7.1.5

Quick Start Deployment: Streaming Cluster in CDP Private Cloud Base

Date published: 2020-10-09
Date modified: 2020-11-25

CLOUDERA

https://docs.cloudera.com/

Legal Notice

Contents

Create a Streams Cluster on CDP Private Cloud Base

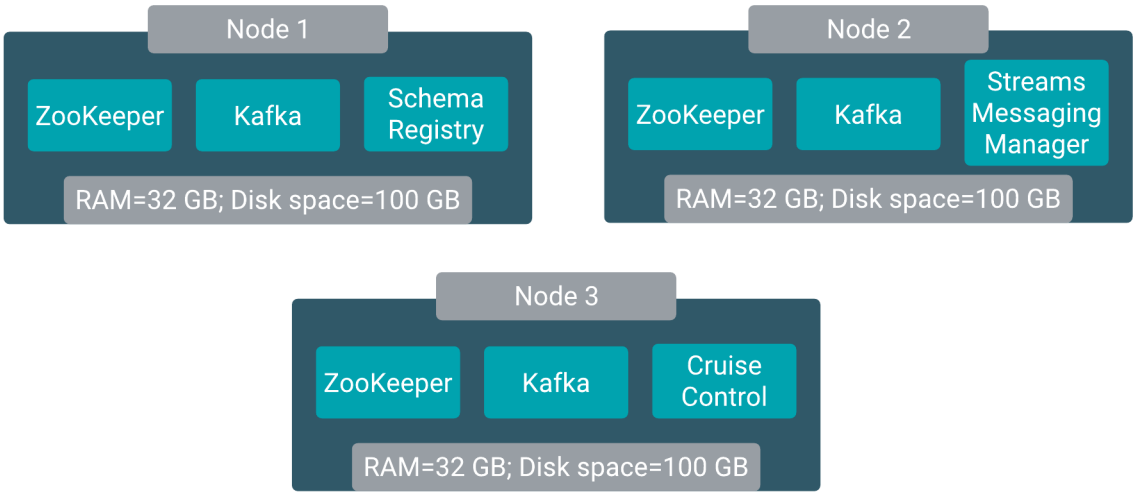
Cloudera's streaming components empower enterprises to handle some of the most complex and sophisticated streaming use cases. You can evaluate the streaming components in CDP Private Cloud Base for up to 60 days. This document walks you through the process of installing the trial software and creating 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

Ensure you are using a Cloudera-supported operating system and version.

Table 1: Supported operating systems for 7.1.5

Operating System	Version (bold=new)
RHEL/CentOS/Oracle	7.6, 7.7, 7.8, 7.9
SLES	12 SP5 Not supported when using the Using the Trial Installer (cloudera-manager-installer.bin) to install Cloudera Manager.
IBM PowerPC on RHEL	Supported on RHEL 7.6, 7.7. The following components are not supported: • Impala • Kudu • Ozone • Navigator Encrypt  **Note:** Ranger KMS is the recommended Key Management Server for PowerPC deployments.
Ubuntu	18.04
Debian	Not Supported

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	/etc/yum.conf	proxy=http://*server*:*port*/
Ubuntu	/etc/apt/apt.conf	Acquire::http::Proxy "http://*server*:*port*";

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
   ```

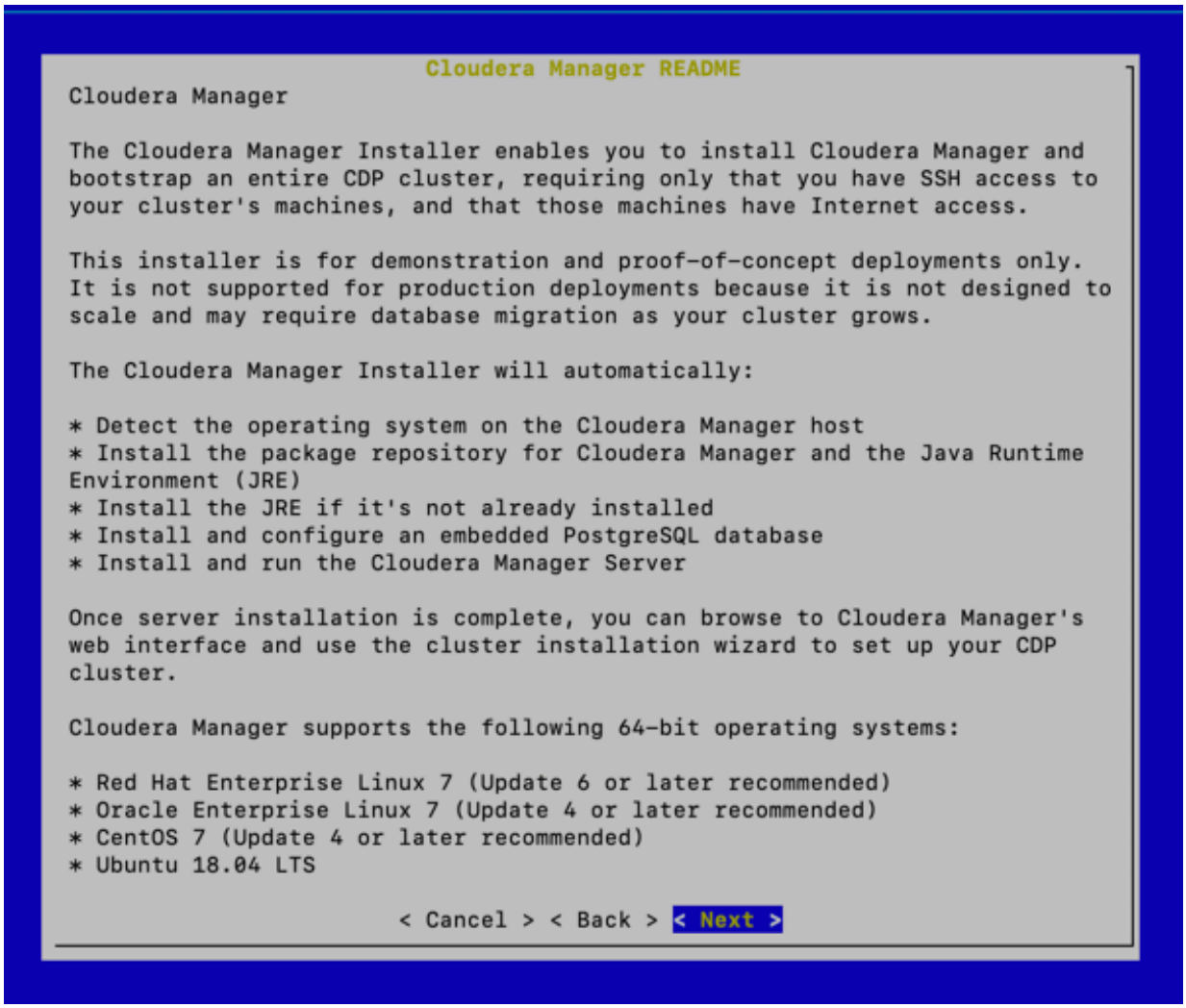
   b) Run the Cloudera Manager Server installer:

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

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

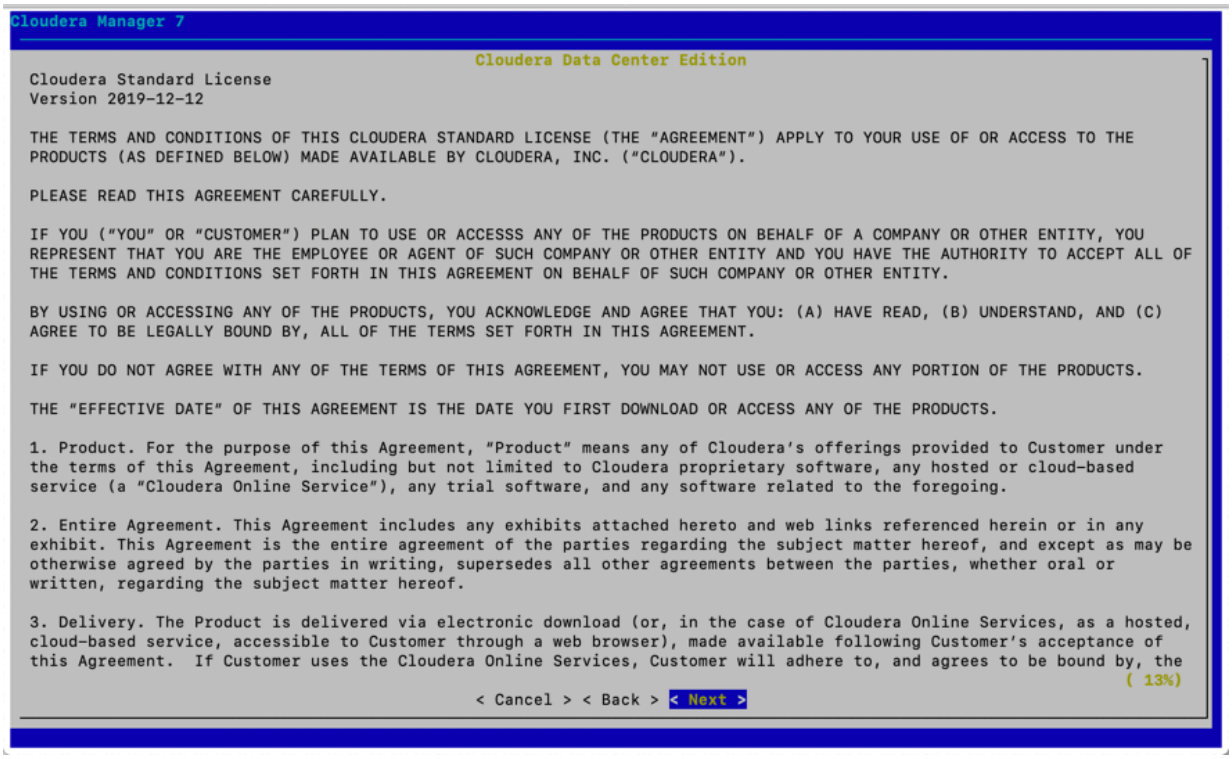
   ```
   sudo ./cloudera-manager-installer.bin --skip_repo_package=1
   ```

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



2. Click Next.

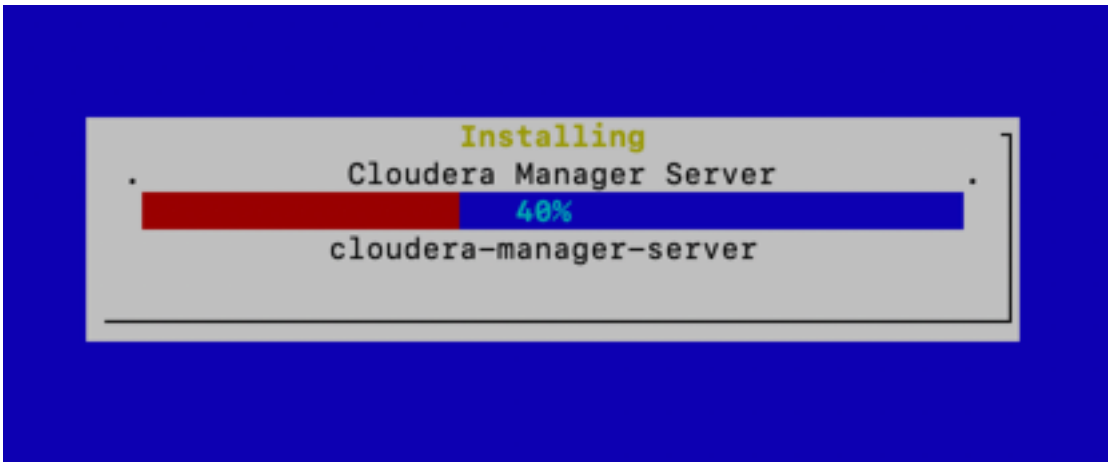
   The **Cloudera Standard License** page appears.



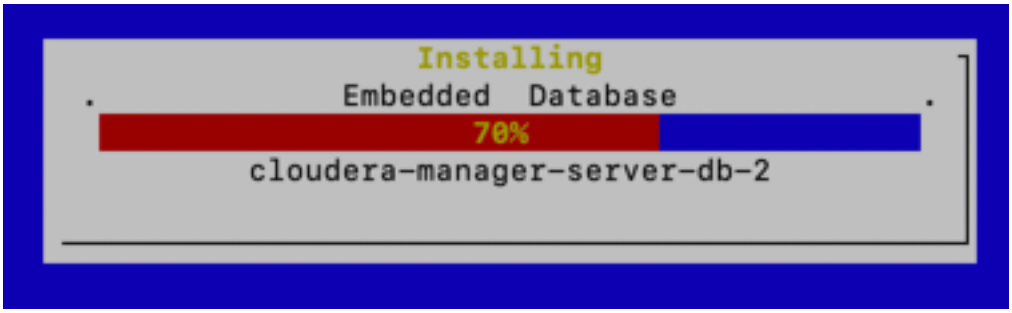
3. Click Next to accept the license agreement.
   The the installer starts and does the following:

   a. Installs Oracle JDK.

   b. Installs the Cloudera Manager Server.



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



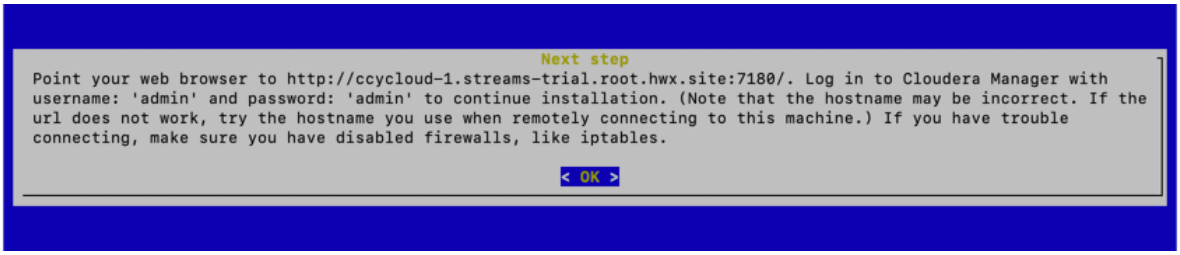
Note: If the installation is interrupted, run the following command on the Cloudera Manager Server host before you retry the installation:

```
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.

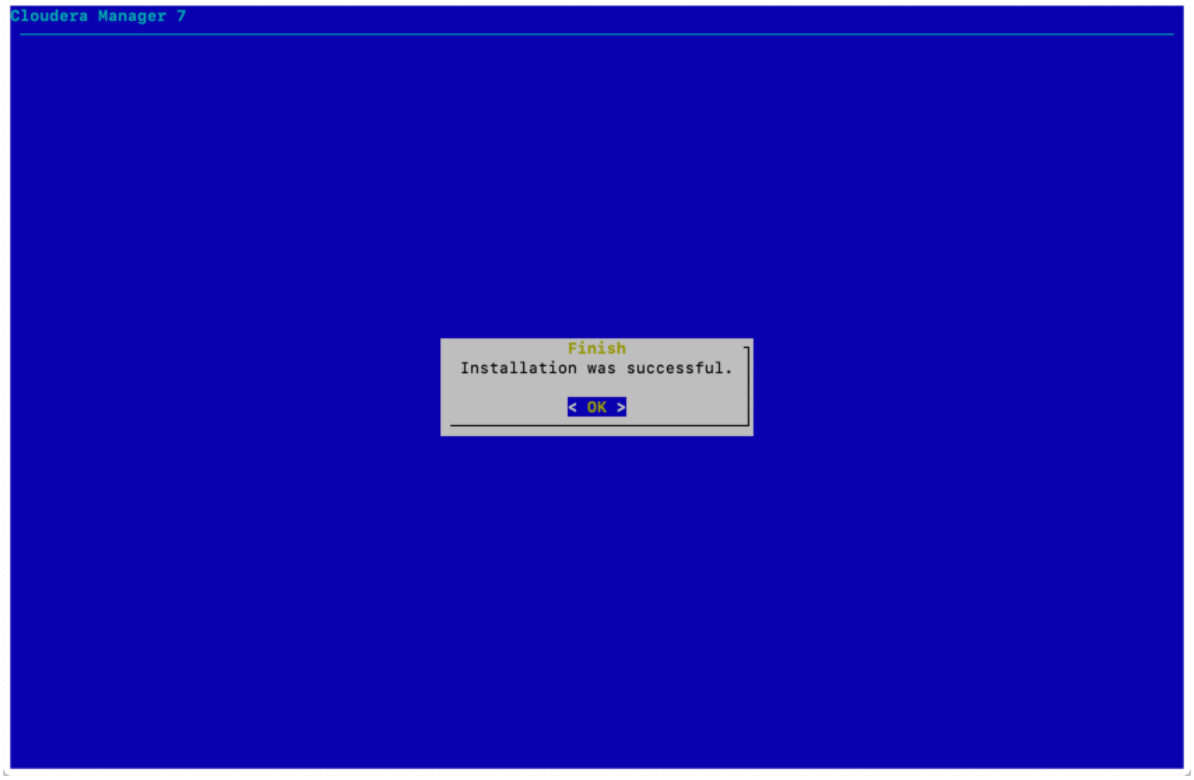
   Important: Make a note of this URL or take a screen capture as you will need it for the next task.

   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

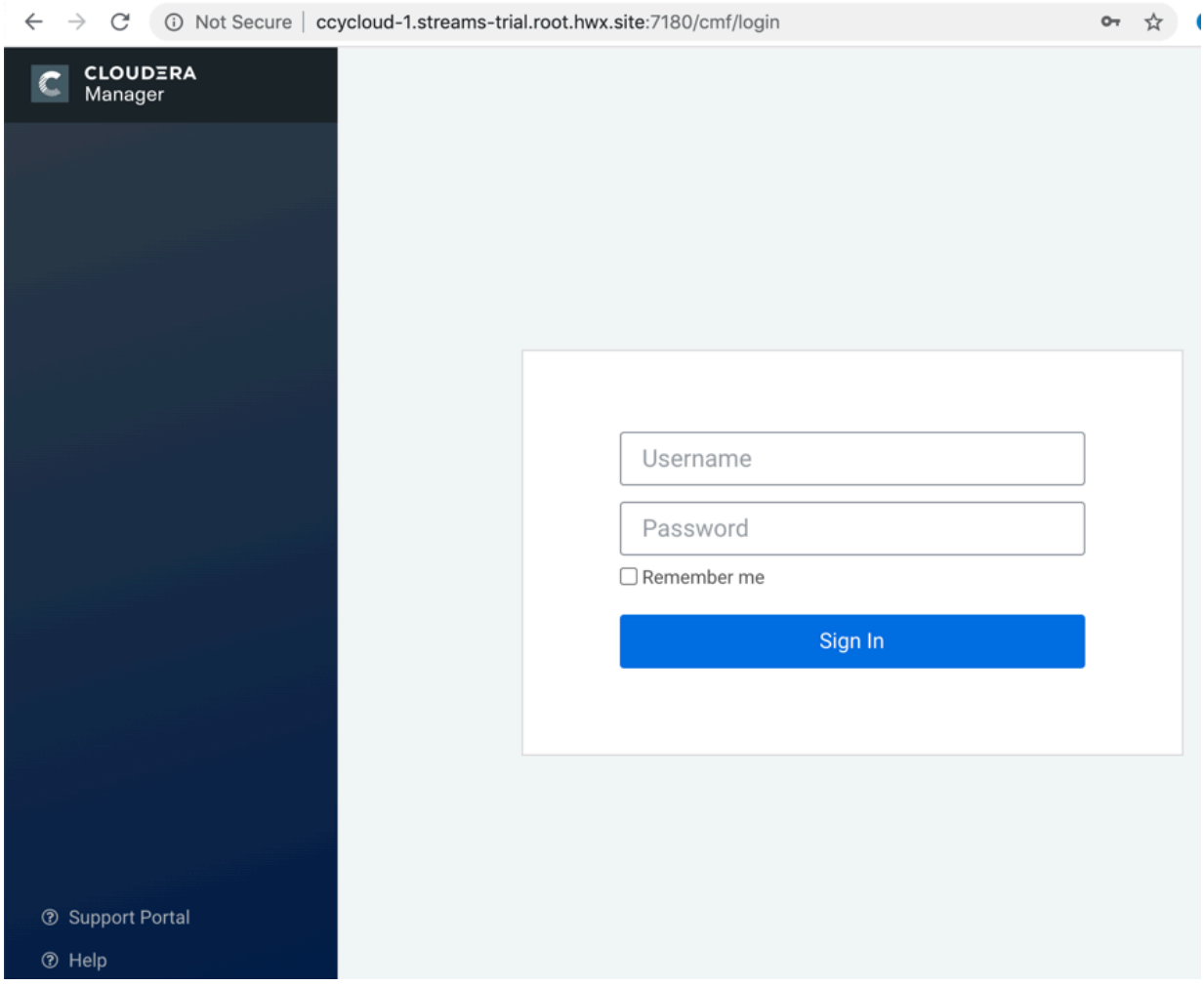
- You have installed Cloudera Manager.

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.

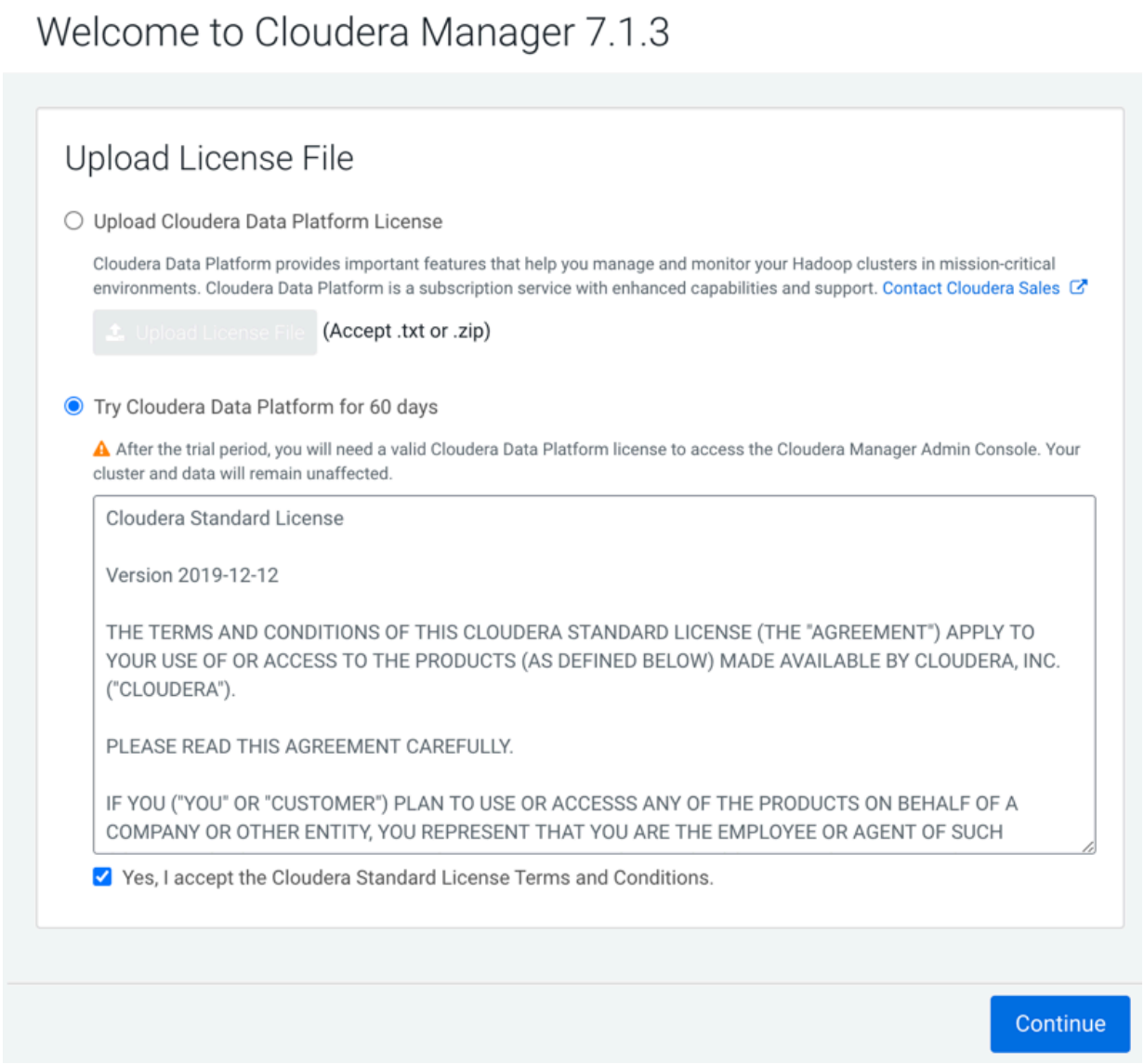


2. Sign in with the default credentials:
   - Username: admin
   - Password: admin

   Click Sign In.

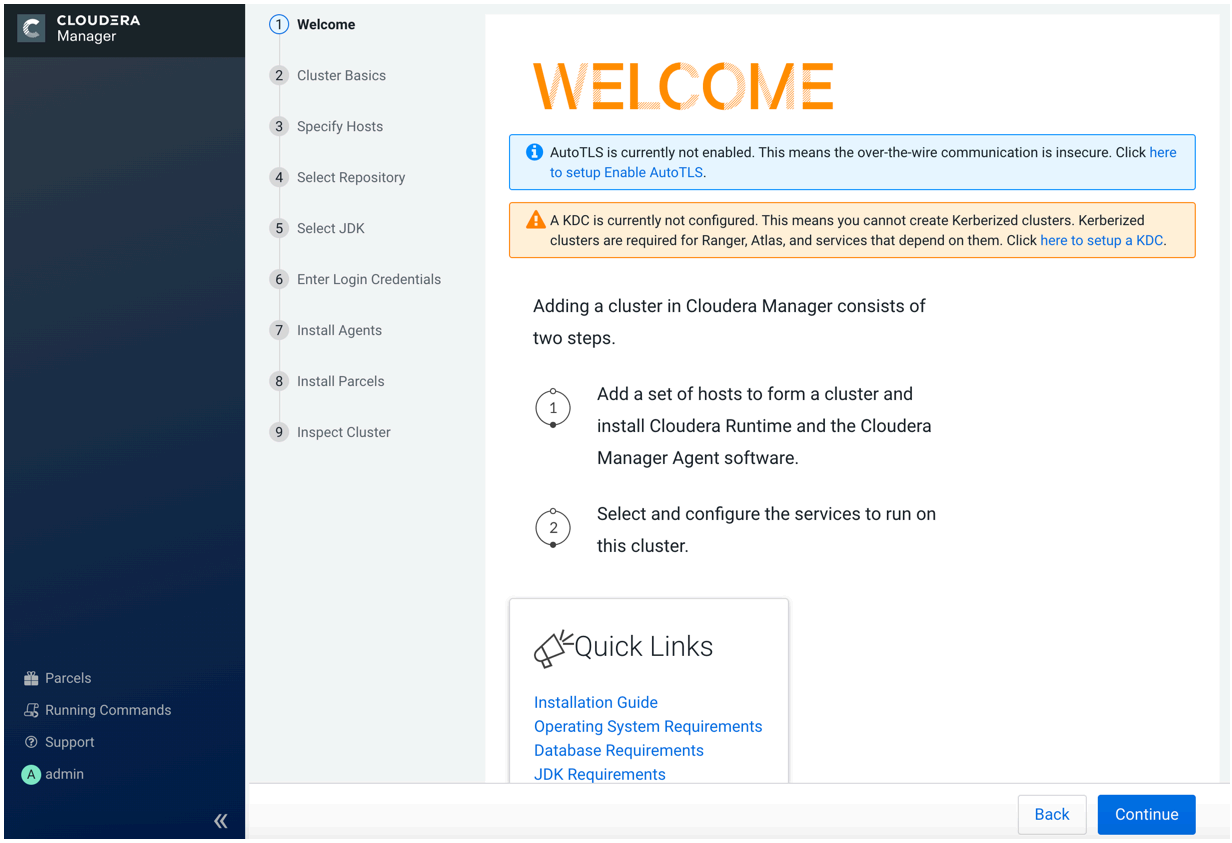
The **Welcome to Cloudera Manager** page appears.

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



4. Click Continue.

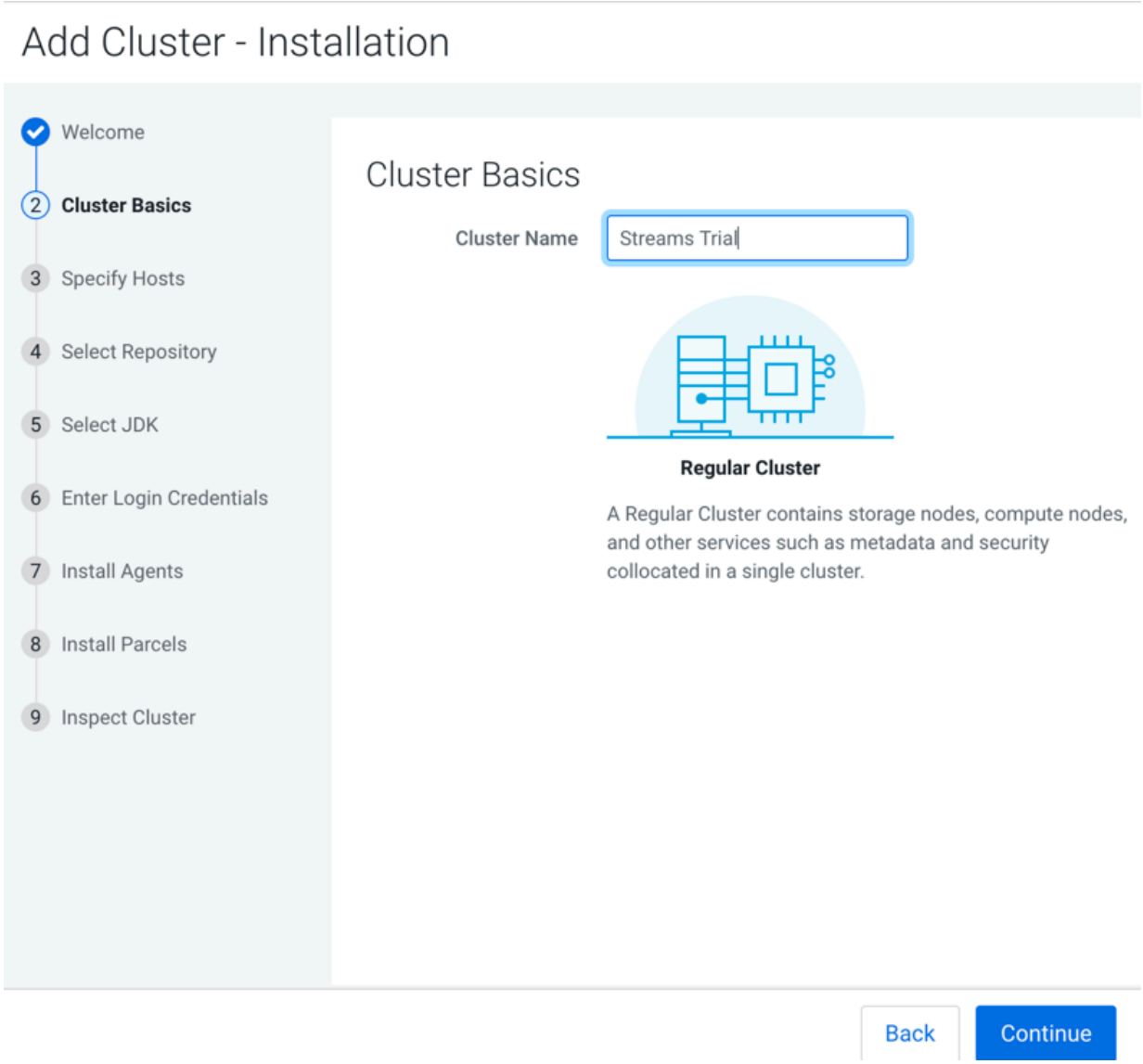
   The **Add Cluster - Installation** page, **Welcome** section appears. The steps on the left let you know where you are in the workflow.



5. Click Continue.

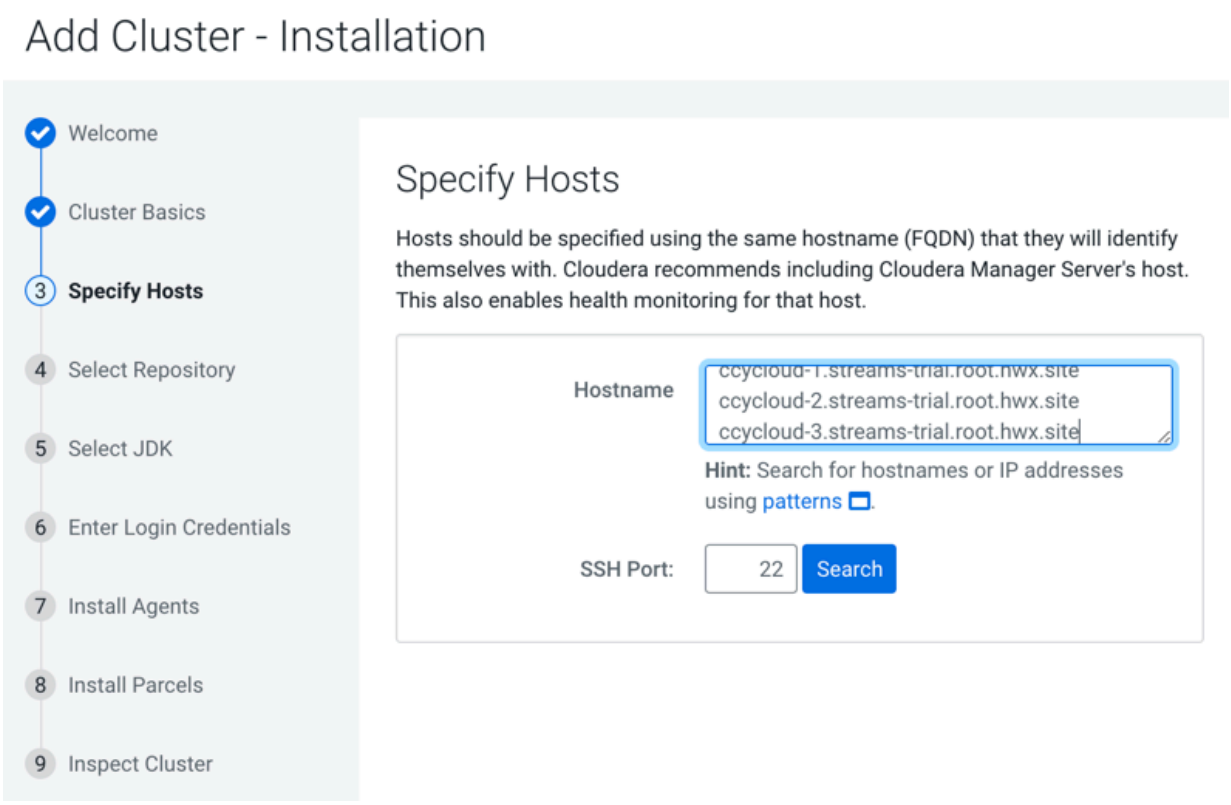
   The Cluster Basics section appears.

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



The **Specify Hosts** section appears.

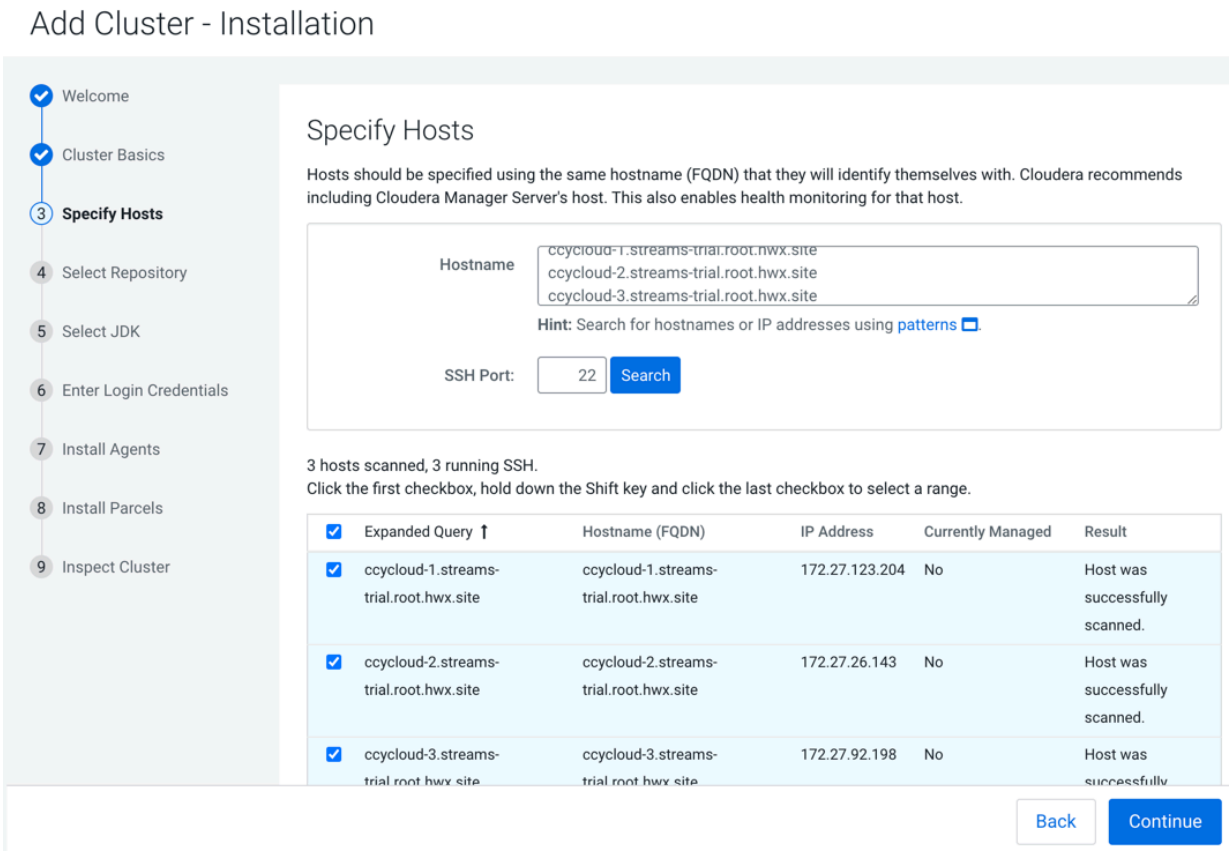
7. Enter the cluster host names or IP addresses in the Hostnames field.



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

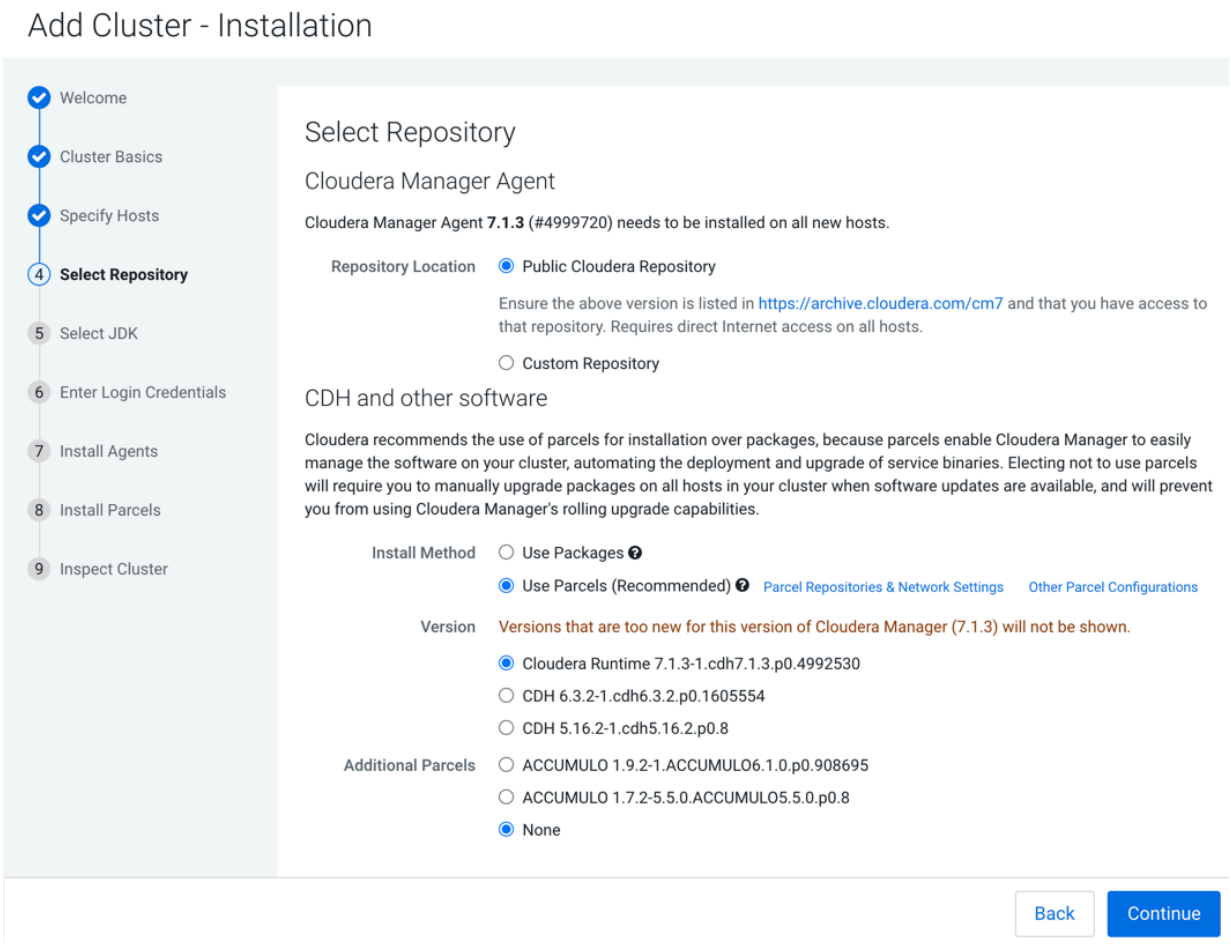
8. Click Search.
   Cloudera Manager discovers the hosts.



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:

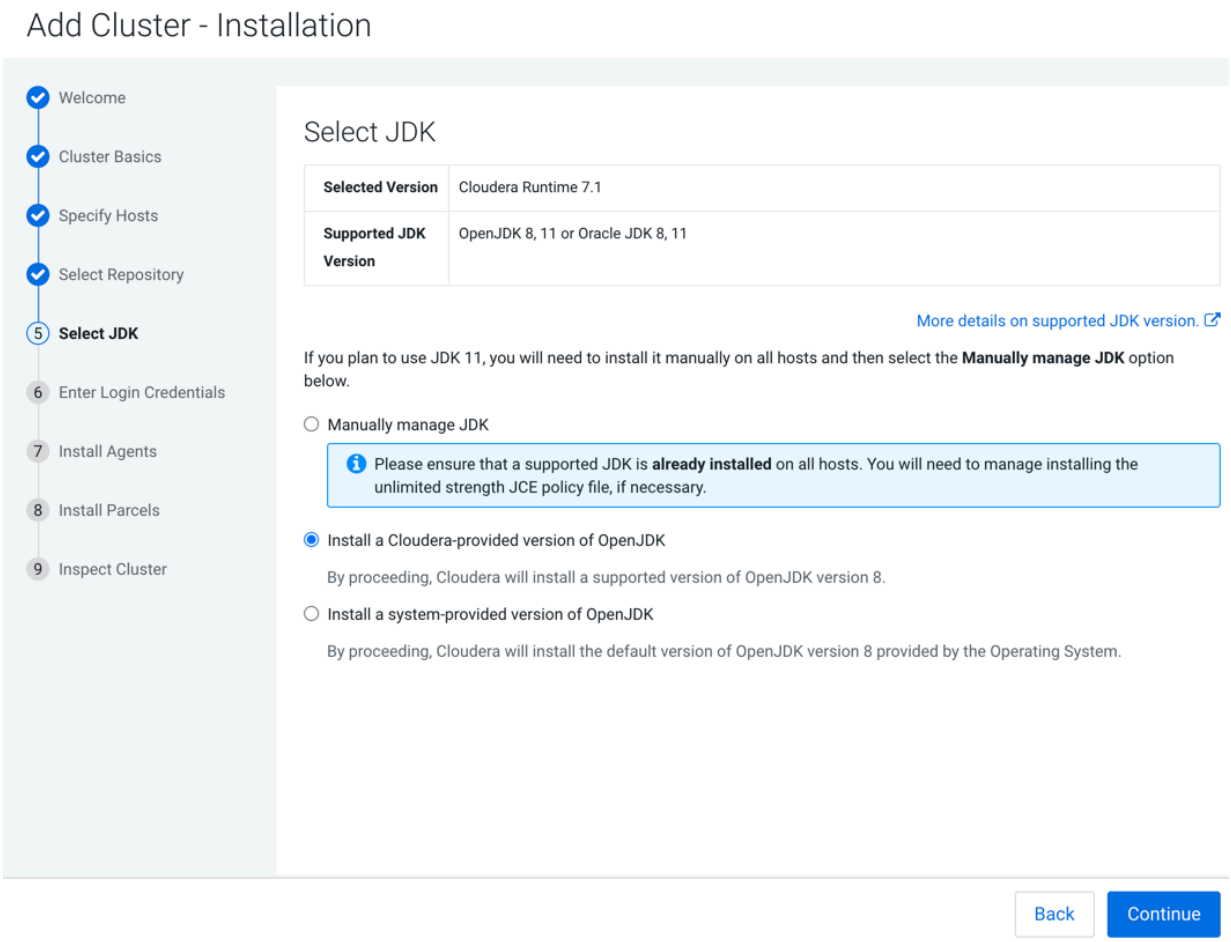
- Public Cloudera Repository
- Use Parcels
- The version of Cloudera Runtime that you want to install.
- In the Additional Parcels section, None.



11. Click Continue.

The **Select JDK** section appears.

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

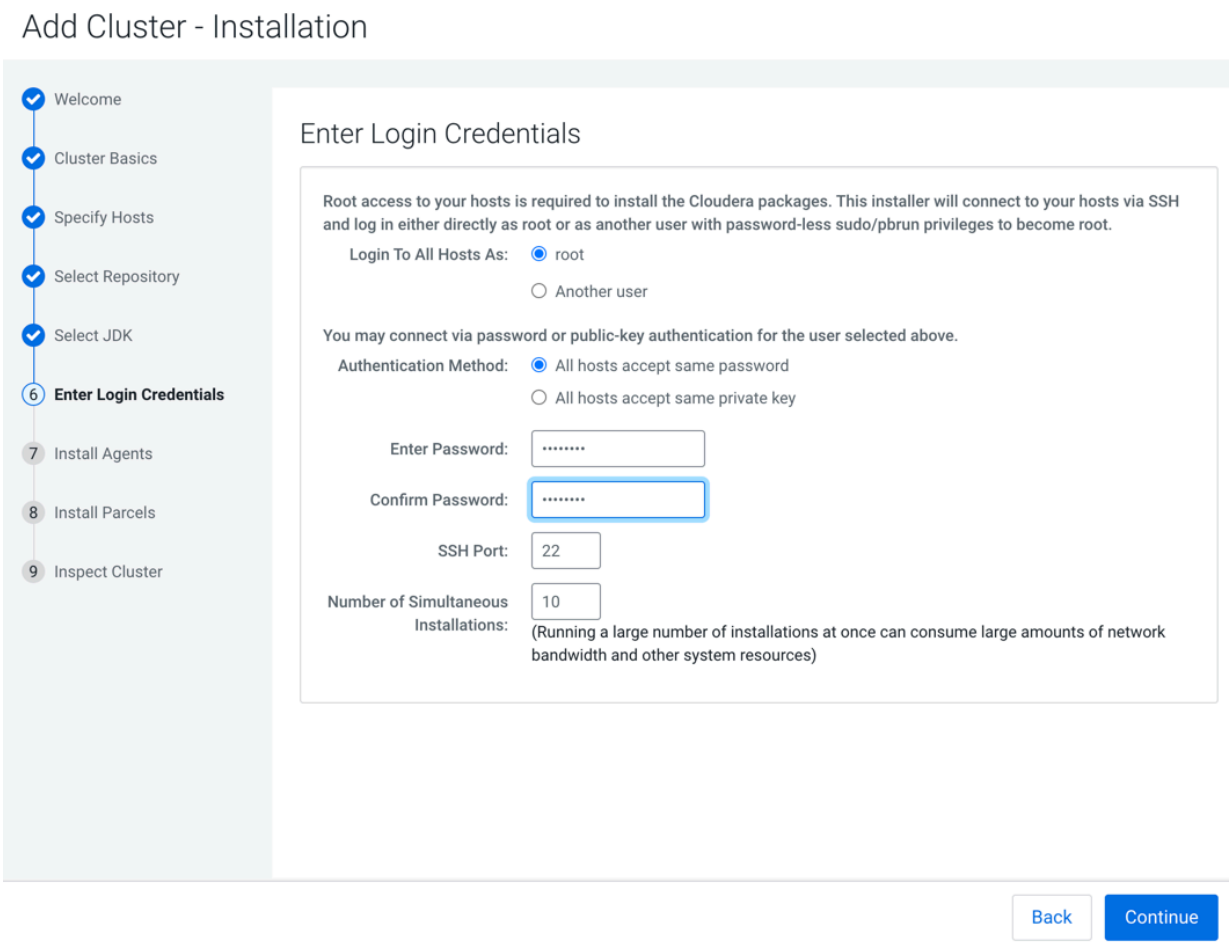


13. Click **Continue**.

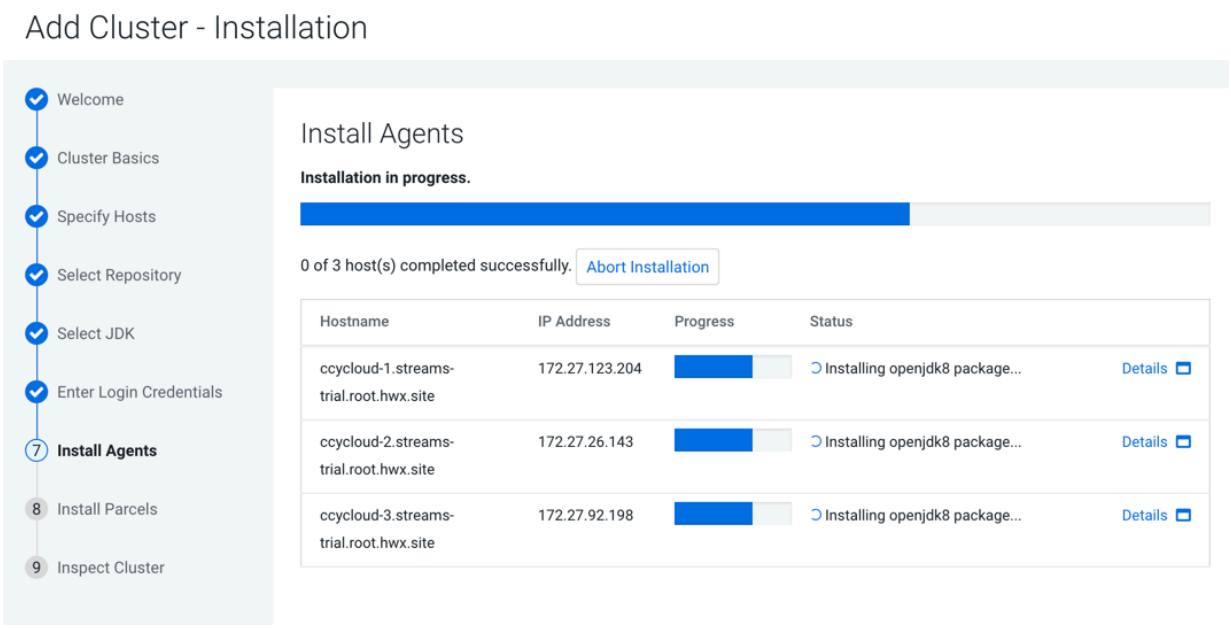
The **Enter Login Credentials** section appears.

14. Do the following:

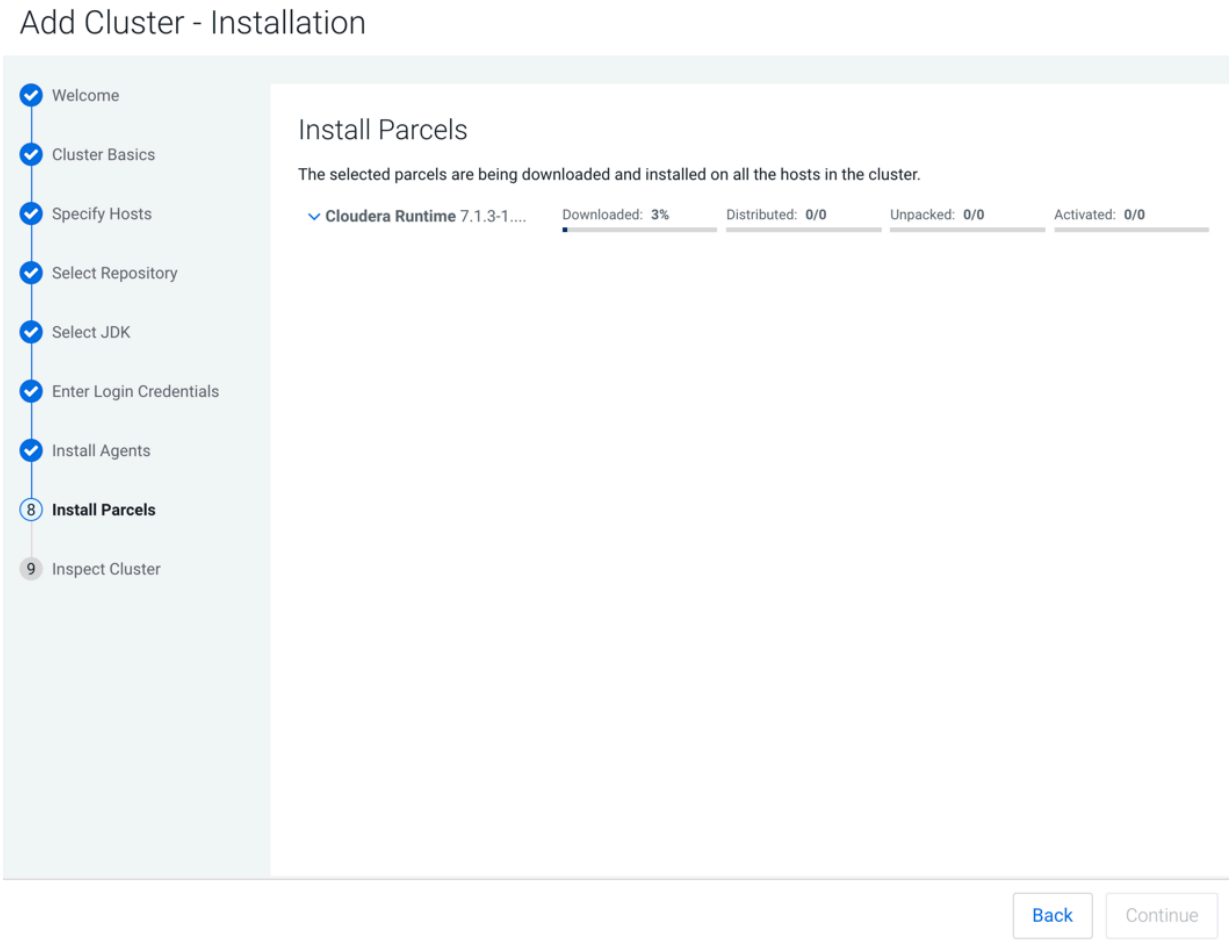
- Select root.
- Select All hosts accept same password.
- Enter the password for the account that allows root access to your hosts.
- Click Continue.



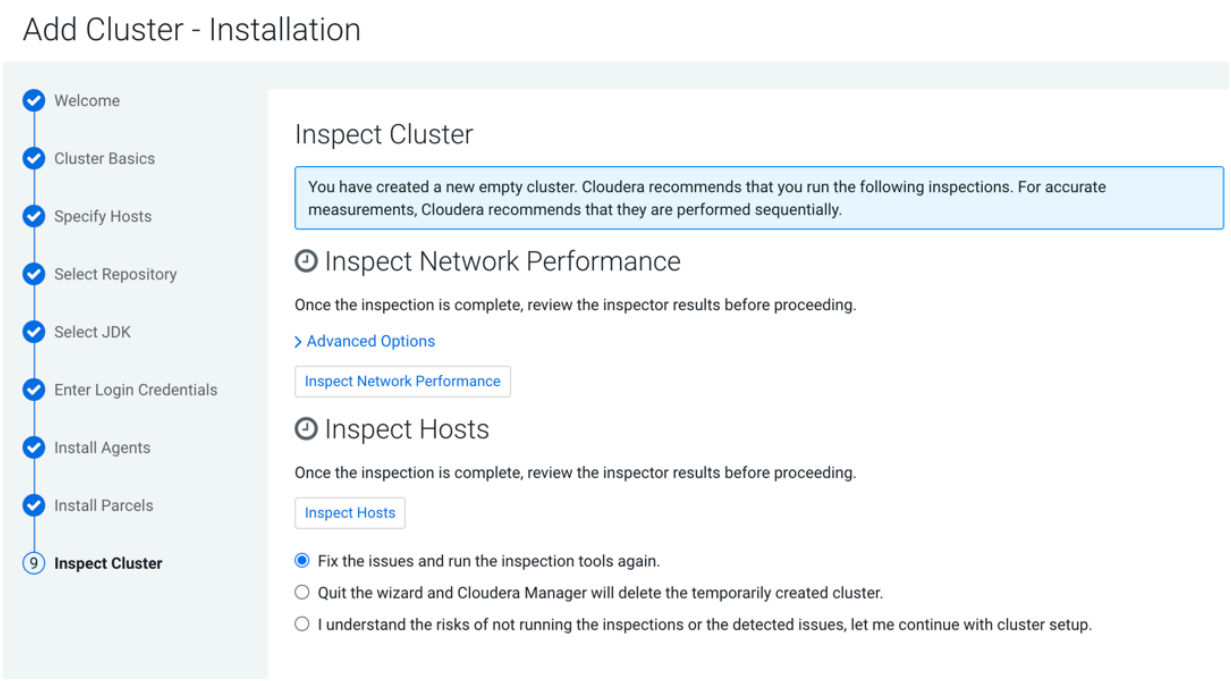
The **Install Agents** section appears showing the progress of the installation.



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



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


Add Cluster - Installation
Welcome
Cluster Basics
Specify Hosts
Select Repository
Select JDK
Enter Login Credentials
Install Agents
Install Parcels
9 Inspect Cluster
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.

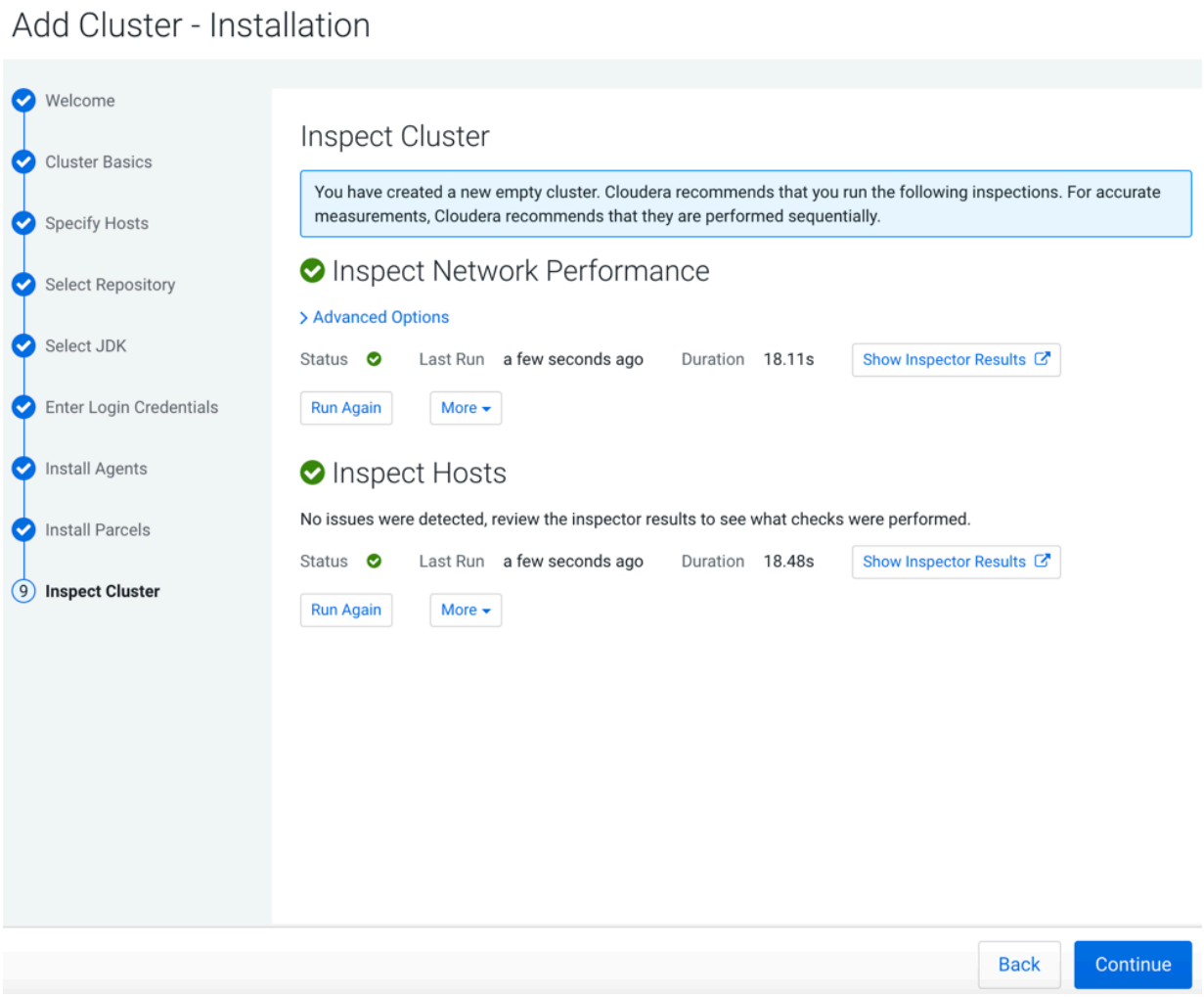
b) After the network inspector completes, click Show Inspector Results to view the results in a new tab.

Address any reported issues, and click Run Again.

c) Click Inspect Hosts.

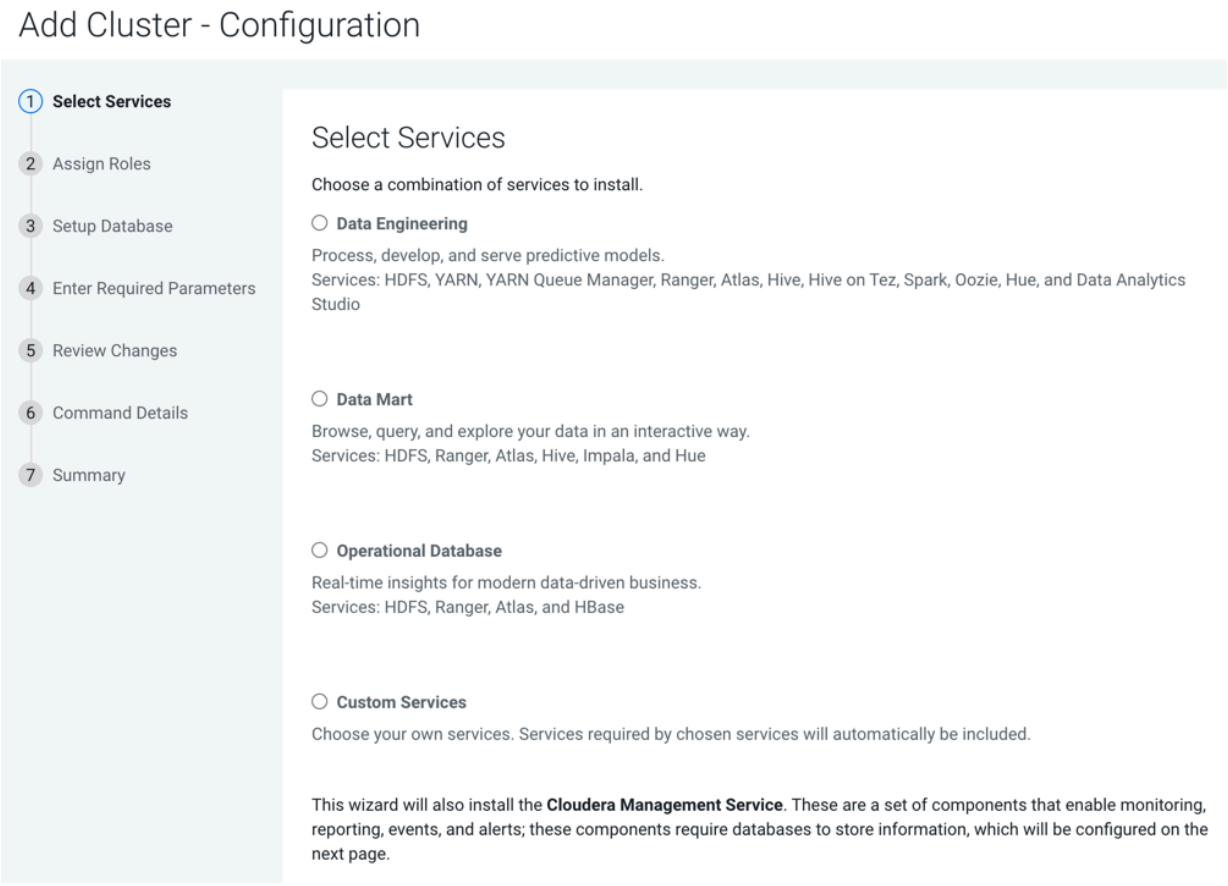
d) After the host inspector completes, click Show Inspector Results to view the results in a new tab.

Address any reported issues, and click Run Again.



16. Click Continue.

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



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.

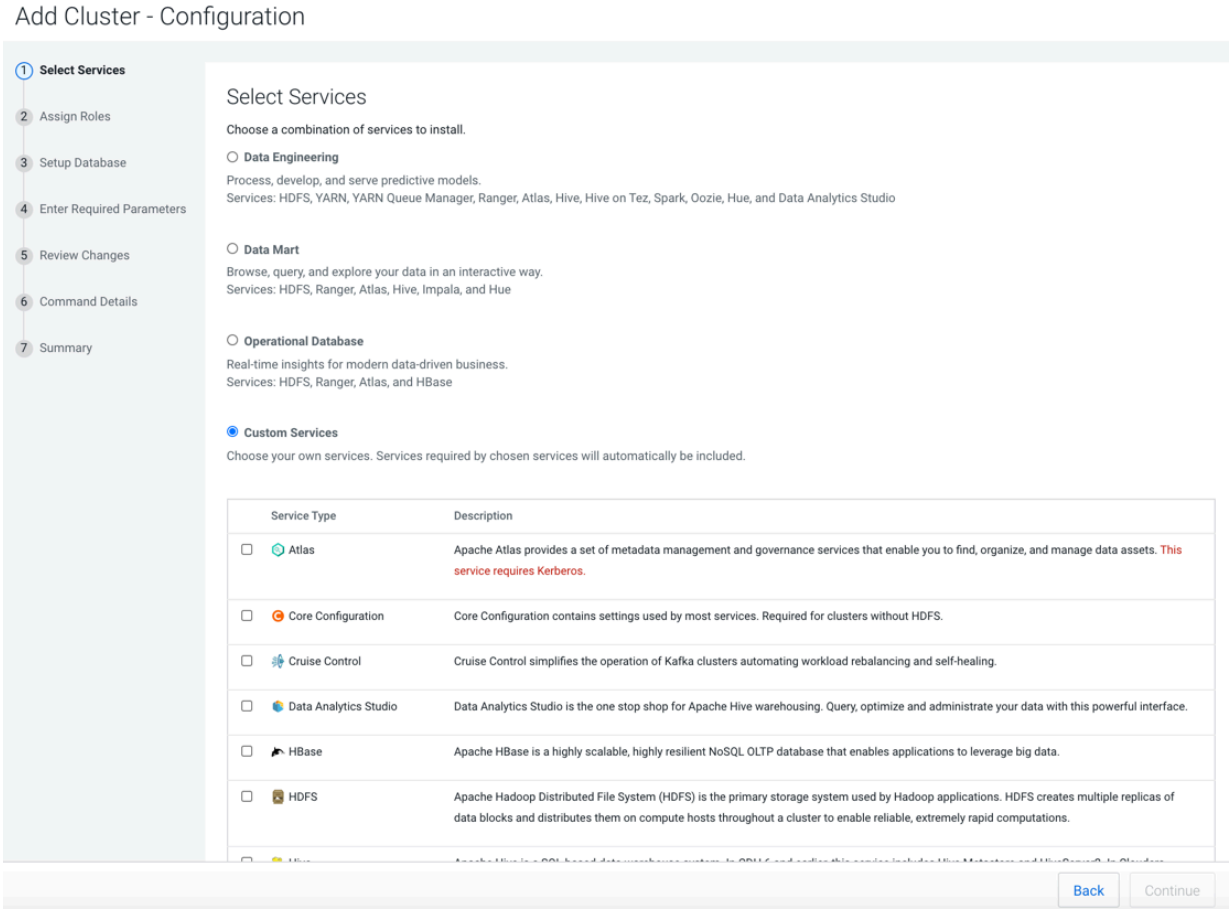
Procedure

1. Verify you are on the **Add Cluster - Configuration** page pf 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.

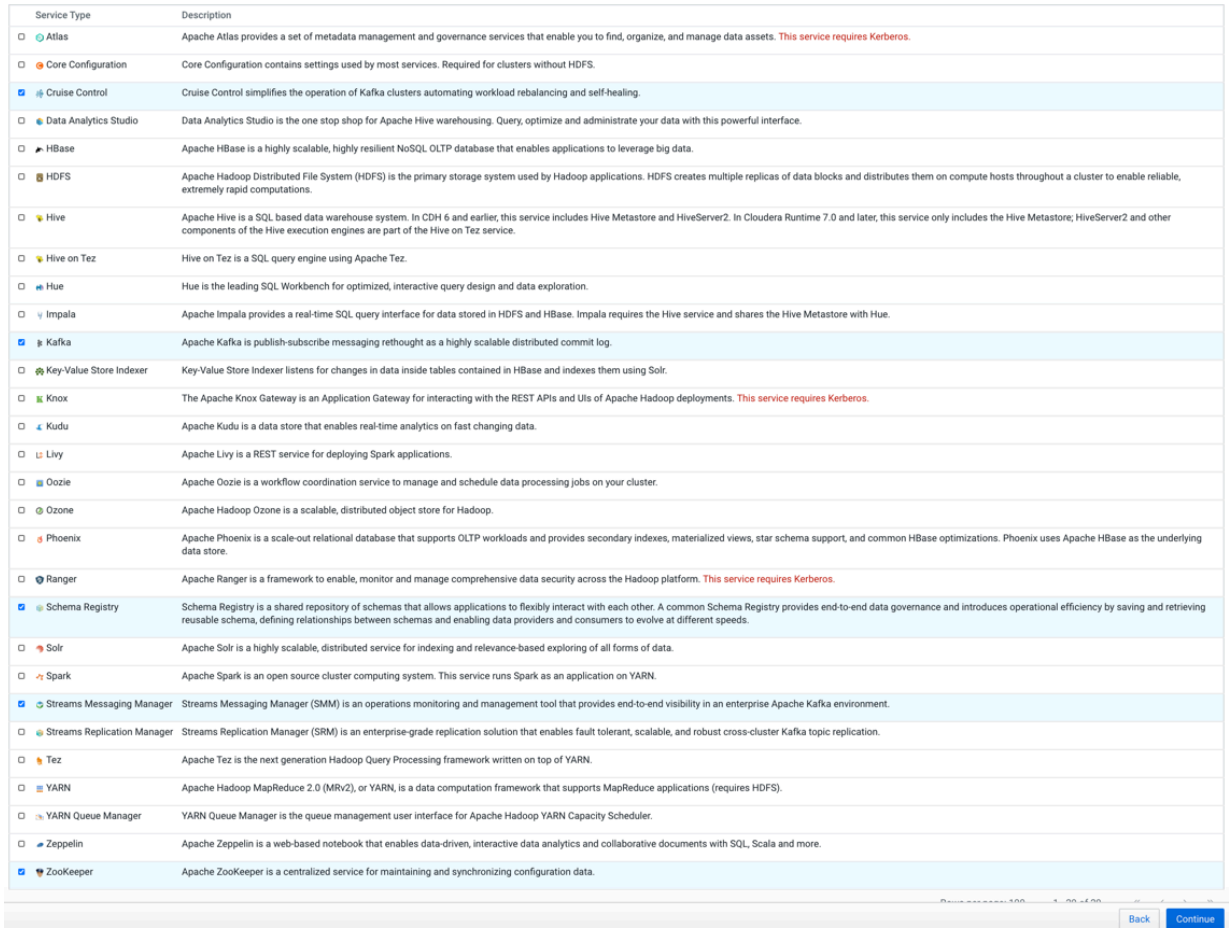
3. Select the Custom Services option.

A list of services appear.

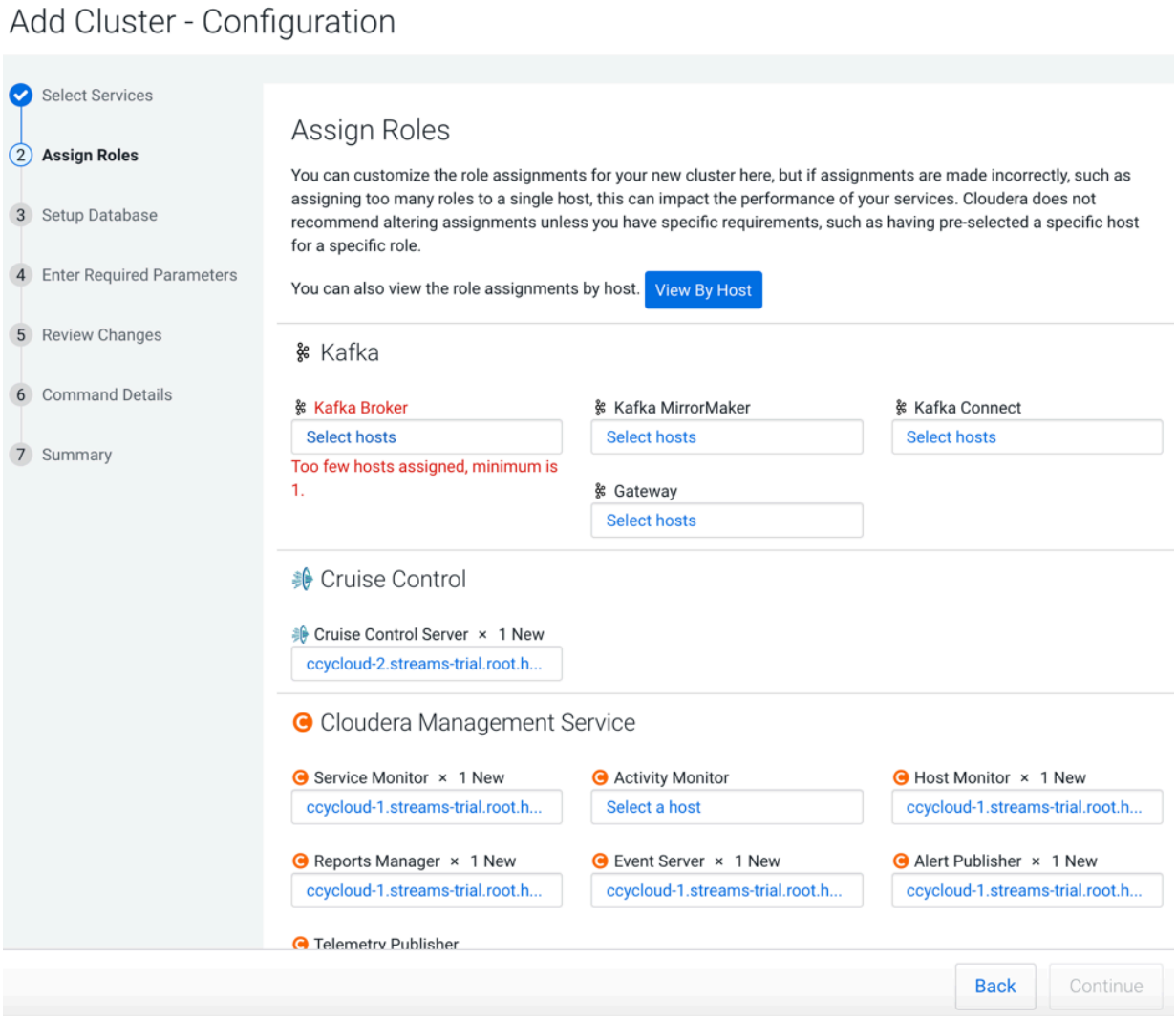


4. Scroll through the list and select the following services:

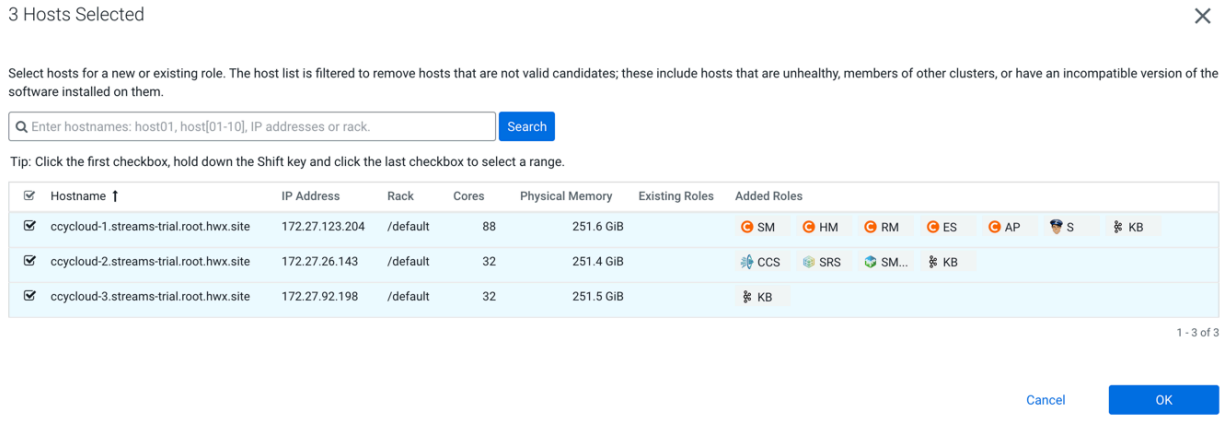
   - Cruise Control
   - Kafka
   - Schema Registry
   - Streams Messaging Manager
   - ZooKeeper



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



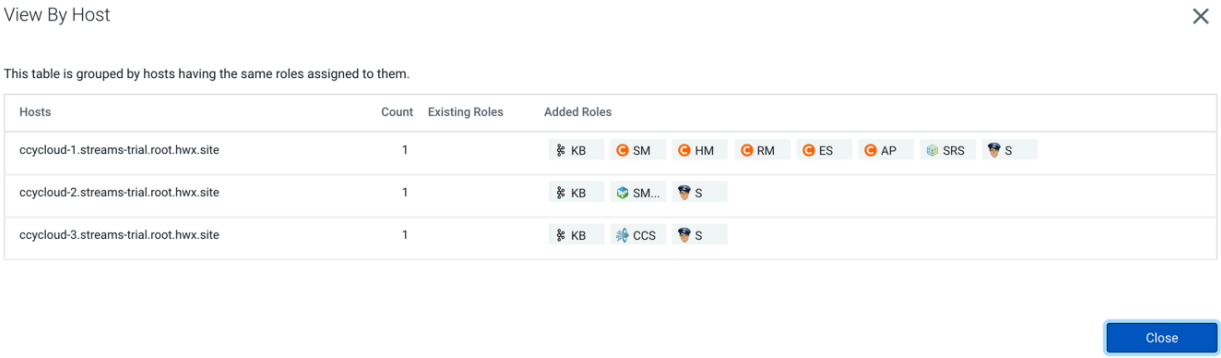
6. In the Kafka Broker field, click Select hosts.
7. Select all hosts for Kafka Broker and click OK.



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.

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

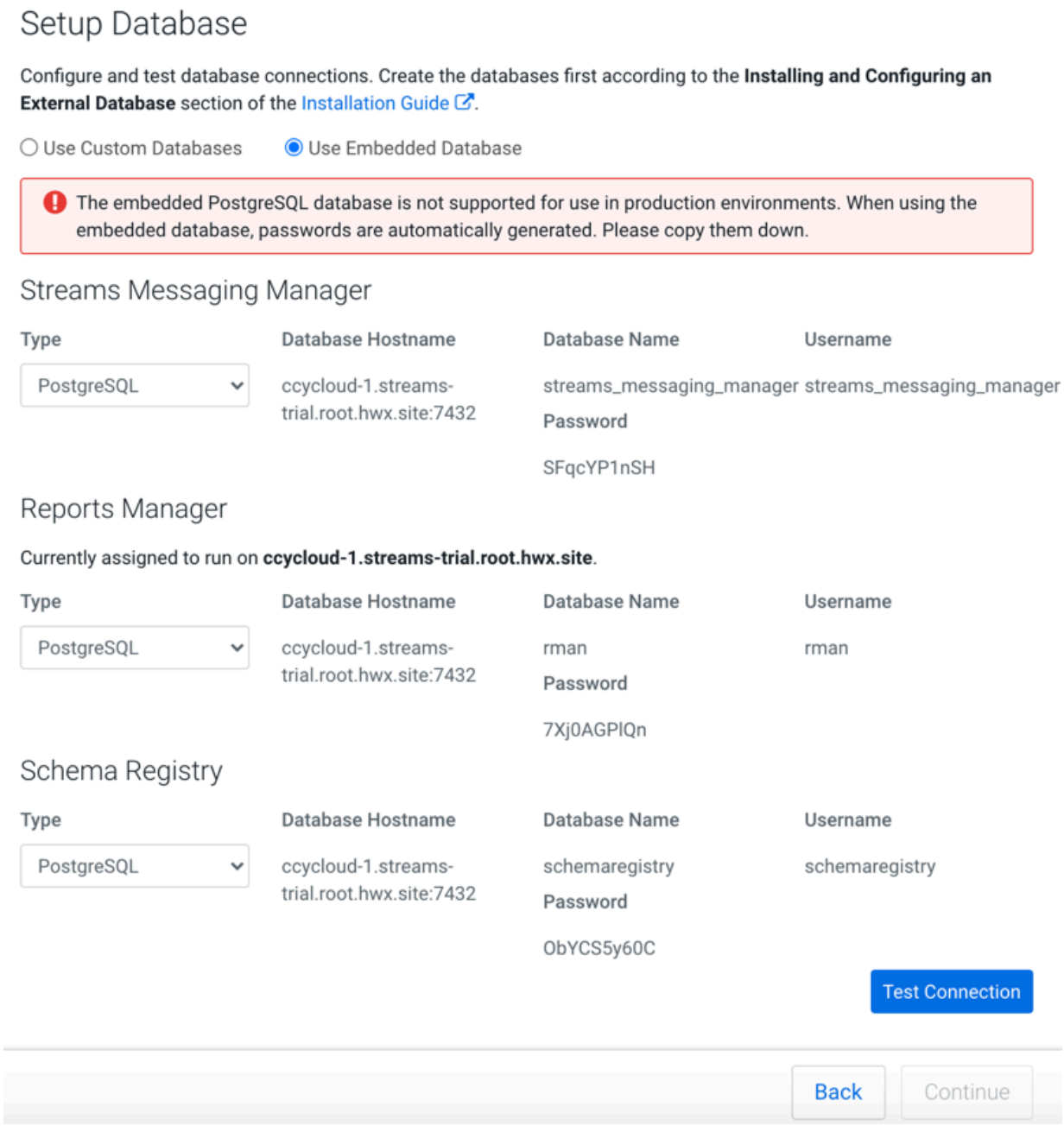
The **View By Host** window appears.



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

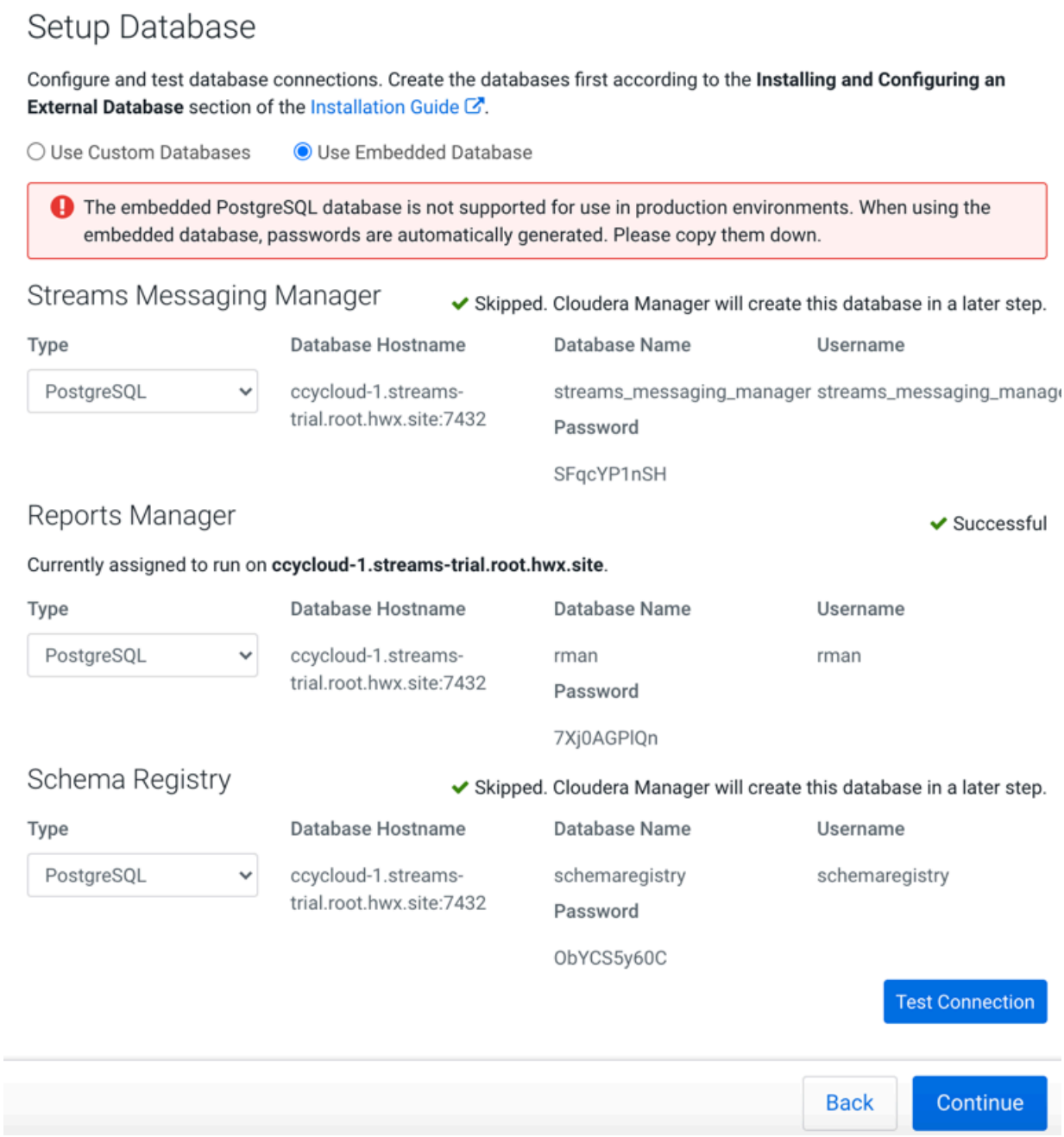
12. Back on the **Assign Roles** section, click Continue.

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



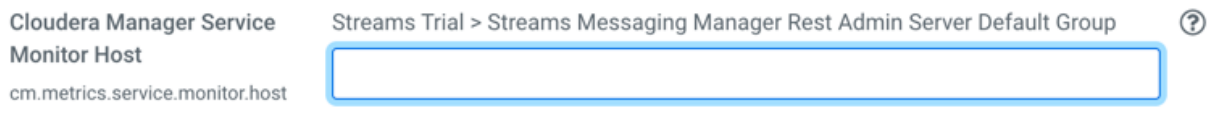
13. Click Test Connection to validate the settings.

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

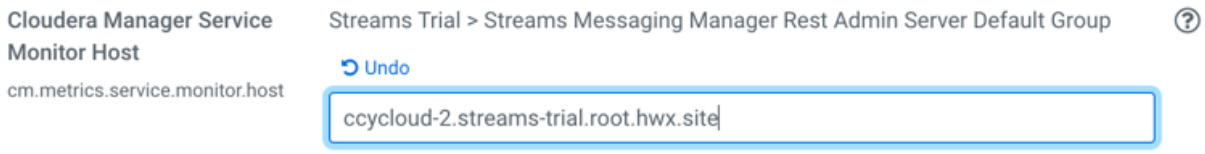


The **Review Changes** section appears with default and suggested settings for several configuration parameters, including data directories.

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



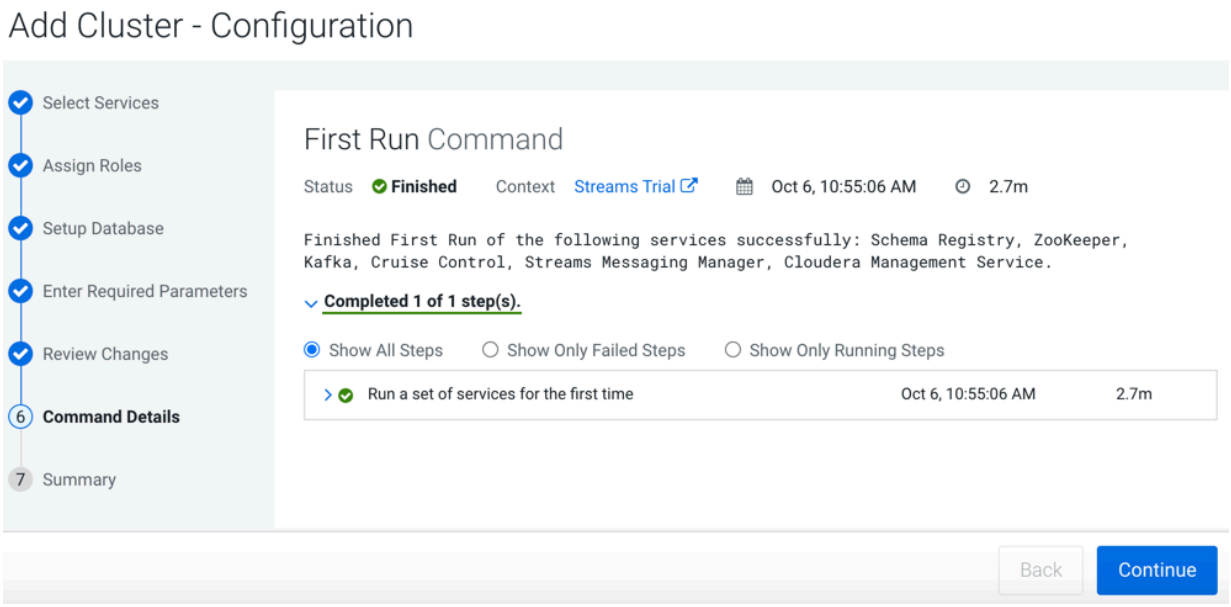
16. Enter the name of the Service Monitor host.



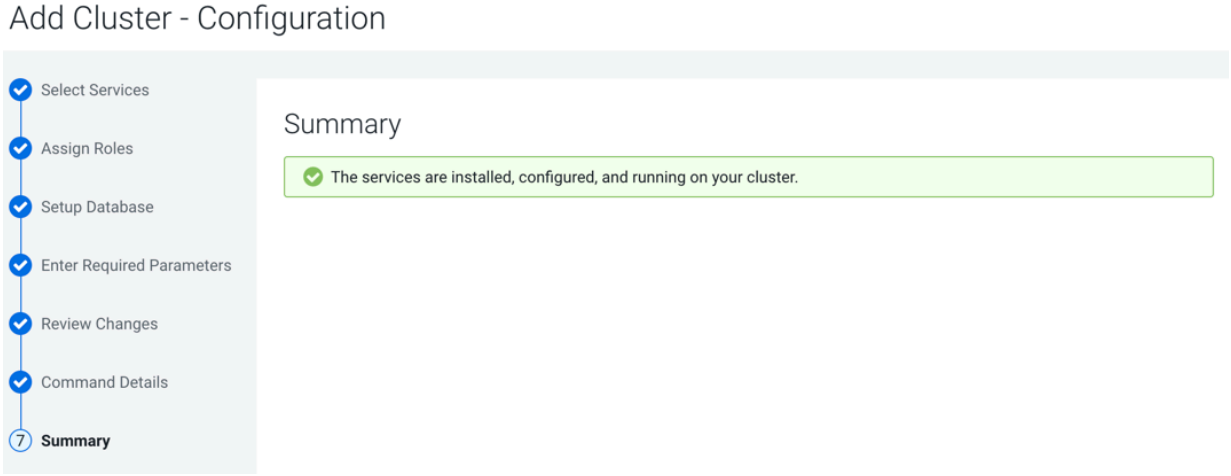
17. Click Continue.

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

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



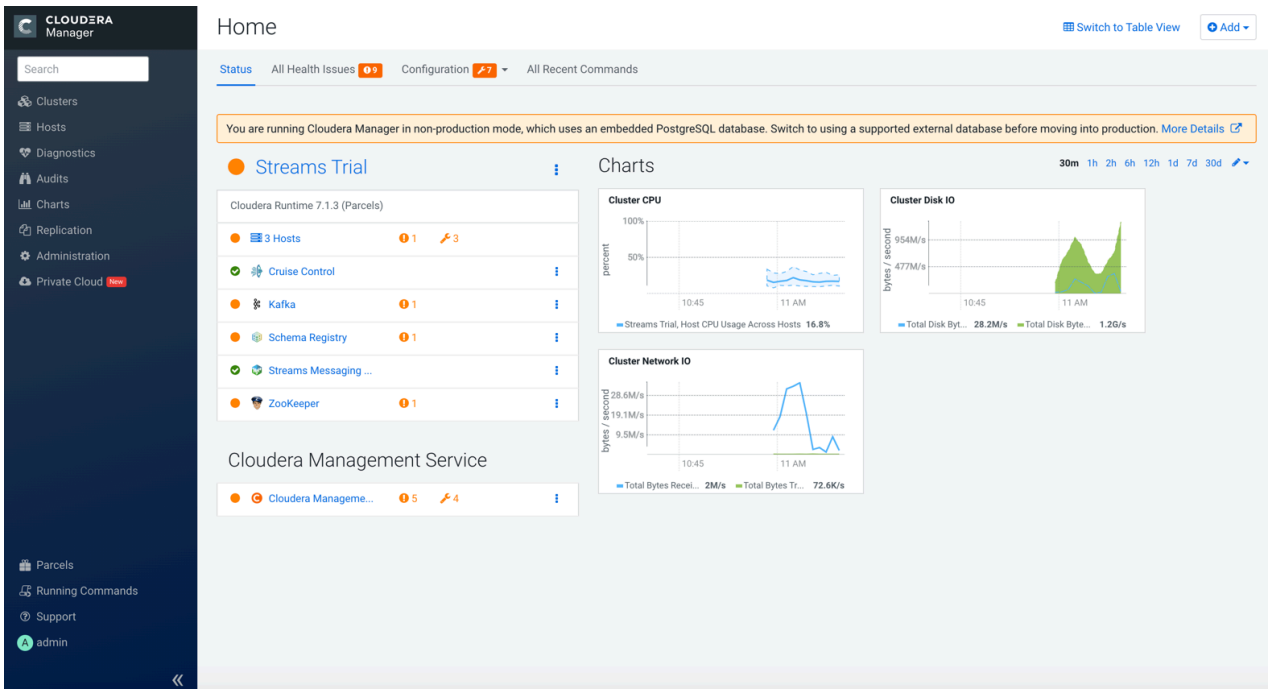
The **Summary** section appears with a success or failure report of the setup wizard.



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 and navigate to the **SMM** > **Topics** page.
   b) Click Add New.
   c) Provide the following information:
      - Topic name
      - Number of partitions
      - Level of availability
      - Cleanup policy
   d) SMM has automatically set Kafka topic configuration parameters. To manually adjust them, click Advanced.
   e) Click Save when done.
2. To create a Kafka topic from the command line:
   a) Open a terminal session and run:

```
$ bin/kafka-topics.sh --create --topic quickstart-events --bootstrap-ser
ver localhost:9092
```

   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-s
erver localhost:9092
```

```
Topic:quickstart-events  PartitionCount:1    ReplicationFactor:1 Configs
:
    Topic: quickstart-events Partition: 0    Leader: 0   Replicas: 0 I
sr: 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.

```
$ bin/kafka-console-producer.sh --topic quickstart-events --bootstrap-se
rver localhost: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:

```
$ bin/kafka-console-consumer.sh --topic quickstart-events --from-beginning
 --bootstrap-server localhost: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 > **SMM** section.
2. Review the information about your Kafka cluster on the **Overview** tab.

   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 Topics.

   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 at the top left of the page.
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

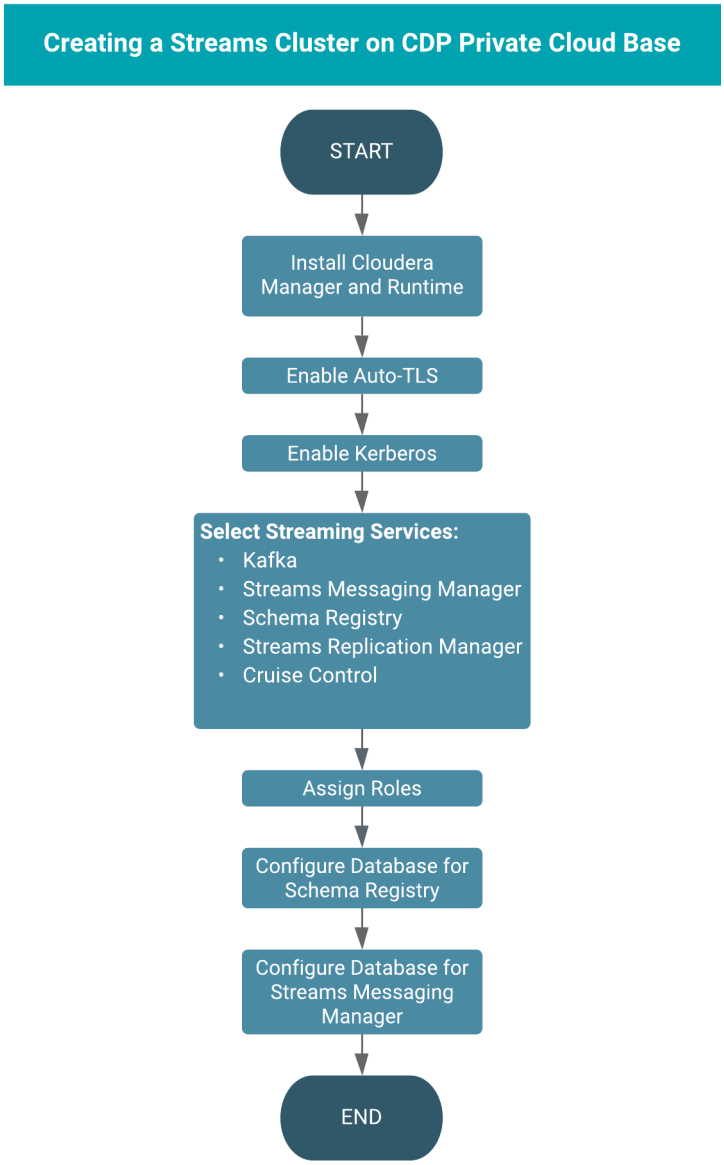
While you use and evaluate CDP Private Cloud Base, you can learn more about the streaming components from our documentation. After evaluation, you can contact Cloudera to request the appropriate license for your production environment.

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