



APACHE KYLIN – STOCKS STREAMING CUBE

Bin Jiang

03/18/2017



Models

Data Source

+ New

Models

streaming_stocks_model



Cubes

Search by name

Name	Status	Cube Size	Source Records	Last Build Time	Owner	Create Time	Actions	Admins	Streaming
streaming_stocks_cube	DISABLED	0.00 KB	0		ADMIN	2017-03-16 11:04:59 GMT+8	Action	Action	true

Total: 1
Storage: 0.00 KB

Tables

DEFAULT

Model Designer



Model Name *

streaming_stocks_model

Description

Next →

Tables

DEFAULT

Model Designer



Fact Table *

DEFAULT.STREAMING_STOCKS_TABLE

+ Add Lookup Table

Prev Next

Tables

DEFAULT

Model Designer



Select dimension columns

ID	Table Name	Columns
1	DEFAULT.STREAMING_STOCKS_TABLE	<div> YEAR_START X QUARTER_START X MONTH_START X WEEK_START X DAY_START X </div> <div> HOUR_START X MINUTE_START X E X T X </div>

← Prev Next →

Tables



DEFAULT

Model Designer



Select measure columns

CP_FIX X L_CUR X C_FIX X

Prev Next

- ID(varchar(256))
- T(varchar(256))
- E(varchar(256))
- L(varchar(256))
- L_FIX(varchar(256))
- L_CUR(double)
- S(date)
- LTT(varchar(256))
- LT(varchar(256))
- LT_DTS(timestamp)
- C(varchar(256))
- C_FIX(double)
- CP(varchar(256))
- CP_FIX(double)
- CCOL(varchar(256))
- PCLS_FIX(varchar(256))
- EL(varchar(256))

Columns

Filter ...

ID ^	Name ^	Data Type ^	Cardinality ^	Comment ^
1	ID	varchar(256)		
2	T	varchar(256)		
3	E	varchar(256)		
4	L	varchar(256)		
5	L_FIX	varchar(256)		
6	L_CUR	double		
7	S	date		
8	LTT	varchar(256)		
9	LT	varchar(256)		
10	LT_DTS	timestamp		
11	C	varchar(256)		
12	C_FIX	double		
13	CP	varchar(256)		

20	ELT	varchar(256)
21	EC	date
22	EC_FIX	date
23	ECP	date
24	ECP_FIX	date
25	ECCOL	varchar(256)
26	DIV	varchar(256)
27	YLD	varchar(256)
28	YEAR_START	date
29	QUARTER_START	date
30	MONTH_START	date
31	WEEK_START	date
32	DAY_START	date
33	HOUR_START	timestamp
34	MINUTE_START	timestamp

Models Data Source

Tables

- ID(varchar(256))
- T(varchar(256))
- E(varchar(256))
- L(varchar(256))
- L_FIX(varchar(256))
- L_CUR(double)
- S(date)
- LTT(varchar(256))
- LT(varchar(256))
- LT_DTS(timestamp)
- C(varchar(256))
- C_FIX(double)
- CP(varchar(256))
- CP_FIX(double)
- CCOL(varchar(256))
- PCLS_FIX(varchar(256))
- EL(varchar(256))

Table Schema:STREAMING_STOCKS_TABLE

Columns Extend Information Streaming Cluster

NAME STREAMING_STOCKS_TABLE

Hive DATABASE DEFAULT

SNAPSHOT TIME

LOCATION

INPUT FORMAT

OUTPUT FORMAT

OWNER

TOTAL FILE NUMBER

TOTAL FILE SIZE

PARTITIONED

PARTITION COLUMNS

- ID(varchar(256))
- T(varchar(256))
- E(varchar(256))
- L(varchar(256))
- L_FIX(varchar(256))
- L_CUR(double)
- S(date)
- LTT(varchar(256))
- LT(varchar(256))
- LT_DTS(timestamp)
- C(varchar(256))
- C_FIX(double)
- CP(varchar(256))
- CP_FIX(double)
- CCOL(varchar(256))
- PCLS_FIX(varchar(256))
- EL(varchar(256))

Kafka Setting

Topic kylinstock

Cluster-1

ID	Host	Port
1	sandbox.hortonworks.com	9092

Parser Setting

Parser Name org.apache.kylin.source.kafka.TimedJsonStreamParser

Parser Properties tsColName=lt_dts;tsParser=org.apache.kylin.source.kafka.DateTimeParser;tsPattern=yyyy-MM-dd'T'HH:mm:ss'Z'

Advanced Setting

Streaming Table And Cluster Info

Need to input streaming source record here, will detect the source schema and create a table schema for streaming.

JSON

```
1 { "id": "304466804484872", "t": "GOOG",
2  "e": "NASDAQ", "l": "852.12", "l_fix": "852.12", "l_cur": "852.12", "s": "2", "lts": "4:00PM EDT", "lt": "Mar 17, 4:00PM EDT", "lt_dts": "2017-03-17T16:00:02Z", "c": "+3.34", "c_fix": "3.34", "cp": "0.39", "cp_fix": "0.39", "ccol": "chg", "pcls_fix": "848.78", "el": "852.15", "el_fix": "852.15", "el_cur": "852.15", "elt": "Mar 17, 7:30PM EDT", "ec": "+0.03", "ec_fix": "0.03", "ecp": "0.00", "ecp_fix": "0.00", "eccol": "chg", "div": "", "yld": "" }
```



1. Choose 'timestamp' type column for streaming table.
2. By default, system will choose 'Default' as database, you can specify database like this 'database.table'
3. derived time dimensions are calculated from timestamp field to help analysis against different time granularities.

Table Name*

database.table

	Column	Column Type	Comment
<input checked="" type="checkbox"/>	id	varchar(256)	
<input checked="" type="checkbox"/>	t	varchar(256)	
<input checked="" type="checkbox"/>	e	varchar(256)	
<input checked="" type="checkbox"/>	l	date	
<input checked="" type="checkbox"/>	l_fix	date	
<input checked="" type="checkbox"/>	l_cur	date	
<input checked="" type="checkbox"/>	s	date	

+ New

Models

streaming_stocks_model

Cube Designer



Model Name *

streaming_stocks_model

Cube Name ⓘ *

streaming_stocks_cube

Notification Email List

Comma Separated

Notification Events ⓘ

ERROR X DISCARDED X SUCCEED X |

Description

Next →

+ New

Models

streaming_stocks_model

Cube Designer



+ Add Dimension Auto Generator Filter ...

ID	Name	Table Name	Type			Actions
1	DAY_START	DEFAULT.STREAMING_STOCKS_TABLE	normal	Column	DAY_START	
2	HOURL_START	DEFAULT.STREAMING_STOCKS_TABLE	normal	Column	HOURL_START	
3	MINUTE_START	DEFAULT.STREAMING_STOCKS_TABLE	normal	Column	MINUTE_START	
4	TICKER	DEFAULT.STREAMING_STOCKS_TABLE	normal	Column	T	
5	EXCHANGE	DEFAULT.STREAMING_STOCKS_TABLE	normal	Column	E	

Prev Next

New

Models

streaming_stocks_model

Cube Designer



Name	Expression	Parameters	Return Type	Actions
COUNT	COUNT	Value:1, Type:constant	bigint	
MAX_PCT_CHANGE	MAX	Value:CP_FIX, Type:column	double	
MIN_PCT_CHANGE	MIN	Value:CP_FIX, Type:column	double	
MAX_PRICE	MAX	Value:L_CUR, Type:column	double	
MIN_PRICE	MIN	Value:L_CUR, Type:column	double	
MAX_PRICE_CHANGE	MAX	Value:C_FIX, Type:column	double	

Models

streaming_stocks_model



Aggregation Groups

Visit [aggregation group](#) for more about aggregation group.

ID Aggregation Groups

1

Includes

MINUTE_START X HOUR_START X DAY_START X T X E X

Mandatory Dimensions

Select Column...

Hierarchy Dimensions

DAY_START X HOUR_START X MINUTE_START X

New Hierarchy+

Joint Dimensions

New Joint+

New Aggregation Group+

Rowkeys i

ID	Column	Encoding	Length	Shard By	
1	MINUTE_START	time	0	false	-
2	HOUR_START	time	0	false	-
3	DAY_START	date	0	false	-
4	T	dict	0	false	-
5	E	dict	0	false	-

New Rowkey Column+

Advanced Dictionaries i

+ New

Models

streaming_stocks_model

Cube Designer



Model Name streaming_stocks_model

Cube Name streaming_stocks_cube

Fact Table DEFAULT.STREAMING_STOCKS_TABLE

Lookup Table 0

Dimensions 5

Measures 7

Description



Jobs

Slow Queries

Cube Name:

Jobs in: LAST ONE WEEK ☐ NEW ☐ PENDING ☐ RUNNING ☐ FINISHED ☐ ERROR ☐ DISCARDED

Job Name ↕	Cube ↕	Progress ↕	Last Modified Time ▼	Duration ↕	Actions	
streaming_stocks_cube - 140_150 - BUILD - GMT+08:00 2017-03-18 09:55:10	streaming_stocks_cube	100%	2017-03-18 10:06:11 GMT+8	10.38 mins	Action ▼	
streaming_stocks_cube - 135_140 - BUILD - GMT+08:00 2017-03-18 09:40:09	streaming_stocks_cube	100%	2017-03-18 09:51:01 GMT+8	10.20 mins	Action ▼	
streaming_stocks_cube - 0_135 - BUILD - GMT+08:00 2017-03-18 09:04:26	streaming_stocks_cube	100%	2017-03-18 09:14:05 GMT+8	9.27 mins	Action ▼	
streaming_stocks_cube - 120_130 - BUILD - GMT+08:00 2017-03-18 08:42:02	streaming_stocks_cube	ERROR	2017-03-18 08:54:01 GMT+8	11.22 mins	Action ▼	
streaming_stocks_cube - 120_125 - BUILD - GMT+08:00 2017-03-18 07:59:07	streaming_stocks_cube	75.00%	2017-03-18 08:12:24 GMT+8	148.15 mins	Action ▼	
streaming_stocks_cube - 110_120 - BUILD - GMT+08:00 2017-03-18 07:50:12	streaming_stocks_cube	E	2017-03-18 07:53:05 GMT+8	2.30 mins	Action ▼	

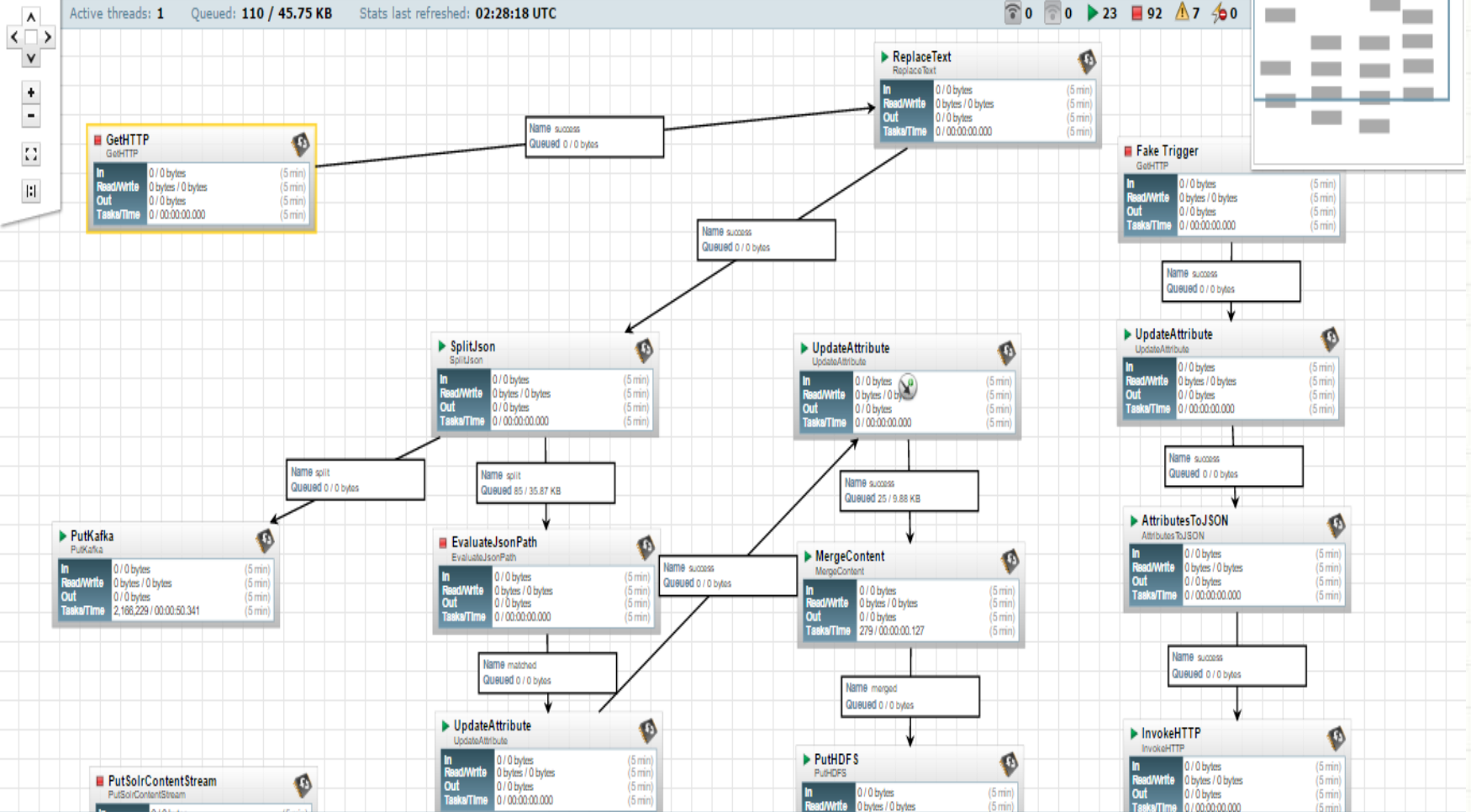
Hortonworks DataFlow

POWERED BY APACHE NIFI Help About



NiFi Flow » Stocks Dashboard

Active threads: 1 Queued: 110 / 45.75 KB Stats last refreshed: 02:28:18 UTC



Hortonworks DataFlow

Nifi in Ambari

[NiFi Flow](#) » [Stocks Dashboard](#)













Active threads: 0 Queued: 110 / 45.7

Configure Processor

Settings Scheduling **Properties** Comments

Required field

[+ New property](#)

Property	Value	
URL	 http://finance.google.com/finance/info?client=...	
Filename	 \${UUID()}	
SSL Context Service	 No value set	
Username	 No value set	
Password	 No value set	
Connection Timeout	 30 sec	
Data Timeout	 30 sec	
User Agent	 No value set	
Accept Content-Type	 No value set	
Follow Redirects	 false	
Proxy Host	 No value set	

Cancel

Apply

Nifi in Ambari

Hortonworks DataFlow

POWERED BY APACHE NIFI

NiFi in Ambari

Search

NiFi Flow » Stocks Dashboard

Active threads: 2 Queued: 110 / 45.7

GetHTTP

PutKafka

UpdateAttribute

PutHDFS

InvokeHTTP

PutSolrContentStream

Processor Details

Settings Scheduling Properties Comments

Required field

Property	Value
Known Brokers	sandbox.hortonworks.com:9092
Topic Name	kylinstock
Partition Strategy	Round Robin
Partition	No value set
Kafka Key	No value set
Delivery Guarantee	Best Effort
Message Delimiter	No value set
Max Buffer Size	5 MB
Max Record Size	1 MB
Communications Timeout	30 secs
Batch Size	200

Ok

root@sandt xroot@sandt xApache Kylin xApache Kylin xKylin xAmbari - Sa xroot@sandt xNiFi Flow xfinance.goo x

localhost:9090/nifi/

star

Nifi in Ambari

HelpAbout

Hortonworks DataFlow

POWERED BY APACHE NIFI

Search

NiFi Flow » Stocks Dashboard

Active threads: 1 Queued: 110 / 45.7

GetHTTP

0 / 0 bytes (5 min)

0 bytes / 0 bytes (5 min)

0 / 0 bytes (5 min)

0 / 00:00:00.000 (5 min)

PutKafka

0 / 0 bytes (5 min)

0 bytes / 0 bytes (5 min)

0 / 0 bytes (5 min)

2,197,469 / 00:00:51.622 (5 min)

UpdateAttribute

0 / 0 bytes (5 min)

0 bytes / 0 bytes (5 min)

0 / 0 bytes (5 min)

0 / 00:00:00.000 (5 min)

PutHDFS

0 / 0 bytes (5 min)

0 bytes / 0 bytes (5 min)

0 / 0 bytes (5 min)

0 / 00:00:00.000 (5 min)

InvokeHTTP

0 / 0 bytes (5 min)

0 bytes / 0 bytes (5 min)

0 / 0 bytes (5 min)

0 / 00:00:00.000 (5 min)

Processor Details

SettingsSchedulingPropertiesComments

Required field

Property	Value
HTTP Method	PUT
Remote URL	http://localhost:7070/kylin/api/cubes/streaming_st...
SSL Context Service	No value set
Connection Timeout	5 secs
Read Timeout	15 secs
Include Date Header	True
Follow Redirects	True
Attributes to Send	No value set
Basic Authentication Username	ADMIN
Basic Authentication Password	Sensitive value set
Proxy Host	No value set

Ok

Nifi in Ambari

10:30 PM 2017-03-17

Tables

No Result.

New Query Saved Queries Query History

```
1 select E as exchange, T as ticker, count(*) as transaction, max(L_CUR) as max_price, min(L_CUR) as min_price from streaming_stocks_table
2 group by E, T
3
```

Tips: Ctrl+Shift+Space or Alt+Space(Windows), Command+Option+Space(Mac) to list tables/columns in query box.

Project: streaming_stocks

LIMIT 50000

Submit

Results

1 x |

Status: All

Query String

Start Time: 2017-03-18 10:30:53 GMT+8

Rerun

Save

Results

1 ⚠️ | 2 ✅ |

Status: All

Query String

Start Time: 2017-03-18 11:44:20 GMT+8 Duration: 1.21s Rerun Save

Status: Success

Project: streaming_stocks

Cubes: streaming_stocks_cube

Results (5)

Visualization Export

E	T	TRANSACTION	MAX_PRICE	MIN_PRICE	
NASDAQ	GOOG	30	852.12	847.2	
NASDAQ	SBUX	30	55.78	54.54	
NASDAQ	AAPL	30	140.69	139.91	
NASDAQ	HDP	30	10.04	9.71	
NYSE	GS	30	248.22	243.94	

root@sandl Xroot@sandl XApache Kylin XApache Kylin XKylin XAmbari - Sa Xroot@sandl XNiFi Flow Xfinance.goo X

localhost:7070/kylin/projects

Kylin

streaming_stocks

Insight

Model

Monitor











System

Help

Welcome, ADMIN

+Project

Projects

Name ^	Owner	Description	Create Time	Actions
streaming_stocks	ADMIN		2017-03-16 10:17:16 GMT+8	 
streaming_sales	ADMIN		2017-03-16 08:58:48 GMT+8	 
sales	ADMIN		2017-03-14 09:49:40 GMT+8	 
learn_kylin				 
DEMO	ADMIN		2017-03-12 00:51:51 GMT+8	 

Total: 5

Apache Kylin | Apache Kylin Community

Windows Taskbar

10:34 PM 2017-03-17

Server Config

```
kylin.hbase.region.cut=5
kylin.job.cubing.inmem.sampling.percent=100
kylin.job.remote.cli.port=22
kylin.job.concurrent.max.limit=10
kylin.rest.servers=localhost:7070
kylin.query.security.enabled=true
kylin.job.mapreduce.mapper.input.rows=1000000
mail.enabled=false
kylin.web.hadoop=
kylin.job.status.with.kerberos=false
kylin.query.mem.budget=3221225472
kylin.job.remote.cli.password=
mail.password=
ldap.server=ldap://ldap_server:389
kylin.owner=whoami@kylin.apache.org
kylin.hdfs.working.dir=/root/kylin
kylin.query.coprocessor.mem.gb=3
crossdomain.enable=true
```

Server Environment

```
TERM = xterm
HBASE_CLASSPATH_PREFIX = /opt/apache-kylin-1.6.0-hbase1.
x-bin/bin/./tomcat/bin/bootstrap.jar:/opt/apache-kylin-
1.6.0-hbase1.x-bin/bin/./tomcat/bin/tomcat-juli.jar:/op
t/apache-kylin-1.6.0-hbase1.x-bin/bin/./tomcat/lib/*:/o
pt/apache-kylin-1.6.0-hbase1.x-bin/conf:/opt/apache-kyli
n-1.6.0-hbase1.x-bin/lib/*:/opt/apache-kylin-1.6.0-hbase
1.x-bin/tool/*:/opt/apache-kylin-1.6.0-hbase1.x-bin/ext/
*:
JAVA_HOME = /usr/lib/jvm/java
ATLAS_HOST = sandbox.hortonworks.com
HBASE_HOME = /usr/hdp/2.4.0.0-169/hbase
HBASE_ENV_INIT = true
SSH_CLIENT = 10.0.2.15 38260 22
MAIL = /var/spool/mail/root
HOSTNAME = sandbox.hortonworks.com
PWD = /opt/apache-kylin-1.6.0-hbase1.x-bin/bin
HBASE_MASTER_OPTS = -Xmx1024m -XX:PermSize=128m -XX:M
```

Actions

Reload Metadata

Calculate Cardinality

Disable Cache

Set Config

Diagnosis

Success
Server environment get successfully

Success
Server config get successfully

Models
Data Source

+ New

Models

DEMO_MODEL
kylin_sales_model
sales_model
streaming_sales_model
streaming_stocks_model

Cubes

Search by name

Name	Status	Cube Size	Source Records	Last Build Time	Owner	Create Time	Actions	Admins	Streaming
streaming_stock_s_cube	READY	12.00 KB	150	2017-03-18 11:40:46 GMT+8	ADMIN	2017-03-16 11:04:59 GMT+8	Action	Action	true
streaming_sales_cube	READY	769.00 KB	986	2017-03-16 09:54:01 GMT+8	ADMIN	2017-03-16 09:12:17 GMT+8	Action	Action	true
sales_cube	READY	2.49 MB	164,558	2017-03-16 08:50:15 GMT+8	ADMIN	2017-03-14 12:10:16 GMT+8	Action	Action	false
kylin_sales_cube	READY	24.65 MB	10,000	2017-03-13 11:44:54 GMT+8			Action	Action	false
DEMO_CUBE	READY	9.00 KB	359,169	2017-03-13 12:14:31 GMT+8	ADMIN	2017-03-12 01:09:48 GMT+8	Action	Action	false