

How to Use New Kylin Streaming

[Architecture](#)

[Environment Setup](#)

[Onboard Real-time Cube](#)

[Create new Streaming Table](#)

[Create Model](#)

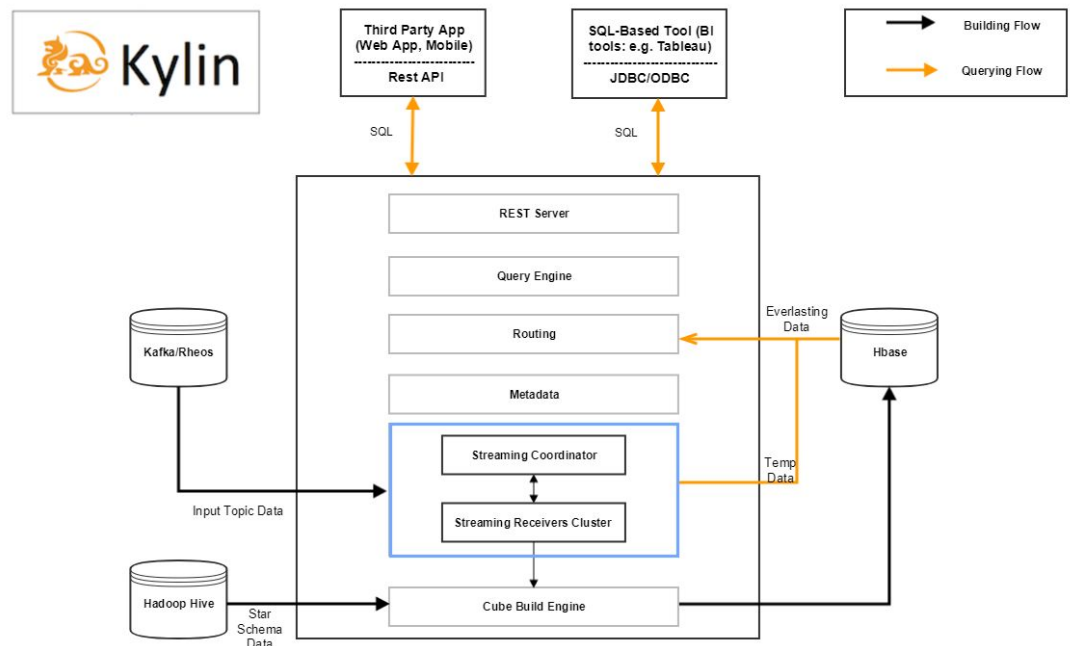
[Create Cube](#)

[Start Consuming](#)

[Monitor](#)

Architecture

The design doc can be found in jira: <https://issues.apache.org/jira/browse/KYLIN-3654>



Environment Setup

To use new kylin streaming, you need to setup the streaming cluster by following steps:

Step1: Start streaming receivers

Streaming receiver is also running in hadoop environment like kylin server, so you must ensure that the hadoop client env is ready, and the metadata store is set to the same as Kylin server. The receiver can be started using the cmd: \$KYLIN_HOME/bin/kylin.sh streaming start

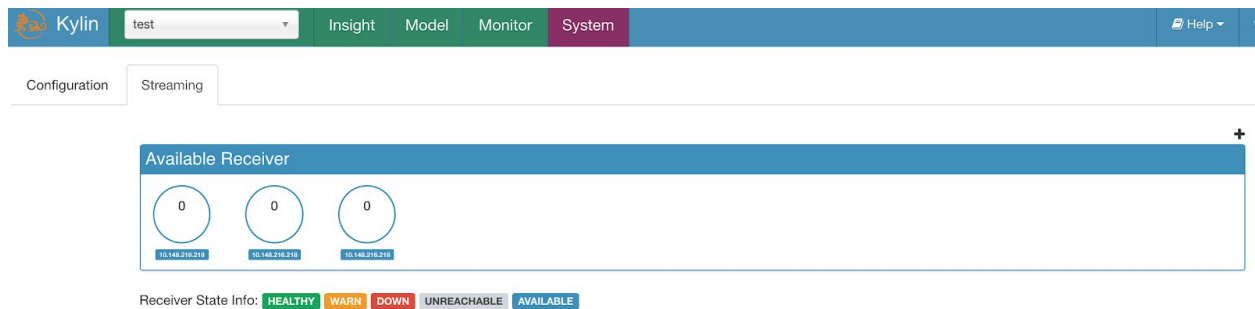
Step2: Ensure that there is at least one streaming coordinator in the cluster

Currently we use kylin server as streaming coordinator, so if there is a Kylin server mode is set to “all” or “stream_coordinator”, then it is ok.

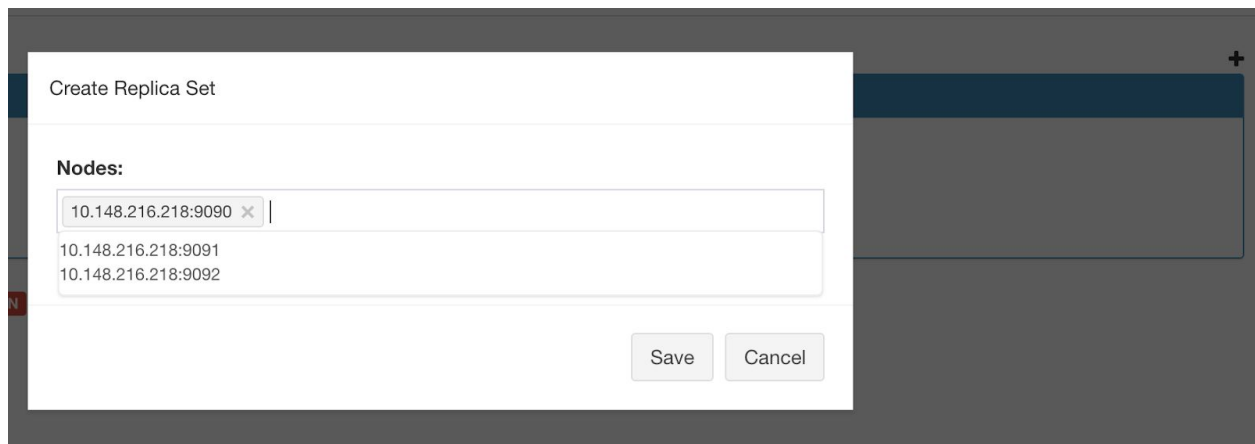
Step3: Create replicaSet

We use replicaSet to ensure the streaming receiver HA capability. The replicaSet can be created via:

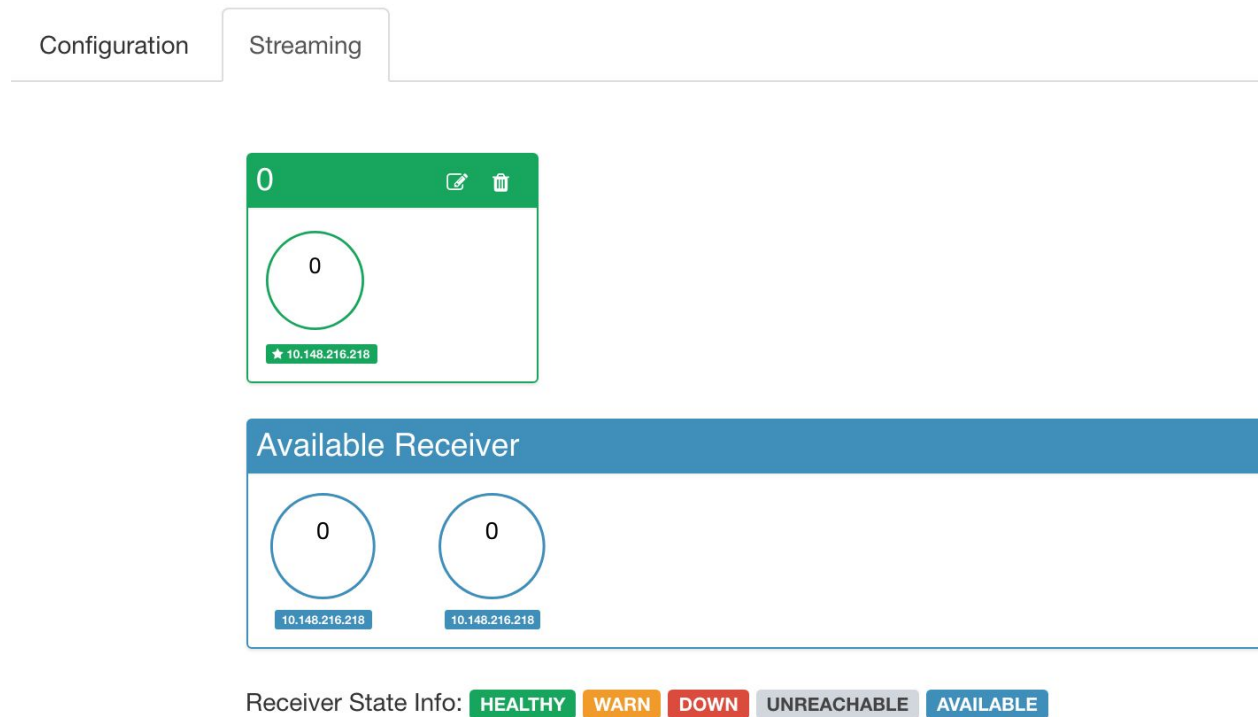
1) go to System -> Steaming tab in Kylin’s portal, you will see all the streaming receivers that have been started



2) click “+” button at the right top, a create replica set dialog will pop up, and you can choose the receivers you want to include in this replica set, click ‘save’



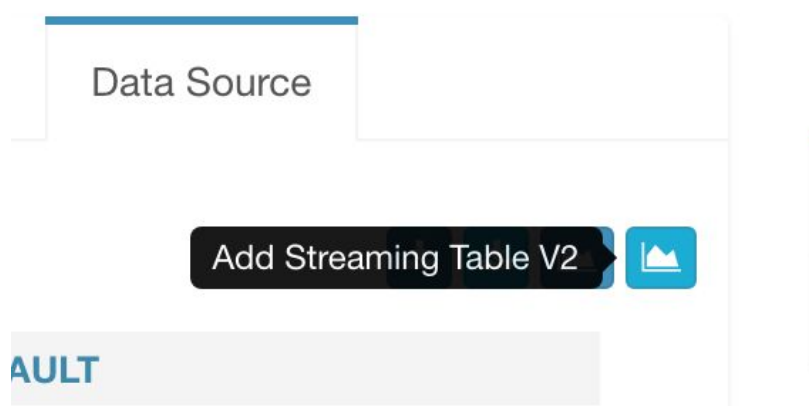
3) a new replicaSet will be created, you can then onboard a real-time cube



Onboard Real-time Cube

Create new Streaming Table

Step1: In DataSource tab, click 'Add Streaming Table V2' button




Step2: In the popup dialog, input the kafka topic name and kafka bootstrap servers, click next.

Topic*	<input type="text" value="test3"/>		
Bootstrap Servers*	Host	Port	Action
	<input type="text" value="10.148.216.117"/>	<input type="text" value="9092"/>	-
			+

Next →

Step3: Click the button in red rectangle, if there is data in related topic, one record will be retrieved as template, otherwise you need to input the json template manually.

Please input topic record sample, or automatically get it from 

JSON

```
1 {  
2   "PRICE": 0,  
3   "SELLER_ID": 337449,  
4   "TRANS_ID": 528265023,  
5   "META_CATEG_NAME": "META_CATEG_NAME-53",  
6   "CATEG_LVL2_NAME": "CATEG_LVL2_NAME-314",  
7   "CATEG_LVL3_NAME": "CATEG_LVL3_NAME-743",  
8   "LEAF_CATEG_NAME": "LEAF_CATEG_NAME-1990",  
9   "PART_DT": 1545126894298  
10 }
```

>>

Step4: click button ">>", Kylin will generate the streaming table schema automatically according to the template, you need specify the streaming table name, and the timestamp column name. If you want to create lambda cube, check the Lambda checkbox, but you must ensure that there is a hive table with the same table name exist in hive, and the hive table schema must be a superset of the schema you define here.

Table Name*

TEST3_TBL

☐ Lambda

TSColumn*

PART_DT

	Column	Type	Mapping	Comment
—	PRICE	int		
—	SELLER_ID	int		
—	TRANS_ID	int		
—	META_CATEG_NAME	varchar(256)		
—	CATEG_LVL2_NAME	varchar(256)		
—	CATEG_LVL3_NAME	varchar(256)		
—	LEAF_CATEG_NAME	varchar(256)		
—	PART_DT	timestamp		
—	year_start	date		derived time dimension
—	quarter_start	date		derived time dimension

Step5: Click 'Submit' button, new streaming table will be created.

Create Model

It is similar as create a batch Kylin model, except that you cannot define a star/snowflake schema, currently it only supports one fact table without any lookup table.

Create Cube

It is similar as create a batch Kylin cube, in 'Configuration Overwrites' tab, we add some streaming related properties as following, the 'kylin.stream.cube.window' property is used to configure the streaming segment length in seconds, by default it is 1 hour.

kylin.stream.cube.duration: is a time to keep data on Streaming Receiver hosts before moving to deep storage (e.g. Hbase). The default value is 3600 (1 hour).

kylin.stream.index.checkpoint.intervals: Time interval to do checkpoint for failure recovery. The default value is 300 (5 min).

✓

Cube Info

✓

Dimensions

✓

Measures

✓

Refresh Setting

✓

Advanced Setting

✓

Configuration Overwrites

Kylin Properties

+ Property

Streaming Properties

kylin.cube.algorithm

INMEM

kylin.stream.cube.window

3600

kylin.stream.cube.duration

3600

kylin.stream.index.checkpoint.intervals

300

kylin.stream.segment.retention.policy

fullBuild

purge

Tips

1. Cube level prop
overwrite config
kylin.properties

Start Consuming

Go to the model tab, open the cube's action drop-down list, click 'Enable' button, then the streaming receiver will start consuming events from kafka.

test3_cube

DISABLED

0.00 KB

0

ADMIN

2018-11-28 03:01:37 UTC

Action

Drop

Edit

Merge

Enable

Purge

Clone

Total: 4

Storage: 290.72 MB

After enable the cube, you can open the 'Streaming' sub-tab for this cube, you can see the consuming rate, how many events are consumed, such information.

test3_cube	READY	0.00 KB	0		ADMIN	2018-11-28 03:01:37 UT C	Action ▾	Action ▾
------------	-------	---------	---	--	-------	-----------------------------	----------	----------

Grid	SQL	JSON(Cube)	Notification	Storage	Planner	Streaming
------	-----	------------	--------------	---------	---------	-----------

Replica Set ID: 0	10.148.216.218:9090						120 msg/s	
	5'	15'	avg	consume	ingest			
	120msg/s	120msg/s	117msg/s	1000	1000			

And after the cube is enabled, it is queriable almost at the same time, you can open kylin's insight tab, to input sql to query the data:

New Query
Saved Queries
Query History

```

1 select minute_start,META_CATEG_NAME,count(1)
2 from TEST3_1
3 where minute_start>=timestamp'2018-12-19 00:00:00' and minute_start<timestamp'2018-12-19 06:00:00'
4 group by META_CATEG_NAME,minute_start
5 order by minute_start

```

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

Project: test
LIMIT 50000
Submit

Results

1 ✓ x | 2 ✓ x | 3 ✓ x | 4 ✓ x | 5 ✓ x | 6 ✓ x | 7 ✓ x | 8 ✓ x

Status: All

Query String
Start Time: 2018-12-19 05:02:39 UTC Duration: 0.21s Rerun Save

Status: Success Project: test Cubes: CUBE[name=test3_cube]

Results (3910) Visualization Export

Monitor