



# Data Migration Platform Pre-Requisite

September 2024

Prepared for: Hume City Council

Author: Chaitra Chandra Mouli (Senior Consultant, Enterprise, KPMG AU)

# Contents

- 1 Introduction..... 3**
  - 1.1 Purpose..... 3
  - 1.2 Decisions ..... 3
  - 1.3 Dependencies ..... 3
- 2 Platform Overview ..... 7**

# Table of Figures

- Figure 1: Data Migration Platform Logical Overview ..... 8

# 1 Introduction

KPMG has been engaged to develop a data migration platform to assist Hume City council with their data migration from their source system to target system.

## 1.1 Purpose

This document describes the detailed landing zone requirements to design and configure a data migration target state data platform that is focused on Azure data services and supporting core services.

The audiences for this document are:

- HCC Technical Stakeholders and
- KPMG Project Team

## 1.2 Decisions

These decisions are based on numerous factors such as data migration platform requirements, constraints, technical considerations, and stakeholder preferences. Following are some key points to understand about design decisions.

Ref #	Category	Description
D-01	Data Ingestion	Azure Synapse Analytics Pipelines will be used for data ingestion purpose
D-02	Data Store	Azure data lake storage will be used for data store purpose
D-03	Data Process	Azure serverless SQL pool, dedicated SQL pool, and spark pools will be used for data processing and enriching purpose.
D-04	Data Ingestion	Azure self-hosted integration runtime will be used to establish the secure network connectivity between all the source systems to data analytics platform
D-05	Data Serve	Azure data gateway will be used to establish the secure network connectivity between Power BI services to data analytics platform. SFTP will be used to securely transfer data between the platform and the target system.
D-06	Environments	Single instance of the environment will be provisioned for the data migration platform considering the data to be migrated is of production level.
D-07	Environments	No disaster recovery environment will be used. The dedicated platform instance is only for the data migration purposes and not for any analytics purpose.
D-08	Subscriptions	The data migration platform will be deployed within the Hume's azure tenancy (subscription)

## 1.3 Dependencies

The following are the dependencies that need to be fulfilled by Hume to complete the implementation of data migration platform.

Ref #	Description		
DP-01	<p>Access to allow deployment of all cloud resources and permissions to provision, troubleshoot and fix deployments without the assistance of the external teams.</p> <p><b>Microsoft Entra ID</b> tenancy for identity and access management of data migration platform resources.</p> <p><b>Microsoft Entra ID Security Groups</b> to control the access to data migration platform resources using RBAC. Refer the RBAC section of the document for requirements.</p>		
DP-02	<p>Identity and Access Management Zone dependency, hosted in the Hume's Azure landing zone:</p> <p><b>Domain Controller:</b> Ensure domain controller are available, to domain join self-hosted integration runtime and data gateway virtual machines.</p> <p><b>DNS Services and Conditional Forwarders:</b> Ensure DNS server with required conditional forwarders are configured to provide name resolution services to translate domain names into IP address for azure data migration platform resources and services from on-premises network. These DNS servers will be configured as a custom DNS for data migration platform virtual network.</p>		
DP-03	<p>Connectivity Zone (Hub) dependency, hosted in the Hume's landing zone:</p> <p><b>Network Connectivity (Express Route or Site-to-Site VPN):</b> Ensure a Site-to-Site VPN or ExpressRoute connection available to connect the on-premises systems for data ingestion. This establishes private connections between Azure data migration platform, on-premises infrastructure, and target system. The connection should support &lt;1 million records of data at any given time. KPMG estimates that a data throughput of 5GB per trial Data Migration run, excluding attachment content. KPMG will monitor the throughputs throughout the project and will adjust on an as needed basis, if required.</p> <p><b>Azure Firewalls:</b> Ensure that Azure firewall or similar is available with required rules and policies to act as a perimeter firewall within hub network to control and inspect network traffic between Azure data migration platform, other Hume systems and Internet. Configure the route table with a user defined route 0.0.0.0/0 traffic to Azure firewall appliance as next Hop.</p> <p><b>Azure Private DNS Zones:</b> Ensure that below azure private DNS zones are available to facilities name resolution of data migration platform services within Azure virtual networks.</p> <p><b>Azure Private DNS Resolver (based on the Hume Azure DNS architecture):</b> Ensure required conditional forwarders are configured to provide name resolution services to translate domain names into IP address for azure data migration platform resources and services from on-premises network.</p> <table> <tr> <th>Private DNS Zone</th><th>Purpose</th></tr> </table>	Private DNS Zone	Purpose
Private DNS Zone	Purpose		

	privatelink.azure.synapse.net	For Synapse web studio endpoints
	privatelink.dev.azure.synapse.net	For Synapse development endpoints
	privatelink.sql.azure.synapse.net	For Synapse Dedicated sql pools
	privatelink.sql.azure.synapse.net	For Synapse Serverless sql pools
	privatelink.blob.core.windows.net	For blob storage accounts
	privatelink.dfs.core.windows.net	For data lake storage accounts
	privatelink.vaultcore.azure.net	For KeyVault
DP-04	<p>Management Zone Dependency, hosted in the Hume's Azure landing zone.</p> <p><b>Centralised Monitoring Resources:</b> Ensure that the centralised monitoring private link and private endpoint resources are configured and available to connect with the data migration platform log analytics workspace.</p> <p><b>Centralised Log Storage Account:</b> Ensure that the centralised storage account is configured and available to store data platform resources logs and telemetry data.</p>	
DP-05	<p>Microsoft Defender Dependency, to secure the data migration platform resources.</p> <p><b>Microsoft Defender:</b> Ensure that the Microsoft defender configurations are completed to secure the below data migration platform resources.</p> <ul style="list-style-type: none"> <li>• Azure Data Lake Storage Account</li> <li>• Virtual Machines</li> <li>• Key Vault Services</li> <li>• Azure Synapse Analytics</li> </ul>	
DP-06	<p>Microsoft DevOps Dependency, hosted in Hume's DevOps organisation.</p> <p><b>Dedicated DevOps project:</b> Ensure that the dedicated DevOps project is configured within the Hume's DevOps organisation to host the data migration platform IaC repositories and CI/CD pipelines.</p> <p><b>DevOps Self-hosted or Microsoft-hosted Agent Pools:</b> Ensure that at least two DevOps Self-hosted or Microsoft-hosted agents and pool resources configured and available to execute the data migration platform IaC builds (bicep code), and CI/CD deployments.</p> <p><b>DevOps Service Connection and service principal:</b> Azure DevOps service connection with appropriate permissions (Contributor access) to data migration platform subscriptions to deploy and configure the Azure data migration platform resources.</p>	

	<b>DevOps Access:</b> Ensure that the KPMG team are having the project administrator access to the DevOps project.
DP-07	<p><b>Power BI workspace:</b> Ensure that the Power BI workspace is configured to generate visualisation reports for reconciliation purposes.</p> <p><b>Power BI Access:</b> Ensure that the KPMG team are provisioned with the appropriate Power BI license and access</p>
DP-08	<b>SFTP services:</b> Ensure that SFTP services are configured and enabled for undertaking file transfers between the platform and the target system.
DP-09	An active Azure subscription with pre-configured policies as per Hume's landing zone standards to host the data platform resources.
DP-10	An Azure resource group with pre-configured policies as per Hume's landing zone standards to host the data platform resources.
DP-11	<p>Ensure that an Azure virtual network with the /25 private endpoints subnet to host the data migration platform resources and private endpoints.</p> <p>Ensure that the virtual network peering to hub network is configured to provide the data migration platform the required network connectivity to Hume's systems.</p> <p>Ensure that the user defined route table is configured and associated with the data migration platform virtual network subnets to route the traffic to azure firewall hosted in the hub network.</p>
DP-12	Standard Azure naming conventions and tagging policies for data migration platform services.

# 2 Platform Overview

The platform overview provides a high-level understanding of the logical components and architecture of the data migration platform. It includes the following key aspects:

**Data Sources:** The platform integrates with various data sources, such as relational databases, SaaS Applications, APIs, and external systems. These sources provide the raw data that is ingested into the platform for processing.

**Data Ingestion:** The platform includes mechanisms for data ingestion, which involve extracting data from the sources, transforming it into a suitable format, and loading it into the platform for further processing. This includes batch processing depending on the specific requirements.

**Data Storage:** The platform provides storage capabilities to securely store and manage the ingested data. This involves using Azure data lakes storage service offered by the underlying cloud infrastructure. The storage is designed to support scalability, performance, and data governance requirements.

**Data Transformation:** The platform includes components for processing and transformation of the data. This involves various data processing technologies such as SQL engines, data pipelines, data transformation tools. These components enable the execution of complex data processing tasks and facilitate the generation of insights and actionable outcomes.

**Data Security:** The platform incorporates security measures to protect data at rest and in transit. This includes access controls, encryption, logging. The platform ensures that only authorised users and processes can access and manipulate the data, maintaining its confidentiality, integrity, and availability.

**Analytics and Visualisation:** The platform provides capabilities for performing analytics and visualising the processed data. This involves tools and frameworks for data exploration, presentation and reconciliation. The goal is to present reports to reconcile migrated data effectively.

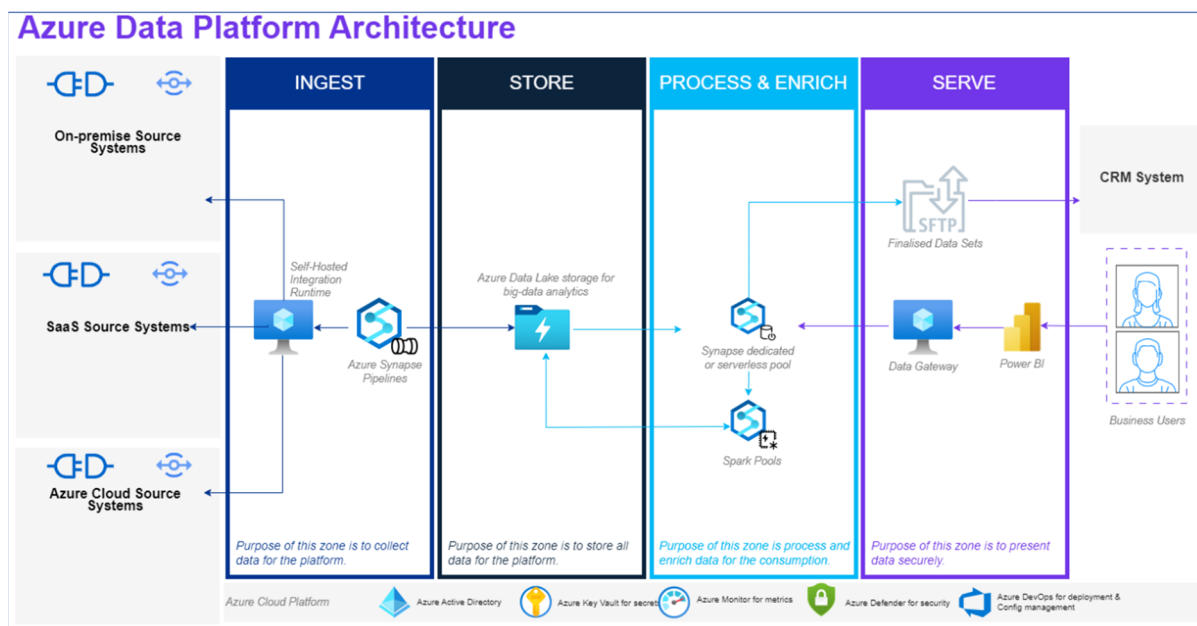


Figure 1: Data Migration Platform Logical Overview

Azure Data Migration Platform provides a comprehensive set of services and tools designed to support the data migration architecture including data ingestion, processing, and visualisation (for reconciliation purposes) capabilities.

Component	Description
<b>Azure Synapse Analytics</b>	Azure Synapse Analytics is a powerful service that supports a wide variety of data sources, including on-premises databases, cloud storage, and third-party applications for data ingestion and processing. Its seamless integration with other Azure services, such as Azure SQL Pools and spark pools, enhances its functionality. It enables data migration team to data exploration, data preparation, at scale using popular programming languages like Python, Scala, and SQL.
<b>Azure Data Lake Storage (Gen2)</b>	Azure Data Lake Storage (Gen2) offers a robust and scalable solution tailored for data migration platforms. It provides seamless integration with Azure's ecosystem, facilitating efficient data ingestion and processing. It offers advanced security features to ensure data remains protected throughout the migration process.
<b>Azure SQL Pool</b>	Azure SQL Pools facilitate efficient data ingestion from various sources, including on-premises databases, cloud storage, and third-party applications. Azure SQL Pools support complex data transformation operations, making it easier to cleanse, integrate, and prepare data for migration. It handles high-performance data processing, by leveraging distributed processing and parallel execution, ensuring minimal downtime and high data availability.
<b>Microsoft Power BI</b>	Power BI is a tool that is used to visualise and explore data through interactive dashboards, reports, and data visualisations. It integrates seamlessly with Azure data migration platform, allowing users to analyse, and reconcile the processed and migrated data effectively.
<b>Self-hosted Integration Runtime (SHIR)</b>	SHIR is a component of Azure Synapse Analytics that allows Hume to securely connect their on-premises data sources and systems to the data analytics platform. It provides a



	hybrid integration solution, enabling data movement and transformation between on-premises data stores and Azure data lake.
<b>Azure Data Gateway</b>	Azure Data Gateway ensures secure connectivity between data analytics platform and Microsoft Power BI by using encryption and authentication mechanisms, protecting data during transit. Azure Data Gateway can be used to schedule and automate the refresh of Power BI reports connected to data migration and analytics platform. This ensures that Power BI visualisations always reflect the latest data.
<b>Microsoft Entra ID</b>	Microsoft Entra ID (previously known as Azure Active Directory) is a comprehensive identity and access management solution that provides secure and efficient management of user identities, access controls, and application integration in the data migration platform.
<b>Azure Key Vault</b>	Azure Key Vault offers centralised management, secure storage, and access control capabilities to protect sensitive information and meet compliance requirements.
<b>SFTP</b>	SFTP (Secure File Transfer Protocol) is an essential component for data migration platforms, offering secure and reliable file transfer capabilities with its ability to handle large file sizes and resume interrupted transfers. It ensures the encryption of data in transit, mitigating the risks associated with data breaches and unauthorized access. It supports various authentication methods, including password-based and public key authentication.
<b>Azure DevOps</b>	Azure DevOps is a cloud-based service offered by Microsoft that provides a set of development tools and services to help the platform teams plan, develop, test, and deliver data analytics platform infrastructure efficiently. It encompasses a wide range of capabilities, including version control, project management, continuous integration/continuous delivery (CI/CD), testing, and release management