

# IST722: Unit 06 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] What is the difference between ETL and ELT?

ETL stands for Extract, Transform and Load. It is the process of retrieving data from OLTP sources, transforming it and placing it into the data warehouse. It is moving the data around the data warehouse and takes 70% of the data warehouse effort. ELT stands for Extract, Load and Transform and the only difference is the location where the transformational logic takes place. It is used mainly in MPP data warehouse as MPP can handle high workloads effectively.

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

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.

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

CET is Current Extraction Timestamp which is the current timestamp of the data warehouse. LSET is the Last Successful Extraction Timestamp which is when the last data extraction happened successfully. They are used for incremental data extraction which is only the data which is modified. You check the timestamp when a record is added and if this is greater than LSET you extract the data or when last updated timestamp is greater than LSET. When extraction is complete, LSET is set to CET.

[4] Explain snapshotting and its purpose in data warehousing.

Snapshotting allows you to see the state of data at regular intervals. For example, if employee changes his address, the warehouse is alerted for a snapshot. It is used to add time variance to data. If data is updated in the OLTP and when the next snapshot is taken, if the last updated data is greater than LSET than you add this data to warehouse to see how data changed over time.

[5] What are 3 ways for matching data that are character strings?

The three ways of matching data that are character strings are:

1. Exact match- all the characters in the string must be exactly the same
2. Fuzzy logic- Generates similarity score and cutoff to determine if there is a match in strings (example- 0.5 for 50% match)
3. Rule-based logic- It decides based on the context. Example- book=novel

WORKS CITED:

Professor Fudge videos and presentations