

BIG DATA WEB APPS FOR INTERACTIVE HADOOP

Enrico Berti

Big Data Spain, Nov 17, 2014

cloudera®



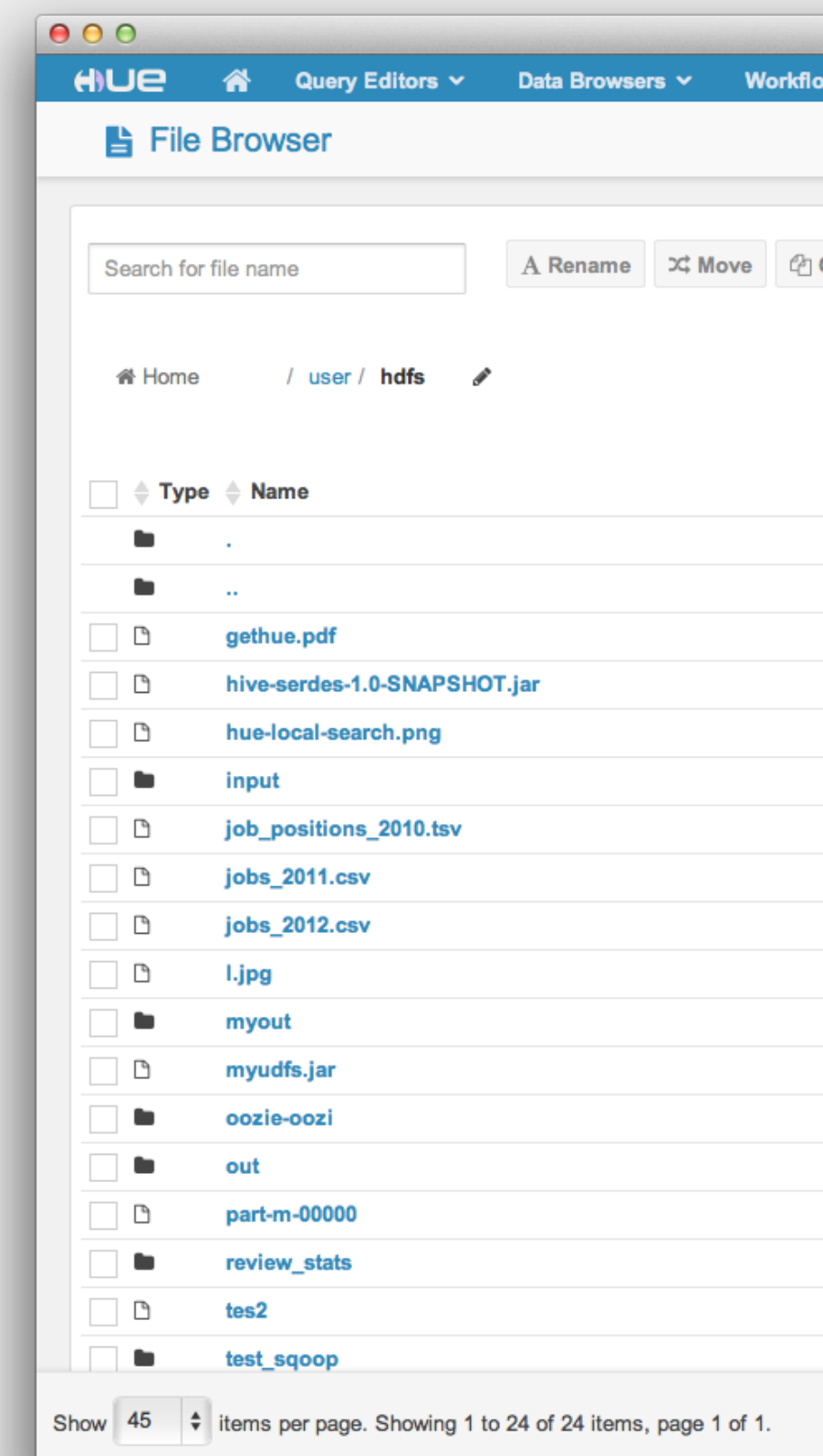
GOAL OF HUE

WEB INTERFACE FOR ANALYZING DATA
WITH APACHE HADOOP

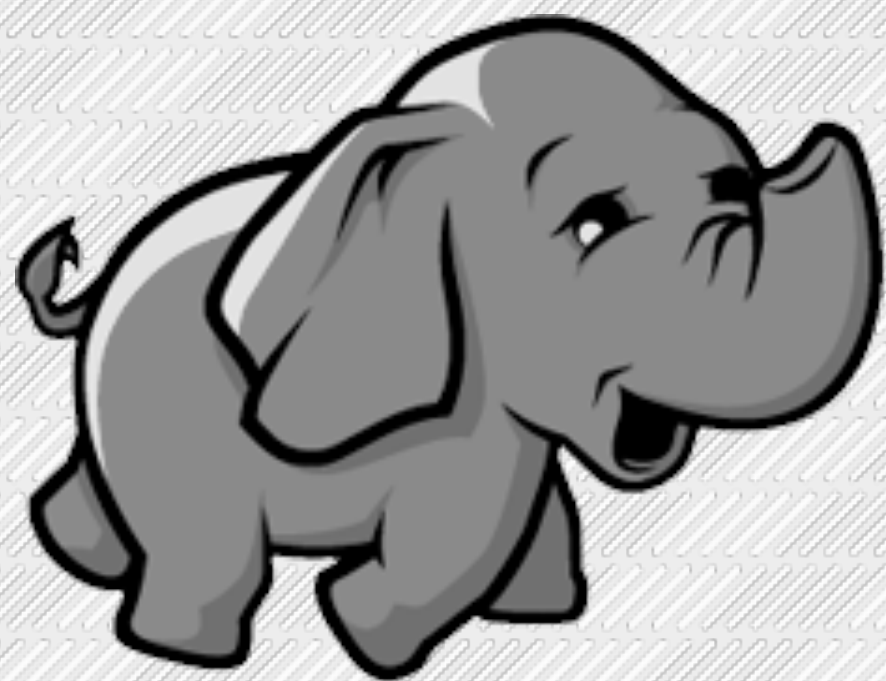
SIMPLIFY AND INTEGRATE

FREE AND OPEN SOURCE

—> OPEN UP BIG DATA



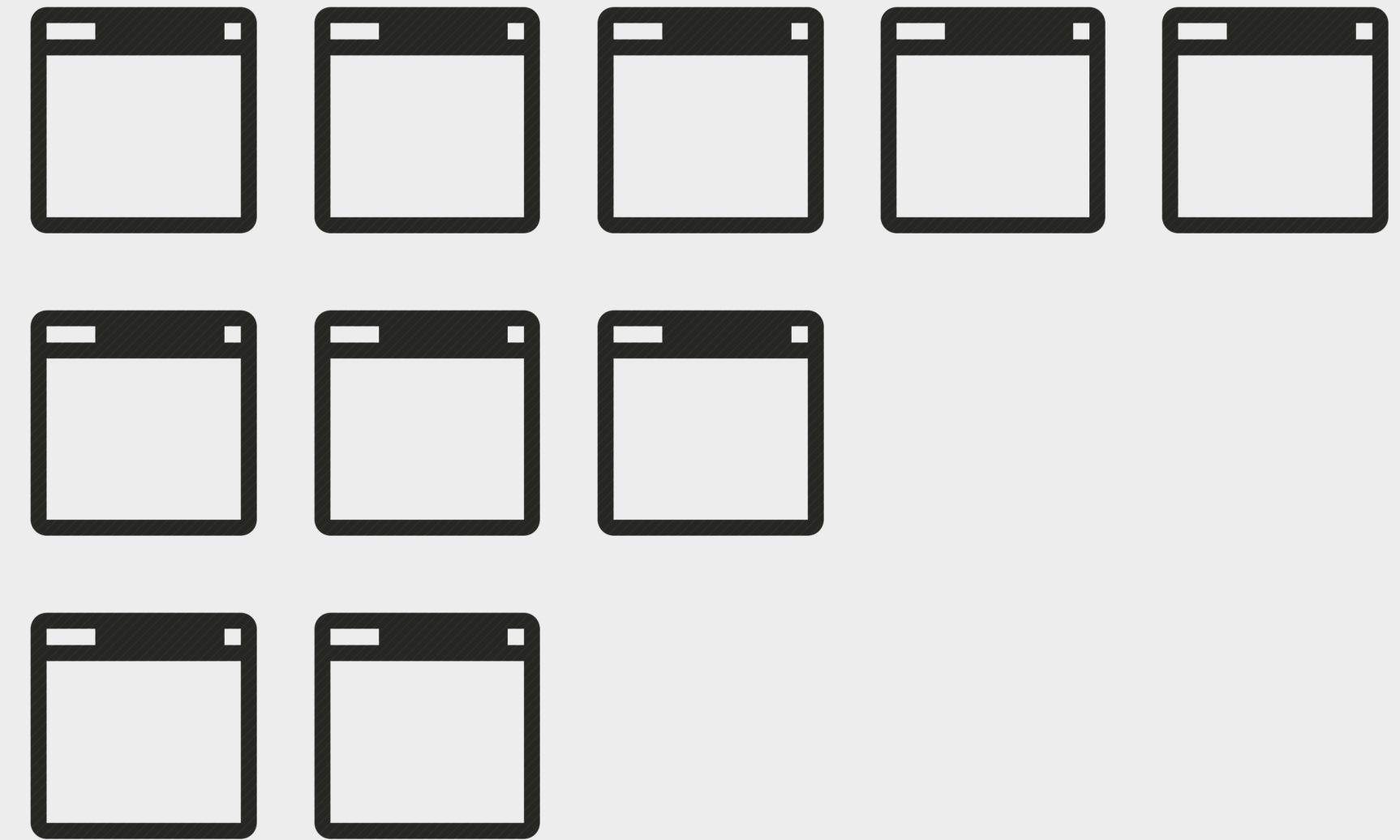
VIEW FROM 30K FEET



Hadoop



Web Server



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

OPEN SOURCE

~4000 **COMMITTS**

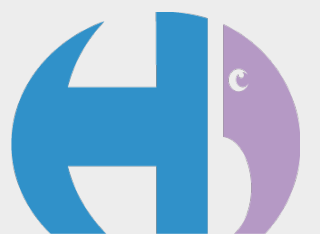
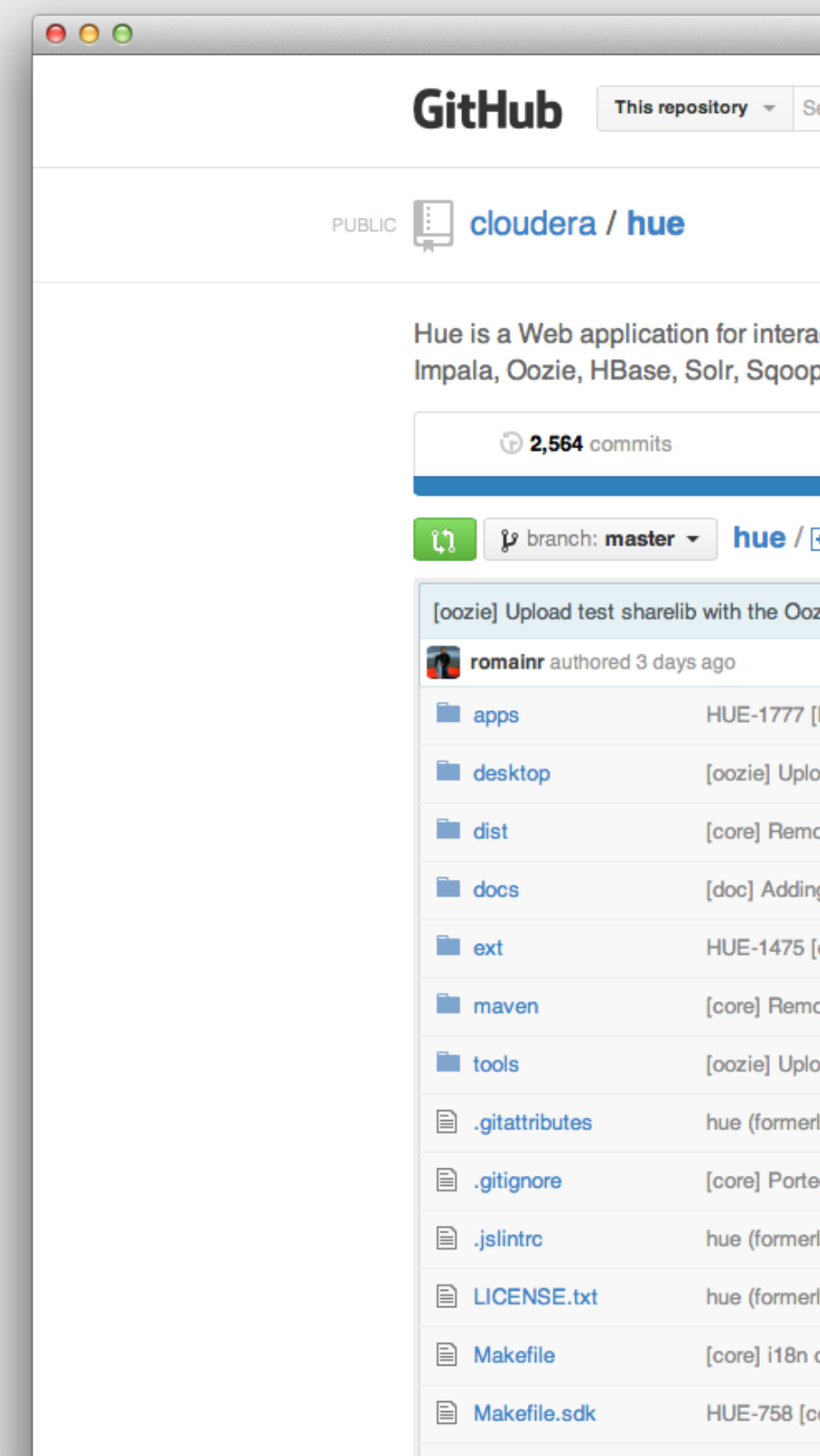
56 **CONTRIBUTORS**

911 **STARS**

337 **FORKS**

github.com/cloudera/hue

cloudera®



AROUND THE WORLD

TALKS

Meetups and events in NYC, Paris, LA, Tokyo, SF, Stockholm, Vienna, San Jose, Singapore, Budapest, DC, Madrid...



RETREATS

Nov 13 Koh Chang, Thailand

May 14 Curaçao, Netherlands Antilles

Aug 14 Big Island, Hawaii

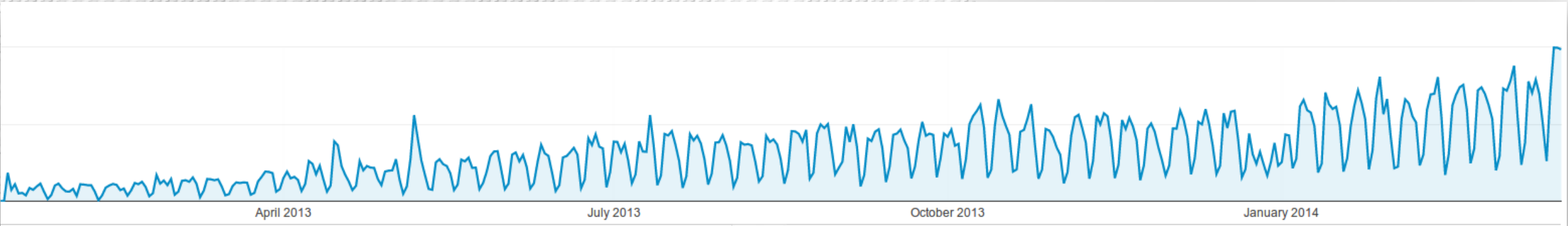
Nov 14 Tenerife, Spain

Nov 14 Nicaragua and Belize

Jan 15 Philippines



TREND: GROWTH



gethue.com

Mar 13, 2014



[impala] Replace non utf8 encoded chars from the resultsets

romainr authored a month ago

d6c869c498

Browse code



[liboauth] Limit oauth login to specific list of domains

dominikgehl authored a month ago

➔ romainr committed a month ago

76064c61ae

Browse code



[liboauth] Show error message after Aouth login failure

dominikgehl authored a month ago

➔ romainr committed a month ago

7307bf69be

Browse code



[liboauth] Indentation fix

dominikgehl authored a month ago

➔ romainr committed a month ago

5cee8ea5eb

Browse code



HUE-2020 [hbase] upload an image sometimes produces an error

abec authored a month ago

49bbfc8786

Browse code



[hbase] Fix new cell upload

abec authored a month ago

cf39ae23b4

Browse code



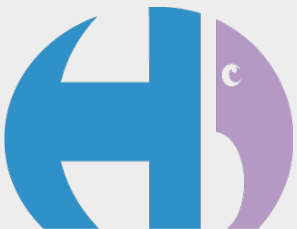








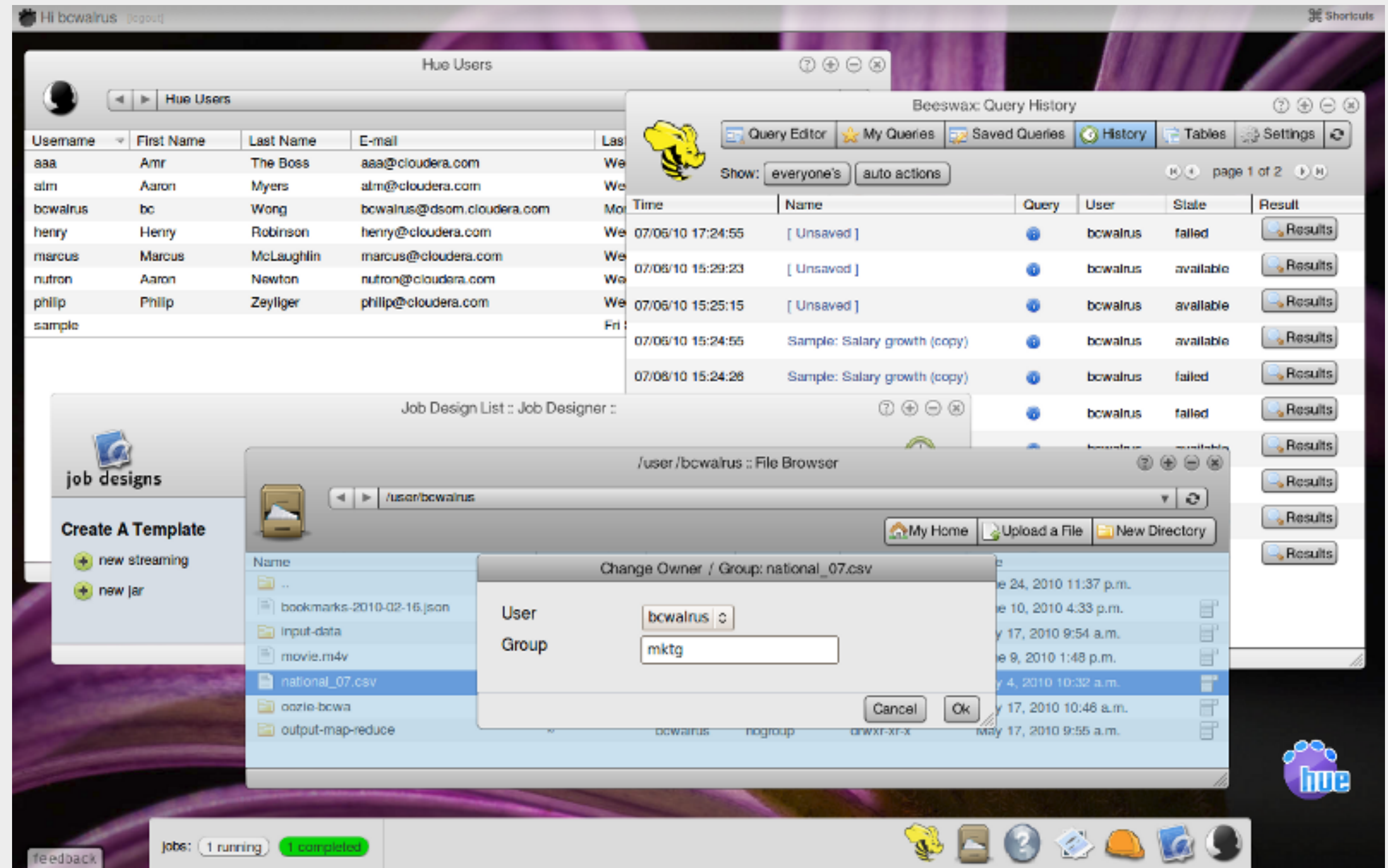




HISTORY

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).



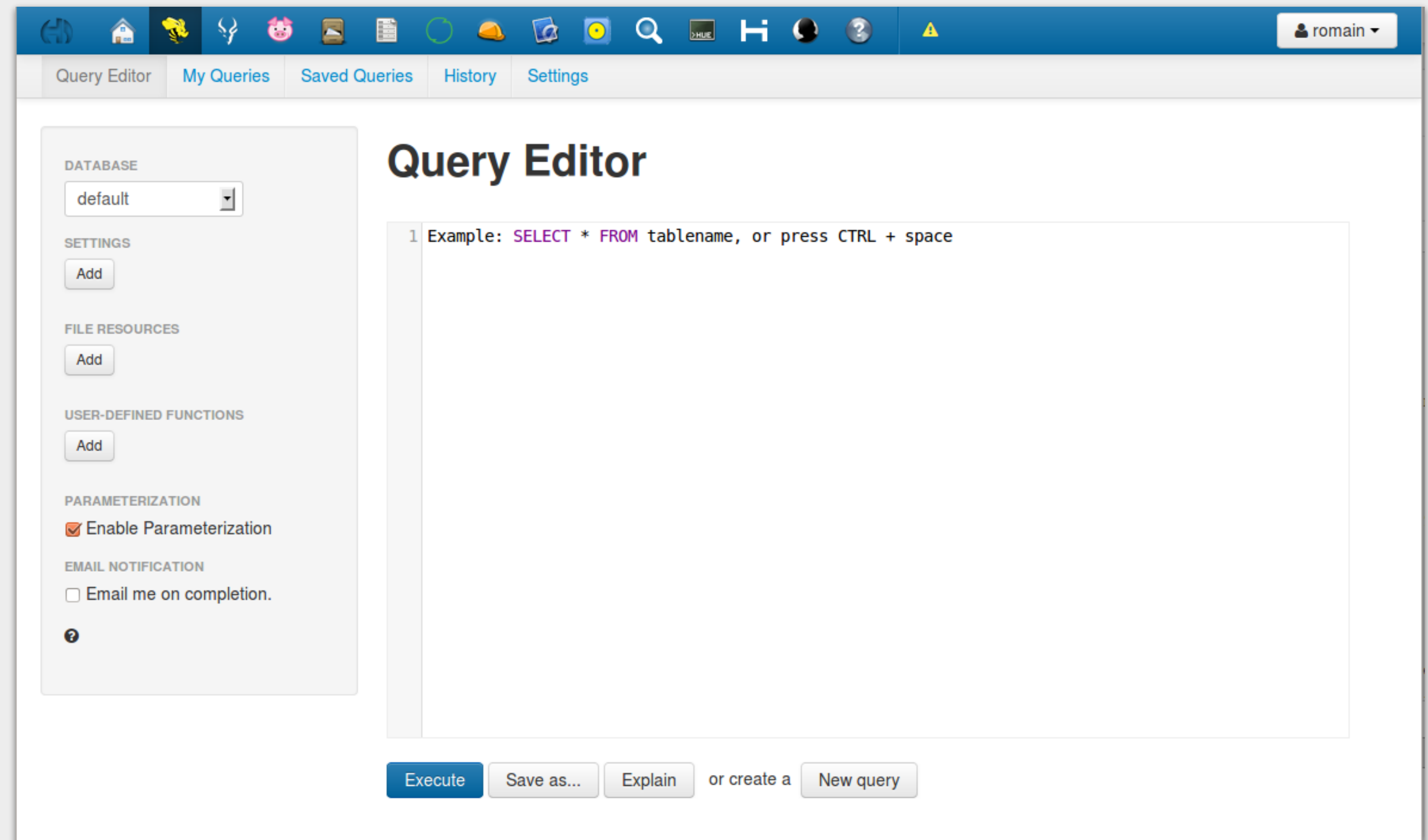
HISTORY

HUE 2

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

HUE 2.5

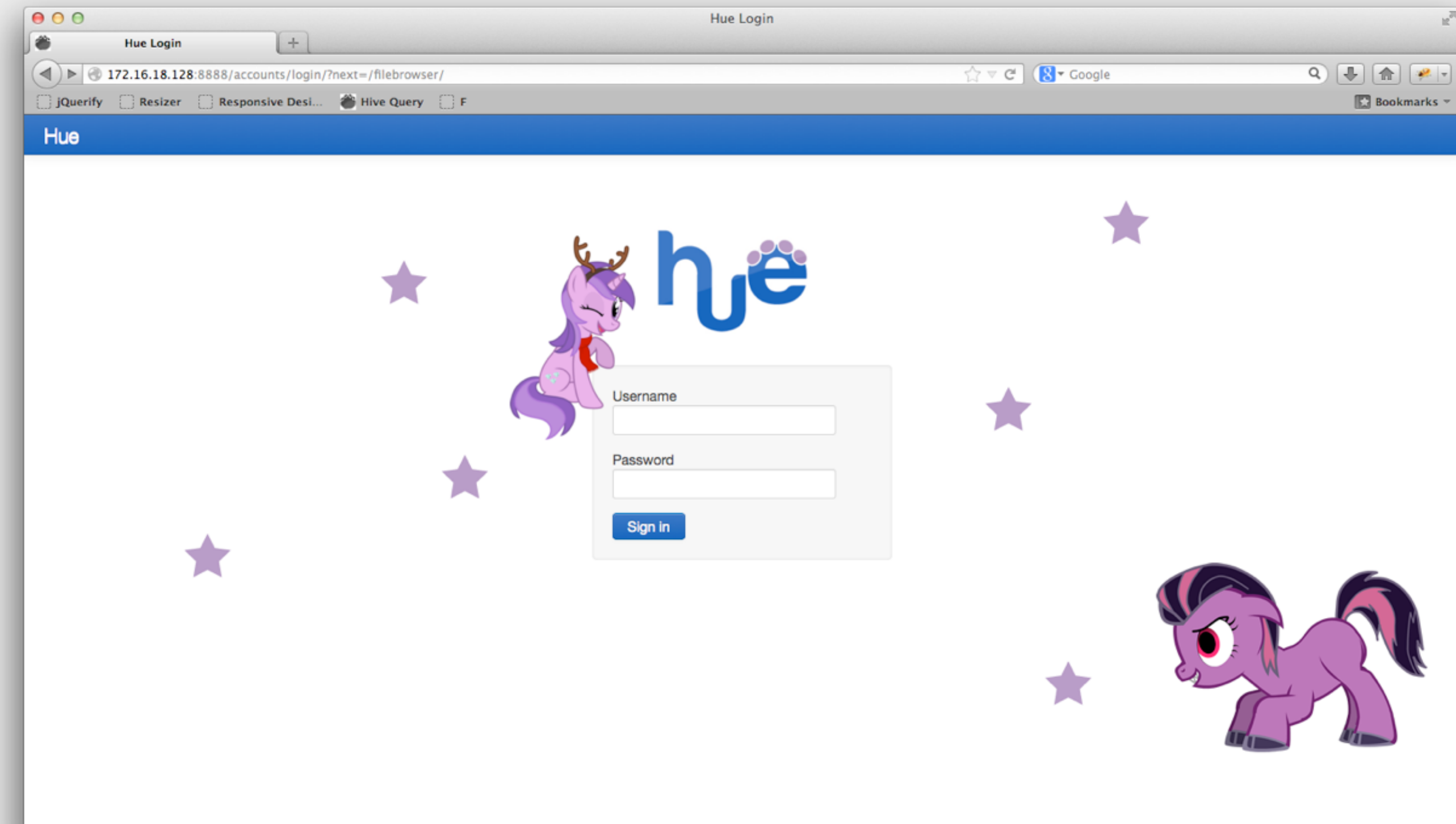
New apps, improved the UX adding new nice functionalities like autocomplete and drag & drop.



HISTORY

HUE 3 ALPHA

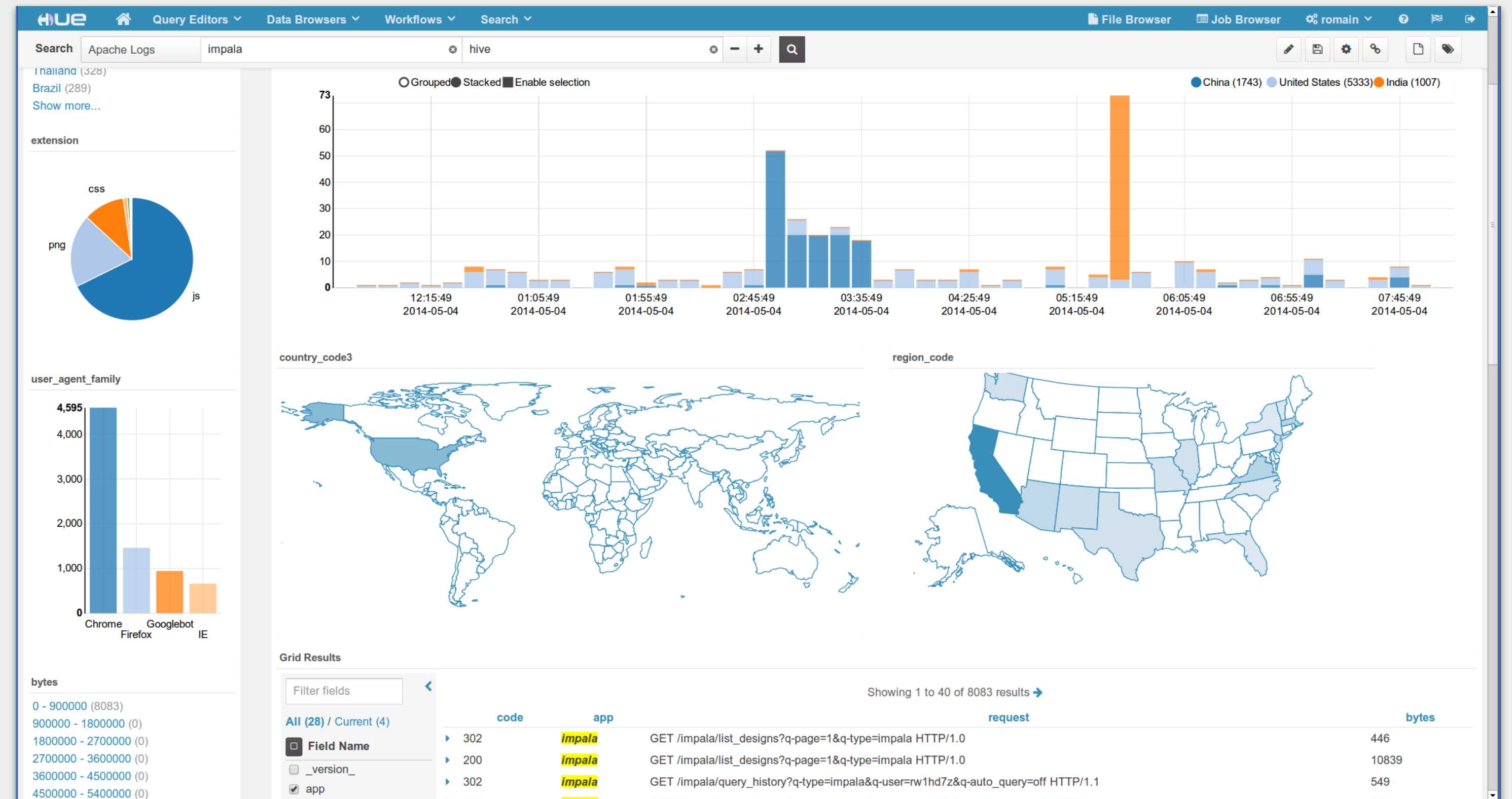
Proposed design, didn't make it.



HISTORY

HUE 3.6+

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



WHICH DISTRIBUTION?

HACKER



GITHUB

Very latest

ADVANCED USER



TARBALL

Advanced preview

NORMAL USER

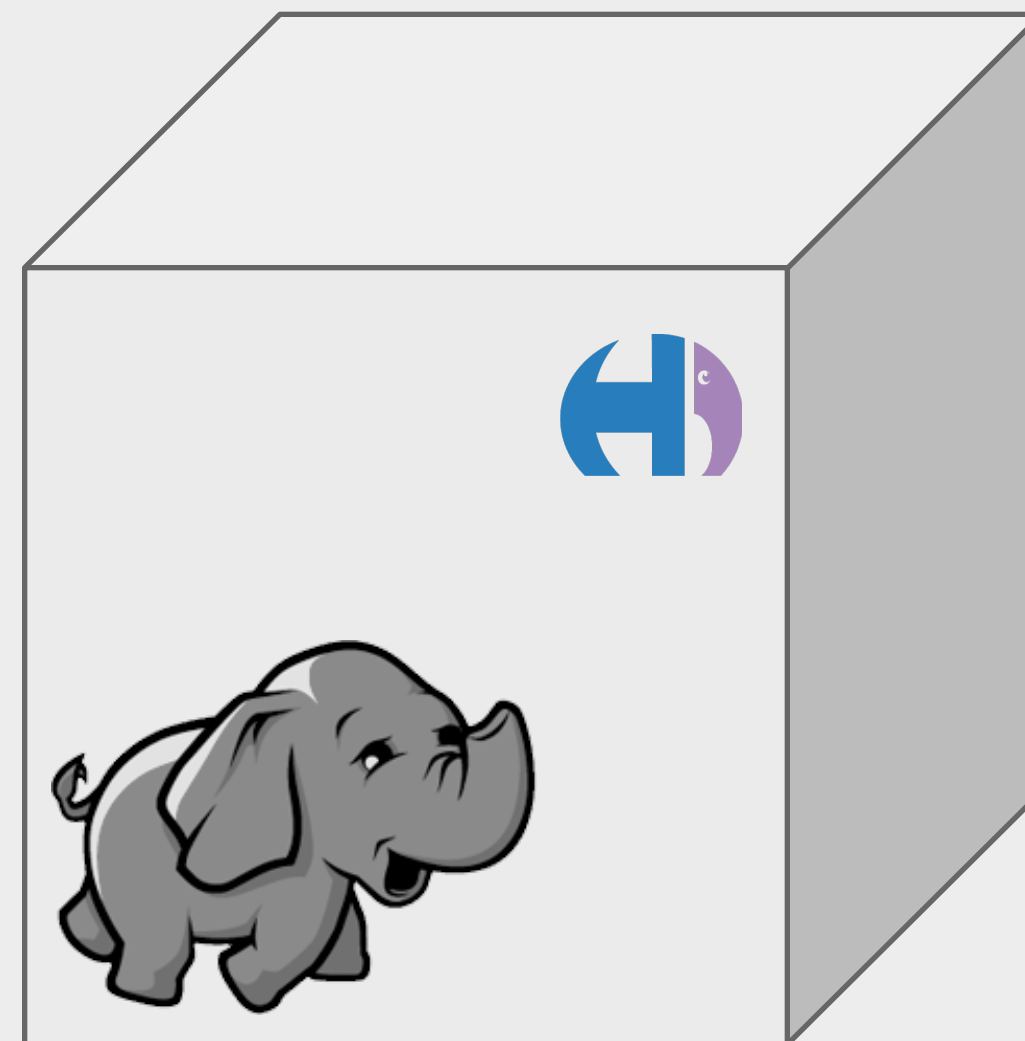


CDH / CM

The most stable and cross
component checked

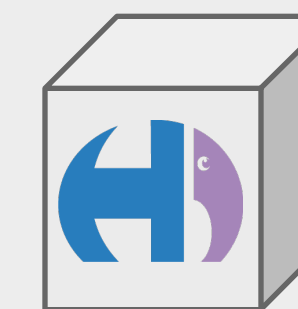
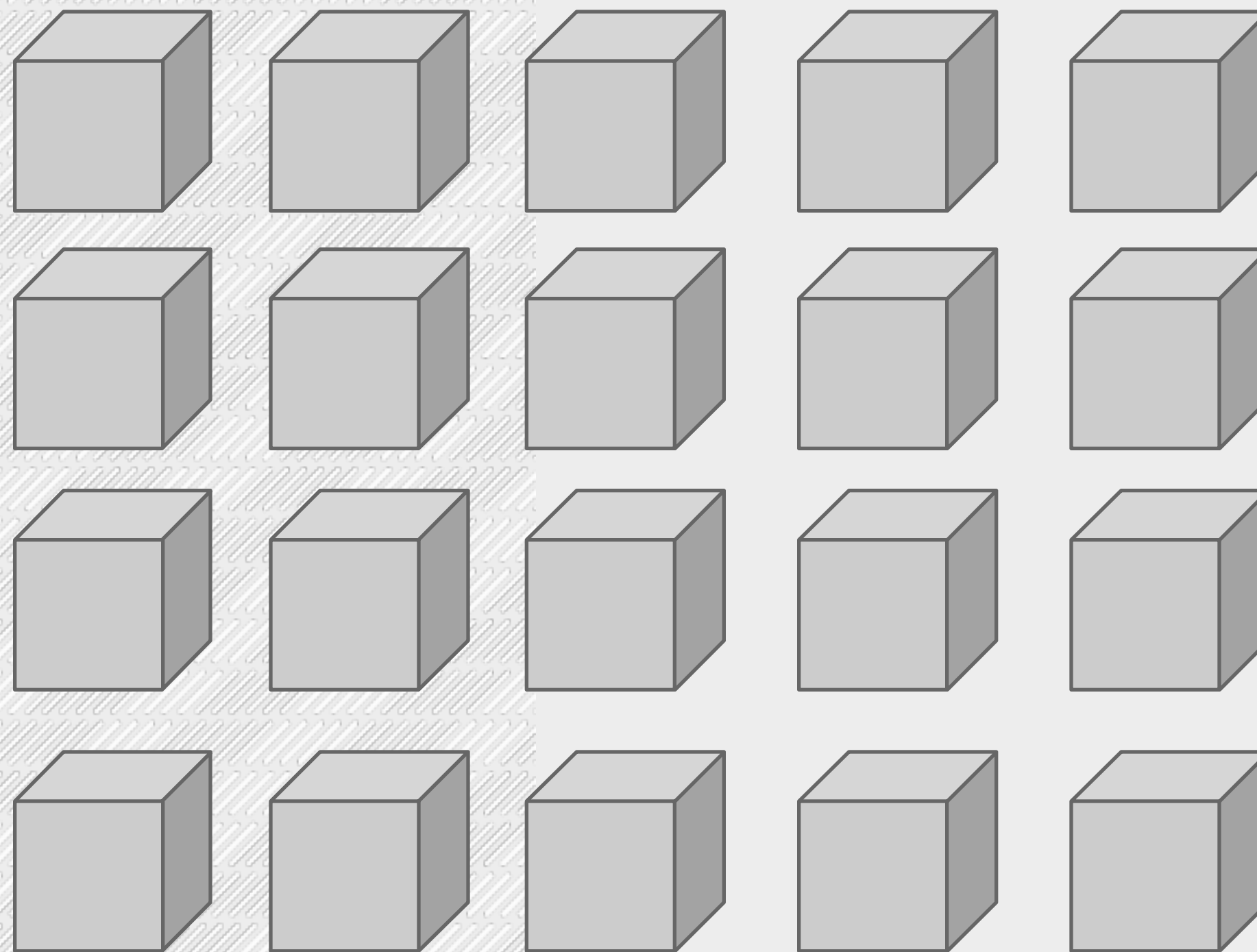
WHERE TO PUT HUE?

IN ONE MACHINE



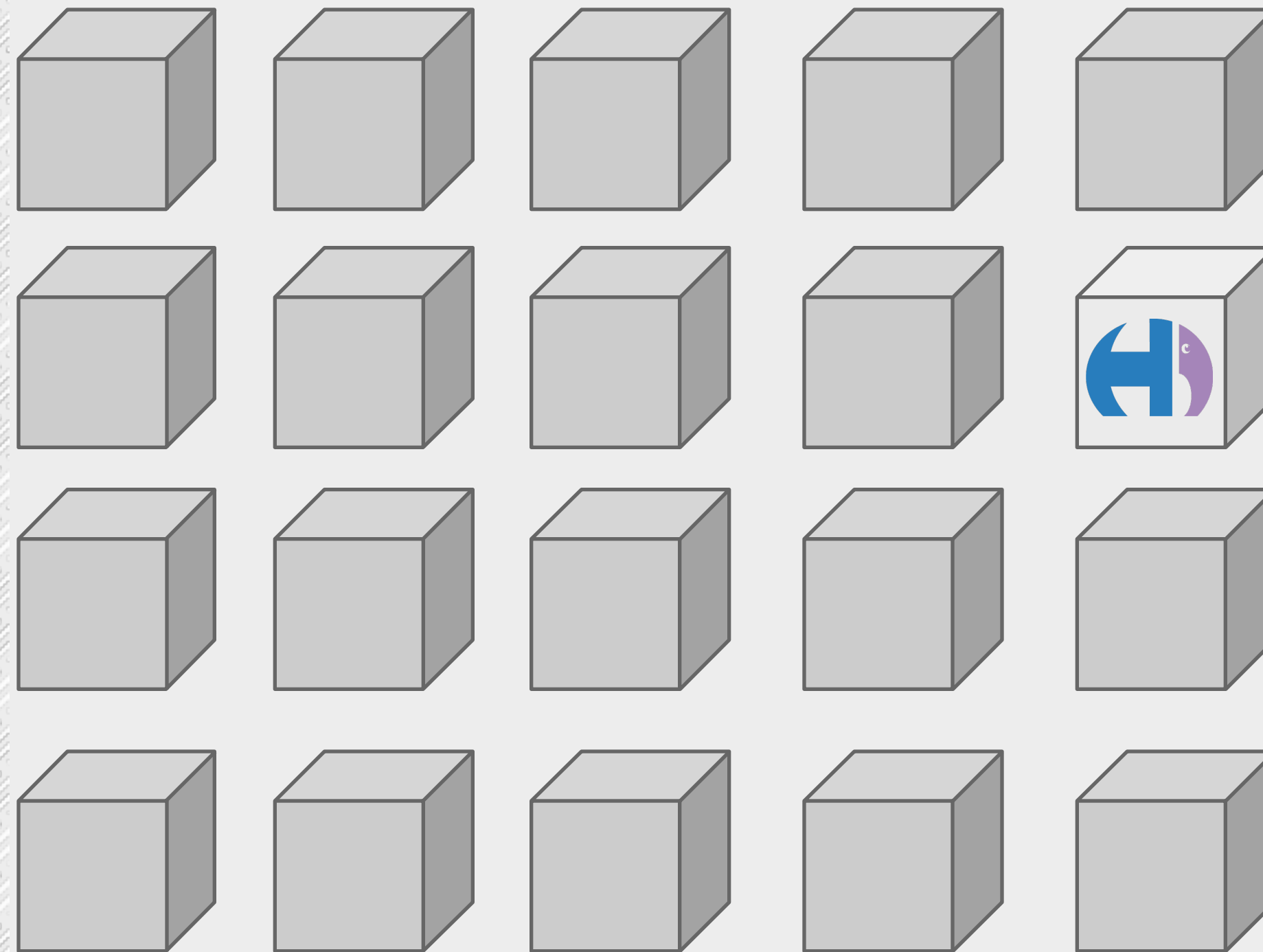
WHERE TO PUT HUE?

OUTSIDE THE CLUSTER



WHERE TO PUT HUE?

INSIDE THE CLUSTER



WHAT DO YOU NEED?

SERVER

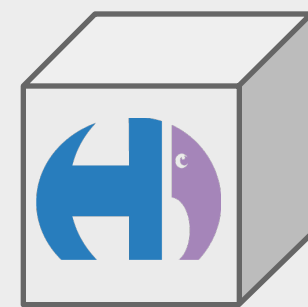
Python ~~2.4~~ 2.6

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.

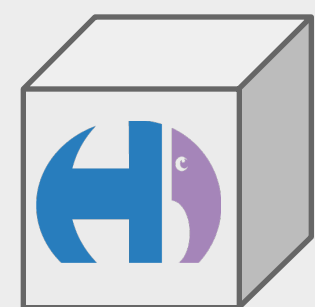
HOW DOES THE HUE SERVICE LOOK LIKE?

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

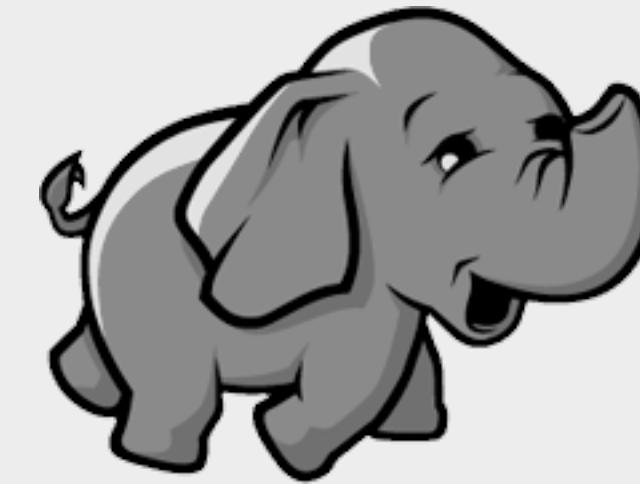
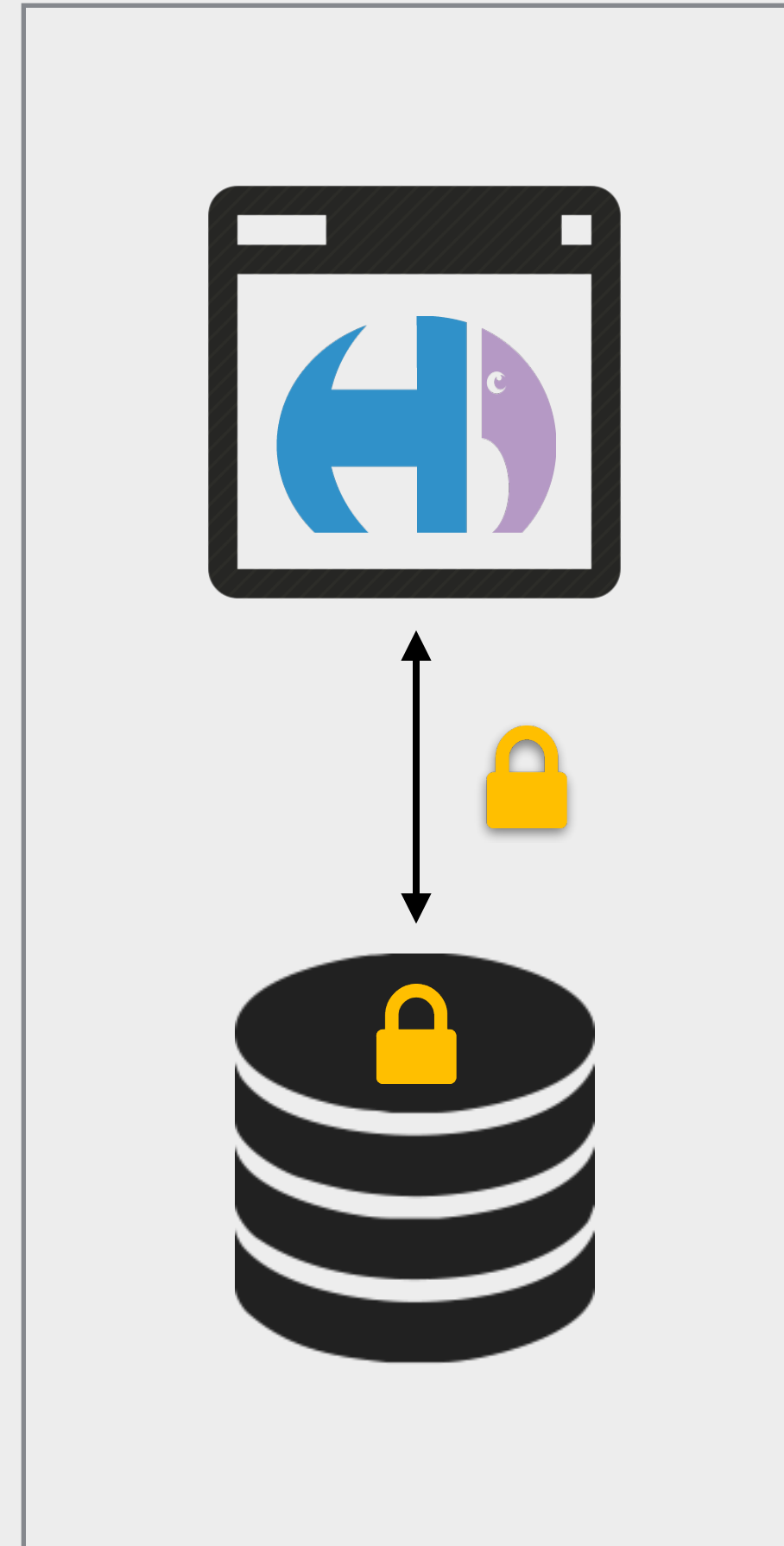
SIMPLE

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

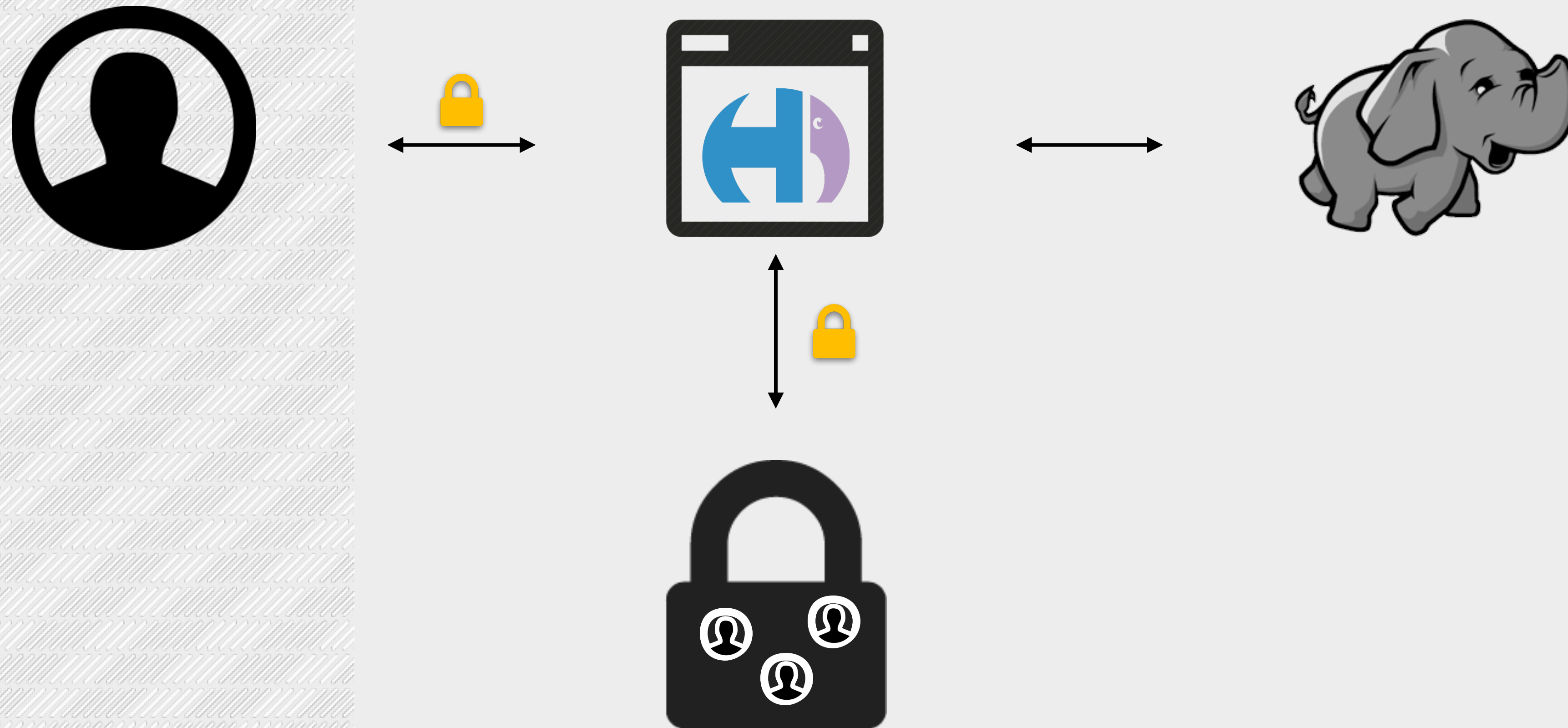
ENTERPRISE

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

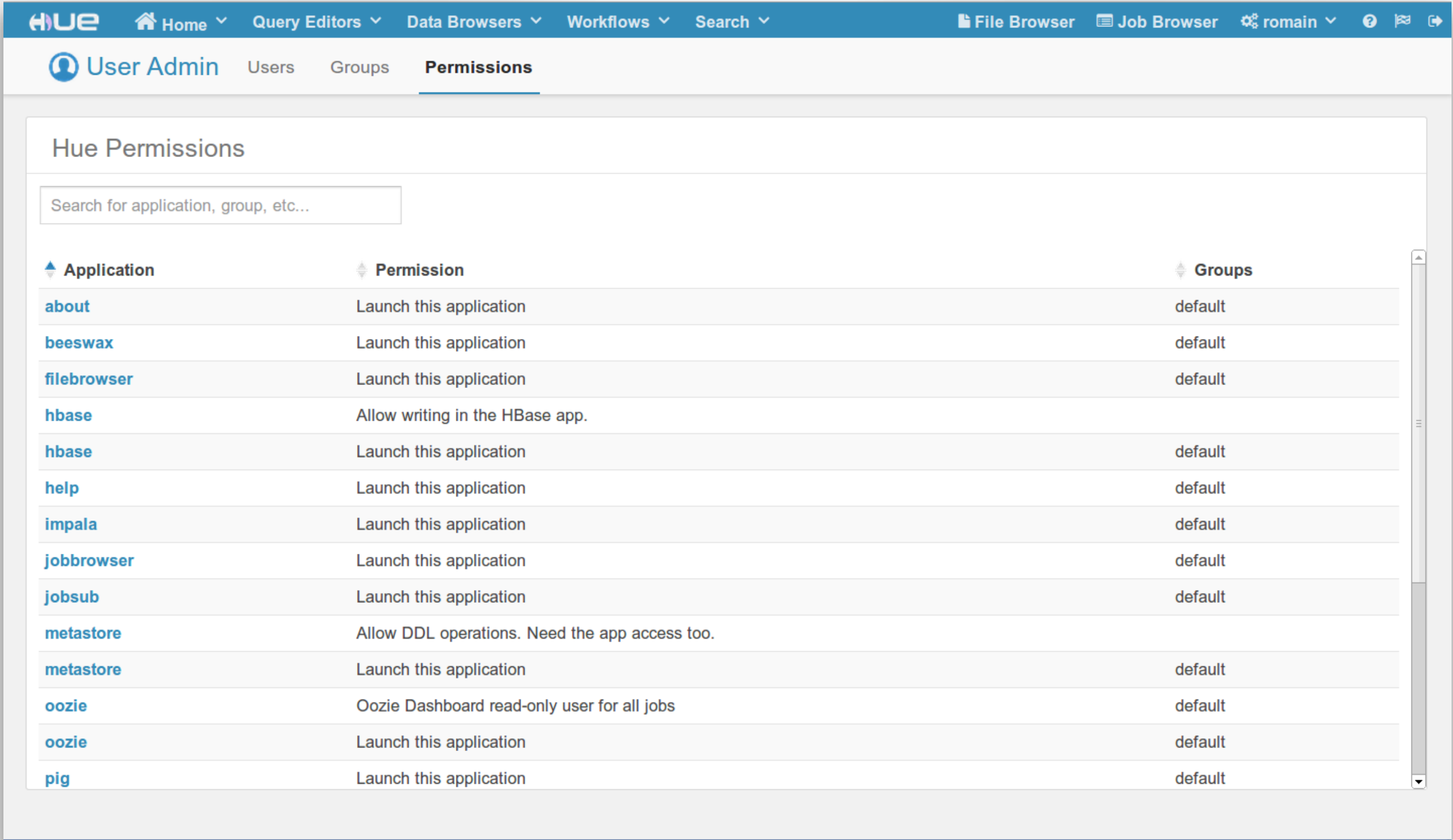
Regular user + permissions

CONFIGURE APPS AND PERMISSIONS

LIST OF GROUPS AND PERMISSIONS

A permission can:

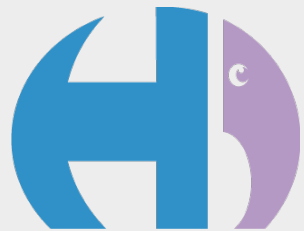
- allow access to one app (e.g. Hive Editor)
- modify data from the app (e.g. drop Hive Tables or edit cells in HBase Browser)



The screenshot shows the Hue User Admin interface with the 'Permissions' tab selected. 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

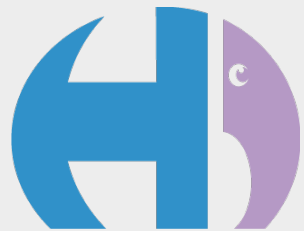
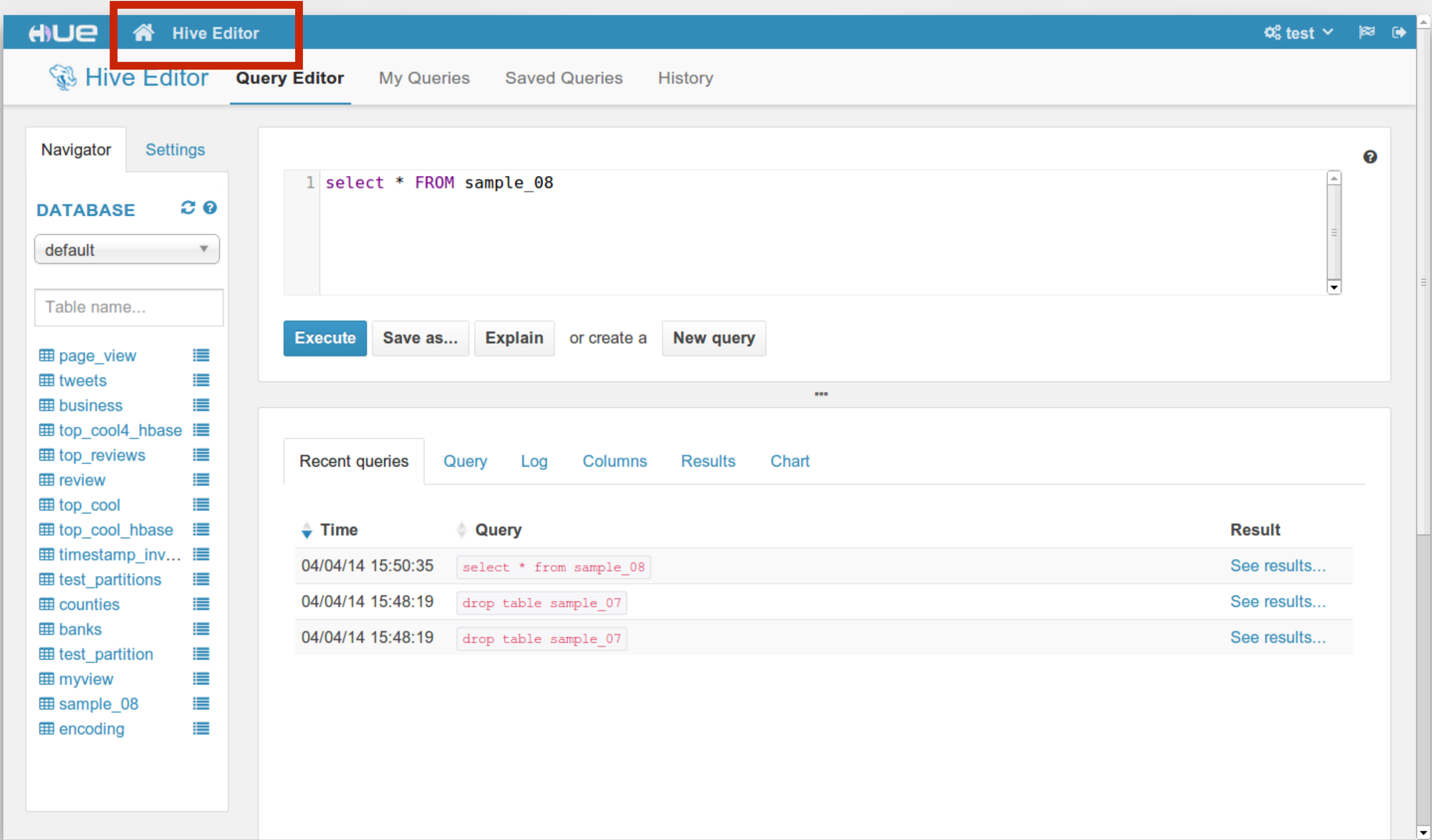
A list of permissions



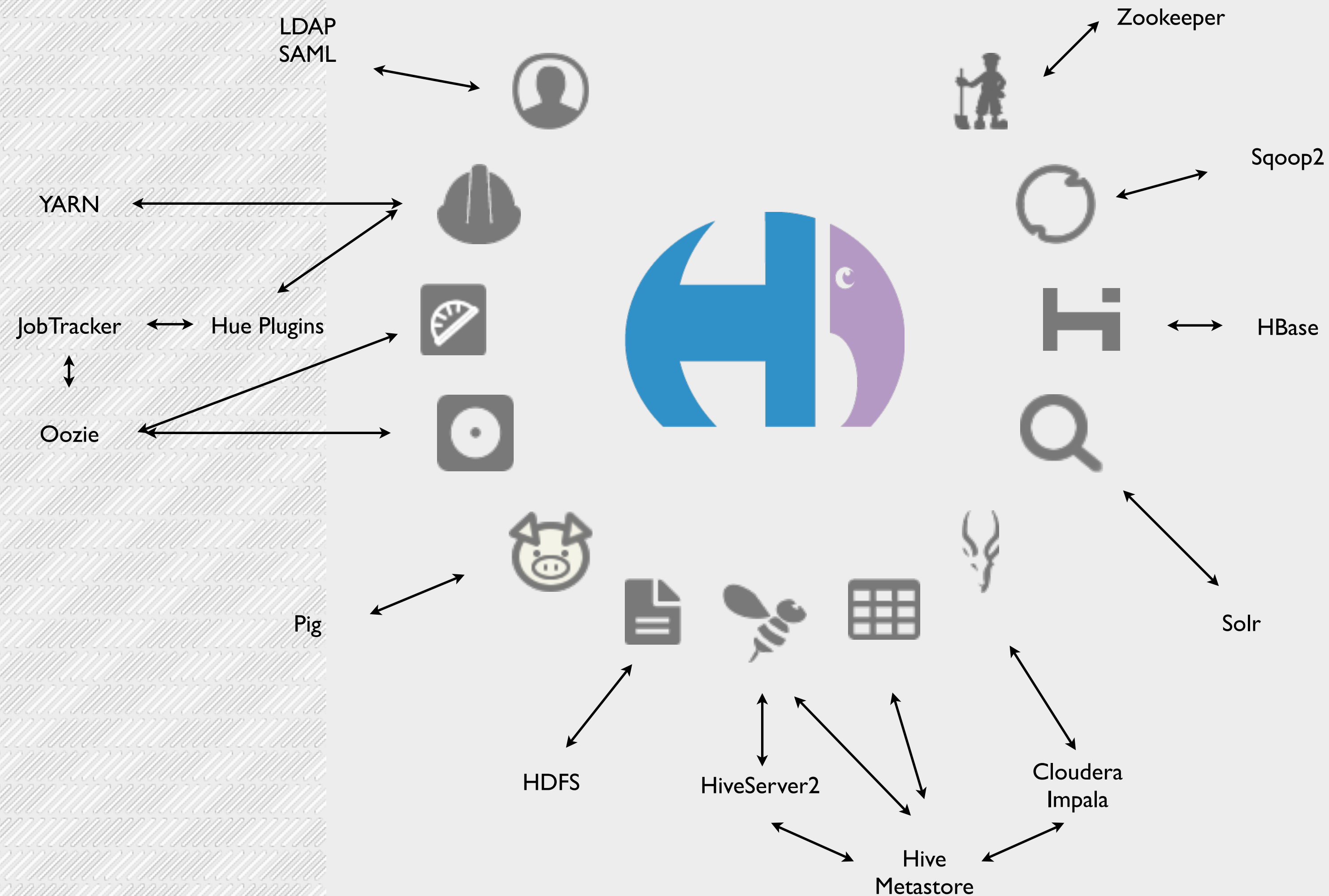
CONFIGURE APPS AND PERMISSIONS

PERMISSIONS IN ACTION

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

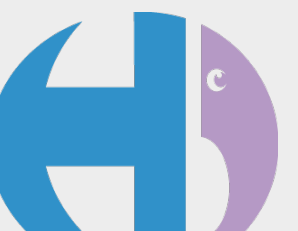
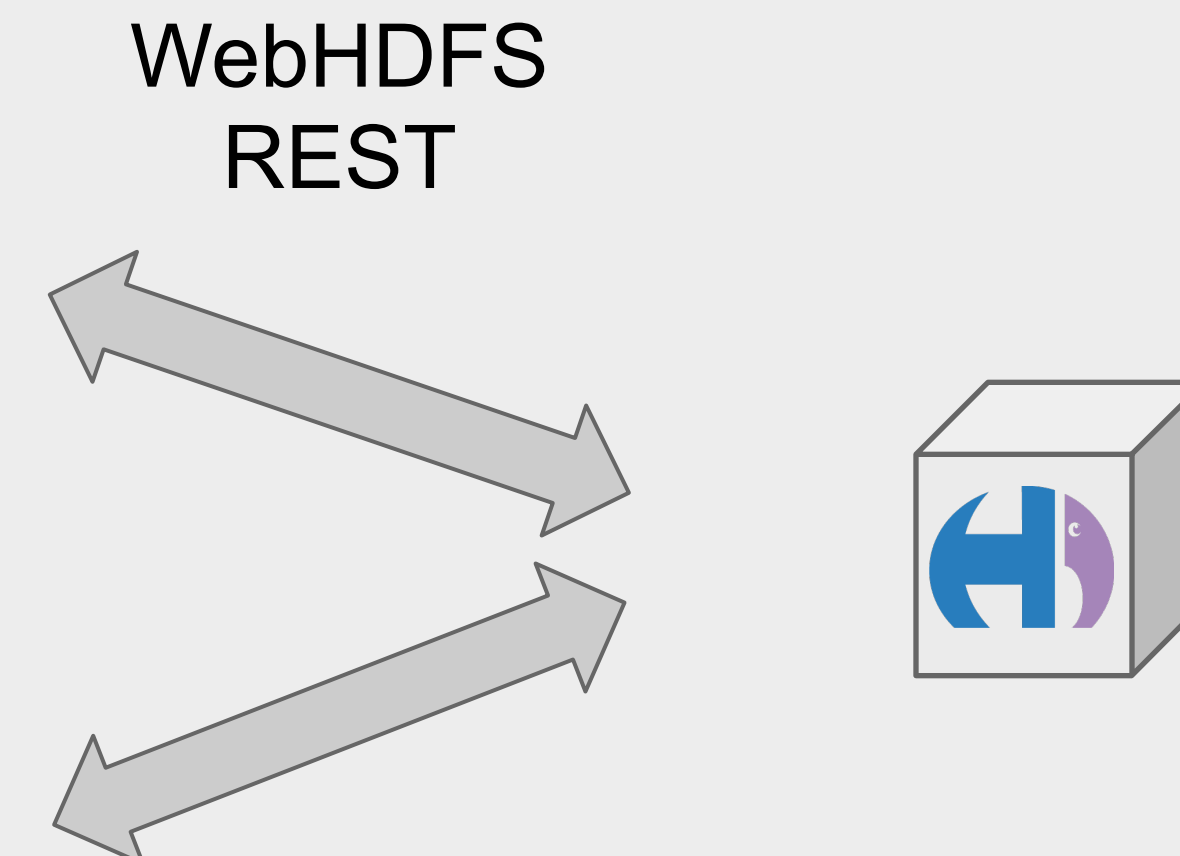
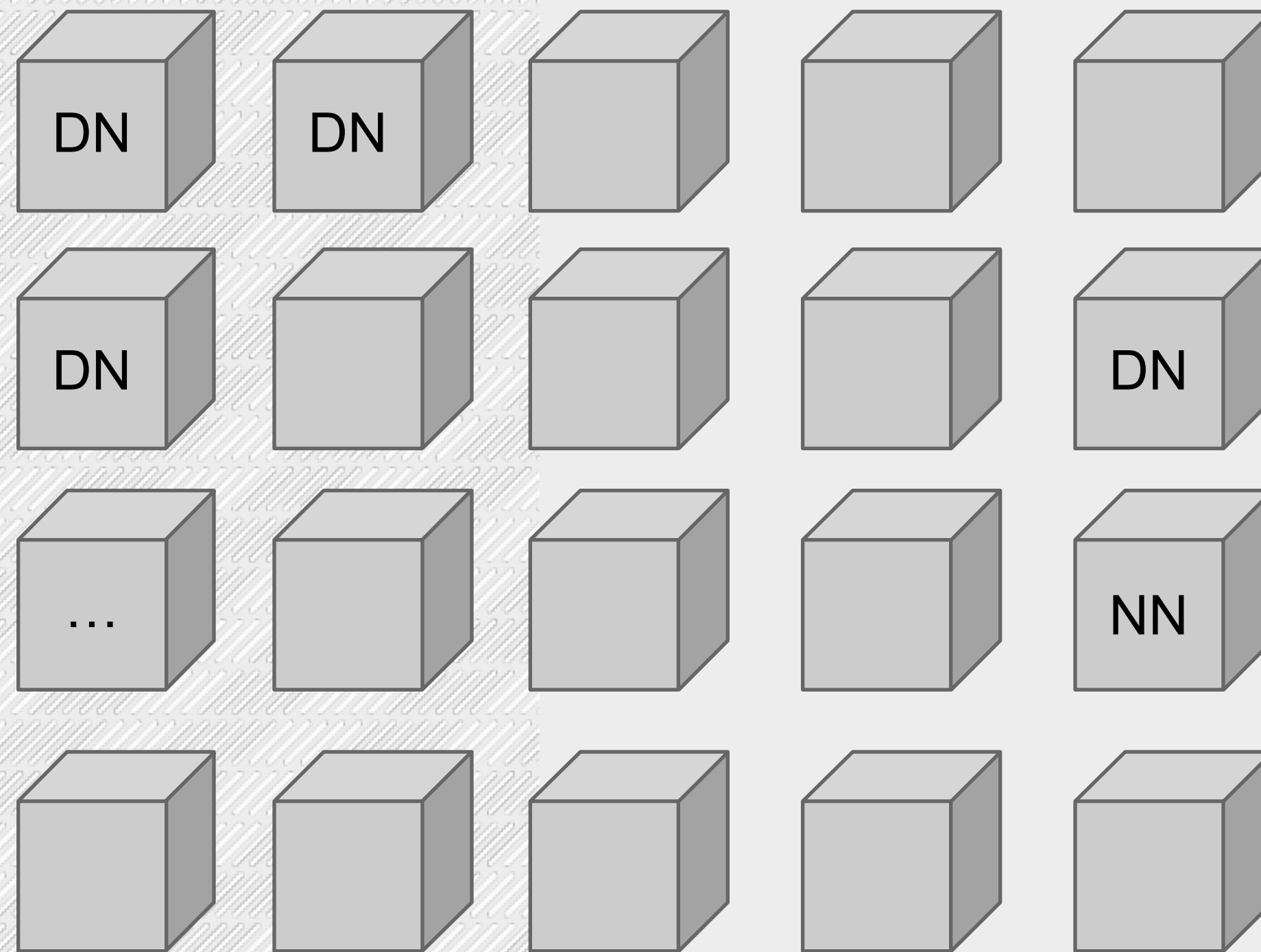


HOW HUE INTERACTS WITH HADOOP



RCP CALLS TO ALL THE HADOOP COMPONENTS

HDFS EXAMPLE



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

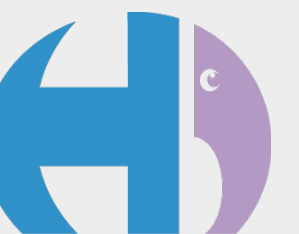
SSL WITH HIVESERVER2

SSL DB

SENTRY

KERBEROS

[READ MORE ...](#)



HIGH AVAILABILITY

HOW

2 Hue instances

HA proxy

Multi DB

Performances: like a website,
mostly RPC calls

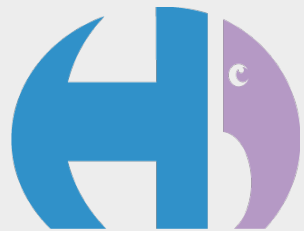
All Services

Add Cluster

Cluster 1 - CDH4

Actions

Name	Status	Role Counts	
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



FULL SUITE OF APPS

Query Editors ▾ **Data Browser**

-  Hive
-  Impala
-  DB Query
-  Pig
-  Job Designer

Data Browsers ▾ **Workflows**

-  Metastore Tables
-  HBase
-  Sqoop Transfer
-  ZooKeeper

Workflows ▾ **Search** ▾ **Security** ▾

-  Dashboards ▸
-  Editors ▸

-  Workflows
-  Coordinators
-  Bundles

o filters applied.

Security ▾

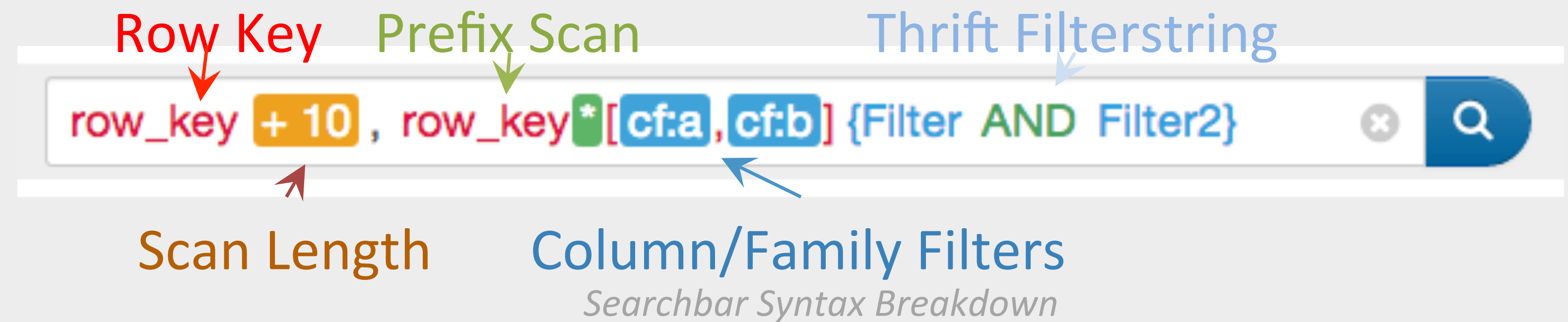
-  Sentry Tables
-  File ACLs

File Browser **Job Browser**  2hx66je ▾

HBASE BROWSER

WHAT

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



SQL

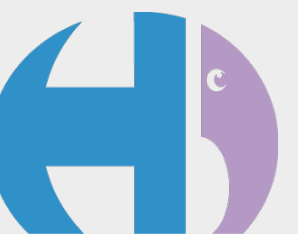
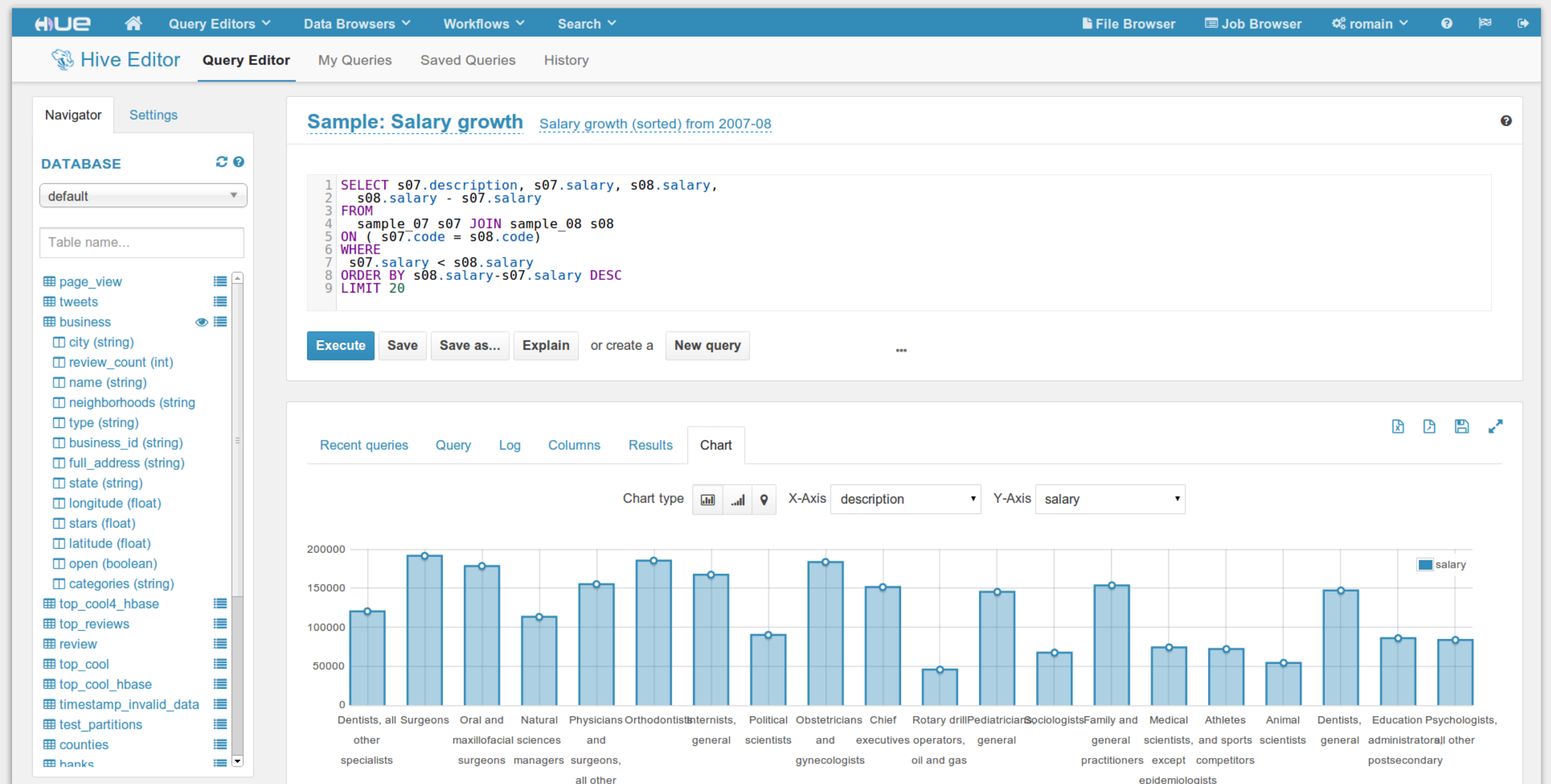
WHAT

Impala, Hive integration, Spark

Interactive SQL editor

Integration with MapReduce,
Metastore, HDFS

cloudera®



SENTRY APP

The screenshot displays the Apache Sentry web interface within the Hue application. The top navigation bar includes links for Query Editors, Data Browsers, Workflows, Search, and Security. The main interface is divided into three sections:

- Left Sidebar:** Contains navigation links for 'PRIVILEGES', 'Browse', 'Roles', and 'GROUPS'.
- Center Panel:** Titled 'Database and Tables privileges', it shows a hierarchical tree of database objects. The selected object is 'default.sample_07'. Below it, a list of roles is shown, including 'server1' and 'default'. The 'default' role is expanded, showing 'sample_07' and 'sample_08'. An 'Impersonate the user' dropdown is set to 'hive'.
- Right Panel:** Titled 'Privileges', it shows a list of roles. The 'read_only' role is selected. Below it, a table shows the privileges for 'read_only' on 'default.sample_07'. The table includes columns for 'Privilege' and 'Action'. The 'SELECT' action is highlighted, and a dropdown menu is open showing options: 'SELECT', 'INSERT', and 'ALL'.

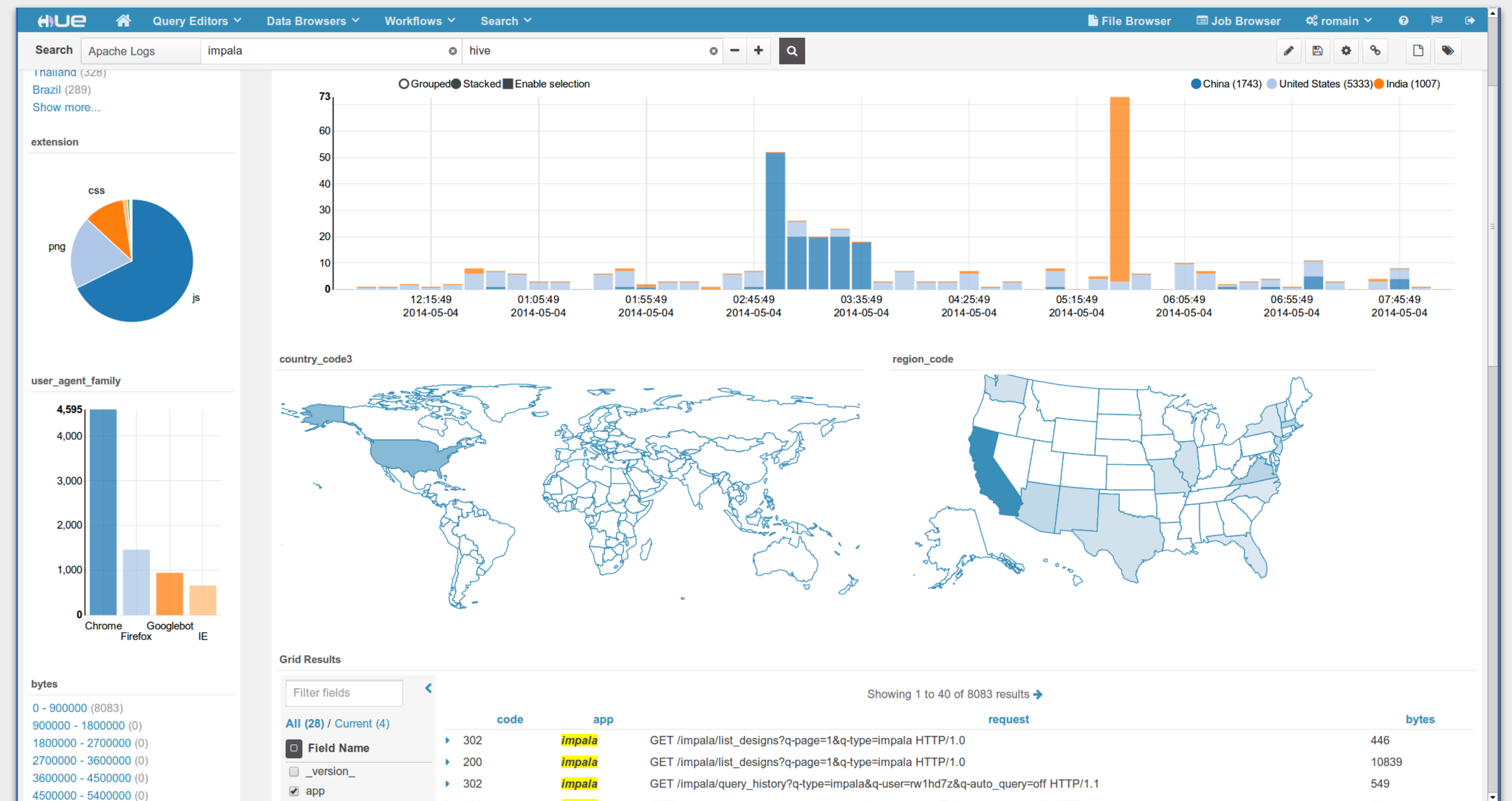
SEARCH

WHAT

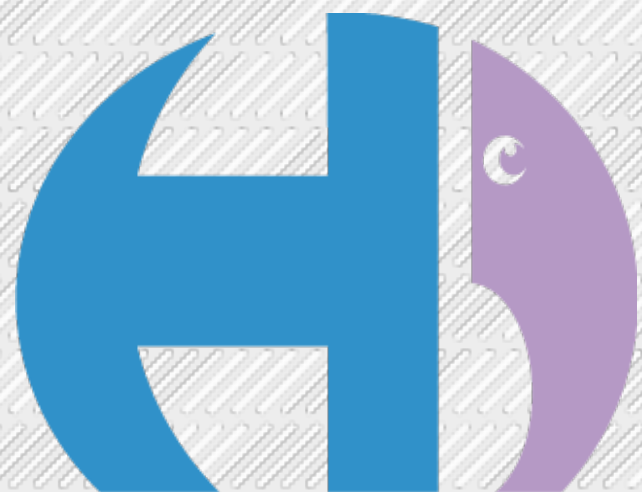
Solr & Cloud integration

Custom interactive dashboards

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



JUST A VIEW ON TOP OF SOLR API

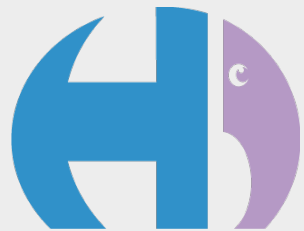


REST



[search]

```
# URL of the Solr Server
## solr_url=http://localhost:8983/solr/
```



HISTORY

V1 USER

The screenshot shows the Cloudera Hue web interface. At the top, there's a navigation bar with various icons and a user profile dropdown for 'hue'. Below this is a search bar with the text 'britney spears' and a search button. To the right of the search bar are links for 'Customize this collection' and 'Collection manager'.

The main content area displays search results for 'britney spears' in 'collection3'. It shows 'Showing 1 - 15 of 43 results' and a 'Sort by' dropdown. The results are a list of tweets:

- Zach (Nickel-Liss)**: Maybe **Britney** can help get me on the show when I'm old enough. Haaaaa That would be extremely
15 hours ago
- connie salinas**: "@britneyspears **Britney Spears** Takes The Stage At Her Niece's Dance Recital!!" http://t.co
15 hours ago
- CeCe Frey**: RT @CelebrityVS_: **Britney Spears** VS Lady Gaga RT: For **Britney** FAV: For Gaga http://t.co
15 hours ago
- ashley. ∞**: RT @FemaleStruggles: I'm never going to forgive Jamie Lynn **Spears** for getting pregnant and ruining
15 hours ago
- ✨Ohh.Tyy.Sonn✨**: I really hate when Brittany **spears** starts playing on Pandora
15 hours ago
- A Faxineira Sasha**: RT @juliana_drg: é fodaeu miro na **britney** e acerto na joelma....nao ta facil nao!
15 hours ago
- Ayuni.**: RT @flopney: Miley and Selena are like the **Britney** and Christina of this decade. Demi is Pink.
15 hours ago

The left sidebar contains filters for the search results:

- TWEETS**: 0 (7), 1000 (2), 2000 (3), 3000 (1), 4000 (1), 5000 (2), 6000 (1), 7000 (3), 8000 (1), 9000 (1)
- LOCATION**: barcelona (2), rio de janeiro (2), a piece land of a\$\$ (1), bandung - indonesia (1), carlow (1), catania, italy (1), croatia (1), finding myself♥ (1), home (1)
- WHEN**: 8 hours ago (2), 9 hours ago (7), 10 hours ago (5), 11 hours ago (2)

HISTORY

V1 ADMIN

Template Editor : collection3

Visual editor Preview Source Advanced

COLLECTION

- Properties
- Search it

TEMPLATE

- Snippet
- Facets
- Sorting
- Highlighting

Visual editor

[[user_name]]

[[text]]

[[#fromnow]][[created_at]][[/fromnow]]

Available Fields

version +

Available Functions

[[#date]] +

Formats date or timestamp in DD-MM-YYYY

Facets

Facets provide an intuitive way to filter the results. Different types of facets can be added on the following steps.

Step 1: General Step 2: Field Facets Step 3: Range Facets Step 4: Date Facets Step 5: Facets Order

There are currently no Date Facets defined.

Field created_at Label when

Range from 10 HOURS before today until 0 before today. Goes up by increments of 1 DAY

YEARS

MONTHS

DAYS

DATE

HOURS

MINUTES

SECONDS

MILLISECONDS

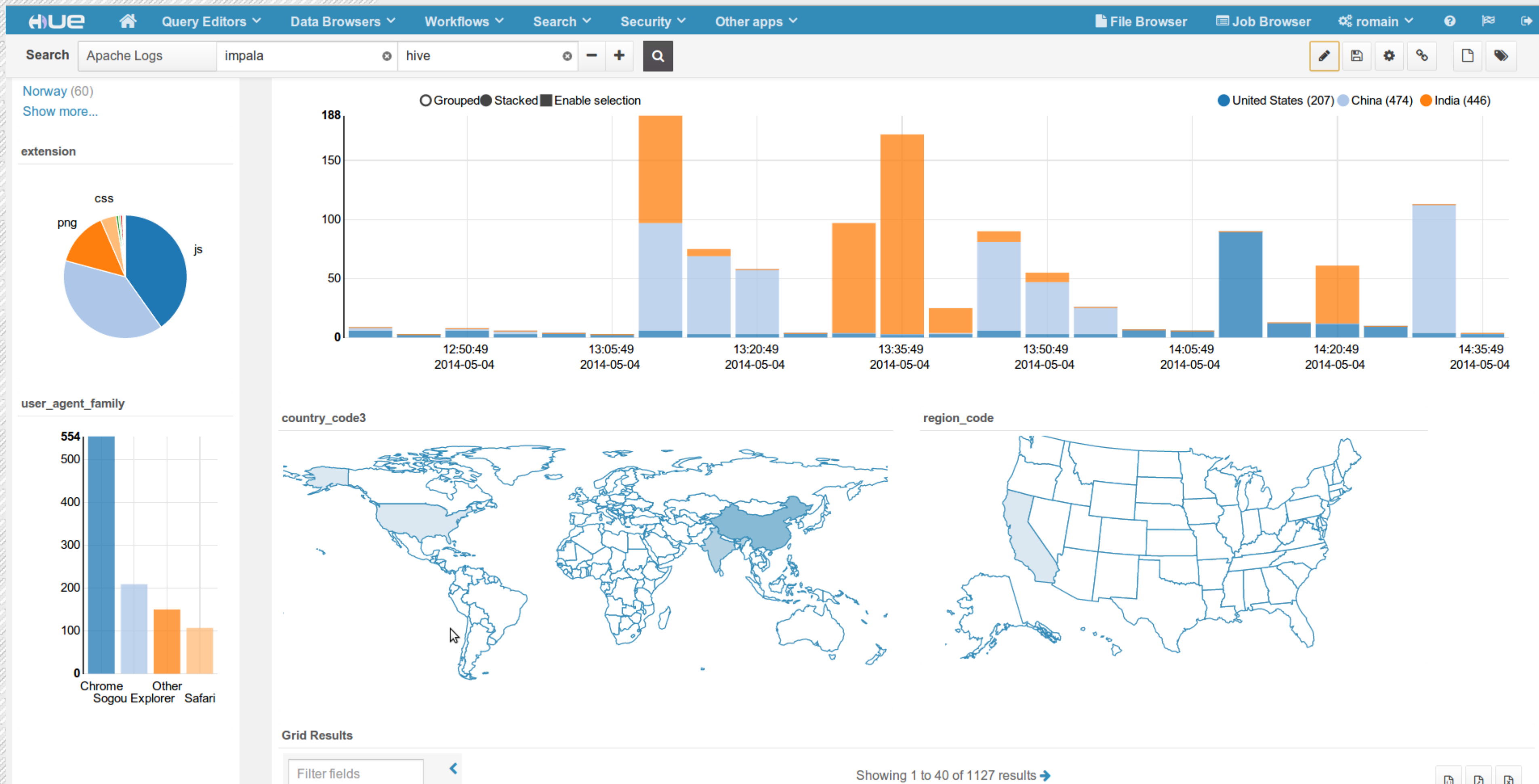
Back Next Save

Visual editor Preview Source Advanced

```
1
2 <div class="row-fluid">
3   <div class="row-fluid">
4     <div class="row-fluid">
5       <div class="span1">
6         
7       </div>
8       <div class="span11">
9         <a href="https://twitter.com/{{user_screen_name}}/status/{{user_name}}>
10          <b>{{user_name}}</b>
11          <br>
12          {{text}}
13          <br>
14          <div class="created">[[#fromnow]][[created_at]][[/fromnow]]</div>
15        </div>
16      </div>
17    </div>
18  </div>
```

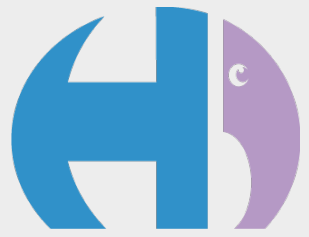
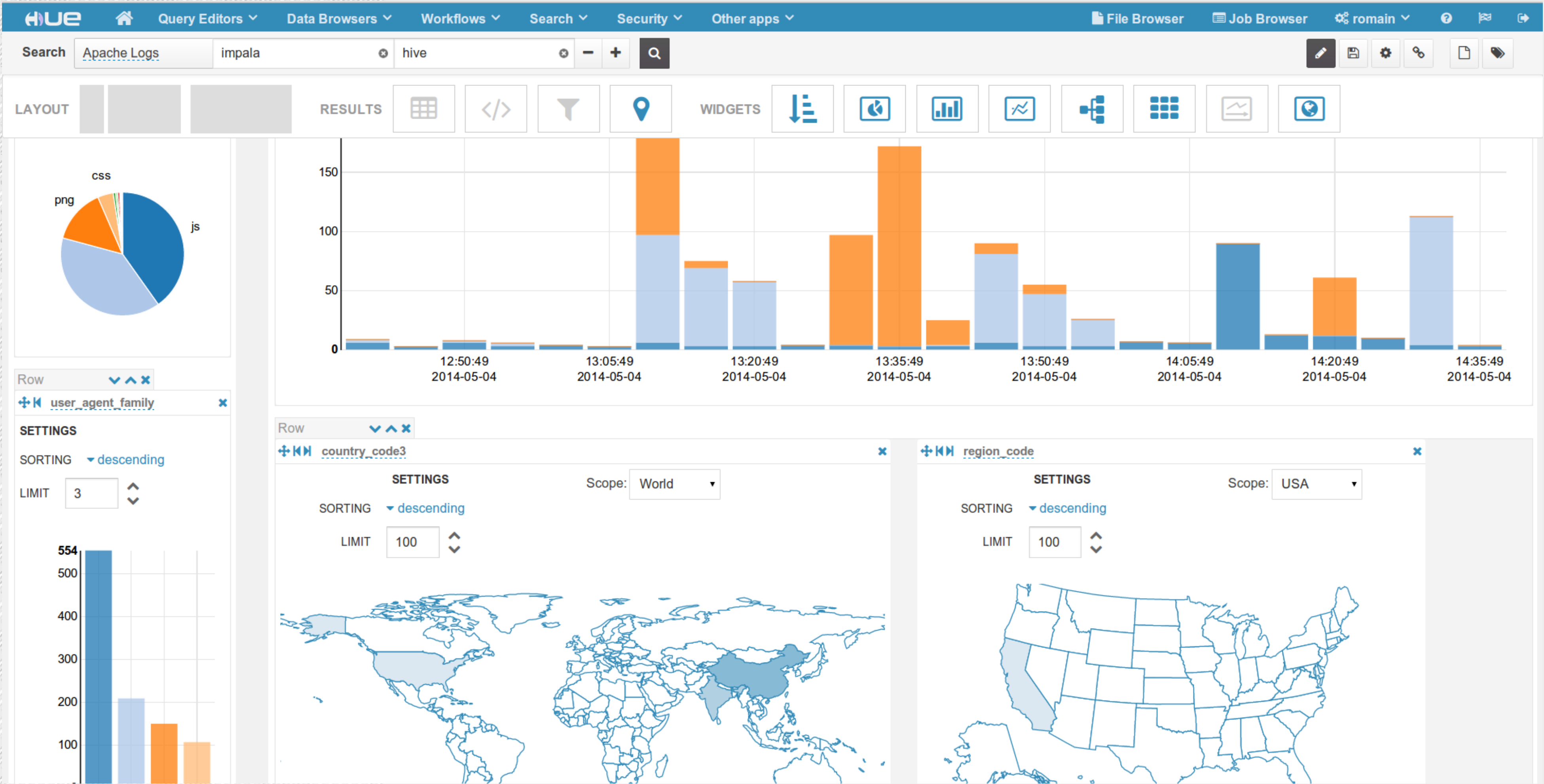
HISTORY

V2 USER



HISTORY

V2 ADMIN

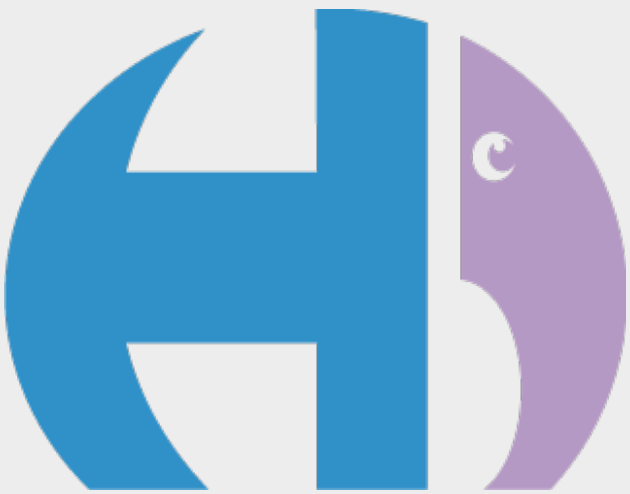


ARCHITECTURE



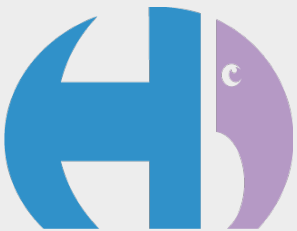
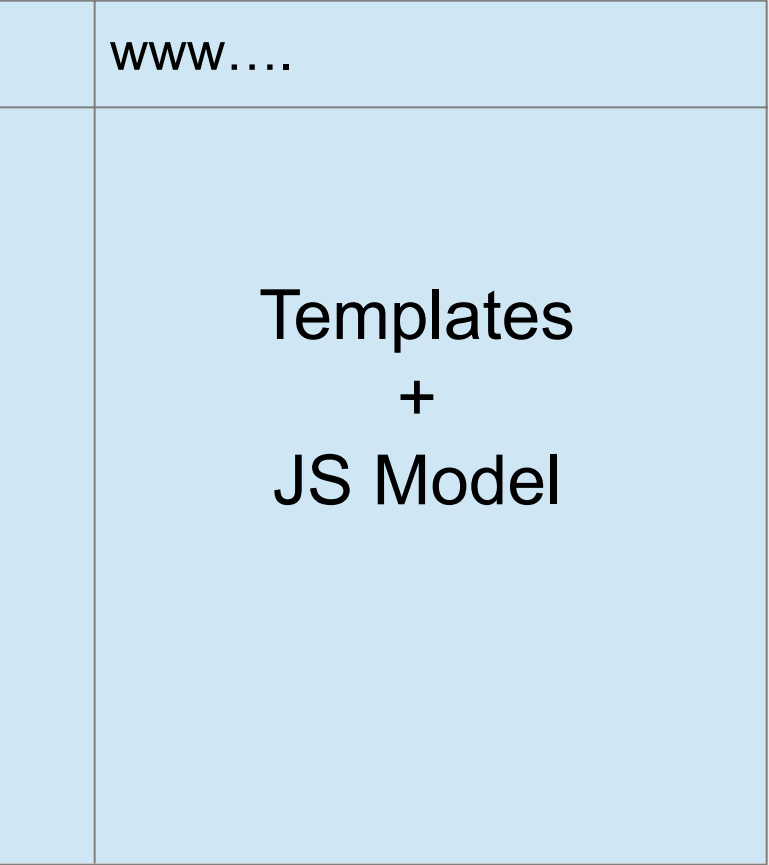
REST

/select
/admin/collections
/get
/luke...



AJAX

/add_widget
/zoom_in
/select_facet
/select_range...



ARCHITECTURE UI FOR FACETS

LAYOUT

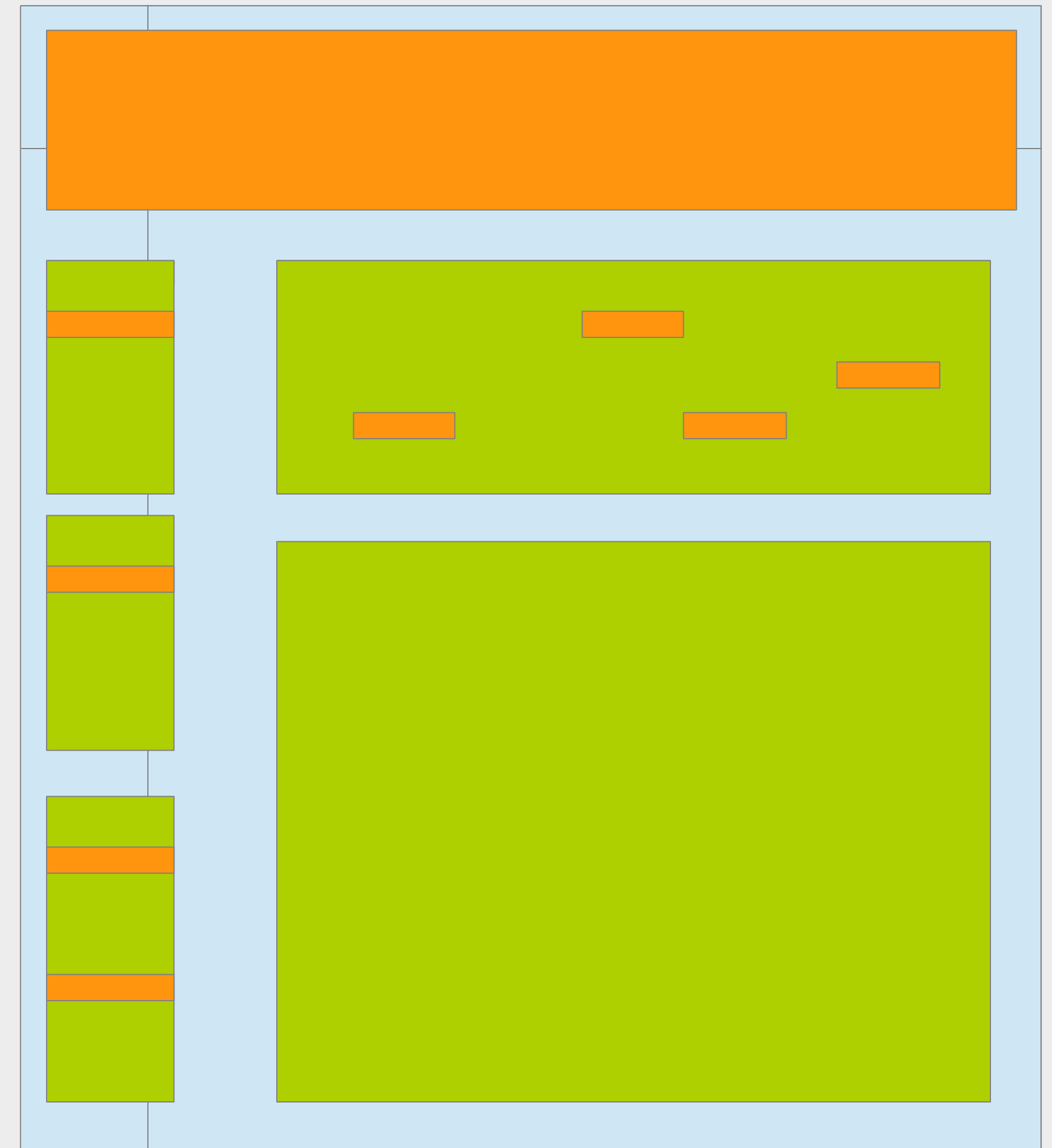
All the 2D positioning (cell ids), visual, drag&drop

COLLECTION

Dashboard, fields, template, widgets (ids)

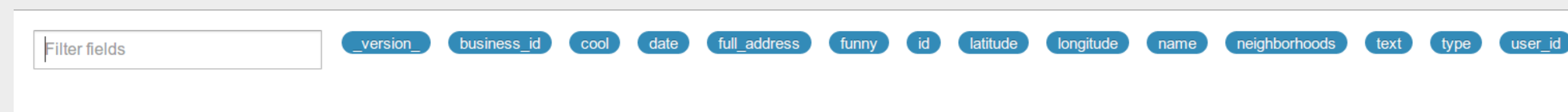
QUERY

Search terms, selected facets (q, fqs)

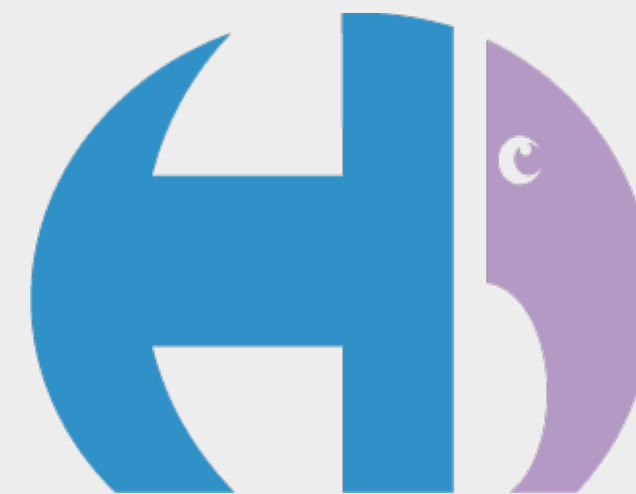


ADDING A WIDGET LIFECYCLE

Load the initial page
Edit mode and Drag&Drop

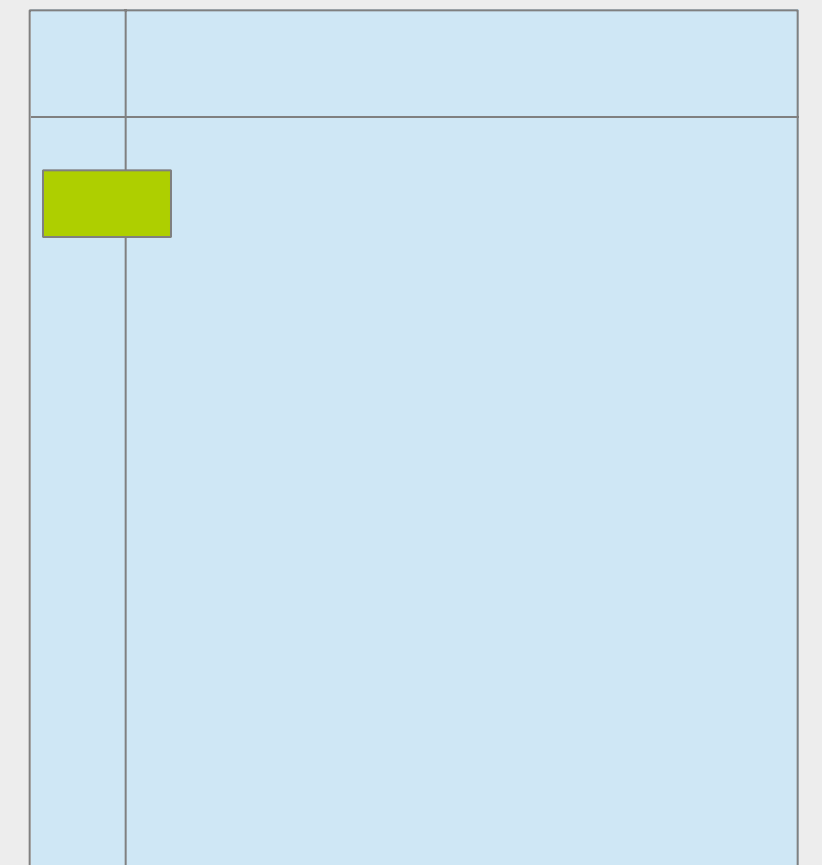


REST



AJAX

/get_collection



ADDING A WIDGET LIFECYCLE

Select the field
Guess ranges (number or dates)
Rounding (number or dates)

A screenshot of the Solr facet settings for the 'bytes' field. The settings are as follows:

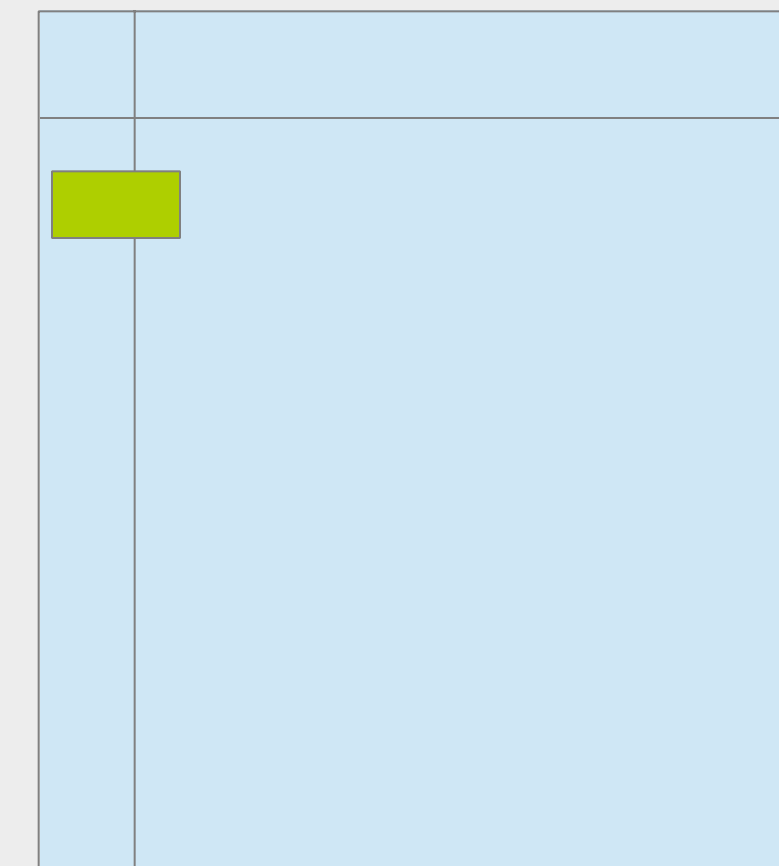
- SETTINGS**
- TYPE**: range
- SORTING**: descending
- Range**: 0 to 9000000
- Facet Range**: 0 to 9045352



REST
/solr/select?stats=true



AJAX
/new_facet



ADDING A WIDGET

LIFECYCLE

Query part 1

facet.range={!ex=bytes}bytes&f.bytes.facet.range.start=0&f.bytes.facet.range.end=9000000&f.bvtes.facet.range.gap=900000&f.bvtes.facet.mincount=0&f.bvtes.facet.limit=10

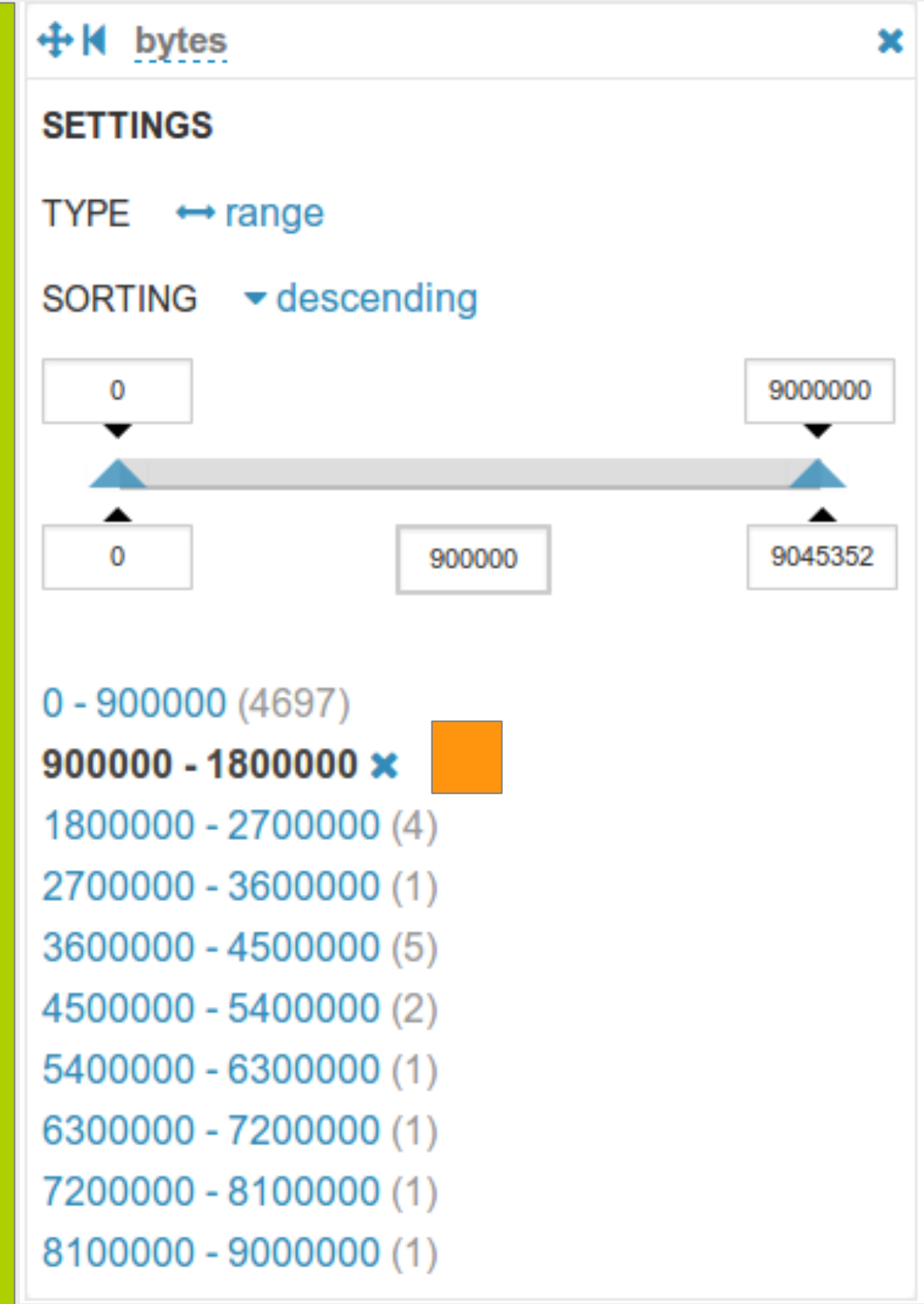
Query Part 2

q=Chrome&fq={!tag=bytes}bytes:[900000+TO+1800000]

```
{
  'facet_counts': {
    'facet_ranges': {
      'bytes': {
        'start': 10000,
        'counts': [
          '900000',
          3423,
          '1800000',
          339,
          ...
        ]
      }
    }
  }
}
```

Augment Solr response

```
{
  ...,
  'normalized_facets': [
    {
      'extraSeries': [
        ],
      'label': 'bytes',
      'field': 'bytes',
      'counts': [
        {
          'from': '900000',
          'to': '1800000',
          'selected': True,
          'value': 3423,
          'field': 'bytes',
          'exclude': False
        }
      ]
    },
    ...
  ]
}
```



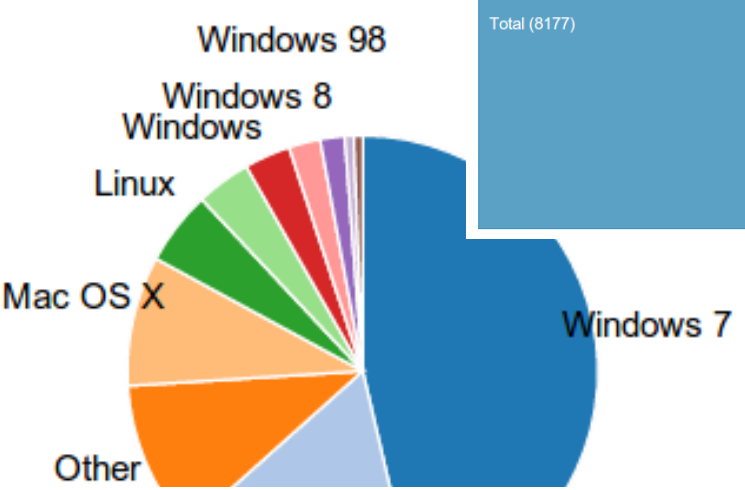
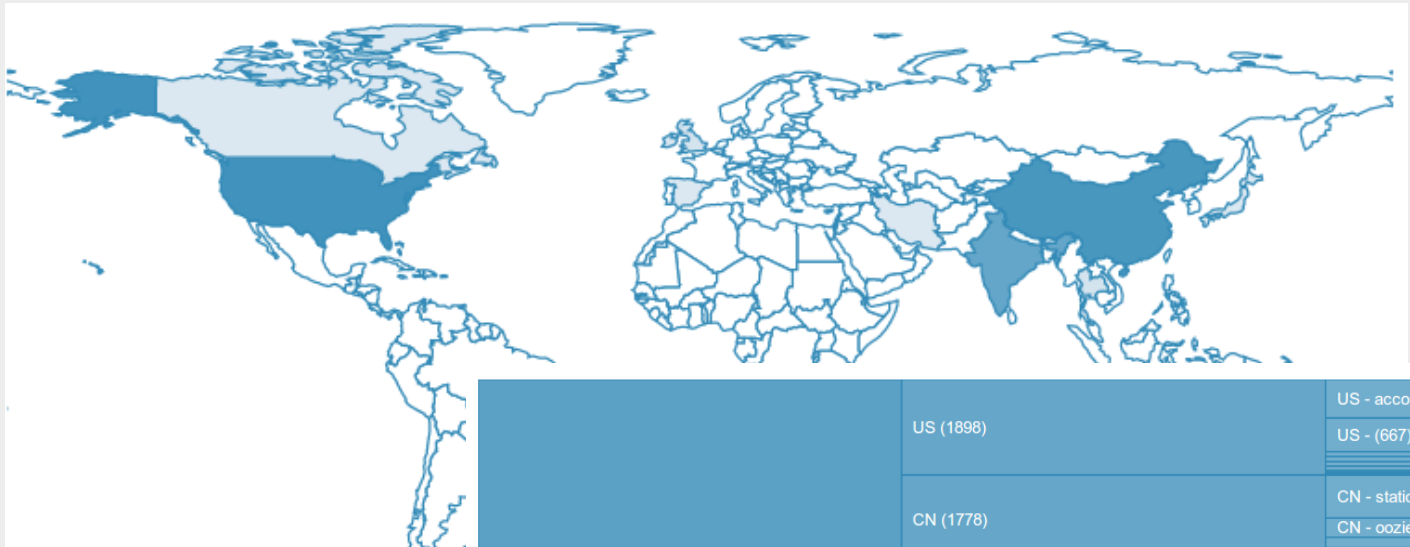
JSON TO WIDGET

```
{
  "field": "rate_code",
  "counts": [
    {
      "count": 97797,
      "exclude": true,
      "selected": false,
      "value": "1",
      "cat": "rate_code"
    }
  ]
} ...

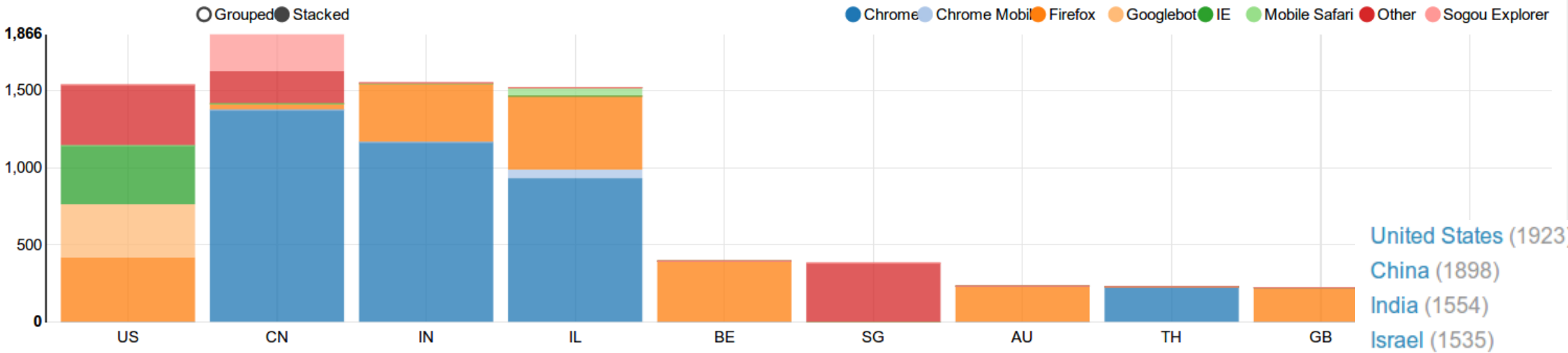
{
  "field": "medallion",
  "counts": [
    {
      "count": 159,
      "exclude": true,
      "selected": false,
      "value": "6CA28FC49A4C49A9A96",
      "cat": "medallion"
    }
  ]
} ...
```

```
{
  "extraSeries": [
    {
      "label": "trip_time_in_secs",
      "field": "trip_time_in_secs",
      "counts": [
        {
          "from": "0",
          "to": "10",
          "selected": false,
          "value": 527,
          "field": "trip_time_in_secs",
          "exclude": true
        }
      ]
    }
  ]
} ...

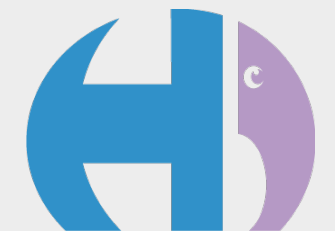
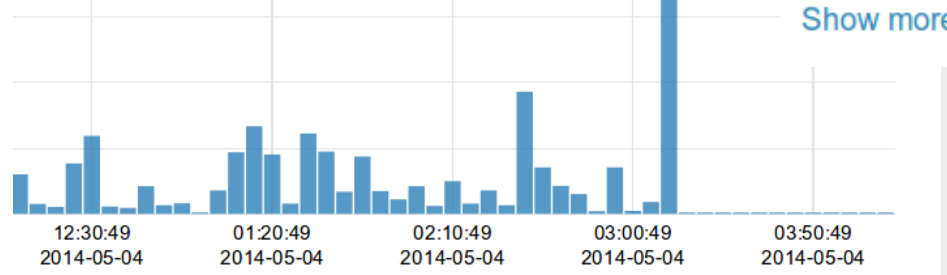
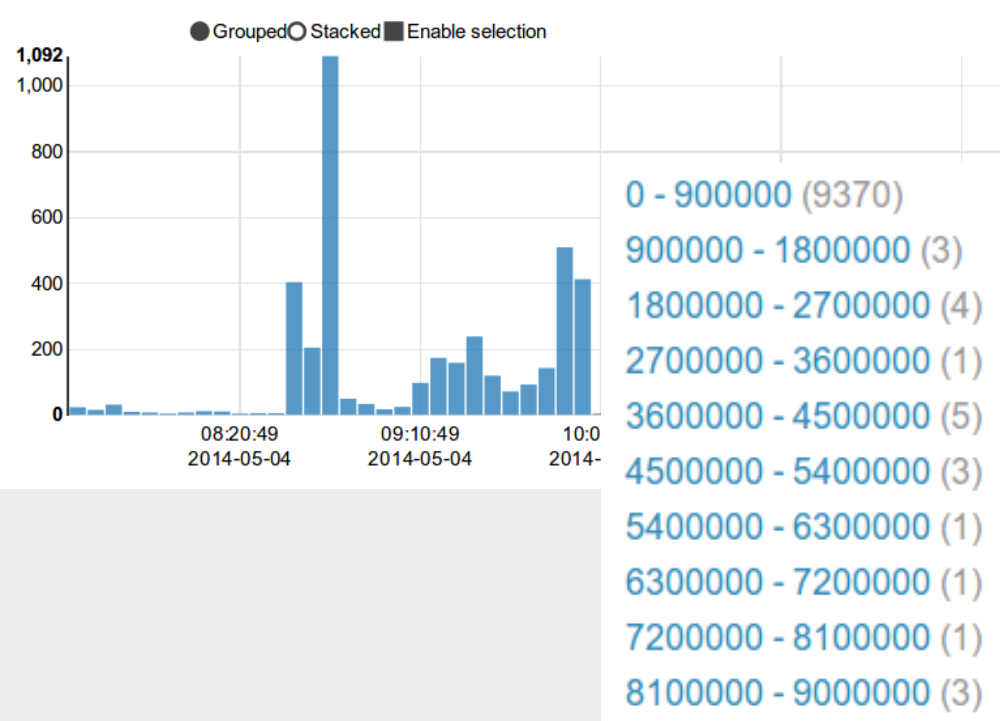
{
  "field": "passenger_count",
  "counts": [
    {
      "count": 74766,
      "exclude": true,
      "selected": false,
      "value": "1",
      "cat": "passenger_count"
    }
  ]
} ...
```



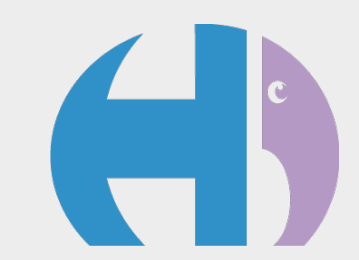
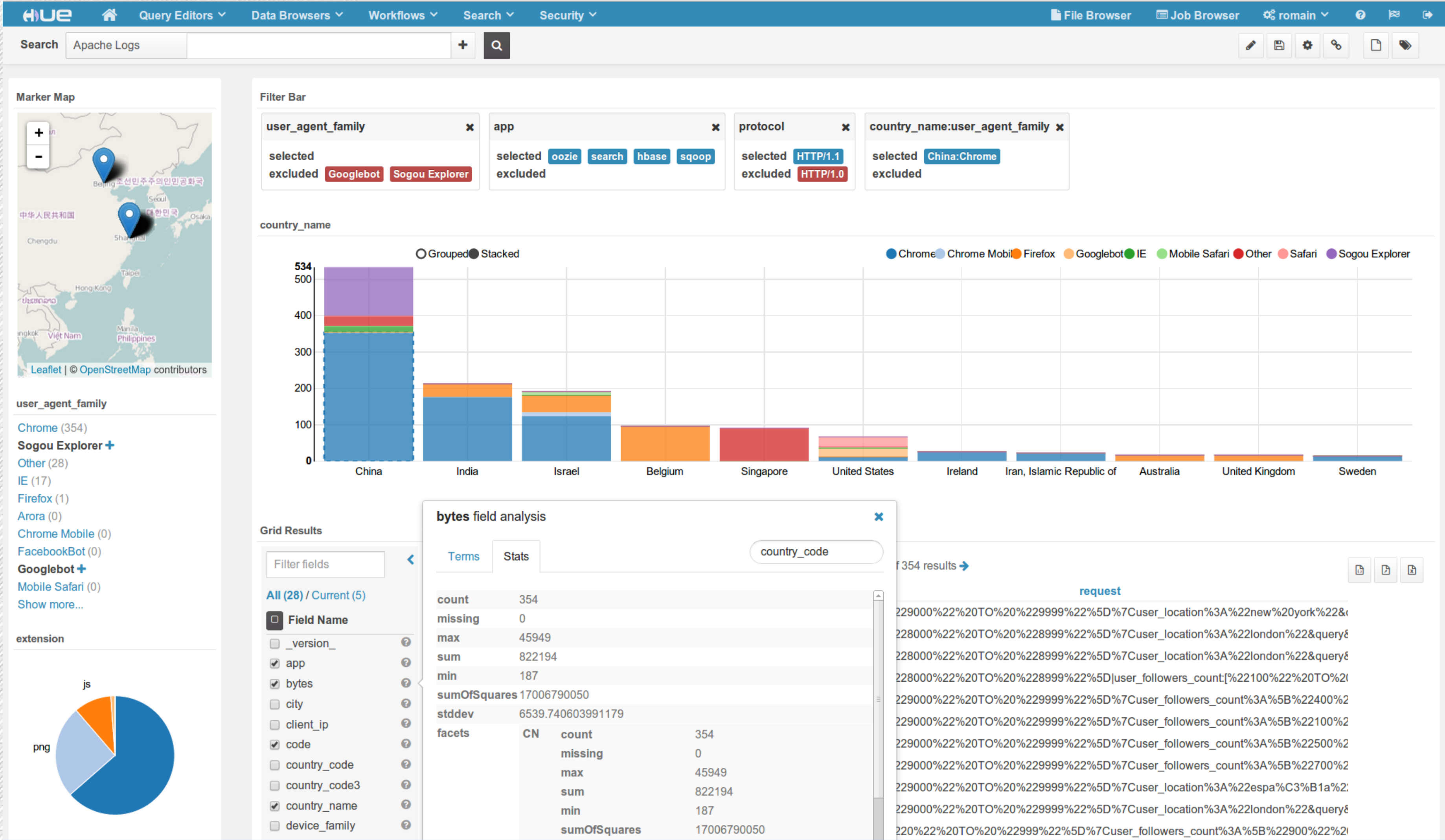
Total (8177)	US (1898)	US - accounts (767)	accounts - login (767)
		US - (667)	- (667)
	CN (1778)	CN - static (856)	static - ext (398)
		CN - oozie (379)	static - js (397)
	IL (1451)	IL - static (738)	static - ext (356)
	IN (1448)	IN - static (786)	static - ext (377)
	BE (393)		static - js (338)
	SG (386)		



- United States (1923)
- China (1898)
- India (1554)
- Israel (1535)
- Belgium (399)
- Singapore (386)
- Australia (237)
- Thailand (232)
- United Kingdom (227)
- Spain (155)
- Show more...



REPEAT UNTIL...



ENTERPRISE FEATURES

- Access to Search App configurable, LDAP/SAML auths
- Share by link
- Solr Cloud (or non Cloud)
- Proxy user
 /solr/jobs_demo/select?user.name=hue&doAs=romain&q=
- Security
 Kerberos
- [Sentry](#)
 Collection level, Solr calls like /admin, /query, Solr UI, ZooKeeper



SPARK IGNITER

HUE

Query Editors

Data Browsers

Workflows

Search

Other apps

File Browser

Job Browser

remain

Spark Igniter

Editor

Applications

Dashboard

Contexts

Uploads

App name

examples

Class path

spark.jobserver.WordCountExample

Context

Automatic

Create one?

Upload app

Create context

Parameters

Name

Value

input.string

a a a b c

..

Delete

Add

Execute

Save as...

or create a

New application

Download

Key

Value

a

3

c

1

b

1

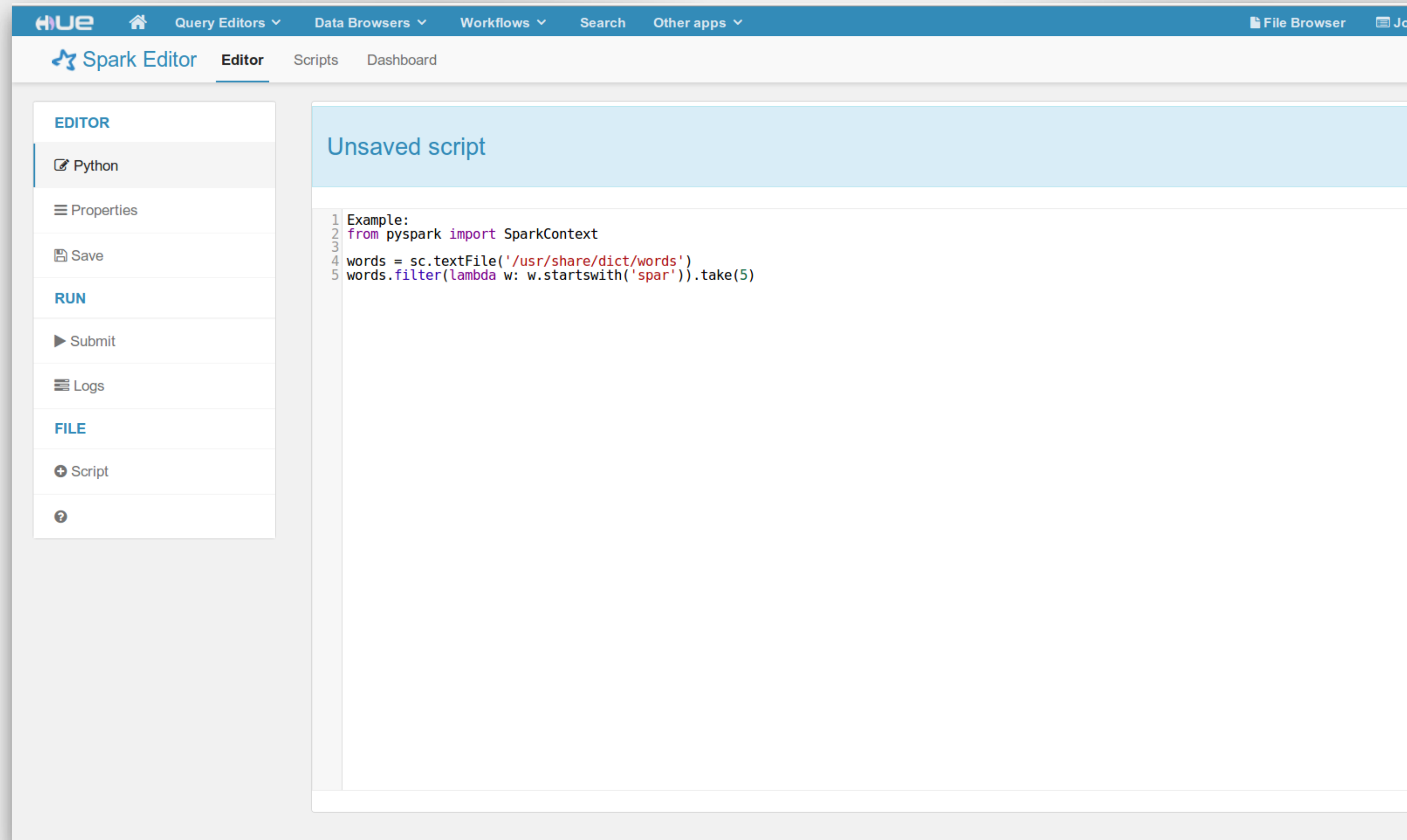
cloudera

HISTORY

OCT 2013

Submit through Oozie

Shell like for Java, Scala, Python



VICTORY

JAN 2014

V2 Spark Igniter

Spark 0.8

Java, Scala with Spark Job Server

APR 2014

Spark 0.9

JUN 2014

Ironing + How to deploy



Home

Query Editors

Data Browsers

Workflows

Search

Other apps

File Browser

Job Browser

romain

Spark Igniter

Editor

Applications

Dashboard

Contexts

Uploads

App name **examples**

Class path

Context ☒ Automatic ☐ Create one?

Upload app

Parameters

Name	Value	
input.string	a a a b c	..Delete

Add

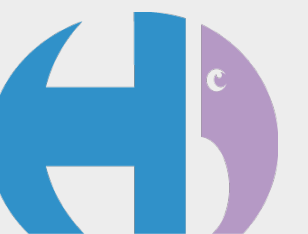
Execute

Save as...

or create a

New application

Key	Value
a	3
c	1
b	1



“JUST A VIEW” ON TOP OF SPARK



Saved script metadata

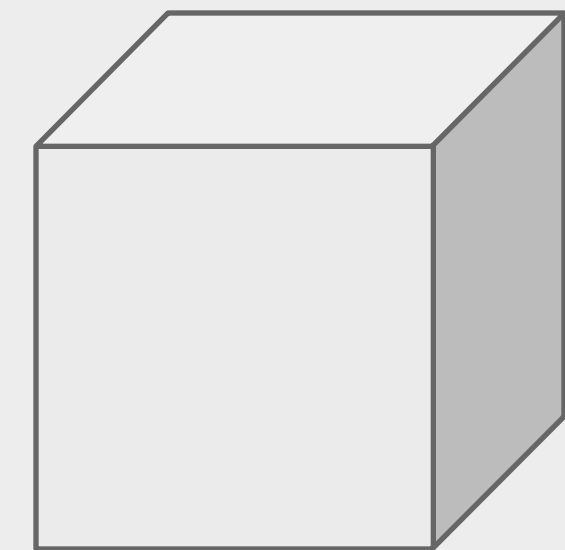
eg. name, args, classname, jar name...



Hue

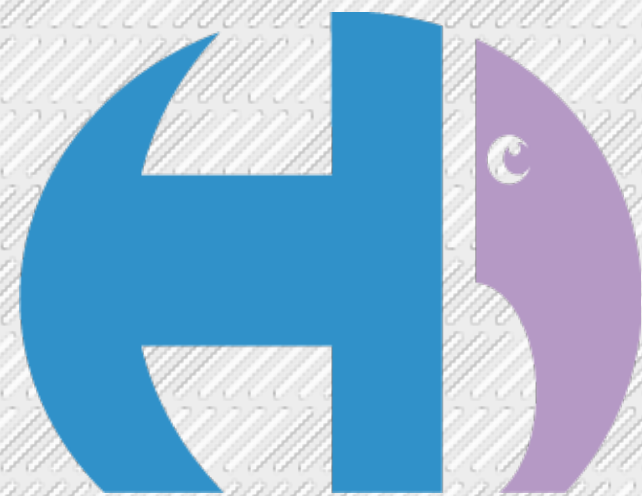


submit
list apps
list jobs
list contexts

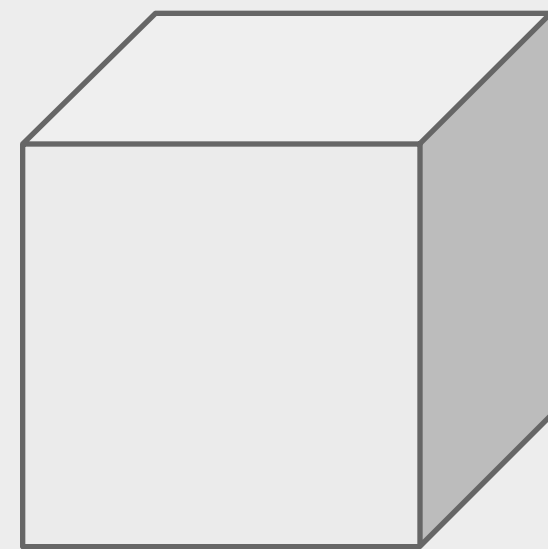


Job Server

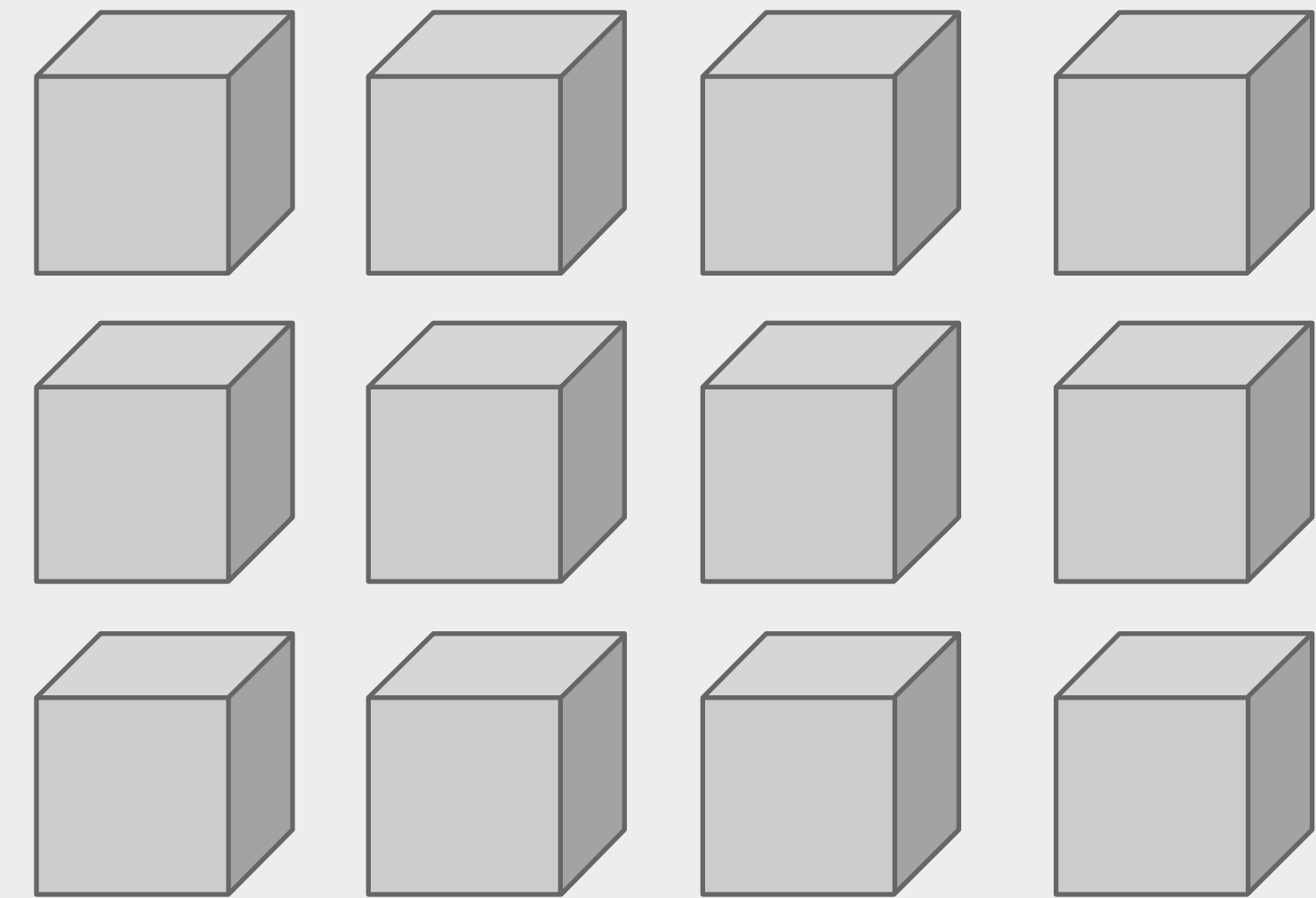
HOW TO TALK TO SPARK?



Hue

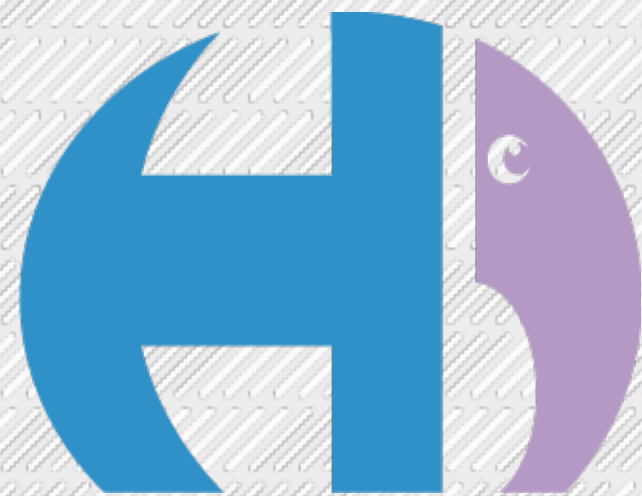


Spark Job Server

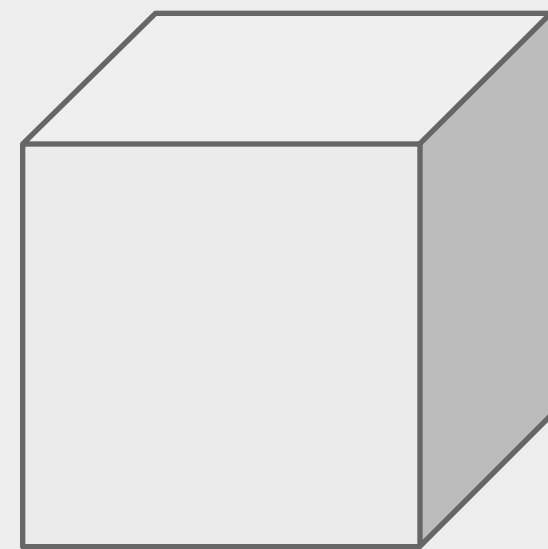


Spark

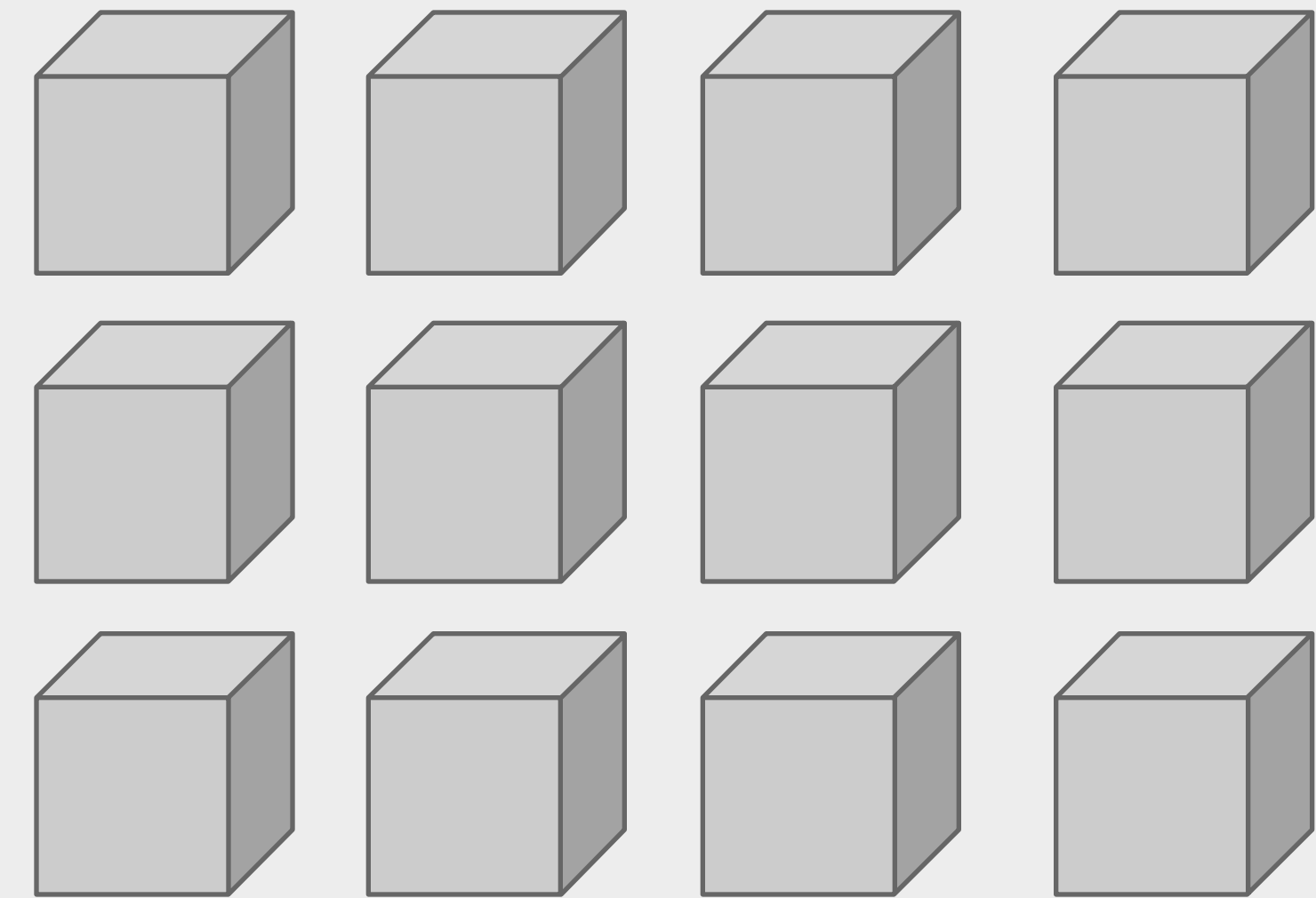
APP LIFE CYCLE



Hue



Spark Job Server



Spark

APP LIFE CYCLE

```
1 package spark.jobserver
2
3 import com.typesafe.config.{Config, ConfigFactory}
4 import org.apache.spark._
5 import scala.util.Try
6 import java.util.{Random, Date}
7
8 /**
9  * A long job for stress tests purpose.
10  * Iterative and randomized algorithm to compute Pi.
11  * Imagine a square centered at (1,1) of length 2 units.
12  * It tightly encloses a circle centered also at (1,1) of radius 1 unit.
13  * Randomly throw darts to them.
14  * We can use the ratio of darts inside the square and circle to approximate the Pi.
15  *
16  * stress.test.Longpijob.duration controls how long it run in seconds.
17  * Longer duration increases precision.
18  *
19  */
20 object LongPiJob extends SparkJob {
21   private val rand = new Random(now)
22
23   def main(args: Array[String]) {
24     val sc = new SparkContext("local[4]", "LongPiJob")
25     val config = ConfigFactory.parseString("")
26     val results = runJob(sc, config)
27     println("Result is " + results)
28   }
29
30   override def validate(sc: SparkContext, config: Config): SparkJobValidation = {
31     SparkJobValid
32   }
33
34   override def runJob(sc: SparkContext, config: Config): Any = {
35     val duration = Try(config.getInt("stress.test.longpijob.duration")).getOrElse(5)
36     var hit:Long = 0
37     var total:Long = 0
```

... extend SparkJob

.scala

sbt _/package

JAR

Upload

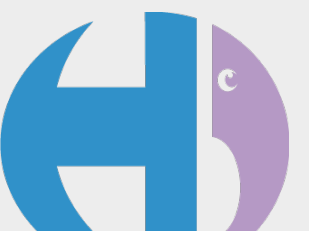
Add

Execute

Save as...

or create a

New a



APP LIFE CYCLE

```
1 package spark.jobserver
2
3 import com.typesafe.config.{Config, ConfigFactory}
4 import org.apache.spark._
5 import scala.util.Try
6 import java.util.{Random, Date}
7
8 /**
9  * A long job for stress tests purpose.
10  * Iterative and randomized algorithm to compute Pi.
11  * Imagine a square centered at (1,1) of length 2 units.
12  * It tightly encloses a circle centered also at (1,1) of radius 1 unit.
13  * Randomly throw darts to them.
14  * We can use the ratio of darts inside the square and circle to approximate the Pi.
15  *
16  * stress.test.Longpijob.duration controls how long it run in seconds.
17  * Longer duration increases precision.
18  *
19  */
20 object LongPiJob extends SparkJob {
21   private val rand = new Random(now)
22
23   def main(args: Array[String]) {
24     val sc = new SparkContext("local[4]", "LongPiJob")
25     val config = ConfigFactory.parseString("")
26     val results = runJob(sc, config)
27     println("Result is " + results)
28   }
29
30   override def validate(sc: SparkContext, config: Config): SparkJobValidation = {
31     SparkJobValid
32   }
33
34   override def runJob(sc: SparkContext, config: Config): Any = {
35     val duration = Try(config.getInt("stress.test.longpijob.duration")).getOrElse(5)
36     var hit:Long = 0
37     var total:Long = 0
```

... extend SparkJob

.scala

Context

sbt _/package

JAR

Upload

Add

Execute

Save as...

or create a

New app

create context: auto or manual

SPARK JOB SERVER

WHERE

<https://github.com/ooyala/spark-jobserver>

WHAT

REST job server for Spark

WHEN

Spark Summit talk Monday 5:45pm:

Spark Job Server: Easy Spark Job

Management by Ooyala

```
curl -d "input.string = a b c a b see" 'localhost:8090/jobs?
appName=test&classPath=spark.jobserver.WordCountExample'
{
  "status": "STARTED",
  "result": {
    "jobId": "5453779a-f004-45fc-a11d-a39dae0f9bf4",
    "context": "b7ea0eb5-spark.jobserver.WordCountExample"
  }
}
```

FOCUS ON UX

The screenshot shows the Cloudera Hue Spark Igniter interface. At the top, there's a navigation bar with tabs like Query Editors, Data Browsers, Workflows, Search, and Other apps. Below this, the 'Spark Igniter' section is active, showing the 'Editor' tab. The main area contains a form for configuring a new application. The 'App name' is set to 'examples', and the 'Class path' is 'spark.jobserver.WordCountExample'. The 'Context' is set to 'Automatic'. There are buttons for 'Upload app' and 'Create context'. Below this, the 'Parameters' section is expanded, showing a table with columns 'Name' and 'Value'. The table has one row with 'input.string' as the name and 'a a a b c' as the value. There are buttons for 'Add', 'Delete', 'Execute', 'Save as...', 'or create a', 'New application', and 'Download'. At the bottom, there's a table with columns 'Key' and 'Value' showing the word counts for the input string: 'a' is 3, 'c' is 1, and 'b' is 1.

Name	Value
input.string	a a a b c

Key	Value
a	3
c	1
b	1

VS

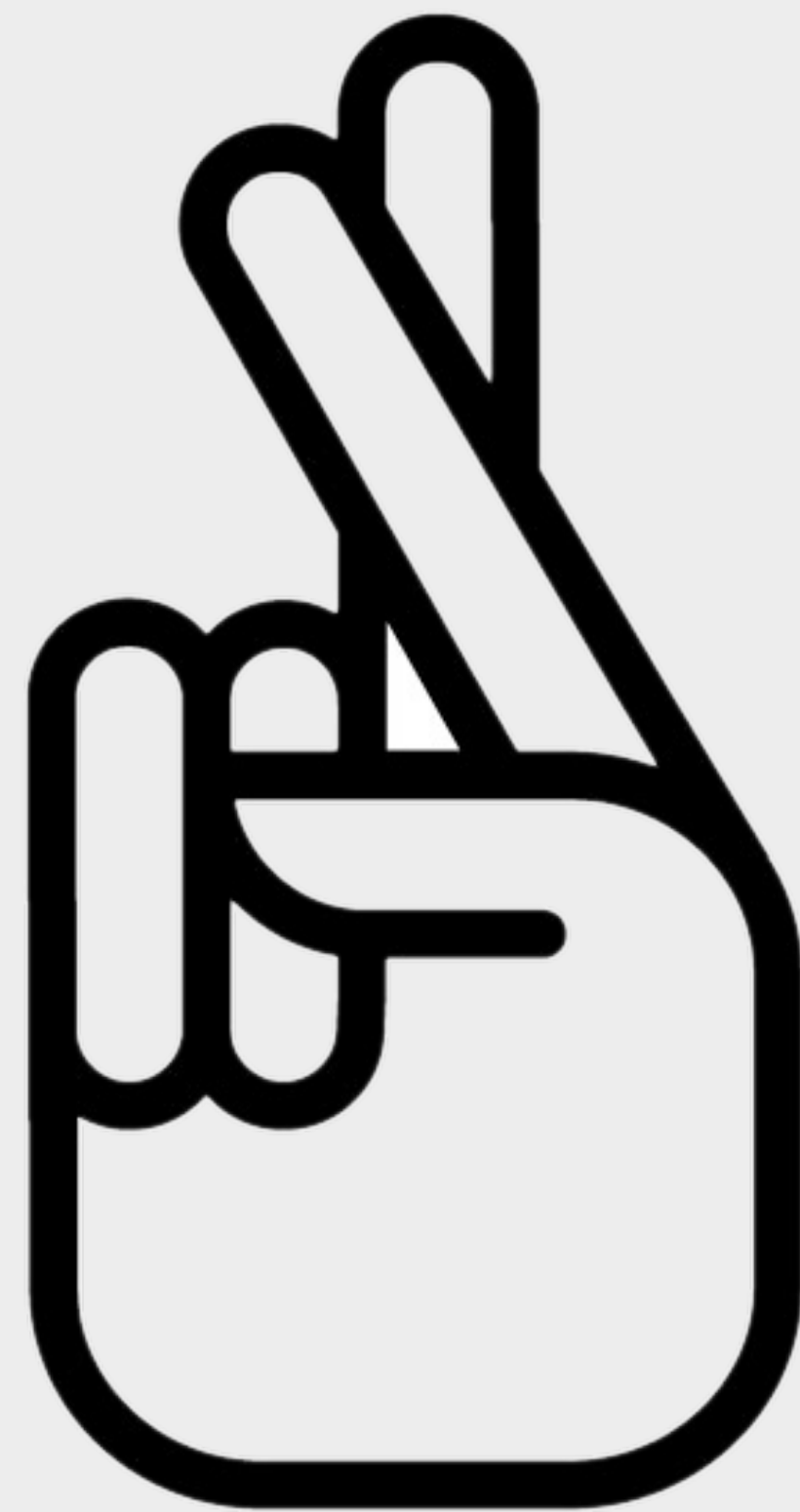
```
curl -d "input.string = a b c a b see" 'localhost:8090/jobs?
appName=test&classPath=spark.jobserver.WordCountExample'
{
  "status": "STARTED",
  "result": {
    "jobId": "5453779a-f004-45fc-a11d-a39dae0f9bf4",
    "context": "b7ea0eb5-spark.jobserver.WordCountExample"
  }
}
```

TRAIT SPARKJOB

```
/**
 * This trait is the main API for Spark jobs submitted to the Job Server.
 */
trait SparkJob {
  /**
   * This is the entry point for a Spark Job Server to execute Spark jobs.
   */
  def runJob(sc: SparkContext, jobConfig: Config): Any

  /**
   * This method is called by the job server to allow jobs to validate their input and reject
   * invalid job requests. */
  def validate(sc: SparkContext, config: Config): SparkJobValidation
}
```

DEMO TIME



SUM-UP

INSTALL

Install Hue on one machine

LDAP

Use an LDAP backend

ENABLE

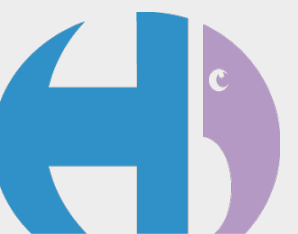
Enable Hadoop Service APIs
for Hue as a proxy user

HELP

Get help on @gethue or hue-
user

CONFIGURE

Configure hue.ini to point to
each Service API



ROADMAP NEXT 6 MONTHS

WHAT

Oozie v2

Spark v2

SQL v2

More dashboards!

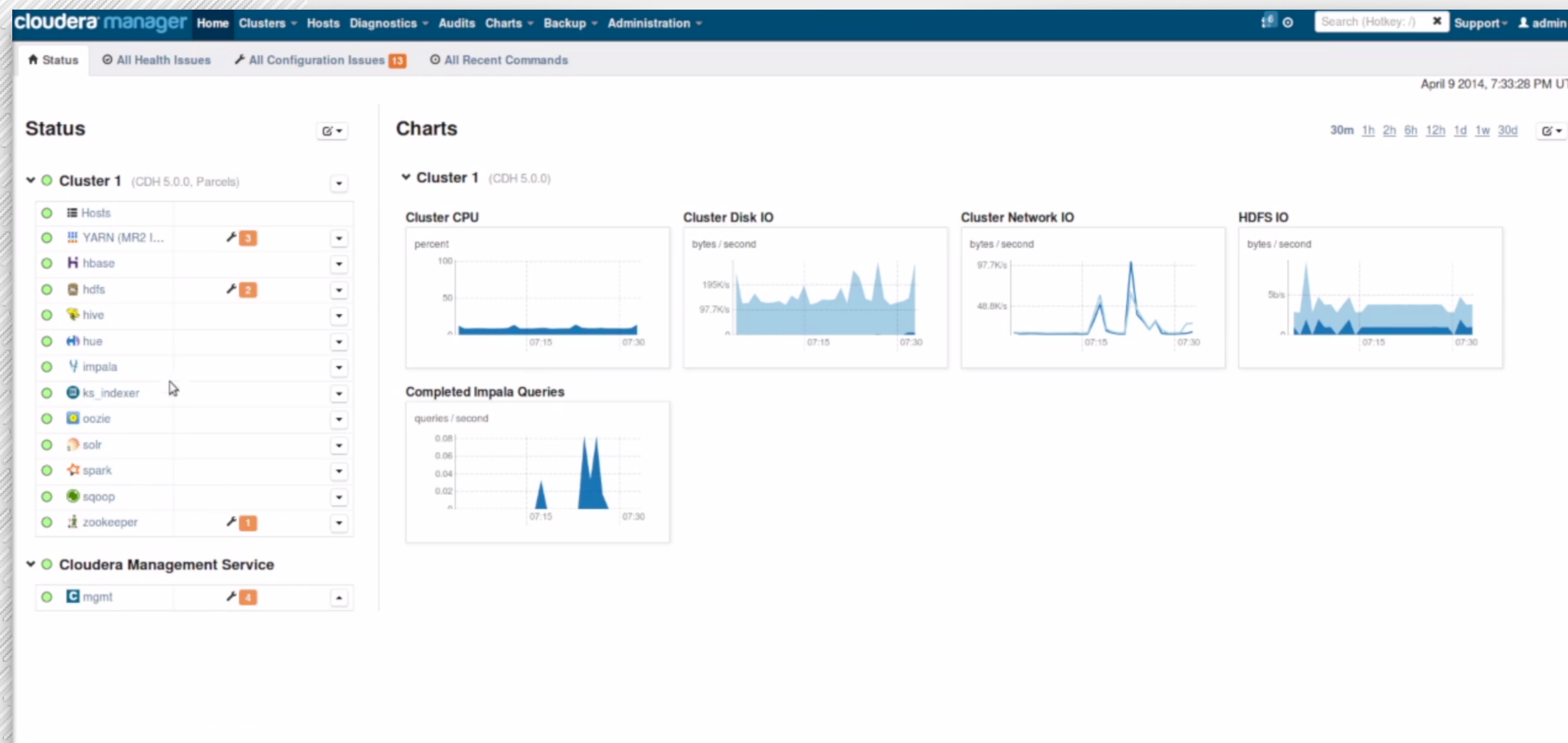
Inter component integrations
(HBase <-> Search, create index
wizards, document permissions),
Hadoop Web apps SDK

Your idea here.



CONFIGURATIONS ARE HARD...

...GIVE CLOUDERA MANAGER A TRY!



vimeo.com/91805055

MISSED
SOMETHING?

learn.gethue.com



HomeDown

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

Hadoop Voice Recognition: use your voice to use Hadoop!
By Hue Team on April 1, 2014

GRACIAS!

WEBSITE

<http://gethue.com>

LEARN

<http://learn.gethue.com>

TWITTER

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

USER GROUP

hue-user@

cloudera®

