# Impala hands-on

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

Create database hive\_vs\_impala;

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

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

**CREATE EXTERNAL TABLE hive\_table\_e**

**(id string, price string, amount string, date\_received string, description string)**

**ROW FORMAT DELIMITED**

**FIELDS TERMINATED BY ','**

**LOCATION '/user/cloudera/imp'** (in this directory is located data sorce file from the next step)

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

**create table hive\_table**

**(id int, price float, amount double, date\_received string, description string)**

**stored as parquet**

**insert into hive\_table**

**select**

**cast(regexp\_replace(id, '"','')as int) as id,**

**cast(regexp\_replace(price, '"', '') as float) as price,**

**cast(regexp\_replace(amount, '"', '') as double) as amount,**

**date\_received as date\_received,**

**description as description**

**from hive\_table\_e**

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

**Select \* from hive\_vs\_impala.hive\_table;**

**Answer: none, Impala doesn’t “see” Hive table yet.**

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

**Invalidate metadata;**

**Select \* from hive\_vs\_impala.hive\_table;**

**Answer: all the rows from the table hive\_table;**

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 EXTERNAL TABLE hive\_table\_2\_e**

**(id string, f\_name string, l\_name string, date\_of\_birth string, comments string)**

**ROW FORMAT DELIMITED**

**FIELDS TERMINATED BY ','**

**LOCATION '/user/cloudera/imp\_';** (in this directory is located data source file from the next step)

**create table hive\_table**

**(id int, price float, amount double, date\_received string, description string)**

**stored as textfile;**

**insert into hive\_table\_2**

**select**

**cast(regexp\_replace(id, '"','')as int) as id,**

**f\_name as f\_name,**

**l\_name as l\_name,**

**cast(regexp\_replace(date\_of\_birth, '"','')as date) as date\_of\_birth,**

**comments as comments**

**from hive\_table\_2\_e;**

1. Load some data in this table (any data, which satisfies data types used in table, at least 10 rows)
2. Open the same database in Impala, query data using SELECT. How many rows you get as output? Why?

**select \* from hive\_vs\_impala.hive\_table\_2;**

**Answer: all the columns except of column “date\_of\_birth”. Impala doesn’t have type “date”.**

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

Hive\_table\_3 Impala output:

