IST722: Unit 07 Participation Questions

This is an individual assignment.

Before you begin, please make sure you've read and understand 1) our class honor code, 2) course policies on late work and 3) participation policies as posted on the syllabus. "I didn't know" is not an excuse.

You should cite your sources in a standard format like MPA or APA and include a list of works cited.

Your Name:	Rashika Singh
Your Email:	rsingh37@syr.edu

Instructions

Answer each of the following questions as concisely as possible. More is not necessarily better. Please justify your answer by citing your sources from the assigned readings from our textbooks, our class lectures, or online if directed to do so. Be sure to cite in text and include a list of works cited. Place your answer below each question. When you're finished, print out this document and bring it to class as part of your participation grade.

Questions

[1] Explain what is meant by the surrogate key pipeline in your own words. Keep this brief.
Ans: Maintain a surrogate key lookup table for each dimension. This table is updated whenever a new record is added or when a SCD-2 update occurs in an existing dimensional entity. The dimensions must be updated with Type-2 updates before any facts are loaded into the Data Warehouse, to guarantee referential integrity. This is known as Surrogate key pipeline.
Late arriving fact:
Suppose we receive today a purchase order that is one month old and our dimensions are SCD type -2 dimensions
We insert this late arriving fact into the correct historical position, even though our sales summary will change
For each dimension, find the corresponding dimension record in effect at the time of the purchase
Using the surrogate keys thus, replace the incoming natural ID's with the surrogate keys. Insert the late arriving record in the correct partition of the table. When we load fact table we assume that all records are current. This operation refers to a surrogate key pipeline where the source data for the fact table contains natural keys that must be replaced by surrogate keys before the records are loaded. They are found through the line of little lookup surrogate tables which are updated every time a new dimension entity is created or changed.

[2] What is the purpose of the lookup transformation? How many attributes must match for the lookup to succeed?
The purpose of lookup transformation is to replace natural keys present in the fact with surrogate keys of that dimension to keep the table updated of any changes. The natural key of all dimension tables present in the fact table must match with the surrogate keys for the lookup to be successful

[3] What is Upsert? When is Loading via Upsert used?
The Upsert wizard combines inserting and updating a record. If a record in your file matches an existing record, the existing record is updated with the values in your file. If no match is found, a new record is created. Loading via Upsert is

used in a SCD type-1 and SCD type-2 change when we need to load data into the fact table .After we have computed the facts and loaded the dimension keys using the surrogate key pipeline we put data into fact table .If there is a match in the records via the wizard lookup the data is updated through a SCD type-1 change. If there is no match then a new insert is made on the existing record and previous record is retained to preserve history.

For example

ID	Name	Department	RowisCurrent
1	Rashika Singh	Product	0
2	Rashika Singh	Sales	1

[4] When you find a data type or length mismatch, where should you fix the issue as best practice: source, stage, dw?

Ans:

If we find a data type or length mismatch the best practice is to fix the issue in the stage and DW .We cannot make change in source and thus it is untouched.

[5] A fellow student reported:

> *I am stuck in my ETL. After troubleshooting I found that it's the DayofYear column not matching in type. Here's the specific error message: "Cannot map columns of different types. Column 'DayOfYear' is of type 'System.Int32' and column 'YearDay' is of type 'System.Int16'." I need help.*
>
> **What would be your advice to your fellow student?**

Ans:

I would suggest my peer to typecast the datatype Dayofyear from int to tinyint and would make the change in the stage or warehouse stage as schema is flexible.

WORKS CITED:

Fudge slides and videos