

1.

a. How many facts are there in this dataset?

There are 2 main facts in this dataset:

Inspection Score

Cost of Inspection (in dollars)

b. For the facts that you identify, what type of facts are they?

In data warehousing and dimensional modeling, “fact” refers only to measurable, quantitative data that you can add, average, or aggregate (e.g., dollars, counts, scores).

Facts (quantitative measures):

COST_OF_INSPECTION_IN_DOLLARS

→ A measurable cost → Additive Fact

INSPECTION_SCORE

→ A numeric score → Semi-additive Fact (can be averaged, not summed)

2.

a. How many dimensions are there in this dataset?

There are 4 clear dimensions in this dataset.

b. Which dimensions do you identify?

1. Date Dimension

Based on: INSPECTION_DATE

Can include: day, month, year, quarter, etc.

2. Public Housing Agency (PHA) Dimension

Based on: PUBLIC_HOUSING_AGENCY_NAME

May include historical tracking (Type 2 SCD)

3. Development Dimension

Based on: INSPECTED_DEVELOPMENT_NAME

May also include: development address, city, state

4. Location Dimension (optional, or combined with Development)

Based on: INSPECTED_DEVELOPMENT_ADDRESS, CITY, STATE

Useful for geographic grouping and filtering

3. Senior management is interested in viewing the facts identified above, at both the inspection level, as well as a periodic summary of inspection costs for each month. Based on this context, if you were to store this data in a set of fact tables, which type (or types) of fact tables would you use and why?

We should use two types of fact tables:

1. Transaction Fact Table

Stores detailed, event-level data—one row per inspection.

This supports inspection-level analysis such as:

What was the cost of inspection X?

What score did inspection Y receive?

Facts stored:

INSPECTION_SCORE

COST_OF_INSPECTION_IN_DOLLARS

Grain: One row per inspection.

2. Periodic Snapshot Fact Table

Stores data captured at regular intervals (e.g., monthly).

This supports time-based aggregation like:

What was the total cost of inspections in May 2015?

What was the average inspection score across agencies last month?

Facts stored (aggregated monthly):

TOTAL_COST_OF_INSPECTIONS

AVERAGE_INSPECTION_SCORE

NUMBER_OF_INSPECTIONS

Grain: One row per PHA per month (or per development per month, depending on needs).

4. Senior Management is also concerned with changes in the names and addresses of the public housing agency names since they tend to get merged with other agencies on a frequent basis. Based on this, how should we handle this slowly changing dimension? Select from types 0, 1, 2, or 3 from the Kimball reading. Justify your answer.

To handle changes in the names and addresses of Public Housing Agencies (PHAs), the best approach is to implement a Type 2 Slowly Changing Dimension (SCD). This method allows us to preserve the full history of changes by creating a new version of the dimension record whenever a change occurs, such as a merger or renaming. Each version of the PHA record would have its own surrogate key and include effective and

expiration dates, as well as a flag to indicate the current version. This ensures that inspection data remains historically accurate and traceable to the correct version of the agency at the time of the inspection. Unlike Type 0 or Type 1, which either freeze or overwrite data, and Type 3, which only retains limited history, Type 2 provides a robust solution for tracking frequent structural changes in PHAs while supporting reliable reporting and trend analysis for senior management.