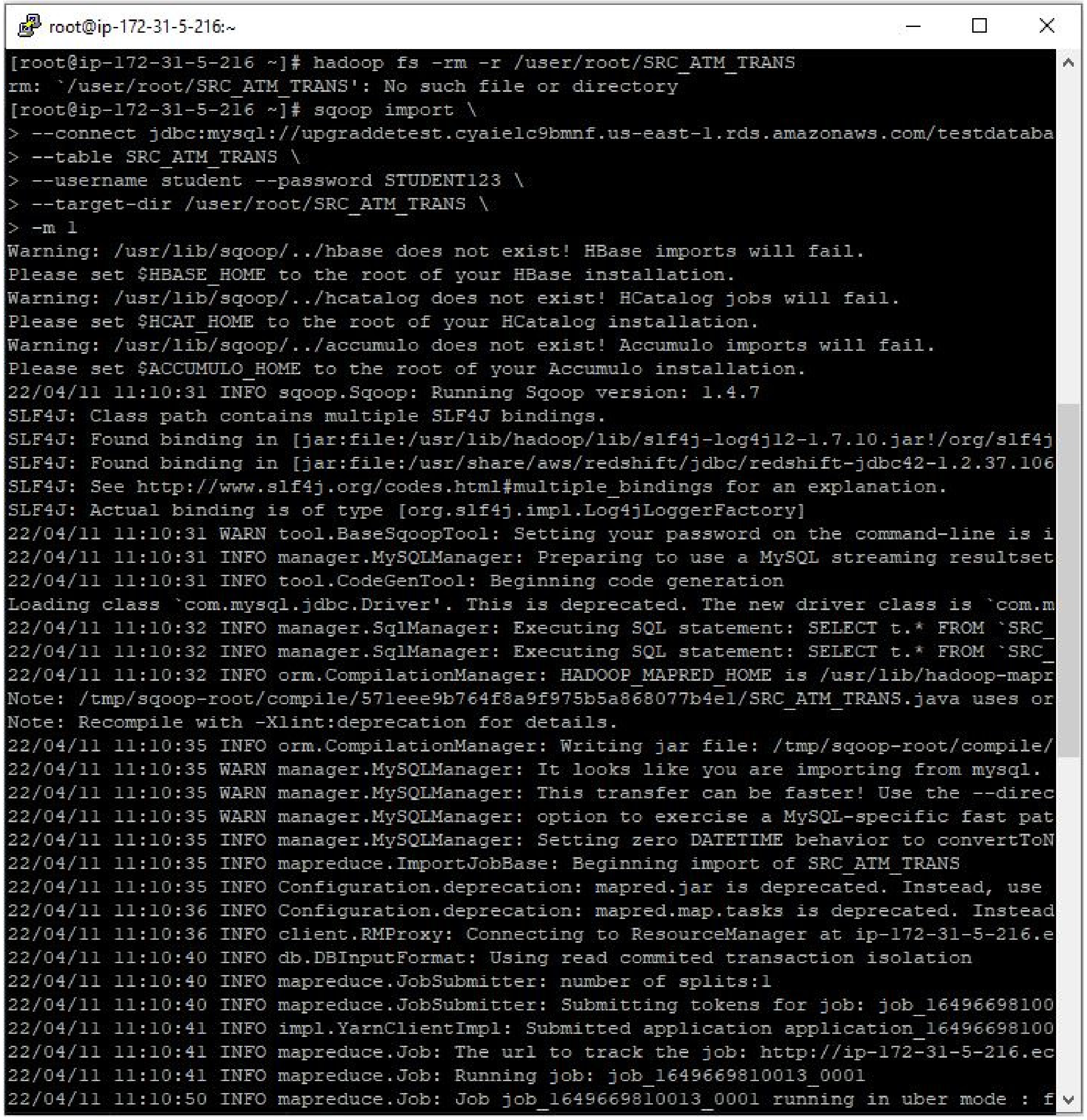
Data Ingestion from the RDS to HDFS using Sqoop

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

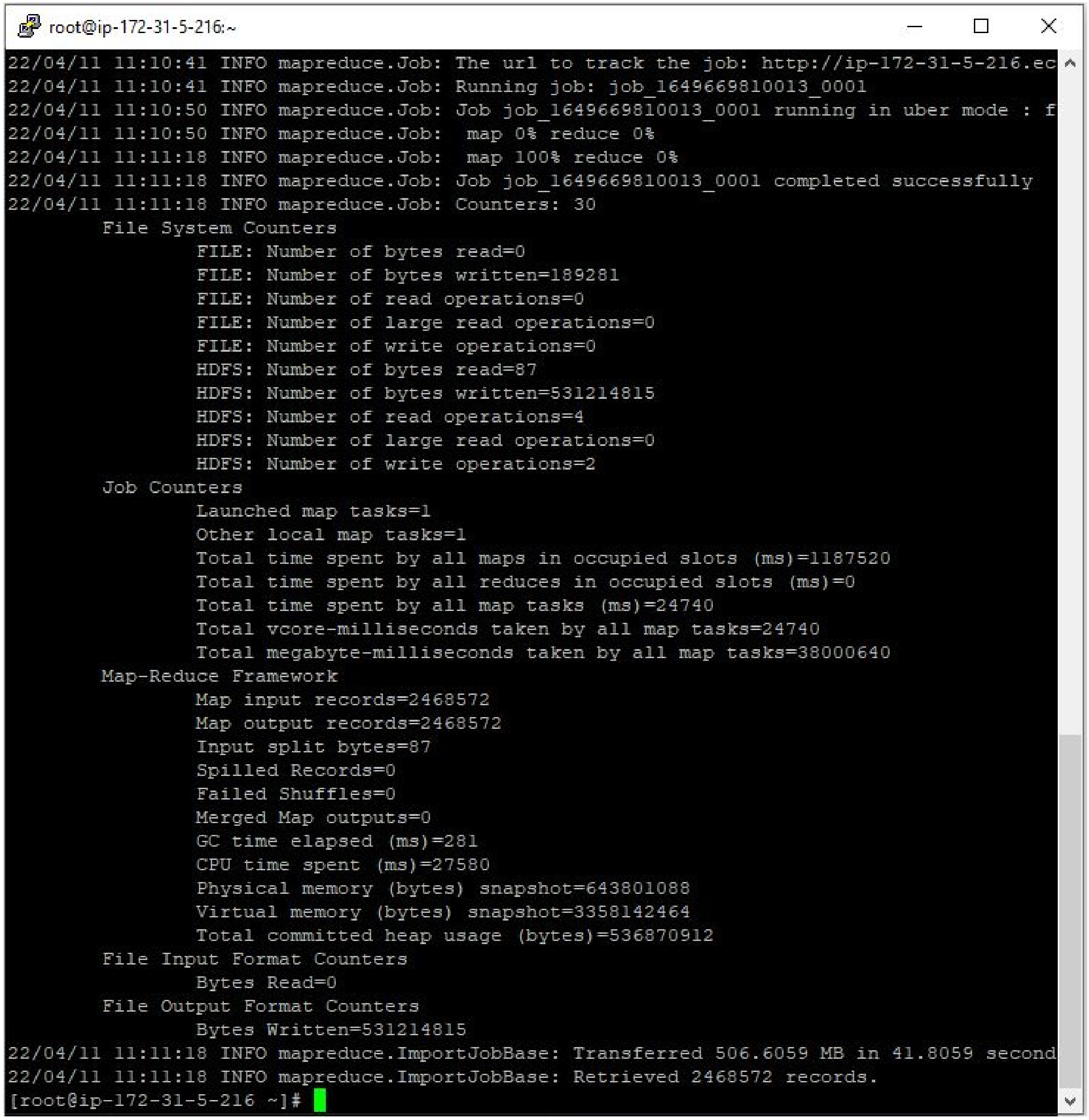
Sqoop Import Command:

```
sqoop import \
--connect jdbc:mysql://upgraddetest.cyaielc9bmnf.us-east-
1.rds.amazonaws.com/testdatabase \
--table SRC_ATM_TRANS \
--username student --password STUDENT123 \
--target-dir /user/root/SRC_ATM_TRANS \
-m 1
```

In the screenshot below, I can see that as a result of Sqoop Import Job, 2468572 records have been retrieved (same as the checkpoint mentioned in the Validation document)



```
[root@ip-172-31-5-216 ~]# hadoop fs -rm -r /user/root/SRC_ATM_TRANS
rm: `/user/root/SRC_ATM_TRANS': No such file or directory
[root@ip-172-31-5-216 ~]# sqoop import \
> --connect jdbc:mysql://upgraddetest.cyaielc9bmnf.us-east-1.rds.amazonaws.com/testdataba
> --table SRC_ATM_TRANS \
> --username student --password STUDENT123 \
> --target-dir /user/root/SRC_ATM_TRANS \
> -m 1
Warning: /usr/lib/sqoop/../hbase does not exist! HBase imports will fail.
Please set $HBASE_HOME to the root of your HBase installation.
Warning: /usr/lib/sqoop/../hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
22/04/11 11:10:31 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j
SLF4J: Found binding in [jar:file:/usr/share/aws/redshift/jdbc/redshift-jdbc42-1.2.37.106
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
22/04/11 11:10:31 WARN tool.BaseSqoopTool: Setting your password on the command-line is i
22/04/11 11:10:31 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset
22/04/11 11:10:31 INFO tool.CodeGenTool: Beginning code generation
Loading class `com.mysql.jdbc.Driver'. This is deprecated. The new driver class is `com.m
22/04/11 11:10:32 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `SRC_
22/04/11 11:10:32 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `SRC_
22/04/11 11:10:32 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapr
Note: /tmp/sqoop-root/compile/571eee9b764f8a9f975b5a868077b4e1/SRC_ATM_TRANS.java uses or
Note: Recompile with -Xlint:deprecation for details.
22/04/11 11:10:35 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-root/compile/
22/04/11 11:10:35 WARN manager.MySQLManager: It looks like you are importing from mysql.
22/04/11 11:10:35 WARN manager.MySQLManager: This transfer can be faster! Use the --direc
22/04/11 11:10:35 WARN manager.MySQLManager: option to exercise a MySQL-specific fast pat
22/04/11 11:10:35 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToN
22/04/11 11:10:35 INFO mapreduce.ImportJobBase: Beginning import of SRC_ATM_TRANS
22/04/11 11:10:35 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use
22/04/11 11:10:36 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead
22/04/11 11:10:36 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-5-216.e
22/04/11 11:10:40 INFO db.DBInputFormat: Using read commited transaction isolation
22/04/11 11:10:40 INFO mapreduce.JobSubmitter: number of splits:1
22/04/11 11:10:40 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_16496698100
22/04/11 11:10:41 INFO impl.YarnClientImpl: Submitted application application_16496698100
22/04/11 11:10:41 INFO mapreduce.Job: The url to track the job: http://ip-172-31-5-216.ec
22/04/11 11:10:41 INFO mapreduce.Job: Running job: job_1649669810013_0001
22/04/11 11:10:50 INFO mapreduce.Job: Job job_1649669810013_0001 running in uber mode : f
```



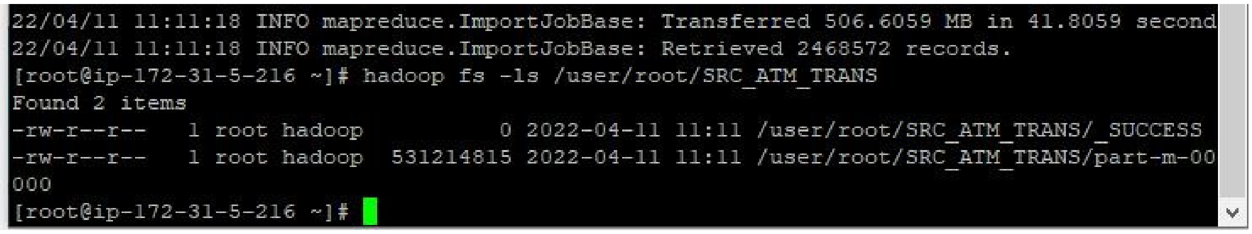
```
22/04/11 11:10:41 INFO mapreduce.Job: The url to track the job: http://ip-172-31-5-216.ec
22/04/11 11:10:41 INFO mapreduce.Job: Running job: job_1649669810013_0001
22/04/11 11:10:50 INFO mapreduce.Job: Job job_1649669810013_0001 running in uber mode : f
22/04/11 11:10:50 INFO mapreduce.Job:  map 0% reduce 0%
22/04/11 11:11:18 INFO mapreduce.Job:  map 100% reduce 0%
22/04/11 11:11:18 INFO mapreduce.Job: Job job_1649669810013_0001 completed successfully
22/04/11 11:11:18 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=189281
                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)=1187520
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=24740
                Total vcore-milliseconds taken by all map tasks=24740
                Total megabyte-milliseconds taken by all map tasks=38000640
        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)=281
                CPU time spent (ms)=27580
                Physical memory (bytes) snapshot=643801088
                Virtual memory (bytes) snapshot=3358142464
                Total committed heap usage (bytes)=536870912
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=531214815
22/04/11 11:11:18 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 41.8059 second
22/04/11 11:11:18 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[root@ip-172-31-5-216 ~]#
```

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

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

- The target directory contains 2 items:
    i) First file is the success file, indicating that the MapReduce job was successful.
    ii) Second file ‘part-m-00000’ contains all of the data imported.

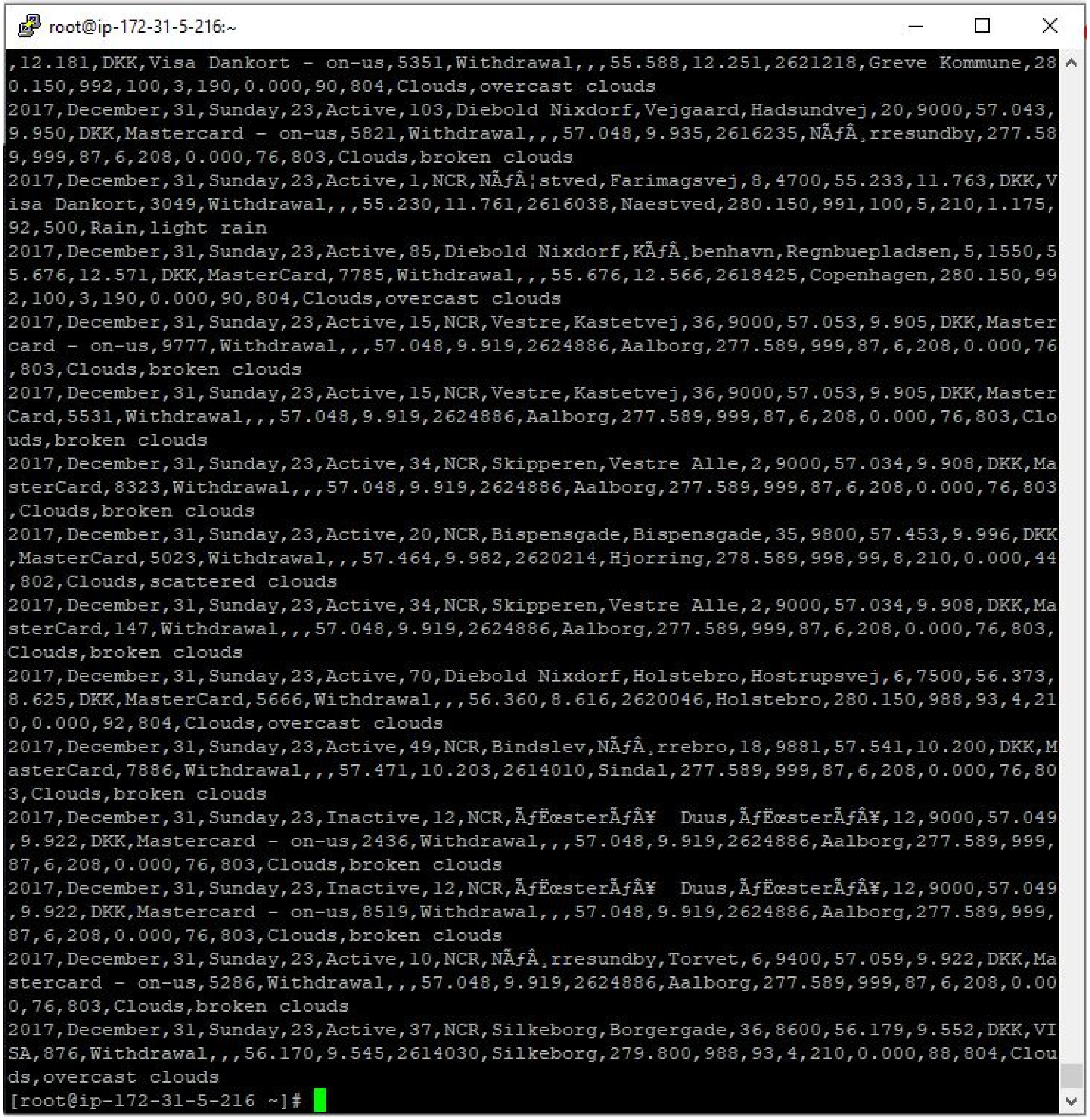
Screenshot of the imported data:



```
22/04/11 11:11:18 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 41.8059 second
22/04/11 11:11:18 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[root@ip-172-31-5-216 ~]# hadoop fs -ls /user/root/SRC_ATM_TRANS
Found 2 items
-rw-r--r--   1 root hadoop          0 2022-04-11 11:11 /user/root/SRC_ATM_TRANS/_SUCCESS
-rw-r--r--   1 root hadoop  531214815 2022-04-11 11:11 /user/root/SRC_ATM_TRANS/part-m-00
000
[root@ip-172-31-5-216 ~]# 
```

Checking the data in ‘part-m-00000’ file using the following command:

```
hadoop fs -cat /user/root/SRC_ATM_TRANS/part-m-00000
```



```
,12.181,DKK,Visa Dankort - on-us,5351,Withdrawal,,,55.588,12.251,2621218,Greve Kommune,28
0.150,992,100,3,190,0.000,90,804,Clouds,overcast clouds
2017,December,31,Sunday,23,Active,103,Diebold Nixdorf,Vejgaard,Hadsundvej,20,9000,57.043,
9.950,DKK,Mastercard - on-us,5821,Withdrawal,,,57.048,9.935,2616235,NÃƒÂ¸rresundby,277.58
9,999,87,6,208,0.000,76,803,Clouds,broken clouds
2017,December,31,Sunday,23,Active,1,NCR,NÃƒÂ¦stved,Farimagsvej,8,4700,55.233,11.763,DKK,V
isa Dankort,3049,Withdrawal,,,55.230,11.761,2616038,Naestved,280.150,991,100,5,210,1.175,
92,500,Rain,light rain
2017,December,31,Sunday,23,Active,85,Diebold Nixdorf,KÃƒÂ¸benhavn,Regnbuepladsen,5,1550,5
5.676,12.571,DKK,MasterCard,7785,Withdrawal,,,55.676,12.566,2618425,Copenhagen,280.150,99
2,100,3,190,0.000,90,804,Clouds,overcast clouds
2017,December,31,Sunday,23,Active,15,NCR,Vestre,Kastetvej,36,9000,57.053,9.905,DKK,Master
card - on-us,9777,Withdrawal,,,57.048,9.919,2624886,Aalborg,277.589,999,87,6,208,0.000,76
,803,Clouds,broken clouds
2017,December,31,Sunday,23,Active,15,NCR,Vestre,Kastetvej,36,9000,57.053,9.905,DKK,Master
Card,5531,Withdrawal,,,57.048,9.919,2624886,Aalborg,277.589,999,87,6,208,0.000,76,803,Clo
uds,broken clouds
2017,December,31,Sunday,23,Active,34,NCR,Skipperen,Vestre Alle,2,9000,57.034,9.908,DKK,Ma
sterCard,8323,Withdrawal,,,57.048,9.919,2624886,Aalborg,277.589,999,87,6,208,0.000,76,803
,Clouds,broken clouds
2017,December,31,Sunday,23,Active,20,NCR,Bispensgade,Bispensgade,35,9800,57.453,9.996,DKK
,MasterCard,5023,Withdrawal,,,57.464,9.982,2620214,Hjorring,278.589,998,99,8,210,0.000,44
,802,Clouds,scattered clouds
2017,December,31,Sunday,23,Active,34,NCR,Skipperen,Vestre Alle,2,9000,57.034,9.908,DKK,Ma
sterCard,147,Withdrawal,,,57.048,9.919,2624886,Aalborg,277.589,999,87,6,208,0.000,76,803,
Clouds,broken clouds
2017,December,31,Sunday,23,Active,70,Diebold Nixdorf,Holstebro,Hostrupsvej,6,7500,56.373,
8.625,DKK,MasterCard,5666,Withdrawal,,,56.360,8.616,2620046,Holstebro,280.150,988,93,4,21
0,0.000,92,804,Clouds,overcast clouds
2017,December,31,Sunday,23,Active,49,NCR,Bindslev,NÃƒÂ¸rrebro,18,9881,57.541,10.200,DKK,M
asterCard,7886,Withdrawal,,,57.471,10.203,2614010,Sindal,277.589,999,87,6,208,0.000,76,80
3,Clouds,broken clouds
2017,December,31,Sunday,23,Inactive,12,NCR,ÃƒËœsterÃƒÂ¥  Duus,ÃƒËœsterÃƒÂ¥,12,9000,57.049
,9.922,DKK,Mastercard - on-us,2436,Withdrawal,,,57.048,9.919,2624886,Aalborg,277.589,999,
87,6,208,0.000,76,803,Clouds,broken clouds
2017,December,31,Sunday,23,Inactive,12,NCR,ÃƒËœsterÃƒÂ¥  Duus,ÃƒËœsterÃƒÂ¥,12,9000,57.049
,9.922,DKK,Mastercard - on-us,8519,Withdrawal,,,57.048,9.919,2624886,Aalborg,277.589,999,
87,6,208,0.000,76,803,Clouds,broken clouds
2017,December,31,Sunday,23,Active,10,NCR,NÃƒÂ¸rresundby,Torvet,6,9400,57.059,9.922,DKK,Ma
stercard - on-us,5286,Withdrawal,,,57.048,9.919,2624886,Aalborg,277.589,999,87,6,208,0.00
0,76,803,Clouds,broken clouds
2017,December,31,Sunday,23,Active,37,NCR,Silkeborg,Borgergade,36,8600,56.179,9.552,DKK,VI
SA,876,Withdrawal,,,56.170,9.545,2614030,Silkeborg,279.800,988,93,4,210,0.000,88,804,Clou
ds,overcast clouds
[root@ip-172-31-5-216 ~]#
```