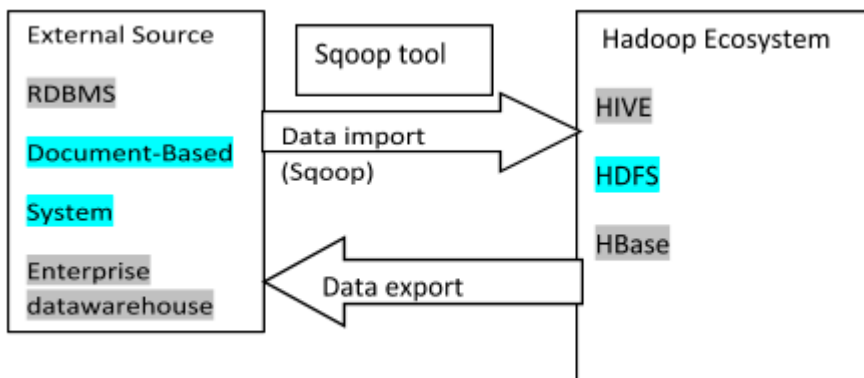# By OnlineInterviewQuestions.com

Sqoop Interview Questions

Q1. What is Sqoop?

The sqoop is an acronym of **SQL-TO-HADOOP**. It is a command-line interface application. It is worked for conveying data between relational databases (**MySQL/ PostgreSQL/Oracle/SQL Server/DB2**) and Hadoop (**HIVE, HDFS, HBase**). The sqoop has community support and contributions, therefore this is robust.

Q2. What is sqoop used for?

The sqoop is primarily conveying the immense data between the relational database and the Hadoop ecosystem. The entire database or individual table import to the ecosystem (**HDFS**) and after modification export to the database. The sqoop helps to support the multiple loads in one database table.

Q3. List some major features of Apache Sqoop?

The following list of apache sqoop features.

- Compression
- Connectors for all major RDBMS databases
- Kerberos security integration
- Full load
- Incremental load
- Import results of database query
- Load data directly into hive and HBase
- Parallel import and export connection
- Support for accumulo

Q4. **What is use of Sqoop Export?**

The data is a transfer from the **HDFS** to **RDBMS** (**relational database**), called sqoop export. Before transforming the data, sqoop tool fetch table from the database. Therefore the table must be available in the database.

The syntax of export is below.

```
$ sqoop export (generic-args) (export-args)
$ sqoop-export (generic-args) (export-args)
```

Q5. **What is --direct mode in sqoop?**

The sqoop is used for **Hadoop** and **database connection** but has some stages. **The --directive mode** in scoop used for directly import multiple table or individual table into **HIVE**, **HDFS**, **HBase**. If we have a specific database connection directly apart from default database connection then –directive mode used.

Q6. **What are reducers in sqoop?**

The reducer is used for **accumulation** or **aggregation**. After mapping, the reducer fetches the data transfer by the database to Hadoop. In the sqoop there is no reducer because import and export work parallel in sqoop.

Q7. **What is boundary query in sqoop?**

The boundary query is used for splitting the value according to **id_no** of the database table.

To boundary query, we can take a minimum value and maximum value to split the value.

To make split using boundary queries, we need to know all the values in the table.

To import data from the database to HDFS using boundary queries.

Example--boundary-query

```
"SELECT min(id_value), max(id_value) from table_name"
```

Q8. What is sqoop import?

The sqoop import is helped to import table data to the Hadoop file system in the form of text or binary.

The syntax of sqoop import is below.

```
$sqoop import (generic-args) (import-args)
$sqoop-import (generic-args) (import-args)
$ sqoop import / --connect jdbc:mysql:
// localhost/ database / --username user_name / --table table_name --m */ --targe
```

Q9. Explain relaxed isolation in sqoop?

The –relaxed-isolation is the argument of import sqoop. This is used to import the data which is read uncommitted for mappers. The sqoop transfer committed data relational database to the Hadoop file system but with this argument, we can transfer uncommitted data in the isolation level.

Q10. How are large objects handled in Sqoop?

If the object is less than **16mb** then it is stored with other common size data. **The large objects** are **CLOBs** and **BLOBs**. The large objects are handled by import the large object into **LobFile** means a large object file. The LobFile is an external storage file that can store records of large objects.

Q11. What is sqoop Metastore?

The metastore is a tool that is used to **share data** or **/and repository**. The multiple users can create a job and load some data other hand remote users access than a job and run again this job. The metastore in sqoop is the central place for job information.

It worked for users and developers to make collaboration.

Q12. What is default file format in Sqoop?

The default file type is a **text file format**. It is the same as specifying –as-textfile clause to sqoop import command.

Q13. List some sqoop commands you have used?

In the sqoop mostly import and export command are used. Apart from these two other commands are used.

This list is below

- codegen
- eval
- import-all-tables
- job, list-database
- list-tables
- Merge
- metastore

Q14. What is accumulo in sqoop?

The Accumulo in sqoop is a sorted, distributed key and value store. It provides robust, extensible data storage and retrieves data. This is stable and it has own security for key and value. A large amount of data store, retrieve and manage the **HDFS** data.

Q15. What is sqoop eval?

The eval tool in sqoop used for sample demo for import data. It is permit user to run the sample RDBMS queries and examine the results on the console. Because of the eval tool, we know what is output and what kind of data import.

Q16. How to check sqoop version in hortonworks?

The following command is used to know the version in hortonworks.

```
# ssh username@127.0.0.1  -p 2222
Enter password: hadoop[username@sandbox ~]# sqoop version
```

Q17. Does sqoop use MapReduce?

Mapreduce used in sqoop for parallel import and export the data between database and Hadoop file system.

It is used for fault resistance.

Q18. What is the use of split by in sqoop?

The sqoop import/export parallel, data can split into multiple chunks to transfer. **The Split** by in sqoop selects the **id_number** to split a column of the table.

```
sqoop import --connect jdbc:mysql:// localhost/database_name --username user_name
--query 'select * from table_name where id=3 AND $CONDITIONS' --split-by table.id
```

the split by helped to proper distribution to make a split of data.

Q19. Why do we use $conditions in sqoop?

The condition comes up with split but split automatically decides which slice of data transfers as every task. Condition force to run only one job ar a time and gives mapper to transfer data without any attack.

```
sqoop import --connect jdbc:mysql:// localhost/database_name --username user_name
--query 'select * from table_name where id=3 AND $CONDITIONS' --split-by table.id
```

Q20. How to do job scheduling in sqoop?

The sqoop can create Oozie workflow jobs Oozie has in-built sqoop actions inside, where the sqoop commands are executed. sqoop **job –create job_name**.

Please Visit OnlineInterviewquestions.com to download more pdfs