

# Databricks Databricks-Certified-Professional-Data-Engineer퍼펙트최신버전공부자료, Databricks-Certified-Professional-Data-Engineer시험대비덤프최신 데모



참고: Pass4Test에서 Google Drive로 공유하는 무료 2025 Databricks Databricks-Certified-Professional-Data-Engineer 시험 문제집이 있습니다: <https://drive.google.com/open?id=1SgXcieROQBhvuqNNdp8-xNTXK8aGMPZu>

우리 Pass4Test에서는 여러분을 위하여 정확하고 우수한 서비스를 제공하였습니다. 여러분의 고민도 덜어드릴 수 있습니다. 빨리 성공하고 빨리Databricks Databricks-Certified-Professional-Data-Engineer인증시험을 패스하고 싶으시다면 우리 Pass4Test를 장바구니에 넣으시죠 . Pass4Test는 여러분의 아주 좋은 합습가이드가 될것입니다. Pass4Test로 여러분은 같고 싶은 인증서를 빠른시일내에 얻게될것입니다.

Databricks Certified Professional Data Engineer 시험은 Databricks를 사용하여 데이터 엔지니어링 솔루션을 구축, 배포 및 유지 관리하는 데이터 전문가의 전문성을 증명하고자하는 자격증 프로그램입니다. 이 시험은 데이터 엔지니어링과 관련된 다양한 주제를 다루며, Databricks 데이터 엔지니어링 개념과 기술을 철저히 이해해야 합니다. 시험은 어렵고, 후보자는 Databricks를 사용하여 특정 작업을 수행할 수 있는 능력을 증명해야 합니다.

Databricks Certified Professional Data Engineer 시험을 통과하는 것은 Databricks 플랫폼과 함께 일하는 데이터 전문가에게 유용한 성과입니다. 인증은 응시자가 데이터를 사용하여 데이터 엔지니어링 작업을 효과적으로 수행하는 데 필요한 기술과 지식을 가지고 있음을 보여줍니다. 또한 고용주가 데이터 엔지니어링 인증을받은 후보자를 점점 더 많이 찾고 있기 때문에 구직 시장에서 경쟁 우위를 제공합니다.

Databricks Certified Professional Data Engineer Exam은 데이터 엔지니어링 작업에 Databricks를 사용하는 데이터 전문가의 지식과 기술을 테스트하도록 설계되었습니다. 시험은 데이터 수집, 데이터 변환, 데이터 저장 및 데이터 분석을 포함한 다양한 주제를 다룹니다. 시험은 또한 후보자의 데이터 사업 도구 및 서비스를 사용하여 이러한 작업을 효과적으로 수행하는 능력을 테스트합니다.

>> Databricks Databricks-Certified-Professional-Data-Engineer퍼펙트 최신버전 공부자료 <<

## Databricks-Certified-Professional-Data-Engineer시험대비 덤프 최신 데모 - Databricks-Certified-Professional-Data-Engineer최신 시험 공부자료

저희가 알아본 데 의하면 많은 IT인사들이Databricks인증Databricks-Certified-Professional-Data-Engineer시험을 위하여 많은 시간을 투자하고 있다고 합니다. 하지만 특별한 학습 반 혹은 인터넷강이 같은건 선택하지 않으셨습니다.때문에 패스는 아주 어렵습니다.보통은 한번에 패스하시는 분들이 적습니다.우리 Pass4Test에서는 아주 믿을만한 학습가이드를 제공합니다.우리 Pass4Test에는Databricks인증Databricks-Certified-Professional-Data-Engineer테스트버전과 Databricks인증Databricks-Certified-Professional-Data-Engineer문제와 답 두 가지 버전이 있습니다.우리는 여러분의 Databricks인증Databricks-Certified-Professional-Data-Engineer시험을 위한 최고의 문제와 답 제공은 물론 여러분이 원하는 모든 IT인증시험자료들을 선사할 수 있습니다.

# 최신 Databricks Certification Databricks-Certified-Professional-Data-Engineer 무료 샘플문제 (Q11-Q16):

## 질문 # 11

The data engineering team maintains the following code:

```
import pyspark.sql.functions as F

(spark.table("silver_customer_sales")
 .groupBy("customer_id")
 .agg(
   F.min("sale_date").alias("first_transaction_date"),
   F.max("sale_date").alias("last_transaction_date"),
   F.mean("sale_total").alias("average_sales"),
   F.countDistinct("order_id").alias("total_orders"),
   F.sum("sale_total").alias("lifetime_value")
 ).write
 .mode("overwrite")
 .table("gold_customer_lifetime_sales_summary")
)
```

Assuming that this code produces logically correct results and the data in the source table has been de-duplicated and validated, which statement describes what will occur when this code is executed?

- A. The `gold_customer_lifetime_sales_summary` table will be overwritten by aggregated values calculated from all records in the `silver_customer_sales` table as a batch job.
- B. An incremental job will leverage running information in the state store to update aggregate values in the `gold_customer_lifetime_sales_summary` table.
- C. An incremental job will detect if new rows have been written to the `silver_customer_sales` table; if new rows are detected, all aggregates will be recalculated and used to overwrite the `gold_customer_lifetime_sales_summary` table.
- D. The `silver_customer_sales` table will be overwritten by aggregated values calculated from all records in the `gold_customer_lifetime_sales_summary` table as a batch job.
- E. A batch job will update the `gold_customer_lifetime_sales_summary` table, replacing only those rows that have different values than the current version of the table, using `customer_id` as the primary key.

정답: A

설명:

This code is using the `pyspark.sql.functions` library to group the `silver_customer_sales` table by `customer_id` and then aggregate the data using the minimum sale date, maximum sale total, and sum of distinct order ids. The resulting aggregated data is then written to the `gold_customer_lifetime_sales_summary` table, overwriting any existing data in that table. This is a batch job that does not use any incremental or streaming logic, and does not perform any merge or update operations. Therefore, the code will overwrite the `gold` table with the aggregated values from the `silver` table every time it is executed. Reference:

<https://docs.databricks.com/spark/latest/dataframes-datasets/introduction-to-dataframes-python.html>

<https://docs.databricks.com/spark/latest/dataframes-datasets/transforming-data-with-dataframes.html>

<https://docs.databricks.com/spark/latest/dataframes-datasets/aggregating-data-with-dataframes.html>

## 질문 # 12

Below sample input data contains two columns, one `cartId` also known as `session id`, and the second column is called `items`, every time a customer makes a change to the cart this is stored as an array in the table, the Marketing team asked you to create a unique list of items that were ever added to the cart by each customer, fill in blanks by choosing the appropriate array function so the query produces below expected result as shown below.

Schema: `cartId INT, items Array<INT>`

Sample Data

Sample Input data:

cartId	items
1	[1,100,200,300]
1	[1,250,300]

Expected Result

cartId	items
1	[1,100,200,300,250]



- 1.SELECT cartId, \_\_\_\_ (\_\_(items)) as items
- 2.FROM carts GROUP BY cartId

Expected result:

cartId items  
1 [1,100,200,300,250]

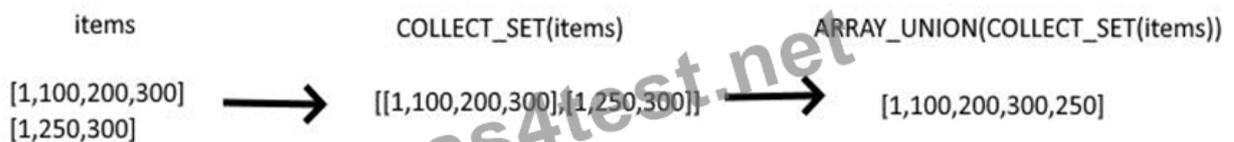
- A. ARRAY\_UNION, FLATTEN
- **B. ARRAY\_UNION, COLLECT\_SET**
- C. ARRAY\_UNION, ARRAY\_DISTINCT
- D. FLATTEN, COLLECT\_UNION
- E. ARRAY\_DISTINCT, ARRAY\_UNION

정답: B

설명:

Explanation

COLLECT SET is a kind of aggregate function that combines a column value from all rows into a unique list ARRAY\_UNION combines and removes any duplicates, Graphical user interface, application Description automatically generated with medium confidence



질문 # 13

What is the purpose of a gold layer in Multi-hop architecture?

- A. Optimizes ETL throughput and analytic query performance
- **B. Powers ML applications, reporting, dashboards and adhoc reports.**
- C. Preserves grain of original data, without any aggregations
- D. Eliminate duplicate records
- E. Data quality checks and schema enforcement

정답: B

설명:

Explanation

The answer is Powers ML applications, reporting, dashboards and adhoc reports.

Review the below link for more info,

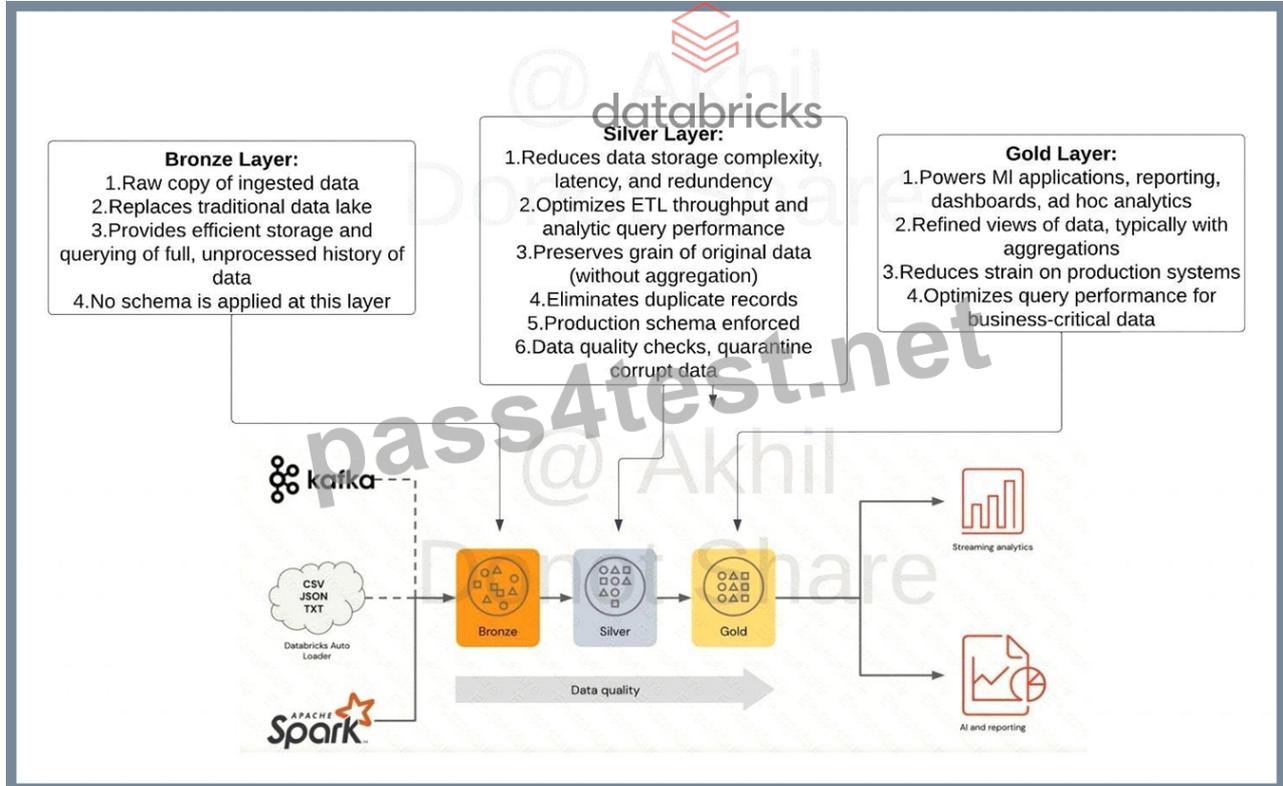
Medallion Architecture - Databricks

Gold Layer:

1. Powers ML applications, reporting, dashboards, ad hoc analytics
2. Refined views of data, typically with aggregations
3. Reduces strain on production systems
4. Optimizes query performance for business-critical data

Exam focus: Please review the below image and understand the role of each layer (bronze, silver, gold) in medallion architecture, you will see varying questions targeting each layer and its purpose.

Sorry I had to add the watermark some people in Udemey are copying my content.



#### 질문 # 14

Each configuration below is identical to the extent that each cluster has 400 GB total of RAM, 160 total cores and only one Executor per VM.

Given a job with at least one wide transformation, which of the following cluster configurations will result in maximum performance?

- A. \* Total VMs:2  
\* 200 GB per Executor  
\* 80 Cores / Executor
- B. \* Total VMs; 1  
\* 400 GB per Executor  
\* 160 Cores / Executor
- C. \* Total VMs: 8  
\* 50 GB per Executor  
\* 20 Cores / Executor
- D. \* Total VMs: 4  
\* 100 GB per Executor  
\* 40 Cores/Executor

정답: C

설명:

This is the correct answer because it is the cluster configuration that will result in maximum performance for a job with at least one wide transformation. A wide transformation is a type of transformation that requires shuffling data across partitions, such as join, groupBy, or orderBy. Shuffling can be expensive and time-consuming, especially if there are too many or too few partitions. Therefore, it is important to choose a cluster configuration that can balance the trade-off between parallelism and network overhead. In this case, having 8 VMs with 50 GB per executor and 20 cores per executor will create 8 partitions, each with enough memory

and CPU resources to handle the shuffling efficiently. Having fewer VMs with more memory and cores per executor will create fewer partitions, which will reduce parallelism and increase the size of each shuffle block. Having more VMs with less memory and cores per executor will create more partitions, which will increase parallelism but also increase the network overhead and the number of shuffle files. Verified References: [Databricks Certified Data Engineer Professional], under "Performance Tuning" section; Databricks Documentation, under "Cluster configurations" section.

### 질문 # 15

The business reporting team requires that data for their dashboards be updated every hour. The total processing time for the pipeline that extracts, transforms, and loads the data for their pipeline runs in 10 minutes. Assuming normal operating conditions, which configuration will meet their service-level agreement requirements with the lowest cost?

- A. Configure a job that executes every time new data lands in a given directory.
- B. Schedule a Structured Streaming job with a trigger interval of 60 minutes.
- C. Schedule a job to execute the pipeline once an hour on a dedicated interactive cluster.
- **D. Schedule a job to execute the pipeline once an hour on a new job cluster.**

정답: D

설명:

Comprehensive and Detailed Explanation From Exact Extract:

\* Exact extract: "Job clusters are created for a job run and terminate when the job completes." References: Cluster types (job vs all-purpose); Databricks Jobs scheduling

### 질문 # 16

.....

Databricks Databricks-Certified-Professional-Data-Engineer 인증 시험 최신버전덤프만 마련하시면 Databricks Databricks-Certified-Professional-Data-Engineer 시험패스는 바로 눈앞에 있습니다. 주문하시면 바로 사이트에서 pdf파일을 다운받을 수 있습니다. Databricks Databricks-Certified-Professional-Data-Engineer 덤프의 pdf버전은 인쇄 가능한 버전이라 공부하기도 편합니다. Databricks Databricks-Certified-Professional-Data-Engineer 덤프 샘플문제를 다운받은 후 굳게 믿고 주문해보세요. 궁금한 점이 있으시면 온라인서비스나 메일로 상담받으시면 됩니다.

**Databricks-Certified-Professional-Data-Engineer 시험대비 덤프 최신 데모**: <https://www.pass4test.net/Databricks-Certified-Professional-Data-Engineer.html>

- Databricks-Certified-Professional-Data-Engineer 최고덤프문제 □ Databricks-Certified-Professional-Data-Engineer 유효한 최신덤프 □ Databricks-Certified-Professional-Data-Engineer 인기공부자료 □ ➔ [www.koreadumps.com](http://www.koreadumps.com) □ 에서 ⇒ Databricks-Certified-Professional-Data-Engineer ⇐ 를 검색하고 무료로 다운로드하세요 Databricks-Certified-Professional-Data-Engineer 최고덤프문제
- Databricks-Certified-Professional-Data-Engineer 인기공부자료 □ Databricks-Certified-Professional-Data-Engineer 퍼펙트 인증덤프 □ Databricks-Certified-Professional-Data-Engineer 자격증문제 □ ✓ [www.itdumps.com](http://www.itdumps.com) □ ✓ □ 에서 검색만 하면 【 Databricks-Certified-Professional-Data-Engineer 】 를 무료로 다운로드할 수 있습니다 Databricks-Certified-Professional-Data-Engineer 높은 통과율 덤프 샘플문제
- 최신 Databricks-Certified-Professional-Data-Engineer 퍼펙트 최신버전 공부자료 공부문제 □ □ [kr.fast2test.com](http://kr.fast2test.com) □ 의 무료 다운로드 ( Databricks-Certified-Professional-Data-Engineer ) 페이지가 지금 열립니다 Databricks-Certified-Professional-Data-Engineer 최신 덤프문제모음집
- Databricks-Certified-Professional-Data-Engineer 최신 덤프문제모음집 □ Databricks-Certified-Professional-Data-Engineer 인증시험 인기 덤프자료 □ Databricks-Certified-Professional-Data-Engineer 높은 통과율 시험덤프 □ 검색만 하면 ✓ [www.itdumps.com](http://www.itdumps.com) □ ✓ □ 에서 ( Databricks-Certified-Professional-Data-Engineer ) 무료 다운로드 Databricks-Certified-Professional-Data-Engineer 인기공부자료
- Databricks-Certified-Professional-Data-Engineer 최고품질 인증 시험 공부자료 □ Databricks-Certified-Professional-Data-Engineer 최고품질 덤프문제 □ Databricks-Certified-Professional-Data-Engineer 최고품질 덤프문제 □ 시험 자료를 무료로 다운로드하려면 [ [kr.fast2test.com](http://kr.fast2test.com) ] 을 통해 [ Databricks-Certified-Professional-Data-Engineer ] 를 검색하십시오 Databricks-Certified-Professional-Data-Engineer 자격증문제
- Databricks-Certified-Professional-Data-Engineer 퍼펙트 최신버전 공부자료 덤프는 시험패스에 가장 좋은 공부자료 □ 무료로 다운로드하려면 > [www.itdumps.com](http://www.itdumps.com) □ 로 이동하여 ⇒ Databricks-Certified-Professional-Data-Engineer ⇐ 를 검색하십시오 Databricks-Certified-Professional-Data-Engineer 높은 통과율 덤프 샘플문제
- 100% 유효한 Databricks-Certified-Professional-Data-Engineer 퍼펙트 최신버전 공부자료 덤프자료 □ ⇒ [www.itdumps.com](http://www.itdumps.com) ⇐ 에서 > Databricks-Certified-Professional-Data-Engineer □ 를 검색하고 무료로 다운로드하세요

