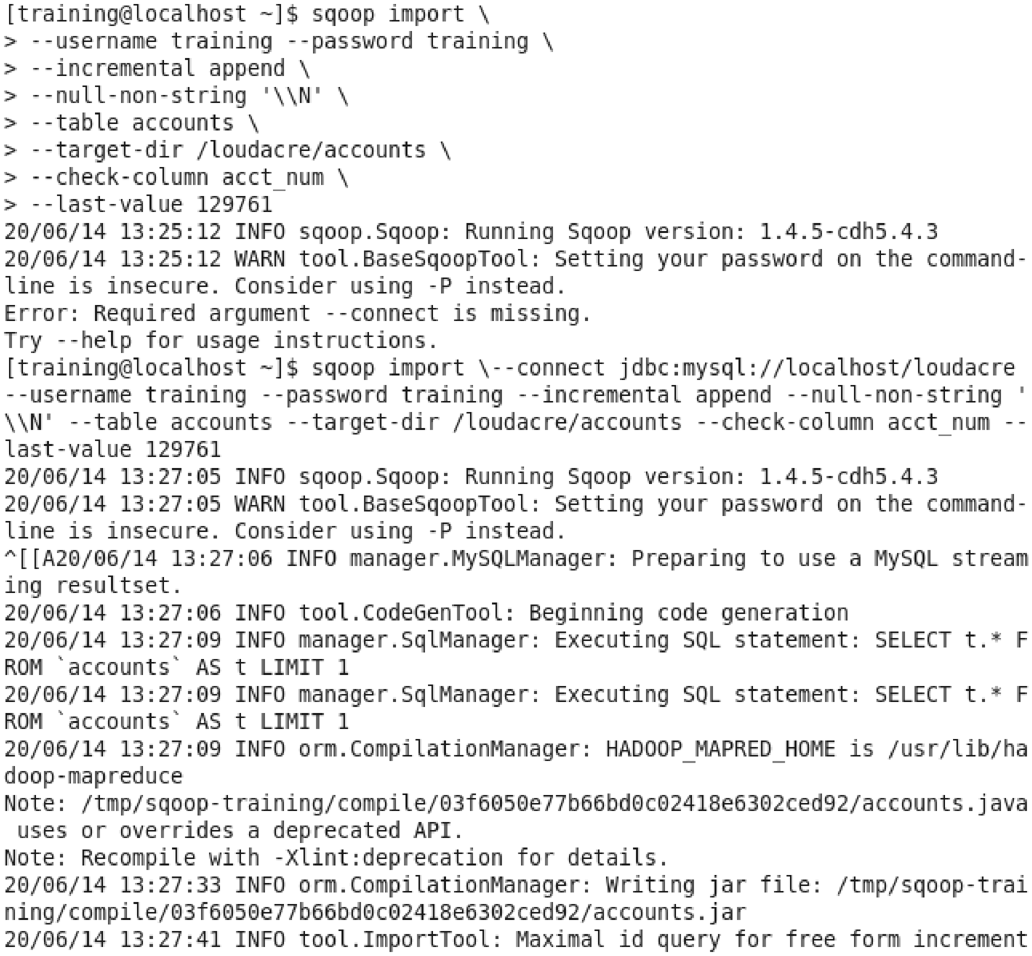
BUAN 6346 Big Data

Homework 4-A

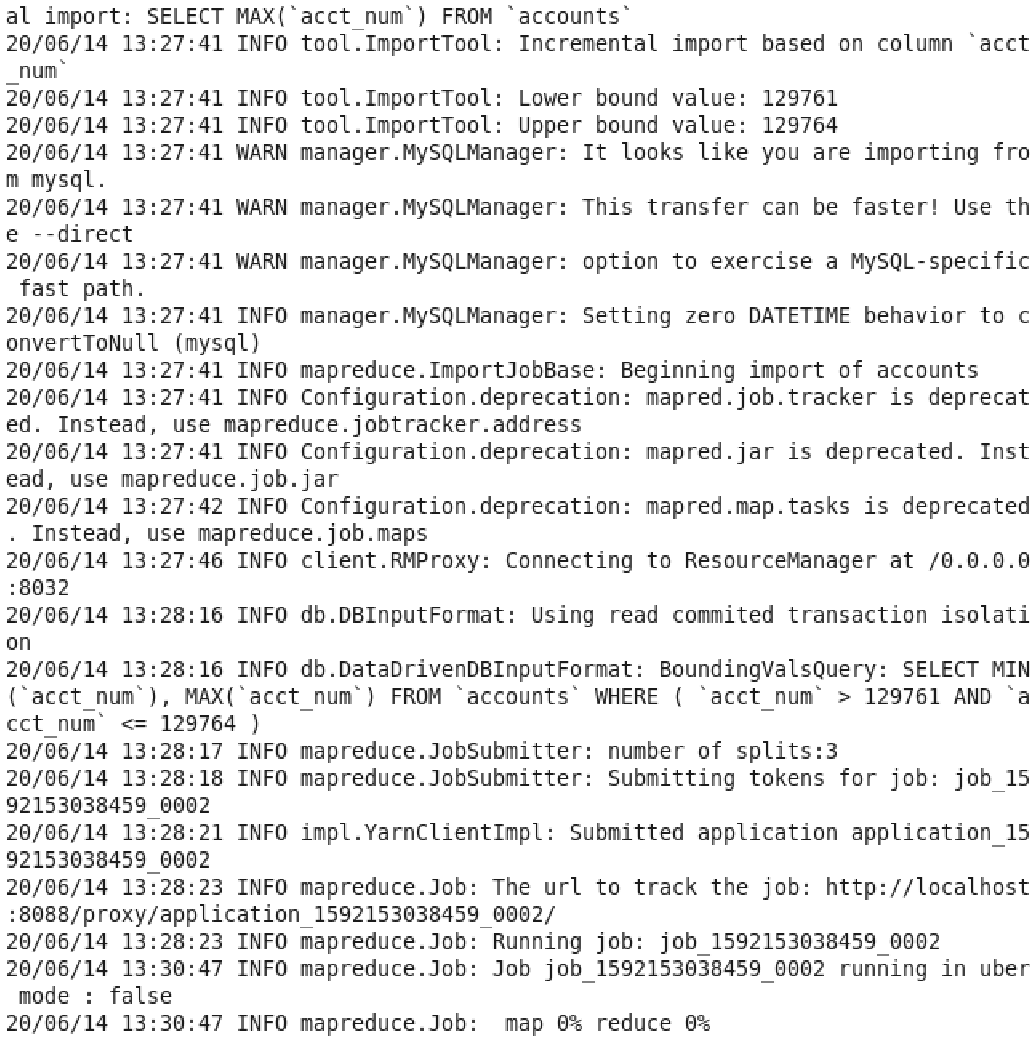
CAP – Developing with Spark and Hadoop

Step11: $sqoop import \

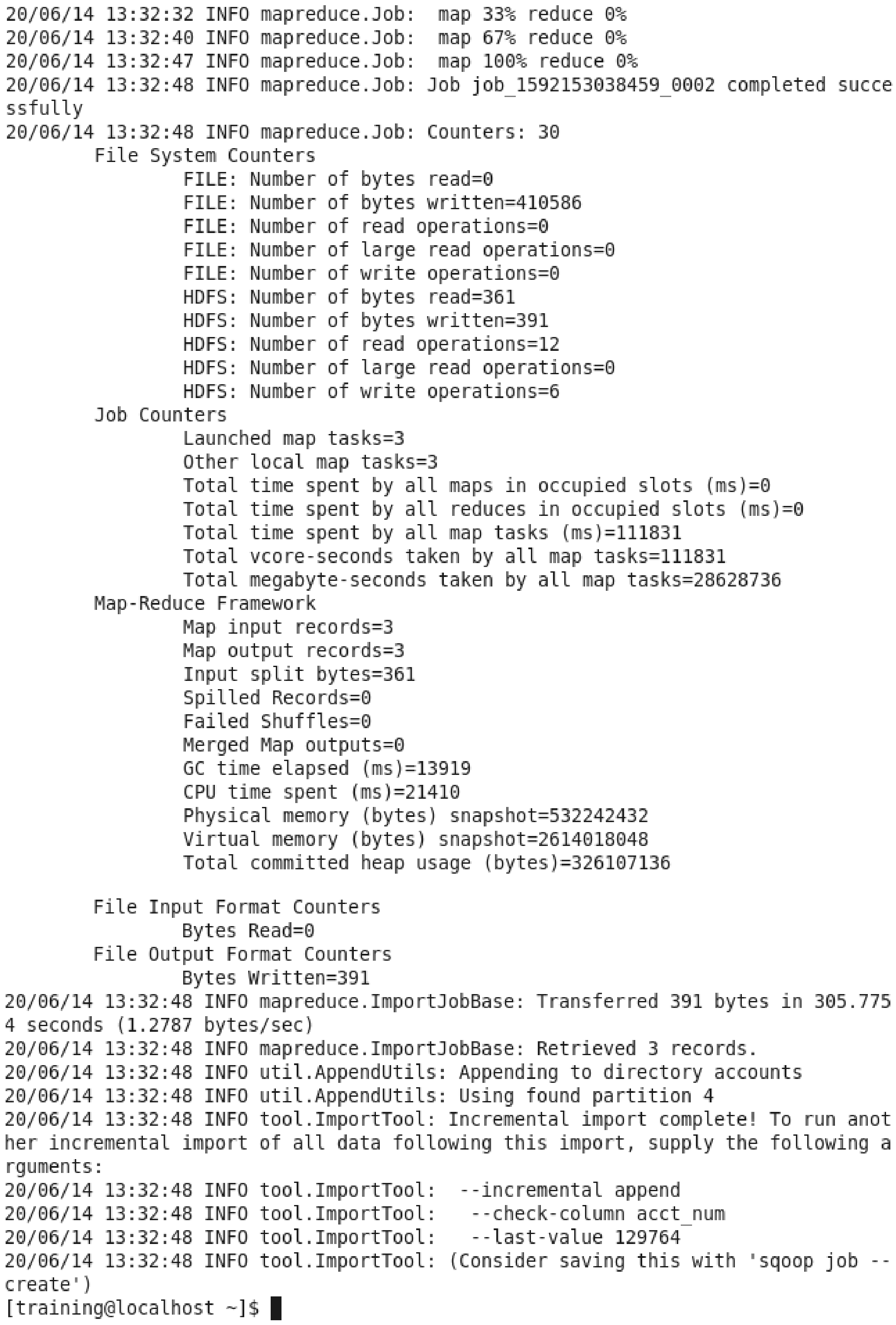
```
--connect jdbc:mysql://localhost/loudacre \
--username training --password training \
--incremental append \
--null-non-string '\\N' \
--table accounts \
--target-dir /loudacre/accounts \
--check-column acct_num \
--last-value <largest_acct_num>
```



```
[training@localhost ~]$ sqoop import \
> --username training --password training \
> --incremental append \
> --null-non-string '\\N' \
> --table accounts \
> --target-dir /loudacre/accounts \
> --check-column acct_num \
> --last-value 129761
20/06/14 13:25:12 INFO sqoop.Sqoop: Running Sqoop version: 1.4.5-cdh5.4.3
20/06/14 13:25:12 WARN tool.BaseSqoopTool: Setting your password on the command-
line is insecure. Consider using -P instead.
Error: Required argument --connect is missing.
Try --help for usage instructions.
[training@localhost ~]$ sqoop import \--connect jdbc:mysql://localhost/loudacre
--username training --password training --incremental append --null-non-string '
\\N' --table accounts --target-dir /loudacre/accounts --check-column acct_num --
last-value 129761
20/06/14 13:27:05 INFO sqoop.Sqoop: Running Sqoop version: 1.4.5-cdh5.4.3
20/06/14 13:27:05 WARN tool.BaseSqoopTool: Setting your password on the command-
line is insecure. Consider using -P instead.
^[[A20/06/14 13:27:06 INFO manager.MySQLManager: Preparing to use a MySQL stream
ing resultset.
20/06/14 13:27:06 INFO tool.CodeGenTool: Beginning code generation
20/06/14 13:27:09 INFO manager.SqlManager: Executing SQL statement: SELECT t.* F
ROM `accounts` AS t LIMIT 1
20/06/14 13:27:09 INFO manager.SqlManager: Executing SQL statement: SELECT t.* F
ROM `accounts` AS t LIMIT 1
20/06/14 13:27:09 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/ha
doop-mapreduce
Note: /tmp/sqoop-training/compile/03f6050e77b66bd0c02418e6302ced92/accounts.java
 uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
20/06/14 13:27:33 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-trai
ning/compile/03f6050e77b66bd0c02418e6302ced92/accounts.jar
20/06/14 13:27:41 INFO tool.ImportTool: Maximal id query for free form increment
```

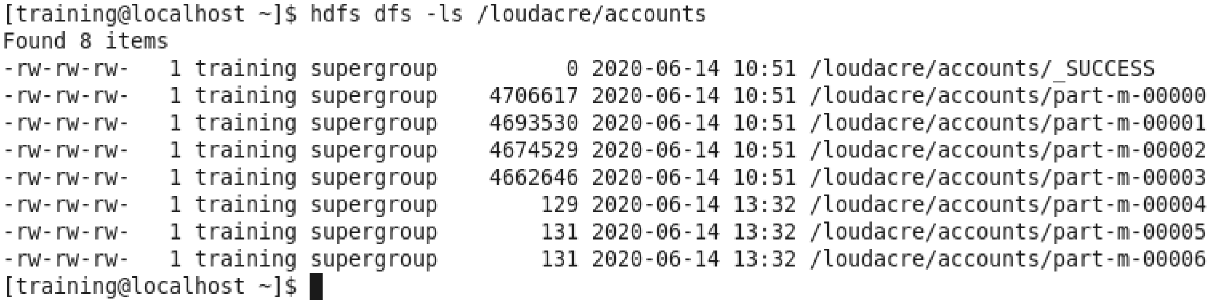


```
al import: SELECT MAX(`acct_num`) FROM `accounts`
20/06/14 13:27:41 INFO tool.ImportTool: Incremental import based on column `acct
_num`
20/06/14 13:27:41 INFO tool.ImportTool: Lower bound value: 129761
20/06/14 13:27:41 INFO tool.ImportTool: Upper bound value: 129764
20/06/14 13:27:41 WARN manager.MySQLManager: It looks like you are importing fro
m mysql.
20/06/14 13:27:41 WARN manager.MySQLManager: This transfer can be faster! Use th
e --direct
20/06/14 13:27:41 WARN manager.MySQLManager: option to exercise a MySQL-specific
 fast path.
20/06/14 13:27:41 INFO manager.MySQLManager: Setting zero DATETIME behavior to c
onvertToNull (mysql)
20/06/14 13:27:41 INFO mapreduce.ImportJobBase: Beginning import of accounts
20/06/14 13:27:41 INFO Configuration.deprecation: mapred.job.tracker is deprecat
ed. Instead, use mapreduce.jobtracker.address
20/06/14 13:27:41 INFO Configuration.deprecation: mapred.jar is deprecated. Inst
ead, use mapreduce.job.jar
20/06/14 13:27:42 INFO Configuration.deprecation: mapred.map.tasks is deprecated
. Instead, use mapreduce.job.maps
20/06/14 13:27:46 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0
:8032
20/06/14 13:28:16 INFO db.DBInputFormat: Using read commited transaction isolati
on
20/06/14 13:28:16 INFO db.DataDrivenDBInputFormat: BoundingValsQuery: SELECT MIN
(`acct_num`), MAX(`acct_num`) FROM `accounts` WHERE ( `acct_num` > 129761 AND `a
cct_num` <= 129764 )
20/06/14 13:28:17 INFO mapreduce.JobSubmitter: number of splits:3
20/06/14 13:28:18 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_15
92153038459_0002
20/06/14 13:28:21 INFO impl.YarnClientImpl: Submitted application application_15
92153038459_0002
20/06/14 13:28:23 INFO mapreduce.Job: The url to track the job: http://localhost
:8088/proxy/application_1592153038459_0002/
20/06/14 13:28:23 INFO mapreduce.Job: Running job: job_1592153038459_0002
20/06/14 13:30:47 INFO mapreduce.Job: Job job_1592153038459_0002 running in uber
 mode : false
20/06/14 13:30:47 INFO mapreduce.Job:  map 0% reduce 0%
```



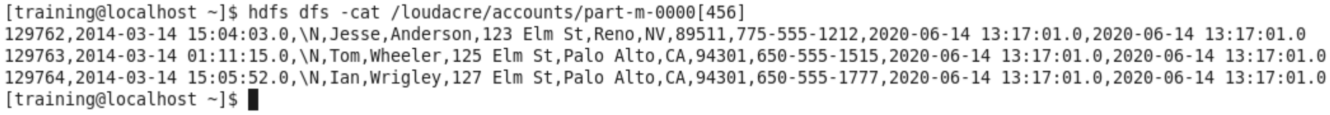
```
20/06/14 13:32:32 INFO mapreduce.Job:  map 33% reduce 0%
20/06/14 13:32:40 INFO mapreduce.Job:  map 67% reduce 0%
20/06/14 13:32:47 INFO mapreduce.Job:  map 100% reduce 0%
20/06/14 13:32:48 INFO mapreduce.Job: Job job_1592153038459_0002 completed succe
ssfully
20/06/14 13:32:48 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=410586
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=361
                HDFS: Number of bytes written=391
                HDFS: Number of read operations=12
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=6
        Job Counters
                Launched map tasks=3
                Other local map tasks=3
                Total time spent by all maps in occupied slots (ms)=0
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=111831
                Total vcore-seconds taken by all map tasks=111831
                Total megabyte-seconds taken by all map tasks=28628736
        Map-Reduce Framework
                Map input records=3
                Map output records=3
                Input split bytes=361
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=13919
                CPU time spent (ms)=21410
                Physical memory (bytes) snapshot=532242432
                Virtual memory (bytes) snapshot=2614018048
                Total committed heap usage (bytes)=326107136

        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=391
20/06/14 13:32:48 INFO mapreduce.ImportJobBase: Transferred 391 bytes in 305.775
4 seconds (1.2787 bytes/sec)
20/06/14 13:32:48 INFO mapreduce.ImportJobBase: Retrieved 3 records.
20/06/14 13:32:48 INFO util.AppendUtils: Appending to directory accounts
20/06/14 13:32:48 INFO util.AppendUtils: Using found partition 4
20/06/14 13:32:48 INFO tool.ImportTool: Incremental import complete! To run anot
her incremental import of all data following this import, supply the following a
rguments:
20/06/14 13:32:48 INFO tool.ImportTool:  --incremental append
20/06/14 13:32:48 INFO tool.ImportTool:   --check-column acct_num
20/06/14 13:32:48 INFO tool.ImportTool:   --last-value 129764
20/06/14 13:32:48 INFO tool.ImportTool: (Consider saving this with 'sqoop job --
create')
[training@localhost ~]$ █
```

Step12: $ `hdfs dfs -ls /loudacre/accounts`



```
[training@localhost ~]$ hdfs dfs -ls /loudacre/accounts
Found 8 items
-rw-rw-rw-   1 training supergroup          0 2020-06-14 10:51 /loudacre/accounts/_SUCCESS
-rw-rw-rw-   1 training supergroup    4706617 2020-06-14 10:51 /loudacre/accounts/part-m-00000
-rw-rw-rw-   1 training supergroup    4693530 2020-06-14 10:51 /loudacre/accounts/part-m-00001
-rw-rw-rw-   1 training supergroup    4674529 2020-06-14 10:51 /loudacre/accounts/part-m-00002
-rw-rw-rw-   1 training supergroup    4662646 2020-06-14 10:51 /loudacre/accounts/part-m-00003
-rw-rw-rw-   1 training supergroup        129 2020-06-14 13:32 /loudacre/accounts/part-m-00004
-rw-rw-rw-   1 training supergroup        131 2020-06-14 13:32 /loudacre/accounts/part-m-00005
-rw-rw-rw-   1 training supergroup        131 2020-06-14 13:32 /loudacre/accounts/part-m-00006
[training@localhost ~]$ 
```

Step13: $ `hdfs dfs -cat /loudacre/accounts/part-m-0000[456]`



```
[training@localhost ~]$ hdfs dfs -cat /loudacre/accounts/part-m-0000[456]
129762,2014-03-14 15:04:03.0,\N,Jesse,Anderson,123 Elm St,Reno,NV,89511,775-555-1212,2020-06-14 13:17:01.0,2020-06-14 13:17:01.0
129763,2014-03-14 01:11:15.0,\N,Tom,Wheeler,125 Elm St,Palo Alto,CA,94301,650-555-1515,2020-06-14 13:17:01.0,2020-06-14 13:17:01.0
129764,2014-03-14 15:05:52.0,\N,Ian,Wrigley,127 Elm St,Palo Alto,CA,94301,650-555-1777,2020-06-14 13:17:01.0,2020-06-14 13:17:01.0
[training@localhost ~]$ 
```