Assignment 3: Public Housing Inspections Star Schema

Background:

Public housing in the United States is administered by local or regional Public Housing Agencies (PHAs), under the supervision of the Federal Department of Housing & Urban Development (HUD). These agencies conduct periodic inspections of the developments in their jurisdiction and assign an inspection score for that development at a certain cost to the US Taxpayers.

You are a Developer hired by HUD to assist with a dimensional model for the inspection data, and to provide a key analysis for Senior Management as your deliverables.

You have access to a flat file of inspections including the PHA name, development name being inspected (along with address), as well as the date of inspection, inspection score which is expressed as a *ratio or percentage* from 0 to 100, and the cost of performing that inspection in $$.

Your boss in IT wants you to answer the following questions based on your inspection of the dataset:

Question 1

How many facts are there in this dataset?
There are 2 facts in dataset.

Which facts do you identify?
"COST_OF_INSPECTION_IN_DOLLARS" and "INSPECTION_SCORE" are facts in the dataset.

For the facts that you identify, what type of facts are they?
They are quantitative facts because they record numerical values that can be used for comparison. At the same time, they are transactional facts because these data capture some transactional information about each inspection event.

Question 2

How many dimensions are there in this dataset?
There are 6 dimensions in this dataset.

Which dimensions do you identify?
They are "PUBLIC_HOUSING_AGENCY_NAME", "INSPECTED_DEVELOPMENT_NAME", "INSPECTED_DEVELOPMENT_ADDRESS", "INSPECTED_DEVELOPMENT_CITY", "INSPECTED_DEVELOPMENT_STATE", and "INSPECTION_DATE".

Question 3

Senior management is interested in viewing the facts identified above, at both the inspection level, as well as a periodic summary of inspection costs for each month. Based on this context, if you were to store these data in a set of fact tables, which type (or types) of fact tables would you use and why?

I think choosing a transactional fact table is more appropriate. We can use this table to record information about each inspection. Additionally, I would also consider an accumulating snapshot fact table because it can clearly help us track and record periodic data.

Question 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 context, how would handle this slowly changing dimension? Select from types 0,1,2, or 3 from the Kimball reading. Justify your answer.

There are typically four types of Slowly Changing Dimensions (SCD). Among them, Type 2 retains historical records by creating a new record for each attribute change. For our situation, I believe this type is the most suitable. Type 2 can record historical information in detail, which is very useful for tracking cases of information integration.

Question 5

Finally, Senior Management is interested in a subset of this data, for only those PHAs that saw an *increase* in the $$ cost of performing an inspection in their jurisdiction. Since none of them are SQL programmers, they've asked your help in performing this analysis by providing a file as your final deliverable with the following columns:

Note that MR stands for "most recent":

PHA_NAME,
MR_INSPECTION_DATE,
MR_INSPECTION_COST,
SECOND_MR_INSPECTION_DATE,
SECOND_MR_INSPECTION_COST,
CHANGE_IN_COST
PERCENT_CHANGE_IN_COST

Management has asked that you perform this function using lead or lag functions in SQL.

This is my answer:

PHA_NAME	MR_INSPECTION_DATE	MR_INSPECTION_COST	SECOND_MR_INSPECTION_DATE	SECOND_MR_INSPECTION_COST	CHANGE_IN_COST	PERCENT_CHANGE_IN_COST
Akron Metropolitan Housing Autho	2014-10-09	25593	10/8/2014	15626	9967	63.7847
Alachua County	2015-01-22	37345	5/1/2014	17019	20326	119.4312
Alaska Housing Finance Corporati	2014-11-14	26342	11/13/2014	21366	4976	23.2893
Albany Housing Authority	2015-01-12	31115	1/9/2015	30247	868	2.8697
Alexander County Housing Authori	2014-11-18	31272	4/24/2014	18855	12417	65.8552
Alexandria Redevelopment & Housi	2014-05-09	29123	4/18/2014	14767	14356	97.2168
ALLEGHENY COUNTY HOUSING AUTHORI	2015-02-02	37108	2/2/2015	36454	654	1.7940
Allentown Housing Authority	2014-11-17	34040	11/14/2014	18989	15051	79.2617
ALTOONA HOUSING AUTHORITY	2014-11-24	25750	9/15/2014	24813	937	3.7762
ANNISTON HA	2014-12-30	31506	8/21/2014	10785	20721	192.1280
Area Housing Commission	2013-06-25	28713	6/24/2013	19114	9599	50.2197
Asbury Park Housing Authority	2014-06-03	35723	5/21/2014	14987	20736	138.3599
Ashland Housing Authority	2014-04-29	29106	4/29/2014	17510	11596	66.2250
ASHTABULA METROPOLITAN HOUSING A	2014-06-03	37948	4/24/2014	13920	24028	172.6149
Athens Metropolitan Housing Auth	2014-05-22	21816	5/21/2014	10996	10820	98.3994
Aurora Housing Authority	2015-02-02	14683	6/24/2014	12831	1852	14.4338
Aurora Housing Authority ofthe C	2014-07-03	14908	6/11/2013	14570	338	2.3198
Austin Housing Authority	2014-06-30	36672	6/26/2014	25920	10752	41.4815
Barre Housing Authority	2014-06-18	19254	6/16/2014	16757	2497	14.9012