Apache Sqoop Interview Questions and Answers

Q: What is Apache Sqoop?
A: A tool to transfer data between Hadoop and relational databases.

Q: What are the key features of Sqoop?
A: Parallel import/export, connectors for major RDBMS, incremental load, Hive/HBase integration,

Q: What is the default file format used by Sqoop?
A: Delimited text file.

Q: What is the use of Sqoop import?
A: To import data from RDBMS to HDFS.

Q: What is the use of Sqoop export?
A: To export data from HDFS to RDBMS.

Q: What is incremental import in Sqoop?
A: Importing only new or updated data using --incremental append or --incremental lastmodified.

Q: What is the difference between append and lastmodified mode?
A: Append imports new rows; lastmodified imports updated rows.

Q: How do you perform a full import using Sqoop?
A: Without using the --incremental option.

Q: What is the use of --split-by in Sqoop?
A: Defines the column used to split data for parallel processing.

Q: What is the use of --target-dir?
A: Specifies the HDFS directory to store imported data.

Q: How can you import data into Hive using Sqoop?
A: Use --hive-import or --hive-table.

Q: How can you import data into HBase using Sqoop?
A: Use --hbase-table and --column-family.

Q: What is the use of --direct mode in Sqoop?
A: Enables database-specific fast import/export.