Assignment 3 — Public Housing Inspections Star Schema

Shumin Tu (NUID: 002848797)

College of Professional Studies, Northeastern University

ALY6030: Data Warehousing and SQL

Professor: Adam Jones

Nov 30, 2024

1.

a. **How many facts are there in this dataset?**

Facts are the measurements/metrics or facts from business process. In this dataset, there are two facts.

b. **Which facts do you identify?**

I identify the columns “COST_OF_INSPECTION_IN_DOLLARS” and “INSPECTION_SCORE” as facts.

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

COST_OF_INSPECTION_IN_DOLLARS is **Additive Fact** because it can be summed across all inspections or grouped by categories such as city, state, or agency to calculate the total cost. It retains meaning when aggregated.

INSPECTION_SCORE is **Non-additive Fact** because the inspection score cannot be summed meaningfully across multiple inspections.

2.

a. **How many dimensions are there in this dataset?**

Dimension attributes are used to organize data, often presenting as a context for measurable facts. Based on the dataset, there are 7 dimensions.

b. **Which dimensions do you identify?**

I identify columns “Inspection ID”, “Public Housing Agency Name”, “Inspected Development Name”, “Inspected Development Address”, “Inspected City”, “Inspected Development State” and “Inspection Date” as dimensions.

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

   Ralph Kimball's dimensional data modeling defines three types of fact tables. These are:

- Transaction fact tables.
- Periodic snapshot tables, and
- Accumulating snapshot tables

In this scenario, I will use **periodic snapshot fact table to** summarize inspection data periodically, especially monthly. I choose this type of fact table because it supports the need for periodic summaries of inspection costs for each month, making it easier for management to analyze trends, compare costs over time, and perform high-level aggregations.

4. **How should we handle this slowly changing dimension? Select from types 0, 1, 2, or 3 from the Kimball reading.**

Type	Definition
0	Retain original
1	Update or overwrite the existing record in the dimension table when changes occur.
2	Preserve historical data by inserting a new row for each change, marking the new row as current and appropriately labeling the previous record as historical.
3	Use a separate column to capture and represent the changes.

From my perspective, SCD Type 2 is the most suitable approach. Using SCD Type 2, we can create a new row in the dimension table to record changes in the agency name or address.

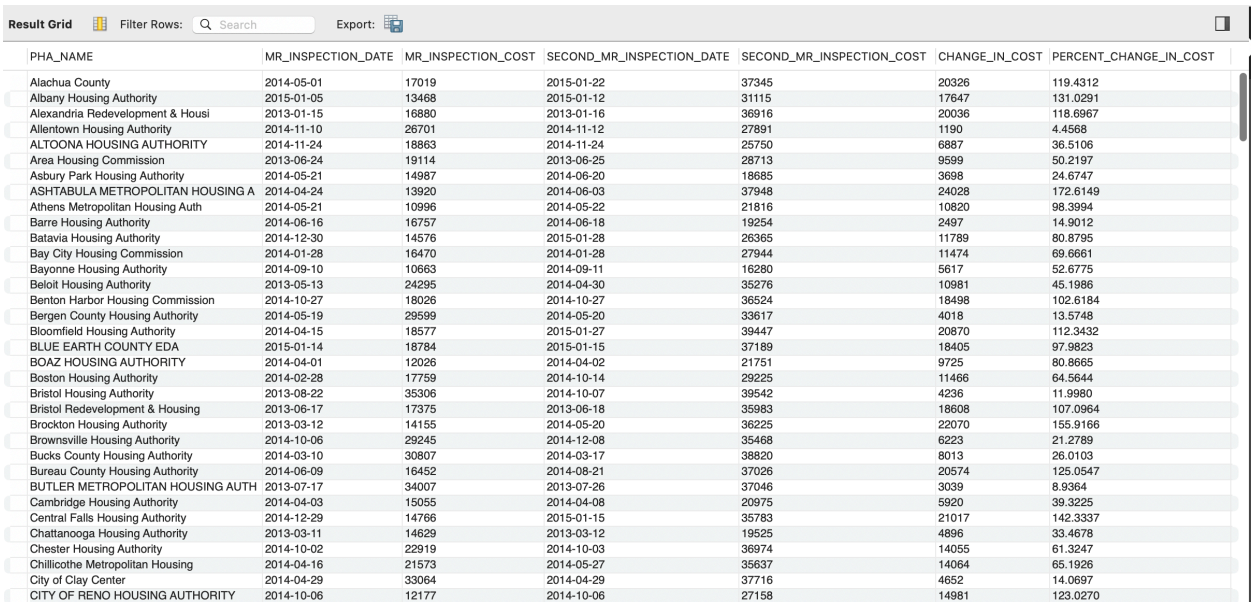
The most important point is that SCD Type 2 preserves the history of changes. If senior management wants to compare historical data before and after changes, or perform cost analysis or inspection score comparisons based on agency changes, SCD Type 2 can provide the data and information needed for further analysis.

5. SQL Results

```sql
# Filter out PHAs that only performed one inspection, convert dates from TEXT to Date format and use date to rank the first inspection and second inspection.
WITH T1 AS (
SELECT
PUBLIC_HOUSING_AGENCY_NAME AS PHA_NAME,
STR_TO_DATE(INSPECTION_DATE, '%m/%d/%Y') AS INSPECTION_DATE,
COST_OF_INSPECTION_IN_DOLLARS AS INSPECTION_COST,
RANK() OVER (PARTITION BY PUBLIC_HOUSING_AGENCY_NAME ORDER BY INSPECTION_DATE ASC) as RNK
FROM public_housing_inspection_data
WHERE PUBLIC_HOUSING_AGENCY_NAME NOT IN (SELECT PUBLIC_HOUSING_AGENCY_NAME
                                         FROM public_housing_inspection_data
                                         GROUP BY PUBLIC_HOUSING_AGENCY_NAME
                                         HAVING COUNT(PUBLIC_HOUSING_AGENCY_NAME) =1
                                         )
),

# Use Lead window function to create the second date and second cost, and remove inspections after the second one.
T2 AS (
SELECT
PHA_NAME,
INSPECTION_DATE AS MR_INSPECTION_DATE,
INSPECTION_COST AS MR_INSPECTION_COST,
LEAD(INSPECTION_DATE) OVER (PARTITION BY PHA_NAME ORDER BY INSPECTION_DATE ASC) AS SECOND_MR_INSPECTION_DATE,
LEAD(INSPECTION_COST) OVER (PARTITION BY PHA_NAME ORDER BY INSPECTION_DATE ASC) AS SECOND_MR_INSPECTION_COST
FROM T1
WHERE RNK <= 2
)

 # List each PHA only once, with no duplicates, and calculate the change in cost and the percent change in cost,
 # restricted to cases where there is an increase in cost between the first and second inspections.
 SELECT
 DISTINCT PHA_NAME,
 MR_INSPECTION_DATE,
 MR_INSPECTION_COST,
 SECOND_MR_INSPECTION_DATE,
 SECOND_MR_INSPECTION_COST,
 (SECOND_MR_INSPECTION_COST - MR_INSPECTION_COST) AS CHANGE_IN_COST,
 ((SECOND_MR_INSPECTION_COST - MR_INSPECTION_COST) / MR_INSPECTION_COST * 100) AS PERCENT_CHANGE_IN_COST
 FROM T2
 WHERE (SECOND_MR_INSPECTION_COST - MR_INSPECTION_COST) > 0
```



Result Grid Filter Rows: Search Export:

PHA_NAME	MR_INSPECTION_DATE	MR_INSPECTION_COST	SECOND_MR_INSPECTION_DATE	SECOND_MR_INSPECTION_COST	CHANGE_IN_COST	PERCENT_CHANGE_IN_COST
Alachua County	2014-05-01	17019	2015-01-22	37345	20326	119.4312
Albany Housing Authority	2015-01-05	13468	2015-01-12	31115	17647	131.0291
Alexandria Redevelopment & Housi	2013-01-15	16880	2013-01-16	36916	20036	118.6967
Allentown Housing Authority	2014-11-10	26701	2014-11-12	27891	1190	4.4568
ALTOONA HOUSING AUTHORITY	2014-11-24	18863	2014-11-24	25750	6887	36.5106
Area Housing Commission	2013-06-24	19114	2013-06-25	28713	9599	50.2197
Asbury Park Housing Authority	2014-05-21	14987	2014-06-20	18685	3698	24.6747
ASHTABULA METROPOLITAN HOUSING A	2014-04-24	13920	2014-06-03	37948	24028	172.6149
Athens Metropolitan Housing Auth	2014-05-21	10996	2014-05-22	21816	10820	98.3994
Barre Housing Authority	2014-06-16	16757	2014-06-18	19254	2497	14.9012
Batavia Housing Authority	2014-12-30	14576	2015-01-28	26365	11789	80.8795
Bay City Housing Commission	2014-01-28	16470	2014-01-28	27944	11474	69.6661
Bayonne Housing Authority	2014-09-10	10663	2014-09-11	16280	5617	52.6775
Beloit Housing Authority	2013-05-13	24295	2014-04-30	35276	10981	45.1986
Benton Harbor Housing Commission	2014-10-27	18026	2014-10-27	36524	18498	102.6184
Bergen County Housing Authority	2014-05-19	29599	2014-05-20	33617	4018	13.5748
Bloomfield Housing Authority	2014-04-15	18577	2015-01-27	39447	20870	112.3432
BLUE EARTH COUNTY EDA	2015-01-14	18784	2015-01-15	37189	18405	97.9823
BOAZ HOUSING AUTHORITY	2014-04-01	12026	2014-04-02	21751	9725	80.8665
Boston Housing Authority	2014-02-28	17759	2014-10-14	29225	11466	64.5644
Bristol Housing Authority	2013-08-22	35306	2014-10-07	39542	4236	11.9980
Bristol Redevelopment & Housing	2013-06-17	17375	2013-06-18	35983	18608	107.0964
Brockton Housing Authority	2013-03-12	14155	2014-05-20	36225	22070	155.9166
Brownsville Housing Authority	2014-10-06	29245	2014-12-08	35468	6223	21.2789
Bucks County Housing Authority	2014-03-10	30807	2014-03-17	38820	8013	26.0103
Bureau County Housing Authority	2014-06-09	16452	2014-08-21	37026	20574	125.0547
BUTLER METROPOLITAN HOUSING AUTH	2013-07-17	34007	2013-07-26	37046	3039	8.9364
Cambridge Housing Authority	2014-04-03	15055	2014-04-08	20975	5920	39.3225
Central Falls Housing Authority	2014-12-29	14766	2015-01-15	35783	21017	142.3337
Chattanooga Housing Authority	2013-03-11	14629	2013-03-12	19525	4896	33.4678
Chester Housing Authority	2014-10-02	22919	2014-10-03	36974	14055	61.3247
Chillicothe Metropolitan Housing	2014-04-16	21573	2014-05-27	35637	14064	65.1926
City of Clay Center	2014-04-29	33064	2014-04-29	37716	4652	14.0697
CITY OF RENO HOUSING AUTHORITY	2014-10-06	12177	2014-10-06	27158	14981	123.0270