Udacity Data Architect Project 2: Design A Datasystem

Business Objective: Analyze the effect of the **weather** on **customer reviews** of restaurants
Objective: Create a data system to crawl data from multiple sources and store it to the Data Warehouse

YELP Datasource

- business.JSON
- review.JSON
- user.JSON
- checkin.JSON
- tip.JSON

Covid YELP Datasource

- covid.JSON

Climate Explorer

- temperature.CSV
- precipitation.CSV

BULK LOADING TO SNOWFLAKE

STAGING

business	
business_info	variant

review	
review_info	variant

user	
user_info	variant

checkin	
checkin_info	variant

tip	
tip_info	variant

covid	
covid_info	variant

temperature	
date	int
min	int
max	int
normal_min	double
normal_max	double

temperature	
date	int
precipitation	varchar
precipitation_normal	double

OPERATIONAL DATA STORE

ODS Stage 1

CONVERT JSON TO MULTI-COLUMNS TABLE

business_raw	
business_id	varchar
name	varchar
address	varchar
city	varchar
state	varchar
postal_code	varchar
latitude	double
longitude	double
stars	double
review_count	int
is_open	int
restaurant_takeout	varchar
business_parking	varchar
category	varchar
hours	varchar

review_raw	
review_id	varchar
user_id	varchar
business_id	varchar
stars	int
date	date
text	varchar
useful	int
funny	int
cool	int

user_raw	
user_id	varchar
name	varchar
review_count	int
yelping_since	date
friends	varchar
useful	int
funny	int
cool	int
fans	int
elite	varchar
average_stars	double
compliment_hot	int
compliment_more	int
compliment_profile	int
compliment_cute	int
compliment_list	int
compliment_note	int
compliment_plain	int
compliment_cool	int
compliment_funny	int
compliment_writer	int
compliment_photos	int

checkin_raw	
business_id	varchar
date	datetime

tip_raw	
text	varchar
date	datetime
compliment_count	int
business_id	varchar
user_id	varchar

covid_raw	
business_id	varchar
highlight	varchar
delivery_or_takeout	varchar
grubhub_enabled	varchar
call_to_action_enabled	varchar
request_a_quote_enable	varchar
covid_banner	varchar
temporary_closed_until	varchar
virtual_services_offered	varchar

CONVERT DATA TYPE, CLEANING DATA

temperature	
date	date
min	int
max	int
normal_min	double
normal_max	double

precipitation	
date	date
precipitation	double
precipitation_normal	double

ODS Stage 2

ADDING CONSTRAINT, DATA CLEANING, TABLE NORMALIZATION, DEFINE SCHEMA

business_covid_feature	
business_id (PK)	varchar
highlight	varchar
delivery_or_takeout	varchar
grubhub_enabled	varchar
call_to_action_enabled	varchar
request_a_quote_enable	varchar
covid_banner	varchar
temporary_closed_until	varchar
virtual_services_offered	varchar

business_state	
state_id (PK)	int
state_name	varchar

business_city	
city_id (PK)	int
city_name	varchar

business_normalized	
business_id (PK)	varchar
state_id (FK)	int
city_id (FK)	int
name	varchar
address	varchar
postal_code	varchar
latitude	double
longitude	double
stars	double
review_count	int
is_open	int
restaurant_takeout	varchar
business_parking	varchar
category	varchar
hours	varchar

weather	
date (FK) (PK)	date
temp_min	int
temp_max	int
normal_temp_min	double
normal_temp_max	double
precipitation	double
precipitation_normal	double

checkin_normalized	
business_id (PK)	varchar
date	datetime

tip_normalized	
tip_id (PK)	int
business_id (FK)	varchar
user_id (FK)	varchar
date (FK)	date
text	varchar
compliment_count	int

review_normalized	
review_id (PK)	varchar
business_id (FK)	varchar
user_id (FK)	varchar
date (FK)	date
stars	int
text	varchar
useful	int
funny	int
cool	int

user_normalized	
user_id (FK) (PK)	varchar
name	varchar
review_count	int
yelping_since	date
useful	int
funny	int
cool	int
fans	int
elite	varchar
average_stars	double
compliment_hot	int
compliment_more	int
compliment_profile	int
compliment_cute	int
compliment_list	int
compliment_note	int
compliment_plain	int
compliment_cool	int
compliment_funny	int
compliment_writer	int
compliment_photos	int

user_friend_info	
user_id (PK)	varchar
friend_id (FK)	varchar

DATA WAREHOUSE

DESIGN STAR-SCHEMA BASED ON BUSINESS REQUIREMENT

dimension_business	
business_id (PK)	varchar
state_name	varchar
city_name	varchar
name	varchar
address	varchar
postal_code	varchar
latitude	double
longitude	double
stars	double
review_count	int
is_open	int
restaurant_takeout	varchar

fact_daily_review	
business_review_id (PK)	varchar
business_id (FK)	varchar
date (FK)	date
avg_stars	varchar
avg_useful	double
avg_funny	double
avg_cool	double

dimension_weather	
date (FK) (PK)	date
temp_min	int
temp_max	int
normal_temp_min	double
normal_temp_max	double
precipitation	double
precipitation_normal	double

dimension_review_date	
date (FK) (PK)	date
review_year	int
review_month	int
review_day	int

REPORTING

SQL QUERY FOR REPORTING

business_rating_with_weather	
business	varchar
date	varchar
avg_stars	varchar
avg_useful	double
avg_funny	double
avg_cool	double
temp_min	int
temp_max	int
precipitation	double