

# Enterprise Azure Modern Data Platform





# Contents

- 1. **Scope ..... 2**
- 2. **Service Descriptions and Prerequisites ..... 4**

## Notice

This document contains confidential information that is proprietary to ANS Group Limited. No part of its contents may be used, copied, disclosed or conveyed to any party in any manner whatsoever without prior written permission from ANS Group Limited. E&OE. © Copyright 2009 – ANS Group Limited, All Rights Reserved.



# 1. Scope

The ANS Enterprise Azure Modern Data Platform service will design and deliver a central and secure data platform for you to ingest and store structured, semi-structured and unstructured data built on top of the Microsoft Azure platform and using the best-in-class services that it offers. ANS will follow the guidelines of the Microsoft Cloud Adoption Framework in designing and delivering this platform. The platform will be bespoke to you and will ensure longer term ambitions are not restricted.

The core components of this solution are listed below, these solutions have been recommended based on us looking at the entire data landscape in Azure and match this to SYP's requirements, taking into consideration service maturity, support status and future roadmap.

**Synapse Pipelines (Ingestion)** - Pipelines manage the seamless transfer of data between varied data stores, both within the Synapse or Fabric workspace and external systems. This encompasses extracting data from source systems, applying necessary transformations, and loading it into desired target destinations such as data warehouses, data lakes, or other designated repositories. Pipelines lay out the precise sequence of activities to be executed, guaranteeing that each step of the data processing workflow occurs in the correct order. These activities encompass data movement, transformation, and even external tasks, such as executing SQL scripts or running custom code.

**Azure Data Lake (Storage)** - The data lake serves as a centralised data reservoir, offering a cohesive and infinitely scalable solution for storing, managing, and harnessing data in its native format. This contrasts strongly with earlier data warehouse architectures where monolithic SQL databases were responsible for both the storage and processing of data, often limited in terms of capacity, using expensive storage arrays to keep data in multiple formats (often redundant), and where it was difficult to store or work with unstructured data. The lake facilitates the ingestion and storage of source data in its original format and provides the staging ground for data preparation and refinement. Data is to be curated, cleaned, and transformed within the data lake before being transferred to downstream systems such as the data warehouse (see below) or analytical platforms.

**Azure Synapse Spark and Serverless SQL Pools (Processing)** - Spark is the underlying platform that provides the parallel processing of data across a cluster of nodes, significantly accelerating the execution of complex computations, data transformations, and analytical queries. Similarly, Serverless SQL Pools provide an on-demand level of compute power, using traditional T-SQL over data in the lake rather than stored in a SQL Server. This gives huge cost benefits but retains familiarity for users and enterprise SQL features.

**Synapse Dedicated SQL Pools (Serving)** - Dedicated SQL pools use Massively Parallel Processing to distribute QL query workloads across numerous processing nodes. This design accelerates query execution and enhances overall data processing efficiency. Unlike Serverless Pools, they include the storage of data (as in a traditional SQL engine). Dedicated Pools can respond to queries extremely quickly and can process enormous quantities of data at speed.

**Azure Monitor** - Azure Monitor maximises the availability and performance of your applications and services by delivering a comprehensive solution for collecting, analysing, and acting on telemetry from the data platform environment. It will help SYP understand how your platform is

performing and proactively identifies issues affecting them and the resources they depend on.

Synapse is the heart of the new platform and handles all of the data orchestration, manipulation, modelling and serving the data out for downstream users. This coupled with a DevOps approach will allow SYP to embrace multiple environments such as Dev, Test & Prod all underpinned with a version-controlled code repository and build / release pipelines. This architecture is highly recommended for a fast moving and innovative data team. It will provide efficiency within the team and in processing of the data.

All data will be stored in a Data Lake which provides the permanent repository for historical data as it is ingested and while it undergoes any transformations before ultimately being served up in the Synapse SQL Pool.

## 2. Service Descriptions and Prerequisites

The following is an outline of what is and is not provided within this Statement of Work:

- ANS Group requires access to the necessary services / applications as required for the duration of the project:
  - Azure Subscriptions (Contributor Role)
  - Usernames & passwords required by ANS consultants to allow them access to deploy services into the environment.
  - Connection details to the data sources
  - IP Addresses
- ANS Group will be delivering this project remotely and all meetings will be arranged in advanced and using Microsoft Teams.
- ANS expect you to make Subject Matter Experts (SMEs) available

What is included within project scope?

- Data Platform and associated security ready for production use
  - Implementation of Zone Configuration and Directory Structure
  - Implementation of Orchestration Tooling
  - Implementation of Model & Serving Tooling
  - Implementation of Synapse
  - Configuration of Synapse workspace
  - Access Control design and configuration
  - Firewall configuration
  - Encryption configuration
  - Logging and auditing configuration
- HA & DR Configuration
- 3-Tier environments
- DevOps & Data innovation Pipelines
- Final Documentation