****

### **Public Housing Inspections Star Schema**

#### ALY6030-Mod4: Assignment 3

#### Integrated Data Warehousing and SQL

Instructor: [Kayal Chandrasekaran](https://northeastern.instructure.com/courses/140410/users/121784)

22 June 2023

Submitted By: Srishti Singh

**INTRODUCTION**

Public housing in the United States is overseen by local Public Housing Agencies (PHAs) under HUD's supervision. PHAs conduct inspections, assign scores, and determine costs for developments. As a Developer hired by HUD, the task is to analyze the inspection data and provide key insights to Senior Management.

The dataset includes PHA names, development details, inspection dates, scores, and costs. The report aims to answer several questions. Firstly, we identify the facts and their types in the dataset. Secondly, we determine the dimensions present. Thirdly, we suggest appropriate fact table types to store the data for inspection-level and periodic cost summary views. We also address slowly changing dimensions related to changes in PHA names and addresses.

Lastly, we perform an analysis to identify PHAs with increased inspection costs. The final deliverable will be a file containing PHA information, most recent inspection date and cost, second most recent inspection date and cost, change in cost, and percentage change. We will utilize lead or lag functions in SQL for this analysis. Additionally, we will investigate converting dates from TEXT to Date format for proper function usage and ensure a clean dataset by filtering out duplicates and PHAs with a single inspection.

By providing comprehensive answers and insights, this analysis will assist Senior Management in making informed decisions and optimizing public housing programs.

**QUESTIONS**

**1.**

In this dataset, we have identified below facts:

COST\_OF\_INSPECTION\_IN\_DOLLARS

INSPECTION\_SCORE

The type of facts for these identified measures are as follows:

COST\_OF\_INSPECTION\_IN\_DOLLARS: Additive Fact

INSPECTION\_SCORE: Additive Fact

**2.**

We can determine seven dimensions from the given dataset:

INSPECTION\_ID

PUBLIC\_HOUSING\_AGENCY\_NAME

INSPECTED\_DEVELOPMENT\_NAME

INSPECTED\_DEVELOPMENT\_ADDRESS

INSPECTED\_DEVELOPMENT\_CITY

INSPECTED\_DEVELOPMENT\_STATE

INSPECTION\_DATE

**3.**

There are two types of fact tables that can be used to store inspection data. The transactional fact table records individual inspection events, while the periodic snapshot fact table captures aggregated data at regular intervals. By utilizing both types of fact tables with appropriate dimension tables, we can provide detailed inspection-level data and a periodic overview of inspection expenses.

**4.**

To handle changes in public housing agency names and addresses, we can implement a Slowly Changing Dimension (SCD) Type 2 approach. SCD Type 2 allows us to track changes over time by creating new records with distinct identifiers, start dates, and end dates for each modification. This preserves the complete history of changes while ensuring accurate reporting of data. For example, if an agency undergoes a name and address change, a new record is created with updated information while the existing record is assigned an end date. SCD Type 2 helps maintain a comprehensive history of changes while upholding data integrity and accuracy.

**5.**

Please refer to the "Srishti\_Singh \_Assignment3.sql" file that is attached for the SQL query and the exported file.

