

HUE 超好用各类数据网页插件

1、介绍

HUE=Hadoop User Experience Hue是一个开源的Apache Hadoop UI系统，由Cloudera Desktop演化而来，最后Cloudera公司将其贡献给Apache基金会的Hadoop社区，它是基于Python Web框架Django实现的。

通过使用Hue我们可以在浏览器端的Web控制台上与Hadoop集群进行交互来分析处理数据，例如操作HDFS上的数据，运行MapReduce Job，执行Hive的SQL语句，浏览Hbase数据库等等。

2、安装

2.1 安装hue依赖的第三方包

```
sudo yum install ant asciidoc cyrus-sasl-devel cyrus-sasl-gssapi cyrus-sasl-plain gcc gcc-c++ krb5-devel libffi-devel libxml2-devel libxslt-devel make mysql mysql-devel openldap-devel python-devel sqlite-devel gmp-devel
```

注意： yum 安装 ant 会自动下载安装 openJDK，这样的话会 java 的版本就会发生变化，可是使用软连接重新改掉 /usr/bin/java，使其指向自己安装的 java 即可。

```
sudo rm /usr/bin/java
sudo ln -s /home/darren/program/java/bin/java /usr/bin/java
```

注意： 本来我是准备安装 hue4.2.0 的，但是编译 hue 需要依赖 python，但是 python2.6 中缺少必要的依赖，不能编译 hue4.2，安装 python2.7 后又导致 yum 失效，由于时间紧任务重，就没有再解决升级 python2.7 后带来的问题，于是就降低了 hue 的版本到 3.7，之后顺利编译通过。如果已经升级了 python2.7，可以尝试安装 hue4。

2.2 解压HUE tar包

```
tar -zxvf hue-3.7.1.tgz
mv hue-3.7.1 program/hue
```

2.3 编译HUE

```
cd program/hue
make apps
```

几分钟就编译好了

3、配置HUE

```
vi program/hue/desktop/conf/hue.ini
```

修改如下配置：


```
secret_key=jFE93j;2[290eiw.KEiwN2s3['d;/.q[eIW^y#e=+Iei*@Mn<qw5o
http_host=master
http_port=8888
time_zone=Asia/Shanghai
```

4、启动HUE

```
program/hue/build/env/bin/supervisor >> /home/darren/program/hue/log/hue.log
2>&1 &
```

5、访问HUE 页面

<http://centos1:8888>

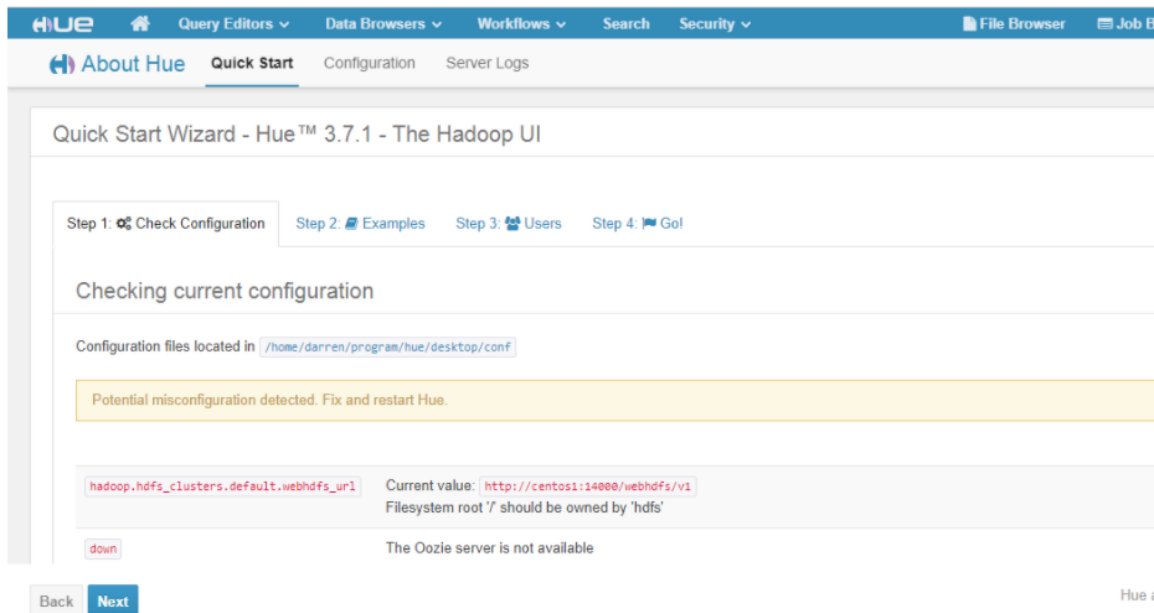


创建您的 Hue 帐户

由于这是您的首次登录，因此请选择任意用户名和密码。务必牢记此用户名和密码，因为它们将成为您的 Hue 超级用户凭据。..

创建帐户

第一次会让你创建账户，登录之后如下所示：



到此，HUE就安装配置好了，接下来进行和hadoop，hive的集成。

注意：HUE默认使用的SQLite数据库，可以更改成其他数据库。

6、配置MySQL数据库

6.1 配置MySQL用户和权限

先使用root用户登录mysql

```
mysql -u root -p
```

添加新用户并授权远程访问和本地访问

```
mysql> GRANT ALL PRIVILEGES ON *.* TO 'hue'@'%' IDENTIFIED BY 'hue';
mysql> GRANT ALL PRIVILEGES ON *.* TO 'hue'@'centos1' IDENTIFIED BY 'hue';
mysql> GRANT ALL PRIVILEGES ON *.* TO 'hue'@'localhost' IDENTIFIED BY 'hue';

mysql> flush privileges;
```

查看权限

```
mysql> select host, user from user;
```

```
mysql> select host, user from user;
+-----+-----+
| host      | user  |
+-----+-----+
| %         | hive  |
| %         | hue   |
| 127.0.0.1 | root  |
| centos1   |       |
| centos1   | hive  |
| centos1   | hue   |
| centos1   | root  |
| localhost |       |
| localhost | hive  |
| localhost | hue   |
| localhost | root  |
+-----+-----+
11 rows in set (0.00 sec)
```

使用hue账户登录，创建database hue

```
mysql -u hue -p
mysql> create database hue;
```

6.2 修改HUE的配置文件

```
vi program/hue/desktop/conf/hue.ini
# Configuration options for specifying the Desktop Database. For more info,
# see http://docs.djangoproject.com/en/1.4/ref/settings/#database-engine
# -----
[[database]]
# Database engine is typically one of:
# postgresql_psycopg2, mysql, sqlite3 or oracle.
#
# Note that for sqlite3, 'name', below is a path to the filename. For other
backends, it is the database name.
# Note for Oracle, options={'threaded':true} must be set in order to avoid
crashes.
# Note for Oracle, you can use the Oracle Service Name by setting "port=0"
and then "name=<host>:<port>/<service_name>".

engine=mysql //数据库
host=centos1 //主机名或IP地址
port=3306 //MySQL 端口
user=hue //MySQL 用户
password=hue //MySQL 密码
name=hue //数据库名字
```

```
## options={}
```

6.3 初始化数据库

该步骤是创建表和插入部分数据。hue的初始化数据表命令由hue/bin/hue syncdb完成，创建期间，需要输入用户名和密码。如下所示：

```
#同步数据库
$> program/hue/build/env/bin/hue syncdb

#导入数据,主要包括oozie、pig、desktop所需要的表
$> program/hue/build/env/bin/hue migrate
```

注意：这里是两个命令，容易忽略第二个

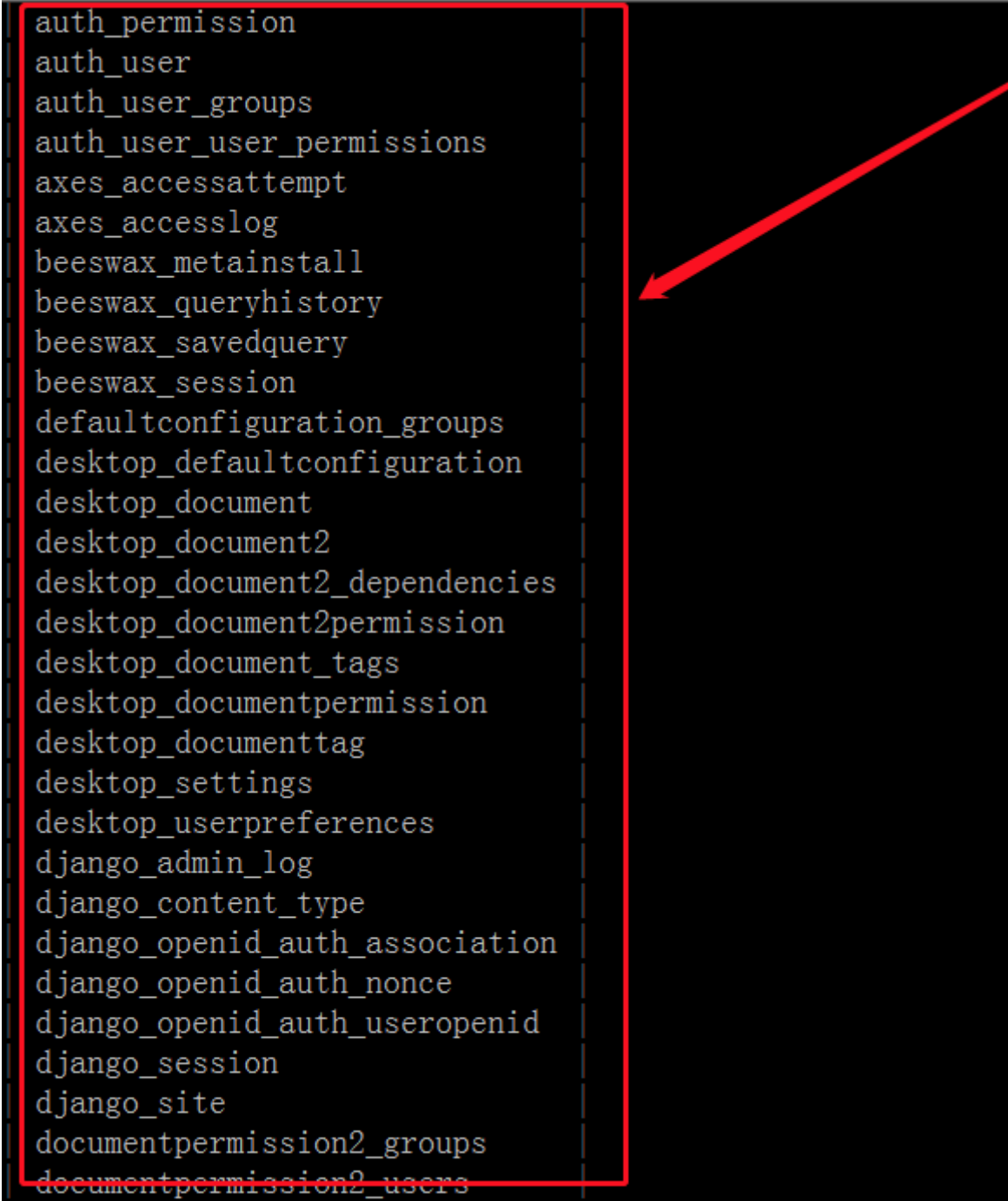
```
[centos@s101 /home/centos]$ ~/hue-3.12.0/build/env/bin/hue syncdb
Syncing...
Creating tables ...
Creating table auth_permission
Creating table auth_group_permissions
Creating table auth_group
Creating table auth_user_groups
Creating table auth_user_user_permissions
Creating table auth_user
Creating table django_openid_auth_nonce
Creating table django_openid_auth_association
Creating table django_openid_auth_useropenid
Creating table django_content_type
Creating table django_session
Creating table django_site
Creating table django_admin_log
Creating table south_migrationhistory
Creating table axes_accessattempt
Creating table axes_accesslog

You just installed Django's auth system, which means you don't have any superusers defined.
Would you like to create one now? (yes/no):
```

输入的是机器的用户名和密码

使用hue登录mysql，查看表的生成情况：

```
mysql -u hue -p
mysql> use hue;
mysql> show tables;
```



```
auth_permission
auth_user
auth_user_groups
auth_user_user_permissions
axes_accessattempt
axes_accesslog
beeswax_metainstall
beeswax_queryhistory
beeswax_savedquery
beeswax_session
defaultconfiguration_groups
desktop_defaultconfiguration
desktop_document
desktop_document2
desktop_document2_dependencies
desktop_document2permission
desktop_document_tags
desktop_documentpermission
desktop_documenttag
desktop_settings
desktop_userpreferences
django_admin_log
django_content_type
django_openid_auth_association
django_openid_auth_nonce
django_openid_auth_useropenid
django_session
django_site
documentpermission2_groups
documentpermission2_users
```

如果发现表没这么多，那么你一定忘记执行如下命令了：

```
#导入数据,主要包括oozie、pig、desktop所需要的表
$> program/hue/build/env/bin/hue migrate
```

查看hue进程，杀掉重启

```
netstat -npl | grep 8888
kill -9 xxx
program/hue/build/env/bin/supervisor >> /home/darren/program/hue/log/hue.log
2>&1 &
```

访问HUE UI界面



Sign in to continue to Hue

Sign in

使用之前的账户登录即可，没有什么问题，数据库就替换完成了。

7、配置hadoop

7.1 修改hadoop配置文件

```
vi program/hadoop/etc/hadoop/core-site.xml
```

添加如下配置，配置hadoop代理用户hadoop.proxyuser.\${user}.hosts，第一个user是安装hadoop的user，或者说可以访问hdfs的用户，从centos1:50070 -》Utilities-》Browse the file system可以看到的Owner信息，第二个hue是给hue这样的权限，第三个是给httpfs这样的权限：

```
<!-- Hue WebHDFS proxy user setting -->
<property>
  <name>hadoop.proxyuser.darren.hosts</name>
  <value>*</value>
</property>
<property>
  <name>hadoop.proxyuser.darren.groups</name>
  <value>*</value>
</property>

<property>
  <name>hadoop.proxyuser.hue.hosts</name>
  <value>*</value>
</property>

<property>
  <name>hadoop.proxyuser.hue.groups</name>
  <value>*</value>
</property>

<property>
  <name>hadoop.proxyuser.httpfs.hosts</name>
  <value>*</value>
</property>
```

```
<property>
  <name>hadoop.proxyuser.httpfs.groups</name>
  <value>*</value>
</property>
```

为什么会有 httpfs，httpfs 是什么？

HUE 与 hadoop 连接，即访问 hadoop 文件，可以使用两种方式。

- webHDFS

提供高速数据传输，client可以直接和 DataNode 通信。

- HttpFS

一个代理服务，方便于集群外部的系统进行集成。**注意：**HA模式下只能使用该中方式。

高可用模式下需要配置 httpfs，否则报错。

7.2 开启运行HUE web访问HDFS

```
vi program/hadoop/etc/hadoop/hdfs-site.xml
<!-- 设置hue web access -->
<property>
  <name>dfs.webhdfs.enabled</name>
  <value>true</value>
</property>
```

7.3 配置httpfs

```
vi program/hadoop/etc/hadoop/httpfs-site.xml

<!-- 配置HUE -->
<property>
  <name>httpfs.proxyuser.hue.hosts</name>
  <value>*</value>
</property>
<property>
  <name>httpfs.proxyuser.hue.groups</name>
  <value>*</value>
</property>
```

7.4 关掉hadoop集群，分发配置文件到其他节点，重新启动

```
stop-all.sh

# 分发,其他省略
scp program/hadoop/etc/hadoop/core-site.xml centos2:~/program/hadoop/etc/hadoop/

# 分发完毕后重启
start-all.sh
```


7.5 启动httpFS

```
httpfs.sh start
```

```
# 启动后检查端口，默认14000
```

```
netstat -anop |grep 14000
```

```
[centos@s101 /soft/hadoop/sbin]$netstat -anop |grep 14000
(Not all processes could be identified, non-owned process info
will not be shown, you would have to be root to see it all.)
tcp6      0      0 :::14000          :::*               LISTEN      11042/java        off (0.00/0/0)
[centos@s101 /soft/hadoop/sbin]$clear
[centos@s101 /soft/hadoop/sbin]$cd ~/hue-3.12.0/build/env/
[centos@s101 ~/home/centos/hue-3.12.0/build/env]$ls
```

hadoop的准备工作完毕，接下来配置HUE的配置文件，完成对hadoop的集成

7.6 配置hue.ini，集成hadoop

```
[[yarn_clusters]]
```

```
[[[default]]]
```

```
# Enter the host on which you are running the ResourceManager
## resourcemanager_host=localhost
```

```
# The port where the ResourceManager IPC listens on
## resourcemanager_port=8032
```

```
# Whether to submit jobs to this cluster
submit_to=True
```

```
# Resource Manager logical name (required for HA)
logical_name=mycluster-yarn
```

```
# Change this if your YARN cluster is Kerberos-secured
## security_enabled=false
```

```
# URL of the ResourceManager API
# 配置resource manager
resourcemanager_api_url=http://centos1:8088
```

```
# URL of the ProxyServer API
## proxy_api_url=http://localhost:8088
```

```
# URL of the HistoryServer API
# 配置 history server
history_server_api_url=http://centos1:19888
```

```
# HA support by specifying multiple clusters
```

```
[[yarn_clusters]]
```

```
[[[default]]]
```

```
# Enter the host on which you are running the ResourceManager
## resourcemanager_host=localhost
```

```
# The port where the ResourceManager IPC listens on
## resourcemanager_port=8032
```

```

# Whether to submit jobs to this cluster
submit_to=True

# Resource Manager logical name (required for HA)
logical_name=mycluster-yarn

# Change this if your YARN cluster is Kerberos-secured
## security_enabled=false

# URL of the ResourceManager API
# 配置resource manager
resourcemanager_api_url=http://centos1:8088

# URL of the ProxyServer API
## proxy_api_url=http://localhost:8088

# URL of the HistoryServer API
# 配置 history server
history_server_api_url=http://centos1:19888

# HA support by specifying multiple clusters

# webserver runs as this user
server_user=hue
server_group=hue

# This should be the Hue admin and proxy user
default_user=hue

# This should be the hadoop cluster admin
default_hdfs_superuser=hue

```

History Server的启动方式:

```
mr-jobhistory-daemon.sh start historyserver
```

重新启动HUE

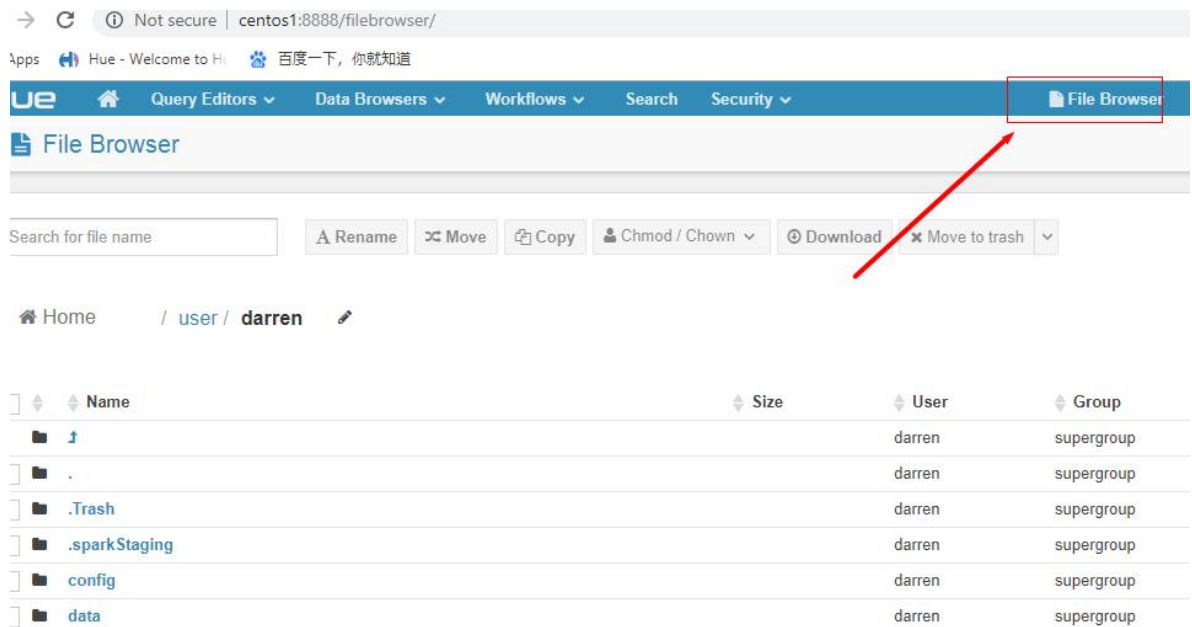
```

netstat -npl | grep 8888

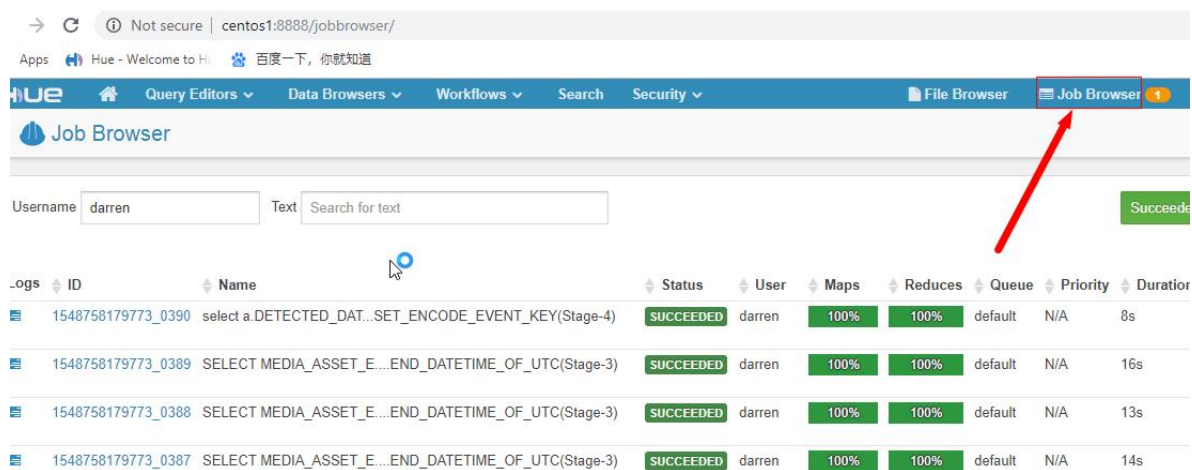
kill -9 xxx

program/hue/build/env/bin/supervisor >> /home/darren/program/hue/log/hue.log
2>&1 &

```



点击File Browser，可以看到如上信息，当然文件夹是我之前建好的，这说明HDFS集成好了。



点击Job Browser，说明resource manager集成好了

注意：可能会遇到的问题

```
RemoteException: User darren is not allowed to impersonate darren (error 500)
```

这个错就需要修改core-site.xml, 如上文所述。

如果报connection refuse，那可能是https没有启动，注意配置https-site.xml如上文所示，并启动https，启动方式如下：

```
https.sh start
```

Hive集成

hive集成就比较简单了，配置主机，端口和配置文件路径即可

[beeswax]

```
# Host where HiveServer2 is running.
# If Kerberos security is enabled, use fully-qualified domain name (FQDN).

hive_server_host=centos1

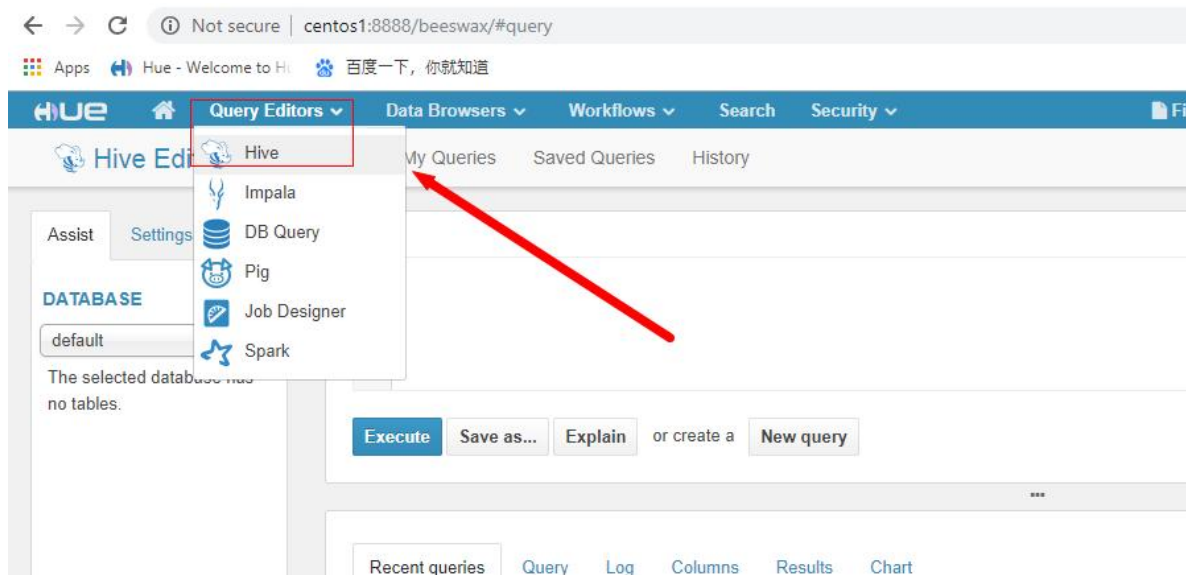
# Port where HiveServer2 Thrift server runs on.
hive_server_port=10000

# Hive configuration directory, where hive-site.xml is located
hive_conf_dir=/home/darren/program/hive/conf
```

注意：集成hive使用的是HiveServer2，所以要求启动HiveServer2服务，启动方式如下：

```
hive --service hiveserver2 >> /home/darren/program/hive/log/hiveserver2.log 2>&1
&
```

重新启动HUE，即可看到Hive的数据库和表



到此就集成完了