Load data from AWS RDS to Hadoop

<Command to import data from AWS RDS to Hadoop>

1. Setting up MySql connector with commands

a. Install the MySQL connector jar file:

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

b. Extract the MySQL connector tar file

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

c. Getting into MySQL Connector directory

cd mysql-connector-java-8.0.25/

d. Copying it to the Sqoop library

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

2. Run the Sqoop import command to import data from AWS RDS to Hadoop

```
sqoop import \
--connect jdbc:mysql://upgraddetest.cyaielc9bmnf.us-east-1.rds.amazonaws.com/testdatabase \
--table bookings \
--username student --password STUDENT123 \
--null-string '\\N' --null-non-string '\\N' \
--target-dir bookings_data \
-m 1
```

<Command to view the imported data>

hadoop fs -ls bookings_data

hadoop fs -cat bookings_data/part-m-00000 | wc -l

<Screenshot of the data>

```
[hadoop@ip-172-31-70-125 ~]$ hadoop fs -ls bookings_data
Found 2 items
-rw-r--r--   1 hadoop hadoop          0 2024-04-26 19:18 bookings_data/_SUCCESS
-rw-r--r--   1 hadoop hadoop     165678 2024-04-26 19:18 bookings_data/part-m-00000
[hadoop@ip-172-31-70-125 ~]$ hadoop fs -cat bookings_data/part-m-00000 | wc -l
1000
[hadoop@ip-172-31-70-125 ~]$
```