

OPEN UP INTERACTIVE BIG DATA ANALYSIS FOR YOUR ENTERPRISE

Enrico Berti

Budapest DW Forum, Jun 4, 2014

cloudera

The logo for HUE, featuring a stylized blue 'H' with a purple and white circular element on its right side, followed by the letters 'U' and 'E' in a blue, sans-serif font.

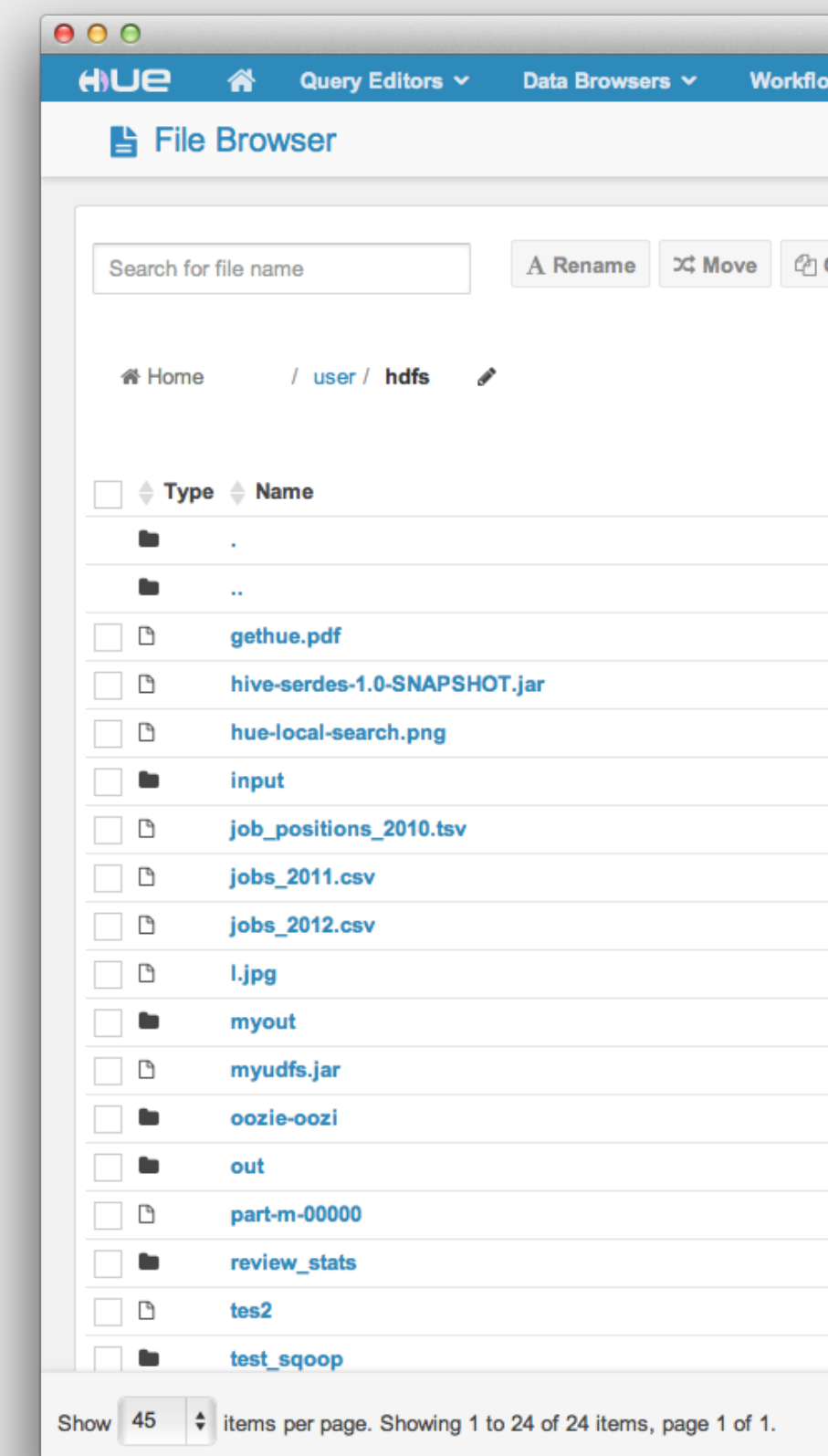
GOAL OF HUE

WEB INTERFACE FOR ANALYZING DATA
WITH APACHE HADOOP

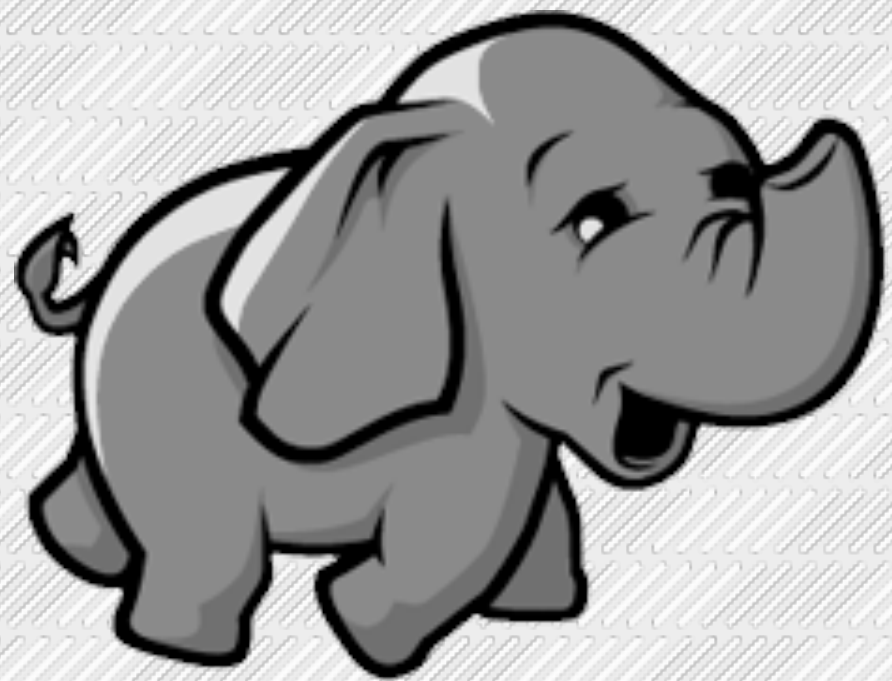
SIMPLIFY AND INTEGRATE

FREE AND OPEN SOURCE

—> OPEN UP BIG DATA



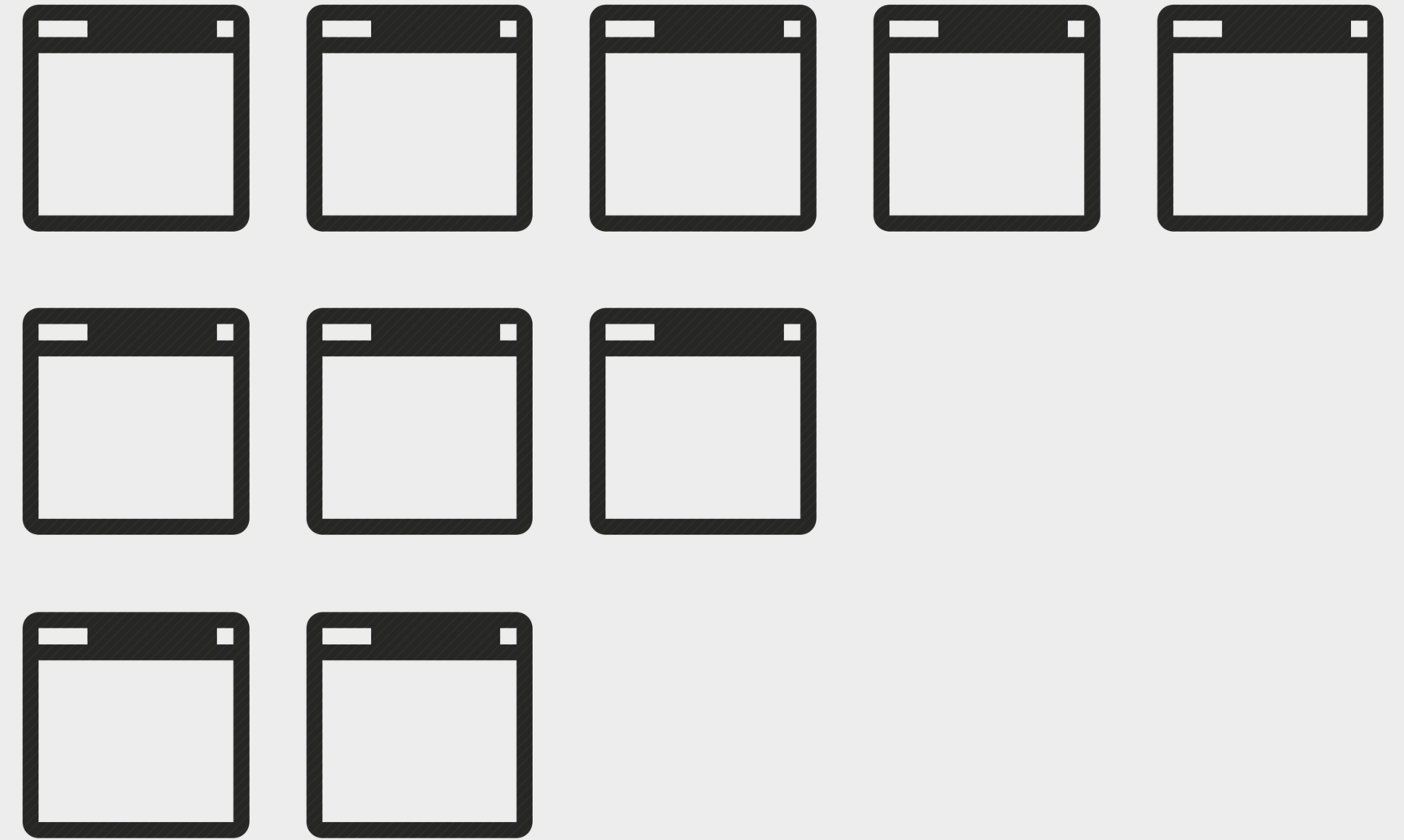
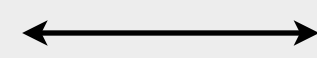
VIEW FROM 30K FEET



Hadoop



Web Server



You, your colleagues and even that friend that uses IE9 ;)

OPEN SOURCE

~3350 **COMMITTS**

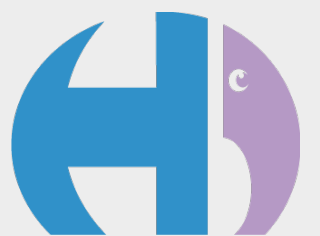
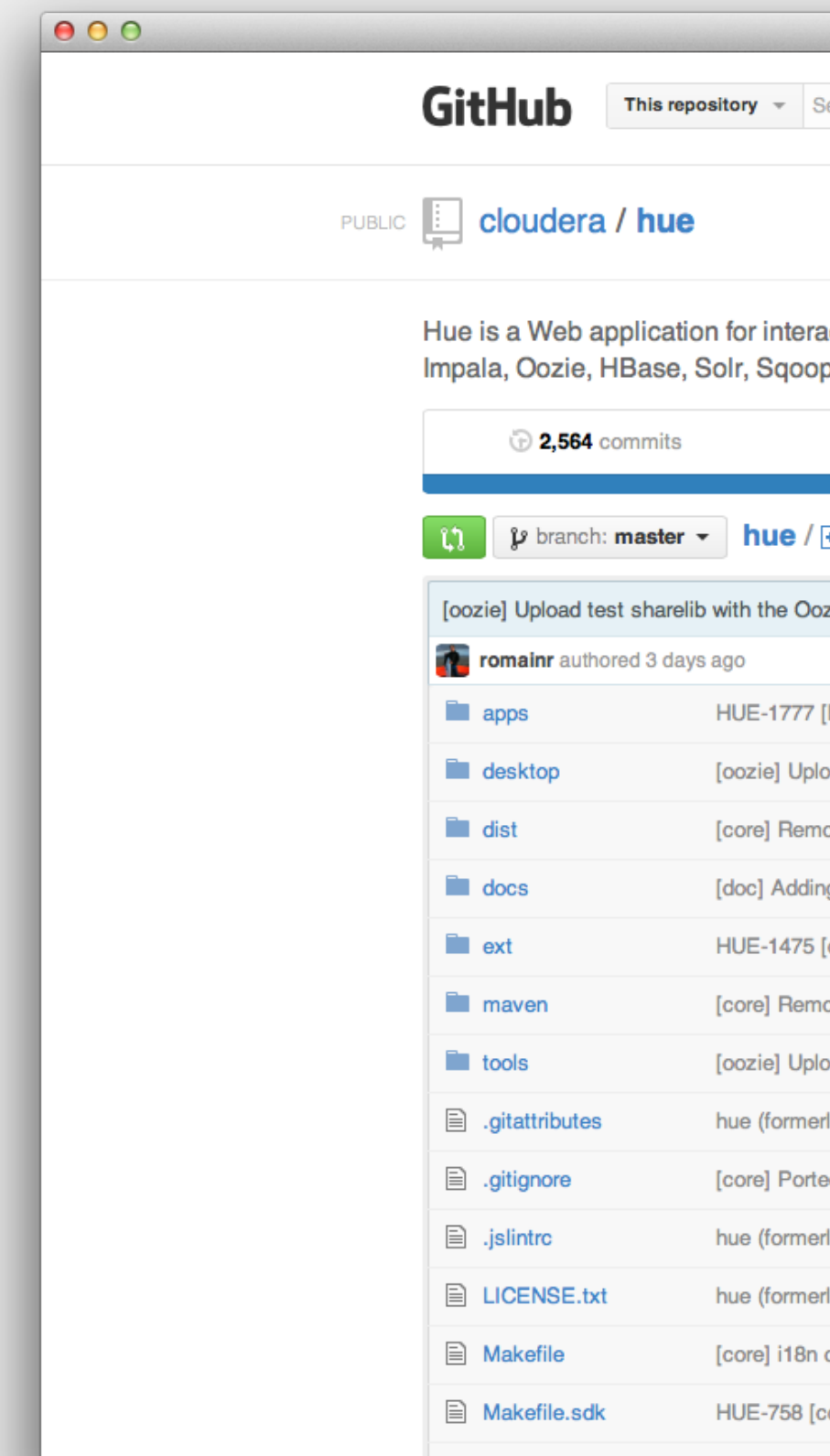
38 **CONTRIBUTORS**

698 **STARS**

245 **FORKS**

github.com/cloudera/hue

cloudera



THE CORE TEAM PLAYERS



Romain Rigaux



Chang



Enrico Berti



Amstel



Abraham Elmahrek

AROUND THE WORLD

TALKS

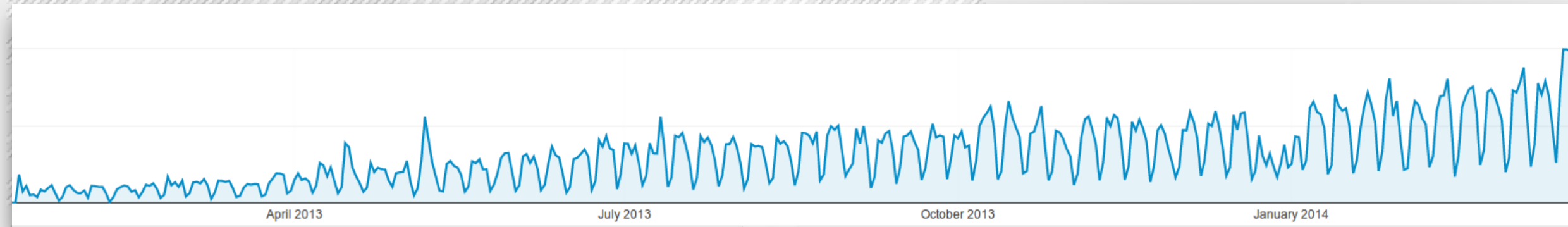
Meetups and events in NYC, Paris, LA, Tokyo, SF, Stockholm, Vienna, San Jose, Singapore, Budapest...
Coming up in London, West coast
...

RETREATS

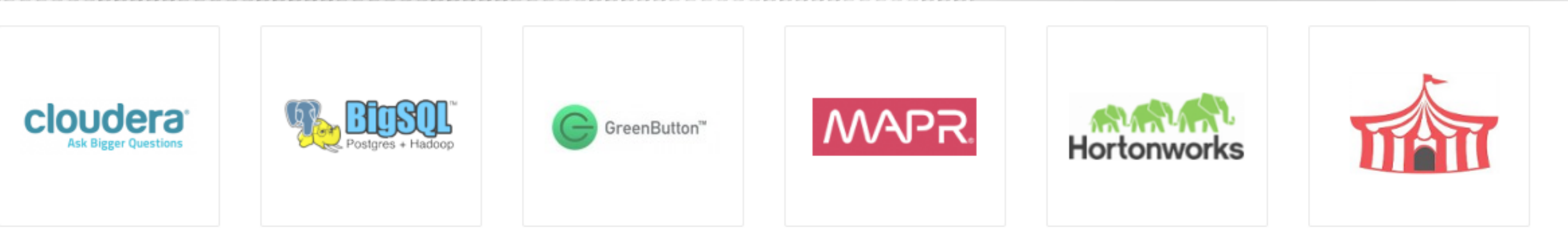
Nov 13 Koh Chang, Thailand
May 14 Curaçao, Netherlands Antilles
Nov 14 Goa, India



TREND: GROWTH









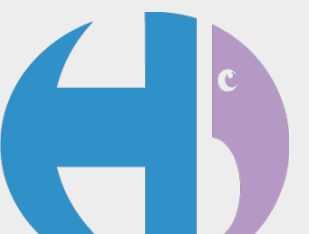
gethue.com



cloudera

Mar 13, 2014

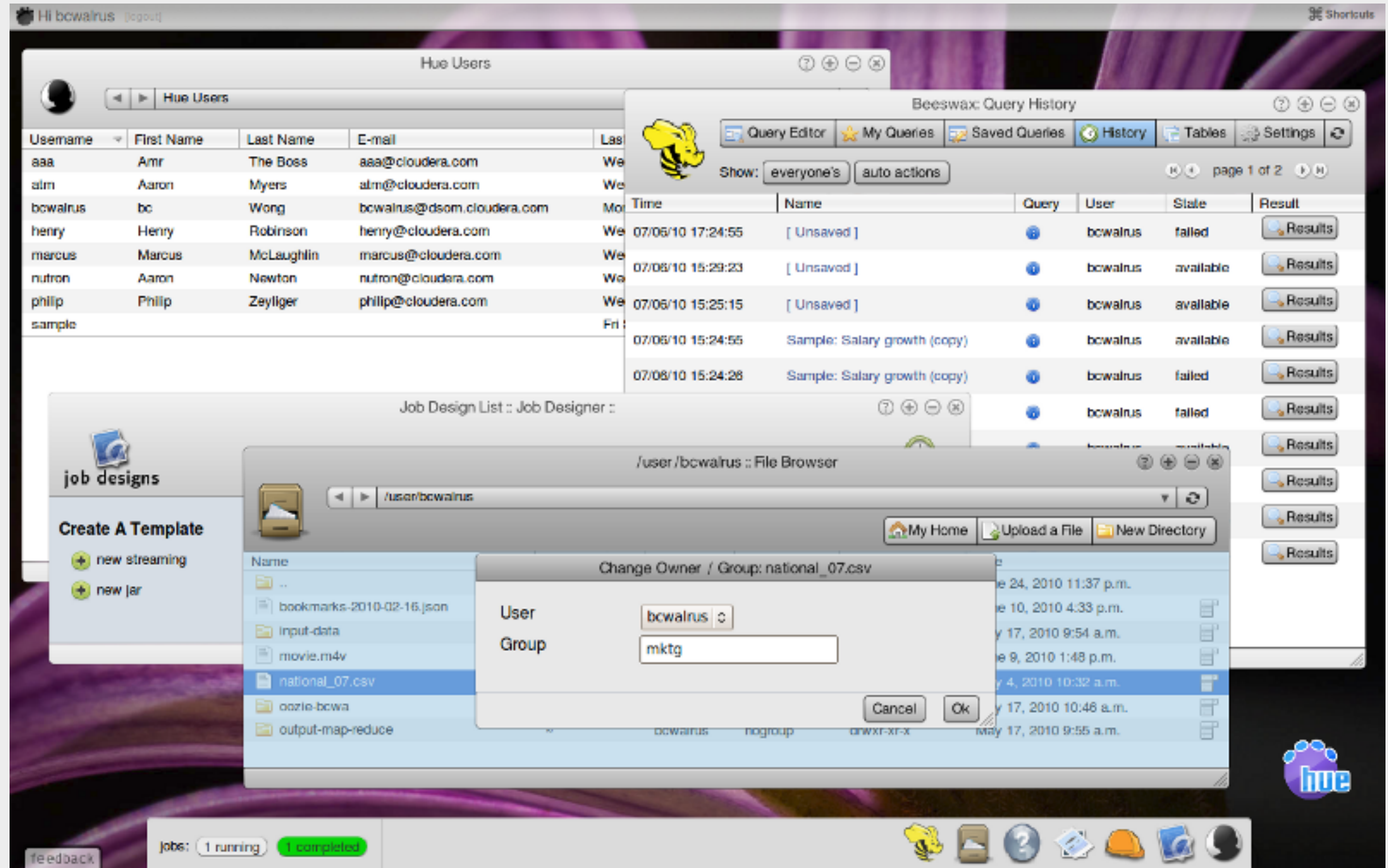
-  **[impala] Replace non utf8 encoded chars from the resultsets** d6c869c498
romainr authored a month ago
[Browse code](#)
-  **[liboauth] Limit oauth login to specific list of domains** 76064c61ae
dominikgehl authored a month ago
→ romainr committed a month ago
[Browse code](#)
-  **[liboauth] Show error message after Aouth login failure** 7307bf69be
dominikgehl authored a month ago
→ romainr committed a month ago
[Browse code](#)
-  **[liboauth] Indentation fix** 5cee8ea5eb
dominikgehl authored a month ago
→ romainr committed a month ago
[Browse code](#)
-  **HUE-2020 [hbase] upload an image sometimes produces an error** 49bbfc8786
abec authored a month ago
[Browse code](#)
-  **[hbase] Fix new cell upload** cf39ae23b4
abec authored a month ago
[Browse code](#)



HISTORY

HUE 1

Desktop-like in a browser, did its job but pretty slow, memory leaks and not very IE friendly but definitely advanced for its time (2009-2010).



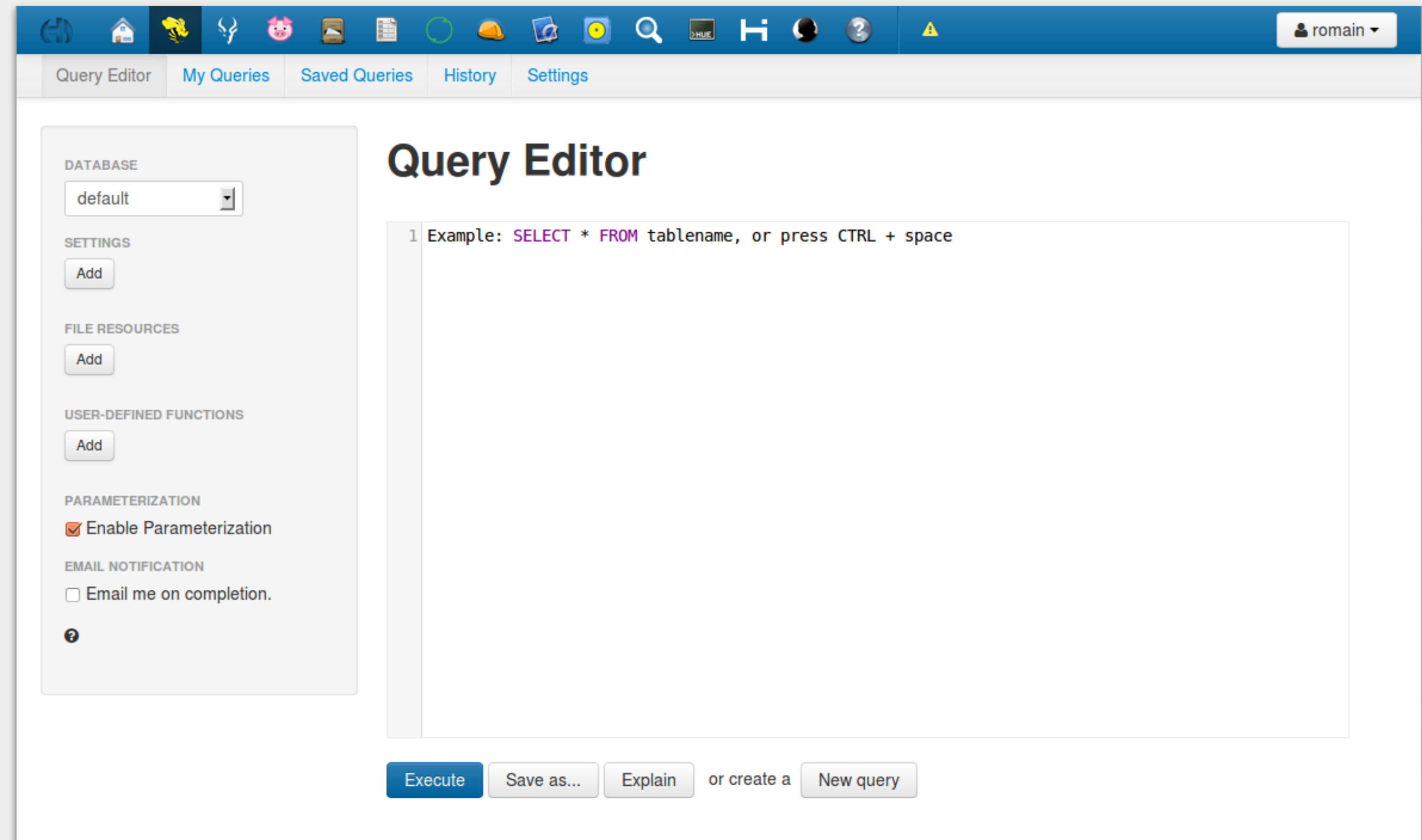
HISTORY

HUE 2

The first flat structure port, with Twitter Bootstrap all over the place.

HUE 2.5

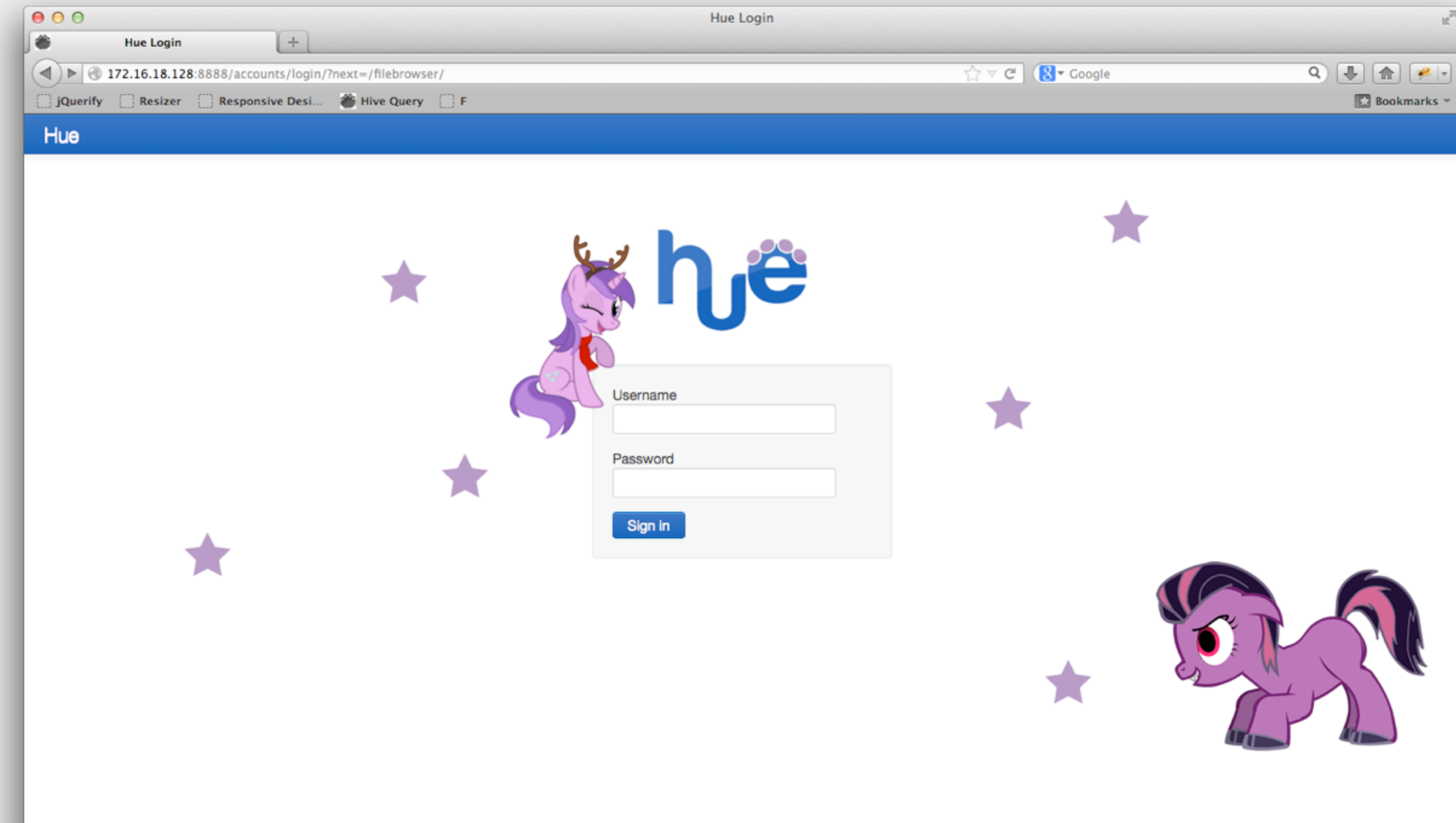
New apps, improved the UX adding new nice functionalities like autocomplete and drag & drop.



HISTORY

HUE 3 ALPHA

Proposed design, didn't make it.



HISTORY

HUE 3.5

New UI, several new apps, the most user friendly features to date.

The screenshot displays the Hue 3.5 interface. At the top, there's a navigation bar with 'HUE' logo and various menu items like 'Query Editors', 'Data Browsers', 'Workflows', 'Search', 'File Browser', 'Job Browser', and 'remain'. Below this is the 'Hive Editor' section with tabs for 'Query Editor', 'My Queries', 'Saved Queries', and 'History'. The 'Query Editor' tab is active, showing a SQL query for 'Sample: Salary growth'.

```
1 SELECT s07.description, s07.salary, s08.salary,
2       s08.salary - s07.salary
3 FROM
4       sample_07 s07 JOIN sample_08 s08
5 ON ( s07.code = s08.code)
6 WHERE
7       s07.salary < s08.salary
8 ORDER BY s08.salary-s07.salary DESC
9 LIMIT 20
```

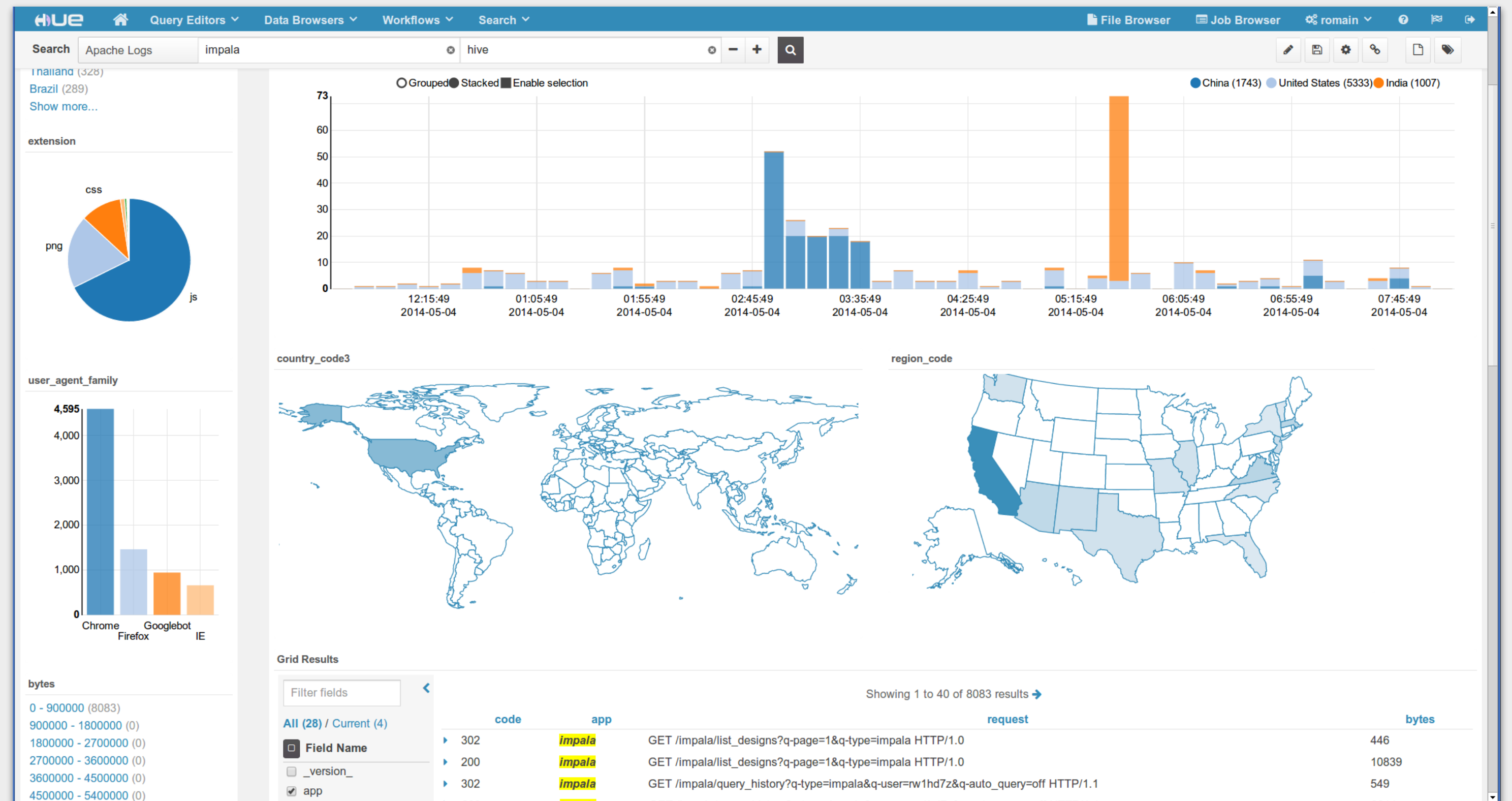
Below the query editor are buttons for 'Execute', 'Save', 'Save as...', 'Explain', and 'New query'. The 'Execute' button is highlighted.

The bottom section shows a 'Chart' view of the query results. The chart is a bar chart with 'description' on the X-axis and 'salary' on the Y-axis. The Y-axis ranges from 0 to 200,000. The X-axis labels include various professions and categories like 'Dentists, all other specialists', 'Surgeons', 'Oral and maxillofacial surgeons', 'Natural and dental managers', 'Physicians and surgeons', 'Orthodontists', 'Internists, general', 'Political scientists', 'Obstetricians and gynecologists', 'Chief executives', 'Rotary drill operators, oil and gas', 'Pediatricians', 'Sociologists', 'Family and general practitioners', 'Medical scientists, except epidemiologists', 'Athletes and sports competitors', 'Animal scientists', 'Dentists, general', 'Education administrators', and 'Psychologists, all other postsecondary'. The bars represent the salary for each category, with the highest salaries seen for 'Surgeons' and 'Oral and maxillofacial surgeons'.

HISTORY

HUE 3.6+

Where we are now, a brand new way to search and explore your data.



WHICH VERSION TO USE?

HUE 2.X

1-2 years old, use only if you depend on Hive < 0.12

HUE 3.X

2500+ commits later, new UI, interactive search / SQL, dashboards...

WHICH DISTRIBUTION?

HACKER



GITHUB

Very latest

ADVANCED USER



TARBALL

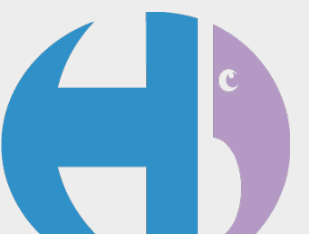
Advanced preview

NORMAL USER



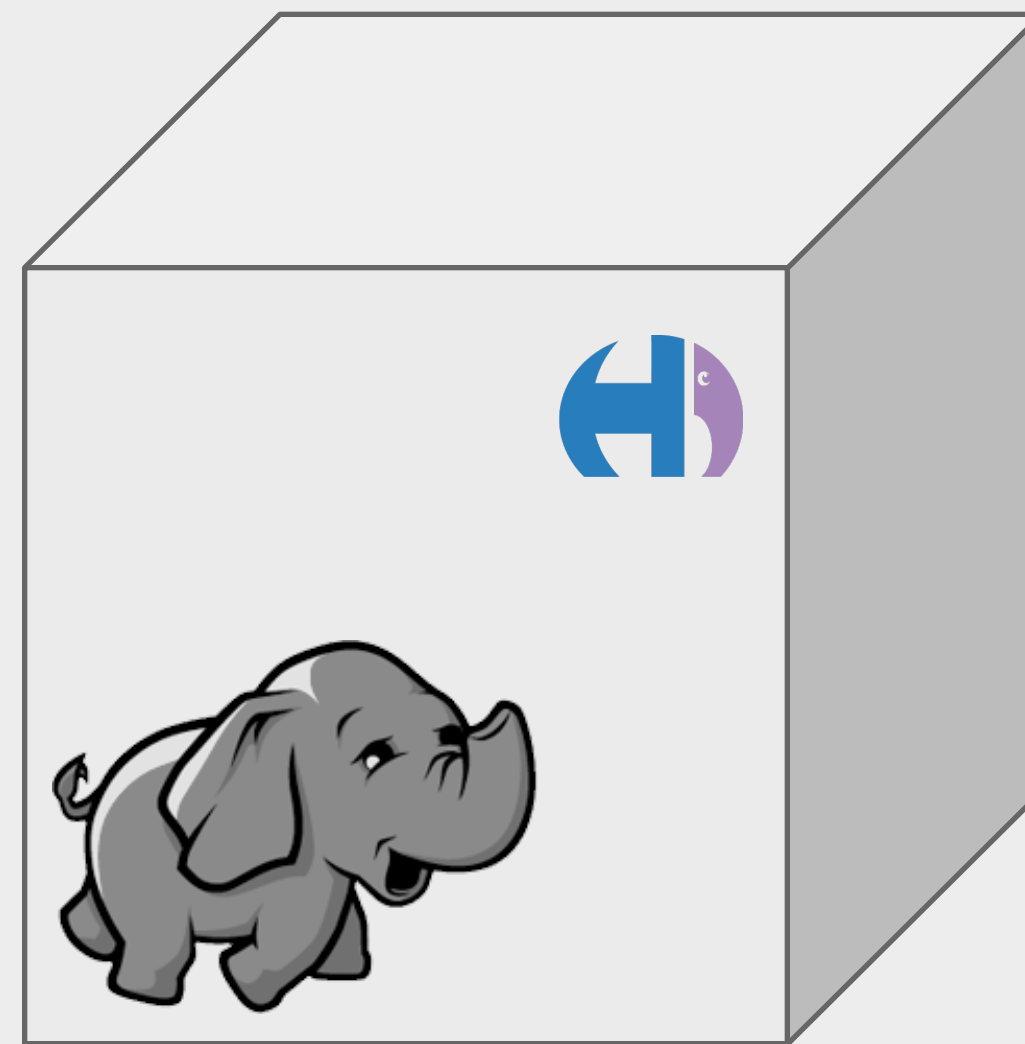
CDH / CM

The most stable and cross component checked



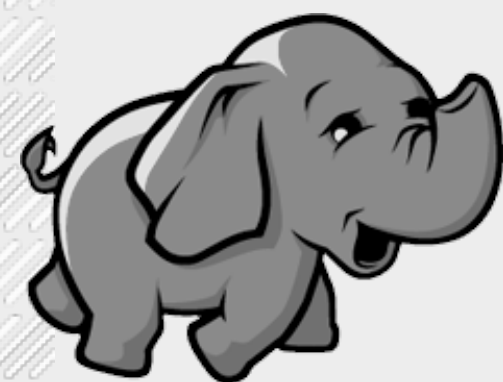
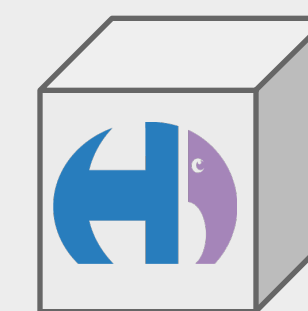
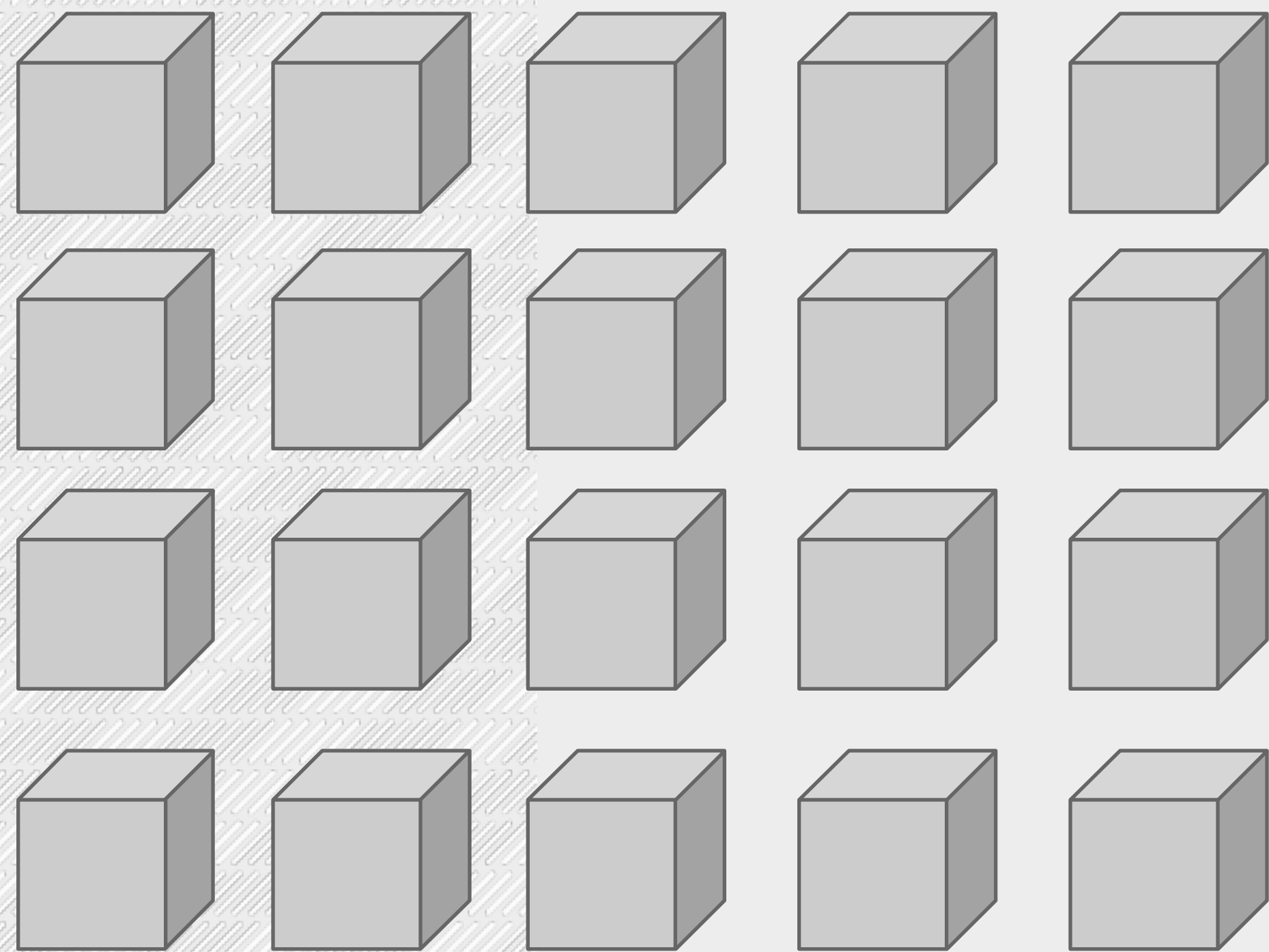
WHERE TO PUT HUE?

IN ONE MACHINE



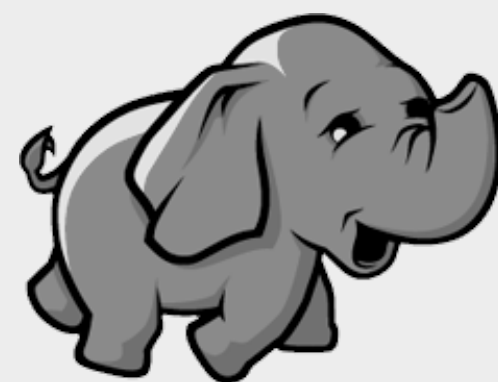
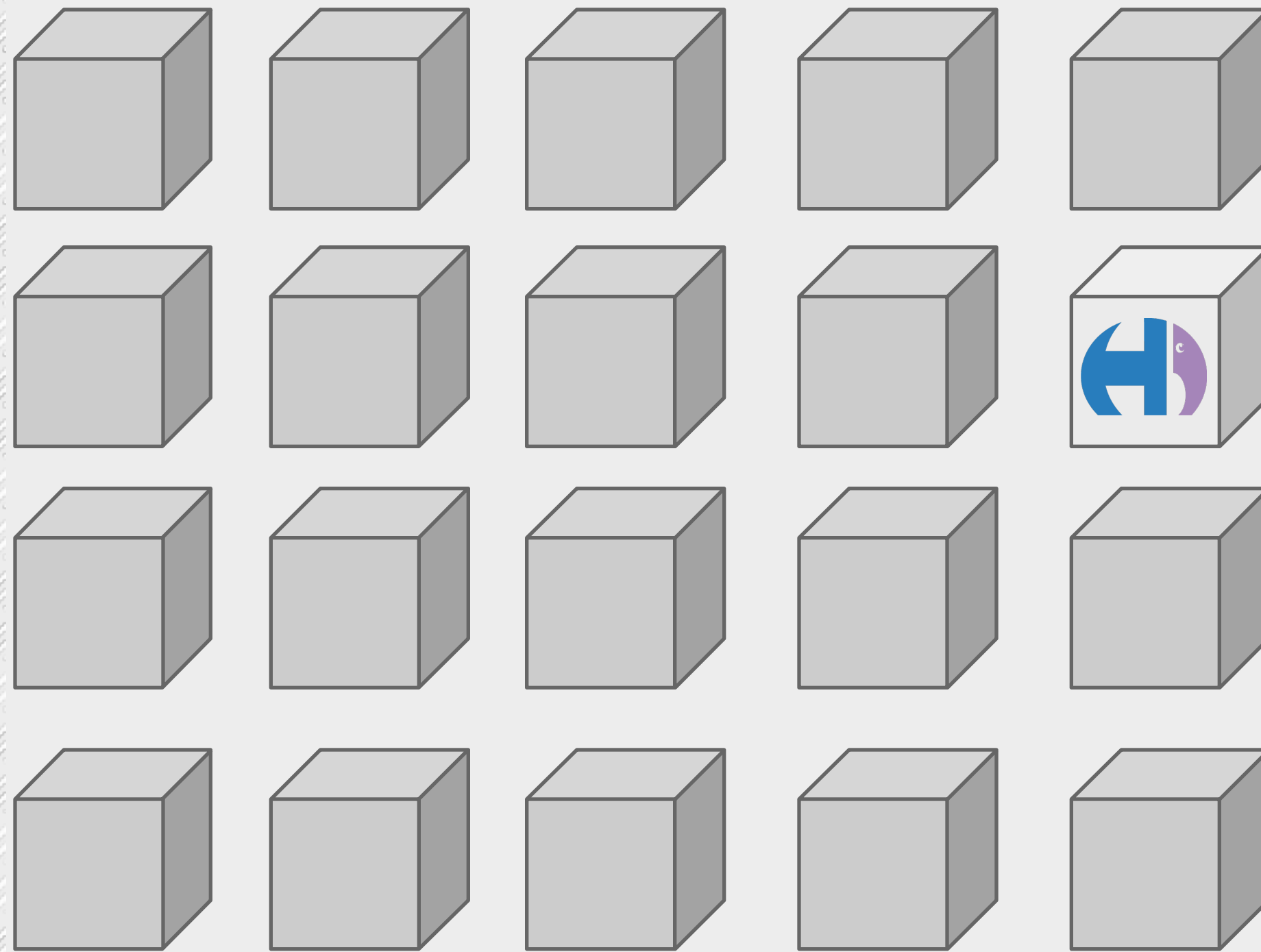
WHERE TO PUT HUE?

OUTSIDE THE CLUSTER



WHERE TO PUT HUE?

INSIDE THE CLUSTER



WHAT DO YOU NEED?

SERVER

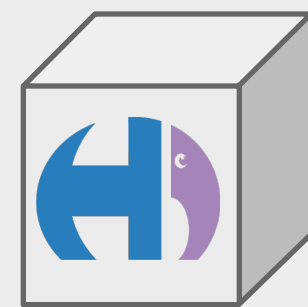
Python ~~2.4~~ 2.6

That's it if using a packaged version. If building from the source, here are the extra packages

CLIENT

Web Browser

IE 9+, FF 10+, Chrome, Safari



Hi there, I'm "just" a web server.

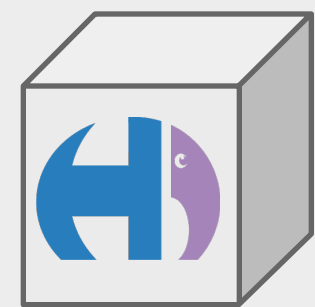
HOW DOES THE HUE SERVICE LOOK LIKE?

1 SERVER

Process serving pages and also static content

1 DB

For cookies, saved queries, workflows, ...



Hi there, I'm "just" a web server.

HOW TO CONFIGURE HUE

HUE.INI

Similar to core-site.xml but
with .INI syntax

Where?

```
/etc/hue/conf/hue.ini
```

or

```
$HUE_HOME/desktop/conf/  
pseudo-distributed.ini
```

[desktop]

[[database]]

```
# Database engine is typically one of:  
# postgresql_psycopg2, mysql, or sqlite3  
engine=sqlite3  
## host=  
## port=  
## user=  
## password=  
name=desktop/desktop.db
```

AUTHENTICATION

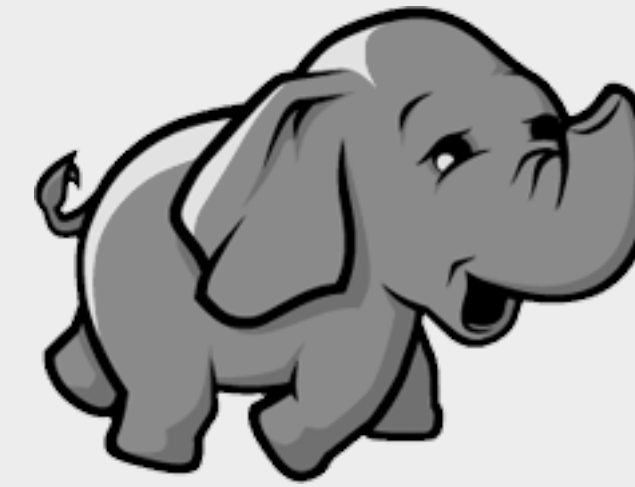
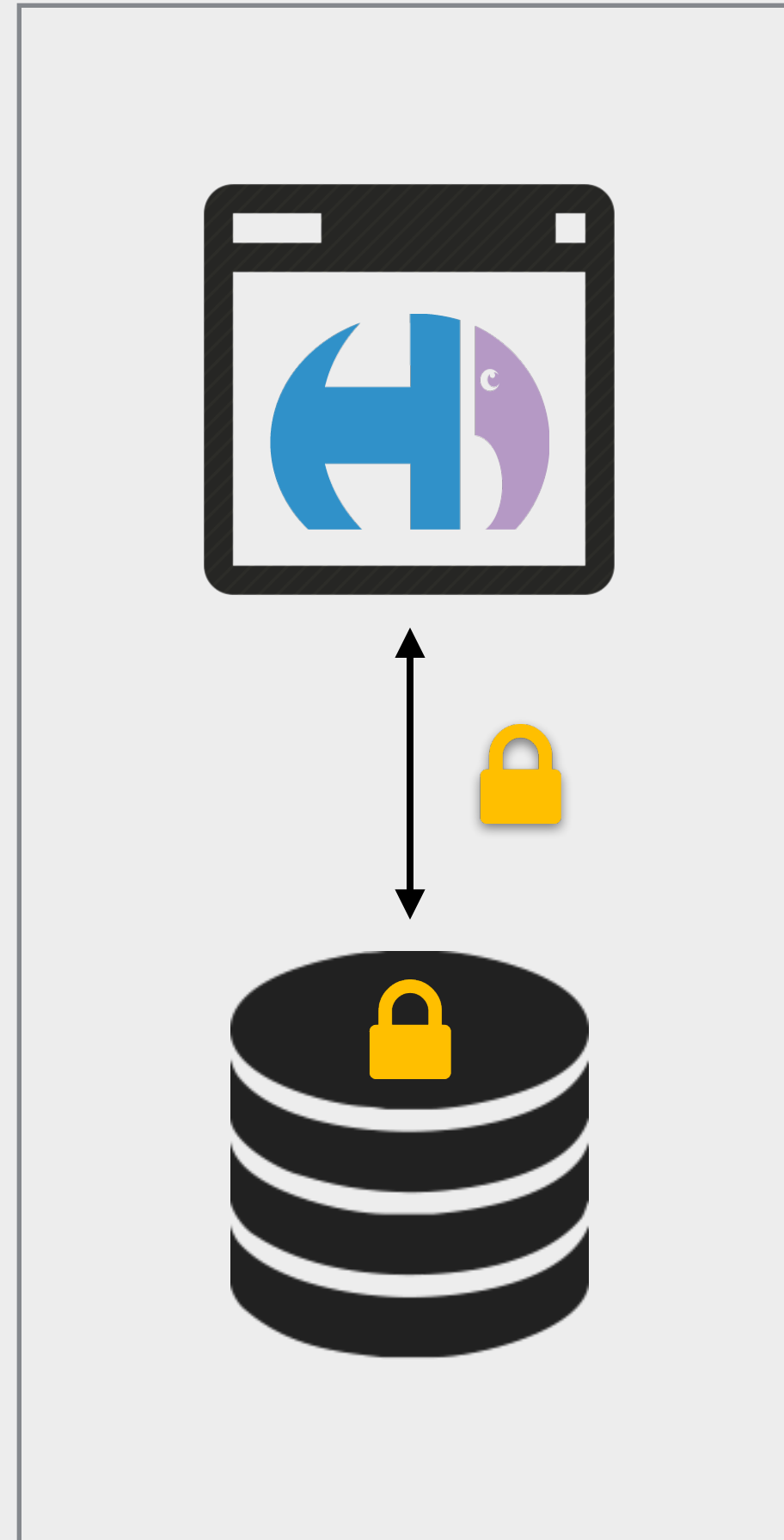
SIMPLE

Login/Password in a Database
(SQLite, MySQL, ...)

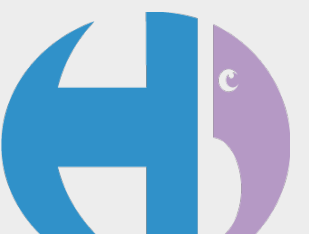
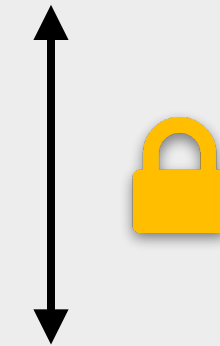
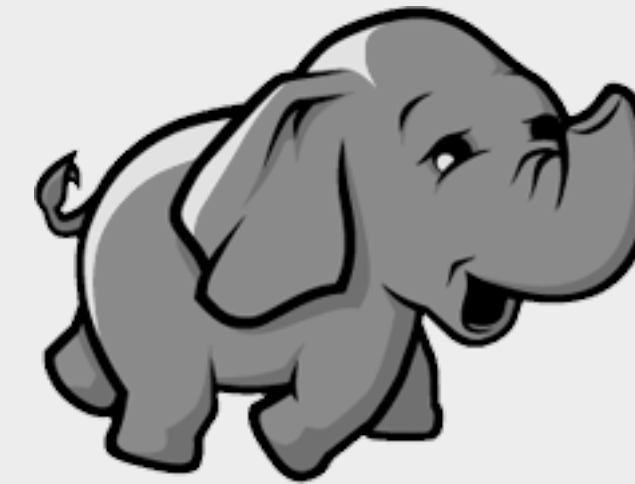
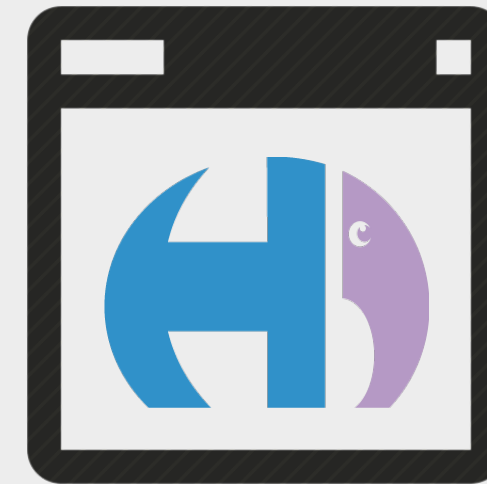
ENTERPRISE

LDAP (most used), OAuth,
OpenID, SAML

DB BACKEND



LDAP BACKEND



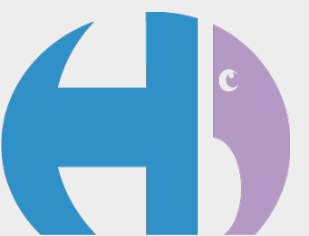
USERS

ADMIN

Can give and revoke permissions to single users or group of users

USER

Regular user + permissions

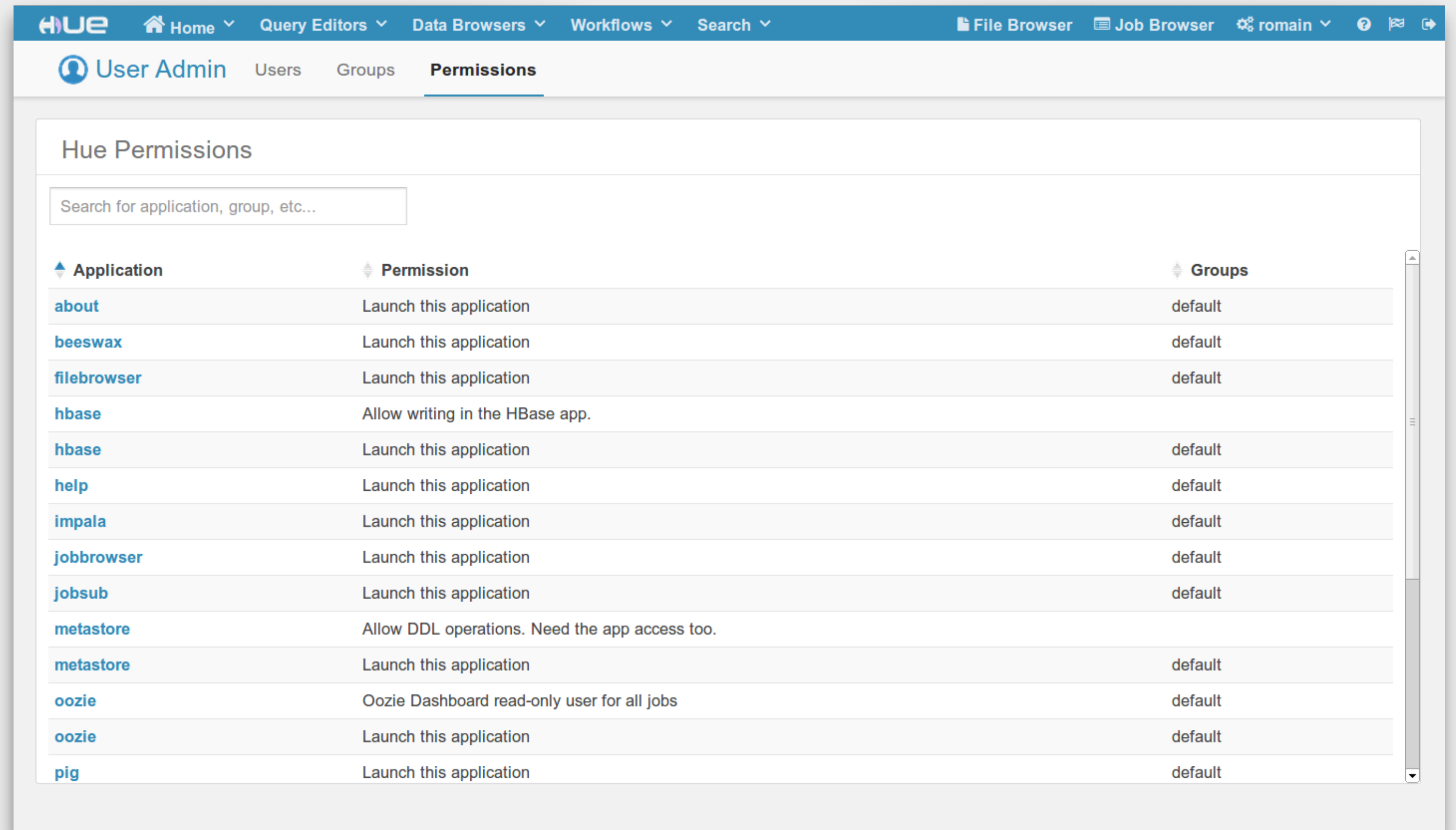


CONFIGURE APPS AND PERMISSIONS

LIST OF GROUPS AND PERMISSIONS

A permission can:

- allow access to one app (e.g. Hive Editor)
- modify data from the app (e.g. drop Hive Tables or edit cells in HBase Browser)



The screenshot shows the Hue User Admin interface, specifically the 'Permissions' tab. The page title is 'Hue Permissions'. There is a search bar with the placeholder text 'Search for application, group, etc...'. Below the search bar is a table with three columns: 'Application', 'Permission', and 'Groups'. The table lists various applications and their associated permissions for the 'default' group.

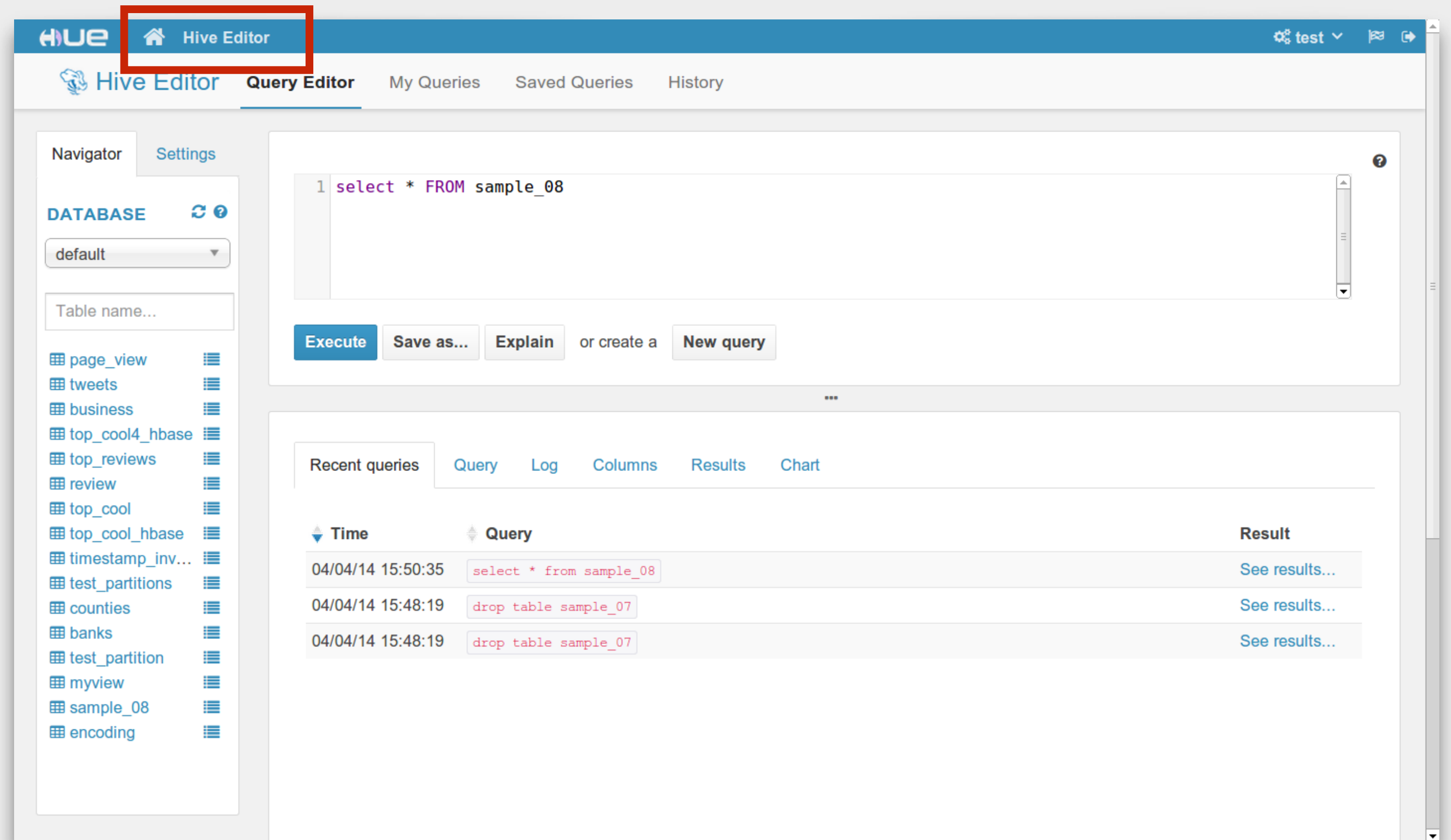
Application	Permission	Groups
about	Launch this application	default
beeswax	Launch this application	default
filebrowser	Launch this application	default
hbase	Allow writing in the HBase app.	
hbase	Launch this application	default
help	Launch this application	default
impala	Launch this application	default
jobbrowser	Launch this application	default
jobsub	Launch this application	default
metastore	Allow DDL operations. Need the app access too.	
metastore	Launch this application	default
oozie	Oozie Dashboard read-only user for all jobs	default
oozie	Launch this application	default
pig	Launch this application	default

A list of permissions

CONFIGURE APPS AND PERMISSIONS

PERMISSIONS IN ACTION

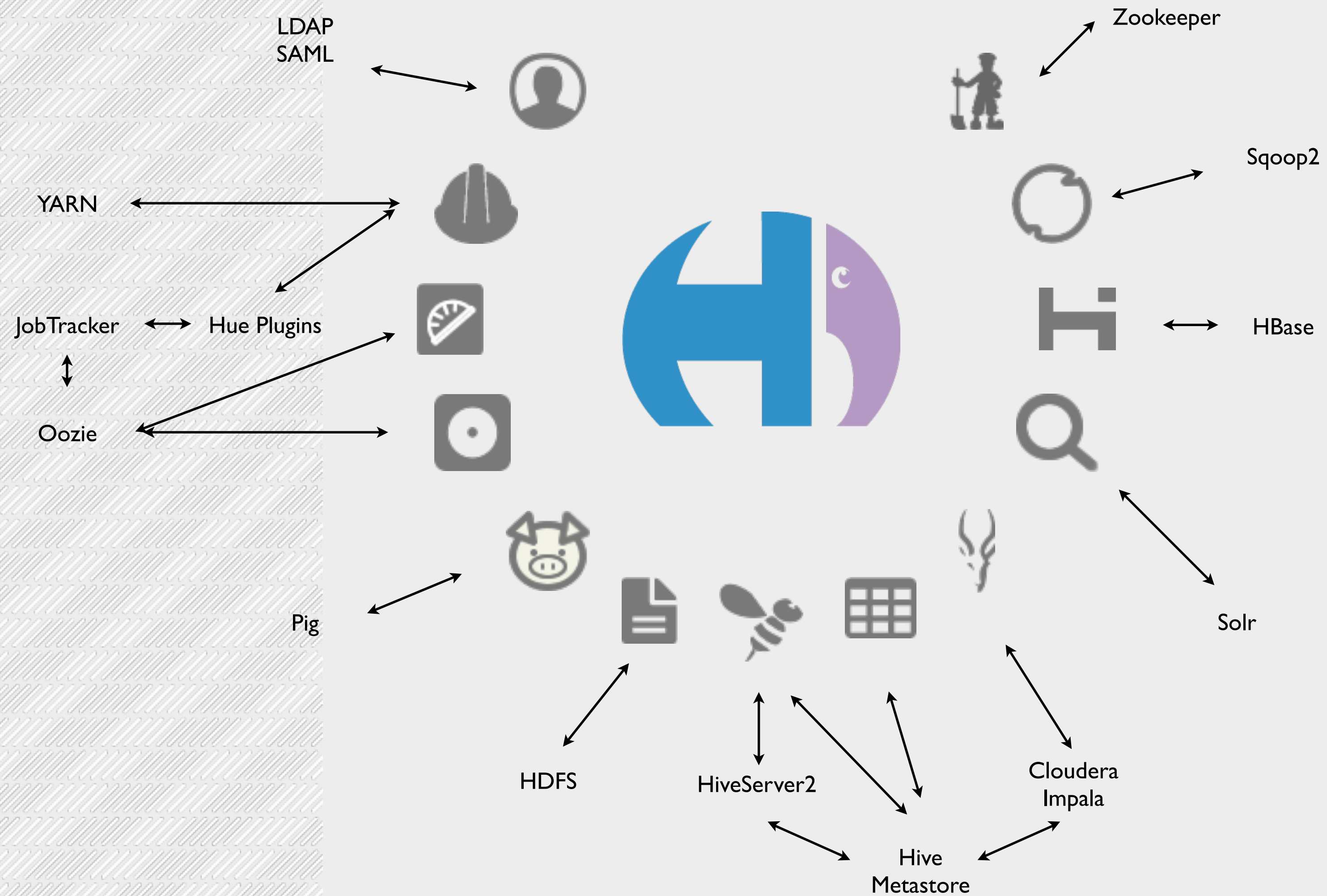
User 'test' belonging to the group 'hiveonly' that has just the 'hive' permissions



The screenshot shows the Cloudera Hue interface. The top navigation bar includes the Hue logo and a 'Hive Editor' tab, which is highlighted with a red box. Below the navigation bar, there are tabs for 'Hive Editor', 'Query Editor', 'My Queries', 'Saved Queries', and 'History'. The main interface is divided into a left sidebar and a main content area. The sidebar contains a 'Navigator' section with a 'DATABASE' dropdown set to 'default' and a 'Table name...' input field. Below this is a list of tables: page_view, tweets, business, top_cool4_hbase, top_reviews, review, top_cool, top_cool_hbase, timestamp_inv..., test_partitions, counties, banks, test_partition, myview, sample_08, and encoding. The main content area features a SQL editor with the query '1 select * FROM sample_08'. Below the editor are buttons for 'Execute', 'Save as...', 'Explain', and 'New query'. At the bottom, the 'Recent queries' section displays a table with the following data:

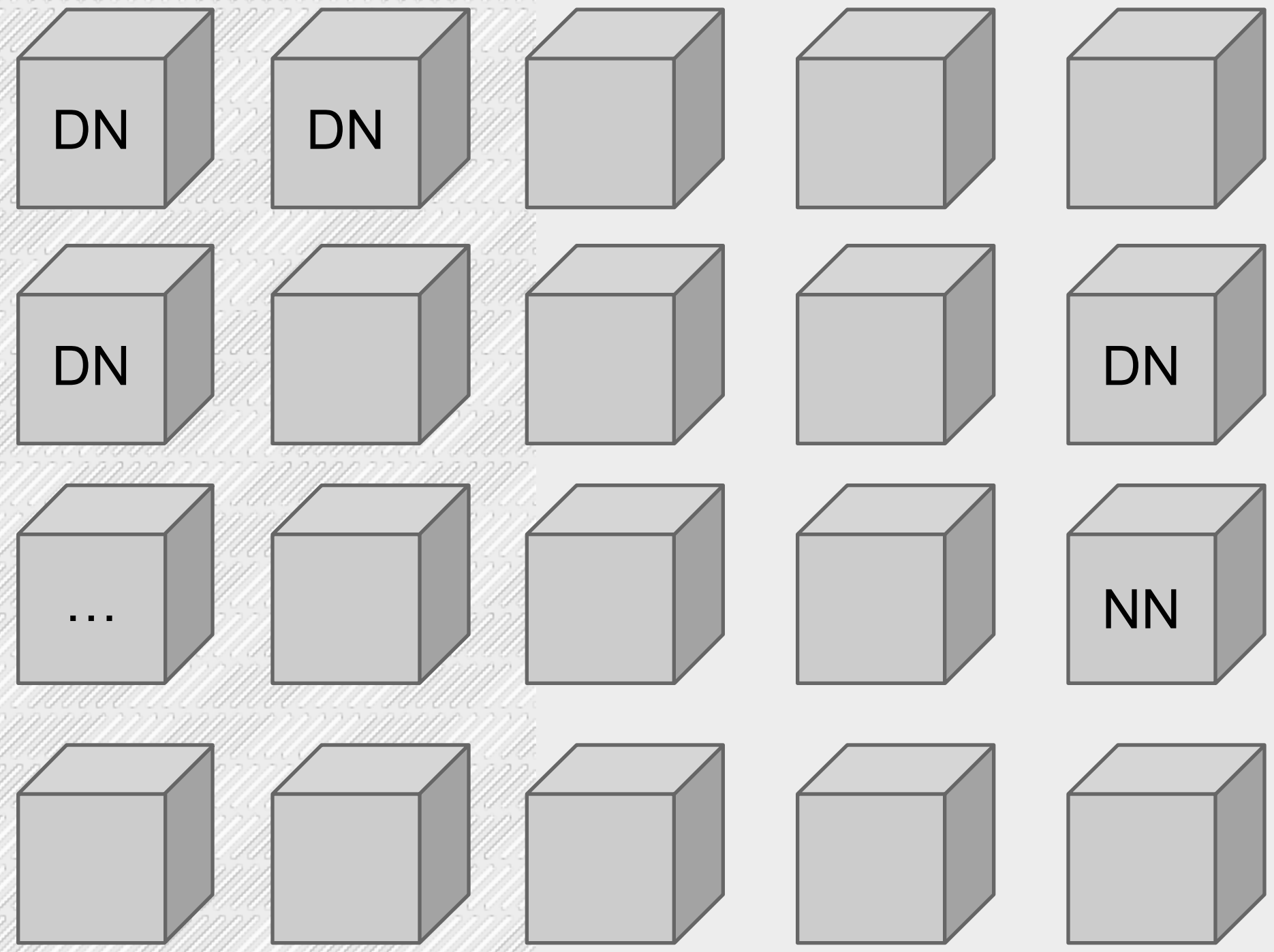
Time	Query	Result
04/04/14 15:50:35	select * from sample_08	See results...
04/04/14 15:48:19	drop table sample_07	See results...
04/04/14 15:48:19	drop table sample_07	See results...

HOW HUE INTERACTS WITH HADOOP

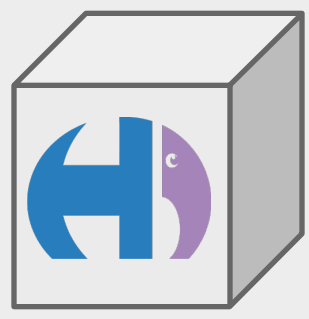


RCP CALLS TO ALL THE HADOOP COMPONENTS

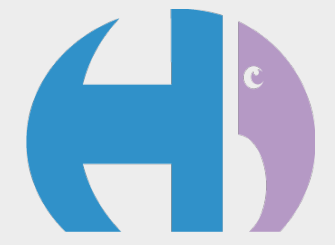
HDFS EXAMPLE



WebHDFS
REST



<http://localhost:50070/webhdfs/v1/<PATH>?op=LISTSTATUS>



RCP CALLS TO ALL THE HADOOP COMPONENTS

HOW

List all the host/port of Hadoop APIs in the hue.ini

For example here HBase and Hive.

[hbase]

```
# Comma-separated list of HBase Thrift servers for  
# clusters in the format of '(name|host:port)'.  
hbase_clusters=(Cluster|localhost:9090)
```

[beeswax]

```
hive_server_host=host-abc  
hive_server_port=10000
```

[Full list](#)

SECURITY FEATURES

HTTPS

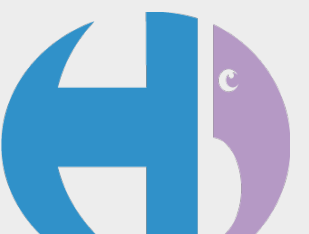
SSL WITH HIVESERVER2

SSL DB

AUDITING

KERBEROS

[READ MORE ...](#)



HIGH AVAILABILITY

HOW

2 Hue instances

HA proxy

Multi DB

Performances: like a website,
mostly RPC calls

All Services Add Cluster

Cluster 1 - CDH4 Actions

Name	Status	Role Counts	Actions
FLUME-1	Concerning Health	2 Agents	Actions
HBASE-1	Concerning Health	2 RegionServers, 1 Master, 1 Gateway	Actions
HDFS-1	Bad Health	1 SecondaryNameNode, 1 NameNode, 1 Balancer, 2 DataNodes, 1 Gateway	Actions
HIVE-1	Good Health	1 Hive Metastore Server, 2 Gateways	Actions
HUE-1	Good Health	1 Beeswax Server, 1 Hue Server	Actions
HUE-2	Good Health	1 Beeswax Server, 1 Hue Server	Actions
IMPALA-1	Concerning Health	2 Impala Daemons, 1 Impala StateStore Daemon	Actions
MAPREDUCE-1	Concerning Health	1 JobTracker, 2 TaskTrackers, 1 Gateway	Actions
OOZIE-1	Good Health	1 Oozie Server	Actions
SOLR-1	Good Health	2 Solr Servers	Actions
SQOOP-1	Good Health	1 Sqoop Server	Actions
YARN-1	Good Health	1 JobHistory Server, 2 NodeManagers, 1 Gateway, 1 ResourceManager	Actions
ZOOKEEPER-1	Concerning Health	1 Server	Actions

SQL

WHAT

Impala, Hive integration, Spark
(Shark too)

Interactive SQL editor

Integration with MapReduce,
Metastore, HDFS

The screenshot displays the Cloudera Hue interface. At the top, there's a navigation bar with 'HUE' logo, home icon, and menu items: 'Query Editors', 'Data Browsers', 'Workflows', 'Search', 'File Browser', 'Job Browser', and 'remain'. Below this is the 'Hive Editor' section with tabs for 'Query Editor', 'My Queries', 'Saved Queries', and 'History'. On the left, a 'Navigator' sidebar shows a 'DATABASE' dropdown set to 'default' and a list of tables including 'page_view', 'tweets', 'business', and various columns like 'city (string)', 'review_count (int)', etc. The main area shows a SQL query titled 'Sample: Salary growth' with the following code:

```
1 SELECT s07.description, s07.salary, s08.salary,  
2       s08.salary - s07.salary  
3 FROM  
4 sample_07 s07 JOIN sample_08 s08  
5 ON ( s07.code = s08.code)  
6 WHERE  
7 s07.salary < s08.salary  
8 ORDER BY s08.salary-s07.salary DESC  
9 LIMIT 20
```

Below the query are buttons for 'Execute', 'Save', 'Save as...', 'Explain', and 'New query'. The bottom section shows a 'Chart' visualization of the query results. The chart is a bar chart with 'description' on the X-axis and 'salary' on the Y-axis. The Y-axis ranges from 0 to 200,000. The X-axis labels include various professions and categories like 'Dentists, all other specialists', 'Surgeons', 'Oral and maxillofacial sciences', 'Natural and dental surgeons', 'Physicians and surgeons', 'Orthodontists', 'Internists, general', 'Political scientists', 'Obstetricians and gynecologists', 'Chief executives', 'Rotary drill operators, oil and gas', 'Pediatricians', 'Sociologists', 'Family and general practitioners', 'Medical scientists, except epidemiologists', 'Athletes and sports competitors', 'Animal scientists', 'Dentists, general', 'Education administrators', and 'Psychologists, all other postsecondary'. The bars represent the salary for each category, with the highest salaries seen for 'Surgeons' and 'Oral and maxillofacial sciences'.

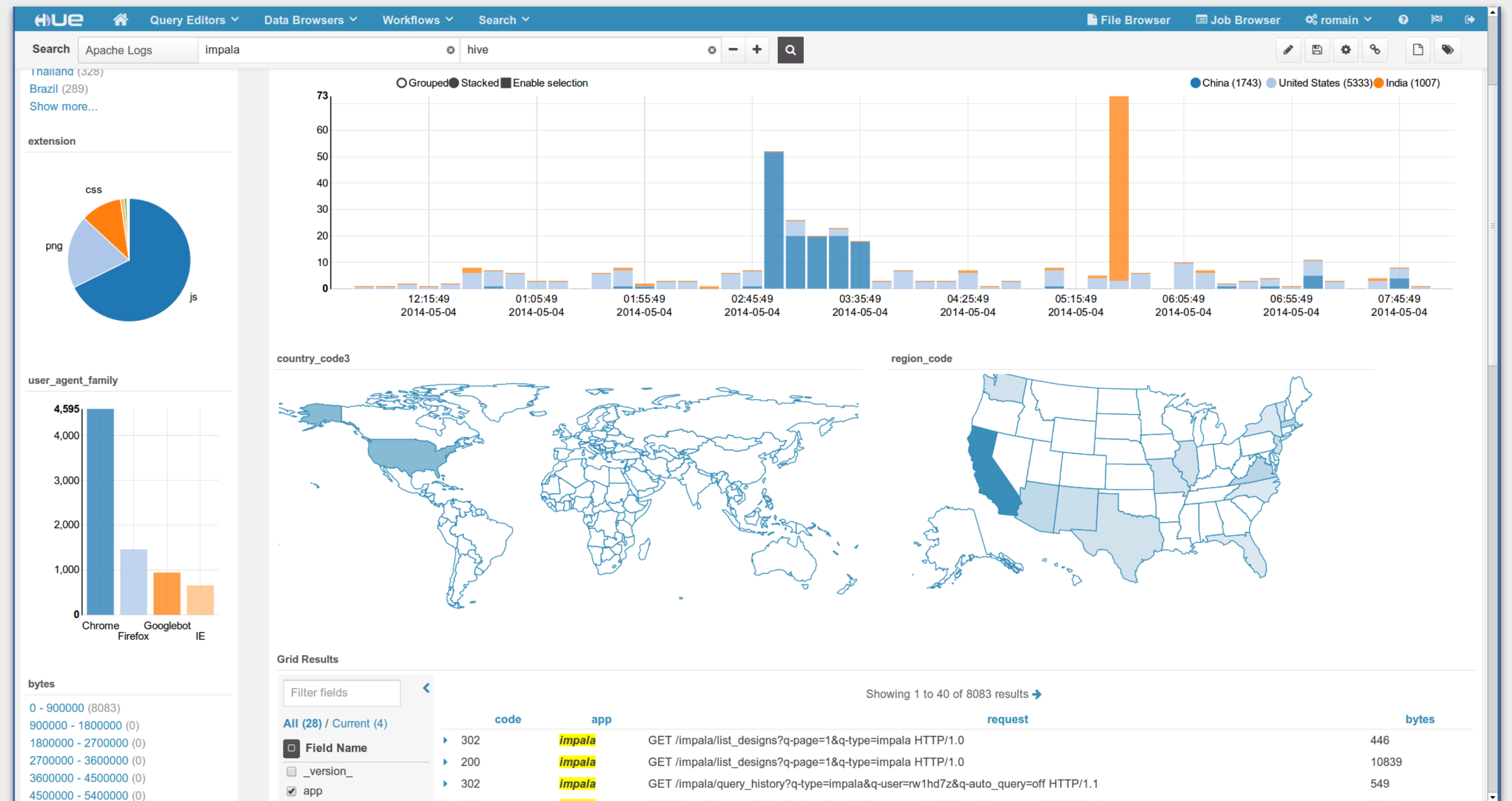
SEARCH

WHAT

Solr & Cloud integration

Custom interactive dashboards

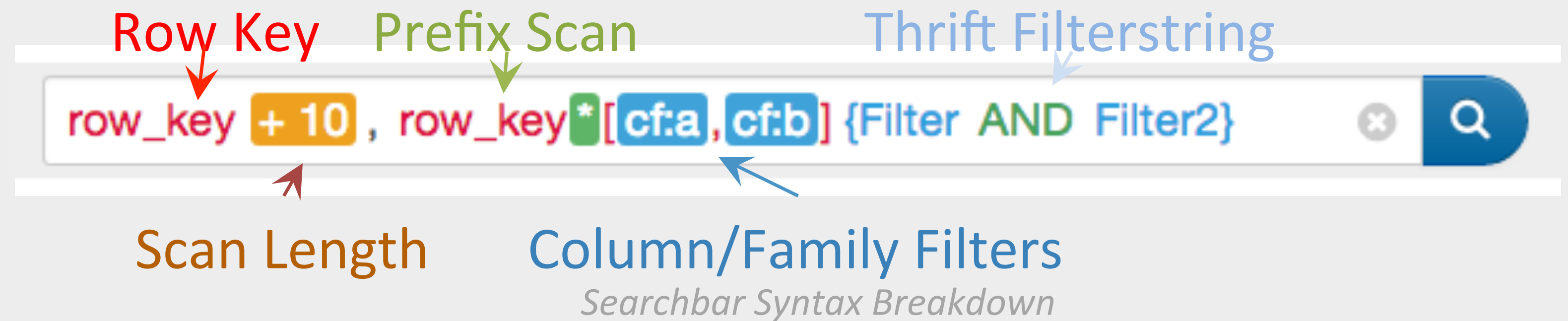
Drag & drop widgets (charts, timeline...)



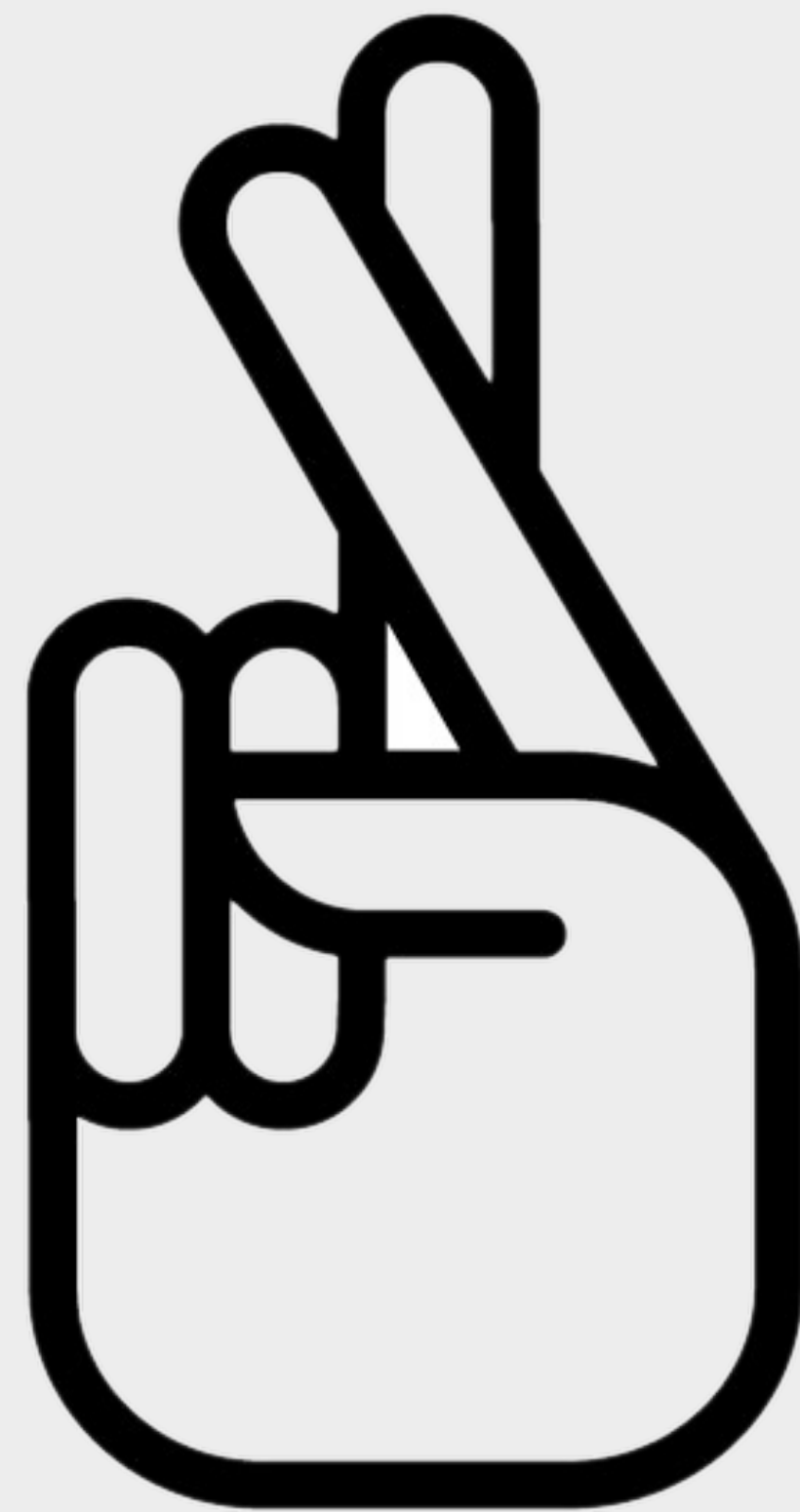
HBASE BROWSER

WHAT

- Simple custom query language
- Supports HBase filter language
- Supports selection & Copy + Paste, gracefully degrades in IE
- Autocomplete Help Menu



DEMO TIME



SUM-UP

INSTALL

Install Hue on one machine

LDAP

Use an LDAP backend

ENABLE

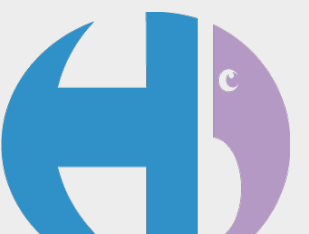
Enable Hadoop Service APIs for Hue as a proxy user

HELP

Get help on @gethue or hue-user

CONFIGURE

Configure hue.ini to point to each Service API



ROADMAP NEXT 6 MONTHS

WHAT

Sentry

Search, Spark, SQL

More dashboards!

Oozie v2

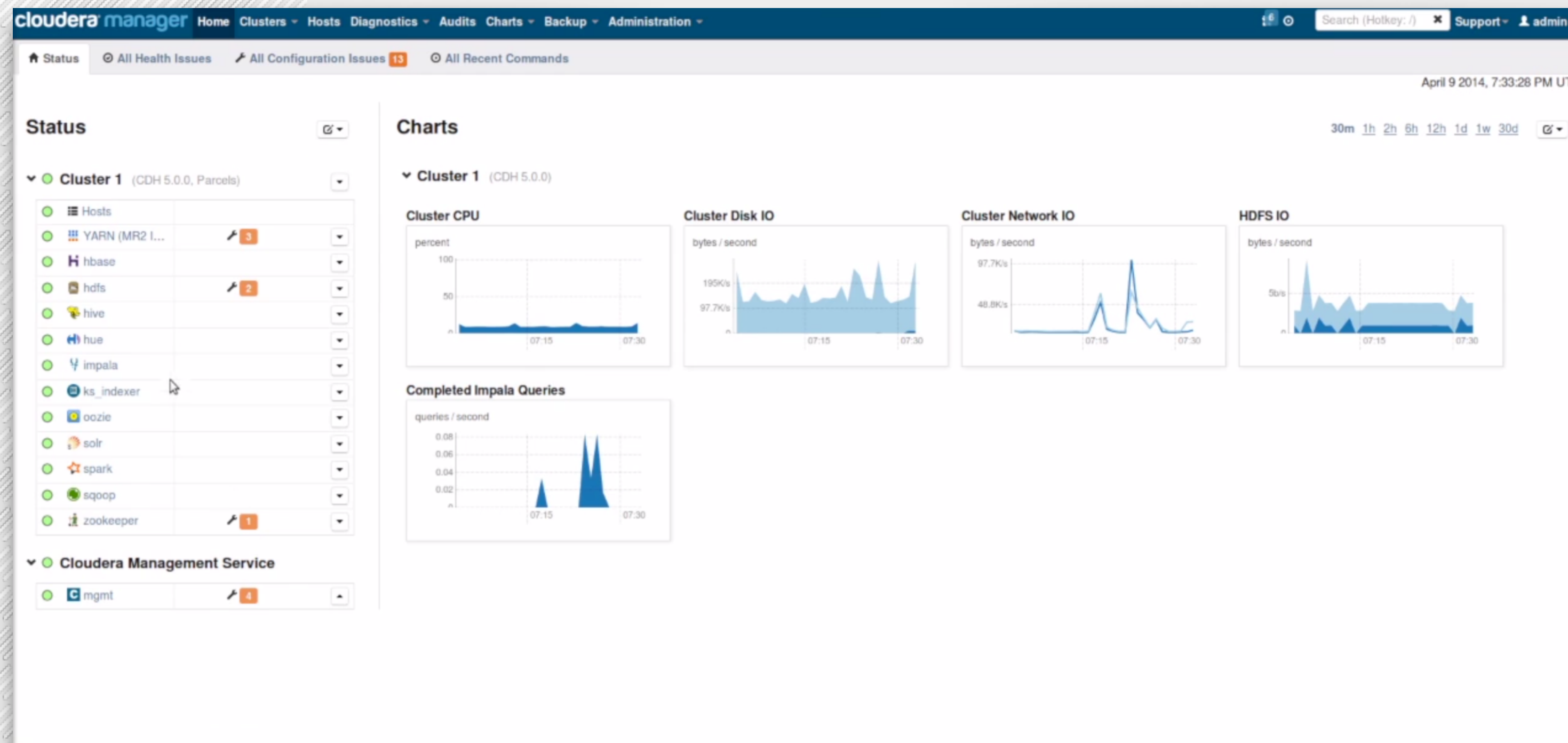
Inter component integrations
(HBase <-> Search, create index
wizards, document permissions),
Hadoop Web apps SDK

Your idea here.



CONFIGURATIONS ARE HARD...

...GIVE CLOUDERA MANAGER A TRY!



vimeo.com/91805055

MISSED
SOMETHING?

learn.gethue.com

Tutorial

Recent Posts

- Interactive SQL on Hadoop is now easier: New Impala and Hive Editors >
April 4, 2014
- How to fix the Map chart on Hue 3.5 >
April 4, 2014
- Monitor and get alerts for your workflows with the Oozie SLAs >
April 3, 2014
- Oozie workflow credentials with a Hive action with Kerberos >
April 2, 2014
- Hadoop Voice Recognition: use your voice to use Hadoop! >
April 1, 2014

Categories

- Administration (5) >
- DBQuery (1) >

Interactive SQL on Hadoop is now easier: New Impala and Hive Editors

By Hue Team on April 4, 2014

The Impala and Hive applications have a new face! In addition to new features, their interfaces in Hue have...

0 0 0

Hadoop Voice Recognition: use your voice to use Hadoop!

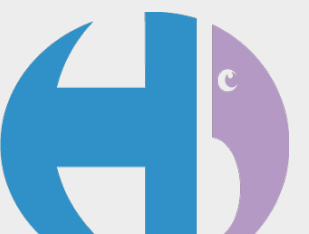
By Hue Team on April 1,

Monitor for your workflows with the Oozie SLAs

By Hue Team on April 3, 2014

Hue now has SLAs for your workflows, you can get alerts when so...

Hadoop Scheduler: Hadoop's intuitive...



THANK YOU!

WEBSITE

<http://gethue.com>

LEARN

<http://learn.gethue.com>

TWITTER

[@gethue](https://twitter.com/gethue)

USER GROUP

hue-user@

