Summary & Highlights

- ETL stands for Extract, Transform, and Load
- Loading means writing the data to its destination environment
- Cloud platforms are enabling ELT to become an emerging trend
- The key differences between ETL and ELT include the place of transformation, flexibility, Big Data support, and time-to-insight
- There is an increasing demand for access to raw data that drives the evolution from ETL, which is still used, to ELT, which enables ad-hoc, self-serve analytics
- Data extraction often involves advanced technology including database querying, web scraping, and APIs
- Data transformation, such as typing, structuring, normalizing, aggregating, and cleaning, is about formatting data to suit the application
- Information can be lost in transformation processes through filtering and aggregation
- Data loading techniques include scheduled, on-demand, and incremental
- Data can be loaded in batches or streamed continuously

1. ETL process consists of Extract > Transform > Load. Which of these three processes is also known as data wrangling?

- (●) Transform

  Correct! This process wrangles the data into the format suitable for destination and use.

- ( ) Extraction
- ( ) Load
- ( ) Data wrangling is a term for another data warehouse process

2. What is the main difference between the ELT and ETL process?

- ( ) ELT is only used for analyzing
- (●) Order of stages

  Correct! ELT and ETL have similar stages but the order they process is different.

- ( ) Data types
- ( ) ETL used primarily for cloud

3. Transformations for ETL happen in the data pipeline. Where do transformations happen for ELT?

- ○ Extraction process
- ○ Load process
- ○ Source environment
- ◉ Destination environment

Correct! ELT transformations happen in the destination environment.

4. Which of the following raw data sources is related to sales?

- ◉ Transactional data

Correct! Transactional data from business, financial, real estate, and point-of sale (POS) transactions.

- ○ Survey data
- ○ Analog data
- ○ Merchandise data

5. Which of the following data transformation techniques will be suitable for 'unlike' data sources?

- ○ Filtering, sorting, aggregation
- ○ Cleaning
- ◉ Joining or merging

Correct! This technique addresses disparate data sources.

- ○ Data structuring

2. The ELT process has no information loss. What is the main reason for this benefit?

- ○ Data source integration
- ◉ Data is acquired and directly loaded, as-is, into its destination environment.
- ○ There is a separation between moving and processing data.
- ○ It separates the data pipeline from processing.

Correct
Feedback: Correct! ELT provides a replica of the source data, and with that, no information loss occurs.

3. Which of the following in an ELT process best compares to the "Staging area" in the ETL process?

- ○ Transformed data storage
- ○ Database servers
- ◉ Data lake in ELT process
- ○ Storage for source data in the ELT process

Correct
Correct! The staging area fits the description of a data lake, which is a modern self-serve repository for storing and manipulating raw data.

Question 2

The ELT process has no information loss. What is the main reason for this benefit?

✅ **Data is acquired and directly loaded, as-is, into its destination environment.**

(ELT loads raw data first without transformation, preserving all original information.)

Question 3

Which of the following in an ELT process best compares to the "Staging area" in the ETL process?

✅ **Data lake in ELT process**

(The data lake serves as the initial storage for raw data, similar to a staging area in ETL.)

4. Which of the following pain points does ELT address?

- (●) Challenges imposed by Big Data
- ( ) Request for fixed processes
- ( ) Transformation in data pipeline
- ( ) Cost-effectiveness

Correct
Correct! Challenges like scalability imposed by Big Data are addressed.

5. There are many techniques for extracting data. What does the choice of technique depend on?

- ( ) Operating system
- ( ) Type of client
- ( ) Optical or analog
- (●) Kind of data source and intended use

Correct
Correct! It depends on the kind of data source and intended use of the data.

Question 4

Which of the following pain points does ELT address?

✅ **Challenges imposed by Big Data**

(ELT is better suited for handling large-scale, unstructured data compared to traditional ETL.)

Question 5

There are many techniques for extracting data. What does the choice of technique depend on?

✅ **Kind of data source and intended use**

(The extraction method varies based on whether the source is a database, API, IoT device, etc., and how the data will be used.)

6. Extracting data from IoT devices involves large volumes of redundant data. What is used to decrease the data volume of redundant data and only extract features of interest from raw data?

- ○ Biometric sensors
- ○ APIs
- ○ SQL languages
- ◉ Edge computing

Correct
Correct! Edge computing reduces the data volumes of redundant data by extracting features of interest from the raw data.

7. ETL uses the schema-on-write approach. What is the biggest disadvantage of this approach?

- ○ Consistency
- ◉ Limited versatility
- ○ Stability
- ○ More data access

Correct
Correct! The ETL approach limits the versatility whereas the ELT approach is versatile since it obtains multiple views of the same source data with ad-hoc schemas.

Question 6

Extracting data from IoT devices involves large volumes of redundant data. What is used to decrease the data volume of redundant data and only extract features of interest from raw data?

✅ **Edge computing**

(Edge computing processes data at the source, reducing redundancy before transmission.)

Question 7

ETL uses the schema-on-write approach. What is the biggest disadvantage of this approach?

✅ **Limited versatility**

(Schema-on-write requires defining structure upfront, making it less flexible for evolving data needs.)

8. Why is there no information loss in ELT unlike ETL where there is loss of information?

- ○ Because ELT uses edge computing
- ◉ Because the data is copied as is
- ○ Because ETL uses aggregation of data but ELT doesn't
- ○ Because ELT involves lossy data compression

Correct
Correct! All the original information content is left intact as the data is copied as is.

9. Which of these is most useful for incremental loading strategy?

- ○ Only stream loading
- ○ Only batch loading
- ◉ Both batch and stream loading
- ○ File partitioning

Correct
Correct! Stream loading is useful for incremental loading and should be real-time. Batch loading is useful when the incremental loading can be in batches.

Question 8

Why is there no information loss in ELT unlike ETL where there is loss of information?

✅ **Because the data is copied as is**

(ELT preserves raw data before transformation, whereas ETL may discard or aggregate data during staging.)

Question 9

Which of these is most useful for incremental loading strategy?

✅ **Both batch and stream loading**

(Incremental loading can use batch for periodic updates or streaming for real-time changes.)

10. Which of the following loading techniques can split a single file into smaller chunks?

- ○ Batch loading
- ○ Scheduled loading
- ○ Stream loading
- ◉ Parallel loading

Correct
Correct! Parallel loading technique splits single files into small chunks and loads them simultaneously.

Question 10

Which of the following loading techniques can split a single file into smaller chunks?

✅ **Parallel loading**

(Parallel loading divides files into parts for faster processing.)