Azure Synapse SQL Pool - Database Connectivity Options

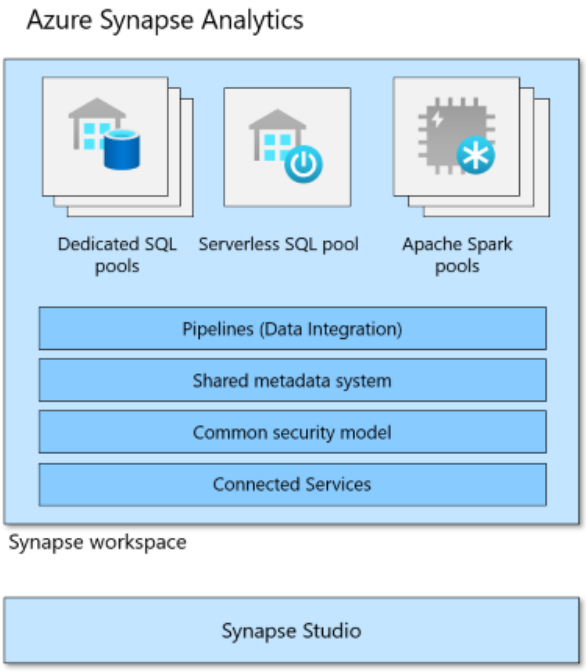
Contents

Objective

This scope of this document is to provide connectivity options for different personas to Azure Synapse SQL and the Azure Data Lake. This document tries to list out core connectivity options for querying, analyzing, and loading data. We also try to point out authentication and authorization options avaliable with AAD which is an integrated identity provider in Azure.

Azure Synapse Overview

Azure Synapse is an integrated 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, CosmosDB, and AzureML.



More Information: [What is Azure Synapse Analytics? - Azure Synapse Analytics | Microsoft Docs]

Azure Synapse Database Access with Tools

Authenticate, Query, and Analyze Data directly

Synapse SQL Pools provides multiple options and tools for connectivity. For user personas who need to manage, engineer, develop, or analyze directly connected to the databases this would be a preferred option.

Suggested User Personas	Database Managers, Developers, Administrators, Data Scientist, and ML Engineers
Recommended Connectivity Tools	• Built-in Synapse Studio (Web): Quickly Get Started with Azure Synapse Studio - Microsoft Tech Community • SSMS, Data Studio, SSDT (VS), VSCode: SQL tools overview - SQL Server | Microsoft Docs . • All the tools listed support ***AAD Authentication***. Most also support AAD with interactive MFA.
Connection strings examples	Homepage for SQL client programming - SQL Server | Microsoft Docs

- Azure Synapse Analytics workspace comes with SQL dedicated and serverless pools - endpoints. Serverless SQL pool is a query service over the data in your data lake by creating virtual databases, tables, and views over file paths.
    - Synapse SQL architecture - Azure Synapse Analytics | Microsoft Docs
- Dedicated SQL pools (formerly SQL DW) refers to the enterprise data warehousing features that let you create physical databases, tables views.
    - What is dedicated SQL pool (formerly SQL DW)? - Azure Synapse Analytics | Microsoft Docs
- Both Serverless and Dedicated SQL pools allows users to query petabytes of data across a data lake or column store indexes.
    - Serverless SQL pool - Azure Synapse Analytics | Microsoft Docs
- Both Serverless and Dedicated SQL pools provides users industry standard SQL and T-SQL language to query and analyze your data. You can use standard ANSI-compliant dialect of SQL language for data analysis. Both support features like views, stored procedures, functions, and schemas.
    - T-SQL feature differences in Synapse SQL - Azure Synapse Analytics | Microsoft Docs
- Both Support **Azure active directory (AAD) authentication** and MFA. Both also support **SQL authentication** for data access.
    - Authorizing database access - Authorize server and database access using logins and user accounts - Azure SQL Database & SQL Managed Instance & Azure Synapse Analytics | Microsoft Docs
    - Synapse Workspace separates component and data level access. If users are using Synapse Workspace Studio they must have necessary workspace access and data level access. How to set up access control for your Synapse workspace - Azure Synapse Analytics | Microsoft Docs

Authenticate and Load Data to database

Azure Synapse dedicated SQL Pools allows users to create relational tables and load data into these tables from various sources such as data lake, Azure blobs, on-prem/cloud databases such as Oracle, MySQL, SQL Server, SAP, and more. SaaS application data such as Salesforce, Marketo, and more. Please reference the Synapse Pipelines/ADF connectors for a full list of services.

Quickstart: to load data into dedicated SQL pool using the copy activity - Azure Synapse Analytics | Microsoft Docs

Suggested User Personas	Database Developers, Administrators, ETL Developers and ML Engineers
Recommended Data Loading Tools	• Synapse Workspace Pipelines: Pipelines and activities in Azure Data Factory - Azure Data Factory | Microsoft Docs • The COPY statement is the recommended loading utility as it enables you to seamlessly and flexibly load data. COPY is the fastest statement level option to load from Azure storage or data lake gen2. COPY INTO (Transact-SQL) - (Azure Synapse Analytics) - SQL Server | Microsoft Docs • Quickstart: Bulk load data using a single T-SQL statement - Azure Synapse Analytics | Microsoft Docs • Synapse Studio also provides a Bulk Load wizard for loading data quickly from Azure storage or Azure data lake. Quickstart: Bulk load data with a dedicated SQL pool - Azure Synapse Analytics | Microsoft Docs • PolyBase with T-SQL requires you to define external data objects. Tutorial: Load New York Taxicab data - Azure Synapse Analytics | Microsoft Docs • PolyBase with Azure Databricks transfers data from a table to a Databricks dataframe and/or writes data from a Databricks dataframe to a table using PolyBase. Tutorial - Perform ETL operations using Azure Databricks | Microsoft Docs
Best Practice for loading data into dedicated SQL Pools	Data loading best practices - Azure Synapse Analytics | Microsoft Docs

The following guide provides examples for how to use the COPY statement to load data from Azure Data Lake Storage to Synapse dedicated SQL Pools.

- Tutorial load data from Azure Data Lake Storage - Azure Synapse Analytics | Microsoft Docs

Guidance on how to quickly load data into Azure Synapse dedicated SQL or Azure data lake from on-premise or other cloud sources using Synapse Pipelines/ADF.

- Visual authoring - Azure Data Factory | Microsoft Docs

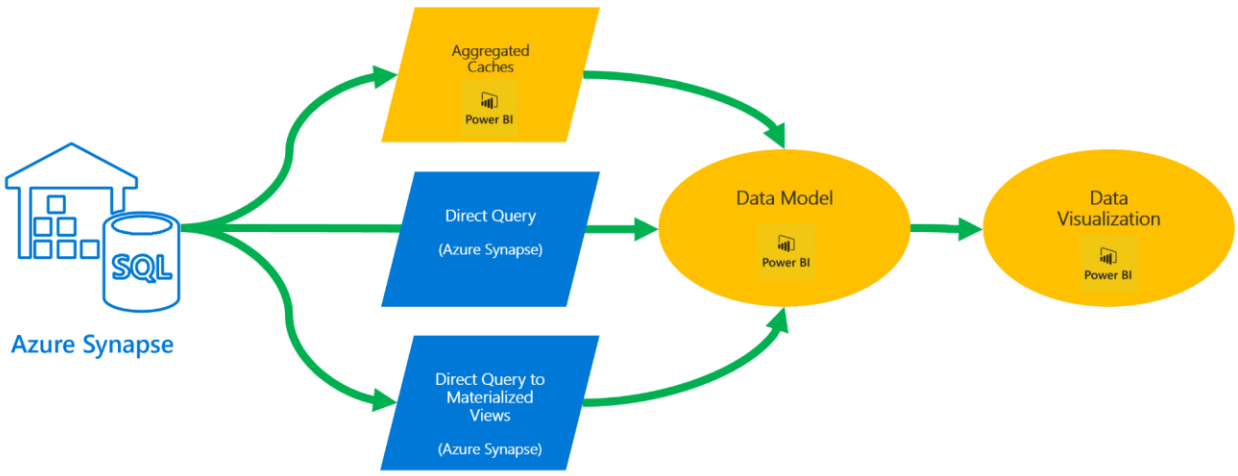
The loading tools support AAD service identity (MSI) based and authentication, SAS key-based authentication, and other authentication options based on what storage or services are being accessed by the data loading process.

Direct Query with Power BI Datasets (AAD Auth)

Users can also directly access and visualize with Synapse SQL databases using Power BI datasets. This allows for users to analyze data stored in Synapse SQL dedicated or Synapse SQL serverless (data lake) using Power BI visualizations. In this scenario after initial connection and dataset setup users can continually use the Power BI interface to do various visual analytics in a secured manner.

Suggested User Personas	Business Users, Power Users, and BI Developers
Recommended Connectivity Tools	• Power BI Cloud – Powerbi.com (preferred option) • Power BI Desktop - Power BI Desktop—Interactive Reports | Microsoft Power BI • Power BI Mobile - Mobile | Microsoft Power BI • XMLA connection to Power BI datasets - Dataset connectivity and management with the XMLA endpoint in Power BI - Power BI | Microsoft Docs • Power BI Embedded - Power BI Embedded Analytics | Microsoft Azure

Store Data in Different Layers using a Power BI Composite Model with Azure Synapse Analytics



Learn how to design and deliver high-performance, scalable, secure, and cost-effective Power BI solutions with Azure Synapse Analytics

- Power BI Pro Guide to Azure Synapse Analytics | Microsoft Azure

Synapse Developers and BI teams can quickly setup Power BI datasets with Direct Query connections to Synapse SQL dedicated and serverless pool databases.

- Connect to Synapse SQL serverless Pool - Connect to Synapse SQL with Power BI Professional - Azure Synapse Analytics | Microsoft Docs
- Connect to a Synapse SQL dedicated Pool (former DW) - Azure SQL Data Warehouse with DirectQuery - Power BI | Microsoft Docs
- During the initial connection setup users can setup connectivity with AAD authentication or use a SQL login for accessing Synapse SQL pools.

With Power BI composite models, users can connect to different kinds of data sources when you use Power BI Desktop or the Power BI service. Reports can seamlessly include data connections from more than one DirectQuery or import data connection, in any combination you choose. based on the storage mode, for model entities, visuals that don't require direct query can be imported which can improve performance and reduce back-end load. Providing optimal user experiences for BI analysts and business users.

- Composite Models in Power BI - Use composite models in Power BI Desktop - Power BI | Microsoft Docs

Materialized views in Azure Synapse SQL pool provide a low maintenance method for complex analytical queries to get fast performance without any query change. Materialized views are self maintained and Power BI can use them as a source just the same a regular database views.

- Performance tune with materialized views - Azure Synapse Analytics | Microsoft Docs

You can also link a Power BI workspace to your Azure Synapse workspace. This capability allows you to easily get data into your Power BI workspace. You can edit your Power BI reports directly in your Azure Synapse workspace.

- Azure Synapse workspace to your new Power BI workspace - Tutorial: Get started with Azure Synapse Analytics - visualize workspace data with Power BI - Azure Synapse Analytics | Microsoft Docs

Access to XMLA endpoints is available for datasets in Power BI Premium. These datasets can be connected to Azure Synapse SQL data sources and many other sources. Excel Users can use Excel PivotTables to summarize, analyze, explore, and present summary data from Power BI datasets.

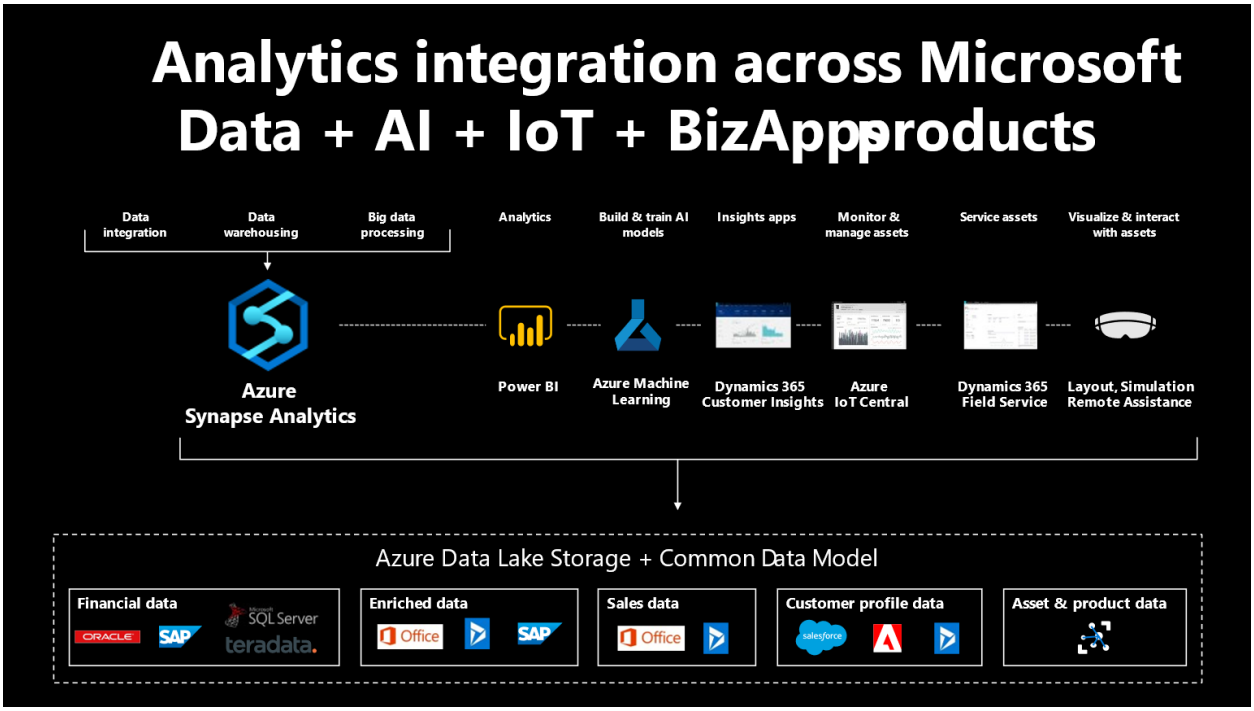
- Announcing read/write XMLA endpoints in Power BI Premium public preview | Microsoft Power BI Blog | Microsoft Power BI

Azure Synapse Spark Access with Tools

Apache Spark is a parallel processing framework that supports in-memory processing to boost the performance of big-data analytic applications. Azure Synapse makes it easy to create and configure a serverless Apache Spark pool in Azure. Spark pools in Azure Synapse are compatible with Azure Storage and Azure Data Lake Generation 2 Storage. The primary use for Spark pools is to process your data stored in Azure.

Apache Spark includes many language features to support preparation and processing of large volumes of data so that it can be made more valuable and then consumed by other services within Azure Synapse Analytics. This is enabled through multiple languages (C#, Scala, PySpark, Spark SQL) and supplied libraries for processing and connectivity.

Suggested User Personas	Data Developers, ML Engineers, Data Scientists
Recommended Connectivity Tools	• Synapse Studio Notebooks - Synapse Studio notebooks - Azure Synapse Analytics | Microsoft Docs • Azure Toolkit for IntelliJ - Tutorial - Azure Toolkit for IntelliJ (Spark application) - Azure Synapse Analytics | Microsoft Docs • Visual Studio Code - Tutorial - Spark & Hive Tools for VSCode (Spark application) - Azure Synapse Analytics | Microsoft Docs
Security	• Synapse uses Azure Active Directory (AAD) passthrough – Secure access credentials with Linked Services in Apache Spark for Azure Synapse Analytics - Azure Synapse Analytics | Microsoft Docs



Organizations often need to process large volumes of data before serving to key business stakeholders. In this tutorial, you will learn how to leverage the integrated experiences in Azure Synapse Analytics to process data using Apache Spark and later serve the data to end-users through Power BI and Serverless SQL.

- Azure Synapse Studio notebooks - Azure Synapse Analytics | Microsoft Docs

Apache Spark includes many language features to support preparation and processing of large volumes of data. Data preparations, transformations require the ability to visualize the results in a Synapse Studio notebook in Azure Synapse Analytics.

- Visualize data with Apache Spark - Azure Synapse Analytics | Microsoft Docs

You can enrich your data in Spark tables with new machine learning models that you train by using automated machine learning. In Azure Synapse Analytics, you can select a Spark table in the workspace to use as a training dataset for building machine learning models, and you can do this in a code-free experience.

- Tutorial: Train a model by using automated machine learning - Azure Synapse Analytics | Microsoft Docs
- Tutorial: Train a model in Python with automated machine learning - Azure Synapse Analytics | Microsoft Docs