Applying the Azure Well-Architected Framework to Azure Synapse Analytics

Andy Cutler

November 2021

Andy Cutler

Independent BI/DW Consultant/Contractor

Azure Data Platform & Power BI

datahai.co.uk/blog

serverlesssql.com

twitter.com/MrAndyCutler

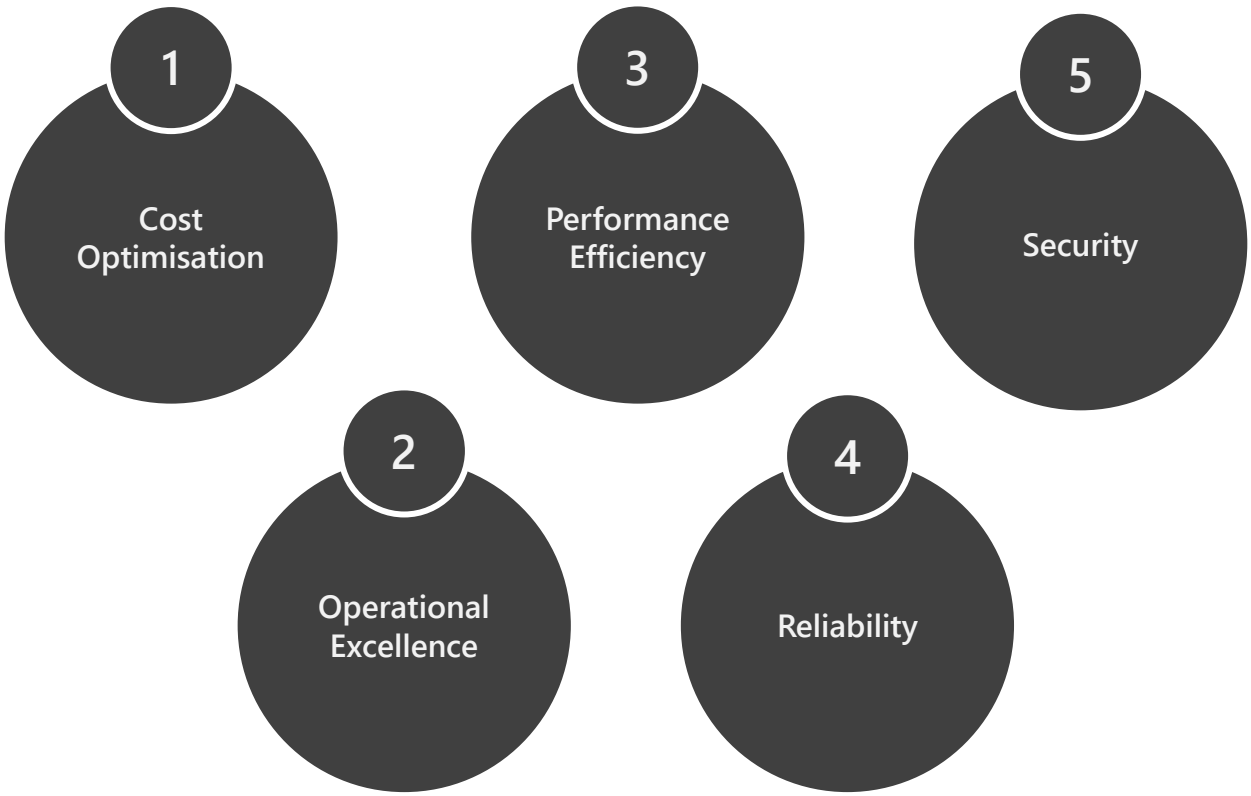
linkedin.com/in/andycutler/

Microsoft Partner | Silver Data Analytics, Silver Data Platform

MVP Microsoft® Most Valuable Professional

SCAN ME

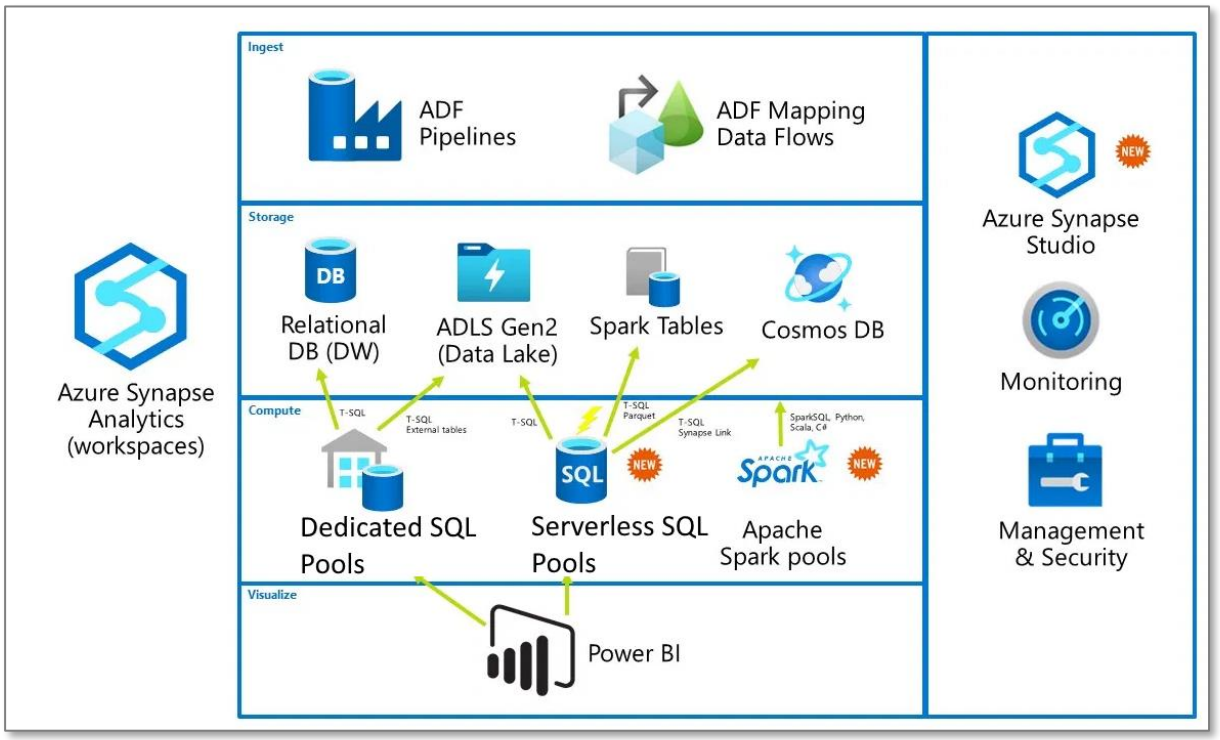
The Microsoft Azure Well-Architected Framework is a **set of 5 pillars** which can be used to help an organisation **improve the quality** of their Azure workloads and infrastructure.



We can apply settings and features from the **Dedicated SQL Pools** and **Serverless SQL Pools** to these pillars.

Synapse Analytics Components

Synapse is a set of services that have been brought together into a unified platform .



- Dedicated SQL Pools
- Serverless SQL Pools
- Spark Pools
- Pipelines (Data Factory)
- Data Explorer Pools
- Power BI integration

Who is the Framework for?

The framework is just as relevant to a large multi-national organisations infrastructure as to a Small/Medium Sized business, and a single person with their own infrastructure.

Who really likes:

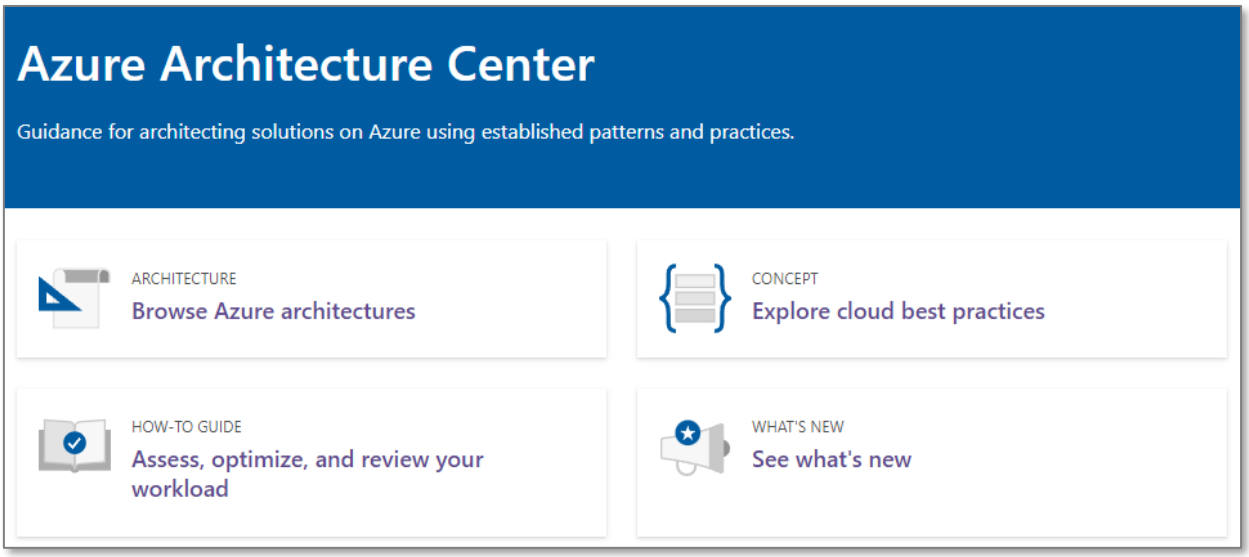
- Spending too much money
- Spending time administrating your infrastructure (unless that is your role!)
- Spending time watching……things….progress
- Deployments and code breaking (unless intentionally!)
- Data being stolen

There is cross-over in terms of features and pillars.

E.G. a feature that helps with cost optimisation can also help with Performance Efficiency.

Azure Architecture Centre

The Azure Architecture Centre is part of the overall Microsoft Documentation resource online .

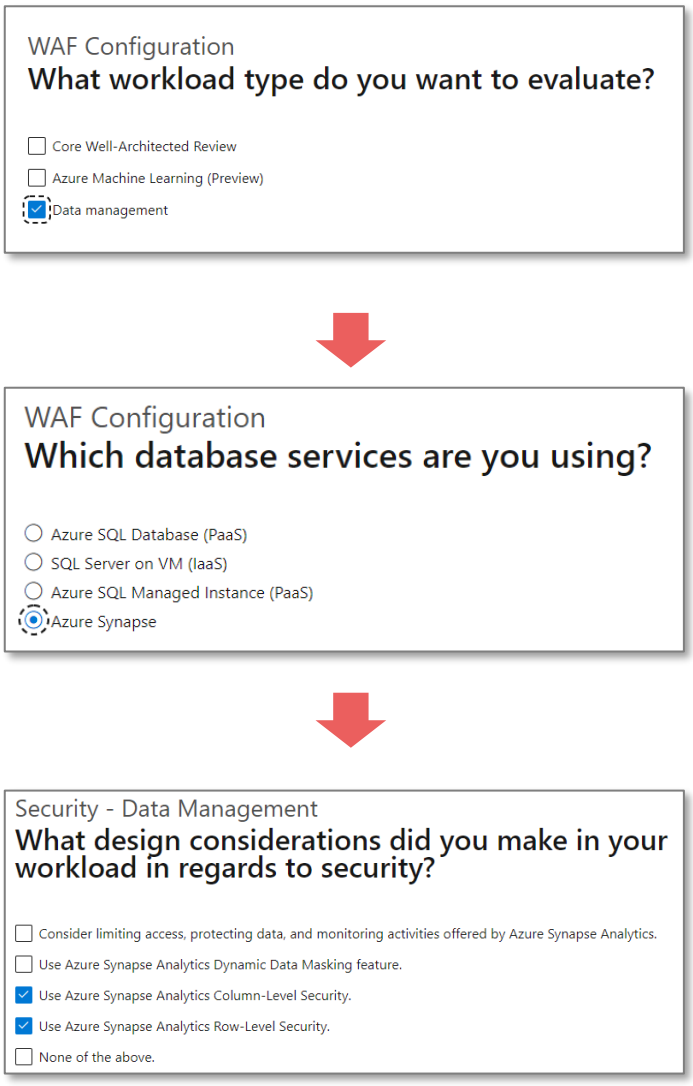


Within the Architecture Centre are areas which look at:

- Reference Architectures
- Well-Architected Framework
- Best Practices

https://docs.microsoft.com/en-us/azure/architecture/

Assessment



Within the Azure Architecture Centre is an Assessment process which is a series of questions across the 5 pillars.

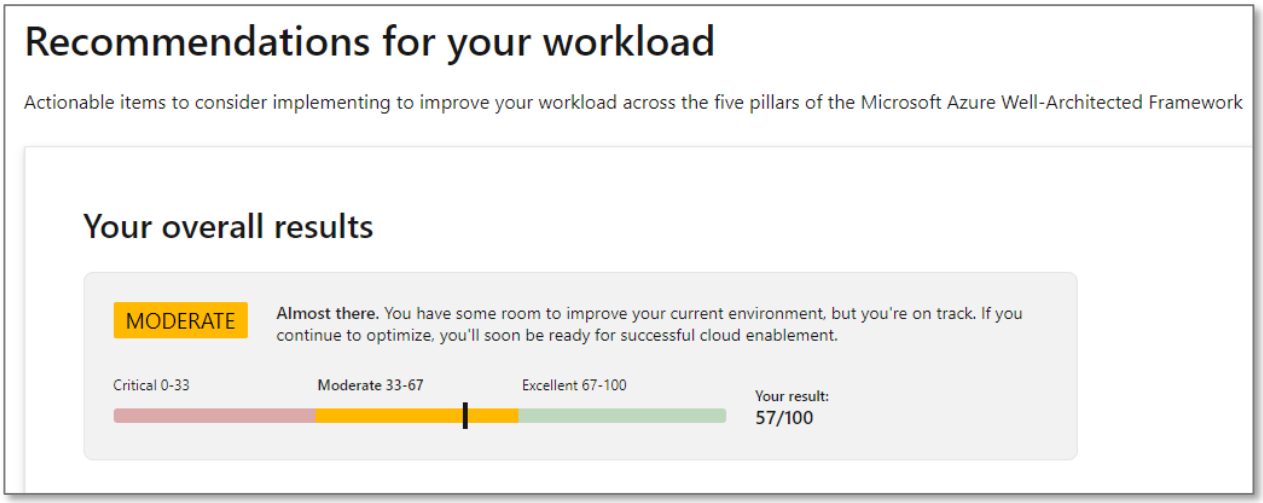
You can choose which pillars to assess.

For each question you are given a series of checkboxes.

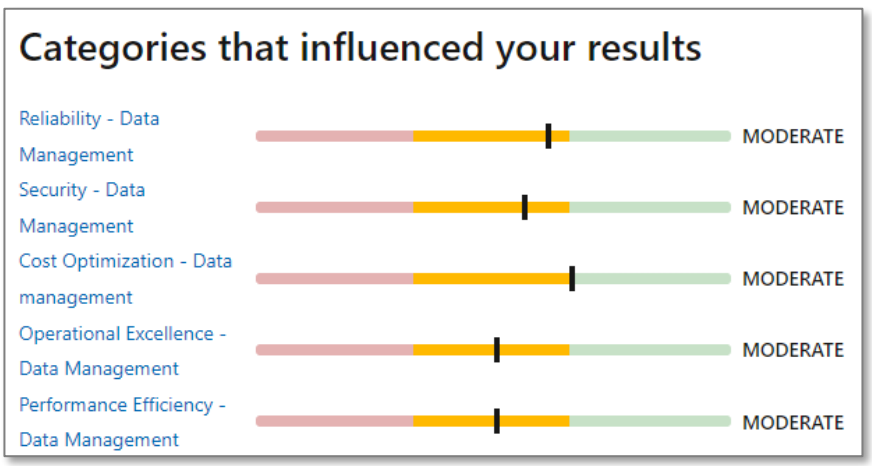
Pillar	Number of Questions
Reliability	12
Security	7
Cost Optimisation	6
Operational Excellence	6
Performance Efficiency	6

Assessment Score

Once the assessment is complete you are given an overall score.



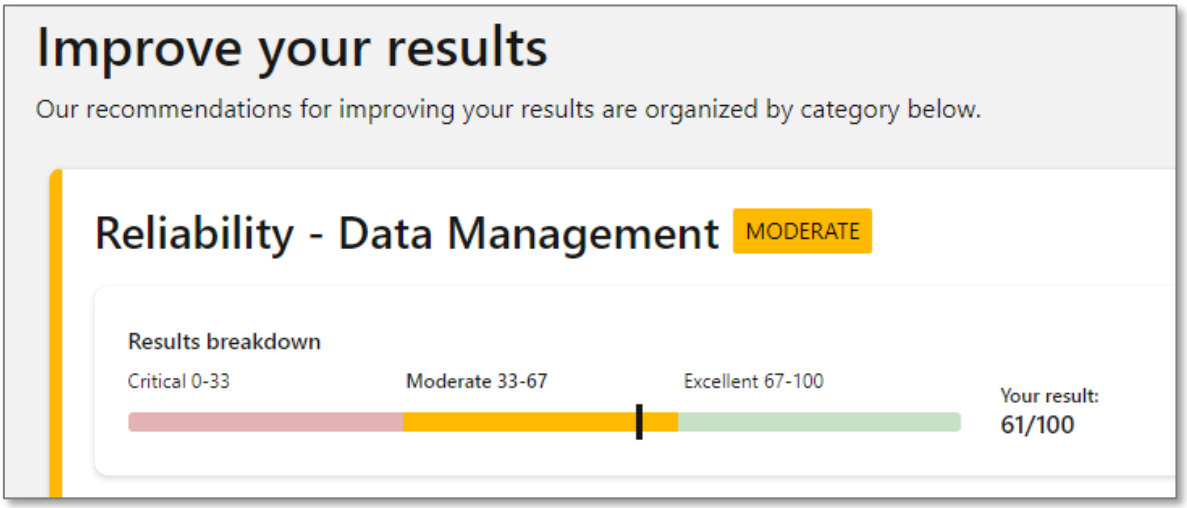
Depending on which pillars (categories) were selected at the beginning of the assessment will dictate the influence on the overall score.



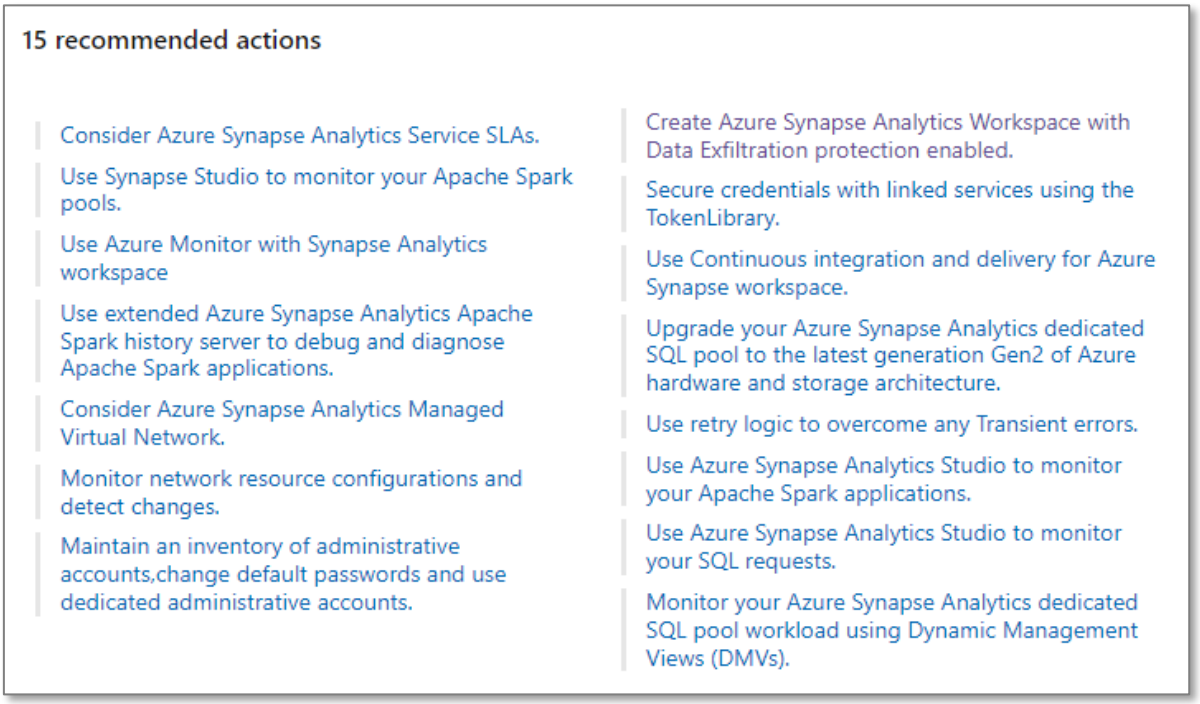
You are able to re-take the assessment at any time once you have addressed an area.

Assessment Breakdown & Resources

To address any concerns across the pillars (categories) each pillar provides a set of resources to review.



For each pillar there are recommended actions to take.



1 Cost Optimisation

In terms of optimising costs we are looking to maximize the value delivered whilst keeping costs as low as possible.

Dedicated SQL Pools

- Assign compute using **DWUs** from DWU100 (60GB) to 30000 (18TB) to suit workload. Potentially use a higher DWU to load data then lower the DWU to query data during the day.
- Scale **Up** and **Down**, and also **Pause** using Azure Automation, Functions, Pipelines.
- 1 & 3 Year Reserve Pricing per DWU100 can **reduce pricing by up to 65%.**

Serverless SQL Pools

- The cost is based on **Data Processed** on-demand which includes both **Reading** data from and **Writing** data to external storage.
- Recommended practice is to **optimise Data Types** and where possible **use Parquet.**
- Data Processed can be seen in Synapse Studio and also in system views
  - Amount of Data Processed
  - Daily, Weekly, and Monthly Limits set

Azure Pricing Calculator includes Dedicated SQL Pools, Serverless SQL Pools, Pipelines, and Spark .

Dedicated SQL Pools DWUs Per Hour

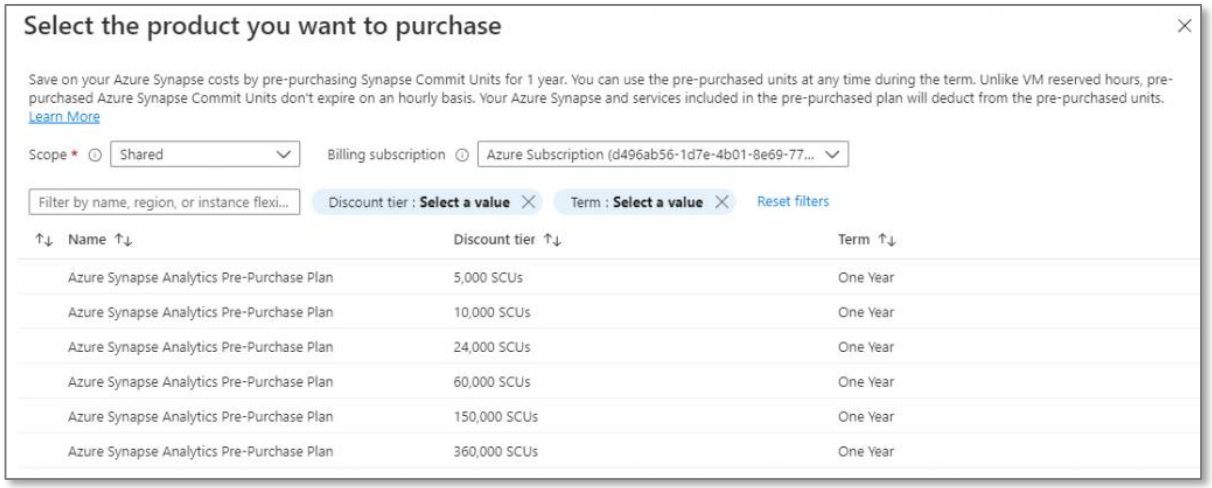
If we set the DWU size to a value and never change it then it may not be optimised for cost.

If analyse our expected workload and match the DWU accordingly then we can optimise our costs.

Pre-Purchase Synapse Commit Units

Microsoft now offer a pre-purchase plan called SCUs (Synapse Commit Units). The units can be used across services within the Synapse eco-system:

- Azure Synapse Analytics Dedicated SQL Pool
- Azure Synapse Analytics Managed VNET
- Azure Synapse Analytics Pipelines
- Azure Synapse Analytics Serverless SQL Pool
- Azure Synapse Analytics Serverless Apache Spark Pool - Memory Optimized
- Azure Synapse Analytics Data Flows



1 SCU = 1 of the currency being charged. E.G. in the UK, 5000 SCUs = 5000 GBP. This is purchased at a discounted rate.

The SCUs can be used at any point within a 12 month period.

Savings start at ~11%

2 Operational Excellence

Ensuring that we have operational processes that keep a system updated and running in production.

Dedicated SQL Pools

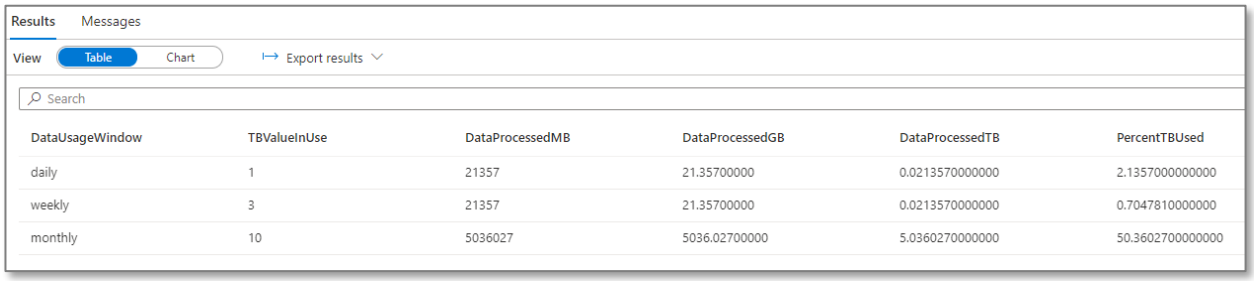
- Database projects are supported and can be created using SSDT in Visual Studio and Azure Data Studio. A full CI/CD process can then be created using Azure DevOps integration to deploy code changes.
- Enable logging using **Azure Log Analytics.**
- Use **Azure Monitor** to surface alerts and metrics to monitor DWU usage, Cache usage, RAM utilisation, and CPU usage.

Serverless SQL Pools

- The is currently no SSDT support, however SQL Scripts to create objects can be source controlled using **Synapse Studio Source Control integration**.
- System Views to track Data Processed volumes by **Day, Week, and Month**. If limits have been set, monitoring data processed usage is vital.
- **Azure Monitor** can also surface metrics to monitor data processed volumes, plus login attempts, and ended SQL requests.

Monitoring Data Processed

An example of monitoring data processed and Daily, Weekly, Monthly limits set:



Results | Messages

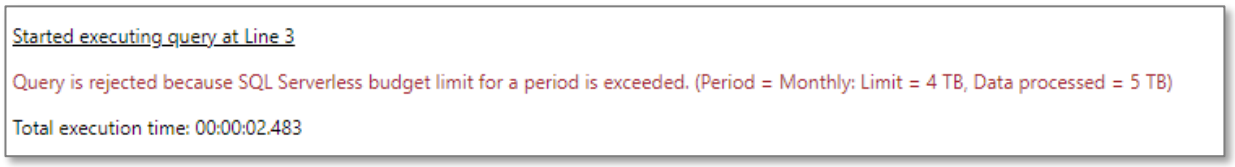
View: Table | Chart — Export results

Search

DataUsageWindow	TBValueInUse	DataProcessedMB	DataProcessedGB	DataProcessedTB	PercentTBUsed
daily	1	21357	21.35700000	0.0213570000000	2.1357000000000
weekly	3	21357	21.35700000	0.0213570000000	0.7047810000000
monthly	10	5036027	5036.02700000	5.0360270000000	50.3602700000000

In this example, we have currently used 50% of the capped 10TB limit.

Changing the monthly cap to 4TB now means no more queries can be executed…



Started executing query at Line 3

Query is rejected because SQL Serverless budget limit for a period is exceeded. (Period = Monthly: Limit = 4 TB, Data processed = 5 TB)

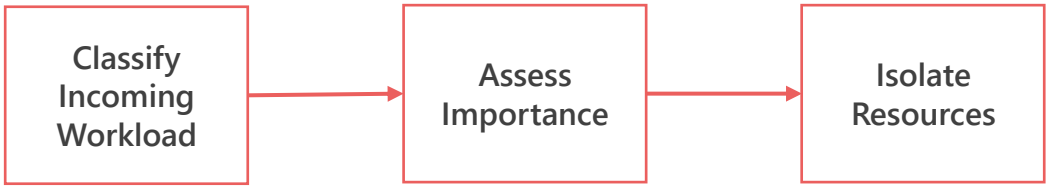
Total execution time: 00:00:02.483

3 Performance Efficiency

The ability of a system to adapt to changes in load and ensure performant operations such as data loading and delivery.

Dedicated SQL Pools

- Currently auto-scaling DWUs "online" is not supported so resizing compute is an offline process.
- The **Workload Classification, Importance, and Isolation** feature allows certain workload sizes to be allocated the required resources.



- There are a set of best practices which include **data loading strategies**, **table distribution** & **partitioning strategies.**

Serverless SQL Pools

- With the Polaris engine there is no need to set any workload settings or classification as the service will **scale** and **allocate resources** during workload execution.
- Consider using Serverless SQL Pools to query data external to the Synapse ecosystem rather than loading/transforming to another data store.
- There are a set of best practices which include **data type optimisation**, **storage co-location**, and using partition pruning functions such as **filepath** & **filename.**

4 Reliability

The ability of a system to recover from failures and return to operation.

Dedicated SQL Pools

- When the Dedicated SQL Pool is running, **Automatic Restore Points** are created periodically during the day and are available for **7 days**. You can also create **User-Defined Restore Points** if you regularly scale/pause the service (limited to 7 Days retention).
- You can enable/disable **Geo-Backup** which will backup to a paired region, E.G. UK South and UK West.
- Enabling Threat Protection to detect anomalous activities which impact security but also workload operations.
- Consider using Managed Virtual Network & Private Endpoints to control and isolate traffic.

Serverless SQL Pools

- **Automatic Fault tolerance** in Polaris engine with an automated query restart process. "Tasks" will be restarted automatically in the event of a failure and this is a seamless process to the user/query executor.
- Consider co-locating Storage Account within the same region as Synapse Analytics.

5 Security

Protecting applications and data from threats and malicious activity from any direction.

Dedicated SQL Pools

- Use Azure Active Directory Groups and Users, MFA, to secure access to databases, database objects and data.
- Enable **Transparent Data Encryption** (TDE) to encrypt data at rest and use your own keys.
- Enable **Microsoft Defender for SQL** with periodic Vulnerability Assessment scans to determine current security state.
- **Managed Vnets, Firewall rules** and **Private Endpoints** can be used to allow access to/from Synapse Analytics (Serverless too).
- **Dynamic Data Masking**, **Column-Level,** and **Row-Level Security** are available.
- **Advanced Threat Protection** for detecting and alerting malicious activity.

Serverless SQL Pools

- As with Dedicated SQL Pools we can use Azure Active Directory Groups and Users to secure database objects such as **Views** and **External Tables**.
- Access to the data itself is based on Azure Storage permissions:
  - **AAD Group/User**
  - **SAS credentials**
  - **Managed Identity**
  - **Database Key (CosmosDB)**

References

Document	Link
Microsoft Azure Well-Architected Framework	https://docs.microsoft.com/en-us/azure/architecture/framework/
Azure Well-Architected Review	https://docs.microsoft.com/en-us/assessments/?mode=pre-assessment
Azure Advisor	https://docs.microsoft.com/en-us/azure/advisor/
Azure Advisor Score	https://docs.microsoft.com/en-us/azure/advisor/azure-advisor-score