Microsoft Azure

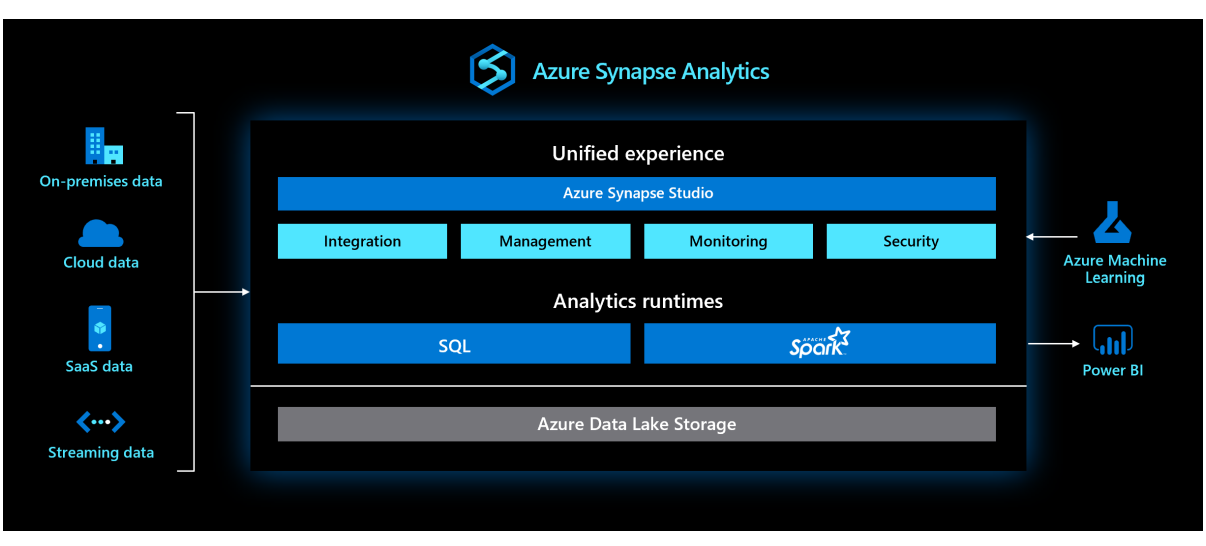
Architecture deep dive: Azure Synapse Analytics

Azure Synapse Analytics brings together enterprise data warehousing and big data analytics with a unified experience to ingest, prepare, manage, and serve data for immediate BI and machine learning needs. Here, we dive into some of the architectural features driving benefits in efficiency, agility, and value.

Tip: Get started with Azure Synapse Analytics in four quick steps.

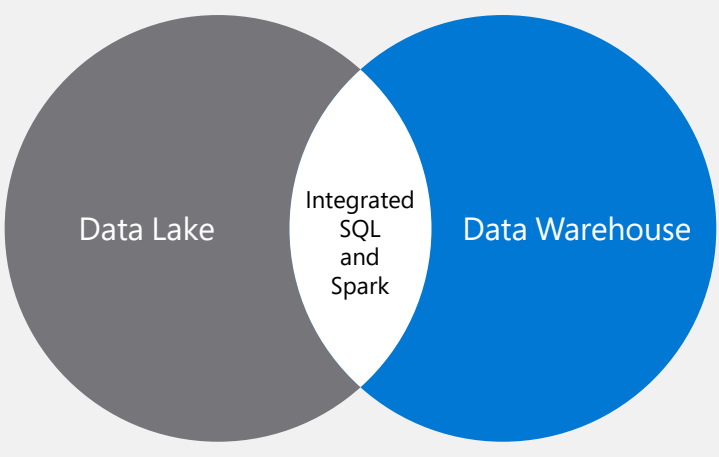
Azure Synapse Studio architecture and features

At the heart of Azure Synapse Analytics is the Azure Synapse Studio, a securable collaboration workspace for implementing and managing cloud-based analytics in Azure. A Studio workspace is deployed in a specific region under a resource group and has an associated Azure Data Lake Storage account and file system for storing temporary data.



Deeply integrated Apache Spark and SQL engines

Azure Synapse Analytics connects various analytics runtimes (such as Apache Spark and SQL) through a single platform to enhance collaboration among data professionals working on advanced analytics solutions.



Azure Synapse Analytics

Fast and easy to explore and analyze data

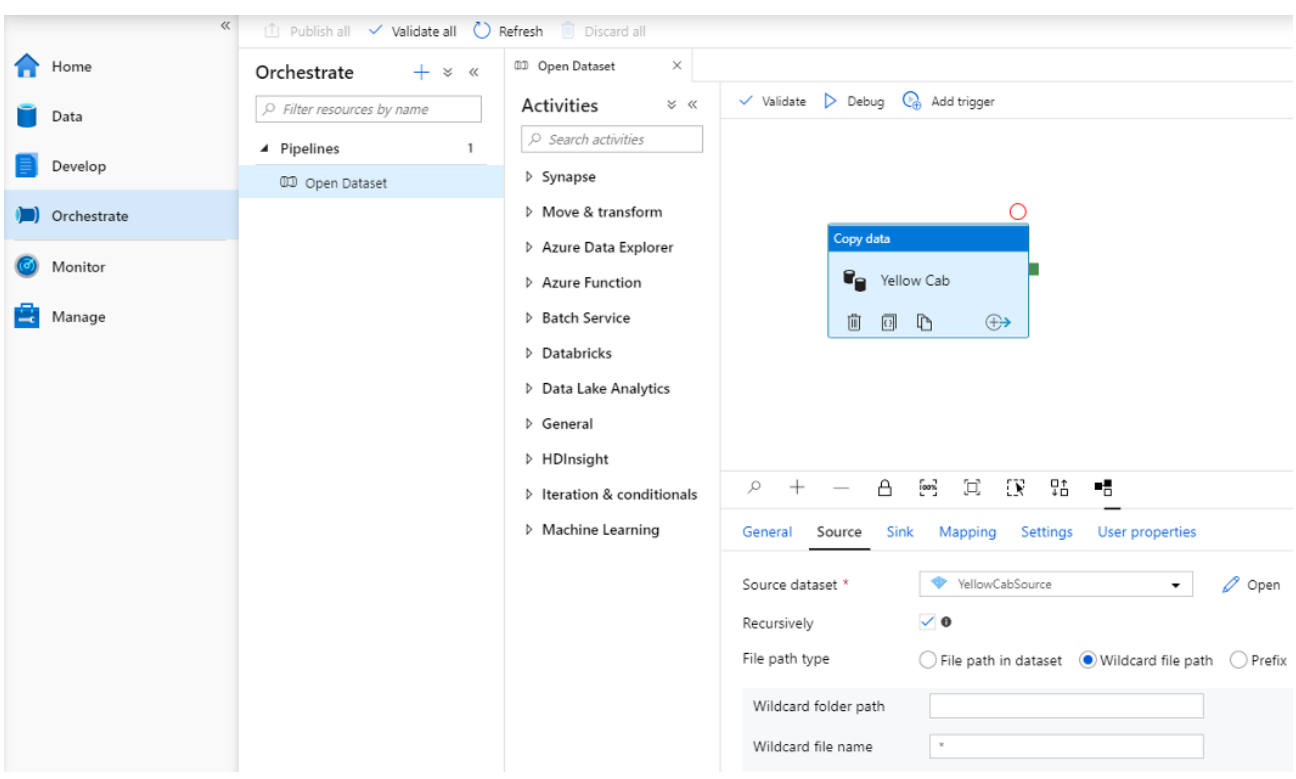
The serverless endpoint in Synapse SQL makes it fast and easy to explore and analyze over data in a data lake—with no infrastructure to set up or manage. With T-SQL, you can run serverless queries over the data lake without provisioning or managing any infrastructure. By eliminating the overhead of data center management and operations for the data warehouse, you can reallocate resources to where value is produced and focus on using the data warehouse to deliver the best information and insight. This lowers overall total cost of ownership and provides better cost control over operating expenses.

Powerful performance

Azure Synapse Analytics offers powerful relational database performance by using techniques such as Massively Parallel Processing (MPP) and automatic in-memory caching. Independent benchmarks, such as this one by GigaOm, show the results in action.

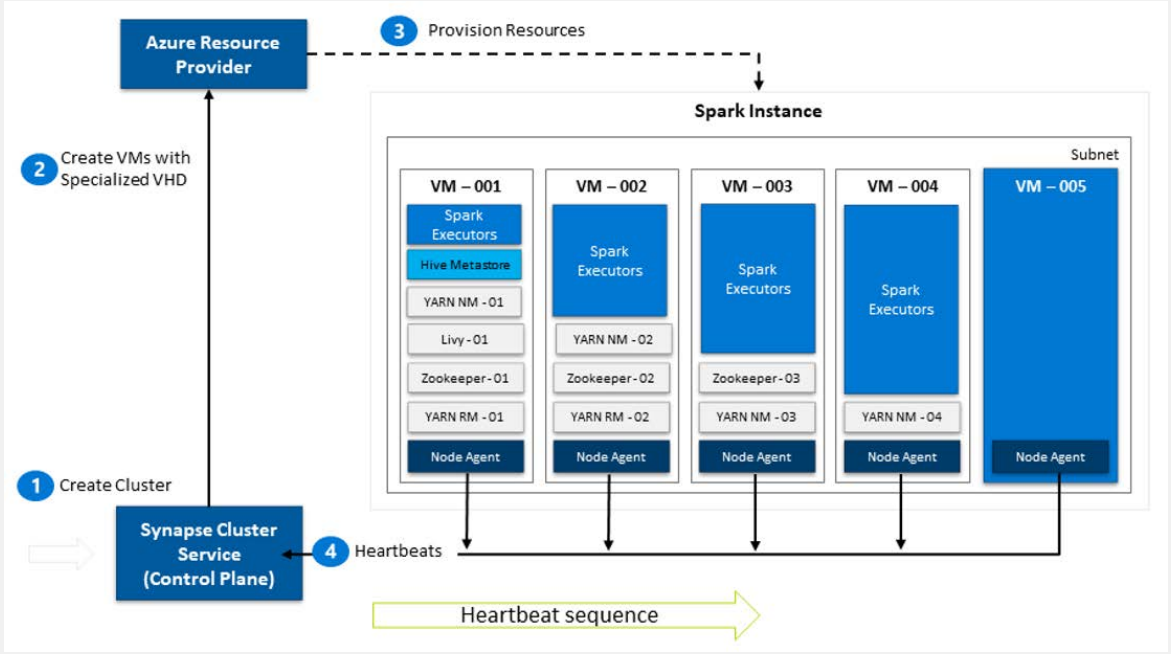
Flexibility to bring together relational and non-relational data

Easily query files in the data lake with the same service used to build data warehousing solutions. Orchestrate pipelines to perform common analytics scenarios without writing a line of code. By defining a pipeline, a data source can be linked from the **Orchestrate** hub and copied into an Azure Data Lake Storage account without any coding.



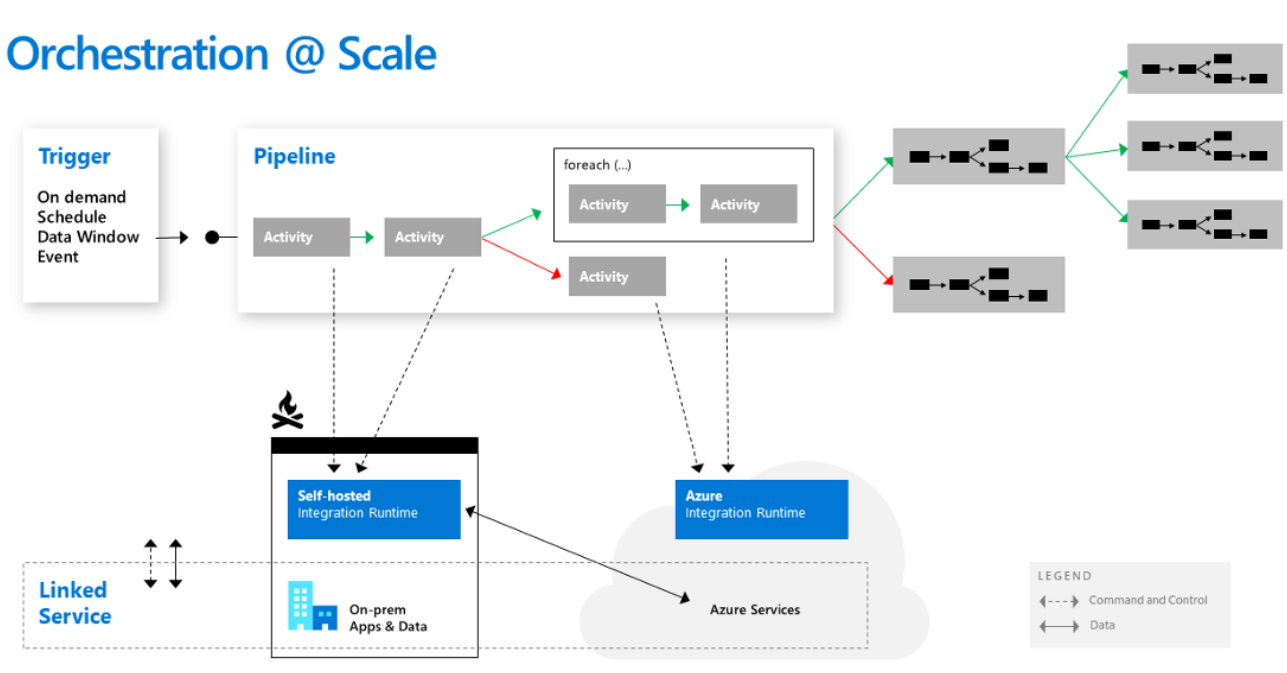
Flexible Spark integration

The Apache Spark engine simplifies the use of big data by removing the complexity of setup and cluster tuning. The power of Spark with built-in support for Azure Machine Learning addresses the full range of analytics needs, from data engineering to data science, using PySpark (Python), Spark (Scala), .NET Spark (C#), and Spark SQL. This enables enhanced collaboration, as you can now use T-SQL on both your data warehouse and embedded Spark engine.



Fast, elastic, and secure data warehousing

SQL pools can process high concurrent and complex T-SQL queries across petabytes of data to serve BI tools and applications. Cloud elasticity enables Azure Synapse Analytics to quickly increase and decrease its capacity according to demand with no impact to infrastructure availability, stability, performance, or security. Best of all, you only pay for your actual usage.



Highly scalable, hybrid data integration capability

Data ingestion and operationalization are accelerated through automated data pipelines. While the volume of data in a data warehouse typically grows with the age of the establishment, the scalability of Azure Synapse matches this by incrementally adding resources as data and workloads increase.

Industry-leading management and security

Azure is a globally available, highly scalable, secure cloud platform and Azure Synapse inherits all of that. In an Azure Synapse workspace, access to workspaces, data, and pipelines is managed granularly. Data is secured using familiar SQL-based security mechanisms. If Spark is used in the data pipeline for data preparation, cleansing, or enrichment, the Spark tables created in the process can be queried directly from Azure Synapse SQL serverless. Access is secured by using Azure Private Link to bring a serverless endpoint into a private virtual network by mapping it to a private IP address.

Get started with Azure Synapse today.

Sign up for an Azure free account

Get more details in a free technical e-book from Packt

Speak to a sales specialist for help with pricing, best practices, and implementing a proof of concept