# Impala hands-on

1. Create a Hive table called “hive\_table” in database “hive\_vs\_impala”

CREATE DATABASE hive\_vs\_impala;

CREATE TABLE hive\_table(Id INT, Price FLOAT, Amount DOUBLE, Date\_received STRING, Description STRING) STORED AS PARQUETE;

CREATE TABLE data\_table(Id INT, Price FLOAT, Amount DOUBLE, Date\_received STRING, Description STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY ‘,’ STORED AS TEXTFILE;

1. “hive\_table” should have the following columns, storage format - parquet:

* Id – Integer
* Price – Float
* Amount – Double
* Date\_received – String
* Description – String

1. Load some data in this table (any data, which satisfies data types used in table, at least 10 rows)

LOAD DATA INPATH '/user/cloudera/data\_impala.csv' OVERWRITE INTO TABLE data\_table;

insert overwrite table hive\_table select id, price, amount, date\_received, description FROM data\_table;

1. Open the same database in Impala, query data using SELECT. How many rows you get as output?

hive\_vs\_impala DB doesn’t show.

1. Use “Invalidate Metadata” option from Hue and query data again. How many rows you get as output?

hive\_vs\_impala DB is available now.

QUERY shows 12 rows as was imported from example data.

1. Create another Hive table “hive\_table\_2” with the following columns, storage format – text:

* Id – Integer
* F\_name – String
* L\_name – String
* Date\_of\_birth– Date
* Comments – String

CREATE TABLE hive\_table2(Id INT, F\_name STRING, L\_name STRING, Date\_of\_birth DATE, Comments STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY ‘,’ STORED AS TEXTFILE;

1. Load some data in this table (any data, which satisfies data types used in table, at least 10 rows)

LOAD DATA INPATH '/user/cloudera/data\_impala2.csv' OVERWRITE INTO TABLE hive\_table2;

1. Open the same database in Impala, query data using SELECT. How many rows you get as output? Why?

0 rows are showed. There is an error – Unsupported type 'DATE'. Because Impala doesn’t have a data type DATE. It only has TIMESTAMP.

1. Make needed changes in Hive table “hive\_table\_2” by creating a new table “hive\_table\_3” so you could query data using Impala

CREATE TABLE hive\_table3(Id INT, F\_name STRING, L\_name STRING, Date\_of\_birth TIMESTAMP, Comments STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY ‘,’ STORED AS TEXTFILE;

LOAD DATA INPATH '/user/cloudera/data\_impala2.csv' OVERWRITE INTO TABLE hive\_table2;