

## Introduction to Cloudbreak

**Date of Publish:** 2018-09-14



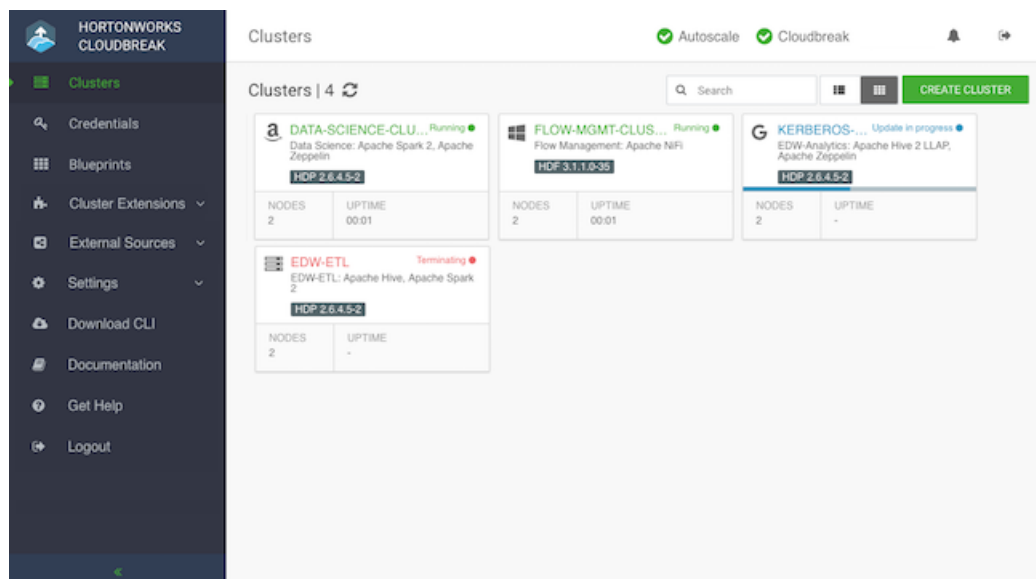
# Contents

- What is Cloudbreak..... 3**
- Primary use cases..... 3**
- Architecture.....3**
  - Cloudbreak deployer architecture.....3
  - Cloudbreak application architecture.....4
- Core concepts..... 5**

## What is Cloudbreak

Cloudbreak simplifies the provisioning, management, and monitoring of on-demand HDP and HDF clusters in virtual and cloud environments. It leverages cloud infrastructure to create host instances, and uses Apache Ambari via Ambari blueprints to provision and manage Hortonworks clusters.

Cloudbreak allows you to create clusters by using the Cloudbreak web UI, Cloudbreak CLI, and Cloudbreak REST API. Clusters can be launched on public cloud infrastructure platforms Amazon Web Services (AWS) including AWS GovCloud, Microsoft Azure, and Google Cloud Platform (GCP), and on the private cloud infrastructure platform OpenStack.



## Primary use cases

Cloudbreak allows you to create, manage, and monitor your HDP and HDF clusters on your chosen cloud platform:

- Dynamically deploy, configure, and manage clusters on public and private clouds (AWS, Azure, Google Cloud, OpenStack).
- Use automated scaling to seamlessly manage elasticity requirements as cluster workloads change.
- Secure your cluster by enabling Kerberos.

## Architecture

Cloudbreak deployer installs Cloudbreak components on a VM. Once these components are deployed, you can use Cloudbreak application or Cloudbreak CLI to create, manage, and monitor clusters.

### Cloudbreak deployer architecture

Cloudbreak deployer installs Cloudbreak components on a VM.

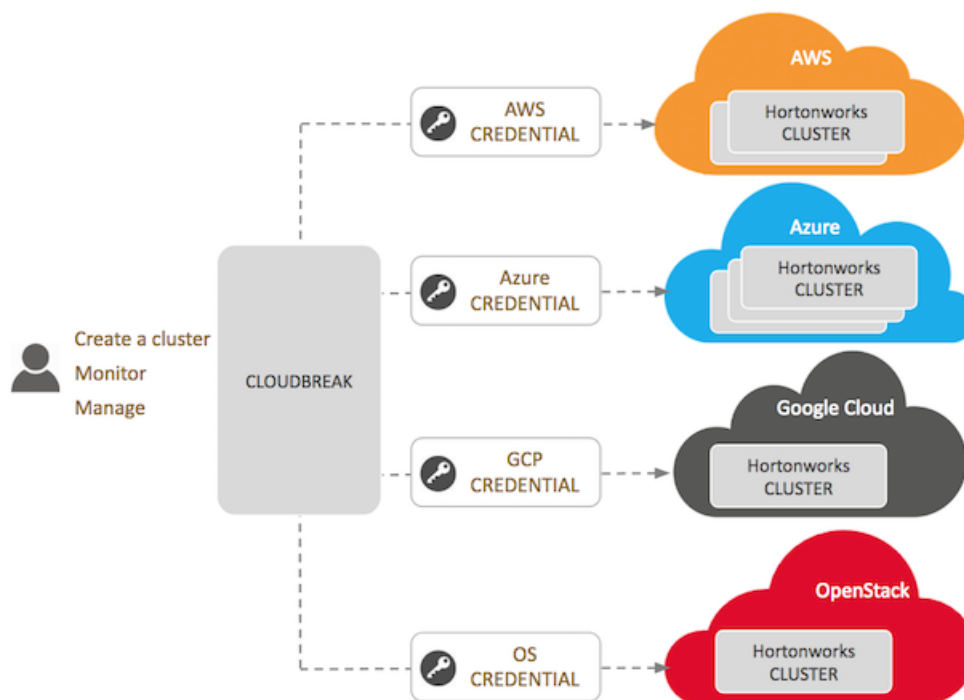
Cloudbreak deployer includes the following components:

Component	Description
Cloudbreak Application	Cloudbreak application is built on the foundation of cloud provider APIs and Apache Ambari.
Uluwatu	This is Cloudbreak web UI, which can be used to create, manage, and monitor clusters.
Cloudbreak CLI	This is Cloudbreak's command line tool, which can be used to create, manage, and monitor clusters.
Identity	This is Cloudbreak's OAuth identity server implementation, which utilizes UAA.
Sultans	This is Cloudbreak's user management system.
Periscope	This is Cloudbreak's autoscaling application, which is responsible for automatically increasing or decreasing the capacity of the cluster when your pre-defined conditions are met.

## Cloudbreak application architecture

Cloudbreak application is a web application which simplifies cluster provisioning in the cloud.

Based on your input, Cloudbreak application provisions all required cloud infrastructure and then provisions a cluster on your behalf within your cloud provider account.



Cloudbreak application is built on the foundation of cloud provider APIs and Apache Ambari:

- Cloudbreak uses cloud provider APIs to communicate with the cloud providers.
- Cloudbreak uses the Cloudbreak credential to authenticate with your cloud provider account and provision cloud resources required for the clusters.
- Cloudbreak uses Apache Ambari and Ambari blueprints to provision, manage, and monitor clusters. Ambari blueprints are a declarative definition of a cluster. With a blueprint, you can specify stack, component layout, and configurations to materialize a cluster instance via Ambari REST API, without having to use the Ambari cluster install wizard.

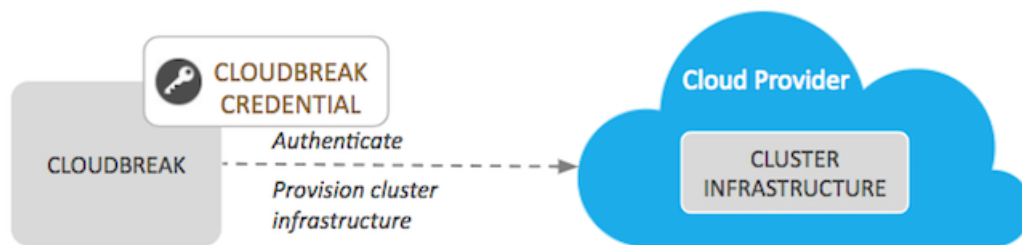
## Core concepts

Cloudbreak includes the following core concepts: credentials, clusters, blueprints, cluster extensions, and external sources.

### Credentials

Cloudbreak credential allows Cloudbreak to authenticate with the cloud provider and create resources on your behalf.

The authentication process varies depending on the cloud provider, but is typically done via assigning a specific IAM role to Cloudbreak which allows Cloudbreak to perform certain actions within your cloud provider account.



After launching Cloudbreak, you are required to create a Cloudbreak credential for each cloud provider on which you would like to provision clusters. Only after you have completed that step, you can start creating clusters.

### Clusters

Cloudbreak allows you to create two types of clusters:

- Workload clusters: These HDP or HDF clusters can be created as needed and then used for running specific workloads.
- Data lake clusters: A data lake provides a way for you to centrally apply and enforce authentication, authorization, and audit policies across multiple ephemeral workload clusters. "Attaching" your workload cluster to the data lake instance allows the attached cluster workloads to access data and run in the security context provided by the data lake.

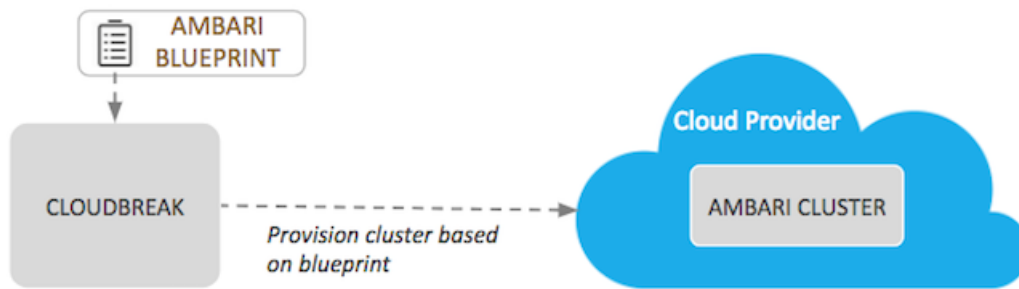
**Note:**

The data lake functionality is technical preview.

### Blueprints

Ambari blueprints are a declarative definition of a cluster. A blueprint allows you to specify stack, component layout, and configurations to materialize a cluster instance via Ambari REST API, without having to use the Ambari cluster install wizard. Ambari blueprints are specified in JSON format. After you provide the blueprint to Cloudbreak, the host groups in the JSON are mapped to a set of instances when starting the cluster, and the specified services and components are installed on the corresponding nodes.

Cloudbreak includes a few default blueprints and allows you to upload your own blueprints.



### Cluster extensions

Cloudbreak allows you to utilize the following cluster extensions:

- **Recipes:** Cloudbreak allows you to upload custom scripts, called "recipes". A recipe is a script that runs on all nodes of a selected node group at a specific time. You can use recipes for tasks such as installing additional software or performing advanced cluster configuration. For example, you can use a recipe to put a JAR file on the Hadoop classpath.
- **Management packs:** Management packs allow you to deploy a range of services to your Ambari-managed cluster. You can use a management pack to deploy a specific component or service, such as HDP Search, or to deploy an entire platform, such as HDF.

### External sources

Cloudbreak allows you to define external sources that are created independently of a cluster – and therefore their lifespan is not limited by the lifespan of any cluster – and that can be reused with multiple clusters. The external sources that can be registered in Cloudbreak include:

- Authentication configurations (LDAP/AD)
- Database configurations
- Image catalogs
- Proxy configurations

### Workspaces

When a Cloudbreak instance is configured with an LDAP or Active Directory, the LDAP/AD users can share Cloudbreak resources with other users via workspaces. Any user can create a workspace and invite other users to that workspace, granting them specific access rights. Once a workspace is created, users who are part of it share all resources, such as clusters, blueprints, recipes, and so on, created as part of that workspace.