

Public Housing Inspections Star Schema



Assignment - 3

Public Housing Inspections Star Schema

Tanveer Syed

ALY - 6030

Data Warehousing & SQL

Dr Adam Jones

23 - March – 2025

Public Housing Inspections Star Schema

Assignment 3: Retail Sales Star Schema

1. How many facts are there in this dataset?

1	SELECT
2	COUNT (*) AS total_facts,
3	COUNT (DISTINCT INSPECTION_ID) AS unique_inspections
4	FROM public_housing_inspection_data;
5	

	total_facts	unique_inspections
1	6464	6464

total_facts = 6464

unique_inspection = 6464

- There are **6464 individual inspections**, which meaning each inspection is a unique fact in the dataset.

Public Housing Inspections Star Schema

- **Which facts do you identify?**

Facts in a dimensional model represent measurable, quantitative data. In this dataset, the key measurable attributes (facts) are:

- **INSPECTION_SCORE** (A percentage from 0 to 100 representing the inspection result)
- **COST_OF_INSPECTION_IN_DOLLARS** (The cost of performing the inspection)
- **For the facts that you identify, what type of facts are they?**
 - **INSPECTION_SCORE → Additive Fact**

This represents a score for each inspection and can be aggregated at different levels (e.g., average inspection score per month).

- **COST_OF_INSPECTION_IN_DOLLARS → Semi-Additive Fact**

This is a monetary value that **can be summed** across inspections but must be handled carefully when aggregating over time (e.g., monthly, or yearly totals).

Public Housing Inspections Star Schema

2.How many dimensions are there in this dataset?

```
4 FROM public_housing_inspection_data;
5 SELECT DISTINCT
6 PUBLIC_HOUSING_AGENCY_NAME,
7 INSPECTED_DEVELOPMENT_NAME,
8 INSPECTED_DEVELOPMENT_ADDRESS,
9 INSPECTED_DEVELOPMENT_CITY,
10 INSPECTED_DEVELOPMENT_STATE
11 FROM public_housing_inspection_data;
```

	PUBLIC_HOUSING_AGENCY_NAME	INSPECTED_DEVELOPMENT_NAME	INSPECTED_DEVELOPMENT_ADDRESS	INSPECTED_DEVELOPMENT_CITY	INSPECTED_DEVELOPMENT_STATE
1	Abbotsford Housing Authority	GREEN VISTA	310 E Pine St	Abbotsford	WI
2	Abingdon Redevelopment and Housi	Kings Mountain	300 Green Spring Rd	Abingdon	VA
3	Ada County Housing Authority	ADA COUNTY HA	Scattered Sites	BOISE	ID
4	ADAMS METROPOLITAN HOUSING AUTHO	ADAMS APARTMENTS	401 E 7th St	Manchester	OH
5	ADAMS METROPOLITAN HOUSING AUTHO	ADAMS APARTMENTS	50 E 8th Ave	Peebles	OH
6	Afton Housing Commission	FAIRVIEW	611 E Folk St	Afton	IA
7	Agra Housing Authority	CITY PARK MANOR	555 Southern Ave	Agra	KS
8	Ahoskie Housing Authority	Ahoskie Housing Un	200 Pierce Ave	Ahoskie	NC
9	Ainsworth Housing Authority	AINSWORTH PARK HOM	554 E 4th St	Ainsworth	NE
10	Akron Metropolitan Housing Autho	SAFENSTEIN TOWERS	585 Diagonal Rd	Akron	OH
11	Akron Metropolitan Housing Autho	Willow Run	555 Fuller St	Akron	OH
12	Akron Metropolitan Housing Autho	PINEWOOD GARDENS	8788 Ray Ct	Twinsburg	OH
13	Akron Metropolitan Housing Autho	VAN BUREN HOMES	244 244 LAMBERTON AVE Avenue	HARBERTON	OH
14	Akron Metropolitan Housing Autho	HONEY LOCUST GARDE	3299 Frange Dr	Cuyahoga Falls	OH
15	Akron Metropolitan Housing Autho	CRIMSON TERRACE	408 Sierra Ct	Barberton	OH
16	Akron Metropolitan Housing Autho	COLONIAL HILLS	27 Colonial Hills Dr	Akron	OH
17	Akron Metropolitan Housing Autho	MAPLEWOOD GARDENS	500 Hardman Dr	Cuyahoga Falls	OH

Which dimensions do you identify?

There are these Dimensions in the dataset:

- **Dimensions:**
 - **PHA (Public Housing Agency)**
 - **Development Name**
 - **Development Address**
 - **City**
 - **State**
 - **Inspection Date** (Time Dimension)

Public Housing Inspections Star Schema

1. Public Housing Agency (PHA) Dimension →

PUBLIC_HOUSING_AGENCY_NAME

2. Development Dimension → INSPECTED_DEVELOPMENT_NAME,

INSPECTED_DEVELOPMENT_ADDRESS,

INSPECTED_DEVELOPMENT_CITY,

INSPECTED_DEVELOPMENT_STATE

3. Date Dimension → INSPECTION_DATE

4. Inspection Result Dimension → INSPECTION_SCORE (since it

classifies inspection quality

3. which type (or types) of fact tables would you use and why?

Based on the requirements, two fact tables will be used:

- **Fact_Inspections Table** (Transactional Fact Table) – Stores individual inspection records.

Inspection-Level Fact Table

- This table captures details at the individual inspection level.
- It includes attributes like INSPECTION_ID, PUBLIC_HOUSING_AGENCY_NAME, INSPECTION_DATE, INSPECTION_SCORE, and COST_OF_INSPECTION_IN_DOLLARS.
- **Why?** This allows management to analyse individual inspections and their associated costs.
- **Fact_Monthly_Inspections Table** (Periodic Snapshot Fact Table) – Aggregates inspection costs monthly.
- **Periodic Summary Fact Table (Monthly Aggregation)**

Public Housing Inspections Star Schema

- This table aggregates inspection costs at a **monthly level** for each Public Housing Agency (PHA).
- It includes PUBLIC_HOUSING_AGENCY_NAME, inspection_month, and total_cost.
- **Why?** This enables management to track trends in inspection costs over time, which helps with budgeting and resource allocation.

Transactional Fact Table: The inspection-level fact table is a transactional fact table because it stores individual inspection events.

Periodic Snapshot Fact Table: The monthly summary fact table is a periodic snapshot fact table because it provides cost data at regular time intervals (monthly).

This approach ensures that management can analyse inspections at both granular and summarized levels, facilitating better decision-making.

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.

Public Housing Agencies (PHAs) frequently merge, causing changes in their names and addresses over time. To properly track these changes, we need to implement a Slowly Changing Dimension (SCD) strategy.

Recommended SCD Type: Type 2 (Preserving History)

The best approach for handling these changes is **SCD Type 2**, which maintains historical records by creating a new row whenever a change occurs. This allows us to **track the history of changes** in agency names and addresses over time.

Explanation for SCD Type 2

1. **Maintains Historical Accuracy:** Since senior management is concerned with tracking changes, **keeping historical data** allows them to analyze how agency structures have evolved.

Public Housing Inspections Star Schema

2. **Prevents Data Loss:** Instead of overwriting the old name/address (Type 1), Type 2 **preserves past data**, making it possible to analyse trends before and after a merger.
3. **Enables Time-Based Analysis:** By adding **effective date columns**, management can query which agency name/address was valid at any given point in time.

Why We Didn't Choose Them

SCD Type	Description	Why Not Chosen?
Type 0 (Retain Original)	Never updates historical data.	Not suitable because agency names/addresses change frequently .
Type 1 (Overwrite Data)	Updates the record but does not keep history.	Causes data loss ; we cannot track past agency names.
Type 3 (Track Partial History)	Adds a "Previous Name" column but keeps only the latest and previous values.	Not sufficient for tracking multiple mergers over time.

For tracking mergers and name changes, SCD Type 2 is the best approach. It ensures historical accuracy, prevents data loss, and allows management to analyse agency structure evolution over time.

5. Analysis of Inspection Costs for Public Housing Agencies (PHAs)

1. Objective

Senior management requested an analysis of inspection costs, specifically identifying PHAs that have experienced an increase in the cost of inspections.

The objectives of this analysis are to:

- Identify PHAs where inspection costs have increased.
- Provide insights into the most recent and second-most recent inspections for each PHA.
- Calculate the absolute and percentage increase in inspection costs.
- Deliver the findings in a structured file format.

2. Methodology

To achieve this, the analysis was performed using SQLite with the following steps:

Public Housing Inspections Star Schema

1. Data Preparation

- The **INSPECTION_DATE** field was converted to the appropriate DATE format to ensure accurate chronological ordering.
- The dataset was structured to allow for comparison of the two most recent inspections per PHA.

2. Identifying Cost Increases

- The **LAG()** window function was used to retrieve the second-most recent inspection data for each PHA.
- The cost difference (**CHANGE_IN_COST**) and percentage increase (**PERCENT_CHANGE_IN_COST**) were calculated.

3. Filtering and Cleaning

- PHAs with only one inspection were removed to ensure valid comparisons.
- Only PHAs where the most recent inspection cost was higher than the second-most recent inspection cost were included.
- The final dataset was sorted in descending order based on percentage increase to highlight the most significant changes.

3. Findings & Deliverables

- The final dataset includes only PHAs that experienced an increase in inspection costs.
- Each PHA is represented only once to avoid duplicate data.
- The report includes the most recent and second-most recent inspection details with absolute and percentage cost changes.
- The extracted data has been exported as a CSV file for further review.

The CSV file contains the following columns:

- **PHA_NAME** – Name of the public housing agency.
- **MR_INSPECTION_DATE** – Date of the most recent inspection.
- **MR_INSPECTION_COST** – Cost of the most recent inspection.
- **SECOND_MR_INSPECTION_DATE** – Date of the second-most recent inspection.
- **SECOND_MR_INSPECTION_COST** – Cost of the second-most recent inspection.
- **CHANGE_IN_COST** – Absolute increase in inspection cost.

Public Housing Inspections Star Schema

- **PERCENT_CHANGE_IN_COST** – Percentage increase in inspection cost.

4. Conclusion

This analysis provides senior management with a clear and structured view of PHAs experiencing increasing inspection costs. The use of SQL functions ensures efficiency, while the output file enables easy decision-making. Future recommendations include further trend analysis to identify cost drivers and predict future cost fluctuations.

Public Housing Inspections Star Schema

SQL Query Used

The following SQL query was executed in **DB Browser for SQLite** to extract the required insights:

```
1 WITH inspection_ranked AS (  
2   SELECT  
3     PUBLIC_HOUSING_AGENCY_NAME AS PHA_NAME,  
4     DATE(INSPECTION_DATE) AS MR_INSPECTION_DATE,  
5     COST_OF_INSPECTION_IN_DOLLARS AS MR_INSPECTION_COST,  
6     LAG(DATE(INSPECTION_DATE)) OVER (  
7       PARTITION BY PUBLIC_HOUSING_AGENCY_NAME  
8       ORDER BY DATE(INSPECTION_DATE) DESC  
9     ) AS SECOND_MR_INSPECTION_DATE,  
10    LAG(COST_OF_INSPECTION_IN_DOLLARS) OVER (  
11      PARTITION BY PUBLIC_HOUSING_AGENCY_NAME  
12      ORDER BY DATE(INSPECTION_DATE) DESC  
13    ) AS SECOND_MR_INSPECTION_COST  
14  FROM public_housing_inspection_data  
15 )  
16 SELECT  
17   PHA_NAME,  
18   MR_INSPECTION_DATE,  
19   MR_INSPECTION_COST,  
20   SECOND_MR_INSPECTION_DATE,  
21   SECOND_MR_INSPECTION_COST,  
22   (MR_INSPECTION_COST - SECOND_MR_INSPECTION_COST) AS CHANGE_IN_COST,  
23   ROUND(((MR_INSPECTION_COST - SECOND_MR_INSPECTION_COST) * 100.0 /  
24     SECOND_MR_INSPECTION_COST), 2) AS PERCENT_CHANGE_IN_COST  
25  FROM inspection_ranked  
26  WHERE SECOND_MR_INSPECTION_COST IS NOT NULL -- Remove PHAs with only one  
27    inspection  
28  AND MR_INSPECTION_COST > SECOND_MR_INSPECTION_COST -- Keep only increasing  
29    costs  
30  ORDER BY PERCENT_CHANGE_IN_COST DESC; -- Sort by highest increase first
```