COHESITY

Version 1.0

November 2019

Protect Your Azure VM Data with Cohesity

Cohesity Platform for Azure Native Backup Architecture Reference

ABSTRACT

In the real world, the cloud is equal to a hosted data center. Like every data center, failures are imminent. With Cohesity Azure Native Backup, you can now protect your Azure Virtual Machines and associated volumes against data loss. Cohesity Platform also provides the ability to restore data at granular levels, to the same or different Azure accounts.

Table of Contents

Figures

Tables

Introduction to Azure VM Data Protection

Azure Virtual Machine (VM) is a service that provides secure and on-demand compute resources in the cloud. With more and more organizations using Azure VMs to deploy enterprise workloads, it's imperative to think about backing up data on Azure VMs. Now you can protect your Azure VM data while enjoying the recovery granularity and flexibility of Cohesity Platform.

With Azure Native Backup, Cohesity supports both Azure and Azure Gov cloud.

Backups Protect Azure VMs from Data Loss

Data loss is inevitable in any data center. The same rule applies to data centers hosted by cloud service providers. Table 1 describes some of the possible reasons for data loss in Azure VMs. Backing up those Azure VMs helps mitigate such risks.

Table 1: Causes of Data Loss in Azure VMs

REASON FOR DATA LOSS	DATA LOSS SCENARIOS
Data loss due to platform issues	Azure is a very robust platform, but, like any other data center, hardware failure scenarios are unavoidable. Although Azure can maintain multiple copies of the data across different availability zones, there can be scenarios that cause data loss.
Administrative Errors/Mistakes	Azure VM provisioning and de-provisioning are day-to-day tasks, and prone to administrative errors. In a highly dynamic environment, the risk of an administrator deleting an Azure VM by mistake can be high.
Insider Threat	An insider with access can delete data residing in an Azure VM, or an entire Azure VM, from your infrastructure.
Application Error	There are cases where an application that was not tested properly wiped out data in the Azure VM disks.
Hackers/Ransomware/Virus Attacks	Azure has checks and balances when it comes to external attacks. However, mistakes resulting from application deployment and network glitches can result in external attacks that compromise data.

Given the above scenarios, it becomes crucial to have automated point-in-time backups of the Azure VMs that can be used to restore your data in cases of data loss. This approach can also be used for other use cases, such as cloning Azure VMs for test and dev projects.

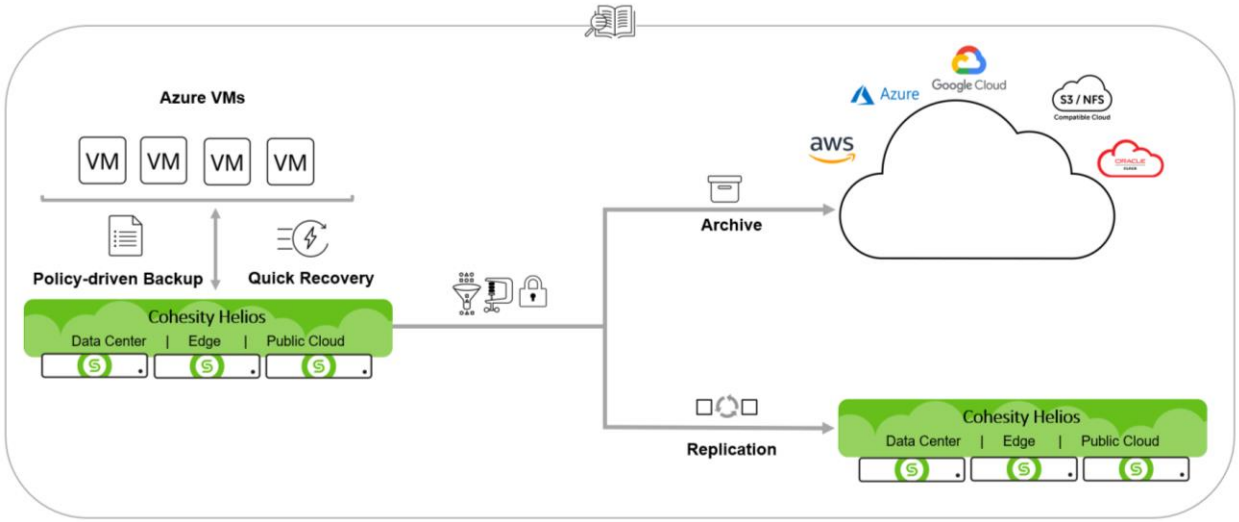
Using Cohesity Platform to perform your Azure VM backups gives you the ability to take control of your Azure infrastructure and associated data by providing the ability to do automated point-in-time backups of the Azure VMs and the associated volumes. It allows you to restore at different levels of granularity, from entire Azure VMs to individual files and folders. What's more, you can use Cohesity Platform to protect and recover your Azure data across multiple Azure accounts.

Cohesity's Azure Native Backup Solution

The Azure native backup feature of Cohesity Platform gives you the ability to back up Azure VMs, along with their attached volumes. The solution uses native Azure APIs for both backup and recovery of Azure VMs, without the need for agents. This helps you protect your entire Azure infrastructure with backup and restore features that keep your data safe when you need to move it around. With your Azure VM backups on Cohesity Platform, you can protect them further with flexible replication and archival to almost any storage platform.

NOTE: Any of the Cohesity editions (Physical, Virtual, or Cloud) can be used for Azure native backups.

Figure 1: Use Cohesity to Protect Azure VMs with Backup, Archive, and Replication to Any Storage



How Azure Native Backups Work with Cohesity

How Azure native backup works depends upon the type of disk used for the Azure VMs. Apart from some of the common steps, the workflow for VMs using managed disks is different from the workflow for VMs using unmanaged disks. Table 2 below compares Azure managed disk and unmanaged disk.

Table 2: Azure Managed Disk vs Azure Unmanaged Disk

AZURE MANAGED DISK	AZURE UNMANAGED DISK
Managed Disks handle the storage account creation and management in the background for you, and ensures that you do not have to worry about the scalability limits of the storage account. You simply specify the disk size and the performance tier (Standard/Premium), and Azure creates and manages the disks for you. As you add disks or scale the VM up and down, you don't have to worry about the storage being used.	Unmanaged disks are the traditional type of disks that have been used by VMs. With these disks, you create your own storage account and specify that storage account when you create the disk. In this option, customers need to be mindful of Input/output operations per second (IOPS) and other limits on the storage account.

Backup Workflow Using Unmanaged Disks

Because unmanaged disks are created and managed by Azure tenants, they have tighter control over their configuration and access.

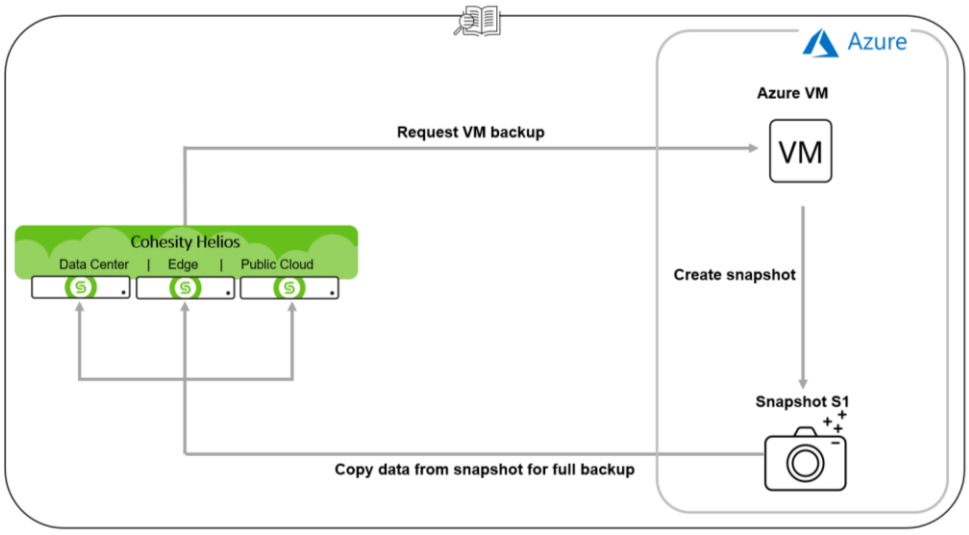
In Azure, unmanaged disks are stored as blobs. Azure Storage provides the capability to take snapshots of blobs. Snapshots capture the blob state at that point in time. For unmanaged disks, Azure provides Azure Storage REST APIs, which makes it possible to take backups directly. Cohesity Azure Native Backup provides the ability to take full as well as incremental backups. The first backup is always full and successive backups can either be full or incremental, based on business needs.

Full Backup Workflow

To take a full backup of VMs that use unmanaged disks, Cohesity DataProtect:

1. Fetches the VM's information, along with details of every resource attached to that VM, such as disks and network details.
2. Creates a snapshot of the unmanaged disk, such as 'Snapshot S1'.
3. Downloads the page ranges of that snapshot.
4. Fetches data from the snapshot. The first time the backup is run, the snapshot ('Snapshot S1') is retained, as it is used to take an incremental backup in the next run. After that, the most recent previous snapshot of the page blob is always retained for the next backup run.

Figure 2: Full Backup of Azure VMs Workflow

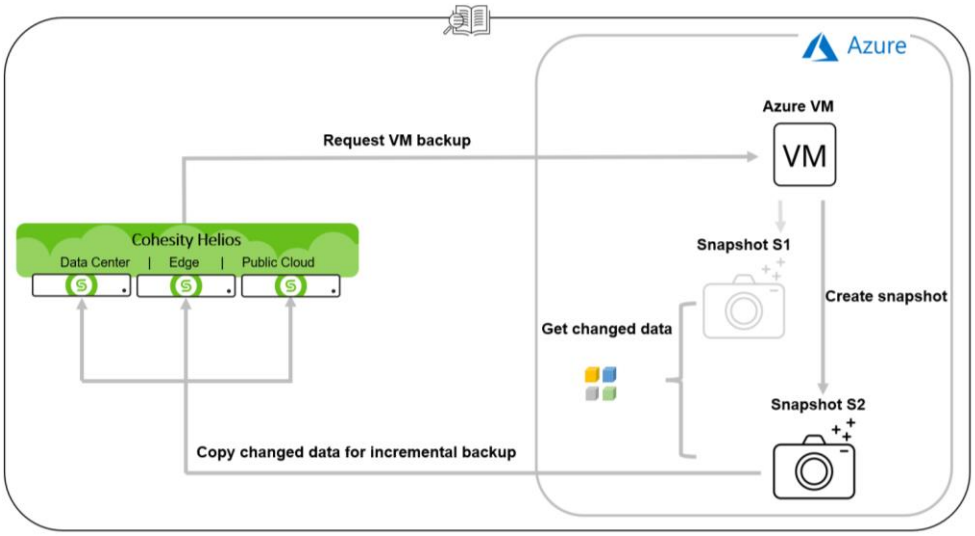


Incremental Backup Workflow

To take an incremental backup of VMs that use unmanaged disks, Cohesity DataProtect:

1. Fetches the VM's information, along with details of every resource attached to that VM, such as disks and network details.
2. Creates a new snapshot ('Snapshot S2').
3. Gets the delta page ranges (the changed data) between the previous snapshot ('Snapshot1') and the new snapshot ('Snapshot S2').
4. Fetches data from only the identified page ranges and downloads them onto the Cohesity cluster.
5. Deletes the previous snapshot ('Snapshot S1') and retains the new snapshot ('Snapshot S2') for the next incremental backup.

Figure 3: Incremental Backup of Azure VMs Workflow



Backup Workflow for VMs Using Managed Disks

A *managed snapshot* is a read-only full copy of a *managed disk*, and is stored as a standard managed disk by default. With snapshots, you can back up a point in time of your managed disks. These snapshots are independent of the source disk and can be used to create new managed disks.

> **NOTE**: Currently, Azure does not provide APIs to take incremental snapshots of managed disks. As a result, every backup of a managed disk is a full back.

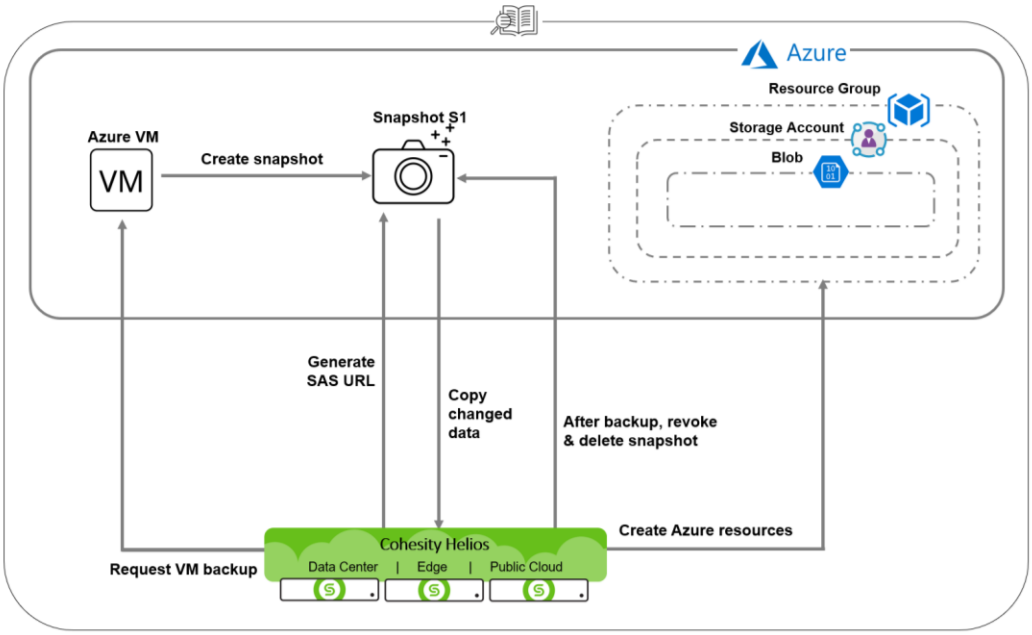
Because managed disks are managed by Azure, there are restrictions on which APIs they expose. They don't provide APIs that directly take backups of managed disks. For that reason, Cohesity takes a different approach to backing up managed disks.

Backup Managed Disks

To take a backup of VMs that use managed disks, Cohesity Platform:

1. Fetches the VM info, such as details of every resource attached to VM e.g. disks, network details.
2. Takes a snapshot (‘Snapshot S1’) of each managed disk attached to the VM.
3. Creates a resource group, storage account, and storage container, which are not used for backup as such, but are used during data recovery. (Azure tenants are not charged for these until and only if they are used.)
4. Generates a SAS URL for the newly created snapshot.
5. Backs up the data using the SAS URL directly to the Cohesity cluster.
6. Downloads only the filled page ranges, not the whole disk.
7. After the backup, revokes the access of the snapshot (‘Snapshot S1’) and then deletes it.

Figure 4: Full Backup of Azure VMs with Managed Disks Workflow



Configure Cohesity Platform for Azure VMs

There are several steps involved in setting up Cohesity Platform to take backups of Azure VMs.

1. Get your Azure subscription details.
2. Register your Azure subscription.
3. Create a Protection Job.

Get Your Azure Subscription Details

You will need the following information for registration.

Table 3: Get Azure Subscription Details

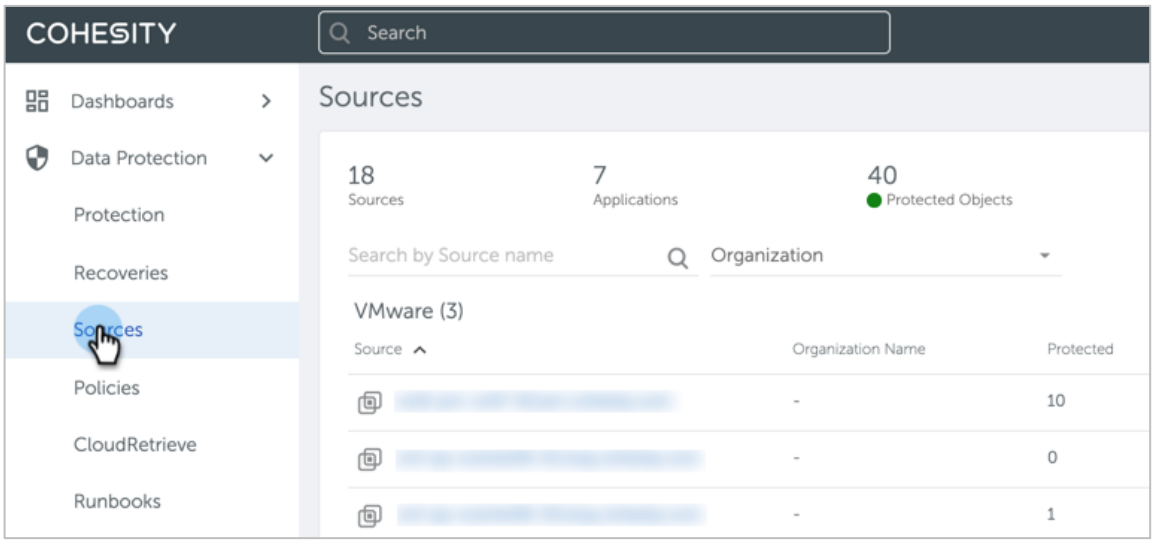
FIELD NAME	DESCRIPTION
Category	Choose Standard Azure or Azure Government. Both categories are supported, but Azure Government provides additional security.
Subscription ID	Enter the Subscription ID for the subscription used. Log in to the Azure portal. In the left panel, click **Subscriptions**. Copy the Subscription ID from the table.
Application ID	Enter the Application ID assigned by Azure during the service principal creation process. Learn how to get an application ID and authentication key in the Microsoft Azure documentation.
Authentication Key	Enter the Azure Authentication Key generated using the Legacy App Registration during the service principal creation process. Learn how to get an application ID and authentication key in the Microsoft Azure documentation.
Tenant ID	Enter the unique Tenant ID assigned by Azure. Learn how to get a tenant ID in the Microsoft Azure documentation.

For details on collecting this information, see Appendix C: Prepare Azure Subscription to Register with Cohesity.

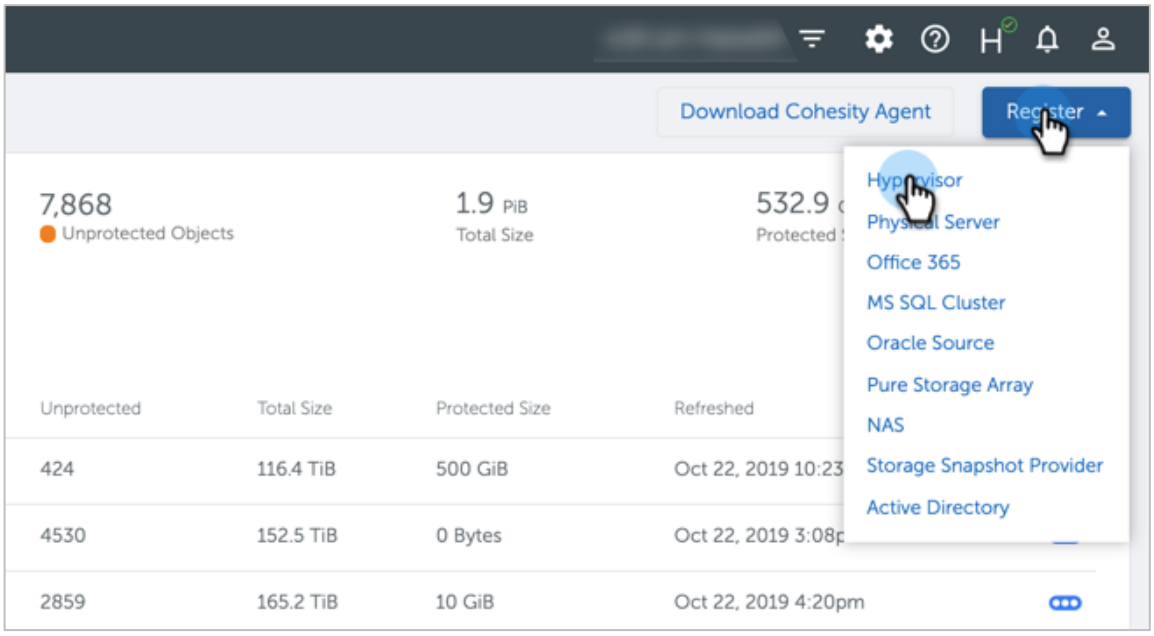
Register Your Azure Subscription

To register your Azure subscription to Cohesity Platform, follow steps below:

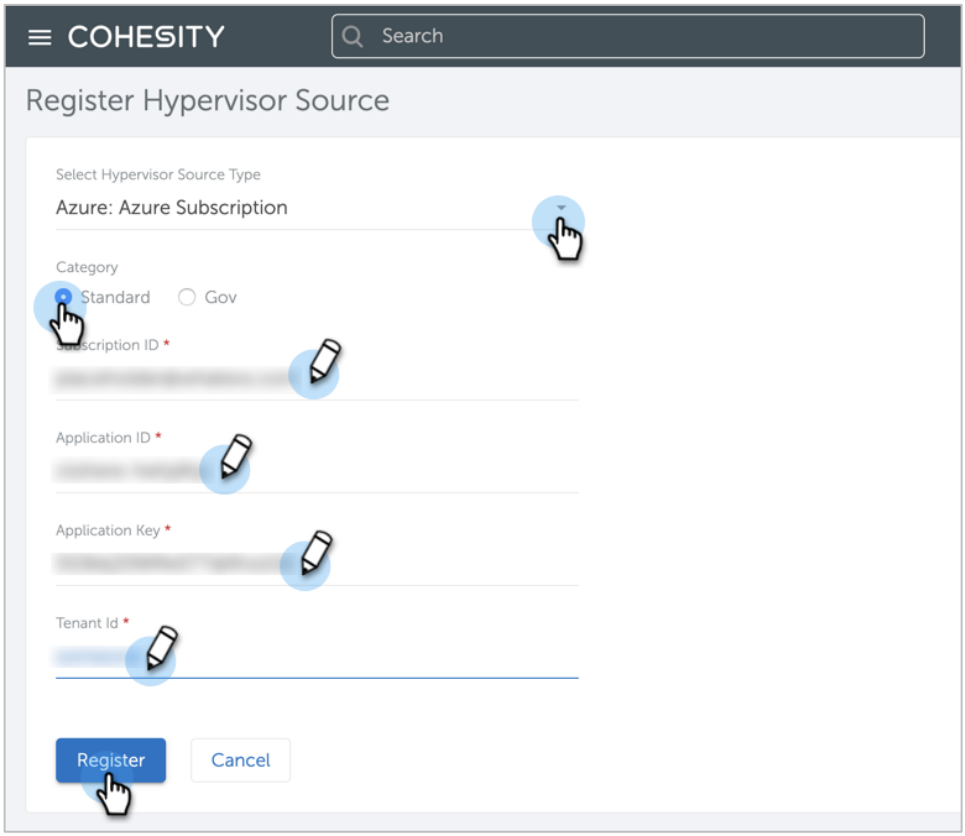
1. Log in to Cohesity Platform and select **Data Protection > Sources** on the left.



2. On the **Sources** page, select **Register > Hypervisor** on the right.



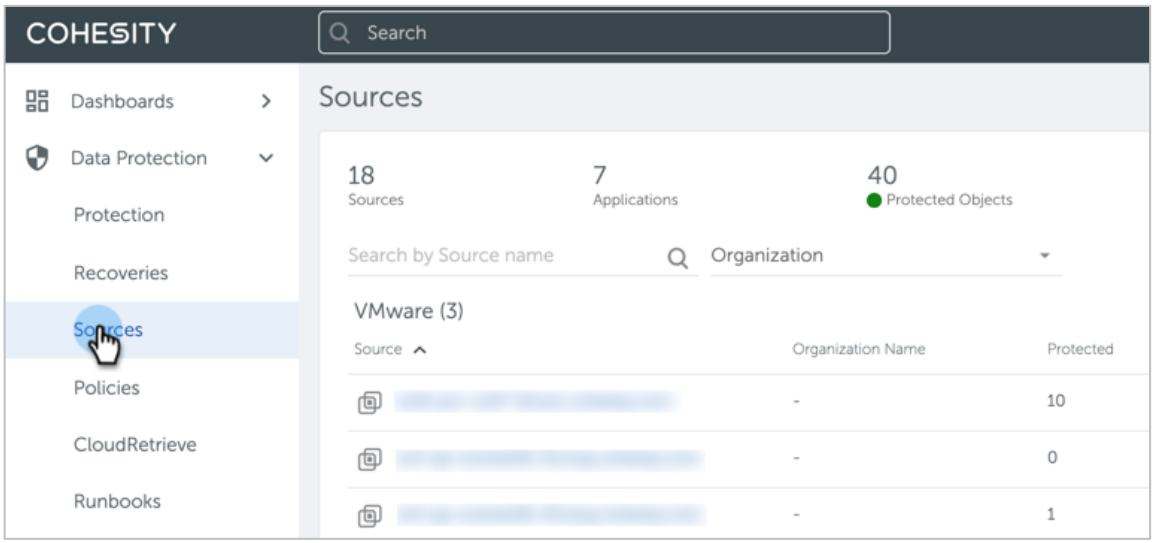
3. Select **Azure: Azure Subscription** under **Select Hypervisor Source Type**, and select the **Category**, **Standard** or **Gov** (both of which are supported)**.** Enter the **Subscription ID, Application ID, Application Key,** and **Tenant Id.** Click **Register** to proceed.



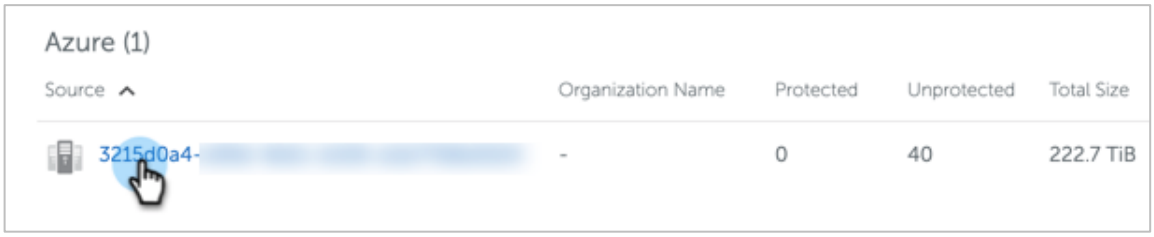
Create a Protection Job

Create a Protection Job by adding objects and assigning a Protection Policy to perform backups:

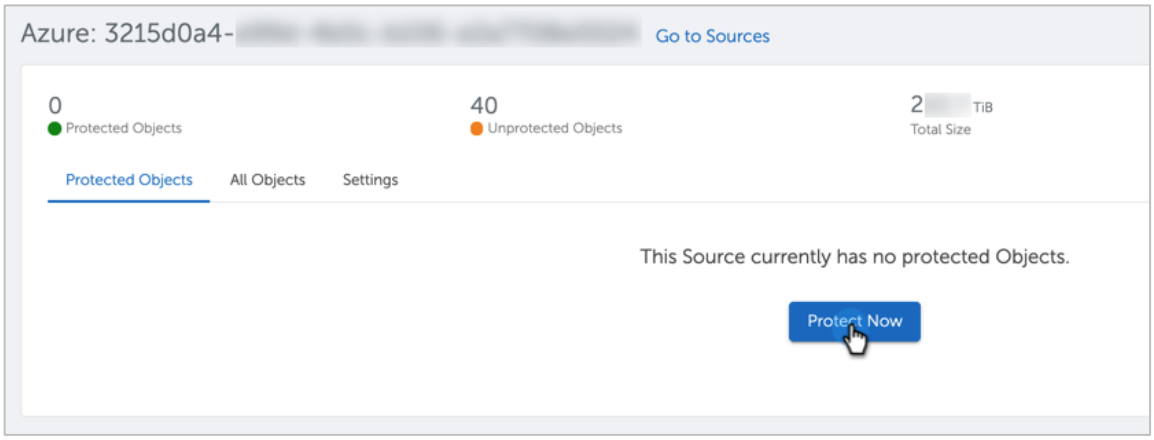
1. Log in to Cohesity Platform and select **Sources**.



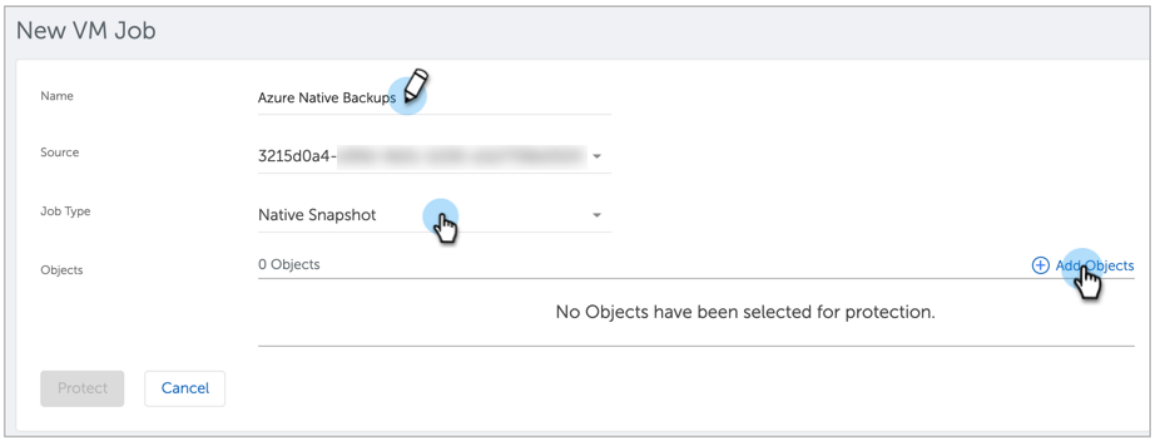
2. Click the Azure subscription ID.



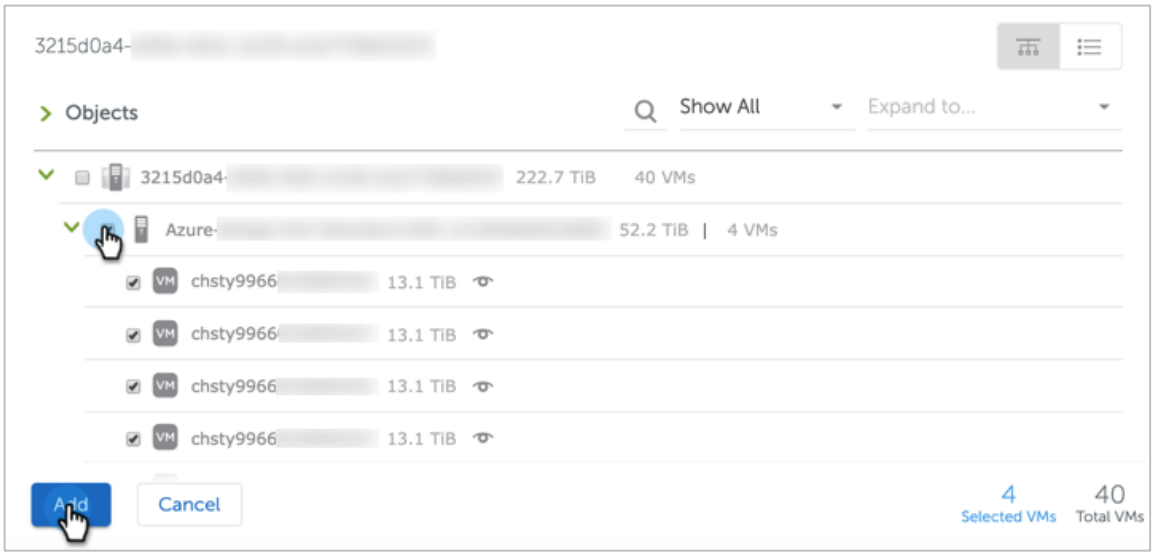
3. Click **Protect Now**.



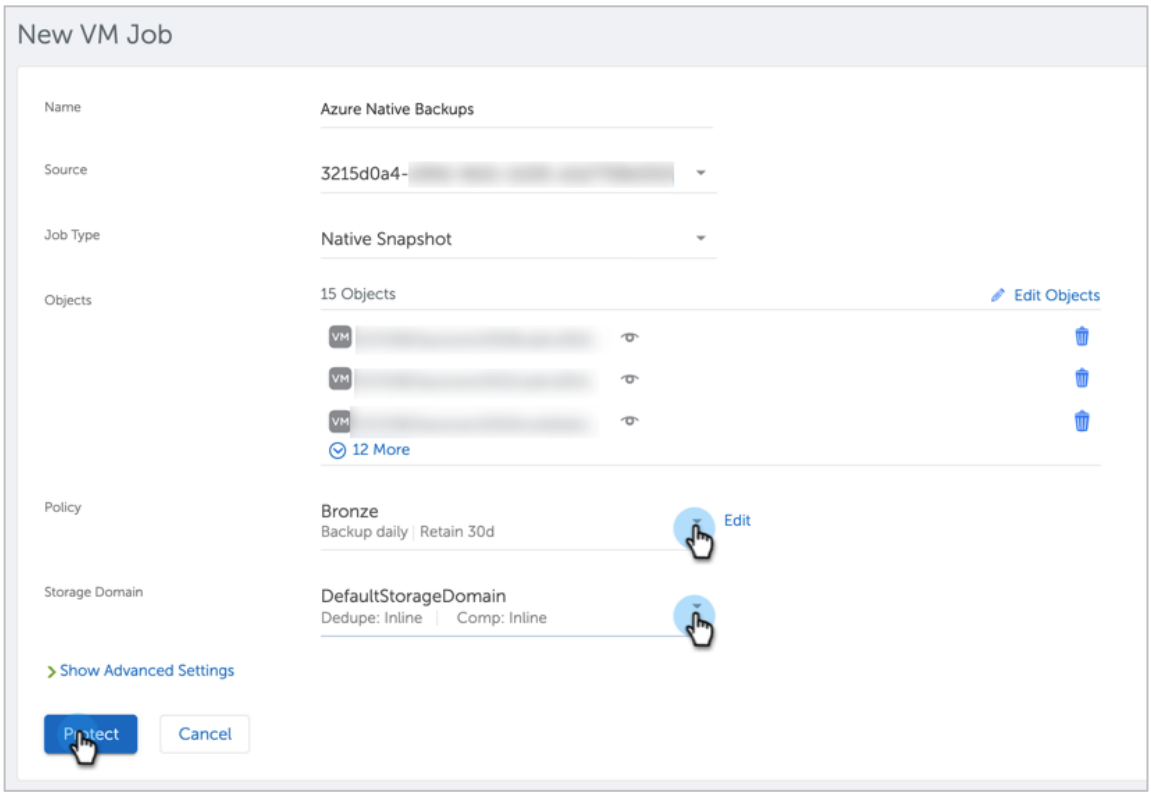
4. Enter the **Name** and **Job Type** (as **Native Snapshot**). Click **Add Objects.**



5. Select the VMs to be backed up and click **Add**.



6. Select the **Policy** and **Storage Domain**, and then click **Protect** to create the Protection Job.



Recover Data with Cohesity Azure Native Backup

Recovery from Cohesity Azure native backups can be performed at different levels of granularity, and to both the original and alternate locations. Data can be restored across Azure accounts and regions.

The restore granularity levels are:

1. **Protection Job**. All Azure VMs that are part of a Protection Job can be recovered to a previous point in time using backup snapshots in a single click.
2. **Azure VM**. Individual Azure VMs can be searched using their names, or the tags associated with them, and recovered from a backup snapshot.
3. **Files and Folders**. Starting with version 6.4, specific files and folders that reside in an Azure VM can be restored.

How Recovery Works with Azure Native Backup

Recovery is the most important aspect of any backup solution, and with Cohesity Platform, you can recover data at every level of granularity.

Although there are different workflows, they all involve these steps:

1. Go to **Protection > Recovery**, click **Recover** and select **VMs**.
2. Search for an Azure VM using its name or assigned tags.
3. Select the Azure VM to be restored and initiate a restore.

Restore Workflows

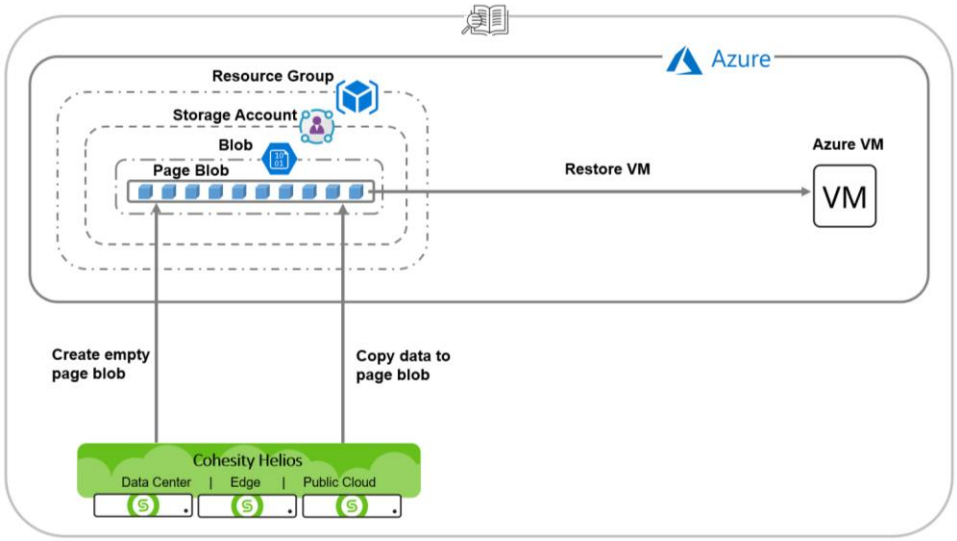
The workflow for data recovery for Azure VMs using an unmanaged disk is different from the recovery workflow for Azure VMs using a managed disk.

Restore Azure VMs Using Unmanaged Disks

The restore workflow for unmanaged disks involves the following steps:

1. You identify the point in time which you wish to restore and select Job run ***<MM DD, YYYY HH:MM>*** from the Protection Job you are restoring. Cohesity Platform creates an empty page blob on Azure.
2. Cohesity DataProtect uploads the data of the snapshot for Job run ***<MM DD, YYYY HH:MM>*** from the Cohesity cluster to that page blob.
3. Finally, it creates the VM with the data uploaded to the page blob.

Figure 5: Restore Azure VM Using Unmanaged Disks

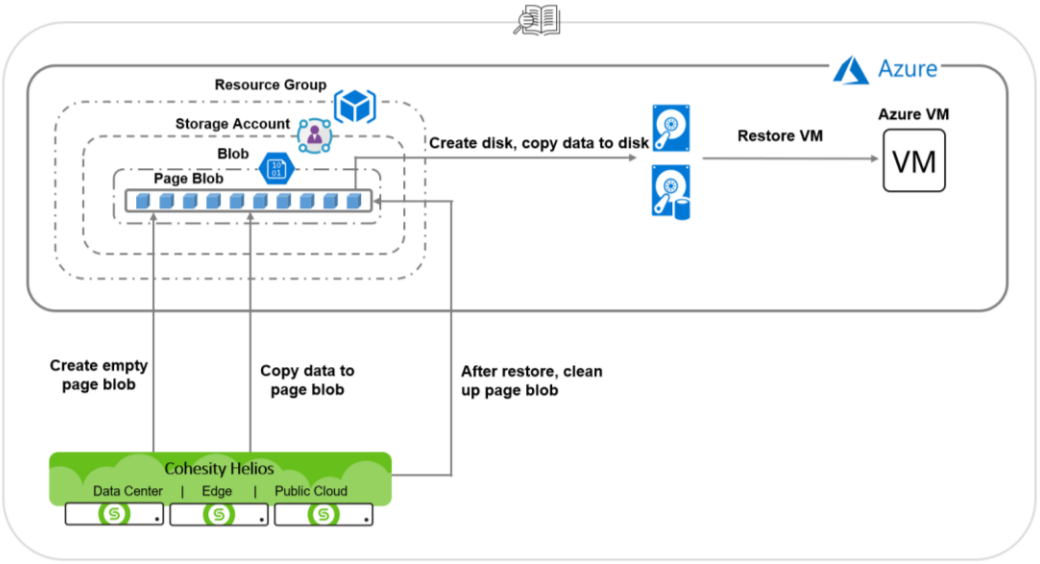


Restore Azure VMs Using Managed Disks

The restore workflow for managed disks involves the following steps:

1. You identify the point in time which you wish to restore and select Job run ***<MM DD, YYYY HH:MM>*** from the Protection job you are restoring. Cohesity DataProtect creates an empty page blob of the same size as the disk size on Azure.
2. Cohesity DataProtect uploads the data of the snapshot for Job run ***<MM DD, YYYY HH:MM>*** from the Cohesity cluster to that page blob.
3. It then creates a managed disk from the new page blob.
4. Next, it retrieves the ID of the created managed disk, and creates the VM using the managed disk.
5. Finally, after the restore completes, it cleans up the page blob.

Figure 6: Restore Azure VM Using Managed Disks



Azure Resources Created by Cohesity Platform

In the process of running Azure native backups, Cohesity Platform creates resources on Azure.

Table 4: Resources Created in Azure by Cohesity Platform

RESOURCE	WORKFLOW	COUNT	REASON	LIFE CYCLE	NAMING CONVENTION
Resource Group	Restore	Per region per subscription	To restore VMs	Permanent	"cohesity" + <region of source vm> + "-rg"
Storage Account	Restore	Per region per subscription per cluster	To copy data from Cohesity to Azure	Permanent	"cohesity" + Hash(<cluster_id> + <region of vm > + <subscription_id>)
Storage Container	Restore	Per region per subscription per cluster	To copy data from Cohesity to Azure	Permanent	"cohesity" + <region of source vm> + "-sc"
VMName	Restore	Per Restore per VM	For VM restore	Permanent	Same as backed up VM name if no prefix and suffix are provided through UI.
Managed Disk Name	Restore	Per Restore per VM	For VM restore	Permanent	Same as backed up disk name.
Blob Name	Restore	Per Restore per VM	For VM restore	Permanent (only for managed disks)	"Cohesity_" + <cluster_id> + <job_id> + <task_id> + <backed_up blob name>
Network Interface Name	Restore	Per Restore per VM	For VM restore	Permanent	**For restoring in same location**: <VM Name > + Hash(<backed_up interface name> + <cluster_id> + <job_id> + <task_id>) **For restoring in different location**: <VM Name> + "-nic"

Appendix A: Azure Native Backup Terminology

There are several terms that are especially important to understand as you learn about the architecture of Azure native backups on Cohesity Platform.

Table 5: Azure Native Backup Terminology

TERM	DEFINITION
Azure Virtual Machines	Compute allocation in the cloud. Azure VM being backed up is addressed as the “Source VM.”
Azure Disk Storage	Block storage allocated in the form of disks to an Azure VM.
Virtual Network	Virtual network dedicated to your Azure account.
Tags	A tag is a label that you or Azure assigns to an Azure resource. Each tag consists of a key and a value.

Appendix B: Egress Cost Considerations

Table 6 below outlines the deployment models and associated egress cost for backup and restore of data using Cohesity's Azure native backup solution.

Table 6: Egress Cost Considerations

COHESITY PLATFORM	REGION	BACKUP	RESTORE
On-Premise	NA	Yes	No
Cloud Edition	Same Region	No	No
	Different Region	Yes	Yes

Appendix C: Prepare Azure Subscription to Register with Cohesity Platform

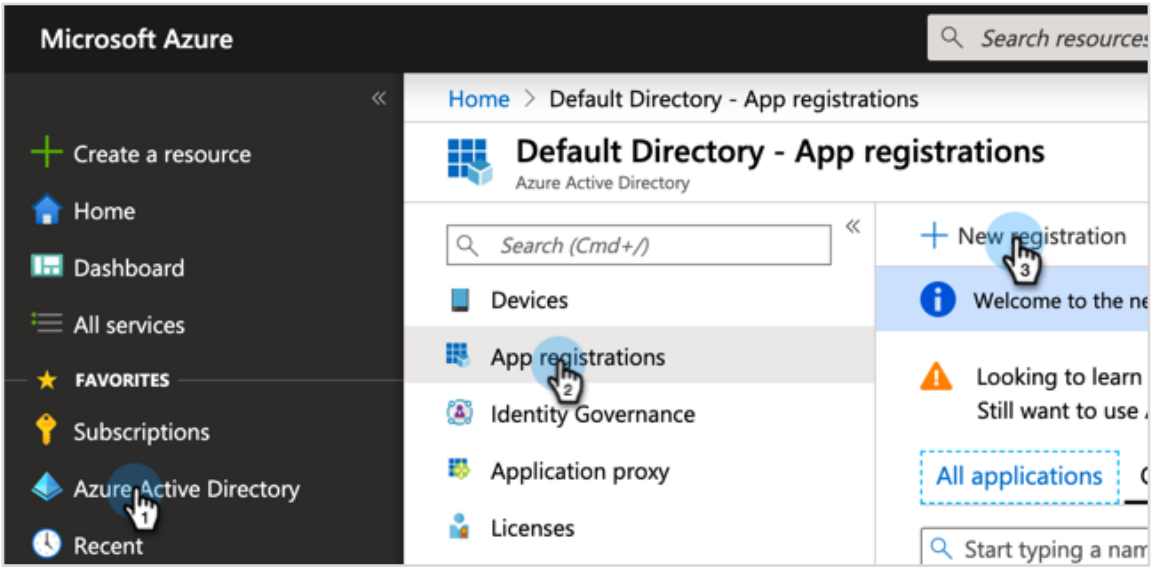
Before you register an Azure subscription with Cohesity Platform, some settings must be configured on Azure.

You'll need to:

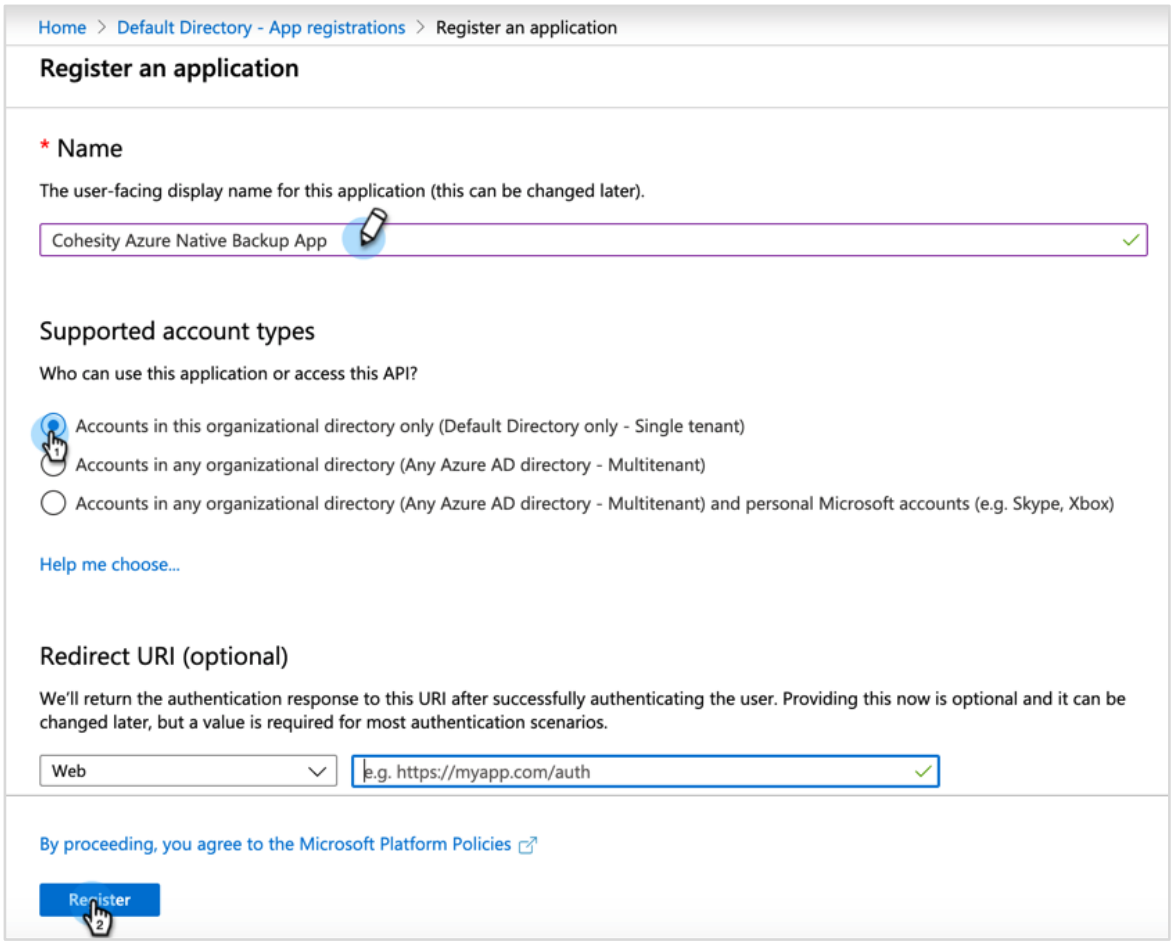
1. Register an App on Azure.
2. Create a custom role and assign permissions using the Azure command-line interface (CLI).
3. Assign the custom role and App to the subscription.

Register an App on Azure

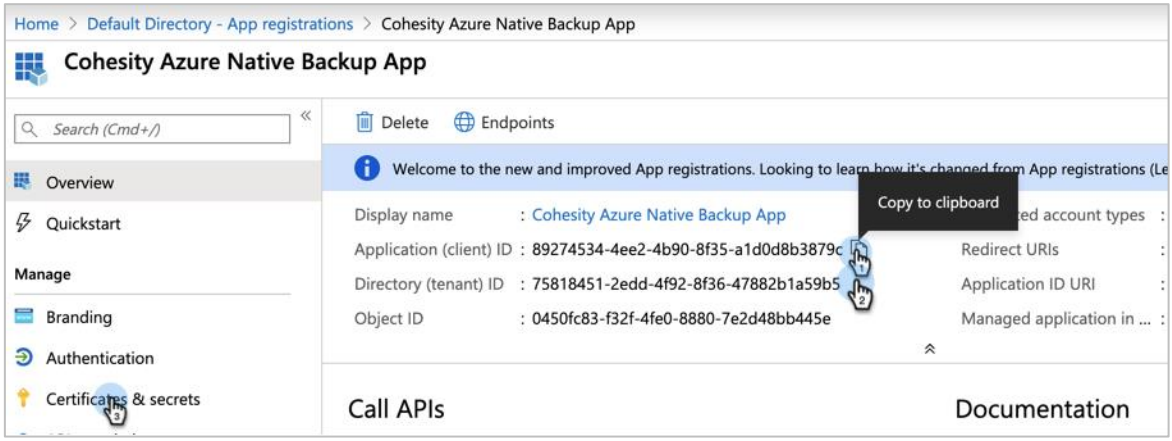
1. Log in to the Azure Portal at: https://portal.azure.com.
2. Navigate to **Azure Active Directory > App registrations**. Click **New registration** to register a new App.



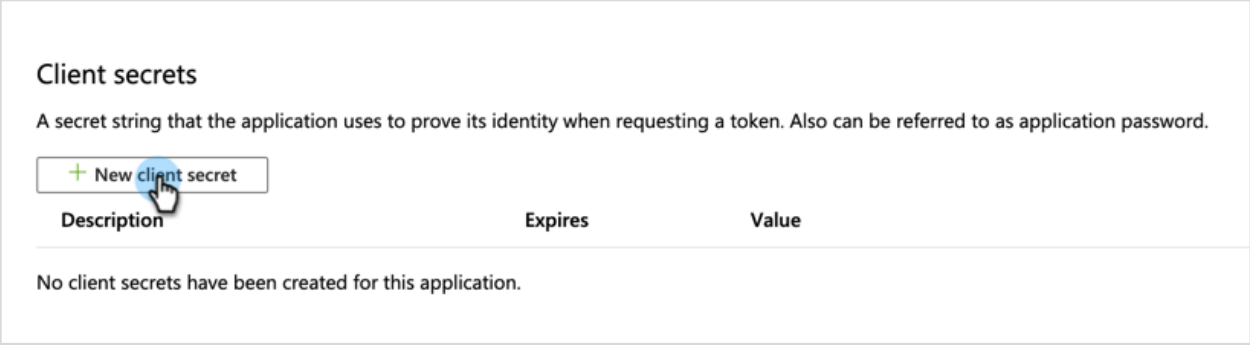
3. Enter the **Name, Supported account types**, and optionally the **Redirect URI**. Click **Register** to continue.



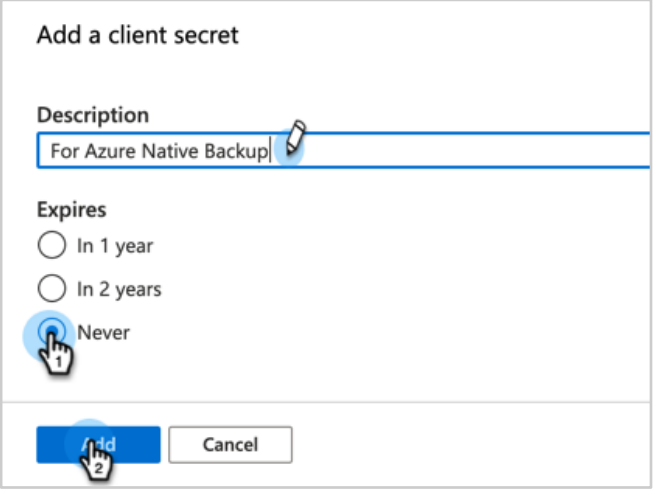
4. Copy the **Application (client) ID** and **Directory (tenant) ID** for the source registration process. Then click **Certificates & secrets**.



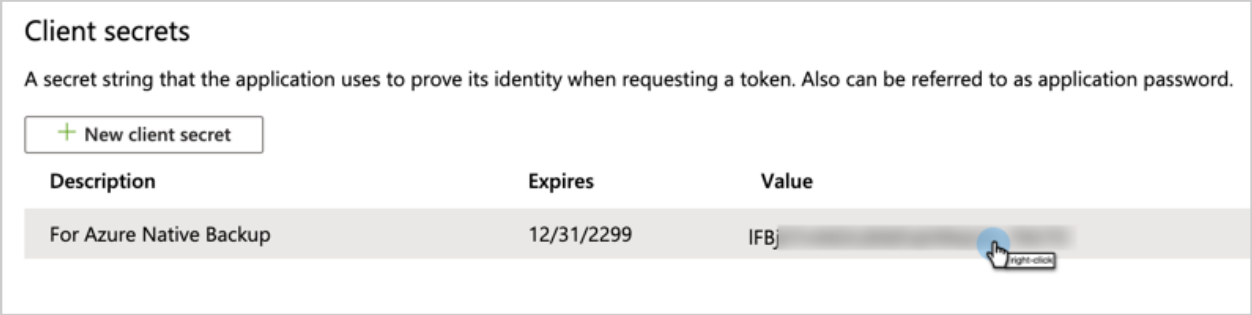
5. Click **New client secret**.



6. Enter a **Description** and select a period for **Expires**, then click **Add**.



7. Copy the **App Secret** for source registration.



Create a Custom Role and Assign Permissions Using the Azure CLI

1. In the Azure portal, navigate to **Subscriptions,** click **Subscription**, and copy the **Subscription ID** for source registration**.**

2. To create a custom role, download, install, and log in to the Azure command-line interface. For instructions, see Install the Azure CLI.

> **NOTE:** Custom roles can only be created using Azure CLI.

3. Log in to the Azure CLI and enter the command below.

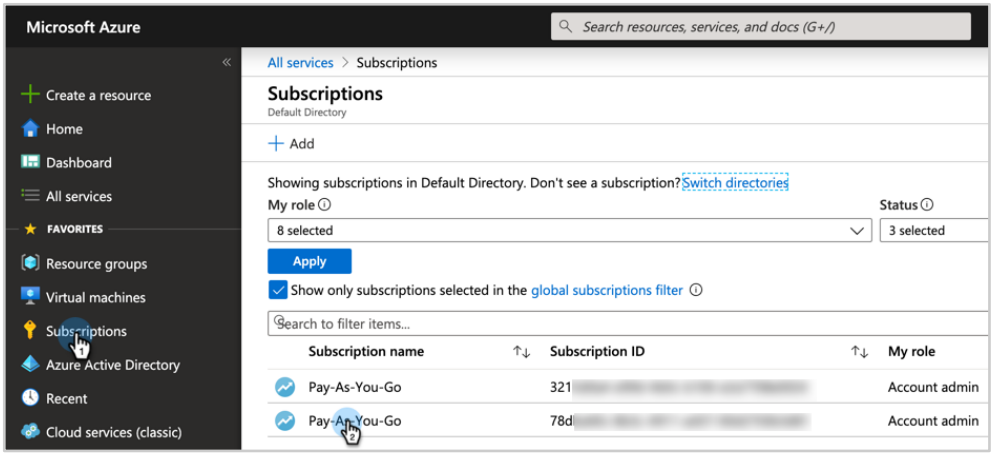
> **IMPORTANT**: In the JSON command below, you must replace ***<SUBSCRIPTION-ID>*** with your actual Subscription ID

```
az role definition create --role-definition '{
"Name": "MinPerAzureNativeBackupRecovery",
"Description": "Azure Native Managed Disk VM Backup & Recovery " ,
"IsCustom": true,
"Actions": [
"Microsoft.Compute/virtualMachines/read",
"Microsoft.Compute/disks/read",
"Microsoft.Compute/snapshots/read",
"Microsoft.Network/networkInterfaces/ipconfigurations/read",
"Microsoft.Network/networkSecurityGroups/securityRules/read",
"Microsoft.Network/networkSecurityGroups/read",
"Microsoft.Network/networkInterfaces/read",
"Microsoft.Resources/subscriptions/resourcegroups/read",
"Microsoft.Resources/subscriptions/resourcegroups/write",
"Microsoft.Storage/storageAccounts/write",
"Microsoft.Storage/storageAccounts/read",
"Microsoft.Storage/storageAccounts/blobServices/containers/write",
"Microsoft.Storage/storageAccounts/blobServices/containers/read",
"Microsoft.Compute/snapshots/write",
"Microsoft.Compute/disks/beginGetAccess/action",
"Microsoft.Storage/storageAccounts/listkeys/action",
"Microsoft.Compute/snapshots/beginGetAccess/action",
"Microsoft.Compute/snapshots/endGetAccess/action",
"Microsoft.Compute/snapshots/delete",
"Microsoft.Compute/disks/endGetAccess/action",
"Microsoft.Compute/disks/write",
"Microsoft.Network/networkInterfaces/write",
"Microsoft.Network/networkInterfaces/join/action",
"Microsoft.Network/virtualNetworks/subnets/join/action",
"Microsoft.Compute/virtualMachines/write",
"Microsoft.Compute/virtualMachines/powerOff/action",
"Microsoft.Compute/virtualMachines/deallocate/action",
"Microsoft.Network/networkInterfaces/delete",
"Microsoft.Compute/disks/delete",
"Microsoft.Network/networkSecurityGroups/join/action",
"Microsoft.Network/virtualNetworks/read",
"Microsoft.Network/virtualNetworks/subnets/read",
"Microsoft.KeyVault/vaults/deploy/action"
],
"DataActions": [],
"NotActions": [],
"NotDataActions": [],
"AssignableScopes": ["/subscriptions/<SUBSCRIPTION-ID>"]
}
```

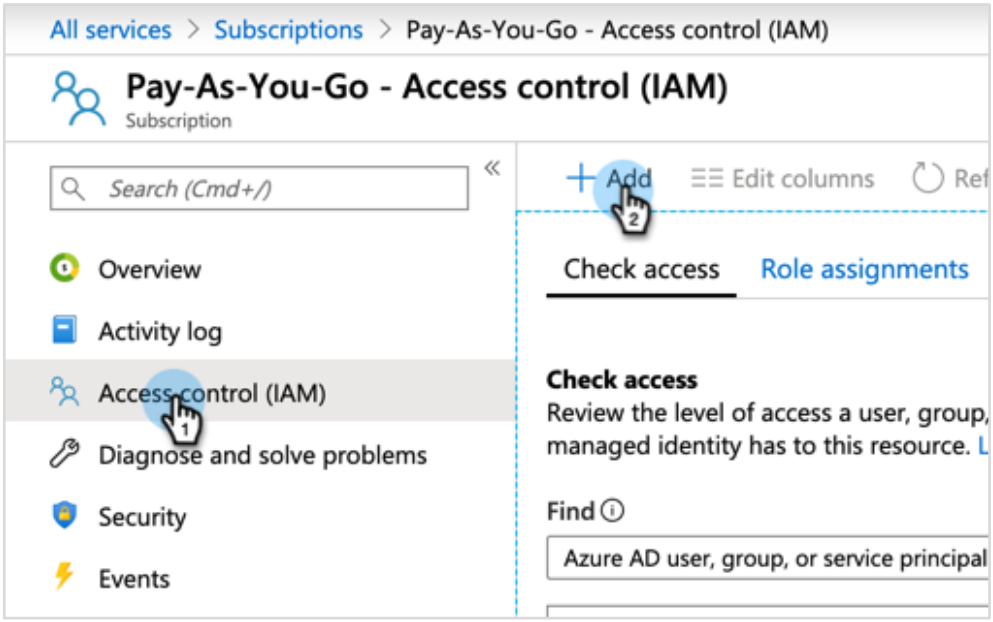
Assign Custom Role to Registered Application for Subscription

The custom role created above specifies the permissions that Cohesity requires to back up Azure VMs. We need to assign this role to the App we created under Register an App on Azure. Hence, when Cohesity uses this App for authorization, it is able to inherit the permissions specified in the custom role.

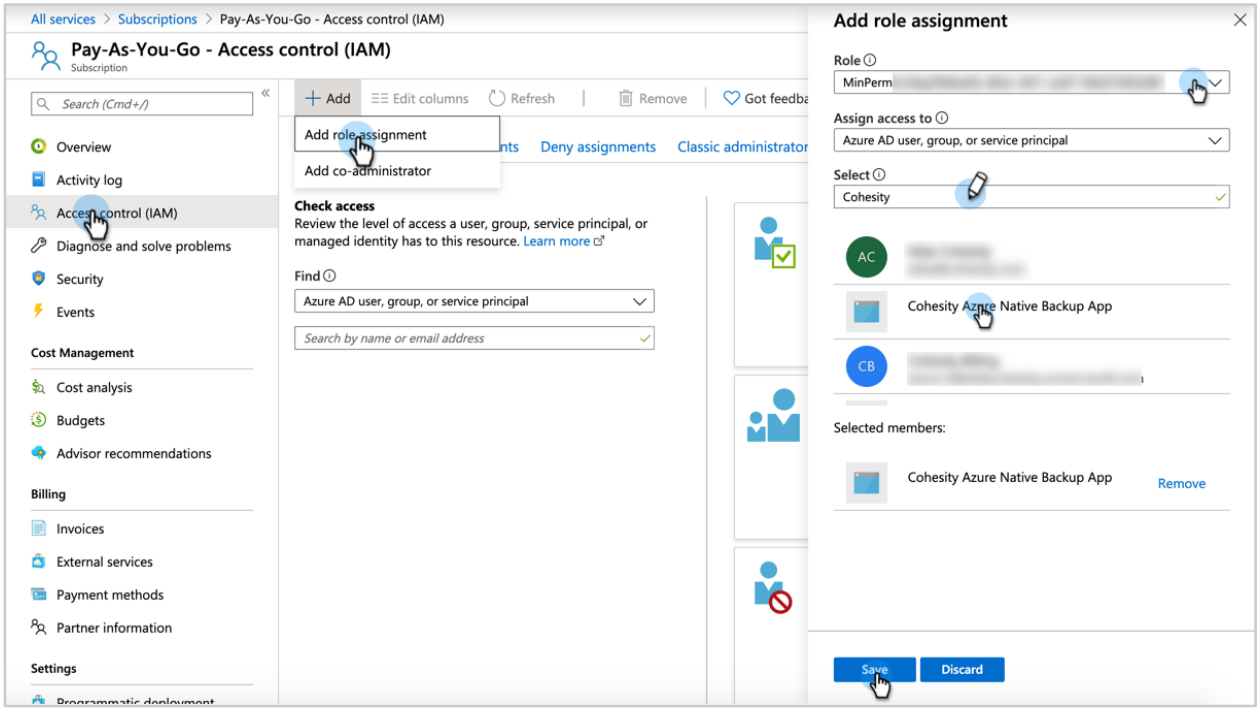
1. In the Azure portal, go to **Subscriptions > Pay-As-you-Go**.



2. Click **Access control (IAM)** and then **Add**.



3. Click **Add role assignment**, select the custom **Role**, and select the App you registered. Click **Save**.



Limitations

Azure Native Snapshotting has the following limitations:

- For Azure VMs created using Azure marketplace images, the restored VM does not support logging in using a username and password present in the backed up VM. Login using SSH keys is supported.
- When the restored VM boots up for the first time, Azure images are pre-installed with cloud-init software that locks user accounts, which is the default behavior.

  TIP: Create an admin user from the Azure portal or log in through the root account if it was enabled in the source VM. Reinstate the user accounts using the following command: `passwd -u username`.

- VMs encrypted through ADE can't boot up after being restored to a different location, unless the user replicates the keys that were used to encrypt the VM into the new location. VMs encrypted using Azure SSE do not have this issue.
- Managed disk VMs that are turned off are shown as 0 bytes in size in the entity hierarchy of Azure Source.
- Backup of Azure VMs that have Ephemeral Volumes is not supported.
- VMs with a static IP will not be recovered back with that static IP.
- Recovery of VMs from an Availability Set to a different location will not inherit the Availability Set parameters, as Availability Set parameters are tied to location.
- Recovery of an unmanaged disk with different SKU types depends on the storage container where the recovery is done.

Your Feedback

Was this document helpful? Send us your feedback!

About the Authors

Saurabh Singh is a Staff Product Manager at Cohesity. In his role, he focuses on product management for Cohesity Data Management as a Service (DMaaS), Security, SaaS Backup, and Secure Multi-tenancy.

Other essential contributors include:

- Nitendra Singh Tomar, Engineering
- Shishir Jindal, Engineering
- Dinesh Pathak, Engineering
- Adaikkappan Arumugam, Sr Manager Technical Marketing & Solution Engineering
- Bart Abicht, Sr. Manager, Technical Marketing & Solution Engineering

Document Version History

VERSION	DATE	DOCUMENT HISTORY
1.0	Nov 2019	First release

ABOUT COHESITY

Cohesity radically simplifies data management. We make it easy to protect, manage, and derive value from data -- across the data center, edge, and cloud. We offer a full suite of services consolidated on one multicloud data platform: backup and recovery, disaster recovery, file and object services, dev/test, and data compliance, security, and analytics -- reducing complexity and eliminating mass data fragmentation. Cohesity can be delivered as a service, self-managed, or provided by a Cohesity-powered partner.

Visit our website and blog, follow us on Twitter and LinkedIn and like us on Facebook.