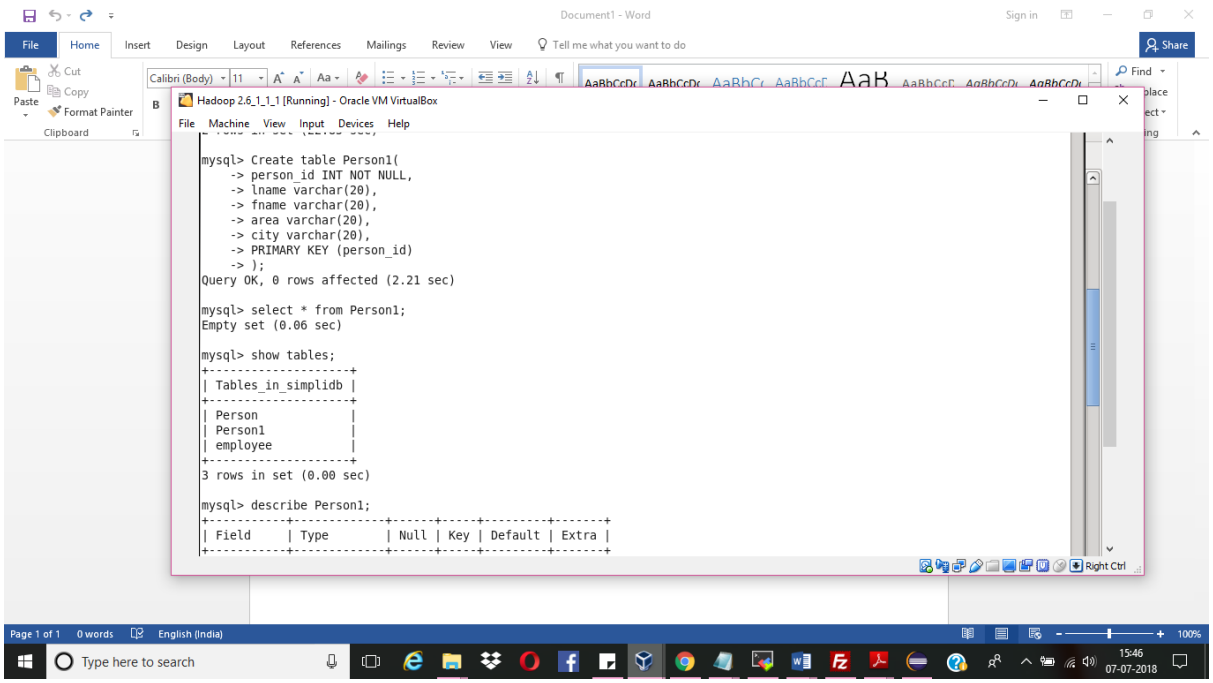
Session 6:
DATA INGESTION TOOL SQOOP & INTRODUCTION TO CASE STUDY

Assignment 1

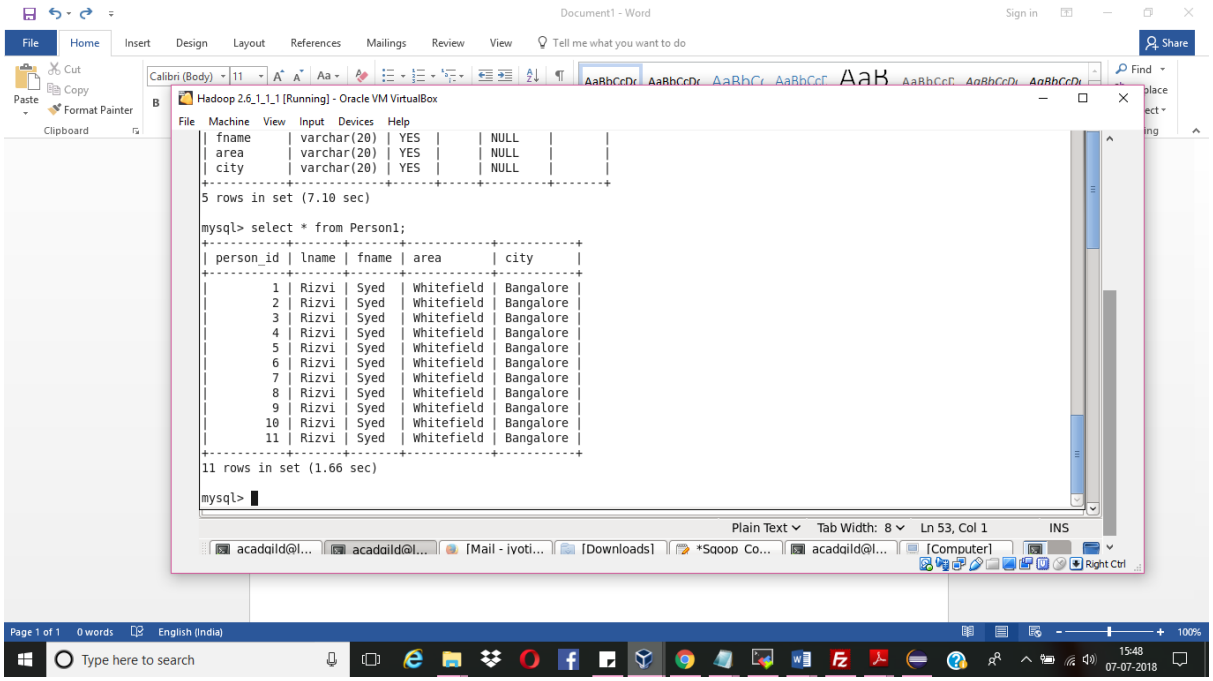
Task 1

Use Sqoop tool to export data present in SQOOPOUT folder made while demo of Import table

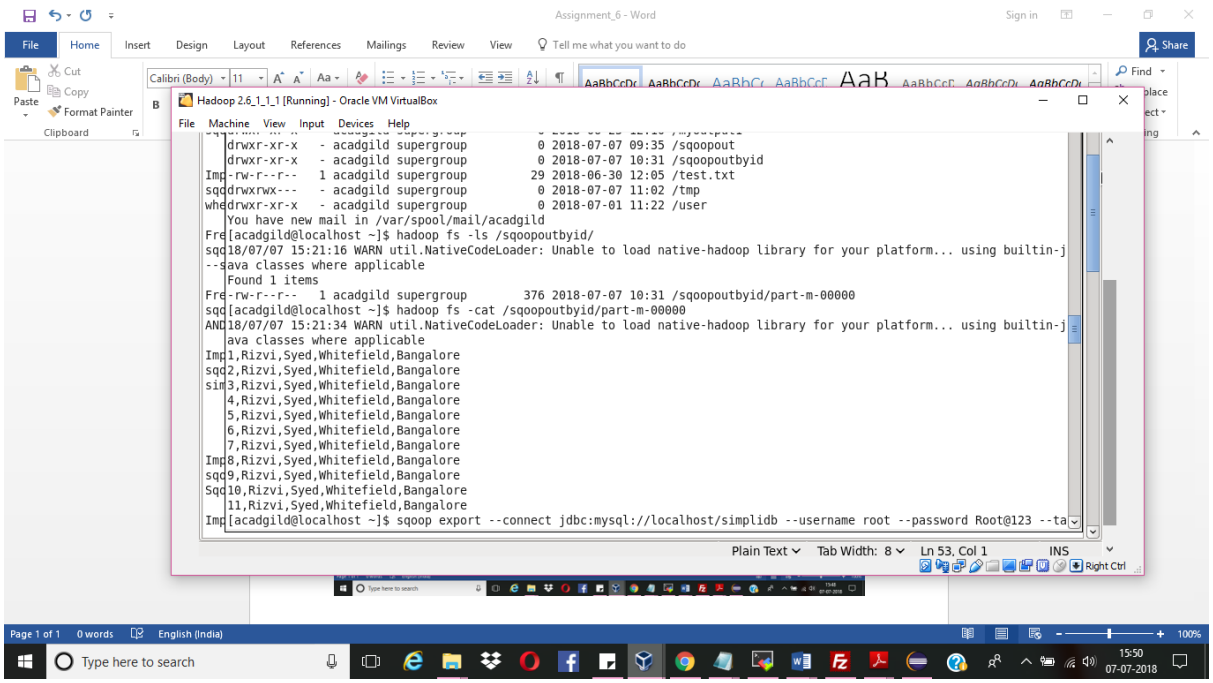
Step 1:



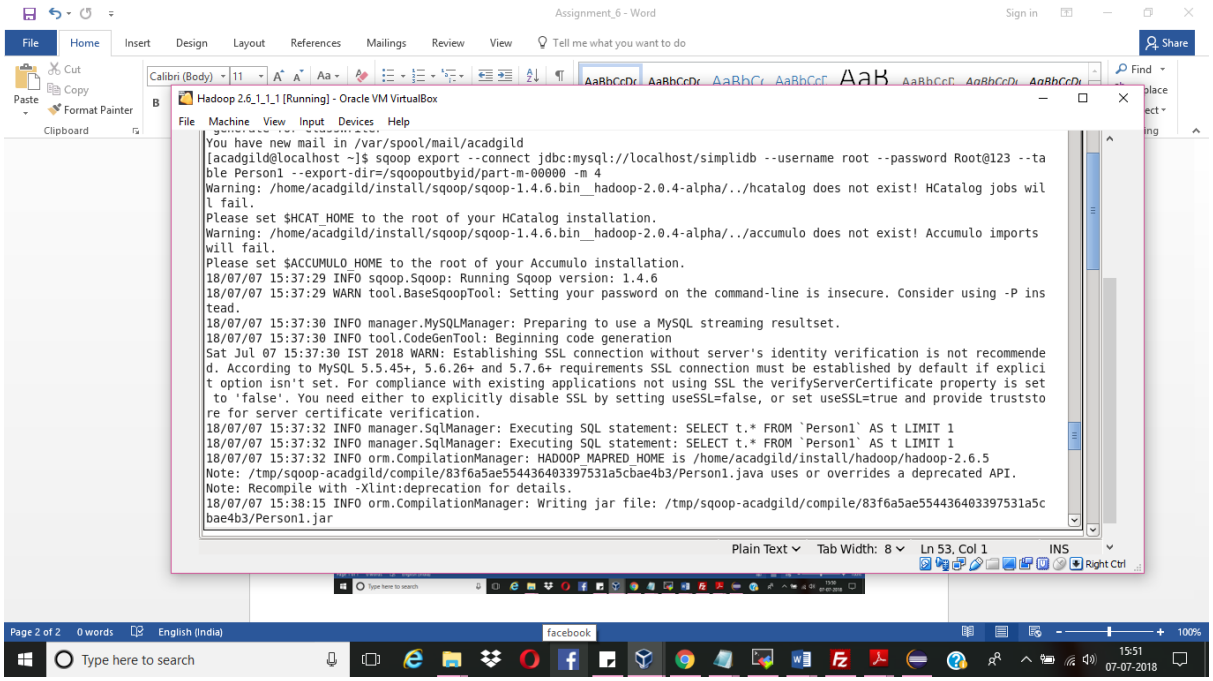
Step 2:


Document1 - Word
Hadoop 2.6_1_1_1 [Running] - Oracle VM VirtualBox
fname	varchar(20)	YES		NULL	
area	varchar(20)	YES		NULL	
city	varchar(20)	YES		NULL	
5 rows in set (7.10 sec)
mysql> select * from Person1;
person_id	lname	fname	area	city
1	Rizvi	Syed	Whitefield	Bangalore
2	Rizvi	Syed	Whitefield	Bangalore
3	Rizvi	Syed	Whitefield	Bangalore
4	Rizvi	Syed	Whitefield	Bangalore
5	Rizvi	Syed	Whitefield	Bangalore
6	Rizvi	Syed	Whitefield	Bangalore
7	Rizvi	Syed	Whitefield	Bangalore
8	Rizvi	Syed	Whitefield	Bangalore
9	Rizvi	Syed	Whitefield	Bangalore
10	Rizvi	Syed	Whitefield	Bangalore
11	Rizvi	Syed	Whitefield	Bangalore
11 rows in set (1.66 sec)
mysql>
Plain Text Tab Width: 8 Ln 53, Col 1 INS
Type here to search
15:48
07-07-2018

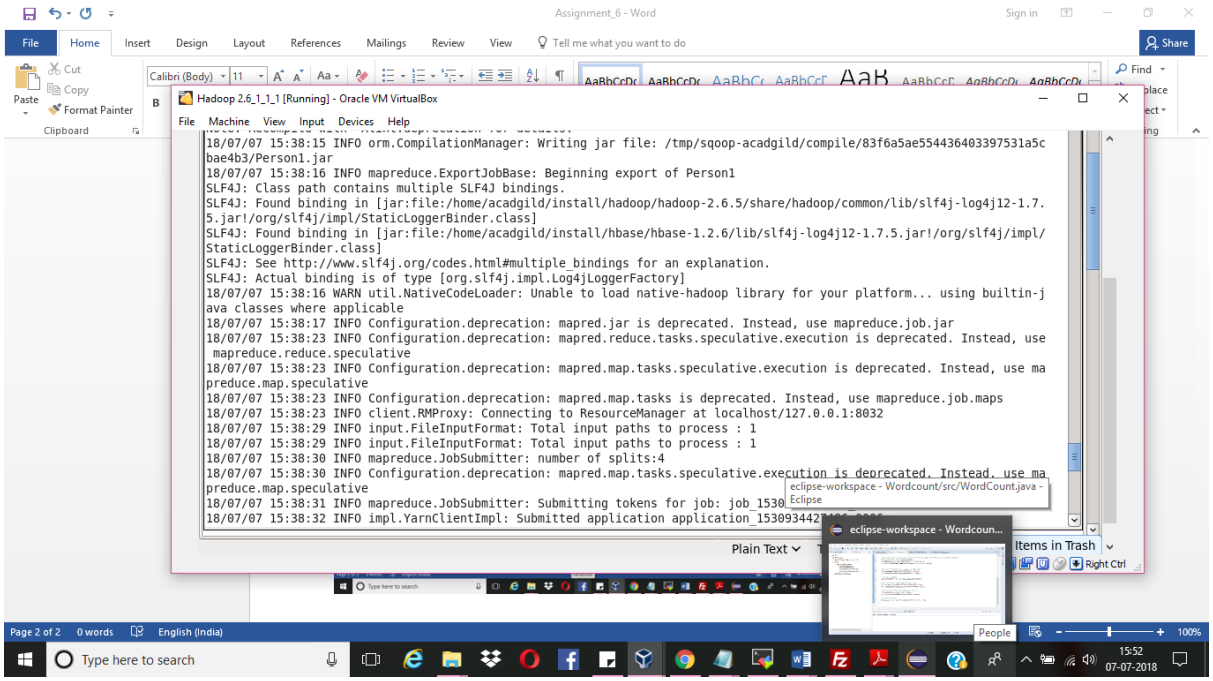
Step 3:

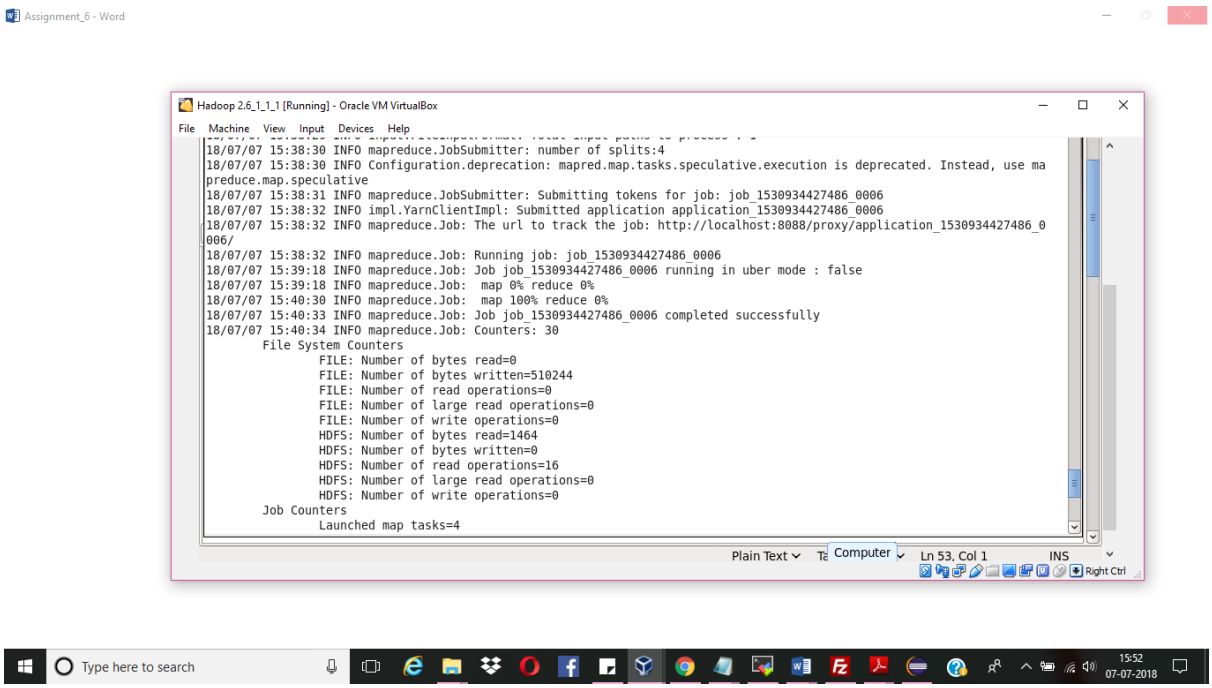

Assignment_6 - Word
Hadoop 2.6_1_1_1 [Running] - Oracle VM VirtualBox
drwxr-xr-x - acadgild supergroup 0 2018-07-07 09:35 /sqoopout
drwxr-xr-x - acadgild supergroup 0 2018-07-07 10:31 /sqoopoutbyid
-rw-r--r-- 1 acadgild supergroup 29 2018-06-30 12:05 /test.txt
drwxrwx--- - acadgild supergroup 0 2018-07-07 11:02 /tmp
drwxr-xr-x - acadgild supergroup 0 2018-07-01 11:22 /user
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -ls /sqoopoutbyid/
18/07/07 15:21:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
-rw-r--r-- 1 acadgild supergroup 376 2018-07-07 10:31 /sqoopoutbyid/part-m-00000
[acadgild@localhost ~]$ hadoop fs -cat /sqoopoutbyid/part-m-00000
18/07/07 15:21:34 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
1,Rizvi,Syed,Whitefield,Bangalore
2,Rizvi,Syed,Whitefield,Bangalore
3,Rizvi,Syed,Whitefield,Bangalore
4,Rizvi,Syed,Whitefield,Bangalore
5,Rizvi,Syed,Whitefield,Bangalore
6,Rizvi,Syed,Whitefield,Bangalore
7,Rizvi,Syed,Whitefield,Bangalore
8,Rizvi,Syed,Whitefield,Bangalore
9,Rizvi,Syed,Whitefield,Bangalore
10,Rizvi,Syed,Whitefield,Bangalore
11,Rizvi,Syed,Whitefield,Bangalore
[acadgild@localhost ~]$ sqoop export --connect jdbc:mysql://localhost/simplidb --username root --password Root@123 --ta
Plain Text Tab Width: 8 Ln 53, Col 1 INS
Type here to search
15:50
07-07-2018

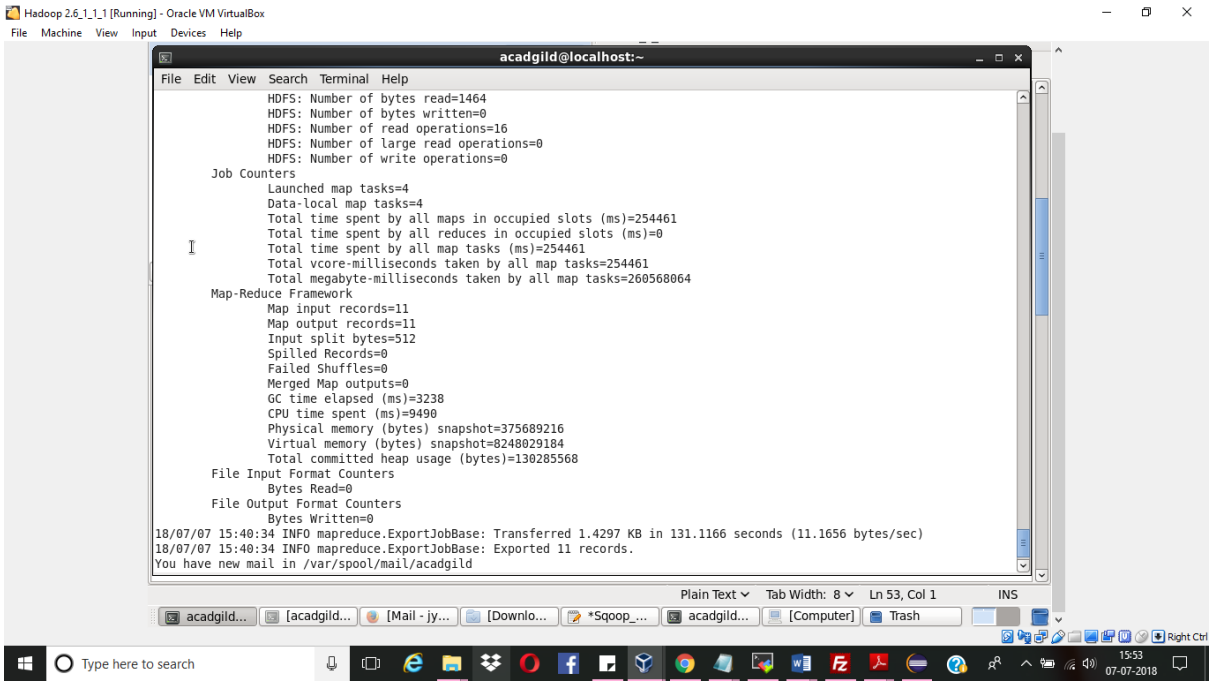
Step 4:


Hadoop 2.6_1_1_1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ sqoop export --connect jdbc:mysql://localhost/simplidb --username root --password Root@123 --table Person1 --export-dir=/sqoopoutbyid/part-m-00000 -m 4
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.
18/07/07 15:37:29 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6
18/07/07 15:37:29 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
18/07/07 15:37:30 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
18/07/07 15:37:30 INFO tool.CodeGenTool: Beginning code generation
Sat Jul 07 15:37:30 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ 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.
18/07/07 15:37:32 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `Person1` AS t LIMIT 1
18/07/07 15:37:32 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `Person1` AS t LIMIT 1
18/07/07 15:37:32 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /home/acadgild/install/hadoop/hadoop-2.6.5
Note: /tmp/sqoop-acadgild/compile/83f6a5ae554436403397531a5cbae4b3/Person1.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
18/07/07 15:38:15 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-acadgild/compile/83f6a5ae554436403397531a5cbae4b3/Person1.jar
Plain Text Tab Width: 8 Ln 53, Col 1 INS
Right Ctrl
Page 2 of 2 0 words English (India)
facebook
Type here to search
15:51
07-07-2018

STEP 5:


Hadoop 2.6_1_1_1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
18/07/07 15:38:15 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-acadgild/compile/83f6a5ae554436403397531a5cbae4b3/Person1.jar
18/07/07 15:38:16 INFO mapreduce.ExportJobBase: Beginning export of Person1
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]
18/07/07 15:38:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/07/07 15:38:17 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
18/07/07 15:38:23 INFO Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce.reduce.speculative
18/07/07 15:38:23 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
18/07/07 15:38:23 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
18/07/07 15:38:23 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/07/07 15:38:29 INFO input.FileInputFormat: Total input paths to process : 1
18/07/07 15:38:29 INFO input.FileInputFormat: Total input paths to process : 1
18/07/07 15:38:30 INFO mapreduce.JobSubmitter: number of splits:4
18/07/07 15:38:30 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
18/07/07 15:38:31 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1530
18/07/07 15:38:32 INFO impl.YarnClientImpl: Submitted application application_153093442
eclipse-workspace - Wordcount/src/WordCount.java - Eclipse
eclipse-workspace - Wordcoun...
Items in Trash
Plain Text
Right Ctrl
Page 2 of 2 0 words English (India)
People
Type here to search
15:52
07-07-2018





Task2

Use Sqoop tool to export data present in SQOOPOUT folder made while demo of Import table with parameter person_id =3.

```
mysql> select * from Person2;
Empty set (0.00 sec)

mysql>
```

Check Sqoopoutbyid folder present in hadoop .

```
[acadgild@localhost ~]$ hadoop fs -ls /sqoopoutbyid
18/08/19 14:12:40 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 2018-08-19 14:12 /sqoopoutbyid/_SUCCESS
-rw-r--r--   1 acadgild supergroup         17 2018-08-19 14:12 /sqoopoutbyid/part-m-00000
[acadgild@localhost ~]$ hadoop fs -cat /sqoopoutbyid/part-m-00000
18/08/19 14:13:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
applicable
USA,NJ,BBB,AAA,3
```

Now we will write export command to fetch data from sqoopoutbyid folder to our mysql PERSON2 table.

```
USA,NJ,BBB,AAA,3
[acadgild@localhost ~]$ sqoop export -m 1 --connect jdbc:mysql://localhost/simplidb --username root --P --table Person2 --export-dir /sqo
opoutbyid/
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.
18/08/19 14:14:02 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6
Enter password:
18/08/19 14:14:07 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
18/08/19 14:14:07 INFO tool.CodeGenTool: Beginning code generation
Sun Aug 19 14:14:07 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. According to My
SQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance w
ith 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.
18/08/19 14:14:09 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `Person2` AS t LIMIT 1
18/08/19 14:14:09 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `Person2` AS t LIMIT 1
18/08/19 14:14:09 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /home/acadgild/install/hadoop/hadoop-2.6.5
Note: /tmp/sqoop-acadgild/compile/1405dbe0807104bed7c6d0353655b2bf/Person2.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
18/08/19 14:14:12 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-acadgild/compile/1405dbe0807104bed7c6d0353655b2bf/Person2.jar
18/08/19 14:14:12 INFO mapreduce.ExportJobBase: Beginning export of Person2
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/i
mpl/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]
18/08/19 14:14:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
applicable
18/08/19 14:14:13 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
18/08/19 14:14:15 INFO Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce.reduce.
speculative
18/08/19 14:14:15 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.specul
ative
18/08/19 14:14:15 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
18/08/19 14:14:15 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
```

```
18/08/19 14:14:15 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
18/08/19 14:14:15 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/08/19 14:14:17 INFO input.FileInputFormat: Total input paths to process : 1
18/08/19 14:14:17 INFO input.FileInputFormat: Total input paths to process : 1
18/08/19 14:14:18 INFO mapreduce.JobSubmitter: number of splits:1
18/08/19 14:14:18 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.specul
ative
18/08/19 14:14:18 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1534667273374_0003
18/08/19 14:14:19 INFO impl.YarnClientImpl: Submitted application application_1534667273374_0003
18/08/19 14:14:19 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1534667273374_0003/
18/08/19 14:14:19 INFO mapreduce.Job: Running job: job_1534667273374_0003
18/08/19 14:14:34 INFO mapreduce.Job: Job job_1534667273374_0003 running in uber mode : false
18/08/19 14:14:34 INFO mapreduce.Job:  map 0% reduce 0%
18/08/19 14:14:44 INFO mapreduce.Job:  map 100% reduce 0%
18/08/19 14:14:45 INFO mapreduce.Job: Job job_1534667273374_0003 completed successfully
18/08/19 14:14:46 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=127548
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=148
                HDFS: Number of bytes written=0
                HDFS: Number of read operations=4
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=0
        Job Counters
                Launched map tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=8048
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=8048
                Total vcore-milliseconds taken by all map tasks=8048
                Total megabyte-milliseconds taken by all map tasks=8241152
        Map-Reduce Framework
                Map input records=1
                Map output records=1
```

```
        Job Counters
                Launched map tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=8048
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=8048
                Total vcore-milliseconds taken by all map tasks=8048
                Total megabyte-milliseconds taken by all map tasks=8241152
        Map-Reduce Framework
                Map input records=1
                Map output records=1
                Input split bytes=128
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=124
                CPU time spent (ms)=1950
                Physical memory (bytes) snapshot=105771008
                Virtual memory (bytes) snapshot=2061332480
                Total committed heap usage (bytes)=32571392
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=0
18/08/19 14:14:46 INFO mapreduce.ExportJobBase: Transferred 148 bytes in 31.0136 seconds (4.7721 bytes/sec)
18/08/19 14:14:46 INFO mapreduce.ExportJobBase: Exported 1 records.
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

Now Check the Exported data in PERSON2 Table;

```
mysql> select * from Person2;
Empty set (0.00 sec)

mysql> select * from Person2;
+-----------+-------+-------+------+------+
| person_id | lname | fname | area | city |
+-----------+-------+-------+------+------+
|         3 | AAA   | BBB   | USA  | NJ   |
+-----------+-------+-------+------+------+
1 row in set (0.00 sec)

mysql>
```