

Lab 1.

1. Create database and table in Mysql according to oil dataset.
2. Sqoop import oil data into HDFS as parquet file.

Hint: `$ sqoop import <options> --as-parquetfile`

3. Verify parquet file is created.

```

cloudera@quickstart:~
File Edit View Search Terminal Help
Query OK, 0 rows affected (0.15 sec)

mysql> drop table oil data;
Query OK, 0 rows affected (0.02 sec)

mysql> create table oil(district varchar(255),distributer varchar(255),buy_rate
int,sell_rate int,volume_in int,volume_out int,year int);
Query OK, 0 rows affected (0.01 sec)

mysql> desc oil;
+-----+-----+-----+-----+-----+-----+
| Field      | Type          | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| district   | varchar(255)  | YES  |     | NULL    |       |
| distributer | varchar(255)  | YES  |     | NULL    |       |
| buy_rate   | int(11)       | YES  |     | NULL    |       |
| sell_rate  | int(11)       | YES  |     | NULL    |       |
| volume_in  | int(11)       | YES  |     | NULL    |       |
| volume_out | int(11)       | YES  |     | NULL    |       |
| year       | int(11)       | YES  |     | NULL    |       |
+-----+-----+-----+-----+-----+-----+
7 rows in set (0.05 sec)

mysql>
[cloudera@quickstart ~]$ sqoop import --connect 'jdbc:mysql://localhost/sqoop' --username root --password cloudera --table oil --target-dir /user/cloudera/sqoopfile --as-parquetfile -m 1
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
19/12/09 10:18:07 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.13.0
19/12/09 10:18:07 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
19/12/09 10:18:08 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
19/12/09 10:18:08 INFO tool.CodeGenTool: Beginning code generation
19/12/09 10:18:08 INFO tool.CodeGenTool: Will generate java class as codegen_oil
19/12/09 10:18:09 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'oil' AS t LIMIT 1
19/12/09 10:18:09 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'oil' AS t LIMIT 1
19/12/09 10:18:09 INFO orm.CompilationManager: HADOOP MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-cloudera/compile/4a3407d166e400d81c576c6d51633ec6/codegen_oil.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
19/12/09 10:18:13 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/4a3407d166e400d81c576c6d51633ec6/codegen_oil.jar
19/12/09 10:18:13 WARN manager.MySQLManager: It looks like you are importing from mysql.
19/12/09 10:18:13 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
19/12/09 10:18:13 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
19/12/09 10:18:13 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
19/12/09 10:18:13 INFO mapreduce.ImportJobBase: Beginning import of oil
19/12/09 10:18:13 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
19/12/09 10:18:13 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
19/12/09 10:18:15 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'oil' AS t LIMIT 1
19/12/09 10:18:15 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'oil' AS t LIMIT 1
19/12/09 10:18:16 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedExcepcion
    at java.lang.Object.wait(Native Method)
import data from Mysql to HDFS by sqoop
- Cloudera Community - Mozilla Firefox
oil_data.csv (~/.Desko... cloudera@quickstart:~ import data from Mys... cloudera@quickstart:~

```

```
FILE: Number of write operations=0
HDFS: Number of bytes read=10530
HDFS: Number of bytes written=11556
HDFS: Number of read operations=68
HDFS: Number of large read operations=0
HDFS: Number of write operations=10
Job Counters
  Launched map tasks=1
  Other local map tasks=1
  Total time spent by all maps in occupied slots (ms)=13946
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=13946
  Total vcore-milliseconds taken by all map tasks=13946
  Total megabyte-milliseconds taken by all map tasks=14280704
Map-Reduce Framework
  Map input records=401
  Map output records=401
  Input split bytes=87
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=280
  CPU time spent (ms)=1580
  Physical memory (bytes) snapshot=161107968
  Virtual memory (bytes) snapshot=1513381888
  Total committed heap usage (bytes)=74973184
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=0
19/12/09 10:18:51 INFO mapreduce.ImportJobBase: Transferred 11.2852 KB in 34.3513 seconds (336.4065 bytes/sec)
19/12/09 10:18:51 INFO mapreduce.ImportJobBase: Retrieved 401 records.
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ hdfs dfs -ls /user/cloudera/sqoopfiles;
Found 5 items
-rw-r--r-- 1 cloudera cloudera 0 2019-12-03 20:25 /user/cloudera/sqoopfiles/_SUCCESS
-rw-r--r-- 1 cloudera cloudera 741614 2019-12-03 20:24 /user/cloudera/sqoopfiles/part-m-00000
-rw-r--r-- 1 cloudera cloudera 753022 2019-12-03 20:25 /user/cloudera/sqoopfiles/part-m-00001
-rw-r--r-- 1 cloudera cloudera 752368 2019-12-03 20:25 /user/cloudera/sqoopfiles/part-m-00002
-rw-r--r-- 1 cloudera cloudera 752940 2019-12-03 20:25 /user/cloudera/sqoopfiles/part-m-00003
[cloudera@quickstart ~]$
17190,2013-11-09 00:00:00.0,0021,COMPLETE
17191,2013-11-09 00:00:00.0,1455,PENDING_PAYMENT
17192,2013-11-09 00:00:00.0,9320,PENDING
17193,2013-11-09 00:00:00.0,1593,PENDING_PAYMENT
17194,2013-11-09 00:00:00.0,2581,CANCELED
17195,2013-11-09 00:00:00.0,621,COMPLETE
17196,2013-11-09 00:00:00.0,3293,COMPLETE
17197,2013-11-09 00:00:00.0,1540,SUSPECTED_FRAUD
17198,2013-11-09 00:00:00.0,642,CLOSED
17199,2013-11-09 00:00:00.0,7246,PENDING_PAYMENT
17200,2013-11-09 00:00:00.0,4846,PENDING_PAYMENT
17201,2013-11-09 00:00:00.0,10506,PENDING_PAYMENT
17202,2013-11-09 00:00:00.0,4145,PROCESSING
17203,2013-11-09 00:00:00.0,6725,COMPLETE
17204,2013-11-09 00:00:00.0,3960,CLOSED
17205,2013-11-09 00:00:00.0,2715,CLOSED
17206,2013-11-09 00:00:00.0,2848,PROCESSING
17207,2013-11-09 00:00:00.0,8986,COMPLETE
17208,2013-11-09 00:00:00.0,1364,CLOSED
17209,2013-11-09 00:00:00.0,336,CLOSED
17210,2013-11-09 00:00:00.0,12143,PENDING_PAYMENT
17211,2013-11-09 00:00:00.0,1595,COMPLETE
17212,2013-11-09 00:00:00.0,11387,COMPLETE
17213,2013-11-09 00:00:00.0,6166,COMPLETE
17214,2013-11-09 00:00:00.0,585,CLOSED
17215,2013-11-09 00:00:00.0,8326,COMPLETE
17216,2013-11-09 00:00:00.0,5729,COMPLETE
17217,2013-11-09 00:00:00.0,486,COMPLETE
17218,2013-11-09 00:00:00.0,1870,CLOSED
17219,2013-11-09 00:00:00.0,7749,CLOSED
17220,2013-11-09 00:00:00.0,4821,COMPLETE
17221,2013-11-09 00:00:00.0,2366,PENDING
[cloudera@quickstart ~]$
```

⚙️ Actions ▾

✖ Move to trash ▾

📶 Upload ▾

➕ New ▾

🏠 Home

/ user / cloudera / sqoopfile

<input type="checkbox"/>	Name	Size	User	Group	Permissions	Date
<input type="checkbox"/>	↑		cloudera	cloudera	drwxr-xr-x	December 09, 2019 11:23 AM
<input type="checkbox"/>	.		cloudera	cloudera	drwxr-xr-x	December 09, 2019 10:18 AM
<input type="checkbox"/>	.metadata		cloudera	cloudera	drwxr-xr-x	December 09, 2019 10:18 AM
<input type="checkbox"/>	.signals		cloudera	cloudera	drwxr-xr-x	December 09, 2019 10:18 AM
<input type="checkbox"/>	22962808-fcab-4c5c-b02b-e576da72c14e.parquet	8.9 KB	cloudera	cloudera	-rw-r--r--	December 09, 2019 10:18 AM

Show 45 of 3 items

Page 1 of 1

⏪ ⏩ ⏴ ⏵

w=/user#

Lab 2.

1. Create a database in hive for storing emp details.
2. Create a mysql database and table for storing emp_details dataset.
3. Sqoop import emp details data into hive table.

Verify if hive table is created and data has been loaded.

```
mysql> create table emp(id varchar(255),job_title varchar(255),dept varchar(255),emp_name varchar(255),salary int);
Query OK, 0 rows affected (0.10 sec)
```

```
mysql> LOAD DATA INFILE '/home/cloudera/Desktop/emp.csv' INTO TABLE emp FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n';
Query OK, 20001 rows affected, 40002 warnings (0.29 sec)
Records: 20001 Deleted: 0 Skipped: 0 Warnings: 20001
```

```
mysql> select * from emp limit 10;
```

id	job_title	dept	emp_name	sal
ID	Job Title	Company	Dept	
9b837fd3-4e6c-4bf6-9762-e4388294c4a3	HR Coordinator	Zepter	Marketing	
947ddace-daf5-444a-a87d-0356812982e6	Loan Officer	Danone	Marketing	
49cef4bc-a392-4c80-8206-d2acbdb5a905	Cashier	Boeing	Accounting	
5e33e850-65e9-4147-9472-d3a82dd445d4	HR Coordinator	Carrefour	Research and Development	
7eb52f54-45dd-4693-9853-80611d9187d3	Treasurer	Team Guard SRL	Marketing	
433d88ea-d651-4c0e-9593-b9da5ae4e31d	Cook	Zepter	Human Resources	
c30538e6-5a15-42a7-abc0-97c1439c6803	Design Engineer	Boeing	IT	
263fe2e6-5263-47c8-9f9a-52dcca09275c	Lecturer	Comodo	Operations	
ffelae7d-ba16-4208-a955-f26d93103876	Software Engineer	Global Print	Management	

10 rows in set (0.00 sec)

```
[cloudera@quickstart ~]$ sqoop import --connect 'jdbc:mysql://localhost/sqoop' --username root --password cloudera --table emp -m 1 --create-hive-table --hive-table sqoop.emp --target-dir /user/cloudera/hiveemp --hive-import
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
19/12/09 11:20:24 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.13.0
19/12/09 11:20:24 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
19/12/09 11:20:24 INFO tool.BaseSqoopTool: Using Hive-specific delimiters for output. You can override
19/12/09 11:20:24 INFO tool.BaseSqoopTool: delimiters with --fields-terminated-by, etc.
19/12/09 11:20:24 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
19/12/09 11:20:24 INFO tool.CodeGenTool: Beginning code generation
19/12/09 11:20:25 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `emp` AS t LIMIT 1
19/12/09 11:20:26 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `emp` AS t LIMIT 1
19/12/09 11:20:26 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
```

```

Total time spent by all reduces in occupied slots (ms)=0
Total time spent by all map tasks (ms)=9410
Total vcore-milliseconds taken by all map tasks=9410
Total megabyte-milliseconds taken by all map tasks=9635840
Map-Reduce Framework
  Map input records=20001
  Map output records=20001
  Input split bytes=87
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=104
  CPU time spent (ms)=3130
  Physical memory (bytes) snapshot=150769664
  Virtual memory (bytes) snapshot=1511235584
  Total committed heap usage (bytes)=74973184
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=1490154
19/12/09 11:23:47 INFO mapreduce.ImportJobBase: Transferred 1.4211 MB in 27.9596 seconds (52.0475 KB/sec)
19/12/09 11:23:47 INFO mapreduce.ImportJobBase: Retrieved 20001 records.
19/12/09 11:23:47 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `emp` AS t LIMIT 1
19/12/09 11:23:47 INFO hive.HiveImport: Loading uploaded data into Hive

Logging initialized using configuration in jar:file:/usr/lib/hive/lib/hive-common-1.1.0-cdh5.13.0.jar!/hive-log4j.properties
OK
Time taken: 2.577 seconds
Loading data to table default.sqoop_emp1
Table default.sqoop_emp1 stats: [numFiles=1, totalSize=1490154]
OK
Time taken: 0.495 seconds
[cloudera@quickstart ~]$

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> show databases;
OK
airlines
default
emp
Time taken: 0.855 seconds, Fetched: 3 row(s)
hive> use default;
OK
Time taken: 0.042 seconds
hive> show tables;
OK
dept
emp
sqoop_emp1
sqoopdata
Time taken: 0.086 seconds, Fetched: 4 row(s)
hive> select * from sqoop_emp1 limit 10;
OK
ID          Job Title      Company Dept    0
9b837fd3-4e6c-4bf6-9762-e4388294c4a3  HR Coordinator  Zepter  Marketing    0
947ddace-daf5-444a-a87d-0356812982e6  Loan Officer   Danone  Marketing    0
49cef4bc-a392-4c80-8206-d2acbdb5a905  Cashier Boeing Accounting    0
5e33e850-65e9-4147-9472-d3a82dd445d4  HR Coordinator Carrefour Research and Development 0
7eb52f54-45dd-4693-9853-80611d9187d3  Treasurer      Team Guard SRL Marketing    0
433d88ea-d651-4c0e-9593-b9da5ae4e31d  Cook           Zepter  Human Resources 0
c30538e6-5a15-42a7-abc0-97c1439c6803  Design Engineer Boeing  IT            0
263fe2e6-5263-47c8-9f9a-52dcca09275c  Lecturer       Comodo  Operations    0
ffelaed7-ba16-4208-a955-f26d93103876  Software Engineer Global Print Management 0
Time taken: 0.602 seconds, Fetched: 10 row(s)
hive>

```