

**Your grade: 100%**Your latest: **100%** • Your highest: **100%** • To pass you need at least 70%. We keep your highest score.**Next item →****1.** Select the correct statement regarding Apache Airflow.**1 / 1 point**

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

 **Correct**

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

**2.** What is the first stage of the ETL process?**1 / 1 point**


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

 **Correct**

The ETL stages are Extraction, Transformation, and Loading.

**3.** Select the correct statement regarding batch processing.**1 / 1 point**

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

 **Correct**

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

**4.** Bash uses \_\_\_\_\_ to turn your file into a Bash shell script.**1 / 1 point**

- ☒ shebang
- ☐ loadstat
- ☐ crontab
- ☐ 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.

**5.** SSIS, Amazon Redshift, IBM InfoSphere Information Server, and Oracle GoldenGate are examples of \_\_\_\_\_.**1 / 1 point**

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

✔ Correct

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

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

1 / 1 point

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

✔ Correct

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

7. 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.
- ☐ Data is fed through a data pipeline in large packets.
- ☒ Overall accuracy of the ETL workflow has been a more important requirement than speed.
- ☐ With conventional ETL pipelines data is processed in real time.

✔ Correct

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

8. Which of these transformations is correctly described?

1 / 1 point

- ☐ Data Structuring: Fixing any errors or missing values
- ☐ Cleaning: merging disparate data sources
- ☒ Normalizing: Converting data to common units
- ☐ 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.

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

1 / 1 point

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

✔ Correct

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

10. ETL jobs can be run on a schedule using \_\_\_\_\_.

1 / 1 point

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

✔ 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.

