Cloudera DataFlow for Data Hub 7.2.15

Planning Your Flow Management Deployment

Date published: 2019-12-16
Date modified: 2022-05-12

https://docs.cloudera.com/

Legal Notice

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.x documentation
HDF	HDF documentation
CDH	CFM 1.x 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

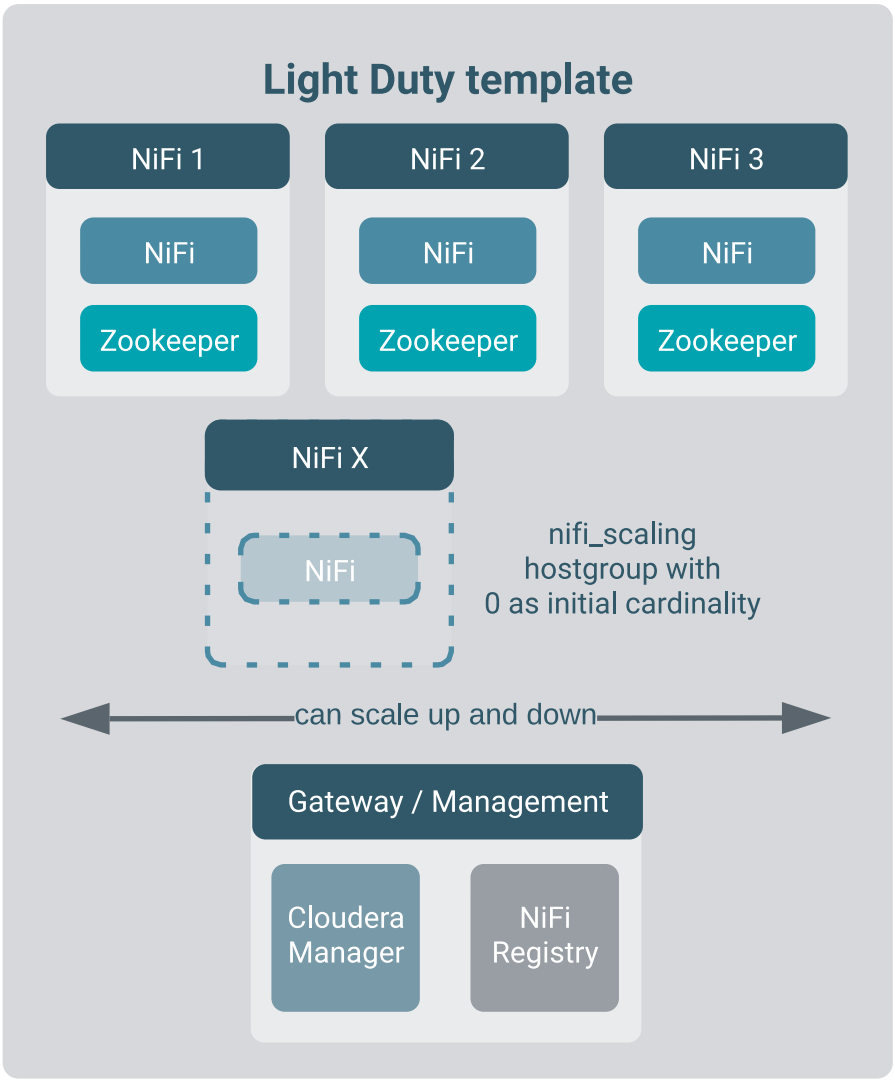
You can use a Flow Management: Light Duty cluster definition in development, testing, or proof of concept scenarios.

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