Synapse Workspace for Data Asset Inventory

Classification: ---UNDISCLOSED-----

Revision History

Revision	Author	Reviewer	Date
1	Architects		04/11/2022
2			

Table of Contents

Revision History 2
Overview and Scope 4
Solution Design Overview 5
Overview Design ..... **Error! Bookmark not defined.**
ESA Domains & Controls 5
Network Security: *9*
Identity Management *10*
Privileged Access Management *12*
Data Protection: *12*
Asset Management ..... ***Error! Bookmark not defined.***
Monitoring & Detection – Logging, Monitoring and Threat Detection *14*
Incident Management ..... ***Error! Bookmark not defined.***
Vulnerability Management *14*
Endpoint Security ..... ***Error! Bookmark not defined.***
Backup and Recovery *15*
Governance & Strategy ..... ***Error! Bookmark not defined.***
Residual Risks 15
Appendix 15
Reference Material 16

Overview and Scope

The purpose of this document is to define the high-level Business objective and proposed technical solution for enabling Azure Synapse Analytics Workspace, connected to enterprise Data Lakes, Power BI, and others.

Azure Synapse is an enterprise analytics service that accelerates time to insight across data warehouses and big data systems. Azure Synapse brings together the best of SQL technologies used in enterprise data warehousing, Spark technologies used for big data, Pipelines for data integration and ETL/ELT, and deep integration with other Azure services such as Power BI, Cosmos DB, and Azure ML.

For ---UNDISCLOSED-----, Azure Synapse Analytics Workspace will provide one single central user-friendly solution platform to maintain and easily access all main Azure Resources that the team harnessing the power of Server-less Pools, ADL and ADF (including Power BI for data visualizations). As of now, our team uses many different applications to connect to those resources. This project aims to address Reporting and Monitoring challenges for ---UNDISCLOSED----- Cloud allowing for better insights on data, resource usage, monitoring, and cost optimization. Inventory Reports provide an improved understanding of data and perform impact analysis, and the project will equip ---UNDISCLOSED-----with the means necessary by creating an Azure Synapse Workspace and enabling the native Azure Blob Inventory feature for the Enterprise data lakes.

Data classification is Official Use Only.

This project is to establish a centralized, single-IDE environment for automation and reporting, where ---UNDISCLOSED-----team will use data store, computing, and integration pipelines to collect, transform, store, and report on data generated by data platforms. Scope of the project is the Corporate Lake (part of the Enterprise Data Lake strategy), Data classification is CSC.

Azure Synapse will be used as the single plain for ---UNDISCLOSED----- Reporting and Automation for internal use. The platform will integrate it with the existing ITSDA managed ecosystem providing insight on data access security and enable a cost-effective server-less compute-based approach to report automation allowing for seamless and secure integration of data sources and Power BI and a single plain for Reporting and Automation. To maintain a consolidated view in big ---UNDISCLOSED-----it is essential to have the ability to execute automated pipelines and notebooks. Pipeline functionality also enables development and configuration of alerting mechanisms for proactive and reactive monitoring of data platform resources.

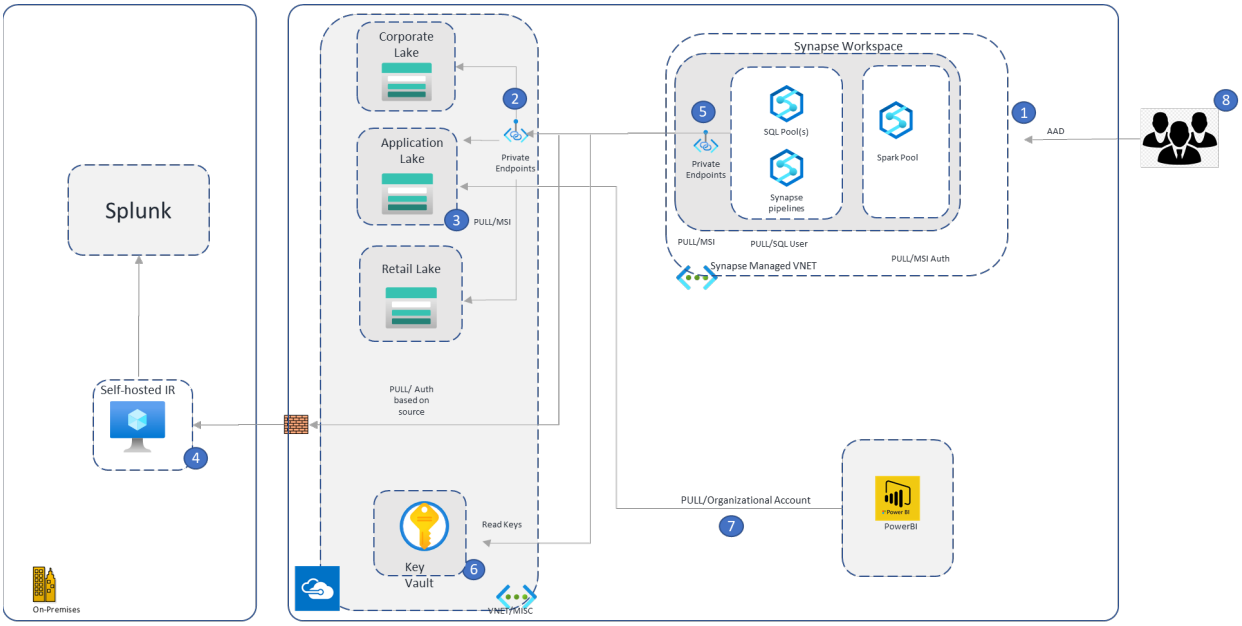
Azure Blob Inventory provides automatic and comprehensive inventory reporting for storage accounts including various attributes including Total Data Size, Age, Access Control, Encryption status, Immutability policy and others. Inventory Reporting has been a major challenge with the Corporate Lakes due to the sheer size of the data and the millions of objects, making it nearly impossible to produce an end-to-end picture of the data inside. Azure Blob Inventory integrated and analyzed with Azure Synapse addresses the immediate reporting needs of ---UNDISCLOSED-----.

On the long term the project aims to establish a ---UNDISCLOSED-----Intelligence Platform for Analytics and near real-time reporting Integrating metrics from sources integrating data from multiple sources Azure Monitor, Application Insights through ---UNDISCLOSED-----and other tools allowing near real time reporting capability. The Project goes in parallel with Synapse Workspace Enablement and related projects incorporating the latest accreditations and practices.

Success Criteria:

- Centralized IDE for Data for ---UNDISCLOSED----- Reporting and Automation
- Integration of **---UNDISCLOSED-----** usage metrics
- Improved Tracking for ---UNDISCLOSED----- Access Control
- ---UNDISCLOSED----- Security Controls
- Acceptable operational services

Solution Design Overview



The approach solves the immediate ---UNDISCLOSED-----Inventory Reporting needs. The required metrics will be retrieved from ---UNDISCLOSED-----and integrated with Inventory data in Synapse. Initially, reports are not expected to produce a lot of traffic and **Synapse Server-less pools** will be used as much as possible. We are not considering Dedicated pools or relational data store at this stage and requirement will be identified in the development process.

Solution components:

1. **New Separate Synapse workspace will be created** for ITSDA Data Asset Inventory Reporting:
2. **Azure Blob Inventory will be Enabled** for all Enterprise Data Lakes.
    - Inventory Reports will be stored in a dedicated container and contain ---UNDISCLOSED-----data.
    - Rules and Filters will be defined for appropriate frequency and consistency of the data across data lakes.
3. **Synapse Pipelines will:**
    - Extract and Process Inventory reports from EDL Stores using MSI.
    - Collect Resource details from Azure ARM REST API.
    - Collect Audit and Access details from ---UNDISCLOSED-----REST API.

4. **Self-Hosted Integration Runtime** will act as a Reverse Proxy for **---UNDISCLOSED-----access.**
5. **Synapse Notebooks and Scripts** will be used to analyze data.

Synapse Workspace:

- **Workspace will use** the Application Lake ---UNDISCLOSED-----as the primary storage container**.**
- **Spark and Server-less pool** considered for the initial setup.
- **Managed Private Endpoints** to the Enterprise Data ---UNDISCLOSED-----
- **Private endpoint connection** to the Synapse Serverless pool.
- **---UNDISCLOSED-----access** via **Self-Hosted Integration Runtime.**
- **Azure ARM REST API** will be used for resource and usage details and metrics
- **Key design decision is to use Azure AD groups to maintain access** to any of the artefacts within the Synapse Workspace.
- **Existing Custom RBAC and Security Controls** based on ---UNDISCLOSED-----

Backup and Restore

The Synapse Workspace will use the Retail Data Lake Storage Accounts for its primary data storage in a dedicated container. The Built-In Server-less pool will use external tables to store processed data in the Data Lake Retail storage account, and the existing Data Lake Backup and DR strategies will be used.

Inventory data can be regenerated from data sources. The required configuration will be backed up in the Data Lake using Pipelines to allow for easy transfer to a new Synapse workspace should it be necessary. The initial approach for recovery use manual recreation of the Synapse Workspace, but with the DevSecOps Integration of Synapse, planned ---UNDISCLOSED-----, Backup and DR options will further improve with version control and pipeline deployment.

Encryption and key/secret management

Encryption using Customer Managed Keys will be used when creating the Synapse workspace. The Solution will integrate a Key Vault, that is being secured with Private Link and FW restrictions. All secrets for Service Principals or SQL Logins will be stored in the Key Vault. All Linked services will use Key Vault to authenticate underlying data lake and on-prem resources (if applicable).

Segregation of Duties

Segregation of Duties will follow the already established model for Synapse Workspace Enablement:

Activity	Responsible
Create Separate Synapse Workspace per use case	---UNDISCLOSED------
Create Synapse Spark Pools and Dedicated Pools	---UNDISCLOSED------

Enable Blob Inventory on Storage Accounts	---UNDISCLOSED-----
Define Blob Inventory Rules	---UNDISCLOSED-----
Pause Synapse Dedicated pools	---UNDISCLOSED-----
Install/manage Synapse packages	---UNDISCLOSED-----
Configure VNETs, Firewall rules, and Network Pre-requisites for Integration Runtimes	Network Admin / ---UNDISCLOSED-----
Create Integration Runtimes in Synapse	---UNDISCLOSED-----
Create credentials, assign workspace item permissions, provide access to developers	---UNDISCLOSED-----
Develop pipelines, notebooks, dedicated pool tables/objects	---UNDISCLOSED------
Promote code between environments (no delete permissions in PROD)	---UNDISCLOSED------ Lead
Create dedicated pool level system objects (credentials, external tables, data sources)	---UNDISCLOSED-----

Scope for this accreditation

- Azure Synapse Workspace with Server-less SQL and Spark Pool
- Azure Blob Storage Inventory
- Synapse Self-hosted Integration Runtime

Components out of Scope for review

- Storage accounts
- Power BI

Authentication/Authorization

SRN	Component	Description	Security Control
1	Azure Synapse Workspace	---UNDISCLOSED-----to access Azure Synapse Workspace using strong AAD authentication with MFA.	AAD with 2FA Encryption using Customer Managed Keys

2	Azure Data Lake	Using Applications lake for Primary Storage. Other Storage accounts will be access as Linked services for integration scenarios (ETL/ELT jobs)	Synapse Workspace will use MSI; Linked Storage access using Managed Private Endpoints. Access will be limited to Blob Inventory Reports Only.
3	Azure Blob Inventory	Storage Account Feature used to generate Inventory Reports on Azure Data Lakes. Blob Inventory will use dedicated container as target for Inventory Reports which will be processed by Synapse.	Synapse Workspace will use MSI; Linked Storage access using Managed Private Endpoints. Access will be limited to Blob Inventory Reports Only.
4	Self-hosted integration runtimes	Installed on Windows VM inside WBG Premises to connect and move data from internal data sources. Synapse use Self IR perform data movement operations using the credentials encrypted and stored locally on Windows VM.	Authentication key is generated by Synapse. This Authentication key is used to register Self Integration Runtime environment. Credentials are encrypted and stored using Windows DPAPI. Multiple nodes are required for HA. Only outbound TCP communication on port 1433 for selected targets.
5	Synapse Pipelines	Data will be moved from source to target using the Synapse Pipelines Synapse will use self IR to access On-prem data sources and push it to Data lake. The DL would have a data structure to absorb the initial (landing) files in the Synapse primary storage container and once validated the files would flow to an invalid (in case of errors) or valid subfolder allowing for further manipulation. New BLOB created on the Valid subfolder would trigger the rest	Using MSI wherever applicable and possible

		of the process in ADF (Synapse).	

Previous accreditation

---UNDISCLOSED-----

ESA Domains & Controls

Cloud_Service_Security_Controls_v202106

Network Security:

NS-1: Protect Azure resources within virtual networks

Use managed virtual networks when you deploy Azure Synapse Workspace. Synapse workspace will have Private Links enabled

Integration Runtimes Installed on Windows VM inside WBG Premises to connect and move data from internal data sources. Integration runtime will have outbound tcp 1433

Set a minimal TLS version 1.2 ensures that newer TLS versions are supported.

NS-2: Establish private network access to cloud services

Use Managed private endpoints, traffic between your Azure Synapse workspace and other Azure resources.

Limit outbound traffic from your workspace to only approved targets. You must create Managed private endpoints to these targets.

Use Private endpoint connection to the Synapse Serverless pool

Provide ---UNDISCLOSED-----access via Self-Hosted Integration Runtime

Use Azure Virtual Network service endpoints or secure endpoints when private link/private endpoint is not available.

Note (Below are not controls just notes):
One common issue people facing is their Spark pools not being able to read files on the storage account. This is because you need to manually create a managed service endpoint the storage account.

Where managed private endpoints allow the workspace to connect to other PaaS services outside of its managed virtual network, private endpoint connections allow for everyone and everything to connect to Synapse endpoints using a private endpoint.

Use Azure Virtual Network service endpoints to provide secure access to Azure Synapse when private link/private endpoint is not available.

NS-4: Enable data exfiltration Protection

Enable data exfiltration protection for workspaces. Access storage accounts through managed private endpoint. The storage account owner will need to approve the connection request to establish the private link.

Identity Management

IM-1: Use Azure Active Directory as the central identity and authentication system

Use Azure AD with strong authentication (mfa) for ---UNDISCLOSED-----accessing synapse workspaces

Use Azure AD groups to maintain access within the Synapse Workspace.

Implement appropriate authentication mechanism based on the WBG tiered authentication framework for systems that process personal data.

IM-2: Manage application identities securely and automatically

Use managed identities for Azure Synapse workspaces. Use private endpoints for Linked storage access

The Self Hosted integration runtimes Authentication key is generated by Synapse. This Authentication key is used to register Self Integration Runtime environment. Credentials are encrypted and stored using Windows DPAPI.

Store All secrets for Service Principals or SQL Logins in KeyVault. All Linked services will use KeyVault to authenticate underlying datalake and on-prem resources (if applicable)

IM-3: Use dedicated admin accounts and strong authentication controls for admin activities

Segregation of Duties will follow the already established model for Synapse Workspace Enablement:
Rbac fill be created as per table below which is in the PA-2 control. (PA2=IM3)

Activity	**Responsible**
Create Separate Synapse Workspace per use case	---UNDISCLOSED----
Create Synapse Spark Pools and Dedicated Pools	---UNDISCLOSED----
Enable Blob Inventory on Storage Accounts	---UNDISCLOSED----
Define Blob Inventory Rules	---UNDISCLOSED----
Pause Synapse Dedicated pools	---UNDISCLOSED----
Install/manage Synapse packages	---UNDISCLOSED----
Configure VNETs, Firewall rules, and Network Pre-requisites for Integration Runtimes	Network Admin / ---UNDISCLOSED----
Create Integration Runtimes in Synapse	---UNDISCLOSED----
Create credentials, assign workspace item permissions, provide access to developers	---UNDISCLOSED----
Develop pipelines, notebooks, dedicated pool tables/objects	---UNDISCLOSED-----
Promote code between environments (no delete permissions in PROD)	---UNDISCLOSED---- Lead
Create dedicated pool level system objects (credentials, external tables, data sources)	---UNDISCLOSED----

IM-4: Do not store credentials in code and rotate keys

Do not store credentials in code use more secure locations such as KeyVaults for storing credentials.

Protect key material and other secrets by using key vaults. Establish RBAC access to keys and Rotate keys at least once per year.

Privileged Access Management

PA-1: Restrict administrative access to business-critical systems

Access will be limited to Blob Inventory Reports.
Blob Inventory will use dedicated container as target for Inventory Reports which will be processed by Synapse

PA-2: Follow just enough administration (least privilege principle) and RBAC

Implement RBAC using the Least Access Privilege approach and segregation of duties as per table. Artifacts can be used as a reference for IM-3 control

Activity	Responsible
Create Separate Synapse Workspace per use case	**---UNDISCLOSED-----**
Create Synapse Spark Pools and Dedicated Pools	**---UNDISCLOSED-----**
Enable Blob Inventory on Storage Accounts	**---UNDISCLOSED-----**
Define Blob Inventory Rules	**---UNDISCLOSED-----**
Pause Synapse Dedicated pools	**---UNDISCLOSED-----**
Install/manage Synapse packages	**---UNDISCLOSED-----**
Configure VNETs, Firewall rules, and Network Pre-requisites for Integration Runtimes	**Network Admin / ---UNDISCLOSED-----**
Create Integration Runtimes in Synapse	**---UNDISCLOSED-----**
Create credentials, assign workspace item permissions, provide access to developers	**---UNDISCLOSED-----**
Develop pipelines, notebooks, dedicated pool tables/objects	**---UNDISCLOSED-------**
Promote code between environments (no delete permissions in PROD)	**---UNDISCLOSED------ Lead**
Create dedicated pool level system objects (credentials, external tables, data sources)	**---UNDISCLOSED-----**

Data Protection:

DP-1: Protect sensitive data

New Separate Synapse workspace will be created for ITSDA Data Asset Inventory Reporting.
Rules and Filters should be defined for appropriate frequency and consistency of the data across data lakes

Use Existing Custom RBAC and Security Controls based on ACN-2021-20567 Synapse Workspace Enablement.

Encryption using Customer Managed Keys will be used when creating the Synapse workspace.

Integrate a Key Vault, that is being secured with Private Link and FW restrictions.

All secrets for Service Principals or SQL Logins will be stored in the Key Vault. All Linked services will use Key Vault to authenticate underlying data lake and on-prem resources (if applicable).

Authentication key is generated by Synapse. This Authentication key is used to register Self Integration Runtime environment. Credentials are encrypted and stored using Windows DPAPI.

DP-2: Monitor for unauthorized transfer of sensitive data

Advanced Threat Protection for Azure Synapse Analytics to detect anomalous activities indicating unusual and potentially harmful attempts to access or exploit databases. (if cost is okay to use ATP)

Advanced Threat Protection is part of the Azure Defender for SQL offering, which is a unified package for advanced SQL security capabilities. (use Azure defender for SQL if cost is okay to use)

DP-3: Encrypt sensitive information in transit

Encrypt all sensitive information in transit. Ensure that inbound and outbound connections to your resources can negotiate TLS 1.2 or greater. If there is incompatibility TLS 1.1 and above is okay.

DP-4: Encrypt sensitive data at rest

Use encryption at rest on all resources. Encrypt using Customer-managed key. Enforce Encryption for the Synapse workspace.

Secure access to key vault with privatelink and FW restriction.

Transparent data encryption should be enabled to protect data-at-rest and meet compliance requirements

If necessary, Implement Transparent Data Encryption (TDE) with your own key provides you with increased transparency and control over the TDE Protector.

Note (not a control):
Transparent data encryption (TDE) helps protect Azure Synapse SQL against the threat of malicious offline activity by encrypting data at rest. It performs real-time encryption and decryption of the database, associated backups, and transaction log files at rest without requiring changes to the application. In Azure, the default setting for TDE is that the DEK is protected by a built-in server certificate. Alternatively, you can use customer-managed TDE, also referred to as Bring Your Own Key (BYOK) support for TDE. In this scenario, the TDE Protector that encrypts the DEK is a customer-managed asymmetric key, which is stored in a

customer-owned and managed Azure Key Vault (Azure's cloud-based external key management system) and never leaves the key vault.

Data Privacy

DPR-1: Tier 3 privacy data should not be logged

Tier 3 Privacy data (e.g Immigration, Salary, SSN, Driver's license, health records, passwords) should not leave outside the application boundaries. Tier 3 privacy data should not be logged in the event logs.

If there is tier 3 data consider to implement Dynamic data masking helps prevent unauthorized access to sensitive data by enabling customers to designate how much of the sensitive data to reveal with minimal impact on the application layer. It's a policy-based security feature that hides the sensitive data in the result set of a query over designated database fields, while the data in the database is not changed.

DPR-2: Ability update personal profile information and receive notification

Data subjects the ability to access, review and update their own personal data when necessary. Data Subjects to be notified in the event of modification of their personal data.

DPR-3: Data Life Cycle Management

Controls must be implemented to retain personal data in accordance with the applicable Records Retention and Disposition Schedules and implement a deletion mechanism for the personal data at the end of the retention period, as determined by WBG standards.

Monitoring & Detection – Logging, Monitoring and Threat Detection

LT-1: Enable threat detection for resources

If cost is not an issue enable Azure Defender for your Azure Synapse Analytics resources. Azure Defender provides a set of advanced SQL security capabilities, including SQL Vulnerability Assessment and Advanced Threat Protection.

Forward any logs from Azure Synapse to WBG SIEM, which can be used to set up custom threat detections.

LT-4: Enable audit logging for resources, collect security logs from OS and 3rd party applications

Integrate WBG SIEM with audit, activity, resource logs for resources to meet the requirements for compliance, threat detection, hunting, and incident investigation.

Collect security logs from OS and applications send to WBG SIEM. Verify that logs are securely transmitted to a preferably WBG SIEM system for analysis, detection, alerting, and escalation. ([C9](https://owasp.org/www-project-proactive-controls/#div-numbering))

Vulnerability Management

PV-1: Perform Vulnerability Assessments on resources, images, registries, codes

Use the latest versions of supported platforms, programming languages, libraries, protocols, and frameworks

Implement OIS DevSecops security hooks and Adopt WBG DevSecOps practice

Perform vulnerability scanning on dependencies, supported platforms, programming languages, libraries, protocols, and frameworks.

Perform SQL vulnerability assessment that can discover, track, and help you remediate potential database vulnerabilities.

Note (Not a Control):
You can use Vulnerability assessment is part of the Azure Defender for SQL offering. Use it to proactively improve your database security Azure Defender provides a set of advanced SQL security capabilities, including SQL Vulnerability Assessment and Advanced Threat Protection.

Backup and Recovery

BR-1: Ensure regular automated backups

Ensure you are backing up systems and data to maintain business continuity after an unexpected event. This should be defined by any objectives for Recovery Point Objective (RPO) and Recovery Time Objective (RTO). Implement a deletion mechanism for these backups at the end of the retention period.

Backup/restore scope is for the following two capabilities:

- Development artefacts – backup for data pipelines and notebooks will be carried over the DevSecOps process, planned for Synapse Workspace in Q3/FY21
- Dedicated Pools – existing Backup configuration of **opsdatamartdw** will be reused, as there are no new requirements on Availability/DR capabilities for the solution. Existing strategy is snapshot backup across pair region every 24 hrs.

BR-2: Mitigate risk of lost keys

Ensure that you have measures in place to prevent and recover from loss of keys. Enable soft delete and purge protection in Azure Key Vault to protect keys against accidental or malicious deletion.

Residual Risks

The following tables contain the residual risks identified during the Security Architecture Review of

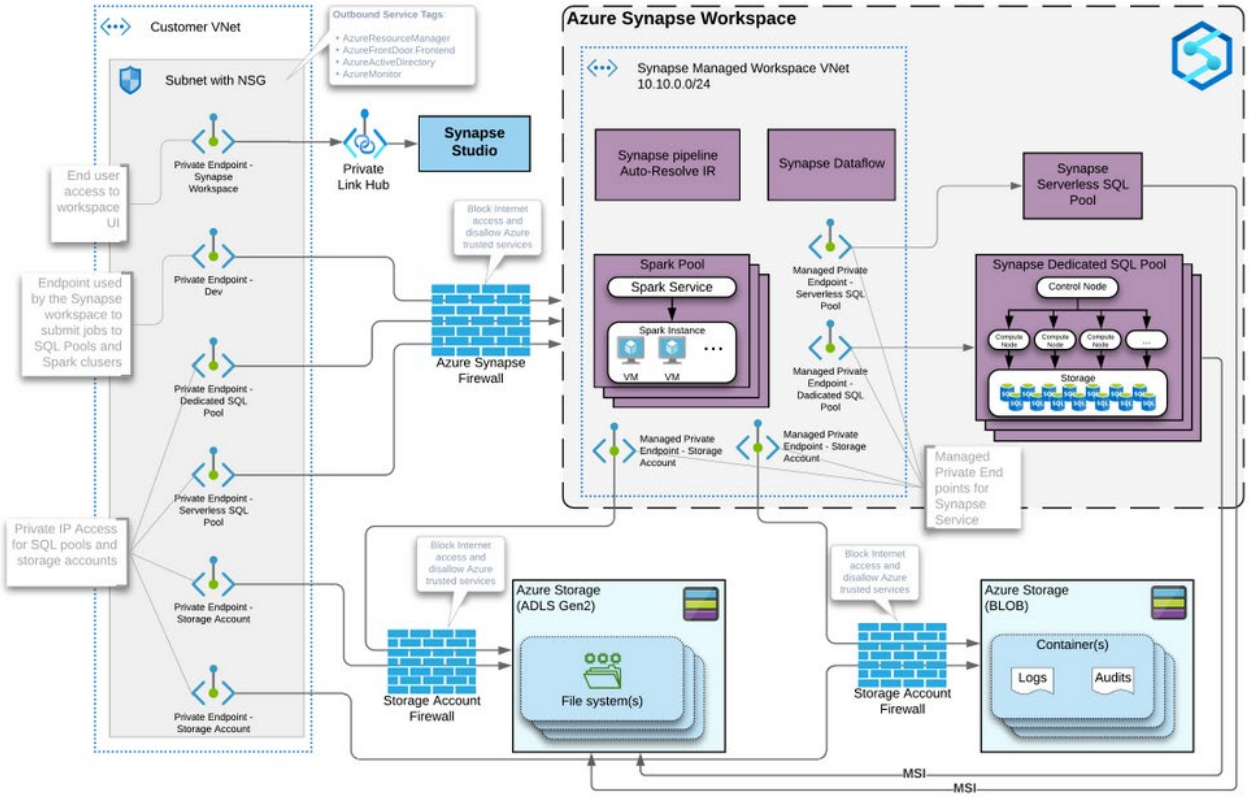
Technical Risks	Description	Likelihood	Impact	Overall
Business Risks	**Description**	**Likelihood**	**Impact**	**Overall**

Appendix

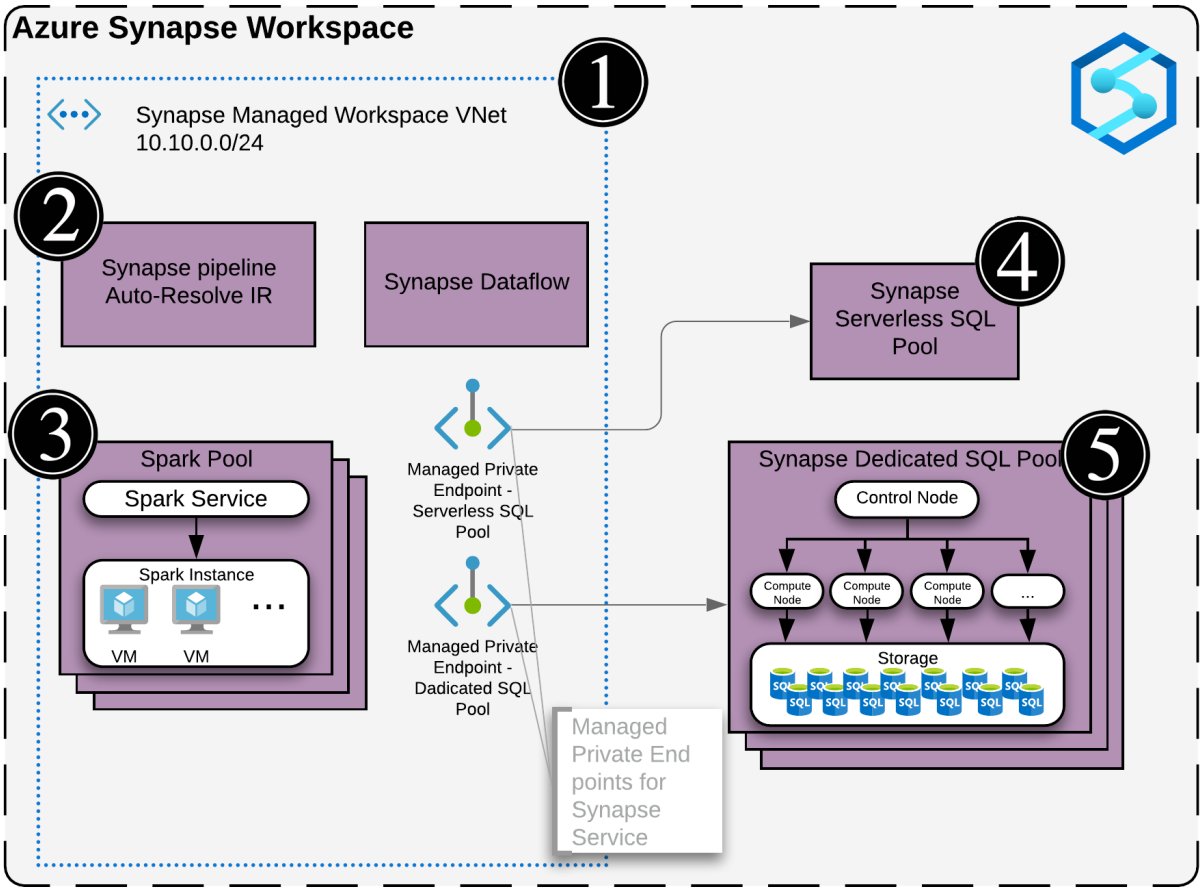
The following documents were leveraged during review of this case:

Reference Material

Network Security Azure Synapse

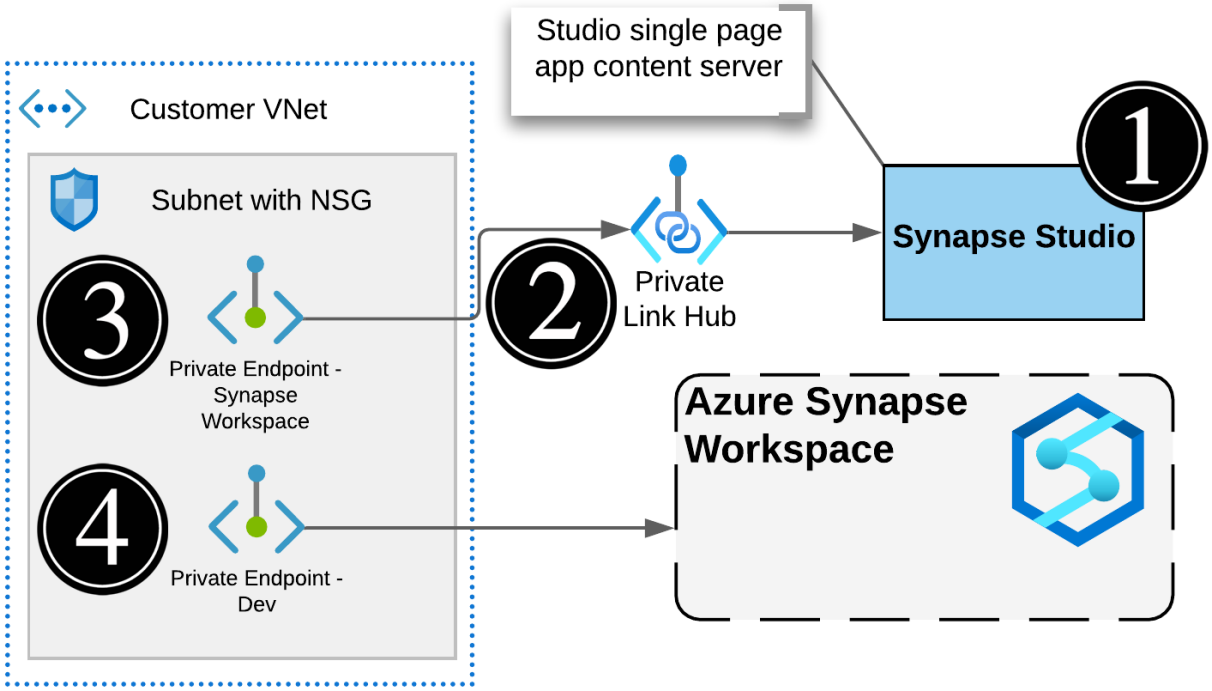


Synapse Workspace



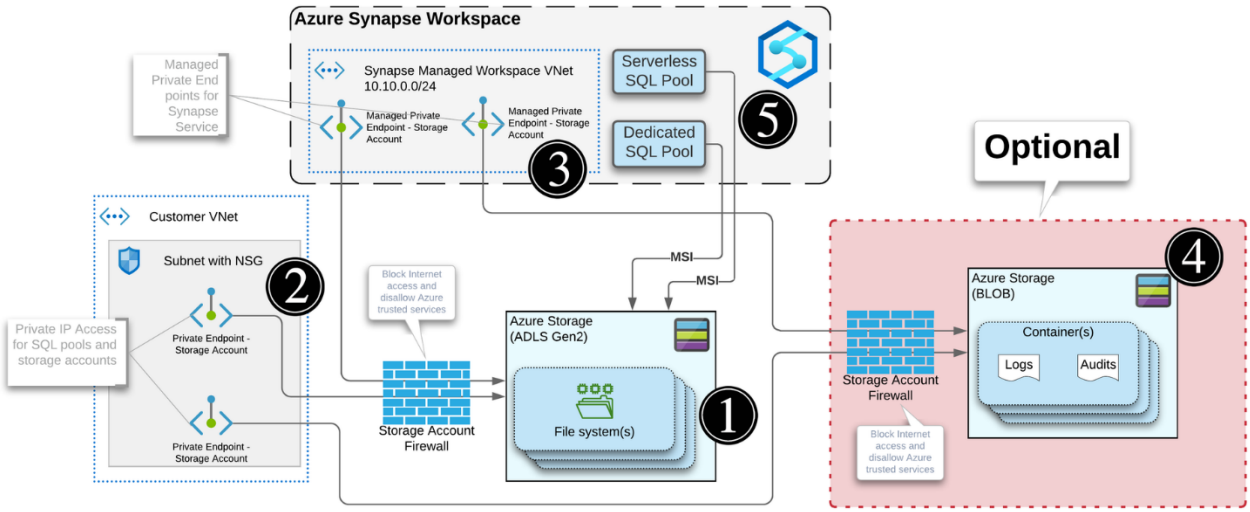
1. **Use managed virtual networks, when you deploy Azure Synapse Workspace.** Virtual network will give you network isolation against other workspaces. This is accomplished by enabling the "Enable managed virtual network" option during the deployment of the workspace. The virtual network created as part of the managed vNet workspace deployment. This vNet is managed by Microsoft and cannot be seen in the Azure portal's resource list.
2. Managed Vnet should contain the compute for the self-hosted integration runtime and the compute for the Synapse Dataflow.
3. Managed Vnet should host Any spark pools which will create virtual machines behind the scenes.
4. Connect Serverless SQL pool service via private endpoints. Serverless SQL pool is a multi-tenant service and will not be physically deployed in the vNet.
5. Connect dedicated SQL pool service via private endpoints. Dedicated SQL pool is a multi-tenant service and will not be physically deployed in the vNet.

Synapse Studio



1. The workspace studio UI is a single-page application (SPA) and is created as part of the Synapse workspace deployment.
2. **Utilize an [Azure Synapse Link Hub]**, you're able to create a private endpoint into the customer's owned vNet.
3. Users can connect to the Studio UI using this private endpoint.
4. Executions like notebooks or SQL scripts made from the studio web interface will submit commands via the DEV private endpoint and ran on the appropriate pool.

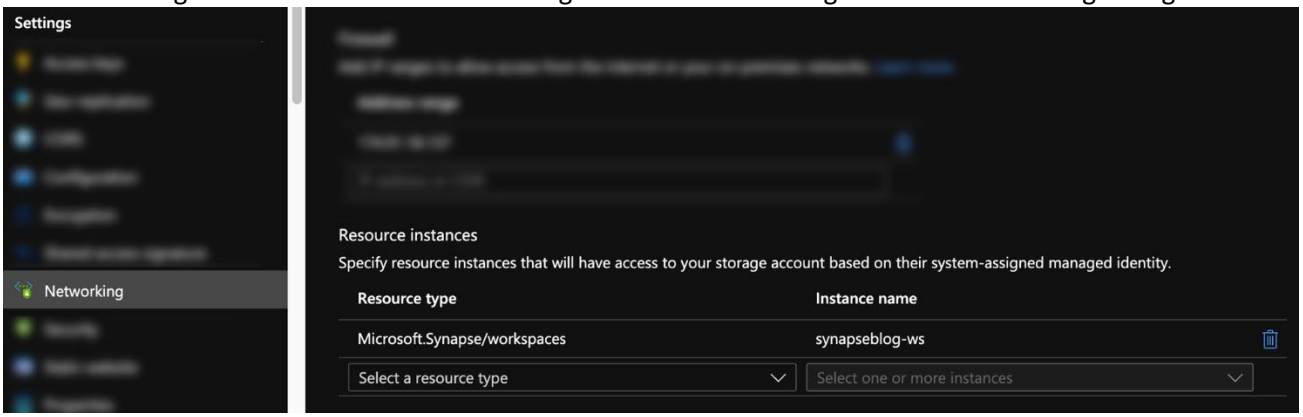
Storage Accounts and Synapse



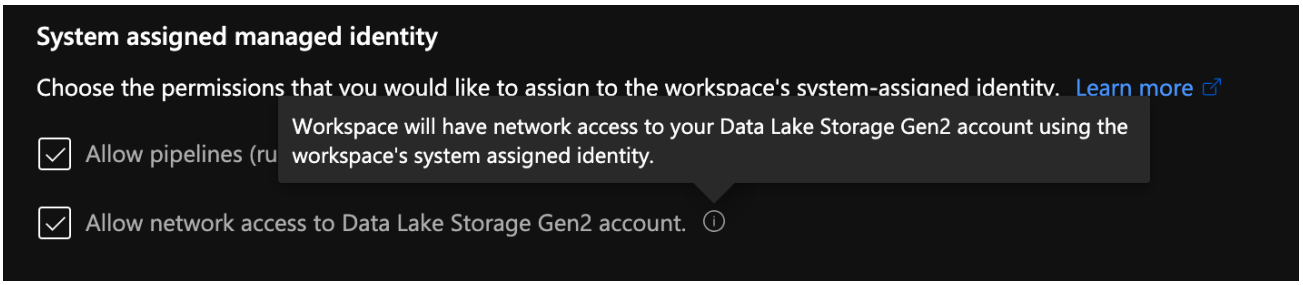
1. For each workspace created, you will need to specify a storage account / file system with hierarchical name space enabled in order for Synapse to store its metadata.
2. When your storage account is configured to limit access to certain vNets, endpoints are needed to allow the connection and authentication. Similar to how Synapse needs private endpoints to communicate with the storage account, any external systems or people that need to read or write to the storage account will require a private endpoint.
3. Every storage accounts that you connect to your Synapse workspace via linked services will need a managed private endpoint like we mentioned previously. This applies to each service within the managed vNet.
4. Optional: You can use another storage account to store any logs or audits.

Note, logs and audits cannot use storage accounts with hierarchical namespace enabled. Hence the reason why we have 2 storage accounts in the diagram.

The SQL pools, which are multi-tenant services, talk to the Storage over public IPs but use trusted service based isolation. However, going over public IPs doesn't mean data is going to the internet. Azure networking implements cold potato routing, so traffic stays on Azure backbone as long as the two entities communicating are on Azure. This can be configured within the storage account networking configuration.



Storage Account Trusted MSIvOr can also be set during the Synapse workspace creation.



Storage Account Trusted MSI

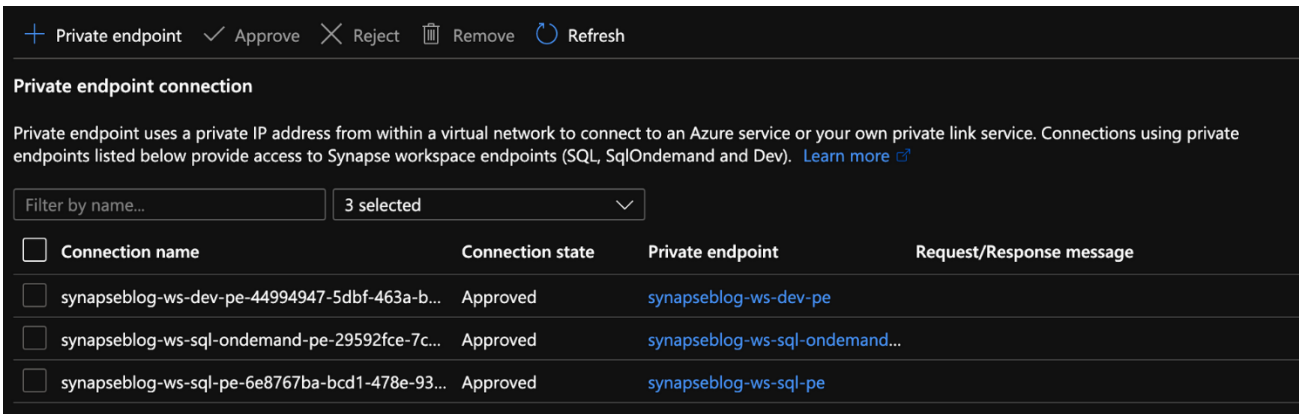
In commercial network routing between autonomous systems which are interconnected in multiple locations, hot-potato routing is the practice of passing traffic off to another autonomous system as quickly as possible, thus using their network for wide-area transit. Cold-potato routing is the opposite, where the originating autonomous system holds onto the packet until it is as near to the destination as possible.

Private endpoints in customer-owned vNet

Like I mentioned previously for the storage accounts, private endpoints need to be created in the customer's vNet for the following:

1. Dedicated SQL Pool
2. Serverless SQL Pool
3. Dev

Like you can see here:



Private Endpoints created in the portal

DNS Setup for Hybrid Scenario DNS resolution is conditionally forwarded to Azure

A typical hub-spoke company setup where part of the infrastructure is on prem and the resources would need to be discoverable across networks.

For this to work you need to also setup a (redundant/resiliant) DNS forwarder (on the hub) and on the on prem existing DNS selectively forward some DNS queries to this DNS forwarder. In addition you would need to add Azure Private Zone to the private links and keep those associated to the VNET/VNETs that should be able to talk to Synapse.

The typical setup is described here: https://docs.microsoft.com/en-us/azure/private-link/private-endpoint-dns (picture below pulled from there), and to complete the setup as you suggested the private zones required would be

- privatelink.sql.azuresynapse.net
- privatelink.dev.azuresynapse.net
- privatelink.web.azuresynapse.net
- privatelink.dfs.core.windows.net
- privatelink.blob.core.windows.net