

HUE

1.简述

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。

Hue 在数据库方面，默认使用的是 SQLite 数据库来管理自身的数据，包括用户认证和授权，另外，可以自定义为 MySQL 数据库、Postgresql 数据库、以及 Oracle 数据库。其自身的功能包含有：

- 对 HDFS 的访问，通过浏览器来查阅 HDFS 的数据。
- Hive 编辑器：可以编写 HQL 和运行 HQL 脚本，以及查看运行结果等相关 Hive 功能。
- 提供 Solr 搜索应用，并对应相应的可视化数据视图以及 DashBoard。
- 提供 Impala 的应用进行数据交互查询。
- 最新的版本集成了 Spark 编辑器和 DashBoard
- 支持 Pig 编辑器，并能够运行编写的脚本任务。
- Oozie 调度器，可以通过 DashBoard 来提交和监控 Workflow、Coordinator 以及 Bundle。
- 支持 HBase 对数据的查询修改以及可视化。
- 支持对 Metastore 的浏览，可以访问 Hive 的元数据以及对应的 HCatalog。
- 另外，还有对 Job 的支持，Sqoop，ZooKeeper 以及 DB（MySQL，SQLite，Oracle 等）的支持。

2.HUE IN HADOOP

2.1 基础软件

在集成 Hue 工具时，我们需要去下载对应的源码，该系统是开源免费的，可以在 GitHub 上下载到对应的源码，下载地址如下所示：

```
git@github.com:cloudera/hue.git
```

我们使用 **Git** 命令将其克隆下来，命令如下所示：

```
git clone git@github.com:cloudera/hue.git
```

在 Hadoop 账号下安装 Hue 需要的依赖环境，命令如下所示：

```
sudo yum install krb5-devel cyrus-sasl-gssapi cyrus-sasl-devel libxml2-devel libxslt-devel mysql mysql-devel openldap-devel python-devel python-simplejson sqlite-devel
```

2.2 编译部署

在基础环境准备完成后，我们开始对 Hue 的源码进行编译，编译的时候，Python 的版本需要是 2.6+ 以上，不然在编译的时候会出现错误，编译命令如下所示：

```
[hadoop@nna ~]$ cd hue
[hadoop@nna ~]$ make apps
```

等待其编译完成，在编译的过程中有可能会出现问题，在出现错误时 Shell 控制台会提示对应的错误信息，大家可以根据错误信息的提示来解决对应的问题，在编译完成后，我们接下来需要对其做对应的配置，Hue 的默认配置是启动本地的 Web 服务，这个我们需要对其修改，供外网或者内网去访问其 Web 服务地址，我们在 Hue 的根目录下的 desktop/conf 文件夹下加 pseudo-distributed.ini 文件，然后我们对新增的文件添加如下内容：

```
vi pseudo-distributed.ini

[desktop]

http_host=10.211.55.28

http_port=8000

[hadoop]
```

```
[[hdfs_clusters]]

  [[default]]

    fs_defaultfs=hdfs://cluster1

    logical_name=cluster1

    webhdfs_url=http://10.211.55.26:50070/webhdfs/v1

    hadoop_conf_dir=/home/hadoop/hadoop-2.6.0/etc/hadoop

[beeswax]

  hive_server_host=10.211.55.17

  hive_server_port=10000

  # hive_conf_dir=/home/hive/warehouse

[hbase]

  hbase_clusters=(cluster1|10.211.55.26:9090)

  hbase_conf_dir=/home/hadoop/hbase-1.0.1/conf
```

关于 Hue 的详细和更多配置需要，大家可以参考官方给的知道文档，连接地址如下所示：

```
http://cloudera.github.io/hue/docs-3.8.0/manual.html
```

Hue 的集成就完成了，下面可以输入启动命令来查看，命令如下所示：

```
[hadoop@nna ~]$ /root/hue-3.7.1/build/env/bin/supervisor &
```

下面，在浏览器中输入对应的访问地址，这里我配置的 Port 是 8000，在第一次访问时，需要输入用户名和密码来创建一个 Administrator，这里需要在一步的时候需要注意下。