← Back Final Quiz Graded Quiz • 1h 40m	▲ Try again once you are ready		<b>Due</b> Sep 17, 11:59 Pt
	Grade Latest Submission To pass 80% or received 75% Grade 75% higher	Try again	
	1. What is meant by the term "data transformation"?  Writing data to some new destination environment  Processing data to make it conform to requirements.	1/1 point	
	<ul> <li>Processing data to make it conform to requirements.</li> <li>Configuring access to the data and reading it into an application</li> <li>Making data readily available for ingestion by analytics applications so that end users can gain value from it.</li> </ul>		
	<ul> <li>∠ Expand</li> <li>✓ Correct</li> <li>Correct! Data transformation, also known as data wrangling, means processing data to make it conform to</li> </ul>		
	Correct! Data transformation, also known as data wrangling, means processing data to make it conform to the requirements of both the target system and the intended use case for the curated data.		
	When data is structured and relational rather than unstructured.  When a very large amount of information is either already recorded or being generated, but is not yet captured, or accessible.	0 / 1 point	
	When moving data is usually more of a bottleneck than processing it.  When data is being live streamed from a single location.		
	Expand		
	3. Which of the following best describes a driving factor of the evolution from ETL to ELT?	0 / 1 point	
	<ul> <li>The need to decouple transformations from the data pipeline.</li> <li>The need to transfer data from a legacy system to an updated one.</li> <li>The demand to release raw data to a wider user base.</li> <li>The need to build a single dashboard from multiple data sources.</li> </ul>		
	Expand		
	Ncorrect Incorrect. Review the Comparing ETL to ELT video.		
	Which best describes optical character recognition (OCR)?      Capturing and digitizing opinions, questionnaires, and vital statistical data obtained through polling and census methods.      Creating through web pages in search of text, images, tables, and hyperlinks.	1/1 point	
	<ul> <li>Crawling through web pages in search of text, images, tables, and hyperlinks.</li> <li>Interpreting and digitizing text scanned from paper documents so it can be stored in a computer readable format.</li> <li>Digitizing analog audio recordings and signals into a computer readable format.</li> </ul>		
	∠ Expand  ✓ Correct		
	Correct! OCR takes scanned and paper documents and translates them into computer readable text.  5. What is stream loading?	0 / 1 point	
	<ul> <li>Stream loading refers to loading an initial history into a database.</li> <li>Stream loading refers to loading data in real time as it becomes available.</li> <li>Stream loading refers to loading data on a schedule.</li> </ul>		
	Stream loading refers to loading data in chunks defined by some time windows of data accumulated by the data source.		
	Expand    Note		
	6. What is the purpose of the Bash shebang?	1/1 point	
	It is used to display live statistics.  It is used to turn your file into a Bash shell script.  It is used to schedule a workflow.  It is used to append live statistics to a log file.		
	∠ <sup>™</sup> Expand		
	Correct! The shebang, "!#", is used to turn your file into a Bash shell script.		
	7. Which of the following best describes latency?  The average amount of time a packet spends at each stage in the pipeline  The sum of the times a packet spends at each stage in the pipeline	1/1 point	
	The amount of data that can be fed through the pipeline per unit of time  The size of a packet		
	<ul> <li>✓ Correct         Correct! Latency is the total time it takes for a single packet of data to pass through the pipeline     </li> </ul>		
	Correct! Latency is the total time it takes for a single packet of data to pass through the pipeline  8. Which of these fall into the category of data pipeline monitoring?	0 / 1 point	
	<ul> <li>Extraction and ingestion</li> <li>Scheduling and maintenance</li> <li>Loading and scheduling</li> </ul>		
	Catency and throughput ∠ <sup>N</sup> Expand		
	Incorrect Review the Key Data Pipeline Processes video.		
	9. When is stream processing used instead of batch processing?  When accuracy is more critical than immediate processing	1/1 point	
	<ul> <li>When results are required with minimal latency, essentially in real time</li> <li>When processing is triggered by the amount of data reaching a certain size</li> <li>When processing must be done on a fixed schedule, ranging from hours to weeks apart</li> </ul>		
	∠ Expand  ✓ Correct		
	Stream processing is designed for ingesting information such as credit card transactions, that need to be processed immediately as they occur.		
	10. Which of the following is popular and versatile programming environment for building data pipelines?  Data Frame  Talend	1/1 point	
	AWS Glue  Pandas		
	<ul> <li>✓ Expand</li> <li>✓ Correct</li> <li>Correct! Pandas is a very popular and highly versatile programming environment for building data</li> </ul>		
	pipelines.  11. Which of the following best describes Apache Airflow?	1/1 point	
	<ul> <li>It is primarily a directed acyclic graph (DAG).</li> <li>It is primarily a data streaming solution.</li> <li>It is primarily a workflow manager.</li> </ul>		
	○ It is primarily a data pipeline monitoring tool. ∠ 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. Each task performed within your DAG is also written in?	1/1 point	
	Java C++ Kafka		
	∠ <sup>™</sup> Expand		
	✓ Correct  Just like the DAG itself, each task performed within your DAG is also written in Python.		
	DAGs are always displayed as trees.  The interface flags graphs that contain loops in red.	1/1 point	
	Clicking the task duration button allows you to see a graph representing the latency of packets in the pipeline.  Clicking on the Code button allows you to see the Python source code that is used to create a DAG in your environment.		
	<ul> <li>✓ Correct</li> <li>Correct! By clicking the Code button, you can also view the complete Python source code that defines</li> </ul>		
	Correct! By clicking the Code button, you can also view the complete Python source code that defines your DAG.  14. In the Apache Airflow DAG, which code block contains the nodes of the DAG?	0 / 1 point	
	Task pipeline      DAG argument specification	, ~ point	
	Task definitions  DAG definition		
	Expand     Incorrect Incorrect. Review the Build DAG Using Airflow video.		
	<ul><li>15. Which of the following is a tool that can be used to search, index, and analyze log files?</li><li>Splunk</li></ul>	1/1 point	
	<ul><li>Splunk</li><li>IBM Cloud</li><li>StatsD</li><li>Prometheus</li></ul>		
	∠ <sup>™</sup> 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.   Events describe a clearly observable entity, such as a car or a hospital patient.	1/1 point	
	<ul> <li>An event is something that no one notices, but it is happening.</li> <li>Events describe an entity's observable state updates over time.</li> <li>The temperature at which water freezes is an example of an event.</li> </ul>		
	∠ Expand ○ Correct		
	An event normally means something worth noticing is happening.  17. In which scenario is it most appropriate to use Kafka?	1/1 point	
	When an enterprise needs high throughput and reliable data transportation services among event sources and destinations.  When an enterprise needs an on-demand ESP-as-a-service.  To log and monitor the status of tasks in DAG runs, and to diagnose and debug issues.		
	To build a streaming data pipeline.		
	Correct Correct! You basically can use Kafka in scenarios when you want high throughput and reliable data transportation services among various event sources and destinations.		
	transportation services among various event sources and destinations.  18. Select the correct statement regarding the main features of Kafka.	1/1 point	
	<ul> <li>Kafka is very fast, but not highly scalable.</li> <li>A Kafka cluster normally has multiple event brokers that can handle event streaming in parallel.</li> <li>Kafka is a full featured, commercial product that is highly reliable.</li> </ul>		
	Kafka stores events temporarily; as such, event consumption must be done by a deadline.		
	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	
	<ul> <li>A core component is Replications, that duplicate partitions into different brokers.</li> <li>A core component is Producers, that divide topics into different brokers.</li> <li>A core component is Servers, which are applications that publish events into topics.</li> </ul>		
	<ul> <li>A core component is Consumers, that receive, store, and process events.</li> <li>∠ Expand</li> </ul>		
	Correct One core component of Kafka is replications, that duplicate partitions into different brokers.		
	20. What is Kafka Streams API?  It helps data engineers through multiple processing of records.	1/1 point	
	<ul> <li>It is a simple client library aiming to facilitate data processing in event-streaming pipelines.</li> <li>It processes and analyzes data stored in Kafka libraries.</li> <li>It facilitates stream processing by focusing on the input of the Steams API.</li> </ul>		
	∠ Expand  ✓ Correct		
	Correct The Streams API facilitates stream processing.		