

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

1 / 1 point

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

 Expand

 **Correct**


Correct! To extract data is to configure access to it and read it into an application.

2. Which of the following best describes a use case for ELT?

0 / 1 point

- ☐ When a very large amount of information is either already recorded or being generated, but is not yet captured, or accessible.
- ☒ When data is being live streamed from a single location.
- ☐ When data is structured and relational rather than unstructured.
- ☐ When moving data is usually more of a bottleneck than processing it.

 Expand

 **Incorrect**

Incorrect. Review the ELT Basics video.

3. Which of the following best describes a driving factor of the evolution from ETL to ELT?

1 / 1 point

- ☐ The need to transfer data from a legacy system to an updated one.
- ☐ The need to decouple transformations from the data pipeline.
- ☒ The demand to release raw data to a wider user base.
- ☐ The need to build a single dashboard from multiple data sources.

 Expand

 **Correct**

Correct! One of the factors driving the evolution of ETL to ELT is the demand to release raw data to a wider user base for the enterprise.

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

1 / 1 point

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

 Expand

 **Correct**

Correct! Analog digital converters digitize analog audio recordings and signals.

5. What is stream loading?

1 / 1 point

- ☐ 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.

 Expand

 **Correct**

Correct! Stream loading is loading data in real time as it becomes available.

6. How can an ETL job be scheduled to run?

1 / 1 point

- ☐ By loading statistics into the reporting system using the supplied API.
- ☒ By creating a cron job for your Bash script.
- ☐ By running a Bash script.
- ☐ By using the touch command on the command line.

 Expand

 **Correct**

Correct! An ETL job can be scheduled to run by creating a cron job for your Bash script.

7. Which of the following best describes throughput?

1 / 1 point

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

 Expand

 **Correct**

Correct! Throughput refers to how much data can be fed through the pipeline per unit of time.

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

1 / 1 point

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

 Expand

 **Correct**

Latency is the time it takes for data packets to flow through the pipeline. Throughput is the volume of data passing through the pipeline over time.

9. Which of the following statements is true in regard to stream processing?

1 / 1 point

- ☐ Stream processing is when a pipeline contains many transformations.
- ☐ Stream processing is when much of the data needs to be cleaned in the pipeline.
- ☐ Stream processing typically operates on a fixed schedule.
- ☒ Records or events are processed immediately as they occur.

[Expand](#)



Correct

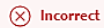
Correct! With streaming pipelines, records or events are processed immediately as they occur.

10. Select the correct statement regarding open source and enterprise ETL and ELT tools.

0 / 1 point

- ☐ Alteryx is specific to ETL pipelines.
- ☐ Panoply is specific to ELT pipelines.
- ☒ Apache Airflow is a useful open-source Python library.
- ☐ Pandas allows you to author, schedule, and monitor Big Data workflows.

[Expand](#)



Incorrect

Review the Data Pipeline Tools and Technologies video.

11. Which of the following best describes Apache Airflow?

1 / 1 point

- ☐ It is primarily a data streaming solution.
- ☒ It is primarily a workflow manager.
- ☐ It is primarily a data pipeline monitoring tool.
- ☐ It is primarily a directed acyclic graph (DAG).

 Expand

 **Correct**

Correct! Unlike Big Data tools such as Apache Kafka, Apache Storm, Apache Spark, or Flink, Apache Airflow is not a data streaming solution. It is primarily a workflow manager.

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

1 / 1 point

- ☐ 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.

 Expand

 **Correct**

Correct! A tree is a commonly used graph for representing directory structures. All trees are DAGs, but not all DAGs are trees.

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

1 / 1 point

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

[Expand](#)

✓ **Correct**

DAGs can be visualized as a Simple Example, Tree View, Graph View, Task Instance Context Menu, and as Code.

14. In the Apache Airflow DAG, in which block includes scheduling instructions?

1 / 1 point

- ☐ Library imports
- ☐ Task definitions
- ☒ DAG definition
- ☐ DAG argument specification

[Expand](#)

✓ **Correct**

Correct! The DAG definition block, also known as instantiation, contains scheduling instructions.

15. Which of the following is a tool that can be used to search, index, and analyze log files?

1 / 1 point

- ☐ Prometheus
- ☐ StatsD
- ☐ IBM Cloud
- ☒ Splunk

 Expand

 **Correct**

Correct! Airflow recommends using Elasticsearch and Splunk, which are two popular document database and search engines, to index, search, and analyze log files.

16. Select the correct statement regarding events.

1 / 1 point

- ☒ Events describe an entity's observable state updates over time.
- ☐ An event is something that no one notices, but it is happening.
- ☐ Events describe a clearly observable entity, such as a car or a hospital patient.
- ☐ The temperature at which water freezes is an example of an event.

 Expand

 **Correct**

An event normally means something worth noticing is happening.

17. Which statement best describes Kafka brokers?

1 / 1 point

- ☐ A Kafka broker is a communication protocol to exchange data between clients and servers.
- ☒ A Kafka broker is a server cluster acting to receive, store, and distribute events.
- ☐ A Kafka broker is a system that ensures events streams are handled in an efficient and collaborative way.
- ☐ A Kafka broker is a collection of shell scripts to communicate with Kafka servers.

 Expand

 **Correct**

Correct! Kafka brokers are clusters with many associated servers acting as the event broker to receive, store, and distribute events.

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

1 / 1 point

☒ Event analysis

 **Correct**

Correct! The third common component is the analytic and query engine which is used for querying and analyzing the stored events.

☒ Event storage

 **Correct**

Correct! The second common component of an ESP is Event Storage, which is used for storing events being received from event sources.

☐ Event pipeline

☐ Event transportation

 Expand

 **Correct**

Great, you got all the right answers.

19. What is the function of the ZooKeeper in Kafka?

1 / 1 point

- ☒ To synchronize and manage brokers
- ☐ To log events
- ☐ To process events
- ☐ To manage an event stream

 Expand

 **Correct**

Correct! Brokers are synchronized and managed by another dedicated server called ZooKeeper.

20. What is Kafka Streams API?

1 / 1 point

- ☐ It facilitates stream processing by focusing on the input of the Steams API.
- ☐ It helps data engineers through multiple processing of records.
- ☐ It processes and analyzes data stored in Kafka libraries.
- ☒ It is a simple client library aiming to facilitate data processing in event-streaming pipelines.

 Expand

 **Correct**

The Streams API facilitates stream processing.