Apache Sqoop Cheat Sheet

Sqoop Import	
| list-databases | ```
sqoop list-databases \
--connect "jdbc:mysql://quickstart.cloudera:3306" \
--username retail_dba \
--password cloudera
``` |
| list-tables | ```
sqoop list-tables \
--connect "jdbc:mysql://quickstart.cloudera:3306/retail_ABC" \
--username retail_dba \
--password cloudera
``` |
| Sqoop eval Command | ```
sqoop eval \
  --connect "jdbc:mysql://quickstart.cloudera:3306/retail_ABC" \
  --username retail_dba \
  --password cloudera \
  --query "select count(1) from order_items"
``` |
| import-all-tables | ```
sqoop import-all-tables \
  -m 12 \
  --connect "jdbc:mysql://quickstart.cloudera:3306/retail_ABC" \
  --username=retail_dba \
  --password=cloudera \
  --as-avrodatafile \
  --warehouse-dir=/user/hive/warehouse/retail_stage.db
``` |
| Import table as text file | ```
sqoop import \
  --connect "jdbc:mysql://quickstart.cloudera:3306/retail_ABC" \
  --username=retail_dba \
  --password=cloudera \
  --table departments \
  --as-textfile \
  --target-dir=/user/cloudera/departments
``` |
| Import table as Sequence file | ```
sqoop import \
  --connect "jdbc:mysql://quickstart.cloudera:3306/retail_ABC" \
  --username=retail_dba --password=cloudera \
  --table departments \
  --as-sequencefile \
  --target-dir=/user/cloudera/departments
``` |

Import table as AVRO data file	```
	sqoop import \
	  --connect "jdbc:mysql://quickstart.cloudera:3306/retail_ABC" \
	  --username=retail_dba \
	  --password=cloudera \
	  --table departments \
	  --as-avrodatafile \
	  --target-dir=/user/cloudera/departments
	``

Import-all-tables with Hive Import	```
	sqoop import-all-tables \
	  --num-mappers 1 \
	  --connect "jdbc:mysql://quickstart.cloudera:3306/retail_ABC" \
	  --username=retail_dba \
	  --password=cloudera \
	  --hive-import \
	  --hive-overwrite \
	  --create-hive-table \
	  --compress \
	  --compression-codec org.apache.hadoop.io.compress.SnappyCodec \
	  --outdir java_files
	``

Import selective columns	```
	sqoop import \
	  --connect "jdbc:mysql://quickstart.cloudera:3306/retail_ABC" \
	  --username=retail_dba \
	  --password=cloudera \
	  --table departments \
	  --target-dir /user/cloudera/departments \
	  -m 2 \
	  --columns department_id,department_name
	``

Import with Query and split by	```
	sqoop import \
	  --connect "jdbc:mysql://quickstart.cloudera:3306/retail_ABC" \
	  --username=retail_dba \
	  --password=cloudera \
	  --query="select * from orders join order_items on
	orders.order_id = order_items.order_item_order_id where
	\$CONDITIONS" \
	  --target-dir /user/cloudera/order_join \
	  --split-by order_id \
	  --num-mappers 4
	``

| Import with Append Mode | ```
sqoop import \
  --connect "jdbc:mysql://quickstart.cloudera:3306/retail_ABC" \
  --username=retail_dba \
  --password=cloudera \
  --table departments \
  --target-dir /user/hive/warehouse/retail_ods.db/departments \
  --append \
  --fields-terminated-by '|' \
  --lines-terminated-by '\n' \
  --num-mappers 1 \
  --outdir java_files
``` |
| Import with Append mode & split-by | ```
sqoop import \
  --connect "jdbc:mysql://quickstart.cloudera:3306/retail_ABC" \
  --username=retail_dba \
  --password=cloudera \
  --table departments \
  --target-dir /user/hive/warehouse/retail_ods.db/departments \
  --append \
  --fields-terminated-by '|' \
  --lines-terminated-by '\n' \
  --split-by department_id \
  --outdir java_files
``` |
| Import with Selective records | ```
sqoop import \
  --connect "jdbc:mysql://quickstart.cloudera:3306/retail_ABC" \
  --username=retail_dba \
  --password=cloudera \
  --table departments \
  --target-dir /user/hive/warehouse/retail_ods.db/departments \
  --append \
  --fields-terminated-by '|' \
  --lines-terminated-by '\n' \
  --split-by department_id \
  --where "department_id > 7" \
  --outdir java_files
``` |

| Import with incremental load | ```
sqoop import \
  --connect "jdbc:mysql://quickstart.cloudera:3306/retail_ABC" \
  --username=retail_dba \
  --password=cloudera \
  --table departments \
  --target-dir /user/hive/warehouse/retail_ods.db/departments \
  --append \
  --fields-terminated-by '|' \
  --lines-terminated-by '\n' \
  --check-column "department_id" \
  --incremental append \
  --last-value 7 \
  --outdir java_files
``` |
| Create HIVE table while sqoop import | ```
sqoop import \
  --connect "jdbc:mysql://quickstart.cloudera:3306/retail_ABC" \
  --username=retail_dba \
  --password=cloudera \
  --table departments \
  --fields-terminated-by '|' \
  --lines-terminated-by '\n' \
  --hive-home /user/hive/warehouse \
  --hive-import \
  --hive-table departments_test \
  --create-hive-table \
  --outdir java_files
``` |
| Overwrite existing Hive table | ```
sqoop import \
  --connect "jdbc:mysql://quickstart.cloudera:3306/retail_ABC" \
  --username=retail_dba \
  --password=cloudera \
  --table departments \
  --fields-terminated-by '|' \
  --lines-terminated-by '\n' \
  --hive-home /user/hive/warehouse/retail_ods.db \
  --hive-import \
  --hive-overwrite \
  --hive-table departments \
  --outdir java_files
``` |

Sqoop Job

Create sqoop job	```sqoop job --create sqoop_job \
-- import \
--connect "jdbc:mysql://quickstart.cloudera:3306/retail_db" \
--username=retail_dba \
--password=cloudera \
--table departments \
--target-dir /user/hive/warehouse/retail_ods.db/departments \
--append \
--fields-terminated-by '|' \
--lines-terminated-by '\n' \
--check-column "department_id" \
--incremental append \
--last-value 7 \
--outdir java_files```
list all sqoop job	```sqoop job -list```
Display the sqoop_job details	```sqoop job --show sqoop_job```
Execute the sqoop job	```sqoop job --exec sqoop_job```
Delete the sqoop job	```Sqoop job -delete sqoop_job```