

Cloudera DataFlow for Data Hub 7.3.1

# Planning Your Flow Management Deployment

Date published: 2019-12-16

Date modified: 2024-12-10

The Cloudera logo is displayed in a bold, orange, sans-serif font. The word "CLOUDERA" is written in all caps, with the letter 'E' in the middle of "UDERA" having a unique design where the top bar is a horizontal line with a small gap in the center.

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

# 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

<b>Flow Management cluster definitions.....</b>	<b>4</b>
<b>Flow Management cluster layout.....</b>	<b>4</b>

## Flow Management cluster definitions

There are six Flow Management cluster definitions available for deploying Apache NiFi and Apache NiFi Registry in CDP Public Cloud. Before you select your cluster definition, it is crucial to consider your cloud provider and operational objectives.

Available cluster definitions:

- 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 facilitate the installation of Flow Management clusters that run Apache NiFi and Apache NiFi Registry.

With Flow Management templates, you can address crucial enterprise use cases, offering high-scale data ingestion, transformation, and management across diverse environments. It supports tasks such as data movement, continuous data ingestion, log data processing, and acquiring various streaming data types, including social, mobile, clickstream, and IoT data.

## Flow Management cluster layout

The Data Hub service provides two default Flow Management cluster definitions: Flow Management: Light Duty and Flow Management: Heavy Duty. Understanding the layout, capacity, and components of these definitions is essential for effective deployment.

### Flow Management: Light Duty cluster layout

The Flow Management: Light Duty cluster definition is suitable for development, testing, or proof of concept scenarios.



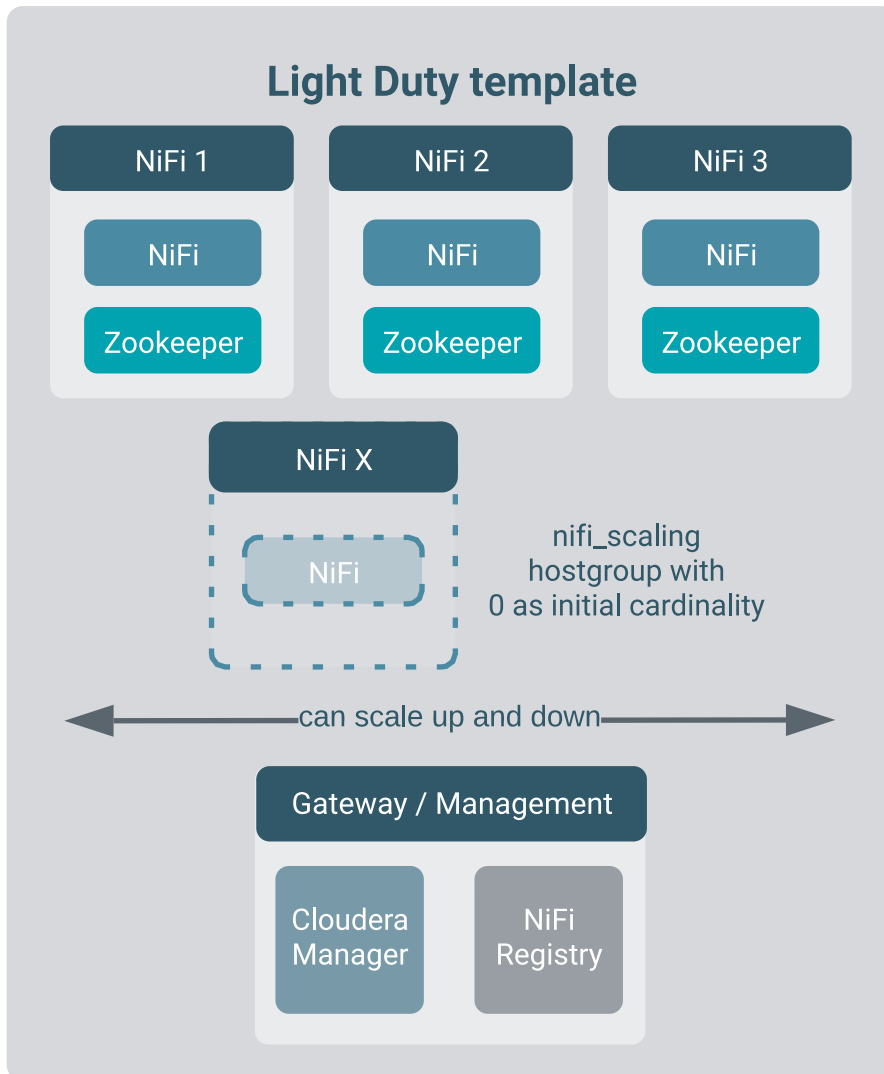
**Note:** When you create a Flow Management Light Duty cluster, a non-HA external RDS instance is provisioned and used by NiFi Registry and Schema Registry.

Each cluster node comprises the following components:

- NiFi and ZooKeeper are co-located on all instances.
- Specifications for nodes hosting NiFi and ZooKeeper:
  - AWS: m5.2xlarge
  - Azure: D8\_v3
  - GCE: e2-standard-8
- Storage requirements per NiFi node:
  - AWS: 4 x 500GB EBS ST1
  - Azure 4 x 500GB Standard SSD
  - GCE: 4 x 500GB PD-Standard

- Each NiFi node hosts:
  - FlowFile repository
  - Content repository
  - Provenance repository
  - Log and Database repository

For more information, see the *Instance types* and *Storage information* specific to your cloud provider.



### Flow Management: Heavy Duty cluster layout

The Flow Management: Heavy Duty cluster definition is intended for production scenarios.



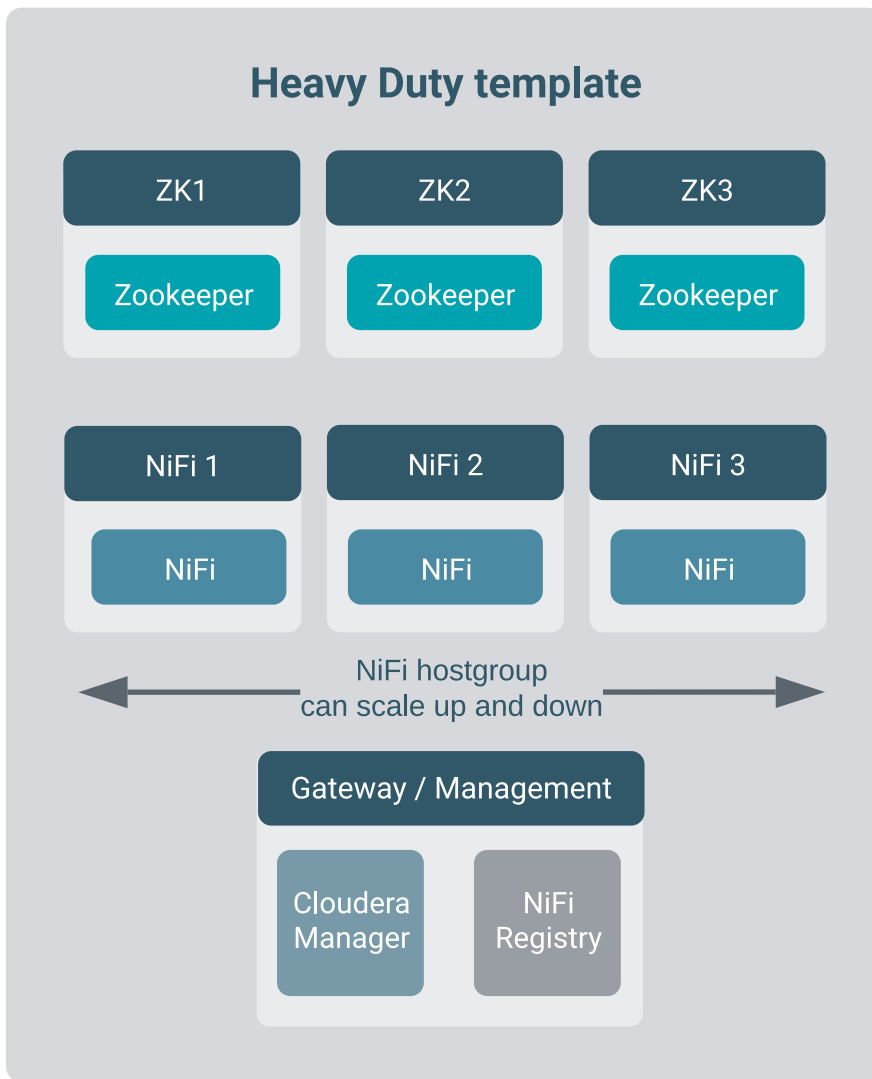
**Note:** When you create a Flow Management Heavy Duty cluster, an HA external RDS instance is provisioned and used by NiFi Registry and Schema Registry.

Each cluster node comprises the following components:

- NiFi and ZooKeeper run on separate nodes.
- NiFi nodes scale independently of ZooKeeper.

- Specifications for each ZooKeeper node:
  - AWS: m5.2xlarge
  - Azure: D8\_v3
  - GCE: e2-standard-8
- Specifications 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:
  - FlowFile repository
  - Content repository
  - Provenance repository
  - Log and Database repository

For more information, see the *Instance types* and *Storage information* specific to your cloud provider.



**Related Information**

[AWS instance types](#)

[Azure instance types](#)

[AWS storage information](#)

[Azure storage information](#)