

Assignment 3: Public Housing Inspections Star Schema

1. Fact Identification

➤ (a) How many facts are there in this dataset?

- By looking at the data, there are mainly **two** numbers that we would measure or analyze.

➤ (b) Which facts do you identify?

- The facts are:
 - **COST_OF_INSPECTION_IN_DOLLARS**
 - **INSPECTION_SCORE**

➤ (c) For the facts that you identify, what type of facts are they?

- **COST_OF_INSPECTION_IN_DOLLARS:** This is an **additive fact** and these inspection costs can be added up. It can add costs for different housing agencies, for different months, or for different states.
- **INSPECTION_SCORE:** This is a **semi-additive fact**, it can be taken as the average of these scores. For example, the average score for a housing agency over a year, or the average score for all properties in a city for a month. But, adding up the scores themselves does not give a total that's easy to understand in the same way that adding costs makes sense.

2. Dimension Identification

➤ (a) How many dimensions are there in this dataset?

- There are about **three** dimensions in this dataset.

➤ **(b) Which dimensions do you identify?**

- **Date Dimension:** This comes from the **INSPECTION_DATE** column. It tells "when" each inspection happened.
- **Public Housing Agency Dimension:** This comes from the **PUBLIC_HOUSING_AGENCY_NAME** column. It tells "who" the agency involved in the inspection is.
- **Inspected Development Dimension:** This describes "what" property was checked and "where" it is. It would use columns like **INSPECTED_DEVELOPMENT_NAME**, **XXX_ADDRESS**, **XXX_CITY**, and **XXX_STATE**.

3. Fact Table Choice

To give senior management both the detailed view and the monthly summary they want, it would be best to use two kinds of fact tables:

➤ **Transactional Fact Table:**

- **Why:** To see facts at the inspection level, a transactional fact table is the way to go. Each row in this table would be for one single inspection. It would hold the **COST_OF_INSPECTION_IN_DOLLARS** and **INSPECTION_SCORE** for that inspection, and link foreign keys to the Date, PHA, and Inspected Development dimension tables.
- **Business Impact:** This lets managers look very closely at any single inspection. For example, if HUD managers want to know why a property got a very low

score, they can find that specific inspection record and see all its details, like the cost, the agency, and when it happened.

➤ **Periodic Snapshot Fact Table:**

- **Why:** For the periodic summary of inspection costs for each month, a periodic snapshot fact table is better. This table would show a summary of the facts at regular times, like each month. Each row might show the total **COST_OF_INSPECTION_IN_DOLLARS** for each housing agency for each month.
- **Business Impact:** This makes it quick and easy to see trends in costs over time. Managers can easily see if costs for an agency are going up or down month by month, without the computer having to add up all the individual inspections every time they run a report.

4. Handling Slowly Changing Dimensions

If Public Housing Agency (PHA) names and addresses change often because of mergers, and HUD needs to keep an accurate history for its reports, the best way to handle this is using SCD Type 2: **Add New Row**.

Why:

- **Type 0 (The passive method):** This won't work because the PHA information does change.
- **Type 1 (Overwrite):** If we just write over the old PHA name with the new one, we'll lose the history. For example, if Eastside PHA becomes part of United PHA, all old

inspections done by Eastside PHA would now look like United PHA did them. This would make it hard to see how Eastside PHA was doing before the change.

- **Type 2 (Add New Record):** This method keeps all the history. When a PHA's name or address changes, we add a new row in our PHA dimension table for the PHA with its new information. The old row is kept but marked as not current anymore. This means old inspection facts stay linked to the PHA information that was correct at that time, and new inspections link to the current PHA information.
- **Type 3 (Add New Column):** This means adding a new column like Previous_PHA_Name. But, since agencies might merge or change names sometimes, just having one previous column might not be enough to keep track of all the changes.

So, **SCD Type 2** is the best choice here. It makes sure all historical inspection data is linked to the correct PHA information for that time.