### **Data Ingestion from the RDS to HDFS using Sqoop**

***Sqoop Import command used for importing table from RDS to HDFS:***

*sqoop import \*

*> --connect jdbc:mysql://upgraddetest.cyaielc9bmnf.us-east-1.rds.amazonaws.com/testdatabase \*

*> --table SRC\_ATM\_TRANS \*

*> --username student --password STUDENT123 \*

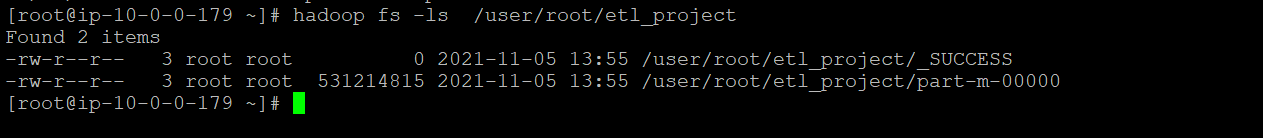
*> --target-dir /user/root/etl\_project \*

*> -m 1 ;*

***Command used to see the list of imported data in HDFS:***

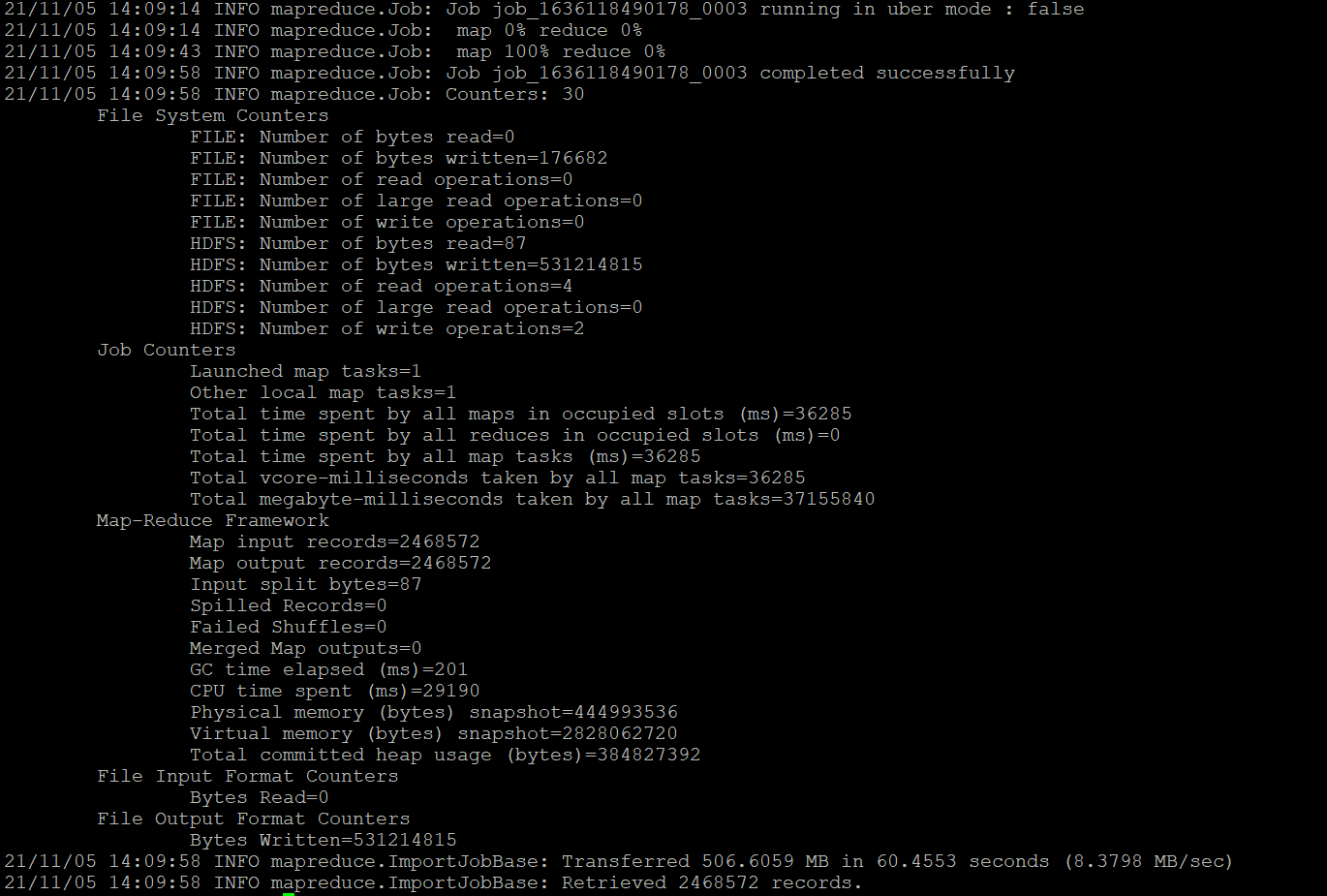
*hadoop fs -ls /user/root/etl\_project*

***Screenshot of the imported data:***

**

***Explanation and comments:***

1. We logged into ec2 instance and switched to root user before starting off with importing of data from RDS.
2. We have imported data from given rds instance to hdfs using sqoop import command with 1 mapper job.
3. We mentioned the target directory in sqoop command as **/user/root/etl\_project**.

Once sqoop finishes importing the data it gives info on number of map jobs and reduce jobs, in this case reduce is always 0. And in the end shows total size of the table transferred, number of rows it has and time taken to transfer.

1. Once sqoop import is done we verified the data is present in hdfs by typing -> hadoop fs -ls /user/root/etl\_project. This shows two outputs -> \_SUCEESS indicating sqoop import was successful and next is part-m-00000 as we used only 1 mapper job there is only 1 part.