

Planning Your Flow Management Deployment

Date published: 2019-12-16

Date modified: 2021-06-14



Legal Notice

© Cloudera Inc. 2024. 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.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

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.

Contents

| | |
|---|----------|
| Deployment scenarios..... | 4 |
| Flow Management cluster definitions..... | 4 |
| Flow Management cluster layout..... | 4 |

Deployment scenarios

Before you get started with a Cloudera Data Flow deployment with Flow Management tools, it is useful to understand which software version is right for your platform and operational objectives. This table helps you to understand the documentation you need to review to get started.

| If you want to install NiFi and NiFi Registry with ... | Review this documentation |
|--|---|
| CDP Public Cloud – Flow Management Clusters | Cloudera Data Flow for Data Hub documentation |
| CDP Private Cloud Base | CFM 2.0.x documentation |
| HDF | HDF documentation |
| CDH | CFM 1.1.0 documentation |

Flow Management cluster definitions

There are four Flow Management cluster definitions available to deploy Apache NiFi and Apache NiFi Registry in CDP Public Cloud. Before you select your cluster definition, it is useful to know your cloud provider and operational objective.

The following cluster definitions are available for Flow Management in CDP Public Cloud:

- Flow Management Light Duty for AWS
- Flow Management Light Duty for Azure
- Flow Management Light Duty for GCP
- Flow Management Heavy Duty for AWS
- Flow Management Heavy Duty for Azure
- Flow Management Heavy Duty for GCP

These cluster definitions support installing Flow Management clusters running Apache NiFi and Apache NiFi Registry.

Flow Management delivers high-scale data ingestion, transformation, and management to enterprises from any-to-any environment. It addresses key enterprise use cases such as data movement, continuous data ingestion, log data ingestion, and acquisition of all types of streaming data including social, mobile, clickstream, and IoT data.

Flow Management templates include a no-code data ingestion and management solution powered by Apache NiFi. With NiFi's intuitive graphical interface and 300+ processors, Flow Management enables easy data ingestion and movement between CDP services, as well as 3rd party cloud services. NiFi Registry is automatically set up and provides a central place to manage versioned data flows.

Flow Management cluster layout

Describes the layout and capacity of the Flow Management light and heavy duty cluster definitions.

Flow Management: Light Duty cluster layout

You can use a Flow Management: Light Duty cluster definition in development, testing, or proof of concept scenarios.

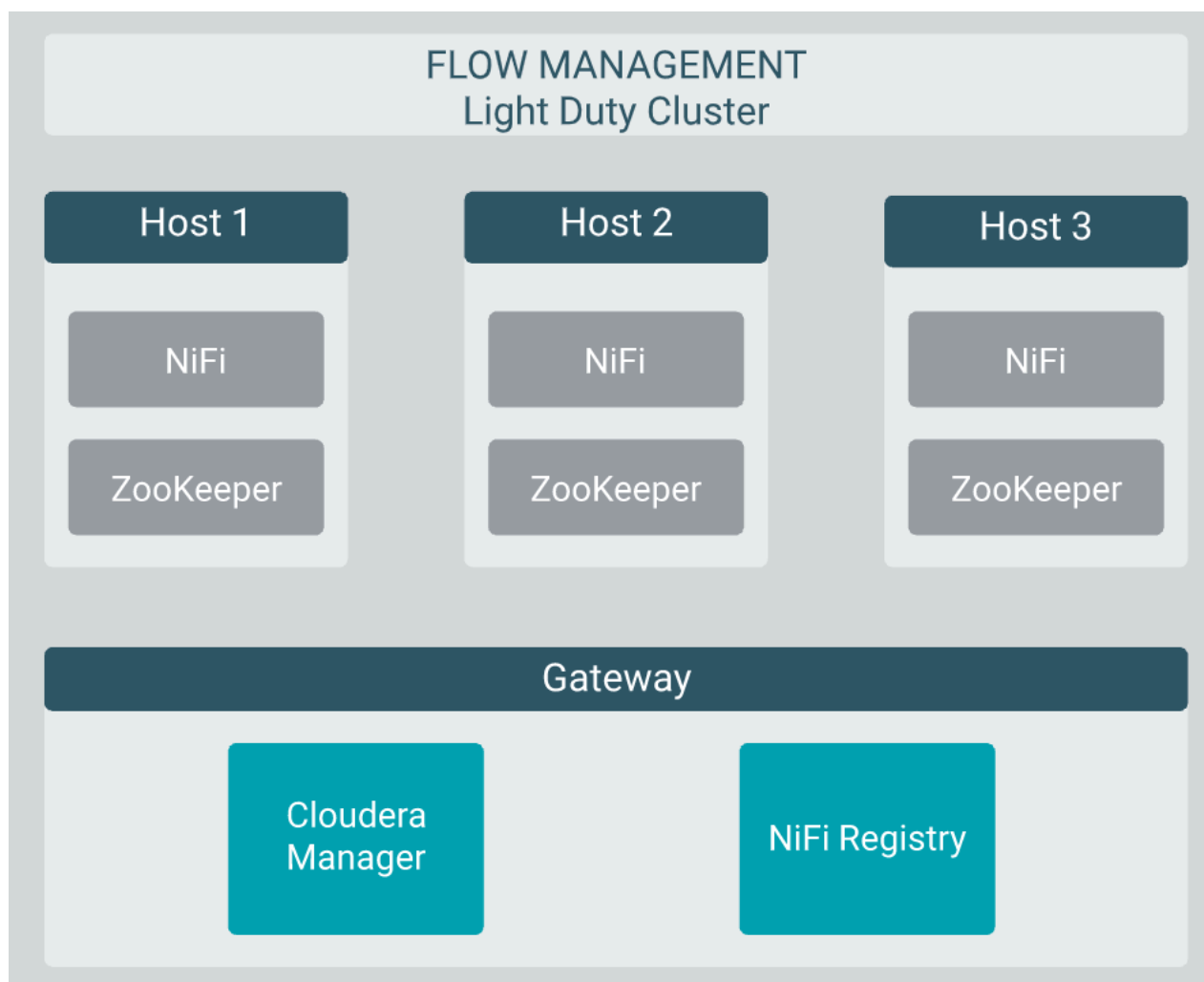


Note: When you create a Flow Management light duty cluster, a non HA external RDS instance is provisioned and used by the NiFi Registry.

Each of the cluster nodes includes:

- NiFi and ZooKeeper co-located on all instances
- For each node hosting NiFi and ZooKeeper:
 - AWS: m5.2xlarge
 - Azure: D8_v3
 - GCE: e2-standard-8
- Storage requirements per NiFi node:
 - AWS: 4x 500GB EBS ST1
 - Azure 4x 500GB Standard SSD
 - GCE: 4x 500GB PD-Standard
- Each NiFi node hosts the following repositories
 - FlowFile repository
 - Content repository
 - Provenance repository
 - Log and Database repository

For more information, see your *cloud provider-specific information about instance types and storage information*.



Flow Management: Heavy Duty cluster layout

You can use the Flow Management: Heavy Duty cluster definition in production scenarios.

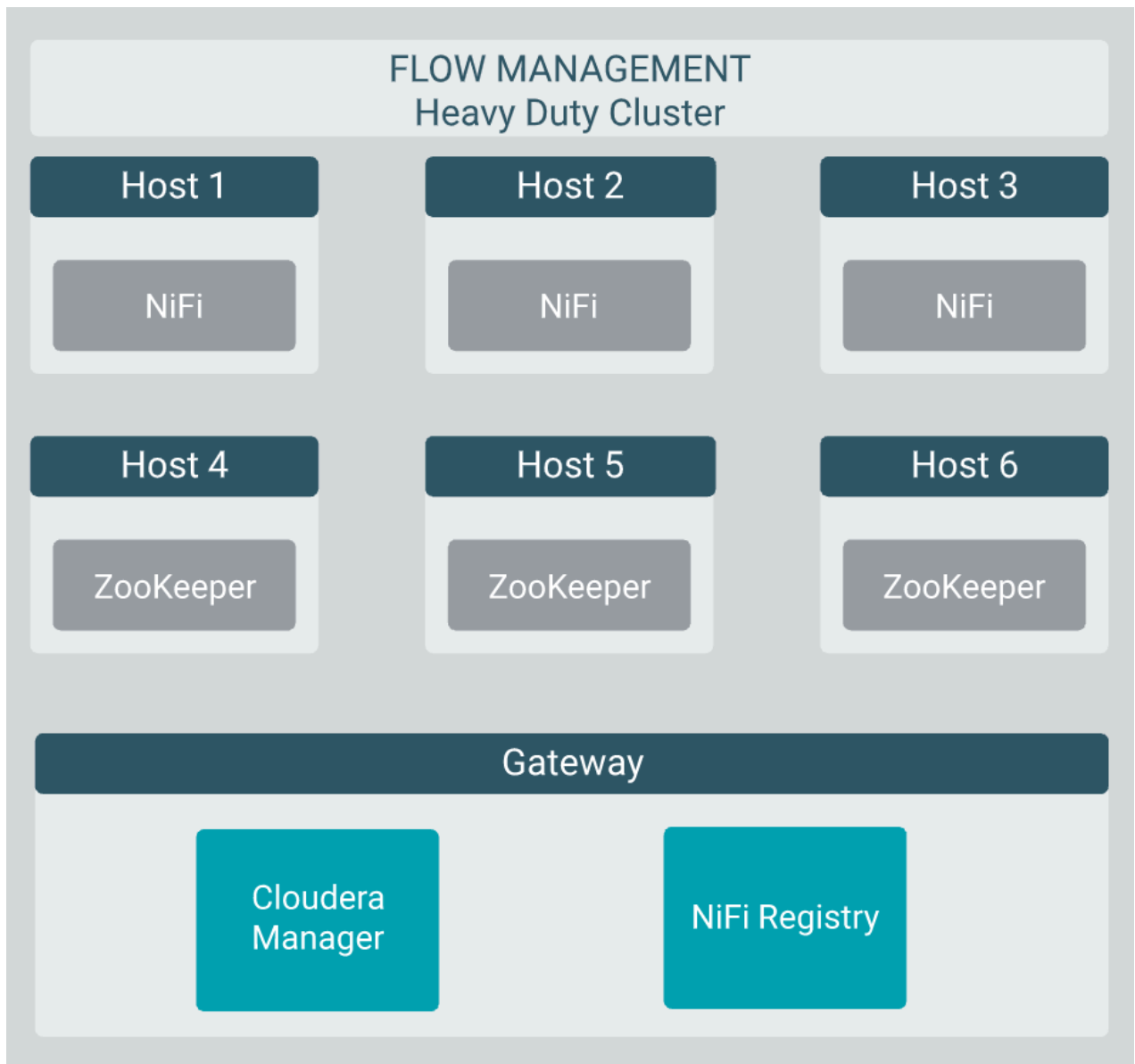


Note: When you create a Flow Management heavy duty cluster, an HA external RDS instance is provisioned and used by the NiFi Registry.

The cluster definition includes:

- NiFi and ZooKeeper on separate nodes
- NiFi nodes scale independently of ZooKeeper
- For each ZooKeeper node:
 - AWS – m5.2xlarge
 - Azure – D8_v3
 - GCE: e2-standard-8
- For each NiFi node:
 - AWS – m5.2xlarge
 - Azure – F16sv2
 - GCE: e2-standard-8
- Storage requirements per NiFi node:
 - AWS – 4x 1TB EBS GP2
 - Azure – 4x 1TB Premium SSD
 - GCE: 4x 1TB PD-SSD
- Each NiFi node hosts the following repositories
 - FlowFile repository
 - Content repository
 - Provenance repository
 - Log and Database repository

For more information, see your *cloud provider-specific information about instance types and storage information*.

**Related Information**[AWS instance types](#)[Azure instance types](#)[AWS storage information](#)[Azure storage information](#)