

Big data

02 December 2024 19:19

Sqoop cmd to import

```
sqoop import \  
--connect "jdbc:mysql://quickstart.cloudera:3306/retail_db" \  
--username root \  
--password cloudera \  
--table categories \  
--target-dir /user/cloudera/nov151
```

To list mysql databases using sqoop:

```
sqoop list-databases --connect "jdbc:mysql://quickstart.cloudera:3306" --username root --password cloudera
```

To list mysql table

```
sqoop list-tables --connect "jdbc:mysql://quickstart.cloudera:3306/retail_db" --username root --password cloudera
```

To list onlu specified columns of a table:

```
sqoop import \  
--connect "jdbc:mysql://quickstart.cloudera:3306/retail_db" \  
--username root \  
--password cloudera \  
--table customers \  
--columns "customer_id,customer_fname,customer_lname" \  
--target-dir /user/cloudera/nov18
```

Transferring table without primary key - specifying mappers:

```
sqoop import \  
--connect "jdbc:mysql://quickstart.cloudera:3306/retail_db" \  
--username root \  
--password cloudera \  
--table sumit \  
--m 1 \  
--target-dir /user/cloudera/nov20
```

Using splitby-by when no primary key is there - specifying id

```
sqoop import \  
--connect "jdbc:mysql://quickstart.cloudera:3306/retail_db" \  
--username root \  
--password cloudera \  
--table sumit \  
--split-by id \  
--target-dir /user/cloudera/oct2
```

To list files in hdfs:

```
Hadoop fs - ls /ser/cloudera/nov15
```

To view file data:

```
Hadoop fs - cat /ser/cloudera/nov15
```

Warehouse dirc:

```
sqoop import \  
--connect "jdbc:mysql://quickstart.cloudera:3306/retail_db" \  
--username root \  
--password cloudera \  
--table customers \  
--columns "customer_id,customer_fname,customer_lname" \  
--warehouse-dir /user/cloudera/oct3
```

Transdfer all tables to hdfs:

```
Sqoop import -all-tables \  
--connect "jdbc:mysql://quickstart.cloudera:3306/retail_db" \  
--username root \  
--password cloudera \  
--columns "customer_id,customer_fname,customer_lname" \  
--warehouse-dir /user/cloudera/oct4
```

Import all and exclude tables:

```
sqoop import-all-tables \  
--connect "jdbc:mysql://quickstart.cloudera:3306/retail_db" \  
--username root \  
--password cloudera \  
--exclude-tables sumit , table1 \
```

Mysql cmds:

```
mysql -u root -p
```

```
Use retails_db;  
Show tables;
```

```
Create table vaishali;  
Select * from vaishali;
```

```
Insert into vaishali(id int, name varchar(10));
```

```
Create table sumit(id int, name varchar(10));  
Insert into sumit values(1,"a"),(2,"b");
```

```
Desc categories
```

```
--warehouse-dir /user/cloudera/alldoc99
```

When id are not in right manner:

```
--boundary -query "select 1,7"
```

Conditional query:

```
--query "select * from sumit where age>75"
```

```
sqoop import \  
--connect "jdbc:mysql://quickstart.cloudera:3306/retail_db" \  
--username root \  
--password cloudera \  
--query "SELECT * from categories WHERE category_id>5 AND\ $CONDITIONS" \  
--split-by category_id \  
--target-dir /user/cloudera/nov1
```

VERBOSE : TO see the logs in details:

```
sqoop import \  
--connect "jdbc:mysql://quickstart.cloudera:3306/retail_db" \  
--username root \  
--password cloudera \  
--verbose \  
--query "SELECT * from categories WHERE category_id>5 AND\ $CONDITIONS" \  
--split-by category_id \  
--target-dir /user/cloudera/nov1
```

Incremental load

Append mode:

```
sqoop import \  
--connect "jdbc:mysql://quickstart.cloudera:3306/retail_db" \  
--username root \  
--password cloudera \  
--table rohan \  
--check-column id \  
--incremental append \  
--last-value 4 \  
--target-dir /user/cloudera/rohan1
```

Last modified mode:

```
CREATE TABLE employees (id INT PRIMARY KEY, name VARCHAR(50), salary INT, last_modified  
TIMESTAMP);  
INSERT INTO employees VALUES (1, 'mohan', 50000, '2023-01-01 10:00:00');  
INSERT INTO employees VALUES (2, 'veena', 60000, '2023-01-02 11:30:00');
```

```
Sqoop import \  
--connect "jdbc:mysql://quickstart.cloudera:3306/retail_db" \  
--username root \  
--password cloudera \  
--table employees \  
--incremental lastmodified \  
--check-column last_modified \  
--last-value "2023-01-01 10:00:00" \  
--merge-key id \  
--target-dir /user/cloudera/oct22
```

```
sqoop import \  
--connect "jdbc:mysql://quickstart.cloudera:3306/retail_db" \  
--username root \  
--password cloudera \  
--table rohan \  
--check-column id \  
--incremental append \  
--last-value 4 \  
--target-dir /user/cloudera/rohan1
```