

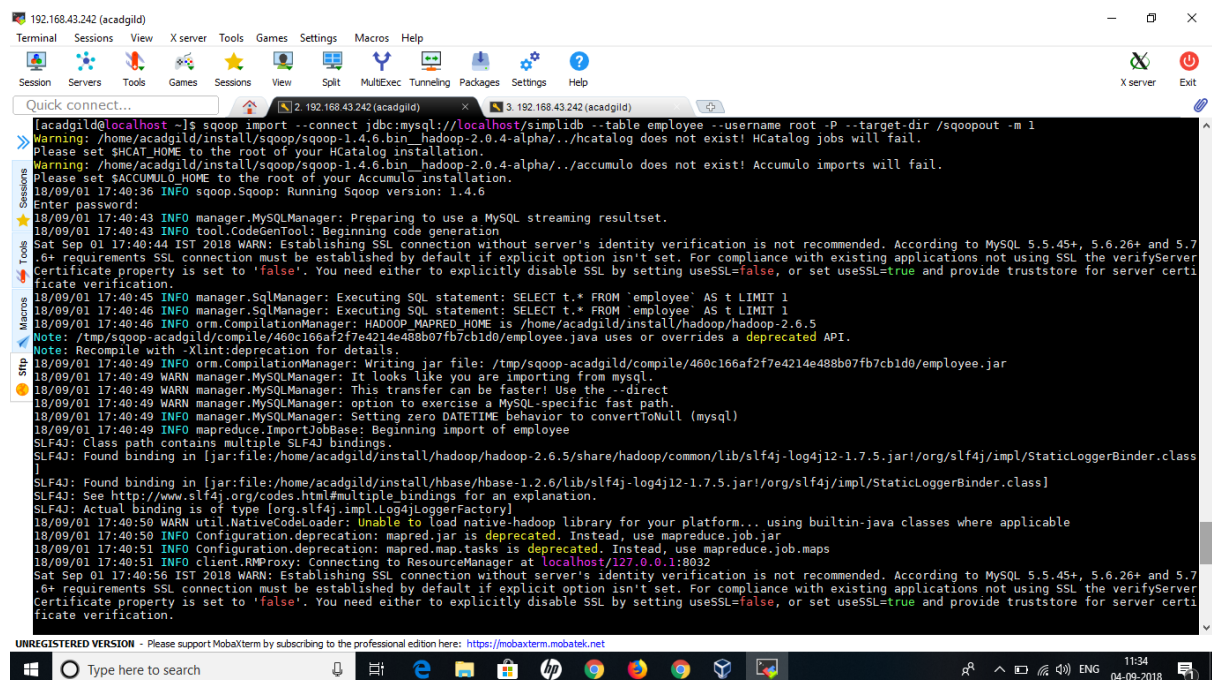
## Assignment No. 6.1

**Task 1:** Use Sqoop tool to import the data from Mysql to HDFS as Sqoopout folder.

**Solution:** We have used the sqoop tool to import the employee table from Mysql to Sqoopout folder of HDFS.

**Command:** `sqoop import --connect jdbc:mysql://localhost/simplidb --table employee --username root -P --target-dir /sqoopout -m 1`

**Command Execution:**

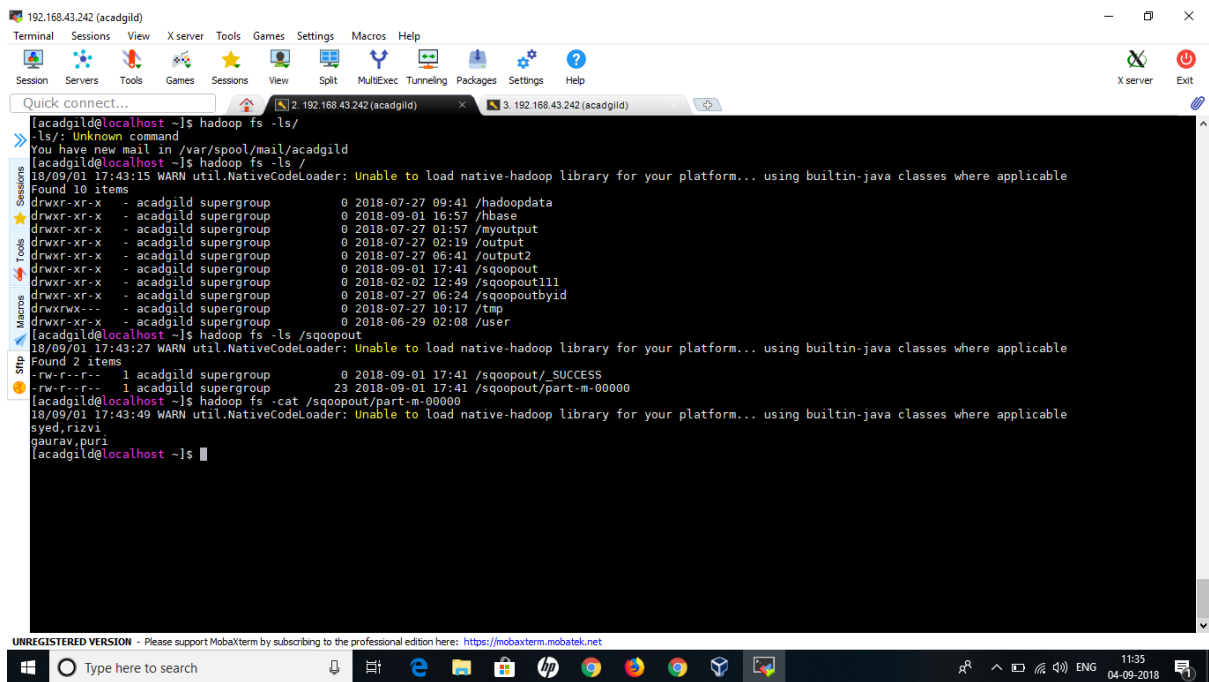


```
192.168.43.242 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect...
[acadgild@localhost ~]$ sqoop import --connect jdbc:mysql://localhost/simplidb --table employee --username root -P --target-dir /sqoopout -m 1
Warning: /home/acadgild/install/sqoop/sqoop-1.4.6.bin_hadoop-2.0.4-alpha/./hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /home/acadgild/install/sqoop/sqoop-1.4.6.bin_hadoop-2.0.4-alpha/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
18/09/01 17:40:36 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6
Enter password:
18/09/01 17:40:43 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
18/09/01 17:40:43 INFO tool.CodeGenTool: Beginning code generation
Sat Sep 01 17:40:44 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7
+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServer
Certificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certi
ficate verification.
18/09/01 17:40:45 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'employee' AS t LIMIT 1
18/09/01 17:40:46 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'employee' AS t LIMIT 1
18/09/01 17:40:46 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /home/acadgild/install/hadoop/hadoop-2.6.5
Note: /tmp/sqoop-acadgild/compile/460c166af2f7e4214e488b07fb7cb1d0/employee.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
18/09/01 17:40:49 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-acadgild/compile/460c166af2f7e4214e488b07fb7cb1d0/employee.jar
18/09/01 17:40:49 WARN manager.MySQLManager: It looks like you are importing from mysql.
18/09/01 17:40:49 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
18/09/01 17:40:49 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
18/09/01 17:40:49 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
18/09/01 17:40:49 INFO mapreduce.ImportJobBase: Beginning import of employee
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class
]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: see http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
18/09/01 17:40:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/09/01 17:40:50 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
18/09/01 17:40:51 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
18/09/01 17:40:51 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
Sat Sep 01 17:40:56 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7
+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServer
Certificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certi
ficate verification.
```

```
192.168.43.242 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect...
2. 192.168.43.242 (acadgild) 3. 192.168.43.242 (acadgild)
18/09/01 17:40:57 INFO db.DBInputFormat: Using read committed transaction isolation
18/09/01 17:40:57 INFO mapreduce.JobSubmitter: number of splits=1
18/09/01 17:40:57 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1535803021312_0001
18/09/01 17:40:58 INFO impl.YarnClientImpl: Submitted application application_1535803021312_0001
18/09/01 17:40:58 INFO mapreduce.Job: The url to track the job: http://localhost:8080/proxy/application_1535803021312_0001/
18/09/01 17:40:58 INFO mapreduce.Job: Running job: job_1535803021312_0001
18/09/01 17:41:19 INFO mapreduce.Job: Job job_1535803021312_0001 running in uber mode : false
18/09/01 17:41:19 INFO mapreduce.Job: map 0% reduce 0%
18/09/01 17:41:32 INFO mapreduce.Job: map 100% reduce 0%
18/09/01 17:41:32 INFO mapreduce.Job: Job job_1535803021312_0001 completed successfully
18/09/01 17:41:33 INFO mapreduce.Job: Counters: 30
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=127705
  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=23
  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)=9212
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=9212
  Total vcore-milliseconds taken by all map tasks=9212
  Total megabyte-milliseconds taken by all map tasks=9433088
Map-Reduce Framework
  Map input records=2
  Map output records=2
  Input split bytes=87
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=86
UNREGISTERED VERSION - Please support Mobaxterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net
Type here to search
11:34 04-09-2018
```

```
192.168.43.242 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect...
2. 192.168.43.242 (acadgild) 3. 192.168.43.242 (acadgild)
FILE: Number of bytes read=0
FILE: Number of bytes written=127705
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=23
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)=9212
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=9212
  Total vcore-milliseconds taken by all map tasks=9212
  Total megabyte-milliseconds taken by all map tasks=9433088
Map-Reduce Framework
  Map input records=2
  Map output records=2
  Input split bytes=87
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=86
  CPU time spent (ms)=2470
  Physical memory (bytes) snapshot=112824320
  Virtual memory (bytes) snapshot=2069131264
  Total committed heap usage (bytes)=48758784
File Input Format Counters
  Bytes Read=0
  Bytes Written=23
18/09/01 17:41:33 INFO mapreduce.ImportJobBase: Transferred 23 bytes in 41.3659 seconds (0.556 bytes/sec)
18/09/01 17:41:33 INFO mapreduce.ImportJobBase: Retrieved 2 records.
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
UNREGISTERED VERSION - Please support Mobaxterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net
Type here to search
11:34 04-09-2018
```

Output:



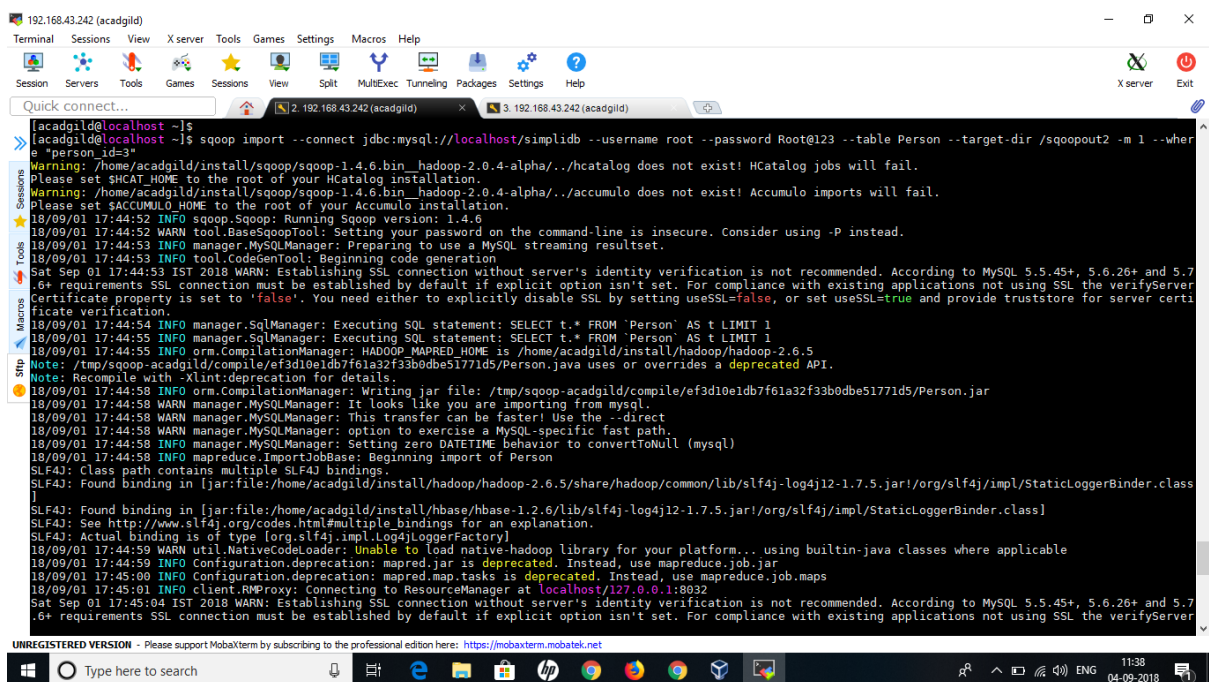
```
[acadgild@localhost ~]$ hadoop fs -ls /
ls: Unknown command
[acadgild@localhost ~]$ hadoop fs -ls /
18/09/01 17:43:15 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 10 items
drwxr-xr-x - acadgild supergroup 0 2018-07-27 09:41 /hbase
drwxr-xr-x - acadgild supergroup 0 2018-09-01 16:57 /hbase
drwxr-xr-x - acadgild supergroup 0 2018-07-27 01:57 /myoutput
drwxr-xr-x - acadgild supergroup 0 2018-07-27 02:19 /output
drwxr-xr-x - acadgild supergroup 0 2018-07-27 06:41 /output2
drwxr-xr-x - acadgild supergroup 0 2018-09-01 17:41 /sqoopout
drwxr-xr-x - acadgild supergroup 0 2018-02-02 12:49 /sqoopout111
drwxr-xr-x - acadgild supergroup 0 2018-07-27 06:24 /sqoopoutbyid
drwxr-xr-x - acadgild supergroup 0 2018-07-27 10:17 /tmp
drwxr-xr-x - acadgild supergroup 0 2018-06-29 02:08 /user
[acadgild@localhost ~]$ hadoop fs -ls /sqoopout
18/09/01 17:43:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup 0 2018-09-01 17:41 /sqoopout/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 23 2018-09-01 17:41 /sqoopout/part-m-00000
[acadgild@localhost ~]$ hadoop fs -cat /sqoopout/part-m-00000
syed,rizvi
gaurav,pur
[acadgild@localhost ~]$
```

**Task 2:** Use Sqoop tool to import the data from Person table of Mysql to HDFS as Sqoopout2 folder where column id = 3.

**Solution:** We have used the sqoop tool to import the Person table from Mysql to Sqoopout2 folder of HDFS where column id = 3.

**Command:** `sqoop import --connect jdbc:mysql://localhost/simplidb --username root --password Root@123 --table Person --target-dir /sqoopout2 -m 1 --where "person_id=3"`

### Command Execution:



```
[acadgild@localhost ~]$ sqoop import --connect jdbc:mysql://localhost/simplidb --username root --password Root@123 --table Person --target-dir /sqoopout2 -m 1 --where "person_id=3"
Warning: /home/acadgild/install/sqoop/sqoop-1.4.6.bin_hadoop-2.0.4-alpha/./hcatalog does not exist! HCatalog jobs will fail.
Please set HCAT_HOME to the root of your HCatalog installation.
Warning: /home/acadgild/install/sqoop/sqoop-1.4.6.bin_hadoop-2.0.4-alpha/./accumulo does not exist! Accumulo imports will fail.
Please set ACCUMULO_HOME to the root of your Accumulo installation.
18/09/01 17:44:52 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6
18/09/01 17:44:52 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
18/09/01 17:44:53 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
18/09/01 17:44:53 INFO tool.CodeGenTool: Beginning code generation
Sat Sep 01 17:44:53 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.16+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.
18/09/01 17:44:54 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'Person' AS t LIMIT 1
18/09/01 17:44:55 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /home/acadgild/install/hadoop/hadoop-2.6.5
Note: /tmp/sqoop-acadgild/compile/ef3d10e1db7f61a32f33b0dbe51771d5/Person.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
18/09/01 17:44:58 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-acadgild/compile/ef3d10e1db7f61a32f33b0dbe51771d5/Person.jar
18/09/01 17:44:58 WARN manager.MySQLManager: It looks like you are importing from mysql.
18/09/01 17:44:58 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
18/09/01 17:44:58 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
18/09/01 17:44:58 INFO mapreduce.ImportJobBase: Beginning import of Person
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
18/09/01 17:44:59 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/09/01 17:44:59 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
18/09/01 17:45:00 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
18/09/01 17:45:01 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
Sat Sep 01 17:45:04 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.16+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServer
```

```
192.168.43.242 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect...
Certificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certi
18/09/01 17:45:04 INFO db.DBInputFormat: Using read committed transaction isolation
18/09/01 17:45:04 INFO mapreduce.JobSubmitter: number of splits:1
18/09/01 17:45:05 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1535803021312_0002
18/09/01 17:45:06 INFO impl.YarnClientImpl: Submitted application application_1535803021312_0002
18/09/01 17:45:06 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1535803021312_0002/
18/09/01 17:45:06 INFO mapreduce.Job: Running job: job_1535803021312_0002
18/09/01 17:45:21 INFO mapreduce.Job: Job job_1535803021312_0002 running in uber mode : false
18/09/01 17:45:21 INFO mapreduce.Job: map 0% reduce 0%
18/09/01 17:45:33 INFO mapreduce.Job: map 100% reduce 0%
18/09/01 17:45:34 INFO mapreduce.Job: Job job_1535803021312_0002 completed successfully
18/09/01 17:45:34 INFO mapreduce.Job: Counters: 30
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=128016
  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=32
  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)=8776
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=8776
  Total vcore-milliseconds taken by all map tasks=8776
  Total megabyte-milliseconds taken by all map tasks=8986624
Map-Reduce Framework
  Map input records=1
  Map output records=1
  Input split bytes=87
  Spilled Records=0
  Failed Shuffles=0
UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net
```

```
192.168.43.242 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect...
FILE: Number of bytes read=0
FILE: Number of bytes written=128016
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=32
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)=8776
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=8776
  Total vcore-milliseconds taken by all map tasks=8776
  Total megabyte-milliseconds taken by all map tasks=8986624
Map-Reduce Framework
  Map input records=1
  Map output records=1
  Input split bytes=87
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=93
  CPU time spent (ms)=2420
  Physical memory (bytes) snapshot=113483776
  Virtual memory (bytes) snapshot=2068512768
  Total committed heap usage (bytes)=48758784
File Input Format Counters
  Bytes Read=0
  Bytes Written=32
18/09/01 17:45:34 INFO mapreduce.ImportJobBase: Transferred 32 bytes in 33.4781 seconds (0.9558 bytes/sec)
18/09/01 17:45:34 INFO mapreduce.ImportJobBase: Retrieved 1 records.
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

Output:

192.168.43.242 (acadgild)

Terminal Sessions View X server Tools Games Settings Macros Help

Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help

Quick connect...

2. 192.168.43.242 (acadgild) 3. 192.168.43.242 (acadgild)

```
[acadgild@localhost ~]$ hadoop fs -ls /
18/09/01 17:46:47 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 11 items
drwxr-xr-x - acadgild supergroup      0 2018-07-27 09:41 /hadoopdata
drwxr-xr-x - acadgild supergroup      0 2018-09-01 16:57 /hbase
drwxr-xr-x - acadgild supergroup      0 2018-07-27 01:57 /myoutput
drwxr-xr-x - acadgild supergroup      0 2018-07-27 02:19 /output
drwxr-xr-x - acadgild supergroup      0 2018-07-27 06:41 /output2
drwxr-xr-x - acadgild supergroup      0 2018-09-01 17:41 /sqoopout
drwxr-xr-x - acadgild supergroup      0 2018-02-02 12:49 /sqoopout111
drwxr-xr-x - acadgild supergroup      0 2018-09-01 17:45 /sqoopout2
drwxr-xr-x - acadgild supergroup      0 2018-07-27 06:24 /sqoopoutbyid
drwxrwx--- - acadgild supergroup      0 2018-07-27 10:17 /tmp
drwxr-xr-x - acadgild supergroup      0 2018-06-29 02:08 /user
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -ls /sqoopout2
18/09/01 17:47:02 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup      0 2018-09-01 17:45 /sqoopout2/ SUCCESS
-rw-r--r-- 1 acadgild supergroup    32 2018-09-01 17:45 /sqoopout2/part-m-00000
[acadgild@localhost ~]$ hadoop fs -cat /sqoopout2/part-m-00000
18/09/01 17:47:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Cybercity,Gurgaon,Puri,Gaurav,3
[acadgild@localhost ~]$
```

UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>

Type here to search

11:39 04-09-2018