

Certified-Data-Engineer-Professional題庫更新， Certified-Data-Engineer-Professional題庫資訊



有人問，成功在哪里？我告訴你，成功就在KaoGuTi。選擇KaoGuTi就是選擇成功。KaoGuTi Databricks的Certified-Data-Engineer-Professional考試培訓資料是幫助所有IT認證的考生通過認證的，它針對Databricks的Certified-Data-Engineer-Professional考試認證的，經過眾多考生反映，KaoGuTi Databricks的Certified-Data-Engineer-Professional考試培訓資料在考生中得到了很大的反響，建立了很好的口碑，說明選擇KaoGuTi Databricks的Certified-Data-Engineer-Professional考試培訓資料就是選擇成功。

Databricks Certified-Data-Engineer-Professional Exam Syllabus Topics:

Section	Objectives
Topic 1: Data Ingestion & Acquisition	- Design and implement data ingestion pipelines • 1. Ingest formats including Delta Lake, Parquet, ORC, AVRO, JSON, CSV, XML, text, and binary data from sources such as message buses and cloud storage • 2. Create an append-only data pipeline capable of handling both batch and streaming data using Delta
Topic 2: Data Transformation, Cleansing, and Quality	- Transform and validate data • 1. Write efficient Spark SQL and PySpark code for advanced transformations including window functions, joins, and aggregations • 2. Develop a quarantining process for bad data with Lakeflow Declarative Pipelines or Auto Loader in classic jobs
Topic 3: Debugging and Deploying	- Deploying CI/CD • 1. Configure and integrate Git-based CI/CD workflows using Databricks Git folders for notebook and code deployment • 2. Build and deploy Databricks resources using Databricks Asset Bundles - Debugging and Troubleshooting • 1. Analyze errors and remediate failed job runs using job repairs and parameter overrides • 2. Use Lakeflow Declarative Pipelines event logs and Spark UI to debug Lakeflow Declarative Pipelines and Spark pipelines • 3. Identify diagnostic information using Spark UI, cluster logs, system tables, and query profiles to troubleshoot errors

Topic 4: Ensuring Data Security and Compliance	- Ensuring Compliance • 1. Implement compliant batch and streaming pipelines that detect and mask PII • 2. Develop data purging solutions that comply with data retention policies - Applying Data Security Mechanisms • 1. Use row filters and column masks to protect sensitive table data • 2. Apply anonymization and pseudonymization methods including hashing, tokenization, suppression, and generalization • 3. Use ACLs to secure workspace objects and enforce the principle of least privilege
Topic 5: Data Modeling	- Design and optimize data models • 1. Design and implement scalable data models using Delta Lake to manage large datasets • 2. Identify the benefits of liquid clustering over partitioning and Z-Ordering • 3. Simplify data layout decisions and optimize query performance using liquid clustering • 4. Design dimensional models for analytical workloads with efficient querying and aggregation
Topic 6: Cost & Performance Optimization	- Optimize cost and performance • 1. Understand Databricks query optimization techniques for large datasets, including data skipping and file pruning • 2. Understand Delta optimization techniques such as deletion vectors and liquid clustering • 3. Use query profiling to identify bottlenecks such as inefficient joins and data shuffling • 4. Apply Change Data Feed to address streaming table limitations and improve latency • 5. Understand how and why Unity Catalog managed tables reduce operational overhead and maintenance burden
Topic 7: Data Governance	- Govern enterprise data • 1. Demonstrate understanding of the Unity Catalog permission inheritance model • 2. Create and add descriptions and metadata to enterprise data to improve discoverability
Topic 8: Monitoring and Alerting	- Alerting • 1. Use SQL Alerts to monitor data quality • 2. Use the Workflows UI and Jobs API to configure notifications for job status and performance issues - Monitoring • 1. Use Databricks REST APIs and Databricks CLI to monitor jobs and pipelines • 2. Use Lakeflow Declarative Pipelines event logs to monitor pipelines • 3. Use system tables for observability of resource utilization, cost, auditing, and workloads • 4. Use Query Profile and Spark UI to monitor workloads
Topic 9: Data Sharing and Federation	- Share and federate data • 1. Demonstrate secure Delta Sharing between Databricks deployments using Databricks-to-Databricks sharing or with external platforms using the open sharing protocol • 2. Configure Lakehouse Federation with appropriate governance across supported source systems • 3. Use Delta Sharing to share live data from the Lakehouse with any computing platform

Topic 10: Developing Code for Data Processing using Python and SQL	- Using Python and Tools for Development 1. Design and implement a scalable Python project structure optimized for Databricks Asset Bundles, enabling modular development, deployment automation, and CI/CD integration 2. Develop User-Defined Functions using Pandas/Python UDF 3. Manage and troubleshoot external third-party library installations and dependencies, including PyPI packages, local wheels, and source archives - Building and Testing an ETL Pipeline with Lakeflow Declarative Pipelines, SQL, and Apache Spark 1. Compare Spark Structured Streaming and Lakeflow Declarative Pipelines to determine the optimal approach for scalable ETL pipelines 2. Create pipeline components using control flow operators such as if/else and foreach 3. Use APPLY CHANGES APIs to simplify CDC in Lakeflow Declarative Pipelines 4. Explain the advantages and disadvantages of streaming tables compared to materialized views 5. Build and manage reliable, production-ready batch and streaming data pipelines using Lakeflow Declarative Pipelines and Auto Loader 6. Develop unit and integration tests using assertEquals, assertSchemaEqual, DataFrame.transform, testing frameworks, and debugging tools 7. Choose appropriate configurations for environments, dependencies, high-memory notebook tasks, and retry behavior 8. Create and automate ETL workloads using Jobs through the UI, APIs, or CLI
---	--

>>> Certified-Data-Engineer-Professional 題庫更新 <<<

Databricks Certified-Data-Engineer-Professional 考試題庫，真正的PDF格式的考試問題和答案，為妳準備的Certified-Data-Engineer-Professional 測試

我們KaoGuTi Databricks的Certified-Data-Engineer-Professional考試認證資料是全球所有網站不能夠媲美的，當然這不僅僅是品質的問題，我們的品質肯定是沒得說，更重要的是我們KaoGuTi Databricks的Certified-Data-Engineer-Professional考試認證資料適合所有的IT考試認證，它的使用性達到各個IT領域，所以我們KaoGuTi網站得到很多考生的關注，他們相信我們，依賴我們，這也是我們KaoGuTi網站所擁有的實力所體現之處，我們的考試培訓資料能讓你買了之後不得不向你的朋友推薦，並讚不絕口，因為它真的對你們有很大的幫助。

最新的 Databricks Certification Certified-Data-Engineer-Professional 免費考試真題 (Q213-Q218):

問題 #213

A platform engineer needs to report the resource consumption, categorized by SKU tier, across all workspaces. The engineer decides to use the `system.billing.usage` system table to create a query. Which SQL query will accurately return the daily usage by product?

- A.
- ```
SELECT
 DATE_TRUNC('day', usage_start_time) AS day,
 sku_name,
 SUM(dbus) AS daily_dbus
FROM
 system.billing.usage
WHERE
 sku_name IS NOT NULL
GROUP BY
 day, sku_name;
```

- B.

```
SELECT
 date_trunc('day', usage_start_time) AS usage_day,
 sku_name,
 SUM(usage_quantity) AS total_daily_dbus
FROM
 system.billing.usage
WHERE
 usage_quantity IS NOT NULL
GROUP BY
 usage_day, sku_name
ORDER BY
 usage_day;
```

```
SELECT
 TO_DATE(usage_start_time) AS usage_day,
 sku_name,
 SUM(usage_quantity) AS total_dbus
FROM
 system.billing.usage
GROUP BY
 usage_day
ORDER BY
 usage_day
```

- C.

```
SELECT
 CAST(usage_start_time AS DATE) AS usage_day,
 sku_name,
 COUNT (usage_quantity) AS sku_count
FROM
 system.billing.usage
GROUP BY
 usage_day, sku_name
ORDER BY
 usage_day;
```

- D.

答案：B

解題說明：

This query correctly aggregates usage at a daily granularity by truncating the usage start timestamp to the day and summing the usage quantity, which represents DBUs. Grouping by both the derived daily value and the SKU name ensures usage is accurately categorized by product tier across all workspaces.

#### 問題 #214

A data engineer wants to create a cluster using the Databricks CLI for a big ETL pipeline. The cluster should have five workers, one driver of type i3.xlarge, and should use the '14.3.x- scala2.12' runtime. Which command should the data engineer use?

- A. databricks compute create 14.3.x-scala2.12 --num-workers 5 --node-type-id i3.xlarge --cluster-name Data Engineer\_cluster
- B. databricks compute add 14.3.x-scala2.12 --num-workers 5 --node-type-id i3.xlarge --cluster-name Data Engineer\_cluster
- C. databricks clusters create 14.3.x-scala2.12 --num-workers 5 --node-type-id i3.xlarge --cluster-name DataEngineer\_cluster
- D. databricks clusters add 14.3.x-scala2.12 --num-workers 5 --node-type-id i3.xlarge --cluster-name Data Engineer\_cluster

答案：C

解題說明：

The correct Databricks CLI command to create a new cluster is `databricks clusters create`. You specify the runtime with `--spark-version` (here '14.3.x-scala2.12'), the number of workers with `--num-workers`, the node type with `--node-type-id`, and the cluster name with `--cluster-name`. This command properly initializes the cluster with the desired configuration.

#### 問題 #215

The data science team has requested assistance in accelerating queries on free form text from user reviews. The data is currently stored in Parquet with the below schema:

`item_id INT, user_id INT, review_id INT, rating FLOAT, review STRING`

The review column contains the full text of the review left by the user. Specifically, the data science team is looking to identify if any of 30 key words exist in this field.

A junior data engineer suggests converting this data to Delta Lake will improve query performance.

Which response to the junior data engineer's suggestion is correct?

- A. Text data cannot be stored with Delta Lake.
- **B. Delta Lake statistics are not optimized for free text fields with high cardinality.**
- C. Delta Lake statistics are only collected on the first 4 columns in a table.
- D. ZORDER ON review will need to be run to see performance gains.
- E. The Delta Lake creates a term matrix for free text fields to support selective filtering.

答案： B

解題說明：

Converting the data to Delta Lake may not improve query performance on free text fields with high cardinality, such as the review column. This is because Delta Lake collects statistics on the minimum and maximum values of each column, which are not very useful for filtering or skipping data on free text fields. Moreover, Delta Lake collects statistics on the first 32 columns by default, which may not include the review column if the table has more columns. Therefore, the junior data engineer's suggestion is not correct. A better approach would be to use a full-text search engine, such as Elasticsearch, to index and query the review column. Alternatively, you can use natural language processing techniques, such as tokenization, stemming, and lemmatization, to preprocess the review column and create a new column with normalized terms that can be used for filtering or skipping data.

#### 問題 #216

A data team is automating a daily multi-task ETL pipeline in Databricks. The pipeline includes a notebook for ingesting raw data, a Python wheel task for data transformation, and a SQL query to update aggregates. They want to trigger the pipeline programmatically and see previous runs in the GUI. They need to ensure tasks are retried on failure and stakeholders are notified by email if any task fails. Which two approaches will meet these requirements? (Choose two.)

- **A. Trigger the job programmatically using the Databricks Jobs REST API (/jobs/run-now), the CLI (databricks jobs run-now), or one of the Databricks SDKs.**
- B. Create a single orchestrator notebook that calls each step with `dbutils.notebook.run()`, defining a job for that notebook and configuring retries and notifications at the notebook level.
- **C. Create a multi-task job using the UI, Databricks Asset Bundles (DABs), or the Jobs REST API (/jobs/create) with notebook, Python wheel, and SQL tasks. Configure task-level retries and email notifications in the job definition.**
- D. Use the REST API endpoint `/jobs/runs/submit` to trigger each task individually as separate job runs and implement retries using custom logic in the orchestrator.
- E. Use Databricks Asset Bundles (DABs) to deploy the workflow, then trigger individual tasks directly by referencing each task's notebook or script path in the workspace.

答案： A,C

解題說明：

Databricks Jobs supports defining multi-task workflows that include notebooks, SQL statements, and Python wheel tasks. These can be configured with retry policies, dependency chains, and failure notifications. The correct practice, as stated in the documentation, is to use the Jobs REST API (`/jobs/create`) or Databricks Asset Bundles to define multi-task jobs, and then trigger them programmatically using `/jobs/run-now`, CLI, or SDK. This allows the team to maintain full job history, handle retries automatically, and receive alerts via configured email notifications. Using `/jobs/runs/submit` creates one-off ad hoc runs without maintaining dependency visibility. Therefore, options B and C together satisfy the operational, automation, and governance requirements.

### 問題 #217

A transactions table has been liquid clustered on the columns product\_id, user\_id, and event\_date. Which operation lacks support for cluster on write?

- A. spark.write.format('delta').mode('append')
- B. INSERT INTO operations
- C. CTAS and RTAS statements
- D. spark.writestream.format('delta').mode('append')

答案：D

解題說明：

Delta Lake's Liquid Clustering is an advanced feature that improves query performance by dynamically clustering data without requiring costly compaction steps like traditional Z-ordering.

When performing writes to a Liquid Clustered table, some write operations automatically maintain clustering, while others do not.

### 問題 #218

.....

如果你想參加Certified-Data-Engineer-Professional認證考試，那麼是使用Certified-Data-Engineer-Professional考試資料是很有必要的。如果你正在漫無目的地到處尋找參考資料，那麼趕快停止吧。如果你不知道應該用什麼資料，那麼試一下KaoGuTi的Certified-Data-Engineer-Professional考古題吧。這個考古題的命中率很高，可以保證你一次就取得成功。與別的考試資料相比，這個考古題更能準確地劃出考試試題的範圍。這樣的話，可以讓你提高學習效率，更加充分地準備Certified-Data-Engineer-Professional考試。

**Certified-Data-Engineer-Professional題庫資訊：** [https://www.kaoguti.com/Certified-Data-Engineer-Professional\\_exam-pdf.html](https://www.kaoguti.com/Certified-Data-Engineer-Professional_exam-pdf.html)

- 最有效的Certified-Data-Engineer-Professional題庫更新，由Databricks權威專家撰寫 □ 到「tw.fast2test.com」搜索 □ Certified-Data-Engineer-Professional □ 輕鬆取得免費下載Certified-Data-Engineer-Professional PDF題庫
- Databricks Certified-Data-Engineer-Professional認證考古題 □ 來自網站 ➡ [www.newdumpsdf.com](http://www.newdumpsdf.com) □ 打開並搜索 ➡ Certified-Data-Engineer-Professional □ 免費下載Certified-Data-Engineer-Professional熱門考題
- Certified-Data-Engineer-Professional軟件版 □ 最新Certified-Data-Engineer-Professional考古題 □ Certified-Data-Engineer-Professional新版題庫上線 □ > [www.kaoguti.com](http://www.kaoguti.com) ◀ 上的 □ Certified-Data-Engineer-Professional □ 免費下載 只需搜尋最新Certified-Data-Engineer-Professional考古題
- Certified-Data-Engineer-Professional考題 □ Certified-Data-Engineer-Professional PDF題庫 □ Certified-Data-Engineer-Professional題庫更新資訊 □ 來自網站 《 [www.newdumpsdf.com](http://www.newdumpsdf.com) 》 打開並搜索 □ Certified-Data-Engineer-Professional □ 免費下載Certified-Data-Engineer-Professional考試重點
- Certified-Data-Engineer-Professional信息資訊 □ Certified-Data-Engineer-Professional考古題更新 □ Certified-Data-Engineer-Professional熱門考古題 □ □ [www.newdumpsdf.com](http://www.newdumpsdf.com) □ 網站搜索 { Certified-Data-Engineer-Professional } 並免費下載Certified-Data-Engineer-Professional認證題庫
- Certified-Data-Engineer-Professional軟件版 □ Certified-Data-Engineer-Professional軟件版 □ Certified-Data-Engineer-Professional題庫資訊 □ 複製網址 《 [www.newdumpsdf.com](http://www.newdumpsdf.com) 》 打開並搜索 > Certified-Data-Engineer-Professional □ 免費下載Certified-Data-Engineer-Professional PDF題庫
- 高效的Certified-Data-Engineer-Professional題庫更新和認證考試的領導者材料和權威的Certified-Data-Engineer-Professional題庫資訊 □ ▶ [www.newdumpsdf.com](http://www.newdumpsdf.com) ◀ 上的免費下載 ➡ Certified-Data-Engineer-Professional □ 頁面立即打開Certified-Data-Engineer-Professional認證題庫
- Certified-Data-Engineer-Professional考題 □ Certified-Data-Engineer-Professional熱門題庫 □ Certified-Data-Engineer-Professional熱門題庫 □ 免費下載“Certified-Data-Engineer-Professional”只需在「[www.newdumpsdf.com](http://www.newdumpsdf.com)」上搜索Certified-Data-Engineer-Professional熱門認證
- 可靠的Certified-Data-Engineer-Professional題庫更新擁有模擬真實考試環境與場境的軟件VCE版本 & 可依賴的Certified-Data-Engineer-Professional題庫資訊 □ 到 ( [www.pdfexamdumps.com](http://www.pdfexamdumps.com) ) 搜索 《 Certified-Data-Engineer-Professional 》 輕鬆取得免費下載Certified-Data-Engineer-Professional PDF題庫
- Certified-Data-Engineer-Professional熱門題庫 □ Certified-Data-Engineer-Professional熱門考古題 □ Certified-Data-Engineer-Professional考題 □ ➡ [www.newdumpsdf.com](http://www.newdumpsdf.com) □ 上的免費下載 《 Certified-Data-Engineer-Professional 》 頁面立即打開Certified-Data-Engineer-Professional題庫資訊
- 根據最新的考試大綱更新得到的Certified-Data-Engineer-Professional考古題 - 是最完整的Certified-Data-Engineer-Professional - Databricks Certified Data Engineer Professional題庫資料 □ [ [tw.fast2test.com](http://tw.fast2test.com) ] 上的免費下載 ⇒ Certified-Data-Engineer-Professional ◀ 頁面立即打開Certified-Data-Engineer-Professional認證考試解析

- [illegible]