

Data Ingestion from the RDS to HDFS using Sqoop

Sqoop Import command used for importing table from RDS to HDFS:

To proceed with Sqoop import command, It should be made sure that the target directory does Not already exists. Otherwise Sqoop import command will throw error.

```
hadoop fs -rm -r /user/root/SRC_ATM_TRANS
```

Sqoop Import Command :

```
sqoop import \  
-connect jdbc:mysql://upgradtest.cyaielec9bmnf.us-east-1.rds.amazonaws.com/testdatabase -  
table SRC_ATM_TRANS \  
-username student -password STUDENT123 \  
-target-dir /user/root/SRC_ATM_TRANS \  
-m 1
```

It can be seen in below screenshot that sqoop import job has fetched **2468572** records (As the checkpoint mentioned in the Validation document):

```
22/09/28 18:46:07 INFO mapreduce.Job: Counters: 30  
  File System Counters  
    FILE: Number of bytes read=0  
    FILE: Number of bytes written=189849  
    FILE: Number of read operations=0  
    FILE: Number of large read operations=0  
    FILE: Number of write operations=0  
    HDFS: Number of bytes read=87  
    HDFS: Number of bytes written=531214815  
    HDFS: Number of read operations=4  
    HDFS: Number of large read operations=0  
    HDFS: Number of write operations=2  
  Job Counters  
    Launched map tasks=1  
    Other local map tasks=1  
    Total time spent by all maps in occupied slots (ms)=2038560  
    Total time spent by all reduces in occupied slots (ms)=0  
    Total time spent by all map tasks (ms)=21235  
    Total vcore-milliseconds taken by all map tasks=21235  
    Total megabyte-milliseconds taken by all map tasks=65233920  
  Map-Reduce Framework  
    Map input records=2468572  
    Map output records=2468572  
    Input split bytes=87  
    Spilled Records=0  
    Failed Shuffles=0  
    Merged Map outputs=0  
    GC time elapsed (ms)=174  
    CPU time spent (ms)=21760  
    Physical memory (bytes) snapshot=1040510976  
    Virtual memory (bytes) snapshot=4625391616  
    Total committed heap usage (bytes)=958398464  
  File Input Format Counters  
    Bytes Read=0  
  File Output Format Counters  
    Bytes Written=531214815  
22/09/28 18:46:07 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 37.578 seconds (13.4815 MB/sec)  
22/09/28 18:46:07 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.  
[hadoop@ip-172-31-8-119 mysql-connector-java-8.0.25]$
```

Command used to see the list of imported data in HDFS:

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

We can see in below screenshot that target directory has 2 files:

- Success file showing that the import job was successful
- Second file part-m-00000 in which all the records are present. As the number of mappers used in command was 1 only 1 file got created.

```
[hadoop@ip-172-31-8-119 mysql-connector-java-8.0.25]$ hadoop fs -ls /user/root/SRC_ATM_TRANS
Found 2 items
-rw-r--r-- 1 hadoop hadoop          0 2022-09-28 18:46 /user/root/SRC_ATM_TRANS/_SUCCESS
-rw-r--r-- 1 hadoop hadoop 531214815 2022-09-28 18:46 /user/root/SRC_ATM_TRANS/part-m-00000
[hadoop@ip-172-31-8-119 mysql-connector-java-8.0.25]$
```

Screenshot of the imported data:

```

2017, January, 16, Monday, 12, Active, 15, NCR, Vestre, Kasen, 36, 9000, 57.853, 9.085, DKK, Visa Dankort, 8457, Withdrawal, --, 57.848, 9.919, 2624886, Aalborg, 271.150, 1028, 79, 3, 360, 0.00, 2801, Clouds, few clouds
2017, January, 16, Monday, 12, Inactive, 89, NCR, Intern Frederikshavn, Danmarksgade, 48, 9900, 57.441, 18.537, DKK, Dankort -- on-us, 6468, Withdrawal, --, 57.441, 18.537, 2621927, Frederikshavn, 271.150, 1028, 79, 3, 360, 0.00, 2, 801, Clouds, few clouds
2017, January, 16, Monday, 12, Active, 34, NCR, Hobø, Adelgade, 31, 9500, 56.638, 9.794, DKK, Visa Dankort, 3020, Withdrawal, --, 56.643, 9.799, 2620167, Hobø, 271.150, 1028, 79, 3, 360, 0.00, 2801, Clouds, few clouds
2017, January, 16, Monday, 12, Active, 34, NCR, Skippersen, Vestre Alle, 2, 9000, 57.858, 9.943, 9.968, DKK, Visa Dankort -- on-us, 4865, Withdrawal, --, 57.848, 9.919, 2624886, Aalborg, 271.150, 1028, 79, 3, 360, 0.00, 2801, Clouds, few clouds
2017, January, 16, Monday, 12, Active, 79, NCR, Middelfart, Brogade, 9, 5500, 55.507, 9.727, DKK, Visa Dankort -- on-us, 1264, Withdrawal, --, 55.586, 9.731, 2616933, Middelfart, 273.730, 1027, 74, 2, 350, 0.00, 0, 800, Clear, Sky is Clear
2017, January, 16, Monday, 12, Active, 35, NCR, Aabybro, ÅFestergade, 6, 9440, 57.162, 9.730, DKK, Visa Dankort -- on-us, 4128, Withdrawal, --, 57.158, 9.737, 2625837, Aabybro, 271.150, 1028, 79, 3, 360, 0.00, 2801, Clouds, few clouds
2017, January, 16, Monday, 12, Active, 99, NCR, Intern Vejle, Havneparken, 4, 7100, 55.718, 11.764, DKK, Visa Dankort -- on-us, 6845, Withdrawal, --, 55.789, 9.536, 2610613, Vejle, 271.150, 1028, 98, 2, 330, 0.00, 0, 800, Clear, Sky is Clear
2017, January, 16, Monday, 12, Active, 26, NCR, Holbæk, ÅF, Slotsvolden, 7, 4300, 55.718, 11.784, DKK, Visa Dankort -- on-us, 1843, Withdrawal, --, 55.780, 11.770, 2620214, Holbæk, 272.850, 1026, 69, 4, 350, 0.00, 0, 800, Clear, Sky is Clear
2017, January, 16, Monday, 12, Active, 64, NCR, Horsens, Gråfå, nlandsvej, 5, 8900, 55.859, 9.854, DKK, Visa Dankort, 8681, Withdrawal, --, 55.861, 9.850, 2621778, Horsens, 272.690, 1029, 92, 1, 360, 0.00, 0, 800, Clear, Sky is Clear
2017, January, 16, Monday, 12, Inactive, 14, NCR, HÅF, ring, NÅF, rrealte, 12, 8362, 56.886, 18.837, DKK, Dankort -- on-us, 7874, Withdrawal, --, 56.157, 18.211, 2626452, Arhus, 273.590, 1028, 74, 1, 270, 0.00, 0, 800, Clear, Sky is Clear
2017, January, 16, Monday, 12, Inactive, 30, NCR, Nykø, ÅF, bing, Mars, Kirkevej, 1, 7900, 56.575, 9.860, DKK, Mastercard -- on-us, 3327, Withdrawal, --, 56.793, 8.663, 2615964, Nykø, 270.150, 1029, 92, 3, 325, 0.00, 0, 800, Clear, Sky is Clear
2017, January, 16, Monday, 12, Active, 104, NCR, Intern ÅFesterÅF, ÅFesterÅF, 12, 9000, 57.049, 9.922, DKK, Visa Dankort -- on-us, 3812, Withdrawal, --, 57.048, 9.919, 2624886, Aalborg, 271.150, 1028, 79, 3, 360, 0.00, 2801, Clouds, few clouds
2017, January, 16, Monday, 12, Active, 38, NCR, Bispensgade, Bispensgade, 35, 9800, 57.453, 9.996, DKK, Mastercard, 5079, Withdrawal, --, 57.444, 9.982, 2620214, Hjørring, 271.150, 1028, 79, 3, 360, 0.00, 2801, Clouds, few clouds
2017, January, 16, Monday, 12, Active, 20, NCR, Aabybro, ÅFestergade, 6, 9440, 57.162, 9.730, DKK, Visa Dankort -- on-us, 6361, Withdrawal, --, 57.158, 9.737, 2625837, Aabybro, 271.150, 1028, 79, 3, 360, 0.00, 2801, Clouds, few clouds
2017, January, 16, Monday, 12, Active, 7, Diebold Nixdorf, Hjallerup, Hjallerup Centret, 18, 9320, 57.168, 18.148, DKK, Visa Dankort -- on-us, 6567, Withdrawal, --, 57.165, 18.146, 2620275, Hjallerup, 271.150, 1028, 79, 3, 360, 0.00, 2, 801, Clouds, few clouds
2017, January, 16, Monday, 12, Active, 58, NCR, StÅF, vring, Baunebakken, 4, 9530, 56.898, 9.836, DKK, Visa Dankort -- on-us, 3753, Withdrawal, --, 56.885, 9.838, 2612500, Stovring, 271.150, 1028, 79, 3, 360, 0.00, 2801, Clouds, few clouds
2017, January, 16, Monday, 12, Inactive, 12, NCR, ÅFesterÅF, BrÅF, ÅFesterÅF, 12, 9000, 57.049, 9.922, DKK, Visa Dankort, 9062, Withdrawal, --, 57.048, 9.919, 2624886, Aalborg, 271.150, 1028, 79, 3, 360, 0.00, 2801, Clouds, few clouds
2017, January, 16, Monday, 12, Active, 48, Diebold Nixdorf, BrÅF, nderslev, Algade, 4, 9700, 57.269, 9.945, DKK, Mastercard -- on-us, 4050, Withdrawal, --, 57.118, 9.857, 2610823, Vadum, 271.150, 1028, 79, 3, 360, 0.00, 2801, Clouds, few clouds
2017, January, 16, Monday, 12, Active, 28, NCR, LÅF, gstÅF, r, ÅFesterbrogade, 8, 9670, 56.967, 9.253, DKK, Visa Dankort -- on-us, 3086, Withdrawal, --, 56.962, 9.258, 2617467, Løgstor, 271.150, 1028, 79, 3, 360, 0.00, 2801, Clouds, few clouds
2017, January, 16, Monday, 12, Active, 57, NCR, HillerÅF, d, KÅF, behavnsvej, 31, 3400, 55.933, 12.314, DKK, Visa Dankort -- on-us, 3178, Withdrawal, --, 55.967, 12.433, 2621394, Gladsaxe Kommune, 272.850, 1026, 69, 4, 350, 0.00, 0, 800, Clear, Sky is Clear
2017, January, 16, Monday, 12, Active, 38, NCR, Hasseris, Hasserisvej, 113, 9000, 57.044, 9.898, DKK, HÅF, ÅFvekor, -- on-us, 6695, Withdrawal, --, 57.048, 9.919, 2624886, Aalborg, 271.150, 1028, 79, 3, 360, 0.00, 2801, Clouds, few clouds
2017, January, 16, Monday, 12, Active, 24, NCR, Hobø, Adelgade, 31, 9500, 56.638, 9.794, DKK, Visa Dankort -- on-us, 3939, Withdrawal, --, 56.643, 9.799, 2620167, Hobø, 271.150, 1028, 79, 3, 360, 0.00, 2801, Clouds, few clouds
2017, January, 16, Monday, 12, Active, 105, Diebold Nixdorf, NÅF, resundby, Torvet, 6, 9400, 57.059, 9.922, DKK, Mastercard, 9611, Withdrawal, --, 57.048, 9.919, 2624886, Aalborg, 271.150, 1028, 79, 3, 360, 0.00, 2801, Clouds, few clouds
2017, January, 16, Monday, 12, Active, 104, NCR, Intern ÅFesterÅF, ÅFesterÅF, 12, 9000, 57.049, 9.922, DKK, Visa Dankort -- on-us, 6238, Withdrawal, --, 57.048, 9.919, 2624886, Aalborg, 271.150, 1028, 79, 3, 360, 0.00, 2801, Clouds, few clouds
2017, January, 16, Monday, 12, Active, 26, NCR, Holbæk, ÅF, Slotsvolden, 7, 4300, 55.718, 11.784, DKK, Visa Dankort, 2357, Withdrawal, --, 55.780, 11.780, 2620214, Holbæk, 272.850, 1026, 69, 4, 350, 0.00, 0, 800, Clear, Sky is Clear
2017, January, 16, Monday, 12, Inactive, 29, NCR, Skelagervej, 15, Skelagervej, 15, 9000, 57.023, 9.891, DKK, Visa Dankort, 3071, Withdrawal, --, 57.048, 9.919, 2624886, Aalborg, 271.150, 1028, 79, 3, 360, 0.00, 2801, Clouds, few clouds
2017, January, 16, Monday, 12, Active, 11, NCR, Sauer
```