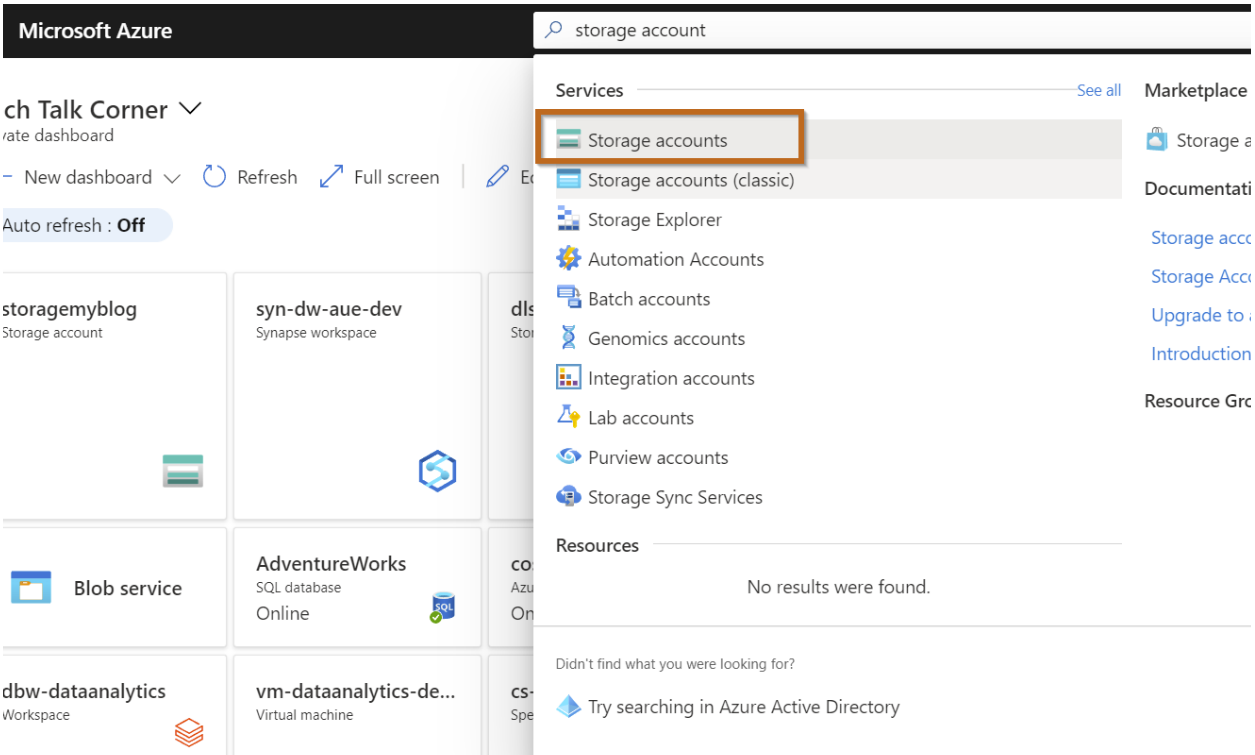
Pre-requisite 1: Create Azure Data Lake

This exercise outlines the steps to follow to create an Azure Data Lake using the Azure portal. An Azure Data Lake is required in Azure Synapse Analytics.

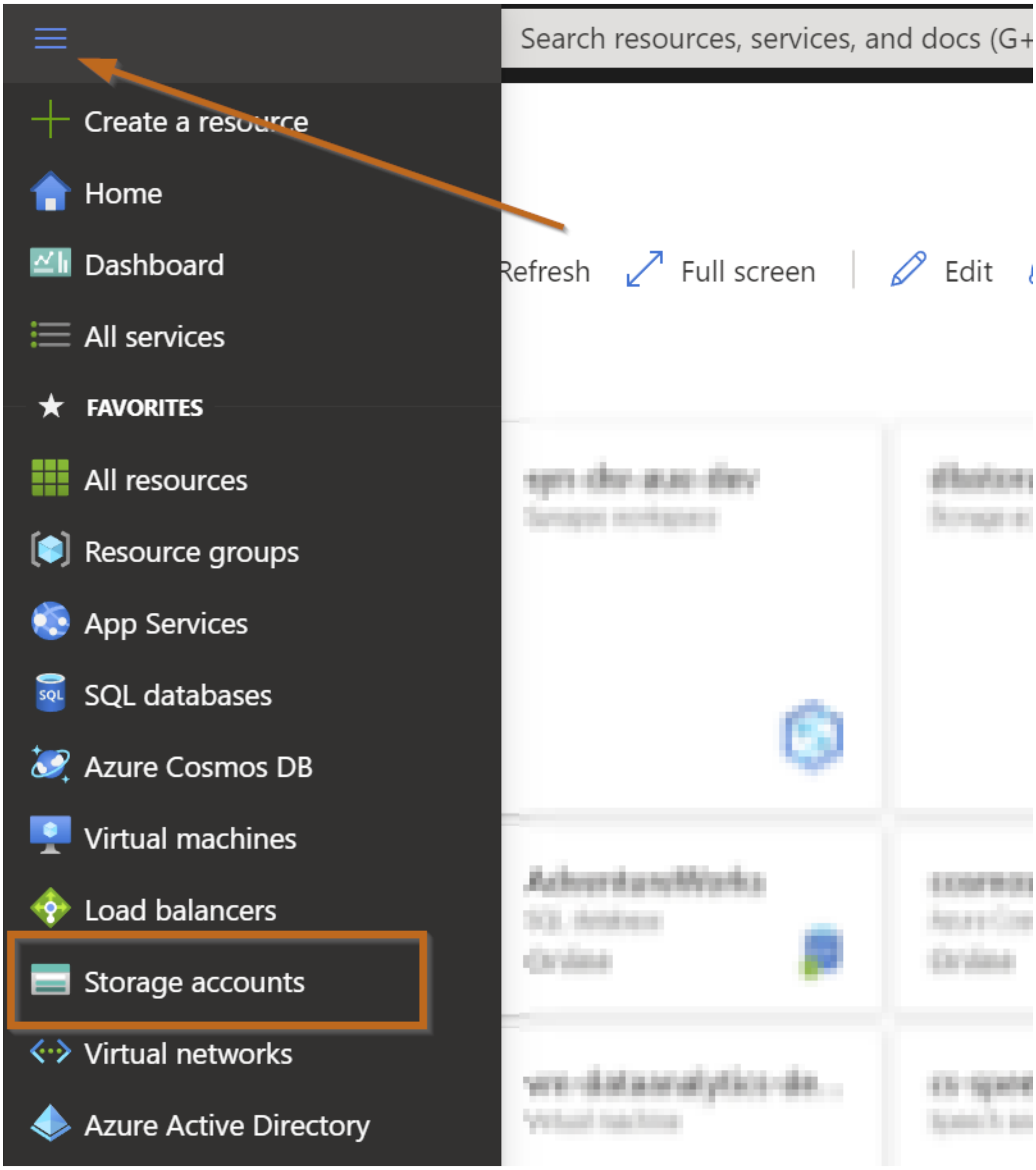
Duration: 10 minutes

Step 1 - Find Azure Data Lake

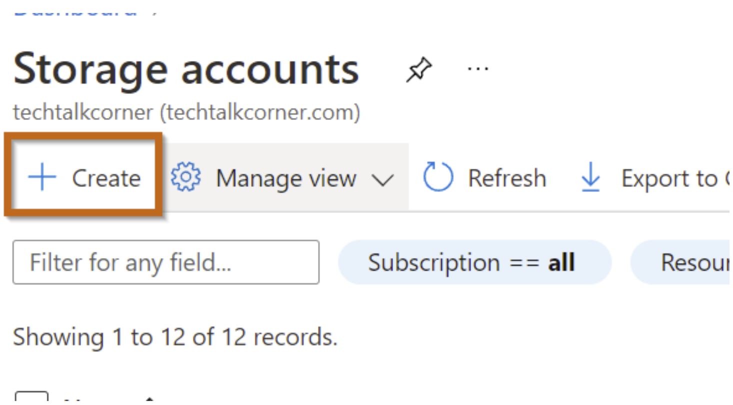
First, find the "Storage accounts" service in the Azure Portal. Azure Data Lake is a configuration of Azure Storage Accounts.



There is also a shortcut on the left navigation pane.



Then select “Create”.


Storage accounts

techtalkcorner (techtalkcorner.com)

Showing 1 to 12 of 12 records.

Finally, you can start configuring the main options to create your Azure Data Lake.

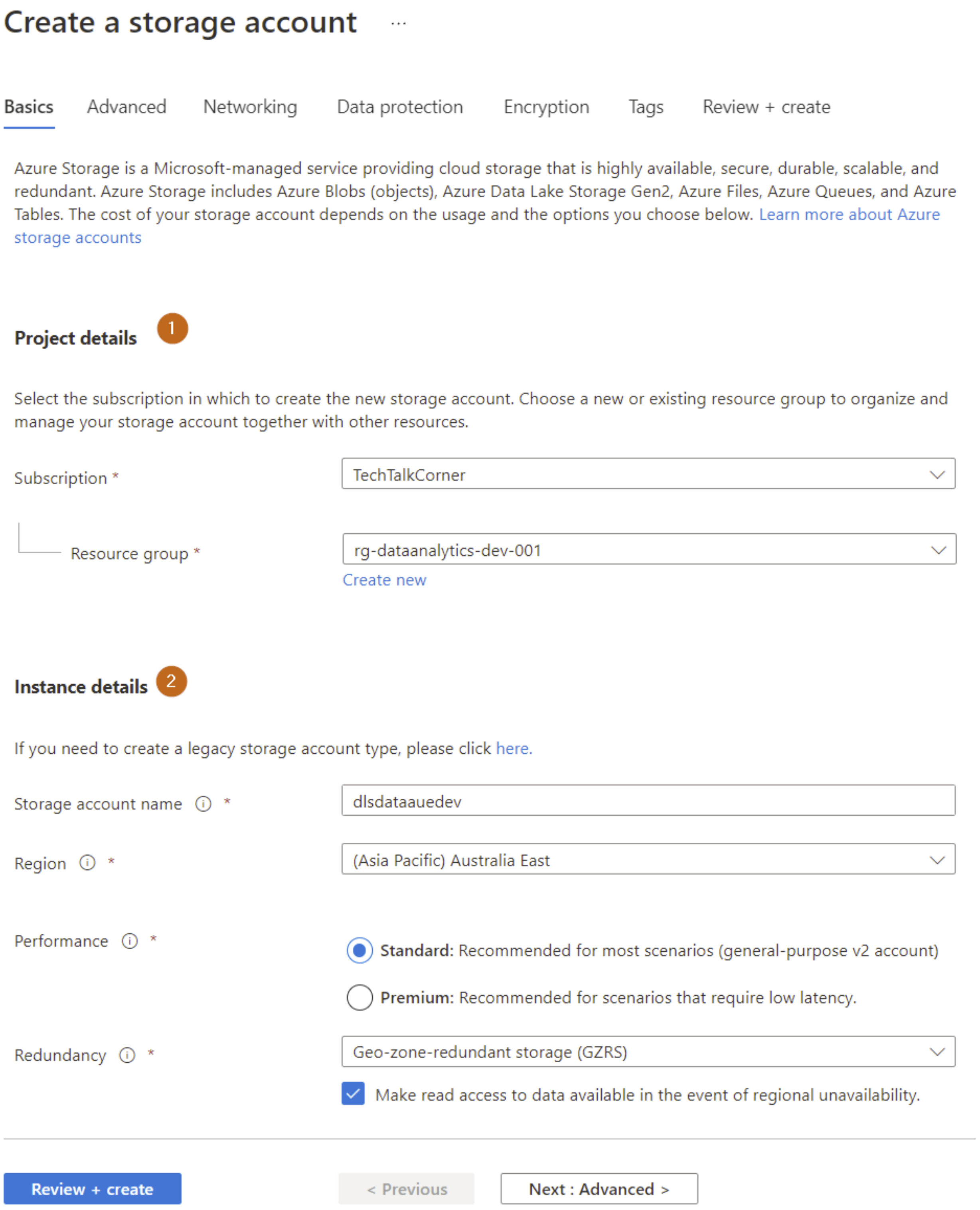
Step 2 - Basic Configuration

1. Project details

Select the subscription and the resource group where you'd like to create the Azure Data Lake.

2. Storage account details

- Storage account name (The field can contain only lowercase letters and numbers. Name must be between 3 and 24 characters). My naming convention:
  - dls (data lake storage) + descriptive short name + 3 letter region + 3 letter environment
- Region, select the region where you would like to host your service.
- Performance and Redundancy



Create a storage account ···

Azure Storage is a Microsoft-managed service providing cloud storage that is highly available, secure, durable, scalable, and redundant. Azure Storage includes Azure Blobs (objects), Azure Data Lake Storage Gen2, Azure Files, Azure Queues, and Azure Tables. The cost of your storage account depends on the usage and the options you choose below. Learn more about Azure storage accounts

Project details 1

Select the subscription in which to create the new storage account. Choose a new or existing resource group to organize and manage your storage account together with other resources.

Subscription *: TechTalkCorner

Resource group *: rg-dataanalytics-dev-001

Create new

Instance details 2

If you need to create a legacy storage account type, please click here.

Storage account name ⓘ *: dlsdataauedev

Region ⓘ *: (Asia Pacific) Australia East

Performance ⓘ *:

● **Standard:** Recommended for most scenarios (general-purpose v2 account)

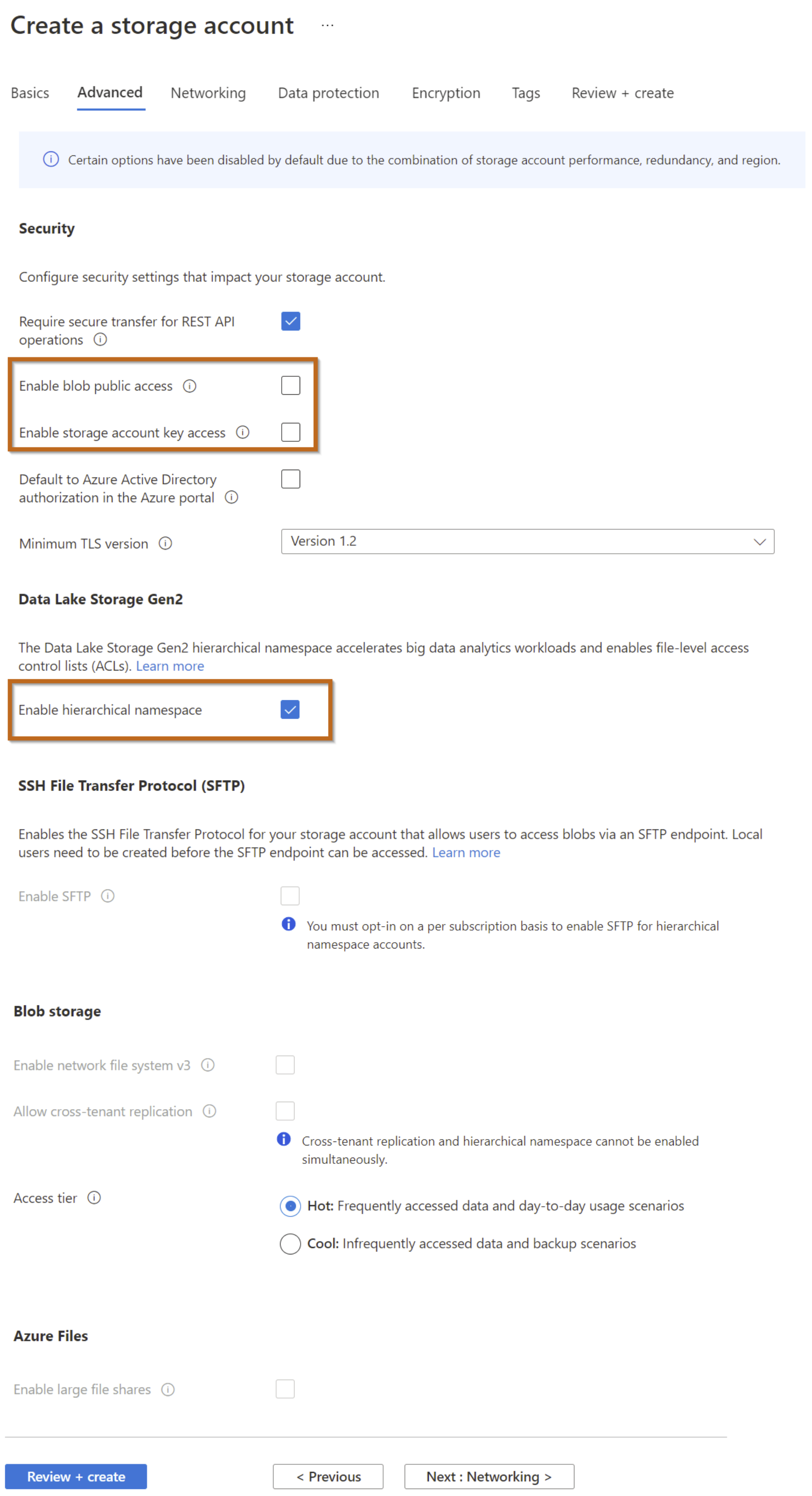
○ **Premium:** Recommended for scenarios that require low latency.

Redundancy ⓘ *: Geo-zone-redundant storage (GZRS)

☑ Make read access to data available in the event of regional unavailability.

Step 3 - Advanced

1. For security reasons, always disable Blob Public Access and Account Key Access. You can enable these options after creating the service if required.
2. Enable Hierarchical Namespaces. **This option needs to be enabled to create an Azure Data Lake**



Create a storage account ···

ⓘ Certain options have been disabled by default due to the combination of storage account performance, redundancy, and region.

Security

Configure security settings that impact your storage account.

Require secure transfer for REST API operations ⓘ ☑

Enable blob public access ⓘ □

Enable storage account key access ⓘ □

Default to Azure Active Directory authorization in the Azure portal ⓘ □

Minimum TLS version ⓘ Version 1.2

Data Lake Storage Gen2

The Data Lake Storage Gen2 hierarchical namespace accelerates big data analytics workloads and enables file-level access control lists (ACLs). Learn more

Enable hierarchical namespace ☑

SSH File Transfer Protocol (SFTP)

Enables the SSH File Transfer Protocol for your storage account that allows users to access blobs via an SFTP endpoint. Local users need to be created before the SFTP endpoint can be accessed. Learn more

Enable SFTP ⓘ □

ⓘ You must opt-in on a per subscription basis to enable SFTP for hierarchical namespace accounts.

Blob storage

Enable network file system v3 ⓘ □

Allow cross-tenant replication ⓘ □

ⓘ Cross-tenant replication and hierarchical namespace cannot be enabled simultaneously.

Access tier ⓘ

● **Hot:** Frequently accessed data and day-to-day usage scenarios

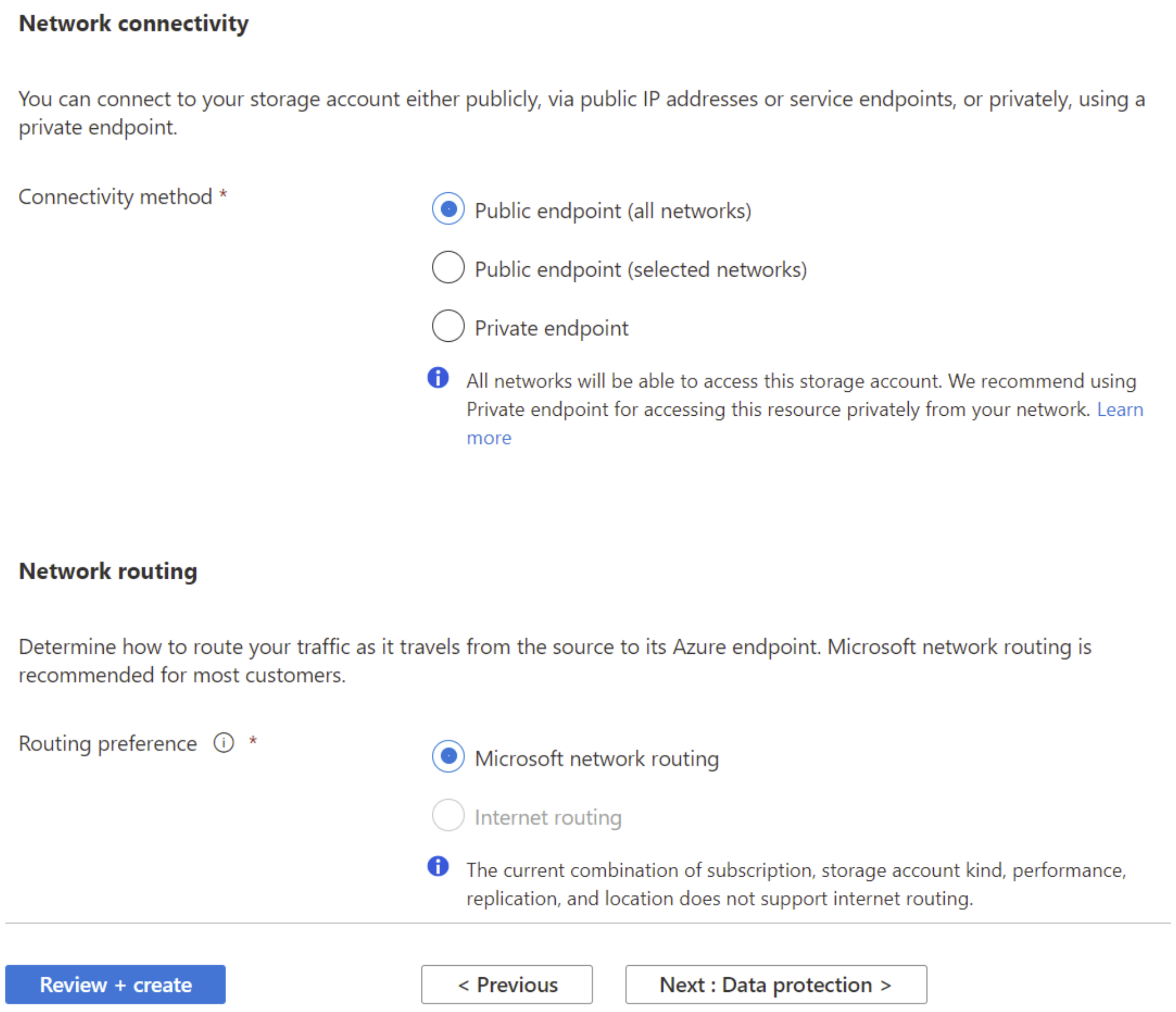
○ **Cool:** Infrequently accessed data and backup scenarios

Azure Files

Enable large file shares ⓘ □

Step 4 - Networking

I highly recommend using Private endpoints if possible when creating your Azure Data Lake, this will require additional configuration steps to enable conectivity with other services.



Network connectivity

You can connect to your storage account either publicly, via public IP addresses or service endpoints, or privately, using a private endpoint.

Connectivity method *

● Public endpoint (all networks)

○ Public endpoint (selected networks)

○ Private endpoint

All networks will be able to access this storage account. We recommend using Private endpoint for accessing this resource privately from your network. Learn more

Network routing

Determine how to route your traffic as it travels from the source to its Azure endpoint. Microsoft network routing is recommended for most customers.

Routing preference *

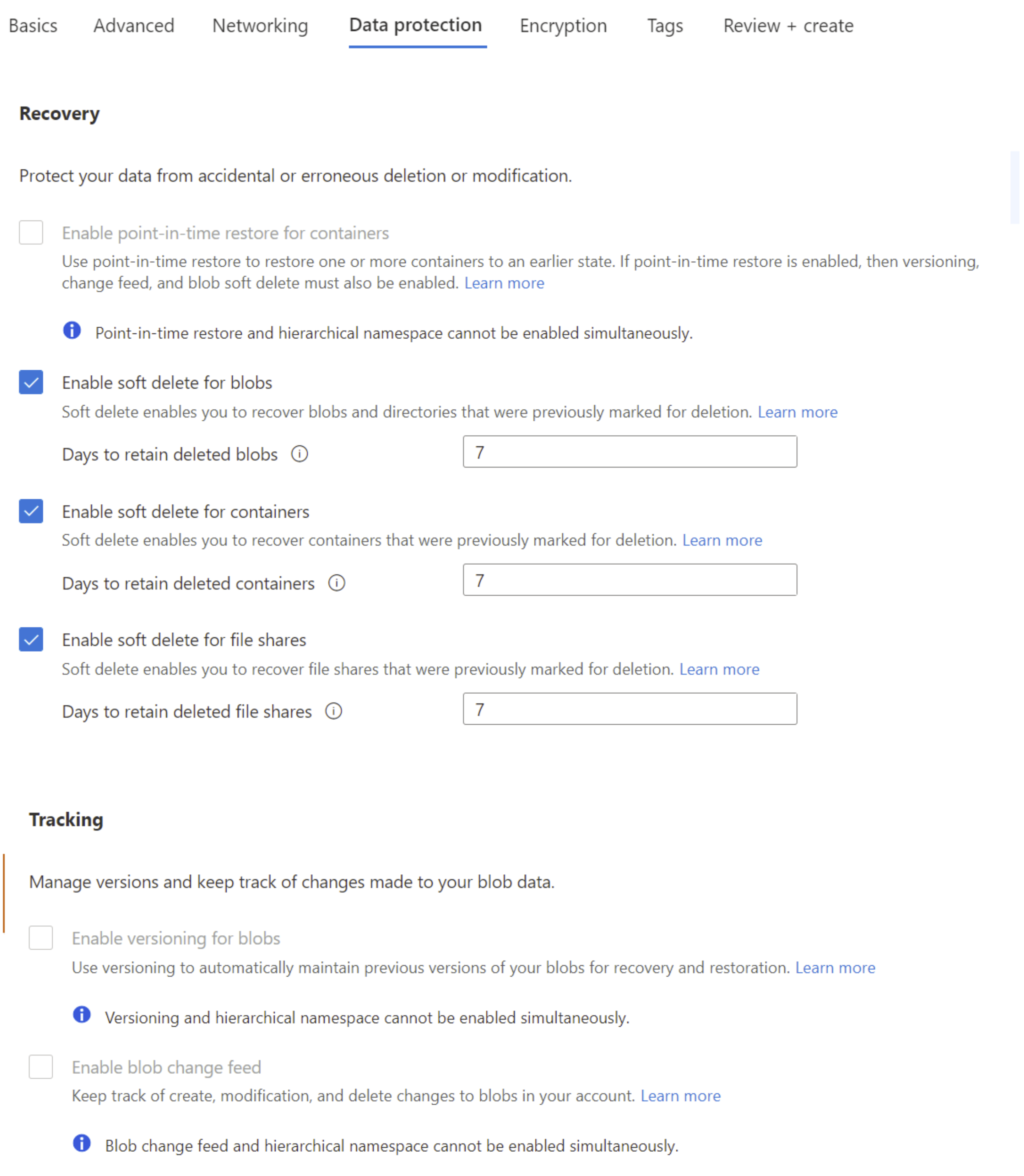
● Microsoft network routing

○ Internet routing

The current combination of subscription, storage account kind, performance, replication, and location does not support internet routing.

Step 5 - Data Protection

Modify the data recovery options as required. Unlike normal storage accounts, tracking capabilities are not available in Azure Data Lake.



Recovery

Protect your data from accidental or erroneous deletion or modification.

□ Enable point-in-time restore for containers
Use point-in-time restore to restore one or more containers to an earlier state. If point-in-time restore is enabled, then versioning, change feed, and blob soft delete must also be enabled. Learn more

Point-in-time restore and hierarchical namespace cannot be enabled simultaneously.

☑ Enable soft delete for blobs
Soft delete enables you to recover blobs and directories that were previously marked for deletion. Learn more

Days to retain deleted blobs: 7

☑ Enable soft delete for containers
Soft delete enables you to recover containers that were previously marked for deletion. Learn more

Days to retain deleted containers: 7

☑ Enable soft delete for file shares
Soft delete enables you to recover file shares that were previously marked for deletion. Learn more

Days to retain deleted file shares: 7

Tracking

Manage versions and keep track of changes made to your blob data.

□ Enable versioning for blobs
Use versioning to automatically maintain previous versions of your blobs for recovery and restoration. Learn more

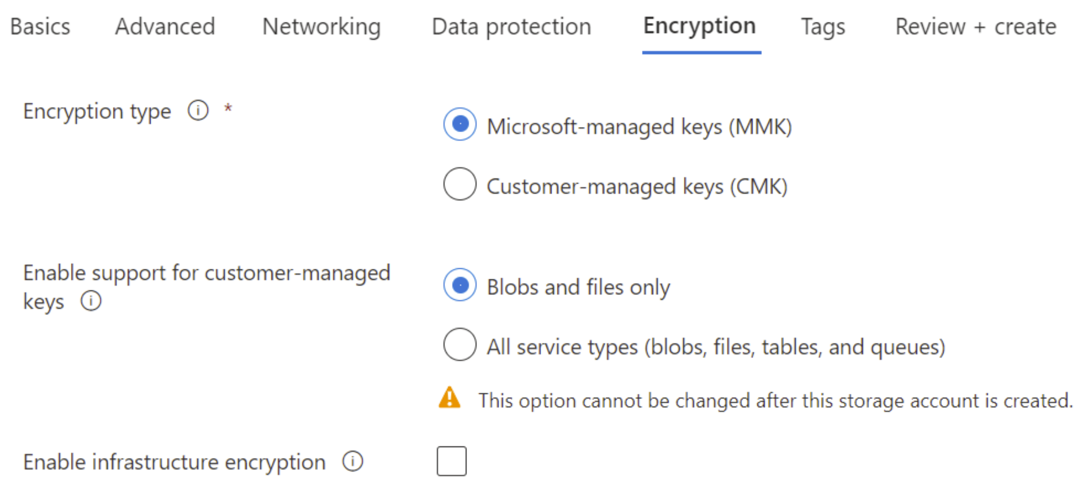
Versioning and hierarchical namespace cannot be enabled simultaneously.

□ Enable blob change feed
Keep track of create, modification, and delete changes to blobs in your account. Learn more

Blob change feed and hierarchical namespace cannot be enabled simultaneously.

Step 6 - Encryption

If you want to use your own keys to encrypt the information at rest, you can configure it under this section.



Encryption type ⓘ *

● Microsoft-managed keys (MMK)

○ Customer-managed keys (CMK)

Enable support for customer-managed keys ⓘ

● Blobs and files only

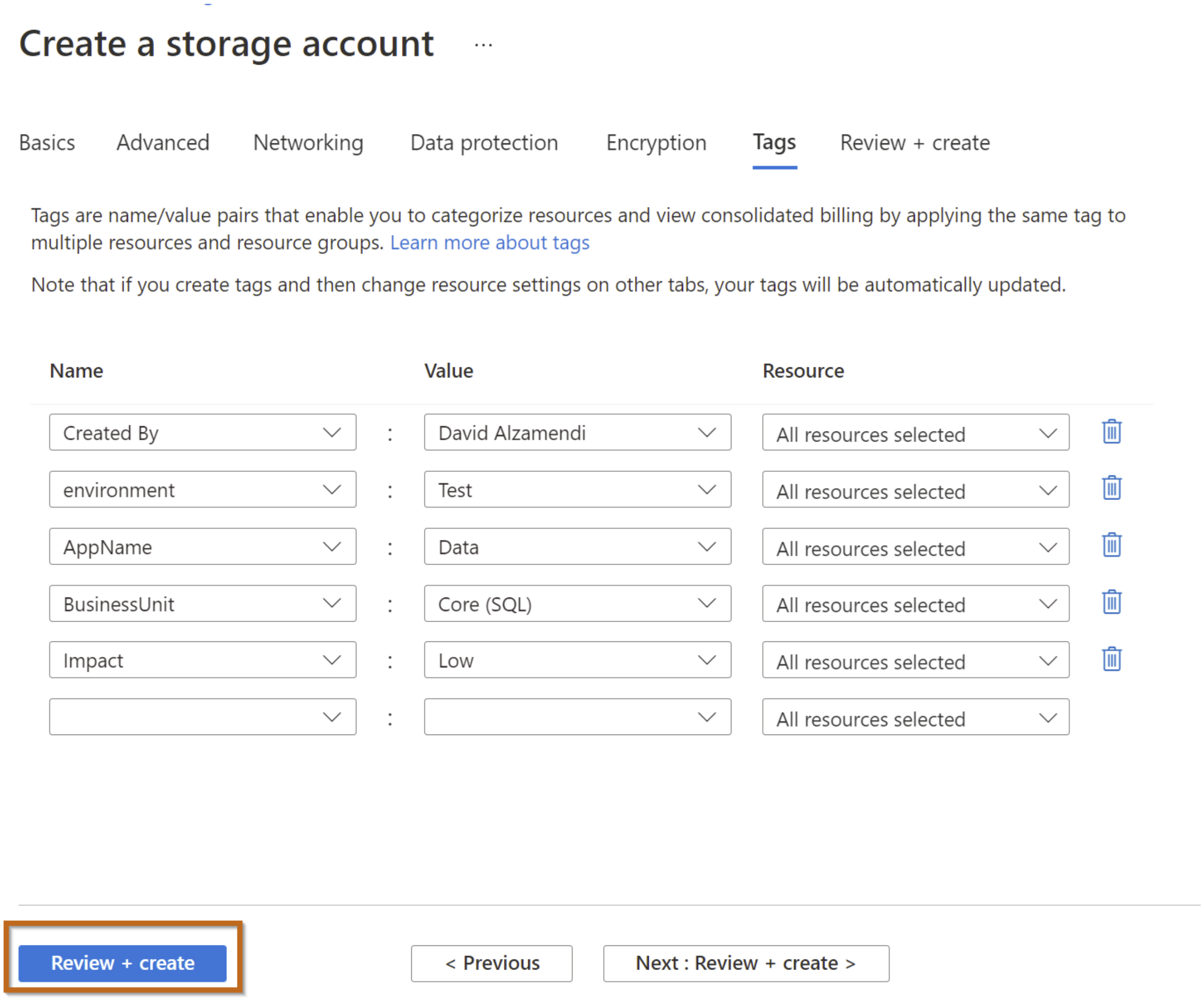
○ All service types (blobs, files, tables, and queues)

⚠ This option cannot be changed after this storage account is created.

Enable infrastructure encryption ⓘ □

Step 7 - Tags

Make sure you create/include some tags that facilitate the administration before you create your Azure Data Lake.



Create a storage account ···

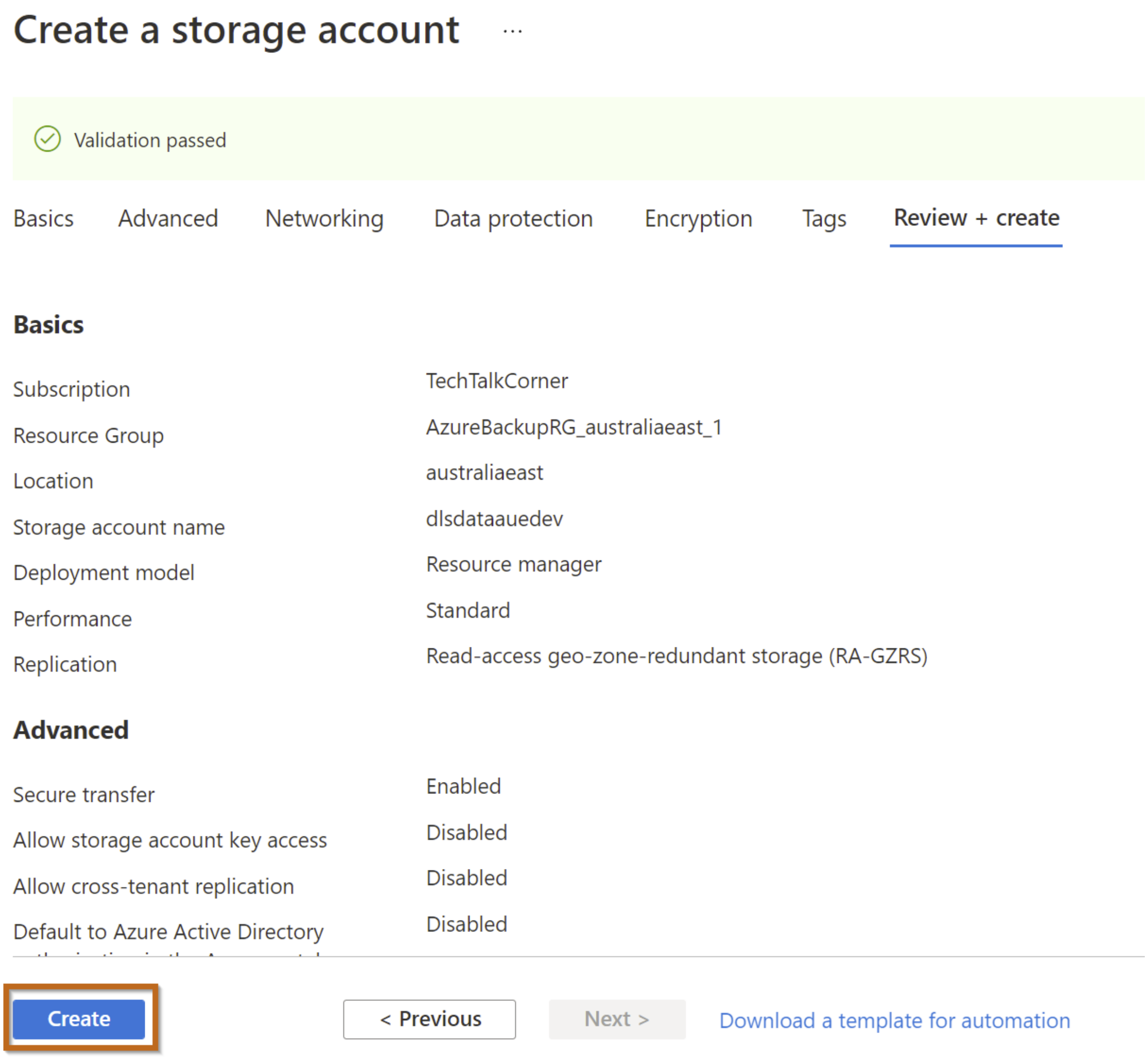
Tags are name/value pairs that enable you to categorize resources and view consolidated billing by applying the same tag to multiple resources and resource groups. Learn more about tags

Note that if you create tags and then change resource settings on other tabs, your tags will be automatically updated.

Name		Value	Resource
Created By	:	David Alzamendi	All resources selected
environment	:	Test	All resources selected
AppName	:	Data	All resources selected
BusinessUnit	:	Core (SQL)	All resources selected
Impact	:	Low	All resources selected
	:		All resources selected

Step 8 - Summary

Last screen before you create your Azure Data Lake!



Create a storage account ···

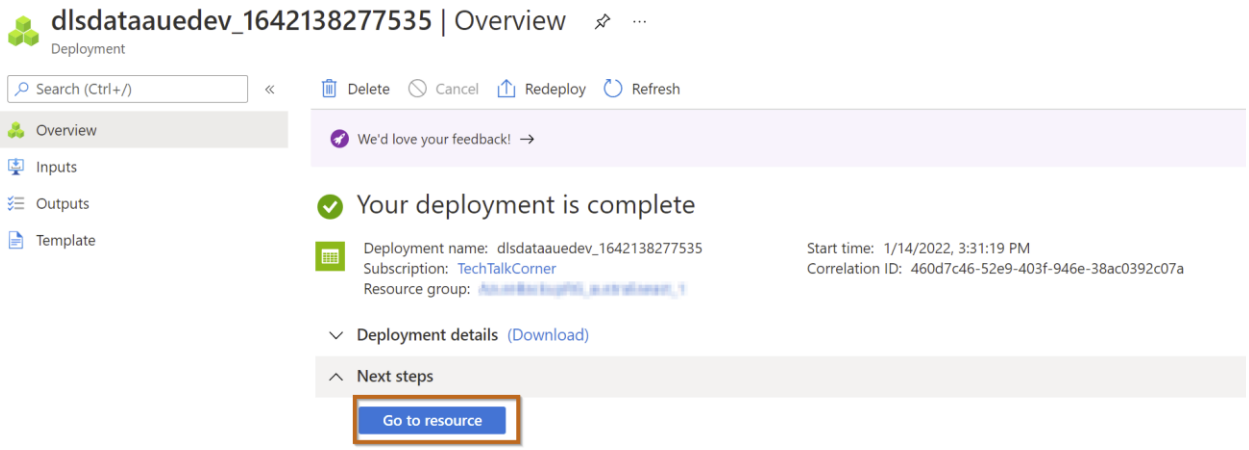
✓ Validation passed

Basics

Subscription	TechTalkCorner
Resource Group	AzureBackupRG_australiaeast_1
Location	australiaeast
Storage account name	dlsdataauedev
Deployment model	Resource manager
Performance	Standard
Replication	Read-access geo-zone-redundant storage (RA-GZRS)

Advanced

Secure transfer	Enabled
Allow storage account key access	Disabled
Allow cross-tenant replication	Disabled
Default to Azure Active Directory	Disabled



dlsdataauedev_1642138277535 | Overview

Deployment

✓ Your deployment is complete

Deployment name: dlsdataauedev_1642138277535
Subscription: TechTalkCorner
Resource group: [redacted]

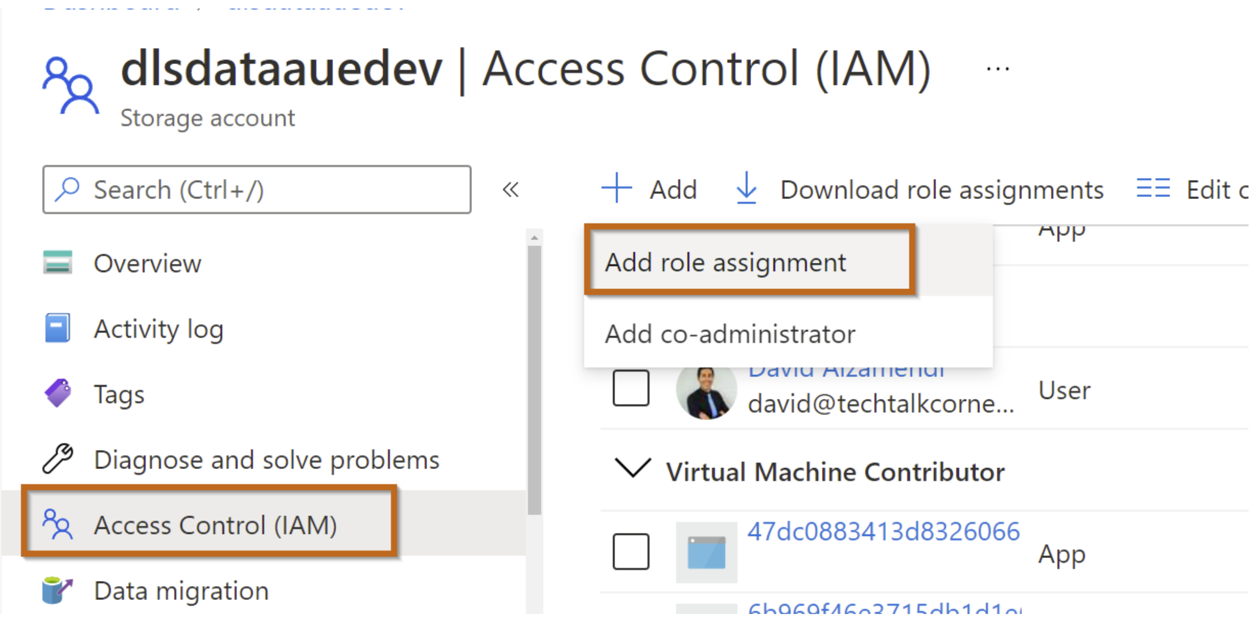
Start time: 1/14/2022, 3:31:19 PM
Correlation ID: 460d7c46-52e9-403f-946e-38ac0392c07a

Step 9 - Access your Azure Data Lake

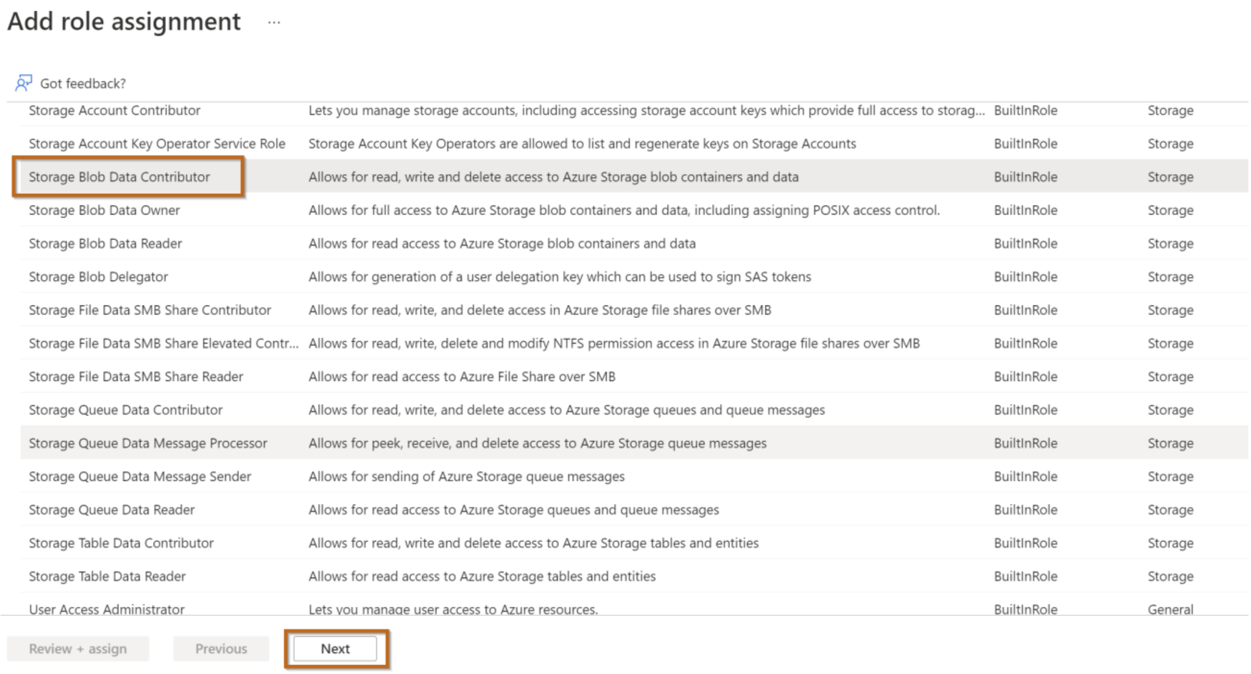
Once Azure Data Lake is created, you can navigate folders using Azure Portal or clients tools like Azure Storage Explorer. In a Data Lake, the access is managed using Role-based access control (RBAC).

Grant Access to your Azure Data Lake

You can grant access by using RBAC.

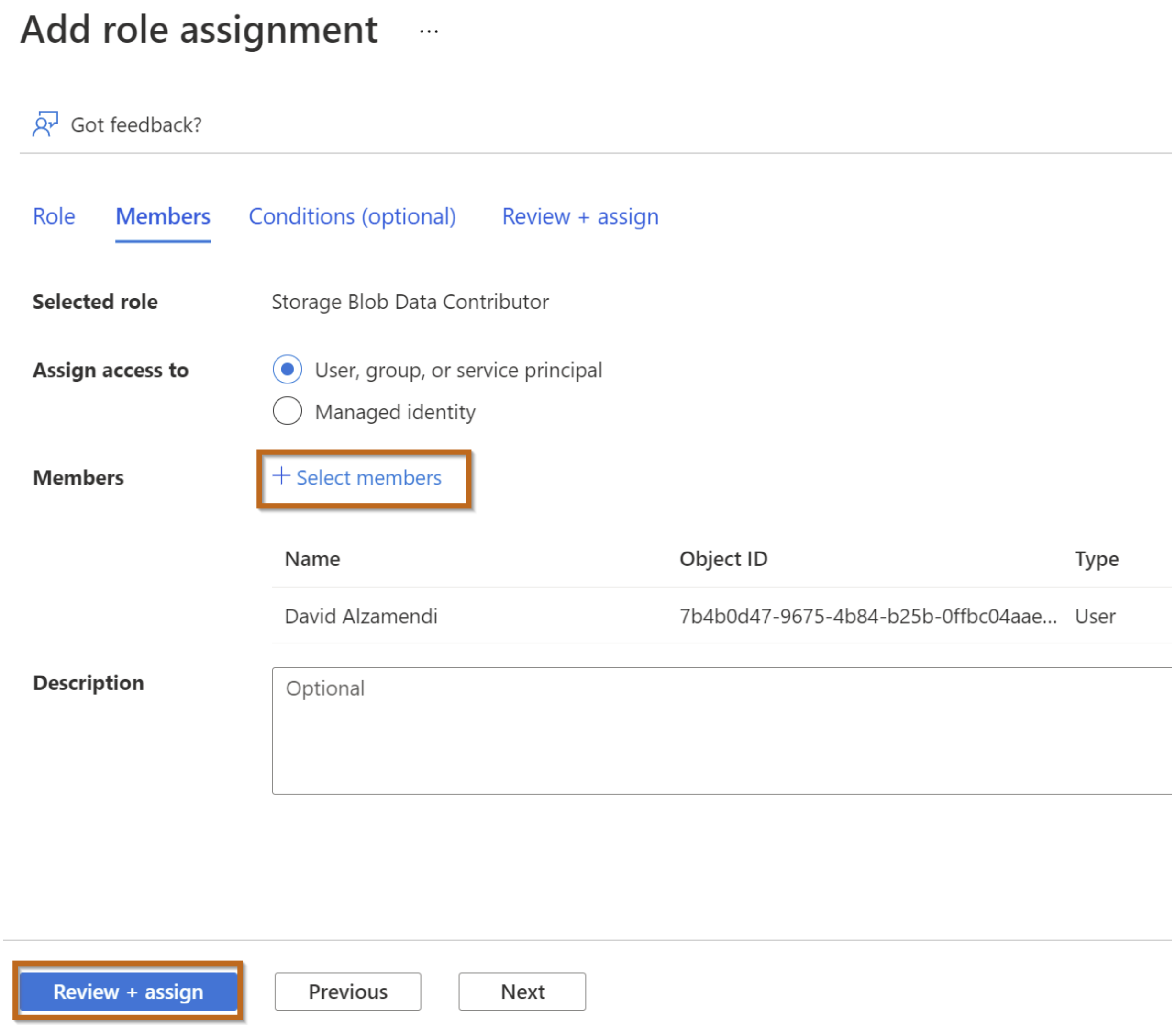


In this case, select Storage Blob Data Contributor.



Storage Account Contributor	Lets you manage storage accounts, including accessing storage account keys which provide full access to storag...	BuiltInRole	Storage
Storage Account Key Operator Service Role	Storage Account Key Operators are allowed to list and regenerate keys on Storage Accounts	BuiltInRole	Storage
Storage Blob Data Contributor	Allows for read, write and delete access to Azure Storage blob containers and data	BuiltInRole	Storage
Storage Blob Data Owner	Allows for full access to Azure Storage blob containers and data, including assigning POSIX access control.	BuiltInRole	Storage
Storage Blob Data Reader	Allows for read access to Azure Storage blob containers and data	BuiltInRole	Storage
Storage Blob Delegator	Allows for generation of a user delegation key which can be used to sign SAS tokens	BuiltInRole	Storage
Storage File Data SMB Share Contributor	Allows for read, write, and delete access in Azure Storage file shares over SMB	BuiltInRole	Storage
Storage File Data SMB Share Elevated Contr...	Allows for read, write, delete and modify NTFS permission access in Azure Storage file shares over SMB	BuiltInRole	Storage
Storage File Data SMB Share Reader	Allows for read access to Azure File Share over SMB	BuiltInRole	Storage
Storage Queue Data Contributor	Allows for read, write, and delete access to Azure Storage queues and queue messages	BuiltInRole	Storage
Storage Queue Data Message Processor	Allows for peek, receive, and delete access to Azure Storage queue messages	BuiltInRole	Storage
Storage Queue Data Message Sender	Allows for sending of Azure Storage queue messages	BuiltInRole	Storage
Storage Queue Data Reader	Allows for read access to Azure Storage queues and queue messages	BuiltInRole	Storage
Storage Table Data Contributor	Allows for read, write and delete access to Azure Storage tables and entities	BuiltInRole	Storage
Storage Table Data Reader	Allows for read access to Azure Storage tables and entities	BuiltInRole	Storage
User Access Administrator	Lets you manage user access to Azure resources.	BuiltInRole	General

Select members as described below.



Add role assignment ...

Selected role Storage Blob Data Contributor

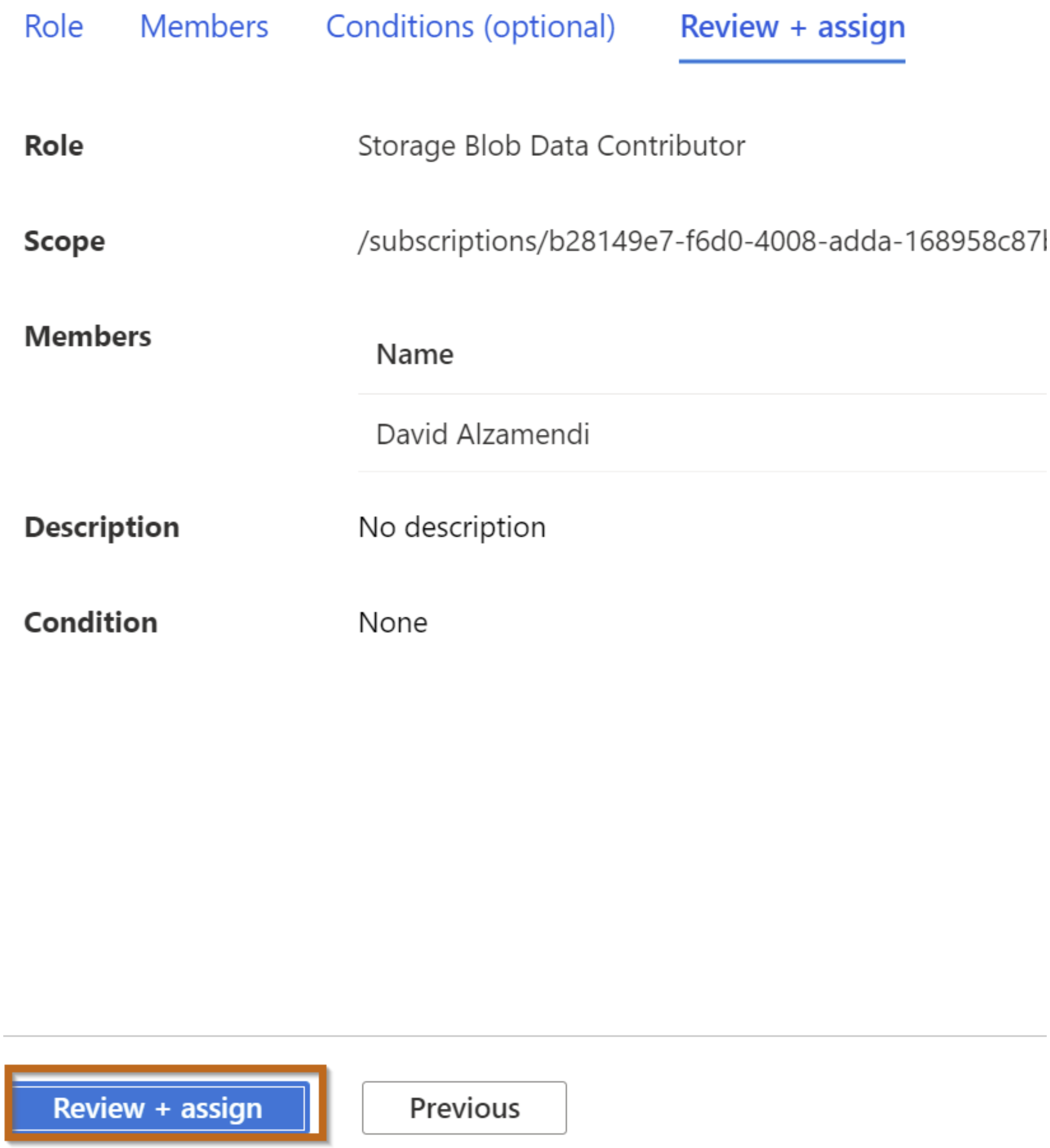
Assign access to
● User, group, or service principal
○ Managed identity

Members + Select members

Name	Object ID	Type
David Alzamendi	7b4b0d47-9675-4b84-b25b-0ffbc04aae...	User

Description Optional

Finally, review and assign.



Role Storage Blob Data Contributor

Scope /subscriptions/b28149e7-f6d0-4008-adda-168958c87

Members

Name
David Alzamendi

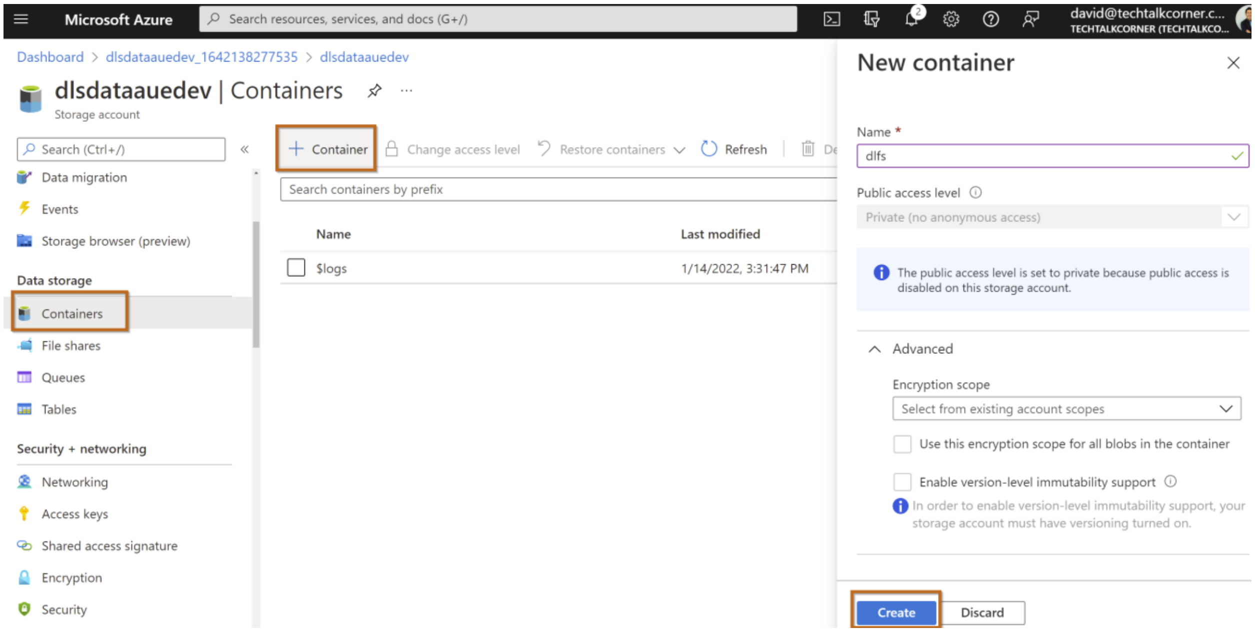
Description No description

Condition None

Create Containers and folders

Now you can create containers and folders.

Create a container as described below:



Dashboard > dlsdataauedev_1642138277535 > dlsdataauedev

dlsdataauedev | Containers

Storage account

New container

Name *: dlfs

Public access level: Private (no anonymous access)

The public access level is set to private because public access is disabled on this storage account.

Advanced

Encryption scope: Select from existing account scopes

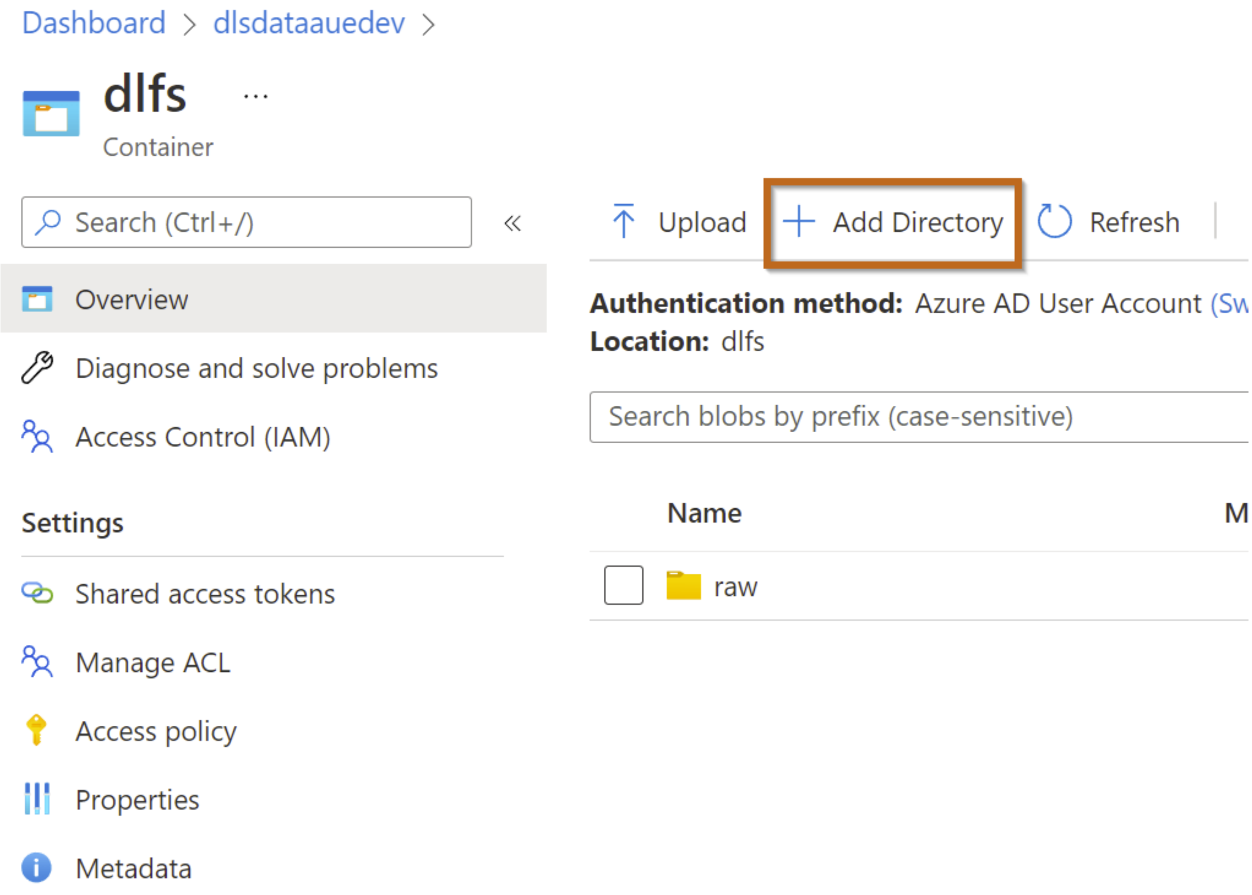
□ Use this encryption scope for all blobs in the container

□ Enable version-level immutability support

In order to enable version-level immutability support, your storage account must have versioning turned on.

Name	Last modified
□ $logs	1/14/2022, 3:31:47 PM

Create a folder inside the container to test permissions.



Dashboard > dlsdataauedev >

dlfs

Container

Authentication method: Azure AD User Account (Sw

Location: dlfs

Name	M
□ raw	