

ACADGILD

Session 11: Sqoop Flume

Assignment 1 Question Session 11:

Assignment 1

Problem Statement

Perform and explain the code flow and the associated result for the below tasks. Candidates should create and use their own employee dataset for the same. Share the screenshot of the commands used and its associated result.

• Transfer data between Mysql and HDFS using Sqoop.

[acadgild@localhost ~]\$ sudo service mysqld start
[sudo] password for acadgild:

Starting mysqld:

[OK]

[acadgild@localhost ~]\$ mysql -u root -p

```
[acadgild@localhost ~]$ mysql -u root -p
Enter password:
Welcome to the MySQL monitor. Commands end with ; or \g.
Your MySQL connection id is 4
Server version: 5.1.73 Source distribution

Copyright (c) 2000, 2013, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its affiliates. Other names may be trademarks of their respective owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.
```

mysql> grant all on *.* to 'acadgild'@'localhost' with grant option;

mysql> show databases;

mysql> create database db1;

```
mysql> create database db1;
Query OK, 1 row affected (0.00 sec)
```

```
mysql> use db1;
```

```
mysql> use db1;
Database changed
```

mysql> GRANT ALL PRIVILEGES ON db1.* TO 'acadgild'@'localhost';

Query OK, 0 rows affected (0.00 sec)

mysql> flush privileges;

Query OK, 0 rows affected (0.00 sec)

mysql> commit;

Query OK, 0 rows affected (0.00 sec)

```
mysql> GRANT ALL PRIVILEGES ON db1.* TO 'acadgild'@'localhost';
Query OK, 0 rows affected (0.00 sec)

mysql> flush privileges;
Query OK, 0 rows affected (0.00 sec)

mysql> commit;
Query OK, 0 rows affected (0.00 sec)

mysql> use db1;
Database changed
```

mysql> CREATE TABLE acad (emp_id int(2),emp_name varchar(10),emp_sal int(10),date date);

```
mysql> CREATE TABLE acad (emp_id int(2),emp_name varchar(10),emp_sal int(10),date date);
Query OK, 0 rows affected (0.01 sec)
```

[acadgild@localhost ~]\$ cat >>input.txt

5,Hadoop,50000,2011-03-21

6,Spark,60000,2011-03-22

```
7,Scala,40000,2011-02-12
```

8,Java,40000,2012-06-11

```
[acadgild@localhost ~]$ cat >>input.txt
5,Hadoop,50000,2011-03-21
6,Spark,60000,2011-02-12
8,Java,40000,2012-06-11
^C
[acadgild@localhost ~]$ cat input.txt
5,Hadoop,50000,2011-03-21
6,Spark,60000,2011-03-22
7,Scala,40000,2011-02-12
8,Java,40000,2012-06-11
```

[acadgild@localhost ~]\$ hadoop fs -put input.txt /user/acadgild/sqoop_sql

```
[acadgild@localhost ~]$ hadoop fs -put input.txt /user/acadgild/sqoop_sql
17/10/24 18:04:55 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes wher
e applicable
[acadgild@localhost ~]$ hadoop fs -cat /user/acadgild/sqoop_sql/input.txt
17/10/24 18:05:17 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes wher
e applicable
5,Hadoop,50000,2011-03-21
6,Spark,60000,2011-03-22
7,Scala,40000,2011-03-22
8,Java,40000,2012-06-11
```

[acadgild@localhost ~]\$ sqoop export --connect jdbc:mysql://localhost/db1 --username root --password root --table acad --export-dir sqoop_sql

[acadgide]coalhost ~]\$ sqoop export --connect jdbc:mysql://localhost/dbl --username root --password root --table acad --export-dir sqo og gql
Marning, /usr/local/sqoop/../hostalog does not exist! Hostalog jobs will fail.
Please set \$HCNT_HOME to the root of your HCatalog installation.
Warning; /usr/local/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set \$ACCUMBNIO_HOME to the root of your Accumulo installation.
Warning; /usr/local/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set \$ACCUMBNIO_HOME to the root of your Accumulo imports will fail.

2017-10-24 18:05:23,476 INFO [main] aqoop \$600P; Running \$600P version: 1.4.5
2017-10-24 18:05:23,499 RNFO [main] apoop \$600P; Running \$600P version: 1.4.5
2017-10-24 18:05:23,499 RNFO [main] manager. MySOUManager: Preparing to use a MySOL streaming resultset.
2017-10-24 18:05:23,698 INFO [main] manager. MySOUManager: Preparing to use a MySOL streaming resultset.
2017-10-24 18:05:23,698 INFO [main] manager. MySOUManager: Preparing to use a MySOL streaming resultset.
2017-10-24 18:05:23,698 INFO [main] manager. SqlManager: Preparing to use a MySOL streaming resultset.
2017-10-24 18:05:24,006 INFO [main] manager. SqlManager: Preparing to use a MySOL streaming resultset.
2017-10-24 18:05:24,006 INFO [main] manager. SqlManager: Preparing to use a MySOL streaming resultset.
2017-10-24 18:05:24,006 INFO [main] manager. SqlManager: Preparing to use a MySOL streaming resultset.
2017-10-24 18:05:25,008 INFO [main] manager. SqlManager: Preparing Polymanager. SplECT t.* PROM 'mond' AS t LIMIT 1
2017-10-24 18:05:24,006 INFO [main] manager. SqlManager: Preparing Polymanager. SplECT t.* PROM 'mond' AS t LIMIT 1
2017-10-24 18:05:25,008 INFO [main] com. CompilationManager: Writing jar file: /tmp/sqoop-acadgild/compile/bocelb9afc2ed9ae9346dadcod18
2017-10-24 18:05:25,668 INFO [main] com. CompilationManager: Writing jar file: /tmp/sqoop-acadgild/compile/bocelb9afc2ed9ae9346dadcod18
2017-10-24 18:05:25,668 INFO [main] manager. SplManager: Prepar

p acadgidelocalhost-

acaugitu@iocaiiosc-	B ^
educe.map.speculative	
2017-10-24 18:05:26,654 INFO	[main] Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
2017-10-24 18:05:26,806 INFO	[main] client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2017-10-24 18:05:27,912 INFO	[main] input.FileInputFormat: Total input paths to process : 1
2017-10-24 18:05:27,918 INFO	[main] input.FileInputFormat: Total input paths to process : 1
2017-10-24 18:05:27,997 INFO	[main] mapreduce.JobSubmitter: number of splits:4
2017-10-24 18:05:28,030 INFO	[main] Configuration.deprecation: mapred.job.name is deprecated. Instead, use mapreduce.job.name
2017-10-24 18:05:28,030 INFO	[main] Configuration.deprecation: mapred.cache.files.timestamps is deprecated. Instead, use mapreduce.job
.cache.files.timestamps	
2017-10-24 18:05:28,030 INFO	[main] Configuration.deprecation: mapreduce.map.class is deprecated. Instead, use mapreduce.job.map.class
2017-10-24 18:05:28,030 INFO	[main] Configuration.deprecation: mapred.input.dir is deprecated. Instead, use mapreduce.input.fileinputf
ormat.inputdir	
2017-10-24 18:05:28,034 INFO	[main] Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapr
educe.map.speculative	
2017-10-24 18:05:28,035 INFO	[main] Configuration.deprecation: mapreduce.inputformat.class is deprecated. Instead, use mapreduce.job.i
nputformat.class	
2017-10-24 18:05:28,035 INFO	[main] Configuration.deprecation: mapreduce.outputformat.class is deprecated. Instead, use mapreduce.job.
outputformat.class	
2017-10-24 18:05:28,035 INFO	[main] Configuration.deprecation: mapred.cache.files is deprecated. Instead, use mapreduce.job.cache.file
S	
2017-10-24 18:05:28,035 INFO	[main] Configuration.deprecation: mapred.working.dir is deprecated. Instead, use mapreduce.job.working.di
r 2017-10-24 18:05:28,035 INFO	[main] Configuration.deprecation: mapred.mapoutput.value.class is deprecated. Instead, use mapreduce.map.
output.value.class	[main] Configuration.deprecation: mapred.mapoutput.value.class is deprecated. instead, use mapreduce.map.
2017-10-24 18:05:28,035 INFO	[main] Configuration.deprecation: mapred.mapoutput.kev.class is deprecated. Instead, use mapreduce.map.ou
tput.key.class	[main] Configuration.deprecation: mapred.mapoutput.key.class is deprecated. Instead, use mapreduce.map.ou
2017-10-24 18:05:28,035 INFO	[main] Configuration.deprecation: mapred.job.classpath.files is deprecated. Instead, use mapreduce.job.cl
asspath.files	[main] Configuration.ueprecation.mapled.job.crasspach.files is deprecated.finstead, use mapleduce.job.cr
2017-10-24 18:05:28,035 INFO	[main] Configuration.deprecation: user.name is deprecated. Instead, use mapreduce.job.user.name
2017-10-24 18:05:28,035 INFO	[main] Configuration.deprecation: mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces
2017-10-24 18:05:28,035 INFO	[main] Configuration, deprecation: mapred.cache.files.filesizes is deprecated. Instead, use mapreduce.job.
cache.files.filesizes	[] 0011-3-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1
2017-10-24 18:05:28,218 INFO	[main] mapreduce.JobSubmitter: Submitting tokens for job: job 1508822481355 0003
2017-10-24 18:05:28,686 INFO	[main] impl.YarnClientImpl: Submitted application application 1508822481355 0003 to ResourceManager at /0
.0.0.0:8032	**
2017-10-24 18:05:28,727 INFO	[main] mapreduce.Job: The url to track the job: http://http://localhost:8088/proxy/application 1508822481

mysql> select * from acad;

Importing sql table to hdfs directory

[acadgild@localhost ~]\$ sqoop import --connect jdbc:mysql://localhost/db1 --username root --password root --table acad --target-dir sqoop_sql_import1 -m 1

```
2017-10-24 18:27:47,347 INFO [main] Configuration.deprecation: mapred.working.dir is deprecated. Instead, use mapreduce.job.working.di
2017-10-24 18:27:47,348 INFO [main] Configuration.deprecation: mapred.job.classpath.files is deprecated. Instead, use mapreduce.job.cl
asspath.files
2017-10-24 18:27:47,348 INFO [main] Configuration.deprecation: user.name is deprecated. Instead, use mapreduce.job.user.name
2017-10-24 18:27:47,348 INFO [main] Configuration.deprecation: mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces 2017-10-24 18:27:47,348 INFO [main] Configuration.deprecation: mapred.cache.files.filesizes is deprecated. Instead, use mapreduce.job
  ache.files.filesizes
017-10-24 18:27:47,348 INFO [main] Configuration.deprecation: mapred.output.key.class is deprecated. Instead, use mapreduce.job.outpu
t.key.class
2017-10-24 18:27:47,477 INFO [main] mapreduce.JobSubmitter: Submitting tokens for job: job_1508822481355_0004
2017-10-24 18:27:47,722 INFO [main] impl.YarnClientImpl: Submitted application application_1508822481355_0004 to ResourceManager at /(
2017-10-24 18:27:47,748 INFO [main] mapreduce.Job: The url to track the job: http://http://localhost:8088/proxy/application 1508822481
355_0004/
2017-10-24 18:27:47,749 INFO
2017-10-24 18:27:47,749 INFO
2017-10-24 18:27:55,977 INFO
2017-10-24 18:27:55,977 INFO
2017-10-24 18:27:55,977 INFO
2017-10-24 18:28:01,309 INFO
2017-10-24 18:28:01,309 INFO
[main] mapreduce.Job: map 0% reduce 0%
2017-10-24 18:28:01,342 INFO
[main] mapreduce.Job: map 10% reduce 0%
2017-10-24 18:28:01,342 INFO
[main] mapreduce.Job: job 1508822481355_0004 running in uber mode: false

2017-10-24 18:28:01,350 INFO
[main] mapreduce.Job: map 10% reduce 0%
2017-10-24 18:28:01,514 ERROR
[main] mapreduce.Job: Job job 1508822481355_0004 completed successfully
2017-10-24 18:28:01,514 ERROR
[main] tool.ImportTool: Imported Failed: No enum constant org.apache.hadoop.mapreduce.JobCounter.MB_MILLI
[acadgild@localhost ~]$ hadoop fs -ls /user/acadgild/
17/10/24 18:30:08 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes when
 e applicable
-rw-r-r- 1 acadgild supergroup 69234933 2017-10-07 15:08 /user/acadgild/crime.csv
drwxr-xr-x - acadgild supergroup 0 2017-09-29 18:33 /user/acadgild/hadoop
drwxr-xr-x - acadgild supergroup 0 2017-10-07 16:33 /user/acadgild/program1
drwxr-xr-x - acadgild supergroup 0 2017-10-07 16:41 /user/acadgild/program2
                                                                                         0 2017-10-07 16:33 /user/acadgild/program1
0 2017-10-07 16:41 /user/acadgild/program2
0 2017-10-07 17:00 /user/acadgild/program3
0 2017-10-07 15:13 /user/acadgild/program4
0 2017-10-24 18:04 /user/acadgild/sqoop_sql
0 2017-10-24 18:24 /user/acadgild/sqoop_sql_import
0 2017-10-24 18:27 /user/acadgild/sqoop_sql_import
15 2017-09-30 20:37 /user/acadgild/student details.tx
                       acadgild supergroupacadgild supergroup
drwxr-xr-x
drwxr-xr-x - acadgild supergroup
drwxr-xr-x - acadgild supergroup
drwxr-xr-x - acadgild supergroup
-rw-r-r-- 1 acadgild supergroup
```

[acadgild@localhost ~]\$ hadoop fs -ls /user/acadgild/sqoop_sql_import1/

• Transfer data between Mysql and Hive (Import only selected columns) using Sqoop.

[acadgild@localhost ~]\$ sqoop import --connect jdbc:mysql://localhost/db1 --username root --password root --table acad --where "emp_name = 'Scala'" --target-dir sqoop_sql_import2 -m 1

```
2017-10-24 18:42:20,250 INFO [main] mapreduce.Job: Running job: job_1508822481355_0005
2017-10-24 18:42:28,443 INFO [main] mapreduce.Job: Job job_1508822481355_0005 running in uber mode: false
2017-10-24 18:42:28,447 INFO [main] mapreduce.Job: map 0% reduce 0%
2017-10-24 18:42:33,555 INFO [main] mapreduce.Job: map 100% reduce 0%
2017-10-24 18:42:34,636 INFO [main] mapreduce.Job: Job job_1508822481355_0005 completed successfully
2017-10-24 18:42:34,773 ERROR [main] tool.ImportTool: Imported Failed: No enum constant org.apache.hadoop.mapreduce.JobCounter.MB_MILLI
S_MAPS
Lacadrild%localbest ~1% badoop_fs_cat_/user/acadrild/
```

[acadgild@localhost ~]\$ hadoop fs -cat /user/acadgild/sqoop_sql_import2/part-m-00000

```
[acadgild@localhost ~]$ hadoop fs -ls /user/acadgild/sqoop_sql_import2
17/10/24 18:43:23 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes whe e applicable
Found 2 items
-rw-r-r--- 1 acadgild supergroup 0 2017-10-24 18:42 /user/acadgild/sqoop_sql_import2/SUCCESS
-rw-r--r-- 1 acadgild supergroup 25 2017-10-24 18:42 /user/acadgild/sqoop_sql_import2/part-m-00000
[acadgild@localhost ~]$ hadoop fs -cat /user/acadgild/sqoop_sql_import2/part-m-00000
17/10/24 18:43:40 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes whe e applicable
7,Scala,40000,2011-02-12
[acadgild@localhost ~]$
```