# Practice Test 4

### **Question 1**

Domain :Implement data storage solutions

A social media start-up is planning to set up Azure - Cosmos DB account. Which consistency level would be the best choice for them if they want High Availability, High Throughput and Low Latency?

]A.

**Strong Consistency**

]B.

**Bounded Staleness**

]C.

**Session Consistency**

]D.

**Consistent Prefix**

]E.

**Eventual Consistency**

**Explanation:**

**Answer: E**

**Option A is incorrect.**Because in a strong consistency level, all the data will be synchronously replicated with every replica in the database. Thus, it will not be possible to achieve high availability, high Throughput and Low Latency performance.

**Option B is incorrect.**Because in bounded Staleness, data is replicated asynchronously with the predefined stillness window defined by several records or time period that means high availability, high throughput and low latency are not possible.

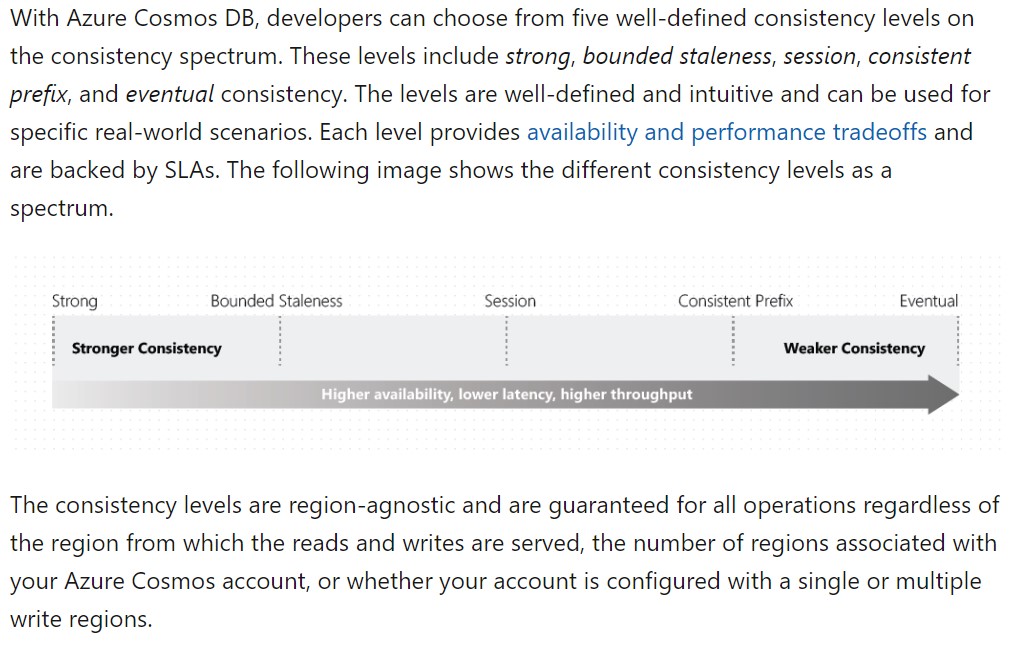
**Option C is incorrect.**Because session has a medium consistency level which uses a session token internally with each request to ensure that the consistency level means high availability, high throughput and low latency is not possible.

**Option D is incorrect.**Because consistent Prefix ensures that updates contain some prefix of all the updates, with no gaps has high availability, high throughput and but latency is higher than eventual consistent.

**Option E is CORRECT.**Because eventual consistency is the best fit because it has the highest availability, lowest latency and highest throughput. Eventual consistency is ideal where the application does not require any ordering guarantees examples of this include things like counting posts or likes or non-threaded comments in a social media post.

The Microsoft documentation mentions the following.

**Diagram:**



**Reference:**For more information on Cosmos DB (Consistency Levels), please refer to the below document link from Azure-

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

### **Question 2**

Domain :Monitor and optimize data solutions

While estimating RUs (Request Units ) per second to provision. Which of the following is not true about estimation option:

]A.

**As the size of an item increases, the number of RUs consumed to read or write the item also increases.**

]B.

**Assume that the default indexing is applied on all properties, the number of Request units(RUs) consumed to write an item decreases as the item property count increases.**

]C.

**Point reads cost significantly fewer request units(RUs) than with queries.**

]D.

**The exact same query on the same data will always cost the same number of RUs even on repeated executions.**

**Explanation:**

**Answer :** **B**

**Option A is incorrect because** The number of RUs consumed to read or write the item is directly proportional to size of the item.

**Option B is CORRECT because** the number of Request units consumed to write an item will always increase as the item property count increases.

**Option C is incorrect because** Queries cost is significantly higher than the Point reads as the latter is a more light weight process.

**Option D is incorrect because** query once executed on the same data on repeated iterations will always cost the same number of Request Units

**Reference :**For more information on Request Units in Azure Cosmos DB. please refer to the below document link from Azure:

* <https://docs.microsoft.com/en-us/azure/cosmos-db/request-units>
* <https://docs.microsoft.com/en-us/azure/cosmos-db/optimize-cost-queries#evaluate-request-unit-charge-for-a-query>

### **Question 3**

Domain :Implement data storage solutions

You are working as a Consultant to a South Korean Insurance company. They want to migrate their On-Premise Cassandra database to cloud. They have finalized Azure’s Cosmos DB offering as it provides zero code change option for their existing applications. But they are now worried about the security of data. They have asked you to check if encryption of data is possible within Azure Cosmos DB. What would be your answer to this?

]A.

**custom-managed Key**

]B.

**Service-managed key**

]C.

**A & B**

]D.

**None of the above**

**Explanation:**

**Answer: C**

**Option A is incorrect.**Because Azure Cosmos DB account provides both service-managed key & custom-managed keys for encryption of data. Custom-managed key provides an additional layer of encryption on top of default service -managed layer.

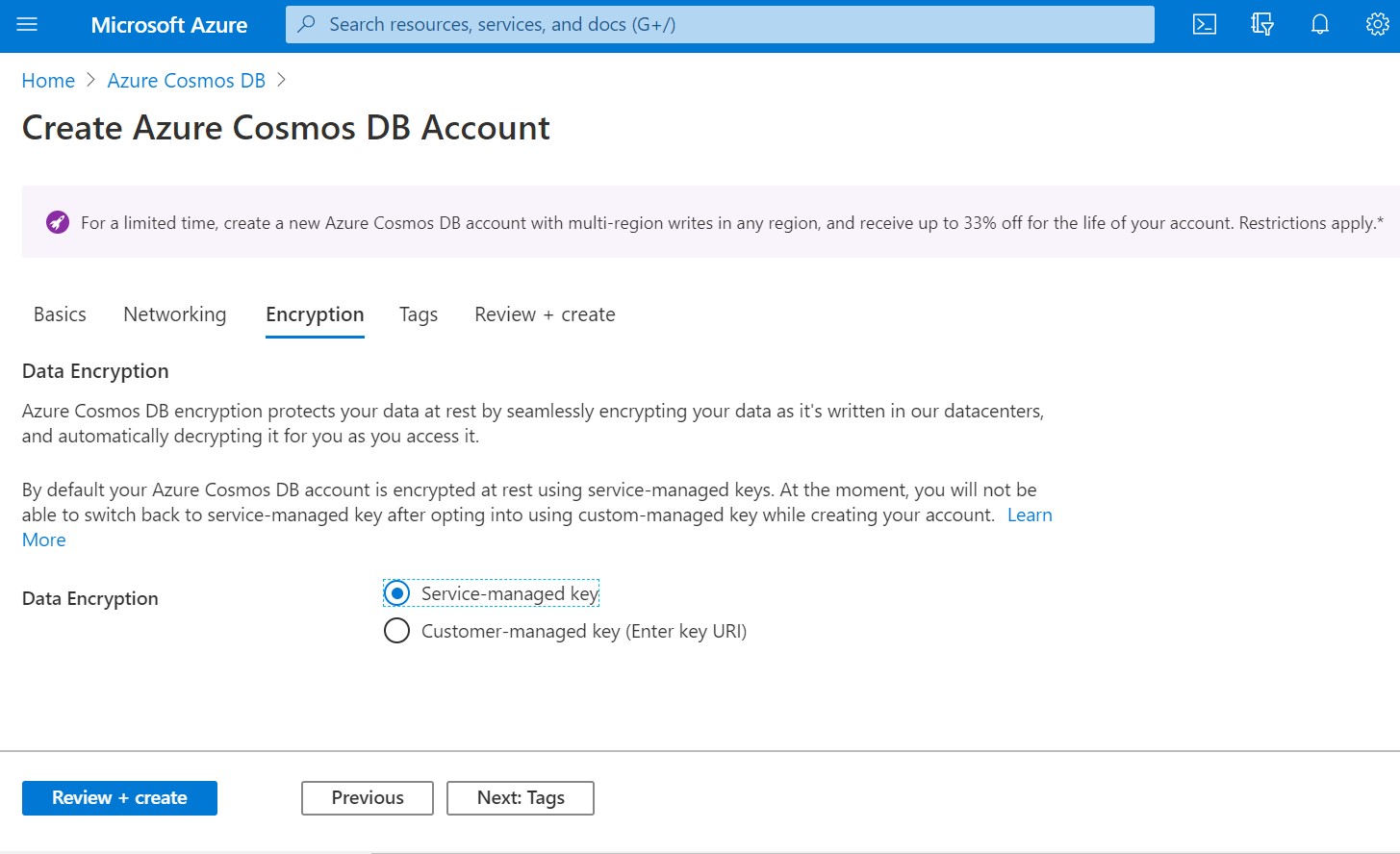
**Option B is incorrect.**Because Azure Cosmos DB account provides both service-managed key & custom-managed keys for encryption of data.

Azure Cosmos DB is encrypted at rest using service-managed keys by default.

**Option C is CORRECT.**Because Azure Cosmos DB account provides both service-managed key & custom-managed keys for encryption of data.

Azure Cosmos DB is encrypted at rest using service-managed keys by default and custom-managed key provides an additional layer of encryption on top of default service-managed layer.

**Diagram:**



**Option D is incorrect.**Because Azure Cosmos DB account provides both service-managed key & custom-managed keys for encryption of data.

**Reference:**For more information on Cosmos DB (Security - Encryption), please refer to the below document link from Azure-

* <https://docs.microsoft.com/en-us/azure/storage/common/storage-introduction>
* <https://devblogs.microsoft.com/cosmosdb/customer-managed-keys-encryption/>

### **Question 4**

Domain :Monitor and optimize data solutions

You are working as a Cloud engineer with a data analytics company, which uses Azure Cosmos DB. It's been observed that the query cost is increasing rapidly because one of the databases which is frequently used for read queries by the team. Which one of the following is the correct way to optimize and reduce the Query Cost?

]A.

**Queries should have an equality or range filter clause on any property**

]B.

**Query should have a filter clause within a single partition key**

]C.

**Implement indexing which will reduce the query cost**

]D.

**Optimize cost with time to live and change feed**

**Explanation:**

**Answer: B**

**Option A is incorrect because** keeping an equality or range filter clause on any property will definitely increase the query cost**.**

**Option B is CORRECT because** having a filter clause in a query with a single partition key is one of the best ways to optimize and reduce query cost.

**Option C is incorrect because**implementing indexing in Azure CosmosDB is a better way to reduce storage cost. But, this will not help in reducing query cost.

**Option D is incorrect because**optimizing cost with time to live and change feed is also a method to optimize storage cost. This will also have no effect on reducing query cost.

**Reference -**For more information on Optimizing Query cost with Azure Cosmos DB. please refer to the below document link from Azure:

* <https://docs.microsoft.com/en-us/azure/cosmos-db/optimize-cost-queries>
* <https://docs.microsoft.com/en-us/azure/cosmos-db/plan-manage-costs>
* <https://docs.microsoft.com/en-us/azure/cosmos-db/optimize-cost-storage>

### **Question 5**

Domain :Manage and develop data processing

A junior developer wants to add a container in Azure Cosmos DB account for SQL API, which all details are mandatory to create new containers as per your knowledge.

]A.

**Database id, Container id, Partition key**

]B.

**Database id, Container id, Throughput**

]C.

**Container id, Partition key**

]D.

**Database id, Container id, Partition key, Throughput**

**Explanation:**

**Answer:** **D**

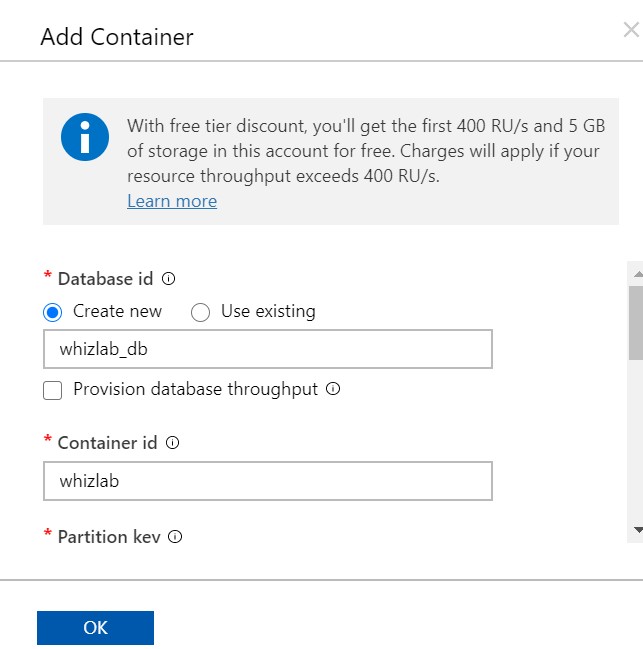
**Option A is incorrect because** to create or add a container, it also needs throughput to configure.

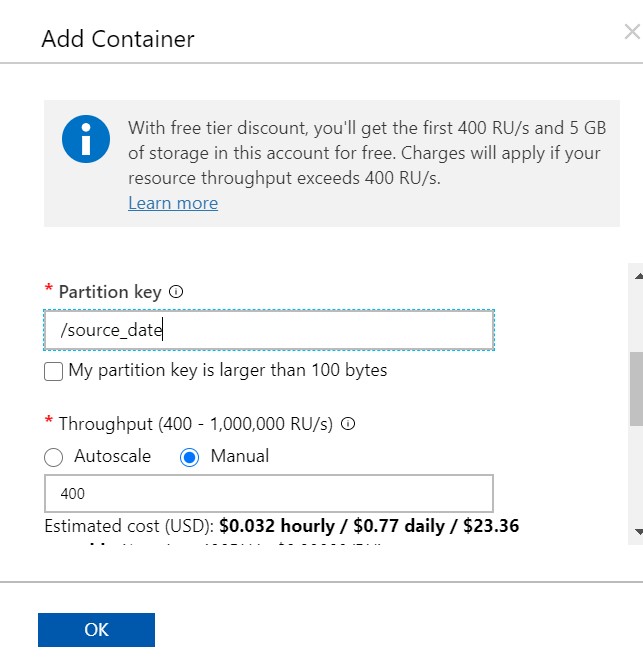
**Option B is incorrect because** to create or add a container, it also needs Partition Key.

**Option C is incorrect because** to create or add container, it also needs Database id & Throughput

**Option D is CORRECT because** to create or add a container for SQL API , these are the mandatory fields - Database id, Container id, Partition key, Throughput.

**Diagram:**





**Reference:**For more information on Cosmos DB (how to create container) please refer to the below document link from Azure :

* <https://docs.microsoft.com/en-us/azure/cosmos-db/how-to-create-container>

### **Question 6**

Domain :Monitor and optimize data solutions

Real Estate company storing enormous amounts of data on Azure Cosmos DB wants to predict the cost of properties worldwide. They observed some sluggishness while querying their data. They are not able to find the issue as directed logs are not enabled on Azure Monitor Log Analytics. They have Professional Direct subscription, so they have decided to directly call the Azure Support team regarding the same.  As a part of Azure Support team you need to quickly identify the issue, there can be possibility of incorrect partition, which option fits best to check partition issue?

]A.

**Log Analytics**

]B.

**Alert**

]C.

**Metrics**

]D.

**None of the above**

**Explanation:**

**Answer: C**

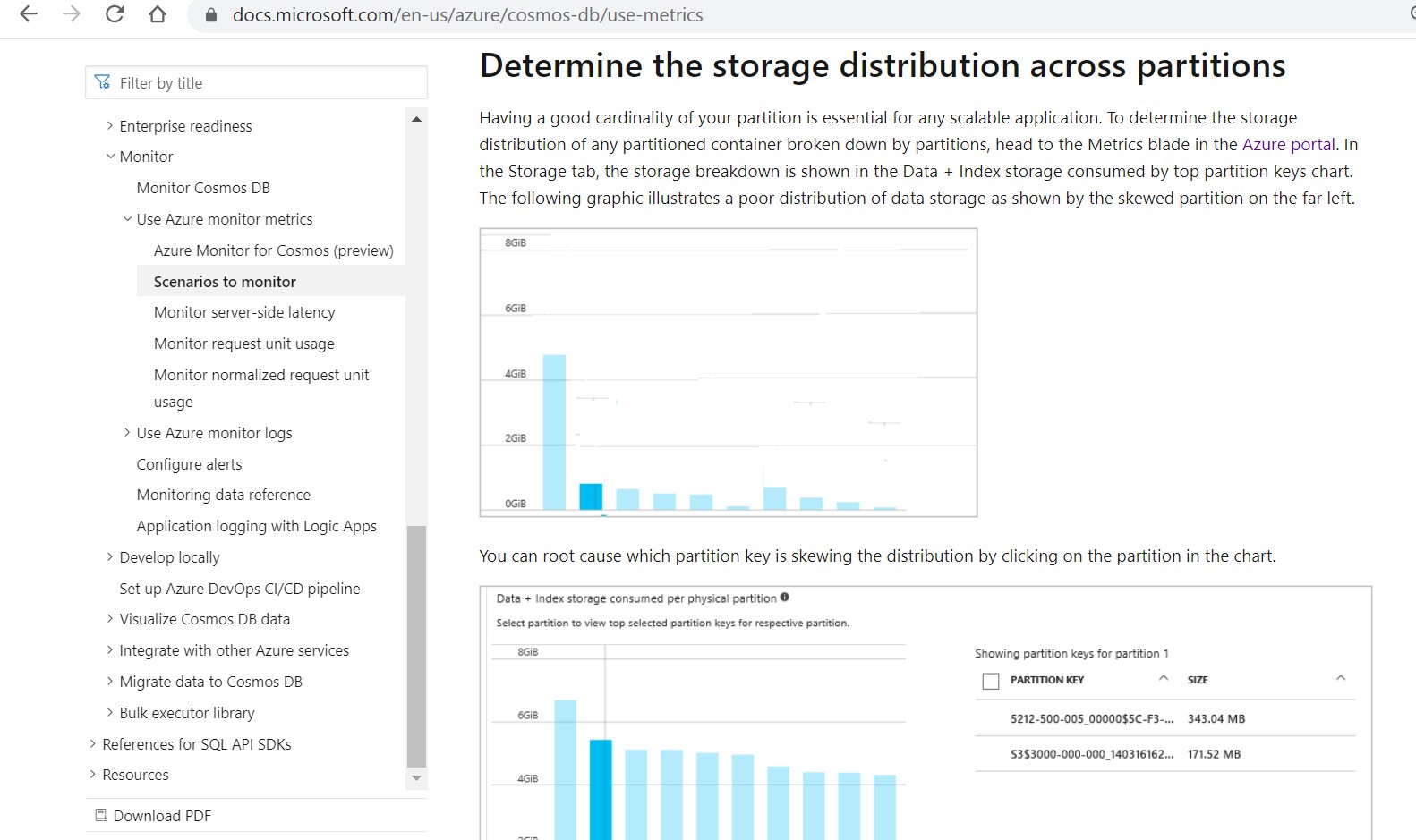
**Option A is incorrect because** they have not directed any logs to Azure Monitor Log Analytics. Log Analytics cannot be used.

**Option B is incorrect because** Alert needs to configure, and this can be used to efficiently monitor Azure resources.

**Option C is CORRECT because** Metrics here will be the best option to quickly identify partition issues.

 In Storage metrics inside Data and Index storage consumed, you can see Top Partition metrics and identify which partition is causing issues.

**Diagram:**



**Option D is incorrect because**using metrics (Storage) we can find the root cause which partition key is skewing the distribution.

**Reference :**For more information on Cosmos DB (Storage distribution across partitions) please refer to the below document link from Azure:

* <https://docs.microsoft.com/en-us/azure/cosmos-db/use-metrics>

Microsoft Azure official YouTube link -

* <https://www.youtube.com/watch?v=UZiccUhvWKE>

### **Question 7**

Domain :Implement data storage solutions

A company wants to store and access geospatial data. Which of the following API could be supported by Azure Cosmos DB?

]A.

**Azure Cosmos DB SQL API**

]B.

**Azure Cosmos DB MongoDB API**

]C.

**Azure Cosmos DB Graph API**

]D.

**Azure Cosmos DB Table API**

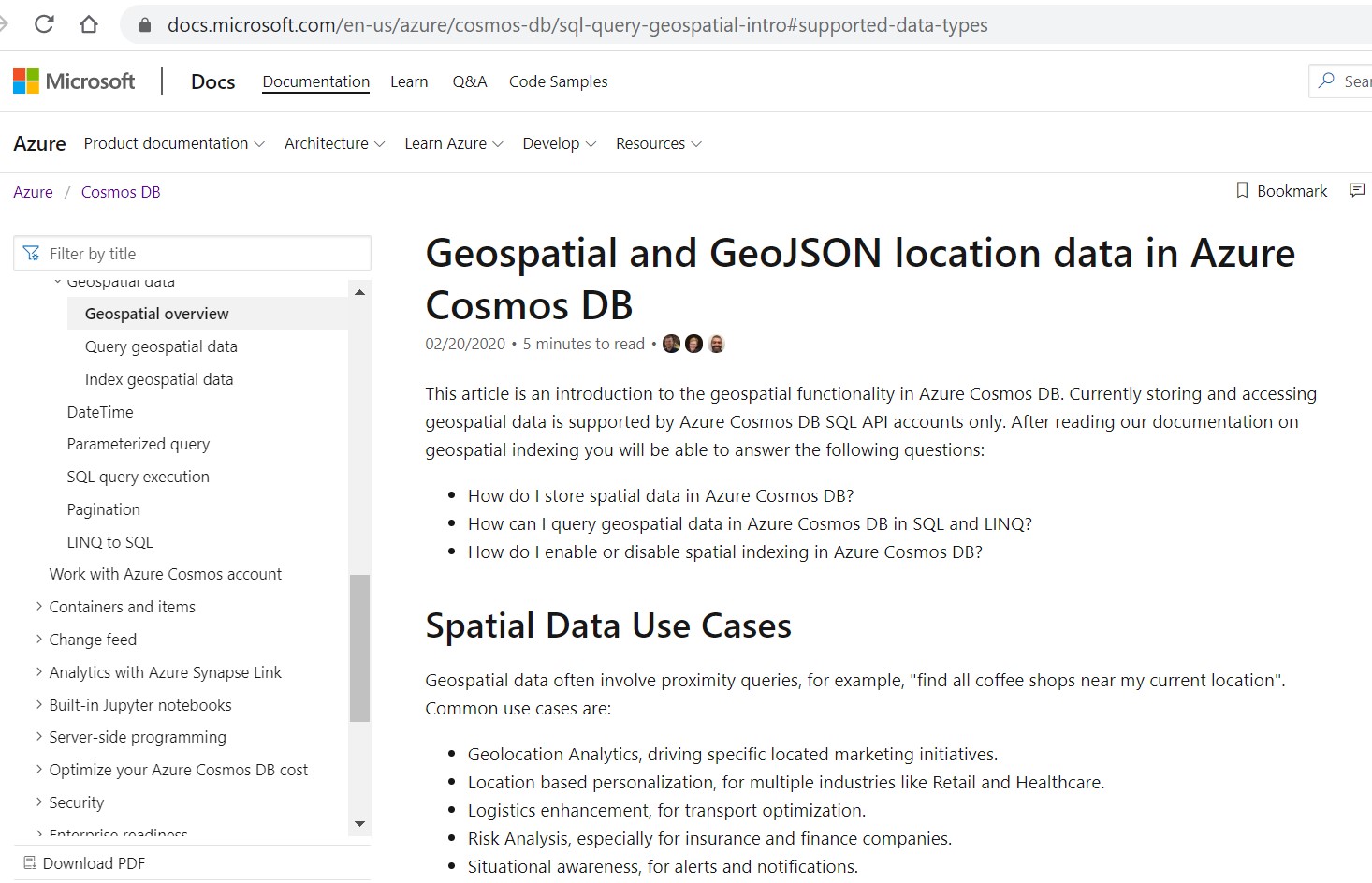
**Explanation:**

**Answer: A**

**Option A is CORRECT.**Because currently storing and accessing geospatial data is supported by Azure Cosmos DB SQL API accounts only.

The Microsoft documentation mentions the following.

**Diagram:**



**Option B is incorrect.**Because geospatial is not supported for MongoDB API.

**Option C is incorrect.**Because geospatial is not supported for Graph API.

**Option D is incorrect.**Because geospatial is not supported for Table API.

**Reference:**For more information on Cosmos DB (geospatial support) please refer to the below document link from Azure.

* <https://docs.microsoft.com/en-us/azure/cosmos-db/sql-query-geospatial-intro#supported-data-types>
* [https://devblogs.microsoft.com/cosmosdb/spatial-geometry-data](https://devblogs.microsoft.com/cosmosdb/spatial-geometry-data/)[/](https://devblogs.microsoft.com/cosmosdb/spatial-geometry-data/)

### **Question 8**

Domain :Implement data storage solutions

A company wants to deploy Azure Cosmos DB. They want to ensure that all databases currently supported by this service.

A.

**Cassandra**

B.

**MongoDB**

C.

**Hbase**

D.

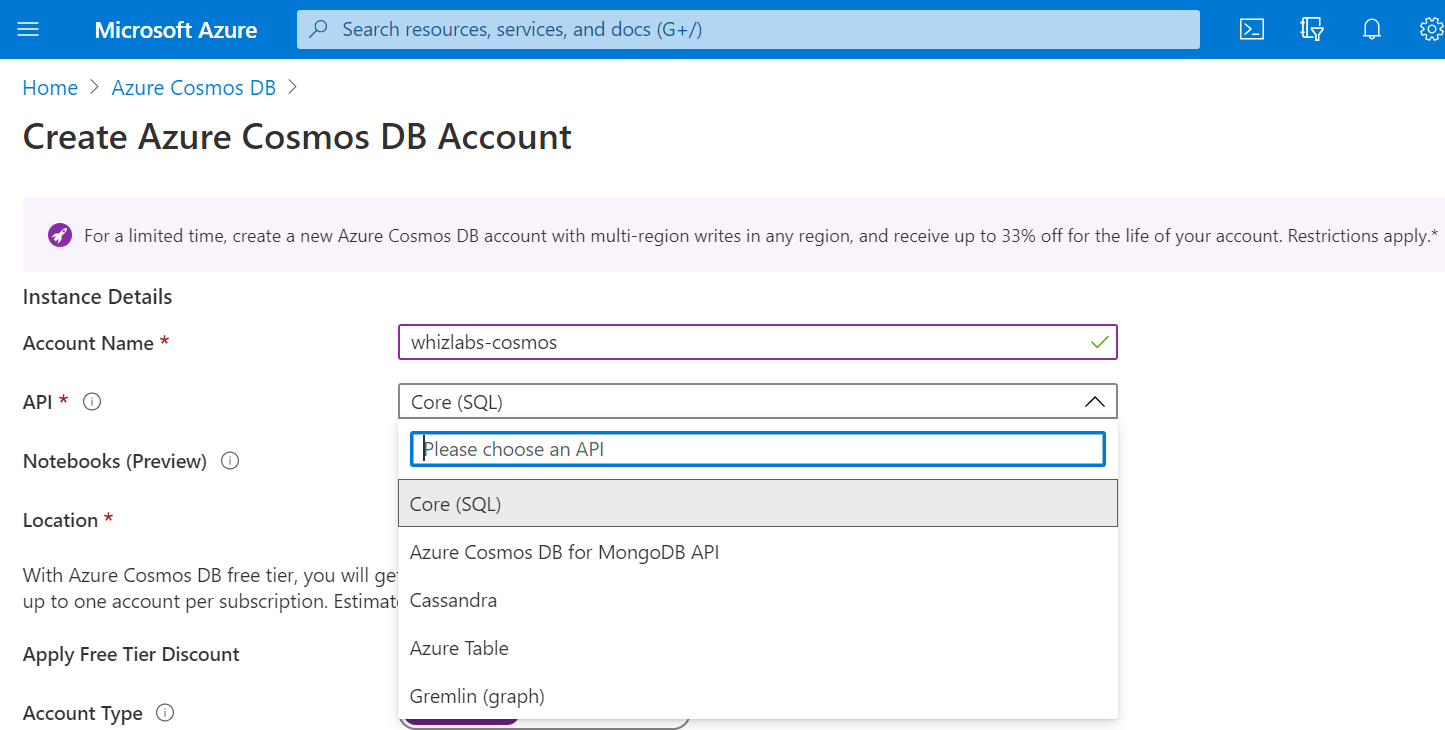
**SQL Server**

**Explanation:**

**Answer: A, B & D**

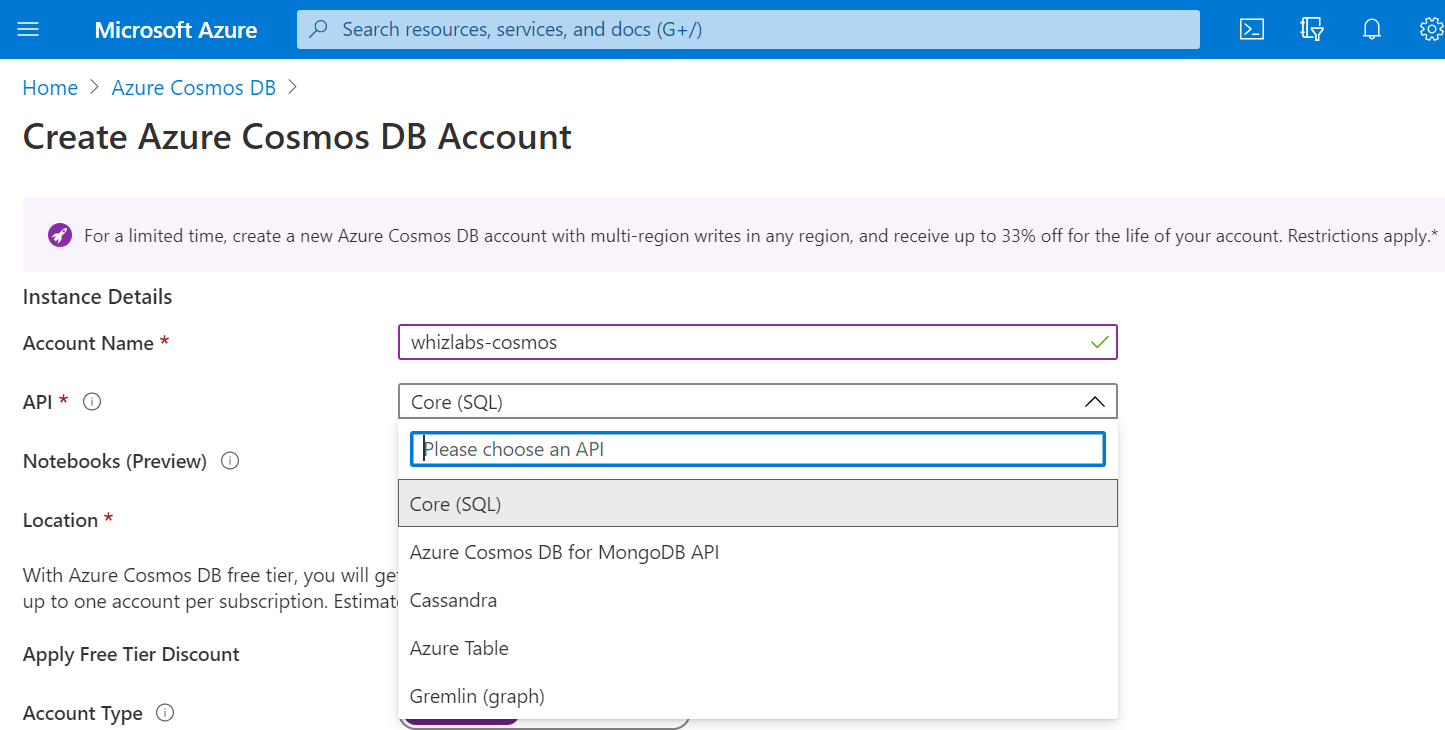
**Option A is CORRECT.**Because this is currently supported using Cassandra API.

The Microsoft documentation mentions the following.



**Option B is CORRECT.**Because this is currently supported using Mongo API.

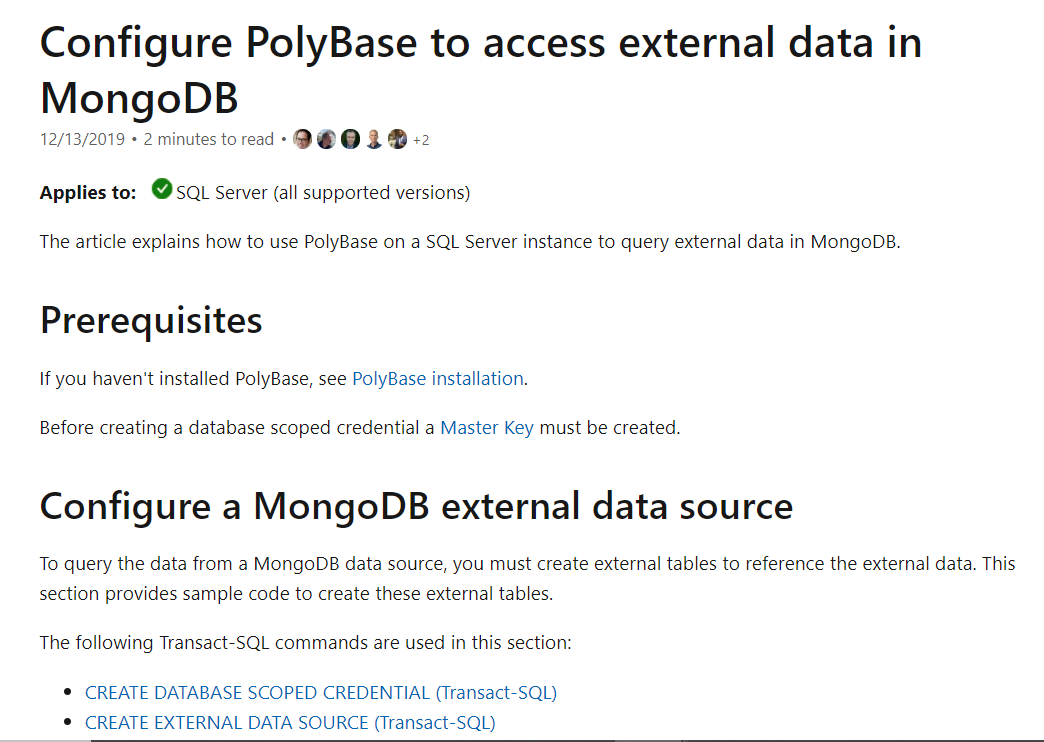
The Microsoft documentation mentions the following.



**Option C is incorrect.**Because this is currently not supported.

**Option D is CORRECT.**Because this is currently supported using SQL server Polybase service and MongoDB API.

The Microsoft documentation mentions the following.



**Reference:**For more information on Cosmos DB (Database support & Polybase services), please refer to the below document link from Azure-

* <https://docs.microsoft.com/en-us/azure/cosmos-db/cassandra-introduction>
* <https://docs.microsoft.com/en-us/azure/cosmos-db/mongodb-introduction>
* <https://docs.microsoft.com/en-us/sql/relational-databases/polybase/polybase-configure-mongodb?view=sql-server-ver15>

### **Question 9**

Domain :Manage and develop data processing

You are working for an international bank as a developer and you have been tasked with implementing a loan approval application which checks the credit score stored in Azure table storage of Azure cosmos db account. What should be the consistency level for this Account?

]A.

**Strong Consistency**

]B.

**Bounded Staleness**

]C.

**Session Consistency**

]D.

**Consistent Prefix**

]E.

**Eventual Consistency**

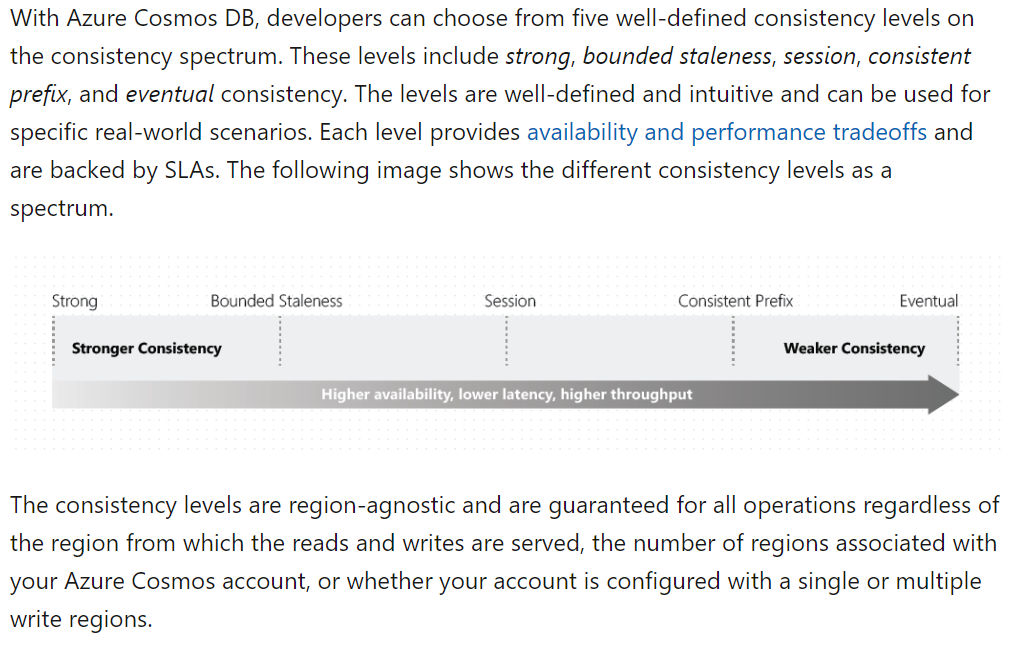
**Explanation:**

**Answer: A**

**Option A is** **CORRECT because** in a strong consistency level,  all the data will be synchronously replicated with every replica in the database and has the highest consistency, only the latest credit score is received anytime.

The Microsoft documentation mentions the following

**Diagram:**



**Option B is incorrect because** bounded staleness data is replicated asynchronously with the predefined stillness window defined by a number of records or time period that means there is possibility of receiving old credit score if data is updated within the stillness window period.

**Option C is incorrect because** session has a medium consistency level which uses a session token internally with each request to ensure the set consistency level means possibility of receiving old credit score.

**Option D is** **incorrect because** consistent Prefix ensures updates contain some prefix of all the updates, with no gaps, and has the possibility of receiving an old credit score.

**Option E is** **incorrect because** Eventual Consistency is the worst fit because it has the highest possibility of receiving an old credit score.

**Reference**: For more information on Cosmos DB (Consistency Levels) please refer to the below document link from Azure :

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

### **Question 10**

Domain :Implement data storage solutions

You are working as a System Administrator in a Fintech company handing various operations and admin tasks on Azure Cosmos DB account of some financial projects of your company.

Which of the following is an inbuilt role of CosmosDB which shall help you provision Azure Cosmos accounts, databases, and containers but cannot access the keys that are required to access the sensitive financial data stored in it?

]A.

**Cosmos Backup Operator**

]B.

**Cosmos DB Account Reader**

]C.

**Create a Custom Role to manage the required provision.**

]D.

**Cosmos DB Operator**

**Explanation:**

**Answer:** **D**

**Option A is incorrect.**Because the Cosmos Backup Operator role will only restore requests for an Azure Cosmos database or a container.

**Option B is incorrect.**Because Cosmos DB Account Reader inbuilt- role will have only read-level Azure Cosmos DB account data. It will not allow provision access to the users.

**Option C is incorrect.**Because Custom role is not an inbuilt role of Azure Cosmos DB.

**Option D is CORRECT.**Because Cosmos DB Operator is an inbuilt role which allows to provision Azure Cosmos accounts, databases, and containers but cannot access the keys that are required to access the data.

**Reference:** For more information on Role-based access control in Azure CosmosDB, please refer to the below document link from Azure-

* <https://docs.microsoft.com/en-us/azure/cosmos-db/role-based-access-control>

### **Question 11**

Domain :Implement data storage solutions

A company is planning on setting up a Real-time recommendation application on the Azure Platform. This application needs guaranteed fast response time anywhere in the world and it’s required to be available always. Which service within Azure will be best fit according to you?

]A.

**Azure SQL Database**

]B.

**Azure Cosmos DB**

]C.

**Azure Data Factory**

]D.

**All of the above**

**Explanation:**

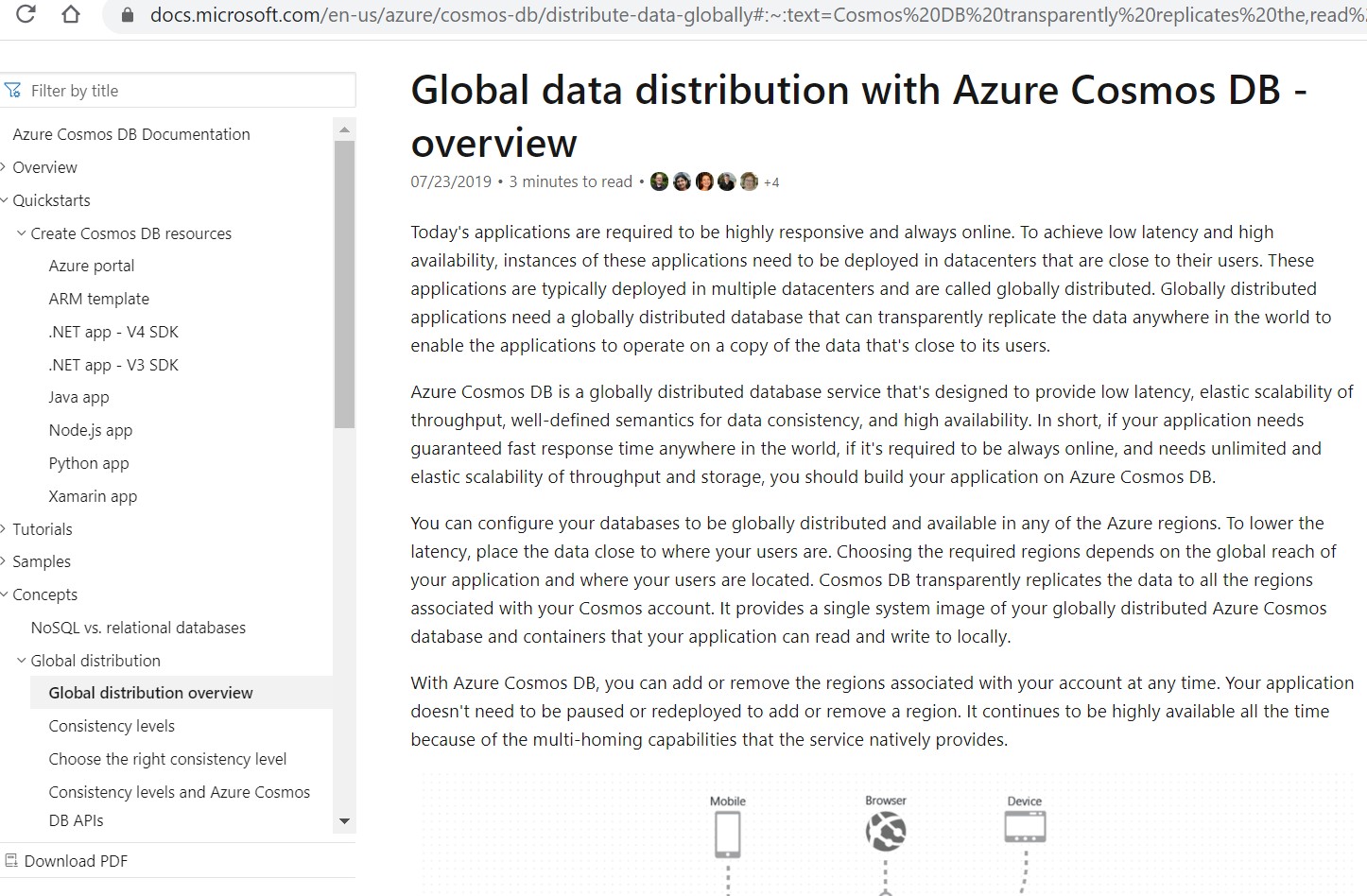
**Answer:** **B**

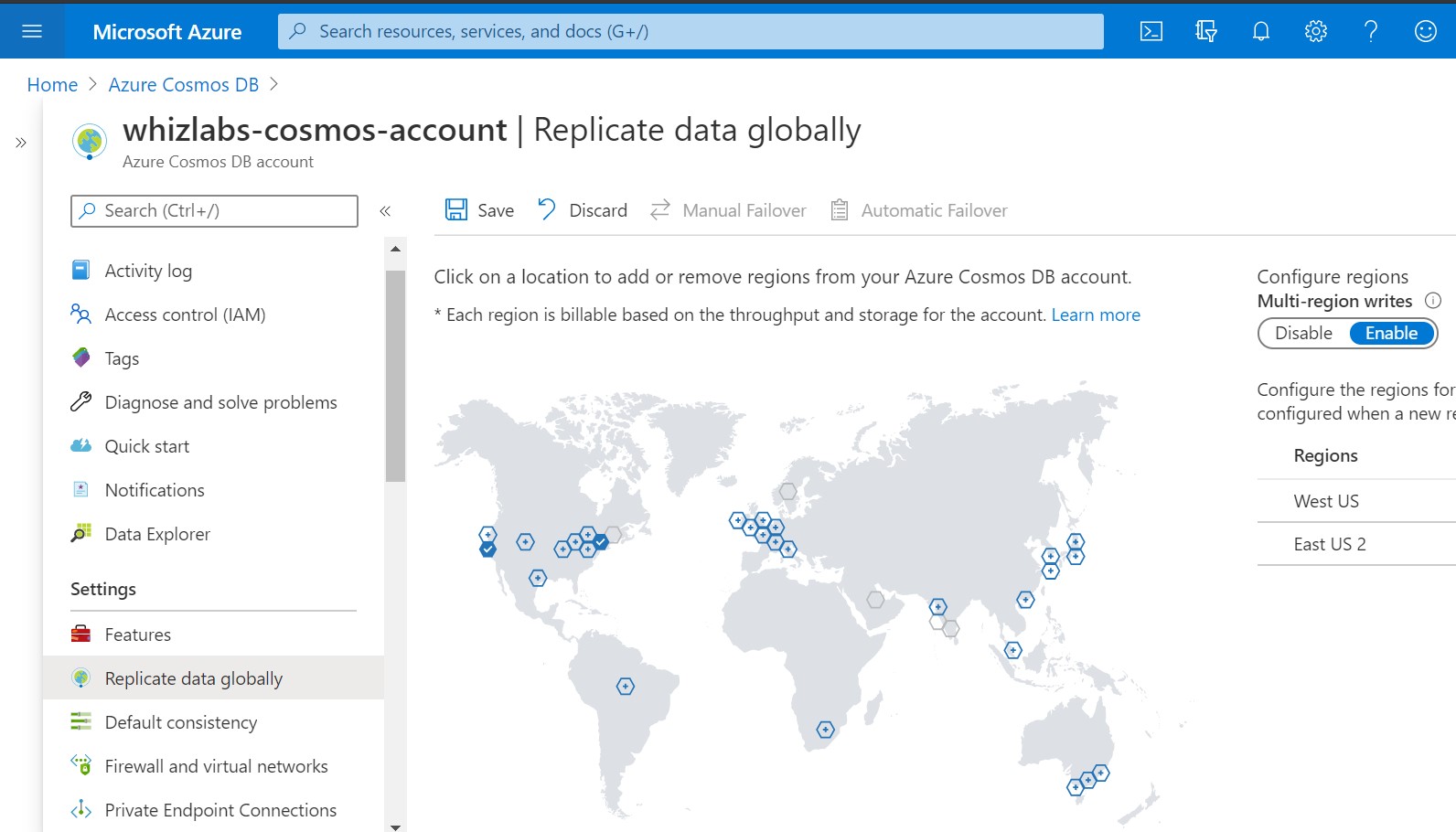
**Option A is incorrect.**Because Azure SQL Database is not suitable for Real-time recommendation application. This type of storage can be used where high transactional volume involves.

**Option B is CORRECT.**Because Azure Cosmos DB is best fit for Real-time recommendation applications where low latency and high availability needed.

The Microsoft documentation mentions the following.

**Diagram:**





**Option C is incorrect.**Because Azure Data Factory is a managed cloud service that provides feasibility to build hybrid extract-transform-load (ETL), extract-load-transform (ELT), and data integration projects.

**Option D is incorrect.**Because we can achieve these requirements using Azure Cosmos DB.

**Reference:**For more information on Cosmos DB (Replication / distribute data globally), please refer to the below document link from Azure-

* <https://docs.microsoft.com/en-us/azure/cosmos-db/distribute-data-globally#:~:text=Cosmos%20DB%20transparently%20replicates%20the,read%20and%20write%20to%20locally.>

### **Question 12**

Domain :Implement data storage solutions

Your company’s primary line of business heavily relies on a globally distributed database solution, which is implemented on Azure Cosmos DB. Business activities span over several regions and, to ensure high availability and timely access to data, your database is distributed over three regions. One of the regions is a write region.

As a data engineer, you have to design for continuity of the operation in case of any outage in the write region.

What should you do?

]A.

**Set manual failover for the Cosmos DB database**

]B.

**Set the database’s consistency level to Session**

]C.

**Enable automatic failover in the Cosmos DB account**

]D.

**Configure all three regions as write regions**

**Explanation:**

**Answer: C**

**Option A is incorrect because** Manual failover wouldn’t bring the expected result since it requires connectivity between the regions to be able to check if there is no data loss during the failover. If the primary region is unavailable, this consistency check fails and the manual failover cannot be completed.

**Option B is incorrect because**setting the consistency level to Session (or any other) won’t result in high availability.

**Option C is CORRECT because** automatic failover must be enabled so that the other regions (in pre-defined priority) can take over the role of the write region in case of any downtime in the write region.

**Option D is incorrect because**multiple write regions generate multiplied cost. This setting alone won’t result in high availability and undisturbed operation.

**Reference:**

* <https://docs.microsoft.com/en-us/azure/cosmos-db/high-availability>
* <https://docs.microsoft.com/en-us/azure/cosmos-db/consistency-levels-tradeoffs>

### **Question 13**

Domain :Implement data storage solutions

Your company’s primary line of business heavily relies on a globally distributed database solution, which is implemented on Azure Cosmos DB. Business activities span over several regions and, to ensure high availability and timely access to data, your database is distributed over three regions. One of the regions is a write region.

As a data engineer, you are tasked to improve the availability of your database to RTO= 0 (Recovery Time Objective).

What should you do?

]A.

**Set partitioning of your database to round-robin**

]B.

**Set all regions to write regions**

]C.

**Set consistency level to Strong**

]D.

**Set all regions to write regions, enable zone redundancy**

**Explanation:**

**Answer: D**

**Option A is incorrect because** partitioning has an effect on the performance and doesn’t contribute to high availability.

**Option B is incorrect because**multiple write regions actually improve the availability but this alone doesn’t fulfil the strict requirement.

**Option C is incorrect because** in case of multi-region writes, Cosmos account cannot be configured to Strong consistency.

**Option D is CORRECT because**zone redundancy, together with multi-region writes, ensures the highest possible data durability in case of regional failures.

**Reference:**

* <https://docs.microsoft.com/en-us/azure/cosmos-db/high-availability>
* <https://docs.microsoft.com/en-us/azure/cosmos-db/consistency-levels-tradeoffs>

### **Question 14**

Domain :Implement data storage solutions

Your company is planning to implement an Azure SQL-based solution in the West US region. Several applications will run on a database and will require it to be available continuously. Highly sensitive data must be defended from unauthorized access, and access rights should be granted with minimal administrative effort.

As a data engineer, you are defining the rules for a secure Azure SQL database solution.

Which option would **not** serve the fulfilment of the above requirements?

]A.

**Enable Transparent Data Encryption (TDE)**

]B.

**On columns containing sensitive data, apply Always Encrypted option**

]C.

**Grant access to all single users who need access to the data**

]D.

**Use built-in or user-defined roles to grant access to data**

**Explanation:**

**Answer: C**

**Option A is incorrect.**Because transparent data encryption (TDE), especially together with Azure Key Vault, is an effective way to separate the data access rights.

**Option B is incorrect.**Because the risk of revealing sensitive data can be mitigated by using Always Encrypted for the given columns.

**Option C is CORRECT.**Because granting access to individual roles results in significant administrative overhead and also implies a risk of failure in a clear separation of responsibilities and duties. Use roles instead.

**Option D is incorrect.**Because it is a common best practice to use roles (either built-in or user-defined) to clearly separate access rights, with the minimal administrative effort. Users then need to be assigned to these roles.

**Reference:**

* <https://docs.microsoft.com/en-us/azure/azure-sql/database/security-best-practice#access-management>

### **Question 15**

Domain :Implement data storage solutions

Your company is planning to implement an Azure SQL-based solution in West US. Several applications will run on a database and will require it to be available continuously. In case of any failures, recovery and continuity of the business must be ensured with the least possible need for manual intervention.

As a data engineer, your task is to design a failure-proof, highly available solution which fulfils this requirement.

Which option would you choose?

]A.

**Configure a failover group with a secondary database in East US.**

]B.

**Set up an active geo-replica in West US.**

]C.

**Configure failover group with a secondary database in West US.**

]D.

**Set up multiple geo-replicas in region West US and East US.**

**Explanation:**

**Answer: A**

**Option A is CORRECT.**Because using failover groups provide automatic switch-over to the secondary server in case of outage of the primary one. Within failover groups, the secondary server always resides in another region, resulting in a HA solution.

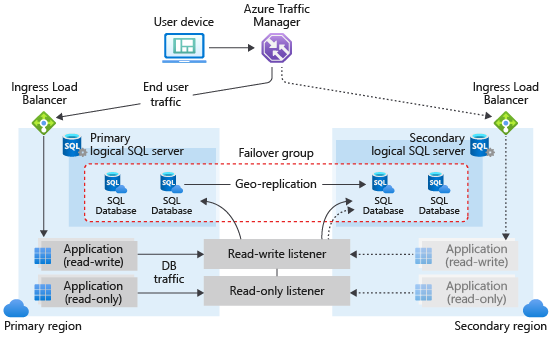
**Option B is incorrect.**Because using geo-replication doesn’t provide automatic failover, as prescribed in the requirements.

**Option C is incorrect.**Because failover groups cannot be configured this way. i.e. when configuring Azure SQL failover groups. The secondary server is always launched in a region different from the primary one.

**Option D is incorrect.**Because using geo-replication (irrespective of the number of replicas) doesn’t provide automatic failover, which is stated in the requirements.

**Reference:**

Schema of high-availability solutions in Azure:



* <https://docs.microsoft.com/en-us/azure/azure-sql/database/business-continuity-high-availability-disaster-recover-hadr-overview>
* <https://docs.microsoft.com/en-us/azure/azure-sql/database/auto-failover-group-overview?tabs=azure-powershell#best-practices-for-sql-database>

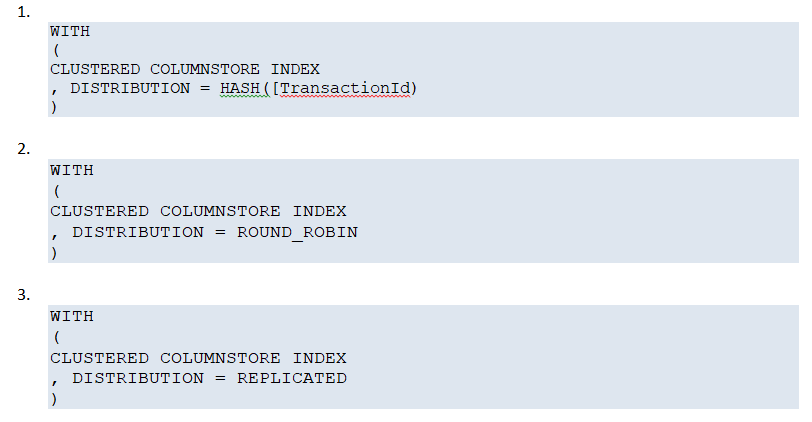
### **Question 16**

Domain :Implement data storage solutions

Your company is going to use Azure Synapse’s data warehousing capabilities to support their regular analytics tasks. Sales transactions from online stores have to be collected continuously and aggregated to *salesdata*table, which is expected to scale to TBs. A large dimension table (*products)* and a small (*product\_categories*) are connected to *salesdata* in a star scheme architecture are maintained and refreshed weekly.

As a data engineer, your task is to implement a solution ensuring that processing of *salesdata* table is optimal and exploits the MPP capabilities of Synapse Analytics.

While creating the database tables you have the following options with CREATE TABLE…



Which code segments above would you use for *salesfact, products, product\_categories*?

]A.

**1,1,3**

]B.

**1,3,1**

]C.

**3,1,1**

]D.

**2,2,2**

**Explanation:**

**Answer: A**

**Option A is CORRECT.**Because distribution method HASH is the optimal solution for large fact tables and for large dimension tables with many joins and common hash key. Small dimension tables are optimally distributed along the REPLICATED pattern.

**Option B is incorrect.**Because REPLICATED should be avoided in case large tables.

**Option C is incorrect.**Because REPLICATED is typically used for small dimension tables. For fact, tables use HASH for optimal performance.

**Option D is incorrect.**Because while ROUND\_ROBIN could be a good choice as a generic solution in many cases for large fact tables, HASH provides the best performance. REPLICATED is recommended for smaller dimension tables.

**Reference:**

* <https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/massively-parallel-processing-mpp-architecture>

### **Question 17**

Domain :Implement data storage solutions

Your company is collecting a huge amount of data from daily transactions and stores it in Azure Blob Storage. Data is available in various formats. In order to serve the analytics department, all the data have to be uploaded to an Azure Synapse Analytics platform. Upload sessions take place weekly, and require quite a long time to complete. An optimized time-effective upload process has to be implemented, which can cope with the growing amount of data, using specialized Azure services.

You, as a data engineer, suggest using PolyBase for loading the data from the Blob Storage to Synapse Analytics.

What steps should you include in your process, in what order?

]A.

**Create external table; Create master key; Create external data source; Create database scoped credential**

]B.

**Create database scoped credential; Create external table; Create master key; Create external data source**

]C.

**Create external data source; Create database scoped credential; Create external table; Create master key**

]D.

**Create database scoped credential; Create master key; Create external data source; Create external table**

**Explanation:**

**Answer: D**

**Option A is incorrect. B**ecause an external data source must be defined before creating an external table.

**Option B is incorrect.**Because an external data source must be defined before creating an external table. master key is needed for the credential.

**Option C is incorrect.**Because creating a master key must precede creating the database scoped credential.

**Option D is CORRECT.**Because after creating the master key for encrypting the database scoped credential, an external data source is needed so that external table could be created.

**Reference:**

* <https://docs.microsoft.com/en-us/sql/relational-databases/polybase/polybase-guide?view=sql-server-ver15>
* [https://docs.microsoft.com/en-us/sql/relational-databases/polybase/polybase-configure-azure-blob-storage?view=sql-server-ver15](https://docs.microsoft.com/en-us/sql/t-sql/statements/create-database-scoped-credential-transact-sql?view=sql-server-ver15)

### **Question 18**

Domain :Implement data storage solutions

Your company is collecting a huge amount of data from daily transactions and stores it in Azure Blob Storage. Data is available in various formats. In order to serve the analytics departments, all the data have to be uploaded to an Azure Synapse Analytics platform. Upload sessions take place weekly, and require quite a long time to complete. An optimized time-effective upload process has to be implemented, which can cope with the growing amount of data, making use of specialized Azure services.

You, as a data engineer, decided to use PolyBase for loading the data to a staging table in Synapse Analytics, maximizing the processing performance by exploiting Synapse’s MPP capabilities.

Which **two**actions could help you to achieve your goal?

A.

**Disable database statistics on the target table.**

B.

**Split your data into 60 equally sized files.**

C.

**Decrease the number of DWU units.**

D.

**Increase the number of DWU units.**

**Explanation:**

**Answer: B, D**

**Option A is incorrect.**On the opposite, it is the best practice to create table statistics and refresh them regularly to improve performance.

**Option B is CORRECT.**Because Azure Synapse’s MPP (Massive Parallel Processing) architecture can distribute the tasks among its 60 compute nodes.

**Option C is incorrect**because decreasing the DWU units will result in a decreased computing performance.

**Option D is CORRECT**because a higher DWU can increase the throughput of the compute nodes (although at significantly higher cost).

**Reference:**

* <https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-best-practices>
* <https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/what-is-a-data-warehouse-unit-dwu-cdwu>

### **Question 19**

Domain :Implement data storage solutions

In order to prevent sensitive data from being exposed to unauthorized persons, Azure SQL provides tools for dynamically masking field contents. There are pre-defined masking patterns (functions) which can be used. But the custom type is also supported. As a data engineer, you are responsible for defining and setting the appropriate Dynamic Data Masking (DDM) rules for user-facing queries.

Which masking functions would you select to satisfy the following rules?

Rule\_1: the domain of Customer’s email address must not be exposed.

Rule\_2: Customer’s salary must not be revealed.

]A.

**Default and Email**

]B.

**Email and Custom**

]C.

**Email and Default**

]D.

**Random and Custom**

**Explanation:**

**Answer: C**

**Option A is incorrect because**rule Default provides a single unified content for the field, regardless of its content. Rule Email is not applicable for numeric data**.**

**Option B is incorrect because**Custom is applicable only for text data.

**Option C is CORRECT because**there is a built-in masking function for exposing only some parts of an email address, which is the right option for Rule\_1. The best option for Rule\_2 is using the field’s default value which completely hides the real content, with the least possible effort.

**Option D is incorrect because**built-in function Random is applicable only for numeric data.

**Reference:**

* <https://docs.microsoft.com/en-us/azure/azure-sql/database/dynamic-data-masking-overview>

### **Question 20**

Domain :Implement data storage solutions

Your company is storing a large amount of data, including sensitive personal data in an Azure SQL database. It is TDE protected. For archiving purposes, the database has to be duplicated to another location, with minimizing the risk of exposing sensitive information at the lowest possible cost. As a data engineer, your task is to implement a solution. You need to secure that sensitive content will not be risked to any unauthorized access during the process.

Which solution would you choose?

]A.

**export database\_A to a BACPAC file, move the file, import to database\_B**

]B.

**copy the content of database\_A using Data Factory to database\_B**

]C.

**export database\_A to BACPAC, set TDE for BACPAC, move the file, import to database\_B**

]D.

**set Always Encrypted for sensitive data in database\_A, export to BACPAC, move the file, import to database\_B**

**Explanation:**

**Answer: D**

**Option A is incorrect because**exporting TDE-protected data, the encryption isn’t transferred to the BACPAC file.

**Option B is incorrect because**using Data Factory pipelines for transferring huge amount of data would be highly expensive.

**Option C is incorrect because**the transparent data encryption (TDE) is applicable only for databases; there is no such option for BACPAC.

**Option D is CORRECT because**only the Always Encrypted feature ensures that the data is encrypted during BACPAC transfer.

**Reference:**

* <https://docs.microsoft.com/en-us/sql/relational-databases/security/encryption/always-encrypted-migrate-using-bacpac?view=sql-server-ver15>
* <https://docs.microsoft.com/en-us/sql/relational-databases/data-tier-applications/data-tier-applications?view=sql-server-ver15>

### **Question 21**

Domain :Implement data storage solutions

Your company is storing its employee records (including sensitive data) in an Azure SQL database. No one but the owners of the content (the HR department) must be allowed to access the sensitive data. The whole database content must be encrypted at rest.

As a data engineer, your task is to implement a solution which ensures that the sensitive content will be accessible only to the owners of the data.

What should you do?

]A.

**On creation of the database, manually set TDE to enabled.**

]B.

**On creation of the database, leave the TDE setting as is, and set    Always Encrypted for the sensitive data.**

]C.

**On creation of the database set TDE disabled and apply dynamic data masking on the sensitive data.**

]D.

**Apply dynamic data masking on the table containing the sensitive data.**

**Explanation:**

**Answer: B**

**Option A is incorrect because** Transparent Data Encryption itself cannot provide the required security. In Azure SQL, TDE is enabled by default, no need to set it manually.

**Option B is CORRECT because**applying the Always Encrypted option to the sensitive data can ensure that the content will not be revealed to unauthorized access. In Azure SQL, TDE is enabled by default, which is a recommended setting.

**Option C is incorrect because**disabling TDE adds vulnerability to the database and dynamic data masking doesn’t prevent operational staff (like DBAs) from accessing sensitive data.

**Option D is incorrect because**dynamic data masking doesn’t prevent operational staff (like DBAs) from accessing sensitive data. It is not recommended to apply DDM to whole tables.

**Reference:**

* <https://docs.microsoft.com/en-us/sql/relational-databases/security/encryption/always-encrypted-database-engine?view=sql-server-ver15>

### **Question 1**

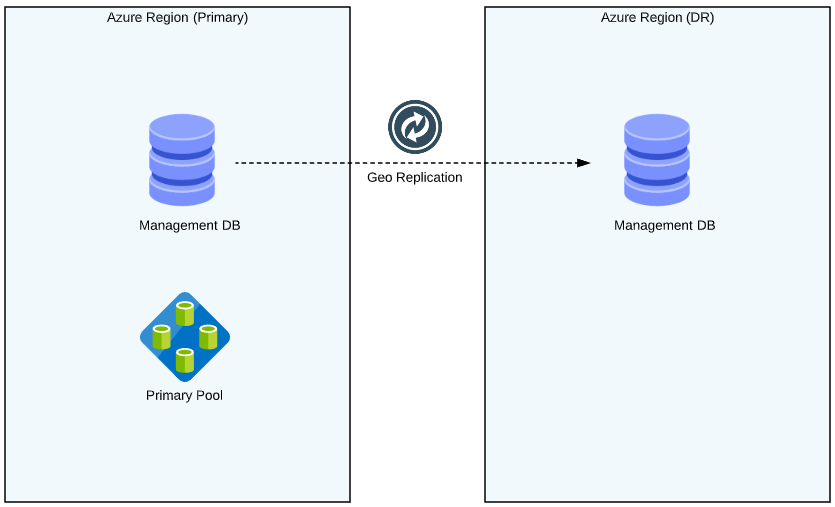
Domain :Implement data storage solutions

A cost-sensitive organization wants to simplify deployment and management of their application. Their objective is to ensure that application as a whole is never offline.

To satisfy the simplicity requirement,

* All tenant databases are deployed into one elastic pool in the Azure region.
* Management databases are deployed as geo-replicated single databases.
* With no additional cost geo-restore is used for the disaster recovery of tenants.
* Availability of the management databases is ensured by geo-replicating them to another region using an auto-failover group.

This configuration is illustrated in the next diagram given below.



Refer to the scenario given above and answer the following.

* In case of disaster in the Primary Azure Region, what recovery steps need to be performed to bring the application online?

Select three choices.

A.

**Manually failover management database to Azure DR region.**

B.

**The elastic pool is created with the same configuration as the original pool.**

C.

**Automatic failover is initiated by failover groups of the management database to the Azure DR region.**

D.

**Geo-restore is used to create copies of the tenant databases.**

E.

**Set all databases in the DR pool to read-only to ensure they cannot be modified in the DR region.**

**Explanation:**

**Answer:** B, C and D

Option A is incorrect because the management database is set up to failover automatically.

Option B is CORRECT because the management database is set up to failover automatically.

Option C is CORRECT because elastic pool in the DR region requires the same configuration as in Primary region in order to handle the load of all tenants.

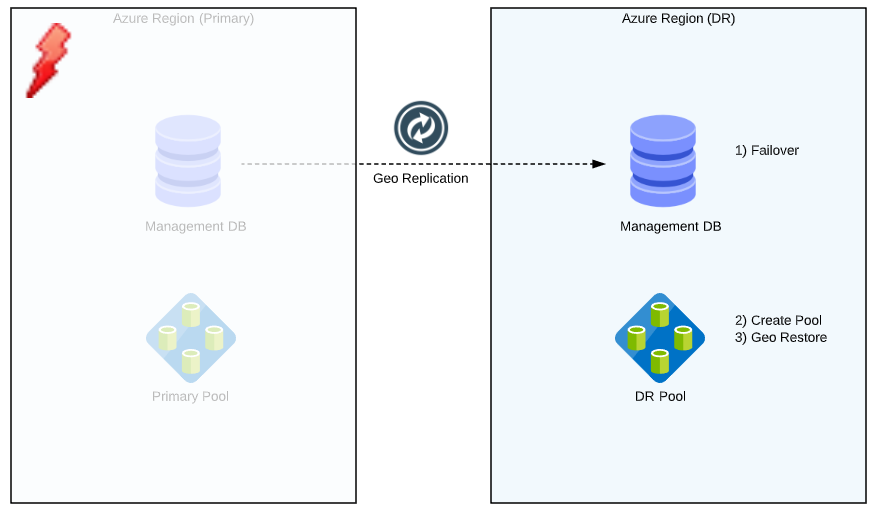
Option D is CORRECT as geo restore is required to recover tenants. Geo-restore comes at no extra cost.

Option E is incorrect as this is a step required to bring the application back to Primary Azure region.

In case of a disaster in Primary Azure Region, the following recovery steps are performed to bring the application online.

1. Automatic failover is initiated by failover groups of the management database to the Azure DR region. The application is automatically reconnected to the new primary and all new accounts and tenant databases are created in the DR region. The existing customers see their data temporarily unavailable.
2. The elastic pool is created with the same configuration as the original pool.
3. Geo-restore is used to create copies of the tenant databases.

**Diagram:**



**Reference:**

* <https://docs.microsoft.com/en-us/azure/sql-database/sql-database-disaster-recovery-strategies-for-applications-with-elastic-pool>

### **Question 2**

Domain :Implement data storage solutions

An organization wants to limit exposure to sensitive data in a SQL server to non-privileged users. Database users are allowed to connect to databases directly and execute queries that provide sensitive data. The developers should be able to query the Production environment and PII information for troubleshooting without violating controls for compliance.

The organization does not want to replace sensitive data permanently by altering it at rest within database copies. Instead, they aim at temporarily hiding or replacing sensitive data in transit, leaving the original at-rest data intact and unaltered.

Refer to the scenario given above and answer the following.

Which data security feature should you use in this scenario?

]A.

**Static Data Masking**

]B.

**Row Level Security**

]C.

**Dynamic Data Masking**

]D.

**Password Policy**

**Explanation:**

**Answer:** C

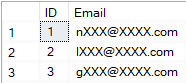
Option A is incorrect because it permanently replaces sensitive data by altering data at rest.

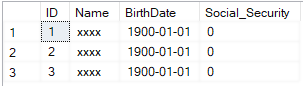
Option B is incorrect because it enables to use execution context to control access to rows in a table.

Option C is CORRECT because it aims to temporarily hide or replace sensitive data in transit, leaving the original at-rest data intact and unaltered.

Option D is incorrect because password policies are designed to deter brute force attacks by increasing the number of possible passwords.

**Diagram:**





**Reference:**

* <https://docs.microsoft.com/en-us/sql/relational-databases/security/dynamic-data-masking?view=sql-server-ver15>

### **Question 3**

Domain :Implement data storage solutions

**Scenario**:

In an organization, you track SQL Server encryption progress by querying ‘percent\_complete’ and ‘encryption\_state’ columns. ‘encryption\_state’ column indicates the encryption status of the database and ‘percent\_complete’ column tells percent complete of the DB encryption state change.

You, as a data engineer, are dealing with a huge database. A decision is made at your organization not to allow TDE encryption scanners to run during business hours. If you see any performance issues, you would like to halt the process temporarily. You do not have an option to lower the priority of TDE scanner.

Refer to scenario given above and answer the following question.

While the workload on the system is heavy, or during business-critical hours, what query would you run to pause the scan that can be resumed later?

]A.

**ALTER DATABASE SET ENCRYPTION ON;**

]B.

**ALTER DATABASE SET ENCRYPTION PAUSE;**

]C.

**ALTER DATABASE SET ENCRYPTION OFF;**

]D.

**ALTER DATABASE SET ENCRYPTION SUSPEND;**

**Explanation:**

**Answer:** D

Option A is incorrect because the objective is to pause the scan.

Option B is incorrect because it will start decryption instead.

Option C is incorrect because the correct syntax is to use SUSPEND.

Option D is CORRECT because this is the correct syntax.

**Reference:**

* [**h**ttps://docs.microsoft.com/en-us/sql/relational-databases/security/encryption/transparent-data-encryption?view=sql-server-ver15](https://docs.microsoft.com/en-us/sql/relational-databases/security/encryption/transparent-data-encryption?view=sql-server-ver15)
* <https://docs.microsoft.com/en-us/sql/relational-databases/system-dynamic-management-views/sys-dm-database-encryption-keys-transact-sql?view=sql-server-2017>

### **Question 4**

Domain :Implement data storage solutions

**Scenario**: You must secure your data using database encryption. Taking advantage of SQL Server database encryption can be a difficult task in a huge database. But once it is configured, it is also important to monitor it. SQL Server keeps track of the encryption progress and we can pull that information by querying database columns. As a data engineer, you are regularly checking database state that you expect them to be in. During audits, you run queries and produce artefacts to support data security and encryption.

**Question Text**: Refer to scenario given above and choose two options to complete the query given below:

SELECT DB\_NAME(database\_id), encryption\_state,

encryption\_state\_desc =

CASE \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

         WHEN '0'  THEN  'No database encryption key present, no encryption'

         WHEN '1'  THEN  'Unencrypted'

         WHEN '2'  THEN  'Encryption in progress'

         WHEN '3'  THEN  'Encrypted'

         WHEN '4'  THEN  'Key change in progress'

         WHEN '5'  THEN  'Decryption in progress'

         WHEN '6'  THEN  'Protection change in progress'

         ELSE 'Not Encrypted'

         END AS 'Desc',

percent\_complete, encryptor\_type  FROM \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

A.

**sys.dm\_database\_encryption\_keys**

B.

**encryption\_state**

C.

**encryptor\_type**

D.

**Encryption\_scan\_state**

E.

**encryption\_scan\_state\_desc**

**Explanation:**

**Answer:** A, B

Option A is CORRECT. Refer to the complete query given below.

Option B is CORRECT. Refer to the complete query given below.

Option C is incorrect. Refer to the complete query given below.

Option D is incorrect. Refer to the complete query given below.

Option E is incorrect. Refer to the complete query given below.

SELECT DB\_NAME(database\_id), encryption\_state,

encryption\_state\_desc =

CASE encryption\_state

         WHEN '0'  THEN  'No database encryption key present, no encryption'

         WHEN '1'  THEN  'Unencrypted'

         WHEN '2'  THEN  'Encryption in progress'

         WHEN '3'  THEN  'Encrypted'

         WHEN '4'  THEN  'Key change in progress'

         WHEN '5'  THEN  'Decryption in progress'

         WHEN '6'  THEN  'Protection change in progress'

         ELSE 'Not Encrypted'

         END AS 'Desc',

percent\_complete, encryptor\_type  FROM sys.dm\_database\_encryption\_keys

**Reference:**

* <https://docs.microsoft.com/en-us/sql/relational-databases/system-dynamic-management-views/sys-dm-database-encryption-keys-transact-sql?view=sql-server-2017>

### **Question 5**

Domain :Implement data storage solutions

**Scenario**: In your organization, you are tasked to achieve peak performance of Azure Cosmos DB by making client-side optimizations. You do not have to make major architecture changes or write complex code.

* You plan to reduce additional network hops.
* You are not limited to the number of socket connections.
* You plan to avoid startup latency.
* You choose regions to avoid higher latency.

Refer the scenario given above and answer the following.

How can you improve your Cosmos database performance? Choose three options.

A.

**Use direct connection mode.**

B.

**Avoid caching document URIs.**

C.

**Call OpenAsync() once during initialization.**

D.

**Collocate clients in same Azure region.**

E.

**Do not increase number of threads.**

F.

**Tune for higher request units/second usage.**

**Explanation:**

**Answer:** A, C and D

Option A is CORRECT  because direct mode reduced the number of network hops.

Option B is incorrect because caching will improve performance.

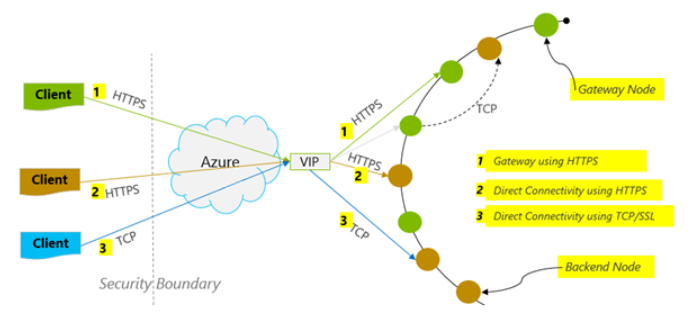
Option C is CORRECT because the first request has a higher latency, reason being, it has to fetch the address routing table.

Option D is CORRECT because collocating clients in the same Azure region will improve performance.

Option E is incorrect because increasing the number of threads will improve performance.

Option F is incorrect because tuning for lower request units/second usage will improve performance.

**Diagram:**



**Reference:**

* <https://docs.microsoft.com/en-us/azure/cosmos-db/performance-tips>

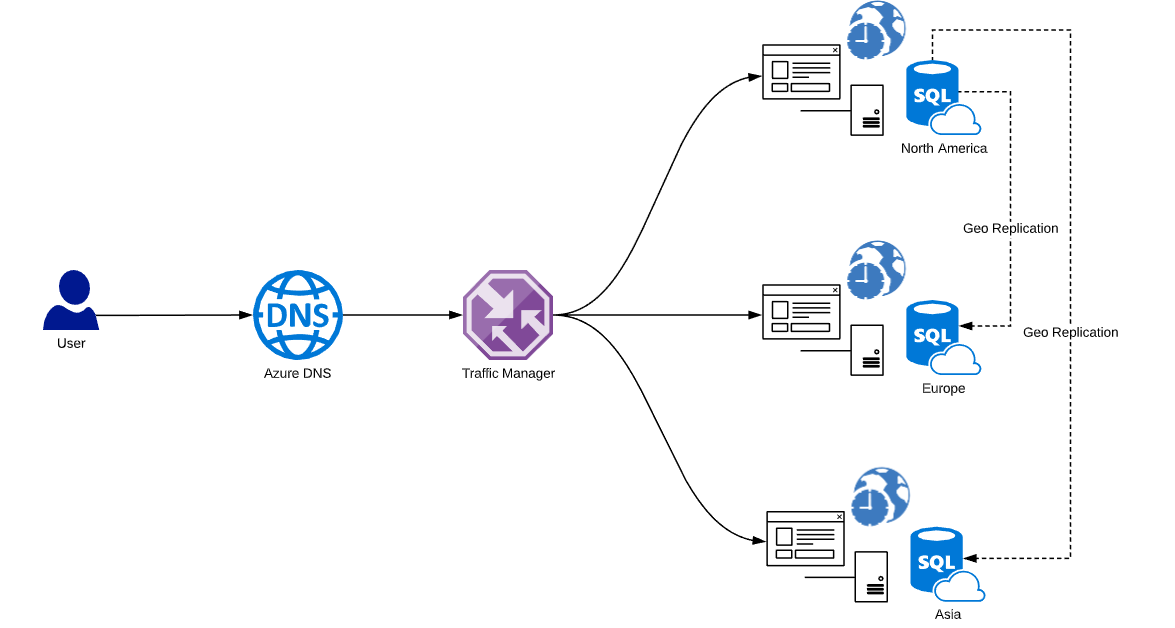
### **Question 6**

Domain :Implement data storage solutions

**Scenario**: You have a global business model and you have deployed applications in different geographies as per "follow the sun" model. In this scenario,

* The end users access the application from different geographies.
* The application includes read-only workloads that do not depend on full synchronization
* Write access is supported in the same geography for regional users
* Read latency is critical for the end user experience

You have designed the user device connectivity to the application deployed in the same geography for the analytical operations. However, transactional operations are processed in the same geography most of the time. The application resources should be deployed in each geography where you have substantial usage demand. For example, if your application is actively used in North America, Europe and Asia the application should be deployed to all of these geographies.



The primary database should be dynamically switched from one geography to the next at the end of the working hours. Geo-replication is setup between Azure SQL databases. Automation is achieved using Azure logic apps. Traffic manager is configured with the performance routing method. It ensures that the end user’s device is connected to the web service in the closest region. Traffic manager should be set up with endpoint monitoring enabled for each web service endpoint. The failover group configuration defines which region is used for failover.

Given the scenario above, state three key benefits of this design:

A.

**The application can survive a loss of one of the regions without any significant downtime.**

B.

**The overall cost is reduced since only resources in local geographies are used during their working hours.**

C.

**The design is simplified since each region gets servers in their local geography.**

D.

**The read-only application workload accesses data in the closest region.**

E.

**The read-write application workload accesses data in the closest region during the period of the highest activity in each geography.**

**Explanation:**

**Answer:** A, D and E

Option A is CORRECT because failover will ensure application availability. There might be a higher latency

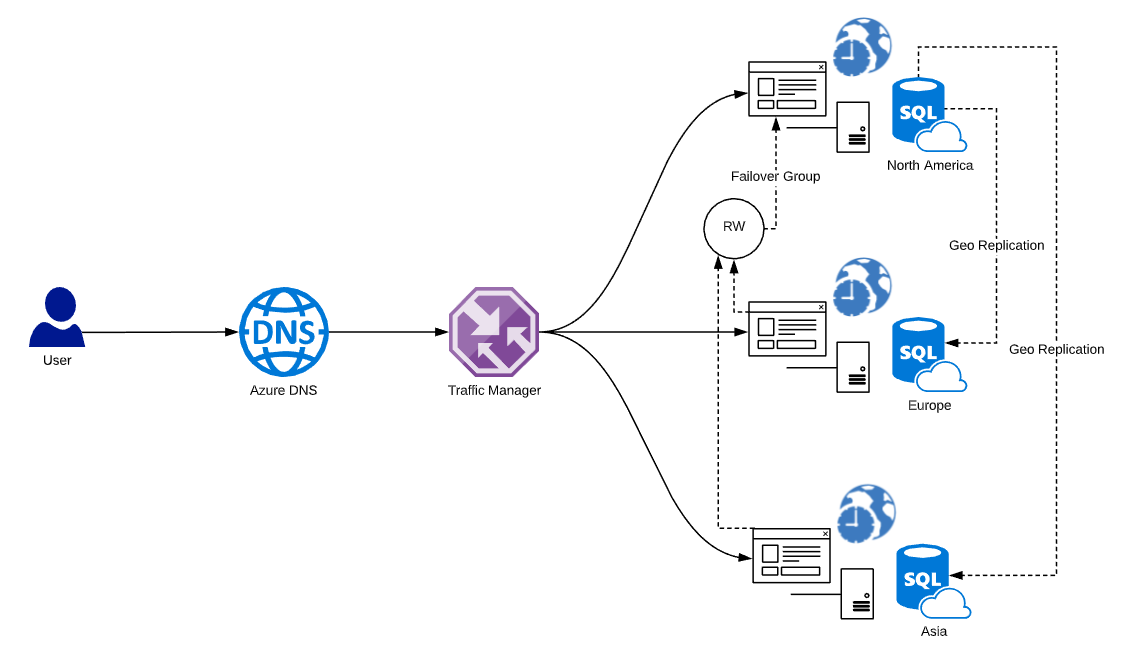
Option B is incorrect because overall cost would increase with additional resources and egress traffic

Option C is incorrect because design gets complex with additional components and automation scenario

Option D is CORRECT lower latency is offered during read-only operations as per the requirement

Option E is CORRECT as write operations’ performance is improved in the closest Azure region

**Diagram:**

****

**Reference:**

* <https://docs.microsoft.com/en-us/azure/sql-database/sql-database-designing-cloud-solutions-for-disaster-recovery>

### **Question 7**

Domain :Monitor and optimize data solutions

**Scenario**: In an organization, you use Azure SQL database monitoring tools to troubleshoot potential problems and make recommendations in order to improve database performance. You receive an incident where the system is running slow. You connect with the end user and get to know that the query is taking longer than expected. As an initial troubleshooting step, you check the monitor resource usage using dynamic management views.

**Text**: Refer to the scenario given above and complete the following query:

SELECT

    AVG(avg\_cpu\_percent) AS 'Average CPU use in percent',

    MAX(avg\_cpu\_percent) AS 'Maximum CPU use in percent',

    AVG(avg\_data\_io\_percent) AS 'Average data IO in percent',

    MAX(avg\_data\_io\_percent) AS 'Maximum data IO in percent',

    AVG(avg\_log\_write\_percent) AS 'Average log write use in percent',

    MAX(avg\_log\_write\_percent) AS 'Maximum log write use in percent',

    AVG(avg\_memory\_usage\_percent) AS 'Average memory use in percent',

    MAX(avg\_memory\_usage\_percent) AS 'Maximum memory use in percent'

FROM \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_;

]A.

**sys.resource\_stats**

]B.

**sys.dm\_tran\_active\_transactions**

]C.

**sys.dm\_db\_resource\_stats**

]D.

**sys.dm\_tran\_database\_transactions**

**Explanation:**

**Answer:** C

Option A is incorrect because it does not have column “avg\_memory\_usage\_percent”

Option B is incorrect because it is used for active transaction details

Option C is CORRECT because it provides resource usage. Refer the query below.

Option D is incorrect because it provides information about transactions at the database level.

SELECT

    AVG(avg\_cpu\_percent) AS 'Average CPU use in percent',

    MAX(avg\_cpu\_percent) AS 'Maximum CPU use in percent',

    AVG(avg\_data\_io\_percent) AS 'Average data IO in percent',

    MAX(avg\_data\_io\_percent) AS 'Maximum data IO in percent',

    AVG(avg\_log\_write\_percent) AS 'Average log write use in percent',

    MAX(avg\_log\_write\_percent) AS 'Maximum log write use in percent',

    AVG(avg\_memory\_usage\_percent) AS 'Average memory use in percent',

    MAX(avg\_memory\_usage\_percent) AS 'Maximum memory use in percent'

FROM sys.dm\_db\_resource\_stats;

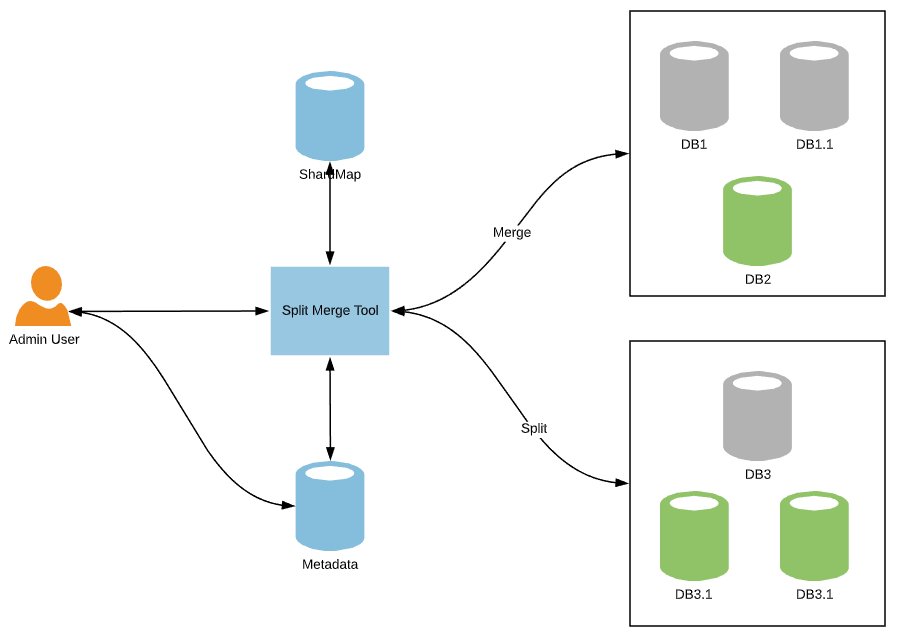
**Reference:**

* <https://docs.microsoft.com/en-us/azure/sql-database/sql-database-monitoring-with-dmvs>

### **Question 8**

Domain :Implement data storage solutions

**Scenario**: In order to accommodate the data growth, you decide to add more shards. Redistribution of the data to the new databases has to be done without disrupting the data integrity. You use the split-merge tool to orchestrate the data movement between shards in combination with the necessary shard map updates. It uses shard map management to maintain the service metadata database and ensure consistent mappings. For reference tables, the split, merge and move operations copy the data from the source to the target shard.



Refer the scenario given above and choose three requirements or limitations for split-merge service.

A.

**Collocate your split-merge service in the region and data center where your databases reside.**

B.

**Shards need to exist and be registered in the shard map before a split-merge operation on these shards can be performed.**

C.

**Service does not create tables or any other database objects automatically as part of its operations.**

D.

**Schema for all sharded tables and reference tables needs to exist on the target shard prior to any split operation.**

E.

**Keep the test tenant data size above the maximum data size of your largest tenant.**

**Explanation:**

**Answer:** B, C and D

Option A is incorrect because it is the best practice, not a requirement or limitation.

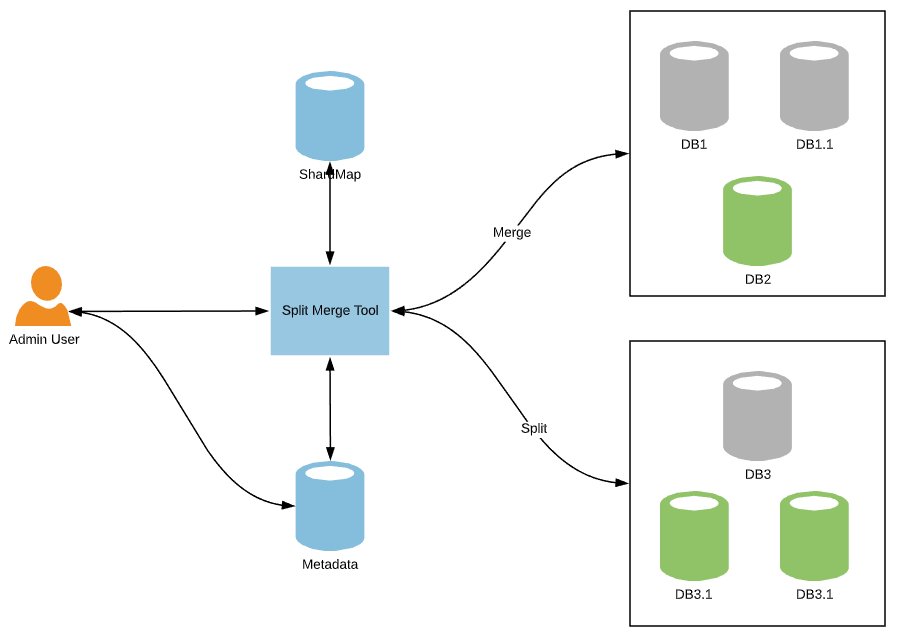
Option B is CORRECT because it is a requirement that shards need to exist and registered in shard map.

Option C is CORRECT because service does not create tables automatically.

Option D is CORRECT because it is a requirement for split-merge service.

Option E is incorrect because it is a best practice, not a requirement or limitation.

**Diagram:**



**Reference:**

* <https://docs.microsoft.com/en-us/azure/sql-database/sql-database-elastic-scale-overview-split-and-merge>

### **Question 9**

Domain :Monitor and optimize data solutions

**Scenario**:Your application team developed an application using Azure App Service and Azure SQL Database as a back-end using Free Tier. After a few days you receive an error message when the connection to Azure SQL Database has failed. You review the log and get the error code as 40544, that is, the database has reached its size quota. Partition or delete data, drop indexes, or consult the documentation for possible resolutions. As the next step, you decide to identify which tables are consuming the most space and are therefore potential candidates for cleanup.

Refer to scenario given above and choose an option to complete the query give below:

*SELECT o.name,*

*SUM(p.row\_count) AS 'Row Count',*

*SUM(p.reserved\_page\_count) \* 8.0 / 1024 AS 'Table Size (MB)'*

*FROM sys.objects o*

*JOIN \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ p on p.object\_id = o.object\_id*

*GROUP BY o.name*

*ORDER BY [Table Size (MB)] DESC*

]A.

**sys.dm\_db\_partition\_stats**

]B.

**sys.dm\_db\_file\_space\_usage**

]C.

**sys.dm\_db\_log\_stats**

]D.

**sys.dm\_db\_session\_space\_usage**

**Explanation:**

**Answer:** A

Option A is CORRECT because sys.dm\_db\_partition\_stats Returns page and row-count information for every partition in the current database

Option B is incorrect because it does not produce desired results on the join with sys.objects

Option C is incorrect because it does not produce desired results on the join with sys.objects

Option D is incorrect because it does not produce desired results on the join with sys.objects

SELECT o.name,

 SUM(p.row\_count) AS 'Row Count',

 SUM(p.reserved\_page\_count) \* 8.0 / 1024 AS 'Table Size (MB)'

FROM sys.objects o

JOIN sys.dm\_db\_partition\_stats p on p.object\_id = o.object\_id

GROUP BY o.name

ORDER BY [Table Size (MB)] DESC

**Reference:**

* <https://docs.microsoft.com/en-us/azure/sql-database/troubleshoot-connectivity-issues-microsoft-azure-sql-database>

### **Question 10**

Domain :Monitor and optimize data solutions

**Scenario**: You have a requirement to configure logging of diagnostics telemetry for Azure SQL databases in order to get performance reports, alerts, and mitigation recommendations. You go ahead and enable streaming of diagnostics telemetry for an elastic pool. However, as you realize the database temetry is missing. You troubleshoot and find out that you also need to configure diagnostics telemetry for each database in the elastic pool.



### **Question Text**: Follow the scenario given above and choose three options to enable streaming of diagnostic telemetry for an elastic pool resource.

A.

**Turn off diagnostics if already enabled**

B.

**Select the checkbox for ResourceUsageStats**

C.

**Select destination resource as "Send to Log Analytics"**

D.

**Choose log telemetry options**

E.

**Select the check box for Basic metrics**

**Explanation:**

**Answer:** C, D and E

Option A is incorrect because diagnostics need to be enabled

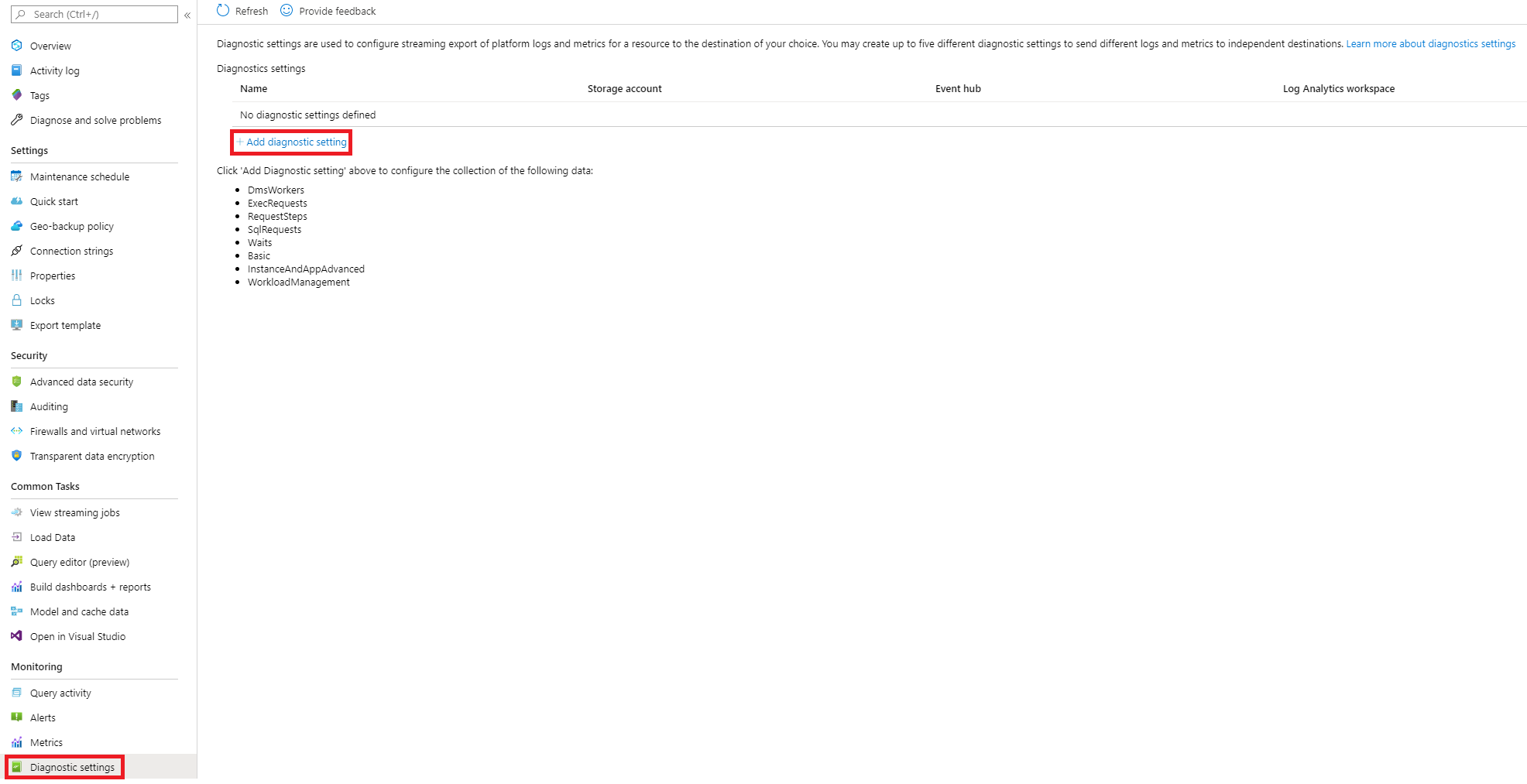
Option B is incorrect because ResourceUsageStats is selected for managed instances

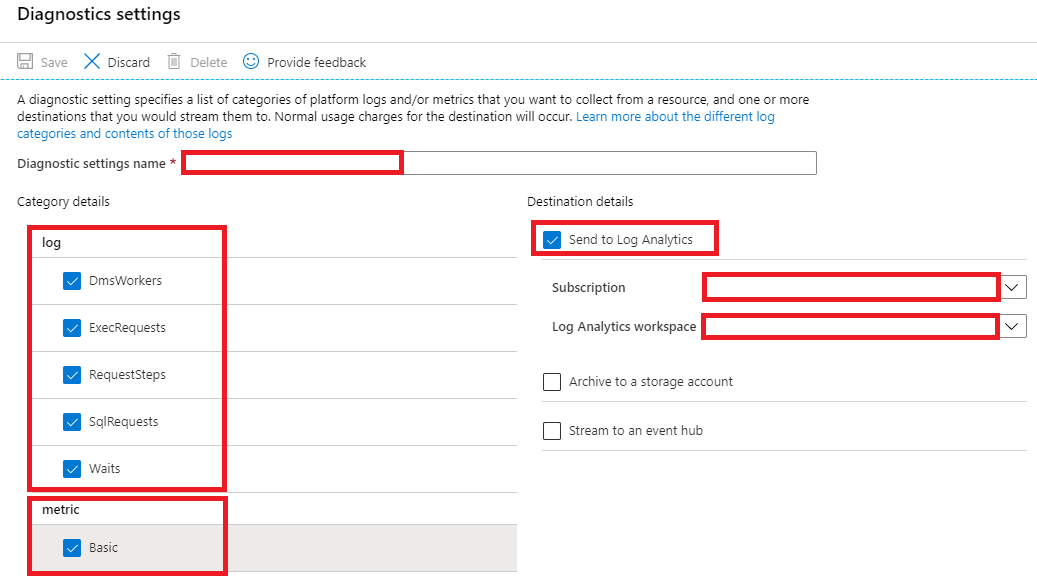
Option C is CORRECT because objective is to send the telemetry details to Log Analytics

Option D is CORRECT because log options need to be selected in order to log  monitoring info

Option E is CORRECT because it is required for advanced one-minute-based monitoring experience

**Diagram:**





**Reference:**

* <https://docs.microsoft.com/en-us/azure/sql-database/sql-database-metrics-diag-logging?tabs=azure-portal>