😊 Loading data from Pipelines – Storage Account

In Azure Synapse Analytics, pipelines are a fundamental concept used in the development and orchestration of data workflows. A pipeline is a logical grouping of activities that perform a series of tasks to move, transform, or analyze data. Azure Synapse provides a robust set of tools for building, scheduling, and monitoring pipelines, allowing organizations to automate and streamline their data integration, ETL (Extract, Transform, Load), and analytics processes.

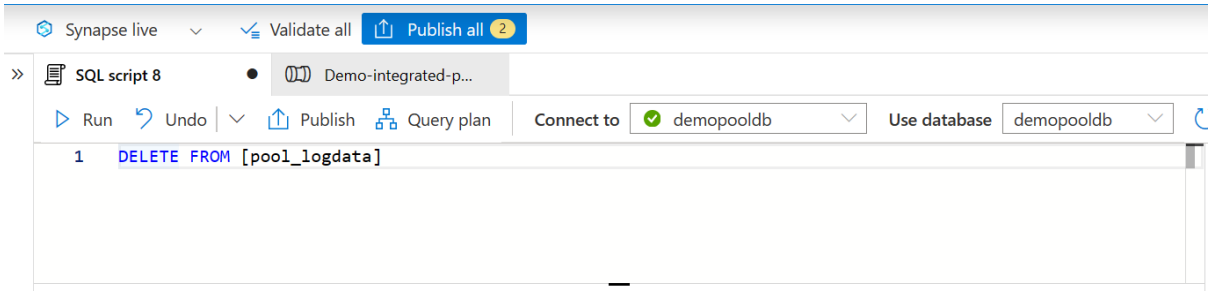
Here are key components and features of pipelines in Azure Synapse:

1. **Activities:** Activities are the building blocks of pipelines. Each activity represents a task or operation that performs a specific action on data, such as copying data, transforming data using SQL or Spark, running a machine learning model, or executing a custom script.
2. **Connectors:** Azure Synapse provides a wide range of connectors to interact with various data sources and destinations, including Azure Blob Storage, Azure Data Lake Storage, Azure SQL Database, Azure Cosmos DB, and more. Connectors simplify data integration by providing pre-built connectivity to different data platforms.
3. **Data Movement and Transformation:** Pipelines enable data movement and transformation operations to extract data from source systems, transform it using SQL or Spark transformations, and load it into target systems. This enables organizations to implement ETL processes for data integration and data preparation tasks.
4. **Monitoring and Logging:** Azure Synapse provides monitoring and logging capabilities to track the execution of pipelines in real-time. You can monitor pipeline runs, view execution statistics, and diagnose issues using built-in monitoring tools and logs, enabling proactive management of data workflows.
5. **Triggers:** Triggers define when a pipeline should be executed. Azure Synapse supports various trigger types, including schedule-based triggers, event-based triggers, and manual triggers. Triggers allow you to automate the execution of pipelines based on predefined schedules or events, ensuring timely data processing and analysis.
6. **Parameters and Variables:** Pipelines can utilize parameters and variables to dynamically configure and control pipeline behavior at runtime. Parameters enable you to pass inputs to pipelines, while variables allow you to store and manipulate data within pipelines, enhancing flexibility and reusability.
7. **Integration with Azure Services:** Azure Synapse integrates seamlessly with other Azure services such as Azure Data Factory, Azure Databricks, Azure Machine Learning, and Azure Monitor. This enables end-to-end data workflows, advanced analytics, and operational insights across the Azure ecosystem.

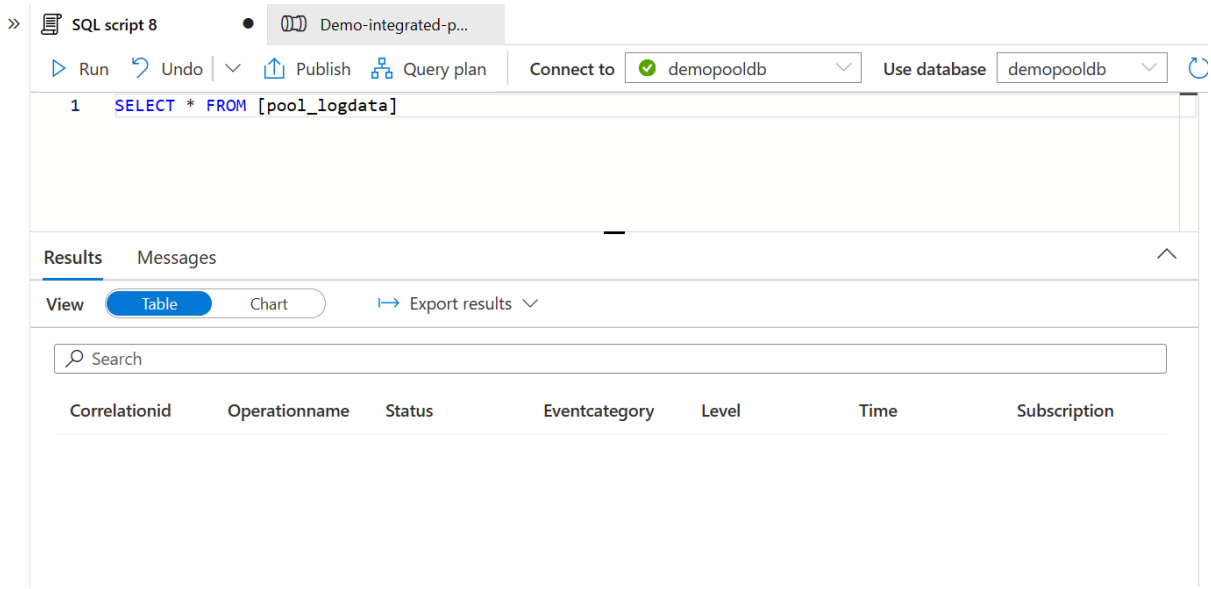
😄 To begin with the Lab:

1. In this lab we are going to see how we copy data from our Data Lake Gen 2 storage account onto Azure Synapse using the pipeline feature that's available in Synapse itself.

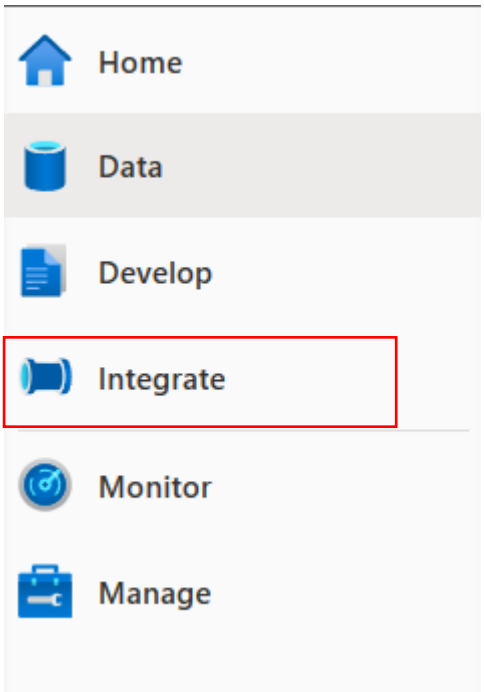
2. First you have to create a new SQL Script.
3. Then you are going to delete the data in your dedicated SQL Pool.



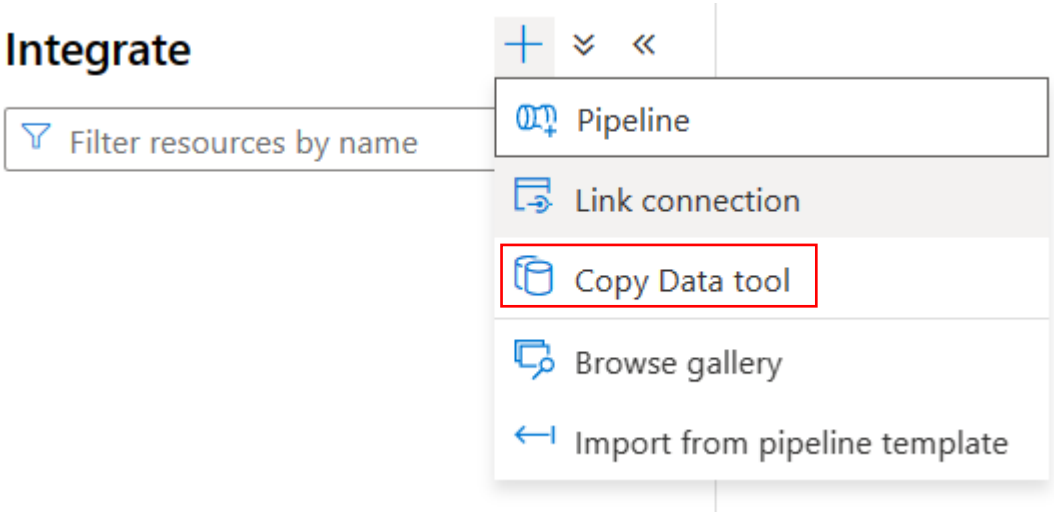
4. Below you can see that there is no data available.



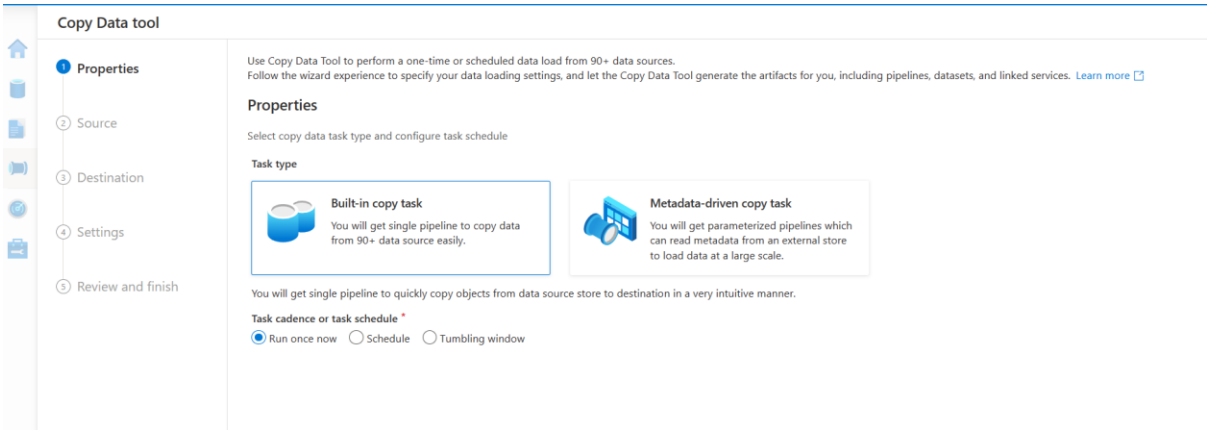
5. Now from the left pane you can see the Integrate section, navigate towards it.



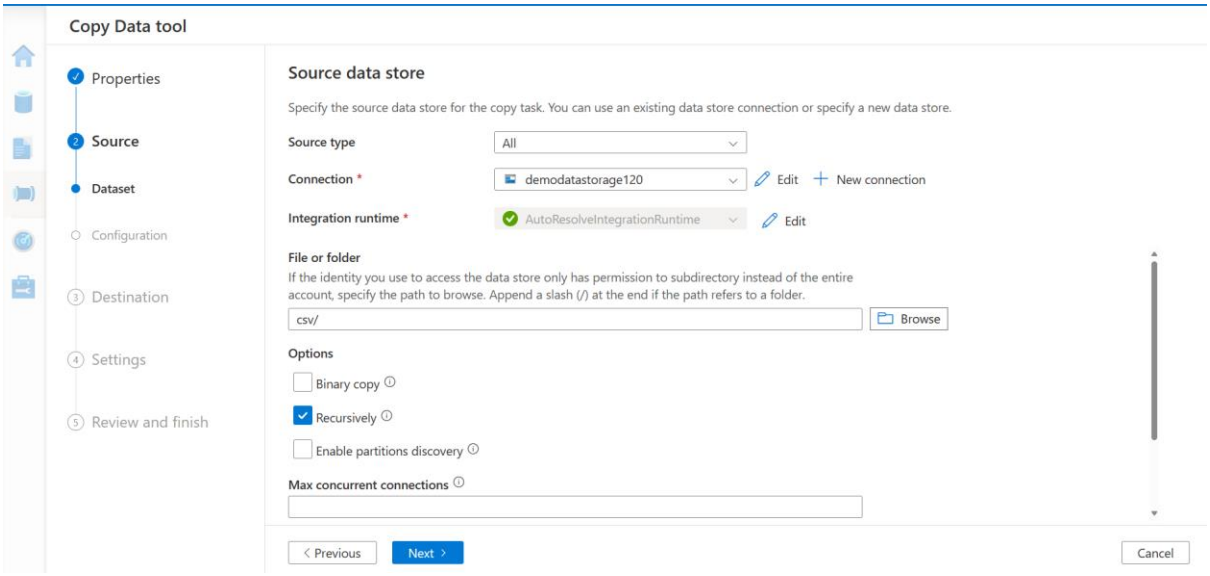
6. Now in Integrate if you click on the plus icon then you will have some options, from these options you must choose the Copy Data tool.



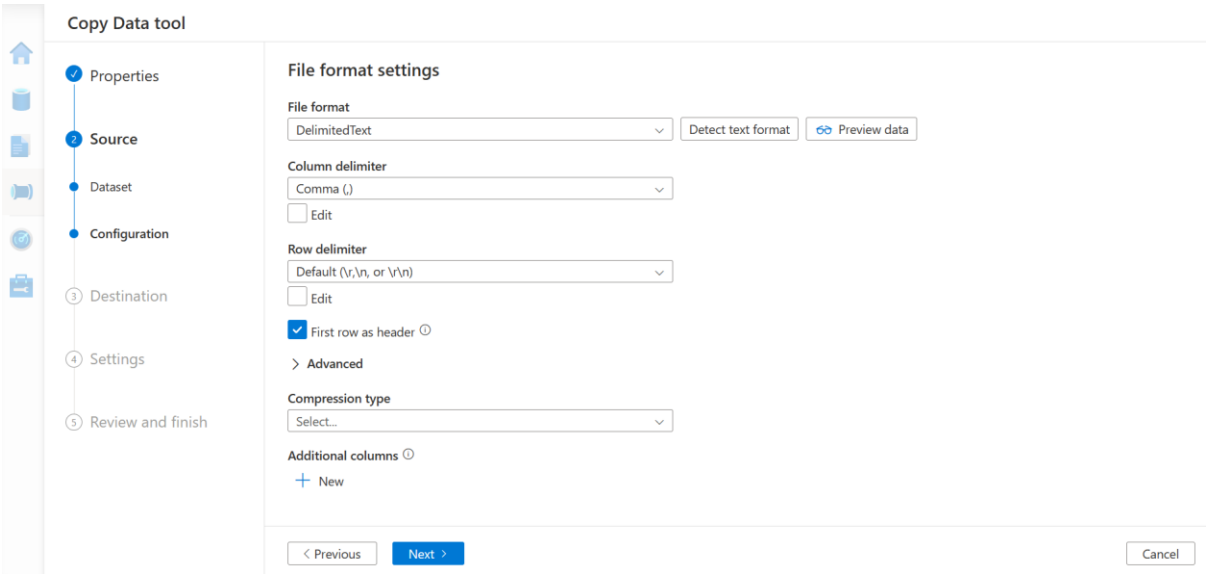
7. Once you have clicked on it you will be on this kind of wizard.
8. Now you must click on next and move to step 2.



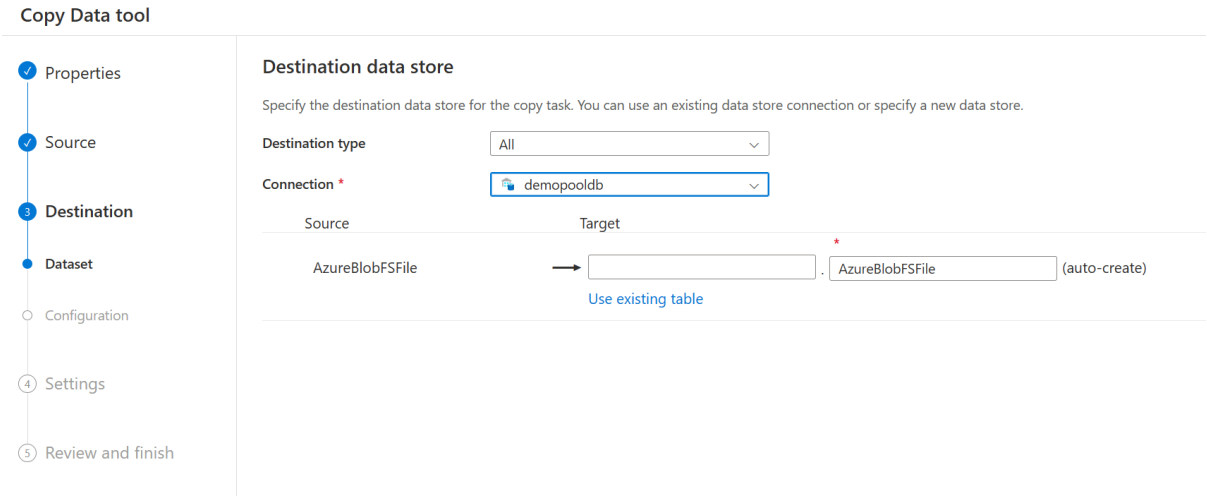
9. Here in the connection, you must choose your storage account and then in the file or folder option choose the CSV container. Then just click on next.



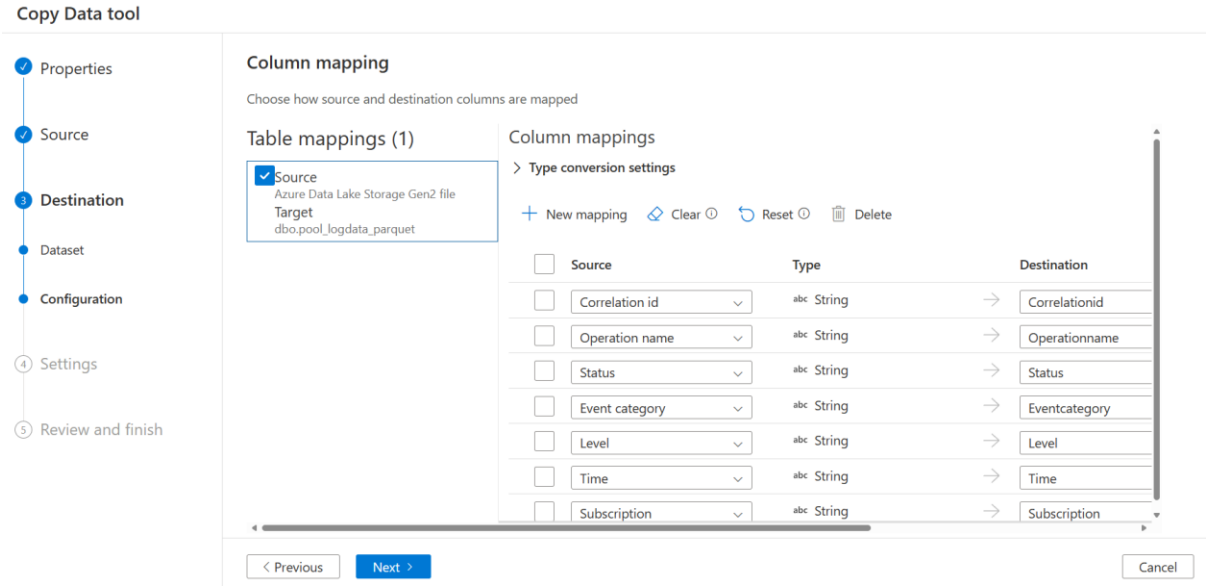
10. Keep everything to default on the next page and move forward.



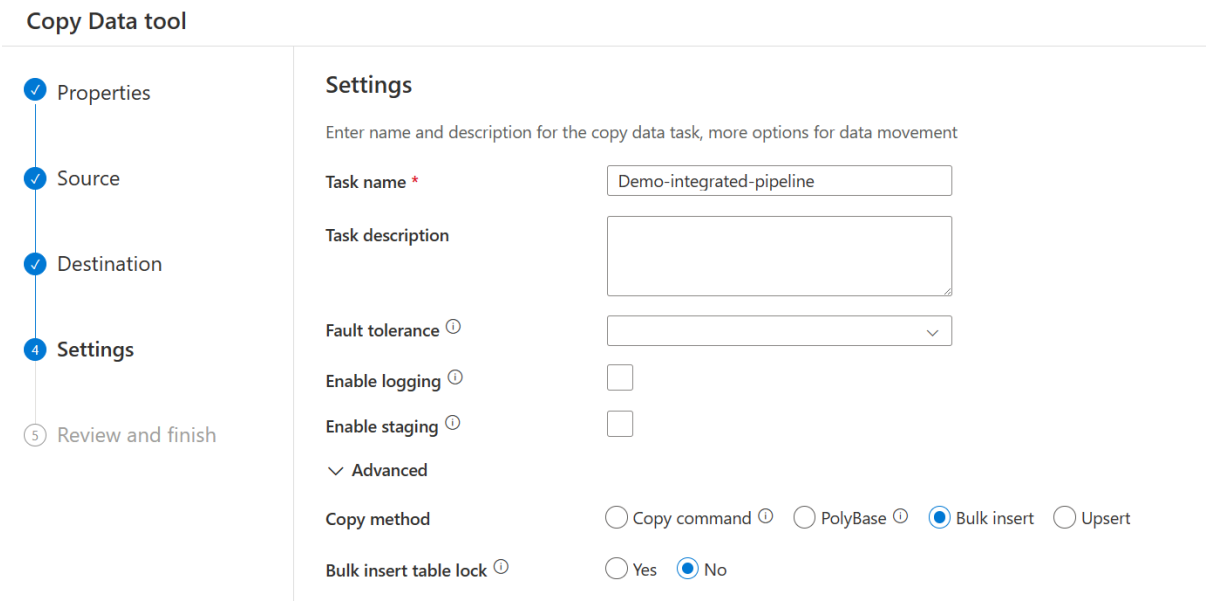
11. Now on the next page you must choose the destination. For that choose your connection as your Pool database and then for file storage click on use existing.
12. So, here you have to choose that existing table in which you just deleted all of your data at the start of this lab. Then click on next.



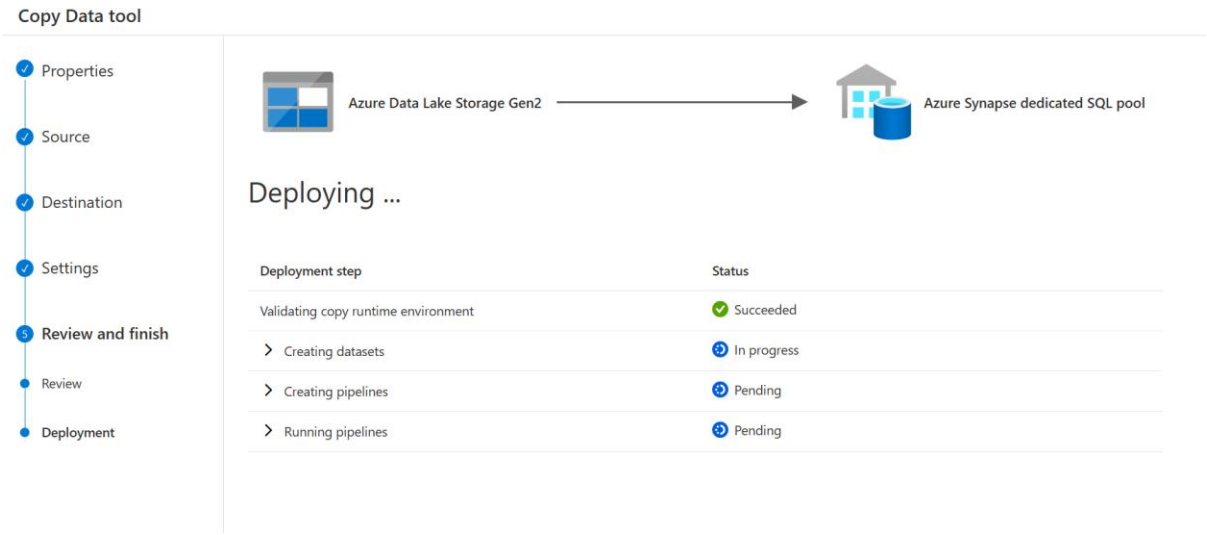
13. On the next page it is doing the table mapping click on next.



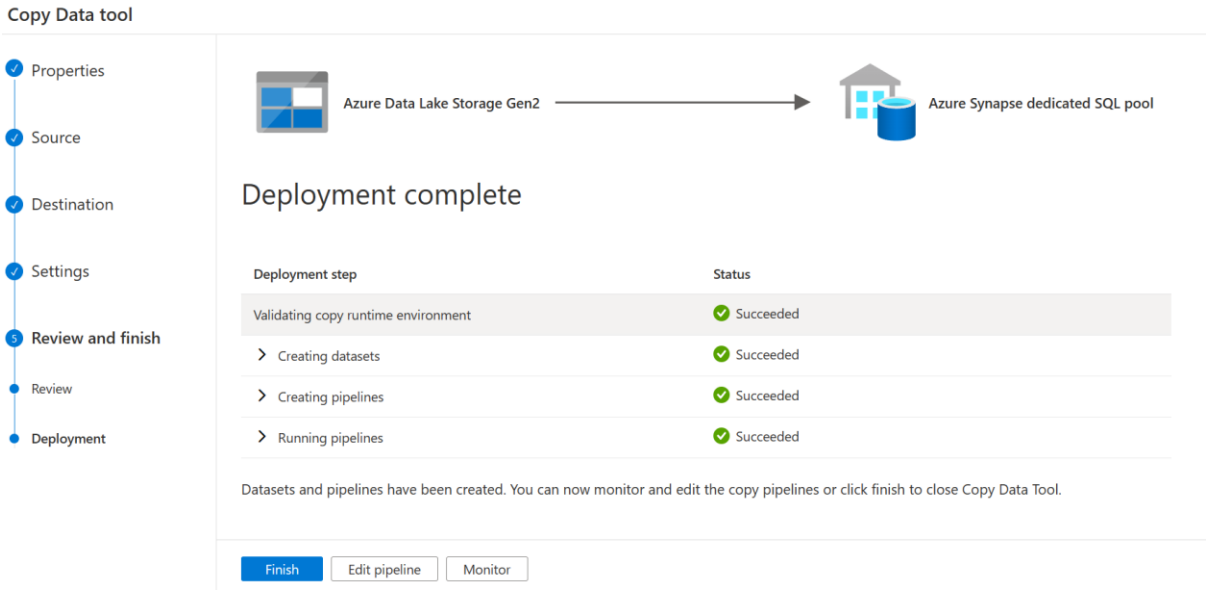
14. Now you must give it the name of your choice and then in the copy method choose Bulk insert.
15. After moving to the review page click on finish.

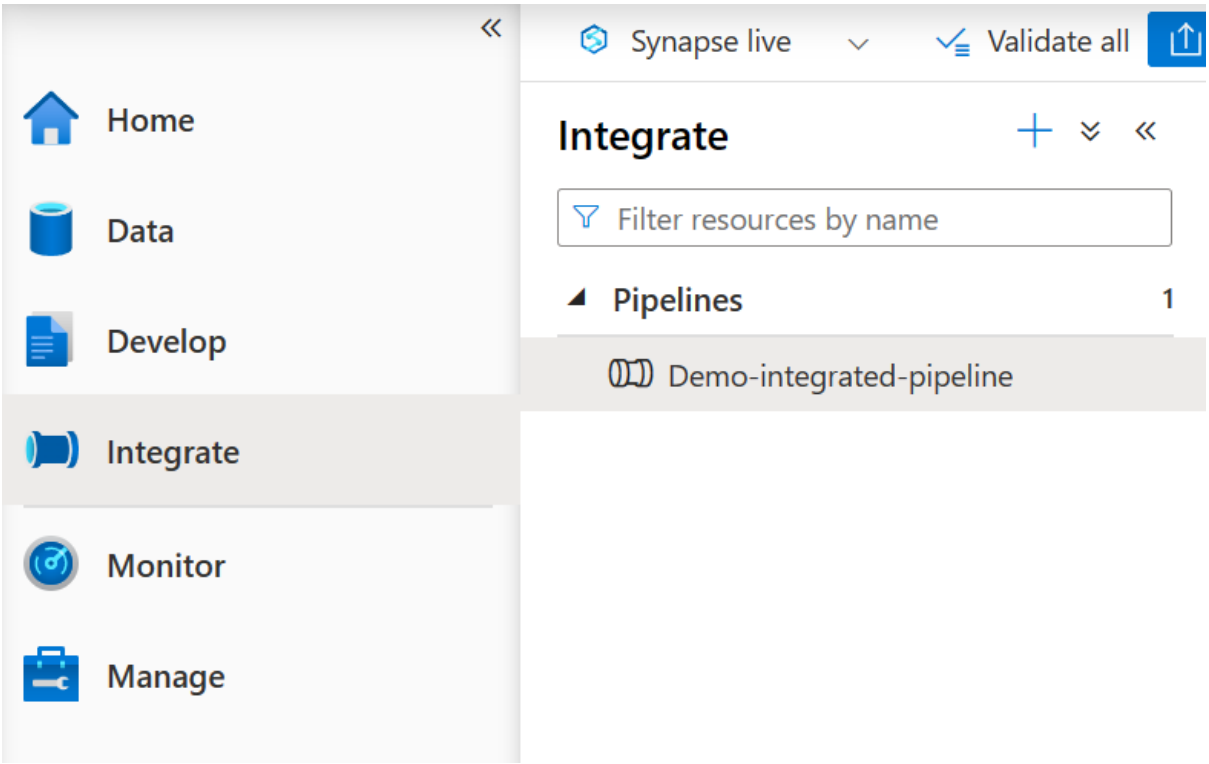


16. Below you can see that it has started the deployment.

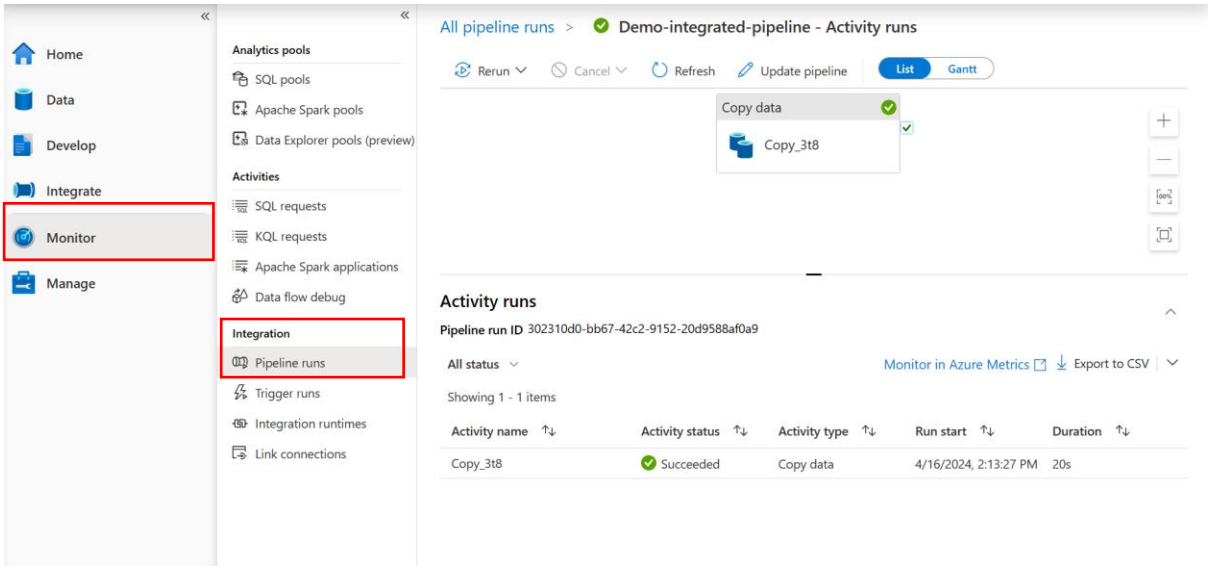


17. Once your deployment is done click on finish. Now in the integrate section, you can see your pipeline.





18. Now if you navigate to the Monitor section from the left pane and go to pipeline runs. Here you can see that you pipeline run was successful.



19. Now come back to your script and run the select command to check whether the data has been copied or not.
20. Below you can see that we have the data.