DBCLOUD
EXAMS

QUESTIONS & ANSWERS

verified



Technology Provider: Microsoft

Exam: Implementing Data Engineering Solutions Using Azure Databricks

Number of questions in the database: 109

Exam Version: NOV.2026

Question #1 Topic 1

Instructions

This is a case study. Case studies are not timed separately from other exam sections. You can use as much exam time as you would like to complete each case study. However, there might be additional case studies or other exam sections. Manage your time to ensure that you can complete all the exam sections in the time provided. Pay attention to the Exam Progress at the top of the screen so you have sufficient time to complete any exam sections that follow this case study.

To answer the case study questions, you will need to reference information that is provided in the case. Case studies and associated questions might contain exhibits or other resources that provide more information about the scenario described in the case.

Information provided in an individual question does not apply to the other questions in the case study.

A Review Screen will appear at the end of this case study. From the Review Screen, you can review and change your answers before you move to the next exam section. After you leave this case study, you will NOT be able to return to it.

To start the case study

To display the first question in this case study, select the "Next" button. To the left of the question, a menu provides links to information such as business requirements, the existing environment, and problem statements. Please read through all this information before answering any questions. When you are ready to answer a question, select the "Question" button to return to the question.

Overview

Company Information

Constoso, Inc. es a renewable energy provider that operates solar and wind farms across North America

Existing Environment

Azure Environment

Contoso has a single Azure Databricks workspace named Workspace1 in the West US Azure region. Workspace1 is enabled for Unity Catalog.

Workspace1 contains all-purpose clusters for both development and production workloads.

https://dbcloud.academy/

The company's Azure environment contains:

• In the West US, Central US, and East US Azure regions, Azure Event Hubs that stream telemetry data and an Azure Data Lake Storage Gen2 account in each region for each hub

• A single Azure SQL database in the West US region that hosts enterprise resource planning (ERP) data

• An Azure Database for PostgreSQL server in the West US region that stores operational maintenance data

Data Environment

Contoso ingests the following operational and business data:

• Telemetry data: More than 40,000 IoT sensors across 28 sites emit JSON telemetry events every few seconds. Each site sends the events to the nearest event hub, which writes the data into the corresponding Data Lake Storage Gen2 account. These files frequently experience schema drift.

• Maintenance logs: Maintenance systems generate historical repair logs, daily incremental updates, technician notes, and unstructured attachments that are stored in the Data Lake Storage Gen2 accounts.

• Operational maintenance data: Structured operational maintenance data is stored on the Azure Database for PostgreSQL server.

• External weather data: Hourly weather forecasts are retrieved from a REST API and written to the Data Lake Storage Gen2 accounts.

• ERP data: Daily CSV extracts of 50 to 100 GB contain equipment metadata, work orders, and purchase order information.

Problem Statements

The company's existing analytics environment has several issues:

Ingestion

• Telemetry pipelines fall behind during peak loads.

• Telemetry ingestion fails when schema drift occurs.

• Streaming pipelines reprocess events after a pipeline restarts.

Compute

• Production and development workloads run on the same all-purpose clusters.

• Production and development workloads do NOT support autoscaling or workload isolation.

Governance

• The ERP data is duplicated across systems and development teams.

https://dbcloud.academy/

• Naming conventions are inconsistent across development teams, regions, and products.

• Ownership of the IoT sensors changes over time, and analysts must track the full history of the ownership.

• Occasionally, equipment manufacturers must correct data-entry mistakes in equipment names. Historical values are NOT required.

Pipeline operations

• Pipelines lack resiliency, alerting, and centralized scheduling.

Requirements

Planned Changes

Contoso plans to implement the following changes:

• Implement scalable data pipeline orchestration.

• Create a managed analytics catalog in Unity Catalog.

• Implement a consistent approach to creating curated datasets.

• Establish a centralized governance model across ingestion, cleansed, and curated layers.

• Grant data engineers access to the ERP tables by using minimal development effort.

• Adopt a compute strategy that isolates production workloads and supports autoscaling.

• Adopt a slowly changing dimension (SCD) approach to address current data modeling issues.

Technical Requirements

Contoso identifies the following environment and compute requirements:

• Ensure that production ingestion workloads run on compute clusters that can scale automatically during telemetry spikes.

• Provide fast and consistent performance for business intelligence (BI) workloads.

• Prevent development activity from affecting production pipelines.

• Production ingestion workloads must run as scheduled, non-interactive pipelines rather than on shared interactive development clusters.

Contoso identifies the following data ingestion and processing requirements:

• Auto-scale ingestion pipelines to handle bursty workloads.

• Handle schema drift for the maintenance and telemetry data.

• Ingest file-based telemetry data by using minimal operational effort.

• Store all the ingested data in a format that supports incremental processing.

https://dbcloud.academy/

• Support the continuous ingestion of telemetry data from the event hubs by using exactly-once semantics.

• Support the ingestion of the structured maintenance data from the Azure Database for PostgreSQL server.

• Build a new telemetry pipeline that ingests raw events from the event hubs, cleanses the data, and publishes curated tables to Unity Catalog.

• Ensure that the Apache Spark Structured Streaming pipelines reading from the event hubs write the data into a managed Delta table named telemetry.raw_events. The pipelines must support schema drift and resume processing after failures without reprocessing the data.

Contoso identifies the following data modeling and optimization requirements:

• Build curated tables that standardize business logic.

• Overwrite equipment metadata attributes, such as name, manufacturer, model, and commissioning date, when the attributes change. Historical values are NOT required.

Contoso identifies the following pipeline deployment and operation requirements:

• Orchestrate multi-step ingestion and transformation workflows.

• Define a clear execution order and dependencies.

• Automatically retry failed steps and notify operators.

• Schedule ingestion and transformation workloads consistently.

Governance Requirements

Contoso identifies the following governance requirements:

• Centralize the metadata catalog.

• Provide isolated development areas that follow standard naming conventions.

• Establish a consistent structure for organizing raw, cleansed, and curated data.

• Provide a read-only mechanism to reference the ERP data through a foreign catalog.

Business Requirements

Contoso identifies the following business requirements:

• Improve ingestion reliability and reduce operational effort.

• Standardize data definitions across development teams.

You need to configure compute for the ingestion of telemetry data. The solution must meet the data ingestion and processing requirements.

What should you do?

https://dbcloud.academy/

A. Increase an all-purpose cluster to a larger fixed node type.

B. Enable Photon acceleration for a job compute cluster.

C. Disable autoscaling for a job compute cluster.

D. Move the ingestion pipelines to shared compute.

Correct Answer: B

Question #2 Topic 1

Instructions

This is a case study. Case studies are not timed separately from other exam sections. You can use as much exam time as you would like to complete each case study. However, there might be additional case studies or other exam sections. Manage your time to ensure that you can complete all the exam sections in the time provided. Pay attention to the Exam Progress at the top of the screen so you have sufficient time to complete any exam sections that follow this case study.

To answer the case study questions, you will need to reference information that is provided in the case. Case studies and associated questions might contain exhibits or other resources that provide more information about the scenario described in the case.

Information provided in an individual question does not apply to the other questions in the case study.

A Review Screen will appear at the end of this case study. From the Review Screen, you can review and change your answers before you move to the next exam section. After you leave this case study, you will NOT be able to return to it.

To start the case study

To display the first question in this case study, select the "Next" button. To the left of the question, a menu provides links to information such as business requirements, the existing environment, and problem statements. Please read through all this information before answering any questions. When you are ready to answer a question, select the "Question" button to return to the question.

Overview

Company Information

Constoso, Inc. es a renewable energy provider that operates solar and wind farms across North America

Existing Environment

Azure Environment

Contoso has a single Azure Databricks workspace named Workspace1 in the West US Azure region. Workspace1 is enabled for Unity Catalog.

https://dbcloud.academy/

Workspace1 contains all-purpose clusters for both development and production workloads.

The company's Azure environment contains:

• In the West US, Central US, and East US Azure regions, Azure Event Hubs that stream telemetry data and an Azure Data Lake Storage Gen2 account in each region for each hub

• A single Azure SQL database in the West US region that hosts enterprise resource planning (ERP) data

• An Azure Database for PostgreSQL server in the West US region that stores operational maintenance data

Data Environment

Contoso ingests the following operational and business data:

• Telemetry data: More than 40,000 IoT sensors across 28 sites emit JSON telemetry events every few seconds. Each site sends the events to the nearest event hub, which writes the data into the corresponding Data Lake Storage Gen2 account. These files frequently experience schema drift.

• Maintenance logs: Maintenance systems generate historical repair logs, daily incremental updates, technician notes, and unstructured attachments that are stored in the Data Lake Storage Gen2 accounts.

• Operational maintenance data: Structured operational maintenance data is stored on the Azure Database for PostgreSQL server.

• External weather data: Hourly weather forecasts are retrieved from a REST API and written to the Data Lake Storage Gen2 accounts.

• ERP data: Daily CSV extracts of 50 to 100 GB contain equipment metadata, work orders, and purchase order information.

Problem Statements

The company's existing analytics environment has several issues:

Ingestion

• Telemetry pipelines fall behind during peak loads.

• Telemetry ingestion fails when schema drift occurs.

• Streaming pipelines reprocess events after a pipeline restarts.

Compute

• Production and development workloads run on the same all-purpose clusters.

• Production and development workloads do NOT support autoscaling or workload isolation.

Governance

https://dbcloud.academy/

• The ERP data is duplicated across systems and development teams.

• Naming conventions are inconsistent across development teams, regions, and products.

• Ownership of the IoT sensors changes over time, and analysts must track the full history of the ownership.

• Occasionally, equipment manufacturers must correct data-entry mistakes in equipment names. Historical values are NOT required.

Pipeline operations

• Pipelines lack resiliency, alerting, and centralized scheduling.

Requirements

Planned Changes

Contoso plans to implement the following changes:

• Implement scalable data pipeline orchestration.

• Create a managed analytics catalog in Unity Catalog.

• Implement a consistent approach to creating curated datasets.

• Establish a centralized governance model across ingestion, cleansed, and curated layers.

• Grant data engineers access to the ERP tables by using minimal development effort.

• Adopt a compute strategy that isolates production workloads and supports autoscaling.

• Adopt a slowly changing dimension (SCD) approach to address current data modeling issues.

Technical Requirements

Contoso identifies the following environment and compute requirements:

• Ensure that production ingestion workloads run on compute clusters that can scale automatically during telemetry spikes.

• Provide fast and consistent performance for business intelligence (BI) workloads.

• Prevent development activity from affecting production pipelines.

• Production ingestion workloads must run as scheduled, non-interactive pipelines rather than on shared interactive development clusters.

Contoso identifies the following data ingestion and processing requirements:

• Auto-scale ingestion pipelines to handle bursty workloads.

• Handle schema drift for the maintenance and telemetry data.

• Ingest file-based telemetry data by using minimal operational effort.

https://dbcloud.academy/

• Store all the ingested data in a format that supports incremental processing.

• Support the continuous ingestion of telemetry data from the event hubs by using exactly-once semantics.

• Support the ingestion of the structured maintenance data from the Azure Database for PostgreSQL server.

• Build a new telemetry pipeline that ingests raw events from the event hubs, cleanses the data, and publishes curated tables to Unity Catalog.

• Ensure that the Apache Spark Structured Streaming pipelines reading from the event hubs write the data into a managed Delta table named telemetry.raw_events. The pipelines must support schema drift and resume processing after failures without reprocessing the data.

Contoso identifies the following data modeling and optimization requirements:

• Build curated tables that standardize business logic.

• Overwrite equipment metadata attributes, such as name, manufacturer, model, and commissioning date, when the attributes change. Historical values are NOT required.

Contoso identifies the following pipeline deployment and operation requirements:

• Orchestrate multi-step ingestion and transformation workflows.

• Define a clear execution order and dependencies.

• Automatically retry failed steps and notify operators.

• Schedule ingestion and transformation workloads consistently.

Governance Requirements

Contoso identifies the following governance requirements:

• Centralize the metadata catalog.

• Provide isolated development areas that follow standard naming conventions.

• Establish a consistent structure for organizing raw, cleansed, and curated data.

• Provide a read-only mechanism to reference the ERP data through a foreign catalog.

Business Requirements

Contoso identifies the following business requirements:

• Improve ingestion reliability and reduce operational effort.

• Standardize data definitions across development teams.

You need to develop the task logic for a new job in Lakeflow Jobs that processes telemetry data.

https://dbcloud.academy/

Each task must contain only the appropriate logic for its step in the pipeline. The solution must support the planned changes and meet the data ingestion and processing requirements.

What should you do?

A. Use a single Databricks notebook task that performs ingestion, cleansing, and curation in one script.

B. Create three tasks that each contains the identical logic and use task retries.

C. Create separate tasks for ingestion, cleansing, and curation.

D. Use a single SQL task that performs ingestion, cleansing, and curation by running MERGE commands.

Correct Answer: C

Question #3 Topic 1

You have an Azure Databricks workspace that is enabled for Unity Catalog and contains a catalog named Catalog1. Catalog1 contains a table named Transactions. Transactions contains the following columns:

• transaction_id

• customer_name

• email_address

• credit_card_number

• transaction_amount

You need to ensure that business analysts can query all the rows in the Transactions table. The solution must meet the following requirements:

• Prevent the analysts from seeing the full values in the email_address and credit_card_number columns.

• Ensure that the analysts can see only the values after the @ character in each email address.

• Ensure that the analysts can see only the last four digits of each credit card number.

• Enable analysts to query the table without errors.

• Follow the principle of least privilege.

What should you do?

A. Grant the analysts the SELECT permission for the Transactions table and apply column-level encryption.

https://dbcloud.academy/