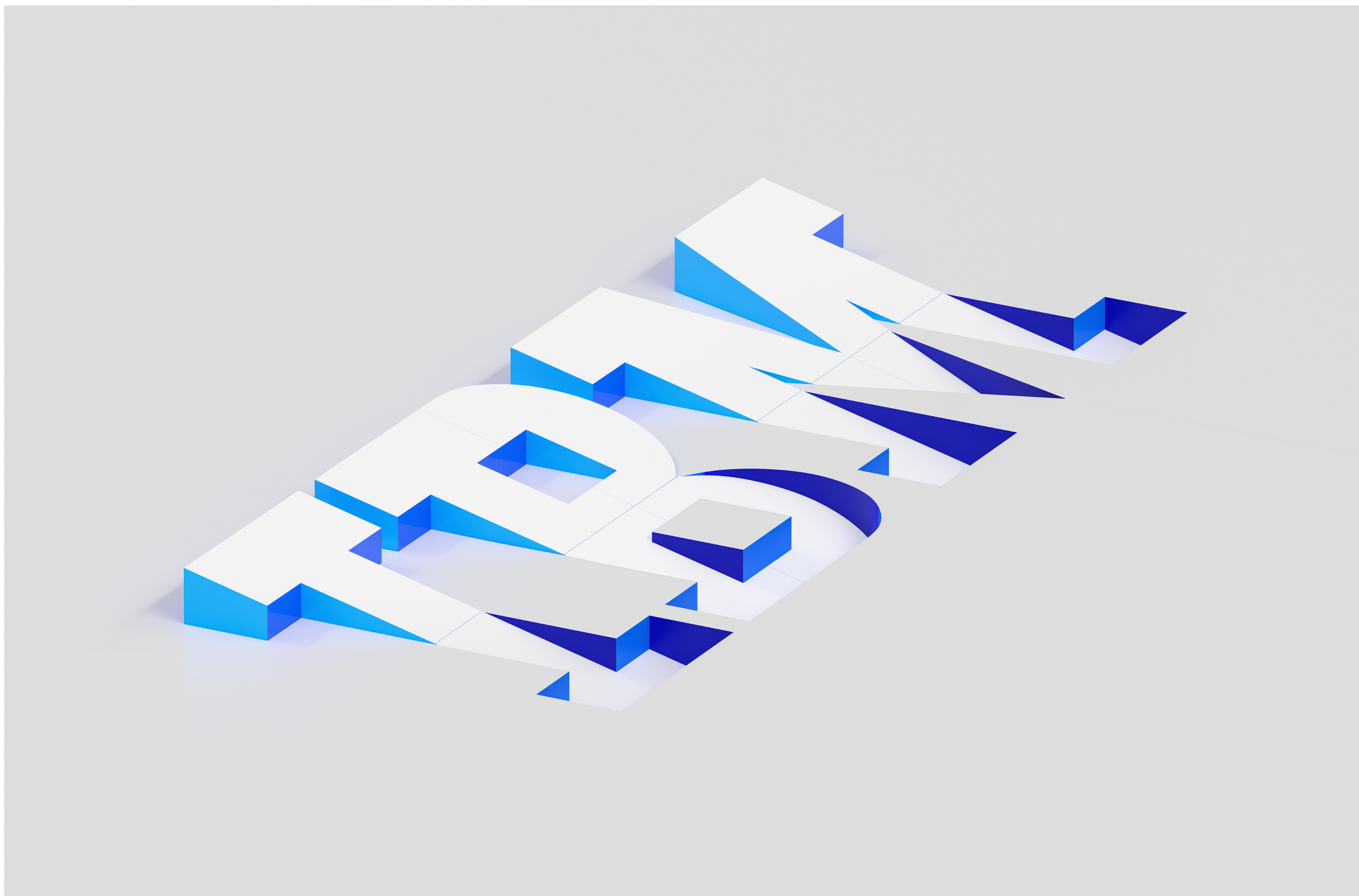# IBM Cloud Pak for Data

Product guide

IBM

Edition notices

This PDF was created on 2024-12-16 as a supplement to *IBM Cloud Pak for Data* in the IBM Cloud docs. It might not be a complete set of information or the latest version. For the latest information, see the IBM Cloud documentation at https://cloud.ibm.com/docs/cloud-pak-data?topic=cloud-pak-data-getting-started.

Getting started with IBM Cloud Pak for Data

Collect, organize, and analyze your data to generate meaningful insight with an extensible, end-to-end platform for governance, analytics, and AI that runs on Red Hat OpenShift on IBM Cloud. With IBM Cloud Pak® for Data, it's easy to find and access trusted data so that you can put your data to work quickly and efficiently. Make data-driven decisions and operationalize AI with trust and transparency throughout your business.

See the IBM Cloud Pak® for Data readme file for detailed information about getting started with Cloud Pak for Data on IBM Cloud.

Learn more about IBM Cloud Pak® for Data by reviewing the product documentation.

What's inside this Cloud Pak

You can choose which services to install when you install IBM Cloud Pak® for Data. For a complete list, see Installing services.

Supported versions

The current release of IBM Cloud Pak for Data on IBM Cloud is IBM Cloud Pak for Data Version 5.0.x.

Before you begin

Before you can install IBM Cloud Pak for Data, you must purchase a license through IBM Passport Advantage or register for a 60-day trial license. See Step 1. Assign the license.

You can deploy Cloud Pak for Data on Virtual Private Cloud (VPC) Gen2 infrastructure. You can use either a single or multi zone deployment. For more information, see Getting started with Red Hat OpenShift on IBM Cloud.

> **Note:** You must ensure that you have sufficient resources for the services that you plan to install. For more information, see the prerequisites for IBM Cloud Pak for Data.

Roles

To install IBM Cloud Pak® for Data on IBM Cloud, a user must have the following IBM Cloud Identity and Access Management (IAM) roles:

Role	Location	Action
Platform Editor	Manage > Account > Licenses and entitlements	Assign a license.
Service Manager	Manage > Access (IAM) > Roles > Kubernetes Service	Run the preinstallation script.
Service Writer	Manage > Access (IAM) > Roles > Kubernetes Service	Run the installation script.
Service Manager in any resource group	Schematics > Workspaces	Create a workspace.
Classic Infrastructure > Services > Storage Manage , Classic Infrastructure > Account > Add/Upgrade Storage	Manage > Access (IAM) > Users	Modify the image registry volume.

IBM Cloud Identity and Access Management roles required for Cloud Pak for Data

For more information, see IAM roles and actions.

Storage

You must ensure that your cluster has sufficient resources and is configured to use supported storage. You can choose one of the following storage options:

- Single zone VPC Gen2 cluster with storage OpenShift® Data Foundation
- Multi zone VPC Gen2 cluster with storage OpenShift® Data Foundation

You must also ensure that your cluster has sufficient resources and is configured to use supported storage.

Step 1. Assign the license

If you don't already have a license, you can:

- Purchase a license through [IBM Passport Advantage](#).
- [Register](#) for a 60-day trial license of IBM Cloud Pak for Data.

Important: The trial is for IBM Cloud Pak for Data software only. The trial does not include entitlement to the Red Hat OpenShift Container Platform.

To assign your license, follow these steps:

1. Log in to your [IBM Cloud account](#).
2. If you don't see any licenses to assign, navigate to Manage > Account and then click Licenses and entitlements in the navigation menu.
3. If there are no licenses to assign on the Licenses and entitlements page, click Check IBM Passport Advantage.
4. Select the appropriate license and click Assign.

Step 2. Configure your installation environment

Specify where you want to install IBM Cloud Pak for Data:

1. Select the Red Hat OpenShift on IBM Cloud cluster where you want to deploy IBM Cloud Pak for Data.
2. Enter or select the Red Hat OpenShift on IBM Cloud project where you want to deploy IBM Cloud Pak for Data.

You can also install IBM Cloud Pak for Data on IBM Cloud Satellite®. For more information about installing IBM Cloud Pak for Data on Satellite locations, see [Cloud deployment environments](#).

Step 3. Configure your workspace

Specify how you will track and manage your installation:

1. Enter or select a name for the installation.
2. Consider changing the default resource group.
3. Specify any tags that you want to use for the installation. Specify multiple tags as a comma-separated list.

Step 4. Complete the preinstallation task

A Red Hat OpenShift on IBM Cloud cluster administrator must complete this step. Specifically, the administrator must have an [access](#) policy in IBM Cloud Identity and Access Management that has an Operator role or higher.

- If you are not an administrator, use the Share link to share the script with your cluster administrator.
- If you are a cluster administrator, click Run script to run the preinstallation set up on your cluster.

The preinstallation script makes the following changes to your Red Hat OpenShift on IBM Cloud cluster:

- Increases the size of the Docker registry to 200 GB. This change increases the cost of your Red Hat OpenShift on IBM Cloud cluster.
- Creates the security context constraints that are required for IBM Cloud Pak® for Data.
- Grants access to the security context constraints to the service accounts that are required for IBM Cloud Pak® for Data.

Confirm that the script completes successfully before you proceed.

If the cluster administrator is not allowed to modify the storage, or the infrastructure account is not the same as the current account, a storage administrator can manually execute the script that is provided in [Complete the preinstallation section](#).

Step 5. Set the deployment values

Choose the Block Storage for VPC ODF storage class that you want to use to provision storage volumes. For multizone clusters, use a storage class with the VolumeBindingMode of WaitForFirstConsumer. For more information, see [Setting up Block Storage for VPC](#).

Specify which services to install when you install IBM Cloud Pak for Data. For example, to install Watson OpenScale, set `aiopenscale` to `true`.

If you don't select any services to install in this step, only the IBM Cloud Pak for Data control plane will be installed.

If you want to install a service later, you can return to the Deployment values section and set the appropriate parameter to true or you can select a service from the IBM Cloud Pak for Data Services catalog and follow the installation instructions for the service.

For more information, see [Installing IBM Cloud Pak for Data](#).

Step 6. Install IBM Cloud Pak for Data

1. Ensure that you have assigned a license for IBM Cloud Pak for Data to the deployment.

2. Confirm that you have read and agree to the license agreements.

3. Click Install.

The IBM Cloud Pak® for Data automated installation makes the following changes to ensure that services can be installed successfully:

- Sets kernel parameters. For more information, see [Changing kernel parameter settings](#).
- Enables `noroot squash` on worker nodes for Network File System (NFS). For more information, see [Storage considerations](#).

Step 7. Launch your instance of IBM Cloud Pak for Data

1. After you click Install, the Schematics > Workspaces page opens. You can watch the progress of the installation in the log.

2. When the installation completes, click Offering dashboard to access your IBM Cloud Pak for Data deployment.

3. Log in to the web client as `admin` using the default password (`password`). Change your password.

4. On the toolbar, click the Services icon and verify that your services are enabled or available.

Next steps

- To add users to your IBM Cloud Pak for Data deployment, see [Managing Cloud Pak for Data users](#).
- To install more services to a deployed cluster, repeat the steps to install from IBM Cloud Catalog and set the required service value to true in the Deployment values section.
- To install other supported services, such as DataStage, MongoDB, Db2, Db2 Big SQL, Cognos Analytics, Decision Optimization, Db2 Data Gate, Execution Engine for Hadoop, Open Pages or SPSS Modeler, which cannot be automatically installed when you install IBM Cloud Pak for Data on IBM Cloud, see [Services](#).
- To install and configure global image pull secrets on your cluster, set the value of configchanges to `Required` and provide your `IBM Cloud API Key` and your `IBM entitlement API` key. See [Updating the global image pull secret for IBM Cloud Pak for Data](#).
- To uninstall IBM Cloud Pak for Data or IBM Cloud Pak for Data services see [Uninstalling](#).

Installing services

If you deploy IBM Cloud Pak for Data on IBM Cloud, you can install a subset of the services that are available in the Cloud Pak for Data services catalog.

You can install services by using two different methods. The supported methods depend on the services that you want to install:

- Some services can be installed from the [Cloud Pak for Data installation page](#) in the IBM Cloud catalog.
- Some services can be installed by creating a catalog source and an operator subscription for each service. For general information and prerequisites, see [Installing IBM Cloud Pak for Data](#). To use this method, you must:
  - Use the Red Hat OpenShift command-line interface (oc CLI) to connect to the cluster. For more information, see [Connecting to the cluster from the CLI](#).
  - Create an operator subscription for each service that you plan to install. For more information, see [Installing an instance of IBM Cloud Pak for Data](#).

Use the links in the following table to learn how to install each of these services.

Service	Install from the Cloud Pak for Data installation page	Install by using the CLI
[Analytics Engine Powered by Apache Spark](#)	✓	✓
[AI Factsheets](#)	✓	✓
[Cognos Analytics](#)		✓
[Cognos Dashboards](#)	✓	✓
[Data Refinery](#)	See Watson Knowledge Catalog or Watson Studio	See Watson Knowledge Catalog or Watson Studio
[DataStage](#)		✓
[Data Virtualization](#)	✓	✓
[Db2 Big SQL](#)		✓
[Db2 Data Management Console](#)	✓	✓
[Db2 Warehouse](#)	✓	✓
[Db2](#)		✓
[Decision Optimization](#)		✓
[IBM Match 360](#)		✓
[MANTA Automated Data Lineage](#)	✓	✓
[OpenPages](#)		✓
[Planning Analytics](#)		✓
[RStudio Runtimes](#)	✓	✓
[SPSS Modeler](#)		✓
[Watson Discovery](#)	✓	✓

Watson Knowledge Catalog	✓	✓
Watson Machine Learning	✓	✓
Watson OpenScale	✓	✓
Watson Pipelines		✓
Watson Speech	✓	✓
Watson Studio	✓	✓
Watsonx Assistant	✓	✓

Methods to install Cloud Pak for Data

Uninstalling

You can uninstall services from IBM Cloud Pak® for Data, or you can uninstall IBM Cloud Pak® for Data.

Uninstalling IBM Cloud Pak® for Data

To uninstall IBM Cloud Pak® for Data, you can use the Schematics Workspaces console or the command-line interface.

Note: The uninstall action deletes the project. Before you delete a project, make sure that the project does not contain any other applications. For more information, see Uninstalling the platform and services.

To uninstall IBM Cloud Pak® for Data from the Schematics Workspaces console, follow these steps:

1. Navigate to Schematics > Workspaces.
2. Select the workspace of your installed IBM Cloud Pak® for Data, and then click Delete.
3. Or, click the actions menu for your workspace, and then click Delete.
4. Choose Delete workspace or Delete all associated resources or both, and enter the name of the workspace to confirm.
5. Click Delete to delete the workspace and wait for the uninstall to complete.
6. Verify that IBM Cloud Pak® for Data is uninstalled by accessing the IBM Managed Red Hat OpenShift Cluster web console and verifying that the components that are related to IBM Cloud Pak® for Data, such as any related pods, are no longer installed.

To uninstall IBM Cloud Pak® for Data by using the command-line interface, see Uninstalling an instance of IBM Cloud Pak for Data.

Uninstalling services in IBM Cloud Pak® for Data

To uninstall a service in IBM Cloud Pak® for Data, see the uninstall topic for Services.

Limitations and known issues

- The Cloud Pak for Data on IBM Cloud installation from the IBM Cloud catalog is not supported on Red Hat OpenShift Satellite clusters.

- The installation does not verify whether you have enough resources on the cluster to install Cloud Pak for Data. If you are running other applications on your Red Hat OpenShift cluster, make sure that you have enough resources on the cluster before you install Cloud Pak for Data.

- When you create a service instance, the service instance initially appears as Failed or Red in the Cloud Pak for Data console until the persistent volume claim is bound to a persistent volume on the cluster.

- On the Settings page in Schematics > Workspaces, you cannot change the Override value from `false` to `true` to install a service. For more information, see [Installing services](#).

- On the Settings page in Schematics > Workspaces, you cannot change the Override value from `true` to `false` to uninstall a service. For more information, see [Uninstalling](#).

For more information, see [Troubleshooting problems in Cloud Pak for Data](#).

IBM