

# Provision Enterprise Security Package enabled HDInsight - Hadoop Cluster

## Provisioning Sequence, and post installation validations

*Prepared by*

**Data SQL Ninja Engineering Team** ([askdmjfordmtools@microsoft.com](mailto:askdmjfordmtools@microsoft.com))

This document is provided "as-is". Information and views expressed in this document, including URL and other Internet Web site references, may change without notice.

Some examples depicted herein are provided for illustration only and are fictitious. No real association or connection is intended or should be inferred.

This document does not provide you with any legal rights to any intellectual property in any Microsoft product. You may copy and use this document for your internal, reference purposes.

© 2019 Microsoft. All rights reserved.

**Note:** The detail provided in this document has been harvested as part of a customer engagement sponsored through the [Azure Data Services Jumpstart Program](#).

# Table of Contents

1	Introduction .....	8
2	Setup Azure Active Directory – Domain Services (AAD-DS) .....	9
2.1	Create AAD-DS Resource.....	9
2.2	Setup AAD-DS - Scoped Synchronization.....	10
2.3	Generate & Configure Self-signed Certificates.....	12
3	Configure Networking .....	15
3.1	Configure NSG Rules to secure LDAP .....	16
3.2	Configure DNS Zone for External Access .....	18
3.3	Configure DNS Server Settings & Enable AAD-DS Password Hash synchronization .....	20
3.4	Update DNS Server Settings for your Virtual Network.....	20
3.5	Configure NSG Rules to support HDInsight Management & Health services .....	23
4	Configure Access Controls .....	25
4.1	Setup User Assigned Managed Identity .....	25
4.2	Create & Configure <i>hadoopadmins</i> AD Group.....	27
4.3	Assign HDInsight Domain Services Contributor Role.....	28
5	Create & Configure ADLS Gen2 Storage Account.....	31
5.1	Create ADLS Gen2 Storage Account .....	31
5.2	Validate network restrictions.....	36
5.3	Configure Storage Blob Data Owner Role .....	38
5.3.1	Add role to User Assigned Managed Identity .....	38
5.3.2	Add role to <i>hadoopadmins</i> group .....	40
5.4	Add HDInsight Cluster Admin User .....	42
5.5	Additional role configurations for group <i>hadoopadmins</i> .....	47
6	Setup HDInsight ESP enabled cluster.....	49
6.1	Create HDInsight + ESP Resource.....	50
6.2	Enable NSG rules for Ambari UI (port 443).....	62
6.3	Verify Ambari UI HTTPS Port access .....	63

6.4	Setup SSH-Tunneling to access Ambari URLs .....	65
6.5	Validate SSH-Tunnel .....	68
6.6	Known Issues – Web UI .....	69
7	Post Installation Review.....	71
7.1	Ambari Landing Page.....	71
7.2	HDFS Service Summary .....	72
7.3	YARN Summary .....	73
7.4	Ranger .....	73
7.5	Storage explorer – HDInsight Container view .....	75
8	Appendix – Reference URLs .....	76
9	Feedback and suggestions.....	76

# Table of Figures

Figure 1: Enable Azure Active Directory - Domain Services .....	9
Figure 2: Configure AAD-DS Synchronization Scope - Add Groups.....	10
Figure 3: AAD-DS Add Group - Group Selection.....	11
Figure 4: Selected AAD-DS Groups for Synchronization scope.....	12
Figure 5: Azure AD Domain Services - Configure Certificate (PFX file).....	14
Figure 6: Azure AD Domain Services - Secure LDAP Enabled .....	14
Figure 7: Azure AD Domain Services - Secure LDAP Brute-force Warning.....	15
Figure 8: HDInsight Networking Interdependencies Overview.....	15
Figure 9: Azure Active Directory - Domain Services Resource (from the listing).....	16
Figure 10: Azure AD Domain Services NSG Rules (Inbound/Outbound) .....	16
Figure 11: Azure AD Domain Services NSG Rules - Add Inbound Security Rule .....	17
Figure 12: Azure AD Domain Services NSG Rules - Updated .....	17
Figure 13: ldp.exe client UI .....	18
Figure 14: ldp.exe - connect error sample .....	19
Figure 15: Azure AD Domain Services - Required Configuration Steps .....	20
Figure 16: HDInsight Cluster Deployment Failure - DomainNotFoundInActiveDirectory .....	21
Figure 17: HDInsight Deployment - Set of Resources Deployed! .....	21
Figure 18: Azure AD Domain Services - Properties.....	22
Figure 19: Azure AD Domain Services - VNET Properties - DNS IP address updates .....	22
Figure 20: Subscription/All Resources View - AAD-DS NSG Resource .....	23
Figure 21: Configured NSG Rules for HDInsight Management & Health Service IP Addresses .....	24
Figure 22: Search for User Assigned Managed Identity Resource type .....	25
Figure 23: Create User Assigned Managed Identity .....	25
Figure 24: User Assigned Managed Identity - Create Form .....	26
Figure 25: User Assigned Managed Identity - Blade View/Overview .....	26
Figure 26: Azure AD - Groups - All Groups view .....	27
Figure 27: Azure AD Domain Services - Add Group to Synchronization scope .....	28
Figure 28: Updated Azure AD Domain Services - Synchronization Scope .....	28
Figure 29: Assign HDInsight Domain Services Contributor role to User Assigned Managed Identity.....	29
Figure 30: Select Managed Identity for the HDInsight Domain Services Contributor Role assignment.....	30
Figure 31: Updated HDInsight Domain Services Contributor Role Assignment .....	30
Figure 32: List of Storage Accounts .....	31
Figure 33: Create Storage Account - Basics Tab .....	32

Figure 34: Create Storage Account - Networking Tab .....	33
Figure 35: Create Storage Account - Advanced Tab .....	33
Figure 36: Create Storage Account - Review + Create - Validation Passed.....	34
Figure 37: Storage Account Deployment (underway).....	35
Figure 38: Storage Account Deployment Complete.....	35
Figure 39: ADLS Gen2 - Overview.....	36
Figure 40: ADLS Gen2 - Validate Network Restrictions - Sample Error Message.....	37
Figure 41: ADLS Gen2 - Validate Network Restrictions - Unable to Load Data (Sample) .....	37
Figure 42: ADLS Gen2 - Add Assign User Assigned Managed Identity.....	38
Figure 43: ADLS Gen2 - Managed Identity selection .....	39
Figure 44: ADLS Gen2 - Save selected Managed Identity.....	39
Figure 45: ADLS Gen2 - Updated Storage Blob Data Owner Role view .....	40
Figure 46: ADLS Gen2 - Add role to <i>hadoopadmins</i> group.....	40
Figure 47: ADLS Gen2 - Updated Role Assignments View .....	41
Figure 48: Add <i>hadoopadmin</i> user to <i>hadoopadmins</i> group .....	42
Figure 49: Add hadoopadmin user to hadoopadmins group (Contd...) .....	43
Figure 50: Azure Active Directory - All Users view .....	43
Figure 51: Hadoop Administrator User Profile (Azure Active Directory).....	44
Figure 52: Azure Portal - Sign-in with a different account .....	44
Figure 53: Azure Portal Login .....	45
Figure 54: Azure Portal First time login - Update Password .....	46
Figure 55: Azure Portal New User Login - Landing Page .....	46
Figure 56: RBAC Roles Application Overview.....	47
Figure 57: <i>hadoopadmins</i> group assigned roles - Access Control (IAM) overview .....	48
Figure 58: Hadoop Admin User - Storage Explorer View - Folder View Sample .....	48
Figure 59: Create Azure HDInsight - Search .....	50
Figure 60: Create Azure HDInsight Resource .....	50
Figure 61: Create HDInsight + ESP - Choose cluster type .....	51
Figure 62: Create HDInsight + ESP cluster - Details .....	52
Figure 63: Create HDInsight + ESP cluster - Storage Tab .....	53
Figure 64: Create HDInsight + ESP cluster - Security + Networking Tab .....	53
Figure 65: Create HDInsight + ESP cluster - Enable ESP.....	54
Figure 66: Create HDInsight + ESP cluster - Select Cluster Admin User .....	55
Figure 67: Create HDInsight + ESP cluster - Price Estimate .....	57
Figure 68: Create HDInsight + ESP cluster - Configuration Summary .....	58

Figure 69: Create HDInsight + ESP cluster - Deployment Underway .....	59
Figure 70: Create HDInsight + ESP cluster - Deployment Complete.....	60
Figure 71: Create HDInsight + ESP cluster - Deployment Operation Details.....	60
Figure 72: HDInsight + ESP cluster overview .....	61
Figure 73: Ambari UI - Cannot reach the page.....	61
Figure 74: Enable Port 443 on NSG for Ambari UI .....	62
Figure 75: NSG Inbound Rule Updated - Ambari UI - HTTPS port enabled .....	63
Figure 76: Retry accessing Ambari Web UI .....	63
Figure 77: Ambari Dashboard View .....	64
Figure 78: Ambari Web UI - HDFS Summary.....	64
Figure 79: Enable SSH access on NSG.....	65
Figure 80: Firefox - SOCKS5 Configuration.....	67
Figure 81: SOCKS5 Proxy Enabled - Whatismyip.com verification .....	68
Figure 82: Ambari Web UI access using SSH-tunneling (port-forwarding).....	69
Figure 83: HDInsight - Ambari - Namenode UI (Known Issue).....	69
Figure 84: Ambari Web UI - Using SSH-tunnel - HTTP Error 401 .....	70
Figure 85: HDInsight YARN Summary .....	73
Figure 86: HDInsight - Ranger Overview .....	73
Figure 87: HDInsight - Ranger - Hive Overview .....	74
Figure 88: HDInsight - Ranger - Hive Sample Policy.....	75
Figure 89: HDInsight Storage Container View using Storage Explorer for <i>hadoopadmin</i> .....	75

## List of Tables

Table 1: hadoopadmins group RBAC roles Overview .....	47
---	----

# 1 Introduction

Enterprise readiness for Azure HDInsight clusters is supported by means of enabling Enterprise Security Package (ESP) as part of the cluster deployment procedure. This document presents a detailed process of setting up interdependent connected components such as enabling AAD-DS, setting up Storage Account, Configuring Access Controls, through to successful creation and validation of ESP enabled HDInsight cluster. References to additional information and relevant public product documentation are made available in the Appendix section.

Please note that the Azure Active Directory (AAD) tenant used here has the Microsoft controlled domain suffix - \*.onmicrosoft.com. The domain name will be different depending on your AAD setup, where you can choose to have your fully owned routable public DNS Domain Name.

## 2 Setup Azure Active Directory – Domain Services (AAD-DS)

A secure HDInsight cluster will depend on Azure Active Directory – Domain Services (AAD-DS) for identity services. For example – authentication. Once enabled there will be one AAD-DS instance per region, and it runs inside a VM instance that offers DNS services.

### 2.1 Create AAD-DS Resource

Follow the below steps to setup Azure Active Directory – Domain Services resource:

- Select Azure AD Domain Services resource from the marketplace.
- Sample create configuration below (note, the Name attribute value is your AAD-DS domain name, that'd be referenced in all the subsequent instruction set; search for `youraadddomainname` string in the following sections of this notes.):

The screenshot shows the 'Enable Azure AD Domain Services' wizard in the Azure portal. The left sidebar lists five steps: 1. Basics (Completed), 2. Network (Completed), 3. Administrator group (Completed), 4. Synchronization (Completed), and 5. Summary (Current Step). The main summary table contains the following data:

Category	Setting	Value
Basics	Name	[REDACTED].onmicrosoft.com
	Subscription	Visual Studio Enterprise
	Resource group	bniexperiments
	Location	(US) West US
Network	Virtual network	bnimvnet
	Virtual network address	10.0.0.0/16
	Subnet	default
	Subnet Address	10.0.0.0/24
Administrator group	Network security group (new)	AADDS-[REDACTED].onmicrosoft.com-NSG
	Administrator group	AAD DC Administrators
Synchronization	Membership Type	Assigned
	Synchronization scope	Scoped

At the bottom, a note states: "By enabling Azure AD Domain Services for this directory, you consent to storing credential hashes required for NTLM and Kerberos authentication in Azure AD." An "OK" button is at the bottom right.

Figure 1: Enable Azure Active Directory - Domain Services

## 2.2 Setup AAD-DS - Scoped Synchronization

AAD-DS synchronizes its identity database with Azure AD. You can choose to synchronize just about everything (the default process) in the AD or choose scoped synchronization. Difference is that if you opt for the defaults the synchronization process may potentially consume significant amount of time. Rather if you are only interested in certain groups and users you can limit the synchronization to just to those entities.

Please note that Users and User Groups that you desire to have access to your HDInsight cluster, must be selected as part of this AAD-DS synchronization process. Else, HDInsight cluster deployment will fail with terminal state error – GroupNotFoundInActiveDirectory. Following is a sample error message for immediate reference:

```
{"\"status\":\"Failed\", \"error\": {\"code\":\"ResourceDeploymentFailure\", \"message\":\"The resource operation completed with terminal provisioning state 'Failed'.\", \"details\":[{\"code\":\"GroupNotFoundInActiveDirectory\", \"message\":\"Group IngressUsers not found in the Active Directory.\"]}}}"
```

Following visual provides a reference to the portal pane from where additional users and user-groups can be added to the AAD-DS synchronization scope:

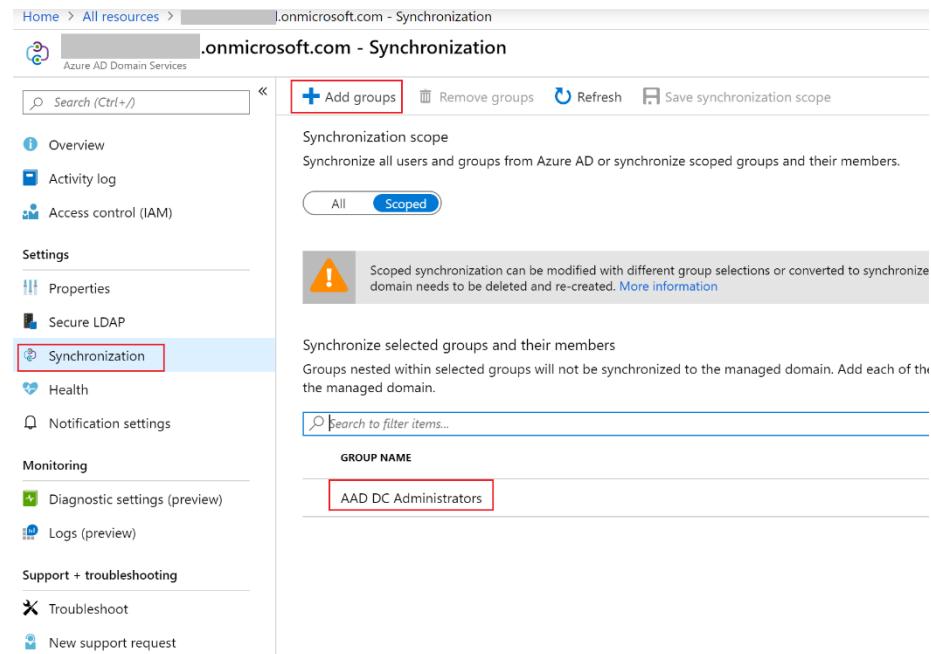


Figure 2: Configure AAD-DS Synchronization Scope - Add Groups

On the similar screen of yours, click on the + Add groups link as highlighted in the visual above.

Choose the groups and or users that must be granted the access, as shown in the following reference visual:

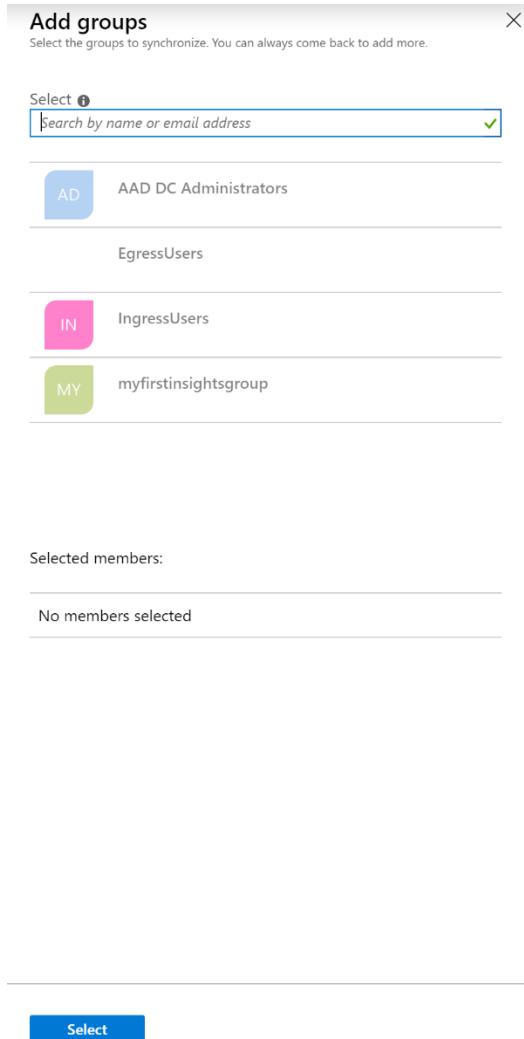


Figure 3: AAD-DS Add Group - Group Selection

Once selected, saved the selections will get added to the synchronization scope as shown below:

The screenshot shows the 'Synchronization scope' configuration page in the Azure portal. On the left, a navigation sidebar includes 'Overview', 'Activity log', 'Access control (IAM)', 'Properties', 'Secure LDAP', **Synchronization**, 'Health', 'Notification settings', 'Diagnostic settings (preview)', 'Logs (preview)', 'Troubleshoot', and 'New support request'. The 'Synchronization' option is highlighted. The main area has a header with 'Add groups', 'Remove groups', 'Refresh', and 'Save synchronization scope' buttons. Below the header, it says 'Synchronization scope' and 'Synchronize all users and groups from Azure AD or synchronize scoped groups and their members.' A 'Scoped' tab is selected. A warning message states: 'Scoped synchronization can be modified with different group selections or converted to synchronize all users and domain needs to be deleted and re-created. [More information](#)'. Below this, it says 'Synchronize selected groups and their members' and 'Groups nested within selected groups will not be synchronized to the managed domain. Add each of the nested groups to the managed domain.' A search bar 'Search to filter items...' is present. A table titled 'GROUP NAME' lists four groups: 'AAD DC Administrators', 'EgressUsers', 'IngressUsers', and 'myfirstinsightsgroup'.

Figure 4: Selected AAD-DS Groups for Synchronization scope

## 2.3 Generate & Configure Self-signed Certificates

- First step is to define a self-signed certificate using the powershell command below:

```
$lifetime=Get-Date  
New-SelfSignedCertificate -Subject youraaddsdomainname.onmicrosoft.com `  
-NotAfter $lifetime.AddDays(365) -KeyUsage DigitalSignature,  
KeyEncipherment `  
-Type SSLServerAuthentication -DnsName  
*.youraaddsdomainname.onmicrosoft.com, youraaddsdomainname.onmicrosoft.com
```

- Warning note about how you'd launch Powershell, if you run as a normal user i.e., non-administrator and attempt to run above command, you'd end up with following error:

```

PS C:\Users\yourid> $lifetime=Get-Date
PS C:\Users\yourid> New-SelfSignedCertificate -Subject
youraddsdomainname.onmicrosoft.com `
>> -NotAfter $lifetime.AddDays(365) -KeyUsage DigitalSignature,
KeyEncipherment `
>> -Type SSLServerAuthentication -DnsName
*.youraddsdomainname.onmicrosoft.com, youraddsdomainname.onmicrosoft.com
New-SelfSignedCertificate : CertEnroll::CX509Enrollment::_CreateRequest:
Access denied. 0x80090010 (-2146893808
NTE_PERM)
At line:1 char:1
+ New-SelfSignedCertificate -Subject youraddsdomainname.onmicrosoft.com `
+ ~~~~~
+ CategoryInfo          : NotSpecified: (:) [New-SelfSignedCertificate],
Exception
+ FullyQualifiedErrorId :
System.Exception,Microsoft.CertificateServices.Commands.NewSelfSignedCertificateCommand

```

- Launch Powershell with the option Run as Administrator. Here's the sample command response on the console output:

```

PS C:\WINDOWS\system32> $lifetime=Get-Date
PS C:\WINDOWS\system32> New-SelfSignedCertificate -Subject
youraddsdomainname.onmicrosoft.com `
>> -NotAfter $lifetime.AddDays(365) -KeyUsage DigitalSignature,
KeyEncipherment `
>> -Type SSLServerAuthentication -DnsName
*.youraddsdomainname.onmicrosoft.com, youraddsdomainname.onmicrosoft.com

PSParentPath: Microsoft.PowerShell.Security\Certificate::LocalMachine\MY

Thumbprint                               Subject
-----                                 -----
3F6F8XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX   CN=youraddsdomainname.onmicrosoft.com

```

- Subsequently, follow the instructions from [here](#) to export the certificate files – PFX and CER. While you export the certificate (PFX) you'd be prompted to password protect the file. Please take a note of the password. It will be used when configuring the same in the portal.
  - Complete necessary steps through to exporting the certificate file (.PFX extension) for Azure Active Directory Domain Services and .CER file for client computers.
- Upload and configure the PFX file to the Secure LDAP settings as shown in the following visual. Click [here](#) for official documentation.

Figure 5: Azure AD Domain Services - Configure Certificate (PFX file)

- Pay attention to those in red boxes. Look forward to section on configuring NSG rules further down in this article. For now, hit Save to secure LDAP.
- Once enabled... here's what it should look like

Figure 6: Azure AD Domain Services - Secure LDAP Enabled

- As a cautionary measure, verify the health of AAD-DS. Note the warning below. It will go away once you configure the NSG rules in the next section.

The screenshot shows the Azure AD Domain Services - Health page. The left sidebar has a 'Health' section selected. The main area displays a warning message: 'The managed domain may be vulnerable to password brute-force attacks' with a severity of 'Warning'. Other sections like Monitors and Alerts are also visible.

Figure 7: Azure AD Domain Services - Secure LDAP Brute-force Warning

## 3 Configure Networking

Following visual offers networking inter-dependencies between various moving parts of the HDInsight cluster. It helps in planning the HDInsight virtual network configurations. Click [here](#) to access the official documentation reference.

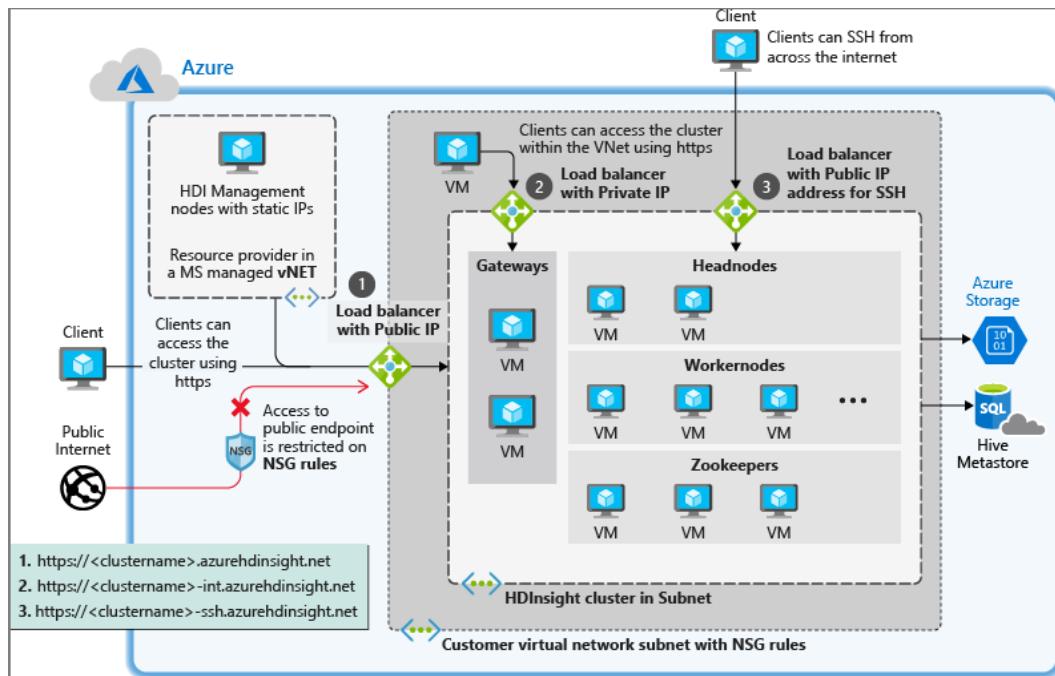


Figure 8: HDInsight Networking Interdependencies Overview

### 3.1 Configure NSG Rules to secure LDAP

To lock down and secure LDAP access over the Internet, in this section we will go through the sequence of steps to configure NSG rules and secure access to LDAP. Once these steps are successfully completed, users seeking access to your LDAP services on port 636 will be enabled. The warning notification that is displayed at the time of certification configuration will be gone. Follow the sequence below:

- Go to NSG resource from the Resource Group's resources list. Naming convention for the NSG resource will have a prefix AADS and suffix NSG as shown in the visual below:

	aadds-[REDACTED]-nic	Network interface	West US	...
	aadds-[REDACTED]-nic	Network interface	West US	...
	aadds-[REDACTED]-lb	Load balancer	West US	...
	aadds-[REDACTED]-pip	Public IP address	West US	...
	AADDS-[REDACTED].onmicrosoft.com-NSG	Network security group	West US	...
	bnmhdiadlsgen	Storage account	West US	...
	bnmmanagedidentity	Managed Identity	West US	...
	bnmvnet	Virtual network	West US	...
	bnmstudioadlsgen2	Storage account	West US	...
	Azure AD Domain Services	Azure AD Domain Services	West US	...

Figure 9: Azure Active Directory - Domain Services Resource (from the listing)

- Select the NSG resource. Here's the reference visual of the overview section:

The screenshot shows the Azure portal's NSG overview page for the 'AADDS-[REDACTED]-onmicrosoft.com-NSG' resource. The left sidebar includes links for Overview, Activity log, Access control (IAM), Tags, and Diagnose and solve problems. The main content area shows the resource group (bnmexperiments), location (West US), and associated subnets. It lists three inbound security rules and three outbound security rules, both prioritized by port and protocol.

PRIORITY	NAME	PORT	PROTOCOL	SOURCE	DESTINATION	ACTION
101	AllowSyncWithAzureAD	443	TCP	AzureActiveDirecto...	Any	Allow
201	AllowRD	3389	TCP	CorpNetSaw	Any	Allow
301	AllowPSRemoting	5986	TCP	AzureActiveDirecto...	Any	Allow
65000	AllowVnetInBound	Any	Any	VirtualNetwork	VirtualNetwork	Allow
65001	AllowAzureLoadBalancerInBound	Any	Any	AzureLoadBalancer	Any	Allow
65500	DenyAllInBound	Any	Any	Any	Any	Deny

PRIORITY	NAME	PORT	PROTOCOL	SOURCE	DESTINATION	ACTION
65000	AllowVnetOutBound	Any	Any	VirtualNetwork	VirtualNetwork	Allow
65001	AllowInternetOutBound	Any	Any	Any	Internet	Allow
65500	DenyAllOutBound	Any	Any	Any	Any	Deny

Figure 10: Azure AD Domain Services NSG Rules (Inbound/Outbound)

- Add Inbound security rule to enable traffic on port 636 for secure LDAP

The screenshot shows the Azure portal interface for managing Network Security Groups (NSGs). On the left, the navigation pane is visible with options like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Settings, Inbound security rules (selected), Outbound security rules, Network interfaces, Subnets, Properties, Locks, Export template, Monitoring, Diagnostic settings, Logs, and NSG flow logs. The main area displays the 'onmicrosoft.com-NSG - Inbound security rules' page, which lists several existing rules:

PRIORITY	NAME	PORT
101	AllowSyncWithAzureAD	443
201	AllowRD	3389
301	AllowPSRemoting	5986
65000	AllowVnetInBound	Any
65001	AllowAzureLoadBalancerInBound	Any
65500	DenyAllInBound	Any

To the right, a modal window titled 'Add inbound security rule' is open, showing the configuration for a new rule:

- Source:** IP Addresses: 10.0.0.0/24
- Source port ranges:** Any
- Destination:** Any
- Destination port ranges:** 636
- Protocol:** TCP
- Action:** Allow
- Priority:** 401
- Name:** AllowLDAPS
- Description:** Setup as part of HDInsight setup

A large 'Add' button is at the bottom of the modal.

Figure 11: Azure AD Domain Services NSG Rules - Add Inbound Security Rule

- Once you successfully have added the rule, the rules list should appear like below

The screenshot shows the 'onmicrosoft.com-NSG' page after the new rule has been added. The 'Inbound security rules' section now includes the 'AllowLDAPS' rule:

PRIORITY	NAME	PORT	PROTOCOL	SOURCE	DESTINATION	ACTION
101	AllowSyncWithAzureAD	443	TCP	AzureActiveDirecto...	Any	Allow
201	AllowRD	3389	TCP	CorpNetSaw	Any	Allow
301	AllowPSRemoting	5986	TCP	AzureActiveDirecto...	Any	Allow
401	AllowLDAPS	636	TCP	10.0.0.0/24	Any	Allow
65000	AllowVnetInBound	Any	Any	VirtualNetwork	VirtualNetwork	Allow
65001	AllowAzureLoadBalancerInBound	Any	Any	AzureLoadBalancer	Any	Allow
65500	DenyAllInBound	Any	Any	Any	Any	Deny

The 'Outbound security rules' section also contains a few rules:

PRIORITY	NAME	PORT	PROTOCOL	SOURCE	DESTINATION	ACTION
65000	AllowVnetOutBound	Any	Any	VirtualNetwork	VirtualNetwork	Allow
65001	AllowInternetOutBound	Any	Any	Any	Internet	Allow
65500	DenyAllOutBound	Any	Any	Any	Any	Deny

Figure 12: Azure AD Domain Services NSG Rules - Updated

## 3.2 Configure DNS Zone for External Access

- DNS Zone entry must be configured on the client resource that intends to interact with the AAD-DS resource. [Here](#)'s the link to the official documentation.
- Configure your external DNS provider to create a host record, such as ldaps, to resolve to the secure external IP address available from the AAD-DS properties blade entry - Secure LDAP external IP address. In this case the IP address resolves to xxx.xxx.xxx.xxx.

xxx.xxx.xxx.xxx	youraddsdomainname.onmicrosoft.com
-----------------	------------------------------------

- Test Queries for the Managed Domain (Note - Work in progress sequence below; needs further NSG configurations to allow remote ldp.exe to be able to reach our LDAPS on the AAD-DS):
  - Install the Remote Server Administrator tools on your local windows machine, following the instruction detail from [here](#).
  - Launch Windows Power-shell as an administrator (i.e., Run as Administrator).
  - And, execute the ldp.exe command as shown in the following visual:

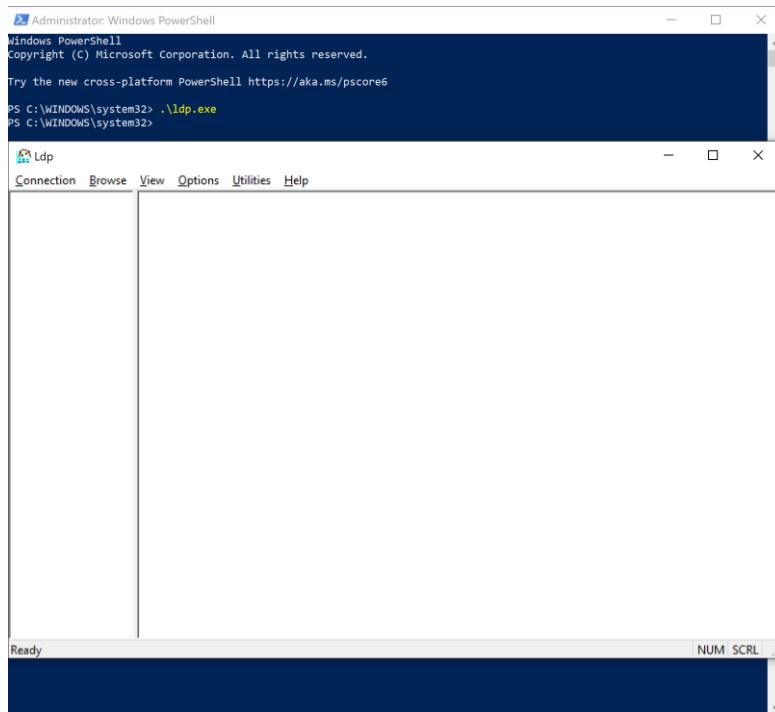
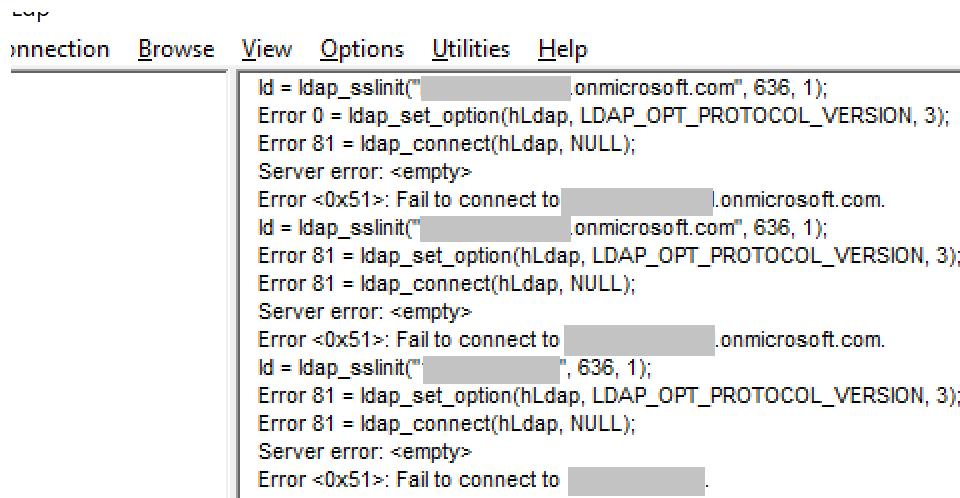


Figure 13: ldp.exe client UI

- It is possible that you'd encounter similar error when attempting to connect and bind with the DNS server using the external IP address as found in the AAD-DS properties.



The screenshot shows a command-line interface for the ldp.exe tool. The menu bar includes 'Connection', 'Browse', 'View', 'Options', 'Utilities', and 'Help'. The main window displays a series of LDAP API calls and their results:

```
Id = ldap_sslinit("[REDACTED].onmicrosoft.com", 636, 1);
Error 0 = ldap_set_option(hLdap, LDAP_OPT_PROTOCOL_VERSION, 3);
Error 81 = ldap_connect(hLdap, NULL);
Server error: <empty>
Error <0x51>: Fail to connect to [REDACTED].onmicrosoft.com.
Id = ldap_sslinit("[REDACTED].onmicrosoft.com", 636, 1);
Error 81 = ldap_set_option(hLdap, LDAP_OPT_PROTOCOL_VERSION, 3);
Error 81 = ldap_connect(hLdap, NULL);
Server error: <empty>
Error <0x51>: Fail to connect to [REDACTED].onmicrosoft.com.
Id = ldap_sslinit("[REDACTED]", 636, 1);
Error 81 = ldap_set_option(hLdap, LDAP_OPT_PROTOCOL_VERSION, 3);
Error 81 = ldap_connect(hLdap, NULL);
Server error: <empty>
Error <0x51>: Fail to connect to [REDACTED].
```

Figure 14: ldp.exe - connect error sample

- Detailed instructions of how to accomplish the steps is accessible from the official documentation section [here](#).

## 3.3 Configure DNS Server Settings & Enable AAD-DS Password Hash synchronization

For password has synchronization click [here](#) for the official documentation. This part of the process is essential for hybrid environment setups and is out of scope for this document.

The screenshot shows the Azure portal interface for managing Azure AD Domain Services. The left sidebar contains navigation links for Home, Resource groups, and the current domain (.onmicrosoft.com). The main content area is titled 'Required configuration steps' and lists two tasks:

- Update DNS server settings for your virtual network**: Describes updating DNS settings to point to IP addresses 10.0.0.4 and 10.0.0.5. It includes a 'Configure' button.
- Enable Azure AD Domain Services password hash synchronization**: Describes enabling synchronization for managed domains. It includes two links: 'Instructions for cloud-only user accounts' and 'Instructions for synced user accounts'.

At the bottom, there is a 'Related links' section with three items: Join a Windows Server VM to your managed domain, Join a Windows Server VM to your managed domain using PowerShell, and Administer your managed domain.

Figure 15: Azure AD Domain Services - Required Configuration Steps

## 3.4 Update DNS Server Settings for your Virtual Network

Getting back to the DNS configuration step, it is important to update the IP addresses as shown under the *Required configuration next steps* section (see the previous visual above). This configuration will allow dependent services like HDInsight Ranger to have a line of sight to your directory services. And such access is verified during the cluster setup process. Failure to do so will result in a terminal provisioning state failure error. Following is the sample of such error message:

```
{
  "code": "DeploymentFailed",
  "message": "At least one resource deployment operation failed. Please list deployment operations for details. Please see https://aka.ms/arm-debug for usage details.",
  "details": [
    {
      "code": "Conflict",
      "message": "\r\n \"status\": \"Failed\", \r\n \"error\": \r\n \r\n \"code\": \"ResourceDeploymentFailure\", \r\n \"message\": \"The resource operation completed with terminal provisioning state 'Failed'.\", \r\n \"details\": [ \r\n \r\n \\"Domain\\youradddsdomainname.onmicrosoft.com not found in the Active Directory.\r\n ]\r\n ]\r\n "
    }
  ]
}
```

Here's the visual reference that details the deployment status:

The screenshot shows the Microsoft Azure portal with the URL [https://portal.azure.com/#blade/Microsoft\\_Azure\\_Templates/TemplateOverviewBlade/resource/Microsoft.Template](#). The page displays a deployment named 'Microsoft.Template' with the status 'Your deployment failed'. The error message states: 'The resource operation completed with terminal provisioning state 'Failed'. Click here for details →'. The error details pane shows the same message: 'The resource operation completed with terminal provisioning state 'Failed'. (Code: ResourceDeploymentFailure)'. It also lists a specific error: 'Domain \\youradddsdomainname.onmicrosoft.com not found in the Active Directory. (Code: DomainNotFoundInActiveDirectory)'. Below the error details, there are troubleshooting options like 'Common Azure deployment errors', 'Check Usage + Quota', and 'New Support Request'.

Figure 16: HDInsight Cluster Deployment Failure - DomainNotFoundInActiveDirectory

Following visual shows the inconsistent state of your cluster deployment process, where the necessary resources are deployed, but are not usable due to above failure condition:

The screenshot shows the Microsoft Azure portal with the URL [https://portal.azure.com/#blade/Microsoft\\_Azure\\_Resources/ResourcesBlade/resourceType/microsoft.resources/resourceGroups](#). The page displays a list of 35 resources under the 'All resources' section. Most resources are listed as 'Not Usable' (indicated by a yellow exclamation mark icon). These include various network interfaces, storage accounts, load balancers, and managed identities. A few resources are marked as 'Usable' (indicated by a green checkmark icon), such as the 'Azure AD Domain Services' and some virtual networks. The resources are grouped by subscription, all of which are 'Visual Studio Enterprise'.

Figure 17: HDInsight Deployment - Set of Resources Deployed!

Remediation steps or the steps that must be performed as part of AAD-DS setup sequence are as follows:

- In the AAD-DS resource properties select the *Available in virtual network/subnet* link as shown below:

The screenshot shows the 'Properties' page for a domain service. On the left, there's a navigation pane with links like Overview, Activity log, Access control (IAM), Settings (Properties is selected), Secure LDAP, Synchronization, Health, Notification settings, Monitoring (Diagnostic settings (preview), Logs (preview)), Support + troubleshooting (Troubleshoot, New support request), and Help & feedback. The main content area has sections for DNS domain name (.onmicrosoft.com), Location (West US), Forest type (User), and Available in virtual network/subnet (bnmvnet/default). Other visible fields include Network security group associated with subnet (AADDSS-.onmicrosoft.com-NSG), IP address on virtual network (10.0.0.4 10.0.0.5), Secure LDAP (Enabled), Secure LDAP certificate thumbprint (redacted), Secure LDAP certificate expires (Thu, 24 Sep 2020 21:53:57 GMT), Secure LDAP external IP address (redacted), and Synchronization.

Figure 18: Azure AD Domain Services - Properties

- As shown in the following visual reference, instead of Azure-provided choice, select Custom and enter both the IP addresses from the previous step.

The screenshot shows the 'DNS servers' configuration page for a virtual network. The left sidebar includes links for Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Address space, Connected devices, Subnets, DDoS protection, Firewall, and Security. The main panel shows the 'DNS servers' section with two options: 'Default (Azure-provided)' (radio button) and 'Custom' (radio button, which is selected). Below this, two IP addresses are listed: 10.0.0.4 and 10.0.0.5, each with an ellipsis button. At the bottom, there's a 'Save' button, a 'Discard' button, and a 'Add DNS server' input field.

Figure 19: Azure AD Domain Services - VNET Properties - DNS IP address updates

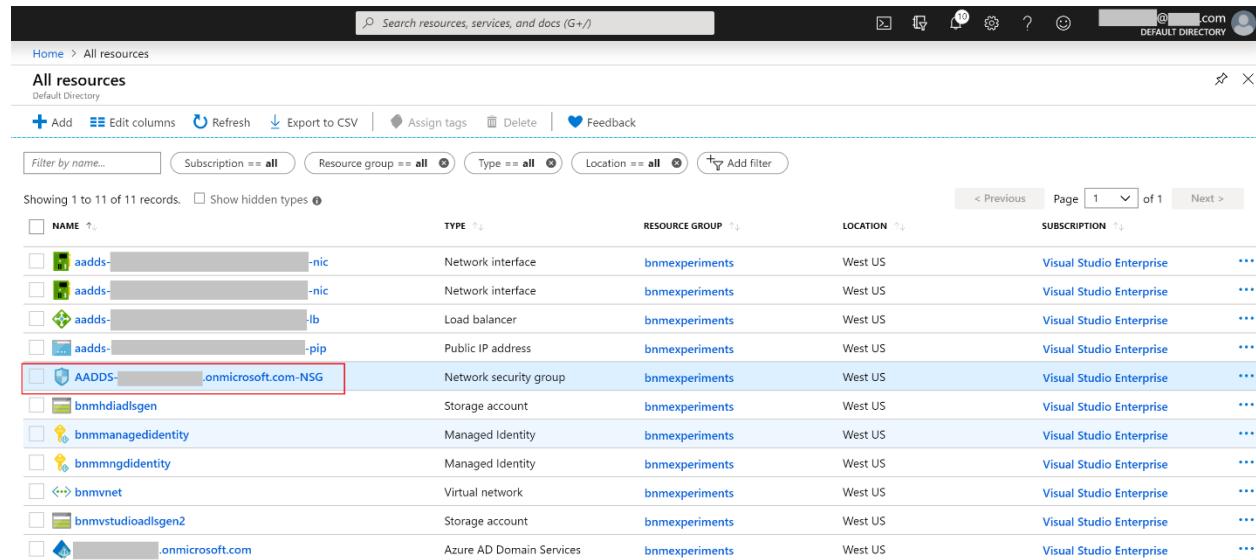
## 3.5 Configure NSG Rules to support HDInsight Management & Health services

By default, access to HDInsight cluster components & services is disabled. Based on your requirements and requirements of the Azure management and health services certain IP and Ports must be enabled and configured using the NSG rules. In this section will go through the process of enabling essential NSG rules.

It is important to note that configurations applied here assume that you are going to leverage Azure-provided DNS service. Click [here](#) to access more on this topic from the official documentation. In this section will enable Azure Management and Health service IP address and ports on the NSG rules. *Skipping this step will result in cluster deployment failures.* Once the cluster setup is complete, we can come back and configure a rule to enable traffic on destination port 443 within the VNET IP address range (CIDR). This will enable secure HTTP access to cluster services. Click [here](#) to learn more about ports used by Apache Hadoop services on HDInsight).

As a next step follow the sequence below:

- In the Azure portal, navigate to the NSG resource. You can either use All Resources link in the left navigation pane of the portal or go to your resource group and select the NSG resource. Here's an example that uses the all resources view:



NAME	TYPE	RESOURCE GROUP	LOCATION	SUBSCRIPTION
aadds-[REDACTED]-nic	Network interface	bnmexperiments	West US	Visual Studio Enterprise
aadds-[REDACTED]-nic	Network interface	bnmexperiments	West US	Visual Studio Enterprise
aadds-[REDACTED]-lb	Load balancer	bnmexperiments	West US	Visual Studio Enterprise
aadds-[REDACTED]-pip	Public IP address	bnmexperiments	West US	Visual Studio Enterprise
AADDS-[REDACTED].onmicrosoft.com-NSG	Network security group	bnmexperiments	West US	Visual Studio Enterprise
bnmhdiadlsgen	Storage account	bnmexperiments	West US	Visual Studio Enterprise
bnmmanagedidentity	Managed identity	bnmexperiments	West US	Visual Studio Enterprise
bnmmngdidentity	Managed identity	bnmexperiments	West US	Visual Studio Enterprise
bnmvnet	Virtual network	bnmexperiments	West US	Visual Studio Enterprise
bnmvstudioadlsgen2	Storage account	bnmexperiments	West US	Visual Studio Enterprise
[REDACTED].onmicrosoft.com	Azure AD Domain Services	bnmexperiments	West US	Visual Studio Enterprise

Figure 20: Subscription/All Resources View - AAD-DS NSG Resource

- You can reference the required HDInsight management and health service IP addresses and ports from the official documentation link [here](#).

- You must configure IP address and ports that apply to all regions as well as those that are specific to a certain region where your HDInsight resource will be deployed.
- Following visual provides a reference example of all the necessary addresses and ports that are enabled in the NSG inbound rule records:

The screenshot shows the Azure portal interface for managing Network Security Groups (NSGs). The left sidebar navigation includes Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Settings (with Inbound security rules selected), Outbound security rules, Network interfaces, Subnets, Properties, Locks, Export template, Monitoring, Diagnostic settings, Logs, NSG flow logs, Support + troubleshooting, Effective security rules, and New support request.

The main content area displays the 'Inbound security rules' for the 'AADDS-1.onmicrosoft.com-NSG'. The table lists the following rules:

PRIORITY	NAME	PORT	PROTOCOL	SOURCE	DESTINATION	ACTION
101	AllowSyncWithAzureAD	443	TCP	AzureActiveDirecto...	Any	<span style="color: green;">Allow</span>
201	AllowRD	3389	TCP	CorpNetSaw	Any	<span style="color: green;">Allow</span>
301	AllowPSRemoting	5986	TCP	AzureActiveDirecto...	Any	<span style="color: green;">Allow</span>
401	AllowLDAPS	636	TCP	10.0.0/24	Any	<span style="color: green;">Allow</span>
411	EndClientLDAP_636	636	Any	255.255.0.0/17	Any	<span style="color: green;">Allow</span>
421	HDInsight-168.61.49.99-443	443	Any	168.61.49.99	Any	<span style="color: green;">Allow</span>
431	HDInsight-23.99.5.239-443	443	Any	23.99.5.239	Any	<span style="color: green;">Allow</span>
441	HDInsight-168.61.48.131-443	443	Any	168.61.48.131	Any	<span style="color: green;">Allow</span>
451	HDInsight-138.91.141.162-443	443	Any	138.91.141.162	Any	<span style="color: green;">Allow</span>
461	HDInsight-13.64.254.98-443-HealthAndMgmtRegionSpecific	443	Any	13.64.254.98	Any	<span style="color: green;">Allow</span>
471	HDInsight-23.101.196.19-443-HealthAndMgmtRegionSpecific	443	Any	23.101.196.19	Any	<span style="color: green;">Allow</span>
481	HDInsight-168.63.129.16-53-AzureDNSService	53	Any	168.63.129.16	Any	<span style="color: green;">Allow</span>
65000	AllowVnetInBound	Any	Any	VirtualNetwork	VirtualNetwork	<span style="color: green;">Allow</span>
65001	AllowAzureLoadBalancerInBound	Any	Any	AzureLoadBalancer	Any	<span style="color: green;">Allow</span>
65500	DenyAllInBound	Any	Any	Any	Any	<span style="color: red;">Deny</span>

Figure 21: Configured NSG Rules for HDInsight Management & Health Service IP Addresses

## 4 Configure Access Controls

### 4.1 Setup User Assigned Managed Identity

User assigned managed identity that we create here will be configured as part of the HDI cluster setup. It will allow HDI services to authenticate and authorize with the ADLS Gen2 storage account, and offer HDFS service to its own components like Hive, etc., as well to the users who'd like to interact with HDI cluster's HDFS service. There are caveats to watch when interacting with HDFS, in comparison to how open source would work. Here, all storage interactions are delegated to the ADLS Gen2 service. Hence, configurations like authentication and authorization must be provisioned following ADLS Gen2 recommended methods. See more about ADLS Gen2 [here](#).

Follow the below sequence to create a User Assigned Managed Identity resource:

- From the portal, click on + Create Resource link on the left navigation and search based on the string - User Assigned Managed Identity, as in the following visual:

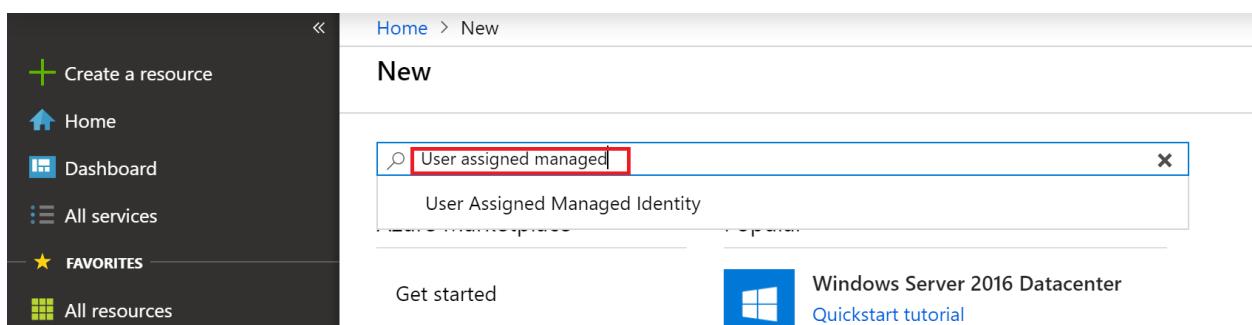


Figure 22: Search for User Assigned Managed Identity Resource type

- In the subsequent view, create the resource by clicking on the Create button.

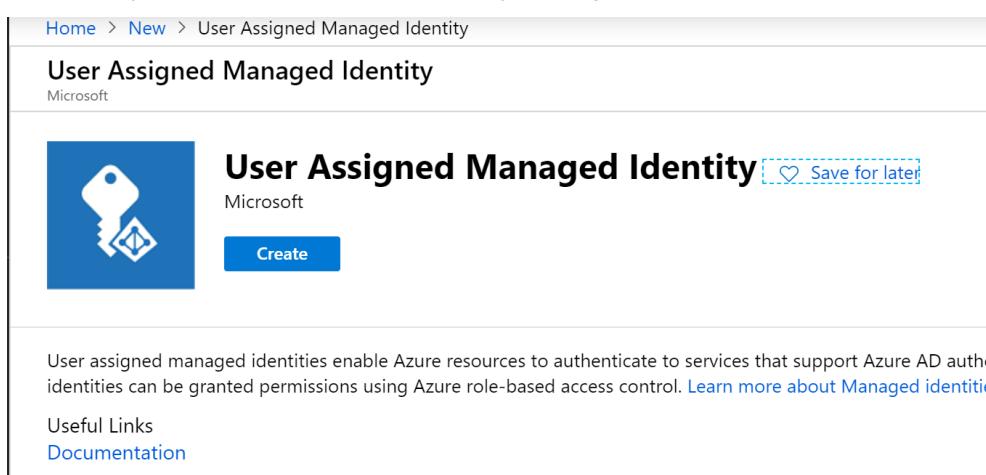


Figure 23: Create User Assigned Managed Identity

- Then, input all the necessary configuration detail as shown in the following visual:

The screenshot shows the 'Create user assigned managed identity' form in the Azure portal. The 'Resource Name' field contains 'bnmmngdentity'. The 'Subscription' dropdown is set to 'Visual Studio Enterprise'. The 'Resource group' dropdown is set to 'bnmexperiments'. The 'Location' dropdown is set to 'West US'. A red box highlights the 'Create' button at the bottom left of the form.

Figure 24: User Assigned Managed Identity - Create Form

- Note - for illustration, I have used a slightly different resource name. In the subsequent section about ADLS Gen2, the name of the resource that is used (and created for the purpose following same instructions) will be *bnmmanagedidentity*.
- You can verify the details of the managed identity from the all resources pane, as shown in the following visual:

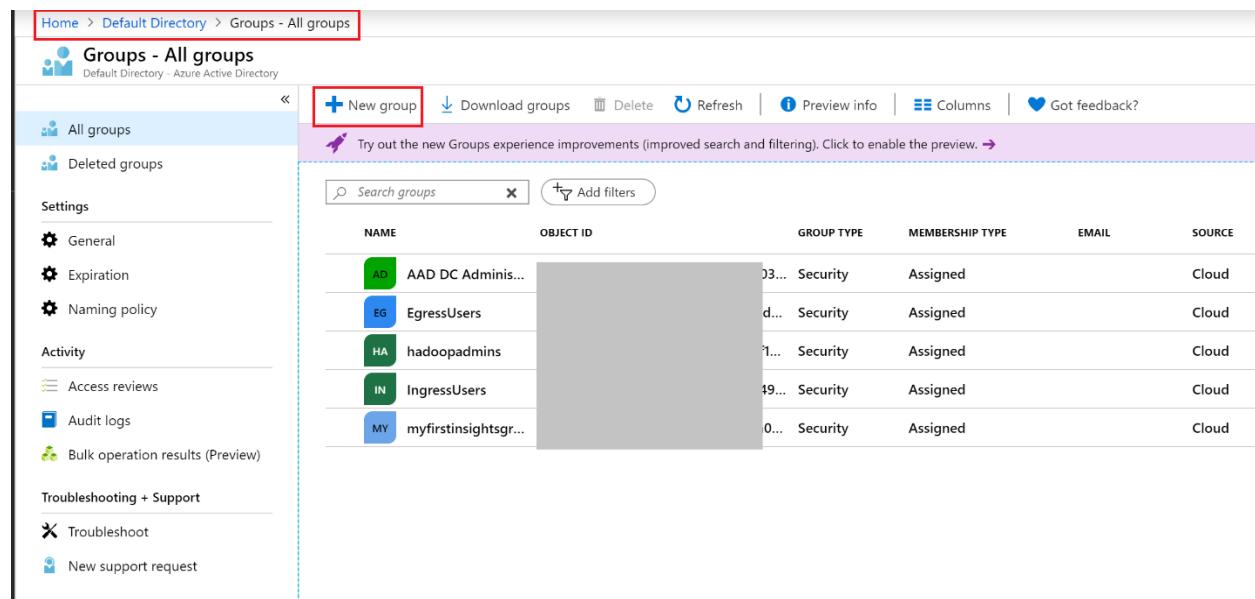
The screenshot shows the blade view for the 'bnmmngdentity' managed identity. The 'Resource group' is 'bnmexperiments', 'Location' is 'West US', 'Subscription' is 'Visual Studio Enterprise', and 'Type' is 'User assigned managed identity'. Other details shown include 'Client ID' and 'Object ID'.

Figure 25: User Assigned Managed Identity - Blade View/Overview

## 4.2 Create & Configure *hadoopadmins* AD Group

While HDInsight depends on a managed identity to represent it against storage layer access, we need similar administration handle to administer user level access controls to the storage. For example, in order for Hadoop users to be able to successfully query data, besides granting them permission to the respective tables, they should also be granted access to the storage path where the table data is maintained and managed. Hence, create an Azure Directory group called *hadoopadmins* using the Azure Active Directory pane for Groups.

Following visual provides a reference to the AD Groups pane (note, the default directory here is the name of the Active Directory):



The screenshot shows the 'Groups - All groups' page in the Azure Active Directory portal. The left sidebar includes links for 'All groups', 'Deleted groups', 'Settings' (General, Expiration, Naming policy), 'Activity' (Access reviews, Audit logs), and 'Troubleshooting + Support' (Troubleshoot, New support request). The main area displays a table of groups with columns: NAME, OBJECT ID, GROUP TYPE, MEMBERSHIP TYPE, EMAIL, and SOURCE. A new group named 'hadoopadmins' is highlighted in the table. The top navigation bar shows the path 'Home > Default Directory > Groups - All groups'. A red box highlights the '+ New group' button in the top navigation bar.

NAME	OBJECT ID	GROUP TYPE	MEMBERSHIP TYPE	EMAIL	SOURCE
AD	AAD DC Adminis...	03...	Security	Assigned	Cloud
EG	EgressUsers	d...	Security	Assigned	Cloud
HA	hadoopadmins	1...	Security	Assigned	Cloud
IN	IngressUsers	49...	Security	Assigned	Cloud
MY	myfirstinsightsgr...	0...	Security	Assigned	Cloud

Figure 26: Azure AD - Groups - All Groups view

As shown in the sample reference above, use the *New group* link to add the group *hadoopadmins*. In subsequent sections will add this group to the storage account with a role that allows the members of the group to administer access control to other users.

Next, we need to ensure the group and its members from Azure Active Directory are also synchronized to the Azure Active Directory – Domain Services (AAD-DS). To satisfy this requirement it is required to add the group *hadoopadmins* to the AD Synchronization scope. Navigate to the AAD-DS Synchronization pane and add the group as shown below:

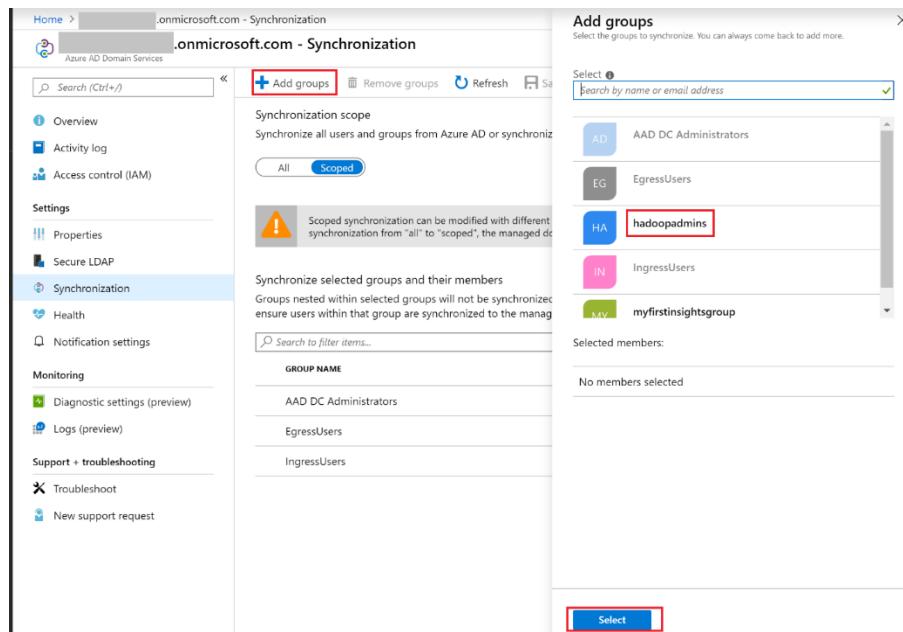


Figure 27: Azure AD Domain Services - Add Group to Synchronization scope

Once added, the group must be part of the list of groups that are enabled for AAD-DS synchronization scope:

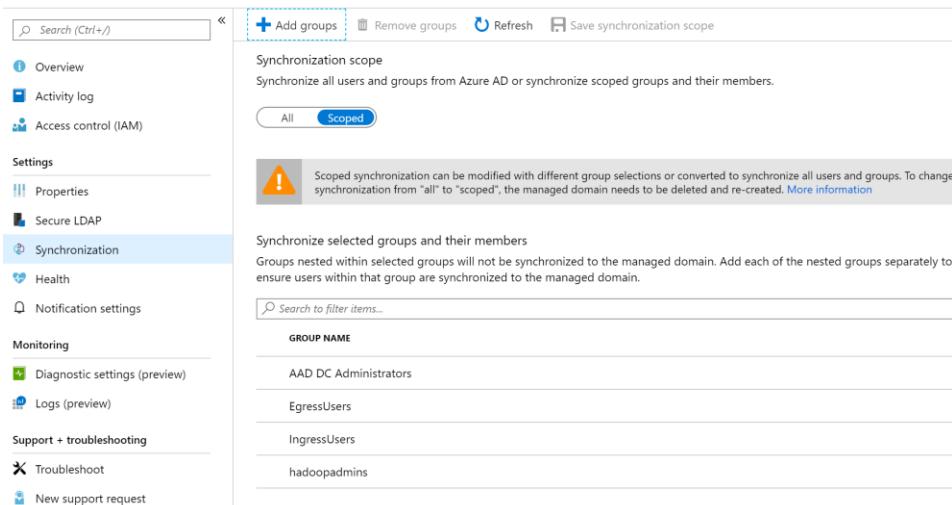


Figure 28: Updated Azure AD Domain Services - Synchronization Scope

## 4.3 Assign HDInsight Domain Services Contributor Role

HDInsight ESP cluster requires a managed identity to be configured, such that it can perform all necessary directory service interactions. HDInsight Domain Services Contributor role is a default role provided by AAD-DS integration to allow any assigned managed identity to perform HDInsight related operations with the AAD-DS. Click [here](#) to learn more on this topic from the official documentation. In this section we will assign the services contributor role to the managed identity that we have created in the previous section.

Here is the sequence of visuals that will guide you to configure the role to the managed identity:

- Navigate to the Access Control (IAM) pane of your AAD-DS resource, and configure the HDInsight Domain Services Contributor role, as shown in the following visual reference:

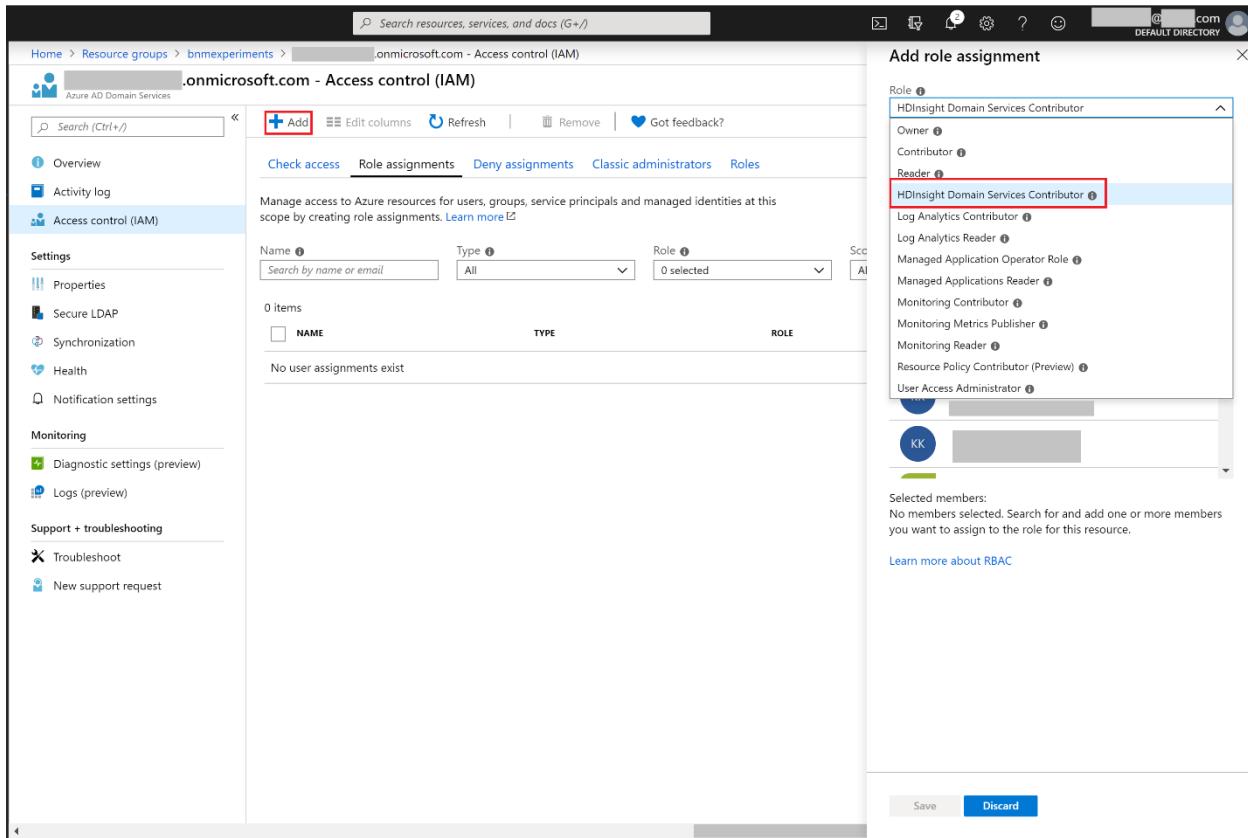


Figure 29: Assign HDInsight Domain Services Contributor role to User Assigned Managed Identity

- In the same view, once you have selected the role, next select the appropriate User assigned Managed Identity (example – `bnnmanagedidentity`) that you have defined in the prior sections. Following visual offers a reference to such selection:

Figure 30: Select Managed Identity for the HDInsight Domain Services Contributor Role assignment

- Subsequently save the identity to role mapping configuration. The saved configuration will be shown as in the following visual:

Figure 31: Updated HDInsight Domain Services Contributor Role Assignment

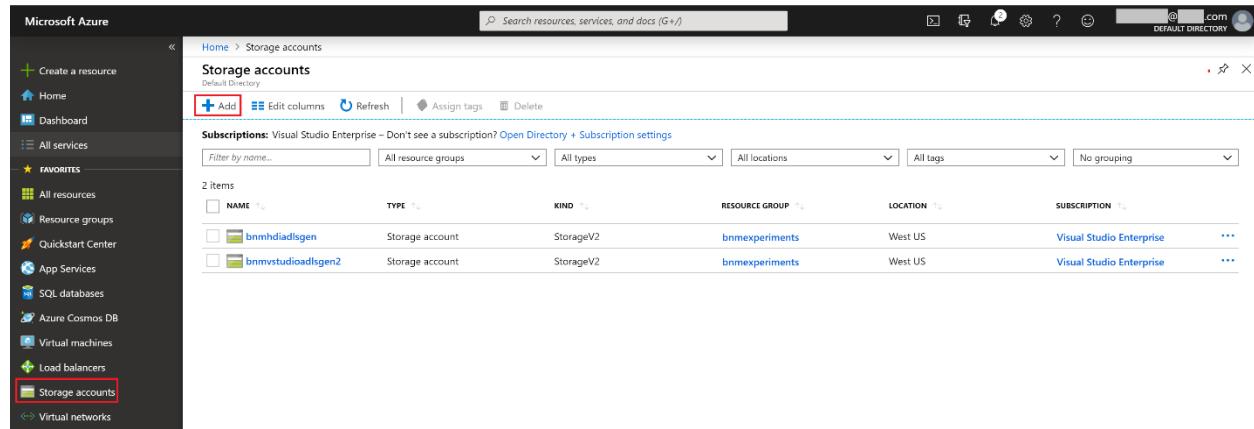
## 5 Create & Configure ADLS Gen2 Storage Account

ADLS Gen2 with support for Hierarchical Namespace (HNS) will be setup as a primary storage layer for the EPS enabled HDInsight cluster. In this section will go through the necessary steps to create an ADLS Gen2 resource and configure required administrator access controls.

### 5.1 Create ADLS Gen2 Storage Account

Using the Azure Portal interface, follow the steps detailed below to create an ADLS Gen2 Storage Account:

- Assuming you are logged in to Azure Portal, navigate to the Storage Accounts blade.



The screenshot shows the Azure Storage accounts blade. On the left, there's a sidebar with various service icons. The 'Storage accounts' icon is highlighted with a red box. The main area has a header 'Storage accounts' with a 'Default Directory' dropdown. Below it is a toolbar with 'Add' (highlighted with a red box), 'Edit columns', 'Refresh', 'Assign tags', and 'Delete'. A search bar 'Search resources, services, and docs (G+)' is at the top right. The main content area shows a table with two items:

NAME	TYPE	KIND	RESOURCE GROUP	LOCATION	SUBSCRIPTION
bnmhdiespadlsgen	Storage account	StorageV2	bnmexperiments	West US	Visual Studio Enterprise
bnmstudioadlsgen2	Storage account	StorageV2	bnmexperiments	West US	Visual Studio Enterprise

Figure 32: List of Storage Accounts

- Click on + Add to start defining the resource configuration. Configuration parameters to specify include:
  - Subscription – choose a subscription where you AAD-DS is enabled and the same where you would have the HDInsight cluster setup.
  - Resource Group – choose a resource group where HDInsight cluster will be setup. For example - `bnmexperiments`.
  - Storage account name – contextually relevant name of the ADLS Gen2 storage account. For example - `bnmhdiespadlsgen2`.
  - Location – same location where you will have HDInsight setup.
  - Performance – if your use case requires low-latent I/O needs, you can boost them by selecting a premium storage option. It provides SSD backed storage layer. If otherwise, you can use a standard choice.

- Account Kind – Select *StorageV2 (general purpose v2)*.
- Replication – you can choose a Locally-redundant Storage (LRS) that can sustain an available zone failure, but not a region-wide failure.
  - Infer the desired replication strategy based on your fault-tolerance requirements.
  - Choices other than LRS will have additional costs. Something to avoid if your business case does not need one.
- Access tier (default) – considering we will have frequent read/write operations on the storage layer, select hot access tier.
- Following visual provides a sample configuration setting:

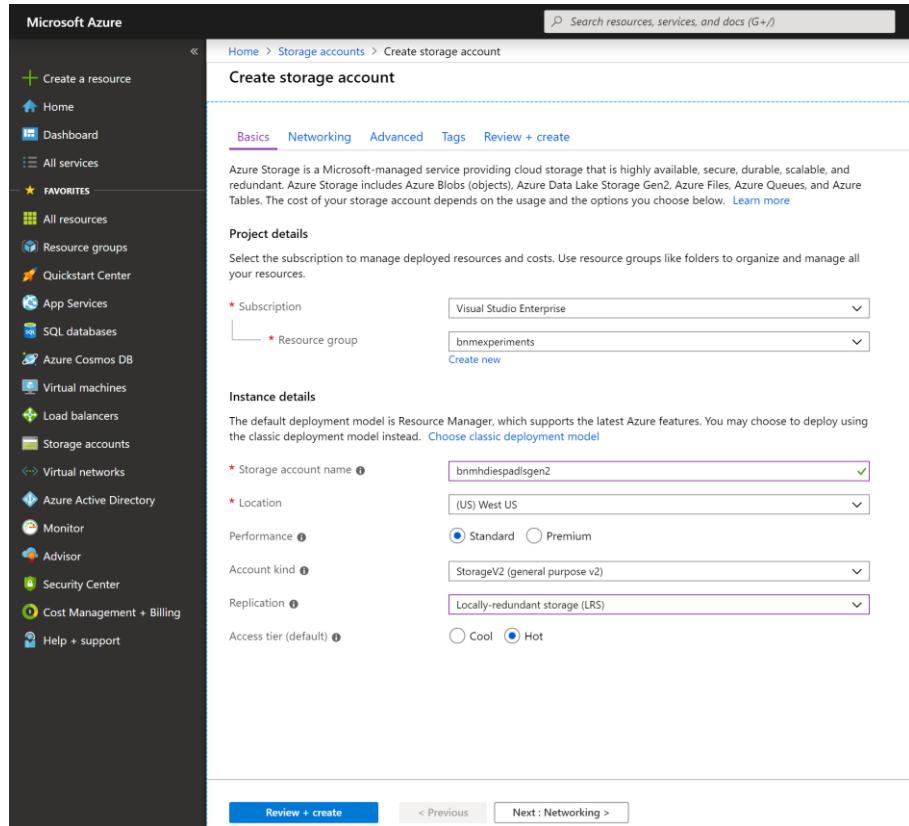


Figure 33: Create Storage Account - Basics Tab

- Click on *Next: Networking* to configure networking options. Access to the storage account can be restricted to select IP addresses or VNETs. For example, we can limit access from just the VNET where HDInsight will be setup.

Home > Create storage account

## Create storage account

Basics Networking Advanced Tags Review + create

### Network connectivity

You can connect to your storage account either publically, 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

### Virtual networks

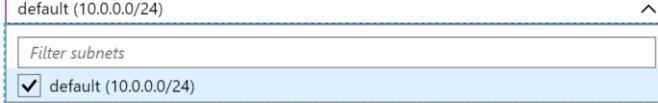
Only the selected network will be able to access this storage account. [Learn more](#)

Virtual network subscription  Visual Studio Enterprise

Virtual network  bnmvnet

[Create virtual network](#)

[Manage selected virtual network](#)

\* Subnets 

default (10.0.0.0/24)
<input type="checkbox"/> default (10.0.0.0/24)

Figure 34: Create Storage Account - Networking Tab

- Then, will configure Advanced configuration settings that include following:
  - Enable *Security – Secure transfer required* to ensure storage account will respect only secured connections. Example, RESTful services that depend on HTTPS rather HTTP, and encrypted SMB connection.
    - Non-secure connection attempts will be rejected.
  - Enable *Hierarchical namespace* to take advantage of the file system semantics and POSIX style support offered by the ADLS Gen2.

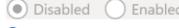
Basics Networking Advanced Tags Review + create

### Security

Secure transfer required 

Disabled  Enabled

### Data protection

Blob soft delete 

Disabled  Enabled

⚠ Blob soft delete and hierarchical namespace cannot be enabled simultaneously.

### Data Lake Storage Gen2

Hierarchical namespace 

Disabled  Enabled

Figure 35: Create Storage Account - Advanced Tab

- Resources can be assigned common tags for manageability reasons. The next step will allow you to add tags. Further, you can review the configuration choices, as shown in the following visual reference:

The screenshot shows the 'Create storage account' review + create page. At the top, there's a green bar with a checkmark and the text 'Validation passed'. Below it, the tabs 'Basics', 'Networking', 'Advanced', 'Tags', and 'Review + create' are visible, with 'Review + create' being the active tab. The 'Basics' section contains the following configuration:

Subscription	Visual Studio Enterprise
Resource group	bnmexperiments
Location	(US) West US
Storage account name	bnmhdiestpadlsgen2
Deployment model	Resource manager
Account kind	StorageV2 (general purpose v2)
Replication	Locally-redundant storage (LRS)
Performance	Standard
Access tier (default)	Hot

The 'Networking' section contains:

Connectivity method	Public endpoint (selected networks)
Virtual network subscription	Visual Studio Enterprise
Virtual network resource group	bnmexperiments
Virtual network	bnmvnet
Subnet	default (10.0.0.0/24)

The 'Advanced' section contains:

Secure transfer required	Enabled
Hierarchical namespace	Enabled
Blob soft delete	Disabled

At the bottom, there are buttons for 'Create' (highlighted in blue), '< Previous' and 'Next >', and a link 'Download a template for automation'.

Figure 36: Create Storage Account - Review + Create - Validation Passed

- Click on create to submit the request to deploy and create the storage account resource. Following is the set of visual references:

The screenshot shows the Microsoft Storage Account Deployment Overview page. The title bar reads "Microsoft.StorageAccount-20191005203707 - Overview". The left sidebar has a "Deployment" icon and lists "Overview", "Inputs", "Outputs", and "Template". The main content area has a heading "Your deployment is underway". It shows deployment details: Deployment name: Microsoft.StorageAccount-20191005203707, Start time: 10/5/2019, 9:06:21 PM, Subscription: Visual Studio Enterprise, Resource group: bnmexperiments. Below this is a table with columns "RESOURCE", "TYPE", "STATUS", and "OPERATION DETAILS". One row shows a resource named "bnmhdiespaldsgen2" of type "Microsoft.Storage/stora..." with status "Accepted" and a link "Operation details". At the bottom is a "Next steps" section.

Figure 37: Storage Account Deployment (underway)

The screenshot shows the Microsoft Storage Account Deployment Overview page. The title bar reads "Microsoft.StorageAccount-20191005203707 - Overview". The left sidebar has a "Deployment" icon and lists "Overview", "Inputs", "Outputs", and "Template". The main content area has a heading "Your deployment is complete" with a checkmark icon. It shows deployment details: Deployment name: Microsoft.StorageAccount-20191005203707, Start time: 10/5/2019, 9:06:21 PM, Subscription: Visual Studio Enterprise, Resource group: bnmexperiments. Below this is a "Deployment details" section with a link "(Download)" and a "Next steps" section. At the bottom is a blue "Go to resource" button.

Figure 38: Storage Account Deployment Complete

- Click on the *Go to resource* button to navigate to the just created ADLS Gen2 storage account.

The screenshot shows the Azure Storage Account Overview page for the account 'bnmhdiespadlsgen2'. The left sidebar lists navigation options like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Data transfer, and Storage Explorer (preview). Under Settings, there are links for Access keys, Geo-replication, CORS, Configuration, Encryption, Shared access signature, Firewalls and virtual networks, and Private endpoint connection. The main content area displays details such as Resource group (bnmexperiments), Status (Primary: Available), Location (West US), Subscription (Visual Studio Enterprise), Subscription ID (05b11ef-40e5-4806-91ae-03dab40f9090), and Tags. Below this, the Services section features four cards: 'Data Lake Gen2 file systems' (highlighted with a red box), 'File shares', 'Tables' (highlighted with a red box), and 'Queues'. The 'Data Lake Gen2 file systems' card describes it as 'Massively scalable data lake storage' and includes a 'Learn more' link.

Figure 39: ADLS Gen2 - Overview

- Note the highlighted links in the above visual. For example, Access control (IAM) allows you to configure security access at the storage account level, while the link to Data Lake Gen2 file systems will help you navigate to the individual storage containers and configure appropriate settings at each container level.

## 5.2 Validate network restrictions

Since we have restricted access to the ADLS Gen2 Storage Account for a select VNET and Subnet, we can quickly verify its impact. This can be quickly validated by logging into your Azure Subscription using Storage Explorer and attempt to access the storage account container resources. Following visual provides a reference to a failure condition that must be expected when you have restricted access to the storage account:

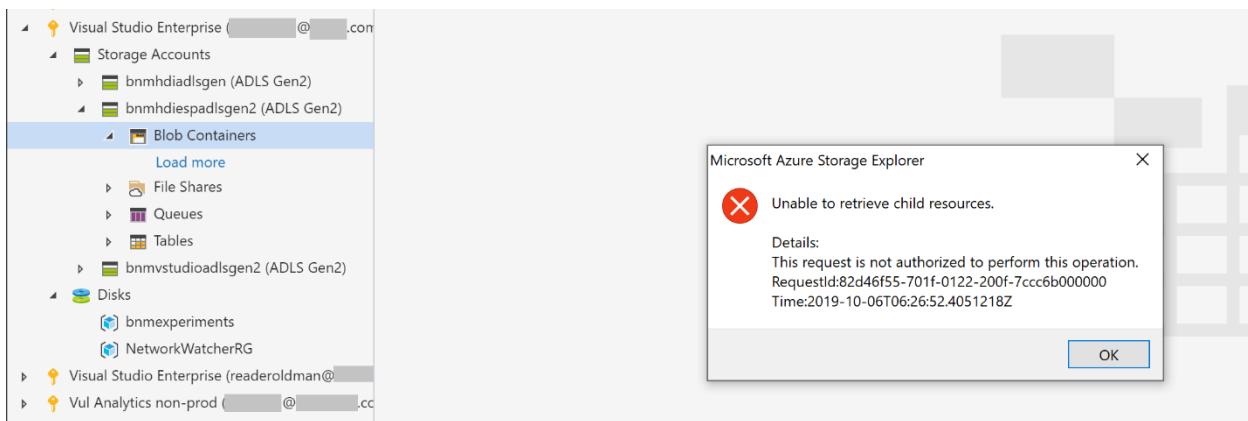


Figure 40: ADLS Gen2 - Validate Network Restrictions - Sample Error Message

Also, the same effect can be observed when attempting to access the storage explorer from within the Azure portal.

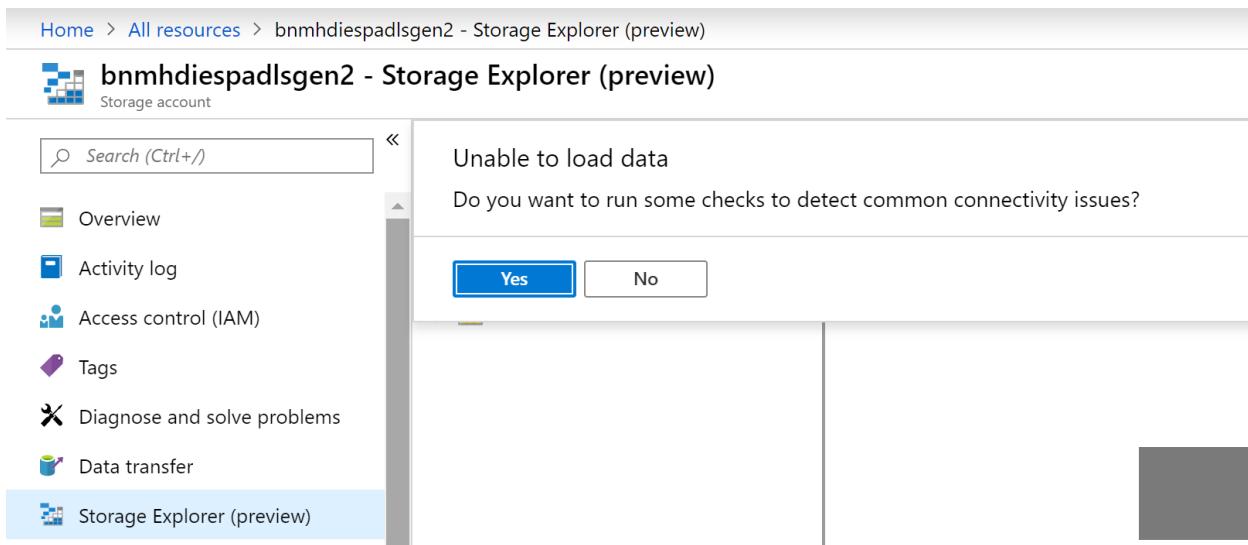


Figure 41: ADLS Gen2 - Validate Network Restrictions - Unable to Load Data (Sample)

## 5.3 Configure Storage Blob Data Owner Role

We need two kinds of owners assigned to the storage account, such that HDInsight services can successful read and write to the underlying storage containers, and that members of the group *hadoopadmins* can administer access to HDFS layer for users intending to query data from data stores like Hive. And, also be able to write to shared folders that are part of the HDInsight storage container.

### 5.3.1 Add role to User Assigned Managed Identity

Following are the necessary steps to assign the Storage Blob Data Owner role to the managed identity – *bnmmanagedidentity*:

- Click on the **+Add** link to configure the desired role.

The screenshot shows the 'Access control (IAM)' blade for a storage account named 'bnmhdiestpadlsgen2'. The left sidebar has 'Access control (IAM)' selected. The main area has tabs for 'Check access', 'Role assignments', 'Deny assignments', 'Classic administrators', and 'Roles'. The 'Check access' tab is active. It includes a search bar for 'Find' (set to 'Azure AD user, group, or service principal') and a search input field. To the right, there's a large callout box titled 'Add a role assignment' with a checked checkbox. Below it, 'View role assignments' and 'View deny assignments' sections are shown, each with a 'View' button and a 'Learn more' link.

Figure 42: ADLS Gen2 - Add Assign User Assigned Managed Identity

- Here's the sample reference visual:

**Add role assignment**

Role **Storage Blob Data Owner**

Assign access to **User assigned managed identity**

\* Subscription **Visual Studio Enterprise**

Select **Search by name**

bnmmngdidentity /subscriptions/05b111ef-40e5-4806-91ae-03dab40f...

bnmmngdidentity /subscriptions/05b111ef-40e5-4806-91ae-03dab40f...

Selected members:  
No members selected. Search for and add one or more members you want to assign the role for this resource.

Learn more about RBAC

**Save** **Discard**

Figure 43: ADLS Gen2 - Managed Identity selection

**Add role assignment**

Role **Storage Blob Data Owner**

Assign access to **User assigned managed identity**

\* Subscription **Visual Studio Enterprise**

Select **Search by name**

bnmmngdidentity /subscriptions/05b111ef-40e5-4806-91ae-03dab40f...

Selected members:  
**bnmmngdidentity** /subscriptions/05b111ef-40e5-4806-91ae-03dab40f... Remove

**Save** **Discard**

Figure 44: ADLS Gen2 - Save selected Managed Identity

The screenshot shows the 'Access control (IAM)' section of the Azure Storage Account for 'bnmhdiespadlsgen2'. The 'Role assignments' tab is active. A table lists a single role assignment:

NAME	TYPE	ROLE	SCOPE
bnmmanagedentity	App	Storage Blob Data Owner	This resource

Figure 45: ADLS Gen2 - Updated Storage Blob Data Owner Role view

- Configuring the role will give the managed identity the ability to define a storage container for the HDInsight cluster, at the time of cluster deployment process.

### 5.3.2 Add role to *hadoopadmins* group

Like the managed identity, add the group *hadoopadmins* the role of a Storage Blob Data Owner. For reference, use the following sample visual:

The screenshot shows the 'Add role assignment' dialog box. The 'Role' dropdown is set to 'Storage Blob Data Owner'. The 'Assign access to' dropdown is set to 'Azure AD user, group, or service principal'. The 'Select' dropdown shows a list of groups, with 'hadoopadmins' highlighted with a red box. The 'Save' button is highlighted with a red box.

Figure 46: ADLS Gen2 - Add role to *hadoopadmins* group

Select the group `hadoopadmins`, then save the selection. Here's the updated list of role assignments:

The screenshot shows the 'Access control (IAM)' blade for the storage account 'bnmhdiestspadlsgen2'. The 'Role assignments' tab is selected. The interface includes search, add, refresh, and remove buttons, along with a feedback link. The main area displays role assignments with filters for name, type, scope, and role. Two items are listed: 'bnmmanagedidentity' (App) and 'hadoopadmins' (Group), both assigned as 'Storage Blob Data Owner' at the 'This resource' scope.

NAME	TYPE	ROLE	SCOPE
bnmmanagedidentity	App	Storage Blob Data Owner	This resource
hadoopadmins	Group	Storage Blob Data Owner	This resource

Figure 47: ADLS Gen2 - Updated Role Assignments View

## 5.4 Add HDInsight Cluster Admin User

In this section will add and configure a dedicated user to administer all Hadoop service activities. The user will also be part of the group *hadoopadmins*. Following sequence describes the steps involved in adding this new user:

- Navigate to the Azure Active Directory in the portal, as shown in the following visual reference, and select the group *hadoopadmins* that will be assigned as part of user creation.

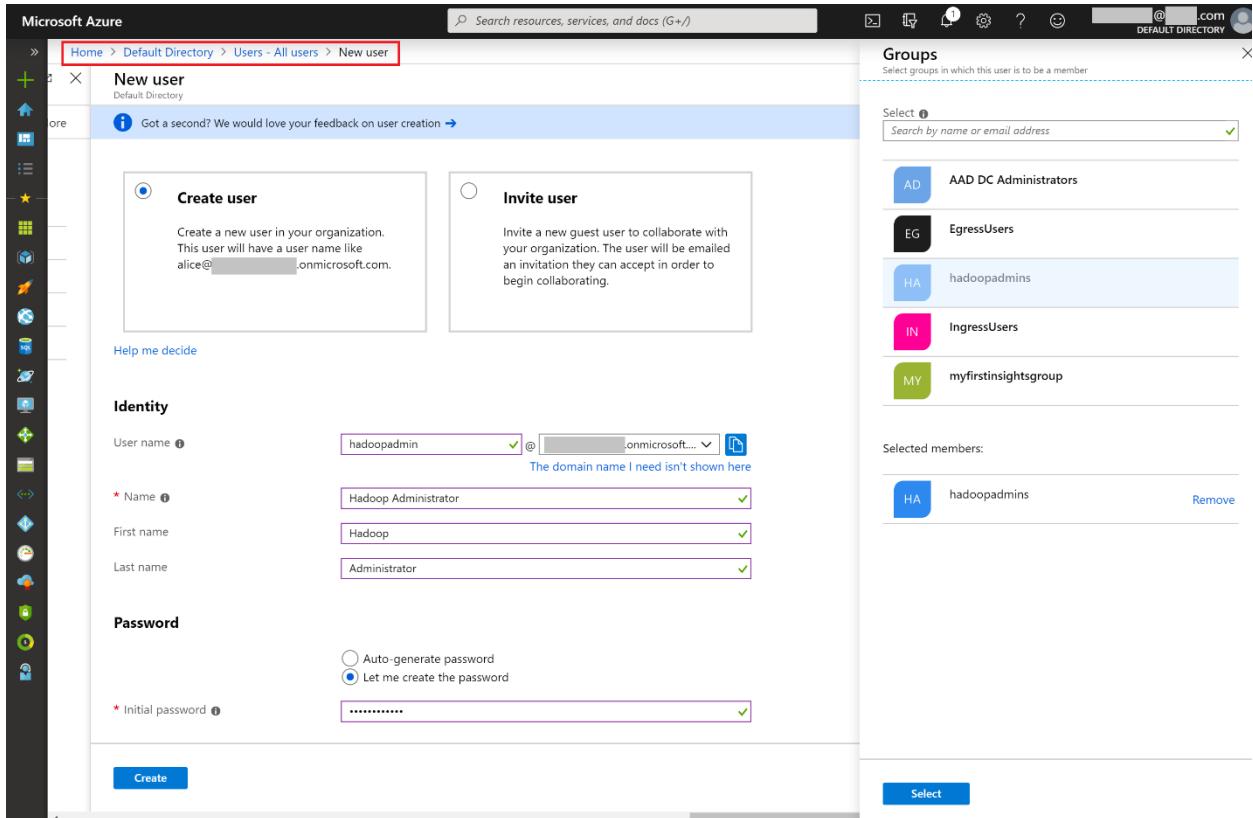


Figure 48: Add *hadoopadmin* user to *hadoopadmins* group

- You can restrict the sign-in of this user by limiting the geo-location from where the user can login into Azure. Ensure to leave the Block Sign-in to No. Else, the user will not be able to sign-in i.e., login to the Azure.

**Settings**

Block sign in	<input type="radio"/> Yes <input checked="" type="radio"/> No
Usage location	United States

**Job info**

Job title	Hadoop Administrator
Department	DevOps

**Create**

Figure 49: Add hadoopadmin user to hadoopadmins group (Contd...)

- Once successfully created, the user would be listed in the Azure Active Directory set of users.

Users - All users					Documentation	X				
All users		+ New user	+ New guest user	Bulk create	Bulk invite	Bulk delete	Download users	Reset password	Delete user	More
All users										
Deleted users										
Password reset										
User settings										
Activity										
Sign-ins										
Audit logs										
Bulk operation results (Preview)										
Troubleshooting + Support										
Troubleshoot										
...										

Search: Name or email  
Search attributes: Name, email (begins with)  
Show: All users

NAME	USER NAME	USER TYPE	SOURCE
KK Kalyan Kadiyala	@.com	Member	Microsoft Account
EO External User		Guest	External Azure Active Directory
RO Reader Oldman	readeroldman@.onmicrosoft.com	Member	Azure Active Directory
WN Writer Newman	writernewman@.onmicrosoft.com	Member	Azure Active Directory
HA Hadoop Administrator	hadoopadmin@.onmicrosoft.com	Member	Azure Active Directory

Figure 50: Azure Active Directory - All Users view

- And, here is a quick look at the Hadoop Administrator's profile:

Figure 51: Hadoop Administrator User Profile (Azure Active Directory)

- The next step is to verify login to Azure subscription. To do so, click on the profile icon on the top right corner of your Azure portal and choose *Sign in with a different account link*.

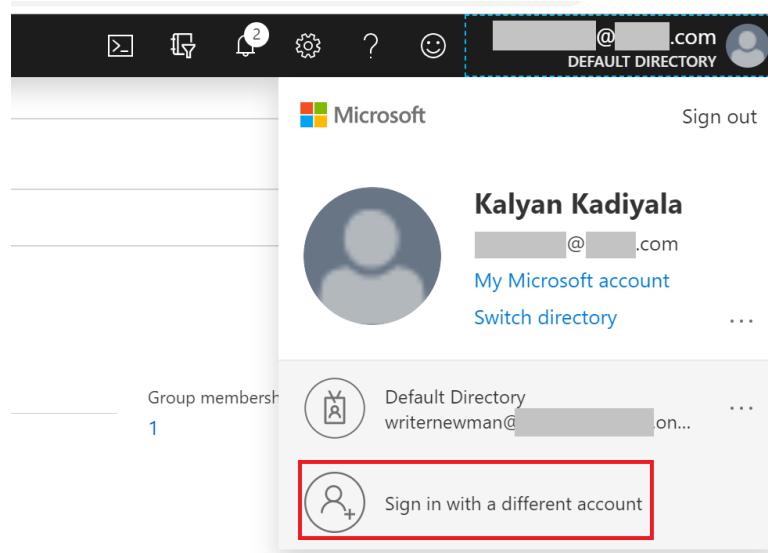
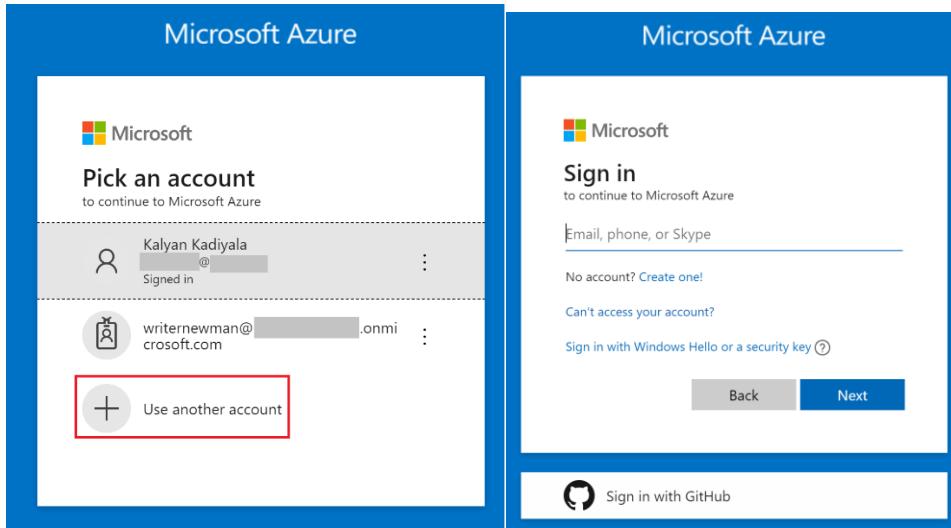


Figure 52: Azure Portal - Sign-in with a different account



- Select + Use another account.
- In the login prompt enter the credentials for our new user –  
hadoopadmin@youraddsdomainname.onmicrosoft.com

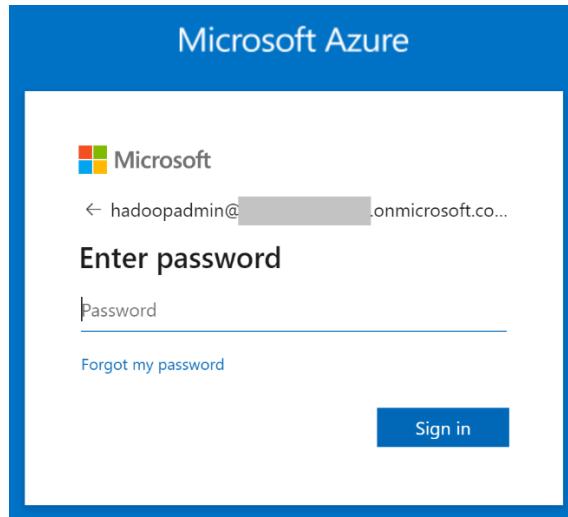


Figure 53: Azure Portal Login

- On first successful login you will be prompted to change your password

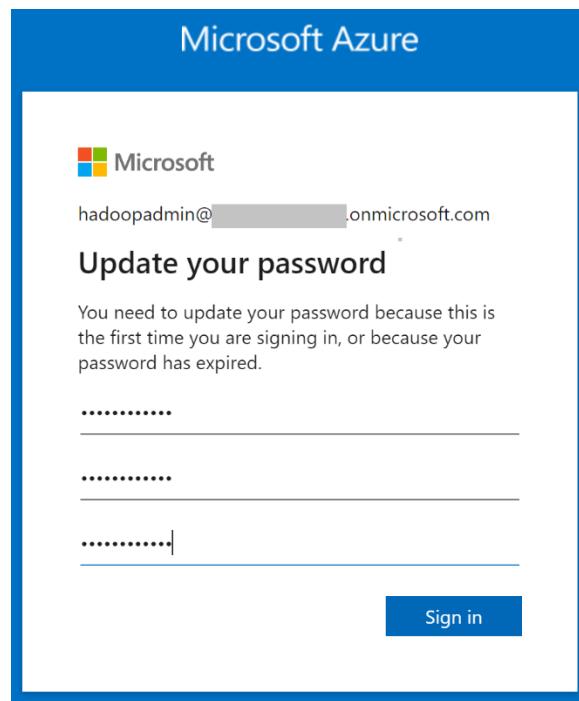


Figure 54: Azure Portal First time login - Update Password

- Visual reference that shows a landing page post successful login (with password change).

Figure 55: Azure Portal New User Login - Landing Page

Note – able to successfully login and being part of the *hadoopadmins* user group does not grant automatic access to the resources like Hadoop Cluster and the underlying ADLS Gen2. This is accomplished by enabling roles. More on this under the section – *Enforce Storage Access-Controls*.

## 5.5 Additional role configurations for group *hadoopadmins*

To be able to administer user access controls, the members of the group *hadoopadmins* must be enabled with additional roles at parent resource levels in the hierarchy. This includes the resource, resource-group and at the subscription. Following is a visual excerpt from the [official](#) documentation that demonstrates how the RBAC roles are enforced at various scope levels of the resource hierarchy:

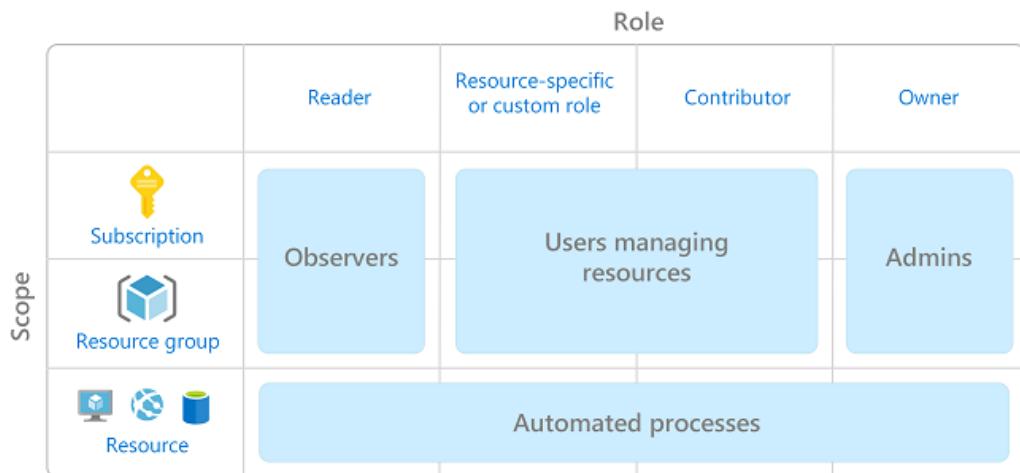


Figure 56: RBAC Roles Application Overview

Based on the above guidance following role assignments must be made for the group *hadoopadmins*:

Role	Scope (resource type)	Additional Description
Reader	Storage Account, Resource Group	Use Access Control (IAM) from respective resources.
Reader and Data Access	Storage Account	Storage Account's Access Control (IAM); allows the respective users to be able to view folder contents (say from storage explorer)
Storage Blob Data Owner	Storage Account	Storage Account's Access Control (IAM)

Table 1: *hadoopadmins* group RBAC roles Overview

Since we have already assigned the Storage Blob Data Owner access to the *hadoopadmins* group, we also need to enable the Reader and the Reader & Data Access roles at respective scopes. The process is similar to that described in the section *Add role to hadoopadmins group*, where you would be using the resource specific Access Control (IAM) pane.

Following visual provides a view of the expected role assignment at the storage account level for the group *hadoopadmins*:

The screenshot shows the 'Access control (IAM)' section of the Azure Storage Account 'bnmhdiespaldigen2'. The 'Role assignments' tab is selected. It displays five items (1 User, 3 Groups, 1 Service Principal) under four categories: OWNER, READER, READER AND DATA ACCESS, and STORAGE BLOB DATA OWNER.

OWNER	READER	READER AND DATA ACCESS	STORAGE BLOB DATA OWNER
Kalyan Kadiyala (User)	hadoopadmins (Group)	hadoopadmins (Group)	bnmmanagedidentity (App)
	hadoopadmins (Group)		hadoopadmins (Group)

Figure 57: *hadoopadmins* group assigned roles - Access Control (IAM) overview

And, here is the expected storage explorer view that a user of the group *hadoopadmins* should see (in this case it is the admin user that we have added in preceding sections):

The screenshot shows the 'Storage Explorer (preview)' section of the Azure Storage Account 'bnmhdiespaldigen2'. The left sidebar shows the 'Storage Explorer (preview)' section is selected. The main pane displays a file system structure with a folder named 'bnmhdiespaldigen2-2019-10-08t2' highlighted.

NAME	LAST MODIFIED	CONTENT TYPE	SIZE
ams	10/8/2019, 2:52:11 PM	Folder	0 B
amshbase	10/8/2019, 2:52:11 PM	Folder	0 B
app-logs	10/8/2019, 2:52:11 PM	Folder	0 B
apps	10/8/2019, 2:52:11 PM	Folder	0 B
atshistory	10/8/2019, 2:52:11 PM	Folder	0 B
example	10/8/2019, 3:41:05 PM	Folder	0 B
hbase	10/8/2019, 2:52:11 PM	Folder	0 B
HdiSamples	10/8/2019, 3:42:08 PM	Folder	0 B
hdp	10/8/2019, 6:01:58 PM	Folder	0 B
hive	10/8/2019, 2:52:11 PM	Folder	0 B
mapred	10/8/2019, 2:52:11 PM	Folder	0 B
mapreduceaging	10/8/2019, 5:31:06 PM	Folder	0 B
mr-history	10/8/2019, 2:52:11 PM	Folder	0 B
ranger	10/8/2019, 3:15:13 PM	Folder	0 B
tezstaging	10/8/2019, 3:13:19 PM	Folder	0 B
tmp	10/8/2019, 2:52:11 PM	Folder	0 B
user	10/8/2019, 6:01:48 PM	Folder	0 B

Figure 58: Hadoop Admin User - Storage Explorer View - Folder View Sample

## 6 Setup HDInsight ESP enabled cluster

Azure HDInsight supports several cluster types depending on your needs. For example, you could have a cluster of type Hadoop that focuses on Hive, HDFS and YARN/Map-reduce. If you intend to bring in Apache Spark workload then there is a cluster type called Spark. And so forth, for HBase if you intend to setup a low-latent, consistent NoSQL backend data store.

By enabling the Enterprise Security Package (ESP) support, HDInsight services can leverage the Azure Active Directory – Domain Services (AAD-DS) to address user & service principal identity management, and handle secure interactions with the cluster services that require Kerberos authentication. It is important to note that storage layer access controls must be managed by the interfaces supported by ADLS Gen2 itself. These include Role-based Access Controls (RBAC) and Access Control Lists (ACLs). RBAC's fine grain support is limited to assigning and managing roles at-most at the container level. While, ACLs allow you to enforce more fine-grained access control at folder paths and files within containers as well.

The sub-sections below will guide through the necessary steps to create and setup HDInsight cluster along with enabling ESP support.

## 6.1 Create HDInsight + ESP Resource

- Assuming you have already logged in to the [Azure Portal](#), click the *+ Create Resource* link from the left navigation pane.

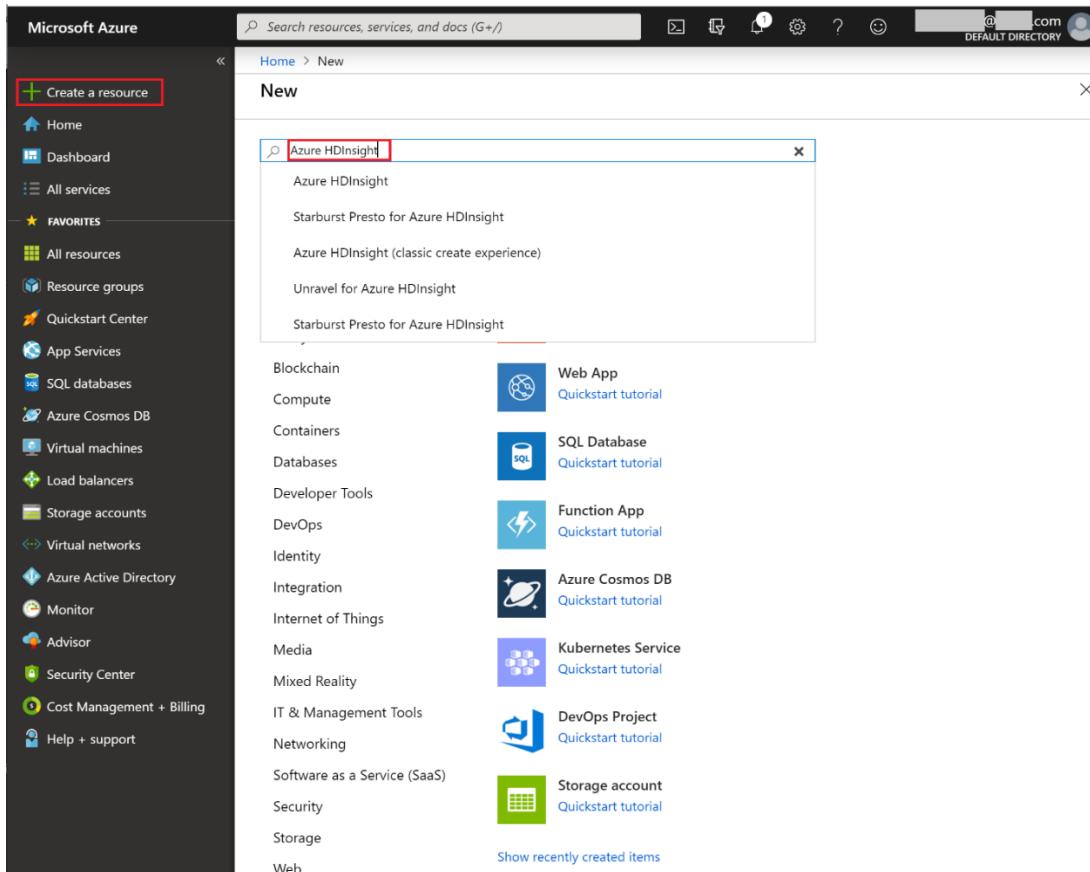


Figure 59: Create Azure HDInsight - Search

- Once you have selected the resource type, the next step is to begin the creation process.

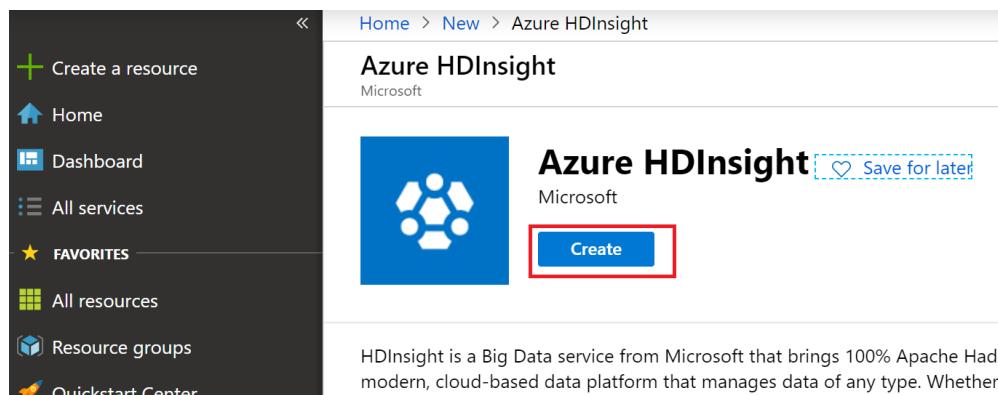


Figure 60: Create Azure HDInsight Resource

- In the *Basics* tab select the Subscription, Resource Group, name for the cluster (of your choice), location (Azure region), the type of the cluster and the cluster credentials. Following visual references provide some contextual examples:

The screenshot displays the Azure portal interface for creating an HDInsight cluster. On the left, the main page shows the 'Create HDInsight cluster' wizard with tabs for Basics, Storage, Security + networking, Configuration + pricing, Review + create, and Go to classic create experience. The Basics tab is active, showing fields for Subscription (Visual Studio Enterprise), Resource group (Unexperiment), Cluster name (bmhdinsightespcluster), Location (West US), Cluster type (Select cluster type), and Version (dropdown menu). Below these are sections for Project details, Cluster details, and Cluster credentials. At the bottom are Review + create and Next: Storage buttons.

On the right, a separate modal window titled 'Select cluster type' lists several options with their descriptions and 'Select' buttons:

- Hadoop**: Petabyte-scale processing with Hadoop components like MapReduce, Hive (SQL on Hadoop), Pig, Sqoop and Oozie. (Select button)
- Spark**: Fast data analytics and cluster computing using in-memory processing. (Select button)
- Kafka**: Build a high throughput, low-latency, real-time streaming platform using a fast, scalable, durable, and fault-tolerant publish-subscribe messaging system. (Select button)
- HBase**: Fast and scalable NoSQL database. Available with both standard and premium (SSD) storage options. (Select button)
- Interactive Query**: Build Enterprise Data Warehouse with in-memory analytics using Hive (SQL on Hadoop) and LLAP (Low Latency Analytical Processing). Note that this feature requires high memory instances. (Select button)
- ElastiCache**: Reliably process infinite streams of data in real-time. (Select button)
- Storm**: Reliably process infinite streams of data in real-time. (Select button)
- ML Services (R Server)**: Analyze data at scale, build intelligent apps and discover valuable insights across your business using both R and Python. Adds 0.016 USD per Core-Hour. (Select button)

Figure 61: Create HDInsight + ESP - Choose cluster type

- In this experiment we are going to select HDI 3.6, which is observed to be a common compatible version in the Azure migration case studies. For example, it is compatible with Hortonworks Distribution version 2.6. Please note, the way connected components are integrated in HDInsight distribution will still vary with that of open source or other vendor frameworks. For example, HDInsight relies heavily on ADLS choices to satisfy HDFS at-rest needs.

Home > New > Azure HDInsight > Create HDInsight cluster

### Create HDInsight cluster

[Go to classic create experience](#)

#### Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

\* Subscription: Visual Studio Enterprise  
 \* Resource group: bnmexperiments  
[Create new](#)

#### Cluster details

Name your cluster, pick a region, and choose a cluster type and version. [Learn more](#)

\* Cluster name: bnmhd3dot6espcluster  
 \* Region: West US  
 \* Cluster type: Hadoop  
[Change](#)  
 \* Version: Hadoop 2.7.3 (HDI 3.6)

#### Cluster credentials

Enter new credentials that will be used to administer or access the cluster.

\* Cluster login username: admin  
 \* Cluster login password:  
 \* Confirm cluster login password:  
 \* Secure Shell (SSH) username: sshuser  
 Use cluster login password for SSH

[Review + create](#)   [< Previous](#)   [Next: Storage >](#)

Figure 62: Create HDInsight + ESP cluster - Details

- Quick note on the cluster login name parameter above. It maps to the HTTP-User who can login to Ambari with administrator privileges. Note – as part of the Security + Network tab configuration we will enable ESP and add a cluster admin user. The identity of this user will be part of the AAD and synchronized by AAD-DS. The configuration parameter here is – *Cluster admin user*. We will need to select the AAD admin user (*hadoopadmin@youraadddomainname.onmicrosoft.com*) that we have added in prior sections above.
- Click on `Next: Storage >>` to configure the storage.
  - Select ADLS Gen2 Storage Account that we have created earlier – *bnmhdiesspadlsgen2*.
  - Then specify the User Assigned Managed Identity *bnmmanagedidentity*, that the cluster will use to represent on its behalf when interacting with the storage containers.

- You can choose to give the file system a name or leave it at its default.

Create HDInsight cluster

Go to classic create experience

Basics Storage Security + networking Configuration + pricing Review + create

Select or create storage accounts that will be used for the cluster's logs, job input, and job output. Configure the cluster's access to these accounts, if needed.

**Primary storage**

Select or create a storage account that will be the default location for cluster logs and other output.

\* Primary storage type: Azure Data Lake Storage Gen2  
 \* Primary storage account: brnvhdiesspadlsgen2  
 \* Filesystem: brnvhdi3dot6tespccluster-2019-10-08t19-53-08-245z

**Identity**

Select a user-assigned managed identity to represent the cluster for Azure Data Lake Gen2 Storage account access. Only identities with access to the selected storage account are listed. [Learn more](#)

\* User-assigned managed identity: brnvhdiesspadlsgen2

**Additional Azure storage**

Link additional Azure storage accounts to the cluster.

ACCOUNT NAME Add Azure storage

**Metastore settings**

To preserve your Hive and/or Oozie metadata outside of this cluster, select a SQL database for this cluster.

SQL database for Hive:   
 SQL database for Oozie:

Review + create < Previous Next: Security + networking >

Figure 63: Create HDInsight + ESP cluster - Storage Tab

- It is recommended to consider using an external datastore of choice for services like Hive and Oozie, when setting up a production cluster. For example, SQL DB offers in-built replication strategies to satisfy DR strategies.
- Next step is to configure the Security and networking, where we also would enable the Enterprise Security Package.

Home > New > Azure HDInsight > Create HDInsight cluster

Create HDInsight cluster

Go to classic create experience

Basics Storage Security + networking Configuration + pricing Review + create

Configure your cluster's security and network settings.

**Enterprise security package**

Connect this cluster with Active Directory Domain Services (AAD-DS) to have finer control of who can access the cluster. [Learn more](#)

Enable enterprise security package (Adds 0.01 USD per Core-Hour)

**Virtual network**

Connect this cluster to a virtual network. [Learn more](#)

Virtual network:

**Identity**

Select a user-assigned service identity to represent your cluster for enterprise security package or Kafka disk encryption. [Learn more](#)

User-assigned managed identity:

Review + create < Previous Next: Configuration + pricing >

Figure 64: Create HDInsight + ESP cluster - Security + Networking Tab

- Enable the Enterprise Security Package.

The screenshot shows the Azure portal interface for creating an HDInsight cluster. The top navigation bar includes 'Home', 'New', 'Azure HDInsight', and 'Create HDInsight cluster'. Below the navigation is a title 'Create HDInsight cluster' with a link to 'Go to classic create experience'. The main content area has tabs for 'Basics', 'Storage', 'Security + networking' (which is underlined in purple), 'Configuration + pricing', and 'Review + create'. Under 'Security + networking', there's a note about connecting to Active Directory Domain Services (AAD-DS) for finer control of access. A checkbox labeled 'Enable enterprise security package' is checked, with a note that it adds 0.01 USD per Core-Hour. A validation status box says 'validating...'. Below this, the 'Virtual network' section shows '(Default) System-managed virtual network' as the selected option. The 'Identity' section allows selecting a user-assigned service identity for enterprise security or Kafka disk encryption. At the bottom are buttons for 'Review + create' (highlighted in blue), '« Previous', and 'Next: Configuration + pricing »'.

Figure 65: Create HDInsight + ESP cluster - Enable ESP

- Once enabled, the configuration process will validate existing AAD-DS setup, that is available in the subscription's tenant. It also derives necessary domain name parameters as shown below:

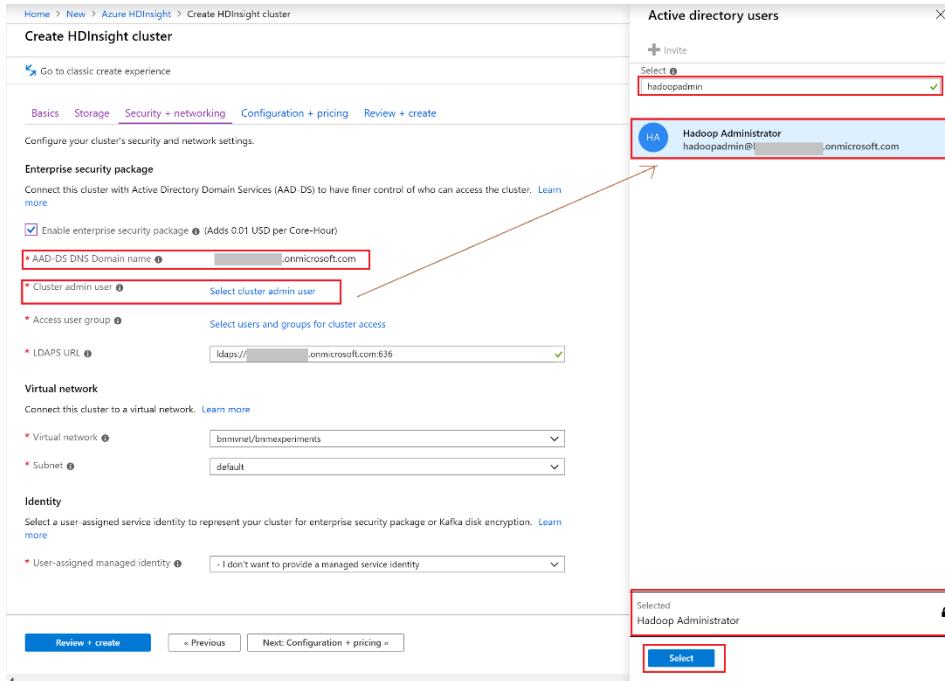


Figure 66: Create HDInsight + ESP cluster - Select Cluster Admin User

- Pay attention to the AAD-DS DNS Domain name and the LDAPS URL. These should map to your AAD-DS DNS name. For example, let's call it *youraddsdomainname.onmicrosoft.com*.
- Next will choose the *cluster admin user*, who will have permissions to manage cluster services as well grant permissions to other users for cluster access. Example, able to use Ranger and grant permission to a Hive table.
  - This user must be defined in the Azure Active Directory. For example, it should be the [hadoopadmin@youraddsdomainname.onmicrosoft.com](mailto:hadoopadmin@youraddsdomainname.onmicrosoft.com) user that we have created in the previous steps.
- Will next choose the additional AD user groups that should have user level access to the cluster. For example – *IngressUsers* and *EgressUsers*.
- Your configuration pane should look like, very similar to the following visual below:

The screenshot shows the 'Create HDInsight cluster' wizard in the Azure portal. The current step is 'Security + networking'. The 'Enterprise security package' section is active, showing a checked checkbox for 'Enable enterprise security package' which adds 0.01 USD per Core-Hour. Below it, the 'AAD-DS DNS Domain name' is set to '.onmicrosoft.com'. The 'Cluster admin user' is listed as 'Hadoop Administrator' with a link to 'Change cluster admin user'. Under 'Access user group', '2 groups selected' are shown with a link to 'Change groups or users'. The 'LDAPS URL' is set to 'ldaps://onmicrosoft.com:636'. The 'Virtual network' section shows 'Virtual network' set to 'bnmvnet/bnmexperiments' and 'Subnet' set to 'default'. The 'Identity' section shows 'User-assigned managed identity' set to 'bnmmmanagedidentity'. Navigation buttons at the bottom include 'Review + create' (highlighted in blue), '< Previous', and 'Next: Configuration + pricing >'.

- Also note the choice of VNET/Subnet choices above. The VNET choice here must be the same as that configured on Storage Account's firewall rules. Otherwise, it can be any other VNET with the exception your storage account does not block access to traffic from such virtual network.
- It is important to ensure the User-assigned Managed Identity is set to the identity that is created under the section *Configure Access Controls* section above. The identity is used by HDInsight cluster service represent on their behalf when interacting with the storage account, as well to handle on-desk encryption for Kafka.
- Next choose the configuration and pricing options, as show in the sample below:

## Create HDInsight cluster

 Go to classic create experience

Configure cluster performance and pricing. [Learn more](#)

### Node configuration

Configure your cluster's size and performance, and view estimated cost information.

The cost estimate represented in the table does not include subscription discounts or costs related to storage, networking, or data transfer.

 This configuration will use 20 of 250 available cores in the West US region.  
[View cores usage](#)

NODE TYPE	NODE SIZE	NUMBER OF NODES	
Head node	D12 v2 (4 Cores, 28 GB RAM), 0.34 USD/hour	2	0.69 USD
Worker node	D12 (4 Cores, 28 GB RAM), 0.34 USD/hour	3	1.03 USD
<input type="checkbox"/> Enable autoscale (preview) <a href="#">Learn more</a>			
Enterprise security package cost/hour			0.20 USD
<b>Total estimated cost/hour</b>			<b>1.91 USD</b>

[Review + create](#)

[« Previous](#)

[Next: Review + create »](#)

Figure 67: Create HDInsight + ESP cluster - Price Estimate

- Choice of nodes here must be derived based on the workloads and required capacity to support them. Head nodes host one or more coordinator services like YARN Resource Manager, HDFS Namenode, and Hive Master.
- Next step is to review the configurations and create the cluster.

Home > New > Azure HDInsight > Create HDInsight cluster

## Create HDInsight cluster

[Go to classic create experience](#)

Validation succeeded.

[Basics](#) [Storage](#) [Security + networking](#) [Configuration + pricing](#) [Review + create](#)

Hadoop 2.7.3 (HDI 3.6) **1.91 USD Total estimated cost/hour**  
 This estimate does not include subscription discounts or costs related to storage, networking, or data transfer.

**Basics**

Subscription	Visual Studio Enterprise
Resource group	bnmexperiments
Region	West US
Cluster name	(new) bnmhdi3dot6espcluster
Cluster type	Hadoop 2.7.3 (HDI 3.6)
Cluster login username	admin
Secure Shell (SSH) username	sshuser
Use cluster login password for SSH	Enabled

**Security + networking**

Enterprise security package	Enabled
AAD-DS DNS Domain name	[REDACTED] onmicrosoft.com
LDAPS URL	ldaps://[REDACTED].onmicrosoft.com:636
Cluster admin user	Hadoop Administrator
Virtual network	bnnvnet
Subnet	default
User assigned managed identity	bnnmanagedidentity

**Storage**

Primary storage type	Azure Data Lake Storage Gen2
Primary storage account	bnmhdi3dot6espcluster-2019-10-13t17-36-45-517z
Filesystem	bnmhdi3dot6espcluster-2019-10-13t17-36-45-517z
User-assigned managed identity	bnnmanagedidentity
Additional Azure storage	None
Data Lake Storage Gen1 access	Disabled

**Cluster configuration**

Head	2 nodes, D12 v2 (4 Cores, 28 GB RAM)
Worker	3 nodes, D12 (4 Cores, 28 GB RAM)

[Create](#) [« Previous](#) [Next](#) [Download a template for automation](#)

Figure 68: Create HDInsight + ESP cluster - Configuration Summary

- Submit configuration for resource creation and deployment.

HDInsight\_2019-10-13T18.30.29.812Z - Overview

Your deployment is underway

Deployment name: HDInsight\_2019-10-13T18.30.29.812Z Start time: 10/13/2019, 11:30:31 AM  
Subscription: Visual Studio Enterprise Correlation ID: 73b961f5-143b-4d65-a836-d3462fc94b19  
Resource group: bnmexperiments

No results.

Figure 69: Create HDInsight + ESP cluster - Deployment Underway

HDInsight\_2019-10-13T18.30.29.812Z - Overview

Your deployment is underway

Deployment name: HDInsight\_2019-10-13T18.30.29.812Z Start time: 10/13/2019, 11:30:31 AM  
Subscription: Visual Studio Enterprise Correlation ID: 73b961f5-143b-4d65-a836-d3462fc94b19  
Resource group: bnmexperiments

RESOURCE	TYPE	STATUS	OPERATION DETAILS
bnmhdi3dot6espcluster	Microsoft.HDInsight/cluster	OK	<a href="#">Operation details</a>

- The deployment will take about 25 to 30 minutes. It would add whole bunch of resources including:
  - Network Interface Cards (NIC) for each of the nodes – gateway, head, zookeeper and worker nodes.
  - Load balancers for the gateway and head nodes.
  - HDInsight cluster
  - Etc. as required for a functional HDInsight cluster with ESP enabled.
- Here is the sample visual that shows the detailed provided by the portal on successful deployment of the cluster.

The screenshot shows the Azure portal's deployment overview for 'HDInsight\_2019-10-13T18.30.29.812Z'. The left sidebar has 'Overview' selected. The main area displays a green checkmark indicating 'Your deployment is complete'. Deployment details show the name 'HDInsight\_2019-10-13T18.30.29.812Z', subscription 'Visual Studio Enterprise', and resource group 'bnmexperiments'. The start time is 10/13/2019, 11:30:31 AM, and the correlation ID is 73b961f5-143b-4d65-a836-d3462fc94b19. A table lists the resource 'bnmhdi3dot6espcluster' as a Microsoft.HDInsight/clusters type in an 'OK' status. A 'Go to resource' button is at the bottom.

Figure 70: Create HDInsight + ESP cluster - Deployment Complete

- And, when you click on the *Operation details* link (see under the deployment details above), you should see detail that is very similar to following:

The screenshot shows the 'Operation details' page for the deployment. It includes a summary table with columns for Operation ID, Tracking ID, Status, Provisioning State, Timestamp, Duration, Type, Resource ID, and Resource. The 'Operation details' link in the sidebar is highlighted.

OPERATION DETAILS	OPERATION ID	944528D65B267804
	TRACKING ID	bd00e105-bdf8-40f2-9f3b-66fd93985be0
	STATUS	OK
	PROVISIONING STATE	Succeeded
	TIMESTAMP	10/13/2019, 11:55:30 AM
	DURATION	24 minutes 57 seconds
<b>OPERATION DETAILS</b>	TYPE	Microsoft.HDInsight/clusters
<b>Operation details</b>	RESOURCE ID	/subscriptions/05b111ef-40e5-4806-91ae-03d...
	RESOURCE	bnmhdi3dot6espcluster

Figure 71: Create HDInsight + ESP cluster - Deployment Operation Details

- Here is the sample visual reference that shows the overview page of the newly created secure HDInsight cluster:

Figure 72: HDInsight + ESP cluster overview

- The URL shown to the top right of the pane above is the Ambari Web UI. Ambari listens on port 443. NSG rules must be configured to allow inbound traffic on port 443. Without which, access to the Ambari Web UI cannot be satisfied by your browser. Here's the sample error for additional details of specific browser error:

Figure 73: Ambari UI - Cannot reach the page

## 6.2 Enable NSG rules for Ambari UI (port 443)

As observed from the previous step, without enabling NSG rule for inbound traffic on secure HTTP port, users will not be able to access the Ambari Web UI. In this section will add NSG rule to enable such access.

- Navigate to the NSG resource using the *All Resources* link available on the left navigation pane of the portal.

The screenshot shows the Microsoft Azure portal interface. On the left, there's a dark sidebar with various navigation options like 'Create a resource', 'Home', 'Dashboard', 'All services', and 'FAVORITES'. Under 'FAVORITES', the 'All resources' link is highlighted with a red box. The main content area is titled 'All resources' and shows a list of 28 records. One item, 'AADDSS-kckadiyalagmail.onmicrosoft.com-NSG', is also highlighted with a red box. The list includes entries for Public IP address, Network security group, HDInsight cluster, Storage account, and Managed Identity.

- Click + Add link to add an entry for the port 443 inbound TCP traffic.

The screenshot shows the 'Inbound security rules' page for the NSG 'AADDSS-kckadiyalagmail.onmicrosoft.com-NSG'. On the left, there's a sidebar with 'Overview', 'Activity log', 'Access control (IAM)', 'Tags', 'Diagnose and solve problems', 'Settings' (with 'Inbound security rules' selected), 'Outbound security rules', 'Network interfaces', 'Subnets', 'Properties', 'Locks', 'Export template', 'Monitoring' (with 'Diagnostic settings' selected), 'Logs', 'NSG flow logs', 'Support + troubleshooting', 'Effective security rules', and 'New support request'. The 'Add' button in the top right of the main table is highlighted with a red box. The right side shows a detailed 'Add inbound security rule' form with the following configuration:

Source	Any
Source port ranges	(empty)
Destination	IP Addresses
Destination IP addresses/CIDR ranges	10.0.0.0/24
Destination port ranges	443
Protocol	TCP
Action	Allow
Name	HDInsightSecureHTTPPortInboundAccess

A large red box highlights the 'Add' button at the bottom right of the form.

Figure 74: Enable Port 443 on NSG for Ambari UI

## 6.3 Verify Ambari UI HTTPS Port access

- Once the rule is added it will be part of the NSG Inbound Security rules list as shown below:

421	HDInsight-168.61.49.99-443	443	Any	168.61.49.99	Any	Allow	...
431	HDInsight-23.99.5.239-443	443	Any	23.99.5.239	Any	Allow	...
441	HDInsight-168.61.48.131-443	443	Any	168.61.48.131	Any	Allow	...
451	HDInsight-138.91.141.162-443	443	Any	138.91.141.162	Any	Allow	...
461	HDInsight-13.64.254.98-443-HealthAndMg...	443	Any	13.64.254.98	Any	Allow	...
471	HDInsight-23.101.196.19-443-HealthAndMg...	443	Any	23.101.196.19	Any	Allow	...
481	HDInsight-168.63.129.16-53-AzureDNSService	53	Any	168.63.129.16	Any	Allow	...
491	168.61.49.99-To-HDInsight-On-443	443	TCP	168.61.49.99	VirtualNetwork	Allow	...
501	HDInsightSecureHTTPPortInboundAccess	443	TCP	Any	10.0.0.0/24	Allow	...

Figure 75: NSG Inbound Rule Updated - Ambari UI - HTTPS port enabled

- Give it few seconds and refresh your browser. You should see following login prompt, where you can use the user name as the [hadoopadmin@youradddomainname.onmicrosoft.com](mailto:hadoopadmin@youradddomainname.onmicrosoft.com) and the password as configured from earlier AD user setup process.

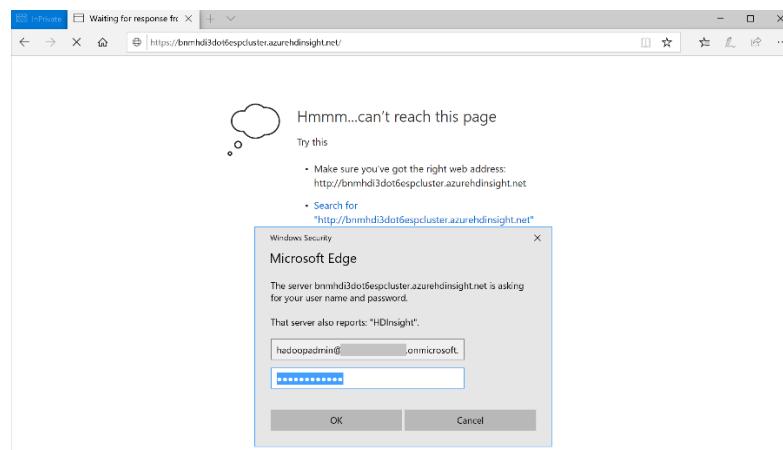


Figure 76: Retry accessing Ambari Web UI

- A successful login would land you at the Ambari Web UI – Dashboard view, as shown below:

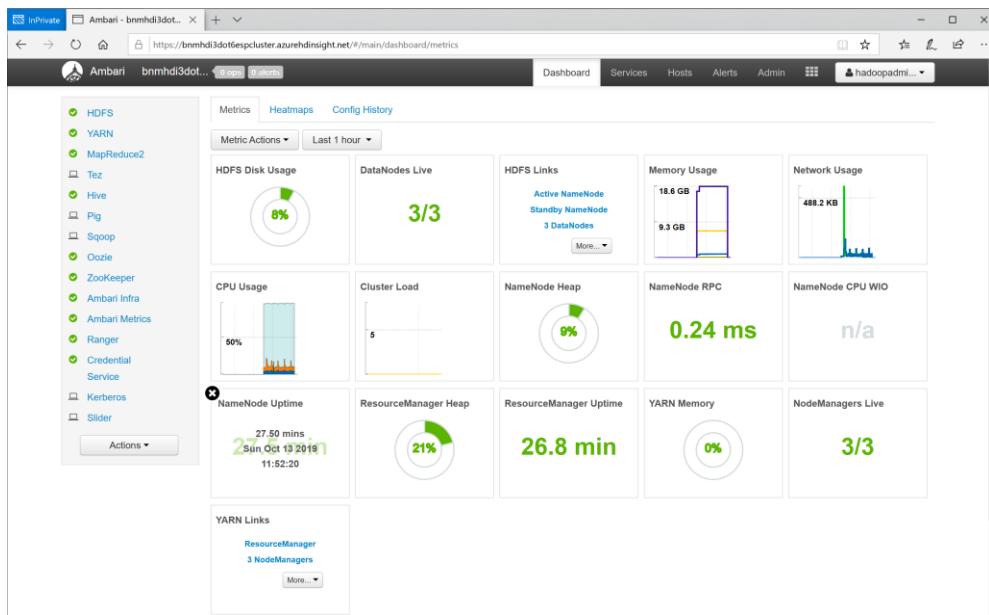


Figure 77: Ambari Dashboard View

- And, here is the visual reference to the HDFS service details:

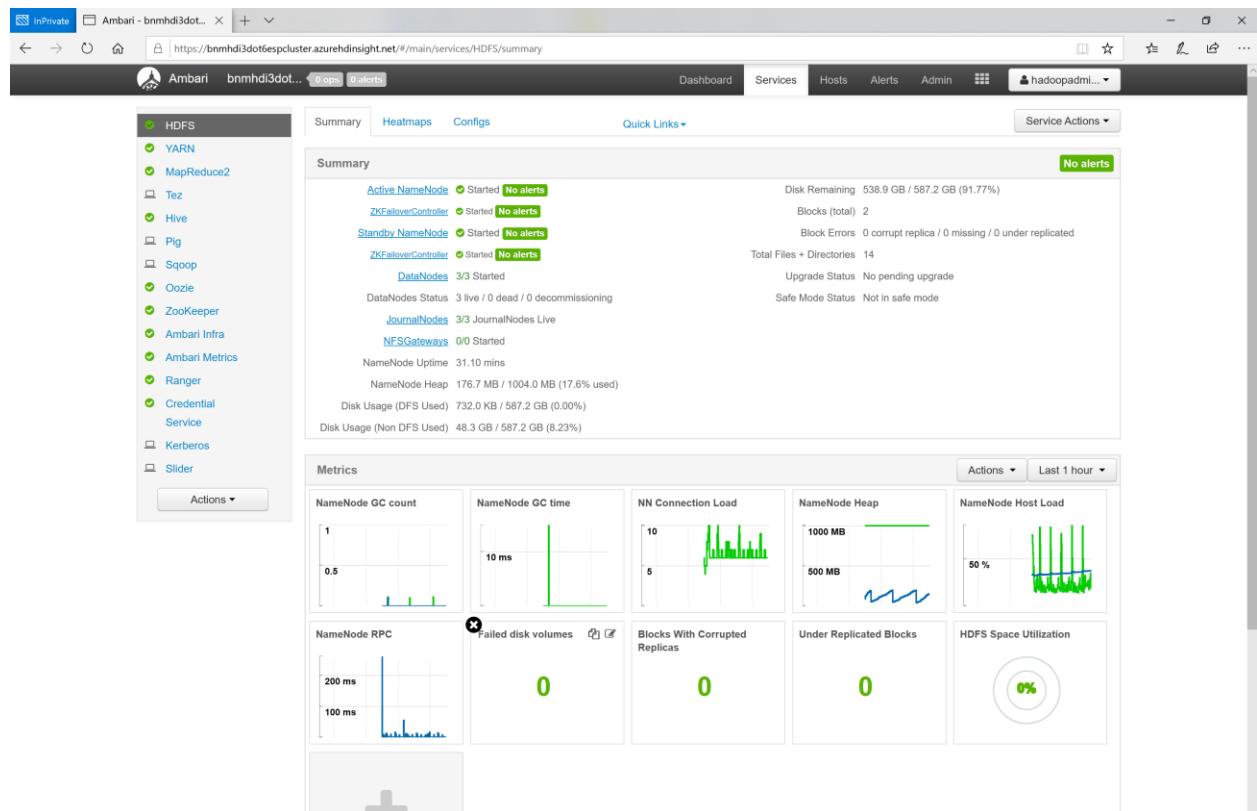


Figure 78: Ambari Web UI - HDFS Summary

## 6.4 Setup SSH-Tunneling to access Ambari URLs

To access resource URL's such as Namenode, Job History, etc., we must be configure SSH-tunneling. Before we attempt SSH-tunneling, the SSH port 22 must be enabled for in-bound traffic using the NSG rules. Such configuration must be limited to destination IP address range of 10.0.0.0/24 and protocol support limited to TCP. See below an entry for quick reference:

511	HDInsightSecureSSHAcess	22	TCP	Any	10.0.0.0/24	<input checked="" type="checkbox"/> Allow	...
...	...	...	...	...	...	...	...

Figure 79: Enable SSH access on NSG

Click [here](#) to know more about methods to enable SSH-tunnel. For this experiment here, will use SSH-tunneling approach, where in a tunnel or port forwarding will be established on localhost @ 9876. All calls will be routed to destination FQDN – bnmhd3dot6espcluster-ssh.azurehdinsight.net, which is the SSH hostname for the head node.

Following is the reference CLI output:

```

kalyan@myhost-name:/mnt/c/Users/userid$ ssh -C2qTnNv -D 9876 sshuser@bnmhdi3dot6espcluster-
ssh.azurehdinsight.net
OpenSSH_7.4p1 Debian-10+deb9u6, OpenSSL 1.0.2r 26 Feb 2019
debug1: Reading configuration data /home/kalyan/.ssh/config
debug1: Reading configuration data /etc/ssh/ssh_config
debug1: /etc/ssh/ssh_config line 19: Applying options for *
debug1: Connecting to bnmhdi3dot6espcluster-ssh.azurehdinsight.net [137.135.47.123] port 22.
debug1: Connection established.
debug1: identity file /home/kalyan/.ssh/id_rsa type 1
debug1: key_load_public: No such file or directory
debug1: identity file /home/kalyan/.ssh/id_rsa-cert type -1
debug1: key_load_public: No such file or directory
debug1: identity file /home/kalyan/.ssh/id_dsa type -1
debug1: key_load_public: No such file or directory
debug1: identity file /home/kalyan/.ssh/id_dsa-cert type -1
debug1: key_load_public: No such file or directory
debug1: identity file /home/kalyan/.ssh/id_ecdsa type -1
debug1: key_load_public: No such file or directory
debug1: identity file /home/kalyan/.ssh/id_ecdsa-cert type -1
debug1: key_load_public: No such file or directory
debug1: identity file /home/kalyan/.ssh/id_ed25519 type -1
debug1: key_load_public: No such file or directory
debug1: identity file /home/kalyan/.ssh/id_ed25519-cert type -1
debug1: Enabling compatibility mode for protocol 2.0
debug1: Local version string SSH-2.0-OpenSSH_7.4p1 Debian-10+deb9u6
debug1: Remote protocol version 2.0, remote software version OpenSSH_7.2p2 Ubuntu-4ubuntu2.8
debug1: match: OpenSSH_7.2p2 Ubuntu-4ubuntu2.8 pat OpenSSH* compat 0x04000000
debug1: Authenticating to bnmhdi3dot6espcluster-ssh.azurehdinsight.net:22 as 'sshuser'
debug1: SSH2_MSG_KEXINIT sent
debug1: SSH2_MSG_KEXINIT received
debug1: kex: algorithm: curve25519-sha256@libssh.org
debug1: kex: host key algorithm: ecdsa-sha2-nistp256
debug1: kex: server->client cipher: aes128-ctr MAC: umac-64-etm@openssh.com compression:
zlib@openssh.com
debug1: kex: client->server cipher: aes128-ctr MAC: umac-64-etm@openssh.com compression:
zlib@openssh.com
debug1: expecting SSH2_MSG_KEX_ECDH_REPLY
debug1: Server host key: ecdsa-sha2-nistp256 SHA256:9xy78s6sdMRCCPLcPGSuypodYstkuD2dPYgijtyyto
The authenticity of host 'bnmhdi3dot6espcluster-ssh.azurehdinsight.net (137.135.47.123)' can't be
established.
ECDSA key fingerprint is SHA256:9xy78s6sdMRCCPLcPGSuypodYstkuD2dPYgijtyyto.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'bnmhdi3dot6espcluster-ssh.azurehdinsight.net' (ECDSA) to the list of
known hosts.
debug1: rekey after 4294967296 blocks
debug1: SSH2_MSG_NEWKEYS sent
debug1: expecting SSH2_MSG_NEWKEYS
debug1: SSH2_MSG_NEWKEYS received
debug1: rekey after 4294967296 blocks
debug1: SSH2_MSG_EXT_INFO received
debug1: kex_input_ext_info: server-sig-algs=<rsa-sha2-256,rsa-sha2-512>
debug1: SSH2_MSG_SERVICE_ACCEPT received
Authorized uses only. All activity may be monitored and reported.
debug1: Authentications that can continue: publickey,password
debug1: Next authentication method: publickey
debug1: Offering RSA public key: /home/kalyan/.ssh/id_rsa
debug1: Authentications that can continue: publickey,password

```

```

debug1: Trying private key: /home/kalyan/.ssh/id_dsa
debug1: Trying private key: /home/kalyan/.ssh/id_ecdsa
debug1: Trying private key: /home/kalyan/.ssh/id_ed25519
debug1: Next authentication method: password
sshuser@bnmhdi3dot6espcluster-ssh.azurehdinsight.net's password:
debug1: Enabling compression at level 6.
debug1: Authentication succeeded (password).
Authenticated to bnmhdi3dot6espcluster-ssh.azurehdinsight.net ([137.135.47.123]:22).
debug1: Local connections to LOCALHOST:9876 forwarded to remote address socks:0
debug1: Local forwarding listening on ::1 port 9876.
debug1: channel 0: new [port listener]
debug1: Local forwarding listening on 127.0.0.1 port 9876.
debug1: channel 1: new [port listener]
debug1: Requesting no-more-sessions@openssh.com
debug1: Entering interactive session.
debug1: pledge: network
debug1: client_input_global_request: rtype hostkeys-00@openssh.com want_reply 0

```

The next step is the configure browser's proxy settings. For the context will use Firefox Browser, where in the proxy settings can be managed using the Browser properties itself. Unlike, Edge or Chrome where the proxy settings are delegated to the user's laptop network configurations. Following visual provides a sample view of how to set the SOCKS5 proxy configuration:

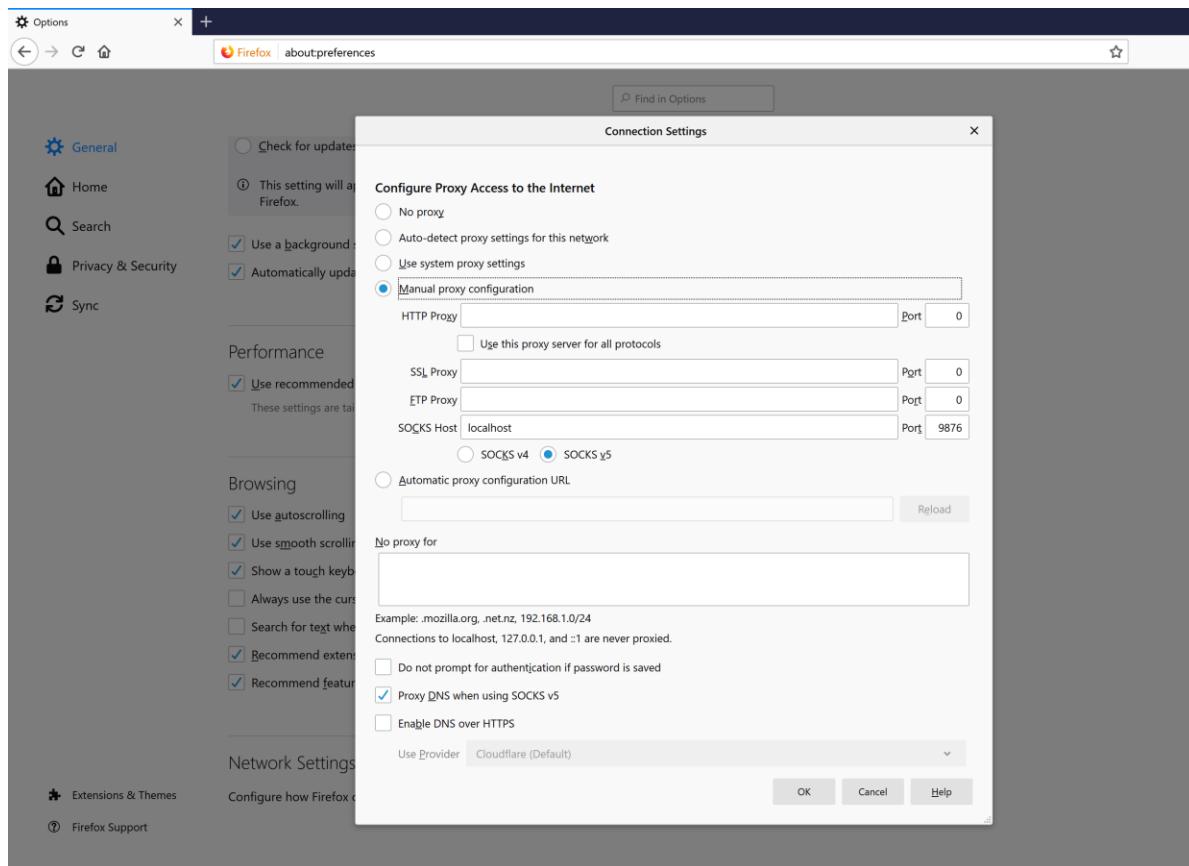


Figure 80: Firefox - SOCKS5 Configuration

## 6.5 Validate SSH-Tunnel

Now that you have updated the proxy settings, check if the tunneling works by accessing the site - <https://www.whatismyip.com/>. The web page should return the IP address of one of the Microsoft data center addresses, as shown in the following visual reference:

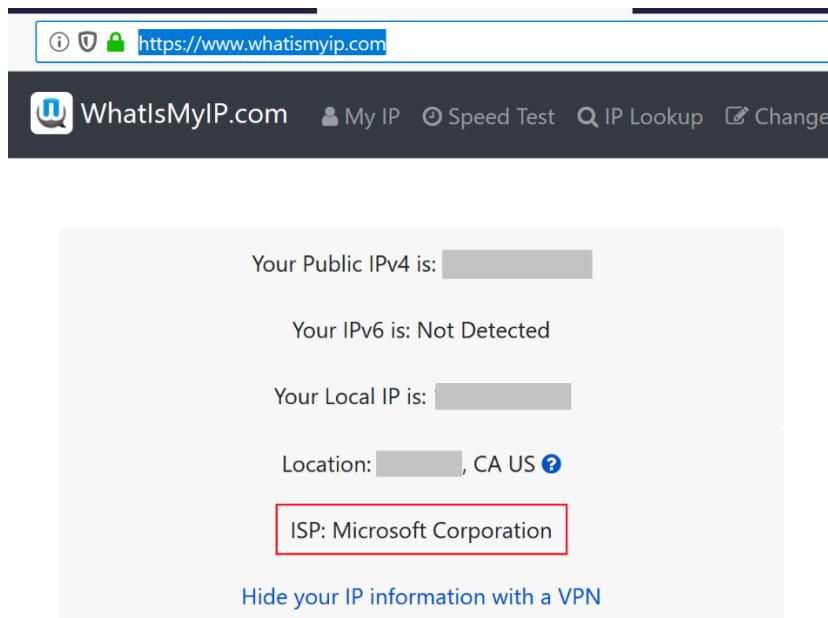


Figure 81: SOCKS5 Proxy Enabled - Whatismyip.com verification

Since we have started the SSH-tunnel (port forwarding) in debug mode, you should see console output that relates to browser activity. Here are the URLs that you can try to connect to the Ambari UI on the head node:

- Access via SSH-tunnel (HTTP) – secure over SSH:
  - <http://hn0-bnmhdi.youraddnsdomainname.onmicrosoft.com:8080>
- Access via public internet (HTTPS) – secure HTTP:
  - <https://bnmhdi3dot6espcluster.azurehdinsight.net>

Let's try with the SSH-tunnel bound HTTP URL. Note all traffic is encrypted and routed through the secure tunnel (port forwarding mechanism). When accessed the browser will prompt you to login to the Ambari UI, as shown in the following visual reference:

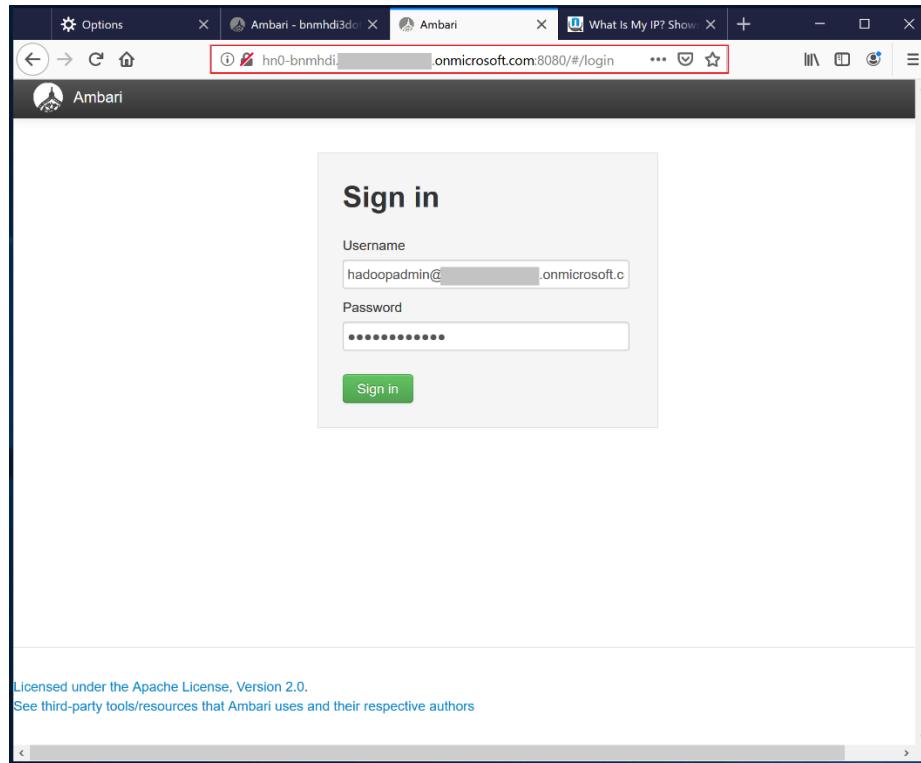


Figure 82: Ambari Web UI access using SSH-tunneling (port-forwarding)

## 6.6 Known Issues – Web UI

At the time, this document is being developed, there are few noted issues associated with ability to access Namenode UI. These are detailed below:

- The links in the Namenode UI does not work when logged in to Ambari UI using the HTTPS route (public internet). The user name provided in the login prompt is in the format – [hadoopadmin@youraddsdomainname.onmicrosoft.com](mailto:hadoopadmin@youraddsdomainname.onmicrosoft.com). Here's the visual reference of the landing page view:



Figure 83: HDInsight - Ambari - Namenode UI (Known Issue)

- SSH tunneling provides an encrypted connection to the target head-node in the HDInsight cluster. Hence, the URL to the Ambari portal will be different as described in the validate

section above. However, the user cannot login to the portal using the complete identity, i.e., [hadoopadmin@youraddsdomainname.onmicrosoft.com](mailto:hadoopadmin@youraddsdomainname.onmicrosoft.com). You can only use the name portion of it i.e., *hadoopadmin*.

- When logged into Ambari using the SSH-tunneling approach, the login works successfully as described in the previous point. However, you cannot access the HDFS Namenode UI page. The UI throws an error that is like the following visual:

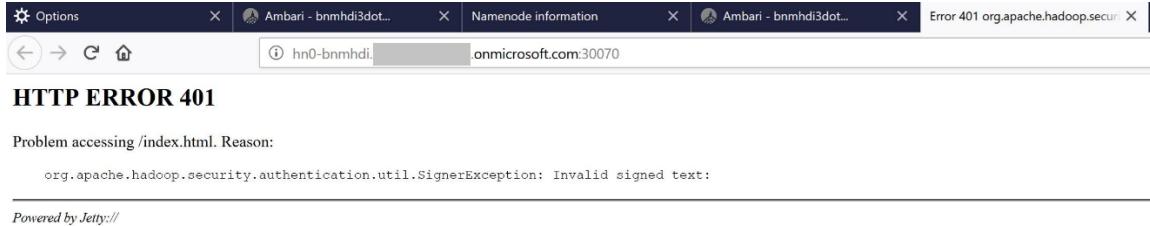
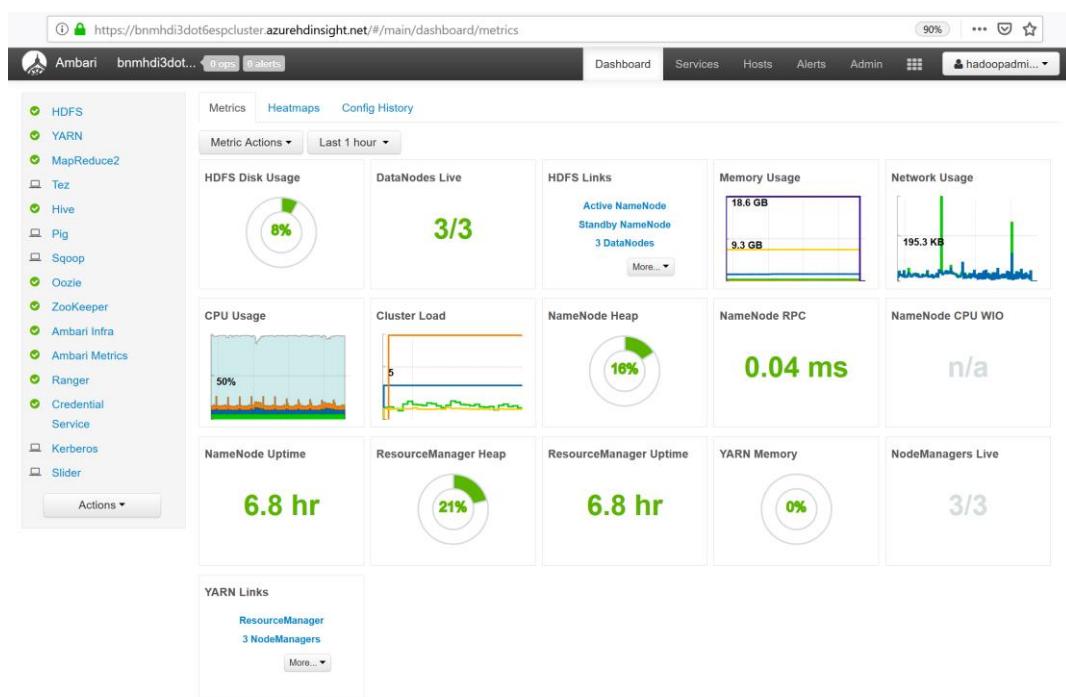


Figure 84: Ambari Web UI - Using SSH-tunnel - HTTP Error 401

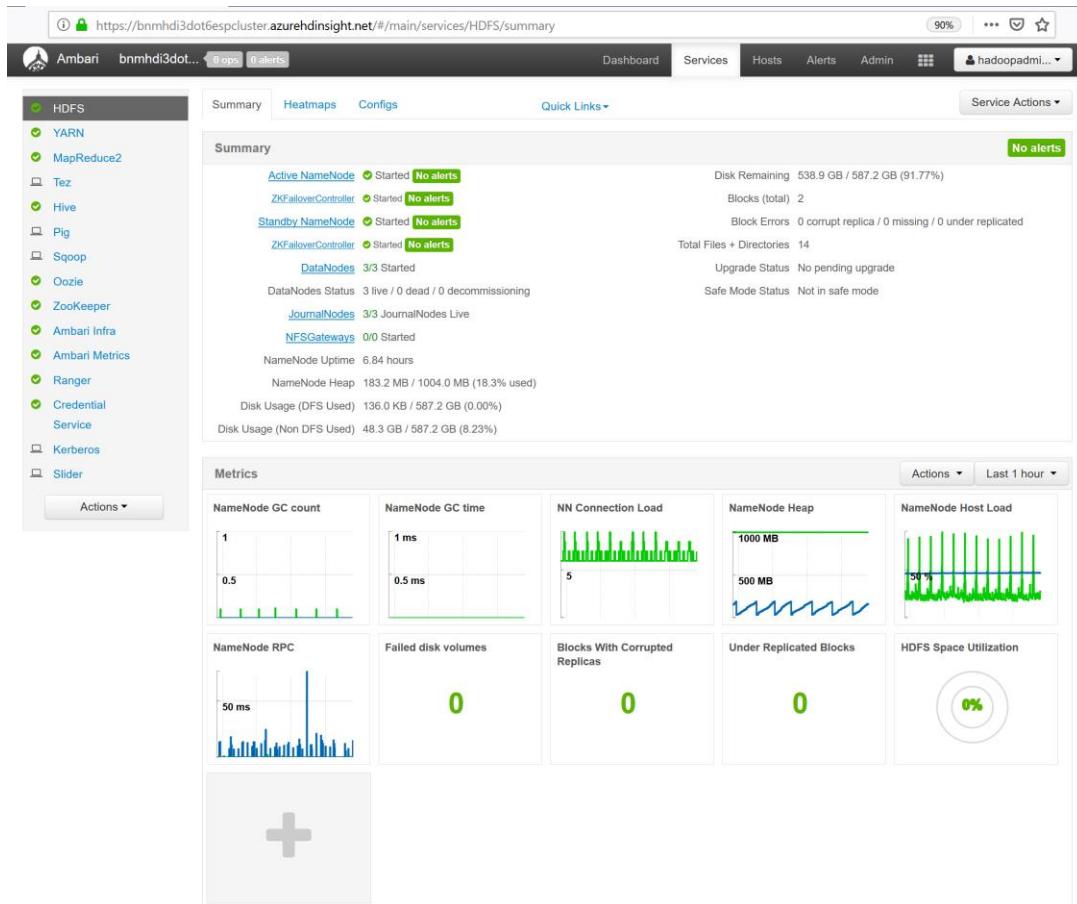
## 7 Post Installation Review

We have observed few issues accessing Namenode UI or the ability to access YARN Resource Manager UI (when using the SSH-tunnel/port-forwarding route), specifically using the Quick Links section. However, one can validate each of the services from the service specific link that is available on the landing page. The previously noted issues do not cause hindrance to access actions that you as an administrator can perform.

### 7.1 Ambari Landing Page



## 7.2 HDFS Service Summary



## 7.3 YARN Summary

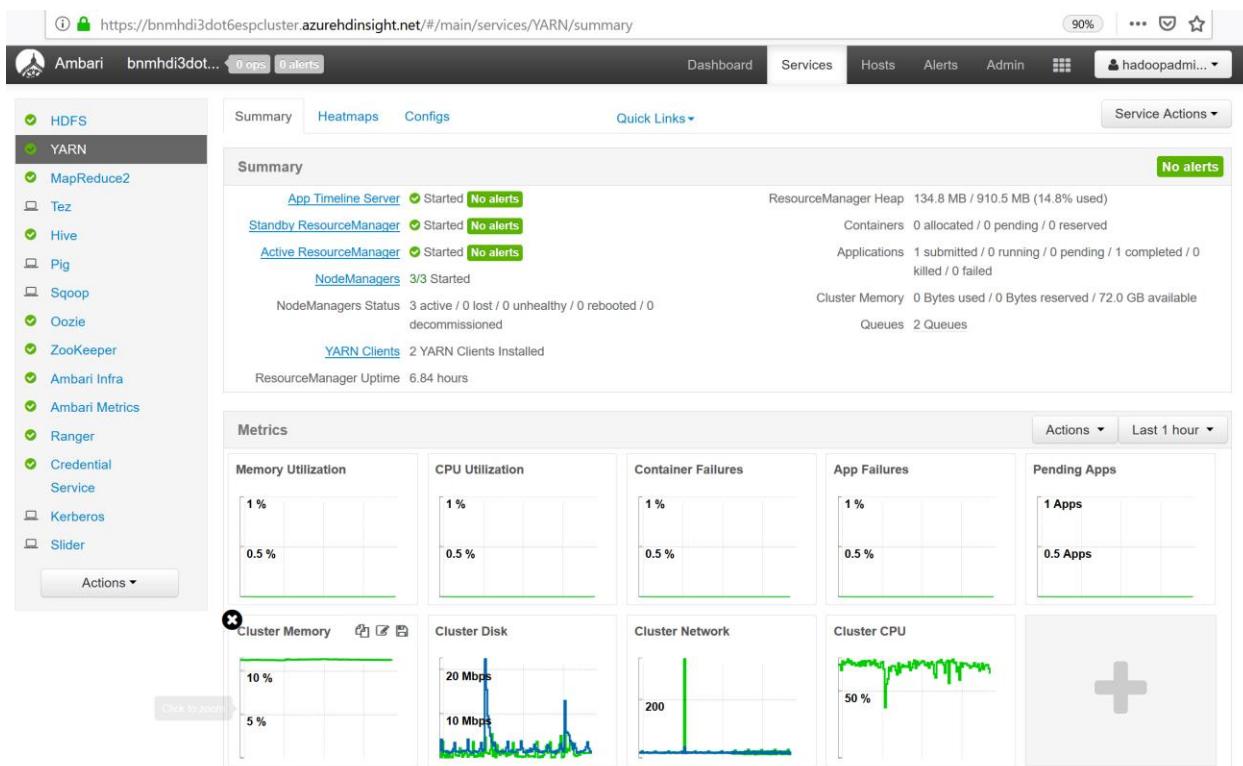


Figure 85: HDInsight YARN Summary

## 7.4 Ranger

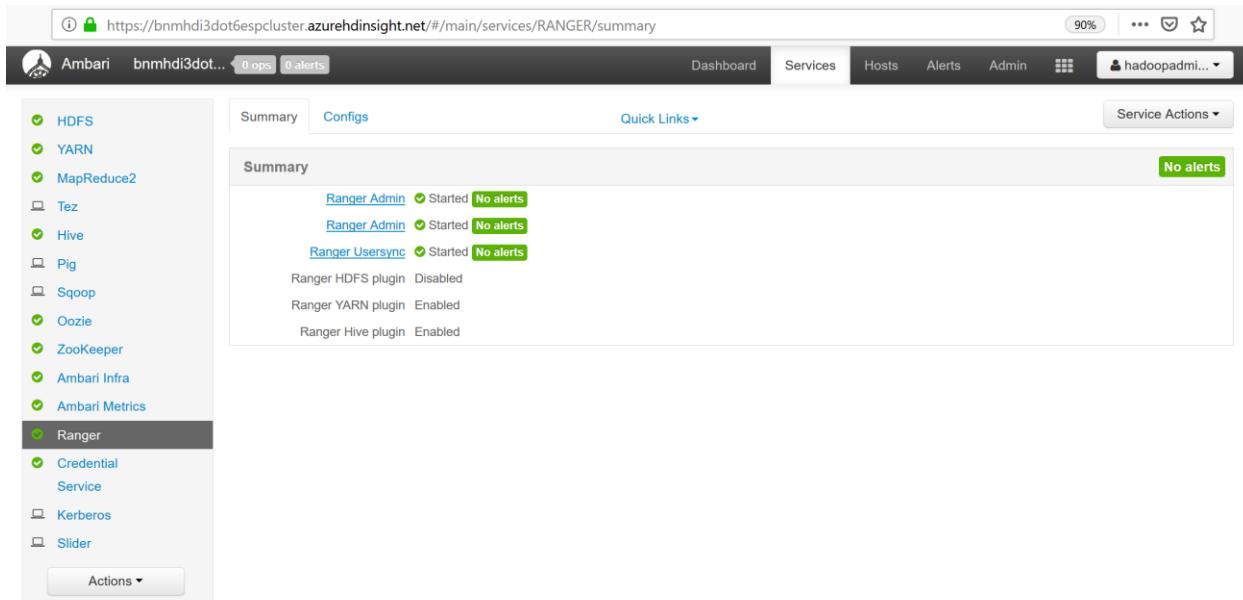


Figure 86: HDInsight - Ranger Overview

The screenshot displays two main sections of the HDInsight - Ranger - Hive Overview interface.

**Service Manager:**

- HDFS:** +, Import, Export
- HBASE:** +, Import, Export
- HIVE:** +, Import, Export, bnmhdi3dot6espcluster\_hive
- YARN:** +, Import, Export, bnmhdi3dot6espcluster\_yarn
- KNOX:** +, Import, Export
- STORM:** +, Import, Export
- SOLR:** +, Import, Export
- KAFKA:** +, Import, Export
- ATLAS:** +, Import, Export
- WASB:** +, Import, Export

Licensed under the Apache License, Version 2.0

**List of Policies:** bnmhdi3dot6espcluster\_hive

Policy ID	Policy Name	Status	Audit Logging	Groups	Users	Action
1	all - url	Enabled	Enabled	--	hdsp12d41d640a5314e, rangerlookup, ambari-qa, hadoopadmin	<input checked="" type="checkbox"/> <input type="button" value="Edit"/>
2	all - database, table, column	Enabled	Enabled	--	hdsp12d41d640a5314e, rangerlookup, ambari-qa, hadoopadmin	<input checked="" type="checkbox"/> <input type="button" value="Edit"/>
3	all - hiveservice	Enabled	Enabled	--	hdsp12d41d640a5314e, rangerlookup, ambari-qa, hadoopadmin	<input checked="" type="checkbox"/> <input type="button" value="Edit"/>
4	all - database, udf	Enabled	Enabled	--	hdsp12d41d640a5314e, rangerlookup, ambari-qa, hadoopadmin	<input checked="" type="checkbox"/> <input type="button" value="Edit"/>
6	hivesamptable_policy	Enabled	Enabled	public	hadoopadmin	<input checked="" type="checkbox"/> <input type="button" value="Edit"/>

Figure 87: HDInsight - Ranger - Hive Overview

**Policy Details :**

- Policy Type: Access
- Policy ID: 2
- Policy Name \*: all - database, table, column
- Audit Logging: YES
- Description: Policy for all - database, table, column

**Allow Conditions :**

Select Group	Select User	Permissions	Delegate Admin
Selected Group	hdisp12d41d640a5314e rangerlookup ambari-qa hadoopadmin	select update Create Drop Alter Index Lock All Read	<input checked="" type="checkbox"/> <input type="checkbox"/>

Save Cancel Delete

Figure 88: HDInsight - Ranger - Hive Sample Policy

## 7.5 Storage explorer – HDInsight Container view

NAME	LAST MODIFIED	CONTENT TYPE	SIZE
ams	10/13/2019, 11:31:53 AM	Folder	0 B
ambase	10/13/2019, 11:31:53 AM	Folder	0 B
app-logs	10/13/2019, 11:31:53 AM	Folder	0 B
apps	10/13/2019, 11:31:53 AM	Folder	0 B
atshistory	10/13/2019, 11:31:53 AM	Folder	0 B
example	10/13/2019, 11:54:19 AM	Folder	0 B
hbase	10/13/2019, 11:31:53 AM	Folder	0 B
HdiSamples	10/13/2019, 11:55:49 AM	Folder	0 B
hdp	10/13/2019, 11:54:31 AM	Folder	0 B
hive	10/13/2019, 11:31:53 AM	Folder	0 B
mapred	10/13/2019, 11:31:53 AM	Folder	0 B
mapreducestaging	10/13/2019, 11:53:53 AM	Folder	0 B
mr-history	10/13/2019, 11:31:53 AM	Folder	0 B
ranger	10/13/2019, 11:50:29 AM	Folder	0 B
tezstaging	10/13/2019, 11:54:08 AM	Folder	0 B
tmp	10/13/2019, 11:31:53 AM	Folder	0 B
user	10/13/2019, 11:54:24 AM	Folder	0 B

Figure 89: HDInsight Storage Container View using Storage Explorer for hadoopadmin

## 8 Appendix – Reference URLs

- [Plan a virtual network for Azure HDInsight](#)
- [Directly connect to Apache Hadoop Services](#)
- [Configure DNS Zone for external access](#)
- [Remote Server Administration Tools](#)
- [Tutorial: Configure secure LDAP for an Azure Active Directory Domain Services managed domain - Test Queries to the managed domain](#)
- [Tutorial: Enable password synchronization in Azure Active Directory Domain Services for hybrid environments](#)
- [HDInsight management IP addresses](#)
- [Ports used by Apache Hadoop services on HDInsight](#)
- [Introduction to Azure Data Lake Storage Gen2](#)
- [Create and Authorize a managed identity](#)
- [What is Azure Private Endpoint?](#)
- [Use Azure Data Lake Storage Gen2 with Apache Hadoop in Azure HDInsight](#)
- [Azure Active Directory – Domain Services \(AAD-DS\)](#)
- [Troubleshoot cluster creation failures with Azure HDInsight](#)
- [Manage access to Azure resources using RBAC and the Azure Portal](#)
- [Best practice for using RBAC](#)
- [Users of HDInsight clusters with ESP](#)
- [Synchronize AAD users with Ambari REST API \(immediate synchronization\)](#)
- [Use SSH tunneling](#)

## 9 Feedback and suggestions

If you have feedback or suggestions for improving this data migration asset, please contact the Data SQL Ninja Team ([askdmjfordmtools@microsoft.com](mailto:askdmjfordmtools@microsoft.com)). Thanks for your support!

Note: For additional information about migrating various source databases to Azure, see the [Azure Database Migration Guide](#).