

Ansh Ranjan

Azure Data

EXERCISE 4: Azure Synapse Analytics

TASK 1: Deploy a sample Azure Synapse Analytics workspace

1. Go to Azure Synapse Analytics > Create > Enter Details

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 *

2. Select or create a Data Lake Gen 2 storage

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

Account name * ⓘ
[Create new](#)

File system name *
[Create new](#)

3. Once again set a sql server admin login and password
4. Click on Review and Create

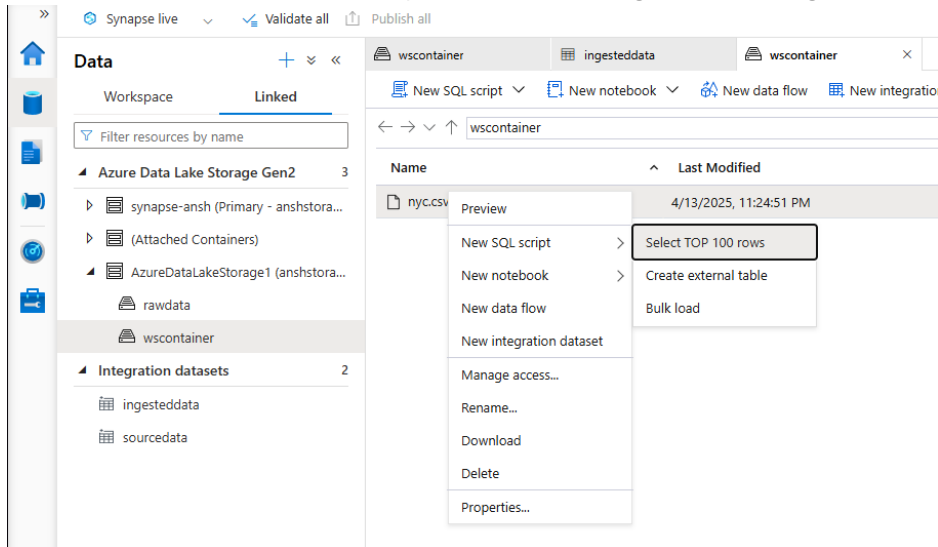
TASK 2: Load a sample dataset and perform basic queries

1. Upload a dataset in your container created

The screenshot shows the Azure Storage browser interface for the 'gen2storageansh' storage account. The left sidebar contains navigation options: Overview, Activity log, Tags, Diagnose and solve problems, Access Control (IAM), Data migration, Events, and Storage browser (selected). The main pane shows the 'ws-container' blob container. The top bar includes 'Add Directory', 'Upload', 'Refresh', and 'Delete' buttons. The 'Authentication method' is set to 'Access key'. A search bar is present with the text 'Search blobs by prefix (case-sensitive)'. Below the search bar, a dropdown menu is set to 'Only show active objects'. The table below shows 'Showing all 1 items' with columns for Name, Last modified, and Access tier. The table contains one row: 'sales_datas...' with a last modified date of '4/10/2025, 3:42:03 PM' and an access tier of 'Hot (Inferred)'.

Name	Last modified	Access tier
sales_datas...	4/10/2025, 3:42:03 PM	Hot (Inferred)

2. Go to Data side tab > browse to your container > right click on ingested data > Select top 100 rows



3. This will write a sql script for you to display first 100 rows of your ingested dataset.

```
1  -- This is auto-generated code
2  SELECT
3      TOP 100 *
4  FROM
5      OPENROWSET(
6          BULK 'https://anshstorageacc1.dfs.core.windows.net/wscontainer/nyc.csv',
7          FORMAT = 'CSV',
8          PARSE_VERSION = '2.0'
9      ) AS [result]
```

4. You can execute the query using the serverless sql pool provided while creating the workspace.
5. However due to firewall limitations we are unable to use either serverless sql pools or dedicated sql pools in this azure account.

TASK 3: Document Azure Synapse Key Benefits and Use cases

Azure Synapse Key Benefits:

1. **Unified Analytics Platform:** Combines big data and data warehousing into a single platform, allowing for streamlined data analysis.
2. **Scalability:** Offers on-demand scalability, enabling you to handle large datasets and workloads without managing infrastructure.
3. **Integrated AI and Machine Learning:** Built-in integration with Azure Machine Learning and cognitive services to run advanced analytics and AI models.
4. **Real-time Analytics:** Supports real-time data streaming and analytics, providing insights with minimal delay.
5. **End-to-End Security:** Features robust security with encryption, firewalls, threat protection, and compliance with industry standards.
6. **Optimized Query Performance:** Leverages in-memory processing, parallel query execution, and caching for faster query performance.
7. **Serverless and Provisioned Models:** Offers both serverless querying and provisioned resources, allowing cost flexibility.
8. **Easy Integration with Other Azure Services:** Seamlessly integrates with tools like Power BI, Azure Data Factory, and Azure Databricks for enhanced data processing and visualization.

Azure Synapse Use Cases:

1. **Data Warehousing:** Store, manage, and analyze large datasets with high-performance query capabilities.

2. **Real-Time Analytics:** Process and analyze real-time streaming data for quick decision-making.
3. **Big Data Processing:** Work with massive amounts of unstructured and structured data using Spark and other big data tools.
4. **Business Intelligence:** Integrate with Power BI for advanced reporting, dashboards, and visual analytics.
5. **Advanced Analytics and AI:** Run predictive analytics and machine learning models directly within the platform.
6. **ETL/ELT Pipelines:** Build data pipelines using Azure Data Factory to move, transform, and load data for further analysis.
7. **Data Lakes:** Store raw, unstructured data in Azure Data Lake and process it using Synapse's integrated tools.
8. **Cost-Effective Storage:** Archive historical data with minimal cost by utilizing the platform's tiered storage model.