



Databricks Data Engineer Professional

**Databricks Lakehouse Data Engineer Professional
Certification Questions & Answers**

Get Instant Access to Vital Exam Acing
Materials | Study Guide | Sample
Questions | Practice Test

DATA ENGINEER PROFESSIONAL

[Databricks Certified Data Engineer Professional](#)

60 Questions Exam – 70% Cut Score – Duration of 120 minutes

Table of Contents:

Discover More about the Databricks Data Engineer Professional Certification.....	2
Databricks Data Engineer Professional Lakehouse Data Engineer Professional Certification Details:.....	2
Databricks Data Engineer Professional Syllabus:	2
Broaden Your Knowledge with Databricks Data Engineer Professional Sample Questions:	5
Avail the Study Guide to Pass Databricks Data Engineer Professional Lakehouse Data Engineer Professional Exam:	9
Career Benefits:	9

Discover More about the Databricks Data Engineer Professional Certification

Are you interested in passing the Databricks Data Engineer Professional exam? First discover, who benefits from the Data Engineer Professional certification. The Data Engineer Professional is suitable for a candidate if he wants to learn about Databricks Certified Data Engineer Professional. Passing the Data Engineer Professional exam earns you the Databricks Certified Data Engineer Professional title.

While preparing for the Data Engineer Professional exam, many candidates struggle to get the necessary materials. But do not worry; your struggling days are over. The Data Engineer Professional PDF contains some of the most valuable preparation tips and the details and instant access to useful [Data Engineer Professional study materials just at one click.](#)

Databricks Data Engineer Professional Lakehouse Data Engineer Professional Certification Details:

Exam Name	Databricks Certified Data Engineer Professional
Exam Code	Data Engineer Professional
Exam Price	\$200 (USD)
Duration	120 mins
Number of Questions	60
Passing Score	70%
Schedule Exam	Kryterion Webassesor
Sample Questions	Databricks Data Engineer Professional Sample Questions
Practice Exam	Databricks Data Engineer Professional Certification Practice Exam

Databricks Data Engineer Professional Syllabus:

Topic	Details	Weights
Databricks Tooling	<ul style="list-style-type: none"> - Explain how Delta Lake uses the transaction log and cloud object storage to guarantee atomicity and durability - Describe how Delta Lake's Optimistic Concurrency Control provides isolation, and which transactions might conflict 	20%

Topic	Details	Weights
	<ul style="list-style-type: none"> - Describe basic functionality of Delta clone. - Apply common Delta Lake indexing optimizations including partitioning, zorder, bloom filters, and file sizes - Implement Delta tables optimized for Databricks SQL service - Contrast different strategies for partitioning data (e.g. identify proper partitioning columns to use) 	
Data Processing (Batch processing, Incremental processing, and Optimization)	<ul style="list-style-type: none"> - Describe and distinguish partition hints: coalesce, repartition, repartition by range, and rebalance - Contrast different strategies for partitioning data (e.g. identify proper partitioning columns to use) - Articulate how to write Pyspark dataframes to disk while manually controlling the size of individual part-files. - Articulate multiple strategies for updating 1+ records in a spark table (Type 1) - Implement common design patterns unlocked by Structured Streaming and Delta Lake. - Explore and tune state information using stream-static joins and Delta Lake - Implement stream-static joins - Implement necessary logic for deduplication using Spark Structured Streaming - Enable CDF on Delta Lake tables and re-design data processing steps to process CDC output instead of incremental feed from normal Structured Streaming read - Leverage CDF to easily propagate deletes - Demonstrate how proper partitioning of data allows for simple archiving or deletion of data - Articulate, how “smalls” (tiny files, scanning overhead, over partitioning, etc) induce performance problems into Spark queries 	30%

Topic	Details	Weights
Data Modeling	<ul style="list-style-type: none"> - Describe the objective of data transformations during promotion from bronze to silver - Discuss how Change Data Feed (CDF) addresses past difficulties propagating updates and deletes within Lakehouse architecture - Apply Delta Lake clone to learn how shallow and deep clone interact with source/target tables. - Design a multiplex bronze table to avoid common pitfalls when trying to productionalize streaming workloads. - Implement best practices when streaming data from multiplex bronze tables. - Apply incremental processing, quality enforcement, and deduplication to process data from bronze to silver - Make informed decisions about how to enforce data quality based on strengths and limitations of various approaches in Delta Lake - Implement tables avoiding issues caused by lack of foreign key constraints - Add constraints to Delta Lake tables to prevent bad data from being written - Implement lookup tables and describe the trade-offs for normalized data models - Diagram architectures and operations necessary to implement various Slowly Changing Dimension tables using Delta Lake with streaming and batch workloads. - Implement SCD Type 0, 1, and 2 tables 	20%
Security & Governance	<ul style="list-style-type: none"> - Create Dynamic views to perform data masking - Use dynamic views to control access to rows and columns 	10%
Monitoring & Logging	<ul style="list-style-type: none"> - Describe the elements in the Spark UI to aid in performance analysis, application debugging, and tuning of Spark applications. - Inspect event timelines and metrics for stages and jobs performed on a cluster 	10%

Topic	Details	Weights
	<ul style="list-style-type: none"> - Draw conclusions from information presented in the Spark UI, Ganglia UI, and the Cluster UI to assess performance problems and debug failing applications. - Design systems that control for cost and latency SLAs for production streaming jobs. - Deploy and monitor streaming and batch jobs 	
Testing & Deployment	<ul style="list-style-type: none"> - Adapt a notebook dependency pattern to use Python file dependencies - Adapt Python code maintained as Wheels to direct imports using relative paths - Repair and rerun failed jobs - Create Jobs based on common use cases and patterns - Create a multi-task job with multiple dependencies - Design systems that control for cost and latency SLAs for production streaming jobs. - Configure the Databricks CLI and execute basic commands to interact with the workspace and clusters. - Execute commands from the CLI to deploy and monitor Databricks jobs. - Use REST API to clone a job, trigger a run, and export the run output 	10%

Broaden Your Knowledge with Databricks Data Engineer Professional Sample Questions:

Question: 1

Why is log rotation important in log management?

- a) To enhance system security
- b) To prevent log files from becoming too large
- c) To speed up the system
- d) To improve user experience

Answer: b

Question: 2

The data architect has mandated that all tables in the Lakehouse should be configured as external, unmanaged Delta Lake tables. Which approach will ensure that this requirement is met?

- a) Whenever a table is being created, make sure that the LOCATION keyword is used.
- b) When tables are created, make sure that the EXTERNAL keyword is used in the CREATE TABLE statement.
- c) When the workspace is being configured, make sure that external cloud object storage has been mounted.
- d) Whenever a database is being created, make sure that the LOCATION keyword is used.
- e) Whenever a table is being created, make sure that the LOCATION and UNMANAGED keywords are used.

Answer: a

Question: 3

Why is it important to have real-time monitoring in IT systems? (Choose all that apply)

- a) To immediately identify and respond to issues
- b) To track historical data trends
- c) To prevent data loss
- d) To ensure system reliability and availability

Answer: a, d

Question: 4

A data engineering team is trying to transfer ownership from its Databricks Workflows away from an individual that has switched teams. However, they are unsure how permission controls specifically for Databricks Jobs work. Which statement correctly describes permission controls for Databricks Jobs?

- a) The creator of a Databricks Job will always have "Owner" privileges; this configuration cannot be changed.
- b) Databricks Jobs must have exactly one owner; "Owner" privileges cannot be assigned to a group.
- c) Other than the default "admins" group, only individual users can be granted privileges on Jobs.
- d) Only workspace administrators can grant "Owner" privileges to a group.
- e) A user can only transfer Job ownership to a group if they are also a member of that group.

Answer: b

Question: 5

Which technologies are commonly used for stream processing? (Choose all that apply)

- a) Apache Kafka
- b) Hadoop
- c) Apache Spark
- d) MongoDB

Answer: a, c

Question: 6

Which of the following features are provided by Databricks Repos? (Choose all that apply)

- a) Version control integration
- b) Automated machine learning
- c) Collaboration and code sharing
- d) Continuous integration/continuous deployment (CI/CD) capabilities

Answer: a, c, d

Question: 7

A data engineer is developing an ETL workflow that could see late-arriving, duplicate records from its single source. The data engineer knows that they can deduplicate the records within the batch, but they are looking for another solution.

Which approach allows the data engineer to deduplicate data against previously processed records as it is inserted into a Delta table?

- a) VACUUM the Delta table after each batch completes.
- b) Rely on Delta Lake schema enforcement to prevent duplicate records.
- c) Set the configuration `delta.deduplicate = true`.
- d) Perform a full outer join on a unique key and overwrite existing data.
- e) Perform an insert-only merge with a matching condition on a unique key.

Answer: e

Question: 8

What are dimensions in data warehousing?

- a) Measurements of data processing speed
- b) Data categories used for organization and filtering
- c) The physical size of the database
- d) The different levels of data redundancy

Answer: b

Question: 9

A Delta Lake table was created with the query:

- CREATE TABLE dev.my_table
- USING DELTA
- LOCATION "/mnt/dev/my_table"

Realizing that the table needs to be used by other and its name is misleading, the below code was executed:

- ALTER TABLE dev.my_table RENAME TO dev.our_table

Which result will occur after running the second command?

- a) The table name change is recorded in the Delta transaction log.
- b) The table reference in the metastore is updated and all data files are moved.
- c) The table reference in the metastore is updated and no data is changed.
- d) A new Delta transaction log is created for the renamed table.
- e) All related files and metadata are dropped and recreated in a single ACID transaction.

Answer: c

Question: 10

A data engineer needs to use a Python package to process data. As a result, they need to install the Python package on all of the nodes in the currently active cluster. What describes a method of installing a Python package scoped at the notebook level to all nodes in the currently active cluster?

- a) Use %pip install in a notebook cell
- b) Use %sh pip install in a notebook cell
- c) Run source env/bin/activate in a notebook setup script
- d) Install libraries from PyPI using the cluster UI
- e) Use b in a notebook cell

Answer: a

Avail the Study Guide to Pass Databricks Data Engineer Professional Lakehouse Data Engineer Professional Exam:

- Find out about the Data Engineer Professional syllabus topics. Visiting the official site offers an idea about the exam structure and other important study resources. Going through the syllabus topics help to plan the exam in an organized manner.
- Once you are done exploring the [Databricks Data Engineer Professional syllabus](#), it is time to plan for studying and covering the syllabus topics from the core. Chalk out the best plan for yourself to cover each part of the syllabus in a hassle-free manner.
- A study schedule helps you to stay calm throughout your exam preparation. It should contain your materials and thoughts like study hours, number of topics for daily studying mentioned on it. The best bet to clear the exam is to follow your schedule rigorously.
- The candidate should not miss out on the scope to learn from the [Lakehouse Data Engineer Professional training](#). Joining the Databricks provided training for this Databricks certification exam helps a candidate to strengthen his practical knowledge base from the certification.
- Learning about the probable questions and gaining knowledge regarding the exam structure helps a lot. Go through the [Databricks Data Engineer Professional sample questions](#) and boost your knowledge
- Make yourself a pro through online practicing the syllabus topics. Data Engineer Professional practice tests would guide you on your strengths and weaknesses regarding the syllabus topics. Through rigorous practicing, you can improve the weaker sections too. Learn well about time management during exam and become confident gradually with practice tests.

Career Benefits:

Passing the Databricks Data Engineer Professional exam, helps a candidate to prosper highly in his career. Having the certification on the resume adds to the candidate's benefit and helps to get the best opportunities.

Here Is the Trusted Practice Test for the Databricks Data Engineer Professional Certification

CertFun.Com is here with all the necessary details regarding the Data Engineer Professional exam. We provide authentic practice tests for the Data Engineer Professional exam. What do you gain from these practice tests? You get to experience the real exam-like questions made by industry experts and get a scope to improve your performance in the actual exam. Rely on CertFun.Com for rigorous, unlimited two-month attempts on the [Data Engineer Professional practice tests](https://www.certfun.com/databricks/databricks-certified-data-engineer-professional), and gradually build your confidence. Rigorous practice made many aspirants successful and made their journey easy towards grabbing the Databricks Certified Data Engineer Professional.

Start Online practice of Databricks Data Engineer Professional Exam by visiting URL

<https://www.certfun.com/databricks/databricks-certified-data-engineer-professional>