

1. What is the first stage of the ETL process?

1 / 1 point

- ☐ Loading
- ☐ Cleaning
- ☐ Transformation
- ☒ Extraction

✓ **Correct**

The ETL stages are Extraction, Transformation, and Loading.

2. Which of these transformations is correctly described?

1 / 1 point

- ☒ Normalizing: Converting data to common units
- ☐ Cleaning: merging disparate data sources
- ☐ Data Structuring: Fixing any errors or missing values
- ☐ Sorting: selecting only what is needed

✓ **Correct**

Cleaning is fixing errors. Sorting is order the data for searching. Data structuring is converting one data format to another.

3. Which of these is NOT an example of a system in the data load phase?

1 / 1 point

- ☐ A data warehouse
- ☒ A scanned medical document
- ☐ An Excel spreadsheet
- ☐ A comma separated file

✓ **Correct**

A scanned medical document is an example of data in the Extraction phase.

4. Select the correct statement regarding ETL workflows as data pipelines.

1 / 1 point

- ☐ Bottlenecks within the pipeline can often be handled by anonymizing slower tasks.
- ☒ Overall accuracy of the ETL workflow has been a more important requirement than speed.
- ☐ With conventional ETL pipelines data is processed in real time.
- ☐ Data is fed through a data pipeline in large packets.

✓ **Correct**

Accuracy is valued over speed, although efficiency is usually an important factor in minimizing resource costs.

5. Select the correct statement regarding batch processing.

1 / 1 point

- ☐ Data is processed in batches, usually on a weekly schedule.
- ☐ Batch processing triggers are rarely on demand.
- ☐ When an event of interest occurs, such as an intruder alert, the interval would be periodic.
- ☒ Batch processing intervals can be triggered by events.

✓ Correct

Examples include when the data reaches a certain size, an event of interest occurs, or on demand.

6. ETL pipelines are frequently used to integrate data from disparate and usually _____ systems within the enterprise.

1 / 1 point

- ☐ aggregating
- ☐ simultaneous
- ☒ siloed
- ☐ batched

✓ Correct

The systems within the enterprise are usually separate, that is siloed.

7. Select the correct statement regarding Apache Airflow.

1 / 1 point

- ☐ Apache Airflow represents the workflow in DAGs, but not in code.
- ☐ Apache Airflow tasks can be expressed as Python, but not Bash.
- ☐ Apache Airflow is a well-known commercial tool.
- ☒ Apache Airflow is a workflow orchestration tool.

✓ Correct

Apache Airflow breaks down details of the workflow into tasks and dependencies.

8. Bash uses _____ to turn your file into a Bash shell script.

1 / 1 point

- ☐ loadstat
- ☐ crontab
- ☒ shebang
- ☐ getstat

✓ Correct

ETL pipelines can be built from basic Bash scripts. In the editor, type in the Bash shebang to turn your file into a Bash shell script.

9. SSIS, Amazon Redshift, IBM InfoSphere Information Server, and Oracle GoldenGate are examples of _____.

1 / 1 point

- ☐ Popular open-source ETL tools.
- ☒ Popular commercial ETL tools.
- ☐ Popular open-source ELT tools
- ☐ Popular commercial ELT tools.

✓ **Correct**

There are many commercial ETL tools available. These of some of the most well-known.

10. ETL jobs can be run on a schedule using _____.

1 / 1 point

- ☐ getstat
- ☐ loadstat
- ☐ shebang
- ☒ crontab

✓ **Correct**

ETL jobs can be run on a schedule using crontab. Open the crontab editor, schedule your job to run, for example, every minute of every day, close the editor, and save your edits.