

12/09/2023

Azure Synapse Analytics SQL Data Warehouse

Azure Synapse Mapping Data Flow

Azure Synapse Pipeline

Data Ingestion and preparation

Data Transformation and Loading

Query Performance Optimization Techniques

- Data Distribution
- Indexing
- Statistics
- Data compression
- Partitioning
- Query design
- Monitoring and Tuning
- Data Movement
- Resource Scaling
- Caching

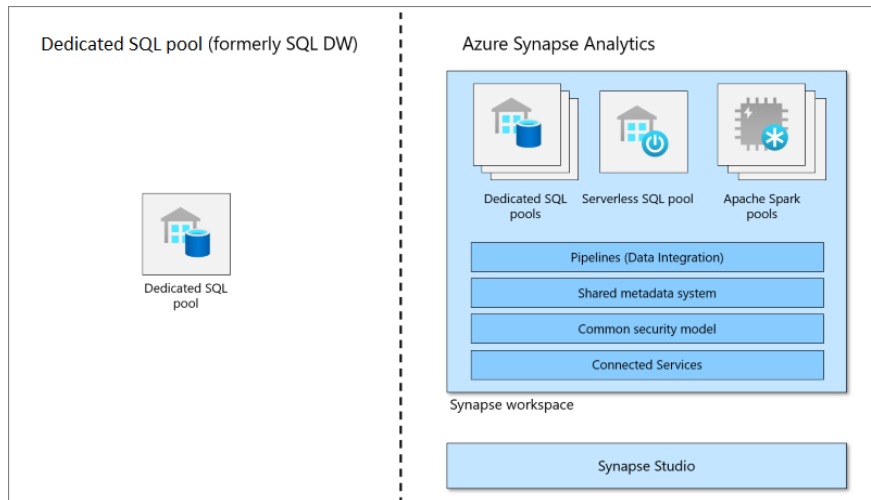
Serverless SQL Pool in Azure Synapse Analytics

Every Azure Synapse Analytics workspace comes with **serverless SQL pool endpoints** that you can use to query data in the Azure Data Lake (Parquet, Delta Lake, Delimited text formats), Azure Cosmos DB, or Dataverse.

Serverless SQL pool is a query service over the data in your data lake. It enables you to access your data through the following functionalities:

- A familiar T-SQL syntax to query data in place without the need to copy or load data into a specialized store. To learn more, see the T-SQL support section.
- Integrated connectivity via the T-SQL interface that offers a wide range of business intelligence and ad-hoc querying tools, including the most popular drivers. To learn more, see the Client tools section.

Dedicated SQL Pool (formerly SQL DW) refers to the enterprise data warehousing feature that are available in the **Azure Synapse Analytics**.



Dedicated SQL Pool represents a collection of analytic resources that are provisioned when using Synapse SQL. The size of a dedicated SQL pool (formerly SQL DW) is determined by Data Warehousing Units (DWU).

Difference between Serverless SQL pool and Dedicated SQL Pool:

1. While serverless doesn't store data, only access data from storage accounts and scale the MPP environment automatically, the dedicated SQL Pool keeps a static number of servers according to the service level we choose and a constant number of distributions - always 60.

Serverless SQL Pools:

- Idle for either structured or unstructured data. Data can be stored in different file formats(raw or parquet)
- The compute associated with the serverless pool is managed and allocated by microsoft
- Seamless data transformation without any infrastructure in place.

Dedicated SQL pools:

- Data is stored in relational tables
- If you need an optimized compute strategy for performance requirements, go with Dedicated(option to increase the Data warehouse units)
- Require dedicated servers before performing any operation
- The performance is based on DWU (Data Warehouse Units)

Create Synapse workspace ...

Subscription * ⓘ

Resource group * ⓘ [Create new](#)

Managed resource group ⓘ

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 * ✓

Region *

Select Data Lake Storage Gen2 * ⓘ ☒ From subscription ☐ Manually via URL

Account name * ⓘ [Create new](#)

File system name * [Create new](#)

[Review + create](#)

< Previous

Next: Security >

Create Synapse workspace ...

Configure security options for your workspace.

Authentication

Choose the authentication method for access to workspace resources such as SQL pools. The authentication method can be changed later on. [Learn more](#)

Authentication method ⓘ ☒ Use both local and Azure Active Directory (Azure AD) authentication ☐ Use only Azure Active Directory (Azure AD) authentication

SQL Server admin login * ⓘ

SQL Password ⓘ ✓

Confirm password ✓

System assigned managed identity permission

Select to grant the workspace network access to the Data Lake Storage Gen2 account using the workspace system identity. [Learn more](#)

☐ Allow network access to Data Lake Storage Gen2 account. ⓘ

i The selected Data Lake Storage Gen2 account does not restrict network access using any network access rules, or you selected a storage account manually via URL under Basics tab. [Learn more](#)

[Review + create](#)

< Previous

Next: Networking >

Microsoft Azure

Search resources, services, and docs (G+)

pioneer23@ravinsofttec...
RAVINSOFTTECH (RAVINSOFTTE...)

Home > Azure Synapse Analytics >

Create Synapse workspace

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

Configure networking options for your workspace.

Managed virtual network
Choose whether to set up a dedicated Azure Synapse-managed virtual network for your workspace. [Learn more](#)

Managed virtual network ☐ Enable ☒ Disable

Create managed private endpoint to primary storage account ☒ Yes ☐ No

Allow outbound data traffic only to approved targets ☐ Yes ☒ No

Public network access
Choose whether to permit public network access to your workspace. You can modify the firewall rules after you enable this setting. [Learn more](#)

Public network access to workspace endpoints ☒ Enable ☐ Disable

Review + create

< Previous

Next: Tags >

Microsoft Azure

Search resources, services, and docs (G+)

pioneer23@ravinsofttec...
RAVINSOFTTECH (RAVINSOFTTE...)

Home >

Microsoft.Azure.SynapseAnalytics-20230912112618 | Overview

Deployment

Search

Delete Cancel Redeploy Download Refresh

Overview

Inputs

Outputs

Template

✓ Your deployment is complete

Deployment name : Microsoft.Azure.Synapse... Start time : 9/12/2023, 11:28:58 AM
Subscription : RDBF Correlation ID : 2dc47062-9341-43af-a19...
Resource group : RDBF

Deployment details

Resource	Type	Status
juhi-workspace/alk	Microsoft.Synapse/workspaces/fi	OK
juhi-workspace	Synapse workspace	OK
juhi-workspace	Synapse workspace	OK
demofile	Deployment	OK

Next steps

Go to resource group

Cost management

Get notified to stay within your budget and prevent unexpected charges on your bill.
[Set up cost alerts >](#)

Microsoft Defender for Cloud

Secure your apps and infrastructure
[Go to Microsoft Defender for Cloud >](#)

Free Microsoft tutorials

[Start learning today >](#)

Work with an expert

Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support.
[Find an Azure expert >](#)

Microsoft Azure | Synapse Analytics | juhi-workspace | Search

We use optional cookies to provide a better experience. [Learn more](#)

Synapse live | Validate all | Publish all

Develop

Filter resources by name

Data flows 1

Dataflow1

source1

Columns: 0 total

Add Source

Source settings | Source options

Dataset *

+ New

Options

New linked service

Azure Blob Storage [Learn more](#)

☒ From Azure subscription ☐ Enter manually

Azure subscription [ⓘ]

RDBF (5464fa55-993e-42e2-b3a4-206cba24f19e)

Storage account name *

juhistorageacc

Managed private endpoint

Not available

[Create new](#)

Additional connection properties

+ New

Test connection [ⓘ]

☒ To linked service ☐ To file path

Annotations

+ New

> Parameters

> Advanced [ⓘ]

Create Cancel

[Test connection](#)

Microsoft Azure | Synapse Analytics | juhi-workspace | Search

We use optional cookies to provide a better experience. [Learn more](#)

Synapse live | Validate all | Publish all

Develop

Filter resources by name

Data flows 1

Dataflow1

source1

Columns: 0 total

Add Source

Source settings | Source options

Dataset *

+ New

Options

Set properties

Name

Parquet1

Linked service *

AzureBlobStorage1

Connect via integration runtime * [ⓘ]

☒ AutoResolveIntegrationRuntime (Managed Virtual Network)

☒ Interactive authoring enabled [ⓘ]

File path

demofile / Directory / NYCTripSmall.parquet

Import schema

☒ From connection/store ☐ From sample file ☐ None

> Advanced

OK Back Cancel

Microsoft Azure

Synapse Analytics

juhi-workspace

Search

1

4

?

pioneer23@ravinsofttech.com

RAVINSOFTTECH

We use optional cookies to provide a better experience. [Learn more](#)

Accept

Reject

More options

Synapse live

Validate all

Publish all 3

Develop

Filter resources by name

Data flows

Dataflow1

Dataflow1

Validate

Data flow debug

source1

Import data from Parquet1

+

sink1

Columns: 23 total

Add Source

Sink

Settings

Errors

Mapping

Optimize

Inspect

Data preview

Sink type *

Integration dataset

Inline

Workspace DB

Cache

Dataset *

Sink

Test connection

Open

New

Skip line count

Options

Allow schema drift

Microsoft Azure

Synapse Analytics

juhi-workspace

Search

1

3

?

pioneer23@ravinsofttech.com

RAVINSOFTTECH

We use optional cookies to provide a better experience. [Learn more](#)

Synapse live

Validate all

Publishing 3

Develop

Filter resources by name

Data flows

Dataflow1

Dataflow1

Validate

Data flow debug

source1

Import data from Parquet1

+

Add Source

Sink

Settings

Errors

Mapping

Sink type *

Dataset *

Skip line count

Options

Publish all

You are about to publish all pending changes to the live environment. [Learn more](#)

Pending changes (3)

NAME	CHANGE	EXISTING
Parquet1	(New)	-
Sink	(New)	-
Dataflow1	(New)	-

Publish

Cancel

Microsoft Azure | Synapse Analytics | rdblsynapse | Search | pioneer23@ravinssofttech.com | RAVINSOFTTECH

Synapse live | Validate all | Publish all 1

Develop

Filter resources by name

- SQL scripts 1
 - SQL script 1

```
35
36 COPY INTO dbo.NYCTaxiTripSmall
37 (DateID 1, MedallionID 2, HackneyLicenseID 3, PickupTimeID 4, DropoffTimeID 5,
38 PickupGeographyID 6, DropoffGeographyID 7, PickupLatitude 8, PickupLongitude 9,
39 PickupLatLong 10, DropoffLatitude 11, DropoffLongitude 12, DropoffLatLong 13,
40 PassengerCount 14, TripDurationSeconds 15, TripDistanceMiles 16, PaymentType 17,
41 FareAmount 18, SurchargeAmount 19, TaxAmount 20, TipAmount 21, TollsAmount 22,
42 TotalAmount 23)
43 FROM 'https://rdblsynapse.dfs.core.windows.net/rdblsynapse/NYCTripSmall.parquet'
44 WITH
45 (
46 FILE_TYPE = 'PARQUET'
47 ,MAXERRORS = 0
48 ,IDENTITY_INSERT = 'OFF'
49 )
```

Results Messages

No results to show
Your query yielded no displayable results

00:00:18 Query executed successfully.

Microsoft Azure | Synapse Analytics | rdblsynapse | Search | pioneer23@ravinssofttech.com | RAVINSOFTTECH

Synapse live | Validate all | Publish all 2

Data

Workspace Linked

Filter resources by name

- SQL database 17
 - DedicatPool (SQL)
 - dedsq!pool (SQL)
 - gowreemanoharsqldb (SQL)
 - gowreemanoharsq!pool (SQL)
 - juhipool (SQL)
 - Tables
 - dbo.NYCTaxiTripSmall

```
1 SELECT TOP (100) [DateID]
2 ,[MedallionID]
3 ,[HackneyLicenseID]
4 ,[PickupTimeID]
5 ,[DropoffTimeID]
6 ,[PickupGeographyID]
7 ,[DropoffGeographyID]
8 ,[PickupLatitude]
9 ,[PickupLongitude]
10 ,[PickupLatLong]
11 ,[DropoffLatitude]
12 ,[DropoffLongitude]
13 ,[DropoffLatLong]
14 ,[PassengerCount]
```

Results Messages

Columns

- DateID (int, null)
- MedallionID (int, null)
- HackneyLicenseID (int, null)
- PickupTimeID (int, null)
- DropoffTimeID (int, null)
- PickupGeographyID (int, null)

Actions: New SQL script, New notebook, New data flow, New integration dataset, Refresh

20131228

Licens...	PickupTimeID	DropoffTimeID	PickupGeogra...	DropoffGeogr.
71681	71979	2834	175410	
31320	31860	254625	81484	

00:00:02 Query executed successfully.

Microsoft Azure | Synapse Analytics | rdblsynapse

Synapse live | Validate all | Publish all

Data

Workspace Linked

Filter resources by name

SQL database 17

- DedicatPool (SQL)
- dedsqpool (SQL)
- gowreemanoharsqldb (SQL)
- gowreemanoharsqpool (SQL)
- juhipool (SQL)
 - Tables
 - dbo.NYCTaxiTripSmall
 - Columns
 - DatelID (int, null)
 - MedallionID (int, null)
 - HackneyLicenseID (int, null)
 - PickupTimeID (int, null)
 - DropoffTimeID (int, null)
 - PickupGeographyID (int, null)

SQL script 1

```
1 SELECT PassengerCount,
2 SUM(TripDistanceMiles) as SumTripDistance,
3 AVG(TripDistanceMiles) as AvgTripDistance
4 INTO dbo.PassengerCountStats
5 FROM dbo.NYCTaxiTripSmall
6 WHERE TripDistanceMiles > 0 AND PassengerCount > 0
7 GROUP BY PassengerCount;
8 SELECT * FROM dbo.PassengerCountStats
9 ORDER BY PassengerCount;
```

Results Messages

View Table Chart Export results

PassengerCount	SumTripDistance	AvgTripDistance
1	190167.34	2.87357339297047
2	49399.74	3.13072691552063

00:00:08 Query executed successfully.

Microsoft Azure | Synapse Analytics | rdblsynapse

Synapse live | Validate all | Publish all

Data

Workspace Linked

Filter resources by name

SQL database 17

- DedicatPool (SQL)
- dedsqpool (SQL)
- gowreemanoharsqldb (SQL)
- gowreemanoharsqpool (SQL)
- juhipool (SQL)
 - Tables
 - dbo.NYCTaxiTripSmall
 - Columns
 - DatelID (int, null)
 - MedallionID (int, null)
 - HackneyLicenseID (int, null)
 - PickupTimeID (int, null)
 - DropoffTimeID (int, null)
 - PickupGeographyID (int, null)

SQL script 1

```
1 SELECT PassengerCount,
2 SUM(TripDistanceMiles)
3 AVG(TripDistanceMiles)
4 INTO dbo.PassengerCountStats
5 FROM dbo.NYCTaxiTripSmall
6 WHERE TripDistanceMiles
7 GROUP BY PassengerCount;
8 SELECT * FROM dbo.PassengerCountStats
9 ORDER BY PassengerCount;
```

Results Messages

View Table Chart

Search

PassengerCount

1

2

00:00:08 Query executed successfully.

Set properties

Name
AzureSynapseAnalyticsTable1

Linked service *
rdblsynapse-WorkspaceDefaultSqlServer

Connect via integration runtime * ⓘ
AutoResolveIntegrationRuntime

Table name
Select...
dbo.NYCTaxiTripSmall
dbo.PassengerCountStats
From connection/store ☐ None ☒

OK Back Cancel

Microsoft Azure

Synapse Analytics

rdblsynapse

Search

1

2

?

pioneer23@ravinsofttech.com

RAVINSOFTTECH

Synapse live

Validate all

Publish all

SQL script 2

AzureSynapseAnalyt...

Dataflow1

AzureSynapseAnalyt...

Develop

Filter resources by name

SQL scripts2

SQL script 1

SQL script 2

Data flows1

Dataflow1

Azure Synapse Analytics

AzureSynapseAnalyticsTable2

Connection

Schema

Parameters

Linked service *

rdblsynapse-WorkspaceDefaultSqlSe...

Test connection

Edit

New

Learn more

Linked service properties

Name	Value	Type
DBName	Value	String

Integration runtime *

AutoResolveIntegrationRuntime

Edit

Table

dbo

target

Edit

Preview data