

# Add Access to External S3 Buckets for CDW Clusters on AWS (Preview)

Date published: 2021-05-12

Date modified: 2021-05-25

## Legal Notice

© Cloudera Inc. 2021. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 ("ASLv2"), the Affero General Public License version 3 (AGPLv3), or other license terms.

Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER'S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*

## Contents

<b>Legal Notice</b>	<b>2</b>
<b>Contents</b>	<b>3</b>
<b>Adding access to external S3 buckets</b>	<b>4</b>
<b>Adding read only or read/write access using default encryption to external buckets in the same account</b>	<b>5</b>
<b>Adding read-only or read/write access using default encryption to external buckets in a different account</b>	<b>7</b>
<b>Adding a custom key for read/write access to external buckets</b>	<b>10</b>

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*

## Adding access to external S3 buckets

Learn how to configure access to external S3 buckets in Cloudera Data Warehouse (CDW) Public Cloud clusters running on AWS environments.

When you create a Virtual Warehouse in the CDW Public Cloud service, a cluster is created in your AWS account. This cluster has two buckets. One bucket is used for managed data and the other is used for external data. The naming convention for these two S3 buckets that are created by the CDW service is described below:

- `<s3-bucket-name>-<random-string>-dwx-managed`

For example, if you specified the bucket name `dwx-data` when you registered your environment with Management Console, the managed data S3 bucket might be named something like: `dwx-data-t8hq-dwx-managed`

- `<s3-bucket-name>-<random-string>-dwx-external`

Continuing the above scenario where you specified `dwx-data` as the bucket name during environment registration, the external S3 bucket might be named: `dwx-data-8nhs-dwx-external`

Access to these two buckets is controlled by [AWS instance profiles](#). To add additional external AWS S3 buckets to your CDW service cluster, you can use the CDW UI.

# Adding read only or read/write access using default encryption to external buckets in the same account

You can configure read only or read/write access using default encryption to external S3 buckets that reside in the same AWS account as the Cloudera Data Warehouse (CDW) Public Cloud cluster.

## About this task

### **Important:**

If you configure read only access to an external S3 bucket, there is no need to restart Virtual Warehouses. However, if you configure read/write access to an external S3 bucket, you must restart Virtual Warehouses by suspending them and starting them again. Alternatively, you can create a new Virtual Warehouse to use the external S3 bucket with read/write access.

**Required role:** DWAdmin

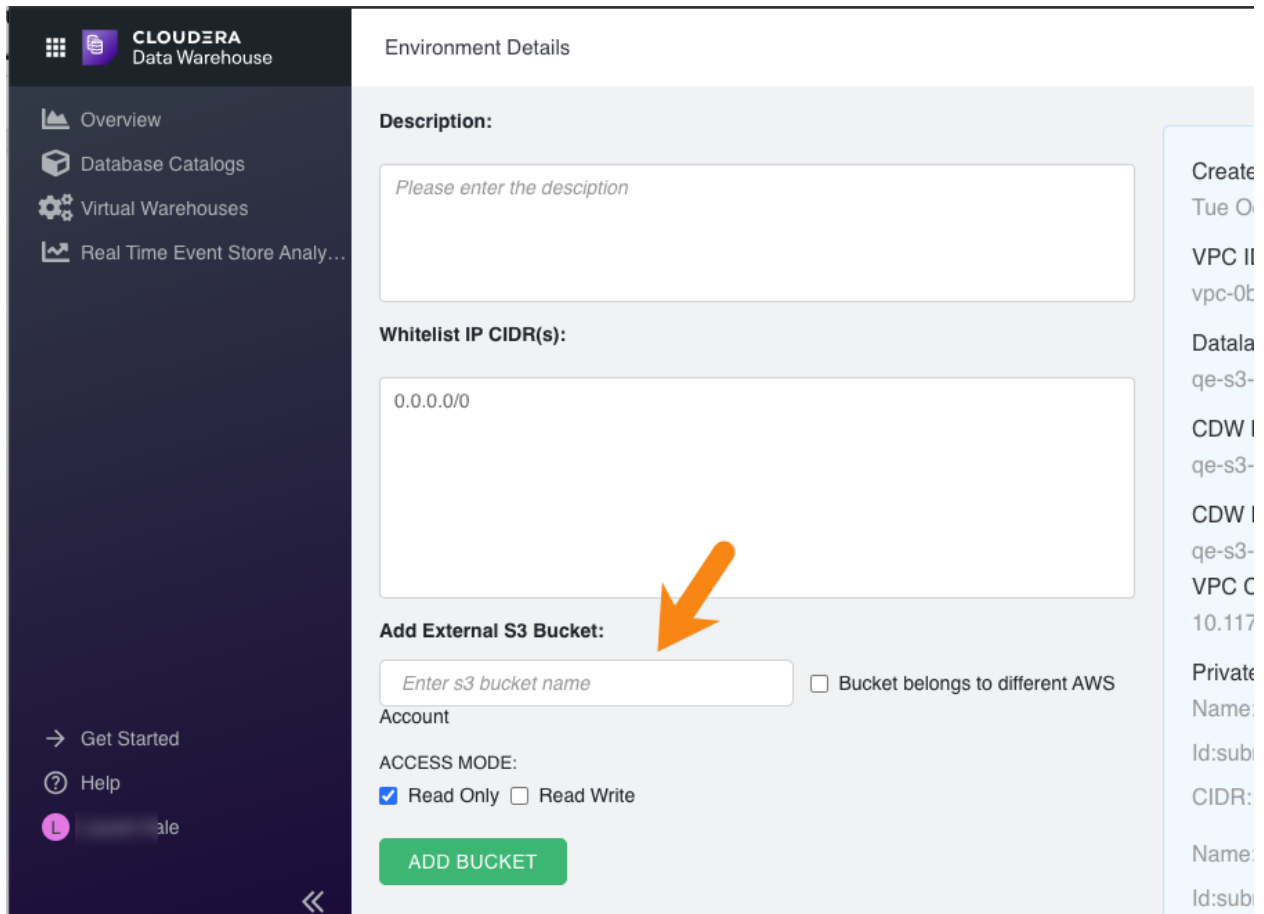
## Before you begin

- Identify which environment you want to configure for access to an external bucket in another AWS account. In the CDW UI, expand the Environments column by clicking the **More...** menu, and then click on the environment tile. This causes the Database Catalog and Virtual Warehouses that use this environment to be highlighted in the CDW UI. Choose the environment that is activated for the Virtual Warehouses you want to use with the external AWS bucket.
- In the AWS Management Console, identify the external S3 bucket you want to configure access to.

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*

**Steps**

1. On the CDW UI Overview page, navigate to the environment tile for the environment for which you want to configure access to an external AWS bucket, and then click the edit icon. This loads the Environment Details page.
2. In the Environment Details page, toward the bottom of the page, type the name of the AWS bucket you want to configure access to in the **Add External S3 Bucket** text box:



3. Specify whether **Read Only** or **Read Write** access is needed.

**Note:**  
**For Read Write access only:** If you do not need to use a custom encryption key, leave the **ENCRYPTION SETTINGS** text box blank. If you do not use a custom key, the system uses AES256 encryption by default. If you need to use a custom encryption key, see the "Related Information" section at the bottom of this page for a link to instructions for adding a custom key when you configure read/write access to an external S3 bucket. **Read Only** access does not involve encryption.

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*

4. Click **Add Bucket** to save the configuration. A success message displays at the top of the page.

**What to do next:**

If you have configured **Read Write** access, you must restart the Virtual Warehouses that are associated with this environment for the configuration changes to take effect.

## Adding read-only or read/write access using default encryption to external buckets in a different account

Learn how you can configure read-only or read/write access using default encryption to external S3 buckets in AWS accounts that are different from the account where the Cloudera Data Warehouse (CDW) Public Cloud cluster resides.

**About this task**

**Important:** If you configure read only access to an external S3 bucket, there is no need to restart Virtual Warehouses. However, if you configure read/write access to an external S3 bucket, you must either restart Virtual Warehouses by suspending them and starting them again. Alternatively, you can create a new Virtual Warehouse to use the external S3 bucket with read/write access.

**Required role:** DWAdmin

**Before you begin**

- Identify which environment you want to configure for access to an external bucket in another AWS account. In the CDW UI, expand the Environments column by clicking the **More...** menu, and then click on the environment tile. This causes the Database Catalog and Virtual Warehouses that use this environment to be highlighted in the CDW UI. Choose the environment that is activated for the Virtual Warehouses you want to use with the external AWS bucket that resides in a different AWS account.
- In the AWS Management Console for the different account, identify the external S3 bucket you want to configure access to.

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*

## Steps

1. On the CDW UI Overview page, navigate to the environment tile for the environment for which you want to configure access to an external AWS bucket, and then click the edit icon. This loads the Environment Details page.
2. In the Environment Details page, toward the bottom of the page, type the name of the AWS bucket you want to configure access to in the **Enter s3 bucket name** text box.
3. Select **Bucket belongs to a different AWS Account**. The CDW bucket policy displays.
4. In the window that displays the CDW bucket policy, click the copy icon on the right margin of the window:

**Add External S3 Bucket:**

Bucket belongs to different AWS Account

ACCESS MODE:  
 Read Only  Read Write


Please update source bucket policy to include the following. Refer this [link](#) for additional instructions.

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "read-write-access-for-cdw-env-xspjph",
      "Effect": "Allow",
      "Principal": {
        "AWS": [
          "arn:aws:iam::123014800043:role/func-qe-weekly-role",
          "arn:aws:iam::123014800043:role/env-xspjph-dwx-stack-NodeInstanceRole-1BR2OOVUXZH9F"
        ]
      },
      "Action": [
        "s3:Get*",
        "s3:ListBucket",
        "s3:GetBucketLocation"
      ],
      "Resource": [
        "arn:aws:s3:::foo",
        "arn:aws:s3:::foo/*"
      ]
    }
  ]
}

```

Click the copy icon.



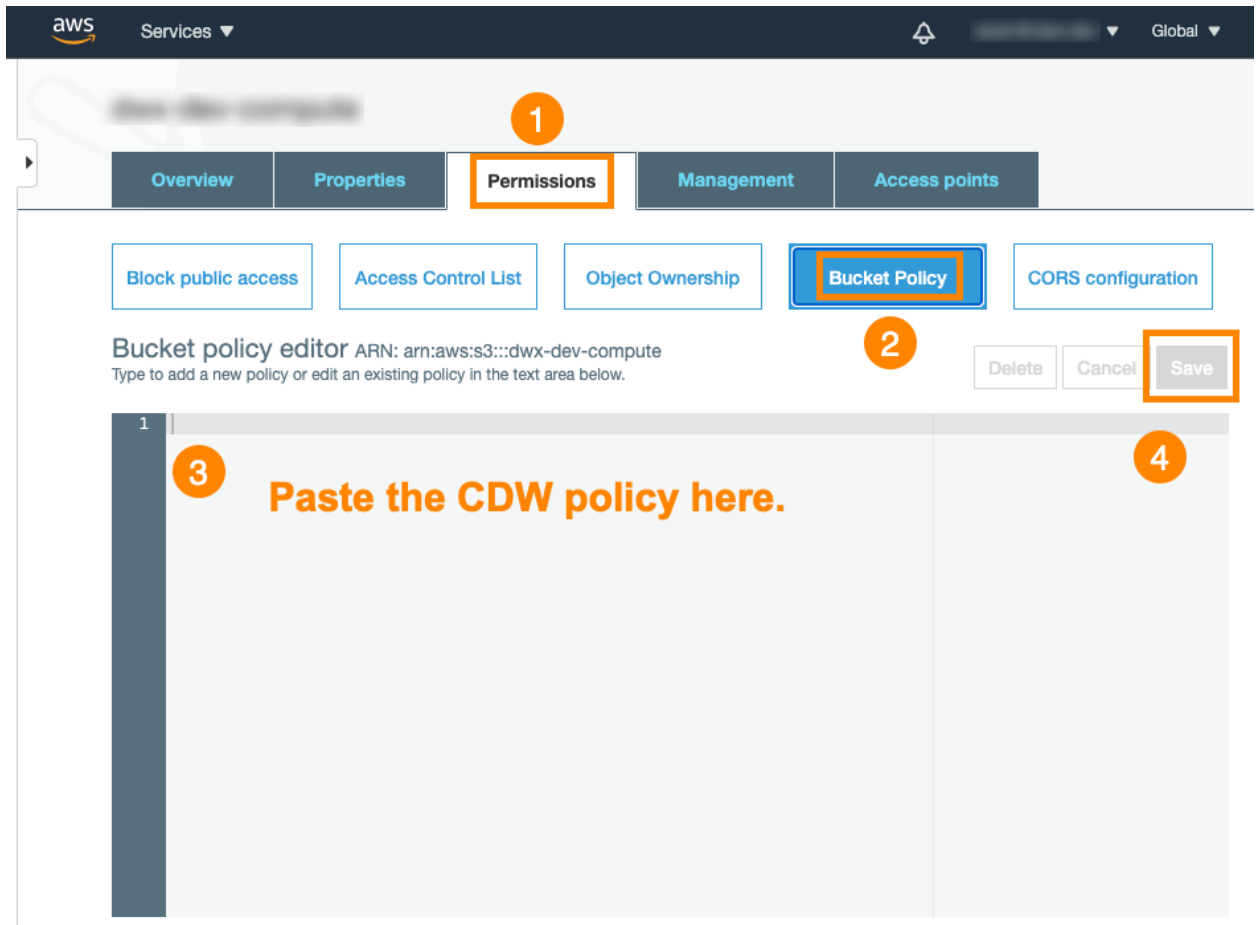
ADD BUCKET

5. Open the AWS Management Console for the different account where the external bucket is located and navigate to the bucket to which you want to configure access.

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*



- On the bucket details page of AWS Management Console, click the **Permissions** tab, click **Bucket Policy**, paste the policy from CDW, and click **Save**:



- In the CDW UI Environment Details page, specify either **Read Only** or **Read Write** access for the external bucket.

**Note:**  
**For Read Write access only:** If you do not need to use a custom encryption key, leave the **ENCRYPTION SETTINGS** text box blank. If you do not use a custom key, the system uses AES256 encryption by default. If you need to use a custom encryption key, see the "Related Information" section at the bottom of this page for a link to instructions for adding a custom key when you configure read/write access to an external S3 bucket. **Read Only** access does not involve encryption.

- Click **Add Bucket** to save the configuration. A success message displays at the top of the page.

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*

# Adding a custom key for read/write access to external buckets

You can also configure a custom key for read/write access to an external S3 bucket from Cloudera Data Warehouse (CDW) Public Cloud on AWS.

## About this task

Sometimes you have a custom encryption key configured for the external S3 bucket you want to access from CDW. In that case, you can perform the steps described in this topic to use your own custom key.

Encryption key configuration is only needed if you are configuring read/write access to the external bucket and you want to use your own custom encryption key.

### Important:

If you want to use a custom key, you must perform the configuration described in this topic whether the bucket is in the same AWS account as CDW or if the bucket resides in a different AWS account.

**Required role:** DWAdmin

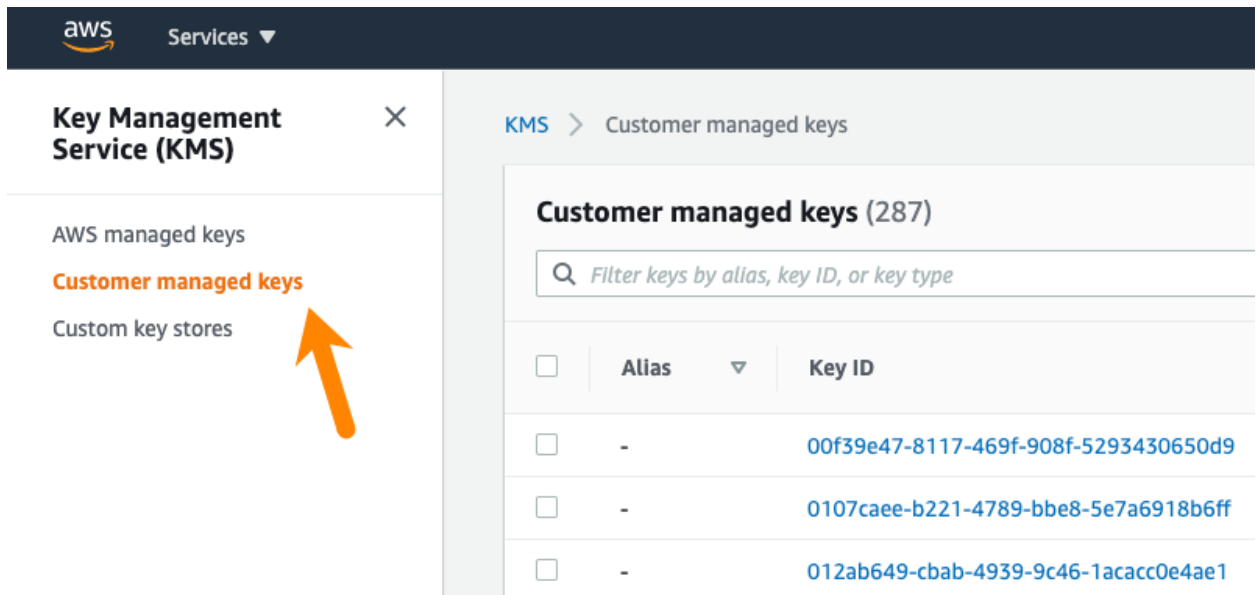
## Before you begin

- Identify which environment you want to configure for access to an external bucket in another AWS account. In the CDW UI, expand the Environments column by clicking the **More...** menu, and then click on the environment tile. This causes the Database Catalog and Virtual Warehouses that use this environment to be highlighted in the CDW UI. Choose the environment that is activated for the Virtual Warehouses you want to use with the external AWS bucket that resides in a different AWS account.
- In the AWS Management Console, identify the external S3 bucket you want to configure access to.

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*

**Steps**

1. On the CDW UI Overview page, navigate to the environment tile for the environment for which you want to configure access to an external AWS bucket, and then click the edit icon. This loads the Environment Details page.
2. In the Environment Details page, toward the bottom of the page, type the name of the AWS bucket you want to configure access to in the **Enter s3 bucket name** text box.
3. (Optional) If you want to configure access to an S3 bucket in a different AWS account, select **Bucket belongs to different AWS Account**.
4. Select **Read Write**. This causes the **ENCRYPTION SETTINGS** text box and a key policy to display.
5. In the AWS Management Console for the account where the S3 bucket resides, navigate to the Key Management Service, and select **Customer Managed Keys** in the left navigation menu:



On the Customer managed keys page, select the key you want to use.

- On the key details page, select the **Key policy** tab in the center panel of the page:

The screenshot shows the AWS KMS console interface. On the left, the 'Key Management Service (KMS)' sidebar is visible with 'Customer managed keys' selected. The main panel displays details for a specific key (Key ID: 00f39e47-8117-469f-908f-5293430650d9). Below the 'General configuration' section, there are four tabs: 'Key policy', 'Cryptographic configuration', 'Tags', and 'Key rotation'. An orange arrow points to the 'Key policy' tab. The 'Key policy' tab is active and shows a JSON policy document with the following content:

```

1 {
2   "Version": "2012-10-17",
3   "Id": "env-nb5pfv-dwx-stack-kms-key",
4   "Statement": [
5     {
6       "Sid": "Allow administration of the key",
7       "Effect": "Allow",
8       "Principal": {
9         "AWS": "arn:aws:iam::581[redacted]:root"
10      },
11      "Action": "kms:*",
12      "Resource": "*"
13    }
14  ]
15 }

```

This displays the key policy for the customer-managed key.

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*

- In the CDW Environment Details page, copy the Amazon Resource Names (ARNs) associated with the environment that displays in the panel:

**Add External S3 Bucket:**

foo  Bucket belongs to different AWS Account

ACCESS MODE:  
 Read Only  Read Write

ENCRYPTION SETTINGS: ⓘ

(Optional) Enter KMS CMK ARN

Please update source encryption key policy to include the following. Refer this [link](#) for additional instructions.

```
{
  "Sid": "encrypt-decrypt-objects-for-cdw-env-xspjph",
  "Effect": "Allow",
  "Principal": {
    "AWS": [
      "arn:aws:iam::123014800043:role/func-qe-weekly-role",
      "arn:aws:iam::123014800043:role/env-xspjph-dwx-stack-NodeInstanceRole-1BR2OOVUXZH9F"
    ]
  },
  "Action": [
    "kms:Encrypt",
    "kms:Decrypt",
    "kms:ReEncrypt*",
    "kms:GenerateDataKey*",
    "kms:DescribeKey"
  ],
  "Resource": ""
}
```

**Copy these 2 roles that are associated with this environment to your system clipboard.**

ADD BUCKET

The actions listed in the above screen image are the minimum set of actions needed by CDW:

```
"Action": [
  "kms:Encrypt",
  "kms:Decrypt",
  "kms:ReEncrypt*",
  "kms:GenerateDataKey*",
  "kms:DescribeKey"
]
```

The key policy you use should allow at least these actions.

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cludera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*

- Return to the key details page in the AWS Management Console, click **Edit** in the upper right corner of the **Key policy** tab, paste the 2 ARNs to append them after the existing ARNs in the key policy, and then click **Save changes**:

**Edit key policy**


**Key policy**

```

1 {
2   "Version": "2012-10-17",
3   "Id": "env-nb5pfv-dwx-stack-kms-key",
4   "Statement": [
5     {
6       "Sid": "Allow administration of the key",
7       "Effect": "Allow",
8       "Principal": {
9         "AWS": "arn:aws:iam::581665743680:root",
10        "arn:aws:iam::123014800043:role/func-qe-weekly-role":
11        "arn:aws:iam::123014800043:role/env-xspjph-dwx-stack-NodeInstanceRole-1BR200VUXZH9F"
12      },
13      "Action": "kms:*",
14      "Resource": "*"
15    }
16  ]
17 }

```

Paste the 2 ARNs under the existing ARNs in the key policy, and then click "Save changes."



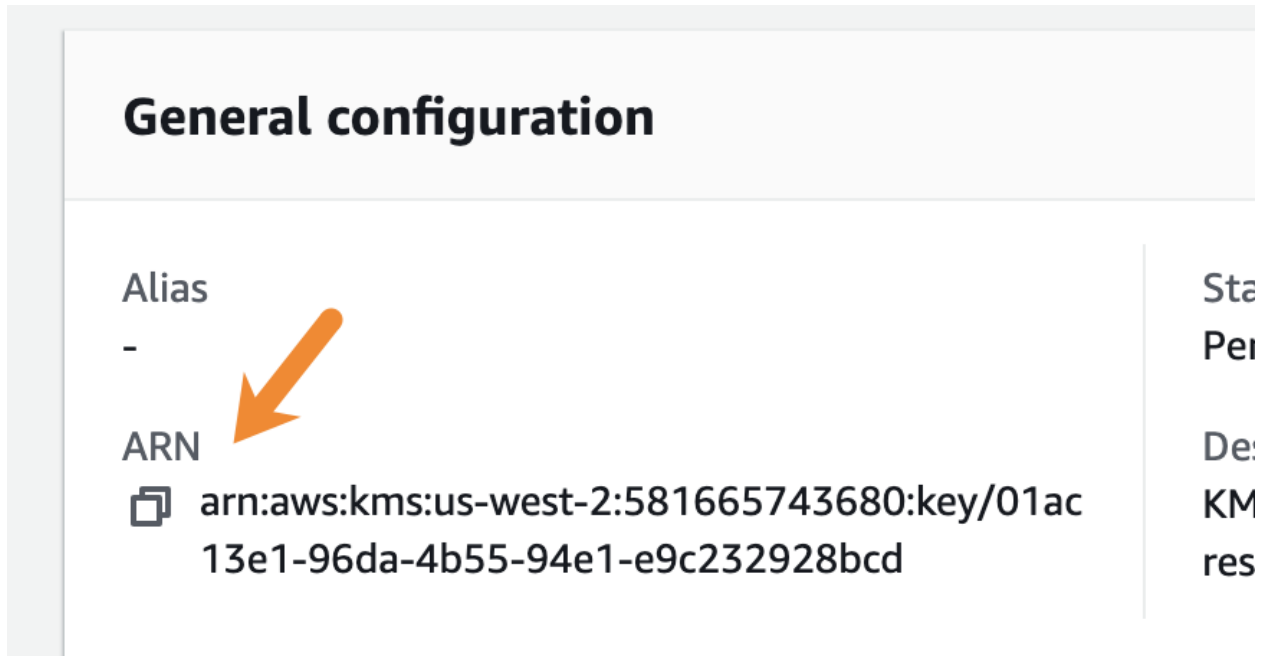
"arn:aws:iam::123014800043:role/func-qe-weekly-role":  
"arn:aws:iam::123014800043:role/env-xspjph-dwx-stack-NodeInstanceRole-1BR200VUXZH9F"

Cancel Save changes

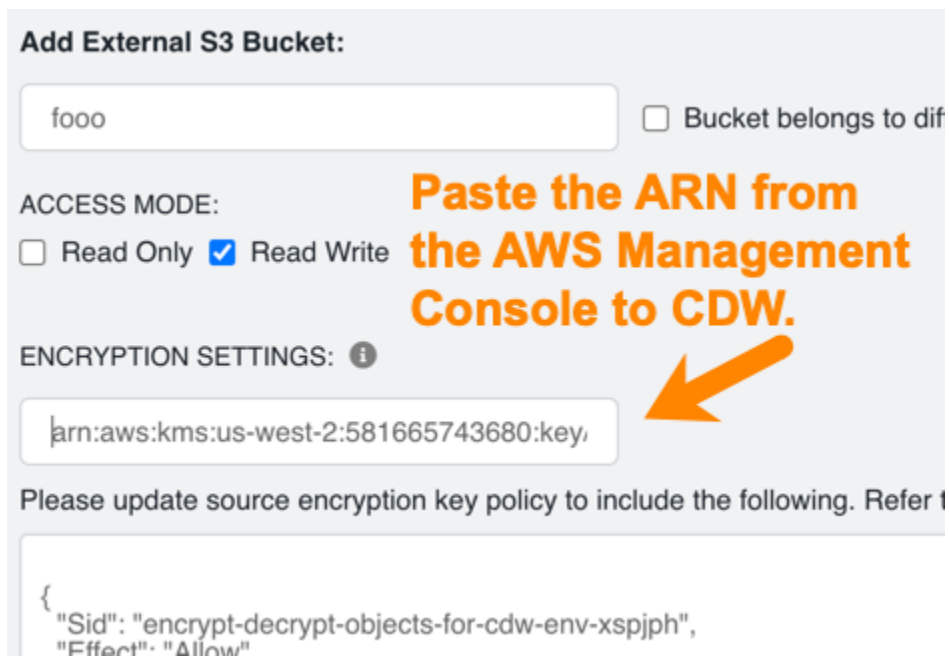
The key policy shown in the above screen capture lists "kms : \*" in the "Action" section of the policy. This indicates that *all actions are allowed*. At minimum, CDW requires the "Encrypt", "Decrypt", "ReEncrypt\*", "GenerateDataKey\*", and the "DescribeKey" actions as shown on the screen capture in Step 7.

If there is no key policy on the **Key Policy** tab of the key details page, copy and paste the entire key policy in the AWS Management Console from the CDW UI.

- After you save the changes to the key policy in the AWS Management Console, copy the ARN from the **General configuration** section of the key details page:



- In the CDW Environment details page, add the ARN you copied in Step 9 to the **ENCRYPTION SETTINGS** text box:



- Click **Add Bucket** to save the configuration. A success message displays at the top of the page.

*This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.*