Sqoop Q&A

1. How to import BLOB and CLOB-like big objects in Sqoop?

We can use JDBC-based imports for BLOB and CLOB, as Sqoop does not support direct import function.

1. Basic sqoop commands

Export

Import

List-databases

Import-all-tables

Versions

Eval

Create-hive-table

List-tables

1. How can you execute a free-form SQL query in Sqoop to import the rows in a sequential manner?

By using the –m 1 option in the Sqoop import command we can accomplish it. Basically, it will create only one MapReduce task which will then import rows serially.

1. Does Apache Sqoop have a default database?

Yes, MySQL is the default database.

1. How to list all tables using sqoop in a particular schema?

sqoop list-tables –connect jdbc:postgresql://localhost/corp –username name -password password –schema payrolldept

1. How to list all columns of a table in sqoop?

Sqoop import

–m 1

–connect ‘jdbc: sqlserver: //nameofmyserver; database=nameofmydatabase; username=DeZyre;

password=mypassword’

–query “SELECT column\_name, DATA\_TYPE FROM INFORMATION\_SCHEMA.Columns WHERE table\_name=’mytableofinterest’ AND \$CONDITIONS”

–target-dir ‘mytableofinterest\_column\_name’

1. What is the use of sqoop-eval?

It is used to run simple sql queries like DML and DDL and preview the output as well.

1. What are the options for incremental load?

Append and lastmodifieddate(attributes : mode, column, last value)

1. When to use –target-dir and when to use –warehouse-dir while importing data?

Basically, we use –target-dir to specify a particular directory in HDFS. Whereas we use –warehouse-dir to specify the parent directory of all the sqoop jobs. So, in this case under the parent directory sqoop will create a directory with the same name as the table.

1. How can you automate passing password to your sqoop commands?

Use a password file instead of -p option

1. How can you control the mapping between SQL data types and Java types?

sqoop import … –map-column-java id = String, value = Integer

1. What is the significance of using –compress-codec parameter?

We use the –compress -code parameter to get the out file of a sqoop import in formats other than .gz like .bz2.

1. Sqoop import with incremental load

sqoop import \

--connect jdbc:mysql://localhost/userdb \

--username root \

--table emp \

--m 1 \

--incremental append \

--check-column id \

-last value 1205

1. How to import all tables from Rdbms to HDFS using sqoop?

$ sqoop import-all-tables \

--connect jdbc:mysql://localhost/userdb \

--username root

1. Sqoop export example

sqoop export \

--connect jdbc:mysql://localhost/db \

--username root \

--password root \

--table employee \

--export-dir /emp/emp\_data

1. Write sqoop statement to update an existing table and also insert if new data is found

sqoop export --connect jdbc:mysql://localhost/db \

--username root \

--password root \

--table employee \

--update-mode allowinsert \

--update-key id \

--export-dir /emp/emp\_data \

--input-fields-terminated-by ',';