Apache Kylin2.3.1对接FusionInsight_HD_C80

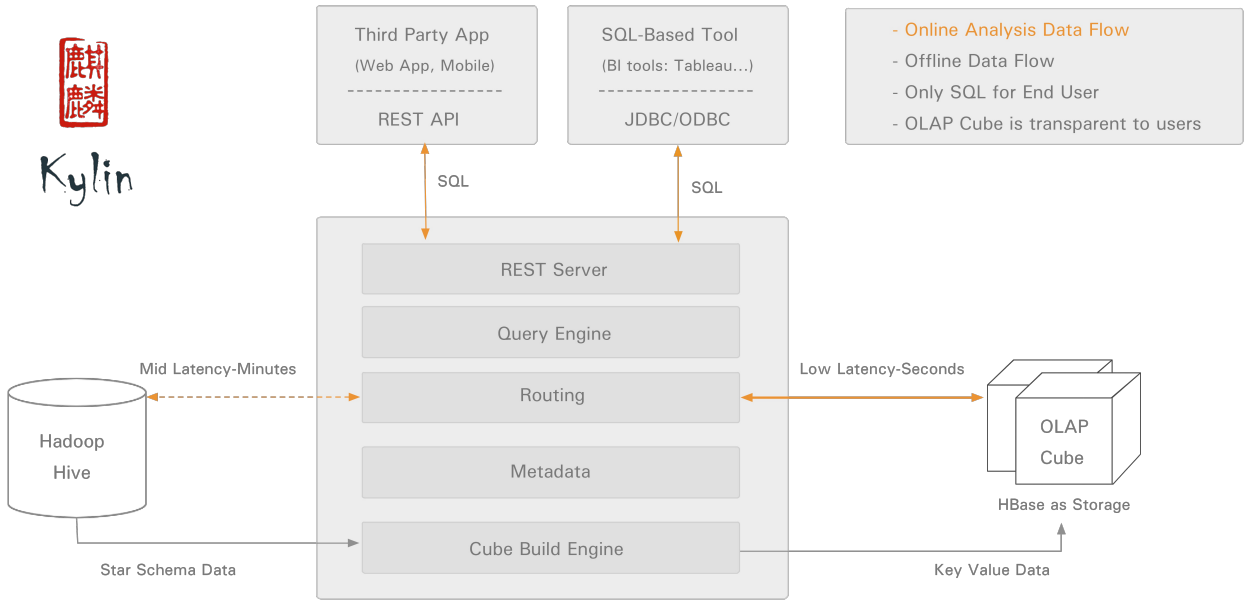
适用场景

> Apache Kylin 2.3.1 <-> FusionInsight HD V100R002C80SPC100

说明

Apache Kylin™是一个开源的分布式分析引擎，提供Hadoop之上的SQL查询接口及多维分析（OLAP）能力以支持超大规模数据，最初由eBay Inc. 开发并贡献至开源社区。它能在亚秒内查询巨大的Hive表。

Apache Kylin主要与FusionInsight的Hive和HBase进行对接



环境准备

- 修改/etc/hosts

  添加本机主机名解析

  ```
  172.16.52.86 kylin
  ```

- 配置NTP服务

  使用vi /etc/ntp.conf增加NTP服务的配置,时间与FusionInsight集群同步

  ```
  server 172.18.0.18 nomodify notrap nopeer noquery
  ```

  启动NTP服务

  ```
  service ntpd start
  chkconfig ntpd on
  ```

- 参考FusionInsight产品文档在Kylin节点安装FusionInsight客户端

  在FusionInsight Manager服务管理页面下载客户端，上传到kylin节点安装FusionInsight客户端到 `/opt/hadoopclient` 目录

  ```
  ./install.sh /opt/hadoopclient
  ```

- 安装JDK1.8

```
rpm -Uvh jdk-8u112-linux-x64.rpm
```

下载Kylin

Fusioninsight配套的HBase是1.3.0，Apache Kylin可直接下载apache-kylin-2.3.1-hbase1x-bin.tar.gz主版本二进制包，无需编译Apache kylin

下载解压Kylin

- 下载Kylin-2.3.1基于HBase1.x版本的二进制包， http://ftp.cuhk.edu.hk/pub/packages/apache.org/kylin/apache-kylin-2.3.1/apache-kylin-2.3.1-hbase1x-bin.tar.gz
- 上传apache-kylin-2.3.1-hbase1x-bin.tar.gz到Apache kylin节点的`/opt`目录
- 解压上一步骤的安装包

```
cd /opt
tar -zxvf apache-kylin-2.3.1-hbase1x-bin.tar.gz -C /opt
```

配置Kylin

配置环境变量

- 配置环境变量：`vi /etc/profile`，增加以下配置

```
export KYLIN_HOME=/opt/apache-kylin-2.3.1-bin
```

- 导入环境变量

```
source /etc/profile
```

- Kylin启动还需要配置HIVE_CONF、HCAT_HOME，使用`vi /opt/hadoopclient/Hive/component_env`，在文件最后增加

```
export HIVE_CONF=/opt/hadoopclient/Hive/config
export HCAT_HOME=/opt/hadoopclient/Hive/HCatalog
```

- 导入环境变量

```
source /opt/hadoopclient/bigdata_env
```

- 进行kerberos认证

```
kinit test
```

- Kylin检查环境设置：

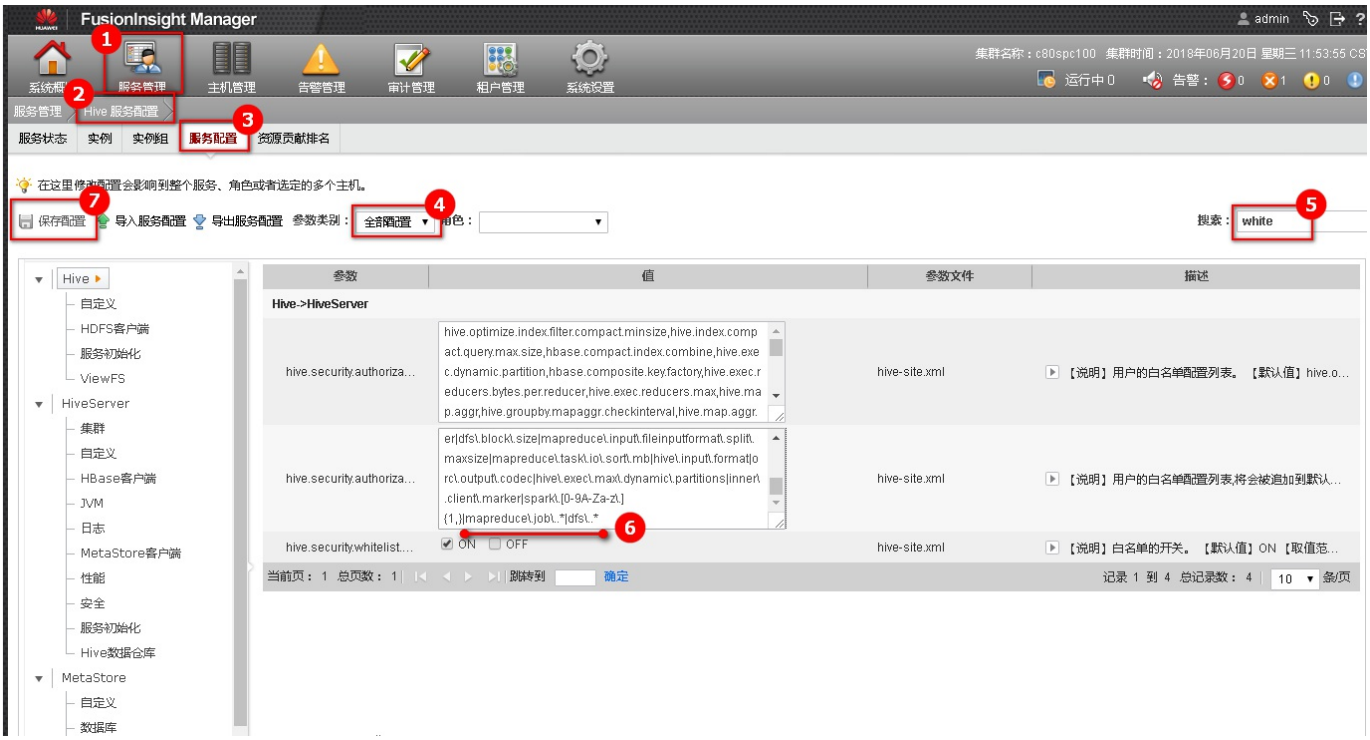
```
cd /opt/apache-kylin-2.3.1-bin/bin
./check-env.sh
```

```
[root@kylin bin]# ./check-env.sh
Retrieving hadoop conf dir...
KYLIN_HOME is set to /opt/apache-kylin-2.3.1-bin
[root@kylin bin]#
```

修改FusionInsight的Hive配置项

- 在hive.security.authorization.sqlstd.confwhitelist.append参数最后追加一下参数配置，保存配置，重启影响的服务

```
|mapreduce\.job\..*|dfs\..*
```



修改Kylin配置

- 获取Hive的JDBC字符串

  执行Beeline查看Hive的JDBC字符串

```
source bigdata_env
kinit test
beeline
```

```
[root@kylin hadoopclient]# source bigdata_env
[root@kylin hadoopclient]# kinit test
Password for test@HADOOP.COM:
[root@kylin hadoopclient]# beeline
Connecting to jdbc:hive2://172.21.3.101:24002,172.21.3.102:24002,172.21.3.103:24002/;serviceDiscoveryMode=zooKee
per;zooKeeperNamespace=hiveserver2;sasl.qop=auth-conf;auth=KERBEROS;principal=hive/hadoop.hadoop.com@HADOOP.COM
Debug is  true storeKey false useTicketCache true useKeyTab false doNotPrompt false ticketCache is null isInitia
tor true KeyTab is null refreshKrb5Config is false principal is null tryFirstPass is false useFirstPass is false
 storePass is false clearPass is false
Acquire TGT from Cache
Principal is test@HADOOP.COM
Commit Succeeded

Connected to: Apache Hive (version 1.2.1)
Driver: Hive JDBC (version 1.2.1)
Transaction isolation: TRANSACTION_REPEATABLE_READ
Beeline version 1.2.1 by Apache Hive
0: jdbc:hive2://172.21.3.101:21066/>
```

- 修改kylin.properties： `vi /opt/apache-kylin-2.3.1-bin/conf/kylin.properties`

  配置Hive client使用beeline：

```
kylin.source.hive.client=beeline
kylin.source.hive.beeline-shell=beeline
kylin.source.hive.beeline-params=-n root -u
'jdbc:hive2://172.21.3.101:24002,172.21.3.102:24002,172.21.3.103:24002/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=hiveserv
er2;sasl.qop=auth-conf;auth=KERBEROS;principal=hive/hadoop.hadoop.com@HADOOP.COM'
```

  JDBC字符串使用上一步骤获取的字符串

```
#### SOURCE ###
#
## Hive client, valid value [cli, beeline]
kylin.source.hive.client=beeline
#
## Absolute path to beeline shell, can be set to spark beeline instead of the default hive beeline on PATH
kylin.source.hive.beeline-shell=beeline
#
## Parameters for beeline client, only necessary if hive client is beeline
kylin.source.hive.beeline-params=-n root -u 'jdbc:hive2://172.21.3.101:24002,172.21.3.102:24002,172.21.3.103:24002/;se
rviceDiscoveryMode=zooKeeper;zooKeeperNamespace=hiveserver2;sasl.qop=auth-conf;auth=KERBEROS;principal=hive/hadoop.had
oop.com@HADOOP.COM'
#
## While hive client uses above settings to read hive table metadata,
```

> 注意：kylin.source.hive.beeline-params参数里面原有的
> `--hiveconf hive.security.authorization.sqlstd.confwhitelist.append='mapreduce.job.*|dfs.*'` 要去掉

- 修改Hive/HBase配置

  将/opt/hadoopclient/Hive/config/hivemetastore-site.xml中的配置合并到hive-site.xml

  将/opt/hadoopclient/HBase/hbase/conf/hbase-site.xml中的配置合并到/opt/apache-kylin-2.3.1-bin/conf/kylin_job_conf.xml

- Hive lib路径

  kylin的/opt/apache-kylin-2.3.1-bin/bin/find-hive-dependency.sh默认Hive lib路径为大数据集群中Hive的安装路径，需要修改为客户端路径

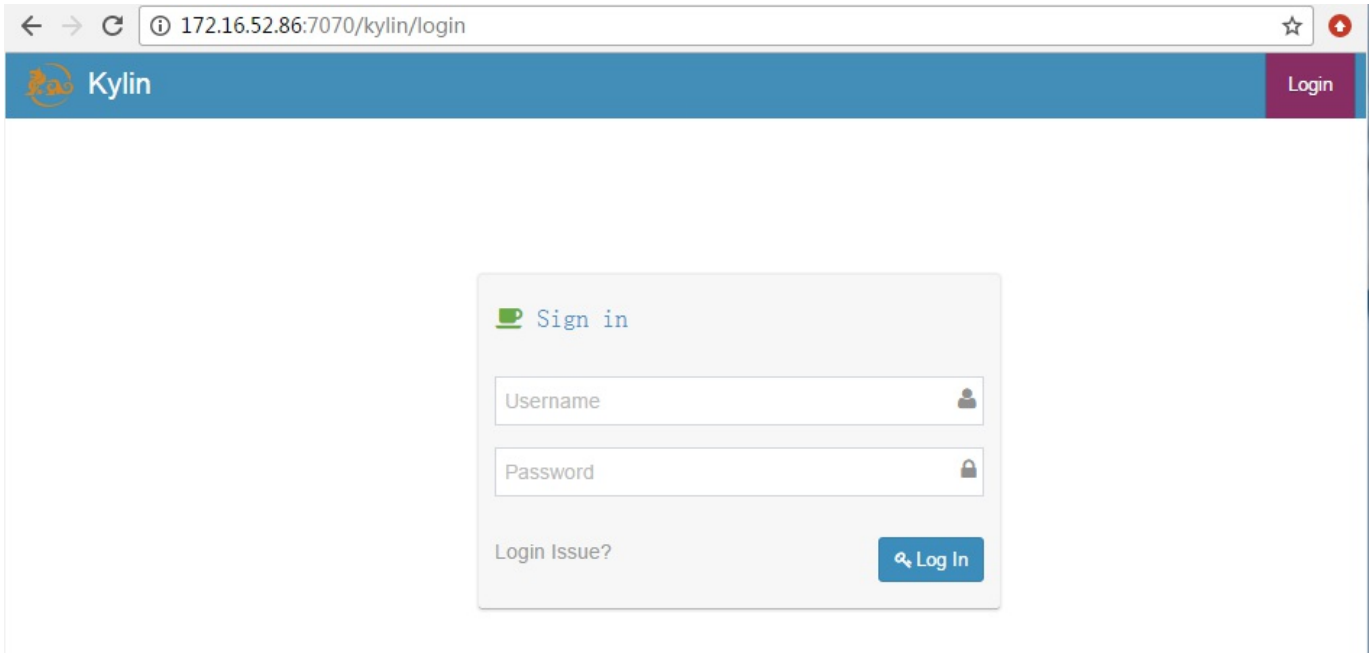
```
if [ -z "$HIVE_LIB" ]
then
    verbose "HIVE_LIB is not set, try to retrieve hive lib from hive_exec_path"
    hive_lib_dir="$(dirname $HCAT_HOME)"
else
    hive_lib_dir="$HIVE_LIB"
fi
hive_lib=`find -L ${hive_lib_dir} -name '*.jar' ! -name '*calcite*' -printf '%p:' | sed 's/:$//'`
```

启动Kylin

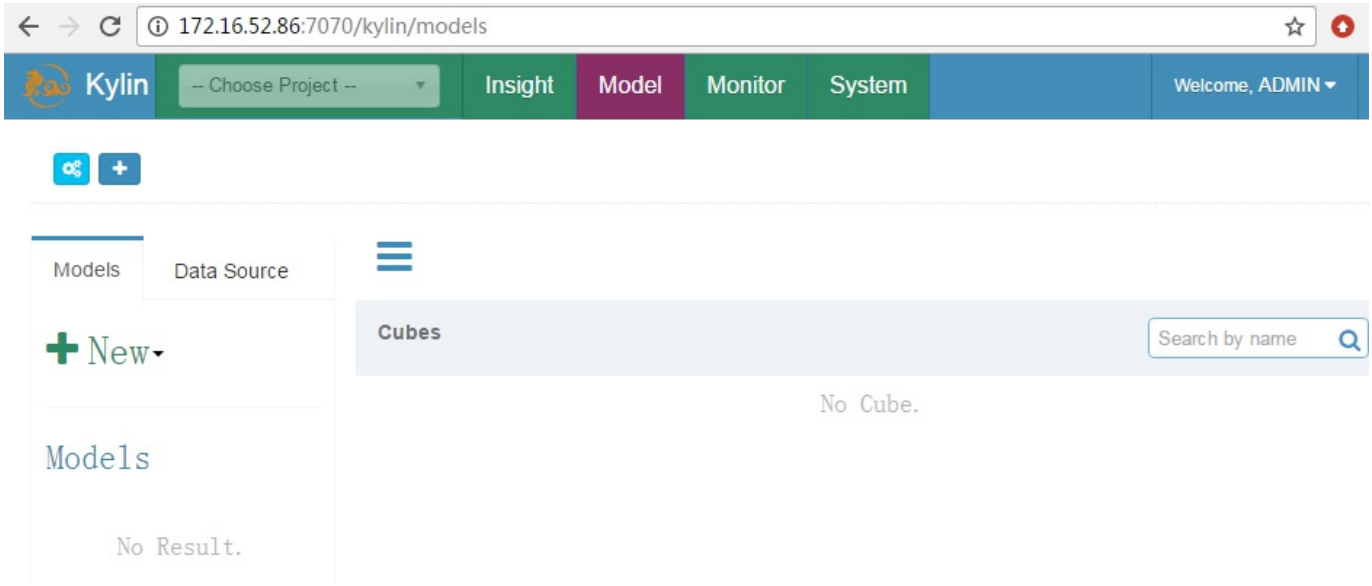
- 使用 `./kylin.sh start` 启动Kylin

```
[root@kylin bin]# ./kylin.sh start
Retrieving hadoop conf dir...
KYLIN_HOME is set to /opt/apache-kylin-2.3.1-bin
Retrieving hive dependency...
Retrieving hbase dependency...
```

```
A new Kylin instance is started by root. To stop it, run 'kylin.sh stop'
Check the log at /opt/apache-kylin-2.3.1-bin/logs/kylin.log
Web UI is at http://<hostname>:7070/kylin
```



输入默认用户名密码：ADMIN/KYLIN登陆



Demo测试

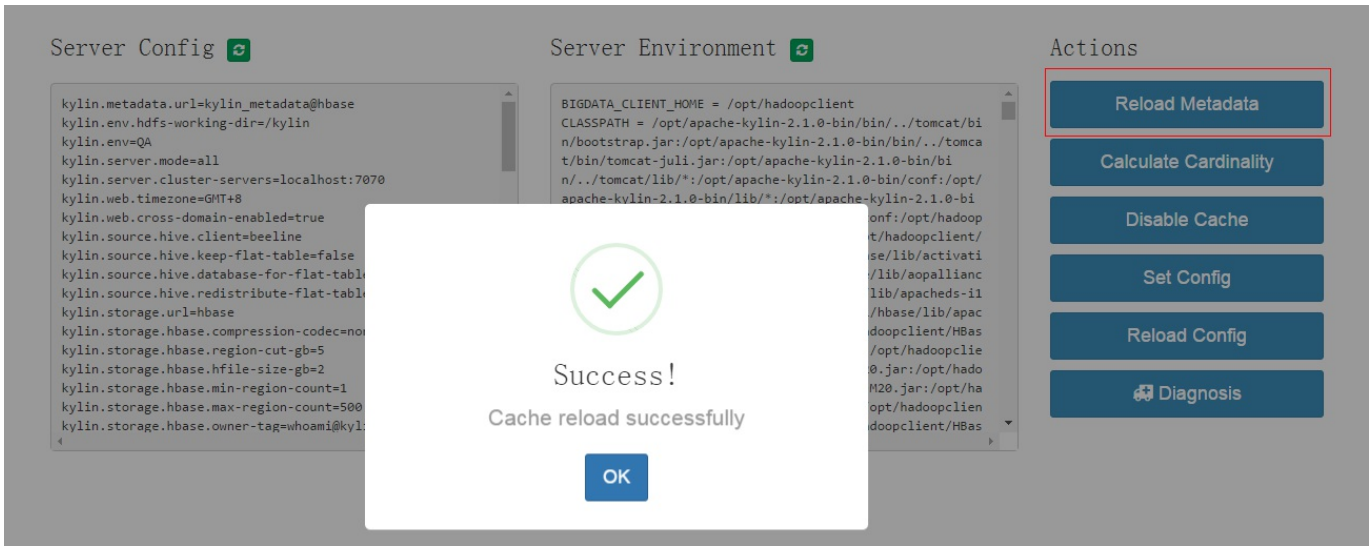
导入Demo数据

- 执行以下命令导入sample数据

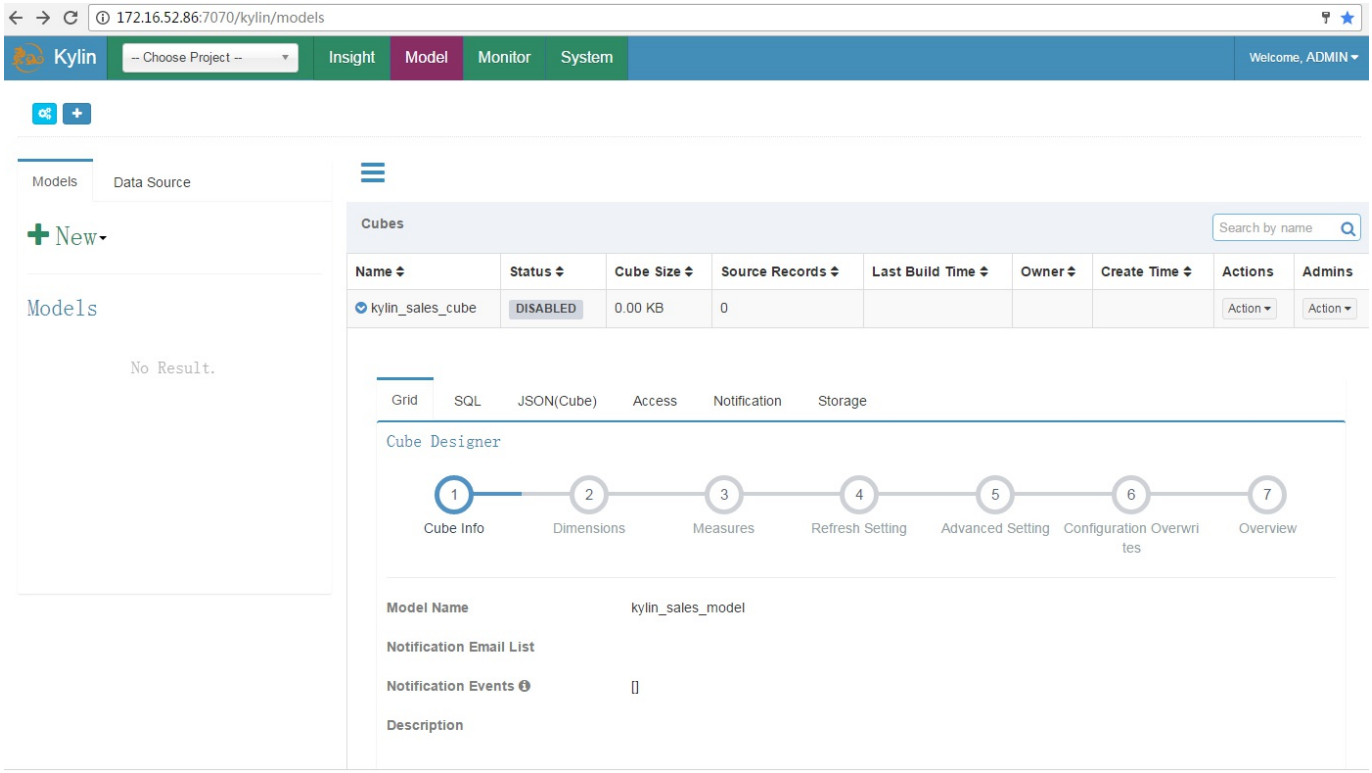
```
cd /opt/apache-kylin-2.3.1-bin/bin
./sample.sh
```

```
2018-06-20 11:26:35,651 INFO  [close-hbase-conn] hbase.HBaseConnection:137 : Closing HBase connections...
2018-06-20 11:26:35,651 INFO  [close-hbase-conn] client.ConnectionManager$HConnectionImplementation:2292 : Closing master protocol: MasterService
2018-06-20 11:26:35,655 INFO  [close-hbase-conn] client.ConnectionManager$HConnectionImplementation:1756 : Closing zookeeper sessionid=0x1400000686169fef
2018-06-20 11:26:35,657 INFO  [main-EventThread] zookeeper.ClientCnxn:614 : EventThread shut down for session: 0x1400000686169fef
2018-06-20 11:26:35,658 INFO  [close-hbase-conn] zookeeper.ZooKeeper:1325 : Session: 0x1400000686169fef closed
Sample cube is created successfully in project 'learn_kylin'.
Restart Kylin Server or click Web UI => System Tab => Reload Metadata to take effect
```

选择菜单 **System** -> **Actions** -> **Reload Metadata**

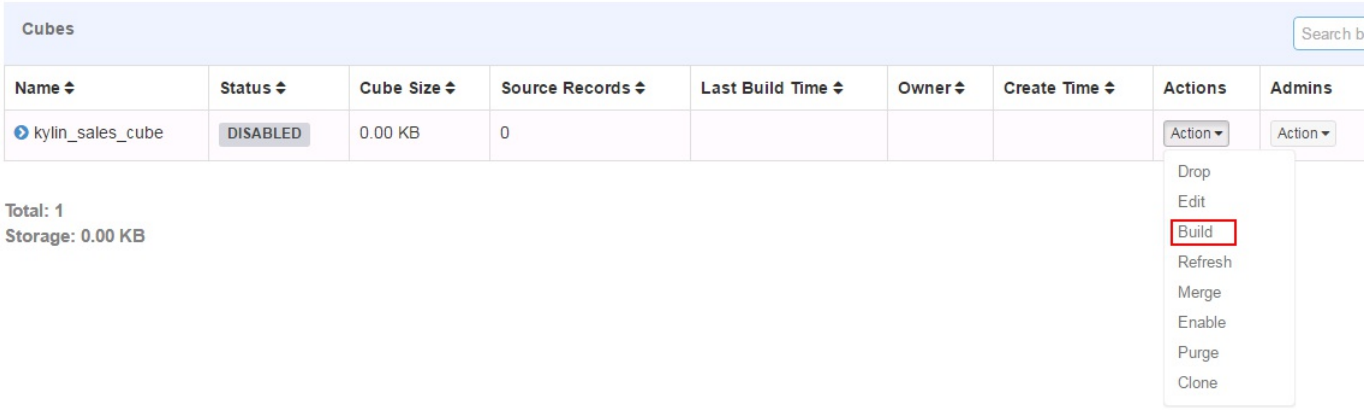


选择菜单 **System** -> **Model**

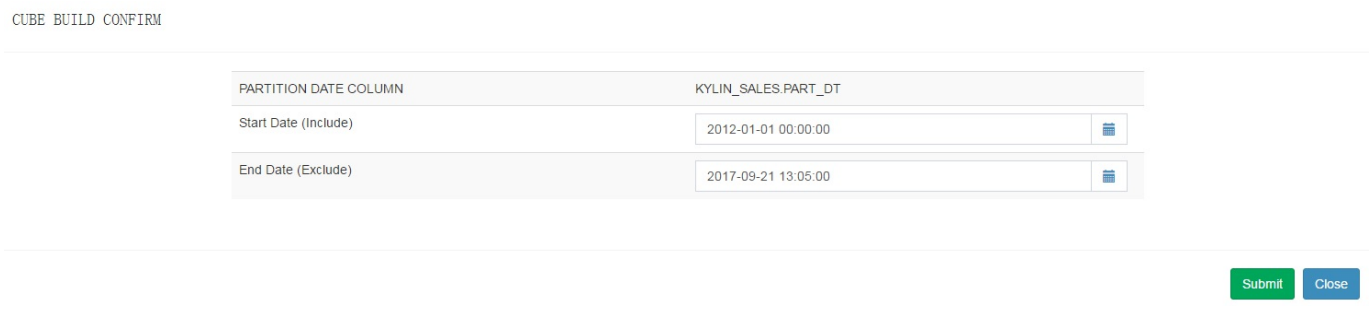


构建Cube

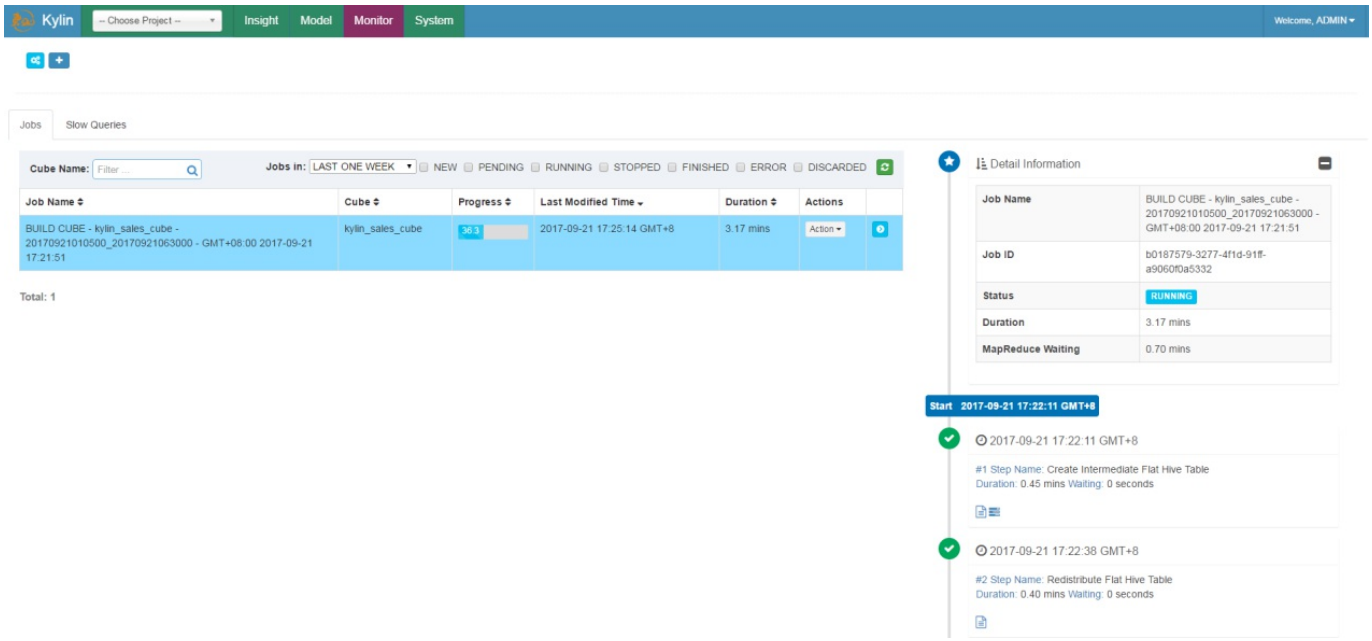
- 构建默认的kylin_sales_cube



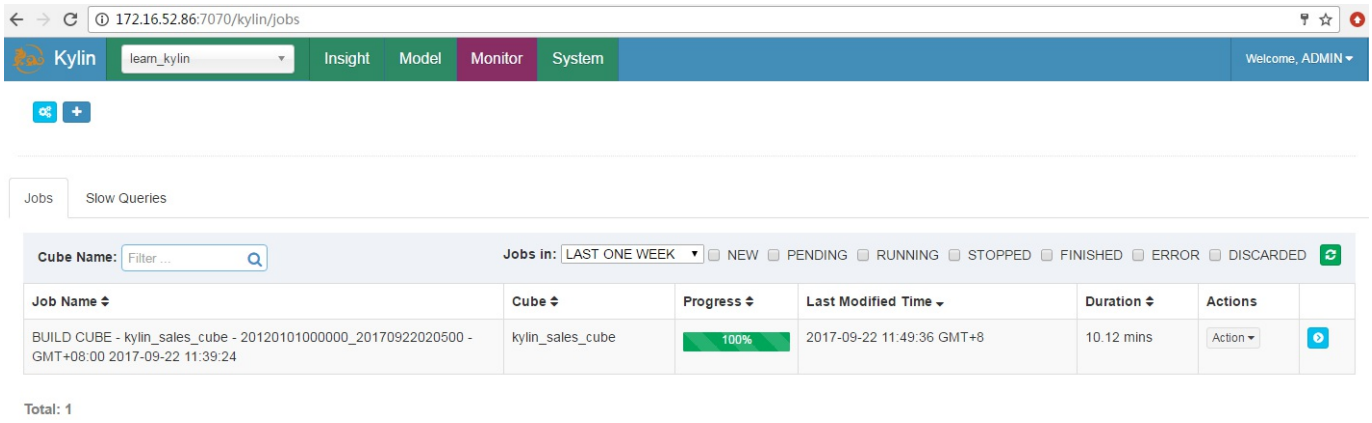
- 选择End Data（Exclude）时间：



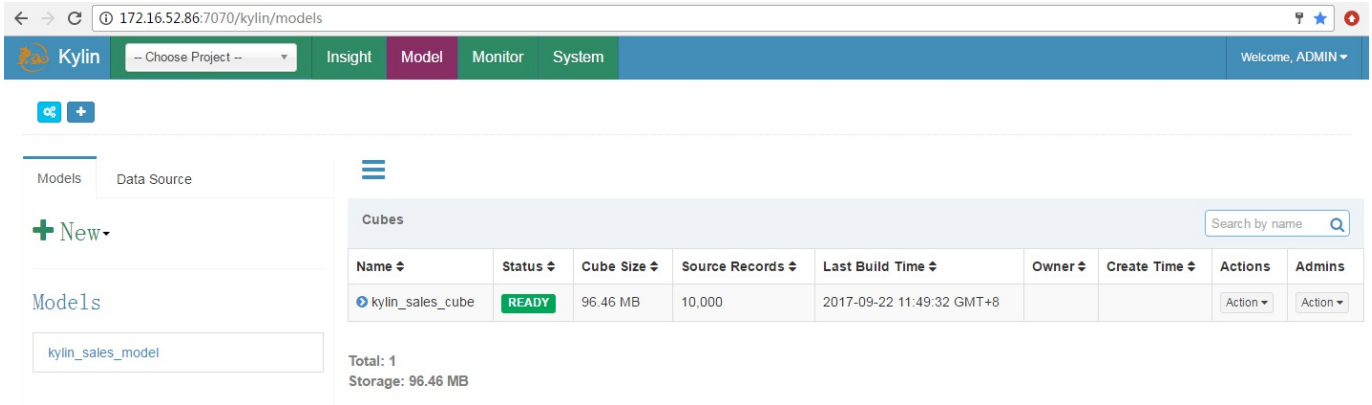
- 点击Monitor可以查看build状态：



- Build完成：

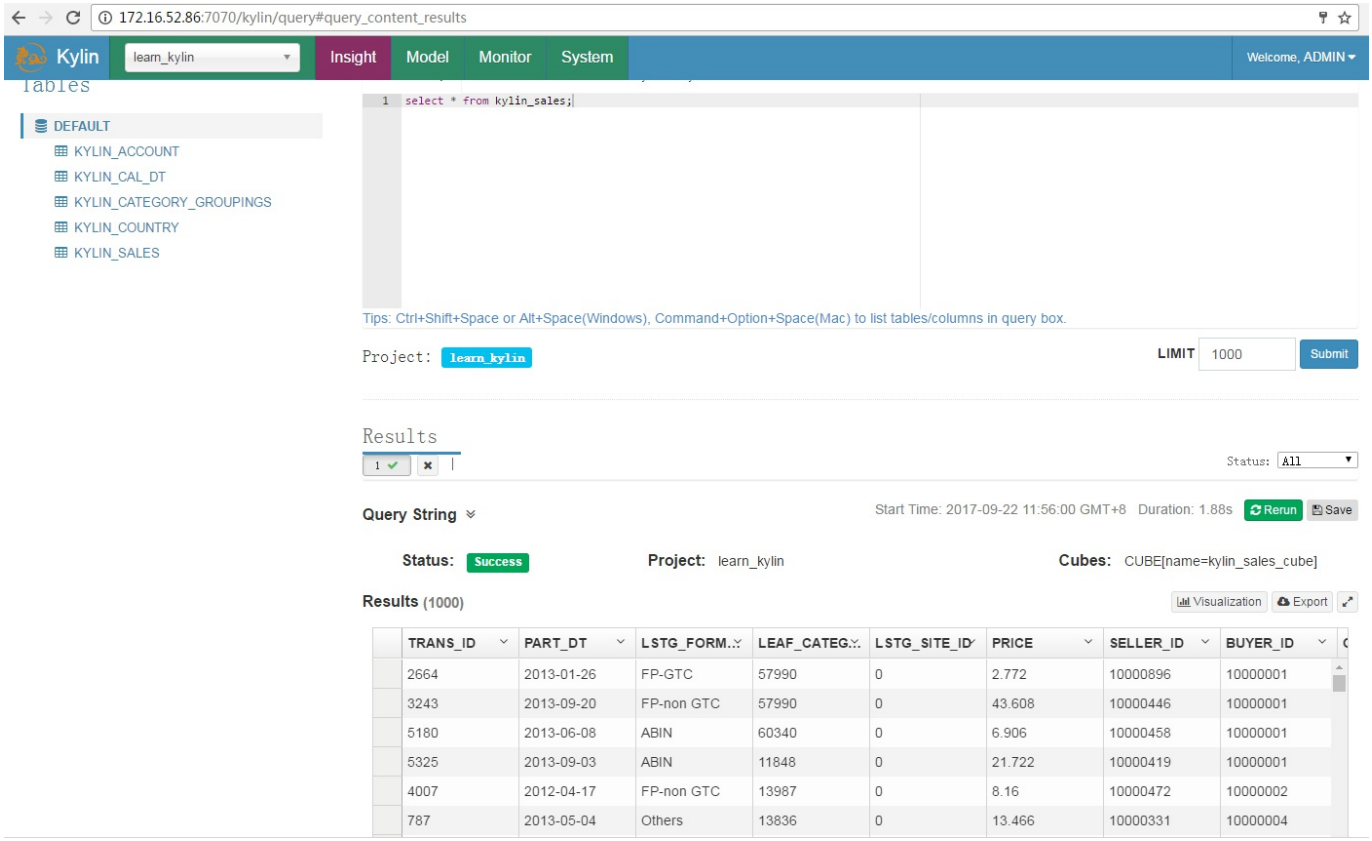


- Cube构建成功，状态变为READY



查询表数据

- 在Insight页面执行查询


172.16.52.86:7070/kylin/query#query_content_results
Kylin
learn_kylin
Insight
Model
Monitor
System
Welcome, ADMIN
Tables
DEFAULT
KYLIN_ACCOUNT
KYLIN_CAL_DT
KYLIN_CATEGORY_GROUPINGS
KYLIN_COUNTRY
KYLIN_SALES
1 select * from kylin_sales;
Tips: Ctrl+Shift+Space or Alt+Space(Windows), Command+Option+Space(Mac) to list tables/columns in query box.
Project: learn_kylin
LIMIT 1000
Submit
Results
Status: All
Query String
Start Time: 2017-09-22 11:56:00 GMT+8 Duration: 1.88s
Rerun
Save
Status: Success
Project: learn_kylin
Cubes: CUBE[name=kylin_sales_cube]
Results (1000)
Visualization
Export
TRANS_ID PART_DT LSTG_FORM.. LEAF_CATEG.. LSTG_SITE_ID PRICE SELLER_ID BUYER_ID
2664 2013-01-26 FP-GTC 57990 0 2.772 10000896 10000001
3243 2013-09-20 FP-non GTC 57990 0 43.608 10000446 10000001
5180 2013-06-08 ABIN 60340 0 6.906 10000458 10000001
5325 2013-09-03 ABIN 11848 0 21.722 10000419 10000001
4007 2012-04-17 FP-non GTC 13987 0 8.16 10000472 10000002
787 2013-05-04 Others 13836 0 13.466 10000331 10000004