ARCHITECTING AND DESIGNING A DATA WAREHOUSE FOR REPORTING AND OLAP

AN ETL PROJECT USING YELP AND CLIMATE DATA

NAME: **PIERRE MORA ERAZO**

DATE: **06/15/2024**

PROJECT OVERVIEW

Objective: To architect and design a Data Warehouse (DWH) for the purpose of reporting and Online Analytical Processing (OLAP) using skills acquired in the Designing Data Systems course.

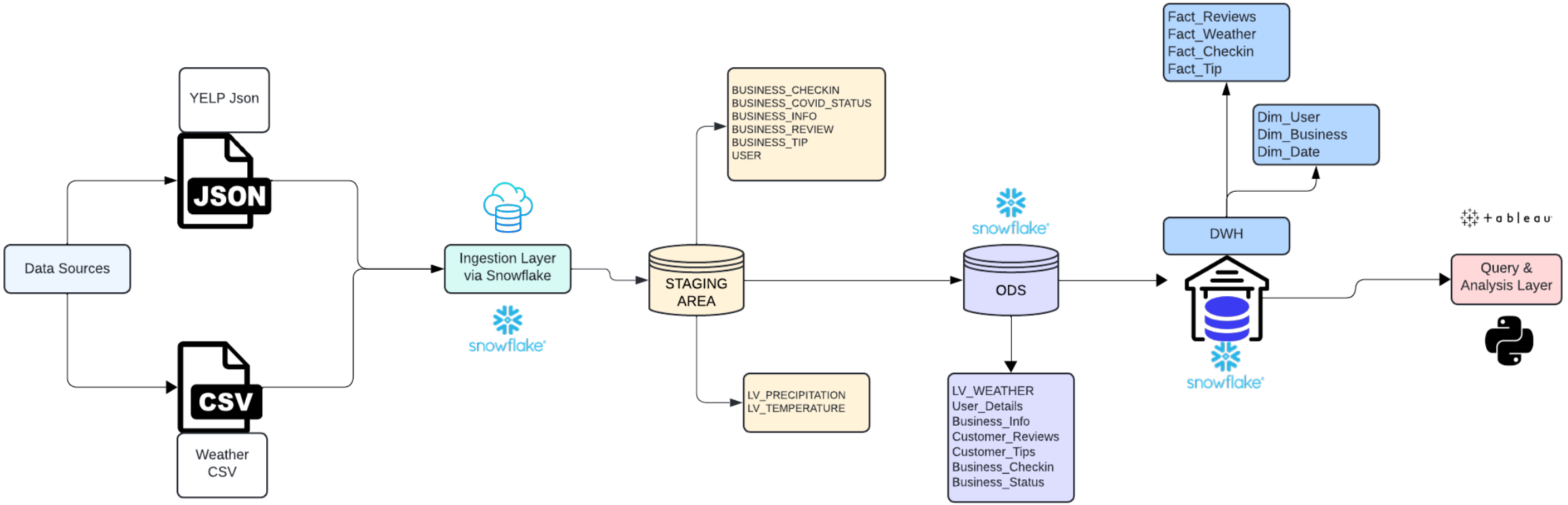
Data Sources:

- **Yelp Data:** Business information, reviews, tips, and check-ins.
- **Climate Data:** Temperature and precipitation observations.

Tools Used: Snowflake for data storage and transformation.

DATA ARCHITECTURE FOR ETL PROCESS

- Ingesting and Migrating Data from Sources to Data Warehouse



CREATING THE STAGING SCHEMA

- Creation of the staging environment in Snowflake and the upload process for Yelp and climate data:

```
PMORAER#COMPUTE_WH@(no database).(no schema)>USE DATABASE YELP_COVID;
+----------------------------------+
| status                           |
|----------------------------------|
| Statement executed successfully. |
+----------------------------------+
1 Row(s) produced. Time Elapsed: 0.114s
PMORAER#COMPUTE_WH@YELP_COVID.PUBLIC>CREATE SCHEMA STAGING;
+-----------------------------------+
| status                            |
|-----------------------------------|
| Schema STAGING successfully created. |
+-----------------------------------+
```

```
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>CREATE OR REPLACE FILE FORMAT mycsvformat
                                      TYPE = 'CSV'
                                      COMPRESSION = 'AUTO'
                                      FIELD_DELIMITER = ','
                                      RECORD_DELIMITER = '\n'
                                      SKIP_HEADER = 1
                                      ERROR_ON_COLUMN_COUNT_MISMATCH = TRUE
                                      NULL_IF = ('NULL', 'null')
                                      EMPTY_FIELD_AS_NULL = TRUE;

+--------------------------------------------+
| status                                     |
|--------------------------------------------|
| File format MYCSVFORMAT successfully created. |
+--------------------------------------------+
1 Row(s) produced. Time Elapsed: 0.243s
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>CREATE OR REPLACE STAGE my_csv_stage
                                      FILE_FORMAT = mycsvformat;

+--------------------------------------------+
| status                                     |
|--------------------------------------------|
| Stage area MY_CSV_STAGE successfully created. |
+--------------------------------------------+
```

```
1 Row(s) produced. Time Elapsed: 1.340s
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>create or replace file format myjsonformat type = 'JSON' strip_o
                                                                    uter_array = true;
001003 (42000): SQL compilation error:
syntax error line 2 at position 48 unexpected 'uter_array'.
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>create or replace file format myjsonformat type = 'JSON'strip_outer_array = true;
+--------------------------------------------+
| status                                     |
|--------------------------------------------|
| File format MYJSONFORMAT successfully created. |
+--------------------------------------------+
1 Row(s) produced. Time Elapsed: 0.170s
```

UPLOADING CSV FILES TO STAGING SCHEMA

- Creating LV_TEMPERATURE table and uploading the CSV. File to the staging schema:

```
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>CREATE TABLE "LV_TEMPERATURE"(
                                     "DATE" VARCHAR(8),
                                     "MIN_TEMP" VARCHAR(10),
                                     "MAX_TEMP" VARCHAR(10),
                                     "NORMAL_MIN_TEMP" FLOAT,
                                     "NORMAL_MAX_TEMP" FLOAT);
+-----------------------------------------+
| status                                  |
|-----------------------------------------|
| Table LV_TEMPERATURE successfully created. |
+-----------------------------------------+
1 Row(s) produced. Time Elapsed: 0.700s
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>PUT file:///Users/pierremora/Desktop/Data_engineering/usw00023169-temperature-degreef.csv  @MY_CSV_STAGE AUTO_COMPRESS=TRUE;
+-------------------------------+-----------------------------------------+-------------+-------------+--------------------+--------------------+----------+---------+
| source                        | target                                  | source_size | target_size | source_compression | target_compression | status   | message |
|-------------------------------+-----------------------------------------+-------------+-------------+--------------------+--------------------+----------+---------|
| usw00023169-temperature-degreef.csv | usw00023169-temperature-degreef.csv.gz |      816174 |      196592 | NONE               | GZIP               | UPLOADED |         |
+-------------------------------+-----------------------------------------+-------------+-------------+--------------------+--------------------+----------+---------+
1 Row(s) produced. Time Elapsed: 1.159s
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>COPY INTO "LV_TEMPERATURE" FROM @MY_CSV_STAGE/usw00023169-temperature-degreef.csv.gz file_format=mycsvformat ON_ERROR = 'CONTINUE' PURGE = TRUE;
+-----------------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-------------------+-------------------------+
| file                                          | status | rows_parsed | rows_loaded | error_limit | errors_seen | first_error | first_error_line | first_error_character | first_error_column_name |
|-----------------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-------------------+-------------------------+
| my_csv_stage/usw00023169-temperature-degreef.csv.gz | LOADED |       28241 |       28241 |       28241 |           0 | NULL        |              NULL |                    NULL | NULL                    |
+-----------------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-------------------+-------------------------+
```

- Creating LV_PRECIPITATION table and uploading the CSV. File to the staging schema:

```
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>CREATE TABLE "LV_PRECIPITATION"(
                                        "DATE" VARCHAR(8),
                                        "PRECIPITATION" VARCHAR(5),
                                        "PRECIPITATION_NORMAL" FLOAT
                                    );

+-----------------------------------------+
| status                                  |
|-----------------------------------------|
| Table LV_PRECIPITATION successfully created. |
+-----------------------------------------+
```

```
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>PUT file:///Users/pierremora/Desktop/Data_engineering/usw00023169-las-vegas-mccarran-intl-ap-precipitation-inch.csv  @MY_CSV_STAGE AUTO_COMPRESS=TRUE;
+-----------------------------------------------+-----------------------------------------------+-------------+-------------+--------------------+--------------------+----------+---------+
| source                                        | target                                        | source_size | target_size | source_compression | target_compression | status   | message |
|-----------------------------------------------+-----------------------------------------------+-------------+-------------+--------------------+--------------------+----------+---------|
| usw00023169-las-vegas-mccarran-intl-ap-precipitation-inch.csv | usw00023169-las-vegas-mccarran-intl-ap-precipitation-inch.csv.gz |      528165 |      118992 | NONE               | GZIP               | UPLOADED |         |
+-----------------------------------------------+-----------------------------------------------+-------------+-------------+--------------------+--------------------+----------+---------+
1 Row(s) produced. Time Elapsed: 1.147s
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>COPY INTO "LV_PRECIPITATION" FROM @MY_CSV_STAGE/usw00023169-las-vegas-mccarran-intl-ap-precipitation-inch.csv.gz file_format=mycsvformat ON_ERROR = 'CONTINUE' PURGE = TRUE;
+-----------------------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-------------------+-------------------------+
| file                                                | status | rows_parsed | rows_loaded | error_limit | errors_seen | first_error | first_error_line | first_error_character | first_error_column_name |
|-----------------------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-------------------+-------------------------+
| my_csv_stage/usw00023169-las-vegas-mccarran-intl-ap-precipitation-inch.csv.gz | LOADED |       28241 |       28241 |       28241 |           0 | NULL        |              NULL |                    NULL | NULL                    |
+-----------------------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-------------------+-------------------------+
```

UPLOADING JSON FILES TO STAGING SCHEMA

- Creating BUSINESS_COVID_STATUS and BUSINESS_INFO Tables and uploading their respective JSON files to the staging schema:

```
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>CREATE TABLE "BUSINESS_COVID_STATUS"("BUSINESS_JSON" VARIANT);
+-----------------------------------------------+
| status                                        |
|-----------------------------------------------|
| Table BUSINESS_COVID_STATUS successfully created. |
+-----------------------------------------------+
1 Row(s) produced. Time Elapsed: 0.466s
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>CREATE OR REPLACE STAGE my_json_stage
                                    FILE_FORMAT = myjsonformat;

+-----------------------------------------------+
| status                                        |
|-----------------------------------------------|
| Stage area MY_JSON_STAGE successfully created. |
+-----------------------------------------------+
1 Row(s) produced. Time Elapsed: 0.434s
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>PUT file:///Users/pierremora/Desktop/Data_engineering/yelp_academic_dataset_covid_features.json @my_json_stage AUTO_COMPRESS=TRUE;

+-----------------------------------+-----------------------------------------+-------------+-------------+--------------------+--------------------+----------+---------+
| source                            | target                                  | source_size | target_size | source_compression | target_compression | status   | message |
|-----------------------------------+-----------------------------------------+-------------+-------------+--------------------+--------------------+----------+---------|
| yelp_academic_dataset_covid_features.json | yelp_academic_dataset_covid_features.json.gz |    64835031 |     6954672 | NONE               | GZIP               | UPLOADED |         |
+-----------------------------------+-----------------------------------------+-------------+-------------+--------------------+--------------------+----------+---------+
1 Row(s) produced. Time Elapsed: 4.947s
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>copy into BUSINESS_COVID_STATUS from @my_json_stage/yelp_academic_dataset_covid_features.json.gz file_format=myjsonformat on_error='skip_file';
+-----------------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-------------------+-----------------------+-------------------------+
| file                                          | status | rows_parsed | rows_loaded | error_limit | errors_seen | first_error | first_error_line | first_error_character | first_error_column_name |
|-----------------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-------------------+-----------------------+-------------------------|
| my_json_stage/yelp_academic_dataset_covid_features.json.gz | LOADED |      209795 |      209795 |           1 |           0 | NULL        |              NULL |                  NULL | NULL                    |
+-----------------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-------------------+-----------------------+-------------------------+
1 Row(s) produced. Time Elapsed: 5.499s
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>CREATE TABLE "BUSINESS_INFO"("BUSINESS_INFO_JSON" VARIANT);
+-----------------------------------------+
| status                                  |
|-----------------------------------------|
| Table BUSINESS_INFO successfully created. |
+-----------------------------------------+
1 Row(s) produced. Time Elapsed: 0.661s
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>PUT file:///Users/pierremora/Desktop/Data_engineering/yelp_dataset/yelp_academic_dataset_business.json @my_json_stage AUTO_COMPRESS= TRUE parallel=4;
+-----------------------------------+-----------------------------------------+-------------+-------------+--------------------+--------------------+----------+---------+
| source                            | target                                  | source_size | target_size | source_compression | target_compression | status   | message |
|-----------------------------------+-----------------------------------------+-------------+-------------+--------------------+--------------------+----------+---------|
| yelp_academic_dataset_business.json | yelp_academic_dataset_business.json.gz |   118863795 |    20340592 | NONE               | GZIP               | UPLOADED |         |
+-----------------------------------+-----------------------------------------+-------------+-------------+--------------------+--------------------+----------+---------+
1 Row(s) produced. Time Elapsed: 13.454s
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>COPY INTO "BUSINESS_INFO" FROM @my_json_stage/yelp_academic_dataset_business.json.gz file_format=myjsonformat on_error= 'skip_file';
+-----------------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-------------------+-----------------------+-------------------------+
| file                                          | status | rows_parsed | rows_loaded | error_limit | errors_seen | first_error | first_error_line | first_error_character | first_error_column_name |
|-----------------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-------------------+-----------------------+-------------------------|
| my_json_stage/yelp_academic_dataset_business.json.gz | LOADED |      150346 |      150346 |           1 |           0 | NULL        |              NULL |                  NULL | NULL                    |
+-----------------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-------------------+-----------------------+-------------------------+
```

UPLOADING JSON FILES TO STAGING SCHEMA

- Creating BUSINESS_CHECKIN Table and uploading the respective JSON files to the staging schema:

```
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>CREATE TABLE "BUSINESS_CHECK_IN"("BUSINESS_CHECK_IN_JSON" VARIANT);
+----------------------------------------+
| status                                 |
|----------------------------------------|
| Table BUSINESS_CHECK_IN successfully created. |
+----------------------------------------+
1 Row(s) produced. Time Elapsed: 0.526s
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>PUT file:///Users/pierremora/Desktop/Data_engineering/yelp_dataset/yelp_academic_dataset_checkin.json @my_json_stage AUTO_COMPRESS= TRUE parallel=6;
+-----------------------------+-----------------------------------+-------------+-------------+--------------------+--------------------+----------+---------+
| source                      | target                            | source_size | target_size | source_compression | target_compression | status   | message |
|-----------------------------+-----------------------------------+-------------+-------------+--------------------+--------------------+----------+---------|
| yelp_academic_dataset_checkin.json | yelp_academic_dataset_checkin.json.gz |   286958945 |    80199328 | NONE               | GZIP               | UPLOADED |         |
+-----------------------------+-----------------------------------+-------------+-------------+--------------------+--------------------+----------+---------+
1 Row(s) produced. Time Elapsed: 87.528s
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>COPY INTO "BUSINESS_CHECK_IN" FROM my_json_stage/yelp_academic_dataset_checkin.json.gz file_format=myjsonformat on_error='skip_file';
001003 (42000): SQL compilation error:
syntax error line 1 at position 35 unexpected 'my_json_stage'.
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>COPY INTO "BUSINESS_CHECK_IN" FROM @my_json_stage/yelp_academic_dataset_checkin.json.gz file_format=myjsonformat on_error='skip_file';
+-----------------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-------------------+--------------------------+-------------------------+
| file                                          | status | rows_parsed | rows_loaded | error_limit | errors_seen | first_error | first_error_line | first_error_character | first_error_column_name |
|-----------------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-------------------+--------------------------+-------------------------|
| my_json_stage/yelp_academic_dataset_checkin.json.gz | LOADED |      131930 |      131930 |           1 |           0 | NULL        |              NULL |                     NULL | NULL                    |
+-----------------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-------------------+--------------------------+-------------------------+
1 Row(s) produced. Time Elapsed: 12.194s
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>CREATE TABLE "BUSINESS_REVIEW"("BUSINESS_REVIEW_JSON" VARIANT);
```

UPLOADING JSON FILES TO STAGING SCHEMA

- Creating BUSINESS_REVIEW Table and uploading the respective JSON file to the staging schema:

```
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>PUT file:///Users/pierremora/Desktop/Data_engineering/yelp_dataset/yelp_academic_dataset_review_4.json @my_json_stage AUTO_COMPRESS= TRUE parallel=6;
+-----------------------------------+-----------------------------------+-------------+-------------+--------------------+--------------------+----------+---------+
| source                            | target                            | source_size | target_size | source_compression | target_compression | status   | message |
|-----------------------------------+-----------------------------------+-------------+-------------+--------------------+--------------------+----------+---------|
| yelp_academic_dataset_review_4.json | yelp_academic_dataset_review_4.json.gz |   534186718 |   218442272 | NONE               | GZIP               | UPLOADED |         |
+-----------------------------------+-----------------------------------+-------------+-------------+--------------------+--------------------+----------+---------+
1 Row(s) produced. Time Elapsed: 104.867s
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>COPY INTO "BUSINESS_REVIEW" FROM @my_json_stage/yelp_academic_dataset_review_4.json.gz file_format=myjsonformat on_error='skip_file';
+-----------------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-----------------+------------------------+-------------------------+
| file                                          | status | rows_parsed | rows_loaded | error_limit | errors_seen | first_error | first_error_line | first_error_character | first_error_column_name |
|-----------------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-----------------+------------------------+-------------------------|
| my_json_stage/yelp_academic_dataset_review_4.json.gz | LOADED |      694506 |      694506 |           1 |           0 | NULL        |            NULL |                   NULL | NULL                    |
+-----------------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-----------------+------------------------+-------------------------+
1 Row(s) produced. Time Elapsed: 25.239s
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>PUT file:///Users/pierremora/Desktop/Data_engineering/yelp_dataset/yelp_academic_dataset_review_5.json @my_json_stage AUTO_COMPRESS= TRUE parallel=6;
+-----------------------------------+-----------------------------------+-------------+-------------+--------------------+--------------------+----------+---------+
| source                            | target                            | source_size | target_size | source_compression | target_compression | status   | message |
|-----------------------------------+-----------------------------------+-------------+-------------+--------------------+--------------------+----------+---------|
| yelp_academic_dataset_review_5.json | yelp_academic_dataset_review_5.json.gz |   534187004 |   218262784 | NONE               | GZIP               | UPLOADED |         |
+-----------------------------------+-----------------------------------+-------------+-------------+--------------------+--------------------+----------+---------+
1 Row(s) produced. Time Elapsed: 104.768s
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>COPY INTO "BUSINESS_REVIEW" FROM @my_json_stage/yelp_academic_dataset_review_5.json.gz file_format=myjsonformat on_error='skip_file';
+-----------------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-----------------+------------------------+-------------------------+
| file                                          | status | rows_parsed | rows_loaded | error_limit | errors_seen | first_error | first_error_line | first_error_character | first_error_column_name |
|-----------------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-----------------+------------------------+-------------------------|
| my_json_stage/yelp_academic_dataset_review_5.json.gz | LOADED |      693888 |      693888 |           1 |           0 | NULL        |            NULL |                   NULL | NULL                    |
+-----------------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-----------------+------------------------+-------------------------+
1 Row(s) produced. Time Elapsed: 24.990s
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>PUT file:///Users/pierremora/Desktop/Data_engineering/yelp_dataset/yelp_academic_dataset_review_6.json @my_json_stage AUTO_COMPRESS= TRUE parallel=6;
+-----------------------------------+-----------------------------------+-------------+-------------+--------------------+--------------------+----------+---------+
| source                            | target                            | source_size | target_size | source_compression | target_compression | status   | message |
|-----------------------------------+-----------------------------------+-------------+-------------+--------------------+--------------------+----------+---------|
| yelp_academic_dataset_review_6.json | yelp_academic_dataset_review_6.json.gz |   534187484 |   218223184 | NONE               | GZIP               | UPLOADED |         |
+-----------------------------------+-----------------------------------+-------------+-------------+--------------------+--------------------+----------+---------+
1 Row(s) produced. Time Elapsed: 105.146s
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>COPY INTO "BUSINESS_REVIEW" FROM @my_json_stage/yelp_academic_dataset_review_6.json.gz file_format=myjsonformat on_error='skip_file';
+-----------------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-----------------+------------------------+-------------------------+
| file                                          | status | rows_parsed | rows_loaded | error_limit | errors_seen | first_error | first_error_line | first_error_character | first_error_column_name |
|-----------------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-----------------+------------------------+-------------------------|
| my_json_stage/yelp_academic_dataset_review_6.json.gz | LOADED |      700857 |      700857 |           1 |           0 | NULL        |            NULL |                   NULL | NULL                    |
+-----------------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-----------------+------------------------+-------------------------+
1 Row(s) produced. Time Elapsed: 26.124s
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>PUT file:///Users/pierremora/Desktop/Data_engineering/yelp_dataset/yelp_academic_dataset_review_7.json @my_json_stage AUTO_COMPRESS= TRUE parallel=6;
+-----------------------------------+-----------------------------------+-------------+-------------+--------------------+--------------------+----------+---------+
| source                            | target                            | source_size | target_size | source_compression | target_compression | status   | message |
|-----------------------------------+-----------------------------------+-------------+-------------+--------------------+--------------------+----------+---------|
| yelp_academic_dataset_review_7.json | yelp_academic_dataset_review_7.json.gz |   534186339 |   218089664 | NONE               | GZIP               | UPLOADED |         |
+-----------------------------------+-----------------------------------+-------------+-------------+--------------------+--------------------+----------+---------+
1 Row(s) produced. Time Elapsed: 104.603s
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>COPY INTO "BUSINESS_REVIEW" FROM @my_json_stage/yelp_academic_dataset_review_7.json.gz file_format=myjsonformat on_error='skip_file';
+-----------------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-----------------+------------------------+-------------------------+
| file                                          | status | rows_parsed | rows_loaded | error_limit | errors_seen | first_error | first_error_line | first_error_character | first_error_column_name |
|-----------------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-----------------+------------------------+-------------------------|
| my_json_stage/yelp_academic_dataset_review_7.json.gz | LOADED |      695442 |      695442 |           1 |           0 | NULL        |            NULL |                   NULL | NULL                    |
+-----------------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-----------------+------------------------+-------------------------+
1 Row(s) produced. Time Elapsed: 25.107s
```

UPLOADING JSON FILES TO STAGING SCHEMA

- Creating BUSINESS_TIP AND USER Tables and uploading their respective JSON files to the staging schema:

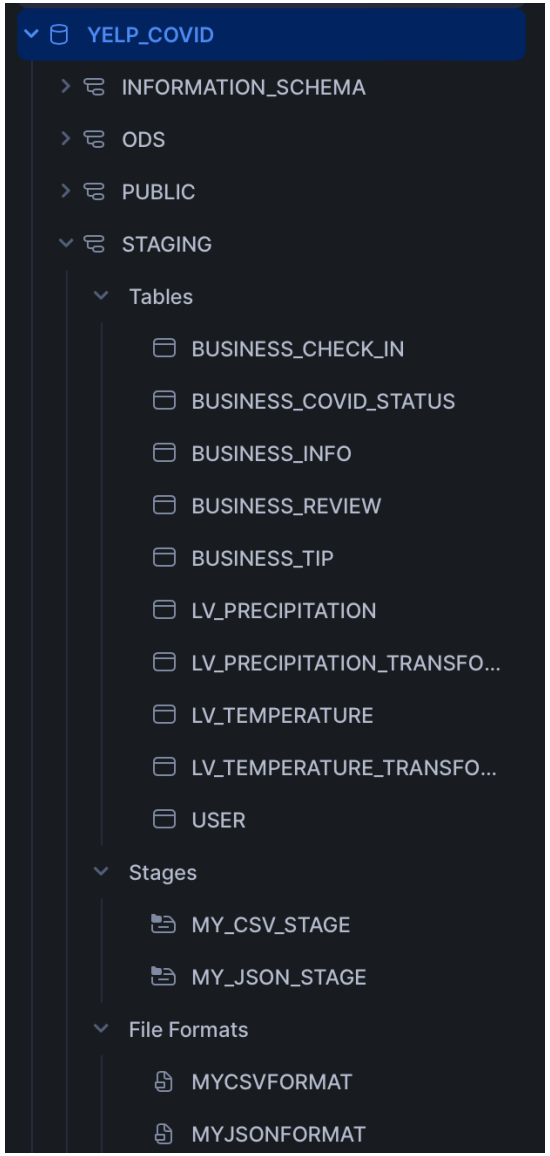
```
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>COPY INTO "BUSINESS_TIP" FROM @my_json_stage/yelp_academic_dataset_tip.json.gz file_format=myjsonformat on_error='skip_file';
+-----------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-----------------+---------------------+-----------------------+
| file                                    | status | rows_parsed | rows_loaded | error_limit | errors_seen | first_error | first_error_line | first_error_character | first_error_column_name |
|-----------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-----------------+---------------------+-----------------------|
| my_json_stage/yelp_academic_dataset_tip.json.gz | LOADED |      908915 |      908915 |           1 |           0 | NULL        |            NULL |                NULL | NULL                  |
+-----------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-----------------+---------------------+-----------------------+
1 Row(s) produced. Time Elapsed: 12.979s
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>CREATE TABLE "USER" ("USER_JSON" VARIANT);
+----------------------------------+
| status                           |
|----------------------------------|
| Table USER successfully created. |
+----------------------------------+
1 Row(s) produced. Time Elapsed: 1.212s
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>PUT file:///Users/pierremora/Desktop/Data_engineering/yelp_dataset/yelp_academic_dataset_user_01.json @my_json_stage AUTO_COMPRESS= TRUE parallel=9;
+---------------------------------+-----------------------------------+-------------+-------------+--------------------+--------------------+----------+---------+
| source                          | target                            | source_size | target_size | source_compression | target_compression | status   | message |
|---------------------------------+-----------------------------------+-------------+-------------+--------------------+--------------------+----------+---------|
| yelp_academic_dataset_user_01.json | yelp_academic_dataset_user_01.json.gz |   840838797 |   517441376 | NONE               | GZIP               | UPLOADED |         |
+---------------------------------+-----------------------------------+-------------+-------------+--------------------+--------------------+----------+---------+
1 Row(s) produced. Time Elapsed: 230.400s
PMORAER#COMPUTE_WH@YELP_COVID.STAGING>COPY INTO "USER" FROM @my_json_stage/yelp_academic_dataset_user_01.json.gz file_format=myjsonformat on_error='skip_file';
+-----------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-----------------+---------------------+-----------------------+
| file                                    | status | rows_parsed | rows_loaded | error_limit | errors_seen | first_error | first_error_line | first_error_character | first_error_column_name |
|-----------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-----------------+---------------------+-----------------------|
| my_json_stage/yelp_academic_dataset_user_01.json.gz | LOADED |      401784 |      401784 |           1 |           0 | NULL        |            NULL |                NULL | NULL                  |
+-----------------------------------------+--------+-------------+-------------+-------------+-------------+-------------+-----------------+---------------------+-----------------------+
1 Row(s) produced. Time Elapsed: 33.144s
```

- This slide demonstrates the creation of the staging environment in Snowflake and the upload process for Yelp and climate data, including handling large JSON files by splitting them into smaller chunks.

STAGING SCHEMA IN SNOWFLAKE

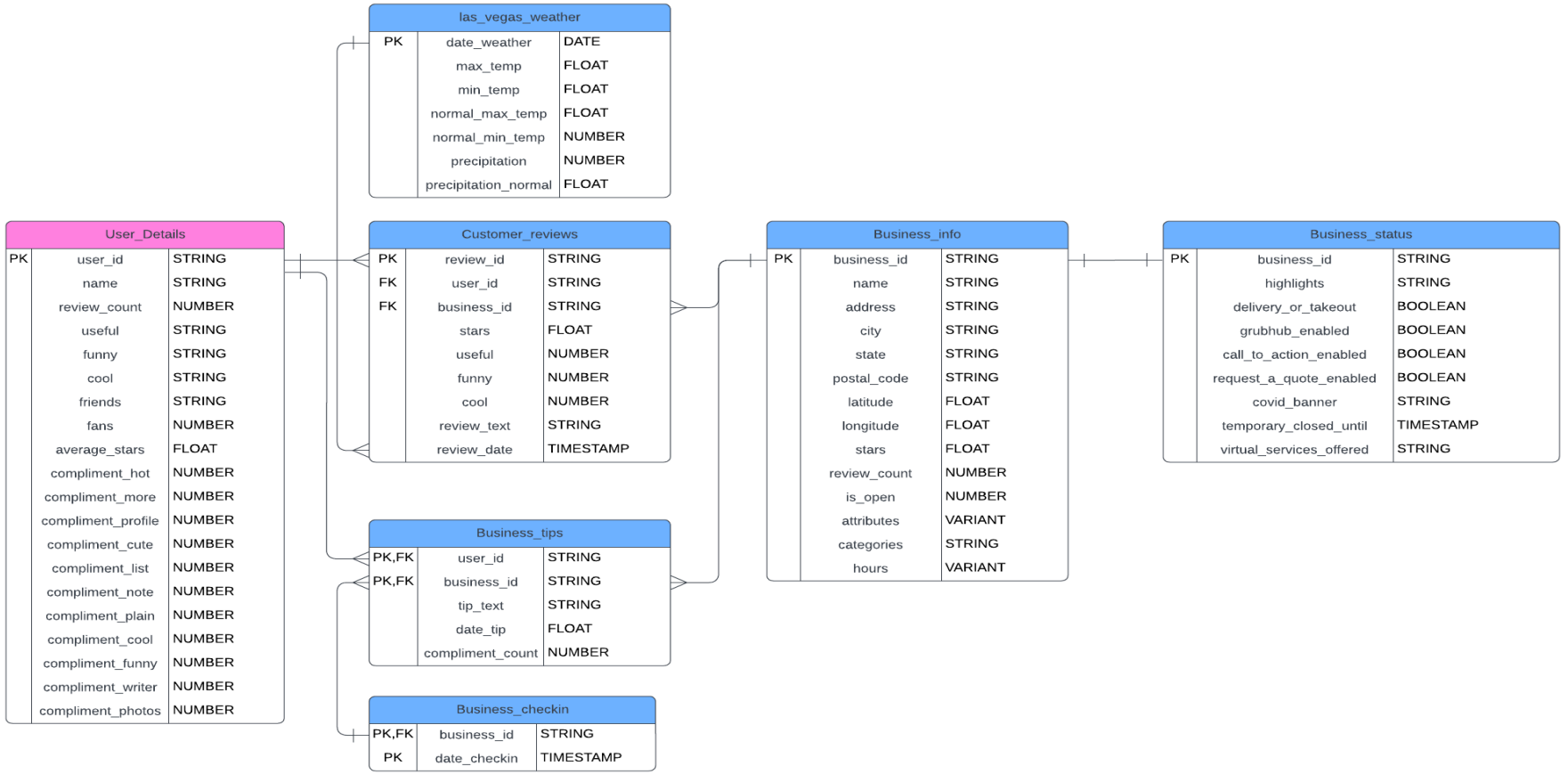
- Overview of the Created Staging Schema and tables

- This screenshot displays the Snowflake UI interface with the created staging schema and its respective tables, showcasing the initial setup before data transformation.



ENTITY-RELATIONSHIP (ER) DIAGRAM

- This ER diagram visualizes the relationships between different entities in the data structure, providing a clear overview of how Yelp and climate data are interconnected.



TRANSFORMING DATA FROM STAGING TO ODS

- This slide demonstrates the SQL queries used to transform data from staging to ODS, including the use of JSON functions to extract and transform data. It also provides a comparison of data sizes in raw files, staging tables, and ODS tables

```sql
CREATE TABLE User_Details(
user_id STRING PRIMARY KEY,
name STRING,
review_count NUMBER,
yelping_since TIMESTAMP,
useful STRING,
funny STRING,
cool STRING,
friends STRING,
fans NUMBER,
average_stars FLOAT,
compliment_hot NUMBER,
compliment_more NUMBER,
compliment_profile NUMBER,
compliment_cute NUMBER,
compliment_list NUMBER,
compliment_note NUMBER,
compliment_plain NUMBER,
compliment_cool NUMBER,
compliment_funny NUMBER,
compliment_writer NUMBER,
compliment_photos NUMBER );



INSERT INTO ODS.User_Details (user_id, name, review_count, yelping_since, useful, funny, cool, friends, fans, average_stars, compliment_hot, compliment_more, compliment_profile, compliment_cute,
compliment_list, compliment_note, compliment_plain, compliment_cool, compliment_funny, compliment_writer, compliment_photos)
SELECT
    "USER_JSON":user_id::STRING,
    "USER_JSON":name::STRING,
    "USER_JSON":review_count::NUMBER,
    "USER_JSON":yelping_since::TIMESTAMP,
    "USER_JSON":useful::NUMBER,
    "USER_JSON":funny::NUMBER,
    "USER_JSON":cool::NUMBER,
    "USER_JSON":friends::STRING,
    "USER_JSON":fans::NUMBER,
    "USER_JSON":average_stars::FLOAT,
    "USER_JSON":compliment_hot::NUMBER,
    "USER_JSON":compliment_more::NUMBER,
    "USER_JSON":compliment_profile::NUMBER,
    "USER_JSON":compliment_cute::NUMBER,
    "USER_JSON":compliment_list::NUMBER,
    "USER_JSON":compliment_note::NUMBER,
    "USER_JSON":compliment_plain::NUMBER,
    "USER_JSON":compliment_cool::NUMBER,
    "USER_JSON":compliment_funny::NUMBER,
    "USER_JSON":compliment_writer::NUMBER,
    "USER_JSON":compliment_photos::NUMBER
FROM STAGING.User;
```

```sql
CREATE TABLE Business_info(
business_id STRING PRIMARY KEY,
name STRING,
address STRING,
city STRING,
state STRING,
postal_code STRING,
latitude FLOAT,
longitude FLOAT,
stars FLOAT,
review_count NUMBER,
is_open NUMBER,
attributes VARIANT,
categories STRING,
hours VARIANT
);


INSERT INTO ODS.Business_Info (
    business_id,
    name,
    address,
    city,
    state,
    postal_code,
    latitude,
    longitude,
    stars,
    review_count,
    is_open,
    attributes,
    categories,
    hours
)
SELECT
    "BUSINESS_INFO_JSON":business_id::STRING,
    "BUSINESS_INFO_JSON":name::STRING,
    "BUSINESS_INFO_JSON":address::STRING,
    "BUSINESS_INFO_JSON":city::STRING,
    "BUSINESS_INFO_JSON":state::STRING,
    "BUSINESS_INFO_JSON":postal_code::STRING,
    "BUSINESS_INFO_JSON":latitude::FLOAT,
    "BUSINESS_INFO_JSON":longitude::FLOAT,
    "BUSINESS_INFO_JSON":stars::FLOAT,
    "BUSINESS_INFO_JSON":review_count::NUMBER,
    "BUSINESS_INFO_JSON":is_open::NUMBER,
    "BUSINESS_INFO_JSON":attributes::VARIANT,
    "BUSINESS_INFO_JSON":categories::STRING,
    "BUSINESS_INFO_JSON":hours::VARIANT
FROM STAGING.BUSINESS_INFO;
```

TRANSFORMING DATA FROM STAGING TO ODS

YELP_COVID.ODS Settings Code Versions

```
CREATE TABLE customer_reviews(
review_id STRING PRIMARY KEY,
user_id STRING,
business_id STRING,
stars FLOAT,
useful NUMBER,
funny NUMBER,
cool NUMBER,
review_text STRING,
review_date TIMESTAMP,
FOREIGN KEY (user_id) REFERENCES ODS.User_Details(user_id),
FOREIGN KEY (business_id) REFERENCES ODS.Business_info(business_id)
);

INSERT INTO ODS.customer_reviews (
    review_id,
    user_id,
    business_id,
    stars,
    useful,
    funny,
    cool,
    review_text,
    review_date
)
SELECT
    "BUSINESS_REVIEW_JSON":review_id::STRING,
    "BUSINESS_REVIEW_JSON":user_id::STRING,
    "BUSINESS_REVIEW_JSON":business_id::STRING,
    "BUSINESS_REVIEW_JSON":stars::FLOAT,
    "BUSINESS_REVIEW_JSON":useful::NUMBER,
    "BUSINESS_REVIEW_JSON":funny::NUMBER,
    "BUSINESS_REVIEW_JSON":cool::NUMBER,
    "BUSINESS_REVIEW_JSON":text::STRING AS review_text,
    "BUSINESS_REVIEW_JSON":date::TIMESTAMP AS review_date
FROM STAGING.BUSINESS_REVIEW;


CREATE TABLE Business_tips(
user_id STRING,
business_id STRING,
tip_text STRING,
date_tip TIMESTAMP,
compliment_count NUMBER,
PRIMARY KEY (user_id,business_id),
FOREIGN KEY (user_id) REFERENCES ODS.User_Details(user_id),
FOREIGN KEY (business_id) REFERENCES ODS.Business_info(business_id)
);
```

YELP_COVID.ODS Settings Code Versions

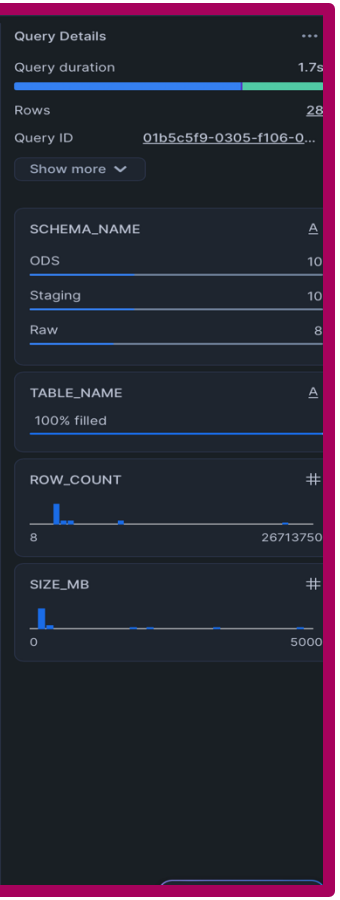
```
INSERT INTO Business_tips(
     user_id,
     business_id,
     tip_text,
     date_tip,
     compliment_count)
SELECT "BUSINESS_TIP_JSON":user_id::STRING,
       "BUSINESS_TIP_JSON":business_id::STRING,
       "BUSINESS_TIP_JSON":text::STRING AS tip_text,
       "BUSINESS_TIP_JSON":date::TIMESTAMP AS date_tip,
       "BUSINESS_TIP_JSON":compliment_count::NUMBER
FROM STAGING.BUSINESS_TIP;


CREATE TABLE Business_checkin(
business_id STRING,
date_checkin TIMESTAMP,
PRIMARY KEY (business_id,date_checkin),
FOREIGN KEY (business_id) REFERENCES ODS.Business_info(business_id)
);

INSERT INTO ODS.Business_checkin (
    business_id,
    date_checkin
)
SELECT
    BUSINESS_CHECK_IN_JSON:business_id::STRING AS business_id,
    TRY_TO_TIMESTAMP(trim(value)) AS date_checkin
FROM STAGING.BUSINESS_CHECK_IN,
LATERAL FLATTEN(INPUT => SPLIT(BUSINESS_CHECK_IN_JSON:date, ',')) AS date_checkin_data;



CREATE TABLE business_status (
    business_id STRING PRIMARY KEY,
    highlights STRING,
    delivery_or_takeout BOOLEAN,
    grubhub_enabled BOOLEAN,
    call_to_action_enabled BOOLEAN,
    request_a_quote_enabled BOOLEAN,
    covid_banner STRING,
    temporary_closed_until TIMESTAMP,
    virtual_services_offered STRING,
    FOREIGN KEY (business_id) REFERENCES ODS.Business_info(business_id)
);
```

TRANSFORMING DATA FROM STAGING TO ODS


```
INSERT INTO business_status (
    business_id,
    highlights,
    delivery_or_takeout,
    grubhub_enabled,
    call_to_action_enabled,
    request_a_quote_enabled,
    covid_banner,
    temporary_closed_until,
    virtual_services_offered
)
SELECT
    BUSINESS_JSON:business_id::STRING,
    IFF(BUSINESS_JSON:highlights::STRING = 'FALSE', NULL, BUSINESS_JSON:highlights::STRING) AS highlights,
    BUSINESS_JSON:"delivery or takeout"::BOOLEAN AS delivery_or_takeout,
    BUSINESS_JSON:"Grubhub enabled"::BOOLEAN AS grubhub_enabled,
    BUSINESS_JSON:"Call To Action enabled"::BOOLEAN AS call_to_action_enabled,
    BUSINESS_JSON:"Request a Quote Enabled"::BOOLEAN AS request_a_quote_enabled,
    IFF(BUSINESS_JSON:"Covid Banner"::STRING = 'FALSE', NULL, BUSINESS_JSON:"Covid Banner"::STRING) AS covid_banner,
    IFF(BUSINESS_JSON:"Temporary Closed Until"::STRING = 'FALSE', NULL, TRY_TO_TIMESTAMP(BUSINESS_JSON:"Temporary Closed Until"::STRING)) AS temporary_closed_until,
    BUSINESS_JSON:"Virtual Services Offered"::STRING AS virtual_services_offered
FROM STAGING.business_covid_status;


CREATE TABLE lv_weather(
date_weather DATE,
min_temp FLOAT,
max_temp FLOAT,
normal_min_temp FLOAT,
normal_max_temp FLOAT,
precipitation NUMBER(38,2),
precipitation_normal FLOAT);
)


INSERT INTO ODS.lv_weather(
date_weather,
min_temp,
max_temp,
normal_min_temp,
normal_max_temp,
precipitation,
precipitation_normal)
SELECT t.DATE,
       t.MIN_TEMP,
       t.MAX_TEMP,
       t.NORMAL_MIN_TEMP,
       t.NORMAL_MAX_TEMP,
       p.PRECIPITATION,
       p.PRECIPITATION_NORMAL
FROM STAGING.LV_TEMPERATURE_TRANSFORMED AS t
INNER JOIN STAGING.LV_PRECIPITATION_TRANSFORMED AS p
ON t.DATE = p.DATE;

ALTER TABLE lv_weather
ADD CONSTRAINT pk_lv_weather PRIMARY KEY (date_weather);
```

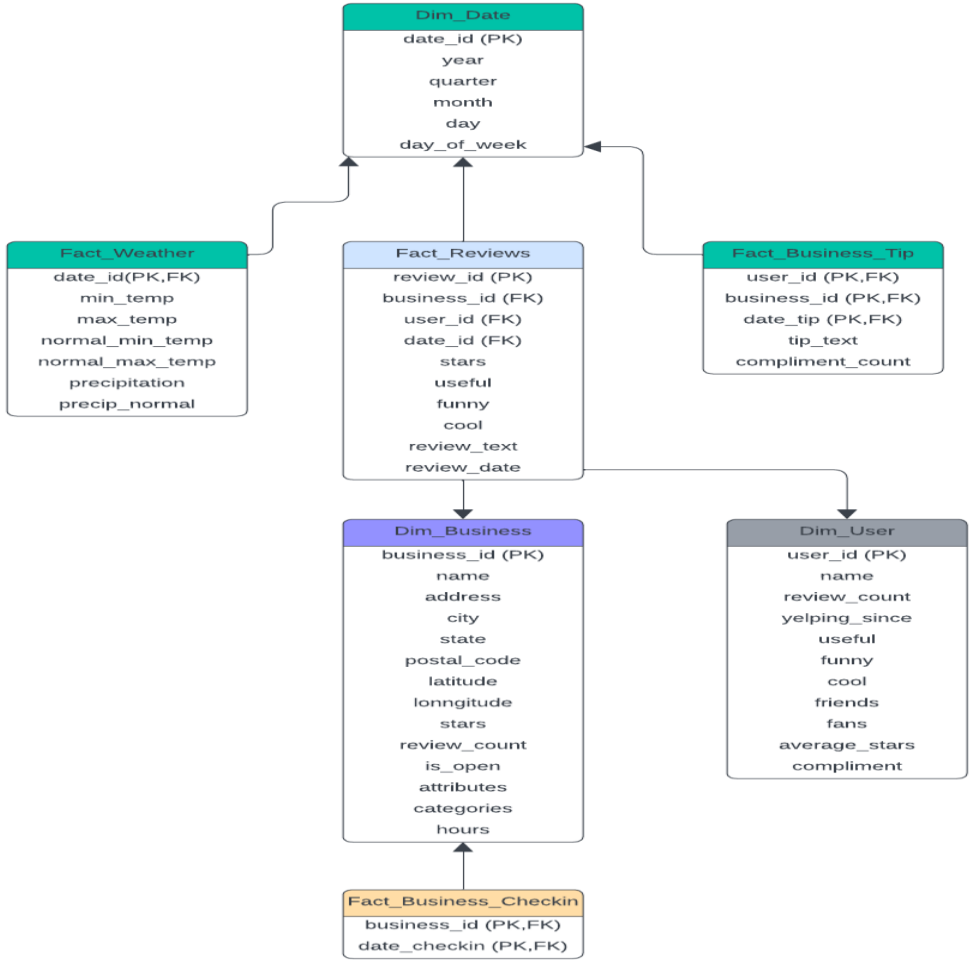
COMPARISON OF DATA SIZES IN RAW FILES, STAGING TABLES, AND ODS TABLES

	SCHEMA_NAME	TABLE_NAME	ROW_COUNT	SIZE_MB
1	ODS	BUSINESS_CHECKIN	26713750	192.7
2	ODS	BUSINESS_INFO	150346	11.5
3	ODS	BUSINESS_STATUS	209795	7.0
4	ODS	BUSINESS_TIPS	908915	45.7
5	ODS	CUSTOMER_REVIEWS	7691076	2153.3
6	ODS	LV_WEATHER	28241	0.2
7	ODS	ODS_TABLE_SIZES	null	null
8	ODS	RAW_FILE_SIZES	8	0.0
9	ODS	STAGING_TABLE_SIZES	null	null
10	ODS	USER_DETAILS	1987897	1891.6
11	Raw	usw00023169-las-vegas-mccarran-intl-ap-precipitation-inch.csv	null	1.0
12	Raw	usw00023169-temperature-degreef.csv	null	1.0
13	Raw	yelp_academic_dataset_business.json	null	118.0
14	Raw	yelp_academic_dataset_checkin.json	null	287.0
15	Raw	yelp_academic_dataset_covid_status.json	null	65.0
16	Raw	yelp_academic_dataset_review.json	null	5000.0
17	Raw	yelp_academic_dataset_tip.json	null	180.0
18	Raw	yelp_academic_dataset_user.json	null	3400.0
19	Staging	BUSINESS_CHECK_IN	131930	80.5
20	Staging	BUSINESS_COVID_STATUS	209795	5.0
21	Staging	BUSINESS_INFO	150346	11.0
22	Staging	BUSINESS_REVIEW	7691076	2156.2
23	Staging	BUSINESS_TIP	908915	46.1
24	Staging	LV_PRECIPITATION	28241	0.1
25	Staging	LV_PRECIPITATION_TRANSFORMED	28241	0.1
26	Staging	LV_TEMPERATURE	28241	0.3
27	Staging	LV_TEMPERATURE_TRANSFORMED	28241	0.2
28	Staging	USER	1987897	1892.8



DATA WAREHOUSE STAR SCHEMA

- This slide illustrates the star schema of the data warehouse, showcasing the relationships between dimension and fact tables for efficient querying and analysis.



DATA WAREHOUSE STAR SCHEMA

Explanation of Why We Have Multiple Fact Tables:

Having multiple fact tables in a star schema design is crucial for several reasons:

1. Different Types of Analysis:

- Fact_Reviews: Focuses on the review data, capturing details about user reviews, ratings, and associated business and user information
- Fact_Weather: Captures weather data, which is essential for analyzing the impact of weather conditions on business reviews.
- Fact_Business_Tip: Stores business tip data, which provides insights into additional user feedback separate from formal reviews.
- Fact_Business_Checkin: Tracks business check-ins, offering another dimension of user interaction with businesses.

2. Separation of Concerns:

- Each fact table represents a distinct type of business process or event. By separating these into different fact tables, we maintain clarity and focus for each type of analysis.

3. Performance Optimization:

- Smaller, specialized fact tables can improve query performance because each table is optimized for specific types of queries and analysis.

4. Easier Maintenance and Scalability:

- Managing and scaling the data warehouse is easier when fact tables are organized by distinct business processes, allowing for more straightforward updates and maintenance.

SQL QUERIES FOR DATA MOVEMENT (ODS TO DWH)

YELP_COVID.WAREHOUSE Settings Code Versions

```sql
CREATE TABLE Dim_User(
user_id STRING PRIMARY KEY,
name STRING,
review_count NUMBER,
yelping_since DATE,
useful STRING,
funny STRING,
cool STRING,
friends STRING,
fans NUMBER,
average_stars FLOAT,
compliment_hot NUMBER,
compliment_more NUMBER,
compliment_profile NUMBER,
compliment_cute NUMBER,
compliment_list NUMBER,
compliment_note NUMBER,
compliment_plain NUMBER,
compliment_cool NUMBER,
compliment_funny NUMBER,
compliment_writer NUMBER,
compliment_photos NUMBER );

INSERT INTO Dim_User (
    user_id,
    name,
    review_count,
    yelping_since,
    useful,
    funny,
    cool,
    friends,
    fans,
    average_stars,
    compliment_hot,
    compliment_more,
    compliment_profile,
    compliment_cute,
    compliment_list,
    compliment_note,
    compliment_plain,
    compliment_cool,
    compliment_funny,
    compliment_writer,
```

YELP_COVID.WAREHOUSE Settings Code Versions

```sql
    compliment_photos
)
SELECT
    user_id,
    name,
    review_count,
    CAST(yelping_since AS DATE) AS yelping_since,
    useful,
    funny,
    cool,
    friends,
    fans,
    average_stars,
    compliment_hot,
    compliment_more,
    compliment_profile,
    compliment_cute,
    compliment_list,
    compliment_note,
    compliment_plain,
    compliment_cool,
    compliment_funny,
    compliment_writer,
    compliment_photos
FROM
    ODS.User_Details;



CREATE TABLE Dim_Business (
    business_id STRING PRIMARY KEY,
    name STRING,
    address STRING,
    city STRING,
    state STRING,
    postal_code STRING,
    latitude FLOAT,
    longitude FLOAT,
    stars FLOAT,
    review_count NUMBER,
    is_open NUMBER,
    attributes VARIANT,
    categories STRING,
    hours VARIANT);
```

SQL QUERIES FOR DATA MOVEMENT (ODS TO DWH)

YELP_COVID.WAREHOUSE Settings Code Versions

```sql

INSERT INTO Dim_Business(
business_id,
name,
address,
city,
state,
postal_code,
latitude,
longitude,
stars,
review_count,
is_open,
attributes,
categories,
hours)
SELECT business_id,
name,
address,
city,
state,
postal_code,
latitude,
longitude,
stars,
review_count,
is_open,
attributes,
categories,
hours
FROM ODS.BUSINESS_INFO
WHERE categories ILIKE '%restaurant%';


```

YELP_COVID.WAREHOUSE Settings Code Versions

```sql
CREATE TABLE Dim_Date (
    date_id DATE PRIMARY KEY,
    year NUMBER,
    month NUMBER,
    day NUMBER,
    quarter NUMBER,
    week NUMBER,
    weekday NUMBER,
    day_name STRING,
    month_name STRING,
    is_weekend BOOLEAN
);

-- Generate and insert date range into Dim_Date
INSERT INTO Dim_Date (date_id, year, month, day, quarter, week, weekday, day_name, month_name, is_weekend)
WITH RECURSIVE date_sequence AS (
    SELECT TO_DATE('1990-01-01') AS date_id
    UNION ALL
    SELECT DATEADD(DAY, 1, date_id)
    FROM date_sequence
    WHERE date_id < TO_DATE('2024-12-31')
)
SELECT
    date_id,
    EXTRACT(YEAR FROM date_id) AS year,
    EXTRACT(MONTH FROM date_id) AS month,
    EXTRACT(DAY FROM date_id) AS day,
    EXTRACT(QUARTER FROM date_id) AS quarter,
    EXTRACT(WEEK FROM date_id) AS week,
    EXTRACT(DAYOFWEEK FROM date_id) AS weekday,
    TO_CHAR(date_id, 'Day') AS day_name,
    TO_CHAR(date_id, 'Month') AS month_name,
    CASE WHEN EXTRACT(DAYOFWEEK FROM date_id) IN (6, 7) THEN TRUE ELSE FALSE END AS is_weekend
FROM
    date_sequence;

```

SQL QUERIES FOR DATA MOVEMENT (ODS TO DWH)

```sql
CREATE TABLE Fact_Reviews (
    review_id STRING PRIMARY KEY,
    user_id STRING,
    business_id STRING,
    date_id DATE,
    stars FLOAT,
    useful NUMBER,
    funny NUMBER,
    cool NUMBER,
    review_text STRING,
    FOREIGN KEY (user_id) REFERENCES Dim_User(user_id),
    FOREIGN KEY (business_id) REFERENCES Dim_Business(business_id),
    FOREIGN KEY (date_id) REFERENCES Dim_Date(date_id)
);

INSERT INTO Fact_Reviews (
    review_id,
    user_id,
    business_id,
    date_id,
    stars,
    useful,
    funny,
    cool,
    review_text
)
SELECT
    r.review_id,
    u.user_id,
    b.business_id,
    d.date_id,
    r.stars,
    r.useful,
    r.funny,
    r.cool,
    r.review_text
FROM
    ODS.Customer_reviews r
JOIN
    Dim_User u ON r.user_id = u.user_id
JOIN
    Dim_Business b ON r.business_id = b.business_id
JOIN
    Dim_Date d ON r.review_date::DATE = d.date_id;


CREATE TABLE Fact_Weather (
    date_id DATE PRIMARY KEY,
    min_temp FLOAT,
    max_temp FLOAT,
    normal_min_temp FLOAT,
    normal_max_temp FLOAT,
    precipitation FLOAT,
    precipitation_normal FLOAT,
    FOREIGN KEY (date_id) REFERENCES Dim_Date(date_id)
);
```

```sql
INSERT INTO Fact_Weather (
    date_id,
    min_temp,
    max_temp,
    normal_min_temp,
    normal_max_temp,
    precipitation,
    precipitation_normal
)
SELECT
    date_weather,
    min_temp,
    max_temp,
    normal_min_temp,
    normal_max_temp,
    precipitation,
    precipitation_normal
FROM
    YELP_COVID.ODS.LV_WEATHER;


CREATE TABLE Fact_Tip (
    tip_id STRING PRIMARY KEY,
    user_id STRING,
    business_id STRING,
    date_id DATE,
    compliment_count NUMBER,
    tip_text STRING,
    FOREIGN KEY (user_id) REFERENCES Dim_User(user_id),
    FOREIGN KEY (business_id) REFERENCES Dim_Business(business_id),
    FOREIGN KEY (date_id) REFERENCES Dim_Date(date_id)
);


INSERT INTO Fact_Tip (
    tip_id,
    user_id,
    business_id,
    date_id,
    compliment_count,
    tip_text
)
SELECT
    CONCAT('tip_', CAST(row_number() over (order by user_id) as varchar)) AS tip_id, -- Generate unique ID for tips
    user_id,
    business_id,
    date_tip::DATE AS date_id,
    compliment_count,
    tip_text
FROM
    YELP_COVID.ODS.BUSINESS_TIPS;
```

```sql
CREATE TABLE Fact_Checkin (
    checkin_id STRING PRIMARY KEY,
    business_id STRING,
    date_id DATE,
    FOREIGN KEY (business_id) REFERENCES Dim_Business(business_id),
    FOREIGN KEY (date_id) REFERENCES Dim_Date(date_id)
);


INSERT INTO Fact_Checkin (
    checkin_id,
    business_id,
    date_id
)
SELECT
    CONCAT(business_id, '_', CAST(row_number() over (order by business_id) as varchar)) AS checkin_id, -- Generate a unique ID
    business_id,
    date_checkin::DATE AS date_id
FROM
    YELP_COVID.ODS.BUSINESS_CHECKIN;
```

QUERYING THE DATA WAREHOUSE FOR INSIGHTS

```sql
WITH TemperatureRanges AS (
    SELECT
        r.date_id,
        r.business_id,
        r.stars,
        CASE
            WHEN w.min_temp <= 32 THEN 'Very Cold (<= 32°F)'
            WHEN w.min_temp > 32 AND w.min_temp <= 50 THEN 'Cold (33-50°F)'
            WHEN w.min_temp > 50 AND w.min_temp <= 70 THEN 'Moderate (51-70°F)'
            WHEN w.min_temp > 70 AND w.min_temp <= 85 THEN 'Warm (71-85°F)'
            ELSE 'Hot (> 85°F)'
        END AS temp_range
    FROM
        Fact_Reviews r
    JOIN
        Fact_Weather w ON r.date_id = w.date_id
)
SELECT
    b.name AS business_name,
    tr.temp_range,
    COUNT(*) AS number_of_reviews,
    ROUND(AVG(tr.stars), 2) AS average_rating
FROM
    TemperatureRanges tr
JOIN
    Dim_Business b ON tr.business_id = b.business_id
GROUP BY
    b.name,
    tr.temp_range
ORDER BY
    b.name,
    tr.temp_range;
```

Results | Chart

	BUSINESS_NAME	TEMP_RANGE	NUMBER_OF_REVIEWS	AVERAGE_RATING
1	"Genuino" Italian Cafe'	Cold (33-50°F)	3	4.33
2	"Genuino" Italian Cafe'	Hot (> 85°F)	22	4.32
3	"Genuino" Italian Cafe'	Moderate (51-70°F)	21	4.48
4	"Genuino" Italian Cafe'	Warm (71-85°F)	23	4.35
5	#1 Mongolian BBQ - Best Stir Fried Noodles In Boise	Cold (33-50°F)	2	5
6	#1 Mongolian BBQ - Best Stir Fried Noodles In Boise	Hot (> 85°F)	57	3.81
7	#1 Mongolian BBQ - Best Stir Fried Noodles In Boise	Moderate (51-70°F)	30	3.87
8	#1 Mongolian BBQ - Best Stir Fried Noodles In Boise	Warm (71-85°F)	24	2.83
9	$5 Fresh Burger Stop	Hot (> 85°F)	1	4
10	$5 Fresh Burger Stop	Moderate (51-70°F)	3	5
11	$5 Fresh Burger Stop	Warm (71-85°F)	1	5
12	&pizza - UPenn	Cold (33-50°F)	2	3
13	&pizza - UPenn	Hot (> 85°F)	27	4.04

Query Details

Query duration 3.2s

Rows 128.1K

Query ID 01b5c649-0305-f101-0...

Show more

BUSINESS_NAME A

100% filled

TEMP_RANGE A

Hot (> 85°F) 36,478

Moderate (51-70°F)

Warm (71-85°F) 34,519

Ask Copilot

CONCLUSIONS AND ACKNOWLEDGEMENTS

1. **Conclusions**:

- **Summary of Data Architecture**:
- Successfully designed a robust data architecture for integrating Yelp and climate datasets using Snowflake.
- Created a well-structured data pipeline from raw data ingestion to a fully-functional data warehouse.
- Implemented staging, ODS, and DWH schemas to facilitate efficient data processing and analysis.
- Ensured data integrity and consistency through proper transformation and normalization techniques.

2. **Project Achievements**:

- Developed a comprehensive data architecture diagram to visualize the data flow and relationships between different layers.
- Utilized Snowflake's capabilities to manage large datasets and perform complex transformations.
- Designed and implemented a star schema for the data warehouse, optimizing for OLAP queries.
- Achieved a scalable and maintainable data warehouse structure that supports future data integration and analysis needs.

3. **Next Steps**:

- Further enhance the data warehouse by integrating additional data sources and refining existing transformations.
- Explore advanced data modeling techniques to improve query performance and data insights.
- Implement data governance and security measures to ensure the protection and privacy of the data.