Cassandra Table Design Documentation

November 20, 2024

Contents

1 Overview

This document presents the table structures for a Cassandra-based system designed for social media analytics. Each table is optimized for specific use cases to ensure performance and scalability.

2 Table Structures

2.1 `hashtag_stats`

- **Structure:**
  - Tracks the usage of hashtags per day (`day_bucket`).
  - Maintains a usage counter for each hashtag.
  - **Primary Key:** (`day_bucket`, `hashtag`), sorted in ascending order by `hashtag`.
- **Use Case:**
  - Analyze daily trends of hashtags.

2.2 `post_aggregates`

- **Structure:**
  - Stores aggregates such as total comments, likes, and shares per post.
  - Incorporates a time window (`window_start`) for fixed-period analysis.
  - **Primary Key:** (`window_start`, `post_id`).
- **Use Case:**
  - Generate reports on post engagement for specific time periods.

2.3 `post_stats`

- **Structure:**
  - Stores detailed data for each post, including comments, likes, shares, and associated hashtags.
  - Orders events by `timestamp` for activity history.
  - **Primary Key:** (`post_id`, `timestamp`), sorted in descending order.
- **Use Case:**
  - Detailed analysis of posts, enabling visualization of engagement trends over time.

3 Indexes

3.1 Secondary Index: `idx_hashtags`

- **Indexed Column:** `hashtag`.
- **Purpose:**
  - Quickly find all posts containing a given hashtag.
- **Note:**
  - Secondary indexes in Cassandra should be used cautiously due to potential performance impacts on large datasets.

4 Summary

The above table designs provide a robust foundation for social media analytics in Cassandra. They balance performance with scalability while addressing key use cases such as daily trend analysis, fixed-period engagement reports, and detailed post analytics.