

Cloudbreak deployment options 2

## Cloudbreak Deployment Options

**Date of Publish:** 2019-05-28



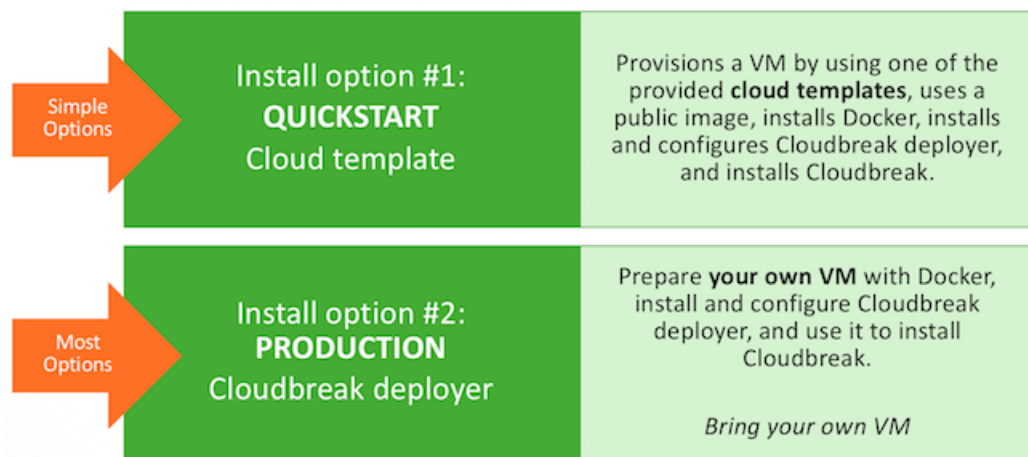
<https://docs.hortonworks.com/>

# Contents

|   |          |
|---|----------|
| <b>Cloudbreak deployment options.....</b>               | <b>3</b> |
| Deployment option cheatsheet.....                       | 3        |
| Quickstart option for AWS.....                          | 4        |
| Quickstart option for Azure.....                        | 4        |
| Quickstart option for GCP.....                          | 5        |
| Production deployment option.....                       | 5        |
| <br>  |          |
| <b>Cluster deployment options.....</b>                  | <b>5</b> |
| Default cluster configurations.....                     | 6        |
| <br>  |          |
| <b>Deploying on AWS GovCloud.....</b>                   | <b>7</b> |
| <br>  |          |
| <b>Using Cloudbreak with special Azure regions.....</b> | <b>8</b> |

## Cloudbreak deployment options

In general, Cloudbreak offers a quickstart option, as well as a production deployment option:



- The quickstart option allows you to get started with Cloudbreak quickly, but offers limited flexibility. Use this option for getting started with Cloudbreak. This option is not suitable a production.
- The production option is less automated, but offers more configurability. This option is recommended when your production environments.

Refer to the following documentation to learn more about Cloudbreak deployment options available for your cloud platform:

### Deployment option cheatsheet

The information provided here summarizes Cloudbreak deployment options.

The following table summarizes the available Cloudbreak deployment options:

|                              | Quickstart  | Production  |
|------------------------------|---|---|
| Operating system             | <ul style="list-style-type: none"> <li>• AWS: Amazon Linux</li> <li>• Azure: CentOS 7</li> <li>• GCP: CentOS 7</li> </ul>               | Bring your own RHEL/CentOS/OEL 7  |
| Networking                   | <ul style="list-style-type: none"> <li>• AWS: New VPC</li> <li>• Azure: New or existing vNet</li> <li>• GCP: New VPC network</li> </ul> | Custom  |
| External Cloudbreak database | Embedded PostgreSQL   | Embedded PostgreSQL is available, but for production you should use an external PostgreSQL. |

The following operating systems are used when launching by using the quickstart option:

|           | Quickstart                                 | Operating system (public image) |
|-----------|--|---------------------------------|
| AWS       | CloudFormation (CFN) template              | Amazon Linux                    |
| Azure     | Azure Resource Manager (ARM) template      | CentOS 7                        |
| GCP       | Cloud Deployment Manager, Google Cloud SDK | CentOS 7                        |
| OpenStack | NOT AVAILABLE                              |                                 |

## Quickstart option for AWS

The quickstart option on AWS allows you to instantiate Cloudbreak quickly from the CloudFormation template.



**Note:**

The quickstart option cannot be used with AWS GovCloud. You must use the production deployment option.

This option utilizes the following AWS services and provisions the following resources:

| Resource                           | Description  | How it is used by Cloudbreak   |
|------------------------------------|--|--|
| <a href="#">AWS CloudFormation</a> | AWS CloudFormation is used to create and manage a collection of related AWS resources.   | Cloudbreak is launched by using a CloudFormation template.   |
| <a href="#">Amazon EC2</a>         | Amazon EC2 is used to launch a virtual machine for Cloudbreak. Security groups are used to control the inbound and outbound traffic to and from the Cloudbreak instance.   | Cloudbreak automatically provisions a new VM that runs Amazon Linux, installs Docker, and launches Cloudbreak. |
| <a href="#">Amazon VPC</a>         | Amazon VPC is used to provision your own dedicated virtual network and launch resources into that network. As part of VPC infrastructure, an internet gateway and a route table are provisioned: An internet gateway is used to enable outbound access to the internet from the control plane and the clusters, and a route table is used to connect the subnet to the internet gateway. | Cloudbreak provisions a new VPC and subnet, and launches the Cloudbreak VM within it.                          |
| <a href="#">AWS IAM</a>            | AWS Identity & Access Management (IAM) is used to control access to AWS services and resources.  | Cloudbreak provisions the CloudbreakQuickstartRole IAM role that is used during the quickstart deployment.     |
| <a href="#">AWS Lambda</a>         | This is a utility service for running code in AWS.   | Cloudbreak uses AWS Lambda is for running code when deploying Cloudbreak.                                      |

### Related Information

[Quickstart on AWS](#)

## Quickstart option for Azure

The quickstart option on Azure allows you to instantiate Cloudbreak quickly from Azure Resource Manager (ARM) template.

On Azure, resources are organized by using resource groups. When you launch Cloudbreak, you may either select to use an existing resource group or a new resource group is created. The following Azure resources are provisioned within the selected resource group:

- [Virtual network](#) (VNet) securely connects Azure resources to each other. You may either launch Cloudbreak into an existing VPC, or a new VPC is created and added to the resource group.
- [Network security group](#) (NSG) defines inbound and outbound security rules, which control network traffic flow.
- [Virtual machine](#) runs Cloudbreak. Based on the ARM template, Azure automatically provisions a new VM that runs CentOS 7, installs Docker, and launches Cloudbreak.
- [Public IP address](#) is assigned to your VM so that it can communicate with other Azure resources.
- [Network interface](#) (NIC) attached to the VM provides the interconnection between the VM and the underlying software network.
- [Blob storage container](#) is created to store Cloudbreak deployer OS disk's data.

### Related Information

[Quickstart on Azure](#)

## Quickstart option for GCP

The quickstart option on Google Cloud Platform allows you to instantiate Cloudbreak quickly from the Cloud Deployment Manager template.

Based on the Cloud Deployment Manager template, GCP automatically provisions a new VM that runs CentOS 7, installs Docker, and launches Cloudbreak. The following resources are provisioned on you GCP account:

- Network
- Subnetwork (subnet)
- Route (routing table)
- Firewall
- VM instance
- Config
- Config waiter

The config and config waiter are scripts used for deploying Cloudbreak. The startup config script is watched by the waiter script and updated when the deployment is up and running or if it failed to start. The startup waiter script keeps the deployment “in-progress” until the startup script that’s running on the created machine update the cbd-deployment-startup-config value to “success” or “failed”.

### Related Information

[Quickstart on GCP](#)

## Production deployment option

The production deployment option allows you to install Cloudbreak deployer manually on your own VM.

This option is available for all cloud providers supported by Cloudbreak.

This option:

- Uses your own VM with CentOS 7, RHEL 7, or Oracle Linux 7
- Allows you to use your custom virtual network (On Azure this is also possible with the quickstart option)
- Requires you to install Docker
- Requires you to download the cbd tarball, extract it, and configure Cloudbreak deployer
- Requires you to configure an external Cloudbreak database
- Allows you to configure advanced options such as LDAP. To review available advanced configurations, refer to Cloudbreak configuration documentation.

### Related Information

[Launch Cloudbreak on AWS](#)

[Launch Cloudbreak on Azure](#)

[Launch Cloudbreak on GCP](#)

[Launch Cloudbreak on OpenStack](#)

[Configuring Cloudbreak](#)

## Cluster deployment options

On a basic level, Cloudbreak offers three cluster deployment options:

- Basic cluster deployment with prescriptive options
- Advanced cluster deployment with customized options
- Enterprise HDP cluster deployment with a data lake and attached workload clusters:



**Note:**

The data lake deployment option is technical preview.

|                                     | Basic ephemeral | Advanced ephemeral      | Data lake                      |
|-------------------------------------|-----------------|-------------------------|--------------------------------|
| Tuned and optimized infrastructure  | ✓               | ✓                       | ✓                              |
| Simplified and automated operations | ✓               | ✓                       | ✓                              |
| Cloud storage integration           | ✓               | ✓                       | ✓                              |
| Gateway                             | ✓               | ✓                       | ✓                              |
| Schema                              | -               | Shared (Hive metastore) | Shared (Hive metastore)        |
| Authentication                      | Single-user     | Single-user             | Single or multi-user (LDAP/AD) |
| Authorization                       | -               | -                       | Security policies (Ranger)     |
| Cloud storage audit                 | -               | -                       | Audit (Ranger)                 |

To learn about advanced cluster options, refer to the documentation linked below:

**Related Information**

[Creating a cluster on AWS](#)

[Creating a cluster on Azure](#)

[Creating a cluster on GCP](#)

[Creating a cluster on OpenStack](#)

[Advanced cluster options on Azure](#)

[Advanced cluster options on GCP](#)

[Advanced cluster options on OpenStack](#)

[Working with data lakes](#)

**Default cluster configurations**

Cloudbreak includes default cluster configurations (in the form of blueprints) and supports using your own custom cluster configurations (in the form of custom blueprints).

The following default cluster configurations are available:

**HDP 3.1**

| Cluster type    | Main services         | Description  | List of all services  |
|-----------------|-----------------------|--|---|
| Data Science    | Spark 2, Zeppelin     | Useful for data science with Spark 2 and Zeppelin. | HDFS, YARN, MapReduce2, Tez, Hive, Pig, Sqoop, ZooKeeper, Ambari Metrics, Spark 2, Zeppelin |
| EDW - Analytics | Hive 2 LLAP, Zeppelin | Useful for EDW analytics using Hive LLAP.          | HDFS, YARN, MapReduce2, Tez, Hive 2 LLAP, Druid, Pig, ZooKeeper, Ambari Metrics, Spark 2    |

### HDP 2.6

| Cluster type    | Main services         | Description   | List of all services  |
|-----------------|-----------------------|---|---|
| Data Science    | Spark 2, Zeppelin     | Useful for data science with Spark 2 and Zeppelin.    | HDFS, YARN, MapReduce2, Tez, Hive, Pig, Sqoop, ZooKeeper, Ambari Metrics, Spark 2, Zeppelin |
| EDW - Analytics | Hive 2 LLAP, Zeppelin | Useful for EDW analytics using Hive LLAP.             | HDFS, YARN, MapReduce2, Tez, Hive 2 LLAP, Druid, Pig, ZooKeeper, Ambari Metrics, Spark 2    |
| EDW - ETL       | Hive, Spark 2         | Useful for ETL data processing with Hive and Spark 2. | HDFS, YARN, MapReduce2, Tez, Hive, Pig, ZooKeeper, Ambari Metrics, Spark 2                  |

### HDF 3.3

| Cluster type         | Main services | Description                                 | List of all services                           |
|----------------------|---------------|---|--|
| Flow Management      | NiFi          | Useful for flow management with NiFi.       | NiFi, NiFi Registry, ZooKeeper, Ambari Metrics |
| Messaging Management | Kafka         | Useful for messaging management with Kafka. | Kafka, ZooKeeper, Ambari Metrics               |

## Deploying on AWS GovCloud

You can deploy Cloudbreak and Cloudbreak-managed clusters on public AWS regions and on the isolated [AWS GovCloud](#) region.



**Note:**

This feature is technical preview: It is not suitable for production.

Although Cloudbreak web UI and CLI treat AWS and AWS CloudGov as separate provider types (by enumerating them separately during the credential creation), the two are actually distinct endpoints of the same provider and there are almost no differences when using Cloudbreak with public AWS regions vs AWS CloudGov.

This is why a single documentation set is available for using Cloudbreak on AWS: That is, any Cloudbreak documentation marked as appropriate for AWS can be used for both public AWS and AWS GovCloud. The few restrictions that exist for AWS GovCloud are pointed out in the documentation related to the affected features.

### Restrictions of AWS GovCloud

When working with AWS GovCloud, it is not possible to use existing AMIs. This causes the following Cloudbreak features to be unavailable to use with AWS GovCloud:

| Category            | Description   |
|---------------------|---|
| Installation        | Cloudbreak can be deployed on AWS by using the quickstart option or the production option. However, only the production option can be used with AWS GovCloud. To install Cloudbreak on AWS GovCloud, refer to <a href="#">Installing on AWS</a> .   |
| Credential creation | Both AWS credential options (key-based and role-based) can be used with AWS GovCloud. However, if your Cloudbreak is running on a non-CloudGov AWS account, you cannot use the role-based credential option.<br><br>If using Cloudbreak CLI to create or modify a credential on AWS GovCloud, make sure to use the aws-gov command options, instead of the aws command options. |

| Category         | Description  |
|------------------|--|
| Cluster creation | By default, Cloudbreak includes a set of prewarmed and base images that can be used for deploying clusters. However, these prewarmed images cannot be used on AWS GovCloud; Instead, you must burn your own custom images prior to creating clusters. To prepare custom images, refer to <a href="#">Using custom images</a> . |

**Related Information**

[Using custom images](#)

[Installing on AWS](#)

## Using Cloudbreak with special Azure regions

Azure includes a few [special Azure regions](#) that you may wish to use for compliance or legal purposes. Cloudbreak can be used with these regions, but the following restrictions apply.

**Restrictions when using special Azure regions**

| Category         | Description   |
|------------------|---|
| Installation     | Cloudbreak can be deployed in special Azure region by using the production option or, if your subscription has access to it, the quickstart option.   |
| Cluster creation | Cloudbreak includes a set of prewarmed and base images that can be used for deploying clusters in all standard Azure regions. However, such images are not created for the special Azure regions, and so in order to launch a cluster in one of these special regions, it is a prerequisite to first burn an image in that region. To prepare custom images, refer to <a href="#">Using custom images</a> . |

**Related Information**

[Special Azure regions \(Azure\)](#)

[Using custom images](#)