IBM Cloud Pak for Data Express on Azure

Deployment Guide

March 2024

Overview

IBM Cloud Pak for Data Express is a set of three pre-built, pre-sized offerings designed to address problems in cataloging, analyzing and integrating data. The IBM Cloud Pak for Data Express offerings give you a choice of three popular data express starting points: IBM Data Governance Express for a datacatalog, IBM Data Science and MLOps Express for analytics and modeling, ELT Pushdown Express for high performance data transformations. Each provides pre-sized, pre-selected services designed to address a current data express need.

This deployment guide provides step-by-step instructions for deploying IBM Cloud Pak for Data Express solutions powered by Cloud Pak for Data onto Azure RedHat OpenShift (ARO) managed RedHat OpenShift Container Platform clusters on the Azure Cloud. You can automatically deploy a production instance of Cloud Pak for Data. These instructions enable you to automatically deploy the Cloud Pak for Data Express services required to implement an instance of these offerings.

IBM Cloud Pak for Data Express on Azure:

IBM Cloud Pak for Data Express offerings use Azure service and feature, including Azure Red Hat OpenShift (ARO) cluster, which includes Azure Virtual Machine, Azure Virtual Network, and Azure Managed Disks. It provides you the opportunity to create or configure - Master and Worker Subnets, Availability Zone, Availability Sets, Custom DNS, Security groups, and User-defined routing to build a reliable and scalable cloud platform.
ELT Pushdown Express is a modern, cloud native and secure data integration solution running on ARO that enables you to create, deploy, and observe trusted data pipelines that use your existing data warehouse for high performance data transformations.
Data Governance Express offers a unified data and AI platform running on ARO. The solution builds a foundation for data governance with business glossary and data cataloging.
Data Science and MLOps Express offers an affordable data modelling and AI platform running on ARO. The solution provides a mix of IBM and open-source innovations, allowing data scientists or other users to operate in their choice of environment (including Python, R, and SPSS). A full spectrum of capabilities from data exploration and modelling through to production usage of AI, including drift and bias detection, allows you to adapt the solution to fit the needs of your business.
This reference deployment provides Azure ARM templates to deploy Data Express Solutions on Cloud Pak for Data onto a new OpenShift cluster. This cluster includes:
- A Red Hat OpenShift Container Platform cluster created in a new or existing Vnet on Red Hat CoreOS (RHCOS) instances, using the [Redhat OpenShift Installer Provisioned Infrastructure](). See the [OpenShift Container Platform Installation overview]() for details about the underlying OpenShift deployment architecture.
- A highly available storage infrastructure with Red Hat OpenShift Data Foundation (ODF) and Portworx Enterprise.
- Scalable OpenShift compute nodes running Cloud Pak for Data services.

For more information about Cloud Pak for Data, see the [IBM Documentation.]()

Cost and licenses

The IBM Cloud Pak for Data Express solution environment is deployed by using Azure ARM template. You are responsible for the cost of the Azure services used for the infrastructure. The Azure ARM template for this deployment includes configuration parameters that you can customize. You can use it to build a new Vnet for your IBM Cloud Pak for Data Express solution on Azure cluster or deploy on an existing Azure Vnet. Some of these settings, such as instance type, will affect the cost of deployment. For cost estimates, see the pricing pages for each Azure service you will be using. Prices are subject to change.

For more information about licensing terms, see the Cloud Pak for Data Express software license agreement.

- [IBM Data Governance Express](#)
- [IBM Data Science and MLOps Express](#)
- [IBM ELT Pushdown Express](#)

Architecture

Deploying the Azure Marketplace template for a new VNet with **default parameters** builds the following Cloud Pak for Data environment in the Azure Cloud.

Figure 2: Deployment architecture for Cloud Pak for Data on Azure

The ARM template sets up the following:
- A highly available architecture that spans up to three Availability Zones.*
- A Virtual network configured with public and private subnets.*
- In the public subnets:
    - a bastion host to allow inbound Secure Shell (SSH) access to compute instances in private subnets.
- In the private subnets:
    - OpenShift Container Platform master instances.
    - OpenShift Container Platform (OCP) compute nodes that combined, contain Cloud Pak for Data Collect, Organize, and Analyse services.
- An Azure Load Balancer spanning the public subnets for accessing Cloud Pak for Data from a web browser.
- Storage disks with Azure Managed Disk mounted on compute nodes for ODF (OpenShift Data Foundation) v4.12.
- An Azure domain as your public Domain Name System (DNS) zone for resolving domain names of the IBM Cloud Pak for Data management console and applications deployed on the cluster.

* The template that deploys into an existing Virtual network skips the components marked by asterisks and prompts you for your existing Virtual network configuration.

Cloud Pak for Data microservices are preconfigured on compute nodes. The following diagram shows the platform architecture.
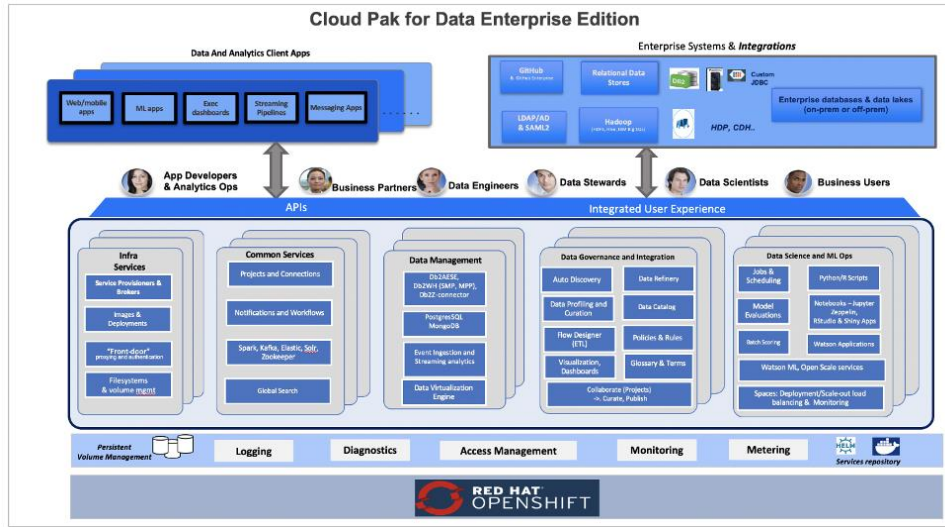

Figure 3: Cloud Pak for Data services

Planning the deployment

Specialized knowledge

This deployment assumes basic familiarity with Cloud Pak for Data components and services. If you're new to Cloud Pak for Data and Red Hat OpenShift, see the Additional resources section. This deployment also assumes familiarity with the OpenShift command line interface and Linux, in addition to a moderate level of familiarity with Azure services.

Technical requirements

For Cloud Pak for Data requirements, see System requirements for Cloud Pak for Data. Red Hat Enterprise Linux CoreOS (RHCOS) is used for the OpenShift compute node instances in this deployment.

Before you launch the template, your account must have resource quotas as specified in the following table.

Resources	If necessary, request service quota increases for the following resources. You might need to do this if an existing deployment uses these resources, and you might exceed the default quotas with this deployment. The Service Quotas console displays your usage and quotas for some aspects of some services. For more information, see the Azure documentation.

Resource	This deployment uses
Virtual Network	1
Public IP addresses	3
Network Load Balancers	2
Standard_D4s_v3 virtual machines (Bootnode)	1
Standard_D8s_v3 virtual machines (Master nodes)	3
Standard_D16s_v3 virtual machines (Compute nodes)	3

	Standard_D16s_v3 virtual machines (Compute nodes)	3(ODF)
	Public IP addresses for ODF storage	2
Regions	This deployment includes 3 Availability Zones (AZ), which isn't currently supported in all Azure Regions. So, please make sure you select the region which supports 3 AZ.	
IAM permissions	To deploy the template, you must log in to the Azure portal with Azure Identity and Access Management (IAM) permissions for the resources and actions the templates will deploy. The *AdministratorAccess* managed policy within IAM provides sufficient permissions, although your organization may choose to use a custom policy with more restrictions.	

Deployment options

This template provides the following deployment options:

- **Deploy Cloud Pak for Data into a new Virtual Network** (end-to-end deployment). This option builds a new Azure environment consisting of the Virtual Network, subnets, NAT gateways, Network Security Groups (NSG), bastion hosts, and other infrastructure components, and then deploys Cloud Pak for Data into this new Virtual Network.
- **Deploy Cloud Pak for Data into an existing Virtual Network**. This option provisions Cloud Pak for Data in your existing Virtual Network infrastructure.

The template also lets you configure CIDR blocks, Virtual Machine types, and Cloud Pak for Data settings, as discussed later in this guide.

Pre-requisites

Ensure the following pre-requisites are in place with your existing Cloud Pak for Data entitlements.

Step 1. IBM Cloud Pak for Data Subscription

When you purchase Cloud Pak for Data from Marketplace, you will get the Cloud Pak for Data entitlement Username and API Key. You should keep it handy as it's a required parameter in the ARM Template

Note: Please make sure you have valid entitlement key before using the offering, you can check your entitlement key is valid or not here. Login to myibm.ibm.com and refer to the name(s) below when accessing your container software as shown in the below screenshot.
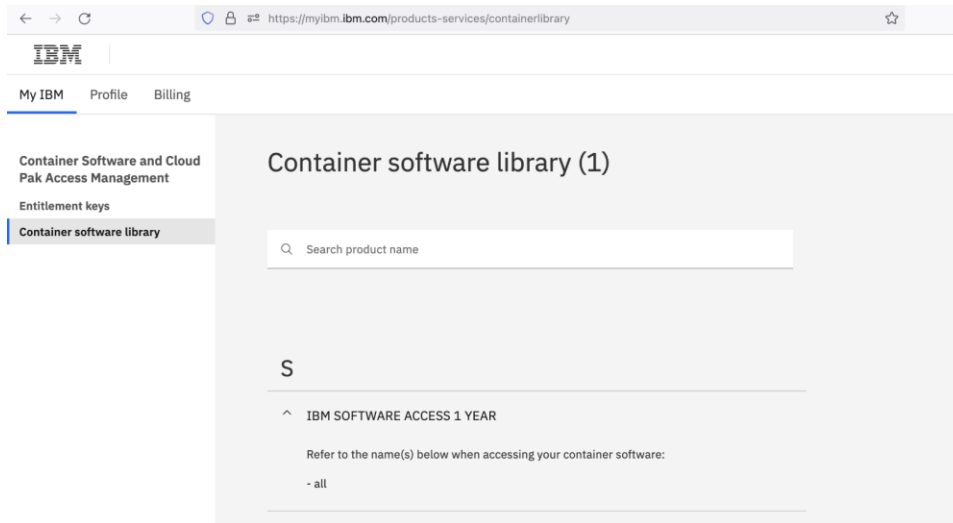
Step 2. Red Hat Subscription

Ensure that you have the Red Hat OpenShift pull secret with your purchase. You should keep it handy as this json is required parameter in the ARM Template. If you are deploying through Quickstart Template or from Azure Marketplace, you can directly enter the pull secret's json value.

Step 3. Storage Subscription

- OpenShift Data Foundation (ODF): The Red Hat ODF license is linked as a separate entitlement to your RedHat subscription. This is the preferred option on Azure Marketplace.
- Portworx: When you select Portworx as the persistent storage layer, you will need to specify the install spec from your Portworx account. You can generate a new spec using the Spec Generator. Note that the Portworx trial edition expires in 30 days after which you need to upgrade to an Enterprise Edition. Here are the steps to genereate Portworx Spec.

Step 4. Sign in to your Azure portal account

1. Sign in to your Azure account at https://portal.azure.com/ with an Azure Identity and Access Management (IAM) user role that has the necessary permissions. For details, see Planning the deployment, earlier in this guide.
2. Make sure that your Azure account is configured correctly, as discussed in Technical requirements, earlier in this guide.
3. Use the Region selector in the navigation bar to choose the Azure Region where you want to deploy Cloud Pak for Data on Azure. An IBM Cloud Pak for Data high availability deployment is restricted to Azure Regions with at least three Availability Zones.
4. The following resources should be made available for Cloud Pak for Data deployment
   - A new or an existing Key Vault location with Red Hat pull secret

Step 5. Azure Service Principal, with Contributor and User Access Administrator Role

The Service Principal can be created by running the azure CLI commands from any host where Azure CLI is installed.

1. Create Service Principal, using your Azure Subscription ID, and save the returned json:
   - az login
   - az ad sp create-for-rbac --role="Contributor" --scopes="/subscriptions/<subscription_id>"
2. Assign User Access Administrator roles, using the App Id:
   - az role assignment create --role "User Access Administrator" --assignee "*<app_id* " --scopes="/subscriptions/<subscription_id>"

Save the ClientID and ClientSecret from above steps. You need this information for the deployment.

If you face any issues while creating Azure Service Principal Client ID, Please refer Microsoft Official Documentation to create Azure Service Principal Client ID and Client Secret

Launch the Deployment

> **Note:** The instructions in this section reflect the current version of the Azure portal. If you're using the redesigned portal, some of the user interface elements might be different.
> You are responsible for the cost of the Azure services used while running this deployment. For full details, see the pricing pages for each Azure service you will be using for this deployment. Prices are subject to change.

1. Launch the Azure ARM template into your Azure account from the Azure Marketplace directly or you can download the template and launch it separately from your account. A Cloud Pak for Data Express standard deployment takes about 3 hours, and a high availability (HA) deployment takes about 4 hours to complete.

2. Create a new Resource Group or specify existing one. This is where all the deployment resources will be stored. There will be two resource group created through ARM templates as specified in the parameters section below. **Note:** Resource group should be unique and same name shouldn't be specified it earlier. The ARM template may fail if the same name resource group created earlier in the past.

3. Check the Region that's displayed in the Region field and change it if necessary. This is where the network infrastructure for Cloud Pak for Data will be built. The template is launched in the 'West US 2' Region by default.

4. Specify Service Principal App ID (Client ID) and Azure AD Client Secret. For more information check the Step 5. It has steps to create service principal and get the client id and client secret.

5. For the Infrastructure settings, OpenShift settings, and Cloud Pak settings pages, review the parameters for the template. Provide values for the parameters that require input. For all other parameters, review the default settings and customize them as necessary.

In the following tables, parameters are listed by category and described.
When you finish reviewing and customizing the parameters, choose **Next**.

PARAMETERS FOR DEPLOYING CLOUD PAK FOR DATA INTO A NEW OR EXISTING VIRTUAL NETWORK

Basic configuration:

Parameter label (name)	Default	Description
Subscription	Microsoft Azure Enterprise	The subscription is necessary to deploy cluster and create resources on the Azure portal. Make sure you have required administrator permission to create cluster resources, secrets, users, etc.
Resource Group	*Requires input*	Resource group is a container that holds related resource for an Azure Solution. Create a new resource group with the unique name (never used it earlier). There will be two Resource group created one with the name of the Resource Group and another one with the cluster name. The Cluster name Resource Group will hold all master and worker nodes related resources. E.g., Virtual Machines, Network Security Gateway (NSG), etc.
Region	West US 2	This is where the network infrastructure for Cloud Pak for Data will be built. The template is launched in the 'West US 2' Region by default.
Service Principal Azure App ID	—	This client ID will be used to create resources on the Azure portal. Check how to create Service Principal Azure App ID
Azure AD Client Secret	—	This client secret will be used to create resources on the Azure portal. Check how to get Azure Client Secret

Infrastructure configuration:

Parameter label (name)	Default	Description
Bootnode Public IP (Attach Public IP to BootnodeVM)	true	Bootnode can be accessible through SSH connection if bootnode has the public IP.
SSH public key	*Requires input*	Your machine's SSH public key will be added in the authorized_key in the bootnode so that you will have the permission to SSH into bootnode.

		Note: *Please store this SSH keypair in the safe place in order to access this bootnode in the future.*
Number of master nodes (NumberOfMaster)	3	The desired capacity for the OpenShift master instances. Must be an odd number. For a development deployment, 1 is sufficient; for production deployments, a minimum of 3 is required.
Bootnode VM size (BootnodeType)	Standard_D4s_v3	The virtual machine type for the OpenShift bootnode VM.
Master VM size (MasterInstanceType)	Standard_D8s_v3	The virtual machine type for the OpenShift master VMs.
Compute VM size (NodesInstanceType)	Standard_D16s_v3	The virtual machine type for the OpenShift compute VMs.

Virtual Network configuration:

Parameter label (name)	Default	Description
Virtual Network	*Requires input*	Create a new Virtual Network or select existing VNet
BootNode Subnet (Bootnode Subnet CIDR)	*Requires input*	Subnets for virtual network
Master Subnet (Master Subnet CIDR)	*Requires input*	Subnets for virtual network
Worker Subnet (Worker Subnet **CIDR**)	*Requires input*	Subnets for virtual network
Single or Multi Zone	Multi Zone	Deploy VMs to Single Zone region or Multiple Zone (Availability Zones). Recommended option is MultiZone. Please make sure that selected region supports AvailabilityZones: https://learn.microsoft.com/en-us/azure/reliability/availability-zones-service-support#azure-regions-with-availability-zone-support

OpenShift hosts configuration:

Parameter label (name)	Default	Description
RedHat subscription pull secret. (RedhatPullSecret)	*Requires input*	Enter json value of the RedHat Openshift Pull Secret, if you are deploying cluster through Azure Marketplace.
Cluster prefix (ClusterName)	*Requires input*	Custom cluster name for kubernetes.io/cluster/tags. The cluster name must be *unique* and should never be used earlier because this name will be also assigned it to another cluster resource group and automation scripts are stored in ARM Staging Artifacts by using this cluster prefix name. This cluster resource group will hold all cluster nodes related resources.
ARO Resource Provider Principal ID (Only required for IBM CPD on ARO – BYOL)	*Requires input*	ARO cluster deployment needs Principal ID of the ARO Service Principal exist in your account in order to create ARO cluster in the given Azure account. Here is the command to fetch the Principal ID (ObjectId) of the ARO Service Principal, Azure CLI command: (Prereqs: Azure CLI installation in your computer.) ` az ad sp list --filter "displayname eq 'Azure Red Hat OpenShift RP'" --query "[?appDisplayName=='Azure Red Hat OpenShift RP'].{name: appDisplayName, objectId: id}"` **E.g.:**

		az ad sp list --filter "displayname eq 'Azure Red Hat OpenShift RP'" --query "[?appDisplayName=='Azure Red Hat OpenShift RP'].{name: appDisplayName, objectId: id}"
		
		This command or command group has been migrated to Microsoft Graph API. Please carefully review all breaking changes introduced during this migration: https://docs.microsoft.com/cli/azure/microsoft-graph-migration
		[
		  {
		    "name": "Azure Red Hat OpenShift RP",
		    "objectId": "464114b9-XXXX-XXXX-a068-35ab76dXXXXX"
		  }
		]
		Please make sure you have "Network Contributor" role assigned to ARO Resource Provider Principal ID
		Note: If this Azure CLI command doesn't work for you then you may need to register the required resource provider. Please follow instructions on this page and execute *"Get the service principal object ID for the Openshift resource provider – Azure CLI" command*: https://learn.microsoft.com/en-us/azure/openshift/quickstart-openshift-arm-bicep-template?pivots=aro-arm#register-the-required-resource-providers---azure-cli

Use Private or Public Endpoints (Public Cluster)	public	To Deploy a private cluster, select "private" and "public" for public cluster.
OpenShift Console Username (OpenShift Username)	—	ARO cluster creates OpenShift Username for us. It is accessible through Azure portal.
OpenShift Console Password (OpenShift Password)	—	ARO cluster creates OpenShift Username for us. It is accessible through Azure portal.
Egress Outbound Type	Load Balancer	ARO cluster creates Load Balancer for us.

Storage Configuration:

Parameter label (name)	Default	Description
Storage type for Cluster (StorageType)	*ODF or Portworx*	OpenShift Data Foundation (ODF) and Portworx storage options are available now. Recommended storage option is ODF for Azure Marketplace.
ODF instance type (ODFInstanceType)	Standard_D16s_v3	Update this value if Storage type selected is ODF. The Virtual Machine type for the ODF instances.
Number of ODF nodes (NumberOfODF)	3	The desired capacity for the ODF instances. Minimum of 3 is required. You don't need to add these 3 ODF nodes in worker nodes count as our template will add 3 ODF nodes for you. Note: These 3 nodes will only be created for ODF but not for the Portworx storage option.
Portworx Spec URL	—	Required field Only if you chose Portworx as storage option. Generated Spec URL for ARO cluster from Portworx spec generator: https://central.portworx.com/specGen/wizard. Here are the steps to genereate Portworx Spec.
Portworx CSI	true	Required field Only if you chose Portworx as storage option. Portworx CSI option to enable CSI environment. Default value for Portworx CSI is true. It is recommended to enable Portworx CSI option.

IBM Cloud Pak for Data configuration:

Parameter label (name)	Default	Description
IBM Cloud Pak for Data Entitled Registry API key Value (APIKey)	—	Enter the IBM Cloud Pak for Data API key to access IBM Container Registry
License agreement (LicenseAgreement)	*Reject*	I have read and agreed to the license terms for IBM Cloud Pak for Data that were provided to me at time of purchase. *You must accept the license to install Cloud Pak for Data services.*
OpenShift project (NameSpace)	zen	The OpenShift project that will be created for deploying Cloud Pak for Data. It can be any lowercase string.
IBM Cloud Pak for Data version (CPDVersion)	4.8.x	The default version of Cloud Pak for Data to be deployed.
IBM Cloud Pak for Data Entitled Registry User (APIUsername)	cp	Enter the IBM Cloud Pak for Data Username value to access IBM Container Registry.

IBM Cloud Pak for Data Express:

Cloud Pak for Data Express Solutions are implemented on IBM Cloud Pak for Data and use Data Science and MLOps Express, Data Governance Express, Cloud Pak for Data software license agreement. By default, each data express solution deploys the services required to deliver use cases within its area of functionality.

IBM Data Governance Express:

Parameter label (name)	Default	Description
Number of compute nodes (NumberOfNodes)	4	The desired capacity for the OpenShift node instances. Minimum of 4 nodes required for the default instance type *Standard_D16s_v3*. **Warning** If the number of node instances exceeds your Red Hat entitlement limits or Azure virtual machine quotas, the stack will fail. Choose a number that is within your limits.
IBM Data Governance Express		Installs the Watson Knowledge Catalog Core, and Analytics Engine services.

IBM Data Science and MLOps Express:

Parameter label (name)	Default	Description
Number of compute nodes (NumberOfNodes)	5	The desired capacity for the OpenShift node instances. Minimum of 5 nodes required for the default instance type *Standard_D16s_v3*. **Warning** If the number of node instances exceeds your Red Hat entitlement limits or Azure virtual machine quotas, the stack will fail. Choose a number that is within your limits or increase your compute quota for the selected region.
IBM Data Science and MLOps Express		Installs the Watson Studio, Watson Machine Learning, and Watson Openscale services.

Optional Services which can be installed with IBM Data Science and MLOps Express Offering:

Data Science and MLOps optional services are available for deployment but are not required for many common use cases. Each Optional Service below delivers additional capability, but consumes some cluster resources. Select them depending upon your business need.

Parameter label (name)	Default	Description
RStudio	No	Choose **Yes** to install the RStudio service.

13

Analytics Engine (AE)	No	Choose **Yes** to install the Analytics Engine (Apache Spark) service.
Decision Optimization (DO)	No	Choose **Yes** to install Decision Optimization service
SPSS Modeler	No	Choose **Yes** to install SPSS Modeler service.
Watson Pipelines	No	Choose **Yes** to install IBM Watson Pipelines service.

IBM ELT Pushdown Express:

Parameter label (name)	Default	Description
Number of compute nodes (NumberOfNodes)	3	The desired capacity for the OpenShift node instances. Minimum of 3 nodes required for the default instance type *Standard_D16s_v3*. **Warning** If the number of node instances exceeds your Red Hat entitlement limits or Azure virtual machine quotas, the stack will fail. Choose a number that is within your limits.
IBM ELT Pushdown Express		Installs DataStage service.

Optional Services which can be installed with IBM ELT Pushdown Express Offering:
ELT Pushdown Express Offering optional services are available for deployment but are not required for many common use cases. The Optional Service below delivers additional capability, but consumes some cluster resources. Select them depending upon your business need.

Parameter label (name)	Default	Description
Watson Pipelines	No	Choose **Yes** to install IBM Watson Pipelines service.

1. On the Openshift and Cloud Pak Settings page, select appropriate values from the above table.
2. On the **Review** + **Create** page, review and confirm the template settings. Make the necessary changes based on your requirement before deploying template.
3. Choose **Create** to deploy the templates.
4. On Azure Portal, monitor the status of the templates *Resource Group > Deployments*. When the status is succeeded for the CPD deployments, the Cloud Pak for Data cluster is ready.
5. Use the URLs displayed in the *Resource Group > Deployments **>** AzureRMSamples > Outputs*. The URL for the *"Cloud Pak for Data U*RL" output key will navigate to the console login page.

(Optional) Edit the network security group
Optional: You might need to edit the Azure network security group to add IP addresses that can access the Cloud Pak for Data web client.
Navigate to Load Balancers on your Azure portal and filter on tags, for example *kubernetes.io/service-name: openshift-ingress/router-default*.

1. In Load Balancers, filter and select the security group.

## 2. Select Security Group and modify the Inbound rules

3. Choose **Add Rule**, and fill in the rule details. For the rule **Type**, select either HTTP or HTTPS in the drop-down menu. Port 80 or 443 is filled in automatically. Add the network CIDR for the group of IP addresses that you want to permit HTTP or HTTPS access to the proxy nodes. To allow any IP address, use 0.0.0.0/0.
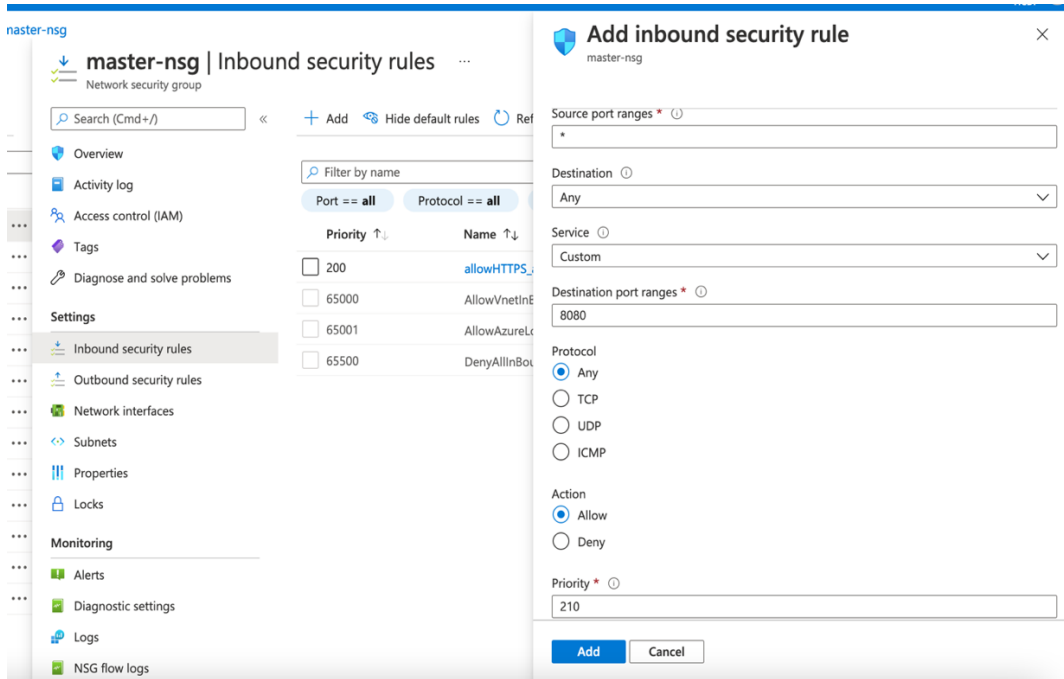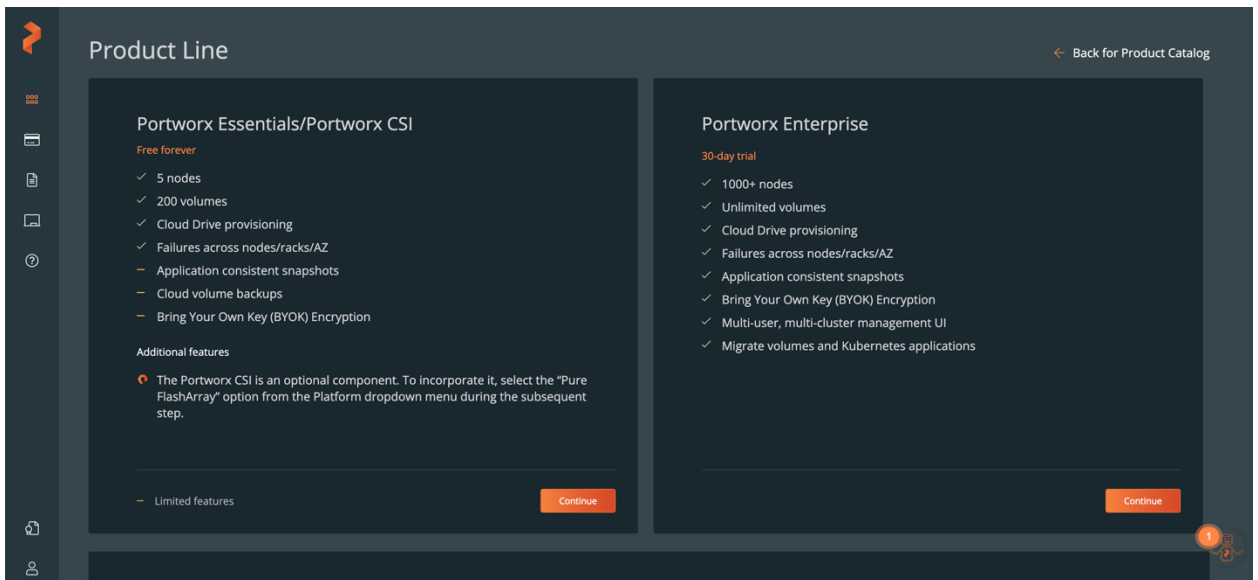
4. In the rule editor window, choose **Save**.

Generating the Portworx Spec URL for ARO cluster

1. Launch the spec generator

2. Select **Portworx Enterprise** and press *Continue*:
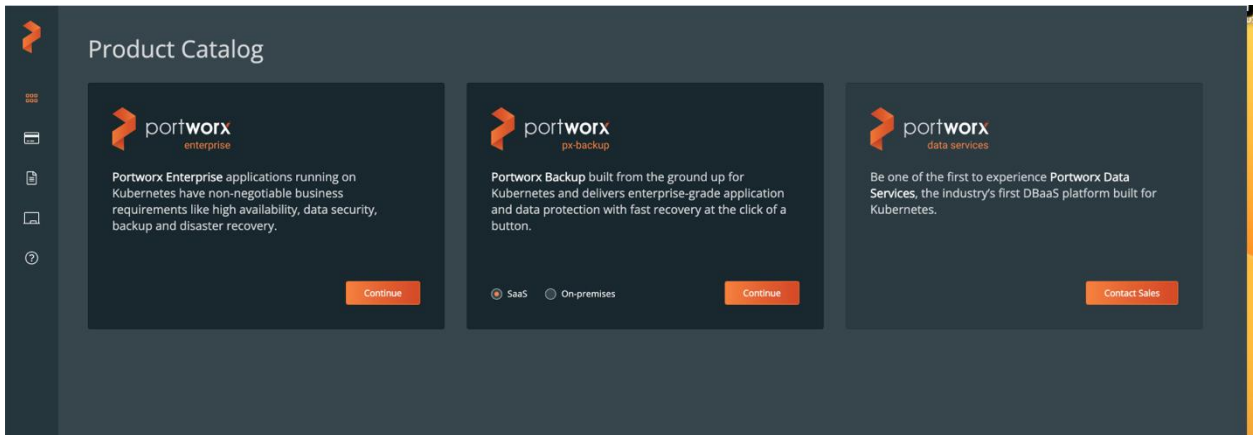
3. Select **Portworx Essentials** or **Portworx Enterprise** and press *Continue*:

4. In generate Spec page, Select Portworx version *2.13,* Select Platform *Azure* and Select Kubernetes Distribution Name ***Azure Red Hat OpenShift (ARO)***, Change Namespace name to ***kube-system*** and press *Customize*
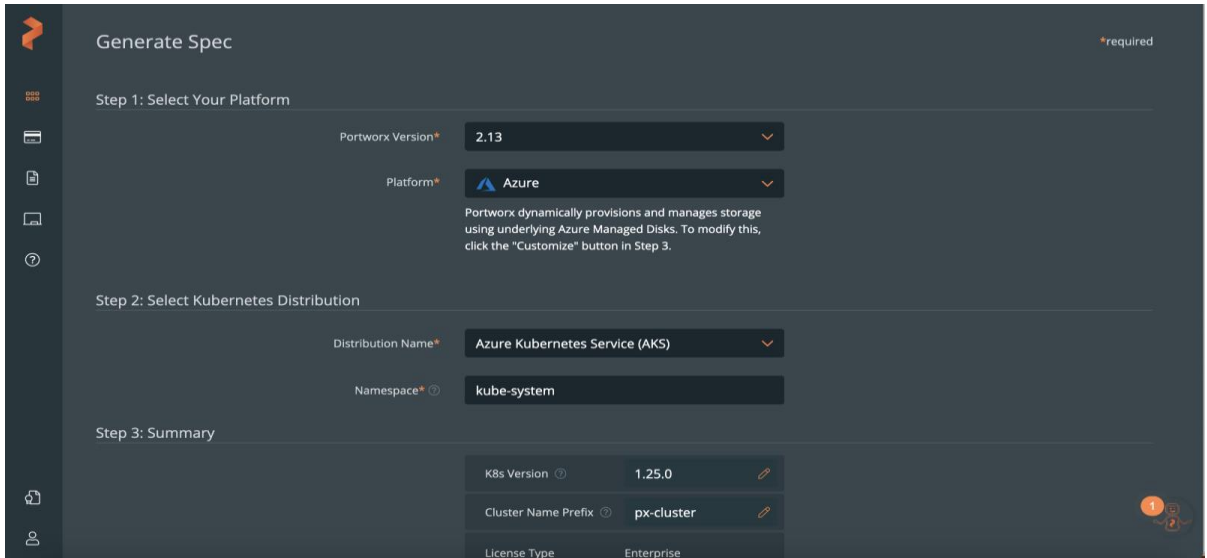
5. Check Use the Portworx Operator box, select the Portworx version as *2.13*, enter Namespace as *kube-system*, select *Built-in* and then press *Next*. (Note: Please make sure all selected options in Step 4 are reflected on this page)
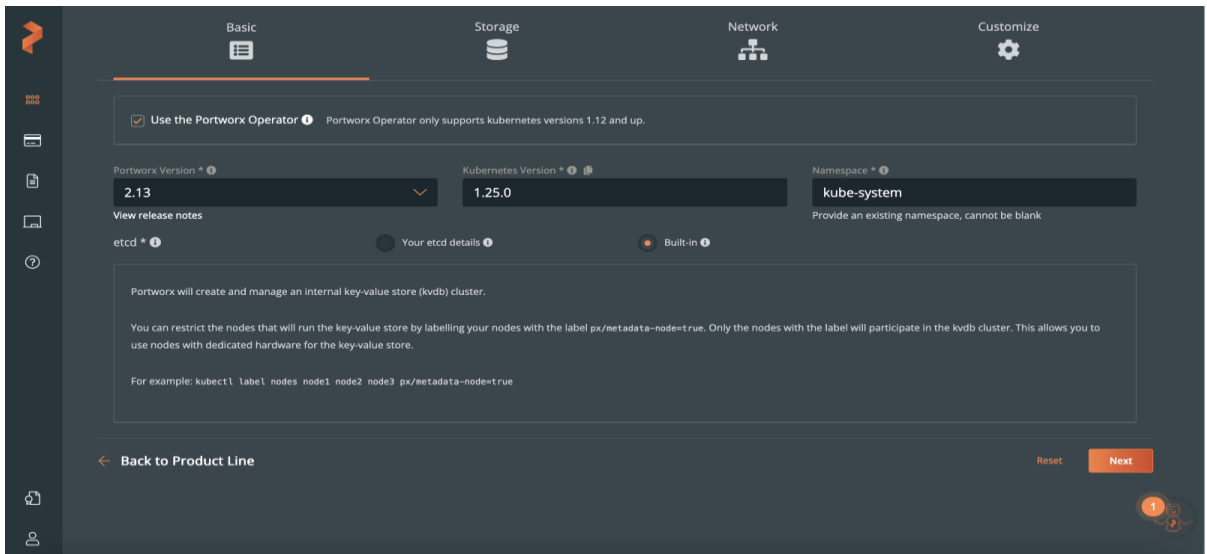
6. Select *Cloud* as your environment. Click on *Azure* and select *Create Using a Spec* option for Select type of disk. Enter value for Size (GB) as **500** and then press *Next*.
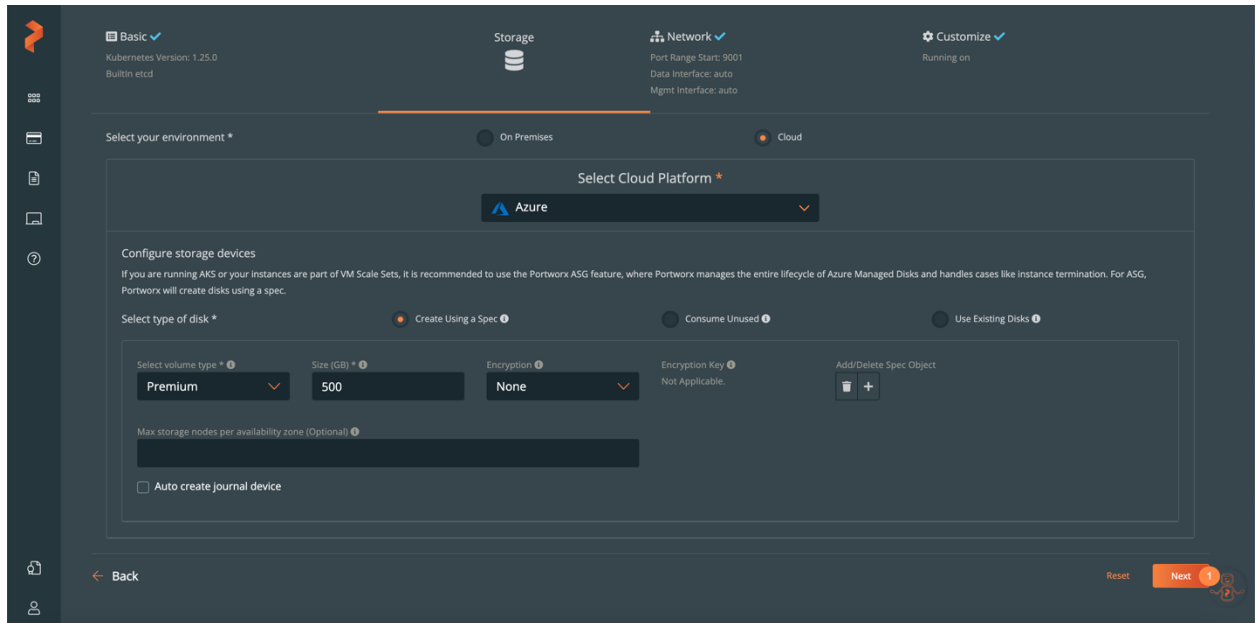
7. Keep the default option ***auto*** for the network interfaces and press *Next*:
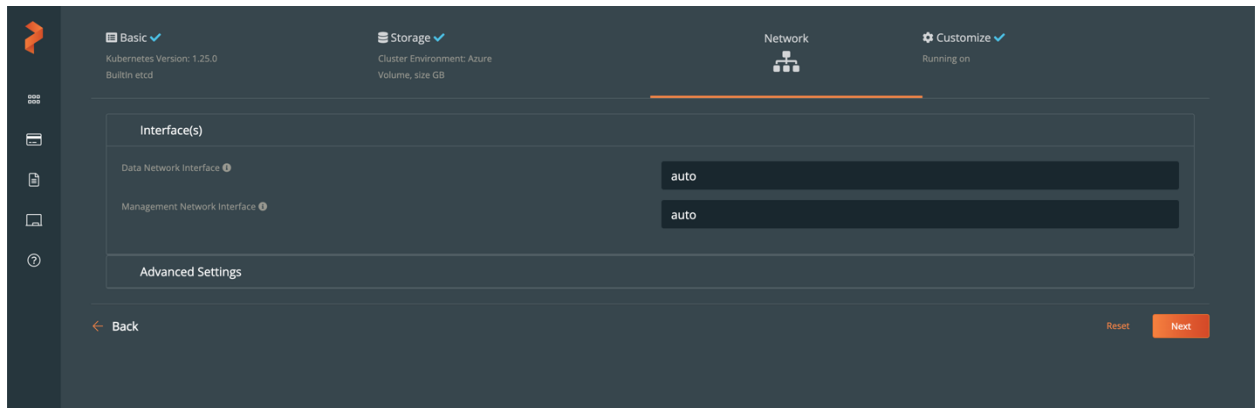
8. Select ***Azure Red Hat OpenShift (ARO)*** as Openshift environment, go to *Advanced Settings*:

(Note: if you don't find the ARO option in the Customize tab, please verify you have selected

Portworx 2.12 or later versions as it is only available for Portworx 2.12 or later versions).
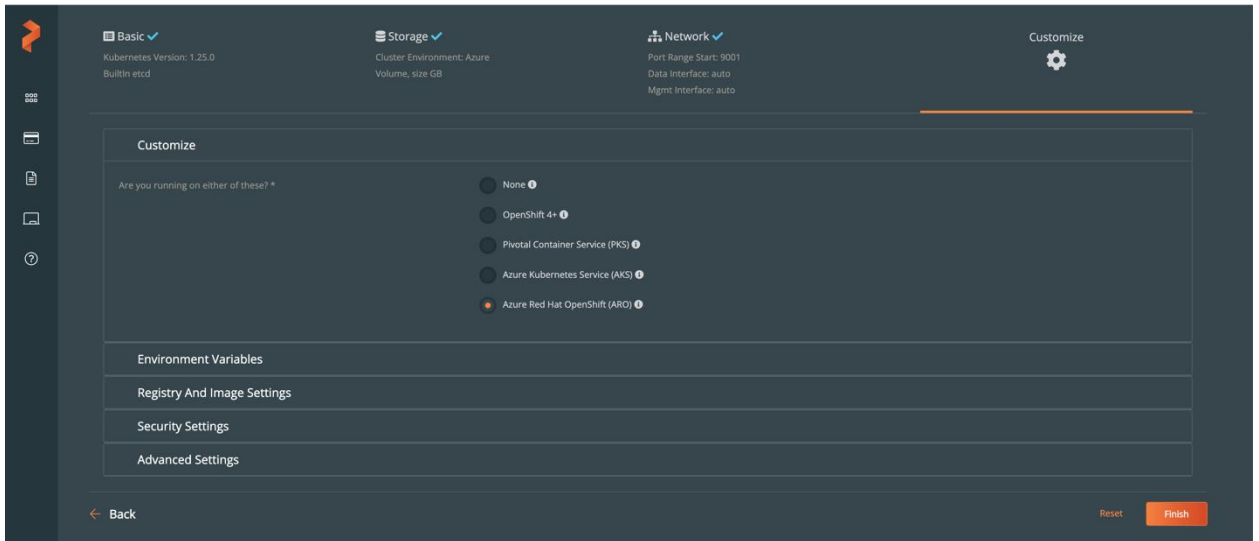
9. In the Advanced Settings tab: *Enable Stork, CSI, Monitoring and Telemetry* and press *Finish*:
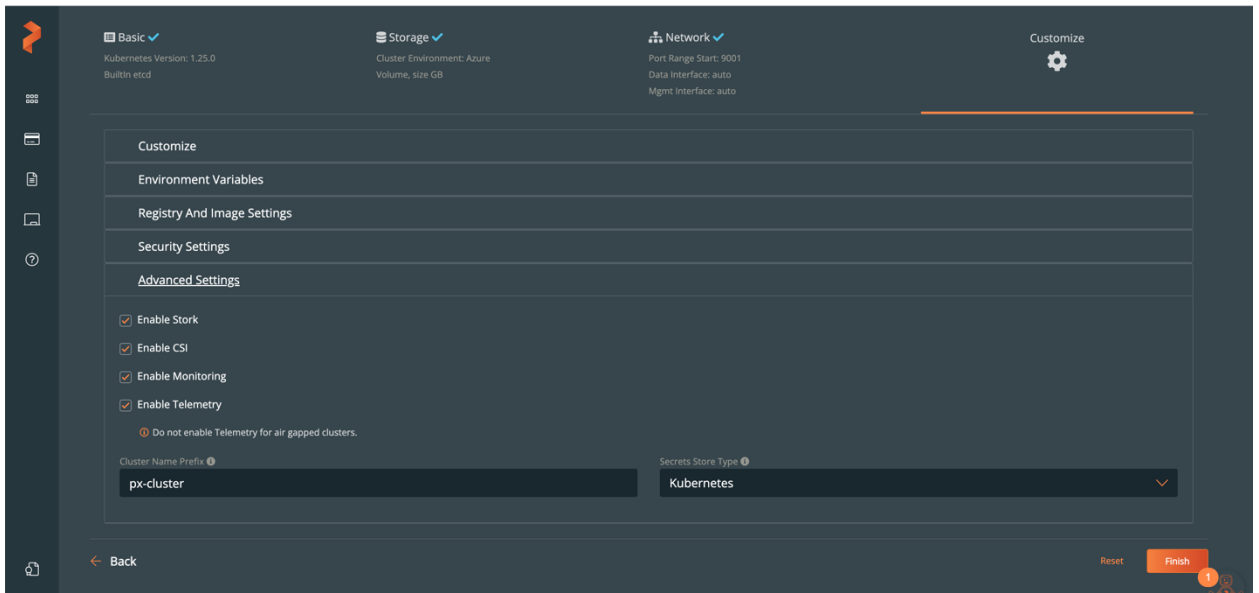
10. Copy the Spec URL and Save Spec in your Portworx account. (Note: Please copy only spec URL but not the whole command or single quote in order to avoid any errors.)
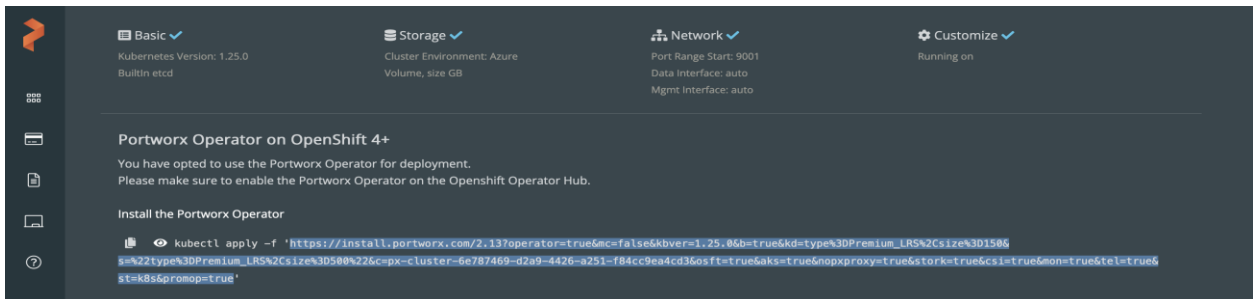
Manage your cluster using the OpenShift Console

1. To access the Openshift console, go to the overview section of the root stack.

2. In the overview section you can get **Openshift URL**, **username** and **password**, please select the Azure Redhat openshift cluster deployment from the **overview** section of root stack as shown in the below screenshot:
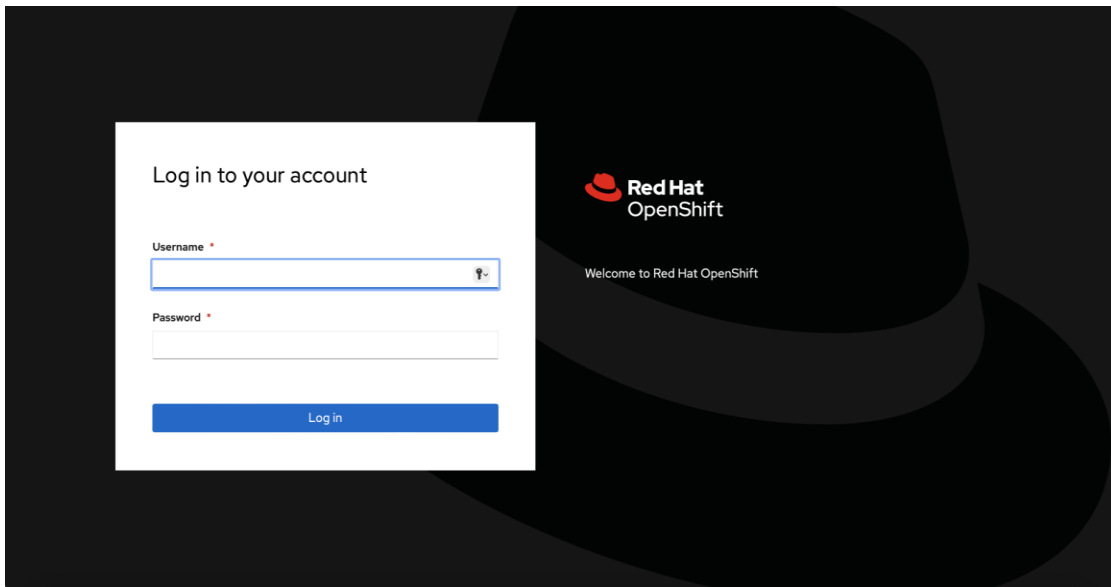
 Home > Resource Groups > Cluster's Resource groups > select Deployments > ibm-usa-ny-armonk-hq-6275750-ibmcloud-asperia.ibm-20231207151241 > overview
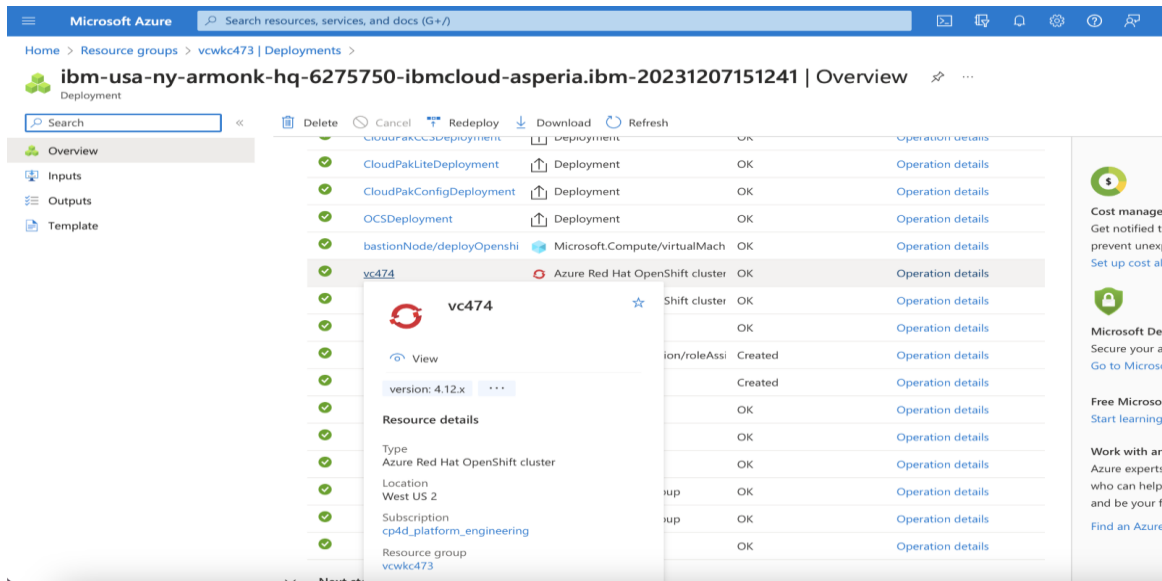
3. Click on the "Connect" and get the **openshift URL, username & password** from the right panel.
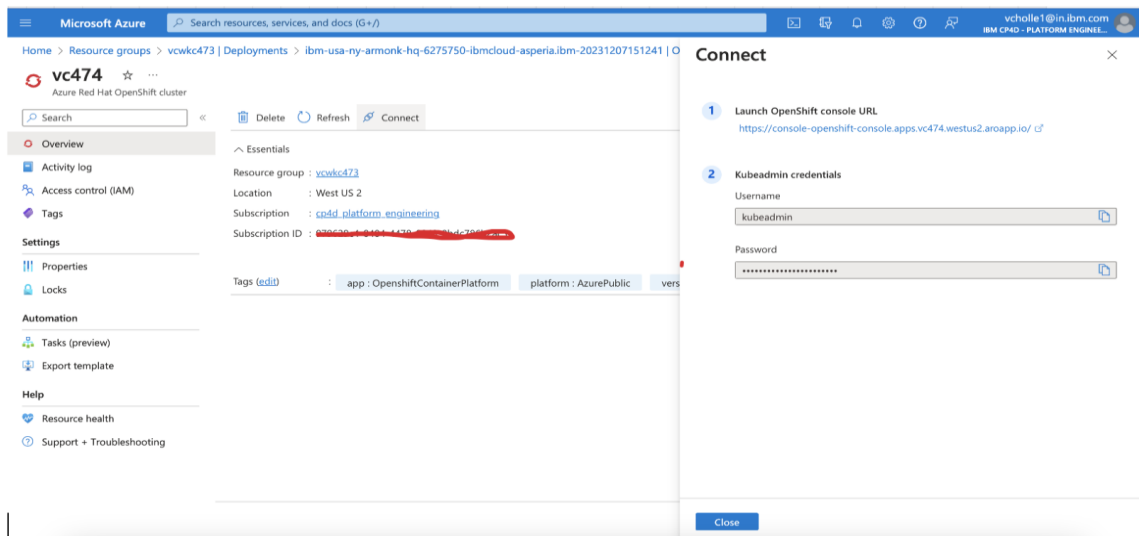
4. Open the OpenShift Console URL in a browser and Login with the **username** and **password** from the previous step.
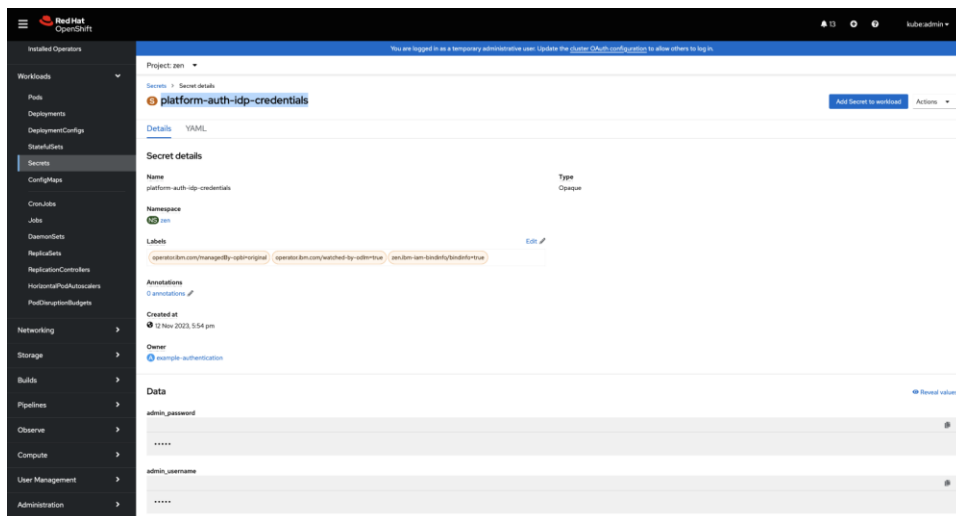
Login to Cloud Pak for Data web client

When the Azure ARM template has successfully created the stack, all server nodes will be running with the software installed in your Azure portal. In the following steps, connect to Cloud Pak for Data web client to verify the deployment, and then use the web client to explore Cloud Pak for Data features.

1. To access the Cloud Pak for Data web client, first get the console URL from the output for key **cloud Pak for Data URL**
    Home > Resource Groups > Cluster's Resource groups > select **Deployments > ibm-alliance-global-1560886.cloud_pak_for_data-20211118184734 > Outputs**


2. To get the password of Cloud Pak for Data web client, please login in to openshift console as mentioned in the previous step-6 and click on the workloads in the left pane
    Workloads > secrets > click on **platform-auth-idp-credentials >** copy the admin_username & admin_password under **Data.**

3. Log in to the Cloud Pak for Data web client by using the user "admin" and the password from the previous step.
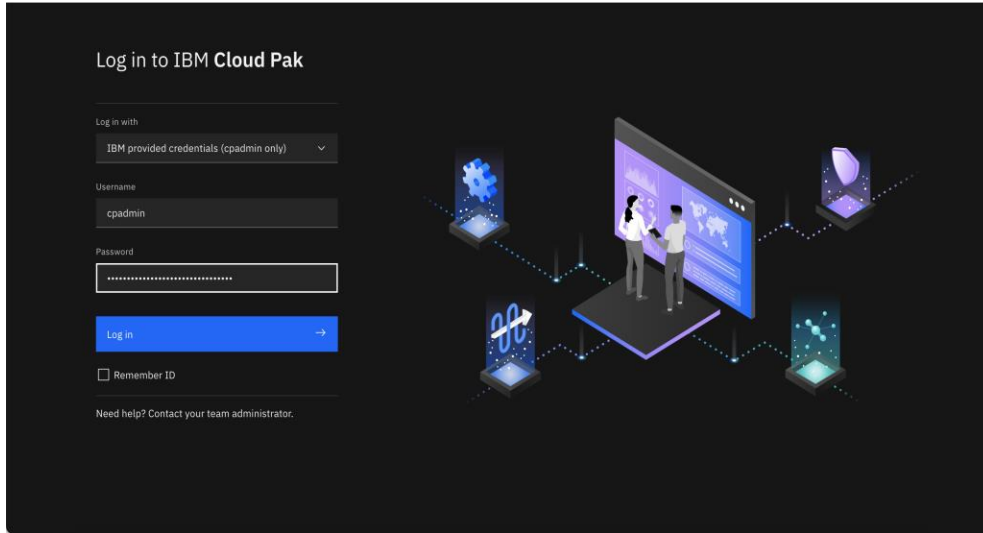
4. Once you log in, the welcome page is displayed.
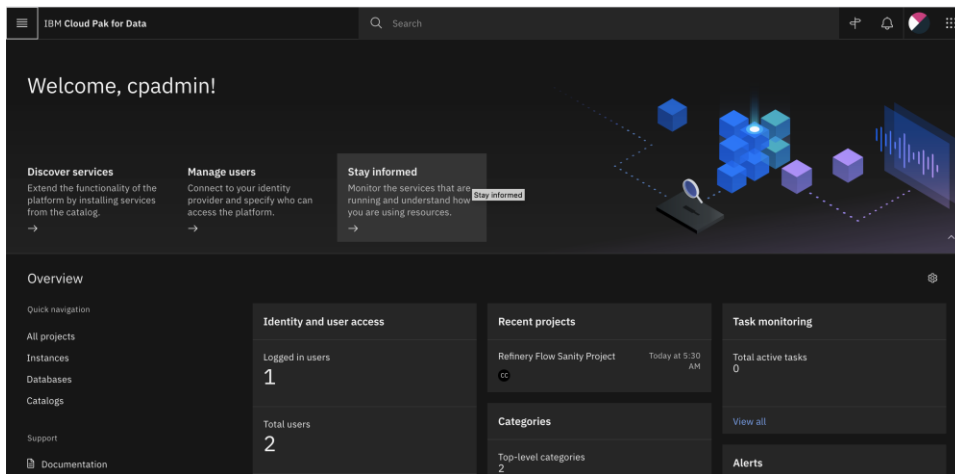
See resources on platform features and capabilities. For a list of supported browsers, see Supported browsers.

(Optional) Provide Boot Node SSH access

The boot node is used for certain command-line cluster administration tasks, such as adding compute nodes. SSH access to the boot node is required for some cluster administrators. After deployment, you only have access to the boot node. Provide the workstation IP address CIDR as the value of the network security group (nsg) inbound rule.

This section describes the steps to modify the network security group (nsg) inbound rules.

1. In the Azure cluster's **Resource Groups** page, select name containing **bastion-nsg**
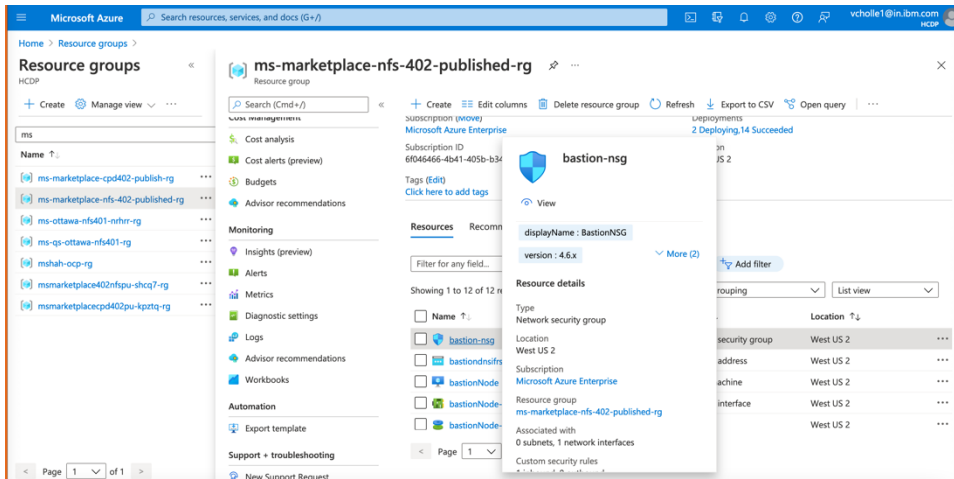

2. The security group window displays the ingress rules. Select the **Inbound** tab, and choose **Edit** to bring up the rule editor, choose **Add Rule**, and fill in the rule details. Add the network CIDR for the group of IP addresses that you want to permit SSH access to the boot nodes. To allow any IP address, use * in the **"Source port ranges"** field.
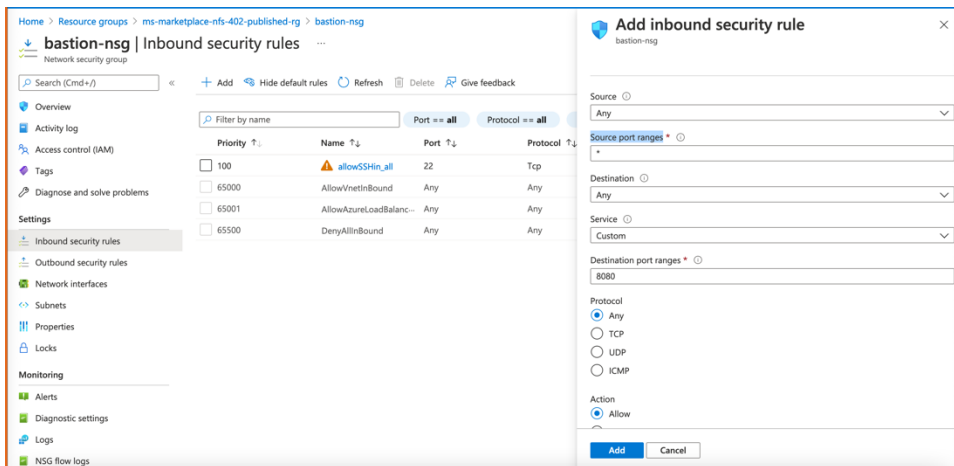
3. In the rule editor window, click on **Add**.

Scaling up your cluster by adding compute nodes

Login to your OpenShift Console, navigate Compute Machine Sets, each machine set can be scaled up.
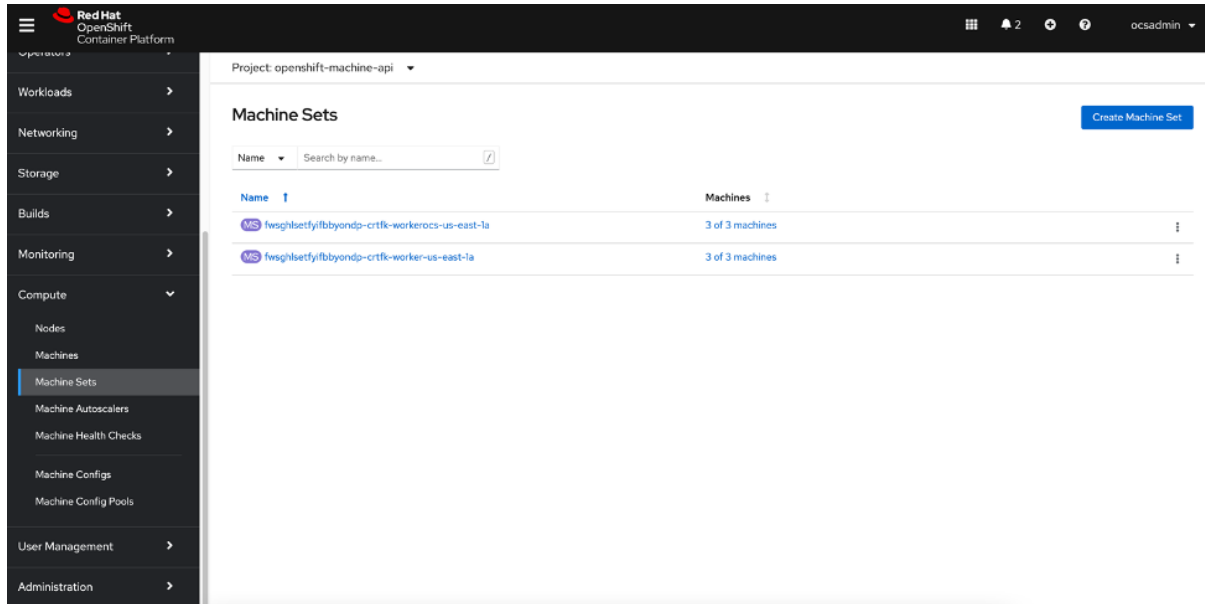
- An Azure instance will be created, and Desired count and current count will get updated to replica value.
- After few mins once the node joins the cluster ready and available count will be updated to replica value

Note
1. If you choose to scale down your cluster or reduce the number of compute nodes, there is a risk of the cluster becoming unstable because pods will need to be rescheduled. Scaling down the worker nodes is not a recommended option.
2. Cluster auto scaler can overrule the scaling activity to maintain the required threshold.

Cloud Pak for Data Express services

You can browse the various services that are available for use by navigating to the Cloud Pak for Data Service Catalog in Cloud Pak for Data.
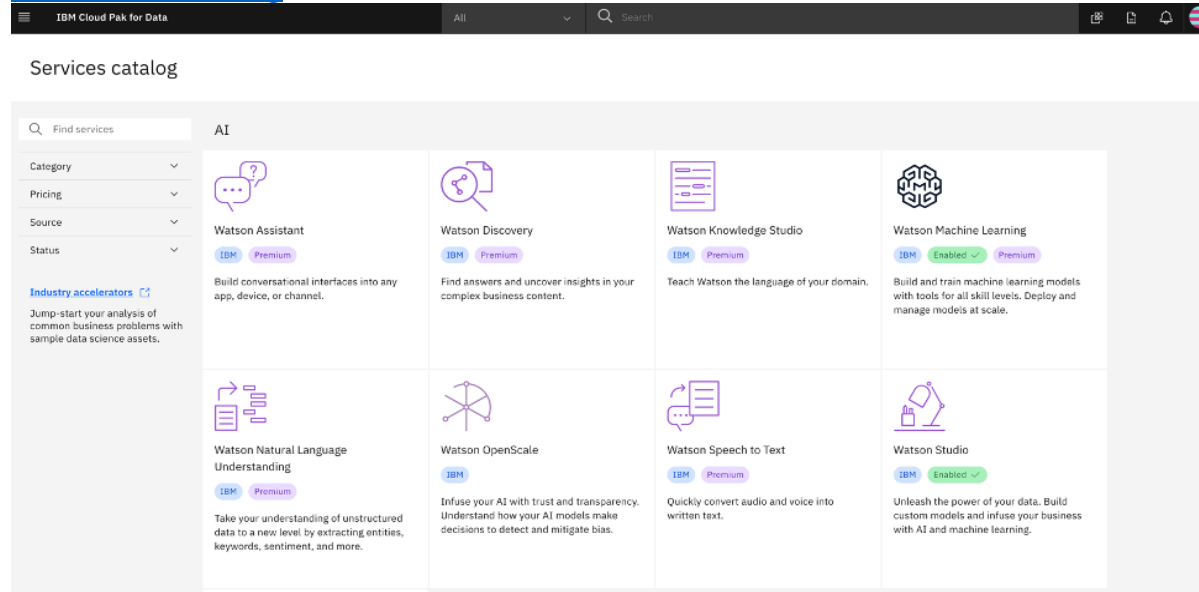

Figure: Services catalog page in Cloud Pak for Data

As part of this installation, the control plane is installed by default, and the following services are enabled by default depending on the Cloud Pak for Data Express solution configured:

IBM Data Governance Express
- Watson Knowledge Catalog
- Analytics Engine

IBM DataScience and MLOps Express
- Watson Studio
- Watson Machine Learning
- Watson Open Scale

IBM ELT Pushdown Express
- DataStage

System requirements for the Cloud Pak for Data Express

Cloud Pak for Data Express	CPU Cores (vCPUs)	Memory
IBM Data Governance Express	64	256 GB
IBM Data Science and MLOps Express	80	320 GB
IBM ELT Pushdown Express	48	192 GB

Note: The resources listed are only for the services and does not include the resources for Cloud Pak for Data control plane and foundational services.

Install a Cloud Pak for Data Service
- Login to your bootnode server.

- Follow the Operator Subscriptions install instructions for the services you are interested.
- Create the custom resources for the service that you are interested in.

To get information on various other services that are available, see the Cloud Pak for Data Service Catalog.

Upgrade Cloud Pak for Data Services

See what new features and improvements are available in the latest release of IBM® Cloud Pak for Data.

· Login to your bootnode server.
  - Follow the Operator Subscriptions upgrade instructions for the services you are interested to upgrade.
  - Upgrade the Cloud Pak for Data control plane
  - Apply upgrades to the custom resources for the service that you are interested in

Limitations & Known Issues

- Review the known issues and limitations for Cloud Pak for Data.

Additional resources

Azure resources
· Getting Started Resource Center
· Azure General Reference

Azure services
- **Azure VM**
- **Azure DNS**
- **Azure Resource groups**

IBM Cloud Pak for Data documentation
- IBM Documentation
- Red Hat OpenShift Container Platform

Document revisions

Date	Change
March 2024	Updated the Architecture diagram and updated the doc to IBM CPD v4.8.x
December 2023	Updated IBM Cloud Pak for Data version to 4.8.x and updated Links, Updated the process of extracting username and password for CPD web console
September 2023	Updated the Portworx version and Bootnode VM type

July 2023	Updated IBM Cloud Pak for Data version 4.7.x and Document links, Portworx Support with CSI option
May 2023	Initial version for Cloud Pak for Data Express