

My Notes on this section

Exam Perspective

Polybase

PolyBase 6 steps are very important. You should remember all 6 steps in sequence. You will get a question, where you will be given many steps, and you will have to choose the right steps and put those steps in sequence.

Azure SQL Data warehouse

Data distribution is very important. You should have a clear understanding about three distribution methods. Scenarios will be given, and you will be asked to choose the right distribution method.

Geo-replication was removed from the syllabus on July 31, 2020. But even after that questions are coming based on this concept, maybe Microsoft has not updated their test yet.

My Notes

Managed instance deployment

The managed instance deployment option is useful if you have an on-premises SQL Server instance with multiple databases that must all be moved to the cloud.

This deployment option is almost 100% compatible with an on-premises instance

All databases in a managed instance deployment share the same resources.

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

Elastic Pool

An elastic pool allows you to deploy multiple databases to a single logical instance and have all databases share a pool of resources.

All on-premises features are not available with Azure SQL Server or Elastic Pool like You cannot take advantage of CLR features with an elastic pool.

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

Data Sync

Data Sync is a service that lets you synchronize data across multiple Azure SQL Databases and on-premises SQL Server instances bi-directionally.

This is not the preferred solution for disaster recovery scenarios.

Further study: [Sync data across multiple cloud and on-premises databases with SQL Data Sync](#)

Geo-replication

Active geo-replication is a disaster recovery solution for Azure SQL Database that allows replicating a database to another Azure region.

The synchronization direction is only from the master to the replica database, and you only have read access to the replica database.

Further study:

- [Creating and using active geo-replication](#)
- [Sync data across multiple cloud and on-premises databases with SQL Data Sync](#)

Data Migration Assistant (DMA)

DMA is an assessment tool for migrating SQL Server instances to Azure SQL Database.

It evaluates incompatibilities and recommends performance improvements for the target database.

Further study: [Overview of Data Migration Assistant](#)

Azure Database Migration Service.

This is a fully managed service to migrate multiple database sources to Azure with minimal downtime.

Further study: [What is Azure Database Migration Service?](#)

Data Redundancy Strategy

LRS: Replicate data in the same availability zone. This redundancy level will expose the application in case of an Azure outage in a specific availability zone.

ZRS: This redundancy strategy replicates data in other availability zones in the same Azure region.

GRS: This redundancy strategy will replicate data to another Azure region.

Further Study: [Azure Storage redundancy](#)

SQL Server Agent.

- This solution allows you to run administrative tasks in a specific database.
- Only supported in on-premises SQL Server instances or Azure SQL managed instances.

Further Study: [SQL Server Agent](#)

Elastic Database Jobs.

- Allows you to run jobs against all the databases in the elastic pool
- supports PowerShell scripts
- Job definition is stored in a job database
- internal system databases in the elastic pool are configured in a target group

Further study:

- [Automate management tasks using database jobs](#)
- [Create, configure, and manage elastic jobs](#)

Azure Alerts action group

- An action group is a collection of notification preferences used by Azure Monitor to respond to an alert.
- These notification preferences could also include automated remediation actions, like executing an Azure function or running an automated runbook written in PowerShell.
- An action group should be attached to an alert.
- You need to manually create all the connection and target servers logic in a runbook script, which increases administrative efforts.

Further study:

[Create and manage action groups in the Azure portal](#)

Columnstore index: is an in-memory table used in operational analytics.

Azure SQL Data warehouse

SQL Data Warehouse allows you to perform parallel queries using MPP architecture on Big Data

Distribution Method

Replicate: A replicated table is copied across all the compute nodes in a data warehouse. This improves the performance of queries for data in small tables.

Hash: Data is sharded across compute nodes by a column that you specify. Useful for large table

Round Robin - A round-robin distribution shards data evenly. Generally use for staging table

Type	Great fit for...
Replicated	✓ Small-dimension tables in a star schema with less than 2GB of storage after compression (~5x compression).
Round-robin (default)	✓ Temporary/Staging table ✓ No obvious joining key or good candidate column.
Hash	✓ Fact tables. ✓ Large-dimension tables.

Further study:

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

Which table should be used in which scenario?

For a large fact table, you should create a hash-distributed table. Query performance improves when the table is joined with a replicated table or with a table that is distributed on the same column. This avoids data movement.

For a staging table with unknown data, you should create a round-robin distributed table. The data will be evenly distributed across the nodes, and no distribution column needs to be chosen.

For a small dimension table, you should use a replicated table. No data movement is involved when the table is joined with another table on any column.

For a table that has queries that scan a date range, you should create a partitioned table. Partition elimination can improve the performance of the scans when the scanned range is a small part of the table.

Data masking functions are important, you will be given an example, and asked which function you should use to get this desired result. Or if you use this function, what will be the output.

You will see few questions based on **TDE and Always encryption**. Make sure you have good understanding about the difference between these two.

Row-level security is not in syllabus, but you may still see questions based on it.

Deterministic Encryption: Encrypted to the same value every time

Randomized encryption: Encrypted to different value every time

Data Masking

Function	Description
Default	<p>Full masking according to the data types of the designated fields.</p> <p>For string data types, use XXXX or fewer Xs if the size of the field is less than 4 characters (char, nchar, varchar, nvarchar, text, ntext).</p> <p>For numeric data types use a zero value (bigint, bit, decimal, int, money, numeric, smallint, smallmoney, tinyint, float, real).</p> <p>For date and time data types use 01.01.1900 00:00:00.0000000 (date, datetime2, datetime, datetimeoffset, smalldatetime, time).</p> <p>For binary data types use a single byte of ASCII value 0 (binary, varbinary, image).</p>
Email	<p>Masking method that exposes the first letter of an email address and the constant suffix ".com", in the form of an email address.</p> <p><code>aXXX@XXXX.com</code>.</p>
Random	<p>A random masking function for use on any numeric type to mask the original value with a random value within a specified range.</p>
Custom String	<p>Masking method that exposes the first and last letters and adds a custom padding string in the middle. <code>prefix,[padding],suffix</code></p> <p>Note: If the original value is too short to complete the entire mask, part of the prefix or suffix will not be exposed.</p>

References

[Dynamic Data Masking](#)

[SQL Database dynamic data masking](#)

Always encryption

Further reference:

[Always Encrypted \(Database Engine\)](#)

[Always Encrypted \(client development\)](#)

Advance Data Security

Advanced Data Security is a package that helps to find and classify sensitive data.

It also identifies potential database vulnerabilities and anomalous activity, which may indicate a threat to your Azure SQL Database

[Advanced data security for Azure SQL Database](#)

Row-level Security

Good link to read about Row-level security

- <https://www.sqlshack.com/introduction-to-row-level-security-in-sql-server/>

Further References: [Row-Level Security](#)