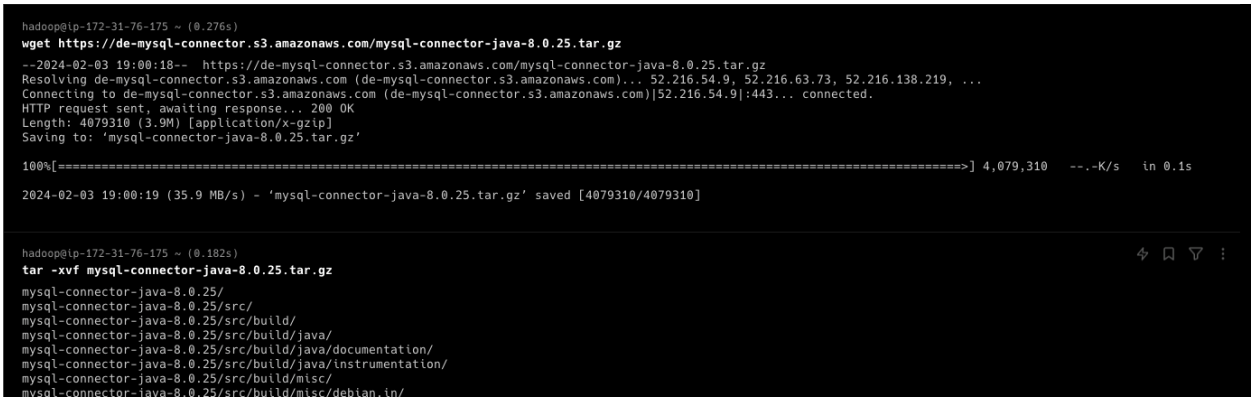
# Load data from AWS RDS to Hadoop

## Data Ingestion with sqoop

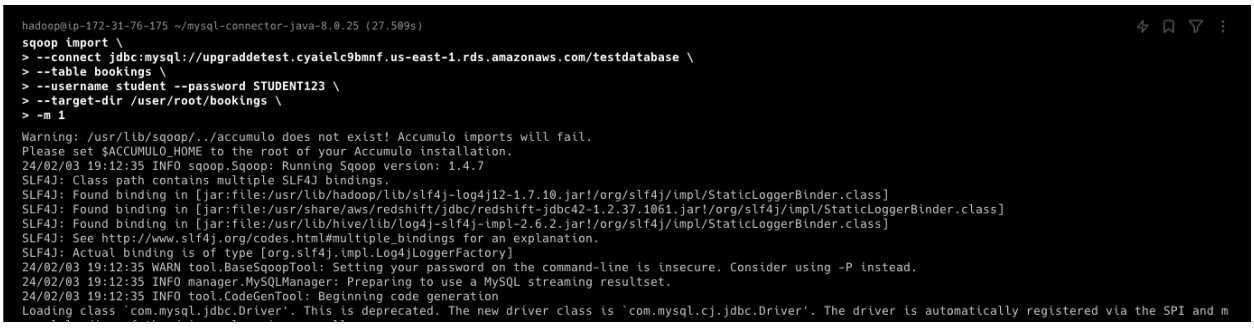
1. Install MySQL connector

```
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/
```




2. Run 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 \
--target-dir /user/root/bookings \
-m 1
```



```
                Total vcore-milliseconds taken by all map tasks=4769
                Total megabyte-milliseconds taken by all map tasks=7325184
        Map-Reduce Framework
                Map input records=1000
                Map output records=1000
                Input split bytes=87
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=99
                CPU time spent (ms)=2930
                Physical memory (bytes) snapshot=286011392
                Virtual memory (bytes) snapshot=3304468480
                Total committed heap usage (bytes)=239599616
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=165678
24/02/03 19:13:00 INFO mapreduce.ImportJobBase: Transferred 161.7949 KB in 20.154 seconds (8.0279 KB/sec)
24/02/03 19:13:00 INFO mapreduce.ImportJobBase: Retrieved 1000 records.
```

1000 records have been fetched

3. Verify files in hdfs

`hadoop fs -ls /user/root/bookings`

```
hadoop@ip-172-31-76-175 ~/mysql-connector-java-8.0.25 (2.295s)
hadoop fs -ls /user/root/bookings
Found 2 items
-rw-r--r--   1 hadoop hadoop          0 2024-02-03 19:12 /user/root/bookings/_SUCCESS
-rw-r--r--   1 hadoop hadoop     165678 2024-02-03 19:12 /user/root/bookings/part-m-00000
```

`hadoop fs -cat /user/root/bookings/part-m-00000 | head -n 5`

```
hadoop@ip-172-31-76-175 ~/mysql-connector-java-8.0.25 (2.628s)
hadoop fs -cat /user/root/bookings/part-m-00000 | head -n 5
BK8968087150,51811359,15055660,2.2.14,Android,-49.4319655,103.917851,-58.8043875,146.477367,2020-06-23 19:33:10.0,2020-06-06 09:02:10.0,534,83,INR,black,054-38-4479,4,
3,3
BK6298519O4,31663218,60872180,3.4.1,iOS,-83.5408405,175.80085,86.20705,128.367238,2020-05-23 12:22:04.0,2020-08-09 19:02:56.0,126,67,INR,lime,796-39-6801,3,2,4
BK1797410350,86869399,94276051,4.1.36,iOS,-67.8930645,55.234128,-51.1079,-31.07475,2020-05-19 14:14:32.0,2020-08-23 18:38:39.0,297,63,INR,olive,748-73-1579,1,3,3
BK5788246325,58230837,45457227,2.4.27,Android,13.707887,113.499943,54.3812915,-18.437751,2020-03-24 01:30:15.0,2020-05-19 11:16:45.0,932,32,INR,white,558-80-6346,3,2,2
BK8342703255,84232510,86494681,4.1.34,Android,-6.091461,-114.649789,22.8449505,70.137827,2020-08-03 19:10:52.0,2020-03-24 08:25:40.0,260,7,INR,blue,068-72-1637,3,3,3
cat: Unable to write to output stream.
```