Module 3 Assignment

Zilu(Lukas) Ji (ID 002838809)

College of Professional Studies, Northeastern University

ALY6030: Data Warehousing and SQL

Professor Adam Jones

Nov 24, 2024

Part 1

Based on the observation, only two variables stratify as facts, **COST_OF_INSPECTION _IN_DOLLARS** and **INSPECTION_SCORES**.

COST_OF_INSPECTION_IN_DOLLARS should be the additive facts, since it can be summed across various dimensions directly, such as development, city, state or time periods.

INSPECTION_SCORES can be averaged but not summed directly or meaningfully across dimensions. For example, summing scores across multiple inspections would not yield a meaningful measures. Nonetheless, we can calculate averages fro meaning insights. Hence, INSPECTION_SCORES should be consider as semi-additive facts.

There is no sign of non-additive facts observed base on the dataset.

Part 2

From the observation, there should be total 6 dimensions, which are

PUBLIC_HOUSING_AGENCY_NAME,

INSPECTED_DEVELOPMENT_NAME,

INSPECTED_DEVELOPMENT_ADDRESS,

INSPECTED_DEVELOPMENT_CITY,

INSPECTED_DEVELOPMENT_STATE,

INSPECTION_DATE.

Part 3

In order to fulfill the request of the Senior Management, two types of facts table will be used in this case.

Transaction Fact Table:

In order to store detailed information and each inspection at the finest level of granularity, which will allow a clear and direct drill down to individual inspections and views all related facts.

Structure:

(Facts)

COST_OF_INSPECTION_IN_DOLLARS,

INSPECTION_SCORE

(Dimensions)

PUBLIC_HOUSING_AGENCY_NAME,

INSPECTED_DEVELOPMENT_NAME,

INSPECTED_DEVELOPMENT_ADDRESS,

INSPECTED_DEVELOPMENT_CITY,

INSPECTED_DEVELOPMENT_STATE,

INSPECTION_DATE

Periodic Aggregate Fact Table:

Storing summarized data by a specific period can provide a higher level view of inspection costs and scores over time, tend to reducing query load and enhancing performance for periodic summaries.

Structure:

(Facts)

Summarized COST_OF_INSPECTION_IN_DOLLARS (total cost),

Average INSPECTION_SCORE

(Dimensions)

MONTH,

INSPECTED_DEVELOPMENT_STATE

Part 4

Under the consideration of the current condition, **type 2 Slowly Changing Dimension (SCD)** should be applied.

The reason for choosing type 2 because it approach enables the ability to view facts in the context of the agency's details at the time of the inspection, also it will retain historical data ensures the integrity and continuity of reporting, it will also allows reporting on how inspection costs or scores were attributed to agencies before and after they merged.

Not choosing type 0 is simply because it retains no history, so it does not address the need for historical tracking. Type 1 will overwrites the existing data, which will erase the history of agency changes and is unsuitable for tracking mergers. Type 3 is kind of the same as the reason not choosing type 0, it stores limited history, which is insufficient for a comprehensive history of multiple mergers and changes.

References

APA Style Table: APA.org. (n.d.). Retrieved from https://apastyle.apa.org/style-grammar-guidelines/tables-figures/tables

Chin, C. (2024, August 28). *The three types of fact tables*. The Holistics Blog. https://www.holistics.io/blog/the-three-types-of-fact-tables/

Gegner, S. (2024, June 19). *Introduction to slowly changing dimensions (SCD)*. IRI. https://www.iri.com/blog/vldb-operations/introduction-to-slowly-changing-dimensions-scd/

Wikimedia Foundation. (2024, October 2). *Slowly changing dimension*. Wikipedia. https://en.wikipedia.org/wiki/Slowly_changing_dimension