Report

Bogdan Negru

November 2025

1 Documentation

Join Keys

The model relies on a consistent set of join keys to connect core entities. Questions and answers are linked through `questions.question_id` and `answers.question_id`. The accepted answer relationship uses `questions.accepted_answer_id` joined to `answers.answer_id`. User interactions are unified via `users.user_id`, which connects to both `questions.owner_user_id` (askers) and `answers.owner_user_id` (answerers). Tags connect through the exploded array of `questions.tag_ids`, mapping each tag to a question using `tags.tag_id`. Finally, the `tag_metrics` table is keyed by `tag_metrics.tag_id`, aligning directly with the Tags dimension.

Hierarchies, Dimensions, Attributes, and Levels

The model contains several dimensions that enable multi-level analysis. The Users dimension provides user-level attributes such as reputation, supporting aggregation across both questions and answers. The Tags dimension supplies categorical identifiers and names, forming the basis for tag-level exploration. The Questions table functions partly as a fact (recording posting events) and partly as a dimension with descriptive attributes including title, view count, and creation timestamp. These attributes support analysis at several levels: event-level (individual question or answer), attribute-level (properties of questions or users), and aggregated tag-level (through the `tag_metrics` table). Tags do not form a formal hierarchy, but the combination of raw tag usage and aggregated metrics represents a logical structure from individual tag instances to aggregated tag behaviour.

Conformed Dimensions

The Users and Tags tables act as conformed dimensions. Both connect consistently to multiple fact-producing tables without changing meaning. Users join to both questions and answers, enabling unified analysis of participation and contribution patterns. Tags join to both raw question usage and aggregated

metrics, allowing a consistent semantic interpretation across analytic outputs. Because these dimensions use stable keys and definitions, reporting remains coherent regardless of the fact table referenced.

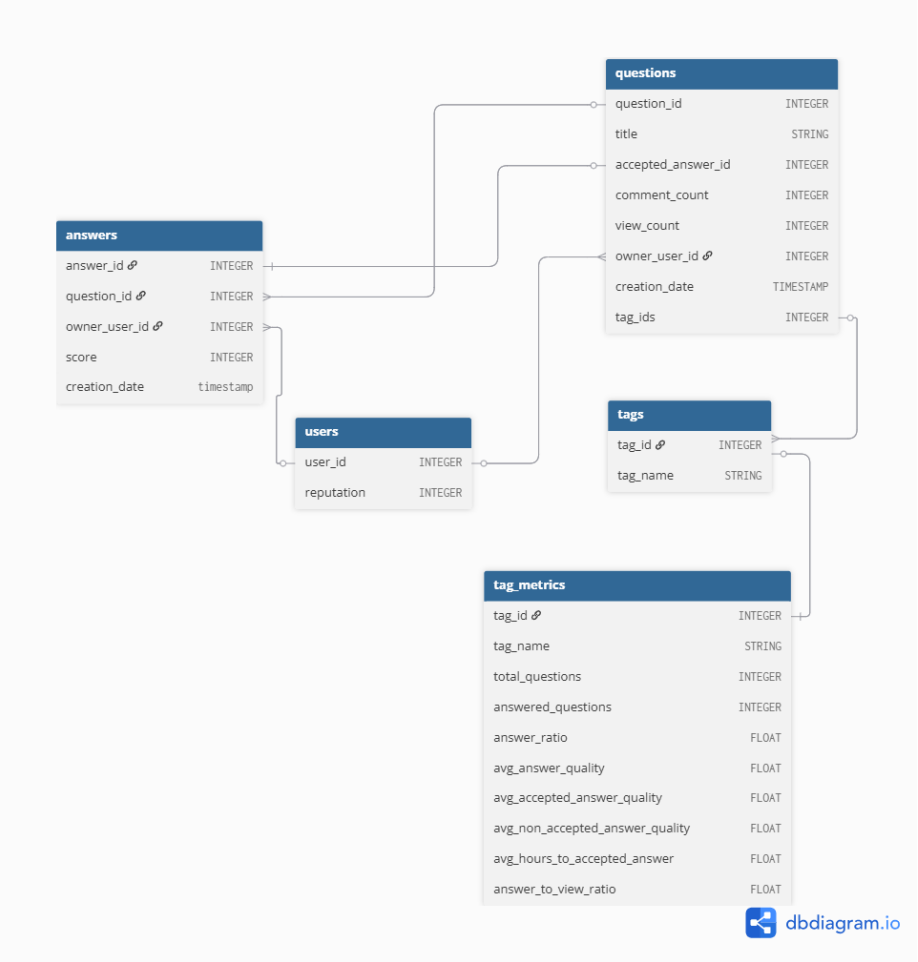
Fact Table Types

The model contains two transactional fact tables: `questions` and `answers`. Each records discrete posting events with associated descriptive attributes. The `tag_metrics` table functions as a periodic snapshot fact table, summarizing tag-level behaviour such as total questions, answer ratios, and average answer quality. This table consolidates historical activity into a snapshot optimized for analytics. No accumulating snapshot facts are present, as none of the tracked processes follow multi-stage lifecycle patterns.

Bridge and Helper Fact Tables

Questions may have multiple associated tags, creating a many-to-many relationship. Exploding the `tag_ids` array produces a bridge that links each question to each tag, enabling analysis at both granular and aggregated levels. The `tag_metrics` table serves as a helper fact table, precomputing key aggregates required for dashboarding. This reduces repetitive heavy joins on raw fact data and supports efficient downstream analytics.

2 Database Schema



3 Useful Links

1. GCP Project Dataset
2. LookerStudio Dashboard