Report on Snowflake Database Setup and Queries

I. Database Creation and Table Setup

In the provided SQL script, the following tasks were performed to set up the Snowflake database:

A) Creation of Tables

The script begins with the creation of various tables in the public schema, including:

Artist

MediaType

Genre

Playlist

Album

Track

PlaylistTrack

These tables define the structure of the database and relationships between different entities.

B) Creation of a Stage

A stage named s3\_data was created to provide a location for data ingestion from an Amazon S3 bucket. This stage allows for the easy loading of data into Snowflake from external sources.

C) Creation of File Format

A file format called CLASSIC\_CSV was created to specify the format and characteristics of the CSV files to be ingested. The format settings include delimiters, date and timestamp formats, error handling, and more.

D) Data Ingestion

Data was copied into the previously created tables from CSV files located in the specified S3 bucket. The process used the CLASSIC\_CSV file format and continued on error.

II. Creation of a Star Schema

After setting up the initial tables, the script proceeds to create a new schema named star\_schema. In this schema, the following dimensions and fact table were created:

DimAlbum: Contains information about albums, including title, production year, CD year, artist name, artist birth year, and country.

DimMediaType: Contains data on media types.

DimGenre: Contains information about genres.

DimPlaylistTrack: Contains details about playlist tracks, including the playlist name.

FactTrack: Serves as the fact table, including information about individual tracks, such as track name, composer, duration, bytes, unit price, and relationships to other dimension tables.

III. Data Queries

The script includes various SQL queries to retrieve information from the database. Here are some of the key queries and their results:

1. Titles of Albums with More than 1 CD

This query retrieves the titles of albums with more than one CD.

2. Tracks Produced in 2000 or 2002

A query to find tracks produced in the years 2000 or 2002; however, the query did not return any results.

3. Names and Composers of Rock and Jazz Tracks

This query returns the names and composers of tracks belonging to the Rock and Jazz genres.

4. Top 10 Longest Albums

A query to find the top 10 albums with the longest total duration in milliseconds.

5. Number of Albums Produced by Each Artist

This query counts the number of albums produced for each artist and displays the results in descending order.

6. Number of Tracks Produced by Each Artist

This query counts the number of tracks produced by each artist and displays the results in descending order.

7. Most Listened-to Music Genre in the 2000s

A query to find the most listened-to music genre in the 2000s based on track plays.

8. Playlists with Tracks Longer Than 4 Minutes

This query retrieves the names of playlists that contain tracks with a duration longer than 4 minutes.

9. Rock Tracks by French Artists

A query to find Rock tracks by artists from France.

10. Average Track Length by Music Genre

This query calculates the average track length for each music genre.

11. Playlists with Tracks by Artists Born Before 1990

This query retrieves the names of playlists that contain tracks by artists born before 1990.

IV. Conclusion

In this report, we have documented the database setup in Snowflake, including the creation of tables, stages, and file formats. We also observed several data queries aimed at extracting meaningful information from the database. Snowflake provides a powerful and scalable environment for data management and analytics.