

### **Problem 1:**

A fact in data warehousing is typically a quantitative or measurable data point used for analysis, often stored in a fact table. In this dataset, facts are the data points that represent measurable outcomes or values related to public housing inspections.

Facts Identified:

1. Cost of Inspection (COST\_OF\_INSPECTION\_IN\_DOLLARS)
  - Type: Numeric Fact (measurable and quantitative)
  - Explanation: Represents the cost incurred for each inspection.
2. Inspection Score (INSPECTION\_SCORE)
  - Type: Numeric Fact (measurable and quantitative)
  - Explanation: Represents the score given to the inspected public housing development.

Cost of inspection comes under Additive Facts and Inspection Score comes under Non-Additive Facts.

### **Problem 2:**

A dimension in data warehousing is a descriptive attribute or field that categorizes and provides context to facts. Dimensions help in filtering, grouping, and analyzing the facts.

Dimensions Identified:

1. Inspection ID:
  - Type: Identifier
  - Role: Uniquely identifies each inspection record.
2. Public Housing Agency Name:
  - Type: Organizational Dimension
  - Role: Represents the agency responsible for the inspection.
3. Inspected Development Name:
  - Type: Entity Dimension
  - Role: Indicates the specific public housing development inspected.
4. Inspected Development Address:
  - Type: Location Dimension
  - Role: Provides the physical address of the inspected development.
5. Inspected Development City:
  - Type: Location Dimension
  - Role: Specifies the city where the development is located.
6. Inspected Development State:

- Type: Location Dimension
  - Role: Specifies the state where the development is located.
7. Inspection Date:
- Type: Time Dimension
  - Role: Captures the date on which the inspection took place.

**Problem 3:**

Recommended Fact Table Types:

1. Transaction Fact Table:

- Purpose: Stores detailed data at the lowest granularity, i.e., at the inspection level.
- Usage: Allows detailed analysis, like examining inspection costs and scores for each inspection.
- Example:

Inspection ID	Inspection Date	Cost of Inspection	Inspection Score
536466	12/10/2014	27217	93

- Why: Provides granular data, enabling drill-down into individual inspections, which is important for audit or detailed analysis.

2. Periodic Snapshot Fact Table:

- Purpose: Stores summarized data for a specific period (e.g., monthly), ideal for management to view aggregate costs over time.
- Usage: Aggregates inspection costs and average scores for each month.
- Example:

Month-Year	Total Inspection Cost	Average Inspection Score
Dec-2014	100,000	85

- Why: Efficient for high-level reporting and performance tracking over time, helping management understand cost trends.

3. Aggregate Fact Table (Optional):

- Purpose: Stores pre-aggregated data to optimize performance when querying large datasets.
- Usage: Pre-calculates total costs or average scores by month or agency.
- Why: Useful when dealing with large volumes of data where real-time aggregation may be computationally expensive.

#### **Problem 4:**

Recommended SCD Type: SCD Type 2 (Versioned Records)

#### **Why SCD Type 2?**

1. **Maintaining History:** SCD Type 2 allows us to preserve historical data when an agency name or address changes. This is essential when agencies merge or rebrand, as it retains the ability to track past inspections under the old name.
2. **Version Control:** Creates a new record for each change, keeping the old data intact. This way, senior management can analyze inspections both before and after a merger or name change.
3. **Accurate Reporting:** When management wants to see costs or scores based on previous agency names, this method ensures data accuracy by linking each inspection to the correct version of the agency name at that point in time.

#### **How It Works:**

- **Example Scenario:**
  - The agency "Adams Housing Authority" merges with "Central Housing Group".
  - Instead of overwriting the old name, a new record is created:
    - **Old Record:**
      - Agency Name: Adams Housing Authority
      - Start Date: 01/01/2010
      - End Date: 03/31/2024
    - **New Record:**
      - Agency Name: Central Housing Group
      - Start Date: 04/01/2024
      - End Date: NULL (current)
  - Each inspection record will now be linked to the correct version of the agency name, preserving history.

#### **Alternative SCD Types and Why Not to Use Them:**

- **SCD Type 0 (Fixed):** Does not accommodate changes, unsuitable for merging agencies.
- **SCD Type 1 (Overwrite):** Updates the name without preserving history, losing context on past inspections.
- **SCD Type 3 (Previous Value):** Stores only the most recent change and the previous value, insufficient when multiple changes occur.