Practice Test 2

Question

Domain: Describe how to work with relational data on Azure

Which of the following is not true about a relational table?

○]A.

Values in the table are atomic

○]B.

All the rows in the table are unique

○]C.

The sequence of rows matters in the table.

○]D.

Each column needs to have a unique name.

Explanation:

Correct Answer: C
The sequence of rows does not matter in a relational table. The rows can be retrieved in any order and sequence. Adding data to a relational table does not affect the other existing queries.

- **Option A is incorrect.** A relational table needs to have atomic values. Having atomic values is the condition for the table, being in first normal form. Atomic values are the values that can't be divided. It means that columns in the relational table are not repeating groups. For example, in the following table, the coloUr column is not having atomic values. The value "Red, Black" for Item _Id 101 is not atomic and can be divided into two values, Red and black.

Item_Id	ColoUr	Price
101	Red, Black	100
102	White	50
103	Green	70

- **Option B is incorrect.** Unique rows is the essential characteristic of a relational table. No two rows in the table can't be the same. They need to vary for values of one or more columns that is ensured with the primary key in a relational table.
- **Option C is CORRECT.** The sequence of rows does not matter in a relational table. The rows can be retrieved in any order and sequence. Adding data to a relational table does not affect the other existing queries.
- **Option D is incorrect.** In a relational table, each column needs to have a unique name as columns are referred in the table by their name, not their position as the position of the columns in the relational table is insignificant. Moreover, the names of the columns must be different in a table not in the database which means that two different tables in a database can have the same name for a column.

Reference:
To know more about relational tables, refer to the following documentation:

- https://www.cs.wcupa.edu/~zjiang/RDB_table.htm

Question

Domain: Describe how to work with relational data on Azure

Which of the following statement is not true about views in a relational database?

○]A.

A view is a virtual table that has the content defined by a query.

○]B.

A view can be created from multiple tables, a single table or even another view

○]C.

One of the most important use of view is to simplify, focus and customize the users' perception of the database.

○]D.

A view is a special lookup table that the database search engine can utilize to speed up the data retrieval.

Explanation:

Correct Answer: D
A view is a virtual table that has the content defined by a query. Like a relational table, a view has a set of named rows and columns of data. A view can be created from multiple tables, a single table or even another view. Views are generally used to simplify, focus and customize the users' perception of the database. Views can help improve query response time.

- **Option A is incorrect** as a view is a virtual table that has the content defined by a query.
- **Option B is incorrect** as a view can be created from multiple tables, a single table or even another view.
- **Option C is incorrect** as Views are virtual tables that are generally used to simplify, focus and customize the users' perception of the database.
- **Option D is CORRECT** as it is the index, not view, that is a **special lookup table** that the database search engine can utilize to speed up the data retrieval.

Reference:
To know more about views in relational databases, refer to the following documentation from Azure:

- https://docs.microsoft.com/en-us/sql/relational-databases/views/views?view=sql-server-ver15

Question

Domain: Describe how to work with relational data on Azure

Statement 1: Azure SQL Database handles updating and patching the SQL Server software automatically without user involvement.
Statement 2: Azure SQL Database offers scalability that ensures that you can enhance the number of resources available to store.
Statement 3: Azure SQL Database has an important feature of point-in-time restore, through which the database can be recovered to the state of any specific point in the past.
Which of the following statements are true about Azure SQL database?

○]A.

Only statement 1 and 3

○]B.

Only Statement 2 and 3

○]C.

Only statement 1 and 2

○]D.

All statements 1, 2 and 3

Explanation:

Correct Answer: D
Azure SQL database is a completely managed PaaS (Platform-as-a-Service) database engine that supports and handles the important database management functions such as upgrading,

patching, monitoring, and backups automatically without the user's involvement. High availability, scalability, and point-in-time restore are other important features of Azure SQL database.

- **Option A is incorrect** as not only statement 1 and 3 but also statement 2 is true about Azure SQL database. Scalability is also an important feature of Azure SQL database.
- **Option B is incorrect** as not only statement 2 and 3 but also statement 1 is true about Azure SQL database. Azure SQL database handles updating and patching the SQL Server software automatically.
- **Option C is incorrect** as not only statement 1 and 2 but also statement 3 is true about Azure SQL database. Azure SQL database allows recovering the database to the state of any specific point in the past through its point-in-time restore feature.
- **Option D is CORRECT** as manually updating and patching, high scalability and point-in-time restore are important features of Azure SQL database.

Reference:

To know more about the features of Azure SQL Database, refer to the following documentation from Azure:

- https://docs.microsoft.com/en-us/azure/azure-sql/database/sql-database-paas-overview

Question

Domain: Describe how to work with relational data on Azure

Complete the statement.
In Azure SQL databases, transparent data encryption encrypts the…

○]A.

Database to protect the data in the motion

○]B.

Server to protect the data in the motion

○]C.

Database to protect the data at rest

○]D.

Server to protect the data at rest

Explanation:

Correct Answer: C

Transparent Data Encryption (TDE) protects Azure SQL database, Azure Synapse Analytics, and Azure SQL Managed Instance against malicious offline activities by encrypting the data at rest. Here, the database, associated backups, and transaction log files at rest can be encrypted and decrypted in real-time without the need for any change in the application.

- **Option A is incorrect**. Transparent Data Encryption (TDE) protects the data at rest; not in the motion. To protect the data in the motion, transport layer security is used.
- **Option B is incorrect.** Transparent Data Encryption (TDE) protects the data at rest; not in the motion. To protect the data in the motion, transport layer security is used.
- **Option C is CORRECT.** Transparent Data Encryption (TDE) encrypts the database to protects the data at rest.
- **Option D is incorrect.** Transparent Data Encryption (TDE) encrypts the database, not the server, to protects the data at rest.

Reference:
To know more about Transparent Data Encryption (TDE), refer to the following documentation from Azure:

- https://docs.microsoft.com/en-us/azure/azure-sql/database/transparent-data-encryption-tde-overview?tabs=azure-portal

Question

Domain: Describe how to work with relational data on Azure

Which of the following is a DML statement?

○]A.

Create

○]B.

Alter

○]C.

Insert

○]D.

Drop

Explanation:

Correct Answer: C
DML statements are the statements that are used to manipulate the data. On the other hand, DDL statements are statements that are used to define the data structure. In simple terms, DDL

statements are used to deal with the structure of the database and DML statement are used just to deal with the data, not the structure.

- **Option A is incorrect.** Create command is used to create a table on the database and therefore is a DDL command.
- **Option B is incorrect.** Alter command is used to alter/ change the structure of a table such as adding the column, dropping the column, change datatype or size of a column and rename the column. Therefore, it is a DDL command, not DML.
- **Option C is CORRECT.** Insert command is used to insert data in the table. The syntax for Insert is :

INSERT INTO table_name (column1, column2, …)

VALUES (value1, value2, …)

Option D is incorrect. Drop is again a DDL command that is used to remove an object from the database. For example, Drop table Student can be used to drop the student table from the database and therefore is a DDL statement.

Reference*:*
To know more about the DML statements, refer to the following documentation from Azure:

- https://docs.microsoft.com/en-us/learn/modules/query-relational-data/2-introduction-to-sql

Question

Domain: Describe how to work with relational data on Azure

What is the significance of Resource governance error "Error: 40544"?

○]A.

The database has reached its size quota

○]B.

The service is currently busy

○]C.

Session has been terminated because you have a long-running transaction

○]D.

The session has been terminated because of excessive memory usage

Explanation:

Correct Answer: A
Error: 40544 indicates that the database has reached its size quota. Some possible resolutions, in this case, are drop the indexes, partition or delete data, or check the documentation for resolutions.

- **Option A is CORRECT.** Error: 40544 indicates that the database has reached its size quota.
- **Option B is incorrect.** It is the error *Error 40501* that indicates that the service is currently busy.
- **Option C is incorrect.** It is the error *Error 40549 that indicates that the* session has been terminated because you have a long-running transaction
- **Option D is incorrect.** It is the error *Error 40553 that indicates that* The session has been terminated because of excessive memory usage

Reference:
To know more about the common error issues, refer to the following documentation from Azure:

- https://docs.microsoft.com/en-us/azure/azure-sql/database/troubleshoot-common-errors-issues

Question

Domain: Describe how to work with relational data on Azure

In the following SQL query:
INSERT INTO Students (StudentName, StudentID, Marks, Address) VALUES ('Ram', 1501, 87, 'Delhi')
What are the *Students* and *Marks* respectively?

○]A.

Column and table

○]B.

Index and Column

○]C.

Table and Primary key

○]D.

Table and Column

Explanation:

Correct Answer: D
INSERT is a command in SQL with syntax

INSERT INTO *table name* (column 1, column 2, column 3, …) VALUES (Value 1, Value 2, Value 3, …). Therefore, in the preceding query, Students is a table and Marks is a column of the Students table.
- **Option A is incorrect** as according to the syntax for INSERT, Students is the table and Marks is the column.
- **Option B is incorrect** as Students is a table, not an index. INSERT command is never applied on the indexes.
- **Option C is incorrect** as in the given SQL query, Marks is a simple column as it is not explicitly defined as the primary key.
- **Option D is CORRECT** as according to the syntax of INSERT command, Students is the table and Marks is the column.

Reference:
To know more, refer to the following documentation from Azure:

- https://docs.microsoft.com/en-us/learn/modules/query-relational-data/2-introduction-to-sql

Question

Domain: Describe how to work with relational data on Azure

Read the Statement:
Both Azure data studio and SQL Server Management Studio support all operating systems such as Windows, macOS and Linux.
The preceding statement is…

⊙]A.**True**

⊙]B.**False**

Explanation:

Correct Answer: B
Not all studios supports all operating systems. Windows is the operating system that is supported by both Azure Data Studio and SQL Server Management Studio. Whereas MacOS and Linux support only Azure data studio, not SQL Server Management Studio.

- **Option A is incorrect** as SQL server management studio is only supported by Windows, not by MacOS and Linux.
- **Option B is CORRECT** as MacOS and Linux support only Azure data studio, not SQL Server Management Studio.

Reference:
To know more about the Azure Data Studio, refer to the following documentation from Azure:

- https://docs.microsoft.com/en-us/sql/azure-data-studio/what-is-azure-data-studio?view=sql-server-ver15

Question

Domain: Describe how to work with relational data on Azure

Complete the sentence:
PostgreSQL has its own query language that is known as…

○]A.

pgsql

○]B.

Oracle

○]C.

SQL

○]D.

MYSQL

Explanation:

Correct Answer: A
PostgreSQL has its own query language known as *pgsql*. This query language is a specific variant of the SQL (which is a standard relational query language) that allows to write stored procedures that run/ execute inside the database.

- **Option A is CORRECT.** Pgsql is a specific variant of SQL that is used by PostgreSQL.
- **Option B is incorrect.** Oracle is a relational database management system, not any query language.
- **Option C is incorrect.** SQL is a standard relational query language while pgsql is a specific variant of SQL with some extra features that is used by PostgreSQL.
- **Option D is incorrect**. MYSQL is an open-source relational database management system, not any query language.

Reference:
To know more about the PostgreSQL, refer to the following documentation from Azure:

- https://docs.microsoft.com/en-us/learn/modules/explore-relational-data-offerings/6-postgresql-mariadb-mysql

Question

How many deployment options are available for *Azure Database for PostgreSQL*?

○]A.

One

○]B.

Two

○]C.

Three

○]D.

Four

Explanation:

Correct Answer: C
There are three deployment options available for Azure Database for PostgreSQL. These options are Single-server, Flexible Server (Preview), and Hyperscale.

Single-server deployment option provides and supports the features similar to the Azure Database for MySQL. With this option, the user can select from three available pricing tiers: basic, general purpose, and memory optimized. These tiers vary in the number of memory, CPUs, and storage sizes, and can be selected depending upon the load.

Flexible Server (Preview) deployment option provides more granular control and flexibility over database management functionalities and configuration settings. It offers better cost optimization control with the capability to stop or start the server and burstable compute tier, ideal for workloads that do not require continuous full compute capacity.

Hyperscale (Citus) is the deployment option that supports large database loads by scaling queries across several server nodes. Here, the database is split across nodes and Data is split into chunks depending upon the value of a sharding or partition key.

- **Option A is incorrect** as Single server, Flexible Server (Preview) and Hyperscale are the three deployment options available for *Azure Database for PostgreSQL*.
- **Option B is incorrect** as Single server, Flexible Server (Preview) and Hyperscale are the three deployment options available for *Azure Database for PostgreSQL*.

- **Option C is CORRECT** as Single server, Flexible Server (Preview) and Hyperscale are the three deployment options available for *Azure Database for PostgreSQL.*
- **Option D is incorrect** as Single server, Flexible Server (Preview) and Hyperscale are the three deployment options available for *Azure Database for PostgreSQL.*

Reference:
To know more about the Azure Database for PostgreSQL, refer to the following documentation from Azure:

- https://docs.microsoft.com/en-us/azure/postgresql/overview

Question

Domain: Describe how to work with relational data on Azure

Which of the following is not the specific feature of Azure Database for MySQL?

○]A.

Elastic scaling

○]B.

Automation and monitoring to streamline the management for large-scale deployments

○]C.

Protecting the data with automatic backups and point-in-time store for up to 35 days

○]D.

Capability of storing and manipulating geometric data, such as polygons, circles, and lines.

Explanation:

Correct Answer: D
Azure Database for MySQL is a relational database service in the Microsoft cloud with the following features:

- Built-in high availability
- Protecting the data with automatic backups and point-in-time store for up to 35 days
- Automated maintenance for the underlying operating system, hardware, and database engine to maintain the services secured and up to date
- Predictable performance through inclusive pay-as-you-go pricing
- Elastic scaling within no significant time
- Cost optimization controls with the ability to stop/start server
- Industry-leading compliance and Enterprise-grade security for protecting sensitive data-in-motion and at-rest

- Automation and monitoring to streamline the management for large-scale deployments
- Industry-leading support experience
- **Option A is incorrect** as Azure Database for MySQL supports Elastic scaling within no significant time.
- **Option B is incorrect** as Azure Database for MySQL supports automation and monitoring to streamline the management for large-scale deployments.
- **Option C is incorrect** as Azure Database for MySQL protects the data with automatic backups and point-in-time store for up to 35 days.
- **Option D is CORRECT** as the Capability of storing and manipulating geometric data, such as polygons, circles, and lines is not the specific feature of Azure database for MYSQL. But it is the feature of Azure Database for PostgreSQL.

Reference:

To know more about the Azure Database for SQL, refer to the following documentation from Azure:

- https://docs.microsoft.com/en-us/azure/mysql/overview
- https://docs.microsoft.com/en-us/learn/modules/explore-relational-data-offerings/6-postgresql-mariadb-mysql

Question

Domain: Describe how to work with relational data on Azure

There is a Student Management System with the following "Student" table:

Student_Name	Registration_No	Address
Rahul	10301	Delhi
Ram	10302	Agra
Shyam	10303	Noida

Now, there is a need to change the name of the "Registration_No" column to "Roll_No". Which statement would you use?

○]A.

UPDATE

○]B.

ALTER

○]C.

CREATE

○]D.

INSERT

Explanation:

Correct Answer: B
For a function such as changing the column name, size, or datatype, the Alter command is used.

Syntax for ALTER:

ALTER TABLE *Tablename*
Rename Column *OldColumnName* TO *NewColumnName*

- **Option A is incorrect** as UPDATE statement is used to modify the data; not the name of the column.
- **Option B is CORRECT** as ALTER statement is used to change the structure of the table. As here we need to rename the column, the following statement can be used:

ALTER TABLE Student

RENAME COLUMN Registration_No TO Roll_no ;

- **Option C is incorrect** as the CREATE statement is used to create the table, not to rename the column.
- **Option D is incorrect** as INSERT is used to insert the data into the table.

Reference:
To know more, refer to the following documentation from Azure:

- https://docs.microsoft.com/en-us/learn/modules/query-relational-data/2-introduction-to-sql

Question

Domain: Describe how to work with relational data on Azure

A relational database is used when…

○]A.

A dynamic schema is needed

○]B.

Data is to be stored in non-tabular form

○]C.

High consistency is needed.

○]D.

There is a need to store large videos and images

Explanation:

Correct Answer: C
Dynamic and flexible schema, non-tabular data storage and capability to store large videos and images are the features of non-relational databases. A relational database stores the data in tables and ensures high consistency.

- **Option A is incorrect** as the dynamic schema is the feature of the non-relational database; not the relational one.
- **Option B is incorrect** as in the relational database, the data is stored in tables. It is a non-relational database where the data is stored in non-tabular form such as key-value pairs, graphs, columnars and documents.
- **Option C is CORRECT** as high consistency is the major feature of Relational databases.
- **Option D is incorrect** as the non-relational database is used to store large images and videos.

References:
To know more about the Properties of RBDMSs and NoSQL databases, refer to the following documentation from Azure:

- https://docs.rackspace.com/support/how-to/properties-of-rdbmss-and-nosql-databases/

Question

Domain: Describe how to work with relational data on Azure

Which of the following Azure tools describes the service(s) to be deployed in JSON (JavaScript Object Notation) format?

○]A.

Azure Portal

○]B.

Azure command-line interface (CLI)

○]C.

Azure PowerShell

○]D.

Azure Resource Manager templates

Explanation:

Correct Answer: D
An Azure Resource Manager template describes the service(s) to be deployed in a text file, in JSON format. The following is an example template that you can be utilized to provision an instance of Azure SQL Database:

```json
JSON

"resources": [
  {
    "name": "sql-server-dev",
    "type": "Microsoft.Sql/servers",
    "apiVersion": "2014-04-01-preview",
    "location": "[parameters('location')]",
    "tags": {
      "displayName": "SqlServer"
    }
         "properties": {}
  }
]
```

- **Option A is incorrect.** The Azure portal shows the sequence of service-specific pages that prompt you for the settings needed and confirms these settings before the service is actually provisioned.
- **Option B is incorrect.** CLI offers a set of commands that can be run from the operating system (OS) command prompt or the Cloud Shell in the Azure portal. These commands can be used for creating and managing the Azure resources.
- **Option C is incorrect.** Azure PowerShell provides the set of commands that can be used for scripting and automating the administrative tasks. Azure offers a series of commandlets, that are Azure-specific commands, that can be utilized in PowerShell for creating and managing the Azure resources.
- **Option D is CORRECT.** An Azure Resource Manager template describes the service(s) to be deployed in a text file, in JSON format.

Reference:
To know more about the relation data services, refer to the following documentation from Azure:

- https://docs.microsoft.com/en-us/learn/modules/explore-provision-deploy-relational-database-offerings-azure/2-describe-provision-relational-data-services

Question

Domain: Describe how to work with non-relational data on Azure

Your organization needs to design a data store that will have employees data. The data needs to be stored in the following format:

Employee_Id	Employee_Information
1001	FName: Aron LName: Smith Email: abc@ajopo.com
1002	FirstName: Ben Last: Stokes Contact: 123-456-389
1003	FirstName: David LastName: Hussain

Which kind of data store would you use?

○]A.

Document

○]B.

Graph

○]C.
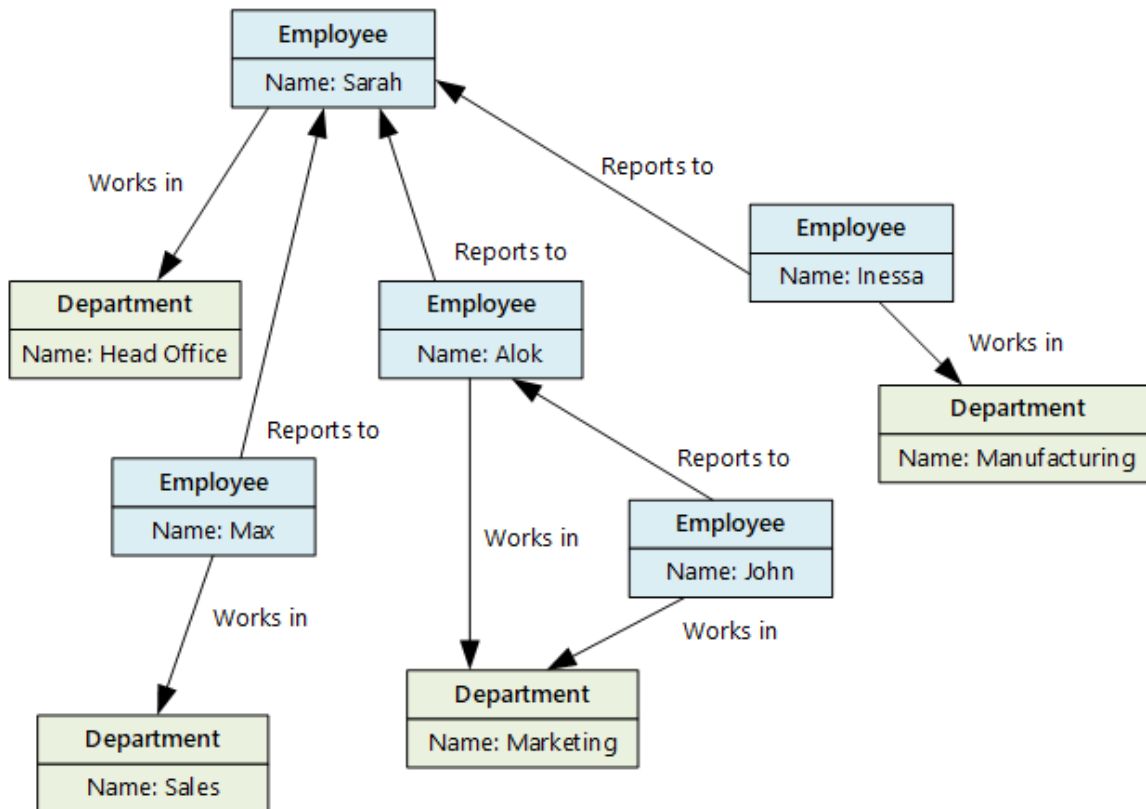
Columnar

○]D.

Key/Value

Explanation:

Correct Answer: C

Columnar or column family data store stores the data in form of tables having rows, and dynamic columns. This data store is more flexible than relational databases as here each row does not need to have the same columns. This data store is preferable when there is a need to store a large amount of data and the query patterns are predictable.

- **Option A is incorrect.** A document data store stores the data in documents in JSON (JavaScript Object Notation) format. Here, each document has pairs of fields and values. A field value can be a scalar item such as number or string, or a compound item such as parent-child collection or list. A document has complete data for an entity. For example, an entity can have the details of the customer, order placed by that customer, or both. The following is an example of document data store:

Key	Document
1001	```{ "CustomerID": 99, "OrderItems": [{ "ProductID": 2010, "Quantity": 2, "Cost": 520 }, { "ProductID": 4365, "Quantity": 1, "Cost": 18 }], "OrderDate": "04/01/2017" }```
1002	```{ "CustomerID": 220, "OrderItems": [{ "ProductID": 1285, "Quantity": 1, "Cost": 120 }], "OrderDate": "05/08/2017" }```

- **Option B is incorrect.** A graph data store provides two types of information in form of nodes and edges. Here, the nodes represent entities, and the edges indicate the relationships among those entities. The following figure shows an example of Graph data store.

- **Option C is CORRECT** as data is available in the format of a table with rows and columns that is the major feature of columnar data type along with additional flexibility in format.
- **Option D is incorrect.** Key/value store is a kind of large hash table where each data value is associated with a unique key. This key is used to store the data with the help of a hashing function. The following is an example of key-value data store.

Key	Value
AAAAA	1101001111010100110101111...
AABAB	1001100001011001101011110...
DFA766	0000000000101010110101010...
FABCC4	1110110110101010100101101...

Reference:
To know more about the different type of data stores, please refer to the following Azure doc:

- https://docs.microsoft.com/en-us/azure/architecture/data-guide/big-data/non-relational-data

Question

Read the following statement:
You should use Azure File Storage for the files that are likely to be written by multiple concurrent processes simultaneously.
Choose the correct option.

○]A.**True**

○]B.**False**

Explanation:

Correct Answer: B
It is never recommended to use Azure File Storage for the files that are likely to be written by multiple concurrent processes simultaneously. Multiple writers need careful synchronization, otherwise, the updates or changes performed by one process are likely to be overwritten by other processes. The possible option is to lock the file while it is in writing mode, and as the write operation is complete release the lock. But, this option can highly impact the concurrency and limit performance.

Reference:
To know more about the Azure File Storage, please refer to the Azure documentation:

- https://azure.microsoft.com/en-in/blog/azure-file-storage-now-generally-available/

Question

At which of the following levels, it is possible to set the throughput for an Azure Cosmos DB account? (Select two options.)

☐ A.

Partition

☐ B.

Database

☐ C.

Container

☐ D.

Item

☐ E.

Value

Explanation:

Correct Answers: B and C
With Azure Cosmos DB, throughput can be provisioned at two granularities: Azure Cosmos containers and Azure Cosmos databases. Provisioned throughput can be of two types - standard (manual) or auto-scale.

- **Option A is incorrect.** With Azure Cosmos DB, the throughput can be provisioned only at two levels: Database and container.
- **Option B is CORRECT**. With Azure Cosmos DB it is possible to provision the throughput at the database level.
- **Option C is CORRECT.** With Azure Cosmos DB it is possible to provision the throughput at the container level.
- **Option D is incorrect.** With Azure Cosmos DB, the throughput can be provisioned only at two levels: Database and container.
- **Option E is incorrect.** With Azure Cosmos DB, the throughput can be provisioned only at two levels: Database and container.

Reference:
To know more about the Azure Cosmos Account, please refer to the Azure documentation:

- https://docs.microsoft.com/en-us/azure/cosmos-db/set-throughput

Question

Domain: Describe how to work with non-relational data on Azure

Which of the following API of Azure Cosmo Database provides a graph-based view on database?

○]A.

MongoDB API

○]B.

Cassandra AP

○]C.

Gremlin

○]D.

Table API

Explanation:

Correct Answer: C
Cosmo DB supports 5 APIs: Core (SQL) API, MongoDB API, Cassandra API, Table API, and Gremlin API.

Core (SQL) API is the default API that provides a view of that data that resembles a traditional NoSQL document store. SQL-like language is used to query the hierarchical JSON documents with a SQL-like language.

MongoDB is a document database that has its own programmatic interface. Therefore, the data structure associated with MongoDB is JSON documents. Cassandra is a column family database management system that supports querying data through Cassandra Query Language (CQL).

Gremlin provides a graph-based view of the database. It means that the data is either a vertex (an individual item in the database), or an edge (a relationship between the database items).

- **Option A is incorrect.** MongoDB is a document database that has its own programmatic interface. Therefore, data structure associated with MongoDB is JSON documents.
- **Option B is incorrect.** Cassandra is a column family database management system that supports querying data through Cassandra Query Language (CQL).
- **Option C is CORRECT.** Gremlin provides a graph-based view of the database. It means that the data is either a vertex (an individual item in the database), or an edge (a relationship between the database items).
- **Option D is incorrect.** Table API allows using the Azure Table Storage API to store and retrieve the documents. The associated data structure with Table API is key-value.

Reference:
To know more about the API for Azure Cosmos Database, please refer to the Azure documentation:

- https://docs.microsoft.com/en-us/learn/modules/choose-api-for-cosmos-db/2-identify-the-technology-options

Question

Domain: Describe how to work with non-relational data on Azure

Match the common error status code with their significance

Status Code	Description / Significance
a) 401	1) Service Unavailable
b) 408	2) Too many Requests
c) 429	3) Request timed out
d) 503	4) Not authorized

○]A.

a1 b2 c3 d4

○]B.

a3 b2 c1 d4

○]C.

a4 b3 c2 d1

○]D.

a2 b4 c3 d1

Explanation:

Correct Answer: C
401- Not authorized

408- Request Timed out

429- Too many Requests

503- Service Unavailable

Error HTTP 401 states that "the MAC signature found in the HTTP request is not the same as the computed signature".

Error HTTP 408 happens if the SDK was unable to complete the request before the timeout limit occurred.

Error HTTP 429 indicates that the requests are being throttled.

Error HTTP 503 indicates that the SDK was not able to connect to Azure Cosmos DB.

Reference:
To know more, please refer to the Azure documentation:

-

Question

Domain: Describe how to work with non-relational data on Azure

Which of the following is the right sequence for CosmosDB levels demonstrating the strongest consistency to lowest consistency (means the first level in sequence having the strongest consistency and the last level showing the least consistency)?

○]A.

Strong, Bounded staleness, Consistent prefix, Session, and Eventual

○]B.

Strong, Session, Consistent prefix, Bounded staleness, and Eventual

○]C.

Strong, Bounded staleness, Session, Consistent prefix and Eventual

○]D.

Strong, Bounded staleness, Session, Eventual and Consistent prefix

Explanation:

Correct Answer: C
Azure Cosmos DB supports five well-defined levels varying in term of consistency, latency, availability, and throughput. Here are the five levels from strongest to weakest: Strong, Bounded staleness, Session, Consistent prefix, and Eventual.

Each level offers performance and availability tradeoffs. The following image demonstrates the various consistency levels as a spectrum:
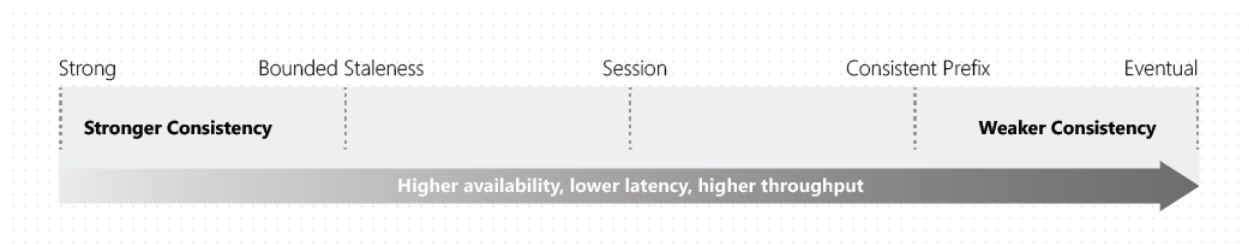
Strong consistency: With this option, the client sees the writes only after the modification/changes are confirmed as written successfully to all replicas. Uncommitted or partial writes are never visible to the clients. Strong consistency option is unavailable if there is a need for distributing the data across multiple global regions.

Bounded Staleness: With this option, there is a lag between writing and reading the updated data. While choosing bounded staleness, staleness can be defined in two ways;

1. number of versions (K) of item
2. The time interval (T) reads might lag behind the writes

Session: During a single client session, reads are certain to honor the consistent-prefix, monotonic reads and writes, read-your-writes, and write-follows-reads guarantees. It considers a single "writer" session or sharing the session token for many writers.

Consistent Prefix: This option guarantees that changes appear in order, although there is a possible delay before they are visible. Within that period, the applications are likely to see old data.

Eventual Consistency: This is weakest consistency option with no ordering guarantee for reads. A client may read the values even older than the ones it had read before. This consistency option is suitable for the applications that don't need any ordering guarantees. Examples are count of Retweets, non-threaded comments or likes.

Reference:
To know more about the different levels supported by Azure Cosmos DB, please refer to the Azure documentation:

- https://docs.microsoft.com/en-us/azure/cosmos-db/consistency-levels

Question

Domain: Describe how to work with non-relational data on Azure

Which of the followings can be used to set the IP firewall on the Azure Cosmos DB account?

○]A.

Azure portal

○]B.

Azure Resource Manager template

○]C.

Azure CLI

○]D.

Azure PowerShell

○]E.

All of the above

Explanation:

Correct Answer: E
IP firewall can be set on Azure Cosmos DB account through any of the following ways:

- Using the Azure portal
- Declaratively through an Azure Resource Manager template
- Programmatically using the Azure CLI or Azure PowerShell by updating the **ipRangeFilter** property

1.Configuring IP firewall through Azure Portal
Go to the Azure Cosmos DB account page and on the navigation menu, choose **Firewall and virtual networks**. update the **Allow access from** value to **Selected networks**, and finally click **Save**.
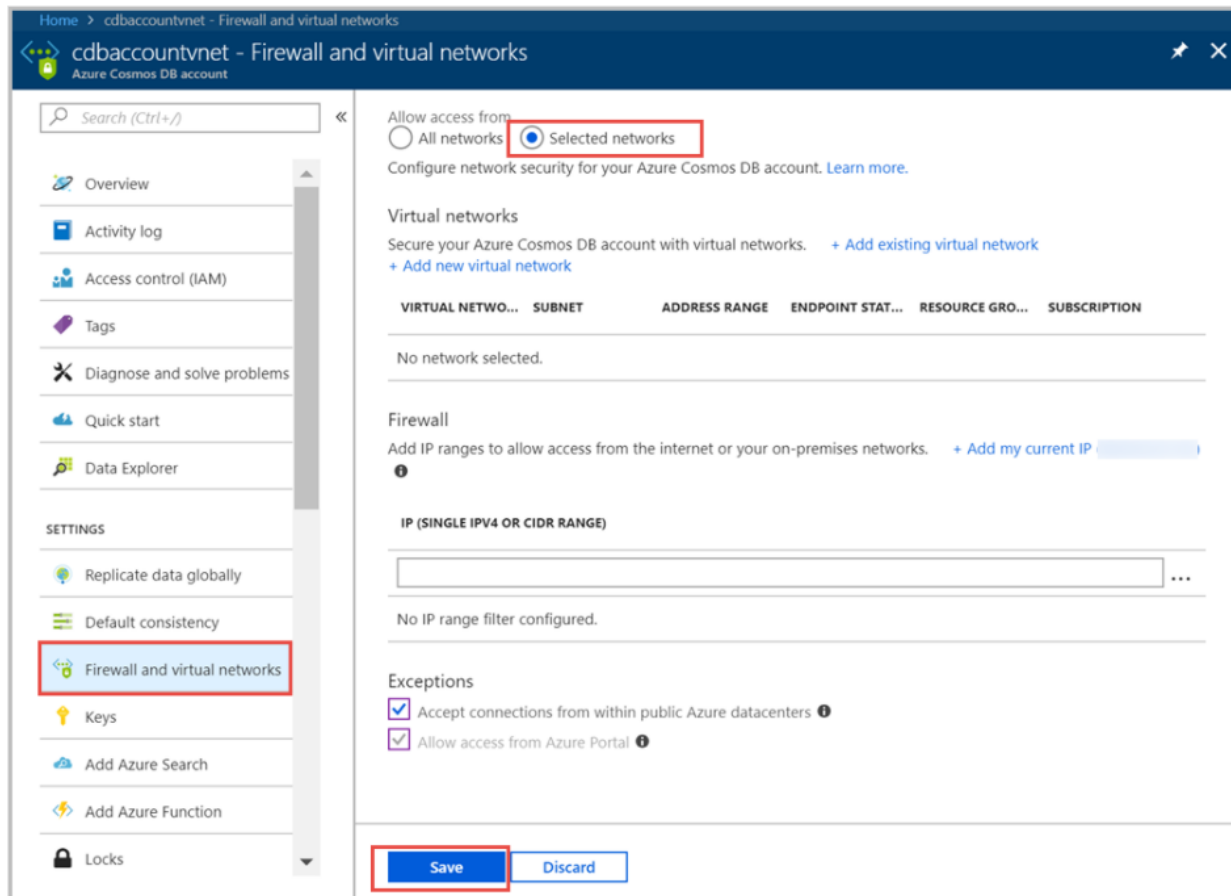
2. Configuring an IP firewall through a Resource Manager template
while configuring access control to Azure Cosmos DB account, ensure that the Resource Manager template specifies the **ipRules** property with an array of allowed IP ranges.
The **ipRules** property was added with API version 2020-04-01. In earlier versions, **ipRangeFilter** property is exposed, that is a list of comma-separated IP addresses.
Example to demonstrate how the **ipRules** property is exposed in API version 2020-04-01 or later:

```json
JSON                                                          Copy
{
  "type": "Microsoft.DocumentDB/databaseAccounts",
  "name": "[variables('accountName')]",
  "apiVersion": "2020-04-01",
  "location": "[parameters('location')]",
  "kind": "GlobalDocumentDB",
  "properties": {
    "consistencyPolicy": "[variables('consistencyPolicy')[parameters('defaultConsistencyLevel')]]",
    "locations": "[variables('locations')]",
    "databaseAccountOfferType": "Standard",
    "enableAutomaticFailover": "[parameters('automaticFailover')]",
    "ipRules": [
      {
        "ipAddressOrRange": "40.76.54.131"
      },
      {
        "ipAddressOrRange": "52.176.6.30"
      },
      {
        "ipAddressOrRange": "52.169.50.45"
      },
      {
        "ipAddressOrRange": "52.187.184.26"
      }
    ]
  }
}
```

Example to demonstrate API version prior to 2020-04-01

```json
JSON                                                          Copy
{
  "type": "Microsoft.DocumentDB/databaseAccounts",
  "name": "[variables('accountName')]",
  "apiVersion": "2019-08-01",
  "location": "[parameters('location')]",
  "kind": "GlobalDocumentDB",
  "properties": {
    "consistencyPolicy": "[variables('consistencyPolicy')[parameters('defaultConsistencyLevel')]]",
    "locations": "[variables('locations')]",
    "databaseAccountOfferType": "Standard",
    "enableAutomaticFailover": "[parameters('automaticFailover')]",
    "ipRangeFilter":"40.76.54.131,52.176.6.30,52.169.50.45,52.187.184.26"
  }
}
```

3. Configuring an IP access control policy through Azure CLI

```
# Create a Cosmos DB account with default values and IP Firewall enabled
resourceGroupName='MyResourceGroup'
accountName='mycosmosaccount'
ipRangeFilter='192.168.221.17,183.240.196.255,40.76.54.131'

# Make sure there are no spaces in the comma-delimited list of IP addresses or CIDR ranges.
az cosmosdb create \
    -n $accountName \
    -g $resourceGroupName \
    --locations regionName='West US 2' failoverPriority=0 isZoneRedundant=False \
    --locations regionName='East US 2' failoverPriority=1 isZoneRedundant=False \
    --ip-range-filter $ipRangeFilter
```

4. Configuring an IP access control policy through PowerShell

```
# Create a Cosmos DB account with default values and IP Firewall enabled
$resourceGroupName = "myResourceGroup"
$accountName = "mycosmosaccount"
$ipRules = @("192.168.221.17","183.240.196.255","40.76.54.131")

$locations = @(
    @{ "locationName"="West US 2"; "failoverPriority"=0; "isZoneRedundant"=False },
    @{ "locationName"="East US 2"; "failoverPriority"=1, "isZoneRedundant"=False }
)

# Make sure there are no spaces in the comma-delimited list of IP addresses or CIDR ranges.
$CosmosDBProperties = @{
    "databaseAccountOfferType"="Standard";
    "locations"=$locations;
    "ipRules"=$ipRules
}

New-AzResource -ResourceType "Microsoft.DocumentDb/databaseAccounts" `
    -ApiVersion "2020-04-01" -ResourceGroupName $resourceGroupName `
    -Name $accountName -PropertyObject $CosmosDBProperties
```

Reference:
To know more about how to configure firewall, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/azure/cosmos-db/how-to-configure-firewall#troubleshoot-ip-firewall

Question

Domain: Describe how to work with non-relational data on Azure

A system is required to be built to monitor the temperature throughout a number of office blocks, and set the air conditioning in every room in every block to have a pleasant temperature. Your system needs to manage the air conditioning in many thousand buildings throughout the country, and each building has at least 110 air-conditioned rooms. What kind of NoSQL data store is most suitable to capture the temperature data so that it can be processed quickly?

○]A.

Storing the data in a file stored in a share, created through Azure File Storage.

○]B.

Send the data to an Azure Cosmos DB database and utilize Azure Functions for processing the data.

○]C.

Use blobs to write the temperatures in Azure Blob storage.

○]D.

Use table to store temperature

Explanation:

Correct Answer: B
Azure Cosmos DB is capable to ingest huge volumes of data quickly. A thermometer in every room can send the temperature data to an Azure Cosmos DB database. An Azure Function can be run as each and every item is stored. This function can check the temperature, and activate a remote process to configure the air conditioning in the rooms.

- **Option A is incorrect.** There will be a need of separate processes in every room or building to capture the data in the file. Although data can be read by multiple processes simultaneously from a file stored in Azure File Storage, it is the multiple writers that can create problems, and the writes done by one process can be overwritten by another that will result in losing the data.
- **Option B is CORRECT.** Azure Cosmos DB is capable to ingest huge volumes of data quickly. A thermometer in every room can send the temperature data to a Azure Cosmos DB database. An Azure Function can be run as each and every item is stored. This function can check the temperature, and activate a remote process for configuring the air conditioning in the rooms.
- **Option C is incorrect**. For the given scenario, blob storage is not a good option.
- **Option D is incorrect**. For the given scenario, table storage is not a good option.

Reference:
To know more, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/azure/cosmos-db/migrate-cosmosdb-data

Question

Which of the following is a benefit of using multi-region replication with Cosmos DB?

◯]A.

Increased Availability

◯]B.

Consistent data in every region

◯]C.

Enhanced data security

◯]D.

None of these

Explanation:

Correct Answer: A
With multi-region replication in Cosmos DB, data is replicated across multiple regions that increases the availability. If a region is inaccessible, the data is still available in other regions and can be retrieved from there.

- **Option A is CORRECT.** With multi-region replication in Cosmos DB data is replicated across multiple regions that increases the availability.
- **Option B is incorrect.** It is the availability, not the consistency, that is the benefit of multi-region replication with Cosmos DB.
- **Option C is incorrect** as it's the encryption that enhances data security. Availability, not the enhanced data security, is the benefit of multi-region replication with cosmos DB.
- **Option D is incorrect** as increased availability is the benefit of using multi-region replication with Cosmos DB.

Reference:
To know more about the benefit of using multi-region replication with Cosmos DB, please refer to the Azure documentation:

- https://docs.microsoft.com/en-us/azure/cosmos-db/distribute-data-globally

Question

Domain: Describe an analytics workload on Azure

Consider the following statements:
S1: Bring data into Power BI Desktop, and create a report
S2: Publish to the Power BI service, where you can create new visualizations or build dashboards
S3: Share dashboards with others, especially people who are on the go
S4: View and interact with shared dashboards and reports in Power BI Mobile apps.
Which of the following sequence of the preceding statements represents the common flow of activities in Power BI?

○]A.

S1-S2-S4-S3

○]B.

S2-S1-S3-S4

○]C.

S1-S2-S3-S4

○]D.

S2-S3-S1-S4

Explanation:

Correct Answer: C
The activities and analyses in Power BI generally follow a common flow as follows:

1. Bring data into Power BI Desktop, and create a report.
2. Publish to the Power BI service, where you can create new visualizations or build dashboards.
3. Share dashboards with others, especially people who are on the go.
4. View and interact with shared dashboards and reports in Power BI Mobile apps.

So, the sequence S1-S2-S3-S4 is the right sequence.

- **Option A is incorrect** as in normal flow, Share dashboards with others, especially people who are on the go comes before View and interact with shared dashboards and reports in Power BI Mobile apps.
- **Option B is incorrect** as it is Bring data into Power BI Desktop, and create a report, not Publish to the Power BI service, where you can create new visualizations or build dashboards, that is the first activity in the normal flow.
- **Option C is correct**. Bring data into Power BI Desktop, and create a report, Publish to the Power BI service, where you can create new visualizations or build dashboards, Share dashboards with others, especially people who are on the go, and finally View

and interact with shared dashboards and reports in Power BI Mobile apps is the common flow in Power BI.

- **Option D is incorrect** as Bring data into Power BI Desktop, and create a report is the first activity in the normal flow. The right sequence of common flow is - Bring data into Power BI Desktop, and create a report, Publish to the Power BI service, where you can create new visualizations or build dashboards, Share dashboards with others, especially people who are on the go and finally View and interact with shared dashboards and reports in Power BI Mobile apps is the common flow in Power BI.

Reference:
To know more about Power BI, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/learn/modules/get-started-with-power-bi/2-using-power-bi

Question

Domain: Describe an analytics workload on Azure

Which of the following element of Azure Synapse enables you to access all the Synapse Analytics tools?

○]A.

Synapse Spark pool

○]B.

Synapse pipelines

○]C.

Synapse Link

○]D.

Synapse Studio

Explanation:

Correct Answer: D
Synapse SQL pool, Synapse Spark pool, Synapse Pipelines, Synapse Link, and Synapse Studio are the elements of Azure Synapse. Out of these elements, Synapse Studio is a web user interface that enables the data engineers to access all the Synapse Analytics tools. Synapse Studio can be used for creating SQL and Spark pools, defining and running pipelines, and configuring links to external data sources.

- **Option A is incorrect.** Synapse Spark Pool is the element that supports Azure Machine Learning through integration with the AzureML and SparkML packages.

- **Option B is incorrect.** A Synapse pipeline represents a logical grouping of activities that collectively perform a task. The activities of the pipeline define the actions to be performed on your data.
- **Option C is incorrect**. Synapse Link is the component that enables you to connect to Azure Cosmos DB. It can be used to perform/run near real-time analytics over the operational data stored in an Azure Cosmos DB.
- **Option D is correct.** Synapse Studio is a web user interface that enables data engineers to access all the Synapse Analytics tools. Synapse Studio can be used for creating SQL and Spark pools, defining and running pipelines, and configuring links to external data sources.

Reference:

To know more about the Azure Synapse, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/learn/modules/explore-data-storage-processing-azure/3-explore-azure-synapse-analytics

Question

Domain: Describe an analytics workload on Azure

Which of the following services is useful to ingest data into Azure Synapse Analytics?

○]A.

Power BI

○]B.

Azure Data Factory

○]C.

Azure Active Directory

○]D.

Azure Databricks

○]E.

Azure Data Lake Storage

Explanation:

Correct Answer: B

Azure Data Factory is a *data integration service* that supports retrieving data from one or multiple data sources, and converting it into the desired format that you process. It can also be used to ingest data into Azure Synapse Analytics from almost every source.

- **Option A is incorrect.** Power BI is an analytics tool that can be utilized to report on the data within Azure Synapse Analytics.
- **Option B is correct.** Azure Data factory can be used to ingest data into Azure Synapse Analytics from almost every source.
- **Option C is incorrect.** Azure Active Directory is an authentication method. It is not a data ingestion framework that could be used to ingest data into Azure Synapse Analytics.
- **Option D is incorrect.** Azure Databricks is an Apache Spark environment that runs on Azure to offer big data processing, streaming, and machine learning.
- **Option E is incorrect.** Azure Data Lake Storage is a secure cloud platform that offers cost-effective and scalable storage for big data analytics.

Reference:
To know more about Azure Data Factory, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/azure/data-factory/introduction

Question

Domain: Describe an analytics workload on Azure

You have retrieved data formatted in a different format from multiple sources. Now, you need to transform the data into a single uniform format. Which of the following data services would you use?

○]A.

Azure Data Factory

○]B.

Azure Data Lake Storage

○]C.

Azure Databricks

○]D.

None of These

Explanation:

Correct Answer: A
Azure Data Factory is a *data integration service* that allows users to retrieve data from one or multiple data sources, and convert it into the desired format. The various data sources might

have different data representation and have *a noise* that is needed to be filtered out. Azure Data Factory allows to extract only the interesting/required data, and discard the rest. Even if the interesting data is not presented in the required format for processing by other services, it can be transformed into the desired format.
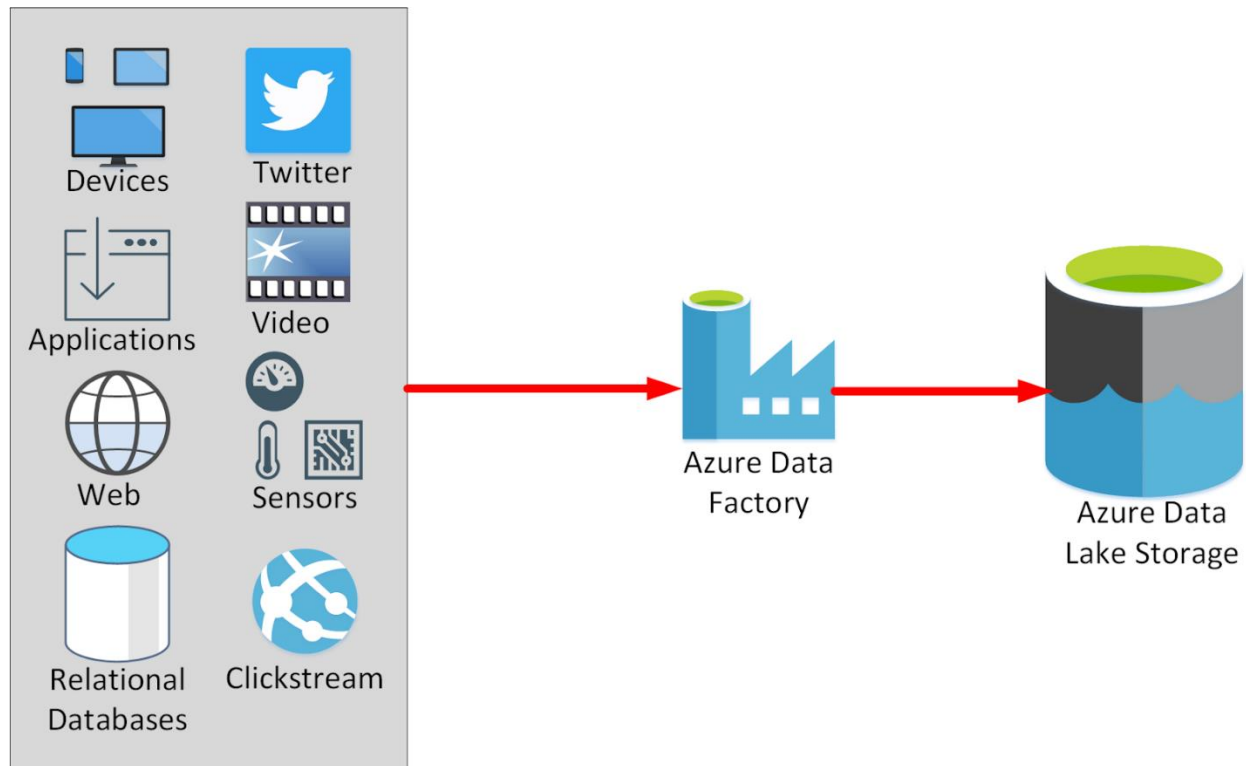
- **Option A is correct.** Azure Data Factory allows transforming the data from multiple sources in the desired uniform format.
- **Option B is incorrect.** Azure Data Lake Storage is a secure cloud platform that offers cost-effective and scalable storage for big data analytics. In an Azure Data Services data warehouse solution, data is generally loaded into Azure Data Lake Storage before it is processed into a structure that supports efficient analysis in Azure Synapse Analytics.
- **Option C is incorrect.** Azure Databricks is an Apache Spark environment that runs on Azure to offer big data processing, streaming, and machine learning.
- **Option D is incorrect**. Azure Data Factory allows users to transform the data from multiple sources in the desired uniform format.

Reference:
To know more about Azure Data Factory, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/learn/modules/examine-components-of-modern-data-warehouse/3-explore-azure-data-services-warehousing

Question

Domain: Describe an analytics workload on Azure

In which of the following case, you would recommend Azure Synapse Analytics? (You can choose multiple options.)

☐ A.

To perform complex queries and aggregations

☐ B.

To create a dashboard from the tabular data

☐ C.

To allow a large number of users to query analytics data

☐ D.

To quickly process very large amounts of data.

☐ E.

To retrieve data from multiple data sources

Explanation:

Correct Answers: A and D
Azure Synapse Analytics incorporates SQL technologies, Transact-SQL query capabilities, and open-source Spark tools to support the quick processing of very large amounts of data. SQL pools in Synapse Analytics are helpful in complex reporting. You can utilize the full power of Transact-SQL to run complex SQL statements or queries that summarize and aggregate the data.

- **Option A is correct**. Azure Synapse Analytics is optimized to perform compute-intensive tasks such as complex queries and aggregations.
- **Option B is incorrect**. Azure Analysis Services, not Azure Synapse, is used to create dashboards.
- **Option C is incorrect.** Azure Synapse Analytics is optimized to perform compute-intensive operations instead of the high level of query concurrency.
- **Option D is correct.** Azure Synapse Analytics incorporates SQL technologies, Transact-SQL query capabilities, and open-source Spark tools to support the quick processing of very large amounts of data.
- **Option D is incorrect.** Azure Data Factory, not Azure Synapse Analytics, is used to retrieve data from one or multiple data sources and convert it into the desired format.

Reference:
To know more about Azure Synapse Analytics, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/learn/modules/explore-data-storage-processing-azure/3-explore-azure-synapse-analytics

Question

Which of the following is the main difference between a data warehouse and a data lake?

○]A.

A data warehouse contains raw business data whereas a data lake holds structured information

○]B.

A data warehouse stores the dynamic data whereas a data lake stores the static data

○]C.

A data warehouse contains structured information whereas a data lake holds raw business data

○]D.

A data warehouse is highly accessible and easy to update whereas data lake is complicated while making changes

Explanation:

Correct Answer: C
A data lake is a repository for a huge amount of raw data. As the data is raw and unprocessed, it supports quick loading and updating but due to lack of structure, it does not support efficient analysis. Also, a data lake is highly accessible and it is easy to update the data lake. A data warehousing solution transforms the raw data of a data lake into meaningful business information in a data warehouse.

- **Option A is incorrect.** A data warehouse contains structured information, not the raw data, whereas a data lake holds the raw business data, not the structured information.
- **Option B is incorrect.** The data and information held by both the data warehouse and data lake will change over time with the ingestion and processing of new data.
- **Option C is correct.** A data warehouse contains structured information whereas a data lake holds raw business data. A data warehousing solution transforms the raw data of a data lake into meaningful business information in a data warehouse.
- **Option D is incorrect.** A data lake is highly accessible and easy to update whereas a data warehouse is complicated and it is costly to make changes in the data warehouse.

Reference:
To learn more about Azure Data Warehouse and Azure Data Lake, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/learn/modules/introduction-to-azure-data-lake-storage/
- https://docs.microsoft.com/en-us/azure/architecture/data-guide/relational-data/data-warehousing

Question

Domain: Describe an analytics workload on Azure

Which of the following statements is/are true for batch processing? (Select multiple options.)

☐ A.

It processes a large volume of data all at once

☐ B.

data is always processed in real-time

☐ C.

There is no or significantly low latency

☐ D.

there is an expected latency

Explanation:

Correct Answers: A and D
Batch data processing is an effective method of processing a huge amount of data where a set of transactions is collected over a time period. It involves collecting, entering, and processing data and eventually producing the batch results. In batch processing, latency is common.
- **Option A is correct.** Batch processing processes the large volume of data all at once.
- **Option B is incorrect.** It is real-time data processing that processes the data in real-time.
- **Option C is incorrect**. It is real-time processing where there is no or very low latency.
- **Option D is correct.** In batch processing, there are the expected latencies.

References:
To know more about Batch Data Processing, please refer to the following URLs:

- https://www.bmc.com/blogs/what-is-batch-processing-batch-processing-explained/
- https://www.7wdata.be/business-analytics/batch-vs-real-time-data-processing/

Question

Which of the following are the features of transactional data? (Select four options.)

☐ A.

Denormalized data

☐ B.

Tabular data

☐ C.

Multidimensional model

☐ D.

Appendable and updateable

☐ E.

Low integrity and consistency

☐ F.

Schema on write

☐ G.

Pessimistic or optimistic Locking strategy

Explanation:

Correct Answers: B, D, F, and G
Transactional data is data or information that tracks the interactions associated with the activities of an organization. These interactions are generally business transactions, such as payments done by the customers, payments done to suppliers, items moving through inventory, orders received, or services delivered. Transactional data has the following features:

- Highly normalized
- Supports user transactions
- Schema on write, strongly enforced
- Updatable and appendable
- Strong consistency and ACID guarantees
- Pessimistic or optimistic locking strategy
- High integrity and flexibility
- Optimized workload with heavy writes and moderate reads
- Primary and secondary indexing

- Small to medium-sized datum
- Relational model with tabular transactional data
- Small (MBs) to large (a few TBs) scaling
- **Option A is incorrect.** Normalization is one of the most common features of transactional data. It is OLAP where data is denormalized for improving the query performance when aggregation is to be done.
- **Option B is correct.** Transactional data is in tabular form.
- **Option C is incorrect**. Transactional data uses the relational model; not the multidimensional model.
- **Option D is correct.** Transactional data is appendable and updatable.
- **Option E is incorrect.** Transactional data has high integrity and consistency.
- **Option F is correct.** Schema on write, strongly enforced is one of the common features of transactional data.
- **Option G is correct.** In transactional data, the locking strategy is pessimistic or optimistic.

Reference:

To know more about transactional data, please refer to the following URL:

- https://docs.microsoft.com/en-us/azure/architecture/data-guide/relational-data/online-transaction-processing

Question

Domain: Describe an analytics workload on Azure

Select the right sequence for implementing the Extract, Load, and Transform (ELT) process.

○]A.

Extract the source data into text files, prepare the data for loading, load the data into staging tables with PolyBase or the COPY command, land the data into Azure Blob storage or Azure Data Lake Store, transform the data, insert the data into production tables.

○]B.

Extract the source data into text files, land the data into Azure Data Lake Store or Azure Blob storage, prepare the data for loading, load the data into staging tables with PolyBase or the COPY command, transform the data, insert the data into production tables.

○]C.

Extract the source data into text files, land the data into Azure Blob storage or Azure Data Lake Store, prepare the data for loading, load the data into staging tables with PolyBase or the COPY command, insert the data into production tables, transform the data.

○]D.

Extract the source data into text files, prepare the data for loading, load the data into staging tables with PolyBase or the COPY command, land the data into Azure Blob storage or Azure Data Lake Store, insert the data into production tables, transform the data.

Explanation:

Correct Answer: B
ELT is a process that extracts the data from a source system, loads it into a SQL pool, and then transforms it.

The basic steps for implementing the Extract, Load, and Transform (ELT) process are:

1. Extract the source data into text files.
2. Land the data into Azure Data Lake Store or Azure Blob storage.
3. Prepare the data for loading.
4. Load the data into staging tables through PolyBase or the COPY command.
5. Transform the data.
6. Insert the data into the production tables.

References:
To know more about the implementation of the Extract, Load, and Transform (ELT) process, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/design-elt-data-loading

Question

Domain: Describe an analytics workload on Azure

…is an open-source and parallel-processing framework that supports in-memory processing to improve the performance of big-data analysis applications.

○]A.

Apache Hadoop

○]B.

Apache Storm

○]C.

Apache Kafka

○]D.

Apache Spark

Explanation:

Correct Answer: D
HDInsight supports specific cluster types and cluster customization capabilities such as the capability to add components, languages, and utilities. The various cluster types offered by HDInsight are Apache Hadoop, Apache Spark, Apache HBase, ML Services, Apache Storm, Apache Interactive Query, and Apache Kafka. Out of these, Apache Spark is an open-source and parallel-processing framework that supports in-memory processing to improve the performance of big-data analysis applications.

- **Option A is incorrect.** Apache Hadoop is a framework that utilizes HDFS, a simple MapReduce programming model, and YARN resource management for processing and analyzing the batch data in parallel.
- **Option B is incorrect.** Apache Storm is a distributed and real-time computation system that processes large streams of data very quickly.
- **Option C is incorrect.** Apache Kafka is an open-source platform that is optimized to build streaming data pipelines and applications.
- **Option D is correct.** Apache Spark is an open-source and parallel-processing framework that supports in-memory processing to improve the performance of big-data analysis applications.

Reference:
To know more about the HDInsight, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-overview

Question

Domain: Describe an analytics workload on Azure

Which of the following statements is false?

○]A.

In a Spark pool, the nodes are replaced with a Spark cluster.

○]B.

The number of nodes is specified while creating the Spark cluster.

○]C.

A Spark pool doesn't have any auto scaling option.

○]D.

The Spark cluster splits the work out into a sequence of parallel tasks that could be performed concurrently.

Explanation:

Correct Answer: C
The statement "A Spark pool doesn't have any auto scaling option" is incorrect. Because Spark pools have an auto scaling option that can be enabled so that pools scale by removing or adding the nodes as per requirements. Autoscaling can occur while processing is active.

- **Option A is incorrect.** It is true that in a Spark pool, the nodes are replaced with a spark cluster.
- **Option B is incorrect**. In the Spark pool, the number of nodes is specified while creating the Spark cluster.
- **Option C is correct.** Spark pools have an auto scaling option that can be enabled so that pools scale by removing or adding the nodes as per requirements. Autoscaling can occur while processing is active.
- **Option D is incorrect**. The Spark cluster indeed splits the work out into a sequence of parallel tasks that could be performed concurrently.

Reference:
To know more about Spark pool, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/learn/modules/examine-components-of-modern-data-warehouse/3-explore-azure-data-services-warehousing

Question

Domain: Describe an analytics workload on Azure

In Power BI, _____ is a collection of preset and ready-made visuals and reports that are shared with the entire organization.

○]A.

An App

○]B.

The Canvas

○]C.

Scheduled Refresh

○]D.

Gateway Connection

○]E.

Parameters

Explanation:

Correct Answer: A
In Power Bi, an app is a collection of preset and ready-made visuals and reports that are shared with an entire organization. The app is as easy to use as ordering a fast-food value meal or microwaving a TV dinner where you just need to make a few comments or press a few buttons and you are provided with a set of entrees designed to go together and all presented in a tidy and readymade package.

- **Option A is correct.** It is an app in Power Bi that is a collection of preset and ready-made visuals and reports that are shared with the entire organization.
- **Option B is incorrect**. It is an app, not canvas, that is a collection of preset and ready-made visuals and reports. Instead, the *canvas* is the area in the center of the Power BI service that displays the available data sources in the Power BI service. This area is blank until visuals or reports are created.
- **Option C is incorrect**. Scheduled refresh is not a collection of preset and ready-made visuals and reports. Rather, it allows the user to automatically update the data for datasets.
- **Option D is incorrect**. Gateway Connection is not a collection of preset and ready-made visuals and reports. Rather, it allows the user to automatically update the data for datasets.
- **Option E is incorrect**. The parameter is not a collection of preset and ready-made visuals and reports. Rather, it allows the user to automatically update the data for datasets.

Reference:
To know more about Power BI, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/learn/modules/get-started-with-power-bi/4-exercise-touring-and-using-power-bi

Question

Domain: Describe an analytics workload on Azure

In Data Lake Analytics, which of the following languages is used to define a job that contains queries to transform data and extract insights?

○]A.

SQL

○]B.

C#

○]C.

U-SQL

○]D.

PL/SQL

Explanation:

Correct Answer: C
Data Lake Analytics provides a framework and collection of tools that is used to analyze data stored in Microsoft Azure Data Lake Store and other repositories. You write the jobs consisting of queries to transform the data and get insights. The language that is used to write these jobs is known as "U-SQL". U-SQL is a hybrid language that has the features of both languages SQL and C#, and offers the procedural and declarative capabilities that can be used for data processing.

- **Option A is incorrect.** It is U-SQL, not SQL, that is used to define the mentioned jobs in data lake analytics.
- **Option B is incorrect**. It is U-SQL, not #C, that is used to define the mentioned jobs in data lake analytics.
- **Option C is correct.** U-SQL is the language that is used to define a job that contains queries to transform data and extract insights into data lake analytics.
- **Option D is incorrect.** It is U-SQL, not PL/SQL, that is used to define the mentioned jobs in data lake analytics.

References:
To know more, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/learn/modules/explore-data-storage-processing-azure/2-describe-azure

Question

Domain: Describe an analytics workload on Azure

Which of the following statement(s) is/are true about Data Lake Store? (Select multiple options.)

☐ A.

Azure Data Lake Store offers a file system that is capable of storing only a limited amount of data.

☐ B.

Azure Data Lake Store is optimized for analytics workloads.

☐ C.

Azure Data Lake Store is not compatible with the Hadoop Distributed File System (HDFS).

☐ D.

Azure Data Lake Store offers granular security over data, through Access Control Lists (ACL).

Explanation:

Correct Answers: B and D
Data Lake Store offers a file system that is capable of storing near-limitless quantities of data. It supports holding massive amounts of raw data (blobs) and structured data. Data Lake Store is optimized for analytics workloads. It has compatibility with the Hadoop Distributed File System (HDFS). Hadoop jobs can be run through Azure HDInsight that can write and read data in the Data Lake Store efficiently. Azure Data Lake Store offers granular security over data, through Access Control Lists (ACL), which specifies which folders or files in the store can be accessed through which accounts.

- **Option A is incorrect**. Data Lake Store is capable of storing near-limitless quantities of data. You can hold huge amounts of raw data (blobs) and structured data.
- **Option B is correct**. Azure Data Lake Store is optimized for analytics workloads.
- **Option C is incorrect.** Azure Data lake Store is compatible with the Hadoop Distributed File System (HDFS). Hadoop jobs can be run through Azure HDInsight that can write and read data in the Data Lake Store efficiently.
- **Option D is correct.** Azure Data Lake Store offers granular security over data, through Access Control Lists (ACL), which specifies which folders or files in the store can be accessed through which accounts.

Reference:
To know more about Azure Data Lake Store, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/learn/modules/explore-data-storage-processing-azure/2-describe-azure

Question

Domain: Describe an analytics workload on Azure

Which of the following objects can be added to a Microsoft Power BI dashboard? (Select three options.)

☐ A.

Microsoft PowerPoint Slide

☐ B.

Report Page

☐ C.

Visualization from a Report

☐ D.

Textbox

☐ E.

Data Flow

Explanation:

Correct Answers: B, C and D
It is possible to add a report page, a visualization from a report, textbox, web content, image, video, and streaming data to a Microsoft Power BI Dashboard. You can't add the objects such as Microsoft PowerPoint slides or data flow to a Microsoft Power BI Dashboard.

- **Option A is incorrect.** It is not possible to add a Microsoft PowerPoint slide to a Microsoft PowerPoint Slide.
- **Option B is correct.** It is possible to add a report page to a Microsoft Power BI dashboard.
- **Option C is correct.** It is possible to add visualization from a report to a Microsoft Power Bi Dashboard.
- **Option D is correct.** A textbox can be easily added to a Microsoft Power BI dashboard.
- **Option E is incorrect.** It is not possible to add a data flow to a Microsoft Power BI dashboard.

References:
To know more about Microsoft Power BI Dashboard, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/power-bi/consumer/end-user-dashboards
- https://docs.microsoft.com/en-us/power-bi/create-reports/service-dashboard-add-widget

Question

Domain: Describe core data concepts

The mathematical terms **Relation** and **Tuple** in relational database respectively represent…

○]A.

Table and Domain

○]B.

Table and Row

○]C.

Table and column

○]D.

Key and Row

Explanation:

Correct Answer: B
In a relational database, data is stored in the form of tables. Entities are modeled as tables and each instance of an entity is represented by a row in the table and the column represents the attributes of that entity. In mathematical terms, a table is known as a Relation, and rows are known as Tuples.

- **Option A is incorrect.** In a relational database, a relation refers to the table whereas a tuple represents a row in the table. A domain is the set of permitted values for an attribute.
- **Option B is correct.** In a relational database, a Relation refers to the table whereas a Tuple refers to the row.
- **Option C is incorrect.** In a relational database, a relation refers to the table whereas a tuple represents a row in the table. A column represents the set of values of the same data type for various instances of an entity.
- **Option D is incorrect.** In a relational database, a relation refers to the table whereas a tuple represents a row in the table. There are many types of keys such as primary key, secondary key, alternate keys, and so on that have their purpose. A primary key is the most common key that uniquely identifies each row in the table.

References:
To know more about the characteristics of the Relational Database, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/learn/modules/describe-concepts-of-relational-data/2-explore-characteristics

Question

Domain: Describe core data concepts

Which of the following property ensures that the concurrent execution of multiple transactions leaves the database in the same state that would have been attained if the transactions were executed sequentially?

○]A.

Atomicity

○]B.

Consistency

○]C.

Isolation

○]D.

Durability

Explanation:

Correct Answer: C
A transactional database needs to adhere to the ACID (Atomicity, Consistency, Isolation, Durability) properties that ensure that the database remains consistent while processing transactions.

Out of these properties, isolation is the one that ensures the concurrent execution of multiple transactions leaves the database in the same state that would have been attained if the transactions were executed sequentially. A concurrent process can't see the data in an inconsistent state. For example, there is a transaction of transferring money from one account to another. The money has been deducted from one account but not yet credited to another.

- **Option A is incorrect.** *Atomicity* ensures that each transaction is considered as a single *unit*, which either succeeds or fails completely. If any statement of a transaction is failed to complete, the whole transaction is considered failed and the state of the database remains unchanged.
- **Option B is incorrect.** *Consistency* ensures that a transaction can only take the data in the database from one valid state to another. A consistent database should never create or lose data in such a way that can't be accounted for.
- **Option C is correct.** Isolation ensures that the concurrent execution of multiple transactions leaves the database in the same state that would have been attained if the transactions were executed sequentially.
- **Option D is incorrect**. Durability ensures that once a transaction has been committed, the changes made by that transaction will survive permanently even if there is a system failure such as a service outage or crash.

Reference:
To know more, please refer to the following Azure documentation:

Question

Domain: Describe core data concepts

Which of the following are the characteristics of a relational database?

☐ A.

All data needs to be stored as integers

☐ B.

The rows in a table consist of one or more columns that represent the properties of the entity.

☐ C.

The different rows in the same table might have different columns.

☐ D.

A tuple in a relation represents a single entity.

☐ E.

A relational database has a flexible schema.

Explanation:

Correct Answers: B and D
In a relational database, the entities are modeled as the tables. Entities are the things for which the information needs to be stored or known. For example, in an organization, you might be interested in storing the information regarding employees, department and project, and so on. All these employees, department and project are known as entity types, and tables are created for those. Each row in a table represents the one instance of that entity type. For example, in the EMPLOYEE table, each row represents the detail of one employee. Here is a snapshot of the database with table EMPLOYEE, DEPARTMENT, and PROJECT:

EMPLOYEE

Employee_ID	First_Name	Last_Name	Hire_Date	Manager	Department_ID
37	Frances	Newton	14-SEP-2005		
1234	Donald	Newton	24-SEP-2006	28	10
7895	Matthew	Michaels	16-MAY-2007	28	10

DEPARTMENT

Department_ID	Name
10	Accounting
20	Payroll

PROJECT

Project_ID	Project_Name
10	Online Market Research
20	Flight Booking

- **Option A is incorrect.** All the data don't need to be stored as integers. Data can be stored as any data type integer, real number, string, date, and so on.
- **Option B is correct.** In a relational database, the rows in a table consist of one or more columns that represent the properties of the entity.
- **Option C is incorrect.** In a table, all the rows need to have the same columns although the values for different rows might be the same or different depending upon the attribute.
- **Option D is correct.** A Row or Tuple in a table represents the one instance or entity of a particular entity type.
- **Option E is incorrect.** Schemaless design in the property of the non-relational database, not the relational database.

Reference:
To know more about the characteristics of a Relational Database, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/learn/modules/describe-concepts-of-relational-data/2-explore-characteristics

Question

Domain: Describe core data concepts

Which of the following is/are the examples of streaming dataset? (Select two options.)

☐ A.

Data from Twitter feeds

☐ B.

Sales data of a company for last month

☐ C.

Data from sensors and devices

☐ D.

List of employees employed in an organization

Explanation:

Correct Answers: A and C
A streaming dataset is the collection of data that is published continuously. In stream processing, each piece of new incoming data is processed when it arrives. Streaming handles the data in real-time.

- **Option A is correct.** Data from Twitter feeds is a streaming dataset as the data on Twitter is published continuously.
- **Option B is incorrect.** Sales data for the last month is not an example of a streaming dataset as this data is a point in time dataset.
- **Option C is correct.** Data from sensors and devices is an example of a streaming dataset as this data is published on a continuous basis.
- **D is incorrect**. In the company, there are a finite number of employees and the number of employees in the organization doesn't change very frequently such as minute to minute. Therefore, the list of employees employed in an organization is not an example of a streaming dataset.

Reference:
To know more about streaming datasets, please refer to the following Azure documentation:
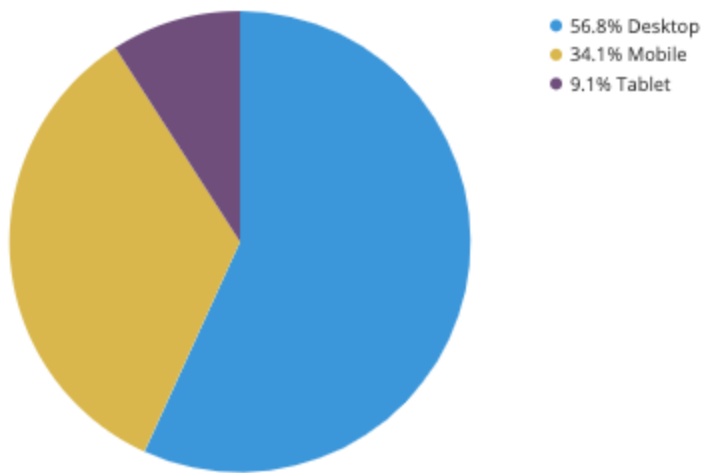
- https://docs.microsoft.com/en-us/learn/modules/explore-core-data-concepts/4-describe-difference

Question

Domain: Describe core data concepts

Choose the right type of the following graph.

Number of Visits by Device

- 56.8% Desktop
- 34.1% Mobile
- 9.1% Tablet

○]A.

Bar chart

○]B.

Pie chart

○]C.

Tree map

○]D.

Line Chart

○]E.

Scatter Chart

Explanation:

Correct Answer: B

A pie chart displays the division of the total amount among different categories as a circle divided into radial slices. A single slice in the circle represents a category and the size of the slice represents the proportion of that category out of the total.

- **Option A is incorrect.** A bar chart displays numeric values against different categories in a 2-axis plot. One axis lists the category levels and for each category, one

bar is plotted, and the length of the bar along the other axis represents the numeric value assigned to that category.
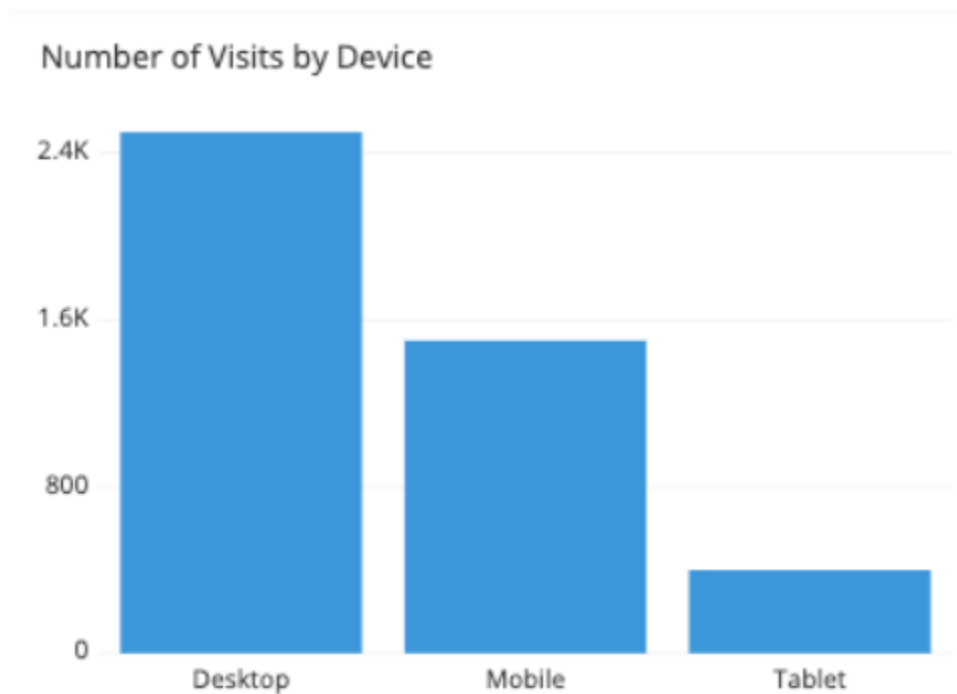
Number of Visits by Device

Figure: An example of bar chart

- **Option B is correct.** As in the preceding diagram, the categories and their share are denoted by the slices of the circle. The preceding diagram is an example of a pie chart.
- **Option C is incorrect.** Treemap is a chart of colored rectangles where size represents the relative value of each item. They can be hierarchical, with rectangles nested within the main rectangles.
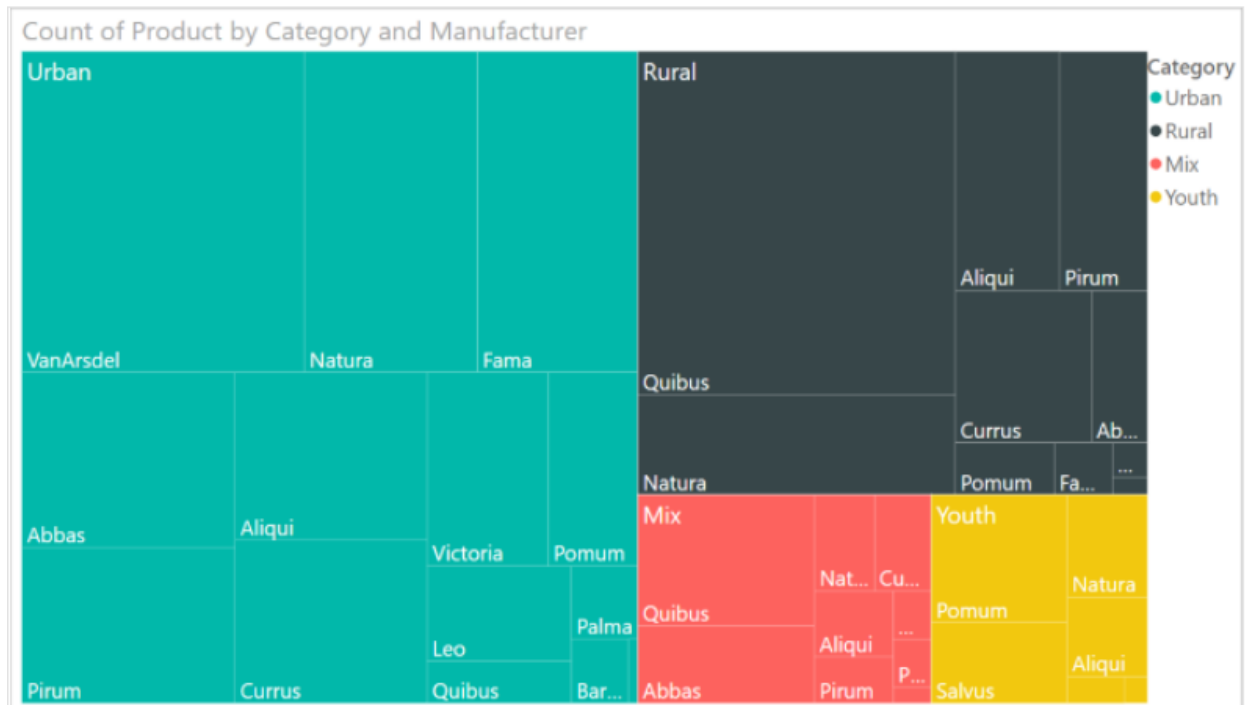
Figure: An example of Treemap

- **Option D is incorrect.** The line chart emphasizes the overall shape of an entire series of values, generally over time.
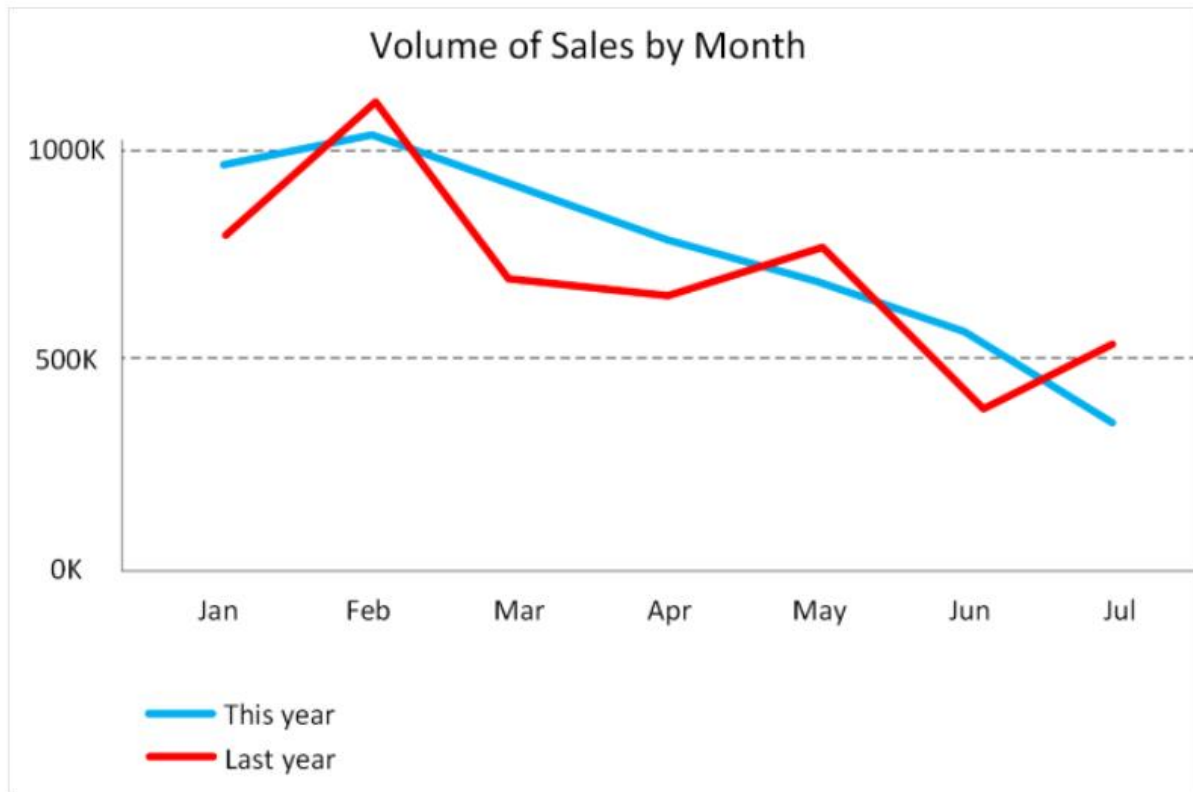

Figure: Example of line chart

- **Option E is incorrect.** A scatter chart displays the relationship between two numerical values.
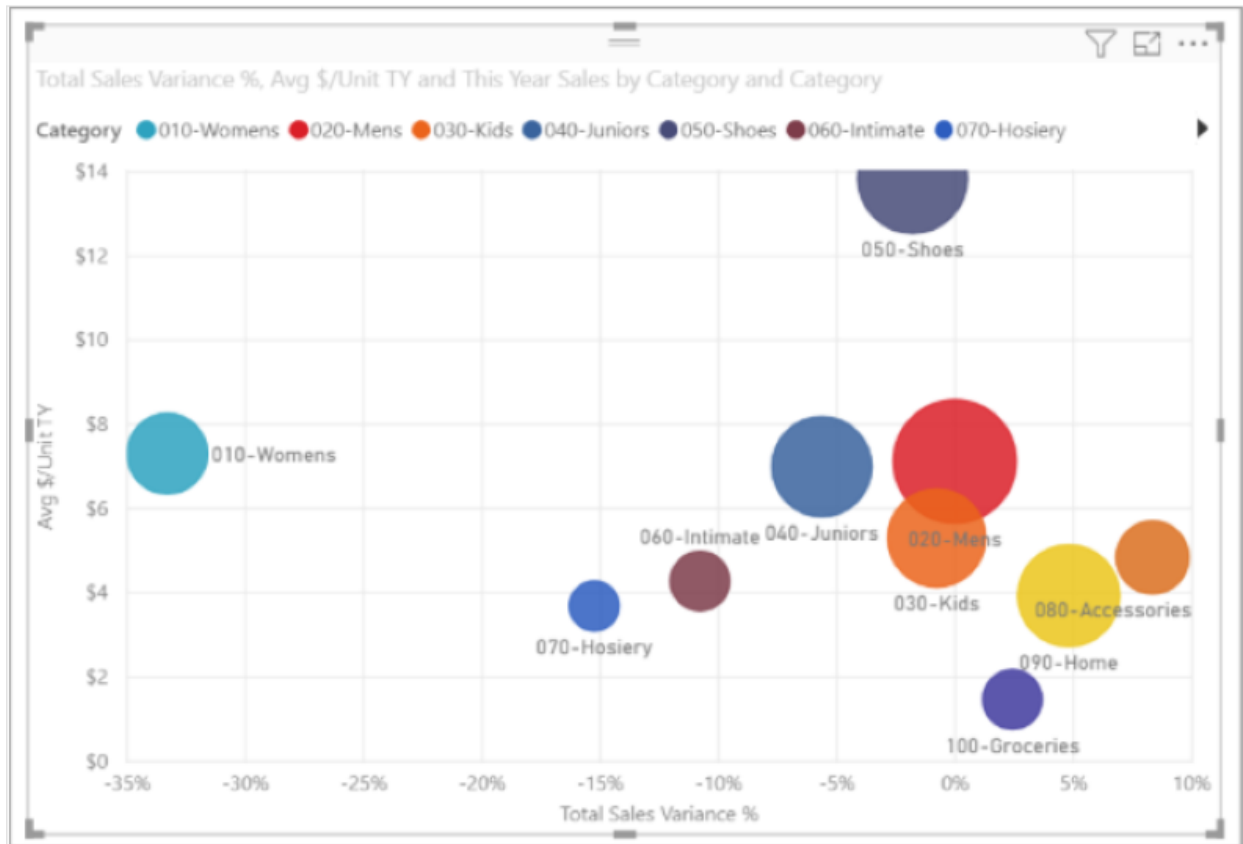

Figure: An example of scatter chart

References:

To know more, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/learn/modules/explore-concepts-of-data-analytics/3-explore-data-visualization
- https://chartio.com/learn/charts/how-to-choose-pie-chart-vs-bar-chart/

Question

Domain: Describe core data concepts

Which type of analytics is appropriate to tell what will happen in the future?

○]A.

Descriptive analytics

○]B.

Diagnostic analytics

○]C.

Predictive analytics

○]D.

Prescriptive analytics

Explanation:

Correct Answer: C
Predictive analytics helps in answering the questions regarding what will happen in the future. Here historical data is utilized to identify trends and determine if they might recur. Predictive analytical tools offer valuable insight into what may occur in the future.

- **Option A is incorrect**. Descriptive analytics helps in answering the questions regarding what has happened, depending upon the historical data.
- **Option B is incorrect.** Diagnostic analytics helps in answering questions regarding why things happened, not about what might happen in the future.
- **Option C is correct.** Predictive analytics helps in answering the questions regarding what will happen in the future.
- **Option D is incorrect.** Prescriptive analytics helps in answering the questions regarding what actions should be taken to reach a goal or target.

References:
To know more, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/learn/modules/explore-concepts-of-data-analytics/4-explore

Question

Domain: Describe core data concepts

In which pipeline does the transformation happen in the target datastore?

○]A.

ETL
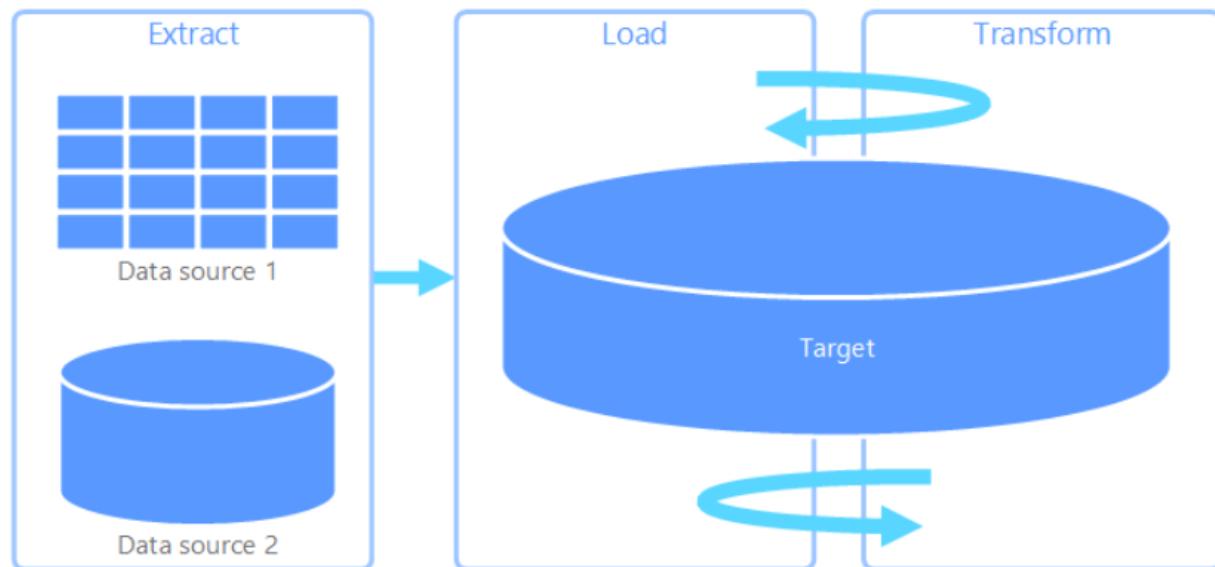
○]B.

ELT

○]C.

None of these

○]D.

None

Explanation:

Correct Answer: B
In the ELT (Extract, Load, and Transform) pipeline, the transformation happens in the target datastore. Rather than utilizing a separate transformation engine, the processing capabilities of the target datastore are utilized to transform data. However, ELT is efficient only when the target system is enough powerful to transform the data efficiently.

- **Option A is incorrect.** Extract, transform, and load (ETL) pipeline is utilized to collect data from multiple sources, transform the data as per business rules, and eventually load it into a destination data store. The transformation process in ETL occurs in a specialized engine, and typically involves utilizing staging tables to temporarily hold data as it is being transformed and eventually loaded to its destination. Therefore, in ETL, transformation does not happen in the target data store.
- **Option B is correct.** ELT (Extract, Load, and Transform) is a data pipeline where the transformation happens in the target datastore. Rather than utilizing a separate transformation engine, the processing capabilities of the target datastore are utilized to transform data.

Reference:
To know more about the ELT (Extract, Load, and Transform), please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/azure/architecture/data-guide/relational-data/etl

Question

_____ is the process of transforming and mapping the raw data into a more useful format for analysis.

○]A.

Data Cleaning

○]B.

Data Ingestion

○]C.

Wrangling

○]D.

Data Analysis

○]E.

Data Modeling

Explanation:

Correct Answer: C
Wrangling is the process of transforming and mapping raw data into a more useful format for analysis. It might include writing code for capturing, filtering, cleaning, combining, and aggregating data from multiple sources.

- **Option A is incorrect.** Data cleaning is a generalized term that comprises a wide range of actions such as removing or deleting anomalies and applying filters and transformations that would be too time-taking to run during the ingestion phase.
- **Option B is incorrect.** Data ingestion is the process that obtains and imports the data for immediate use or storage in a database.
- **Option C is correct.** Wrangling is the process of transforming and mapping raw data into a more useful format for analysis.
- **Option D is incorrect.** Data analysis is the process of inspecting, cleansing, transforming, and modeling data to get useful information, supporting decision-making, and informing conclusions.
- **Option E is incorrect.** Data modeling is a process of creating a conceptual representation of the data objects, in association with other data objects, and the rules between them.

Reference:
To know more about wrangling, please refer to the following Azure documentation:

-

Question

<div align="right">Domain: Describe core data concepts</div>

Choose the visual that depicts the major contributors to a selected value or result.

○]A.

Bar and Column chart

○]B.

Matrix Chart

○]C.

Treemap

○]D.

Key Influencers

Explanation:

Correct Answer: D
A key influencer chart depicts the major contributors to a selected value or result. Key influencers help a lot in understanding the factors that influence a key metric. For example, it is a good choice to know why sales were so high in last month or what influences the clients to place a 2nd order.

Figure: An example of a key influencer chart

- **Option A is incorrect.** Bar and column charts are a good choice to know how a set of variables varies across different categories. These are the key influencers, not bar or column charts, that depict the major contributors to a selected value or result.
- **Option B is incorrect.** A matrix is a tabular visual that summarizes data. It does not help in depicting the major contributors to a selected value or result.
- **Option C is incorrect**. A treemap is a chart of colored rectangles where size represents the relative value of each item. They can be hierarchical, with rectangles nested within the main rectangles.

- **Option D is correct. A** key influencer chart depicts the major contributors to a selected value or result. Key influencers help a lot to understand the factors that influence a key metric.

Reference:
To know more, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/learn/modules/explore-concepts-of-data-analytics/3-explore-data-visualization

Question

Domain: Describe core data concepts

Which of the following statement is true?

○]A.

Prescriptive Analytics helps in learning what might occur if circumstances change, and how to handle these situations while cognitive analytics help in answering the questions about what steps or actions should be taken to reach a goal or target.

○]B.

Cognitive Analytics helps in learning what might occur if circumstances change, and how to handle these situations while Prescriptive Analytics helps in answering the questions about what steps or actions should be taken to reach a goal or target.

○]C.

Predictive Analytics helps in answering the questions about what steps or actions should be taken to reach a goal or target while cognitive analytics helps in learning what might occur if circumstances change, and how to handle these situations.

○]D.

Cognitive Analytics helps in answering the questions about what steps or actions should be taken to reach a goal or target while Predictive Analytics helps in learning what might occur if circumstances change, and how to handle these situations.

Explanation:

Correct Answer: B
Data Analytics activities can be categorized as *descriptive*, *diagnostic*, *predictive*, *prescriptive*, and *cognitive* analytics. The various data analytics activities with their specific purposes can be understood from the following table:

Data Analytics Activity	Purpose
Descriptive Analytics	Helps answer questions about what has happened, based on historical data.
Diagnostic Analytics	Helps answer questions about why things happened.
Predictive Analytics	Helps answer questions about what will happen in the future.
Prescriptive Analytics	Helps answer questions about what actions should be taken to reach to a goal or target.
Cognitive Analytics	Helps to draw inferences from existing data and patterns and to learn what might occur if circumstances change, and how to handle these situations.

- **Option A is incorrect**. It is cognitive analytics, not prescriptive analytics, that helps in learning what might occur if circumstances change, and how to handle these situations while it is prescriptive analytics, not cognitive analytics, that help in answering the questions about what steps or actions should be taken to reach to a goal or target.
- **Option B is correct.** Cognitive analytics helps in learning what might occur if circumstances change, and how to handle these situations while Prescriptive analytics helps in answering the questions about what steps or actions should be taken to reach a goal or target.
- **Option C is incorrect.** It is prescriptive analytics, not predictive analytics, that helps in answering the questions about what steps or actions should be taken to reach to a goal or target while cognitive analytics helps in learning what might occur if circumstances change, and how to handle these situations.
- **Option D is incorrect.** It is prescriptive analytics, not cognitive analytics, that helps in answering the questions about what steps or actions should be taken to reach to a goal or target while cognitive analytics, not predictive analytics, helps in learning what might occur if circumstances change, and how to handle these situations.

References:
To know more about different types of analytics activities, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/learn/modules/explore-concepts-of-data-analytics/4-explore

Question

Domain: Describe core data concepts

Which of the following are the roles and responsibilities of a database administrator? (Choose three options.)

☐ A.

To install and upgrade the database server and application tools.

☐ B.

To enroll the users and maintain system security.

☐ C.

To develop, construct, test, and maintain the databases and data structures.

☐ D.

To monitor and optimize the performance of the database

☐ E.

To Prepare the data for prescriptive and predictive modeling

Explanation:

Correct Answers: A, B, and D
Database Administrator is tasked to manage and organize the databases. The main job of a database administrator is to ensure that data is available and protected from loss, theft, and corruption and is easily accessible whenever needed. The following figure shows the common roles and responsibilities of a database administrator:

- To install and upgrade the database server and application tools
- To allocate system storage and plan storage needs for the database system
- To modify the database structure, as necessary, from information provided by application developers
- To enroll the users and maintain system security
- To ensure compliance with the database vendor license agreement
- To control and monitor user access to the database
- To monitor and optimize the performance of the database
- To do Planning for backup and recovery of database information
- To maintain archived data
- To ensure Backing up and restoring databases
- To contact database vendor for technical support
- To generate various reports from the database using appropriate queries as per needs
- To manage and monitor data replication

- **Option A is correct**. To install and upgrade the database server and application tools is one of the common roles and responsibilities of a database administrator.
- **Option B is correct**. To enroll the users and maintain system security is one of the common roles and responsibilities of a database administrator.

- **Option C is incorrect.** To develop, construct, test, and maintain the databases and data structures is the responsibility of the data engineer, not the data administrator.
- **Option D is correct.** To monitor and optimize the performance of the database is one of the common roles and responsibilities of a database administrator.
- **Option E is incorrect.** To prepare the data for prescriptive and predictive modeling is the responsibility of the data engineer, not the data administrator.

References:

To know more about the roles and responsibilities of a data administrator, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/learn/modules/explore-roles-responsibilities-world-of-data/3-review-tasks-tools-for-database-administration
- https://docs.microsoft.com/en-us/learn/modules/explore-roles-responsibilities-world-of-data/4-review-tasks-tools-for-data-engineering

Question

<div align="right">Domain: Describe core data concepts</div>

Which of the following is not a data job role?

○]A.

Data Engineer

○]B.

Data Analyst

○]C.

System Administrator

○]D.

Database Administrator

Explanation:

Correct Answer: C

Systems administrators are the ones who deal with infrastructure components such as networks, virtual machines, and other physical devices in a data center.

- **Option A is incorrect.** Data Engineer is the one who works with data, apply data filtering and cleaning algorithms, identify business rules, and convert the data into useful information.

- **Option B is incorrect.** The main role of a data analyst is exploring and analyzing the data to create charts and other visualizations that help an organization to make informed decisions.
- **Option C is correct**. A system administrator is the one who deals with infrastructure components, not the data.
- **Option D is incorrect**. The Database Administrator is responsible to manage and organize the databases. The main job of a database administrator is to ensure that data is available and protected from loss, theft, and corruption and is easily accessible whenever needed.

Reference:
To know more about data job roles, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/learn/modules/explore-roles-responsibilities-world-of-data/2-explore-job-roles

Question

Domain: Describe how to work with non-relational data on Azure

How many performance tiers are offered by Azure File Storage?

○]A.

1

○]B.

2

○]C.

3

○]D.

4

Explanation:

Correct Answer: B
Standard tier and Premium tier are two performance tiers offered by Azure File Storage. These tiers vary in terms of throughput and cost. The *Standard* tier utilizes hard disk-based hardware in a data center while the *Premium* tier utilizes solid-state disks. The *Premium* tier provides higher throughput than the standard tier but the cost of a premium tier is also higher than the standard one.
- **Option A is incorrect.** Azure File Storage offers two performance tiers, not one.
- **Option B is correct**. Standard tier and Premium tier are two performance tiers offered by Azure File Storage.

- **Option C is incorrect.** Azure File Storage offers two performance tiers—standard tier and Premium tier, not three.
- **Option D is incorrect.** Azure File Storage offers two performance tiers—standard tier and Premium tier, not four.

Reference:

To know more about Azure File Storage, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/learn/modules/explore-non-relational-data-offerings-azure/4-explore-azure-file-storage

Question

Domain: Describe how to work with non-relational data on Azure

Name the elements of an Azure Table storage key.

○]A.

Row key and column key

○]B.

Partition key and Row key

○]C.

Column name and Table name

○]D.

Column name and Column key

Explanation:

Correct Answer: B

Two elements of an Azure Table Storage key are the partition key and the Row key. The partition key is the key that identifies the partition having the row and a row key is unique for each row in a given partition. In a partition, the items are stored in row key order.

- **Option A is incorrect.** Columns in an Azure table storage table are anonymous and have no name or key. Partition key and Row key are two elements of an Azure Table Storage Key.
- **Option B is correct**. Partition key and Row key are two elements of an Azure Table Storage Key where the partition key identifies a specific partition and the row key identifies the row in a given partition.

- **Option C is incorrect**. Columns in an Azure table storage table are anonymous and have no name or key. The partition key and row key are the elements of an Azure Table storage key.
- **Option D is incorrect**. Columns in an Azure table storage table are anonymous and have no name or key.

Reference:

To know more about Azure Table storage, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/learn/modules/explore-non-relational-data-offerings-azure/2-explore-azure-table-storage

Question

Domain: Describe how to work with non-relational data on Azure

Which of the following are the advantages of using Azure Table Storage tables over other methods of data storage? (Select 3 Options)

☐ A.

It ensures consistency as transactional updates across multiple entities are guaranteed

☐ B.

There is no requirement of mapping and maintaining the complex relationships typically required by a normalized relational database

☐ C.

Referential integrity is the major benefit of Azure Table Storage tables that maintains the relationships between rows

☐ D.

Tables can store semi-structured data

☐ E.

It is simpler to scale. It consumes the same time to insert data in an empty table or a table with billions of entries. An Azure storage account can consist of up to 500 TB of data

Explanation:

Correct Answers: B, D, and E

The tables in Azure Table Storage are schemaless. Data can be easily adapted as per the need of the application. Tables can hold flexible datasets such as user data for web applications, device

information, address books, or other kinds of metadata your service needs. The important thing is to select the partition and row keys carefully.

The main advantages of using Azure Table Storage tables over other methods of data storage are as follows:

- It is simpler to scale. It consumes the same time to insert data in an empty table or a table with billions of entries. An Azure storage account can consist of up to 500 TB of data.
- Tables can store semi-structured data.
- There is no requirement of mapping and maintaining the complex relationships typically required by a normalized relational database.
- Row insertion is easy and fast.
- Data retrieval is also fast if the row and partition keys are specified as query criteria

The main disadvantages of using Azure Table Storage tables are as follows:

- Consistency might be an issue as it does not guarantee transactional updates across multiple entities.
- There is no concept of referential integrity. The relationships between rows are maintained externally to the table
- Filtering and sorting on non-key data is quite difficult. Queries, where non-key fields are searching criteria, might result in full table scans.

- **Option A is incorrect**. Consistency is not the benefit, rather an issue in Azure table storage tables as it does not guarantee transactional updates across multiple entities.
- **Option B is correct.** In Azure table storage tables, there is no requirement of mapping and maintaining the complex relationships typically required by a normalized relational database.
- **Option C is incorrect**. In Azure table storage tables, there is no concept of referential integrity. The relationships between rows are maintained externally to the table.
- **Option D is correct**. Azure table storage tables can hold semi-structured data which is also an advantage of Azure table storage tables.
- **Option E is correct**. Scaling is easy in Azure table storage tables. It consumes the same time to insert data in an empty table or a table with billions of entries. An Azure storage account can consist of up to 500 TB of data.

Reference:
To know more about Azure Table storage, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/learn/modules/explore-non-relational-data-offerings-azure/2-explore-azure-table-storage

Question

Domain: Describe how to work with non-relational data on Azure

Choose the correct statement(s).

☐ A.

In an on-Premises solution, Application level controls and Network Controls are the responsibility of cloud customers and physical security is the responsibility of cloud providers.

☐ B.

In SaaS solutions, Application level controls and Network Controls are the responsibility of cloud customers and physical security is the responsibility of cloud providers.

☐ C.

In SaaS solutions, Application level controls, Network Controls, and physical security are the responsibilities of the cloud provider.

☐ D.

In IaaS solution, Application level controls are the responsibility of cloud customers while it is the responsibility of the cloud provider in the SaaS solution.

Explanation:

Correct Answers: C and D
Data security is a shared responsibility between cloud customers and database providers. Depending upon the chosen database provider, the amount of responsibility varies. The following diagram depicts the responsibilities of cloud customers and cloud providers in different types of solutions.

Responsibility	On-Prem	IaaS	PaaS	SaaS
Data classification & accountability	Cloud Customer	Cloud Customer	Cloud Customer	Cloud Customer
Client & end-point protection	Cloud Customer	Cloud Customer	Cloud Customer	Shared
Identity & access management	Cloud Customer	Cloud Customer	Shared	Shared
Application level controls	Cloud Customer	Cloud Customer	Shared	Cloud Provider
Network controls	Cloud Customer	Shared	Cloud Provider	Cloud Provider
Host infrastructure	Cloud Customer	Shared	Cloud Provider	Cloud Provider
Physical security	Cloud Customer	Cloud Provider	Cloud Provider	Cloud Provider

Cloud Customer ▣   Cloud Provider ▣

- **Option A is incorrect.** As shown in the preceding diagram, in the case of an on-premises solution, everything from end-point protection to physical security is the responsibility of cloud customers.
- **Option B is incorrect**. As shown in the preceding diagram, in a SaaS solution, application level controls, network controls, and physical security is the responsibility of the cloud provider.
- **Option C is correct.** As shown in the preceding diagram, in a SaaS solution, application level controls, network controls, and physical security are the responsibilities of the cloud provider.
- **Option D is correct.** As shown in the preceding diagram, in the IaaS solution, application level controls are the responsibility of cloud customers while it is the responsibility of cloud providers in SaaS solutions.

Reference:

To know more, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/azure/cosmos-db/database-security

Question

Domain: Describe how to work with non-relational data on Azure

Which of the following statements is false about Azure Cosmos DB?

⊙]A.

It supports providing or restricting access to the Cosmos account, database, container, and offers (throughput) using Access control (IAM) in the Azure portal.

⊙]B.

It supports monitoring the account for normal and abnormal activities.

⊙]C.

All connections to Azure Cosmos DB support HTTPS.

⊙]D.

The data once deleted can't be recovered in any way.

Explanation:

Correct Answer: D
Azure Cosmos DB secures the database through the following:

- Network security and firewall settings
- User authentication and fine-grained user controls
- Replicating data globally for regional failures
- Ability to failover from one data center to another
- Local data replication within a data center
- Automatic data backups
- restoring the deleted data from backups
- isolating and protecting the sensitive data
- Responding to attacks
- Monitoring for attacks
- Geo-fencing data to adhere to data governance restrictions
- Physical protection of servers in protected data centers
- Certifications

- **Option A is incorrect.** In Azure Cosmos DB, you can provide or restrict access to the Cosmos account, container, database, and offers (throughput) through Access control (IAM) in the Azure portal. IAM offers role-based access control and integrates with Active Directory. You can utilize built-in or even the custom roles for groups as well as individuals.

- **Option B is incorrect.** Azure Cosmos DB allows monitoring the account for normal and abnormal activities using audit logging and activity logs. You can see what operations have been carried out on your resources, who started the operation, when the operation happened and the status of the operation, and so on.
- **Option C is incorrect**. In Azure Cosmos DB, all connections to Azure Cosmos DB support HTTPS.
- **Option D is correct.** It is false that the data once deleted can't be recovered in Azure Cosmos DB. It supports recovering data you might have accidentally deleted up to ~30 days after the event through automated online backups.

Reference:
To know more about Azure Cosmos DB, please refer to the following Azure documentation:

- https://docs.microsoft.com/en-us/azure/cosmos-db/database-security