

Supersets Install 및 차트 생성

(이삭엔지니어링 인턴 김형근 fnfn9087@gmail.com)
(VM 에서 진행, Hive 테이블은 DIA 클러스터의 것을 사용)

제공 환경 (VM) :

Core : **16 Processor**

RAM : **60GB**

설치 위치 : /opt 폴더

목차

1. 설치 과정
2. Supersets 차트 생성
 - 1) 데이터 가져오기
 - 2) 차트 생성하기

1. 설치 과정

1) 다음과 같이 입력한다.

```
sudo yum upgrade python-setuptools
```

```
sudo yum install gcc gcc-c++ libffi-devel python-devel python-pip python-wheel openssl-  
devel libsass2-devel openldap-devel
```

2) Python 3.6.x 설치

```
sudo yum install https://centos7.iuscommunity.org/ius-release.rpm  
sudo yum install python36u
```

```
python3.6 -V //python 3.6.x 와 같이 상세버전이 표시된다.
```

```
sudo yum install python36u-pip  
sudo yum install python36u-devel
```

3) Python virtualenv(가상환경)을 설정한다. -> superset 은 python3.6 이상 지원.
(supersets 설치가 끝날때까지 가상환경을 유지할 것!)

```
pip install virtualenv  
python3.6 -m venv venv //python3.6 -m venv (생성할 디렉토리 명)  
source PATH/venv/bin/activate
```

위와 같이 가상환경을 작동시키면 프롬프트 앞에 “(디렉토리명)” 표시가 뜬다.

ex) (venv)[root@worker1~] \$

그리고, `$python -V` 를 입력하면 python 3.6.7 로 뜨는 것을 확인할 수 있다.
본래환경에서는 `$python3.6 -V` 라고 입력해야 python3 이상의 버전확인이 가능했다.

본래환경 : python 2.7.5

가상환경 : python 3.6.7

가상환경을 종료시키고 싶을땐, `$deactivate` 라고 입력하면 된다.

Python2가 설치되어 있지만 Python3 환경에서만 실행되는 프로그램도 있을 수 있다.

(Python2에서의 요소들이 Python3에서 새로 변경되었거나 지원을 안 하기 때문이다.)

그럴 경우에는 virtualenv로 가상환경을 설정하여 2버전과 3버전을 따로 쓸 수 있는 독립적인 환경을 구성해주는 것이다.

4) Python' s setup tools and pip

```
pip install --upgrade setuptools pip
```

5) Supersets installation

```
# Install superset  
pip install superset
```

```
# Create an admin user (you will be prompted to set a username, first and last name  
before setting a password)
```

```
fabmanager create-admin --app superset
```

```
# Initialize the database
```

```
superset db upgrade
```

```
# Load some data to play with
superset load_examples

# Create default roles and permissions
superset init

# To start a development web server on port 8088, use -p to bind to another port
superset runserver
```

6) SuperSets 접속(<http://VM IP address:8088>) 후, Hive connect error 해결

"Error in sasl_client_start (-4) SASL(-4): no mechanism available: No worthy mechs found"
라는 에러가 발생할 수 있다.

```
sudo yum install cyrus-sasl-devel cyrus-sasl-gssapi cyrus-sasl-md5 cyrus-sasl-plain
```

를 입력한 후, superset 프로세스를 재시작한다.

```
ps -ef | grep superset
kill -9 "process id"
superset runserver
```

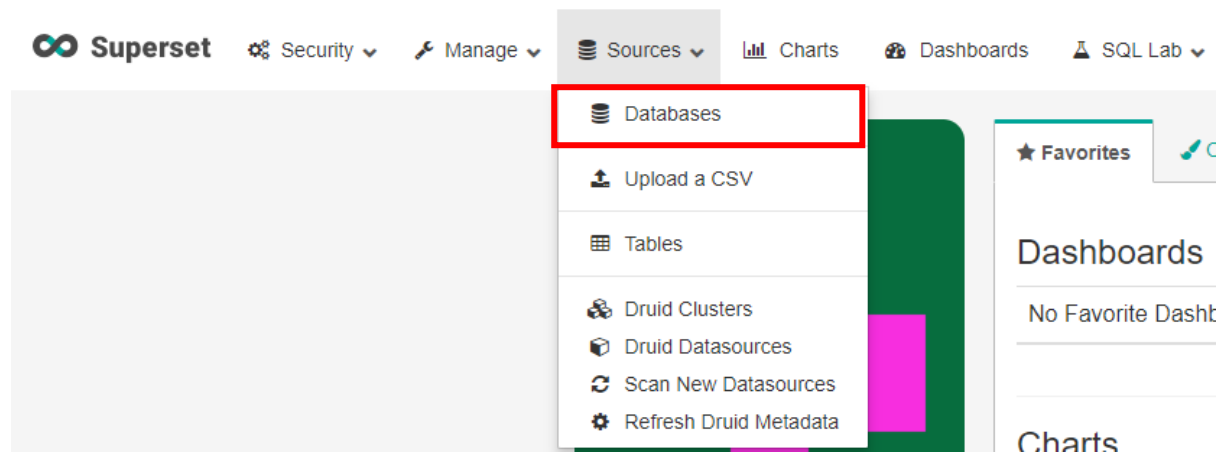
(superset stopserver, superset restartserver 등을 사용해서 서버를 재시작하고 싶었으나 적절한 명령어를 찾지 못해서, 결국 process를 죽이고 다시 시작하는 방식으로 택했다.)

출처 : <https://superset.incubator.apache.org/installation.html>

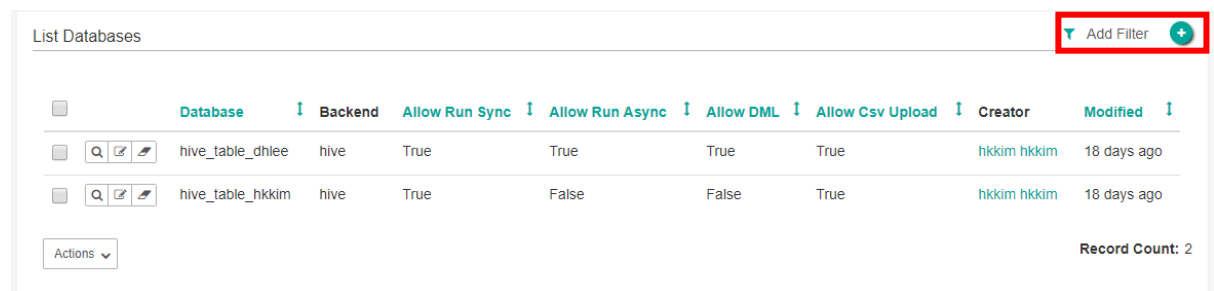
2. Supersets 차트 생성

1) 데이터 가져오기

(1) "Source" -> "Databases"



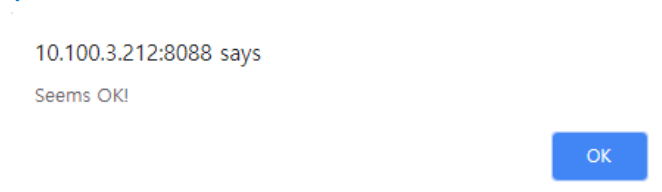
(2) "Add Filter" 옆의 "+" 눌러서 새 DB를 추가한다.



(3) Database란에는 본인이 짓고 싶은 이름을, URI란에 DB주소를 적는다.

Database	<input type="text" value="test_db"/>
SQLAlchemy URI	<input type="text" value="hive://10.100.3.210:10000/hkkim"/> <small>Refer to the SQLAlchemy docs for more information on how to structure your URI.</small> <button>Test Connection</button>
Chart Cache Timeout	<input type="text" value="Chart Cache Timeout"/> <small>Duration (in seconds) of the caching timeout for charts of this database. A timeout of 0 indicates that the cache never expires. Note this defaults to the global timeout if undefined.</small>

(Test Connection을 클릭하여 연결이 성공하면 아래와 같이 뜬다.)



(4) 하단에 있는

Save

를 눌러주면 test_db라는 이름의 DB가 생성된다.

List Databases Add Filter

	Database	Backend	Allow Run Sync	Allow Run Async	Allow DML	Allow Csv Upload	Creator	Modified
<input type="checkbox"/>	test_db	hive	True	False	False	False	admin admin	51 seconds ago
<input type="checkbox"/>	hive_table_dhlee	hive	True	True	True	True	hkkim hkkim	18 days ago
<input type="checkbox"/>	hive_table_hkkim	hive	True	False	False	True	hkkim hkkim	18 days ago

Actions Record Count: 3

(5) "Source" -> "Tables" -> "+"

Superset Security Manage Sources Charts Dashboards SQL Lab

- Databases
- Upload a CSV
- Tables**
- Druid Clusters
- Druid Datasources
- Scan New Datasources
- Refresh Druid Metadata

Equal to ☒

Databases

List Tables Add Filter

SQL Lab View Not Equal to ☒

Search

	Table	Database	Changed By	Modified
<input type="checkbox"/>	test	hive_table_hkkim	hkkim hkkim	18 days ago
<input type="checkbox"/>	tweetsbl	hive_table_dhlee	hkkim hkkim	18 days ago
<input type="checkbox"/>	tweetsbl_new	hive_table_hkkim	hkkim hkkim	18 days ago

Actions Record Count: 3

(6) Database 란에는 (4)에있는 "test_db" DB를 넣어준다.

Table Name에는 test_db와 연결되어 있는 hive 테이블을 하나 가져온다.

Import a table definition

Database *	<div>test_db</div>
Schema	<div>Schema</div> <div>Schema, as used only in some databases like Postgres, Redshift and DB2</div>
Table Name	<div>hkkim_tweetsbi</div> <div>Name of the table that exists in the source database</div>

	Table	Database	Changed By	Modified	1
<input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> hkkim_tweetsbi	test_db	admin admin	now	
<input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> test	hive_table_hkkim	hkkim hkkim	18 days ago	
<input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> tweetsbi	hive_table_dhlee	hkkim hkkim	18 days ago	
<input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> tweetsbi_new	hive_table_hkkim	hkkim hkkim	18 days ago	

2) 차트 생성하기

Superset Security Manage Sources Charts Dashboards SQL Lab

List Charts ▼ Add Filter +

	Chart	Visualization Type	↑ Datasource	↑ Creator	Last Modified	↓
<input type="checkbox"/>	hkkim_worldmap	world_map	tweetsbi_new	hkkim hkkim	4 days ago	
<input type="checkbox"/>	hkkim_test2	world_map	tweetsbi_new	admin admin	4 days ago	
<input type="checkbox"/>	hkkim_msg_visual	table	test	hkkim hkkim	18 days ago	
<input type="checkbox"/>	hkkim_test	dist_bar	tweetsbi_new	hkkim hkkim	18 days ago	

Actions ▼ Record Count: 4

Create a new chart

Choose a datasource

hkkim_tweetsbi ▼

If the datasource your are looking for is not available in the

Choose a visualization type

Table View ▼

Create new chart

(1) msg_visual 차트 ("Table View" 타입 사용)

▼ Time ⓘ

Time Column: ts ▼ Time Grain: day × ▼

Time range: 2018-11-26 : 2018-12-04

▼ NOT GROUPED BY ⓘ

Columns: × tweet_id × msg × ts × ▼

Ordering: 12 option(s) ▼

Row limit: 100 × ▼

msg_visual			
tweet_id	msg	ts	
1066966068680253440	Today I washed my face with my moisturizer. Twice.	2018-11-26T17:05:08	
1066994824656543744	That time it was #Shoaib and #Sami in this time it is Yasir alone.	2018-11-26T18:59:24	
1067000156975964160	What a story!! 🤔	2018-11-26T19:20:35	
1067029410497421317	I won't make the same mistakes twice anymore	2018-11-26T21:16:50	
1067034911192825857	Shhh 🤫 please don't let @TheBlueProphet see this ... lol	2018-11-26T21:38:42	
1067035006596333568	@daLionTamer Yes. Wouldn't think twice.	2018-11-26T21:39:04	
1067051835876392960	Thanks and an honour to be part of #C4PV making the Conservative case for #PeoplesVote	2018-11-26T22:45:57	
1067052427327741952	My gran saw @BigGirlsDC on Saturday night and is still raving about it. Proudly shown me her picture with the cast twice too.	2018-11-26T22:48:18	
1067063714413604864	@neiltyson @17HashBrownsPls These days he might have thought twice. Can you say The Toronto Project?	2018-11-26T23:33:09	
1067064891100549121	Just watch the game film and see how bad FSU's D played. Peep the busted coverages, how they got manhandled in the... https://t.co/3Uf5h1LEUb	2018-11-26T23:37:49	
1067083902508888064	I normally don't market, but I'm incredibly happy with the quality f stickers made by @StickerYou! I've ordered fro... https://t.co/0arGr76Vvt	2018-11-27T00:53:22	

(2) word_cloud 차트 (“Word Cloud” 타입 사용)

Time ⓘ

Time Column: Time Grain:

Time range:

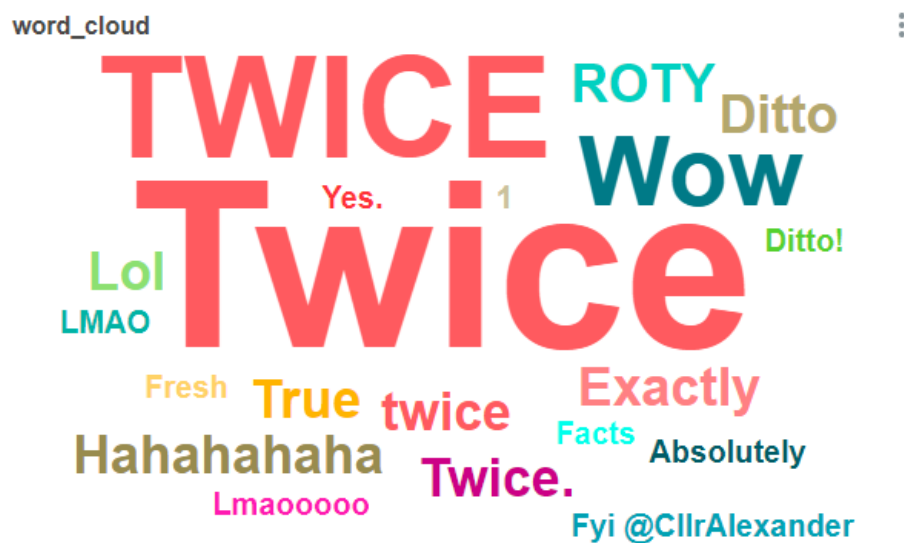
Query

Series:

Metric ⓘ:

Filters:

Row limit:



(3) Pie_chart 차트 (“Pie_chart” 타입 사용)

▼ Time ⓘ

Time Column

ts

Time Grain

day

Time range

2018-11-26 : 2018-12-04

▼ Query

Metric

COUNT(*) ⓘ

Filters

choose a column or metric

Group by

* country

Row limit

10

Pie_chart



United Kingdom

United States

(4) Bar_chart 차트 (“Distrubution Bar” 타입 사용)

▼ Time ⓘ

Time Column

ts

Time Grain

day

Time range

2018-11-26 : 2018-12-04

▼ Query

Metrics

× COUNT(*) ⓘ

Filters

choose a column or metric

Series

× lang

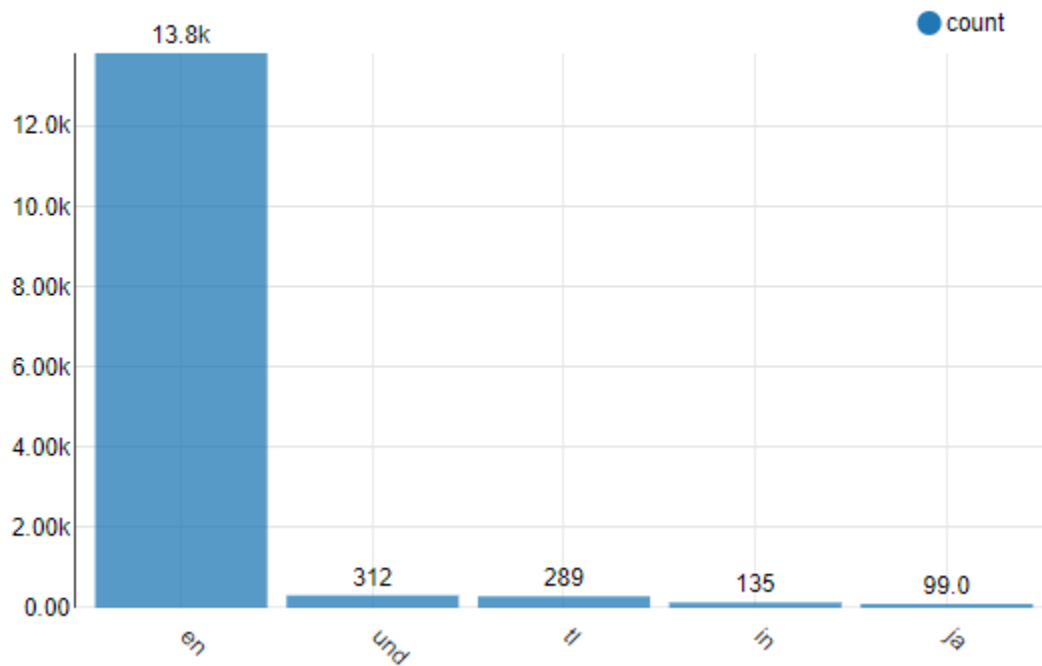
Breakdowns

4 option(s)

Row limit

5

Bar_chart



* Dashboard의 모습

test_db_dashboard ☆

Edit dashboard ▾

