

Sqoop Export Practical - Cloudera:

Sqoop can be used to export data seamlessly from HDFS into RDBMS systems. Sqoop provides many options to handle different scenarios. Many of them will be explained in multiple tutorials with examples. This Sqoop tutorial will cover simple export, export with delimiter, export specific columns, export with update only and export with upsert.

Generic Arguments to export command

Attribute	Description
--export-dir	This is used to specify HDFS directory from where data need to be exported.
--table	This is used to specify RDBMS table name where data need to be exported.

Scenario 1:

1. Create orders_data table in mysql
2. run the export command

```
sqoop export --connect jdbc:mysql://localhost:3306/retail_db --username root --password cloudera --table orders_data --export-dir /sqoop_practice_m1/part-m-00000
```

3. go to mysql and check the data available there.

Scenario 2:

"--input-fields-terminated-by" can be used to specify specific delimiter present in input data file. I

```
sqoop export \  
--connect jdbc:mysql://localhost:3306/retail_db \  
--username root \  
--password cloudera \  
--table cards_Transactions \  

```

--export-dir /exportdata/cards_transactions.csv

--input-fields-terminated-by ","

Scenario 3:

--columns parameter to fetch the subset of columns:

```
sqoop export --connect jdbc:mysql://localhost:3306/retail_db --username root --password  
cloudera --table orders_data --export-dir /sqoop_practice_m1/part-m-00000 --columns  
order_id,order_Status
```

Scenario 4:

If "--update-key" attribute is used without "--update-mode" then it will only update existing data on specified column but it will not insert any new rows

```
sqoop export --connect jdbc:mysql://localhost:3306/retail_db --username root --password  
cloudera --table orders_data --export-dir /sqoop_practice_m1/part-m-00000 --update-key order_id
```

scenario 5:

If "--update-key" attribute is used along with "--update-mode allowinsert" then it will not only update existing data on specified column but also insert new rows.

```
sqoop export --connect jdbc:mysql://localhost:3306/retail_db --username root --password  
cloudera --table orders_data --export-dir /sqoop_practice_m1/part-m-00000 --update-key order_id  
--update-mode allowinsert
```

Mysql Database :

Creating a table from existing table for practical:

```
mysql> create table orders_data as select * from orders;
```

To see the schema in Mysql :

```
mysql> describe orders;
```

Field	Type	Null	Key	Default	Extra
order_id	int(11)	NO	PRI	NULL	auto_increment
order_date	datetime	NO		NULL	
order_customer_id	int(11)	NO		NULL	
order_status	varchar(45)	NO		NULL	

```
4 rows in set (0.11 sec)
```

Sqoop Export command Execution:

Exporting the data to Mysql orders_Data table from the file created in Sqoop_practice_1 folder

```
[cloudera@quickstart ~]$ sqoop export --connect jdbc:mysql://localhost:3306/retail_db --username root --password cloudera
--table orders_data --export-dir /sqoop_practice_m1/part-m-00000
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
22/06/16 08:56:26 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.13.0
22/06/16 08:56:26 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
22/06/16 08:56:26 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
22/06/16 08:56:26 INFO tool.CodeGenTool: Beginning code generation
22/06/16 08:56:27 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `orders_data` AS t LIMIT 1
22/06/16 08:56:27 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `orders_data` AS t LIMIT 1
22/06/16 08:56:27 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-cloudera/compile/a0c71c9e57bf0585d255fac898bee458/orders_data.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
22/06/16 08:56:29 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/a0c71c9e57bf0585d255fac898bee458
/orders_data.jar
22/06/16 08:56:29 INFO mapreduce.ExportJobBase: Beginning export of orders_data
22/06/16 08:56:29 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
22/06/16 08:56:31 INFO Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapred
duce.reduce.speculative
22/06/16 08:56:31 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapred
uce.map.speculative
```

Sqoop Export logs:

Login to the highlighted URL to see the Sqoop Job status

```

File Edit View Search Terminal Help
22/06/16 08:56:27 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `orders_data` AS t LIMIT 1
22/06/16 08:56:27 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `orders_data` AS t LIMIT 1
22/06/16 08:56:27 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-cloudera/compile/a0c71c9e57bf0585d255fac898bee458/orders_data.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
22/06/16 08:56:29 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/a0c71c9e57bf0585d255fac898bee458
/orders_data.jar
22/06/16 08:56:29 INFO mapreduce.ExportJobBase: Beginning export of orders_data
22/06/16 08:56:29 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
22/06/16 08:56:31 INFO Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapre
duce.reduce.speculative
22/06/16 08:56:31 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduc
e.map.speculative
22/06/16 08:56:31 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
22/06/16 08:56:31 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
22/06/16 08:56:41 INFO input.FileInputFormat: Total input paths to process : 1
22/06/16 08:56:41 INFO input.FileInputFormat: Total input paths to process : 1
22/06/16 08:56:41 INFO mapreduce.JobSubmitter: number of splits:4
22/06/16 08:56:41 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduc
e.map.speculative
22/06/16 08:56:42 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1655378496806_0012
22/06/16 08:56:44 INFO impl.YarnClientImpl: Submitted application application_1655378496806_0012
22/06/16 08:56:44 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_16553784968
06_0012/
22/06/16 08:56:44 INFO mapreduce.Job: Running job: job_1655378496806_0012
22/06/16 08:56:52 INFO mapreduce.Job: Job job_1655378496806_0012 running in uber mode : false
22/06/16 08:56:52 INFO mapreduce.Job: map 0% reduce 0%
22/06/16 08:57:09 INFO mapreduce.Job: map 25% reduce 0%
22/06/16 08:57:15 INFO mapreduce.Job: map 75% reduce 0%
22/06/16 08:57:18 INFO mapreduce.Job: map 100% reduce 0%
22/06/16 08:57:20 INFO mapreduce.Job: Job job_1655378496806_0012 completed successfully
22/06/16 08:57:20 INFO mapreduce.Job: Counters: 30

```

Once you login to the URL below page will be opened and click on Tasktype →map→logs

MapReduce Job job_1655378496806_0012 - Mozilla Firefox

Cloudera Live : Welcom... x MapReduce job job_165... x

quickstart.cloudera:19888/jobhistory/job/job_1655378496806_0012

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

Logged in as: dr.who

MapReduce Job job_1655378496806_0012

Application
Job
Overview
Counters
Configuration
Map tasks
Reduce tasks
Tools

Job Overview

Job Name: orders_data.jar

User Name: cloudera

Queue: root.cloudera

State: SUCCEEDED

Uberized: false

Submitted: Thu Jun 16 08:56:43 PDT 2022

Started: Thu Jun 16 08:56:51 PDT 2022

Finished: Thu Jun 16 08:57:16 PDT 2022

Elapsed: 24sec

Diagnostics:

Average Map Time 17sec


ApplicationMaster			
Attempt Number	Start Time	Node	Logs
1	Thu Jun 16 08:56:47 PDT 2022	quickstart.cloudera:8042	logs

Task Type	Total	Complete	
Map	4	4	
Reduce	0	0	
Attempt Type	Failed	Killed	Successful
Maps	0	0	4

Cloudera Live : Welcom... x Map Tasks for job_1655... x

quickstart.cloudera:19888/jobhistory/tasks/job_1655378496806_001 Search

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started



Map Tasks for job_1655378496806_0012

Application

Job

- Overview
- Counters
- Configuration
- Map tasks
- Reduce tasks

Tools


Show 20 entries

Task						Suc
Name	State	Start Time	Finish Time	Elapsed Time	Start Time	
task_1655378496806_0012_m_000000	SUCCEEDED	Thu Jun 16 08:56:54 -0700 2022	Thu Jun 16 08:57:07 -0700 2022	13sec	Thu Jun 16 08:56:54 -0700 2022	
task_1655378496806_0012_m_000001	SUCCEEDED	Thu Jun 16 08:56:55 -0700 2022	Thu Jun 16 08:57:13 -0700 2022	18sec	Thu Jun 16 08:56:55 -0700 2022	
task_1655378496806_0012_m_000002	SUCCEEDED	Thu Jun 16 08:56:56 -0700 2022	Thu Jun 16 08:57:13 -0700 2022	17sec	Thu Jun 16 08:56:56 -0700 2022	
task_1655378496806_0012_m_000003	SUCCEEDED	Thu Jun 16 08:56:57 -0700 2022	Thu Jun 16 08:57:16 -0700 2022	19sec	Thu Jun 16 08:56:57 -0700 2022	

ID State Start Time Finish Time Elapsed Start Time

Showing 1 to 4 of 4 entries

Click logs to see the errors on both succeeded or failure cases doing sqoop export execution.



Attempts for task_1655378496806_0012_m_0012

Application

Job

Task

- Task
- Overview
- Counters

Tools

Show 20 entries

Attempt	State	Status	Node	Logs	Start Time
attempt_1655378496806_0012_m_000000_0	SUCCEEDED	map	/default-rack/quickstart.cloudera:8042	logs	Thu Jun 16 08:56:54 -0700 2022

Attempt State Status Node Logs Start Time

Showing 1 to 1 of 1 entries