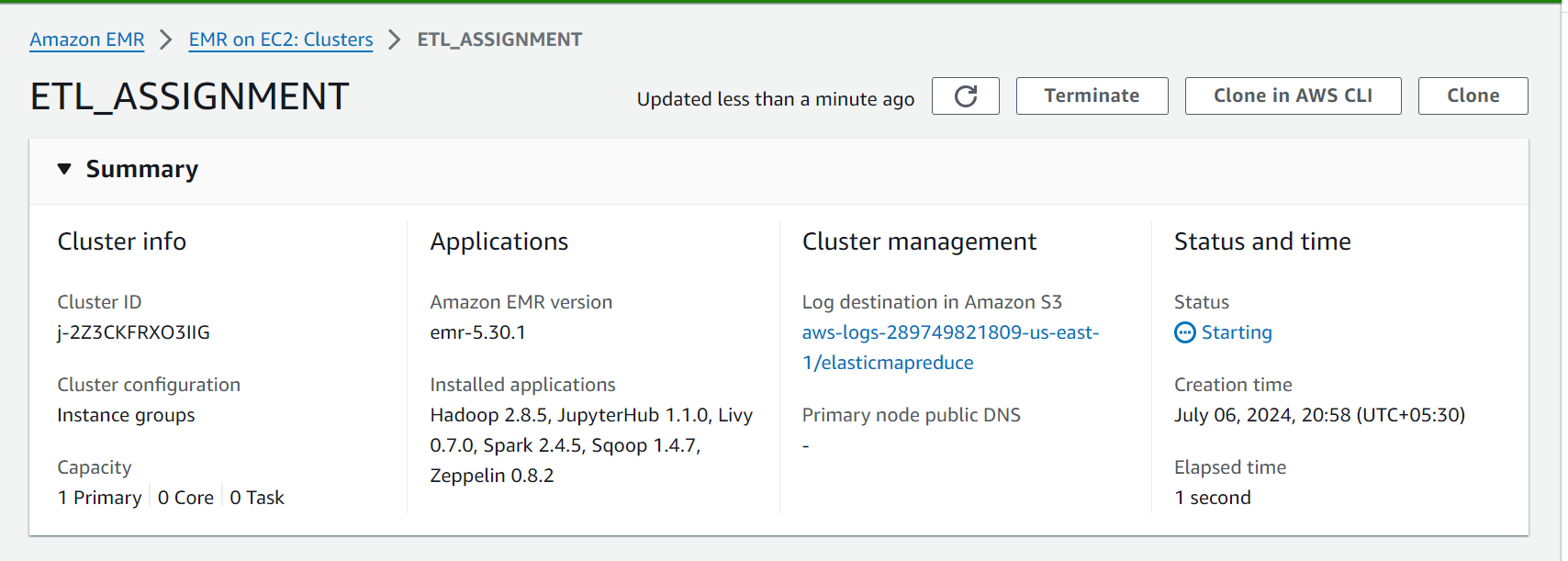
Apache Sqoop SQL to HADOOP

Sqoop used MapReduce to import and export the data, which provides parallel operation as well as fault tolerance.

EMR CLUSTER CREATION



SQL CONNECTOR INSTALLATION

wget <https://de-mysql-connector.s3.amazonaws.com/mysql-connector-java-8.0.25.tar.gz>

tar -xvf mysql-connector-java-8.0.25.tar.gz

cd mysql-connector-java-8.0.25/

sudo cp mysql-connector-java-8.0.25.jar /usr/lib/sqoop/lib/

mysql\_secure\_installation

mysql -u root -p

MariaDB [(none)]> GRANT ALL PRIVILEGES ON \*.\* TO 'root'@'%' identified by '1234' WITH GRANT OPTION;

Query OK, 0 rows affected (0.00 sec)

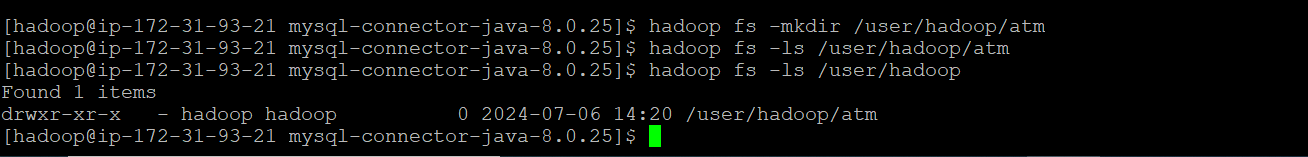
[hadoop@ip-172-31-93-21 mysql-connector-java-8.0.25]$ sudo service mysqld restart

Redirecting to /bin/systemctl restart mysqld.service

hadoop fs -mkdir /user/hadoop/atm

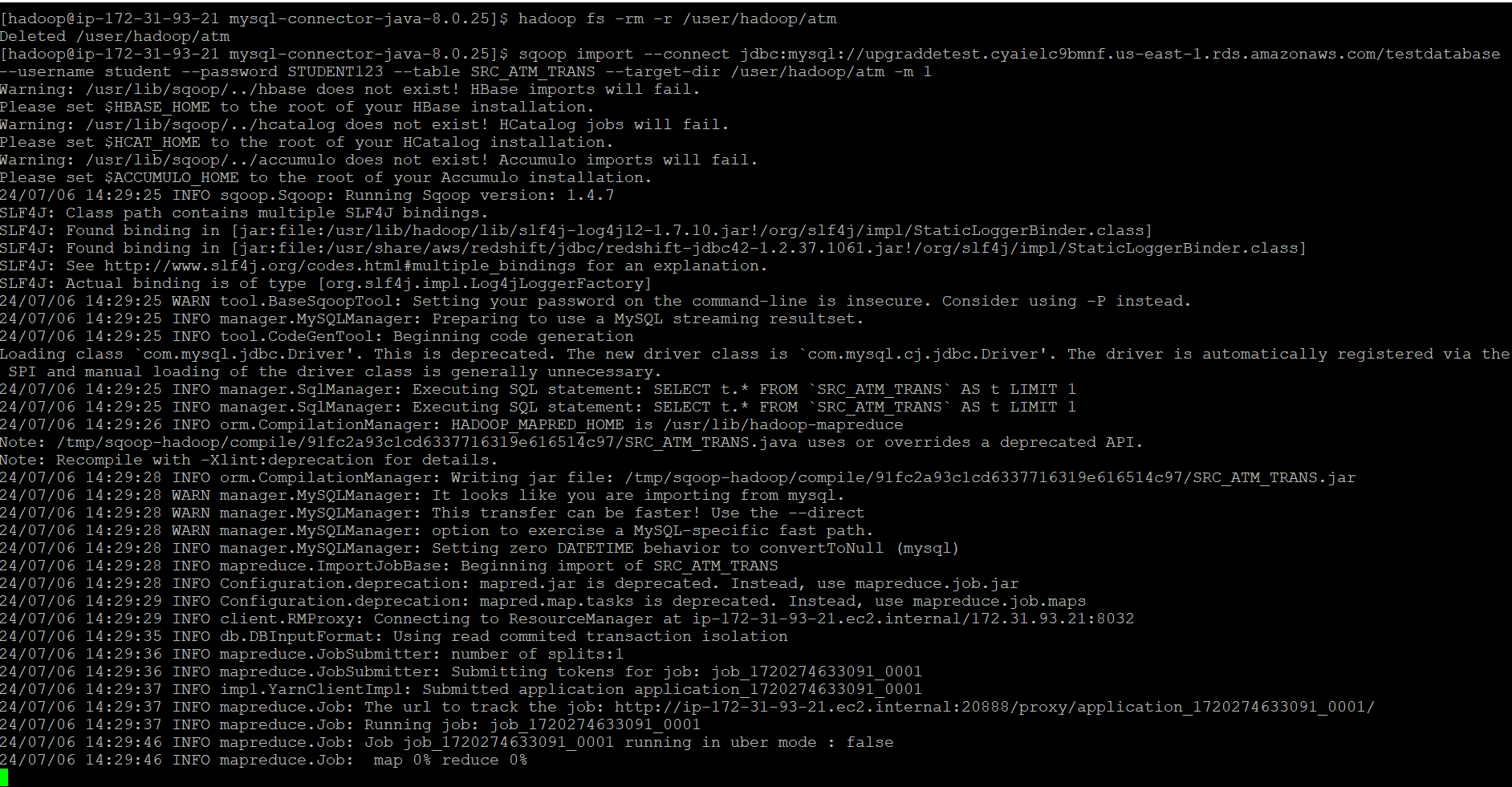
hadoop fs -ls /user/Hadoop

24/07/06 14:22:34 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7



**Note**: Before running the Sqoop import command, you need to ensure that the target directory does not already exist in HDFS. Otherwise, an error will be thrown that the target directory does not exist.

sqoop import --connect jdbc:mysql://upgraddetest.cyaielc9bmnf.us-east-1.rds.amazonaws.com/testdatabase --username student --password STUDENT123 --table SRC\_ATM\_TRANS --target-dir /user/hadoop/atm -m 1



A screenshot of a computer

Description automatically generated

24/07/06 14:30:18 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 49.4649 seconds (10.2417 MB/sec)

24/07/06 14:30:18 INFO mapreduce.ImportJobBase: Retrieved **2468572** records.

**Data Ingestion with Sqoop ● Please check the number of records that are imported after the Sqoop Job. Count of Records – 2468572**

[hadoop@ip-172-31-93-21 mysql-connector-java-8.0.25]$ hadoop fs -cat /user/hadoop/atm/part-m-\* | head -n 10

