**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.

1. **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.

1. **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:

1. Azure Subscription
2. Azure storage accounts
3. Azure Synapse Analytics
4. Azure Synapse Studio
5. SQL pools
6. Visualization tools
7. SQL query tools
8. Documentation tools

3.1 Azure Subscription:

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

* 1. Azure Storage Accounts:

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

* 1. Azure Synapse Analytics:

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

* 1. Azure Synapse Studio:

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

* 1. Azure SQL Pools:

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

* 1. 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.

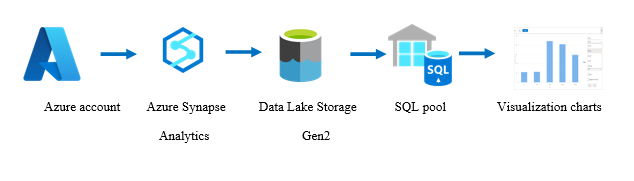
* 1. SQL Query Tools:

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

* 1. 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.

1. **Architecture diagram:**
2. **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.

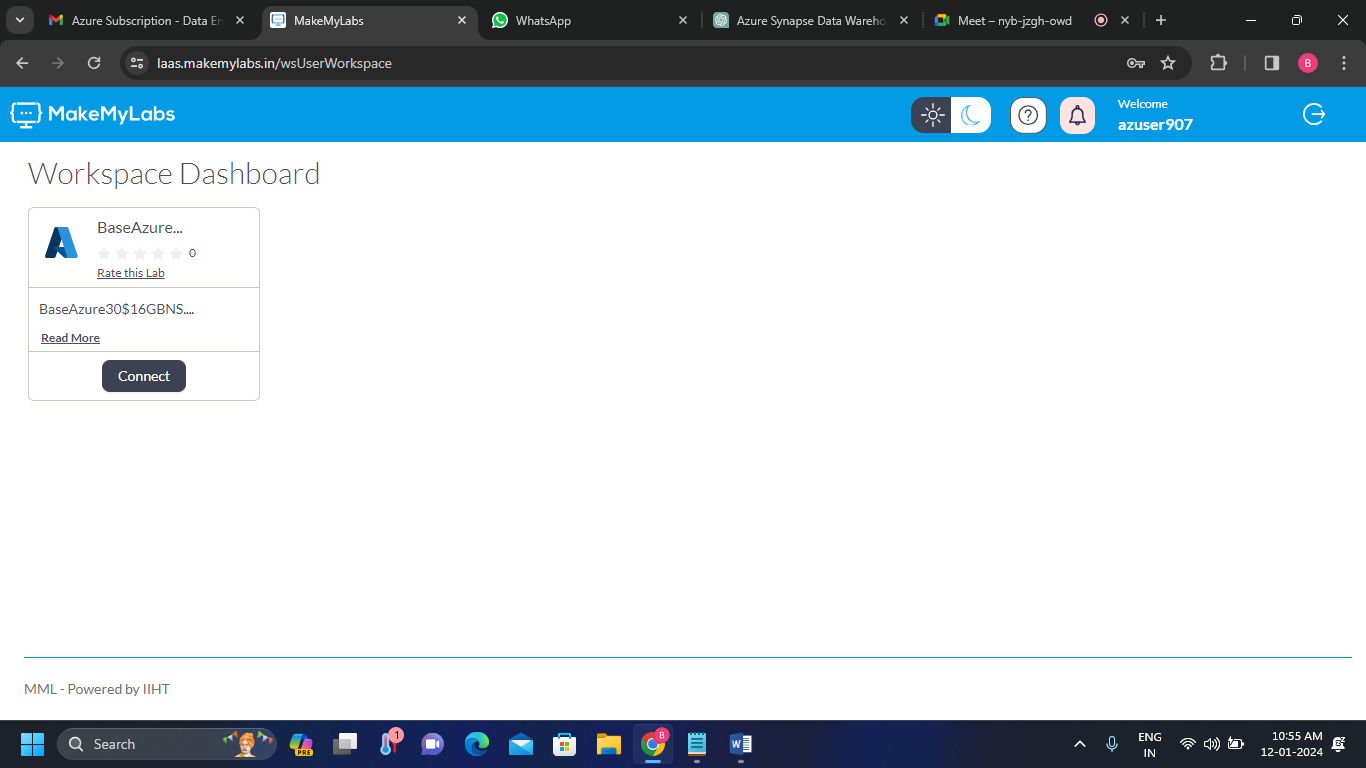
1. **Source Date Files:**

Here we took Source Data files of type ‘CSV’

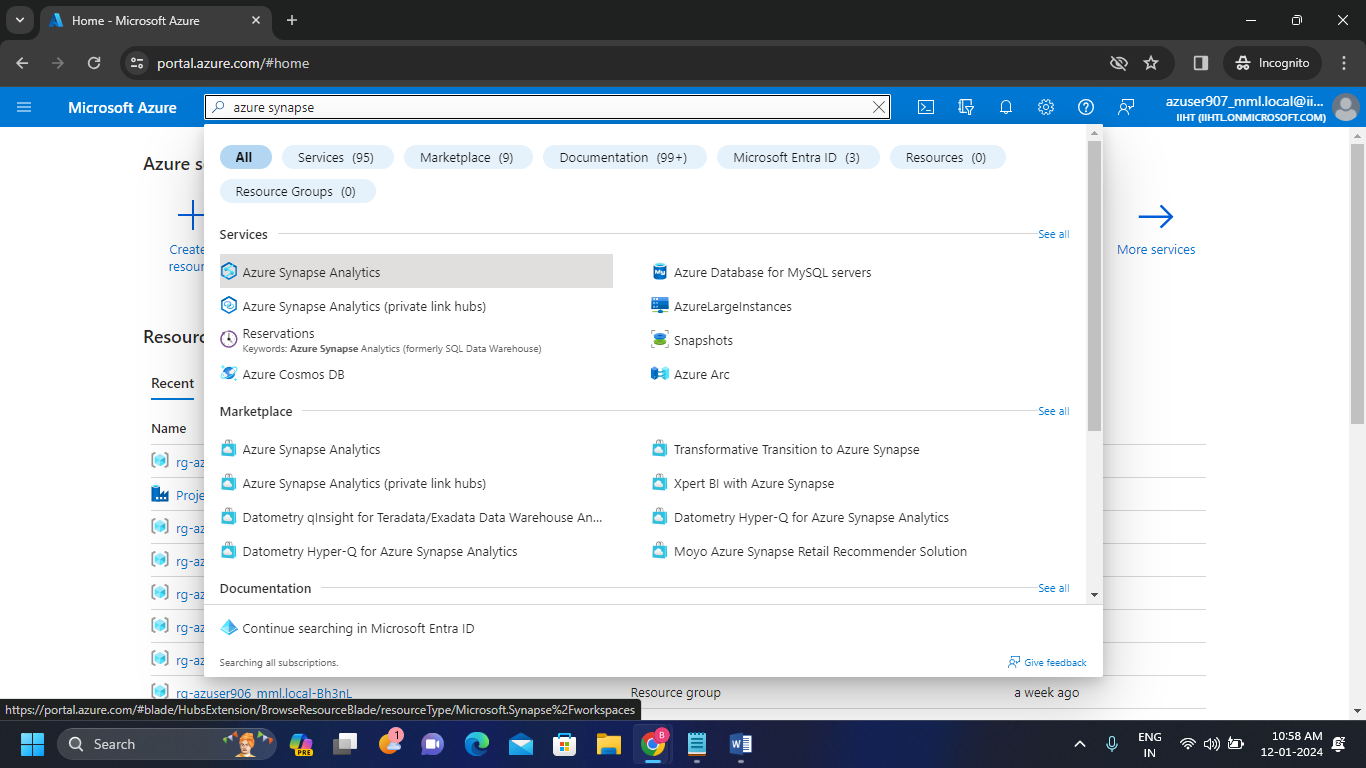
File1: Project.csv

File2: Project1.csv

1. **Project Implementation - Tasks performed**
   1. **Azure account**

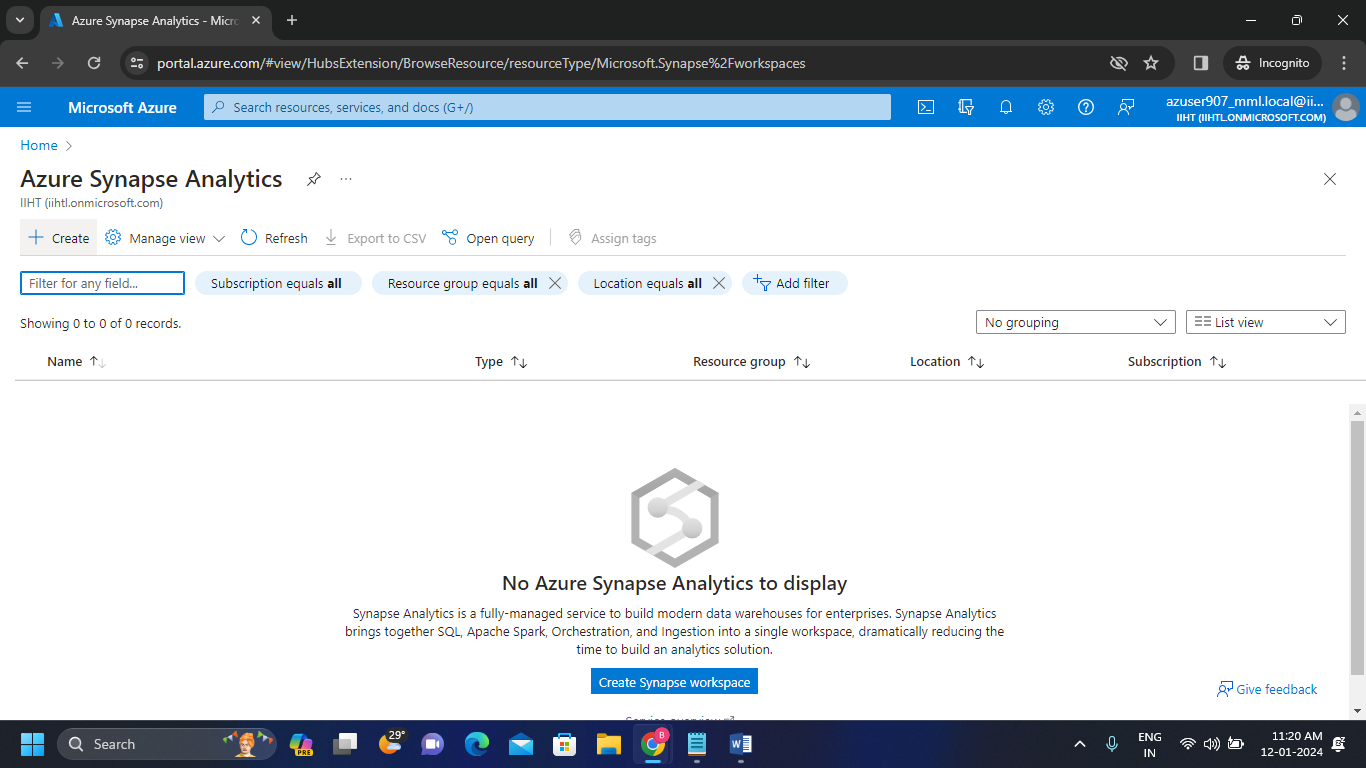
* ****Login to the Azure account
  1. **Azure Synapse Analytics**

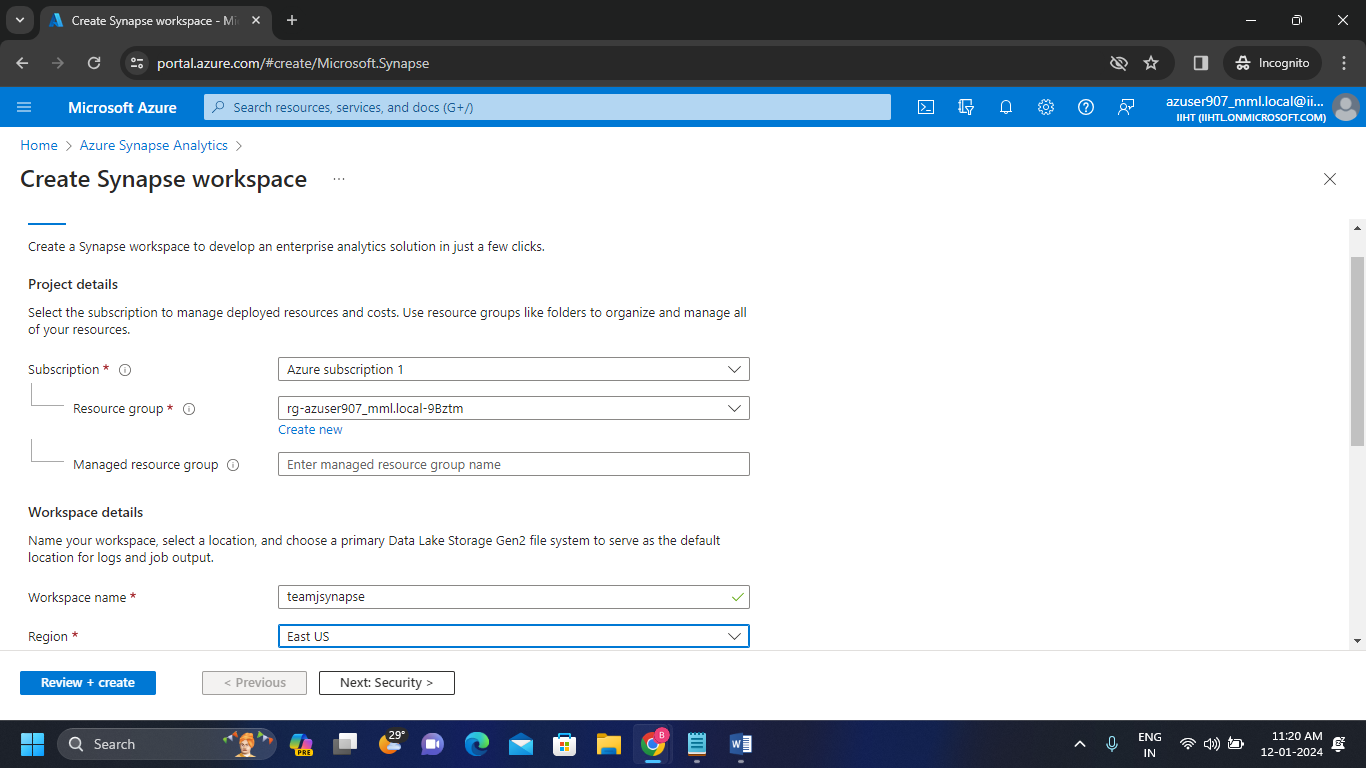
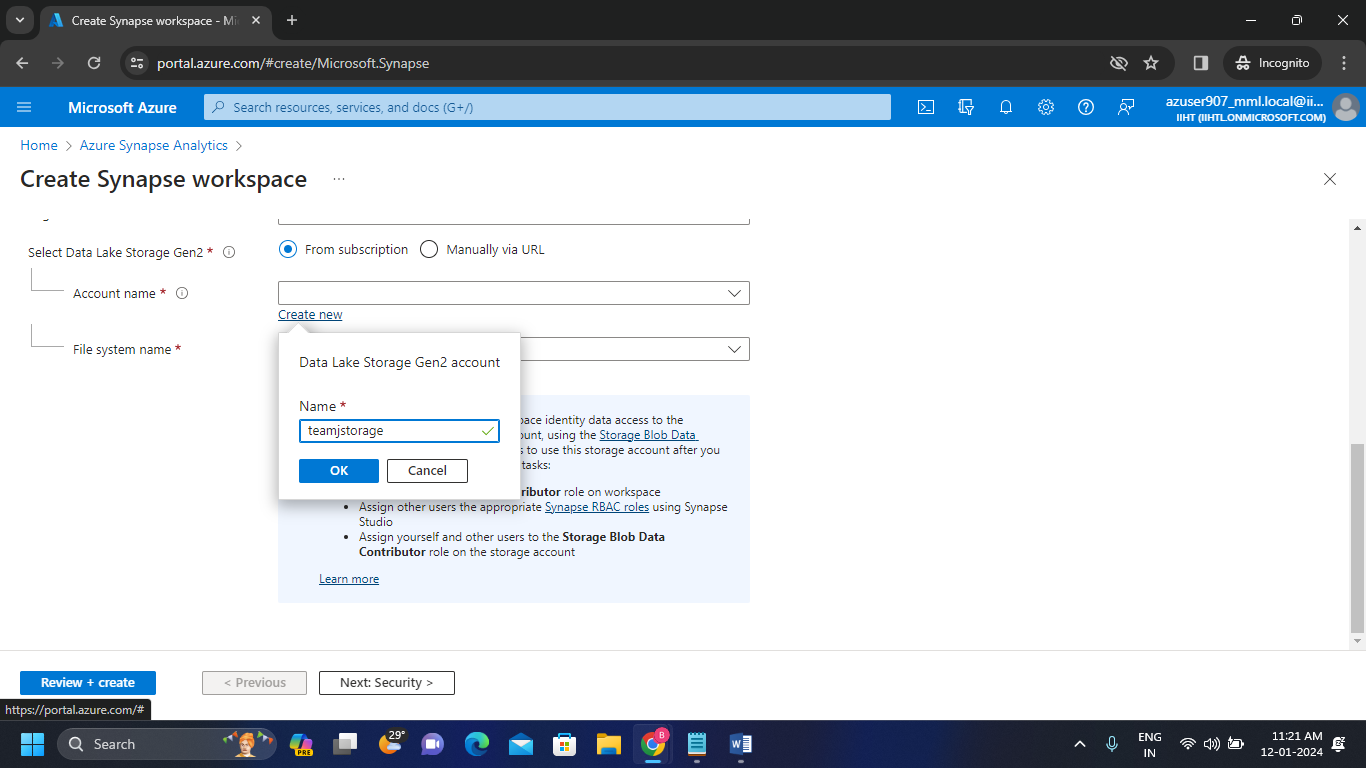
**Step-1:**

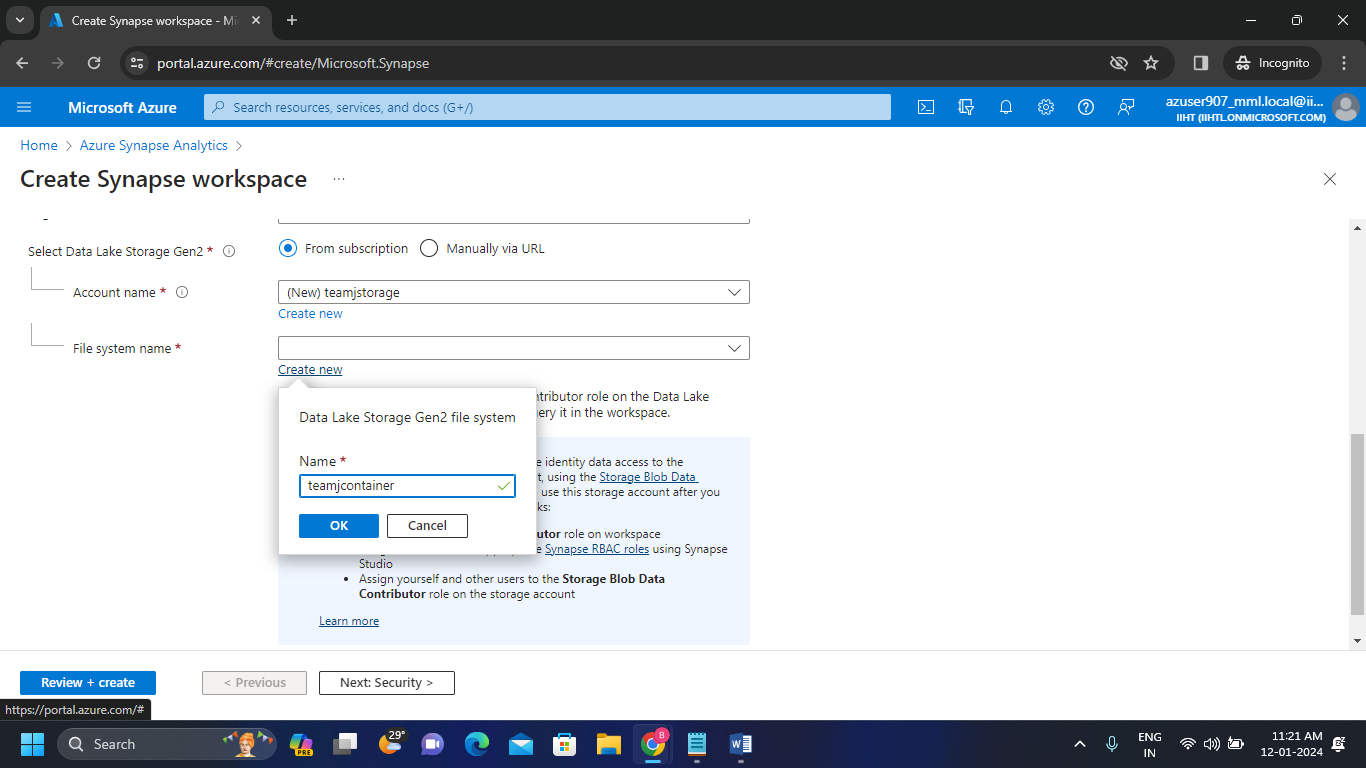
* Open the Azure Portal and search for Azure Synapse Analytics in search bar.

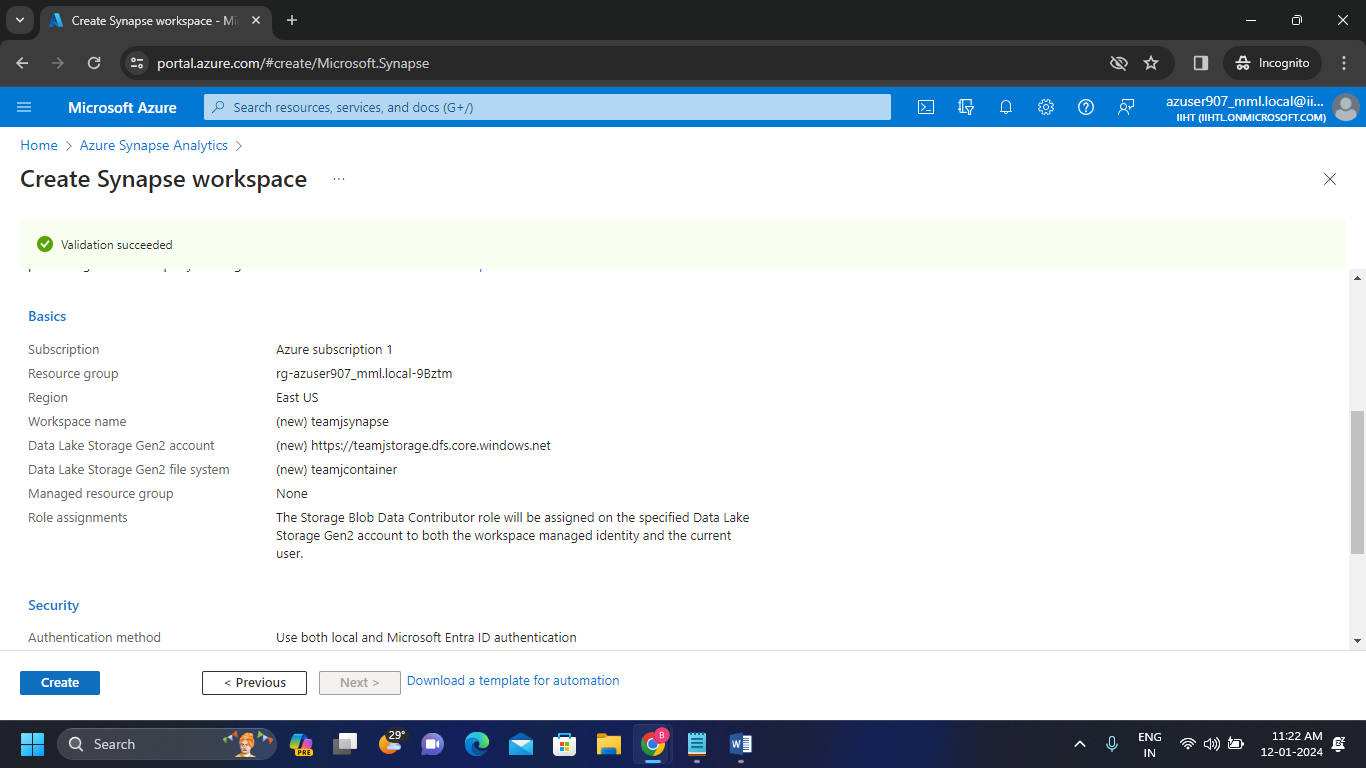
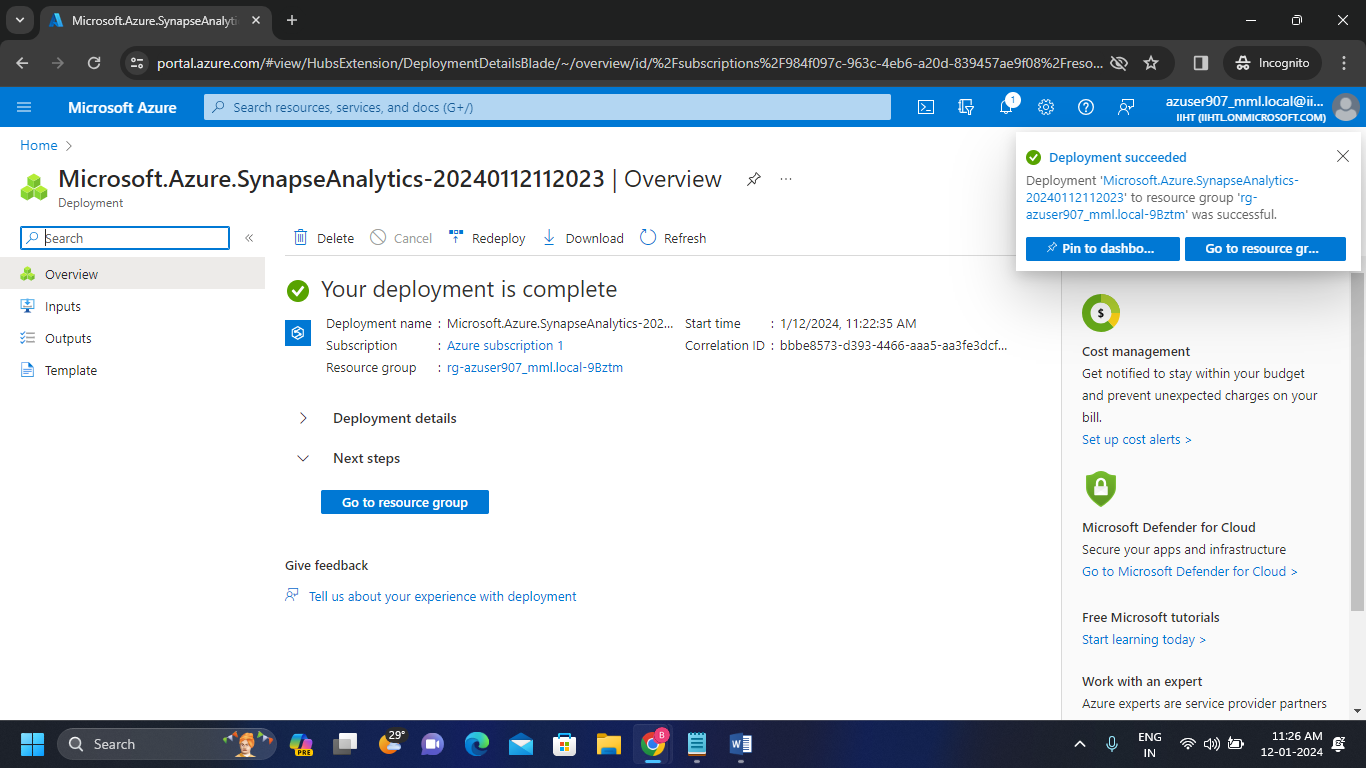
**Step-2:**

* Create a new Azure Synapse Analytics workspace.



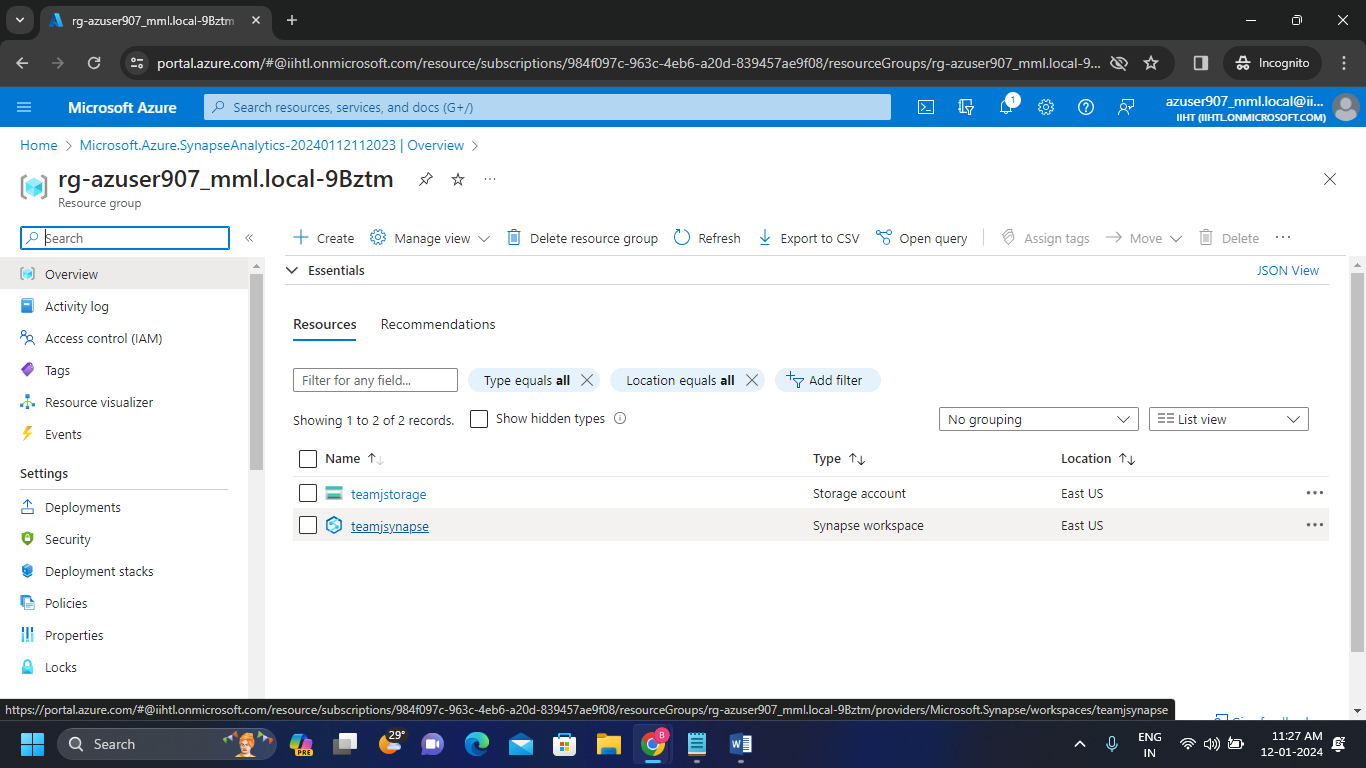
* Fill the necessary data. Here I gave the synapse workspace name as “teamjsynapse”.
* Select the Data Lake Storage Gen2 account if already exists.
* If not exists create new one.
* Here we created new Data Lake Storage Gen2 account as “teamjstorage”.
* We named the file system (container) as “teamjcontainer”.



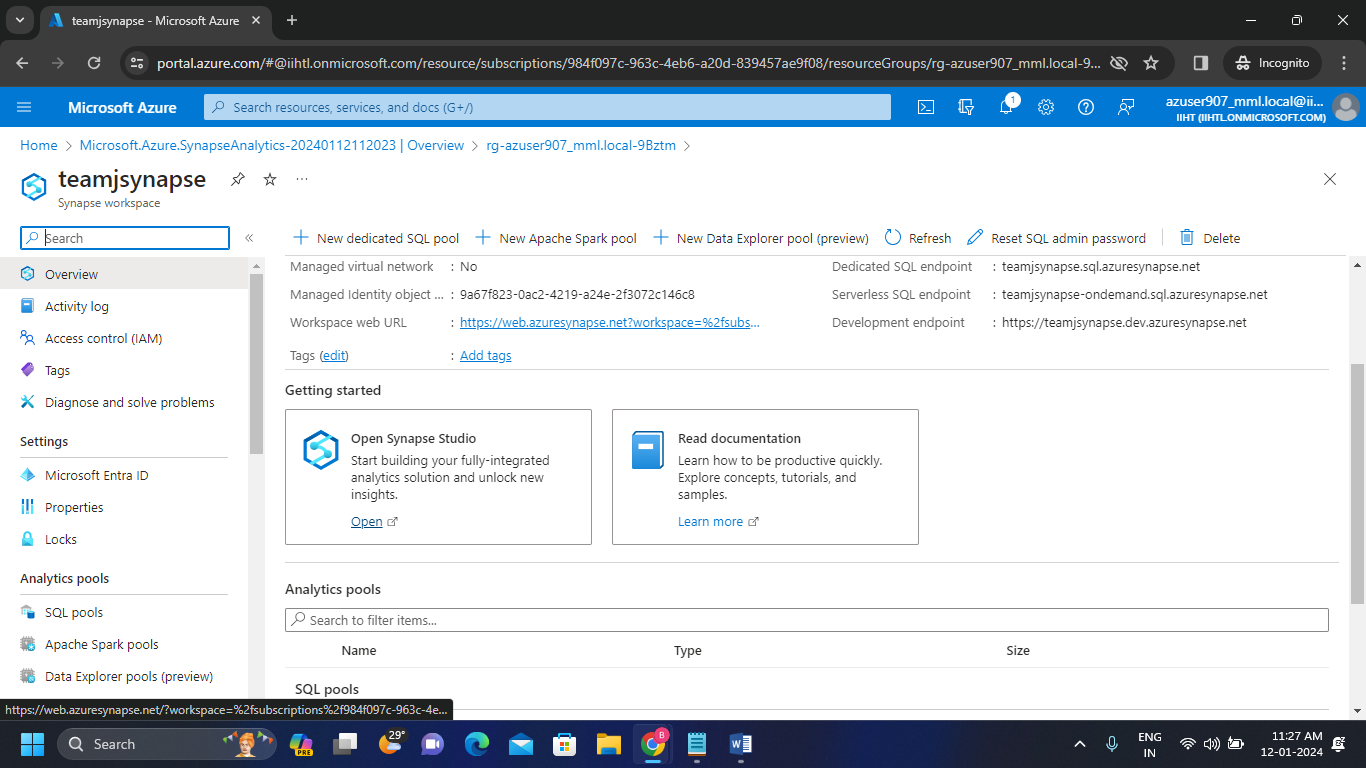
* Review and click on create to create the Azure Synapse workspace.
* Here we can see that the Azure synapse is successfully created.

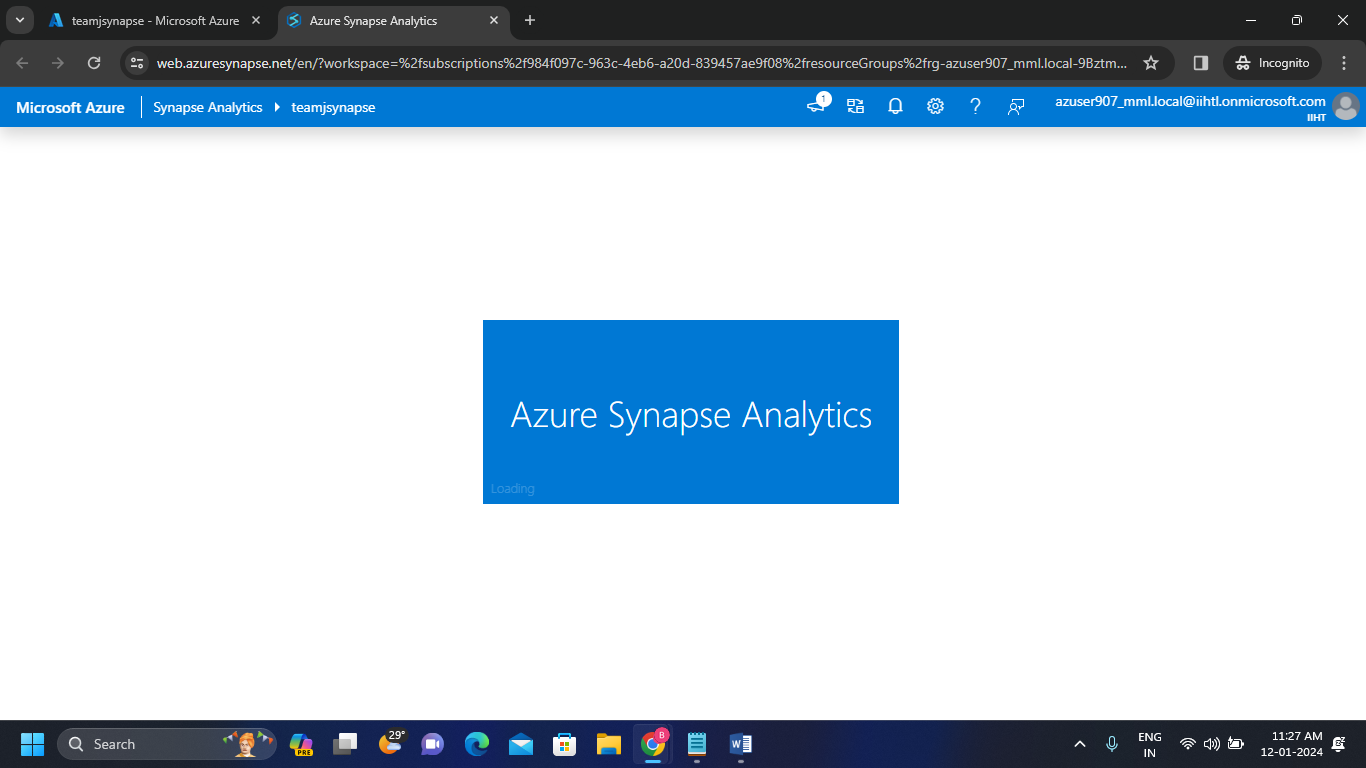
**Step-3:**

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

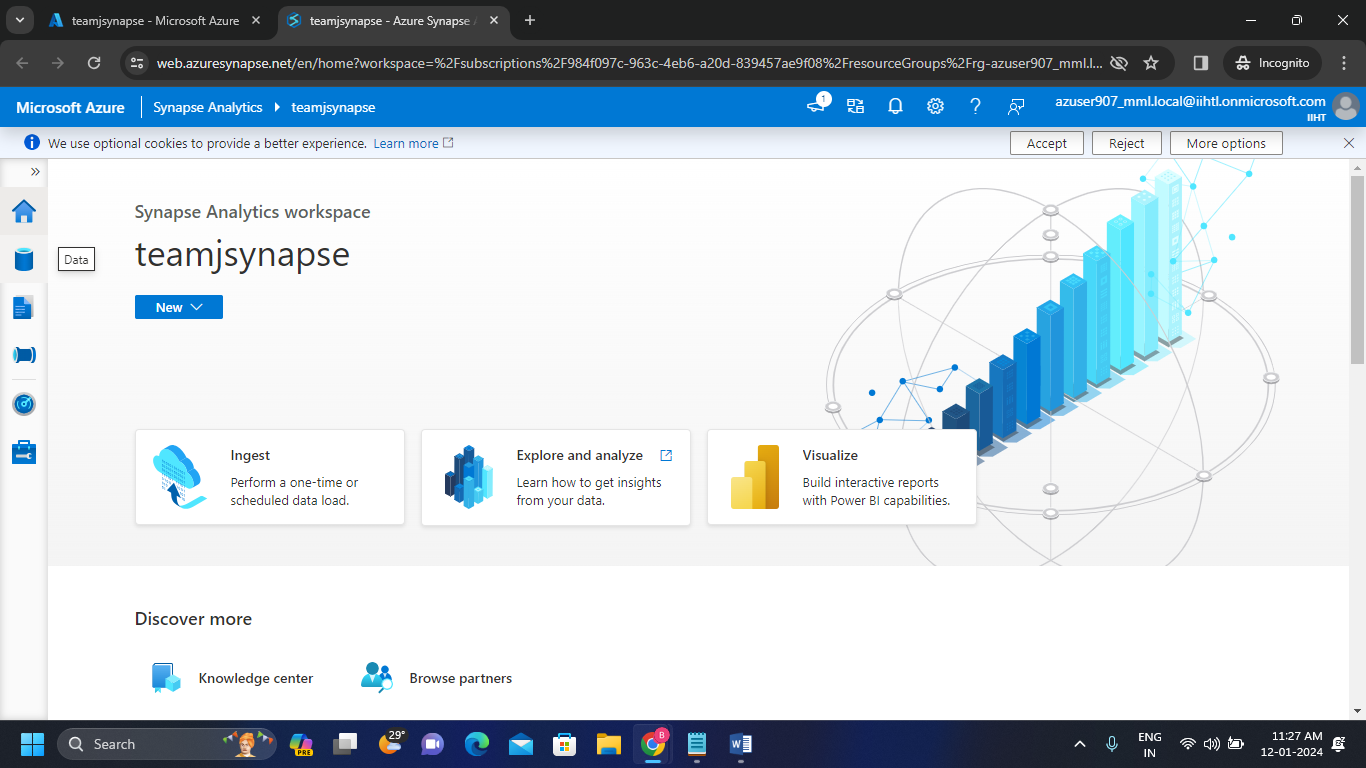
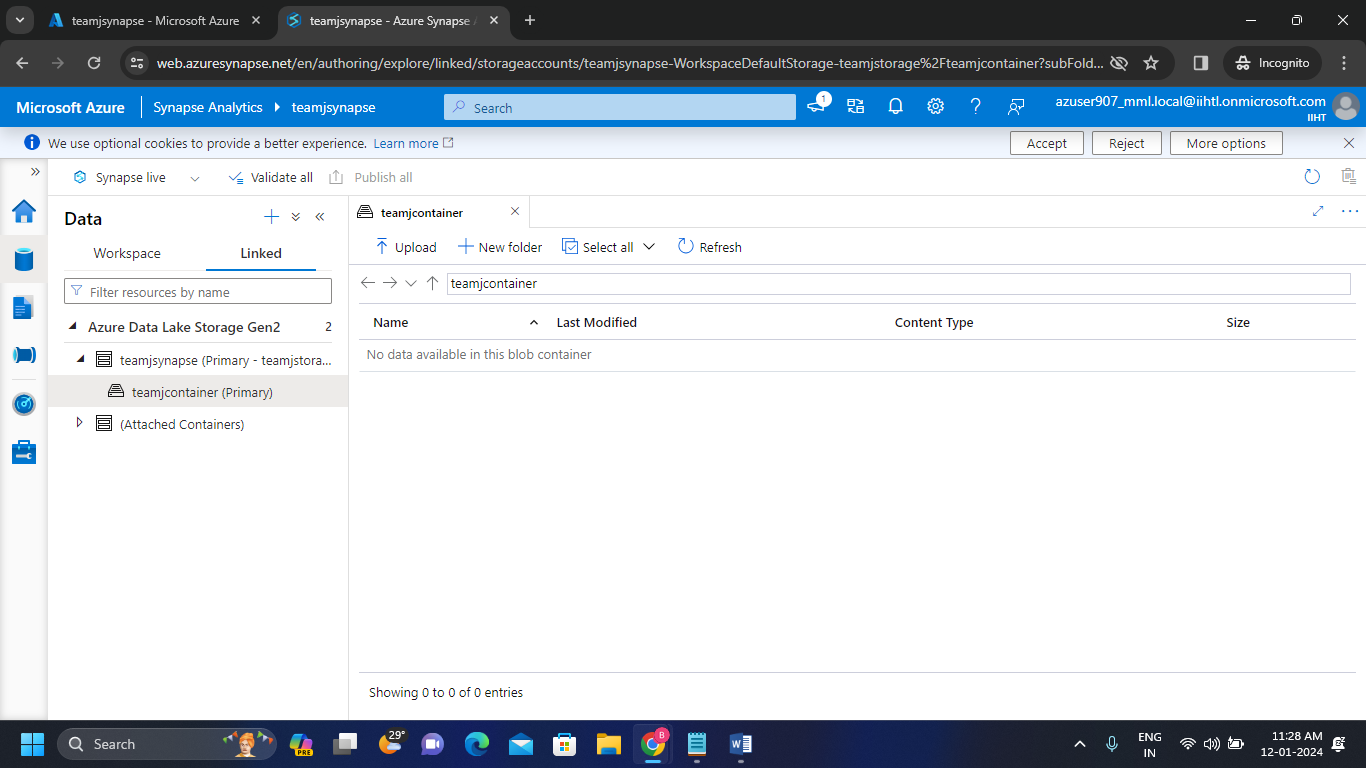


**Step-4:** Now open the Synapse studio.

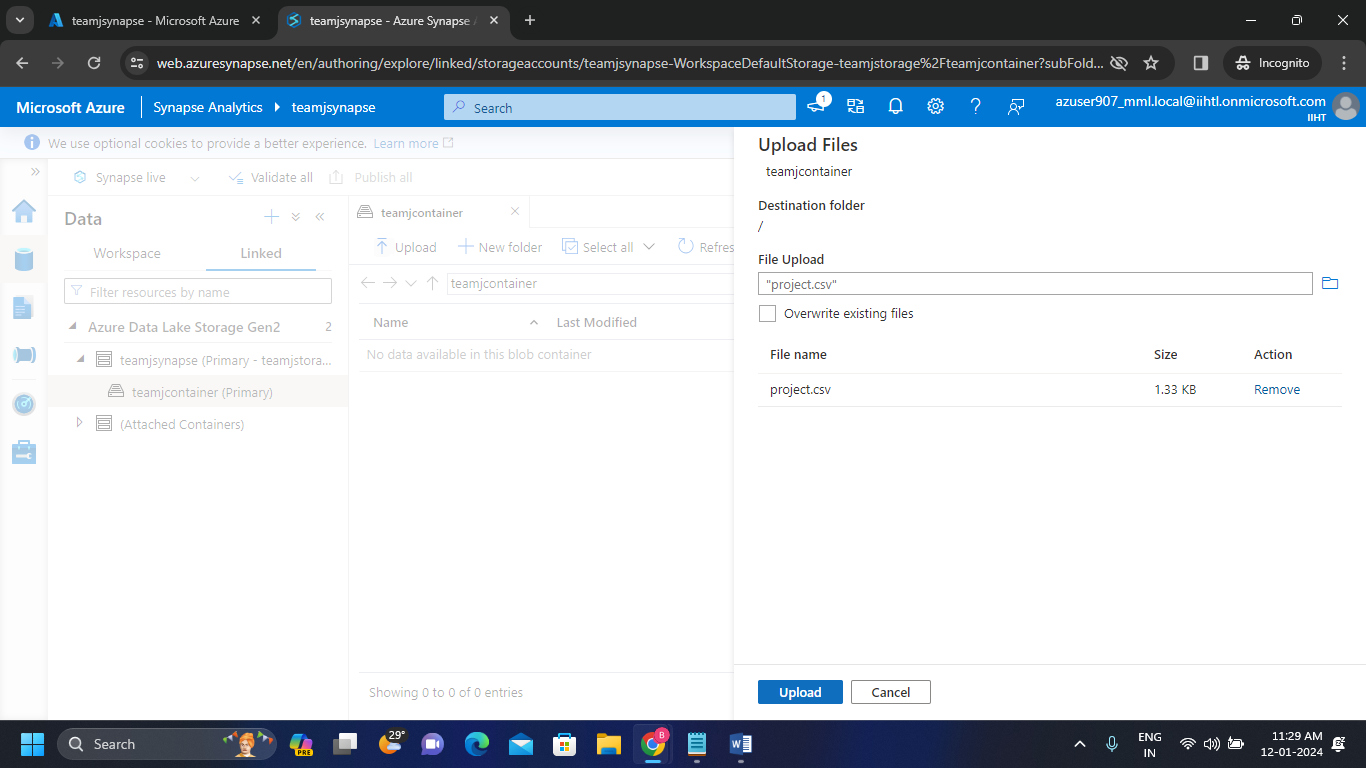
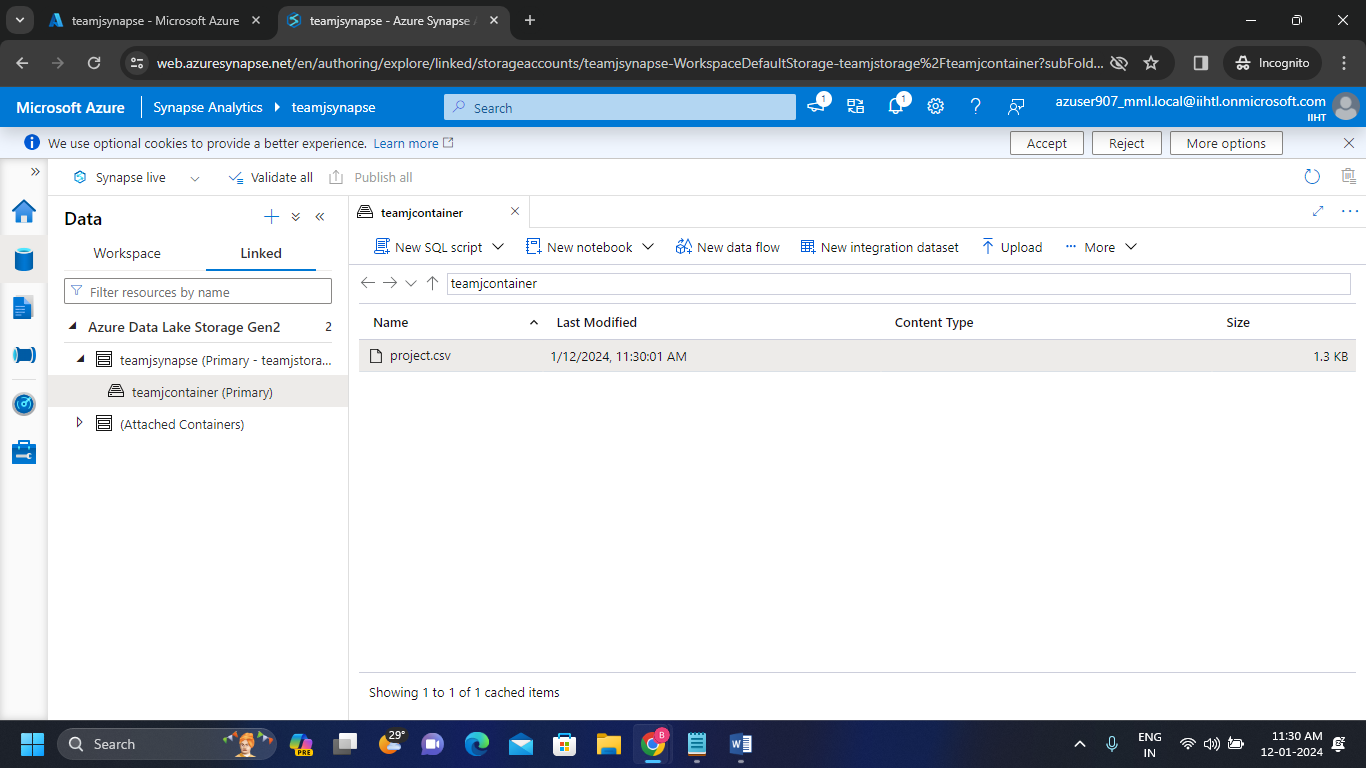




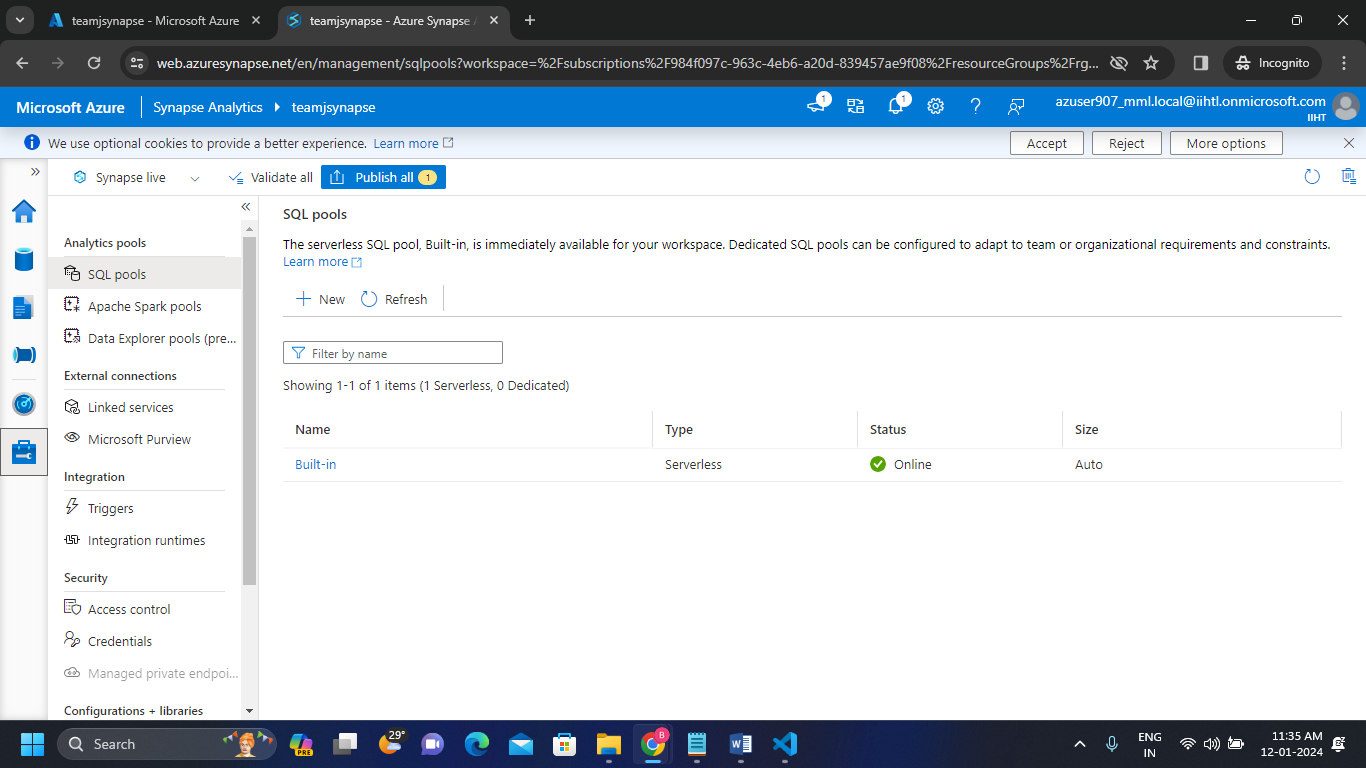
**Step-5:**

* Below we can see the Synapse workspace.
* By clicking Data navigator we can able to see the linked data storage accounts in the container.
* Here we can see the teamjsynapse Data lake storage Gen2 account that we created and also the teamjcontainer.

**Step-6:**

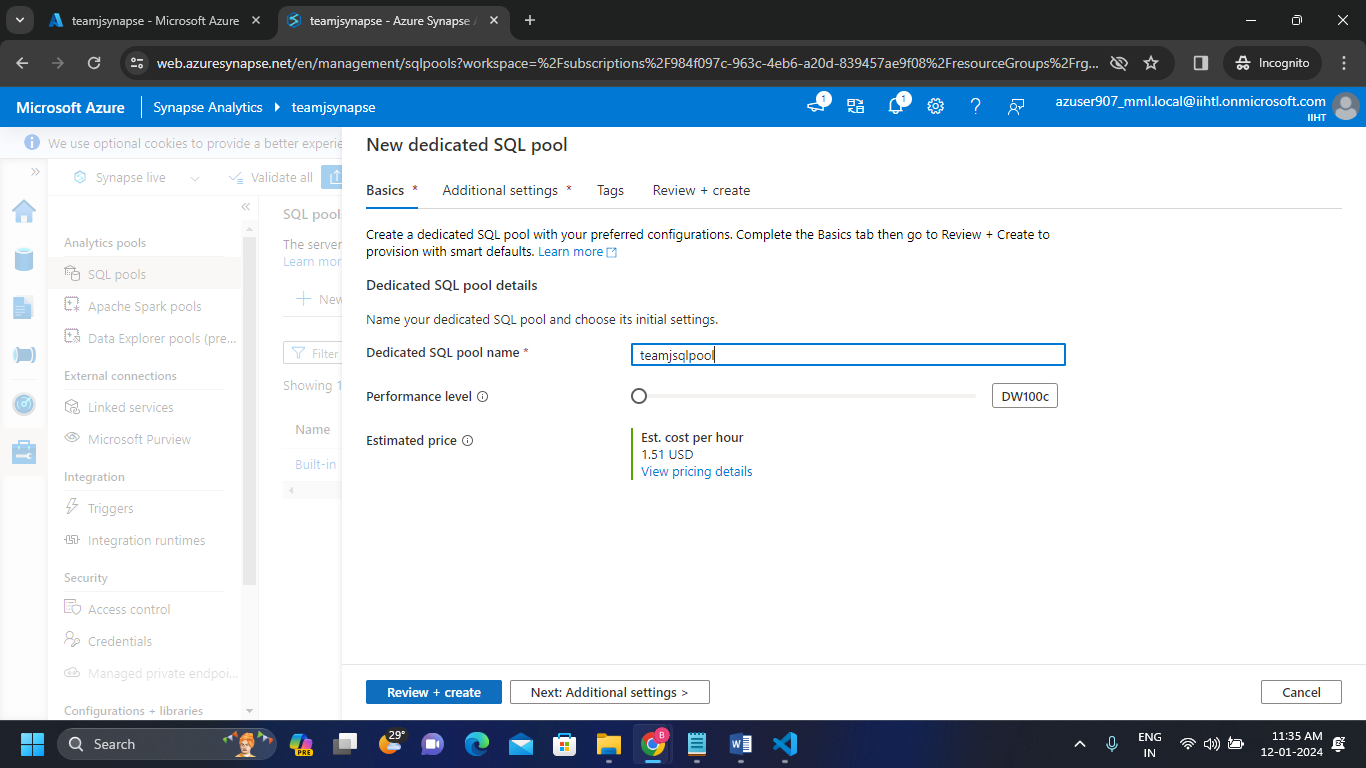
* After opening the container we uploaded our “project.csv” file in the container.
* We can see the uploaded file inside the container.

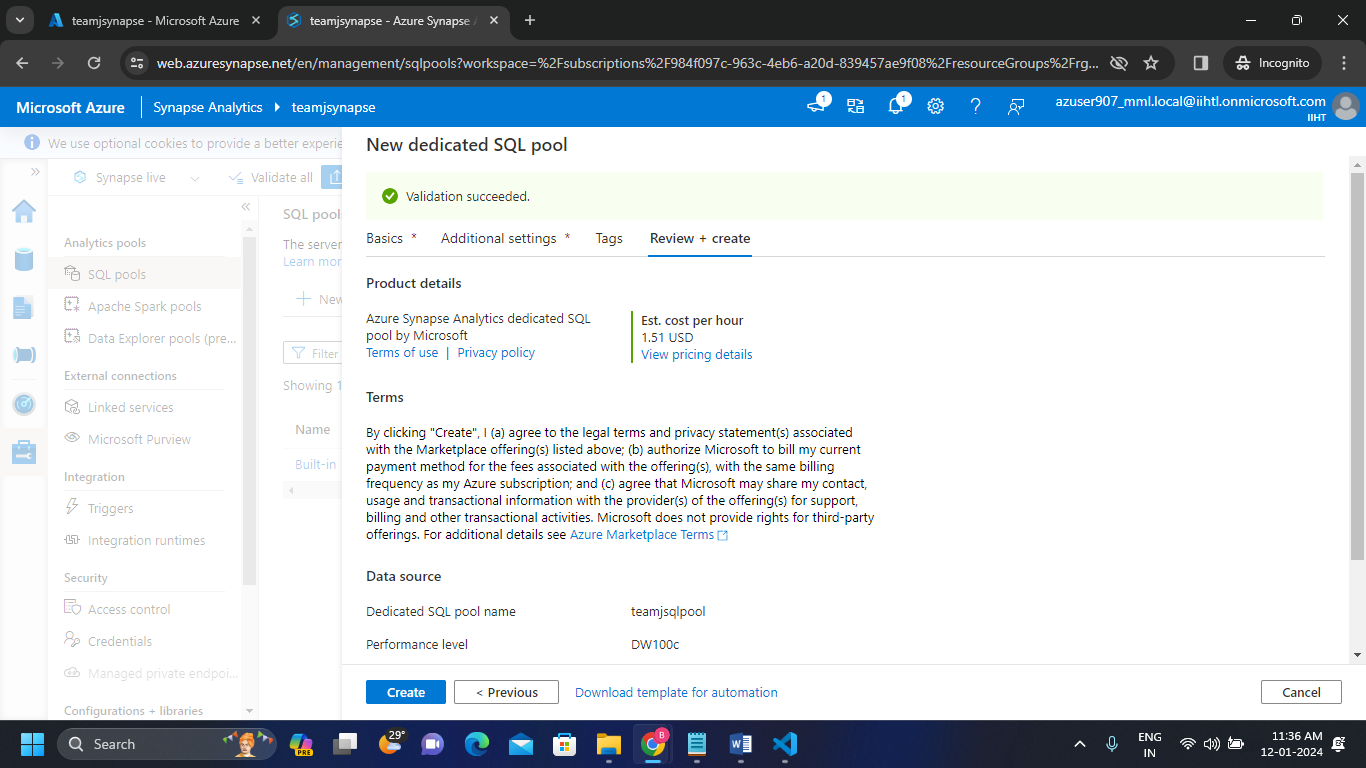
**Step-7:**

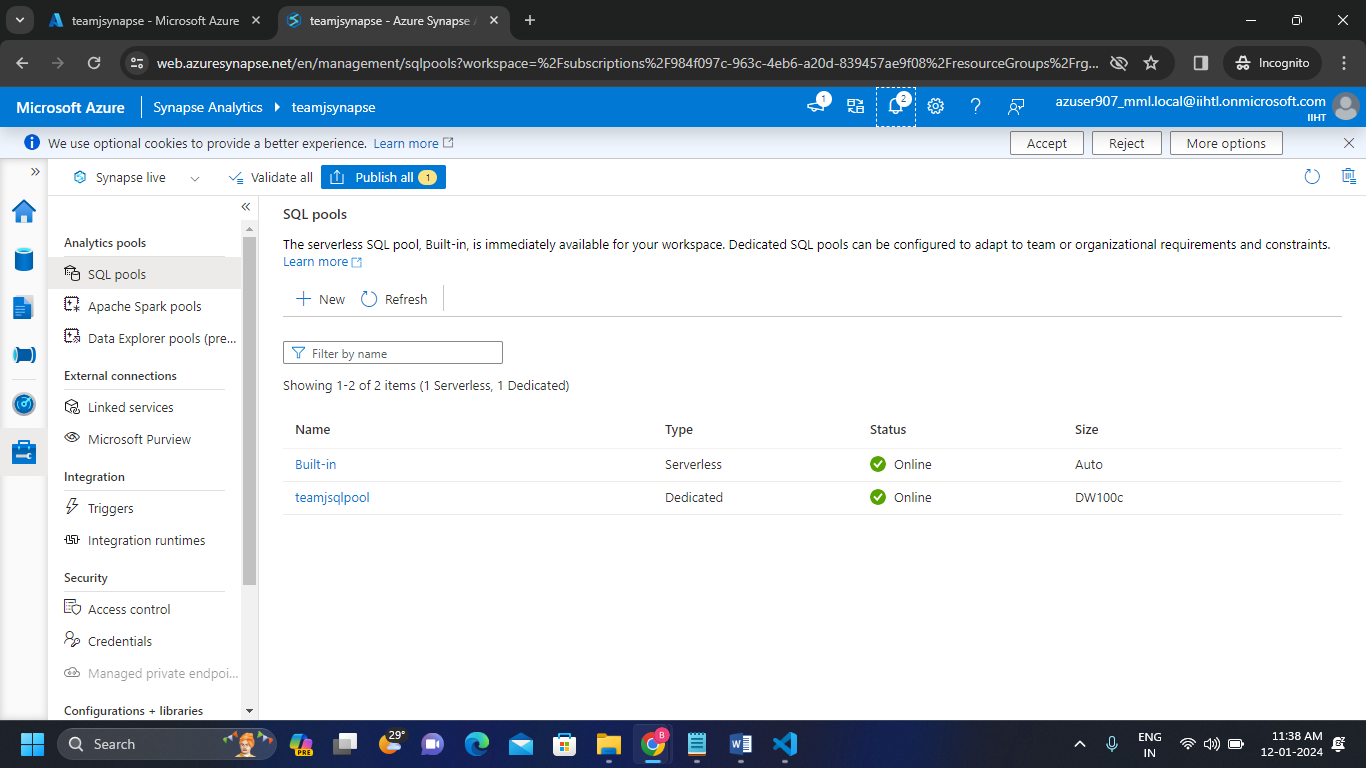
* After uploading csv file into the container open the manage navigator to see SQL pools.
* Here we can see one sql pool which is buit in sql pool used to perform serverless integrations.

**Step-8:**

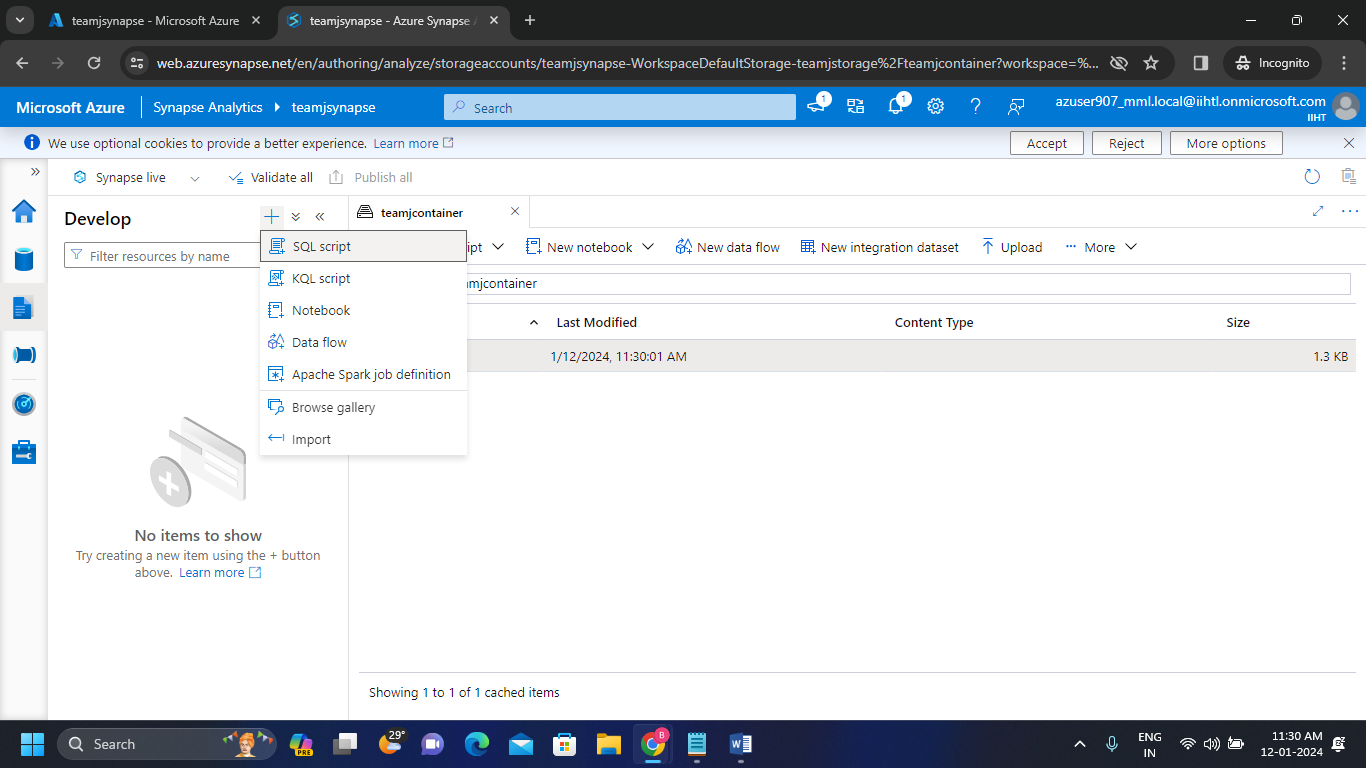
* Here we created Dedicated Sql pool to perform dedicated integrations.





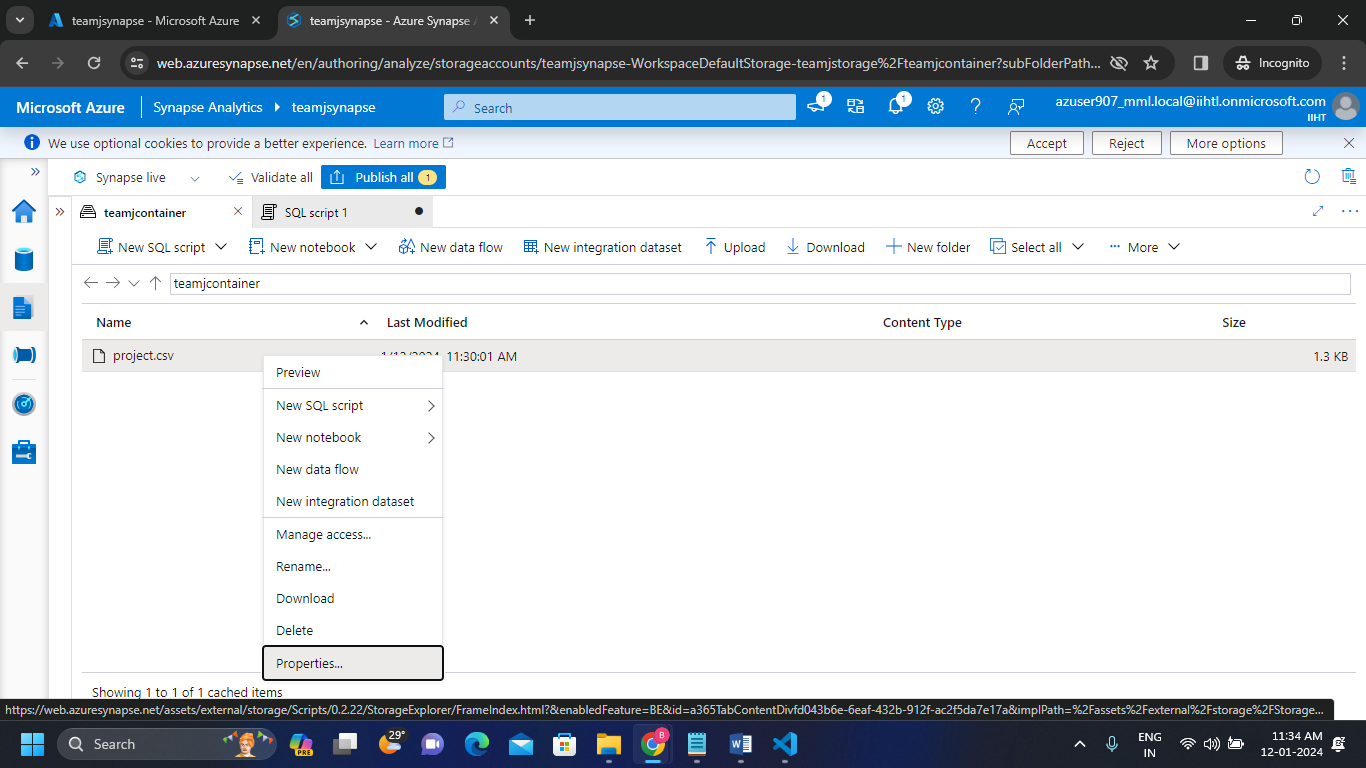
* We can see the Dedicated Sql pool created here.

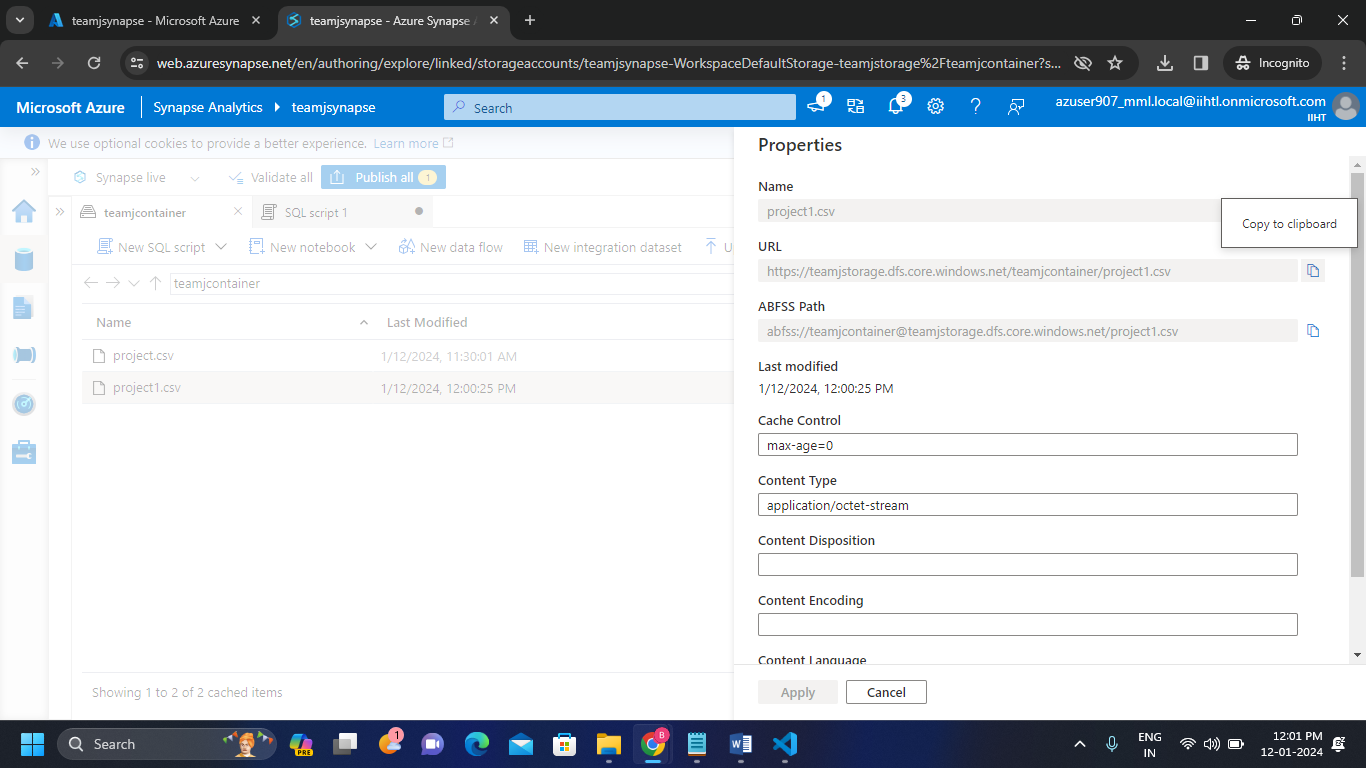
**Step-9:**

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

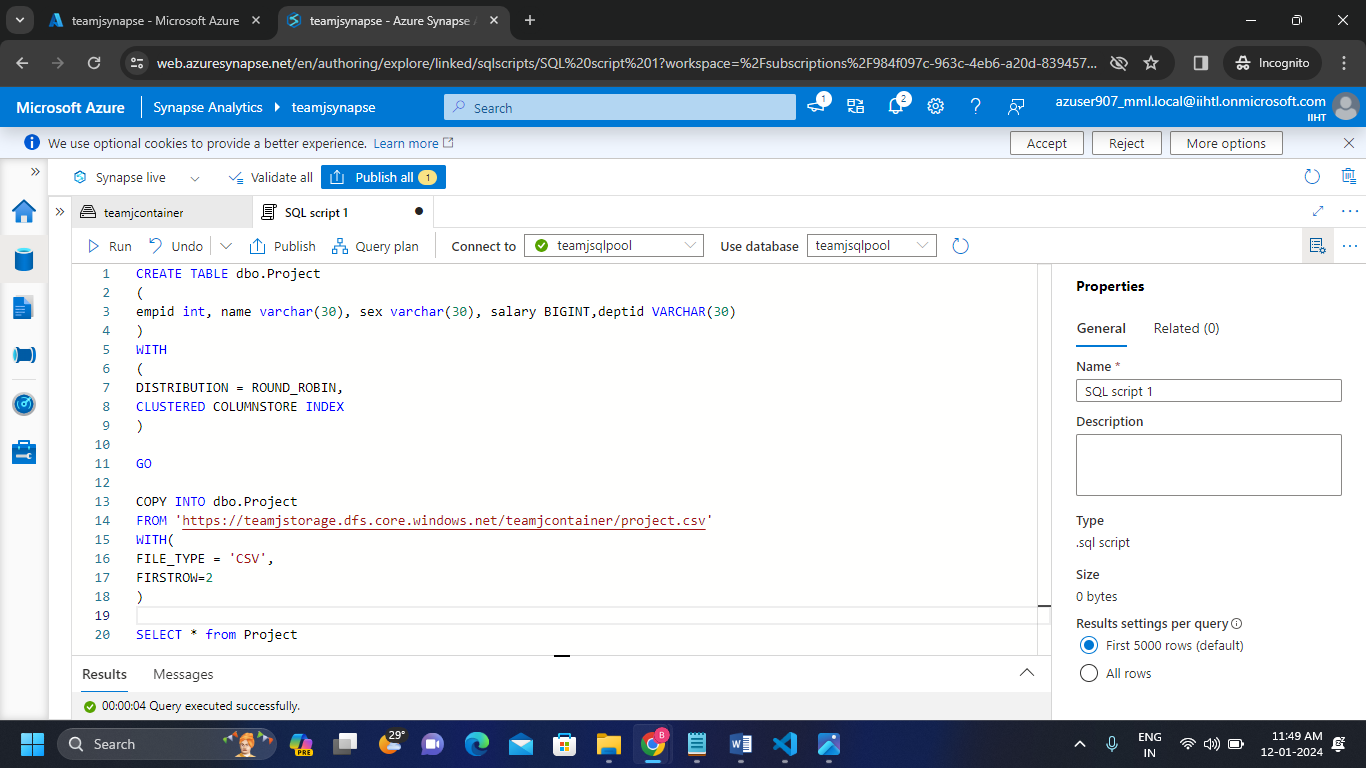
**Step-10:**

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



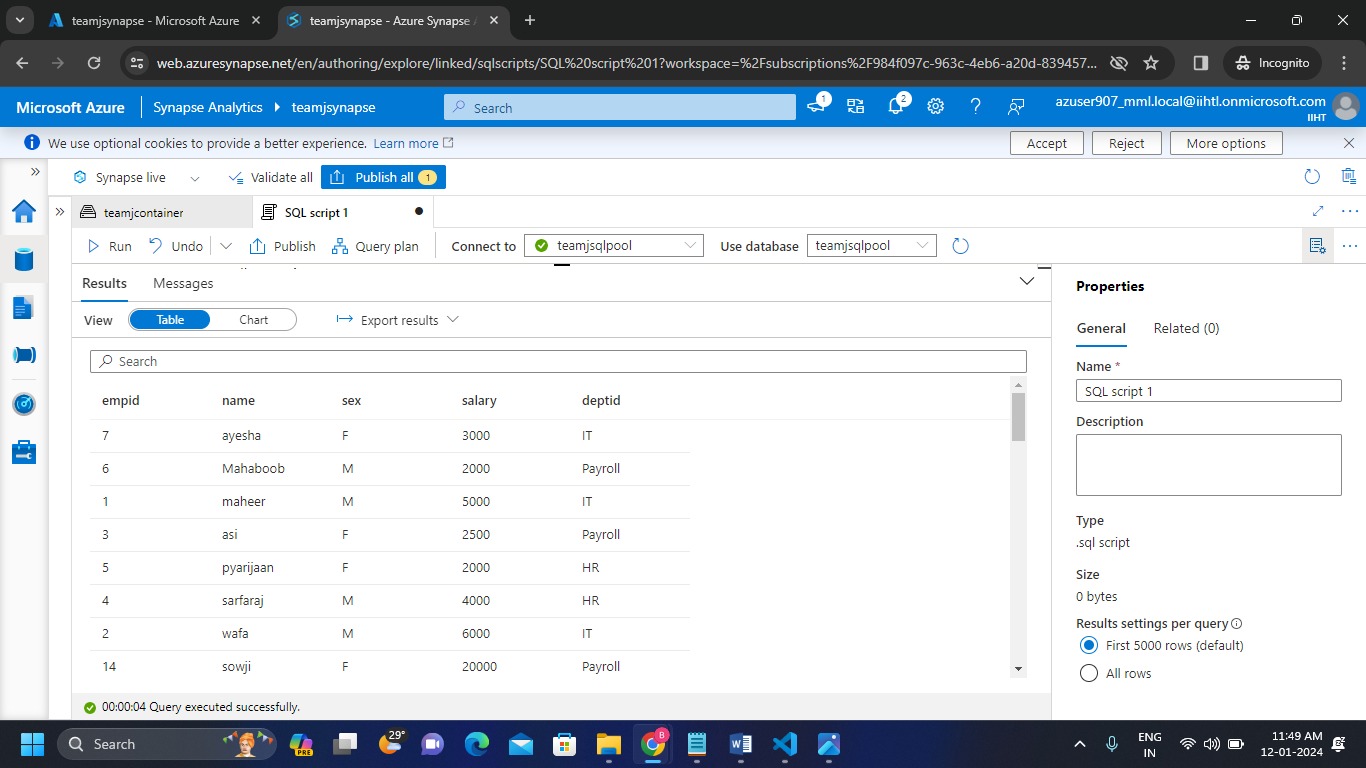
* From here we copied the url of the file.

**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.

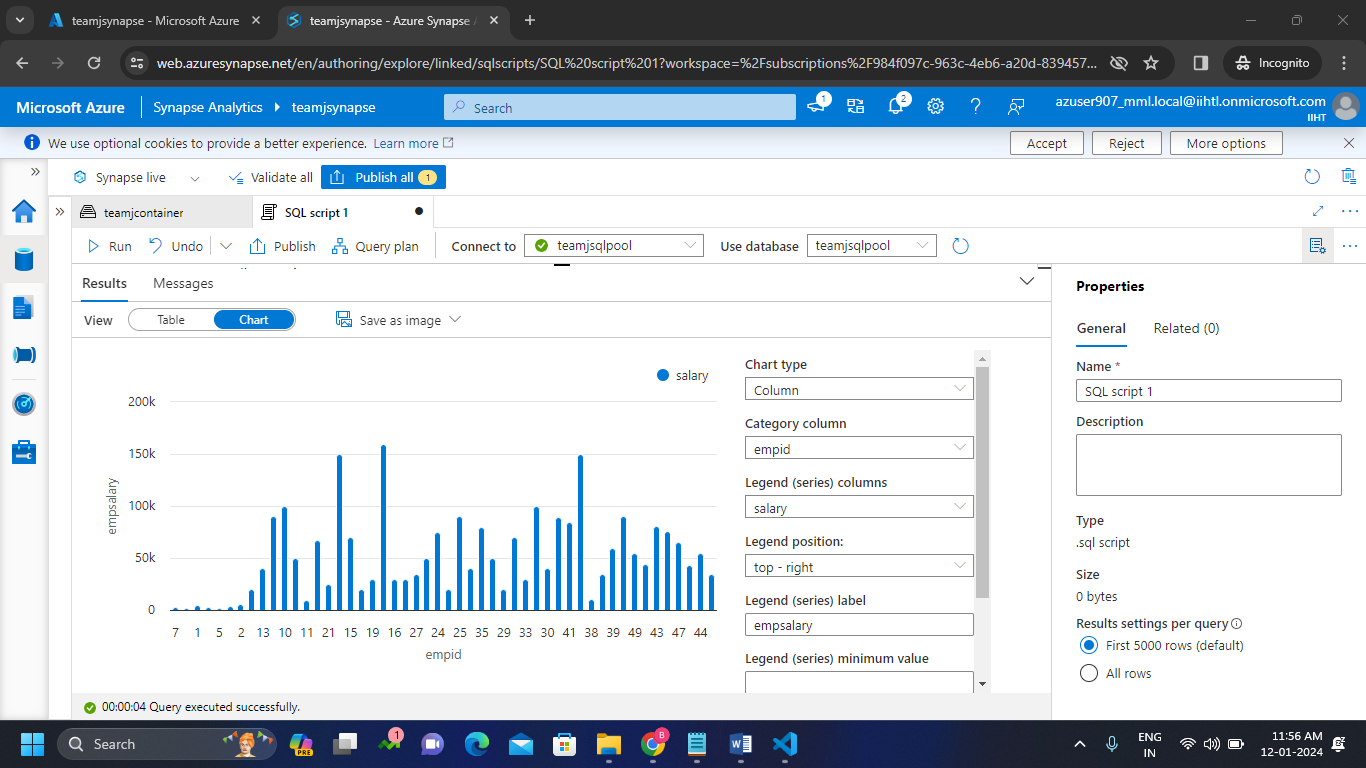
**Step-1:**

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



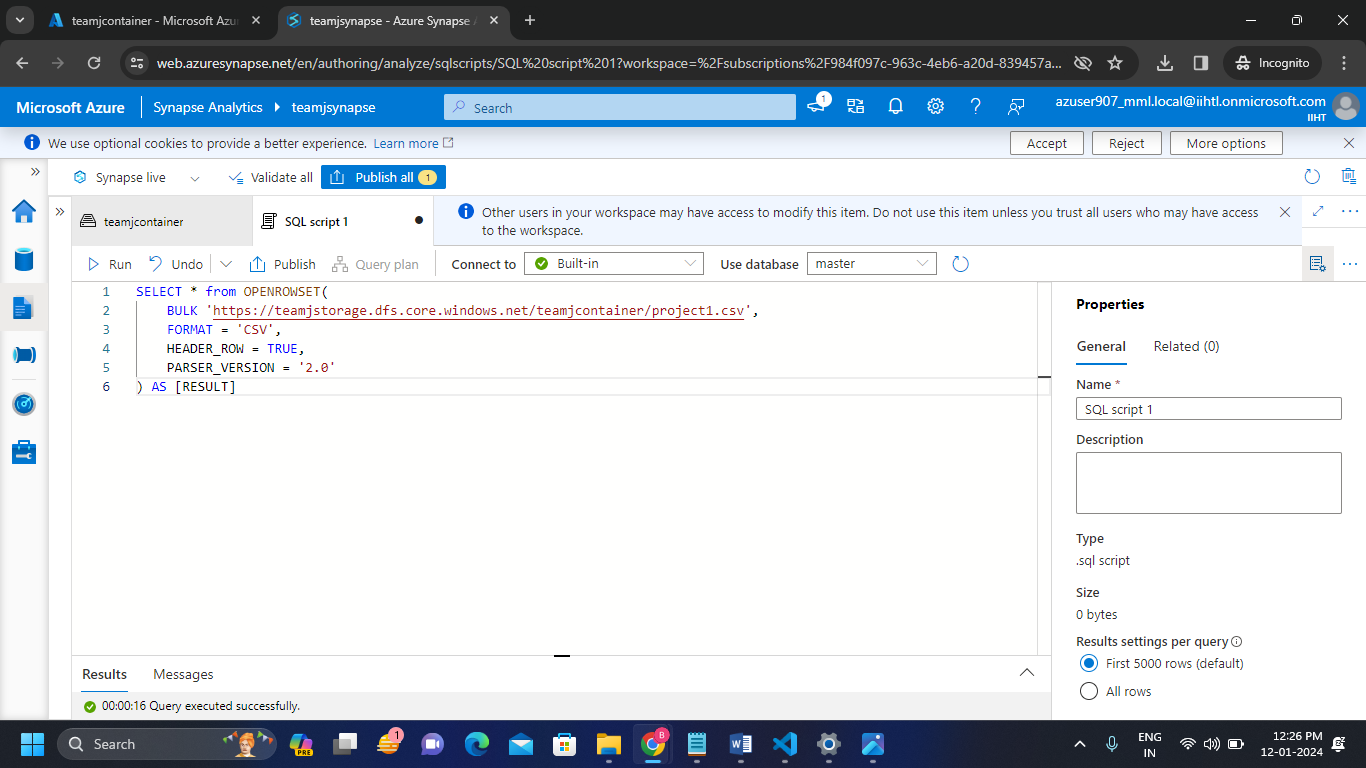
**Step-2:**

* 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.

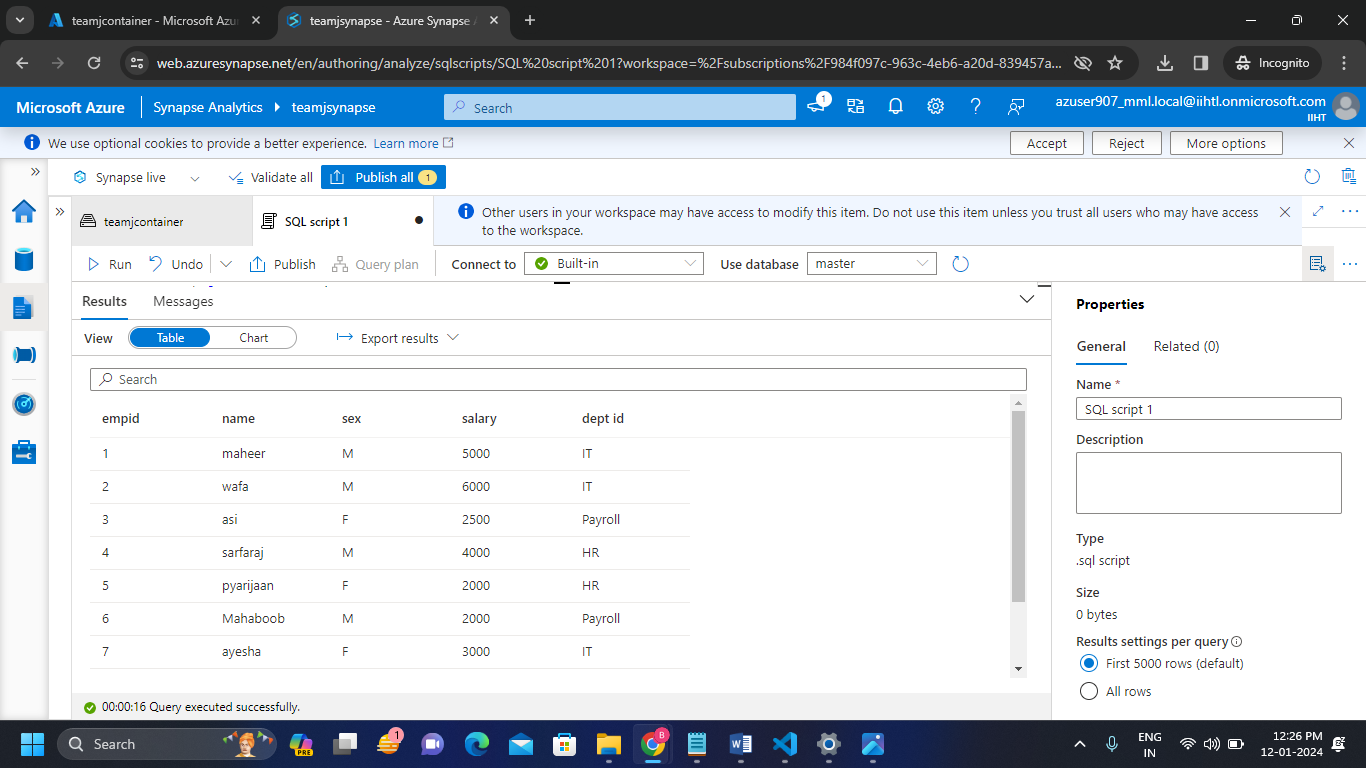


**7.4 Running SQL script in Dedicated SQL pool**

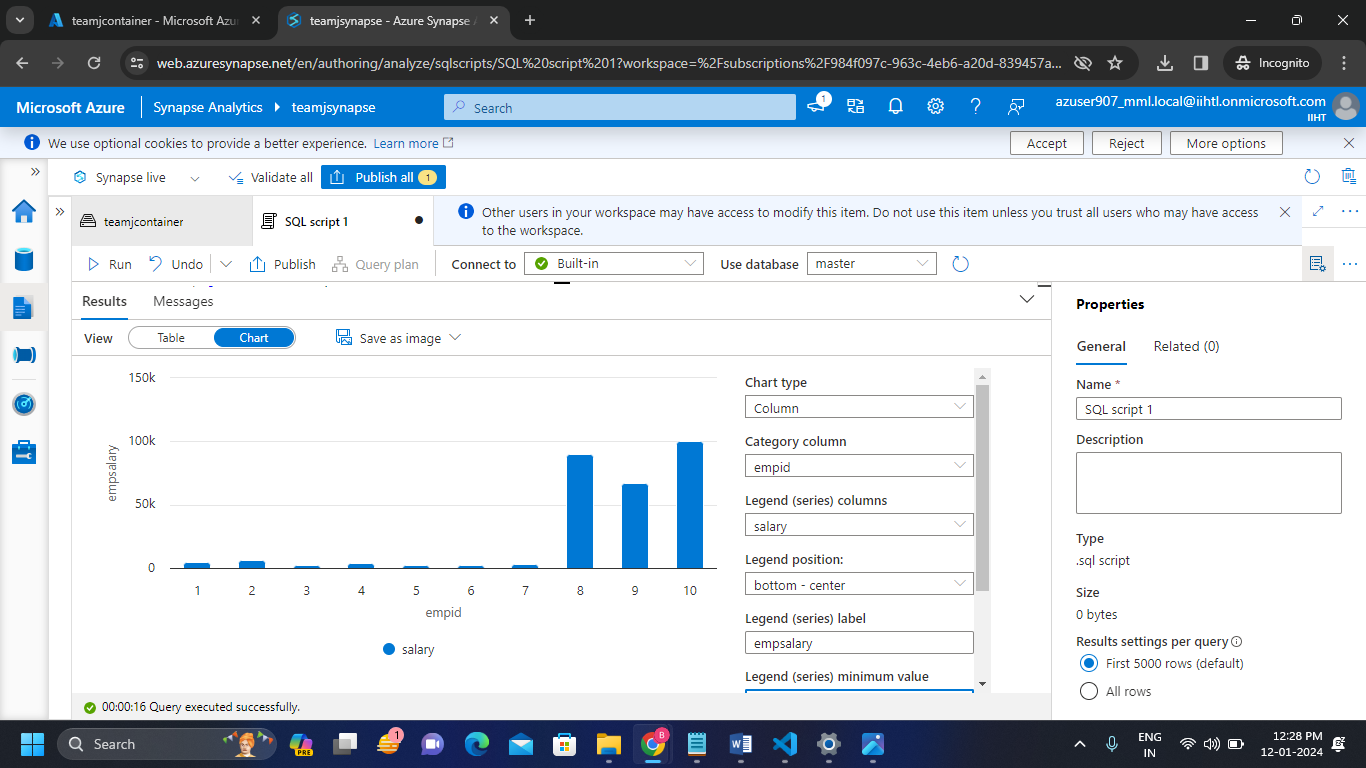
* And also we performed with built-in sql pool.
* We performed some queries in sql script by giving the path of the file.
* Here we used OPENROWSET function which works for serverless sql pools.
* And we connected to Built – in pool then automatically it will use master databse.
* Then Click on Run button.



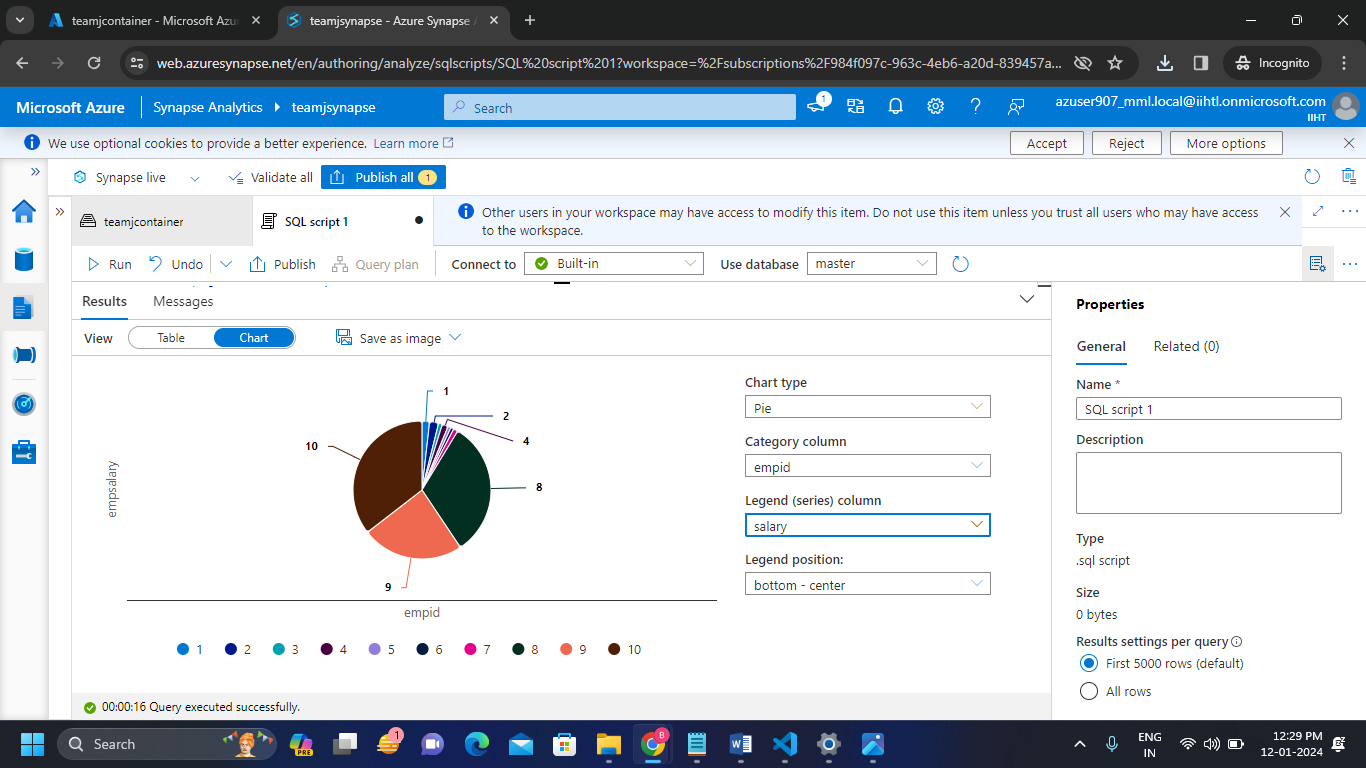
* Here We can see the output of the above code which is in Tabular Format.



* Visualizing the data in bar graph format.



* Visualizing the data using Pie-chart.



**Conclusion:** Hence I successfullySet 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.