大数据Web工具Hue

谭唐华

课程大纲

- Hue介绍
- Hue部署

Cloudera Hue

Let's big data.

Hue is a Web interface for analyzing data with Apache Hadoop.

- 参考文档:
  - http://gethue.com/
  - http://archive.cloudera.com/cdh5/cdh/5/hue-3.7.0-cdh5.3.6/manual.html#_install_hue
  - https://github.com/cloudera/hue

Cloudera Hue

Free & Open Source

Public source code with an active community

Be productive
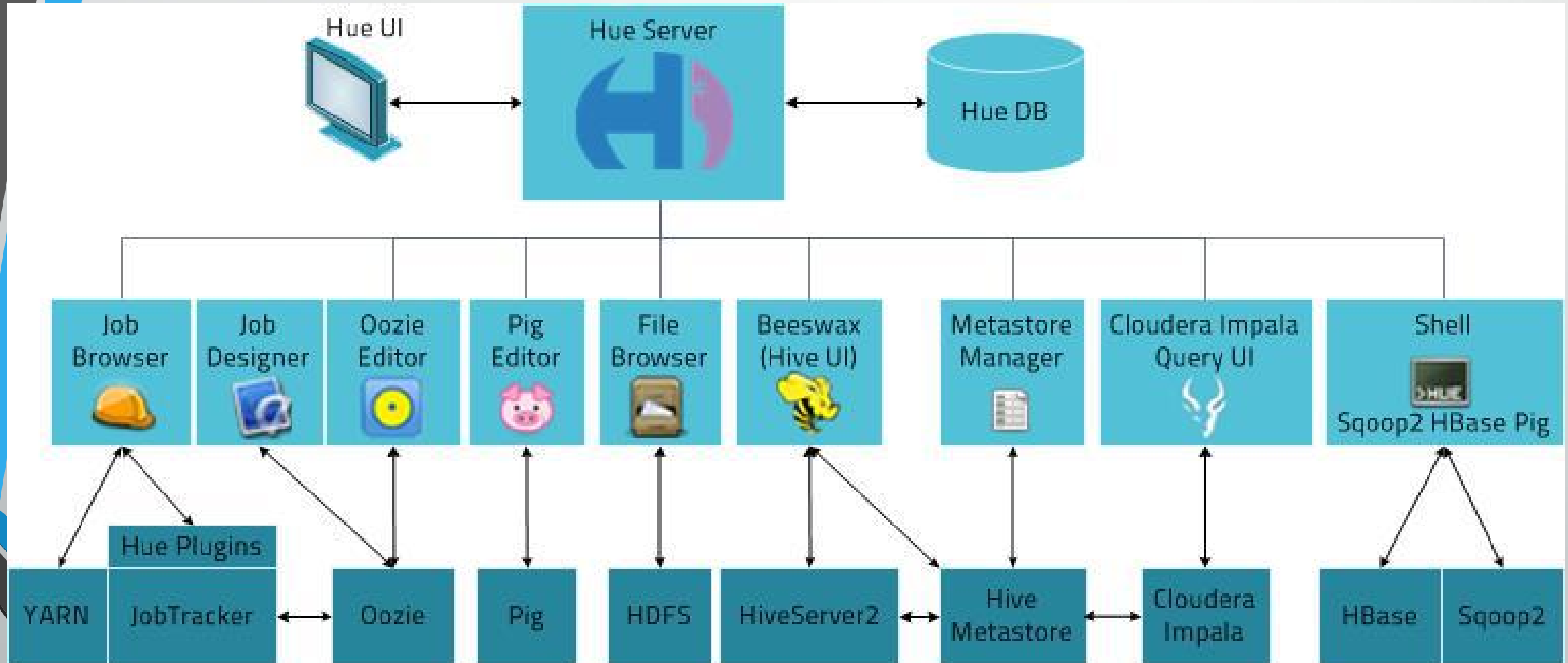
Query and visualize data directly from your Browser

100% Compatible

Works with ANY Hadoop

Hue Architecture

课程大纲

- Hue介绍
- Hue部署

Hue Installation

- CDH 5.3.6版本，下载地址：

  http://archive.cloudera.com/cdh5/cdh/5/hue-3.7.0-cdh5.3.6.tar.gz

- Hue 编译

  - 在编译Hue之前，要做如何两件准备工作：

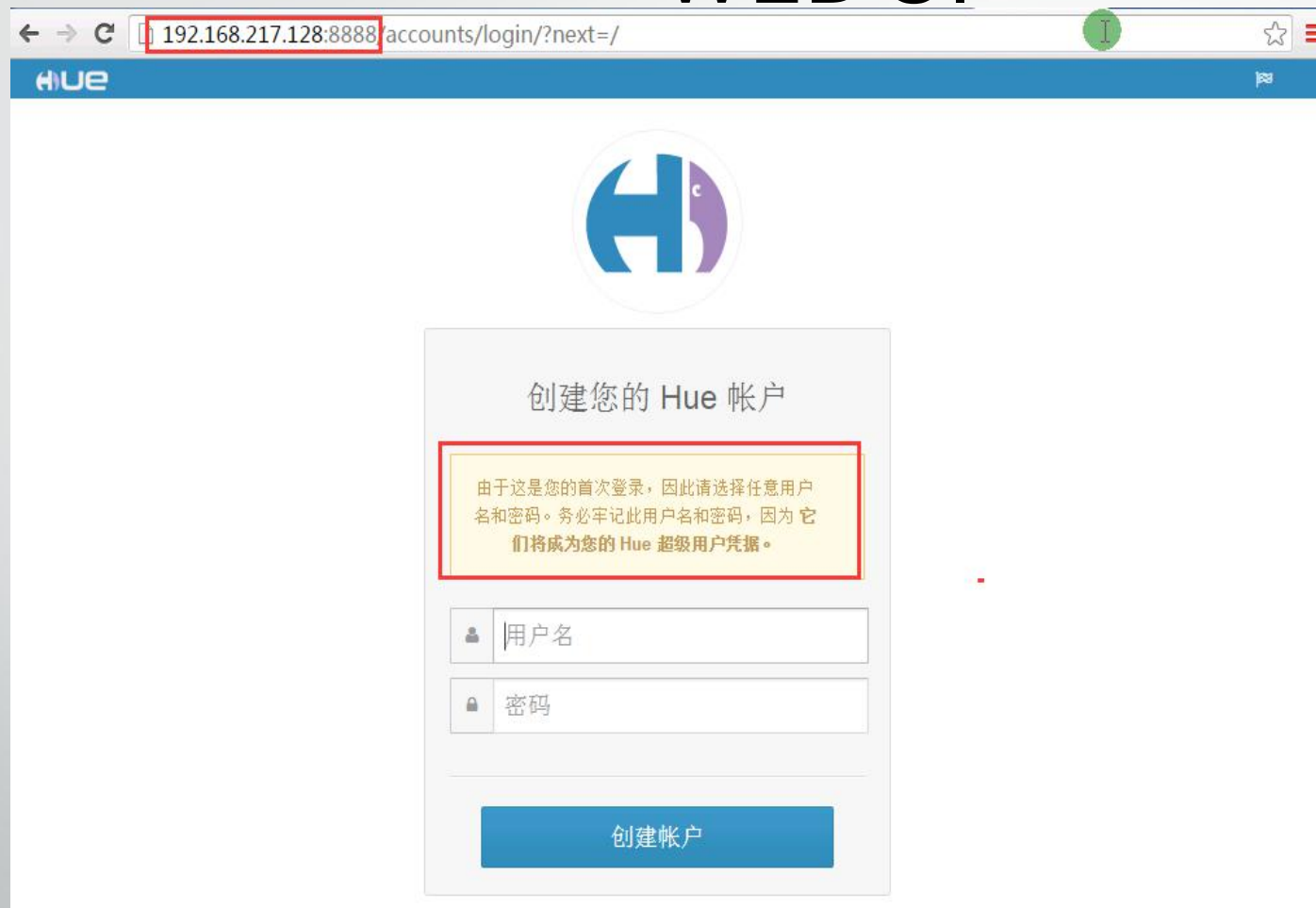 - 虚拟机要连接互联网

  - 安装Hue所依赖的系统包，针对不同Unix系统

Hue Installation

- 1、安装依赖包[联网]

- yum -y install ant asciidoc cyrus-sasl-devel cyrus-sasl-gssapi gcc gcc-c++ krb5-devel libtidy libxml2-devel libxslt-devel openldap-devel python-devel sqlite-devel openssl-devel gmp-devel

- **卸载安装的jdk:卸载系统自带的JDK

- # rpm -qa | grep -i java

- # rpm -e --nodeps tzdata-java-2012j-1.el6.noarch

- 2、解压、编译hue

  - $ tar zxf /home/xx/hue-3.7.0-cdh5.3.6.tar.gz –C ./

  - $ cd hue-3.7.0-cdh5.3.6/

  - $ make apps

Hue Installation

- 3、修改desktop/conf/hue.ini

  - [desktop]

  - secret_key=jFE93j;2[290-eiw.KEiwN2s3['d;/.q[eIW^y#e=+Iei*@Mn<qW5o'

  - http_host=bigdata.xx.com

  - http_port=8888

  - time_zone=Asia/Shanghai

- 4、启动Hue

- $ build/env/bin/supervisor

WEB UI

第一次登陆Hue时，会有提示，随便输入用户和密码，必须要记住，这就是以后登陆Hue管理员的用户名和密码。

[Hue与Hadoop集成]

- 配置Hadoop目录的配置文件：
- **hdfs-site.xml**
  - \<property>
    - \<name>dfs.webhdfs.enabled\</name>
    - \<value>true\</value>
  - \</property>
- **core-site.xml**
  - \<property>
    - \<name>hadoop.proxyuser.hue.hosts\</name>
    - \<value>*\</value>
  - \</property>
  - \<property>
    - \<name>hadoop.proxyuser.hue.groups\</name>
    - \<value>*\</value>
  - \</property>
- ！！！ 重启Hadoop集群服务

修改hue.ini

- 修改hue.ini

- [[hdfs_clusters]]

-   fs_defaultfs=hdfs://bigdata.xx.com:8020

-   webhdfs_url=http://bigdata.xx.com:50070/webhdfs/v1

- hadoop_conf_dir=/opt/modules/hadoop-2.5.0-cdh5.3.6/etc/hadoop

- hadoop_hdfs_home=/opt/modules/hadoop-2.5.0-cdh5.3.6/

- hadoop_bin =/opt/modules/hadoop-2.5.0-cdh5.3.6/bin/

修改hue.ini

- [[yarn_clusters]]
- resourcemanager_host=bigdata.xx.com
- resourcemanager_port=8032
- submit_to=True
  - resourcemanager_api_url=http://bigdata.xx.com:8088
  - proxy_api_url=http://bigdata.xx.com:8088
  - history_server_api_url=http://bigdata.xx.com:19888

Hue与Hive集成

- Hive准备：

- 1、安装配置好hive

- 2、修改hive-site.xml：启用Hiveserver2服务

  - \<property>

    - \<name>hive.server2.thrift.port\</name>

    - \<value>10000\</value>

  - \</property>

  - \<property>

    - \<name>hive.server2.thrift.bind.host\</name>

    - \<value>bigdata.xx.com\</value>

  - \</property>

  - \<property>

    - \<name>hive.server2.long.polling.timeout\</name>

    - \<value>5000\</value>

  - \</property>

Hue与Hive集成

- 3、启用thriftserver服务
  - <property>
  - <name>hive.metastore.uris</name>
  - <value>thrift://bigdata.xx.com:9083</value>
  - </property>

- 4、启动服务:
  - bin/hive --service hiveserver2 &
  - bin/hive --service metastore &

Hue与Hive集成

- 配置Hue.ini
- [beeswax]
-   hive_server_host=bigdata.xx.com
    - hive_server_port=10000
    - hive_conf_dir=/opt/modules/hive-0.13.1-cdh5.3.6/conf

Hue与mysql集成

- [[[mysql]]]
  - nice_name="My SQL DB "
  - engine=mysql
  - host=bigdata.xx.com
  - port=3306
  - user=root
  - password=123456