

Exploring BigQuery Information Schema with SQL Queries

In BigQuery, the `INFORMATION_SCHEMA` views provide metadata about datasets, tables, columns, and constraints. This article explores a series of SQL queries that help retrieve and analyze metadata in the `bigquery-public-data.new_york_subway` dataset. Each query serves a specific purpose, providing valuable insights into the structure and constraints of the tables within the dataset.

1. Listing All Tables in the Dataset

```
SELECT*  
FROM bigquery-public-data.new_york_subway.INFORMATION_SCHEMA.TABLES;
```

This query retrieves all tables within the `new_york_subway` dataset. The `INFORMATION_SCHEMA.TABLES` view contains metadata such as table names, creation times, and other relevant details. This query is useful for getting an overview of the dataset's structure.

Result: A list of all tables in the `new_york_subway` dataset, including their names and additional metadata.

2. Listing All Columns in the Dataset

```
SELECT*  
FROM bigquery-public-data.new_york_subway.INFORMATION_SCHEMA.COLUMNS;
```

This query retrieves all columns within the `new_york_subway` dataset. The `INFORMATION_SCHEMA.COLUMNS` view provides details about each column, such as column names, data types, and their respective tables. This query helps understand the schema of each table.

Result: A comprehensive list of all columns in the dataset, along with their data types and other metadata.

3. Listing Columns for Specific Tables

```
SELECT  
    table_name,
```

```

column_name,
data_type
FROM bigquery-public-data.new_york_subway.INFORMATION_SCHEMA.COLUMNS
WHERE
table_name IN ('stations', 'station_entrances');

```

This query retrieves the column names and data types for the `stations` and `station_entrances` tables. By filtering the `INFORMATION_SCHEMA.COLUMNS` view, it focuses on these two specific tables, providing a detailed schema overview.

Result: A list of columns, including their names and data types, for the `stations` and `station_entrances` tables.

3. Finding Common Columns Between Two Tables

```

SELECT column_name
FROM bigquery-public-data.new_york_subway.INFORMATION_SCHEMA.COLUMNS
WHERE table_name = 'stations'
AND column_name IN (
    SELECT column_name
    FROM bigquery-public-data.new_york_subway.INFORMATION_SCHEMA.COLUMNS
    WHERE table_name = 'station_entrances'
);

```

This query finds the common column names between the `stations` and `station_entrances` tables. It uses a subquery to identify columns that exist in both tables. This query is useful for identifying shared attributes between related tables.

Result: A list of column names that are common to both the `stations` and `station_entrances` tables.

4. Comprehensive Side-by-Side Column Comparison

```
SELECT
    t1.column_name AS table1_column,
    t2.column_name AS table2_column
FROM
    `bigquery-public-data.new_york_subway.INFORMATION_SCHEMA.COLUMNS` AS t1
LEFT JOIN
    `bigquery-public-data.new_york_subway.INFORMATION_SCHEMA.COLUMNS` AS t2
ON
    t1.column_name = t2.column_name
    AND t2.table_name = 'station_entrances'
WHERE
    t1.table_name = 'stations'

UNION ALL

SELECT
    t1.column_name AS table1_column,
    t2.column_name AS table2_column
FROM
    `bigquery-public-data.new_york_subway.INFORMATION_SCHEMA.COLUMNS` AS t1
RIGHT JOIN
    `bigquery-public-data.new_york_subway.INFORMATION_SCHEMA.COLUMNS` AS t2
ON
    t1.column_name = t2.column_name
    AND t1.table_name = 'stations'
WHERE
    t2.table_name = 'station_entrances'
```

Description: This query provides a comprehensive side-by-side comparison of columns from the `stations` and `station_entrances` tables. It uses a combination of left and right joins to ensure that all columns from both tables are included, even if there is no match in the other table. The `UNION ALL` combines the results of both joins.

Result: A detailed comparison of columns from both tables, including all columns from `stations` and `station_entrances`, with `NULL` values where there is no match.

Conclusion: These queries leverage BigQuery's `INFORMATION_SCHEMA` views to explore and analyze the metadata of the `new_york_subway` dataset. Whether you're listing all tables and columns, identifying constraints, or comparing schemas, these queries provide a robust toolkit for understanding and managing your dataset's structure.

