IT认证电子书

质　量　更　高　　服　务　更　好

半年免费升级服务

http://www.itrenzheng.com

Exam : **Databricks Certified Data Engineer Associate**

Title : Databricks Certified Data Engineer Associate Exam

Version : DEMO

1.A data organization leader is upset about the data analysis team's reports being different from the data engineering team's reports. The leader believes the siloed nature of their organization's data engineering and data analysis architectures is to blame.
Which of the following describes how a data lakehouse could alleviate this issue?
A. Both teams would autoscale their work as data size evolves
B. Both teams would use the same source of truth for their work
C. Both teams would reorganize to report to the same department
D. Both teams would be able to collaborate on projects in real-time
E. Both teams would respond more quickly to ad-hoc requests
Answer: B
Explanation:
A data lakehouse is a data management architecture that combines the flexibility, cost-efficiency, and scale of data lakes with the data management and ACID transactions of data warehouses, enabling business intelligence (BI) and machine learning (ML) on all data12. By using a data lakehouse, both the data analysis and data engineering teams can access the same data sources and formats, ensuring data consistency and quality across their reports. A data lakehouse also supports schema enforcement and evolution, data validation, and time travel to old table versions, which can help resolve data conflicts and errors1.
Reference: 1: What is a Data Lakehouse? - Databricks 2: What is a data lakehouse? | IBM

2.Which of the following describes a scenario in which a data team will want to utilize cluster pools?
A. An automated report needs to be refreshed as quickly as possible.
B. An automated report needs to be made reproducible.
C. An automated report needs to be tested to identify errors.
D. An automated report needs to be version-controlled across multiple collaborators.
E. An automated report needs to be runnable by all stakeholders.
Answer: A
Explanation:
Databricks cluster pools are a set of idle, ready-to-use instances that can reduce cluster start and auto-scaling times. This is useful for scenarios where a data team needs to run an automated report as quickly as possible, without waiting for the cluster to launch or scale up. Cluster pools can also help save costs by reusing idle instances across different clusters and avoiding DBU charges for idle instances in the pool.
Reference: Best practices: pools | Databricks on AWS, Best practices: pools - Azure Databricks | Microsoft Learn, Best practices: pools | Databricks on Google Cloud

3.Which of the following is hosted completely in the control plane of the classic Databricks architecture?
A. Worker node
B. JDBC data source
C. Databricks web application
D. Databricks Filesystem
E. Driver node
Answer: C
Explanation:

The Databricks web application is the user interface that allows you to create and manage workspaces, clusters, notebooks, jobs, and other resources. It is hosted completely in the control plane of the classic Databricks architecture, which includes the backend services that Databricks manages in your Databricks account. The other options are part of the compute plane, which is where your data is processed by compute resources such as clusters. The compute plane is in your own cloud account and network.
Reference: Databricks architecture overview, Security and Trust Center

4.Which of the following benefits of using the Databricks Lakehouse Platform is provided by Delta Lake?
A. The ability to manipulate the same data using a variety of languages
B. The ability to collaborate in real time on a single notebook
C. The ability to set up alerts for query failures
D. The ability to support batch and streaming workloads
E. The ability to distribute complex data operations
Answer: D
Explanation:
Delta Lake is the optimized storage layer that provides the foundation for storing data and tables in the Databricks lakehouse. Delta Lake is fully compatible with Apache Spark APIs, and was developed for tight integration with Structured Streaming, allowing you to easily use a single copy of data for both batch and streaming operations and providing incremental processing at scale1. Delta Lake supports upserts using the merge operation, which enables you to efficiently update existing data or insert new data into your Delta tables2. Delta Lake also provides time travel capabilities, which allow you to query previous versions of your data or roll back to a specific point in time3.
Reference:
1: What is Delta Lake? | Databricks on AWS
2: Upsert into a table using merge |
Databricks on AWS
3: [Query an older snapshot of a table (time travel) | Databricks on AWS]
Learn more
1learn.microsoft.com2medium.com3slideshare.net4docs.databricks.com5github.com6key2consulting.com

5.Which of the following describes the storage organization of a Delta table?
A. Delta tables are stored in a single file that contains data, history, metadata, and other attributes.
B. Delta tables store their data in a single file and all metadata in a collection of files in a separate location.
C. Delta tables are stored in a collection of files that contain data, history, metadata, and other attributes.
D. Delta tables are stored in a collection of files that contain only the data stored within the table.
E. Delta tables are stored in a single file that contains only the data stored within the table.
Answer: C
Explanation:
Delta Lake is the optimized storage layer that provides the foundation for storing data and tables in the Databricks lakehouse. Delta Lake is open source software that extends Parquet data files with a file-based transaction log for ACID transactions and scalable metadata handling1. Delta Lake stores its data

and metadata in a collection of files in a directory on a cloud storage system, such as AWS S3 or Azure Data Lake Storage2. Each Delta table has a transaction log that records the history of operations performed on the table, such as insert, update, delete, merge, etc. The transaction log also stores the schema and partitioning information of the table2. The transaction log enables Delta Lake to provide ACID guarantees, time travel, schema enforcement, and other features1.
Reference: What is Delta Lake? | Databricks on AWS
Quickstart — Delta Lake Documentation