

Apache Kylin2.1.0对接FusionInsight_HD_C70

适用场景

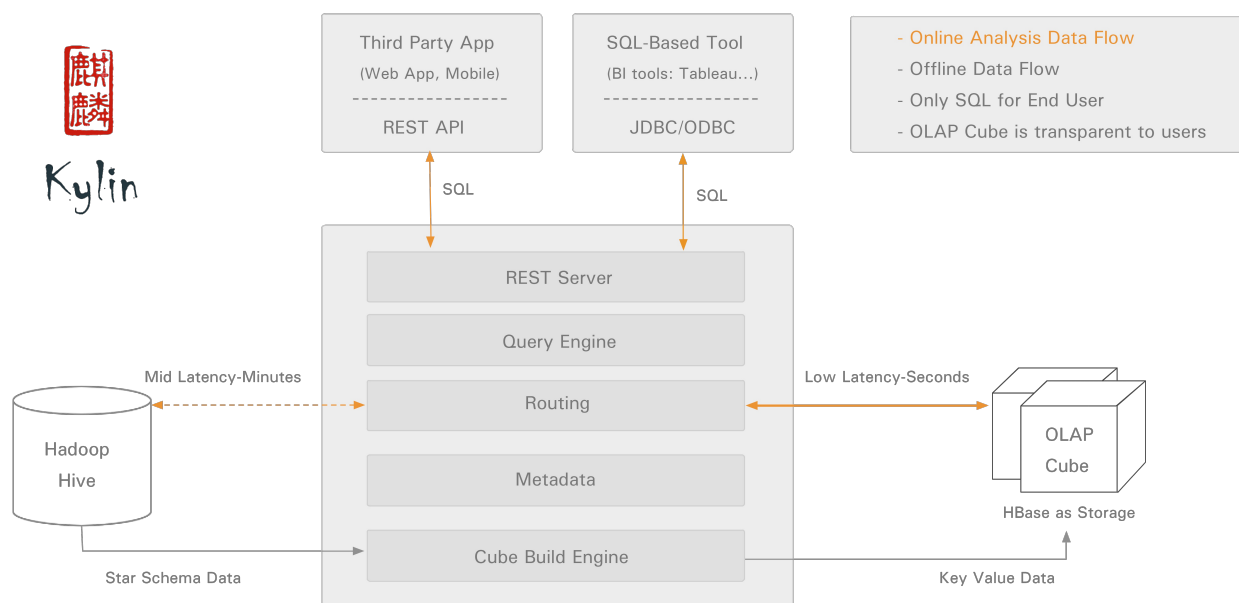
Apache Kylin 2.1.0 <-> FusionInsight HD V100R002C70SPC100

Apache Kylin 2.1.0 <-> FusionInsight HD V100R002C70SPC200

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

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

启动NTP服务

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

- 安装Hadoop client

在FusionInsight Manager服务管理页面下载客户端，上传到kylin安装主机，解压

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

- 安装JDK

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

编译Kylin

可直接下载的二进制文件的Kylin-2.1.0主版本是基于HBase1.1.1编译的，而FusionInsight使用的HBase版本是1.0.2，这两个版本部分类和方法不兼容，需要重新编译Kylin。

- 下载Kylin-2.1.0基于HBase1.1.1版本的源代码<https://github.com/apache/kylin/tree/2.1.x>得到kylin-2.1.x.zip
- 安装编译工具

安装maven:

```
wget http://apache.osuosl.org/maven/maven-3/3.3.9/binaries/apache-maven-3.3.9-bin.tar.gz
tar -xzf apache-maven-3.3.9-bin.tar.gz -C /opt/
```

修改profile文件 `vi /etc/profile` ,增加以下配置

```
export PATH=$PATH:/opt/apache-maven-3.3.9/bin
```

导入环境变量

```
source /etc/profile
```

执行 `mvn -v`

```
[root@kylin bin]# mvn -v
Apache Maven 3.3.9 (bb52d8502b132ec0a5a3f4c09453c07478323dc5; 2015-11-11T00:41:47+08:00)
Maven home: /opt/apache-maven-3.3.9
Java version: 1.8.0_112, vendor: Oracle Corporation
Java home: /opt/hadoopclient/JDK/jdk/jre
Default locale: en_US, platform encoding: UTF-8
OS name: "linux", version: "2.6.32-431.el6.x86_64", arch: "amd64", family: "unix"
```

- 安装git

```
yum install -y git
```

- 安装nodejs:

```
wget https://nodejs.org/dist/v6.10.0/node-v6.10.0-linux-x64.tar.xz --no-check-certificate
tar -xvf node-v6.10.0-linux-x64.tar.xz -C /opt/
```

修改profile文件 `vi /etc/profile` ,增加以下配置

```
export PATH=$PATH:/opt/apache-maven-3.3.9/bin:/opt/node-v6.10.0-linux-x64/bin
```

导入环境变量

```
source /etc/profile
```

执行 `npm -v`

```
[root@kylin bin]# npm -v
3.10.10
```

- 修改kylin源码

修改HiveMRInput.java

```
vi /opt/kylin-2.1.x/source-hive/src/main/java/org/apache/kylin/source/hive/HiveMRInput.java
```

```

private void redistributeTable(KylinConfig config, int numReducers) throws IOException {
    final HiveCmdBuilder hiveCmdBuilder = new HiveCmdBuilder();
    hiveCmdBuilder.overrideHiveProps(config.getHiveConfigOverride());
    hiveCmdBuilder.addStatement(getInitStatement());
    //hiveCmdBuilder.addStatement("set mapreduce.job.reduces=" + numReducers + ";\n");
    hiveCmdBuilder.addStatement("set hive.merge.mapredfiles=false;\n");
    hiveCmdBuilder.addStatement(getRedistributeDataStatement());
    final String cmd = hiveCmdBuilder.toString();

    stepLogger.log("Redistribute table, cmd: ");
    stepLogger.log(cmd);

    Pair<Integer, String> response = config.getCliCommandExecutor().execute(cmd, stepLogger);
    getManager().addJobInfo(getId(), stepLogger.getInfo());

    if (response.getFirst() != 0) {
        throw new RuntimeException("Failed to redistribute flat hive table");
    }
}

```

修改pom.xml

```
vi /opt/kylin-2.1.x/pom.xml
```

将HBase、Hive、Hadoop版本改成与FusionInsight HD对应的版本

```

<!-- Hadoop versions -->
<hadoop2.version>2.7.2</hadoop2.version>
<yarn.version>2.7.2</yarn.version>

<!-- Hive versions -->
<hive.version>1.2.1</hive.version>
<hive-hcatalog.version>1.2.1</hive-hcatalog.version>

<!-- HBase versions -->
<hbase-hadoop2.version>1.0.2</hbase-hadoop2.version>

<!-- Kafka versions -->
<kafka.version>0.10.1.0</kafka.version>

<!-- Spark versions -->
<spark.version>2.1.0</spark.version>
<kryo.version>4.0.0</kryo.version>

```

- 编译打包

```

unzip kylin-2.1.x.zip
cd kylin-2.1.x
sed -i "s/2.1.0-SNAPSHOT/2.1.0/g" `grep 2.1.0-SNAPSHOT -r1 *`
sh build/script/package.sh

```

```

[root@kylin script]# ./package.sh
Checking maven...
maven check passed
Checking git...
git check passed
Checking npm...
npm check passed

```

等待编译完成，得到Kylin二进制安装包

```

apache-kylin-2.1.0-bin/lib/
apache-kylin-2.1.0-bin/lib/kylin-coprocessor-2.1.0.jar
apache-kylin-2.1.0-bin/lib/kylin-job-2.1.0.jar
Package ready:dist/apache-kylin-2.1.0-bin.tar.gz
[root@zeeplin kylin-2.1.x]#

```

启动Kylin

解压二进制包

- 解压上一步骤生成的安装包

```
tar -xzf apache-kylin-2.1.0-bin.tar.gz -C /opt
```

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-2.1.0-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-2.1.0-bin/bin
./check-env.sh
```

```
[root@kylin210 bin]# ./check-env.sh
Retrieving hadoop conf dir...
KYLIN_HOME is set to /opt/new/apache-kylin-2.1.0-bin
```

修改Kylin配置

- 修改kylin.properties: `vi /opt/apache-kylin-2.1.0-bin/conf/kylin.properties`

配置Hive client使用beeline:

```
kylin.source.hive.client=beeline
kylin.source.hive.beeline-params=-n root -u
'jdbc:hive2://172.21.42.30:24002,172.21.42.31:24002,172.21.42.32:24002//;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=hiveserv
er2;sasl.qop=auth-conf;auth=KERBEROS;principal=hive/hadoop.hadoop.com@HADOOP.COM'
```

配置获取任务状态时使用kerberos 鉴权:

```
kylin.job.status.with.kerberos=true
```

- 去掉不允许修改的配置

FusionInsight不允许修改dfs.replication, mapreduce.job.split.metainfo.maxsize的参数, 需要注释掉Kylin所有配置文件中的相关参数, 否则构建Cube时会报如下错误:

```

Connected to: Apache Hive (version 1.3.0)
Driver: Hive JDBC (version 1.3.0)
Transaction isolation: TRANSACTION_REPEATABLE_READ
0: jdbc:hive2://162.1.61.42:21066/> No rows affected (0.347 seconds)
0: jdbc:hive2://162.1.61.42:21066/> Error: Error while processing statement: Cannot modify dfs.replication at runtime. It is not in list of params that are allowed to be
modified at runtime (state=42000,code=1)

```

需要修改以下文件:

- kylin_hive_conf.xml
- kylin_job_conf_inmem.xml
- kylin_job_conf.xml

- 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-2.1.0-bin/conf/kylin_job_conf.xml

- Hive lib路径

kylin的/opt/apache-kylin-2.1.0-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, 若Kylin安装在集群节点上不会有问題, 否则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 $(cat $HCAT_HOME))"
else
    hive_lib_dir="$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@kylin210 bin]# ./kylin.sh start
Retrieving hadoop conf dir...
KYLIN_HOME is set to /opt/apache-kylin-2.1.0-bin

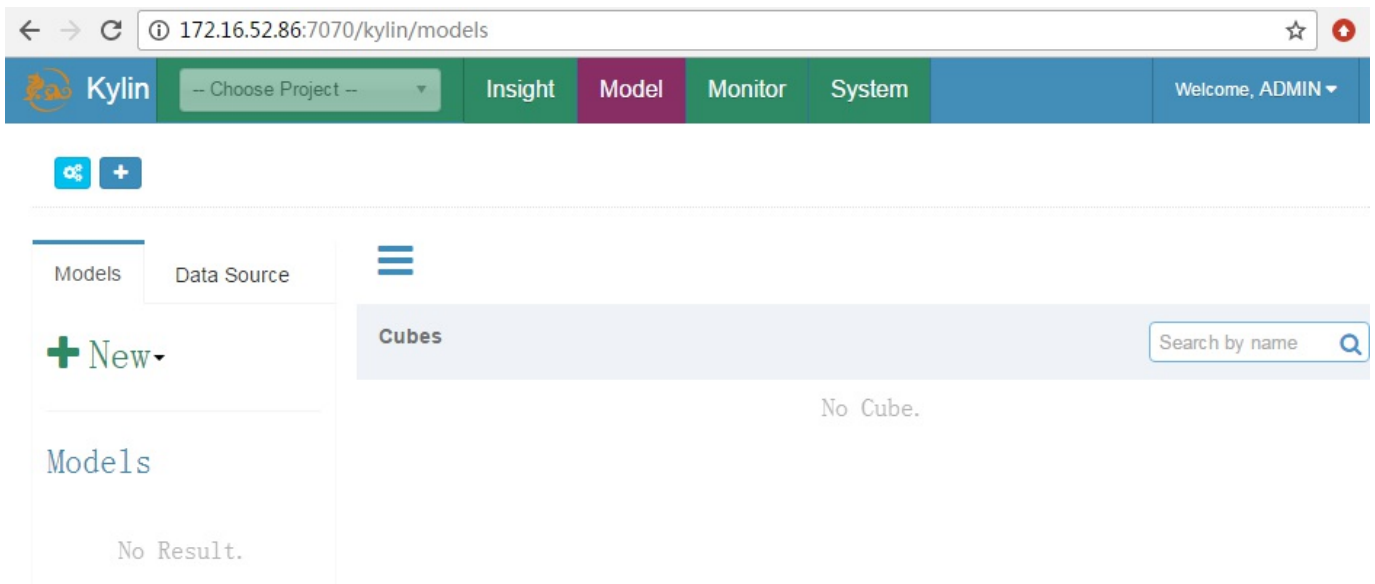
```

```

A new Kylin instance is started by root. To stop it, run 'kylin.sh stop'
Check the log at /opt/apache-kylin-2.1.0-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

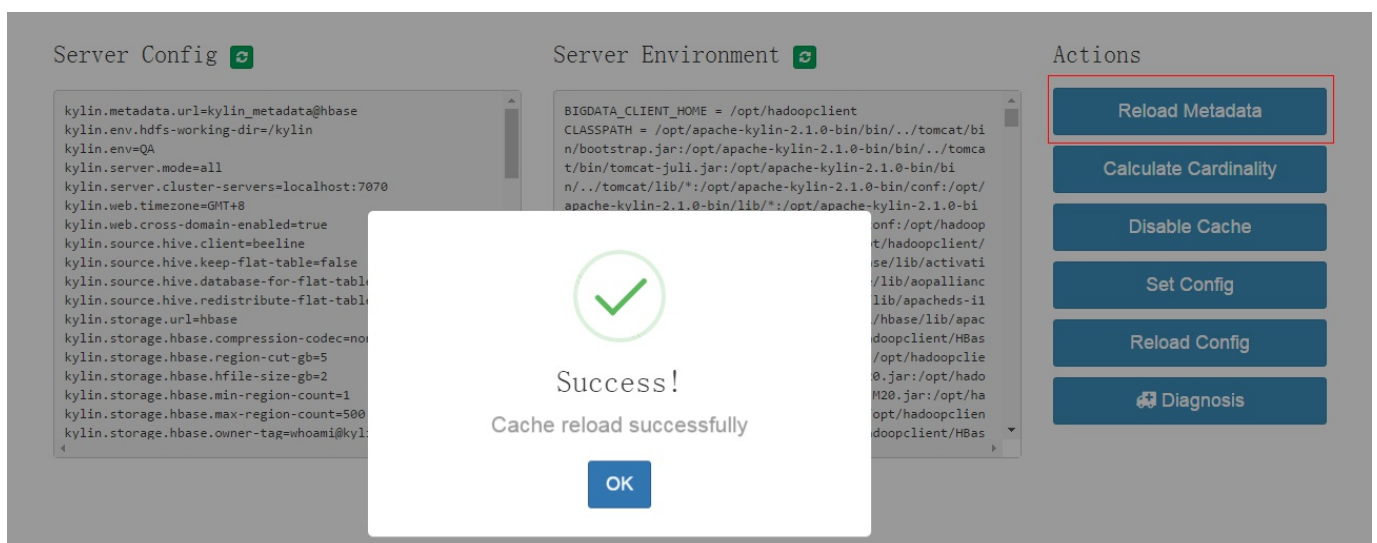
导入Demo数据

- 执行以下命令导入sample数据

```
cd /opt/apache-kylin-2.1.0-bin/bin
./sample.sh
```

```
2017-09-21 16:03:25,971 INFO [main-SendThread(172-21-3-102:24002)] zookeeper.Login:310 : successfully logged in.
2017-09-21 16:03:25,972 INFO [Thread-4] zookeeper.Login:139 : TGT refresh thread started.
2017-09-21 16:03:25,975 INFO [Thread-4] zookeeper.Login:328 : TGT valid starting at: Thu Sep 21 14:54:12 CST 2017
2017-09-21 16:03:25,975 INFO [Thread-4] zookeeper.Login:329 : TGT expires: Fri Sep 22 14:54:08 CST 2017
2017-09-21 16:03:25,976 INFO [Thread-4] zookeeper.Login:193 : TGT refresh sleeping until: Fri Sep 22 10:48:34 CST 2017
2017-09-21 16:03:25,976 INFO [main-SendThread(172-21-3-102:24002)] client.ZooKeeperSaslClient:327 : Client will use GSSAPI as SASL mechanism.
2017-09-21 16:03:26,025 INFO [main-SendThread(172-21-3-102:24002)] zookeeper.ClientCnxn:1327 : Opening socket connection to server 172-21-3-102/172.21.3.102:24002. Will attempt to SASL-authenticate using Login Context section 'Client'
2017-09-21 16:03:26,031 INFO [main-SendThread(172-21-3-102:24002)] zookeeper.ClientCnxn:1053 : Socket connection established, initiating session, client: /172.16.52.86:38466, server: 172-21-3-102/172.21.3.102:24002
2017-09-21 16:03:26,037 INFO [main-SendThread(172-21-3-102:24002)] zookeeper.ClientCnxn:1601 : Session establishment complete on server 172-21-3-102/172.21.3.102:24002, sessionId = 0x13002cf2d3943e9a, negotiated timeout = 90000
2017-09-21 16:03:27,183 INFO [hconnection-0x4b8729ff-shared--pool1-tl] ipc.AbstractRpcClient:817 : RPC Server Kerberos principal name for service=ClientService is hbase/hadoop.hadoop.com@HADOOP.COM
2017-09-21 16:03:27,258 DEBUG [main] hbase.HBaseConnection:283 : HTable 'kylin_metadata' already exists
2017-09-21 16:03:27,280 INFO [main] ipc.AbstractRpcClient:817 : RPC Server Kerberos principal name for service=MasterService is hbase/hadoop.hadoop.com@HADOOP.COM
2017-09-21 16:03:27,508 INFO [main] persistence.ResourceTool:163 : Copy from /opt/apache-kylin-2.1.0-bin/sample_cube/metadata to kylin_metadata@hbase
2017-09-21 16:03:27,629 INFO [Thread-1] client.ConnectionManager$HConnectionImplementation:2188 : Closing master protocol: MasterService
2017-09-21 16:03:27,632 INFO [Thread-1] client.ConnectionManager$HConnectionImplementation:1707 : Closing zookeeper sessionId=0x13002cf2d3943e9a
2017-09-21 16:03:27,635 INFO [main-EventThread] zookeeper.ClientCnxn:614 : EventThread shut down for session: 0x13002cf2d3943e9a
2017-09-21 16:03:27,635 INFO [Thread-1] zookeeper.ZooKeeper:1325 : Session: 0x13002cf2d3943e9a closed
Sample cube is created successfully in project 'learn_kylin'.
Restart Kylin Server or click Web UI => System Tab => Reload Metadata to take effect
[root@kylin210 bin]#
```

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



选择菜单 **System** -> **Model**

172.16.52.86:7070/kylin/models

Kylin

-- Choose Project --

Insight

Model

Monitor

System

Welcome, ADMIN

Models

Data Source

+ New

Models

No Result.

Cubes

Search by name

Name	Status	Cube Size	Source Records	Last Build Time	Owner	Create Time	Actions	Admins
kylin_sales_cube	DISABLED	0.00 KB	0				Action	Action

Grid

SQL

JSON(Cube)

Access

Notification

Storage

Cube Designer

1

 Cube Info

2

 Dimensions

3

 Measures

4

 Refresh Setting

5

 Advanced Setting

6

 Configuration Overwrites

7

 Overview

Model Name

 kylin_sales_model

Notification Email List

Notification Events

 []

Description

构建Cube

- 构建默认的kylin_sales_cube

Cubes

Search by name

Name	Status	Cube Size	Source Records	Last Build Time	Owner	Create Time	Actions	Admins
kylin_sales_cube	DISABLED	0.00 KB	0				Action	Action

Total: 1
Storage: 0.00 KB

Drop

Edit

Build

Refresh

Merge

Enable

Purge

Clone

- 选择End Data（Exclude）时间：

CUBE BUILD CONFIRM

PARTITION DATE COLUMN

KYLIN_SALES.PART_DT

Start Date (Include)

2012-01-01 00:00:00

End Date (Exclude)

2017-09-21 13:05:00

Submit

Close

- 点击Monitor可以查看build状态：

Kylin-- Choose Project --InsightModelMonitorSystemWelcome, ADMIN

JobsSlow Queries

Cube Name:FilterJobs in:LAST ONE WEEKNEWPENDINGRUNNINGSTOPPEDFINISHEDERRORDISCARDED

Job Name	Cube	Progress	Last Modified Time	Duration	Actions
BUILD CUBE - kylin_sales_cube - 20170921010500_20170921063000 - GMT+08:00 2017-09-21 17:21:51	kylin_sales_cube	3.3	2017-09-21 17:25:14 GMT+8	3.17 mins	Action

Total: 1

Detail Information

Job NameBUILD CUBE - kylin_sales_cube - 20170921010500_20170921063000 - GMT+08:00 2017-09-21 17:21:51

Job IDb0187579-3277-4f1d-91f8-a9060f0a5332

StatusRUNNING

Duration3.17 mins

MapReduce Waiting0.70 mins

Start 2017-09-21 17:22:11 GMT+8

2017-09-21 17:22:11 GMT+8

#1 Step Name: Create Intermediate Flat Hive Table
Duration: 0.45 mins Waiting: 0 seconds

2017-09-21 17:22:38 GMT+8

#2 Step Name: Redistribute Flat Hive Table
Duration: 0.40 mins Waiting: 0 seconds

- Build完成:

172.16.52.86:7070/kylin/jobs

Kylinlearn_kylinInsightModelMonitorSystemWelcome, ADMIN

JobsSlow Queries

Cube Name:FilterJobs in:LAST ONE WEEKNEWPENDINGRUNNINGSTOPPEDFINISHEDERRORDISCARDED

Job Name	Cube	Progress	Last Modified Time	Duration	Actions
BUILD CUBE - kylin_sales_cube - 20120101000000_20170922020500 - GMT+08:00 2017-09-22 11:39:24	kylin_sales_cube	100%	2017-09-22 11:49:36 GMT+8	10.12 mins	Action

Total: 1

- Cube构建成功，状态变为READY

172.16.52.86:7070/kylin/models

Kylin-- Choose Project --InsightModelMonitorSystemWelcome, ADMIN

ModelsData Source

+ New

Models

kylin_sales_model

Cubes

Search by name

Name	Status	Cube Size	Source Records	Last Build Time	Owner	Create Time	Actions	Admins
kylin_sales_cube	READY	96.46 MB	10,000	2017-09-22 11:49:32 GMT+8			Action	Action

Total: 1
Storage: 96.46 MB

查询表数据

- 在Insight页面执行查询

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

LIMIT1000

Submit

Results

1

✓

✕

|

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	