Dbt Data Mart Documentation

1. Introduction

This document provides a comprehensive overview of the dbt models used in the data mart transformations for our analytics platform. The purpose of these models is to aggregate and transform raw data from various sources into structured formats that drive insights and decision-making across the organization.

2. Configuration Overview

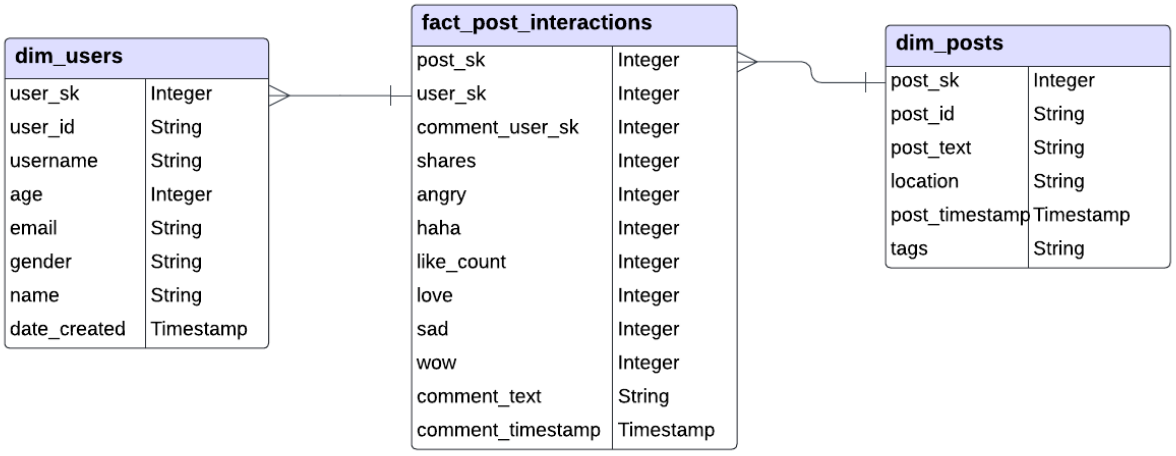
Project Configuration:

- **Name**: dbt_project
- **Version**: 1.0
- **Profile**: postgres_dw
- **Paths**:
    - Models are located in `models/marts`
    - Build artifacts are stored in the `target` directory

3. Data Sources Overview

Our models pull data from the following sources within our PostgreSQL data warehouse:

- **dim_posts**: Contains metadata about posts including ID, text, and timestamps.
- **fact_post_interactions**: Captures user interactions with posts, such as likes, comments, and shares.



4. Model Documentation

Content Performance Mart

File: content_performance_mart.sql

Purpose: Aggregates interaction metrics to assess the performance of content across social media platforms.

Sources:

- `dim_posts`
- `fact_post_interactions`

Transformations:

- Aggregate counts of likes, shares, and comments.
- Use `COALESCE` to handle possible NULL values ensuring robust data handling.

Metrics Produced:

- `total_likes`
- `total_shares`
- `total_comments`
- Interaction specifics like `total_angry`, `total_haha`, etc.

Column Name	Data Type	Description
post_sk	Integer	Surrogate key from dim_posts
post_id	String	Original post ID
post_text	String	Text content of the post
post_timestamp	Timestamp	Time when the post was created
tags	String	Tags associated with the post
location	String	Location where the post was made
total_likes	Integer	Total number of likes
total_shares	Integer	Total number of shares
total_comments	Integer	Total number of comments
total_angry	Integer	Total count of 'angry' reactions
total_haha	Integer	Total count of 'haha' reactions
total_love	Integer	Total count of 'love' reactions
total_sad	Integer	Total count of 'sad' reactions
total_wow	Integer	Total count of 'wow' reactions
total_users_interacted	Integer	Total distinct users who interacted

Location Analysis Mart

File: location_analysis_mart.sql

Purpose: Provides geographic insights into user interactions by aggregating data based on post locations.

Transformations:

- Summarize engagement metrics per location.
- Normalize location data to handle entries like 'Unknown Location'.

Metrics Produced:

- Total interactions per location
- Unique user counts per location

Column Name	Data Type	Description
post_sk	Integer	Surrogate key from dim_posts
post_id	String	Original post ID
post_text	String	Text content of the post
post_timestamp	Timestamp	Time when the post was created
tags	String	Tags associated with the post
location	String	Location where the post was made
total_likes	Integer	Total number of likes
total_shares	Integer	Total number of shares
total_comments	Integer	Total number of comments
total_angry	Integer	Total count of 'angry' reactions
total_haha	Integer	Total count of 'haha' reactions
total_love	Integer	Total count of 'love' reactions
total_sad	Integer	Total count of 'sad' reactions
total_wow	Integer	Total count of 'wow' reactions
total_users_interacted	Integer	Total distinct users who interacted

Tag Analysis Mart

File: tag_analysis_mart.sql

Purpose: Analyzes post interactions grouped by tags to determine trends in content engagement.
Transformations:

- Aggregate total and average likes per tag.

Metrics Produced:

- `total_likes`
- `total_posts_with_tag`
- `avg_likes_per_tag`

Column Name	Data Type	Description
tags	String	Tags associated with posts
total_likes	Integer	Total likes on posts with the tag
total_shares	Integer	Total shares on posts with the tag
total_comments	Integer	Total comments on posts with the tag
total_angry	Integer	Total 'angry' reactions for the tag
total_haha	Integer	Total 'haha' reactions for the tag
total_love	Integer	Total 'love' reactions for the tag
total_sad	Integer	Total 'sad' reactions for the tag
total_wow	Integer	Total 'wow' reactions for the tag
total_posts_with_tag	Integer	Total posts tagged with specific tags
avg_likes_per_tag	Float	Average likes per post tagged with tag

User Engagement Mart

File: user_engagement_mart.sql

Purpose: Ranks users based on their engagement with content, identifying key influencers.
Transformations:

- Sum and rank user interactions across multiple metrics.

Metrics Produced:

- Comprehensive interaction metrics per user
- `engagement_rank`

Column Name	Data Type	Description
tags	String	Tags associated with posts
total_likes	Integer	Total likes on posts with the tag
total_shares	Integer	Total shares on posts with the tag
total_comments	Integer	Total comments on posts with the tag
total_angry	Integer	Total 'angry' reactions for the tag
total_haha	Integer	Total 'haha' reactions for the tag
total_love	Integer	Total 'love' reactions for the tag
total_sad	Integer	Total 'sad' reactions for the tag
total_wow	Integer	Total 'wow' reactions for the tag
total_posts_with_tag	Integer	Total posts tagged with specific tags
avg_likes_per_tag	Float	Average likes per post tagged with tag

5. Schema Definitions

Each table and column in the models is documented in our `schema.yml`, ensuring clarity on data types and responsibilities:

- **Columns**:
    - Descriptions detail the role and data type of each column.
    - Tests validate data integrity, such as uniqueness and not-null constraints.

6. Best Practices and Conventions

We adhere to rigorous coding standards and best practices in our dbt projects to ensure maintainability and scalability:

- **Materialization**: Models are primarily materialized as tables to optimize performance.
- **Testing**: Rigorous data testing is conducted to ensure accuracy.

7. Conclusion

The documented dbt models play a crucial role in structuring our data architecture, providing reliable and scalable data marts for our business intelligence and analytics needs. This documentation serves as a guide for maintaining and extending our data infrastructure.