



# IMPALA

## **open source massively parallel processing (MPP) SQL query engine**



**Cloudera Impala is a query engine that runs on Apache Hadoop. Impala brings scalable parallel database technology to Hadoop, enabling users to issue low-latency SQL queries to data stored in HDFS and Apache HBase without requiring data movement or transformation.**

# What is Impala ?

**General--- purpose SQL engine**

**Real--time queries in Apache Hadoop**

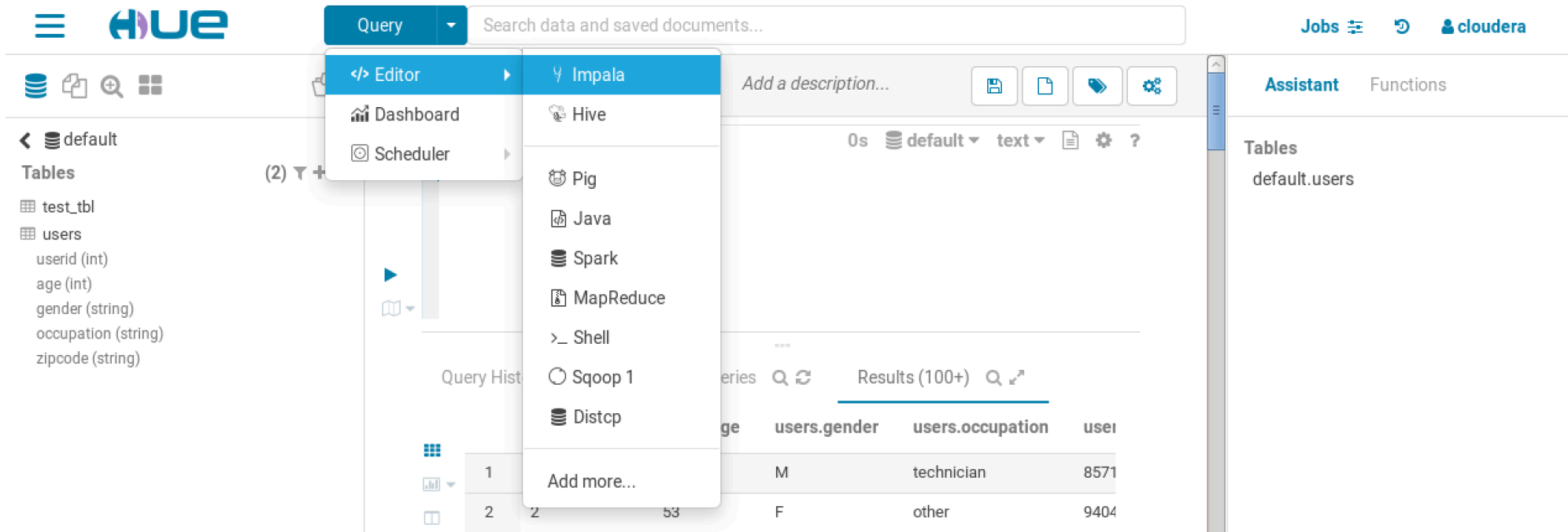
**Opensource under Apache License**

**Runs directly within Hadoop**

**High performance**

- C++ instead of Java**
- Runtime code generator**
- Roughly 4-100 x Hive**

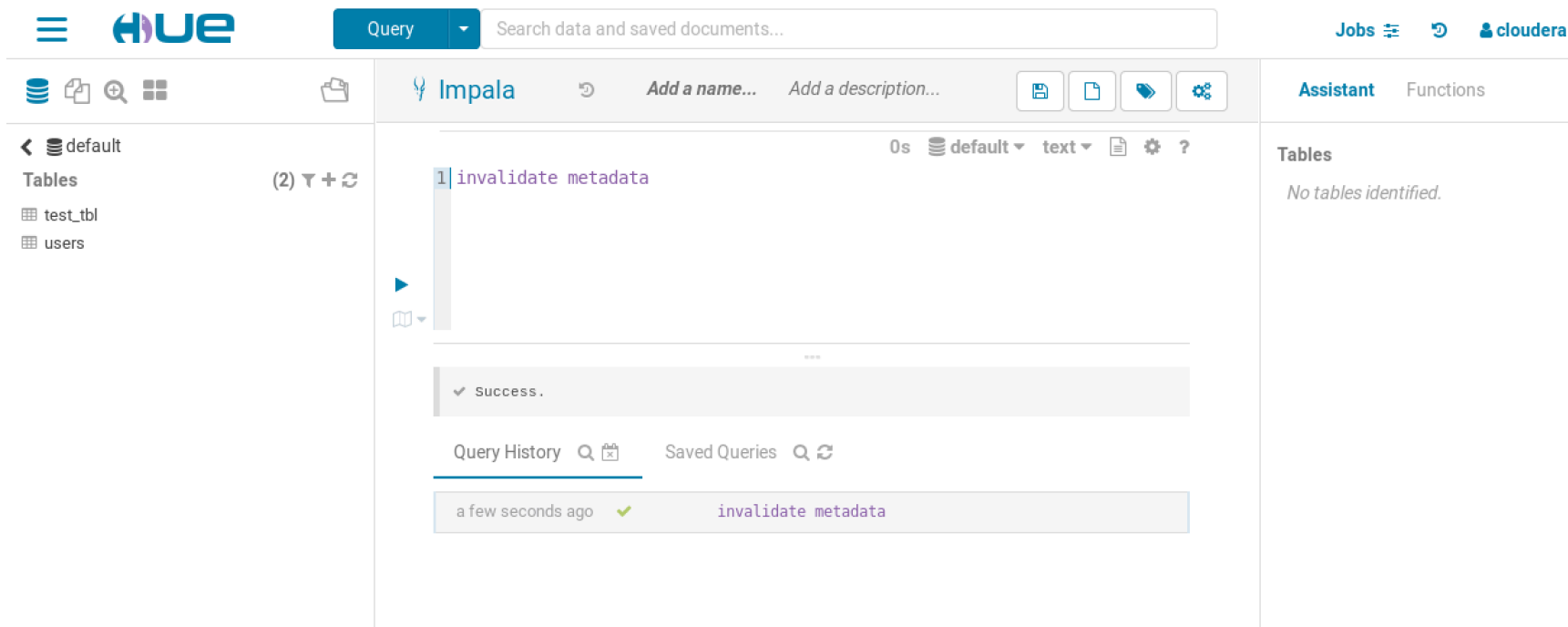
# Start Impala Query Editor



The screenshot shows the Hue web interface. The top navigation bar includes the Hue logo, a 'Query' dropdown menu, a search bar, and links for 'Jobs', 'cloudera', and 'Assistant'. The left sidebar shows a tree view of tables, including 'test\_tbl' and 'users' (with fields: userid (int), age (int), gender (string), occupation (string), zipcode (string)). The main content area displays the 'Query' dropdown menu, which is open, showing options: 'Editor', 'Dashboard', 'Scheduler', 'Impala', 'Hive', 'Pig', 'Java', 'Spark', 'MapReduce', 'Shell', 'Sqoop 1', 'Distcp', and 'Add more...'. The 'Impala' option is highlighted. Below the menu, a table of query results is visible, showing columns 'users.gender', 'users.occupation', and 'userid'. The table contains two rows of data.

users.gender	users.occupation	userid
M	technician	8571
F	other	9404

## Update the list of tables/metadata by execute the command **invalidate metadata**



The screenshot displays the Hue web interface for managing Hadoop data. The top navigation bar includes the Hue logo, a 'Query' dropdown menu, a search bar, and links for 'Jobs', 'cloudera', and 'Assistant'. The main workspace is divided into three panels. The left panel, titled 'Tables', shows a list of tables under the 'default' database: 'test\_tbl' and 'users'. The center panel, titled 'Impala', contains a text editor with the command 'invalidate metadata' entered. Below the editor, a 'Success.' message indicates the command was executed successfully. The bottom panel, titled 'Query History', shows a list of recent queries, including the 'invalidate metadata' command. The right panel, titled 'Tables', shows 'No tables identified.'.

**Query** Search data and saved documents...

**Jobs** **cloudera**

**Assistant** Functions

**Tables**  
No tables identified.

**Impala** Add a name... Add a description...

0s default text

1 invalidate metadata

Success.

Query History Saved Queries

a few seconds ago invalidate metadata

# Run Impala On Terminal

Open Terminal

Type command > impala-shell

[quickstart.cloudera:21000] > select \* from users;

```
[[quickstart.cloudera:21000] > select * from users;
```

```
Query: select * from users
```

userid	age	gender	occupation	zipcode
1	24	M	technician	85711
2	53	F	other	94043
3	23	M	writer	32067
4	24	M	technician	43537
5	33	F	other	15213
6	42	M	executive	98101
7	57	M	administrator	91344
8	36	M	administrator	05201
9	29	M	student	01002
10	53	M	lawyer	90703
11	39	F	other	30329
12	28	F	other	06405
13	47	M	educator	29206
14	45	M	scientist	55106
15	49	F	educator	97301
16	21	M	entertainment	10309