

Cloudera Data Flow for Data Hub 7.2.0

Planning Your Streams Messaging Deployment

Date published: 2020-05-01

Date modified: 2020-06-16

The Cloudera logo is displayed in a bold, orange, sans-serif font. The word "CLOUDERA" is written in all caps, with a stylized 'E' that has three horizontal bars.

<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

- Deployment scenarios..... 4**
- Data Hub cluster definitions..... 4**
- Streams Messaging cluster layout..... 5**

Deployment scenarios

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

If you want ...	Running with the following additional components	Review this documentation
CDK powered by Apache Kafka for CDH 5.x		CDK 4.1 Powered by Apache Kafka documentation
	<ul style="list-style-type: none"> • Schema Registry 	Cloudera Streams Processing 1.0.0 documentation
	<ul style="list-style-type: none"> • Schema Registry • Streams Messaging Manager • Streams Replication Manager 	Cloudera Streams Processing 2.0.x documentation
Kafka in CDH 6.x		CDH 6.x Kafka documentation
	<ul style="list-style-type: none"> • Schema Registry • Streams Messaging Manager • Streams Replication Manager 	Cloudera Streams Processing 2.0.x documentation
Kafka for HDF and HDP		HDF 3.4.x documentation
Kafka for CDP Private Cloud Base		Streaming libraries in Cloudera Runtime 7.0.3 CDP Private Cloud Base
Kafka for CDP in Data Hub	n/a	Streaming libraries in Cloudera Runtime 7.0.2
	<ul style="list-style-type: none"> • Schema Registry • Streams Messaging Manager 	Cloudera Data Flow for Data Hub documentation

Data Hub cluster definitions

The Streams Messaging templates include Kafka, Schema Registry, Streams Messaging Manager, and ZooKeeper. You may choose from the following template options, depending on your operational objectives:

- Streams Messaging Heavy Duty for AWS
- Streams Messaging Heavy Duty for Azure
- Streams Messaging Light Duty for AWS
- Streams Messaging Light Duty for Azure

Streams Messaging provides advanced messaging and real-time processing on streaming data using Apache Kafka, centralized schema management using Schema Registry, as well as management and monitoring capabilities powered by Streams Messaging Manager.

These templates set up fault-tolerant standalone deployments of Apache Kafka and supporting Cloudera components (Schema Registry and Streams Messaging Manager), which can be used for Kafka workloads in the cloud or as a disaster recovery instance for on-premises Kafka clusters.

**Note:**

Streams Messaging cluster definitions and templates available in CDP Public Cloud do not support the following Cloudera Runtime Streaming components and features.

- Streams Replication Manager
- Cruise Control
- Kafka Connect

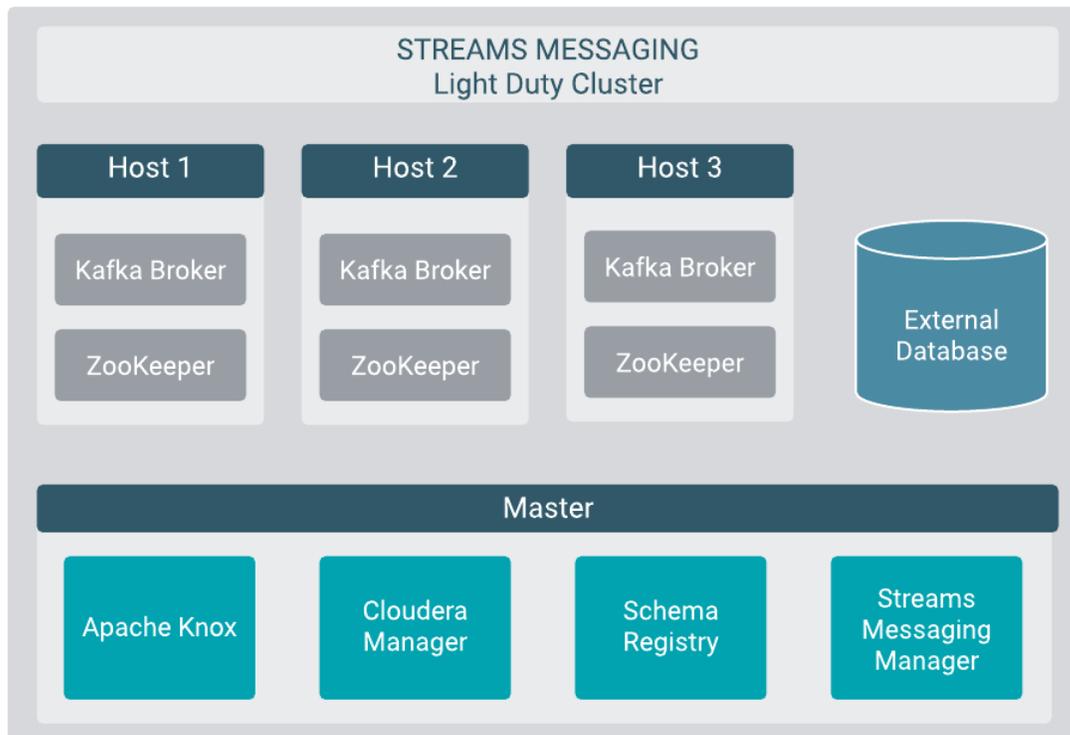
Streams Messaging cluster layout

Describes the layout and capacity of the Streams Messaging: Light Duty and Streams Messaging: Heavy Duty cluster definitions

Streams Messaging: Light Duty cluster layout

You can use a Streams Messaging: Light Duty cluster definition in development, testing, or proof of concept scenarios. Each of the cluster nodes includes:

- Kafka Broker and ZooKeeper co-located on all instances
- For each node hosting Kafka Broker and ZooKeeper instances:
 - AWS: m5.2xlarge
 - Azure: D8_v3
- Storage configuration per Kafka Broker node:
 - AWS: 1 TB Volume EBS ST1
 - Azure: 4x 500GB Standard SSD
- Master node instance
 - AWS: m5.2xlarge
 - Azure: Standard_D16_v3
- Storage configuration for Master node:
 - AWS: 100 GB Volume EBS Magnetic
 - Azure: 100 GB Standard Locally-redundant SSD storage



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

Streams Messaging: Heavy Duty cluster layout

You can use the Streams Messaging: Heavy Duty cluster definition in production scenarios. The cluster definition includes:

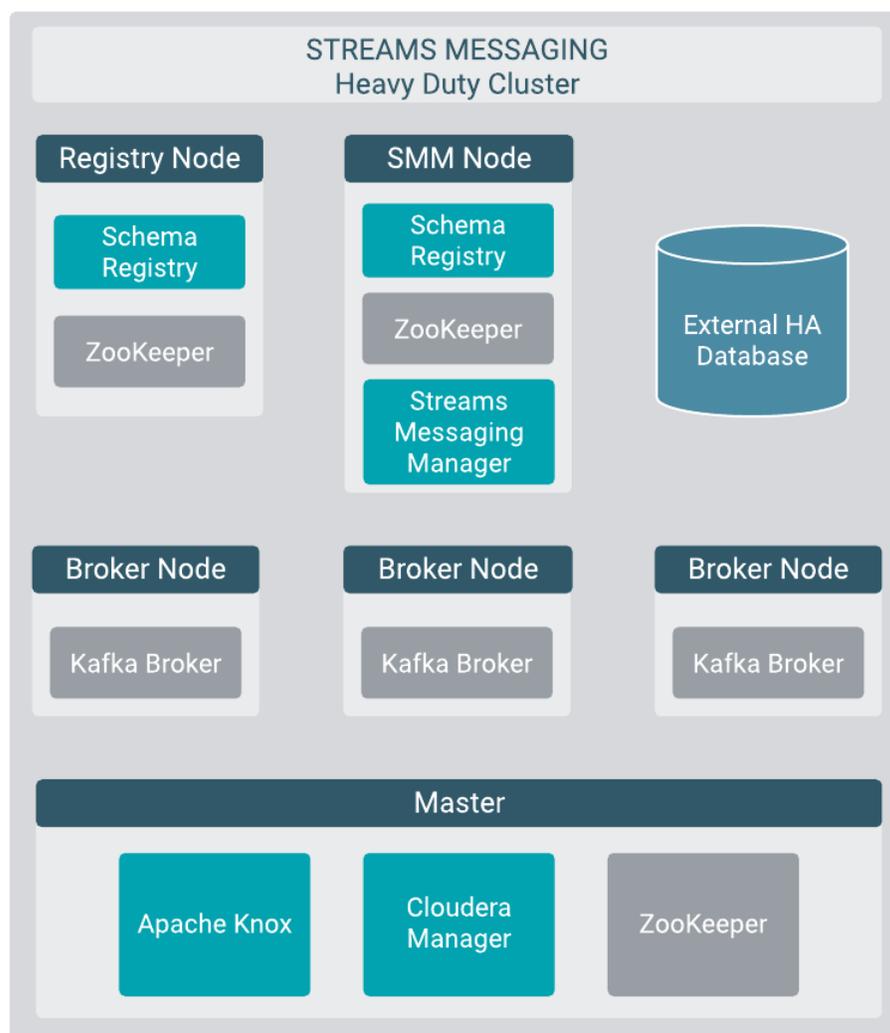
Azure

- Master Node – Containing Knox, Cloudera Manager, ZooKeeper
 - Instance type – Standard_D16_v3
 - Storage configuration – 100 GB Standard Locally-redundant SSD storage
- Registry Nodes – Containing Schema Registry, ZooKeeper
 - Instance type – Standard_D8_v3
 - Storage configuration – 100 GB Standard Locally-redundant SSD storage
- SMM Nodes – Containing SMM, Schema Registry, ZooKeeper
 - Instance type – Standard_D8_v3
 - Storage configuration – 100 GB Standard Locally-redundant SSD storage
- Broker Nodes – Containing a Kafka Broker
 - Instance type – Standard_D8_v3
 - Storage configuration – 2 x 500 GB Standard Locally-redundant SSD storage

AWS

- Master Node – Containing Knox, Cloudera Manager, ZooKeeper
 - Instance type – m5.2xlarge
 - Storage configuration – 100 GB EBS ST1

- Registry Nodes – Containing Schema Registry, ZooKeeper
 - Instance type – m5.2xlarge
 - Storage configuration – 100 GB EBS ST1
- SMM Nodes – Containing SMM, Schema Registry, ZooKeeper
 - Instance type – m5.2xlarge
 - Storage configuration – 100 GB EBS ST1
- Broker Nodes – Containing a Kafka Broker
 - Instance type – m5.2xlarge
 - Storage configuration – 1 TB EBS ST1



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](#)