1. A data pipeline has three main processes. When working with data, how should a cloud data analyst follow these processes?

- ✅ **The analyst should tailor the processes for each data problem.**
  Correct: Data pipelines should be adapted to the specific context and business needs, not followed rigidly.
- The analyst should follow the processes in a strict order: extract, load, and transform.
- The analyst should design a data problem to align with the order of the processes.
- The analyst should follow the processes using the pipeline methodology.

---

2. What does the “E” stand for in the ETL pipeline technique?

- Expand
- ✅ **Extract**
  Correct: "E" stands for Extract — the first step in ETL where data is pulled from source systems.
- Elaborate
- Explore

---

3. A cloud data analyst is designing a data pipeline. What is the first step they need to design?

- Exploration
- ✅ **Ingestion**
  Correct: Ingestion is the first step, where raw data is collected from various sources.
- Inclusion
- Elaboration

---

4. Which technique for collecting and moving raw data sources to a staging area requires less computing power and storage space?

- ✅ **Batch ingestion**
  Correct: Batch ingestion processes data in scheduled chunks, using fewer resources than streaming.
- Streaming ingestion
- Bundle ingestion
- Schedule ingestion

---

5. A cloud data team is working on a data pipeline. What is the next step in the data pipeline after the team collects and moves the data into a staging area?

- Transforming the data
- Exploring the data
- Ingesting the data
- ✅ **Mapping the data**
  Correct: Mapping defines how data fields relate to each other and prepares them for transformation.

---

6. A cloud data analyst is in the process of fixing data quality issues. What process did the cloud data analyst use to identify those issues?

- Data mapping
- Data ingestion
- ✅ **Data profiling**
  Correct: Data profiling helps identify anomalies, patterns, and quality issues in the dataset.
- Data cleaning

---

7. A cloud data analyst is working with a dataset. They decide to join some of the fields and to add a new field to the dataset. What is the cloud data analyst doing?

- Manipulating the data using data conversion
- Manipulating the data using data cleaning

- ✅ **Manipulating the data using data enrichment**
  Correct: Data enrichment involves enhancing the dataset by adding new or derived fields.
- Manipulating the data using data standardization

---

8. A data analyst is in the loading stage of the data pipeline. They decide to load only the data that has changed since the last loading stage. What type of loading is the cloud data analyst using?

- Batch
- ✅ **Incremental**
  Correct: Incremental loading updates only new or modified data, improving efficiency.
- Schedule
- Streaming

---

9. What process can be performed at any stage of a data pipeline but is especially important in the load stage?

- Data exploration
- Data ingestion
- ✅ **Data validation**
  Correct: Validation ensures data integrity and correctness before final storage or analysis.
- Data transformation

---

10. A cloud data analyst is checking the dataset before loading it into their final destination for analysis. They verify that there are no empty or missing values. What is the cloud data analyst doing?

- Validating the data using duplicate validation
- ✅ **Validating the data using null validation**
  Correct: Null validation checks for missing or empty values in the dataset.
- Validating the data using range validation
- Validating the data using format validat

---

11. What does the “L” stand for in the ETL pipeline technique?

- Lead
- Logs
- ✅ **Load**
  Correct: "L" stands for Load — the final step where transformed data is stored in its destination.
- Learn

---

12. A data team is working on designing a pipeline. As the first step, they need to choose a technique that allows them to process the data as soon as it is available. What technique should the data team use to move the raw data into the staging area?

- Batch ingestion
- ✅ **Streaming ingestion**
  Correct: Streaming ingestion allows real-time data processing as soon as data is available.
- Bundle ingestion
- Schedule ingestion

---

13. A cloud data analyst is working with a dataset. They are standardizing and converting the data. What is the cloud data analyst doing?

- Exploring the data
- Ingesting the data
- ✅ **Manipulating the data**
  Correct: Standardizing and converting data are forms of manipulation to prepare it for analysis.
- Profiling the data

---

14. What is the term for the process of checking the quality of data to verify that it is complete, accurate, secure, and consistent?

- Data transformation
- ✅ **Data validation**
  Correct: Validation ensures data meets quality standards before use.
- Data exploration
- Data ingestion

---

15. A cloud data analyst is checking the dataset before loading it into their final destination for analysis. They verify that all the data in the date fields has the format "MM/DD/YYYY". What is the cloud data analyst doing?

- Validating the data using null validation
- Validating the data using duplicate validation
- Validating the data using range validation
- ✅ **Validating the data using format validation**
  Correct: Format validation ensures values follow expected patterns, like date formats