

✔ Congratulations! You passed!

Grade
received 80%

Latest Submission
Grade 80%

To pass 80% or
higher

Go to next item

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

1 / 1 point

- ☐ Processing data to make it conform to requirements.
- ☒ Configuring access to the data and reading it into an application.
- ☐ 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.

↗ Expand

✔ Correct

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

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

1 / 1 point

- ☐ Blt 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 curated data.
- ☒ There is no information loss with ELT because you are working with a replica of the data.
- ☐ With ELT, the data is acquired and prepared for subsequent use in an analytics environment.

↗ Expand

✔ Correct

Correct! Because you are working with a replica of the source data, there is no information loss.

3. Select the correct statement regarding ETL.

1 / 1 point

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

↗ Expand

✔ Correct

Transformations for ETL pipelines take place within the data pipeline before the data reaches its destination.

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. When is it appropriate to use micro-batch loading?

1 / 1 point

- ☒ When imminent processes need access to a small window of recent data.
- ☐ When a specified amount of data is accumulated by the data source.
- ☐ When the source data reaches a specified size.
- ☐ When an event is detected by the source system.

↗ Expand

Expand



Correct

Correct! Micro-batch loading is used when imminent processes need access to a small window of recent data.

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

1 / 1 point

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

Expand



Correct

ETL pipelines can be built from basic Bash scripts, and 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 average amount of time a packet spends at each stage in the pipeline
- ☐ The size of a packet
- ☐ The sum of the 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

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?

0 / 1 point

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

Expand



Incorrect

Review the Key Data Pipeline Processes video.

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

1 / 1 point

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

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

Expand



Incorrect

Review the Data Pipeline Tools and Technologies video.

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

1 / 1 point

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

[Expand](#)

✓ **Correct**

Apache Airflow pipelines are built on four main principles. They are scalable, dynamic, extensible, and lean.

12. Each task performed within your DAG is also written in _____?

1 / 1 point

- ☐ C++
- ☐ Kafka
- ☐ Java
- ☒ Python

[Expand](#)

✓ **Correct**

Just like the DAG itself, each task performed within your DAG is also written in Python.

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

[Expand](#)

✓ **Correct**

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

14. Which of these is part of the Apache Airflow DAG definition script?

1 / 1 point

- ☐ DAG nodes
- ☐ DAG pipeline or instantiation
- ☐ DAG argument imports
- ☒ Python library imports

[Expand](#)

✓ **Correct**

An Apache Airflow DAG is a Python script that consists of logical blocks, including Python library imports.

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

0 / 1 point

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

[Expand](#)

✗ **Incorrect**

Incorrect. Review the Airflow Logging and Monitoring video.

16. Select the correct statement regarding events.

1 / 1 point

- ☐ The event is a record of a task's execution.

- ☐ The temperature at which water freezes is an example of an event.
- ☒ 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.

[Expand](#)

☒ **Correct**
An event normally means something worth noticing is happening.

17. Which statement best describes Kafka brokers?

0 / 1 point

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

[Expand](#)

☒ **Incorrect**
Incorrect. Review the Apache Kafka Overview video.

18. Select the correct statement regarding the main features of Kafka.

1 / 1 point

- ☒ A Kafka cluster normally has multiple event brokers that can handle event streaming in parallel.
- ☐ Kafka is a full featured, commercial product that is highly reliable.
- ☐ Kafka stores events temporarily; as such, event consumption must be done by a deadline.
- ☐ Kafka is very fast, but not highly scalable.

[Expand](#)

☒ **Correct**
A Kafka cluster normally has multiple event brokers that can handle event streaming in parallel. As such, Kafka is very fast and highly scalable.

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

1 / 1 point

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

[Expand](#)

☒ **Correct**
One core component of Kafka is replications, that duplicate partitions into different brokers.

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

1 / 1 point

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

[Expand](#)

☒ **Correct**
Correct! The Kafka Streams API processes and analyzes data stored in Kafka topics, ensures that each record will only be processed once, and processes only one record at a time.