1. **HOW DO YOU ACHIEVE THE INCREMENTAL LOADING USING SQOOP?**

Sqoop supports two types of incremental imports: *append* and *lastmodified*. You can use the *–incremental* argument to specify the type of incremental import to perform.

* *append* : You should specify the append mode, where new rows are continually added with increasing row id values. You must specify the column containing the row’s id with *–check-column*. Sqoop imports rows where the check column has a value greater than the one specified with –last-value.
* *lastmodified* : This should be used when rows of the source table is updated, and each such update will set the value of a last-modified column to the current timestamp. Rows where the check column holds a timestamp more recent than the timestamp specified with *–last-value* are imported.

1. **WHEN TO USE "--split-by" OPTION OF SQOOP...EXPLAIN THE SCENARIO?**

If your table has no index column, or has a multi-column key, then you must also manually choose a splitting column. If a table does not have a primary key defined and the --split-by <col> is not provided, then import will fail unless the number of mappers is explicitly set to one with the --num-mappers 1.

1. **WHAT IS THE USE OF "--update-key" IN SQOOP...WHERE DO WE USE THAT?**

By default, sqoop-export appends new rows to a table; each input record is transformed into an INSERT statement that adds a row to the target database table. If your table has constraints and already contains data, you must take care to avoid inserting records that violate these constraints.

If you specify the --update-key argument, Sqoop will instead modify an existing dataset in the database. Each input record is treated as an UPDATE statement that modifies an existing row. The row a statement modifies is determined by the column name(s) specified with --update-key.

1. **WHAT IS THE USE OF "--direct" IN SQOOP....CAN WE MAKE USE OF THIS OPTION IN ALL DATABASES...WHAT ARE THE LIMITIONS**

For MySQL and PostgreSQL it relates to bulk loader utilities (i.e. completely bypassing JDBC) and for Oracle it relates to "direct path INSERT" i.e. with JDBC but in a non-transactional mode (so you'd better use a temp table, or you might end up with duplicates in a PK and a corrupt table). To be short, its the mode for fast import which doesn't runs any mappers or reducers.

LIMITATIONS:

1. --direct is only supported in mysql and postgresql.
2. Sqoop’s direct mode does not support imports of BLOB, CLOB, or LONGVARBINARY columns.
3. **HOW DO YOU HANDLE THE SQOOP IMPORT LEVEL FAILURES DUE TO NETWORK FAILURES?**

If sqoop fails in the middle of the data transfer, the job has to start from first. Before starting the job existing “target directory” need to delete or need to give new “target directory”.

1. **TO IMPROVE THE PERFOMANCE OF SQOOP IMPORT , WHAT ARE THE OPTIONS IN SQOOP?**

* **–num-mappers**: Use “n” map tasks to import in parallel.
* **–split-by**: Column of the table used to split work units across the mappers.
* **–boundary-query:** Boundary query to use for creating splits.
* **–fetch-size**: Number of entries to read from database at once (Default fetch size is 1000 records at a time)
* **–direct**: Use direct connector if exists for the database

1. **CAN WE DIRECTLY EXPORT THE DATA TO A "STAGING TABLE" USING SQOOP?**

Yes we can export the data to “STAGING TABLE” with mentioning the respective DB and table information.

1. **HOW DO WE ACHIEVE THE PASSWORD LESS SQOOP JOB EXECUTION...? IF WE WANT TO HIDE THE PASSWORD, ARE THERE ANY OPTIONS?**

* --password-file & --options-file: Used to read password from file
* -P: Used to read password from console at runtime (User need to enter while running the sqoop job).
* Oozie: Configure the password in oozie “job.properties” file, while running the sqoop job with oozie.

1. **WHAT IS THE USE OF "--lastmodified" OPTION IN SQOOP?**

This should be used when rows of the source table is updated, and each such update will set the value of a last-modified column to the current timestamp. Rows where the check column holds a timestamp more recent than the timestamp specified with *–last-value* are imported.