



Creating SQL Pool for Parquet Data

Azure SQL Pool, formerly known as SQL Data Warehouse, is a cloud-based data warehousing service provided by Microsoft Azure. It's part of the Azure Synapse Analytics suite, which includes various data integration, analytics, and visualization tools.

Azure SQL Pool is designed to handle large volumes of data and complex queries by distributing data across multiple nodes. It uses a massively parallel processing (MPP) architecture to distribute and process data across multiple compute nodes, allowing for high performance and scalability.

Key features of Azure SQL Pool include:

1. **Massively Parallel Processing (MPP):** Data is distributed across multiple compute nodes, enabling parallel processing of queries for improved performance.
2. **Columnar Storage:** Data is stored in a columnar format, which enhances query performance, especially for analytics workloads that involve aggregations and scans of large datasets.
3. **Integration with Azure Ecosystem:** Azure SQL Pool integrates seamlessly with other Azure services such as Azure Data Factory, Azure Databricks, Azure Synapse Studio, and Power BI, enabling end-to-end data analytics and reporting solutions.
4. **Scalability:** Azure SQL Pool allows you to scale compute resources up or down based on your workload requirements, providing flexibility and cost optimization.
5. **Security and Compliance:** It offers robust security features including encryption, authentication, role-based access control (RBAC), and compliance certifications to meet regulatory requirements.
6. **Data Loading and Integration:** Azure SQL Pool supports various data loading techniques including PolyBase, Azure Data Factory, and Azure Databricks, allowing you to ingest data from different sources into your data warehouse.



Use cases of SQL Pool:

Azure SQL Pool (formerly SQL Data Warehouse) is a powerful tool for handling large-scale data warehousing and analytics workloads in the cloud. Here are some common use cases where Azure SQL Pool can be beneficial:

1. **Enterprise Data Warehousing:** Azure SQL Pool is well-suited for building enterprise data warehouses that consolidate and manage large volumes of data from multiple sources. It allows organizations to store structured, semi-structured, and unstructured data in a centralized repository for analysis and reporting.
2. **Big Data Analytics:** Organizations dealing with massive datasets can leverage Azure SQL Pool's parallel processing capabilities to perform complex analytics tasks such as ad-hoc queries, data mining, predictive analytics, and machine learning on large volumes of data.
3. **Business Intelligence and Reporting:** Azure SQL Pool integrates seamlessly with business intelligence (BI) and reporting tools such as Power BI, Tableau, and Microsoft




Excel. It enables users to create interactive dashboards, generate reports, and gain insights from data stored in the data warehouse.

4. **Real-time Analytics:** Azure SQL Pool supports near real-time data ingestion and analytics, allowing organizations to analyze streaming data from sources like IoT devices, sensors, and social media in real-time. This enables timely decision-making and actionable insights based on the latest data.
5. **Data Exploration and Discovery:** Data scientists and analysts can use Azure SQL Pool to explore and discover patterns, trends, and correlations in large datasets. With its scalable architecture and support for advanced analytics functions, Azure SQL Pool facilitates exploratory data analysis and hypothesis testing.
6. **Data Integration and Transformation:** Azure SQL Pool supports integration with Azure Data Factory, Azure Databricks, and other data integration services, enabling organizations to ingest, transform, and load data from various sources into the data warehouse. This facilitates data integration and consolidation for reporting and analysis purposes.
7. **Advanced Analytics and Machine Learning:** Azure SQL Pool can be integrated with Azure Machine Learning and other advanced analytics platforms to build and deploy machine learning models directly within the data warehouse environment. This enables organizations to derive predictive insights and drive intelligent decision-making based on data-driven models.
8. **Cost-effective Data Storage and Processing:** Azure SQL Pool offers a pay-as-you-go pricing model, allowing organizations to scale compute resources up or down based on workload demands. This helps optimize costs by only paying for the resources consumed, making it cost-effective for both small and large-scale data warehousing projects.

To begin with the Lab:

1. Now we're going to look at the other compute option that's available with Azure Synapse and that is the SQL pool. With the help of the SQL pool, you can now build tables in your SQL data warehouse.
2. Up till now with the serverless SQL pool, we could only build external tables wherein the structure of the table is defined in Azure Synapse whereas the data itself resides in an Azure storage account. But if you want to persist data within Azure Synapse itself, you know this makes it much more efficient when you're working with large data tables, and if you want to perform analysis across those tables, then you should build a SQL data warehouse with the help of the dedicated SQL pool.
3. Now in your Azure Portal you have to go to your Synapse Workspace and from the left pane you will see an option for SQL Pools click on it.

Analytics pools

-  **SQL pools**
-  Apache Spark pools
-  Data Explorer pools (preview)

- Here you can see your built-in SQL pool which is serverless. Now you have to click on the plus icon and create your dedicated SQL Pool.

+ New	Refresh	Assign tags	Delete
<input type="text" value="Search to filter items..."/>			
Name	↑↓ Type	↑↓ Status	↑↓ Size
<input type="checkbox"/> Built-in	Serverless	N/A	Auto

- Now here you just need to give it a name and for the performance level keep it to the lowest.
- Then you can see the pricing mentioned which means that this service is quite expensive. So, we need to be very careful while working with this service.

*** Basics** * Additional settings Tags Review + create

Create a dedicated SQL pool with your preferred configurations. Complete the Basics tab then go to Review + Create to provision with smart defaults, or visit each tab to customize. [Learn more](#)

Dedicated SQL pool details

Name your dedicated SQL pool and choose its initial settings.

Dedicated SQL pool name *	<input type="text" value="demopooldb"/>
Performance level ⓘ	<div><div></div><div>DW100c</div><div>DW100c</div></div>
Estimated price ⓘ	<div>Est. Cost Per Hour 115.64 INR View pricing details</div>

- Once it is done then move to the review page and create your SQL Pool. Then wait till the deployment is completed.
- Now you need to go back to your Synapse Workspace Studio and on data in your workspace you will see two databases one is your serverless database which is demodb and the other one is your demopooldb which is your dedicated SQL pool.

Data



Workspace

Linked

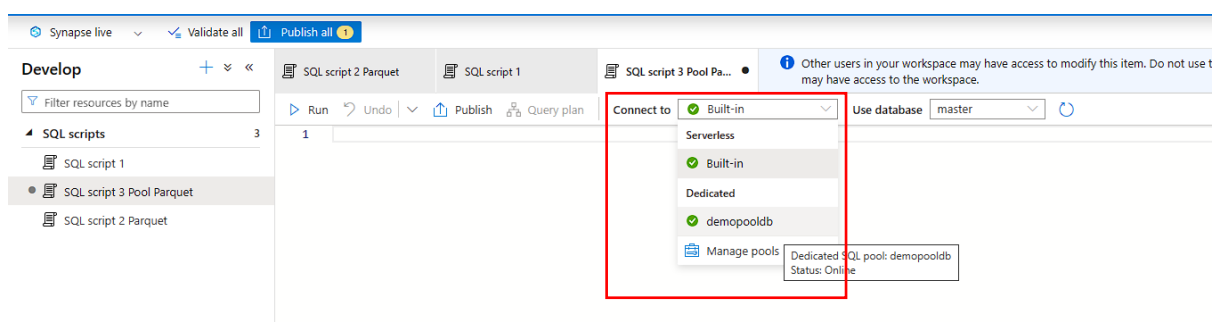
Filter resources by name

SQL database

2

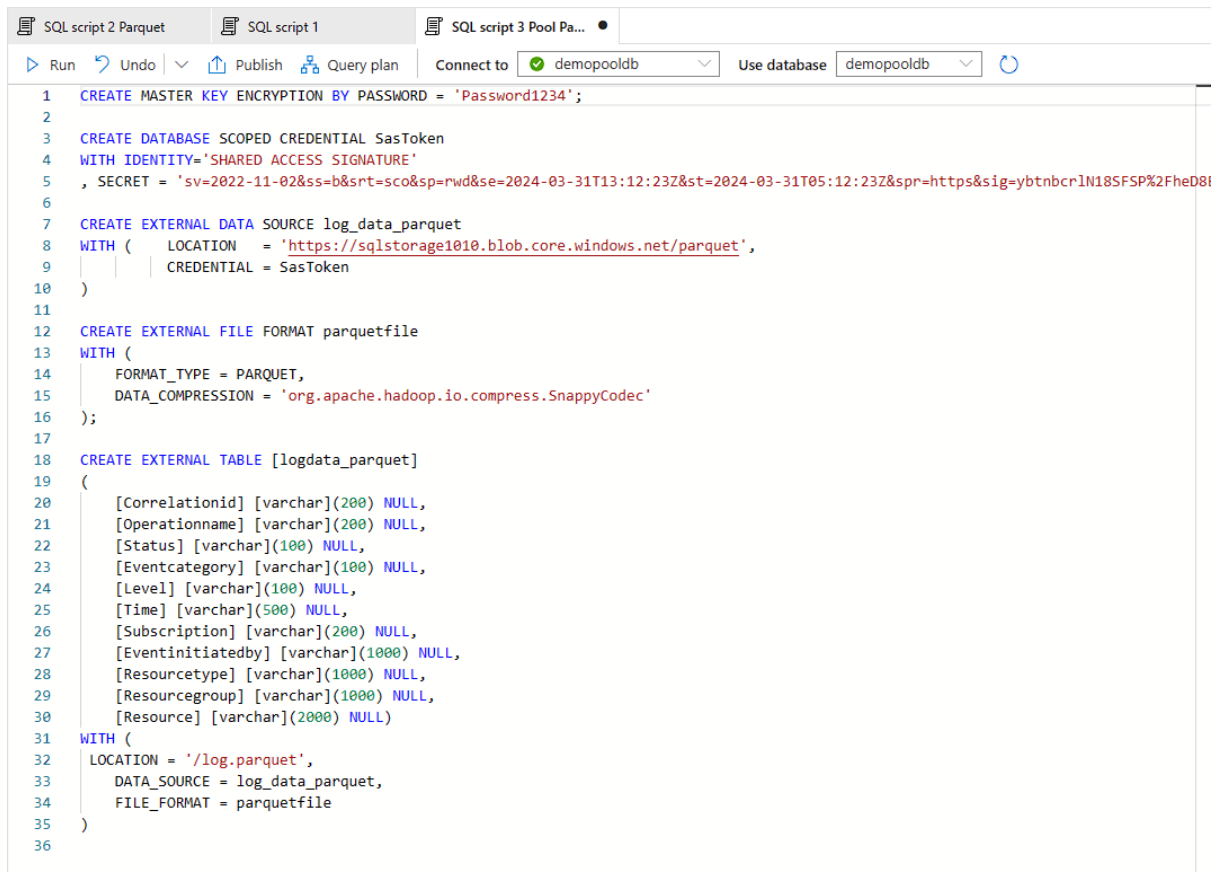
- demodb (SQL)
 - External tables
 - External resources
 - Views
 - Schemas
 - Security
- demopooldb (SQL)
 - Tables
 - External tables
 - External resources
 - Views
 - Programmability
 - Schemas
 - Security

9. Now in your develop section create a new SQL Script. In there, you have to change the connection from built-in to demopooldb and your database will change automatically.



10. Now while creating the external table most of the code will be same as we'd seen earlier on.
11. Here first thing is that we don't need to create a database. But we are going to master key.
12. Then we have to create SAS token, after that external data source and external file format.

13. If in case your SAS token has expired then you can create a new one or you can simply copy the command from previous script and paste it here.
14. Then we are going to create our external table and run it as usual.
15. Once you have pasted all the data then you can execute them one by one and get the results.



```
1 CREATE MASTER KEY ENCRYPTION BY PASSWORD = 'Password1234';
2
3 CREATE DATABASE SCOPED CREDENTIAL SasToken
4 WITH IDENTITY='SHARED ACCESS SIGNATURE'
5 , SECRET = 'sv=2022-11-02&ss=b&srt=sco&sp=rwd&se=2024-03-31T13:12:23Z&st=2024-03-31T05:12:23Z&spr=https&sig=ybtncbrlN18SFSP%2FheD8!
6
7 CREATE EXTERNAL DATA SOURCE log_data_parquet
8 WITH (
9     LOCATION = 'https://sqlstorage1010.blob.core.windows.net/parquet',
10    CREDENTIAL = SasToken
11 )
12
13 CREATE EXTERNAL FILE FORMAT parquetfile
14 WITH (
15     FORMAT_TYPE = PARQUET,
16     DATA_COMPRESSION = 'org.apache.hadoop.io.compress.SnappyCodec'
17 );
18
19 CREATE EXTERNAL TABLE [logdata_parquet]
20 (
21     [Correlationid] [varchar](200) NULL,
22     [Operationname] [varchar](200) NULL,
23     [Status] [varchar](100) NULL,
24     [Eventcategory] [varchar](100) NULL,
25     [Level] [varchar](100) NULL,
26     [Time] [varchar](500) NULL,
27     [Subscription] [varchar](200) NULL,
28     [Eventinitiatedby] [varchar](1000) NULL,
29     [Resourcetype] [varchar](1000) NULL,
30     [Resourcegroup] [varchar](1000) NULL,
31     [Resource] [varchar](2000) NULL)
32 WITH (
33     LOCATION = '/log.parquet',
34     DATA_SOURCE = log_data_parquet,
35     FILE_FORMAT = parquetfile
36 )
```

SQL script 2 ParquetSQL script 1SQL script 3 Pool Pa...

RunUndoPublishQuery plan

Connect to demopoolddbUse database demopoolddb

35)

36

37 SELECT*FROM [logdata_parquet]

38

ResultsMessages

ViewTableChartExport results

Search

Correlationid	Operationname	Status	Eventcategory	Level	Time	Subscription
c6f121ec-3674-...	'audit' Policy ac...	Succeeded	Policy	Warning	2023-04-10T13:...	6912d7a0-bc28...
9da21fd9-22d3-...	'auditIfNotExist...	Succeeded	Policy	Informational	2023-04-10T13:...	6912d7a0-bc28...
65b2093f-1ef5-...	Delete Disk	Failed	Administrative	Error	2023-04-10T13:...	6912d7a0-bc28...
39356cb1-ec3e-...	'audit' Policy ac...	Succeeded	Policy	Warning	2023-04-06T12:...	6912d7a0-bc28...
e57cb6fa-0e18-...	'audit' Policy ac...	Succeeded	Policy	Warning	2023-04-06T11:...	6912d7a0-bc28...
e57cb6fa-0e18-...	'audit' Policy ac...	Succeeded	Policy	Warning	2023-04-06T11:...	6912d7a0-bc28...
5c58edea-94d0-...	'audit' Policy ac...	Succeeded	Policy	Warning	2023-04-06T11:...	6912d7a0-bc28...
56fd2173-cf0e-...	Create or Upda...	Started	Administrative	Informational	2023-04-06T11:...	6912d7a0-bc28...
4085a982-8b9c-...	'auditIfNotExist...	Succeeded	Policy	Informational	2023-04-06T08:...	6912d7a0-bc28...
99fe9c3a-e36e-...	Update SQL da...	Succeeded	Administrative	Informational	2023-04-25T03:...	6912d7a0-bc28...
3eb6ed06-9208-...	Delete Public Ip...	Started	Administrative	Informational	2023-04-17T15:...	6912d7a0-bc28...
84fb2d66-c063-...	'auditIfNotExist...	Succeeded	Policy	Informational	2023-04-17T15:...	6912d7a0-bc28...
2127d1c8-ba1f-...	'audit' Policy ac...	Succeeded	Policy	Warning	2023-04-17T15:...	6912d7a0-bc28...
0dc66a10-870b-...	Delete An App	Failed	Administrative	Error	2023-04-17T14:...	6912d7a0-bc28...

00:00:14 Query executed successfully.