

NAME : AKULA SHARATH CHANDRA

BATCH: DATA ENGINEERING

DATE :15-02-2024

**TOPICS: DATABASES VS DATA WAREHOUSES VS
DATA LAKES, CREATING A STORAGE ACCOUNT,
CREATING A CONTAINER , GENERATING A SAS AND
CONNECTING STRING,CONNECTING TO AZURE
RESOURCE**

1) DATABASES: A database is a collection of data or information. Databases are typically accessed electronically and are used to support Online Transaction Processing (OLTP). Database Management Systems (DBMS) store data in the database and enable users and applications to interact with the data. The term "database" is commonly used to reference both the database itself as well as the DBMS.

DATABASE CHARACTERISTICS:

All databases store information, but each database will have its own characteristics, Relational databases store data in tables with fixed rows and columns. Non- relational databases (also known as NoSQL databases) store data in a variety of models including JSON (JavaScript Object Notation), BSON (Binary JSON), key- value pairs, tables with rows and dynamic columns, and nodes and edges.

Databases store structured and/or semi-structured data, depending on the type.

→You may also find database characteristics like:

- i)Security features to ensure the data can only be accessed by authorised users. ACID (Atomicity, Consistency, Isolation, Durability) transactions to ensure data
- ii)Integrity
- iii)Query languages and APIs to easily interact with the data in the database.
- iv)Indexes to optimise query performance.
- v)Full-text search.
- vi)Optimizations for mobile devices.

vii)Flexible deployment topologies to isolate workloads (e.g., analytics workloads) to a specific set of resources.
vii)On-premises, private cloud, public cloud, hybrid cloud, and/or multi-cloud hosting options.

WHY USE A DATABASE?

If your application needs to store data (and nearly every interactive application does), your application needs a database. Applications across industries and use cases are built on databases. Many types of data can be stored in databases, Including:

- i)Patient medical records
- ii)Items in an online store
- iii)Financial records
- iv)Articles and blog entries
- v)Sports scores and statistics
- vi)Online gaming information.
- vii)Student grades and scores
- viii)IoT device readings
- ix)Mobile application information

DATABASE EXAMPLES:

—>A myriad of databases exist.

Examples include:

- >Relational databases:Oracle, MYSQL. Microsoft SQL Server, and PostgreSQL.
- >Document databases: MongoDB and CouchDB
- >Key-value databases: Redis and DynamoDB
- >Wide-column stores: Cassandra and HBase
- > Graph databases: Amazon Neptune,Neo4j

OLAP + DATA WAREHOUSES AND DATA LAKES:

OLAP systems are typically used to collect data from a variety of sources. The data is then used to power a range of analytical use cases ranging from business intelligence and reporting (eg. quarterly sales reports by store) to forecasting (eg, predicting home sales for the next six months based on historical trends)

2) DATA WAREHOUSE: A data warehouse is a system that stores highly structured information from various sources. Data warehouses typically store current and historical data from one or more systems. The goal of using a data warehouse is to combine disparate data sources in order to analyze the data, look for insights, and create business intelligence (BI) in the form of reports and dashboards.

Is a data warehouse a database?

Yes, a data warehouse is a giant database that is optimised for analytics.

DATA WAREHOUSE CHARACTERISTICS:

—>Data warehouses contain a range of data, from raw ingested data to highly curated, cleansed, filtered, and aggregated data.

—>Extract, transform, load (ETL) processes move data from its original source to the data warehouse.

—>Data warehouses typically have a pre-defined and fixed relational schema. Therefore, they work well with structured data.

—>Once the data is in the warehouse, business analysts can connect data warehouses with BI tools. These tools allow business analysts and data scientists to explore the data, look for insights, and generate reports for business stakeholders.

WHY USE A DATA WAREHOUSE?

When you need to store large amounts of historical data and/or perform in-depth analysis of your data to generate business intelligence. Due to their highly structured nature, analysing the data in data warehouses is relatively straightforward and can be performed by business analysts and data scientists.

DATA WAREHOUSE EXAMPLES:

Examples of data warehouses include:

- >Amazon Redshift.
- >Google BigQuery.
- >IBM Db2 Warehouse.
- >Microsoft Azure Synapse.
- >Oracle Autonomous Data Warehouse.
- >Snowflake.
- >Teradata Vantage.

3)WHAT IS A DATA LAKE?

A data lake is a repository of data from disparate sources that is stored in its original, raw format. Like data warehouses, data lakes store large amounts of current and historical data. What sets data lakes apart is their ability to store data in a variety of formats including JSON, BSON, CSV, TSV, Avro, ORC, and Parquet.

The primary purpose of a data lake is to analyse the data to gain insights. However, organisations sometimes use data lakes simply for their cheap storage with the idea that the data may be used for analytics in the future.

IS A DATA LAKE A DATABASE?

A data lake is a repository for data stored in a variety of ways including databases. With modern tools and technologies, a data lake can also form the storage layer of a database. Tools like Starburst, Presto, Dremio, and Atlas Data Lake can give a database-like view into the data stored in your data lake. In many cases, these tools can power the same analytical workloads as a data warehouse.

DATA LAKE CHARACTERISTICS:

- >Data lakes store large amounts of structured, semi-structured, and unstructured data. They can contain everything from relational data to JSON documents In PDFs to audio files.
- >Data does not need to be transformed in order to be added to the data lake, which means data can be added (or "ingested") incredibly efficiently without upfront planning.
- >The primary users of a data lake can vary based on the structure of the data. Business analysts will be able to gain insights when the data is more structured. When the data is more unstructured, data analysis will likely require the expertise of developers, data scientists, or data engineers.
- >The flexible nature of data lakes enables business analysts and data scientists to look for unexpected patterns and insights. The raw nature of the data combined with its volume allows users to solve problems they may not have been aware of when they initially configured the data lake.
- >Data in data lakes can be processed with a variety of OLAP systems and visualised with BI tools.

WHAT ARE THE KEY DIFFERENCES BETWEEN A DATABASE, DATA WAREHOUSE, AND DATA LAKE?

Feature	Database	Data Lake	Data Warehouse
Workloads	Operational and transactional	Analytical	Analytical
Data Type	Structured or semi-structured	Structured, semi-structured, and/or unstructured	Structured and/or semi-structured
Schema Flexibility	Rigid or flexible	No schema	Pre-defined and fixed
Data Freshness	Real time	May not be up-to-date	May not be up-to-date
Users	Application developers	Business analysts, data scientists	Business analysts, data scientists
Pros	Fast queries, good for storing and updating data	Easy data storage	Easy to analyze data
Cons	Limited analytics, requires data preparation	Requires effort to organize data	Difficult to design schema, not flexible

CREATING A STORAGE ACCOUNT :

- >To create a storage account in Microsoft Azure, start by signing in to the Azure Portal using azure account.
- >In the All services search box, type "Storage Accounts" and select it from the results.
- >In the "Storage Accounts" blade, click on the "Add" button to start the process of creating a new storage account.
- >After providing the resource group,storage account name,region,account kind we need to click the "Review + create" button.
- >After reviewing the summary, click the "Create" button to start the deployment process.

CREATING A CONTAINER IN IT:

- >In the "All services" search box, type "Storage Accounts" and select it from the results.Click on the relevant storage account for which you want to create a container.
- >Click on the "+ Container" button to create a new container.Provide a unique name for the container in the "Name" field. Container names must be lowercase and can include only letters, numbers, and hyphens.Choose the access level for the container (e.g., Private, Blob, Container, or Blob + Container).
- >Review the configuration settings.Click the "Create" button to create the new container.
- >Once the container is created, you can access it in the "Containers" section under the Blob service.
- >Once you create the container we can upload the files in it .

Microsoft Azure

Home > hexadeb1069_1707992455063 | Overview > hexadeb1069

hexadeb1069 | Containers

Storage account

Search

Events

Storage browser

Storage Mover

Data storage

Containers

File shares

Queues

Tables

Security + networking

Networking

Front Door and CDN

Access keys

Shared access signature

Encryption

Microsoft Defender for Cloud

Search containers by prefix

Show deleted containers

Name	Last modified	Anonymous access level	Lease state
<input type="checkbox"/> \$logs	15/02/2024, 15:51:26	Private	Available
<input type="checkbox"/> mycontainer	15/02/2024, 16:31:31	Private	Available

31°C Sunny

Search

ENG US

4:37 PM 2024-02-15

GENERATING A SAS AND CONNECTING STRING:

—>To generate a sas and connecting string Firstly, Look for an option or section related to "Shared Access Signature" or "SAS" in the Azure Portal. This is where you'll configure the settings for the SAS.

—>Then allow the required permissions of resource type (service,container,object)

—> Then give an ip address and generate a sas and connecting string.

—> Then the Blob service sas url will be generated.

portal.azure.com/#@iihtl.onmicrosoft.com/resource/subscriptions/984f097c-963c-4eb6-a20d-839457ae9f08/resourcegroups/rg-azuser...

Microsoft Azure

Search resources, services, and docs (G+/)

azuser1069_mml.local@...
IIHT (IIHTLONMICROSOFT.COM)

Home > hexadeb1069_1707992455063 | Overview > hexadeb1069

hexadeb1069 | Shared access signature ☆ ...

Storage account

Search

Give feedback

Data storage

Containers

File shares

Queues

Tables

Security + networking

Networking

Front Door and CDN

Access keys

Shared access signature

Encryption

Microsoft Defender for Cloud

Data management

Redundancy

Data protection

A shared access signature (SAS) is a URI that grants restricted access rights to Azure Storage resources. You can provide a shared access signature to clients who should not be trusted with your storage account key but whom you wish to delegate access to certain storage account resources. By distributing a shared access signature URI to these clients, you grant them access to a resource for a specified period of time.

An account-level SAS can delegate access to multiple storage services (i.e. blob, file, queue, table). Note that stored access policies are currently not supported for an account-level SAS.

[Learn more about creating an account SAS](#)

Allowed services ⓘ
☒ Blob ☒ File ☒ Queue ☒ Table

Allowed resource types ⓘ
☒ Service ☒ Container ☒ Object

Allowed permissions ⓘ
☒ Read ☒ Write ☒ Delete ☒ List ☒ Add ☒ Create ☒ Update ☒ Process ☒ Immutable storage ☒ Permanent delete

Blob versioning permissions ⓘ
☒ Enables deletion of versions

Allowed blob index permissions ⓘ
☒ Read/Write ☒ Filter

31°C Sunny

Search

4:06 PM 2024-02-15

portal.azure.com/#@iihtl.onmicrosoft.com/resource/subscriptions/984f097c-963c-4eb6-a20d-839457ae9f08/resourcegroups/rg-azuser...

Microsoft Azure

Search resources, services, and docs (G+/)

azuser1069_mml.local@...
IIHT (IIHTLONMICROSOFT.COM)

Home > hexadeb1069_1707992455063 | Overview > hexadeb1069

hexadeb1069 | Shared access signature ☆ ...

Storage account

Search

Give feedback

Data storage

Containers

File shares

Queues

Tables

Security + networking

Networking

Front Door and CDN

Access keys

Shared access signature

Encryption

Microsoft Defender for Cloud

Data management

Redundancy

Data protection

Start and expiry date/time ⓘ

Start 15/02/2024 16:01:04

End 16/02/2024 00:01:04

(UTC+05:30) Chennai, Kolkata, Mumbai, New Delhi

Allowed IP addresses ⓘ
For example, 168.1.5.65 or 168.1.5.65-168.1.5.70

Allowed protocols ⓘ
☒ HTTPS only ☐ HTTPS and HTTP

Preferred routing tier ⓘ
☒ Basic (default) ☐ Microsoft network routing ☐ Internet routing

Some routing options are disabled because the endpoints are not published.

Signing key ⓘ
key1

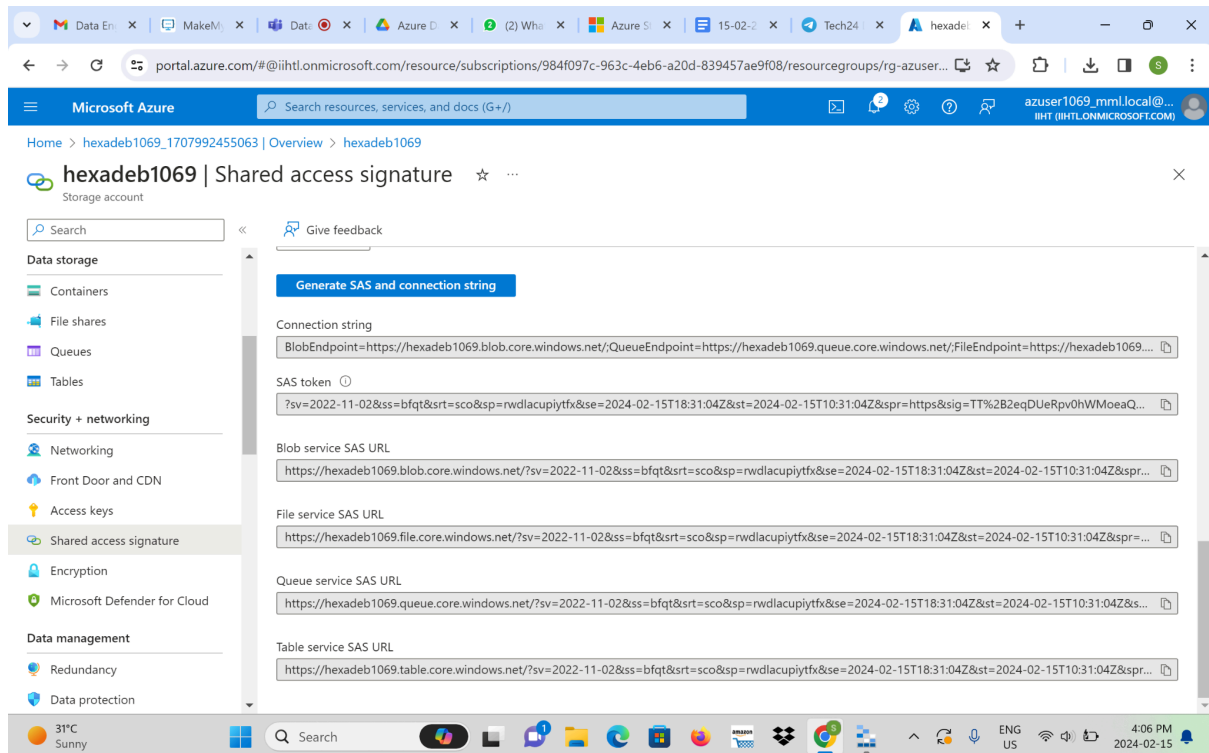
Generate SAS and connection string

Connection string
BlobEndpoint=https://hexadeb1069.blob.core.windows.net/QueueEndpoint=https://hexadeb1069.queue.core.windows.net/FileEndpoint=https://hexadeb1069...

31°C Sunny

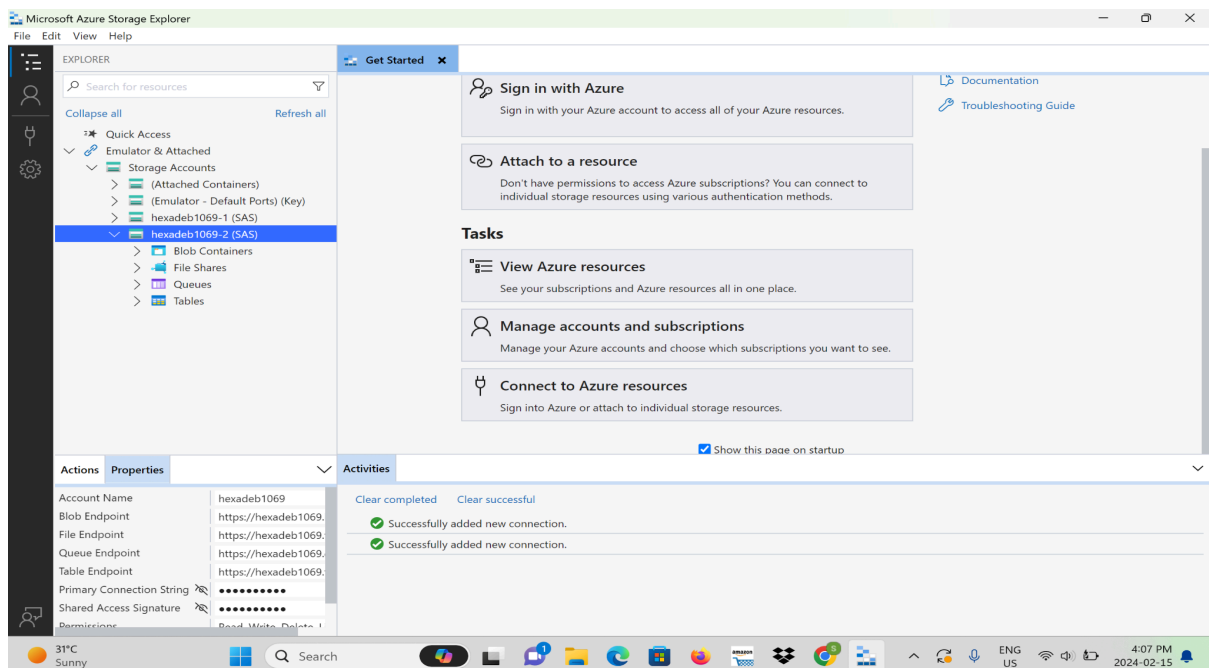
Search

4:06 PM 2024-02-15

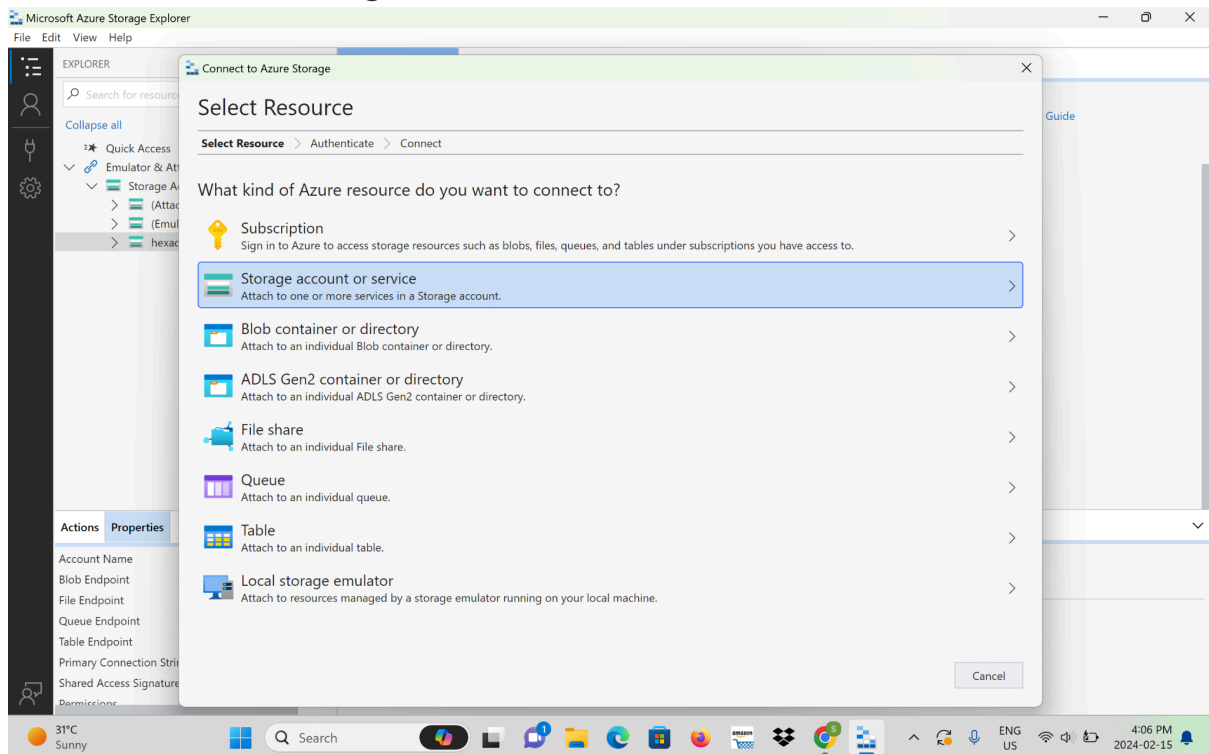


CONNECTING TO AZURE RESOURCE:

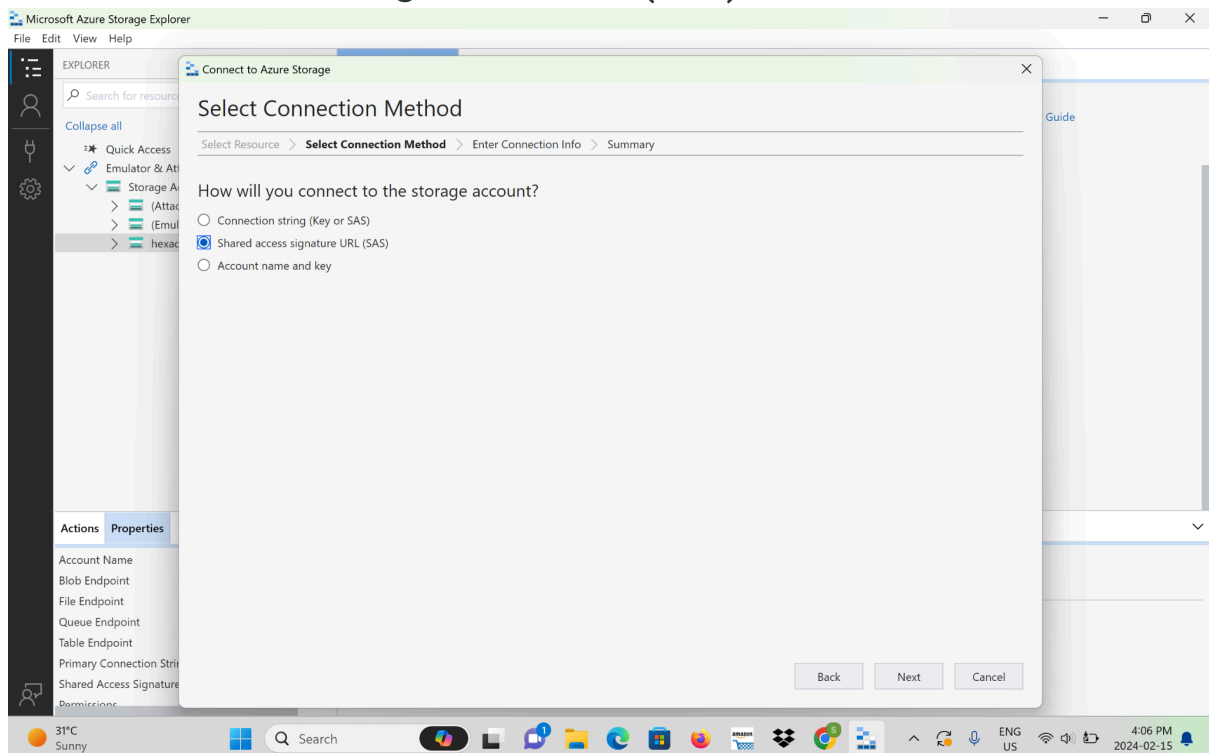
- >For connecting to azure resource Firstly,we need to install Microsoft Azure storage explorer.
- >Then we need to set up the Microsoft Azure Storage Explorer.
- >Then we need to click on connect to azure resources.



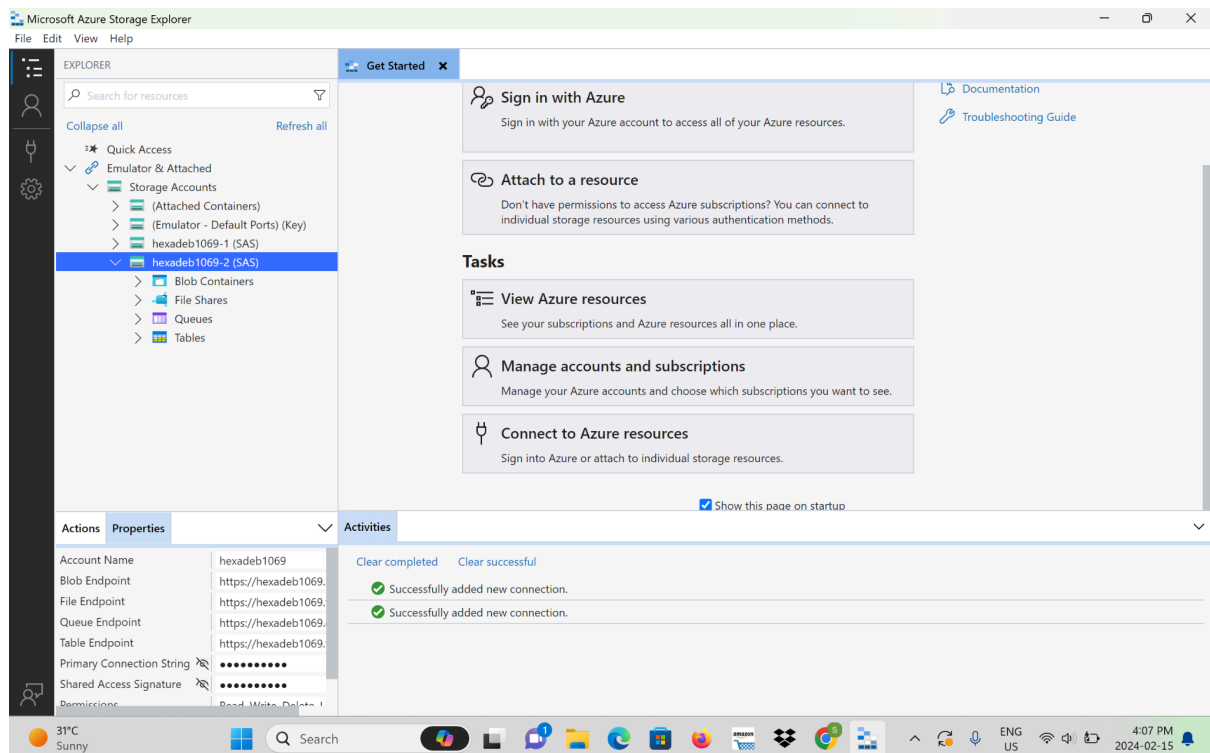
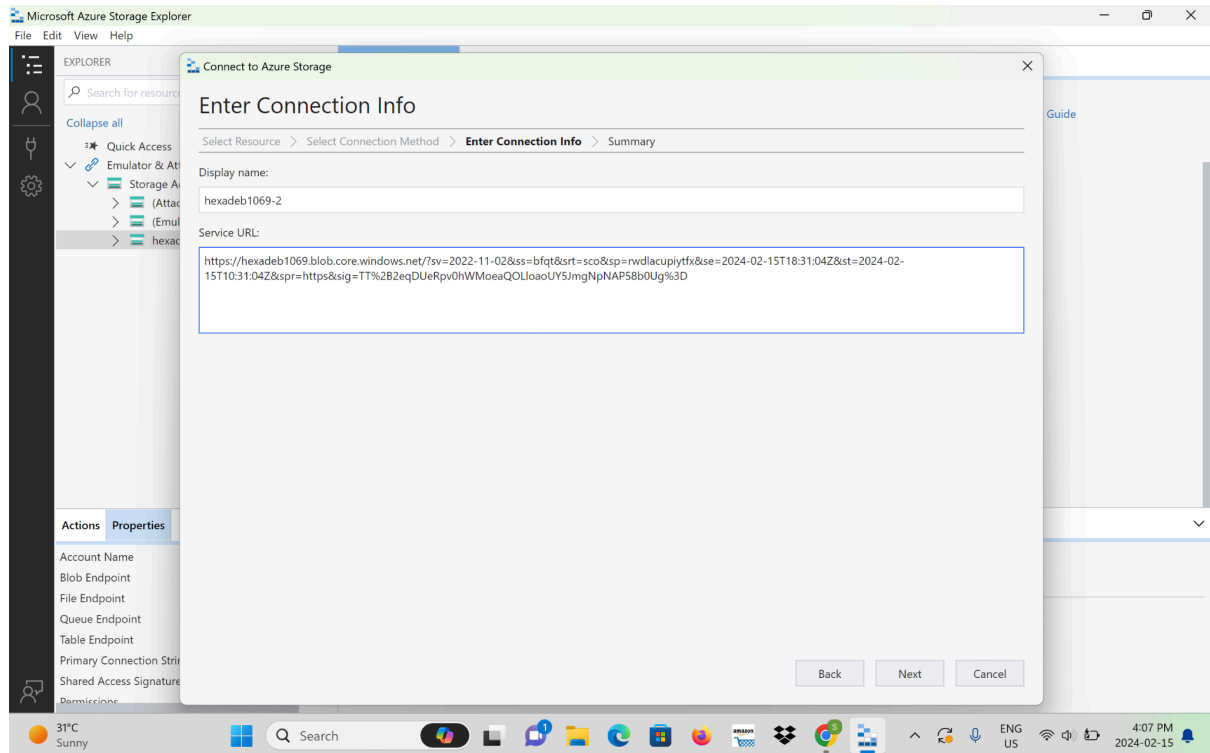
—>After that we need to select the resource then we need to click on a storage account or services .



—> Then we need to click on the Select connection method to shared access signature URL(sas).



—>Then we need enter the connection information (i.e the link which had copy the Blob service sas url and paste it.)and click on next and then we will get to see successfully added a new connection.



—>We can see the files which are in the container after adding the connection

The screenshot displays the Microsoft Azure Storage Explorer application. The left sidebar shows the Explorer view with a tree structure of storage resources. The main pane shows the 'mycontainer' blob container, which contains a single file named 'employee_data.csv'. The file's properties are listed in the table below:

Name	Access Tier	Access Tier Last Modified	Last Modified
employee_data.csv	Hot (inferred)		2024-02-15 4:32 PM

The bottom pane shows the Properties tab for the selected file, displaying various metadata fields such as URL, Custom Domain, Type, HNS Enabled, Lease State, Lease Status, Public Read Access, and Last Modified. The Activities pane on the right shows a list of successful operations, including 'Successfully added new connection.'.