

Data Engineering in the Cloud

Azure Synapse Analytics

Xuemao Zhang
East Stroudsburg University

January 18, 2025

Outline

- Introduction to Azure Synapse Analytics
- Lab: Creating an Azure Synapse Workspace

Introduction

- You are hired as an Azure data engineer. Your organization wants you to ingest, explore, prepare, and serve the data for instant BI and machine learning requirements to derive some decisions out of it.
 - ▶ For this task, you have decided to utilize Azure Synapse Analytics and Synapse Workspace, which is a complex mechanism.
- **Azure Synapse Analytics** is an integrated analytics service that accelerates time to insight across data warehouses and big data systems. It brings together data integration, enterprise data warehousing, and big data analytics to offer a single unified experience for ingesting, preparing, managing, and serving data for immediate business intelligence and machine learning needs.

Introduction

- It includes processes like data integration, enterprise data warehousing, and big data analytics.
- It can be used for data ingestion, exploration, management, and serving of data.
- In summary, Azure Synapse Analytics offers a unified platform for handling large-scale data querying and analytics, integrating various data processing and management capabilities to drive business intelligence and insights.

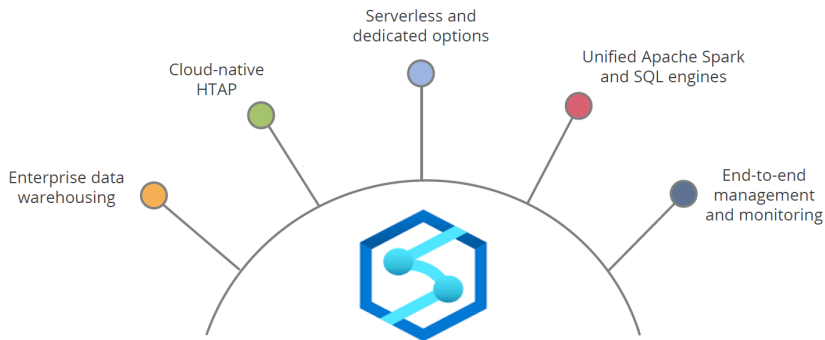
Introduction

- **Data Warehouse** is designed for **structured data** storage, retrieval, and analysis. It is optimized for querying and reporting.
- Enterprise Data Warehousing (EDW) is a **centralized repository** that consolidates data from different sources within an organization.
 - ▶ It is designed to support decision-making processes by providing a unified, consistent, and comprehensive view of the organization's data.
 - ▶ EDWs enable complex queries and analysis, facilitate reporting, and support business intelligence activities.

Use Cases of EDW

- Executive Reporting: Provides management with comprehensive reports and dashboards to monitor business performance.
- Operational Efficiency: Analyzes operational data to identify inefficiencies and areas for improvement.
- Customer Insights: Consolidates customer data to provide a 360-degree view, enabling personalized marketing and improved customer service.
- Financial Analysis: Aggregates financial data to support budgeting, forecasting, and financial reporting.
- Compliance and Auditing: Stores and manages data to ensure compliance with regulatory requirements and facilitate audits.

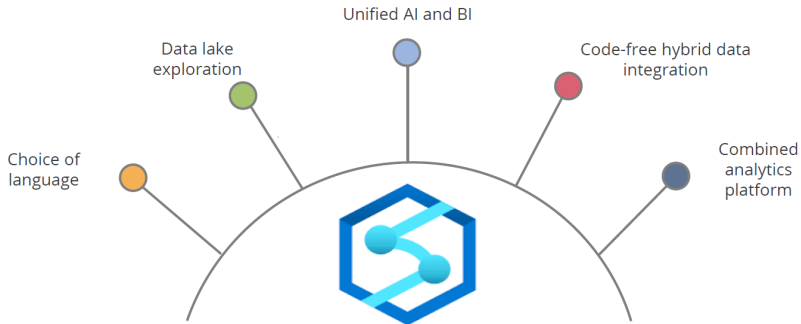
Azure Synapse Analytics: Features



Azure Synapse Analytics: Features

- Enterprise data warehousing: Robust, scalable data warehousing capabilities.
- Cloud native HTAP: Hybrid Transactional and Analytical Processing.
- Serverless and dedicated options: Flexible computing options to suit various workloads.
 - ▶ Serverless is a cloud computing model in which the cloud provider dynamically manages the allocation and provisioning of servers. In a serverless architecture, developers can focus on writing code without worrying about the underlying infrastructure.
- Unified Apache Spark and SQL engines: Seamless integration of big data and relational data.
- End to end management and monitoring: Comprehensive tools for data management and performance monitoring

Azure Synapse Analytics: Features



Azure Synapse Analytics: Features

- Choice of language: Support for multiple languages: SQL, Python, R, Scala.
- Data lake exploration: Efficient data exploration and analytics within the data lake.
- Unified AI and BI: Integration with Azure Machine Learning and Power BI.
- Code free hybrid data integration: Simplified data integration with visual tools.
- Combined analytics platform: A single platform for all data analytics needs.

Azure Synapse Workspace

- Azure Synapse workspace combines enterprise data warehousing and big data analytics.
- Microsoft's Azure Synapse Analytics service is a software as a service (SaaS). It can be utilized as required and consists of four parts:



Azure Synapse Workspace

- Connectors: Facilitate integration with various data sources.
- SQL Cluster: Enables scalable and high-performance SQL-based data warehousing.
- Spark: Supports big data processing and analytics with Apache Spark.
- SQL On Demand: Allows for serverless querying of data, enabling on-demand analytics without the need to provision resources.


Labs: Creating an Azure Synapse Workspace

- Prerequisites (Lab in Lecture03):
 - ▶ Create a datalake (storage account with hierarchical namespace enabled)
 - ▶ Create a container
 - ▶ Upload a data set to the container
 - ▶ Create a data base
- Objective: Creating an Azure Synapse Workspace

Lab 1: SQL server settings

- Open the database and turn on the server

Home >

 **SQLdatabase (serverxzhang2/SQLdatabase)** ✕ ☆ ...

SQL database

Search

Copy Restore Export Set server firewall Delete Connect with... Feedback

Overview

- Activity log
- Tags
- Diagnose and solve problems
- Query editor (preview)
- Settings
- Data management
 - Replicas
 - Sync to other databases
- Integrations
- Power Platform
- Security
- Intelligent Performance
- Monitoring


Essentials


Resource group (move)	: xzhang2	Server name	: serverxzhang2.database.windows.net
Status	: Online	Elastic pool	: No elastic pool
Location	: East US	Connection strings	: Show database connection strings
Subscription (move)	: Azure for Students	Pricing tier	: General Purpose: Gen5, 2 vCores
Subscription ID	: b5db3bdc-dae9-4e53-888f-8241b09dc786	Earliest restore point	: 2024-06-12 01:54 UTC
Tags (edit)	: Add tags		


Getting started Monitoring Properties Features Notifications (0) Integrations Tutorials

Start working with your database

Connect to your database and start working with data with a few simple steps. [Learn more](#)


Configure access



Connect to application


Start developing

Lab 1: SQL server settings

• Check the IAM

Home > SQLdatabase (serverxzhang2/SQLdatabase) > serverxzhang2

 **serverxzhang2** | Access control (IAM) ☆ ...
SQL server

Search ◊ « + Add ▾ ⬇ Download role assignments ≡ Edit columns ↻ Refresh | ✕ Remove | 🗨 Feedback

- Overview
- Activity log
- Access control (IAM)**
- Tags
- Quick start
- Diagnose and solve problems
 - Settings
 - Data management
 - Security
 - Intelligent Performance
 - Monitoring
 - Automation
 - Help

Check access | Role assignments | Roles | Deny assignments | Classic administrators

My access
View my level of access to this resource.
[View my access](#)

Check access
Review the level of access a user, group, service principal, or managed identity has to this resource. [Learn more](#)

[Check access](#)

Grant access to this resource

Grant access to resources by assigning a role.
[Learn more](#)

[Add role assignment](#)

View access to this resource

View the role assignments that grant access to this and other resources.
[Learn more](#)

[View](#)

View deny assignments

View the role assignments that deny access to this and other resources.
[Learn more](#)

[View](#)

Lab 1: SQL server settings

- Check the settings of the database

Home > serverxzhang2

serverxzhang2 | SQL databases ☆ ...

SQL server

Search

Search to filter databases...

Database	Status
SQLdatabase	Online

Showing 1 - 1 of 1 results.

- Overview
- Activity log
- Access control (IAM)
- Tags
- Quick start
- Diagnose and solve problems
- Settings
 - Microsoft Entra ID
 - SQL databases**
 - SQL elastic pools
 - DTU quota
 - Properties
 - Locks
- > Data management
- > Security
- > Intelligent Performance
- > Monitoring
- > Automation

Lab 1: SQL server settings

- Then you will see a similar screen

Home > serverxzhang2 | SQL databases >

SQL database (serverxzhang2/SQLdatabase)

SQL database

Search

Copy Restore Export Set server firewall Delete Connect with... Feedback

- Overview
- Activity log
- Tags
- Diagnose and solve problems
- Query editor (preview)
- Settings
- Data management
 - Replicas
 - Sync to other databases
- Integrations
- Power Platform
- Security
- Intelligent Performance
- Monitoring
- Automation

Essentials


Resource group	(move) : xzhang2	Server name	: serverxzhang2.database.windows.net
Status	: Online	Elastic pool	: No elastic pool
Location	: East US	Connection strings	: Show database connection strings
Subscription	(move) : Azure for Students	Pricing tier	: General Purpose: Gen5, 2 vCores
Subscription ID	: b5db3bdc-dae9-4e53-888f-8241b09dc786	Earliest restore point	: 2024-06-12 01:54 UTC
Tags	(edit) : Add tags		

Getting started

Monitoring Properties Features Notifications (0) Integrations Tutorials


Start working with your database

Connect to your database and start working with data with a few simple steps. [Learn more](#)




Configure access

Configure network access to your SQL database.



Connect to application

Use connecting strings to connect to your SQL database.



Start developing

Work in your database by using tools to add data.

Lab 1: SQL server settings

- Click on the server name, you will see

Home > SQLdatabase (servernzhang2/SQLdatabase) >

servernzhang2 SQL server

Search ◊ ◀ + Create database + New elastic pool + New dedicated SQL pool (formerly SQL DW) ↓ Import database ↻ Reset password → Move 🗑 Delete 💬 Feedback

Overview

- Activity log
- Access control (IAM)
- Tags
- Quick start
- Diagnose and solve problems
- Settings
- Microsoft Entra ID
- SQL databases
- SQL elastic pools
- DTU quota
- Properties
- Locks
- Data management
- Security
- Intelligent Performance
- Monitoring
- Automation
- Help

Essentials

Resource group (move)	: xzhang2	Server admin	: sqladmin
Status	: Available	Networking	: Show networking settings
Location	: East US	Microsoft Entra admin	: Not configured
Subscription (move)	: Azure for Students	Server name	: servernzhang2.database.windows.net
Subscription ID	: b5db3bdc-dae9-4e53-888f-8241b09dc786		
Tags (edit)	: Add tags		

Notifications (0) **Features (6)**

All Security (4) Performance (1) Recovery (1)

Microsoft Entra admin

Allows you to centrally manage identity and access to your Azure SQL databases.

NOT CONFIGURED

Microsoft Defender for SQL

Vulnerability Assessment and Advanced Threat Protection.

NOT CONFIGURED

Automatic tuning

Monitors and tunes your database automatically to optimize performance.

CONFIGURED

Audit

Track & log in

Failover groups

Automatically manages replication, connectivity and failover for a set of databases.

NOT CONFIGURED

Transparent data encryption

Encryption at rest for your databases, backups, and logs.

SERVICE-MANAGED KEY

Available resources Filter by name All times

Lab 1: SQL server settings

- Check data management
 - ▶ Failover group are a SQL server feature designed to manage replication, connectivity and failover of a set of databases.
 - ▶ If you want to create a Failover group, just click Add group
 - ★ Read [Configure a failover group for Azure SQL Database](#) for more information.

Home > SQLdatabase (serverxzhang2/SQLdatabase) > serverxzhang2

serverxzhang2 | Failover groups ☆ ...

SQL server

Search ◯ « + Add group ↻ Refresh 🗨 Feedback

Overview

Activity log

Access control (IAM)

Tags

Quick start

Diagnose and solve problems

Settings

Microsoft Entra ID

SQL databases

SQL elastic pools

DTU quota

Properties

Locks

Data management

Backups

Deleted databases

Failover groups

Want to enable disaster recovery at lower price? Create standby replica for General Purpose or Business Critical service tier, with savings of up to 40%. [Learn more](#)

Failover group are a SQL server feature designed to manage replication, connectivity and failover of a set of databases.

Name	Primary server	Secondary server	Read/Write failover
You have no group created			

Lab 1: SQL server settings

- Check Automatic tuning under Intelligent tuning
 - ▶ It helps optimize database performance by automatically applying performance tuning recommendations.

Home > serverxzhang2

serverxzhang2 | Automatic tuning ☆ ...

SQL server

Search ◦ ◀ ▶ Apply Revert to defaults

Overview

Activity log

Access control (IAM)

Tags

Quick start

Diagnose and solve problems

Settings

Data management

Security

Intelligent Performance

Automatic tuning ☆

Recommendations

Monitoring

Automation

Help




1 Azure SQL Database built-in intelligence automatically tunes your databases to optimize performance. Click here to learn more about automatic tuning.

Inherit from: ◯

Azure defaults Don't inherit

1 The database is inheriting automatic tuning configuration from Azure defaults.

Configure the automatic tuning options ◯

Option	Desired state			Current stat
 FORCE PLAN	ON	OFF	INHERIT	ON Inherited frc
 CREATE INDEX	ON	OFF	INHERIT	OFF Inherited frc
 DROP INDEX	ON	OFF	INHERIT	OFF Inherited frc

Lab 2: Creating an Azure Synapse Workspace

- Now go back to Home and search for Synapse, then choose Azure Synapse Analytics → Create Synapse workspace
- Basics
 - ▶ Specify a Workspace name. (Mine is synapse611)
 - ▶ Let's add a new container called rawdata

[Home](#) > [Azure Synapse Analytics](#) >

Create Synapse workspace ...

*** Basics** * Security Networking Tags Review + create

Create a Synapse workspace to develop an enterprise analytics solution in just a few clicks.

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all of your resources.

Subscription *	<input type="text" value="Azure for Students"/>
Resource group *	<input type="text" value="xzhang2"/>
	Create new
Managed resource group	<input type="text" value="Enter managed resource group name"/>

Workspace details

Name your workspace, select a location, and choose a primary Data Lake Storage Gen2 file system to serve as the default location for logs and job output.

Workspace name *	<input type="text" value="Enter workspace name"/>
Region *	<input type="text" value="East US"/>

Lab 2: Creating an Azure Synapse Workspace

- In the security tab, specify the password (username is sqladminuser)
- keep the other settings to default and go to review + create

[Home](#) > [Azure Synapse Analytics](#) >

Create Synapse workspace ...

✓ Validation succeeded

* Basics * Security Networking Tags Review + create

Product Details

Azure Synapse Analytics workspace
by Microsoft
[Terms of use](#) | [Privacy policy](#)

Serverless SQL est. cost/TB ⓘ
\$5.00 USD

Terms



By clicking Create, I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to bill my current payment method for the fees associated with the offering(s), with the same billing frequency as my Azure subscription; and (c) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. For additional details see [Azure Marketplace Terms](#). ⓘ

Basics








Subscription	Azure for Students
Resource group	xzhang2
Region	East US
Workspace name	(new) synapse611
Data Lake Storage Gen2 account	https://azdatalake611.dfs.core.windows.net
Data Lake Storage Gen2 file system	(new) rawdata
Managed resource group	None

Lab 2: Creating an Azure Synapse Workspace

[Home](#) >

 **Microsoft.Azure.SynapseAnalytics-20240612131259** | Overview  ...

Deployment


   Delete  Cancel  Redeploy  Download  Refresh

 Overview

 Inputs


 Outputs











 Template

 Your deployment is complete

 Deployment name : Microsoft.Azure.SynapseAnalytics-20240612131259
Subscription : [Azure for Students](#)
Resource group : [xzhang2](#)

Start time : 6/12/2024, 1:17:53 PM
Correlation ID : d44e693a-b410-47ed-b06e-fd6ba43d7816

 Deployment details

Resource	Type	Status	Operation details
 StorageRoleDeploymentResource	 Deployment	OK	Operation details
 synapse611/allowAll	 Microsoft.Synapse/workspaces/f	OK	Operation details
 synapse611	 Synapse workspace	OK	Operation details
 synapse611	 Synapse workspace	OK	Operation details
 rawdata	 Deployment	OK	Operation details

 Next steps

[Go to resource group](#)

Lab 2: Creating an Azure Synapse Workspace

- Once the deployment is completed, click on the go to resource button
- Then, open the Synapse workspace we just created
- Now in the synapse workspace, go to Analytic pools

Home > synapse611
synapse workspace

Search

+ New dedicated SQL pool + New Apache Spark pool + New Data Explorer pool (preview) Refresh Reset SQL admin password Delete

Essentials

Resource group (rg) : [azbap92](#)

Status : Succeeded

Location : East US

Subscription (mgmt) : [Azure for Students](#)

Subscription ID : b5db3bdc-dae9-4e53-888f-8241b09dc786

Managed virtual network : No

Managed Identity object ... : 01337133-9f5e-4f89-a86b-8f0de2cda5f4

Workspace web URL : <https://web.azure.synapse.com/Workspace%3D%27b5db3bdc-dae9-4e53-888f-8241b09dc786%27>

Tags (add) : [Add tags](#)

Networking : [Show firewall settings](#)

Primary ADLS Gen2 acco... : <https://azdatalake611.dfs.core.windows.net>

Primary ADLS Gen2 file s... : rawdata

SQL admin username : sqladminuser

SQL Microsoft Entra admin : azbap92@asu.edu

Dedicated SQL endpoint : synapse611.sql.azure.synapse.net

Serverless SQL endpoint : synapse611-ondemand.sql.azure.synapse.net

Development endpoint : <https://synapse611.dev.azure.synapse.net>

Getting started

Open Synapse Studio

Start building your fully-integrated analytics solution and unlock new insights.

[Open it](#)

Read documentation

Learn how to be productive quickly. Explore concepts, tutorials, and samples.

[Learn more](#)

Analytics pools

Search to filter items...

Name	Type	Size
------	------	------

Lab 2: Creating an Azure Synapse Workspace

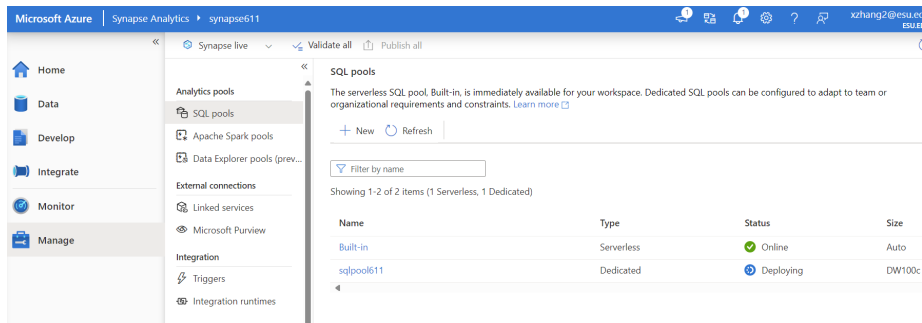
- Click on Open in Open Synapse Studio box
 - In the Synapse Studio, you can see the SQL pools

The screenshot displays the Microsoft Azure Synapse Studio interface. The top navigation bar shows 'Microsoft Azure | Synapse Analytics > synapse611'. A cookie consent banner is visible below the navigation bar. The left sidebar contains a menu with options: Home, Data, Develop, Integrate, Monitor (selected), and Manage. The main content area is titled 'SQL pools' and includes a 'Refresh' button and an 'Edit columns' link. Below these, there is a filter dropdown set to 'All'. The text 'Showing 1 - 1 of 1 items' is displayed above a table. The table has four columns: 'Pool name', 'Type', 'Version', and 'Status'. It contains one row with the values 'Built-in', 'Serverless', 'v2', and 'Online' (indicated by a green checkmark icon).

Pool name	Type	Version	Status
Built-in	Serverless	v2	Online

Lab 2: Creating an Azure Synapse Workspace

- The serverless SQL pool, Built-in, is immediately available for your workspace. Dedicated SQL pools can be configured to adapt to team or organizational requirements and constraints.
- Let's add a new Dedicated (serverless) SQL pool
 - ▶ Specify the name for the SQL pool.
 - ▶ Choose the performance level. **Note:** the pricing will depend on the performance level being selected. Set the value for Performance Level as 'DW100c'.



The screenshot displays the Microsoft Azure Synapse Analytics portal. The left sidebar contains navigation links: Home, Data, Develop, Integrate, Monitor, and Manage. The main content area is titled 'SQL pools' and includes a description: 'The serverless SQL pool, Built-in, is immediately available for your workspace. Dedicated SQL pools can be configured to adapt to team or organizational requirements and constraints. [Learn more](#)'. Below this, there are '+ New' and 'Refresh' buttons. A 'Filter by name' input field is present. The text 'Showing 1-2 of 2 items (1 Serverless, 1 Dedicated)' is displayed above a table. The table has four columns: Name, Type, Status, and Size. It lists two items: 'Built-in' (Serverless, Online, Auto) and 'sqlpool611' (Dedicated, Deploying, DW100c).

Name	Type	Status	Size
Built-in	Serverless	Online	Auto
sqlpool611	Dedicated	Deploying	DW100c

Lab 2: Creating an Azure Synapse Workspace

- SQL pool is the method for querying data in Azure Synapse.
 - ▶ The data is stored in a relational table structure with column storage.
- The serverless SQL pool is defined as the mechanism of querying data for a data lake.
 - ▶ The storage and serverless SQL pool must be in the **same region**.

Lab 2: Creating an Azure Synapse Workspace

- Dedicated SQL pool provides a dedicated set of resources for data warehousing. It's a fully managed, enterprise-grade cloud data warehouse.
- Dedicated SQL pool is suitable for consistent, large-scale data warehousing needs, offering high performance and dedicated resources at a higher cost and with more management overhead.
- Serverless SQL Pool is ideal for ad-hoc, on-demand querying, and data exploration without the need to manage resources. Cost-effective for intermittent or unpredictable workloads.

Lab 2: Creating an Azure Synapse Workspace

- The following is what we have done so far

Azure services



Create a
resource



Azure Synapse
Analytics



Cost alerts



SQL databases



Storage
accounts



Quickstart
Center



Azure AI
services



Kubernetes
services








Virtual
machines



More service

Resources

Recent Favorite

Name	Type	Last Viewed
 azdatalake611	Storage account	a few seconds ago
 synapse611	Synapse workspace	a few seconds ago
 xzhang2	Resource group	2 hours ago
 serverxzhang2	SQL server	3 hours ago
 SQLdatabase (serverxzhang2/SQLdatabase)	SQL database	3 hours ago

See all

Lab 2: Creating an Azure Synapse Workspace

- Delete everything we created.

License



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](#).