

# Final Quiz

Quiz ID: 1234567

Submit your assignment

Due Jul 16, 11:59 PM EDT Attempts 2 every 24 hours

Receive grade

To Pass: 80% or higher

Like Dislike Report an issue

Start assignment

Your grade

-

1. What is meant by the term "data transformation"?

- ☒ Moving data ready available for ingestion by analytics applications so that end users can gain value from it.
- ☐ Processing data to make it conform to requirements.
- ☐ Writing data to some new destination environment.
- ☐ Configuring access to the data and reading it into an application.

23/47

1 point

Expand

2. Which of the following is an advantage of ELT compared to ETL?

- ☒ There is no information loss with ELT because you are working with a replica of the data.
- ☐ It is easier to fix errors or spot missing values with ELT.
- ☐ With ELT, processing the data makes it conform to the requirements of both the target system and the intended use case for the current data.
- ☐ With ELT, the data is acquired and prepared for subsequent use in an analytics environment.

23/47

1 point

Expand

3. Select the correct statement regarding ETL.

- ☐ ETL happens in the destination environment at all.
- ☐ Transformations for ETL are decoupled from the data pipeline.
- ☐ ETL is flexible, making data readily available for self-serve analytics.
- ☒ Transformations happen within the data pipeline.

23/47

1 point

Expand

4. Which of the following best describes the function of analog-to-digital conversion?

- ☒ Digitizing analog audio recordings and signals into a computer readable format.
- ☐ Interpreting and digitizing text scanned from paper documents so it can be stored in a computer readable format.
- ☐ Capturing and digitizing comments, questionnaires, and vital statistical data obtained through polling and survey methods.
- ☐ Crawling through web pages in search of text, images, tables, and hyperlinks.

23/47

1 point

Expand

5. What is stream loading?

- ☐ Stream loading refers to loading data in chunks defined by some time windows of data accumulated by the data source.
- ☒ Stream loading refers to loading data in real time as it becomes available.
- ☐ Stream loading refers to loading data on a schedule.
- ☐ Stream loading refers to loading an initial history into a database.

23/47

1 point

Expand

6. How is shell scripting used to implement an ETL pipeline?

- ☒ By building a basic Bash script.
- ☐ By using the command editor.
- ☐ By using Python to build the shell.
- ☐ By using an API.

23/47

1 point

Expand

7. Which of the following best describes latency?

- ☒ The sum of the times a packet spends at each stage in the pipeline.
- ☐ The average amount of time a packet spends at each stage in the pipeline.
- ☐ The amount of data that can be fed through the pipeline per unit of time.
- ☐ The size of a packet.

23/47

1 point

Expand

8. Which of these fall into the category of data pipeline monitoring?

- ☐ Extraction and ingestion.
- ☐ Loading and scheduling.
- ☐ Latency and throughput.
- ☒ Scheduling and maintenance.

23/47

1 point

Expand

9. When is stream processing used instead of batch processing?

- ☐ When processing must be done on a fixed schedule, ranging from hours to weeks apart.
- ☐ When accuracy is more critical than immediate processing.
- ☐ When processing is triggered by the arrival of data meeting a certain size.
- ☒ When results are required with minimal latency, essentially in real time.

23/47

1 point

Expand

10. Which of the following is popular and versatile programming environment for building data pipelines?

- ☐ Data Frame
- ☐ PySpark
- ☐ Talend
- ☒ AWS Glue

23/47

1 point

Expand

11. Which of the following are the four principles Apache Airflow is built upon?

- ☐ Robust, scalable, effective, dynamic.
- ☒ Scalable, dynamic, schedulable, lean.
- ☐ Effective, simple to use, scalable, agile.
- ☐ Sustainable, competitive, agile, simple to use.

23/47

1 point

Expand

12. Which statement is true about directed acyclic graphs (DAGs)?

- ☐ Some DAGs have loops.
- ☒ All DAGs are trees but not all trees are DAGs.
- ☐ All DAGs have loops.
- ☐ All trees are DAGs but not all DAGs are trees.

23/47

1 point

Expand

13. Which of these are ways to visualize DAGs in Apache Airflow?

- ☐ DAGs can be visualized in Tree View and Dashboard View.
- ☐ DAGs can be visualized as a Basic Flow and as a Task Instance.
- ☒ DAGs can be visualized in Tree View and Graph View.
- ☐ DAGs can be visualized as Stage View and as Code.

23/47

1 point

Expand

14. In the Apache Airflow DAG, which code block contains the nodes of the DAG?

- ☐ DAG argument specification
- ☐ Task definitions
- ☐ Task pipeline
- ☒ DAG definition

23/47

1 point

Expand

15. In Airflow, which is an example of a counter metric?

- ☒ The latency of a task.
- ☐ The number of failed tasks.
- ☐ The amount of time it takes to complete a task.
- ☐ The number of currently running tasks.

23/47

1 point

Expand

16. Which statement best describes event streaming?

- ☐ Event streaming is the middle layer among various event sources and destinations.
- ☐ Event streaming is the act of storing events being received from event sources.
- ☐ Event streaming is the generation of a large event volume at a short time interval or nearly in real time.
- ☒ Event streaming is the continual transportation between an event source and an event destination.

23/47

1 point

Expand

17. Select the correct statement regarding Apache Kafka.

- ☐ Kafka is used primarily for measures in the educational sector.
- ☐ Kafka is used primarily for measuring keyboard strokes, page views, and screen time.
- ☐ Kafka is a niche streaming platform, used almost exclusively in banking.
- ☒ Kafka is a multi-tenant, used to track user activities such as mouse clicks, but is now suitable for other metrics streaming.

23/47

1 point

Expand

18. Select two common components of an event streaming platform (ESP).

- ☐ Event pipeline
- ☒ Event transportation
- ☒ Event storage
- ☐ Event analysis

23/47

1 point

Expand

19. Select the correct statement regarding Kafka's core components.

- ☐ A core component is Producers, that divide topics into different brokers.
- ☐ A core component is Consumers, that receive, store, and process events.
- ☒ A core component is Replicators, that duplicate partitions into different brokers.
- ☐ A core component is Servers, which are applications that publish events into topics.

23/47

1 point

Expand

20. Which of the following describes a function of the Kafka Streams API?

- ☒ It ensures that a record is only processed once.
- ☐ It publishes a received stream to a Kafka topic.
- ☐ It stores raw data.
- ☐ It publishes events.

23/47

1 point

Expand

Update to submit