IST722: Class Exercise 7

**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: | Yunhan Zhang |
| Your Email: | yzhan297@syr.edu |

# Instructions (Refer Units 6 & 7)

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] How can you identify changes to a business entity when there is no natural key?

**You can identify changes to a business entity using CRC/Hash which allows to perform compare for assigning a cryptographic hash to data and comparing it to incoming data to detect changes if there is no natural key.**

[2] What are CET and LSET? How are they used in data warehousing?

**CET is the current timestamp of data warehouse and LSET is when the last data extraction happened successfully. They are used for incremental data extraction which is only the data which is modified. And when EXTRACTION is done, LSET is set to CET.**

[3] Briefly explain late-arriving dimensions and late-arriving facts, and how you would manage them.

**The late-arriving dimensions only occurs when you have all of the dimension data arriving in data warehouse later than the fact data that references that dimension. A fact row is late arriving if the most current dimensional context for new fact rows does not match the incoming row. This happens when the fact row is arriving late. We can find every column which holds the dimension natural key for managing late arriving late dimensions.**

[4] Explain what is meant by the surrogate key pipeline in your own words. Keep this brief.

**Surrogate key pipeline maintains a surrogate key lookup table for each of the dimension. The table is updated whenever a new record is added or when a SCD-2 update occur in an existing dimension entity. And the dimensions must be updated with Type -2 updates before any facts are loaded into the data warehouse, to guarantee integrity.**

[5] What is the purpose of the lookup transformation? How many attributes must match for the lookup to succeed?

**The lookup transformation is used to replace keys present in the fact with surrogate keys of that dimension to keep the table updated of any changes. And the natural key of all dimension tables will present in the fact table must match with the surrogate keys for the lookup to be working.**

WORKS CITED:

**Professor Fudge’s video and slides.**

**Wikipedia.**