

Intro to Cloudbreak 2

Introduction to Cloudbreak

Date of Publish: 2019-05-28



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

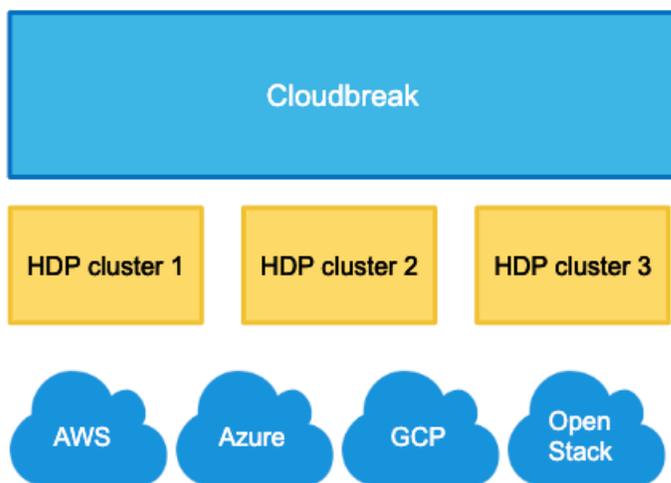
Contents

What is Cloudbreak.....	3
Primary use cases.....	3
Interfaces.....	3
Core concepts.....	4
Architecture.....	7
Cloudbreak deployer architecture.....	7
Cloudbreak application architecture.....	8

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 web UI, CLI, and 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

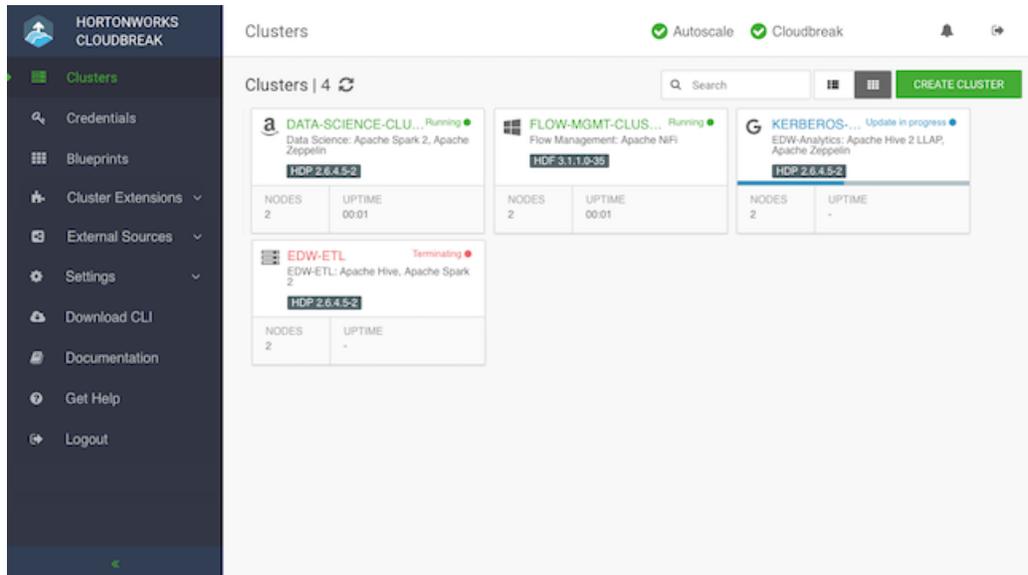
Cloudbreak allows you to create, manage, and monitor your HDP and HDF clusters on your chosen cloud platform:

- 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.

Interfaces

Cloudbreak includes a web user interface (web UI), command line interface (CLI), and REST APIs which allow you to perform administrative tasks such as registering your cloud provider credentials, registering external sources, and provisioning and managing clusters.

The Cloudbreak web UI can be accessed via a browser. The Cloudbreak CLI can be downloaded from Cloudbreak web UI and installed on your computer.



Core concepts

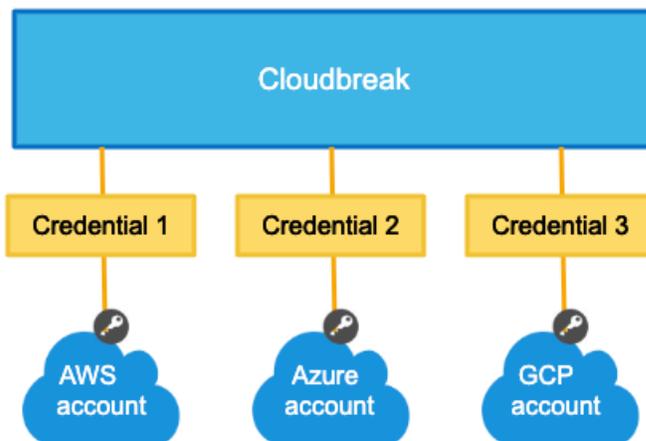
Cloudbreak includes the following core concepts: credentials, clusters, blueprints, cluster extensions, and external sources.

Credentials

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.

Cloudbreak credential allows Cloudbreak to authenticate with the cloud provider and obtain authorization to provision cloud provider resources on your behalf.

The authentication and authorization 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.



Clusters

Cloudbreak allows you to create two types of clusters: workload clusters (referred to as "workload clusters", "ephemeral clusters", or "clusters") and data lake clusters (referred to as "data lakes").

Workload clusters: Multiple HDP or HDF clusters can be provisioned as needed and then used for storing and processing data. These clusters are typically ephemeral. By default, Cloudbreak includes a few default HDP cluster blueprints that allow you to provision clusters including three node types: master, worker, and compute.

- A master node runs the components for managing the cluster (including Ambari), storing data, processing tasks, as well as other master components.
- A worker node runs the components that are used for executing processing tasks and handling storing data in HDFS.
- A compute node can optionally be used for running data processing tasks.

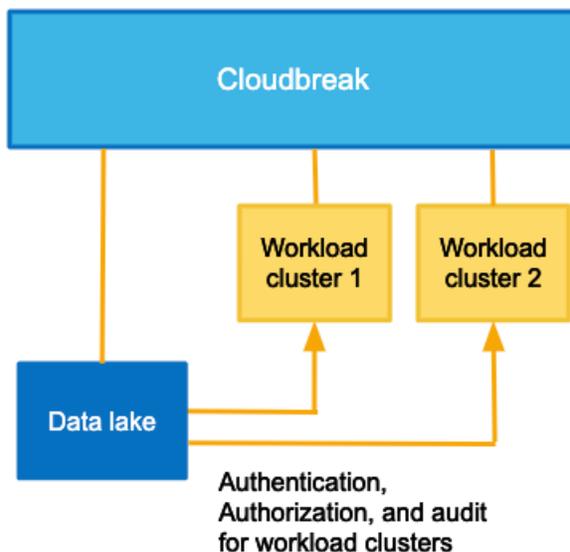
**Note:**

Default HDF cluster blueprints do not follow this pattern. Furthermore, you can upload your own HDP and HDF workload cluster blueprints with custom node types.

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:**

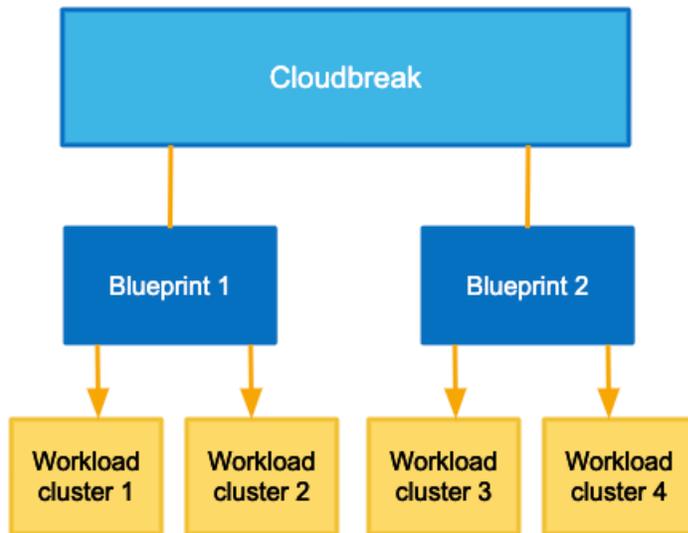
The data lake functionality is technical preview.



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

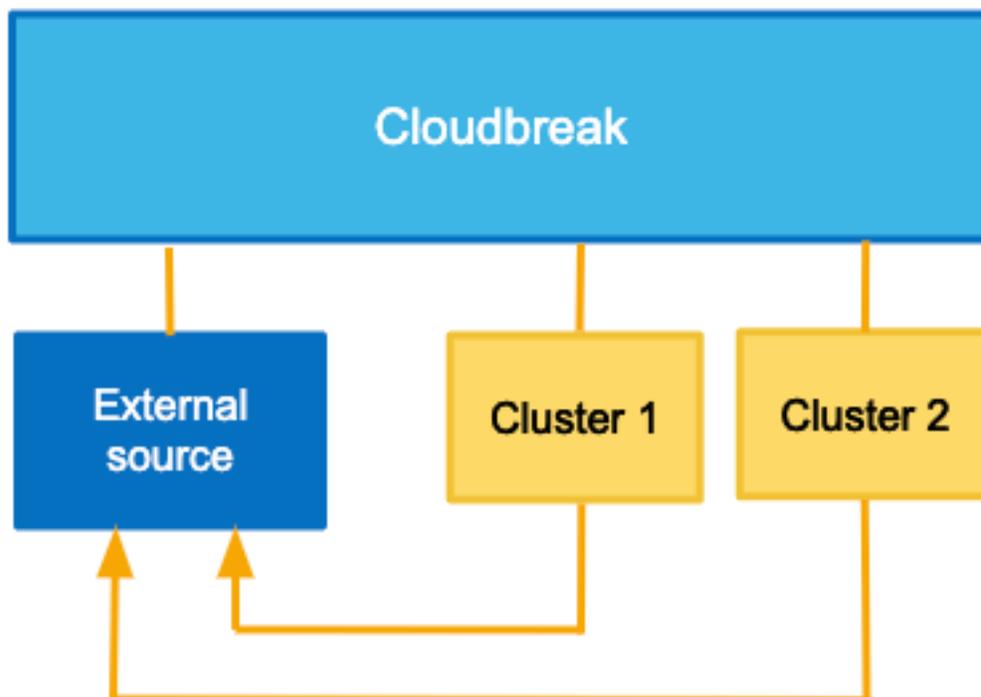
Cloudbreak allows you to utilize the following cluster extensions:

- **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:

External source	Description
Authentication configurations (LDAP/AD)	Cloudbreak allows you to register an existing LDAP/AD instance as an external source and use it for multiple clusters.
Database configurations	Cloudbreak allows you to register existing RDBMS instances as external sources to be used as databases for specific cluster components (such as Ambari, Ranger, Hive, and so on), and then reuse them for multiple clusters.
Image catalogs	Cloudbreak provides a default image catalog with default prewarmed and base images, and allows you to create custom image catalogs with custom base images. While base images include only OS and tools, prewarmed images also include Ambari and stack.
Proxy configurations	Cloudbreak allows you to register an existing proxy to be used for your clusters.



Workspaces

When 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.



Note:

It is not possible to share resources between workspaces; That is, every resource is attached to the workspace in which it was created and it cannot be shared with or moved to another workspace.

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.

To learn more, refer to the following documentation:

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.

Component	Description
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 the Cloudbreak credential to authenticate with your cloud provider account and provision cloud resources required for the clusters.
- Cloudbreak uses cloud provider APIs to communicate with the cloud providers.
- 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.