

7.1.7

Quick Start Deployment: Streaming Cluster in CDP Private Cloud Base

Date published: 2020-10-09

Date modified: 2021-10-08

The Cloudera logo, featuring the word "CLUDERA" in a bold, orange, sans-serif font. The letter "E" is stylized with a horizontal bar through its 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

Create a Streams Cluster on CDP Private Cloud Base.....	4
Before You Install.....	4
System Requirements for POC Streams Cluster.....	4
Disable the Firewall.....	5
Enable an NTP Service.....	6
Installing a Trial Streaming Cluster.....	7
Download the Trial version of CDP Private Cloud Base.....	7
Run the Cloudera Manager Server Installer.....	8
Install Cloudera Runtime.....	13
Set Up a Streaming Cluster.....	26
Getting Started on your Streams Cluster.....	34
Create a Kafka Topic to Store your Events.....	34
Write a few Events into the Topic.....	35
Read the Events.....	36
Monitor your Cluster from the SMM UI.....	37
After Evaluating Trial Software.....	37

Create a Streams Cluster on CDP Private Cloud Base

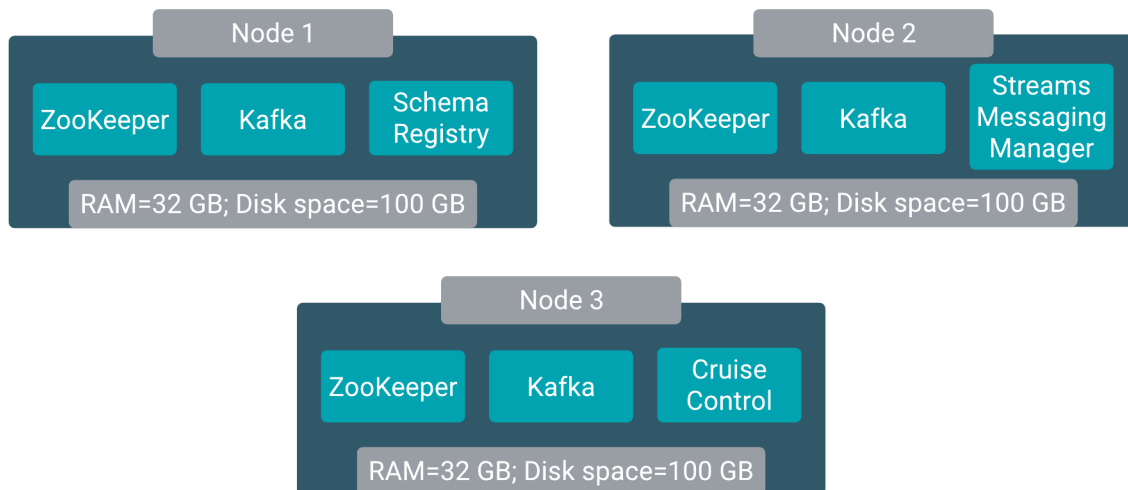
Cloudera's streaming components empower enterprises to handle some of the most complex and sophisticated streaming use cases. You can evaluate the streaming components in CDP Private Cloud Base for up to 60 days. This document walks you through the process of installing the trial software and creating a streams cluster for your proof-of-concept (POC) or sandbox environment. After you evaluate CDP Private Cloud Base, you can contact Cloudera to request a license for your production environment.

The CDP Private Cloud Base streaming components provide advanced messaging, real-time processing, and analytics on real-time streaming data. The components are:

- Cruise Control
- Apache Kafka
- Schema Registry
- Streams Messaging Manager (SMM)
- Streams Replication Manager

This document guides you through the steps to create a three-node cluster with all the streaming components with the exception of Streams Replication Manager which you do not need for a trial evaluation. You will need Streams Replication Manager for production-level performance and availability.

The following diagram shows the layout of the streaming components across the three-node cluster:



Note:

- Do not install the trial software on a NAS device. Use your local disk.
- You can use the trial version of CDP Private Cloud Base for up to 60 days.

Before You Install

Before you start the installation process, verify that your system meets the prerequisites for the trial install.

System Requirements for POC Streams Cluster

Understand the hardware, operating system, database, and other requirements for the trial CDP Private Cloud Base software.

Hardware

For each node in your cluster, allocate:

- 32 GB RAM
- 100 GB disk space



Important: Do not install the trial software on a NAS device. Use your local disk.

Operating System

See the [Cloudera Support Matrix](#) for detailed information about supported operating systems.

HTTP Proxy

The Cloudera Manager installer accesses `archive.cloudera.com` by using `yum` on RHEL systems. If your hosts access the Internet through an HTTP proxy, you can configure `yum` system-wide, to access `archive.cloudera.com` through a proxy.

To do so, modify the system configuration on the host node as follows:

OS	File	Property
RHEL-compatible	<code>/etc/yum.conf</code>	<code>proxy=http://server:port/</code>
Ubuntu	<code>/etc/apt/apt.conf</code>	<code>Acquire::http::Proxy "http://server:port";</code>

SELinux

If you are using SELinux in enforcing mode, you must disable SELinux for the Cloudera Manager installer to work.

Configure `SELINUX=disabled` in the `/etc/selinux/config` file.

Set:

```
setenforce 0
```

Cluster Host

The hosts you intend to use must satisfy the following requirements:

- You must be able to log in to the Cloudera Manager Server host using the root user account or an account that has passwordless `sudo` privileges.
- The Cloudera Manager Server host must have uniform SSH access on the same port to all hosts.
- All hosts must have access to standard package repositories for the operating system and either `archive.cloudera.com` or a local repository with the required installation files.

Disable the Firewall

To install the trial CDP Private Cloud Base software, you must disable the firewall on each node in your cluster.

Procedure

1. For `iptables`, save the existing rule set:

```
sudo iptables-save > ~/firewall.rules
```

2. Disable the firewall.

- RHEL 7:

```
sudo systemctl disable firewalld
sudo systemctl stop firewalld
```

- SLES:

```
sudo chkconfig SuSEfirewall2_setup off
sudo chkconfig SuSEfirewall2_init off
sudo rcSuSEfirewall2 stop
```

- Ubuntu:

```
sudo service ufw stop
```

Enable an NTP Service

You must configure a Network Time Protocol (NTP) service on each node in your cluster. Most operating systems include the `ntpd` service for time synchronization.

About this task

RHEL 7 compatible operating systems use `chronyd` by default instead of `ntpd`. If `chronyd` is running (on any OS), Cloudera Manager uses it to determine whether the host clock is synchronized. Otherwise, Cloudera Manager uses `ntpd`.

To use `ntpd` for time synchronization:

Procedure

1. Install the `ntp` package:

- RHEL compatible:

```
yum install ntp
```

- Ubuntu:

```
apt-get install ntp
```

2. Edit the `/etc/ntp.conf` file to add NTP servers, as in the following example:

```
server 0.pool.ntp.org
server 1.pool.ntp.org
server 2.pool.ntp.org
```

3. Start the `ntpd` service:

- RHEL 7 Compatible:

```
sudo systemctl start ntpd
```

- Ubuntu:

```
sudo service ntpd start
```

4. Configure the ntpd service to run at boot:

- RHEL 7 Compatible:

```
sudo systemctl enable ntpd
```

- Ubuntu:

```
chkconfig ntpd on
```

5. Synchronize the system clock to the NTP server:

```
ntpdate -u <ntp_server>
```

6. Synchronize the hardware clock to the system clock:

```
hwclock --systohc
```

Installing a Trial Streaming Cluster

When you install the CDP Private Cloud Base trial software, Cloudera Manager automates the installation of the Oracle JDK, Cloudera Manager Server, embedded PostgreSQL database, Cloudera Manager Agent, Runtime, and managed service software on cluster hosts. Cloudera Manager also configures databases for the Cloudera Manager Server and Hive Metastore and optionally for Cloudera Management Service roles.



Important: This procedure is intended for trial and proof-of-concept deployments only. It is not supported for production deployments because it is not designed to scale.

Refer to the following steps to install a trial cluster.

Download the Trial version of CDP Private Cloud Base

You can download the trial version of CDP Private Cloud Base from the [Cloudera Download](#) site.

About this task

You can use the trial software for 60 days without obtaining a license key file. The trial installation includes an embedded PostgreSQL database and is not suitable for a production environment.

Procedure

1. Go to the trial [download page](#) for CDP Private Cloud Base.
2. Click Try Now.
3. Follow the download-instructions.

What to do next

Run the Cloudera Manager Server Installer.

Related Information

[CDP Private Cloud Trial Download](#)

Run the Cloudera Manager Server Installer

Run the Cloudera Manager installer to the cluster host to which you are installing the Cloudera Manager Server. By default, the automated installer binary (cloudera-manager-installer.bin) installs the highest version of Cloudera Manager.

Before you begin

- Download the trial software.

Procedure

1. Run the Cloudera Manager installer:

- a) Change cloudera-manager-installer.bin to have execute permissions:

```
chmod u+x cloudera-manager-installer.bin
```

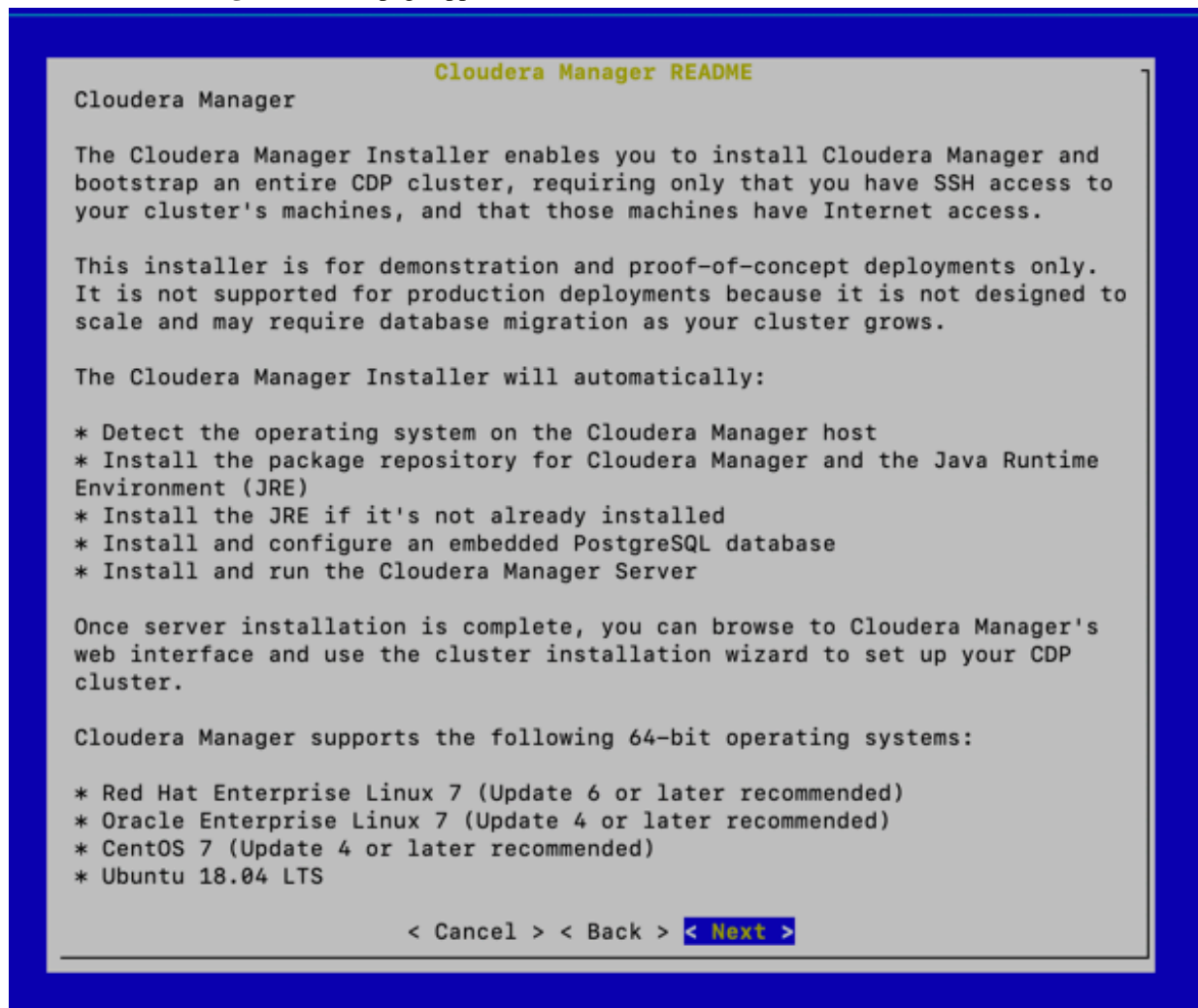
- b) Run the Cloudera Manager Server installer:

```
sudo ./cloudera-manager-installer.bin
```

- c) For clusters without Internet access: Install Cloudera Manager packages from a local repository:

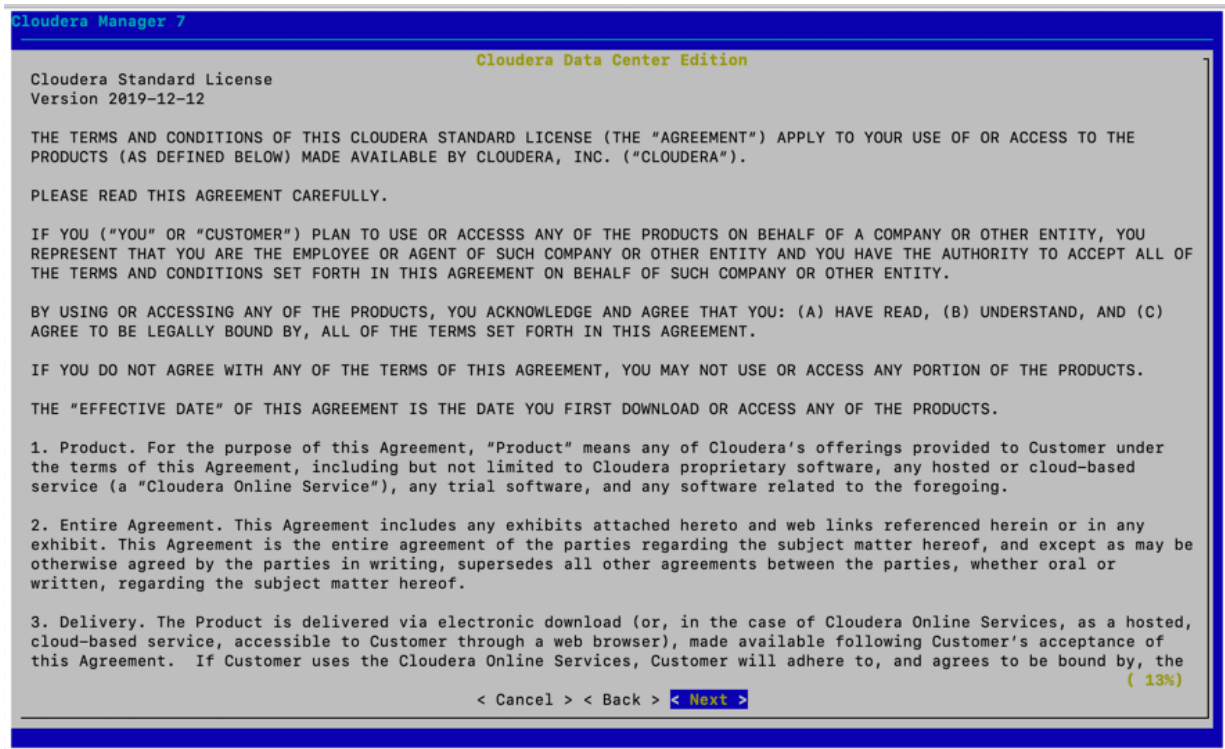
```
sudo ./cloudera-manager-installer.bin --skip_repo_package=1
```

The **Cloudera Manager Read Me** page appears.



2. Click Next.

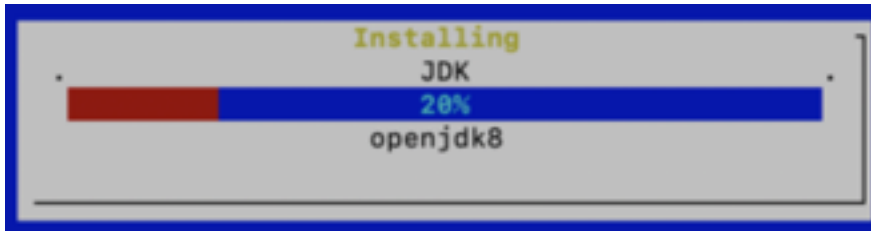
The **Cloudera Standard License** page appears.



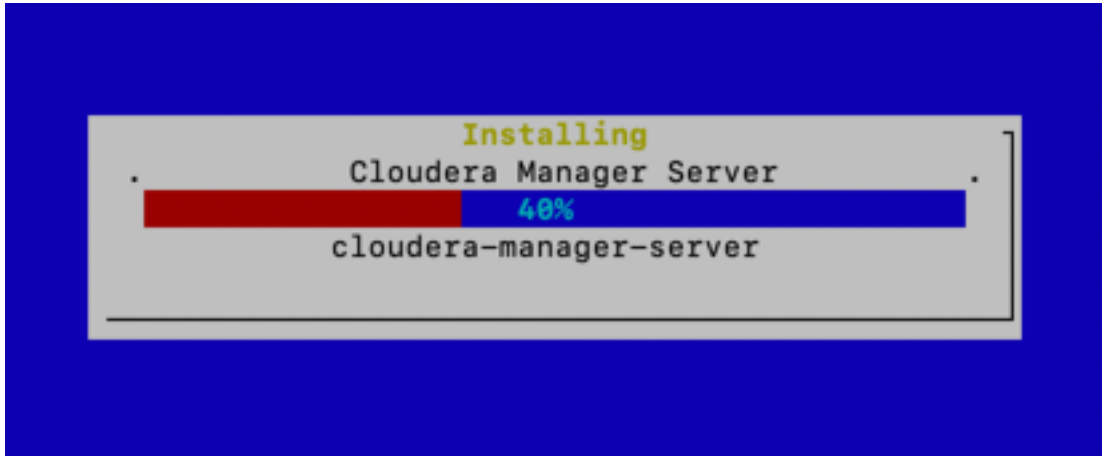
3. Click Next to accept the license agreement.

The the installer starts and does the following:

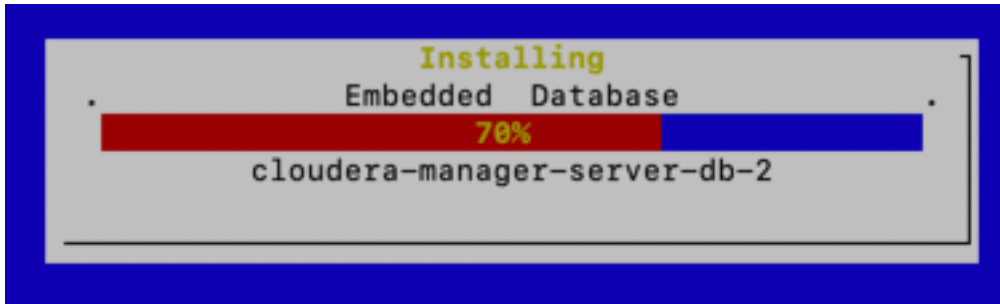
- a. Installs Oracle JDK.



- b. Installs the Cloudera Manager Server.



- c. Installs the embedded PostgreSQL packages and starts the database and Cloudera Manager Server.



Note: If the installation is interrupted, run the following command on the Cloudera Manager Server host before you retry the installation:

```
sudo /usr/share/cmf/uninstall-cloudera-manager.sh
```

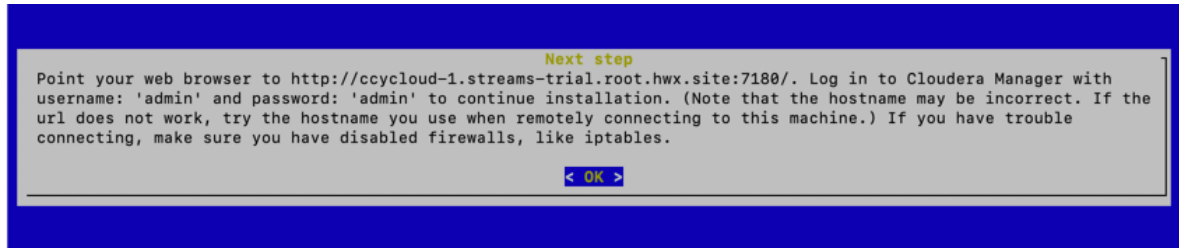
The log files for the installer are stored in `/var/log/cloudera-manager-installer/`.

4. Exit the installer:

- a) When the installation completes, the complete URL for the Cloudera Manager Admin Console displays, including the default port number: 7180.



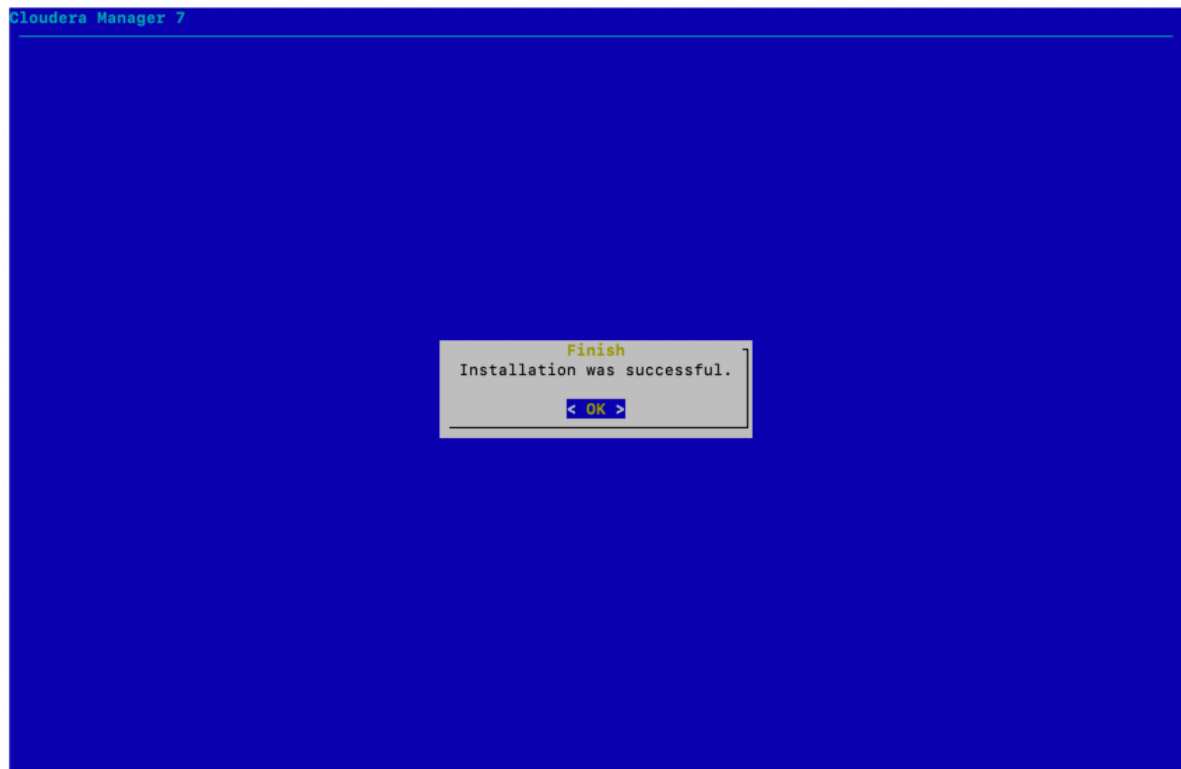
Important: Make a note of this URL or take a screen capture as you will need it for the next task.



- b) Click OK.

The success message appears.

- c) Click OK to exit the installer.



- d) Wait a few minutes for the Cloudera Manager Server to start. To observe the startup process, run `sudo tail -f /var/log/cloudera-scm-server/cloudera-scm-server.log` on the Cloudera Manager Server host. When you see the following log entry, the Cloudera Manager Admin Console is ready:

```
INFO WebServerImpl:com.cloudera.server.cmf.WebServerImpl: Started Jetty server.
```

What to do next

Install Cloudera Runtime

Install Cloudera Runtime

After you have installed Cloudera Manager, log in to Cloudera Manager to access the **Add Cluster - Installation** wizard. Here you will add hosts to form a cluster and install Cloudera Runtime and Cloudera Manager Agent software.

Before you begin

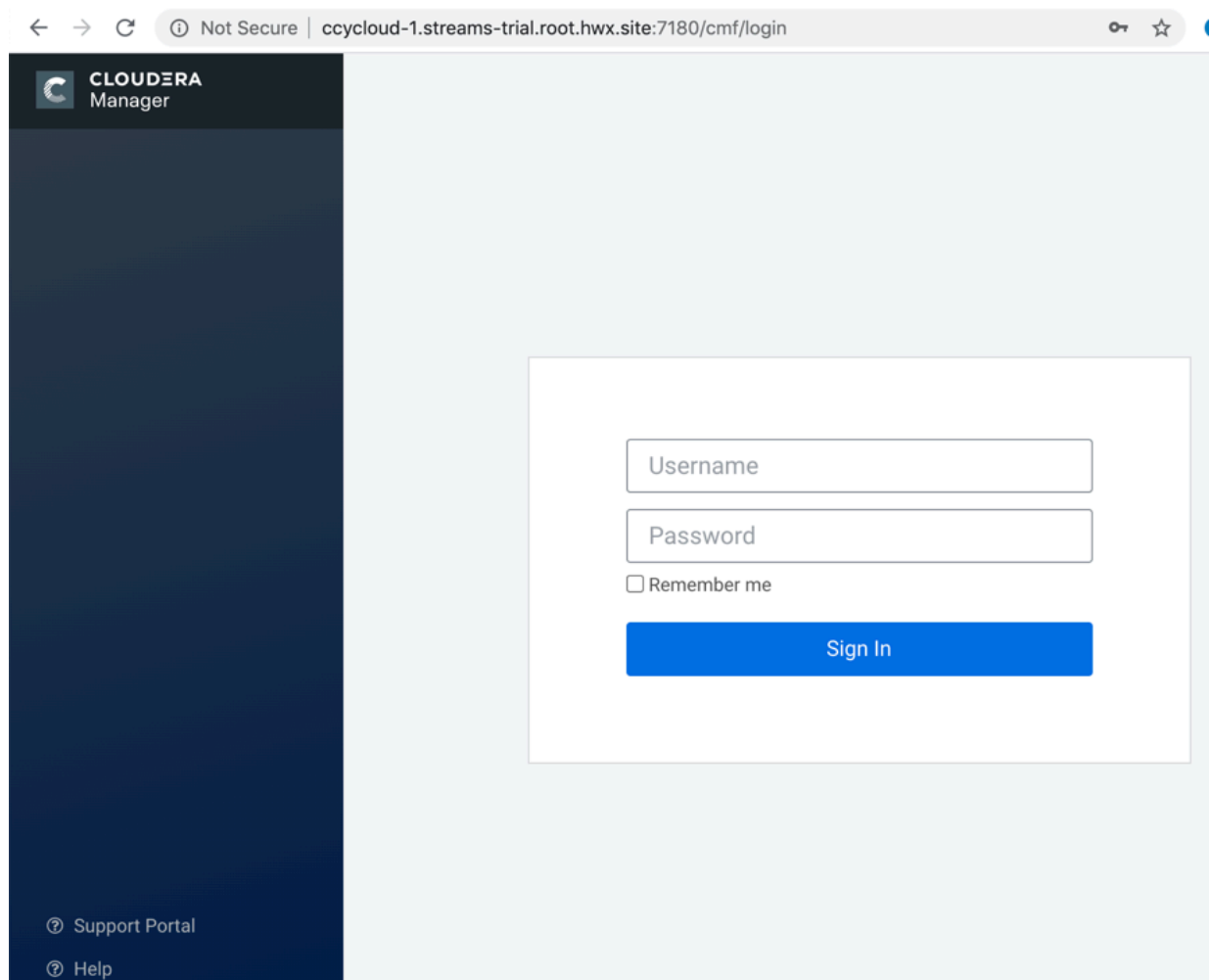
- You have installed Cloudera Manager.

Procedure

1. In a web browser, enter the URL that the Cloudera Manager Installer displayed in the previous task: `http://<server_host>:7180`, where `<server_host>` is the FQDN or IP address of the host where the Cloudera Manager Server is running.

For example: `http://ccycloud-1.streams-trial.root.hwx.site:7180`

The **Cloudera Manager Sign In** page appears.



← → ↻ ⓘ Not Secure | ccycloud-1.streams-trial.root.hwx.site:7180/cm/login

CLOUDERA
Manager

Support Portal
Help

Username

Password

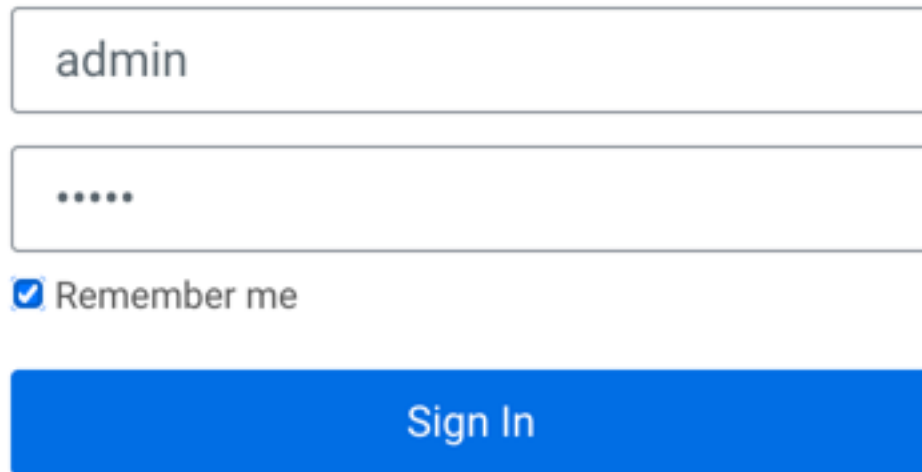
☐ Remember me

Sign In

2. Sign in with the default credentials:

- Username: admin
- Password: admin

Click Sign In.



A sign-in form with two input fields. The first field contains the text 'admin'. The second field contains five dots, indicating a password. Below the password field is a checkbox labeled 'Remember me' which is checked. At the bottom is a blue button with the text 'Sign In'.

The **Welcome to Cloudera Manager** page appears.

3. Select:

- Try Cloudera Data Platform for 60 days
- Yes, I accept the Cloudera Standard License Terms and Conditions

Welcome to Cloudera Manager 7.1.3

Upload License File

☐ Upload Cloudera Data Platform License

Cloudera Data Platform provides important features that help you manage and monitor your Hadoop clusters in mission-critical environments. Cloudera Data Platform is a subscription service with enhanced capabilities and support. [Contact Cloudera Sales](#)

 Upload License File (Accept .txt or .zip)

☒ Try Cloudera Data Platform for 60 days

 After the trial period, you will need a valid Cloudera Data Platform license to access the Cloudera Manager Admin Console. Your cluster and data will remain unaffected.

Cloudera Standard License

Version 2019-12-12

THE TERMS AND CONDITIONS OF THIS CLOUDERA STANDARD LICENSE (THE "AGREEMENT") APPLY TO YOUR USE OF OR ACCESS TO THE PRODUCTS (AS DEFINED BELOW) MADE AVAILABLE BY CLOUDERA, INC. ("CLOUDERA").

PLEASE READ THIS AGREEMENT CAREFULLY.

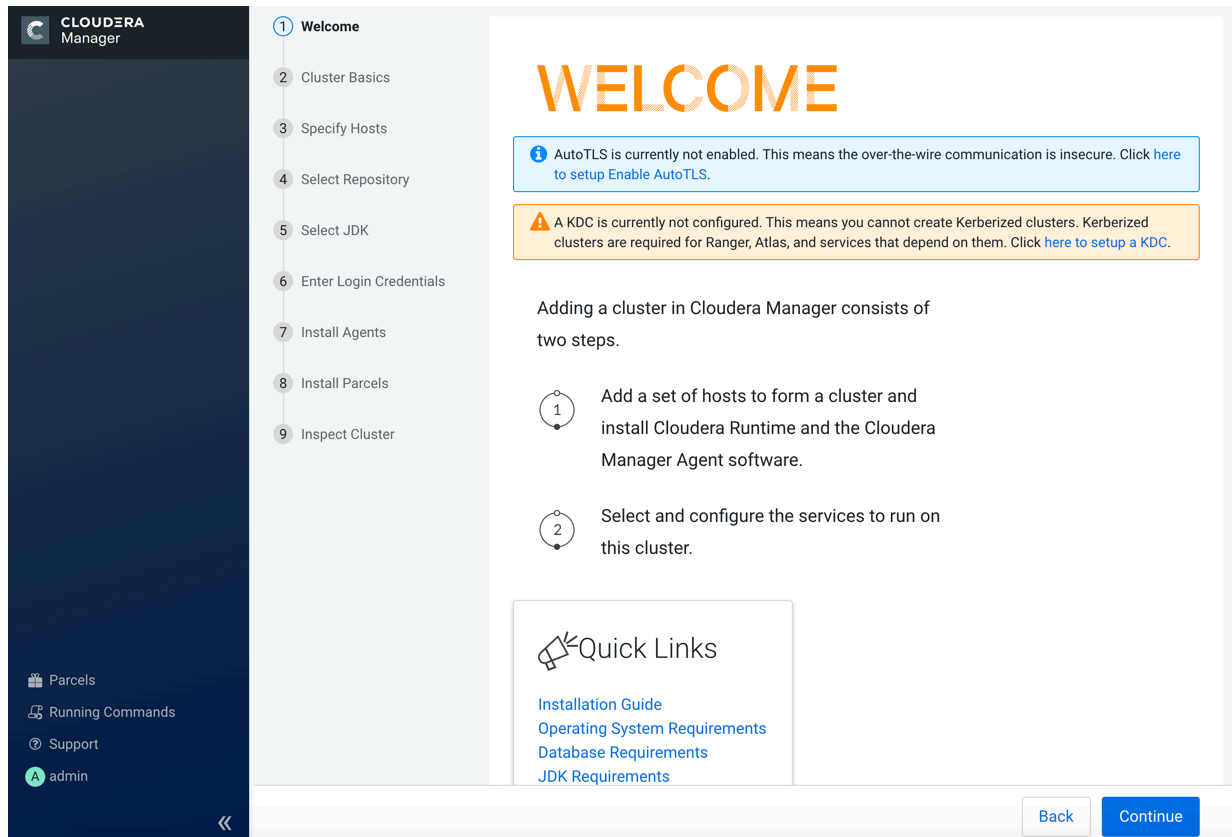
IF YOU ("YOU" OR "CUSTOMER") PLAN TO USE OR ACCESS ANY OF THE PRODUCTS ON BEHALF OF A COMPANY OR OTHER ENTITY, YOU REPRESENT THAT YOU ARE THE EMPLOYEE OR AGENT OF SUCH

☒ Yes, I accept the Cloudera Standard License Terms and Conditions.

Continue

4. Click Continue.

The **Add Cluster - Installation** page, **Welcome** section appears. The steps on the left let you know where you are in the workflow.



5. Click Continue.

The Cluster Basics section appears.

6. Enter a name for the cluster and click Continue.

Add Cluster - Installation

✓ Welcome

2 **Cluster Basics**

3 Specify Hosts

4 Select Repository

5 Select JDK

6 Enter Login Credentials

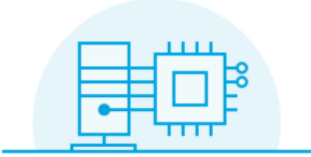
7 Install Agents

8 Install Parcels

9 Inspect Cluster

Cluster Basics

Cluster Name



Regular Cluster

A Regular Cluster contains storage nodes, compute nodes, and other services such as metadata and security collocated in a single cluster.

Back

Continue

The **Specify Hosts** section appears.

- Enter the cluster host names or IP addresses in the Hostnames field.

Add Cluster - Installation

✓ Welcome

✓ Cluster Basics

3 Specify Hosts

4 Select Repository

5 Select JDK

6 Enter Login Credentials

7 Install Agents

8 Install Parcels


9 Inspect Cluster

Specify Hosts

Hosts should be specified using the same hostname (FQDN) that they will identify themselves with. Cloudera recommends including Cloudera Manager Server's host. This also enables health monitoring for that host.

Hostname

ccycloud-1.streams-trial.root.hwx.site
ccycloud-2.streams-trial.root.hwx.site
ccycloud-3.streams-trial.root.hwx.site

Hint: Search for hostnames or IP addresses using [patterns](#) 

SSH Port:

22

Search

You can specify host name and IP address ranges as follows:

Expansion Range	Matching Hosts
10.1.1.[1-4]	10.1.1.1, 10.1.1.2, 10.1.1.3, 10.1.1.4
host[1-3].example.com	host1.example.com, host2.example.com, host3.example.com
host[07-10].example.com	host07.example.com, host08.example.com, host09.example.com, host10.example.com

8. Click Search.

Cloudera Manager discovers the hosts.

Add Cluster - Installation

✓ Welcome

✓ Cluster Basics

3 Specify Hosts

4 Select Repository

5 Select JDK

6 Enter Login Credentials

7 Install Agents

8 Install Parcels


9 Inspect Cluster

Specify Hosts

Hosts should be specified using the same hostname (FQDN) that they will identify themselves with. Cloudera recommends including Cloudera Manager Server's host. This also enables health monitoring for that host.

Hostname

ccycloud-1.streams-trial.root.hwx.site
ccycloud-2.streams-trial.root.hwx.site
ccycloud-3.streams-trial.root.hwx.site

Hint: Search for hostnames or IP addresses using [patterns](#) 

SSH Port:

3 hosts scanned, 3 running SSH.
Click the first checkbox, hold down the Shift key and click the last checkbox to select a range.

<input checked="" type="checkbox"/>	Expanded Query ↑	Hostname (FQDN)	IP Address	Currently Managed	Result
<input checked="" type="checkbox"/>	ccycloud-1.streams-trial.root.hwx.site	ccycloud-1.streams-trial.root.hwx.site	172.27.123.204	No	Host was successfully scanned.
<input checked="" type="checkbox"/>	ccycloud-2.streams-trial.root.hwx.site	ccycloud-2.streams-trial.root.hwx.site	172.27.26.143	No	Host was successfully scanned.
<input checked="" type="checkbox"/>	ccycloud-3.streams-trial.root.hwx.site	ccycloud-3.streams-trial.root.hwx.site	172.27.92.198	No	Host was successfully scanned.

9. Verify host entries, deselect any that you do not want to install services on, and click Continue.

The **Select Repository** section appears.

10. Select the following options:

- Public Cloudera Repository
- Use Parcels
- The version of Cloudera Runtime that you want to install.
- In the Additional Parcels section, None.

Add Cluster - Installation

Select Repository

Cloudera Manager Agent

Cloudera Manager Agent 7.1.3 (#4999720) needs to be installed on all new hosts.

Repository Location ☒ Public Cloudera Repository

Ensure the above version is listed in <https://archive.cloudera.com/cm7> and that you have access to that repository. Requires direct Internet access on all hosts.

☐ Custom Repository

CDH and other software

Cloudera recommends the use of parcels for installation over packages, because parcels enable Cloudera Manager to easily manage the software on your cluster, automating the deployment and upgrade of service binaries. Electing not to use parcels will require you to manually upgrade packages on all hosts in your cluster when software updates are available, and will prevent you from using Cloudera Manager's rolling upgrade capabilities.

Install Method ☐ Use Packages

☒ Use Parcels (Recommended) [Parcel Repositories & Network Settings](#) [Other Parcel Configurations](#)

Version **Versions that are too new for this version of Cloudera Manager (7.1.3) will not be shown.**

☒ Cloudera Runtime 7.1.3-1.cdh7.1.3.p0.4992530

☐ CDH 6.3.2-1.cdh6.3.2.p0.1605554

☐ CDH 5.16.2-1.cdh5.16.2.p0.8

Additional Parcels ☐ ACCUMULO 1.9.2-1.ACCUMULO6.1.0.p0.908695

☐ ACCUMULO 1.7.2-5.5.0.ACCUMULO5.5.0.p0.8

☒ None

[Back](#) [Continue](#)

11. Click Continue.

The **Select JDK** section appears.

12. Select Install a Cloudera-provided version of OpenJDK.

Add Cluster - Installation

Select JDK

Selected Version	Cloudera Runtime 7.1
Supported JDK Version	OpenJDK 8, 11 or Oracle JDK 8, 11

[More details on supported JDK version.](#)

If you plan to use JDK 11, you will need to install it manually on all hosts and then select the **Manually manage JDK** option below.

☐ Manually manage JDK

Please ensure that a supported JDK is **already installed on all hosts. You will need to manage installing the unlimited strength JCE policy file, if necessary.**

☒ **Install a Cloudera-provided version of OpenJDK**

By proceeding, Cloudera will install a supported version of OpenJDK version 8.

☐ **Install a system-provided version of OpenJDK**

By proceeding, Cloudera will install the default version of OpenJDK version 8 provided by the Operating System.

[Back](#) [Continue](#)

13. Click Continue.

The **Enter Login Credentials** section appears.

14. Do the following:

- Select root.
- Select All hosts accept same password.
- Enter the password for the account that allows root access to your hosts.
- Click Continue.

Add Cluster - Installation

The screenshot shows the 'Enter Login Credentials' step of the Cloudera installation wizard. On the left is a vertical sidebar with a progress indicator showing steps from 'Welcome' to 'Inspect Cluster'. Step 6, 'Enter Login Credentials', is currently selected and highlighted. The main content area is titled 'Enter Login Credentials' and contains the following text and form elements:

Root access to your hosts is required to install the Cloudera packages. This installer will connect to your hosts via SSH and log in either directly as root or as another user with password-less sudo/pbrun privileges to become root.

Login To All Hosts As: ☒ root
☐ Another user

You may connect via password or public-key authentication for the user selected above.

Authentication Method: ☒ All hosts accept same password
☐ All hosts accept same private key

Enter Password:

Confirm Password:

SSH Port:

Number of Simultaneous Installations:
(Running a large number of installations at once can consume large amounts of network bandwidth and other system resources)

At the bottom right of the window are two buttons: 'Back' and 'Continue'.

The **Install Agents** section appears showing the progress of the installation.

Add Cluster - Installation

✓ Welcome

✓ Cluster Basics

✓ Specify Hosts

✓ Select Repository

✓ Select JDK

✓ Enter Login Credentials

7 Install Agents

8 Install Parcels

9 Inspect Cluster

Install Agents

Installation in progress.

0 of 3 host(s) completed successfully. [Abort Installation](#)

Hostname	IP Address	Progress	Status
ccycloud-1.streams-trial.root.hwx.site	172.27.123.204		Installing openjdk8 package... Details
ccycloud-2.streams-trial.root.hwx.site	172.27.26.143		Installing openjdk8 package... Details
ccycloud-3.streams-trial.root.hwx.site	172.27.92.198		Installing openjdk8 package... Details

After the agents are installed, the **Install Parcels** section appears showing the progress of the parcel installation.

Add Cluster - Installation

✓ Welcome

✓ Cluster Basics

✓ Specify Hosts

✓ Select Repository

✓ Select JDK

✓ Enter Login Credentials

✓ Install Agents

8 Install Parcels

9 Inspect Cluster

Install Parcels

The selected parcels are being downloaded and installed on all the hosts in the cluster.

✓ Cloudera Runtime 7.1.3-1....

Downloaded: 3%

Distributed: 0/0

Unpacked: 0/0

Activated: 0/0

Back

Continue

After the parcels are installed the **Inspect Cluster** section appears.

Add Cluster - Installation

The screenshot shows the 'Inspect Cluster' step in the Cloudera Manager installation wizard. On the left is a vertical progress bar with steps: Welcome, Cluster Basics, Specify Hosts, Select Repository, Select JDK, Enter Login Credentials, Install Agents, Install Parcels, and Inspect Cluster (the current step, marked with a '9'). The main content area is titled 'Inspect Cluster' and contains a blue informational box stating: 'You have created a new empty cluster. Cloudera recommends that you run the following inspections. For accurate measurements, Cloudera recommends that they are performed sequentially.' Below this are two sections: 'Inspect Network Performance' and 'Inspect Hosts'. Each section has a clock icon, a description, a link to 'Advanced Options', and a button to start the inspection. At the bottom, there are three radio button options for handling detected issues.

Inspect Cluster

You have created a new empty cluster. Cloudera recommends that you run the following inspections. For accurate measurements, Cloudera recommends that they are performed sequentially.

Inspect Network Performance

Once the inspection is complete, review the inspector results before proceeding.

[Advanced Options](#)

[Inspect Network Performance](#)

Inspect Hosts

Once the inspection is complete, review the inspector results before proceeding.

[Inspect Hosts](#)

☒ Fix the issues and run the inspection tools again.

☐ Quit the wizard and Cloudera Manager will delete the temporarily created cluster.

☐ I understand the risks of not running the inspections or the detected issues, let me continue with cluster setup.

15. Do the following:

- a) Select Inspect Network Performance.
You can click Advanced Options to customize some ping parameters.
- b) After the network inspector completes, click Show Inspector Results to view the results in a new tab.
Address any reported issues, and click Run Again.
- c) Click Inspect Hosts.
- d) After the host inspector completes, click Show Inspector Results to view the results in a new tab.
Address any reported issues, and click Run Again.

Add Cluster - Installation

The screenshot shows the 'Inspect Cluster' step in the Cloudera installation wizard. On the left is a vertical sidebar with a list of steps: Welcome, Cluster Basics, Specify Hosts, Select Repository, Select JDK, Enter Login Credentials, Install Agents, Install Parcels, and Inspect Cluster (which is highlighted with a circled '9'). The main content area is titled 'Inspect Cluster' and contains a blue informational box stating: 'You have created a new empty cluster. Cloudera recommends that you run the following inspections. For accurate measurements, Cloudera recommends that they are performed sequentially.' Below this, there are two inspection sections. The first is 'Inspect Network Performance', marked with a green checkmark, with a link to '> Advanced Options'. It shows a status of 'Success' (green checkmark), 'Last Run' as 'a few seconds ago', and 'Duration' as '18.11s'. It includes buttons for 'Run Again' and 'More', and a 'Show Inspector Results' button with an external link icon. The second section is 'Inspect Hosts', also marked with a green checkmark. It shows a status of 'Success' (green checkmark), 'Last Run' as 'a few seconds ago', and 'Duration' as '18.48s'. It includes buttons for 'Run Again' and 'More', and a 'Show Inspector Results' button with an external link icon. At the bottom right of the wizard are 'Back' and 'Continue' buttons.

16. Click Continue.

The **Add Cluster - Configuration** page appears.

Add Cluster - Configuration

1 Select Services

2 Assign Roles

3 Setup Database

4 Enter Required Parameters

5 Review Changes

6 Command Details

7 Summary

Select Services

Choose a combination of services to install.

☒ **Data Engineering**
Process, develop, and serve predictive models.
Services: HDFS, YARN, YARN Queue Manager, Ranger, Atlas, Hive, Hive on Tez, Spark, Oozie, Hue, and Data Analytics Studio

☐ **Data Mart**
Browse, query, and explore your data in an interactive way.
Services: HDFS, Ranger, Atlas, Hive, Impala, and Hue

☐ **Operational Database**
Real-time insights for modern data-driven business.
Services: HDFS, Ranger, Atlas, and HBase

☐ **Custom Services**
Choose your own services. Services required by chosen services will automatically be included.

This wizard will also install the **Cloudera Management Service**. These are a set of components that enable monitoring, reporting, events, and alerts; these components require databases to store information, which will be configured on the next page.

Results

This completes the **Add Cluster - Installation** wizard.

What to do next

Set up a cluster.

Set Up a Streaming Cluster

After completing the **Add Cluster - Installation** wizard, the **Add Cluster - Configuration** wizard automatically starts. Here you will select the streaming services, specify the host to run each service on, test the connection to the database, and run the command to set up your cluster.

Before you begin

- You have installed Cloudera Manager.
- You have installed Cloudera Runtime.

Procedure

1. Verify you are on the **Add Cluster - Configuration** page of the Cloudera Manager UI.
The list of steps on the left let you know where you are in the workflow.
2. Verify that you are on the **Select Services** section.

3. Select the Custom Services option.

A list of services appear.

Add Cluster - Configuration

1 Select Services

2 Assign Roles

3 Setup Database

4 Enter Required Parameters

5 Review Changes

6 Command Details

7 Summary

Select Services

Choose a combination of services to install.

☐ **Data Engineering**
Process, develop, and serve predictive models.
Services: HDFS, YARN, YARN Queue Manager, Ranger, Atlas, Hive, Hive on Tez, Spark, Oozie, Hue, and Data Analytics Studio

☐ **Data Mart**
Browse, query, and explore your data in an interactive way.
Services: HDFS, Ranger, Atlas, Hive, Impala, and Hue

☐ **Operational Database**
Real-time insights for modern data-driven business.
Services: HDFS, Ranger, Atlas, and HBase

☒ **Custom Services**
Choose your own services. Services required by chosen services will automatically be included.

Service Type	Description
<input type="checkbox"/> Atlas	Apache Atlas provides a set of metadata management and governance services that enable you to find, organize, and manage data assets. This service requires Kerberos.
<input type="checkbox"/> Core Configuration	Core Configuration contains settings used by most services. Required for clusters without HDFS.
<input type="checkbox"/> Cruise Control	Cruise Control simplifies the operation of Kafka clusters automating workload rebalancing and self-healing.
<input type="checkbox"/> Data Analytics Studio	Data Analytics Studio is the one stop shop for Apache Hive warehousing. Query, optimize and administrate your data with this powerful interface.
<input type="checkbox"/> HBase	Apache HBase is a highly scalable, highly resilient NoSQL OLTP database that enables applications to leverage big data.
<input type="checkbox"/> HDFS	Apache Hadoop Distributed File System (HDFS) is the primary storage system used by Hadoop applications. HDFS creates multiple replicas of data blocks and distributes them on compute hosts throughout a cluster to enable reliable, extremely rapid computations.
<input type="checkbox"/> Hive	Apache Hive is a SQL-based data warehouse system for PaaS and self-managed clusters. Includes Hive Metastore and HiveServer2.

Back

Continue

4. Scroll through the list and select the following services:

- Cruise Control
- Kafka
- Schema Registry
- Streams Messaging Manager
- ZooKeeper

Service Type	Description
<input type="checkbox"/> Atlas	Apache Atlas provides a set of metadata management and governance services that enable you to find, organize, and manage data assets. This service requires Kerberos.
<input type="checkbox"/> Core Configuration	Core Configuration contains settings used by most services. Required for clusters without HDFS.
<input checked="" type="checkbox"/> Cruise Control	Cruise Control simplifies the operation of Kafka clusters automating workload rebalancing and self-healing.
<input type="checkbox"/> Data Analytics Studio	Data Analytics Studio is the one stop shop for Apache Hive warehousing. Query, optimize and administrate your data with this powerful interface.
<input type="checkbox"/> HBase	Apache HBase is a highly scalable, highly resilient NoSQL OLTP database that enables applications to leverage big data.
<input type="checkbox"/> HDFS	Apache Hadoop Distributed File System (HDFS) is the primary storage system used by Hadoop applications. HDFS creates multiple replicas of data blocks and distributes them on compute hosts throughout a cluster to enable reliable, extremely rapid computations.
<input type="checkbox"/> Hive	Apache Hive is a SQL based data warehouse system. In CDH 6 and earlier, this service includes Hive Metastore and HiveServer2. In Cloudera Runtime 7.0 and later, this service only includes the Hive Metastore; HiveServer2 and other components of the Hive execution engines are part of the Hive on Tez service.
<input type="checkbox"/> Hive on Tez	Hive on Tez is a SQL query engine using Apache Tez.
<input type="checkbox"/> Hue	Hue is the leading SQL Workbench for optimized, interactive query design and data exploration.
<input type="checkbox"/> Impala	Apache Impala provides a real-time SQL query interface for data stored in HDFS and HBase. Impala requires the Hive service and shares the Hive Metastore with Hue.
<input checked="" type="checkbox"/> Kafka	Apache Kafka is publish-subscribe messaging rethought as a highly scalable distributed commit log.
<input type="checkbox"/> Key-Value Store Indexer	Key-Value Store Indexer listens for changes in data inside tables contained in HBase and indexes them using Solr.
<input type="checkbox"/> Knox	The Apache Knox Gateway is an Application Gateway for interacting with the REST APIs and Uis of Apache Hadoop deployments. This service requires Kerberos.
<input type="checkbox"/> Kudu	Apache Kudu is a data store that enables real-time analytics on fast changing data.
<input type="checkbox"/> Livy	Apache Livy is a REST service for deploying Spark applications.
<input type="checkbox"/> Oozie	Apache Oozie is a workflow coordination service to manage and schedule data processing jobs on your cluster.
<input type="checkbox"/> Ozone	Apache Hadoop Ozone is a scalable, distributed object store for Hadoop.
<input type="checkbox"/> Phoenix	Apache Phoenix is a scale-out relational database that supports OLTP workloads and provides secondary indexes, materialized views, star schema support, and common HBase optimizations. Phoenix uses Apache HBase as the underlying data store.
<input type="checkbox"/> Ranger	Apache Ranger is a framework to enable, monitor and manage comprehensive data security across the Hadoop platform. This service requires Kerberos.
<input checked="" type="checkbox"/> Schema Registry	Schema Registry is a shared repository of schemas that allows applications to flexibly interact with each other. A common Schema Registry provides end-to-end data governance and introduces operational efficiency by saving and retrieving reusable schema, defining relationships between schemas and enabling data providers and consumers to evolve at different speeds.
<input type="checkbox"/> Solr	Apache Solr is a highly scalable, distributed service for indexing and relevance-based exploring of all forms of data.
<input type="checkbox"/> Spark	Apache Spark is an open source cluster computing system. This service runs Spark as an application on YARN.
<input checked="" type="checkbox"/> Streams Messaging Manager	Streams Messaging Manager (SMM) is an operations monitoring and management tool that provides end-to-end visibility in an enterprise Apache Kafka environment.
<input type="checkbox"/> Streams Replication Manager	Streams Replication Manager (SRM) is an enterprise-grade replication solution that enables fault tolerant, scalable, and robust cross-cluster Kafka topic replication.
<input type="checkbox"/> Tez	Apache Tez is the next generation Hadoop Query Processing framework written on top of YARN.
<input type="checkbox"/> YARN	Apache Hadoop MapReduce 2.0 (MRV2), or YARN, is a data computation framework that supports MapReduce applications (requires HDFS).
<input type="checkbox"/> YARN Queue Manager	YARN Queue Manager is the queue management user interface for Apache Hadoop YARN Capacity Scheduler.
<input type="checkbox"/> Zeppelin	Apache Zeppelin is a web-based notebook that enables data-driven, interactive data analytics and collaborative documents with SQL, Scala and more.
<input checked="" type="checkbox"/> ZooKeeper	Apache ZooKeeper is a centralized service for maintaining and synchronizing configuration data.

Back Continue

5. Click Continue.

The **Assign Roles** section appears with suggested role assignments for the hosts in your cluster.

Add Cluster - Configuration

✓ Select Services

2 Assign Roles

3 Setup Database

4 Enter Required Parameters

5 Review Changes

6 Command Details

7 Summary

Assign Roles

You can customize the role assignments for your new cluster here, but if assignments are made incorrectly, such as assigning too many roles to a single host, this can impact the performance of your services. Cloudera does not recommend altering assignments unless you have specific requirements, such as having pre-selected a specific host for a specific role.

You can also view the role assignments by host. [View By Host](#)

Kafka

Kafka Broker

Select hosts

Too few hosts assigned, minimum is 1.

Kafka MirrorMaker

Select hosts

Kafka Connect

Select hosts

Gateway

Select hosts

Cruise Control

Cruise Control Server × 1 New

ccycloud-2.streams-trial.root.h...

Cloudera Management Service

Service Monitor × 1 New

ccycloud-1.streams-trial.root.h...

Activity Monitor

Select a host

Host Monitor × 1 New

ccycloud-1.streams-trial.root.h...

Reports Manager × 1 New

ccycloud-1.streams-trial.root.h...

Event Server × 1 New

ccycloud-1.streams-trial.root.h...

Alert Publisher × 1 New

ccycloud-1.streams-trial.root.h...

Telemetry Publisher

Back

Continue

6. In the Kafka Broker field, click Select hosts.

7. Select all hosts for Kafka Broker and click OK.

3 Hosts Selected

Select hosts for a new or existing role. The host list is filtered to remove hosts that are not valid candidates; these include hosts that are unhealthy, members of other clusters, or have an incompatible version of the software installed on them.

Q Enter hostnames: host01, host[01-10], IP addresses or rack. [Search](#)

Tip: Click the first checkbox, hold down the Shift key and click the last checkbox to select a range.

<input checked="" type="checkbox"/> Hostname ↑	IP Address	Rack	Cores	Physical Memory	Existing Roles	Added Roles
<input checked="" type="checkbox"/> ccycloud-1.streams-trial.root.hwx.site	172.27.123.204	/default	88	251.6 GiB	SM	HM RM ES AP S KB
<input checked="" type="checkbox"/> ccycloud-2.streams-trial.root.hwx.site	172.27.26.143	/default	32	251.4 GiB	CCS	SRS SM... KB
<input checked="" type="checkbox"/> ccycloud-3.streams-trial.root.hwx.site	172.27.92.198	/default	32	251.5 GiB	KB	

1 - 3 of 3

Cancel

OK

8. Assign Cruise Control, Schema Registry, and SMM to separate hosts.

9. Assign ZooKeeper to all hosts.

ZooKeeper must be on an odd number of hosts.

29

10. Click View By Host to see the host and role pairing.

The **View By Host** window appears.

View By Host

This table is grouped by hosts having the same roles assigned to them.

Hosts	Count	Existing Roles	Added Roles
ccycloud-1.streams-trial.root.hwx.site	1		KB SM HM RM ES AP SRS S
ccycloud-2.streams-trial.root.hwx.site	1		KB SM... S
ccycloud-3.streams-trial.root.hwx.site	1		KB CCS S

Close

11. Verify that the services are on the right hosts and click Close to close the **View By Host** window.



Note: Before you close, make a note of the node where Service Monitor is running. You will need this information later.

12. Back on the **Assign Roles** section, click Continue.

The **Setup Database** section appears with pre-populated database names and passwords.

Setup Database

Configure and test database connections. Create the databases first according to the **Installing and Configuring an External Database** section of the [Installation Guide](#).

☐ Use Custom Databases ☒ Use Embedded Database

! The embedded PostgreSQL database is not supported for use in production environments. When using the embedded database, passwords are automatically generated. Please copy them down.

Streams Messaging Manager

Type	Database Hostname	Database Name	Username
PostgreSQL	ccycloud-1.streams-trial.root.hwx.site:7432	streams_messaging_manager	streams_messaging_manager
		Password	
		SFqcYP1nSH	

Reports Manager

Currently assigned to run on **ccycloud-1.streams-trial.root.hwx.site**.

Type	Database Hostname	Database Name	Username
PostgreSQL	ccycloud-1.streams-trial.root.hwx.site:7432	rman	rman
		Password	
		7Xj0AGPIQn	

Schema Registry

Type	Database Hostname	Database Name	Username
PostgreSQL	ccycloud-1.streams-trial.root.hwx.site:7432	schemaregistry	schemaregistry
		Password	
		ObYCS5y60C	

Test Connection

Back

Continue

13. Click Test Connection to validate the settings.

14. After verifying that each connection is successful, click Continue.

Setup Database

Configure and test database connections. Create the databases first according to the **Installing and Configuring an External Database** section of the [Installation Guide](#).

☐ Use Custom Databases ☒ Use Embedded Database

! The embedded PostgreSQL database is not supported for use in production environments. When using the embedded database, passwords are automatically generated. Please copy them down.

Streams Messaging Manager

✓ Skipped. Cloudera Manager will create this database in a later step.

Type	Database Hostname	Database Name	Username
PostgreSQL	ccycloud-1.streams-trial.root.hwx.site:7432	streams_messaging_manager	streams_messaging_manager
		Password	SFqcYP1nSH

Reports Manager

✓ Successful

Currently assigned to run on **ccycloud-1.streams-trial.root.hwx.site**.

Type	Database Hostname	Database Name	Username
PostgreSQL	ccycloud-1.streams-trial.root.hwx.site:7432	rman	rman
		Password	7Xj0AGPIQn

Schema Registry

✓ Skipped. Cloudera Manager will create this database in a later step.

Type	Database Hostname	Database Name	Username
PostgreSQL	ccycloud-1.streams-trial.root.hwx.site:7432	schemaregistry	schemaregistry
		Password	ObYCS5y60C

Test Connection

Back

Continue

The **Review Changes** section appears with default and suggested settings for several configuration parameters, including data directories.

15. Find the Cloudera Manager Service Monitor Host field for SMM.

Cloudera Manager Service
Monitor Host
cm.metrics.service.monitor.host

Streams Trial > Streams Messaging Manager Rest Admin Server Default Group



16. Enter the name of the Service Monitor host.

Cloudera Manager Service Monitor Host Streams Trial > Streams Messaging Manager Rest Admin Server Default Group ?

cm.metrics.service.monitor.host

[Undo](#)

ccycloud-2.streams-trial.root.hwx.site

17. Click Continue.

The **Command Details** section appears with the details of the First Run command.

18. After the First Run Command completes, click Continue.

Add Cluster - Configuration

First Run Command

Status ✔ Finished Context [Streams Trial](#) 📅 Oct 6, 10:55:06 AM ⌚ 2.7m

Finished First Run of the following services successfully: Schema Registry, ZooKeeper, Kafka, Cruise Control, Streams Messaging Manager, Cloudera Management Service.

✓ **Completed 1 of 1 step(s).**

☒ Show All Steps ☐ Show Only Failed Steps ☐ Show Only Running Steps

> ✔ Run a set of services for the first time	Oct 6, 10:55:06 AM	2.7m
--	--------------------	------

[Back](#) [Continue](#)

The **Summary** section appears with a success or failure report of the setup wizard.

Add Cluster - Configuration

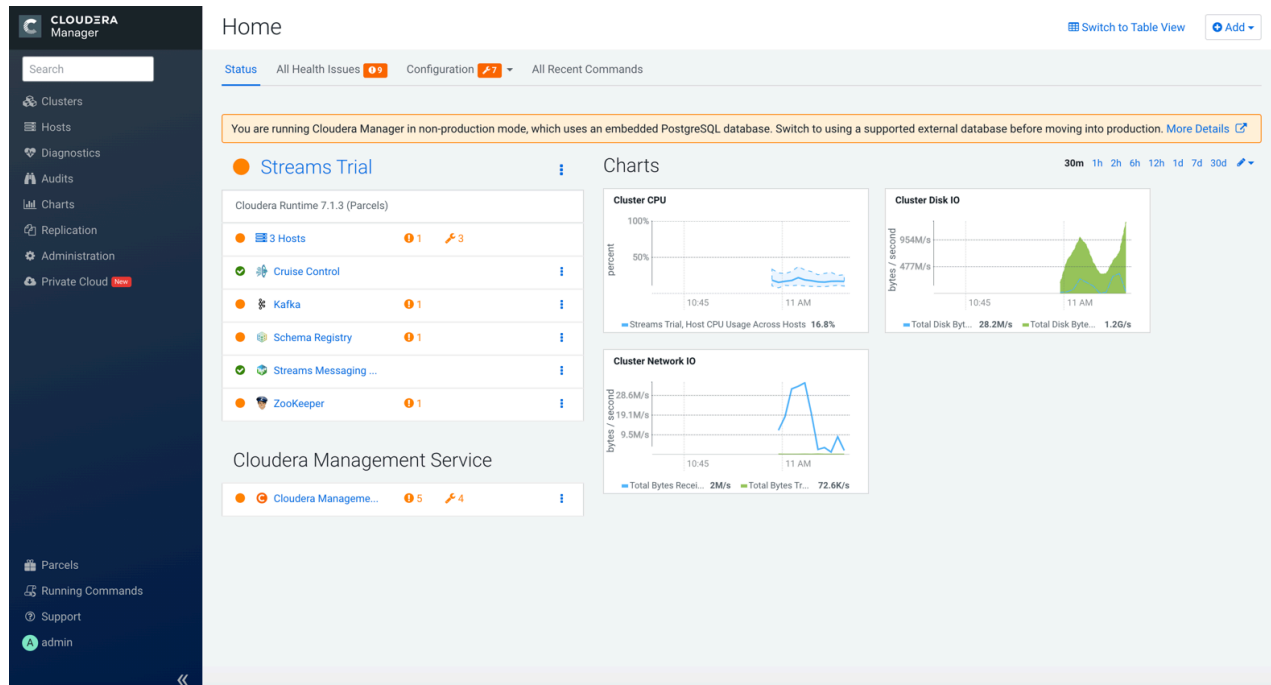
Summary

✔ The services are installed, configured, and running on your cluster.

19. Click Finish to complete the cluster setup.

Cloudera recommends that you change the default password as soon as possible by clicking the logged-in username at the top right of the home screen and clicking Change Password.

Results



What to do next

Perform simple streams-related tasks on your cluster to evaluate streaming services on CDP Private Cloud Base.

Getting Started on your Streams Cluster

Now that you have a streams cluster, you can evaluate the streaming services on CDP Private Cloud Base. To get started, you can create a Kafka topic, write events into it, and then read those events. Then use the SMM UI to monitor your cluster and view topic details.

Create a Kafka Topic to Store your Events

You must create a Kafka topic to store the events that you plan to stream. You can create a topic from the command line or the from the SMM UI.

About this task

Kafka is a distributed event streaming platform that lets you read, write, store, and process events (also called records or messages) across many machines.

Example events are payment transactions, geolocation updates from mobile phones, shipping orders, sensor measurements from IoT devices or medical equipment, and much more. These events are organized and stored in topics. Very simplified, a topic is similar to a folder in a filesystem, and the events are the files in that folder.

So before you can write your first events, you must create a topic.

Before you begin

- You have installed the trial version of CDP Private Cloud Base and setup the streaming cluster.

Procedure

1. To create a Kafka topic from the SMM UI:

- a) Go to the **Cloudera Manager** UI, click the **Streams Messaging Manager** service, and select the **Streams Messaging Manager Web UI**.
- b) Click the Topics icon on the left bar.
The **Topics** page appears.
- c) Click Add New.
- d) Provide the following information:
 - Topic name
 - Number of partitions
 - Level of availability
 - Cleanup policy. Cloudera recommends the delete option.
- e) SMM has automatically set Kafka topic configuration parameters. To manually adjust them, click Advanced.
- f) Click Save when done.

2. To create a Kafka topic from the command line:

- a) Open a terminal session and run:

```
$ kafka-topics --create --topic quickstart-events --bootstrap-server <hostname>:9092
--partitions 10 --replication-factor 3
```

- b) Run the kafka-topics.sh command without any arguments to display usage information.
- c) You can view details such as the partition count of the new topic:

```
$ bin/kafka-topics.sh --describe --topic quickstart-events --bootstrap-server <hostname>:9092
```

```
Topic:quickstart-events PartitionCount:1 ReplicationFactor:1 Configs:
Topic: quickstart-events Partition: 0 Leader: 0 Replicas: 0 Isr: 0
```

For more Kafka command-line tools, see *Kafka Command Line Tools*.

What to do next

Write a few events into the topic.

Related Information

[Kafka Command Line Tools](#)

Write a few Events into the Topic

After you create a topic, populate the topic with one or more events that you want to stream.

About this task

A Kafka client communicates with the Kafka brokers via the network for writing or reading events. Once received, the brokers will store the events in a durable and fault-tolerant manner for as long as you need.

Before you begin

- You have created a Kafka topic.

Procedure

1. Run the console producer client to write a few events into your topic. By default, each line you enter will result in a separate event being written to the topic.

```
$ kafka-console-producer --topic quickstart-events --broker-list <hostname>:9092
This is my first event
This is my second event
```

You can stop the producer client with Ctrl-C at any time.

2. Optionally, write more messages to the topic.

For more Kafka command-line tools, see *Kafka Command Line Tools*.

What to do next

Read the events.

Related Information

[Kafka Command Line Tools](#)

Read the Events

Consumers are client applications that subscribe to read and process events. You can simulate the subscription process by running the console consumer client to read the events you just created.

About this task

Events in a topic can be read as often as needed and by as many consumers as necessary. Events are not deleted after consumption.

Before you begin

- You have a topic with events in it.

Procedure

1. Open another terminal session and run the console consumer client to read the events you just created:

```
$ kafka-console-consumer --topic quickstart-events --from-beginning --broker-list <hostname>:9092
This is my first event
This is my second event
```

You can stop the consumer client with Ctrl-C at any time.

2. Feel free to experiment: for example, switch back to your producer terminal (previous step) to write additional events, and see how the events immediately show up in your consumer terminal.

For more Kafka command-line tools, see *Kafka Command Line Tools*.

What to do next

Monitor your cluster from the SMM UI.

Related Information

[Kafka Command Line Tools](#)

Monitor your Cluster from the SMM UI

Use the SMM UI to monitor your cluster. You can quickly check the number of producers, brokers, topics, and consumer groups on the Overview tab. From the Topics tab, you can view topic details such as the producers and consumers that are connected to the topic or the number of events that are written into the topic in a certain time frame.

Before you begin

- You have a topic with events in it.

Procedure

1. Go to the **Cloudera Manager** UI, click the **Streams Messaging Manager** service, and select the **Streams Messaging Manager Web UI**.
2. Review the information about your Kafka cluster on the **Overview** icon.

The **Overview** shows the total number of producers, brokers, topics, and consumer groups. It also provides more detailed metrics about producers and consumers.

Click the drop-down arrow in any of the boxes to view a list of Kafka resource. Select one or more Kafka resource to filter your view to just that resource. You can also search for a specific resource. You can click clear at any time to return to the unfiltered view.

3. From the left navigation pane, click the Topics icon.

The **Topic** page contains a number of useful details about your Kafka topics. This page helps you answer the following questions:

- How can I see if the replicas in this topic are in sync?
- How do I see this topic's retention rate?
- How can I see the replication factor for this topic?
- How do I see the producers and consumers that are connected to this topic?
- How do I find the total number of messages going into this topic, over a specified time range?

4. Select the topic you are interested in. You can either scroll through the list of topics, or use the Search bar.
5. Click the green hexagon at the left of the topic to view details.

To perform more tasks in SMM, review the following documents:

- *Monitoring Kafka Clusters using Streams Messaging Manager*
- *Managing Alert Policies using Streams Messaging Manager*
- *Managing Kafka Topics using Streams Messaging Manager*
- *Monitoring End-to-End Latency using Streams Messaging Manager*

Related Information

[Monitoring Kafka Clusters using Streams Messaging Manager](#)

[Managing Alert Policies using Streams Messaging Manager](#)

[Managing Kafka Topics using Streams Messaging Manager](#)

[Monitoring End-to-End Latency using Streams Messaging Manager](#)

After Evaluating Trial Software

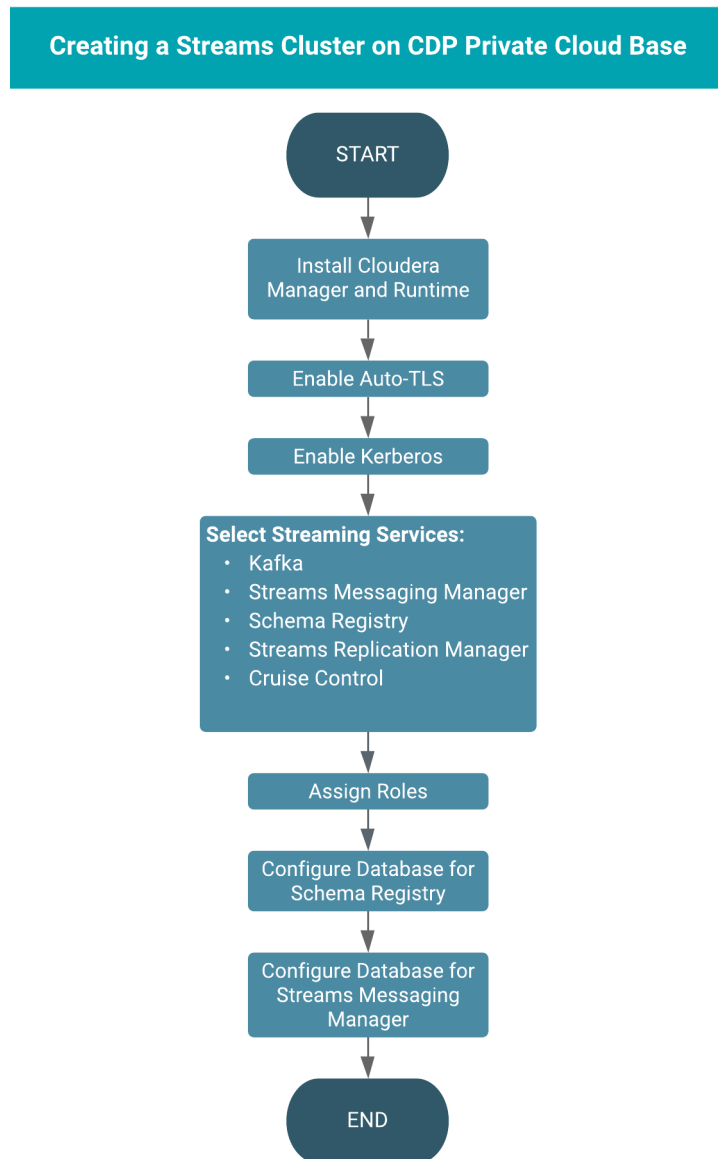
While you use and evaluate CDP Private Cloud Base, you can learn more about the streaming components from our documentation. After evaluation, you can contact Cloudera to request the appropriate license for your production environment.

To obtain a licence for CDP Private Cloud Base for your production environment, fill in the [Contact Us](#) form.

To understand more about the Cloudera Data Platform Runtime streaming components, see the following documentation:

- [Apache Kafka Overview](#)
- [Cruise Control Overview](#)
- [Schema Registry Overview](#)
- [Streams Messaging Manager Overview](#)
- [Streams Replication Manager Overview](#)

To install the production software, follow the installation instructions in the *CDP Private Cloud Base Installation Guide*. The following diagram shows the main steps involved in a standard production installation:



Related Information

[CDP Private Cloud Base Production Installation Guide](#)

[Apache Kafka Overview](#)

[Cruise Control Overview](#)

[Schema Registry Overview](#)

[Streams Messaging Manager Overview](#)

[Streams Replication Manager Overview](#)
[Contact Cloudera](#)