Data Engineering in the Cloud

Azure Synapse Analytics

Xuemao Zhang
East Stroudsburg University

January 18, 2025

Outline

- Introduction to Azure Synapse Analytics
- Lab: Creating an Azure Synapse Workspace

Introduction

- You are hired as an Azure data engineer. Your organization wants you to ingest, explore, prepare, and serve the data for instant BI and machine learning requirements to derive some decisions out of it.
  - For this task, you have decided to utilize Azure Synapse Analytics and Synapse Workspace, which is a complex mechanism.
- **Azure Synapse Analytics** is an integrated analytics service that accelerates time to insight across data warehouses and big data systems. It brings together data integration, enterprise data warehousing, and big data analytics to offer a single unified experience for ingesting, preparing, managing, and serving data for immediate business intelligence and machine learning needs.

Introduction

- It includes processes like data integration, enterprise data warehousing, and big data analytics.
- It can be used for data ingestion, exploration, management, and serving of data.
- In summary, Azure Synapse Analytics offers a unified platform for handling large-scale data querying and analytics, integrating various data processing and management capabilities to drive business intelligence and insights.

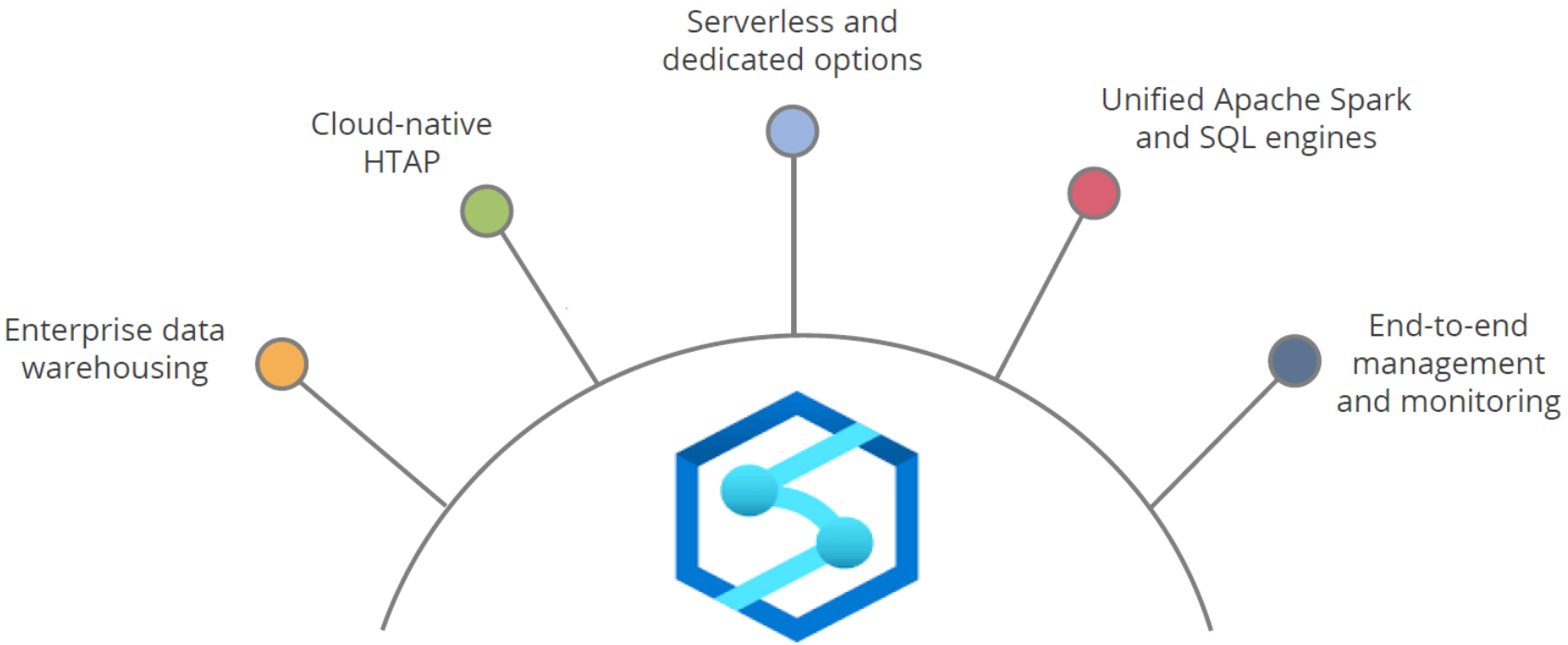
Introduction

- **Data Warehouse** is designed for **structured data** storage, retrieval, and analysis. It is optimized for querying and reporting.
- Enterprise Data Warehousing (EDW) is a **centralized repository** that consolidates data from different sources within an organization.
  - It is designed to support decision-making processes by providing a unified, consistent, and comprehensive view of the organization's data.
  - EDWs enable complex queries and analysis, facilitate reporting, and support business intelligence activities.

Use Cases of EDW

- Executive Reporting: Provides management with comprehensive reports and dashboards to monitor business performance.
- Operational Efficiency: Analyzes operational data to identify inefficiencies and areas for improvement.
- Customer Insights: Consolidates customer data to provide a 360-degree view, enabling personalized marketing and improved customer service.
- Financial Analysis: Aggregates financial data to support budgeting, forecasting, and financial reporting.
- Compliance and Auditing: Stores and manages data to ensure compliance with regulatory requirements and facilitate audits.

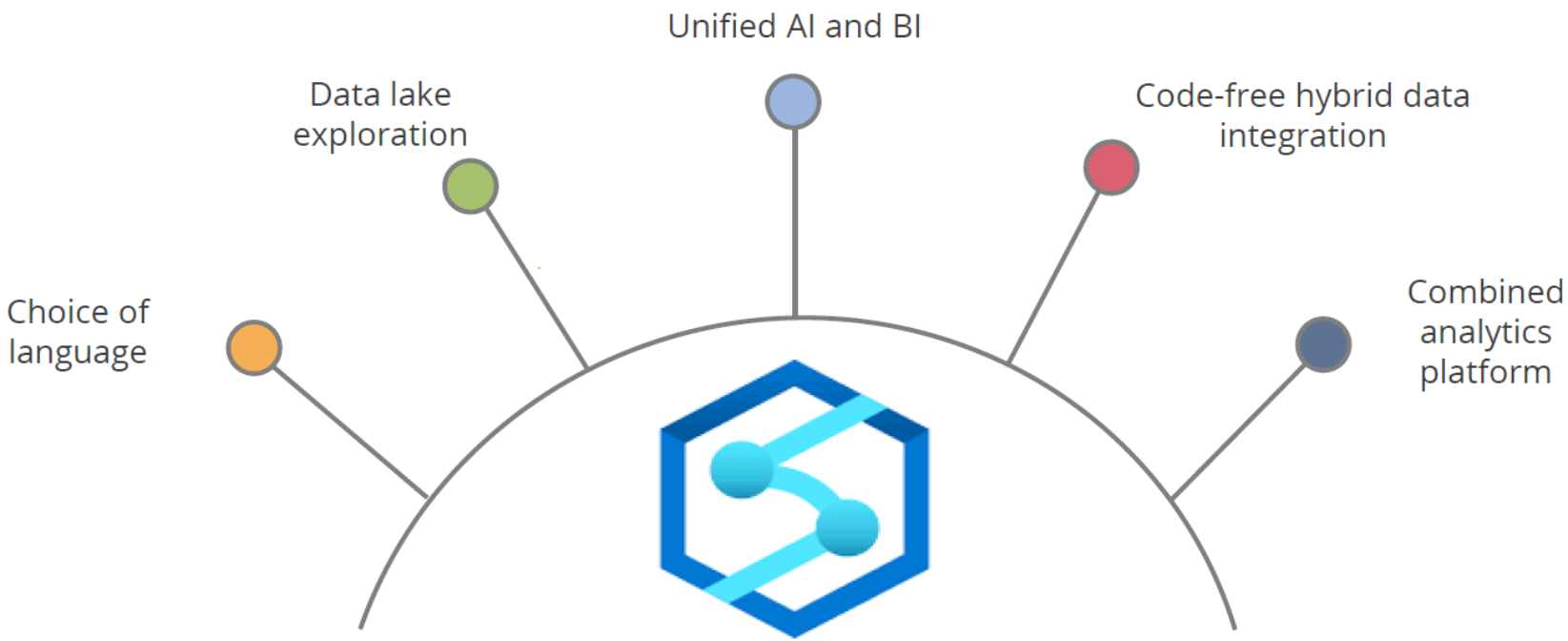
Azure Synapse Analytics: Features



Azure Synapse Analytics: Features

- Enterprise data warehousing: Robust, scalable data warehousing capabilities.
- Cloud native HTAP: Hybrid Transactional and Analytical Processing.
- Serverless and dedicated options: Flexible computing options to suit various workloads.
  - Serverless is a cloud computing model in which the cloud provider dynamically manages the allocation and provisioning of servers. In a serverless architecture, developers can focus on writing code without worrying about the underlying infrastructure.
- Unified Apache Spark and SQL engines: Seamless integration of big data and relational data.
- End to end management and monitoring: Comprehensive tools for data management and performance monitoring

Azure Synapse Analytics: Features



Azure Synapse Analytics: Features

- Choice of language: Support for multiple languages: SQL, Python, R, Scala.
- Data lake exploration: Efficient data exploration and analytics within the data lake.
- Unified AI and BI: Integration with Azure Machine Learning and Power BI.
- Code free hybrid data integration: Simplified data integration with visual tools.
- Combined analytics platform: A single platform for all data analytics needs.

Azure Synapse Workspace

- Azure Synapse workspace combines enterprise data warehousing and big data analytics.
- Microsoft's Azure Synapse Analytics service is a software as a service (SaaS). It can be utilized as required and consists of four parts:

Azure Synapse Workspace

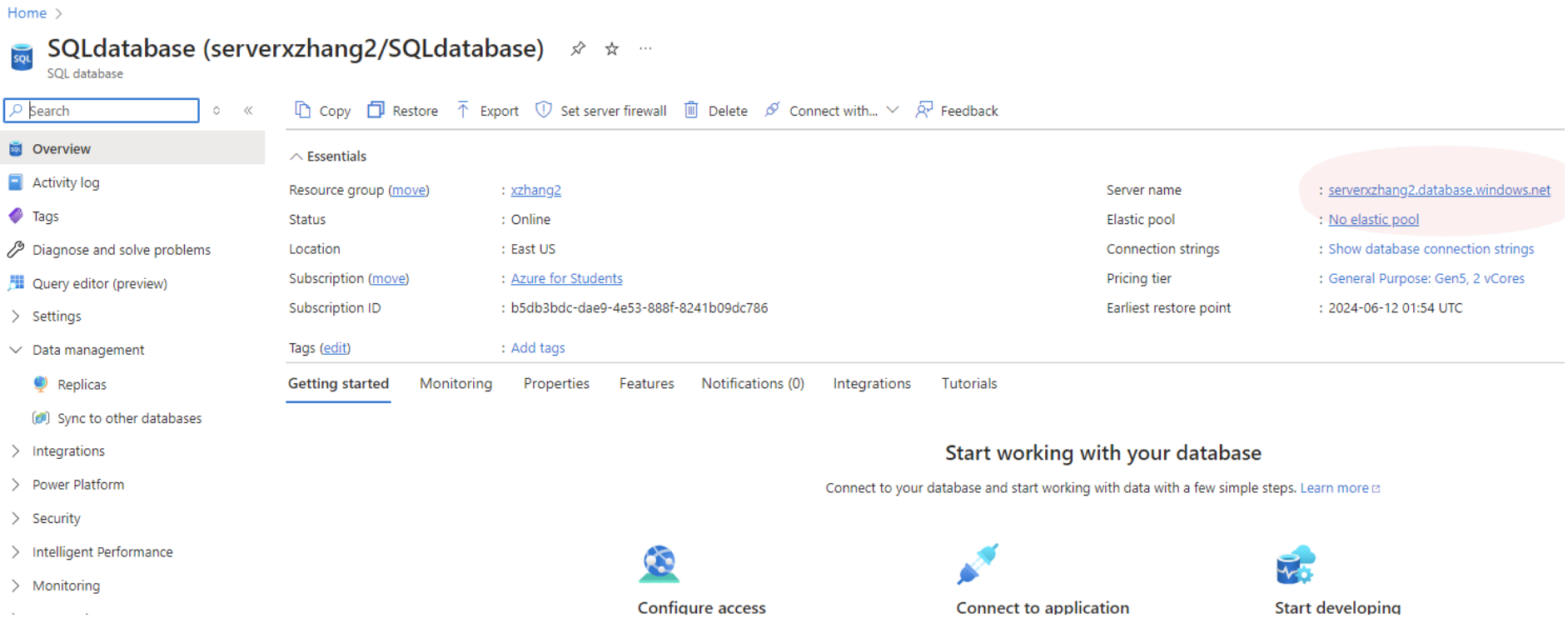
- Connectors: Facilitate integration with various data sources.
- SQL Cluster: Enables scalable and high-performance SQL-based data warehousing.
- Spark: Supports big data processing and analytics with Apache Spark.
- SQL On Demand: Allows for serverless querying of data, enabling on-demand analytics without the need to provision resources.

Labs: Creating an Azure Synapse Workspace

- Prerequisites (Lab in Lecture03):
  - Create a datalake (storage account with hierarchical namespace enabled)
  - Create a container
  - Upload a data set to the container
  - Create a data base
- Objective: Creating an Azure Synapse Workspace

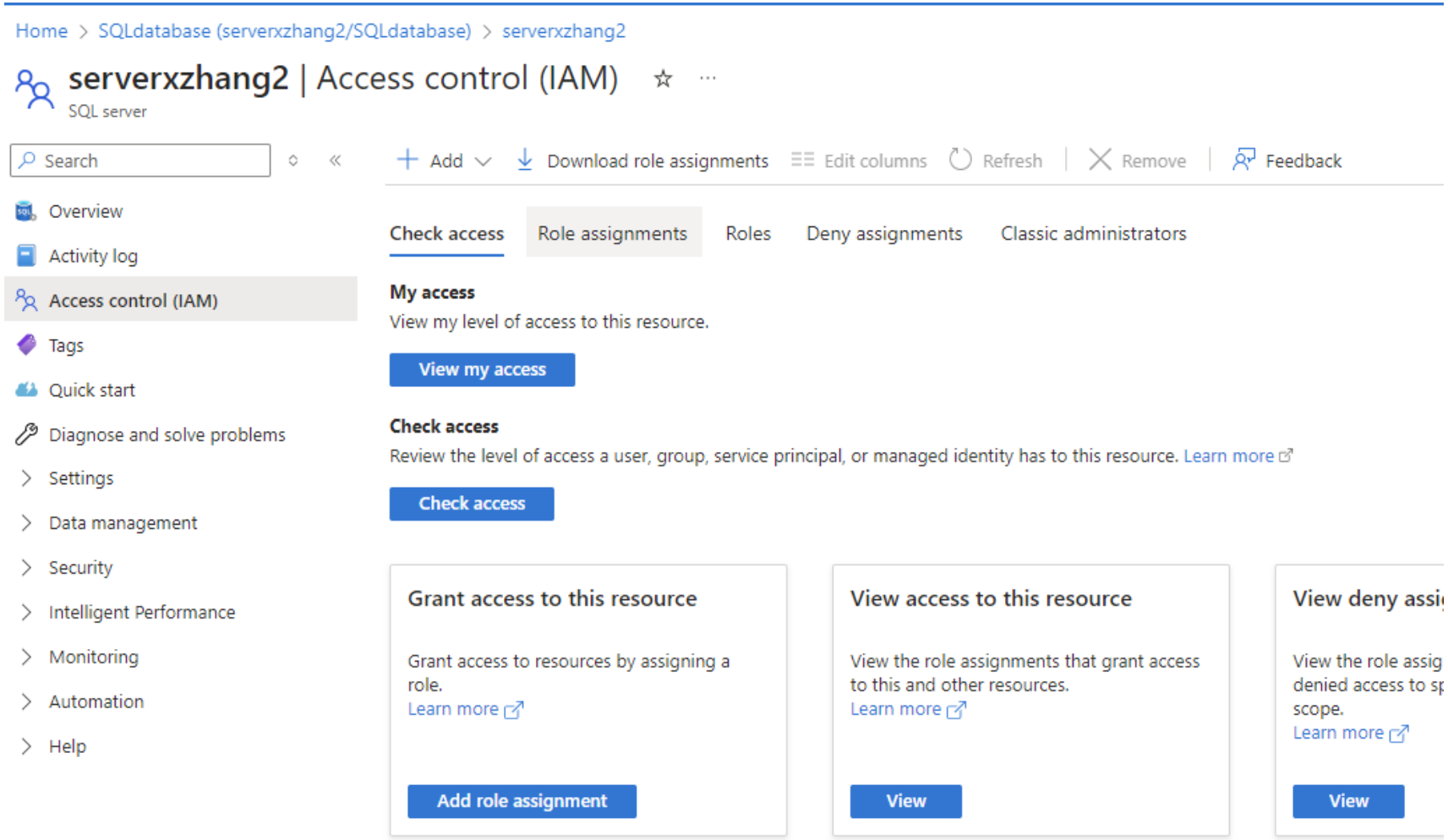
Lab 1: SQL server settings

- Open the database and turn on the server



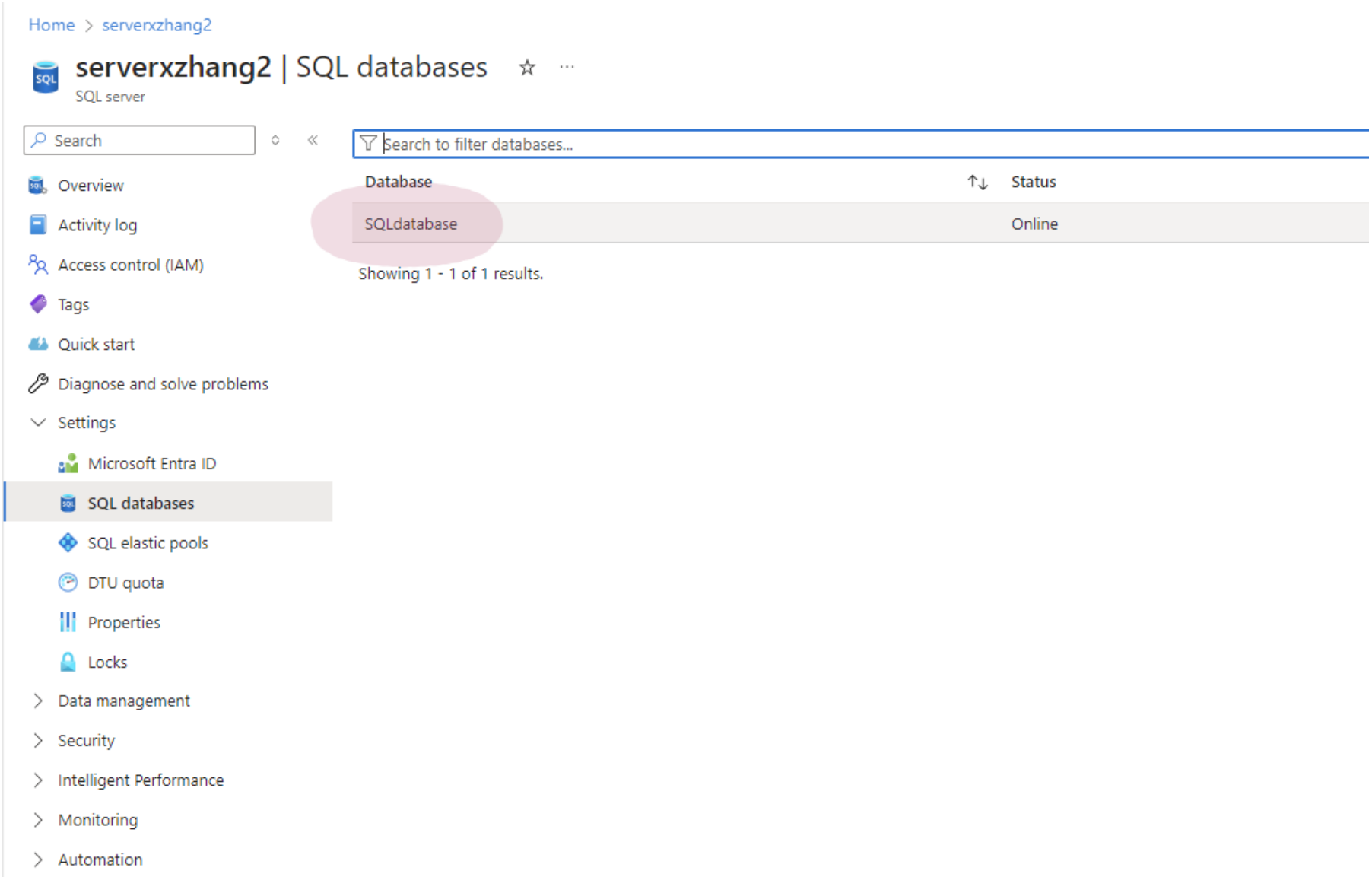
Lab 1: SQL server settings

- Check the IAM



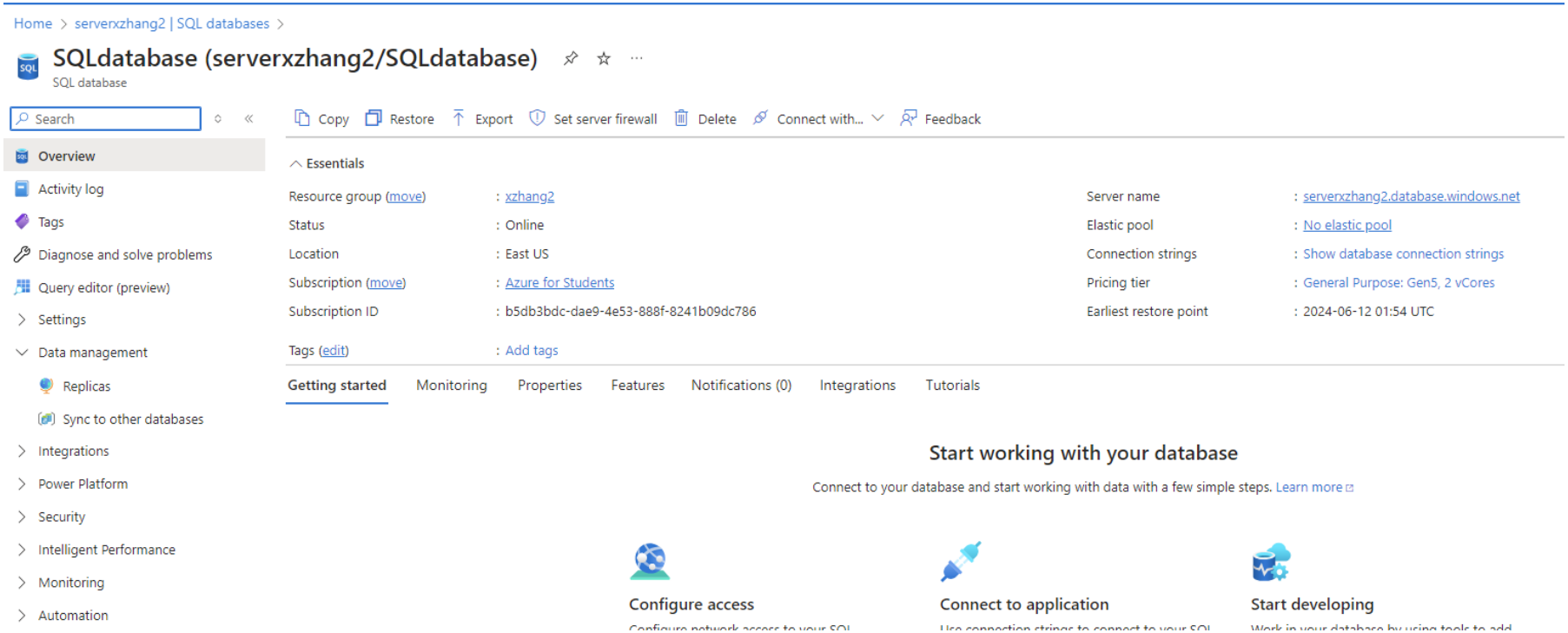
Lab 1: SQL server settings

- Check the settings of the database



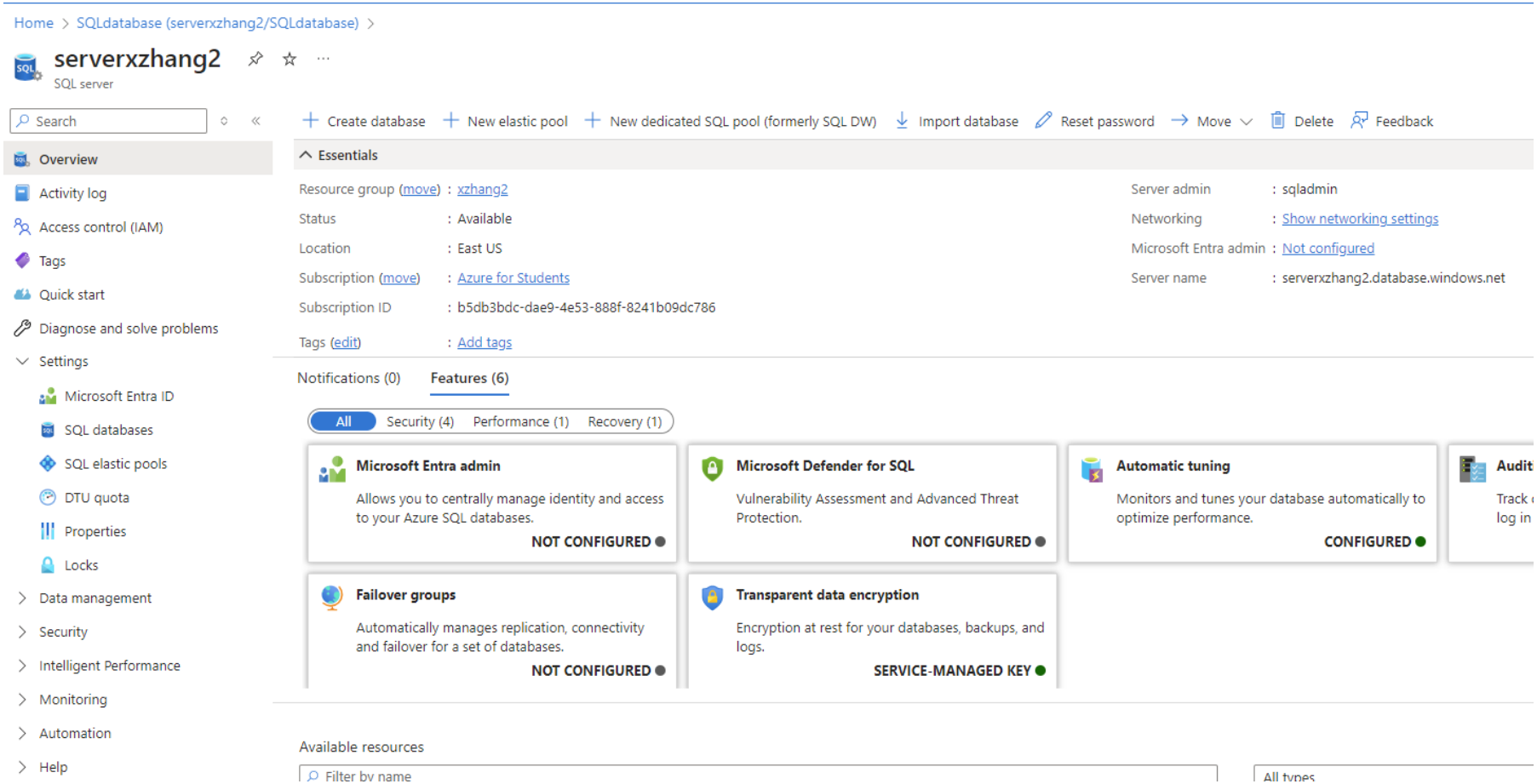
Lab 1: SQL server settings

- Then you will see a similar screen



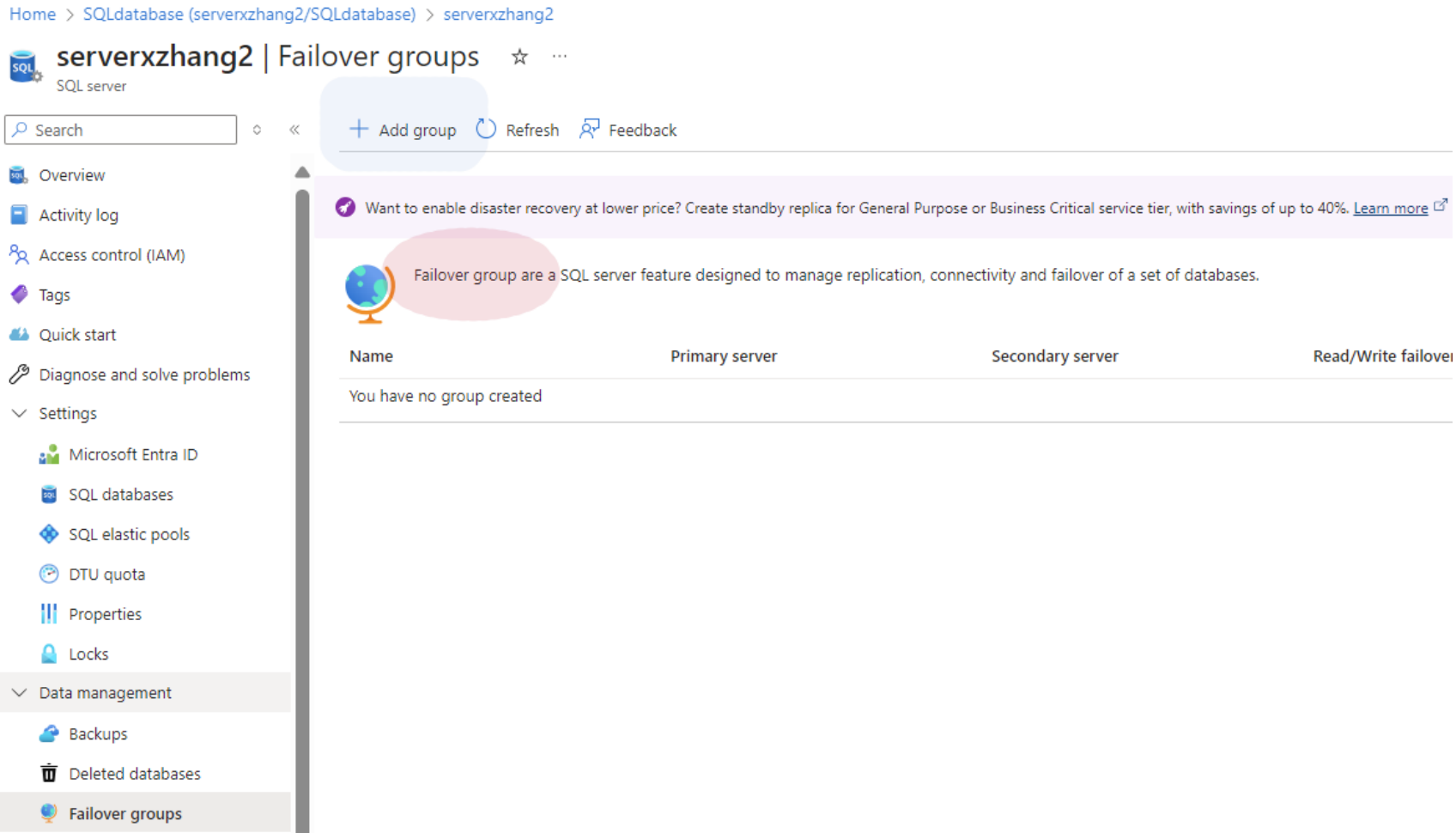
Lab 1: SQL server settings

- Click on the server name, you will see



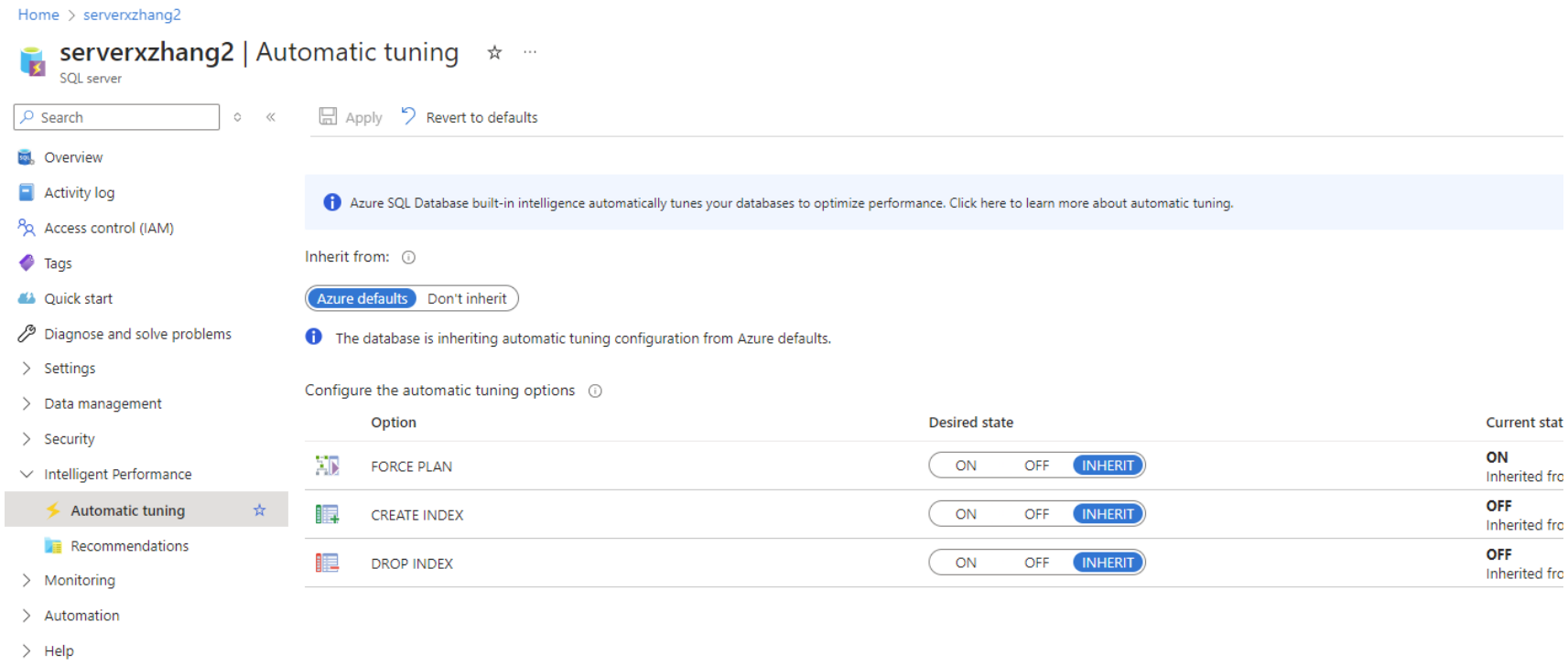
Lab 1: SQL server settings

- Check data management
  - Failover group are a SQL server feature designed to manage replication, connectivity and failover of a set of databases.
  - If you want to create a `Failover group`, just click `Add group`
    - ★ Read Configure a failover group for Azure SQL Database for more information.



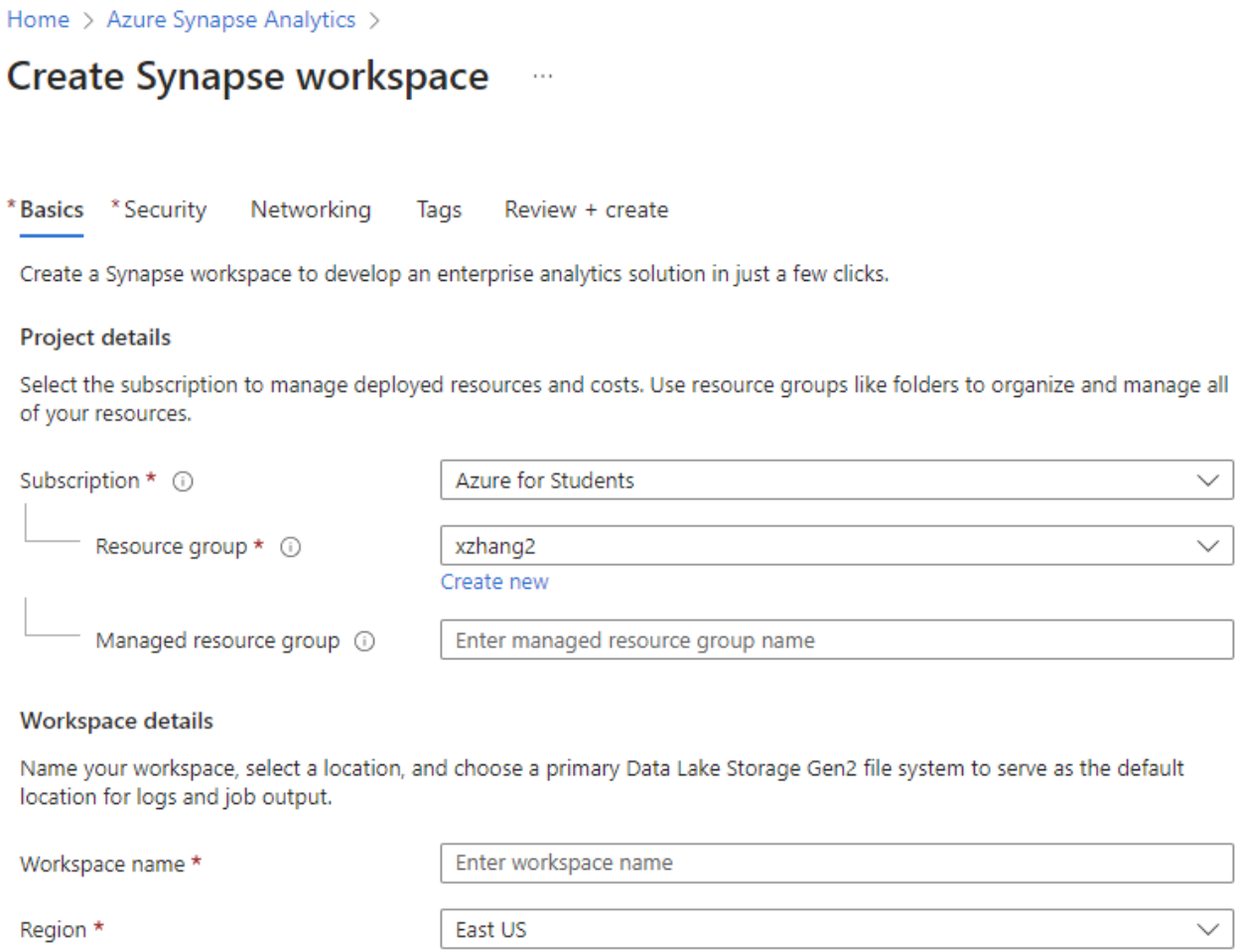
Lab 1: SQL server settings

- Check `Automatic tuning` under `Intelligent tuning`
  - It helps optimize database performance by automatically applying performance tuning recommendations.



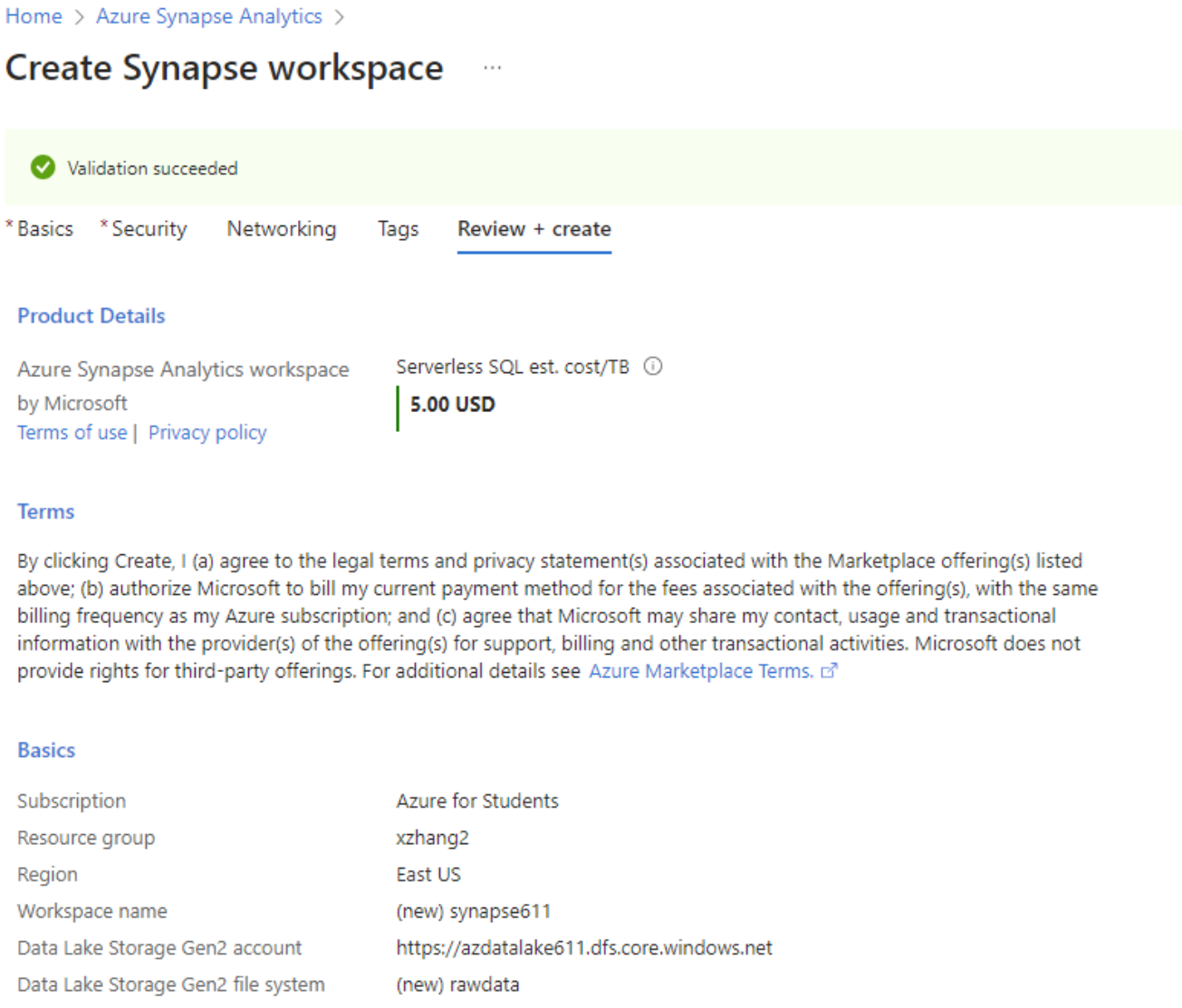
Lab 2: Creating an Azure Synapse Workspace

- Now go back to `Home` and search for `Synapse`, then choose `Azure Synapse Analytics` → `Create Synapse workspace`
- Basics
  - Specify a Workspace name. (Mine is synapse611)
  - Let's add a new container called `rawdata`

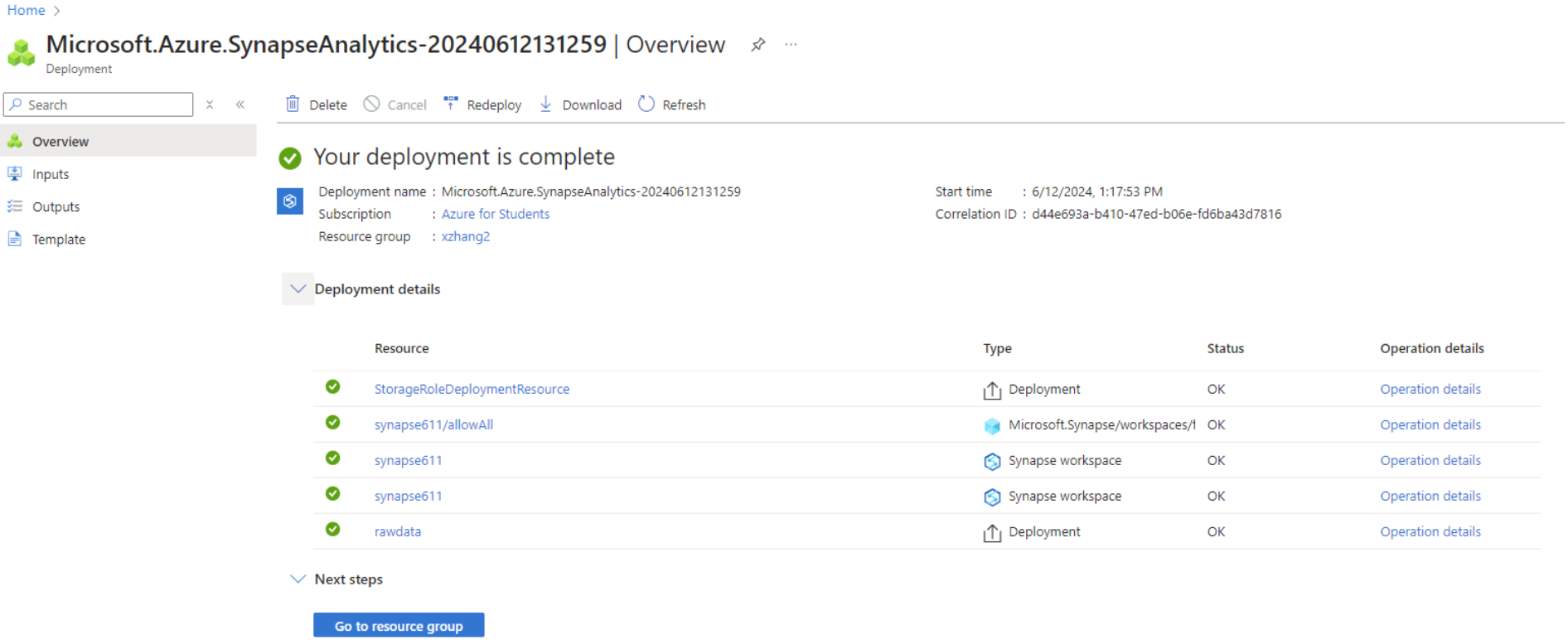


Lab 2: Creating an Azure Synapse Workspace

- In the security tab, specify the password (username is sqladminuser)
- keep the other settings to default and go to `review + create`

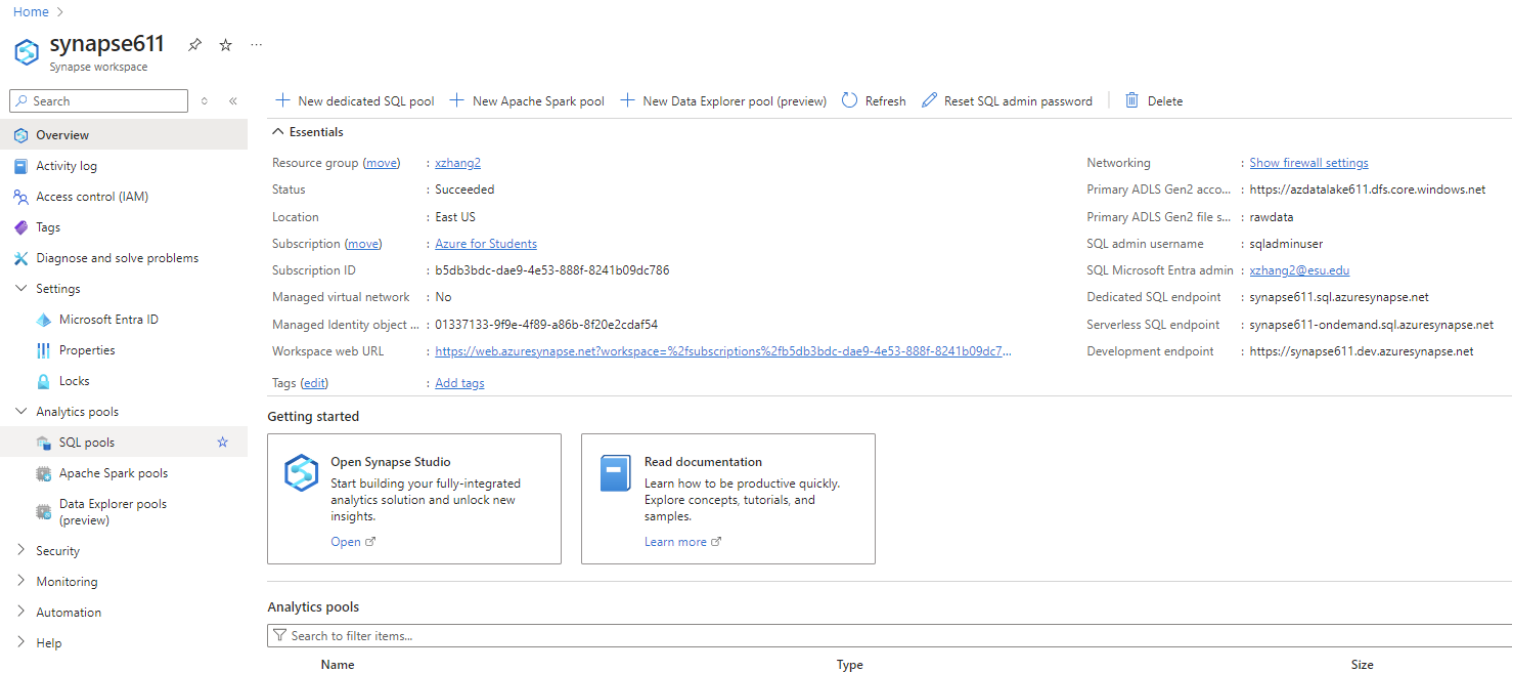


Lab 2: Creating an Azure Synapse Workspace



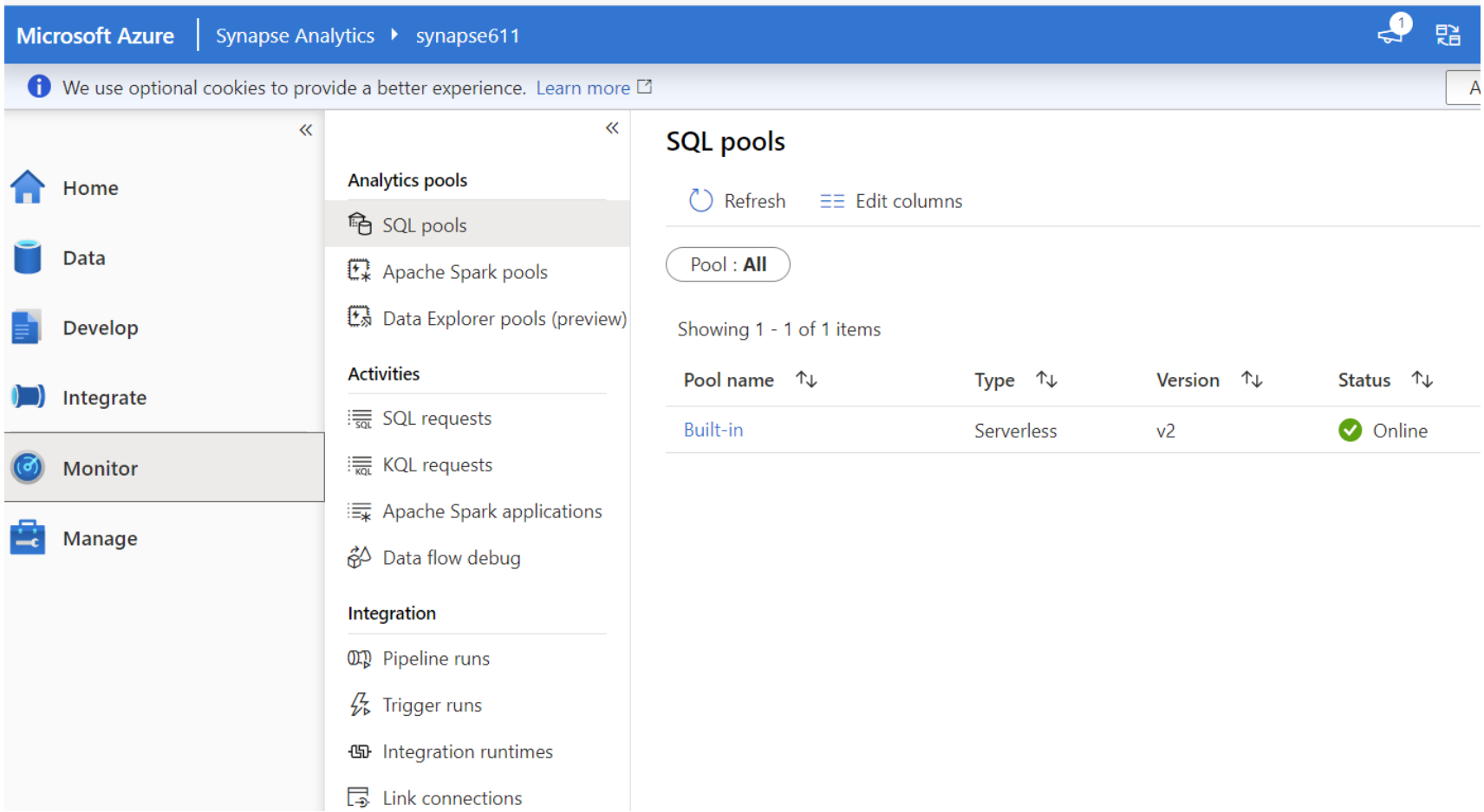
Lab 2: Creating an Azure Synapse Workspace

- Once the deployment is completed, click on the `go to resource` button
- Then, open the Synapse workspace we just created
- Now in the synapse workspace, go to `Analytic pools`



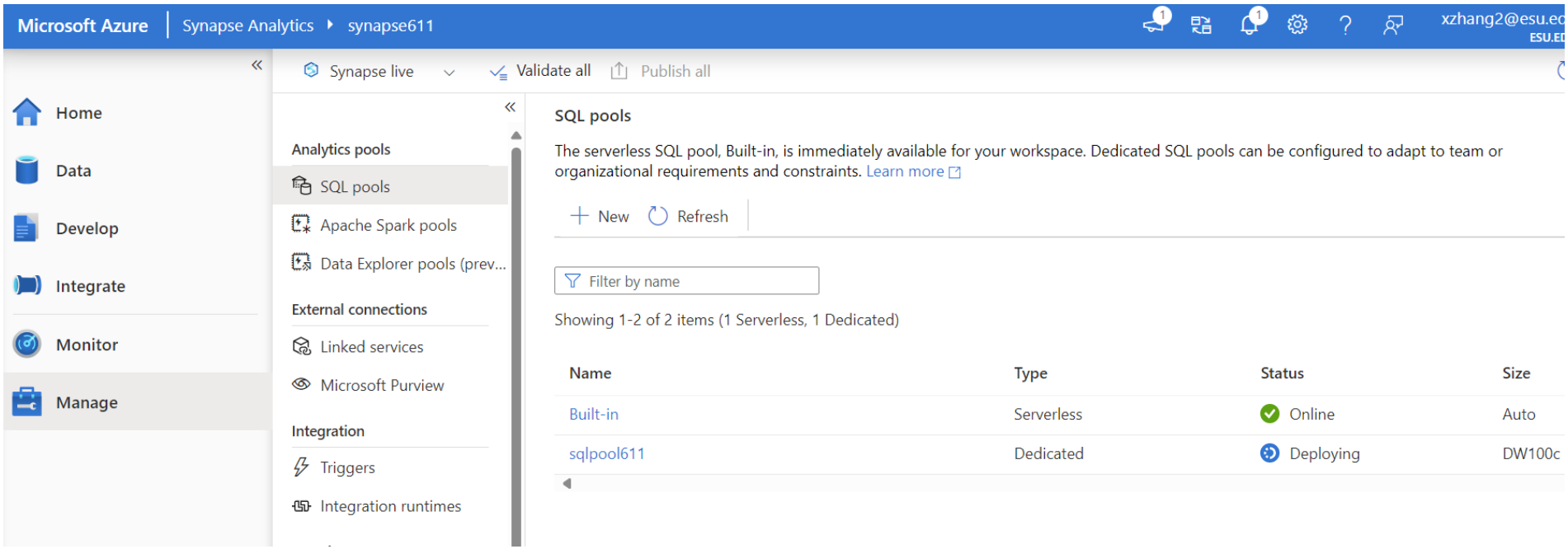
Lab 2: Creating an Azure Synapse Workspace

- Click on `Open` in `Open Synapse Studio` box
  - In the Synapse Studio, you can see the SQL pools



Lab 2: Creating an Azure Synapse Workspace

- The serverless SQL pool, Built-in, is immediately available for your workspace. Dedicated SQL pools can be configured to adapt to team or organizational requirements and constraints.
- Let's add a new Dedicated (serverless) SQL pool
  - Specify the name for the SQL pool.
  - Choose the performance level. **Note**: the pricing will depend on the performance level being selected. Set the value for Performance Level as 'DW100c'.



Lab 2: Creating an Azure Synapse Workspace

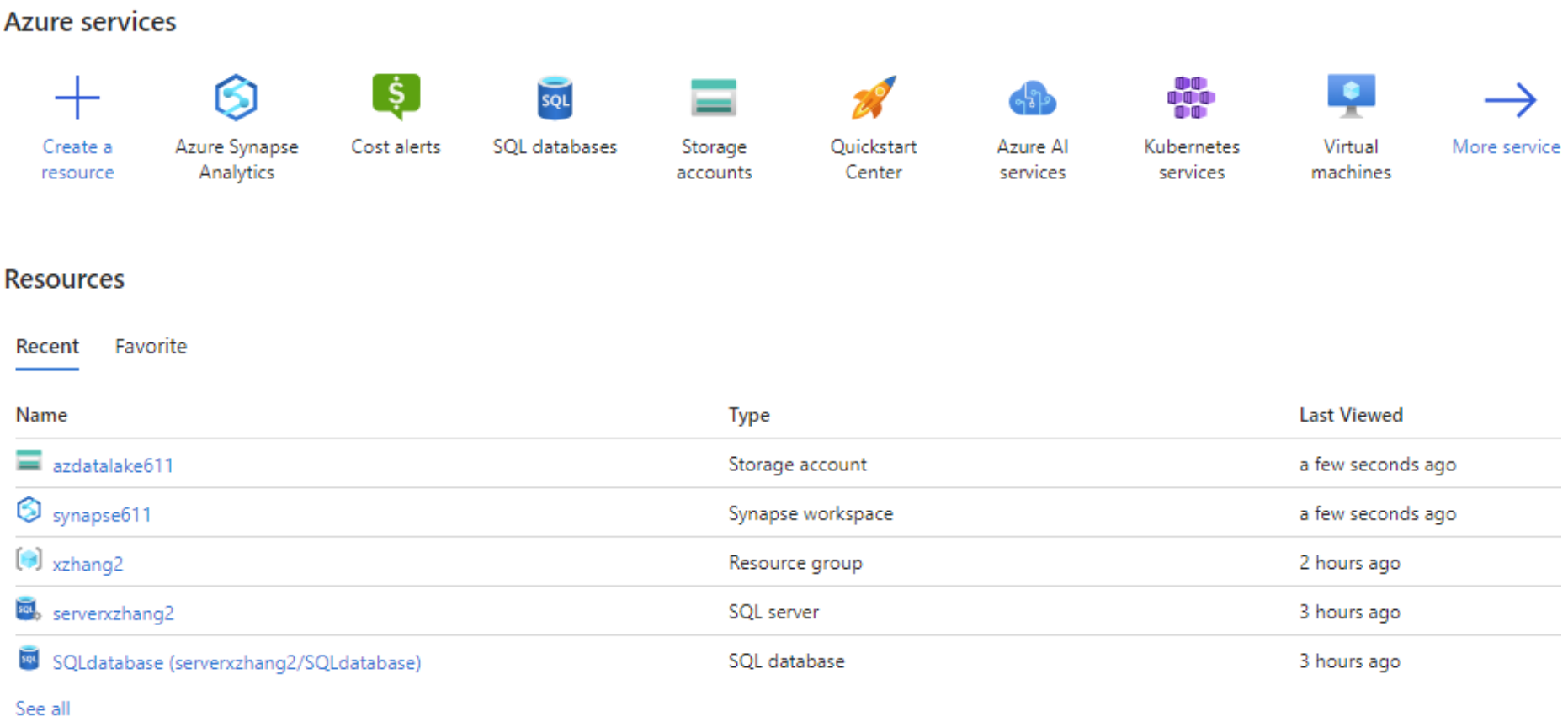
- SQL pool is the method for querying data in Azure Synapse.
  - The data is stored in a relational table structure with column storage.
- The serverless SQL pool is defined as the mechanism of querying data for a data lake.
  - The storage and serverless SQL pool must be in the **same region**.

Lab 2: Creating an Azure Synapse Workspace

- Dedicated SQL pool provides a dedicated set of resources for data warehousing. It's a fully managed, enterprise-grade cloud data warehouse.
- Dedicated SQL pool is suitable for consistent, large-scale data warehousing needs, offering high performance and dedicated resources at a higher cost and with more management overhead.
- Serverless SQL Pool is ideal for ad-hoc, on-demand querying, and data exploration without the need to manage resources. Cost-effective for intermittent or unpredictable workloads.

Lab 2: Creating an Azure Synapse Workspace

- The following is what we have done so far



Lab 2: Creating an Azure Synapse Workspace

- Delete everything we created.

License