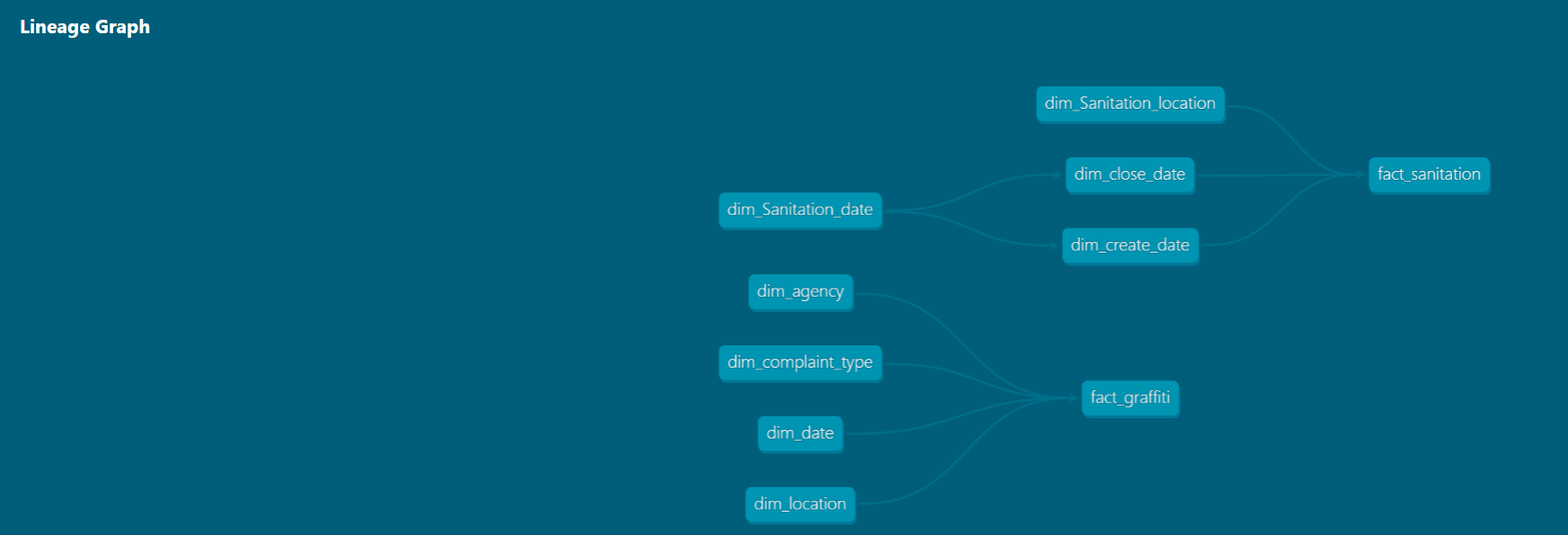
**Documentation:**

**Lineage Graph**

### 

### 

### 

### **1. Data Ingestion:**

* **Source Data Loading:** We successfully loaded our source datasets (Graffiti and Sanitation) into BigQuery.
* **dbt Project Setup:** We linked our dbt project to your BigQuery project using a service account key. This connection allows dbt to interact with your data and execute SQL transformations.

### **2. Data Modeling and Transformation:**

* **Model Creation:** We defined data models (dimensions and facts) for both Graffiti and Sanitation data. These models outline the structure and relationships between different data elements.
* **SQL Queries:** We wrote the SQL queries within these models to transform and clean the raw data, creating a data warehouse-ready structure.

Fact Graffiti

SQL Query:

{{ config(

materialized='view'

) }}

SELECT

`Unique Key` AS unique\_key,

d.date\_dimension\_id,

a.agency\_dim\_id,

ct.complain\_dim\_type\_id,

l.location\_dim\_id,

COUNT(\*) AS complaint\_count

FROM

`cis9440finalproject-442001.source\_dataset.grafitti\_complaints\_dataset` c

JOIN {{ ref('dim\_date') }} d ON c.`Created Date` = TIMESTAMP(d.full\_date)

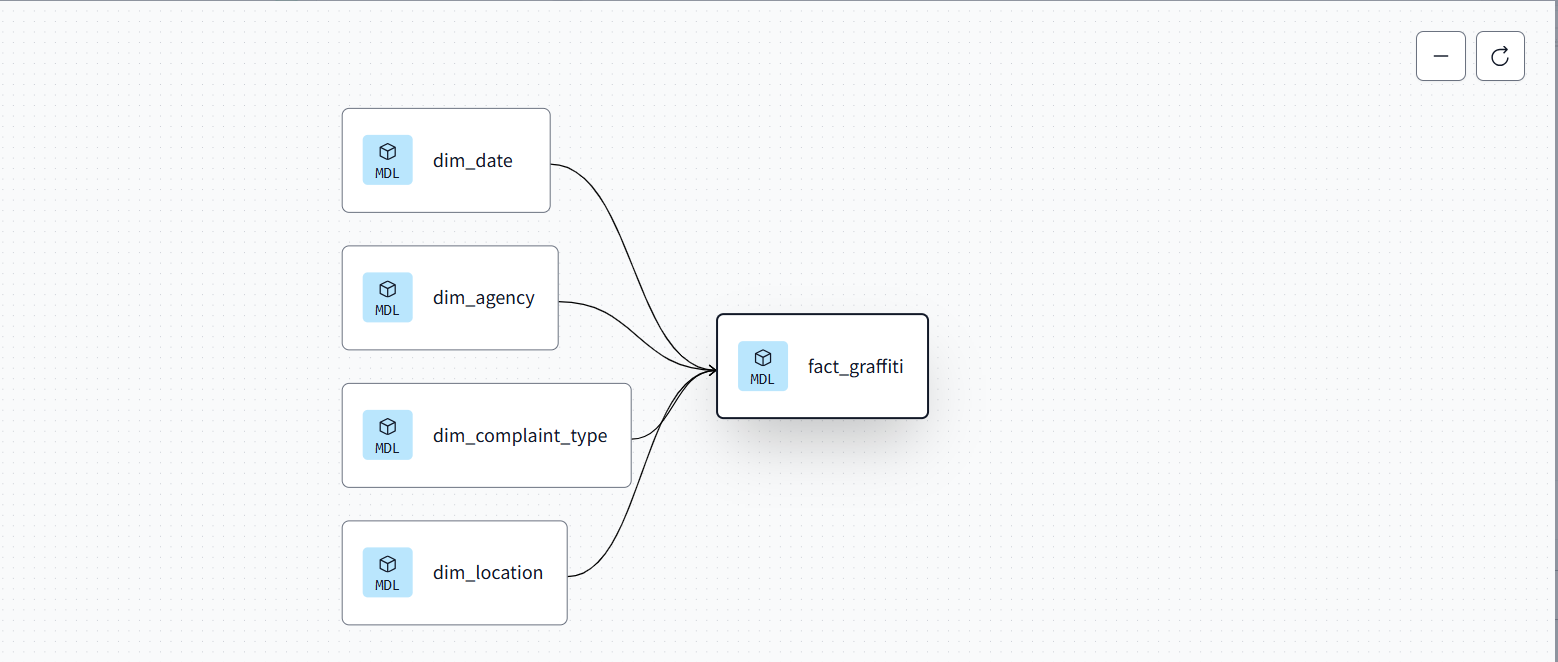
JOIN {{ ref('dim\_agency') }} a ON c.Agency = a.agency\_name

JOIN {{ ref('dim\_complaint\_type') }} ct ON CONCAT(c.`Complaint Type`, '\_', c.`Descriptor`) = ct.complain\_dim\_type\_id

JOIN {{ ref('dim\_location') }} l ON CONCAT(c.`Location Type`, '\_', c.`Incident Zip`, '\_', c.`Borough`) = l.location\_dim\_id

GROUP BY

unique\_key, d.date\_dimension\_id, a.agency\_dim\_id, ct.complain\_dim\_type\_id, l.location\_dim\_id



Fact Sanitation

SQL Query:

WITH base\_data AS (

SELECT

BID\_Name,

Borough,

`Incident Address`,

CONCAT(`BID\_Name`, '\_', `Borough`, '\_', `Incident Address`) AS location\_dim\_id,

SAFE\_CAST(`Sanitation Staff Employed` AS INT64) AS sanitation\_worker\_employed,

SAFE\_CAST(`Public Safety Employed` AS INT64) AS public\_safety\_officers\_employed,

SAFE\_CAST(`Sanitation hours Worked` AS FLOAT64) AS sanitation\_hours\_worked,

SAFE\_CAST(`Public Safety Hours Worked` AS FLOAT64) AS public\_safety\_hour\_worked,

SAFE\_CAST(`Trash Bags Collected` AS INT64) AS trash\_collected,

SAFE\_CAST(`Trash Receptacles Services` AS INT64) AS trash\_and\_recycling\_receptacles,

SAFE\_CAST(`Graffiti Incidents Removed` AS INT64) AS incidents\_graffiti\_removal,

SAFE\_CAST(`Sanitation Expenses` AS FLOAT64) AS sanitation\_expenses,

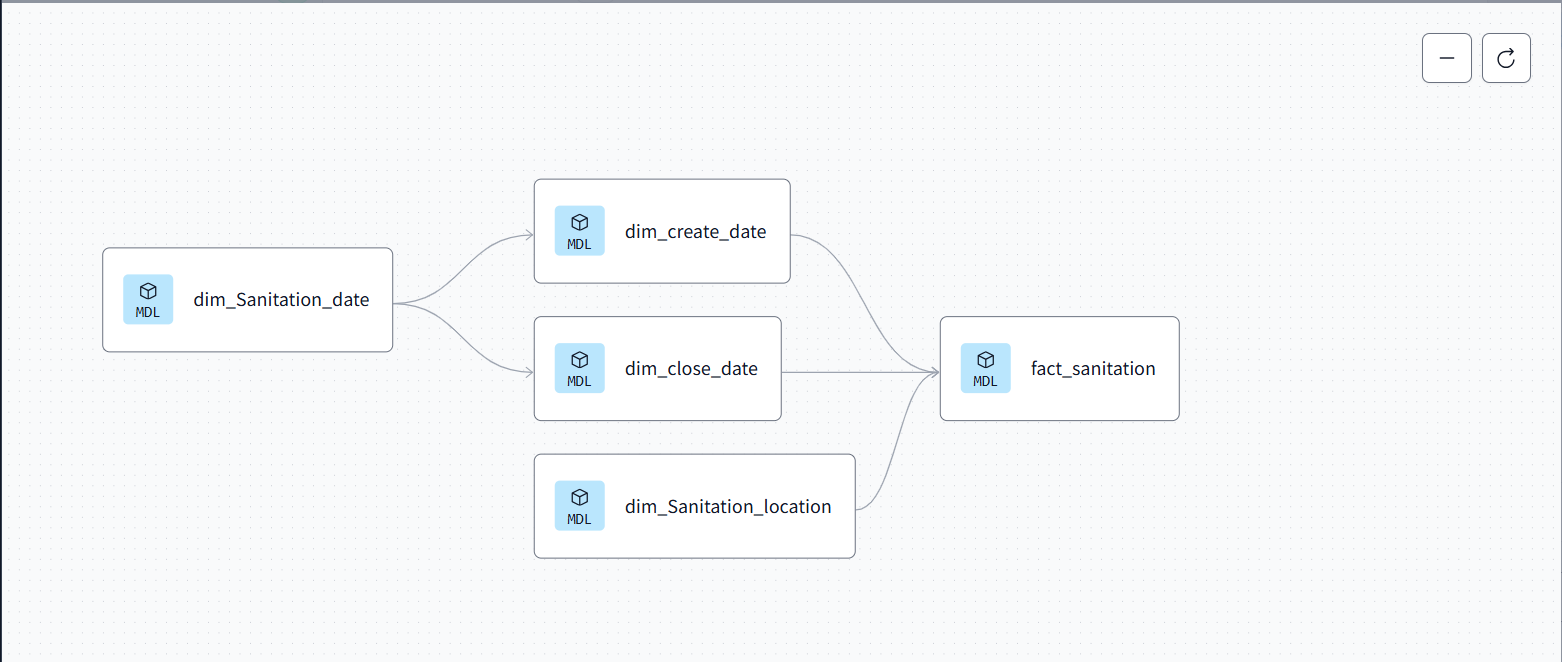
SAFE\_CAST(`Public Safety Expenses` AS FLOAT64) AS public\_safety\_expenses,

PARSE\_DATE('%Y-%m-%d', FORMAT\_DATETIME('%Y-%m-%d', `Incident Date`)) AS create\_date\_id,

PARSE\_DATE('%Y-%m-%d', FORMAT\_DATETIME('%Y-%m-%d', `Resolution Update Action Date`)) AS close\_date\_id

FROM

`cis9440finalproject-442001.source\_dataset.bid\_trends\_sanitation`



### 

### 

### **3. Testing and Deployment:**

* **Testing:** We set up a testing environment to run tests on our models, ensuring data quality and consistency.
* **Deployment:** Once testing was successful, we deployed our dbt project to production, triggering the execution of the defined transformations.

**Testing Query:**

version: 2

models:

- name: stg\_agency

description: Staging model for containing agency-related information

columns:

- name: agency\_dim\_id

description: Unique identifier for an agency

tests:

- unique

- not\_null

- name: stg\_complaint\_type

description: Staging model for storing the type of complaints

columns:

- name: complain\_dim\_type\_id

description: Unique identifier for a complaint type

tests:

- unique

- not\_null

- name: stg\_date

description: Staging model for date dimension

columns:

- name: date\_dimension\_id

description: Unique identifier for the date dimension

tests:

- unique

- not\_null

- name: stg\_location

description: Staging model for location dimensions

columns:

- name: location\_dim\_id

description: Unique identifier for the location dimension

tests:

- unique

- not\_null

- name: stg\_Sanitation\_date

description: Staging model for the sanitation date dimension

columns:

- name: date\_dimension\_id

description: Unique identifier for sanitation date

tests:

- unique

- not\_null

- name: stg\_Sanitation\_location

description: Staging model for sanitation location dimensions

columns:

- name: location\_dim\_id

description: Unique identifier for sanitation locations

tests:

- unique

- not\_null

- name: stg\_close\_date

description: Staging model for sanitation close dates

columns:

- name: closed\_date\_id

description: Unique identifier for close dates in sanitation

tests:

- unique

- not\_null

- name: stg\_create\_date

description: Staging model for sanitation creation dates

columns:

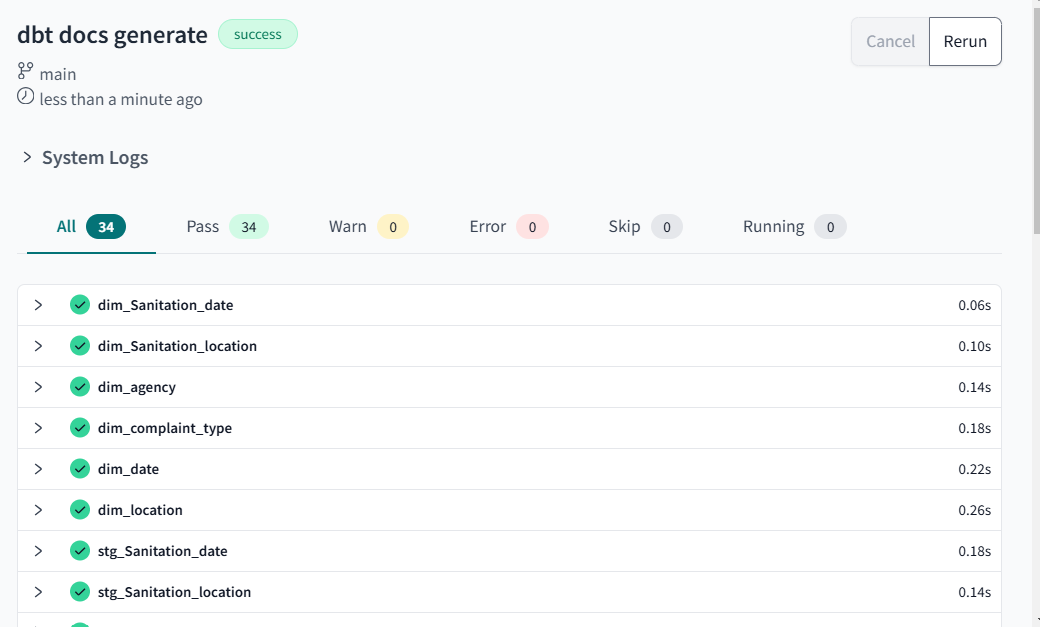
- name: created\_date\_id

description: Unique identifier for create date in sanitation

tests:

- unique

- not\_null



**Deployment:**

