```
-- MySQL Commands

Login to mysql
==================================
mysql -u root -p
mysql -u retail_dba -p

-- SQOOP Commands

List the dataBases
==================================
sqoop-list-databases \
--connect "jdbc:mysql://quickstart.cloudera:3306" \
--username retail_dba \
--password cloudera

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


Show tables
==================================
sqoop-list-tables \
--connect "jdbc:mysql://quickstart.cloudera:3306/retail_db" \
--username retail_dba \
--password cloudera


SQOOP eval
==================================
sqoop-eval \
--connect "jdbc:mysql://quickstart.cloudera:3306" \
--username retail_dba \
--password cloudera \
--query "select * from retail_db.customers limit 10"

Display Schema
==================================
sqoop-eval \
--connect "jdbc:mysql://quickstart.cloudera:3306" \
--username retail_dba \
--password cloudera \
--query "describe retail_db.orders"

or

in mysql
------------
```

```
describe tablename;

Import table from MYSQL to HDFS[With Primary Key]
=================================================================
sqoop import \
--connect "jdbc:mysql://quickstart.cloudera:3306/retail_db" \
--username retail_dba \
--password cloudera \
--table orders \
--target-dir /queryresult

Import table from MYSQL to HDFS[Without Primary Key]
=================================================================
** It will through an error
sqoop import \
--connect "jdbc:mysql://quickstart.cloudera:3306/trendytech" \
--username root \
--password cloudera \
--table people \
--target-dir /peopleresult

Changing mapprer to 1
-----------------------------------------------------------
sqoop import \
--connect "jdbc:mysql://quickstart.cloudera:3306/trendytech" \
--username root \
--password cloudera \
--table people \
-m 1 \
--target-dir /peopleresult


Import All tables
================================================
sqoop-import-all-tables \
--connect "jdbc:mysql://quickstart.cloudera:3306/retail_db" \
--username retail_dba \
--password cloudera \
--as-sequencefile \
--warehouse-dir /user/cloudera/sqoopdir1


Store logs messages in some file
==============================================
sqoop import \
--connect "jdbc:mysql://quickstart.cloudera:3306/retail_db" \
--username retail_dba \
--password cloudera \
--table orders \
--warehouse-dir /queryresult4 1>query.out 2>query.err
```

Compress the file
==========
```
sqoop import \
--connect jdbc:mysql://quickstart.cloudera:3306/retail_db \
--username root \
--password cloudera \
--table orders \
--compress \
--warehouse-dir Allmyfiles

sqoop import \
--connect jdbc:mysql://quickstart.cloudera:3306/retail_db \
--username root \
--password cloudera \
--table orders \
--compression-codec BZip2Codec \
--warehouse-dir Allmyfiles
```

Import only selected columns
==========
```
sqoop import \
--connect jdbc:mysql://quickstart.cloudera:3306/retail_db \
--username root \
--password cloudera \
--table customers \
--columns customer_id,customer_fname,customer_city \
--warehouse-dir Allmyfiles
```

Use Where class

```
sqoop-import \
--connect jdbc:mysql://quickstart.cloudera:3306/retail_db \
--username retail_dba \
--password cloudera \
--table orders \
--columns order_id,order_customer_id,order_status \
--where "order_status='complete'" \
--target-dir UsingWhareClass
```

Customize boundary query
==========
```
sqoop import \
--connect jdbc:mysql://localhost:3306/retail_db \
--username retail_dba \
--password cloudera \
```

```
--table orders \
--boundary-query "select 1,68883" \
--warehouse-dir /orderboundaryval


Customize boundary query with non primary key column
=====================================================
sqoop import \
--connect jdbc:mysql://localhost:3306/retail_db \
--username retail_dba \
-P \
--table order_items \
--boundary-query "select min(order_item_order_id),max(order_item_order_id)
from order_items where order_item_order_id > 10000" \
--warehouse-dir /user/cloudera/bvqresult


Sqoop import using split by
=====================================================
** Create a table without primary key
create table orders_no_pk(
order_id int (11) not null,
order_date datetime not null,
order_customer_id int(11) not null,
order_status varchar (45) not null
);

** Copy table content

insert into orders_no_pk
select order_id, order_date,order_customer_id, order_status from orders;

** Below query fails bz there is no primary key present

sqoop import \
--connect jdbc:mysql://localhost:3306/retail_db \
--username retail_dba \
-P \
--table orders_no_pk \
--warehouse-dir /ordersnopk

sqoop import \
--connect jdbc:mysql://localhost:3306/retail_db \
--username retail_dba \
-P \
--table orders_no_pk \
--split-by order_id \
--warehouse-dir /ordersnopk

Split by on non numeric columns
```

```
==================================
sqoop import \
-Dorg.apache.sqoop.splitter.allow_text_splitter=true \
--connect jdbc:mysql://localhost:3306/retail_db \
--username retail_dba \
-P \
--table orders_no_pk \
--split-by order_status \
--warehouse-dir /ordersnopk
--delete-target-dir


Auto reset to one mapper
========================================
sqoop-import-all-tables \
--connect jdbc:mysql://localhost:3306/retail_db \
--username retail_dba \
-P \
--autoreset-to-one-mapper \
--warehouse-dir /autoresetresult

Delimeters
====================================
sqoop import \
--connect jdbc:mysql://localhost:3306/retail_db \
--username retail_dba \
--password cloudera \
--table orders \
--fields-terminated-by '|' \
--lines-terminated-by ';' \
--warehouse-dir /delimitedresult


Create hive table
==========================================
sqoop create-hive-table \
--connect jdbc:mysql://localhost:3306/retail_db \
--username retail_dba \
--password cloudera \
--table orders \
--hive-table emps \
--fields-terminated-by ','


Sqoop verbose
=========================================
sqoop import \
--connect jdbc:mysql://localhost:3306/retail_db \
--username retail_dba \
--password cloudera \
```

```
--table orders \
--verbose \
--target-dir /verboseresult
```

Sqoop Append

```
==================================================
sqoop import \
--connect jdbc:mysql://localhost:3306/retail_db \
--username retail_dba \
--password cloudera \
--table orders \
--verbose \
--target-dir /user/cloudera/appendresult \
--append
```

Dealing with NULL while importing data

```
==================================================
sqoop import \
--connect jdbc:mysql://localhost:3306/retail_db \
--username retail_dba \
--password cloudera \
--table orders \
--verbose \
--warehouse-dir /user/cloudera/nullstringresult \
--delete-target-dir \
--null-non-string "-1"

sqoop import \
--connect jdbc:mysql://localhost:3306/retail_db \
--username retail_dba \
--password cloudera \
--table orders \
--verbose \
--warehouse-dir /user/cloudera/nullstringresult \
--delete-target-dir \
--null-string "Nullispresent"


create table people(
PersonID int,
LastName varchar(255),
FirstName varchar(255),
Address varchar(255),
City varchar(255)
);
```

```
insert into people values
(101, 'rao', 'mohan', 'whitefield', 'bangalore'),
(102, 'reddy', 'srinivash', 'habala', 'hydetabad'),
(103, 'sharma', 'amit', 'marathali', 'bangalore'),
(104, 'sharma', 'amit', 'majestic', 'hyderabad');
```

SQOOP EXPORT

****

****

```
create database banking;

create table card_transactions(
card_id bigint,
member_id bigint,
amount int(10),
post_id bigint,
transaction_dt varchar(255),
status varchar(255),
primary key (card_id, transaction_dt)
);
```

Sqoop export

==========

```
sqoop export \
--connect jdbc:mysql://localhost:3306/banking \
--username root \
--password cloudera \
--table card_transactions \
--export-dir /data/cardmembers.csv \
--fields-terminated-by ','
```

Staging Table

==========

```
create table card_transactions(
transaction_id int(10),
card_id bigint,
member_id bigint,
amount int(10),
postcode int(10),
pos_id bigint,
transaction_dt varchar(255),
status varchar(255),
primary key (transaction_id)
);
```

```
create table card_transactions_stage(
transaction_id int(10),
card_id bigint,
member_id bigint,
amount int(10),
postcode int(10),
pos_id bigint,
transaction_dt varchar(255),
status varchar(255),
primary key (transaction_id)
);


sqoop export \
--connect jdbc:mysql://localhost:3306/banking \
--username root \
--password cloudera \
--table card_transactions \
--staging-table card_transactions_stage \
--export-dir /data/cardmembers.csv \
--fields-terminated-by ','
```

Sqoop Incremental
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

Append
==========
```
sqoop import \
--connect jdbc:mysql://localhost:3306/retail_db \
--username root \
--password cloudera \
--table orders \
--warehouse-dir /data \
--incremental append \
--check-column order_id \
--last-value 0


21/06/20 06:19:36 INFO tool.ImportTool:  --incremental append
21/06/20 06:19:36 INFO tool.ImportTool:   --check-column order_id
21/06/20 06:19:36 INFO tool.ImportTool:   --last-value 68883
```

```
-- Add below columns and run query

insert into orders values(68884,'2014-07-23 00:00:00',5522,'COMPLETE');
insert into orders values(68885,'2014-07-23 00:00:00',5522,'COMPLETE');
insert into orders values(68886,'2014-07-23 00:00:00',5522,'COMPLETE');
insert into orders values(68887,'2014-07-23 00:00:00',5522,'COMPLETE');
insert into orders values(68888,'2014-07-23 00:00:00',5522,'COMPLETE');
insert into orders values(68889,'2014-07-23 00:00:00',5522,'COMPLETE');
commit;



sqoop import \
--connect jdbc:mysql://localhost:3306/retail_db \
--username root \
--password cloudera \
--table orders \
--warehouse-dir /data \
--incremental append \
--check-column order_id \
--last-value 68883


hadoop fs -cat /data/orders/* | wc -l --> to see No.of lines
hadoop fs -cat /data/orders/* | tail --> to see last the lines which are exist in
last

LastModified
=====================================================================
sqoop import \
--connect jdbc:mysql://localhost:3306/retail_db \
--username root \
--password cloudera \
--table orders \
--warehouse-dir /data \
--incremental lastmodified \
--check-column order_date \
--last-value 0 \
--append


-- Here we are modifying the data and collecting the current time stamp.

insert into orders values(68884,current_timestamp,5523,'COMPLETE');
insert into orders values(68885,current_timestamp,5523,'COMPLETE');
insert into orders values(68886,current_timestamp,5523,'COMPLETE');
insert into orders values(68887,current_timestamp,5523,'COMPLETE');
insert into orders values(68888,current_timestamp,5523,'COMPLETE');
insert into orders values(68889,current_timestamp,5523,'COMPLETE');
```

```
update orders set order_status='COMPLETE'
and order_date = current_timestamp
WHERE ORDER_ID = 68862;

commit;


21/06/20 07:31:41 INFO tool.ImportTool:  --incremental lastmodified
21/06/20 07:31:41 INFO tool.ImportTool:   --check-column order_date
21/06/20 07:31:41 INFO tool.ImportTool:   --last-value 2021-06-20 07:30:54.0


sqoop-import \
--connect jdbc:mysql://quickstart.cloudera:3306/retail_db \
--username root \
--password cloudera \
--table orders \
--warehouse-dir /data \
--incremental lastmodified \
--check-column order_date \
--last-value '2021-06-20 07:30:54.0' \
--append

hadoop fs -cat /data/orders/* | grep 68862 --> it will display 2 records

21/06/20 07:44:58 INFO tool.ImportTool:  --incremental lastmodified
21/06/20 07:44:58 INFO tool.ImportTool:   --check-column order_date
21/06/20 07:44:58 INFO tool.ImportTool:   --last-value 2021-06-20 07:44:14.0


Merge
================================================================
sqoop-import \
--connect jdbc:mysql://quickstart.cloudera:3306/retail_db \
--username root \
--password cloudera \
--table orders \
--warehouse-dir /data \
--incremental lastmodified \
--check-column order_date \
--last-value '2021-06-20 08:39:35.0' \
--merge-key order_id

21/06/20 08:28:48 INFO tool.ImportTool:  --incremental lastmodified
21/06/20 08:28:48 INFO tool.ImportTool:   --check-column order_date
21/06/20 08:28:48 INFO tool.ImportTool:   --last-value 2021-06-20 08:27:04.0


Sqoop Job
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
```

```
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
-- For append

sqoop job \
--create job_orders \
-- import \
--connect jdbc:mysql://localhost:3306/retail_db \
--username root \
--password cloudera \
--table orders \
--warehouse-dir /data \
--incremental append \
--check-column order_id \
--last-value 0

sqoop job --exec job_orders

-- For lastmodified

sqoop job \
--create job_orders_lastmodified \
-- import \
--connect jdbc:mysql://localhost:3306/retail_db \
--username root \
--password cloudera \
--table orders \
--warehouse-dir /data \
--incremental lastmodified \
--check-column order_date \
--last-value 0 \
--append

sqoop job --exec job_orders_lastmodified

Create Password file
============================================================

echo -n "cloudera" >> .password-file


sqoop job \
--create job_orders \
-- import \
--connect jdbc:mysql://quickstart.cloudera:3306/retail_db \
--username root \
--password-file file:///home/cloudera/.password-file \
--table orders \
--warehouse-dir /data \
--incremental append \
--check-column order_id \
```

```
--last-value 0


Create alias
================================================================

hadoop credential create mysql.banking.password -provider
jceks://hdfs/user/cloudera/mysql.password.jceks

--To see password
hadoop fs -cat /user/cloudera/mysql.password.jceks

sqoop eval \
-Dhadoop.security.credential.provider.path=jceks://hdfs/user/cloudera/mysql.passwor
d.jceks \
--connect jdbc:mysql://quickstart.cloudera:3306/retail_db \
--username root \
--password-alias mysql.banking.password \
--query "select count(*) from orders"
```