1. Features of sqoop

Sqoop is used to transfer data to and fro hdfs,hive - mysql,rdbms

1. Purpose of sqoop import

To bring data from rdbms to hdfs

1. What is the default file format to import data using Apache Sqoop?

Delimited and sequence file

1. How can I import large objects (BLOB and CLOB objects) in Apache Sqoop?

Without using direct, via jdbc

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

By using only one mapper (--m -1)

## Sqoop Import Syntax

Sqoop import

--connect

--username

--password

--target-dir

1. Sqoop Import – Import control arguments
2. Sqoop import example
3. Sqoop import hive connection - https://data-flair.training/blogs/sqoop-import/
4. Default database of apache sqoop
5. How will you list all the columns of a table using Apache Sqoop?
6. Incremental parameter in sqoop
7. How can we import data from particular row or column? What is the destination types allowed in Sqoop import command?
8. When to use –target-dir and when to use –warehouse-dir while importing data?
9. What is the significance of using –compress-codec parameter?
10. Can free-form SQL queries be used with Sqoop import command? If yes, then how can they be used?
11. Eval tool
12. How to use where command in sqoop query
13. What is sqoop hcatalog
14. How to parameterise a password and pass it separately
15. Use of jdbc driver
16. Is JDBC driver enough to connect sqoop to the databases? - <https://data-flair.training/blogs/sqoop-connectors-drivers/>
17. How to divide input files in to smaller parts - InputSplit
18. Sqoop export - <https://data-flair.training/blogs/sqoop-export/>
19. Sqoop codegen - <https://data-flair.training/blogs/sqoop-codegen/>
20. How to see all available commands in sqoop - help
21. Sqoop schedule in oozie - <https://data-flair.training/blogs/sqoop-job/>
22. What is the importance of — the split-by clause in running parallel import tasks in sqoop?
23. What is a sqoop metastore?
24. What is the purpose of sqoop-merge?
25. How can you see the list of stored jobs in sqoop metastore?
26. Which database the sqoop metastore runs on? Where can it be hosted
27. Give the sqoop command to see the content of the job named myjob?
28. Is it possible to add a parameter while running a saved job?
29. What is the usefulness of the options file in sqoop.
30. How to import all tables at once - https://data-flair.training/blogs/sqoop-import-all-tables/
31. How can you control the number of mappers used by the sqoop command?
32. What is the default extension of the files produced from a sqoop import using the –compress parameter? What are other common formats?
33. What is a disadvantage of using –direct parameter for faster data load by sqoop?
34. How will you update the rows that are already exported?
35. All commands in sqoop
36. Sqoop = sql+hadoop
37. What is Sqoop Validation? - <https://data-flair.training/blogs/sqoop-validation/>
38. What is Sqoop Job?
39. Difference Between Apache Sqoop vs Flume.
40. How to choose no of mappers for a job

<https://data-flair.training/blogs/sqoop-interview-questions/>

Sqoop is map only job

Commands:

Sqoop Import

--------------

Read RDBMS(MYSQL) data

bin/sqoop-eval

--connect jdbc:mysql://localhost/test

-username root

-password root

--query "SELECT \* FROM demo"

MYSQL to HDFS

bin/sqoop import

--connect jdbc:mysql://localhost/test

-username root

-password root

--table test2 -m1

hadoop fs -ls

--target-dir

--sqoop import-all-tables

MYSQL to Hive

bin/sqoop-import

--connect jdbc:mysql://localhost/test

-username root

-password root

--table test

--hive-table mysqltohive2

--create-hive-table

--hive-import -m1

MYSQL to HBase

bin/sqoop import

--connect jdbc:mysql://localhost/test

--username root

--password root

--table demo

--hbase-table testhbase

--column-family cf

--hbase-row-key pid

--hbase-create-table -m1

Sqoop Export

------------

HDFS to MYSQL

bin/sqoop export

--connect jdbc:mysql://localhost/test

--username root

-password root

--table name\_3

--export-dir /user/hive/warehouse/test.db/name.txt -m1

Hive to MYSQL

bin/sqoop export

--connect jdbc:mysql://localhost/test

--table test\_hive\_1

--export-dir /user/hive/warehouse/test.db/demo

--username root

-password root -m1

Connection-manager

Driver

P

Password-file

Validate

Append

As-avrofile

As-sequencefile

As-textfile

As-parquetfile

Fetch-size

M,num mappers

E,query

-target-dir

-warehouse-dir

-where

-compress

-compression-codec

Import export is done in parallel

It can do incremental and full load

How to check no of records in rdbms before importing

Sqoop eval

--connect jdbc:localhost/databasename

--username

--password

--query

How to test a connection

Use sqoop eval

How to use eval

How to import all tables

Sqoop-import

--connect jdbc:localhost/databasename

--username

--password

--import-all-tables

How to export all tables

Sqoop-export

--connect jdbc:localhost/databasename

--username

--password

--table

--export-dir

How to exclude certain tables while importing/exporting

--except table names

What are widely used commands in sqoop

How to change no of mappers

--num-of-mappers

--m

What compression formats are supported and how to use it

.gzip,snappy,bzip

sqoop import \

--connect jdbc:mysql://localhost:3306/retail\_db \

--username root \

--password mysqlrootpassword \

--driver com.mysql.cj.jdbc.Driver \

--table orders \

--target-dir hdfs://localhost:9000/user/username/scoop\_import/avro\_zip \

--append \

--compression-codec Gzip

Purpose of sequence file

How to pass password secretly

--password-file

Diff between warehouse and target dir

Target-dir is location of file and warehouse is where all tables are stored in sqoop

--append

To append new data in existing

--boundary-query is used to fetch min and max and it requires single mapper or split-by clause

Incremental:

--check-column order\_id \

--incremental append \

--last-value 1022 \