← Back Final Quiz Graded Quiz • 1h 40m
Final Quiz Graded Quiz • 1h 40m

Graded Quiz • 1h 40m	▲ Try again once you are ready Grade Latest Submission To pass 80% or Retake the assign	nment in 23h 40m
	received 75% Grade 75% higher	
	1. What is the ETL process?	0 / 1 point
	ETL is processing data to include cleaning, filtering, and formatting. ETL is curating data from multiple sources, conforming it, and loading it.	
	ETL is extracting data from web pages using an application such as Python. ETL is making data readily available for analytics, dashboards, and reports.	
	∠ [¬] Expand	
	Neview the ETL Fundamentals video.	
	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 moving data is usually more of a bottleneck than processing it.	
	When data is structured and relational rather than unstructured. When data is being live streamed from a single location.	
	∠ [™] Expand	
	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 happens in the destination environment at will. ETL is flexible, making data readily available for self-serve analytics.	
	∠ [¬] Expand	
	Correct Transformations for ETL pipelines take place within the data pipeline before the data reaches its destination.	
	4. What are two examples of raw data sources?	1/1 point
	Social media and artificial intelligence Analytics and human genomes data	
	Paper documents and weather station networks. Calculations and web pages	
	∠ [™] Expand	
	Correct Raw data sources include (among others) paper documents, weather station networks, web pages, social media, and human genomes data. Processed data is not raw data.	
	5. When is it appropriate to use micro-batch loading?	1 / 1 point
	When the source data reaches a specified size. When imminent processes need access to a small window of recent data.	
	When an event is detected by the source system. When a specified amount of data is accumulated by the data source.	
	∠ [¬] Expand	
	Correct Correct! Micro-batch loading is used when imminent processes need access to a small window of recent	
	data.	
	Out in used to expend live statistics to a log file. Out in used to expend live statistics to a log file.	1/1 point
	 It is used to append live statistics to a log file. It is used to schedule a workflow. It is used to turn your file into a Bash shell script. 	
	∠ Expand	
	Correct Correct! The shebang, "!#", is used to turn your file into a Bash shell script.	
	7. What is the definition of a data pipeline?	0 / 1 point
	 7. What is the definition of a data pipeline? Data pipelines are systems that specifically move or modify data. Data pipelines are any series of connected processes. 	
	Data pipelines are any series of connected processes. Data pipelines are units of data queued for ingestion. Data pipelines are processes in chains.	
	∠ Expand	
	Expand	
	8. Which of these data pipeline stages tends to be optional?	0 / 1 point
	Loading data into a destination facility Ingestion of data into the pipeline	
	Data transformation Extraction of the data from one or more sources	
	∠ [¬] Expand	
	Name of the New York of the Ne	
	9. Which of the following statements is true in regard to stream processing?	1/1 point
	 Stream processing typically operates on a fixed schedule. 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.	
	∠ [™] Expand	
	 ✓ Correct Correct! With streaming pipelines, records or events are processed immediately as they occur. 	
	10. Which of the following is a stream-processing technology?	1/1 point
	Apache SparkPanoply	
	O Pandas O Alteryx	
	∠ ⁷ Expand	
	Correct Correct! Stream-processing technologies include Apache Kafka, IBM Streams, SQLstream, and Apache Spark.	
	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	
	 Scalable, dynamic, extensible, lean Effective, simple to use, scalable, agile 	
	∠ [¬] Expand	
	Correct Apache Airflow pipelines are built on four main principles. They are scalable, dynamic, extensible, and lean.	
	12. Which of the following is an example of a directed acyclic graph (DAG)?	1/1 point
	∠ Expand	
	 ✓ Correct Correct! A directed acyclic graph has no loops, and each edge has a single specified direction. 	
	13. Which statement best describes the default "DAGs View" in the Apache Airflow UI?	1/1 point
	It's a table of quick links to drill down into more information related to each DAG. It's an interactive table displaying a thumbnail of each DAG in your environment. It's an interactive table containing data about each DAG in your environment.	
	 It's an interactive table containing data about each DAG in your environment. It's a static table containing each DAG's name, its run schedule, and a thumbnail of the DAG. 	
	∠ Z Expand	
	Correct Correct! The 'DAGs View,' is a table containing data about each DAG in your environment. Each row displays interactive information about a DAG in your environment, such as each DAG's name; its run schedule, in this case in the crontab format; and the DAG's owner.	
	14. In the Apache Airflow DAG, which code block contains the nodes of the DAG?	0 / 1 point
	14. In the Apache Airflow DAG, which code block contains the nodes of the DAG? DAG definition Task definitions	0 / 1 point
	Task definitions DAG argument specification Task pipeline	
	∠ [™] Expand	
	Incorrect Incorrect. Review the Build DAG Using Airflow video.	
	15. Which of the following is a tool that can be used to search, index, and analyze log files?	1/1 point
	○ IBM Cloud○ Prometheus	
	StatsD 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
	The temperature at which water freezes is an example of an event. An event is something that no one notices, but it is happening.	
	 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. Events describe an entity's observable state updates over time. 	
	∠ [¬] Expand	
	Correct An event normally means something worth noticing is happening.	
	17. Select the correct statement regarding Apache Kafka.	1/1 point
	 Kafka was originally used to track user activities such as mouse clicks, but is now suitable for other metric-streaming. 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.	
	∠ Expand	
	Correct Kafka is used for sensor readings, GPS, and hardware and software monitoring.	
	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 very fast, but not highly scalable. 	
	 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. 	
	∠ [¬] 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 Consumers, that receive, store, and process events. A core component is Servers, which are applications that publish events into topics.	
	A core component is Replications, that duplicate partitions into different brokers. 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 ensures that a record is only processed once. It publishes a received stream to a Kafka topic. It publishes events	
	It publishes events. It filters raw data.	
	∠ ^N 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.	