

Twitter Data를 이용한 감성분석

(이삭엔지니어링 인턴 김형근 fnfn9087@gmail.com)

(StreamSets, Hive, Kibana 모두 CDH 5.15에서 진행.)

Index

1. StreamSets로 트위터 데이터 받기

- (1) 트위터 데이터를 받아오는 파이프라인
- (2) 받아온 트위터들을 다시 재정리 하는 파이프라인

2. Hue에서 Hive Table 만들기

- (1) Hue에서 Hive Table 만들기

3. Hive Table을 Kibana에서 분석

- (1) Kibana에서 사용할 external 테이블을 생성
- (2) Dashboard 생성
- (3) Dashboard에 들어갈 Visualize(그래프) 생성

4. Excel 3D Map에서 분석 (부록 먼저 실행할 것!!)

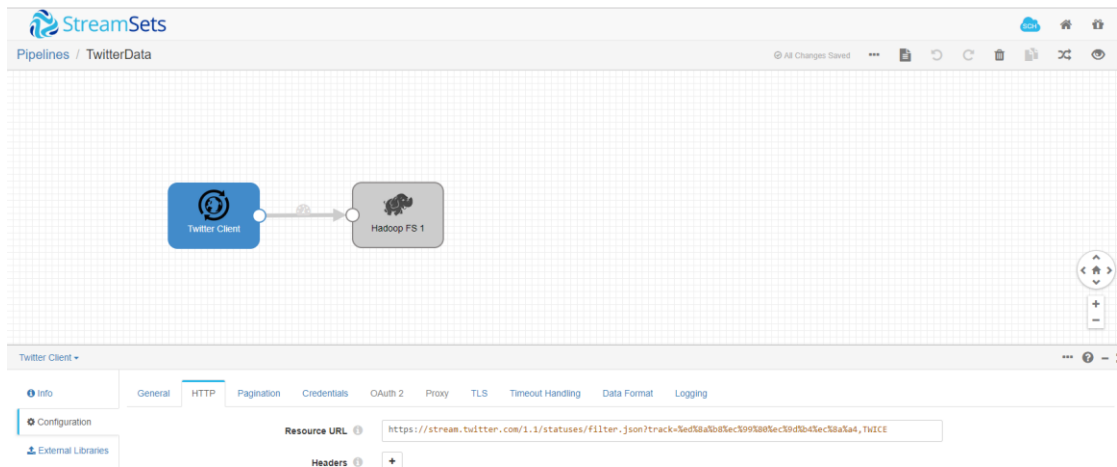
- (1) Hive Table 가져오기 : "데이터 탭" -> "새 쿼리" -> "기타 원본에서" - > "ODBC에서"
- (2) Hive Table 가져오기 : "데이터 탭" -> "기타 원본에서" - > "데이터 연결 마법사에서"
- (3) Excel 3D Map 사용

+ 부록 1 : Cloudera ODBC Connector 설치하는 방법

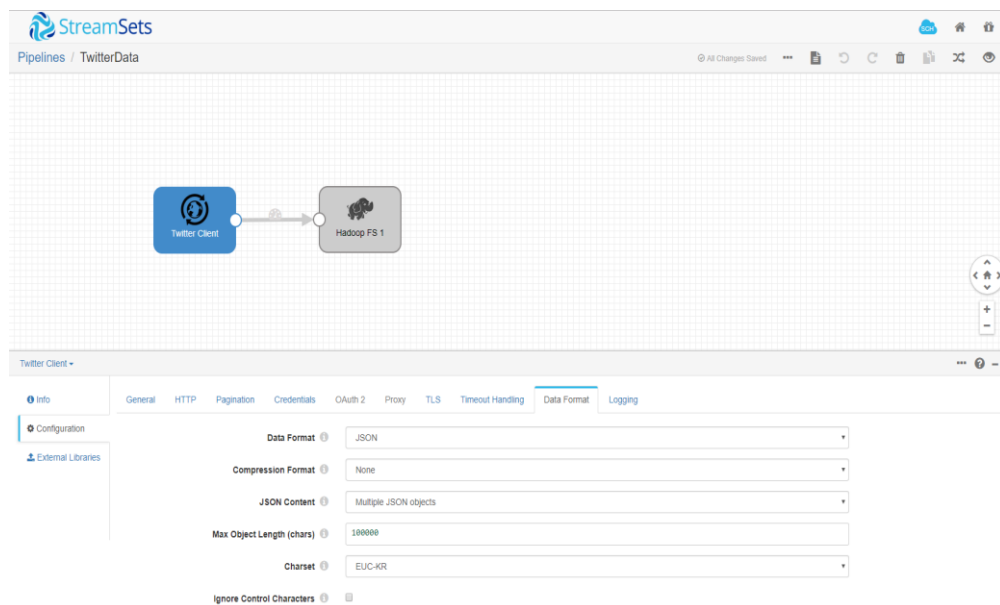
+ 부록 2 : Excel 3D Map 활성화

1. Streamsets로 트위터 데이터 받기(Topic : TWICE(걸그룹))

(1) 트위터 데이터를 받아오는 파이프라인



Resource : <https://stream.twitter.com/1.1/statuses/filter.json?track=%ed%8a%b8%ec%99%80%ec%9d%b4%ec%8a%a4,TWICE>



Data Format : JSON

File당 사이즈 : 500MB

Streaming한 데이터를 저장하는 위치 : /tmp/out/hkkm/twitter

General Hadoop FS Output Files Late Records Data Format

File Type ⁱ Text files

Files Prefix ⁱ sdc-\${sdc:id()}

Files Suffix ⁱ

Directory in Header ⁱ ☐

Directory Template ⁱ /tmp/out/hkim/twitter

Data Time Zone ⁱ +00:00 UTC (UTC)

Time Basis ⁱ \${time:now()}

Max Records in File ⁱ 0

Max File Size (MB) ⁱ 500

(2) 받아온 트위터들을 다시 재정리 하는 파이프라인

- 1) 받아온 Twitter Data들을 JSON파일로 정리하고, 다시 JSON의 Field들 중에서 몇몇 요소들만 골라낸다.

StreamSets

Pipelines / TwitterData_Level2

© All Changes Saved

TwitterData_Level2

Info Configuration Rules History

General Parameters Notifications Error Records Cluster EMR Test Origin

Pipeline ID TwitterData_Level234bf4e3c-21ce-43fd-a693-3d0f3944790e

Title TwitterData_Level2

Description

Labels

Execution Mode ⁱ Standalone

Delivery Guarantee ⁱ At Least Once

General Hadoop FS Files Post Processing Data Format

Name Hadoop FS Standalone 1

Description

Stage Library CDH 5.15.0

“General” 탭 -> “Stage Library” (CDH 5.15.0)

2) “Field Remover”에서 하단에 기입한 Field들만 남겨놓는다.

붉은색 field는 대부분이 NULL이지만, 가끔씩 값이 들어있는 경우가 있다.

(/id, /text, /created_at, /timestamp_ms,

/lang, /'user.location', /'user.screen_name',/'user.name', /'user.id', /'extended_tweet.full_text',

/'retweeted_status.text', /'retweet_status.extended_tweet.full_text', /'place.location', /'place.country')

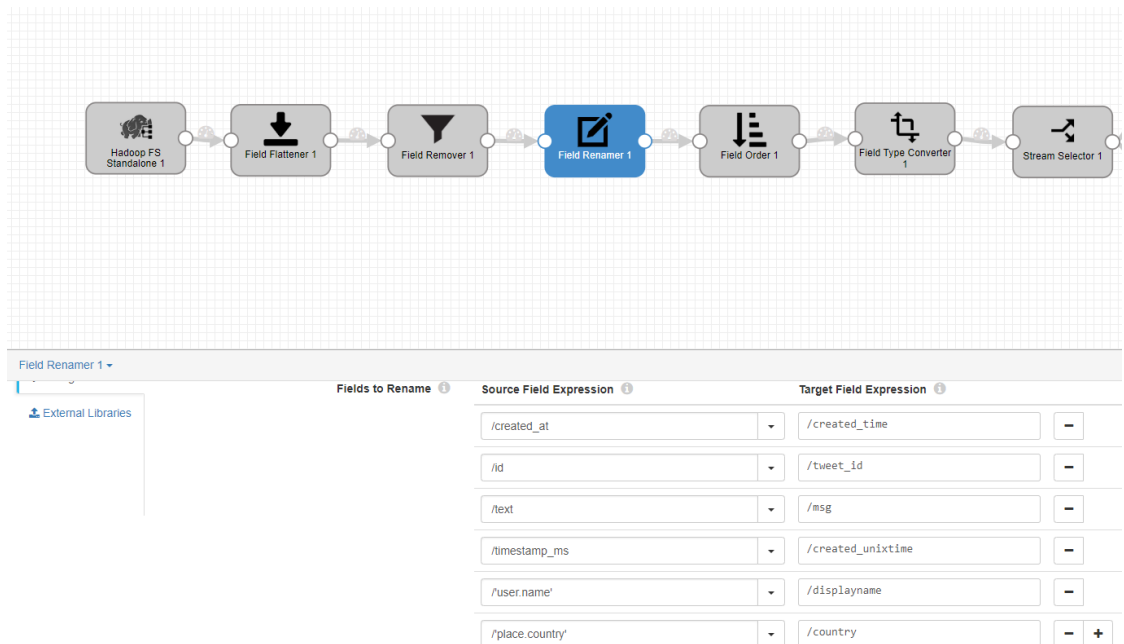
General Remove/Keep

Action ⓘ Keep Listed Fields

Fields ⓘ /created_at × /lang × /timestamp_ms × /text × /place.country × /id × /user.name ×

[Select Fields Using Preview Data](#)

3) Field Renamer에서 Field 이름을 변경한다.



1	tweet_id
2	msg
3	created_time
4	created_unixtime
5	lang
6	user_location
7	user_screen_name
8	user_name
9	user_id
10	extd_full_msg
11	rt_msg
12	rt_full_msg
13	location
14	country

/id -> tweet_id, **/text** -> msg,

/created_at -> created_time, **/timestamp_ms** -> created_unixtime

/place.location -> location, **/place.country** -> country

/extended_tweet.full_text -> extd_full_msg, **/retweeted_status.text** -> rt_msg

/retweet_status.extended_tweet.full_text -> rt_full_msg

나머지 Field에는 '.' 가 들어가 있다면 '_'로 대체한다. (예 : **/user.location** -> /user_location)

4) "Field Order" : 위 1 ~ 14 순서대로 Field를 정렬한다.



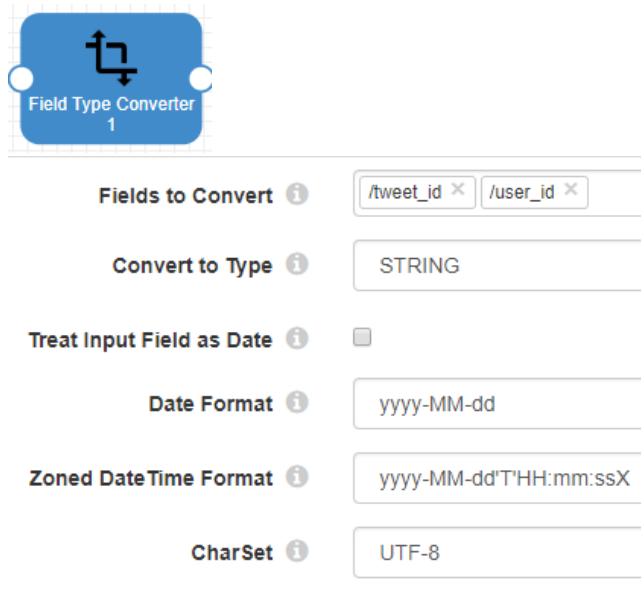
Fields to Order ⓘ

/tweet_id × /msg × /created_time × /created_unixtime × /lang × /user_location × /user_screen_name × /user_name × /user_id ×
 /extd_full_msg × /rt_msg × /rt_full_msg × /location × /country ×

Select Fields Using Preview Data

5) "Field Type Converter" : /tweet_id, /user_id를 STRING으로 바꾸어준다.

(STRING으로 바꾸지 않으면 Preview시 Type 에러로 간주한다.)

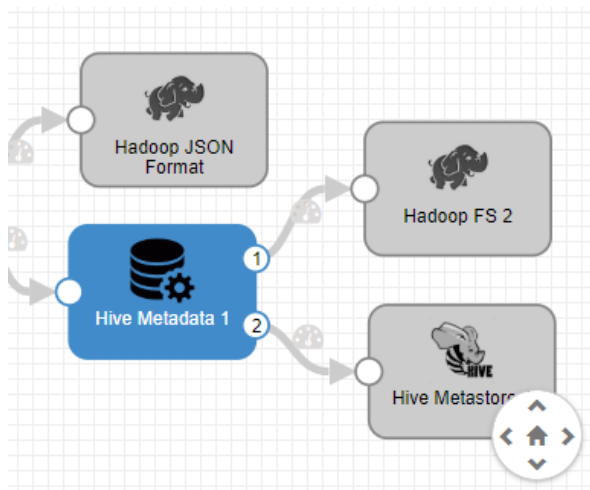


The image shows the configuration interface for the "Field Type Converter" tool. It has a blue header with a circular icon and the text "Field Type Converter 1". Below the header, there are several configuration fields:

- Fields to Convert**: A list of fields to be converted, currently showing "/tweet_id" and "/user_id".
- Convert to Type**: A dropdown menu set to "STRING".
- Treat Input Field as Date**: A checkbox that is currently unchecked.
- Date Format**: A text input field containing "yyyy-MM-dd".
- Zoned DateTime Format**: A text input field containing "yyyy-MM-dd'T'HH:mm:ssX".
- CharSet**: A text input field containing "UTF-8".

6) 정리한 데이터들을 JSON파일로도 저장해두고, Hive Table도 생성해준다.

(Table명 : tweets_text2)



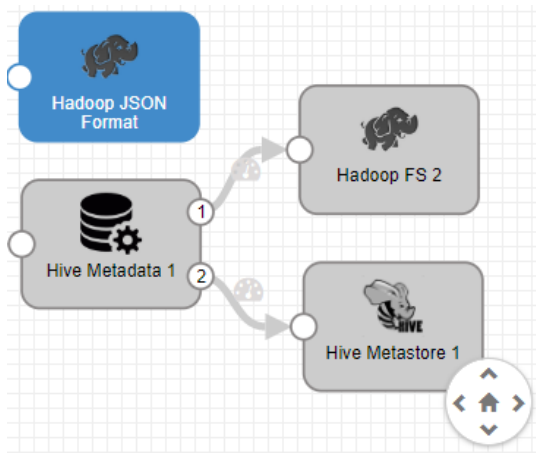
"General" 탭 -> "Stage Library" (CDH 5.15.0)

"Hive" 탭 -> "JDBC URL" **jdbc:hive2://10.100.3.210:10000/(Database)**

"Table" 탭 -> "Database Expression" (DB이름), "Table Name" (테이블 이름 짓기)

"Data Format" 탭 -> "Data Format" (Avro)

(Hive DB는 반드시 먼저 생성되어 있어야한다!!)

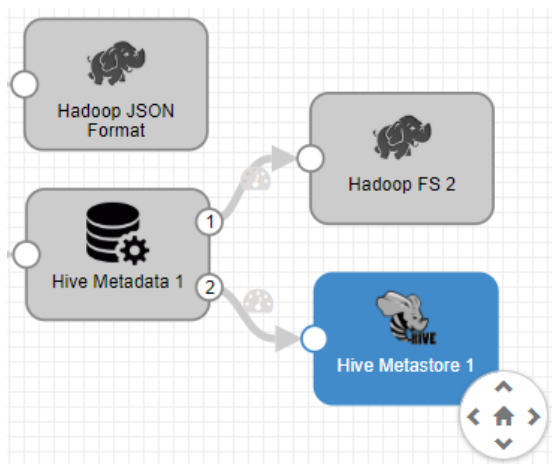


"General" 탭 -> "Stage Library" (CDH 5.15.0)

"Hadoop FS" 탭 -> "Hadoop FS Configuration Directory" (hadoop-conf)

"Output Files" 탭 -> "Directory Template" (저장하고 싶은 곳의 경로를 넣는다.)

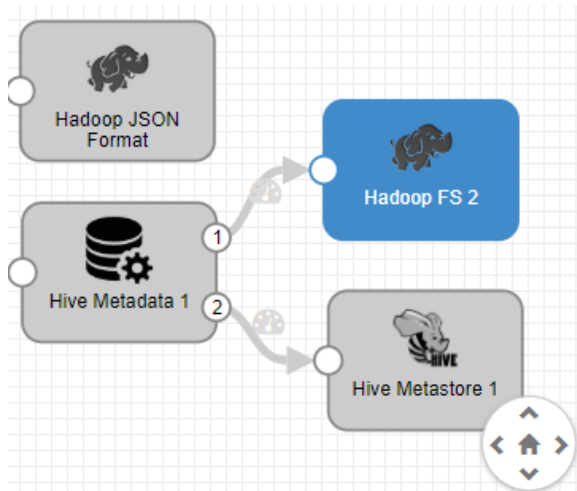
"Data Format" 탭 -> "Data Format" (JSON)



"General" 탭 -> "Stage Library" (CDH 5.15.0)

"Hive" 탭 -> "JDBC URL" **jdbc:hive2://10.100.3.210:10000/(Database)**

"Advanced" 탭 -> "Stored as Avro" (체크하기)



"General" 탭 -> "Stage Library" (CDH 5.15.0)

"Hadoop FS" 탭 -> "Hadoop FS Configuration Directory" (hadoop-conf)

"Output Files" 탭 -> "Directory in Header" (체크하기)

"Data Format" 탭 -> "Data Format" (Avro)

홈 / user / hive / warehouse / hkkim.db

<input type="checkbox"/>	이름	크기	사용자	그룹
<input type="checkbox"/>	↓		hive	hive
<input type="checkbox"/>	.		impala	hive
<input type="checkbox"/>	_tmp.twitterbi_hkkim		hkkim	hive
<input type="checkbox"/>	tt		hkkim	hive
<input type="checkbox"/>	tweets_sentiment		hkkim	hive
<input type="checkbox"/>	tweets_text		sdc	hive
<input type="checkbox"/>	tweets_text2		sdc	hive
<input type="checkbox"/>	tweets_text_v3		sdc	hive
<input type="checkbox"/>	tweetsbi		hkkim	hive
<input type="checkbox"/>	tweetsbi_new		hkkim	hive
<input type="checkbox"/>	tweetsbi_notnull		hkkim	hive
<input type="checkbox"/>	twitterbi_hkkim		hkkim	hive

2. Hue 에서 Hive Table 만들기

(1) Hue 에서 Hive Table 만들기

dictionary 테이블 만들기

(영어 단어에 대해 긍정, 중립, 부정 척도를 판별해주는 table)

```
CREATE EXTERNAL TABLE if not exists dictionary
(
  type string,
  length int,
  word string,
  pos string,
  stemmed string,
  polarity string )
ROW FORMAT DELIMITED FIELDS TERMINATED BY 't'
STORED AS TEXTFILE LOCATION '/user/hkkim/twitter/data/tables/dictionary';
```

tweets_clean 뷰 만들기

(사용자가 만든 tweets_text2 에서 tweet_id, timestamp, message, country 만 뽑은 View 를 만든다.)

```
CREATE VIEW IF NOT EXISTS tweets_clean AS
SELECT tweet_id,
cast(from_unixtime(cast(substring(created_unixtime,1,10) as bigint )) as timestamp) ts, msg, country FROM tweets_text2;
```

l1, l2, l3 뷰 만들기

(l1 : msg 의 문장들을 단어 하나하나씩 쪼갬다.)

```
create view IF NOT EXISTS l1 as select tweet_id, words from tweets_text2 l
ateral view explode(sentences(lower(msg))) dummy as words;
```

(l2 : l1 에서 쪼갬 단어들 각각에 대해 나열한 view 를 만든다.)

```
create view IF NOT EXISTS l2 as select tweet_id, word from l1 lateral view
explode( words ) dummy as word;
```

(l3 : l2 에 나열된 각 단어들에 대한 긍정, 중립, 부정의 척도를 1,0,-1 로 나타낸다.)

```
create view IF NOT EXISTS l3 as select
tweet_id, l2.word,
case d.polarity
when 'negative' then -1
when 'positive' then 1
else 0 end as polarity
from l2 left outer join dictionary d on l2.word = d.word;
```

tweets_sentiment 테이블 만들기

(tweet_id. 즉, 사용자 별로 msg의 내용이 긍정, 중립, 부정 인지를 나타내는 표이다.)

```
create table IF NOT EXISTS tweets_sentiment as select
tweet_id,
case
when sum( polarity ) > 0 then 'positive'
when sum( polarity ) < 0 then 'negative'
else 'neutral' end as sentiment
from l3 group by tweet_id;
```

tweetsbi 테이블 만들기

(앞서 만들었던 tweets_clean table 과 tweets_sentiment table 을 서로 join 시킨다.)

```
CREATE TABLE IF NOT EXISTS tweetsbi AS SELECT
t.*,
case s.sentiment
when 'positive' then 2
when 'neutral' then 1
when 'negative' then 0
end as sentiment
FROM tweets_clean t LEFT OUTER JOIN tweets_sentiment s on t.tweet_id = s.tweet_id;
```

tweetsbi_notnull 테이블 만들기

(sentiment 및 country 에 NULL 이 다수 포함되어 있어 NULL 을 없앤 table 을 다시 만들.)

```
create table tweetsbi_notnull as  
select * from tweetsbi where country != 'NULL' and sentiment is not null
```

tweetsbi_new 테이블 만들기

(tweets_text2 의 lang Column 도 포함시킨 table 을 만들.)

```
create table tweetsbi_new as  
select tweetsbi_notnull.*, tweets_text2.lang  
from tweetsbi_notnull left outer join tweets_text2  
on tweetsbi_notnull.tweet_id = tweets_text2.tweet_id
```

3. Hive Table을 Kibana에서 분석

1) Kibana에서 사용할 external 테이블을 생성한다.

```
CREATE EXTERNAL TABLE twitterbi_hkkm_kibana (  
  tweet_id string,  
  ts string,  
  msg string,  
  country string,  
  sentiment int,  
  lang string  
)  
STORED BY 'org.elasticsearch.hadoop.hive.EsStorageHandler'  
TBLPROPERTIES('es.resource' = 'twitterbi_hkkm_kibana/doc', 'es.nodes'='10.1  
00.3.210');  
INSERT OVERWRITE TABLE twitterbi_hkkm_kibana SELECT * FROM tweetsbi_new;
```

10.100.3.210:5601 Kibana로 접속

The screenshot shows the Kibana Management interface. On the left is a sidebar with navigation links: Discover, Visualize, Dashboard, Timelion, Canvas, Machine Learning, Infrastructure, Logs, APM, Dev Tools, Monitoring, and Management (highlighted with a red box). The main content area is titled 'Management / Kibana' and includes a 'Create index pattern' button. Below this button, a list of index patterns is shown: es_bi_twitter, kibana_sample_data_logs, and twitterbi_hkkm. The 'Create index pattern' dialog is open, showing 'Step 1 of 2: Define index pattern'. It contains a text input field for the index pattern, a list of indices to match, and a 'Rows per page' dropdown menu (set to 10, also highlighted with a red box).

Management / Kibana

Index Patterns Saved Objects Spaces Reporting Advanced Settings

Create index pattern

★ es_bi_twitter
kibana_sample_data_logs
twitterbi_hkkm

Create index pattern

Kibana uses index patterns to retrieve data from Elasticsearch indices for things like visualizations.

Step 1 of 2: Define index pattern

Index pattern

index-name-*

You can use a * as a wildcard in your index pattern.
You can't use spaces or the characters \, /, ?, ", <, >, |.

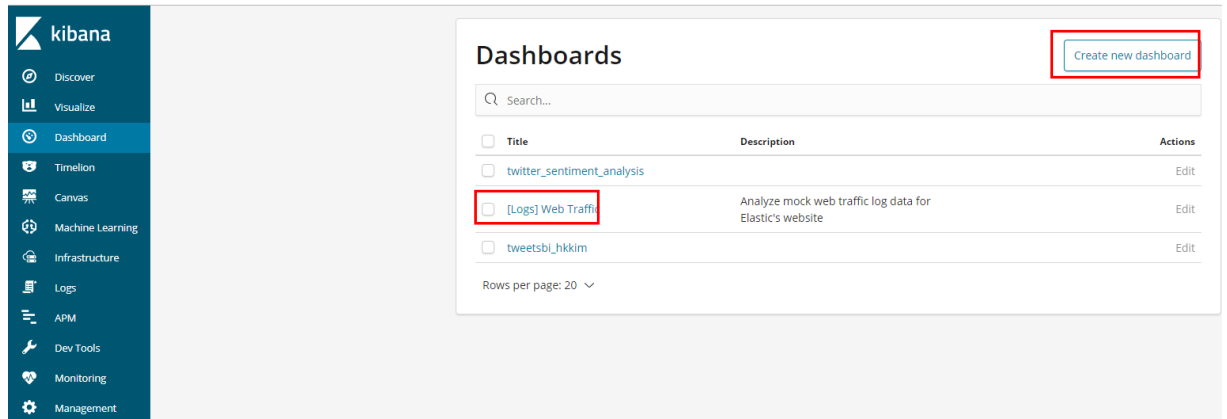
Your index pattern can match any of your **5 indices**, below.

es_bi_twitter
es_bi_twitter_christmas
es_bi_twitter_schol
radio
twitterbi_hkkm

Rows per page: 10 ▾

2) Dashboard 생성

“tweetsbi_hkkim”이란 이름의 Dashboard 생성

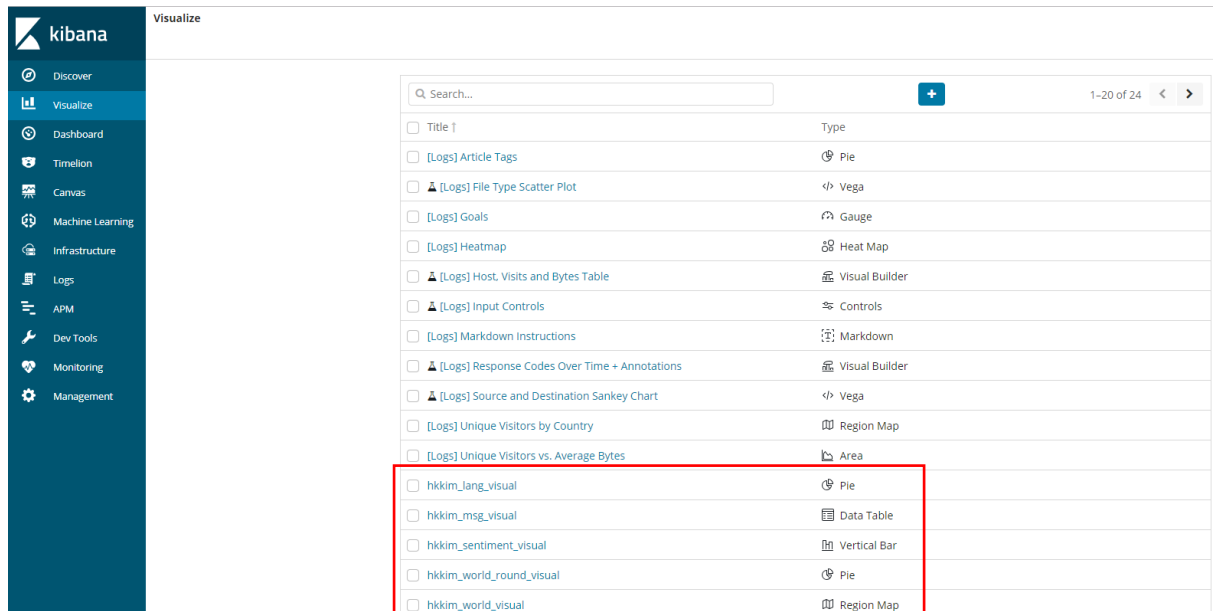


The screenshot shows the Kibana interface. On the left is a dark blue sidebar with the Kibana logo and a list of navigation items: Discover, Visualize, Dashboard (selected), Timelion, Canvas, Machine Learning, Infrastructure, Logs, APM, Dev Tools, Monitoring, and Management. The main area is titled 'Dashboards' and contains a search bar, a table of existing dashboards, and a 'Create new dashboard' button in the top right corner. The table has columns for Title, Description, and Actions. The following table represents the data shown in the screenshot:

Title	Description	Actions
twitter_sentiment_analysis		Edit
[Logs] Web Traffic	Analyze mock web traffic log data for Elastic's website	Edit
tweetsbi_hkkim		Edit

Below the table, it says 'Rows per page: 20' with a dropdown arrow.

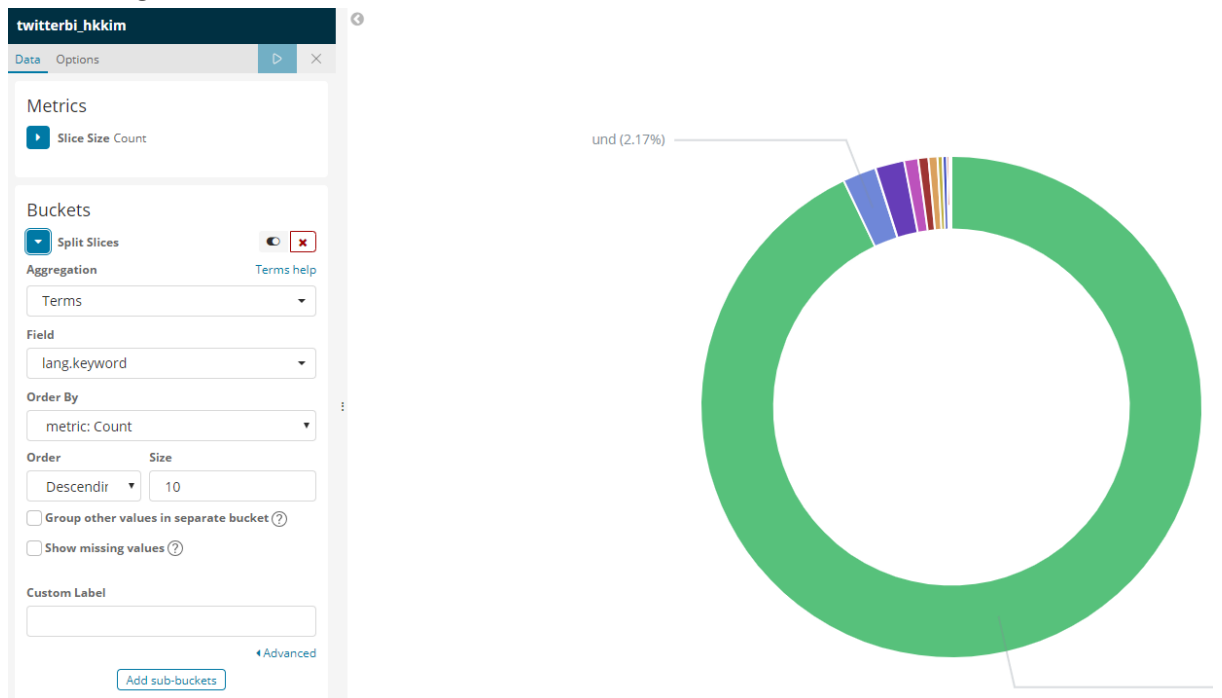
3) Dashboard에 들어갈 Visualize(그래프) 생성



The screenshot shows the Kibana 'Visualize' page. The left sidebar is the same as in the previous screenshot, but 'Visualize' is selected. The main area is titled 'Visualize' and contains a search bar, a list of visualizations, and a '+ Add' button. The visualizations are listed in a table with columns for Title, Type, and a '+' icon. The following table represents the data shown in the screenshot:

Title	Type
[Logs] Article Tags	Pie
[Logs] File Type Scatter Plot	Vega
[Logs] Goals	Gauge
[Logs] Heatmap	Heat Map
[Logs] Host, Visits and Bytes Table	Visual Builder
[Logs] Input Controls	Controls
[Logs] Markdown Instructions	Markdown
[Logs] Response Codes Over Time + Annotations	Visual Builder
[Logs] Source and Destination Sankey Chart	Vega
[Logs] Unique Visitors by Country	Region Map
[Logs] Unique Visitors vs. Average Bytes	Area
hkkim_lang_visual	Pie
hkkim_msg_visual	Data Table
hkkim_sentiment_visual	Vertical Bar
hkkim_world_round_visual	Pie
hkkim_world_visual	Region Map

hkkim_lang_visual : 언어별 사용자에게 대한 통계를 원그래프로 표현

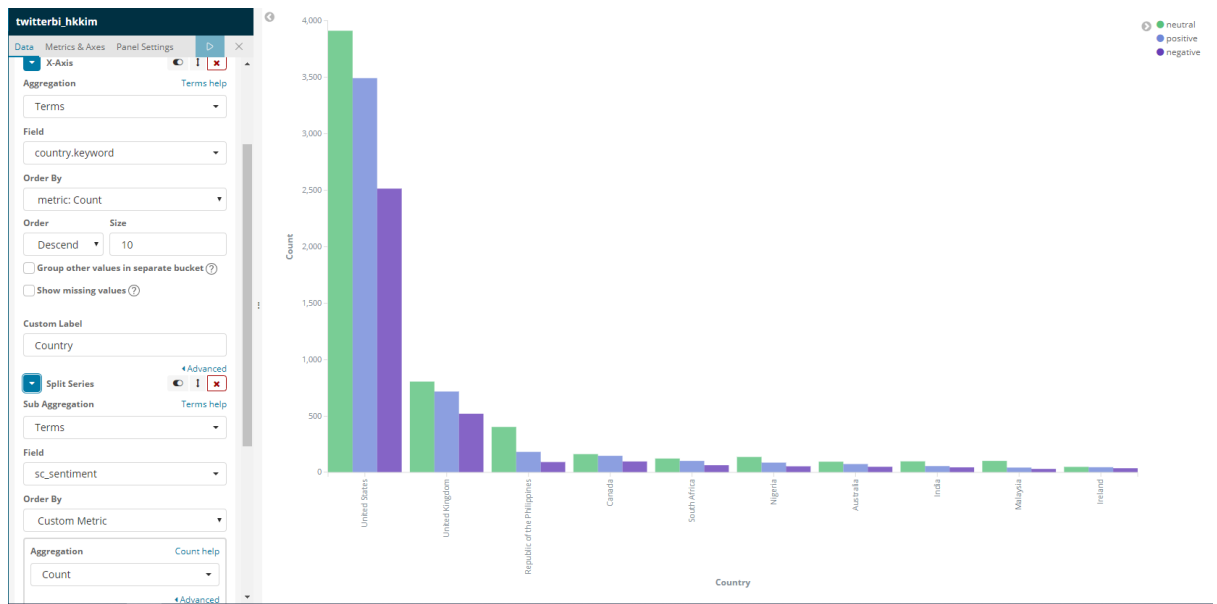


hkkim_msg_visual : 사용자들이 작성한 트위터 메시지들을 표시

The screenshot shows the configuration for a table view in the twitterbi_hkkim interface. The Metrics section is set to 'Split Rows', 'Aggregation' set to 'Terms', 'Field' set to 'msg.keyword', 'Order By' set to 'Alphabetical', and 'Order' set to 'Descend' with a size of 10. The Buckets section is also configured with 'Split Rows', 'Sub Aggregation' set to 'Terms', 'Field' set to 'ts.keyword', 'Order By' set to 'metric: Count', and 'Order' set to 'Descend' with a size of 5. The resulting table displays tweets with columns for Message, Time, and Count.

Message	Time	Count
well she wasted her money twice	Sat Dec 01 00:40:19 +0000 2018	1
sometimes ya gotta try twice https://t.co/W1Wb1PNLXX	Thu Nov 29 02:42:15 +0000 2018	1
Bastards. Punishment should be like for like. That would make them and others think twice about doing it in the fi... https://t.co/DoaA6jnWg0	Mon Nov 26 14:01:39 +0000 2018	1
smdh	Sun Dec 02 08:01:15 +0000 2018	1
ahh cause if he did it once i probably did it twice back. Facts	Fri Nov 30 02:02:18 +0000 2018	1
not again	Tue Nov 27 17:18:34 +0000 2018	1
that's a sign! i think quant is your calling @joannadrivalas	Thu Nov 29 16:40:22 +0000 2018	1
went through a nigga phone more than twice & saw all tyla shit these Nigga's really don't be for u like they say they are.	Sat Dec 01 04:34:29 +0000 2018	1
can't wait to hear what other silly benefits our leaders get that don't trickledown to us. Thanks for keeping it... https://t.co/WelSr4zogz	Sun Dec 02 13:11:36 +0000 2018	1
y'all ever had a friend who just inspired you to work twice as hard just to keep up with their greatness.... https://t.co/cgxl8rYWWE	Tue Nov 27 23:56:45 +0000 2018	1

hkkim_sentiment_visual : 사용자들의 긍정, 중립, 부정의 정도를 표현



(+ value명을 변경하는 법.)

(1) Management -> twitterbi_hkkim_kibana -> Scripted fields -> Add scripted field

twitterbi_hkkim_kibana

This page lists every field in the **twitterbi_hkkim_kibana** index and the field's associated core type as recorded by Elasticsearch. To change a field type, use the Elasticsearch [Mapping API](#).

Fields (16)

Scripted fields (1)

Source filters (0)

Q Filter

All field types

Name	Type	Format	Searchable	Aggregatable	Excluded
_id	string		•	•	
_index	string		•	•	
_score	number				
_source	_source				
_type	string		•	•	
country	string		•		
country.keyword	string		•	•	
lang	string		•		
lang.keyword	string		•	•	
msg	string		•		

Fields (16)

Scripted fields (1)

Source filters (0)

Q Filter

All languages

Scripted fields

You can use scripted fields in visualizations and display them in your documents. However, you cannot search scripted fields.

Add scripted field

Name	Lang	Script	Format
			<div><div>Edit</div><div>Delete</div></div>

(2) Script 작성

Create scripted field

⚠ Proceed with caution
Please familiarize yourself with [script fields](#) and with [scripts in aggregations](#) before using scripted fields.

Scripted fields can be used to display and aggregate calculated values. As such, they can be very slow, and if done incorrectly, can cause Kibana to be unusable. There's no safety net here. If you make a typo, unexpected exceptions will be thrown all over the place!

Name

New scripted field

Name is required

Language

painless

Type

number

Format (Default: Number)

- Default -

Formatting allows you to control the way that specific values are displayed. It can also cause values to be completely changed and prevent highlighting in Discover from working.

Popularity

0

Script

Script is required

Name : scripted field 이름 짓기 ex) *str_sentiment*

Language : Painless, Expression 2개가 있는데, default는 Painless.

필자는 Painless로 설정되어 있는 것으로 바로 진행.

Type : 잘 모르겠음.

Format : 잘 모르겠음. (Type이 바뀌면 format의 default값 또한 변경된다.)

Popularity : 잘 모르겠음.

Script : Script내용 작성

(ex) sentiment field는 중립, 긍정, 부정을 '1', '2', '0'으로 표현. 그러나 그 표현방식을 "Neutral", "Positive", "Negative"로 바꾸어보는 script.)

꼭 알아야할 점 : `doc['field_name'].value`을 통해 해당 field에 있는 value들에게 access할 수 있다.

```
if (doc['sentiment'].value == 1) return "Neutral";  
else if (doc['sentiment'].value == 2) return "Positive";  
else return "Negative";
```


DataMetrics & AxesPanel Settings

Metrics

Y-Axis Count

Add metrics

Buckets

X-Axis

AggregationTerms help

Terms

Field

Select a field

_type

country.keyword

lang.keyword

msg.keyword

ts.keyword

tweet_id.keyword

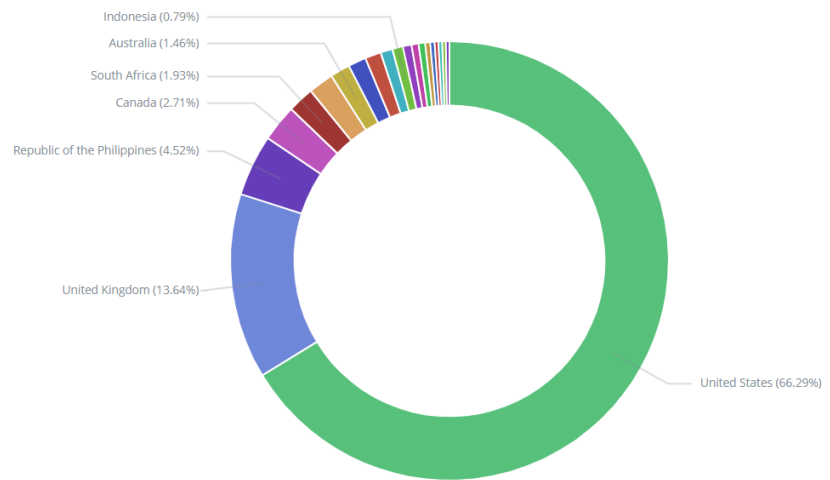
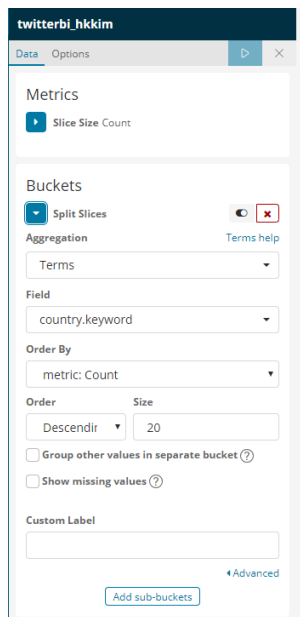
str_sentiment

Custom Label

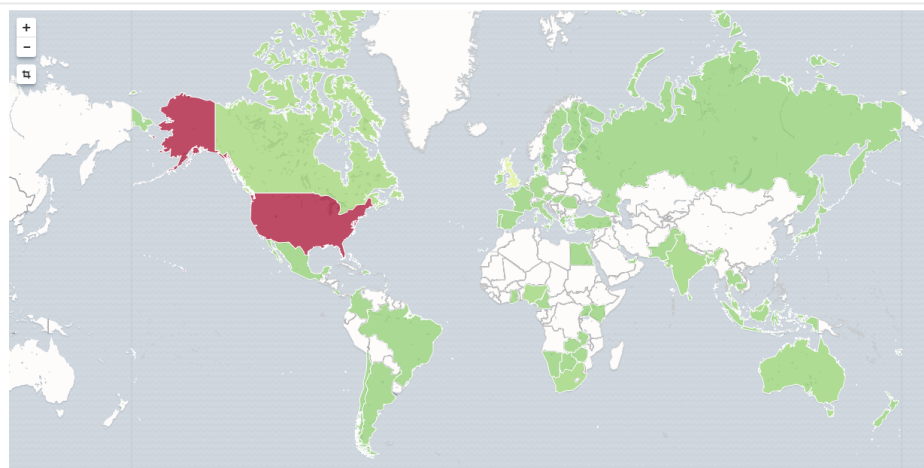
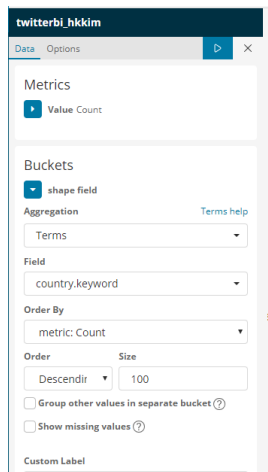
Country

위 과정을 거쳐서 "str_sentiment"란 이름을 가진 scripted field를 새로 생성하였다.

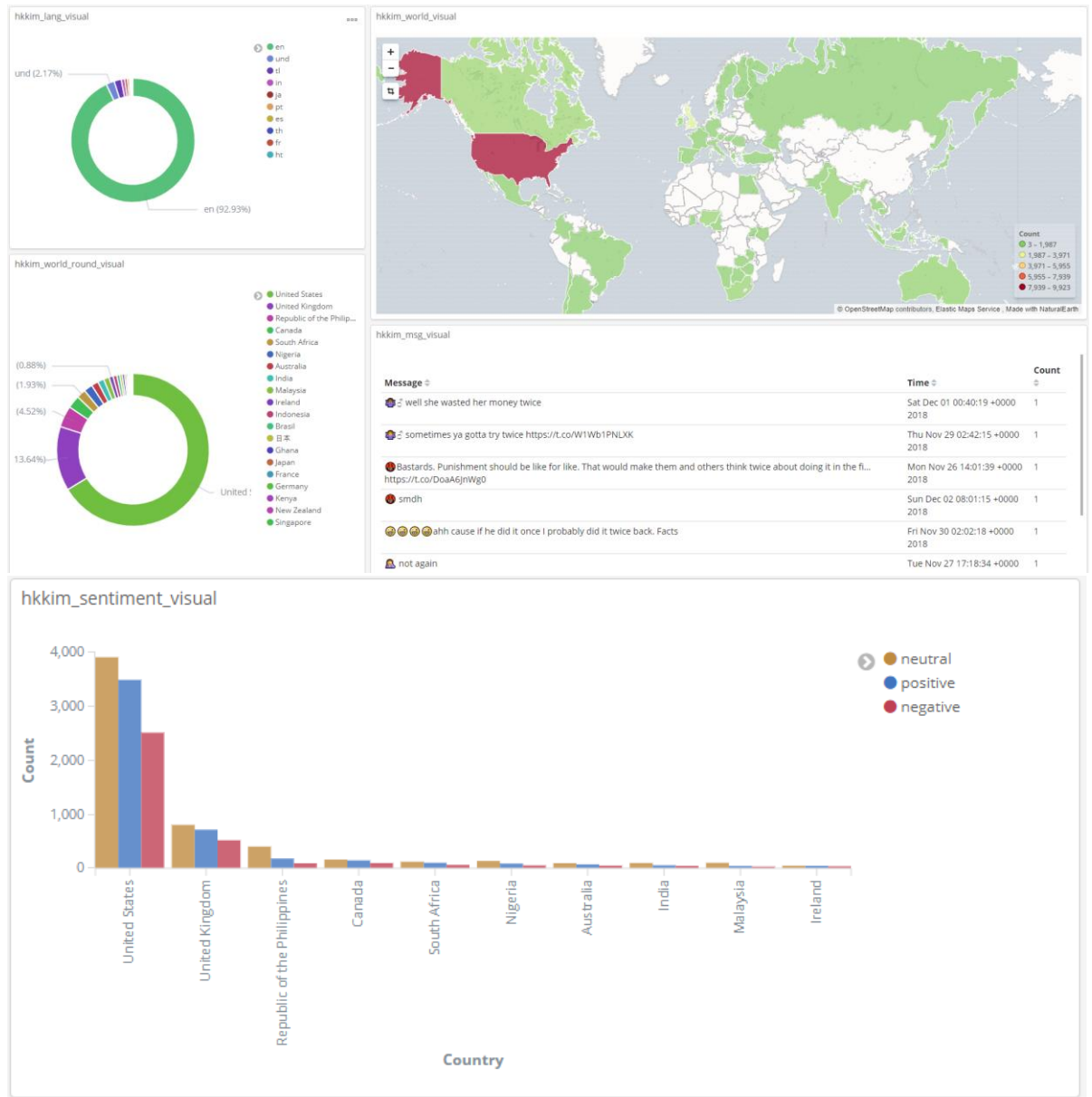
hkkim_world_round_visual : 나라별 사용자에게 대한 통계를 원그래프로 표현



hkkim_world_visual : 나라별 사용자에게 대한 통계를 세계지도에서 표현

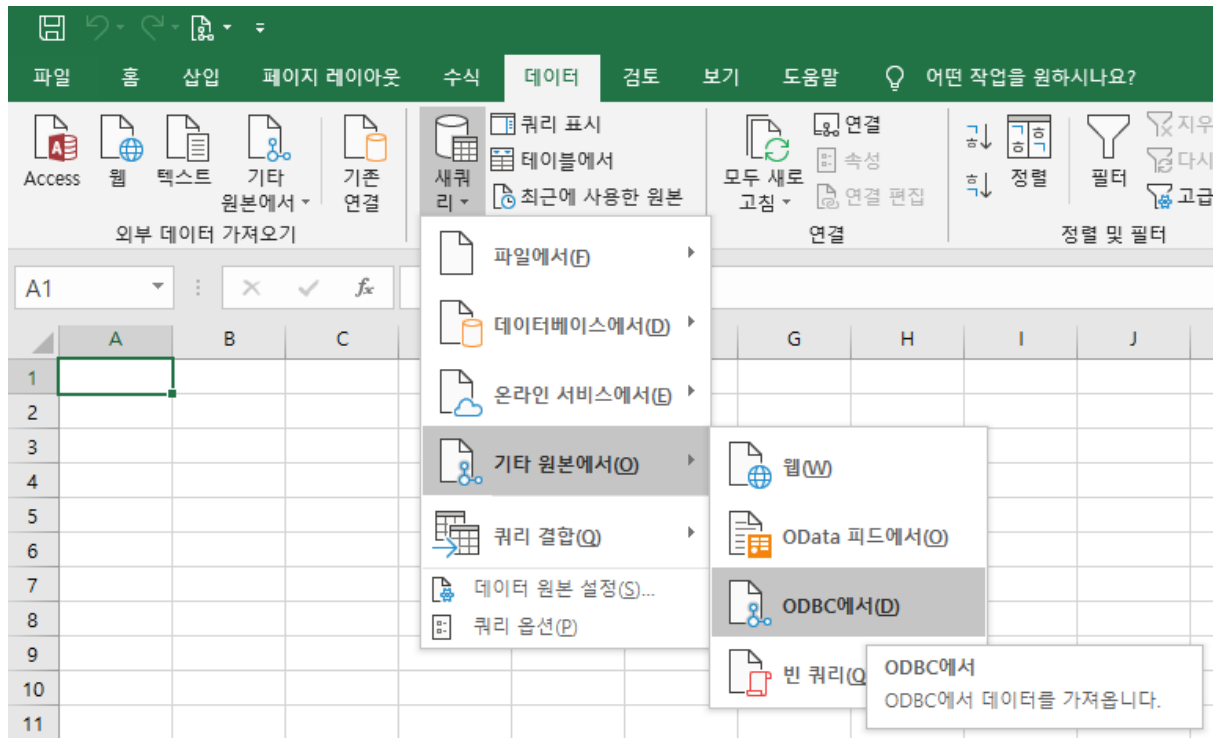


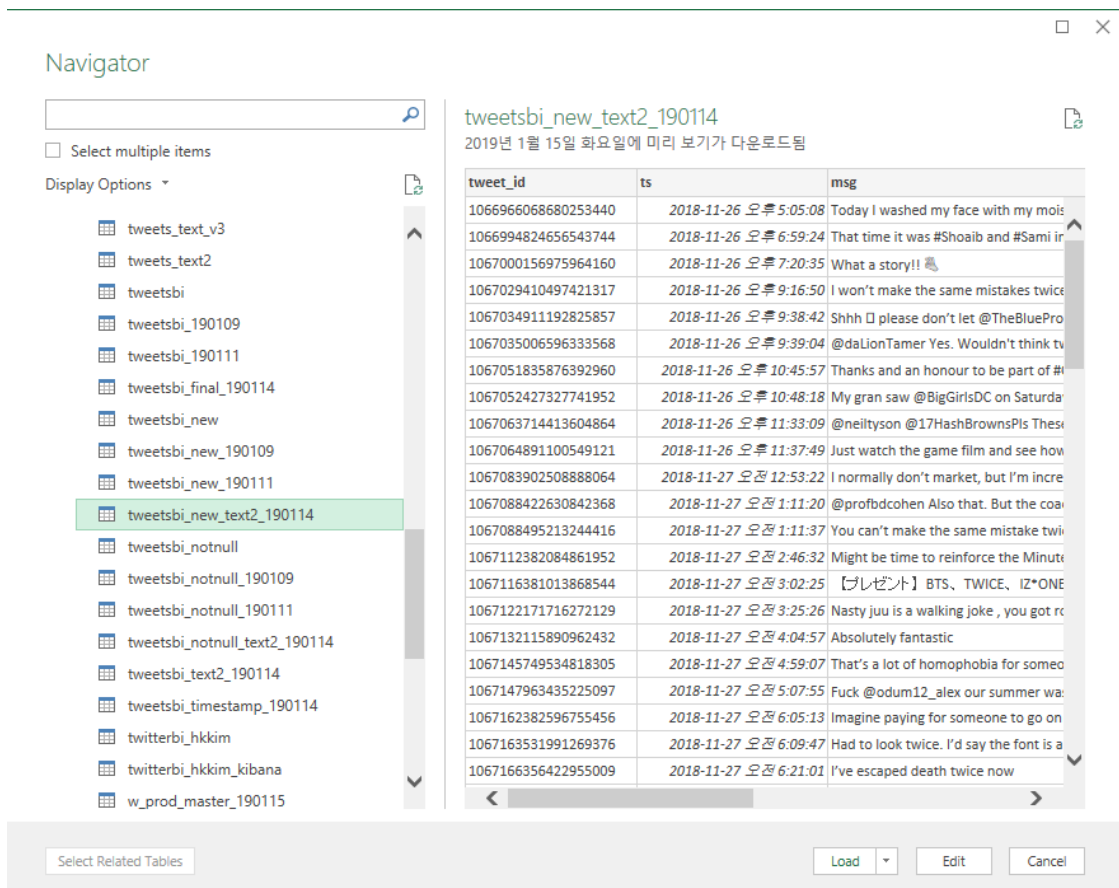
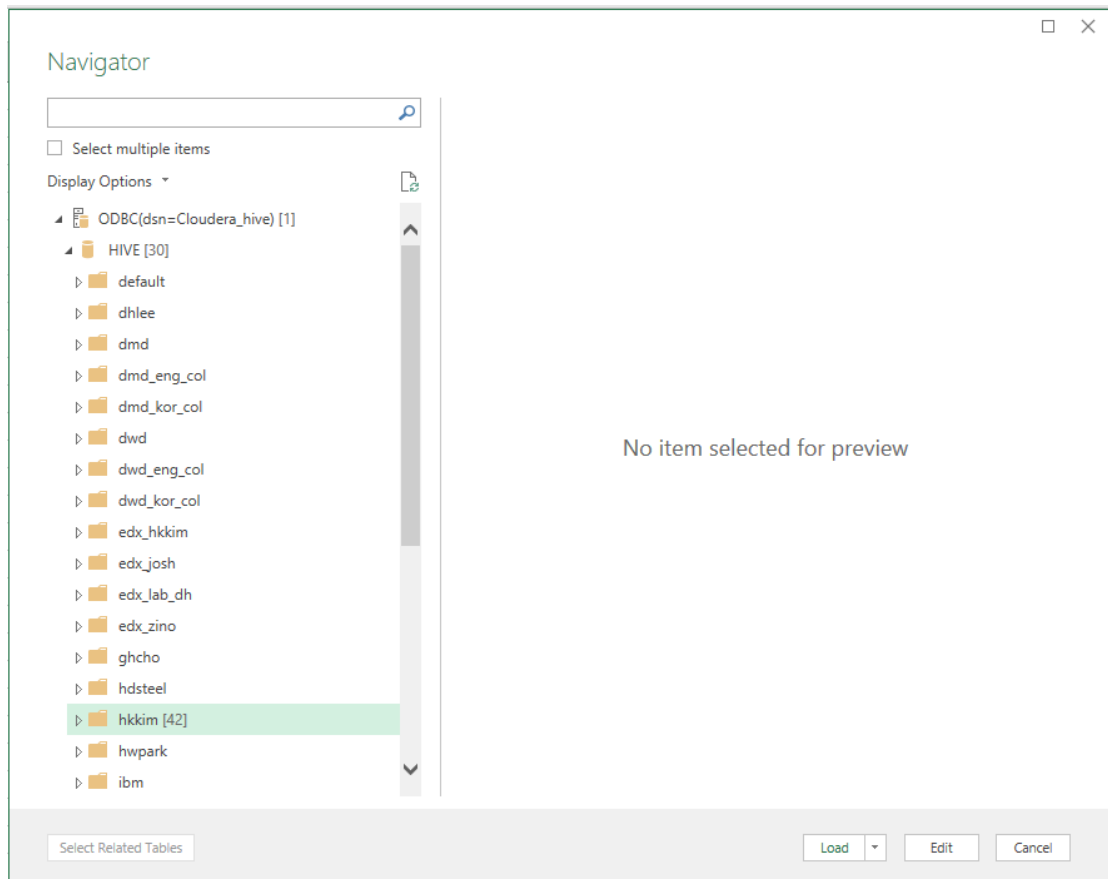
(Dashboard에 적용된 모든 그래프들)



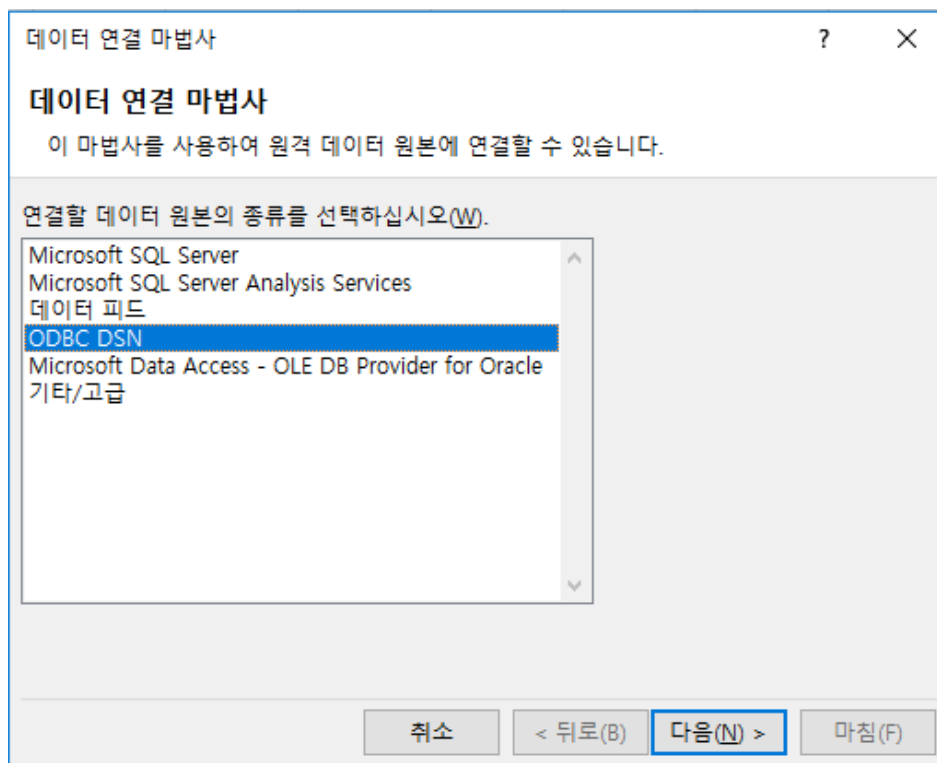
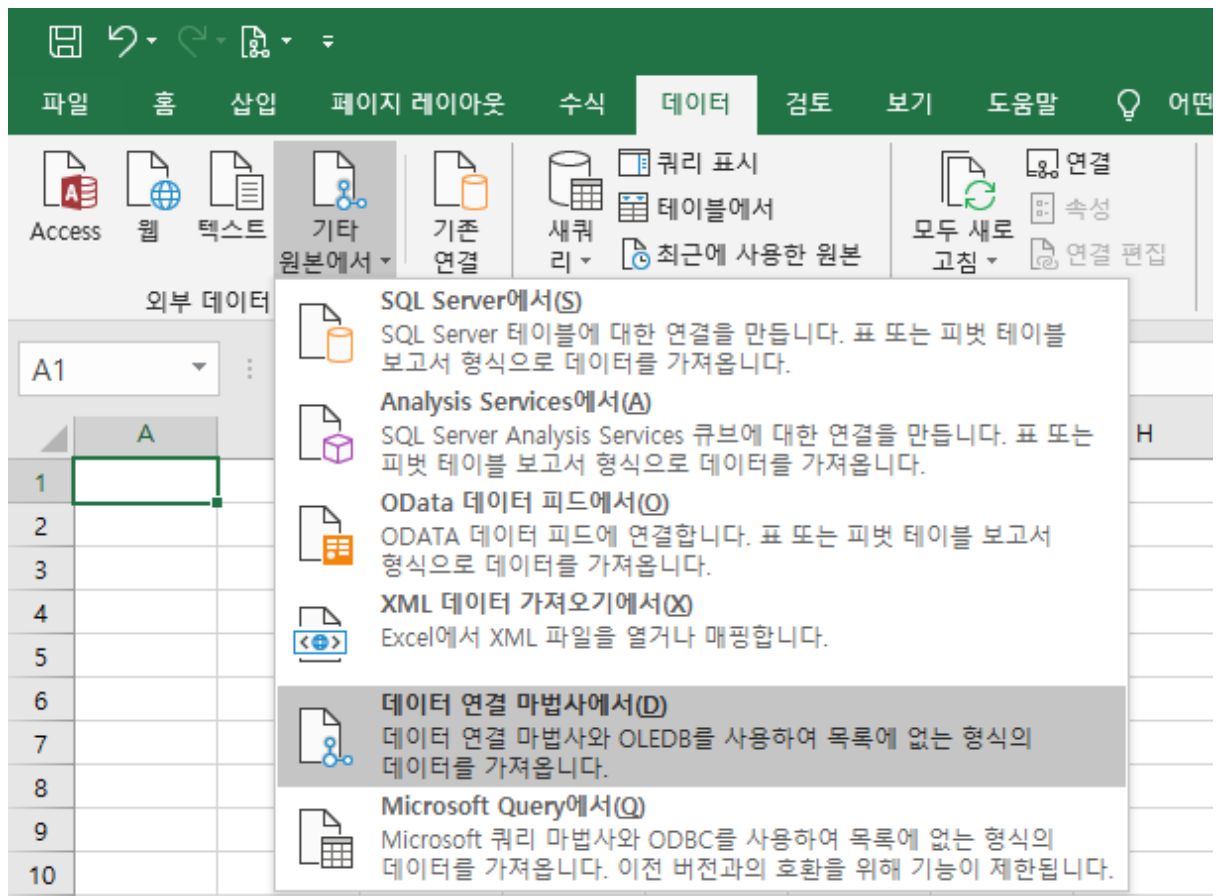
4. Excel 3D Map 사용 (부록 먼저 실행할 것!!)

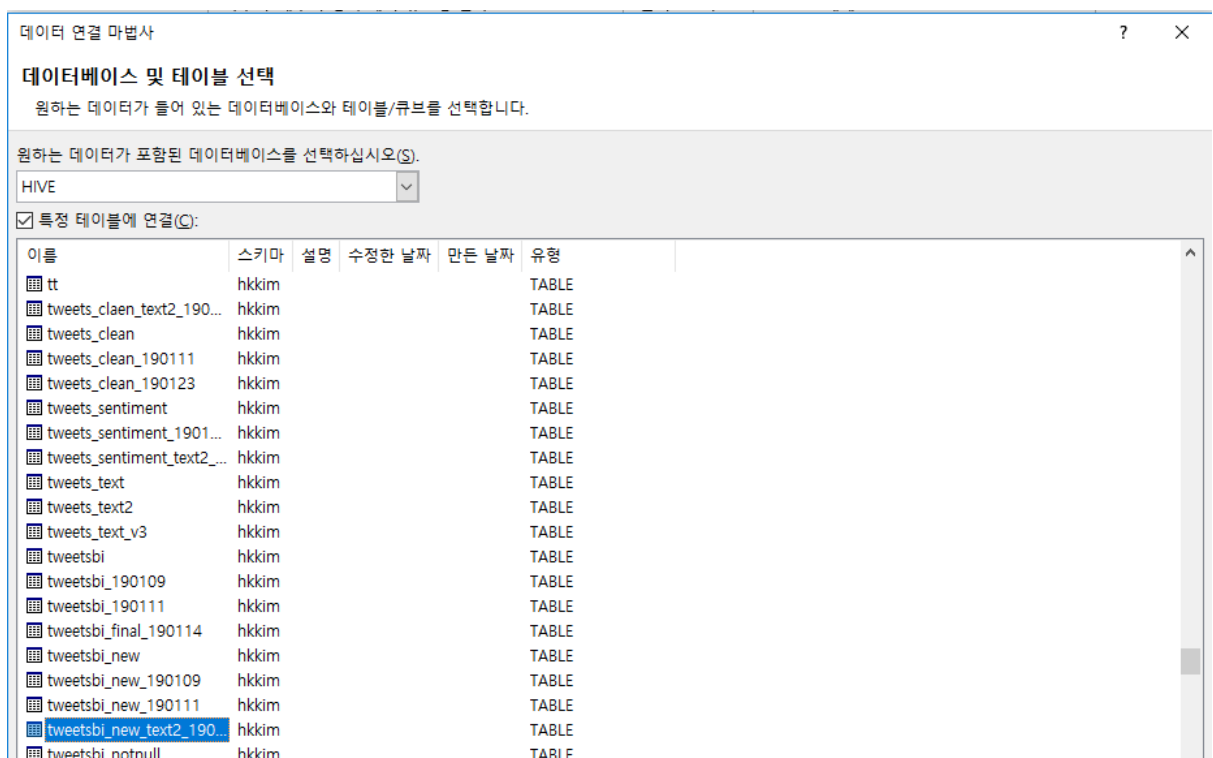
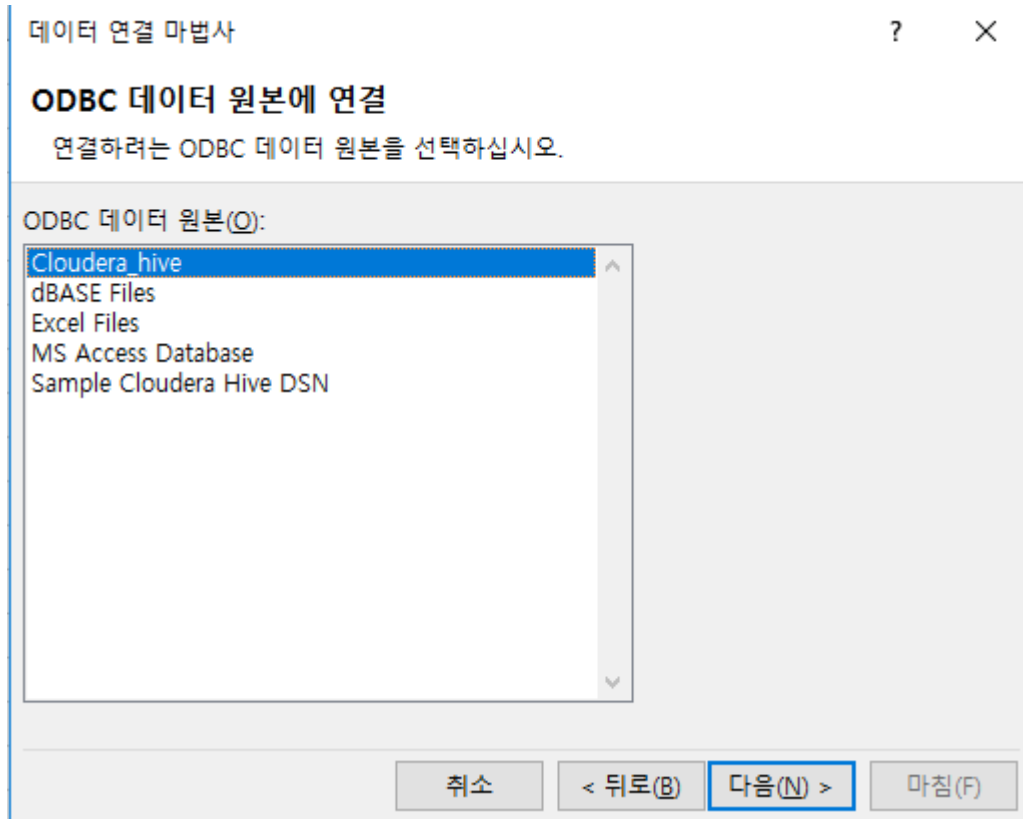
(1) Hive Table 가져오기 : “데이터 탭” -> “새 쿼리” -> “기타 원본에서” -> “ODBC에서”





(2) Hive Table 가져오기 : “데이터 탭” -> “기타 원본에서” - > “데이터 연결 마법사에서”

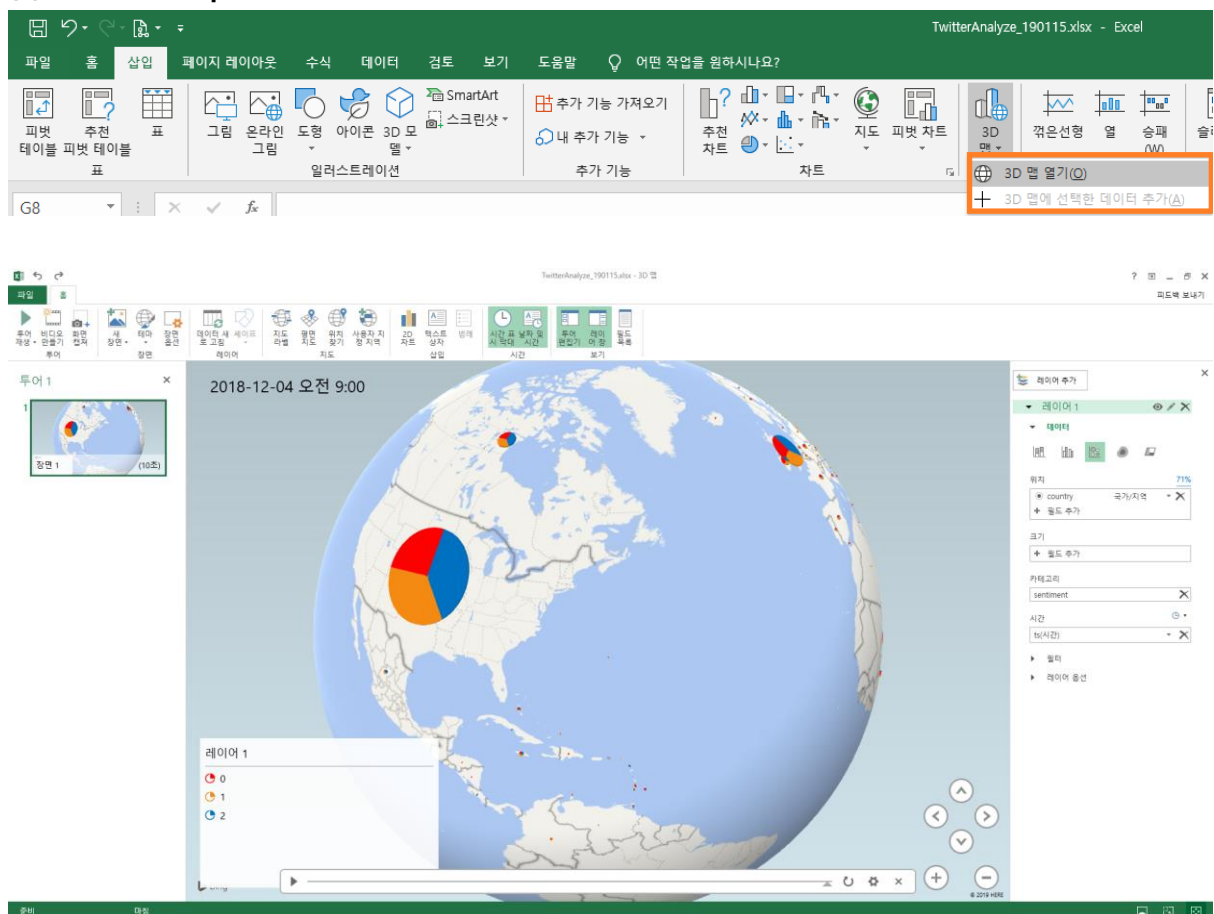


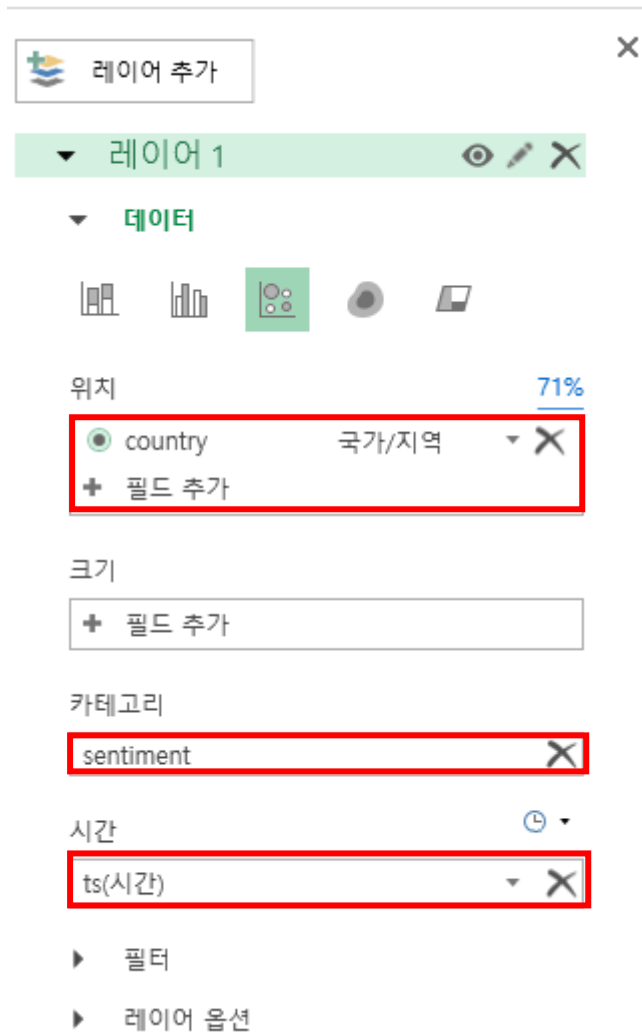


결과 : 데이터를 Excel에 Import 했을 시의 모습

tweet id	ts	msg	country	sentiment	lang
106696060680253440	2018-11-26 17:05	Today I washed my face with my moisturizer. Twice.	United States	1	en
106699482456543744	2018-11-26 18:59	That time it was #Shoaib and #Sami in this time it is Yasir alone.		1	en
1067000156975964160	2018-11-26 19:20	What a story!! 📖	United Kingdom	1	en
1067029410497421317	2018-11-26 21:16	I won't make the same mistakes twice anymore	United States	0	en
1067034911192825857	2018-11-26 21:38	Shhh 🤫 please don't let @TheBlueProphet see this ... lol	Ghana	2	en
1067035006596333568	2018-11-26 21:39	@daLionTamer Yes. Wouldn't think twice.	Hong Kong	2	en
1067051835876392960	2018-11-26 22:45	Thanks and an honour to be part of #C4PV making the Conservative case for #PeoplesVote	United Kingdom	1	en
1067052427327741952	2018-11-26 22:48	My gran saw @BigGirlsDC on Saturday night and is still raving about it. Proudly shown me he	United Kingdom	0	en
1067063714413604864	2018-11-26 23:33	@nellytson @17HashBrownsPls These days he might have thought twice. Can you say The Tor	United States	2	en
1067064891100549121	2018-11-26 23:37	Just watch the game film and see how bad FSU's D played. Peep the busted coverages, how t	United States	2	en
1067083902508880664	2018-11-27 0:53	I normally don't market, but I'm incredibly happy with the quality f stickers made by @Sticker	United States	2	en
106708422630842368	2018-11-27 1:11	@profdbcohen Also that. But the coaching situation had to be changed. Embarrassing losses t	United States	0	en
1067088495213244416	2018-11-27 1:11	You can't make the same mistake twice. The second time you make it, it's no longer a mistake	Germany	0	en
1067112382084861952	2018-11-27 2:46	Might be time to reinforce the Minuteman movement! Organized, United, Strong. Ethical	Volunteer United States	2	en
1067116381013868548	2018-11-27 3:02	【プレゼント】BTS、TWICE、IZ*ONE、Wanna 日本公演！12/12開催「2018 MAMA」日本公演に5日	日本	1	ja
106712217171672129	2018-11-27 3:25	Nasty juu is a walking joke , you got rocked on ur own side lines N started crying in front of e	United States	2	en
1067132115890962432	2018-11-27 4:04	Absolutely fantastic	United Kingdom	2	en
1067145749534818305	2018-11-27 4:59	That's a lot of homophobia for someone who would get their shit ran by dwight howard. Thin	United States	1	en
1067147963435225097	2018-11-27 5:07	Fuck @odum12_alex our summer wasn't good enough. 🙄🙄	United States	2	en
1067162382596755456	2018-11-27 6:05	Imagine paying for someone to go on holiday but they can't even hold off a gig /: of someon	United Kingdom	2	en
1067163531991269376	2018-11-27 6:09	Had to look twice. I'd say the font is a little too close together. Would make an excellent	stock United States	0	en
1067166356422955009	2018-11-27 6:21	I've escaped death twice now	The Netherlands	1	en
1067171716894474240	2018-11-27 6:42	and the eagles won both games. isn't that the part that matters?	United States	0	en
1067177692716589058	2018-11-27 7:06	Really thought this was just me for ages. @TonyMartiallah	United Kingdom	2	en
1067202785249293684	2018-11-27 8:45	@Roggy_2016 I told koe I just wanted my hair done BUT I also want this really pretty ring	that United States	2	en
1067204108057305088	2018-11-27 8:51	I drop my phone and got it replaced twice (2x!!!!) this weekend should i just get a pink motor	United States	2	en
1067209020866363393	2018-11-27 9:10	@elonmusk @margrethempossi Sounds like a permanent, manic state. The light that burns twic	United States	2	en
10672120918419800064	2018-11-27 9:11	@Bungie fix this please I got error code guitar twice tonight 😊 https://t.co/DbSAhTqwGw	Ireland	0	en
1067210619475148800	2018-11-27 9:16	don't gotta tell me twice 😊	United States	1	en
1067221204182937600	2018-11-27 9:58	Eh blue pound	United Kingdom	1	en
1067222671451201536	2018-11-27 10:04	@Blondiegrl @nytimes Hard to imagine this to happen. Once a candidate lost he or she has	United States	0	en
1067230750255267841	2018-11-27 10:36	@AbbyRose/ I've been at myrtle where the j/m intersect for 6 months now but i've only taken	United States	1	en
1067237365985488896	2018-11-27 11:03	@BravenaBlog oh dear God. I mean, I love me some beto. Think he's super. But now I have	United States	2	en

(3) Excel 3D Map 사용





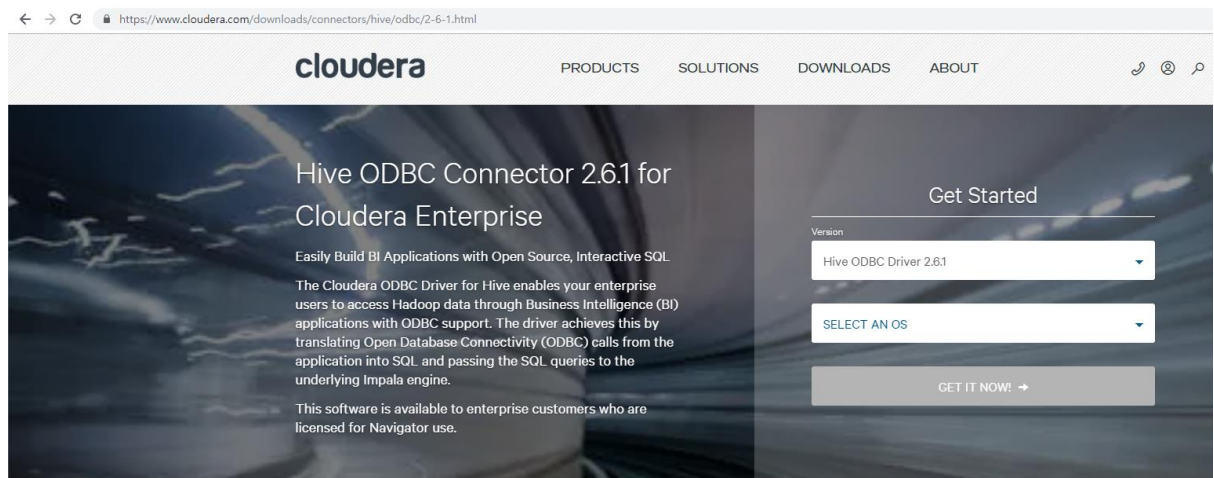
위치 : 각 나라별 위치를 뜻한다,

sentiment : 각 나라별로 응답의 긍정,중립,부정의 정도를 나타낸다.

ts : 시간의 흐름에 따라 각 나라별 sentiment의 정도의 변화를 볼 수 있다.

참고 : <https://github.com/nberah/tutorials/blob/hdp/tutorials/hortonworks/analyzing-social-media-sentiment-data/tutorial.md>

+ 부록1 : Cloudera ODBC Connector 설치하는 방법



다운로드 경로 : <https://www.cloudera.com/downloads/connectors/hive/odbc/2-6-1.html>

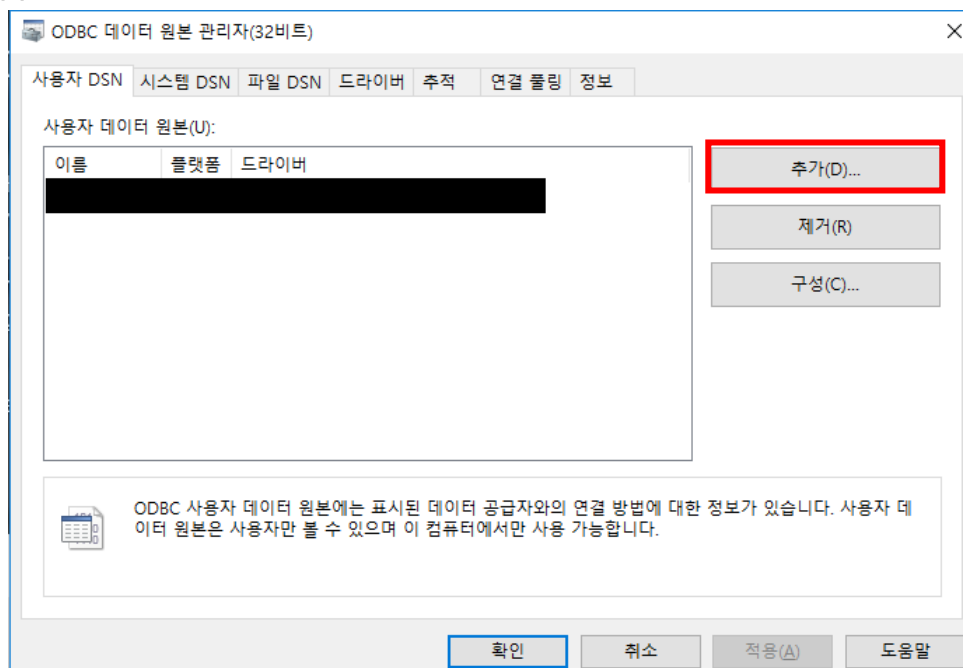
(32bit, 64bit 2가지가 있는데 본인에게 알맞은 bit로 선택한다.)

작성자의 경우에는 32bit 버전을 다운받았다.)

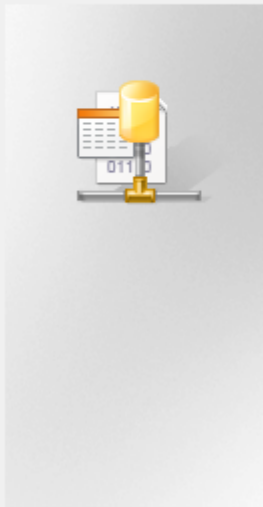
(1) 다운을 받아 ClouderaHiveODBC32(2.6).msi 파일을 열어 설치 진행.

(2) “시작” -> “ODBC 데이터 원본 (32비트)”

(3) 아래와 같이 진행



새 데이터 원본 만들기



데이터 원본을 설정할 드라이버를 선택하십시오(S).

이름	버
Cloudera ODBC Driver for Apache Hive	2.0
Driver da Microsoft para arquivos texto (*.txt; *.csv)	10
Driver do Microsoft Access (*.mdb)	10
Driver do Microsoft dBase (*.dbf)	10
Driver do Microsoft Excel(*.xls)	10
Driver do Microsoft Paradox (*.db)	10
IBM DB2 ODBC DRIVER	11
IBM DB2 ODBC DRIVER - DB2COPY1	11

< 뒤로(B)

마침

취소

Cloudera ODBC Driver for Apache Hive DSN Setup



Data Source Name:

Description:

Hive Server Type:

Service Discovery Mode:

Host(s):

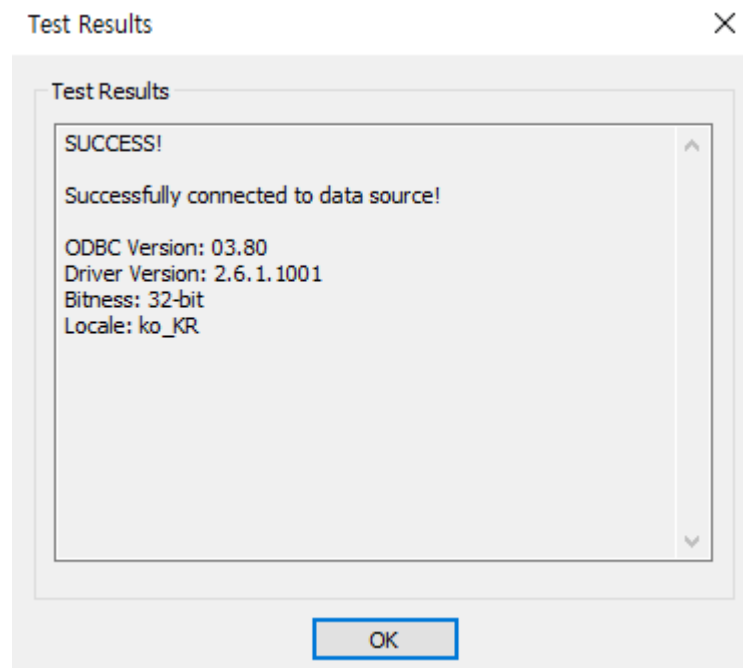
Port:

Database:

ZooKeeper Namespace:

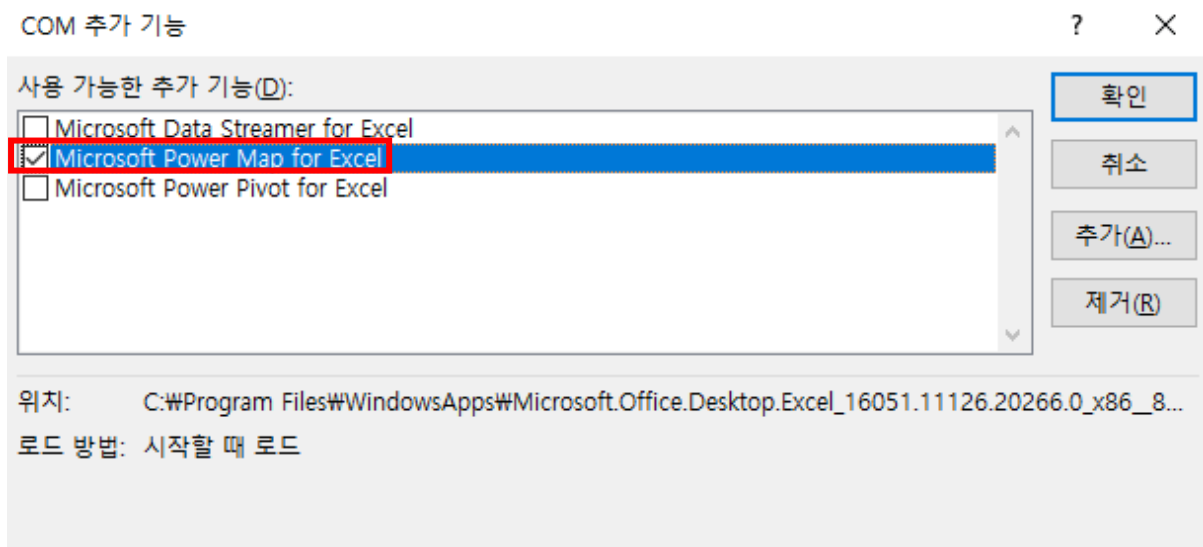
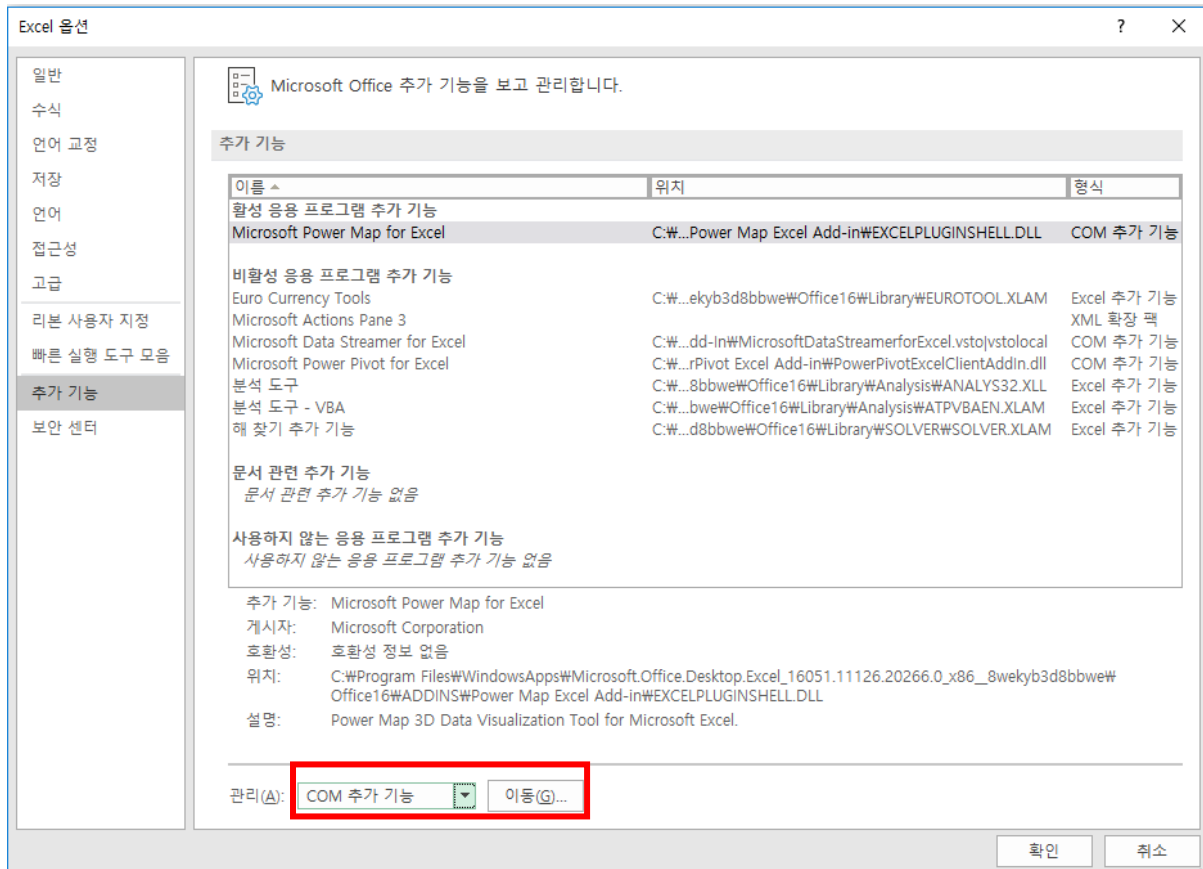
테스트 결과가 아래와 같이 나오면 연결 성공.

Excel “데이터” 탭에서 Hive테이블로 연결하고 싶을 시, Cloudera Hive Driver가 목록에 뜬다.



+ 부록2 : Excel 3D Map 활성화

“파일” -> “옵션” 을 클릭하면 아래와 같이 나온다.



Reference : <https://github.com/nberah/tutorials/blob/hdp/tutorials/hortonworks/analyzing-social-media-sentiment-data/tutorial.md>