

Synapse Analytics Data Warehousing

~By

Alekhya Krishna Balivada

alekhya0830@gmail.com

Table of Contents

Sno.		Topic	Pg.No.
1		Project Statement	1
2		Project Overview	1
3		Project Requirements	1
4		Architecture Diagram	3
5		Execution Overview	3
6		Source Data Files	3
7		Project Implementation-Tasks performed	4
	7.1	Azure account	4
	7.2	Azure Synapse Analytics	4
	7.3	Running SQL script in Dedicated SQL pool	16
	7.4	Running SQL script in Dedicated SQL pool	17
		Conclusion	20

1. Project Statement

Synapse Analytics Data Warehousing:

Set up a simple data warehousing project using Azure Synapse Analytics and integrate with Azure SQL pools and run the queries to visualize the data in charts.

2. Project Overview

In this project, we will establish a streamlined data warehousing system utilizing Azure Synapse Analytics. The foundation will include seamless integration with Azure SQL pools, ensuring a robust and scalable data infrastructure. By leveraging the power of Synapse Analytics, we aim to enhance data processing capabilities and optimize storage for efficient querying. The project's core involves running targeted queries on integrated Azure SQL pools to extract valuable insights from the stored data. To visualize these insights comprehensively, we will employ charting tools, transforming raw data into meaningful graphical representations. This holistic approach not only underscores the efficiency of Azure Synapse Analytics in data management but also highlights the project's ultimate goal of delivering actionable insights through visually compelling charts.

3. Project Requirements

To initiate a basic data warehousing project using Azure Synapse Analytics and integrate it with Azure SQL pools, along with running queries to visualize data through charts, you'll need the following software components:

- a. Azure Subscription
- b. Azure storage accounts
- c. Azure Synapse Analytics
- d. Azure Synapse Studio
- e. SQL pools
- f. Visualization tools
- g. SQL query tools
- h. Documentation tools

3.1 Azure Subscription:

Ensure you have access to an Azure subscription to set up and manage the necessary services.

3.2 Azure Storage Accounts:

Utilize tools like Azure Storage Accounts to facilitate the smooth integration of data from source into Azure Synapse Analytics.

3.3. Azure Synapse Analytics:

Establish an instance of Azure Synapse Analytics to act as the central hub for your data warehouse.

3.4 Azure Synapse Studio:

Launch Azure Synapse Studio, a unified analytics and data integration environment, to efficiently manage your data warehousing project

3.5 Azure SQL Pools:

Configure and seamlessly integrate Azure SQL pools within Azure Synapse Analytics to ensure efficient data storage and retrieval.

3.6 Visualization Tools:

Select visualization tools compatible with Azure, such as Power BI or Azure Synapse Studio's built-in capabilities, to craft insightful charts and dashboards.

3.7 SQL Query Tools:

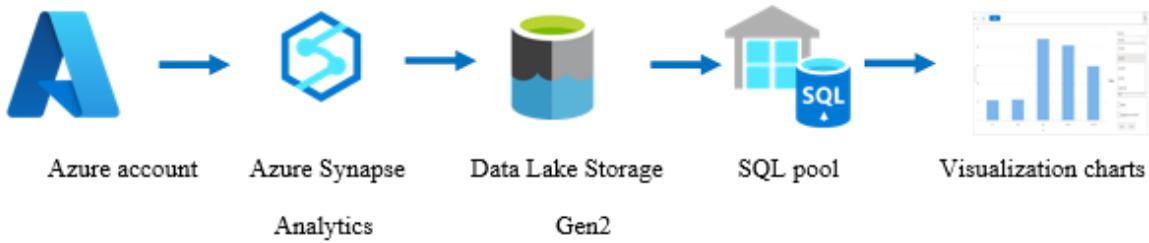
Employ SQL query tools like Azure Synapse Studio or other compatible options to execute queries on the integrated Azure SQL pools.

3.8 Documentation Tools:

Utilize documentation tools like Microsoft Word, Markdown, or other collaborative platforms to comprehensively document the project setup, configurations, and procedures.

- Ensure the compatibility and proper configuration of all software components according to your project's specifications. Regularly update and patch the software to take advantage of the latest features and security updates.

4. Architecture diagram:



5. Execution Overview:

- Here we did using Azure Synapse Analytics and integrate with azure sql pools and run the queries to visualize the data in charts.
- First we created a synapse analytics workspace and also created datalake gen2 storage account with one container while creating the Azure Synapse workspace.
- After creation of workspace we launched the workspace and by clicking on data navigator we can able to see the linked data storage accounts and containers which we created at the time of workspace creation.
- In that container we uploaded two csv files named project.csv and project1.csv.
- After uploading files we opened Sql pools inside the manage navigator there we can find builtin Sql pool to perform serverless integration.
- And also we created Dedicated Sql pool to perform dedicated integration.
- After creation of Sql pools we created Sql script by clicking on manage navigator to run the sql queries and visualize the data in charts.

6. Source Date Files:

Here we took Source Data files of type ‘CSV’

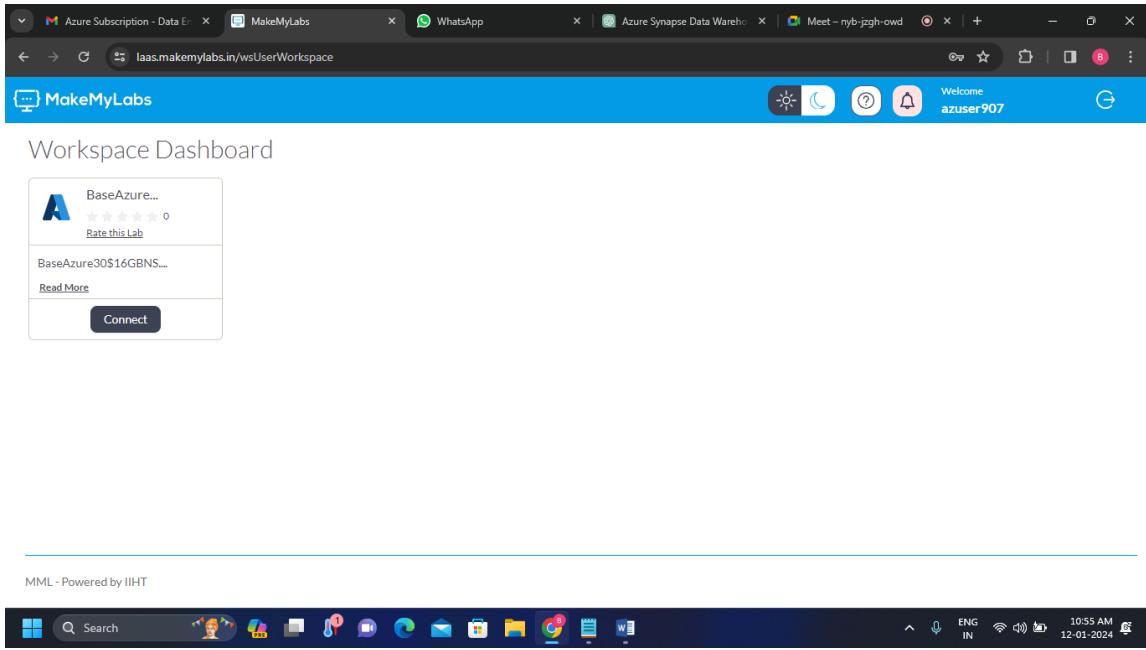
File1: Project.csv

File2: Project1.csv

7. Project Implementation - Tasks performed

7.1 Azure account

- Login to the Azure account



7.2 Azure Synapse Analytics

Step-1:

- Open the Azure Portal and search for Azure Synapse Analytics in search bar.

Step-2:

- Create a new Azure Synapse Analytics workspace.

- Fill the necessary data. Here I gave the synapse workspace name as “teamjsynapse”.

Create Synapse workspace

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 * 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 *

Review + create < Previous Next: Security >

- Select the Data Lake Storage Gen2 account if already exists.
- If not exists create new one.
- Here we created new Data Lake Storage Gen2 account as “teamjstorage”.

Create Synapse workspace

Select Data Lake Storage Gen2 * From subscription Manually via URL

Account name *

Data Lake Storage Gen2 account

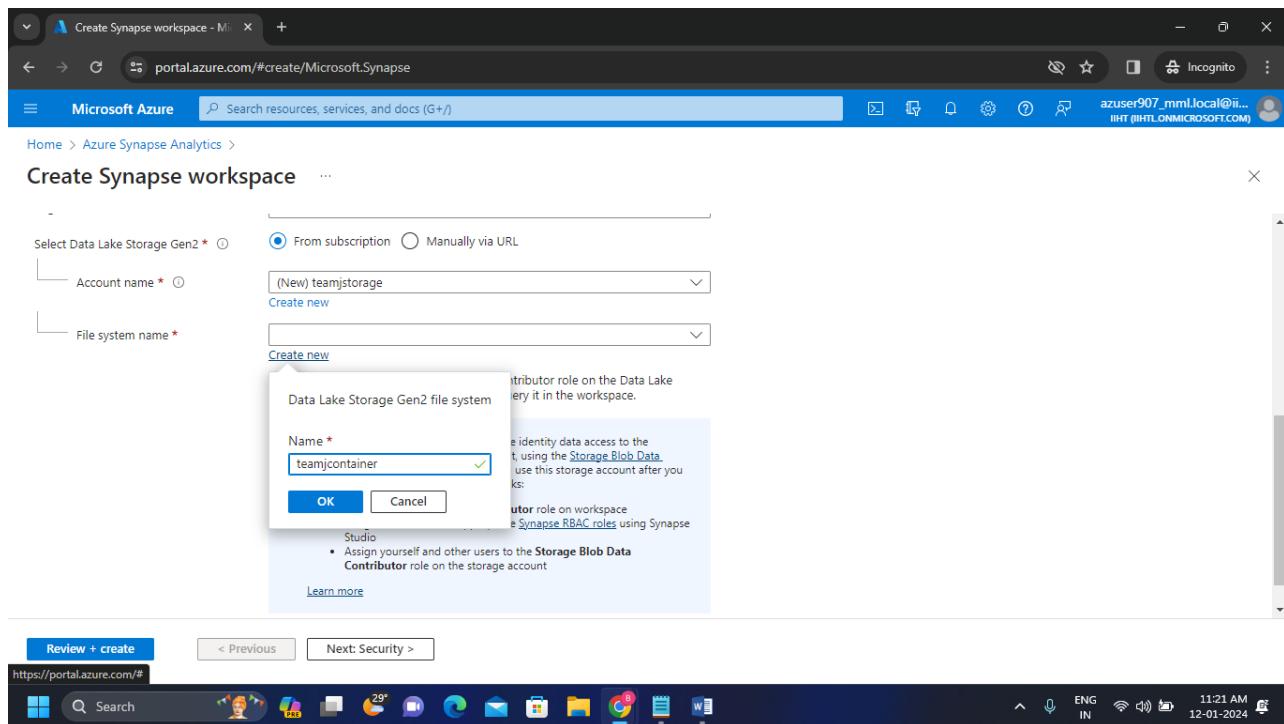
Name * OK Cancel

Grant workspace identity data access to the storage account, using the [Storage Blob Data Contributor](#) role on workspace and [Storage Blob Data Contributor](#) role on the storage account.

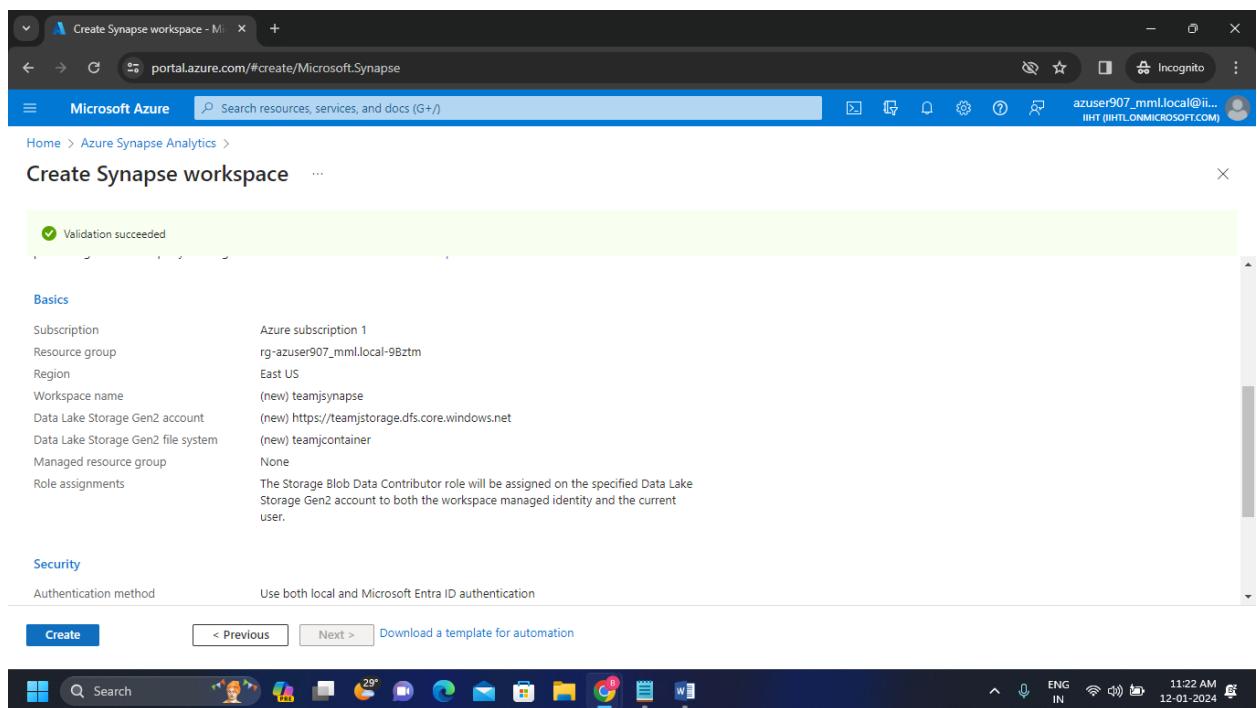
Learn more

Review + create < Previous Next: Security >

- We named the file system (container) as “teamjcontainer”.



- Review and click on create to create the Azure Synapse workspace.



- Here we can see that the Azure synapse is successfully created.

The screenshot shows the Microsoft Azure portal with the URL <https://portal.azure.com/#view/HubsExtension/DeploymentDetailsBlade/~/overview/id/%2Fsubscriptions%2F984f097c-963c-4eb6-a20d-839457ae9f08%2Fres...>. The page title is "Microsoft.Azure.SynapseAnalytics-20240112112023 | Overview". A prominent green checkmark icon indicates "Deployment succeeded". The deployment details show the name "Microsoft.Azure.SynapseAnalytics-2023", start time "1/12/2024, 11:22:35 AM", subscription "Azure subscription 1", correlation ID "bbbe8573-d393-4466-aaa5-aa3fe3dcf...", and resource group "rg-azuser907_mml.local-9Bztm". Below the deployment summary, there are sections for "Deployment details" and "Next steps", with a "Go to resource group" button. To the right, there are promotional cards for "Cost management", "Microsoft Defender for Cloud", "Free Microsoft tutorials", and "Work with an expert". The status bar at the bottom shows the date "12-01-2024" and time "11:26 AM".

Step-3:

- After deployment of workspace we go to the Synapse resource. Then we clicked on the synapse workspace.

The screenshot shows the Microsoft Azure portal with the URL https://portal.azure.com/#@ihtlonmicrosoft.com/resource/subscriptions/984f097c-963c-4eb6-a20d-839457ae9f08/resourceGroups/rg-azuser907_mml.local-9.... The page title is "rg-azuser907_mml.local-9Bztm | Resource group". The left sidebar shows navigation options like "Overview", "Activity log", "Access control (IAM)", "Tags", "Resource visualizer", "Events", "Settings", "Deployments", "Security", "Deployment stacks", "Policies", "Properties", and "Locks". The main content area displays a table of resources under the "Essentials" section. The table has columns for "Name", "Type", and "Location". Two resources are listed: "teamjstorage" (Storage account, Type: blob, Location: East US) and "teamjsynapse" (Synapse workspace, Type: Synapse workspace, Location: East US). The status bar at the bottom shows the date "12-01-2024" and time "11:27 AM".

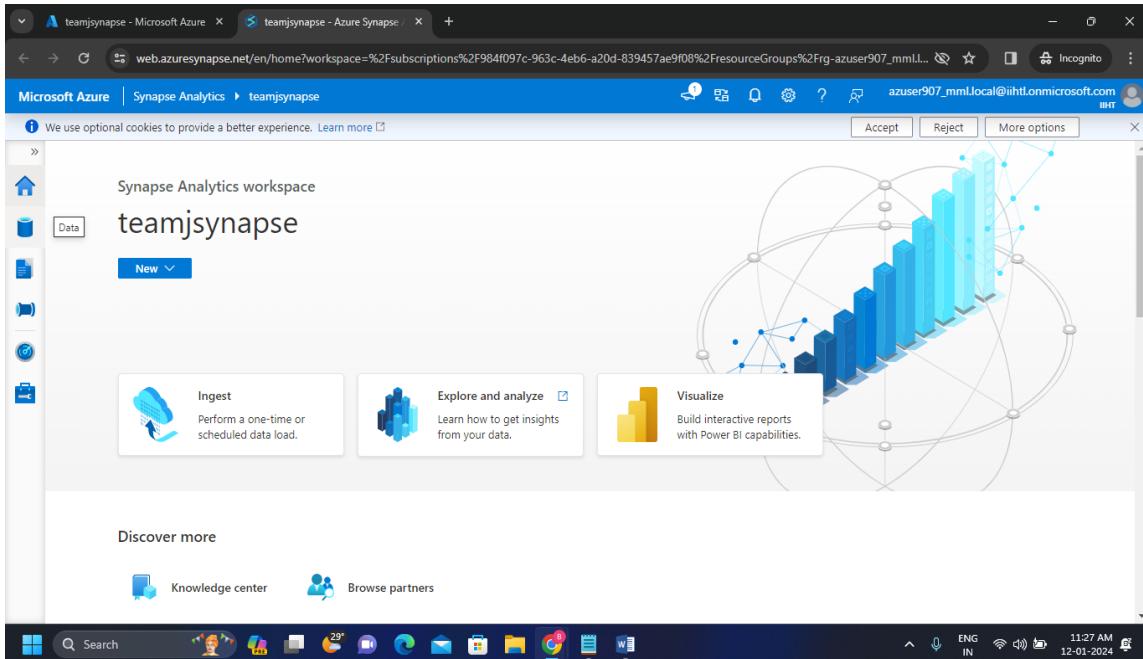
Step-4: Now open the Synapse studio.

The screenshot shows the Microsoft Azure portal interface for a Synapse workspace named 'teamjsynapse'. The top navigation bar includes the workspace name, a search bar, and various icons for account management. The main content area displays workspace settings such as Managed virtual network (No), Managed Identity object ID (9a67f823-0ac2-4219-a24e-2f3072c146c8), Workspace web URL (https://web.azuresynthesize.net?workspace=%2bsubscriptions%2f984f097c-963c-4eb6-a20d-839457ae9f08%2frg-azuser907_mml.local-9Bztm), and Dedicated SQL endpoint (teamjsynapse.on-demand.sql.azuresynapse.net). A 'Getting started' section offers links to 'Open Synapse Studio' and 'Read documentation'. On the left, a sidebar provides navigation links for Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Settings (Microsoft Entra ID, Properties, Locks), and Analytics pools (SQL pools, Apache Spark pools, Data Explorer pools (preview)). The bottom status bar shows the URL https://web.azuresynthesize.net/?workspace=%2bsubscriptions%2f984f097c-963c-4eb6-a20d-839457ae9f08%2frg-azuser907_mml.local-9Bztm..., the date (12-01-2024), and time (11:27 AM).

The screenshot shows the Azure Synapse Analytics workspace home page. The top navigation bar includes the workspace name, a search bar, and various icons for account management. The main content area features a large blue banner with the text 'Azure Synapse Analytics' and 'Loading'. On the left, a sidebar provides navigation links for Microsoft Azure, Synapse Analytics, and teamjsynapse. The bottom status bar shows the URL https://web.azuresynthesize.net/en/?workspace=%2bsubscriptions%2f984f097c-963c-4eb6-a20d-839457ae9f08%2frg-azuser907_mml.local-9Bztm..., the date (12-01-2024), and time (11:27 AM).

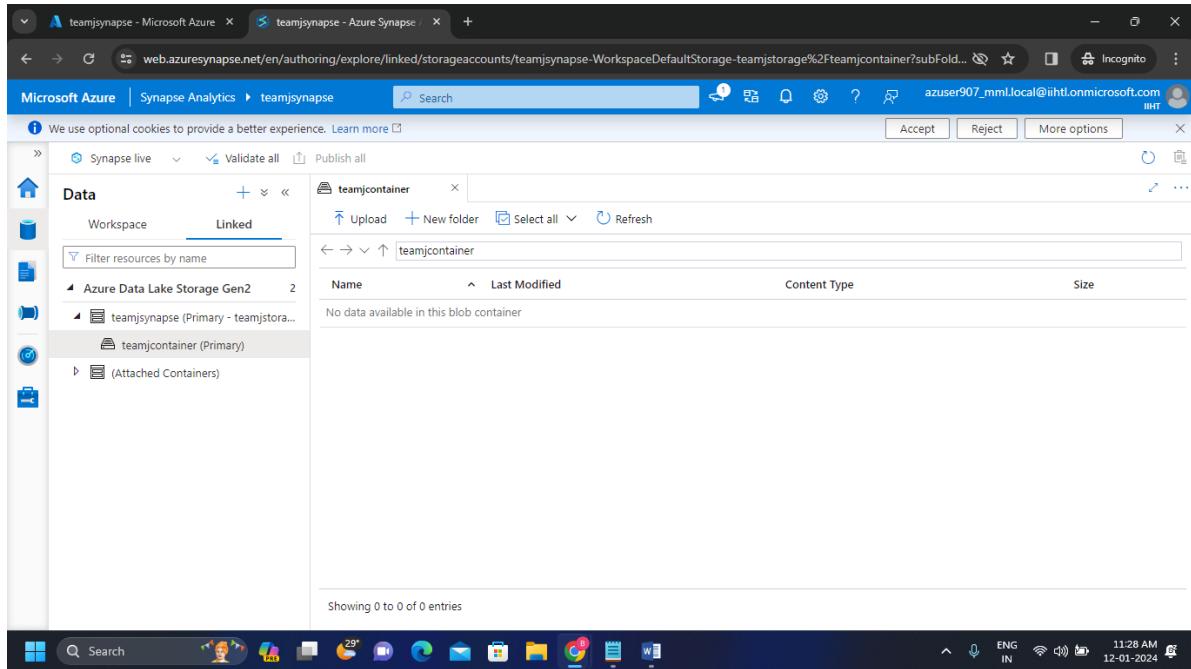
Step-5:

- Below we can see the Synapse workspace.
- By clicking Data navigator we can able to see the linked data storage accounts in the container.



The screenshot shows the Microsoft Azure Synapse Analytics workspace titled "teamjsynapse". On the left, there's a sidebar with icons for Home, Data, and New. The "Data" icon is selected. Below it, there are three cards: "Ingest" (Perform a one-time or scheduled data load), "Explore and analyze" (Learn how to get insights from your data), and "Visualize" (Build interactive reports with Power BI capabilities). To the right of these cards is a large, 3D bar chart visualization. At the bottom of the workspace, there are links for "Discover more", "Knowledge center", and "Browse partners".

- Here we can see the teamjsynapse Data lake storage Gen2 account that we created and also the teamjcontainer.



The screenshot shows the Microsoft Azure Data storage blade for the "teamjsynapse" workspace. The "Data" section has two tabs: "Workspace" and "Linked". The "Linked" tab is selected. Under "Linked", there is a list of storage accounts: "teamjstorage" (Primary - teamjstorage) and "teamjcontainer" (Primary). The "teamjcontainer" entry is expanded, showing its primary container and attached containers. The blade includes a search bar, filter by name, upload and new folder buttons, and a refresh button. The status bar at the bottom indicates "Showing 0 to 0 of 0 entries".

Step-6:

- After opening the container we uploaded our “project.csv” file in the container.

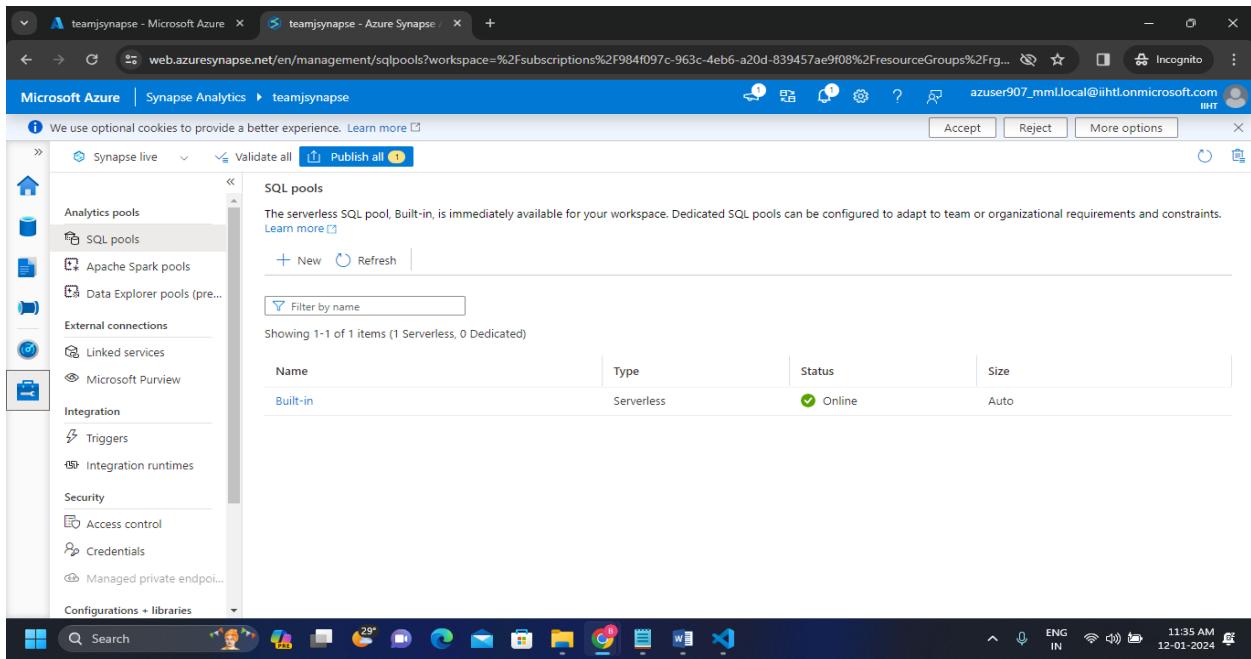
The screenshot shows the Microsoft Azure Synapse Analytics Data Explorer interface. On the left, the navigation pane shows 'Data' selected, with 'Workspace' and 'Linked' options. Under 'Linked', there is a section for 'Azure Data Lake Storage Gen2' which includes 'teamjsynapse (Primary - teamjstorage)' and 'teamjcontainer (Primary)'. The main workspace shows a 'teamjcontainer' folder. On the right, there is an 'Upload Files' panel for 'teamjcontainer'. It shows a 'File Upload' field containing 'project.csv', with an unchecked 'Overwrite existing files' checkbox. Below it is a table with one row for 'project.csv', showing its size as 1.33 KB and an 'Action' column with a 'Remove' link. At the bottom of the panel are 'Upload' and 'Cancel' buttons. The status bar at the bottom of the browser window shows the date and time as 12-01-2024 11:29 AM.

- We can see the uploaded file inside the container.

The screenshot shows the Microsoft Azure Synapse Analytics Data Explorer interface. The left navigation pane shows 'Data' selected, with 'Workspace' and 'Linked' options. Under 'Linked', there is a section for 'Azure Data Lake Storage Gen2' which includes 'teamjsynapse (Primary - teamjstorage)' and 'teamjcontainer (Primary)'. The main workspace shows a 'teamjcontainer' folder. On the right, the data grid displays a single item: 'project.csv' with a 'Content Type' of 'Text' and a 'Size' of 1.3 KB. The status bar at the bottom of the browser window shows the date and time as 12-01-2024 11:30 AM.

Step-7:

- After uploading csv file into the container open the manage navigator to see SQL pools.
- Here we can see one sql pool which is built in sql pool used to perform serverless integrations.



The screenshot shows the Microsoft Azure Synapse Analytics management interface. The left sidebar has a tree view with 'Analytics pools' expanded, showing 'SQL pools' selected. The main content area is titled 'SQL pools' with a sub-instruction: '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.' Below this are buttons for '+ New' and 'Refresh'. A 'Filter by name' input field is present. A table lists one item: 'Built-in' (Type: Serverless, Status: Online, Size: Auto). The top navigation bar shows the URL 'web.azuresynthesize.net/en/management/sqlpools?workspace=%2Fsubscriptions%2F984f097c-963c-4eb6-a20d-839457ae9f08%2FresourceGroups%2Frg...' and the user 'azuser907_mml.local@ihtl.onmicrosoft.com'. The bottom taskbar includes icons for various Microsoft services like Power BI, Data Factory, and Stream Analytics.

Step-8:

- Here we created Dedicated Sql pool to perform dedicated integrations.

The screenshot shows the Microsoft Azure portal interface for creating a new dedicated SQL pool. The URL in the address bar is <https://web.azure-synapse.net/en/management/sql/pools?workspace=%2Fsubscriptions%2F984f097c-963c-4eb6-a20d-839457ae9f08%2FresourceGroups%2Frg...>. The user is signed in as azuser907_mml.local@iht.onmicrosoft.com.

New dedicated SQL pool

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. [Learn more](#)

Dedicated SQL pool details

Name your dedicated SQL pool and choose its initial settings.

Dedicated SQL pool name * teamsqlpool

Performance level DW100c

Estimated price Est. cost per hour 1.51 USD [View pricing details](#)

Review + create **Next: Additional settings >** **Cancel**

The screenshot shows the Microsoft Azure portal interface for creating a new dedicated SQL pool, now on the 'Review + create' tab. The URL in the address bar is <https://web.azure-synapse.net/en/management/sql/pools?workspace=%2Fsubscriptions%2F984f097c-963c-4eb6-a20d-839457ae9f08%2FresourceGroups%2Frg...>. The user is signed in as azuser907_mml.local@iht.onmicrosoft.com.

New dedicated SQL pool

Validation succeeded.

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

Product details

Azure Synapse Analytics dedicated SQL pool by Microsoft
Est. cost per hour 1.51 USD [View pricing details](#)
[Terms of use](#) | [Privacy policy](#)

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](#).

Data source

Dedicated SQL pool name: teamsqlpool
Performance level: DW100c

Create **< Previous** **Download template for automation** **Cancel**

- We can see the Dedicated Sql pool created here.

The screenshot shows the Microsoft Azure Synapse Analytics interface. The left sidebar navigation includes options like Analytics pools, SQL pools, Apache Spark pools, Data Explorer pools, External connections, Linked services, Microsoft Purview, Integration, Triggers, Integration runtimes, Security, Access control, Credentials, and Managed private endpoints. The main content area is titled "SQL pools" and contains a table with the following data:

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

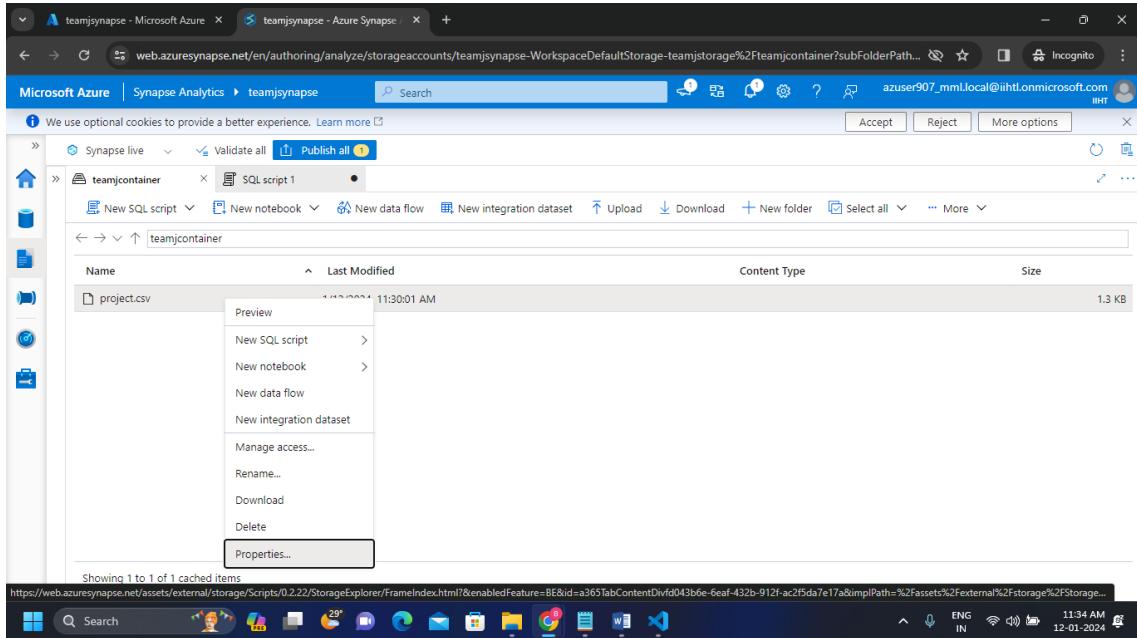
Step-9:

- Now we created Sql script by clicking develop navigator.
- Created Sql script to run the queries and for visualizing the data in charts.

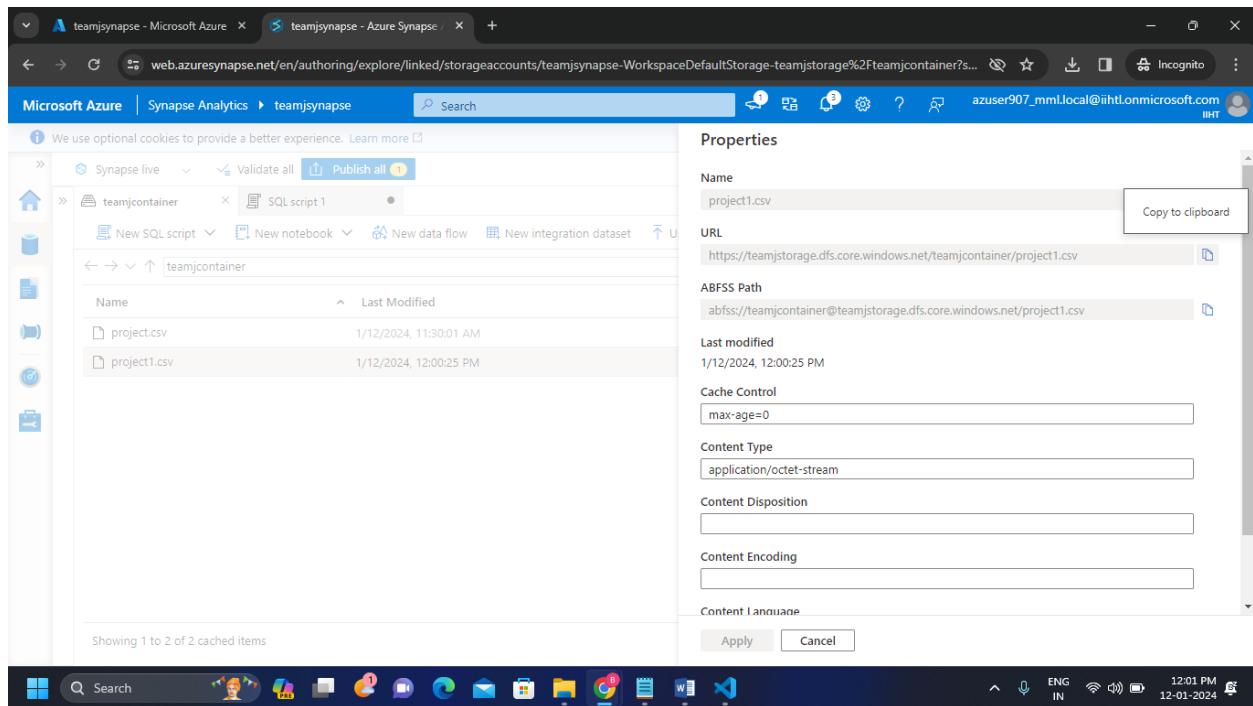
The screenshot shows the Microsoft Azure Synapse Analytics interface under the "Develop" section. The left sidebar navigation includes options like Synapse live, Validate all, Publish all, and a "teamjcontainer" item. A dropdown menu is open, listing resources: SQL script, KQL script, Notebook, Data flow, Apache Spark job definition, Browse gallery, and Import. The "teamjcontainer" item is highlighted. Below the dropdown, a message says "No items to show" and "Try creating a new item using the + button above." The status bar at the bottom shows "11:30 AM 12-01-2024".

Step-10:

- Now open the storage account and by right clicking on that data file we can see the properties dropdown.



- From here we copied the url of the file.



7.3 Running SQL script in Dedicated SQL pool

- Now we run the commands to create a table by accessing the file which we already uploaded in storage account.
- Here we accessed that file by pasting that url of the file which we copied in step-9.
- And here we performed by using Dedicated sql pool which is created in step-7.
- We connected Sql script to Dedicated sql pool then automatically it will use the database.

The screenshot shows the Microsoft Azure Synapse Analytics studio interface. A SQL script is being run against a Dedicated SQL pool named 'teamjsqlpool'. The script creates a table 'dbo.Project' and copies data from a CSV file in blob storage into it. The execution message at the bottom indicates '00:00:04 Query executed successfully.'

```
1 CREATE TABLE dbo.Project
2 (
3     empid int, name varchar(30), sex varchar(30), salary BIGINT,deptid VARCHAR(30)
4 )
5 WITH
6 (
7     DISTRIBUTION = ROUND_ROBIN,
8     CLUSTERED COLUMNSTORE INDEX
9 )
10
11 GO
12
13 COPY INTO dbo.Project
14 FROM 'https://teamjstorage.dfs.core.windows.net/teamjcontainer/project.csv'
15 WITH(
16     FILE_TYPE = 'CSV',
17     FIRSTROW=2
18 )
19
20 SELECT * from Project
```

Step-1:

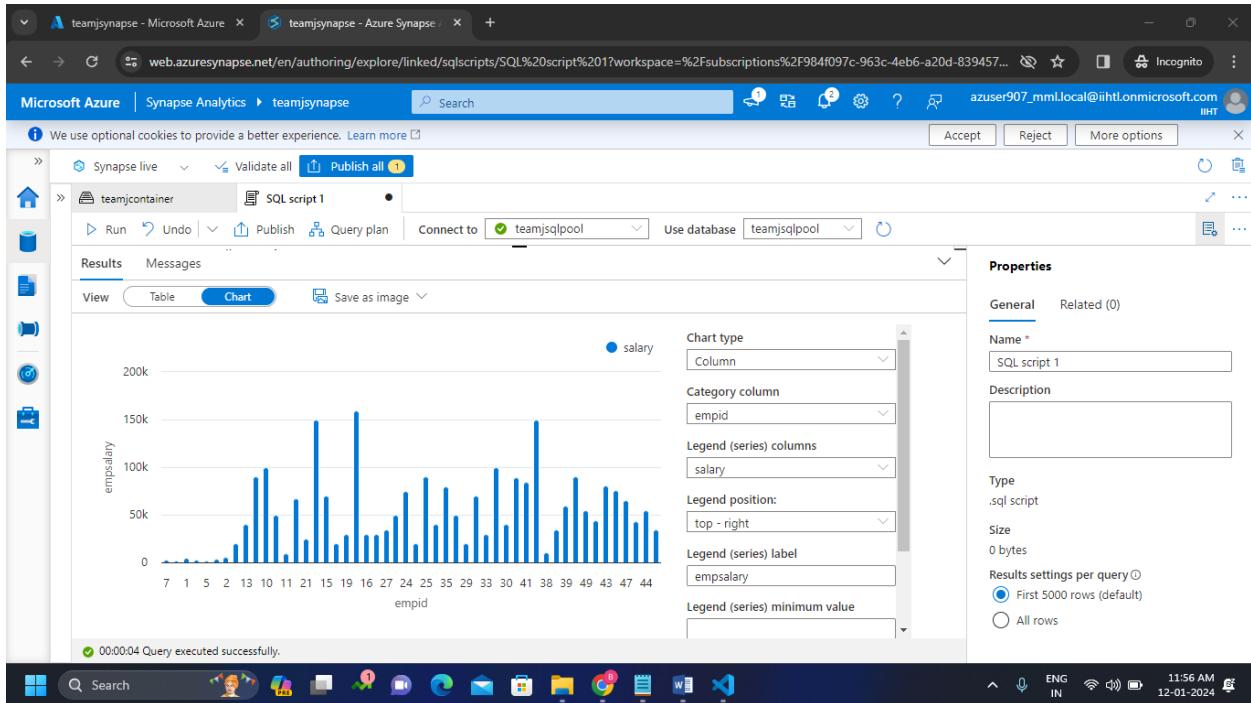
- Run the SQL Script and we can see the execution output as below.

The screenshot shows the execution results of the SQL script. The results tab displays a table with data from the 'Project' table. The columns are 'empid', 'name', 'sex', 'salary', and 'deptid'. The data includes rows for employees like ayesha, Mahaboob, maheer, asi, pyarijaan, sarfaraj, wafa, and sowji, with their respective details.

empid	name	sex	salary	deptid
7	ayesha	F	3000	IT
6	Mahaboob	M	2000	Payroll
1	maheer	M	5000	IT
3	asi	F	2500	Payroll
5	pyarijaan	F	2000	HR
4	sarfaraj	M	4000	HR
2	wafa	M	6000	IT
14	sowji	F	20000	Payroll

Step-2:

- In this step we did visualization for the data which we got output.
- We represented the data in the form of column chart by visualizing the data.



7.4 Running SQL script in Dedicated SQL pool

- And also we performed with built-in sql pool.
- We performed some queries in sql script by giving the path of the file.
- Here we used OPENROWSET function which works for serverless sql pools.
- And we connected to Built – in pool then automatically it will use master database.
- Then Click on Run button.

The screenshot shows the Microsoft Azure Synapse Analytics workspace interface. A BULK INSERT query is being run:

```

1 SELECT * from OPENROWSET(
2   BULK 'https://teamjstorage.dfs.core.windows.net/teamjcontainer/project1.csv',
3   FORMAT = 'CSV',
4   HEADER_ROW = TRUE,
5   PARSE_VERSION = '2.0'
6 ) AS [RESULT]

```

The results pane shows a message: "00:00:16 Query executed successfully." To the right, the "Properties" panel is open for "SQL script 1".

General	Related (0)
Name *	SQL script 1
Description	
Type	.sql script
Size	0 bytes
Results settings per query	<input checked="" type="radio"/> First 5000 rows (default) <input type="radio"/> All rows

- Here We can see the output of the above code which is in Tabular Format.

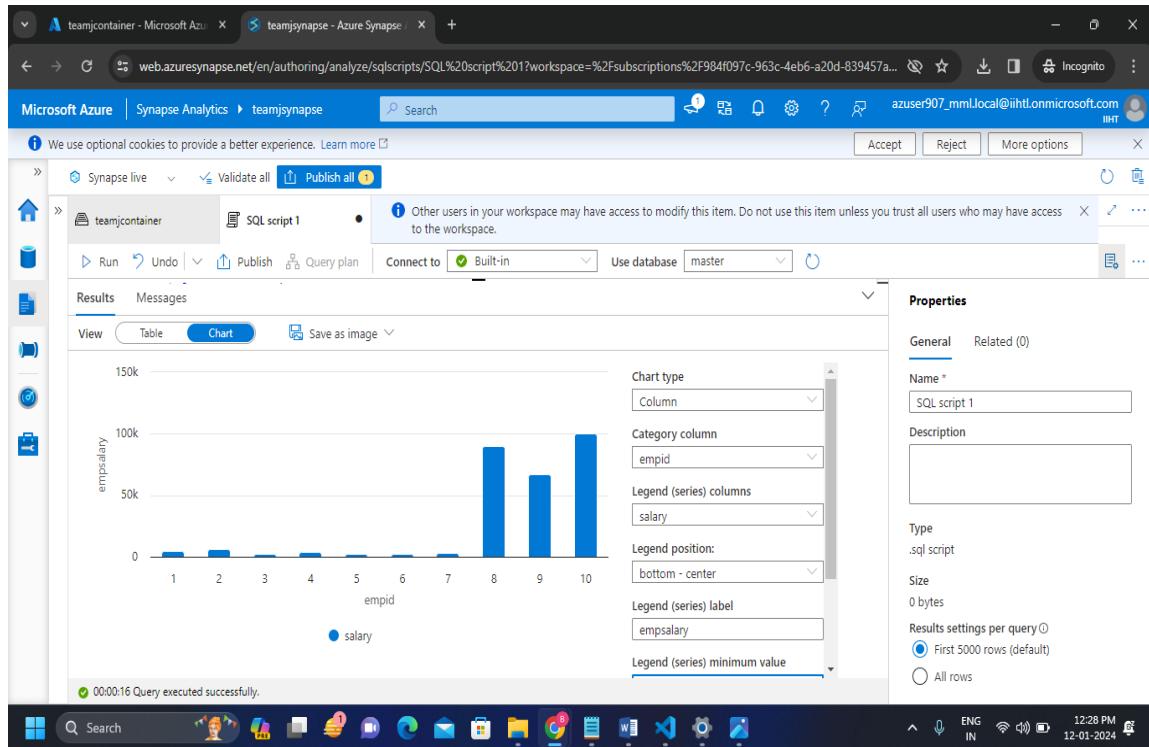
The screenshot shows the Microsoft Azure Synapse Analytics workspace interface displaying the results of the BULK INSERT query. The results are presented in a table format:

empid	name	sex	salary	dept id
1	maheer	M	5000	IT
2	wafa	M	6000	IT
3	asi	F	2500	Payroll
4	sarfaraj	M	4000	HR
5	pyariaan	F	2000	HR
6	Mahaboob	M	2000	Payroll
7	ayesha	F	3000	IT

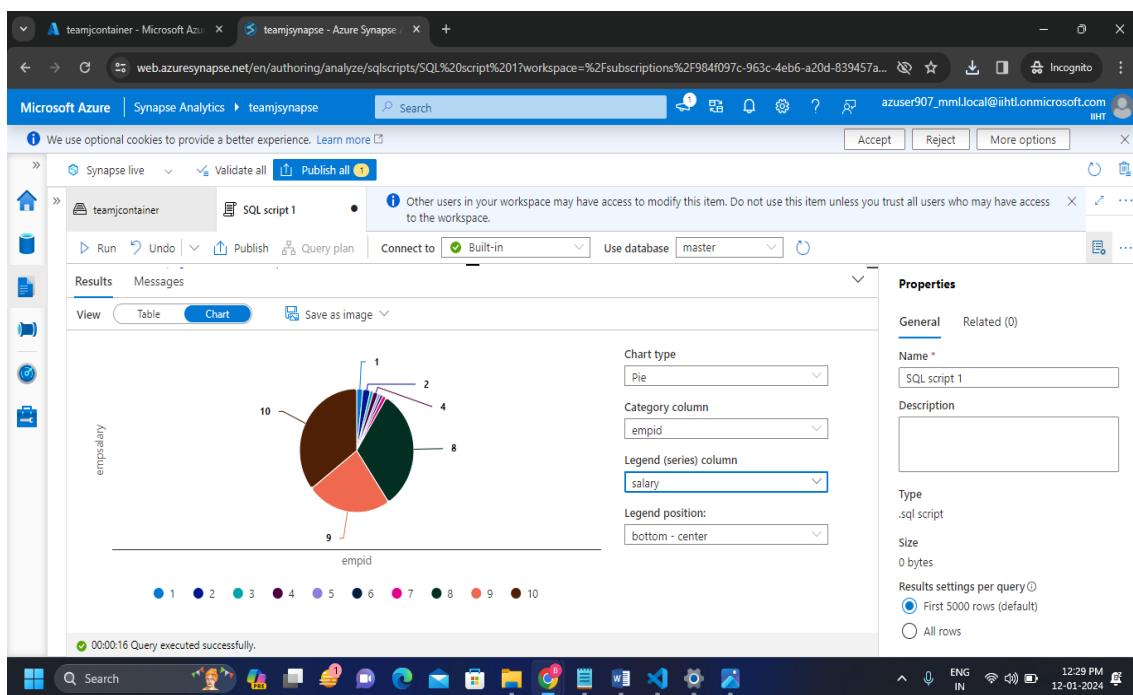
The results pane shows a message: "00:00:16 Query executed successfully." To the right, the "Properties" panel is open for "SQL script 1".

General	Related (0)
Name *	SQL script 1
Description	
Type	.sql script
Size	0 bytes
Results settings per query	<input checked="" type="radio"/> First 5000 rows (default) <input type="radio"/> All rows

- Visualizing the data in bar graph format.



➤ Visualizing the data using Pie-chart.



Conclusion: Hence I successfully Set up a simple data warehousing project using Azure Synapse Analytics and integrate with azure sql pools and run the queries to visualize the data in charts.