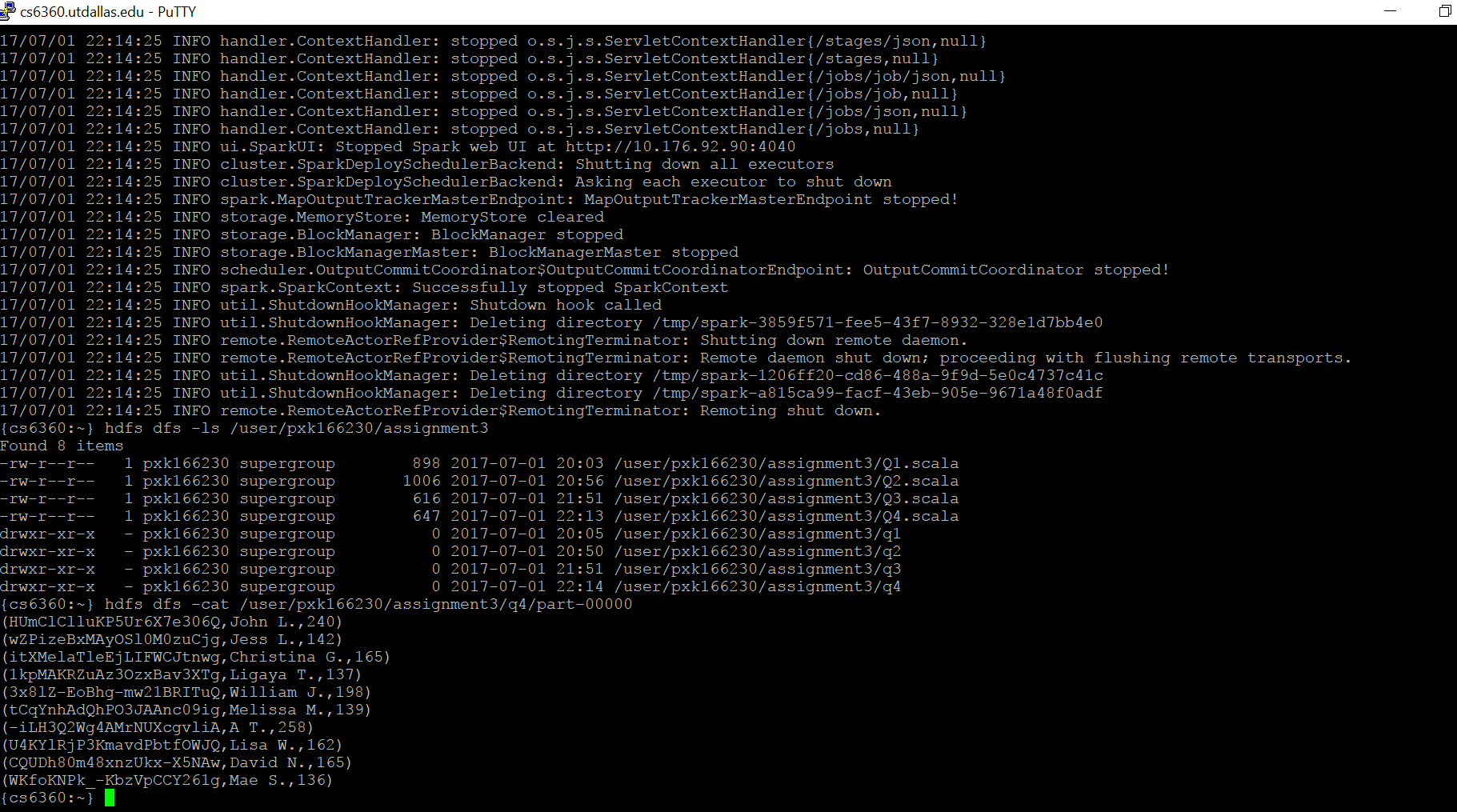
****

Q4: spark-shell -i Q4.scala

Output is stored in folder q4. Retrieve it by using the command hdfs dfs –cat <output-folder>

hdfs dfs -cat /user/pxk166230/assignment3/q4/part-00000

Screenshot of the output:

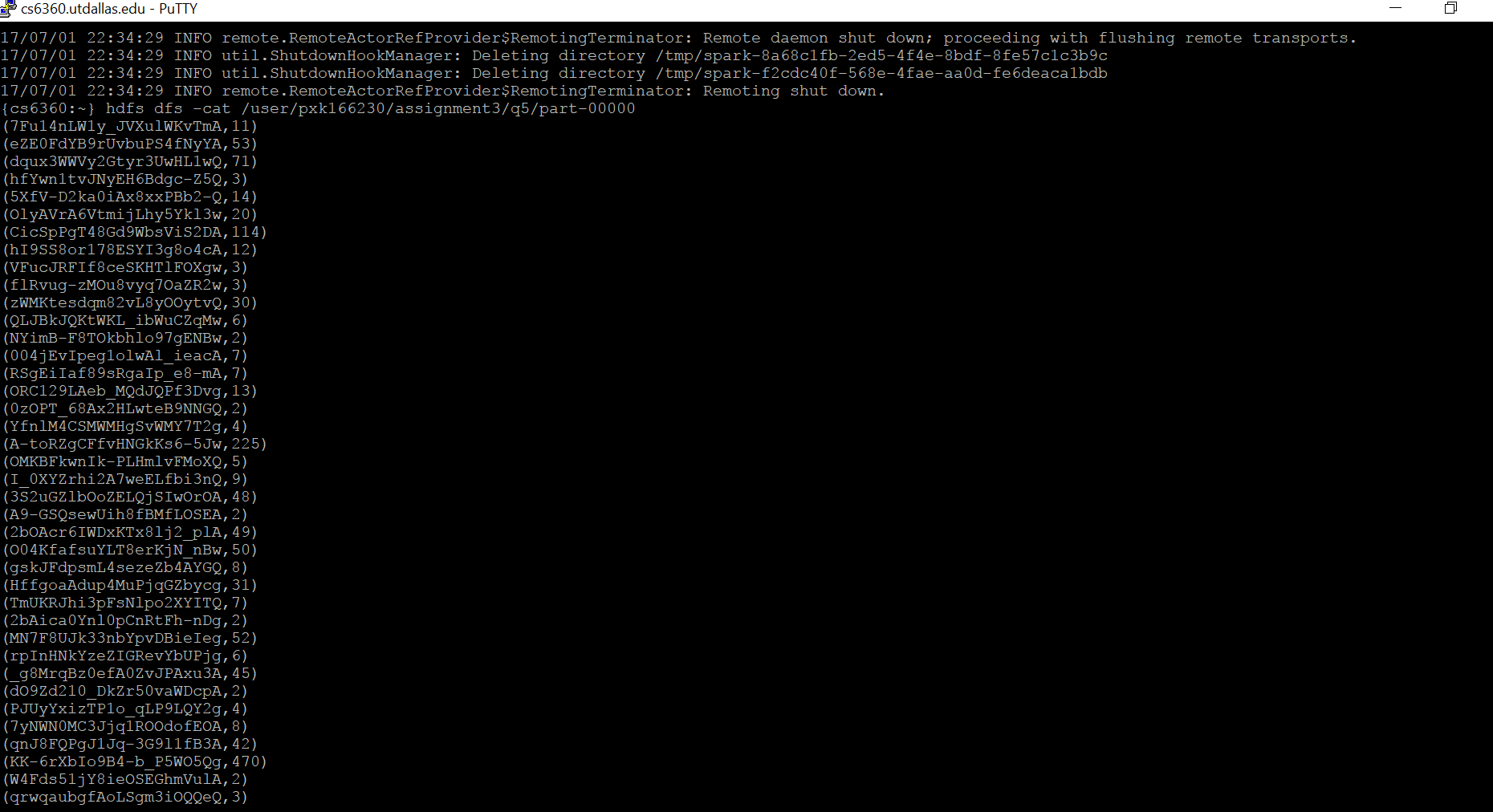
****

Q5: spark-shell -i Q5.scala

Output is stored in folder q5. Retrieve it by using the command hdfs dfs –cat <output-folder>

hdfs dfs -cat /user/pxk166230/assignment3/q5/part-00000

Screenshot of the output:

****