'hich method for renaming a DataFrame column is incorrect?

df.select(col("timestamp").alias("dateCaptured"))

All are incorrect.

df.alias("timestamp", "dateCaptured")

All are correct.

df.toDF("dateCaptured")

Report Error

TEST MODE

1	2
8	9
15	16
22	23
29	30
36	37
43	44

EV

NEXT >

rio: You need to provision a data store that will store but not query data.

is the least expensive option which will meet the requirement?

- Azure Queue Storage
- Azure Cosmos DB Storage
- Azure Database Server Storage
- Azure Blob Storage
- Azure File Storage
- Azure Table Storage

Report Error

TEST MODE

1 2
8 9
15 16
22 23
29 30
36 37
43 44

EV

NEXT >

rep training

Q. 3

Identify the missing word(s) in the following sentence within the context of Microsoft Azure.

Azure Synapse Analytics is a cloud-based data platform that brings together enterprise data warehousing and Big Data analytics. It can process massive amounts of data and answer complex business questions with limitless scale.

Azure Synapse Analytics uses the [?] approach for bulk data.

- ○ Extract, Transform, and Load (ETL)
- ○ Atomicity, Consistency, Isolation, and Durability (ACID)
- ○ Automated Data Processing Equipment (ADPE)
- ○ Extract, Load, and Transform (ELT)

Report Error

TEST MODE

1 2
8 9
15 16
22 23
29 30
36 37
43 44

< PREV

NEXT >

rep training

Q. 4 In which modes does Azure Databricks provide data encryption?

- At-rest and in-transit
- None of the listed options.
- In-transit only
- At-rest only

Report Error

TEST MODE

1 2
8 9
15 16
22 23
29 30
36 37
43 44

PREV NEXT

rep training

Q. 5

Scenario: ACME Books Inc uses Azure Cosmos DB to store user profile data from their eCommerce site. The NoSQL document store provided by the Azure Cosmos DB SQL API provides the familiarity of managing their data using SQL syntax, while being able to read and write the files at a massive, global scale.

While ACME is happy with the capabilities and performance of Azure Cosmos DB, they are concerned about the cost of executing a large volume of analytical queries over multiple partitions (cross-partition queries) from their data warehouse. They want to efficiently access all the data without needing to increase the Azure Cosmos DB request units (RUs).

They have looked at options for extracting data from their containers to the data lake as it changes, through the Azure Cosmos DB change feed mechanism. The problem with this approach is the extra service and code dependencies and long-term maintenance of the solution. They could perform bulk exports from a Synapse Pipeline, but then they won't have the most up-to-date information at any given moment.

Required: Pick a solution will ensure all transactional data is automatically stored in a fully isolated column store without impacting the transactional workloads or incurring resource unit (RU) costs.

- Enable Spark Pools for SQL Datawarehouse and enable the analytical store on their Azure Cosmos DB containers.
- Enable Azure Synapse Link for Cosmos DB and enable the analytical store on their Azure Cosmos DB containers.
- Enable Azure Private Link for SQL Database and enable the analytical store on their SQL Database containers.
- None of the listed options.

Report Error

TEST MODE

1 2
8 9
15 16
22 23
29 30
36 37
43 44

PREV NEXT

rep training

TEST MODE

1 2
8 9
15 16
22 23
29 30
36 37
43 44

Q. 6

There are two concepts within Apache Spark Pools in Azure Synapse Analytics, namely Spark Pools and Spark Instances.

Which of the following attributes belong to Spark Instances? (Select three)

- [ ] Creates a Spark Instance
- [ ] Created when connected to Spark Pool, Session, or Job
- [ ] Multiple users can have access
- [ ] Permissions can be applied
- [ ] Exists as Metadata
- [ ] Reusable

Report Error

PREV NEXT

• Permissions can be applied
• Best practices
Spark Instances:
• Created when connected to Spark Pool, Session, or Job
• Multiple users can have access
• Reusable
A Spark pool is created in the Azure portal. It is the definition of a Spark pool that, when instantiated, is used to create a Spark instance that processes data. When a Spark pool is created, it exists only as metadata; no resources are consumed, running, or charged for. A Spark pool has series of properties that control the characteristics of a Spark instance; these characteristics include but are not limited to name, size, scaling behaviour, time to live.
As there is no resource cost associated with creating Spark pools, any number of pools can be created with any number of different configurations. Permissions can also be applied to Spark pools allowing users only to have access to some and not others.
A best practice is to create smaller Spark pools that may be used for development and debugging and then larger ones for running production workloads.
An example of Spark Pools:
• You create a Spark pool called SP1; it has a fixed cluster size of 20 nodes.
• You submit a notebook job, J1 that uses 10 nodes, a Spark instance, SI1 is created to process the job.
• You now submit another job, J2, that uses 10 nodes because there is still capacity in the pool and the instance, the J2, is processed by SI1.
• If J2 had asked for 11 nodes, there would not have been capacity in SP1 or SI1. In this case, if J2 comes from a notebook, then the job will be rejected; if J2 comes from a batch job, then it will be queued.
Spark instances are created when you connect to a Spark pool, create a session, and run a job. As multiple users may have access to a single Spark pool, a new Spark instance is created for each user that connects.
When you submit a second job, then if there is capacity in the pool, the existing Spark instance also has capacity then the existing instance will process the job; if not and there is capacity at the pool level, then a new Spark instance will be created.
An example of a Spark Instance:
• You create a Spark pool call SP2; it has an autoscale enabled 10 – 20 nodes
• You submit a notebook job, J1 that uses 10 nodes, a Spark instance, SI1, is created to process the job.

TEST MODE :

1 2
8 9
15 16
22 23
29 30
36 37
43 44

PREV NEXT

others.
A best practice is to create smaller Spark pools that may be used for development and debugging and then larger ones for running production workloads.
An example of Spark Pools:
• You create a Spark pool called SP1; it has a fixed cluster size of 20 nodes.
• You submit a notebook job, J1 that uses 10 nodes, a Spark instance, SI1 is created to process the job.
• You now submit another job, J2, that uses 10 nodes because there is still capacity in the pool and the instance, the J2, is processed by SI1.
• If J2 had asked for 11 nodes, there would not have been capacity in SP1 or SI1. In this case, if J2 comes from a notebook, then the job will be rejected; if J2 comes from a batch job, then it will be queued.
Spark instances are created when you connect to a Spark pool, create a session, and run a job. As multiple users may have access to a single Spark pool, a new Spark instance is created for each user that connects.
When you submit a second job, then if there is capacity in the pool, the existing Spark instance also has capacity then the existing instance will process the job; if not and there is capacity at the pool level, then a new Spark instance will be created.
An example of a Spark Instance:
• You create a Spark pool call SP2; it has an autoscale enabled 10 – 20 nodes
• You submit a notebook job, J1 that uses 10 nodes, a Spark instance, SI1, is created to process the job.
• You now submit another job, J2, that uses 10 nodes, because there is still capacity in the pool the instance auto grows to 20 nodes and processes J2.
https://docs.microsoft.com/en-us/azure/synapse-analytics/spark/apache-spark-concepts

- [ ] Permissions can be applied
- [ ] Exists as Metadata
- [ ] Reusable

Report Error

TEST MODE :

1 2
8 9
15 16
22 23
29 30
36 37
43 44

PREV NEXT

rep training

Q. 7

Identify the missing word(s) in the following sentence within the context of Microsoft Azure.

A common use with [?] is to take the data that is shared and use it as a source into Azure Data Factory pipelines to use with your own internal data.

- Azure SQL Database
- Azure Data Lake Storage
- Azure Data Share
- Azure Databricks
- Azure Managed SQL Warehouse

Report Error

TEST MODE

1 2
8 9
15 16
22 23
29 30
36 37
43 44

PREV NEXT

rep training

Q. 8

Identify the missing word(s) in the following sentence within the context of Microsoft Azure.

Synapse dedicated SQL Pools supports JSON format data to be stored using [?]. The JSON format enables representation of complex or hierarchical data structures in tables. It allows to transform arrays of JSON objects into table format. The performance of JSON data can be optimized by using columnstore indexes and memory optimized tables.

- Standard NVARCHAR table columns
- Standard VARCHAR table columns
- Standard CHAR table columns
- Standard NCHAR table columns

Report Error

TEST MODE

1 2
8 9
15 16
22 23
29 30
36 37
43 44

PREV NEXT

rep training

Q. 9 Which transformation is used to load data into a data store or compute resource?

- Cache
- Window
- Sink
- Field
- Source

Report Error

TEST MODE

1	2
8	9
15	16
22	23
29	30
36	37
43	44

PREV NEXT

rep training

Q. 10

Azure provides many ways to store your data. There are multiple database options like Azure SQL Database, Azure Cosmos DB, and Azure Table Storage. Azure offers multiple ways to store and send messages, such as Azure Queues and Event Hubs. You can even store loose files using services like Azure Files and Azure Blobs.

A storage account defines a policy that applies to all the storage services in the account.

Which are the settings that are controlled by a storage account. (Select all that apply)

- Virtual networks
- Access tier
- Secure transfer required
- All of these
- Location
- Subscription

Report Error

TEST MODE

1	2
8	9
15	16
22	23
29	30
36	37
43	44

PREV NEXT

rep
raining

Q. 11 Within Azure Synapse Link for Azure Cosmos DB, which Column-oriented store is optimized for queries?

- Cosmos DB store
- Query store
- Analytical store
- Transactional store

Report Error

TEST MODE

1	2
8	9
15	16
22	23
29	30
36	37
43	44

PREV NEXT

ep
ining

Q. 12

Continuous integration is the practice of testing each change made to your codebase automatically and as early as possible. Continuous delivery follows the testing that happens during continuous integration and pushes changes to a staging or production system.

Below is a sample overview of the CI/CD lifecycle in an Azure data factory that's configured with Azure Repos Git.

The order of the activities has been shuffled.
a. A developer creates a feature branch to make a change. They debug their pipeline runs with their most recent changes.
b. A development data factory is created and configured with Azure Repos Git. All developers should have permission to author Data Factory resources like pipelines and datasets.
c. After a pull request is approved and changes are merged in the master branch, the changes get published to the development factory.
d. After the changes have been verified in the test factory, deploy to the production factory by using the next task of the pipelines release.
e. When the team is ready to deploy the changes to a test or UAT (User Acceptance Testing) factory, the team goes to their Azure Pipelines release and deploys the desired version of the development factory to UAT. This deployment takes place as part of an Azure Pipelines task and uses Resource Manager template parameters to apply the appropriate configuration.
f. After a developer is satisfied with their changes, they create a pull request from their feature branch to the master or collaboration branch to get their changes reviewed by peers.

Select the correct sequence of events in the CI/CD lifecycle.

- a,b,c,f,d,e
- a,f,d,b,c,e
- b,a,f,d,e,c
- b,a,f,c,e,d

Report Error

TEST MO

1	
8	
15	
22	
29	
36	
43	

PREV NEXT

Q. 12

Continuous integration is the practice of testing each change made to your codebase automatically and as early as possible. Continuous delivery follows the testing that happens during continuous integration and pushes changes to a staging or production system.

Below is a sample overview of the CI/CD lifecycle in an Azure data factory that's configured with Azure Repos Git.

The order of the activities has been shuffled.
a. A developer creates a feature branch to make a change. They debug their pipeline runs with their most recent changes.
b. A development data factory is created and configured with Azure Repos Git. All developers should have permission to author Data Factory resources like pipelines and datasets.
c. After a pull request is approved and changes are merged in the master branch, the changes get published to the development factory.
d. After the changes have been verified in the test factory, deploy to the production factory by using the next task of the pipelines release.
e. When the team is ready to deploy the changes to a test or UAT (User Acceptance Testing) factory, the team goes to their Azure Pipelines release and deploys the desired version of the development factory to UAT. This deployment takes place as part of an Azure Pipelines task and uses Resource Manager template parameters to apply the appropriate configuration.
f. After a developer is satisfied with their changes, they create a pull request from their feature branch to the master or collaboration branch to get their changes reviewed by peers.

Select the correct sequence of events in the CI/CD lifecycle.

- ○ a,b,c,f,d,e
- ○ a,f,d,b,c,e
- ○ b,a,f,d,e,c
- ○ b,a,f,c,e,d

Report Error

TEST MO

1 8 15 22 29 36 43

PREV NEXT

Q. 13 What is an example of a branching activity used in control flows in Azure Data Factory?

- ○ Lookup- condition
- ○ Where-condition
- ○ Having-condition
- ○ Until-condition
- ○ If-condition

Report Error

TEST MODE

1 2 8 9 15 16 22 23 29 30 36 37 43 44

PREV NEXT

rep training

TEST MODE

1 2
8 9
15 16
22 23
29 30
36 37
43 44

Q. 14

Azure Synapse Studio is the primary tool to use to interact with the many components that exist in the service. It organizes itself into hubs which allow you to perform a wide range of activities against your data.

Which of the following are the referenced hubs on Azure Synapse Studio?

- ○ Monitor
- ○ Data
- ○ Integrate
- ○ All of these
- ○ Home
- ○ Manage

Report Error

PREV NEXT

rep training

TEST MODE

1 2
8 9
15 16
22 23
29 30
36 37
43 44

Q. 15

Azure Data factory can accommodate organizations that are embarking on data integration projects from differing starting point. Typically, many data integration workflows must consider existing pipelines that have been created on previous projects, with different dependencies and using different technologies.

Which of the following are ingestion methods that can be used to extract data from a variety of sources?

- ☐ Linked Services
- ☐ Compute resources
- ☐ Datasets
- ☐ SSIS packages
- ☐ Copy Activity
- ☐ Self-hosted

Report Error

PREV NEXT

rep training

Q. 17 Correct or Incorrect: Unique credential creation is always required when utilizing the Azure Synapse Apache Spark Pool to Synapse SQL connector to enforce access control.

- ○ Incorrect
- ○ Correct

Report Error

TEST MODE

1	2
8	9
15	16
22	23
29	30
36	37
43	44

PREV NEXT

rep training

Q. 18

Scenario: Pennyworth's Haberdashery is a clothing retailer based in London. The company has 2,000 retail stores across the EU and an emerging online presence. The network contains an Active Directory forest named pennyworths.com. The forest it integrated with an Azure Active Directory (Azure AD) tenant named pennyworths.com. Pennyworth's has an Azure subscription associated to the pennyworths.com Azure AD tenant.

Pennyworth's has three years of customer, transactional, operational, sourcing, and supplier data comprised of 10 billion records stored across multiple on-premises. Microsoft SQL Server servers. The SQL Server instances contain data from various operational systems. The data is loaded into the instances by using SQL Server Integration Services (SSIS) packages.

You have been hired as a consultant by Alfred Pennyworth to advise on very important projects within the company.

During your assessment of the IT environment, you estimate that combining all product sales transactions into a company-wide sales transactions dataset will result in a single table that contains 5 billion rows, with one row per transaction.

Most queries targeting the sales transactions data will be used to identify which products were sold in retail stores and which products were sold online during different time periods. Sales transaction data that is older than three years will be removed monthly.

The IT team plans to create a retail store table that will contain the address of each retail store. The table will be approximately 2 MB. Queries for retail store sales will include the retail store addresses.

They also plan to create a promotional table that will contain a promotion ID. The promotion ID will be associated to a specific product. The product will be identified by a product ID. The table will be approximately 5 GB.

The e-commerce department at Pennyworth's develops an Azure logic app that captures trending Twitter feeds referencing the company's products and pushes the products to Azure Event Hubs.

Planned Changes and Requirements
Pennyworth's plans to implement the following changes:
• Load the sales transaction dataset to Azure Synapse Analytics.
• Integrate on-premises data stores with Azure Synapse Analytics by using SSIS packages.
• Use Azure Synapse Analytics to analyze Twitter feeds to assess customer sentiments about products.
Sales Transaction Dataset Requirements

TEST MODE :

1	2	
8	9	
15	16	
22	23	
29	30	
36	37	
43	44	

PREV NEXT

Pennyworth's identifies the following requirements for the sales transaction dataset:
• Partition data that contains sales transaction records. Partitions must be designed to provide efficient loads by month. Boundary values must belong to the partition on the right.
• Ensure that queries joining and filtering sales transaction records based on product ID complete as quickly as possible.
• Implement a surrogate key to account for changes to the retail store addresses.
• Ensure that data storage costs and performance are predictable.
• Minimize how long it takes to remove old records.

Customer Sentiment Analytics Requirements
Pennyworth's identifies the following requirements for customer sentiment analytics:
• Allow Pennyworth's users to use PolyBase in an Azure Synapse Analytics dedicated SQL pool to query the content of the data records that host the Twitter feeds.
• Data must be protected by using row-level security (RLS). The users must be authenticated by using their own Azure AD credentials.
• Maximize the throughput of ingesting Twitter feeds from Event Hubs to Azure Storage without purchasing additional throughput or capacity units.
• Store Twitter feeds in Azure Storage by using Event Hubs Capture. The feeds will be converted into Parquet files.
• Ensure that the data store supports Azure AD-based access control down to the object level.
• Minimize administrative effort to maintain the Twitter feed data records.
• Purge Twitter feed data records that are older than two years.

Data Integration Requirements
Pennyworth's identifies the following requirements for data integration:
• Use an Azure service that leverages the existing SSIS packages to ingest on-premises data into datasets stored in a dedicated SQL pool of Azure Synapse Analytics and transform the data.
• Identify a process to ensure that changes to the ingestion and transformation activities can be version-controlled and developed independently by multiple data engineers.

The IT team has come up with a list of commands they are considering to execute which is shown below:
a. CREATE EXTERNAL DATA SOURCE
b. CREATE EXTERNAL FILE FORMAT
c. CREATE EXTERNAL TABLE
d. CREATE EXTERNAL TABLE AS SELECT
e. CREATE DATABASE SCOPED CREDENTIAL

PREV NEXT

• Use an Azure service that leverages the existing SSIS packages to ingest on-premises data into datasets stored in a dedicated SQL pool of Azure Synapse Analytics and transform the data.
• Identify a process to ensure that changes to the ingestion and transformation activities can be version-controlled and developed independently by multiple data engineers.

The IT team has come up with a list of commands they are considering to execute which is shown below:
a. CREATE EXTERNAL DATA SOURCE
b. CREATE EXTERNAL FILE FORMAT
c. CREATE EXTERNAL TABLE
d. CREATE EXTERNAL TABLE AS SELECT
e. CREATE DATABASE SCOPED CREDENTIAL

The Ask:
Alfred places a great importance on this project and asks you to work closely with the team to ensure that the Twitter feed data can be analyzed in the dedicated SQL pool. The solution must meet the customer sentiment analytic requirements.

As the Azure expert, the team looks to you for direction with regards to the proper path forward.

Which Transact-SQL DDL commands should you recommend to be run in sequence?

- ○ d,a,b,e
- ○ a,b,d
- ○ e,d,e,a,c
- ○ c,d,a,e

Report Error

PREV NEXT

rep training

TEST MODE

1 2
8 9
15 16
22 23
29 30
36 37
43 44

Q. 19

You can natively perform data transformations with Azure Data Factory code free using the Mapping Data Flow task. Mapping Data Flows provide a fully visual experience with no coding required. Your data flows will run on your own execution cluster for scaled-out data processing.

Clicking Debug will provision the Spark clusters required to interact with the Mapping Data Flow transformations. If you select AutoResolveIntegrationRuntime, what will be the result? (Select all that apply)

- [ ] It typically takes 5-7 minutes for the cluster to spin up.
- [ ] A cluster with eight cores that will be available with a time to live value of 60 minutes.
- [ ] The number of rows that are returned within the data previewer are fixed by the AutoResolve Agent.
- [ ] None of the listed options.
- [ ] Data engineers can develop data transformation logic with or without writing code.
- [ ] All the listed options.

Report Error

PREV NEXT

rep training

TEST MODE

1 2
8 9
15 16
22 23
29 30
36 37
43 44

Q. 20

While Agile, CI/CD, and DevOps are different, they support one another

What does DevOps focus on?

- ○ Development process
- ○ Culture
- ○ Strategy
- ○ Practices

Report Error

PREV NEXT

rep training

TEST MODE

1 2
8 9
15 16
22 23
29 30
36 37
43 44

Q. 21 In order to create a Spark pool in Azure Synapse Analytics, what needs to be created to do so?

- Synapse Analytics Workspace
- Azure Databricks
- A Spark Instance
- HDI

Report Error

PREV NEXT

rep training

TEST MODE

1 2
8 9
15 16
22 23
29 30
36 37
43 44

Q. 22

Identify the missing word(s) in the following sentence within the context of Microsoft Azure.

Apache Spark Structured Streaming is a fast, scalable, and fault-tolerant stream processing API. You can use it to perform analytics on your streaming data in [?].

- Historic batches
- Prediction mode
- Near real-time
- Real-time

Report Error

PREV NEXT

rep training

TEST MODE

1 2
8 9
15 16
22 23
29 30
36 37
43 44

Q. 23

Scenario: You have been contracted by Wayne Enterprises, a company owned by Bruce Wayne with market value of over twenty seven million dollars. Bruce founded Wayne Enterprises shortly after he created the Wayne Foundation and he became the president and chairman of the company.
Bruce has come to you because his IT team needs advice on which API to use for the database model and type based on the following information.

Specifications:
• The application uses a NoSQL database to store data
• The database uses the key-value and wide-column NoSQL database type.

Required: Developers need to access data in the database using an API.

Which of the following APIs should you recommend to Bruce and his team?

- [ ] MongoDB API
- [ ] Cassandra API
- [ ] SQL API
- [ ] Table API
- [ ] Gremlin API

Report Error

PREV NEXT

rep training

TEST MODE :

1 2
8 9
15 16
22 23
29 30
36 37
43 44

Q. 24

Scenario: The Daily Bugle is a news organization led by J. Jonah Jameson which has a meager beginning and now has become a household name in news reporting. The company has grown well, despite some technology issues and now Jonah has hired you as an IT consultant to advise on several IT projects geared to improving the company's efficiencies.

One of the current projects is geared towards the design of an enterprise data warehouse in Azure Synapse Analytics.

Specifications:
• There will be millions of rows of data loaded to the data warehouse daily
• Staging tables must be optimized for data loading

Required:
• Design the staging tables

The team has come up with a list of options they are considering to use for the task at had but are not sure which to utilize. As you are the Azure expert, the team leader has asked for your opinion on which to use.

Which of the following should you recommend for the staging table design?

- ○ External table
- ○ Replicated table
- ○ Round-robin distributed table
- ○ Hash-distributed table

Report Error

PREV NEXT

prep training

Q. 25

Usually when you see 'df' in some code it refers to a [?].

1.Python
2.new_rows =
3.demo_df =

- ○ Dataflow
- ○ Datafeature
- ○ Dataformat
- ○ Dateformat
- ○ Dataframe

Report Error

prep training

Q. 26 What are the two prerequisites for connecting Azure Databricks with Azure Synapse Analytics that apply to the Azure Synapse Analytics instance?

- ○ Generate a OTP to verify the account credentials, then set a master endpoint then configure the endpoint firewall to enable Azure services to connect.
- ○ Use a correctly formatted ConnectionString and create a database master key
- ○ Create a database master key and configure the firewall to enable Azure services to connect
- ○ Add the client IP address to the firewall's allowed IP addresses list and use the correctly formatted ConnectionString

Report Error

prep training

Q. 27 What's the purpose of linked services in Azure Data Factory?

- ○ To represent a data store or a compute resource that can host execution of an activity.
- ○ To represent a processing step in a pipeline.
- ○ To link data storage devices between on-prem and cloud environments.
- ○ To link data stores or computer resources together for the movement of data between resources.

Report Error

prep training

Q. 28

Scenario: O'Shaughnessy's is a fast food restaurant. The chain has stores nationwide and is rivalled by Big Belly Burgers. You have been hired by the company to advise on working with Microsoft Azure Synapse Analytics.

At the moment, you are leading a meeting where the topic at hand is designing an enterprise data warehouse.

The IT team at O'Shaughnessy's is working on a project to design and create an enterprise data warehouse in Azure Synapse Analytics which will contain a table named Customers. Customers will contain credit card information.

Because security is critical to O'Shaughnessy's, they have asked you to recommend a solution to provide salespeople with the ability to view all the entries in Customers but prevent all the salespeople from viewing or inferring the credit card information.

Which of the following techniques should you propose in your recommendation?

- ○ Always Encrypted
- ○ Row-level security
- ○ Data masking
- ○ Column-level security

Report Error

prep training

Q. 29

Scenario: While working on a project using Azure Data Factory, you are planning to load data into a data store or compute resource.

Which transformation in Mapping Data Flow is used to do this?

- ○ Window
- ○ Cache
- ○ Source
- ○ Sink
- ○ Field mapping

Report Error

prep training

Q. 30

Spark is a distributed computing environment. Therefore, work is parallelized across executors.

At which two levels does this parallelization occur?

- ○ The Executor and the Slot
- ○ The Slot and the Task
- ○ The Executor and the Task
- ○ The Driver and the Executor

Report Error

prep training

Q. 31

Identify the missing word(s) in the following sentence within the context of Microsoft Azure.

[?] is an encryption mechanism to help you protect Azure Synapse Analytics. It will protect Azure Synapse Analytics against threats of malicious offline activity. The [?] way will do so by is encrypting data at rest. [?] performs real-time encryption as well as decryption of the database, associated backups, and transaction log files at rest without you having to make changes to the application.

- ○ Table-level security
- ○ Column-level security
- ○ Row-level security
- ○ Database Encryption Key
- ○ Transparent Data Encryption
- ○ Dynamic Data Masking

Report Error

prep training

Q. 32

Azure provides many ways to store your data. A Storage account defines a policy that applies to all the storage services in the account. One of the settings within the Storage account is the Storage account kind, which is a set of policies that determine which data services you can include in the account and the pricing of those services.

Which of the following are valid kinds of Storage accounts? (Select three)

- [ ] General Purpose v2
- [ ] Block blobs Storage
- [ ] Container Storage
- [ ] General Purpose v1
- [ ] Data Pool Storage
- [ ] Blob Storage

Report Error

prep training

Q. 33

Identify the missing word(s) in the following sentence within the context of Microsoft Azure.

As a Data Engineer, you can transfer and move data in several ways. The most common tool is [?], which provides robust resources and nearly 100 enterprise connectors. [?] also allows you to transform data by using a wide variety of languages.

- ○ Azure Stream Analytics
- ○ Azure Databricks
- ○ Azure Data Lake Storage
- ○ Azure Data Factory
- ○ Azure Data Catalogue

Report Error

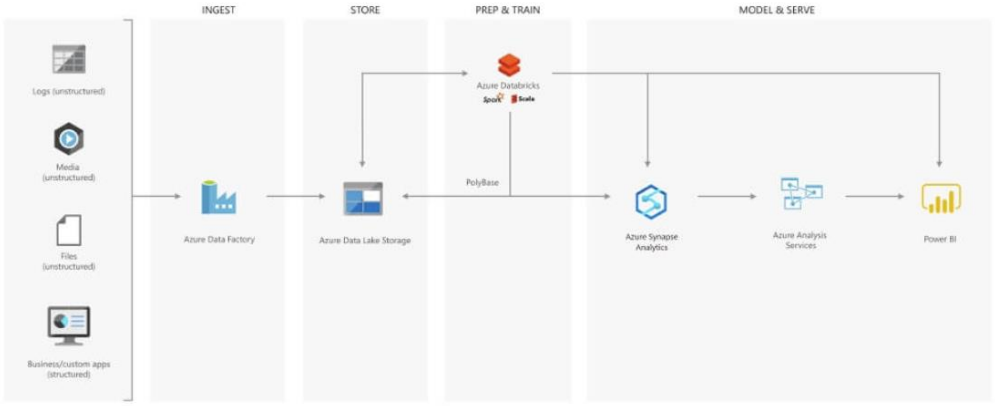
Q. 34

Scenario: Azure Data Lake Storage plays an important role in providing a large-scale data store. Your skills are needed by Hydra Corporation, which is a global seller of bicycles and cycling components through a chain of resellers and on the internet. As their customers browse the product catalogue on their websites and add items to their baskets, a recommendation engine that is built into Azure Databricks recommends other products.

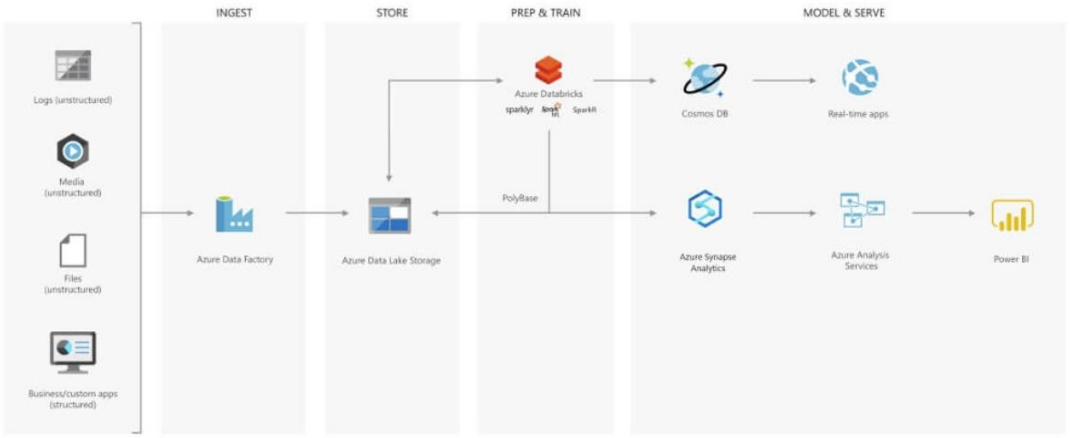
They need to make sure that the results of their recommendation engine can scale globally. The recommendations are based on the web log files that are stored on the web servers and transferred to the Azure Databricks model hourly. The response time for the recommendation should be less than 1 ms.

Review the following architecture designs.

Design A:



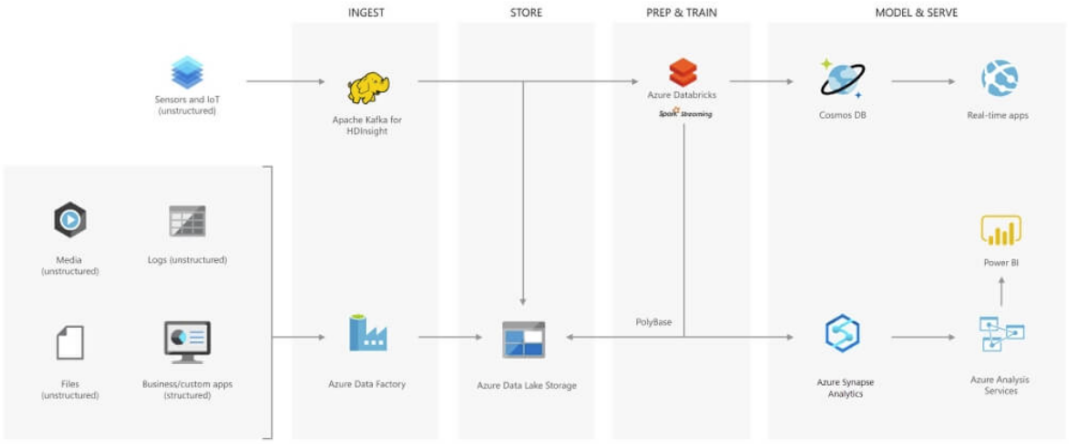
Design B:



Design C:

prep training

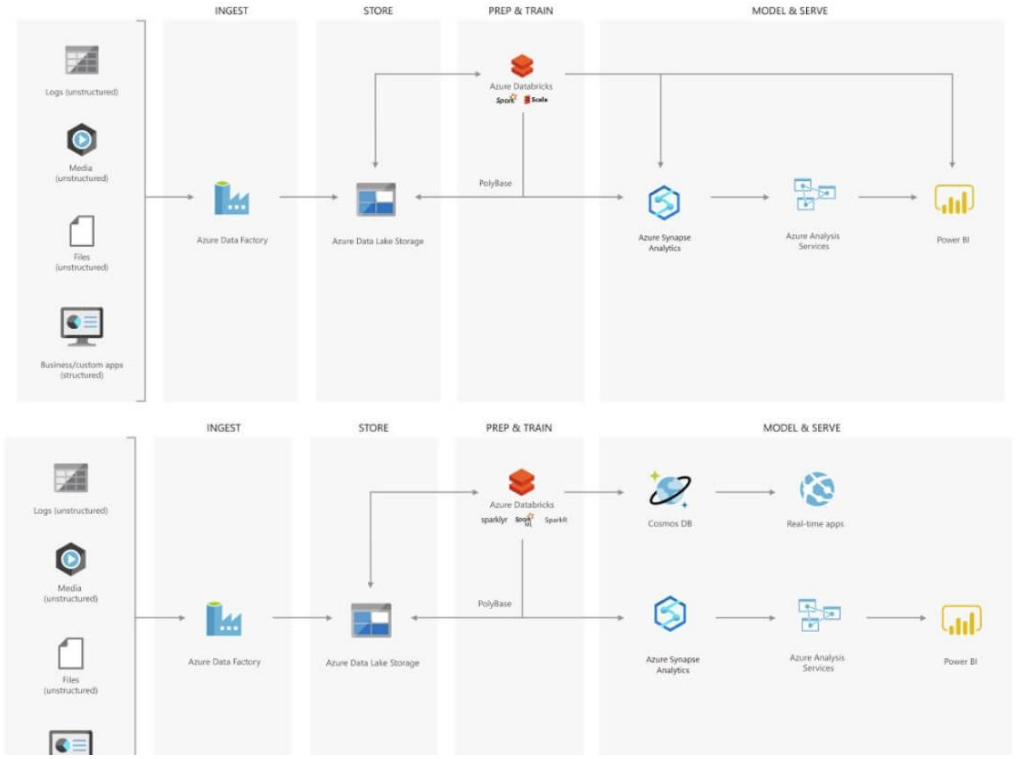
Design C:

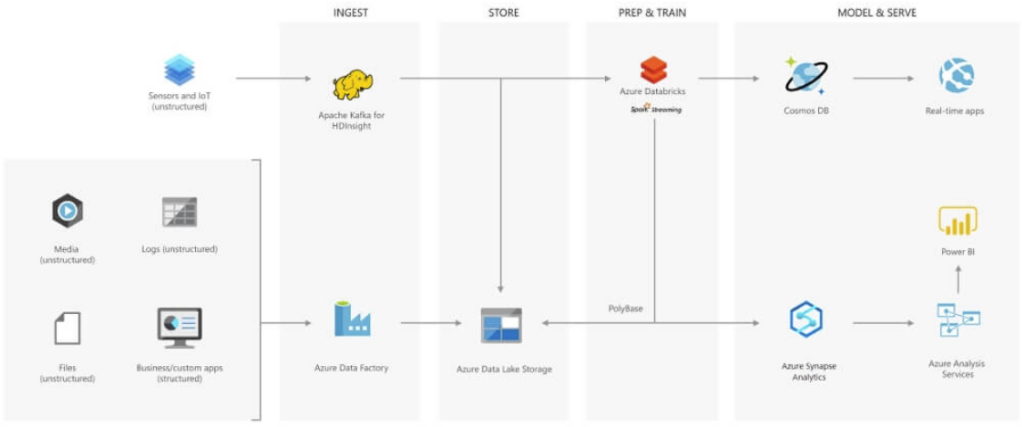


Which architecture would be best suited for the need?

prep
training

Which architecture would be best suited for the need?





- Design A
- Design C
- Design B
- None of the listed options

Report Error

Q. 35

Identify the missing word(s) in the following sentence within the context of Microsoft Azure.

Azure Synapse Analytics is an integrated analytics platform, which combines data warehousing, big data analytics, data integration, and visualization into a single environment. Azure Synapse Analytics empowers users of all abilities to gain access and quick insights across all of their data, enabling a whole new level of performance and scale.

Descriptive analytics answers the question [?].

- ○ Why is it happening?
- ○ When will the modification made meet my goals?
- ○ What is happening in my business?
- ○ What is likely to happen in the future based on previous trends and patterns?"

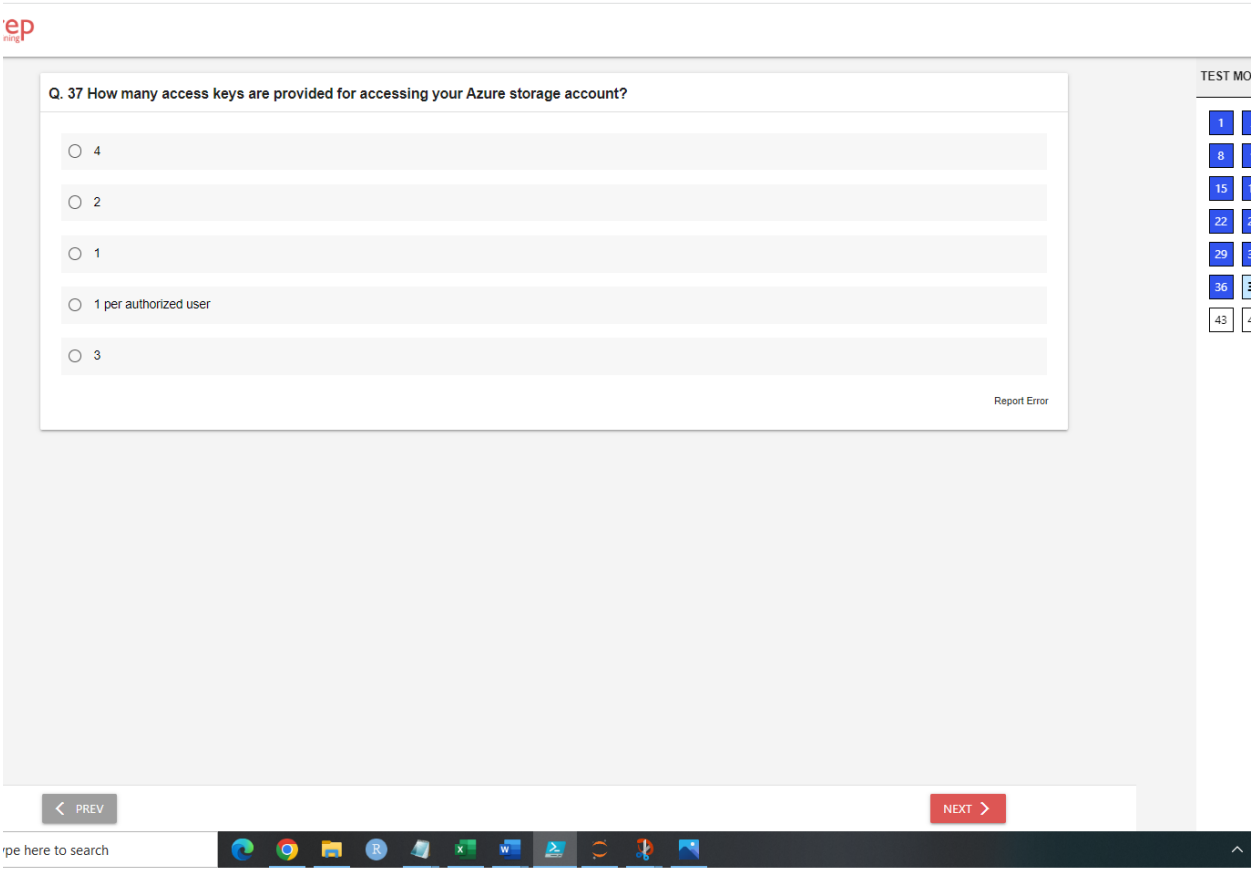
Report Error

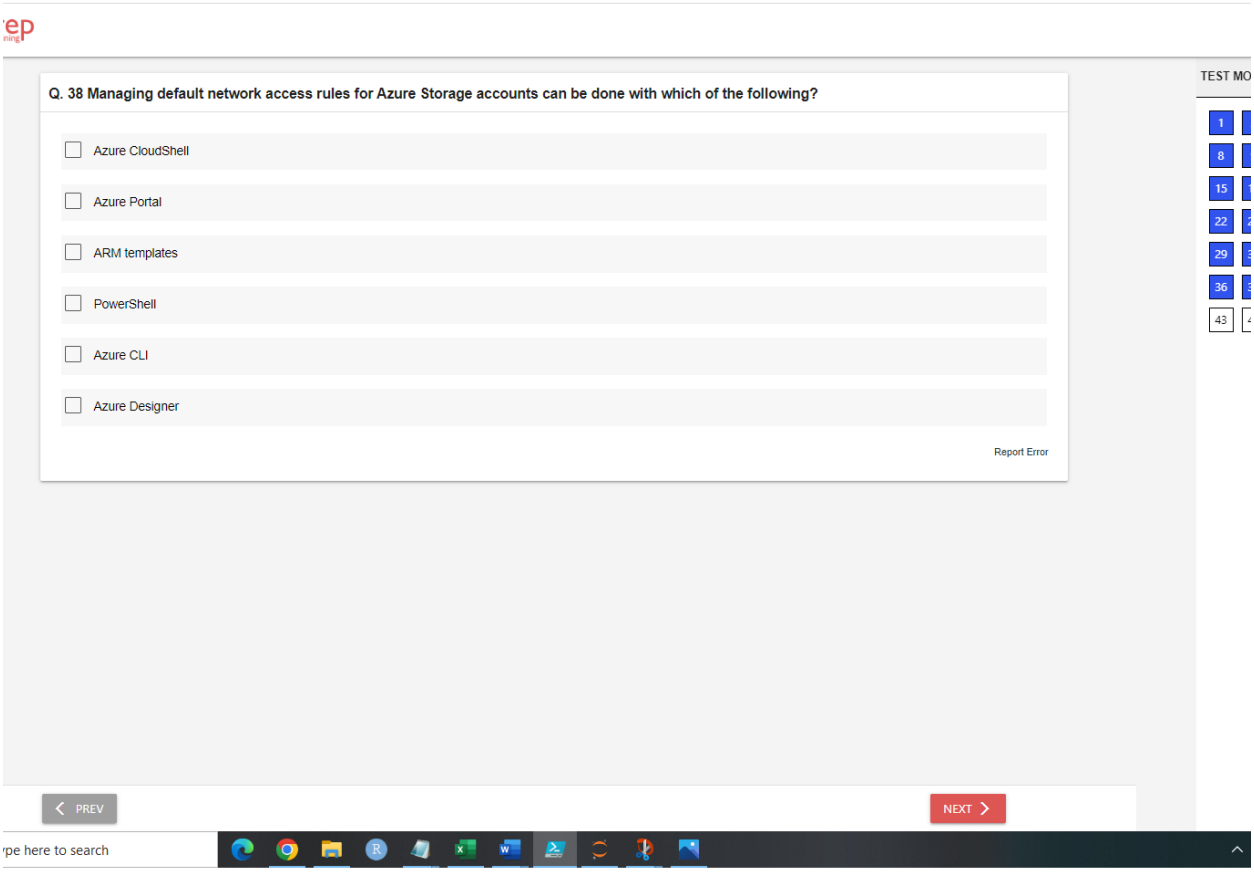
prep
training

Q. 36 Where is the best place to monitor spark pools?

- ○ Any of the listed options are equally proficient to monitor spark pools
- ○ Azure Monitor from the Azure Portal linked to your Azure Synapse Workspace
- ○ Monitor tab in Azure Advisor linked to your Azure Synapse Workspace
- ○ Monitor tab in Azure Synapse Studio within your Azure Synapse Workspace

Report Error


Q. 37 How many access keys are provided for accessing your Azure storage account?
4
2
1
1 per authorized user
3
Report Error
TEST MO
PREV
NEXT



Q. 39

Init Scripts provide a way to configure cluster's nodes. It is recommended to favour Cluster Scoped Init Scripts over Global and Named scripts. Which of the following is best described by:

"You can limit the init script to run only on for a specific cluster's creation and restarts by placing it in /databricks/init/ folder."

- ○ Global
- ◉ Cluster Named

Q. 40

Scenario: You are working at OZcorp which is a multi-million dollar company run by Mayor Norman Osborn. Profits from the company are used to fund Norman's operatives, such as a police task force.

At the moment, you have been hired by OZcorp as a Microsoft Azure Synapse Analytics SME.

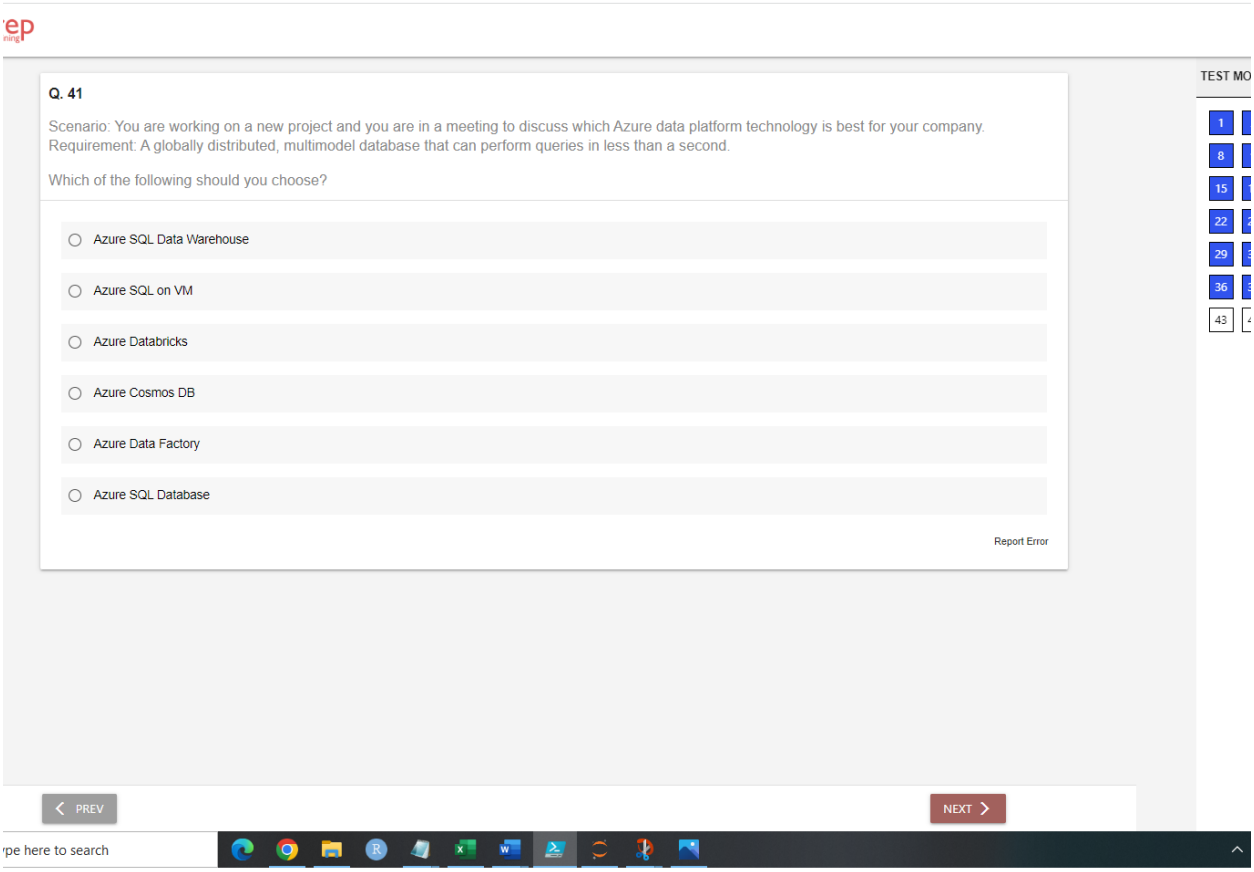
Given:
OZcorp has an on-premises data warehouse that includes the following fact tables. Both tables have the following columns: DateKey, ProductKey, RegionKey.
• Table - Sales: The table is 600 GB in size. DateKey is used extensively in the WHERE clause queries. ProductKey is used extensively in join operations. RegionKey is used for grouping. Seventy-five percent of the records relate to one of forty regions.
• Table - Invoice: The table is 6 GB in size. DateKey and ProductKey are used extensively in the WHERE clause queries. RegionKey is used for grouping.
• There are 120 unique product keys and 65 unique region keys.
• Queries that use the data warehouse take a long time to complete.

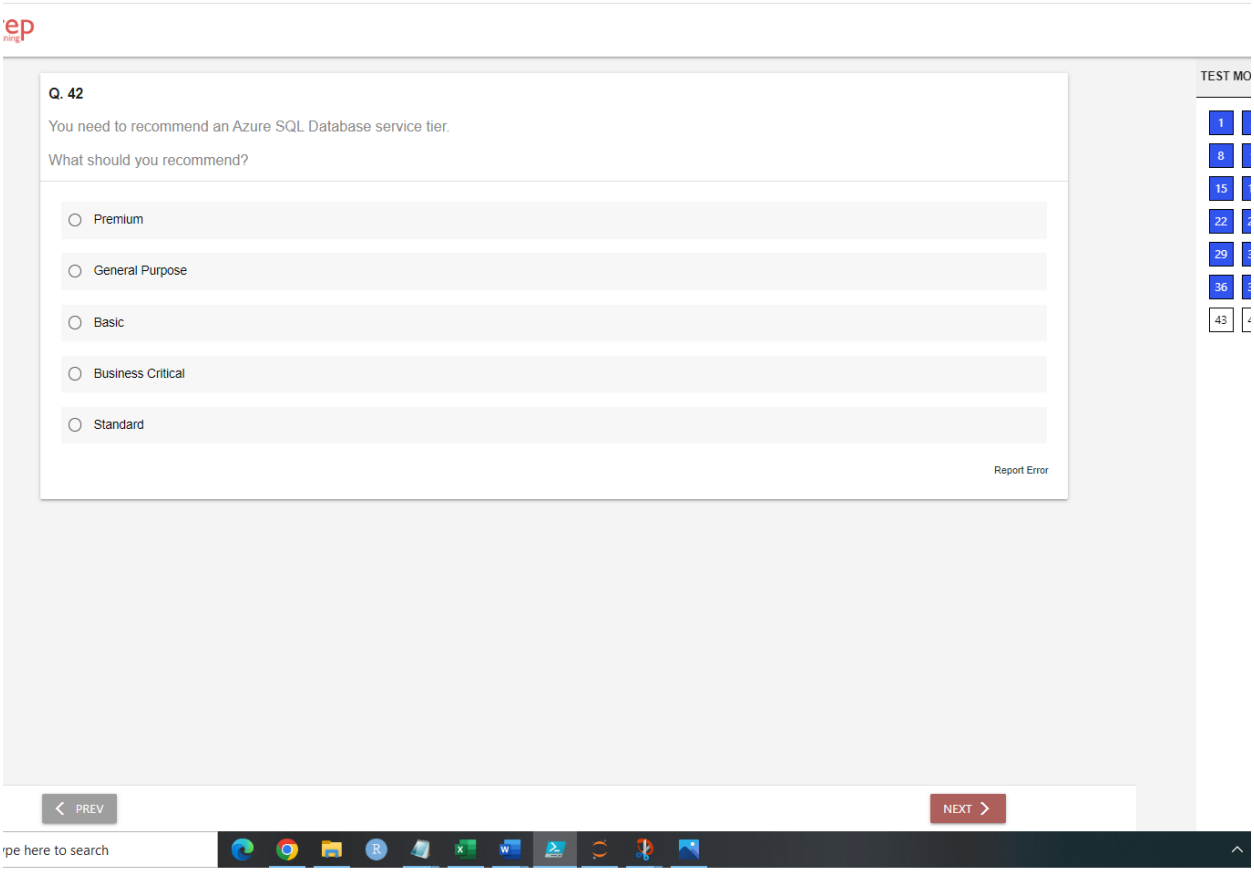
Required:
The team plans to migrate the solution to use Azure Synapse Analytics and they need to ensure that the Azure-based solution optimizes query performance and minimizes processing skew.

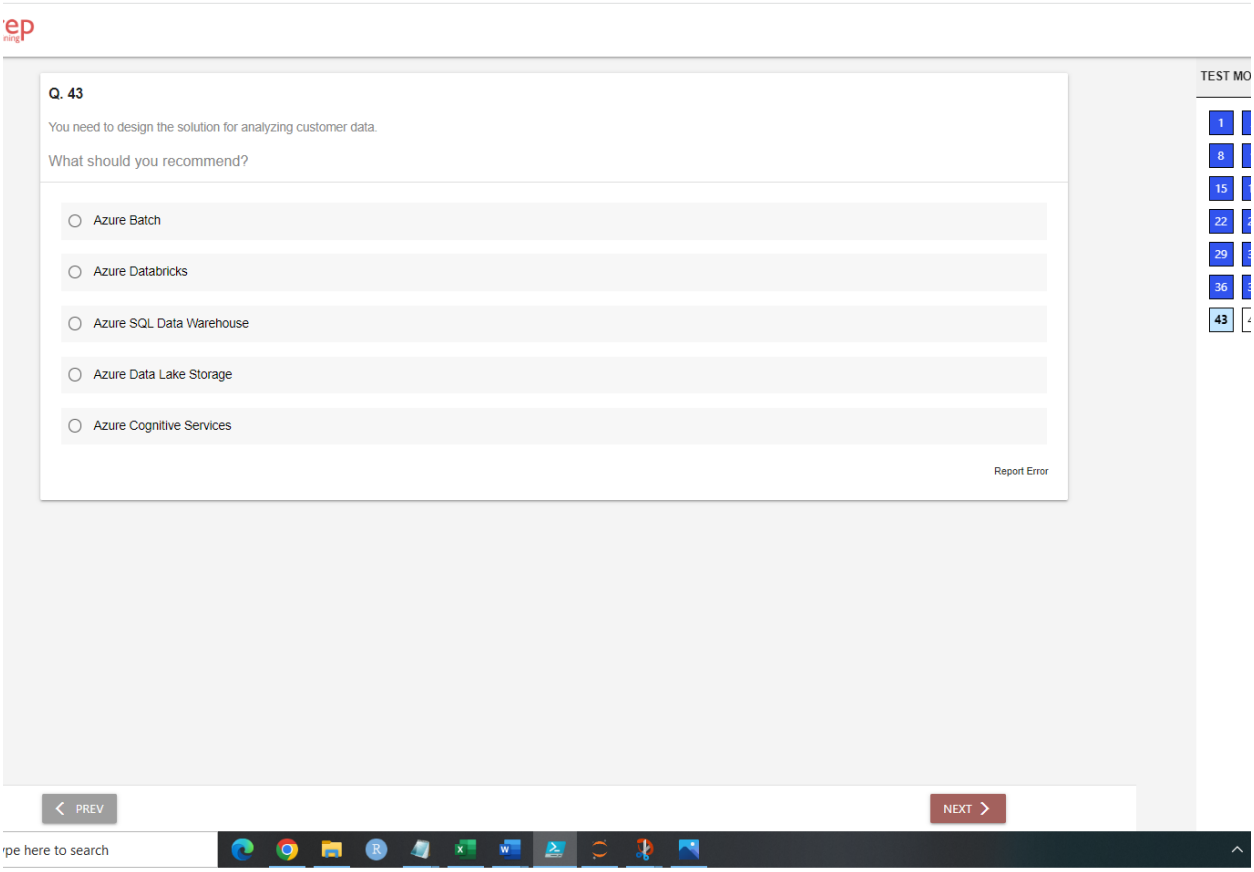
Azure Synapse Analytics SME, the team looks to you for the best way forward. Which of the following should you recommend for the Sales table?

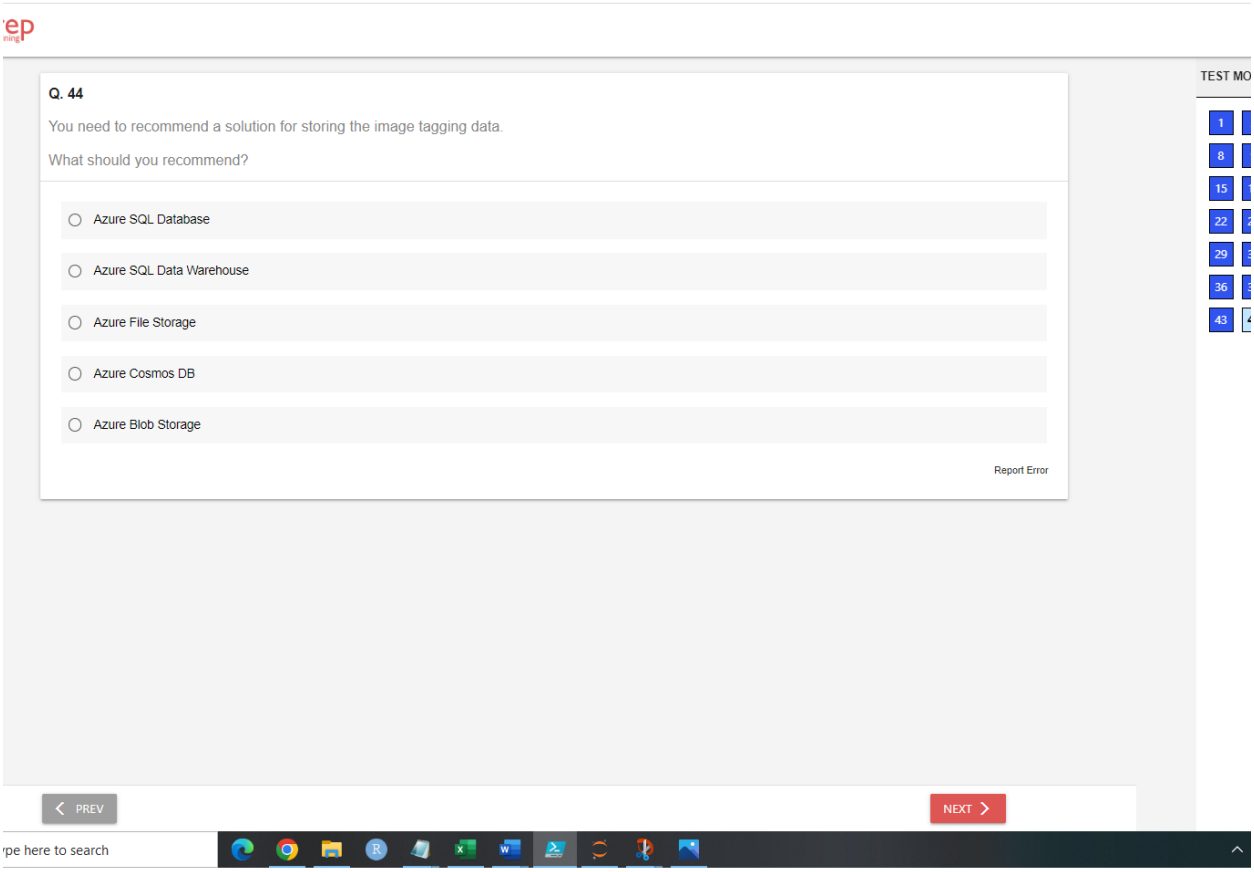
- Distribution type: Hash-distributed, Distribution column: RegionKey
- Distribution type: Round-robin, Distribution column: RegionKey
- Distribution type: Round-robin, Distribution column: ProductKey
- Distribution type: Hash-distributed, Distribution column: ProductKey

Report Error


ep
TEST MO
Q. 41
Scenario: You are working on a new project and you are in a meeting to discuss which Azure data platform technology is best for your company.
Requirement: A globally distributed, multimodel database that can perform queries in less than a second.
Which of the following should you choose?
Azure SQL Data Warehouse
Azure SQL on VM
Azure Databricks
Azure Cosmos DB
Azure Data Factory
Azure SQL Database
Report Error
1
8
15
22
29
36
43
PREV
NEXT
pe here to search


Q. 42
You need to recommend an Azure SQL Database service tier.
What should you recommend?
Premium
General Purpose
Basic
Business Critical
Standard
Report Error
TEST MO
1
8
15
22
29
36
43
PREV
NEXT
ype here to search


Q. 43
You need to design the solution for analyzing customer data.
What should you recommend?
Azure Batch
Azure Databricks
Azure SQL Data Warehouse
Azure Data Lake Storage
Azure Cognitive Services
Report Error
TEST MO
1
8
15
22
29
36
43
PREV
NEXT
pe here to search


Q. 44
You need to recommend a solution for storing the image tagging data.
What should you recommend?
Azure SQL Database
Azure SQL Data Warehouse
Azure File Storage
Azure Cosmos DB
Azure Blob Storage
Report Error
TEST MO
1
8
15
22
29
36
43
PREV
NEXT
pe here to search

Q. 45

You need to design a sharding strategy for the Planning Assistance database.

What should you recommend?

- a list mapping shard map on the binary representation of the License Plate column
- a range mapping shard map on the binary representation of the speed column
- a list mapping shard map on the location column
- a range mapping shard map on the time column

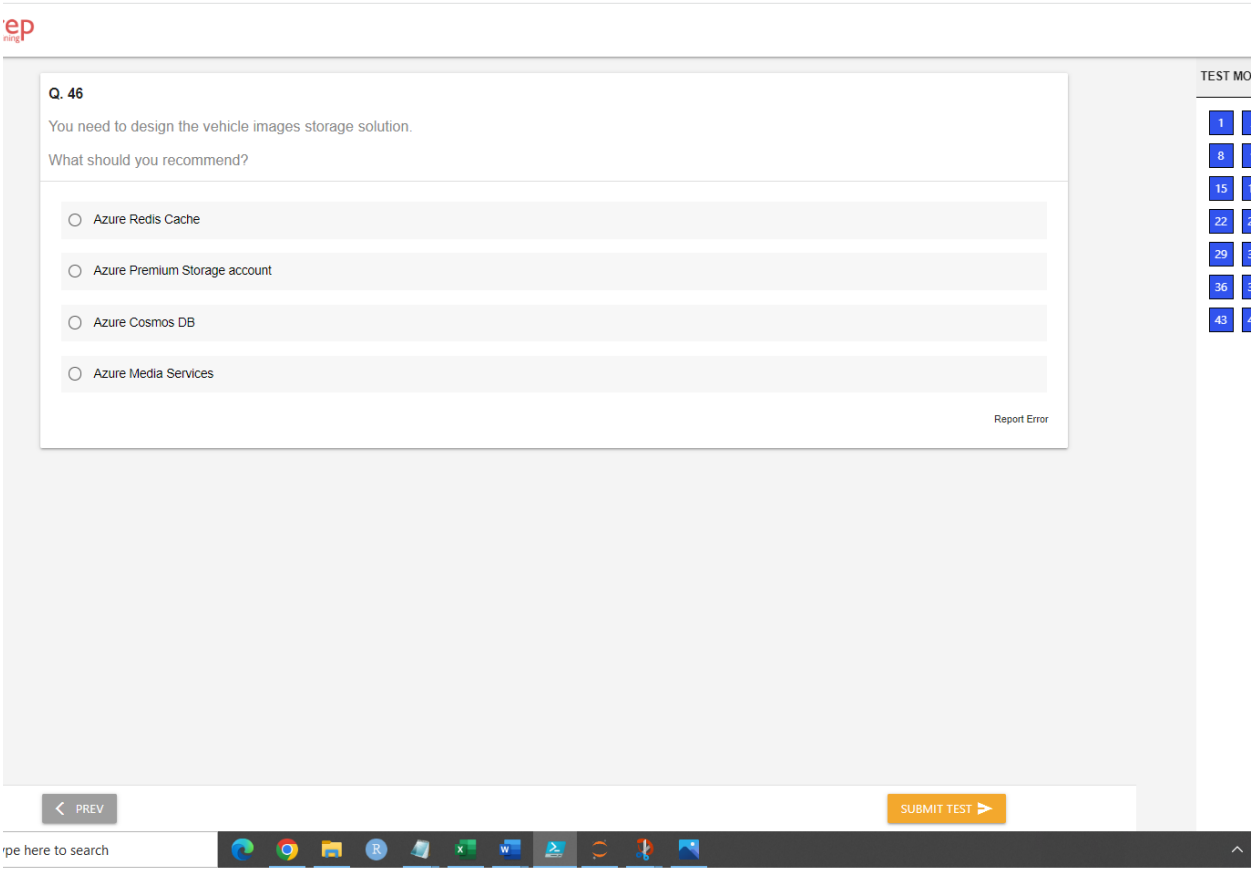
Report Error

TEST MO

1 8 15 22 29 36 43

PREV NEXT

ype here to search


Q. 46
You need to design the vehicle images storage solution.
What should you recommend?
Azure Redis Cache
Azure Premium Storage account
Azure Cosmos DB
Azure Media Services
Report Error
TEST MO
1
8
15
22
29
36
43
PREV
SUBMIT TEST
ype here to search