Question1 Recommend an Azure Data solution based on requirements

Case Study

Complete the Case Study

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3

Instructions

This case study contains a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

Note: You cannot go back or review questions of this type on the actual certification exam.

Question 1.1

Your company supports a software as a service (SaaS) application with a large number of customers. You are designing a support database to use elastic pools and elastic queries to retrieve customer information. Customer information is stored in a table that includes values for CustomerID and RegionalID.

You need to partition data to optimize queries by customer sorted by geographic location. The solution should minimize support costs.

You need to recommend a partitioning strategy.

Solution: You configure horizontal partitioning based on CustomerID.

Does this solution meet the goal?

Complete the Case Study

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3

No

Yes

This solution does not meet the requirements. You need to use sharding, which is partitioning data horizontally to distribute data across multiple databases in a scaled-out design, but CustomerID is not the best choice in this scenario. Sharding by RegionalID will make sorting by geographic location more efficient.

Sharding requires that the schema is the same on all of the databases involved. Sharding helps to minimize the size of individual databases, which in turn helps to improve transactional process performance. Hardware support requirements are minimized, which helps to reduce related costs. Elastic queries let you run queries across multiple shards. You can configure and manage sharding through the elastic database tools libraries or through self-sharding.

References

[Scaling out with Azure SQL Database](#)

[What is database sharding?](#)

[Azure SQL Database elastic query overview (preview)](#)

Question 1.2

Your company supports a software as a service (SaaS) application with a large number of customers. You are designing a support database to use elastic pools and elastic queries to retrieve customer information. Customer information is stored in a table that includes values for CustomerID and RegionalID.

You need to partition data to optimize queries by customer sorted by geographic location. The solution should minimize support costs.

You need to recommend a partitioning strategy.

Solution: You configure vertical partitioning based on CustomerID.

Does this solution meet the goal?

Complete the Case Study

No
Yes

Explanation

This solution does not meet the requirements. Vertical partitioning does not provide for data partitioning in the way needed by this scenario. Vertical partitioning is a way to organize data columns and can be used to make retrieving data through JOIN operations more efficient.

You need to use sharding, which is partitioning data horizontally to distribute data across multiple databases in a scaled-out design. This requires that the schema is the same on all of the databases involved. Sharding helps to minimize the size of individual databases, which in turn helps to improve transactional process performance. Sharding by RegionalID will make sorting by geographic location more efficient. Hardware support requirements are minimized, which helps to reduce related costs. Elastic queries let you run queries across multiple shards. You can configure and manage sharding through the elastic database tools libraries or through self-sharding.

References

Scaling out with Azure SQL Database

What is database sharding?

Azure SQL Database elastic query overview (preview)

Question 1.3

Your company supports a software as a service (SaaS) application with a large number of customers. You are designing a support database to use elastic pools and elastic queries to retrieve customer information. Customer information is stored in a table that includes values for CustomerID and RegionalID.

You need to partition data to optimize queries by customer sorted by geographic location. The solution should minimize support costs.

You need to recommend a partitioning strategy.

Solution: You configure sharding by RegionalID.

Does this solution meet the goal?

No
Yes

Explanation

This solution meets the requirements. Sharding by RegionalID, which would segregate items by region, would help optimize sorting by geographic area. Another reason you might want to use RegionalID for sharding is that it might be necessary to have different parts of the database configured to reside in different geographies for compliance or geopolitical reasons.

Sharding partitions data horizontally to distribute data across multiple databases in a scaled-out design. This requires that the schema is the same on all of the databases involved. Sharding helps to minimize the size of individual databases, which in turn helps to improve transactional process performance. Hardware support requirements are minimized, which helps to reduce related costs. Elastic queries let you run queries across multiple shards. You can configure and manage sharding through the elastic database tools libraries or through self-sharding.

References

Scaling out with Azure SQL Database

What is database sharding?

Azure SQL Database elastic query overview (preview)

Case Study

Complete the Case Study

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3

Instructions

This case study contains a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

Note: You cannot go back or review questions of this type on the actual certification exam.

Question 4.1

You are migrating an application and its on-premises SQL Server Enterprise Edition database to the cloud. Application and database changes should be kept to a minimum during migration. You want to choose the exact amount of compute resources dedicated to the workload. Management overhead should be kept to a minimum.

You need to choose an appropriate deployment and purchase model to meet your needs.

Solution: You choose an elastic pool deployment and eDTU pricing model.

Does this solution meet the goal?

Complete the Case Study

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3

No
Yes

Explanation

The solution does not meet the requirements. You should not choose an elastic pool deployment and eDTU pricing model. An elastic pool is designed to provide an optimum environment across multiple databases with varying resource requirements. The more databases included and the more random (and non-overlapping) the requirements, the more cost-effective this model becomes. The eDTU pricing model lets you set guidelines for resources that can be scaled and shared between the databases as needed. The eDTU pricing model does not let you choose the exact amount of compute resources dedicated to the workload.

You need to deploy a managed instance because this deployment best meets your needs. A managed instance provides the environment that is most compatible with SQL Server Enterprise Edition and helps to keep the changes to the application and database to a minimum. Supported features that help to reduce the management overhead include automatic patching, automatic version updates, automated backups, and high availability. The vCore pricing model lets you configure specific computing and storage resources for the deployment.

References

Azure SQL Database pricing

Choose between the vCore and the DTU purchasing models

What is Azure SQL Database managed instance?

Elastic pools help you manage and scale multiple Azure SQL databases

Question 4.2

You are migrating an application and its on-premises SQL Server Enterprise Edition database to the cloud. Application and database changes should be kept to a minimum during migration. You want to choose the exact amount of compute resources dedicated to the workload. Management overhead should be kept to a minimum.

You need to choose an appropriate deployment and purchase model to meet your needs.

Solution: You choose a managed instance deployment and vCore pricing model.

Does this solution meet the goal?

No
Yes

Explanation

The solution meets the requirements. You need to deploy a managed instance because this deployment best meets your needs. A managed instance provides the environment that is most compatible with SQL Server Enterprise Edition and helps to keep the changes to the application and database to a minimum. Supported features that help to reduce the management overhead include automatic patching, automatic version updates, automated backups, and high availability. The vCore pricing model lets you configure specific computing and storage resources for the deployment.

References

Azure SQL Database pricing

Choose between the vCore and the DTU purchasing models

What is Azure SQL Database managed instance?

Question 4.3

You are migrating an application and its on-premises SQL Server Enterprise Edition database to the cloud. Application and database changes should be kept to a minimum during migration. You want to choose the exact amount of compute resources dedicated to the workload. Management overhead should be kept to a minimum.

You need to choose an appropriate deployment and purchase model to meet your needs.

Solution: You choose a single database deployment and vCore pricing model.

Does this solution meet the goal?

Complete the Case Study

- Solution Evaluation
    - Question 1
    - Question 2
    - Question 3

Yes
No

Explanation

The solution does not meet the requirements. You should not choose a single database deployment and vCore pricing model. A single database deployment does not meet the same level of compatibility as a managed instance, the preferred deployment in this scenario. This deployment creates an Azure SQL Database instance with its own resources.

You need to deploy a managed instance because this deployment best meets your needs. A managed instance provides the environment that is most compatible with SQL Server Enterprise Edition and helps to keep the changes to the application and database to a minimum. Supported features that help to reduce the management overhead include automatic patching, automatic version updates, automated backups, and high availability. The vCore pricing model lets you configure specific computing and storage resources for the deployment.

References

Azure SQL Database pricing

Choose between the vCore and the DTU purchasing models

What is Azure SQL Database managed instance?

What is a single database in Azure SQL Database

Case Study

Complete the Case Study

- Solution Evalutation
  - Question 1
  - Question 2
  - Question 3
  - Question 4

Instructions

This case study contains a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

Note: You cannot go back or review questions of this type on the actual certification exam.

Question 7.1

You are a data architect for a grocery delivery company. The development team is building an app that allows grocery stores to delivery groceries. Individual grocery stores sell different types of grocery products. You want to implement a solution that does not restrict the product attributes that are used by each grocery store. You want to access product information by using OData queries.

You need to create the appropriate data store.

Solution: You create an Azure table.

Does this solution meet the goal?

Complete the Case Study

- Solution Evalutation
  - Question 1
  - Question 2
  - Question 3
  - Question 4

No

Explanation

This solution meets the goal. An Azure table stores entities as key-value pairs. Each entity can have an arbitrary number of attributes, and it does not have to have the same number of attributes as other entities. It also supports OData queries.

References

Introduction to Azure Storage

Question 7.2

You are a data architect for a grocery delivery company. The development team is building an app that allows grocery stores to delivery groceries. Individual grocery stores sell different types of grocery products. You want to implement a solution that does not restrict the product attributes that are used by each grocery store. You want to access product information by using OData queries.

You need to create the appropriate data store.

Solution: You create an Azure Cosmos DB account with the Table API.

Does this solution meet the goal?

Complete the Case Study

- Solution Evalutation
  - Question 1
  - Question 2
  - Question 3
  - Question 4

No
Yes

Explanation

This solution meets the goal. Azure Cosmos DB is a multi-model, non-relational data store that allows you to query data by using one of five APIs. The Table API allows you to use OData and Language Integrated Query (LINQ) to query data. You can issue LINQ queries with .NET.

References

[Introduction to Azure Cosmos DB: Table API](#)

[Welcome to Azure Cosmos DB](#)

Question 7.3

You are a data architect for a grocery delivery company. The development team is building an app that allows grocery stores to delivery groceries. Individual grocery stores sell different types of grocery products. You want to implement a solution that does not restrict the product attributes that are used by each grocery store. You want to access product information by using OData queries.

You need to create the appropriate data store.

Solution: You create an Azure SQL Data Warehouse.

Does this solution meet the goal?

Complete the Case Study

- Solution Evalutation
  - Question 1
  - Question 2
  - Question 3
  - Question 4

No
Yes

Explanation

This solution does not meet the goal. Azure SQL Data Warehouse is an Enterprise Data Warehouse (EDW) that uses Massively Parallel Data Processing (MPPD). It allows you to run complex queries across compute nodes for fast response. It is a relational data store. Therefore, it is not suitable for this scenario. In this scenario, product data has varying attributes, which indicates that it is non-relational. In this scenario, each grocery store has different types of products. Azure Table storage or Cosmos DB with a Table API supports this.

References

[Introduction to Azure Storage](#)

Question 7.4

You are a data architect for a grocery delivery company. The development team is building an app that allows grocery stores to delivery groceries. Individual grocery stores sell different types of grocery products. You want to implement a solution that does not restrict the product attributes that are used by each grocery store. You want to access product information by using OData queries.

You need to create the appropriate data store.

Solution: You create an Azure SQL Database.

Does this solution meet the goal?

Complete the Case Study

- Solution Evalutation
  - Question 1
  - Question 2
  - Question 3
  - Question 4

Yes
No

Explanation

This solution does not meet the goal. Azure SQL Database is the Platform-as-a-Service (PaaS) equivalent of an on-premises Microsoft SQL Server deployment. SQL Database is a relational data store. In this scenario, the solution must use a non-relational data store because product attributes vary.

References

Case Study

Complete the Case Study

- Background

Company A is a merchant integration solutions company. The company works with website owners and merchants to enable merchants to advertise their products. The company has an existing on-premises solution that uses pools of servers and storage. You plan to develop a new cloud-based solution.

- Existing Process

Each merchant uploads product files to your company's FTP server over the weekend. A Windows service invokes merchant-specific components that process those files. Some merchants upload XML files, some upload CSV files, some upload JSON files, and some use proprietary file formats. Some of the files are over 1 petabyte.

- Technical Requirements

The conversion process must be moved to the cloud. Customers should be able to continue to send their files to the FTP server. However, another service should then move those files to Azure. Once the files are moved to Azure, parallel processes should transform the files from merchant-specific formats to a common relational format. The resulting data should be spread across compute nodes by vendor so that parallel queries can be run. There are 60 vendors.

- Product Table

The common relational format of the resulting Product table is defined as follows:

- Id: This field uniquely identifies a product. It is an alphanumeric string that has variable length, with a maximum of 20 characters.
- Price: This field represents the cost of one unit of the product. It is a currency value.
- Name: This field represents the name of a product. It is an alphanumeric string that has variable length, with a maximum of 60 characters.
- Description: This field describes the product. It is an alphanumeric string that has variable length, with a maximum of 200 characters.
- Vendor: This field represents the name of one of the 60 vendors.

The Product table will primarily be read-only during the week. The only time that it will be updated is on the weekend, when new files are retrieved. It is expected that only 10 percent of the table will change. The total number of rows is expected to be around one million.

- Vendor Table

The common relational format of the resulting Vendor table is defined as follows:

- Id: This field uniquely identifies a vendor. It is an alphanumeric string that has variable length, with a maximum of 20 characters.
- Name: This field represents the name of a product. It is an alphanumeric string that has variable length, with a maximum of 60 characters.

The Vendor table will be read-only all the time. It will be used in joins with the Product table. Reads against this table must be fast.

- Question 1
- Question 2
- Question 3
- Question 4
- Question 5
- Question 6

Question 11.1
You need to choose the service for copying product files from the FTP server to Azure.

Which service should you use?

Complete the Case Study

- Background
- Existing Process
- Technical Requirements
- Product Table
- Vendor Table
- Question 1
- Question 2
- Question 3
- Question 4
- Question 5
- Question 6

Azure Databricks
Azure SQL Database

Azure Data Migration Assistant

Explanation

You should use Azure Data Factory. Azure Data Factory allows you to create data-driven workflows. You can copy data from an on-premises server to the cloud. It supports data from multiple sources.

You should not use Azure SQL Database. Azure SQL Database is a Platform-as-a-Service (PaaS) offering that is the equivalent of on-premises SQL Server. You cannot use it to copy data from an on-premises FTP server to Azure.

You should not use Azure Data Migration Assistant. This helps you detect compatibility issues when upgrading SQL Server or SQL Data Warehouse. You cannot use it to copy data from an on-premises FTP server to Azure.

You should not use Azure Databricks. Azure Databricks is a Spark-based analytics platform that allows you to easily view data graphically. You cannot use it to copy data from an on-premises FTP server to Azure.

References

[What is Azure Data Factory?](#)

[Copy data from FTP server by using Azure Data Factory](#)

[What is the Azure SQL Database service?](#)

[Overview of Data Migration Assistant](#)

[What is Azure Databricks?](#)

Question 11.2

You need to choose the data store that serves as the destination when copying product files from the FTP server.

Which data source should you choose?

Complete the Case Study

Azure SQL Database
Azure Data Warehouse

Azure Databricks

Azure Data Lake

Explanation

You should choose Azure Data Lake. Azure Data Lake is a repository for big data analytics workloads. It allows you to store data of any type and size.

You should not choose Azure Databricks. Azure Databricks is a Spark-based analytics platform that allows you to easily view data graphically.

You should not choose Azure Data Warehouse. Azure Data Warehouse is an Enterprise Data Warehouse (EDW) that uses Massively Parallel Processing (MPP) to process data across compute nodes. You cannot use it to copy data from an on-premises FTP server to Azure.

You should not use Azure SQL Database. Azure SQL Database is a Platform-as-a-Service (PaaS) offering that is the equivalent of on-premises SQL Server. You cannot use it to copy data from an on-premises FTP server to Azure.

References

What is Azure SQL Data Warehouse?

What is Azure Databricks?

What is the Azure SQL Database service?

Question 11.3

You need to choose the data source for the final product data.

Which data source should you choose?

Complete the Case Study

- Background
- Existing Process
- Technical Requirements
  - Product Table
  - Vendor Table
    - Question 1
    - Question 2
    - Question 3
    - Question 4
    - Question 5
    - Question 6

Azure SQL Data Warehouse

Azure Table

Azure Cosmos DB

Azure Databricks

Explanation

You should choose Azure SQL Data Warehouse. Azure Data Warehouse is an Enterprise Data Warehouse (EDW) that uses Massively Parallel Processing (MPP) to process data across compute nodes. You can run parallel queries across multiple compute nodes.

You should not use Azure Table storage. Azure Table storage is a NoSQL structured data store that allows you to store entities (rows) with varying attributes (columns). It is not a relational data store. In this scenario, you need to store relational data.

You should not use Azure Cosmos DB. Azure Cosmos DB is a multi-model, non-relational data store that supports one of five APIs. In this scenario, you need to store relational data.

You should not use Azure Databricks. Azure Databricks is a Spark-based analytics

platform that allows you to easily view data graphically. You cannot use it to copy data from an on-premises FTP server to Azure.

References

[What is Azure SQL Data Warehouse?](#)

[Introduction to Azure Storage](#)

[Azure Table storage overview](#)

[What is the Azure SQL Database service?](#)

[Welcome to Azure Cosmos DB](#)

Question 11.4

You need to design the partition distribution scheme for the Product table.

Which scheme should you use?

Complete the Case Study

- Background
- Existing Process
- Technical Requirements
  - Product Table
  - Vendor Table
  - Question 1
  - Question 2
  - Question 3
  - Question 4
  - Question 5
  - Question 6

Round-robin distribution with a clustered columnstore index
Replicated distribution with a non-clustered columnstore index

Round-robin distribution with a non-clustered columnstore index

Hash distribution with a clustered columnstore index

Explanation

You should use hash distribution with a columnstore index. Hash distribution is recommended for large fact tables. Table rows are distributed across compute nodes by a hash function, which is usually performed against a column. In this case, the column should be one that identifies a vendor. This allows table rows to be distributed across compute nodes by vendor. Each compute node contains products from a specific vendor. There are a total of 60 distributions of rows in SQL Data Warehouse. Because there are 60 vendors, it distributes products from each vendor across the 60 distributions.

You should not use replicated distribution with a non-clustered columnstore index. Replicated distribution places a replicated table across each compute node. This means that the entire table will be distributed across the 60 distributions.

You should not use round-robin distribution. Round-robin distribution distributes table rows evenly across compute nodes. In this scenario, you want to distribute rows across compute nodes by vendor.

References

Create tables and perform queries

Guidance for designing distributed tables in Azure SQL Data Warehouse

Columnstore Indexes Described

Design guidance for using replicated tables in Azure SQL Data Warehouse

Question 11.5

You need to design the partition distribution scheme for the Vendor table.

Which scheme should you use?

Complete the Case Study

- Background
- Existing Process
- Technical Requirements
  - Product Table
  - Vendor Table
  - Question 1

Replicated table with a clustered columnstore index

Round-robin distribution with a non-clustered columnstore index

Hash distribution with a non-clustered columnstore index

Round-robin distribution with a clustered columnstore index

Explanation

You should use a replicated table with a clustered columnstore index. Replicated distribution places a replicated table across each compute node. This means that the entire table will be distributed across the 60 vendor nodes. This is recommended for small tables. There are only 60 vendors, so the Vendor table contains only 60 rows. By copying the Vendor table to each compute node, you allow joins against the Product table to be performed quickly. The clustered columnstore index causes reads against the table to be fast.

You should not use hash distribution with a non-clustered columnstore index. Hash distribution is recommended for large fact tables. Table rows are distributed across compute nodes by a hash function, which is usually performed against a column. In this case, the Vendor table is small.

You should not use round-robin distribution. Round-robin distribution distributes table rows evenly across compute nodes. In this scenario, you should copy the entire Vendor table across all distributions.

References

[Create tables and perform queries](#)

[Guidance for designing distributed tables in Azure SQL Data Warehouse](#)

[Columnstore indexes: Overview](#)

[Design guidance for using replicated tables in Azure SQL Data Warehouse](#)

You need to recommend a solution to implement the components that transform data from merchant-specific formats.

What should you use?

- Background
- Existing Process
- Technical Requirements
  - Product Table
  - Vendor Table
  - Question 1
  - Question 2
  - Question 3
  - Question 4
  - Question 5
  - Question 6

SQL Server Integration Services (SSIS)

PowerShell

Azure CLI

Polybase

Explanation

You should use Polybase. Polybase allows you to use SQL to import data from Azure Data Lake and Azure blob storage containers. In this scenario, data should be stored in Azure Data Lake. Azure Data Lake allows you to store data of any type and of any size.

You should not use SSIS. SSIS cannot read data from Azure Data Lake, which is where the files should be copied to from the FTP server.

You should not use PowerShell or Azure CLI. Both allow you to connect to and access files from Azure Data Lake. However, neither allow you to transform those files.

References

Load data by using PolyBase

What is PolyBase?

Case Study

Complete the Case Study

- Solution Evalutation

- **Instructions**

  This case study contains a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

  Note: You cannot go back or review questions of this type on the actual certification exam.

- Data Model
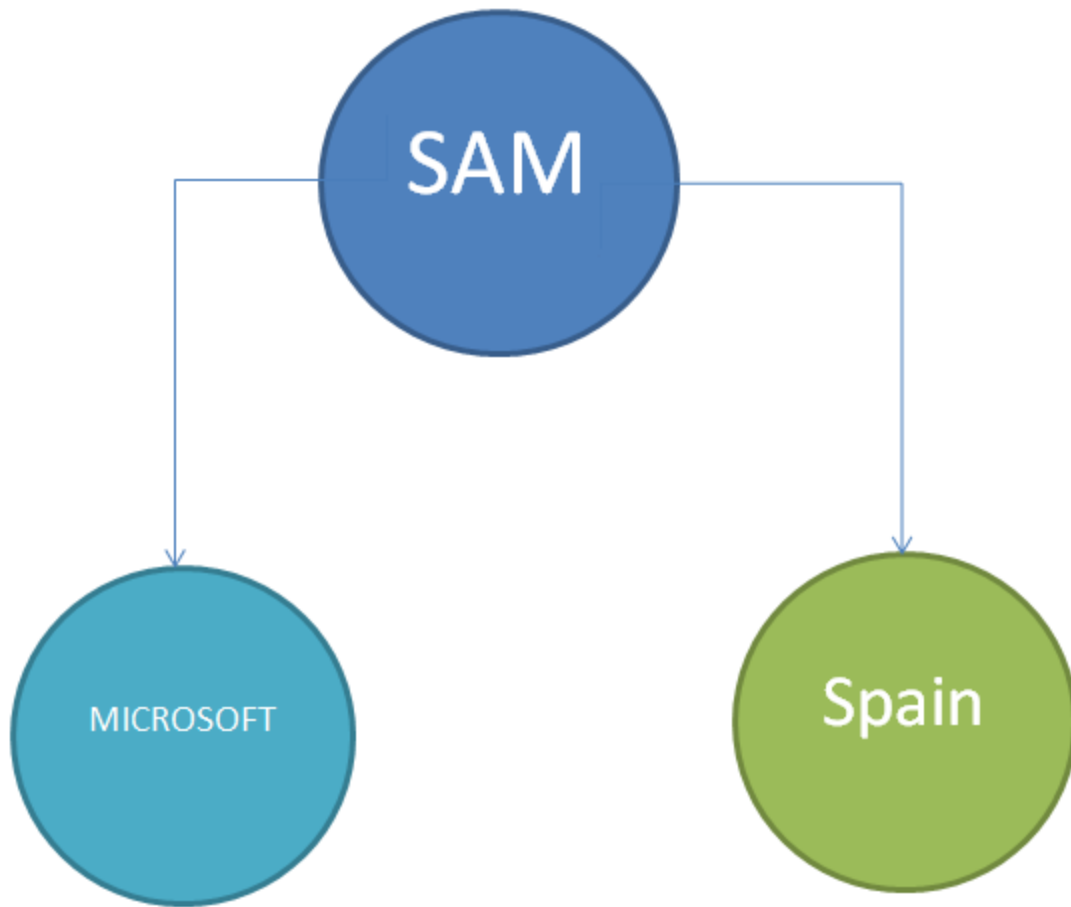
Question 17.1

You are a data architect for your company. The company's application development team is upgrading an application and migrating it to the cloud. The current application uses a graph data model that represents data as vertices and edges, as shown in the Data Model exhibit.

You need to choose a cloud data store for the model.

Solution: You choose Azure Cosmos DB with the Table API.

Does this solution meet the goal?

Complete the Case Study

- Solution Evalutation
  - Data Model
  - Question 1
  - Question 2

Yes
No

Explanation

You should not use Azure Cosmos DB with the Table API. This API allows you to store data as key-attribute pairs. It is not suitable for graph data.

Cosmos DB with the Gremlin API meets the goal. Azure Cosmos DB is a multi-model database that supports one of five APIs. The Gremlin API allows you to store, manage, and access graph data.

References

[Introduction to Azure Cosmos DB: Gremlin API](#)

[Introduction to Azure Cosmos DB: Table API](#)

Question 17.2

You are a data architect for your company. The company's application development team is upgrading an application and migrating it to the cloud. The current application uses a graph data model that represents data as vertices and edges, as shown in the Data Model exhibit.

You need to choose a cloud data store for the model.

Solution: You choose Azure Cosmos DB with the Gremlin API.

Does this solution meet the goal?

Complete the Case Study

- Solution Evalutation
  - Data Model
  - Question 1
  - Question 2
  - Question 3
  - Question 4

Yes
No

Explanation

This solution meets the goal. Azure Cosmos DB is a multi-model database that supports one of five APIs. The Gremlin API allows you to store, manage, and access graph data.

References

[Introduction to Azure Cosmos DB: Gremlin API](#)

[Graph data modeling for Azure Cosmos DB Gremlin API](#)

You are a data architect for your company. The company's application development team is upgrading an application and migrating it to the cloud. The current application uses a graph data model that represents data as vertices and edges, as shown in the Data Model exhibit.

You need to choose a cloud data store for the model.

Solution: You choose Azure Cosmos DB with the SQL API.

Does this solution meet the goal?

Complete the Case Study

- Solution Evalutation
  - Data Model
  - Question 1
  - Question 2
  - Question 3
  - Question 4

Yes
No

Explanation

This solution does not meet the goal. The SQL API allows you to access relational data, not graph data.

Cosmos DB with the Gremlin API meets the goal. Azure Cosmos DB is a multi-model database that supports one of five APIs. The Gremlin API allows you to store, manage,

and access graph data.

References

Question 17.4

You are a data architect for your company. The company's application development team is upgrading an application and migrating it to the cloud. The current application uses a graph data model that represents data as vertices and edges, as shown in the Data Model exhibit.

You need to choose a cloud data store for the model.

Solution: You choose Azure Cosmos DB with the MongoDB API.

Does this solution meet the goal?

Complete the Case Study

- Solution Evalutation
  - Data Model
  - Question 1
  - Question 2
  - Question 3
  - Question 4

Yes
No

Explanation

This solution does not meet the goal. The MongoDB API allows you to access JSON documents, not graph data.

Cosmos DB with the Gremlin API meets the goal. Azure Cosmos DB is a multi-model database that supports one of five APIs. The Gremlin API allows you to store, manage, and access graph data.

References

Case Study

Complete the Case Study

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3

Instructions

This case study contains a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

Note: You cannot go back or review questions of this type on the actual certification exam.

Question 21.1

You are a data architect for your company. Your company manages data from customers all over the world. Each customer supplies your company with data in a variety of formats. Your company must transform the data after it receives it. The total size of all customer data is under one pebibyte (PiB).

You need to recommend a data storage solution for customer data.

Solution: You recommend Azure Table storage.

Does this solution meet the goal?

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3

Yes
No

Explanation

This solution does not meet the goal. Azure Table storage allows you to store data as key-attribute pairs. In this scenario, you want to store the data in a variety of formats. Azure blob storage and Azure Data Lake support data of any format, such as XML, CSV, and binary.

References

Choosing a big data storage technology in Azure

Azure Table storage overview

Question 21.2

You are a data architect for your company. Your company manages data from customers all over the world. Each customer supplies your company with data in a variety of formats. Your company must transform the data after it receives it. The total size of all customer data is under one pebibyte (PiB).

You need to recommend a data storage solution for customer data.

Solution: You recommend Azure Blob storage.

Does this solution meet the goal?

Complete the Case Study

- Solution Evaluation
  - Question 1
  - Question 2

Yes

No

Explanation

This solution meets the goal. Azure Blob Storage allows you to store data in any format. Each blob storage account supports up to two PiB of data.

References

[Choosing a big data storage technology in Azure](#)

[Azure subscription and service limits, quotas, and constraints](#)

Question 21.3

You are a data architect for your company. Your company manages data from customers all over the world. Each customer supplies your company with data in a variety of formats. Your company must transform the data after it receives it. The total size of all customer data is under one pebibyte (PiB).

You need to recommend a data storage solution for customer data.

Solution: You recommend Azure Data Lake.

Does this solution meet the goal?

Complete the Case Study

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3

Yes

No

Explanation

This solution meets the goal. Azure Data Lake is a hyperscale repository for big data analytics workloads. It supports data of any type and size.

References

Question22

Question23

Question24 Design non-relational cloud data stores

Case Study

Complete the Case Study

- Solution Evaluation

- **Instructions**

    This case study contains a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

    Note: You cannot go back or review questions of this type on the actual certification exam.

- Sample Data

Division	Team	Wins	Losses
East	Sharks	4	3
East	Whales	2	5
East	Goats	2	5
South	Birds	5	2
South	Reptiles	7	0
South	Dogs	5	2
West	Bears	3	4
West	Lions	1	6
West	Tigers	3	4
North	Dinosaurs	6	1
North	Cats	1	6
North	Zebras	7	0

You are a data architect. You are designing a data solution to manage statistics for a world-wide sports league. You plan to store the data in an Azure table. Every team in the league has a unique name, and each team is part of a division. There are a total of four divisions. The statistics keep track of the wins and losses for each team in the division. Sample data is shown in the Sample Data exhibit. The production table will contain over 4,000 rows.

You need to recommend an appropriate partition key.

Solution: You choose Division as the partition key.

Does this solution meet the goal?

Complete the Case Study

- Solution Evaluation
  - Sample Data
    - Question 1
  - Question 2
  - Question 3
  - Question 4

No
Yes

Explanation

This solution meets the goal. In Azure Tables, partition keys separate data into different servers. All data that shares the same partition key resides on the same server, although the same server can store multiple partitions. By choosing Division as the partition key, you can speed up searches by division. All rows that have the same division are stored on the same physical server.

References

Designing a Scalable Partitioning Strategy for Azure Table Storage

You are a data architect. You are designing a data solution to manage statistics for a world-wide sports league. You plan to store the data in an Azure table. Every team in the league has a unique name, and each team is part of a division. There are a total of four divisions. The

statistics keep track of the wins and losses for each team in the division. Sample data is shown in the Sample Data exhibit. The production table will contain over 4,000 rows.

You need to recommend an appropriate partition key.

Solution: You choose Team as the partition key.

Does this solution meet the goal?

No
Yes

Explanation

This solution does not meet the goal. In this scenario, team names are unique through the league. This means that every row would be stored on a different physical server, which would actually negatively impact performance.

References

[Designing a Scalable Partitioning Strategy for Azure Table Storage](#)

Question 24.3

You are a data architect. You are designing a data solution to manage statistics for a world-wide sports league. You plan to store the data in an Azure table. Every team in the league has a unique name, and each team is part of a division. There are a total of four divisions. The statistics keep track of the wins and losses for each team in the division. Sample data is shown in the Sample Data exhibit. The production table will contain over 4,000 rows.

You need to recommend an appropriate partition key.

Solution: You choose Wins as the partition key.

Does this solution meet the goal?

Yes
No

Explanation

This solution meets the goal. In Azure Tables, partition keys separate data into different servers. All data that shares the same partition key resides on the same server, although the same server can store multiple partitions. By choosing Wins as the partition key, you can speed up searches by wins. All rows that have the same number of wins are stored on the same physical server. However, it is also possible that Azure will create range partitions, causing all rows that have wins within a particular range to be stored on the same physical server. For example, Azure might create two range partitions: one that contains wins between 0 and 3, and another that contains wins between 4 and 7.

References

[Designing a Scalable Partitioning Strategy for Azure Table Storage](#)

Question 24.4

You are a data architect. You are designing a data solution to manage statistics for a world-wide sports league. You plan to store the data in an Azure table. Every team in the league has a unique name, and each team is part of a division. There are a total of four divisions. The statistics keep track of the wins and losses for each team in the division. Sample data is shown in the Sample Data exhibit. The production table will contain over 4,000 rows.

You need to recommend an appropriate partition key.

Solution: You choose Losses as the partition key.

Does this solution meet the goal?

- Solution Evaluation
  - Sample Data
  - Question 1
  - Question 2
  - Question 3
  - Question 4

No
Yes

Explanation

This solution meets the goal. In Azure Tables, partition keys separate data into different servers. All data that shares the same partition key resides on the same server, although the same server can store multiple partitions. By choosing Losses as the partition key, you can speed up searches by losses. All rows that have the same number of losses are stored on the same physical server. However, it is also possible that Azure will create range partitions, causing all rows that have losses within a particular range to be stored on the same physical server. For example, Azure might create two range partitions: one that contains losses between 0 and 3, and another that contains losses between 4 and 7.

References

Designing a Scalable Partitioning Strategy for Azure Table Storage

Question25

Question26

Question27

Question28 Design non-relational cloud data stores
Case Study

- Scenario Evaluation

Instructions

This case study contains a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

Note: You cannot go back or review questions of this type on the actual certification exam.

Question 28.1

You are designing an HDInsight application with a clustered Apache Hadoop file system that uses Azure Data Lake Gen1 storage. The application design specifies the following requirements:

- Azure Active Directory (Azure AD) multi-factor authentication (MFA)
- Access control through POSIX permissions
- Support for auditing through diagnostics logging
- Automatic data encryption before persistent storage

You need to configure your solution to optimize storage performance.

Solution: You batch process files into file sizes of at least 256 MB before writing to Data Lake Storage Gen1.

Does this solution meet the goal?

Complete the Case Study

- Scenario Evaluation
  - Question 1
  - Question 2
  - Question 3

No
Yes

Explanation

The solution meets the requirements. You should batch smaller files into larger files. Depending on how the data is used, a file size of at least 256 MB and less than 2 GB is recommended. You can batch the files when landing in Data Lake Storage Gen1. If this

is not possible, you can run a separate compaction job to combine smaller files into larger files. This is to prevent the possible impact to performance with supporting the overhead associated with POSIX permissions and auditing. This also helps to reduce the number of authentication checks needed and reduces the number of open files, which means you have fewer files to process when updating POSIX permissions. This also allows for faster copying and replication.

References

[What is Azure Data Lake Storage Gen1?](#)

[Best practices for using Azure Data Lake Storage Gen1](#)

[Tuning Azure Data Lake Storage Gen1 for performance](#)

[Encryption of data in Azure Data Lake Storage Gen1](#)

Question 28.2

You are designing an HDInsight application with a clustered Apache Hadoop file system that uses Azure Data Lake Gen1 storage. The application design specifies the following requirements:

- Azure Active Directory (Azure AD) multi-factor authentication (MFA)
- Access control through POSIX permissions
- Support for auditing through diagnostics logging
- Automatic data encryption before persistent storage

You need to configure your solution to optimize storage performance.

Solution: You preprocess files to ensure that they are smaller than 256 MB before writing to Data Lake Storage Gen1.

Does this solution meet the goal?
Complete the Case Study

- Scenario Evaluation
  - Question 1
  - Question 2
  - Question 3

No
Yes

Explanation

The solution does not meet the requirements. Using file sizes smaller than 256 MB can impact performance, especially when considering the overhead associated with POSIX permissions and auditing.

Depending on how the data is used, a file size of at least 256 MB and less than 2 GB is recommended. You can batch the files when landing in Data Lake Storage Gen1. If this is not possible, you can run a separate compaction job to combine smaller files into larger files. This also helps to reduce the number of authentication checks needed and reduces the number of open files, which means you have fewer files to process when updating POSIX permissions. This also allows for faster copying and replication.

References

[What is Azure Data Lake Storage Gen1?](#)

[Best practices for using Azure Data Lake Storage Gen1](#)

[Tuning Azure Data Lake Storage Gen1 for performance](#)

[Encryption of data in Azure Data Lake Storage Gen1](#)

Question 28.3

You are designing an HDInsight application with a clustered Apache Hadoop file system that uses Azure Data Lake Gen1 storage. The application design specifies the following requirements:

- Azure Active Directory (Azure AD) multi-factor authentication (MFA)
- Access control through POSIX permissions
- Support for auditing through diagnostics logging
- Automatic data encryption before persistent storage

You need to configure your solution to optimize storage performance.

Solution: You implement compaction jobs to combine smaller files into files that are 2 GB in size or larger.

Does this solution meet the goal?
Complete the Case Study

No
Yes

Explanation

The solution does not meet the requirements. File compaction is a viable solution to reduce the number of smaller files supported, but a maximum file size of no more than 2 GB is recommended.

Depending on how the data is used, a file size of at least 256 MB and less than 2 GB is recommended. You can batch the files when landing in Data Lake Storage Gen1. If this is not possible, you can run a separate compaction job to combine smaller files into larger files. This is to prevent the possible impact to performance with supporting the overhead associated with POSIX permissions and auditing. This also helps to reduce the number of authentication checks needed and reduces the number of open files, which means you have fewer files to process when updating POSIX permissions. This also allows for faster copying and replication.

References

What is Azure Data Lake Storage Gen1?

Best practices for using Azure Data Lake Storage Gen1

Tuning Azure Data Lake Storage Gen1 for performance

Encryption of data in Azure Data Lake Storage Gen1

Question31 Design non-relational cloud data stores

Case Study

Complete the Case Study

- Overview

You are the database administrator for CompanyA. The company is a reseller of internet information about financial and distribution markets. The company is planning an investment into Azure and is looking for the right type of data platform to be able to consume data from all the various incoming data feeds. The feeds consist of a variety of formats including CSV files, JSON documents, XML files and others. The data is presently copied to Azure using a custom script.

The company wants to find the simplest way to ingest the data, transform it into a relational format, and make it readily available for consumers to view as well as ensuring efficient query latency.

Question 31.1

What type of data platform in Azure should be used as the initial destination of the incoming data feeds?

Complete the Case Study

- Overview
- Question 1
- Question 2
- Question 3
- Question 4

Azure SQL Database
Azure Synapse

Azure Data Lake

Azure PolyBase

Explanation

> You should use Azure Data Lake as the initial incoming destination for the incoming data feed. Azure Data Lake offers a platform that supports different and varying data types to be collated together for consumption by other Azure data services.

You should not use Azure PolyBase. Azure PolyBase is used to transform the data into a format that is consumable by a relational database platform, but it does not support storing multiple data formats.

You should not use Azure Synapse. Azure Synapse, formally known as Azure Data Warehouse, is a platform used to process and visualize the data, but it does not support unstructured incoming data formats.

You should not use Azure SQL Database. Azure SQL Database requires a referential data format that can be mapped to a fixed schema. In this case there are multiple formats that need to be supported.

References

[What is Azure Data Factory?](#)

[Data Lake](#)

[PolyBase in SQL Server 2016](#)

[Azure Synapse Analytics](#)

[Azure SQL Database](#)

Question 31.2

What solution should you use to transform the data from the various formats and import the data into a relational platform?

Complete the Case Study

- Overview
- Question 1
- Question 2
- Question 3
- Question 4

Azure PolyBase

Azure Data Lake

Azure Synapse Analytics

Azure Data Factory

Explanation

You should use Azure PolyBase. Azure PolyBase is used to transform the data into a format that is consumable by a relational database platform.

You should not use Azure Data Lake. Azure Data Lake offers a platform that supports different and varying data types to be collated together for consumption by other Azure data services.

You should not use Azure Synapse. Azure Synapse, formally known as Azure Data Warehouse, is a platform used for processing the data, but it does not support unstructured incoming data formats.

You should not use Azure Data Factory. Azure Data Factory can be used to bring sources of data to a central location for processing. Azure PolyBase can then be used to transform the data into a workable format for querying and analysis.

References

What is Azure Data Factory?

Data Lake

What is PolyBase?

Azure Synapse Analytics

Question 31.3

For the final step in the processing of the data, you need to select the right destination Azure repository for the transformed data.

Which destination should you use?

Complete the Case Study

- Overview
- Question 1
- Question 2
- Question 3
- Question 4

Azure Data Factory
Azure Data Lake

Azure Synapse

Azure PolyBase

Explanation

You should use Azure Synapse. Azure Synapse, formally known as Azure Data Warehouse, is a platform used for processing multiple structured data formats.

You should not use Azure Polybase. Azure Polybase is used to transform the data into a format that is consumable by a relational database platform.

You should not use Azure Data Lake as the destination for the transformed data. In this situation, you would use it as the initial incoming destination for the incoming data feed. Azure Data Lake offers a platform that supports different and varying data types to be collated together for consumption by other Azure data services.

You should not use Azure Data Factory. Azure Data Factory can be used to bring sources of data to a central location for processing. Azure PolyBase can then be used to transform the data into a workable format for querying and analysis.

References

[What is Azure Data Factory?](#)

[Data Lake](#)

[What is PolyBase?](#)

[Azure Synapse Analytics](#)

Question 31.4

The company wants to automate the copying of data feeds to Azure to gain more efficiency and bring the data to their users faster.

What Azure service or component should the company use to move the data from on-premises to Azure?

Azure PolyBase

Azure DataFactory

Azure Synapse

Azure Data Lake

Explanation

You should use Azure Data Factory. Azure Data Factory provides a number of methods for copying data to repositories into locations where that same data can be consumed by other Azure Services.

You should not use Azure Synapse. Azure Synapse, formally known as Azure Data Warehouse, is a platform used for processing multiple structured data formats.

You should not use Azure PolyBase. Azure Polybase is used to transform the data into a format that is consumable by a relational database platform.

You should not use Azure Data Lake. Azure Data Lake offers a platform that supports different and varying data types to be collated together for consumption by other Azure data services.

References

What is Azure Data Factory?

Data Lake

What is Polybase?

Azure Synapse Analytics

Case Study

Complete the Case Study

- Solution Evaluation

Instructions

This case study contains a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

Note: You cannot go back or review questions of this type on the actual certification exam.

Question 35.1

You are a data scientist at CompanyB. The company has a strict policy of not allowing inbound network connectivity from the internet or allowing outbound connectivity over anything except HTTP/HTTPS. You want to use Azure DataBricks to transform your on-premises data and collaborate and share visualizations with partners in other companies.

You need to recommend a solution to copy the data to a location where Azure DataBricks can be used to process it.

Solution: Use a self-hosted Integration Runtime (IR).

Does this solution meet the goal?

Complete the Case Study

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3

No
Yes

Explanation

This solution meets the goal. A self-hosted Integration Runtime (IR) supports outbound connectivity over HTTP/HTTPS and can be used to move data in this scenario.

A self hosted Integrated runtime can support activities such as copying data between a data store in a private network and a cloud data store. Supported transformation activities include HDInsight MapReduce activity (Bring Your Own Cluster (BYOC)), HDInsight Spark activity (BYOC), HDInsight Streaming activity (BYOC), HDInsight Hive activity (BYOC) and HDInsight Pig activity (BYOC). Also supported are Machine Learning Batch Execution activity, Machine Learning Update Resource activities, Stored Procedure activity, Data Lake Analytics U-SQL activity, Custom activity (runs on Azure Batch), Lookup activity, and Get Metadata activity. For applications such as SAP, customer drivers will be required to connect to the data.

In addition to a self-hosted IR , there is the concept of a regular IR hosted entirely on Azure.

An Azure- only hosted integration runtime can run data flows in Azure, copy activity between cloud data stores, and dispatch the following transform activities in a public network: Databricks Notebook/ Jar/ Python activity, HDInsight Hive activity, HDInsight Pig activity, HDInsight MapReduce activity, HDInsight Spark activity, HDInsight Streaming activity, Machine Learning Batch Execution activity, Machine Learning Update Resource activities, Stored Procedure activity, Data Lake Analytics U-SQL activity, .NET custom activity, Web activity, Lookup activity, and Get Metadata activity.

The Azure IR element is the component used to connect the data source endpoint to enable Data Factory activities to access the data.

Azure Data Factory is the platform of choice to solve data scenarios in which you need to move, copy, and/or transform data during processing. It is driven by cloud-based workflows to collect and drive movement and transformation.

References

[Copy Activity in Azure Data Factory](#)

[Integration runtime in Azure Data Factory](#)

You are a data scientist at CompanyB. The company has a strict policy of not allowing inbound network connectivity from the internet or allowing outbound connectivity over anything except HTTP/HTTPS. You want to use Azure DataBricks to transform your on-premises data and collaborate and share visualizations with partners in other companies.

You need to recommend a solution to copy the data to a location where Azure DataBricks can be used to process it.

Solution: Use an Azure-SSIS Integration Runtime.

Does this solution meet the goal?

Complete the Case Study

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3

No
Yes

Explanation

This solution does not meet the goal. An Azure-SSIS Integration Runtime supports connectivity from publicly accessible endpoints only. Since access to the corporate network is limited in this case, this option cannot work.

References

Integration runtime in Azure Data Factory

You are a data scientist at CompanyB. The company has a strict policy of not allowing inbound network connectivity from the internet or allowing outbound connectivity over anything except HTTP/HTTPS. You want to use Azure DataBricks to transform your on-premises data and collaborate and share visualizations with partners in other companies.

You need to recommend a solution to copy the data to a location where Azure DataBricks can be used to process it.

Solution: Use an Azure Integration Runtime.

Does this solution meet the goal?

Complete the Case Study

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3

Yes
No

Explanation

This solution does not meet the goal. An Azure Integration Runtime can only operate on datasets and infrastructure that are hosted in Azure. This precludes a scenario like this, in which on-premises data sets are involved.

References

[Integration runtime in Azure Data Factory](#)

Question36

Question37

Question38 Design non-relational cloud data stores
Case Study

Complete the Case Study

- Overview

You are a data scientist for CompanyC, which is a very large retail company.

You are asked to help manage reordering the most popular products. The company allows the managers of franchised locations to use their own delimited text formats for submitting sales. At a minimum, they must include the current date, item stock keeping unit (SKU) number, price, and sales quantity.

CompanyC has had issues with manual ordering processes and wants you to assist with

sales forecasting, inventory management, and stock control.

Overall, the plan is to reduce the net Cost Of Goods sold (COGS) over time. Ten years of historical sales data with over a billion records in total must be viewable graphically via dashboards.

What solution should you choose to store the initial intake of data?

Complete the Case Study

- Overview
- Question 1
- Question 2
- Question 3
- Question 4

Azure Synapse Analytics

Azure Data Lake

Power BI

Azure PolyBase

Explanation

You should use Azure Data Lake to store multiple data format types for consumption. Azure Data Lake can consume different data formats up to petaByte scale and contains trillions of objects.

You should not use Azure PolyBase. Azure PolyBase is a service that can take existing large disparate data sources and import them into a structured format for regular T-SQL queries to be built.

You should not use Azure Synapse Analytics. While Azure Synapse is a large scale data warehousing solution, it does not support importing disparate data sources without transforming them into a supported format.

You should not use Power BI. Power BI is a collection of services that can pull from different data sources and create visual displays and dashboards to aid visualization.

References

Data Lake

What solution should you use to convert the different incoming data formats to a normalized relational format?

Complete the Case Study

- Overview
- Question 1
- Question 2
- Question 3
- Question 4

Azure Synapse Analytics

Azure PolyBase

Azure Data Lake

Power BI

Explanation

You should use Azure PolyBase. Azure PolyBase is a service that can take existing large disparate data sources and import them into a structured format for regular T-SQL queries to be built.

You should not use Azure Data Lake. Azure Data Lake can consume different data formats up to petaByte scale and contains trillions of objects, but it is not used to transform data.

You should not use Azure Synapse Analytics. While Azure Synapse is a large scale data warehousing solution, it does not support importing disparate data sources without transforming them into a supported format and is not used to transform data.

You should not use Power BI. Power BI is a collection of services that can pull from different data sources and create visual displays and dashboards to aid visualization, but it is not used to transform data.

References

[Data Lake](#)

[What is Polybase?](#)

[Azure Synapse Analytics](#)

[What is Power BI?](#)

Question 38.3

What platform should you use to support historical data queries?

Complete the Case Study

- Overview
- Question 1
- Question 2
- Question 3
- Question 4

Power BI
Azure Data Lake

Azure Synapse Analytics

Azure PolyBase

Explanation

You should use Azure Synapse Analytics. Azure Synapse Analytics is a large scale data warehousing solution that is used for large scale queries and Big Data analytics.

You should not use Azure Data Lake. Azure Data Lake can consume different data formats up to petaByte scale and contains trillions of objects, but it is not an analytics solution. Azure Data Lake Analytics is the related analytics solution.

You should not use Azure PolyBase. Azure PolyBase is a service that can take existing large disparate data sources and import them into a structured format for regular T-SQL queries to be built.

You should not use Power BI. Power BI is a collection of services that can pull from different data sources and create visual displays and dashboards to aid visualization.

References

[Data Lake](#)

[What is Polybase?](#)

[Azure Synapse Analytics](#)

[What is Power BI?](#)

Question 38.4

What solution should you use for visualizing the data by using dashboards?

Complete the Case Study

- Overview
- Question 1
- Question 2
- Question 3
- Question 4

<mark>Power BI</mark>

Azure Data Lake

Azure PolyBase

Azure Synapse Analytics

Explanation

You should use Power BI. Power BI is a collection of services that can pull from different data sources and create visual displays and dashboards to aid visualization.

You should not use Azure Synapse Analytics. Azure Synapse Analytics is a large scale data warehousing solution that is used for large scale queries and Big Data analytics. It would act as a source for the Power BI dashboards.

You should not use Azure Data Lake. Azure Data Lake can consume different data formats up to petaByte scale and contains trillions of objects, but it is not a dashboard solution.

You should not use Azure PolyBase. Azure PolyBase is a service that can take existing

large disparate data sources and import them into a structured format for regular T-SQL queries to be built.

References

Case Study

Complete the Case Study

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3

Instructions

This case study contains a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

Note: You cannot go back or review questions of this type on the actual certification exam.

You are a data architect for your company. You plan to import data into Azure SQL Data Warehouse. You want to import a dimension table named Manufacturer that has about 200 rows. The size of the table is about 100 kilobytes (KB). 90 percent of the manufacturers are located in the same postal code. This table will be used in joins for most queries.

You need to recommend the partitioning strategy for the table.

Solution: You use a replicated table.

Does this solution meet the goal?

Complete the Case Study

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3

Yes
No

Explanation

This solution meets the goal. A replicated table is recommended for small dimension tables. This method copies the table across all compute nodes. There are 60 compute nodes to which tables are distributed in SQL Data Warehouse. By copying the Manufacturer table to all compute nodes, you can increase the performance of queries that join that table with other tables.

References

Designing tables in Azure SQL Data Warehouse

You are a data architect for your company. You plan to import data into Azure SQL Data Warehouse. You want to import a dimension table named Manufacturer that has about 200 rows. The size of the table is about 100 kilobytes (KB). 90 percent of the manufacturers are located in the same postal code. This table will be used in joins for most queries.

You need to recommend the partitioning strategy for the table.

Solution: You use a hash-distributed table with the column that represents the postal code.

Does this solution meet the goal?

Complete the Case Study

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3

<mark>No</mark>
Yes

Explanation

This solution does not meet the goal. Hash distribution copies table rows across distributions by using a hashing function on a column. Rows that contain the same value of a column are copied to the same compute node. When using hash distribution, it is recommended to use a column that contains many distinct values. In this scenario, 90 percent of the manufacturers (180) are in the same postal code. This means that there are at most 21 distinct postal codes.

You should use a replicated table. A replicated table is recommended for small dimension tables. This method copies the table across all compute nodes.

References

[Designing tables in Azure SQL Data Warehouse](#)

Question 42.3

You are a data architect for your company. You plan to import data into Azure SQL Data Warehouse. You want to import a dimension table named Manufacturer that has about 200 rows. The size of the table is about 100 kilobytes (KB). 90 percent of the manufacturers are located in the same postal code. This table will be used in joins for most queries.

You need to recommend the partitioning strategy for the table.

Solution: You use round-robin distribution.

Does this solution meet the goal?

- Solution Evaluation
    - Question 1
    - Question 2
    - Question 3

No
Yes

Explanation

This solution does not meet the goal. Round-robin distribution copies the 200 rows evenly across all distributions. This will not improve performance because queries might have to search all of the distributions to find the desired data.

You should use a replicated table. A replicated table is recommended for small dimension tables. This method copies the table across all compute nodes.

References

Designing tables in Azure SQL Data Warehouse

Question43

Question44

Question45 Design relational cloud data stores
Case Study
Complete the Case Study

- Overview

Company1 partners with other organizations and government entities to develop, test, and deploy data collection and analysis solutions.

Company1 is working with a large metropolitan area to gather detailed statistics on both public transportation and private vehicular traffic. Immediate goals include improving public transit performance, improving customer satisfaction, and reducing management

expenses. Long-term goals include planning for a gradual change-over to self-driving vehicles.

Data is collected from sensors installed on buses and railway transport systems. Each public vehicle is given a unique identifying number for tracking purposes. Ground-level sensors are used at locations throughout the city to capture license plate numbers and vehicle images. All items receive a date/time stamp.

Much of the data is maintained in separate databases targeted at the data contained. General and reference information about buses and transit trains is partially denormalized, with information grouped into a column family and references by identifying number.

- Data Collection and Analysis

Company1 has not determined which data will be required by the final solution and is collecting more data than will probably be needed. A large amount of the data is being written to large Azure Tables maintained in a premium tier general-purpose v2 (GPv2) storage account. During initial development, the highest possible levels of reliability and immediate availability of this data are key concerns, including in case of regional failures.

Rider data is collected in an Azure SQL Database. The city wants to receive reports related to public transit includes ridership, time to load and unload, and so forth. Data is highly normalized. Individual file sizes, transaction query time, and reporting time for each transit vehicle should be minimized.

An Azure SQL Database is used as a reference database to support both real-time and batch processing activities. The data in some columns is proprietary to Company1 and considered confidential. Direct access to this data should be limited to applications accessing the data. Other database columns do not require the same protection.

- Technical Requirements

You have the following technical requirements:

- You need to optimize processing rider data by vehicle.
- Backups for rider data should be maintained for 90 days.
- You need to choose an appropriate storage type for general vehicle data.
- You need to ensure that the availability and reliability requirements for Azure Table data are met.

Question 45.1

You need to select a data store option best suited to general vehicle data for buses and transit trains.

What should you choose?

<mark>HBase in HDInsight</mark>

Azure SQL Database

Azure Storage blobs

Cosmos DB

Explanation

You should choose HBase in HDInsight. The storage requirements specify a column-family database. This is a column-data database that organizes data into rows and columns using a denormalized approach to data organization and storage. Data is typically stored in key order.

You should not choose Cosmos DB as your storage solution. Cosmos DB is a globally distributed multi-model database that is suited to various storage scenarios, including:

- Key/value stores in which each data value is associated with a unique key, effectively creating a large hash table
- Document databases with documents made up of a collection of named fields and data and supporting a variety of encoding methods
- Graph databases with data represented using nodes and relationships in graph structures

Cosmos DB does not support the specific data format required for storing and processing the general vehicle data.

You should not choose Azure Storage blobs. Blobs are designed for object-based storage for text- and binary-based data. It is optimized for storage and retrieval of large binary objects such as image files and audio streams.

You should not choose Azure SQL Database as your storage solution. An Azure SQL

Database is used to store highly normalized data in related tables. The denormalized nature of the data makes this an inappropriate solution.

References

[Choose the right data store](#)

[What is Apache HBase in Azure HDInsight](#)

Question 45.2

You need to choose an Azure Table storage redundancy option that meets Company1's requirements.

What should you choose?

Complete the Case Study

- Overview
  - Data Collection and Analysis
    - Technical Requirements
      - Question 1
      - Question 2
      - Question 3
      - Question 4

locally redundant storage (LRS)
read-access geo-redundant storage (RA-GRS)

zone-redundant storage (ZRS)

read-access geo-zone-redundant storage (RA-GZRS)

Explanation

You should choose RA-GZRS. This provides the highest possible level of reliability, recoverability, and availability. GPv2 is the only type of storage account that supports RA-GZRS. RA-GZRS combines features of zone-redundant storage (ZRS) for high availability and geo-redundant storage (GRS) to protect in case of regional failure. If a failure occurs, secondary data copies are available for read access even before failing over to a different region.

You should not choose RA-GRS. This does not provide the same levels of high availability as RA-GZRS.

You should not choose LRS. Data is redundant across different servers within a datacenter, but this does not protect data in case of a datacenter failure.

You should not choose ZRS. This provides high-availability but does not protect against regional failures.

References

[Azure storage account overview](#)

[Azure Table storage overview](#)

[Build highly available Azure Storage applications with geo-zone-redundant storage (GZRS) (preview)](#)

[Azure Storage redundancy](#)

Question 45.3

You need to ensure that rider data is structured to meet file size and processing requirements.

What should you use?

Complete the Case Study

- Overview
  - Data Collection and Analysis
    - Technical Requirements
      - Question 1
      - Question 2
      - Question 3
      - Question 4

Active geo-replication
Memory-optimized clustered columnstore indexing

Nonclustered columnstore indexing

Sharding

Explanation

You should use sharding. Sharding is a method of scaling out the data from a database across multiple databases through horizontal partitioning. Sharding by vehicle identification will help to minimize file sizes. Because the data volume is minimized, transaction processing is optimized.

You should not use nonclustered columnstore indexing or memory-optimized clustered columnstore indexing. These are both methods that are used to optimize performance when a database is used for both online transaction processing (OLTP) and online analytic processing (OLAP). Memory-optimized clustered columnstore indexing is specifically used when optimizing for concurrent OLTP and OLAP processing.

You should not use active geo-replication. Active geo-replication lets you to create readable secondary databases for individual databases in either the same or a different datacenter. It is designed primarily as business continuity solution for quick disaster recovery.

References

Scaling out with Azure SQL Database

Creating and using active geo-replication

What is database sharding?

Optimize performance by using In-Memory technologies in SQL Database

Question 45.4

You need to ensure that retention requirements are met for the rider data. Management effort to implement the solution should be minimized.

What should you recommend?

Complete the Case Study

- Overview
  - Data Collection and Analysis
    - Technical Requirements
      - Question 1
      - Question 2

Geo-redundant storage (GRS)

<mark>Long-term backup retention (LTR)</mark>

AzCopy

Automated backups

Explanation

You should recommend LTR. LTR lets you archive full backups for up to 10 years. You can configure how often backups are written to LTR. Because LTR maintains copies of full backups only, it does not support point-in-time recovery.

You should not recommend automated backups. By itself, using automated backups does not provide the solution needed. Automated backups are maintained for no more than 35 days.

You should not recommend GRS as a way to archive backups. Azure SQL backups use read-access geo-redundant storage (RA-GRS) by default. This is a way to provide for disaster recovery in case of a regional failure and is not a long-term archiving strategy.

You should not recommend AzCopy. AzCopy is used to copy blobs or files to or from a storage account. It is not a backup archive method.

References

[Manage Azure SQL Database long-term backup retention](#)

[Get started with AzCopy](#)

[Store Azure SQL Database backups for up to 10 years](#)

[Automated backups](#)

Case Study

Complete the Case Study

• Background

Company A develops inventory management software. Its flagship product allows employees to scan product barcodes in stores and have that data delivered to a central repository. A website allows supervisors to view the data sent to the repository.

• Inventory Data

Inventory data consists of the following fields:

- UPC
- Name
- Description
- Quantity
- Store ID
- Aisle Number
- Price
- Expiration Date

• Technical Solution

Inventory data is currently stored in two Microsoft SQL Server databases. One database resides in California, and the other database resides in New York. Over 200 terabytes (TB) of total data is stored across the two databases. The scanners submit inventory data to an application server over HTTPS. A service on the application server then analyzes the data and sends it to the databases.

The new solution must allow processing of the inventory data in batches every hour. After the data is processed, it must be kept for at least two years. It must be stored in such a way that parallel queries can be run against the data.

Business stakeholders must be able to graphically visualize the data without writing any code. Data engineers must be able to graphically visualize the data by using Python.

- Data Engineering Requirements

The data engineers at your company are familiar with C#, Python and SQL. Any recommended solution must take advantage of their existing skills.

You need to design a solution for storing the initial inventory data.

Which resource should you use?

Complete the Case Study

- Background
- Inventory Data
- Technical Solution
- Data Engineering Requirements
  - Question 1
  - Question 2
  - Question 3
  - Question 4
  - Question 5

Power BI
Azure SQL Data Warehouse

Azure Data Lake

Event Hub

Explanation

You should use Azure Data Lake. This is a big data store that supports data of any type and size. It supports receiving data in batches.

You should not choose Event Hub. Event Hub supports streaming millions of events per second. You can submit events to Event Hub by using HTTPS or AMQP.

You should not use Azure SQL Data Warehouse. This is a big data Enterprise Data Warehouse (EDW) that processes data in parallel across 60 compute nodes.

You should not use Power BI. Power BI allows you to turn data into visual insights.

References

What is Azure Data Lake Storage Gen1?

Question 49.2

You need to design a solution for analyzing the inventory data by using C# user-defined functions (UDFs) when it arrives.

Which resource should you use?

Complete the Case Study

- Background
  - Inventory Data
- Technical Solution
  - Data Engineering Requirements
    - Question 1
    - Question 2
    - Question 3
    - Question 4
    - Question 5

Power BI

Azure Data Lake Analytics

IoT Hub

Azure Databricks

Explanation

You should use Azure Data Lake Analytics. This is a batch processing service that integrates with Azure Data Lake. It allows developers to write code by using U-SQL. U-SQL is a language that combines declarative SQL with imperative C#.

You should not use Power BI. Power BI allows you to turn data into visual insights.

You should not use IoT Hub. This is a bi-directional central repository that allows data to be sent between an IoT device and Azure.

You should not use Azure Databricks. This is a platform based on Apache Spark clusters

that allows you to create big data workflows.

References

Question 49.3

You need to design a solution for storing the data long-term.

Which resource should you use?

Complete the Case Study

- Background
- Inventory Data
- Technical Solution
- Data Engineering Requirements
  - Question 1
  - Question 2
  - Question 3
  - Question 4
  - Question 5

Azure SQL Data Warehouse

Azure SQL Database

Azure Data Factory

Azure Databricks

Explanation

You should use Azure SQL Data Warehouse. This is a big data Enterprise Data Warehouse (EDW) that processes data in parallel across 60 compute nodes.

You should not use Azure Data Factory. This is a cloud service that allows you to

extract, transform and load big data. It only stores pipeline data for 45 days.

You should not use Azure Databricks. This is a platform based on Apache Spark clusters that allows you to create big data workflows.

You should not use Azure SQL Database. This is a Platform-as-a-Service (PaaS) offering of SQL Server. Azure SQL Database does not allow parallel processing of big data.

References

[What is Azure Databricks?](#)

[What is the Azure SQL Database service?](#)

[Programmatically monitor an Azure data factory](#)

Question 49.4

You need to design a solution to allow data engineers to visualize the data.

What should you use?

Complete the Case Study

- Background
- Inventory Data
- Technical Solution
- Data Engineering Requirements
  - Question 1
  - Question 2
  - Question 3
  - Question 4
  - Question 5

Azure Databricks

Stream Analytics

Power BI

Azure Data Lake

Explanation

You should use Azure Databricks. This is a platform based on Apache Spark clusters that allows you to create big data workflows. Data engineers can visualize data by using Python. Databricks also supports Scala, R, and SQL.

You should not use Power BI. Power BI allows you to turn data into visual insights. It allows business stakeholders to perform this task without writing any code.

You should not use Stream Analytics. This is a real-time event processing service that processes events simultaneously from different sources.

You should not use Azure Data Lake. This is a repository for big data workloads that supports data of any type and size.

References

[What is Power BI?](#)

[What is Azure Stream Analytics?](#)

[What is Azure Data Lake Storage Gen1?](#)

[What is Azure Databricks?](#)

Question 49.5

You need to design a solution to allow business stakeholders to visualize the data.

What should you use?

Complete the Case Study

- Background
- Inventory Data
- Technical Solution
- Data Engineering Requirements
  - Question 1
  - Question 2
  - Question 3
  - Question 4
  - Question 5

Power BI

Stream Analytics

Azure Databricks

Azure Data Lake

Explanation

You should use Power BI. Power BI allows you to turn data into visual insights. It allows business stakeholders to perform this task without writing any code.

You should not use Stream Analytics. This is a real-time event processing service that processes events simultaneously from different sources.

You should not use Azure Data Lake. This is a repository for big data workloads that supports data of any type and size.

You should not use Azure Databricks. This is a platform based on Apache Spark clusters that allows you to create big data workflows.

References

[What is Power BI?](#)

[What is Azure Stream Analytics?](#)

[What is Azure Data Lake Storage Gen1?](#)

[What is Azure Databricks?](#)

Question50

Question51

Question52

Question53

Question54

Case Study

Complete the Case Study

You are the database administrator for your company. The company is a reseller of internet information about financial and distribution markets.

The company is planning an investment into Azure and is looking for the right type of data platform to be able to consume data from all the various incoming data feeds. The feeds consist of Comma Separated Values (CSV) files and Parquet format files.

The company wants to find the simplest way to ingest the data, transform it into a relational format, and make it readily available for consumers to view as well as ensuring efficient query latency.

Question 54.1

You need to choose the data platform Azure to use as the initial destination of the incoming data feeds.

Which platform should you use?

Complete the Case Study

Azure Synapse Analytics
Azure Data Factory

Azure PolyBase

Azure Data Lake

Explanation

You should use Azure Data Lake to store multiple data format types for consumption. Azure Data Lake can consume different data formats up to PetaByte scale and can contain trillions of objects.

You should not use Azure PolyBase. Azure PolyBase is a feature that can take existing large disparate data sources and import them into a structured format for regular T-SQL queries to be built.

You should not use Azure Synapse Analytics. While Azure Synapse is a large scale data warehousing solution, it does not support importing disparate data sources without

transforming them into a supported format.

You should not use Azure Data Factory. Azure Data Factory provides multiple ways of transferring data and transforming data at scale, but in this case, Azure Data Lake is the better option for the data destination due its ability to consume and store multiple different source data types.

References

Data Lake

What is PolyBase?

Azure Synapse Analytics

What is Azure Data Factory?

Question 54.2

You need to choose the solution to transform the data from the various formats and import the data into a relational platform. You must select the simplest implementation path.

Which solution should you use?

Complete the Case Study

- Overview
- Question 1
- Question 2
- Question 3
- Question 4

<mark>PolyBase</mark>

Azure Synapse Analytics

Azure Data Factory

Azure Data Lake

Explanation

You should use Azure PolyBase. Azure PolyBase is a service that can take existing large disparate data sources and import them into a structured format for regular T-SQL queries to be built.

You should not use Azure Data Lake. Azure Data Lake can consume different data formats up to PetaByte scale and can contain trillions of objects, but it is not used as a transformation service for disparate sources of data.

You should not use Azure Synapse Analytics. While Azure Synapse is a large scale data warehousing solution, it does not support importing disparate data sources without transforming them into a supported format for Synapse to work on. In this case Azure PolyBase is the preferred option because Synapse is a non-relational datastore.

You should not use Azure Data Factory. Azure Data Factory provides multiple ways of transferring data and transforming data at scale, but in this case, Azure PolyBase is the better option for the data transformation to a structured data schema. Azure Data Factory can consume the Azure PolyBase service, but that would require additional work to complete the scenario. The question calls for the simplest implementation pathway.

References

Create External File Format (Transact-SQL)

Data Lake

What is PolyBase?

Azure Synapse Analytics

What is Azure Data Factory?

Question 54.3

The company wants to automate the copying of data feeds to Azure to gain more efficiency and bring the data to their users faster.

You need to choose the Azure solution to move the data from on-premises to Azure.

Which solution should you use?

Complete the Case Study

Azure Data Factory

Azure Synapse Alalytics

Azure Data Lake

Azure Polybase

Explanation

You should use Azure Data Factory. Azure Data Factory provides multiple ways of transferring data and transforming data at scale between cloud sources and on-premises sources.

You should not use Azure Data Lake for this high efficiency data transfer. Azure Data Lake can consume different data formats up to PetaByte scale and can contain trillions of objects, but it does not perform gathering or transformation on those objects.

You should not use Azure PolyBase. Azure PolyBase is a service that can take existing large disparate data sources and import them into a structured format for regular T-SQL queries to be built. It is not used for highly efficient data transfer.

You should not use Azure Synapse Analytics. While Azure Synapse is a large scale data warehousing solution, it does not support importing disparate data sources without transforming them into a supported format.

References

[What is Azure Data Factory?](#)

[Data Lake](#)

[What is PolyBase?](#)

[Azure Synapse Analytics](#)

For the final step in the processing of the data, you need to select the right destination Azure repository for the transformed data.

Which solution should you use for the final repository of the data?

- Overview
- Question 1
- Question 2
- Question 3
- Question 4

Azure Polybase
Azure Data Lake

Azure Data Factory

Azure Synapse Analytics

Explanation

You should use Azure Synapse Analytics. Azure Synapse Analytics is a large scale data warehousing solution and can consume data from multiple sources. It also ensures efficient data query latency and is a solution that can be consumed by multiple users across any business.

You should not use Azure Data Lake as the destination data platform. Azure Data Lake can consume different data formats up to PetaByte scale and can contain trillions of objects, but it does not provide the structured analytics required.

You should not use Azure PolyBase. Azure PolyBase is a service that can take existing large disparate data sources and import them into a structured format for regular T-SQL queries to be built.

You should not use Azure Data Factory. Azure Data Factory provides multiple ways of transferring data and transforming data at scale, but in this case, a structured analytics platform is needed.

References

Azure Synapse Analytics

Case Study

Complete the Case Study

- Solution Evaluation

Instructions

This case study contains a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

Note: You cannot go back or review questions of this type on the actual certification exam.

Question 58.1

You are a data architect for a polling company. Each pollster submits data over Advanced Message Queuing Protocol (AMQP). You want to retrieve the data in real time so that you can extract relevant information, transform it, and then send it to Power BI.

You need to implement the solution.

Solution: You do the following:

- Create an Event Hub instance.
- Create a Stream Analytics job that uses a query to extract data.

Does this solution meet the goal?
Complete the Case Study

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3
  - Question 4

Yes
No

Explanation

This solution meets the goal. Event Hubs is an Azure resource that allows you to stream millions of events per second. It accepts streaming data over HTTPS and AMQP. A Stream Analytics job can read data from Event Hubs, transform it, and store the transformed data in a variety of output data sources, including Power BI.

References

Stream Analytics and Power BI: A real-time analytics dashboard for streaming data

Question 58.2

You are a data architect for a polling company. Each pollster submits data over Advanced Message Queuing Protocol (AMQP). You want to retrieve the data in real time so that you can extract relevant information, transform it, and then send it to Power BI.

You need to implement the solution.

Solution: You do the following:

- Create an IoT Hub instance.
- Create a Stream Analytics job that uses a query to extract data.

Does this solution meet the goal?
Complete the Case Study

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3

<mark>Yes</mark>
No

Explanation

This solution meets the goal. IoT Hub is an Azure resource that allows you to stream big data to the cloud. It supports per-device provisioning and device identity. It accepts streaming data over HTTPS, AMQP, and Message Queue Telemetry Transport (MQTT). A Stream Analytics job can read data from IoT Hub, transform it, and store the transformed data in a variety of output data sources, including Power BI.

References

[Get started with Azure Stream Analytics to process data from IoT devices](#)

[Stream Analytics and Power BI: A real-time analytics dashboard for streaming data](#)

Question 58.3

You are a data architect for a polling company. Each pollster submits data over Advanced Message Queuing Protocol (AMQP). You want to retrieve the data in real time so that you can extract relevant information, transform it, and then send it to Power BI.

You need to implement the solution.

Solution: You do the following:

- Create an Azure Databricks instance.
- Create an Azure Automation runbook that extracts and queries data from Databricks.

Does this solution meet the goal?
Complete the Case Study

Yes

Explanation

This solution does not meet the goal. Azure Databricks is an analytics platform that uses Spark clusters. Notebooks contain code that run on clusters of nodes. Automation runbooks execute code written in PowerShell and Python. The problem with this solution is that sensor data cannot be streamed to either of these resources.

The correct solution should use Event Hub or IoT Hub and Stream Analytics. Event Hub and IoT Hub ingest streaming data. Streaming Analytics allows you to write code to analyze the data.

References

Azure Automation runbook types

Connecting IoT Devices to Azure: IoT Hub and Event Hubs

Stream Analytics and Power BI: A real-time analytics dashboard for streaming data

What is Azure Databricks?

Question 58.4

You are a data architect for a polling company. Each pollster submits data over Advanced Message Queuing Protocol (AMQP). You want to retrieve the data in real time so that you can extract relevant information, transform it, and then send it to Power BI.

You need to implement the solution.

Solution: You do the following:

- Create an Azure Relay service.
- Create an Azure Function app that extracts and queries data from Azure Relay.

Does this solution meet the goal?

Complete the Case Study

- Solution Evaluation
  - Question 1

Yes
No

Explanation

This solution does not meet the goal. Azure Relay allows client applications to connect to services hosted on a private network over the internet. An Azure Function app contains one or more functions that are exposed over HTTP. These functions can be invoked by triggers or on a schedule. Neither Azure Relay nor Azure Function apps accept messages over AMPQ.

The correct solution should use Event Hub or IoT Hub and Stream Analytics. Event Hub and IoT Hub ingest streaming data. Streaming Analytics allows you to write code to analyze the data.

References

An introduction to Azure Functions

Connecting IoT Devices to Azure: IoT Hub and Event Hubs

Stream Analytics and Power BI: A real-time analytics dashboard for streaming data

What is Azure Relay?

Question59

Question60

Question61

Question62

Case Study

Complete the Case Study

- Solution Evaluation

Instructions

This case study contains a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

Note: You cannot go back or review questions of this type on the actual certification exam.

Question 62.1

You are designing a real-time processing solution using Azure Stream Analytics. You need to ingest data from IoT sensors installed in driverless vehicles. You also need to support a 200 MB reference data input to correlate related static values.

Solution: You use Azure Event Hubs for stream data input and Azure SQL Database for reference input.

Does this solution meet the goal?

Complete the Case Study

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3

Yes
No

Explanation

This solution meets the goal. You can use Azure Event Hubs, IoT hub, and Azure Blob storage for streaming data input. Because you want real-time processing for a large amount of incoming data, Event Hubs or IoT Hub are your best choices in this scenario. Azure Stream Analytics supports Azure SQL Database or Azure Blob storage for reference data input.

References

Understand inputs for Azure Stream Analytics

Choose a real-time analytics and streaming processing technology on Azure

What is Azure Stream Analytics?

You are designing a real-time processing solution using Azure Stream Analytics. You need to ingest data from IoT sensors installed in driverless vehicles. You also need to support a 200 MB reference data input to correlate related static values.

Solution: You use Azure IoT Hub for stream data input and Azure Blob storage for reference input.

Does this solution meet the goal?

Complete the Case Study

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3

No
Yes

Explanation

This solution meets the goal. You can use Azure Event Hubs, IoT hub, and Azure Blob storage for streaming data input. Because you want real-time processing for a large amount of incoming data, Event Hubs or IoT Hub are your best choices in this scenario. Azure Stream Analytics supports Azure SQL Database or Azure Blob storage for reference data input.

References

Understand inputs for Azure Stream Analytics

Choose a real-time analytics and streaming processing technology on Azure

What is Azure Stream Analytics?

Question 62.3

You are designing a real-time processing solution using Azure Stream Analytics. You need to ingest data from IoT sensors installed in driverless vehicles. You also need to support a 200 MB reference data input to correlate related static values.

Solution: You use Azure Event Hubs for stream data input and Azure Cosmos DB for reference input.

Does this solution meet the goal?

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3

No
Yes

Explanation

This solution does not meet the goal. You can use Azure Event Hubs for data ingestion, but Azure Streaming Analytics does not support using Azure Cosmos DB for reference input.

You can use Azure Event Hubs, IoT hub, and Azure Blob storage for streaming data input. Because you want real-time processing for a large amount of incoming data, Event Hubs or IoT Hub are your best choices in this scenario. Azure Stream Analytics supports Azure SQL Database or Azure Blob storage for reference data input.

References

Understand inputs for Azure Stream Analytics

Choose a real-time analytics and streaming processing technology on Azure

What is Azure Stream Analytics?

Question63

Question64

Question65
Case Study
Complete the Case Study

Instructions

This case study contains a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

Note: You cannot go back or review questions of this type on the actual certification exam.

Question 65.1

You create a blob storage account. You must be able to administer the storage account via the REST API.

You need to provide secure access to the storage account.

Solution: You create a service shared access signature (SAS).

Does this solution meet the goal?

Complete the Case Study

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3
  - Question 4

Yes
No

Explanation

This solution does not meet the goal. A service SAS allows access to a resource in one of the three storage services: table, blob, or file. It does not allow you to administer a storage account.

Instead, you should use a primary or secondary access key, which would allow you to administer the storage account remotely. When you first create a storage account, these two keys are generated automatically. Afterwards, you can choose to regenerate these keys.

References

[Manage storage account settings in the Azure portal](#)

[Grant limited access to Azure Storage resources using shared access signatures (SAS)](#)

Question 65.2

You create a blob storage account. You must be able to administer the storage account via the REST API.

You need to provide secure access to the storage account.

Solution: You create an account shared access signature (SAS).

Does this solution meet the goal?

Complete the Case Study

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3
  - Question 4

Yes
No

Explanation

This solution does not meet the goal. An account SAS allows access to one or more resources in a storage account. It does not allow you to administer a storage account.

Instead, you should use a primary or secondary access key, which would allow you to administer the storage account remotely. When you first create a storage account, these two keys are generated automatically. Afterwards, you can choose to regenerate these keys.

References

[Manage storage account settings in the Azure portal](#)

[Grant limited access to Azure Storage resources using shared access signatures (SAS)](#)

You create a blob storage account. You must be able to administer the storage account via the REST API.

You need to provide secure access to the storage account.

Solution: You use the primary access key.

Does this solution meet the goal?

Complete the Case Study

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3
  - Question 4

No
Yes

Explanation

This solution meets the goal. Both a primary access key and a secondary access key grant access to remotely administer a storage account. Two keys are provided so that they can be rotated regularly.

References

Manage storage account settings in the Azure portal

You create a blob storage account. You must be able to administer the storage account via the REST API.

You need to provide secure access to the storage account.

Solution: You use the secondary access key.

Does this solution meet the goal?

Complete the Case Study

Yes

No

Explanation

This solution meets the goal. Both a primary access key and a secondary access key grant access to remotely administer a storage account. Two keys are provided so that they can be rotated regularly.

References

[Manage storage account settings in the Azure portal](#)

Question66

Question67

Question68

Question69

Case Study

Complete the Case Study

Instructions

This case study contains a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

Note: You cannot go back or review questions of this type on the actual certification exam.

Question 69.1

Your company is developing an Azure data solution that uses Azure Blob Storage. A select set of Azure Active Directory (Azure AD) accounts should be provided access to storage during

the initial test period. Access should expire automatically after 30 days. You want to avoid using the storage account key to provide access. You must be able to manage access specifics through role-based access control (RBAC).

You need to configure authentication and access to the storage blob.

Solution: You request a user delegation key and configure the account level shared access signature (SAS).

Does this solution meet the goal?

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3

Yes
No

Explanation

The solution does not meet the requirements. You should not request a user delegation key and configure the account level SAS. Access through account level is based on the storage account key. You can request user delegation keys, but you would not be using them for this type of access. You can configure expiration, but RBAC is not supported.

You should request a user delegation key and configure user delegation SAS. Access is based on the user Azure AD account and a delegation key you request for the account. You can set an expiration date to automatically expire or revoke the delegation key to remove access. User delegation SAS supports fine-grained security management through RBAC.

References

Grant limited access to Azure Storage resources using shared access signatures (SAS)

Delegate access with a shared access signature

Create a user delegation SAS (preview)

Question 69.2

Your company is developing an Azure data solution that uses Azure Blob Storage. A select set of Azure Active Directory (Azure AD) accounts should be provided access to storage during the initial test period. Access should expire automatically after 30 days. You want to avoid using the storage account key to provide access. You must be able to manage access specifics through role-based access control (RBAC).

You need to configure authentication and access to the storage blob.

Solution: You request a user delegation key and configure shared key access.

Does this solution meet the goal?

Complete the Case Study

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3

Yes
No

Explanation

The solution does not meet the requirements. You should not request a user delegation key and configure shared key access. Shared key access is based on the storage account key. You cannot configure automatic account expiration, and there is no support for RBAC.

You should request a user delegation key and configure user delegation SAS. Access is based on the user Azure AD account and a delegation key you request for the account. You can set an expiration date to automatically expire or revoke the delegation key to remove access. User delegation SAS supports fine-grained security management through RBAC.

References

Your company is developing an Azure data solution that uses Azure Blob Storage. A select set of Azure Active Directory (Azure AD) accounts should be provided access to storage during the initial test period. Access should expire automatically after 30 days. You want to avoid using the storage account key to provide access. You must be able to manage access specifics through role-based access control (RBAC).

You need to configure authentication and access to the storage blob.

Solution: You request a user delegation key and configure the user delegation shared access signature (SAS).

Does this solution meet the goal?

Complete the Case Study

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3

<mark>Yes</mark>
No

Explanation

The solution meets the requirements. You should create a user delegation key and configure the user delegation SAS. Access is based on the user Azure AD account and a delegation key you request for the account. You can set an expiration date to automatically expire or revoke the delegation key to remove access. User delegation SAS supports fine-grained security management through RBAC.

References

Case Study

Complete the Case Study

- Solution Evaluation

Instructions

This case study contains a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

Note: You cannot go back or review questions of this type on the actual certification exam.

Question 72.1

You are designing an Azure SQL Database solution. You are designing a table named Sales that will contain sales records for the company.

A column named SalesRep will represent the sales representative associated with the sale. The SalesRep column will be of type SYSNAME. Only the sales representative associated with a sale should be able to view sales data.

You need to choose the security mechanism.

Solution: You implement row-level security (RLS).

Does this solution meet the goal?

Complete the Case Study

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3
  - Question 4

Yes
No

Explanation

This solution meets the goal. RLS allows you to define a security policy to determine when a row should be retrieved. A filter predicate filters rows that are retrieved during SELECT, UPDATE, and DELETE statements. A filter predicate is simply a function associated with a security policy.

References

[Row-Level Security](#)

Question 72.2

You are designing an Azure SQL Database solution. You are designing a table named Sales that will contain sales records for the company.

A column named SalesRep will represent the sales representative associated with the sale. The SalesRep column will be of type SYSNAME. Only the sales representative associated with a sale should be able to view sales data.

You need to choose the security mechanism.

Solution: You implement Transparent Data Encryption (TDE).

Does this solution meet the goal?

Complete the Case Study

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3
  - Question 4

Yes
No

Explanation

This solution does not meet the goal. TDE encrypts the entire database, including data files and transaction logs. This does not help in this scenario.

In this scenario, you want to implement row-level security (RLS), which allows you to create a security policy that specifies when rows should be retrieved.

References

Azure encryption overview

Row-Level Security

You are designing an Azure SQL Database solution. You are designing a table named Sales that will contain sales records for the company.

A column named SalesRep will represent the sales representative associated with the sale. The SalesRep column will be of type SYSNAME. Only the sales representative associated with a sale should be able to view sales data.

You need to choose the security mechanism.

Solution: You implement Always Encrypted.

Does this solution meet the goal?

Complete the Case Study

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3
  - Question 4

Yes
No

Explanation

This solution does not meet the goal. This technology allows you to encrypt and decrypt column-level data at the client. The data then remains encrypted in the database at the server. This does not help to choose which rows are returned for specific sales people. You must implement row-level security (RLS) to achieve that.

References

Row-Level Security

Always Encrypted (Database Engine)

Question 72.4

You are designing an Azure SQL Database solution. You are designing a table named Sales that will contain sales records for the company.

A column named SalesRep will represent the sales representative associated with the sale. The SalesRep column will be of type SYSNAME. Only the sales representative associated with a sale should be able to view sales data.

You need to choose the security mechanism.

Solution: You implement column-level encryption (CLE).

Does this solution meet the goal?

Complete the Case Study

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3
  - Question 4

Yes
No

Explanation

This solution does not meet the goal. CLE uses symmetric encryption to encrypt data directly in the database. This does not meet the need of ensuring that only specific

rows are returned for specific sales people. You should instead implement row-level security (RLS).

References

[Row-Level Security](#)

[Azure encryption overview](#)

Case Study

Complete the Case Study

- Solution Evaluation

Instructions

This case study contains a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

Note: You cannot go back or review questions of this type on the actual certification exam.

Question 76.1

Your company is developing an inventory data solution for specialty retail sales shops that are part of a partner organization. The application uses Azure Data Warehouse as its data store. The solution will perform sales and trend analysis and feed order point suggestions to the shops.

Shops will upload data from local on-premises storage to the data warehouse each week. Data corruption checks should run each time data is uploaded. The upload should be reversed if corruption is detected. You should be able to reverse the upload as quickly as possible. The process should not impact the performance of other analysis and reporting supported by the

data warehouse.

Solution: You create a user-defined restore point before uploading the data. You run the corruption check separately and delete the restore point after data corruption checks complete.

Does this solution meet the goal?

No
Yes

Explanation

This solution meets the goals. You should create a user-defined restore point before uploading the data. This gives you a way to restore the database in case of corrupted data and reverse the changes. You should run the corruption check separately and delete the restore point after data corruption checks complete. You can have up to 42 restore points.

References

Backup and restore in Azure SQL Data Warehouse

User-defined restore points

Question 76.2

Your company is developing an inventory data solution for specialty retail sales shops that are part of a partner organization. The application uses Azure Data Warehouse as its data store. The solution will perform sales and trend analysis and feed order point suggestions to the shops.

Shops will upload data from local on-premises storage to the data warehouse each week. Data corruption checks should run each time data is uploaded. The upload should be reversed if corruption is detected. You should be able to reverse the upload as quickly as possible. The

process should not impact the performance of other analysis and reporting supported by the data warehouse.

Solution: You configure database-level auditing in Azure SQL Data Warehouse and set retention to 7 days.

Does this solution meet the goal?

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3

Yes
No

Explanation

This solution does not meet the goals. You should not configure database-level auditing in Azure SQL Data Warehouse and set retention to 7 days. Auditing lets you:

- Support compliance-related activities
- Gather insights about database activity
- Identify suspected security violations

While this would not significantly impact performance, it does nothing to provide the rollback support needed.

Instead, you should create a user-defined restore point before uploading the data. This gives you a way to restore the database in case of corrupted data and reverse the changes. You should run the corruption check separately and delete the restore point after data corruption checks complete. You can have up to 42 restore points.

References

Backup and restore in Azure SQL Data Warehouse

User-defined restore points

Azure SQL Data Warehouse introduces Data Auditing

Your company is developing an inventory data solution for specialty retail sales shops that are part of a partner organization. The application uses Azure Data Warehouse as its data store. The solution will perform sales and trend analysis and feed order point suggestions to the shops.

Shops will upload data from local on-premises storage to the data warehouse each week. Data corruption checks should run each time data is uploaded. The upload should be reversed if corruption is detected. You should be able to reverse the upload as quickly as possible. The process should not impact the performance of other analysis and reporting supported by the data warehouse.

Solution: You create a stored procedure that performs an INSERT transaction for each record and then performs a corruption check for the inserted row. You COMMIT or ROLLBACK the transaction depending on the result of the check.

Does this solution meet the goal?

Complete the Case Study

- Solution Evaluation
  - Question 1
  - Question 2
  - Question 3

No
Yes

Explanation

This solution does not meet the goals. You should not create a stored procedure that performs an INSERT transaction for each record and then performs a corruption check for the inserted row. This additional processing could impact other activity in the data warehouse. Also, using COMMIT or ROLLBACK by transaction removes that row only, not the update.

You should create a user-defined restore point before uploading the data. This gives you a way to restore the database in case of data corruption and reverse the changes. You should run the corruption check separately and delete the restore point after data corruption checks complete. You can have up to 42 restore points.

References

[Backup and restore in Azure SQL Data Warehouse](#)

[User-defined restore points](#)

[Using transactions in SQL Data Warehouse](#)

Question77

Question78

Question79 Design non-relational cloud data stores

You are designing the storage requirements for your company's automatic process tracking system. IoT sensors scan components at various points in the automated manufacturing process and record the product number, serial number, time, and manufacturing line location. Several hundred pieces are scanned by each sensor each day.

Sensor data is written to a Cosmos DB account in a collection named Tracking. The insertion rate for tracking data must be maximized. Data should be partitioned to optimize organizing information by part type.

You need to configure the default consistency level and partition key property.

What configuration should you use? To answer, select the appropriate values from the drop-down menus.

Choose the correct options

Configuration setting	Value
Default consistency level	
Partition key	

[dropdown]

eventual

[dropdown]

Product number

Choose the correct options

Configuration setting	Value
Default consistency level	Eventual
Partition key	Product number

⟨ Previous ☆ Next ⟩

79 of 172

Explanation

You should configure the default consistency level as eventual and use the serial number as the partition key. This will minimize latency while loading data into the collection.

Cosmos DB supports the following consistency levels, from strongest consistency to most relaxed:

- Strong
- Bounded staleness
- Session
- Consistent prefix
- Eventual

Eventual consistency does not provide any ordering guarantee on reads, but it provides the highest throughput.

Partitioning based on product number helps to optimize reporting grouped by product type. The partition key is combined with an internally generated item ID value.

You should not partition by serial number. This would result in a separate partition for each item.

You should not partition by time or location. This would do nothing toward grouping items by type.

References

Consistency levels in Azure Cosmos DB

Consistency, availability, and performance tradeoffs

Partitioning in Azure Cosmos DB

Partitioning and horizontal scaling in Azure Cosmos DB

Overview of Azure Cosmos DB

Question80 Design non-relational cloud data stores

You are designing a database solution for an application under development. You plan to use Cosmos DB for data storage. The application requires graph database model support.

You need to choose the appropriate Cosmos DB API.

Which one should you choose?

Choose the correct answer

MongoDB API

Gremlin API

Cassandra API

SQL API

Explanation

> You should choose the Gremlin API. The Gremlin API is based on and is compatible with the Apache TinkerPop graph database standard. Gremlin uses the Gremlin query language. This is the most widely adopted graph query standard.
>
> Cosmos DB is a globally distributed, multi-model database service. As a multi-model database, it supports different data models through the use of Cosmos DB APIs. Supported APIs include:
>
> - Cassandra API - Column-family data model
> - Gremlin API - Graphic data model
> - MongoDB API - Document data model
> - SQL API - SQL-compatible document data model
> - Table API - Key-value pair data model

This lets you choose an appropriate data model that best meets your application requirements or when migrating existing applications.

References

Overview of Azure Cosmos DB

Frequently asked questions about different APIs in Azure Cosmos DB

Introduction to Azure Cosmos DB: Gremlin API

Question81 Design non-relational cloud data stores

You are asked to design a database solution to support a low-latency global application. You plan to use Cosmos DB as the Azure solution.

You need to choose a consistency level that supports the lowest possible Recovery Point Objective (RPO) and a Recovery Time Objective (RTO) of no more than 15 minutes for a multiple region outage.

Which consistency level should you use?

Bounded staleness

Consistent prefix

<mark>Strong</mark>

Session

Explanation

You should use the strong consistency level to ensure that you have the lowest possible Recovery Time Objective (RTO) and Recovery Point Objective (RPO). Strong consistency ensures that the data is fully committed to the database and is recoverable from the exact point of failure, so the RPO would be 0 seconds. Data availability post-failure is guaranteed to be less than 15 minutes with strong consistency, so the RTO of less than 15 minutes is achievable with this consistency level. The RPO is the maximum amount of data loss than can be tolerated in the event of a failover or a disaster. The RTO is the total amount of time that the system can be offline before access is available again.

You should not use bounded staleness consistency. Bounded staleness supports data availability of less than 15 minutes, but the recovery time is dependent on the number of versions of the record and the time between writes. It cannot be guaranteed to be 0 seconds.

You should not use session consistency. Depending on the master model, this has different recovery time values (0 seconds for single master, and less than 15 minutes for multi-master), so the RPO cannot be guaranteed to be zero seconds.

You should not use constant prefix consistency. Constant prefix consistency provides a guarantee that data will not be read out of order and will always be consistent with the order the data was written to the database even if not all the data has replicated to the replica instances. Like session consistency, it has different RTO values (0 seconds for single master and less than 15 minutes for multi-master), but the RTO cannot be guaranteed to be zero seconds.

References

[Consistency Levels in Azure Cosmos DB](#)

[Consistency, availability, and performance tradeoffs](#)

Question82 Design non-relational cloud data stores

You are designing your company's datastore retention policies. Items uploaded to the company's Cosmos DB data store are retained for only 24 hours unless they have a specific retention policy override set.

You need to plan the correct Time to Live policy for the database design. The design must require the least administrative intervention from the solution admins once it is deployed.

What policy should you include in the design?

Choose the correct answer

Set the DefaultTimeToLive to 24.

Set the TimeToLive to 86400.

Set the TimeToLive to 24.

Set the DefaultTimeToLive to -1.

Set the DefaultTimeToLive to 86400.

Explanation

You should incorporate a DefaultTimeToLive of 86400. By setting the DefaultTimeToLive to 86400 seconds, you are applying a 24-hour policy to the whole container. This can be overridden by setting an explicit TTL value on an individual item.

You should not set DefaultTimeToLive to 24. This would apply a 24-second expiration to the items in the container.

You should not use a TimeToLive value of 86400. This would apply a 24-hour expiration at the item level in the container, but this value is not applied if no DefaultTimeToLive value exists at the container level. It would also require significantly more administrative effort because it would need to be set for all items.

You should not use a TimeToLive value of 24. This would apply a 24-second expiration at the item level in the container, but this value is not applied if no DefaultTimeToLive value exists at the container level. It would also require significantly more administrative effort because it would need to be set for all items.

You should not set DefaultTimeToLive to -1. This would disable container-lever default values and instead require the solution users to apply an item-level policy to each and every item. This breaks the lowest administrative requirement of the design.

References

[Configure time to live in Azure Cosmos DB](#)

[Time To Live (TTL) In Azure Cosmos DB](#)

Question83 Design non-relational cloud data stores

You are asked to design a solution that uses a trigger to read blob storage contents and writes them to a new Cosmos DB document overnight. The items are uploaded to blob storage during the daily working hours of 8 A.M. to 9 P.M.

You need to choose the trigger type that requires the least custom functionality to meet the requirements.

What Azure trigger type should you choose?

Choose the correct answer

<mark>Timer</mark>

Queue

Event Grid

HTTP

Explanation

You should use a Timer trigger. A Timer trigger runs an Azure function as a scheduled task and has input binding to Azure Blob store and Cosmos DB. It can be scheduled to run any time of the day and can create Cosmos DB documents from blob store contents.

You should not use an Event Grid trigger. Event Grid triggers are used to respond to

events and send notifications. A typical use is that when an item is uploaded to the blob storage account, the Event Grid can detect it and send a notification or trigger additional custom code. Event Grid triggers respond when the event fires and not on a schedule, so events would fire as the items are uploaded and not as scheduled.

You should not use a Queue trigger. A Queue trigger will run custom functions as messages are added to an Azure Queue. In this case, you are using blob storage, and so a queue trigger does not suit the purpose.

You should not use an HTTP Trigger. HTTP triggers respond to HTTP requests or webhooks. In this case, you need blob storage and Cosmos DB bindings.

References

[Azure Functions triggers and bindings concepts](#)

[Timer trigger for Azure Functions](#)

[Azure Functions HTTP trigger](#)

[Azure Event Grid trigger for Azure Functions](#)

Question84 Design non-relational cloud data stores

You are designing your company's Cosmos DB platform. The applications that use the data store are distributed globally. Users need fast access to the latest version of each other's data. Users are performing write operations through the 24-hour day.

Applications querying the data must have the lowest latency possible. Results returned to the application in the non-primary region can have missing data that has not yet replicated but cannot have out-of-order data.

You need to choose the consistency level to meet the requirements.

Which consistency level should you use?

Choose the correct answer

Consistent prefix

Bounded staleness

Eventual

Session

Explanation

You should use consistent prefix consistency. With consistent prefix consistency, there is a guarantee that any read operation will never see out-of-order writes. Geo-distributed applications may see more or less of the written data subject to the extent of replication, but they will never see writes that happen out of order.

You should not use eventual consistency. With eventual consistency, while the latency will be low, there is no assurance that the applications will read data in the correct order until the total number of writes has completed and the database replicas eventually converge.

You should not use session consistency. Session consistency is a very common consistency level choice, but it only ensures consistency within the context of a user session. If you are performing writes and reads, it will ensure that your data is consistent and the response latency is low. However, if multiple users or applications are writing and reading data, there is now no guarantee that you can see the data a colleague has written. This breaks the fast sharing of user data requirement. The database replicas will eventually converge, but this will take time while writes are still being performed.

You should not use bounded staleness consistency. Bounded staleness consistency has a version and time interval element that offers total global order within the staleness window. The staleness window is the configured period within which results returned to a query can be behind the latest version of the record. It is configured by settings version number and time limits to when a record can be returned.

References

[Consistency levels in Azure Cosmos DB](#)

Question85 Design non-relational cloud data stores

You are designing your company's Cosmos DB platform to support a globally distributed application. Responses to queries to the data store do not need to return the very latest version of a record. You want to have a flexible configuration model for the data with respect to how up-to-date the responses to read requests from applications are. Responses should be

configurable by the number of item versions and the time interval between reads and writes.

You need to choose a consistency level to meet the requirements.

Which consistency level should you use?

Choose the correct answer

Session

Eventual

Consistent prefix

Bounded staleness

Explanation

You should use bounded staleness consistency. Bounded staleness consistency has a version and time interval element that offers total global order within the staleness window. The staleness window is the configured period within which results returned to a query can be behind the latest version of the record. It is configured by setting the maximum version number and time limits to when a response to a read query can lag behind the writes.

You should not use consistent prefix consistency. With consistent prefix consistency, there is a guarantee that read operation will never see out-of order-writes. However, there is no option to configure the number of versions or time delay on the reads.

You should not use eventual consistency. With eventual consistency, while the latency will be low, there is no assurance that the applications will read data in the correct order until the total number of writes has completed and the database replicas eventually converge. There is also no option to configure the number of versions or time delay on the reads.

You should not use session consistency. Session consistency is a very common consistency level choice, but it only ensures consistency within the context of a user session. There is also no option to configure the number of versions or time delay on the reads.

References

Consistency Levels in Azure Cosmos DB

Question86 Design non-relational cloud data stores

You are designing the partitioning methodology for an Azure Synapse Analytics solution.

One of your fact tables has more than 2 billion rows and must contain a maximum of 37 months of sales data. Month 37 is deleted at the end of each month. Extract, Transform, and Load (ETL) processes will be used to insert data to the main fact table, which contains the following column names and data types:

ProductNumber of type BigInt
OrderDate of type Date
OrderQuantity of type BigInt
UnitPrice of type Decimal
TotalSale of Type Decimal

You need to design the most appropriate way to partition across this data.

Which column should you use?

Choose the correct answer

ProductNumber

TotalSale

OrderDate

OrderQuantity

UnitPrice

Explanation

You should use OrderDate. When partitioning very large tables, it is important to consider the data and potential usage. Azure Synapse Analytics will distribute the data into 60 databases. Any partition added is in addition to this distribution. Best practices dictate that for very large tables, the preferred partition column is date-based. The advantage of date-based partitioning, especially in a sales scenario, is that delete operations, which tend to be expensive to perform one row at a time, can be carried out by partition.

The requirement is for the solution to contain 37 months of data with the last month being removed at the end of each month. To remove an entire month of data would take time and processing expense. By using a date-based partition, an entire month could be dropped in one operation.

The other column types do not lend themselves well to partitioning. The likelihood of there being a very large set of matching values in the other columns is low compared to date.

References

[Cheat sheet for Azure Synapse Analytics (formerly SQL DW)](#)

[Partitioning tables in Synapse SQL pool](#)

Question87 Design non-relational cloud data stores

You are building out the mapping data flow for an Azure Data Factory transformation.

You need to route rows of data to different streams based on rule matches against that data. The routing of data should follow only the first successfully resolved rule and not all rules.

Which mapping data flow transformation should you select?

Choose the correct answer

<mark>Conditional split</mark>

Filter

New branch

Unpivot

Explanation

You should use a conditional split. A conditional split uses a case matching algorithm to determine which particular path a data stream should take.

You should not use a filter. A filter uses a conditional matching process. If something about the data matches a specific condition, the output flow will include all items matching that condition. The key difference between a conditional split and a filter is that the conditional split only returns data from the first successfully executed condition. If additional conditions are met, they are not executed on. A filter, on the other hand, returns results from all records that satisfy the filter. In programming flow, the conditional split is equivalent to a CASE statement, and the filter is equivalent to a WHERE clause.

You should not use new branch. New branch is a very powerful option when you want

to perform multiple operations against the same set of data. This is especially useful when you want to perform multiple sinks or for self-joining data together.

You should not use unpivot. An Azure Data Factory mapping unpivot data flow is a way to turn an unnormalized dataset into a more normalized version. It works by expanding values from multiple records into a single column to generate a multi-column single record with the same values in a single column.

References

[Conditional split transformation in mapping data flow](#)

[Creating a new branch in mapping data flow](#)

[Filter transformation in mapping data flow](#)

[Azure Data Factory Unpivot Transformation](#)

Question88 Design non-relational cloud data stores

You are helping a large retail company with managing reordering of their most popular products, sales forecasting, inventory management, and stock control.

The company allows the managers of its franchise to use their own formats for submitting sales. These must include the current date, item stock keeping unit (SKU) number, price, and sales quantity.

The company wants to reduce the net Cost Of Goods sold (COGS) over time. Ten years of historical sales data must be viewable graphically via dashboards.

You need to propose a solution for storing the initial intake of data.

What should you use?

Choose the correct answer

Azure PolyBase

Azure Synapse

Power BI

Azure Data Lake

You should use Azure Data Lake. Azure Data Lake provides a platform that can intake multiple sources of data in different formats and provide a consumable source to drive data driven experiments and investigations.

You should not use Azure PolyBase here. It is most useful in a data transformation scenario.

You should not use Azure Synapse. Azure Synapse is a large scale analytics engine, but it needs data transformation to take place before analytical processing can begin.

Power BI is a data consumption engine that can drive visibility of data once it is in a consumable format, but it does not help with data formatting or standardization.

References

[What is Polybase?](#)

[What is Azure Synapse Analytics (formerly SQL DW)?](#)

[Data Lake](#)

[What is Power BI?](#)

Question89 Design relational cloud data stores

You plan to use Azure Blob Storage to store recent backups. Your company policy specifies that backups should remain available for at least 60 days. Data availability should be 99% or better. Backups are accessed only if needed for recovery and should be quickly and readily available. Solution costs should be kept to a minimum.

You need to recommend a blob storage tier solution.

Which solution should you recommend?

Choose the correct answer

Archive

Cool

Hot

Premium

You should recommend a cool storage tier. A cool storage tier has lower storage costs than a hot storage tier. Access costs are higher, but the data will be accessed intermittently, if at all. The cool storage tier is designed for data that will be in storage at least 30 days with an additional charge if data is deleted early. It provides quick access to data, usually in milliseconds.

You should not recommend an archive storage tier. This tier has the lowest storage cost but the highest access cost. It is designed for data that will remain in storage for at least 180 days with an additional fee for early deletion. Access times are rated in hours, typically several hours to get access to the data, so this would not meet the access requirements.

You should not recommend a hot storage tier. This has the highest cost for storage. It is designed for interactive data with the lowest cost for access and no additional cost for early deletion.

Premium does not refer to a storage tier but to a storage account level. Premium storage does not support hot, cool, or archive storage tiers.

References

[Hot Vs Cool Vs Archive access tiers – Azure Blob Storage](#)

[Azure Blob Storage Tiers Explained](#)

[The Essential Guide to Azure Blob Storage Pricing](#)

[Storage Options for Lower Azure Storage Costs and Azure Backup Costs](#)

Question90 Design relational cloud data stores

Your company deploys an Azure SQL single database to support cloud-based applications. You need to recommend a disaster recovery solution that meets the following requirements:

- Supports ongoing operations after a regional failure.
- Secondary instances support read-only queries.

- All solution components are cloud-based.

Which solution should you recommend?

Choose the correct answer

SQL Server replication

Geo-replication

Geo-redundant storage (GRS)

Locally redundant storage (LRS)

Explanation

You should recommend geo-replication. This feature of Azure SQL Database replicates a full SQL Server instance to another location. You can configure up to four secondaries in different geographic locations. All secondaries support read-only queries. Should a failure occur, you can fail over to a secondary, making it the new primary with read/write access.

You should not recommend SQL Server replication. SQL Server replication does not support placing replication agents on an Azure SQL Database. This solution cannot be implemented solely in the cloud and is commonly used to support replication between a cloud-based and on-premises database. SQL Server replication does not provide for failover in case of failure.

You should not recommend GRS or LRS. These are both methods to provide disaster recovery for Azure storage, but they do not replicate the SQL Server instance.

References

What is a single database in Azure SQL Database

Azure Storage redundancy

Creating and using active geo-replication

Replication to SQL Database single and pooled databases

Question91 Design relational cloud data stores

You implement an Azure SQL Database managed instance to support a business-critical application. Your must design and implement a disaster recovery strategy to ensure operations in case of a regional failure in the primary region. Your solution must support automatic failover configured through user-defined policy.

You need to identify the best solution to meet your requirements.

What should you recommend?

Choose the correct answer
SQL Server replication
Geo-replication
Geo-redundant storage (GRS)
Failover group

Explanation

You should recommend a failover group. You can use a failover group to replicate a managed instance to geographically diverse locations, protecting against a regional failure or unavailability. Failover can be manual or can be automatic through a user-defined policy. Secondary instances support read-only access.

You should not recommend geo-replication. Even though a failover group is based on the same underlying technology as geo-replication, geo-replication does not support Azure SQL Database managed instances.

You should not recommend SQL Server replication. This does provide for replicated instances of SQL Server, but it does not provide for automated failover in case of failure.

You should not recommend GRS. GRS provides disaster recovery for Azure storage, but it does not replicate the SQL Server instance.

References

Use auto-failover groups to enable transparent and coordinated failover of multiple databases

Creating and using active geo-replication

Question92 Design non-relational cloud data stores

You are setting up storage support for an application that uses Azure Data Lake Storage Gen2. A primary concern is disaster recovery.

You need to implement a disaster recovery solution that helps to ensure recovery after a catastrophic event resulting in widespread failure throughout a region.

What should you use?

Choose the correct answer

Geo-replication

Locally redundant storage (LRS)

Zone redundant storage (ZRS)

Geo-redundant storage (GRS)

Explanation

You need to implement GRS. With GRS, data remains available even if an entire datacenter becomes unavailable or if there is a widespread regional failure. Alternately, you could implement read-access geo-redundant storage (RA-GRS), which provides read-access to the data in alternate locations.

You should not implement LRS or ZRS. LRS does not provide protection in case of a zone or regional failure. ZRS protects against a zone failure but not a regional failure.

You should not implement geo-replication. This is a feature of Azure SQL Database and is not supported by Azure Data Lake Storage Gen2. Geo-replication replicates a full SQL Server instance to another location.

References

Best practices for using Azure Data Lake Storage Gen2

Introduction to Azure Data Lake Storage Gen2

Azure Storage redundancy

Question93 Design relational cloud data stores

You are a data architect for your company. You are designing the disaster recover strategy for an Azure SQL Database. Data in the database changes approximately every four hours. You want to use point-in-time recovery so that in the event of failure, you can restore the database as quickly as possible. You also want to use the least amount of disk space required for the backup solution. You consider the following backup solutions:

- Differential backups every day
- Differential backups every 12 hours
- Differential backups every week
- Full backups every day
- Full backups every 12 hours
- Full backups every week
- Transaction log backups

You need to recommend the type of backups to use and the order in which they should be restored.

To answer, select the appropriate backup types from the drop-down menus.
Choose the correct options

Restore Order	Backup Type

First

Second

Third

fully weekly backup

[dropdown] ▾

differential backup from the last 12 hours

[dropdown] ▾

transaction log backup backs ups since last differential backup

You need to recommend the type of backups to use and the order in which they should be restored.

To answer, select the appropriate backup types from the drop-down menus.

Choose the correct options

Restore Order	Backup Type
First	full weekly backup
Second	differential backup from last 12 hours
Third	transaction log backups since last differential backup

‹ Previous ☆ Next ›   93 of 172

Explanation

You should use a combination of a full weekly backup, a differential backup, and transaction log backups. A full backup backs up the entire database. This type of backup includes the transaction log so that the entire database can be restored if the full backup is restored. A differential backup backs up data since the most recent full backup. It captures only the data that has changed. Therefore, they are a lot smaller than full backups. A transaction log backup backs up the transaction logs. A transaction log consists of all the transactions that have occurred. This solution uses the least amount of disk space because only one full backup occurs every week, and differential backups use only enough disk space to capture changes since the last full backup. By first restoring the previous full weekly backup, you restore the database to the state it was in the previous week. By next restoring the differential backup from the last 12 hours, you restore all the changes that occurred between last week and the last 12 hours. By finally restoring the transaction log since the last differential backup, you restore the database to the point of failure. These backups are supported by automatic backups.

You should not use full daily backups. This requires the most disk space because a full

backup consists of the entire database.

You should not use full backups every 12 hours. This requires more disk space than full weekly backups.

You should not restore a differential backup first. You must restore a full backup before you can restore a differential backup.

You should not restore transaction log backups since the last full backup. This takes longer than restoring them since the last differential backup.

References

Manage Azure SQL Database long-term backup retention

Transaction Log Backups (SQL Server)

Differential Backups (SQL Server)

Full Database Backups (SQL Server)

Question94 Design non-relational cloud data stores

You are a data architect for your company. You plan to deploy two Azure Cosmos DB accounts. The accounts must meet the following requirements:

- CosmosDB1 - You want to configure the time by which reads can lag behind writes.
- CosmosDB2 - You want to ensure that each client application reads the same values that it wrote with minimal latency.

You need to choose the most appropriate consistency level for each account.

Which consistency levels should you use? To answer, drag the appropriate consistency level to each Cosmos DB account. Each consistency level may be used once, more than once, or not at all.

Drag and drop the answers

session

bonded staleness

session

bonded staleness

eventual

consistent prefix

strong

pts.measureup.com/web/PBS/LMS/index.php?l=1&k=6834331&tam=959810&role=1&course=at_test202006251135141593084914

Apps   Outlook Web App   IXL - Skip-counting...   Faircost Pouch for L...   uanmembers.epfos...   Creating a macro th...   Bitcoin for beginner...   Test user sign up |...   Opportunities Dash...

Cosmos DB account. Each consistency level may be used once, more than once, or not at all.

Drag and drop the answers

CosmosDB1 | bonded staleness |

CosmosDB2 | session |

| session | bonded staleness | eventual |
| consistent prefix | strong | |

▲ Explanation

You should use bonded staleness for CosmosDB1. This is the second strongest consistency level. With this consistency level, you can configure the staleness of a read. You can configure the staleness with a specific

‹ Previous   ☆ Next ›

94   of 172

8:20 PM
8/7/2020

Explanation

You should use bonded staleness for CosmosDB1. This is the second strongest consistency level. With this consistency level, you can configure the staleness of a read. You can configure the staleness with a specific number of versions, or you can configure it with a specific amount of time. In this scenario, you want to configure the time that reads lag behind writes.

You should use session consistency for CosmosDB2. This is one level weaker than bonded staleness. It ensures that the same client always reads the same data.

You should use not strong consistency. This consistency level ensures that all client applications always read the same values. A replica is not available unless it contains

the same data as the node where a data write occurred. This is the strongest consistency level, but it offers the lowest availability.

You should not use eventual consistency. This consistency level allows you to achieve the highest availability and the lowest latency. Replicas do not necessarily contain the same data. However, the replicas will eventually converge. This is the weakest consistency level, but it offers the highest availability.

You should not use consistent prefix. This is one level weaker than bonded session. This consistency level guarantees that reads from any client never see out-of-order writes.

References

[Consistency levels in Azure Cosmos DB](#)

Question95 Design relational cloud data stores

You plan to use Azure SQL Database to store data in the cloud. You are creating a backup and restore strategy. You use the backup plan scenario in the exhibit. Assuming that the scenario in the exhibit occurs, you want to determine how to restore the database to its corruption point by using the quickest process and the fewest number of restores.

You need to perform the restore.

Which five actions should you perform in sequence? To answer, move the appropriate actions from the list of possible actions to the answer area and arrange them in the correct order.

Create a list in the correct order

Restore Tasks

Restore Tasks in Order

- Restore the first transaction log backup from Monday.
- Restore the last transaction log backup from Monday.
- Create a tail-log backup of the current transaction log.
- Restore the full backup from Sunday.
- Restore the differential backup from Monday.
- Restore the transaction log backup from Tuesday.
- Restore the tail-log backup.

Explanation

You should perform the following tasks in order:

1. Create a tail-log backup of the current transaction log.
2. Restore the full backup from Sunday.
3. Restore the differential backup from Monday.
4. Restore the transaction log backup from Tuesday.
5. Restore the tail-log backup.

You should first create a backup of the current transaction log. The transaction log is separate from the database itself. It contains all the transactions up to the current point of failure. Next, you should restore the full backup. A full backup consists of the entire database. Before you restore any other type of backup, you must restore the full backup. Next, you should restore the differential backup. A differential backup contains only the data changes since the last backup. Then, you should restore the transaction log backup from Tuesday. This rolls forward all committed transactions up to 11:00 AM on Tuesday. Finally, you should restore the tail log backup. This rolls forward all committed transactions between 11:00 AM on Tuesday and the current point of corruption.

You should not restore the transaction log backups from Monday. This is unnecessary since you can restore all data by restoring the differential backup from Monday night.

References

Transaction Log Backups (SQL Server)

Apply Transaction Log Backups (SQL Server)

Question96 Design relational cloud data stores

You plan to use Azure SQL Database to store data in the cloud. You are creating a backup and restore strategy. You use the backup plan scenario in the exhibit. Assuming that the scenario in the exhibit occurs, you want to determine how to restore the database to its corruption point by using the quickest process and the fewest number of restores.

You need to perform the restore.

Which five actions should you perform in sequence? To answer, move the appropriate actions from the list of possible actions to the answer area and arrange them in the correct order.

Create a list in the correct order

Possible actions

Actions in order

- Create a differential backup.
- Create a tail-log backup of the current transaction log.
- Restore the full backup.
- Restore the 11:00 AM transaction log backup.
- Restore the 2:00 PM transaction log backup.
- Restore the tail-log backup.

Explanation

You should perform the following tasks in order:

1. Create a tail-log backup of the current transaction log.
2. Restore the full backup.
3. Restore the 11:00 AM transaction log backup.
4. Restore the 2:00 PM transaction log backup.
5. Restore the tail-log backup.

You should first create a tail-log backup of the current transaction log. The transaction log is separate from the database itself. It contains the transactions up to the current point of failure. Next, you should restore the full backup. A full backup consists of the entire database. Before you restore any other type of backup, you must restore the full backup. Next, you should restore the 11:00 AM Monday transaction log backup. This rolls forward all committed transactions up to 11:00 AM on Monday. Then, you should restore the 2:00 PM Monday transaction log backup. This rolls forward all transactions between 11:00 AM and 2:00 PM. Finally, you should restore the tail log backup. This rolls forward all committed transactions between 2:00 PM on Monday and the current point of corruption.

You should not create a differential backup. A differential backup backs up all data that has changed since the last full backup.

References

Transaction Log Backups (SQL Server)

Apply Transaction Log Backups (SQL Server)

Question97 Design relational cloud data stores

You are a data architect for your company. You are designing distributions for an Azure SQL Data Warehouse. You plan to import data into two tables.

You need to determine the partitioning method to use for each table.

To answer, drag the appropriate partitioning method to each table. A partitioning method may be used once, more than once, or not at all.

Drag and drop the answers

Round-robin

Replicated

Round-robin

Replicated

Hash

Explanation

You should use round-robin distribution for the first table. Round-robin allows you to distribute rows evenly across the distributions in SQL Data Warehouse. This speeds up loading of data.

You should use a replicated table for the second table. This partitioning method is recommended for small tables that are used frequently in joins. A copy of the entire table is replicated across all distributions.

You should not use hash distribution for either table. With hash distribution, a hashing function is applied against a column. All columns that have the same value are stored on the same distribution.

References

[Azure SQL Data Warehouse - Massively parallel processing (MPP) architecture](#)

Question98 Design relational cloud data stores

Your company is building a web application that accesses a database. Data in the database changes frequently, and it should be consistent for all users. Data must also be immediately available to be read after it is written. You want to deploy the web application and database to Azure. You want to use the Azure SQL Database service to host the database. The application must be deployed so that it has minimal downtime and latency. 90 percent of the application's users are in the US West region, and 10 percent are in the US East region. You plan to use DNS-based load balancing for the web application.

You need to design the deployment.

To answer, select the appropriate configurations from the drop-down menus.

Choose the correct options

Deploy the web application to   region(s).

Deploy the database to   region(s).

geo-replication.

[▼]

both

Explanation

You should deploy the web application to both regions. This allows the DNS-based load balancer to route web requests to the web application nearest the user.

You should deploy the database to the US West region and enable geo-replication. This allows the database to be replicated from the US West region to the US East region. In the event of a regional outage, a fail-over can occur.

You should not deploy the web application to only one region. This increases latency for users on the opposite region.

You should not deploy the database to both regions. This would require instantaneous synchronization between the two regions, which is impossible. In this scenario, the data must be consistent for all users.

You should not disable geo-replication. This prevents a fail-over from occurring if a regional outage occurs, which increases downtime.

References

[Designing globally available services using Azure SQL Database](#)

Question99 Design relational cloud data stores

You are a data architect for your company. You plan to deploy Azure SQL Data Warehouse. A large file exists in Azure blob storage.

You need to design a table that allows you to reference the data from blob storage.

Which type of table should you use?

Choose the correct answer
Hash-distributed
Round-robin
Replicated
<mark>External</mark>

Explanation

You should use an external table. An external table stores its schema in Azure SQL Data Warehouse. However, the actual data for the table remains in Azure blob storage.

You should not use a round-robin table. A round-robin table distributes rows evenly across the 60 distributions in Azure SQL Data Warehouse. This occurs after data is loaded from blob storage and moved to Azure SQL Data Warehouse.

You should not use a replicated table. A replicated table copies all rows across all 60 distributions in Azure SQL Data Warehouse. This occurs after data is loaded from blob storage and moved to Azure SQL Data Warehouse.

You should not use a hash-distributed table. A hash-distributed table distributes rows based on the hash value of a column. Rows with columns that contain the same hash value are placed on the same distribution. This occurs after data is loaded from blob storage and moved to Azure SQL Data Warehouse.

References

[Guidance for designing distributed tables in Azure SQL Data Warehouse](#)

Question100 Design relational cloud data stores

You deploy a single instance of Azure SQL Database. The database is configured for automated backup with the default configuration.

You need to determine the point-in-time recovery procedures in case of failure. You must identify the backups from which you need to restore and in what order.

Which three backups should you perform in sequence? To answer, move the appropriate backups from the list of possible backups to the answer area and arrange them in the correct order.

Create a list in the correct order

Possible backups

Backups in order

- Full daily backup
- All log backups since the last full backup
- All differential backups since the last full backup
- Full weekly backup
- Most recent differential backup
- All log backups since the last differential backup

Explanation

A point-in-time recovery requires you to restore from backups in the following order:

1. Full weekly backup
2. Most recent differential backup
3. All log backups since the last differential backup

By default, automated backup runs the following backups:

- Weekly full backup
- Differential backup every 12 hours
- Log backup every 5 to 10 minutes

Automated backups are supported for Basic, Simple, and Premium databases. A full backup backs up all the data from the database. Each differential backup backs up changes since the most recent full backup. The longer the time from the full backup, the larger the differential backup file. Log backups back up and clear committed transactions from the transaction log.

You should start by restoring the most recent full backup. Next, you should restore the most recent differential backup, which will include all changes since the last full backup. Finally, you should restore the log backups made after the most recent differential backup to bring the database up-to-date.

Because a differential backup includes all of the data from previous differential backups, it is only necessary to restore from the most recent backup. After this, you need to apply the changes made after the differential backup so that only the log backups made after that differential backup are needed.

Automated backups do not run daily backups by default.

References

Automated backups

Recover an Azure SQL database using automated database backups

Question101 Design relational cloud data stores

A company has 3D fabrication plants in New York, Mexico City, Quebec, and San Francisco. Images are captured of projects in progress and tagged with the time and project number.

Image information is maintained in an Azure storage blob in the region nearest to New York. Tagging data is collected from all sites to New York and uploaded to an Azure SQL Database. Tagging data should be replicated through geo-replication to locations near each of the fabrication plants. The storage solution must support failover for tagging information.

How should you configure replication?

To answer, select the appropriate setting from the drop-down menus.

Choose the correct options

New York

Mexico City

Quebec

San Francisco

[dropdown menu]

[dropdown menu]

[dropdown menu]

[dropdown menu]

How should you configure replication?

To answer, select the appropriate setting from the drop-down menus.

Choose the correct options

New York	Write region
Mexico City	Read region
Quebec	Read region
San Francisco	Read region

‹ Previous ☆ Next ›

101 | of 172

Explanation

You need to configure geo-replication with the region nearest to New York hosting the primary server and Mexico City, Quebec, and San Francisco as secondary locations. Geo-replication is a business continuity solution that provides for rapid failover and continued operation in case of a regional failure or a large scale outage. You can

configure one primary and up to four secondaries.

The primary is the only writeable instance, so it would be the write region. This is necessary because data is uploaded from New York. All secondaries, the remaining regions, are read-only. You can failover to a secondary at any time.

References

[Creating and using active geo-replication](#)

[Configure active geo-replication for Azure SQL Database in the Azure portal and initiate failover](#)

Question102 Design relational cloud data stores

Your company is migrating several applications and their supporting independent SQL Server databases from on-premises to the cloud. Databases have different resource requirements and different peak usage periods. You want to set maximum resource limits on the databases as a group and provide the ability to scale to meet peak requirements.

Which solution should you implement?

Choose the correct answer

Elastic pool

Single instances

Clustering

Managed instances

Explanation

You should implement an elastic pool. In an elastic pool, several databases share the same resources. You can set resource limits on the pool and, within those limits, databases can scale to meet resource needs for peak requirements. This solution is well suited because the databases have different peak usage periods.

You should not implement Azure SQL Database managed instances. You can set resource limits on managed instances, but these are controlled on an individual database level with no resource sharing between the databases. Managed instances are recommended when you need near 100% compatibility between an on-premises deployment and your cloud deployment.

You should not implement Azure SQL Database single instances. As with managed instances, resources are managed individually by a database and not shared with a group of databases. You can move a single instance into an elastic pool or choose to make a database that is part of a pool a single instance, if necessary. Deploying a database as a single instance makes it independent from other instances and easily portable.

Clusters are an auto-failover configuration that replicates multiple copies of a database to support quick failover in case of failure. It is for disaster avoidance and automatic recovery rather that database resource management.

References

[Elastic pools help you manage and scale multiple Azure SQL databases](#)

[Scaling out with Azure SQL Database](#)

[What is Azure SQL Database managed instance?](#)

[What is a single database in Azure SQL Database](#)

[Use auto-failover groups to enable transparent and coordinated failover of multiple databases](#)

Question103 Design relational cloud data stores

You deploy a new Azure SQL Database as a single instance under the vCore-based purchasing model. You closely monitor database activity during the initial deployment to determine resource requirements and peak usage periods.

You need to manually adjust database resources or the service tier without downtime or performance impact.

Which technology should you implement?

Choose the correct answer

Autoscaling

Database sharding

Read scale-out

Explanation

You should implement dynamic scalability. You can manually change resource limits, purchasing model, and database tier support with little or no downtime and negligible impact on applications or users. Changing between purchase tiers often has a short (no more than 4 seconds) downtime. You can automate this process somewhat by using PowerShell to monitor and manage database instances.

You should not implement autoscaling. Autoscaling is a way to apply automatic adjustments to resources available to a database, but it is not supported in a single instance implementation. Autoscaling is used to share resources between databases in an elastic pool.

You should not use database sharding. Database sharding is not a resource management technology. It is a method of scaling out the data from a database across multiple databases through horizontal partitioning.

You should not use read scale-out. You can use read scale-out to improve reporting performance when you have a read-only replica of the database, such as in geo-replication. The read-only workload can be offloaded to a secondary (read-only) instance to relieve some of the load from the primary database.

References

[Dynamically scale database resources with minimal downtime](#)

[Azure SQL Database scheduled autoscaling](#)

[What is a single database in Azure SQL Database](#)

Question104 Design relational cloud data stores

You are configuring an Azure SQL database to support a large online transaction processing (OLTP) application. The database must support a typical load of 25,000 input/output operations per second (IOPS) with a peak load of up to 35,000 IOPS. You want to minimize storage costs.

You need to select the appropriate type of disk.

What should you select?

Choose the correct answer

Standard SSD Managed Disk

Ultra SSD Managed Disk

Premium SSD Managed Disk

Standard HDD Managed Disk

Explanation

You should select an Ultra SSD Managed Disk solution. You would need to choose a disk size of at least 128 GB, which can support up to 38,400 IOPS.

Even though they are less expensive, neither Premium SSD, Standard SSD, nor Standard HDD can meet the performance baseline.

References

What disk types are available in Azure?

OLTP (online transaction processing)

IOPS (input/output operations per second)

Question105 Design relational cloud data stores

You have a 900 GB MySQL database that supports your company's online retail operations. You need to migrate the database to an Azure MySQL database. Interruptions to applications that use the database must be kept to a minimum.

What should you use to migrate the database?

Choose the correct answer

Azure Database Migration Service

Import and export

MySQL Workbench

Data Migration Assistant

Explanation

You should use Azure Database Migration Service (DMS) to minimize downtime during migration. DMS performs an initial load of on-premises data to Azure MySQL and then continually syncs any new transactions to the target Azure database. When ready to complete the migration, you would stop the application and then update the connection string to Azure.

You should not use MySQL Workbench. You can use MySQL Workbench as part of your migration procedures, but this does not minimize downtime. You would use MySQL Workbench to migrate the database schema and then use the mysqldump utility to dump and restore the data. This procedure is best suited to small databases when downtime is not a major concern.

You should not use import and export. Import and export using the mysqlimport utility does not minimize downtime. Import and export is recommended when:

- Migrating a select subset of tables
- Migrating database objects other than tables
- Moving data from non-MySQL external data sources

You should not use Data Migration Assistant (DMA). You would use this when migrating from on-premises SQL to an Azure SQL database. DMA does not support MySQL migration.

References

[Minimal-downtime migration to Azure Database for MySQL](#)

[Migrate your MySQL database by using import and export](#)

[Migrate MySQL to Azure Database for MySQL](#)

[Overview of Data Migration Assistant](#)

Question106 Recommend an Azure Data solution based on requirements

An in-house team is developing a new application. The design document specifies that data should be represented using nodes and relationships in graph structures. Individual data elements are relatively small.

You need to recommend an appropriate data storage solution.

Which solution should you recommend?

Choose the correct answer

Cosmos DB

HBase in HDInsight

Azure Data Lake Store

Azure Storage Blobs

Explanation

You should choose Cosmos DB as your storage solution. Cosmos DB is a globally distributed multi-model database that is suited to various storage scenarios, including:

- Key/value stores in which each data value is associated with a unique key, effectively creating a large hash table.
- Document databases with documents made up of a collection of named fields and data and supporting a variety of encoding methods.
- Graph databases with data represented by nodes and relationships in graph structures.

You should not choose HBase in HDInsight. This is a column-data database that organizes data into rows and columns using a denormalized approach to data organization and storage. Data is typically stored in key order.

You should not choose Azure Storage blobs. Blobs are designed for object-based storage for text- and binary-based data. It is optimized for storage and retrieval of large binary objects such as image files and audio streams.

You should not choose Azure Data Lake Store. This storage solution is a data analytics solution that is an enterprise-wide hyperscale repository for big data analytic workloads supporting parallel solutions for ingesting, storing, and analyzing data. It supports a wide variety processing and analytics options across platforms and languages.

References

Choose the right data store

Question107 Design relational cloud data stores

You are designing the security elements of your company's application that uses Azure SQL Database. As a part of the information security requirements, no data must be directly accessible from database backups or transaction log backups. All data is considered highly confidential and must accessible only through the application or via a trusted set of administrators.

You need to choose a security design for the database that will meet the requirements.

Which Azure SQL Database capability should you use?

Choose the correct answer

<mark>Transparent Database Encryption (TDE)</mark>

Transport Layer Security (TLS) encryption

Azure Information Protection (AIP)

Always Encrypted

Explanation

You should use TDE. TDE encrypts data at rest on disk and decrypts it as it is read into memory for consumption. It uses a certificate-based encryption method. The certificate is required to perform any restore operation, so the backups are not accessible unless they are restored to a server that uses the same certificate.

You should not use AIP. AIP allows you to apply labels to documents and emails to classify and optionally protect their content. Administrators define policies that can be applied automatically, applied by end users, or both.

You should not use Always Encrypted. Always Encrypted is a method that allows you to classify sensitive data in certain columns of a database as confidential. It is not used to encrypt the whole database. It would also be possible for authorized users on the server side to decrypt the data and move it to unencrypted columns, thereby not meeting the requirement that only the application can see the data.

You should not use TLS encryption. TLS is an encryption method for data in transit that operates between two applications. Usually applications use TLS encryption to ensure the integrity and source of the data and responses between two applications. It evolved from the earlier Secure Sockets Layer (SSL) security protocol, which is no longer recommend because TLS is much more secure.

References

[Transparent data encryption for SQL Database, SQL Managed Instance, and Azure Synapse Analytics](#)

[What is Azure Information Protection?](#)

[Always Encrypted](#)

[Transport Layer Security (TLS)](#)

Question108 Design relational cloud data stores

You are setting up the storage requirements of your company's SQL Managed Instance dataset backups. Data is expected to be retained for a maximum of 30 days in the event of a restore, but there also a requirement for storage that could require immediate access when it is requested. The storage must use the most cost-effective storage pricing.

You need to choose which type of blob storage you should use for the SQL Managed Instance dataset backups.

Which storage type should you use?

Choose the correct answer

Premium

Hot

Archive

Cool

Explanation

You should use cool storage. Cool storage is ideal for short-term storage of data while gathering more data for processing and retention of database or dataset backups in

the event of restoration within a defined period. It is also suitable for immediate retrieval of files when they are needed quickly. The storage costs are much lower than hot or premium but higher than archive.

You should not use archive storage. Archive data is not immediately accessible and is generally used for long-term backup and retention of datasets. To retrieve an item from an archive storage tier, it must be extracted from a repository and made accessible for access to an online storage tier (rehydrated). Typical access times are up to 15 minutes. Archive is the cheapest storage option but does not meet the access requirements.

You should not use hot storage. Hot storage has much better access times than cool storage, but it has much higher storage costs. It is typically used for data that is accessed frequently.

You should not use premium storage. Premium storage is used for highly interactive workloads and workloads that require very high performance. It also costs more than the other options.

References

[Azure Blob storage: hot, cool, and archive access tiers](#)

[Performance tiers for block blob storage](#)

Question109 Design relational cloud data stores

You are designing the data solution for your customer support database. The database runs on Azure SQL Database. The primary table structure needs to include columns for customer identity, product SKU, and customer location.

You need to optimize the performance of queries against the database by product.

Which partitioning strategy should you use?

Choose the correct answer

Vertical partitioning based on product SKU

Horizontal partitioning based on customer identity

Functional partitioning based on product SKU

Explanation

You should use horizontal partitioning based on product SKU. In horizontal partitioning, each partition is a separate data store, but all partitions have the same schema. Each partition or shard will contain a specific subset of the data (in this case including the product SKU) With the SKU as the partition key, you can optimize all queries to the specific database partition that contains that column.

You should not use vertical partitioning based on product SKU. In this case partitioning vertically would not optimize queries because the partitions would not be optimally distributed to support querying based on product SKU. Vertical partitioning is most commonly used to reduce IO by storing the most commonly accessed elements of a record in a single partition and the less commonly retrieved in another, thereby reducing the potential number and sizes or retrieved records.

You should not use functional partitioning based on product SKU. In this case functional partitioning would not optimize queries because the partitions would not be optimally distributed. Functional partitioning uses multiple column types such as product SKU or invoice date to distribute data. In this instance, you want to optimize querying against the product SKU, so maximizing the number of partitions based on product SKU makes the most sense.

You should not use vertical partitioning based on customer identity. In this case partitioning against the product SKU is the required scenario, so partitioning against any other column will not result in the queries being optimally distributed.

References

[Horizontal, vertical, and functional data partitioning](#)

Question110

You manage an on-premises Microsoft SQL Server 2019 instance and an Azure SQL Database Managed instance.

A data engineer is writing a query in an on-premises application that needs to join data stored in a database on the Azure SQL Database managed instance. This query will be used to submit data to an Event Hub every minute for batch processing.

You need to recommend a solution to prepare the environment.

What should you recommend?

Create a master key on the Azure SQL Database managed instance.

Create a linked server on the on-premises instance.

Enable PolyBase on the Azure SQL Database managed instance.

Enable PolyBase on the on-premises instance.

Explanation

You should recommend that the data engineer create a linked server on the on-premises instance. To do this, you must run the sp_addlinkedserver stored procedure. A linked server allows you to access data from a separate SQL Server instance.

You should not recommend that the data engineer create a master key on the Azure SQL Database managed instance. A master key is responsible for encrypting and decrypting other encryption keys. A master key is not necessary to access a linked server. To create a master key, you need to run the CREATE MASTER KEY statement.

You should not recommend that the data engineer enable PolyBase on either instance. PolyBase is a technology that allows SQL Server to access data from external data sources, such as blob storage, using the same queries that you use to access data from other SQL Server databases. Polybase is automatically enabled on Azure SQL Data Warehouse. To enable Polybase, you should run the following statements:

exec sp_configure @configname = 'polybase enabled', @configvalue = 1;
RECONFIGURE [WITH OVERRIDE] ;

References

[Install PolyBase on Windows](#)

[What is PolyBase?](#)

[Create Linked Servers (SQL Server Database Engine)](#)

[CREATE MASTER KEY (Transact-SQL)](#)

Question111

You are designing a solution that allows sports leagues to send scores to the cloud. The solution must allow streaming of over 5,000 HTTPS requests each day. However, the payload of each request is small.

You need to choose a resource for initially receiving the data. Your solution must be cost effective.

Which resource should you use?

Choose the correct answer

Data Factory

Data Lake

Event Hub

IoT Hub

Explanation

You should use Event Hub. It supports up to 5,000 concurrent connections. Supported protocols include HTTPS and AMQP.

You should not use IoT Hub. IoT Hub is similar to Event Hub. In addition to those features, it supports per-device identity and device provisioning. IoT Hub is more expensive than Event Hub.

You should not use Databricks. Databricks uses Spark-based clusters to process big data.

You should not use Data Factory. Data Factory allows you to create a pipeline for copying and transforming data.

References

[Transform data by using Databricks in Azure Data Factory](#)

[Connecting IoT Devices to Azure: IoT Hub and Event Hubs](#)

Question112

You are designing a big data batch processing and streaming solution.

You need to choose the most appropriate resource for different scenarios.

Which resources should you choose? To answer, drag the appropriate resource to each scenario. A resource may be used once, more than once, or not at all.

Drag and drop the answers

Stream Analytics

Databricks
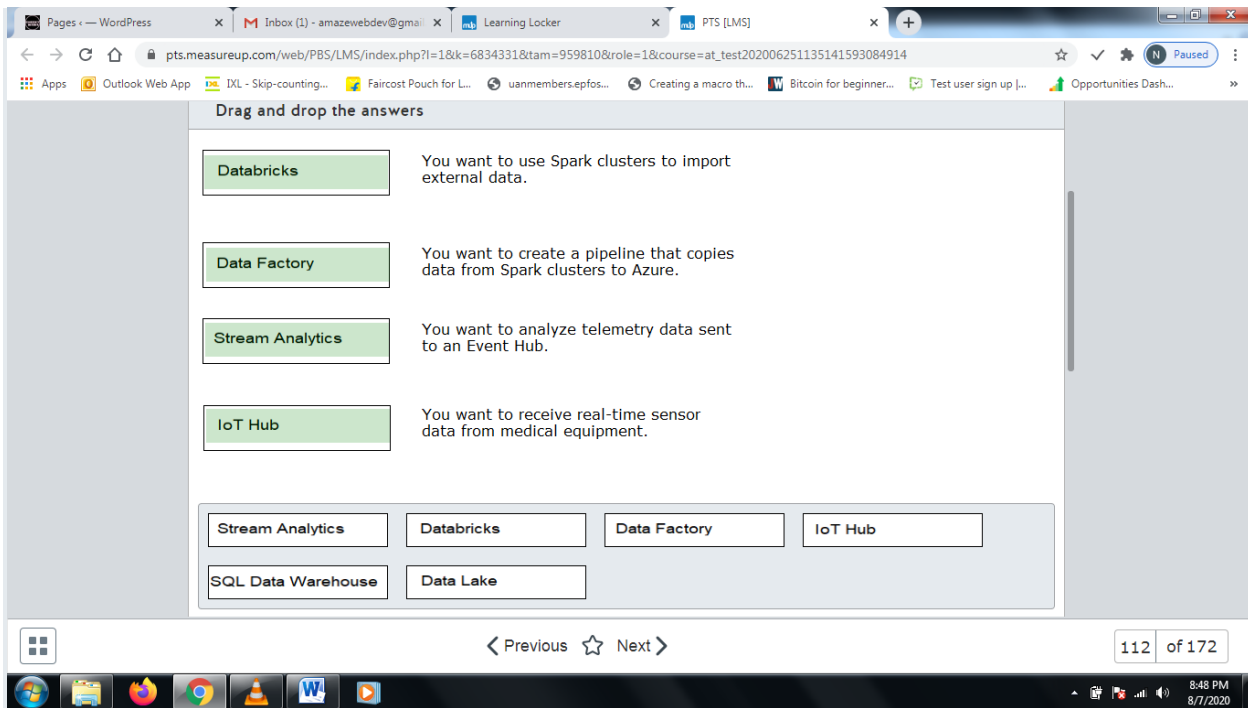
Data Factory

IoT Hub

Stream Analytics	
Databricks	
Data Factory	
IoT Hub	
SQL Data Warehouse	
Data Lake	

Drag and drop the answers

| Databricks | You want to use Spark clusters to import external data. |

| Data Factory | You want to create a pipeline that copies data from Spark clusters to Azure. |

| Stream Analytics | You want to analyze telemetry data sent to an Event Hub. |

| IoT Hub | You want to receive real-time sensor data from medical equipment. |

| Stream Analytics | Databricks | Data Factory | IoT Hub |
| SQL Data Warehouse | Data Lake | | |

‹ Previous ☆ Next ›   112 of 172

8:48 PM
8/7/2020

Explanation

You should use Databricks to import external data into Spark clusters. Databricks is a big data analytics solution that allows you to import data into Spark clusters and transform it.

You should use Data Factory to create a pipeline that copies data from Spark clusters to Azure. Data Factory allows you to create a pipeline that allows you to copy and transform data in the cloud or between on-premises and the cloud.

You should use Stream Analytics to analyze telemetry data sent to an Event Hub. Stream Analytics is a real-time analytics solution that allows you to process streaming events from multiple sources simultaneously.

You should use IoT Hub to receive real-time sensor data from medical equipment. IoT Hub allows you to stream billions of requests per day to Azure. It supports HTTPS and AMQP. It also supports device provisioning and per-device identity.

You should not use SQL Data Warehouse. SQL Data Warehouse is a distributed database system designed for processing petabytes of data in parallel.

You should not use Azure Data Lake. Data Lake allows you to store data of any size and type in the cloud until it is ready for processing.

References

Question113

You are designing a big data streaming solution.

You need to choose the most appropriate resource for different scenarios.

Which Azure resource should you choose? To answer, drag the appropriate resource to each scenario. A resource may be used once, more than once, or not at all.

Drag and drop the answers

Stream Analytics

SQL Data Warehouse

Stream Analytics

SQL Data Warehouse

Databricks

IoT Hub

Data Lake

Data Factory

You need to choose the most appropriate resource for different scenarios.

Which Azure resource should you choose? To answer, drag the appropriate resource to each scenario. A resource may be used once, more than once, or not at all.

Drag and drop the answers

| SQL Data Warehouse | You want to run parallel procesing on 4-terabytes (TB) of data. |

| Stream Analytics | You want to analyze telemetry data sent to an IoT Hub. |

| Stream Analytics | SQL Data Warehouse | Databricks | IoT Hub |
| Data Lake | Data Factory | | |

Explanation

You should use SQL Data Warehouse to run parallel processing on 4-TB data. SQL Data Warehouse is a distributed database system designed for processing petabytes of data in parallel.

Explanation

You should use SQL Data Warehouse to run parallel processing on 4-TB data. SQL Data Warehouse is a distributed database system designed for processing petabytes of data in parallel.

You should use Stream Analytics to analyze telemetry data sent to an IoT Hub. Stream Analytics is a real-time analytics solution that allows you to process streaming events from multiple sources simultaneously.

You should not use IoT Hub. IoT Hub allows you to stream billions of requests per day to Azure. It supports HTTPS and AMQP. It also supports device provisioning and per-device identity.

You should not use Azure Data Lake. Data Lake allows you to store data of any size and type in the cloud until it is ready for processing.

You should not use Databricks. Databricks is a big data analytics solution that allows you to import data into Spark clusters and transform it.

You should not use Data Factory. Data Factory allows you to create a pipeline that allows you to copy and transform data in the cloud or between on-premises and the cloud.

References

Question114 Recommend an Azure Data solution based on requirements

You are a data architect for your company. The development team is planning to develop an application that stores data as nodes in a graph.

You need to choose the best data storage solution to meet this requirement.

Which data storage should you use?

Choose the correct answer

<mark>Cosmos DB</mark>

Blob storage container

Azure Data Lake

Azure Table

Explanation

You should use Cosmos DB. This is a multi-model database storage solution that supports one of five APIs, including SQL, MongoDB, Table, Gremlin, and Cassandra. The API that you should use is Gremlin, which manages and operates on graph data.

You should not use Azure Table. This solution stores data as key-attribute pairs.

You should not use Azure Data Lake. This is a big data repository. It stores data of any type. It serves as an intermediary data store. You cannot natively navigate graph data stored in Azure Data Lake. You must move it to Cosmos DB.

You should not use a blob storage container. This allows you to store unstructured, binary data.

References

[Welcome to Azure Cosmos DB](#)

[Introduction to Azure Cosmos DB: Gremlin API](#)

[Azure Table storage overview](#)

Question115

You are helping design a big data batch processing solution for long-running batch jobs. You need to select a batch processing technology that supports relational data storage and a batch job pricing model. The data store should support Azure Active Directory (Azure AD) authentication.

Which should you choose?

Choose the correct answer
Azure Databricks
Azure SQL Data Warehouse
Azure Data Lake Analytics
HDInsight

Explanation

You should choose Azure Data Lake Analytics. This is an on-demand analytics job service that is optimized for processing big data sets stored in Azure Data Lake Store. Data is stored as relational data, and pricing is by batch job. Azure Data Lake Analytics supports Azure AD authentication.

You should not choose Azure SQL Data Warehouse. This is a relational store that supports processing for big data solutions but is billed by cluster hour rather than batch job. It supports Azure AD and SQL authentication.

You should not choose HDInsight, which is a managed Hadoop service. It supports Spark, Hive, Hive LLAP, and MapReduce for batch processing. It is a non-relational store that bills by cluster hour. The solution does support Azure AD authentication with Hive, Hive LLAP, and MapReduce, but not with Spark.

You should not choose Azure Databricks, which is an Apache Spark-based analytics platform. It is based on a non-relational data store, and pricing is based on Databricks Unit (DBU) (unit of processing capability per hour) and cluster hour. It does support Azure AD authentication.

References

[Choosing a batch processing technology in Azure](#)

[Develop large-scale parallel compute solutions with Batch](#)

[Batch processing](#)

Question116

You are designing a data processing solution that will process data from several diverse on-premises and cloud-based data stores.

You need to identify a data load solution to load data from the sources into an Azure SQL Data Warehouse.

What should you use?

Choose the correct answer

Azure Data Factory

Oozie on HDInsight

SQL Server Integration Services (SSIS)

Azure Data Migration Assistant

Explanation

You should use Azure Data Factory. Azure Data Factory supports data copy and custom transformations from a wide variety of data sources. It supports copying data from both on-premises and cloud-based data sources.

You should not choose Oozie on HDInsight. This solution does not have access to on-premises data.

You should not choose Azure Data Migration Assistant. This is designed to help with data transfer from and to SQL data sources, but it does not support other data sources.

You should not choose SSIS. SSIS supports on-premises SQL only.

References

Question117

You are designing a batch processing solution that will process large quantities of data daily and load the results into an Azure SQL Data Warehouse store.

You need to ensure high availability for your solution. You want to minimize costs and management overhead.

Which two actions should you perform? Each correct answer presents a complete solution.

Choose the correct answers

Deploy multiple instances in different regions configured for failover with the same batch and storage accounts in each region.

Deploy multiple instances in different regions configured for failover with different batch and storage accounts in each region.

Deploy multiple instances in different regions configured as a shared workload with different batch and storage accounts in each region.

Deploy multiple instances in the same region configured for failover with different batch and storage accounts for each instance.

Deploy multiple instances in different regions configured as a shared workload with the same batch and storage accounts in each region.

Explanation

You should configure your solution based on one of these two models:

- Deploy multiple instances in different regions configured for failover with different batch and storage accounts in each region.
- Deploy multiple instances in different regions configured as a shared workload with different batch and storage accounts in each region.

You need to deploy multiple instances in two or more regions, either in a failover cluster configuration or as a shared workload, to ensure high availability. This requires you to configure different accounts in each region.

You should not choose a solution that uses the same account across multiple regions. This is not supported as a solution.

You should not deploy a solution that relies on a single region. This does not provide high availability in case of a regional failure.

References

[Design your application for high availability](#)

Question118

You are using an Azure Databricks interactive cluster for batch processing. The cluster is configured for auto-termination.

You need to ensure that cluster configuration remains available for at least 180 days after termination, including cluster permissions.

What should you do?

Choose the correct answer

Configure the cluster for autostart.

Pin the cluster.

Clone the cluster after termination.

Manually terminate the cluster.

Explanation

You should pin the cluster. By default, the cluster configuration is retained for 30 days after termination. To keep an interactive cluster configuration longer than 30 days, you must pin the cluster. This makes it possible to retain the cluster configuration indefinitely.

You should not configure the cluster for autostart. This does nothing to increase the retention period for the cluster configuration. Configuring cluster autostart does not interfere with auto-termination and lets the cluster restart automatically for scheduled jobs.

You should not manually terminate the cluster. Whether terminated automatically or

manually, the cluster configuration retention period is still 30 days.

References

Question119

You are designing a Spark batch job to process daily log activity. The job should be scheduled to run each day. The job should display status information on the company intranet when it runs.

You need to choose the appropriate technologies to meet the requirements.

To answer, drag the appropriate technologies to each requirement. A technology may be used once, more than once, or not at all.

Drag and drop the answers

Livy

Beeline

Livy	
Beeline	
Azure API App	
Azure Logic App	

Explanation

You should Livy for job triggering and Beeline for job monitoring.

You can use Livy to submit jobs to be run on Spark so it provides an appropriate way to schedule the daily job runs. You can also use Livy to run interactive Spark shells.

You should use Apache Beeline for job monitoring. You can use Beeline to monitor Apache Spark jobs through Apache Hive queries. Beeline is designed to access Hive on HDInsight remotely over the internet and can be used for monitoring over an internal intranet.

You should not use Azure Logic Apps or Azure API Apps for triggering or monitoring. Neither is designed to support either of these activities. Azure Logic Apps are designed to help you schedule, automate, and orchestrate tasks. Azure API Apps provide a means for creating, hosting and using API's for cloud and on premises infrastructure. They do not support managing or monitoring Spark batches.

References

[Use Apache Spark REST API to submit remote jobs to an HDInsight Spark cluster](#)

[Use the Apache Beeline client with Apache Hive](#)

[Overview - What is Azure Logic Apps?](#)

Question120

You are designing an Azure Data Factory pipeline that will be used for processing data. The pipeline must support processing data that is stored in general-purpose standard Azure storage. The environment for data processing should be created on-demand and should be removed when processing is complete.

You need to recommend the appropriate transformation activity.

What should you recommend?

Choose the correct answer

Data Lake Analytics U-SQL activity

HDInsight Pig activity

Databricks Python

Databricks Notebook

Explanation

You should recommend HDInsight Pig activity. You can use Pig activity to execute Pig queries on an on-demand HDInsight cluster and remove the resources when they are no longer needed. HDInsight Pig activity supports general-purpose standard Azure storage.

You should not recommend Databricks Notebook or Databricks Python. Both run on your Databricks cluster and do not support creating an on-demand processing environment.

You should not recommend Data Lake Analytics U-SQL activity. This runs a U-SQL script on your existing Azure Data Lake Analytics cluster and does not support on-demand resources.

References

[Transform data in Azure Data Factory](#)

Question121

You are designing a data analytics solution for use with large data sets. The solution should use Azure Notebooks and support Notebook scheduling. The solution needs to provide visualization through Power BI visualization. The solution should also support cluster automation to run scheduled jobs on a new cluster and terminate the cluster when processing is complete.

Which solution should you recommend?

Choose the correct answer

Azure Databricks

Azure Batch

Azure Machine Learning (ML) Studio

Azure Stream Analytics

Explanation

You should recommend Azure Databricks. You can use Databricks to schedule a notebook. You would use a Azure Databricks job cluster, which lets you schedule workloads. Databricks also supports interactive clusters, which let you work with interactive notebooks. Azure Databricks support data visualization with Power BI Desktop.

You should not recommend Azure Batch. Azure Batch does not provide the features needed by the scenario. Azure Batch provides a way of implementing, scheduling, and managing batch jobs but it does not provide direct notebook support for the visualization needed.

You should not recommend Azure Stream Analytics. Azure Stream Analytics is not used for scheduled batch processing. It is designed to process high-volume streaming data performing real-time analytics.

You should not recommend Azure ML Studio. Azure ML Studio provides a collaborative, drag-and-drop environment for building, testing, and deploying

predictive analytics solutions.

References

[Transform data in Azure Data Factory](#)

[Transform data by running a Databricks notebook](#)

[Clusters](#)

[What is Azure Databricks?](#)

[Jobs](#)

[What is Azure Batch?](#)

[What is Azure Machine Learning Studio?](#)

[What is Azure Stream Analytics?](#)

Question122

You are developing a batch processing solution that collects input from IoT device events for analysis and presentation.

You need to identify the appropriate technologies for each of the remaining process steps.

What technology should you use for each process step? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

Data ingestion:

Transformation/Analysis:

Service storage:

Presentation:

	▼

	▼

	▼

	▼

What technology should you use for each process step? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

Data ingestion: HDInsight Kafka

Transformation/Analysis: Azure Databricks

Service storage: Cosmos DB

Presentation: Power BI

< Previous ☆ Next >

122 of 172

Explanation

You should choose HDInsights Kafka for data ingestion. This lets you collect data streaming in from IoT devices. You can also use IoT Hub or Azure EventHub for this purpose.

You should choose Azure Databricks for data transformation and analysis. Databricks

notebooks use a Kafka consumer API to read the data collected by Kafka.

You should choose Cosmos DB for service storage. Depending on the specific application, other storage options might also be supported, such as Azure SQL Database.

You should choose Power BI for presentation. Power BI is designed for analysis and presentation of warehouse data. This role might also be covered by a custom application.

References

[Ingestion and processing of real-time automotive IoT data](#)

[What is Apache Kafka in Azure HDInsight](#)

Question123

You want to test a batch processing solution that supports ingested data through Azure Data Factory batches, performs data analysis, and stores the result in Azure SQL Data Warehouse.

You need to suggest a solution that performs Spark analyses and creates and deletes clusters on-demand.

What should you use?

Choose the correct answer

<mark>Azure Databricks</mark>

Azure Kafka in HDInsight

Azure Data Lake Storage

Azure Cosmos DB

Explanation

You need to suggest Azure Databricks. You can have Spark run on Databricks. You can have Databricks create an automated cluster at the beginning of a batch job and remove the cluster when finished.

You should not suggest Data Lake Storage. Data Lake Storage is designed as a blob-based big data analytics environment. Data Lake Storage does not in itself provide an

environment for Spark job execution.

You should not suggest Azure Kafka in HDInsight. This is a streaming processing environment and is not used for batch processing with Spark jobs.

You should not suggest Azure Cosmos DB. This is a storage solution and not a batch processing environment.

References

[What is Azure Databricks?](#)

[Quickstart: Run a Spark job on Azure Databricks using the Azure portal](#)

[Clusters](#)

[What is Apache Kafka in Azure HDInsight](#)

[Welcome to Azure Cosmos DB](#)

[Introduction to Azure Data Lake Storage Gen2](#)

Question124

You are designing a batch processing solution that uses Azure Data Lake Storage as its data store. You estimate that the solution must support at least 4000 IOPS.

You need to select the appropriate disk type of your solution. You want to minimize costs.

What should you choose?

Choose the correct answer

Standard SSD

Standard HDD

Premium SSD

Ultra SSD

Explanation

You should choose standard SSD (solid state drive). Standard SSD supports up to 6000 IOPS across multiple drives.

You should not choose premium SSD or ultra SSD. Both exceed the scenario requirements but are more expensive than using standard SSD.

You should not choose standard HDD (hard disk drive). Standard HDD is limited to a maximum of 2000 IOPS and would not meet your requirements.

Hard disk drives are traditional drives that store data on rotating magnetic media. Solid state drives have no moving parts and use flash memory for storage. Access times are significantly less for SSD media than for HDD media.

References

[What disk types are available in Azure?](#)

Question125

You are asked to implement a batch processing system in Azure that will support R and Spark SQL. The users of the system will also require connected notebook support. Support for fast cluster startup times and autoscaling is also required.

You need to choose a batch processing solution to meet the requirements.

Which batch processing solution should you use?

Choose the correct answer

Azure Data Lake Analytics

HDInsight

Azure Synapse Analytics

Azure Databricks

Explanation

You should use Azure Databricks. Azure DataBricks is a spark-based platform that can readily be thought of as Spark as a service. It supports the required languages and includes fast startup, auto-termination, and autoscaling. Azure DataBricks also provides support for Scala and Jupyter based connected notebooks.

You should not use HDInsight. HDInsight is a managed service and supports multiple languages, but it is designed for Hadoop cluster management, not autoscaling.

You should not use Azure Synapse Analytics. Azure Synapse Analytics is a distributed system to query a distributed large data system. It is based on Massive Parallel Processing (MPP), and is used for running high-performance analytics.

You should not use Azure Data Lake Analytics. Azure Data Lake Analytics supports on-demand jobs for distributed processing of very large data sets, but does not support the required languages.

References

[Notebooks](#)

[Choosing a batch processing technology in Azure](#)

Question126 Recommend an Azure Data solution based on requirements

Your company has a large library of training videos that are currently stored on multiple on-premises file servers. You want to move the video files to cloud-based storage in a single data store to make is easier to support streaming video to offices in different locations.

You need to recommend an appropriate storage solution. You want to minimize storage costs.

What should you recommend?

Choose the correct answer

Azure Storage Blob

SQL Data Warehouse

Azure Cosmos DB

Azure File Storage

Explanation

You should recommend an Azure Storage Blob. Blob storage is an object storage solution that is optimized for storing large amounts of unstructured data such as text and binary files. It is well suited for files such as video files. Objects include the stored data, select metadata, and an unique ID that provides for access to the data.

You should not recommend SQL Data Warehouse. SQL Data Warehouse is a relational database management system (RDBMS) designed for storing large amounts of structured data as two-dimensional tables made up of rows and columns.

You should not recommend Azure File Storage. Storage cost for Azure File Storage per GB is over 50 times that of Azure Storage Blob. Azure File Storage is designed for storing files for access through the Storage Message Block (SMB) protocol. Azure Files lets you configure highly available file shares organized through data directories.

You should not recommend Azure Cosmos DB. Cosmos DB is a globally distributed multi-model database that supports various storage scenarios but is not well suited for unstructured data object storage and management.

References

Azure Storage Overview pricing

Introduction to Azure Blob storage

Choose the right data store

Criteria for choosing a data store

Question127 Design non-relational cloud data stores

You are designing a processing solution based on Azure Functions. The manufacturing process and quality control data is collected and uploaded directly into Azure Blob Storage. You must design a solution that:

- Optimizes the processing of new data as it is loaded into blob storage.
- Dynamically scales resources to meet processing requirements.
- Minimizes costs.

You need to recommend the hosting plan and trigger type for the Azure function.

Which hosting plan and trigger type should you recommend? To answer, select the correct configuration settings from the drop-down menus.

Choose the correct options

Hosting plan:

Trigger type:

[dropdown]

[dropdown]

Which hosting plan and trigger type should you recommend? To answer, select the correct configuration settings from the drop-down menus.

Choose the correct options

Hosting plan: Consumption plan

Trigger type: BlobTrigger
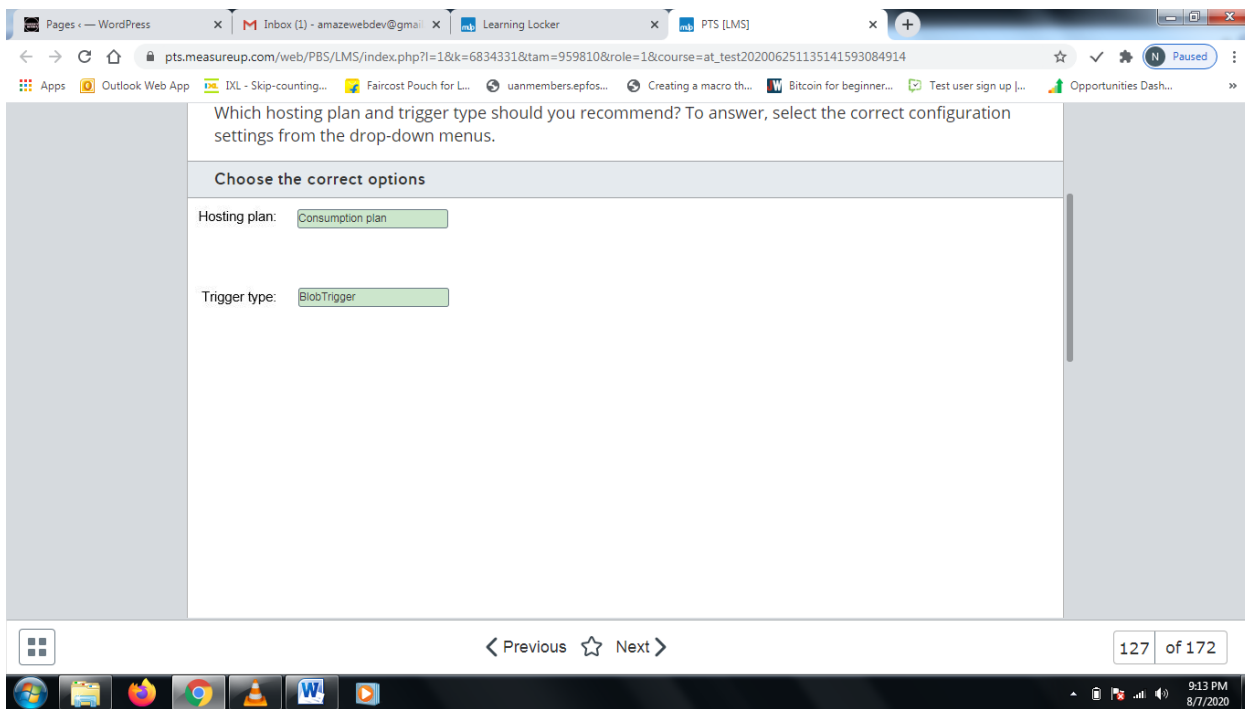
Explanation

You should create a function configured for the consumption hosting plan and a trigger type based on the BlobTrigger template. The BlobTrigger template is triggered when files are uploaded or updated in a blob. The consumption plan, by default, adds resources dynamically as needed for processing. This is a serverless hosting design, so you pay only for the time during which the function runs.

You should not choose the App Service plan, also known as the dedicated plan. You are charged for app service functions based on the same pricing as other App Service resources. The default is manual scaling by adding VM instances to the function. This

plan is more costly and less automated than the consumption hosting plan.

You should not choose the Premium plan. This plan is similar to the consumption plan but includes a fixed monthly cost, making it more expensive.

You should not configure a trigger type based on the EventGridTrigger template. This type of trigger is designed for event-based architectures and responds to events delivered to a subscription in Azure Event Grid.

You should not configure a trigger type based on the QueueTrigger template. This is used to respond to messages when they enter an Azure Storage queue.

References

[An introduction to Azure Functions](#)

[Create a function triggered by Azure Blob storage](#)

[Azure Functions scale and hosting](#)

Question128

You are designing a real-time processing solution by using Stream Analytics. You want to group streaming events that arrive at a similar time and filter out time periods where no data exists.

You need to choose the appropriate windowing function.

Which windowing function should you choose?

Choose the correct answer

Tumbling

Hopping

Sliding

Session

Explanation

You should use the session windowing function. Windowing functions are native to Stream Analytics, which is what you should use to analyze the data. The session

windowing function allows you to group streaming events that arrive at a similar time and filter out time periods where no data exists.

You should not use the tumbling windowing function. This function allows you to segment data into distinct time segments. This does not help in this scenario.

You should not use the hopping windowing function. A hopping window function looks backwards to determine when an event occurs.

You should not use the sliding windowing function. This function produces output only when an event occurs.

References

[Introduction to Stream Analytics windowing functions](#)

Question129

You are designing a real-time processing solution by using Stream Analytics. You want to count the number of weather reports that are received per time zone every minute.

You need to choose the appropriate windowing function.

Which windowing function should you choose?

Choose the correct answer
Hopping
Session
Tumbling
Sliding
Explanation

You should use the tumbling windowing function. This function allows you to segment data into distinct time segments. A tumbling windowing function is then applied to the data in each segment. An example is as follows:

SELECT TimeZone, Count(*) as Count
FROM WeatherStream TIMESTAMP BY ReportedAt
GROUP BY TimeZone, TumblingWindow(minute, 1)

You should not use the session windowing function. The session windowing function allows you to group streaming events that arrive at a similar time and filter out time periods where no data exists.

You should not use the hopping windowing function. A hopping window function looks backwards to determine when an event occurs.

You should not use the sliding windowing function. This function produces output only when an event occurs.

References

[Introduction to Stream Analytics windowing functions](#)

Question130

You are designing a real-time processing solution by using Stream Analytics. You want to count the number of weather reports received in a time zone during the last 30 seconds. If there are no weather reports, then no output should be displayed.

You need to choose the appropriate windowing function.

Which windowing function should you choose?

Choose the correct answer
Session
Tumbling
Hopping
Sliding

Explanation

You should use the sliding windowing function. This function produces output only when an event occurs. You can use the following query:

SELECT TimeZone, Count(*)
FROM WeatherStream TIMESTAMP BY ReportedAt
GROUP BY TimeZone, SlidingWindow(second, 30)

You should not use the session windowing function. Windowing functions are native to Stream Analytics, which is what you should use to analyze the data. The session windowing function allows you to group streaming events that arrive at a similar time, and filter out time periods where no data exists.

You should not use the tumbling windowing function. This function allows you to segment data into distinct time segments. This does not help in this scenario.

You should not use the hopping windowing function. A hopping window function looks backwards to determine when an event occurs.

References

[Introduction to Stream Analytics windowing functions](#)

Question131

You are designing a real-time processing solution by using Stream Analytics. Every five seconds you want to retrieve the number of weather reports by time zone during the last 30 seconds. It is fine for the events to overlap. Output should be generated even if there are no events.

You need to choose the appropriate windowing function.

Which windowing function should you choose?

Choose the correct answer
Tumbling
Sliding
Hopping
Session

Explanation

You should use the hopping windowing function. Windowing functions are native to Stream Analytics, which is what you should use to analyze the data. A hopping window function looks backwards to determine when an event occurs. Events can overlap. You can use the following query:

SELECT TimeZone, Count(*)

FROM WeatherStream TIMESTAMP BY ReportedAt
GROUP BY TimeZone, HoppingWindow(second, 30, 5)

You should not use the session windowing function. The session windowing function allows you to group streaming events that arrive at a similar time and filter out time periods where no data exists.

You should not use the tumbling windowing function. This function allows you to segment data into distinct time segments. Tumbling window events cannot overlap. In this scenario, the events should be allowed to overlap.

You should not use the sliding windowing function. This function produces output only when an event occurs.

References

[Introduction to Stream Analytics windowing functions](#)

Question132

You are designing a real-time processing solution by using Azure Databricks. Source data exists in an Azure blob storage account. You must be able to query the data by using the following query:

SELECT * FROM source LIMIT 10

You need to determine the type of notebook to create.

Which type of notebook should you create?

Choose the correct answer

Scala

SQL

Python

R

Explanation

You should use Python. To access data from blob storage, you must first write Python code to set the Azure storage container and credentials. Next, you must create a

DataFrame to represent a temporary SQL View. Finally, you can issue queries by using SQL.

You should not use SQL, R, or Scala. Although Databricks notebooks support these languages, you cannot use them to access blob storage.

References

Use Notebooks

Quickstart: Run a Spark job on Azure Databricks using the Azure portal

Question133

You are designing a real-time processing solution by using Azure Databricks. The data engineer that will implement the solution is familiar with object-oriented programming languages. The data engineer wants to take advantage of static type-checking.

You need to choose the language for the notebook that you are creating.

Which language should you use?

Choose the correct answer

Python

R

SQL

Scala

Explanation

You should use Scala. Scala is an object-oriented language that uses static type-checking.

You should not use SQL. SQL is an interpreted language used to query databases.

You should not use Python. Python is an object-oriented language that uses dynamic type-checking.

You should not use R. R supports both object-oriented functions and procedural functions. However, it uses dynamic type-checking.

References

[PySpark](#)

[SQL Guide](#)

[Introduction To Scala Programming](#)

[R programming on big data](#)

Question134

Your company manages a large fleet of delivery vehicles. All delivery vehicles have IoT sensors installed. Data from the sensors is collected and sent to Azure Event Hub. Data should go to Azure Power BI to generate real-time visualizations. This must be implemented through a high-speed, high-volume processing environment.

You need to recommend a solution.

What should you recommend?

Choose the correct answer

Azure HDInsight with Storm

Azure HDInsight with Spark Streaming

Apache Spark in Azure Databricks

<mark>Azure Stream Analytics</mark>

Explanation

You should recommend Azure Stream Analytics. Stream Analytics is a real-time analytics engine designed to process and analyze fast-streaming data from multiple resources, and it is the best solution for the type of solution called for in this scenario. Stream Analytics is designed to receive data from an IoT hub like Azure Event Hub and deliver the data to Power BI.

You should not use Azure HDInsight with Spark Streaming or Azure HDInsight with Storm. Neither is well-suited as a solution for this scenario. Both support receiving data from an IoT Hub, but neither is designed to deliver data to Power BI.

You should not use Apache Spark in Azure Databricks. This does not provide the type of processing environment required by this scenario. This solution would be appropriate for an interactive reporting application but not for the data visualization needed.

References

[Visualize real-time sensor data from Azure IoT Hub using Power BI](#)

[What is Azure HDInsight?](#)

[What is Azure Databricks?](#)

[What is Apache Spark in Azure HDInsight](#)

[What is Azure Stream Analytics?](#)

Question135

Your company uses an application that collects and processes scientific data and streams its output results. You are designing a cloud-based solution that performs additional real-time analysis of the data. Final results should be delivered to a real-time dashboard.

You need to design the logical flow for an Azure Stream Analytics solution.

Which three steps must you include in your logical flow? To answer, move the appropriate steps from the list of possible steps to the answer area and arrange them in the correct order.

Create a list in the correct order

Possible steps

Steps in order

- Build queries that output to Azure Data Lake Storage.
- Ingress data from the Azure Data Lake Storage instance.
- Send application data to an Azure Data Lake Storage instance.
- Send application data to an Azure Event Hub instance.
- Ingress data from the Azure Event Hub instance.
- Build queries that output to Power BI.

Explanation

Your solution should include the following steps in order:

1. Send application data to an Azure Event Hub instance.
2. Ingress data from the Azure Event Hub instance.
3. Build queries that output to Power BI.

You will build an Azure Stream Analytics job as a cloud solution. Stream Analytics is a real-time analytics engine designed to process and analyze fast-streaming data. Stream Analytics is designed to receive data from an IoT hub like Azure Event Hub and delivery the data to Power BI, so you need to send the data from an Azure Event Hub instance. You would then ingress the data from your Event Hub and build queries that output to Power BI as part of the Stream Analytics job.

There is no role for Azure Data Lake Storage in this solution. Azure Data Lake Storage is an enterprise-wide data storage solution that can be used with both interactive queries and batch processing. You can have a Stream Processing job deliver data to Azure Data Lake Storage for archiving and long-term storage, but it does not have a place in processing real-time presentation.

References

[Visualize real-time sensor data from Azure IoT Hub using Power BI](#)

[What is Azure Stream Analytics?](#)

[What is Azure Data Lake Storage Gen1?](#)

Question136

You have an Azure Function-based real-time streaming solution. Data for processing is uploaded into Azure Blob Storage, and new blobs should be processed with minimal delay. The solution should scale automatically to meet processing requirements, and function instances should be removed when they are no longer needed. Solution cost should be minimized.

You need to recommend solutions for the function hosting plan and trigger type.

What should you recommend? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

Hosting plan:

Trigger type:

[dropdown]

[dropdown]

You need to recommend solutions for the function hosting plan and trigger type.

What should you recommend? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

Hosting plan: Consumption plan

Trigger type: Blob

Explanation

You should create and deploy the Azure function in a Consumption plan and use a blob trigger. A Consumption plan creates function instances as needed and removes the instances when no longer needed. The solution can scale automatically to meet processing requirements. The only plans that meet the scalability requirement are the Consumption plan and Premium plan. Of these two plans, the Consumption plan has lower costs. Billing is based on the number of executions, execution time, and memory used.

You should not recommend the App service plan (also known as the Dedicated plan).

This plan does not support automatic scaling. Function instances are paused rather than removed after a time-out period.

You should not recommend the Premium plan. As of this time, the Premium plan is in preview release and is not recommended for use in a production environment. Also, this plan is more expensive than the Consumption plan.

References

[Create a function triggered by Azure Blob storage](#)

[Determine the best trigger for your Azure function](#)

[Azure Functions scale and hosting](#)

Question137

You are designing a real-time IoT data analysis solution using Azure Stream Analytics. This will be a mission critical application.

You need to ensure that your solution remains available during Azure service plan updates even if an update error occurs.

Which two actions should you recommend? Each correct answer presents part of the complete solution.

Choose the correct answers

Deploy identical jobs to two unpaired regions.

Deploy identical jobs to three unpaired regions.

Monitor jobs in both regions for failures.

Monitor the jobs in the primary region for failure.

Deploy identical jobs to both regions in a paired region.

Explanation

You should take the following two actions when implementing your solution:

- Deploy identical jobs to both regions in a paired region.
- Monitor jobs in both regions for failures.

Service updates are issued for Stream Analytics at least weekly and can occur more often. When you deploy your jobs in two paired regions, Stream Analytics guarantees that updates occur to the regions in separate batches. This gives you time to respond if an error occurs during update and continue operations. You need to monitor the jobs in both regions to determine if a failure occurs.

You should not deploy in either two or three unpaired regions. There is no guarantee that updates will not be applied at the same time, resulting in loss of availability if there is a problem.

You should not monitor the primary region only. You must monitor both regions.

References

[What is Azure Stream Analytics?](#)

[Guarantee Stream Analytics job reliability during service updates](#)

[Business continuity and disaster recovery (BCDR): Azure Paired Regions](#)

[Get started with Azure Stream Analytics to process data from IoT devices](#)

Question138

Your company uses IoT devices to collect data about delivery vehicles. Data is identified by vehicle number and time. For security reasons, data should expire and automatically be removed from storage after one week. Management overhead should be kept to a minimum.

Your data solution must support on-demand real-time reporting. Reports must execute as quickly as possible.

You need to recommend the Cosmos DB features that meet these requirements.

What should you recommend? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

Report optimization:

Data security requirement:

[▼]

[▼]

What should you recommend? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

Report optimization: [Cosmos DB indexes]

Data security requirement: [Cosmos DB TTL]

‹ Previous ☆ Next ›

138 of 172

Explanation

You should use Cosmos DB indexes to optimize query performance for on-demand reporting and generate your reports as quickly as possible. Indexing lets you retrieve data more quickly and efficiently during queries.

You should use Cosmos DB TTL (Time to Live) to meet the data security requirement. You can configure Cosmos DB TTL to expire and delete data after a configured time period. Expiration can be configured at the container or the individual item basis. Configuring expiration at the container helps to minimize management requirements.

You should not use Cosmos DB change feed. Change feed is used to automatically trigger Azure functions based on Cosmos DB events. It is not used with on-demand reporting.

You should not use Cosmos DB transactions. Transactions are used to ensure that an operation executes as a single logical unit of work to ensure Atomicity, Consistency, Isolation, and Durability (ACID) property guarantees.

You should not use Cosmos DB triggers. Triggers are used to perform operations on a Cosmos DB item either when creating an item (pre-trigger) or after an item operation (post-trigger). Triggers do nothing to meet the requirements.

References

[Time to Live (TTL) in Azure Cosmos DB](#)

[Indexing in Azure Cosmos DB - Overview](#)

[Serverless event-based architectures with Azure Cosmos DB and Azure Functions](#)

[Stored procedures, triggers, and user-defined functions](#)

Question139

You are developing a real-time processing solution to detect and report telecommunications fraud activity. You are using Azure Stream Analytics.

You need to create an appropriate data stream input to support the solution.

What should you use?

Choose the correct answer

Kafka

Azure SQL Database

Azure Event Hubs

Azure Cosmos DB

Explanation

You should use Azure Event Hubs. Event Hubs support real-time data ingestion for real-time processing. Azure Stream Analytics also supports IoT Hub and storage blobs for data ingestion.

You should not use Azure SQL Database. SQL data is not supported for data stream ingestion. You can use SQL Database for reference data input that contains static or near-static reference data.

You should not use Azure Cosmos DB. Azure Cosmos DB is not supported as a data stream input for Azure Stream Analytics.

You should not use Kafka. Kafka can be used for streaming data, but it is not supported by Azure Stream Analytics.

References

Get started using Azure Stream Analytics: Real-time fraud detection

Understand inputs for Azure Stream Analytics

Question140 Design non-relational cloud data stores

You are designing a Cosmos DB solution to support a global application. Your solution should minimize the following:

- Recovery point objective (RPO) - The time period of updates that you can afford to lose in case of regional failure.
- Recovery time objective (RTO) - The time to fully recover an application.

You need to configure a single master replication model across multiple regions to meet these requirements.

Which consistency level should you configure?

Choose the correct answer

Eventual

Consistent prefix

Strong

Session

You should configure strong consistency. This is the strongest consistency level support by Cosmos DB and guarantees that read requests return the most recently committed version of an item. This consistency level guarantees an RPO of 0 and an RTO of less than 15 minutes.

Cosmos DB supports the following consistency levels, from strongest consistency to most relaxed:

- Strong
- Bounded staleness
- Session
- Consistent Prefix
- Eventual

You should not configure Session, Consistent prefix, or Eventual consistency. Each of these will give you an RPO of less than 15 minutes and an RTO of less than 15 minutes, so RPO is not minimized.

The RTO improves with multiple master replication with Bounded staleness, Session, Consistent Prefix, and Eventual with a value of 0.

You should avoid single region configurations. Single or multiple master replication with a single region model and any consistency level results in an RPO of up to 240 minutes and an RTO of up to one week.

References

Consistency, availability, and performance tradeoffs

Consistency levels in Azure Cosmos DB

Question141

You are designing a real-time processing solution using Azure Streaming Analytics. Data is loaded into an Azure Blob Storage container every minute for immediate processing automated through the use of Azure Functions. Your company has IoT sensors throughout its manufacturing facility to collect information for quality analysis. Data should be placed in

long-term storage after processing for later review.

You use Azure Blob Storage for data ingested and create a new blob for each upload. The blob is deleted after the data is processed. You create the functions using the App Service hosting plan and a blob trigger. Processed data is written to an Azure Cosmos DB database and sent to Power BI to support a real-time dashboard.

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

Statement	Yes	No
The streaming data input is appropriate for the application and allows for automation.	○	○
Azure Functions will be able to scale automatically to meet processing needs, and functions are removed when not needed.	○	○
You can both store the processed data in Cosmos DB and send the data to Power BI.	○	○

Explanation

The streaming data input is appropriate for the application and allows for automation. You can use Blob Storage, Azure Event Hubs, or IoT Hubs for streaming data ingestion. You can use a blob trigger to automatically launch function instances when a blob is added or updated.

Azure Functions will not be able to scale automatically to meet processing needs, and functions are removed when not needed. The App Service hosting plan requires manual scaling. You must configure either the Consumption plan or Premium plan to support automatic scaling.

You can both store the processed data in Cosmos DB and send the data to Power BI. Output to Power BI is supported. You can use multiple options for long-term storage, including Cosmos DB, Azure SQL Database, Azure Blob Storage, and other storage options.

References

[What is Azure Stream Analytics?](#)

[Understand inputs for Azure Stream Analytics](#)

[Understand outputs from Azure Stream Analytics](#)

[Determine the best trigger for your Azure function](#)

[Azure Functions scale and hosting](#)

Question142

You are designing a real-time processing solution by using Azure Databricks. The source data is in a local CSV file. You want to use SQL to manually query the CSV file and display a pie chart by running a cell.

You need to create the necessary resources.

Which resources should you create in sequence? To answer, move the appropriate resources from the list of possible resources to the answer area and arrange them in any order.

Create a list in any order

Possible resources

Resources

- Job
- Storage account
- Event Hub
- Databricks workspace
- Databricks cluster
- Databricks notebook

Explanation

You should create a Databricks workspace. A workspace allows you to organize all of the assets used in your solution, including notebooks.

HYou should also create a Databricks notebook. A notebook contains the code that you actually run. Notebooks can use Scala, R, SQL, or Python.

You should also create a Databricks cluster. A cluster is the environment that allows notebooks to run.

You should not create a storage account. In this scenario, the source data is in a local CSV file that you can upload to Databricks.

You should not create a job. A job allows you to run a notebook immediately or on a schedule. In this scenario, you want to manually query the data by running a cell.

You should not create an Event Hub. An Event Hub allows you to stream millions of events per second to the cloud. It is not necessary for running a Databricks cluster.

References

Introduction to Azure Blob storage

Azure Event Hubs — A big data streaming platform and event ingestion service

What is Azure Databricks?

Question143

You are a data architect for your company. You create an Azure SQL Database server with one database. You keep the default firewall settings.

You need to determine how the database can be accessed.

To answer, select Yes if the statement is true. Otherwise, select No.

Statement	Yes	No
You can add data to the database by using the Azure portal.	◌	◌

You can read data from the database by using SQL Server Management Studio (SSMS) on your local computer.	○	○
You can update data in the database from an Azure function.	○	○

Explanation

You can add data to the database by using the Azure portal. By default, you can only access the database from within the Azure portal. The default firewall rules do not allow access to the database from other Azure services or from the internet.

You cannot read data from the database by using SSMS on your local computer. The default firewall rules do not allow access to the database from other Azure services or from the internet.

You cannot update data in the database from an Azure function. The default firewall rules do not allow access to the database from other Azure services or from the internet.

References

Azure SQL Database and Azure SQL Data Warehouse IP firewall rules

Question144

You are creating an Azure SQL Database managed instance that must be accessed by an on-premises application that is load-balanced across four servers. A data engineer must be able to access the database by using SQL Server Management Studio (SSMS) from a personal laptop computer.

You need to configure the environment so that both the application and the data engineer can access the database.

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

Statement	Yes	No
You should create a point-to-site connection for the application.	○	○
You should create a site-to-site connection for the application.	○	○
You should create a point-to-site connection for the developer.	○	○

Explanation

You should create a site-to-site connection for the application. A site-to-site connection is a virtual private networking (VPN) connection that allows you to connect an on-premises private network to Azure. An Azure SQL Database managed instance uses a private IP address only. You cannot connect to a public endpoint.

You should create a point-to-site connection for the developer. A point-to-site connection is a VPN connection that allows you to connect a single computer to Azure. This is necessary because an Azure SQL Database managed instance accepts connections to a private IP address only.

You should not create a point-to-site connection for the application. You should create a site-to-site connection.

References

Create a Site-to-Site connection in the Azure portal

Connect your application to Azure SQL Database managed instance

You are designing a cloud data solution for your company. You create an Azure SQL Database server with a single database. The database will be used by a developer to test queries.

You need to ensure that the developer can successfully access the database.

What should you do?

Choose the correct answer

Create a server-level firewall rule and ensure that inbound TCP port 1433 is open on the developer's computer.

Create an ExpressRoute connection between Azure and the developer's computer.

Create a server-level firewall rule and ensure that outbound TCP port 1433 is open on the developer's computer.

Create a point-to-site connection between Azure and the developer's computer.

Explanation

You should create a server-level firewall rule and ensure that outbound TCP port 1433 is open on the developer's computer. Port 1433 is the default port that Azure uses to allow inbound connections to Azure SQL Database. By default, public access to Azure SQL Database is disallowed. The server-level firewall rule allows you to configure the internet IP addresses that are allowed to connect to the database.

You should not ensure that inbound TCP port 1433 is open on the developer's computer. This exposes the developer's computer to malicious attacks.

You should not create a point-to-site connection between Azure and the developer's computer. A point-to-site connection is necessary when you need to access Azure SQL Database managed instance from a computer. This connection is a virtual private networking (VPN) connection that allows you to access the private IP address used by Azure SQL Database managed instance.

You should not create an ExpressRoute connection between Azure and the developer's computer. ExpressRoute allows you to extend your on-premises private network to Azure over a connectivity provider.

References

Configure a Point-to-Site VPN connection to a VNet using native Azure certificate authentication: Azure portal

ExpressRoute overview

Connect your application to Azure SQL Database managed instance

Azure SQL Database and Azure SQL Data Warehouse IP firewall rules

Question146

You run the following SQL statement on an Azure SQL Database Single Database instance:

EXECUTE sp_set_database_firewall_rule N'Sample Rule', '0.0.0.0', '0.0.0.0';

You need to determine what sources are allowed to access the database.

What should you conclude?

Choose the correct answer

All Azure services

Only the Query Editor in the Azure portal

The entire internet

Only your local computer

Explanation

You should conclude that all Azure services can access the database.
The sp_set_database_firewall_rule stored procedure creates a database-level firewall rule. The first parameter specifies the name of the rule. The second and third parameters specify the start and end IP addresses, respectively.

You should not conclude that only your computer can access the database. To do this, you must add the client IP address of your computer's connection to Azure. This is not necessarily the IP address in your IP configuration due to network address translation (NAT).

You should not conclude that only the Query Editor in the Azure portal can access the database. This would be true if there are no server-level and database-level firewall rules set.

You should not conclude that the entire internet can access the database. It is not

possible to allow the entire internet to access the database.

References

[Azure SQL Database and Azure SQL Data Warehouse IP firewall rules](#)

[Azure SQL Database and SQL Data Warehouse access control](#)

Question147

You create an Azure storage account that contains a table and a blob container. You want to allow two IP addresses the ability to read from the table. The users of the IP addresses must not be able to modify or delete the storage account. They also must not be able to read blobs in the blob container.

You need to provide access to the storage account to the clients of the IP addresses.

What should you provide?

Choose the correct answer

Secondary access key

Service shared access signature (SAS)

Account shared access signature (SAS)

Primary access key

Explanation

You should provide a service SAS. A service SAS allows you to give clients access to a storage account resource. In this scenario, the resource is the Azure table. An SAS is a URL that embeds the resource, permissions, IP range, protocol, access times, and token.

You should not provide an account SAS. An account SAS allows you to give clients access to multiple resources in one or more storage services. In this scenario, you only want to allow the clients to access a table, not the blob container.

You should not provide the primary and secondary access keys. These keys allow full access to a storage account, including the ability to modify and delete resources in it.

References

Question148

You are a data architect for a record distributor. You are designing a solution that allows record labels to submit songs to the cloud. Each record company must be able to submit and delete its own songs only. You plan to create a blob container for each record label. An on-premises application will be used by company managers to browse, download, and delete songs.

You need to design a solution so that the songs are secure.

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

Statement	Yes	No
You should provide the secondary access key to each record company.	○	○
You should provide a service shared access signature (SAS) to each record company.	○	○
You should provide the primary access key to the application.	○	○

Explanation

You should not provide the secondary access key to each record company. The primary and secondary access keys allow full access to the blob storage account.

You should provide a service SAS to each record company. This ensures that each record company can upload and delete songs in its own container only.

You should provide the primary access key to the application. This allows the

application to browse, download, and delete the songs. A primary and secondary access key are created when you create the storage account. After creation, you can have the storage account generate new keys.

References

[Manage storage account settings in the Azure portal](#)

[Grant limited access to Azure Storage resources using shared access signatures (SAS)](#)

Question149 Recommend an Azure Data solution based on requirements

You are designing an application that will use Azure SQL Database as its storage backend. The application extracts data from the database and generates text documents that should be made available to users through Storage Message Block (SMB) shares from a cloud-based location.

You need to choose an appropriate cloud-based storage solution.

Which data storage type should you choose?

Choose the correct answer

Azure Queues

<mark>Azure Files</mark>

Azure Blob

Azure Tables

Explanation

You should choose Azure Files. Azure Files lets you configure highly available file shares that are accessible through the SMB protocol. Shares can be configured as cloud-based or on-premises shares through a URL and allow for read and write access to the data. Data access requires a valid URL and shared access signature (SAS) token.

Azure Blobs, Azure Queues, and Azure Tables do not support SMB access.

You should not choose Azure Blobs. You would use Azure Blobs for storing text and binary objects when you have requirements for storing massive amounts of unstructured data.

You should not choose Azure Queues. Azure Queues provide a reliable message store to handle messaging between applications.

You should not choose Azure Tables. Azure Tables has been integrated into Azure Cosmos DB and supports schemaless storage of large amounts of structured data in throughput-optimized tables.

References

[Choose the right data store](#)

[Azure Files](#)

[Introduction to Azure Storage](#)

Question150

Your network is configured as an Active Directory Domain Services (AD DS) domain. All on-premises user accounts also have synchronized Azure Active Directory (Azure AD) users.

The database server Cust_SQL is used as a staging database to collect data relating to customer sales activity. A batch data solution reads data from Cust_SQL, performs extract, transform, and load, and then passes the result to Azure Power BI for sales analysis.

You need to protect access to the data by limiting access to a non-standard port.

Which port should you choose, and what tool should you use to configure the SQL database? To answer, select the appropriate option from the drop-down menus.

Choose the correct options

Configuration tool:

Network port:

select the appropriate option from the drop-down menus.

Choose the correct options

Configuration tool: SQL Server Configuration Manager

Network port: 8080

⟨ Previous ☆ Next ⟩   150 of 172
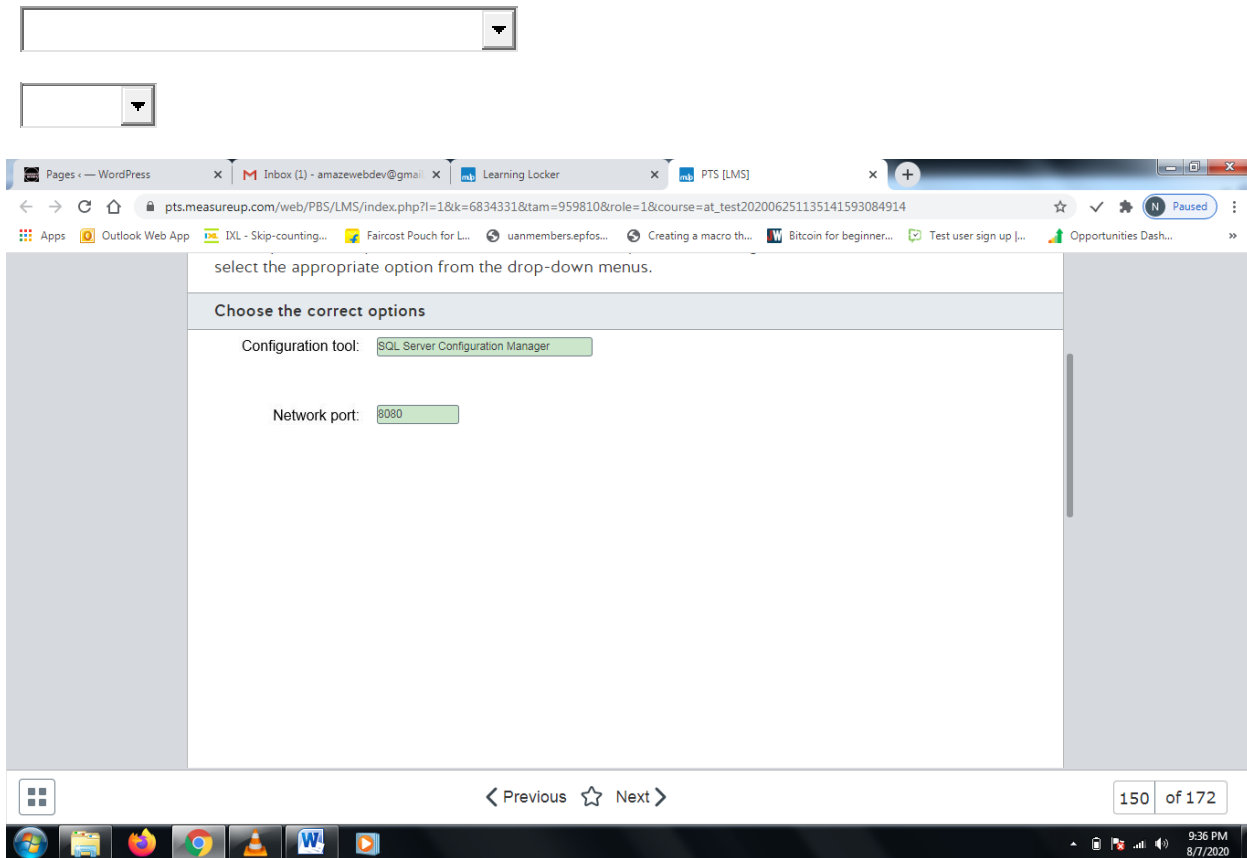
Explanation

You should use SQL Server Configuration Manager to configure SQL Server to listen on a specific port. This gives you a way to override the default port assignment. Some companies use this as a security measure to help limit access to SQL Server, usually in combination with other security measures.

You should not use Azure Data Studio. Azure Data Studio can be used for managing SQL Server but does not provide a way to alter the default port.

You should not use SQL Server Data Tools. These are tools that work with Visual Studio to provide a powerful development environment for SQL Server Database, Azure SQL Database, and Azure SQL Data Warehouse.

You should not use SQL Server Management Study (SSMS). SSMS is used for configuring, managing, and administering SQL Server, but it does not provide a way to change the default communication port.

You should choose 8080 as the alternate port. Port 8080 is occasionally (but rarely) used as an alternate port for HTTP communications and is typically available. You should not specify port 1433. This is the default SQL Server port. You should not use

port 443. This is the port for HTTPS for communication with secure web servers and could result in conflicts.

References

[Configure a Server to Listen on a Specific TCP Port](#)

[SQL Server Management Studio](#)

[List of TCP and UDP port numbers](#)

[SQL query and management tools for SQL Server](#)

Question151

You are designing a new data project. One of the project requirements is to allow access to one set of data without requiring any authentication. Someone accessing the data must already know the data store they want to access and should be unable to enumerate available stores.

You need to determine which Azure storage offering to use.

Which storage offering supports this access option?

Choose the correct answer

Azure Files

Azure Queues

Azure Tables

Azure Blobs

Explanation

You need to configure anonymous public read access. Azure Blobs is the only storage offering that offers support for anonymous public read access. Anonymous access is configured through container configuration settings. The following options are supported:

- No public read access - This is the default setting for all new containers.

- Public read access for blobs only - Anonymous access is granted to blobs within the container, but no access is given to container data, and anonymous users are prevented from enumerating the blobs in the container.
- Public read access for container and its blobs - Anonymous access is granted to blob and contain data, including the ability to enumerate the blobs within a container.

Azure Files, Azure Queues, and Azure Tables do not support and cannot be configured for anonymous access.

References

[Authorizing access to Azure Storage](#)

[Manage anonymous read access to containers and blobs](#)

Question152

You are reviewing access security requirements for new data storage solutions. You need to recommend an authorization option that supports:

- Azure Blob
- Azure Files (SMB)
- Azure Files (REST)
- Azure Queues
- Azure Tables

Which authorization option should you recommend?

Choose the correct answer

Azure Active Directory (Azure AD)

Anonymous public read access

Shared access signature (SAS)

Shared Key (storage account key)

Explanation

You should recommend Shared Key (storage account key). This is the only authentication option that supports all of the Azure storage offerings listed. Every access must be signed using the storage account access key. Azure generates two 512-bit storage access keys when you create a storage account.

You should not recommend SAS. SAS does not support Azure Files (SMB). The other listed offerings are supported. SAS is used to provide limited delegated access to a storage account. You can configure the SAS to be valid for a limited time period.

You should not recommend Azure AD. Azure AD does not support Azure Files (REST) or Azure Tables. Azure AD is recommended when possible because it is easier to manage and provides better security than other methods. Azure AD also supports the use of role-based access control (RBAC).

Anonymous public read access is supported for Azure containers and blobs only. It is recommended that its use be very limited.

References

[Authorizing access to Azure Storage](#)

[Authorize with Shared Key](#)

[Manage storage account settings in the Azure portal](#)

Question153

You are developing an Azure data solution for a non-profit organization that works with United States military veterans. Users should be able to view only the last four digits of a veteran's social security number (SSN) when querying records.

You need to configure an Azure SQL database to provide this protection.

Which security feature should you use?

Choose the correct answer

Dynamic data masking (DDM)

Always Encrypted

Row-level security (RLS)

Transparent Data Encryption (TDE)

Explanation

You should use DDM. DDM lets you limit the exposure of sensitive data to users who do not need full access to the data. You can use this to enable displaying the last four

digits only, using masking characters for the remaining digits.

You should not use Always Encrypted, also known as encrypted-in-use. This is used to protect specific database columns from access. This includes blocking access to database administrators or other database users. Data is decrypted for processing by client applications only.

You should not use RLS. RLS is used to limit access to specific rows within a database based on the access context.

You should not use TDE. TDE encrypts the entire database and also prevents offline access of data through raw files or backup files. All Azure SQL Databases are encrypted by default.

References

An overview of Azure SQL Database security capabilities

Azure database security best practices

Question154

You migrate an Azure data solution from an on-premises database to an Azure SQL Database managed instance. You discover that data can be accessed in clear text when it is accessed offline from database backups.

You need to ensure that the database is encrypted and cannot be accessed from backups. You need to minimize the management effort necessary to implement and maintain the solution.

What should you configure?

Choose the correct answer

Dynamic data masking (DDM)

Transport Layer Security (TLS)

Transparent Data Encryption (TDE)

Always Encrypted

Explanation

You should configure TDE. TDE encrypts the entire database and prevents offline access of data through raw files or backup files. It also encrypts transaction log files. TDE be manually enabled for older Azure SQL Database, Azure SQL managed instance, or Azure SQL Data Warehouse databases.

You should not configure TLS. TLS is used for encryption in transit to ensure that data is encrypted when data is transferred between a client and server. It does not ensure at rest encryption or that backups are encrypted.

You should not configure Always Encrypted. This is used to protect specific database columns from access and must be enabled individually for each protected column. This includes blocking access to database administrators or other database users. Data is decrypted for processing by client applications only.

You should not configure DDM. DDM lets you limit the exposure of sensitive data to users who do not need full access to the data so that only select data is displayed.

References

An overview of Azure SQL Database security capabilities

Transparent data encryption for SQL Database and Data Warehouse

Question155

You are configuring access to Azure Blob storage.

You need to configure fine-grained access control through role-based access control (RBAC).

Which authorization option should you choose?

Choose the correct answer

Access key

Shared key

Azure Active Directory (AD)

Account shared access signature (SAS)

Explanation

You should configure Azure AD authorization. Azure AD authorization provides support for RBAC for Azure Blobs and Azure Queues. This gives you a way to configure fine-grained client access control to a storage account.

You should not configure shared key. Shared key (storage account key) authorization does not support RBAC controls.

Access key is not an authorization method but is used with shared key authorization. A shared key access request must be signed using the storage account access key. Azure generates two 512-bit storage access keys when you create a storage account.

Account SAS does not support RBAC. SAS is used to provide limited delegated access to a storage account. You can configure the SAS to be valid for a limited time period. Access is secured with the storage account key.

References

[Authorizing access to Azure Storage](#)

[Authorize access to Azure blobs and queues using Azure Active Directory](#)

Question156

You are implementing an Azure SQL Data Warehouse to support a data application. You use Azure Active Directory (Azure AD) for user authentication. You are assigning roles based on job requirements:

- Customer account managers must be able to create new databases.
- IT support personnel must be able to assign users to databases.

Role assignments should follow the principle of least privilege.

You need to make role membership assignments.

How should you assign roles? To answer, select the best role from the drop-down menu.
Choose the correct options

Customer account manager:

IT support:

[dropdown ▼]

[dropdown ▼]

How should you assign roles? To answer, select the best role from the drop-down menu.

Choose the correct options

Customer account manager: dbmanager

IT support: db_accessadmin

‹ Previous ☆ Next ›   156 of 172

Explanation

You should assign the dbmanager role to customer account managers. This role is supported for Azure SQL Database and Azure SQL Data Warehouse, but not for on-premises SQL Server. Members of this role can create new databases. This does not necessarily give the users permission to access the databases they create.

You should assign the db_accessadmin role to IT support personnel. Members of this

role can add or remove user access to a database.

You should not assign the db_securityadmin role. Members of this role can modify membership for custom roles and assign permissions. This is not required by the scenario.

You should not assign the dbcreator role. This is supported for on-premises SQL Server instances only.

You should not assign the loginadmin role. This is supported for Azure SQL Database and Azure SQL Data Warehouse, but not for on-premises SQL Server. Members of this role can create and delete logins but they cannot create databases or assign users to databases.

You should not assign the sysadmin role. Members of this role can perform any activity on the server. This would enable the users to meet their job requirements but it would assign more permissions than needed, so the principle of least privilege requirement would not be met.

References

[Server-Level Roles](#)

[Database-Level Roles](#)

[Managing databases and logins in Azure SQL Database](#)

Question157 Design non-relational cloud data stores

You are designing storage support for an HDInsight cluster with access to Azure Data Lake Gen1 storage. Your solution must support copying data between regions from Azure Storage Blob to Azure Data Lake Gen1. The solution must support copying update deltas only.

You need to recommend a copy solution.

What should you recommend?

Choose the correct answer

Distcp

AdlCopy

Sqoop

Azure Data Factory

Explanation

You should use Distcp to copy data between two regional locations. This command supports copying from a blob as a source into Azure Data Lake Gen1 and is generally accepted as the fastest way to move big data. Distcp supports copying changed information only, known as updating deltas between the locations.

You should not use AdlCopy. AdlCopy is a supported copy program, but it does not support copying between regions and it does not support delta updates.

You should not use Azure Data Factory. Azure Data Factory is designed to support hybrid extract-transform-load (ETL), extract-load-transform (ELT), and data integration projects rather than acting as a simple copy utility. Azure Data Factory does not support delta updates.

You should not use Sqoop. The Sqoop copy utility is used to copy data from Azure SQL Database to Azure Data Lake Gen1. Its use would be inappropriate in this scenario.

References

What is Azure Data Lake Storage Gen1?

Best practices for using Azure Data Lake Storage Gen1

Tuning Azure Data Lake Storage Gen1 for performance

Using Azure Data Lake Storage Gen1 for big data requirements

Use Distcp to copy data between Azure Storage Blobs and Azure Data Lake Storage Gen1

Copy data between Data Lake Storage Gen1 and Azure SQL Database using Sqoop

What is Azure Data Factory?

Your company has an Active Directory Domain Services (AD DS) on-premises domain and an Azure Active Directory (Azure AD) domain. The domains are not kept synchronized.

You develop an Azure data solution that uses an Azure SQL Database single instance for data storage. You want to use Azure authentication to control database access.

You need to configure support for Azure AD authentication. You want to minimize the administrative overhead necessary to accomplish this.

What should you do?

Choose the correct answer

Configure automatic synchronization between AD DS and Azure AD.

Create an Azure AD admin account on the Azure SQL Database instance.

Disable SQL Server authentication on the Azure SQL Database instance.

Migrate the Azure SQL Database to an Azure SQL Database managed instance.

Explanation

You should create an Azure AD admin account on the Azure SQL Database instance. This is required before you can configure access for any Azure AD user accounts to the database. You can configure a user or security group account as the Azure AD admin account.

You should not migrate the Azure SQL Database to an Azure SQL Database managed instance. This action is supported but is not necessary to support Azure AD authentication.

You should not configure automatic synchronization between AD DS and Azure AD. Azure AD user accounts are used for authentication, whether or not they are synchronized with AD DS. This is optional, but it not a required action.

You should not disable SQL Server authentication on the Azure SQL Database instance. The instance can support both Azure AD authentication and SQL Server authentication at the same time.

References

Azure database security best practices

Question159

You use various Azure storage authentication methods throughout your organization, often configuring different authentication methods for the same data store to meet different requirements.

You must regenerate an access key for an Azure Blob store. You need to identify which authorization types can be impacted by this action.

Which three authorization methods will this impact? Each correct answer presents part of the solution.

Choose the correct answers

<mark>Service shared access signatures (SAS)</mark>

Anonymous public read access

<mark>Shared key</mark>

User delegation shared access signatures (SAS)

Azure Active Directory (Azure AD)

<mark>Account shared access signatures (SAS)</mark>

Explanation

The following authentication methods are based on the storage access key:

- Shared key
- Service SAS
- Account SAS

These can be impacted when you regenerate an access key. Two 512-bit storage account access keys are generated when you first create a storage account. Extreme care should be taken to protect these account keys to prevent unauthorized access. Microsoft recommends regenerating access keys periodically as part of your ongoing security procedures.

The following authentication methods do not use the storage access key:

- User delegation SAS
- Azure AD
- Anonymous public read access

These are not impacted when you regenerate the access key. User delegation SAS uses a user delegation key that you request for each user. Azure AD uses an Azure AD account. Anonymous access does not use authentication keys.

References

Authorizing access to Azure Storage

Authorize with Shared Key

Grant limited access to Azure Storage resources using shared access signatures (SAS)

Manage storage account settings in the Azure portal

Question160

Your company stores sensitive customer data in a table named Customer in an Azure SQL Database managed instance. You want to protect specific columns in the Customer table. Data in the other tables is not sensitive. An on-premises application accesses the database.

You need to ensure that the customer data remains protected while in transit and at rest, while still allowing the application to decipher it.

What solution should you recommend?

Choose the correct answer

Azure Disk Encryption

Always Encrypted

Transparent Data Encryption (TDE)

Column-level encryption (CLE)

Explanation

You should recommend Always Encrypted. This allows you to encrypt specific columns within a table. The data remains encrypted in transit between the client application and the database server. This requires .NET 4.6 or greater. Encryption and decryption operate at the client.

You should not recommend Azure Disk Encryption. This technology encrypts virtual machine (VM) disks, which is not suitable in this scenario.

You should not recommend TDE. This allows you to encrypt the entire database on the server. However, the actual data does not remain encrypted while in transit.

You should not recommend CLE. This technology uses a symmetric key to encrypt data at the column level inside a database. This does not allow the client application to decipher it.

References

[Always Encrypted (Database Engine)](#)

[Azure encryption overview](#)

[Azure Storage security overview](#)

Question161 Design non-relational cloud data stores

You are a data architect for your company. You plan to deploy two Azure Cosmos DB accounts. The accounts must meet the following requirements:

- CosmosDB1 - You want to achieve the highest availability and the lowest latency.
- CosmosDB2 - You want to ensure that all client applications always read the same values.

You need to choose the most appropriate consistency level for each account.

What consistency levels should you use? To answer, drag the appropriate consistency level to each Cosmos DB account. Each consistency level may be used once, more than once, or not at all.

Drag and drop the answers

strong

eventual

strong

eventual

session

consistent prefix

bounded staleness

Drag and drop the answers

CosmosDB1   eventual

CosmosDB2   strong

strong   eventual   session

consistent prefix   bounded staleness

∧ Explanation

You should use eventual consistency for CosmosDB1. This consistency level allows you to achieve the highest availability and the lowest latency. Replicas do not necessarily contain the same data. However, the replicas will eventually converge. This is the weakest consistency level, but it offers the highest availability.

Explanation

You should use eventual consistency for CosmosDB1. This consistency level allows you to achieve the highest availability and the lowest latency. Replicas do not necessarily contain the same data. However, the replicas will eventually converge. This is the weakest consistency level, but it offers the highest availability.

You should use strong consistency for CosmosDB2. This consistency level ensures that all client applications always read the same values. A replica is not available unless it contains the same data as the node where a data write occurred. This is the strongest consistency level, but it offers the lowest availability.

You should not use bounded staleness. This is one level weaker than the strong consistency level. With this consistency level, you can configure the staleness of a read. You can configure the staleness with a specific number of versions, or you can configure it with a specific amount of time.

You should not use session consistency. This is one level weaker than bounded staleness. It ensures that the same client always reads the same data.

You should not use consistent prefix. This is one level weaker than bounded session. This consistency level guarantees that reads from any client never see out-of-order writes.

References

[Consistency levels in Azure Cosmos DB](#)

Question162

You are a data architect. You want to design a solution that shows only the last four digits of a customer's phone number when you query the number from the database. The remaining part of the number should be returned with X's in place of the other digits.

You need to implement the solution.

What should you implement?

Choose the correct answer

Dynamic data masking (DDM)

Transparent Data Encryption (TDE)

Always Encrypted

Row-level security (RLS)

Explanation

You should implement DDM. This limits exposure of data to non-privileged users by masking parts of it.

You should not implement row-level security (RLS). This allows you to filter rows that are returned by using a security policy and filter predicate.

You should not implement Always Encrypted. This technology allows you to encrypt and decrypt data on the client. The data remains encrypted in the database at the server.

You should not implement TDE. TDE allows you to encrypt the entire database, including the database files and transaction logs.

References

Row-Level Security

Transparent Data Encryption (TDE)

Always Encrypted (Database Engine)

Dynamic Data Masking

Question163

You are a data architect for a property records management company. You are planning to implement dynamic data masking (DDM) for a database solution.

You need to determine the scenarios for which DDM is beneficial.

For each scenario, select Yes if DDM would be beneficial. Otherwise, select No.

Scenario	Yes	No
You want to allow property deeds to be retrieved only by specific personnel.	○	○
You want to ensure that only the last four digits of the property owner's phone number are visible.	○	○

You want to encrypt the loan number associated with the property.	○	○

Explanation

DDM is not beneficial when you want to allow property deeds to be retrieved only by specific personnel. This can be achieved with row-level security (RLS), which uses a security policy and filter predicate to determine when rows should be retrieved.

DDM is beneficial when you want to ensure that only the last four digits of the property owner's phone number are visible. This allows you to define a masking function to specify how specific parts of data is hidden.

DDM is not beneficial when you want to encrypt the loan number associated with the property. You can accomplish this with column-level encryption (CLE), which uses symmetric encryption to encrypt specific columns at the database.

References

[Row-Level Security](#)

[Encrypt a Column of Data](#)

[Dynamic Data Masking](#)

Question164

Your company has strict policies regarding data. You deploy an Azure SQL Database.

You need to design a solution that allows every deletion of data to be stored in Log Analytics.

What should you do?

Choose the correct answer

Add a database metric.

Configure an alert rule.

Enable auditing.

Create a diagnostic setting.

Explanation

You should enable auditing. This allows you to log specific database events to one or more of the following: blob storage, Event Hub, or Log Analytics.

You should not create a diagnostic setting. Diagnostic settings allow you to collect deadlocks, errors, among other events. You cannot use it to log data deletions to Log Analytics.

You should not add a database metric. Metrics allow you to visually monitor a database. Metrics include the number of Database Transaction Units (DTU) used, CPU percentage, and others.

You should not configure an alert rule. Alerts allow you to be notified when a specific metric meets or exceeds a defined threshold.

References

Azure SQL Database metrics and diagnostics logging

Get started with SQL database auditing

Question165

You want to have database events from an Azure SQL Database managed instance logged to a blob storage account.

You need to design the solution for logging the events.

Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of possible actions to the answer area and arrange them in the correct order.

Create a list in the correct order

Possible Actions

Actions in Order

- Enable auditing in the Azure portal.
- Copy the primary access key of the storage account.
- Copy the secondary access key of the storage account.
- Generate a service shared access signature (SAS).

Explanation

You should first generate a service SAS. An SAS embeds the access permissions within it. You should copy the token part of the SAS.

Next, you should use T-SQL to create a credential. A credential specifies an identity and a secret. The secret should be the SAS token.

Next, you should use T-SQL to create a server audit. The audit should specify the blob container URL, and optionally the retention days. This specifies where database events should be logged.

You should not copy the primary access key or secondary access key. These keys are used to administer a blob storage account.

You should not enable auditing in the Azure portal. In this scenario, you are using SQL Database Managed Instance, which means that you should use T-SQL statements to enable auditing.

References

Get started with Azure SQL Database managed instance auditing

Manage storage account settings in the Azure portal

Get started with SQL database auditing

Question166

You are a data architect for your company. You plan to deploy Azure SQL Database to support a customer service application.

You need to identify sensitive data within the database.

What should you do?

Choose the correct answer

Implement Transparent Data Encryption (TDE).

Configure dynamic data masking (DDM).

Enable database auditing.

<mark>Enable Data Discovery and Classification.</mark>

Explanation

You should enable Data Discovery and Classification. This allows you to scan your database for columns that contain potentially sensitive information.

You should not configure DDM. This allows you to hide portions of data in columns. However, it does not help you to identify which columns contain sensitive data.

You should not enable database auditing. This allows you to send database events, such as deletions, to Log Analytics, Event Hub, or blob storage.

You should not implement TDE. TDE allows you to encrypt the entire database, including data files and transaction logs.

References

[SQL Data Discovery and Classification](#)

[Transparent Data Encryption (TDE)](#)

[Dynamic Data Masking](#)

[Get started with Azure SQL Database managed instance auditing](#)

[Azure SQL Database and SQL Data Warehouse data discovery & classification](#)

Question167

Your application generates a large amount of blob data about manufactured devices. After 90 days, the data is no longer accessed, but it should remain available for analysis purposes. Analysis is planned in advance. Access latency is not a significant issue.

The data is not often used, but it is considered critical for history and reference purposes and must be protected against potential regional failures.

You need to choose a storage option for the data that:

- Provides for extended storage (10+ years)
- Supports globally redundant storage (GRS)
- Minimizes ongoing storage costs

Which blob storage tier should you recommend?

Choose the correct answer

Cool

<mark>Archive</mark>

Hot

Premium

Explanation

You should recommend archive storage. This has the lowest ongoing storage costs and provides for indefinite storage. Archive storage supports GRS to protect against regional failures or outages. Access to data can take up to 15 hours. To avoid early deletion changes, data must remain in archive storage for at least 180 days.

You should not recommend cool storage. This storage option is less expensive than hot storage and more expensive than archive storage. There is an early deletion period of 30 days associated with this storage option. Data is available for access in milliseconds.

You should not recommend hot storage. This is designed for interactive storage in a production environment and is more expensive than cool or archive storage.

You should not recommend premium storage. This option should be used with applications with very high-performance requirements. This option has the least data latency but the highest storage costs.

References

[Azure Blob storage: hot, cool, and archive access tiers](#)

[Block blob pricing](#)

[Azure Blob Storage Tiers Explained](#)

Question168

Your company maintains a graphic image reference library in an Azure Blob storage account. You need to select a storage redundancy option that meets the following requirements:

- Data protection in case of loss of a data center or regional failure
- At least 99.999% availability for reads and writes
- Minimum cost

Which storage redundancy option should you choose?

Choose the correct answer

Geo-zone-redundant storage (GZRS)

Zone redundant storage (ZRS)

<mark>Globally redundant storage (GRS)</mark>

Locally redundant storage (LRS)

Explanation

You should recommend GRS. GRS meets or exceeds the data protection and availability requirements. Data is protected in case of a datacenter or region failure. This is the least expensive option that supports blob storage and meets these requirements.

You should not recommend LRS. With LRS, data is replicated between servers in the same datacenter. Protection against a datacenter failure or catastrophe is not guaranteed. This option does meet the other scenario requirements. This is the least expensive option.

You should not recommend ZRS. This meets the scenario's technical requirements and is a relatively low-cost storage option, but ZRS does not support blob storage.

You should not recommend GZRS. This is the highest performing and most secure storage option, but it is also the most expensive and does not support blob storage.

References

Azure Storage redundancy

Geo-redundant storage (GRS): Cross-regional replication for Azure Storage

Question169

You are deploying an Azure SQL Database to support a retail sales application. The SQL Database instance will use the DTU-based purchasing model.

Once in place, this database will be part of a business-critical application. You must ensure that database backups are available for up to 30 days and support point-in-time restoration any time during that period.

You need to recommend a backup and recovery option.

Which two methods should you recommend using? Each correct answer presents a complete solution.

Choose the correct answers

Premium tier with the default backup retention policy

Basic tier with the default backup retention policy

Standard tier with a long-term retention (LTR) backup policy

Premium tier with a long-term retention (LTR) backup policy

Standard tier with the default backup retention policy

Explanation

You should use one of the following:

* Standard tier with the default backup retention policy
* Premium tier with the default backup retention policy

The default backup retention policy keeps backups on hand for up to 5 weeks (35 days). Default automated backups support point-in-time restore. Point-in-time restore requires restoring from a full backup, differential backup, and transaction log backups.

Basic tier default backups retain backups for only 7 days, so you should not choose a basic tier option.

You should not choose an LTR backup policy option. LTR extends the period that backups are retained. LTR retains copies of full backups only, so point-in-time restoration is not supported.

References

Question170 Design non-relational cloud data stores

Performance for an on-premises data warehouse has degraded over time. You discover that the problem is related to two fact tables. Queries using the SalesHistory table take longer than expected to complete. Data loads into the invoice table often suffer performance issues and impact data processing. Both tables include related columns named SalesDate, ProductID, and RegionID.

The SalesHistory table is approximately 1 TB in size and is used for several analysis purposes, including sales analysis and trend predictions. The SalesDate column is often used in queries, and the ProductID column is used extensively in JOIN operations. The RegionID is used for grouping results.

The Invoice table is used primarily as a staging table with data loaded into Invoice before being processed and loaded into SalesHistory. The table size seldom exceeds 1 GB. Table content is processed daily by region with the RegionID used to group data.

You are moving the on-premises data warehouse to Azure SQL Data Warehouse. You need to determine the distribution type and distribution column configurations that best meet your processing requirements. Total storage space requirements should be minimized.

Which distribution types and columns should you use? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

Table	Distribution type	Distribution column
SalesHistory		
Invoice		

[dropdown boxes]

Which distribution types and columns should you use? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

Table	Distribution type	Distribution column
SalesHistory	Hash-distributed	ProductID
Invoice	Round-robin	None

< Previous ☆ Next >

10:25 PM
8/7/2020

Explanation

SQL Data Warehouse uses distributed tables across distributions for fact table data storage. You can help optimize performance through your distribution type and distribution column choices when designing warehouse tables.

You should configure hash-distribution as the distribution type for SalesHistory. Hash-distribution helps to improve query performance for large tables. Because optimization can be recognized in JOIN operations, you should choose the ProductID

column as the distribution column because it is used extensively in joins. Hash-distribution is recommended for tables over 2 GB in size and tables used extensively in INSERT, UPDATE, and DELETE operations.

You should configure round-robin distribution for the Invoice table. Round-robin distribution is well-suited for smaller tables when you want to optimize the table for data loads. It is the recommended distribution for staging tables. Round-robin distribution is determined randomly and does not use a distribution column.

You should not configure either table as replicated distribution. This distributes full copies of the tables, increasing the storage space requirements without significantly improving the performance issues in the scenario.

Neither table would receive its best benefit from using SalesDate as the distribution column.

Question171

You collect streaming data from monitoring devices on a manufacturing floor into an Azure Cosmos DB database. Data should be removed from the database after one day.

You need to configure this to happen automatically. The effort required to maintain the solution should be minimized.

What Cosmos DB feature should you use?

Choose the correct answer

Time to Live (TTL)

Diagnostic logging

Advanced Threat Protection (ATP)

Triggers

Explanation

You should use Cosmos DB TTL. TTL lets you configure expiration by container or by item. To minimize maintenance requirements, you should configure TTL by container. The TTL period is configured in seconds.

You should not use Cosmos ATP. ATP is used to detect attempts to access or exploit Cosmos DB accounts. Security alerts are triggered when anomalies are detected and

are integrated with Azure Security Center and emailed to the subscription administrator.

You should not use Cosmos DB diagnostic logging. Diagnostic logging gives you a way to monitor how and when your databases are accessed.

You should not use Cosmos DB triggers. Triggers are used to automatically invoke Azure functions for data processing.

References

[Time to Live (TTL) in Azure Cosmos DB](#)

[Diagnostic logging in Azure Cosmos DB](#)

[Advanced Threat Protection for Azure Cosmos DB](#)

[Create a function triggered by Azure Cosmos DB](#)

Question172

Your company is developing a data solution to collect, analyze, and report traffic statistics with a focus on public transit use. The solution has both real-time and batch processing requirements.

An Azure SQL Database is used as a reference database used to support both real-time and batch processing activities. The data in some columns is proprietary to the company and is considered confidential. Direct access to this data should be limited to applications accessing the data. Other database columns do not require the same protection.

You need to ensure security for the reference database.

What should you use?

Choose the correct answer

Transport Layer Security (TLS)

Always Encrypted

Advanced Threat Protection (ATP)

Transparent Data Encryption (TDE)

You should use Always Encrypted, also known as encryption-in-use. Always Encrypted is used to protect specific database columns from access. This includes blocking access to database administrators or other database users. Data is decrypted for processing by client applications only.

You should not use ATP. ATP is a way to detect unusual behavior or attempts to access or exploit a database. This is done through analysis of SQL Server logs. This is not an encryption mechanism.

You should not use TLS. TLS is used for encryption in transit, ensuring that data is encrypted when data is transferred between a client and server. It does not ensure at rest encryption or that backups are encrypted.

You should not use TDE. TDE encrypts the entire database and prevents offline access of data through raw files or backup files. It also encrypts transaction log files. TDE be manually enabled for Azure SQL Database, Azure SQL Database managed instance, or Azure SQL Data Warehouse databases.

References

An overview of Azure SQL Database security capabilities

Azure database security best practices