My Notes on this section

Exam Perspective

Cosmos DB
5 APIs are very important, you will see a few questions based on these

- You will get a scenario, and you will have to choose the best API for that given scenario

- If you found “graph” word in question, and Gremlin API is one of the options, that is the answer

- If SQL Like query is given in question - SQL API is the answer

Partition is the next most important topic, you will definitely see a few questions based on this.

- You will get the scenario and you will have to choose the right partition key.

- That’s why I have explained partition in great detail, please make sure you understand it well

Consistency level – You will see at least a few questions on this topic

- If you see “most recent committed version” in question – Strong consistency is the answer

- And if you see “Lowest Latency” in question – “Eventual consistency” is the answer

Cosmos CLI is also important, you may see 1 or 2 question

- You will see a half-filled query to create a cosmos DB account, and you will have to choose the right option drop-down list to complete this query.

Data Lake

- This is not a very important topic for the exam, You may see 1 question

- If you see “hierarchical namespace” in question, and one of the options is Data Lake, that is probably the right answer.

My Notes

Please note that these are not comprehensive notes, but only based on concepts I see repeatedly asked in the exam.

Blob Storage

- Always use a general-purpose v2 account.

  - This account type incorporates all general-purpose v1 features, including blob storage.

  - It delivers the lowest per-gigabyte capacity prices for Azure storage.

Block Blob Storage

- This is a specialized account type used to store block blobs and appends blobs.
- Low latency
- Higher transaction rates.
- Block blob storage accounts only support premium tiers, which are more expensive than general-purpose v2 account types

Azure Data Lake

- Azure Data Lake is a big data storage solution that allows you to store data of any type and size.
- Repository for Big Data analytics workloads.
- Azure Data Lake Storage Gen2 has the capabilities of both Azure Blob Storage and Azure Data Lake Storage Gen1
- Supports hierarchical namespaces.

Cosmos Table API

- Product with different number of attributes can be saved
- Allows you to use OData and Language Integrated Query (LINQ) to query data.
- You can issue LINQ queries with .NET to query data
- Does not support SQL-like queries from web applications.

- Further study

  o Introduction to Azure Cosmos DB: Table API

Cosmos SQL API

- Allows you to use SQL to query data as JavaScript Object Notation (JSON) documents.
- You can use .NET to query Cosmos DB data that uses the SQL API.
- Supports a schema-less data store.

Mongo API

- This API does not allow you to use SQL-like queries to access and filter data.

Graph API

- This API does not support SQL-like queries.
- This API uses the Gremlin Graph Traversal Language to query data from a graph database.

Table Storage

- Azure Table storage uses NoSQL, which allows you to store keys and attributes in a schema-less data store.
- This is similar to Cosmos DB with the Table API.
- Each entity (row) can store a varying number of attributes (fields). This allows different vendors to upload products with varying attributes.
- You can also use .NET to query the data.
- Further study

  o Azure Table storage overview

Cosmos DB Partition key

- This partition key will distribute all the documents evenly across logical partitions.
- Further Study::

  Create a synthetic partition key
  Partitioning in Azure Cosmos DB

Cosmos DB CLI

- Use GlobalDocumentDB to provision a Cosmos DB with the SQL API

Cosmos DB consistency level

- Strong: This level is guaranteed to return the most recent committed version of a records.
- Eventual:

  o Lowest latency

  o No guarantee of reading operations using the latest committed write.
- Session: the same user is guaranteed to read the same value within a single session.

  o Even before replication occurs, the user that writes the data can read the same value.

  o The user at the same location does not mean, they will be in the same session
- Further study: Consistency levels in Azure Cosmos DB

Automatic failover: This is used to automatically failover Cosmos DB in disaster recovery scenarios.

Shared Access Signature

- SAS delegates access to blob containers in a storage account with granular control over how the client accesses your data.
- You can define a SAS token to allow access only to a specific blob container with a defined expiration.
- Further study:

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

Shared Key authorization (Access Keys)

- Gives full administrative access to storage accounts, sometimes more access than necessary.
- Shared keys could be regenerated
- They do not expire automatically

- Further study:

  o Authorize with Shared Key

Azure Managed Disks - These are virtual hard disks intended to be used as a part of Virtual Machine (VM) related storage.

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]

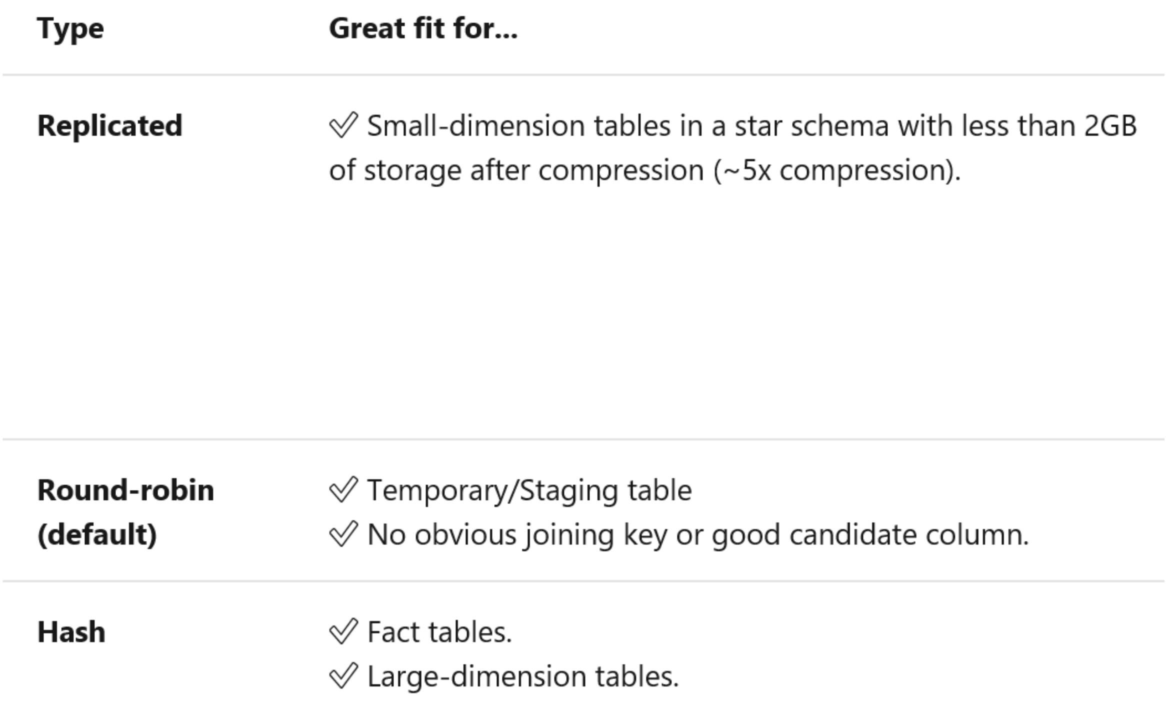
Columnstore index: is an in-memory table used in operational analytics.

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 desire result. Or if you use this function, what will be the output.
You will see few question based on **TDE and Always encryption**. Make sure you have good understanding about 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	Full masking according to the data types of the designated fields. 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**). For numeric data types use a zero value (**bigint**, **bit**, **decimal**, **int**, **money**, **numeric**, **smallint**, **smallmoney**, **tinyint**, **float**, **real**). For date and time data types use 01.01.1900 00:00:00.0000000 (**date**, **datetime2**, **datetime**, **datetimeoffset**, **smalldatetime**, **time**). For binary data types use a single byte of ASCII value 0 (**binary**, **varbinary**, **image**).
Email	Masking method that exposes the first letter of an email address and the constant suffix ".com", in the form of an email address. `aXXX@XXXX.com`.
Random	A random masking function for use on any numeric type to mask the original value with a random value within a specified range.
Custom String	Masking method that exposes the first and last letters and adds a custom padding string in the middle. `prefix,[padding],suffix` Note: If the original value is too short to complete the entire mask, part of the prefix or suffix will not be exposed.

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

My Notes on this section

Exam Perspective

Data Factory Integration runtime is important topic, you should know difference between diff runtime, and in which scenario which runtime should be used.

Self-hosted runtime is more important.

My Notes

Integration Runtime
The integration runtime is the execution environment that provides the compute infrastructure for Data Factory.

Further study: Integration runtime in Azure Data Factory

Self-hosted runtime
When you use the Copy activity to copy data between Azure and a private network, you must use the self-hosted integration runtime.

Further study: Create and configure a self-hosted integration runtime
Azure integration runtime.
This is required when you need to copy data between Azure and public cloud services.

Further study: How to create and configure Azure Integration Runtime
Azure-SSIS integration runtime.
This is required when you want to run existing SSIS packages natively.

Further study: Create Azure-SSIS Integration Runtime in Azure Data Factory

Linked Service
A linked service stores the connection information from the source dataset, like user credentials, server address and database name.

Linked service will be used by the dataset.

Linked services in Azure Data Factory

Activity

An activity is the task that is executed, like copying data or performing a lookup. Activities use datasets to read or write data as the result of a pipeline.

Pipeline

A pipeline is a group of activities linked together to form a data pipeline.

Pipelines and activities in Azure Data Factory
Datasets in Azure Data Factory

Triggers

A tumbling window can define the starting time in the WindowStart setting and the ending time in the WindowEnd setting, defining a time frame to run the data pipeline.

Manual trigger – allow you to manually start pipelines

Schedule trigger – schedule execution of pipeline

Pipeline execution and triggers in Azure Data Factory
Create a trigger that runs a pipeline in response to an event
Create a trigger that runs a pipeline on a schedule
Create a trigger that runs a pipeline on a tumbling window

Databricks

This is an Apache Spark-based technology that allows you to run code in notebooks.

Code can be written in SQL, Python, Scala, and R.

You can have data automatically generate pie charts and bar charts when you run a notebook.

You can override the default language by specifying the language magic command `%<language>` at the beginning of a cell.

My Notes on this Section

Exam Perspective

Windowing functions will be going to dominate this section in the exam.

Also, you should understand the difference between the event hub and IoT hub.

one question can be on Reference data, which can be put on SQL Server or Blob.

My Notes

Stream Analytics is a big data analytics solution that allows you to analyze real-time events simultaneously.
The input data source can be an event hub, an IoT hub, blob storage or SQL Server.

The reference input is data that never or rarely changes. Reference data can be saved in Azure SQL Database or Blob storage

Further reference What is Azure Stream Analytics?

Event Hubs
Event Hub is an Azure resource that allows you to stream big data to the cloud.

Event Hub accepts streaming telemetry data from other sources. It is basically a big data pipeline. It allows you to capture, retain, and replay telemetry data

It accepts streaming data over HTTPS and AMQP.

A Stream Analytics job can read data from Event Hubs and store the transformed data in a variety of output data sources, including Power BI.

IOT Hub
IoT Hub is an Azure resource that allows you to stream big data to the cloud.

It supports per-device provisioning.

It accepts streaming data over HTTPS, AMQP, and Message Queue Telemetry Transport (MQTT).

A Stream Analytics job can read data from IOT Hubs and store the transformed data in a variety of output data sources, including Power BI.

Choosing a real-time message ingestion technology in Azure

Choosing a stream processing technology in Azure
Choose between Azure messaging services - Event Grid, Event Hubs, and Service Bus

Windowing function
Introduction to Stream Analytics windowing functions

Tumbling window
Tumbling windows are a series of fixed-sized, non-overlapping and contiguous time intervals.

Each event is only counted once.

However, they do not check the time duration between events and do not filter out periods of time when no events are streamed.

Tumbling Window (Azure Stream Analytics)

Hopping windows
Hopping windows are a series of fixed-sized and contiguous time intervals. They hop forward by a specified fixed time. If the hop size is less than a size of the window, hopping windows overlap, and that is why an event may be part of several windows.

Hopping windows do not check the time duration between events and do not filter out periods of time when no events are streamed.

Hopping Window (Azure Stream Analytics)

Sliding windows

Sliding windows are a series of fixed-sized and contiguous time intervals. They produce output only when an event occurs, so you can filter out periods of times where no events are streamed.

However, they may overlap and that is why an event may be included in more than one window. Sliding windows also do not check the time duration between events.

Sliding Window (Azure Stream Analytics)

Session windows

Session windows begin when the defect detection event occurs, and they continue to extend, including new events occurring within the set time interval (timeout).

If no further events are detected, then the window will close. The window will also close if the maximum duration parameter is set for the session window, and then a new session window may begin.

The session window option will effectively filter out periods of time where no events are streamed. Each event is only counted once.

Session window (Azure Stream Analytics)

Other Concepts

Event Grid

Event Grid is a publish-subscribe platform for events. Event publishers send the events to Event Grid. Subscribers subscribe to the events they want to handle.

Azure Relay

Azure Relay allows client applications to access on-premises services through Azure.

HDInsight

HDInsight is a streaming technology that allows you to use C#, F#, Java, Python, and Scala.

It does not allow you to use a SQL-like language.

WebJob

WebJob runs in the context of an Azure App Service app.

It can be invoked on a schedule or by a trigger.

You can use C#, Java, Node.js, PHP, Python to implement WebJobs.

However, you cannot use a SQL-like language.

Function App
A function app is similar to a WebJob in that it can be invoked on a schedule or by a trigger.

You can use many different languages to create a function in a function app.

However, you cannot use a SQL-like language.

My Notes on this Section

Exam Perspective

SQL Server monitoring have been removed from syllabus on 31st July, but still questions are coming in exam. So, please make sure you do not ignore this.

My Notes

Log Analytics
Log Analytics allows you to write queries to analyze logs in Azure.

Further reference: Get started with Log Analytics in Azure Monitor

SQL Server Monitoring
Query Performance Insight allows you to view database queries that consume the most resources and those that take the longest to run.
It does not suggest when to create or drop and index.

References:
Query Performance Insight for Azure SQL Database

Azure SQL Database – Diagnostic logging options
SQLInsights gathers performance information and provides recommendations.
QueryStoreRuntimeStatistics provides information about CPU usage and query duration. Basic metrics provides CPU and DTU usage and limits.

QueryStoreWaitStatistics. This provides information about the resources that caused queries to wait, such as the CPU, logs, or locks.

DatabaseWaitStatistics. This provides information about the time a database spent on waiting.
Azure SQL Database metrics and diagnostics logging

SQL Database Advisor
It allows you to review recommendations for creating and dropping indexes, fixing schemas, and parameterizing queries.

References
Find and apply performance recommendations

Azure Advisor: provides recommendations for availability, security, performance, and cost. It integrates with SQL Database Advisor to provide recommendations for creating and dropping indexes.
References:
Improve performance of Azure applications with Azure Advisor
Introduction to Azure Advisor

Query Store - provides statistics on query performance. It helps you identify performance differences that are caused by query changes. It is disabled by default.
References:
Monitoring performance by using the Query Store
Operating the Query Store in Azure SQL Database
Query Store Usage Scenarios

SET SHOWPLAN_TEXT ON
This statement allows you to display query execution information without actually executing the query. This statement is intended for applications that display text.

SET SHOWPLAN_TEXT (Transact-SQL)

SET SHOWPLAN_ALL ON
This statement allows you to display query execution information without actually executing the query. This statement is intended for applications that can display text. It provides additional columns of information for each row that is output.

SET SHOWPLAN_ALL (Transact-SQL)

sys.dm_pdw_exec_requests
This view returns all queries that are currently running or that were recently running. You can use the following SQL statement to return the top 10 longest running queries:

```
SELECT TOP 10 * FROM sys.dm_pdw_exec_requests ORDER BY
total_elapsed_time DESC;
```

Monitor your workload using DMVs

LABEL
LABEL option to assign a comment to the query. This adds a label to the query. For example, you can add the label to a query as follows:

SELECT * FROM FactStoreSales OPTION (LABEL = 'Q4');

You can then easily locate the query's execution steps with the following statement:

SELECT * FROM sys.dm_pdw_exec_requests WHERE [label] = 'Q4';
References:

OPTION Clause (Transact-SQL)

VIEW DATABASE STATE permission.
This permission is required to access Dynamic Management Views (DMVs), which allow you to investigate query execution in Azure SQL Data Warehouse. The view that contains logins is sys.dm_pdw_exec_sessions. It actually contains the last 10,000 logins.

VIEW DEFINITION permission.
This allows the employee to view metadata of an object. For example, the employee can view table metadata in the sys.objects catalog.

ALTER ANY CONNECTION permission
This allows the employee to manage the database server.

ALTER ANY USER permission
This allows the employee to manage database users.

GRANT Database Permissions (Transact-SQL)

My Notes on this Section

Exam Perspective

In this section, mostly you will be asked which metrics you will monitor in particular given situation.

What are diff types of log you will be sending to Log Analytics.
It is good to be aware of commonly used metrics and logs in Diagnostic settings and Metrics page of Monitoring.

My Notes

Ganglia
Default metrics available for Databricks.

How to transfer Databricks logs to Log Analytics/Azure Monitor?
You should use a third-party library to transfer Azure Databricks metrics to Azure Monitor because it is not supported natively at the time of writing.

You should use Azure Log Analytics workspace as the target destination for uploaded Azure Databricks metrics in Azure Monitor. Each workspace has its own data repository, and data sources like Azure Databricks can be configured to store their metrics in a particular Azure Log Analytics workspace.

Monitoring Azure Databricks

Solution for sending application metrics to Azure Monitor?
Dropwizard is a Java library. Spark, which is the cluster engine that is used to run Databricks, uses a configurable metrics system that is based on the Dropwizard Metrics Library.

Diagnostic logs
You should configure diagnostics logs to send data to a blob storage account.

By default, Azure Data Pipeline stores run-data for only 45 days. To store the data longer than that, you must configure diagnostics logs. With diagnostics logs, you can choose to store the run-data in a blob storage account, an event hub, or a Log Analytics workspace.

You can query using KQL in the Log Analytics workspace tables.

ADFPipelineRun table contains rows for status changes like InProgress and Succeeded.
AzureMetrics contains metrics like PipelineSucceededRuns

Azure Data Factory inbuilt monitoring
Azure Data Factory includes monitoring capabilities for your pipeline runs with execution metrics and pipeline status. You can define alerts directly in Azure Data Factory Monitor.

However, Azure Data Factory data retention is limited to 45 days. You need to use Azure Monitor for longer retention.

Azure Stream analytics – Metrics to monitor
SU monitoring
You should use the SU % utilization metric. This metric indicates how much of the provisioned SU % is in use. If this indicator reaches 80%, there is a high probability that the job will stop.

Input events metric - This metric counts the number of records deserialized from the input events.
Runtime errors metric - This metric is the total number of errors during query processing.
Understand Stream Analytics job monitoring and how to monitor queries

Azure Network Watcher
It is a centralized tool for monitoring Azure networking.

My Notes on this Section

Exam Perspective

SQL Server optimization has been removed from the syllabus on 31st July 2020, but you can still see questions based on this topic.

There are many questions in this section that will check once again your Azure SQL Data warehouse distribution knowledge, if you understand the difference between Round-robin, hash, and replicated.

You will see one question on TTL And maybe 1 or 2 questions on the access tier, both topics are very easy, you should not miss these questions.

My Notes

Time to live TTL, Azure Cosmos DB provides the ability to delete items automatically from a container after a certain time period.

By default, you can set time to live at the container level and override the value on a per-item basis.

After you set the TTL at a container or at an item level, Azure Cosmos DB will automatically remove these items after the time period, since the time they were last modified.

Time to live value is configured in seconds.

Examples

This section shows some examples with different time to live values assigned to container and items:

Example 1

TTL on container is set to null (DefaultTimeToLive = null)

TTL on item	Result
ttl = null	TTL is disabled. The item will never expire (default).
ttl = -1	TTL is disabled. The item will never expire.
ttl = 2000	TTL is disabled. The item will never expire.

Example 2

TTL on container is set to -1 (DefaultTimeToLive = -1)

TTL on item	Result
ttl = null	TTL is enabled. The item will never expire (default).
ttl = -1	TTL is enabled. The item will never expire.
ttl = 2000	TTL is enabled. The item will expire after 2000 seconds.

Example 3

TTL on container is set to 1000 (DefaultTimeToLive = 1000)

TTL on item	Result
ttl = null	TTL is enabled. The item will expire after 1000 seconds (default).
ttl = -1	TTL is enabled. The item will never expire.
ttl = 2000	TTL is enabled. The item will expire after 2000 seconds.

Time to Live (TTL) in Azure Cosmos DB
Configure time to live in Azure Cosmos DB

Stream analytics best practices

You should start with six SUs for queries that do not use PARTITION BY. This is considered a best practice.

You should allocate more SUs than you need. This is another best practice.

You should keep the SU metric below 80 percent. This allows Streaming Analytics to account for usage spikes
Understand and adjust Streaming Units

ExpressRoute.
This creates a dedicated link between your on-premises datacenter and Azure. This improves performance when copying data to Azure.

Tuning Azure Data Lake Storage Gen2 for performance

AD Connect allows you to synchronize user accounts between on-premises AD and Azure AD.

Access Tiers
The **Archive tier** is optimized for storing data that is rarely accessed and that is kept for at least 180 days.

The **Cool tier** is optimized for storing data that is accessed infrequently and that is kept for at least 30 days.

The **Hot tier** is optimized for storing data that is accessed frequently.
Azure Blob storage: hot, cool, and archive access tiers

Azure SQL Server
A columnstore index stores column values and increases the aggregate queries that use theses indexes.
Memory-optimized tables store all data and schema in memory, increasing the performance for queries.
Partitioned view can be used to split large tables across multiple smaller tables.
Non-clustered index is generally used to increase filter performance and to lookup rows with specific values.
A heap is a table without a clustered index with table data stored without any specific order. Every query in a heap should perform a table scan. A full table scan has better performance than a table with a clustered index for small tables.

Azure SQL Data Warehouse

Fasted loading process is Polybase.

You can write and run PolyBase T-SQL commands. This is a fully parallel operation and is the fastest option.

You can also use a Copy Activity in Azure Data Factory with the copy method set to PolyBase. This option creates and executes the Polybase commands automatically. This also offers the fastest performance.

The following options are not parallel operations and are thus slower:

- BCP
- SQL BulkCopy API
- SSIS
- Azure Data Factory using a Copy Activity and the bulk insert option

References

Data loading strategies for data warehousing
Tutorial: Load the New York Taxicab dataset