Date

Loading json data

# **Goal**

In this assignment, we will try to understand approaches to load unstructured data in snowflake.

### Upload data to s3.

### Push it to snowflake stage.

### Load data to snowflake.

# **Preparation**

Please download json data from below location,

<https://drive.google.com/file/d/184SsnWNUfyoksqjXwl6WodXG2rzn0YNM/view?usp=sharing>

Once downloaded, upload data to your s3 location.

I assume, you have,

### Integration object.

### Stage object.

### Json file format object.

**Write command to upload data from local to s3 using aws cli.**

**aws s3 cp <your local path> s3://<your s3 path>**

**Create json file format,**

**CREATE OR REPLACE FILE FORMAT "DEMO\_DB"."PUBLIC".JSON\_FORMAT**

**type=JSON**

**Create table,**

**CREATE OR REPLACE TRANSIENT TABLE BOOK\_JSON\_DATA**

**(**

**OID VARCHAR,**

**AUTHOR VARCHAR,**

**TITLE VARCHAR,**

**BOOKTITLE VARCHAR,**

**YEAR NUMBER,**

**TYPE VARCHAR)**

# **Load data**

As we have discussed earlier in the class, we can take below approaches to load unstructured data in snowflake,

#### Load directly to variant column from staging area and parse directly.

#### Load to variant column as staging table and parse and load to structured table.

#### Parse and load from external stage to snowflake internal stage and then copy to structured table.

We have also discussed advantages and disadvantages of each approach.

Let’s take the 3rd approach to load data to snowflake table.

**Sample data,**

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

insert into book

select

parse\_json('{

"\_id":{

"$oid":"595c2c59a7986c0872002043"

},

"mdate":"2017-05-24",

"author":[

"Injoon Hong",

"Seongwook Park",

"Junyoung Park",

"Hoi-Jun Yoo"

],

"ee":"https://doi.org/10.1109/ASSCC.2015.7387453",

"booktitle":"A-SSCC",

"title":"A 1.9nJ/pixel embedded deep neural network processor for high speed visual attention in a mobile vision recognition SoC.",

"pages":"1-4",

"url":"db/conf/asscc/asscc2015.html#HongPPY15",

"year":"2015",

"type":"inproceedings",

"\_key":"conf::asscc::HongPPY15",

"crossref":[

"conf::asscc::2015"

]

}')

**Try this command below,**

**select**

**$1:"\_id":"$oid" as OID,**

**$1:"author"::array as AUTHOR ,**

**$1:"title"::string as TITLE,**

**$1:"booktitle"::string as BOOKTITLE ,**

**$1:"year"::int as YEAR,**

**$1:"type"::string as TYPE**

**from @MY\_S3\_UNLOAD\_STAGE/dblp.json**

**(file\_format =>JSON\_FORMAT)**

**What is your observation,**

Let’s try to copy this parsed data directly to snowflake table,

**copy into BOOK\_JSON\_DATA**

**from(**

**select**

**$1:"\_id":"$oid" as OID,**

**$1:"author"::array as AUTHOR ,**

**$1:"title"::string as TITLE,**

**$1:"booktitle"::string as BOOKTITLE ,**

**$1:"year"::int as YEAR,**

**$1:"type"::string as TYPE**

**from @MY\_S3\_UNLOAD\_STAGE/dblp.json**

**(file\_format =>JSON\_FORMAT)**

**)**

**ON\_ERROR='CONTINUE'**

**What is your observation ?**

**Let’s copy data first to table stage of snowflake internal stage,**

**copy into @%BOOK\_JSON\_DATA**

**from(**

**select**

**$1:"\_id":"$oid" as OID,**

**$1:"author"::string as AUTHOR ,**

**$1:"title"::string as TITLE,**

**$1:"booktitle"::string as BOOKTITLE ,**

**$1:"year"::string as YEAR,**

**$1:"type"::string as TYPE**

**from @MY\_S3\_UNLOAD\_STAGE/dblp.json**

**(file\_format =>JSON\_FORMAT)**

**)**

**Copy from table stage to snowflake table,**

**copy into BOOK\_JSON\_DATA**

**from @%BOOK\_JSON\_DATA**

**file\_format =(type='csv')**

**ON\_ERROR='CONTINUE'**

**Something wrong with two step copy process. Can you mention it below,**

**Truncate the table and again execute copy command below,**

**TRUNCATE TABLE BOOK\_JSON\_DATA;**

**copy into BOOK\_JSON\_DATA**

**from @%BOOK\_JSON\_DATA**

**file\_format =(type='csv')**

**ON\_ERROR='CONTINUE'**

**PURGE = TRUE**

**Now collect all rejected records and store it in separate table.**

**Create or replace json\_reject\_records**

**As**

**select \* from table(validate(BOOK\_JSON\_DATA , job\_id=>'<query\_id>'));**