

# Apache Kylin对接FusionInsight

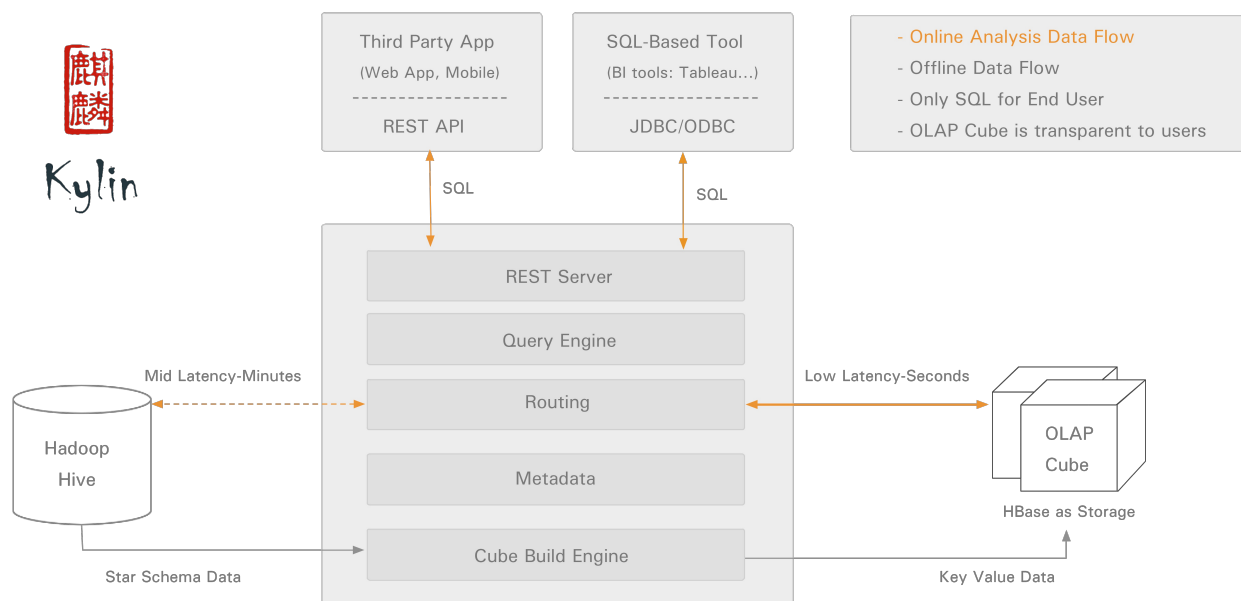
## 适用场景

Apache Kylin 1.6.0 (基于HBase1.0.2的分支版本: yang21-hbase102) <-> FusionInsight HD V100R002C60U20

## 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

```
162.1.115.89 kylin
```

- 配置NTP服务

使用vi /etc/ntp.conf增加NTP服务的配置

```
server 162.2.200.200 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

```
cd FusionInsight_V100R002C60U20_Services_ClientConfig/
./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-1.6.0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，需要配套1.0.2的HBase重新编译Kylin。

- 下载Kylin-1.6.0基于HBase1.0.2版本的源码<https://github.com/apache/kylin/tree/yang21-hbase102>得到kylin-yang21-hbase102.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
```

- 编译打包

```
unzip kylin-yang21-hbase102.zip
cd kylin-yang21-hbase102/
sed -i "s/1.6.0-SNAPSHOT/1.6.0/g" `grep 1.6.0-SNAPSHOT -rl *`
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-1.6.0-bin/lib/
apache-kylin-1.6.0-bin/lib/kylin-coprocessor-1.6.0.jar
apache-kylin-1.6.0-bin/lib/kylin-jdbc-1.6.0.jar
apache-kylin-1.6.0-bin/lib/kylin-joh-1.6.0.jar
Package ready: dist/apache-kylin-1.6.0-bin.tar.gz
[root@kylin kylin-yang21-hbase102]# █
```

## 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-1.6.0-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-1.6.0-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_cn
```

```
[root@kylin ~]# kinit test_cn
Password for test_cn@HADOOP.COM:
[root@kylin ~]#
[root@kylin ~]# klist
Ticket cache: FILE:/tmp/krb5cc_0
Default principal: test_cn@HADOOP.COM

Valid starting    Expires          Service principal
03/14/17 22:50:30 03/15/17 22:50:30  krbtgt/HADOOP.COM@HADOOP.COM
renew until 03/14/17 23:20:30
```

- Kylin检查环境设置：

```
cd /opt/apache-kylin-1.6.0-bin/bin
./check-env.sh
```

```
[root@kylin bin]# ./check-env.sh
KYLIN_HOME is set to /opt/apache-kylin-1.6.0-bin
```

## 修改Kylin配置

- 修改kylin.properties: `vi /opt/apache-kylin-1.6.0-bin/conf/kylin.properties`

配置Hive client使用beeline:

```
kylin.hive.client=beeline
kylin.hive.beeline.params=-n root -u
'jdbc:hive2://162.1.93.103:24002,162.1.93.102:24002,162.1.93.101:24002;/serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=hiveserv
er2;sasl.qop=auth-conf;auth=KERBEROS;principal=hive/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.splitmetainfo.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)
```

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-1.6.0-bin/conf/kylin\_job\_conf.xml

- Hive lib路径

kylin的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误, 否则需要修改为客户端路径。

编辑find-hive-dependency.sh: `vi /opt/apache-kylin-1.6.0-bin/bin/find-hive-dependency.sh`

```
hive_lib=`find -L "$(dirname $HCAT_HOME)" -name '*.jar' ! -name '*calcite*' -printf '%p:' | sed 's:/:$//'`
```

```
#hive_lib=`find -L "$(dirname $hive exec_path)" -name '*.jar' ! -name '*calcite*' -printf '%p:' | sed 's:/:$//'`
hive_lib=`find -L "$(dirname $HCAT_HOME)" -name '*.jar' ! -name '*calcite*' -printf '%p:' | sed 's:/:$//'`
hive_dependency=${hive_conf_path}:${hive_lib}:${hcatalog}
verbose "hive dependency: $hive_dependency"
export hive_dependency
```


## 启动Kylin


- 使用 `./kylin.sh start` 启动Kylin


```
[root@kylin bin]# ./kylin.sh start
KYLIN_HOME is set to /opt/apache-kylin-1.6.0-bin
Retrieving hive dependency...
KYLIN_JVM_SETTINGS is -Xms1024M -Xmx4096M -Xss1024K -XX:MaxPermSize=128M -verbose:gc -XX:+PrintGCDetails -XX:+PrintGCDateSt
amps -Xloggc:/opt/apache-kylin-1.6.0-bin/logs/kylin.gc.11305 -XX:+UseGCLogFileRotation -XX:NumberOfGCLogFiles=10 -XX:GCLogF
ileSize=64M


A new Kylin instance is started by root. To stop it, run 'kylin.sh stop'
Check the log at /opt/apache-kylin-1.6.0-bin/logs/kylin.log
Web UI is at http://<hostname>:7070/kylin
```

← → ↻ ⓘ 不安全 | 162.1.115.89:7070/kylin/login ☆ ⋮

 Kylin Help ▾ Login

 Sign in








[Login Issue?](#) Log In


输入默认用户名密码：ADMIN/KYLIN登陆

← → ↻ ⓘ 162.1.115.89:7070/kylin/models

 Kylin -- Choose Project -- ▾ Insight Model Monitor System


 

Models Data Source

 New ▾

## Models

No Result.



Cubes

## Demo测试

### 导入Demo数据

- 执行以下命令导入sample数据

```
cd /opt/apache-kylin-1.6.0-bin/bin
./sample.sh
```

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

Kylin

learn\_kylin

InsightModelMonitorSystem

HelpWelcome, 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	Streaming
kylin_sales_cube	DISABLED	0.00 KB	0				Action	Action	false

GridSQLJSON(Cube)AccessNotificationStorage

Cube Designer

1234567

Cube InfoDimensionsMeasuresRefresh SettingAdvanced SettingConfiguration OverwritessOverview

Model Name

kylin\_sales\_model

Cube Name

kylin\_sales\_cube\_desc

## 构建Cube

- 选择learn\_kylin工程，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

- Cube构建成功，状态变为READY

Cubes									
Search by name									
Name	Status	Cube Size	Source Records	Last Build Time	Owner	Create Time	Actions	Admins	Streaming
kylin_sales_cube	READY	25.21 MB	10,000	2017-03-13 19:57:08 GMT+8			Action	Action	false

## 查询表数据

- 在Insight页面执行查询



# Tables



DEFAULT

- KYLIN\_CAL\_DT
- KYLIN\_CATEGORY\_GROUPINGS
- KYLIN\_SALES

New Query

Saved Queries

1 select \* from kylin\_sales;

## Results

1 ✓ ✕ |

Query String

Start Time: 2017-03-14 22:41:18 GMT+8 Duration: 0.000s

Status: Success

Project: learn\_kylin

Cubes: kylin\_sales\_cube

Results (9945)

	PART_DT	LEAF_CATEG...	LSTG_SITE_ID	LSTG_FORMA...	USER_ID	REGION	PRICE	SELLER_ID	M
	2012-01-01	1349	0	Others	ADMIN	Shanghai	0.7227		
	2012-01-01	1357	0	Others	ANALYST	Beijing	16.7271		
	2012-01-01	24760	0	FP-non GTC	ADMIN	Shanghai	13.281		
	2012-01-01	53064	0	FP-non GTC	ADMIN	Shanghai	13.7727		
	2012-01-01	57990	0	Others	ADMIN	Shanghai	26.259		
	2012-01-01	63861	0	ABIN	ADMIN	Shanghai	76.1393		
	2012-01-01	63889	0	Others	MODELER	Hongkong	4.2157		
	2012-01-01	80287	0	FP-non GTC	ANALYST	Beijing	37.0823		
	2012-01-01	80287	0	FP-non GTC	ADMIN	Shanghai	37.0823		