Load data from AWS RDS to Hadoop

Ingesting data from AWS EDS to HDFS using Sqoop

1. **Installing MySQL connector**

   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/

   ```
   [hadoop@ip-172-31-44-42 ~]$ wget https://de-mysql-connector.s3.amazonaws.com/mysql-connector-java-8.0.25.tar.gz
   --2025-02-12 05:19:09--  https://de-mysql-connector.s3.amazonaws.com/mysql-connector-java-8.0.25.tar.gz
   Resolving de-mysql-connector.s3.amazonaws.com (de-mysql-connector.s3.amazonaws.com)... 3.5.27.34, 52.217.125.41, 3.5.29.50, ...
   Connecting to de-mysql-connector.s3.amazonaws.com (de-mysql-connector.s3.amazonaws.com)|3.5.27.34|:443... connected.
   HTTP request sent, awaiting response... 200 OK
   Length: 4079310 (3.9M) [application/x-gzip]
   Saving to: 'mysql-connector-java-8.0.25.tar.gz.1'

   mysql-connector-java-8.0.25.tar.gz.1              0%[
   mysql-connector-java-8.0.25.tar.gz.1            100%[===================================================================

   2025-02-12 05:19:09 (64.3 MB/s) - 'mysql-connector-java-8.0.25.tar.gz.1' saved [4079310/4079310]
   ```

   ```
   [hadoop@ip-172-31-44-42 ~]$ cd mysql-connector-java-8.0.25/
   [hadoop@ip-172-31-44-42 mysql-connector-java-8.0.25]$ sudo cp mysql-connector-java-8.0.25.jar /usr/lib/sqoop/lib/
   [hadoop@ip-172-31-44-42 mysql-connector-java-8.0.25]$ []
   ```

2. **Executing sqoop import command ingest the relevant bookings data from AWS RDS to HDFS**

   sqoop import \

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

   --username student --password STUDENT123 \

   --table bookings \

   --target-dir /user/hadoop/bookings \

   --driver com.mysql.cj.jdbc.Driver \

   --fields-terminated-by ',' \

   --as-textfile \

   -m 1

```
[hadoop@ip-172-31-44-42 ~]$ sqoop import \
--connect jdbc:mysql://upgraddetest.cyaielc9bmnf.us-east-1.rds.amazonaws.com/testdatabase \
--username student --password STUDENT123 \
--table bookings \
--target-dir /user/hadoop/bookings \
--driver com.mysql.cj.jdbc.Driver \
--fields-terminated-by ',' \
--as-textfile \
-m 1
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.

        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=165678
2025-02-12 05:23:54,963 INFO mapreduce.ImportJobBase: Transferred 161.7949 KB in 18.5619 seconds (8.7165 KB/sec)
2025-02-12 05:23:54,967 INFO mapreduce.ImportJobBase: Retrieved 1000 records.
[hadoop@ip-172-31-44-42 ~]$
```

1000 records retrieved and imported to HDFS.

3. **Verifying the imported file in HDFS.**
   hadoop fs -ls /user/hadoop/bookings

```
[hadoop@ip-172-31-44-42 ~]$ hadoop fs -ls /user/hadoop/bookings
Found 2 items
-rw-r--r--   1 hadoop hdfsadmingroup          0 2025-02-12 05:23 /user/hadoop/bookings/_SUCCESS
-rw-r--r--   1 hadoop hdfsadmingroup     165678 2025-02-12 05:23 /user/hadoop/bookings/part-m-00000
[hadoop@ip-172-31-44-42 ~]$
```

4. **Verifying imported file in HDFS**
   hadoop fs -cat /user/hadoop/bookings/part-m-00000 | head -10

```
[hadoop@ip-172-31-44-42 ~]$ hadoop fs -cat /user/hadoop/bookings/part-m-00000 | head -10
BK8968087150,51811359,15055660,2.2.14,Android,-49.4319655,103.917851,-58.8043875,146.477367,2020-06-23 19:33
BK629851904,31663218,60872180,3.4.1,iOS,-83.5408405,175.80085,86.20705,128.367238,2020-05-23 12:22:04.0,2020
BK1797410350,86869399,94276051,4.1.36,iOS,-67.8930645,55.234128,-51.1079,-31.07475,2020-05-19 14:14:32.0,202
BK5788246325,58230837,45457227,2.4.27,Android,13.707887,113.499943,54.3812915,-18.437751,2020-03-24 01:30:15
BK8342703255,84232510,86494681,4.1.34,Android,-6.091461,-114.649789,22.8449505,70.137827,2020-08-03 19:10:52
BK6015582453,11981042,35862658,2.4.39,iOS,-18.910034,-70.193103,-10.182921,173.877213,2020-07-17 05:33:48.0,
BK4529355854,60071878,78022360,2.1.9,iOS,1.215274,-56.014903,35.152876,104.324905,2020-01-02 01:48:40.0,2020
BK9720088219,14327312,94427067,3.1.2,Android,-55.4822225,173.362256,65.0121265,51.390751,2020-04-10 15:11:07
BK7157532607,46407210,43160003,1.3.4,Android,46.005843,-16.826146,7.6126015,-156.428577,2020-06-09 05:56:31.
BK5014871433,65861573,64708618,1.3.28,iOS,-29.565326,64.843709,84.068109,-49.820835,2020-08-14 20:43:42.0,20
cat: Unable to write to output stream.
[hadoop@ip-172-31-44-42 ~]$
```