

# Reserving nodes for auto-scaling (Preview)

Date published: 2022-07-26

Date modified: 2022-07-26

## Legal Notice

© Cloudera Inc. 2022. 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>Reserving nodes for auto-scaling</b>                           | <b>4</b> |
| Getting the Availability Zone used by the cluster overprovisioner | 5        |
| Changing the Availability Zone used by the overprovisioner        | 5        |
| Creating the Virtual Warehouse                                    | 6        |

*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.*

# Reserving nodes for auto-scaling

To speed up Virtual Warehouse startup and autoscaling, you need to keep some number of compute instances on standby. You need to know how to configure extra buffer nodes to stand by, ready to join a new compute or autoscaled cluster.

Important information about this feature is:

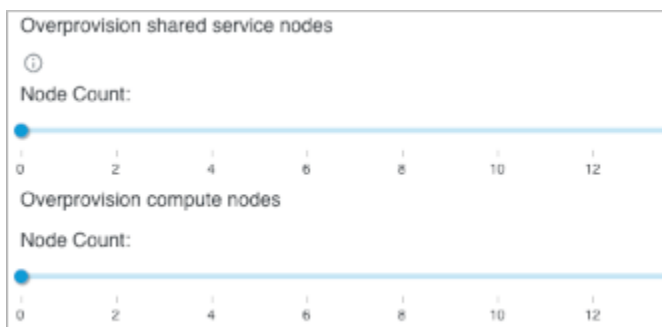
- This feature supports AWS and Azure.
- This feature is behind an entitlement: CDW\_CLUSTER\_OVERPROVISIONER.
- Adding more reserved nodes increases your cloud costs.
- This feature is a technical preview.

## Before you begin:

- You are logged into one of the following CDP environments in Cloudera Data Warehouse (CDW) Kubernetes Public Cloud service.
  - AWS
  - Azure
- You have the CDW\_CLUSTER\_OVERPROVISIONER entitlement to use this feature.

**Required role:** DWAdmin

**About this task:** You increase reserved nodes by configuring the **Overprovision compute nodes or Overprovision shared service nodes** when you activate the CDW environment. Use the slider to set the Node Count to the additional number of nodes you want to reserve. Compute nodes are used by executors and coordinators, while the shared service nodes are used by other services in the cluster. Compute nodes are relevant to autoscaling events as executors or coordinators are scaled up/down during autoscaling.



Adding more reserved nodes increases your cloud costs.

*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.*

## Getting the Availability Zone used by the cluster overprovisioner

When using this feature for compute nodes in AWS, make sure the Availability Zone selected during Virtual Warehouse creation is the same as the one being used by the cluster overprovisioner.

You can determine Availability Zone used by the cluster overprovisioner by following the steps:

1. Run the command to get the compute-reserved-node deployment spec.

```
kubectl get deployment compute-reserved-node -n cluster -o yaml
```

2. Open the YAML and locate the spec > affinity > nodeAffinity.

```
spec:
  affinity:
    nodeAffinity:
      requiredDuringSchedulingIgnoredDuringExecution:
        nodeSelectorTerms:
          - matchExpressions:
              - key: failure-domain.beta.kubernetes.io/zone
                operator: In
                values:
                  - us-west-2a
              - key: compute
                operator: In
                values:
                  - "true"
```

In this example, you see us-west-2a is the availability zone being used.

## Changing the Availability Zone used by the overprovisioner

Generally, you do not change the Availability Zone used by the overprovisioner, but it is possible to do so. For example, if the cloud provider runs out of nodes in an Availability Zone, or for some other reason, you do not want to configure the Virtual Warehouse in the default Availability Zone of the overprovisioner, change the Availability Zone used by the overprovisioner.

In this task, you edit the compute-reserved-node deployment script to change the value for the key [failure-domain.beta.kubernetes.io/zone](#) to the Availability Zone you want, and save the changes.

*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.*

1. Edit the deployment script.

```
kubectl edit deployment compute-reserved-node -n cluster
```

2. Update the value for [failure-domain.beta.kubernetes.io/zone](https://failure-domain.beta.kubernetes.io/zone) with the AZ you want, such as *us-west-2b* or *us-west-2c*, associated with your CDW env.
3. Save the changes to the deployment spec.

## Creating the Virtual Warehouse

1. Create a Virtual Warehouse as described in [Adding a new Virtual Warehouse](#), selecting a Hive or Impala SQL engine type.
2. Configure the Virtual Warehouse properties: Name, Type, Database Catalog.
3. Select the Availability Zone that the cluster overprovisioner uses, as described above. Alternatively, change the Availability Zone of the cluster overprovisioner, as described above, and select that Availability Zone for the Virtual Warehouse.

### New Virtual Warehouse

Name \*

Type \*

HIVE
IMPALA

Database Catalog \*

Availability Zone ⓘ

✓ -- select an option --
us-west-2b (subnet-05bbb3a6c02d0a76a)
us-west-2a (subnet-0ac6a804221711fc2)
us-west-2c (subnet-0ee2ba7b5bb2fcc29)

User Groups ⓘ

4. Click Create.

*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.*