Data Engineering in the Cloud

Azure Storage

Xuemao Zhang
East Stroudsburg University

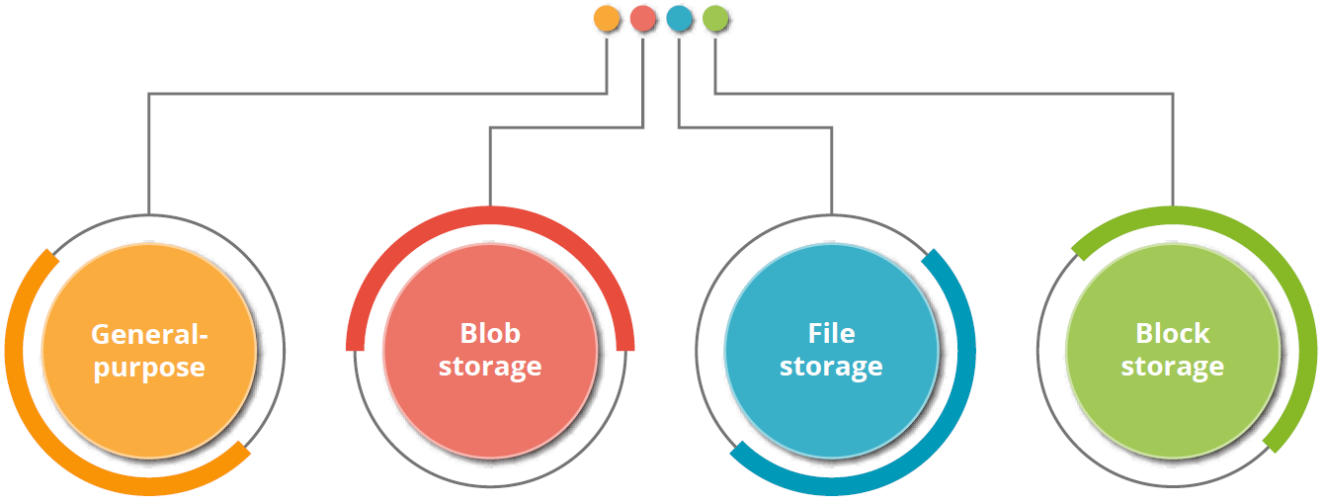
January 18, 2025

Outline

- Types of Azure Storage Account
- Types of Access Tiers
- Azure Storage Replication
- Lab: Create a Storage Account

Azure Storage Account

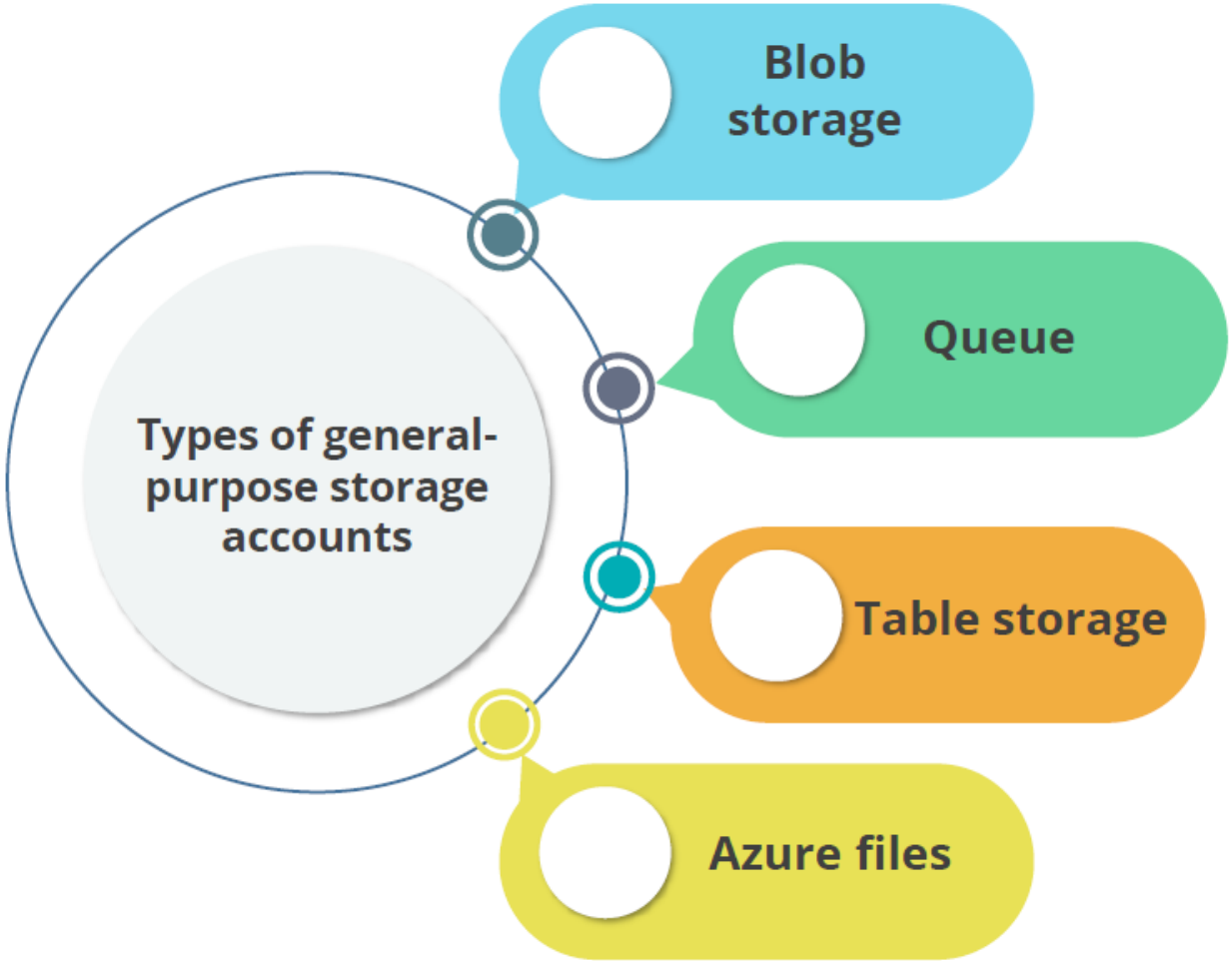
- Azure Storage is a cloud storage service designed to meet the demands of applications that require high availability, scalability, and durability.
- Types of Storage Accounts



General Purpose Storage Account

- It provides access to blobs, queues, files, and tables in one single account.
- It can store object data, act as a NoSQL data store, and develop and implement message processing queues.
- It aids in the establishment of cloud based file sharing.

General Purpose Storage Account



General Purpose Storage Account - Blob Storage

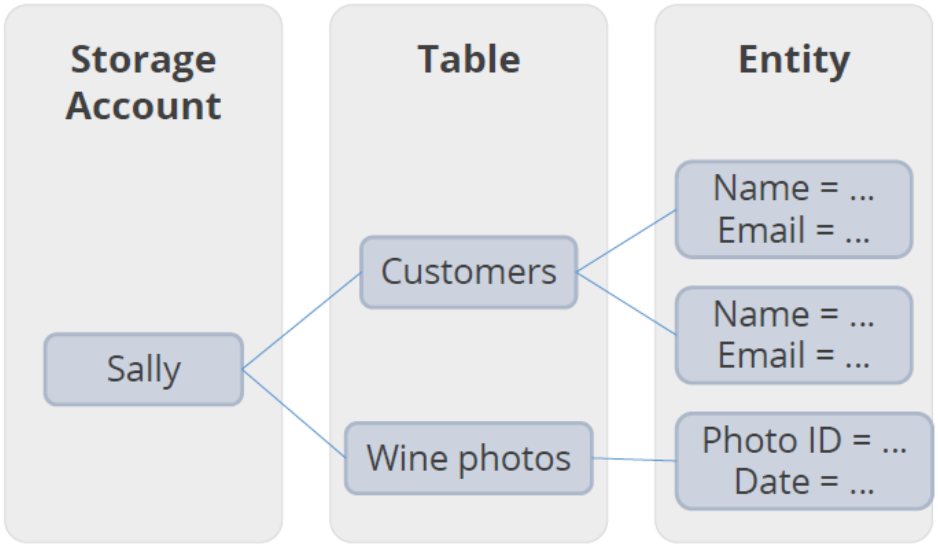
- Blobs (Binary Large Objects) are the units of storage in Azure Blob Storage. They can store any type of text or binary data.
- Blob storage stores blobs of unstructured data such as media, metadata, and more.
- Blobs are organized into containers, which are like directory structures.
  - Containers are logical containers for blobs, similar to directories in a file system.
  - Containers help organize blobs and provide a security boundary. All blobs must reside in a container.

General Purpose Storage Account - Queue Storage

- Azure Queue Storage is a service that provides cloud messaging between application components. It is designed for storing large volumes of messages and ensuring that they can be accessed in a reliable and efficient manner.
- Queues are used to store a large number of messages. Messages in the queue are processed in a First-In-First-Out (FIFO) order, ensuring orderly message handling.
- Non-Ordered Information: Although queues are FIFO, they don't impose strict ordering on messages, making them flexible for various use cases.
- Azure Queue storage enables customers to store large volumes of messages and then consume the data as required.
  - Each message can be up to 64 KB in size, and a queue can hold millions of messages.
- Since Azure Queue storage is based on a pay as you go model, it is cost efficient.

General Purpose Storage Account - Table Storage

- Table storage is a scalable technique for storing large amounts of data in the cloud that involves sorting rows as a key value pair in each table.
- Tables are ideal for holding both structured and non relational data.



General Purpose Storage Account - Azure Files

- Azure File Storage provides fully managed **file shares** in the cloud that are accessible via the Server Message Block (SMB) and Network File System (NFS) protocols. It is designed to enable easy file sharing and storage for various applications and services.
  - File shares are the core units of storage within Azure File Storage. They can be mounted concurrently by cloud or on-premises deployments of Windows, macOS, and Linux.
- The Azure File service facilitates the management and access of files for applications running on Azure Virtual Machines (VMs), providing a seamless storage solution.

Blob Storage Account

- Blob storage accounts are storage accounts to store blob data and regulate access.
- It can build a data access tier that specifies how to manage data.
- Block storage is a type of data storage commonly used in cloud computing environments and traditional storage area networks (SANs). It involves storing data in fixed-sized blocks and managing them individually. Each block can be controlled as an individual hard drive and accessed independently, allowing for high performance and flexibility.

Types of Access Tiers

Hot Access Tier

- The Hot Access Tier in Azure Blob Storage is designed for data that is accessed frequently. This tier offers the lowest latency and highest availability, making it ideal for workloads that require quick and frequent access to data.
  - It is tailored for scenarios where performance and accessibility are paramount.
- It provide the shortest feasible latency.
- It should be used with frequently accessed data.
- It is expensive owing to the feasible latency it provides.

Cold Access Tier

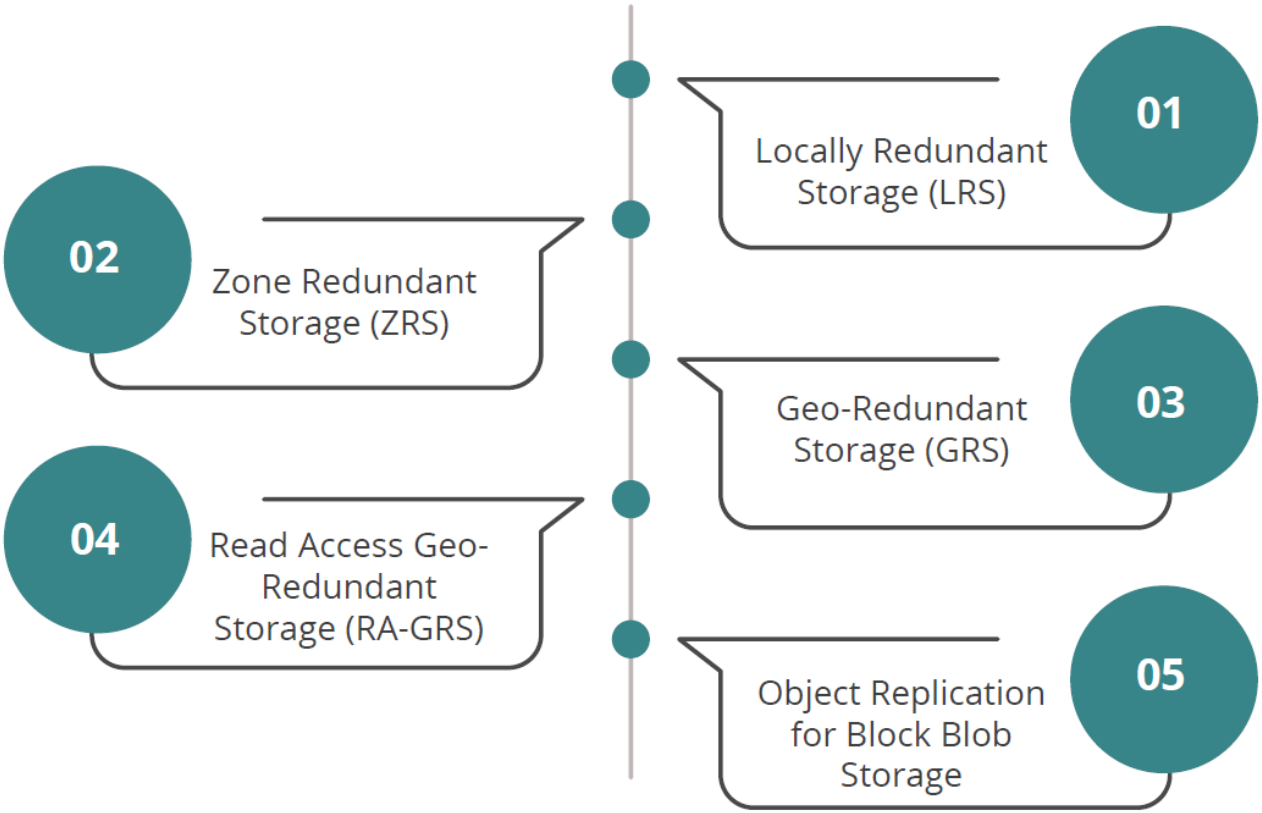
- The Cold Access Tier in Azure Blob Storage, often referred to as the Cool tier, is designed for data that is infrequently accessed. It offers a balance between performance and cost, providing a cost-effective solution for storing data that does not require the low latency and high availability of the Hot tier.
- It is less efficient as compared to the hot access tier.
- It is reserved for less accessed data.
- It is cost effective as compared to the hot access tier.

Azure Storage Replication

- Storage replication is a feature of Azure storage that creates and stores copies of data across several locations to ensure data availability and durability.
- It aims to offer redundancy for data protection against hardware failures, power outages, or network outages.
- There are several redundancy options available. The data can be replicated within a single region or into a secondary region that is geographically distant.
  - Read access to replicated data in a secondary region can be enabled to ensure availability in the event of a disaster.

Types of Azure Storage Replication

- storage-redundancy



Locally Redundant Storage (LRS)

- LRS is a high availability replication mechanism that replicates data in real time to three discs, that is, three times in the data center.
- It ensures at least 99.999999999% (11 nines) durability of objects over a given year.
- It is cost effective and durable. It provides data security by storing data in the case of a server rack.

Zone Redundant Storage

- The ZRS replication technique replicates data across three availability zones in the primary region, each of which is a separate physical location.
- ZRS offers durability for storage resources of at least 99.9999999999% (12 9's) over a given year.
- It is expensive as compared to Locally Redundant Storage.
- It provides high resiliency.

Geo Redundant Storage

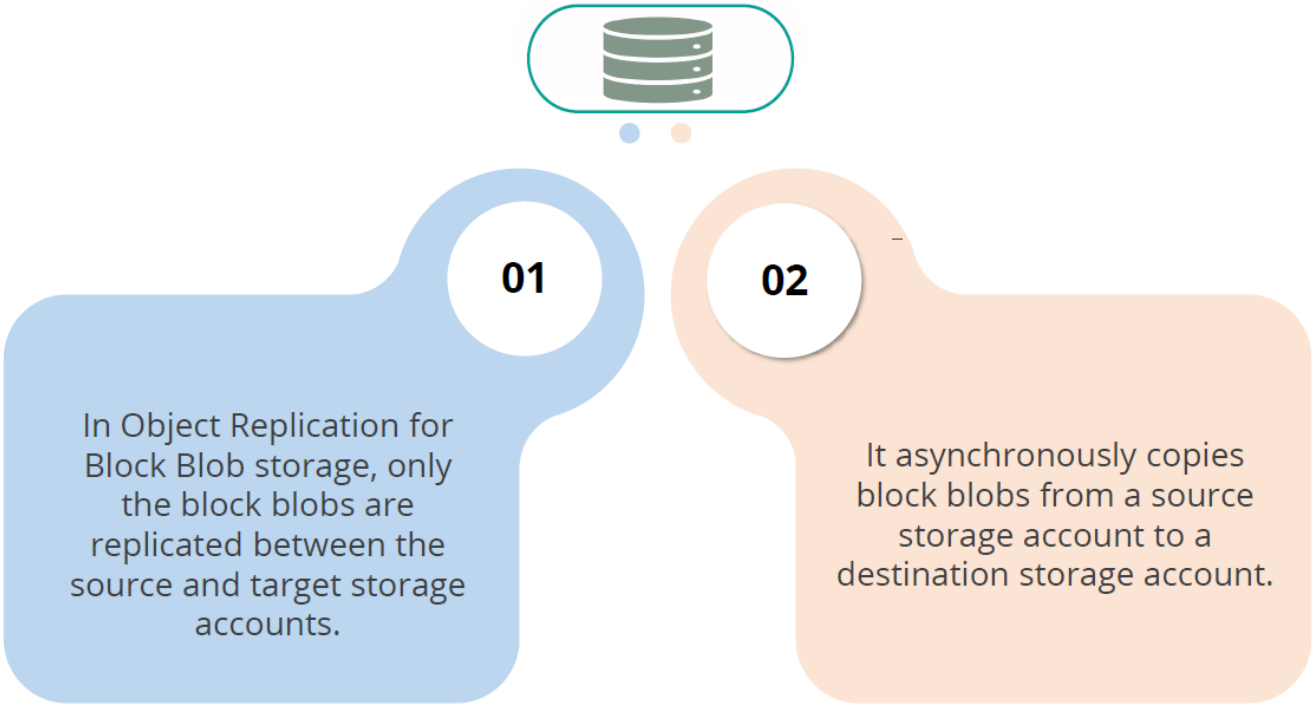
- In GRS, the data is replicated asynchronously three times in a paired Azure region, that is, in the primary region.
- The secondary region is hundreds of kilometers away from the first region.
- GRS offers durability for storage resources of at least 99.9999999999999% (16 9's) over a given year.

Read Access Geo Redundant Storage (RA-GRS)

- Unlike GRS, the RA-GRS allows the user to access the data from both the Azure regions.
- It enables users to examine secondary data in the corresponding Azure region.
- If the applications are configured to support this, it implies they have numerous readable endpoints to utilize.

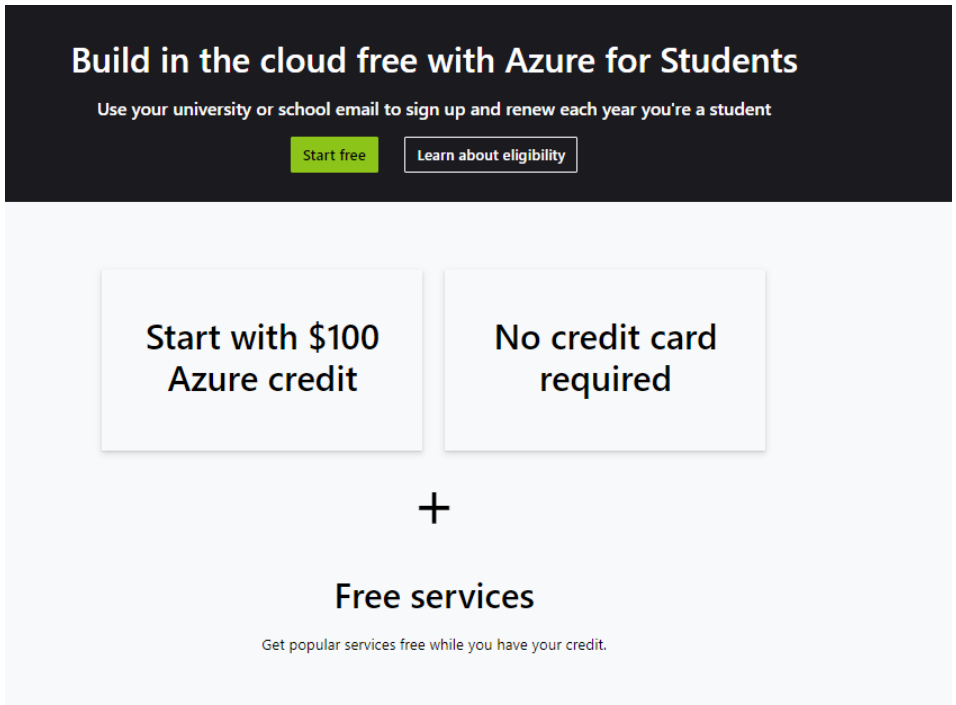
Object Replication for Block Blob Storage

- Object Replication for Block Blob Storage is a feature provided by Azure Blob Storage that allows asynchronous replication of block blobs from a source storage account to a destination storage account.



Lab: Create a Storage Account

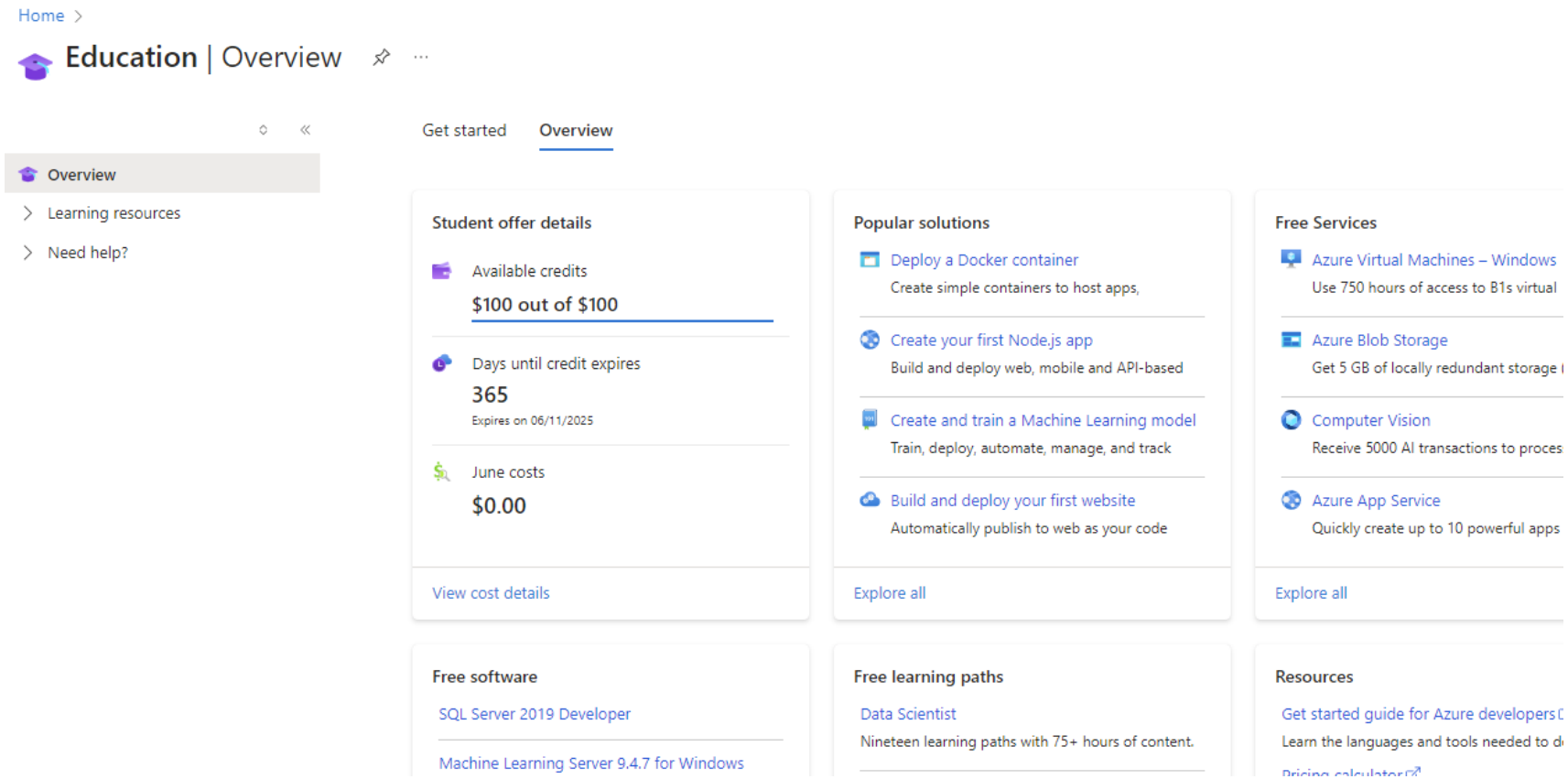
- First, create an **Azure for Students** account using your **esu email** https://azure.microsoft.com/en-us/free/students



- For more information, click **Learn about eligibility** from https://azure.microsoft.com/en-us/free/students
- You can upgrade to a pay-as-you-go subscription - A credit card is needed.

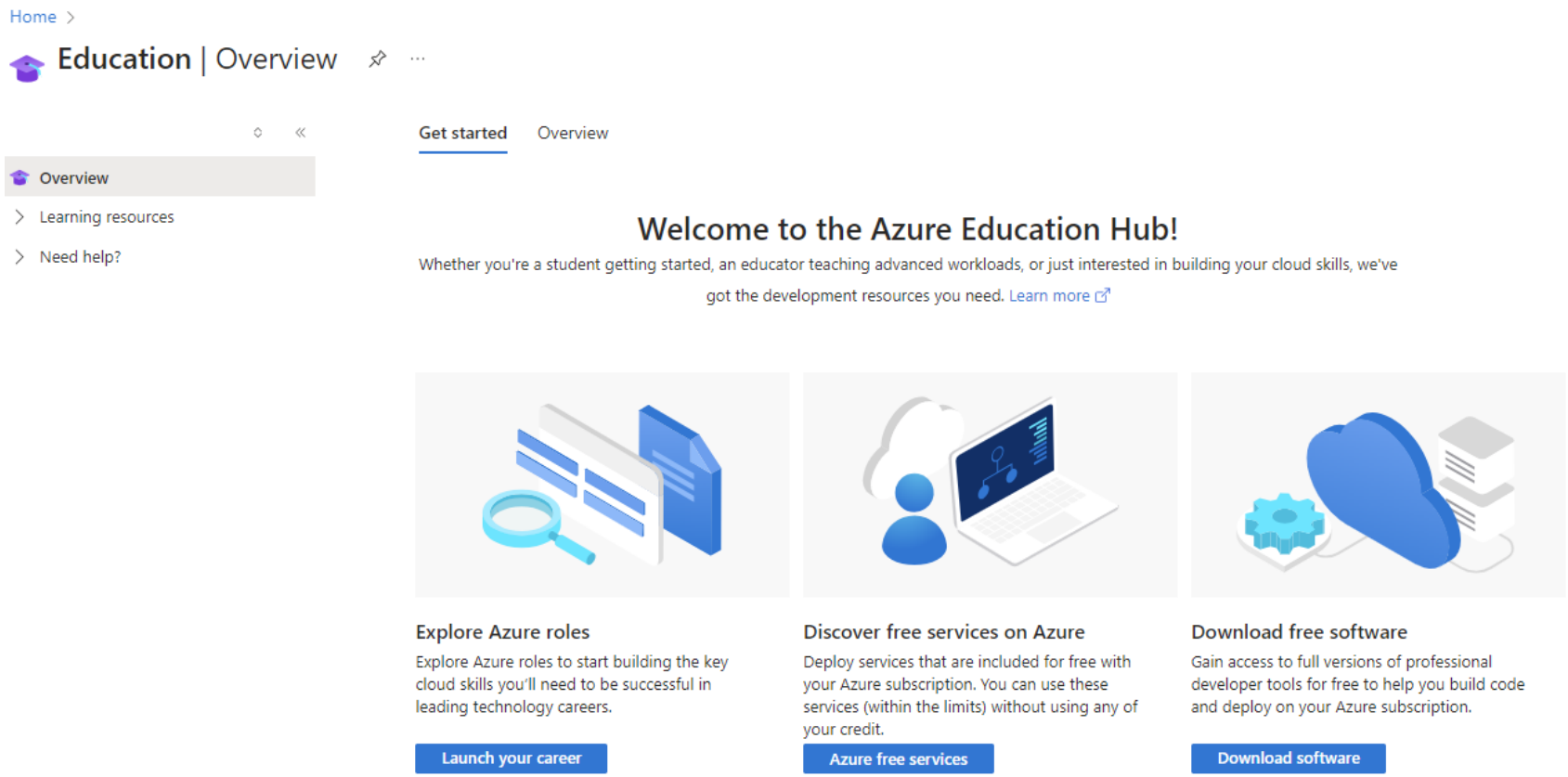
Lab: Create a Storage Account

- You should be able to see a similar screen



Lab: Create a Storage Account

- You should be able to see a similar screen



Lab: Create a Storage Account

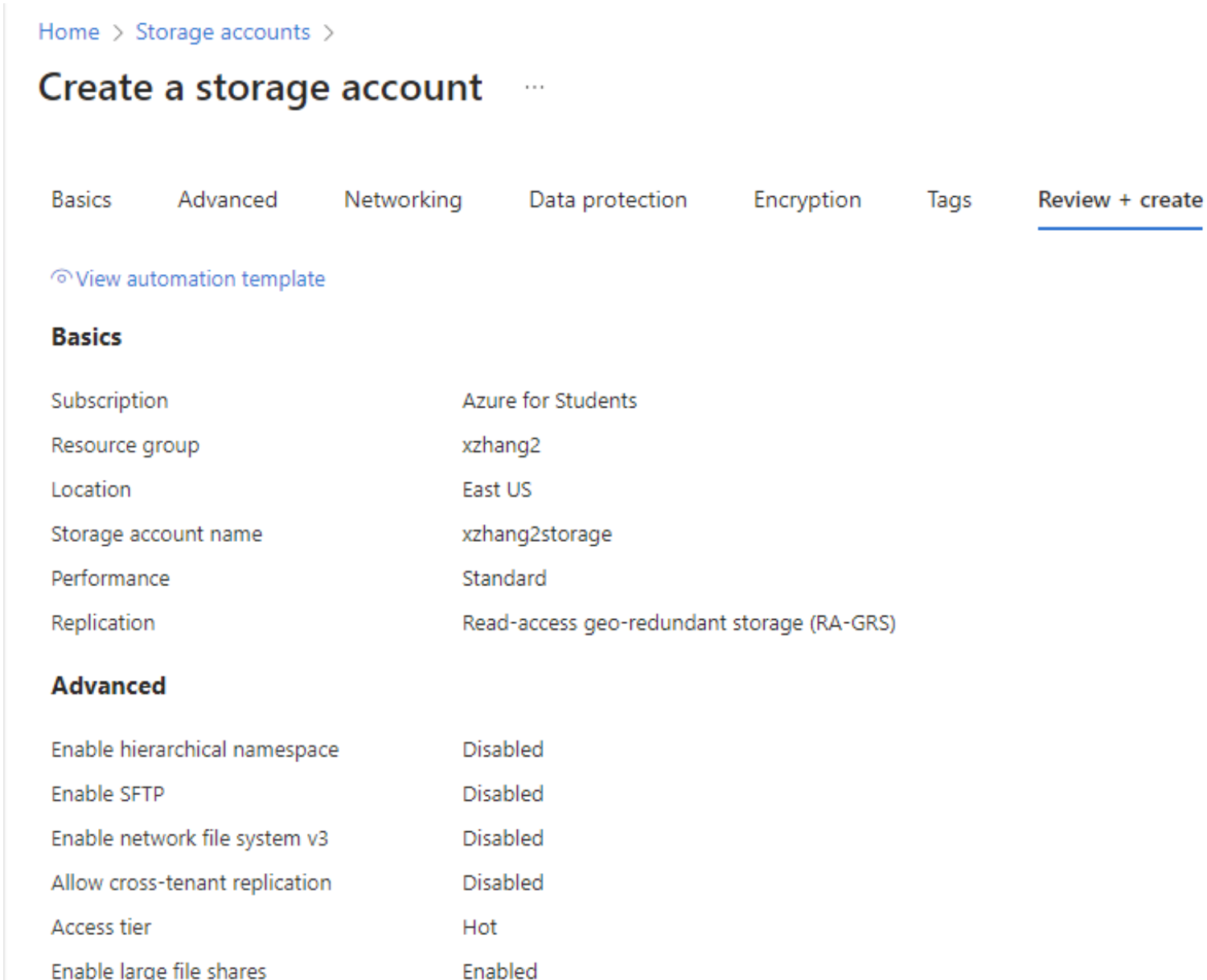
- Search for 'Storage account' or click on 'Azure Blob Storage'
- Basic:
  - Subscription: Azure for students
  - Resource Group: click on create **resource group**.
    - ⋆ Resource group is a logical container that holds related resources for an Azure solution. These resources might include virtual machines (VMs), storage accounts, virtual networks, web apps, databases, and more.
  - Name of the storage account: xzhang2storage
  - Region: East US
  - Performance: Standard
  - Redundany: GRS

Lab: Create a Storage Account

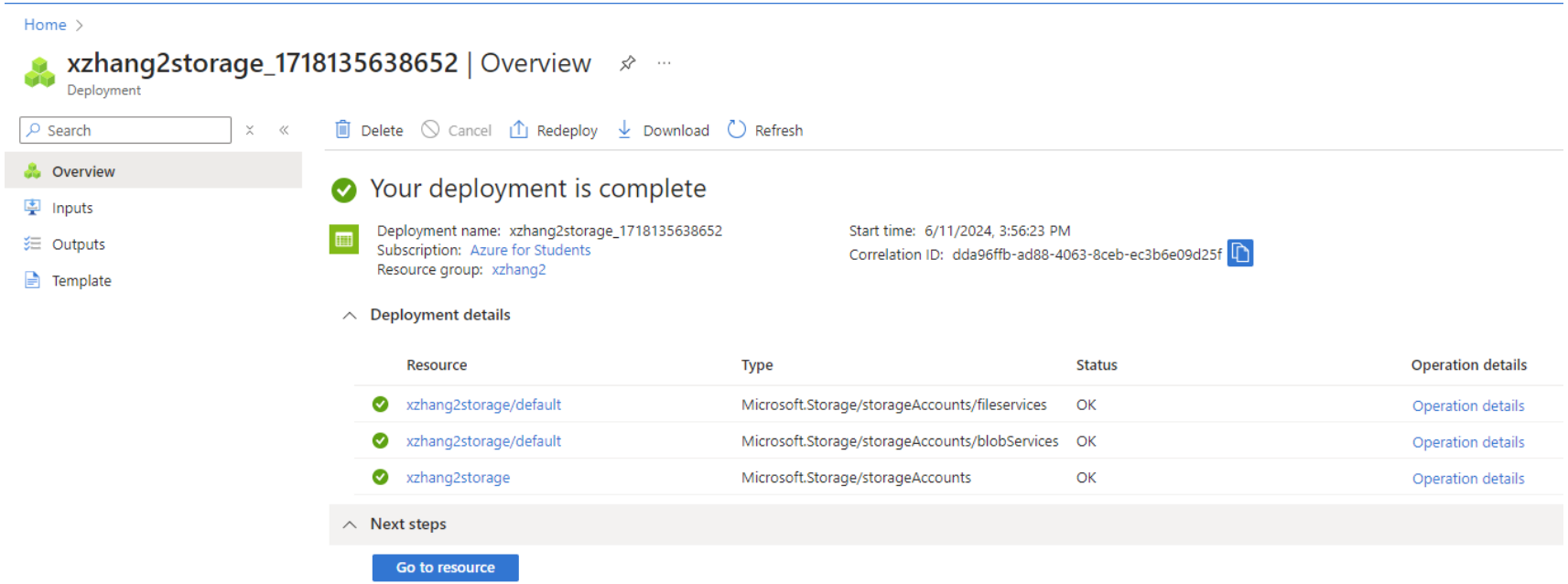
- Advanced: NO changes.
  - If you need to create datalake storage account, need to check `Enable hierarchical namespace` option
- Networking: No changes
- Data protection: No changes
- Encryption: No changes
- Tags: No changes

Lab: Create a Storage Account

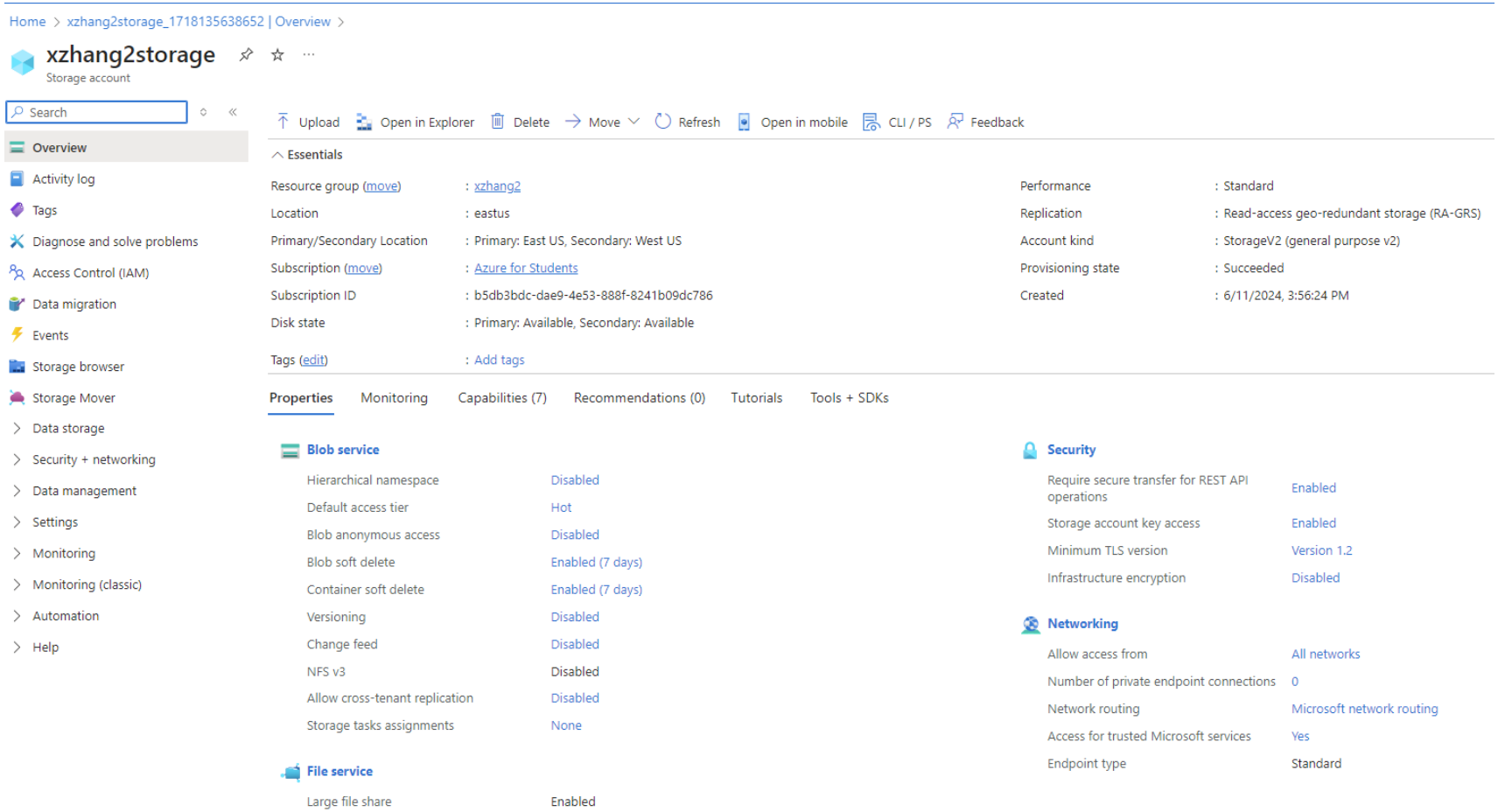
- Review and create it



Lab: Create a Storage Account



Lab: Create a Storage Account



Lab: Create a Storage Account

- Delete the storage account you just created.

License