

CERT-PASS

Databricks Data Engineer Associate

Free Practice Questions Preview

Here are 35 sample questions to help you get started. Unlock the full exam to access all 1030+ questions with detailed explanations.

Question 1 : Databricks Intelligence Platform

For a retail analytics workload, during a platform design review, architects compare implementation choices. A team needs a reliable lakehouse foundation for CRM exports that supports rollback after bad writes, consistent access for SQL analysts, and strong Unity Catalog governance. Which approach best fits Databricks Data Intelligence Platform principles? Which platform capability best fits the requirement?

- A. Use temporary views for all curated data because views automatically provide physical rollback and audit history.
- B. Keep only Parquet files in object storage and grant users direct cloud IAM access.
- C. Use Delta Lake tables governed by Unity Catalog so ACID transactions, time travel, lineage, and shared access controls support BI and AI workloads.**
- D. Store curated data as unmanaged CSV files in DBFS and rely on folder naming for versions.

Answer: C

Delta Lake with Unity Catalog provides transactional reliability, governed discovery, lineage, and consistent access. The strongest distractor, raw Parquet or CSV in storage, misses transaction logs, platform governance, and simple rollback.

Question 2 : Databricks Intelligence Platform

For a retail analytics workload, analysts query curated invoices tables during the day while scheduled transformations run overnight. The platform team wants isolation and cost-aware compute. What should they choose?

- A. Use one all-purpose interactive cluster for every analyst query and production ETL job.
- B. Use a single-node cluster for dashboards because it has the lowest possible startup time.
- C. Use model serving compute because it is optimized for SQL dashboard concurrency.
- D. Use a SQL warehouse for analyst SQL workloads and job compute for scheduled ETL tasks.**

Answer: D

SQL warehouses are designed for SQL analytics concurrency, while job compute is suitable for scheduled workloads. A single all-purpose cluster mixes use cases and weakens isolation and cost control.

Question 3 : Databricks Intelligence Platform

For a retail analytics workload, a company is standardizing access to gold_sales_mart and wants discoverability, ownership, and governed permissions across teams. Which organization pattern is most appropriate?

- A. Create one workspace folder per business unit and store production tables as notebook outputs.
- B. Create a cluster per table and rely on cluster permissions to define table ownership.
- C. Use DBFS root paths as the main security boundary for governed datasets.
- D. Create catalogs and schemas in Unity Catalog and organize data by environment, domain, and sensitivity.**

Answer: D

Unity Catalog provides the catalog/schema/table hierarchy for governance and discovery. Workspace folders or cluster permissions do not govern data objects consistently.

Question 4 : Databricks Intelligence Platform

A BI team needs fast startup and concurrency for ad hoc SQL on gold_sales_mart. Which compute service is the best fit?

- A. Use a job cluster because job clusters are designed for long-running interactive BI sessions.
- B. Use Lakeflow Jobs as the query engine because Jobs replace SQL warehouses for ad hoc analytics.
- C. Use Databricks SQL warehouse/serverless SQL when the workload is interactive SQL analytics with many concurrent users.**
- D. Use an all-purpose cluster with a single worker to guarantee the highest concurrency.

Answer: C

SQL warehouses are optimized for SQL analytics and concurrent BI users. Job clusters are best for scheduled job runs, not always-on interactive analytics.

Question 5 : Databricks Intelligence Platform

For a retail analytics workload, a bad load introduced duplicate records into silver_events. The team needs to recover a previous valid state quickly. What platform feature is most relevant?

- A. Change the cluster runtime version and rerun the dashboard query.
- B. Use Delta time travel or restore capabilities after identifying the bad version.**
- C. Grant SELECT to fewer users because permissions automatically remove corrupted rows.
- D. Delete all files in the table path and reload from a local CSV backup.

Answer: B

Delta Lake transaction history enables time travel and restore workflows. Permissions or runtime changes do not revert data contents.

Question 6 : Databricks Intelligence Platform

For a retail analytics workload, a data engineer must select the safest implementation for the requirement. A dashboard based on gold_device_health suddenly changed after a production deployment. The engineer must identify upstream dependencies and recent modifications. What should be used first?

A. Use Unity Catalog lineage and table history to trace upstream tables and recent writes.

B. Inspect only the notebook revision history because it always shows every data write.

C. Search DBFS file names because table lineage is encoded in file names.

D. Ask every user to manually report which query they ran.

Answer: A

Unity Catalog lineage and Delta history help trace dependencies and modifications. Notebook history alone may miss writes from jobs, SQL, or pipelines.

Question 7 : Databricks Intelligence Platform

For a retail analytics workload, a new security telemetry workload includes interactive exploration, scheduled ETL, and BI consumption. Which decision rule best reflects Databricks compute best practice?

A. Always use serverless SQL for streaming ingestion because it replaces streaming compute.

B. Always use the largest all-purpose cluster to avoid tuning decisions.

C. Select compute based on workload characteristics such as interactivity, concurrency, scheduling, startup needs, and cost model.

D. Always use single-node clusters for production because they are simpler to govern.

Answer: C

Compute choice depends on workload type and cost/performance needs. The top wrong answer overprovisions and ignores workload isolation.

Question 8 : Databricks Intelligence Platform

For a retail analytics workload, the data engineering team is creating a governed source of truth for events data. Which storage and registration approach is most aligned to the lakehouse?

A. Store only JSON files in the workspace Files area.

B. Store data in notebook-scoped temporary views and export query results to CSV for consumers.

C. Place all data in the DBFS root and manage sharing with cluster attach permissions.

D. Store curated data in Delta tables registered in Unity Catalog.

Answer: D

Delta tables in Unity Catalog are governed, discoverable, and reliable. DBFS root or workspace files are not the right governed source-of-truth pattern.

Question 9 : Databricks Intelligence Platform

A team needs a reliable lakehouse foundation for IoT sensor drops that supports rollback after bad writes, consistent access for SQL analysts, and minimal custom code. Which approach best fits Databricks Data Intelligence Platform principles?

- A. Use temporary views for all curated data because views automatically provide physical rollback and audit history.
- B. Keep only Parquet files in object storage and grant users direct cloud IAM access.
- C. Store curated data as unmanaged CSV files in DBFS and rely on folder naming for versions.

D. Use Delta Lake tables governed by Unity Catalog so ACID transactions, time travel, lineage, and shared access controls support BI and AI workloads.

Answer: D

Delta Lake with Unity Catalog provides transactional reliability, governed discovery, lineage, and consistent access. The strongest distractor, raw Parquet or CSV in storage, misses transaction logs, platform governance, and simple rollback.

Question 10 : Databricks Intelligence Platform

Analysts query curated transactions tables during the day while scheduled transformations run overnight. The platform team wants isolation and cost-aware compute. What should they choose?

- A. Use model serving compute because it is optimized for SQL dashboard concurrency.
- B. Use a SQL warehouse for analyst SQL workloads and job compute for scheduled ETL tasks.**
- C. Use one all-purpose interactive cluster for every analyst query and production ETL job.
- D. Use a single-node cluster for dashboards because it has the lowest possible startup time.

Answer: B

SQL warehouses are designed for SQL analytics concurrency, while job compute is suitable for scheduled workloads. A single all-purpose cluster mixes use cases and weakens isolation and cost control.

Question 11 : Databricks Intelligence Platform

For a retail analytics workload, a platform team needs a handoff-ready approach that another team can operate reliably. A company is standardizing access to gold_ops_dashboard and wants discoverability, ownership, and governed permissions across teams. Which organization pattern is most appropriate?

- A. Create a cluster per table and rely on cluster permissions to define table ownership.
- B. Create catalogs and schemas in Unity Catalog and organize data by environment, domain, and sensitivity.**
- C. Use DBFS root paths as the main security boundary for governed datasets.
- D. Create one workspace folder per business unit and store production tables as notebook outputs.

Answer: B

Unity Catalog provides the catalog/schema/table hierarchy for governance and discovery. Workspace folders or cluster permissions do not govern data objects consistently.

Question 12 : Databricks Intelligence Platform

For a retail analytics workload, a BI team needs fast startup and concurrency for ad hoc SQL on gold_ops_dashboard. Which compute service is the best fit? What is the best platform-level choice?

A. Use Databricks SQL warehouse/serverless SQL when the workload is interactive SQL analytics with many concurrent users.

B. Use Lakeflow Jobs as the query engine because Jobs replace SQL warehouses for ad hoc analytics.

C. Use an all-purpose cluster with a single worker to guarantee the highest concurrency.

D. Use a job cluster because job clusters are designed for long-running interactive BI sessions.

Answer: A

SQL warehouses are optimized for SQL analytics and concurrent BI users. Job clusters are best for scheduled job runs, not always-on interactive analytics.

Question 13 : Databricks Intelligence Platform

For a retail analytics workload, a bad load introduced duplicate records into silver_claims. The team needs to recover a previous valid state quickly. What platform feature is most relevant?

A. Use Delta time travel or restore capabilities after identifying the bad version.

B. Grant SELECT to fewer users because permissions automatically remove corrupted rows.

C. Delete all files in the table path and reload from a local CSV backup.

D. Change the cluster runtime version and rerun the dashboard query.

Answer: A

Delta Lake transaction history enables time travel and restore workflows. Permissions or runtime changes do not revert data contents.

Question 14 : Databricks Intelligence Platform

A dashboard based on gold_ops_dashboard suddenly changed after a production deployment. The engineer must identify upstream dependencies and recent modifications. What should be used first?

A. Search DBFS file names because table lineage is encoded in file names.

B. Inspect only the notebook revision history because it always shows every data write.

C. Use Unity Catalog lineage and table history to trace upstream tables and recent writes.

D. Ask every user to manually report which query they ran.

Answer: C

Unity Catalog lineage and Delta history help trace dependencies and modifications. Notebook history alone may miss writes from jobs, SQL, or pipelines.

Question 15 : Databricks Intelligence Platform

A new retail POS feeds workload includes interactive exploration, scheduled ETL, and BI consumption. Which decision rule best reflects Databricks compute best practice?

- A. Always use the largest all-purpose cluster to avoid tuning decisions.
- B. Always use single-node clusters for production because they are simpler to govern.
- C. Select compute based on workload characteristics such as interactivity, concurrency, scheduling, startup needs, and cost model.**
- D. Always use serverless SQL for streaming ingestion because it replaces streaming compute.

Answer: C

Compute choice depends on workload type and cost/performance needs. The top wrong answer overprovisions and ignores workload isolation.

Question 16 : Databricks Intelligence Platform

For a retail analytics workload, the solution must remain reliable after ownership transfers to operations. The data engineering team is creating a governed source of truth for invoices data. Which storage and registration approach is most aligned to the lakehouse?

- A. Store data in notebook-scoped temporary views and export query results to CSV for consumers.
- B. Store curated data in Delta tables registered in Unity Catalog.**
- C. Store only JSON files in the workspace Files area.
- D. Place all data in the DBFS root and manage sharing with cluster attach permissions.

Answer: B

Delta tables in Unity Catalog are governed, discoverable, and reliable. DBFS root or workspace files are not the right governed source-of-truth pattern.

Question 17 : Databricks Intelligence Platform

A team needs a reliable lakehouse foundation for IoT sensor drops that supports rollback after bad writes, consistent access for SQL analysts, and cost-aware compute selection. Which approach best fits Databricks Data Intelligence Platform principles?

- A. Store curated data as unmanaged CSV files in DBFS and rely on folder naming for versions.
- B. Keep only Parquet files in object storage and grant users direct cloud IAM access.
- C. Use temporary views for all curated data because views automatically provide physical rollback and audit history.
- D. Use Delta Lake tables governed by Unity Catalog so ACID transactions, time travel, lineage, and shared access controls support BI and AI workloads.**

Answer: D

Delta Lake with Unity Catalog provides transactional reliability, governed discovery, lineage, and consistent access. The strongest distractor, raw Parquet or CSV in storage, misses transaction logs, platform governance, and simple rollback.

Question 18 : Databricks Intelligence Platform

For a retail analytics workload, analysts query curated customers tables during the day while scheduled transformations run overnight. The platform team wants isolation and cost-aware compute. What should they choose? What is the best platform-level choice?

- A. Use a SQL warehouse for analyst SQL workloads and job compute for scheduled ETL tasks.
- B. Use a single-node cluster for dashboards because it has the lowest possible startup time.
- C. Use one all-purpose interactive cluster for every analyst query and production ETL job.
- D. Use model serving compute because it is optimized for SQL dashboard concurrency.

Answer: A

SQL warehouses are designed for SQL analytics concurrency, while job compute is suitable for scheduled workloads. A single all-purpose cluster mixes use cases and weakens isolation and cost control.

Question 19 : Databricks Intelligence Platform

A company is standardizing access to gold_ops_dashboard and wants discoverability, ownership, and governed permissions across teams. Which organization pattern is most appropriate?

- A. Create a cluster per table and rely on cluster permissions to define table ownership.
- B. Create one workspace folder per business unit and store production tables as notebook outputs.
- C. Use DBFS root paths as the main security boundary for governed datasets.
- D. Create catalogs and schemas in Unity Catalog and organize data by environment, domain, and sensitivity.

Answer: D

Unity Catalog provides the catalog/schema/table hierarchy for governance and discovery. Workspace folders or cluster permissions do not govern data objects consistently.

Question 20 : Databricks Intelligence Platform

For a retail analytics workload, a BI team needs fast startup and concurrency for ad hoc SQL on gold_finance_kpi. Which compute service is the best fit?

- A. Use an all-purpose cluster with a single worker to guarantee the highest concurrency.
- B. Use Databricks SQL warehouse/serverless SQL when the workload is interactive SQL analytics with many concurrent users.
- C. Use Lakeflow Jobs as the query engine because Jobs replace SQL warehouses for ad hoc analytics.
- D. Use a job cluster because job clusters are designed for long-running interactive BI sessions.

Answer: B

SQL warehouses are optimized for SQL analytics and concurrent BI users. Job clusters are best for scheduled job runs, not always-on interactive analytics.

Question 21 : Databricks Intelligence Platform

For a retail analytics workload, a new data pipeline is moving from development into production. A bad load introduced duplicate records into silver_sessions. The team needs to recover a previous valid state quickly. What platform feature is most relevant?

- A. Grant SELECT to fewer users because permissions automatically remove corrupted rows.
- B. Delete all files in the table path and reload from a local CSV backup.
- C. Change the cluster runtime version and rerun the dashboard query.
- D. Use Delta time travel or restore capabilities after identifying the bad version.**

Answer: D

Delta Lake transaction history enables time travel and restore workflows. Permissions or runtime changes do not revert data contents.

Question 22 : Databricks Intelligence Platform

For a retail analytics workload, a dashboard based on gold_sales_mart suddenly changed after a production deployment. The engineer must identify upstream dependencies and recent modifications. What should be used first?

- A. Search DBFS file names because table lineage is encoded in file names.
- B. Inspect only the notebook revision history because it always shows every data write.
- C. Ask every user to manually report which query they ran.
- D. Use Unity Catalog lineage and table history to trace upstream tables and recent writes.**

Answer: D

Unity Catalog lineage and Delta history help trace dependencies and modifications. Notebook history alone may miss writes from jobs, SQL, or pipelines.

Question 23 : Databricks Intelligence Platform

A new CRM exports workload includes interactive exploration, scheduled ETL, and BI consumption. Which decision rule best reflects Databricks compute best practice? Which Databricks platform design should be used?

- A. Always use single-node clusters for production because they are simpler to govern.
- B. Select compute based on workload characteristics such as interactivity, concurrency, scheduling, startup needs, and cost model.**
- C. Always use the largest all-purpose cluster to avoid tuning decisions.
- D. Always use serverless SQL for streaming ingestion because it replaces streaming compute.

Answer: B

Compute choice depends on workload type and cost/performance needs. The top wrong answer overprovisions and ignores workload isolation.

Question 24 : Databricks Intelligence Platform

For a retail analytics workload, the data engineering team is creating a governed source of truth for products data. Which storage and registration approach is most aligned to the lakehouse?

- A. Store data in notebook-scoped temporary views and export query results to CSV for consumers.
- B. Store only JSON files in the workspace Files area.
- C. Place all data in the DBFS root and manage sharing with cluster attach permissions.
- D. Store curated data in Delta tables registered in Unity Catalog.**

Answer: D

Delta tables in Unity Catalog are governed, discoverable, and reliable. DBFS root or workspace files are not the right governed source-of-truth pattern.

Question 25 : Databricks Intelligence Platform

A team needs a reliable lakehouse foundation for security telemetry that supports rollback after bad writes, consistent access for SQL analysts, and strong Unity Catalog governance. Which approach best fits Databricks Data Intelligence Platform principles?

- A. Use Delta Lake tables governed by Unity Catalog so ACID transactions, time travel, lineage, and shared access controls support BI and AI workloads.**
- B. Use temporary views for all curated data because views automatically provide physical rollback and audit history.
- C. Keep only Parquet files in object storage and grant users direct cloud IAM access.
- D. Store curated data as unmanaged CSV files in DBFS and rely on folder naming for versions.

Answer: A

Delta Lake with Unity Catalog provides transactional reliability, governed discovery, lineage, and consistent access. The strongest distractor, raw Parquet or CSV in storage, misses transaction logs, platform governance, and simple rollback.

Question 26 : Databricks Intelligence Platform

For a retail analytics workload, an engineer reviews a teammate's proposed Databricks implementation. Analysts query curated shipments tables during the day while scheduled transformations run overnight. The platform team wants isolation and cost-aware compute. What should they choose?

- A. Use one all-purpose interactive cluster for every analyst query and production ETL job.
- B. Use a single-node cluster for dashboards because it has the lowest possible startup time.
- C. Use a SQL warehouse for analyst SQL workloads and job compute for scheduled ETL tasks.**
- D. Use model serving compute because it is optimized for SQL dashboard concurrency.

Answer: C

SQL warehouses are designed for SQL analytics concurrency, while job compute is suitable for scheduled workloads. A single all-purpose cluster mixes use cases and weakens isolation and cost control.

Question 27 : Databricks Intelligence Platform

In a finance reporting pipeline, a company is standardizing access to gold_sales_mart and wants discoverability, ownership, and governed permissions across teams. Which organization pattern is most appropriate? Scenario variant 2.

- A. Create one workspace folder per business unit and store production tables as notebook outputs.
- B. Use DBFS root paths as the main security boundary for governed datasets.
- C. Create catalogs and schemas in Unity Catalog and organize data by environment, domain, and sensitivity.**
- D. Create a cluster per table and rely on cluster permissions to define table ownership.

Answer: C

Unity Catalog provides the catalog/schema/table hierarchy for governance and discovery. Workspace folders or cluster permissions do not govern data objects consistently.

Question 28 : Databricks Intelligence Platform

For a retail analytics workload, a BI team needs fast startup and concurrency for ad hoc SQL on gold_revenue_daily. Which compute service is the best fit?

- A. Use Lakeflow Jobs as the query engine because Jobs replace SQL warehouses for ad hoc analytics.
- B. Use Databricks SQL warehouse/serverless SQL when the workload is interactive SQL analytics with many concurrent users.**
- C. Use an all-purpose cluster with a single worker to guarantee the highest concurrency.
- D. Use a job cluster because job clusters are designed for long-running interactive BI sessions.

Answer: B

SQL warehouses are optimized for SQL analytics and concurrent BI users. Job clusters are best for scheduled job runs, not always-on interactive analytics.

Question 29 : Databricks Intelligence Platform

In a finance reporting pipeline, a bad load introduced duplicate records into silver_events. The team needs to recover a previous valid state quickly. What platform feature is most relevant? Scenario variant 2.

- A. Change the cluster runtime version and rerun the dashboard query.
- B. Delete all files in the table path and reload from a local CSV backup.
- C. Grant SELECT to fewer users because permissions automatically remove corrupted rows.
- D. Use Delta time travel or restore capabilities after identifying the bad version.**

Answer: D

Delta Lake transaction history enables time travel and restore workflows. Permissions or runtime changes do not revert data contents.

Question 30 : Databricks Intelligence Platform

For a retail analytics workload, a dashboard based on `gold_device_health` suddenly changed after a production deployment. The engineer must identify upstream dependencies and recent modifications. What should be used first?

- A. Inspect only the notebook revision history because it always shows every data write.
- B. Ask every user to manually report which query they ran.
- C. Use Unity Catalog lineage and table history to trace upstream tables and recent writes.**
- D. Search DBFS file names because table lineage is encoded in file names.

Answer: C

Unity Catalog lineage and Delta history help trace dependencies and modifications. Notebook history alone may miss writes from jobs, SQL, or pipelines.

Question 31 : Databricks Intelligence Platform

For a retail analytics workload, in a production readiness review, the data engineering team evaluates competing designs. A new payment gateway extracts workload includes interactive exploration, scheduled ETL, and BI consumption. Which decision rule best reflects Databricks compute best practice?

- A. Always use the largest all-purpose cluster to avoid tuning decisions.
- B. Always use serverless SQL for streaming ingestion because it replaces streaming compute.
- C. Select compute based on workload characteristics such as interactivity, concurrency, scheduling, startup needs, and cost model.**
- D. Always use single-node clusters for production because they are simpler to govern.

Answer: C

Compute choice depends on workload type and cost/performance needs. The top wrong answer overprovisions and ignores workload isolation.

Question 32 : Databricks Intelligence Platform

The data engineering team is creating a governed source of truth for invoices data. Which storage and registration approach is most aligned to the lakehouse?

- A. Store data in notebook-scoped temporary views and export query results to CSV for consumers.
- B. Store curated data in Delta tables registered in Unity Catalog.**
- C. Store only JSON files in the workspace Files area.
- D. Place all data in the DBFS root and manage sharing with cluster attach permissions.

Answer: B

Delta tables in Unity Catalog are governed, discoverable, and reliable. DBFS root or workspace files are not the right governed source-of-truth pattern.

Question 33 : Databricks Intelligence Platform

A team needs a reliable lakehouse foundation for CRM exports that supports rollback after bad writes, consistent access for SQL analysts, and strong Unity Catalog governance. Which approach best fits Databricks Data Intelligence Platform principles?

- A. Store curated data as unmanaged CSV files in DBFS and rely on folder naming for versions.
- B. Use temporary views for all curated data because views automatically provide physical rollback and audit history.
- C. Use Delta Lake tables governed by Unity Catalog so ACID transactions, time travel, lineage, and shared access controls support BI and AI workloads.**
- D. Keep only Parquet files in object storage and grant users direct cloud IAM access.

Answer: C

Delta Lake with Unity Catalog provides transactional reliability, governed discovery, lineage, and consistent access. The strongest distractor, raw Parquet or CSV in storage, misses transaction logs, platform governance, and simple rollback.

Question 34 : Databricks Intelligence Platform

For a retail analytics workload, analysts query curated shipments tables during the day while scheduled transformations run overnight. The platform team wants isolation and cost-aware compute. What should they choose? Which platform capability best fits the requirement?

- A. Use a single-node cluster for dashboards because it has the lowest possible startup time.
- B. Use one all-purpose interactive cluster for every analyst query and production ETL job.
- C. Use a SQL warehouse for analyst SQL workloads and job compute for scheduled ETL tasks.**
- D. Use model serving compute because it is optimized for SQL dashboard concurrency.

Answer: C

SQL warehouses are designed for SQL analytics concurrency, while job compute is suitable for scheduled workloads. A single all-purpose cluster mixes use cases and weakens isolation and cost control.

Question 35 : Databricks Intelligence Platform

For a retail analytics workload, a company is standardizing access to gold_customer_360 and wants discoverability, ownership, and governed permissions across teams. Which organization pattern is most appropriate? Which Databricks platform design should be used?

- A. Use DBFS root paths as the main security boundary for governed datasets.
- B. Create a cluster per table and rely on cluster permissions to define table ownership.
- C. Create one workspace folder per business unit and store production tables as notebook outputs.
- D. Create catalogs and schemas in Unity Catalog and organize data by environment, domain, and sensitivity.**

Answer: D

Unity Catalog provides the catalog/schema/table hierarchy for governance and discovery. Workspace folders or cluster permissions do not govern data objects consistently.

Unlock All 1030+ Questions

Get the complete Q&A package with detailed explanations, topic analytics, and exam-accurate practice.

From €29.00

Visit: <https://cert-pass.com/exams/databricks-data-engineer-associate>

CERT-PASS

cert-pass.com

© 2026 Cert-Pass. This material is for personal use only. Do not distribute.