

IBM Cloud Pak for Data on the AWS Marketplace

Deployment Guide

December 2022

Contents

Overview	2
Cloud Pak for Data on AWS	2
Cost and licenses	3
Architecture	3
Planning the deployment	5
Specialized knowledge	5
Technical requirements	6
Deployment options	7
Pre-requisites	7
Step 1. IBM Cloud Pak for Data Subscription	7
Step 2. Red Hat Subscription	7
Step 3. Sign in to your AWS account	8
Step 4. Launch the Deployment	9
Option 1: Parameters for deploying Cloud Pak for Data into a new VPC	9
Option 2: Parameters for deploying Cloud Pak for Data into an existing VPC	14
Step 5. (Optional) Edit the application security group	18
Step 6. Login to Cloud Pak for Data web client	20
Step 7. Manage your cluster using the OpenShift Console	22
Step 8. (Optional) Provide Boot Node SSH access	24
Accessing the control plane through the Boot Node	25
Scaling up your cluster by adding compute nodes	27
Cloud Pak for Data services	28

System requirements for each of the services	29
Install a Cloud Pak for Data Service	30
Upgrade Cloud Pak for Data Services.....	30
Limitations	30
Troubleshooting	30
Additional resources	31
Document revisions	31

Overview

This deployment guide provides step-by-step instructions for deploying IBM Cloud Pak for Data on a Red Hat OpenShift Container Platform cluster on the AWS Cloud. You can automatically deploy a multi-master, production instance of Cloud Pak for Data.

IBM Cloud Pak for Data enables enterprise users to connect, catalog, govern, transform, and analyze the data regardless of where the data is.

Cloud Pak for Data on AWS

Cloud Pak for Data is an analytics platform that helps prepare data for artificial intelligence (AI). It enables data engineers, data stewards, data scientists, and business analysts to collaborate using an integrated multiple-cloud platform.

Cloud Pak for Data uses IBM's deep analytics portfolio to help organizations meet data and analytics challenges. The required building blocks (collect, organize, analyze, infuse) for information architecture are available using Cloud Pak for Data on AWS.

Cloud Pak for Data uses AWS services and features, including virtual private clouds (VPCs), Availability Zones, security groups, Elastic Block Storage, and Elastic Load Balancing to build a reliable and scalable cloud platform.

This reference deployment provides AWS CloudFormation templates to deploy Cloud Pak for Data on a new OpenShift cluster. This cluster includes:

- A Red Hat OpenShift Container Platform cluster created in a new or existing VPC 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 Redhat OpenShift Container Storage.
- Scalable OpenShift compute nodes running Cloud Pak for Data services.

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

This product requires an internet connection to deploy properly. The following packages are downloaded on deployment:

git, wget, httpd-tools, python3, python3-pip, jq, unzip, podman, terraform, awscliv2, openshift-client-linux-4.10.37

Git repo: <https://github.com/IBM/cp4d-deployment/>

Cost and licenses

The Cloud Pak for Data environment is deployed by using AWS CloudFormation template. You are responsible for the cost of the AWS services used for the infrastructure.

The AWS CloudFormation template for this deployment includes configuration parameters that you can customize. You can use it to build a new VPC for your Cloud Pak for Data on AWS cluster or deploy on an existing AWS VPC. Some of these settings, such as instance type, will affect the cost of deployment. For cost estimates, see the pricing pages for each AWS service you will be using. Prices are subject to change.

For more information about licensing terms, see the Cloud Pak for Data [software license agreement](#).

Upgrading to the latest version of Cloud Pak for Data indicates your acceptance of any new terms that may be applicable for the new version. To determine if new terms apply and to review them, please visit <https://www-03.ibm.com/software/sla/sladb.nsf/search?OpenForm>, execute a search for LI documents, and locate the LI applicable for the version you wish to upgrade to. To locate the LI for the Cloud Pak for Data Standard edition, type, 'IBM Cloud Pak for Data Standard Edition' and for the Enterprise edition type, 'IBM Cloud Pak for Data Enterprise Edition.'

In the event there are no results for the version you are upgrading to, review the LI associated to the previous version which would apply in this scenario (for example, if you are upgrading to version 4.0.1 and there is no LI for this version, you will need to review the LI for version 4.0)

Architecture

Deploying the [AWS Marketplace](#) template for a new VPC with **default parameters** builds the following Cloud Pak for Data environment in the AWS Cloud.

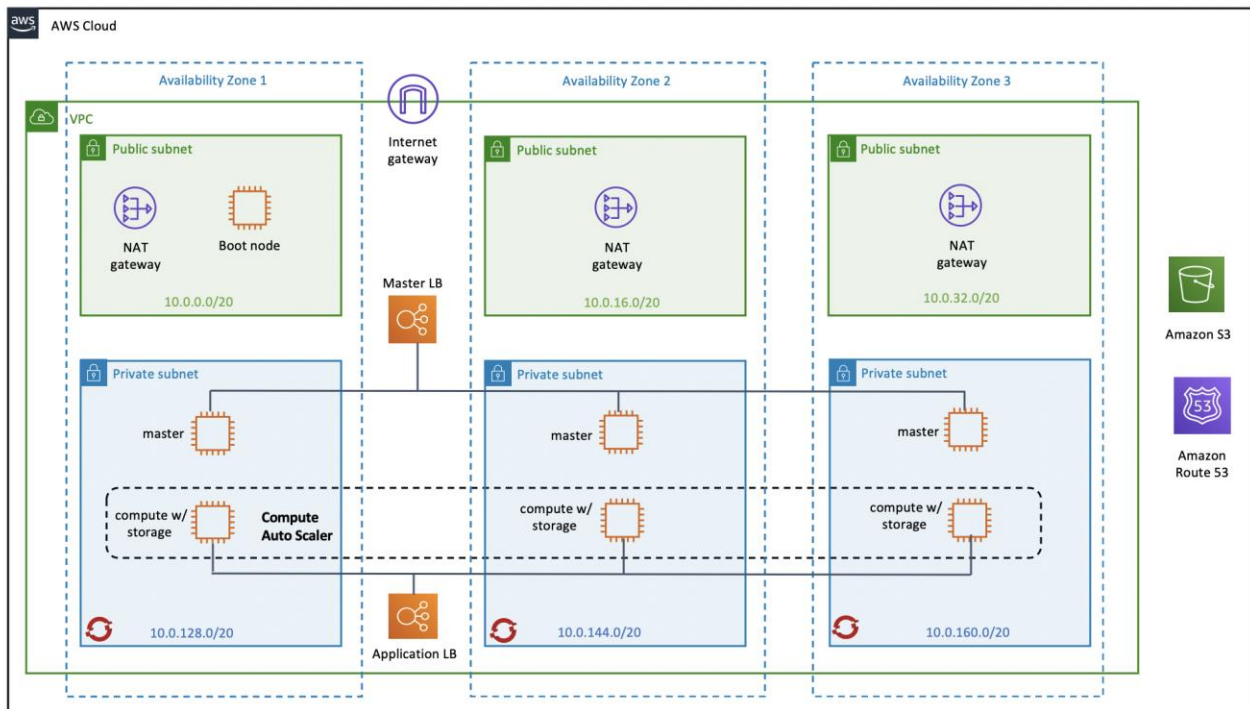


Figure 2: Deployment architecture for Cloud Pak for Data on AWS

The Cloud Formation Stack sets up the following:

- A highly available architecture that spans one or three Availability Zones.*
- A VPC configured with public and private subnets according to AWS best practices, to provide you with your own virtual network on AWS. *
- In the public subnets:
 - Managed network address translation (NAT) gateways to allow outbound internet access for resources in the private subnets.*
 - A boot node Amazon Elastic Compute Cloud (Amazon EC2) instance that also serves as a bastion host to allow inbound Secure Shell (SSH) access to EC2 instances in private subnets.
- In the private subnets:
 - OCP master instances up to three Availability Zones
 - OpenShift Container Platform (OCP) compute nodes that combined, contain Cloud Pak for Data Collect, Organize, and Analyze services.
 - Elastic Block Storage disks that are mounted on the compute nodes for container persistent data

- A Classic Load Balancer spanning the public subnets for accessing Cloud Pak for Data from a web browser.
- A Classic Load Balancer spanning the public subnets for accessing the OCP master instances.
- A Network Load Balancer spanning the private subnets for routing internal OpenShift application programming interface (API) traffic to the OCP master instances.
- Amazon Route 53 as your public Domain Name System (DNS) for resolving domain names of the IBM Cloud Pak for Data web console and applications deployed on the cluster.

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

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

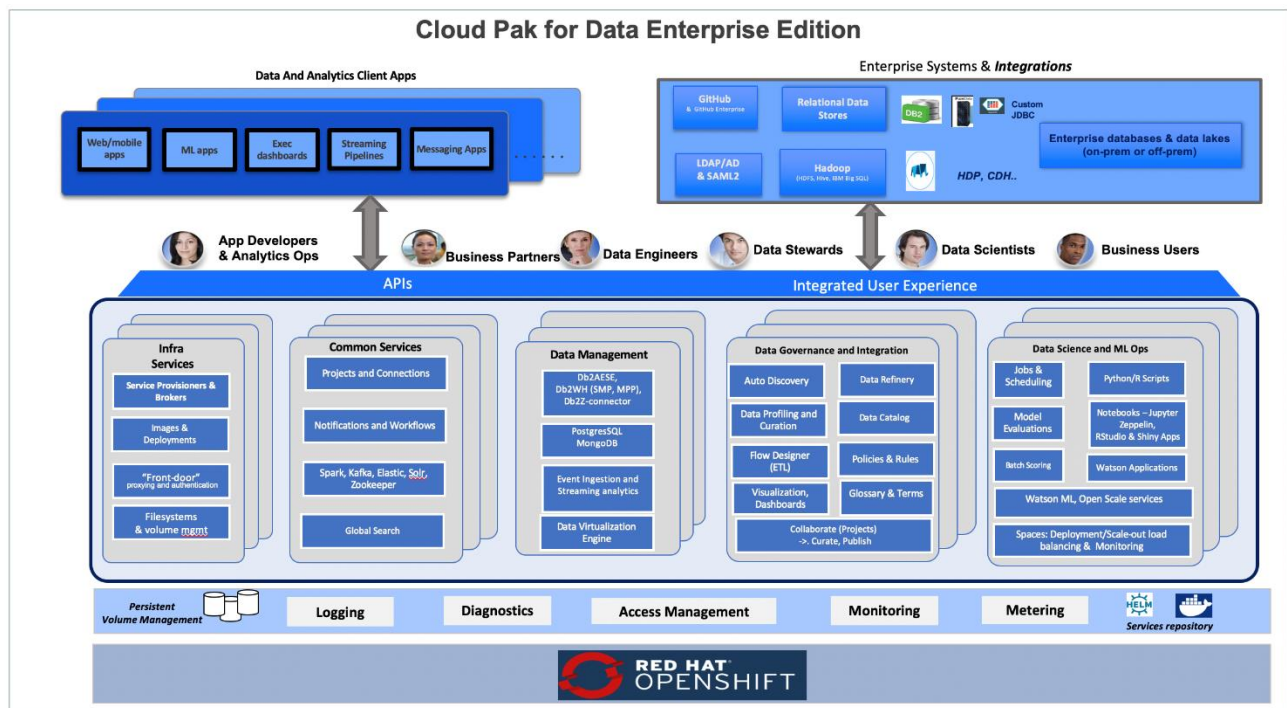


Figure 3: Cloud Pak for Data services

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 AWS 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 [AWS documentation](#).

Resource	This deployment uses
VPCs	1
Elastic IP addresses	3
Network Load Balancers	1
Classic Load Balancers	2
i3.large instances	1
m5.xlarge instances	3
m5.4xlarge instances	3
m4.4xlarge instances	3(OCS)

[Regions](#)

This deployment includes 3 Availability Zones, which isn't currently supported in all AWS Regions. For a current list of supported Regions, see [Service Endpoints and Quotas](#) in the AWS documentation.

[Key pair](#)

Make sure that at least one Amazon EC2 key pair exists in your AWS account in the Region where you are planning to deploy the template. Make note of the key pair name. You'll be prompted for this information during deployment. To create a key pair, follow the [instructions in the AWS documentation](#).

If you're deploying for testing or proof-of-concept purposes, we recommend that you create a new key pair instead of specifying a key pair that's already being used by a production instance.

[IAM permissions](#)

To deploy the template, you must log in to the AWS Management Console with AWS Identity and Access Management (IAM) permissions for the resources and actions the templates will deploy. The *AdministratorAccess* managed policy within IAM provides sufficient permissions.

The AdministratorAccess is required for the creation of openshift cluster. The info is present under the section creating an IAM user in the below link, Please review:

https://docs.openshift.com/container-platform/4.10/installing/installing_aws/installing-aws-account.html

Deployment options

This template provides the following deployment options:

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

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

Pre-requisites

Ensure the following subscriptions are in place with your existing Cloud Pak for Data entitlements or obtained from [AWS Marketplace](#)

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 Cloud Formation Template

Step 2. Red Hat Subscription

Ensure that you have the Red Hat [OpenShift pull secret](#) with your purchase. It should be uploaded to an S3 bucket. The Cloud Formation Template will pull this secret file from the specified S3 bucket location to provision the cluster.

Step 3. Sign in to your AWS account

Notes:

The customers should consider deploying their product to new AWS accounts because the permissions allow your application to create Administrative roles as well as it utilizes a long-term access key and secret key.

1. Sign in to your AWS account at <https://aws.amazon.com> with an AWS 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 AWS 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 AWS Region where you want to deploy Cloud Pak for Data on AWS. An IBM Pak for Data high availability deployment is restricted to AWS Regions with at least three Availability Zones.
4. The following resources should be made available for Cloud Pak for Data deployment
 - A key pair to SSH to the cluster nodes. In the navigation pane of the [Amazon EC2 console](#), choose **Key Pairs**, and then choose your key pair from the list.
 - A new or an existing S3 Bucket location with Red Hat pull secret
 - A new or an existing [Route 53](#) domain and a Hosted Zone Id which will be used for adding DNS records for Cloud Pak for Data applications.
 - An existing S3 Bucket to store deployment log files.

Step 4. Launch the Deployment

Notes The instructions in this section reflect the current version of the AWS CloudFormation console. If you're using the redesigned console, some of the user interface elements might be different.

You are responsible for the cost of the AWS services used while running this deployment. For full details, see the pricing pages for each AWS service you will be using for this deployment. Prices are subject to change.

1. Launch the AWS CloudFormation template into your AWS account from the AWS Marketplace directly or you can download the template and launch it separately from your account.

A Cloud Pak for Data standard deployment takes about 3 hours, and a high availability (HA) deployment takes about 4 hours to complete.

2. Check the Region that's displayed in the upper-right corner of the navigation bar 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 US East (N. Virginia) Region by default.
3. On the **Create stack** page, keep the default setting for the template URL, and then choose **Next**.
4. On the **Specify stack details** page, change the stack name if needed. 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 separately for these deployment options:

- [Parameters for deploying Cloud Pak for Data into a new VPC](#)
 - [Cloud Pak for Data Base Services](#)
 - [Cloud Pak for Data Cartridges](#)
- [Parameters for deploying Cloud Pak for Data into an existing VPC](#)
 - [Cloud Pak for Data Base Services](#)
 - [Cloud Pak for Data Cartridges](#)

When you finish reviewing and customizing the parameters, choose **Next**.

OPTION 1: PARAMETERS FOR DEPLOYING CLOUD PAK FOR DATA INTO A NEW VPC

VPC network configuration:

Parameter label (name)	Default	Description
Number of Availability Zones (NumberOfAZs)	3	The number of Availability Zones to be used for the deployment. Keep in mind that some Regions may be limited to two Availability Zones. For a single IBM Cloud Pak for Data cluster to be highly available, three Availability Zones are needed to avoid a single point of failure when using three, five, or seven master nodes or etcd nodes. With fewer than three Availability Zones, one of the Availability Zones will have more master nodes (or etcd nodes) than can be lost without losing a majority of the etcd instances. Allowed values: 1, 3
Availability Zones (AvailabilityZones)	<i>Requires input</i>	The list of Availability Zones to use for the subnets in the VPC. The template uses one or three Availability Zones and preserves the logical order you specify.
VPC CIDR (VPCCIDR)	10.0.0.0/16	The CIDR block for the VPC to be created.
Private subnet 1 CIDR (PrivateSubnet1CIDR)	10.0.0.0/19	The CIDR block for the private subnet located in Availability Zone 1.
Private subnet 2 CIDR (PrivateSubnet2CIDR)	10.0.32.0/19	The CIDR block for the private subnet located in Availability Zone 2.
Private subnet 3 CIDR (PrivateSubnet3CIDR)	10.0.64.0/19	The CIDR block for the private subnet located in Availability Zone 3.
Public subnet 1 CIDR (PublicSubnet1CIDR)	10.0.128.0/20	The CIDR block for the public subnet located in Availability Zone 1.
Public subnet 2 CIDR (PublicSubnet2CIDR)	10.0.144.0/20	The CIDR block for the public subnet located in Availability Zone 2.
Public subnet 3 CIDR (PublicSubnet3CIDR)	10.0.160.0/20	The CIDR block for the public subnet located in Availability Zone 3.
BootNodeAccessCIDR (BootNodeAccessCIDR)	<i>Requires input</i>	The CIDR IP range that is permitted to access the Boot Node instance. We recommend that you set this value to a trusted IP range. The value `0.0.0.0/0` permits all IP addresses to access. Additional values can be added post-deployment from the Amazon EC2 console.
Cluster Network CIDR (ClusterNetworkCIDR)	<i>Requires input</i>	The Cluster Network CIDR IP range that is used as IP address pools for pods.

DNS configuration:

Parameter label (name)	Default	Description
Domain name (DomainName)	<i>Requires input</i>	The domain name configured for the cluster.

Amazon EC2 configuration:

Parameter label (name)	Default	Description
Key pair name (KeyPairName)	<i>Requires input</i>	The name of an existing public/private key pair, which allows you to securely connect to your instance after it launches.

OpenShift hosts configuration:

Parameter label (name)	Default	Description
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.
Number of compute nodes (NumberOfNodes)	3	The desired capacity for the OpenShift node instances. Minimum of 3 nodes required. Warning If the number of node instances exceeds your Red Hat entitlement limits or AWS instance quotas, the stack will fail. Choose a number that is within your limits.
Master instance type (MasterInstanceType)	m5.xlarge	The EC2 instance type for the OpenShift master instances.
Compute instance type (NodesInstanceType)	m5.4xlarge	The EC2 instance type for the OpenShift node instances.
Cluster name (ClusterName)	<i>Requires input</i>	Custom cluster name for kubernetes.io/cluster/tags. If left blank, will use the stack name suffixed with the AWS Region.
Enable Fips (EnableFips)	false	Enable Fips for OpenShift
Publish (PrivateCluster)	false	To Deploy a Private cluster select true and false for Public cluster
Choose OpenShift Version (OpenShiftVersion)	4.10.37	Choose a supported OpenShift Version .
Choose username to login to OpenShift Console (OpenShiftUser)	ocsadmin	Username to be used to login to OpenShift Console.
Choose password to login to OpenShift and Cloud Pak for Data consoles (OpenShiftUser)	<i>Requires input</i>	Password to be used to login to OpenShift Console.

Storage Configuration:

Parameter label (name)	Default	Description
Storage type for Cluster (StorageType)	<i>efs-efs</i>	efs, efs-efs, ocs storage are available.
OCS instance type (OCSInstanceType)	m4.4xlarge	Update this value if Storage type selected is OCS. The EC2 instance type for the OpenShift Container Storage instances.
Number of OCS nodes (NumberOfOCS)	3	Update this value if Storage type selected is OCS. The desired capacity for the OpenShift container storage instances. Minimum of 3 is required.

Red Hat subscription information:

Parameter label (name)	Default	Description
RedHat subscription pull secret. (RedhatPullSecret)	<i>Requires input</i>	S3 path of OpenShift Installer Provisioned Infrastructure pull secret(e.g., s3://my-bucket/path/to/pull-secret).

IBM Cloud Pak for Data configuration:

Parameter label (name)	Default	Description
IBM Cloud Pak for Data version (ICPDVersion)	4.6.0	The 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 Entitled Registry API key Value (APIKey)	—	Enter the IBM Cloud Pak for Data API key to access IBM Container Registry
License agreement (LicenseAgreement)	—	I have read and agreed to the license terms for IBM Cloud Pak for Data that were provided to me at time of purchase.
OpenShift project (NameSpace)	zen	The OpenShift project that will be created for deploying Cloud Pak for Data. It can be any lowercase string.
Output S3 bucket name	<i>Requires Input</i>	The Existing S3 bucket name where the zip file output should be placed.

Parameter label (name)	Default	Description
(ICPDDeploymentLogsBucketName)		

IBM Cloud Pak for Data Base Services:

Parameter label (name)	Default	Description
Watson Knowledge Catalog service (WKC)	False	Choose True to install the Watson Knowledge Catalog service.
Watson Machine Learning service (WML)	False	Choose True to install the Watson Machine Learning service.
Data Virtualization service (DV)	False	Choose True to install the Data Virtualization service.
Watson Studio service (WSL)	False	Choose True to install the Watson Studio service.
Watson OpenScale and Watson Machine Learning services (OSWML)	False	Choose True to install the Watson OpenScale and Watson Machine Learning services.
Spark Service (SPARK)	False	Choose True to install the Apache Spark Engine
Cognos Dashboard (CDE)	False	Choose True to install the Cognos Dashboard Engine service

IBM Cloud Pak for Data Cartridge (This option is enabled only for Cartridge subscriptions)

Parameter label (name)	Default	Description
Cloud Pak for Data Premium Service (PremiumService)	Blank	Choose the premium service that you want to install. You can select from Db2, Db2 with Management Console, DataStage, Cognos Analytics, Planning Analytics, Watson Discovery, OpenPages, or Open Data for Industries.

OPTION 2: PARAMETERS FOR DEPLOYING CLOUD PAK FOR DATA INTO AN EXISTING VPC

Network configuration:

Parameter label (name)	Default	Description
Number of Availability Zones (NumberOfAZs)	3	The number of Availability Zones to be used for the deployment. Keep in mind that some Regions may be limited to two Availability Zones. For a single IBM Cloud Pak for Data cluster to be highly available, three Availability Zones are needed to avoid a single point of failure when using three, five, or seven master nodes or etcd nodes. With fewer than three Availability Zones, one of the Availability Zones will have more master nodes (or etcd nodes) than can be lost without losing a majority of the etcd instances. Allowed values: 1, 3
VPC ID (VPCID)	<i>Requires input</i>	The ID of your existing VPC for deployment.
VPC CIDR (VPCCIDR)	<i>Requires input</i>	The CIDR block for the VPC to be created.
Private subnet 1 ID (PrivateSubnet1ID)	<i>Requires input</i>	The ID of the private subnet in Availability Zone 1 for the workload (e.g., subnet-a0246dcd).
Private subnet 2 ID (PrivateSubnet2ID)	<i>Requires input</i>	The ID of the private subnet in Availability Zone 2 for the workload (e.g., subnet-b1f432cd).
Private subnet 3 ID (PrivateSubnet3ID)	<i>Requires input</i>	The ID of the private subnet in Availability Zone 3 for the workload (e.g., subnet-b1f4a2cd).
Public subnet 1 ID (PublicSubnet1ID)	<i>Requires input</i>	The ID of the public subnet in Availability Zone 1 for the ELB load balancer (e.g., subnet-9bc642ac).
Public subnet 2 ID (PublicSubnet2ID)	<i>Requires input</i>	The ID of the public subnet in Availability Zone 2 for the ELB load balancer (e.g., subnet-e3246d8e).
Public subnet 3 ID (PublicSubnet3ID)	<i>Requires input</i>	The ID of the public subnet in Availability Zone 3 for the ELB load balancer (e.g., subnet-e324ad8e).
BootNodeAccessCIDR (BootNodeAccessCIDR)	<i>Requires input</i>	The CIDR IP range that is permitted to access the Boot Node instance. We recommend that you set this value to a trusted IP range. The value `0.0.0.0/0` permits all IP addresses to access. Additional values can be added post-deployment from the Amazon EC2 console.
Cluster Network CIDR (ClusterNetworkCIDR)	<i>Requires input</i>	The Cluster Network CIDR IP range that is used as IP address pools for pods.

DNS configuration:

Parameter label (name)	Default	Description
Domain name (DomainName)	<i>Requires input</i>	The domain name configured for the cluster.

Amazon EC2 configuration:

Parameter label (name)	Default	Description
Key pair name (KeyPairName)	<i>Requires input</i>	The name of an existing public/private key pair, which allows you to securely connect to your instance after it launches.

OpenShift hosts configuration:

Parameter label (name)	Default	Description
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.
Number of compute nodes (NumberOfNodes)	3	The desired capacity for the OpenShift node instances. Minimum of 3 nodes required. Warning If the number of node instances exceeds your Red Hat entitlement limits or AWS instance quotas, the stack will fail. Choose a number that is within your limits.
Master instance type (MasterInstanceType)	m5.xlarge	The EC2 instance type for the OpenShift master instances.
Compute instance type (NodesInstanceType)	m5.4xlarge	The EC2 instance type for the OpenShift node instances.
Cluster name (ClusterName)	<i>Requires input</i>	Custom cluster name for kubernetes.io/cluster/tags. If left blank, will use the stack name suffixed with the AWS Region.
Enable Fips (EnableFips)	false	Enable Fips for OpenShift
Publish (PrivateCluster)	false	To Deploy a Private cluster select true and false for Public cluster
Choose OpenShift Version (OpenShiftVersion)	4.10.34	Choose OpenShift Version.
Choose username to login to OpenShift Console (OpenShiftUser)	<i>Requires input</i>	Username to be used to login to OpenShift Console.
Choose password to login to OpenShift and Cloud Pak for Data consoles (OpenShiftUser)	<i>Requires input</i>	Password to be used to login to OpenShift Console.

Storage Configuration:

Parameter label (name)	Default	Description
OCS instance type (OCSInstanceType)	m4.4xlarge	Update this value if Storage type selected is OCS. The EC2 instance type for the OpenShift Container Storage instances.
Number of OCS nodes (NumberOfOCS)	3	Update this value if Storage type selected is OCS. The desired capacity for the OpenShift container storage instances. Minimum of 3 is required.

Red Hat subscription information:

Parameter label (name)	Default	Description
RedHat subscription pull secret (RedhatPullSecret)	<i>Requires input</i>	S3 path of OpenShift Installer Provisioned Infrastructure pull secret(e.g., s3://my-bucket/path/to/pull-secret).

IBM Cloud Pak for Data configuration:

Parameter label (name)	Default	Description
IBM Cloud Pak for Data version (ICPDVersion)	4.6.0	The 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 Entitled Registry API key Value (APIKey)	—	Enter the IBM Cloud Pak for Data API key to access IBM Container Registry
License agreement (LicenseAgreement)	—	I have read and agreed to the license terms for IBM Cloud Pak for Data that were provided to me at time of purchase.
OpenShift project (NameSpace)	zen	The OpenShift project that will be created for deploying Cloud Pak for Data. It can be any lowercase string.
Output S3 Bucket (ICPDDeploymentLogsBucketName)		The name of the S3 bucket where IBM Cloud Pak for Data deployment logs are to be exported. The deployment logs provide a record of the boot strap scripting actions and are useful for problem determination if the deployment fails in some way

IBM Cloud Pak for Data Base Services

Parameter label (name)	Default	Description
Watson Knowledge Catalog service (WKC)	False	Choose True to install the Watson Knowledge Catalog service.
Watson Machine Learning service (WML)	False	Choose True to install the Watson Machine Learning service.
Data Virtualization service (DV)	False	Choose True to install the Data Virtualization service.
Watson Studio service (WSL)	False	Choose True to install the Watson Studio service.
Watson OpenScale and Watson Machine Learning services (OSWML)	False	Choose True to install the Watson OpenScale and Watson Machine Learning services.
Spark Service (SPARK)	False	Choose True to install the Apache Spark Engine
Cognos Dashboard (CDE)	False	Choose True to install the Cognos Dashboard Engine service

IBM Cloud Pak for Data Cartridge (This option is enabled only for Cartridge subscriptions)

Parameter label (name)	Default	Description
Cloud Pak for Data Premium Service (PremiumService)	<i>Requires input</i>	Choose the premium service that you want to install. You can select from Db2, Db2 with Management Console, DataStage, Cognos Analytics, Planning Analytics, Watson Discovery, OpenPages

- On the options page, you can [specify tags](#) (key-value pairs) for resources in your stack and [set advanced options](#). When you're done, choose **Next**.
- On the **Review** page, review and confirm the template settings. Under **Capabilities**, select the check box to acknowledge that the template will create IAM resources.
- Choose **Create stack** to deploy the stack.
- Monitor the status of the stack. When the status is **CREATE_COMPLETE**, the Cloud Pak for Data cluster is ready.

- Use the URLs displayed in the **Outputs** tab for the stack to view the resources that were created. The URL for the **ICPDWebClientURL** output key will navigate to the console login page.

Step 5. (Optional) Edit the application security group

Optional: You might need to edit the AWS application group to add IP addresses that can access the Cloud Pak for Data web client.

Navigate to Load Balancers on your AWS console and filter on tags, for example *kubernetes.io/service-name: openshift-ingress/router-default*.

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

The screenshot shows the AWS Management Console interface for a Load Balancer. The left sidebar contains navigation options like Launch Templates, Spot Requests, Savings Plans, Reserved Instances, Dedicated Hosts, Scheduled Instances, Capacity Reservations, Images, AMIs, Elastic Block Store, Network & Security, Load Balancing, and Auto Scaling. The main content area shows the 'Create Load Balancer' page with a search filter 'kubernetes.io/service-name : openshift-ingress/router-default'. A table lists the load balancer details. Below the table, the 'Basic Configuration' tab is selected, showing fields for Name, DNS name, Type, Scheme, Availability Zones, Creation time, Hosted zone, Status, and VPC. The 'Port Configuration' section shows two ports: 80 (TCP) forwarding to 32019 (TCP) and 443 (TCP) forwarding to 32137 (TCP). The 'Security' section shows the 'Source Security Group' as 'sg-09ddcb0748bca3ba2', which is highlighted with a red box.

- Select Security Group and modify the Inbound rules

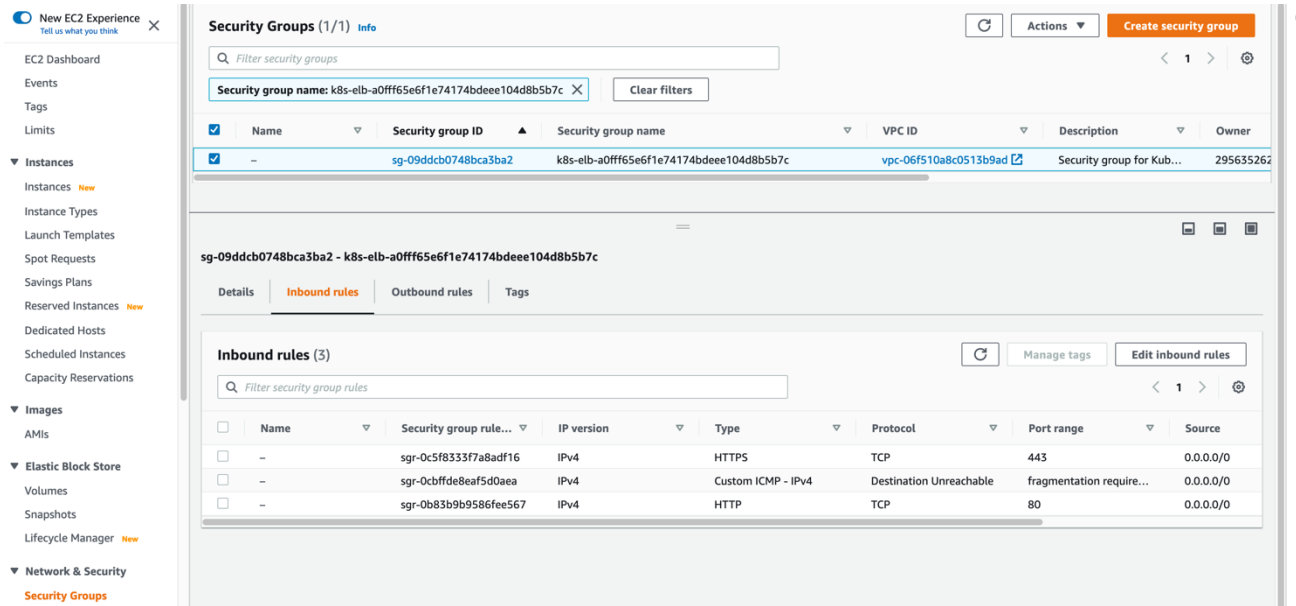


Figure 5: OpenShift node security group

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.

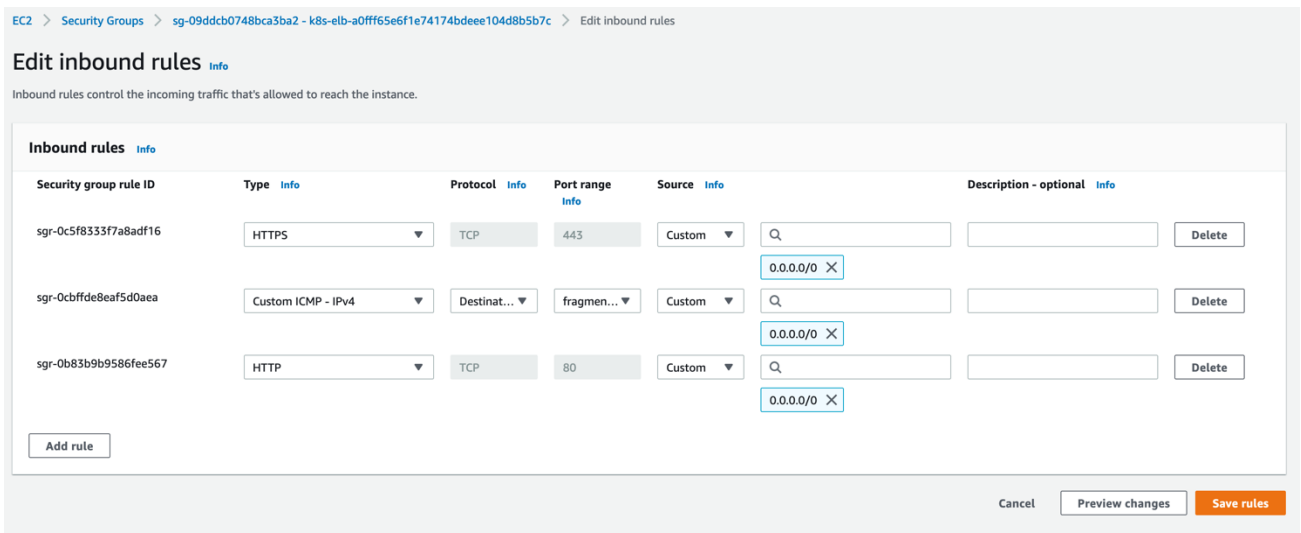


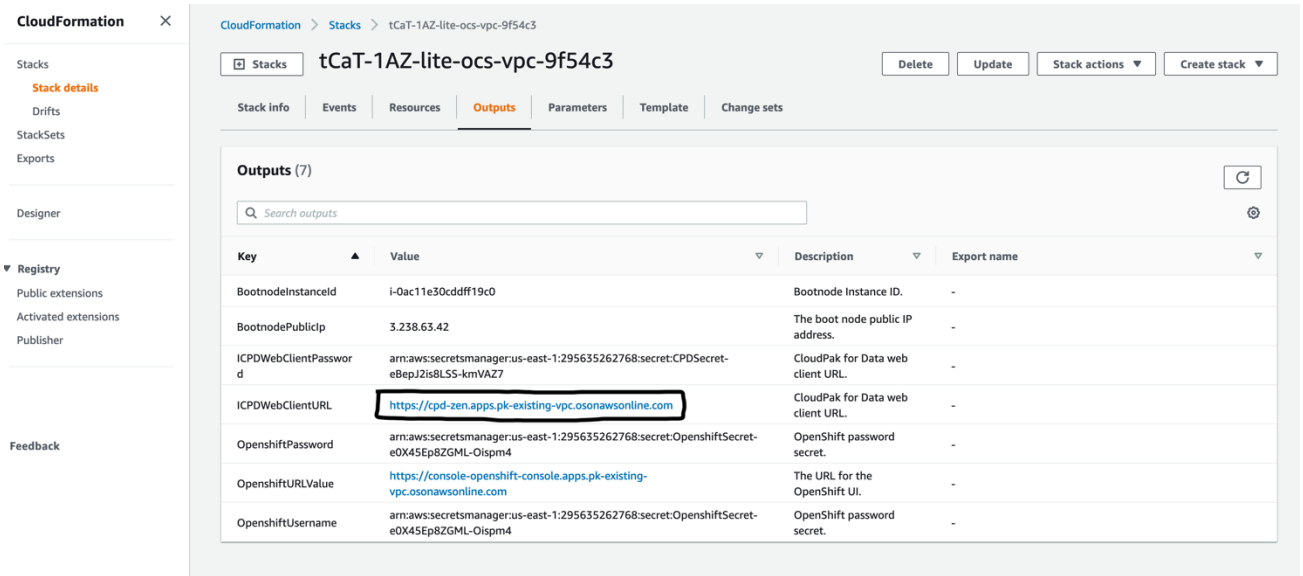
Figure 7: Supplying rule details

4. In the rule editor window, choose **Save**.

Step 6. Login to Cloud Pak for Data web client

When the AWS CloudFormation template has successfully created the stack, all server nodes will be running with the software installed in your AWS account. 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 and password from the Cloud Formation Stack output for key ICPDWebClientURL



2. from the Secret Manager

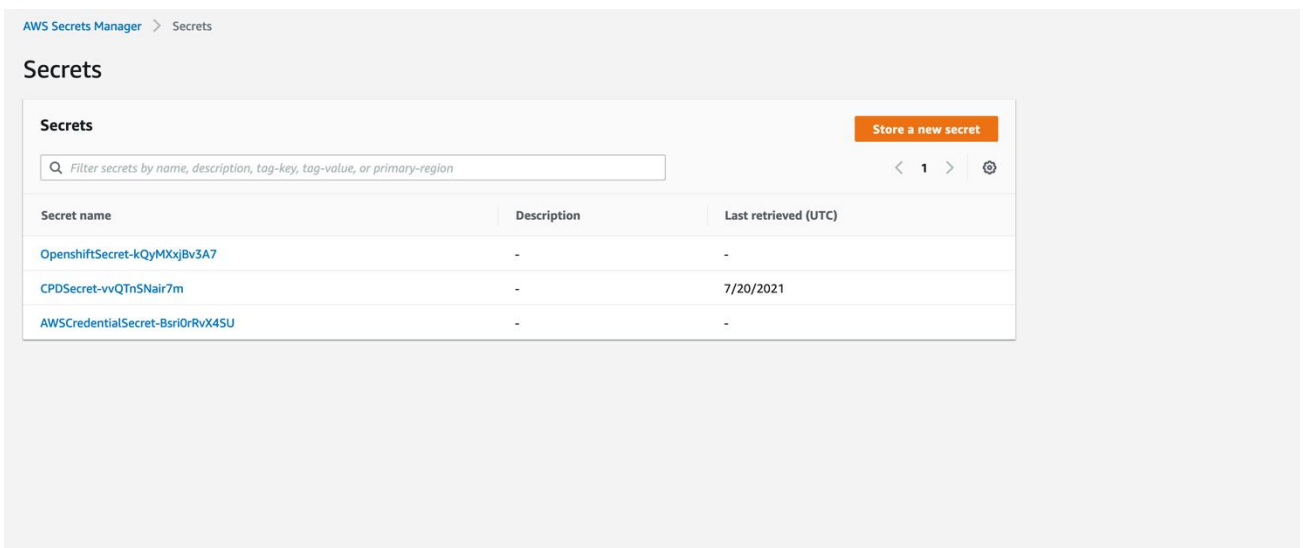
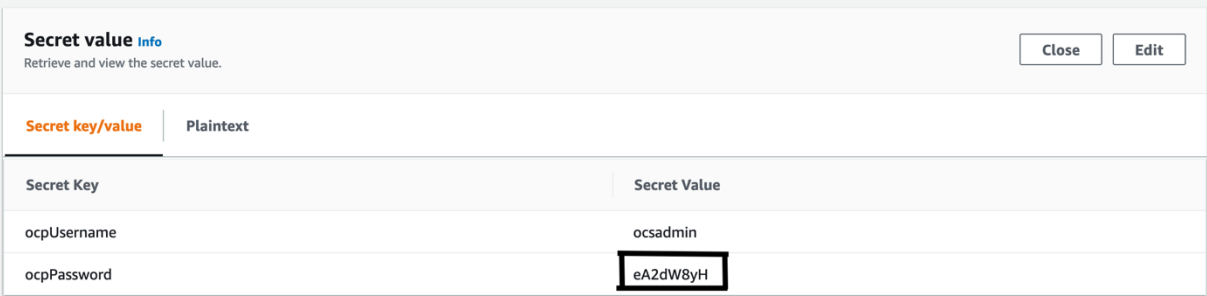


Figure 8: OpenShift Secret in Secret Manager

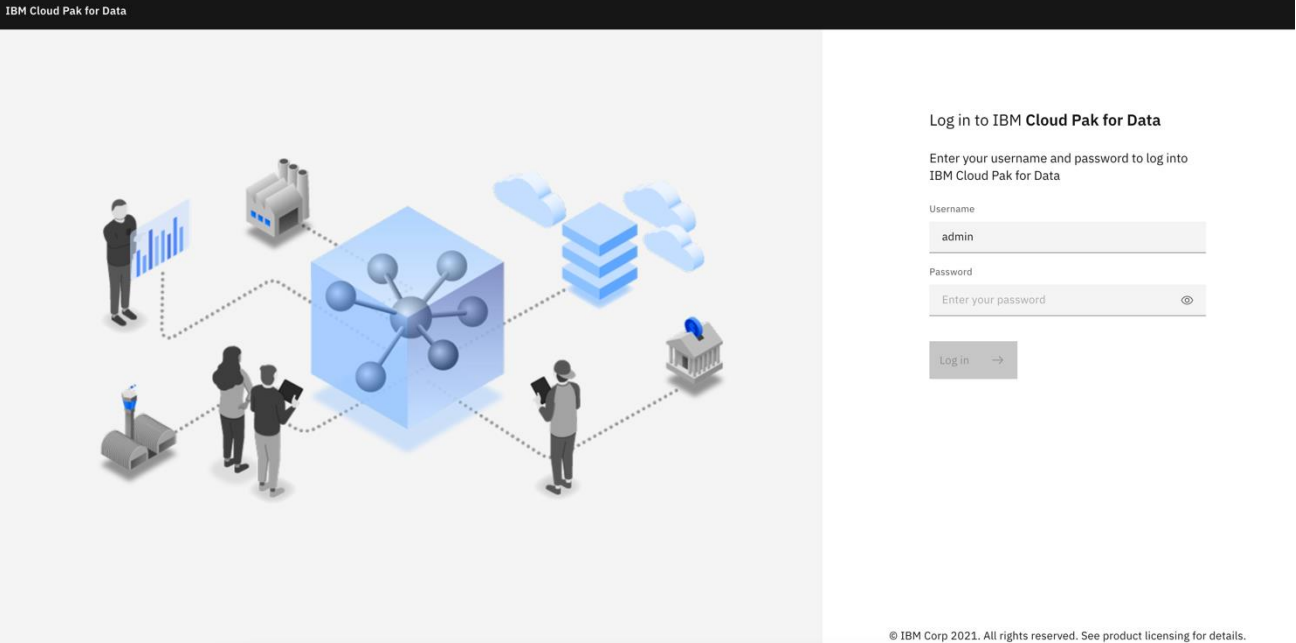
- Click the secret prefixed with “OpenShiftSecret” and retrieve the value of the secret. Take a note of the the value of the “ocpPassword”, this will be the Cloud Pak for Data password.



The screenshot shows the 'Secret value Info' page in OpenShift. It includes a 'Close' and 'Edit' button. Below is a table with two columns: 'Secret Key' and 'Secret Value'. The 'ocpPassword' key has the value 'eA2dW8yH' highlighted with a black box.

Secret Key	Secret Value
ocpUsername	ocsadmin
ocpPassword	eA2dW8yH

- Log in to the Cloud Pak for Data web client by using the user “admin” and the password from the previous step.



The screenshot shows the login page for IBM Cloud Pak for Data. The page title is 'Log in to IBM Cloud Pak for Data'. It prompts the user to 'Enter your username and password to log into IBM Cloud Pak for Data'. The 'Username' field contains 'admin'. The 'Password' field contains the password 'eA2dW8yH'. A 'Log in' button is visible. The page also features a large illustration of a central blue cube with nodes, surrounded by icons representing data, people, and infrastructure.

IBM Cloud Pak for Data

Log in to IBM Cloud Pak for Data

Enter your username and password to log into IBM Cloud Pak for Data

Username
admin

Password
Enter your password

Log in →

© IBM Corp 2021. All rights reserved. See product licensing for details.

- Once you log in, the welcome page is displayed.

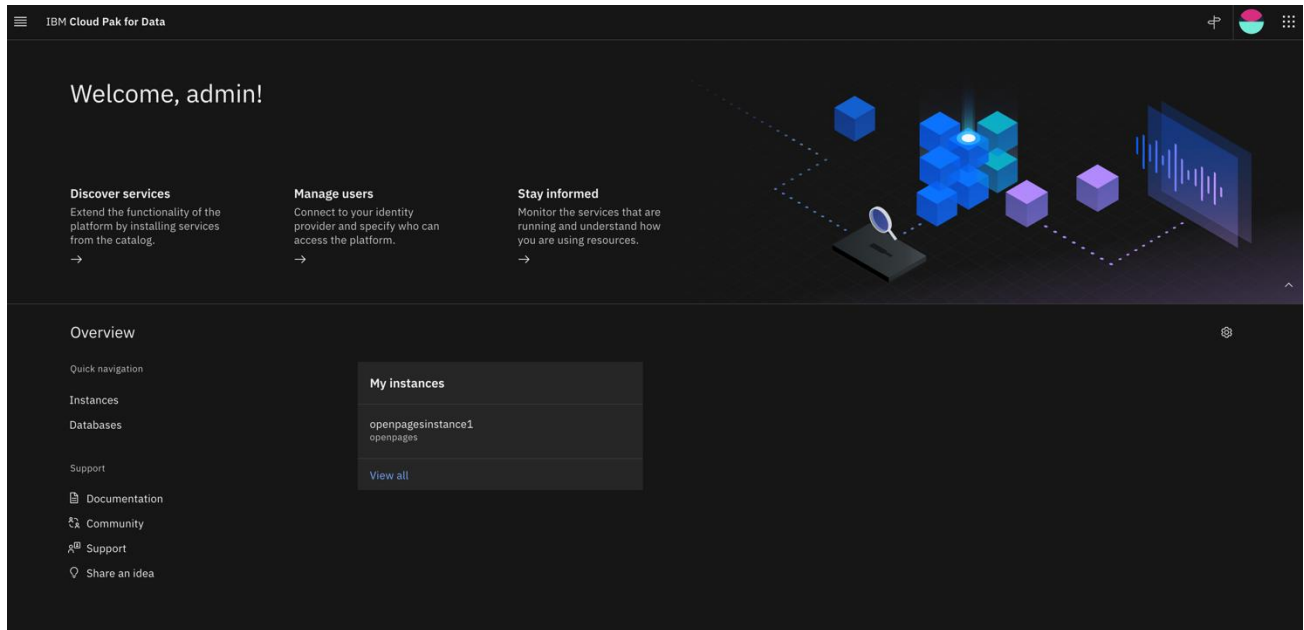
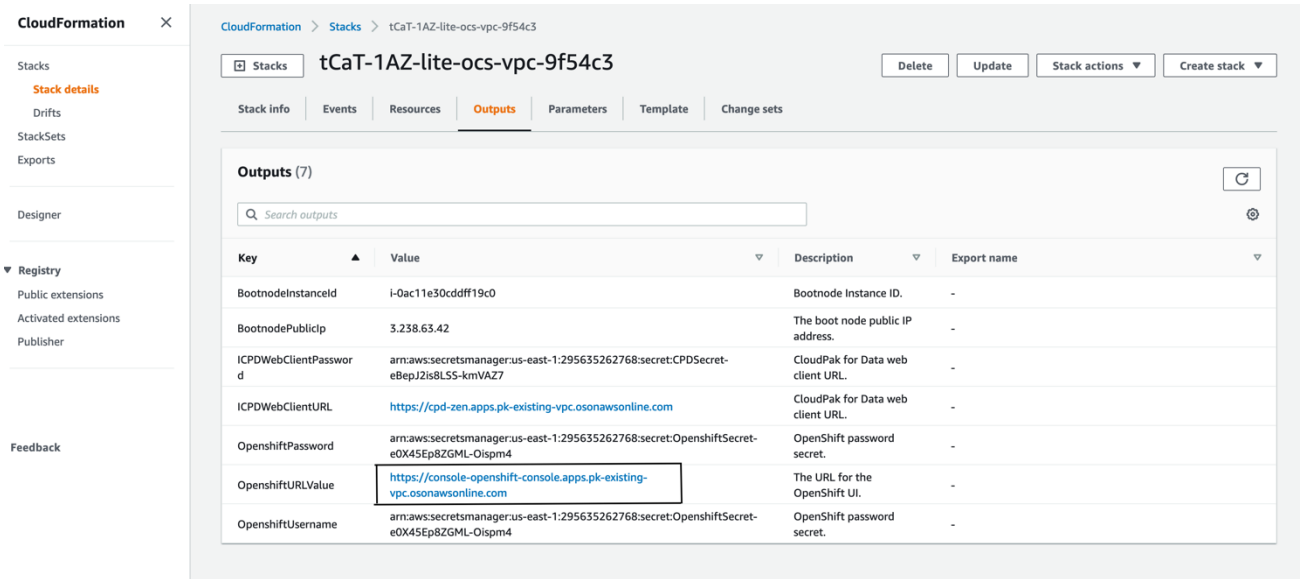


Figure 9: Welcome page for Cloud Pak for Data web client

See [resources](#) on platform features and capabilities. For a list of supported browsers, see [Supported browsers](#).

Step 7. Manage your cluster using the OpenShift Console

1. To access the Cloud Pak for Data web client, go to the **OpenShiftURLValue** output of the root stack



The default OpenShift administrative user is **ocsadmin**.

You can retrieve the secret value by clicking Retrieve secret value as shown, and use the value for the OpenShift console administrative password.

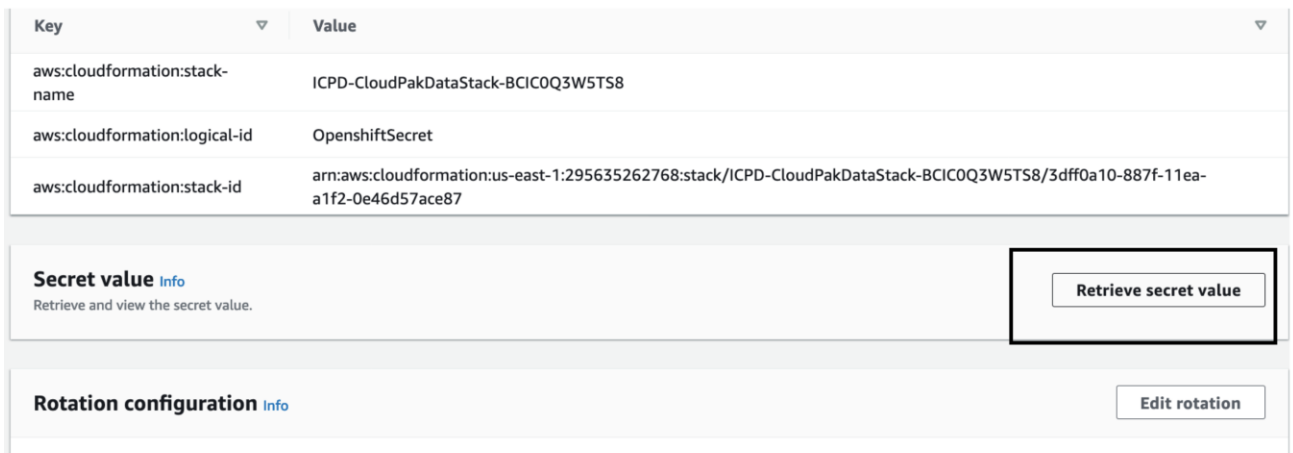
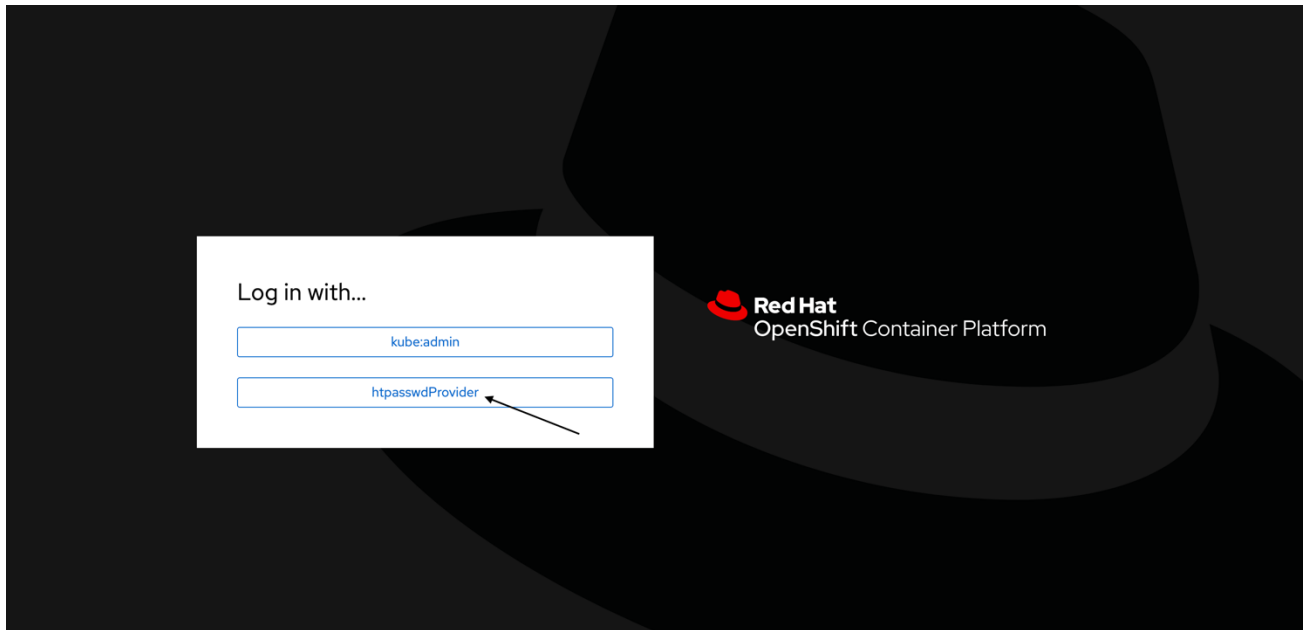


Figure 10.1: Retrieve secret value for console password

- Open the OpenShift Console URL in a browser, select htpasswdProvider at Log in and Login with the username and password from the previous step.



Step 8. (Optional) Provide Boot Node SSH access

The boot node EC2 instance 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 `BootNodeSecurityGroup` rule.

This section describes the steps to modify the `BootNodeSecurityGroup` inbound rules.

Note These steps assume access to the AWS CloudFormation console for the IBM Cloud Pak for Data deployment.

1. In the AWS Security Groups page, filter name containing `BootNodeSecurityGroup`

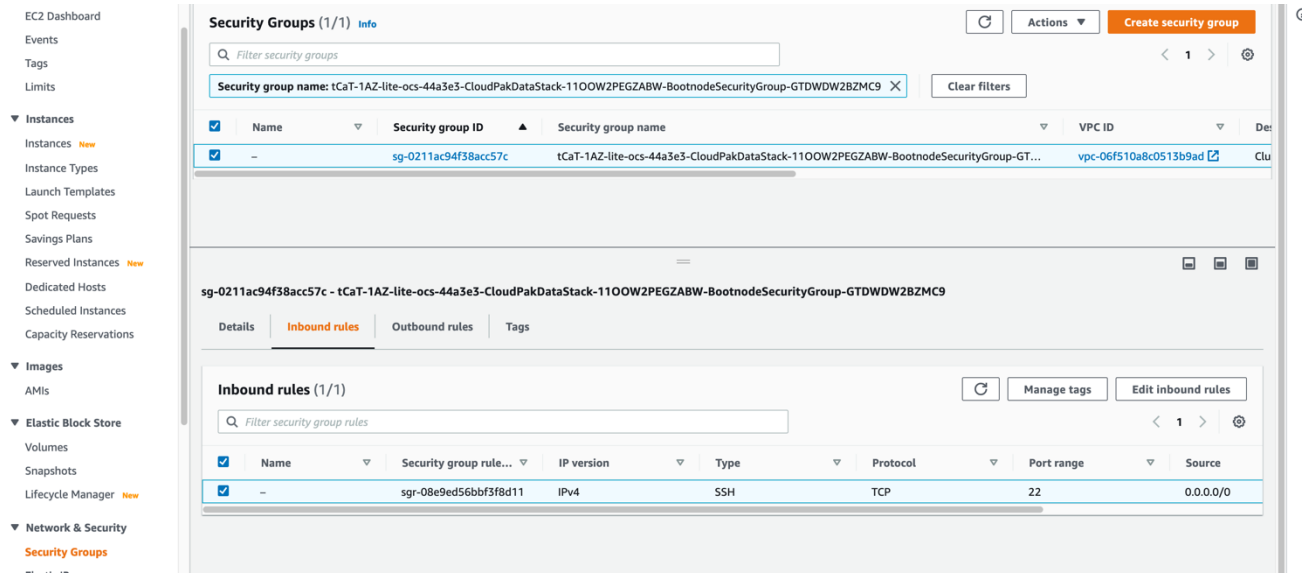


Figure 11: Security Groups

- 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 0.0.0.0/0.

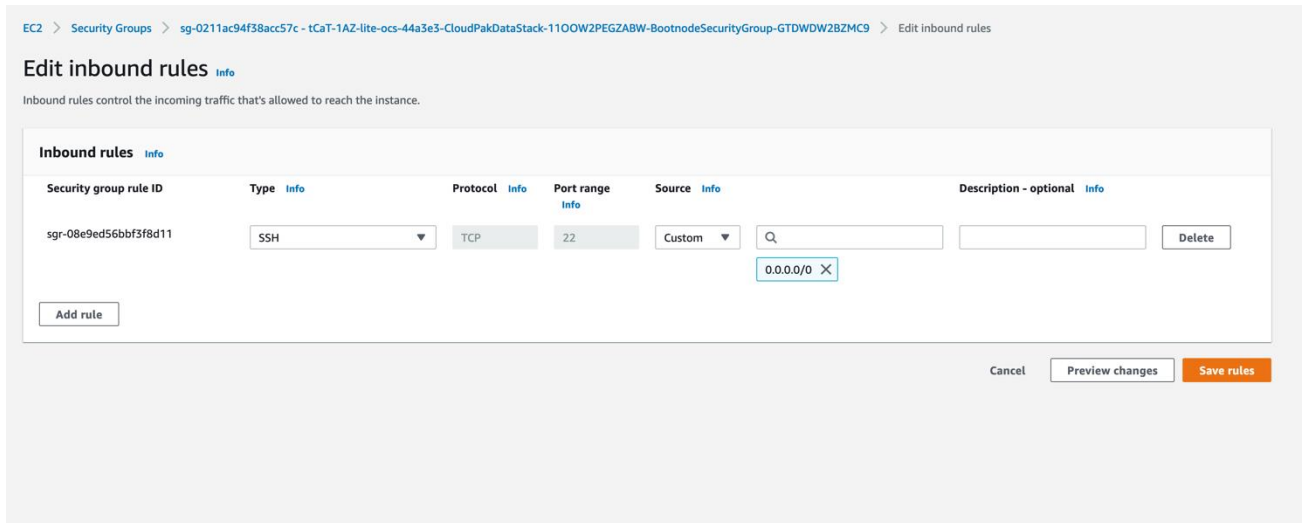


Figure 14: Supplying rule details

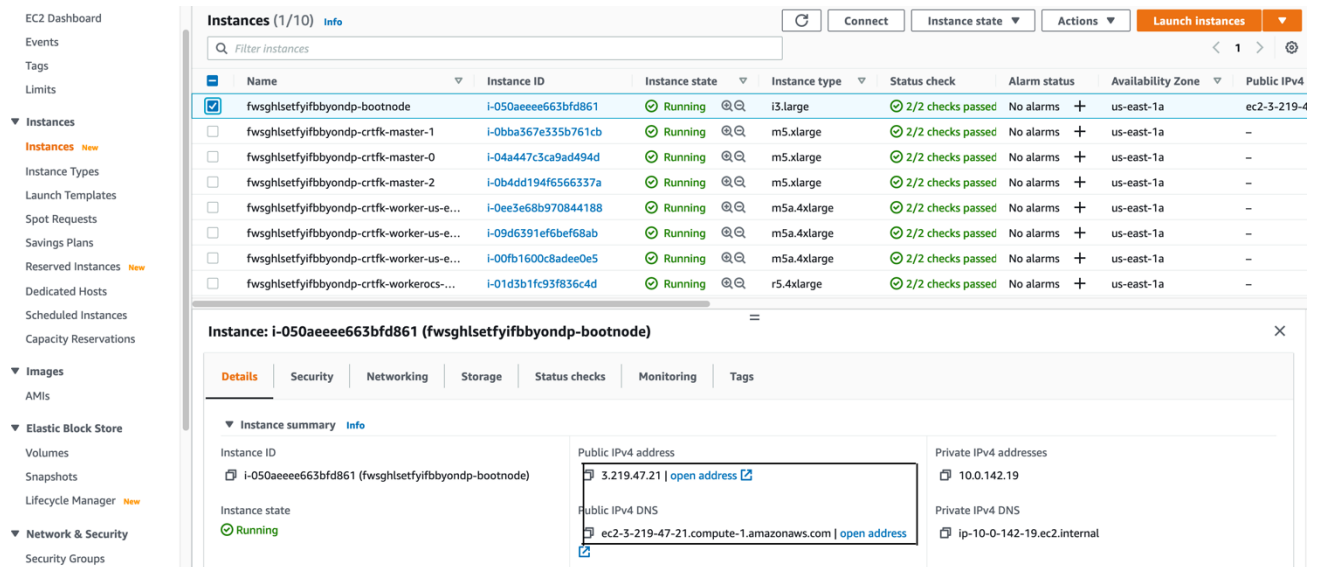
- In the rule editor window, choose **Save Rules**.

Accessing the control plane through the Boot Node

The recommended method of SSH access to the OpenShift cluster instances via the bastion host is by using SSH agent forwarding, as in the following Bash instructions:

- Run the command `ssh-add -K <your-key.pem>` to store the key in your keychain. On Linux, you might need to omit the `-K` flag.

2. Retrieve the host name of the Boot Node from the Amazon EC2 console.



The screenshot shows the Amazon EC2 console. On the left is a navigation menu with categories like Instances, Images, Elastic Block Store, and Network & Security. The main area displays a table of instances. The first instance, 'fwsghlsetfyfbyondp-bootnode' with ID 'i-050aeeee663bfd861', is selected. Below the table, the details for this instance are shown, including its state (Running) and network information. A red box highlights the Public IPv4 address '3.219.47.21' and the Public IPv4 DNS 'ec2-3-219-47-21.compute-1.amazonaws.com'.

Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IPv4
fwsghlsetfyfbyondp-bootnode	i-050aeeee663bfd861	Running	i3.large	2/2 checks passed	No alarms	us-east-1a	ec2-3-219-4
fwsghlsetfyfbyondp-crtfk-master-1	i-0bba367e335b761cb	Running	m5.xlarge	2/2 checks passed	No alarms	us-east-1a	-
fwsghlsetfyfbyondp-crtfk-master-0	i-04a447c3ca9ad494d	Running	m5.xlarge	2/2 checks passed	No alarms	us-east-1a	-
fwsghlsetfyfbyondp-crtfk-master-2	i-0b4dd194f6566337a	Running	m5.xlarge	2/2 checks passed	No alarms	us-east-1a	-
fwsghlsetfyfbyondp-crtfk-worker-us-e...	i-0ee3e68b970844188	Running	m5a.4xlarge	2/2 checks passed	No alarms	us-east-1a	-
fwsghlsetfyfbyondp-crtfk-worker-us-e...	i-09d6391ef6bef68ab	Running	m5a.4xlarge	2/2 checks passed	No alarms	us-east-1a	-
fwsghlsetfyfbyondp-crtfk-worker-us-e...	i-00fb1600c8adee0e5	Running	m5a.4xlarge	2/2 checks passed	No alarms	us-east-1a	-
fwsghlsetfyfbyondp-crtfk-workeroc...	i-01d3b1fc93f836c4d	Running	r5.4xlarge	2/2 checks passed	No alarms	us-east-1a	-

Instance: i-050aeeee663bfd861 (fwsghlsetfyfbyondp-bootnode)

Instance summary

Field	Value
Instance ID	i-050aeeee663bfd861 (fwsghlsetfyfbyondp-bootnode)
Instance state	Running
Public IPv4 address	3.219.47.21 open address
Public IPv4 DNS	ec2-3-219-47-21.compute-1.amazonaws.com open address
Private IPv4 addresses	10.0.142.19
Private IPv4 DNS	ip-10-0-142-19.ec2.internal

Figure 15: Host name of the Boot Node

3. To log in to the bastion host, run

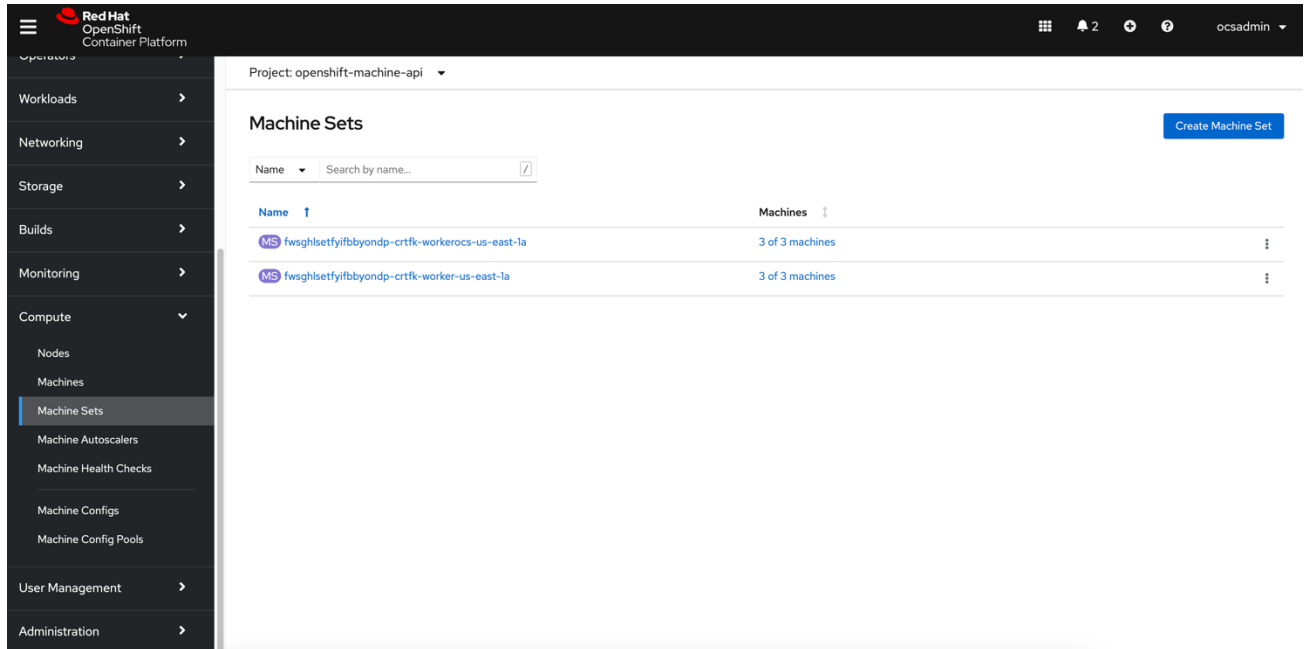
```
$ ssh -I <path-to-private-ssh-key> ec2-user@<bootnode-host-name>
```

4. Run oc login to authenticate with OpenShift and verify that services are in a running state:

```
$ oc get pods
```

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 AWS 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 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.

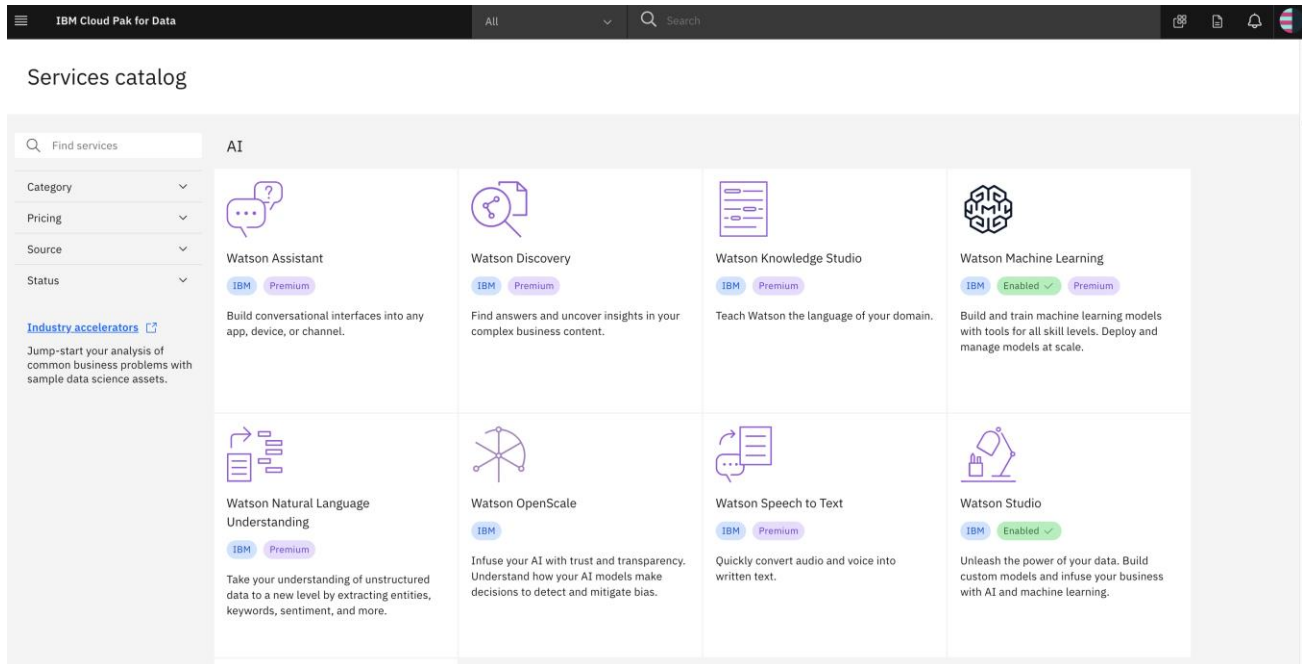


Figure 16: Services catalog page in Cloud Pak for Data

As part of this installation, the control plane is installed by default, and the following services can be enabled while creating the stack:

- Watson Studio
- Watson Knowledge Catalog
- Watson Machine Learning
- Data Virtualization
- Watson OpenScale
- Cognos Dashboards Embedded
- Analytics Engine for Apache Spark
- Db2 Cartridge
- DataStage Cartridge
- Cognos Analytics Cartridge

- Planning Analytics Cartridge
- Watson Discovery Cartridge
- OpenPages Cartridge
- Open Data for Industries Cartridge

System requirements for each of the services

Service Name	CPU cores(vCPUs)	Memory
Watson Studio	48	167GB
Watson Knowledge Catalog	64	180GB
Watson Machine Learning	48	176GB
Data Virtualization	64	198GB
Watson Openscale (including WML)	64	250GB
Spark Engine	7	170GB
Db2	16	198GB
DataStage	6	110GB
Cognos Analytics	27	61GB
Planning Analytics	64	256GB
Watson Discovery	15	508GB
OpenPages	7	250GB
Open Data for Industries	48	144GB

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.

Exception:

After deploying `Watson Discovery` on cluster, execute following command to set the memory limit of RabbitMQ pod (wd-rabbitmq-discovery-0) to 2GB:

```
oc patch wd wd --type=merge --patch '{"spec": {"rabbitmq": {"resource": {"limits": {"memory": "2Gi"}}}}'
```

After you execute the `oc patch` command, the RabbitMQ pod (`wd-rabbitmq-discovery-0`) will be restarted automatically. You can check the completion of the restarting of the pod with `"oc get po wd-rabbitmq-discovery-0"` where you can see its readiness new age.

Install a Cloud Pak for Data Service

- Login to your bootnode server as mentioned in [access section](#).
- Follow the [install](#) instructions
- 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 as mentioned in [access section](#).
- Follow [upgrade](#) instructions

Limitations

- Review the [known issues and limitations](#) for Cloud Pak for Data.

Troubleshooting

Q. I encountered a `CREATE_FAILED` error when I launched the CF Template

A. If AWS CloudFormation fails to create the stack, we recommend that you relaunch the template with **Rollback on failure** set to **No**. (This setting is under **Advanced** in the AWS CloudFormation console, **Options** page.) With this setting, the stack's state will be retained and the instance will be left running, so you can troubleshoot the issue. (Look at the log files in `%ProgramFiles%\Amazon\EC2ConfigService` and `C:\cfn\log`.)

Important When you set **Rollback on failure** to **No**, you will continue to incur AWS charges for this stack. Make sure to delete the stack when you finish troubleshooting.

For additional information, see [Troubleshooting AWS CloudFormation](#) on the AWS website.

Q. I encountered a size limitation error when I deployed the AWS CloudFormation templates.

A. We recommend that you launch the CF templates from the links in this guide or from another S3 bucket. If you deploy the templates from a local copy on your computer or from a non-S3 location, you might encounter template size limitations when you create the stack. For more information about AWS CloudFormation limits, see the [AWS documentation](#).

Q. I am not able to delete the stack that I created from the CF Template

A. We recommend that you clean up the registry bucket in your S3 account and retry deleting the stack.

Additional resources

AWS resources

- [Getting Started Resource Center](#)
- [AWS General Reference](#)
- [AWS Glossary](#)

AWS services

- [Amazon EC2](#)
- [Amazon Route 53](#)
- [Amazon S3](#)
- [Amazon VPC](#)
- [AWS CloudFormation](#)

IBM Cloud Pak for Data documentation

- [IBM Knowledge Center](#)
- [Product Hub](#)
- [Red Hat OpenShift Container Platform](#)

Document revisions

Date	Change	In sections
August 2021	First Marketplace Release	
September 2021	Cloud Pak for Data v4.0.1 (including Db2 and DataStage Cartridges)	Launch the Deployment Cloud Pak for Data Services

November 2021	Cloud Pak for Data v4.0.2 (including Cognos Analytics and Planning Analytics Cartridges)	Launch the Deployment Cloud Pak for Data Services
December 2021	Watson Discovery and OpenPages Cartridges	Launch the Deployment Cloud Pak for Data Services
June 2022	Added Open Data for Industries Cartridge	Launch the Deployment Cloud Pak for Data Services
October 2022	Cloud Pak for Data v4.5.3 Refresh	
December 2022	Cloud Pak for Data v4.6.0 Refresh	Launch the Deployment Cloud Pak for Data Services