Date

Loading unstructured data

# **Goal**

In this assignment, we will be loading unstructured.

### You will understand how to parse and load unstructured data to snowflake.

# **Preparation**

Let’s assume a business scenario, where you got twitter data in aws s3.

Your job is to load this data to snowflake database.

Data can be found in below location,

<https://drive.google.com/file/d/12dWBwgWLReXncB5omUYi7jF_5ekDm_Yb/view?usp=sharing>

Your job is to,

#### Copy this data to aws s3.

#### From s3 copy this data to snowflake table.

#### Structure of the table is as below,

CREATE OR REPLACE TRANSIENT TABLE TWITTER\_PARSED

(

OID VARCHAR(16777216) ,

TEXT VARCHAR(16777216) ,

USER\_NAME VARCHAR(16777216) ,

FOLLOWERS\_COUNT NUMBER(38,0) ,

FRIENDS\_COUNT NUMBER(38,0) ,

LOCATION VARCHAR(16777216) ,

DESCRIPTION VARCHAR(16777216)

)

**Sample data,**

**create or replace table twitter (v variant )**

insert into twitter

select

parse\_json('{

"\_id":{

"$oid":"59435062a7986c085b072088"

},

"text":"@shaqdarcy hahaha.. soya talaga bro.. :)",

"in\_reply\_to\_user\_id\_str":"238156878",

"id\_str":"133582462910611456",

"contributors":null,

"in\_reply\_to\_user\_id":238156878,

"created\_at":"Mon Nov 07 16:31:55 +0000 2011",

"in\_reply\_to\_status\_id":{

"$numberLong":"133582357465792513"

},

"entities":{

"hashtags":[

],

"user\_mentions":[

{

"screen\_name":"shaqdarcy",

"indices":[

0,

10

],

"id\_str":"238156878",

"name":"Darcy Nicolas",

"id":238156878

}

],

"urls":[

]

},

"geo":null,

"source":"web",

"place":null,

"favorited":false,

"truncated":false,

"coordinates":null,

"retweet\_count":0,

"in\_reply\_to\_screen\_name":"shaqdarcy",

"user":{

"profile\_use\_background\_image":true,

"favourites\_count":13,

"screen\_name":"JaybeatBolido",

"id\_str":"255006912",

"default\_profile\_image":false,

"geo\_enabled":false,

"profile\_text\_color":"333333",

"statuses\_count":467,

"profile\_background\_image\_url":"http://a0.twimg.com/images/themes/theme1/bg.png",

"created\_at":"Sun Feb 20 13:31:09 +0000 2011",

"friends\_count":92,

"profile\_link\_color":"0084B4",

"description":"I want to be a JEDI.",

"follow\_request\_sent":null,

"lang":"en",

"profile\_image\_url\_https":"https://si0.twimg.com/profile\_images/1614465172/jp\_normal.jpg",

"profile\_background\_color":"C0DEED",

"url":null,

"contributors\_enabled":false,

"profile\_background\_tile":false,

"following":null,

"profile\_sidebar\_fill\_color":"DDEEF6",

"protected":false,

"show\_all\_inline\_media":false,

"listed\_count":1,

"location":"Phillipines-Manila",

"name":"Japhette Pulido",

"is\_translator":false,

"default\_profile":true,

"notifications":null,

"profile\_sidebar\_border\_color":"C0DEED",

"id":255006912,

"verified":false,

"profile\_background\_image\_url\_https":"https://si0.twimg.com/images/themes/theme1/bg.png",

"time\_zone":null,

"utc\_offset":null,

"followers\_count":42,

"profile\_image\_url":"http://a1.twimg.com/profile\_images/1614465172/jp\_normal.jpg"

},

"retweeted":false,

"id":{

"$numberLong":"133582462910611456"

},

"in\_reply\_to\_status\_id\_str":"133582357465792513"

} ')

# **Approach**

In our previous demo, we had already gone through similar exercise on **book data**.

Keeping that in mind, can you define your approach to load this data to snowflake table ?

# **Solution**

Create,

#### File format object.

#### Stage object.

#### Execute copy command.

Paste your solution below,