



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

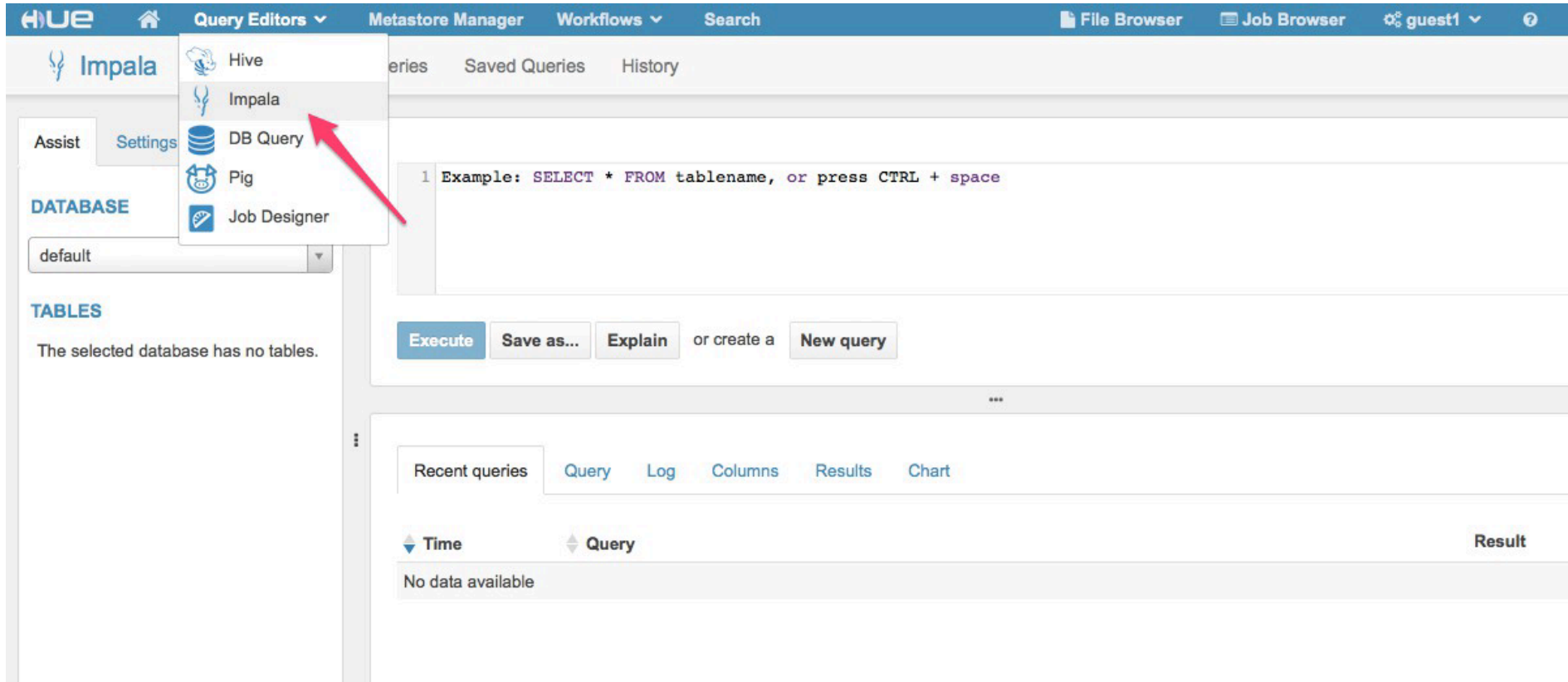
Opensource under Apache License

Runs directly within Hadoop

High performance

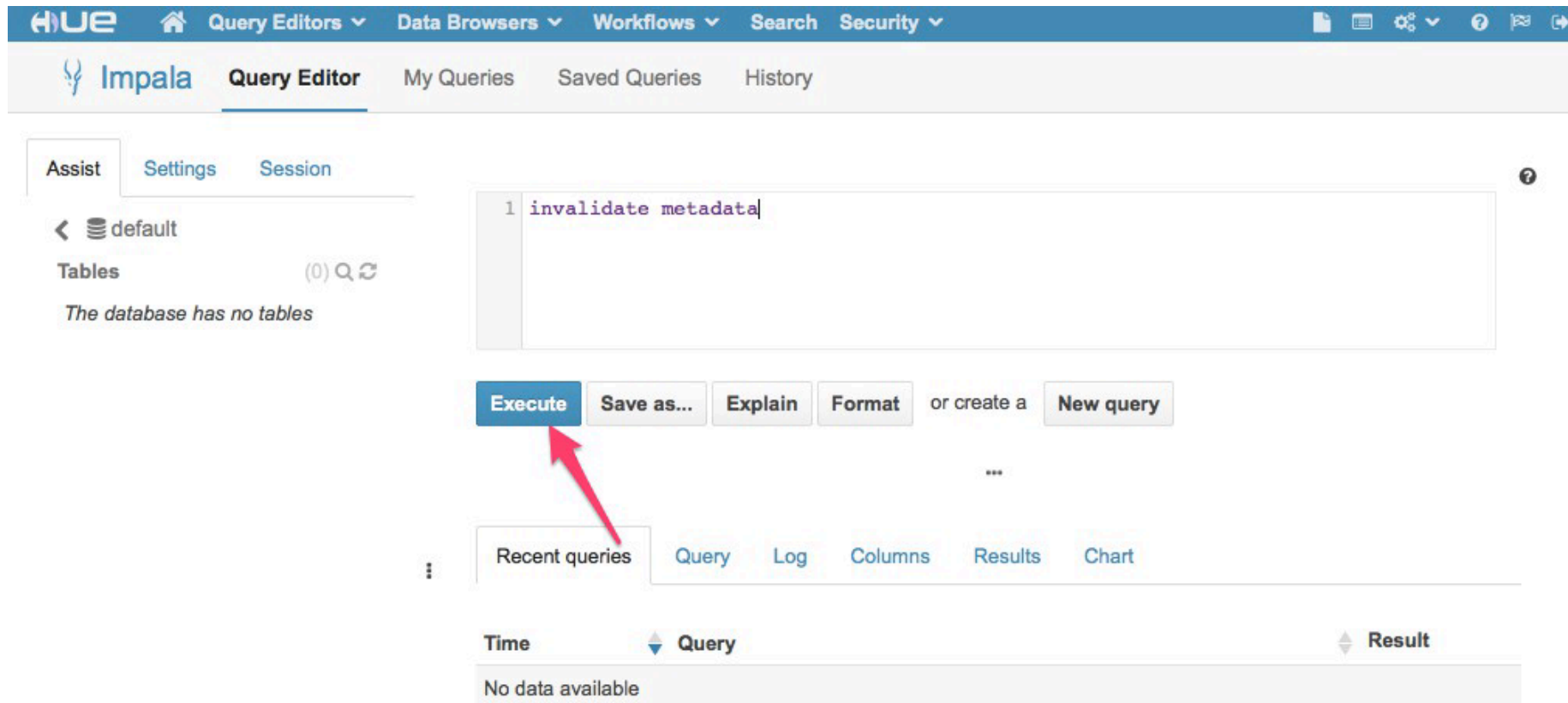
- C++ instead of Java**
- Runtime code generator**
- Roughly 4-100 x Hive**

Start Impala Query Editor



The screenshot displays the Hue web interface for the Impala Query Editor. The top navigation bar includes links for Home, Query Editors, Metastore Manager, Workflows, Search, File Browser, Job Browser, and a user profile dropdown. The 'Query Editors' dropdown menu is expanded, showing options for Hive, Impala, DB Query, Pig, and Job Designer. A red arrow points to the 'Impala' option. The main query editor area contains a sample query: `1 Example: SELECT * FROM tablename, or press CTRL + space`. Below the query editor are buttons for 'Execute', 'Save as...', 'Explain', and 'New query'. The bottom section shows a table with columns 'Time', 'Query', and 'Result', and a message 'No data available'.

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



The screenshot displays the Hue Impala Query Editor interface. The top navigation bar includes 'HUE', a home icon, and dropdown menus for 'Query Editors', 'Data Browsers', 'Workflows', 'Search', and 'Security'. Below this, the 'Impala' section is active, showing 'Query Editor', 'My Queries', 'Saved Queries', and 'History' tabs. On the left sidebar, the 'Assist' tab is selected, showing 'Settings' and 'Session' sub-tabs. Under 'Tables', it indicates '(0)' tables and states 'The database has no tables'. The main query editor area contains a single line of SQL: '1 invalidate metadata'. Below the query editor, a row of buttons includes 'Execute' (highlighted with a red arrow), 'Save as...', 'Explain', 'Format', 'or create a', and 'New query'. At the bottom, there are tabs for 'Recent queries', 'Query', 'Log', 'Columns', 'Results', and 'Chart'. The 'Results' tab is active, showing a table with columns 'Time', 'Query', and 'Result', and a message 'No data available'.

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