

**A PROJECT REPORT
ON
NEXUS DATALENS – AN END-TO-END DATA ENGINEERING
PROJECT**

Submitted in partial fulfilment of the
Requirements for the award of the Degree of

**MASTER OF SCIENCE (INFORMATION TECHNOLOGY) 2024-25
PREPARED BY
RUKSANA SHAIKH (6716)**

**UNDER THE ESTEEMED GUIDANCE OF
PROF. JEBA ROSELET**



**DEPARTMENT OF INFORMATION TECHNOLOGY
RAMNIRANJAN JHUNJHUNWALA COLLEGE OF ARTS, SCIENCE &
COMMERCE (EMPOWERED AUTONOMOUS), GHATKOPAR (W) 400
086
MUMBAI, MAHARASHTRA**

(Affiliated to University of Mumbai)
Hindi Vidya Prachar Samiti's
Ramniranjan Jhunjhunwala College of Arts, Science and Commerce
(Empowered Autonomous)
Ghatkopar (West), Mumbai – 86

**Hindi Vidya Prachar Samiti's
Ramniranjan Jhunjhunwala College of Arts, Science and Commerce
(Empowered Autonomous)
Ghatkopar (West), Mumbai 400 086**



CERTIFICATE

*This is to certify that the project entitled “**Nexus DataLens**” is bonafide work of **Ruksana Shaikh (6716)** submitted in partial fulfillment of the requirements for the award of the degree of Master of Science in Information Technology from University of Mumbai during the academic year 2024-2025*

Internal Guide

External Examiner

Coordinator

CollegeSeal



R. J. COLLEGE of Arts, Science & Commerce (Empowered Autonomous)

(Hindi Vidya Prachar Samiti's RAMNIRANJAN JHUNJHUNWALA COLLEGE of Arts, Science & Commerce)

Opposite Ghatkopar Railway Station, Ghatkopar (West), Mumbai - 400086, Maharashtra, INDIA.

Website: www.rjcollege.edu.in • Email : rjcollege@rjcollege.edu.in • Tel No. + 91 22 25151763

College is recognized under Section 2(f) & 12(B) of the UGC Act, 1956

Affiliated to UNIVERSITY OF MUMBAI II NAAC Re-Accredited 'A' Grade (CGPA: 3.10)

Ref: 1322

04th December 2024

To,
Mr. Gaurav Bhoi and Ms. Ruksana Shaikh
Department of Information Technology
Ramniranjan Jhunjhunwala College
Ghatkopar (W),
Mumbai 400 086

Re: Approval of funding for your Master's Research Project

Dear Mr. Bhoi & Ms. Shaikh,

We are pleased to inform you that the Research Advisory Committee has approved your proposal for a Master's Research Project titled "**Nexus DataLens (An end-to-end data analytics project using AWS)**" under the guidance of Ms. Jeba Roselet. Consequently, a seed money of not more than Rs. 24,850/- has been granted which you have to spend before 15th March 2025 as much as possible.

Under this project, you are entitled to travel in public transport (second class) within the city. Also, there is an upper limit of Rs. 500/- per day to be spent on lodging and boarding. The grant money cannot be used to meet the publication cost of the research papers.

It is necessary to strictly adhere to the guidelines for the release/ reimbursements of grants attached along with this letter.

Wishing you the best for your research work.

Sincerely Yours,


Principal
PRINCIPAL
RAMNIRANJAN JHUNJHUNWALA COLLEGE
OF ARTS, SCIENCE & COMMERCE (AUTONOMOUS)
Ghatkopar (W), Mumbai - 400 086, Maharashtra, INDIA

2019: Star College Status by DBT

2008: Best College by University of Mumbai 2010: IMC RBNQ Award 'Performance Excellence' for the year 2009

2011: 'Best Teacher Award' by Government of Maharashtra 2013: DST-FIST 2014: DBT STAR College

2013 & 2014: 'Jagar Jaanivancha Award' by Govt. of Maharashtra 2016: ISO 14001:2015 2016: ISO 9001:2015 2017: ISO 27001:2013

2018: Autonomous Status by University Grants Commission (No. F. 22-1/2018(AC) - 28.05.2018) & by University of Mumbai (No.Aff./ICD/18-19/440 - 08.06.2018)

ABSTRACT

In the modern era of data-driven decision-making, organizations require efficient data pipelines to ingest, transform, and analyze data at scale. Traditional data processing methods often involve manual intervention, inefficiencies, and scalability challenges, making automated solutions essential for ensuring accuracy, speed, and reliability.

This project, Nexus DataLens, is an end-to-end data analysis and transformation platform that leverages Microsoft Azure services to automate data workflows. The system allows users to upload structured datasets (CSV files) through a web-based frontend, which are then processed via an automated Azure-based data pipeline. Key services include Azure Data Factory (ADF) for data ingestion, Azure Data Lake for storage, and Azure Databricks for data transformation using Apache Spark. The transformed data is stored back in Azure Data Lake, allowing users to securely download the processed dataset through the frontend.

The primary objective of this project is to enhance data processing efficiency, scalability, and security while minimizing manual effort. The system enables seamless automation of ETL (Extract, Transform, Load) processes, ensuring secure, scalable, and optimized data management. The cloud-based architecture supports large datasets while maintaining data security, compliance, and accessibility.

Nexus Data Lens has applications in business intelligence, reporting, data warehousing, and enterprise-level data integration. It can be integrated with SQL databases, NoSQL stores, and AI/ML pipelines to facilitate advanced analytics and decision-making. The solution provides a robust, scalable, and automated approach to handling business-critical data, making it a powerful tool for organizations seeking efficiency in data processing and management.

ACKNOWLEDGEMENT

The success and outcome of this project were possible by the guidance and support from many people. I am incredibly privileged to have got this all along with the achievement of my project. It required a lot of effort from everyone involved in this project with me and I would like to thank them.

I express my sincere gratitude to **Mrs. Bharthi Bhole**, Head of the Department, for providing me with the opportunity to undertake this project. Her unwavering support, leadership, and encouragement have been instrumental in the successful completion of this work. The valuable resources and guidance provided by her played a crucial role in shaping the direction of this project.

I extend my heartfelt appreciation and deep respect to my esteemed professor, **Prof. Jeba Roselet**, whose keen interest and meticulous attention to detail ensured the precision and quality of this project. Her invaluable insights, constant guidance, and constructive feedback throughout every phase of the project have been pivotal in refining the outcome. Her dedication and mentorship have been a source of immense motivation and learning.

I am truly grateful for the enriching experience and knowledge gained through this project, which would not have been possible without the support and encouragement of my mentors.

Then I would like to thank my parents and friends who have helped me with their valuable suggestions and guidance has been helpful in various phases of the completion of the project.

Gaurav Ramesh Bhoi
Ruksana Shaikh

Date: _____

Index

Sr. No.	Title	Page No
1	INTRODUCTION	7
2	LITERATURE SURVEY	8
3	USE CASE SUMMARY	11
4	PROJECT OVERVIEW	12
5	CLOUD SERVICES	15
6	ARCHITECTURE	22
7	SYSTEM DESIGN	25
8	IMPLEMENTATION <ul style="list-style-type: none"> • Setup ssms • Create user in ssms • Resource group • Data injection • Transformation • Synapse • Power BI 	28
9	PIPELINE TESTING	157
10	FUTURE SCOPE	163
11	CONCLUSION	167
12	REFERENCES	168

INTRODUCTION

Data is one of the most valuable assets for businesses and organizations, serving as the foundation for informed decision-making, strategic planning, forecasting, and trend analysis. However, raw data is often unstructured, inconsistent, and requires extensive transformation before it can deliver meaningful insights. Manual data processing methods introduce challenges such as errors, inefficiencies, delays, and scalability limitations, making them unsuitable for modern, data-driven enterprises.

To address these challenges, this project presents Nexus DataLens—a fully automated, scalable, and cloud-based data pipeline that enables users to seamlessly upload, transform, and retrieve structured datasets. The system provides a web-based frontend where users can upload CSV files, which are then processed through an Azure-based data workflow. Key components include:

- Azure Data Factory (ADF) for automated data ingestion
- Azure Data Lake for secure and scalable storage
- Azure Databricks for powerful data transformation and ETL (Extract, Transform, Load) operations

Once the transformation is complete, the processed data is stored back in Azure Data Lake, allowing users to download the refined dataset from the frontend. The solution ensures data security, accuracy, and efficiency, while eliminating the need for manual intervention.

By leveraging cloud computing and big data technologies, this project aims to streamline data workflows, enhance scalability and performance, and provide organizations with a robust tool for managing large datasets. With applications in business intelligence, analytics, and data warehousing, Nexus Data Lens serves as a cutting-edge solution for enterprises looking to optimize their data transformation processes.

LITERATURE SURVEY

1. Introduction

In today's data-driven landscape, organizations are increasingly adopting cloud-based solutions to address the challenges of traditional Extract, Transform, Load (ETL) processes. Manual intervention, scalability limitations, and inefficiencies in handling large datasets have led to the development of modern, automated data pipelines. This literature survey focuses on existing methodologies, tools, and best practices for incremental ETL pipelines and cloud-based data integration, drawing insights from official Microsoft documentation and resources. It also highlights gaps in current approaches and justifies the adoption of Microsoft Azure services for the proposed solution.

2. Existing Solutions in Data Processing

a. Incremental Data Processing with Azure Data Factory

Incremental data processing has become a cornerstone of efficient data pipelines, enabling organizations to transfer only new or updated records instead of reloading entire datasets. According to Microsoft's official documentation on Azure Data Factory (ADF), incremental loads can be implemented using techniques like Change Data Capture (CDC), watermark columns, or timestamp-based filtering. ADF's ability to orchestrate complex workflows ensures seamless data movement between on-premises systems and cloud storage. Additionally, the integration of Self-Hosted Integration Runtime allows secure connectivity for hybrid environments, making it ideal for scenarios where data resides both on-premises and in the cloud.

b. Scalable Storage with Azure Data Lake Gen2

Azure Data Lake Gen2 is designed to handle massive volumes of structured and unstructured data while maintaining cost efficiency and performance. As outlined in Microsoft's Azure Data Lake Gen2 documentation, its hierarchical file system supports a layered architecture (Bronze, Silver, Gold), ensuring clarity and governance throughout the data pipeline. This approach enables organizations to store raw data, perform transformations, and generate business-ready datasets in a scalable and secure manner. Furthermore, the integration of role-based access control (RBAC) and encryption ensures compliance with industry standards.

c. Advanced Transformations with Azure Databricks

Azure Databricks, powered by Apache Spark, provides a collaborative environment for performing high-performance data transformations. According to Microsoft's Azure Databricks documentation, it supports distributed processing for large datasets, enabling operations like filtering, aggregations, joins, and machine learning model training. The platform's interactive notebooks allow teams to write code in Python, Scala, or SQL, fostering collaboration and flexibility. Additionally, its seamless integration with Azure Data Lake Gen2 ensures efficient data ingestion and output storage.

d. Unified Analytics with Azure Synapse Analytics

Azure Synapse Analytics combines big data processing with enterprise data warehousing, offering a unified platform for analytics. As described in Microsoft's Synapse Analytics documentation , it supports both batch and real-time queries, enabling users to analyze data at scale. Its integration with Power BI facilitates the creation of interactive dashboards and reports, empowering stakeholders to make data-driven decisions. Synapse Analytics also supports advanced analytics use cases, such as predictive modeling and trend analysis, making it a versatile tool for modern data pipelines.

e. Visualization with Power BI

Power BI serves as a powerful visualization tool for transforming processed data into actionable insights. According to Microsoft's Power BI documentation , it enables users to create interactive dashboards and reports that can be shared across teams. Its real-time connectivity to Azure services like Synapse Analytics and Data Lake Gen2 ensures timely and accurate visualizations. Additionally, Power BI's self-service capabilities empower non-technical users to explore data independently, fostering a data-driven culture within organizations.

3. Identification of Gaps

While Microsoft's ecosystem offers robust tools for building data pipelines, certain challenges remain:

- Complexity in Implementation : Configuring and integrating services like Azure Data Factory, Data Lake, and Databricks requires technical expertise, which may pose a barrier for smaller teams.
- Cost Management : While Azure services are scalable, managing costs for large-scale data pipelines can be challenging without proper optimization.
- Security Concerns : Although Azure provides tools like Azure Active Directory and Azure Key Vault for securing sensitive data, ensuring compliance with regulations like GDPR and HIPAA requires careful planning and implementation.
- Real-Time Processing Limitations : While incremental loads reduce latency, achieving true real-time analytics often requires additional configurations and optimizations.

These gaps highlight the need for a well-designed, end-to-end solution that addresses these challenges while leveraging the strengths of Microsoft Azure.

4. Justification of the Proposed Approach

The proposed solution leverages Microsoft Azure's ecosystem to build a comprehensive, scalable, and secure data pipeline. By utilizing the following components, the project ensures efficient data processing and actionable insights:

- Azure Data Factory : Automates incremental data ingestion from an on-premises SQL Server database to Azure Data Lake Gen2, reducing manual effort and ensuring data consistency.
- Azure Data Lake Gen2 : Provides a scalable storage solution with a layered architecture (Bronze, Silver, Gold), ensuring data governance and accessibility.
- Azure Databricks : Performs advanced transformations using Apache Spark, enabling high-performance processing of large datasets.
- Azure Synapse Analytics : Supports both batch and real-time analytics, bridging the gap between data engineering and business intelligence.
- Power BI : Visualizes insights through interactive dashboards, facilitating data-driven decision-making.
- Azure Active Directory and Key Vault : Ensure secure access control and protection of sensitive data, addressing compliance requirements.

This approach aligns with Microsoft's best practices and addresses the limitations of traditional ETL processes by integrating automation, scalability, and security into a unified framework.

5. Conclusion

The literature reviewed in this section highlights the capabilities of Microsoft Azure services in addressing modern data engineering challenges. By leveraging Azure Data Factory, Data Lake Gen2, Databricks, Synapse Analytics, and Power BI, organizations can build efficient, scalable, and secure data pipelines. While certain challenges remain, such as cost management and real-time processing, the proposed solution aims to bridge these gaps by implementing a robust, end-to-end architecture. This ensures that the pipeline meets the demands of today's data-driven organizations while adhering to industry standards and best practices.

USE CASE SUMMARY

Industry-level use cases include real-time inventory management in retail, fraud detection in finance, predictive maintenance in manufacturing, smart grid analytics in energy, and network performance monitoring in telecommunications. These use cases leverage the scalability, automation, and advanced analytics capabilities of Azure to drive business value.

The incremental load process from SSMS to Azure is highly versatile and can be applied across various industries to address specific business challenges. Whether it's real-time inventory management in retail, fraud detection in finance, or predictive maintenance in manufacturing, this architecture supports scalable, secure, and automated data pipelines.

Use Case: Real-Time Movie Inventory Management

- Scenario: A company needs to synchronize its on-premises inventory database with a cloud-based analytics platform to monitor levels in near real-time.
- How It Works:
 - Incremental loads transfer only updated inventory records (e.g., new stock arrivals, sold items) to Azure Data Lake.
 - Azure Databricks processes the data to calculate metrics like low stock alerts, reorder points, and sales trends.
 - Power BI visualizes the data for store managers and supply chain teams.
- Benefits:
 - Reduces overstock and stockouts by providing timely insights.
 - Enables predictive analytics for demand forecasting using AI/ML models.

PROJECT OVERVIEW

Nexus DataLens is an end-to-end data pipeline that uses Azure Data Factory for orchestration, Azure Data Lake Gen2 for storage, Azure Databricks for transformation, and Azure Synapse Analytics and Power BI for analysis and reporting. It supports incremental loads from an on-premises SQL Server database, ensuring scalability, automation, security, and real-time insights. By implementing an incremental load process, you ensure efficient data transfer while maintaining data integrity. The layered architecture (Bronze, Silver, Gold) ensures data quality and governance, while Azure Databricks and Azure Synapse Analytics provide robust transformation and analytics capabilities. Finally, Power BI enables interactive visualization for actionable insights, making this solution highly valuable for data-driven decision-making across industries.

1. Source: On-Premises SQL Server Database

- Purpose: The starting point of the data pipeline is an on-premises SQL Server database, which serves as the source of raw data.
- Data Characteristics: This database likely contains structured data that needs to be processed and analysed.
- Incremental Load: Instead of performing a full data transfer, the solution focuses on incremental loading, where only new or updated records are transferred to the cloud. This approach reduces bandwidth usage and processing time.

2. Orchestration: Azure Data Factory (ADF)

- Role: Azure Data Factory acts as the orchestrator for the entire data pipeline.
- Tasks:
 - Triggering Incremental Loads: ADF schedules or triggers the incremental load process based on predefined intervals (e.g., hourly, daily) or events.
 - Data Movement: It transfers data from the on-premises SQL Server to Azure Data Lake Gen2.
 - Pipeline Monitoring: ADF provides visibility into the status of the pipeline, including success or failure notifications.

3. Storage: Azure Data Lake Gen2

- Purpose: Azure Data Lake Gen2 serves as the primary storage layer for the raw and transformed data.
- Architecture Layers:
 - Bronze Layer: Raw data ingested from the SQL Server database is stored here without any transformations. This ensures that the original data is preserved for future use.
 - Silver Layer: Data is cleaned and standardized in this layer. For example, missing values may be handled, duplicates removed, and data formatted consistently.

- Gold Layer: Business-ready data is stored here after further transformations. This layer contains curated datasets that are optimized for analytics and reporting.

4. Transformation: Azure Databricks

- Role: Azure Databricks is used for data transformation and enrichment.
- Tasks:
 - Processing Bronze Layer Data: Raw data from the Bronze Layer is read and transformed into the Silver Layer.
 - Advanced Transformations: Complex operations like filtering, aggregating, joining datasets, or applying machine learning models can be performed using Apache Spark.
 - Scalability: Azure Databricks offers scalable compute resources to handle large datasets efficiently.

5. Analysis and Reporting: Azure Synapse Analytics and Power BI

- Azure Synapse Analytics:
 - Purpose: Serves as a unified analytics platform for querying and analyzing the transformed data.
 - Tasks:
 - Supports both batch and real-time analytics.
 - Enables SQL queries on the Gold Layer data.
 - Integrates with other Azure services for advanced analytics.
- Power BI:
 - Purpose: Provides interactive dashboards and reports for business users.
 - Tasks:
 - Visualizes insights derived from the Gold Layer data.
 - Allows users to explore data through interactive charts, graphs, and tables.
 - Supports self-service analytics for non-technical stakeholders.

6. Security and Governance

- Azure Active Directory (AAD):
 - Purpose: Manages user identities and access control.
 - Tasks:

- Provides single sign-on (SSO) for secure access to Azure resources.
- Enforces role-based access control (RBAC) to restrict access to sensitive data.
- Azure Key Vault:
 - Purpose: Securely stores encryption keys, secrets, and certificates.
 - Tasks:
 - Encrypts data at rest and in transit.
 - Manages authentication credentials for various services in the pipeline.

7. Incremental Load Process

The core functionality of your project is the incremental load , which ensures that only new or updated data is transferred from the SQL Server database to Azure Data Lake Gen2. This process involves:

1. Identifying Changes: Using techniques such as change data capture (CDC), timestamp columns, or log files to detect new or modified records.
2. Extracting Data: Querying the SQL Server database to retrieve only the changed data.
3. Loading Data: Transferring the extracted data to Azure Data Lake Gen2 using Azure Data Factory.
4. Updating Layers: Updating the Bronze, Silver, and Gold layers incrementally to reflect the latest changes.

8. Benefits of the Architecture

- Scalability: Handles large volumes of data with ease due to Azure's cloud infrastructure.
- Automation: Reduces manual intervention through automated ETL processes.
- Security: Ensures data privacy and compliance with tools like Azure Active Directory and Azure Key Vault.
- Real-Time Insights: Supports near-real-time data processing and analytics.
- Flexibility: Easily integrates with other Azure services for advanced analytics and reporting.

Cloud Services



1. SQL Server Management Studio (SSMS)

SSMS is a comprehensive tool for managing and querying on-premises SQL Server databases. It enables users to execute complex SQL queries, manage database schemas, and track changes using features like Change Data Capture (CDC). SSMS ensures the source database is optimized for incremental data extraction, making it a critical starting point for your ETL pipeline.

Overview:

- Purpose: SSMS is a tool used to manage and interact with SQL Server databases. It serves as the interface for querying, managing, and maintaining the source data stored in the on-premises SQL Server database.

Key Features:

- Database Management:
 - Execute SQL queries to extract, update, or delete data.
 - Manage tables, views, stored procedures, and indexes.
- Data Exploration:
 - Explore and analyze data using graphical query builders or custom SQL scripts.
- Change Data Capture (CDC):
 - Enables tracking of changes (inserts, updates, deletes) in the database, which is critical for implementing incremental loads.
- Backup and Recovery:

- Perform regular backups to prevent data loss.
- Restore databases from backups in case of failure.

Role in Project :

- SSMS is used to manage the on-premises SQL Server database, ensuring that it is properly configured for incremental data extraction.
- It provides the mechanism to identify new or updated records using techniques like CDC or timestamp-based filtering.

2. Azure Data Factory (ADF)

Azure Data Factory automates and orchestrates the movement and transformation of data across various environments. It supports incremental loads by identifying new or updated records using watermark columns or CDC mechanisms. ADF's scheduling and monitoring capabilities ensure seamless data flow from on-premises systems to Azure services.

Overview :

- Purpose : ADF is a cloud-based data integration service that orchestrates and automates data movement and transformation workflows.

Key Features :

- Pipeline Orchestration :
 - Automates the ETL process by scheduling and triggering pipelines.
 - Supports event-based triggers (e.g., when new data arrives) or time-based triggers (e.g., daily or hourly).
- Data Movement :
 - Transfers data between on-premises systems (via the Self-Hosted Integration Runtime) and Azure services.
 - Handles large-scale data transfers efficiently.
- Monitoring :
 - Provides a visual interface to monitor pipeline runs, track errors, and view performance metrics.
- Incremental Load Support :
 - Facilitates incremental data extraction by leveraging watermark columns (e.g., timestamps) or CDC mechanisms.

Role in Project :

- ADF extracts incremental data from the SQL Server database and loads it into Azure Data Lake Gen2.

- It orchestrates the entire ETL workflow, ensuring seamless data flow between services.

3. Azure Data Lake Gen2

Azure Data Lake Gen2 provides scalable, cost-efficient storage for structured and unstructured data. Its hierarchical file system organizes data into Bronze, Silver, and Gold layers, ensuring clarity and governance. With built-in security and integration with analytics tools, it serves as the backbone of your data pipeline.

Overview :

- Purpose : Azure Data Lake Gen2 is a scalable storage solution designed for big data analytics. It combines the capabilities of Azure Blob Storage with a hierarchical file system.

Key Features :

- Scalability :
 - Stores petabytes of structured, semi-structured, and unstructured data.
 - Automatically scales to accommodate growing data volumes.
- Layered Architecture:
 - Bronze Layer: Raw data ingested from the source (SQL Server).
 - Silver Layer: Cleaned and standardized data after initial transformations.
 - Gold Layer: Curated, business-ready data optimized for analytics.
- Cost Efficiency:
 - Offers tiered storage options (hot, cool, archive) to optimize costs based on data access patterns.
- Security:
 - Supports encryption at rest and in transit.
 - Integrates with Azure Active Directory for fine-grained access control.

Role in Project:

- Acts as the primary storage layer for raw, intermediate, and transformed data.
- Ensures data is securely stored and readily accessible for downstream processing.

4. Azure Databricks

Azure Databricks leverages Apache Spark to perform high-performance data transformations and advanced analytics. It processes raw data from the Bronze Layer into enriched datasets in the Silver Layer, supporting operations like filtering, aggregations, and joins. Its collaborative notebooks enable teams to build and refine workflows efficiently.

Overview :

- Purpose : Azure Databricks is a collaborative analytics platform powered by Apache Spark. It is used for high-performance data transformation and advanced analytics.

Key Features :

- Apache Spark :
 - Performs distributed data processing for large datasets.
 - Supports batch and real-time transformations.
- Notebooks :
 - Interactive notebooks (Python, Scala, SQL) allow users to write and execute code collaboratively.
- Machine Learning :
 - Integrates with MLflow for building, training, and deploying machine learning models.
- Scalability :
 - Automatically scales compute resources based on workload demands.
- Integration :
 - Seamlessly integrates with Azure Data Lake Gen2 for reading and writing data.

Role in Project :

- Transforms raw data from the Bronze Layer into cleaned and enriched data in the Silver Layer.
- Performs advanced transformations, such as aggregations, joins, and feature engineering, to prepare data for analysis.
- Can be used to build AI/ML models if needed.

5. Azure Synapse Analytics

Azure Synapse Analytics combines big data processing with enterprise data warehousing for unified analytics. It allows users to run SQL queries on transformed data and supports real-time analytics. By integrating with Power BI, it bridges the gap between data analysis and visualization for actionable insights.

Overview :

- Purpose : Azure Synapse Analytics is a unified analytics service that combines big data processing with enterprise data warehousing.

Key Features :

- SQL Analytics :

- Allows users to run SQL queries on large datasets stored in Azure Data Lake.
- Supports both serverless and provisioned SQL pools.
- Big Data Processing :
 - Integrates with Apache Spark for distributed data processing.
- Integration :
 - Connects seamlessly with Power BI for visualization.
 - Works with Azure Data Lake Gen2 for data storage.
- Scalability :
 - Scales automatically to handle massive workloads.
- Real-Time Analytics :
 - Supports streaming data ingestion and analysis.

Role in Project :

- Analyzes the Gold Layer data to generate insights.
- Supports ad-hoc queries and reporting for business stakeholders.
- Serves as a bridge between the data pipeline and visualization tools like Power BI.

6. Power BI

Power BI transforms processed data into interactive dashboards and reports for business stakeholders. Its self-service capabilities empower non-technical users to explore trends and patterns visually. With real-time connectivity to Azure services, Power BI ensures timely and accurate decision-making through intuitive visualizations.

Overview :

- Purpose : Power BI is a business analytics tool used to create interactive dashboards and reports.

Key Features :

- Visualization :
 - Build charts, graphs, and tables to visualize data trends and patterns.
 - Supports drill-down and cross-filtering for deeper insights.
- Self-Service Analytics :
 - Empowers non-technical users to explore data without requiring extensive coding skills.

- Integration :
 - Connects directly to Azure Synapse Analytics or Azure Data Lake Gen2 for real-time data access.
- Collaboration :
 - Share dashboards and reports with teams via the Power BI Service.
- Customization :
 - Create custom visuals or use pre-built templates.

Role in Project :

- Visualizes insights derived from the Gold Layer data.
- Enables business users to make data-driven decisions through interactive dashboards.

7. Azure Active Directory (AAD)

Azure Active Directory provides secure identity management and access control for Azure resources. It enforces role-based permissions and supports multi-factor authentication to protect sensitive data. AAD ensures compliance with industry standards while enabling seamless single sign-on across services.

Overview:

- Purpose: AAD is Microsoft's identity and access management service, providing secure access to Azure resources.

Key Features:

- Single Sign-On (SSO):
 - Allows users to log in once and access multiple Azure services.
- Role-Based Access Control (RBAC)
 - Assigns permissions based on user roles (e.g., admin, contributor, reader).
- Multi-Factor Authentication (MFA):
 - Adds an extra layer of security by requiring additional verification steps.
- Audit Logs:
 - Tracks user activities and access attempts for compliance purposes.

Role in Your Project:

- Ensures secure access to Azure services like ADF, Data Lake, and Synapse Analytics.
- Prevents unauthorized access to sensitive data.

8. Azure Key Vault

Azure Key Vault securely stores encryption keys, secrets, and certificates to safeguard sensitive information. It integrates with other Azure services to manage credentials and API keys without exposing them. By centralizing secret management, Key Vault enhances the security and compliance of your data pipeline.

8. Azure Key Vault

Overview:

- Purpose: Azure Key Vault is a service for securely storing and managing secrets, encryption keys, and certificates.

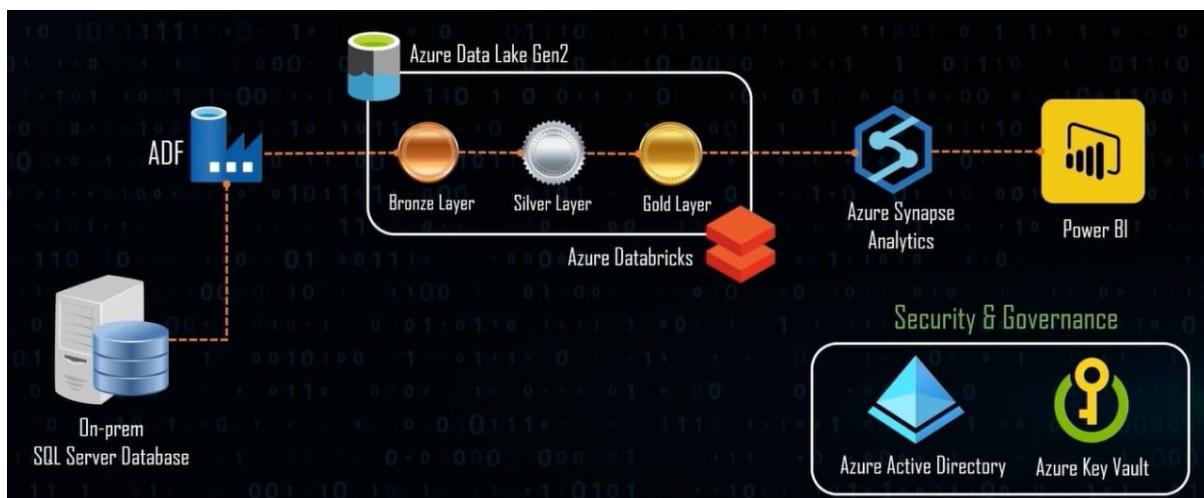
Key Features:

- Secret Management :
 - Stores credentials, API keys, and connection strings securely.
- Encryption Keys :
 - Manages encryption keys for data at rest and in transit.
- Access Control :
 - Restricts access to secrets and keys based on user roles.
- Integration :
 - Works seamlessly with other Azure services like ADF and Data Lake.

Role in Project :

- Protects sensitive information such as database credentials and API keys.
- Ensures compliance with security standards like GDPR or HIPAA.

ARCHITECTURE



1. Data Source: On-Premises SQL Server Database

- The journey begins with an on-premises SQL Server database , which serves as the source of raw data. This database contains structured data that is critical for business operations and analytics.
- To ensure efficiency, the system leverages incremental loading techniques such as Change Data Capture (CDC) or timestamp-based filtering to identify only new or updated records. This minimizes the volume of data transferred and reduces processing overhead.

2. Data Ingestion via Azure Data Factory (ADF)

- Azure Data Factory (ADF) acts as the orchestrator of the entire data pipeline. It automates the extraction of incremental data from the SQL Server database and ensures secure transfer to Azure Data Lake Gen2 .
- ADF supports both scheduled and event-driven triggers, enabling real-time or near-real-time data ingestion. Additionally, it uses the Self-Hosted Integration Runtime to facilitate seamless connectivity between on-premises systems and the cloud.
- During this phase, ADF ensures data integrity by logging pipeline runs, monitoring errors, and providing visibility into the status of data movement.

3. Storage and Layered Architecture in Azure Data Lake Gen2

- Once the data reaches Azure Data Lake Gen2 , it is stored in a layered architecture to maintain clarity, governance, and scalability:
 - **Bronze Layer :** This layer stores raw, unprocessed data ingested directly from the SQL Server database. It serves as a "single source of truth" for all incoming data.
 - **Silver Layer :** Raw data from the Bronze Layer is cleaned, standardized, and enriched using tools like Azure Databricks . Operations such as removing

duplicates, handling missing values, and formatting inconsistencies are performed here.

- Gold Layer : The Silver Layer data is further transformed into curated, business-ready datasets optimized for analytics and reporting. This layer contains high-quality data tailored for specific use cases.
- Azure Data Lake Gen2's hierarchical file system and scalability ensure efficient storage and retrieval of large datasets.

4. Transformation with Azure Databricks

- Azure Databricks , powered by Apache Spark , performs advanced data transformations on the Bronze and Silver Layer data. It processes large volumes of data in parallel, enabling high-performance operations such as filtering, aggregations, joins, and feature engineering.
- Azure Databricks' collaborative notebooks allow data engineers and analysts to write code in languages like Python, Scala, or SQL, making it flexible for various transformation needs.
- The transformed data is then written back to Azure Data Lake Gen2, specifically into the Gold Layer, ready for downstream analytics.

5. Analysis with Azure Synapse Analytics

- The Gold Layer data is analyzed using Azure Synapse Analytics , a unified analytics service that combines big data processing with enterprise data warehousing.
- Synapse Analytics supports both batch and real-time analytics , enabling users to run complex SQL queries or perform exploratory data analysis on large datasets.
- It integrates seamlessly with Power BI , allowing for smooth transitions between data analysis and visualization.

6. Visualization with Power BI

- Insights derived from Azure Synapse Analytics are visualized using Power BI , a powerful business intelligence tool.
- Power BI creates interactive dashboards and reports, enabling business users to explore trends, patterns, and key metrics through charts, graphs, and tables.
- Its self-service capabilities empower non-technical stakeholders to make data-driven decisions without requiring extensive technical expertise.

7. Security and Governance

- Throughout the pipeline, security and governance are ensured using two key services:
 - Azure Active Directory (AAD) : Manages user identities and enforces role-based access control (RBAC). It ensures that only authorized users can access sensitive data and resources.

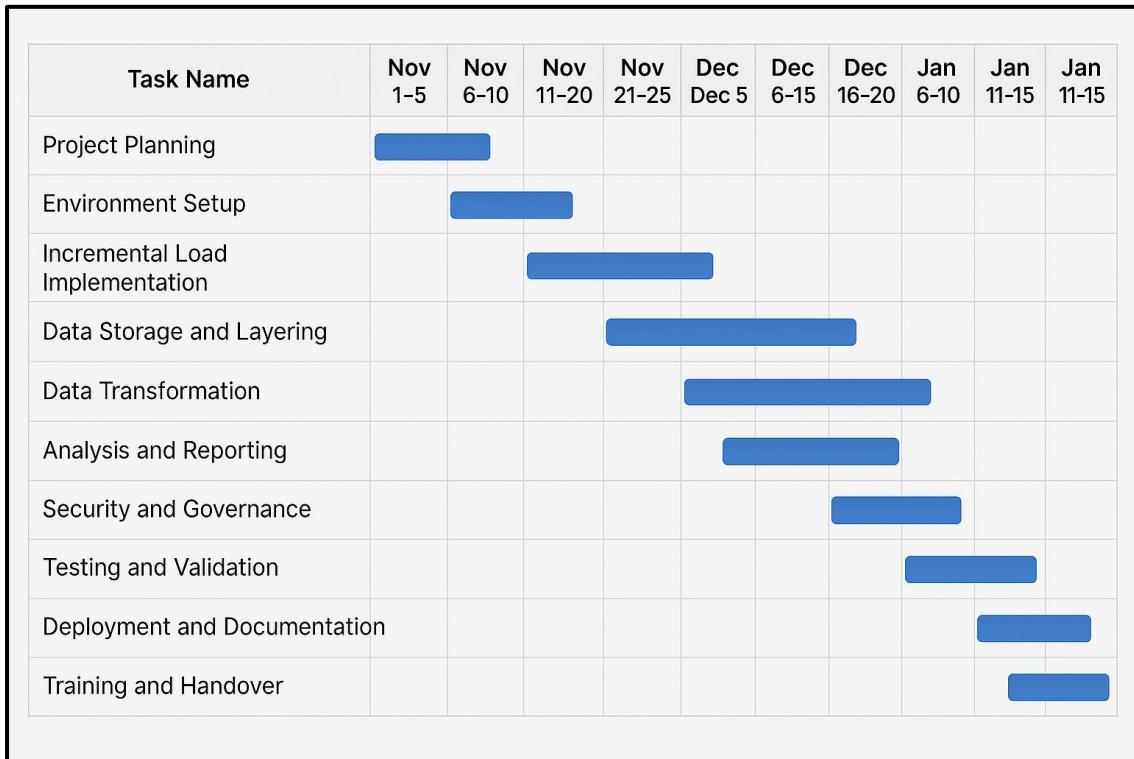
- Azure Key Vault : Securely stores encryption keys, secrets, and certificates, protecting sensitive information such as database credentials and API keys. This ensures compliance with industry standards like GDPR, HIPAA, or SOC 2.

8. End-to-End Workflow Summary

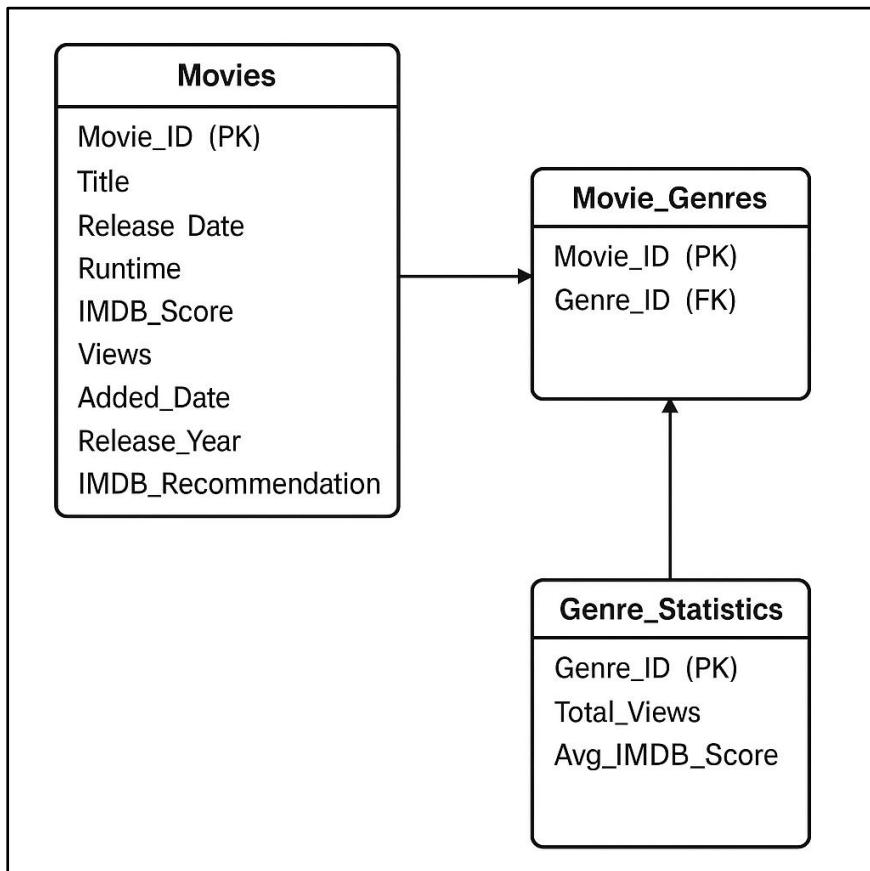
- The entire workflow starts with data extraction from the on-premises SQL Server database using Azure Data Factory.
- Incrementally loaded data is stored in Azure Data Lake Gen2, where it progresses through the Bronze, Silver, and Gold layers.
- Azure Databricks transforms the data, while Azure Synapse Analytics analyzes it for deeper insights.
- Finally, Power BI visualizes the results, enabling stakeholders to make informed decisions.
- Security and governance are maintained throughout the pipeline using Azure Active Directory and Azure Key Vault.

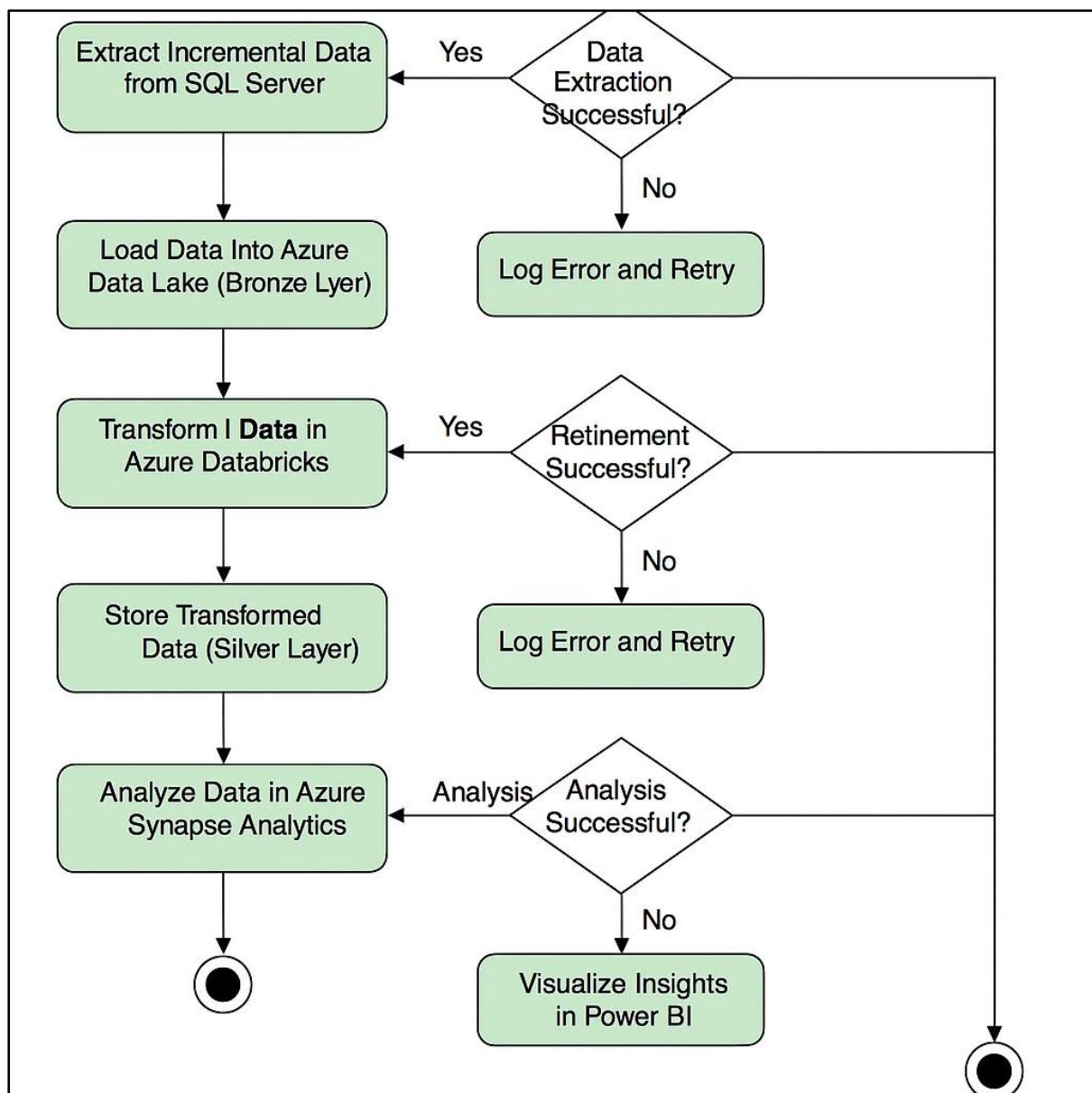
SYSTEM DESIGN

Gantt Chart

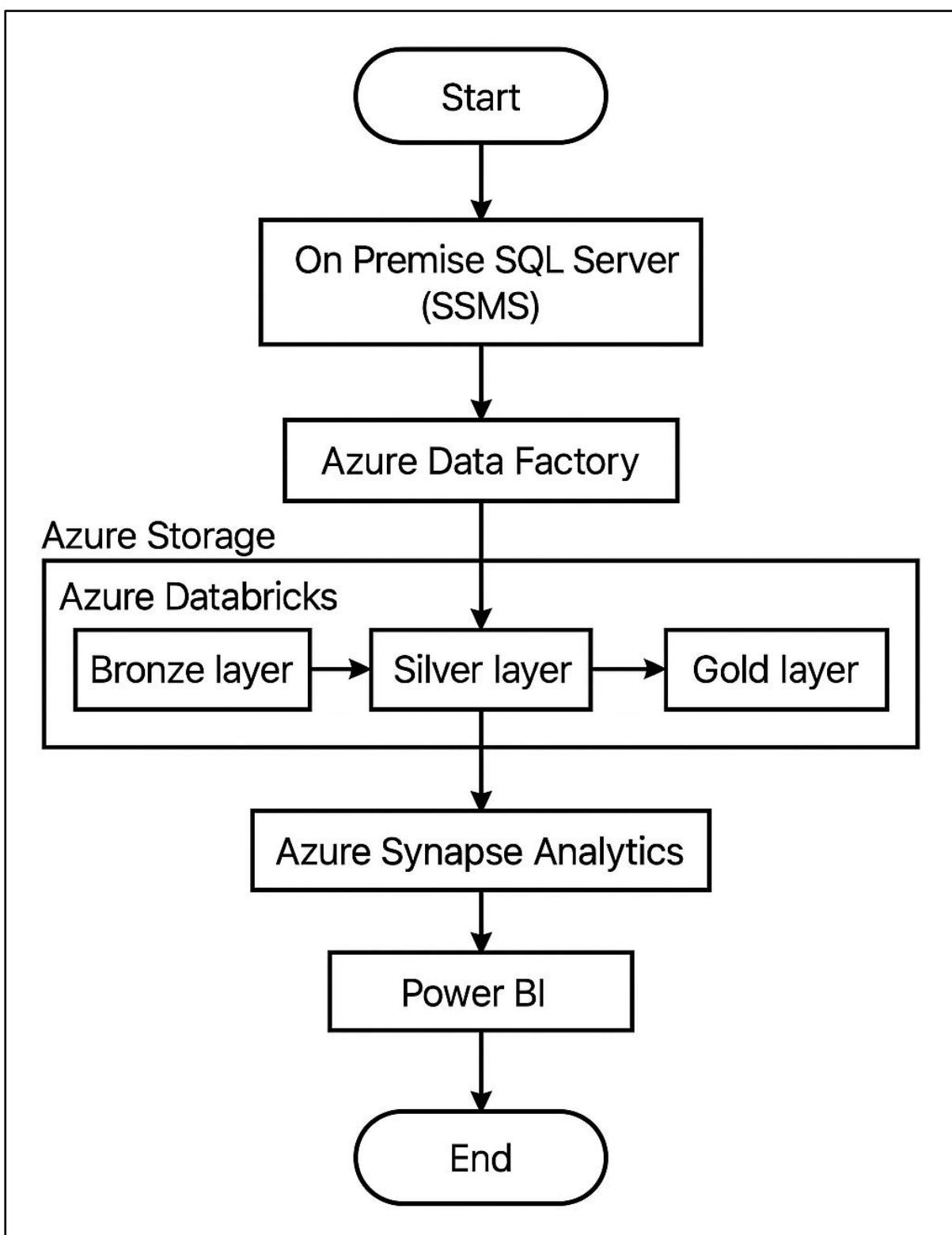


ER Diagram



UML Diagram

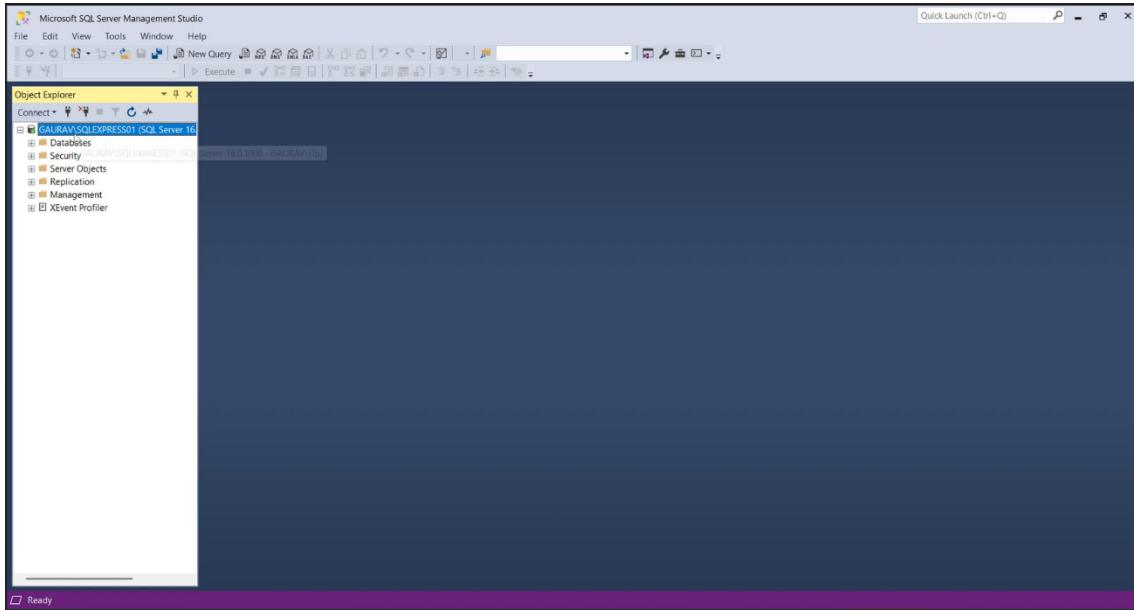
Activity Diagram



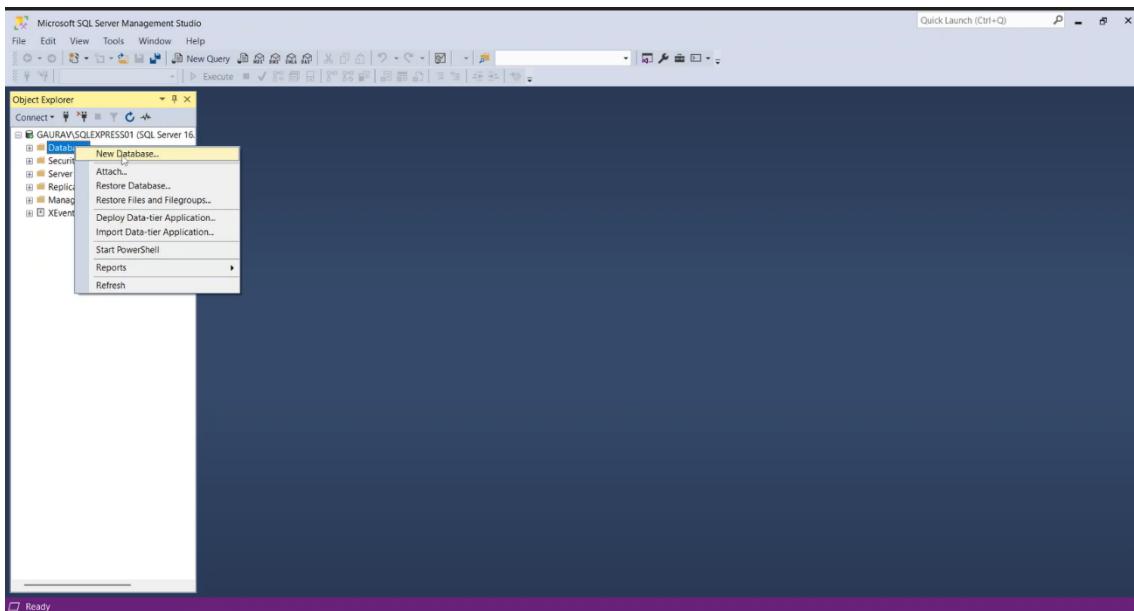
IMPLEMENTATION

Setting Up SQL server Management Server

Microsoft SQL Server Management Studio (SSMS) connected to the local SQL Server instance "GAURAV\SQLEXPRESS01". The Object Explorer displays key components like Databases, Security, Server Objects, and Management.

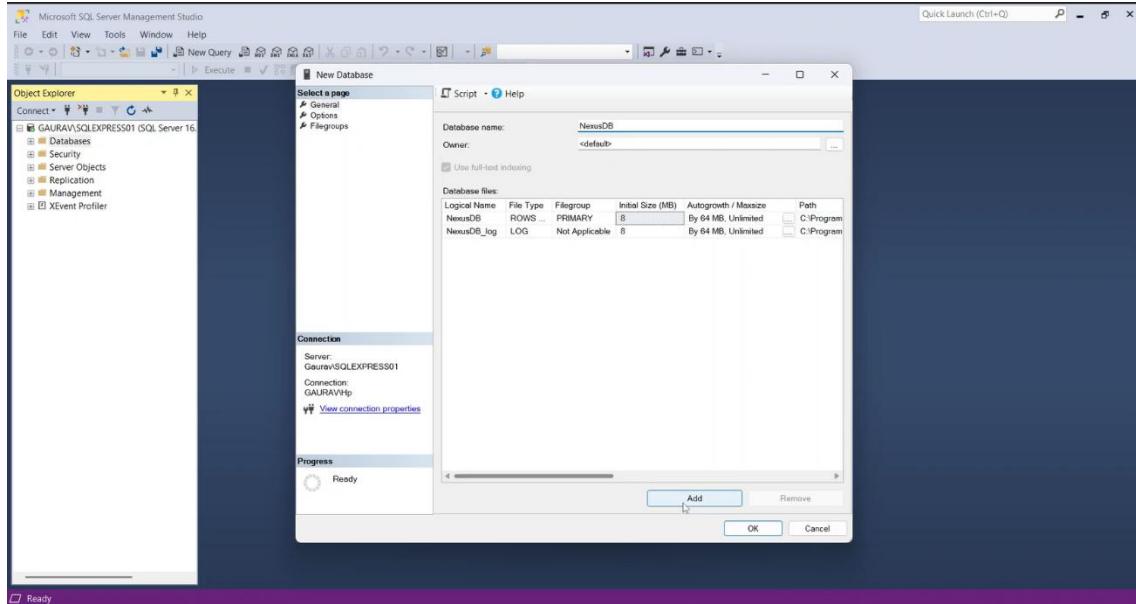


In SQL Server Management Studio, the context menu shows the option to create a new database on the connected server "GUARAV\SQLEXPRESS01".

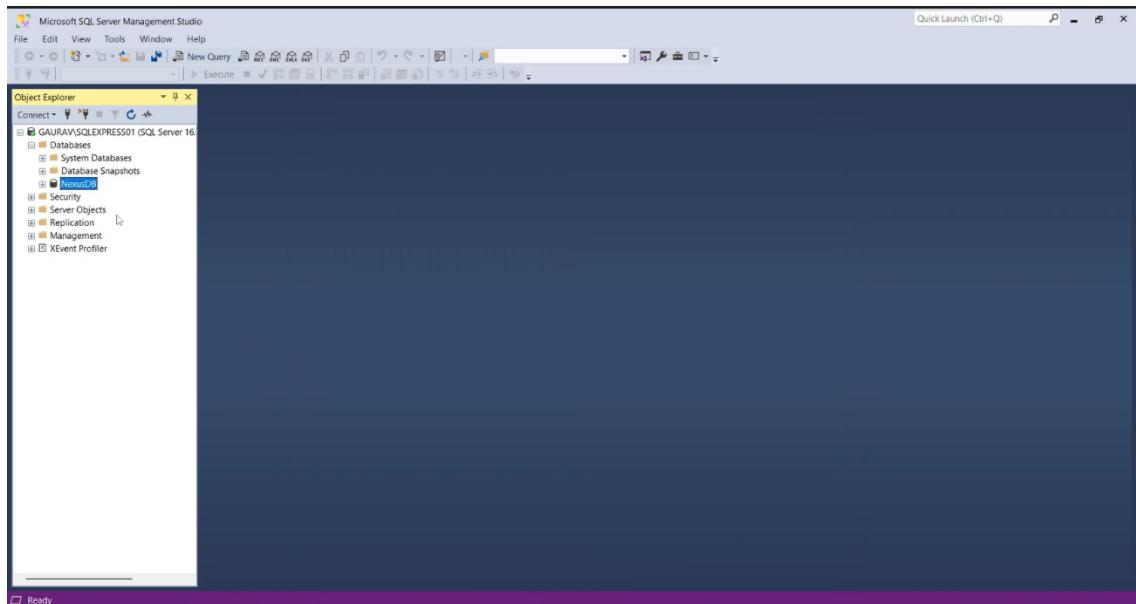


Nexus DataLens

A new database named "NexusDB" is being created in SQL Server Management Studio, with default data and log file settings configured.

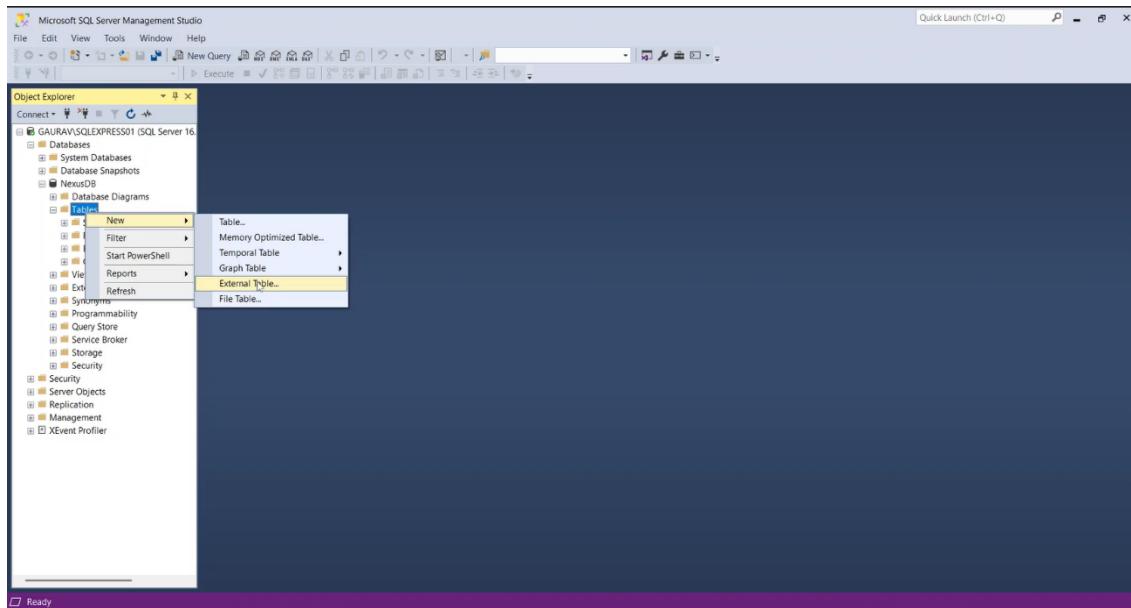


The newly created database "NexusDB" is now visible under the Databases section in SQL Server Management Studio.



Nexus DataLens

A right-click on the "Tables" folder in the NexusDB database shows the context menu to create a new external table.

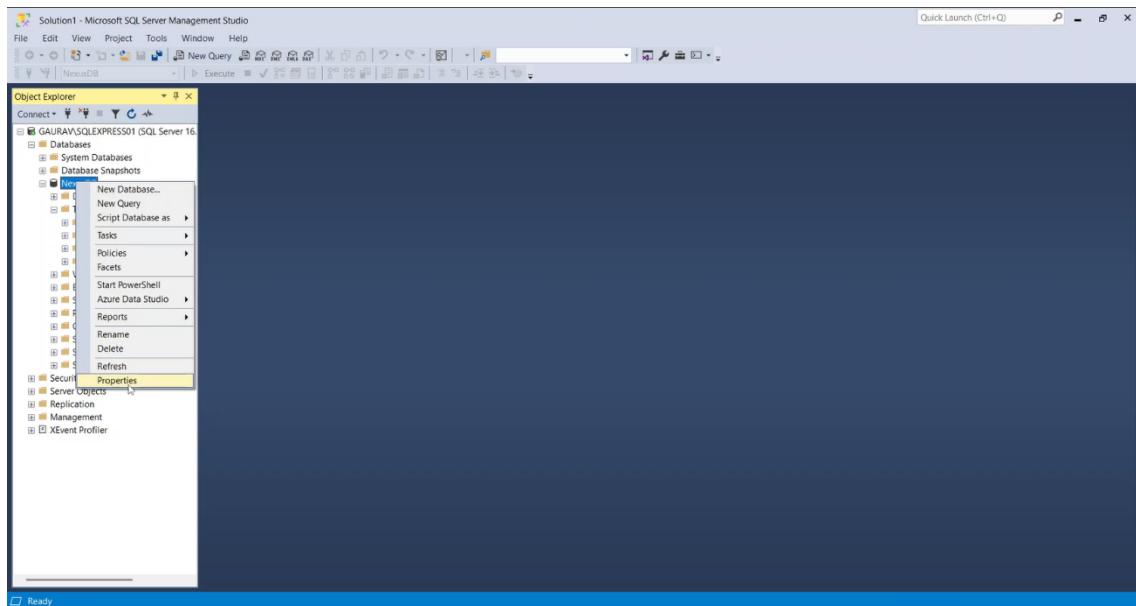


The SQL script shown is a template for creating an external table in SQL Server, including optional steps to drop the table if it exists.

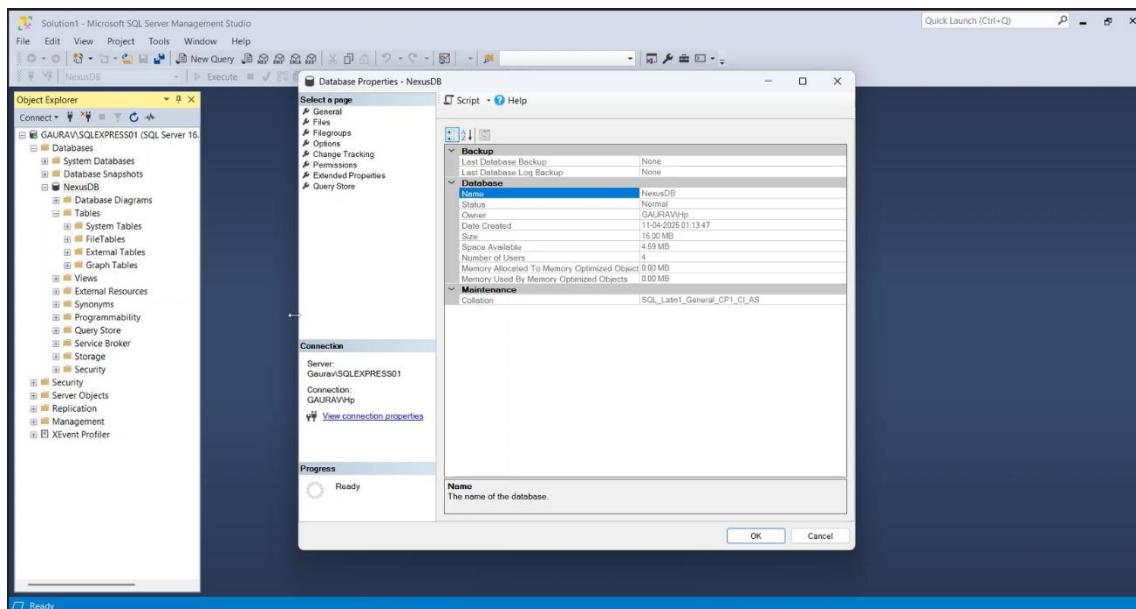
A screenshot of Microsoft SQL Server Management Studio (SSMS) showing the Object Explorer on the left and a query editor window on the right. The query editor contains a SQL script template for creating an external table. The script starts with a comment block `-- Create External Table Template` followed by the `USE` statement, `IF` statement to drop the table if it exists, and the `CREATE EXTERNAL TABLE` statement with its parameters. The script ends with a `GO` statement at the end of the `CREATE` block. The status bar at the bottom of the screen shows 'Connected: (1/1)' and other system information.

Nexus DataLens

The user is right-clicking on the server node in SQL Server Management Studio to access the Properties option from the context menu.

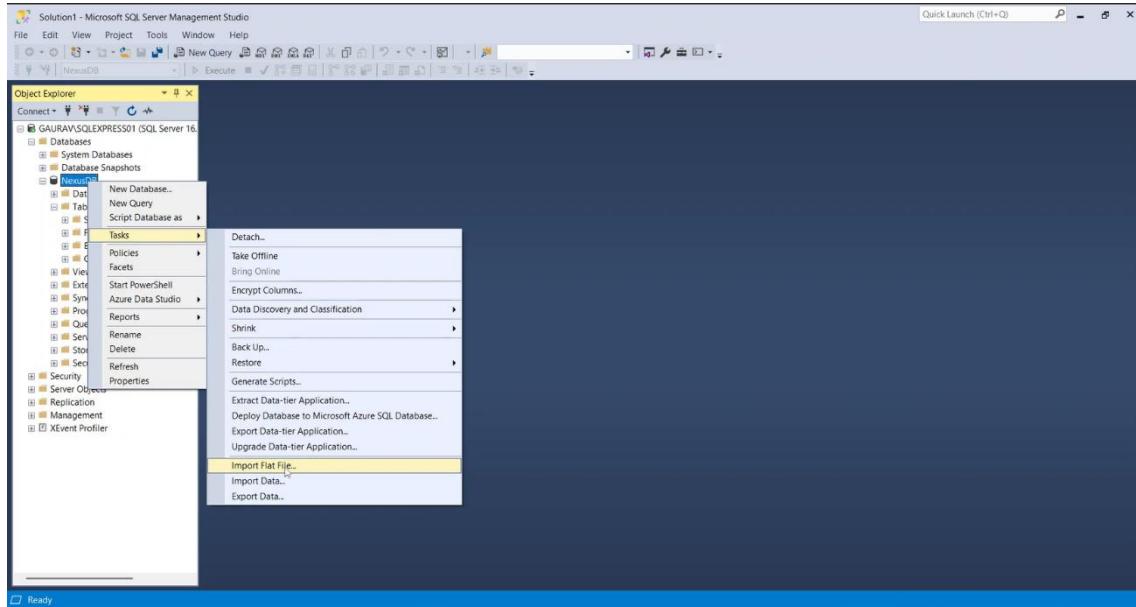


The Database Properties window for *NexusDB* is open in SQL Server Management Studio, showing detailed information like size, owner, creation date, and collation settings.

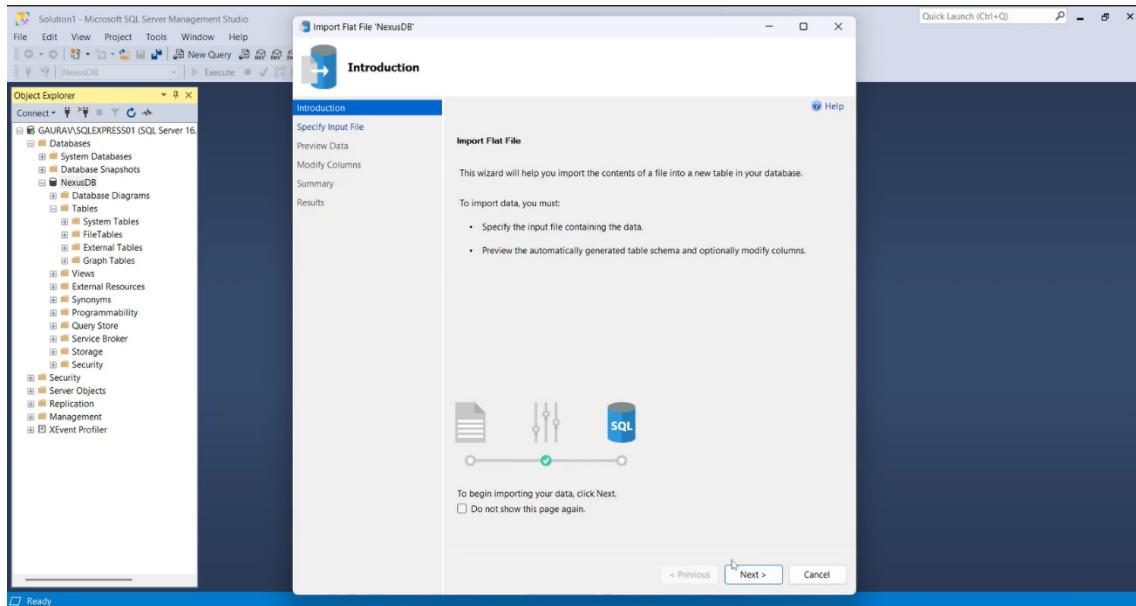


Nexus DataLens

The Tasks menu for the *NexusDB* database is open in SQL Server Management Studio, with the Import Flat File option highlighted to import data from a flat file.

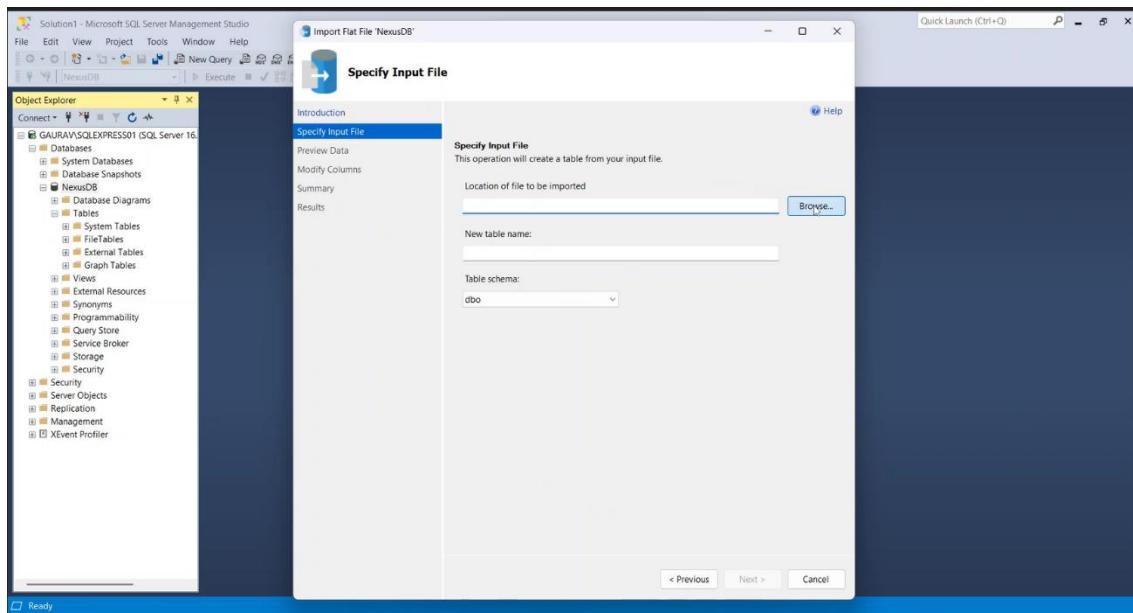


The Import Flat File wizard has opened in SQL Server Management Studio to guide the user through importing a file into the NexusDB database.

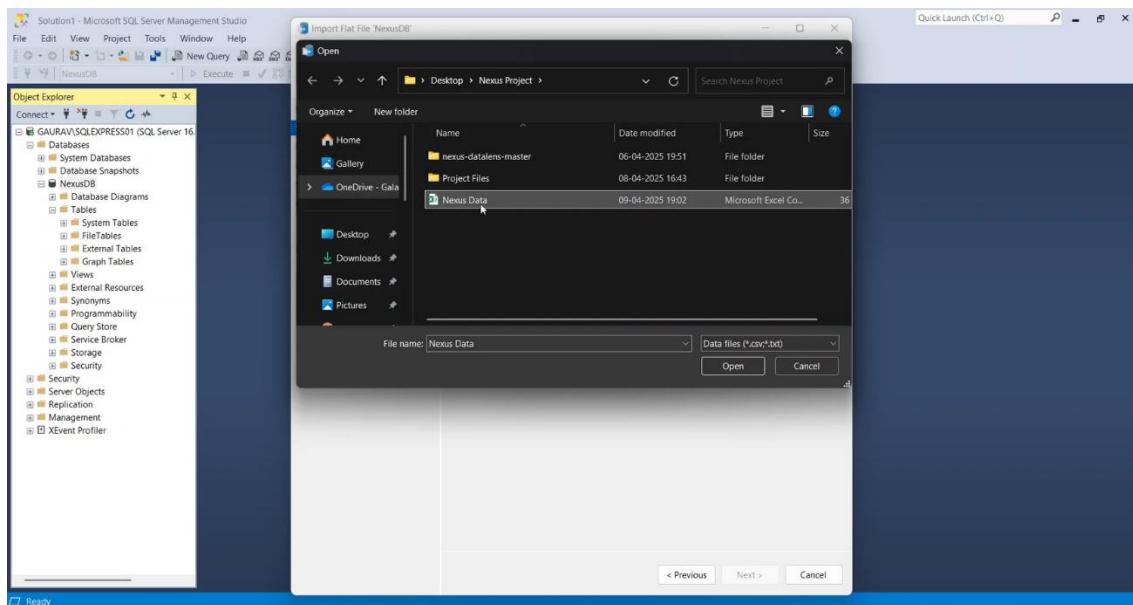


Nexus DataLens

The user to browse and select a flat file to import, specify a new table name, and choose the table schema for the NexusDB database.

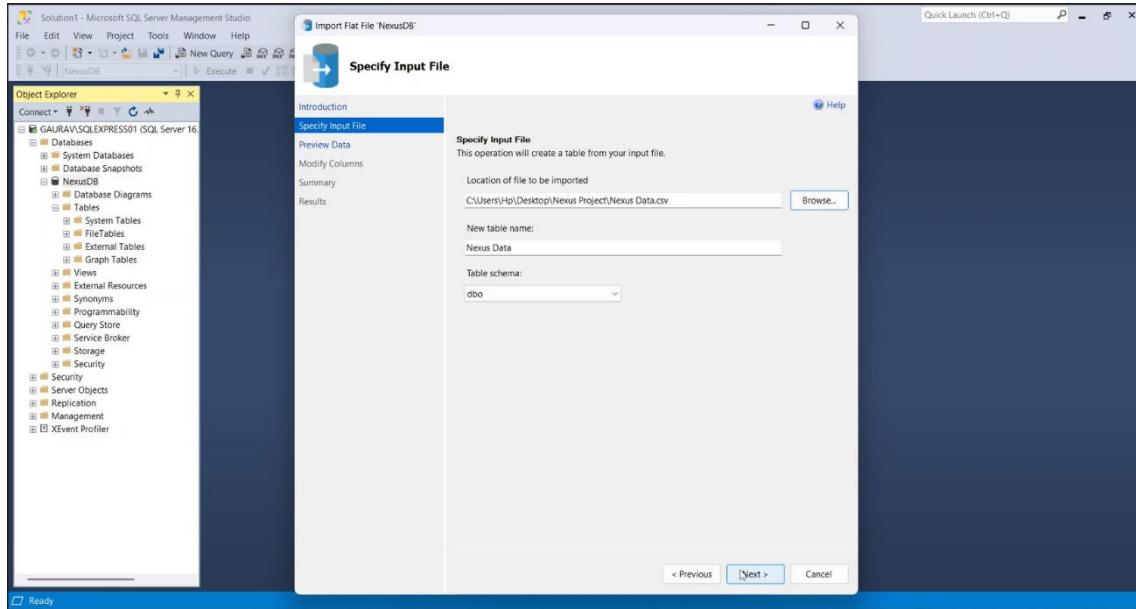


This window shows the user selecting an Excel file named "Nexus Data" to import into the SQL Server database.

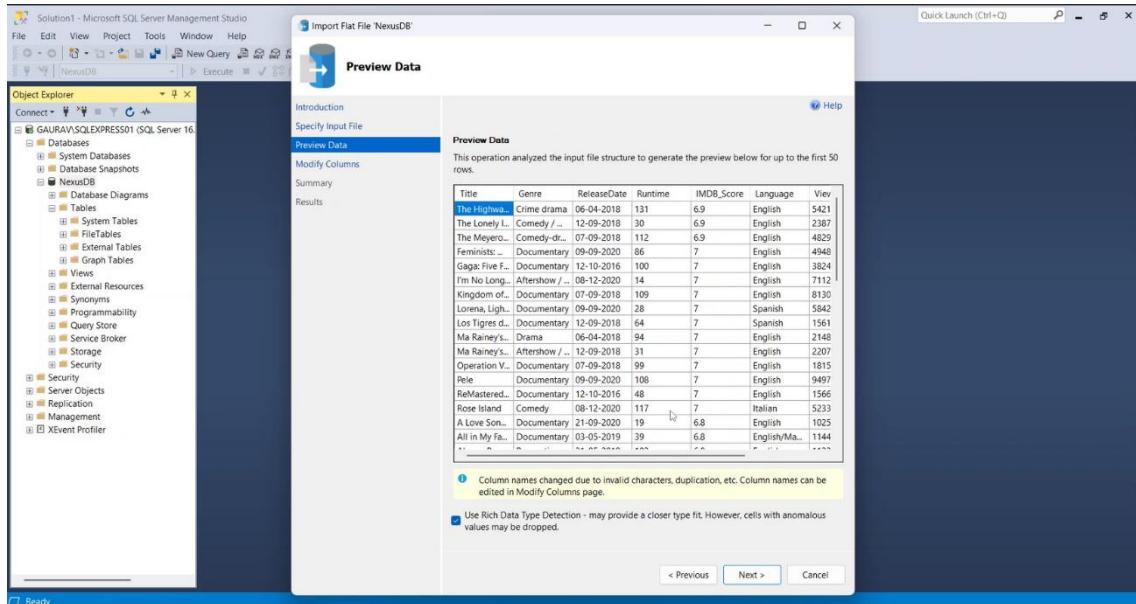


Nexus DataLens

The user has selected the CSV file "Nexus Data.csv" for import and is ready to proceed to the next step to create a new table named "Nexus Data" in the dbo schema.

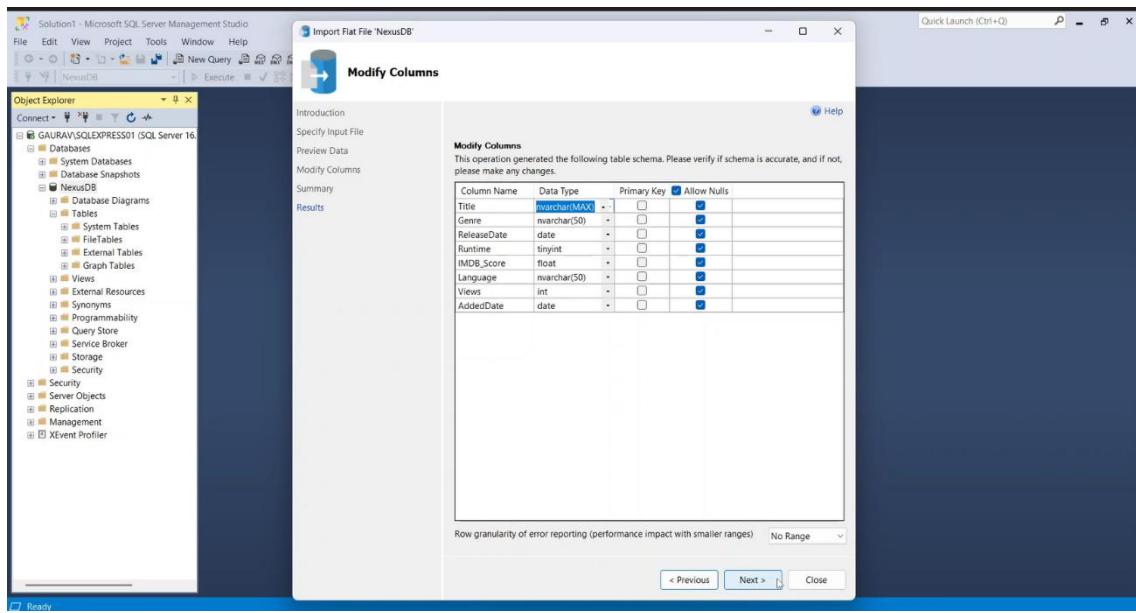


The import wizard displays a preview of the CSV data, showing column names and sample rows while notifying that some column names were adjusted for compatibility.

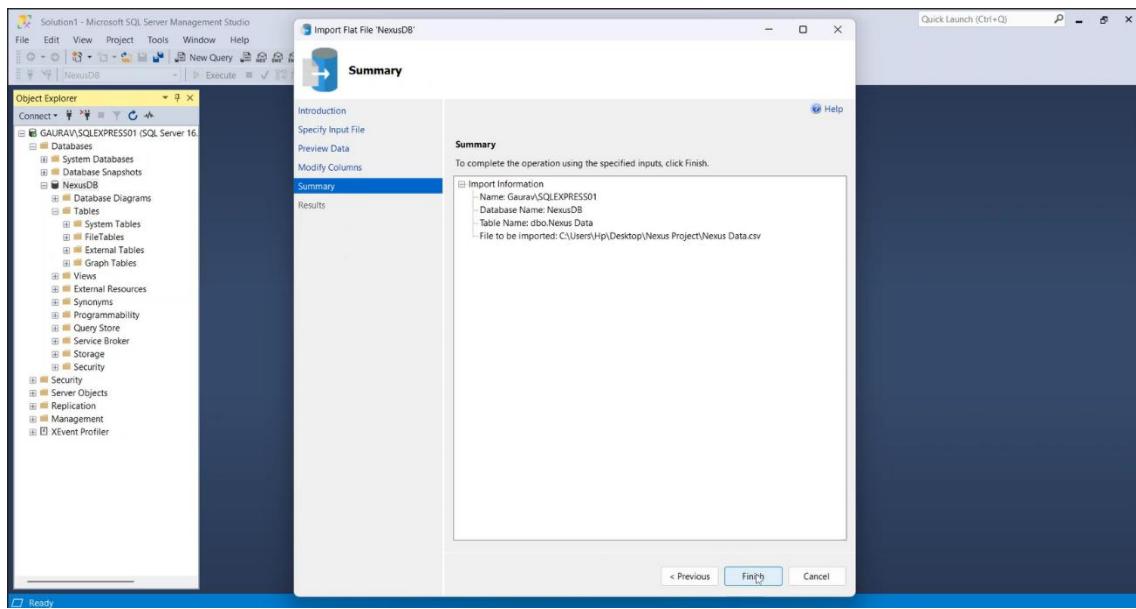


Nexus DataLens

This step lets you review and modify column names, data types, nullability, and primary key settings before importing the data into the SQL table.

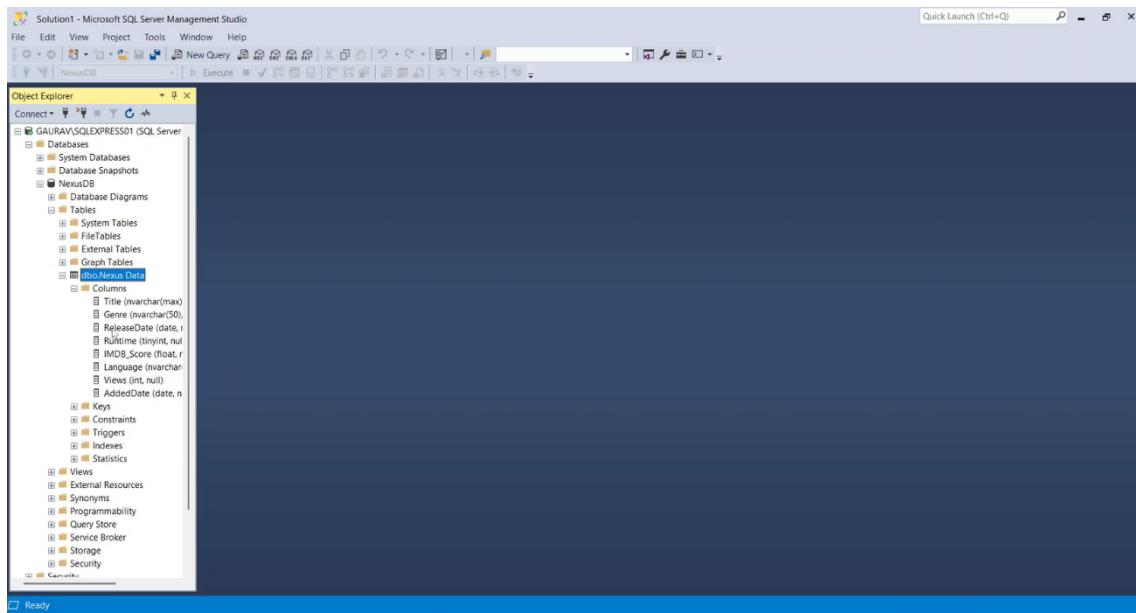


This summary screen confirms all import settings before finalizing the process to load the CSV file into the specified SQL Server table.



Nexus DataLens

The table Nexus Data with its columns has been successfully created and is now listed under Tables in the NexusDB database.



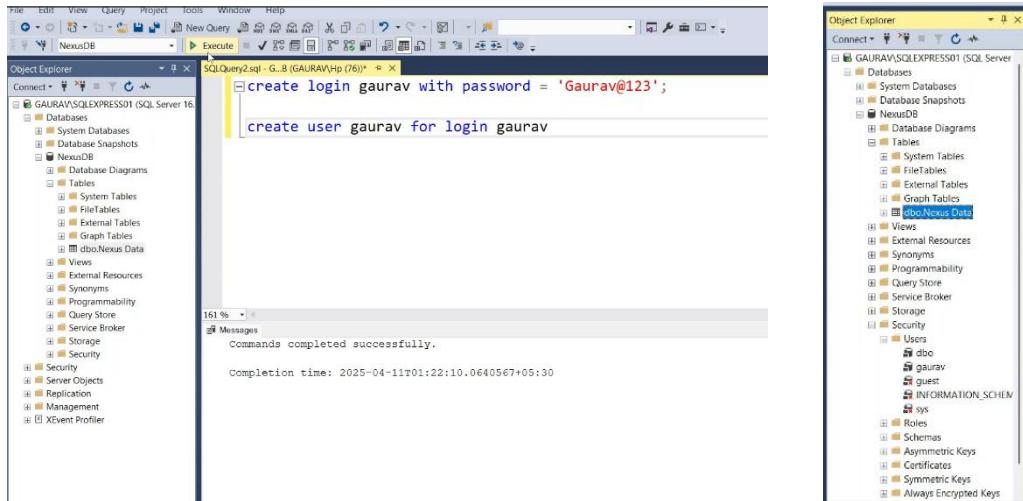
Create User in SSMS

Open SQL Server Management Studio (SSMS) connected to a server named GAURAV\SQLEXPRESS01.

```
create login gaurav with password = 'Gaurav@123';
```

Create a Database User for the Login

```
create user Gaurav for login gaurav
```



SETUP KEYVAULT

“**NexusProject**” resource group in Microsoft Azure, listing its associated resources like Databricks, Data Factory, Storage Account, Key Vault, and Synapse Workspace. All resources are deployed in the Central India region under the Azure for Students subscription.

The screenshot shows the Azure portal interface for the "NexusProject" resource group. The left sidebar lists other resource groups: "databricks-rg-nexusdatabricks-bn", "NetworkWatcherRG", "NexusProject" (selected), and "synapsesworkspace-manageddg-d5". The main content area is titled "Overview" and shows the following details:

- Subscription (move):** Azure for Students
- Subscription ID:** cb194a87-26b0-47d7-827d-3471677a4a02
- Deployments:** 5 Succeeded
- Location:** Central India

The "Resources" section lists the following resources:

Name	Type	Location	Actions
nexusdatabricks	Azure Databricks Service	Central India	...
nexusdatabricksADF	Data factory (V2)	Central India	...
nexusdatamstorage	Storage account	Central India	...
nexuskeyvault	Key vault	Central India	...
nexussynapse	Synapse workspace	Central India	...

At the bottom, there are navigation links for "Previous", "Page 1 of 1", and "Next >".

The image shows the Azure Key Vault "nexuskeyvault" overview page under the "**NexusProject**" resource group, displaying its configuration and security management options.

The screenshot shows the Azure portal interface for the "nexuskeyvault" key vault. The left sidebar lists other objects: "Keys", "Secrets", "Certificates", "Monitoring", "Automation", and "Help". The main content area is titled "Overview" and shows the following details:

- Resource group (move):** NexusProject
- Location:** Central India
- Subscription (move):** Azure for Students
- Subscription ID:** cb194a87-26b0-47d7-827d-3471677a4a02

The "Manage keys and secrets used by apps and services" section contains three buttons:

- Control access to key vault**: Assign access policy and determine whether a given service principal, namely an application or user group, can perform different operations on key vault keys, secrets or certificates.
- Enable logging and set up alerts**: Enable logging to monitor how, when and by whom your key vault is accessed, monitor performance and configure alerts for key vault metrics e.g. service API latency, error code, throttling.
- Turn on recovery options**: For protection against accidental or malicious deletion, soft delete is enabled. Turn on purge protection to guard against manual purging of deleted key vaults and hence prevent loss.

At the bottom, there are links for "Access configuration", "View", and "Explore".

Nexus DataLens

The image shows the "Secrets" section of the Azure Key Vault "nexuskeyvault," listing two enabled secrets named "password" and "username."

A screenshot of the Microsoft Azure portal showing the "Secrets" section of the "nexuskeyvault" key vault. The page displays a table with two rows of secrets. The first row has "password" as the name, "String" as the type, and "Enabled" as the status. The second row has "username" as the name, "String" as the type, and "Enabled" as the status. The left sidebar shows navigation options like Overview, Activity log, Access control (IAM), Tags, and Secrets.

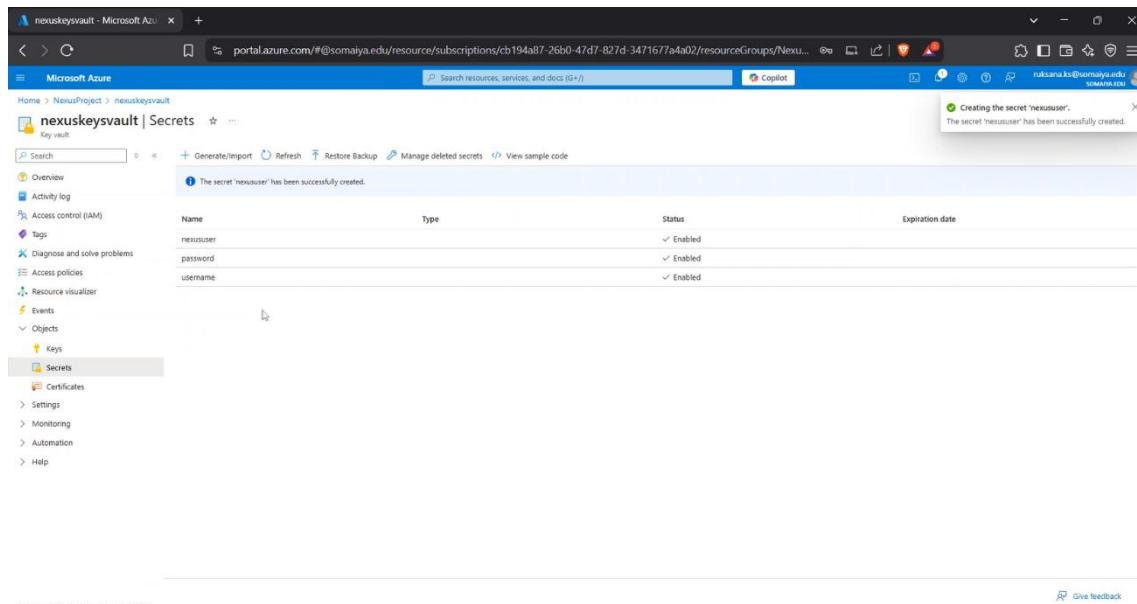
Name	Type	Status	Expiration date
password	String	Enabled	
username	String	Enabled	

The image shows the creation of a new secret named "newuser" in the Azure Key Vault, with manual upload and enabled status.

A screenshot of the "Create a secret" dialog in the Azure portal. The "Name" field is set to "newuser". The "Secret value" field contains "*****". The "Upload options" dropdown is set to "Manual". The "Enabled" switch is turned on. At the bottom, the "Create" button is highlighted with a mouse cursor.

Nexus DataLens

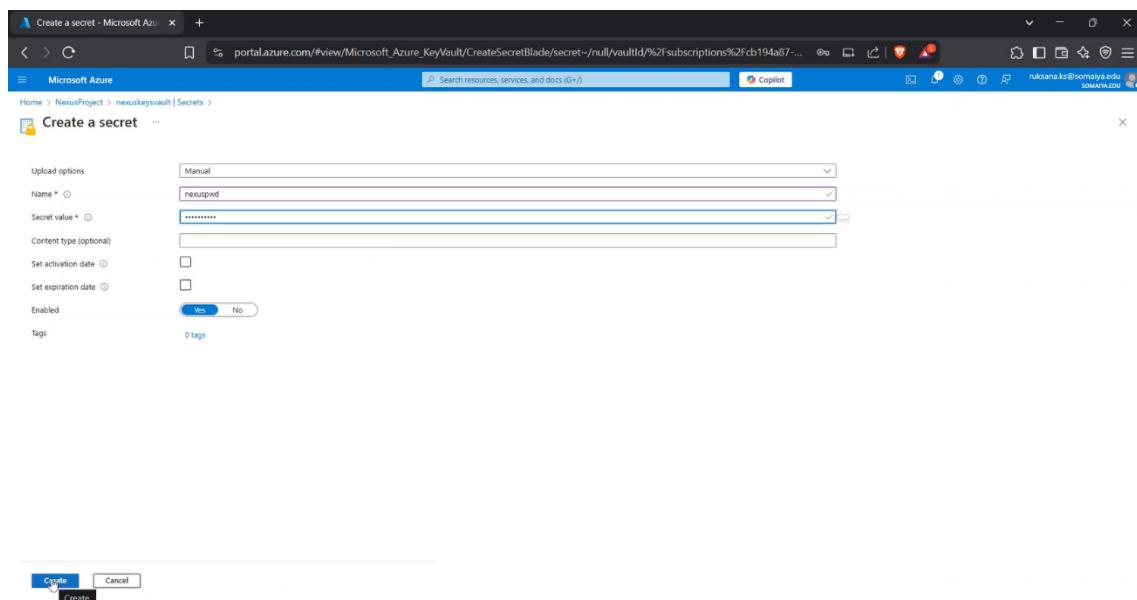
The secret "**newuser**" has been successfully created and added to the list of enabled secrets in the Azure Key Vault.



The screenshot shows the Azure Key Vault interface with the 'Secrets' blade selected. A success message at the top right states: "Creating the secret 'newuser'. The secret 'newuser' has been successfully created." The table lists three secrets:

Name	Type	Status	Expiration date
newuser		✓ Enabled	
password		✓ Enabled	
username		✓ Enabled	

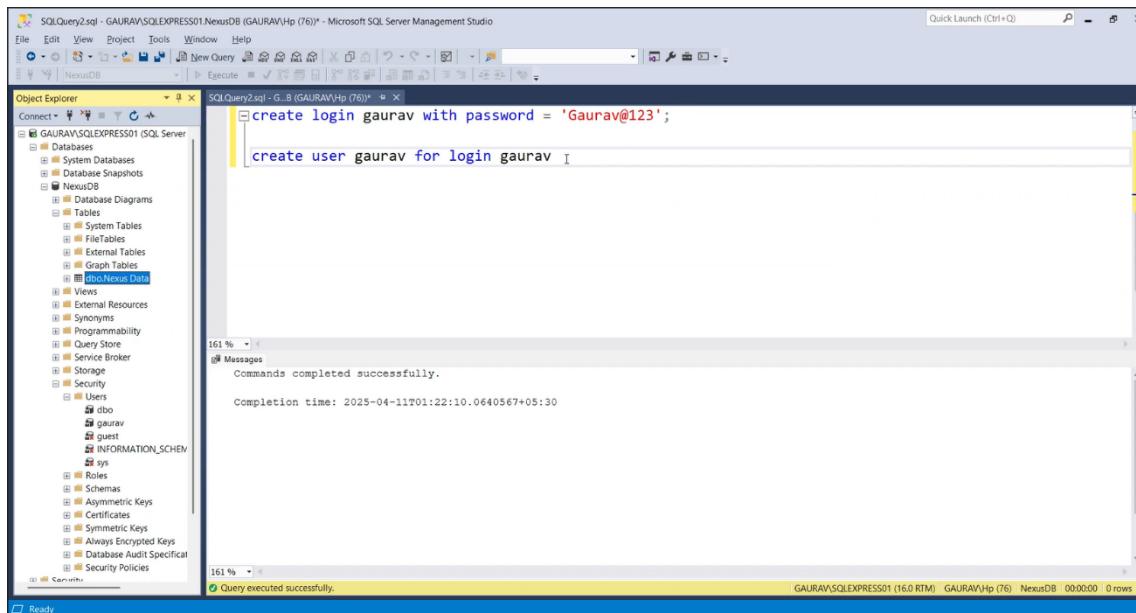
The creation of a new secret named "**newuserpwd**" with a specified value in Azure Key Vault, ready to be saved by clicking "**Create**".



The screenshot shows the 'Create a secret' dialog. The 'Name' field is set to 'newuserpwd' and the 'Secret value' field contains '*****'. The 'Content type (optional)' field is empty. The 'Enabled' switch is set to 'Yes'. At the bottom, the 'Create' button is highlighted in blue, while the 'Cancel' button is in grey.

Nexus DataLens

The SQL script creates a login named "gaurav" with a password and then creates a corresponding user for that login in the **NexusDB** database.



SQLQuery2.sql - GAURAV\SQLEXPRESS01.NexusDB (GAURAV\Hp (76)) - Microsoft SQL Server Management Studio

File Edit View Project Tools Window Help

Object Explorer

GAURAV\SQLEXPRESS01 (SQL Server)

Databases

System Databases

NexusDB

Database Snapshots

Tables

System Tables

FileTables

External Tables

Graph Tables

Views

External Resources

Synonyms

Programmability

Query Store

Service Broker

Storage

Security

Users

dbo

gaurav

guest

INFORMATION_SCHEMA

sys

Roles

Schemas

Asymmetric Keys

Certificates

Symmetric Keys

Always Encrypted Keys

Database Audit Specifications

Security Policies

Script

Ready

SQLQuery2.sql - G_B (GAURAV\Hp (76))

```
create login gaurav with password = 'Gaurav@123';
create user gaurav for login gaurav;
```

161% 161% 161%

Messages

Commands completed successfully.

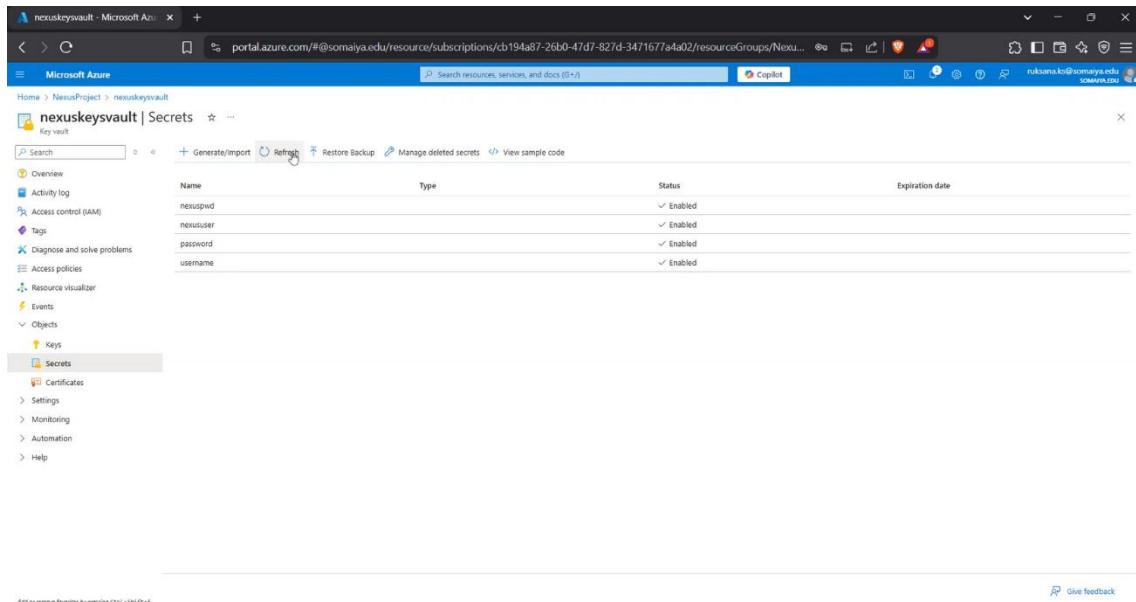
Completion time: 2025-04-11T01:22:10.0640567+05:30

161% 161% 161%

Query executed successfully.

GAURAV\SQLEXPRESS01 (16.0 RTM) GAURAV\Hp (76) NexusDB 00:00:00 0 rows

The Azure Key Vault now contains four enabled secrets: **nexuspwd**, **nexususer**, **password**, and **username**.



Microsoft Azure

Home > NexusProject > nexuskeysvault

Key vault

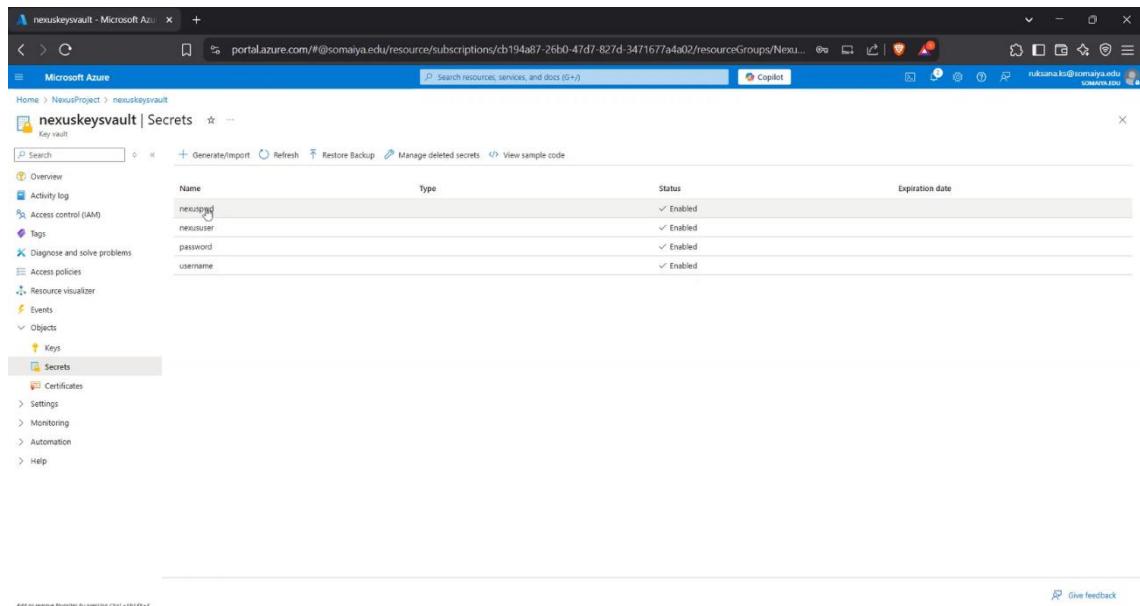
Secrets

Name	Type	Status	Expiration date
nexuspwd		✓ Enabled	
nexususer		✓ Enabled	
password		✓ Enabled	
username		✓ Enabled	

Give feedback

Nexus DataLens

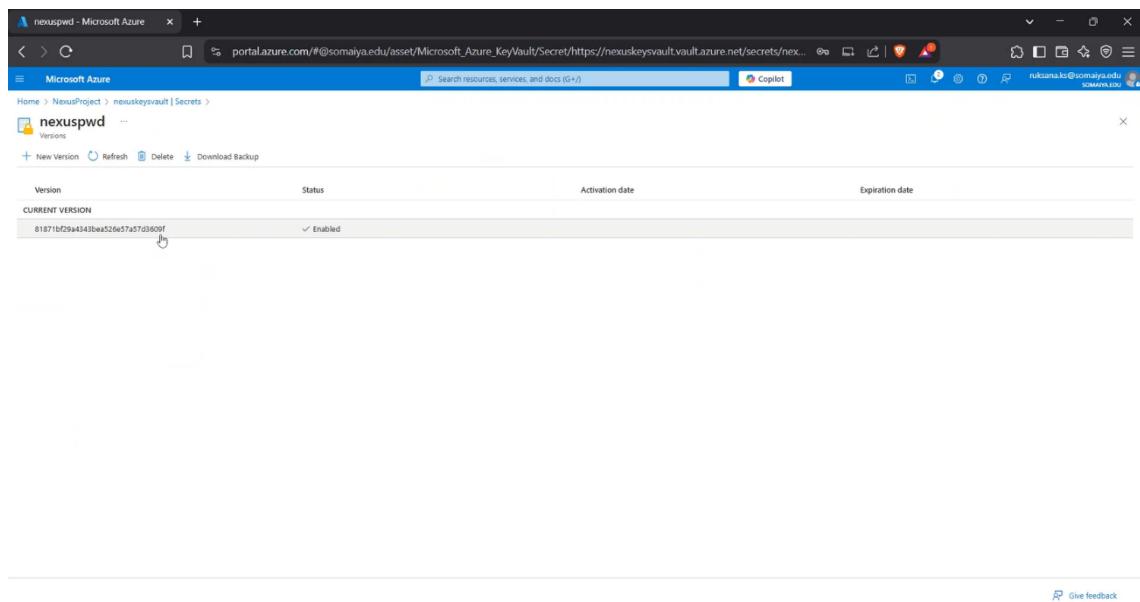
The secret named **nexuspwd** in the Azure Key Vault is selected, with all four listed secrets currently enabled.



A screenshot of the Microsoft Azure portal showing the 'Secrets' page for the 'nexuskeysVault' key vault. The page displays four secrets:

Name	Type	Status	Expiration date
nexuspwd		✓ Enabled	
nexususer		✓ Enabled	
password		✓ Enabled	
username		✓ Enabled	

The secret **nexuspwd** in Azure Key Vault has one enabled version with no set activation or expiration date.



A screenshot of the Microsoft Azure portal showing the 'Versions' page for the 'nexuspwd' secret in the 'nexuskeysVault' key vault. The page displays one current version:

Version	Status	Activation date	Expiration date
B1871bf29a4343bea526e57a57d03609f	✓ Enabled		

Nexus DataLens

The **Access control (IAM)** page for the **Azure Key Vault nexuskeyvault** allows managing role-based access, including checking access, assigning roles, and viewing permissions.

The screenshot shows the 'Access control (IAM)' page for the 'nexuskeyvault' key vault. The left sidebar includes options like Overview, Activity log, Tags, Diagnose and solve problems, and Access policies. The main content area has tabs for Check access, Role assignments, Roles, Deny assignments, and Classic administrators. The 'Check access' tab is selected, displaying sections for 'My access' (viewing level of access to the resource), 'Check access' (reviewing access for users, groups, service principals, or managed identities), and four cards: 'Grant access to this resource', 'View access to this resource', 'View deny assignments', and 'New! Permissions Management'.

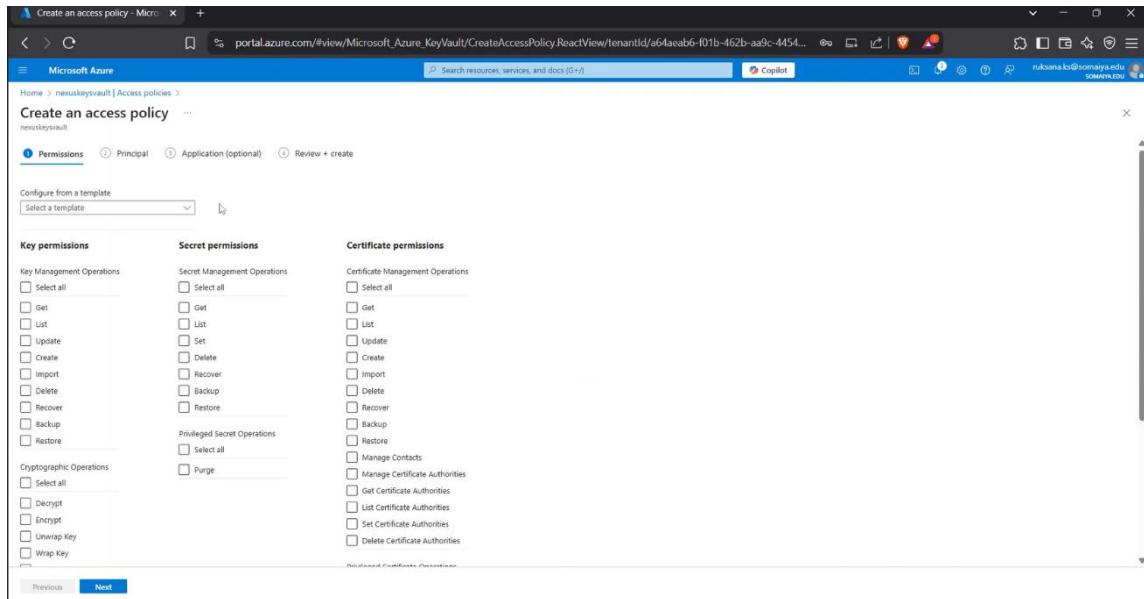
The Access policies page of **nexuskeyvault** displays users with specific permissions to manage keys, secrets, and certificates, with an option to create new policies.

The screenshot shows the 'Access policies' page for the 'nexuskeyvault' key vault. The left sidebar includes options like Overview, Activity log, Tags, Diagnose and solve problems, and Access policies. The main content area shows a table of access policies. The table has columns for Name, Email, Key Permissions, Secret Permissions, and Certificate Permissions. It lists two entries: one for 'UNKNOWN' with a GUID email and another for 'USER' with the email 'rukana.kr@somaiya.edu'. Both entries have the same permission set: Get, List, Set, Delete, Recover, Backup, Restore.

Name	Email	Key Permissions	Secret Permissions	Certificate Permissions
UNKNOWN	d05d52c-548e-4d9b-b524-6dc79aa3e17	Get, List, Set, Delete, Recover, Backup, Restore		
USER	2115112-SHAILESH RUKANA KHATUN	Get, List, Update, Create, Import, Delete, Recover, Ba...	Get, List, Set, Delete, Recover, Backup, Restore	Get, List, Update, Create, Import, Delete, Recover, Ba...

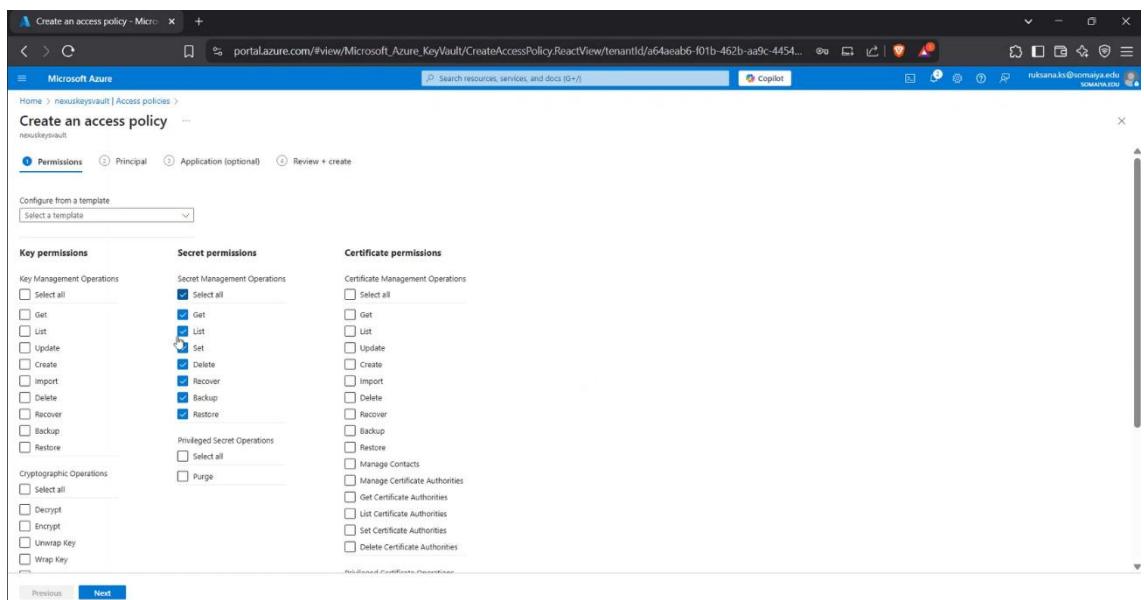
Nexus DataLens

Azure allows configuring permissions for a new Key Vault access policy, including key, secret, and certificate operations.



The screenshot shows the 'Create an access policy' wizard in the Azure portal. The current step is 'Permissions'. In the 'Secret permissions' section, the 'Get', 'List', and 'Delete' checkboxes under 'Secret Management Operations' are selected. Other checkboxes like 'Select all' and 'Set' are also present but not selected.

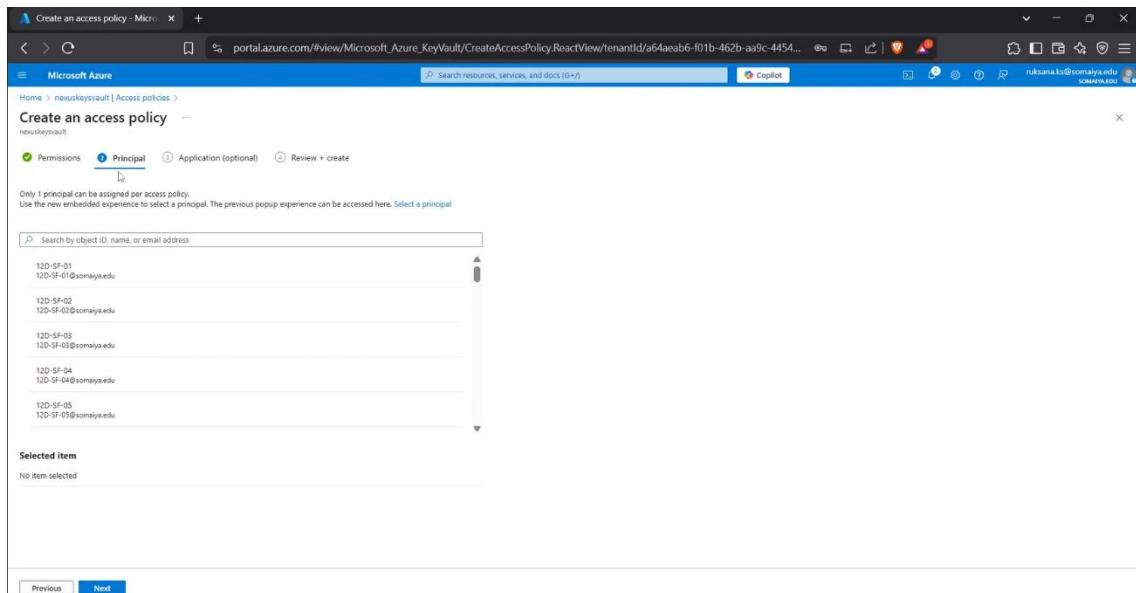
Secret permissions "**Get**," "**List**," and "**Delete**" are selected while creating a Key Vault access policy in Azure.



The screenshot shows the 'Create an access policy' wizard in the Azure portal. The current step is 'Permissions'. In the 'Secret permissions' section, multiple checkboxes under 'Secret Management Operations' are selected: 'Get', 'List', 'Delete', 'Recover', 'Backup', and 'Restore'. The 'Select all' checkbox is also checked. Other sections like 'Key Management Operations' and 'Certificate permissions' are visible but have fewer selected checkboxes.

Nexus DataLens

It allows selecting a principal (user, app, or group) to assign the access policy in Azure Key Vault.



Create an access policy - Microsoft Azure

portal.azure.com/#view/Microsoft_Azure_KeyVault/CreateAccessPolicy/ReactView/tenantId/a64aeab6-f01b-462b-aa9c-4454... Copilot nukusaka.s@somalya.edu SOMALYA.EDU

Microsoft Azure

Home > nexuskeyvault | Access policies > Create an access policy ...

Only 1 principal can be assigned per access policy.
Use the new embedded experience to select a principal. The previous popup experience can be accessed here. Select a principal.

Search by object ID, name, or email address

12D-SF-01
12D-SF-01@somalya.edu

12D-SF-02
12D-SF-02@somalya.edu

12D-SF-03
12D-SF-03@somalya.edu

12D-SF-04
12D-SF-04@somalya.edu

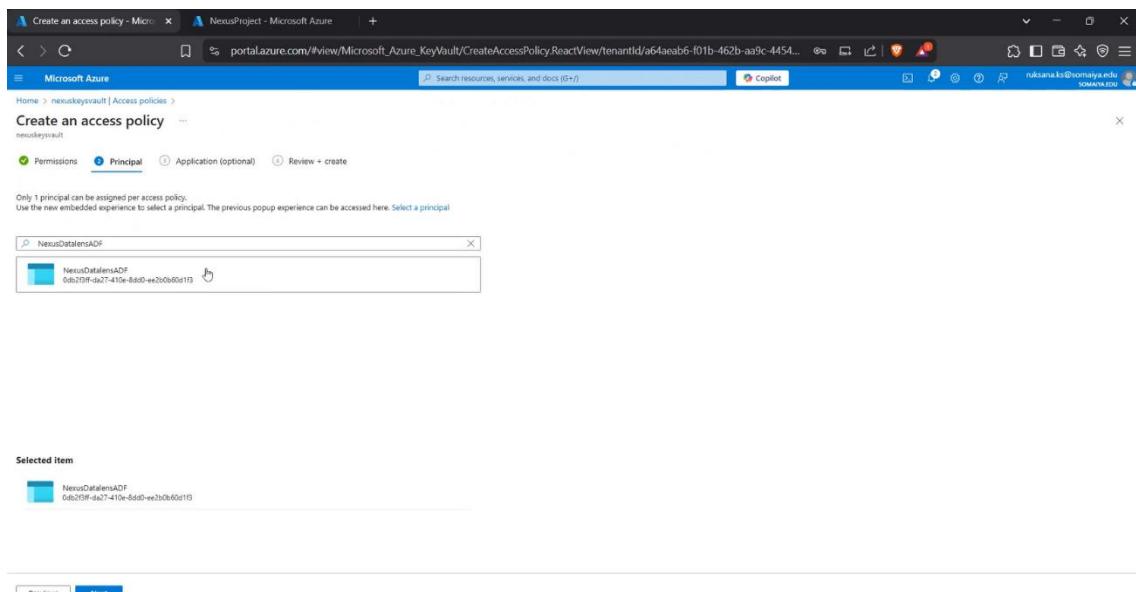
12D-SF-05
12D-SF-05@somalya.edu

Selected item

No item selected

Previous Next

The principal **NexusDataLensADF** is selected to assign permissions in the Azure Key Vault access policy.



Create an access policy - Microsoft Azure

portal.azure.com/#view/Microsoft_Azure_KeyVault/CreateAccessPolicy/ReactView/tenantId/a64aeab6-f01b-462b-aa9c-4454... Copilot nukusaka.s@somalya.edu SOMALYA.EDU

Microsoft Azure

Home > nexuskeyvault | Access policies > Create an access policy ...

Only 1 principal can be assigned per access policy.
Use the new embedded experience to select a principal. The previous popup experience can be accessed here. Select a principal.

NexusDataLensADF

NexusDataLensADF
0db20ff-d27-410e-8dd0-e220a0d1f3

Selected item

NexusDataLensADF

Previous Next

Nexus DataLens

The final review screen confirms that all secret permissions are granted to the principal **NexusDataLensADF**, and the access policy is ready to be created.

Key Permissions

Key Management Operations	None selected
Cryptographic Operations	None selected
Privileged Key Operations	None selected
Rotation Policy Operations	None selected

Secret Permissions

Secret Management Operations	All selected
Privileged Secret Operations	None selected

Certificate Permissions

Certificate Management Operations	None selected
Privileged Certificate Operations	None selected

Principal

Principal name	NexusDataLensADF
Object ID	2dd93ee4-09a3-4a8e-9557-3c0587209913

Application

Authorized application	None selected
Object ID	None selected

Previous Create

A new access policy for the **NexusDataLensADF** application has been successfully added with full secret permissions (Get, List, Set, Delete, Recover, Backup, Restore).

Access policies enable you to have fine grained control over access to vault items. Learn more

Name	Email	Key Permissions	Secret Permissions	Certificate Permissions
NeusDataLensADF		Get, List, Set, Delete, Recover, Backup, Restore		
d05602c548e4d36-b524-6da79aa3e17		Get, List, Set, Delete, Recover, Backup, Restore		
2115112-SHAIOH RIUKSANA KHATUN	rulkhana.k@somaiya.edu	Get, List, Update, Create, Import, Delete, Recover, Ba...	Get, List, Set, Delete, Recover, Backup, Resto...	Get, List, Update, Create, Import, Delete, Reco...

< Previous Page 1 of 1 Next >

Add or remove (invites) by pressing Ctrl + Shift + F

Nexus DataLens

A confirmation dialog is shown to delete one selected access policy from the Azure Key Vault.

The screenshot shows the 'Access policies' page in the Azure portal. On the left, there's a sidebar with options like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, and Access policies. The main area displays a table of access policies. One row for 'NexusDataInesADF' is selected. A confirmation dialog box titled 'Delete access policies' appears in the center, asking 'Are you sure you want to delete (1) access policies?' with 'Delete' and 'Cancel' buttons.

The access policy was successfully deleted, leaving two remaining entries for the application "NexusDataInesADF" and the user "SHAIKH RUKSANA KHATUN."

This screenshot shows the same 'Access policies' page after the deletion. The table now lists only two items: 'NexusDataInesADF' and '2115112-SHAIKH RUKSANA KHATUN'. Both entries have their respective email addresses listed next to them.

SETTING RESOURCE GROUP

The "NexusProject" resource group currently has no visible resources due to active filters, with an option to create new resources or clear filters.

The screenshot shows the Microsoft Azure portal interface for the 'NexusProject' resource group. The top navigation bar includes 'Microsoft Azure', 'Search resources, services, and docs (Q+)', 'Copilot', and user information 'rukshanak@somaiya.edu SOMAIYA.EDU'. The main content area displays the 'Resource group' details for 'NexusProject'. It shows the subscription (Abuse for Students), deployment (bio.deployments), location (Central India), and tags ('Add tags'). The 'Resources' tab is selected, showing a search bar and filter options ('Filter for any field...', 'Type equals all', 'Location equals all', 'Add filter'). A message indicates 'Showing 0 to 0 of 0 records.' Below this, there's a large placeholder image and the text 'No resources match your filters' with a note 'Try changing or clearing your filters.' Buttons for '+ Create resources' and 'Clear filters' are present. At the bottom right, there's a 'Give feedback' link.

The Azure Marketplace displays various Synapse Analytics-related services and solutions available for creation or subscription within the "NexusProject" environment.

The screenshot shows the Microsoft Azure Marketplace search results for 'synapse analytics'. The search bar at the top contains 'synapse analytics'. The results list 102 items, with the first few shown in detail:

- Azure Synapse Analytics** (Microsoft): Azure Service. Limitless analytics service with unmatched time to insight. Options: Create, Subscribe.
- Azure Synapse Analytics (private link hubs)** (Microsoft): Azure Service. Connect to Azure Synapse Studio using private endpoints. Options: Create, Subscribe.
- Synapse Data Fabric** (Spectra Systems LLC): SaaS. Synapse Data Fabric is a comprehensive data management platform that unifies disparate data sources. Starts at \$0.684/month. Options: Subscribe.
- Datometry Hyper-Q for Azure Synapse Analytics** (Datometry): SaaS. Re-platform from Teradata and Essbase to Azure Synapse in a fraction of the time, cost, and risk. Starts at \$2,392,251.04/month. Options: Subscribe.
- Datometry Hyper-Q for Azure Synapse Analytics** (Datometry): SaaS. Run existing Teradata applications natively on Azure Synapse Analytics. Starts at \$2,392,251.04/month. Options: Subscribe.
- Moyo Azure Synapse Retail Recommender Solution** (Moyo Business Advisory): Azure Application. Create recommendations utilizing the power of Azure Synapse and trained modeling. Options: Create.
- Xpert BI with Azure Synapse** (Xpert BI): BI Builders as Virtual Machine. Data warehouse automation software, Xpert BI with Azure Synapse. Options: Create.

Nexus DataLens

The form is being filled to create a new Synapse workspace under the "**NexusProject**" resource group with specified subscription, region, and linked Data Lake Storage Gen2 account.

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 *

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 *

Select Data Lake Storage Gen2 * From subscription Manually via URL

Account name *

File system name *

A Synapse workspace named "**nexussynapse**" is being created in Central India, linked to the "**nexusdatalensstorage**" Data Lake Storage with "**bronze**" as the file system.

Home > NexusProject > Marketplace > Create Synapse workspace

Resource group *

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 *

Select Data Lake Storage Gen2 * From subscription Manually via URL

Account name *

File system name *

Assign myself the Storage Blob Data Contributor role on the Data Lake Storage Gen2 account to interactively query it in the workspace.

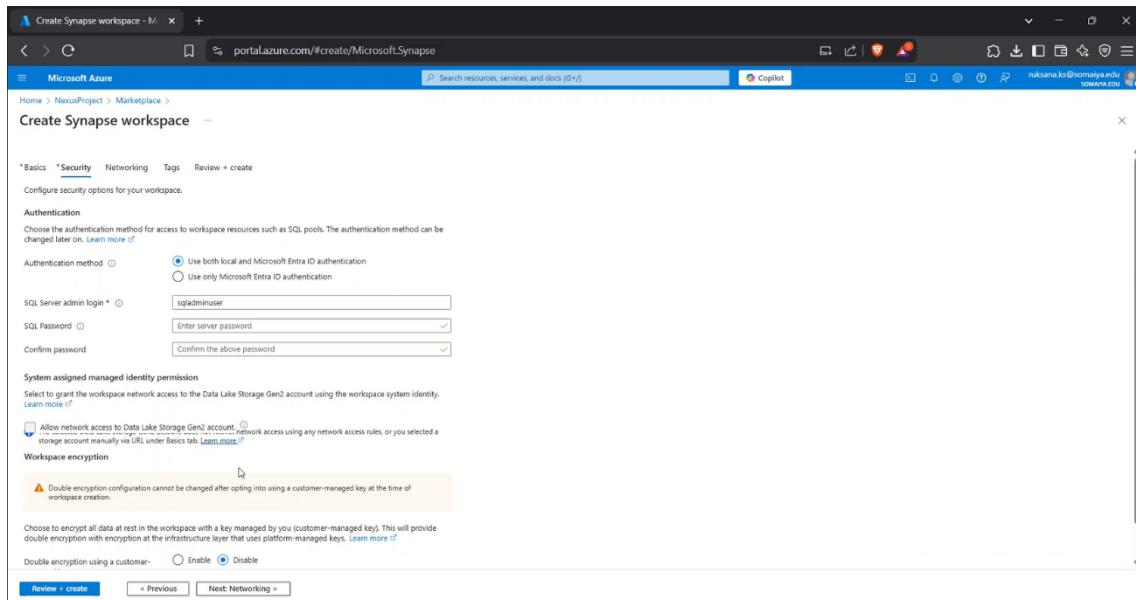
We will automatically grant the workspace Identity data access to the specified Data Lake Storage Gen2 account, using the Storage Blob Data Contributor role. To enable other users to use this storage account after you create your workspace, perform these tasks:

- Assign other users to the Contributor role on the workspace.
- Assign other users to the appropriate Azure AD role using Synapse Studio.
- Assign yourself and other users to the Storage Blob Data Contributor role on the storage account.

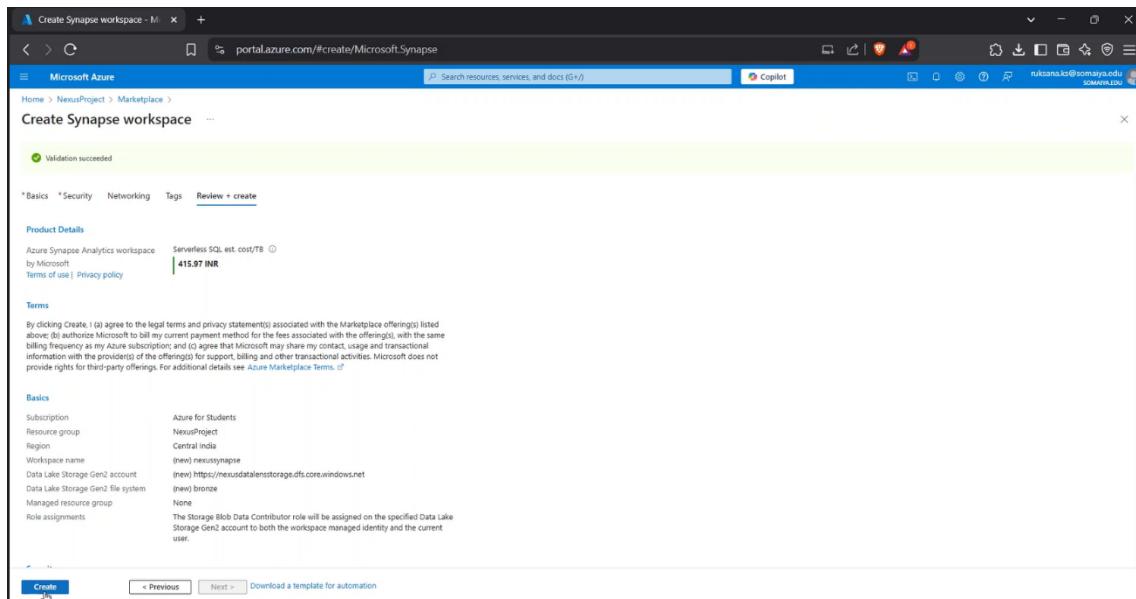
[Learn more](#)

Nexus DataLens

The Synapse workspace is being configured with SQL authentication, network access to Data Lake enabled, and default platform-managed encryption selected.

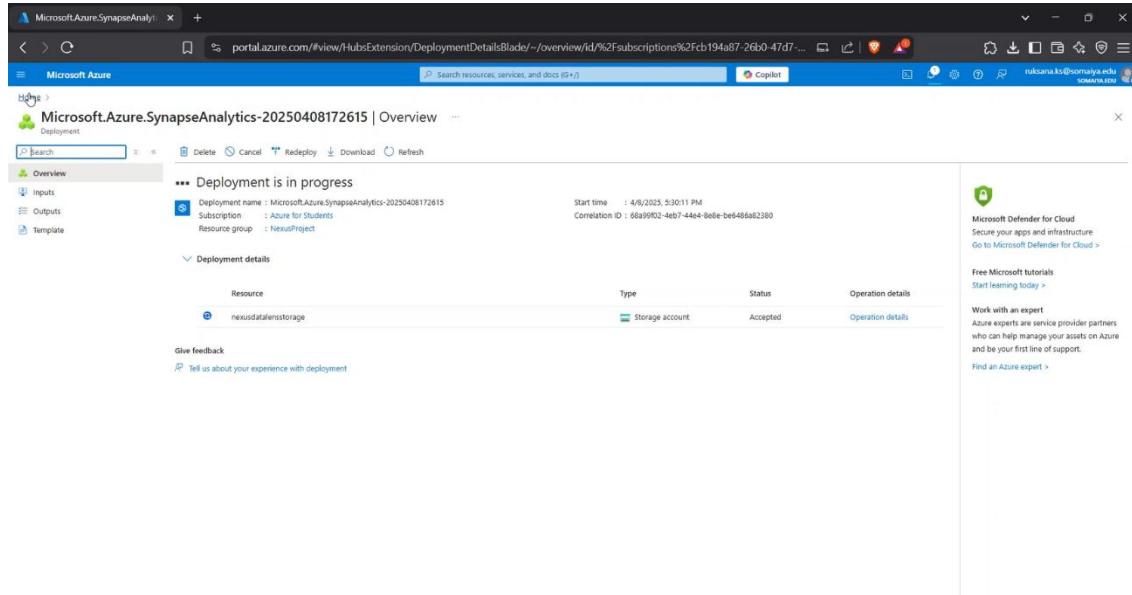


The **Synapse workspace setup review has succeeded**, and the user is ready to create it with specified configuration and pricing.



Nexus DataLens

The Synapse workspace deployment is currently in progress with the storage account resource accepted under the **NexusProject** resource group.

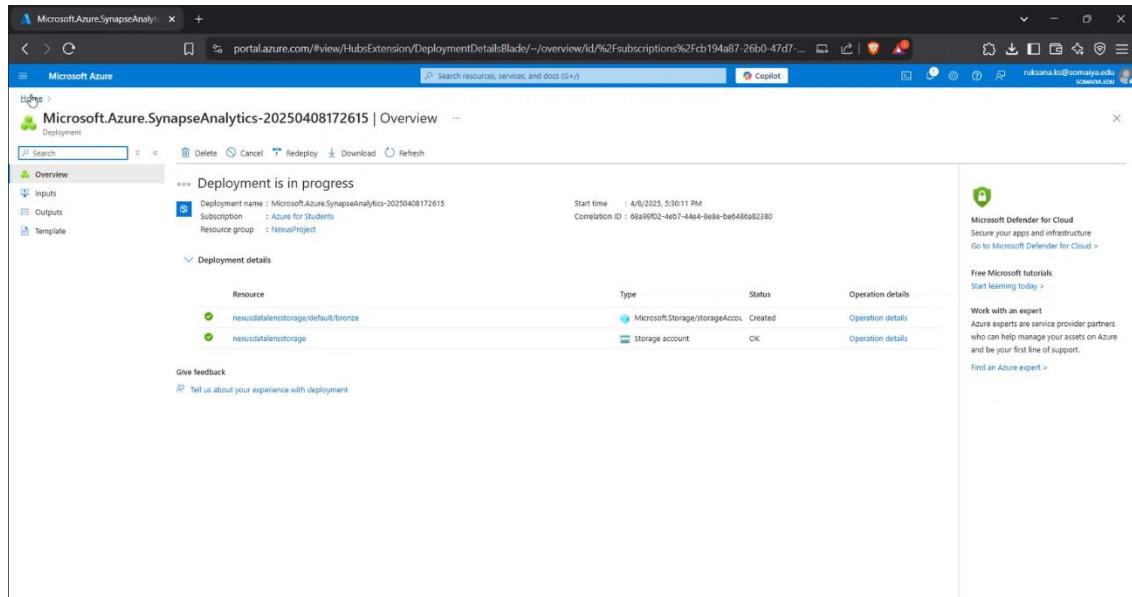


Microsoft Azure-SynapseAnalytics-20250408172615 | Overview

Deployment is in progress

Resource: nexusdatalensstorage, Type: Storage account, Status: Accepted

The Synapse workspace deployment is **in progress**, with the storage container bronze created and the storage account status marked as OK.



Microsoft Azure-SynapseAnalytics-20250408172615 | Overview

Deployment is in progress

Resource: nexusdatalensstorage/default/bronze, Type: Microsoft.Storage/storageAccounts, Status: Created

Resource: nexusdatalensstorage, Type: Storage account, Status: OK

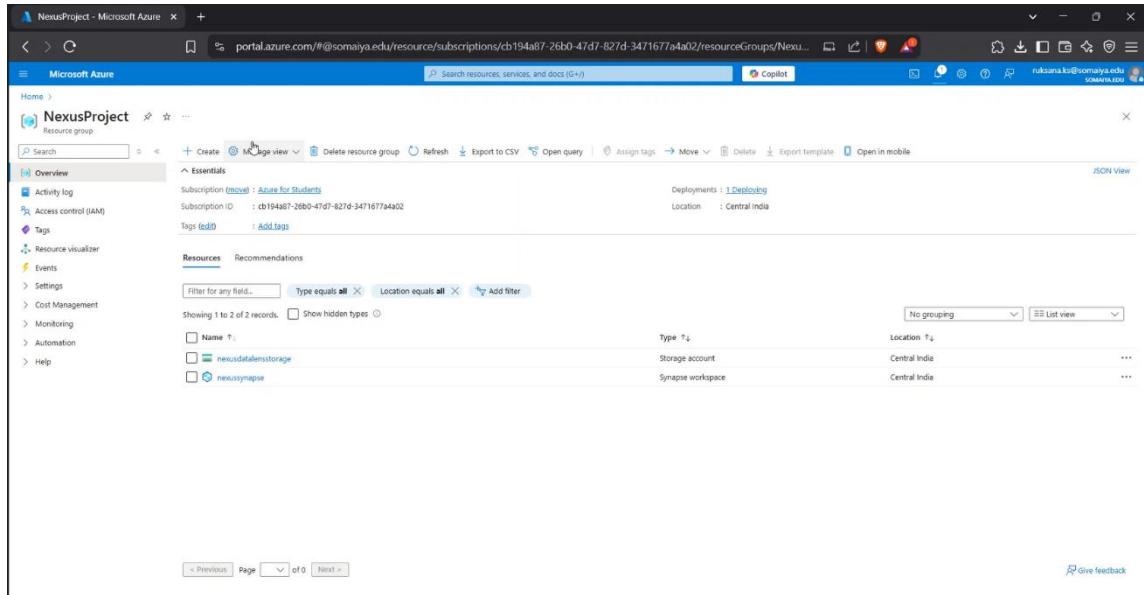
The Azure portal home page displays recent resources, with "NexusProject" (a resource group) being the most recently accessed.

The image shows an Azure portal page displaying the "NexusProject" resource group with one storage account resource named "nexusdatastenstorage" located in Central India.

Name	Type	Location
nexusdatastenstorage	Storage account	Central India

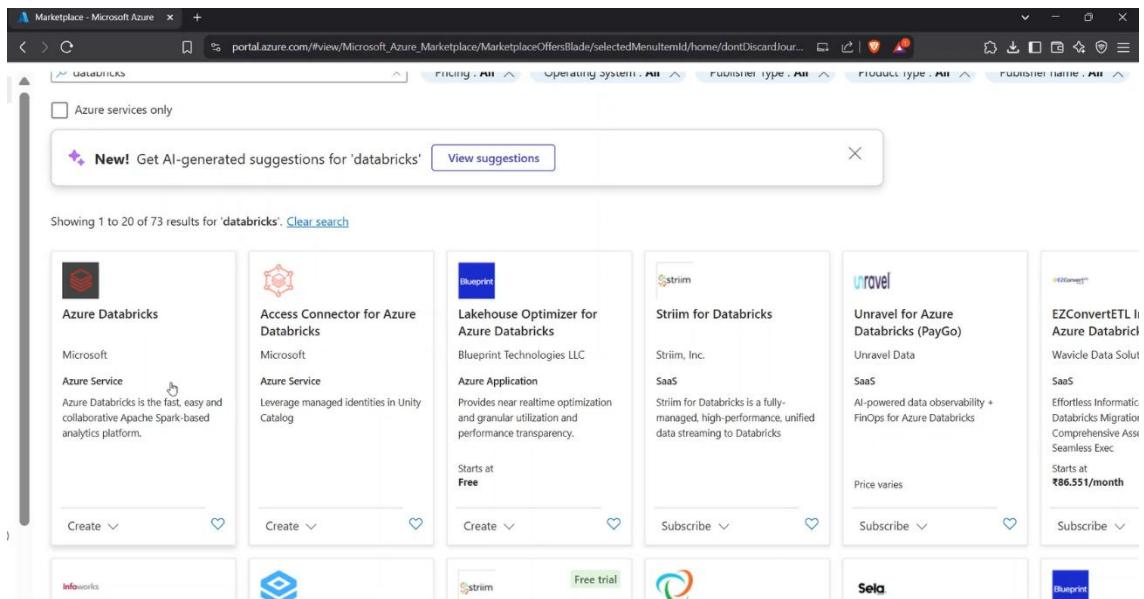
Nexus DataLens

The updated Azure portal view shows the "NexusProject" resource group now containing two resources: a storage account ("nexusdatalensstorage") and a Synapse workspace ("nexussynapse"), both located in Central India.



A screenshot of the Microsoft Azure portal showing the "NexusProject" resource group. The portal displays the "Overview" tab with basic information like Subscription ID, Tags, and Resource types. Under "Resources", two items are listed: "nexusdatalensstorage" (Storage account) and "nexussynapse" (Synapse workspace), both located in Central India.

The Azure Marketplace displays various Databricks-related services, including "**Azure Databricks**" by Microsoft for Spark-based analytics and other integration and optimization tools.



A screenshot of the Microsoft Azure Marketplace search results for "databricks". The search results show various services related to Databricks, including Azure Databricks by Microsoft, Access Connector for Azure Databricks, Lakehouse Optimizer for Azure Databricks, Striim for Databricks, Unravel for Azure Databricks (PayGo), and EZConvertETL for Azure Databricks.

Nexus DataLens

The user is about to create an "**Azure Databricks**" workspace from the Azure Marketplace by clicking the "**Create**" button under the Azure Databricks service.

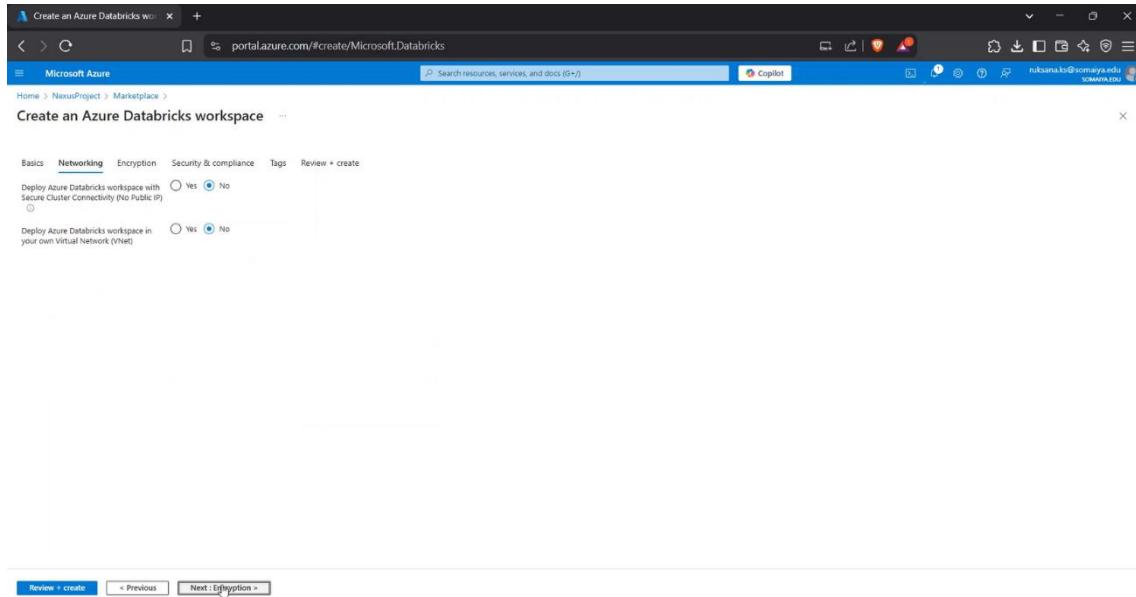
A screenshot of the Microsoft Azure Marketplace interface. The search bar at the top contains the text 'databricks'. Below the search bar, a message says 'New! Get AI-generated suggestions for 'databricks''. The results section shows 'Showing 1 to 20 of 73 results for 'databricks''. The first result is 'Azure Databricks' by Microsoft, described as an 'Azure Service'. It has a brief description: 'Azure Databricks is the fast, easy and collaborative Apache Spark-based analytics platform.' Below the description are 'Create' and 'Azure Databricks' buttons. To the right of this are other marketplace items like 'Access Connector for Azure Databricks', 'Lakehouse Optimizer for Azure Databricks', 'Striim for Databricks', 'Unravel for Azure Databricks (PayGo)', and 'EZConvert™ Azure Databricks'. Each item has its own thumbnail, name, provider, description, and creation options.

The user is configuring the basic settings to create an Azure Databricks workspace named "**nexusdatabricks**" in the "**NexusProject**" resource group within the Central India region.

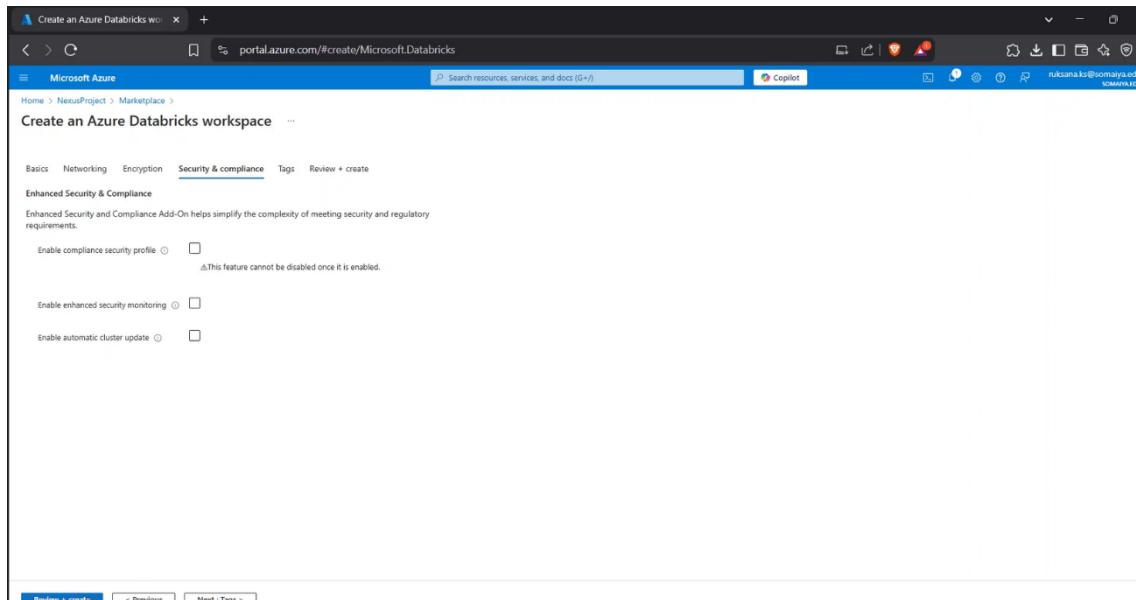
A screenshot of the 'Create an Azure Databricks workspace' wizard on the Azure portal. The page title is 'Create an Azure Databricks workspace'. The 'Basics' tab is selected. Under 'Project Details', the 'Subscription' dropdown is set to 'Azure for Students' and the 'Resource group' dropdown is set to 'NexusProject'. Under 'Instance Details', the 'Workspace name' is 'nexusdatabricks', 'Region' is 'Central India', and 'Pricing Tier' is 'Premium (+ Role-based access controls)'. At the bottom, there are 'Review + create' and 'Next : Networking >' buttons.

Nexus DataLens

The user is configuring networking options for the Azure Databricks workspace, choosing whether to enable secure cluster connectivity and virtual network deployment.

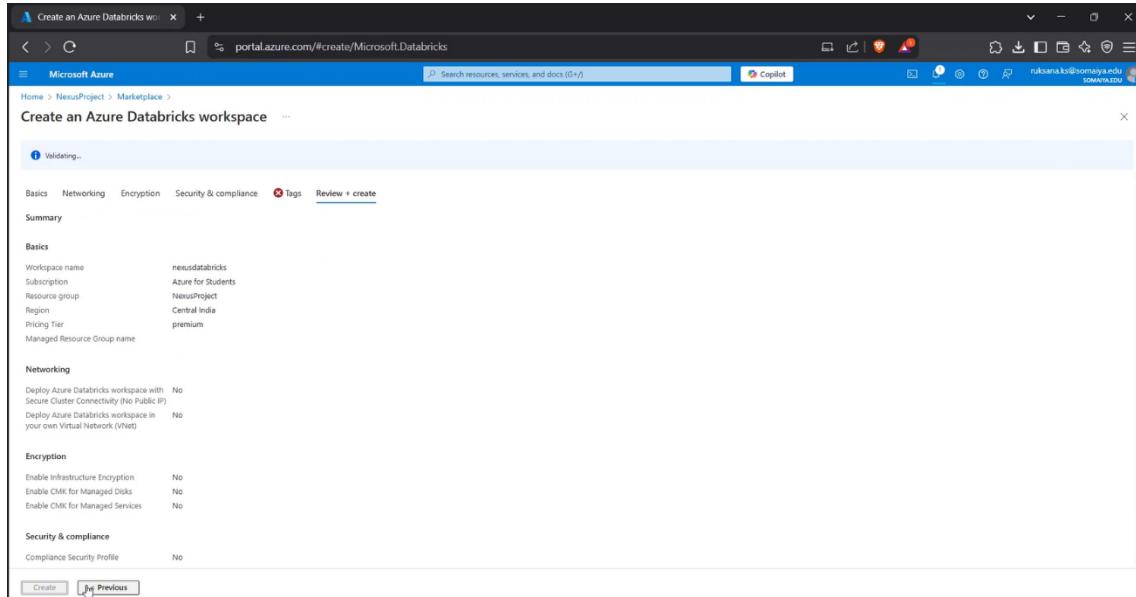


The user is on the Security & Compliance tab while creating an Azure Databricks workspace, with optional settings for compliance, monitoring, and automatic cluster updates before proceeding to review and create.

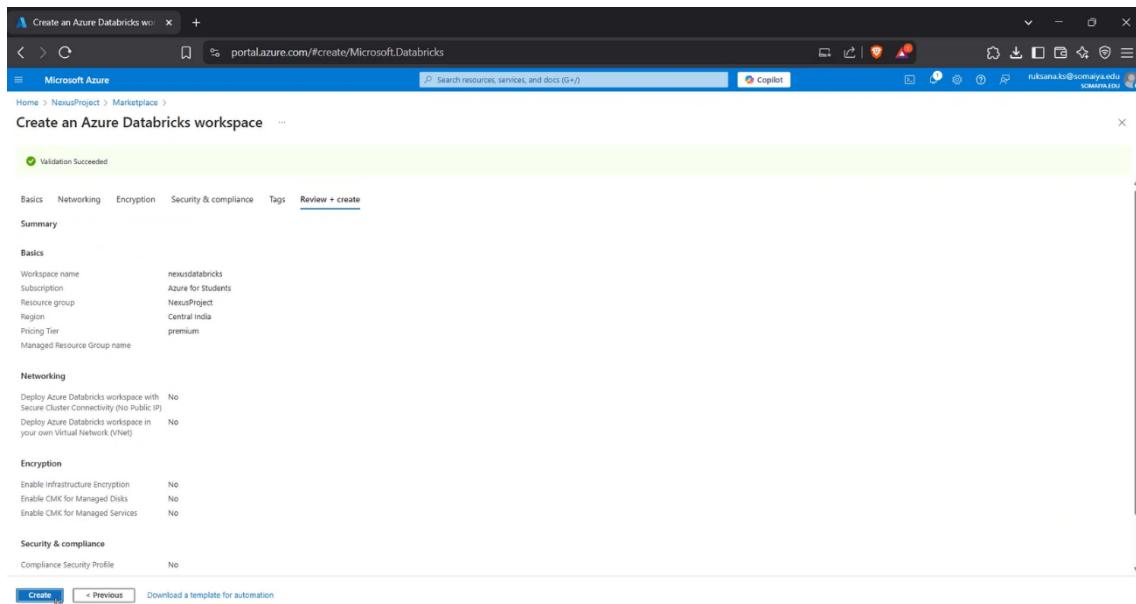


Nexus DataLens

The user is on the final "Review + create" step for setting up an Azure Databricks workspace, reviewing all configuration details before deployment.



The user has successfully validated the configuration and is ready to create the **Azure Databricks workspace** by clicking the "Create" button.



Nexus DataLens

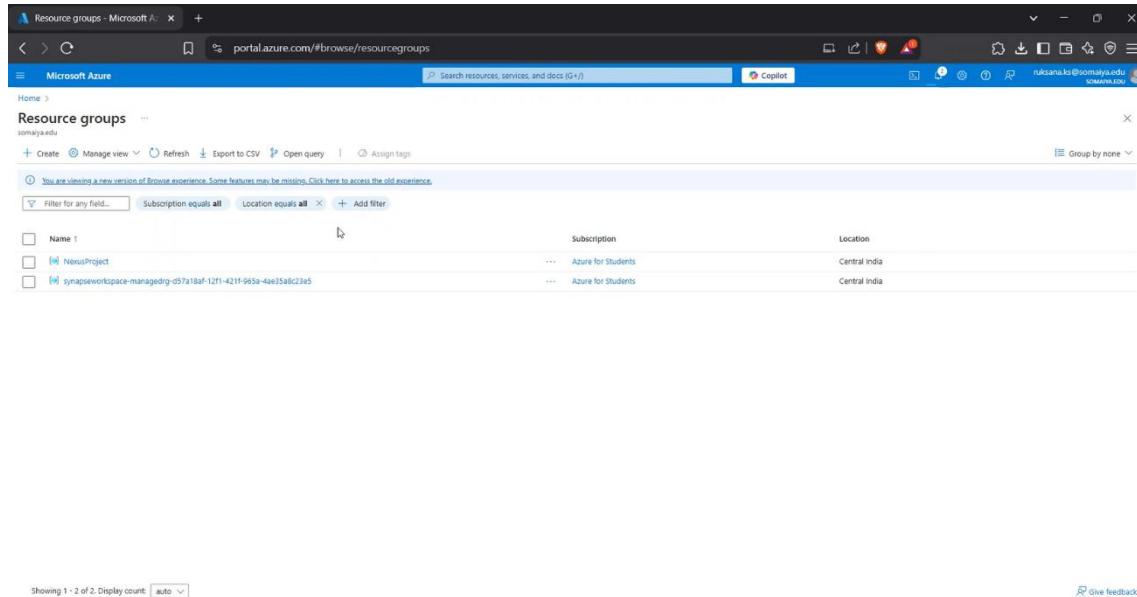
The deployment of the Azure Databricks workspace named "**nexusdatabricks**" is currently in progress within the "**NexusProject**" resource group.

The screenshot shows the Azure portal's deployment details for the 'nexusdatabricks' workspace. The deployment is currently in progress, having started at 5:33:23 PM on 4/6/2025. The workspace is associated with the 'Azure for Students' subscription and the 'NexusProject' resource group. The deployment details table lists the resource 'nexusdatabricks' as an 'Azure Databricks Service' that has been 'Created'. On the right side of the page, there are promotional links for Microsoft Defender for Cloud, Free Microsoft tutorials, and Work with an expert.

On the Azure portal home screen click on the "**Resource groups**" option to view or **create a new resource group**.

The screenshot shows the Azure portal's home screen. The 'Resource groups' option under the 'Azure services' section is highlighted with a yellow box. Below the services section, there is a table of resources, a 'Navigate' section with links to Subscriptions, Resource groups, All resources, and Dashboard, and a 'Tools' section with links to Microsoft Learn, Azure Monitor, Microsoft Defender for Cloud, and Cost Management. There is also a 'Useful links' section and an 'Azure mobile app' section.

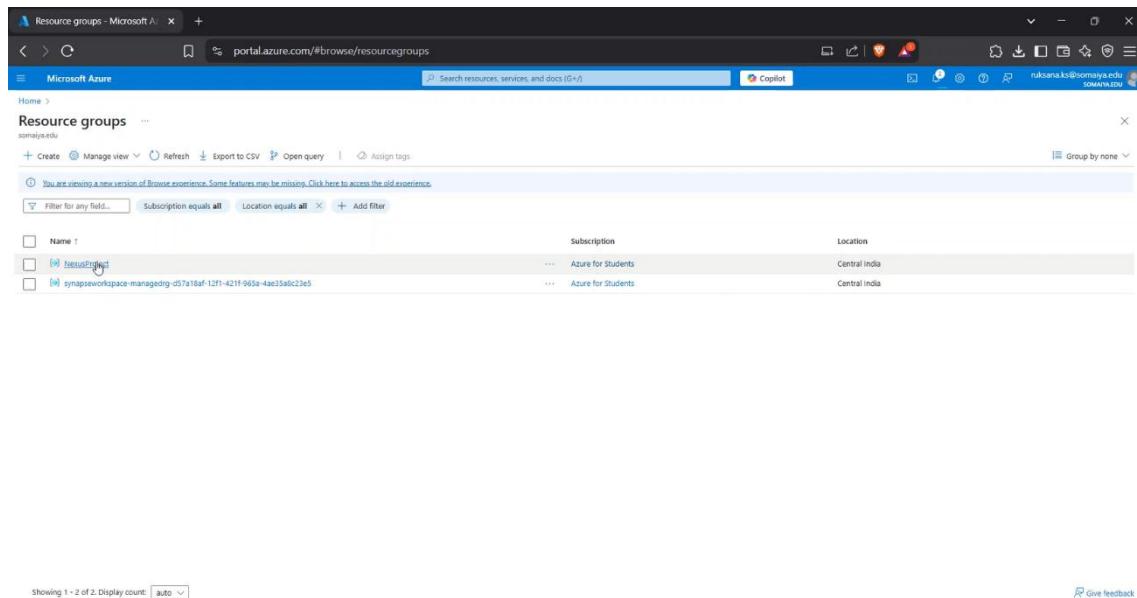
The "**Resource groups**" page in the Azure portal displays existing resource groups under the "**Azure for Students**" subscription, with both located in Central India.



The screenshot shows the Microsoft Azure Resource groups page. The URL is portal.azure.com/#browse/resourcegroups. The page title is "Resource groups". There are two resource groups listed:

Name	Subscription	Location
NexusProject	Azure for Students	Central India
synapseworkspace-managedrg-d57a18af-12f1-421f-965a-4ae35a8c23e5	Azure for Students	Central India

The user selects the "**NexusProject**" resource group from the list of available groups under the "**Azure for Students**" subscription in the Central India region.



The screenshot shows the Microsoft Azure Resource groups page. The URL is portal.azure.com/#browse/resourcegroups. The page title is "Resource groups". The "NexusProject" resource group is highlighted in the list:

Name	Subscription	Location
NexusProject	Azure for Students	Central India
synapseworkspace-managedrg-d57a18af-12f1-421f-965a-4ae35a8c23e5	Azure for Students	Central India

Nexus DataLens

The user searches for "Key Vault" in the Azure Marketplace under the NexusProject resource group to create a new key vault service.

A screenshot of the Microsoft Azure Marketplace search results. The search bar at the top contains the text "key vault". Below the search bar, there are several filters: "Pricing : All", "Operating System : All", "Publisher Type : All", "Product Type : All", and "Publisher Name : All". A checkbox labeled "Azure services only" is checked. A modal window titled "New! Get AI-generated suggestions for 'keyvault'" is open, showing "View suggestions". The main search results show five items:

- Key Vault** (Microsoft): Described as "Safeguard cryptographic keys and other secrets used by cloud apps and services". It has a "Create" button.
- Commvault Cloud for Sentinel** (Commvault): Described as "Azure Application". It has a "Create" button.
- AZ-204 Workshop | Developing Solutions for Opisity, LLC**: Described as "5-day workshop with Onelabue and Cloud Sandbox focused on Azure Development". It has a "Subscribe" button.
- HSM Ingress Controller** (Strd Tech): Described as "Container (Kubernetes App)". It has a "Create" button.
- Unified Contacts Pro – Companion to Microsoft** (glueckjanja AG): Described as "Finding All your Contacts in Teams - Azure Active Directory, Exchange Online and SharePoint". It has a "Subscribe" button.

At the bottom of the page, there are navigation buttons: "Previous", "Page 1 of 1", and "Next".

The user is configuring and creating a new Azure Key Vault named "nexuskeyvault" under the "NexusProject" resource group in the Central India region.

A screenshot of the "Create a key vault" wizard, step 1: "Create a key vault". The page title is "Create a key vault - Microsoft Azure". The URL is "portal.azure.com/#create/Microsoft.KeyVault". The left sidebar shows the navigation path: Home > Resource groups > NexusProject > Marketplace > Create a key vault.

The main form fields are:

- Subscription ***: Azure for Students
- Resource group ***: NexusProject
- Key vault name ***: nexuskeyvault
- Region ***: Central India
- Pricing tier ***: Standard

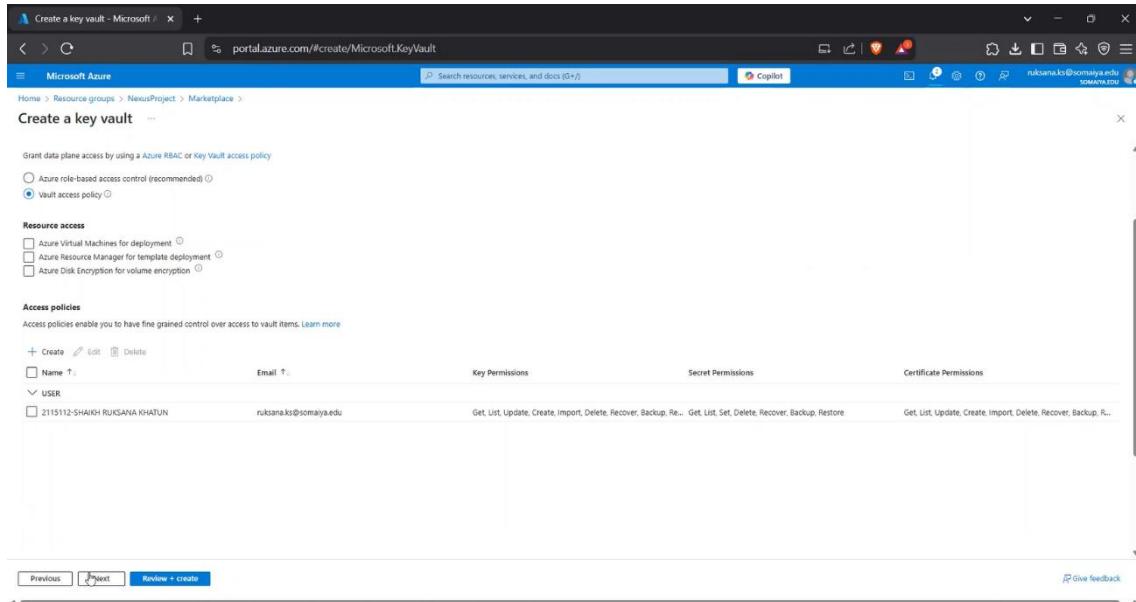
Below these fields are "Recovery options" settings:

- Soft-delete**: Enabled
- Days to retain deleted vaults ***: 90
- Purge protection**:
 - Disable purge protection (allow key vault and objects to be purged during retention period)
 - Enable purge protection (enforce a mandatory retention period for deleted vaults and vault objects)

At the bottom of the page are "Previous", "Next", and "Review + create" buttons. There is also a "Give feedback" link at the bottom right.

Nexus DataLens

The user is assigning full key, secret, and certificate permissions to a specified user while setting access policies for the new Azure Key Vault before final review and creation.



Grant data plane access by using a Azure RBAC or Key Vault access policy

Azure role-based access control (recommended) Vault access policy

Resource access

Azure Virtual Machines for deployment Azure Resource Manager for template deployment Azure Disk Encryption for volume encryption

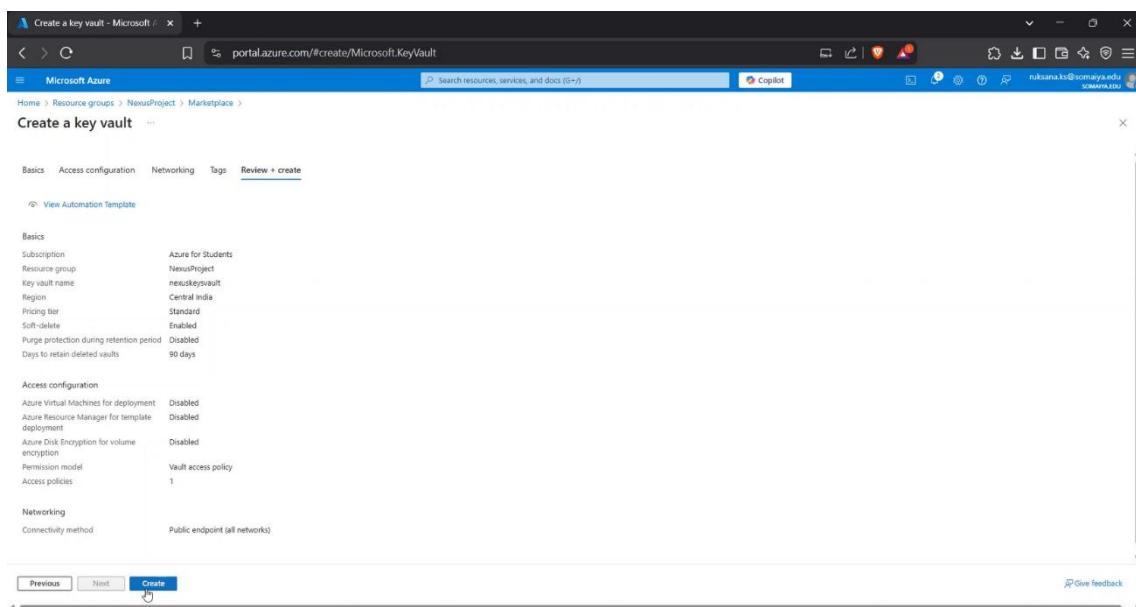
Access policies

Access policies enable you to have fine grained control over access to vault items. Learn more

Name	Email	Key Permissions	Secret Permissions	Certificate Permissions
2115112-SHAIKH RUKSANA KHATUN	rukhsana.k@somaiya.edu	Get, List, Update, Create, Import, Delete, Recover, Backup, Re...	Get, List, Set, Delete, Recover, Backup, Resto...	Get, List, Update, Create, Import, Delete, Recover, Backup, Resto...

Previous Next Review + create Give feedback

The user is finalizing the configuration and clicking "**Create**" to deploy a new Azure Key Vault with specified settings and access policies.



Basics

Subscription: Azure for Students
Resource group: NexusProject
Key vault name: neekukeyvault
Region: Central India
Pricing tier: Standard
Soft delete: Enabled
Purge protection during retention period: Disabled
Days to retain deleted vaults: 90 days

Access configuration

Azure Virtual Machines for deployment: Disabled
Azure Resource Manager for template deployment: Disabled
Azure Disk Encryption for volume encryption: Disabled
Permission model: Vault access policy
Access policies: 1

Networking

Connectivity method: Public endpoint (all networks)

Previous Next Create Give feedback

Nexus DataLens

The deployment of the Azure Key Vault named "**nexuskeyvault**" is currently in progress within the "**NexusProject**" resource group.

This screenshot shows the Microsoft Azure portal interface for a deployment named "nexuskeyvault". The deployment status is listed as "Deployment is in progress". Deployment details include a name of "nexuskeyvault", a subscription of "Azure for Students", and a resource group of "NexusProject". The start time was 4/9/2025, 5:37:03 PM, and the Correlation ID is 1b0046ce-5b5a-4918-83b0-a7bf2a4cd4a2. A message indicates "There are no resources to display". On the right side, there are promotional links for Microsoft Defender for Cloud, Free Microsoft tutorials, and Work with an expert.

The deployment of the "**nexuskeyvault**" in the "**NexusProject**" resource group has completed successfully

This screenshot shows the Microsoft Azure portal interface for the same deployment, now marked as "Deployment succeeded". The deployment status is now "Deployment 'nexuskeyvault' to resource group 'NexusProject' was successful". The deployment details remain the same. A message indicates "Your deployment is complete". On the right side, there are promotional links for Cost management, Microsoft Defender for Cloud, Free Microsoft tutorials, and Work with an expert.

Nexus DataLens

The Azure portal home screen shows the recently accessed resources, including the newly created Key Vault "nexuskeyvault" under the "NexusProject" resource group.

The screenshot shows the Microsoft Azure portal home screen. In the 'Recently' accessed resources section, there are four items listed:

Name	Type	Last Viewed
nexuskeyvault	Key vault	a few seconds ago
NeusProject	Resource group	a few seconds ago
nexusdatabricks	Azure Databricks Service	4 minutes ago
Azure for Students	Subscription	16 hours ago

The "NexusProject" resource group in Azure contains four resources: a Databricks service, a storage account, a key vault, and a Synapse workspace, all located in Central India.

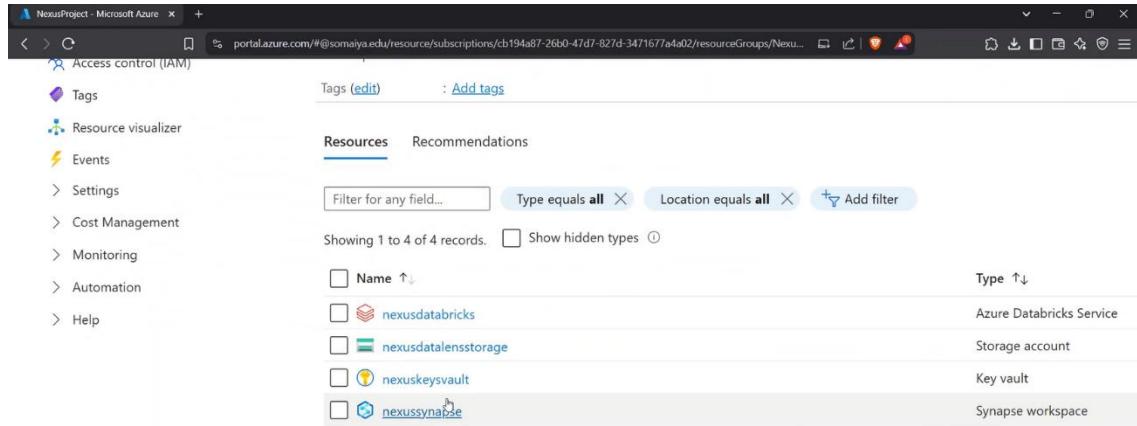
The screenshot shows the 'Overview' page for the 'NexusProject' resource group. The page displays the following information:

- Essentials:** Subscription (Owner) : Azure for Students, Subscription ID : cb194a87-26b0-47d7-827d-3471677a4a02, Tags (edit) : Add tags, Deployments : 4 succeeded, Location : Central India.
- Resources:** A table listing four resources:

Name	Type	Location	Actions
nexusdatabricks	Azure Databricks Service	Central India	...
nexusdatalakestorage	Storage account	Central India	...
nexuskeyvault	Key vault	Central India	...
nexussynapse	Synapse workspace	Central India	...
- Filtering:** Filter for any field... Type equals all, Location equals all, Add filter.
- Pagination:** Showing 1 to 4 of 4 records, Display auto, Page 1 of 1, Next >, Previous <.

Nexus DataLens

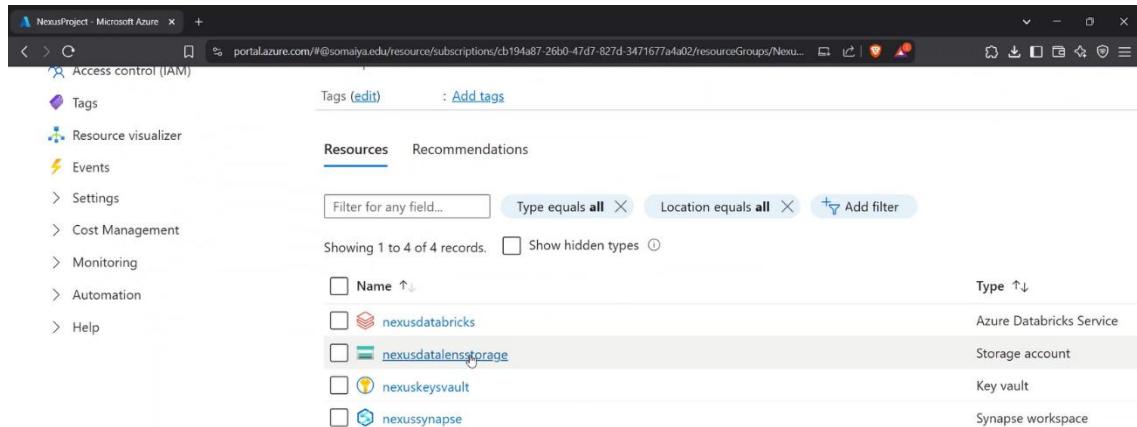
The Azure portal displaying four resources under the "NexusProject" resource group, including Databricks, a storage account, a key vault, and a Synapse workspace.



The screenshot shows the Azure portal interface for the "NexusProject" resource group. On the left, a sidebar lists various monitoring and management options like Access control (IAM), Tags, Resource visualizer, Events, Settings, Cost Management, Monitoring, Automation, and Help. The main area is titled "Resources" and displays four items:

Name	Type
nexusdatabricks	Azure Databricks Service
nexusdatalensstorage	Storage account
nexuskeyvault	Key vault
nexussynapse	Synapse workspace

The Azure portal displays four resources in the "NexusProject" resource group: Databricks, a storage account, a key vault, and a Synapse workspace.



This screenshot is identical to the one above, showing the Azure portal interface for the "NexusProject" resource group. It displays the same four resources: Azure Databricks Service, Storage account, Key vault, and Synapse workspace, listed in a table format.

Nexus DataLens

The Azure Marketplace with search results for "Data Factory" ,displaying various related services and solutions available for deployment.

This screenshot shows the Microsoft Azure Marketplace search results for 'data factory'. The search bar at the top contains 'data factory'. Below the search bar, there are filters: Pricing : All, Operating System : All, Publisher Type : All, Product Type : All, and Publisher name : All. A checkbox for 'Azure services only' is unchecked. The results show 1 to 20 of 111 results for 'data factory'. The results are displayed in a grid of cards:

- Data Factory** by Microsoft: Hybrid data integration service that simplifies ETL at scale. Starts at £1,744.10/month. Create button.
- Data Factory Monitoring Dashboard** by In2Intel: A lightweight web-based dashboard for real-time monitoring and analytics of your Azure Data factories. Starts at £1,744.10/month. Create button.
- Modern Data Mart** by Ceteris AG: Small Data-Warehouse Architecture integrating a single source using Azure Data Factory. Starts at €99.441.615/1 year. Create button.
- On-demand USBA4SAP add-on for Azure Data Factory for SAP** by Ecosever Inc: With USBA4SAP+Azure, access SAP raw data eg tables with delta and modified information. Starts at €253,001.250/month. Subscribe button.
- CloudAtlas AI Factory** by UnifyCloud LLC: Accelerates AI integration with 200+ use cases for proof of concept development in just days. Starts at €253,001.250/month. Subscribe button.
- Astadia FastTrack Factory** by Astadia: Astadia's Mainframe Migration FastTrack Factory is a set of automated refactoring and testing tools. Subscribe button.
- Data#3 Azure Optimiser** by Data#3 Limited: Data Optimiser is an all-in-one solution to optimise usage and prevent unnecessary overspending. Create button.
- 9A Connected Factory & Insights** by Sight Machine, Inc.: A native application to connect your factory and get real-time visibility. Create button.
- Cluedin MDM and Data Quality - PaaS** by Cluedin: Cluedin is a native solution for Master Data Management and Data Quality. Starts at €253,001.250/month. Create button.
- Cluedin MDM and Data Quality - SaaS** by Cluedin: Cluedin is a native solution for Master Data Management and Data Quality. Starts at €253,001.250/month. Create button.
- Altizon Digital Factory (DFX) Platform** by Altizon Inc: Altizon's AI platform to leverage machine data for driving enterprise supply chain efficiency. Starts at €253,001.250/month. Create button.
- Digital Factory | Scheduler** by Profuse: Profuse SaaS Enterprise Master Data Management. Starts at €253,001.250/month. Create button.

The "Create Data Factory" setup page in Azure, where the user is configuring a new Data Factory named "nexusdatainsADF" under the "NexusProject" resource group in the Central India region.

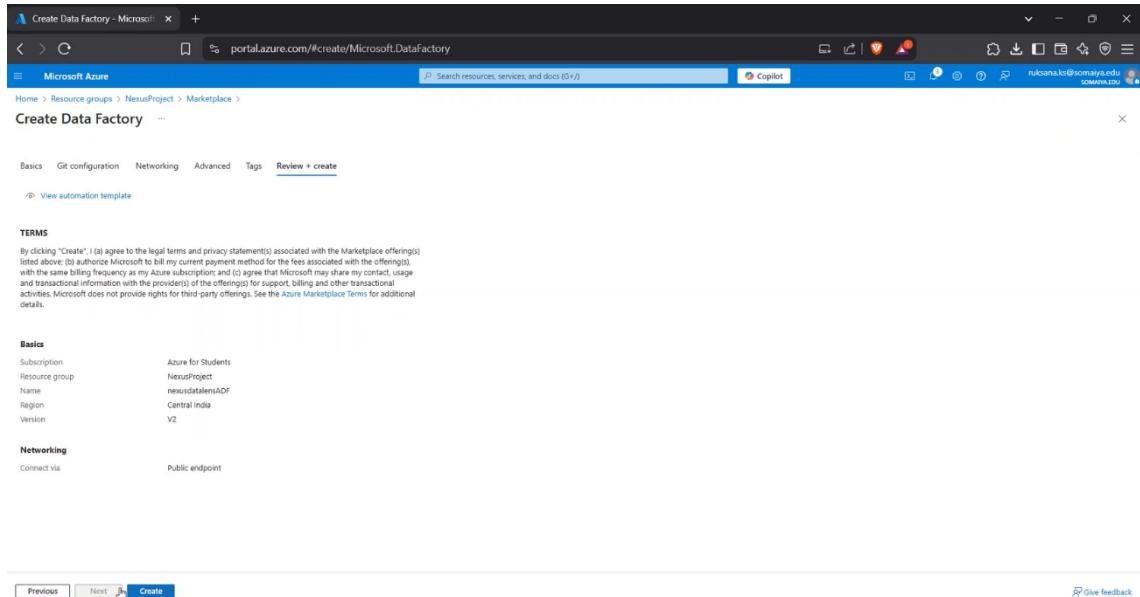
This screenshot shows the 'Create Data Factory' setup page in the Azure portal. The page title is 'Create Data Factory - Microsoft'. The URL is 'portal.azure.com/#create/Microsoft.DataFactory'. The page header includes 'Microsoft Azure', 'Search resources, services, and docs (G+)', 'Copilot', and a user account 'rukshanak@somayya.edu SOMAYYA.EDU'. The main content area is titled 'Create Data Factory' and has tabs: Basics, Git configuration, Networking, Advanced, Tags, Review + create. The 'Basics' tab is selected. It shows the following configuration:

- Subscription ***: Azure for Students
- Resource group ***: NexusProject (Create new)
- Name ***: nexusdatainsADF
- Region ***: Central India
- Version ***: V2

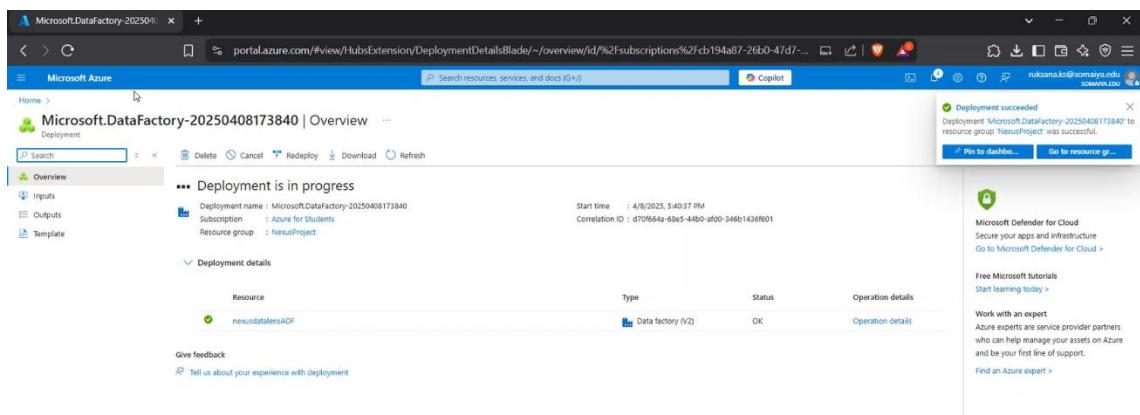
At the bottom, there are buttons: Previous, Next, Review + create, and Give feedback.

Nexus DataLens

The image shows the final "Review + create" step in Azure for deploying a Data Factory instance named "**nexusdatainsADF**" under the "**NexusProject**" resource group using a public endpoint in the Central India region.



The image shows a successful deployment notification for the Data Factory resource "**nexusdatainsADF**" under the "**NexusProject**" resource group in Azure.



Nexus DataLens

The image shows the Azure portal listing four resource groups under the "Azure for Students" subscription, all located in the Central India region.

Name	Subscription	Location
databricks-rg	Azure for Students	Central India
NetworkWatcherRG	Azure for Students	Central India
NexusProject	Azure for Students	Central India
synapsespace-managedrg-d5	Azure for Students	Central India

Complete setup of resource group

Name	Type	Location
neusdatabricks	Azure Databricks Service	Central India
neusdatabricksADF	Data factory (V2)	Central India
neusdatadlstorage	Storage account	Central India
neusdkyvault	Key vault	Central India
neusynapse	Synapse workspace	Central India

SETTING RESOURCE GROUP

The "NexusProject" resource group currently has no visible resources due to active filters, with an option to create new resources or clear filters.

NexusProject - Microsoft Azure

Microsoft Azure

NexusProject

Resource group

Overview

Subscription (most recent): Azure for Students

Subscription ID: cb194a87-26b0-47d7-827d-3471677a4a02

Tags (edit): Add tags

Resources Recommendations

No resources match your filters

+ Create resources | Clear filters

The Azure Marketplace displays various Synapse Analytics-related services and solutions available for creation or subscription within the "NexusProject" environment.

Marketplace - Microsoft Azure

Microsoft Azure

Marketplace

Get Started

Service Providers

Management

Private Marketplace

Private Offer Management

My Marketplace

Favorites

My solutions

Recently created

Private plans

Categories

Analytics (34)

Databases (32)

AI & Machine Learning (25)

T & M Management Tools (13)

Storage (9)

Compute (8)

Developer Tools (5)

Security (4)

Internet of Things (2)

synapse analytics

Pricing: All

Operating System: All

Publisher Type: All

Product Type: All

Publisher name: All

Show 1 to 20 of 102 results for 'synapse analytics'

Azure Synapse Analytics

Azure Synapse Analytics (private link hubs)

Synapse Data Fabric

Datometry Hyper-Q for Azure Synapse Analytics

Mayo Azure Synapse Retail Recommender Solution

Xpert BI with Azure Synapse

The form is being filled to create a new Synapse workspace under the "NexusProject" resource group with specified subscription, region, and linked Data Lake Storage Gen2 account.

Nexus DataLens

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 * 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 * Select Data Lake Storage Gen2 * From subscription Manually via URL Account name * File system name *

A Synapse workspace named "**nexussynapse**" is being created in Central India, linked to the "**nexusdatalensstorage**" Data Lake Storage with "**bronze**" as the file system.

Microsoft Azure Copilot

Home > NexusProject > Marketplace > Create Synapse workspace

Resource group * 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 * Select Data Lake Storage Gen2 * From subscription Manually via URL Account name * File system name * Assign myself the Storage Blob Data Contributor role on the Data Lake Storage Gen2 account to interactively query it in the workspace.

We will automatically grant the workspace identity data access to the specified storage account and assign the Storage Blob Data Contributor role. To enable other users to use the storage account after you create your workspace, perform these tasks:

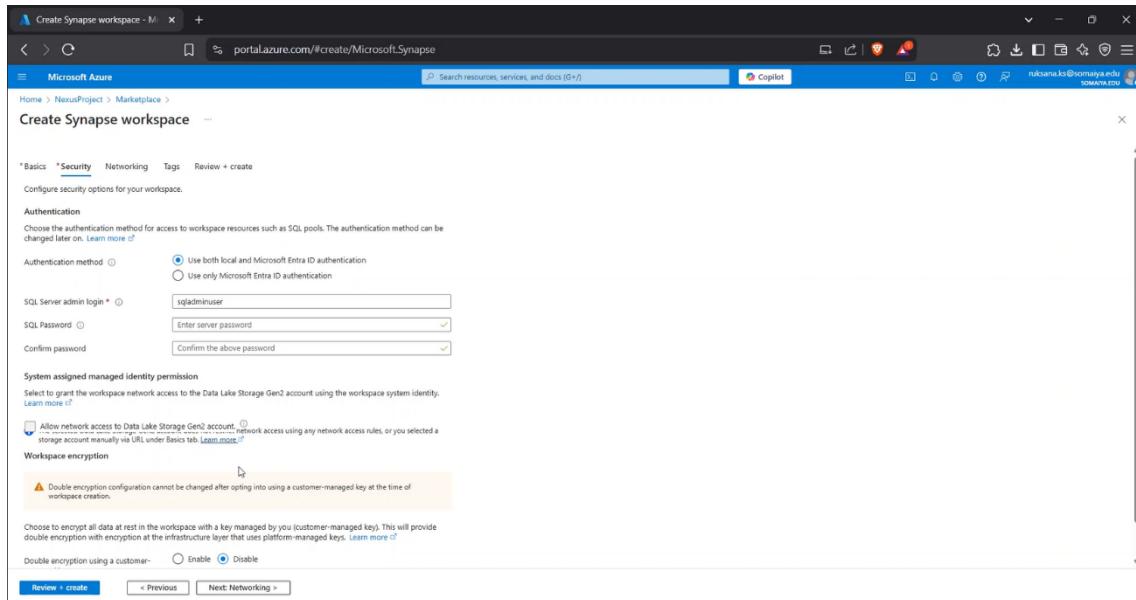
- Assign other users to the **Contributor** role on the workspace
- Assign other users the appropriate **Synapse Data Reader** role using Synapse Studio
- Assign yourself and other users to the **Storage Blob Data Contributor** role on the storage account

[Learn more](#)

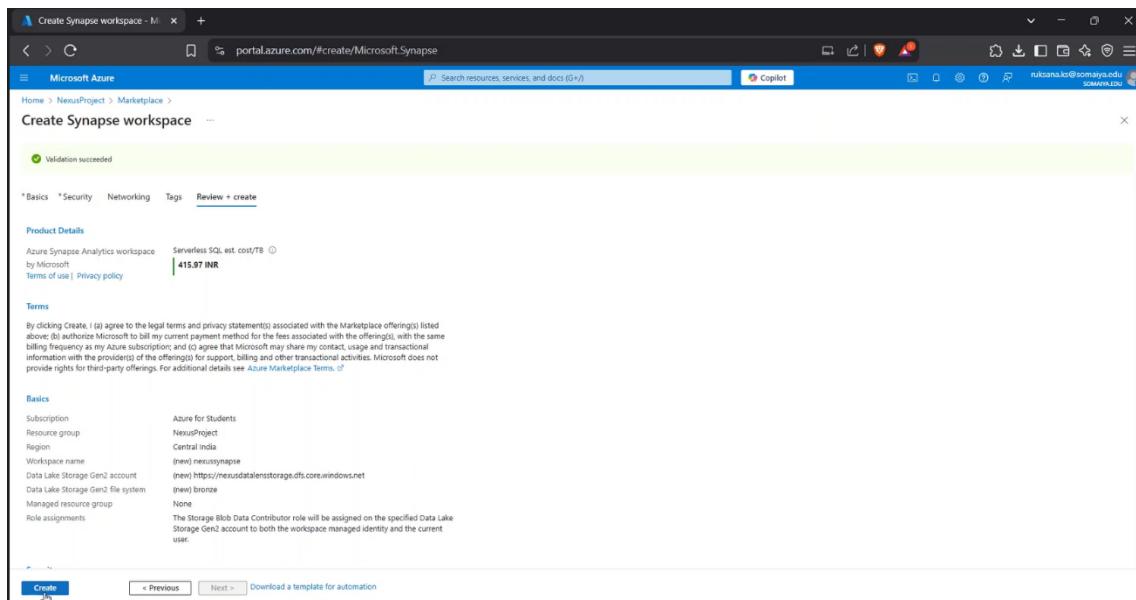
[Review + create](#) [Next: Security](#)

Nexus DataLens

The Synapse workspace is being configured with SQL authentication, network access to Data Lake enabled, and default platform-managed encryption selected.

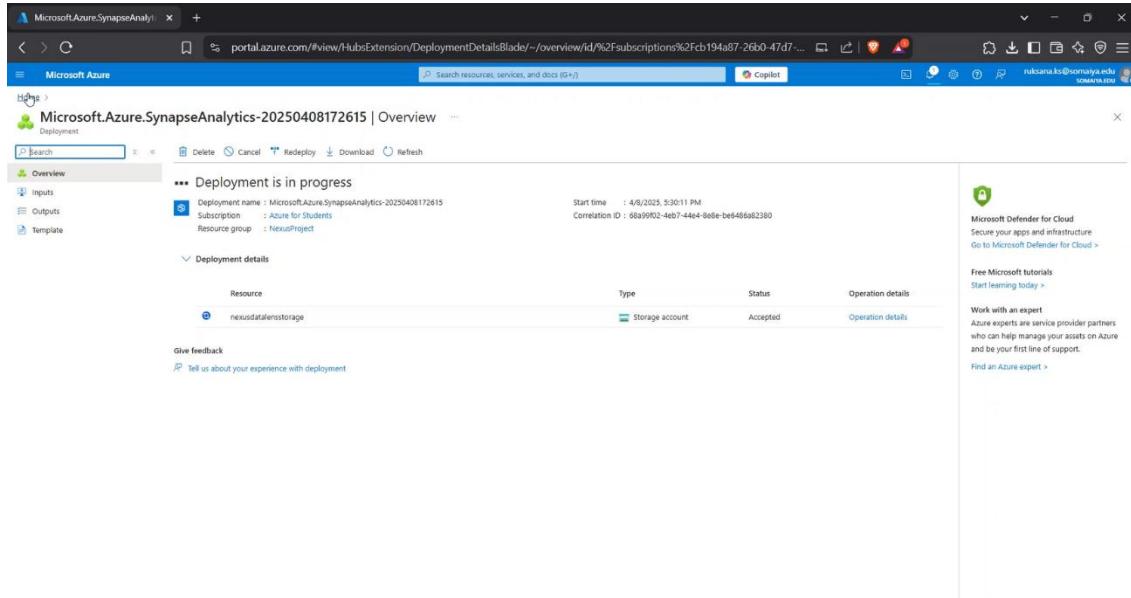


The **Synapse workspace setup review** has succeeded, and the user is ready to create it with specified configuration and pricing.



Nexus DataLens

The Synapse workspace deployment is currently in progress with the storage account resource accepted under the **NexusProject** resource group.



Microsoft.Azure.SynapseAnalytics-20250408172615 | Overview

Deployment

Deployment is in progress

Deployment name: Microsoft.Azure.SynapseAnalytics-20250408172615

Subscription: Azure for Students

Resource group: NexusProject

Start time: 4/8/2025, 5:30:11 PM

Correlation ID: 6ba99f02-4eb7-44e4-8e8e-beb486a82380

Resource Type Status Operation details

nexusdatalensstorage Storage account Accepted Operation details

Give feedback

Tell us about your experience with deployment

Microsoft Defender for Cloud

Secure your apps and infrastructure

Go to Microsoft Defender for Cloud >

Free Microsoft tutorials

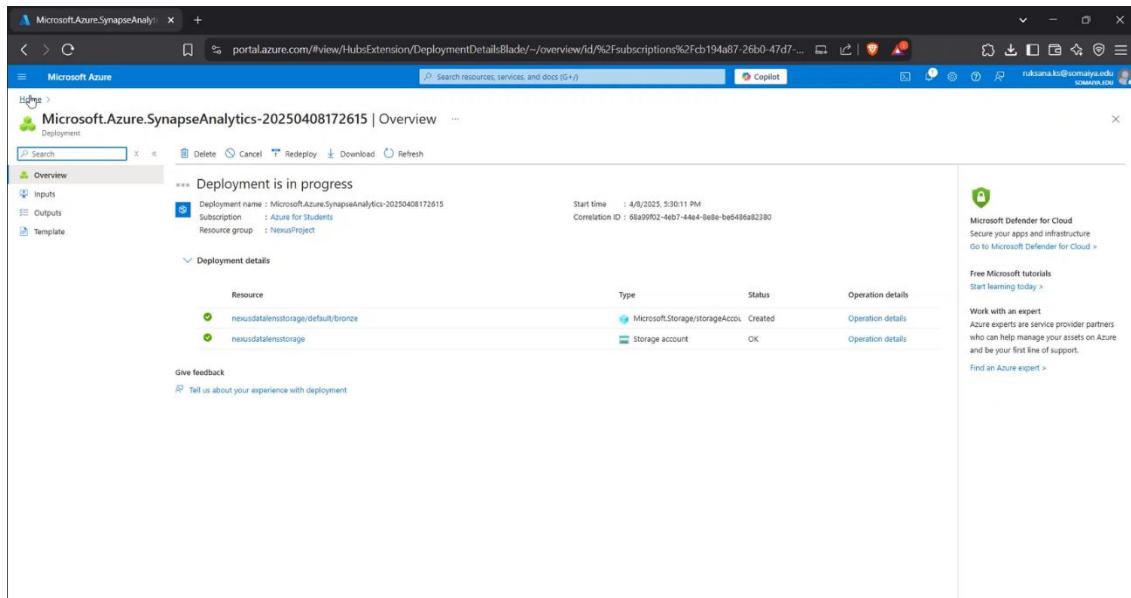
Start learning today >

Work with an expert

Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support.

Find an Azure expert >

The Synapse workspace deployment is in **progress**, with the storage container bronze created and the storage account status marked as OK.



Microsoft.Azure.SynapseAnalytics-20250408172615 | Overview

Deployment

Deployment is in progress

Deployment name: Microsoft.Azure.SynapseAnalytics-20250408172615

Subscription: Azure for Students

Resource group: NexusProject

Start time: 4/8/2025, 5:30:11 PM

Correlation ID: 6ba99f02-4eb7-44e4-8e8e-beb486a82380

Resource Type Status Operation details

nexusdatalensstorage/default/bronze Microsoft.Storage/storageAccount Created Operation details

nexusdatalensstorage Storage account OK Operation details

Give feedback

Tell us about your experience with deployment

Microsoft Defender for Cloud

Secure your apps and infrastructure

Go to Microsoft Defender for Cloud >

Free Microsoft tutorials

Start learning today >

Work with an expert

Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support.

Find an Azure expert >

Nexus DataLens

The Azure portal home page displays recent resources, with "NexusProject" (a resource group) being the most recently accessed.

The screenshot shows the Microsoft Azure portal home page. At the top, there's a navigation bar with icons for Home, Search, Copilot, and user information. Below the bar, the "Azure services" section features a grid of icons for Create a resource, Resource groups, Subscriptions, Azure Synapse Analytics, Budgets, Bot Services, Projects, Quickstart Center, Azure AI services, and More services. The "Resources" section displays a table of recent resources, with "NexusProject" listed as a Resource group last viewed 7 minutes ago. The "Navigate" section includes links for Subscriptions, Resource groups, All resources, and Dashboard. The "Tools" section lists Microsoft Learn, Azure Monitor, Microsoft Defender for Cloud, and Cost Management. The "Useful links" section provides links to Technical Documentation, Azure Services, Recent Azure Updates, and mobile app download links for App Store and Google Play. The "Azure mobile app" section shows download links for both platforms.

The image shows an Azure portal page displaying the "NexusProject" resource group with one storage account resource named "nexusdatastenstorage" located in Central India.

The screenshot shows the Microsoft Azure portal page for the "NexusProject" resource group. The left sidebar lists options like Overview, Activity log, Access control (IAM), Tags, Resource visualizer, Events, Settings, Cost Management, Monitoring, Automation, and Help. The main content area shows the "Overview" tab for the resource group. It displays the subscription (Azure for Students), deployment count (1 Deploying), and location (Central India). A table lists the resources in the group, showing one entry: "nexusdatastenstorage" (Storage account) located in Central India. The bottom of the page includes navigation controls for previous, next, and search.

Nexus DataLens

The updated Azure portal view shows the "NexusProject" resource group now containing two resources: a storage account ("nexusdatastenstorage") and a Synapse workspace ("nexussynapse"), both located in Central India.

The screenshot shows the Microsoft Azure portal interface. The top navigation bar includes the title "NexusProject - Microsoft Azure", the URL "portal.azure.com/#@somaiya.edu/resource/subscriptions/ch194a87-26b0-47d7-827d-3471677a4a02/resourceGroups/NexusProject", and the user "rukshanika@somaiya.edu somaiya.edu". The main content area displays the "Resource group" page for "NexusProject". The "Essentials" section shows the subscription information: "Subscription (Owner) : Azure for Students", "Subscription ID : ch194a87-26b0-47d7-827d-3471677a4a02", and "Location : Central India". Below this, the "Resources" section lists two resources: "nexusdatastenstorage" (Storage account, Central India) and "nexussynapse" (Synapse workspace, Central India). The left sidebar contains links for Activity log, Access control (IAM), Tags, Resource visualizer, Events, Settings, Cost Management, Monitoring, Automation, and Help.

The Azure Marketplace displays various Databricks-related services, including "**Azure Databricks**" by Microsoft for Spark-based analytics and other integration and optimization tools.

The screenshot shows the Microsoft Azure Marketplace search results for "databricks". The search bar at the top contains the query "databricks". The results are displayed in a grid format, showing 20 out of 73 total results. Each result card includes the service name, publisher, price, and a brief description. The first result is "Azure Databricks" by Microsoft, described as "Azure Databricks is the fast, easy and collaborative Apache Spark-based analytics platform." Other visible results include "Access Connector for Azure Databricks" by Microsoft, "Lakehouse Optimizer for Azure Databricks" by Blueprint Technologies LLC, "Strim for Databricks" by Strim, Inc., "Unravel for Azure Databricks (PayGo)" by Unravel Data, and "EZConvertETL for Azure Databricks" by Wavicle Data Solut.

Nexus DataLens

The user is about to create an "**Azure Databricks**" workspace from the Azure Marketplace by clicking the "**Create**" button under the Azure Databricks service.

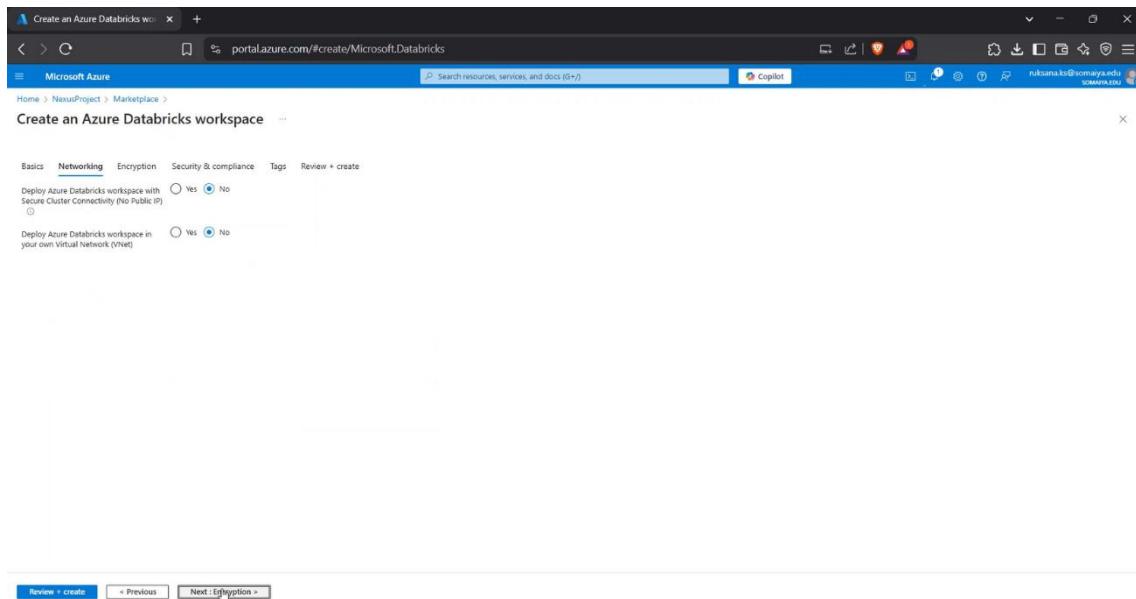
A screenshot of the Microsoft Azure Marketplace search results for 'databricks'. The search bar at the top shows 'databricks'. Below it, a message says 'New! Get AI-generated suggestions for 'databricks'' with a 'View suggestions' button. The results section shows 1 to 20 of 73 items. The first item is 'Azure Databricks' by Microsoft, described as an 'Azure Service'. It has a 'Create' button. The second item is 'Access Connector for Azure Databricks' by Microsoft, also an 'Azure Service'. The third item is 'Lakehouse Optimizer for Azure Databricks' by Blueprint Technologies LLC, an 'Azure Application'. The fourth item is 'Striim for Databricks' by Striim, Inc., a 'SaaS' application. The fifth item is 'Unravel for Azure Databricks (PayGo)' by Unravel Data, a 'SaaS' application. The sixth item is 'EZConvertETL I Azure Databricks' by Wavecle Data Solut, a 'SaaS' application. Other items like 'InfoWorks' and 'Sela' are also listed. The interface includes filters for 'Azure services only' and various sorting and search options.

The user is configuring the basic settings to create an Azure Databricks workspace named "**nexusdatabricks**" in the "**NexusProject**" resource group within the Central India region.

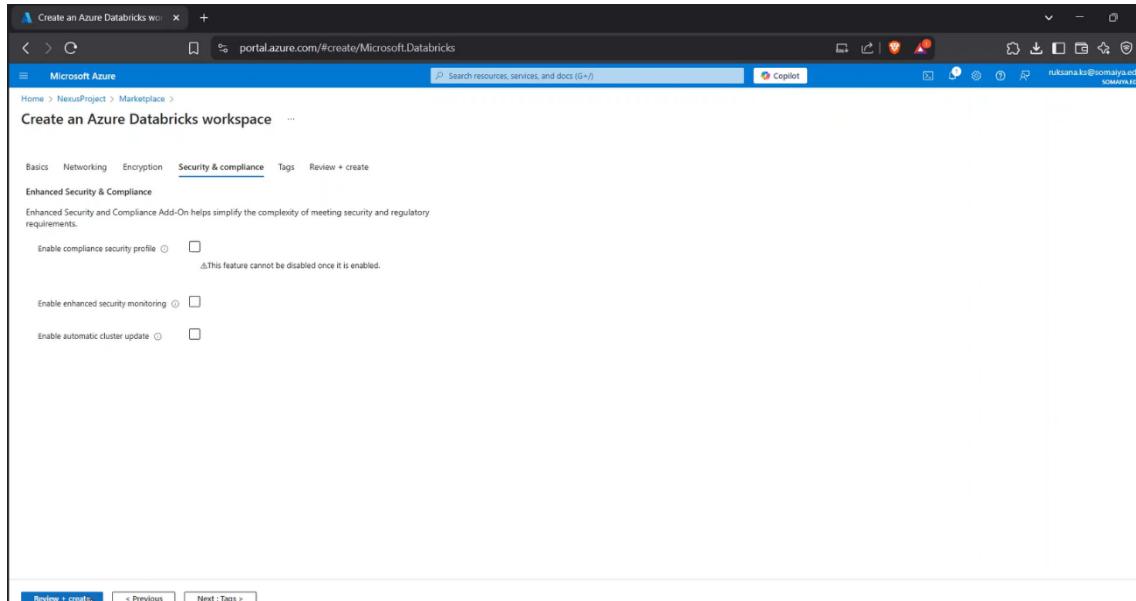
A screenshot of the 'Create an Azure Databricks workspace' wizard on the Azure portal. The page title is 'Create an Azure Databricks workspace'. The 'Basics' tab is selected. Under 'Project Details', the 'Subscription' dropdown is set to 'Azure for Students' and the 'Resource group' dropdown is set to 'NexusProject'. Under 'Instance Details', the 'Workspace name' is 'nexusdatabricks', 'Region' is 'Central India', and 'Pricing Tier' is 'Premium (+ Role-based access controls)'. The 'Managed Resource Group name' field is empty. At the bottom, there are navigation buttons: 'Review + create' (disabled), '< Previous', and 'Next : Networking >'.

Nexus DataLens

The user is configuring networking options for the Azure Databricks workspace, choosing whether to enable secure cluster connectivity and virtual network deployment.



The user is on the Security & Compliance tab while creating an Azure Databricks workspace, with optional settings for compliance, monitoring, and automatic cluster updates before proceeding to review and create.



Nexus DataLens

The user is on the final "**Review + create**" step for setting up an Azure Databricks workspace, reviewing all configuration details before deployment.

This screenshot shows the 'Create an Azure Databricks workspace' review + create step in the Azure portal. The page displays configuration details across several tabs: Basics, Networking, Encryption, Security & compliance, Tags, and Review + create. The 'Review + create' tab is selected. The configuration includes:

- Basics:** Workspace name: nexusdatabricks, Subscription: Azure for Students, Resource group: NexusProject, Region: Central India, Pricing Tier: premium, Managed Resource Group name: (empty).
- Networking:** Deploy Azure Databricks workspace with Secure Cluster Connectivity (No Public IP): No, Deploy Azure Databricks workspace in your own Virtual Network (VNet): No.
- Encryption:** Enable Infrastructure Encryption: No, Enable CMK for Managed Disks: No, Enable CMK for Managed Services: No.
- Security & compliance:** Compliance Security Profile: No.

At the bottom, there are 'Create' and 'Go Previous' buttons. The 'Create' button is highlighted with a mouse cursor.

The user has successfully validated the configuration and is ready to create the **Azure Databricks workspace** by clicking the "**Create**" button.

This screenshot shows the 'Create an Azure Databricks workspace' review + create step in the Azure portal, similar to the previous one but with a green validation bar at the top indicating 'Validation Succeeded'. The configuration details remain the same as in the previous screenshot. The 'Create' button is now highlighted with a blue background and white text, indicating it is the active button.

Nexus DataLens

The deployment of the Azure Databricks workspace named "**nexusdatabricks**" is currently in progress within the "**NexusProject**" resource group.

The screenshot shows the Azure portal's deployment details for the 'NexusProject_nexusdatabricks' workspace. The deployment is currently in progress. Key information includes:

- Deployment name:** NexusProject_nexusdatabricks
- Subscription:** Azure for Students
- Resource group:** NexusProject
- Start time:** 4/8/2025 5:33:23 PM
- Correlation ID:** b9e9249e-12b4-4440-a5d6-c28909ff2d2f

The 'Deployment details' table lists the resource 'nexusdatabricks' with a Type of 'Azure Databricks Service', Status of 'Created', and Operation details showing 'Operation details'.

On the Azure portal home screen click on the "**Resource groups**" option to view or **create a new resource group**.

The screenshot shows the Azure portal's home screen with the 'Resource groups' option highlighted in the top navigation bar. Below the navigation bar, there are sections for 'Azure services' (with 'Create a resource' and 'Resource groups' buttons), 'Resources' (listing 'nexusdatabricks', 'NexusProject', and 'Azure for Students'), 'Navigate' (links to 'Subscriptions', 'Resource groups', 'All resources', and 'Dashboard'), 'Tools' (links to 'Microsoft Learn', 'Azure Monitor', 'Microsoft Defender for Cloud', and 'Cost Management'), and 'Useful links' (links to 'Azure mobile app' and other resources).

Nexus DataLens

The "**Resource groups**" page in the Azure portal displays existing resource groups under the "**Azure for Students**" subscription, with both located in Central India.

The screenshot shows the Microsoft Azure Resource groups page. The URL is <https://portal.azure.com/#browse/resourcegroups>. The page title is "Resource groups". There are two resource groups listed:

Name	Subscription	Location
NexusProject	Azure for Students	Central India
synapseworkspace-managedrg-d57a18af-12f1-42ff-965a-4ae35a8c23e5	Azure for Students	Central India

At the bottom left, it says "Showing 1 - 2 of 2. Display count: auto". At the bottom right, there is a "Give feedback" link.

The user selects the "**NexusProject**" resource group from the list of available groups under the "**Azure for Students**" subscription in the Central India region.

The screenshot shows the Microsoft Azure Resource groups page. The URL is <https://portal.azure.com/#browse/resourcegroups>. The page title is "Resource groups". The "NexusProject" resource group is highlighted with a blue selection bar. The other resource group, "synapseworkspace-managedrg-d57a18af-12f1-42ff-965a-4ae35a8c23e5", is shown below it.

Nexus DataLens

The user searches for "**Key Vault**" in the Azure Marketplace under the **NexusProject** resource group to **create a new key vault service**.

A screenshot of the Microsoft Azure Marketplace search results. The search bar at the top contains the text "key vault". Below the search bar, there are several filters: "Pricing: All", "Operating System: All", "Publisher Type: All", "Product Type: All", and "Publisher name: All". A checkbox labeled "Azure services only" is checked. A search suggestion "New! Get AI-generated suggestions for 'keyvault'" is displayed. The results show five items:

- Key Vault** (Microsoft Azure Service): Described as "Safeguard cryptographic keys and other secrets used by cloud apps and services." It has a "Create" button.
- CommVault Cloud for Sentinel** (CommVault Azure Application): Described as "Enables CommVault users to ingest alerts and other data into their Sentinel instance." It has a "Create" button.
- AZ-204 Workshop | Developing Solutions for Opisility, LLC** (Opisility, LLC Azure Application): Described as "5-day workshop with OneDrive and Cloud Sandbox focused on Azure Development." It has a "Subscribe" button.
- HSM Ingress Controller** (Strd Tech Container (Kubernetes App)): Described as "Ingress that integrates with AKV or mTLS to handle TLS offload in AKS for FIPS 140-3 compliance." It has a "Create" button.
- Unified Contacts Pro – Companion to Microsoft** (glaesckanje AG SeoS): Described as "Finding ALL your Contacts in Teams - Azure Active Directory, Exchange Online and SharePoint." It has a "Subscribe" button.

At the bottom of the page, there are navigation buttons: "Previous", "Page 1 of 1", and "Next".

The user is configuring and creating a new Azure Key Vault named "**nexuskeyvault**" under the "**NexusProject**" resource group in the Central India region.

A screenshot of the "Create a key vault" wizard, step 1: "Select resource group". The page title is "Create a key vault" and the URL is "portal.azure.com/#create/Microsoft.KeyVault". The left sidebar shows the user's profile and recent activity. The main form has the following fields:

- Subscription ***: Azure for Students
- Resource group ***: NexusProject (with a "Create new" option)
- Instance details**
 - Key vault name ***: nexuskeyvault
 - Region ***: Central India
 - Pricing tier ***: Standard
- Recovery options**

Soft delete protection will automatically be enabled on this key vault. This feature allows you to recover or permanently delete a key vault and secrets for the duration of the retention period. This protection applies to the key vault and the secrets stored within the key vault.

To enforce a mandatory retention period and prevent the permanent deletion of key vaults or secrets prior to the retention period elapsing, you can turn on purge protection. When purge protection is enabled, secrets cannot be purged by users or by Microsoft.

Soft delete: Enabled (radio button selected).
Days to retain deleted vaults *: 90.
Purge protection:
 - Disable purge protection (allow key vault and objects to be purged during retention period)
 - Enable purge protection (enforce a mandatory retention period for deleted vaults and vault objects)

At the bottom, there are "Previous", "Next", and "Review + create" buttons.

Nexus DataLens

The user is assigning full key, secret, and certificate permissions to a specified user while setting access policies for the new Azure Key Vault before final review and creation.

Screenshot of the Azure portal showing the 'Create a key vault' wizard step 3: Set access policies. The page displays a table of access policies for a user named '2115112-SHAIKH RUKSANA KHATUN'. The user's email is listed as 'rukksana.k@somaya.edu'. The table includes columns for Name, Email, Key Permissions, Secret Permissions, and Certificate Permissions. The 'Key Permissions' column shows 'Get, List, Update, Create, Import, Delete, Recover, Backup, Restore'. The 'Secret Permissions' and 'Certificate Permissions' columns show 'Get, List, Update, Create, Import, Delete, Recover, Backup, Restore'.

The user is finalizing the configuration and clicking "**Create**" to deploy a new Azure Key Vault with specified settings and access policies.

Screenshot of the Azure portal showing the 'Create a key vault' wizard step 4: Review + create. The 'Review + create' tab is selected. The page displays a summary of the key vault configuration, including the subscription ('Azure for Students'), resource group ('NexusProject'), key vault name ('neuskeyvault'), region ('Central India'), pricing tier ('Standard'), soft delete ('Enabled'), purge protection during retention period ('Disabled'), days to retain deleted vaults ('90 days'), access configuration (disabled for all), networking (public endpoint), and tags (''). The 'Create' button is highlighted at the bottom of the page.

Nexus DataLens

The deployment of the Azure Key Vault named "**nexuskeyvault**" is currently in progress within the "**NexusProject**" resource group.

A screenshot of the Microsoft Azure portal showing the deployment status of a key vault named "nexuskeyvault". The deployment is currently in progress. The deployment details table shows the following information:

Resource	Type	Status	Operation details
There are no resources to display.			

On the right side of the page, there are promotional banners for Microsoft Defender for Cloud, Free Microsoft tutorials, Work with an expert, and Find an Azure expert.

The deployment of the "**nexuskeyvault**" in the "**NexusProject**" resource group has completed successfully

A screenshot of the Microsoft Azure portal showing the deployment status of a key vault named "nexuskeyvault". The deployment is now complete. The deployment details table shows the following information:

Resource	Type	Status	Operation details
Deployment succeeded Deployment 'nexuskeyvault' to resource group 'NexusProject' was successful.			

On the right side of the page, there are promotional banners for Deployment succeeded, Cost management, Microsoft Defender for Cloud, Free Microsoft tutorials, Work with an expert, and Find an Azure expert.

Nexus DataLens

The Azure portal home screen shows the recently accessed resources, including the newly created Key Vault "nexuskeyvault" under the "NexusProject" resource group.

The screenshot shows the Microsoft Azure portal home screen. In the 'Recently accessed' section, there are four items listed:

Name	Type	Last Viewed
nexuskeyvault	Key vault	a few seconds ago
NexusProject	Resource group	a few seconds ago
nexusalldatabricks	Azure Databricks Service	4 minutes ago
Azure for Students	Subscription	16 hours ago

The "NexusProject" resource group in Azure contains four resources: a Databricks service, a storage account, a key vault, and a Synapse workspace, all located in Central India.

The screenshot shows the 'Overview' page for the 'NexusProject' resource group. On the left, there is a sidebar with navigation links like 'Name', 'databricks-rg-nexusalldatabricks-bm', 'NetworkWatcherRG', 'NexusProject', and 'synapseworkspace-managedrg-d5'. The main area displays the following details:

- Subscription (Owner):** Azure for Students
- Subscription ID:** cb194a87-26b0-47d7-827d-3471677a4a02
- Deployments:** 4 Succeeded
- Location:** Central India

Below this, there is a table titled 'Resources' showing the following data:

Name	Type	Location	Actions
nexusalldatabricks	Azure Databricks Service	Central India	...
nexusalldatabricksstorage	Storage account	Central India	...
nexuskeyvault	Key vault	Central India	...
nexussynapse	Synapse workspace	Central India	...

Nexus DataLens

The Azure portal displaying four resources under the "NexusProject" resource group, including Databricks, a storage account, a key vault, and a Synapse workspace.

The screenshot shows the Microsoft Azure portal interface. The left sidebar has a tree view with 'Access control (IAM)' selected. The main area is titled 'Tags (edit)' with a 'Resources' tab selected. There are filters for 'Type equals all' and 'Location equals all'. The results table shows four items:

Name	Type
nexusdatabricks	Azure Databricks Service
nexusdatalensstorage	Storage account
nexuskeyvault	Key vault
nexussynapse	Synapse workspace

The Azure portal displays four resources in the "NexusProject" resource group: Databricks, a storage account, a key vault, and a Synapse workspace.

This screenshot is identical to the one above, showing the Microsoft Azure portal interface with the 'Access control (IAM)' blade for the 'NexusProject' resource group. The results table lists the same four resources: 'nexusdatabricks', 'nexusdatalensstorage', 'nexuskeyvault', and 'nexussynapse', each with its corresponding type listed to the right.

Nexus DataLens

The Azure Marketplace with search results for "Data Factory" ,displaying various related services and solutions available for deployment.

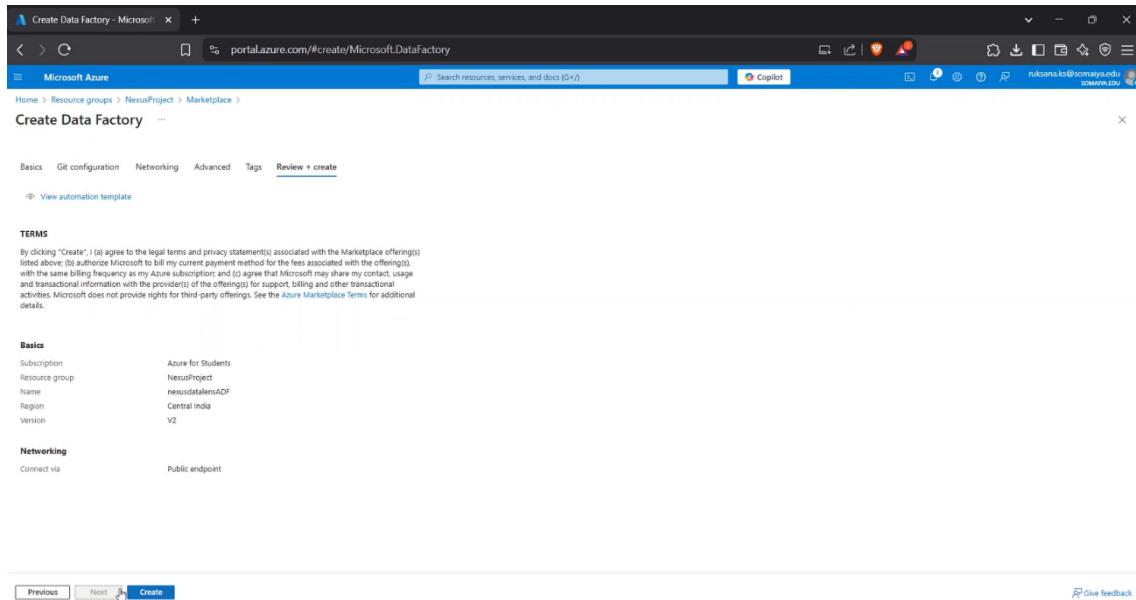
A screenshot of the Microsoft Azure Marketplace interface. The search bar at the top contains the query 'data factory'. Below the search bar, there are several filters: Pricing : All, Operating System : All, Publisher Type : All, Product Type : All, and Publisher name : All. A message box says 'New! Get AI-generated suggestions for "data factory"' with a 'View suggestions' button. The main area shows a grid of 20 results from various publishers. The first result is 'Data Factory' by Microsoft, described as a 'Hybrid data integration service that simplifies ETL at scale'. Other results include 'Data Factory Monitoring Dashboard' by In2Intel, 'Modern Data Mart' by Ceteris AG, 'On-demand USR4SAP add-on to Azure Data Factory for' by Ecossity Inc, 'CloudAtlas AI Factory' by UniCloud LLC, 'Astadia FastTrack Factory' by Astadia, 'Data#3 Azure Optimiser' by Data#3 Limited, 'Sight Machine Factory CONNECT' by Sight Machine, Inc., '9A Connected Factory & Insights' by 9Altitudes, 'CluedIn MDM and Data Quality - PaaS' by CluedIn, 'CluedIn MDM and Data Quality - SaaS' by CluedIn, 'Altizon Digital Factory (DFX) Platform' by Altizon Inc, 'Digital Factory | Scheduler' by Profisee, and 'Profisee SaaS Enterprise Master Data Management' by Profisee. Each listing includes a small icon, the publisher name, the product name, a brief description, and a 'Create' or 'Subscribe' button.

The "Create Data Factory" setup page in Azure, where the user is configuring a new Data Factory named "nexusdatainsADF" under the "NexusProject" resource group in the Central India region.

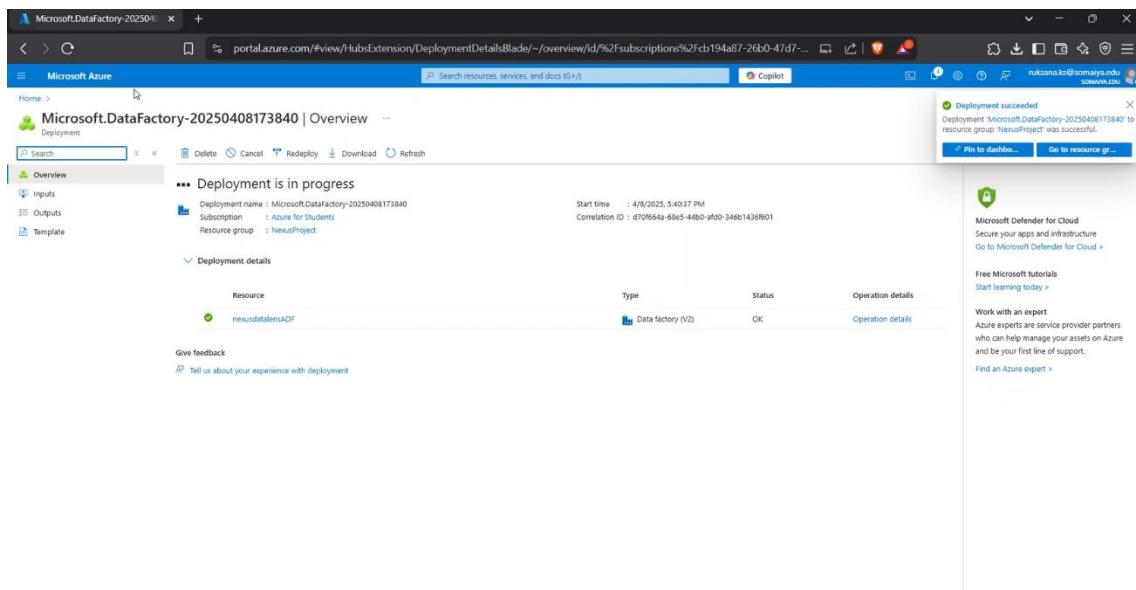
A screenshot of the 'Create Data Factory' setup page in the Azure portal. The top navigation bar shows 'Create Data Factory' and the URL 'portal.azure.com/#create/Microsoft.DataFactory'. The page has tabs for 'Basics', 'Git configuration', 'Networking', 'Advanced', 'Tags', and 'Review + create'. The 'Basics' tab is selected. It shows a summary: 'One-click to create data factory with sample pipeline and datasets. Try it'. Below this, the 'Project details' section asks to 'Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.' It shows 'Subscription' set to 'Azure for Students' and 'Resource group' set to 'NexusProject'. The 'Instance details' section includes fields for 'Name' (set to 'nexusdatalensADF'), 'Region' (set to 'Central India'), and 'Version' (set to 'V2'). At the bottom, there are buttons for 'Previous', 'Next', 'Review + create', and 'Give feedback'.

Nexus DataLens

The image shows the final "Review + create" step in Azure for deploying a Data Factory instance named "**nexusdatainsADF**" under the "**NexusProject**" resource group using a public endpoint in the Central India region.



The image shows a successful deployment notification for the Data Factory resource "**nexusdatainsADF**" under the "**NexusProject**" resource group in Azure.



Nexus DataLens

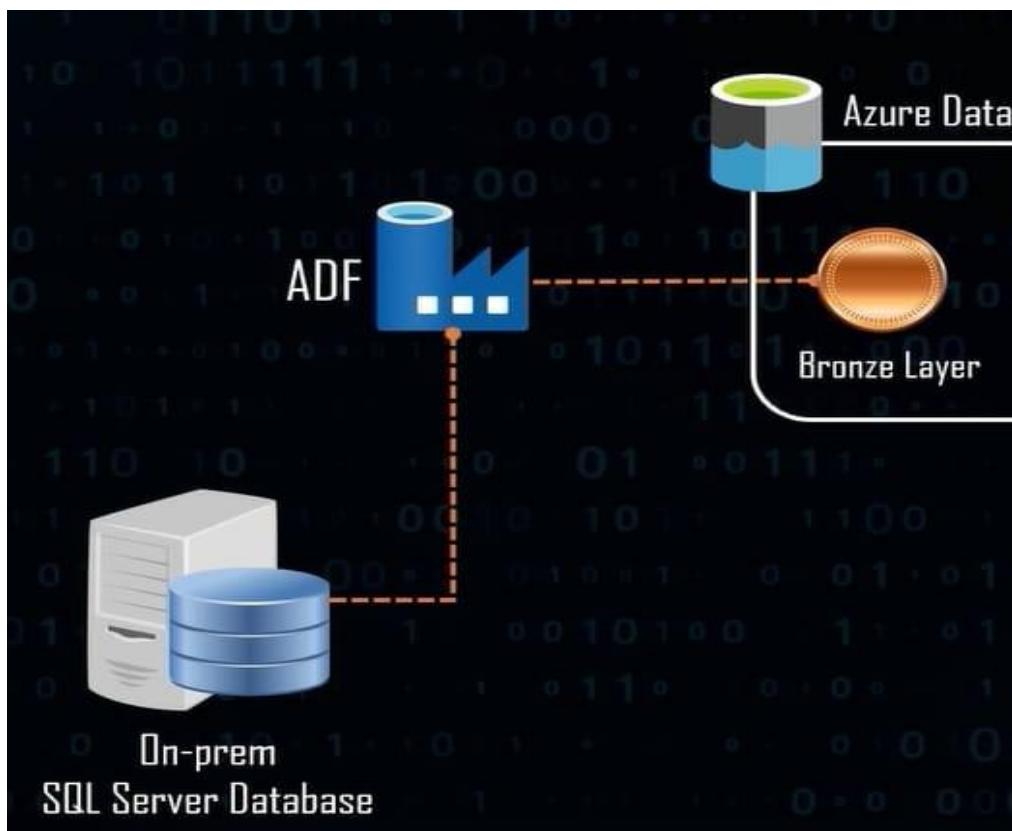
The image shows the Azure portal listing four resource groups under the "Azure for Students" subscription, all located in the Central India region.

The screenshot displays the Azure portal's Resource Groups page. At the top, there are navigation links for Home, Microsoft Azure, and somaya.edu. Below the header, a message indicates that a new version of the Browse experience is available. The main area shows a table of resource groups:

Name	Subscription	Location
databricks-rg	Azure for Students	Central India
NetworkWatcherRG	Azure for Students	Central India
nexusPoint	Azure for Students	Central India
synapsaworkspace-manage-rg	Azure for Students	Central India

At the bottom of the page, it says "Showing 1 - 4 of 4. Display count: auto".

DATA INJECTION - Using SQL Server Management Studio (SSMS)



Open SQL Server Management Studio (SSMS)

Open Azure Portal

- Go to <https://portal.azure.com>.
- Sign in with your Azure credentials.

Navigate to Azure Data Factory

- In the Azure portal search bar at the top, type Data Factories.
- Click on Data Factories from the search results.

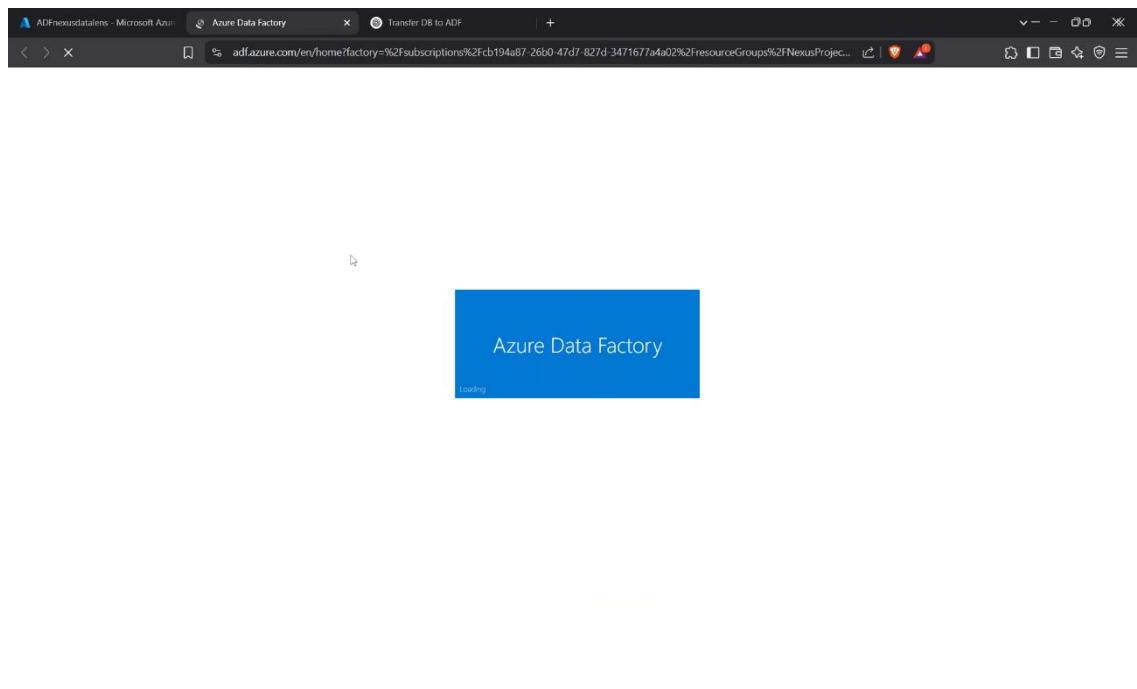
Select Your Data Factory Instance

- In the list of existing Data Factories, select your Data Factory instance (in this case, probably one associated with NexusProject as seen in the URL).

Launch Azure Data Factory Studio

- After selecting the instance, click the 'Open Azure Data Factory Studio' button from the resource overview page.

Nexus DataLens



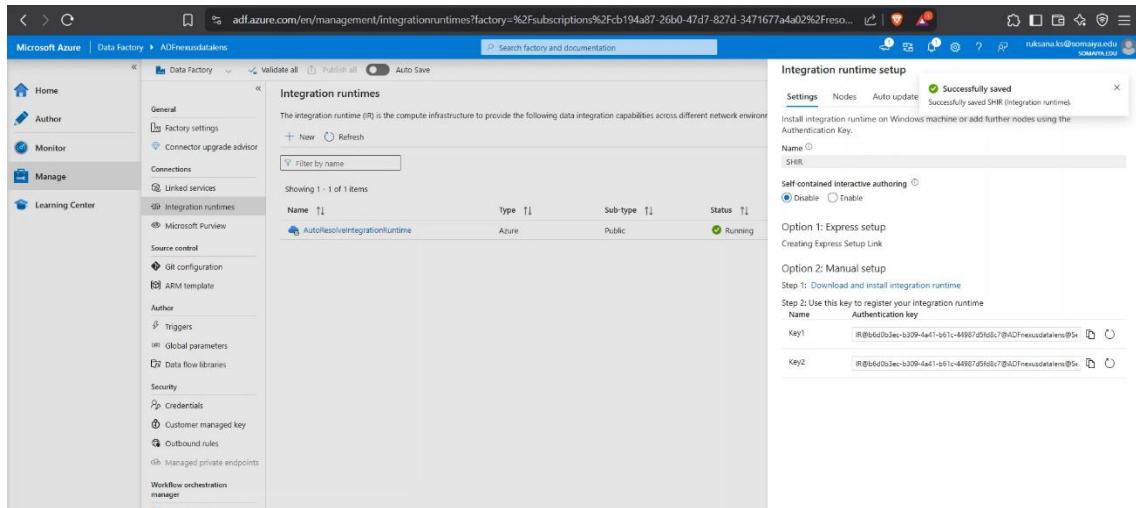
A screenshot of the Azure Data Factory 'Integration runtime setup' page. The left sidebar shows 'Data Factory' under 'Author'. The main area displays 'Integration runtimes' with one item listed: 'AutoResolveIntegrationRuntime' (Type: Azure, Sub-type: Public, Status: Running). To the right, there's a section titled 'Integration runtime setup' with three options: 'Azure, Self-Hosted', 'Azure-SSIS', and 'Airflow (Preview)'. The 'Azure, Self-Hosted' option is highlighted.

Create a new Self-hosted Integration Runtime in Azure Data Factory by entering name **SHIR** and type Self-Hosted clicking Create.

A screenshot of the 'Integration runtime setup' creation form. It shows a warning about private network support. The 'Name' field contains 'shir', the 'Description' field is empty, and the 'Type' is set to 'Self-Hosted'. At the bottom are 'Create' and 'Back' buttons.

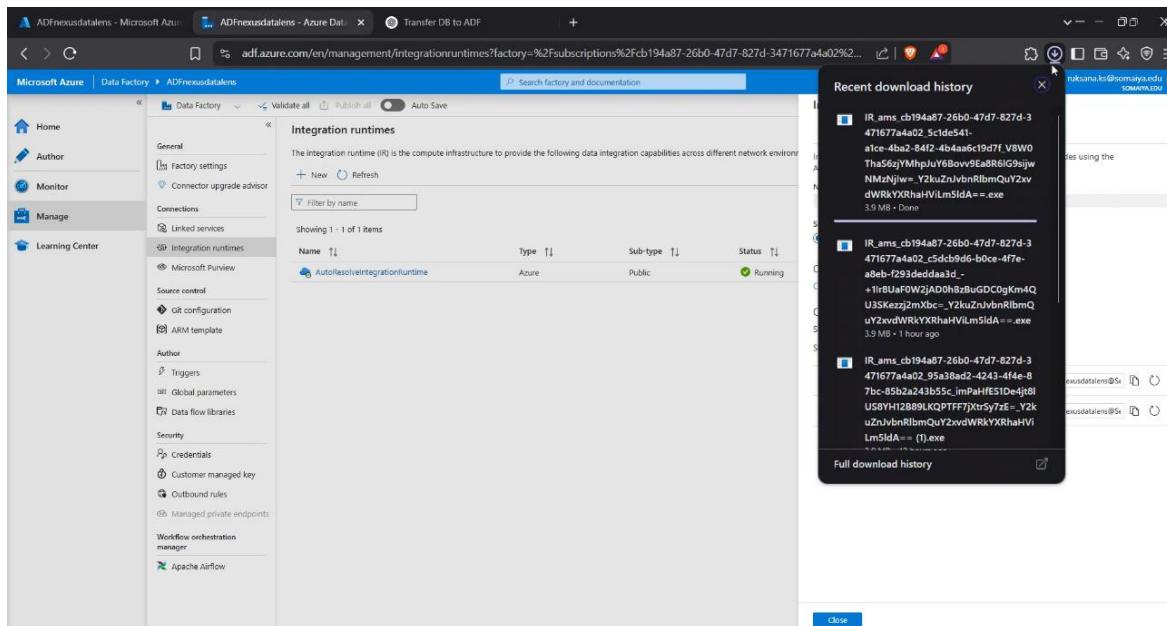
Nexus DataLens

Self-hosted Integration Runtime (SHIR) is created and provides authentication keys for registering it on a Windows machine via express or manual setup.



The screenshot shows the 'Integration runtimes' section of the Azure Data Factory management interface. A single runtime named 'AutoResolveIntegrationRuntime' is listed as Azure, Public, and Running. On the right, a modal window titled 'Integration runtime setup' is open, showing a successful save message. It includes options for 'Express setup' and 'Manual setup' with step-by-step instructions and download links for .exe files.

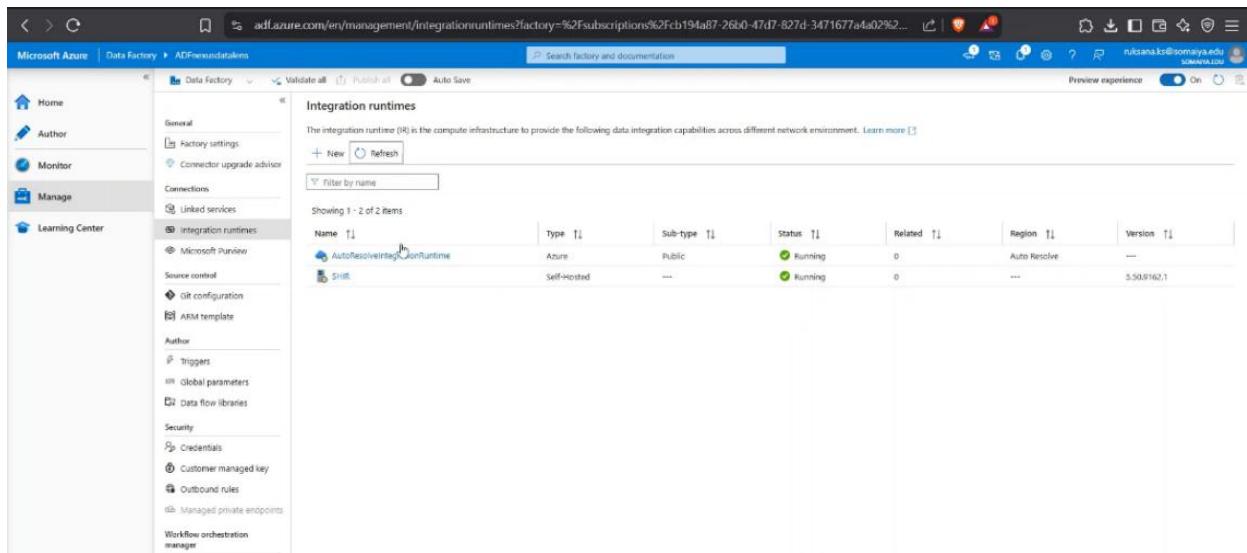
Azure Data Factory's Integration runtimes tab with a running Azure runtime. On the right, the recent downloads panel shows multiple .exe files, likely installers for a Self-hosted Integration Runtime (SHIR).



The screenshot shows the same 'Integration runtimes' section as before, but with a large overlay on the right side. This overlay is the 'Recent download history' pane, which lists several recently downloaded .exe files, likely corresponding to the SHIR installers shown in the previous screenshot.

Nexus DataLens

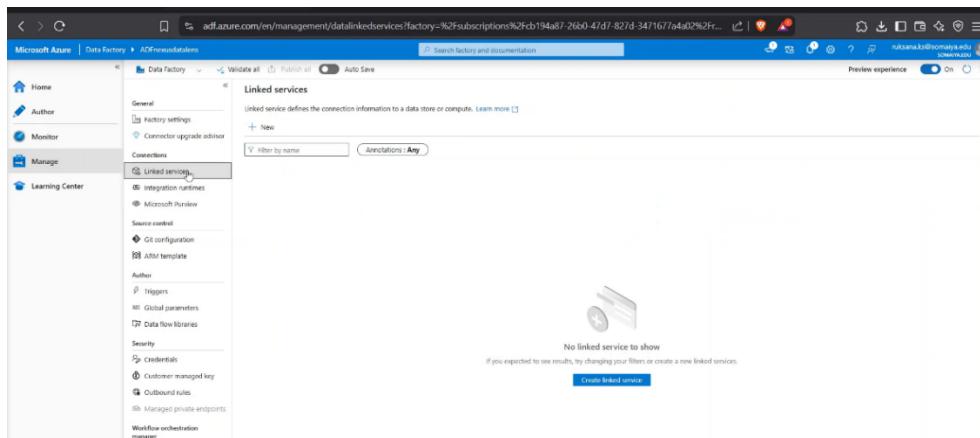
The Azure Data Factory now shows two integration runtimes:
AutoResolveIntegrationRuntime (Azure-hosted) and **SHIR (Self-hosted)**, both in a running state. This setup enables data integration both in the cloud and on-premises.



The screenshot shows the 'Integration runtimes' page in the Azure Data Factory. The left sidebar has 'Manage' selected. The main area displays a table with two rows:

Name	Type	Sub-type	Status	Related	Region	Version
AutoResolveIntegrationRuntime	Azure	Public	Running	0	Auto Resolve	---
SHIR	Self-hosted	---	Running	0	---	5.50.8162.1

The Linked services tab in Azure Data Factory, which currently has no connections configured. Linked services are used to define connection info for data sources or compute environments.



The screenshot shows the 'Linked services' page in the Azure Data Factory. The left sidebar has 'Manage' selected. The main area displays a message: 'No linked service to show'. It includes a note: 'If you expected to see results, try changing your filters or create a new linked services.' and a 'Create linked service' button.

Nexus DataLens

The process of creating a new linked service for a SQL Server in Azure Data Factory. The user is required to input connection details such as server name, database name, and authentication type.

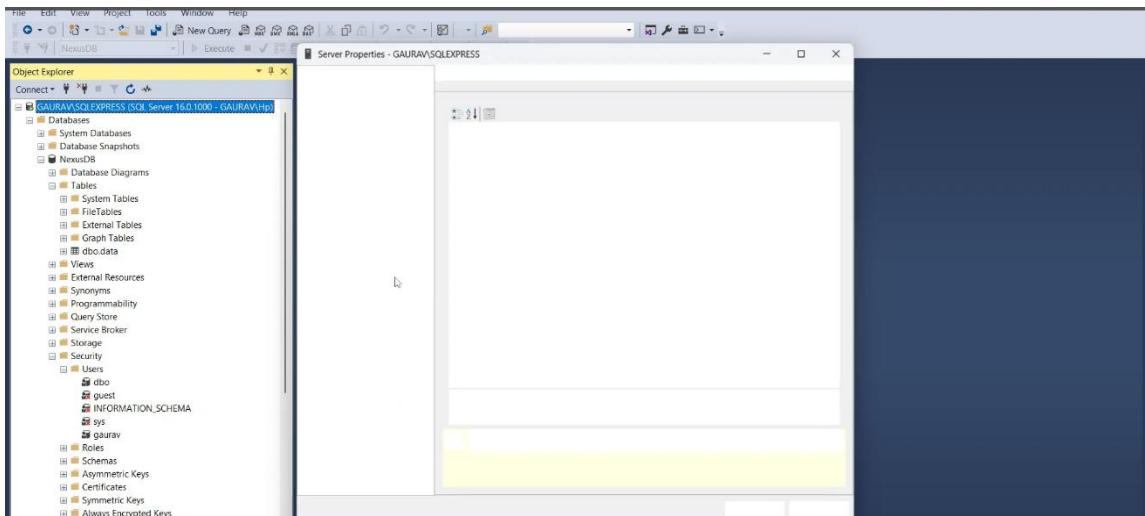
This screenshot shows the 'Linked services' creation page in the Azure Data Factory interface. On the left, a navigation sidebar lists various options like Home, Author, Monitor, Manage, and Learning Center. Under 'Manage', 'Linked services' is selected. The main area is titled 'New linked service' and specifies 'SQL server' as the type. The 'Name' field is populated with 'SqlServer1'. The 'Description' field is empty. The 'Connect via integration runtime' dropdown is set to 'SHIR'. The 'Version' radio button is selected for '2.0'. The 'Import from connection string' checkbox is unchecked. The 'Server name' field is empty. The 'Database name' field contains the placeholder 'Please fill out this field.' The 'Authentication type' dropdown is set to 'SQL authentication'. The 'User name' and 'Password' fields are empty, with 'Password' being highlighted in blue. The 'Azure Key Vault' button is visible but disabled. The 'Always encrypted' checkbox is unchecked. A note at the bottom says 'No linked service to show' and 'If you expected to see results, try changing your filters or create one'.

The Connection String Assistant in Azure Data Factory while creating a new linked service. It prompts the user to input a connection string to auto-populate SQL Server connection details securely.

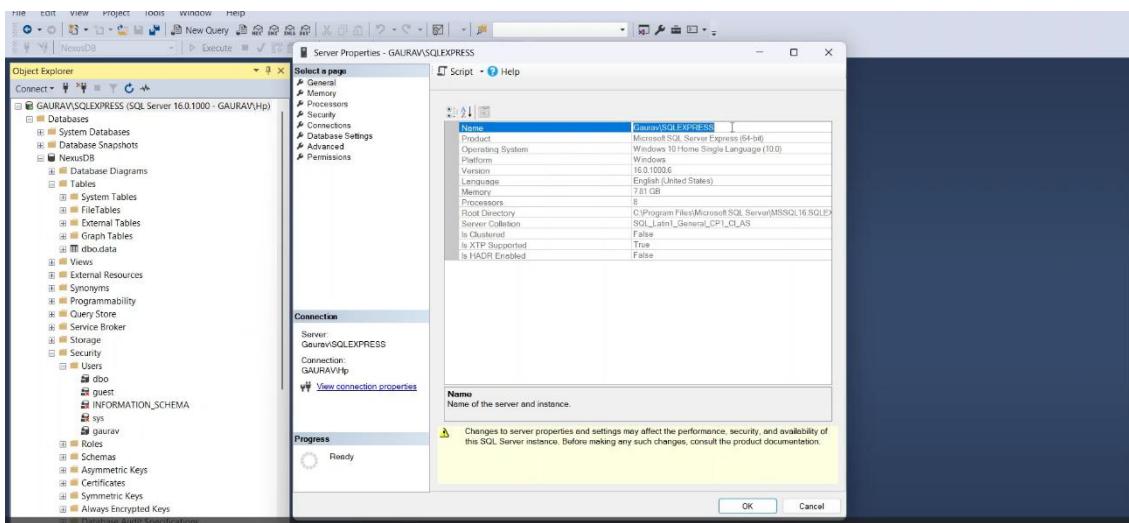
This screenshot shows the 'Connection string assistant' dialog box overlaid on the 'New linked service' page. The dialog has a warning message: 'The credential in the connection string will also be displayed. Please omit it if you prefer not to reveal it.' Below this is a large text input field labeled 'Connection string' with the placeholder 'Please put your connection string here, and we'll assist in populating the fields based on it.' At the bottom of the dialog are 'Import' and 'Cancel' buttons. The background page remains the same as the previous screenshot, showing the 'New linked service' configuration.

Nexus DataLens

SQL Server Management Studio (SSMS) with the NexusDB database expanded, displaying its tables, views, and other objects.



The Server Properties window in SSMS for the instance "GAURAV\SQLEXPRESS"(Server Name), showing system details like SQL Server version, operating system, memory, and collation. It helps users verify configuration and environment settings.



Nexus DataLens

The configuration of a new linked service in Azure Data Factory, connecting to a SQL Server instance named "**Gaurav\SQLEXPRESS**" with the database "**NexusDB**". SQL authentication is used with the username "**gaurav**", and an option to select an Azure Key Vault for the password is being displayed.

This screenshot shows the 'Linked services' configuration page in the Azure Data Factory portal. On the left, the navigation menu is visible with options like Home, Author, Monitor, Manage, Learning Center, and Data Flow. The 'Connections' section is selected, and 'Linked services' is chosen under it. In the center, there's a list of existing linked services. On the right, a form for creating a new linked service is displayed. The 'Name' field is set to 'SqlServer1'. Under 'Connect via Integration runtime', 'SIR' is selected. The 'Version' is set to 2.0. The 'Server name' is 'Gaurav\SQLEXPRESS', and the 'Database name' is 'NexusDB'. For 'Authentication type', 'SQL authentication' is selected, and the 'User name' is 'gaurav'. The 'Password' field has 'Azure Key Vault' selected. A dropdown menu for 'AKV Linked service' is open, showing 'Select...' and a search bar with 'Filter: 1'.

The process of creating a new linked service in Azure Data Factory to connect with an Azure Key Vault. The right pane captures the configuration details, including selecting the subscription, key vault name, and using system-assigned managed identity for authentication.

This screenshot shows the same 'Linked services' configuration page in the Azure Data Factory portal. The 'Name' field is now set to 'AzureKeyVault1'. Under 'Azure key vault selection method', 'From Azure subscription' is selected. The 'Azure subscription' dropdown shows 'Azure for Students (cb194a87-26b0-47d7-827d-3471677a402)' and 'Azure key vault name' is 'nexuskeyvault'. The 'Authentication method' is set to 'System-assigned managed identity'. At the bottom, the 'Managed identity name' is listed as 'ADFneudatalens' with object ID '0261fce8-14af-4925-8ab1-88cf8c8e0130'. There are also sections for 'Test connection', 'Annotations', 'Parameters', and 'Advanced'.

Nexus DataLens

The Access policies section of the Azure Key Vault named “**nexuskeyvault**”. Here, you can configure permissions for users and services to access secrets, keys, and certificates within the vault.

This screenshot shows the Microsoft Azure portal interface. The left sidebar is collapsed. The main navigation bar at the top includes 'Search resources, services, and docs (G+)', 'Copilot', and user information 'rukana.ls@somaiya.edu SOMAIYA EDU'. The current page title is 'nexuskeyvault | Access policies'. The left sidebar has several sections: Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, and Access policies (which is currently selected). Under 'Access policies', there are sub-options: Resource visualizer, Events, Objects (with sub-options Keys, Secrets, Certificates), Settings, Monitoring, Automation, and Help. The main content area is currently empty, showing a placeholder message: 'Access policies enable you to have fine grained control over access to vault items. Learn more'.

The Access policies list for the nexuskeyvault, showing users and their assigned permissions. It includes detailed rights such as Get, List, Set, Delete, Recover, Backup, and Restore for secrets, keys, and certificates.

This screenshot shows the Microsoft Azure portal interface, similar to the previous one but with a different URL in the address bar: 'portal.azure.com/#@somaiya.edu/resource/subscriptions/cb194a87-26b0-47d7-827d-3471677a4a02/resourceGroups/Nexu...'. The main content area now displays the 'Access policies' list for the 'nexuskeyvault' key vault. The table shows two entries:

Name	Email	Key Permissions	Secret Permissions	Certificate Permissions
UNKNOWN	2b0939e4-c9a3-4a8e-9557-3c9887229913		Get, List, Set, Delete, Recover, Backup, Restore	
USER	2115112-SHAikh RUKANA KHATUN rukana.ls@somaiya.edu	Get, List, Update, Create, Import, Delete, Recover, Ba...	Get, List, Set, Delete, Recover, Backup, Restore	Get, List, Update, Create, Import, Delete, Recover, Ba...

Nexus DataLens

Create an access policy page for Azure Key Vault, where you can assign specific permissions for keys, secrets, and certificates. It allows fine-grained control, such as selecting operations like Get, Set, Delete, or Purge under each category.

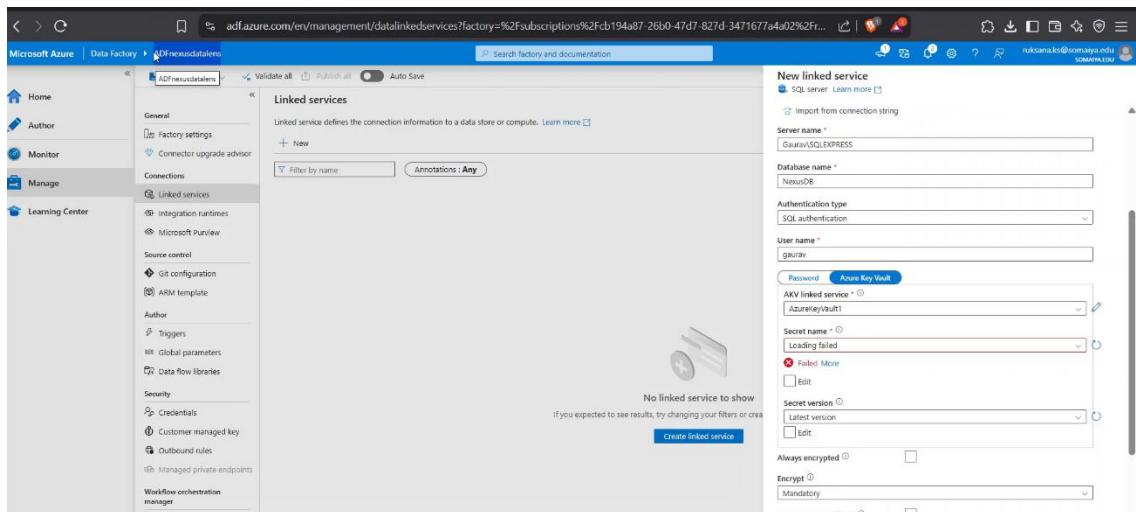
The screenshot shows the 'Create an access policy' page in the Microsoft Azure portal. At the top, there are tabs for 'Permissions', 'Principal', 'Application (optional)', and 'Review + create'. The 'Permissions' tab is selected. Below the tabs, there's a dropdown menu 'Configure from a template' with 'Select a template' as an option. The main area is divided into three columns: 'Key permissions', 'Secret permissions', and 'Certificate permissions'. Each column contains a list of checkboxes for various operations. In the 'Secret permissions' column, the 'Select all' checkbox is checked. In the 'Certificate permissions' column, the 'Select all' checkbox is also checked. At the bottom of the page, there's a large 'Next Step' button.

The Principal selection step while creating an access policy in Azure Key Vault. Here, you choose a user or service principal by searching with name, object ID, or email to assign the defined permissions.

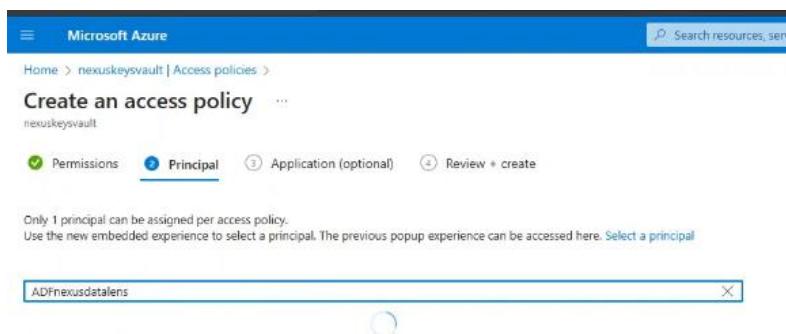
The screenshot shows the 'Create an access policy' page in the Microsoft Azure portal, specifically the 'Principal' selection step. At the top, there are tabs for 'Permissions', 'Principal' (which is selected), 'Application (optional)', and 'Review + create'. Below the tabs, there's a note: 'Only 1 principal can be assigned per access policy. Use the new embedded experience to select a principal. The previous popup experience can be accessed here. Select a principal.' A search bar 'Search by object ID, name, or email address' is present. A list of principals is shown, with the first item '12D-SF-01@somayaya.edu' selected. At the bottom, there's a 'Selected item' section with the message 'No item selected'.

Nexus DataLens

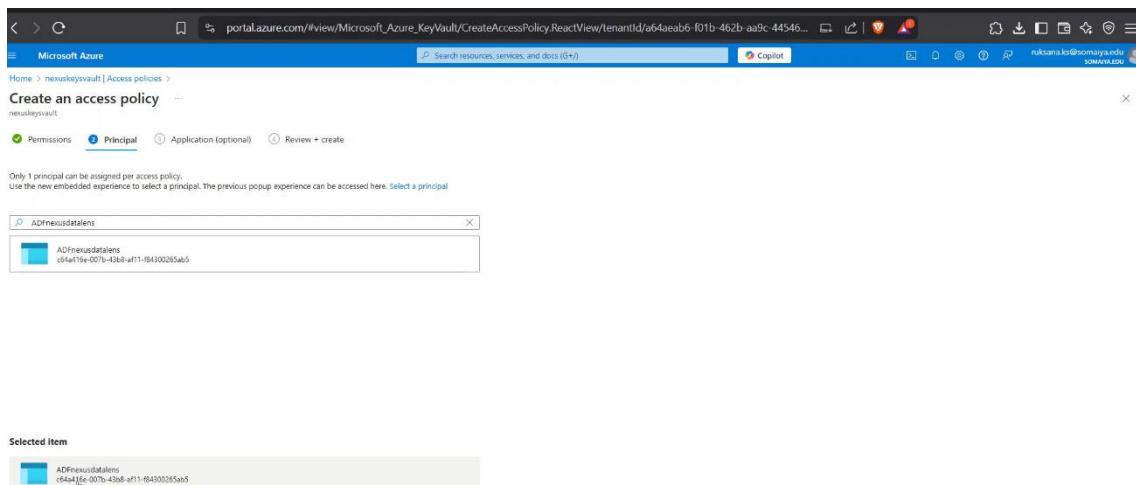
The process of linking a SQL Server in Azure Data Factory using SQL authentication with the password fetched from Azure Key Vault.



The selection of the “**ADFnexusdatalens**” Data Factory as a principal while creating an access policy in Azure Key Vault.



The Data Factory instance **ADFnexusdatalens** has been successfully selected as the principal for the Key Vault access policy.



Nexus DataLens

This step confirms that an application (**ADFnexusdatalens**) has already been selected as the principal to authorize Key Vault access.

The screenshot shows the Azure portal interface for creating an access policy. The URL is https://portal.azure.com/#view/Microsoft_Azure_KeyVault/CreateAccessPolicy/ReactView/tenantId/a64aeab6-f01b-462b-aa9c-44546.... The page title is "Create an access policy" under "nexuskeyvault". The tabs at the top are "Permissions", "Principal", "Application (optional)" (which is selected), and "Review + create". A note below the tabs says "An application has already been chosen as a principal." and "Authorizes this application to perform the specified permissions on the User's or Group's behalf."

The final review of the access policy, showing that only **secret permissions** are granted to the **ADFnexusdatalens principal**.

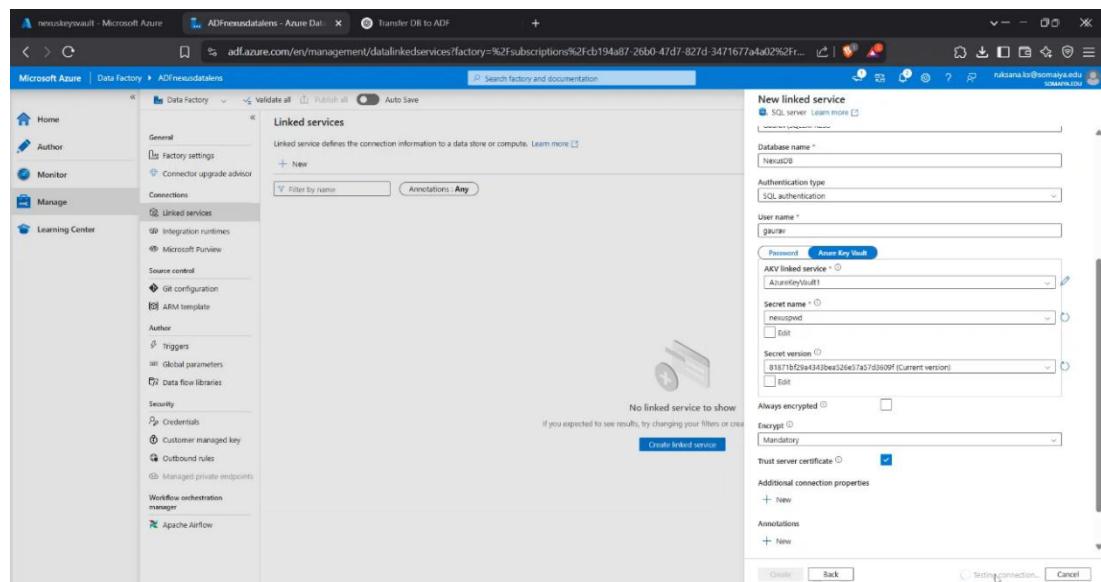
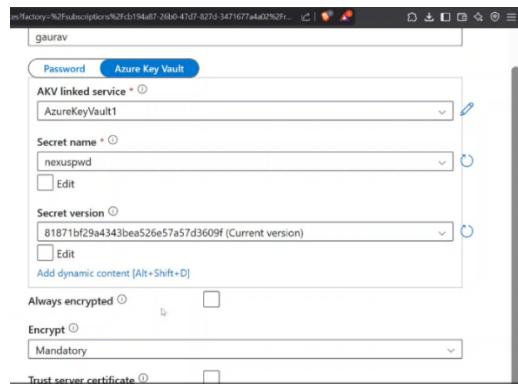
The screenshot shows the final review of the access policy. The URL is https://portal.azure.com/#view/Microsoft_Azure_KeyVault/CreateAccessPolicy/ReactView/tenantId/a64aeab6-f01b-462b-aa9c-44546.... The page title is "Create an access policy" under "nexuskeyvault". The tabs are the same as the previous screenshot. The "Secret Permissions" section is expanded, showing "All selected" for "Secret Management Operations". Other sections like "Key Permissions" and "Certificate Permissions" are also visible.

The access policy for the Key Vault "**nexuskeyvault**" has been successfully updated, granting secret permissions to **ADFnexusdatalens**.

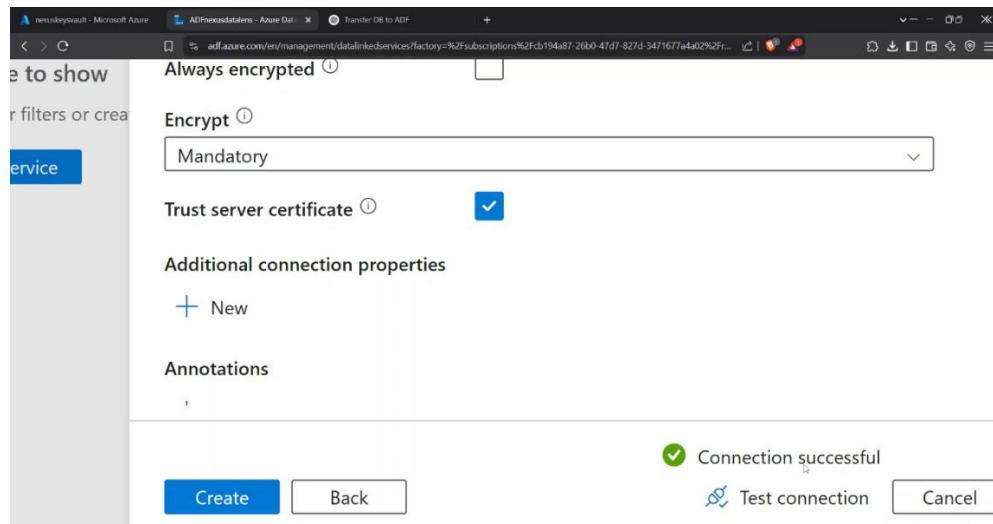
The screenshot shows the updated access policy for the "nexuskeyvault" key vault. The URL is <https://portal.azure.com/#@somaiya.edu/resource/subscriptions/cb194a87-26b0-47d7-827d-3471677a4a02/resourceGroups/Nexu...>. The page title is "nexuskeyvault | Access policies". A success message says "Updating the key vault 'nexuskeyvault': The key vault 'nexuskeyvault' has been successfully updated." The left sidebar shows the navigation menu for the Azure portal. The main area displays the access policy details, including the principal name (ADFnexusdatalens), object ID (0261fce1-14ef-4925-8ab1-83cfbcde0130), and the updated secret permissions for the application.

Nexus DataLens

The Azure Key Vault secret named "**nexuspwd**" with a specified version is selected from **AzureKeyVault1** for secure integration in the linked service.

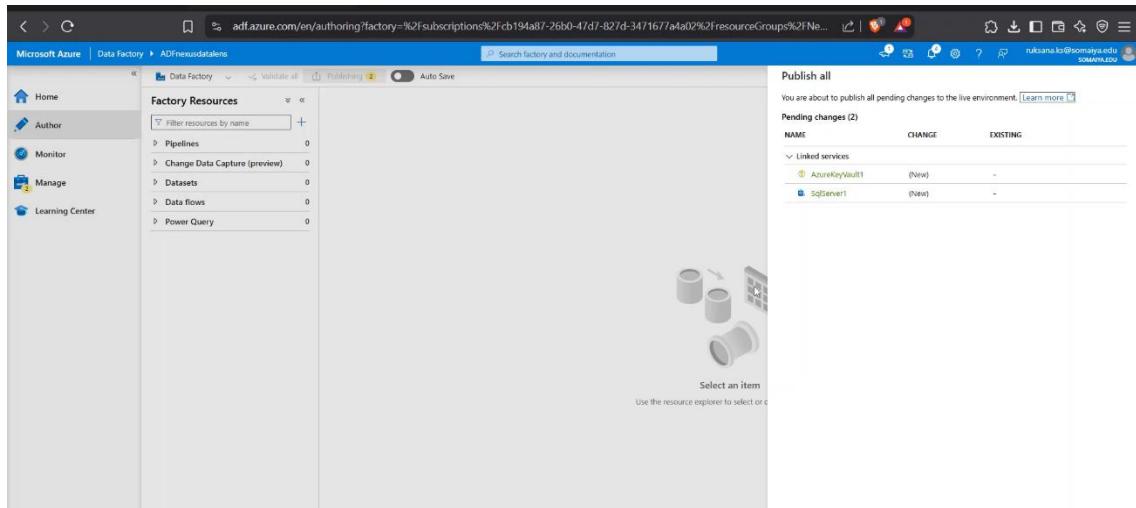


The database connection configuration is complete with encryption and certificate trust, and the **connection test was successful**.

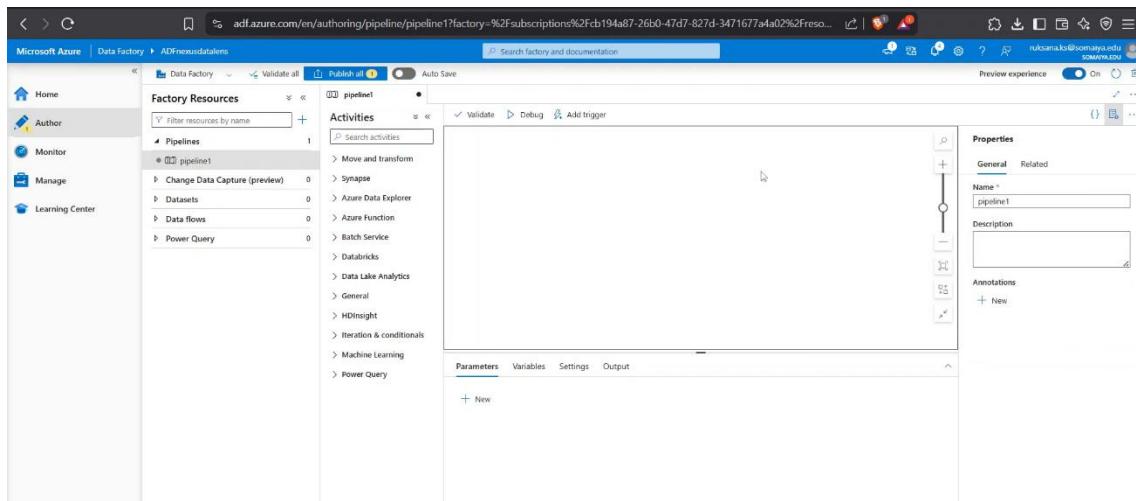


Nexus DataLens

Two new linked services, **AzureKeyVault1** and **SqlServer1**, are ready to be published to the live Azure Data Factory environment.

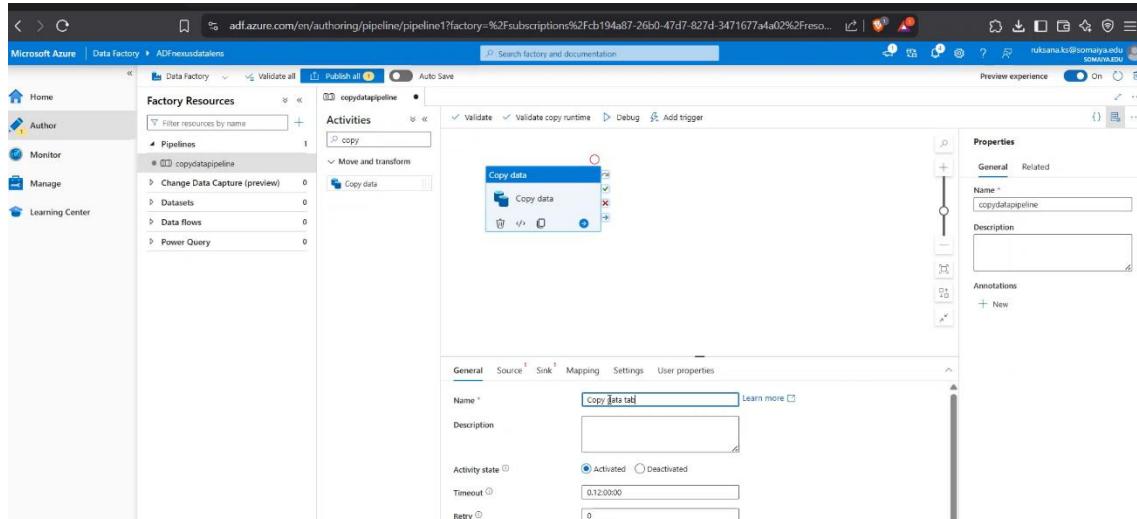


A new pipeline named "**pipeline1**" is being created in Azure Data Factory, ready for activities and configurations.

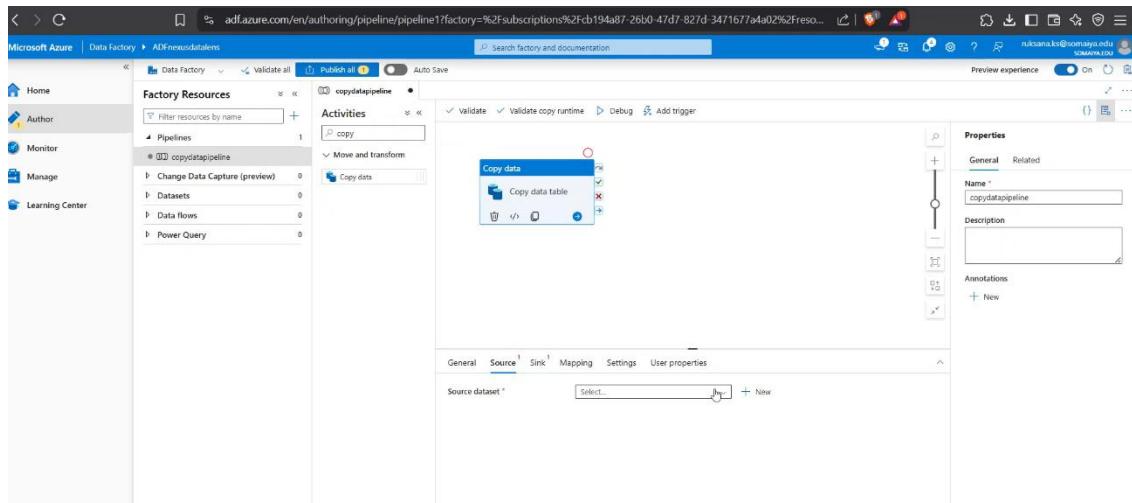


Nexus DataLens

A pipeline named "**copydatipeline**" with a "**Copy data**" activity titled "**Copy data task**" is being configured in Azure Data Factory.

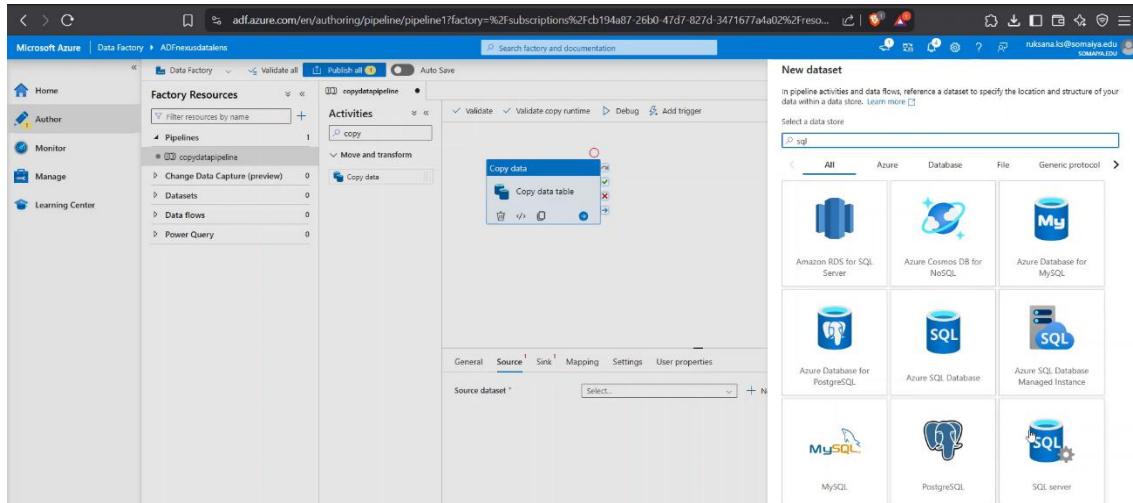


In Azure Data Factory, the source dataset for the "**Copy data table**" activity in the "**copydatipeline**" is being selected.



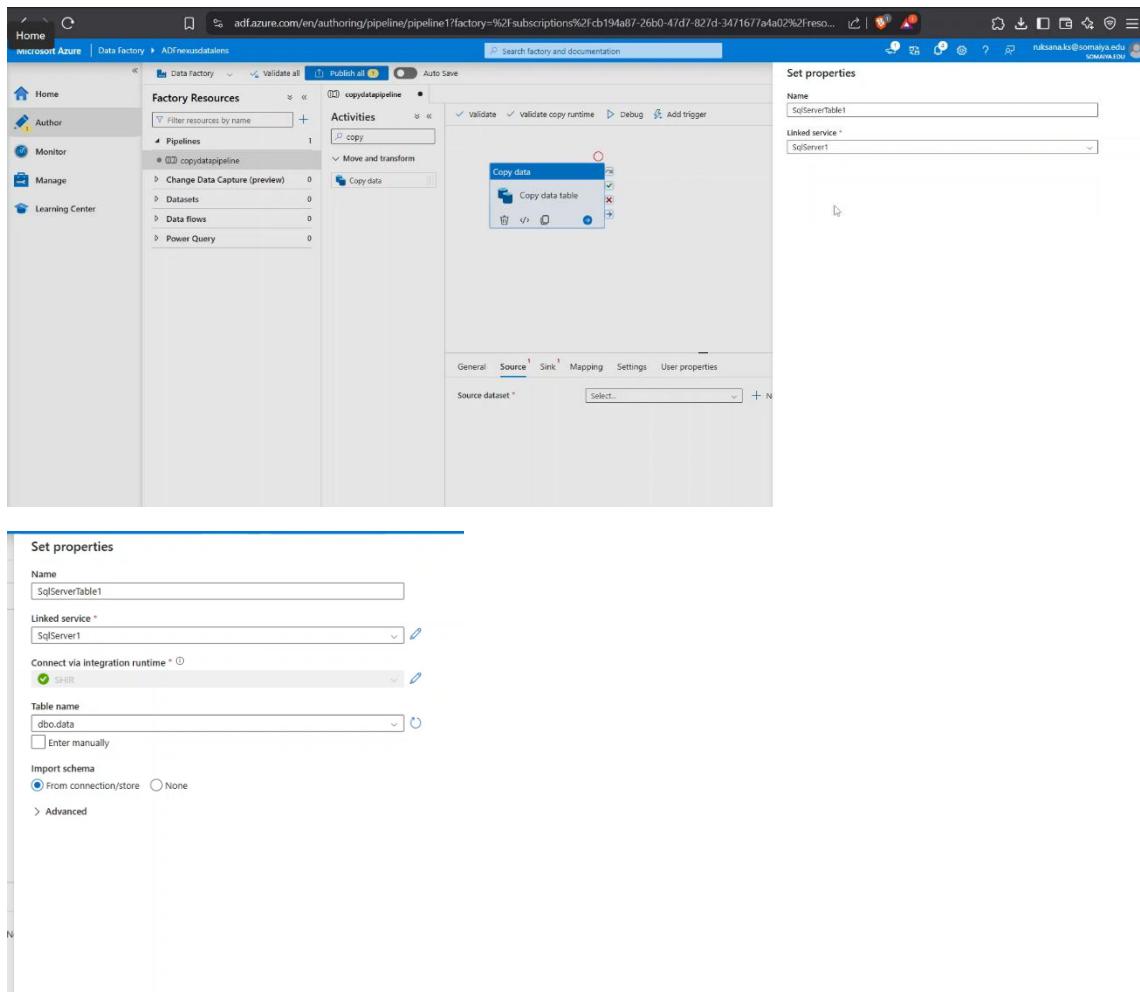
Nexus DataLens

In Azure Data Factory, a new source dataset is being created by selecting a data store type, with options like SQL Server, MySQL, and Azure SQL Database shown.



The screenshot shows the Azure Data Factory pipeline editor. On the left, the 'Factory Resources' sidebar lists Pipelines, Datasets, Data flows, and Power Query. In the center, a pipeline named 'copydatipeline' is displayed with a 'Copy data' activity selected. To the right, a 'New dataset' pane is open, showing a search bar with 'sql' typed in. Below the search bar is a grid of icons representing various data stores: Amazon RDS for SQL Server, Azure Cosmos DB for NoSQL, Azure Database for MySQL, Azure Database for PostgreSQL, Azure SQL Database, Azure SQL Database Managed Instance, MySQL, PostgreSQL, and SQL server.

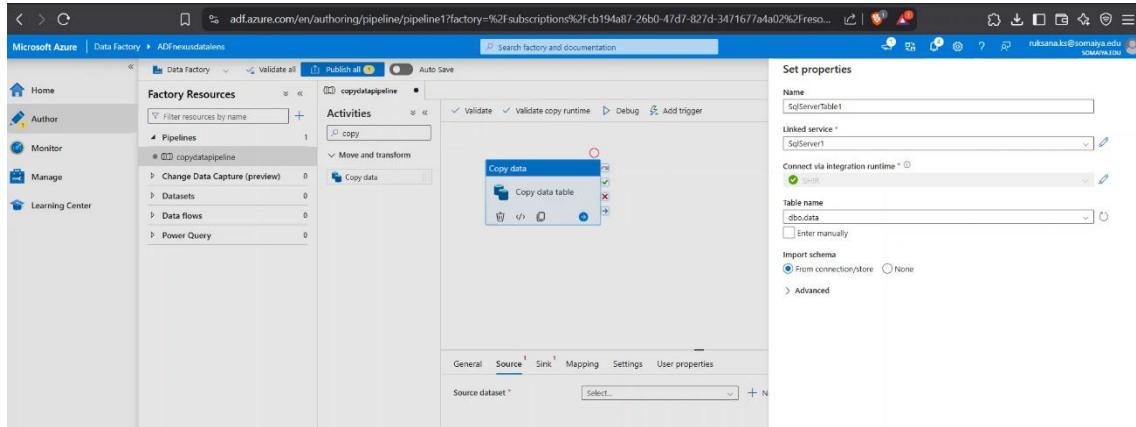
In Azure Data Factory, a source dataset named "**SqlServerTable1**" is being created using the linked service "**SqlServer1**".



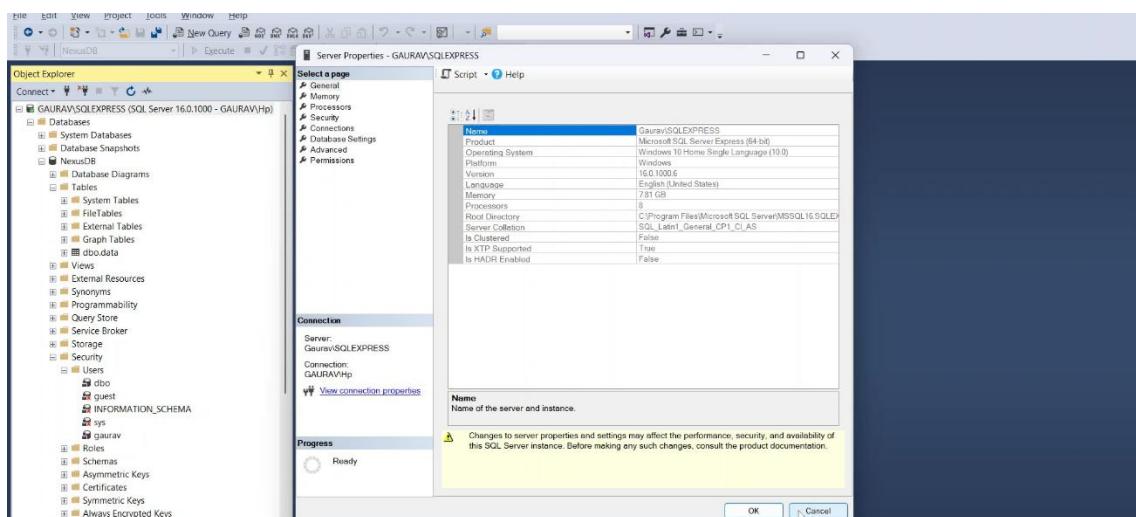
The screenshot shows the 'Source' tab of the 'Copy data' activity configuration. On the left, the 'Factory Resources' sidebar lists Pipelines, Datasets, Data flows, and Power Query. In the center, a pipeline named 'copydatipeline' is displayed with a 'Copy data' activity selected. To the right, a 'Set properties' pane shows 'Name' set to 'SqlServerTable1' and 'Linked service' set to 'SqlServer1'. The main configuration pane shows 'Table name' set to 'dbo.data' and 'Import schema' set to 'From connection/store'.

Nexus DataLens

The source dataset "SqlServerTable1" is configured to connect to the "SqlServer1" linked service and pull data from the "db.data" table using integration runtime.

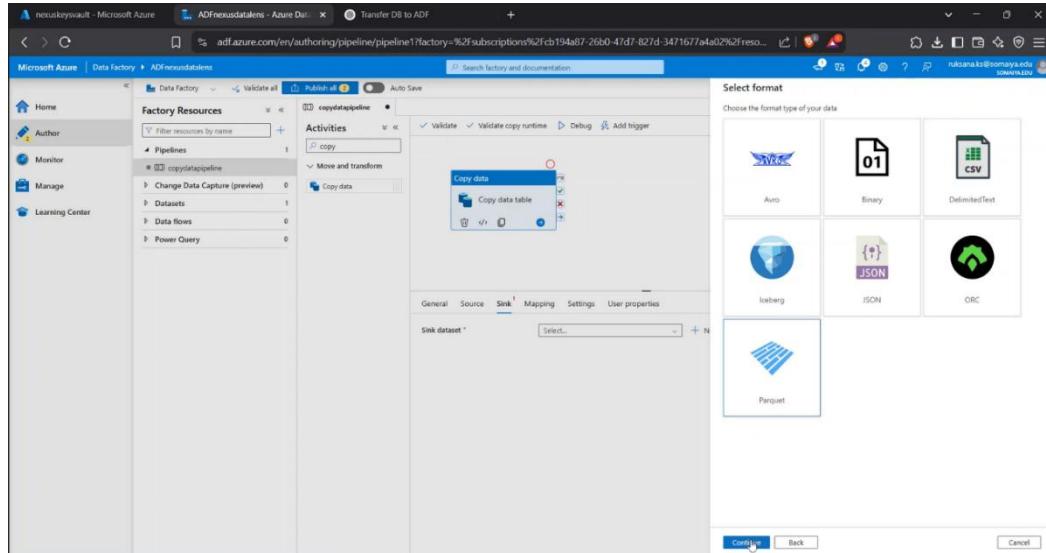


The SQL Server Management Studio (SSMS) is displaying the server properties for the local instance "GAURAV\SQLEXPRESS" running SQL Server Express on a Windows machine.

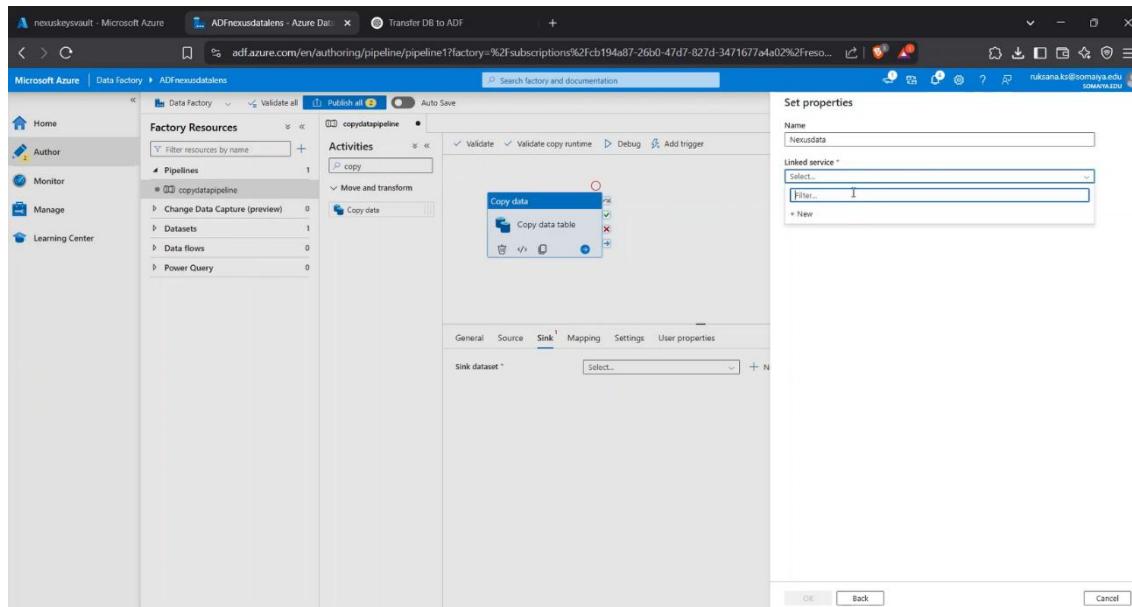


Nexus DataLens

In Azure Data Factory, the user is selecting the **Parquet** format for the sink dataset while configuring a Copy Data activity in the pipeline.

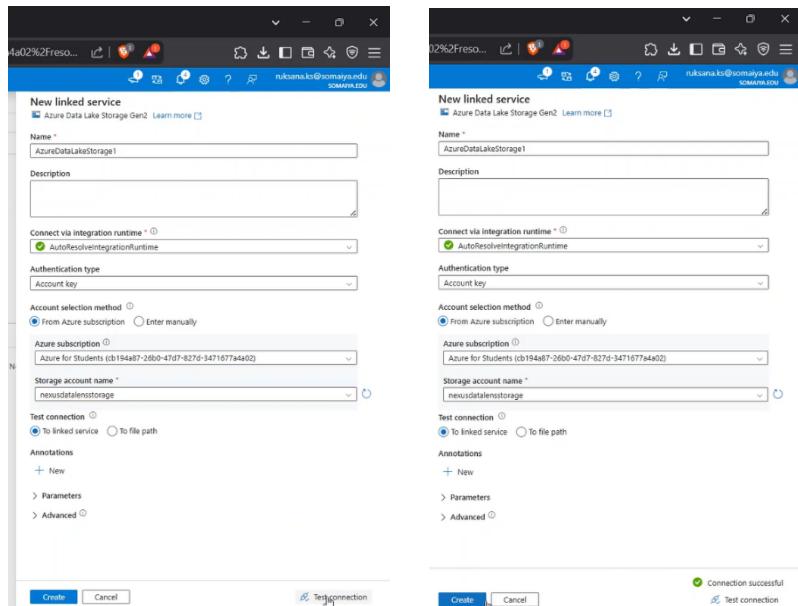


The user is setting properties for the sink dataset named **Nexusdata** by selecting or creating a linked service in Azure Data Factory.

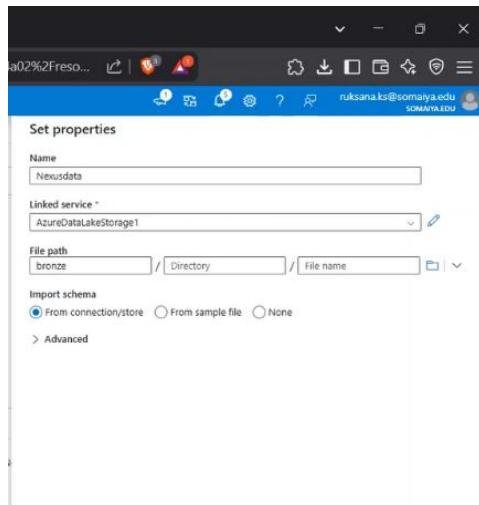


Nexus DataLens

The user is configuring a new linked service to connect **Azure Data Factory** with **Azure Data Lake Storage Gen2** using an account key and is testing the connection.

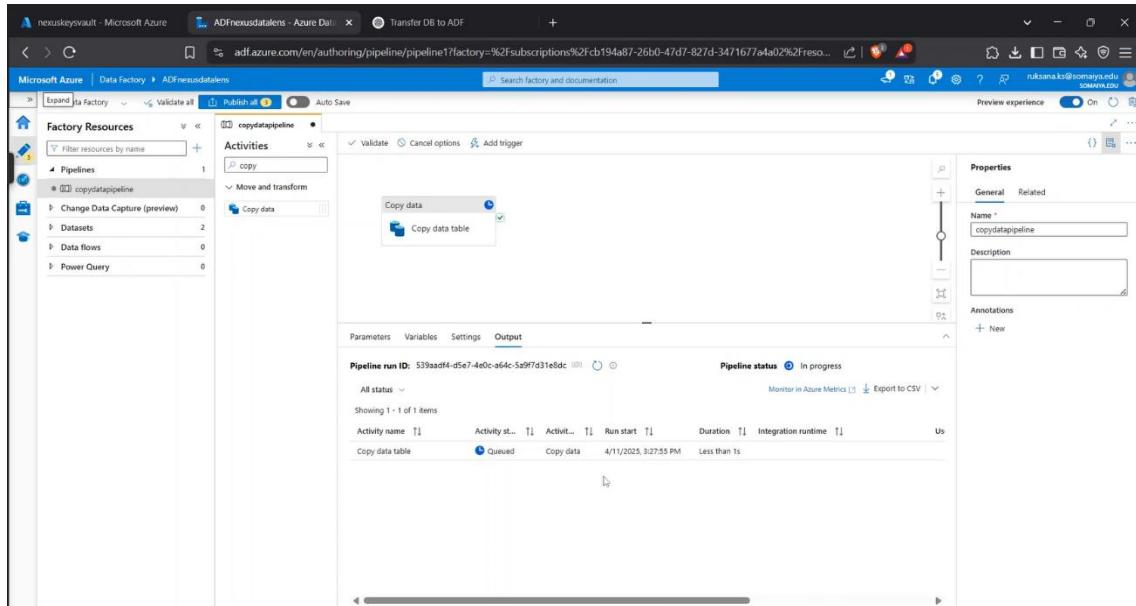


The user is setting the sink dataset properties to write data into the "**bronze**" directory of **Azure Data Lake Storage Gen2** using the linked service.

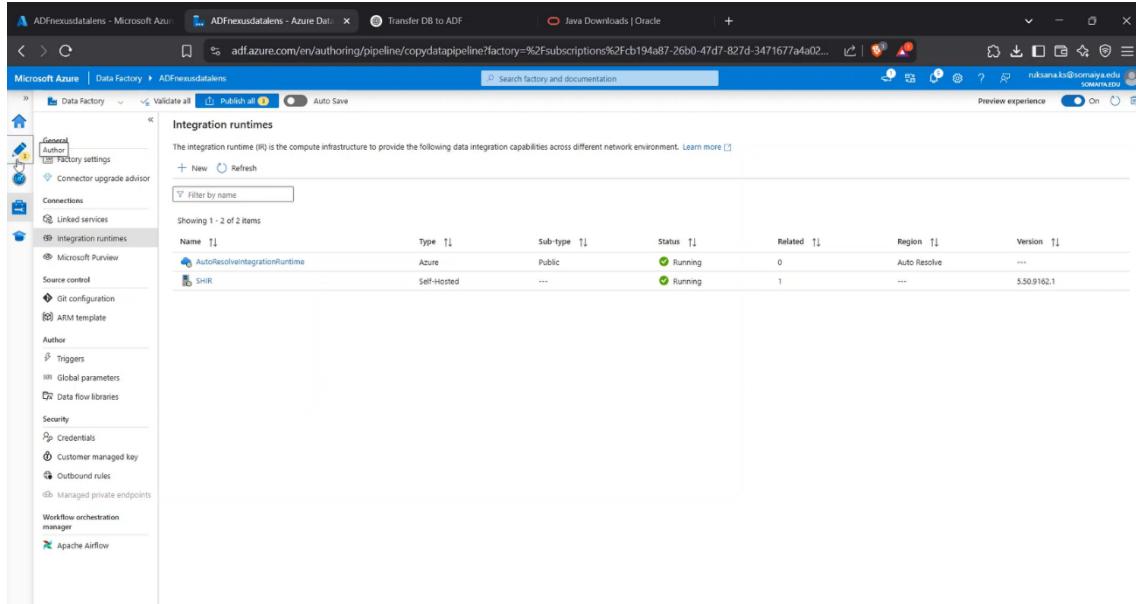


Nexus DataLens

The pipeline run for copying data has started and the "**Copy data table**" activity is currently in the **Queued state**.

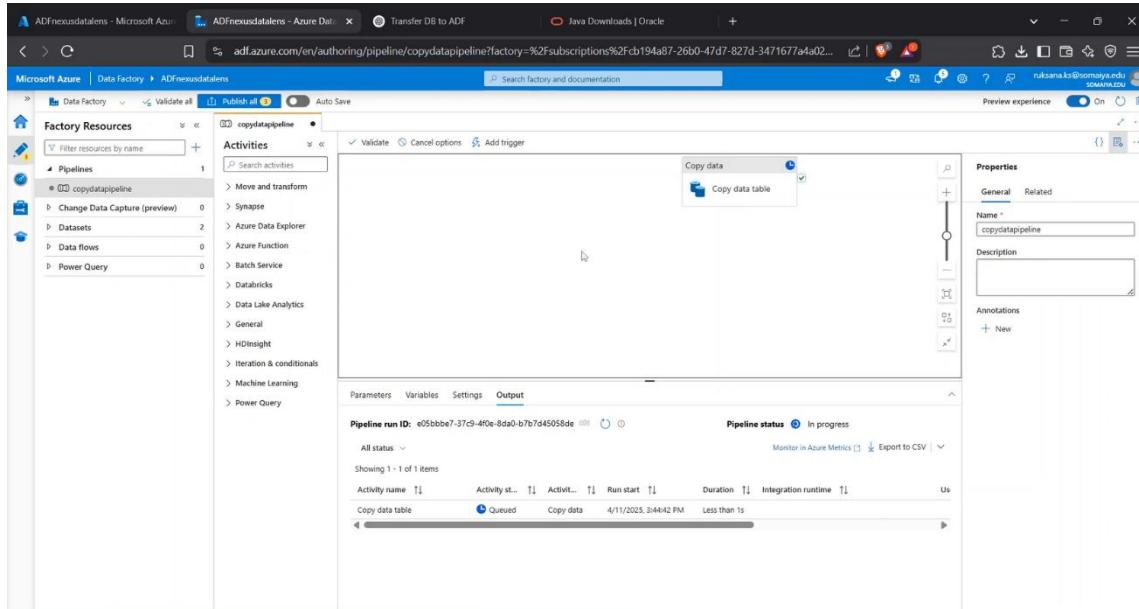


The Integration Runtimes panel shows both **Azure** and **Self-hosted runtimes** are running, enabling data movement across cloud and on-premises environments.

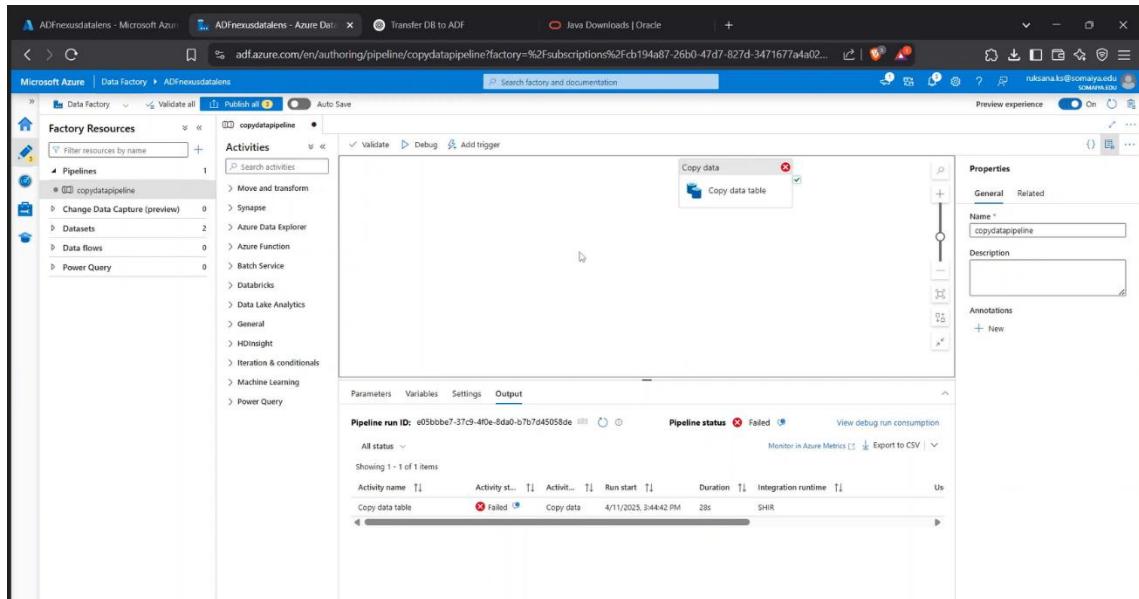


Nexus DataLens

The pipeline run in Azure Data Factory is currently in progress, with the "**Copy data table**" activity queued for execution.



The Azure Data Factory pipeline run has failed during the "**Copy data table**" activity using the self-hosted integration runtime (SHIR).



Nexus DataLens

The Azure Data Factory pipeline **successfully completed** the "Copy data table" activity using the self-hosted integration runtime (SHIR).

This screenshot shows the Azure Data Factory pipeline named 'copydatapipeline'. The pipeline consists of a single activity, 'Copy data', which has completed successfully. The pipeline status is 'Succeeded'. The run details show a single activity named 'Copy data table' with a duration of 14 seconds. The pipeline run ID is 09b8c837-ecaf-4efb-b573-82dd94a6103.

The "**copydatapipeline**" in Azure Data Factory successfully completed the data transfer using SHIR in 18 seconds.

This screenshot shows the Azure Data Factory pipeline named 'copydatapipeline'. The pipeline consists of a single activity, 'Copy data', which has completed successfully. The pipeline status is 'Succeeded'. The run details show a single activity named 'Copy data table' with a duration of 18 seconds. The pipeline run ID is 8638091e-7fcf-44b3-a74a-3489d50b115.

Nexus DataLens

The Azure portal home page displays active resources including a Data Factory, Key Vault, Resource Group, and Databricks service under the user's subscription.

The screenshot shows the Microsoft Azure portal home page. At the top, there are several tabs: 'ADFnxusdataLens - Microsoft', 'ADFnxusdataLens - Azure Data Factory', 'Transfer DB to ADF', 'Java Downloads | Oracle', and 'Search resources, services, and docs (S+?)'. Below the tabs, the URL 'portal.azure.com/#/home' is visible. The main content area is titled 'Azure services' and features a row of icons for 'Create a resource', 'Subscriptions', 'Resource groups', 'Azure Synapse Analytics', 'Budgets', 'Bot Services', 'Projects', 'Quickstart Center', 'Azure AI services', and 'More services'. Below this is a section titled 'Resources' with a 'Recent' tab selected. It lists the following resources:

Name	Type	Last Viewed
ADFnxusdataLens	Data factory (V2)	35 minutes ago
nexuskeyvault	Key vault	an hour ago
NexusProject	Resource group	an hour ago
nexusdatabricks	Azure Databricks Service	3 days ago
Azure for Students	Subscription	4 days ago

Below the resources is a 'See all' link. Under 'Navigate', there are links for 'Subscriptions', 'Resource groups', 'All resources', and 'Dashboard'. Under 'Tools', there are links for 'Microsoft Learn', 'Azure Monitor', 'Microsoft Defender for Cloud', and 'Cost Management'. The URL at the bottom of the page is <https://portal.azure.com/#@somaya.edu/asset/HubExtension/ResourceGroups/subscriptions...>.

The NexusProject resource group in Azure contains various deployed resources including Data Factory, Databricks, a Storage Account, a Key Vault, and a Synapse workspace in the Central India region.

The screenshot shows the Microsoft Azure portal page for the 'NexusProject' resource group. The URL is <https://portal.azure.com/#@somaya.edu/resource/subscriptions/cb194a87-26b0-47d7-827d-3471677a4a02/resourceGroups/NexusProject>. The page title is 'NexusProject - Microsoft Azure'. On the left, there is a navigation sidebar with links for 'Overview', 'Activity log', 'Access control (IAM)', 'Tags', 'Resource visualizer', 'Events', 'Settings', 'Cost Management', 'Monitoring', 'Automation', and 'Help'. The main content area has a 'Search' bar and a toolbar with 'Create', 'Manage view', 'Delete resource group', 'Refresh', 'Export to CSV', 'Open query', 'Assign tags', 'Move', 'Delete', 'Export template', and 'Open in mobile'. Below the toolbar, there is an 'Essentials' section showing 'Subscription (Owner) : Azure for Students', 'Subscription ID : cb194a87-26b0-47d7-827d-3471677a4a02', and 'Tags (edit) : All tags'. The 'Resources' section shows a table of deployed resources:

Name	Type	Location	Actions
ADFnxusdataLens	Data factory (V2)	Central India	...
nexusdatabricks	Azure Databricks Service	Central India	...
nexusdatstorage	Storage account	Central India	...
nexuskeyvault	Key vault	Central India	...
nexussynapse	Synapse workspace	Central India	...

At the bottom of the page, there are navigation links for '< Previous', 'Page 1 of 1', and 'Next >'. The URL at the bottom of the page is <https://portal.azure.com/#@somaya.edu/resource/subscriptions/cb194a87-26b0-47d7-827d...>.

Nexus DataLens

The **nexusdatalakestorage** is a successfully provisioned Azure Data Lake Storage Gen2 account with hierarchical namespace enabled and various features like blob soft delete and versioning disabled.

Essentials

Resource group (Subscription)	: NexusProject
Location	: centralindia
Primary/Secondary Location	: Primary: Central India, Secondary: South India
Subscription (Subscription)	: Azure for Students
Subscription ID	: cb194a87-26b0-47d7-827d-3471677a4a02
Disk state	: Primary: Available, Secondary: Available
Tags (Tags)	: Add tags

Data Lake Storage

Hierarchical namespace	Enabled
Default access tier	Hot
Blob anonymous access	Disabled
Blob soft delete	Disabled
Container soft delete	Disabled
Versioning	Disabled
Change feed	Disabled
NFS v3	Disabled
SFTP	Disabled
Storage tasks assignments	None

File service

Large file share	Enabled
Identity-based access	Not configured
Default share-level permissions	Disabled

Security

Require secure transfer for REST API operations	Enabled
Storage account key access	Enabled
Minimum TLS version	Version 1.2
Infrastructure encryption	Disabled

Networking

Allow access from	All networks
Private endpoint connections	0
Network routing	Microsoft network routing
Access for trusted Microsoft services	Yes
Endpoint type	Standard

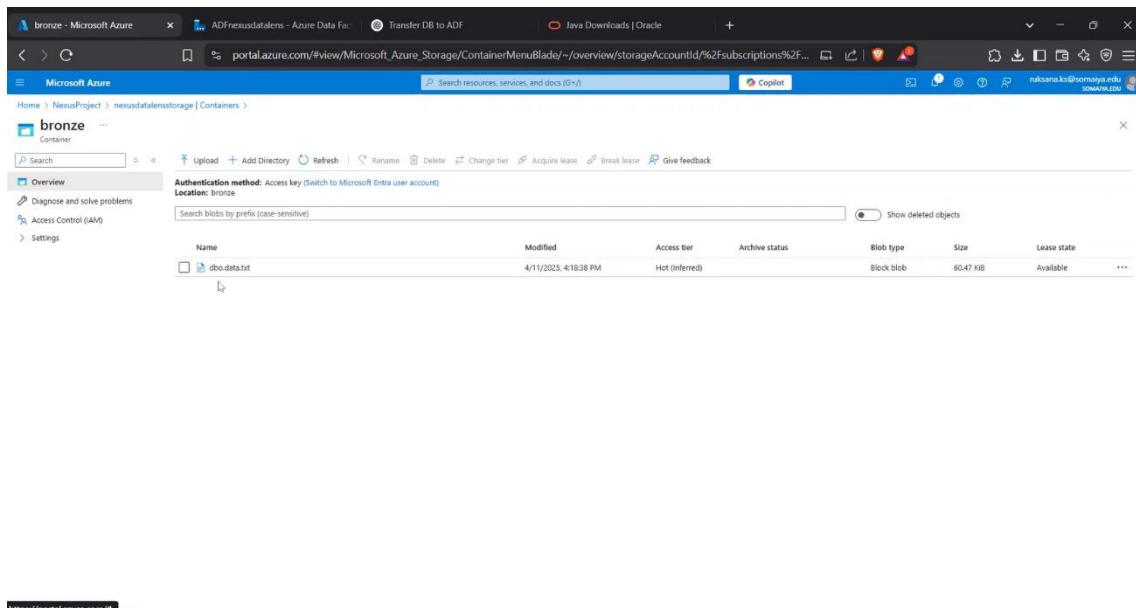
The **nexusdatalakestorage** account contains a single container named **bronze** with private access and is currently in an available lease state.

Containers

Name	Last modified	Anonymous access level	Lease state
bronze	4/8/2025, 5:30:40 PM	Private	Available

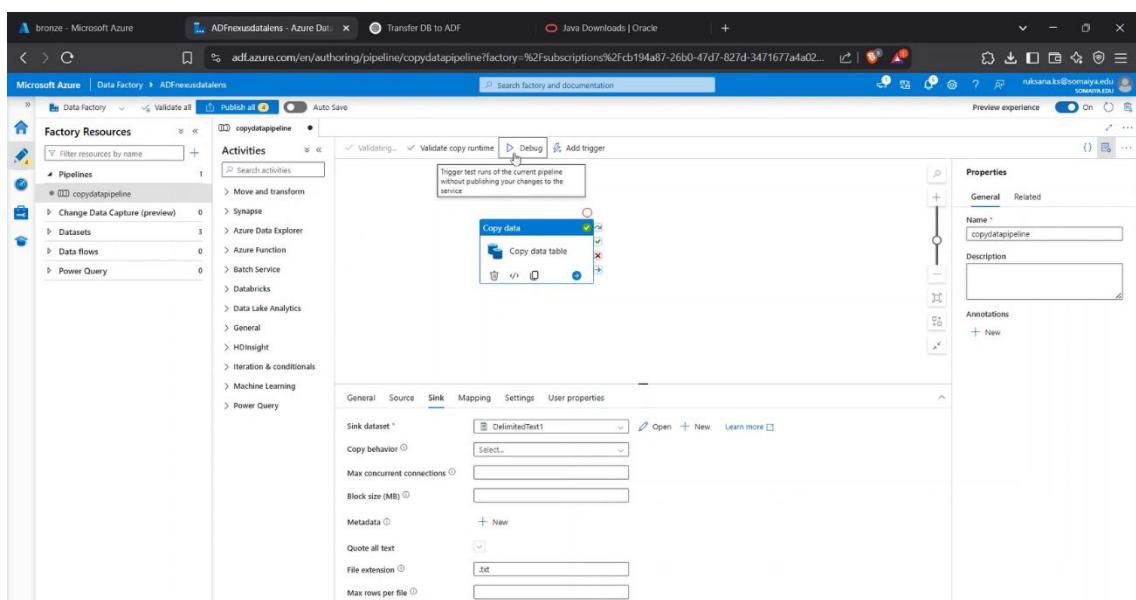
Nexus DataLens

The bronze container in the nexusdatalakestorage account contains a single block blob file named **dbo.data.txt**, stored in the hot access tier and currently available.



A screenshot of the Microsoft Azure Storage Explorer interface. The left sidebar shows the 'bronze' container under 'Containers'. The main area displays a table of blobs. One blob is listed: 'dbo.data.txt'. The details for this blob are: Name: dbo.data.txt, Modified: 4/11/2025, 4:18:38 PM, Access tier: Hot (inferred), Archive status: Not yet archived, Blob type: Block blob, Size: 60.47 KB, Lease state: Available. There are also 'Upload', 'Add Directory', 'Refresh', 'Rename', 'Delete', 'Change tier', 'Acquire lease', 'Break lease', and 'Give feedback' buttons at the top of the blob list.

The sink settings of the "**Copy data**" activity in the Azure Data Factory pipeline are configured to write data to a dataset named **DelimitedText1**.



A screenshot of the Azure Data Factory pipeline editor. On the left, the 'Factory Resources' pane shows a pipeline named 'copydatapipeline'. In the center, the 'Activities' pane shows a 'Copy data' activity selected. The 'Properties' pane on the right shows the activity's name as 'copydatapipeline'. The 'Sink' tab of the activity configuration is selected, showing the 'Sink dataset' dropdown set to 'DelimitedText1'. Other sink settings include 'Copy behavior' (Select...), 'Max concurrent connections' (1), 'Block size (MB)' (1), 'Metadata' (+ New), 'Quote all text' (checked), 'File extension' (.txt), and 'Max rows per file' (1).

Nexus DataLens

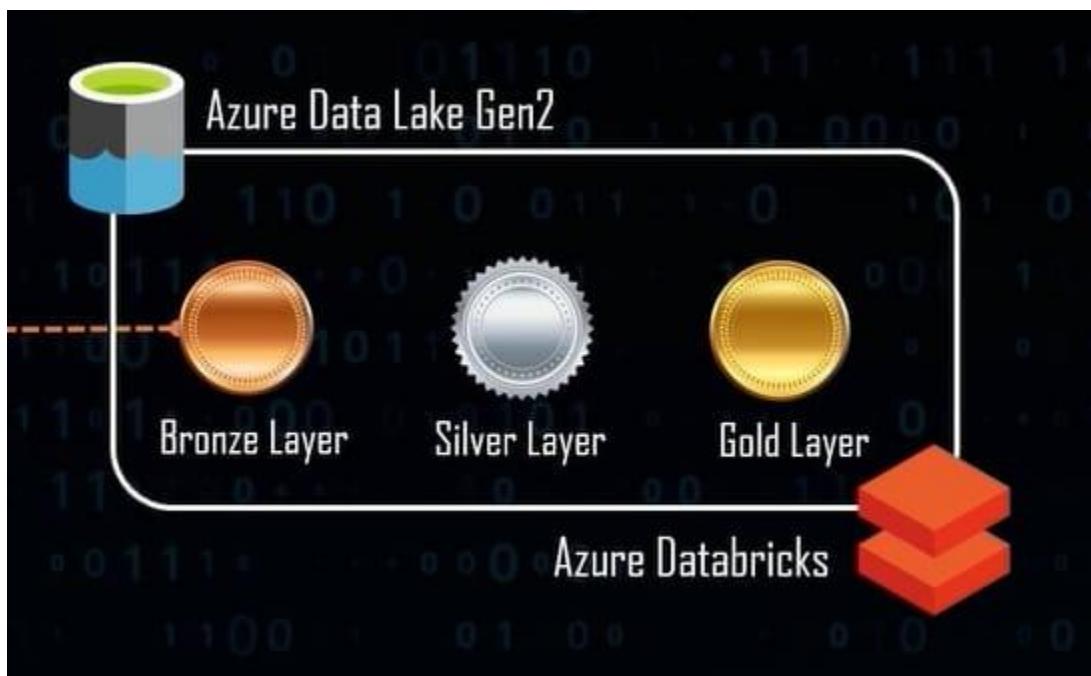
The Azure Data Factory pipeline "**copydatapipeline**" successfully completed the "**Copy data table**" activity in 22 seconds.

The screenshot shows the Azure Data Factory pipeline 'copydatapipeline'. The pipeline consists of a single 'Copy data' activity, which has completed successfully. The pipeline run ID is e968fb41-a1e1-4737-996c-ed8408d6708. The duration of the run was 22 seconds. The pipeline status is marked as 'Succeeded'.

The Azure Storage container "**bronze**" contains a file named **dbo.data.txt**, uploaded on 4/11/2025, with a size of 60.47 KB and access tier set to hot.

The screenshot shows the Azure Storage container 'bronze'. It contains a single file named 'dbo.data.txt'. The file was uploaded on 4/11/2025 at 4:20:20 PM. The access tier is set to 'Hot (Inferred)'. The blob type is 'Block blob' and the size is 60.47 KB. The lease state is 'Available'.

Transformation



The Azure portal screen shows a list of recently accessed resources, including a Data Factory, Storage Account, and Databricks service. Click on the “**NexusDatabricks**” resource group to view or manage related Azure resources.

The screenshot shows the Microsoft Azure portal interface. The top navigation bar includes tabs for 'ADFnexusdatalens - Microsoft', 'bronze - Microsoft Azure', 'ADFnexusdatalens - Azure Data Factory', and 'Data Transformation for Power ...'. The main content area displays the 'Azure services' ribbon with various icons for creating resources, managing subscriptions, and navigating to resource groups, Azure Synapse Analytics, Budgets, Bot Services, Projects, Quickstart Center, Azure AI services, and more.

The 'Resources' section is titled 'Recent' and lists the following resources:

Name	Type	Last Viewed
ADFnexusdatalens	Data factory (V2)	a few seconds ago
Nexusteststorage	Storage account	8 hours ago
NexusProject	Resource group	8 hours ago
nekuskeyvault	Key vault	9 hours ago
nekusdatabricks	Azure Databricks Service	3 days ago
Azure for Stud	Subscription	4 days ago

Below the resources, there are sections for 'Navigate' (Subscriptions, Resource groups, All resources, Dashboard) and 'Tools' (Microsoft Learn, Azure Monitor, Microsoft Defender for Cloud, Cost Management).

Nexus DataLens

The Azure Databricks service overview for the “**nexusdatabricks**” resource, with an active status, **launch the workspace**.

This screenshot shows the Azure portal interface for the 'nexusdatabricks' Azure Databricks Service. The left sidebar includes links for Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Resource visualizer, Settings, Monitoring, Automation, and Help. The main content area displays the 'Essentials' section with the following details:

- Status: Active
- Resource group: [nexusProject](#)
- Location: Central India
- Subscription: [Azure for Students](#)
- Subscription ID: [d194a87-26b0-47d7-827d-3471677a4a02](#)
- Tags: [Add tags](#)

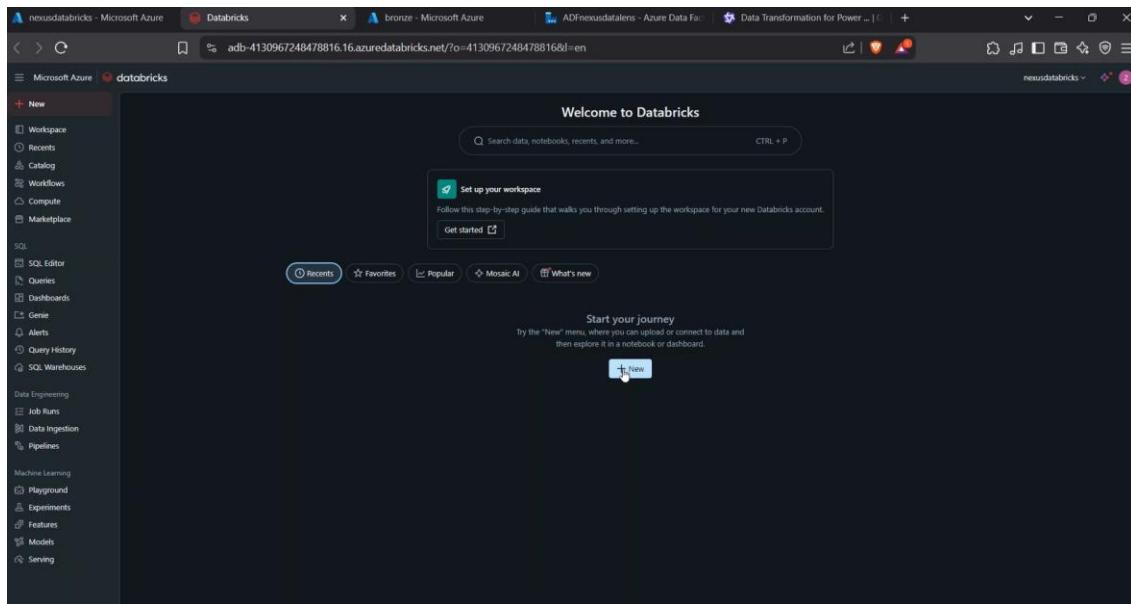
Managed Resource Group: [databricks-rg-nexusdatabricks-bmfozadip0k](#)
URL: <https://adb-4130967248478816.16.azuredatabricks.net>
Pricing Tier: [Premium + Role-based access control \(Click to change\)](#)

A large red 'Launch Workspace' button is prominently displayed. Below it, there are six cards with icons and labels: Documentation, Getting Started, Import Data from file, Import Data from Azure Storage, notebook, Admin Guide, and Link Azure ML workspace.

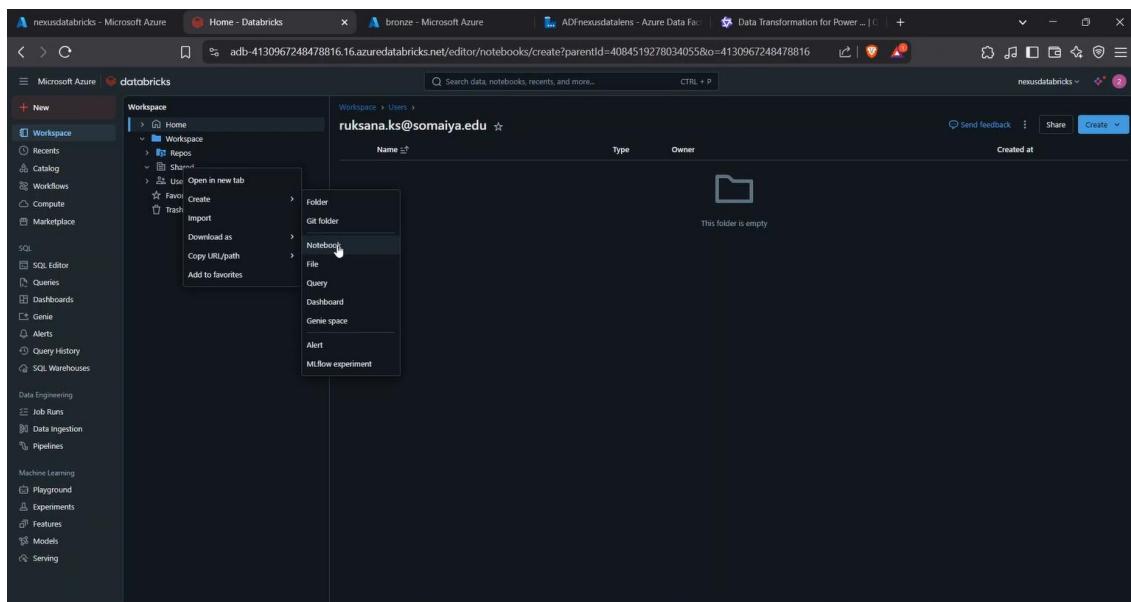
This screenshot shows the 'Sign In to Databricks' page. It features a logo and the text 'Azure Databricks'. Below that is a 'Sign In to Databricks' button with the sub-instruction 'Sign in using Microsoft Entra ID Single Sign On.' A note at the bottom says 'Contact your site administrator to request access.' At the very bottom are links for 'Privacy policy' and 'Terms of use'.

Nexus DataLens

The Databricks workspace home screen where users can start their data journey by creating notebooks, dashboards, or other resources via the "New" button.

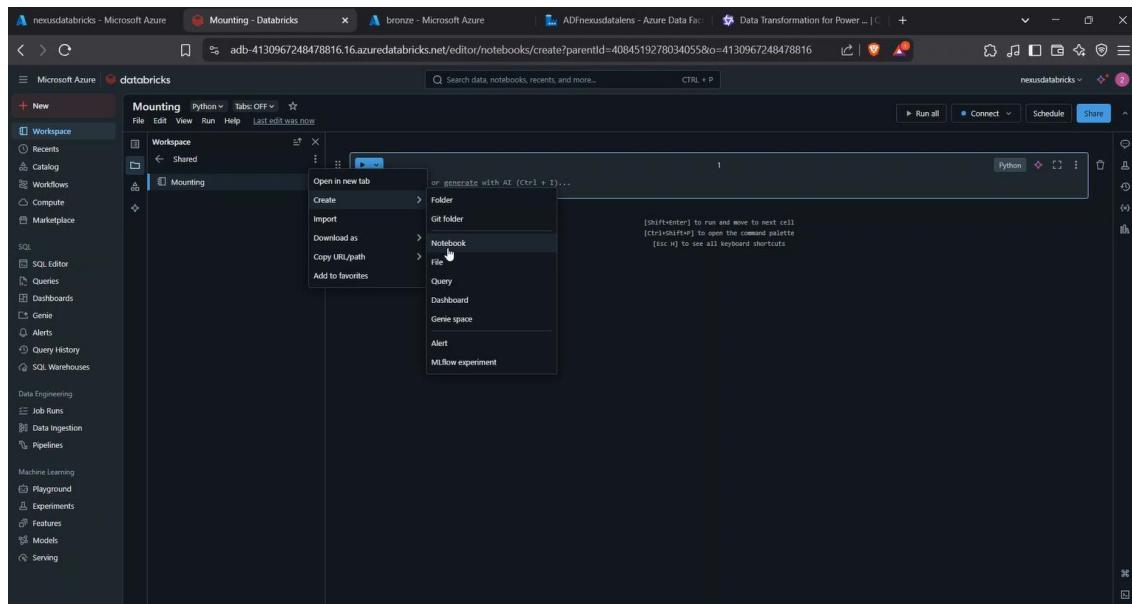


Create a new Databricks notebook from the Workspace section by selecting "Create > Notebook."

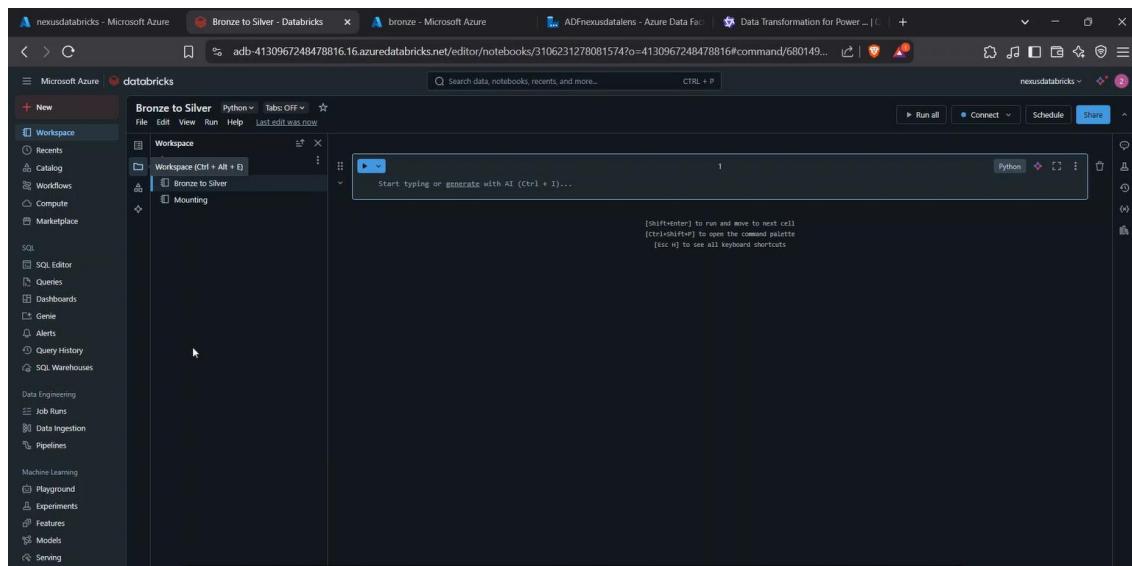


Nexus DataLens

Open Databricks notebook then naming the notebook as “**Mounting**” and using the workspace sidebar menu to create another new notebook.

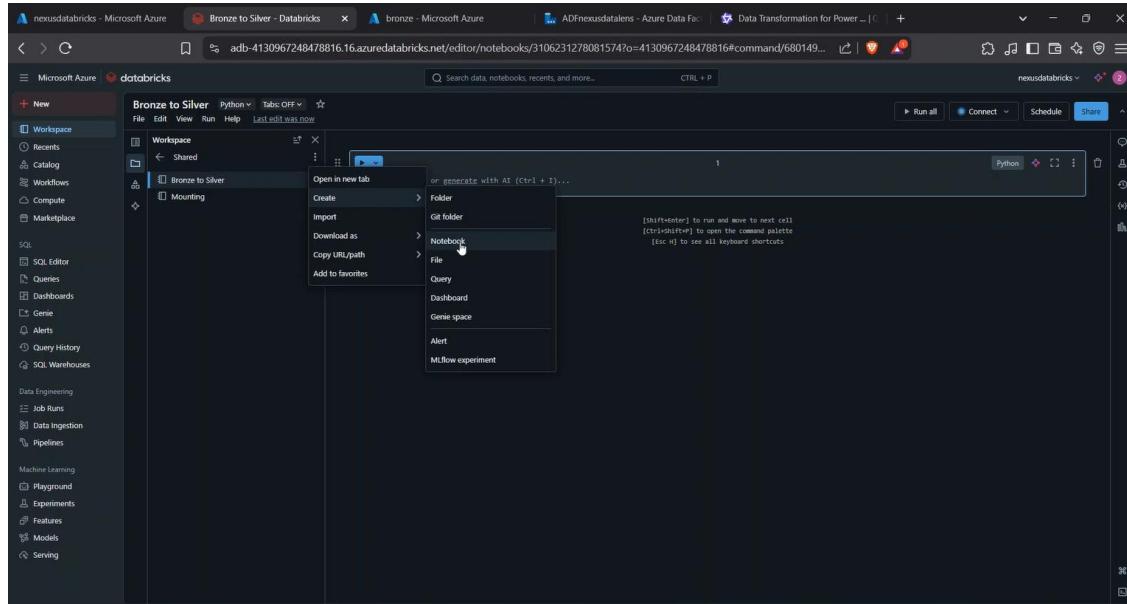


A Databricks notebook named "**Bronze to Silver**" open and ready for code input in the Python language.

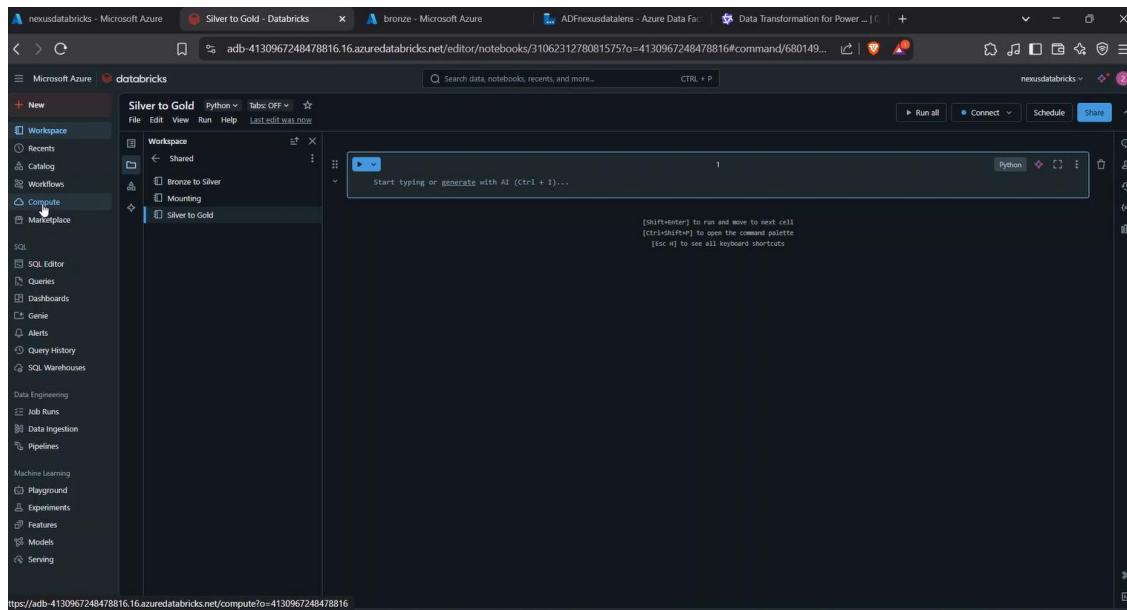


Nexus DataLens

The user in Databricks selecting the option to create a new notebook from the "Create" menu under the Workspace.

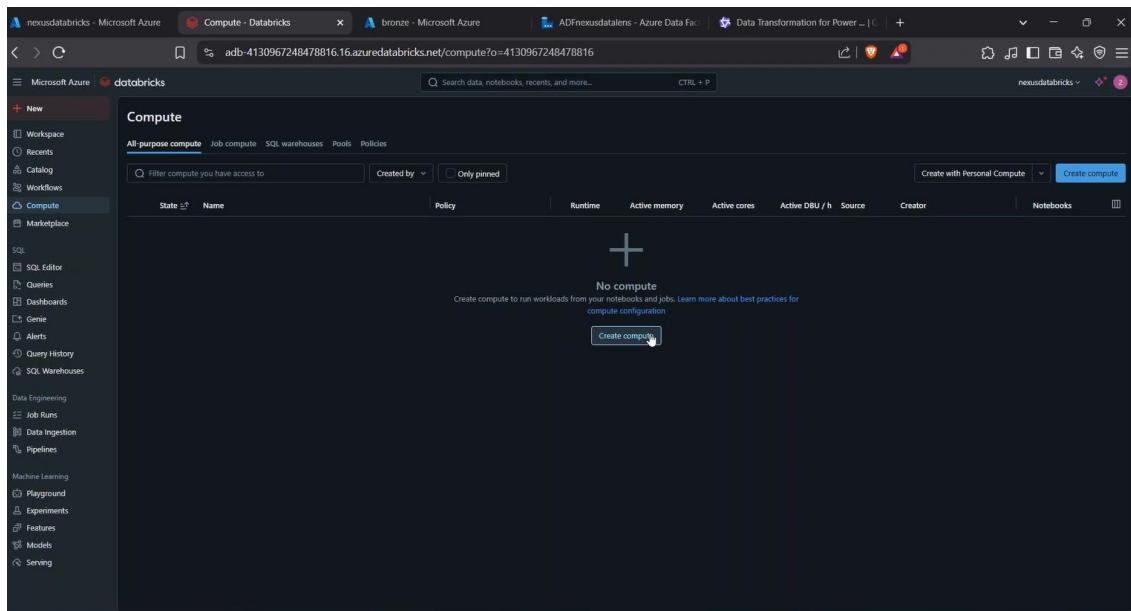


A Databricks notebook named "**Silver to Gold**" open and ready for code input in the Python language.

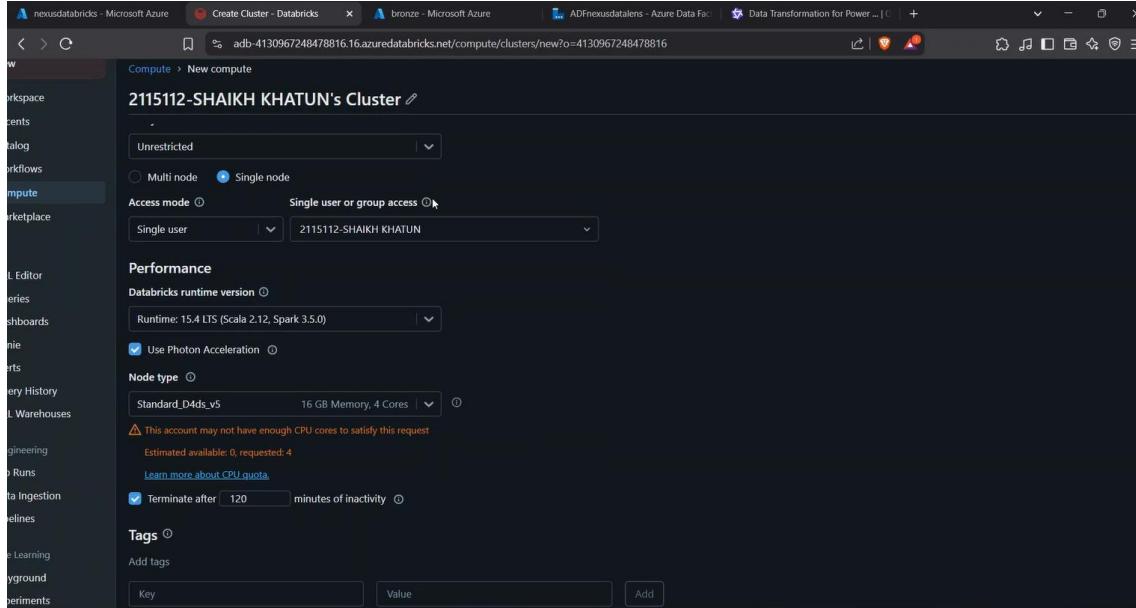


Nexus DataLens

The Databricks **Compute** page in Azure, where a user is about to create a new compute cluster by clicking **Create compute**.

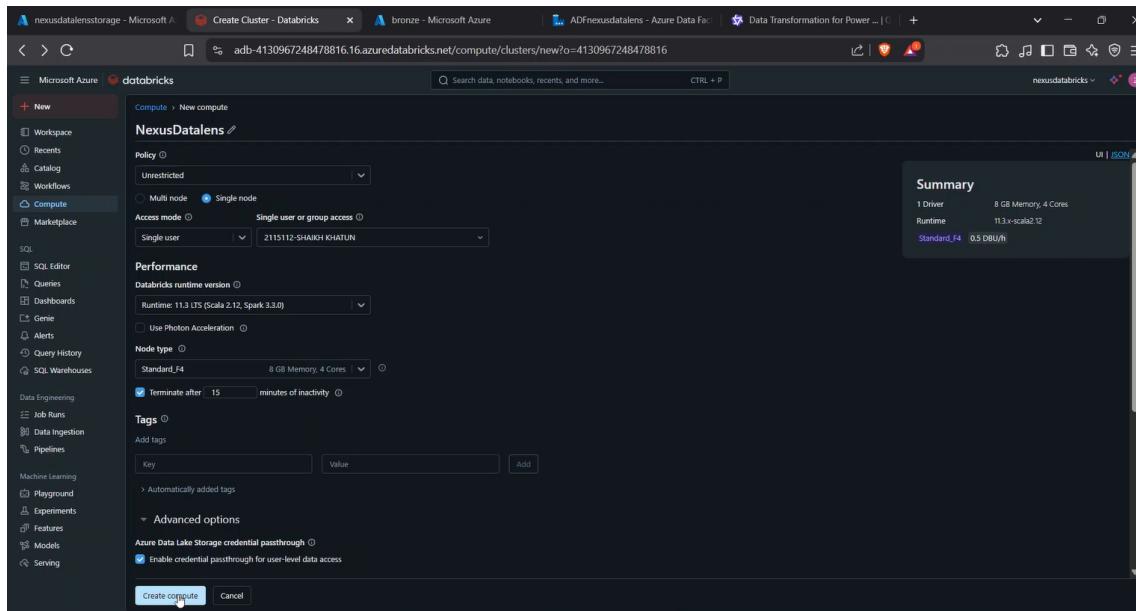


The configuration page for creating a new Databricks cluster in Azure, where the user is setting options like cluster name, access mode, runtime version, and node type. Select “**Single node**”

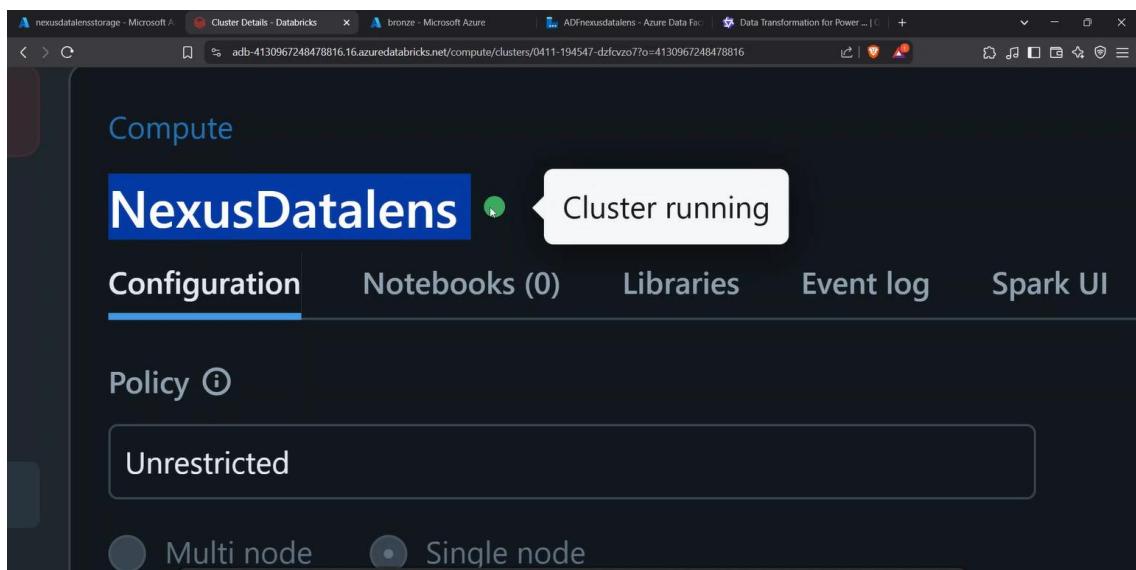


Nexus DataLens

The final step of creating a single-node Azure Databricks cluster named **NexusDataLens** with **Runtime 11.3 LTS (Scala 2.12, Spark 3.3.0)**, **Standard_F4** node type (8 GB memory, 4 cores), 15-minute auto-termination, and Azure Data Lake credential passthrough enabled before clicking **Create Compute**.

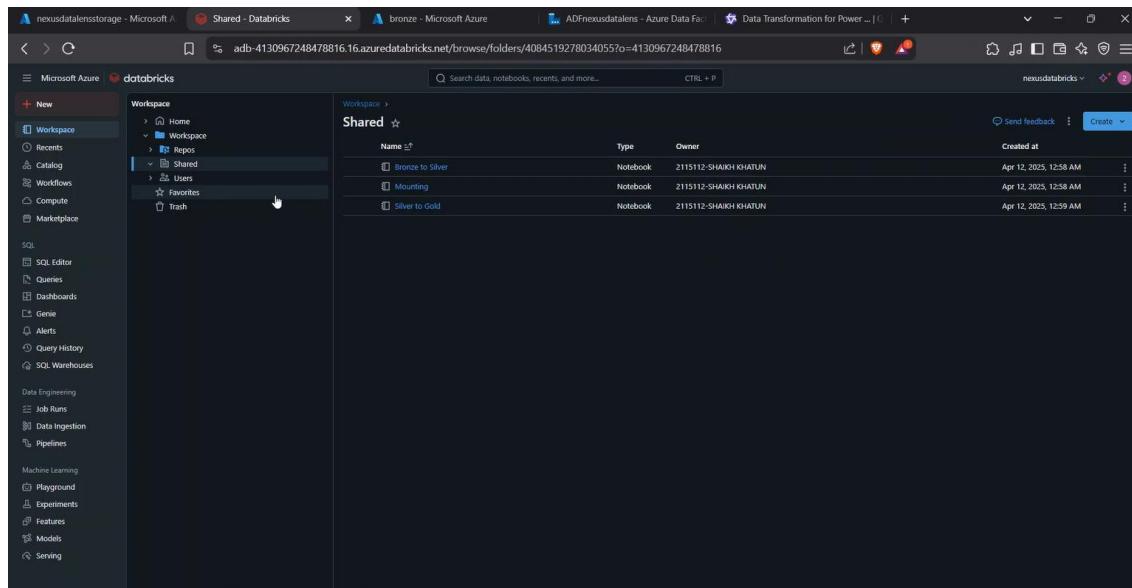


This confirms that the **NexusDataLens** Databricks cluster is successfully running with an **Unrestricted** policy and accessible configuration options.



Nexus DataLens

The **Shared** workspace folder in Azure Databricks containing three notebooks: **Bronze to Silver**, **Mounting**, and **Silver to Gold**, all owned by the same user.



The screenshot shows the Azure Databricks workspace interface. On the left, there is a sidebar with various navigation options like Workspace, Catalog, Workflows, Compute, Marketplace, SQL, Machine Learning, and Data Engineering. The main area shows a list of notebooks in the 'Shared' workspace. The list includes:

Name	Type	Owner	Created at
Bronze to Silver	Notebook	2115112-SHAIKH KHATUN	Apr 12, 2025, 12:58 AM
Mounting	Notebook	2115112-SHAIKH KHATUN	Apr 12, 2025, 12:58 AM
Silver to Gold	Notebook	2115112-SHAIKH KHATUN	Apr 12, 2025, 12:59 AM

Mounting Code

```
# Databricks notebook source
# MAGIC %md
# MAGIC Mounting Bronze container

# COMMAND -------

# Databricks notebook source
configs = {
    "fs.azure.account.auth.type": "CustomAccessToken",
    "fs.azure.account.custom.token.provider.class": spark.conf.get("spark.databricks.passthrough.adls.gen2.tokenProviderClassName")
}

# COMMAND -------

# Optionally, you can add <directory-name> to the source URI of your mount point.
dbutils.fs.mount(
    source = "abfss://bronze@nexusdatalensstorage.dfs.core.windows.net/",
    mount_point = "/mnt/bronze",
    extra_configs = configs)

# COMMAND -------

dbutils.fs.ls("/mnt/bronze/")

# COMMAND -------

# MAGIC %md
# MAGIC Mounting Silver container

# COMMAND -------

configs = {
    "fs.azure.account.auth.type": "CustomAccessToken",
    "fs.azure.account.custom.token.provider.class": spark.conf.get("spark.databricks.passthrough.adls.gen2.tokenProviderClassName")
}

# Optionally, you can add <directory-name> to the source URI of your mount point.
dbutils.fs.mount(
    source = "abfss://silver@nexusdatalensstorage.dfs.core.windows.net/",
    mount_point = "/mnt/silver",
    extra_configs = configs)

# COMMAND -------

dbutils.fs.ls("/mnt/silver/")

# COMMAND -------

# MAGIC %md
# MAGIC Mounting Gold Container

# COMMAND -------

configs = {
    "fs.azure.account.auth.type": "CustomAccessToken",
    "fs.azure.account.custom.token.provider.class": spark.conf.get("spark.databricks.passthrough.adls.gen2.tokenProviderClassName")
}

# Optionally, you can add <directory-name> to the source URI of your mount point.
dbutils.fs.mount(
    source = "abfss://gold@nexusdatalensstorage.dfs.core.windows.net/",
    mount_point = "/mnt/gold",
    extra_configs = configs)
```

Nexus DataLens

Code Screenshot:

The image displays three screenshots of an Azure Databricks notebook interface, illustrating the process of handling data from a bronze layer to a silver layer.

Screenshot 1 (Top): Shows the initial setup where a bronze layer directory is mounted to a silver layer directory. The code executed is:

```
dbutils.fs.ls("/mnt/bronze/")
Out[1]: [FileInfo(path="dbfs:/mnt/bronze/dbo.data.txt", name='dbo.data.txt', size=61921, modificationTime=1744368628000)]
```

Screenshot 2 (Middle): Shows the extraction of a CSV file from the mounted bronze layer. The code executed is:

```
configs = {
    "fs.azure.account.auth.type": "CustomAccessToken",
    "fs.azure.account.custom.token.provider.class": spark.conf.get("spark.databricks.passthrough.adls.gen2.tokenProviderClassName")
}

# Optionally, you can add <directory-name> to the source URI of your mount point.
dbutils.fs.mount(
    source = "abfss://silver@nexusdatalensstorage.dfs.core.windows.net/",
    mount_point = "/mnt/silver",
    extra_configs = configs)
Out[2]: True
```

Screenshot 3 (Bottom): Shows the creation of a DataFrame from the extracted CSV file and basic data cleaning. The code executed is:

```
df_csv = spark.read.format("csv").option("header", "true").option("inferSchema", "true").option("delimiter", ",").load(csv_file_path)

# Display the schema and some data to verify
from pyspark.sql import SparkSession
from pyspark.sql.functions import col, when, regexp_replace, to_date, year, avg, round
(df_csv.show(2))
Out[3]:
+---+---+
|Title|Genre|
+---+---+
|  1 |  A |
|  2 |  B |
+---+---+
```

Screenshot 4 (Second Middle): Shows the transformation of the DataFrame to include IMDB scores. The code executed is:

```
df_csv = df_csv.withColumn("IMDB_Score", round(col("IMDB_Score").cast("double"), 1))
(df_csv.show(2))
Out[4]:
+---+---+-----+
|Title|Genre|IMDB_Score|
+---+---+-----+
|  1 |  A |      1.0 |
|  2 |  B |      2.0 |
+---+---+-----+
```

Screenshot 5 (Bottom Middle): Shows the handling of missing values in the DataFrame. The code executed is:

```
# Step 1: Handle Missing Values
# Replace NULL values in the 'Language' column with "English"
df_csv = df_csv.withColumn("Language", when(col("Language").isNull(), "English").otherwise(col("Language")))

# Replace NULL values in the 'Runtime' column with 0
df_csv = df_csv.withColumn("Runtime", when(col("Runtime").isNull(), 0).otherwise(col("Runtime")))

# Replace NULL values in the 'Views' column with 0
df_csv = df_csv.withColumn("Views", when(col("Views").isNull(), 0).otherwise(col("Views")))

# Replace NULL values in the 'Title' column with "NexusDataLens null value"
df_csv = df_csv.withColumn("Title", when(col("Title").isNull(), "NexusDataLens null value").otherwise(col("Title")))
(df_csv.show(2))
Out[5]:
+---+---+-----+-----+
|Title|Genre|IMDB_Score|Language|
+---+---+-----+-----+
|  1 |  A |      1.0 | English |
|  2 |  B |      2.0 | English |
+---+---+-----+-----+
```

Screenshot 6 (Bottom): Shows the final step of saving the transformed DataFrame as a Parquet file. The code executed is:

```
output_path_parquet = "/mnt/silver/transformed_data.parquet"
df_csv.write.format("parquet").mode("overwrite").save(output_path_parquet)
(1) Spark Jobs
```

Silver to Gold Code

```

# Databricks notebook source
# Define the path to the Parquet file in the Silver container
parquet_file_path = "/mnt/silver/transformed_data.parquet"
# Load the Parquet file into a DataFrame
df_silver = spark.read.format("parquet").load(parquet_file_path)

# COMMAND -----
from pyspark.sql.functions import col, when

# Add a column to classify movies as "Best" or "Worst" based on IMDB Score
df_silver = df_silver.withColumn(
    "IMDB Recommendation",
    when(col("IMDB_Score") >= 7, "Best")
    .when(col("IMDB_Score") <= 5, "Worst")
    .otherwise("Average")
)
# COMMAND -----

from pyspark.sql.functions import mean, when
# Calculate average IMDB Score
avg_imdb_score = df_silver.select(mean("IMDB_Score")).collect()[0][0]

# Fill missing IMDB scores with the average
df_silver = df_silver.withColumn("IMDB_Score", when(df_silver["IMDB_Score"].isNull(), avg_imdb_score).otherwise(df_silver["IMDB_Score"]))

# COMMAND -----
from pyspark.sql.functions import split, explode

# Split genres by "/"
df_silver = df_silver.withColumn("Genres", split("Genre", "/"))
df_silver = df_silver.withColumn("Genre", explode("Genres")).drop("Genres")

# COMMAND -----

from pyspark.sql.functions import col, split

# Split the 'Language' column by '/' and retain only the first part
df_silver = df_silver.withColumn(
    "Language",
    split(col("Language"), "/").getItem(0)
)
# COMMAND -----

import re

# Function to clean column names
def clean_column_name(name):
    # Remove leading/trailing whitespace
    name = name.strip()
    # Replace spaces with underscores
    name = name.replace(" ", "_")
    # Remove invalid characters (e.g., invisible ones)
    name = re.sub(r'[^\w]', '', name) # Keep only alphanumeric and underscores
    return name.lower() # Optional: Convert to lowercase for consistency

# Rename all columns
cleaned_columns = {col: clean_column_name(col) for col in df_silver.columns}
for old_name, new_name in cleaned_columns.items():
    df_silver = df_silver.withColumnRenamed(old_name, new_name)
# Verify the cleaned column names

# COMMAND -----

dbutils.fs.rm(output_path_delta, recurse=True)
# Define the output path in the Gold container
output_path_delta = "/mnt/gold/transformed_data.delta"
df_silver.write.format("delta").save(output_path_delta)

```

Nexus DataLens

Code Screenshot:

The screenshot displays a sequence of six code cells (labeled 1 through 6) in a Python notebook, illustrating the transformation of movie data from a Silver container to a Gold container.

- Cell 1:** Loads a Parquet file from the Silver container into a DataFrame.

```
# Define the path to the Parquet file in the Silver container
parquet_file_path = "/mnt/silver/transformed_data.parquet"
# Load the Parquet file into a DataFrame
df_silver = spark.read.format("parquet").load(parquet_file_path)
```
- Cell 2:** Adds a column to classify movies based on IMDB Score.

```
from pyspark.sql.functions import col, when

# Add a column to classify movies as "Best" or "Worst" based on IMDB Score
df_silver = df_silver.withColumn(
    "IMDB Recommendation",
    when(col("IMDB_Score") >= 7, "Best")
    .when(col("IMDB_Score") <= 5, "Worst")
    .otherwise("Average")
)
```
- Cell 3:** Calculates the average IMDB score for each movie.

```
from pyspark.sql.functions import mean, when
# Calculate average IMDB Score
avg_imdb_score = df_silver.select(mean("IMDB_Score")).collect()[0][0]

# Fill missing IMDB scores with the average
df_silver = df_silver.withColumn("IMDB_Score", when(df_silver["IMDB_Score"].isNotNull(), avg_imdb_score).otherwise(df_silver["IMDB_Score"]))
```
- Cell 4:** Splits the genres column by '/' and retains only the first part.

```
from pyspark.sql.functions import split, explode

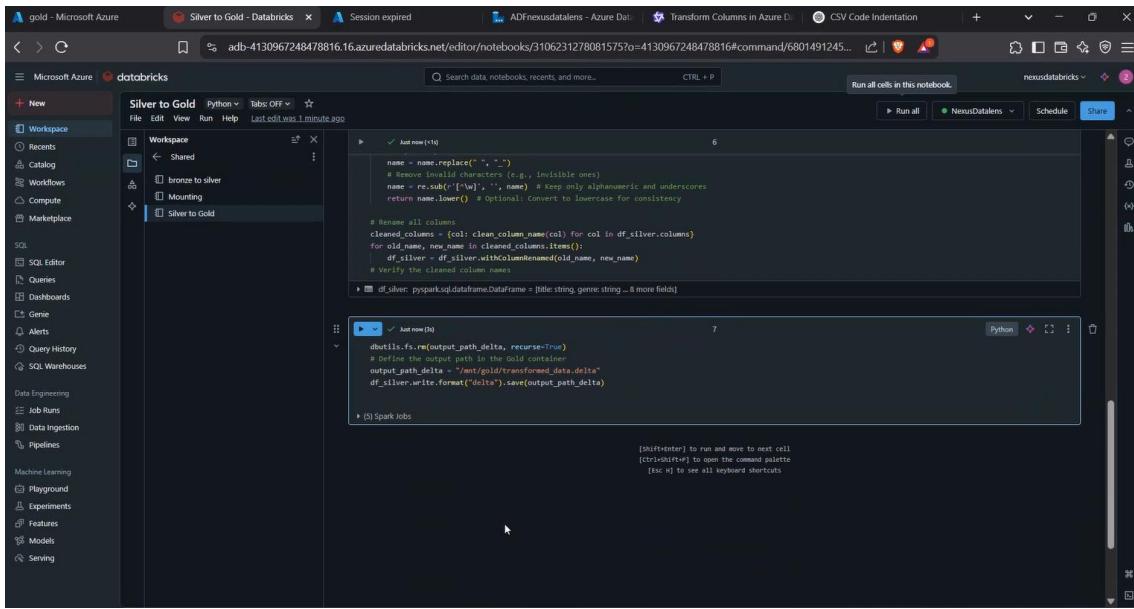
# Split genres by "/"
df_silver = df_silver.withColumn("Genres", split("Genre", "/"))
df_silver = df_silver.withColumn("Genre", explode("Genres")).drop("Genres")
```
- Cell 5:** Splits the 'Language' column by '/' and retains only the first part.

```
from pyspark.sql.functions import col, split

# Split the 'Language' column by '/' and retain only the first part
df_silver = df_silver.withColumn(
    "Language",
    split(col("Language"), "/").getItem(0)
)
```
- Cell 6:** Drops duplicate rows based on the 'Title' column.

```
df_silver = df_silver.dropDuplicates(["Title"])
```

Nexus DataLens



Bronze to Silver Code:

```
# Databricks notebook source
dbutils.fs.ls("/mnt/bronze/")
csv_file_path = f"/mnt/bronze/dbo.data.txt"
# Load the CSV file into a DataFrame
df_csv = spark.read.format("csv").option("header", "true").option("inferSchema", "true").option("delimiter", ",").load(csv_file_path)
# Display the schema and some data to verify

from pyspark.sql import SparkSession
from pyspark.sql.functions import col, when, regexp_replace, to_date, year, avg, round

# COMMAND -----
df_csv = df_csv.withColumn("IMDB_Score", round(col("IMDB_Score").cast("double"), 1))

# COMMAND -----
df_csv = df_csv.withColumn("ReleaseDate", to_date(col("ReleaseDate"), "yyyy-MM-dd"))
df_csv = df_csv.withColumn("AddedDate", to_date(col("AddedDate"), "yyyy-MM-dd"))
df_csv = df_csv.withColumn("ReleaseYear", year(col("ReleaseDate")))

# COMMAND -----
# Step 1: Handle Missing Values
# Replace NULL values in the 'Language' column with "English"
df_csv = df_csv.withColumn("Language", when(col("Language").isNull(), "English").otherwise(col("Language")))

# Replace NULL values in the 'Runtime' column with 0
df_csv = df_csv.withColumn("Runtime", when(col("Runtime").isNull(), 0).otherwise(col("Runtime")))

# Replace NULL values in the 'Views' column with 0
df_csv = df_csv.withColumn("Views", when(col("Views").isNull(), 0).otherwise(col("Views")))

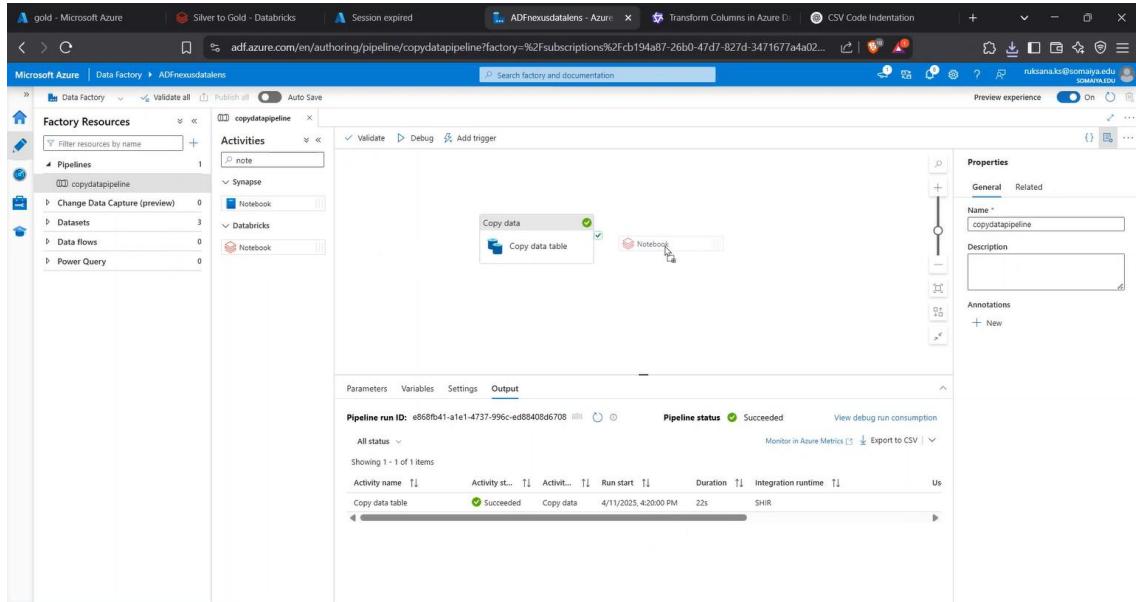
# Replace NULL values in the 'Title' column with "NexusDataLens null value"
df_csv = df_csv.withColumn("Title", when(col("Title").isNull(), "No Title").otherwise(col("Title")))

# COMMAND -----
output_path_parquet = "/mnt/silver/transformed_data.parquet"

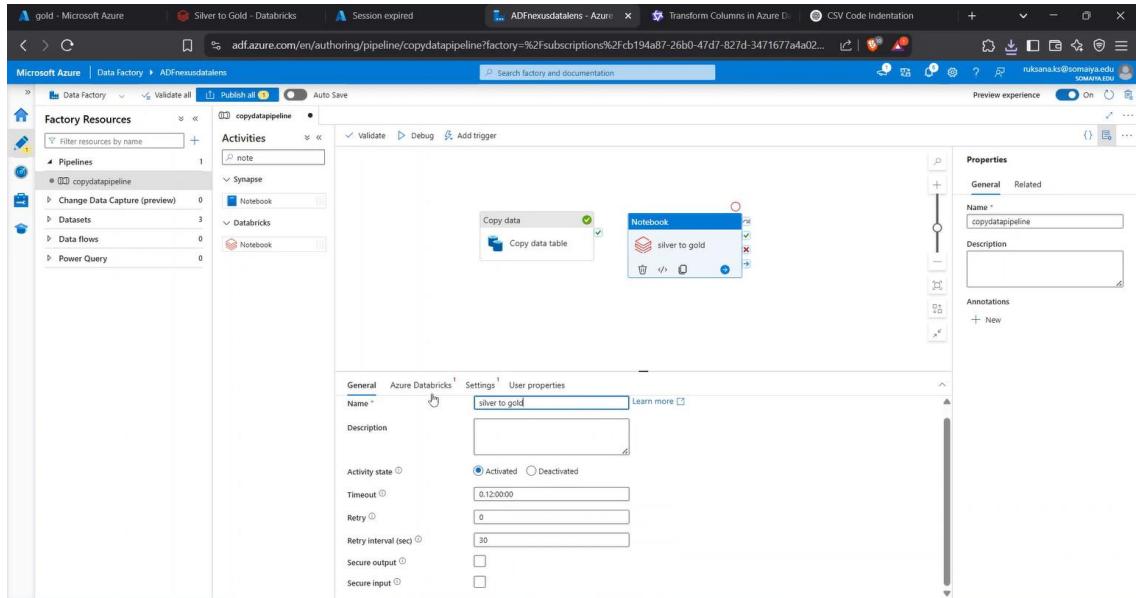
# Save the transformed DataFrame as Parquet
df_csv.write.format("parquet").mode("overwrite").save(output_path_parquet)
```

Nexus DataLens

An Azure Data Factory pipeline named **copydatapipeline**, where a **Copy data** activity has successfully run to copy a data table in 22 seconds.

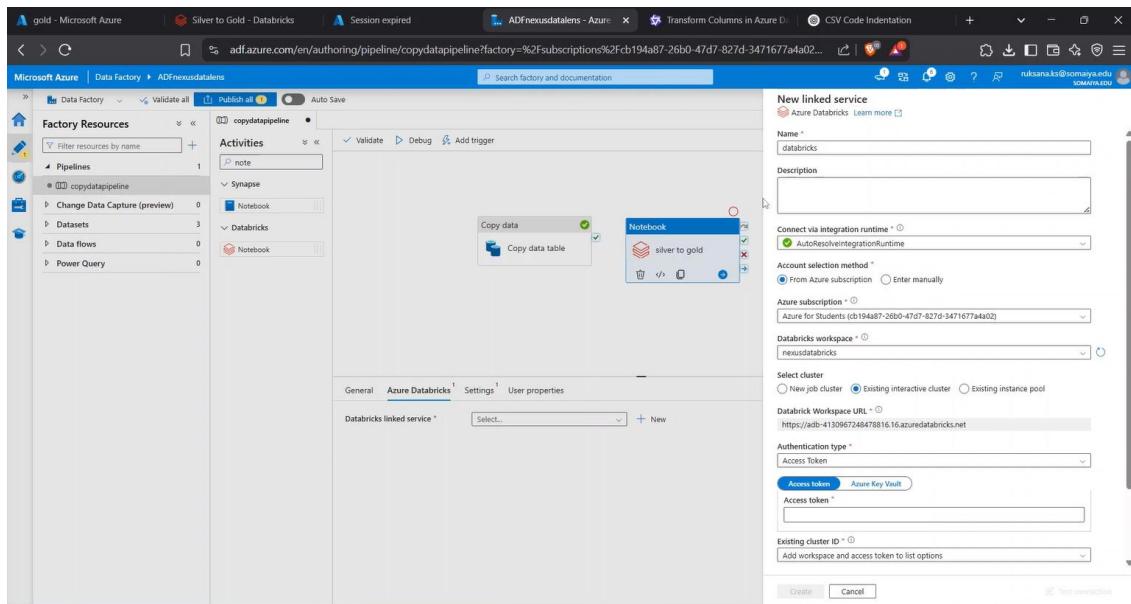


An Azure Data Factory pipeline where a **Databricks Notebook activity** named **Silver to Gold** is being configured. It follows a successfully executed **Copy data** activity, enabling a notebook-driven data transformation step.

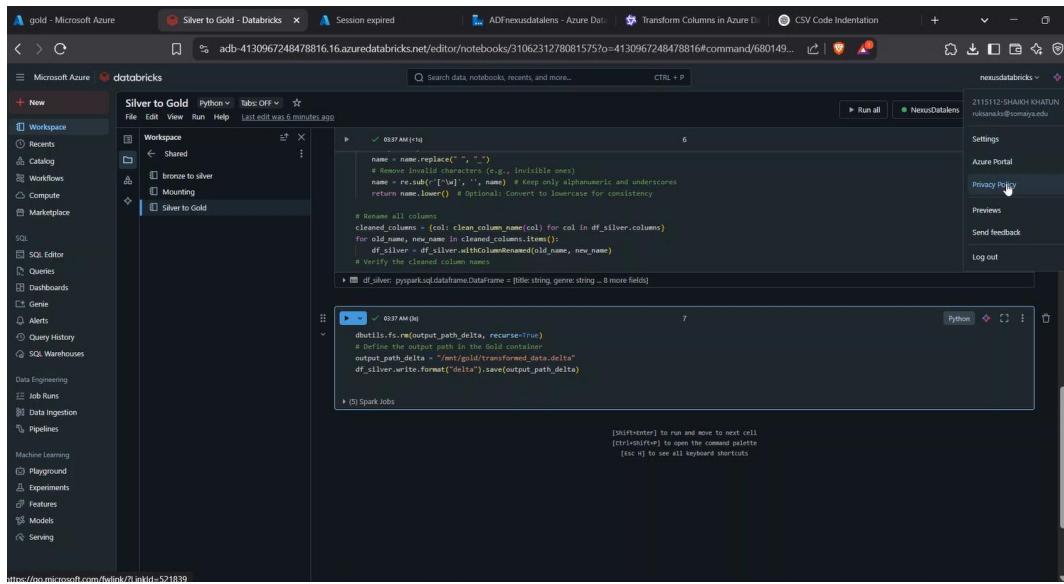


Nexus DataLens

The setup of a new Azure Databricks linked service in Azure Data Factory, including configuration of workspace, authentication, and cluster options.



The image shows a Databricks notebook transforming and writing a cleaned DataFrame from Silver to Gold storage using PySpark.



Nexus DataLens

The Developer settings page in Databricks where users can manage access tokens and configure editor settings like Spark tips and dark mode.

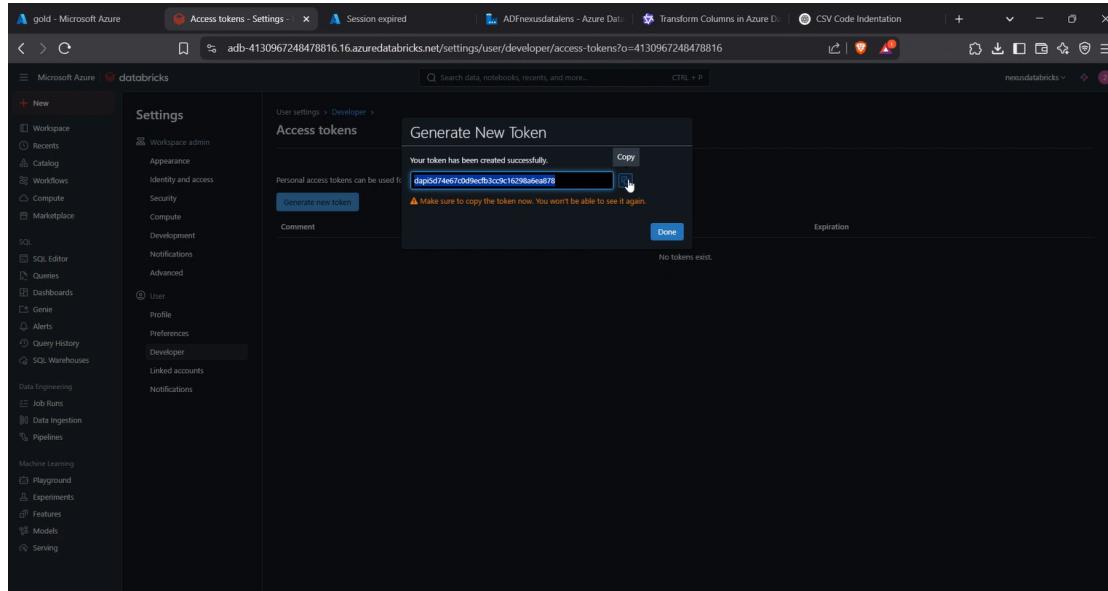
This screenshot shows the 'Developer' settings page in Databricks. The left sidebar has 'Developer' selected under 'User'. The main area is titled 'Developer' and 'Manage your development settings'. It includes sections for 'Access tokens' (with a 'Manage' button), 'SQL query snippets' (with a 'SQL Editor' link), and 'Editor settings'. Under 'Editor settings', there are several toggle switches: 'Spark tips' (On), 'Databricks Advisor' (On), 'Automatically launch and attach to clusters' (Off), 'Command delete confirmation' (On), and 'Dark mode' (Off). A URL at the bottom of the page is <https://adb-4130967248478816.16.azuredatabricks.net/settings/user/developer/access-tokens>.

The image shows the process of generating a new personal access token in Databricks, which is used for secure authentication in integrations like Azure Data Factory.

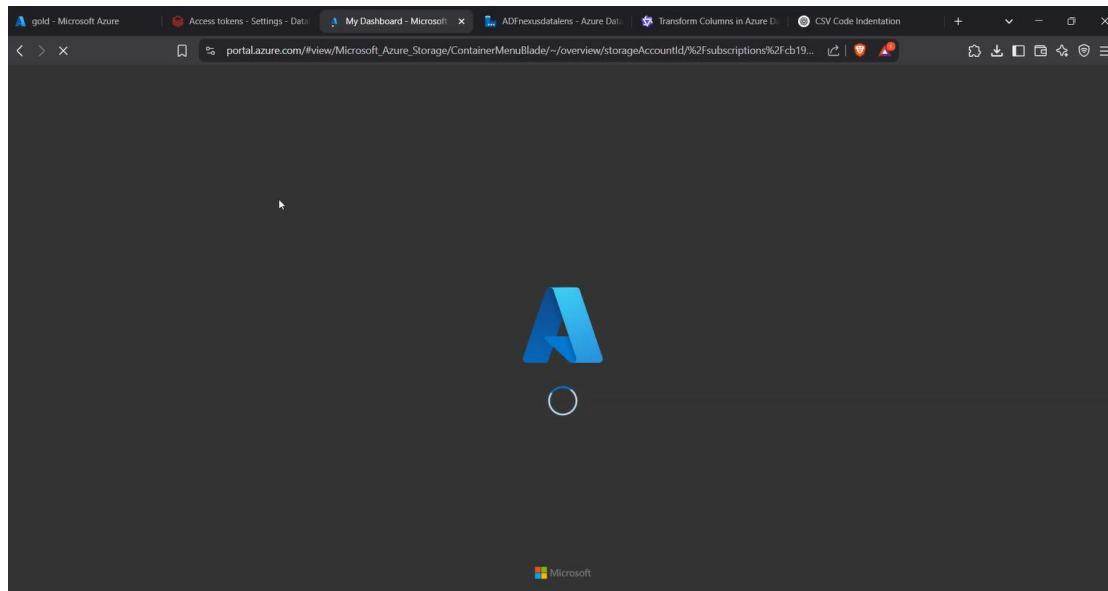
This screenshot shows the 'Generate new token' dialog in Databricks. It's a modal window with fields for 'Comment' (containing 'ADF') and 'Lifetime (days)' (set to 90). There are 'Cancel' and 'Generate' buttons at the bottom. The background shows the 'Access tokens' section of the developer settings page.

Nexus DataLens

The image shows a newly generated Databricks access token, which must be copied immediately as it won't be visible again for future use.

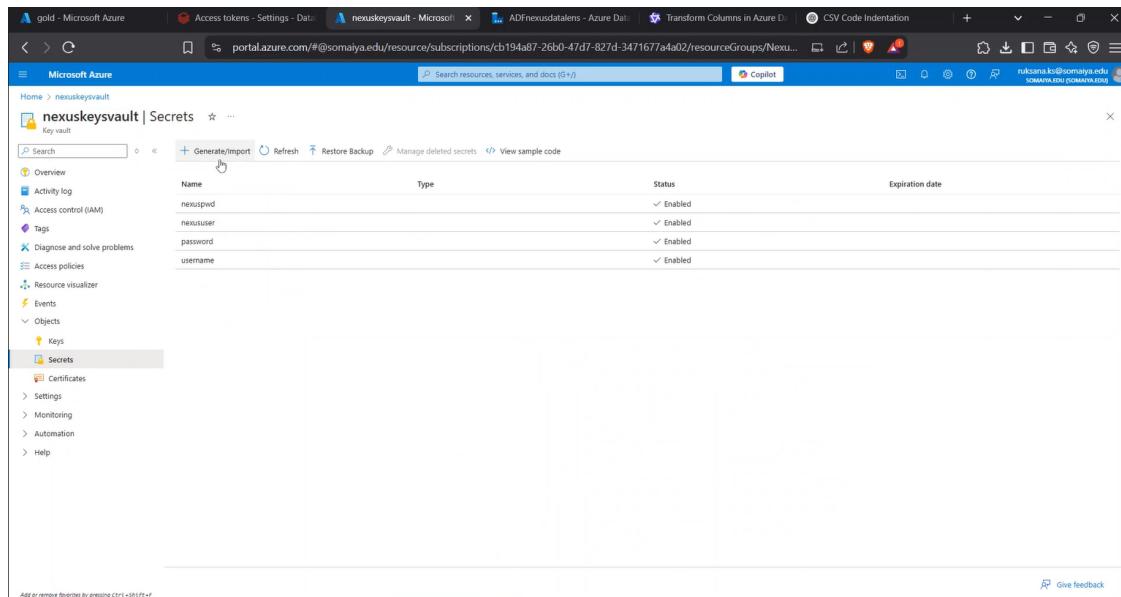


The Azure portal loading screen, likely indicating access to a storage container or resource for integration with Databricks or Data Factory.



Nexus DataLens

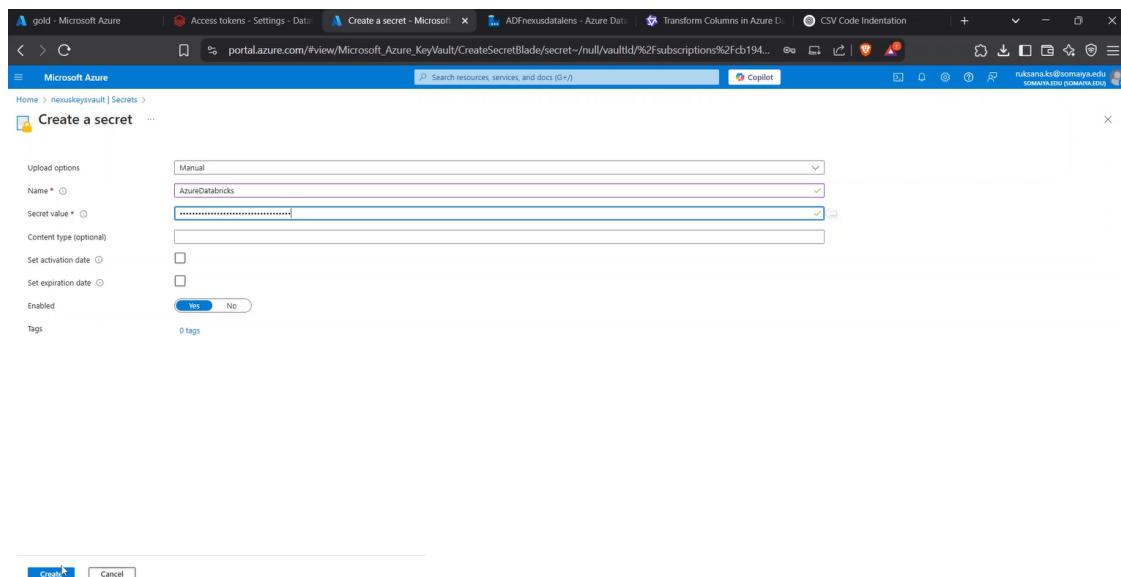
The image shows Azure Key Vault with secrets (like usernames and passwords) stored and enabled for secure access by integrated services like Azure Data Factory and Databricks.



A screenshot of the Microsoft Azure portal showing the 'Secrets' blade for the 'nexuskeysvault'. The page displays four secrets:

Name	Type	Status	Expiration date
nexuspwd		✓ Enabled	
nexususer		✓ Enabled	
password		✓ Enabled	
username		✓ Enabled	

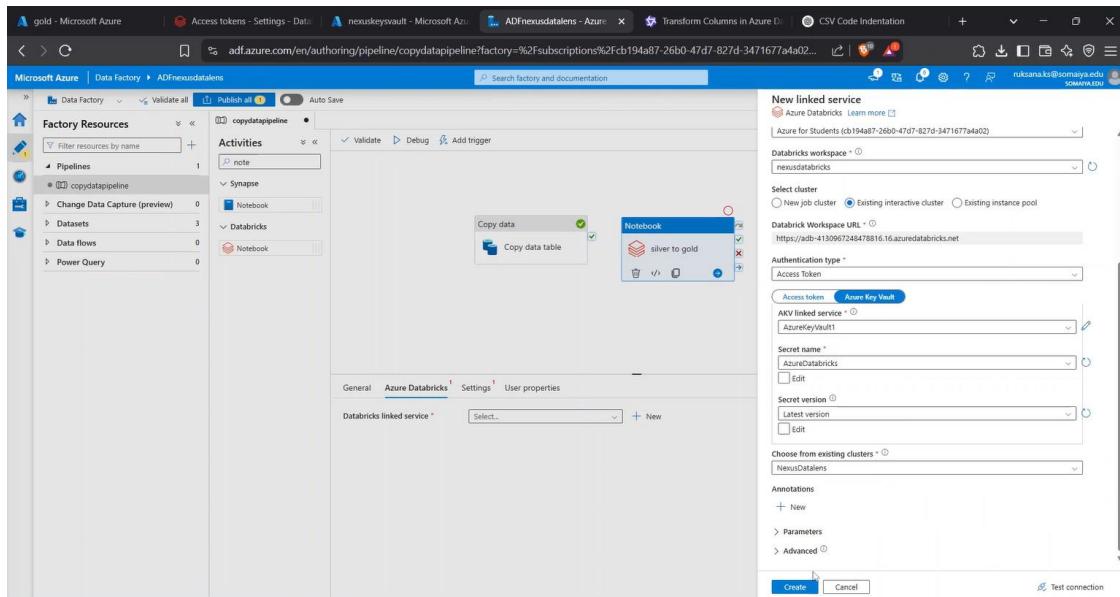
The image shows the creation of a new secret named "AzureDatabricks" in Azure Key Vault with manual upload and the secret value entered.



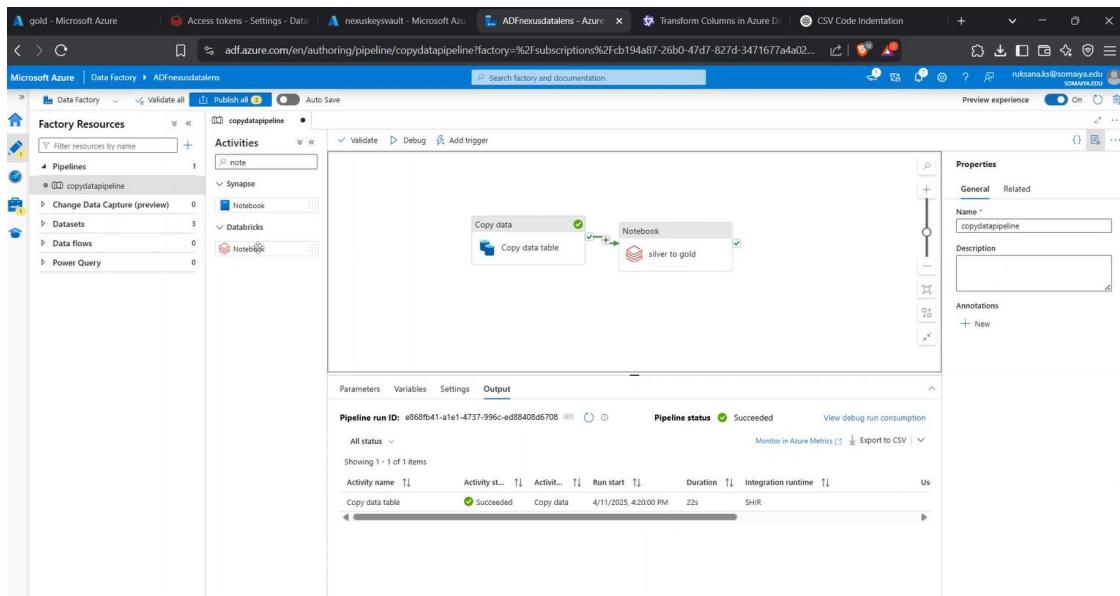
A screenshot of the 'Create a secret' blade in the Microsoft Azure portal. The 'Name' field is populated with 'AzureDatabricks'. The 'Content type (optional)' field is empty. Other fields include 'Upload options' (set to 'Manual'), 'Set activation date' (unchecked), 'Set expiration date' (unchecked), 'Enabled' (set to 'Yes'), and 'Tags' (empty). At the bottom are 'Create' and 'Cancel' buttons.

Nexus DataLens

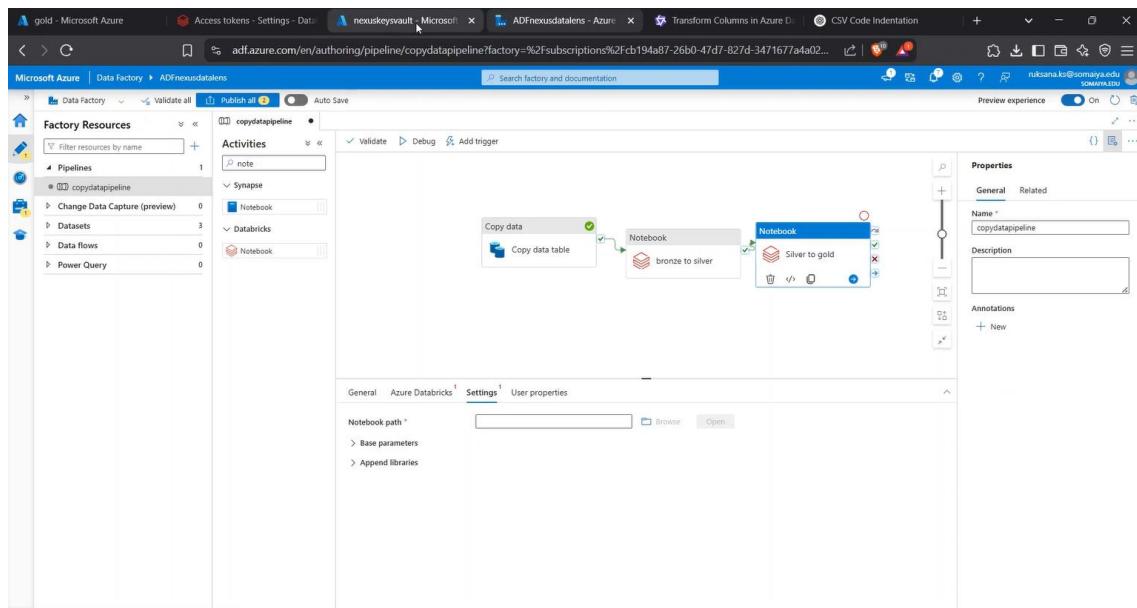
The creation of a new linked service in Azure Data Factory to connect with an Azure Databricks workspace (**nexusdatabricks**) using Azure Key Vault for token authentication. The Key Vault (**Azurekeyvault1**) is configured to retrieve the token stored under the secret name **AzureDatabricks**.



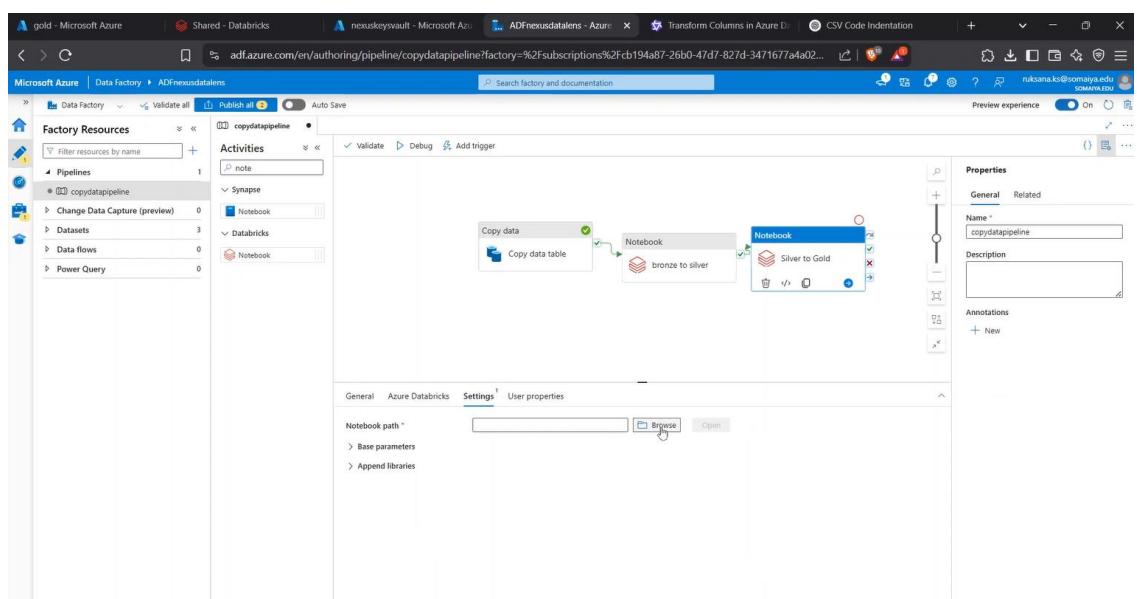
The screenshot shows a successful connection test to the Databricks linked service (**databricks**) in the Azure Data Factory pipeline named **copydatapipeline**.



Nexus DataLens

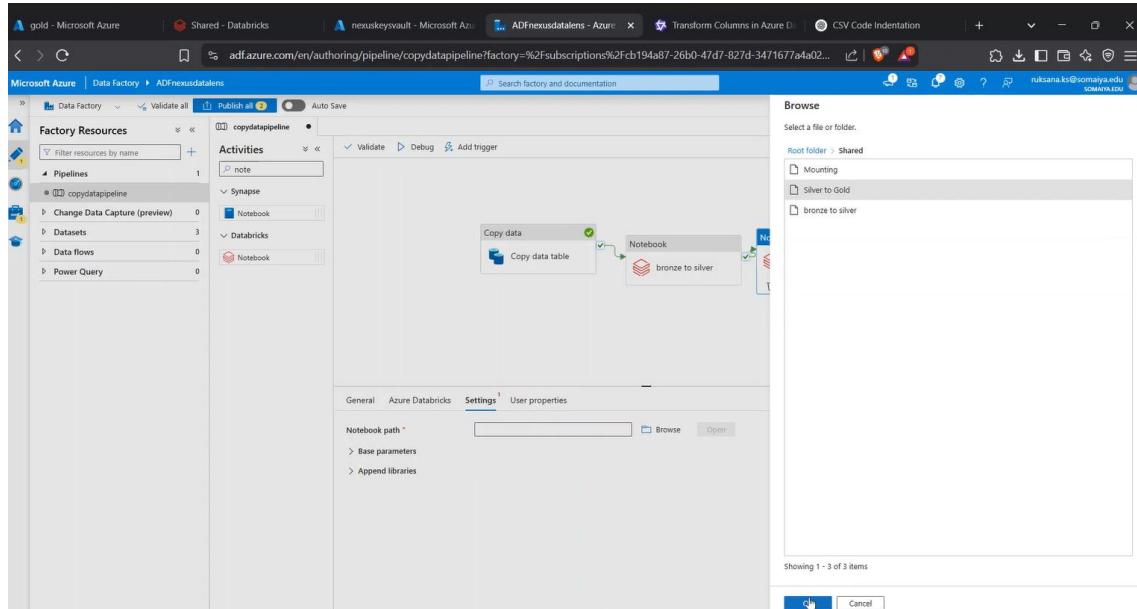


The Azure Data Factory pipeline `copydatapipeline` successfully connects to the databricks linked service to run two sequential notebooks: "**bronze to silver**" and "**silver to gold**"

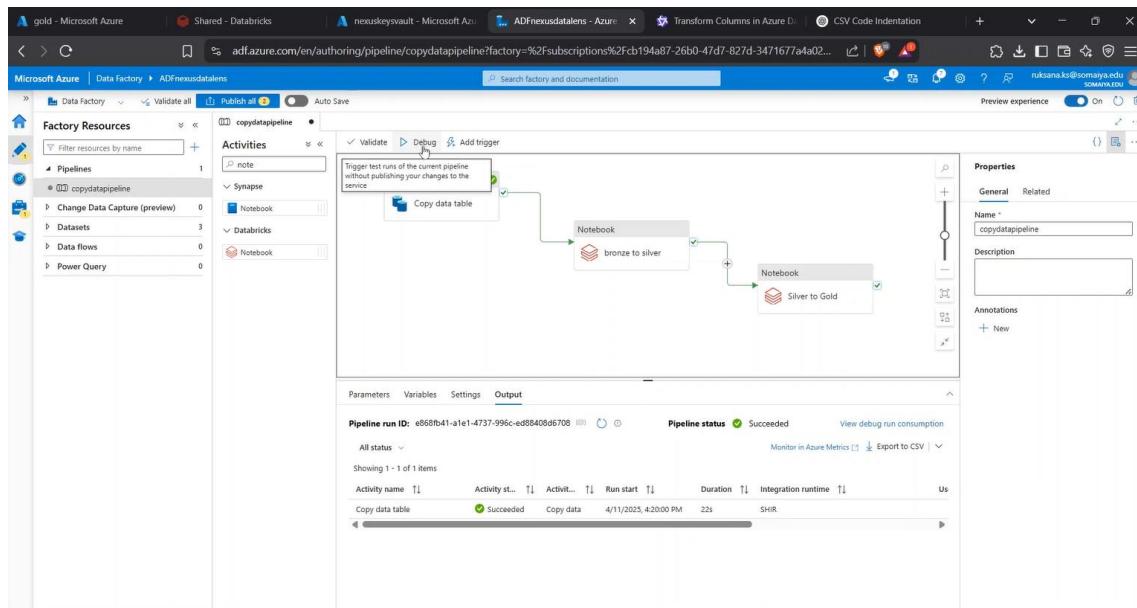


Nexus DataLens

The selection of the **bronze to silver** Databricks notebook from the Shared folder in Azure Data Factory for use in the pipeline.

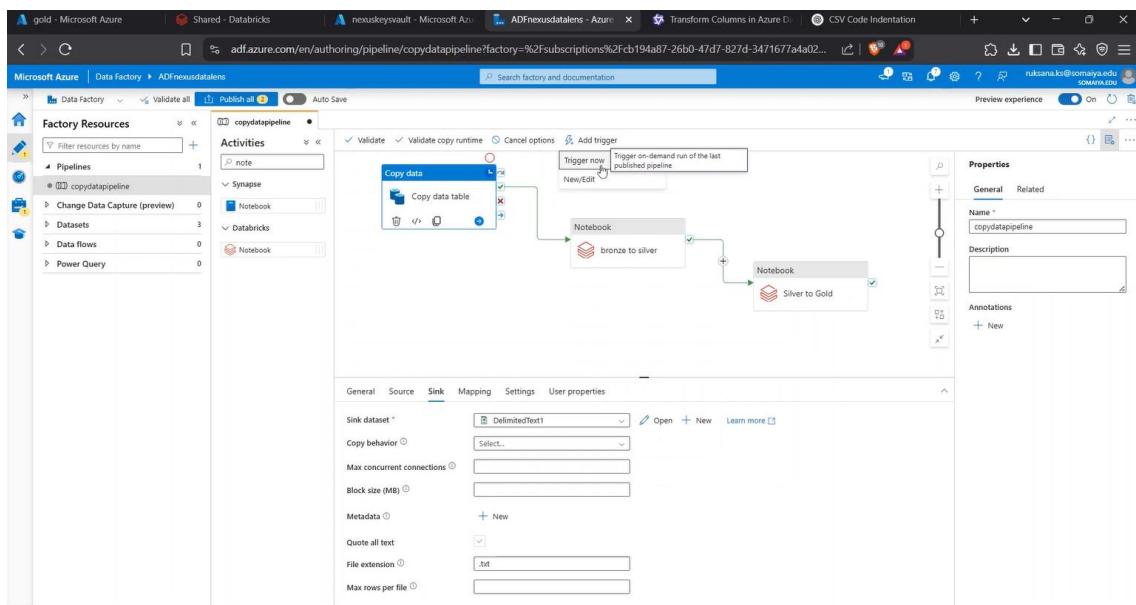


The Azure Data Factory pipeline **copydatapipeline** has successfully run all activities—copying data, and executing the **bronze to silver** and **silver to gold** Databricks notebooks.



Nexus DataLens

This Azure Data Factory pipeline copies data to a table, then processes it through two Databricks notebooks to transform it from bronze to silver, and finally to gold.



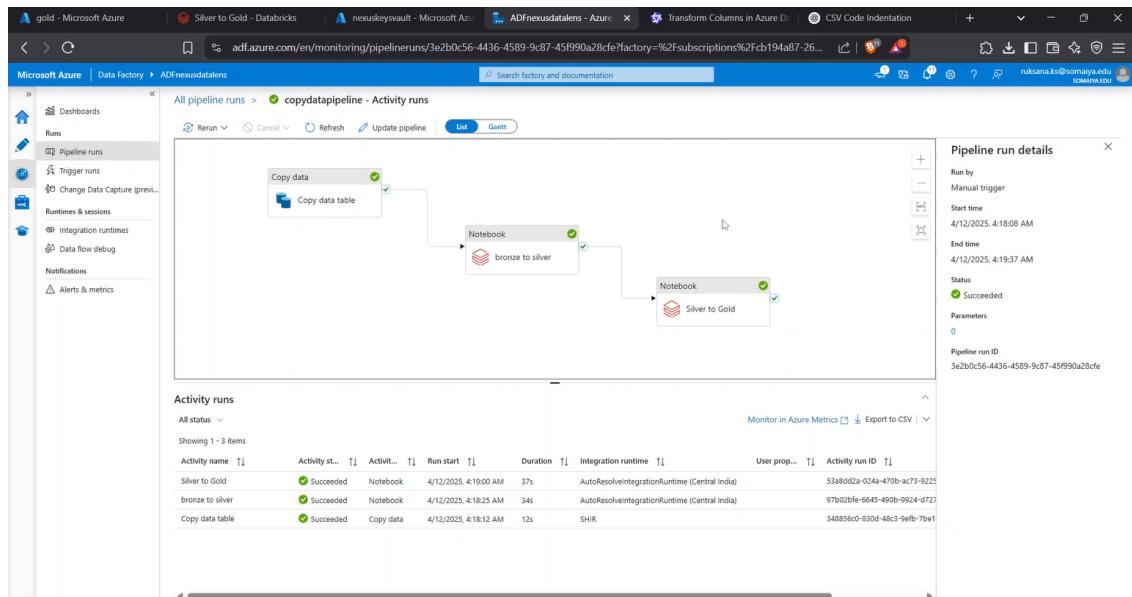
The pipeline successfully executed all three activities—data copy, bronze to silver, and silver to gold transformations—within a span of one minute.

The screenshot shows the 'Activity runs' section of the Azure Data Factory pipeline monitoring interface. It lists three successful runs:

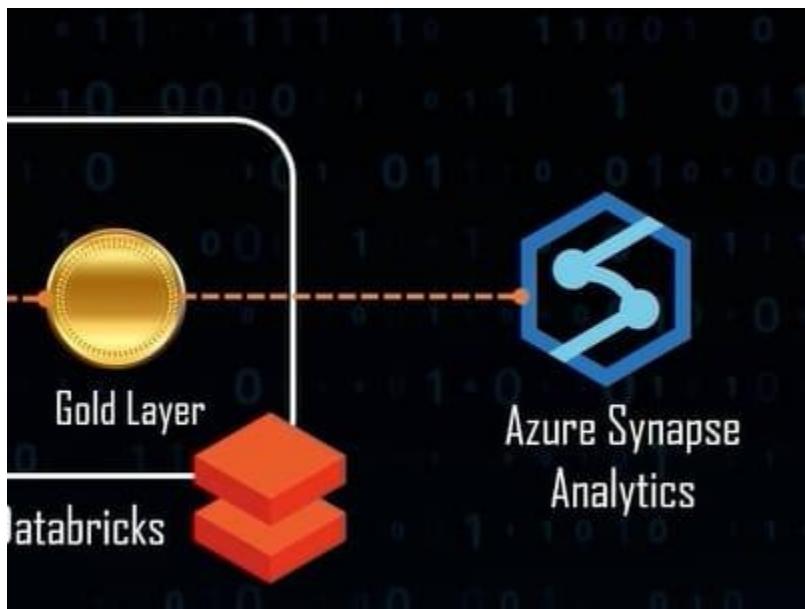
Activity name	Activity st...	Activit...	Run start	Duration	Integration runtime
Silver to Gold	✓ Succeeded	Notebook	4/12/2025, 4:19:00 AM	37s	AutoResolveIntegrationRuntime (Central India)
bronze to silver	✓ Succeeded	Notebook	4/12/2025, 4:18:25 AM	34s	AutoResolveIntegrationRuntime (Central India)
Copy data table	✓ Succeeded	Copy data	4/12/2025, 4:18:12 AM	12s	SHIR

Nexus DataLens

The manually triggered ADF pipeline completed successfully, executing the copy, bronze-to-silver, and silver-to-gold notebook transformations in sequence.



SYNAPSE



The Azure portal homepage displays recent resources, highlighting the selected Data Factory instance “**NexusProject**” for further access or management.

The screenshot shows the Microsoft Azure portal homepage with the following details:

- Azure services:** A row of icons for creating a resource, Subscriptions, Resource groups, Azure Synapse Analytics, Budgets, Bot Services, Projects, Quickstart Center, Azure AI services, and More services.
- Resources:** A table of recent resources:

Name	Type	Last Viewed
nexusalensstorage	Storage account	16 minutes ago
NexusProject	Resource group	3 hours ago
nexudatabricks	Azure Databricks Service	4 hours ago
ADFnexusdatalens	Data factory (V2)	4 hours ago
neuskeyvault	Key vault	13 hours ago
Azure for Students	Subscription	4 days ago
- Tools:** A row of links for Microsoft Learn, Azure Monitor, Microsoft Defender for Cloud, and Cost Management.

Nexus DataLens

The NexusProject resource group in Azure contains five resources, including Data Factory, Databricks, Storage Account, Key Vault, and Synapse Workspace, all located in Central India.

Azure portal screenshot showing the NexusProject resource group details. The 'Resources' tab is selected, displaying five resources:

Name	Type	Location	Actions
ADFNexusdatalens	Data factory (V2)	Central India	...
nexusdatabricks	Azure Databricks Service	Central India	...
nexusdatalenstorage	Storage account	Central India	...
nexuskeyvault	Key vault	Central India	...
nexussynapse	Synapse workspace	Central India	...

The Synapse workspace "**nexussynapse**" is ready for use with serverless SQL enabled, offering quick access to Synapse Studio for data analytics and development.

Azure portal screenshot showing the nexussynapse Synapse workspace details. The 'Getting started' section features two cards:

- Open Synapse Studio: Start building your fully-integrated analytics solution and unlock new insights.
- Read documentation: Learn how to be productive quickly. Explore concepts, tutorials, and samples.

The 'Analytics pools' section shows a table with one entry:

Name	Type	Size
Built-in	Serverless	Auto

Nexus DataLens

The Synapse Analytics workspace "nexusynapse" dashboard offers options to ingest, explore, and visualize data, with navigation to data, development, and integration features.

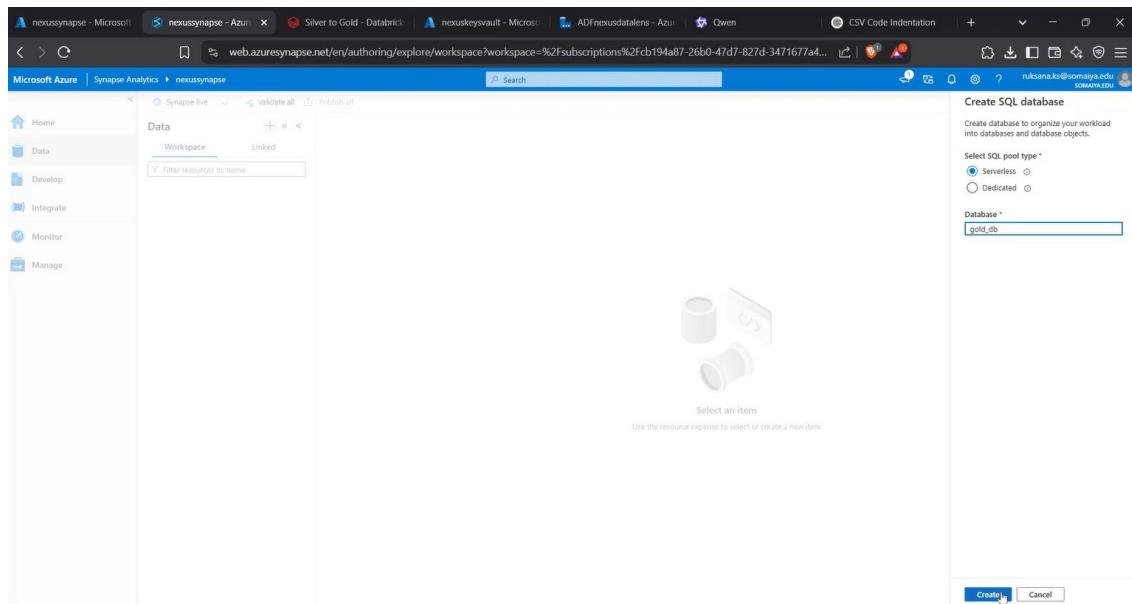
This screenshot shows the Azure Synapse Analytics workspace "nexusynapse". The left sidebar includes links for Home, Data (which is currently selected), Develop, Integrate, Monitor, and Manage. The main area displays a "Synapse Analytics workspace" header and a large, colorful 3D bar chart visualization. Below the chart are three buttons: "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.). A "Discover more" section contains links for "Knowledge center" and "Browse partners". A "Recent resources" section indicates "No recent resources" available. At the bottom, there is a "Feature showcase" section.

The image shows the "Data" pane in Azure Synapse Studio, where the user is selecting to add a new SQL or Lake database under the Workspaces tab.

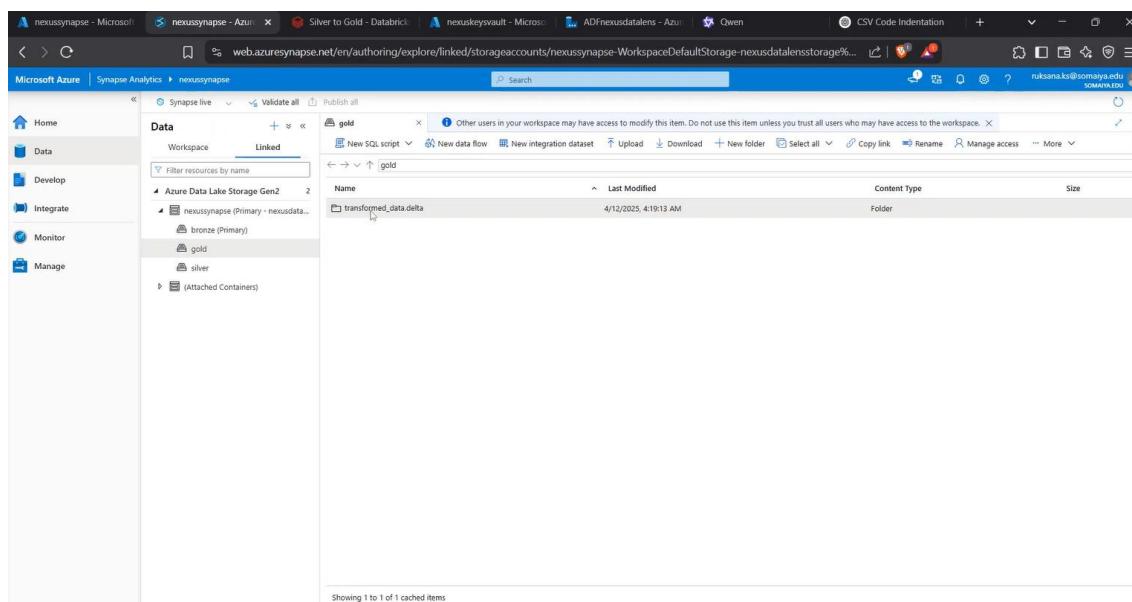
This screenshot shows the "Data" pane in Azure Synapse Studio, specifically the "Workspaces" tab. The left sidebar lists Home, Data (selected), Develop, Integrate, Monitor, and Manage. The main pane shows a resource explorer with "Workspace" selected. Under "Workspace", there are options for "SQL database" and "Lake database", with "Data Explorer database (preview)" listed below them. Other sections include "Linked", "Connect to external data", "Integration dataset", and "Browse gallery". A central area displays two cylindrical icons representing databases, with the text "Select an item" and the instruction "Use the resource explorer to select or create a new item".

Nexus DataLens

The user creating a new serverless SQL database named **gold_db** in Azure Synapse Studio by clicking the **Create** button.

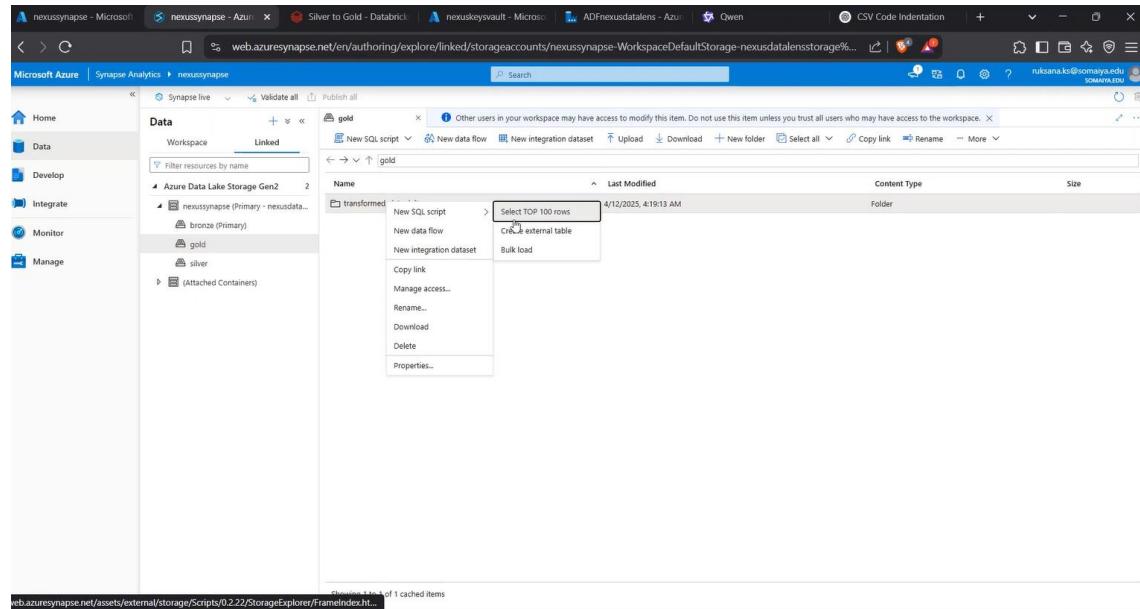


The image shows the gold container in Azure Data Lake Storage Gen2 within Synapse Studio, containing a folder named **transformed_data.delta**.

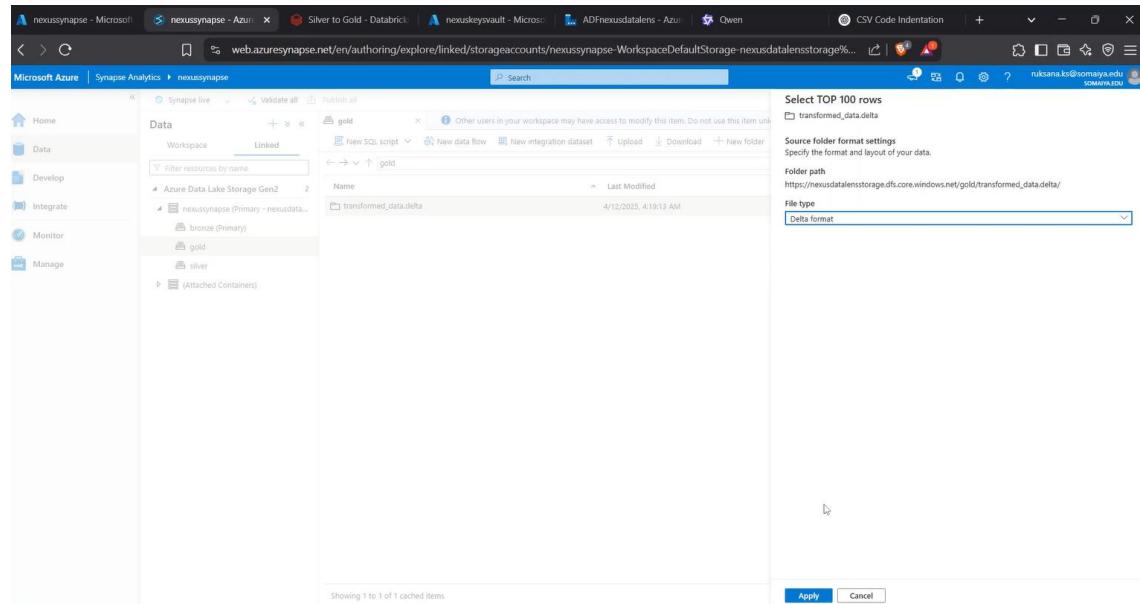


Nexus DataLens

The user right-clicking on the transformed_data.delta folder in Azure Synapse Studio and selecting the "Select TOP 100 rows" SQL script option.



The image shows the user configuring the file type as Delta format for the folder **transformed_data.delta** before querying the top 100 rows in Azure Synapse Studio.



Nexus DataLens

The image shows a SQL script in Azure Synapse Studio that creates a view to select the top 100 rows from a Delta format file located in the **gold/transformed_data.delta** folder.

A screenshot of the Microsoft Azure Synapse Analytics studio interface. The left sidebar shows 'Data' selected under 'Workspace'. In the main area, a 'gold' folder is selected. A SQL script editor window is open with the following code:

```
1 CREATE VIEW [gold].[top100]
2   SELECT TOP 100 *
3   FROM OPENROWSET(
4     BULK 'https://nexusdatalensstorage.dfs.core.windows.net/gold/transformed_data.delta',
5     FORMAT = 'DELTA'
6   ) AS [result]
```

The 'Properties' panel on the right shows the script is named 'SQL script 1'. The results settings per query are set to 'First 5000 rows (default)'. The browser tab at the top is 'web.azure.synapse.net/en/authoring/explore/linked/sqlscripts/SQL%20script%201?workspace=%2Fsubscriptions%2Fc194a87-26b0-4...'. The URL in the address bar is 'https://web.azure.synapse.net/en/authoring/explore/linked/sqlscripts/SQL%20script%201?workspace=%2Fsubscriptions%2Fc194a87-26b0-4...'. The user's email 'rukama.ks@somaya.edu' is visible in the top right corner.

The user again right-clicking on the **transformed_data.delta** folder in Azure Synapse Studio and selecting "Select TOP 100 rows" to preview data from the Delta table.

A screenshot of the Microsoft Azure Synapse Analytics studio interface. The left sidebar shows 'Data' selected under 'Workspace'. In the main area, a 'gold' folder is selected. A context menu is open over a 'transformed_data.delta' file, with the 'Select TOP 100 rows' option highlighted. Other options in the menu include 'New SQL script', 'New data flow', 'New integration dataset', 'Copy link', 'Manage access...', 'Rename...', 'Download', 'Delete', and 'Properties...'. The browser tab at the top is 'web.azure.synapse.net/en/authoring/explore/linked/storageaccounts/nexusynapse-WorkspaceDefaultStorage-nexusdatalensstorage%...'. The URL in the address bar is 'https://web.azure.synapse.net/en/authoring/explore/linked/storageaccounts/nexusynapse-WorkspaceDefaultStorage-nexusdatalensstorage%...'. The user's email 'rukama.ks@somaya.edu' is visible in the top right corner.

Nexus DataLens

The image shows the process of selecting the top 100 rows from a Delta format file (**transformed_data.delta**) in Azure Synapse Analytics.

This screenshot shows the Microsoft Azure Synapse Analytics DataLens interface. On the left, there's a navigation sidebar with options like Home, Data, Develop, Integrate, Monitor, and Manage. The main area is titled 'Data' and shows a 'Workspace' view. A folder structure is displayed under 'Azure Data Lake Storage Gen2': 'nexusynapse (Primary - nexusdata...)' contains 'bronze (Primary)', 'gold', and 'silver'. The 'gold' folder is selected. Inside 'gold', there's a file named 'transformed_data.delta'. To the right of the file list, a modal dialog is open titled 'Select TOP 100 rows'. It shows the path 'https://nexusdatalensstorage.dfs.core.windows.net/gold/transformed_data.delta' and specifies 'File type: Delta format'. At the bottom of the dialog are 'Apply' and 'Cancel' buttons.

The image shows a SQL script in Azure Synapse that creates a view (**transformed_data**) by querying Delta format data from a specified path in the Data Lake using **OPENROWSET**.

This screenshot shows the Microsoft Azure Synapse Analytics DataLens interface. The left sidebar shows the same workspace structure as the previous screenshot. In the center, there's a 'SQL script 1' tab with the following SQL code:

```
1 CREATE VIEW transformed_data
2 AS
3 SELECT *
4 FROM
5 OPENROWSET(
6     BULK 'https://nexusdatalensstorage.dfs.core.windows.net/gold/transformed_data.delta',
7     FORMAT = 'DELTA'
8 ) AS [result]
```

The 'Results' tab at the bottom shows a message: 'No results to show' and 'Your query yielded no displayable results'. The 'Properties' pane on the right shows details for 'SQL script 1': Name is 'SQL script 1', Type is 'sql script', Size is 212 bytes, and Results settings per query is set to 'First 5000 rows (default)'. The status bar at the bottom indicates '00:00:04 Query executed successfully.'

Nexus DataLens

The image shows the "**Linked services**" section in Azure Synapse Analytics where users can add new connections to external resources by clicking the "+ New" button.

A screenshot of the Microsoft Azure portal showing the 'Linked services' section in the 'Synapse Analytics' blade. The page displays two existing linked services: 'nexusynapse-WorkspaceDefaultSqlServer' (Type: Azure Synapse Analytics) and 'nexusynapse-WorkspaceDefaultStorage' (Type: Azure Data Lake Storage Gen2). At the top left, there is a '+ New' button. The left sidebar contains navigation links for Home, Data, Develop, Integrate, Monitor, and Manage, with 'Manage' currently selected.

The image shows the selection of "**Azure SQL Database**" as a new linked service in Azure Synapse Analytics, with the user about to click "**Continue**" to proceed with the setup.

A screenshot of the 'New linked service' dialog box overlaid on the 'Linked services' page. The 'Database' tab is selected, showing two options: 'Azure SQL Database' and 'Azure SQL Database Managed Instance'. At the bottom right of the dialog, there is a 'Continue' button.

Nexus DataLens

The configuration of a new linked service in Azure Synapse for connecting to an Azure SQL Database using SQL authentication.

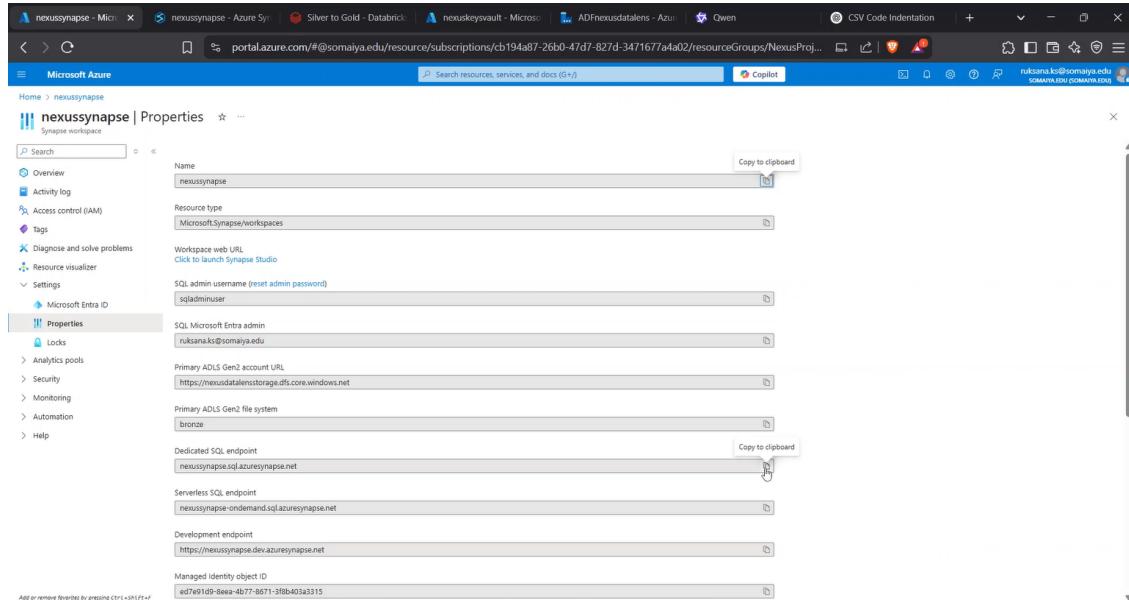
The screenshot shows the Azure portal interface for configuring a linked service. The left sidebar navigation includes Home, Data, Develop, Integrate, Monitor, and Manage sections. Under the 'Integrate' section, 'Linked services' is selected. On the right, a 'New linked service' dialog is open for an 'Azure SQL Database'. The 'Name' field is set to 'ServerlessDB'. The 'Type' dropdown is set to 'Azure Synapse Analytics'. The 'Version' is set to '2.0'. The 'Account selection method' is set to 'From Azure subscription'. The 'Azure subscription' dropdown shows 'Azure for Students (cb194a87-26b0-47d7-827d-3471677a4a02)'. The 'Server name' dropdown shows 'nexusynapse'. The 'Database name' dropdown shows 'Loading...'. The 'Authentication type' is set to 'SQL authentication'. The 'User name' field is empty. The 'Password' field is set to 'Azure Key Vault'. At the bottom, there are 'Create', 'Back', and 'Cancel' buttons.

The Azure portal home displays recently accessed resources, highlighting a Synapse workspace named "nexusynapse" as the most recently viewed.

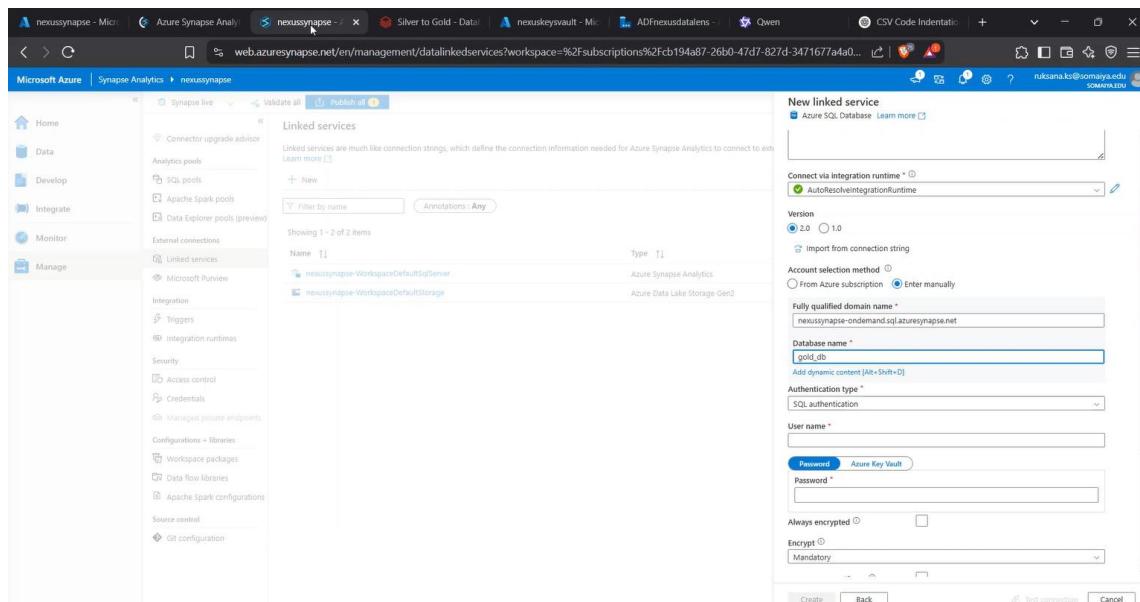
The screenshot shows the Azure portal home page. The top navigation bar includes links for 'bronze - Microsoft A', 'nexusynapse - Azure Sy...', 'Silver to Gold - Databrick...', 'nexuskeyvault - Micro...', 'ADFnxusdatalens - Azu...', 'Qwen', 'CSV Code Indentation', and user 'rukansa.ks@somaiya.edu SOMAIYA.EDU'. Below the navigation is a search bar and a 'Copilot' button. The main area features 'Azure services' with icons for 'Create a resource', 'Subscriptions', 'Resource groups', 'Azure Synapse Analytics...', 'Budgets', 'Bot Services', 'Projects', 'Quickstart Center', 'Azure AI services', and 'More services'. The 'Resources' section shows a list of recently accessed resources, with 'nexusynapse' highlighted. The 'Navigate' section includes links for 'Subscriptions', 'Resource groups', 'All resources', and 'Dashboard'. The 'Tools' section includes links for 'Microsoft Learn', 'Azure Monitor', 'Microsoft Defender for Cloud', and 'Cost Management'.

Nexus DataLens

The properties page of the Synapse workspace "nexusynapse" displays key configuration details like workspace URL, SQL admin, endpoints, and managed identity.

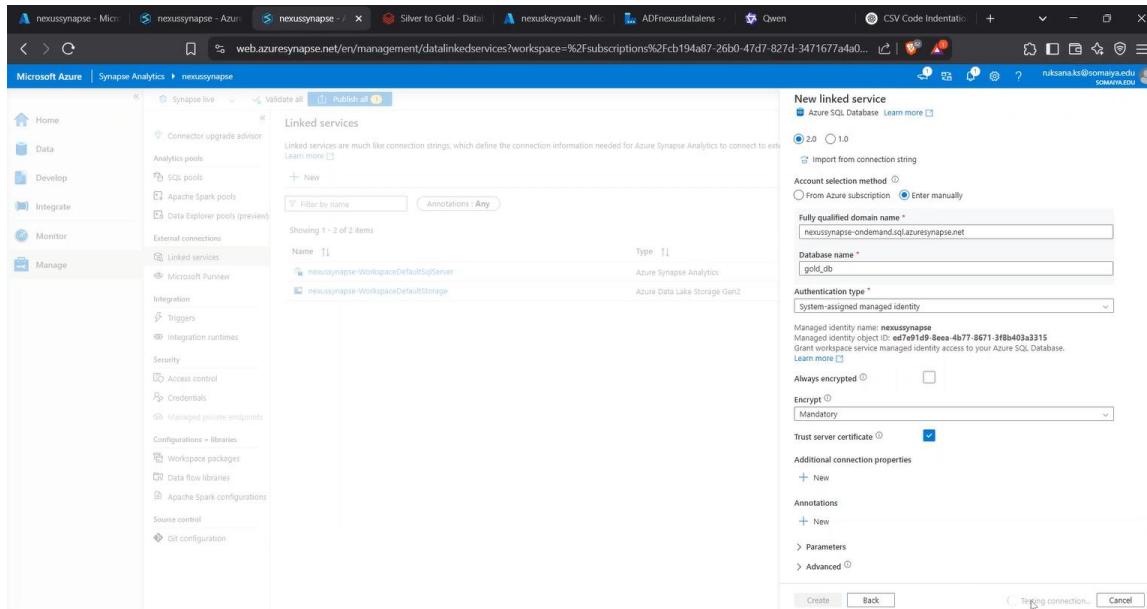


Creating a new linked service using SQL authentication by specifying the domain, database name (**gold_db**), and credentials for Azure SQL Database connection.



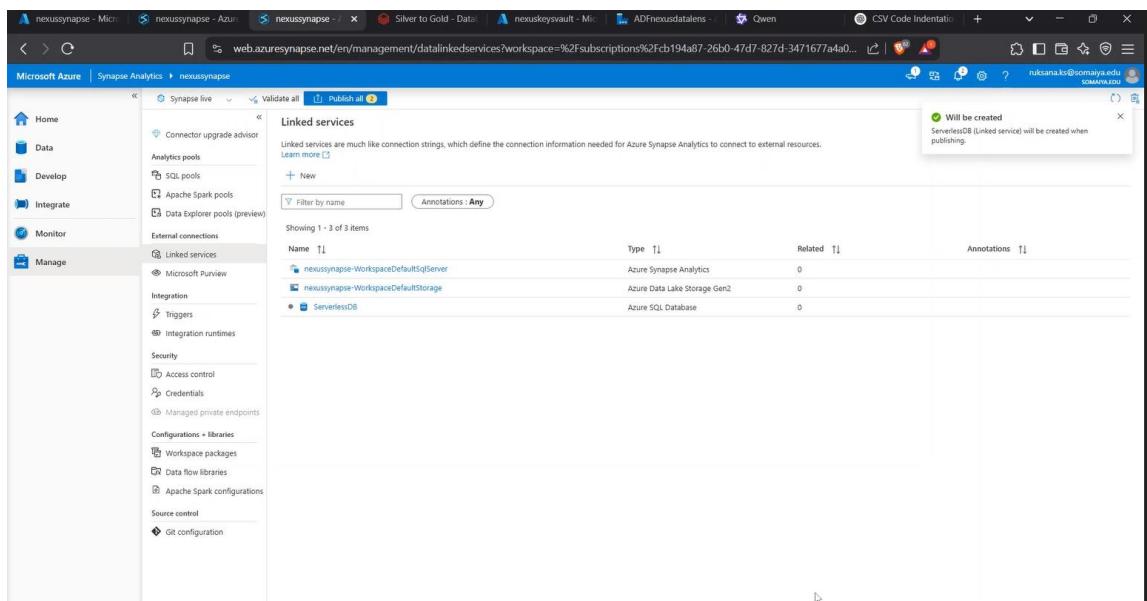
Nexus DataLens

A new linked service for an Azure SQL Database named **gold_db** is being configured using a managed identity for authentication.



The screenshot shows the 'Linked services' configuration page in the Microsoft Azure portal. A new linked service is being created for an Azure SQL Database named 'gold_db'. The 'Authentication type' is set to 'System-assigned managed identity'. The 'Create' button is visible at the bottom.

The "Linked services" view shows that the new linked service ServerlessDB will be created upon publishing.



The screenshot shows the 'Linked services' view in the Microsoft Azure portal after publishing. A tooltip indicates that 'ServerlessDB (Linked service) will be created when publishing.' The table lists three linked services:

Name	Type	Related	Annotations
nexussynapse-WorkspaceDefaultSqlServer	Azure Synapse Analytics	0	
nexussynapse-WorkspaceDefaultStorage	Azure Data Lake Storage Gen2	0	
ServerlessDB	Azure SQL Database	0	

Nexus DataLens

The user is about to publish changes including a new linked service **ServerlessDB** and a new SQL script **SQL script 2** to the Synapse workspace.

The screenshot shows the Microsoft Azure Synapse Analytics portal. On the left, the navigation menu includes Home, Data, Develop, Integrate, Monitor, and Manage. Under the Integrate section, there are options for Connector upgrade advisor, Linked services, External connections, Integration, Security, Configurations + libraries, and Source control. In the center, under 'Linked services', a table lists three items: 'nexusynapse-WorkspaceDefaultSqlServer' (Azure Synapse Analytics), 'nexusynapse-WorkspaceDefaultStorage' (Azure Data Lake Storage Gen2), and 'ServerlessDB' (Azure SQL Database). A 'Pending changes (2)' section on the right shows two items: 'ServerlessDB' (New) under 'Linked services' and 'SQL script 2' (New) under 'SQL script'. Below this is a 'Publish all' button.

The SQL script failed with an error indicating that the table `transformed_data` already exists in the `gold_db` database.

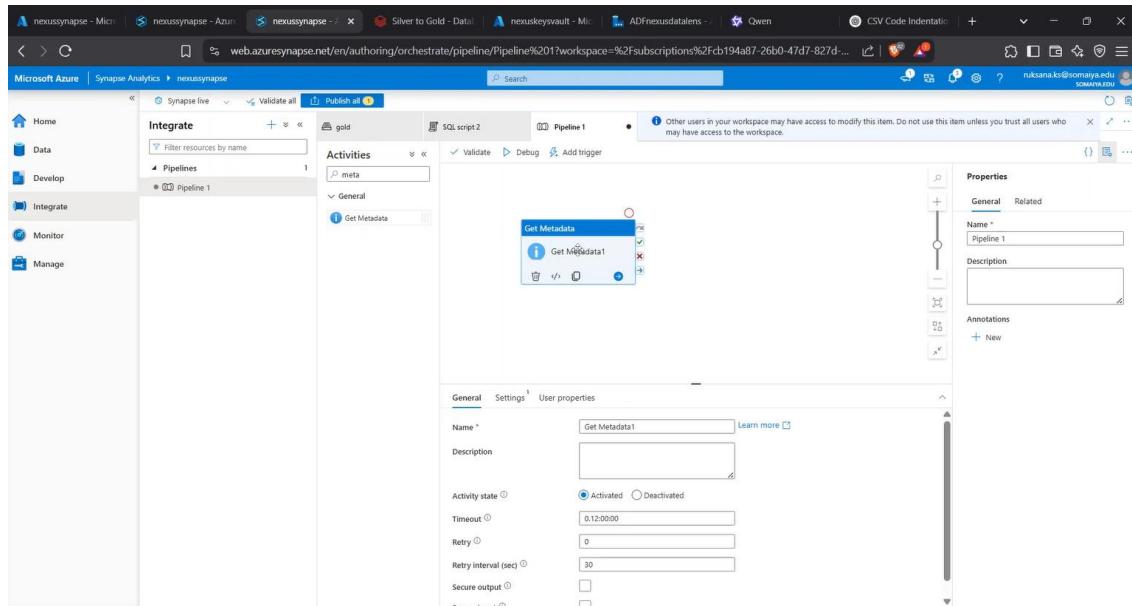
The screenshot shows the Microsoft Azure Synapse Analytics portal with the 'Integrate' section selected. A pipeline named 'gold' is currently running, indicated by a progress bar. The pipeline editor shows a single step: 'Import from pipeline template' with the following SQL script:

```
1 INSERT
2   INTO "https://nexusdatalensstorage.dfs.core.windows.net/gold/transformed_data.delta"
3     FORMAT = 'DELTA'
4   ) AS [result]
```

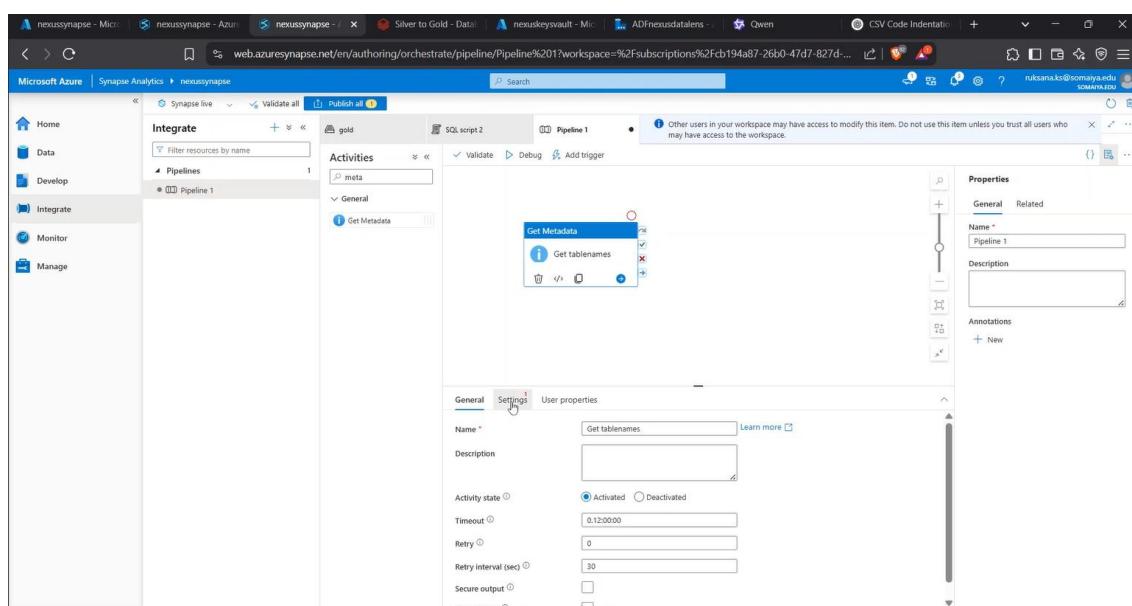
In the 'Messages' pane, an error message is displayed: 'There is already an object named 'transformed_data' in the database.' At the bottom, a note says '00:00:07 Query completed with errors.'

Nexus DataLens

The "Get Metadata1" activity is activated in the pipeline with default execution settings like timeout and retry configured.

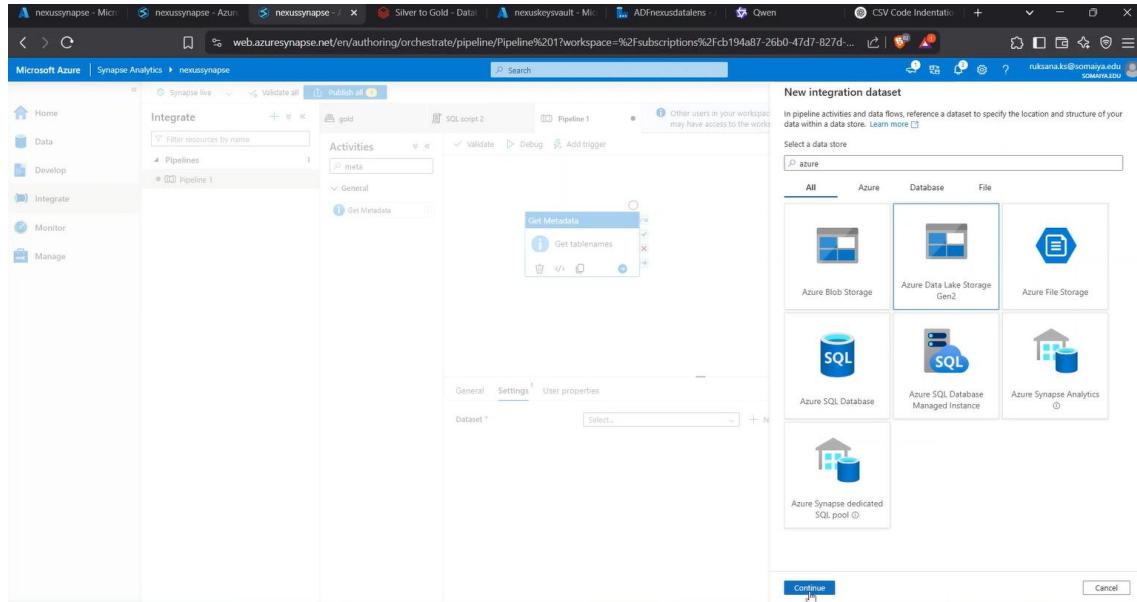


The "Get Metadata" activity is named "**Get tablenames**" and is configured with default timeout, retry, and output settings.

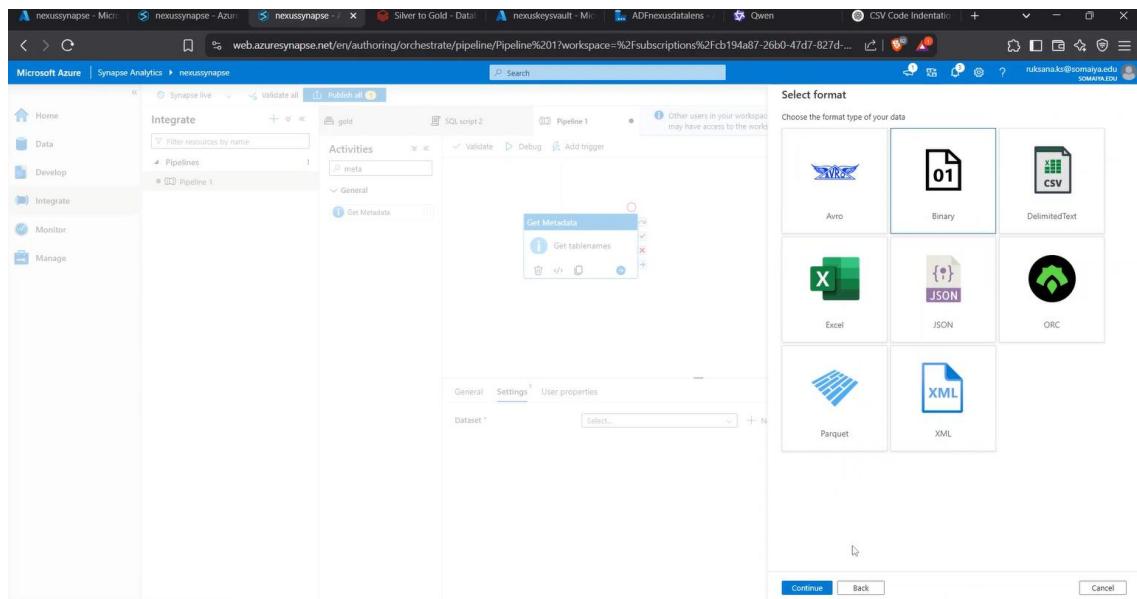


Nexus DataLens

The user is creating a new integration dataset by selecting "**Azure Data Lake Storage Gen2**" as the **data store type**.

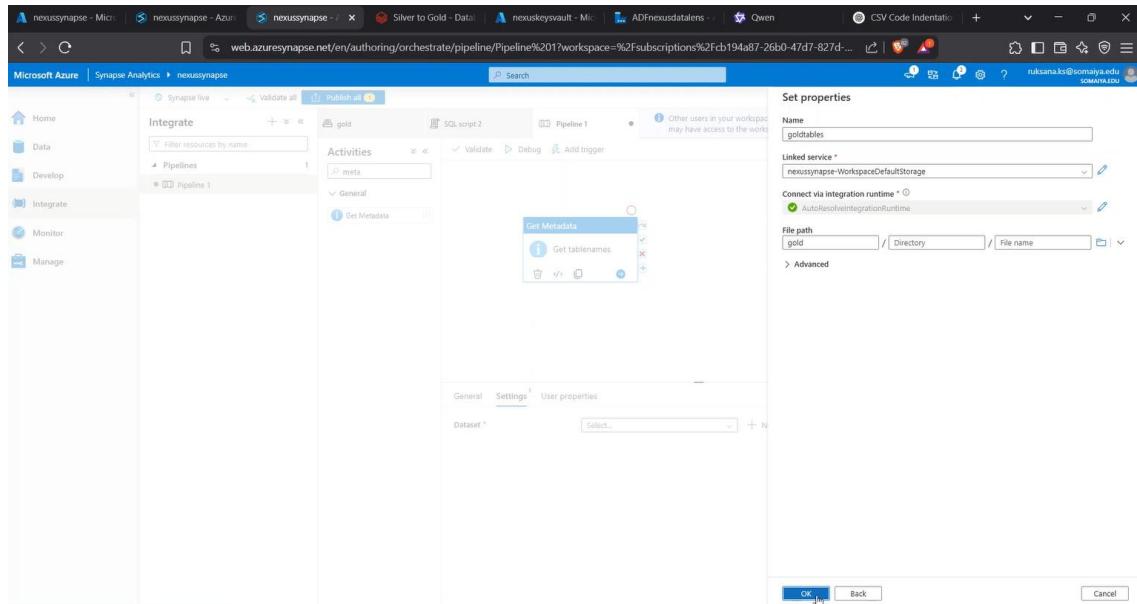


The data format selection screen shows various options, with "**Binary**" currently selected for the dataset configuration.

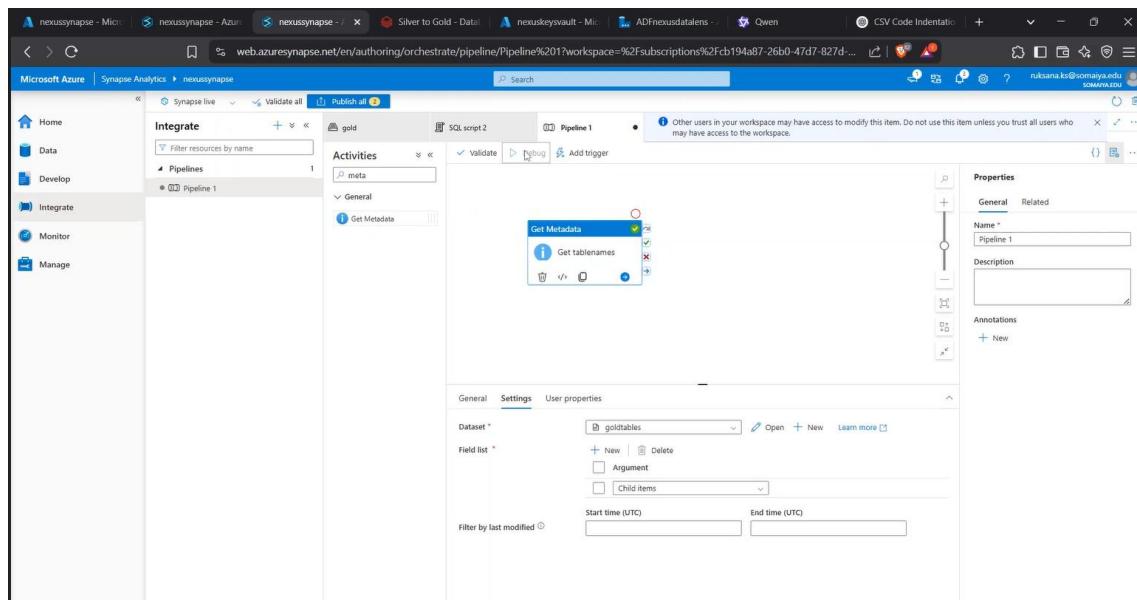


Nexus DataLens

The "goldtables" dataset is configured to access the "gold" directory in Azure Data Lake Storage via the specified linked service.

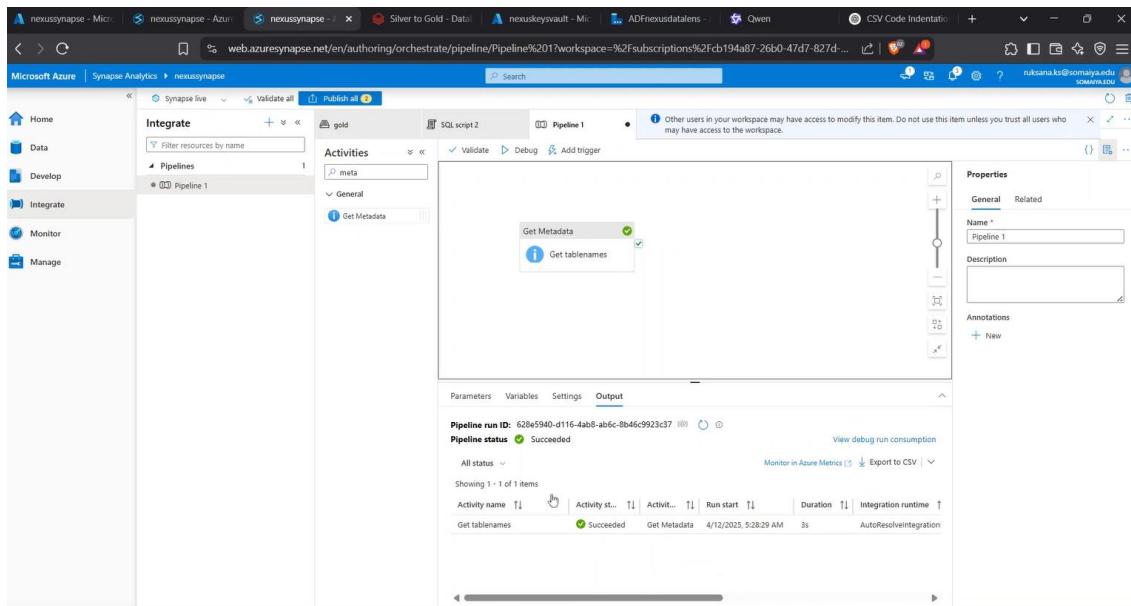


The pipeline's "Get Metadata" activity is configured to retrieve child items (table names) from the "goldtables" dataset.



Nexus DataLens

The Synapse pipeline "**Pipeline 1**" ran successfully, retrieving table names using a "**Get Metadata**" activity.



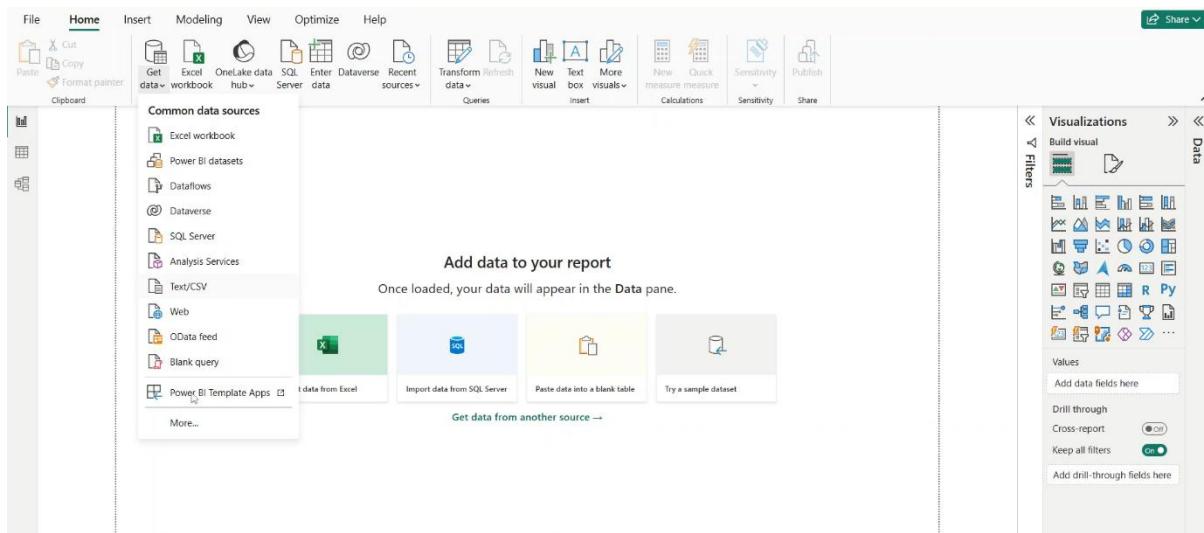
The screenshot shows the Microsoft Azure Synapse Analytics pipeline editor. On the left, the navigation menu includes Home, Data, Develop, Integrate, Monitor, and Manage. Under Integrate, Pipelines is selected, showing Pipeline 1. The main workspace displays the Pipeline 1 interface. The Activities pane on the left lists "Get Metadata" under the "Get" category. The central area shows a single activity named "Get tablenames". The Properties pane on the right shows the General tab with "Name" set to "Pipeline 1". The Output tab shows a table with one row:

Activity name	Activity status	Run start	Duration	Integration runtime	
Get tablenames	Succeeded	Get Metadata	4/12/2025, 5:28:29 AM	3s	AutoResolveIntegration

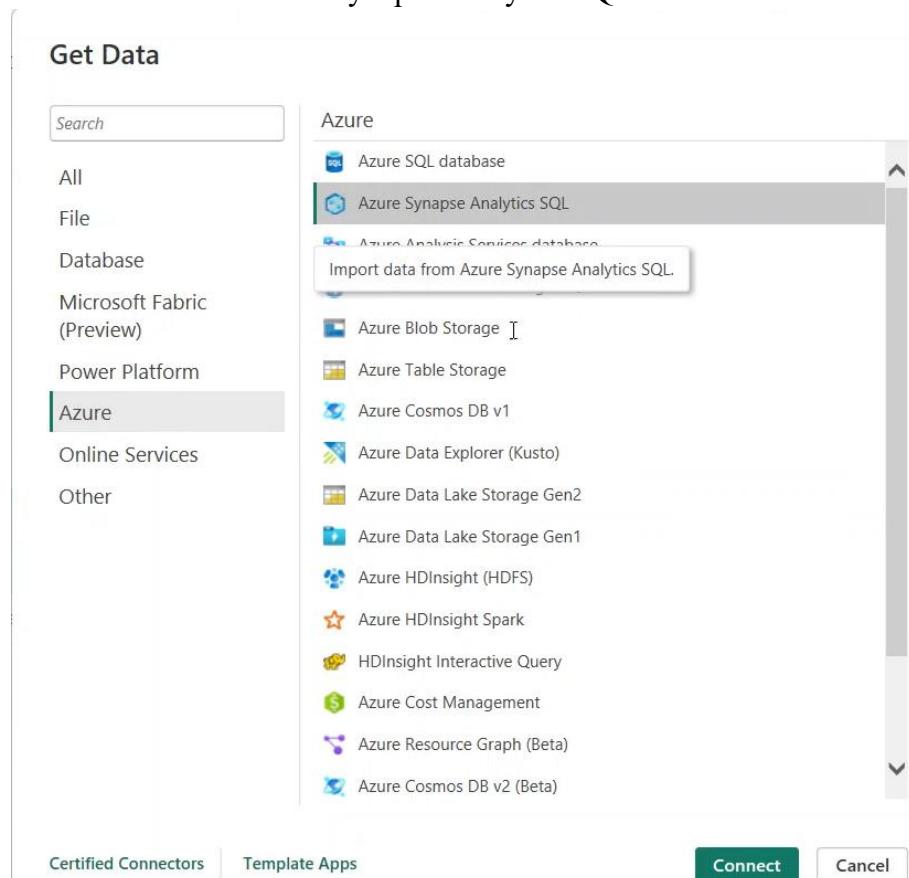
Power Bi Setup



Install Power Bi and open it
in Navbar You will see click on Get Data > More



Click On Azure > Azure Synapse Analytics SQL > Connect



This Window will open You need to paste serverless endpoint of Azure synapse analytics for that

SQL Server database

Server ⓘ

Database (optional)

Data Connectivity mode ⓘ

Import

DirectQuery

▷ Advanced options

OK

Cancel

Nexus DataLens

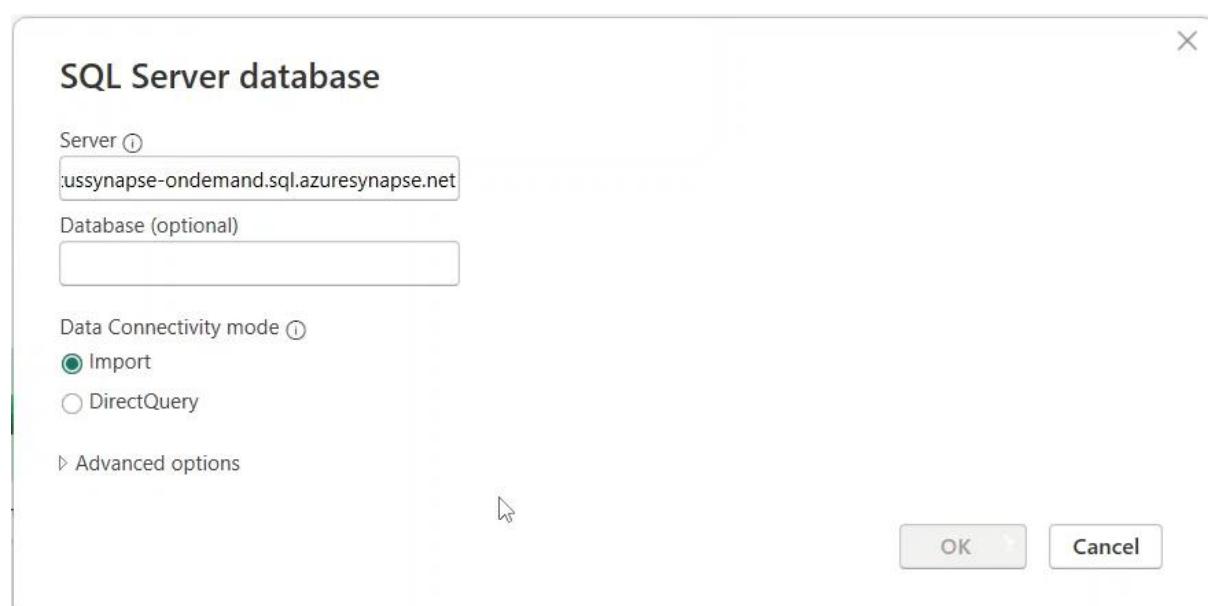
Go to resource group and click on > nexussynapse

The screenshot shows the Microsoft Azure portal interface. At the top, there's a blue header bar with the 'Microsoft Azure' logo and a search bar that says 'Search resources, services, and...'. Below the header, the title 'NexusProject' is displayed, followed by 'Resource group'. On the left, a sidebar menu includes 'Overview' (which is selected), 'Activity log', 'Access control (IAM)', 'Tags', 'Resource visualizer', 'Events', 'Settings', 'Cost Management', 'Monitoring', 'Automation', and 'Help'. The main content area has a heading 'Essentials' with information about the subscription (move to 'Azure for Students', ID: cb194a87-26b0-47d7-827d-3471677a4a02, tags 'Add tags'). Below this is a 'Resources' section with tabs for 'Resources' and 'Recommendations'. It features a search bar with filters for 'Type equals all' and 'Location equals all'. A message says 'Showing 1 to 5 of 5 records.' There are checkboxes for 'Name ↑' and 'Type'. A list of resources is shown, including 'ADFnexusdatalens', 'nexusdatabricks', 'nexusdatalensstorage', 'nexuskeyvault', and 'nexussynapse', with the last one being the target of a mouse cursor.

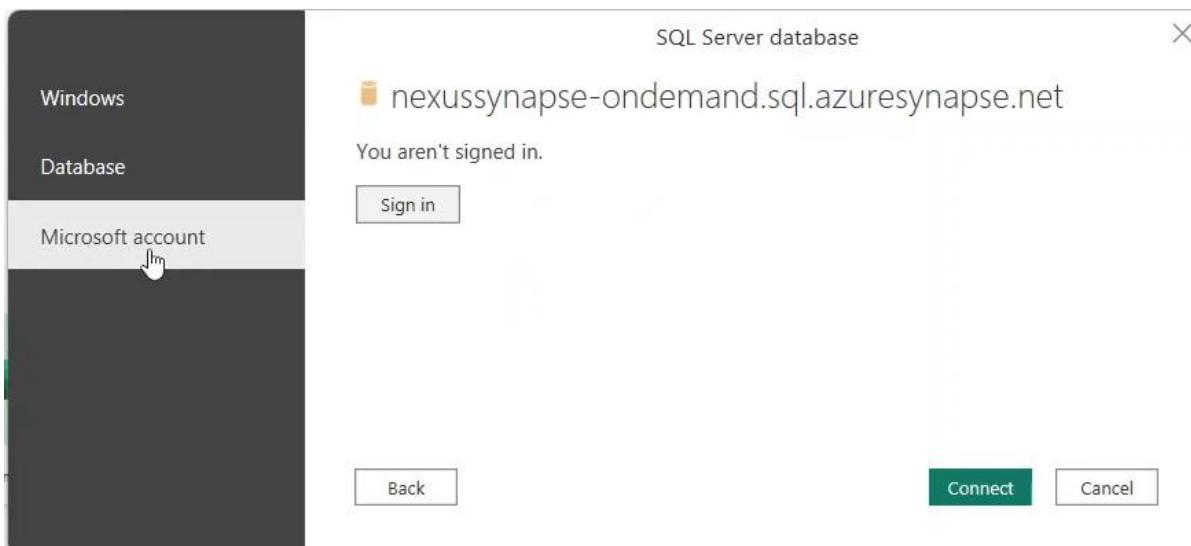
Got to properties you will get link > copy that link



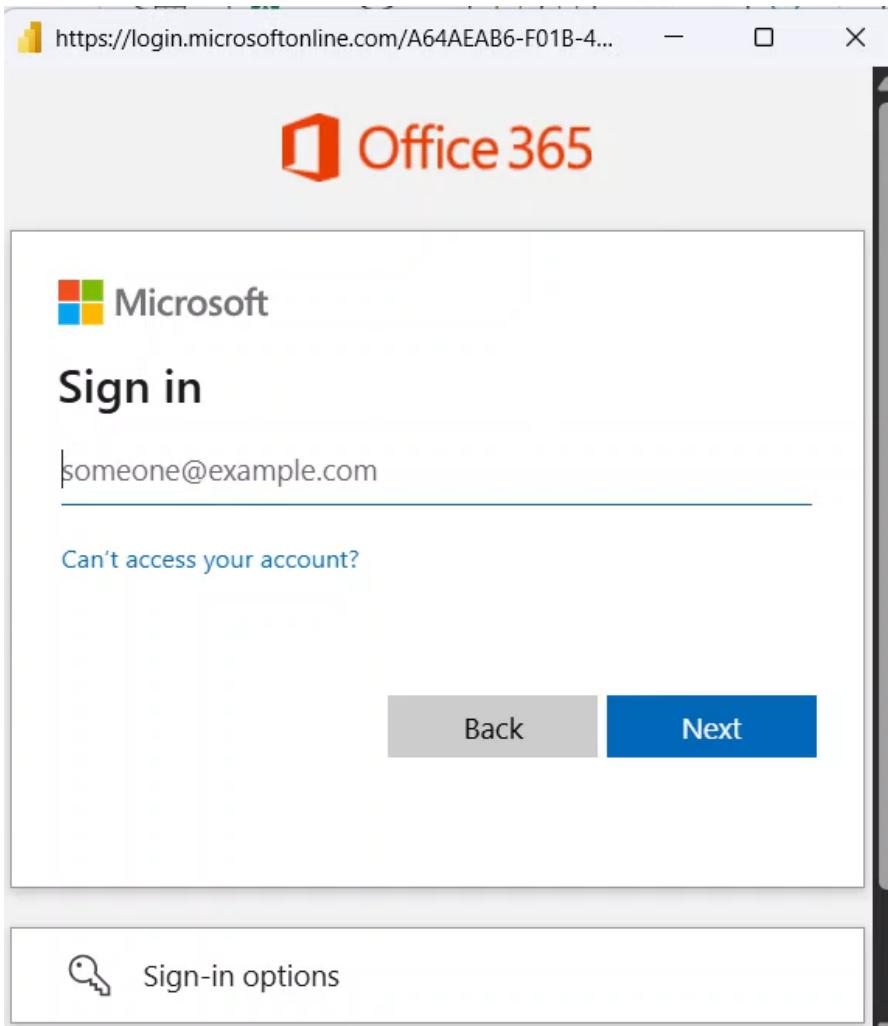
Paste it over here > OK



Click on Microsoft Account > Sign in

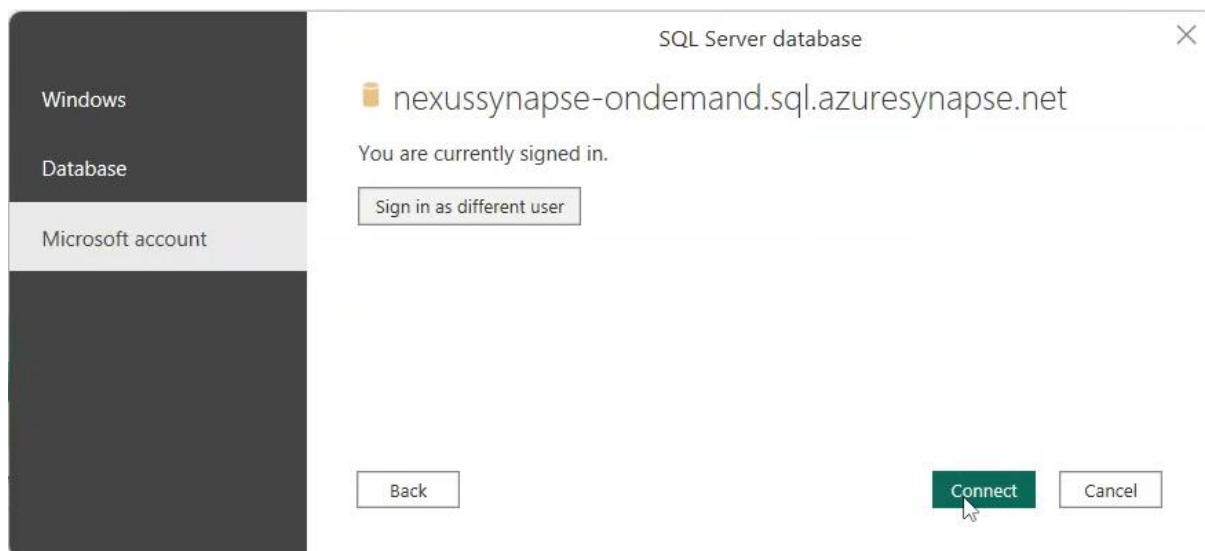


Sign in with same Azure account in which you are making Pipeline

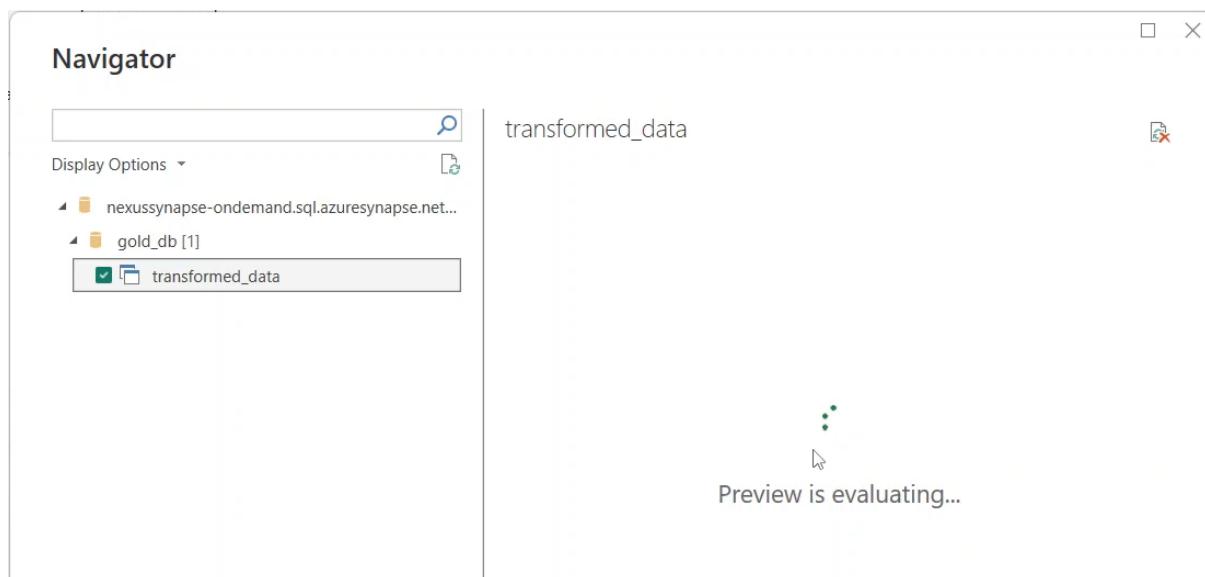


Nexus DataLens

Click on connect



Navigator Tab will get open in which you will able to see Transform Data table Check the checkbox



Wait for some time to load the data

Load

transformed_data
Creating connection in model...

Cancel

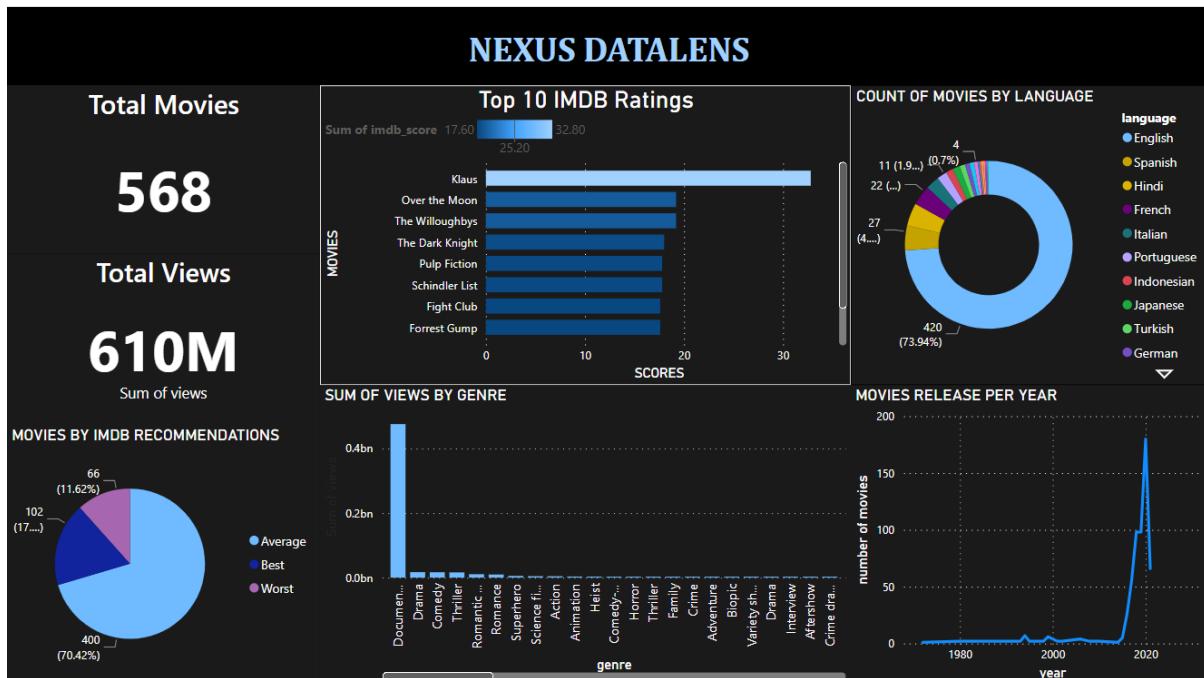
Once it get loaded it will show us Data Coloums in data Tab

The screenshot shows the Nexus DataLens interface. On the left, there's a sidebar with 'Visualizations' and 'Filters'. The main area is titled 'Data' and contains a search bar and a tree view of data fields. The tree view under 'transformed_data' includes fields such as genre, imdb_score, language, releasedate, runtime, title, and views.

You can see the transform data table in table view

title	genre	releasedate	runtime	imdb_score	language	views	addeddate	releaseyear	imdb_recommend
The Highwaymen	Crime drama	06 April 2018	131	6.9	English	54216	16 February 2023	2018	Average
The Lonely Island Presents: The Unauthorized Bash Brothers Experience	Comedy	12 September 2018	30	6.9	English	238789	16 February 2023	2018	Average
The Lonely Island Presents: The Unauthorized Bash Brothers Experience	Musical	12 September 2018	30	6.9	English	238789	16 February 2023	2018	Average
The Meyerowitz Stories (New and Selected)	Comedy-drama	07 September 2018	112	6.9	English	48297	16 February 2023	2018	Average
A Love Song for Latasha	Documentary	21 September 2020	19	6.8	English	1025587	09 February 2023	2020	Average
All In My Family	Documentary	03 May 2019	39	6.8	English	114482	09 February 2023	2019	Average
Always Be My Maybe	Romantic comedy	31 May 2019	102	6.8	English	112288	09 February 2023	2019	Average
Have a Good Trip: Adventures in Psychedelics	Documentary	11 May 2020	85	6.8	English	478240	09 February 2023	2020	Average
Heroin(e)	Documentary	12 September 2017	39	6.8	English	93972	09 February 2023	2017	Average
Mercury 13	Documentary	20 April 2018	79	6.8	English	410607	09 February 2023	2018	Average
The Other Side of the Wind	Drama	02 November 2018	122	6.8	English	74589	09 February 2023	2018	Average
To the Bone	Drama	14 July 2017	107	6.8	English	108306	09 February 2023	2017	Average
Cops and Robbers	Animation	28 December 2020	7	6.9	English	202890	09 February 2023	2020	Average
Cops and Robbers	Short	28 December 2020	7	6.9	English	202890	09 February 2023	2020	Average
I Don't Feel at Home in This World Anymore	Drama	24 February 2017	96	6.9	English	357213	09 February 2023	2017	Average
Our Souls at Night	Romance	29 September 2017	103	6.9	English	8914127	09 February 2023	2017	Average
Outlaw King	Historical-epic	09 November 2018	121	6.9	English	35003	09 February 2023	2018	Average
Dude	Teen comedy-drama	20 April 2018	97	5.1	English	102177	10 February 2023	2018	Average
Father of the Year	Comedy	20 July 2018	94	5.2	English	89849	10 February 2023	2018	Average
Rim of the World	Science fiction adventure	24 May 2019	98	5.2	English	1045588	10 February 2023	2019	Average
The Week Of	Comedy	27 April 2018	116	5.2	English	636243	10 February 2023	2018	Average
Things Heard & Seen	Horror	29 April 2021	121	5.3	English	431185	10 February 2023	2021	Average
Clinical	Thriller	13 January 2017	104	5.1	English	472307	28 January 2023	2017	Average

Create the Visualized Dashboard



PIPELINE TESTING

Perquisites

- Make sure your cluster is connected and attach to the Notebookbook.

```

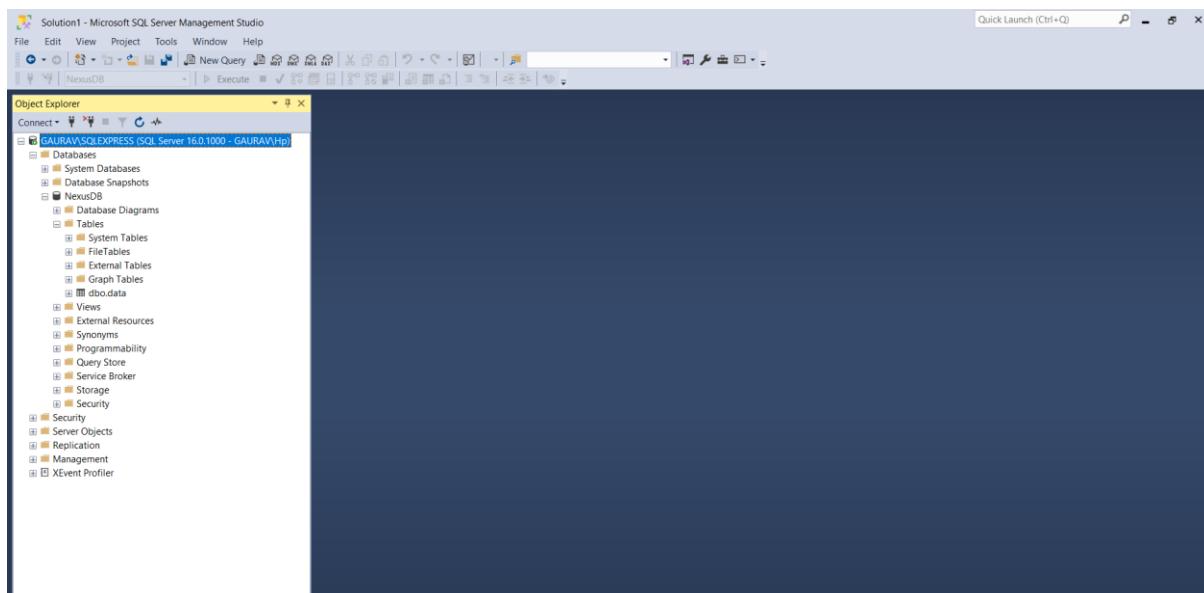
Silver to Gold Python ▾ Tabs: OFF ▾
File Edit View Run Help Last edit was 2 days ago
Workspace Shared Mounting Silver to Gold
1
parquet_file_path = "/mnt/silver/transformed_data.parquet"
# Load the Parquet file into a DataFrame
df_silver = spark.read.format("parquet").load(parquet_file_path)
1 Spark Jobs
df_silver: pyspark.sql.DataFrame = [Title: string, Genre: string ... 7 more fields]

2
from pyspark.sql.functions import col, when
# Add a column to classify movies as "Best" or "Worst" based on IMDB Score
df_silver = df_silver.withColumn(
    "IMDB Recommendation",
    when(col("IMDB_Score") >= 7, "Best")
    .when(col("IMDB_Score") <= 5, "Worst")
    .otherwise("Average")
)
2 days ago (1s)
df_silver: pyspark.sql.DataFrame = [Title: string, Genre: string ... 8 more fields]

3
from pyspark.sql.functions import mean, when
# Calculate average IMDB Score
avg_imdb_score = df_silver.select(mean("IMDB_Score")).collect()[0][0]
# Fill missing IMDB scores with the average
df_silver = df_silver.withColumn("IMDB_Score", when(df_silver["IMDB_Score"].isNull(), avg_imdb_score).otherwise(df_silver["IMDB_Score"]))
3 Spark Jobs
df_silver: pyspark.sql.DataFrame = [Title: string, Genre: string ... 8 more fields]

```

- SSMS Server should be online.



Nexus DataLens

- On-premises Integration runtime is Connected

The screenshot shows the Microsoft Integration Runtime Configuration Manager interface. At the top, there's a navigation bar with Home, Settings, Diagnostics, Update, and Help. A prominent green checkmark icon indicates that a self-hosted node is connected to the cloud service. Below this, it displays the following information:

Data Factory:	ADFnexusdatalens
Integration Runtime:	SHIR
Node:	GAURAV

Below this, there's a button labeled "Stop Service".

Under "Data Source Credential", it shows:

Credential store:	On-premises
Credential status:	In sync
Last backup time:	N/A

With buttons for "Generate Backup" and "Import Backup".

At the bottom, a message says "Connected to the cloud service (Data Factory V2)" with a refresh icon.

To run the Pipeline you need to Auther > copydatapipeline > Trigger > Trigger Now

The screenshot shows the Azure Data Factory pipeline editor. At the top, there are buttons for Validate, Debug, and Add trigger. A dropdown menu is open over the "Add trigger" button, showing "Trigger now" and "New/Edit".

Below this, there's a "Pipeline runs" section with the following details:

Triggered	Debug	Rerun	Cancel options	Refresh	Edit columns	List	Gantt
-----------	-------	-------	----------------	---------	--------------	------	-------

Filter options include: Filter by run ID or name (Chennai, Kolkata, Mu... : Last 24 hours), Pipeline name: All, Status: All, Runs: Latest runs, Triggered by: All, and Add filter.

Showing 1 - 4 items:

Pipeline name ↑↓	Run start ↑↓	Duration	Triggered by	Status ↑↓	Parameters
copydatapipeline	4/14/2025, 3:24:20 AM	27s	Manual trigger	⌚ In progress	
copydatapipeline	4/14/2025, 12:09:02 AM	1m 17s	Manual trigger	✓ Succeeded	
copydatapipeline	4/14/2025, 12:08:05 AM	1m 10s	Manual trigger	✓ Succeeded	
copydatapipeline	4/13/2025, 11:55:33 PM	11m 54s	Manual trigger	✓ Succeeded	

Nexus DataLens

You can see the progress in activity runs
copy data table is Succeeded

Activity runs

All status ▾

Showing 1 - 2 items

Activity name ↑↓	Activity st... ↑↓	Activit... ↑↓	Run start ↑↓	Duration ↑↓	Integration runtime ↑↓
bronze to silver	⌚ In progress	Notebook	4/14/2025, 3:24:46 AM	37s	
Copy data table	✅ Succeeded	Copy data	4/14/2025, 3:24:21 AM	25s	SHIR

Bronze to silver notebook is succeeded

Activity runs

All status ▾

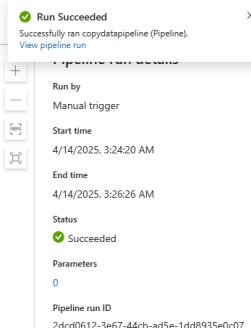
Showing 1 - 3 items

Activity name ↑↓	Activity st... ↑↓	Activit... ↑↓	Run start ↑↓	Duration ↑↓	Integration runtime ↑↓	User prop... ↑↓	Activity run ID ↑↓
Silver to Gold	⌚ In progress	Notebook	4/14/2025, 3:25:37 AM	5s		6138e2bb-0915-466d-9a95-de9	
bronze to silver	✅ Succeeded	Notebook	4/14/2025, 3:24:46 AM	50s	AutoResolveIntegrationRuntime (Central India)	0269c6e0-9658-4775-8ecf-4596	
Copy data table	✅ Succeeded	Copy data	4/14/2025, 3:24:21 AM	25s	SHIR	99badbb7-7181-4bbc-a05c-38d	

Silver to Gold notebook succeeded

All pipeline runs > copydatapipeline - Activity runs

Rerun ▾ Cancel ▾ Refresh Update pipeline List Gantt



Activity runs

All status ▾

Showing 1 - 3 items

Activity name ↑↓	Activity st... ↑↓	Activit... ↑↓	Run start ↑↓	Duration ↑↓	Integration runtime ↑↓	User prop... ↑↓	Activity run ID ↑↓
Silver to Gold	✅ Succeeded	Notebook	4/14/2025, 3:25:37 AM	49s	AutoResolveIntegrationRuntime (Central India)	6138e2bb-0915-466d-9a95-de9	
bronze to silver	✅ Succeeded	Notebook	4/14/2025, 3:24:46 AM	50s	AutoResolveIntegrationRuntime (Central India)	0269c6e0-9658-4775-8ecf-4596	
Copy data table	✅ Succeeded	Copy data	4/14/2025, 3:24:21 AM	25s	SHIR	99badbb7-7181-4bbc-a05c-38d	

No need to rerun the synapse analytics pipeline it will get the views automatically with the help of scripts

Validate Debug Add trigger

Get Metadata

metadata gold_db

Parameters Variables Settings **Output**

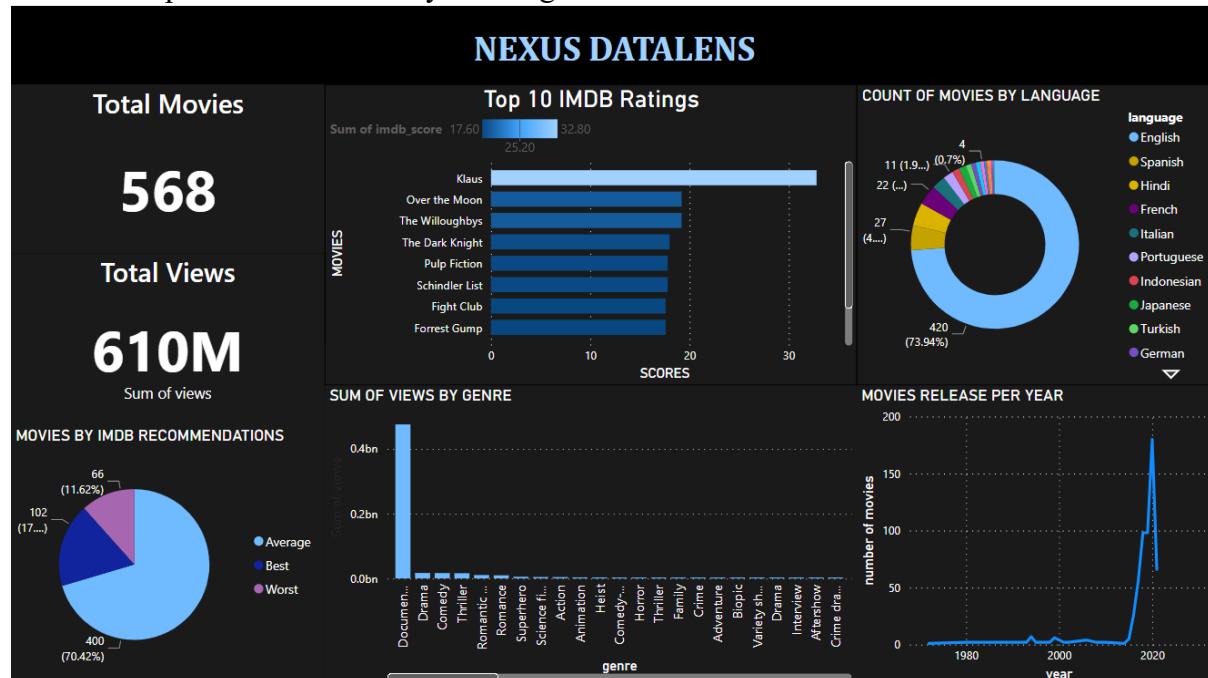
Pipeline run ID: be2b6cd6-9b06-48a5-9d40-e11c8905ebff Pipeline status Succeeded

All status

Showing 1 - 1 of 1 items

Activity name	Activity st...	Activit...	Run start	Duration	Integration runtime
metadata gold_db	Succeeded	Get Metadata	4/14/2025, 3:34:29 AM	3s	AutoResolveIntegrationRuntime (Central India)

Refresh the power bi dashboard you will get the Visualized data



Nexus DataLens

Got to SQL Server Management Studio (SSMS) for Incremental Data

```
SQLQuery4.sql - GAURAV\SQLEXPRESS.NexusDB (GAURAV\Hp (65)) - Microsoft SQL Server Management Studio
File Edit View Query Project Tools Window Help
New Query Execute
Object Explorer
Connect Connect ...
GAURAV\SQLEXPRESS (SQL Server 16.0.1000 - GAURAV\Hp)
Databases
System Databases
Database Snapshots
NexusDB
Database Diagrams
Tables
System Tables
FileTables
External Tables
Graph Tables
dbo.data
Views
External Resources
Synonyms
Programmability
Query Store
Service Broker
Storage
Security
Security
Server Objects
Replication
Management
XEvent Profiler
SQLQuery4.sql - G...B (GAURAV\Hp (65))
INSERT INTO dbo.data (Title, Genre, ReleaseDate, Runtime, IMDB_Score, Language, Views, AddedDate)
VALUES
('Inception', 'Sci-Fi / Thriller', '2010-07-16', 148, 8.8, 'English', 320000, '2023-09-15'),
('Parasite', 'Thriller / Drama', '2019-05-21', 132, 8.6, 'Korean', 250000, '2023-09-15'),
('The Grand Budapest Hotel', 'Comedy / Adventure', '2014-03-07', 99, 8.1, 'English', 180000, '2023-09-15'),
('La La Land', 'Musical / Romance', '2016-12-09', 128, 8.0, 'English', 170000, '2023-09-15'),
('Spirited Away', 'Animation / Fantasy', '2001-07-20', 125, 8.6, 'Japanese', 160000, '2023-09-15'),
('The Revenant', 'Adventure / Drama', '2015-12-25', 156, 8.0, 'English', 150000, '2023-09-15'),
('Whiplash', 'Drama / Music', '2014-10-10', 106, 8.5, 'English', 140000, '2023-09-15'),
('Mad Max: Fury Road', 'Action / Sci-Fi', '2015-05-15', 120, 8.1, 'English', 130000, '2023-09-15'),
('Spider-Man: Into the Spider-Verse', 'Animation / Action', '2018-12-14', 117, 8.4, 'English', 120000, '2023-09-15'),
('Get Out', 'Horror / Thriller', '2017-02-24', 104, 7.7, 'English', 110000, '2023-09-15'),
('Roma', 'Drama / Historical', '2018-11-21', 135, 7.9, 'Spanish', 100000, '2023-09-15'),
('Joker', 'Crime / Drama', '2019-10-04', 122, 8.4, 'English', 90000, '2023-09-15'),
('1917', 'War / Drama', '2019-12-25', 119, 8.3, 'English', 80000, '2023-09-15'),
('The Shape of Water', 'Romance / Fantasy', '2017-12-01', 123, 7.3, 'English', 70000, '2023-09-15'),
('Coco', 'Animation / Family', '2017-11-22', 105, 8.4, 'Spanish/English', 60000, '2023-09-15'),
('A Star is Born', 'Musical / Drama', '2018-10-05', 136, 7.6, 'English', 50000, '2023-09-15'),
('Blade Runner 2049', 'Sci-Fi / Thriller', '2017-10-06', 164, 8.0, 'English', 40000, '2023-09-15'),
('Black Panther', 'Action / Adventure', '2018-02-16', 134, 7.3, 'English', 30000, '2023-09-15'),
('Bohemian Rhapsody', 'Biography / Music', '2018-10-24', 134, 7.9, 'English', 20000, '2023-09-15'),
('Knives Out', 'Mystery / Comedy', '2019-11-27', 130, 7.9, 'English', 10000, '2023-09-15');
```

Run the Insert Query

```
SQLQuery4.sql - G...B (GAURAV\Hp (65))
INSERT INTO dbo.data (Title, Genre, ReleaseDate, Runtime, IMDB_Score, Language, Views, AddedDate)
VALUES
('Inception', 'Sci-Fi / Thriller', '2010-07-16', 148, 8.8, 'English', 320000, '2023-09-15'),
('Parasite', 'Thriller / Drama', '2019-05-21', 132, 8.6, 'Korean', 250000, '2023-09-15'),
('The Grand Budapest Hotel', 'Comedy / Adventure', '2014-03-07', 99, 8.1, 'English', 180000, '2023-09-15'),
('La La Land', 'Musical / Romance', '2016-12-09', 128, 8.0, 'English', 170000, '2023-09-15'),
('Spirited Away', 'Animation / Fantasy', '2001-07-20', 125, 8.6, 'Japanese', 160000, '2023-09-15'),
('The Revenant', 'Adventure / Drama', '2015-12-25', 156, 8.0, 'English', 150000, '2023-09-15'),
('Whiplash', 'Drama / Music', '2014-10-10', 106, 8.5, 'English', 140000, '2023-09-15'),
('Mad Max: Fury Road', 'Action / Sci-Fi', '2015-05-15', 120, 8.1, 'English', 130000, '2023-09-15'),
('Spider-Man: Into the Spider-Verse', 'Animation / Action', '2018-12-14', 117, 8.4, 'English', 120000, '2023-09-15'),
('Get Out', 'Horror / Thriller', '2017-02-24', 104, 7.7, 'English', 110000, '2023-09-15'),
('Roma', 'Drama / Historical', '2018-11-21', 135, 7.9, 'Spanish', 100000, '2023-09-15'),
('Joker', 'Crime / Drama', '2019-10-04', 122, 8.4, 'English', 90000, '2023-09-15'),
('1917', 'War / Drama', '2019-12-25', 119, 8.3, 'English', 80000, '2023-09-15'),
('The Shape of Water', 'Romance / Fantasy', '2017-12-01', 123, 7.3, 'English', 70000, '2023-09-15'),
('Coco', 'Animation / Family', '2017-11-22', 105, 8.4, 'Spanish/English', 60000, '2023-09-15'),
('A Star is Born', 'Musical / Drama', '2018-10-05', 136, 7.6, 'English', 50000, '2023-09-15'),
('Blade Runner 2049', 'Sci-Fi / Thriller', '2017-10-06', 164, 8.0, 'English', 40000, '2023-09-15'),
('Black Panther', 'Action / Adventure', '2018-02-16', 134, 7.3, 'English', 30000, '2023-09-15'),
('Bohemian Rhapsody', 'Biography / Music', '2018-10-24', 134, 7.9, 'English', 20000, '2023-09-15'),
('Knives Out', 'Mystery / Comedy', '2019-11-27', 130, 7.9, 'English', 10000, '2023-09-15')

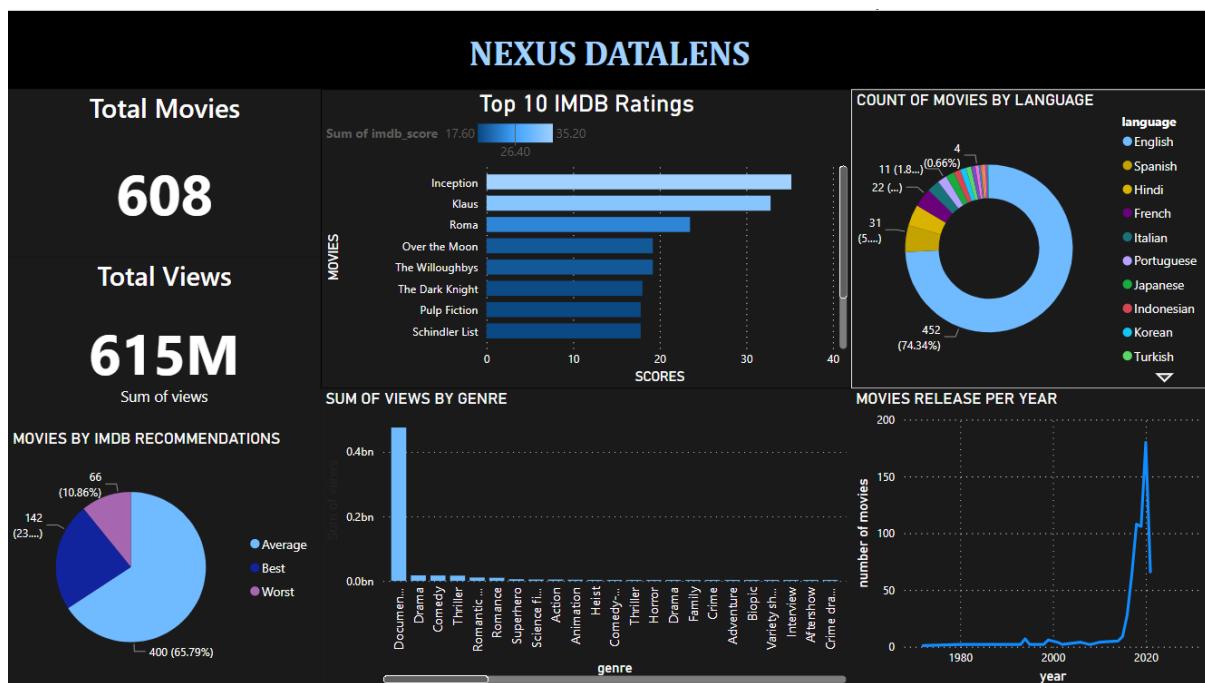
(20 rows affected)

Completion time: 2025-04-14T03:39:12.6628906+05:30
```

Go to Azure Data factory and re run the Pipeline

Pipeline name	Run start	Duration	Triggered by	Status	Parameters
copydatapipeline	4/14/2025, 3:42:04 AM	1m 11s	Manual trigger	Succeeded	
copydatapipeline	4/14/2025, 3:40:35 AM	1m 17s	Manual trigger	Succeeded	

Updated Data In the visualized format



FUTURE SCOPE

The current implementation of an incremental ETL pipeline using Azure services lays a strong foundation for data integration, transformation, and analytics. However, there are numerous opportunities to extend and enhance the system to meet evolving business requirements and leverage cutting-edge technologies. The following areas represent the future scope of the project:

1. Integration with Advanced Analytics and Artificial Intelligence

- Machine Learning Models : Extend the pipeline to incorporate predictive analytics and anomaly detection using Azure Machine Learning or AutoML . This will enable proactive decision-making based on historical trends and real-time insights.
- Natural Language Processing (NLP) : Integrate NLP capabilities to analyze unstructured data such as customer feedback, logs, or social media feeds stored in Azure Data Lake.
- Recommendation Systems : Build recommendation engines tailored to specific industries, such as personalized product suggestions in retail or treatment recommendations in healthcare.

2. Real-Time Data Processing

- Streaming Pipelines : Transition to real-time data processing by integrating Azure Event Hubs or Apache Kafka . This will support use cases like IoT data ingestion, fraud detection, and live monitoring of operational metrics.
- Delta Lake : Implement Delta Lake in Azure Data Lake Gen2 to enable ACID transactions and incremental updates for large-scale datasets, ensuring consistency and reliability in real-time workflows.

3. Scalability and Performance Optimization

- Serverless Architecture : Adopt serverless components such as Azure Functions or Databricks Serverless to reduce infrastructure management overhead and optimize costs.
- Auto-Scaling : Enable auto-scaling for compute resources in Azure Databricks and Azure Synapse Analytics to handle peak workloads efficiently.
- Data Partitioning : Optimize storage and query performance by partitioning data in Azure Data Lake based on attributes like date, region, or category.

4. Multi-Cloud and Hybrid Cloud Support

- Multi-Cloud Integration : Extend the pipeline to integrate with other cloud platforms like AWS or Google Cloud, providing redundancy and flexibility for global operations.
- Hybrid Cloud Enhancements : Strengthen hybrid cloud capabilities by supporting additional on-premises systems or edge devices for data ingestion and processing.

5. Enhanced Security and Compliance

- Advanced Encryption : Implement advanced encryption techniques, such as Homomorphic Encryption , to secure sensitive data during processing.
- Data Masking : Use Azure SQL Database Dynamic Data Masking to protect sensitive information during visualization and reporting.
- Compliance Automation : Automate compliance checks for regulations like GDPR, HIPAA, and CCPA using tools like Azure Policy and Compliance Manager .

6. User-Friendly Interfaces and Self-Service Analytics

- Custom Dashboards : Develop role-specific dashboards in Power BI tailored to executives, analysts, and engineers, ensuring relevant insights for each user group.
- Self-Service Tools : Enable non-technical users to create their own reports and analyses using Power Query or Power BI Templates .
- Chatbots for Insights : Integrate chatbots powered by Azure Bot Service to allow users to query data and receive insights via natural language.

7. Blockchain for Data Integrity

- Immutable Logs : Use blockchain technology to maintain immutable logs of data transformations, ensuring transparency and traceability throughout the pipeline.
- Smart Contracts : Implement smart contracts to automate workflows such as data validation, approval processes, and compliance checks.

8. Edge Computing and IoT Integration

- Edge Analytics : Process data at the edge using Azure IoT Edge to reduce latency and bandwidth usage, enabling real-time insights for IoT devices.
- IoT Data Pipelines : Extend the pipeline to ingest and process data from IoT devices, supporting use cases like predictive maintenance, smart city analytics, and environmental monitoring.

9. Cross-Industry Applications

The project can be adapted to serve diverse industries, each with unique requirements:

- Retail : Inventory optimization, demand forecasting, and personalized marketing.
- Healthcare : Patient data analytics, disease prediction, and telemedicine insights.
- Finance : Fraud detection, risk analysis, and regulatory reporting.
- Manufacturing : Predictive maintenance, supply chain optimization, and quality control.

10. Cost Optimization

- Cost Monitoring : Use Azure Cost Management to monitor and optimize resource usage, ensuring cost efficiency.
- Reserved Instances : Leverage reserved instances for predictable workloads to reduce costs.
- Spot Instances : Use Azure Spot VMs for non-critical tasks like batch processing or testing.

11. Collaboration and DevOps Enhancements

- CI/CD Pipelines : Implement continuous integration and deployment (CI/CD) for ETL pipelines using Azure DevOps or GitHub Actions , ensuring seamless updates and version control.
- Collaborative Notebooks : Use collaborative notebooks in Azure Databricks to enable teams to work together on data transformations and analyses.

12. Sustainability and Green Computing

- Energy-Efficient Workflows : Optimize workflows to minimize energy consumption, aligning with sustainability goals.
- Carbon Tracking : Use Azure's Sustainability Calculator to track and reduce the carbon footprint of the data pipeline.

13. Data Governance and Metadata Management

- Data Catalog : Implement Azure Purview to create a centralized data catalog for metadata management and data discovery.
- Data Lineage : Track data lineage to understand how data flows through the pipeline and identify dependencies.

14. Global Expansion

- Multi-Region Deployment : Deploy the pipeline across multiple Azure regions to ensure high availability and low latency for global users.
- Localization : Support localization for dashboards and reports to cater to international audiences.

15. Open Source Contributions

- Open Source Tools : Contribute to open-source projects related to data engineering, such as Apache Spark or Delta Lake, to enhance community-driven innovation.
- Reusable Components : Package reusable components of the pipeline (e.g., transformation scripts) as open-source libraries for broader adoption.

CONCLUSION

Nexus DataLens offers a comprehensive, end-to-end data analysis solution designed to simplify the way users interact with and extract value from their data. With an intuitive, user-friendly interface, it allows seamless data uploads, efficient processing, and insightful visualizations, ensuring accessibility for both technical and non-technical users. Advanced security measures safeguard sensitive information, while robust analytical capabilities enable businesses to uncover trends, identify patterns, and make data-driven decisions with confidence. By leveraging cutting-edge technologies such as AI-driven insights, automation, and scalable cloud infrastructure, Nexus DataLens optimizes complex data workflows, enhances operational efficiency, and unlocks the full potential of data, driving innovation and business growth.

REFERENCES

1. SQL Server Management Studio (SSMS)
<https://learn.microsoft.com/en-us/ssms/sql-server-management-studio-ssms>
2. Azure Data Factory
<https://learn.microsoft.com/en-us/azure/data-factory/>
3. Azure Data Lake Storage Gen 2
<https://learn.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-introduction>
4. Azure Data Bricks
<https://learn.microsoft.com/en-us/azure/databricks/>
5. Azure Synapse Analytics
<https://learn.microsoft.com/en-us/azure/synapse-analytics/>
6. Power Bi
<https://learn.microsoft.com/en-us/power-bi/>
7. Key Vault
<https://learn.microsoft.com/en-us/azure/key-vault/>
8. Active Directory
<https://learn.microsoft.com/en-us/troubleshoot/windows-server/active-directory/active-directory-overview>
9. Udemy Courses
 - End to End Azure Data Engineering Real Time Project
<https://www.udemy.com/course/end-to-end-azure-data-engineering-real-time-project/?couponCode=NVDIN35>
 - Azure Data Factory +Synapse Analytics End to End ETL project
<https://www.udemy.com/course/azure-data-factory-synapse-analytics-end-to-end-etl-project/?couponCode=NVDIN35>