

Load data from AWS RDS to Hadoop

Command to import data from AWS RDS to Hadoop

Command:

```
sqoop import \  
--connect jdbc:mysql://#####.#####.us-east-1.rds.amazonaws.com/testdatabase \  
--table bookings \  
--username student \  
--password ##### \  
--target-dir /user/hadoop/bookings_data \  
-m 1
```

Explanation: This is the sqoop import command used to import the data from the AWS RDS to the HDFS using the RDS connection string, username, password and table name provided in the problem statement of the project. I have also provided the target directory where the data will be copied in the HDFS (/user/hadoop/bookings_data).

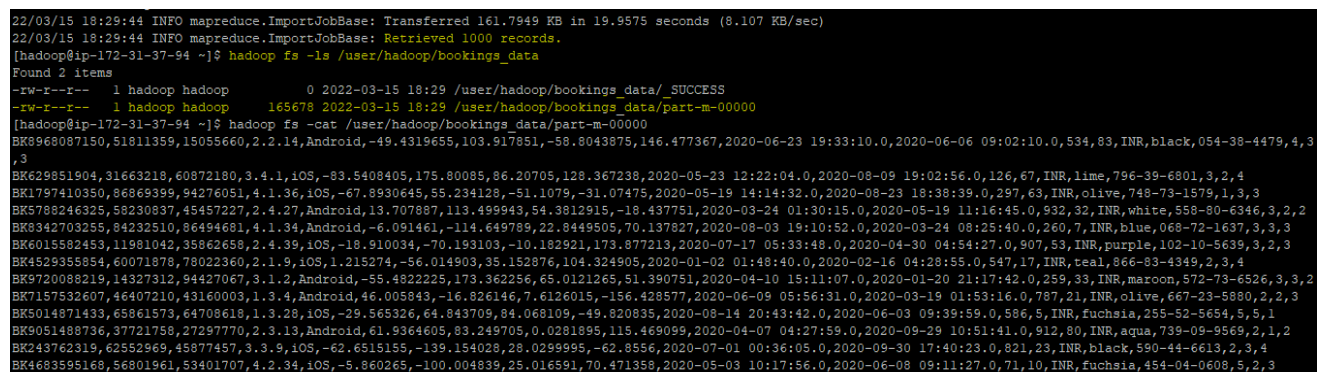
Command to view the imported data

Command:

```
hadoop fs -ls /user/hadoop/bookings_data
```

Explanation: This is the command used to list the imported data in the HDFS. This command will show the files present in the target directory in the HDFS (/user/hadoop/bookings_data). The command will show two files named **_SUCCESS** and **part-m-00000** in the output.

Screenshot of the data



```
22/03/15 18:29:44 INFO mapreduce.ImportJobBase: Transferred 161.7949 KB in 19.9575 seconds (8.107 KB/sec)
22/03/15 18:29:44 INFO mapreduce.ImportJobBase: Retrieved 1000 records.
[hadoop@ip-172-31-37-94 ~]$ hadoop fs -ls /user/hadoop/bookings_data
Found 2 items
-rw-r--r-- 1 hadoop hadoop 0 2022-03-15 18:29 /user/hadoop/bookings_data/_SUCCESS
-rw-r--r-- 1 hadoop hadoop 165678 2022-03-15 18:29 /user/hadoop/bookings_data/part-m-00000
[hadoop@ip-172-31-37-94 ~]$ hadoop fs -cat /user/hadoop/bookings_data/part-m-00000
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
BK629851904,31663218,60872180,3.4.1.1,IOS,-83.5408405,175.80085,86.20705,128.367238,2020-05-23 12:22:04.0,2020-08-09 18:02:56.0,126,67,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 18:39:39.0,297,63,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
BK6015582453,11981042,35862658,2.4.39,IOS,-18.910034,-70.193103,-10.182921,173.877213,2020-07-17 05:33:48.0,2020-04-30 04:54:27.0,907,53,INR,purple,102-10-5639,3,2,3
BK4529355854,60071878,78022360,2.1.9,IOS,1.215274,-56.014903,35.152876,104.324905,2020-01-02 01:48:40.0,2020-02-16 04:28:55.0,547,17,INR,teal,866-83-4349,2,3,4
BK9720088219,14327312,94427067,3.1.2,Android,-55.4822225,173.362256,65.0121265,51.390751,2020-04-10 15:11:07.0,2020-01-20 21:17:42.0,259,33,INR,maroon,572-73-6526,3,3,2
BK7157532607,46407210,43160003,1.3.4,Android,46.005843,-16.826146,7.6126015,-156.428577,2020-06-09 05:56:31.0,2020-03-19 01:53:16.0,787,21,INR,olive,667-23-5880,2,2,3
BK5014871433,65861573,64708618,1.3.28,IOS,-29.565326,64.843709,84.068109,-49.820835,2020-08-14 20:43:42.0,2020-06-03 09:39:59.0,586,5,INR,fuchsia,255-52-5654,5,5,1
BK9051488736,37721758,27297770,2.3.13,Android,61.9364605,83.249705,0.0281895,115.469099,2020-04-07 04:27:59.0,2020-09-29 10:51:41.0,912,80,INR,aqua,739-09-9569,2,1,2
BK243762319,62552969,45877457,3.3.9,IOS,-62.6515155,-139.154028,28.0299995,-62.8556,2020-07-01 00:36:05.0,2020-09-30 17:40:23.0,821,23,INR,black,590-44-6613,2,3,4
BK4683595168,56801961,53401707,4.2.34,IOS,-5.860265,-100.004839,25.016591,70.471358,2020-05-03 10:17:56.0,2020-06-08 09:11:27.0,71,10,INR,fuchsia,454-04-0608,5,2,3
```

Explanation: Above is the screenshot of the data imported from the AWS RDS to the HDFS using Sqoop import command.