

Using the Hue Web Interface





Summer Elasady

Updating Account Info

- Click on your user name
- Edit Profile
- Follow the steps to update your password and info



The screenshot shows the HUE web interface. The top navigation bar includes links for Home, Query Editors, Data Browsers, Workflows, Search, Security, File Browser, Job Browser, and a user profile dropdown for 'selasady'. Below the navigation bar, the 'My documents' section is visible. On the left, there is a sidebar with 'ACTIONS' including 'New document', 'history' (500), and 'trash' (20). The main area displays a table of documents with columns for Name, Description, Last Modified, Project, and Sharing. A search bar is located above the table. The 'Edit Profile' button is highlighted in the top right corner.

Name	Description	Last Modified	Project	Sharing
 Ensembl_annotate	Annotate variants that fall within ensembl gene regions.	04/03/15 10:50:34	default	
 Ensembl_annotate	Annotate variants that fall within ensembl gene regions.	04/06/15 10:04:24	default	

Query Editor- Impala

- Click on Query Editor
- Click on impala

The screenshot displays the Hue Query Editor interface for Impala. A red arrow points to the 'Query Editor' tab in the top navigation bar. The interface includes a left sidebar with 'Assist' and 'Settings' tabs, a 'DATABASE' dropdown set to 'training', a 'Table name...' input field, and a 'global_vars' section. The main area shows a query editor with an example query: 'SELECT * FROM tablename, or press CTRL + space'. Below the editor are buttons for 'Execute', 'Save as...', 'Explain', 'or create a', and 'New query'. At the bottom, there is a 'Recent queries' section with tabs for 'Query', 'Log', 'Columns', 'Results', and 'Chart'. A table lists recent queries with columns for 'Time', 'Query', and 'Result'.

Time	Query	Result
10/27/15 14:55:24	select * from global_variants where chrom='1' order by pos	
10/27/15 14:55:19	FYI select * from global_variants where chrom='1' order by pos	
10/26/15 20:37:06	SELECT * from global_vars where chrom = '2' and pos = 48027775	

Query Editor- Impala

- Query window where you write and execute queries

The screenshot shows the Hue Impala Query Editor interface. The top navigation bar includes the Hue logo and menu items: Query Editors, Data Browsers, Workflows, Search, Security, File Browser, Job Browser, and a user profile 'selasady'. Below this is a sub-navigation bar with 'Impala', 'Query Editor', 'My Queries', 'Saved Queries', and 'History'. The main area is divided into a left sidebar and a central query editor. The sidebar has 'Assist' and 'Settings' tabs, a 'DATABASE' dropdown set to 'training', a 'Table name...' input field, and a 'global_vars' link. The central query editor is highlighted with a red rounded rectangle and contains a text area with the example query 'SELECT * FROM tablename, or press CTRL + space'. Below the text area are buttons for 'Execute', 'Save as...', 'Explain', 'or create a', and 'New query'. At the bottom, there is a 'Recent queries' section with tabs for 'Query', 'Log', 'Columns', 'Results', and 'Chart'. It displays a table with columns 'Time', 'Query', and 'Result'.

Time	Query	Result
10/27/15 14:55:24	select * from global_variants where chrom='1' order by pos	
10/27/15 14:55:19	FYI select * from global_variants where chrom='1' order by pos	
10/26/15 20:37:06	SELECT * from global_vars where chrom = '2' and pos = 48027775	

Query Editor- Impala

- View recent queries and query results

The screenshot displays the Hue Impala Query Editor interface. The top navigation bar includes links for Home, Query Editors, Data Browsers, Workflows, Search, Security, File Browser, Job Browser, and a user profile for 'selasady'. Below this, the 'Impala' section is active, with sub-links for Query Editor, My Queries, Saved Queries, and History.

On the left sidebar, the 'Assist' tab is selected, showing a 'DATABASE' dropdown set to 'training' and a 'Table name...' input field. Below these are 'global_vars' and a menu icon.

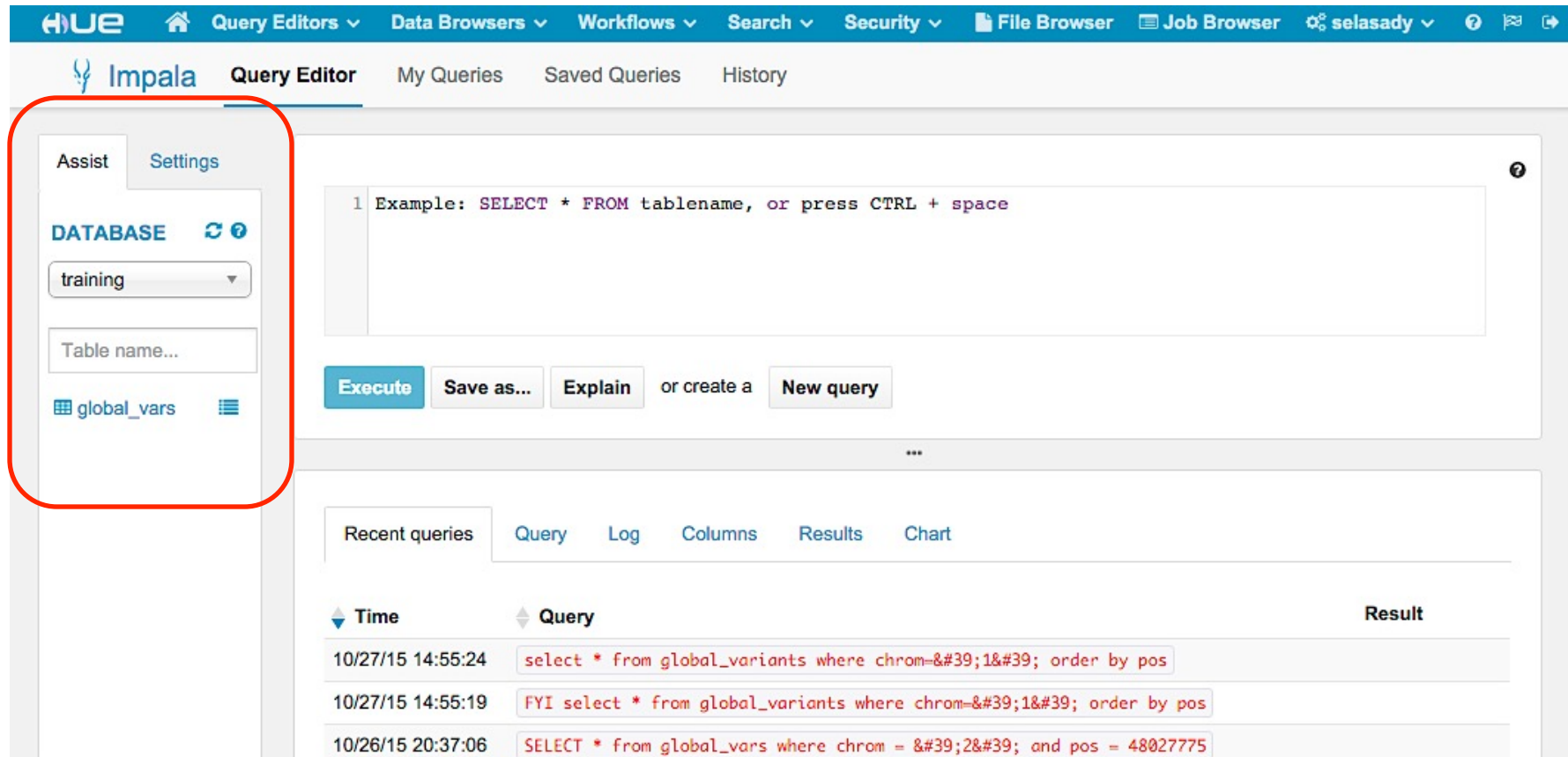
The main area contains a query editor with a text input field showing an example query: `1 Example: SELECT * FROM tablename, or press CTRL + space`. Below the input field are buttons for 'Execute', 'Save as...', 'Explain', 'or create a', and 'New query'.

Below the query editor, a red rounded rectangle highlights the 'Recent queries' section. This section has tabs for 'Query', 'Log', 'Columns', 'Results', and 'Chart'. It contains a table with the following data:

Time	Query	Result
10/27/15 14:55:24	<code>select * from global_variants where chrom=&#39;1&#39; order by pos</code>	
10/27/15 14:55:19	<code>FYI select * from global_variants where chrom=&#39;1&#39; order by pos</code>	
10/26/15 20:37:06	<code>SELECT * from global_vars where chrom = &#39;2&#39; and pos = 48027775</code>	

Selecting and Viewing Database

- Click on the drop-down menu to view and select databases



The screenshot displays the HUE Impala Query Editor interface. The top navigation bar includes links for Home, Query Editors, Data Browsers, Workflows, Search, Security, File Browser, Job Browser, and a user profile 'selasady'. The main header shows 'Impala' and 'Query Editor' tabs, along with 'My Queries', 'Saved Queries', and 'History'.

A red box highlights the left sidebar, which contains the 'Assist' and 'Settings' tabs. Under the 'Settings' tab, the 'DATABASE' dropdown menu is set to 'training'. Below this is a 'Table name...' input field and a link to 'global_vars'.

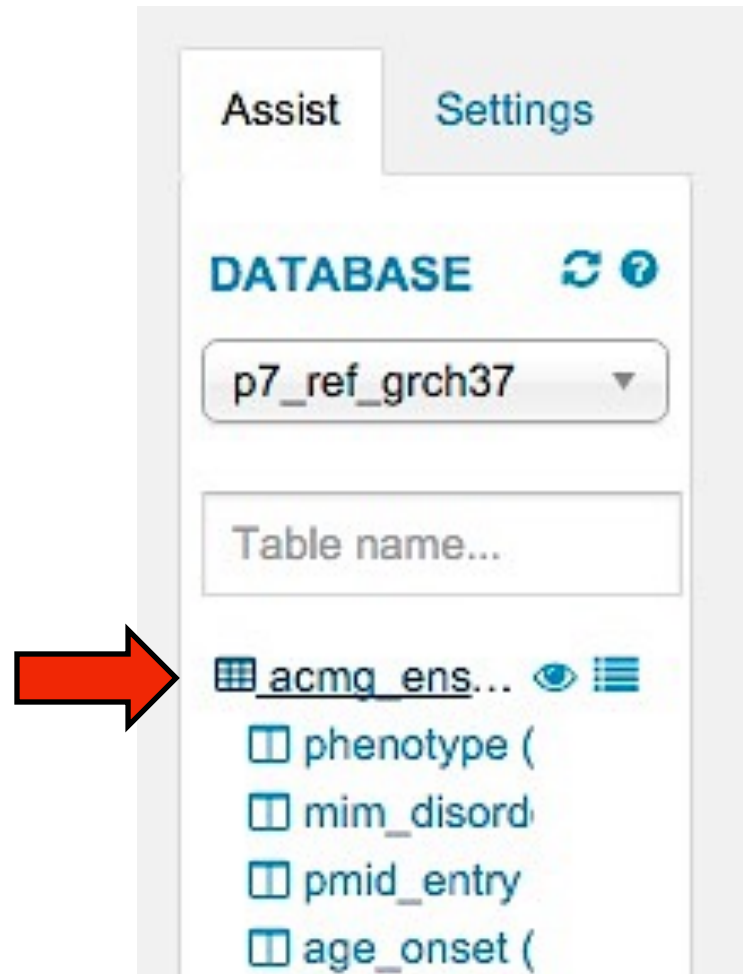
The main query editor area shows a sample query: `1 Example: SELECT * FROM tablename, or press CTRL + space`. Below the query editor are buttons for 'Execute', 'Save as...', 'Explain', and 'New query'.

At the bottom, there is a 'Recent queries' table with columns for Time, Query, and Result.

Time	Query	Result
10/27/15 14:55:24	<code>select * from global_variants where chrom=&#39;1&#39; order by pos</code>	
10/27/15 14:55:19	<code>FYI select * from global_variants where chrom=&#39;1&#39; order by pos</code>	
10/26/15 20:37:06	<code>SELECT * from global_vars where chrom = &#39;2&#39; and pos = 48027775</code>	

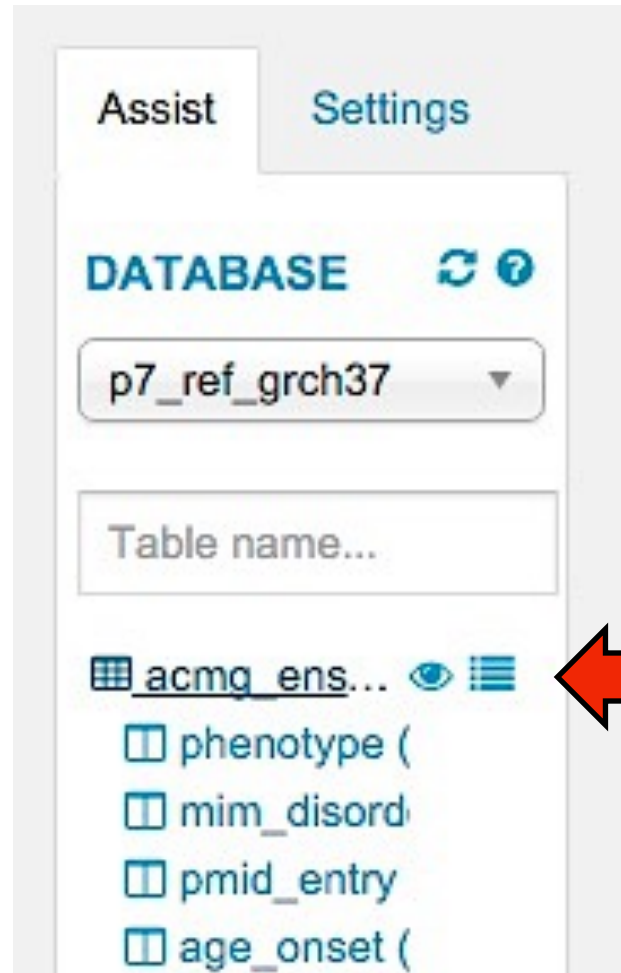
View Database Information

- Click on the database icon to view column names and data types



View Database Information

- Click on the eye to view information and a data sample in a new window
- Click on the list icon to view this information in the same window



Describe Database Columns

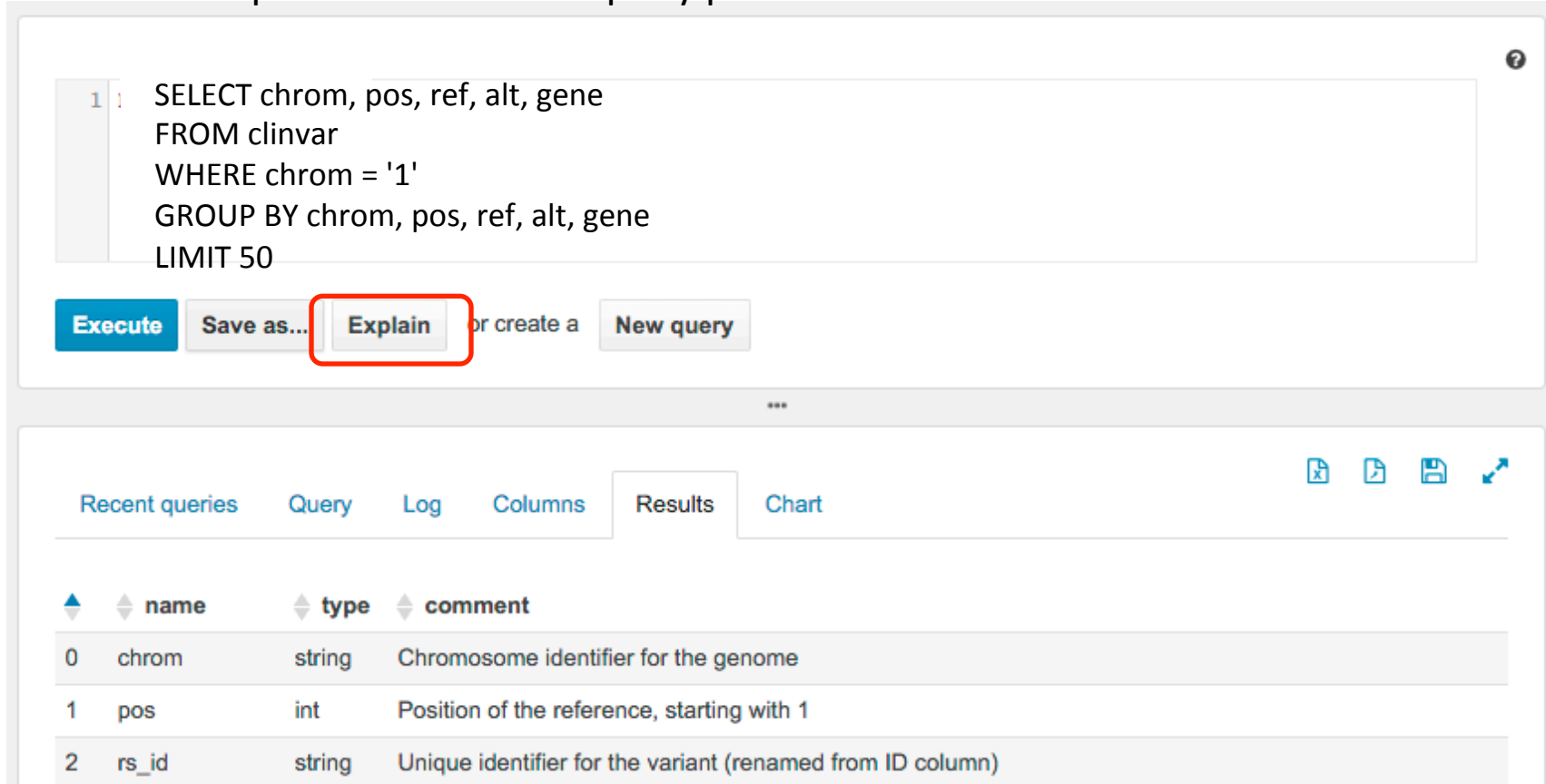
- Type DESCRIBE table_name and click on Execute
- View descriptions of what is in each column

The screenshot shows a database interface with a query editor at the top. The query `1 DESCRIBE clinvar` is entered and highlighted with a red box. Below the editor, the **Execute** button is also highlighted with a red box. Other buttons include **Save as...**, **Explain**, **or create a**, and **New query**. Below the buttons, there is a separator line with three dots. The main area shows a tabbed interface with **Recent queries**, **Query**, **Log**, **Columns**, **Results** (selected), and **Chart**. The **Results** tab displays a table with the following columns: **name**, **type**, and **comment**. The table contains three rows of data:

	name	type	comment
0	chrom	string	Chromosome identifier for the genome
1	pos	int	Position of the reference, starting with 1
2	rs_id	string	Unique identifier for the variant (renamed from ID column)

Create a Query

- Select the p7_ref_grch37 database
- Enter the query shown
- Click on Explain to review the query plan



The screenshot shows a query editor interface. At the top, a text area contains the following SQL query:

```
1 | SELECT chrom, pos, ref, alt, gene
   | FROM clinvar
   | WHERE chrom = '1'
   | GROUP BY chrom, pos, ref, alt, gene
   | LIMIT 50
```

Below the query text area, there are four buttons: "Execute" (blue), "Save as..." (light gray), "Explain" (light gray, highlighted with a red rectangle), and "New query" (light gray). To the right of the "Explain" button is the text "or create a".

Below the buttons, there is a horizontal bar with three dots "...".

Below the horizontal bar, there is a tabbed interface with the following tabs: "Recent queries", "Query", "Log", "Columns", "Results" (selected), and "Chart". To the right of the tabs are four icons: a document with an 'x', a document, a document with a checkmark, and a document with a magnifying glass.

Below the tabs, there is a table with the following columns: "name", "type", and "comment". The table contains three rows of data:

	name	type	comment
0	chrom	string	Chromosome identifier for the genome
1	pos	int	Position of the reference, starting with 1
2	rs_id	string	Unique identifier for the variant (renamed from ID column)

Review Explain Plan and Execute Query

- Read in numerical order
- Look for ways to optimize
- Identify bottlenecks
- Click Execute to run the query

Recent queries

Query

Explanation

Estimated Per-Host Requirements: Memory=90.00MB VCores=2

04:EXCHANGE [UNPARTITIONED]

| limit: 50

|

03:AGGREGATE [FINALIZE]

| group by: chrom, pos, ref, alt, gene

| limit: 50

|

02:EXCHANGE [HASH(chrom,pos,ref,alt,gene)]

|

01:AGGREGATE

| group by: chrom, pos, ref, alt, gene

|

00:SCAN HDFS [p7_ref_grch37.clinvar]

partitions=1/1 files=1 size=5.41MB

predicates: chrom = '1'

Save Queries

- After executing a query click on Save as
- Give the query a name and description

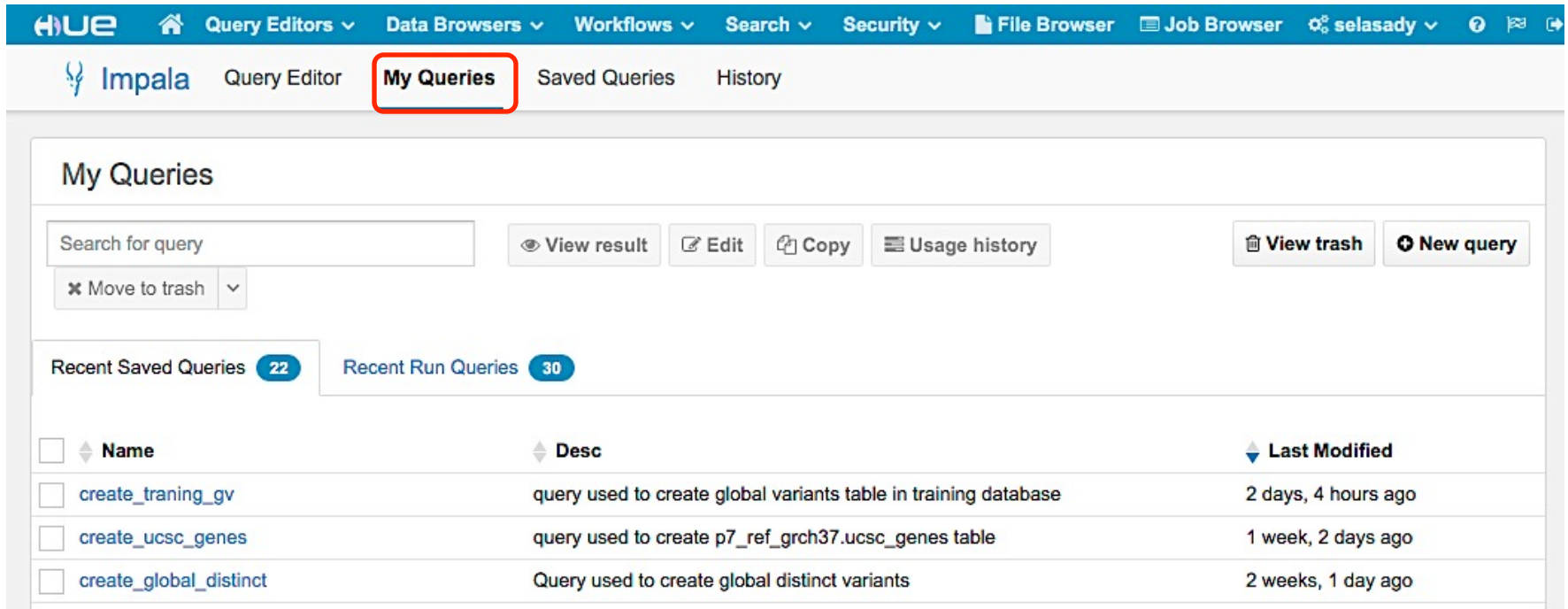
The screenshot displays the Hue Impala Query Editor interface. The top navigation bar includes links for Home, Query Editors, Data Browsers, Workflows, Search, Security, File Browser, Job Browser, and a user profile for 'selasady'. Below this, the 'Impala' section is active, showing 'Query Editor', 'My Queries', 'Saved Queries', and 'History' tabs. On the left sidebar, the 'Assist' tab is selected, showing a 'DATABASE' dropdown set to 'p7_ref_grch37' and a 'Table name...' input field. Below these are database tables: 'acmg_ens...', 'acmg_genes', and 'cadd'. The main query editor area contains a SQL query:

```
1 SELECT *  
2 FROM clinvar  
3 WHERE chrom = '1'  
4 LIMIT 5
```

 At the bottom of the query editor, there are buttons for 'Execute', 'Save as...' (which is highlighted with a red rectangle), 'Explain', and 'New query'. The text 'or create a' is positioned between 'Explain' and 'New query'.

Run Saved Queries

- Click on My Queries to view your saved queries
- Click on the query name to run it in the impala query editor



The screenshot shows the Hue web interface. The top navigation bar includes links for Query Editors, Data Browsers, Workflows, Search, Security, File Browser, Job Browser, and a user profile. Below this, the 'Impala' section has tabs for Query Editor, My Queries (highlighted with a red box), Saved Queries, and History.

The 'My Queries' section displays a search bar and several action buttons: View result, Edit, Copy, Usage history, View trash, and New query. Below these, there are two tabs: Recent Saved Queries (22) and Recent Run Queries (30).

A table lists the recent saved queries:

Name	Desc	Last Modified
<input type="checkbox"/> create_traning_gv	query used to create global variants table in training database	2 days, 4 hours ago
<input type="checkbox"/> create_ucsc_genes	query used to create p7_ref_grch37.ucsc_genes table	1 week, 2 days ago
<input type="checkbox"/> create_global_distinct	Query used to create global distinct variants	2 weeks, 1 day ago

Save Results

- Click on one of the icons shown in the red box to save your results as a table on impala, or to download as a csv or excel file

The screenshot shows the HUE Impala Query Editor interface. The top navigation bar includes links for Query Editors, Data Browsers, Workflows, Search, and Security. The main header shows the Impala logo and tabs for Query Editor, My Queries, Saved Queries, and History. On the left, there's a sidebar with 'Assist' and 'Settings' tabs, a database selector (p7_ref_grch37), and a list of tables. The main area contains a SQL query editor with the following query:

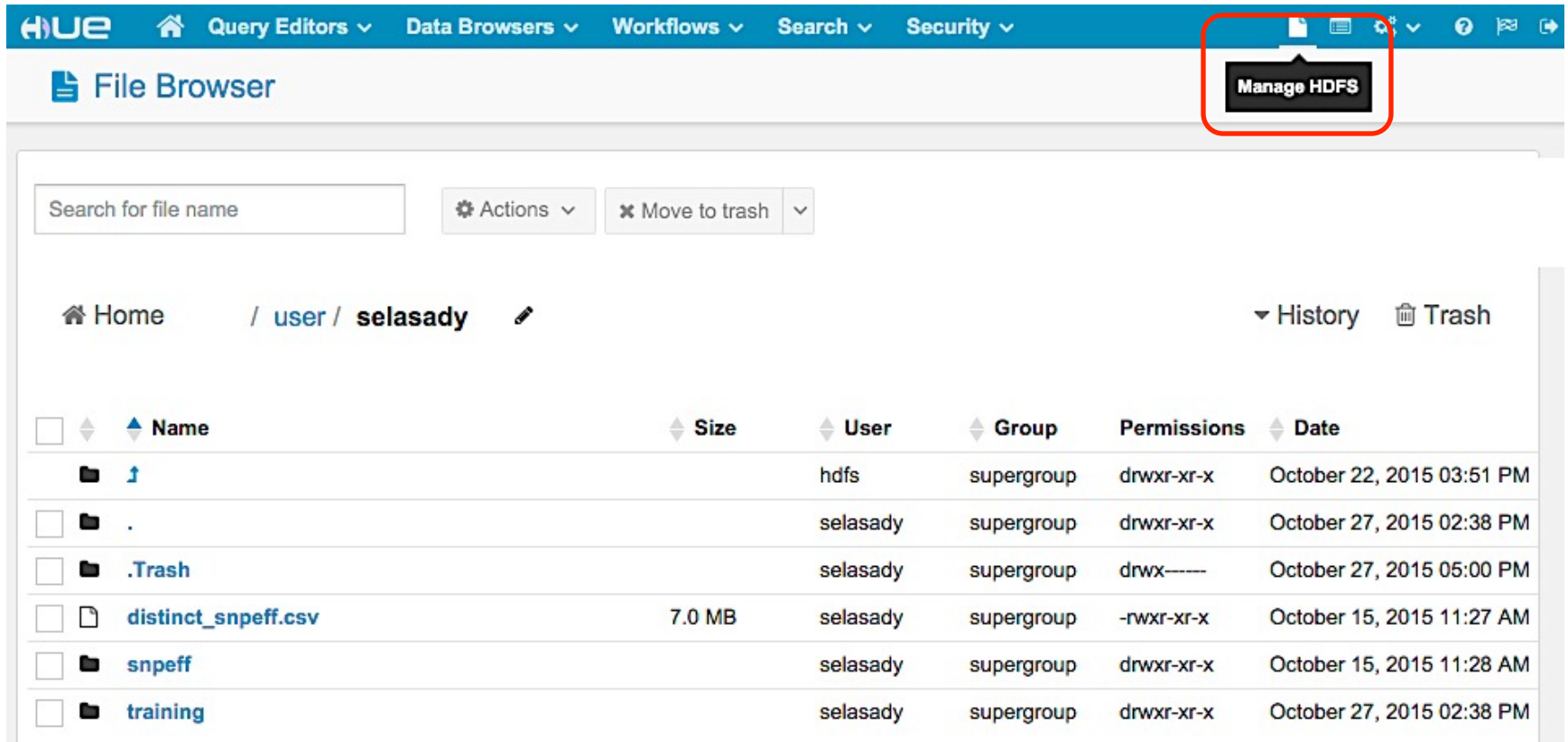
```
1 SELECT *
2 FROM clinvar
3 WHERE chrom = '1'
4 LIMIT 5
```

Below the query editor are buttons for 'Execute', 'Save', 'Save as...', 'Explain', and 'New query'. The 'Results' tab is active, displaying a table of query results. A red box highlights three icons (Excel, CSV, and Save to HDFS) in the top right corner of the results area. A tooltip points to these icons with the text: 'Save the results to HDFS or a new Hive table'.

	chrom	pos	rs_id	ref	alt	qual	filter	rs_pos	rv	vp	gene
0	1	883516	rs267598747	G	A	NULL	NULL	267598747	NULL	NULL	0x050060000305000002
1	1	891344	rs267598748	G	A	NULL	NULL	267598748	NULL	NULL	0x050060000305000002
2	1	906168	rs267598759	G	A	NULL	NULL	267598759	NULL	NULL	0x050060080a05000002
3	1	949696	rs672601345	C	CG	NULL	NULL	672601345	NULL	NULL	0x050060001205000002

Working with HDFS

- Click on the file icon
- Click on manage HDFS

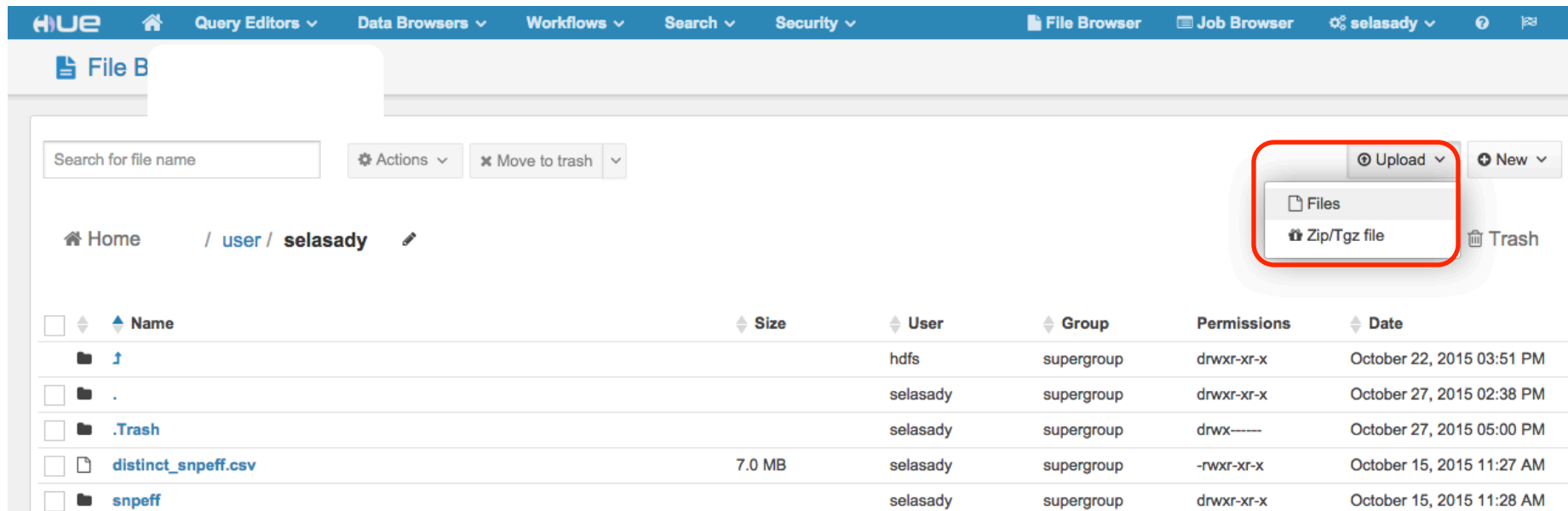


The screenshot shows the HUE File Browser interface. The top navigation bar includes links for Query Editors, Data Browsers, Workflows, Search, and Security. A red box highlights the 'Manage HDFS' button in the top right corner. Below the navigation bar, the 'File Browser' section is visible. It includes a search bar for file names, an 'Actions' dropdown menu, and a 'Move to trash' button. The main content area displays a file listing for the user 'selasady'. The listing includes columns for Name, Size, User, Group, Permissions, and Date. The files listed are 'hdfs', '.', '.Trash', 'distinct_snpeff.csv' (7.0 MB), 'snpeff', and 'training'.

Name	Size	User	Group	Permissions	Date
hdfs		hdfs	supergroup	drwxr-xr-x	October 22, 2015 03:51 PM
.		selasady	supergroup	drwxr-xr-x	October 27, 2015 02:38 PM
.Trash		selasady	supergroup	drwx----	October 27, 2015 05:00 PM
distinct_snpeff.csv	7.0 MB	selasady	supergroup	-rwxr-xr-x	October 15, 2015 11:27 AM
snpeff		selasady	supergroup	drwxr-xr-x	October 15, 2015 11:28 AM
training		selasady	supergroup	drwxr-xr-x	October 27, 2015 02:38 PM

Uploading a TSV/CSV to HDFS

- Click on the file icon
- Click on manage HDFS



The screenshot shows the HUE File Browser interface. At the top, there is a navigation bar with tabs for Query Editors, Data Browsers, Workflows, Search, Security, File Browser, and Job Browser. The 'File Browser' tab is active. Below the navigation bar, there is a search bar and a toolbar with buttons for Actions, Move to trash, and Upload. The 'Upload' button is highlighted with a red box, and its dropdown menu is open, showing options for 'Files' and 'Zip/Tgz file'. The main area displays a file list with columns for Name, Size, User, Group, Permissions, and Date. The file list includes a folder named 'distinct_snpeff.csv' (7.0 MB) and a folder named 'snpeff'.

Name	Size	User	Group	Permissions	Date
distinct_snpeff.csv	7.0 MB	selasady	supergroup	-rwxr-xr-x	October 15, 2015 11:27 AM
snpeff		selasady	supergroup	drwxr-xr-x	October 15, 2015 11:28 AM

Viewing uploaded files

- Once a file is uploaded to HDFS click on that file
- Make sure everything looks correct before converting to a table

The screenshot shows the HUE File Browser interface. The top navigation bar includes links for Home, Query Editors, Data Browsers, Workflows, Search, Security, File Browser, Job Browser, and a user profile for 'selasady'. The main content area displays the file 'query_result (3).csv' under the path 'user / selasady'. On the left sidebar, the 'ACTIONS' menu includes 'View as binary', 'Edit file', 'Download', 'View file location', and 'Refresh'. The 'INFO' section shows the file was last modified on Oct. 29, 2015 at 9:58 a.m., owned by 'selasady', in the 'supergroup' group, with a size of 2.2 KB and mode 100644. The main pane shows the first 20 lines of the CSV file, which contains genomic data with columns for chromosome, position, reference, alternate, and gene.

ACTIONS

- View as binary
- Edit file
- Download
- View file location
- Refresh

INFO

Last modified
Oct. 29, 2015 9:58 a.m.

User
selasady

Group
supergroup

Size
2.2 KB

Mode
100644

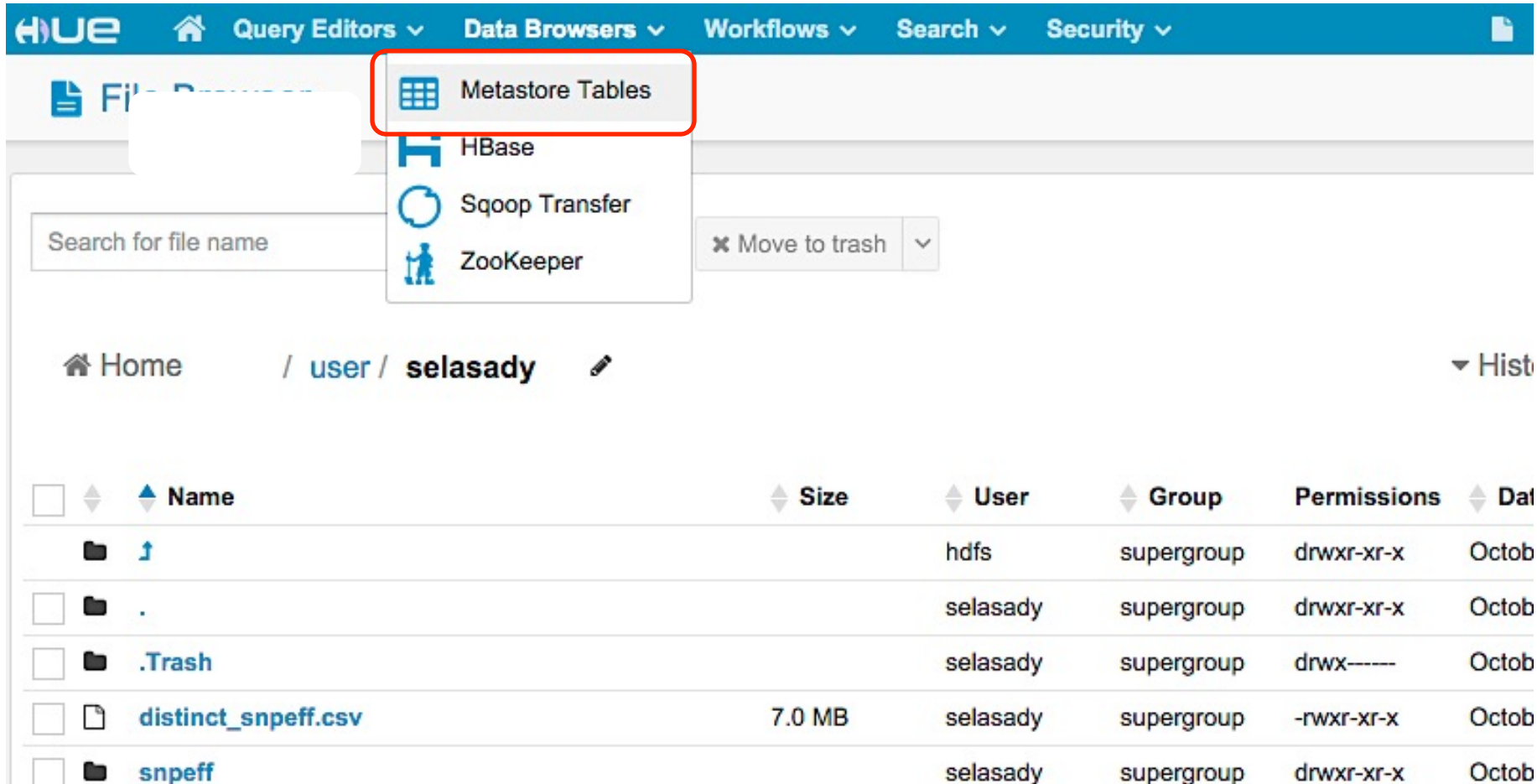
Home / user / selasady / query_result (3).csv

Page 1 of 1

```
chrom,pos,ref,alt,gene
1,248004844,C,T,0x050260080a05000002100120
1,102002045,G,C,0x050060000005000002110120
1,94577102,C,A,0x050160000a05000026110100
1,156108383,C,T,0x050160000305000002100100
1,157067690,C,T,0x050068000a05000002100120
1,94480100,C,G,0x050160000a05000002100100
1,201334762,G,C,0x050060000305000026100100
1,159272169,G,A,0x050068000a05000002100120
1,94476399,GGAA,G,0x050160000005000002100200
1,156106208,T,C,0x050360000a05000002100100
1,40737681,C,G,0x050168000a05000002110100
1,63347836,A,G,0x050060000005000002110120
1,197073936,C,T,0x050060080a05000002110100
1,161599693,T,C,0x050160000a05170126100140
1,190286862,G,C,0x050060080005000002110120
1,169696946,G,A,0x050168000a0515053f110100
1,65307190,T,C,0x050260000a05040036100100
1,5935162,A,T,0x05012820000517053f000101
1,160100264,C,T,0x050360000305150537100100
1,46656458,T,C,0x050068200005000002110100
1,223164921,G,A,0x050060000305000002100120
-----
```

Converting an HDFS file to a table

- Once a file has been uploaded to HDFS, click on Data Browsers
- Click on Metastore Tables



The screenshot shows the Hue web interface. The top navigation bar includes 'HUE', a home icon, and dropdown menus for 'Query Editors', 'Data Browsers', 'Workflows', 'Search', and 'Security'. The 'Data Browsers' dropdown is open, showing options: 'Metastore Tables' (highlighted with a red box), 'HBase', 'Sqoop Transfer', and 'ZooKeeper'. Below the dropdown is a search bar labeled 'Search for file name' and a 'Move to trash' button. The main content area shows a breadcrumb path 'Home / user / selasady' and a table of files in the HDFS directory.

<input type="checkbox"/>	Name	Size	User	Group	Permissions	Date
<input type="checkbox"/>	↑		hdfs	supergroup	drwxr-xr-x	Octob
<input type="checkbox"/>	.		selasady	supergroup	drwxr-xr-x	Octob
<input type="checkbox"/>	.Trash		selasady	supergroup	drwx-----	Octob
<input type="checkbox"/>	distinct_snpeff.csv	7.0 MB	selasady	supergroup	-rwxr-xr-x	Octob
<input type="checkbox"/>	snpeff		selasady	supergroup	drwxr-xr-x	Octob

Select a database to place the table

- Click on the file icon
- Click on manage HDFS

The screenshot shows the Hue Metastore Manager interface. The top navigation bar includes links for Query Editors, Data Browsers, Workflows, Search, Security, File Browser, Job Browser, and a user profile for 'selasady'. The main header is 'Metastore Manager'. On the left, a 'DATABASE' dropdown menu is open, showing a list of databases. The database 'users_selasady' is highlighted at the bottom of the list. The main content area is titled 'Databases > users_selasady'. It features a search bar for table names and three buttons: 'View', 'Browse Data', and 'Drop'. Below these, a table lists various tables with checkboxes for selection. The tables listed are: acmg_annotated, acmg_kaviar, acmg_ucsc, admix_test, clarity_acmg, clarity_clinkav, clarity_clinsig, and clarity_clinsig_results.

DATABASE

users_selasady

oerauit

__ibis_tmp

p7_archive

p7_itmi

p7_platform

p7_ref_grch37

p7_staging

training

users_hrishi

users_selasady

Databases > users_selasady

Search for table name

<input type="checkbox"/> Table Name
<input type="checkbox"/> acmg_annotated
<input type="checkbox"/> acmg_kaviar
<input type="checkbox"/> acmg_ucsc
<input type="checkbox"/> admix_test
<input type="checkbox"/> clarity_acmg
<input type="checkbox"/> clarity_clinkav
<input type="checkbox"/> clarity_clinsig
<input type="checkbox"/> clarity_clinsig_results

Create a new table from a file

HUE | Home | Query Editors | Data Browsers | Workflows | Search | Security | File Browser | Job Browser | selasady | ? | [Icons]

Metastore Manager

DATABASE

users_selasady

ACTIONS

- Create a new table from a file
- Create a new table manually

Databases > users_selasady > Create a new table from a file

Step 1: Choose File | Step 2: Choose Delimiter | Step 3: Define Columns

Name Your Table and Choose A File

Table Name
Name of the new table. Table names must be globally unique. Table names tend to correspond to the directory where the data will be stored.

Description
Use a table comment to describe the table. For example, note the data's provenance and any caveats users need to know.

Input File ..
The HDFS path to the file on which to base this new table definition. It can be compressed (gzip) or not.

Import data from file ☒
Check this box to import the data in this file after creating the table definition. Leave it unchecked to define an empty table.

Warning: The selected file is going to be moved during the import.

Next

Choose your delimiter

- Choose how your file is delimited

Databases > users_selasady > Create a new table from a file

Step 1: Choose File **Step 2: Choose Delimiter** Step 3: Define Columns

Choose a Delimiter

Beeswax has determined that this file is delimited by **commas**.

Delimiter Comma (,) **Preview**

Enter the column delimiter which must be a single character. Use syntax like "\"001\" or \"t\" for special characters.

Table preview	col_1	col_2	col_3	col_4	col_5
	chrom	pos	ref	alt	gene
	1	248004844	C	T	0x050260080a0500000210012...
	1	102002045	G	C	0x05006000000500000211012...
	1	94577102	C	A	0x050160000a0500002611010...
	1	156108383	C	T	0x05016000030500000210010...
	1	157067690	C	T	0x050068000a0500000210012...
	1	94480100	C	G	0x050160000a0500000210010...
	1	201334762	G	C	0x05006000030500002610010...
	1	159272169	G	A	0x050068000a0500000210012...
	1	94476399	GGAA	G	0x05016000000500000210020...

Previous **Next**

Define your columns

- Check that the proper data type is selected for each column

Databases > users_selasady > Create a new table from a file

Step 1: Choose File Step 2: Choose Delimiter Step 3: Define Columns

Define your columns

Use first row as column names

Bulk edit column names



Column name	Column Type	Sample Row #1	Sample Row #2
chrom	bigint	1	1
pos	int	248004844	102002045
ref	string	C	G
alt	string	T	C
gene	bigint	0x050260080a05000002100120	0x0500600000005000002110

Previous

Create Table