Final Quiz | Coursera

6/23/23, 8:52 PM Final Assignment Final Quiz Quiz: Final Quiz
20 questions

Submit your assignment

 \otimes Receive grade

To Pass 80% or higher

Due Jul 16, 11:59 PM EDT Attempts 3 every 24 hours

🖒 Like 📿 Dislike 🏳 Report an issue

Try again once you are ready Grade Latest Submission received 65% Grade 65% To pass 80% or

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

Making data readily available for ingestion by analytics applications so that end users can gain value from it.

Processing data to make it conform to requirements.

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

Blt is easier to fix errors or spot missing values with ELT.

3. Select the correct statement regarding ETL.

5. What is stream loading?

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

cron job for your Bash script.

 By building a basic Bash script By using the crontab editor By using Python to build the shell

Writing data to some new destination environment Onfiguring access to the data and reading it into an application

1/1 point

0 / 1 point

Your grade View Feedback 65% We keep your highest score

Due Jul 16, 11:59 PM EDT

 There is no information loss with ELT because you are working with a replica of the data. 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. With ELT, the data is acquired and prepared for subsequent use in an analytics environment. Correct
 Correct! Because you are working with a replica of the source data, there is no information loss.

ETL happens in the destination environment at will. 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. 4. Which of the following best describes the function of analog-to-digital converters? 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 compute Capturing and digitizing opinions, questionnaires, and vital statistical data obtained through poll Crawling through web pages in search of text, images, tables, and hyperlinks.

 ⊙ Correct
 Correct! Analog digital converters digitize analog audio recordings and signals. Stream loading refers to loading data in chunks defined by some time windows of data accumul 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. Correct
 Correct! Stream loading is loading data in real time as it becomes available.

 By using an API Correct
 ETL pipelines can be built from basic Bash scripts, and an ETL job can be scheduled to run by creating a 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

 Correct
 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 Extraction and ingestion Loading and scheduling Scheduling and maintenance Incorrect
 Review the Key Data Pipeline Processes video. 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 amount of data reaching a certain size

 Correct
 Stream processing is designed for ingesting information such as credit card transactions, that need to be 10. Which of the following is popular and versatile programming environment for building data pipelines? Oata Frame Pandas Talend AWS Glue

When results are required with minimal latency, essentially in real time

11. Which of the following are the four principles Apache Airflow is built upon?" Robust, scalable, effective, dynamic Scalable, dynamic, extensible, lean Effective, simple to use, scalable, agile Sustainable, competitive, agile, simple to use Expand

 Correct
 Apache Airflow pipelines are built on four main principles. They are scalable, dynamic, extensible, and
 12. Which statement is true about directed acyclic graphs (DAGs)? O 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. ⊗ Incorrect
Incorrect. Review the Advantages of Using Data Pipelines as DAGs in Apache Airflow video.

13. Which of these are ways to visualize DAGs in Apache Airflow? O DAGs can be visualized in Tree View and Dashboard View O DAGS can be visualized as a Static 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.

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

14. In the Apache Airflow DAG, which code block contains the nodes of the DAG? 0 / 1 point OAG argument specification Task definitions

 ⊗ Incorrect
 Incorrect, Review the Build DAG Using Airflow video. 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.

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 rea Event steaming is the continual transportation between an event source and an event destination

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 was originally used to track user activities such as mouse clicks, but is now suitable for or ∠⁷ Expand

Correct
 Kafka is used for sensor readings, GPS, and hardware and software monitoring.

⊗ Incorrect
 You didn't select all the correct answers

18. Select two common components of an event streaming platform (ESP). Event broker, event storage, Analytics and Query Engine Event transportation This should not be selected Incorrect. Review the Distributed Event Streaming Platform Components video. Correct! The second common component of an ESP is Event Storage, which is used for stori event sources. ∠⁷ 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 Replications, that duplicate partitions into different brokers. A core component is Servers, which are applications that publish events into topics. 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.

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.

https://www.coursera.org/learn/etl-and-data-pipelines-shell-airflow-kafka/exam/OphJ8/final-quiz/attempt?redirectToCover=true for the properties of the pro