

*Technical preview: Installing the Data
Virtualization service patch*



Contents

- Overview.....1**
- Preparing to install the Data Virtualization service patch..... 1**
- Installing the Watson Knowledge Catalog service patch.....2**
- Installing the Data Virtualization service..... 4**
- Provisioning the service..... 7**
- Installing the Data Virtualization service patch.....9**

Installing the Data Virtualization service patch

You can use the Data Virtualization service to create data sets from disparate data sources, so that you can query and use the data as if it came from a single source.

About this task

This document describes how to install the Data Virtualization service patch v1.3.0.0-284. This patch contains fixes and features that are required to use Watson™ Knowledge Catalog in Data Virtualization.

These instructions assume that you have the following products already installed in your Red Hat OpenShift cluster:

- IBM® Cloud Pak for Data. For more information, see [Installing Cloud Pak for Data on a Red Hat OpenShift cluster](#).
- Watson Knowledge Catalog. For more information, see [Installing the Watson Knowledge Catalog service](#).

As an overview, to install a Data Virtualization service patch, you must complete the following tasks:

1. [“Preparing to install the Data Virtualization service patch” on page 1.](#)
2. [“Installing the Watson Knowledge Catalog service patch ” on page 2.](#)
3. [“Installing the Data Virtualization service” on page 4.](#)
4. [“Provisioning the service” on page 7.](#)
5. [“Installing the Data Virtualization service patch” on page 9.](#)

Preparing to install the Data Virtualization service patch

Use this information to plan and prepare to install the Data Virtualization service patch v1.3.0.0-284.

Procedure

1. Add the `dv-dedicated=dv` label to a compute node on the IBM Cloud Pak for Data cluster:
 - a) Run the following command to get a list of nodes:

```
oc get nodes
```

- b) From the list of nodes, choose one node that has the compute role.
For example, you can select a node on which Red Hat OpenShift can schedule pods.
- c) Run the following command to add the label to the selected node:

```
oc label nodes node_name dv-dedicated=dv
```

Replace *node_name* with the name of the selected compute node.

2. Set semaphore parameters to meet the minimum required values on the selected compute node:
 - a) Edit the `/etc/sysctl.conf` file to add the following parameter:

```
kernel.sem="250 256000 100 4096"
```

For more information, see [Setting up semaphore parameters](#).

- b) Run the following command to apply your changes:

```
sysctl -p
```

3. [“Installing the Watson Knowledge Catalog service patch”](#) on page 2

Installing the Watson Knowledge Catalog service patch

The Watson Knowledge Catalog service patch 3.0.0.1 contains defect fixes and must be installed before you install the Data Virtualization patch.

Before you begin

Required role: To complete this task, you must be a cluster administrator.

Before you can install this patch, you must have installed IBM Cloud Pak for Data 2.5 and the Watson Knowledge Catalog service.

Additionally, before you install this patch, you must perform the following steps:

1. Log in to your Red Hat OpenShift cluster as a project administrator:

```
oc login OpenShift_URL:port
```

2. Open the service YAML file by running the following command:

```
oc edit deployment omag
```

3. If not present in the file, add the following lines:

```
env:
  - name: PORT
    value: "8080"
  - name: XMETA_DB_DRIVER
    value: com.ibm.db2.jcc.DB2Driver
  - name: XMETA_DB_URL
    value: jdbc:db2://is-xmetadocker:50000/xmeta
  - name: XMETA_DB_USER
    value: xmeta
  - name: XMETA_DB_PASSWORD
    valueFrom:
      secretKeyRef:
        name: iis-usr-secrets
        key: xmeta_password
```

4. Ensure that the `initContainers` file section contains the following lines. The value of `'image: docker-registry.default.svc:5000/icp4d-test/wkc-init-container:1.0.44'` is an example. Replace it with your location to store the images in the registry server. For more information, see [Setting up your registry server](#).

```
initContainers:
  - args:
    - sleep 60; while [ ! -f /tmp/jwtkey.cer ]; do sleep 2; done;
    command:
    - sh
    - -c
    - --
    image: docker-registry.default.svc:5000/icp4d-test/wkc-init-container:1.0.44
    imagePullPolicy: IfNotPresent
    name: wait-services
    resources:
      limits:
        cpu: "0"
        memory: "0"
      requests:
        cpu: "0"
        memory: "0"
    terminationMessagePath: /dev/termination-log
    terminationMessagePolicy: File
    volumeMounts:
    - mountPath: /tmp
      name: secrets-pv-volume
  - args:
    - kafka_status=1; while [ $kafka_status != 0 ]; do sleep 2; kafka_status=`nc kafka 9092 < /dev/null; echo $?`; done;
```

```

command:
- sh
- -c
- --
image: docker-registry.default.svc:5000/icp4d-test/wkc-init-container:1.0.44
imagePullPolicy: IfNotPresent
name: wait-kafka
resources:
  limits:
    cpu: "0"
    memory: "0"
  requests:
    cpu: "0"
    memory: "0"
terminationMessagePath: /dev/termination-log
terminationMessagePolicy: File
- args:
- redis_status=1; while [ $redis_status != 0 ]; do sleep 2; redis_status=`nc
  redis-ha 6379 < /dev/null; echo $?`; done;
command:
- sh
- -c
- --
image: docker-registry.default.svc:5000/icp4d-test/wkc-init-container:1.0.44
imagePullPolicy: IfNotPresent
name: wait-redis
resources:
  limits:
    cpu: "0"
    memory: "0"
  requests:
    cpu: "0"
    memory: "0"
terminationMessagePath: /dev/termination-log
terminationMessagePolicy: File
restartPolicy: Always
schedulerName: default-scheduler
securityContext: {}
terminationGracePeriodSeconds: 30

```

5. Save changes and quit by using the **:wq** command.

Procedure

To install the Watson Knowledge Catalog service patch 3.0.0.1:

1. Go to the Cloud Pak for Data Github repository and fork the following repositories:

- [wkc-base](#).
- [wkc-lite](#).

2. Change to the directory for the Cloud Pak for Data command-line interface.

3. Run the following command to download the required files to your local machine:

```

./cpd-Operating_System preloadImages --repo repo.yaml \
--assembly wkc \
--patch-name wkc-patch-3.0.0.1 \
--action download \

```

4. From the machine that can connect to the cluster, run the following commands to push the images to the registry server.

Important: If you are using the internal registry server on Red Hat OpenShift:

a) Log in to the **oc** command-line interface before you push the images:

```
oc login
```

b) If you are using the default self-signed certificate, specify the **--insecure-skip-tls-verify** parameter to prevent x509 errors.

```

./cpd-Operating_System preloadImages \
--assembly wkc \
--patch-name wkc-patch-3.0.0.1 \

```

```
--action push \
--ask-push-registry-credentials \
--load-from Image_Directory_Location \
--transfer-image-to Registry_Location \
```

Replace the following values:

Variable	Replace with
<i>Image_Directory_location</i>	The location of the cpd-Operating_System-workspace directory.
<i>Registry_location</i>	The location to store the images in the registry server.

5. Run the following command to install the patch:

```
./cpd-linux patch \
--namespace Project \
--load-from Image_directory_location \
--assembly wkc \
--patch-name wkc-patch-3.0.0.1 \
--transfer-image-to=Registry_location \
--ask-push-registry-credentials \
--action push
```

Replace the following values:

Variable	Replace with
<i>Project</i>	The project (namespace) where the IBM Cloud Pak for Data control plane is installed.
<i>Image_Directory_location</i>	The location of the cpd-Operating_System-workspace directory.
<i>Registry_location</i>	The location to store the images in the registry server.

Installing the Data Virtualization service

You can install the Data Virtualization service on IBM Cloud Pak for Data.

Before you begin

Ensure that you meet the requirements for this service. For more information, see [“Preparing to install the Data Virtualization service patch”](#) on page 1.

Required role: To complete this task, you must be an administrator of the project (namespace) where you will deploy the service.

Before you can install the service, a cluster administrator must complete the steps in [Setting up the cluster for the Data Virtualization service](#).

Ensure that the Mac OS or Linux machine where you will run the commands meets the appropriate requirements for your environment:

Ensure that you have the following information from your Red Hat OpenShift cluster administrator:

Required information	Description
<i>OpenShift_URL:port</i>	The URL and port number to use when logging in to your Red Hat OpenShift cluster. Ensure that you have the appropriate credentials to log into the cluster using <code>oc login</code> . Value:

Required information	Description
	Your cluster administrator should tell you whether your cluster is connected to the internet or is air-gapped.
<i>Assembly_version</i> Needed for air-gapped installations only.	The version of the assembly to install. Value:
<i>Storage_class_name</i>	The name of the storage class to use to provision storage for the service. If your cluster is not set up to use dynamic storage provisioning, work with an IBM Support representative to determine how you can specify persistent volume claims when you install the service. Value:
<i>Registry_location</i>	The location to store the images in the registry server. If you are installing the service when you are connected to the internet, ensure that you have the appropriate credentials to <i>push</i> images to the registry server. Value: Guidance for Red Hat OpenShift registry users: <ul style="list-style-type: none"> This is the external route to the location in the registry. The default external route is: <pre>docker-registry-default.9.87.654.321.nip.io/project</pre>Where <code>default.9.87.654.321.nip.io</code> is your public IP address. When you specify a value for the <i>Registry_location</i> variable, ensure that you include the <i>project</i> name.
<i>Registry_from_cluster</i>	The location from which pods on the cluster can <i>pull</i> images. Value: Guidance for Red Hat OpenShift registry users: <ul style="list-style-type: none"> This is the internal name of the registry service. The default service name is: <pre>docker-registry.default.svc:5000/project</pre> When you specify a value for the <i>Registry_from_cluster</i> variable, ensure that you include the <i>project</i> name.
<i>Project</i>	The project (namespace) where the IBM Cloud Pak for Data control plane is installed. Value:

Procedure

To install the service, complete the following tasks:

1. Change to the directory where you placed the Cloud Pak for Data command-line interface and the `repo.yaml` file.

Tip: For a list of all available options, enter the command: `./cpd-Operating_System --help`.

2. Run the `cpd adm` command with the following parameters to *preview* the list of resources that must be created on the cluster:

```
./cpd-Operating_System adm --repo repo.yaml \  
--assembly dv \  
--namespace Project
```

Replace the following values:

Variable	Replace with
<i>Operating_System</i>	For Linux, specify <code>linux</code> . For Mac OS, specify <code>darwin</code> .
<i>Project</i>	The project where you will install the Cloud Pak for Data control plane.

The command returns a list of the changes that you must make to your cluster to ensure that the control plane can run on your cluster, including the creation of service accounts and SCC bindings and the configuration of all of the necessary resources.

3. Re-run the `cpd adm` command with the `--apply` flag:

```
./cpd-Operating_System adm --repo repo.yaml \  
--assembly dv \  
--namespace Project \  
--apply
```

Replace the variables with the same values that you used the last time you ran the command.

4. Log in to your Red Hat OpenShift cluster as a project administrator:

```
oc login OpenShift_URL:port
```

5. Change to the project where you will install the Cloud Pak for Data control plane:

```
oc project Project
```

6. Run the following command to install the service:

```
./cpd-Operating_System --repo ./repo.yaml \  
--assembly dv \  
--namespace Project \  
--storageclass Storage_class_name \  
--transfer-image-to Registry_location \  
--cluster-pull-prefix Registry_from_cluster \  
--ask-push-registry-credentials
```

Variable	Replace with
<i>Operating_System</i>	For Linux, specify <code>linux</code> . For Mac OS, specify <code>darwin</code> .
<i>Project</i>	Use the value provided by your cluster administrator.
<i>Storage_class_name</i>	Use the value provided by your cluster administrator.
<i>Registry_location</i>	Use the value provided by your cluster administrator.
<i>Registry_from_cluster</i>	Use the value provided by your cluster administrator.

7. Enable the **hostIPC** parameter to allow pod containers to share the host IPC namespace:

a) Edit the `scc dv-scc` file by running the following command:

```
oc edit scc dv-scc
```

b) Find the `allowHostIPC` parameter and ensure that this parameter is set to `true`.

```
allowHostIPC: true
```

c) Get the name of the pod where the Data Virtualization service is installed by running the following command:

```
DV_SERVICE_POD=$(oc get pods | grep -i dv-addon | cut -d' ' -f1)
```

d) Download the main Helm chart for the Data Virtualization service by running the following command:

```
oc cp $DV_SERVICE_POD:/content/charts/ibm-dv-1.0.0.tgz ibm-dv-1.0.0-original.tgz
```

e) Untar the downloaded Helm chart by running the following command:

```
tar -zxf ibm-dv-1.0.0-original.tgz
```

f) Find the `hostIPC` parameter in the `headPodSecurityContext` section of the `Edit ibm-dv/templates/_sch-chart-config.tpl` file. Ensure that this parameter is set to `true`.

```
hostIPC: true
```

g) Package the service Helm chart by running the following command:

```
tar -pczf ibm-dv-1.0.0.tgz ibm-dv
```

h) Upload the service Helm chart by running the following command:

```
oc cp ibm-dv-1.0.0.tgz $DV_ADDON_POD:/content/charts/ibm-dv-1.0.0.tgz
```

What to do next

[“Provisioning the service” on page 7.](#)

Provisioning the service

Before you use Data Virtualization, you must provision the service to your IBM Cloud Pak for Data.

About this task

The following instructions describe how to provision the Data Virtualization service by using the Cloud Pak for Data user interface. The URL of the Cloud Pak for Data user interface must have the following format:

```
https://Project-cpd-Project.apps.FQDN-LB-node/zen.
```

Replace the following values:

Variable	Replace with
<i>Project</i>	The project where you will install the Cloud Pak for Data control plane.
<i>FQDN-LB-node</i>	The fully qualified domain name of the load balancer node.

The Data Virtualization service is provisioned to any node in the Cloud Pak for Data cluster that has the specified resources (cores and memory) available.

If you have multiple nodes that match the specified criteria, it is recommended that you create the persistent storage on each of those nodes. One or more persistent volumes and associated persistent volume claims are required for external libraries, cache entries, and queries. You can use any physical storage in your environment for the persistent volumes. The persistent volume claim must be scoped to the namespace or OpenShift project that you choose.

External libraries

External libraries (that is, libraries that are not included in the Data Virtualization service) are stored on a persistent volume. You can have the provisioning process create the persistent volume claim by specifying a storage class or you can choose an existing persistent volume claim. The benefits of choosing an existing persistent volume claim:

- You can use the persistent volume claim from a Data Virtualization service instance that is deleted to create a new service instance.
- You can use the same persistent volume claim for other Data Virtualization service instances, which can decrease the time that is required to provision the service instance.

The persistent volume claim for external libraries must have at least 10 GB available.

Cache storage

A data cache holds temporary data that is used very frequently. By using data cache, you can reduce processing and loading time required when you use this data.

Important: To complete this task, you must have the **Provision Databases** permission. The default Cloud Pak for Data administrator role, **Admin**, has this permission.

Procedure

To provision the Data Virtualization service:

1. Click the **Services** icon () from the Cloud Pak for Data web user interface.
2. From the list of services, locate the Data Virtualization service under the **Data sources** category. Click the action menu and select **Provision instance**.
3. Ensure that the **Set up semaphore parameters** box is unchecked.
4. To configure the service, specify the resources that you want to allocate to the head node and worker nodes in the **Nodes** section.

A head node is the main server node where the Data Virtualization service runs. The head node includes, among other components, one worker node. Thus, the amount of memory allocated to the head node is defined by the amount of resources that you allocate to your worker nodes.

- a) Specify the number of worker nodes to allocate to the service.

Recommended: One node is sufficient for most workloads.

- b) Specify the number of cores to allocate to each worker node.

You are constrained by the total number of available cores on the node. If your node has 16 available cores, you can specify up to 16 cores. The maximum number of cores you can specify is 64.

- c) Specify the amount of memory to allocate to each worker node.

You are constrained by the total amount of memory on the node. If your node has 32 GB of memory, you can specify up to 32 GB. The maximum memory that you can specify is 512 GB.

The amount of memory allocated to the head node is defined automatically based on the memory allocated to the worker nodes.

5. In the **Storage** section, specify the resources that you want to use for persistent storage and cache storage.

In the **Cache storage** section, you can configure storage for your data caches.

Note: Part of the total cache storage space is reserved to refresh active caches. This impacts the storage space that is available for creating new cache entries.

In the **Persistent storage** section, you can configure persistent storage for external libraries.

- a) If you want to create a new persistent volume claim, select **Create new claim**. Then, select the storage class to use and specify the amount of storage to allocate to the persistent volume.

If you use Portworx for your persistent storage, select `portworx-shared-gp` for the **Storage class** option. For more information about storage considerations for Cloud Pak for Data, see [Storage considerations](#)

- b) If you want to use an existing persistent volume claim, select **Use existing claim**. Then, select the persistent volume claim that you want to associate with the service.

6. Click **Next**.

7. Ensure that the summary is correct and click **Provision**.

8. Optional: It can take approximately 10 minutes for the service to be provisioned. If you want to use Cloud Pak for Data while you wait for the provisioning process to complete, click **Home**.

What to do next

- You can start using the Data Virtualization service. For more information, see [Virtualizing data](#).
- When you provision the Data Virtualization service you are automatically assigned the **Data Virtualization Admin** role. After you provision the service, you must give other users access to the service. For more information, see [Managing users in Data Virtualization](#).
- To connect to the Data Virtualization service, use the JDBC URL that is provided in the **Connection details** page for the service. Additionally, if you have a load balancer, you must open the port in your load balancer and your firewall.

Installing the Data Virtualization service patch

The v1.3.0.0-284 service patch contains fixes required to use Watson Knowledge Catalog in Data Virtualization.

About this task

You can install the service patch by running the **cpd patch** command. The service patch file must be created under the Data Virtualization service module of a Cloud Pak for Data build.

Procedure

1. Go to the [Cloud Pak for Data Github repository](#) and fork the service patch repository.
2. Run the following command to download the required files to your local machine:

```
./cpd-Operating_System patch --repo repo.yaml \  
--assembly dv \  
--version v1.3.0.0 \  
--patch-name Patch_version \  

```

```
--action download
```

Replace the following values:

Required information	Description
<i>Operating_System</i>	For Linux, specify <code>linux</code> . For Mac OS, specify <code>darwin</code> .
<i>Patch_version</i>	The version of the patch to install. The current service patch is <code>v1.3.0.0-284</code>
<i>Image_directory_location</i>	The location of the <code>cpd-Operating_System-workspace</code> directory.

3. Edit the `patchCommands.txt` file to add or modify the following elements:

- a) Modify the values for the `dv-uc` and `dv-api` deployments to their corresponding names in your cluster:

```
edit cpd-Operating_System-workspace/modules/dv/x86_64/v1.3.0.0/patch/v1.3.0.0-284/patchCommands.txt
```

Replace the following values:

Required information	Description
<i>Operating_System</i>	For Linux, specify <code>linux</code> . For Mac OS, specify <code>darwin</code> .

For example, the `unified-console` and `dv-api` deployment names can have the following format:

```
dv-1-1572975697166-ibm-unified-console  
dv-3-1572975697166-ibm-dv-api
```

- b) Add the **dv-caching** patching command:

```
patch deployment dv-caching -p '{"spec": {"template": {"spec":{"initContainers": [{"name": "init-caching", "image":"{{.DockerRegistryPrefix }}/dv-caching:v1.3.0.0-284"}}]}}}'
```

- c) If your Data Virtualization service instance has more than one worker node, add **dv-worker** patching command:

```
patch statefulset dv-worker -p '{"spec": {"template": {"spec":{"containers": [{"name": "dv-worker", "image":"{{.DockerRegistryPrefix }}/dv-engine:v1.3.0.0-284"}}]}}}'
```

4. Change to the directory where you placed the Cloud Pak for Data command-line interface and the `repo.yaml` file.

5. Log in to your Red Hat OpenShift cluster as a project administrator:

```
oc login OpenShift_URL:port
```

6. Run the following command:

```
oc exec -it dv-0 -c dv-server -- sudo -u bigsql -i bigsql stop
```

7. Run the following command to install the service patch:

Important: If you are using the internal Red Hat OpenShift registry, do not specify the `--ask-pull-registry-credentials` parameter.

```
./cpd-Operating_System patch \  
--namespace Project \  
--load-from Image_Directory_Location \  
--assembly dv \  
--patch-name Patch_version \  
--transfer-image-to=Registry_Location \  
--ask-push-registry-credentials \  

```

```
--action push
```

Replace the following values:

Required information	Description
<i>Project</i>	The project (namespace) where the IBM Cloud Pak for Data control plane is installed.
<i>Patch_version</i>	The version of the assembly to install.
<i>Registry_from_cluster</i>	<p>The location from which pods on the cluster can <i>pull</i> images.</p> <p>Guidance for Red Hat OpenShift registry users:</p> <ul style="list-style-type: none">• This is the internal name of the registry service. The default service name is: <pre>docker-registry.default.svc:5000/Project</pre>• When you specify a value for the <i>Registry_from_cluster</i> variable, ensure that you include the <i>project</i> name.
<i>Image_directory_location</i>	The location of the <i>cpd-Operating_System-workspace</i> directory.
<i>Registry_location</i>	<p>The location to store the images in the registry server.</p> <p>If you are installing the service when you are connected to the internet, ensure that you have the appropriate credentials to <i>push</i> images to the registry server.</p> <p>Guidance for Red Hat OpenShift registry users:</p> <ul style="list-style-type: none">• This is the external route to the location in the registry. The default external route is: <pre>docker-registry-default.9.87.654.321.nip.io/project</pre> <p>Where <code>default.9.87.654.321.nip.io</code> is your public IP address.</p> <ul style="list-style-type: none">• When you specify a value for the <i>Registry_location</i> variable, ensure that you include the <i>project</i> name.

