Interset

Micro Focus Interset 5.9.2 Installation and Configuration Guide

Contents

Introduction

This guide describes how to install and configure Interset 5.9.2. It also provides an overview of the Interset cluster requirements, dependencies, components, and best practices information.

Interset uses data science and advanced analytics to identify the top risky entities and behaviors occurring in your organization. Using your organization's data, Interset establishes the *normal* behavior for your organizational entities and then, using advanced analytics, identifies the *anomalous* behaviors that constitute potential risks such as compromised accounts, insider threats, or other cyber threats.

Interset's innovative user experience, true machine learning, and big data platform easily identify and prioritize high risk anomalies, allowing your security practitioners to instantly explore the underlying raw event data. The Interset analytical models apply risk scores to individual users to provide security teams with relevant, prioritized information quickly enough to stop the activity before data loss occurs.

Interset is a server-based product that is deployed in a clustered configuration. This means that the software is distributed across multiple machines, where each machine (which can be a physical machine or a virtual machine running on a VM server such as VMware ESX) is called a node. The distribution of load and responsibilities across multiple nodes is what makes the Interset solution a scalable system that can handle large amounts of data: the more nodes in your deployment, the more data Interset can handle.

Important:

This guide provides instructions for the secure installation and configuration of Interset software and its associated platform. For information about installing Interset in an unsecured environment, contact Micro Focus Interset Support (interset.support@microfocus.com).

We recommend that you deploy the Interset product and platform in a segregated network with as little external access as possible (for example, exposing only the ports required for installation to the IP address(es) performing the installation -- TCP 22, 443, 8080). Following the installation, additional exceptions should be created for management and end-user networks.

Whenever possible, we recommend that end-users have access only to the Reporting node on TCP 443; this provides access to the Interset UI and API. Administrators should have significantly broader access, and it is generally recommended that this be handled via firewall rules and/or SSH tunnelling.

Supported Environments

Interset Analytics 5.9.2 is supported in the following x86_64 environments:

- CentOS 7.6
- Red Hat Enterprise 7.6

Interset 5.9.2 is supported with the following third-party components:

- Oracle OpenJDK 8u201/211
- Elasticsearch 6.8.1

Interset 5.9.2 is supported with HDP 3.1.0, including the following components:

Ambari	2.7.3.0
AsyncHBase	1.8.2
Avro	1.8.2
Hadoop	3.1.1
HBase	2.0.2
Hortonworks Schema Registry	0.5.3
Kafka	2.0.0
NiFi	1.10.0
Phoenix	5.0.0
Scala	2.11.8
Spark	2.3.2
Storm	1.2.1
ZooKeeper	3.4.6

Interset 5.9.2 is supported with CDH 6.1.1, including the following components:

AsyncHBase	1.8.2
Avro	1.8.2
Hadoop	3.0.0
HBase	2.1.1
Hortonworks Schema Registry	0.5.3
Kafka	2.0
NiFi	1.10.0
Phoenix	5.0.0-HBase-2.0-cdh6.1.1
Scala	2.11.8
Spark	2.4
Storm	1.2.1
ZooKeeper	3.4.5

Interset 5.9.2 supports the following Web browsers:

- Google Chrome 74 and above
- Mozilla Firefox 67 and above

Supported Data Sources

Interset 5.9.2 supports the following data sources. For .csv data sources, the delimiter can be customized.

- Active Directory
  - Active Directory event logs stored in McAfee® Enterprise Security Manager (ESM)
  - Active Directory event logs stored in Splunk®
  - Active Directory event logs stored in Micro Focus ArcSight Logger
  - Active Directory event logs stored in IBM QRadar
  - Windows Security event logs (.csv)
  - Interset-extracted Windows event logs (.csv)
  - Universal Windows event logs (.csv)
  - Windows Event Viewer-extracted event logs (.csv)
  - Active Directory authentication logs
- Universal Alerts stored in third-party DLP systems (.csv)
- NetFlow
  - Version 5
  - Version 9
  - Version 10 (IPFIX)
- Repository
  - Perforce
    - P4AUDIT logs
    - Perforce Structured Server audit logs
  - GitHub Enterprise audit logs
  - Universal repository logs (.csv)
- Pluggable Authentication Module (PAM) AuditD logs (.csv)
- Printer logs
  - Windows printer events stored in Splunk
  - Windows event logs (.csv)
  - Universal logs (.csv)
- Universal Web Proxy (.csv)
- Violations
- Expense Data
- Email Data

Interset 5.9.2 data ingest uses NiFi for data extraction, transformation, and loading. It supports the processing of data set files in the following compression formats:

- tar
- gzip
- tar gzip

To ingest packaged data from other containers such as Arcsight, IBM QRadar, McAfee ESM, and Splunk, please contact Micro Focus Interset Support at interset.support@microfocus.com.

Intended Audience

This Guide assumes that you are an experienced system administrator with sound Linux skills and are familiar with your organization's server environment, security infrastructure, and data sources.

You should also be familiar with the business needs of your organization.

Installation

The Interset installation has several parts:

- installing and configuring components common to all machines in the Interset configuration
- installing and configuring the software on the
    - Monitoring node

      Where we refer to the Monitoring node in this document, we are referring to the node in your Interset cluster where Ambari or Cloudera Manager is installed.
    - Master node(s)
    - Stream node(s)
    - Compute node(s)
    - Search node(s)
    - Reporting node
- tenant configuration

How to Use This Guide

The end-to-end installation and configuration described in this document provides all of the information required to run your Interset cluster.

The instructions in this Guide refer to a ***<mirror>*** location, a repository with the installation packages and files required for this Interset upgrade:

- In an offline installation scenario, this ***<mirror>*** location must be created on a local Web server accessible to all of the servers in your Interset cluster.

  For information about creating an offline mirror repository, please contact Micro Focus Interset Support at interset.support@microfocus.com.

- In an online installation scenario, you configure this ***<mirror>*** location to be the online Interset repository location.

  If you require access credentials, please contact Micro Focus Interset Support at interset.support@microfocus.com.

- In an offline installation scenario, this ***<mirror>*** location must be created on a local Web server accessible to all of the servers in your Interset cluster.

  For information about creating an offline mirror repository, please contact Micro Focus Interset Support at interset.support@microfocus.com.

- In an online installation scenario, you configure this ***<mirror>*** location to be the online Interset repository location.

  If you require access credentials, please contact Micro Focus Interset Support at interset.support@microfocus.com.

As you use this Guide to prepare for and execute your Interset installation, we recommend that you work through the sections in the order in which they are presented.

Important: The scripts and commands provided throughout the installation and configuration instructions are designed to be copied from this Guide to the command prompt in your console. However, you may experience that some scripts and commands don't copy correctly when using different PDF viewer applications. As a result, Interset recommends that, as you copy text to your command console, you quickly review it to ensure that it has copied correctly.

Additional Support

Should you experience unexpected results or identify issues that are not addressed in this document, please contact Micro Focus Interset Support at interset.support@microfocus.com.

Prerequisites

Before you begin, ensure that you review and comply with the following important prerequisites:

- Ensure that your current environment is supported. For more information, please see "Supported Environments" on page 7.
- The umask on your system must be set to 027 or a more permissive value to ensure a successful installation.

  We recommend that you ensure the umask value for your Interset cluster is set to 027 or more permissive prior to beginning the installation. If the umask value is not properly set prior to the installation, you will have to suspend the installation during the base configuration of all nodes and resolve the issue prior to continuing.

  For assistance setting the umask value for your operating system, please contact your operating system administrator.
- If you are performing an offline installation:
  - You must create an offline repository by downloading and extracting the following files from https://public-repo.interset.cloud/5.9.2/interset/download/ (contact Micro Focus Interset Support for required access credentials) to a Web server where the files can be accessed by all of the nodes in your cluster:
    - **interset-592.tar** AND
    - **interset-HDP-592.tar** OR **interset-CDH-592.tar**

    > ✓ **Tips:**
    >
    > - Within this Installation Guide, the server where you extract this file (the mirror server) is referred to as **<mirror_fqdn>**.
    > - You can use `wget` or `cURL` to download the tarball (.tar) file from the Interset online repository and `tar -xvf` to extract it to the server hosting your local mirror.
    >
    >   For instructions on downloading and extracting the **.tar** files, please see "Create a Local YUM Repository for Offline Installation" on page 21.

  - The Interset **<mirror_fqdn>** also requires access to the CentOS / RHEL base repository to retrieve multiple dependencies that are not bundled with the Interset installation packages. Please ensure that this access is available either by your offline **<mirror_fqdn>** repository for the relevant operating system, or through the use of an RHEL satellite server.
  - Each node in the Interset cluster must be able to access the local repositories on the <***mirror_fqdn***> server.
  - Additionally, the **mirror** setting in the **deploy.sh** file must be updated to reflect your offline repository.

For example,

```
mirror="http://<mirror_fqdn>/tmp/5.9.2/interset"
```

- If you are performing an online installation:
  - You do not need to download any archives. Instead, on each machine in your Interset cluster, you will run a bash script to copy and untar the installer archive to your local cluster. "Copy and Untar the Interset Installer Archive" on page 26
  - Each node in your Interset cluster must be able to access the Interset online repository.

Important: For information about recommended disk partitioning and volume sizing for the Interset cluster and data storage, please consult Micro Focus Interset Support at interset.support@microfocus.com for guidance.

Interset Cluster Components

This section provides information about:

- the Interset components
- the third-party components
- the recommended distribution for the Single-instance Node and Multi-instance Node production installations described in this Guide

Interset Components

The Interset components, which will be installed on different nodes in the recommended configurations, include:

- **Interset Analytics**

  This component performs the vital task of determining individual behavioral baselines, and then discovering and ranking deviations from those baselines.

  Interset Analytics is installed on the Master node(s).

- **Interset Reporting**

  This component provides the REST API, as well as the rich user interface that allows the analytics results and raw data to be explored visually.

  Interset Reporting is installed on the Reporting node.

- **Interset Workflow**

  This component applies user-defined rules to highlight specific events and trigger follow-up actions. These user-defined events contribute to the analytics.

  Workflow is installed on the Master node(s).

Third-party Components

The Interset cluster third-party components, also distributed among multiple nodes in the recommended configurations, include:

- **Apache Ambari Server**

  The Apache Ambari project simplifies Apache Hadoop management with the development of software for provisioning, managing, and monitoring Hadoop clusters. Ambari provides an intuitive, central Hadoop management user interface backed by its REST APIs.

  Apache Ambari server is installed on the Monitoring node.

- **Apache Ambari Metrics**

  The Ambari Metrics System (AMS) collects, aggregates, and serves up Hadoop and cluster metrics in Ambari-managed clusters.

  Apache Ambari Metrics is installed on the Ambari, Compute, Master, and Stream nodes.

- **Apache Ambari Client**

  An Ambari client is the node in the cluster that provides the client libraries for any services managed by Ambari, and supports installed client applications (such as Interset Analytics).

  Apache Ambari Client is installed on the Stream node(s).

- **Cloudera Manager Server**

  Cloudera Manager is an end-to-end application for managing CDH clusters. Cloudera Manager provides granular visibility into and control over every part of the CDH cluster—empowering operators to improve performance, enhance quality of service, increase compliance, and reduce administrative costs. With Cloudera Manager, you can easily deploy and centrally operate the complete CDH stack and other managed services.

  Cloudera Manager Server is installed on the Monitoring node.

- **Cloudera Manager Agents**

  The Cloudera Manager Agent is a Cloudera Manager component that works with the Cloudera Manager Server to manage the processes that map to role instances.

  Cloudera Manager Agents are installed on the Monitoring, Compute, Master, and Stream nodes.

- **Apache HDFS™**

  The Hadoop Distributed File System (HDFS) is a distributed file system that provides high-throughput access to application data. All Interset Analytics data, residing in the HBase database, is stored in HDFS.

  Apache HDFS is installed on the Data node.

- **Apache HBase™**

  HBase is a scalable, distributed database that supports structured data storage for large tables. HBase stores the Analytics data for the Interset cluster.

  Apache HBase is installed across the Compute and Master node(s).

- **Apache Storm™**

  Apache Storm reliably processes unbounded streams of data, doing for real-time data processing what Hadoop does for batch processing.

  In a Single Instance Node configuration, Apache Storm is installed on the Master and Stream node.

  In a Multi-instance Node configuration, Apache Storm is installed on multiple Master nodes and the Stream node.

- **Apache Spark2™**

  Spark2 is a fast, general computing engine for Hadoop data. Spark2 executes the Analytics, providing a simple and expressive programming model to support a wide range of applications, including ETL, machine learning, stream processing, and graph computation.

  Apache Spark2 is installed on the Master node.

- **Apache NiFi**

  Apache NiFi is an easy to use, powerful, and reliable system to process and distribute data. It supports powerful and scalable directed graphs of data routing, transformation, and system mediation logic. NiFi is used for the Interset Analytics data ingest.

  Apache NiFi is installed on the NiFi node.

- **Apache Kafka**

  Apache Kafka is a distributed publish-subscribe messaging system that is designed to be fast, scalable, and durable. In the Interset cluster, Kafka is used for data transport to Storm.

  Apache Kafka server is installed on the Stream node.

- **Apache ZooKeeper™**

  ZooKeeper is a high-performance coordination service for distributed applications. In the Interset cluster, ZooKeeper manages the coordination of the various component configurations.

  ZooKeeper is installed on the Master node(s).

- **Elasticsearch**

  Elasticsearch is an open source, broadly-distributable and easily-scalable enterprise-grade search engine. Elasticsearch houses all of the Interset Analytics raw events, and provides all of the data that drives the user interface.

  Elasticsearch is installed on the Search node.

- **Kibana**

  Kibana is an open source data visualization plug-in for Elasticsearch. Kibana serves as the user interface and data exploration mechanism for Elasticsearch.

  Kibana is installed on the Reporting node.

- **Nginx**

  Nginx is a free, open-source, high-performance HTTP and reverse proxy server, as well as an IMAP/POP3 proxy server. Nginx is recognized for its high performance, stability, rich feature set, simple configuration, and low resource consumption.

  Nginx is installed on the Reporting nodes.

- **FreeType**

  FreeType is a public software library for rendering fonts. Interset uses FreeType when rendering PDF Reports.

  FreeType is installed on the Reporting node.

- **Chromium**

  Chromium is an open-source Web browser with PDF generation capability in headless mode.

  Chromium is installed on the Reporting node.

Component Distribution

The Interset production installation distributes the components across separate machines, or nodes, identified as follows:

- Monitoring node

  The Monitoring node is where the cluster management server (Ambari or Cloudera Manager) is installed. Ambari and Cloudera Manager provide convenient Web user interfaces that simplify the deployment and management of the different components that make up the Interset solution.

- Master node(s)

  The Master node(s) is used for various infrastructure components, and for starting the Interset Analytics process.

- Search node(s)

  The Search node(s) is used for the Elasticsearch cluster and, in turn, are used by the Interset Reporting components.

- Stream node(s)

  The Stream node is used for moving data to Storm for determining violations.

- NiFi node(s)

  The NiFi node is used for ingesting data, moving data to the compute nodes for the Analytics, and later to the Search node(s) for Interset Reporting.

- Compute node(s)

  The compute node(s) is used for running the Interset machine learning algorithms on the ingested data to detect anomalous behaviour and score entities. This process uses big-data components for analysis and storage.

- Reporting node(s)

  The Reporting node provides a Web interface for Interset Reporting, and for further exploring and investigating anomalies identified by Interset Analytics.

Interset Configuration

Your Interset configuration will depend primarily on the amount of data to be analyzed. Interset recommends two basic configurations:

- Single-instance Node Configuration

  In this configuration, there is only one instance of each node type.
- Multi-instance Node Configuration

  In this configuration, there are multiple instances of various node types, depending on your data volumes.

Single-Instance Node Configuration

In a single-instance node configuration, the Interset and third-party components are distributed as follows:

Monitoring node	Master Node	Stream Node	Compute Node	Search Node	NiFi Node	Reporting Node
Monitoring Server	ZooKeeper	Metrics Monitor	Metrics Monitor	Elasticsearch	Apache NiFi	Nginx
	MetricsMonitor	Kafka Broker	Yarn Node Manager			Interset Reporting (Kibana, Nginx, Reporting Config, Chromium)
	Metrics Collector	Hadoop Client Components	HBase RegionServer			
	Yarn App Timeline Server	Ingest .jar files for Streaming	HDFS DataNode			
	Yarn Resource Manager		Storm Supervisor			
	Yarn History Server					
	HBase Master					
	HDFS (S)NameNode					
	Spark History Server					
	Hadoop Client Components					
	Storm DRPC Server					
	Storm Nimbus					
	Storm UI Server					
	Interset Analytics (including Workflow)					

Multi-Instance Node Configuration

To maximize redundancy in the infrastructure and performance of the overall cluster, Interset recommends the Multi-instance Node configuration illustrated below. This configuration can also be set up as a high availability (HA) system.

In this multi-instance node configuration, the Interset and third-party components are distributed as follows:

Monitoring node	Master Node 1	Master Node 2	Master Node 3	NiFi Node (s)	Stream Node	Compute Node	Search Node	Reporting Node
Monitoring Server	ZooKeeper	ZooKeeper	ZooKeeper	Apache NiFi	Metrics Monitor	Metrics Monitor	Elasticsearch	Nginx
	Metrics Monitor	Metrics Monitor	Metrics Monitor		Kafka Broker	Yarn Node Manager		Interset Reporting (Kibana, Nginx, Reporting Config, Chromium)
	Metrics Collector	Hadoop Client Components	Hadoop Client Components		Hadoop Client Components	HBase RegionServer		
	Hadoop Client Components	Yarn App Timeline Server	Yarn Resource Manager		Interset Ingest .jar files for Streaming	HDFS DataNode		
	HBase Master	Yarn Resource Manager	HBase Master			Storm Supervisor		
	HDFS NameNode	Yarn MapReduce2 History Server	HDFS JournalNode					
	HDFS JournalNode	HDFS JournalNode	Storm Passive Nimbus					
	ZooKeeperFC	HDFS NameNode						
	Storm DRPC Server	ZooKeeperFC						
	Storm Active Nimbus	Spark History Server						
	Storm UI Server							
	Interset Analytics (including Workflow)							

For information and assistance calculating the optimal Interset topology for your organization, please contact Micro Focus Interset Support at interset.support@microfocus.com.

Create a Local YUM Repository for Offline Installation

To perform an offline installation, you must create a local YUM repository that mirrors the Interset installation repository on a server accessible to your cluster machines, and where a Web server has been configured.

✓ **Tips:**

- If you do not already have an HTTP daemon installed on a local server, you can download and configure an HTTPD service such as Apache, nginx, or GWS on any machine accessible to your Interset cluster.
- We recommend that the Web server not be installed on the cluster Reporting or Endpoint node, as port conflicts may result.
- The instructions below are written for an Apache Web server. If your Web server is not an Apache Web server, please refer to the appropriate documentation for your component.

1. In a Web browser, navigate to https://public-repo.interset.cloud/5.9.2/interset/download/.

   If you require access credentials, please contact Micro Focus Interset Support at interset.support@microfocus.com.
2. Download the following archives to your Web server machine, for example using `wget`:
   - **interset-592.tar** AND
   - **interset-HDP-592.tar** OR **interset-CDH-592.tar**
3. Extract the contents of the **.tar** files to the **/var/www/html** directory, using the following command.

   ```
   sudo tar -xvf <filename>.tar -C /var/www/html
   ```

   This may take several minutes to complete.

   Extracting the **interset-592.tar** file creates the following directory structure:

   ```
   /var/www/html/5.9.2/interset
   ```

4. Run the following command to install the Apache Web server (**httpd**).

   ```
   sudo yum install httpd -y
   ```

5. Ensure that all extracted files and directories are owned by **apache:apache**.

   ```
   sudo chown -R apache:apache 5.9.1
   ```

6. Navigate to **/etc/httpd/conf**, and then open the **httpd.conf** file for editing.

   ```
   sudo vi httpd.conf
   ```

7. In the **<Directory "/var/www/html">** definition, add the following parameters and values.

   These parameters allow for anonymous access, and directory browsing of the root level of the repository.

   ```
   Options Indexes FollowSymLinks
   AllowOverride All
   Order allow,deny
   Allow from all
   ```

8. Run the following command to ensure that the configuration is working properly.

   ```
   apachectl configtest
   ```

   The optimum response is:

   ```
   Syntax OK
   ```

   If the response identifies issues with the configuration, correct them as appropriate.

   > Depending on the configuration of your cluster, you may need to set Selinux to permissive mode so that network requests will be allowed.

9. Run the following command to restart the Web server.

   ```
   sudo service httpd restart
   ```

10. In a Web browser, navigate to **http://<*repository_server_fqdn*>/5.9.1** to verify that you can successfully access the Interset 5.9.1 repository.
11. Back on the mirror computer, navigate to the **/var/www/html/5.9.1/interset** directory, and locate the **deploy.sh** file.

12. Open the **deploy.sh** file, and edit the **mirror** setting to reflect your offline repository URL.

    ```
    version=5.9.2
    theme="interset"
    mirror="http://<repository_server_fqdn>/$[version]"
    ```

    With the offline repository directories in place, and the offline <***repository_server_fqdn***> location defined in the **deploy.sh** file, you are now ready to begin your offline installation.

    > **Tip:** From this point forward, the references to your offline repository server will appear in this Guide as <***mirror_fqdn***>.

13. Save and close the **deploy.sh** file.

Install a New Interset Cluster

These instructions detail the installation and configuration of Interset 5.9.2.

Important:

- This document describes the installation procedure for a secured cluster. If you are installing an unsecured cluster, the numbers in the Interset installation menu will be different. For more information about installing Interset on an unsecured cluster, contact Micro Focus Interset Support.
- For information about recommended disk partitioning and volume sizing for the Interset cluster and data storage, please refer to the Knowledge Base article, *System Requirement - Partition Sizing Guidelines* and consult Micro Focus Interset Support for guidance.

You can contact Micro Focus Interset Support at interset.support@microfocus.com.

Installing the Interset system includes the following tasks:

- copying and extracting the Interset installer archive
- editing the Interset installer configuration file
- editing the Secure Properties file
- downloading the JDK and JCE
- performing the base configuration of all nodes
- installing Ambari OR installing Cloudera
- configuring Kerberos by:
  - generating the TLS certificates
  - setting up the KDC
  - configuring Kerberos on Ambari OR configuring Kerberos on Cloudera
- installing the schema registry
- configuring the Master node
- configuring the Stream node
- configuring the Search node
- configuring Reporting
- setting up authentication
- configuring Workflow
- configuring NiFi
- adding a new data source
- running Analytics

- setting up the search indexes in Elasticsearch
- enabling Windowed Analytics

Copy and Untar the Interset Installer Archive

1. On all machines in the cluster, run the following command as a user with sudo access, substituting <*mirror_fqdn*> with the fully-qualified domain hostname for your installation repository.

```
bash <(curl -ks http://<mirror_fqdn>/5.9.2/interset/deploy.sh)
```

> **Tip:** In an online installation, you will be required to include your access credentials. For example,
>
> ```
> username:password@<mirror_fqdn>/5.9.2/interset/deploy.sh
> ```

This script copies the Interset installer package, untars contents to the **/opt/interset/installer** directory, creates the **interset** user, and gives ownership of the **interset_installer** folder to **interset** user.

> **Tip:** At this time, the **interset** user password is **interset**. You will be prompted to change this password in the early stages of the cluster installation, during the base configuration of all nodes.

The Interset repository will be fully configured after running this script.

Troubleshooting

As you execute each installation script on the **Monitoring node**, detailed installation logs are written to the **/opt/interset/log/** directory on the relevant node in the cluster. For example, when you run **Installer option 5 for Stream node(s) installation**, the installer writes the log file to **/opt/interset/log/intersetlog_installIngestLog_install.txt** on the Stream node. If any of the node configurations fail during the install action process, check for errors in the log files located in the following directory, where **<*node_fqdn*>** corresponds to the server name of the node where the installation failed:

```
/tmp/interset_installer/interset_install_logs/<node_fqdn>
```

You can monitor these logs on the relevant node in the cluster throughout the installation process.

> **Tip:** At any time in the installation, you can force log retrieval without quitting the installation process by using menu option **l** (L).

Edit the Interset Installer Configuration File

1. On the machine designated as the Monitoring node in your Interset cluster, log in as the **interset** user.
2. Navigate to the **/opt/interset/installer/etc** directory, and then locate and open the **config** file for editing.

   This file has a line for each role, or node, in an Interset cluster (for example, **MONITORING** or **STREAM**).
3. In the **config** file, edit the following settings:
   - Verify that the **INTERSET_VERSION**, **THEME**, AND **INTERSET_REPO** settings are configured as follows:

     ```
     INTERSET_VERSION="5.9.1"
     THEME="interset"
     INTERSET_REPO="http://<mirror_fqdn>"
     ```

     For an offline installation, *<mirror_fqdn>* is the server where you installed the offline mirror repository.

     For an online installation, *<mirror_fqdn>* is the URL of the Interset repository (http://repo.interset.com/5.9.2), and must include your access credentials. For example,

     ```
     INTERSET_REPO="http://<username>:<password>@repo.interset.com/5.9.2"
     ```
   - Enter the appropriate Hadoop monitoring system for the cluster, **HDP** or **CDH**. For example,

     ```
     HADOOP_ENV="HDP"
     ```
   - Set the Hadoop security protocol to **KERBEROS**:

     ```
     # HADOOP_SECURITY should be NONE or KERBEROS
     HADOOP_SECURITY=KERBEROS
     ```
   - For each role, enter the fully-qualified domain name of the server(s) that you want to allocate to that role.

     > ✓ **Tip:** For roles that have more than one physical node, enter the multiple fully-qualified domain names as a space-separated list.

     ```
     # 1 value only
     MONITORING="monitoring.interset.com"
     REPORTING="reporting.interset.com"
     POSTGRES="postgres.interset.com"
     PERFMON="perfmon.interset.com"

     # between 1 and many (space-separated) values
     COMPUTE="compute.interset.com"
     MASTER="master.interset.com"
     SEARCH="search.interset.com"
     STREAM="stream.interset.com"
     NIFI="nifi.interset.com"
     ```

> **Notes:**
>
> - **PERFMON** is optional and can be left blank.

4. Save and close the **config** file.

Edit the Secure Properties File

1. If you haven't already, on the machine designated as the Monitoring node in your Interset cluster, log in as the **interset** user.
2. Navigate to the **/opt/interset/installer/etc/secure** directory, and then locate and open the **secure.properties** file for editing.
3. Edit the `ROOT_SUBJ`, `INTERM_SUBJ`, and `SERVER_SUBJ` entries to reflect your location and organization:

```
ROOT_SUBJ="/C=US/ST=California/L=Irvine/O=Interset/CN=Root-ca"
INTERM_SUBJ="/C=US/ST=California/L=Irvine/O=Interset/CN=Intermediate-ca"
SERVER_SUBJ="/C=US/ST=California/L=Irvine/O=Interset/CN"
```

4. Enter the default realm for Kerberos to use:

```
DEFAULT_REALM="INTERSET.COM"
```

5. Save and close the **secure.properties** file.

Download Oracle JDK 8 and the Java Cryptography Extension (JCE)

Interset 5.9.2 requires the Oracle JDK 8 and the Java Cryptography Extension (JCE), which can be downloaded from Oracle.

> If you are performing an online installation, you can skip this step and allow the Interset installer to download and install OpenJDK.

To download the JDK and JCE:

1. On a local machine in your environment with Web access, open a browser and then navigate to http://www.oracle.com/technetwork/java/javase/downloads/index-jsp-138363.html.
2. At the top of the page, click **Sign In** to log into your Oracle account (or create one if necessary).
3. On the **Java SE Downloads** page, locate the **Java SE 8uXXX** section and then, in the **JDK** box, click **Download** to download the Oracle JDK.
4. In the **Java SE Development Kit 8 Downloads** page, scroll down and accept the license agreement, and then locate and download the JDK RPM package for your operating system (**jdk-8u202-linux-x64.rpm** or later).
5. Click the title of the **Downloads** tab to return to the **Java SE Downloads** page.
6. In the **Additional Resources** section, locate the **Java Cryptography Extension (JCE) Unlimited Strength Jurisdiction Policy Files for JDK/JRE** and then, to the right of that entry, click **Download** to download the JCE.
7. In the **Java Cryptography Extension (JCE) Unlimited Strength Jurisdiction Policy Files Download** page, accept the license agreement, and then download the **JCE Unlimited Strength Jurisdiction Policy Files 8**.
8. Transfer the two downloaded archives to the Monitoring node in your Interset cluster.

   > ✓ **Tip:** Make note of the location on the Monitoring node where you transfer the JDK and JCE installation archives, as these paths are required in the next section.

Perform the Base Configuration of All Nodes

1. As the **interset** user, ssh to the Monitoring node in the Interset cluster.

```
ssh interset@<monitoring_node_fqdn>
```

> **Tip:** The Monitoring node is the node in your cluster where Ambari or Cloudera Manager is installed.

2. Run the following commands to navigate to the **/opt/interset/installer** director and launch the installation menu.

```
cd /opt/interset/installer
```

```
./install.sh
```

3. When prompted to review the End User License Agreement (EULA), type **READ** at the prompt, and then space through the agreement as you read it until you reach the end.
4. At the end of the EULA text, type **ACCEPT**, and then press **Enter**.

> **Important:** If you are installing Interset on Amazon Web Services (AWS), please consult Micro Focus Interset Support at interset.support@microfocus.com for guidance before you proceed with **Installer option 1**.

5. From the available installation options, type **1** to select installation **[1] Initial server config (set up SSH, offline_setup script, database ...)**, and then press **Enter**.

```
Please select an installation option from the list above: 1
```

6. From the available installation options in the Installation menu, select:

```
Installer option: 1 for initial server config (set up SSH, offline_setup script, etc...)
```

The output will appear similar to:

```
Running ssh-copy-id to configure passwordless SSH access for local interset user to all interset servers
```

7. When prompted, enter and then re-enter a new password for the **interset** user.
8. You may be prompted to enter the password again for each node. When prompted to continue connecting, enter `yes`:

```
Changing password for 'interset' user on <node_fqdn>...
The authenticity of host '<node_fqdn> (10.31.20.80)' can't be established.
ECDSA key fingerprint is SHA256:t0K8JFJZ7l+bqBskkReVwoA4gFeoRQv4Yl3LpRR0m1Y.
ECDSA key fingerprint is MD5:df:57:6f:58:24:27:6d:4b:74:2a:06:9e:e7:bc:bd:b8.
Are you sure you want to continue connecting (yes/no)? yes
```

9. If you receive the following warning about the system umask setting, you must resolve this issue before you proceed with the Interset 5.9.2 installation.

```
We detected that the umask value for user <user> on <host> is incorrect: <umask>
Ambari, HDP, and HDF support umask values of: 022, 0022, 027, 0027.
Once this is addressed, type CONTINUE...
```

Important: The system **umask** setting restricts the permissions for new files on Linux operating systems. Interset 5.9.2 requires a minimum **umask** setting of **027**, or a more permissive value. If you receive this warning, please contact your operating system administrator to set the cluster **umask** value to **027**.

10. You are prompted to install OpenJDK.

    ```
    Install Openjdk via yum? [y/n]:
    ```

    Do one of the following:

    - If you are performing an offline installation:

      a. Enter `n` at the prompt.

      b. Enter the path and filename of the Oracle JDK RPM package that you downloaded in "Download Oracle JDK 8 and the Java Cryptography Extension (JCE)" on page 30

      ```
      A 'Java 8' RPM must be provided. Please enter the fully-qualified filename (e.g. /tmp/jdk-8u202-
      linux-x64.rpm):
      ```

      c. When prompted, enter the path and filename of the Oracle JCE package that you downloaded.

      ```
      A 'Java 8 Cryptography Extensions' ZIP must be provided. Please enter the fully-qualified
      filename (e.g. /tmp/jce_policy-8.zip):
      ```

    - If you are performing an online installation, enter `y` at the prompt to install OpenJDK.

      ```
      Install Openjdk via yum? [y/n]:y
      ```

    As the initial server configuration continues, the following response appears.

    ```
    Waiting for installation to complete on all hosts...
    ```

11. When prompted, enter – and then re-enter – new passwords for the **Postgres** database user, the cluster manager database user (**HDP Monitoring** or **CDH Monitoring**) , and the **Hive** database user.

    ```
    Please enter password for DB user - Postgres:
    Please enter password for DB user - <CDH or HDP> Monitoring:
    Please enter password for DB user - Hive:
    ```

    Important: Ensure that you save these passwords in a reference document, as these passwords are not exposed within the log files, and you will need them later in the installation process.

When the initial server configuration completes on all nodes, you are returned to the Installation menu and you can proceed to the next section.

✓ **Tip:** As each installation option is completed and you are returned to the installation menu, a green check mark appears beside the installation option to indicate its successful completion. Use these check marks as a guide to track your progress through the installation.

Should a red **X** appear beside an installation option, this indicates that the installation option did not successfully complete. In this situation, run the installation option again.

Install the Ambari Server

> If you are installing Cloudera, skip this section and proceed to "Install Cloudera Manager" on page 52.

Having just completed the base configuration of all nodes in your Interset cluster, you should still be in a command console on the Monitoring node.

1. In a command console on the Monitoring node, run the following commands to navigate to the **/opt/interset/installer** directory and launch the Interset installation menu.

   ```
   cd /opt/interset/installer
   ```

   ```
   ./install.sh
   ```

2. From the available installation options, type **2** to select installation **[2] Ambari node installation**, and then press **Enter**.

   ```
   Please select an installation option from the list above: 2
   ```

3. When prompted that SELinux is to be temporarily disabled and asked to continue, type **y**:

   ```
   Setup ambari-server
   Checking SELinux...
   SELinux status is 'enabled'
   SELinux mode is 'permissive'
   WARNING: SELinux is set to 'permissive' mode and temporarily disabled.
   ```

   ```
   OK to continue [y/n] (y)? y
   ```

4. When asked to customize the user account for the ambari-server daemon, type **y**:

   ```
   Customize user account for ambari-server daemon [y/n] (n)? y
   ```

5. When you are prompted for the user account of the ambari-server daemon, type **interset**:

   ```
   Enter user account for ambari-server daemon (root): interset
   ```

   ```
   Adjusting ambari-server permissions and ownership...
   Checking firewall status...
   ```

6. When to select a JDK, type **2** for **Custom JDK**.

   ```
   Checking JDK...
   [1] Oracle JDK 1.8 + Java Cryptography Extension (JCE) Policy Files 8
   [2] Custom JDK
   >==============================================================================
   ```

   ```
   Enter choice (1): 2
   ```

7. When prompted for the **Path to JAVA_HOME**, enter the path to your JDK installation:

   - for Oracle JDK: **/usr/java/jdk1.8.0_<*xxx*>-amd64**
   - for OpenJDK: **/usr/jdk64/jdk1.8.0_<*xxx*>**

   ```
   WARNING: JDK must be installed on all hosts and JAVA_HOME must be valid on all hosts.
   WARNING: JCE Policy files are required for configuring Kerberos security. If you plan to use Kerberos,please
   make sure JCE Unlimited Strength Jurisdiction Policy Files are valid on all hosts.
   ```

   ```
   Path to JAVA_HOME:/usr/java/jdk1.8.0_202-amd64
   ```

Important:

If you do not know the path to JAVA_HOME, use the command in tip below.

```
Validating JDK on Ambari Server...done.
Completing setup...
Configuring database..
```

Tips:

- To confirm the **Path to JAVA_HOME**, in a new command console, at the prompt, type the following:

  ```
  cat /opt/interset/java_home.sh
  ```

- When you enter the **Path to Java Home**, replace **<*xxx*>** in the file name above with the version you downloaded from the Oracle Web site.

  The installer will install the version of the Oracle JDK supported by Interset 5.9.2.

8. When prompted to **Enable Ambari Server to download and install GPL Licensed LZO packages [y/n]**, type **n**:

```
Enable Ambari Server to download and install GPL Licensed LZO packages y/n (n)? n
```

9. When prompted to **Enter advanced database configuration [y/n]**, type **y**:

```
Enter advanced database configuration [y/n] (n)? y
```

10. When prompted to choose a database option, type **[4] - PostgreSQL**.

```
==============================================================================
Choose one of the following options:
[1] - PostgreSQL (Embedded)
[2] - Oracle
[3] - MySQL / MariaDB
[4] - PostgreSQL
[5] - Microsoft SQL Server (Tech Preview)
[6] - SQL Anywhere
[7] - BDB
==============================================================================
```

```
Enter choice (1): 4
```

After choosing **[4] - PostgreSQL**, you are prompted to enter the following information:

```
Hostname (localhost):
Port (5432):
Database name (ambari):
Postgres schema (ambari):
Username (ambari):
Enter Database Password (bigdata):
Re-enter password:
```

11. Enter the following values:

    - For **Hostname**, enter the fully-qualified domain name of the Monitoring node.

      ```
      Hostname (localhost): <monitoring_node_fqdn>
      ```

    - For **Port**, enter **5432**

      ```
      Port (5432): 5432
      ```

    - For **Database name**, enter **ambari**

      ```
      Database name (ambari): ambari
      ```

    - For **Postgres schema**, enter **ambari**

      ```
      Postgres schema (ambari): ambari
      ```

    - For **Username**, enter **ambari**

      ```
      Username (ambari): ambari
      ```

      This username is for the HDP Monitoring (i.e., Ambari) database user.

    - For **Database password**, enter the password you selected for the HDP Monitoring database in "Perform the Base Configuration of All Nodes" on page 31.

      ```
      Enter Database Password (bigdata): <password>
      ```

    - Re-enter the same HDP Monitoring database password.

      ```
      Re-enter password: <password>
      ```

    The following message appears.

    ```
    WARNING: Before starting Ambari Server, you must run the following DDL against the database to create the
    schema: /var/lib/ambari-server/resources/Ambari-DDL-Postgres-CREATE.sql
    ```

12. Open a second command console for the Monitoring node and, in that console, run the following commands to set up the PostgreSQL JDBC driver and create the database schema.

    ```
    sudo ambari-server setup --jdbc-db=postgres --jdbc-driver=/usr/share/java/postgresql-42.2.2.jar
    ```

    ```
    psql -h <Postgres_host_name> -U ambari -d ambari -a -f /var/lib/ambari-server/resources/Ambari-DDL-Postgres-
    CREATE.sql
    ```

13. When prompted, enter the password for the Ambari database user.

14. When you are returned to the command prompt in the second Monitoring node console, close that console.

15. In the original Monitoring node console (where you are executing the installation), when prompted to proceed with configuring the remote database connection properties, type **y**.

    ```
    Proceed with configuring remote database connection properties [y/n] (y)? y
    ```

    Information will be displayed regarding your RSA private key, your HDP base URL, your HDP-UTIL base URL, and your Ambari server.

16. Copy the RSA private key, the HDP baseURL, and the HDP-UTILS baseURL to a text file. You will be required to provide these values in the next section during the Ambari server configuration.

    Private key:

    ```
    -----BEGIN RSA PRIVATE KEY-----
    XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
    -----END RSA PRIVATE KEY-----
    ```

✓ **Tip:** When copying the RSA private key, be sure to copy the entire text, including the **-----BEGIN RSA PRIVATE KEY-----** and **-----END RSA PRIVATE KEY-----** lines.

HDP baseURL:

```
You will need the HDP baseurl to configure repository for Ambari cluster installation, it is:
http://<mirror_fqdn>/5.9.2/HDP/centos7/3.x/updates/3.1.0.0-78
```

HDP-UTILS baseURL:

```
You will need the HDP-UTILS baseurl to configure repository for Ambari cluster installation, it is:
http://<mirror_fqdn>/5.9.2/HDP-UTILS/HDP-UTILS-1.1.0.22/repos/centos7
```

Your Ambari server:

```
Please navigate to http://<ambari_fqdn>:8080 in a browser to install your cluster before proceeding further.
```

⚠ At this point in the Interset installation, you are leaving the command console temporarily and launching the Ambari user interface to install the Ambari cluster.

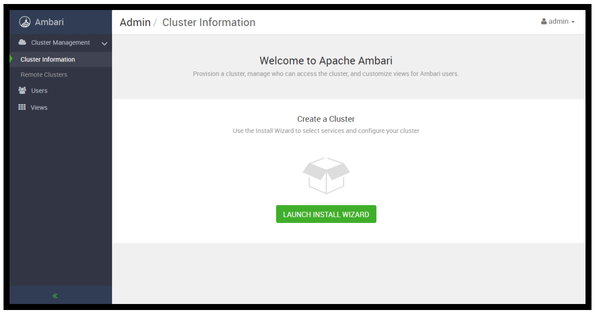
Install the Ambari Cluster

Follow the steps below to set up an Ambari Cluster.

1. Log in to Apache Ambari:
   - Open a Web browser, and go to **http://<*ambari_fqdn*>:8080**, where **<*ambari_fqdn*>** is the name of your Ambari node.
   - Type the following login credentials, then press **Enter**.

     Username: **admin**

     Password: **admin**
2. In the Apache Ambari **Admin / Cluster Information** page, click **Launch Install Wizard**.



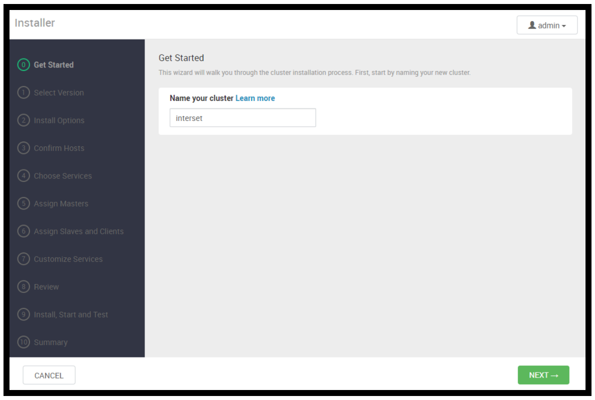
Welcome to Apache Ambari

Provision a cluster, manage who can access the cluster, and customize views for Ambari users.

Create a Cluster

Use the Install Wizard to select services and configure your cluster

3. In the **Get Started** page, in the **Name your cluster** box, type **interset**, and then click **Next**.



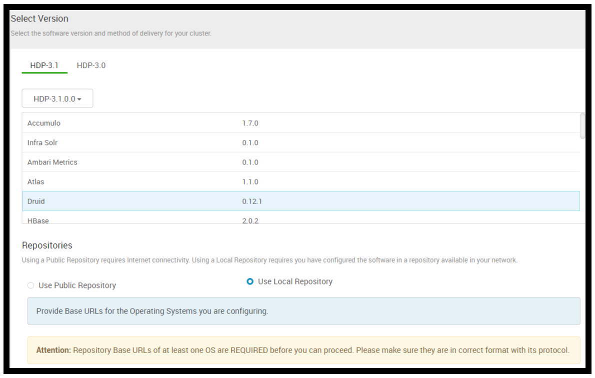
Get Started

This wizard will walk you through the cluster installation process. First, start by naming your new cluster.

Name your cluster: interset

> **Important:** The Ambari cluster name must be **interset**, with all lower-case characters. If the Ambari cluster name is set to something other than **interset**, the deployment will fail.

4. In the **Select Version** page, click the **HDP-3.1** tab.
5. Do one of the following:
   - If you are performing an online installation, from the **HDP-3.1** dropdown menu, select **HDP-3.1.0.0**.



Select Version

Select the software version and method of delivery for your cluster.

Accumulo	1.7.0
Infra Solr	0.1.0
Ambari Metrics	0.1.0
Atlas	1.1.0
Druid	0.12.1
HBase	2.0.2

Repositories

Using a Public Repository requires Internet connectivity. Using a Local Repository requires you have configured the software in a repository available in your network.

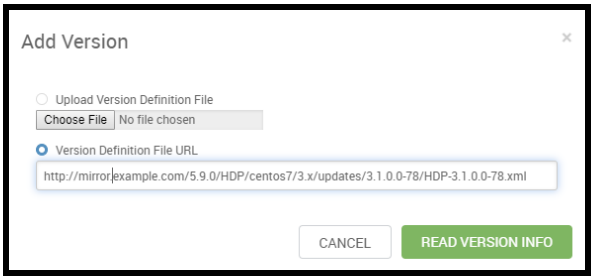
○ Use Public Repository ● Use Local Repository

Provide Base URLs for the Operating Systems you are configuring.

Attention: Repository Base URLs of at least one OS are REQUIRED before you can proceed. Please make sure they are in correct format with its protocol.

- If you are performing an offline installation, do the following:
  - from the **HDP-3.1** dropdown menu, scroll to the bottom of the list, and then select **Add Version ...**
  - In the **Add Version** dialog box, select the **Version Definition File URL** radio button, and then enter the following URL in the text box:

    ```
    http://<mirror_fqdn>/5.9.2/HDP/centos7/3.x/updates/3.1.0.0-78/HDP-3.1.0.0-78.xml
    ```

    Add Version

    ○ Upload Version Definition File

    ● Version Definition File URL

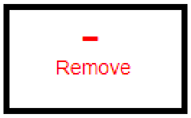
    http://mirror.example.com/5.9.0/HDP/centos7/3.x/updates/3.1.0.0-78/HDP-3.1.0.0-78.xml

    Tip: Ambari does not support TLS (HTTPS) for the Version Definition File URL. Therefore, the Web server hosting the HDP repo must be HTTP for an offline installation.

  - Click **Read Version Info**.

6. Scroll down in the page, and select the **Use Local Repository** radio button.

   In the **Repositories** table, operating systems are listed along with their HDP repository names and baseURLs. In the right-most column of the table, the **Remove** option allows you to remove specific operating systems from the configuration.

7. For each operating system other than **redhat7**, click **Remove.**

   This removes all operating systems that are not relevant to your Interset installation.

Tip: If your operating system is CentOS 7, select **redhat7**.

When finished, only the **redhat7** operating system should remain.

8. In the **Base URL** boxes, set the local repositories using the lines that you copied in the previous section.

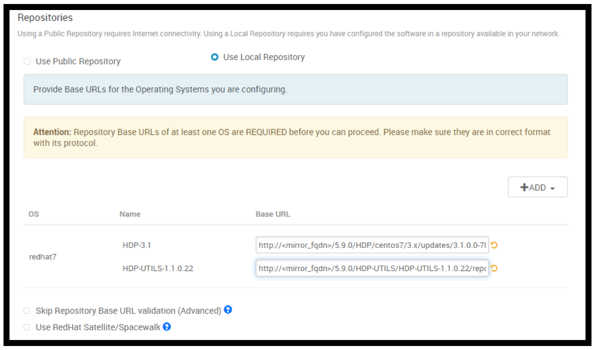
   For example:

   - HDP-3.1 base URL:

     ```
     http://<mirror_fqdn>/5.9.2/HDP/centos7/3.x/updates/3.1.0.0-78/
     ```

   - HDP-UTILS-1.1.0.22 base URL:

     ```
     http://<mirror_fqdn>/5.9.2/HDP-UTILS/HDP-UTILS-1.1.0.22/repos/centos7/
     ```



Repositories

Using a Public Repository requires Internet connectivity. Using a Local Repository requires you have configured the software in a repository available in your network.

○ Use Public Repository ● Use Local Repository

Provide Base URLs for the Operating Systems you are configuring.

Attention: Repository Base URLs of at least one OS are REQUIRED before you can proceed. Please make sure they are in correct format with its protocol.

+ADD

OS	Name	Base URL
redhat7	HDP-3.1	http://<mirror_fqdn>/5.9.0/HDP/centos7/3.x/updates/3.1.0.0-78
	HDP-UTILS-1.1.0.22	http://<mirror_fqdn>/5.9.0/HDP-UTILS/HDP-UTILS-1.1.0.22/repo

□ Skip Repository Base URL validation (Advanced)

□ Use RedHat Satellite/Spacewalk

Tips:

- To retrieve your HDP base URL at any time, run the following command:

  ```
  cat /etc/yum.repos.d/HDP.repo | grep -o 'baseurl=.*' | cut -f2- -d=
  ```

- To retrieve your HDP-UTIL base URL at any time, run the following command:

  ```
  cat /etc/yum.repos.d/HDP-UTILS.repo | grep -o 'baseurl=.*' | cut -f2- -d=
  ```

Note: If you are using an SSL-enabled repository (i.e., https://), select the **Skip Repository Base URL validation (Advanced)** check box.

For additional information about configuring offline repository base URLs, please go to the Hortonworks site at https://docs.hortonworks.com/ and then use the **Search** box to locate the article, "*Using a Local Repository*".

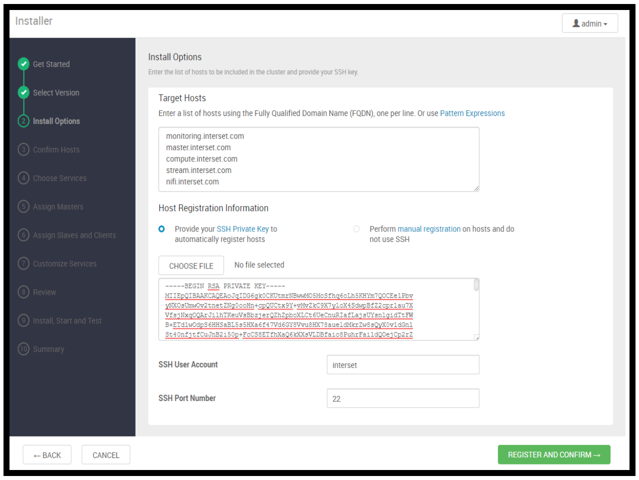
9. Click **Next**.

10. In the **Install Options** page:

    - In the **Target Hosts** box, enter the fully-qualified domain names (FQDNs) of the **Monitoring**, **Master**, **Compute**, **Stream**, and **NiFi** nodes in your Interset cluster.
    - Under **Host Registration Information**, select the **Provide your SSH Private Key** radio button to automatically register hosts, and then copy and paste the RSA private key created in the previous section into the **ssh private key** text box.

      > ✓ **Tip:** To retrieve the RSA key at any time, run the following command:
      >
      > ```
      > cat /home/interset/.ssh/id_rsa
      > ```

    - In the **SSH User Account** text box, type **interset**, and then click **Register and Confirm**.



Install Options

Enter the list of hosts to be included in the cluster and provide your SSH key.

Target Hosts

Enter a list of hosts using the Fully Qualified Domain Name (FQDN), one per line. Or use Pattern Expressions

```
monitoring.interset.com
master.interset.com
compute.interset.com
stream.interset.com
nifi.interset.com
```

Host Registration Information

● Provide your SSH Private Key to automatically register hosts

○ Perform manual registration on hosts and do not use SSH

CHOOSE FILE No file selected

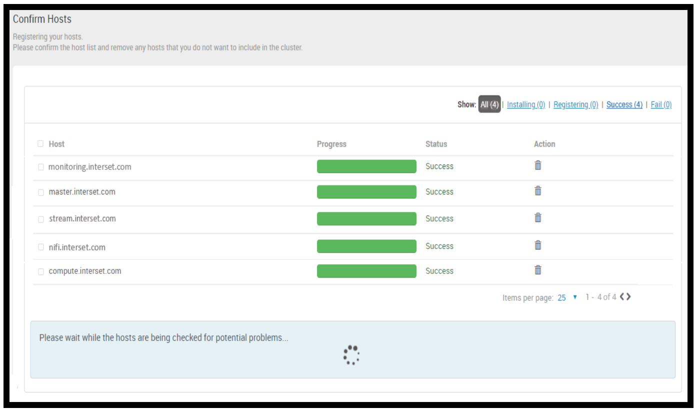
```
-----BEGIN RSA PRIVATE KEY-----
MIIEpQIBAAKCAQEAoJgIDG6gk0CKUtmrNBwwMO5MoSfhg6oLh5KHYm7QOCEe1Pbv
yNXOsUmwOv2tnetZNg0ooHn+cpQUCtx9Y+vMvZkC9X7yloX4SdwpBfZ2cprlau7X
VfsjNxgOQArJilhTKeuVsBbzjerQZhZpboXLCt6UeCnuRIafLajsUYsnlgidTtFW
B+ETd1wOdpS6HHSaBL5s5HXa6f47Vd6GY8Vvu8HX78aueldMkrZw8sQyX0v1dGn1
St40nfjtfCuJnB2i50p+FcCS8ETfhXaQ6kXXsVLDBfaio8PuhrFai1dQOejCp2rZ
```

SSH User Account: interset

SSH Port Number: 22

It takes a few minutes for Ambari to register the hosts and check for potential issues.

11. In the **Confirm Hosts** page, confirm that the list includes all of the hosts.



Confirm Hosts

Registering your hosts.
Please confirm the host list and remove any hosts that you do not want to include in the cluster.

Show: All (4) | Installing (0) | Registering (0) | Success (4) | Fail (0)

□ Host	Progress	Status	Action
□ monitoring.interset.com	🟩	Success	🗑
□ master.interset.com	🟩	Success	🗑
□ stream.interset.com	🟩	Success	🗑
□ nifi.interset.com	🟩	Success	🗑
□ compute.interset.com	🟩	Success	🗑

Items per page: 25 ▾ 1 - 4 of 4 ‹ ›

Please wait while the hosts are being checked for potential problems...

Notes:

- If you are using a time sync service other than NTP (such as VMware tools), you might see a warning about not running NTP. You can ignore this warning.
- If you receive warnings about Snappy and Snappy-devel, these warnings can be ignored.

  Do not remove these packages manually, as these are the exact versions required by Ambari. Any conflicting packages will have already been removed by the installation scripts before you installed the Ambari server.
- If the registration of more than one agent fails and the log indicates that the agent was able to communicate with the Ambari server using "localhost", this is because the Ambari server host name could not be determined. To resolve this issue, run the following command, ensuring that a fully-qualified host name is returned:

  ```
  hostname -f
  ```

  Ensure **HOSTNAME** is set to the fully-qualified server host name, for example:

  ```
  HOSTNAME=ambariserver.interset.com
  ```

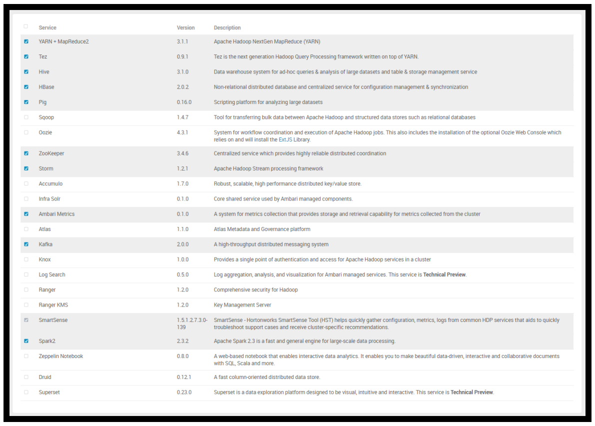
  Ensure **NETWORKING** is set to yes.

  ```
  NETWORKING=yes
  ```

  After you make these changes, run the Ambari agent registration again.
- The Ambari server does not currently offer a supported redundant or High Availability (HA) configuration. If the Ambari server or its corresponding database should become corrupt, the Hadoop stack will continue to function; however, centralized management of the cluster will be lost and cannot be recovered. As a result, the Ambari server should be backed up on a frequent basis. Standard IT system back-up and restore procedures can be used to back up and recover the Ambari server as required.

12. When the host checks have passed and you are satisfied that the list of hosts is correct, click **Next**.

    The **Choose Services** page is displayed.



□	Service	Version	Description
☑	YARN + MapReduce2	3.1.1	Apache Hadoop NextGen MapReduce (YARN)
☑	Tez	0.9.1	Tez is the next generation Hadoop Query Processing framework written on top of YARN.
☑	Hive	3.1.0	Data warehouse system for ad-hoc queries & analysis of large datasets and table & storage management service
☑	HBase	2.0.2	Non-relational distributed database and centralized service for configuration management & synchronization
☑	Pig	0.16.0	Scripting platform for analyzing large datasets
□	Sqoop	1.4.7	Tool for transferring bulk data between Apache Hadoop and structured data stores such as relational databases
□	Oozie	4.3.1	System for workflow coordination and execution of Apache Hadoop jobs. This also includes the installation of the optional Oozie Web Console which relies on and will install the ExtJS Library.
☑	ZooKeeper	3.4.6	Centralized service which provides highly reliable distributed coordination
☑	Storm	1.2.1	Apache Hadoop Stream processing framework
□	Accumulo	1.7.0	Robust, scalable, high performance distributed key/value store.
□	Infra Solr	0.1.0	Core shared service used by Ambari managed components.
☑	Ambari Metrics	0.1.0	A system for metrics collection that provides storage and retrieval capability for metrics collected from the cluster
□	Atlas	1.1.0	Atlas Metadata and Governance platform
☑	Kafka	2.0.0	A high-throughput distributed messaging system
□	Knox	1.0.0	Provides a single point of authentication and access for Apache Hadoop services in a cluster
□	Log Search	0.5.0	Log aggregation, analysis, and visualization for Ambari managed services. This service is **Technical Preview**.
□	Ranger	1.2.0	Comprehensive security for Hadoop
□	Ranger KMS	1.2.0	Key Management Server
☑	SmartSense	1.5.1.2.7.3.0-139	SmartSense - Hortonworks SmartSense Tool (HST) helps quickly gather configuration, metrics, logs from common HDP services that aids to quickly troubleshoot support cases and receive cluster-specific recommendations.
☑	Spark2	2.3.2	Apache Spark 2.3 is a fast and general engine for large-scale data processing.
□	Zeppelin Notebook	0.8.0	A web-based notebook that enables interactive data analytics. It enables you to make beautiful data-driven, interactive and collaborative documents with SQL, Scala and more.
□	Druid	0.12.1	A fast column-oriented distributed data store.
□	Superset	0.23.0	Superset is a data exploration platform designed to be visual, intuitive and interactive. This service is **Technical Preview**.

13. In the **Choose Services** page, select only the following services, and then click **Next.**
    - YARN + MapReduce2
    - Tez
    - Hive
    - HBase
    - Pig
    - ZooKeeper
    - Storm
    - Ambari Metrics
    - Kafka
    - SmartSense
    - Spark2

    Notes:

    - **SmartSense**, while not used by Interset, is required by Ambari and its selection cannot be cleared.
    - When you click **Next**, Ambari may inform you that additional services are required. If prompted, install any additional services.
    - After clicking **Next**, one or more **Limited Functionality Warning** messages may appear. If prompted with this warning, click **Proceed Anyway**.

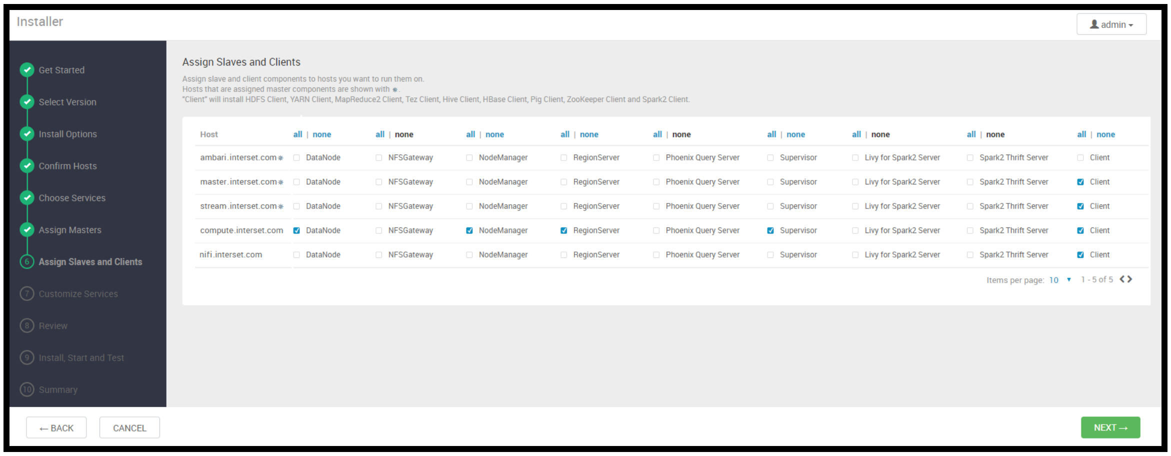
14. In the **Assign Masters** page, assign the components as follows:
    - Hive (all) -> **Monitoring** node
    - Kafka broker -> **Stream** node
    - everything else -> **Master** node

Notes:

- Three (3) ZooKeeper servers will appear in the list, with one assigned to each host. Remove the ZooKeeper servers assigned to the Compute and Stream nodes, leaving only the ZooKeeper server assigned to the Master node.
- If there are duplicate components that cannot be assigned to a **Master** node, remove the duplicate(s).

15. Click **Next**.

    The **Assign Slaves and Clients** page is displayed.



Assign Slaves and Clients

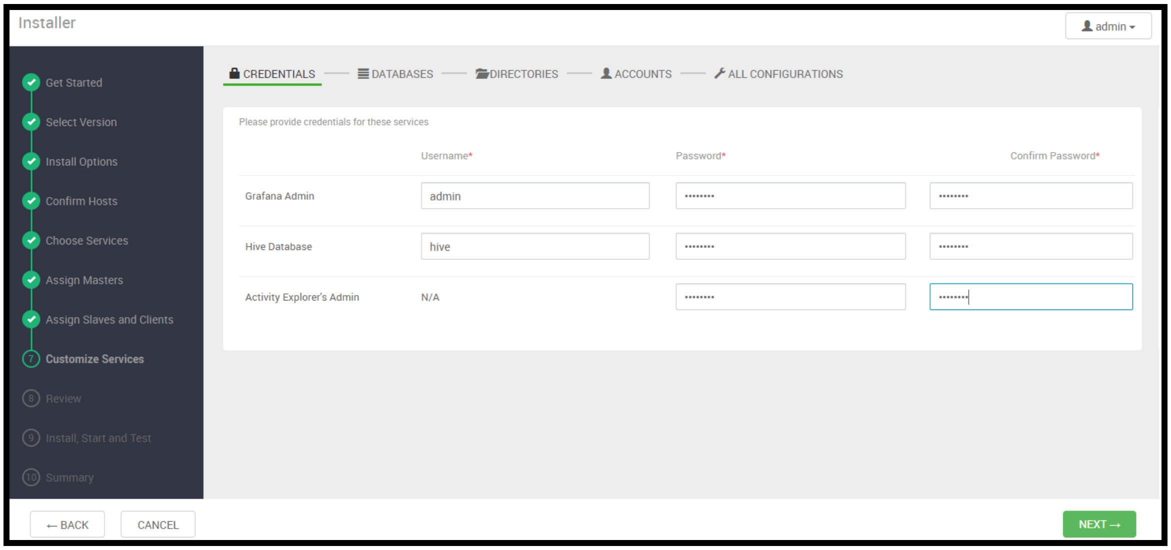
Assign slave and client components to hosts you want to run them on.
Hosts that are assigned master components are shown with ✱.
"Client" will install HDFS Client, YARN Client, MapReduce2 Client, Tez Client, Hive Client, HBase Client, Pig Client, ZooKeeper Client and Spark2 Client.

Host	all | none	all | none	all | none	all | none	all | none	all | none	all | none	all | none	all | none
ambari.interset.com✱	□ DataNode	□ NFSGateway	□ NodeManager	□ RegionServer	□ Phoenix Query Server	□ Supervisor	□ Livy for Spark2 Server	□ Spark2 Thrift Server	□ Client
master.interset.com✱	□ DataNode	□ NFSGateway	□ NodeManager	□ RegionServer	□ Phoenix Query Server	□ Supervisor	□ Livy for Spark2 Server	□ Spark2 Thrift Server	☑ Client
stream.interset.com✱	□ DataNode	□ NFSGateway	□ NodeManager	□ RegionServer	□ Phoenix Query Server	□ Supervisor	□ Livy for Spark2 Server	□ Spark2 Thrift Server	☑ Client
compute.interset.com	☑ DataNode	□ NFSGateway	☑ NodeManager	☑ RegionServer	□ Phoenix Query Server	☑ Supervisor	□ Livy for Spark2 Server	□ Spark2 Thrift Server	☑ Client
nifi.interset.com	□ DataNode	□ NFSGateway	□ NodeManager	□ RegionServer	□ Phoenix Query Server	□ Supervisor	□ Livy for Spark2 Server	□ Spark2 Thrift Server	☑ Client

Items per page: 10 ▾ 1 - 5 of 5 < >

16. On the **Assign Slaves and Clients** page, do the following:
    - Remove all roles from all nodes, and then assign the **Client** role to all node(s) except **Monitoring**.
    - Assign the **DataNode**, **NodeManager**, **RegionServer**, and **Supervisor** roles to the **Compute** node(s).
17. Click **Next**.

    The **Customized Services** page is displayed.

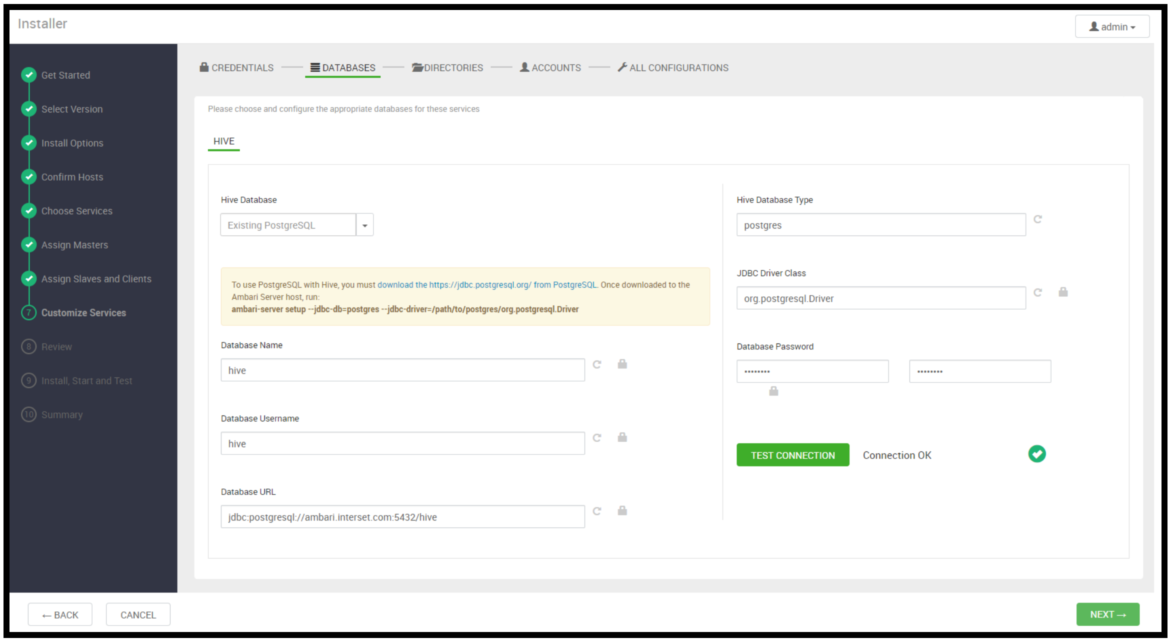


Please provide credentials for these services

	Username*	Password*	Confirm Password*
Grafana Admin	admin	••••••••	••••••••
Hive Database	hive	••••••••	••••••••
Activity Explorer's Admin	N/A	••••••••	••••••••

18. On the **Customized Services** page, in the **Credentials** tab, enter the following passwords:
    - a new password for Grafana Admin
    - the password you chose earlier for the Hive database
    - a new password for Activity Explorer's Admin
19. Click **Next**.

    The **Databases** tab is displayed.



Please choose and configure the appropriate databases for these services

Hive Database: Existing PostgreSQL

To use PostgreSQL with Hive, you must download the https://jdbc.postgresql.org/ from PostgreSQL. Once downloaded to the Ambari Server host, run:
ambari-server setup --jdbc-db=postgres --jdbc-driver=/path/to/postgres/org.postgresql.Driver

Database Name: hive

Database Username: hive

Database URL: jdbc:postgresql://ambari.interset.com:5432/hive

Hive Database Type: postgres

JDBC Driver Class: org.postgresql.Driver

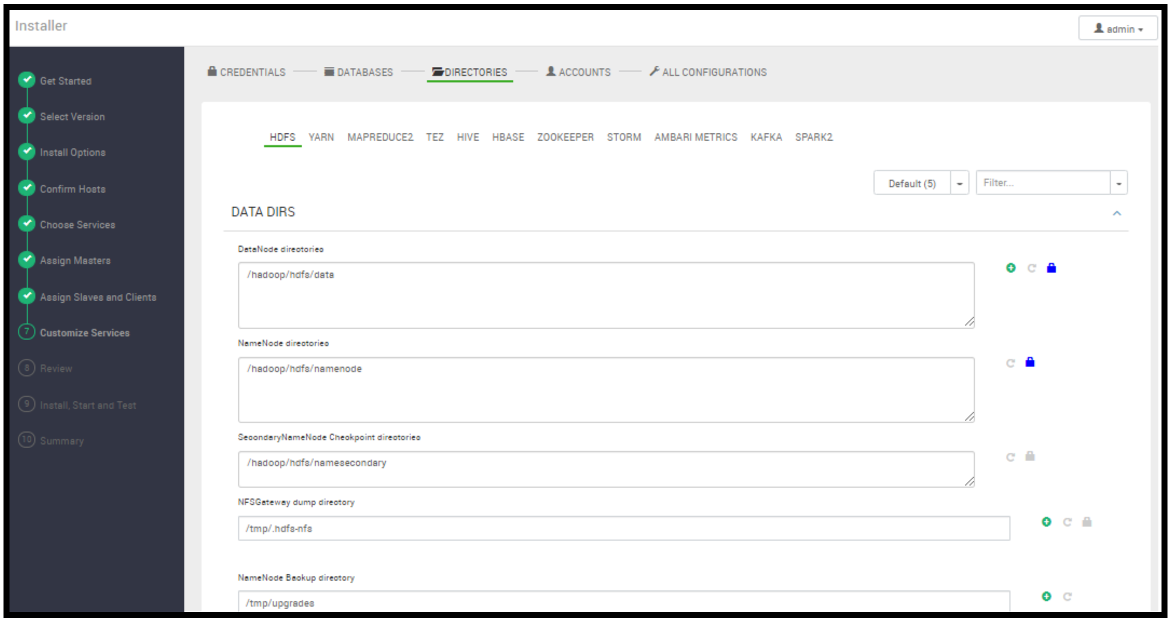
Database Password: •••••••• ••••••••

Connection OK ✓

20. In the **Hive** section, under **Hive Database**, select **Existing PostgreSQL**.
21. For the **Database Name** and **Database Username** fields, enter `hive`.
22. In the **Database URL** text box, ensure that the path reflects the FQDN of the **Monitoring** node.

23. In the **Database Password** fields, enter and confirm the password that you chose for the `DB user - Hive` in "Perform the Base Configuration of All Nodes" on page 31.
24. Click **Test Connection** to validate the remote database connection.
25. When the connection is successful, click **Next**.

    The **Directories** tab is displayed.



DATA DIRS

DataNode directories: /hadoop/hdfs/data

NameNode directories: /hadoop/hdfs/namenode

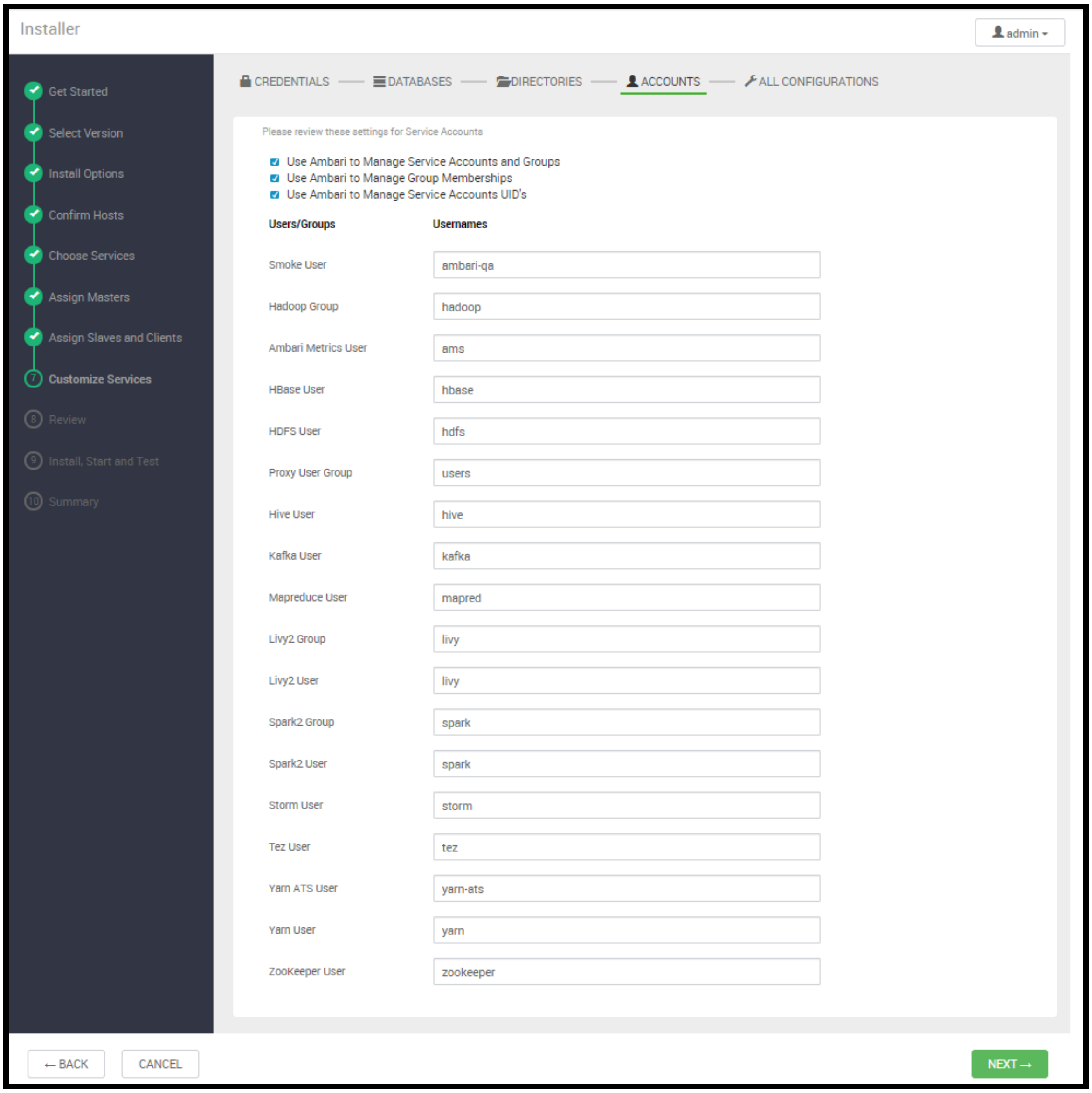
SecondaryNameNode Checkpoint directories: /hadoop/hdfs/namesecondary

NFSGateway dump directory: /tmp/.hdfs-nfs

NameNode Backup directory: /tmp/upgrades

26. On the **Directories** tab, click through each sub-tab (HDFS, Yarn, MapReduce2, and so on) and ensure that the directories listed are correct. In particular, if errors are displayed for HDFS and YARN, complete the following steps to correct them:
    a. Click the **HDFS** tab, and then do the following:
        - Set the path for **NameNode directories** to a data directory on a volume or partition where disk space is readily available
        - In the **DataNode directories** box, set the path to a data directory on a volume or partition where disk space is readily available
    b. Click the **YARN** tab, and then do the following:
        - In the **DATA DIRS** section, set the path for **YARN NodeManager Local directories** to **/hadoop/yarn/local**
        - In the **LOG DIRS** section, set the path for **YARN NodeManager Log directories** to **/hadoop/yarn/log**
27. When you are satisfied with the directory settings, click **Next**.

    The **Accounts** tab is displayed.



Please review these settings for Service Accounts

☑ Use Ambari to Manage Service Accounts and Groups
☑ Use Ambari to Manage Group Memberships
☑ Use Ambari to Manage Service Accounts UID's

Users/Groups	Usernames
Smoke User	ambari-qa
Hadoop Group	hadoop
Ambari Metrics User	ams
HBase User	hbase
HDFS User	hdfs
Proxy User Group	users
Hive User	hive
Kafka User	kafka
Mapreduce User	mapred
Livy2 Group	livy
Livy2 User	livy
Spark2 Group	spark
Spark2 User	spark
Storm User	storm
Tez User	tez
Yarn ATS User	yarn-ats
Yarn User	yarn
ZooKeeper User	zookeeper

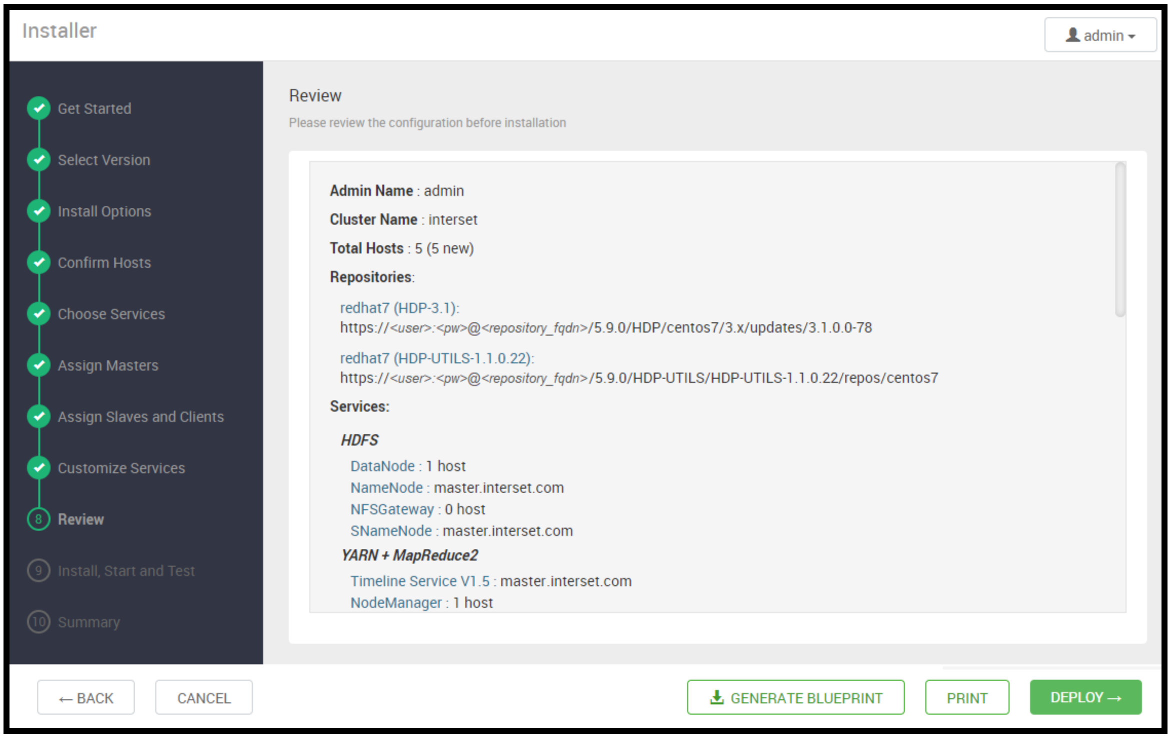
28. Ensure that default settings in the Accounts page appear correct, then click **Next**.

    The **All Configurations** tab is displayed.

29. The **All Configurations** tab gives you opportunity to review and revise the remaining configurations for your services. Browse through each configuration tab, hovering your cursor over each of the properties to displays a brief description of what the property does. Any service with configuration issues that require attention show up in the bell icon with the number of properties that need attention.

30. When you have finished in the **All Configurations** tab, click **Next**.

    The **Review** page is displayed.



Review

Please review the configuration before installation

Admin Name : admin

Cluster Name : interset

Total Hosts : 5 (5 new)

Repositories:

redhat7 (HDP-3.1):
https://<user>:<pw>@<repository_fqdn>/5.9.0/HDP/centos7/3.x/updates/3.1.0.0-78

redhat7 (HDP-UTILS-1.1.0.22):
https://<user>:<pw>@<repository_fqdn>/5.9.0/HDP-UTILS/HDP-UTILS-1.1.0.22/repos/centos7

Services:

HDFS

DataNode : 1 host
NameNode : master.interset.com
NFSGateway : 0 host
SNameNode : master.interset.com

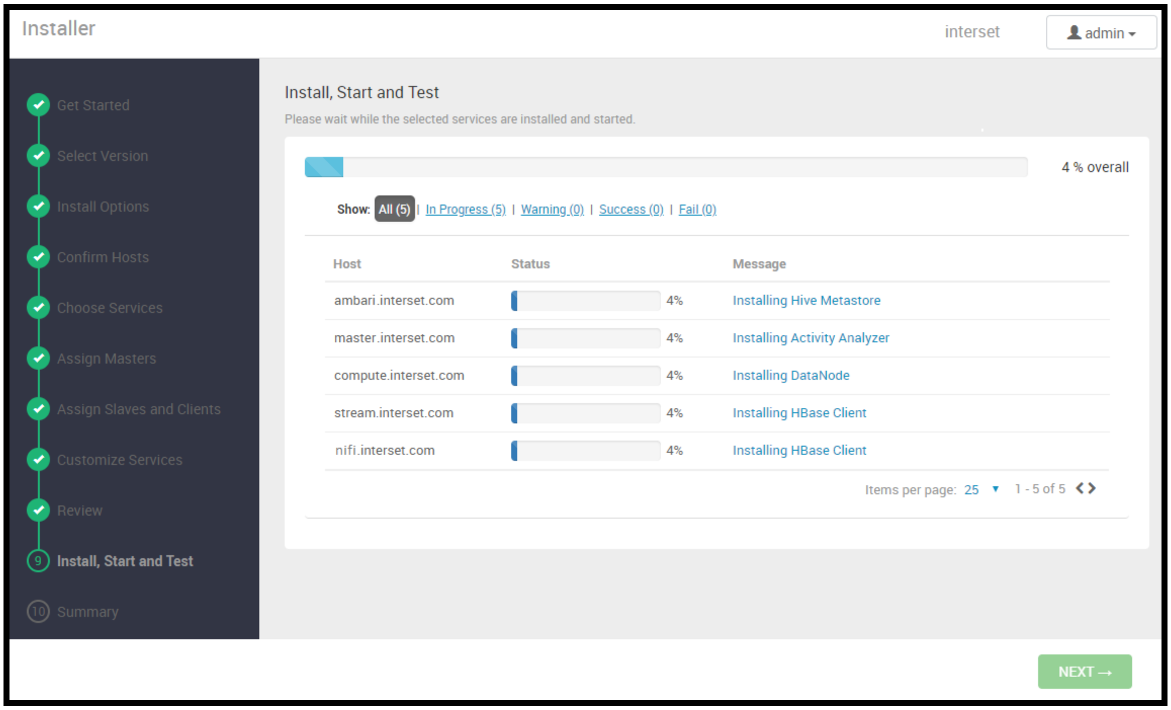
YARN + MapReduce2

Timeline Service V1.5 : master.interset.com
NodeManager : 1 host

31. On the **Review** page, check to make sure everything is correct. If you need to make changes, use the left navigation bar to return to the appropriate screen. To print your information for later reference, click **Print**. To export the blueprint for this cluster, click **Generate Blueprint**.

    When you are satisfied with your choices, click **Deploy**.

    The **Install, Start and Test** page is displayed, showing the progress of the installation.



Install, Start and Test

Please wait while the selected services are installed and started.

4 % overall

Show: All (5) | In Progress (5) | Warning (0) | Success (0) | Fail (0)

Host	Status	Message
ambari.interset.com	4%	Installing Hive Metastore
master.interset.com	4%	Installing Activity Analyzer
compute.interset.com	4%	Installing DataNode
stream.interset.com	4%	Installing HBase Client
nifi.interset.com	4%	Installing HBase Client

Items per page: 25 ▾ 1 - 5 of 5

The deployment will take about 15-45 minutes to complete.

When the installation is complete, this page also shows the warnings and errors encountered during the install. You can make corrections and retry the installation untill all issues are resolved.

32. When the installation is complete, click **Next**.The **Summary** page is displayed.
33. On the **Summary** page, click **Complete**.
34. In the **Services** list on the left side of the Ambari page, verify that all services are running; start any services that are not running.

> **Note:** Tez, Pig, and Slider will not be running, and this is expected. These components are installed by Ambari by default but are not required by Interset.

Change the Ambari Administrator Password

Ambari uses the following default administrator user credentials:

```
Username: admin
Password: admin
```

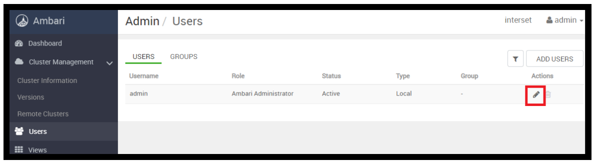
To prevent an outside attacker or malicious insider from gaining access to this account, it is imperative that you change the admin user account password as soon as your Interset cluster is installed and running.

Important: Ensure that you select a new password in accordance with the National Institute of Standards and Technology (NIST) guidelines.

1. Log in to Apache Ambari:
   - Open a Web browser, and go to **http://<*ambari_fqdn*>:8080**, where **<*ambari_fqdn*>** is the name of your Ambari node.
   - Type the following login credentials, then press **Enter**.

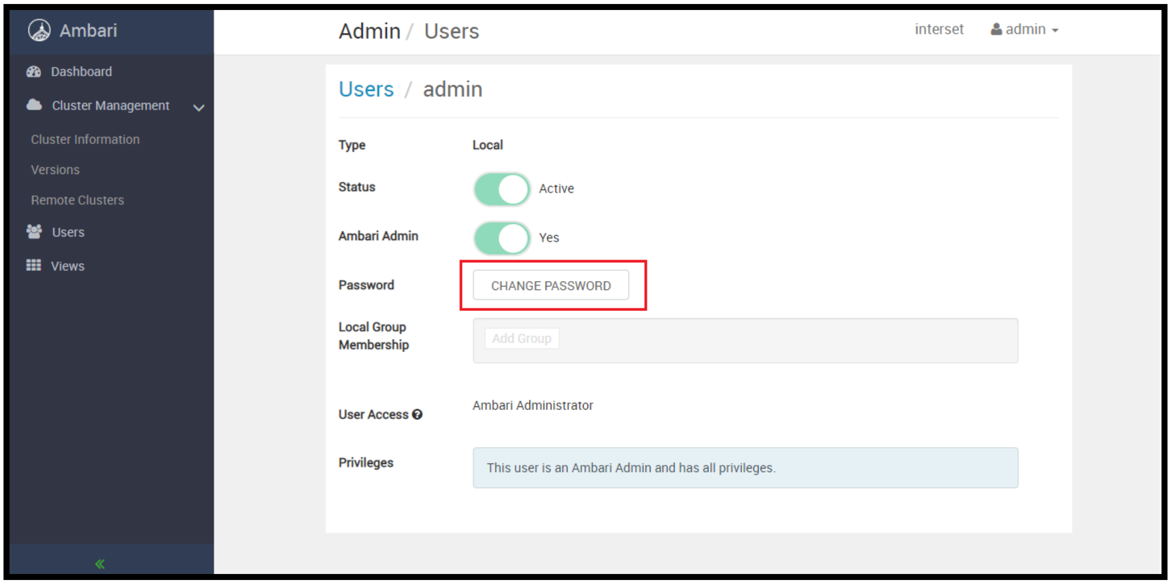
     Username: **admin**

     Password: **admin**
2. From the **Admin** drop-down menu, select **Manage Ambari**.
3. In the menus that appear on the left, select **Users**.
4. In the **Admin / Users** page, click pencil icon next to the **admin** user.



Username	Role	Status	Type	Group	Actions
admin	Ambari Administrator	Active	Local	-	✎

5. On the **Users / admin** page, click **Change Password**.



Type: Local

Status: Active

Ambari Admin: Yes

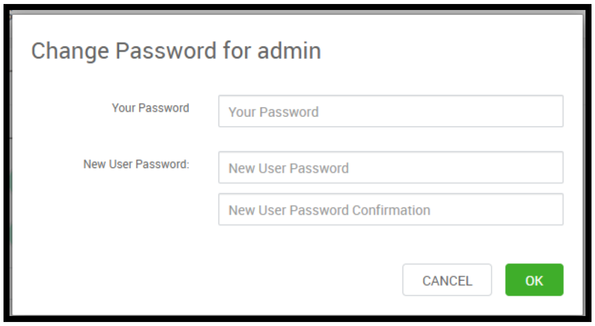
Password

Local Group Membership

User Access: Ambari Administrator

Privileges: This user is an Ambari Admin and has all privileges.

6. In the **Change Password for admin** dialog box, do the following:
   - In the **Your Password** box, enter the current admin user password.
   - In the **New User Password** box, enter the new, NIST-standard password for the admin user.
   - In the **New User Password Confirmation** box, re-enter the new password.
   - Click **OK**.


Change Password for admin

Your Password: Your Password

New User Password: New User Password

New User Password Confirmation

A message appears, confirming the password change.

Configure Ambari Managed Services

1. In a command console on the Monitoring node, run the following commands to navigate to the **/opt/interset/installer** directory and launch the Interset installation menu.

```
cd /opt/interset/installer
```

```
./install.sh
```

2. From the available installation options, type **3** to select option **[3] Ambari managed service(s) configuration**, and then press **Enter**.

```
Please select an installation option from the list above: 3
```

When the configuration is complete, the expected result is:

```
Ambari Configuration Complete!
```

⚠ To continue your installation, proceed to "Generate the TLS Certificates" on page 62.

Install Cloudera Manager

> If you are installing Ambari, skip this section and return to "Install the Ambari Server" on page 34.

Having just completed the base configuration of all nodes in your Interset cluster, you should still be in a command console on the Monitoring node.

1. In a command console on the Monitoring node, run the following commands to navigate to the **/opt/interset/installer** directory and launch the Interset installation menu.

   ```
   cd /opt/interset/installer
   ```

   ```
   ./install.sh
   ```

2. From the available installation options, type **2** to select installation **[2] CDH node installation**, and then press **Enter**.

   ```
   Please select an installation option from the list above: 2
   ```

3. Enter the password for the Cloudera SCM database:

   ```
   Please enter your existing password for DB user - CDH Monitoring (SCM DB):
   ```

The step installs Cloudera Manager server on the Monitoring node, configures the SCM database, and installs the Cloudera agent on the other nodes. It may take several minutes to complete, during which time a series of status messages appear in the command console.

After Cloudera Manager has been installed, the expected result is:

```
Cloudera Manager URL is http://<monitoring_node_fqn>:7180
Please log into Cloudera Manager to complete your CDH cluster installation.
Cloudera Install Complete!
```

> At this point in the Interset installation, you leave the command console temporarily and launch the Cloudera Manager user interface to install the CDH cluster.

Install the Cloudera Cluster

✓ **Tip:** The Cloudera Manager server will be running, as it was started at the end of the installation process in the preceding section.

The Cloudera Manager Server URL takes the form **http://<*monitoring_server_fqdn*>:<*port*>** where **monitoring_server_fqdn** is the fully qualified domain name (FQDN) or IP address of the Monitoring node, and **port** is the port configured for the Cloudera Manager Server. The default port is 7180.

Log into the Cloudera Manager Console:

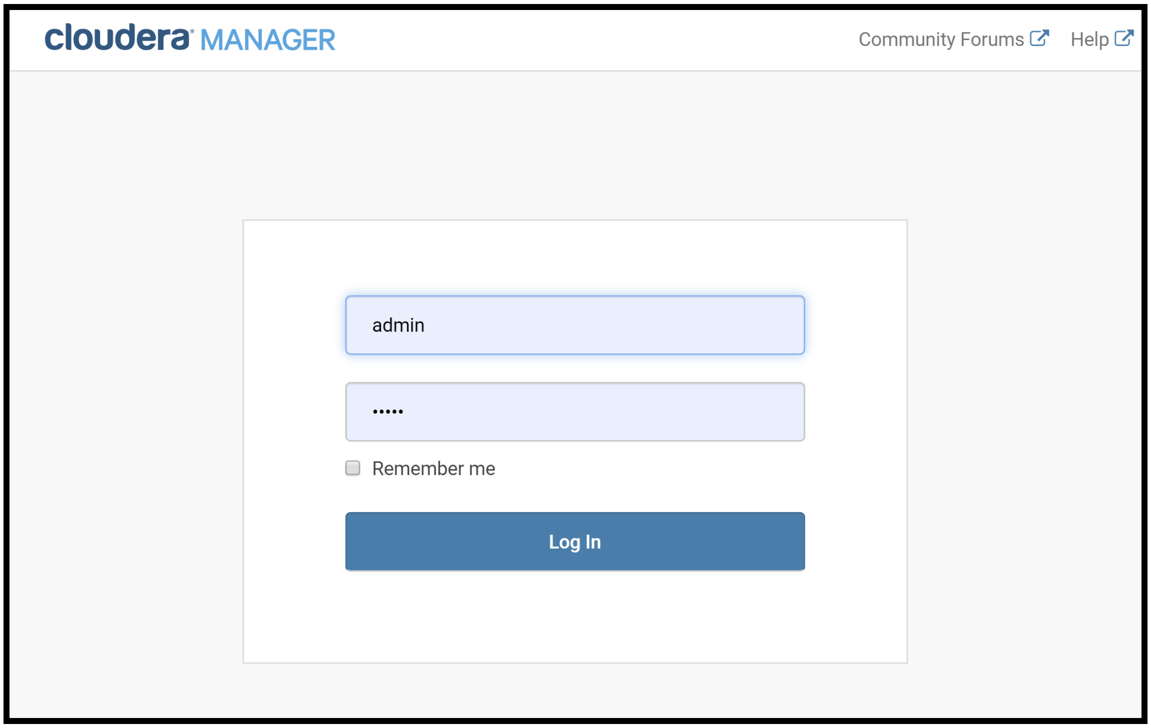
1. Wait several minutes for the Cloudera Manager Server to start.

   To observe the startup process, run the following command on the Monitoring node.

```
tail -f /var/log/cloudera-scm-server/cloudera-scm-server.log
```

2. In a Web browser, navigate to **http://<*monitoring_server_fqdn*>:7180**.

   The login screen for **Cloudera Manager Admin Console** is displayed.



3. Log in to the **Cloudera Manager Admin Console**.

   The default credentials are: **Username:** `admin` **Password:** `admin`.

> Cloudera Manager does not support changing the **admin** username for the installed account. You can change the password using Cloudera Manager after you run the installation wizard. Although you cannot change the **admin** username, you can add a new user, assign administrative privileges to the new user, and then delete the default **admin** account.

After you log in, the **Welcome** page appears.

4. On the **Welcome** page, click **Continue**.

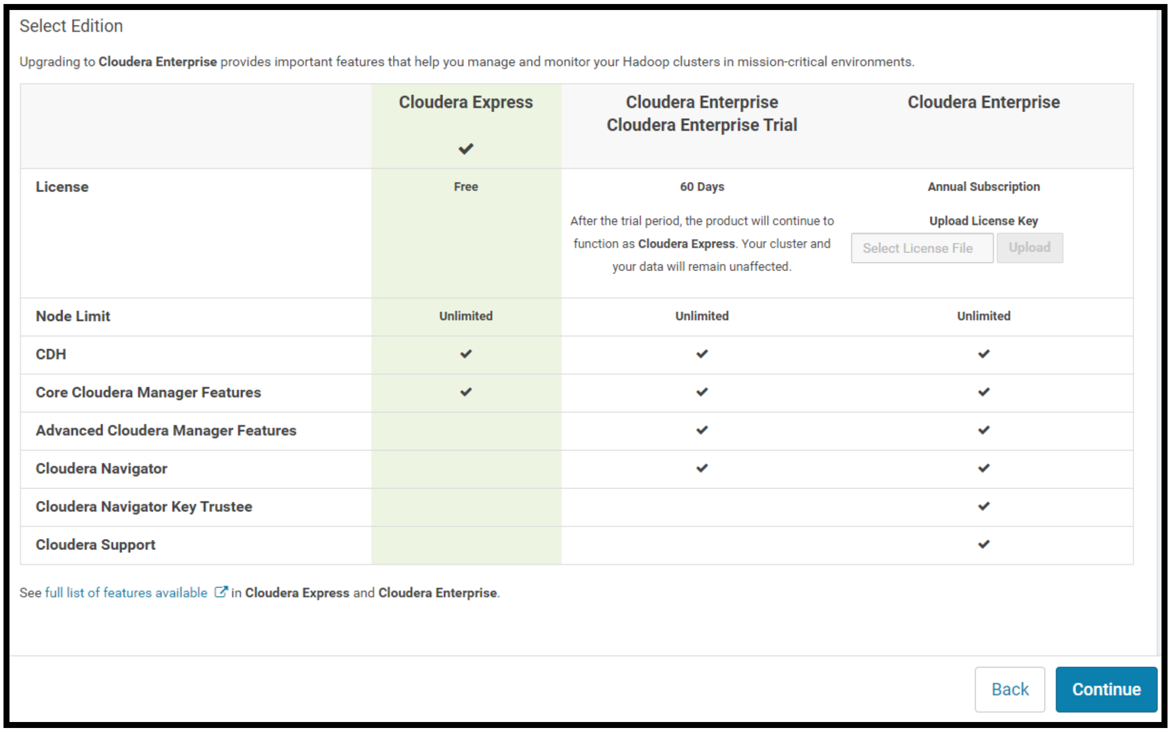
   The **Cloudera Manager End User License Terms and Conditions** page is displayed.

5. Read the terms and conditions, and then select **Yes, I accept the End User License Terms and Conditions.** to accept them.

6. Click **Continue**.

   The **Select Edition** page is displayed.

7. Choose to install **Cloudera Express**.



Select Edition

Upgrading to **Cloudera Enterprise** provides important features that help you manage and monitor your Hadoop clusters in mission-critical environments.

	Cloudera Express ✓	Cloudera Enterprise Cloudera Enterprise Trial	Cloudera Enterprise
License	Free	60 Days After the trial period, the product will continue to function as **Cloudera Express**. Your cluster and your data will remain unaffected.	Annual Subscription Upload License Key Select License File Upload
Node Limit	Unlimited	Unlimited	Unlimited
CDH	✓	✓	✓
Core Cloudera Manager Features	✓	✓	✓
Advanced Cloudera Manager Features		✓	✓
Cloudera Navigator		✓	✓
Cloudera Navigator Key Trustee			✓
Cloudera Support			✓

See full list of features available in **Cloudera Express** and **Cloudera Enterprise**.

8. Click **Continue** to proceed with the installation.

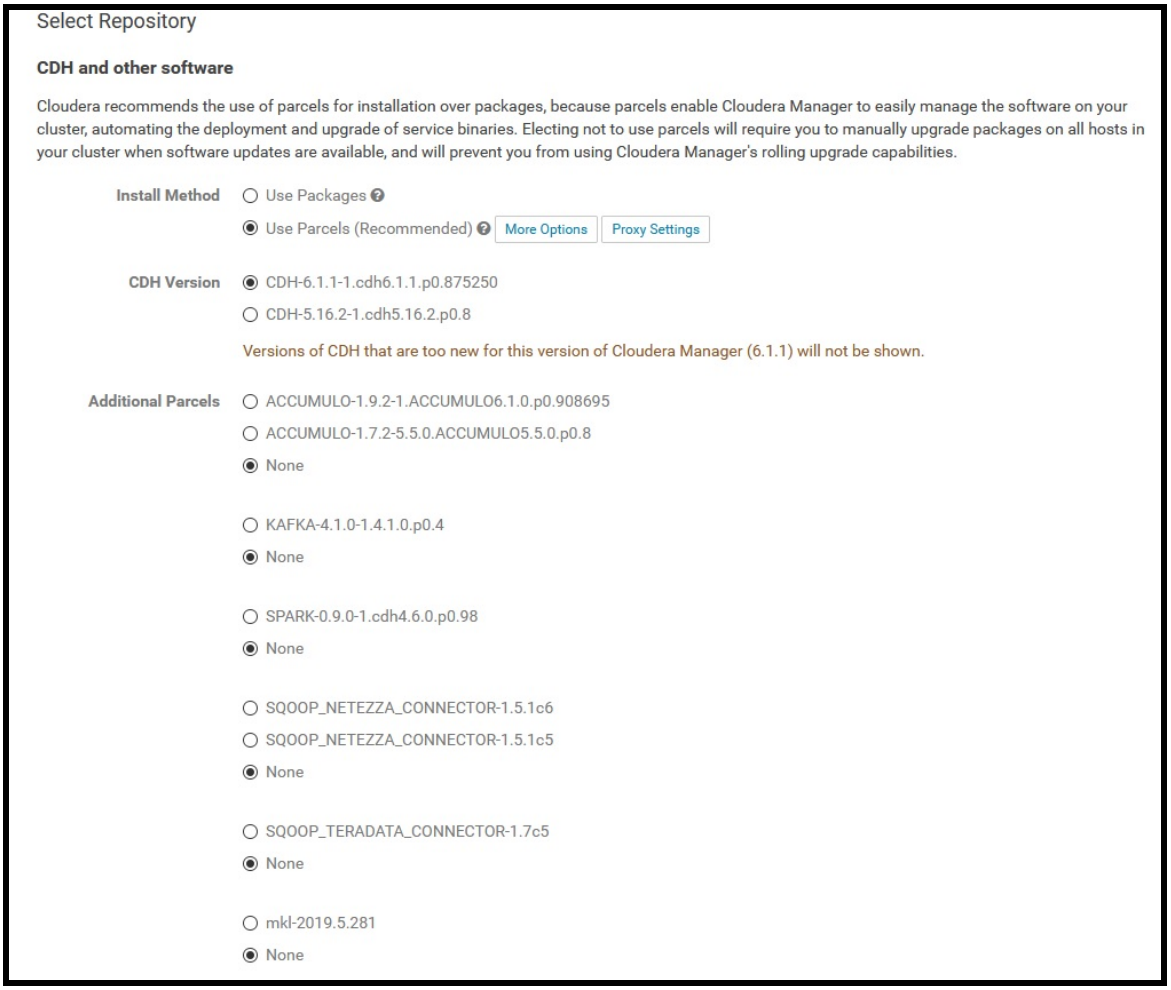
   The **Welcome** page of the **Add Cluster - Installation** wizard appears.

9. Click **Continue**.

10. On the **Specify Hosts** screen, click the **Currently Managed Hosts** tab, and ensure that the the **Monitoring**, **Master**, **Compute**, **Stream**, and **NiFi** nodes in your cluster are listed and selected as required.

11. Verify that the number of hosts shown matches the number of hosts where you want to install services, clearing any host entries that do not apply.
12. Click **Continue** to proceed with the installation.

    The **Select Repository** screen is displayed.



Select 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)

CDH Version ● CDH-6.1.1-1.cdh6.1.1.p0.875250
○ CDH-5.16.2-1.cdh5.16.2.p0.8

Versions of CDH that are too new for this version of Cloudera Manager (6.1.1) will not be shown.

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

○ KAFKA-4.1.0-1.4.1.0.p0.4
● None

○ SPARK-0.9.0-1.cdh4.6.0.p0.98
● None

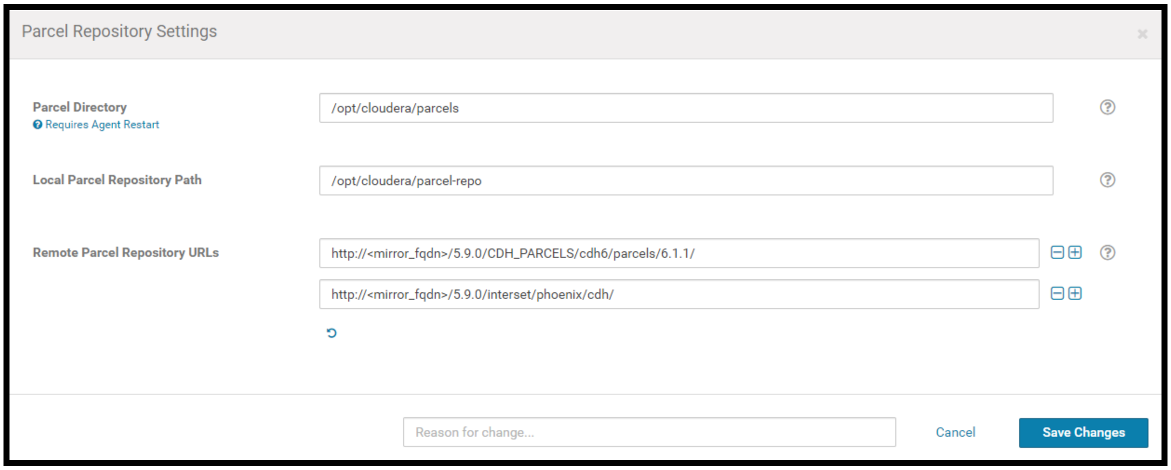
○ SQOOP_NETEZZA_CONNECTOR-1.5.1c6
○ SQOOP_NETEZZA_CONNECTOR-1.5.1c5
● None

○ SQOOP_TERADATA_CONNECTOR-1.7c5
● None

○ mkl-2019.5.281
● None

13. In the **CDH and other software** section, next to **Install Method**, select **Use Parcels (Recommended)**, and then click **More Options**.

    The **Parcel Repository Settings** screen is displayed.



Parcel Repository Settings

Parcel Directory
Requires Agent Restart
/opt/cloudera/parcels

Local Parcel Repository Path
/opt/cloudera/parcel-repo

Remote Parcel Repository URLs
http://<mirror_fqdn>/5.9.0/CDH_PARCELS/cdh6/parcels/6.1.1/
http://<mirror_fqdn>/5.9.0/interset/phoenix/cdh/

14. In the **Remote Parcel Repository URLs** field, click ⊟ to remove any URLs that already present. These will not be used.

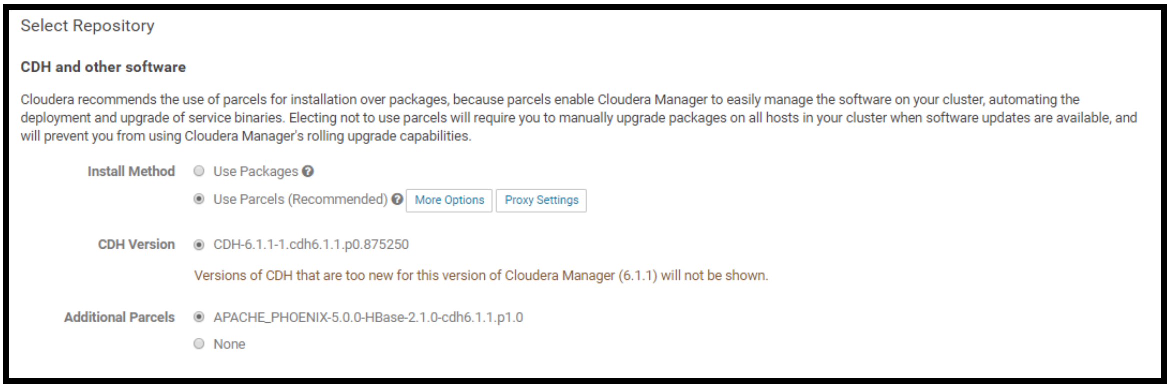
15. In the **Remote Parcel Repository URLs** field, click ⊞, and then enter the URL of the relevant repository substituting the correct value (either your offline repository, or the Interset online repository) for **<mirror_fqdn>**.

Parcel	Installation URL
CDH	http://<mirror_fqdn>/5.9.2/CDH_PARCELS/cdh6/parcels/6.1.1/
Phoenix	http://<mirror_fqdn>/5.9.2/interset/phoenix/cdh/

The URLs you specify are added to the list of repositories listed in the **Configuring Cloudera Manager Server Parcel Settings** page, and each parcel is added to the list of parcels on the **Select Repository** page.

If you have multiple repositories configured, you see all the unique parcels contained in all your repositories.

16. Click **Save Changes**. You are returned to the **Select Repository** screen.



Select 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)

CDH Version ● CDH-6.1.1-1.cdh6.1.1.p0.875250

Versions of CDH that are too new for this version of Cloudera Manager (6.1.1) will not be shown.

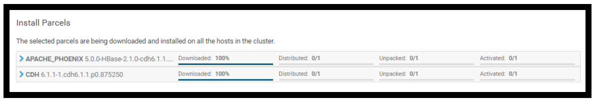
Additional Parcels ● APACHE_PHOENIX-5.0.0-HBase-2.1.0-cdh6.1.1.p1.0
○ None

17. Select **CDH-6.1.1-1.cdh6.1.1.p0.875250** for **CDH Version**.

18. In the **Additional Parcels** list, select the Apache Phoenix parcel.

19. Click **Continue**.

    The **Install Parcels** screen is displayed, and shows the progress of the installation.

    Install Parcels

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

APACHE_PHOENIX 5.0.0-HBase-2.1.0-cdh6.1.1....	Downloaded: **100%**	Distributed: **0/1**	Unpacked: **0/1**	Activated: **0/1**
CDH 6.1.1-1.cdh6.1.1.p0.875250	Downloaded: **100%**	Distributed: **0/1**	Unpacked: **0/1**	Activated: **0/1**

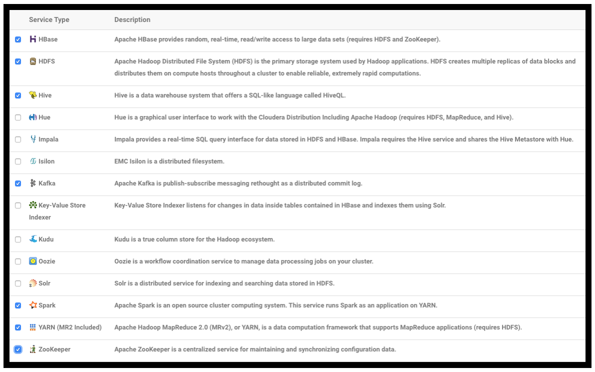
20. When the installation has finished, click **Continue**. The **Inspect Hosts for Correctness** screen is displayed.
21. The **Validations** list is shown after a few moments. You can ignore errors regarding swappiness and transparent huge pages.
22. Review the results and then click **Finish** to complete the installation.

Add Services

At this stage, the **Select Services** screen is displayed. You can also select **Add Services** from the **Clusters** menu on the main page of Cloudera Manager.

1. On the first page of the **Select Services** wizard, click **Custom Services**.

   The list of services is displayed.

	Service Type	Description
☑	HBase	Apache HBase provides random, real-time, read/write access to large data sets (requires HDFS and ZooKeeper).
☑	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.
☑	Hive	Hive is a data warehouse system that offers a SQL-like language called HiveQL.
□	Hue	Hue is a graphical user interface to work with the Cloudera Distribution Including Apache Hadoop (requires HDFS, MapReduce, and Hive).
□	Impala	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.
□	Isilon	EMC Isilon is a distributed filesystem.
☑	Kafka	Apache Kafka is publish-subscribe messaging rethought as a distributed commit log.
□	Key-Value Store Indexer	Key-Value Store Indexer listens for changes in data inside tables contained in HBase and indexes them using Solr.
□	Kudu	Kudu is a true column store for the Hadoop ecosystem.
□	Oozie	Oozie is a workflow coordination service to manage data processing jobs on your cluster.
□	Solr	Solr is a distributed service for indexing and searching data stored in HDFS.
☑	Spark	Apache Spark is an open source cluster computing system. This service runs Spark as an application on YARN.
☑	YARN (MR2 Included)	Apache Hadoop MapReduce 2.0 (MRv2), or YARN, is a data computation framework that supports MapReduce applications (requires HDFS).
☑	ZooKeeper	Apache ZooKeeper is a centralized service for maintaining and synchronizing configuration data.

2. Select the following services to install:
   - HBase
   - HDFS
   - Hive
   - Kafka
   - Spark
   - YARN (MR2 Included)
   - ZooKeeper
3. Click **Continue** to customize the assignment of role instances to hosts. The **Assign Roles** page is displayed.
4. Click a field below a role to select the host(s) for that role.

   We recommend the following host assignments:

- Kafka broker -> **Stream** node
- HBase Master -> **Master** node
- HBase Region Server -> Set to **Same as DataNode**
- HDFS Name Node -> **Master** node
- HDFS Secondary Name Node -> **Master** node
- HDFS Balancer -> Do not assign a node.
- HDFS DataNode -> **Compute** node(s)
- HDFS HttpFS -> Do not assign a node
- HDFS NFS Gateway -> Do not assign a node
- Hive Gateway -> **Master** and **Compute** nodes
- Hive Metastore Server -> **Master** node
- Hive WebHCat Server -> **Master** node
- Hive HiveServer2 -> **Master** node
- Hive (all remaining) -> **Monitoring** node
- Cloudera Management Services (all except Activity Monitor, Reports Manager, and Telemetry Publisher) -> **Monitoring** node
- Spark History Server -> **Master** node
- Spark Gateway --> **Master** and **Compute** nodes
- YARN Resource Manager -> **Master** node
- YARN JobHistory Server -> **Master** node
- YARN NodeManager -> Set to **Same as DataNode**
- Zookeeper Server -> **Master** node

> **Important**:You must assign HBase Gateway to the **NiFi** node(s). There is no HBase Gateway on the Wizard.This needs to be done after the completion of Step 3 (**Configure Cloudera Managed Services**)

5. Click **View By Host** for an overview of the role assignment by hostname ranges.

> **Important**: Do not assign the following roles:
>
> - HBase REST Server
> - HBase Thrift Server
> - HttpFS
> - NFS Gateway
> - Cloudera Activity Monitor, Reports Manager, and Telemetry Publisher

6. When you have finished with the assignments, click **Continue**. The **Setup Database** page is displayed.
7. Ensure the following properties are set for Hive:
   - **Database Host Name**: **Postgres** node FQDN
   - **Database Type**: PostgreSQL
   - **Database Name**: metastore
   - **Username**: metastore
   - **Password**: The password you entered for the **Hive** database user in "Perform the Base Configuration of All Nodes" on page 31.
8. Click **Test Connection**.

   If the test succeeds, click **Continue**; otherwise, check and correct the information you provided for the database and then try the test again.

   The **Review Changes** screen is displayed.

Review Configuration Changes and Start Services

1. Review the configuration changes to be applied.

   Make any changes you require for the HDFS DataNode path or other paths. The file paths required vary based on the services to be installed.

   > **Warning:** Do not place DataNode data directories on NAS devices. When resizing an NAS, block replicas can be deleted, which will result in reports of missing blocks.

2. Click **Continue**.

   The wizard starts the services.
3. When all of the services are started, click **Continue**.

   A message is displayed indicating that your cluster has been successfully started.
4. Click **Finish** to proceed to the **Cloudera Manager Admin Console Home Page**.

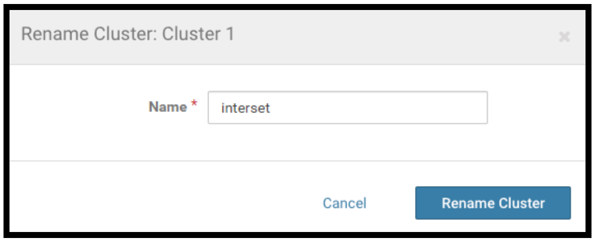
Rename the Cloudera Cluster

> You must rename your cluster before you can continue with the Interset installation.

To rename the cluster:

1. On the **Cloudera Manager Admin Console Home Page**, click the down-arrow next to the cluster name (usually **Cluster1** by default).
2. Click **Rename Cluster**. The **Rename Cluster** dialog is displayed.

3. In the **Name** field, enter `interset` (lowercase) and then click **Rename Cluster**.



Rename Cluster: Cluster 1

Name * interset

You are returned to the **Cloudera Manager Admin Console Home Page**.

Change the Default Administrator Password

As soon as possible, change the default administrator password:

1. On the **Cloudera Manager Admin Console Home Page**, click the logged-in username at the far right of the top navigation bar, and then select **Change Password**.
2. Enter the current password and a new password twice, and then click **OK**.

Configure Cloudera Managed Services

1. In a command console on the Monitoring node, run the following commands to navigate to the **/opt/interset/installer** directory and launch the Interset installation menu.

```
cd /opt/interset/installer
```

```
./install.sh
```

2. From the available installation options, type **3** to select option **[3] CDH managed service(s) configuration**, and then press **Enter**.

```
Please select an installation option from the list above: 3
```

When the configuration is complete, the expected result is:

```
Cloudera Configuration Complete!
```

Assign Hbase Gateway Service to the NiFi Node

At this stage, to assign HBase Gateway role to the NiFi node follow these steps:

1. On CDH Manager page, go to the "**interset**" cluster services and select "**HBase**" service
2. From HBase service "**Action**"; select "**Add Role Instances**"
3. Set HBase Gateway to the **NiFi** node, and then click the"**Continue**" button
4. After this is completed, on the "interset" cluster "HBase" service, click ⇨ to **deploy client configuration**. ..

5. To verify it was deployed, click "**Hosts**" > "**Roles**"
6. You will see "**HG**" (HBase Gateway) symbol role assigned to the **NiFi** node.

Note: :For more details refer to Cloudera documentation at: https://docs.cloudera.com/documentation/enterprise/latest/topics/cm_ mc_ role_ instances.html

To continue your installation, proceed to "Generate the TLS Certificates" on page 62.

Generate the TLS Certificates

1. In a command console on the Monitoring node, run the following commands to navigate to the **/opt/interset/installer** directory and launch the Interset installation menu.

   ```
   cd /opt/interset/installer
   ```

   ```
   ./install.sh
   ```

2. From the available installation options, type **4** to select installation **[4] Generate TLS Certificate**, and then press **Enter**.

   ```
   Please select an installation option from the list above: 4
   ```

3. Enter and confirm a password for the TLS key store:

   ```
   Please enter new password for tls:keystore_password:
   ```

4. Enter a confirm a password for the TLS trust store:

   ```
   Please enter new password for tls:truststore_password:
   ```

When the certificate creation has completed, the expected result is:

```
=========== function main_create_tls_cert Complete! ===========
```

Configure the Key Distribution Centre (KDC)

This step installs the KDC servers and creates the Kerberos server instance.

> **Tip:** We currently recommend **key based / certificate based** authentication for setting up a secure Interset environment. The section below will guide you on configuring KDC.

For more information about KDC and Kerberos for Interset, contact Micro Focus Interset Support at interset.support@microfocus.com.

1. In a command console on the Monitoring node, run the following commands to navigate to the **/opt/interset/installer** directory and launch the Interset installation menu.

   ```
   cd /opt/interset/installer
   ```

   ```
   ./install.sh
   ```

2. From the available installation options, type **5** to select installation step **[5] Set up KDC**, and then press **Enter**.

   ```
   Please select an installation option from the list above: 5
   ```

3. Enter the realm for Kerberos to use, or press Enter to use the realm shown:

   ```
   Enter Realm (<realm>):
   ```

4. Enter and confirm a new password for the Kerberos master user:

   ```
   Please enter new password for kerberos:master:
   ```

5. Enter and confirm a new password for the Kerberos root user:

   ```
   Please enter new password for kerberos:root:
   ```

When the configuration is complete, the expected result is similar to the following:

```
2019-09-17 20:52:42: Use following information in Kerberos configurations
2019-09-17 20:52:42: REALM = <realm>.COM
2019-09-17 20:52:42: admin id = root/admin
2019-09-17 20:52:42: kdc host = <monitoring_node_fqdn>
2019-09-17 20:52:42: Kerberos Encryption types = aes256-cts-hmac-sha1-96
=========== Complete! ===========
```

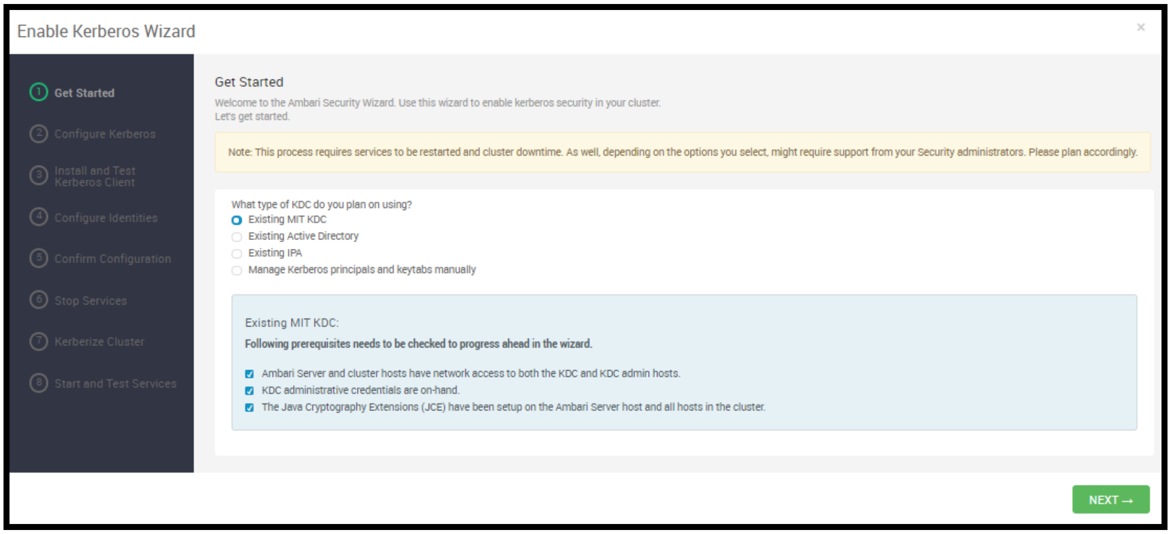
> - To continue your installation on Ambari, proceed to "Configure Kerberos on Ambari" on page 64.
> - To continue your installation on Cloudera, proceed to "Configure Kerberos on Cloudera" on page 69.

Configure Kerberos on Ambari

1. Log in to Apache Ambari:
   - Open a Web browser, and go to **http://<*ambari_fqdn*>:8080**, where **<*ambari_fqdn*>** is the name of your Ambari node.
   - Type the following login credentials, then press **Enter**.

     Username: **admin**

     Password: **admin**
2. From the left-hand menu, click **Cluster Admin** and then click **Kerberos**.
3. Click **Enable Kerberos**.
4. On the **Getting Started** page, select **Existing MIT KDC**, and confirm that the prerequisites are met.



Get Started

Welcome to the Ambari Security Wizard. Use this wizard to enable kerberos security in your cluster.
Let's get started.

Note: This process requires services to be restarted and cluster downtime. As well, depending on the options you select, might require support from your Security administrators. Please plan accordingly.

What type of KDC do you plan on using?

● Existing MIT KDC
○ Existing Active Directory
○ Existing IPA
○ Manage Kerberos principals and keytabs manually

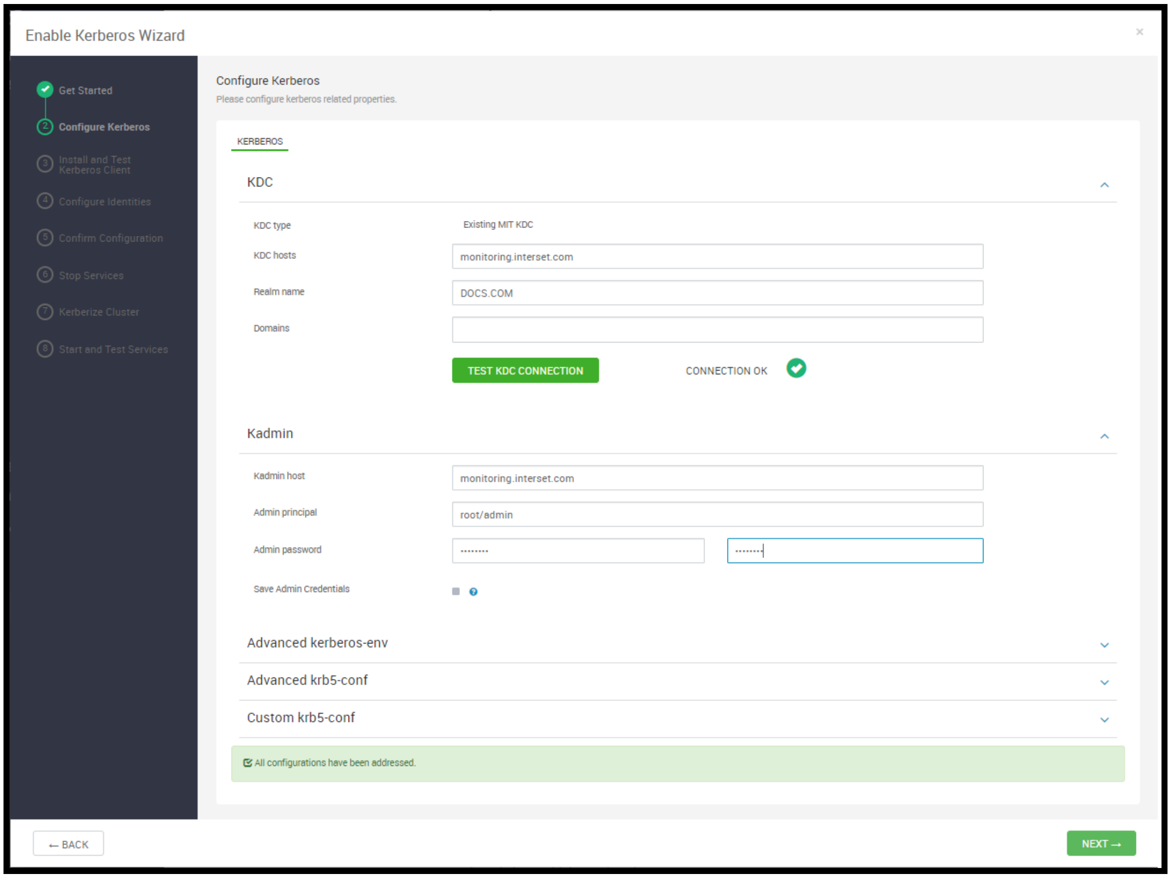
Existing MIT KDC:

Following prerequisites needs to be checked to progress ahead in the wizard.

☑ Ambari Server and cluster hosts have network access to both the KDC and KDC admin hosts.
☑ KDC administrative credentials are on-hand.
☑ The Java Cryptography Extensions (JCE) have been setup on the Ambari Server host and all hosts in the cluster.

5. Click **Next**.

6. On the **Configure Kerberos** page, provide information about the KDC and admin account.



Configure Kerberos

Please configure kerberos related properties.

KDC

KDC type: Existing MIT KDC

KDC hosts: monitoring.interset.com

Realm name: DOCS.COM

Domains:

Kadmin

Kadmin host: monitoring.interset.com

Admin principal: root/admin

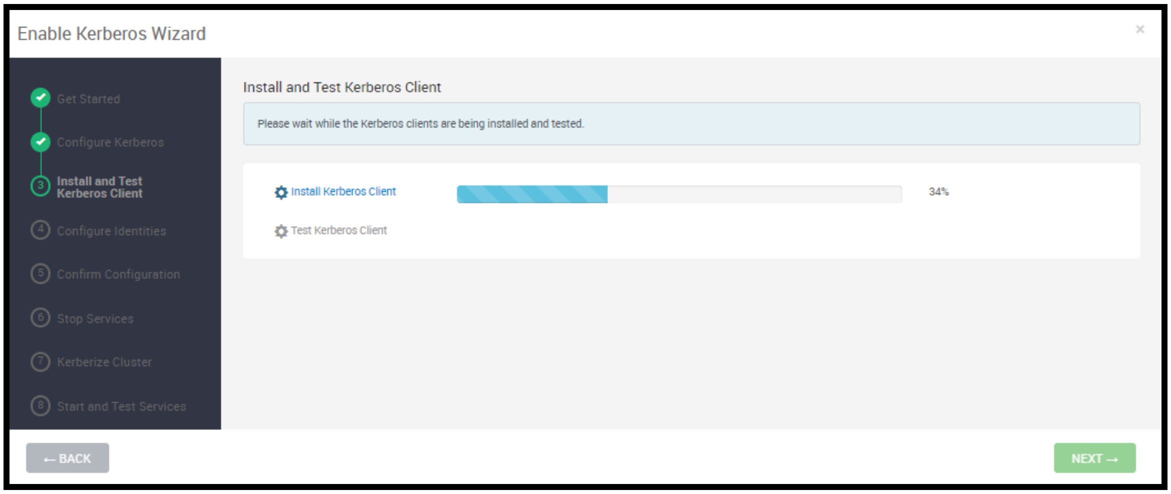
Admin password: •••••••• ••••••••

Save Admin Credentials: □

All configurations have been addressed.

7. Test the KDC connection, and if it is successful, click **Next** to install the Kerberos client. If the connection is unsuccessful, verify the host(s) and realm and try again.

8. The **Install and Test Kerberos Client** page shows you the progress of the installation, but you can also see the progress in the **/var/log/ambari-server/ambari-server.log** file.



Install and Test Kerberos Client

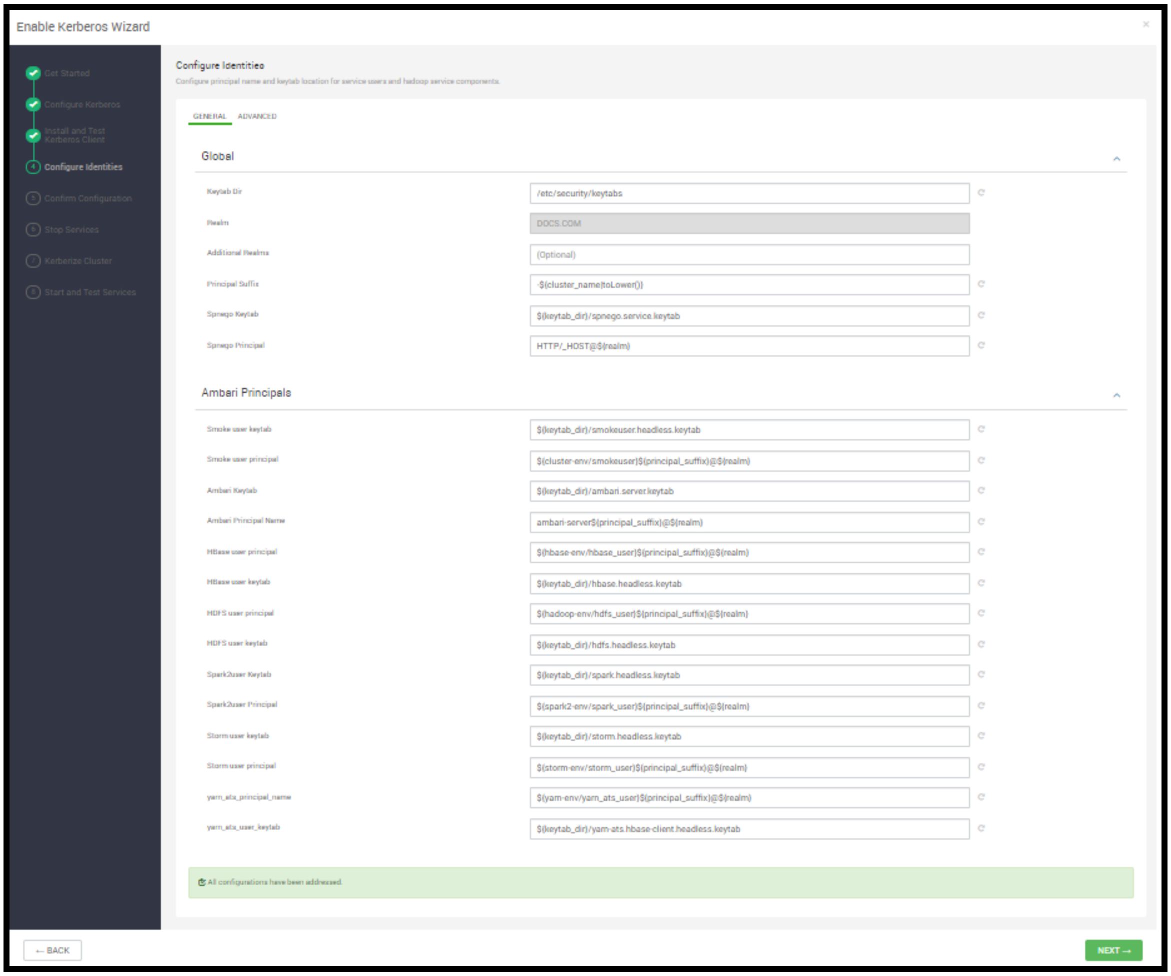
Please wait while the Kerberos clients are being installed and tested.

Install Kerberos Client 34%

Test Kerberos Client

The Kerberos clients are installed on the hosts and the access to the KDC is tested by verifying that Ambari can create a principal, generate a keytab and distribute that keytab.

9. Click **Next**.
10. On the **Configure Identities** page, verify the Kerberos identities that are used by Hadoop, and modify them as required.



Configure Identities

Configure principal name and keytab location for service users and hadoop service components.

Global

Keytab Dir	/etc/security/keytabs
Realm	DOCS.COM
Additional Realms	(Optional)
Principal Suffix	-${cluster_name|toLower()}
Spnego Keytab	${keytab_dir}/spnego.service.keytab
Spnego Principal	HTTP/_HOST@${realm}

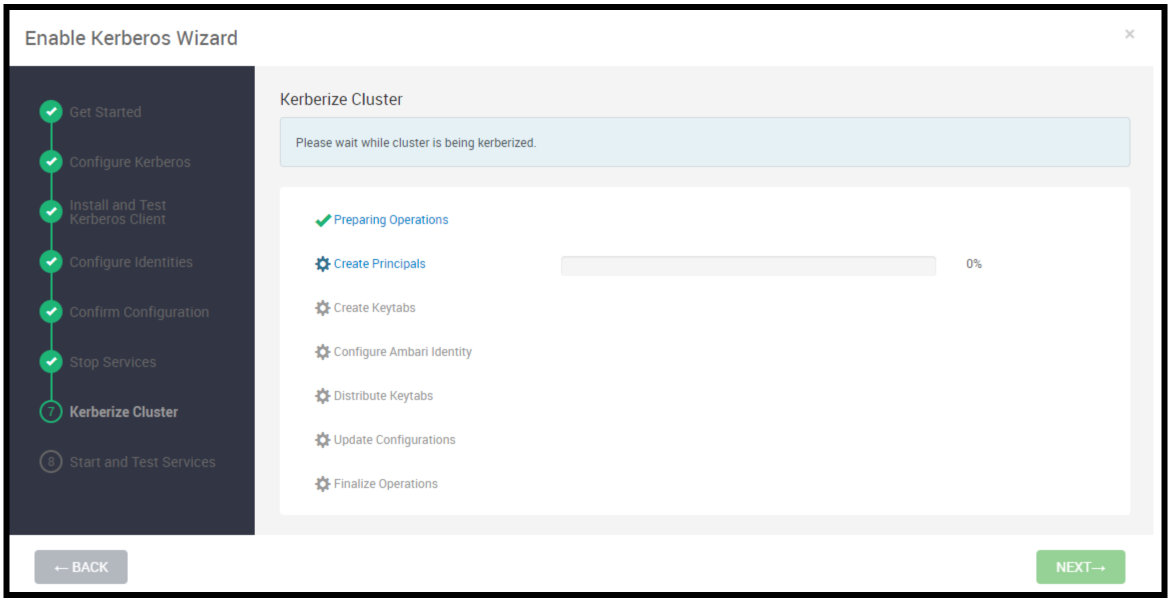
Ambari Principals

Smoke user keytab	${keytab_dir}/smokeuser.headless.keytab
Smoke user principal	${cluster-env/smokeuser}${principal_suffix}@${realm}
Ambari Keytab	${keytab_dir}/ambari.server.keytab
Ambari Principal Name	ambari-server${principal_suffix}@${realm}
HBase user principal	${hbase-env/hbase_user}${principal_suffix}@${realm}
HBase user keytab	${keytab_dir}/hbase.headless.keytab
HDFS user principal	${hadoop-env/hdfs_user}${principal_suffix}@${realm}
HDFS user keytab	${keytab_dir}/hdfs.headless.keytab
Spark2user Keytab	${keytab_dir}/spark.headless.keytab
Spark2user Principal	${spark2-env/spark_user}${principal_suffix}@${realm}
Storm user keytab	${keytab_dir}/storm.headless.keytab
Storm user principal	${storm-env/storm_user}${principal_suffix}@${realm}
yarn_ats_principal_name	${yarn-env/yarn_ats_user}${principal_suffix}@${realm}
yarn_ats_user_keytab	${keytab_dir}/yarn-ats.hbase-client.headless.keytab

All configurations have been addressed.

11. Click **Next**.
12. The **Confirm Configuration** page gives you a final opportunity review your configuration. If you are satisfied with the settings, click **Next**.
13. The **Stop Services** page shows you the progress of the services being stopped. After the services have stopped, click **Next**.

    The next page shows the progress of the cluster being kerberized. This may take several minutes.

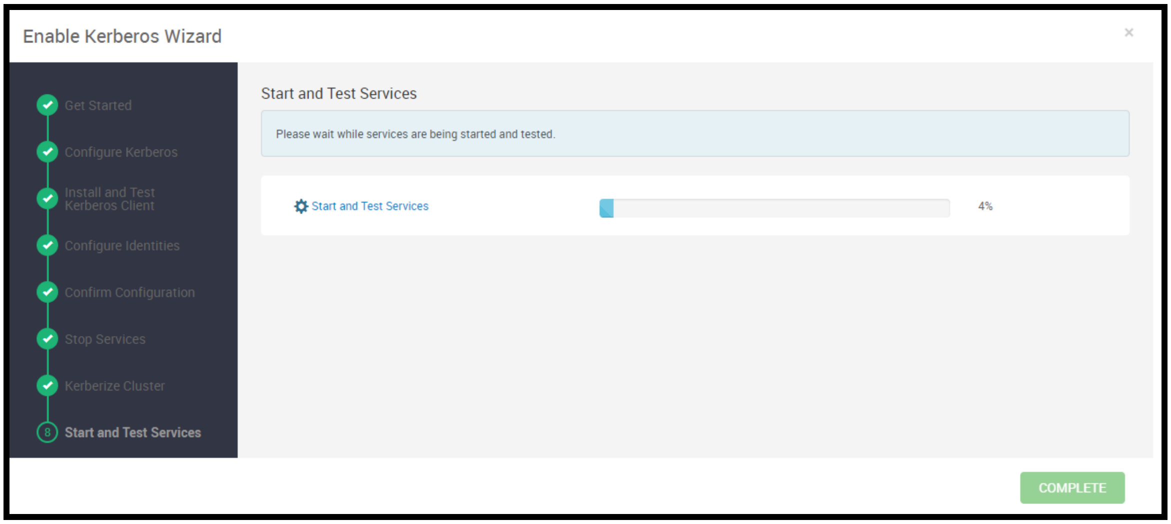


Kerberize Cluster

Please wait while cluster is being kerberized.

14. Click **Next** when it has finished.

    The **Start and Test Services** page is displayed.



Start and Test Services

Please wait while services are being started and tested.

    After principals have been created and keytabs have been generated and distributed, Ambari updates the cluster configurations, then starts and tests the Services in the cluster.

15. Click **Complete** to finish the configuration of Kerberos and return to the Ambari home page.

Configure Ambari with TLS

1. In a command console on the Monitoring node, run the following commands to navigate to the **/opt/interset/installer** directory and launch the Interset installation menu.

```
cd /opt/interset/installer
```

```
./install.sh
```

2. From the available installation options, type **6** to select installation step **[6] Configure HDP/TLS**, and then press **Enter**.

   ```
   Please select an installation option from the list above: 6
   ```

   This step configures Ambari to use TLS.

   When TLS is configured, the expected result is:

   ```
   AMBARI URL: https://<monitoring_node_fqdn>:8443
   ```

Configure Kerberos on Cloudera

The Interset installer automatically configures Kerberos on Cloudera. If you need to manually change the configuration, see "Run the Enable Kerberos Wizard on Cloudera" on page 117.

Configure Cloudera with TLS

1. In a command console on the Monitoring node, run the following commands to navigate to the **/opt/interset/installer** directory and launch the Interset installation menu.

   ```
   cd /opt/interset/installer
   ```

   ```
   ./install.sh
   ```

2. From the available installation options, type **6** to select installation step **[6] Configure HDP/TLS**, and then press **Enter**.

   ```
   Please select an installation option from the list above: 6
   ```

3. Enter and then confirm a new password for the Cloudera admin user.

When the TLS is configured, the new Cloudera Manager URL is displayed:

```
https://<monitoring_node_fqdn>:7183
```

Install the Interset Schema Registry

Interset 5.9.2 uses NiFi for the data extraction and transformation portion of data ingest. To use the Interset schemas in NiFi, you first install the schema registry. Later in the installation, after you install NiFi, you configure the NiFi processors to recognize the Confluent schema registry.

1. In a command console on the Monitoring node, run the following commands to navigate to the **/opt/interset/installer** directory and launch the Interset installation menu.

   ```
   cd /opt/interset/installer
   ```

   ```
   ./install.sh
   ```

2. From the available installation options, type **7** to select installation **[7] Schema Registry Setup**, and then press **Enter**.

   ```
   Please select an installation option from the list above: 7
   ```

   Note: If you see the following error, delete the specified files:

   ```
   Configuring schema registry logging...
   ```

   ```
   /opt/interset/installer/tmp/install_registry.sh: line 179: /etc/rsyslog.conf: Permission denied
   ```

   Delete the following files:

   intersetlog_install_dbs_registryLog.txt

   intersetlog_installSchemaRegistryLog.txt

   From the installation options, type **7** again to continue with schema registration.

3. Enter the password (selected in step 1 of the installation) for the Postgres database user:

   ```
   Please enter your existing password for DB user - Postgres:
   ```

4. Enter and then re-enter a password for the schema registry database user:

   ```
   Please enter new password for DB user - Registry (must be alphanumeric upper/lowercase):
   Please re-enter password for DB user - Registry:
   ```

   After the registry is installed, the expected result is:

   ```
   SSH Successful and install_dbs_registry.sh started on monitoring-1.docs.qa.interset.com!
   Install_dbs_registry.sh on monitoring-1.docs.qa.interset.com Complete!
   ```

5. When prompted to upload Interset schemas to the schema registry, type **y** and then press Enter.

   After the schema upload to the Master node in the Interset cluster is complete, the expected result is:

   ```
   Registry Setup Is Complete!
   ```

✓ You can access the schema registry at the following URL:

https://<master_node_fqdn>:9190/ui/

Configure the Master Node

1. In a command console on the Monitoring node, run the following commands to navigate to the **/opt/interset/installer** directory and launch the Interset installation menu.

   ```
   cd /opt/interset/installer
   ```

   ```
   ./install.sh
   ```

2. From the available installation options, type **8** to select installation **[8] Analytics installation**, and then press **Enter**.

   ```
   Please select an installation option from the list above: 8
   ```

   When the Analytics installation is complete on the Master node, the expected result is:

   ```
   Analytics Node Configuration Complete!
   ```

Important: Only during Cloudera setup, if you get the following error; use the steps below to resolve the issue:

```
==== CONFIGURING INTERSET.CONF =====
zkPhoenix = <master_node_fqdn>:2181:/hbase
esHost=<search_node_fqdn>
mkdir: SIMPLE authentication is not enabled.  Available:[TOKEN, KERBEROS]
```

1. Open the Cloudera Manager
2. Click on Deploy all Configs
3. Once the deployment completes, **Restart** all the services
4. Rerun step 9: **Configure the Master Node**

Configure the Stream Node

1. In a command console on the Monitoring node, run the following commands to navigate to the **/opt/interset/installer** directory and launch the Interset installation menu.

   ```
   cd /opt/interset/installer
   ```

   ```
   ./install.sh
   ```

2. From the available installation options, type **9** to select installation **[9] Stream node(s) installation**, and then press **Enter**.

   ```
   Please select an installation option from the list above: 9
   ```

   When the Stream node installation is complete, the expected result is:

   ```
   Stream Node Configuration Complete!
   ```

Configure the Search Node

1. In a command console on the Monitoring node, run the following commands to navigate to the **/opt/interset/installer** directory and launch the Interset installation menu.

   ```
   cd /opt/interset/installer
   ```

   ```
   ./install.sh
   ```

2. From the available installation options, type **10** to select installation **[10] Search node(s) installation**, and then press **Enter**.

   ```
   Please select an installation option from the list above: 10
   ```

   When the Search node installation is complete, the expected result is:

   ```
   Search Node Configuration Complete!
   ```

   > ✓ **Tip:** Although the Search node is configured at this point in the installation, you can only configure the Search functionality after data is ingested and you have run Analytics. For more information, see "Configure the Search Indexes" on page 108.

   To configure additional Search nodes in your Interset cluster, please contact Micro Focus Interset Support at interset.support@microfocus.com for guidance.

Configure the Reporting Node

1. In a command console on the Monitoring node, run the following commands to navigate to the **/opt/interset/installer** directory and launch the Interset installation menu.

   ```
   cd /opt/interset/installer
   ```

   ```
   ./install.sh
   ```

2. From the available installation options, type **11** to select installation **[11] Reporting node(s) installation**, and then press **Enter**.

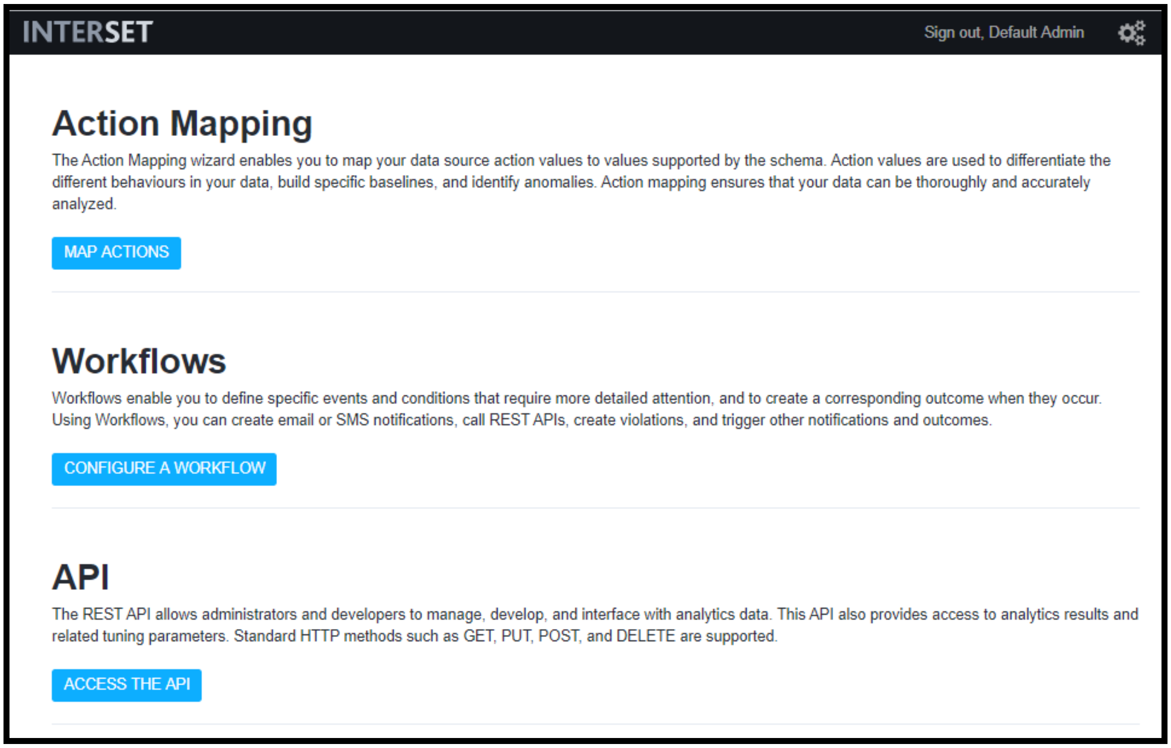
   ```
   Please select an installation option from the list above: 11
   ```

   When the Reporting node installation is complete, the expected result is:

   ```
   Reporting Node Configuration Complete!
   Please confirm that you can log in by accessing https://<reporting_fqdn> in a browser using credentials
   admin/password.
   ```

3. In a Web browser, navigate to **https://*<reporting_fqdn>***, and then log in to Interset using the default username and password, **admin**:**password**.

   The Interset user interface opens to the **Settings** page.



Action Mapping

The Action Mapping wizard enables you to map your data source action values to values supported by the schema. Action values are used to differentiate the different behaviours in your data, build specific baselines, and identify anomalies. Action mapping ensures that your data can be thoroughly and accurately analyzed.

Workflows

Workflows enable you to define specific events and conditions that require more detailed attention, and to create a corresponding outcome when they occur. Using Workflows, you can create email or SMS notifications, call REST APIs, create violations, and trigger other notifications and outcomes.

API

The REST API allows administrators and developers to manage, develop, and interface with analytics data. This API also provides access to analytics results and related tuning parameters. Standard HTTP methods such as GET, PUT, POST, and DELETE are supported.

✓ **Tips:**

- The only settings options that appear in the **Settings** page at this point in the installation are: **Action Mapping**, **Workflows**, and **API**.
- After you ingest data into the Interset system and run Analytics against the data, the Analytics **Overall Risk** page becomes the landing page when you log in.

Notes:

- Reporting may take several minutes to start after the installation completes.
- If you try to log in to Reporting before the Reporting service has completed the startup, you will receive the following message:

  ```
  Wrong username or password. Please try again.
  ```

- If you receive the message **"The server is not responding, please contact your administrator"**, ElasticSearch may have failed to start. To resolve this situation, run the following command to restart the Elasticsearch service, and then run installation option **8** of the installer again.

  ```
  systemctl restart elasticsearch
  ```

Important: We recommend that you change the root user password after a successful installation of reporting. The default user is **root**, with password **root**. The root user can change their own password. After you have changed the root password, you will be redirected to the login page and will have to reenter your credentials using your new root user and password.

To configure additional Reporting nodes in your Interset cluster, please contact Micro Focus Interset Support at interset.support@microfocus.com for guidance.

Configure Authentication

Interset provides the following authentication configuration options:

- local Interset authentication
- LDAP authentication
- SAML authentication

You can configure authentication at this point in the installation of your Interset cluster, or you can wait until the installation is complete.

> ✓ **Tip:** If you plan to configure Workflows in your Interset cluster, you should configure authentication prior to configuring the Workflow component (see "Configure Workflow" on page 93).
>
> For more information about Workflows, please see "Create Workflows" in the Interset 5.9.2*Administrator Guide*.

Local Interset Authentication

When you first install Interset, two tenants are created: the system **Administrative Tenant** (tenant ID: **adm**), responsible for the management of all tenants and users, and the **Default On-Premise Tenant** (tenant ID, or TID **0**), the first tenant available for data ingest configuration and data analysis.

- The **Administrative Tenant** (**adm**) has one default user, the **root user** (username and password **root**/**root**).
  - The Administrative tenant root user is the administrator for the system, creating and managing tenants and users.
- The **Default Tenant0** (Tenant ID 0) has three default users: **admin**, **user**, and **workflow_0**.
  - The default tenant **admin** user is essentially the administrator for the tenant, configuring the data ingest, creating tenant Workflows to highlight specific behaviors, and creating any desired customizations for the Analytics consumers.

    The default username and password for the **admin** user are **admin**:**password**.
  - The default tenant **user** is an Interset Analytics consumer. This user has access to the Analytics data in the Interset Analytics user interface for the tenant.

    The default username and password for the **user** user are **user**:**password**.
  - The default tenant **workflow_0** user is an administrative user that enables the Reporting and Workflow components to communicate without prompting for access credentials.

    A Workflow is specific to a tenant; the Workflow component is configured for an associated tenant via the tenant **rules.conf** configuration file, which also defines the Workflow user for the tenant.

    The default username and password for the **workflow_0** user are **user**:**password**.

If you plan to use the Interset local authentication in your cluster, the default tenant **0** with its three default users (**admin**, **user**, and **workflow_0**) are all you require to complete the installation. When you are ready to configure more tenants and users, please see "Administer the Interset System" in the Interset 5.9.2*Administrator and User Guide.*

LDAP Authentication

Interset can be configured to use an external LDAP provider (such as a Microsoft Active Directory AD server) for authentication.

To configure Interset to use LDAP authentication, you will

- review the Interset account roles
- verify there's a valid LDAP account for the cluster
- connect to an SSL-enabled LDAP server
- configure Interset to use LDAP

Review the Account Roles

There are three types of user account roles in Interset 5.9.2: root, administrator, and user.

When Interset is configured for LDAP authentication and a user successfully logs in using their LDAP credentials, a local account is created and assigned the **user** role. By default, accounts assigned the **user** role are given limited permissions.

The Interset **root** user is the cluster superuser. This user has API permissions to assign the **administrator** role to user accounts.

> ⚠ When LDAP is enabled, the **root** user must be manually configured in the **investigator.yml** file.

Enabling LDAP changes the default system behavior.

For information about using the API, see the Interset 5.9.2*Developer Guide*.

Verify There's a Valid LDAP Account Provisioned for Interset

Interset 5.9.2 requires access to an LDAP service account so that it can query the LDAP server when Interset users log in. We recommend that the LDAP service account:

- be used only by Interset
- be configured to not expire
- have unthrottled search capabilities, and
- have read-only access to your LDAP server

The credentials for this LDAP service account can be verified using a tool such as **ldapsearch**. For example:

```
ldapsearch -v -x -W -h ad.example.com -p 389 -D "CN=Bobby Clobber,OU=New York,OU=Example Users,DC=ad,DC=example,DC=com"
-b "dc=ad,dc=example,dc=com" "(&(objectclass=user)(samaccountname=bclobber))" DN
```

The source that provides the LDAP service account can also provide instructions for, and explain, how to correctly set the **bind** user used to bind to the LDAP directory (**-D** in the command above, **ldapSearchDN** in the configuration file below) and the starting point for the directory search (**-b** in the command above, **ldapSearchBaseDN** in the configuration file below).

Connect to an SSL-enabled LDAP Server

These instructions assume that your LDAP server is configured with SSL. To connect Interset to your LDAP server, you add the LDAP server certificate to the Reporting node Java keystore.

1. Run the following command to make a certificate query and verify that your LDAP server has SSL enabled.

   ```
   openssl s_client -connect ad.example.com:636 </dev/null 2>/dev/null | sed -n '/^-----BEGIN/,/^-----END/p'
   ```

   If the certificate query does not return a result similar to the example below, this means that either SSL is not set up on your LDAP server or it has not been properly configured.

   ```
   -----BEGIN CERTIFICATE-----
   MXIGFzCCBP+gXwIBAgITJQAAAAVZDMkXeqtEpwAAAAAABTANXgkqhkiG9w0BAQUX
   ...
   ...
   ---END CERTIFICATE-----
   ```

   In this case, please contact the source that provided you with the LDAP information.

2. If the command in the previous step returned a valid certificate, save and install the certificate using the following commands:

   ```
   sudo openssl s_client -connect ad.example.com:636 </dev/null 2>/dev/null | sed -n '/^-----BEGIN/,/^-----END/p' >
   $JAVA_HOME/jre/lib/security/cert.pem
   ```

   ```
   sudo keytool -import -file $JAVA_HOME/jre/lib/security/cert.pem -alias ldap -keystore $JAVA_
   HOME/jre/lib/security/cacerts
   ```

   Note: If the Domain Controller LDAP certificate was issued from a private Certificate Authority, please ensure the issuing certificate chain is imported into the the java cacerts store.

3. When prompted for the keystore password, enter the password.

   Tip: The default password is **changeit**

4. When prompted whether to trust this certificate, type **yes**.

   ```
   Trust this certificate? [no]:  yes
   ```

   The response should be similar to the following:

   ```
   Certificate was added to keystore
   ```

5. Run the following command to verify that the certificate was added correctly.

   ```
   sudo keytool -list -keystore $JAVA_HOME/jre/lib/security/cacerts | grep ldap
   ```

The response should be similar to the following:

```
ldap, Jul 18, 2017, trustedCertEntry,
```

Configure Interset to Use LDAP Authentication

To configure Interset to use your LDAP authentication, you must modify the **investigator.yml** file.

1. Open **/opt/interset/etc/investigator.yml**, locate the **# LDAP Authentication** section, and then set the following variables to the LDAP information you verified at the beginning of this section:
   - **ldapSearchBaseDN** is the starting point for the LDAP directory search.
   - **ldapSearchDN** is the LDAP service account's Distinguished Name (bind user).
   - **ldapSearchDNPassword** is the plain text password of the LDAP service account.
   - **ldapUrl** is the LDAP server, such as ldaps://ad.example.com:636.

   > ✓ **Tip:** If your LDAP server isn't configured to use SSL, use ldap://ad.example.com:389.

2. Set the following parameters:
   - **Enabled** to **true**
   - **rootUser** to the username of the LDAP account that you want to be the Interset superuser.

   The Interset superuser is able to assign administrative privileges to other Interset accounts and should not be the LDAP service account (**bind** user).

3. To change the default tenant that Interset users are logged in to, edit the variable **"ldapDefaultTenantId"** with the new tenant's ID.
   - If you want to enforce membership in an LDAP group to restrict who can log in to Interset, edit the **ldapSearchFilter** variable.

     For example, **(memberOf=CN=<>,CN=<>,DC=example,DC=com))**

   > **Notes:**
   >
   > - The Interset superuser (**root**) is restricted from logging in to the Interset user interface.
   > - Users with the role of **user** are restricted from logging into the Interset user interface until the tenant(s) they are assigned to have been configured by an **administrator** user.

4. On the **Reporting** node, restart Reporting using the following command:

   ```
   systemctl restart reporting
   ```

   > ✓ **Tip:** Troubleshooting information is logged in **/opt/interset/log/reporting.log**

> ⚠ **Important:** When you set the investigator.yml **rootUser** to the username of the LDAP account, the original default **root** user is not removed from the **adm** tenant. As a result, the original default **root** user will continue to have root access (as will the new root user) unless the new LDAP root user manually removes it.

5. In a Web browser, log in to Interset as your new LDAP **rootUser.**

   In the **Tenants** page, under **Administrative Tenant**, your new LDAP **rootUser** appears in the list of users.
6. In the **Administrative Tenant** page, disable the system (local) **root** user.

SAML Authentication

If your organization uses SAML for authentication, you can use your SAML deployment for authentication to Interset. For more information, see the **Modify a User's Role** section in the Interset 5.9.2 *Administrator and User Guide*.

> If you need to configure SAML authentication for more than one tenant, please contact Interset support at interset.support@microfocus.com.

When you configure SAML authentication for your Interset system:

- your SAML application will manage authentication for Interset. However, the API will continue to manage the permissions assigned to each user.
- any new user that logs in to Interset will be created automatically in the Reporting server database.

  By default, any new user is assigned the **user** role, the role with the fewest privileges. You can modify the new user's role on the **Tenants** page when you are logged in as the **rootUser** configured in **investigator.yml**.

> ⚠ **Important:** Upon initial installation and setup of your Interset cluster, only the SAML account defined in the **investigator.yml** file as **rootUser** is able to log in and configure the system.

Configuring SAML authentication for your Interset system involves:

- editing the **SAML Authentication** section of the **investigator.yml** file to enable SAML authentication
- configuring an Active Directory Federation Services (AD FS) Relying Party Trust to create the connection between Active Directory and Interset
- creating the claim rules
- restarting Reporting

You can also configure single sign-on with SAML and Okta.

Edit the investigator.yml file to Enable SAML Authentication

1. On the **Reporting** node, navigate to the **/opt/interset/etc** directory, and open the **investigator.yml** file.
2. Scroll to the SAML Authentication section, and update the **saml-auth** parameters as follows:

```
enabled: true
rootUser: <rootUser_email>
defaultTenantId: <TID>
relyingPartyIdentifier: Interset_Investigator_identifier
assertionConsumerServiceURL: https://<reporting_node_fqdn>/api/actions/login/saml/sso
metadataUrl: https://<adfs_server_fqdn>/FederationMetadata/2007-06/FederationMetadata.xml
```

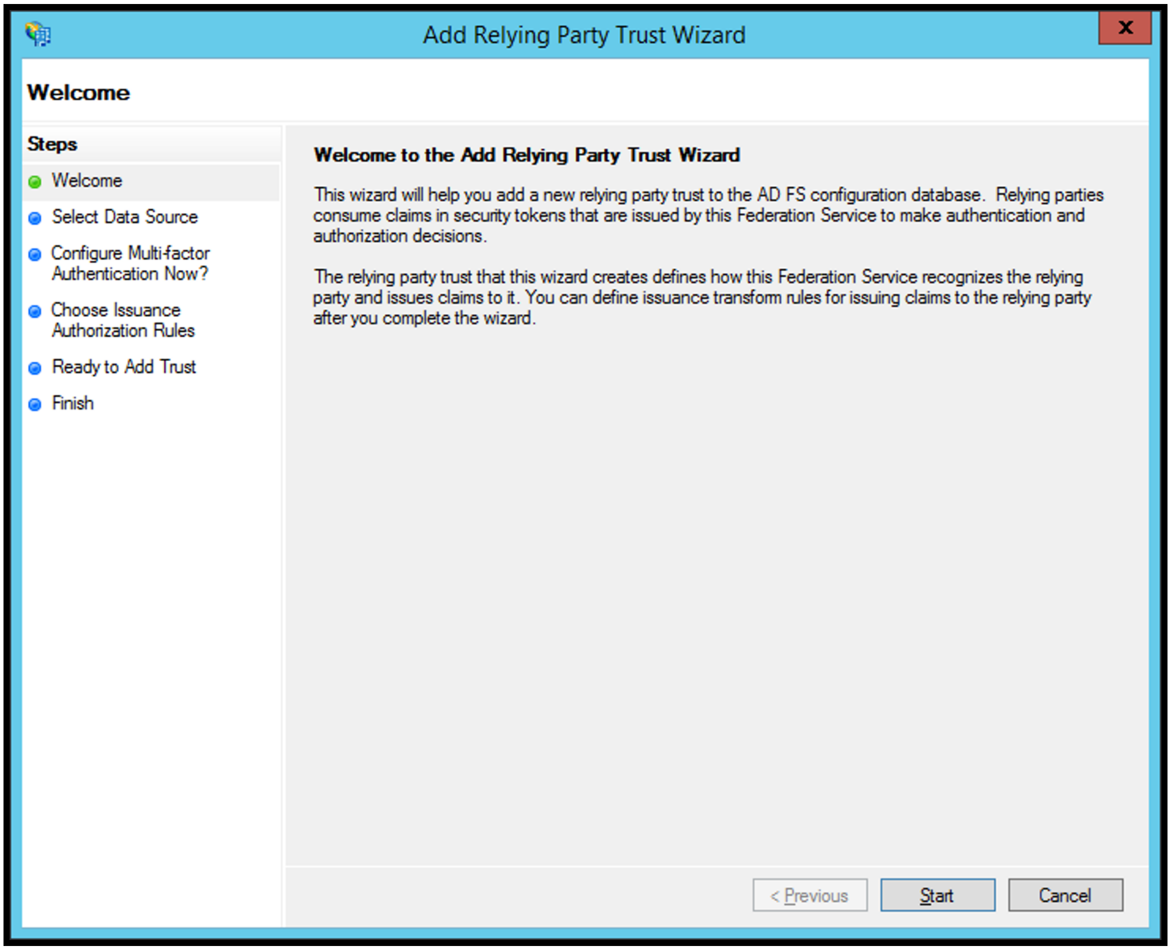
> **Important:** When you set the investigator.yml **rootUser** to the username of the SAML account, the original default **root** user is not removed from the **adm** tenant. As a result, the original default **root** user will continue to have root access (as will the new root user) unless the new SAML root user manually removes it.

3. Save the updated **investigator.yml** file.

Configure the AD FS Relying Party Trust (RPT)

1. Open Active Directory Federation Services.
2. In Server Manager, click **Tools**, and then select **AD FS Management**.
3. In the **Actions** pane, click **Add Relying Party Trust**.

   This launches the **Add Relying Party Trust** wizard and opens the **Welcome** page.



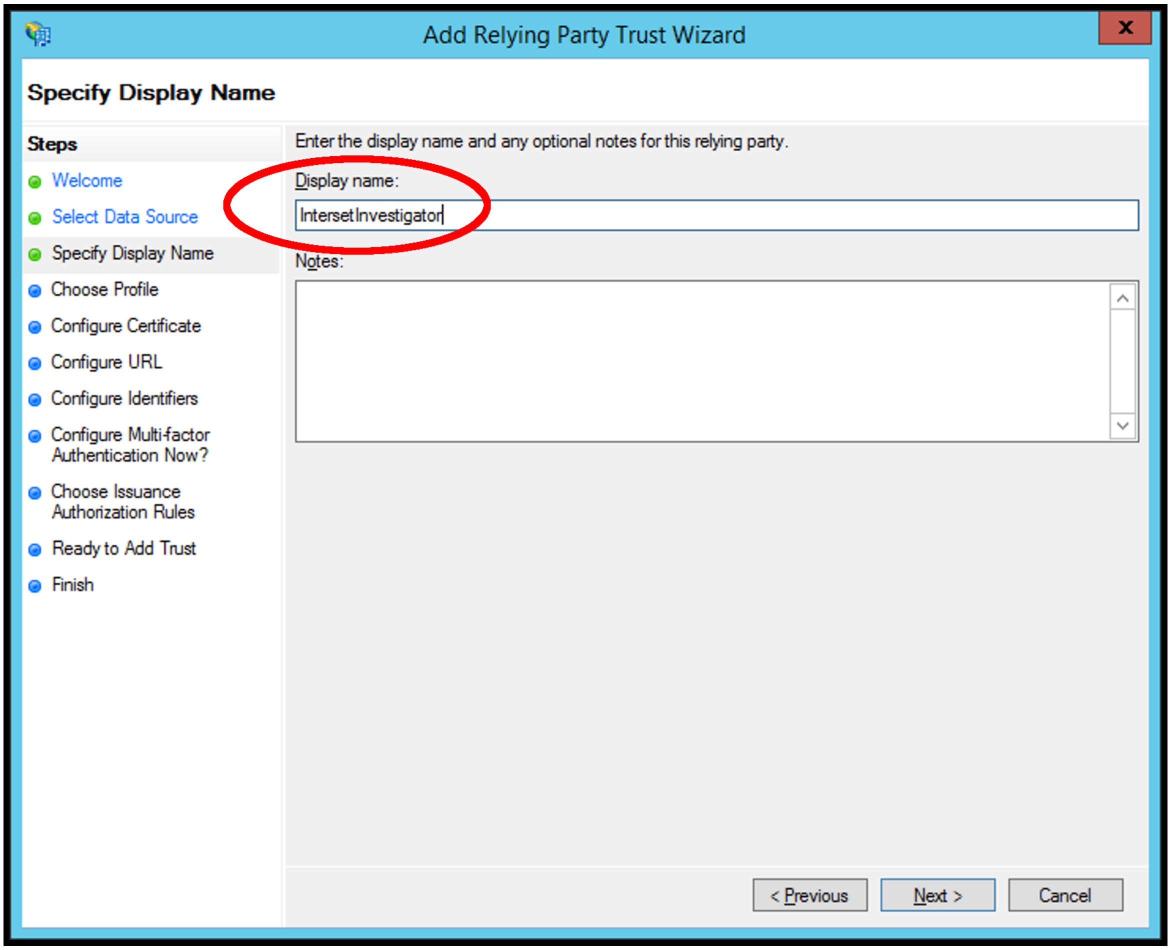
Add Relying Party Trust Wizard
Welcome
Steps
Welcome
Select Data Source
Configure Multi-factor Authentication Now?
Choose Issuance Authorization Rules
Ready to Add Trust
Finish
< Previous
Start
Cancel

Welcome to the Add Relying Party Trust Wizard

This wizard will help you add a new relying party trust to the AD FS configuration database. Relying parties consume claims in security tokens that are issued by this Federation Service to make authentication and authorization decisions.

The relying party trust that this wizard creates defines how this Federation Service recognizes the relying party and issues claims to it. You can define issuance transform rules for issuing claims to the relying party after you complete the wizard.

4. On the **Select Data Source** page, select the **Enter data about the relying party manually** radio button, and then click **Next**.
5. In the **Specify Display Name** page, in the **Display name** box, enter the value of the **relyingPartyIdentifier** parameter you set in the **investigator.yml** file.

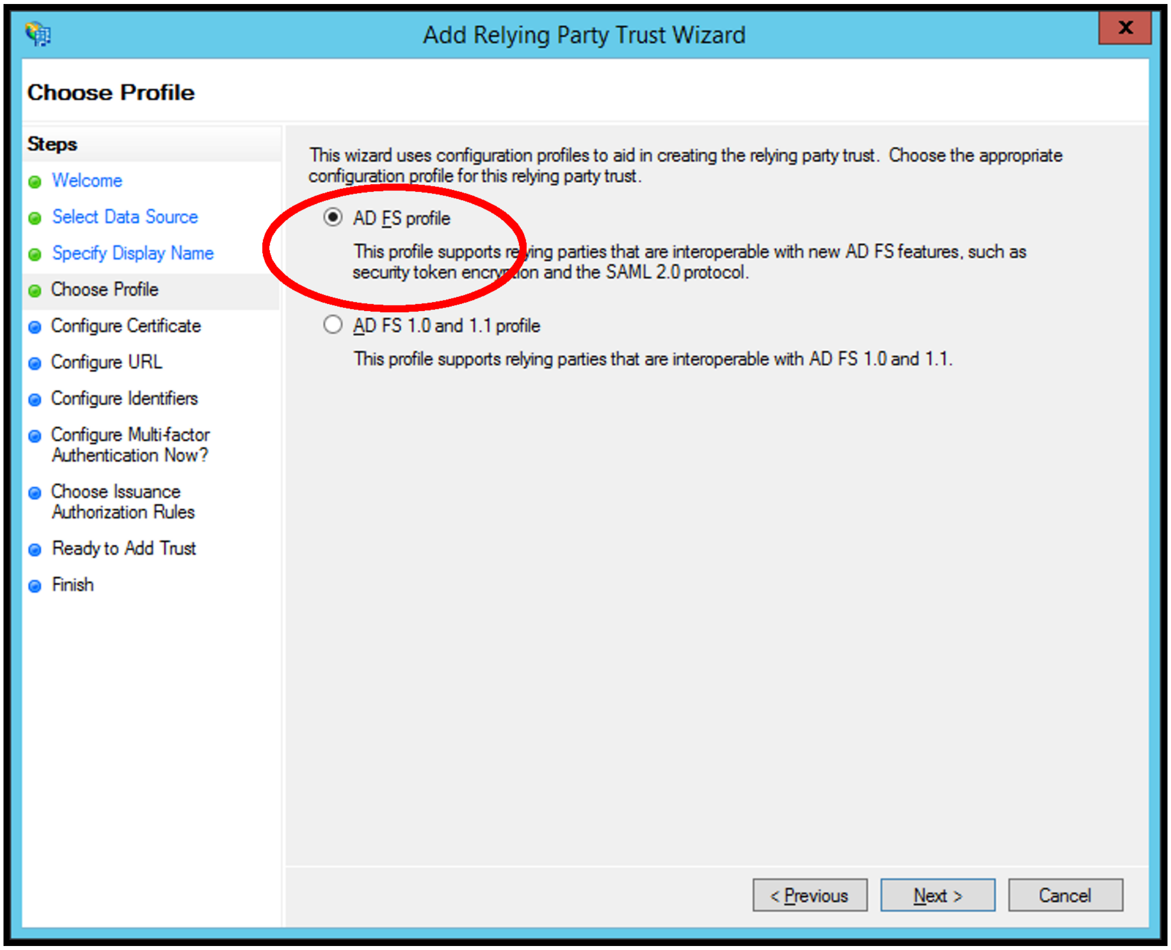


Enter the display name and any optional notes for this relying party.

Display name: Interset Investigator

Notes:

6. On the **Choose Profile** page, select the **AD FS profile** radio button, and then click **Next**.



This wizard uses configuration profiles to aid in creating the relying party trust. Choose the appropriate configuration profile for this relying party trust.

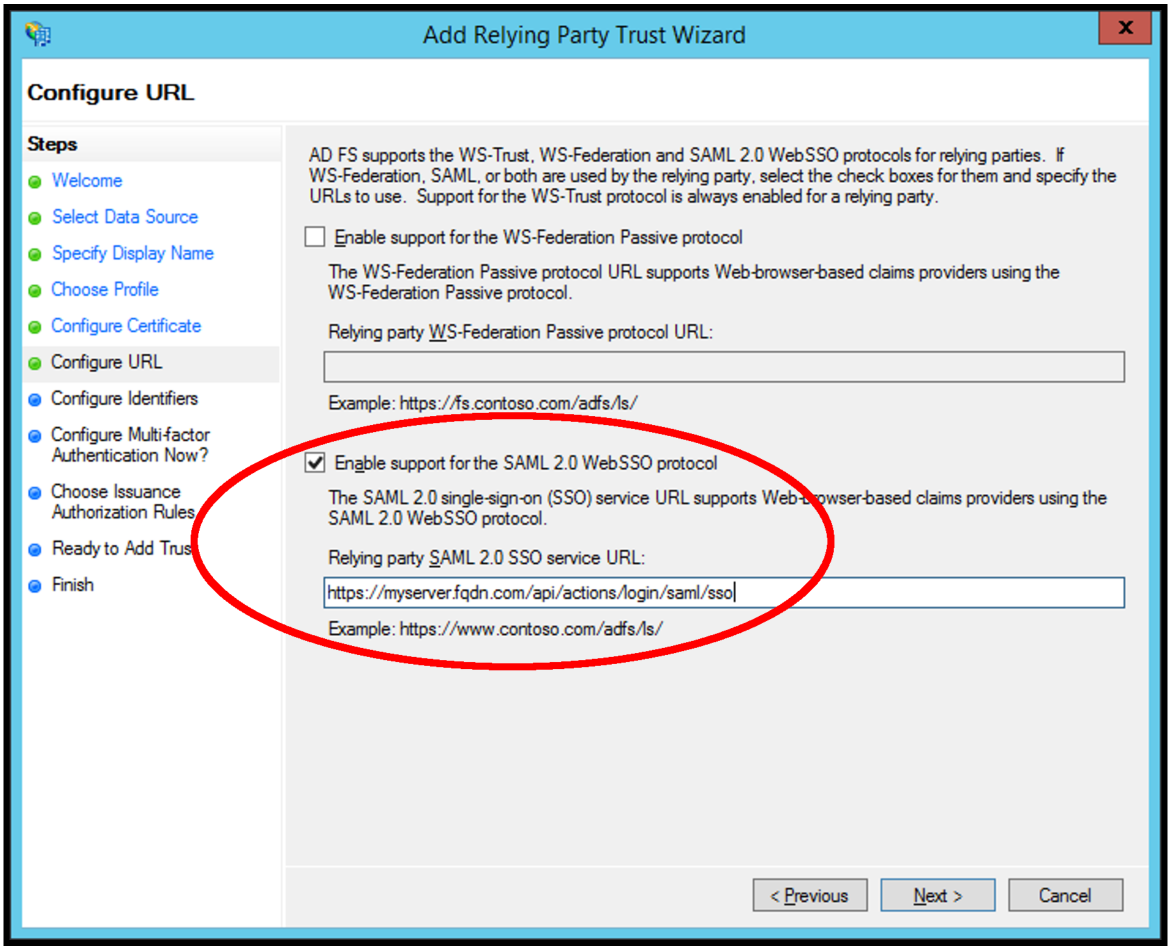
● AD FS profile

This profile supports relying parties that are interoperable with new AD FS features, such as security token encryption and the SAML 2.0 protocol.

○ AD FS 1.0 and 1.1 profile

This profile supports relying parties that are interoperable with AD FS 1.0 and 1.1.

7. On the **Configure Certificate** page, accept the defaults and then click **Next**.
8. On the **Configure URL** page, select the **Enable Support for the SAML 2.0 WebSSO protocol** checkbox.



Configure URL

AD FS supports the WS-Trust, WS-Federation and SAML 2.0 WebSSO protocols for relying parties. If WS-Federation, SAML, or both are used by the relying party, select the check boxes for them and specify the URLs to use. Support for the WS-Trust protocol is always enabled for a relying party.

□ Enable support for the WS-Federation Passive protocol

The WS-Federation Passive protocol URL supports Web-browser-based claims providers using the WS-Federation Passive protocol.

Relying party WS-Federation Passive protocol URL:

Example: https://fs.contoso.com/adfs/ls/

☑ Enable support for the SAML 2.0 WebSSO protocol

The SAML 2.0 single-sign-on (SSO) service URL supports Web-browser-based claims providers using the SAML 2.0 WebSSO protocol.

Relying party SAML 2.0 SSO service URL:

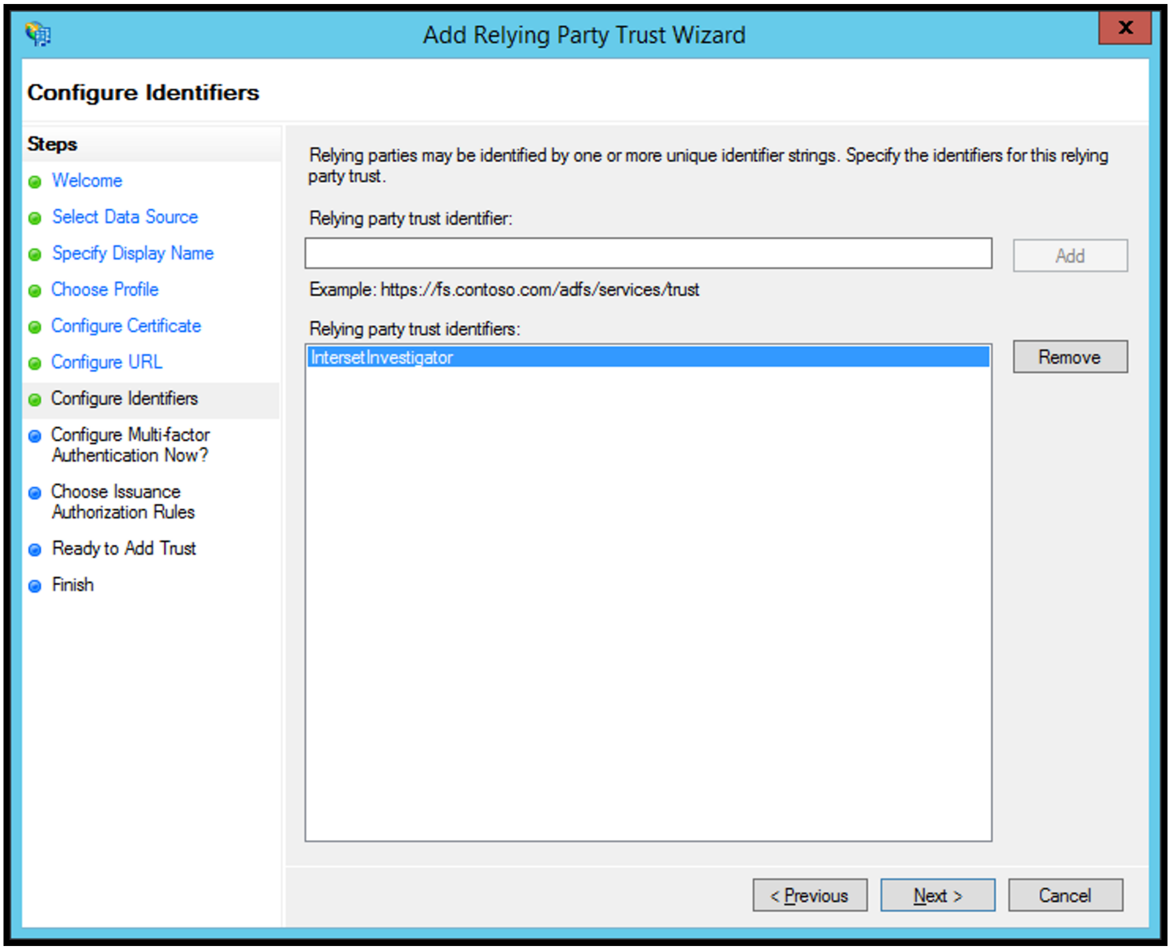
https://myserver.fqdn.com/api/actions/login/saml/sso

Example: https://www.contoso.com/adfs/ls/

In the **Relying party SAML 2.0 SSO service URL** box, enter the value of the **assertionConsumerServiceUrl** parameter you set in the **investigator.yml** file.

9. On the **Configure Identifiers** page, in the **Relying party trust identifier** box, enter the value of the **relyingPartyIdentifier** parameter you set in the **investigator.yml** file, and then click **Add**.

   The new Relying Party Identifier appears in the **Relying party trust identifiers** box.



Configure Identifiers

Relying parties may be identified by one or more unique identifier strings. Specify the identifiers for this relying party trust.

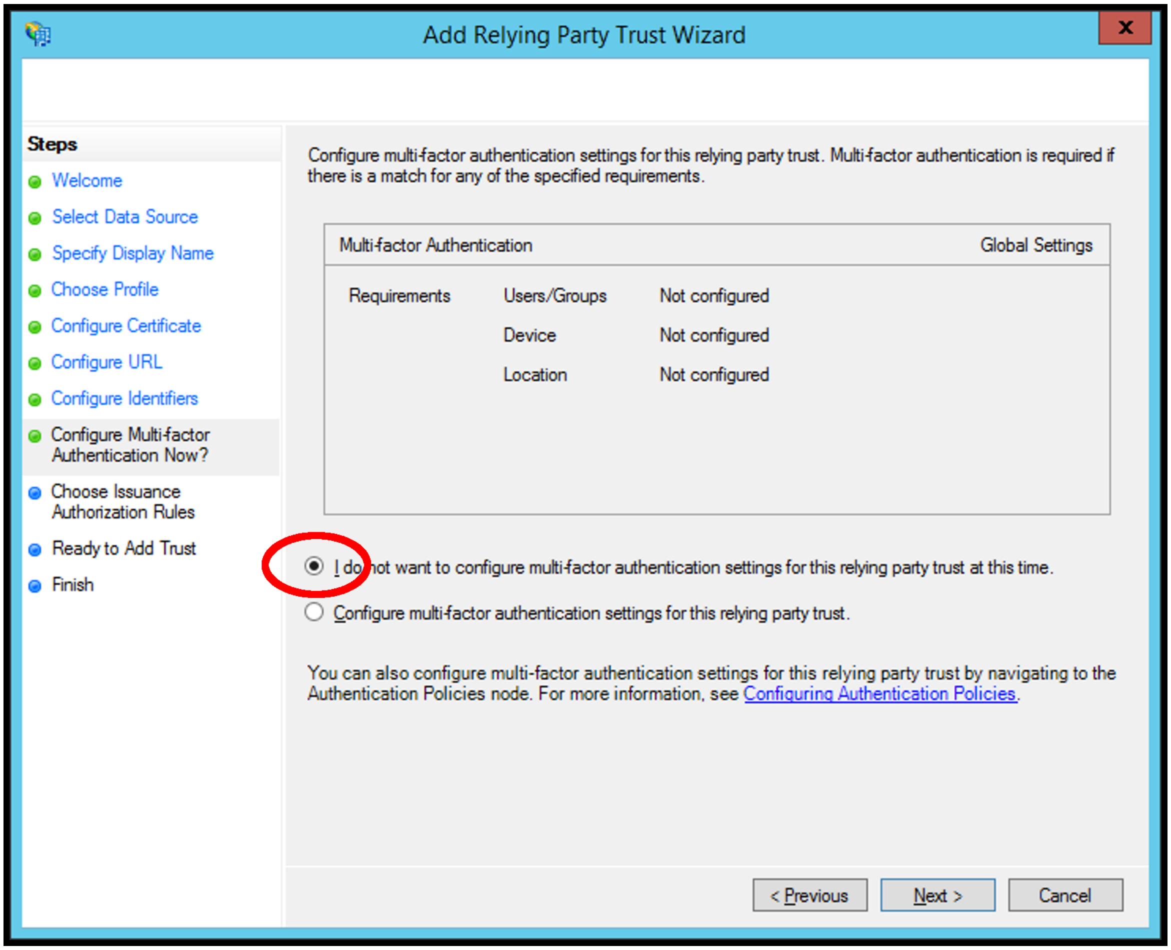
Relying party trust identifier:

Example: https://fs.contoso.com/adfs/services/trust

Relying party trust identifiers:

IntersetInvestigator

10. Click **Next**.
11. On the **Configure Multi-factor Authentication Now?** page, select the **I do not want to configure multi-factor authentication settings for this relying party trust at this time** radio button, and then click **Next**.



Configure multi-factor authentication settings for this relying party trust. Multi-factor authentication is required if there is a match for any of the specified requirements.

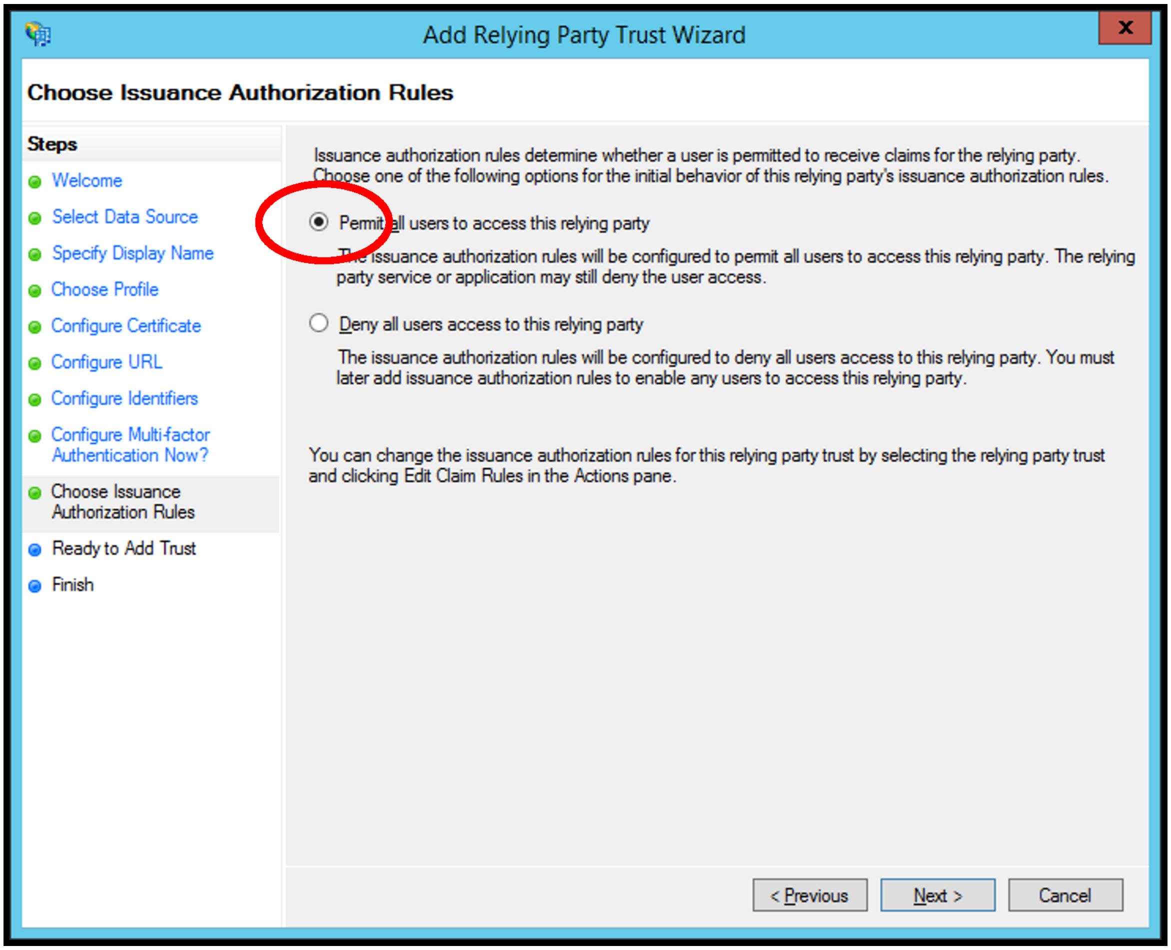
Multi-factor Authentication		Global Settings
Requirements	Users/Groups	Not configured
	Device	Not configured
	Location	Not configured

● I do not want to configure multi-factor authentication settings for this relying party trust at this time.

○ Configure multi-factor authentication settings for this relying party trust.

You can also configure multi-factor authentication settings for this relying party trust by navigating to the Authentication Policies node. For more information, see Configuring Authentication Policies.

12. On the **Choose Issuance Authorization Rules** page, select the **Permit all users to access this relying party** radio button, and then click **Next.**



Choose Issuance Authorization Rules

Issuance authorization rules determine whether a user is permitted to receive claims for the relying party. Choose one of the following options for the initial behavior of this relying party's issuance authorization rules.

● Permit all users to access this relying party

The issuance authorization rules will be configured to permit all users to access this relying party. The relying party service or application may still deny the user access.

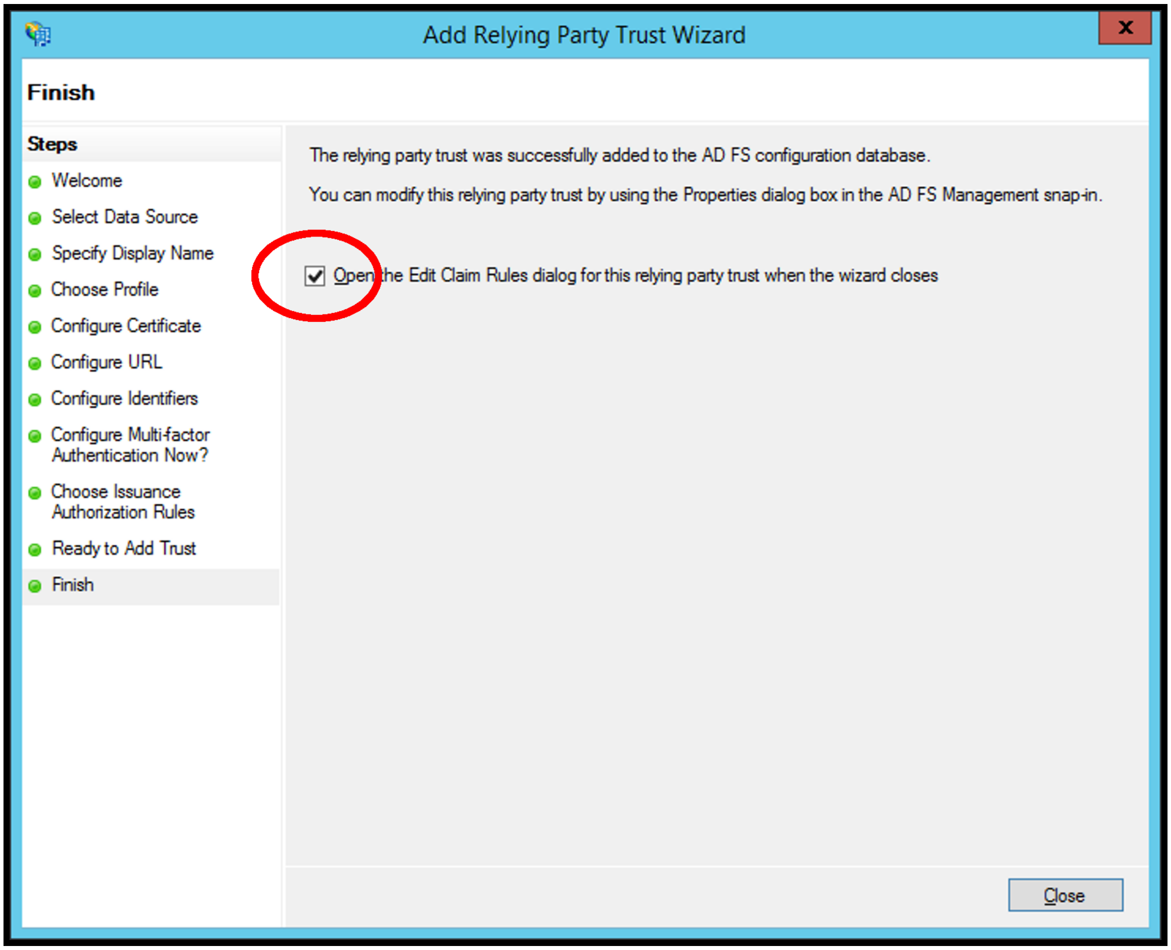
○ Deny all users access to this relying party

The issuance authorization rules will be configured to deny all users access to this relying party. You must later add issuance authorization rules to enable any users to access this relying party.

You can change the issuance authorization rules for this relying party trust by selecting the relying party trust and clicking Edit Claim Rules in the Actions pane.

13. On the **Ready to Add Trust** page, click **Next**.
14. On the **Finish** page, select the **Open the Edit Claim Rules dialog for this relying party trust when the wizard closes** checkbox, and then click **Close**.

    The **Edit Claim Rules** wizard opens.



Add Relying Party Trust Wizard

Finish

The relying party trust was successfully added to the AD FS configuration database.

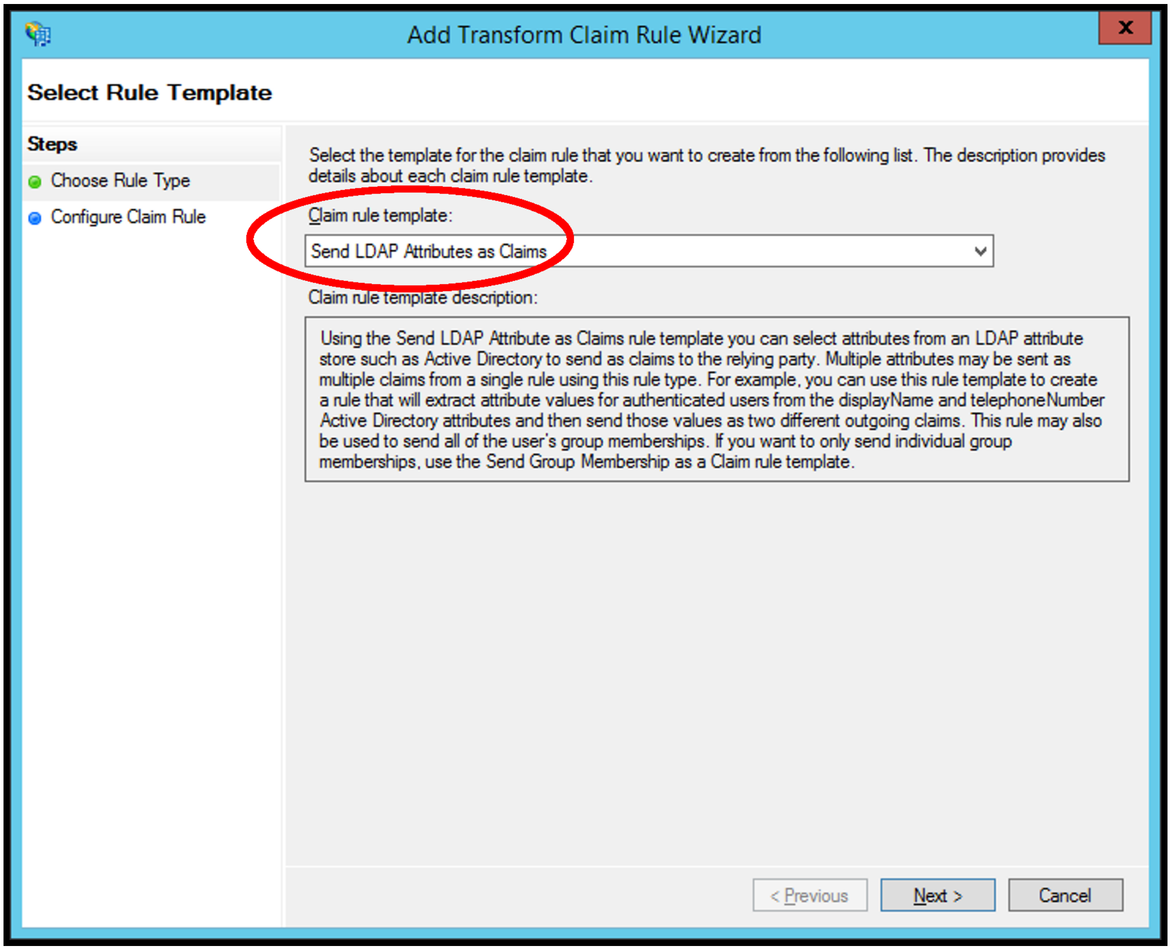
You can modify this relying party trust by using the Properties dialog box in the AD FS Management snap-in.

☑ Open the Edit Claim Rules dialog for this relying party trust when the wizard closes

Create the Interset Claim Rules

After the relying party has been created, you create the claim rules and update the trust with additional settings not configured in the wizard.

1. In the **Edit Claim Rules** dialog box, click **Add Rule**.
2. On the **Select Rule Template** page, click the **Claim rule template** dropdown arrow, select **Send LDAP Attributes as Claims** from the list, and then click **Next**.



Add Transform Claim Rule Wizard

Select Rule Template

Select the template for the claim rule that you want to create from the following list. The description provides details about each claim rule template.

Claim rule template: Send LDAP Attributes as Claims

Claim rule template description:

Using the Send LDAP Attribute as Claims rule template you can select attributes from an LDAP attribute store such as Active Directory to send as claims to the relying party. Multiple attributes may be sent as multiple claims from a single rule using this rule type. For example, you can use this rule template to create a rule that will extract attribute values for authenticated users from the displayName and telephoneNumber Active Directory attributes and then send those values as two different outgoing claims. This rule may also be used to send all of the user's group memberships. If you want to only send individual group memberships, use the Send Group Membership as a Claim rule template.

3. On the **Configure Rule** page, do the following:
   a. In the **Claim rule name** box, enter the value of the **relyingPartyIdentifier** parameter you set in the **investigator.yml** file.
   b. In the **Attribute store** box, click the dropdown arrow and then select **Active Directory** from the list.
   c. In the **Mapping of LDAP attributes to outgoing claim types** section, select **Name ID**.
   d. In the **LDAP Attribute** column, click the dropdown arrow in the first row and then select **SAM-Account-Name**.
   e. In the **Outgoing Claim Type** column, click the dropdown arrow in the first row and then select **Name ID**.
   f. Click **Finish**.

Restart Reporting

On the **Reporting** node, restart Reporting using the following command:

```
systemctl restart reporting
```

Configure Single Sign-on with SAML and Okta

Configuring single sign-on with SAML and Okta involves:

- creating a new Okta/SAML application integration
- editing the metadataUrl parameter in the investigator.yml file
- restarting Reporting

Create a New Okta/SAML Application Integration

1. Log in to the Okta Administrator Dashboard.
2. If you are using the Developer Console, you must switch to the Admin Console (Classic UI). If you see Developer Console in the top left of the page, click it and select Classic UI to switch.
3. Click **Applications** in the menu bar.
4. Click **Add Application**, and then click **Create New App**.
5. On the **Create a New Application Integration** page, do the following:
    - In the **Platform** box, click the dropdown arrow and then select **Web** from the list.
    - In the **Sign on method** section of the page, select the **SAML 2.0** radio button.
    - Click **Create**.

    The **Create SAML Integration** window is displayed and opens at the **General Settings** tab.
6. On the **General Settings** tab, enter a name for the app in the **App name** field, and then click **Next**.
7. On the **Configure SAML** tab, do the following:
    - In the **Single sign on URL** box, enter `<https://<reporting_node_fqdn/api/actions/login/saml/sso>`. This must match the URL you set in the **assertionConsumerServiceURL** parameter the **investigator.yml** file.
    - Select the **Use this for Recipient URL and Destination URL** checkbox.
    - In the **Audience URI (SP Entity ID)** box, enter `<https://<reporting_node_fqdn/api/actions/login/saml/sso>`. This must match the URL you set in the **assertionConsumerServiceURL** parameter in the **investigator.yml** file.
8. Scroll to the bottom of the page, and click **Next**.
9. On the **Feedback** tab, do the following:
    - Beside **Are you a customer or partner?**, select the **I'm an Okta customer adding an internal app** radio button.
    - Optionally set the **App type**.
10. Click **Finish**.

    The application's **Sign on** dialog box appears.
11. Select **Identity Provider metadata** in the yellow application metadata box, and copy the URL to your clipboard.

    In the next section, you will copy the URL to the **metadataUrl** parameter in the **investigator.yml** file.

Edit the investigator.yml File

1. On the **Reporting** node, navigate to the **/opt/interset/etc** directory, and open the **investigator.yml** file.
2. Scroll to the **SAML Authentication** section, and replace the **metadataUrl** parameter with the **Identity Provider metadata** URL you copied in the previous section.

   ```
   enabled: true
   rootUser: <rootUser_email>
   defaultTenantId: <TID>
   relyingPartyIdentifier: Interset_Investigator_identifier
   assertionConsumerServiceURL: https://<reporting_node_fqdn>/api/actions/login/saml/sso
   metadataUrl:
   ```

3. Update the **logoutUrl** parameter with the URL that you want users taken to when they click **Sign out, <*username*>** in the Interset user interface.

   For example, enter the logout URL configured in Okta.

   ```
   logoutUrl: <signout_destination_URL>
   ```

4. Save and close the updated **investigator.yml** file.

Restart Reporting

1. On the **Reporting** node, restart Reporting using the following command:

   ```
   systemctl restart reporting
   ```

2. In a Web browser, log in to Interset as your new SAML **rootUser.**

   In the **Tenants** page, under **Administrative Tenant**, your new SAML **rootUser** appears in the list of users.

3. In the **Administrative Tenant** page, disable the system (local) **root** user.

Configure Workflow

Workflow uses Storm, a distributed real-time computation system, to apply rules to the data ingested into Interset. Workflow also uses the Reporting service, which is why the Reporting server is validated before you begin the Workflow setup.

> **Important**: For the current version of Interset we do not recommend setting up Storm in HA environment.

1. In a command console on the Monitoring node, run the following commands to navigate to the **/opt/interset/installer** directory and launch the Interset installation menu.

   ```
   cd /opt/interset/installer
   ```

   ```
   ./install.sh
   ```

2. From the available installation options, type **12** to select installation **[12] Workflow configuration**, and then press **Enter**.

   ```
   Please select an installation option from the list above: 12
   ```

3. On the Master node, navigate to the **/opt/interset/rules/conf** directory, and open the **rules.conf** configuration file.

4. If you plan to use the schema registry, do the following:

   - In the the **Service Connections** section of the **rules.conf** file, under **Schema Registry**, verify that the **ConfluentRegistry** parameter is not commented out.

     The **ConfluentRegistry** parameter must be enabled when you configure Interset to use the schema registry.

     For more information about the schema registry, please contact Micro Focus Interset Support.

   - In the **Event Processing** section, uncomment and then enter the required values for the **SchemaSubjectEvents** parameter(s), using the values defined in the table below.

     ```
     #SchemaSubjectEvents.<schemaName1> = <archType1>
     ```

     For example, for repository data, the **SchemaSubjectEvents** parameter is entered as:

     ```
     SchemaSubjectEvents.RepositoryRecord = repo
     ```

     > **Tips:**
     >
     > - The ***<schemaName>*** is the value of the name field from within the Avro schema itself, for example, **AccessRecord**.
     > - Enter as many **SchemaSubjectEvents** parameters as required for your data source types.

Data Source Type	<archType>
Access	access
Active Directory	ad, active_directory, activedirectory
Email	email
Endpoint	endpoint, wdc
Expense	expense
Alert	interset-standard, intersetstandard, alert, uaf
Linux AuditD	auditd, linuxauditd
NetFlow	netflow
Printer	printer
Repository	repo, repository
Sensor	sensor
VPN	vpn
Web Proxy	proxy, webproxy, webproxy
Windows Printer	winprinter, windows_printer, windowsprinter

5. If you do not plan to use the schema registry, do the following:
   - In the the **Service Connections** section of the **rules.conf** file, under **Schema Registry**, comment out the **ConfluentRegistry** parameter.
   - In the **Event Processing** section, update the **non-Schema Registry path** parameters by:

     ```
     #TopicEvents.<kafkaTopicName1> = <archType1>
     ```

     - removing the hash (**#**) symbol from the **TopicEvents** parameter to enable this setting,
     - creating as many **#TopicEvents.<kafkaTopicName1>** entries as necessary for your tenant and Workflow data source types,
     - entering the appropriate **<kafkaTopicName>** for the data source type, and
     - entering the relevant **<archType>** value from the table below.

Data Source Type	<archType>
Access	access
Active Directory	ad, active_directory, activedirectory
Email	email
Expense	expense
Alert	interset-standard, intersetstandard, alert, uaf
Linux AuditD	auditd, linuxauditd
NetFlow	netflow
Printer	printer
Repository	repo, repository
Sensor	sensor
VPN	vpn
Web Proxy	proxy, webproxy, webproxy
Windows Printer	winprinter, windows_printer, windowsprinter

   - To configure Workflow for an authentication data source other than Active Directory, do the following:

- Locate the KafkaSpoutTopics parameter.
- Edit the **KafkaSpoutTopics** parameter to include the spout topics for the other authentication data source as a comma-separated list.

  For example, to enable auditD and VPN, add the following spout topics:

  ```
  KafkaSpoutTopics = interset_auditd_events_0_0, interset_vpn_events_0_0
  ```

  and then save the **rules.conf** file.
- Still on the Master node, run the following commands to kill the currently-running topology, and then redeploy it:

  ```
  /opt/interset/rules/bin/workflow.sh --kill /opt/interset/rules/conf/rules.conf
  ```

  ```
  /opt/interset/rules/bin/workflow.sh --deploy /opt/interset/rules/conf/rules.conf
  ```

  Use these same steps to enable other datasources such as repository, WebProxy, and so on.

6. To configure Workflow email notifications, do the following:
   - In the **#Notification Settings** section, change the value of the **EmailOutputEnabled** parameter to **true**.

     ```
     EmailOutputEnabled = true
     ```
   - Enter the appropriate email configuration information:

     ```
     EmailServer = <email_server_name>
     EmaiServerPort = <email_server_SMTP_port>
     EmailSetSSLOnConnect = true
     EmailUser = <email_login>
     EmailPassword = <email_password>
     EmailFromAddress = <your_FROM_address>
     ```
7. To receive text message notifications via your smart phone, enter your credentials as follows:

   > ✓ Tip: You must have a Twilio account (www.twilio.com).

   ```
   SmsSid = <your_twilio_SID>
   SmsToken = <your_twilio_API_token>
   SmsFromNumber = <your_twilio_FROM_number>
   ```
8. Save the updated **rules.conf** file.
9. To deploy the updated Workflow configuration, run the following command:

   ```
    /opt/interset/rules/bin/workflow.sh --deploy /opt/interset/rules/conf/rules.conf
   ```
10. To ensure that the current topology is running, run the status command:

    ```
    /opt/interset/rules/bin/workflow.sh --status /opt/interset/rules/conf/rules.conf
    ```

    The expected response includes information on **Topology_name**, **STATUS**, **Num_tasks**, **Num_workers**, and **Uptime-secs**.
11. To validate the currently running topology, run the following command:

    ```
    /opt/interset/rules/bin/workflow.sh --validate /opt/interset/rules/conf/kafka-config.properties
    /opt/interset/rules/conf/rules.conf
    ```

Install Apache Nifi

Before setting up NiFi, navigate to **/opt/interset/installer/etc/nifi_config** file on the **Monitoring node** and set the following properties for **NIFI_UI_SECURE** and **NIFI_LOAD_MARKETPLACE**.

```
# NIFI_UI_SECURE can be either not secured (N), or secured with TLS and UI authentication via Kerberos (K),
# Yes (Y) is NOT a valid option
NIFI_UI_SECURE=N
# NIFI_LOAD_MARKETPLACE can be loaded for development/testing (Y) or not for production (N)
NIFI_LOAD_MARKETPLACE=N
```

Interset 5.9.2 uses NiFi for the data extraction, transformation, and loading. After you install NiFi, you configure a process group to ingest a new data source.

1. In a command console on the Monitoring node, run the following commands to navigate to the **/opt/interset/installer** directory and launch the Interset installation menu.

   ```
   cd /opt/interset/installer
   ```

   ```
   ./install.sh
   ```

2. From the available installation options, type **13** to select installation **[13] NiFi installation on NiFi node**, and then press **Enter**.

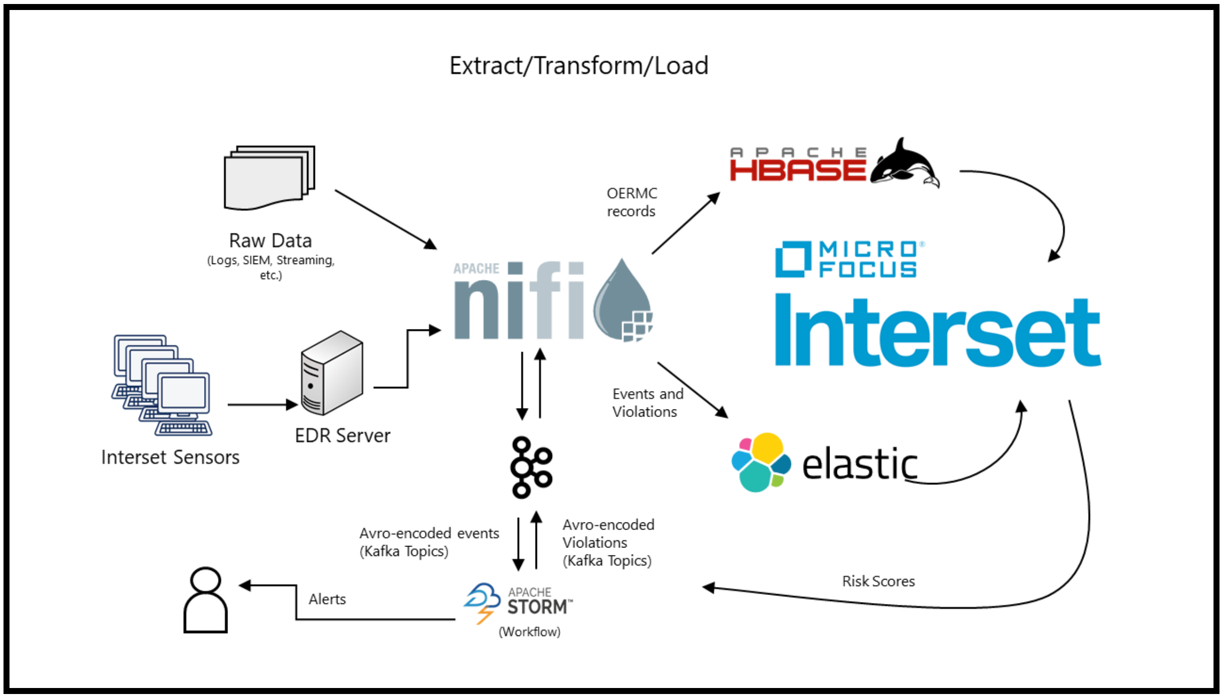
   ```
   Please select an installation option from the list above: 13
   ```

   When the NiFi installation is complete, the expected result is:

   ```
   NiFi installation Complete!
   ```

Configure a New Data Source

With Interset 5.9.2, the data ingest process uses Apache NiFi to extract, transform, and load data. The following diagram illustrates the data flow in Interset 5.9.2.



When you configure data ingest in Interset 5.9.2, you:

- configure an SSL controller service, which is required by some of the processors in the the **Interset Marketplace** template
- create a process group to separate the new data flow from the **Interset Marketplace** template
- specify the location of the Interset schema registry
- enable the controller services
- configure the processors in the NiFi flow that are responsible for extracting, transforming, and loading the data
- start the data flow

For this data ingest, we use the sample AD dataset (**ad_sample.csv.gz**) provided by Interset.

If you don't already have the sample authentication dataset, you can obtain it from the Interset online customer repository. If you require access credentials, please contact Micro Focus Interset Support at interset.support@microfocus.com.

Ensure that you copy the sample data to a directory accessible to NiFi. Make a note of this location; you will need it later.

Configure an SSL Context Service

Some of the processors in the **Interset Marketplace** template require an SSL context service to be specified. To configure an SSL context service, do the following:

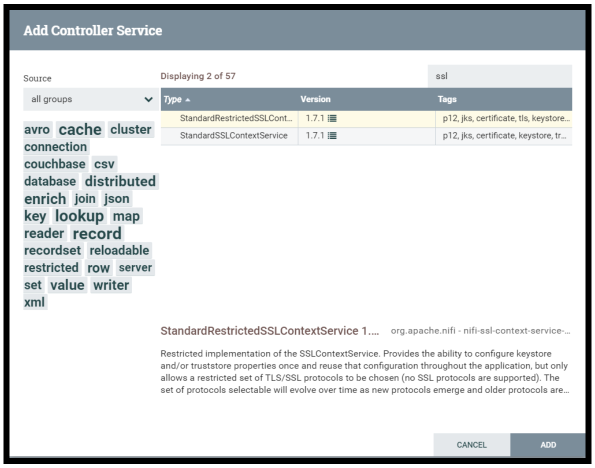
1. Open a browser and navigate to **<*nifi_node_fqdn*>:8085/nifi/**.

   > In a secure cluster, the NiFi URL is **https://<*nifi_node_fqdn*>:8085/nifi/**

2. Right-click the canvas, then click **Configure**.
3. In the **NiFi Flow Configuration** dialog, select the **Controller Services** tab.
4. Click **Create a New Controller Service** ✚. The **Add Controller Service** dialog opens.
5. In the **Filter** text box (above the **Tags** column), type **ssl** to filter the list of services.

   You should see the following services:

   - StandardRestrictedSSLContextService
   - SSLContextService

   Add Controller Service

   Source: all groups

   Displaying 2 of 57

   ssl

Type ▲	Version	Tags
StandardRestrictedSSLCont...	1.7.1	p12, jks, certificate, tls, keystore...
StandardSSLContextService	1.7.1	p12, jks, certificate, keystore, tr...

   StandardRestrictedSSLContextService 1.... org.apache.nifi - nifi-ssl-context-service-...

   Restricted implementation of the SSLContextService. Provides the ability to configure keystore and/or truststore properties once and reuse that configuration throughout the application, but only allows a restricted set of TLS/SSL protocols to be chosen (no SSL protocols are supported). The set of protocols selectable will evolve over time as new protocols emerge and older protocols are...

6. Select **StandardRestrictedSSLContextService**, and then click **Add**.

   The service is added to the list of Controller Services in the **NiFi Flow Configuration** dialog.

7. In the right-most column of the Controller Services list, click **Configure** ⚙.
8. In the **Configure Controller Services** dialog, set the properties as follows:
   - **Keystore Filename**: The path to the Keystore file. The default location is **/etc/security/interset/sslcert/localhost-keystore.jks**.
   - **Keystore Password**:The password for the Keystore.
   - **Key Password**: The password for the key. If this is not specified, but the Keystore Filename, Password, and Type are specified, then the Keystore Password will be assumed to be the same as the Key Password.
   - **Keystore Type**: JKS

- **Truststore Filename**: The path to the trust store file. The default location is **/etc/security/interset/sslcert/all-truststore.jks**.
- **Truststore Password**:The password for the Truststore.
- **Truststore Type**: JKS
- **TLS Protocol**: TLS

9. Click **Apply**.

Create a Process Group

> We strongly recommend that you create a new process group to build your data flow. This ensures that the processors in the **Interset Marketplace** template will not be unintentionally started.

To create a new process group:

1. If you haven't already, open a browser and navigate to **<*nifi_node_fqdn*>:8085/nifi/**.

   > In a secure cluster, the NiFi URL is **https://<*nifi_node_fqdn*>:8085/nifi/**

2. In NiFi, double-click the **Interset Marketplace** template to open it.
3. In the **Additional Deliverables - For ingesting data into HBase and Elasticsearch** (yellow) canvas group, right-click the **AD Loading (Generic CSV) Configuration** process group and click **Copy**.
4. Right-click the **Interset Marketplace** canvas and select **Leave Group**.
5. Right-click the **NiFi Flow** canvas and select **Paste**. A copy of the **AD Loading (Generic CSV) Configuration** process group appears on your canvas next to **Interset Marketplace** template.

You can optionally rename the new process group:

1. Right-click the **AD Loading (Generic CSV) Configuration** process group you just created, and then click **Configure**.
2. On the **General** tab, enter a new name in the **Process Group Name** field.
3. Click **Apply**, and click **OK** when the change is confirmed.
4. Close the configuration dialog to return to the main canvas.

Configure the Schema Registry

1. Right-click the process group you just created, and select **Configure**.
2. On the **Controller Services** tab, scroll down to the **ConfluentSchemaRegistry - Interset Avro Schemas** controller service, and click anywhere in the row to select it.

3. In the right-most column of the table, in the selected row, click the right-arrow →, then click **View Configuration** ⚙.
4. The **Configure Controller Service** dialog opens.
5. On the **Properties** tab, do the following:
   - In the **Schema Registry URLs** field, enter the URL and port of the Interset schema registry.

     > The default location is **https://<*master_node_fqdn*>**:9190

   - In the **SSL Context Service** field, select the SSL context service that you set up in the Configure an SSL Context Service section.
6. Click **Apply**.
7. Close the **NiFi Flow Configuration** dialog to return to the main canvas.

Enable the Controller Services

Before you can run any processors in NiFi, you must enable the controller services for the process group you just created.

To enable controller services:

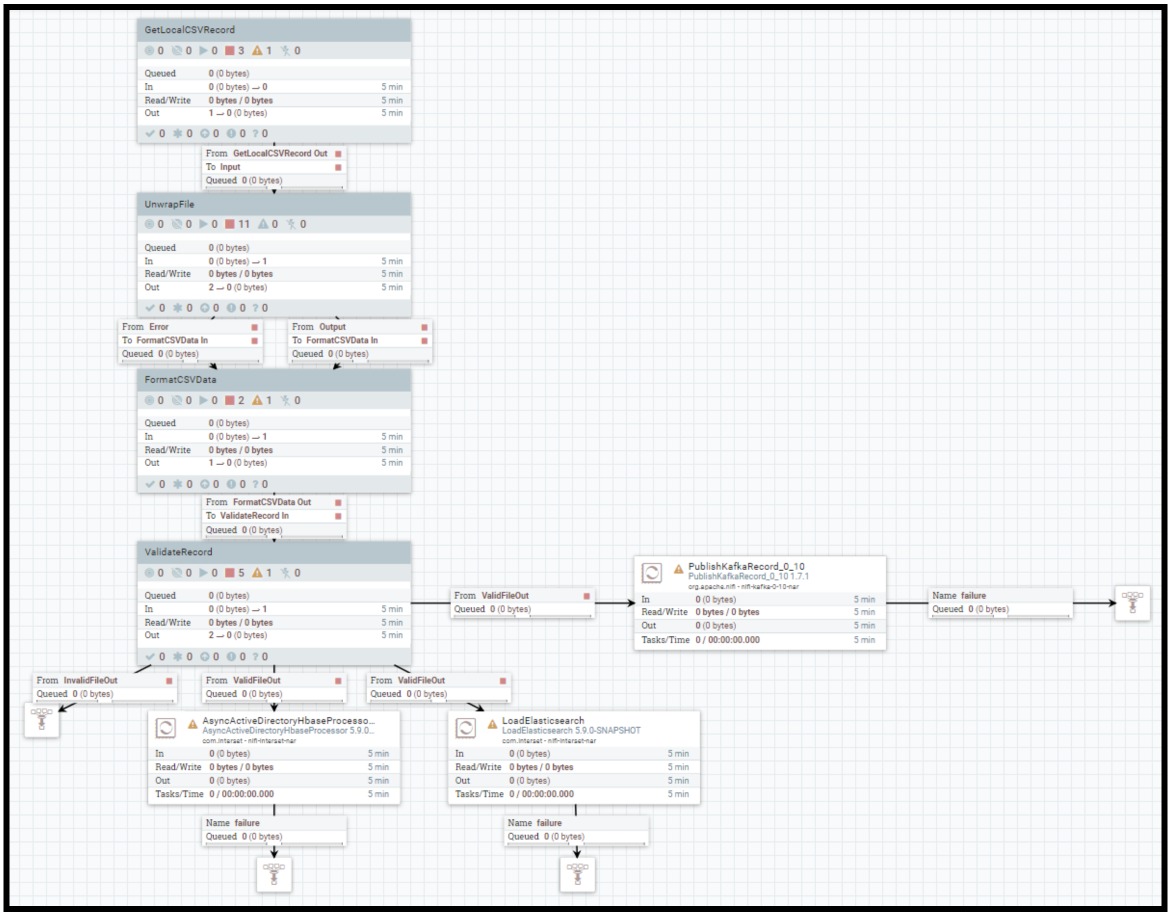
1. Right-click the process group, and then click **Configure**.
2. Select the **Controller Services** tab.
3. For each controller service displayed, ensure that **Enabled** is displayed in the **State** column.
4. For any disabled controller services, click the right-arrow →, and then click Enable ϟ. The **Enable Controller Service** dialog is displayed.
5. Click **Enable**, and then click **Close**.
6. Repeat for all disabled controller services, and then close the configuration dialog.

> **Important**:It may happen that some services won't start because of dependencies.
>
> If that happens, following the right-arrow → to dependent service to **Start** or **Configure** as required.

Configure the Data Flow

In this section, you configure various processors in the NiFi flow that are responsible for extracting, transforming, and loading the data into Interset.

Enter the process group you just created by double-clicking it. You should see the following:



To configure the data flow, you edit properties in the following process groups:

- GetLocalCSVRecord
- FormatCSVData
- ValidateRecord

You must also set properties in the following processors:

- AsyncActiveDirectoryHbaseProcessor
- LoadElasticsearch
- PublishKafkaRecord_0_10

Configure the GetLocalCSVRecord Process Group

1. Double-click the **GetLocalCSVRecord** process group to enter it.
2. Right-click the **ListFile** processor and select **Configure**.
3. On the **Properties** tab, next to **Input Directory**, enter the path to the directory where your data is stored (for example, **/opt/interset/data/**), and then click **Apply**.
4. Right-click the canvas and select **Leave Group** to return to the top level of the flow.

Configure the FormatCSVData Process Group

1. Double-click the **FormatCSVData** process group to enter it.
2. Right-click the **UpdateRecord - Format CSV Fields** processor and select **Configure**.
3. On the **Properties** tab, verify that the expression in the **/time** field contains a date format for the incoming records that matches the date format in the incoming file. For example, the format `"yyyy/MM/dd HH:mm:ssXXX"` matches dates like `2017/04/01 08:01:15-05:00`, whereas the format `"yyyy-MM-ddTHH:mm:ssXXX"` matches date like `2017-04-01T08:01:15-05:00`.
4. If you made changes to the date format, click **Apply**; otherwise, click **Cancel**.
5. Right-click the canvas and click **Leave Group** to return to the top level of the flow.

Configure the AsyncActiveDirectoryHbaseProcessor

To set up this Hbase processor, you must:

- configure the NiFi node environment
- configure the NiFi processor properties

Configure the NiFi Node Environment

1. Run the following command on the **Nifi** node (replace <KERBEROS_DOMAIN.COM> with the realm you set when you configured Kerberos) to ensure that the NiFi user has an active authentication ticket through Kerberos:

   ```
   sudo -u nifi kinit -kt /etc/security/interset/keytab/i_nifi.service.keytab i_nifi/<nifi_node_fqdn>@<REALM>
   ```

   Verify the Validity of the ticket by running the following command. This will provide you with the ticket details about the default principal, start and Expiry date along with the due date for next renewal.

   ```
   sudo -u nifi klist
   ```

2. For CDH, run the following command on the **Compute** node to determine the path to the Region-server Key tab and initialize the key :

   ```
   sudo -u hbase kinit -kt /var/run/cloudera-scm-agent/process/`sudo ls -lrt /var/run/cloudera-scm-agent/process/ |
   awk '{print $9}' |grep 'hbase-MASTER$\|hbase-REGIONSERVER$'| tail -1`/hbase.keytab hbase/<compute_node_
   fqdn>@<REALM>
   ```

   Verify the Validity of the ticket by running the following command. This will provide you with the ticket details about the default principal, start and Expiry date along with the due date for next renewal.

   ```
   sudo -u hbase klist
   ```

3. For HDP, run the following command to initialize the key:

   ```
   sudo -u hbase kinit hbase/<compute_node_fqdn>@<REALM> -kt /etc/security/keytabs/hbase.service.keytab
   ```

4. Run the following command on the **NiFi** node to restart NiFi:

   ```
   service nifi restart
   ```

> **Important**:For ticket renewal use the following commands within the crontab file. These commands will ensure logging for any errors during the renewal of the ticket.
>
> Run on the **Compute node(s)**
>
> ```
> sudo -u hbase sh -c "kinit -R > /tmp/hbase_kinit.log 2>&1"
> ```
>
> Run on the **Nifi node(s)**
>
> ```
> sudo -u nifi sh -c "kinit -R > /var/log/nifi/kinit_cron.log  2>&1"
> ```

Configure the AsyncActiveDirectoryHbaseProcessor Properties

1. Right-click the **AsyncActiveDirectoryHbaseProcessor** processor and select **Configure**.
2. Select the **Properties** tab.
3. In text box next to the **Zookeeper Quorum** property, enter the name of your Zookeeper hosts. The default location is ***<master_node_fqdn>***, for example, **master.interset.com**.
4. Ensure the **Zookeeper Node Parent** property is correctly set:
   - for HDP in an unsecured environment: `/hbase-unsecure`
   - for HDP in a Kerberos environment: `/hbase-secure`
   - for CDH: `/hbase`
5. Set value of the **Tenant ID (tid)** property to the tenant for which you are ingesting data.
6. Set the value of the **Data Instance ID (did)** to the appropriate Data Instance ID. The default is 0.
7. If you are running Interset on a secure cluster, set the properties required for Kerberos on the **Compute** node:
   - **Security Auth Enabled**: `true`
   - **Security Authentication**:`kerberos` (to enable regular Kerberos authentication) or `protected` (to enable encryption on RPC payloads).
   - **Security Kerberos Principal**:The name of the Hbase principal to use for Kerberos authentication.

     ```
     hbase/compute_node_fqdn>@<REALM>
     ```

   - **Security RPC Protection**: This property indicates whether to encrypt RPCs while they are traversing the network. Possible values are `authentication` (no encryption), `integrity` (no encryption) or `privacy` (encrypted). Requires authentication to be enabled. The default is `privacy`.
   - **Security SASL Client Config**:Set this property to the name of the section (for example, Client) in the JAAS configuration file you created in the previous section. This name is used to look up the configuration when NiFi is authenticating a user against a region server.

- **Append Classpath**: This property extends the classpath that the processor will run with, allowing the processor to access the **hbase-site.xml** client configuration file to establish the client connection with HBase. You can set this by using the following command:

  ```
  CDH: /etc/hbase/conf:/etc/hadoop/conf:/opt/cloudera/parcels/CDH/lib/hbase/lib/*:/opt/interset/phoenix_
  lib_links/*
  ```

  ```
  HDP: /etc/hbase/conf:/etc/hadoop/conf:/usr/hdp/current/hbase-client/lib/*:/opt/interset/phoenix_lib_
  links/*
  ```

8. Click **Apply** to close the processor properties and return to the main flow.

Configure the LoadElasticsearch Processor

1. Right-click the **LoadElasticsearch** processor and select **Configure**.
2. Select the **Properties** tab.
3. In text box next to the **Elasticsearch Hostnames** property, enter the name(s) and port(s) of your Elasticsearch hosts.

   > The default location is ***<search_node_fqdn>*****:9200**.

4. Edit the **Index Name** property to include the correct tenant ID. For example, if your tenant is 234, the value of the **Index Name** property must be set to `interset_ad_rawdata_234`
5. Set value of the **Tenant ID (tid)** property to the tenant for which you are ingesting data.
6. Set the value of the **Data Instance ID (did)** to the appropriate Data Instance ID. The default is 0.
7. If you are running Interset on a secure cluster, set the properties required for Kerberos as follows:
   - **Security User**: The Elasticsearch user and password. Must be in the format `<user>:<password>`. The default user is **elastic**.
   - **SSL Enabled**: `true`
   - **Truststore Path**:The path to the trust store file. The default location is **/etc/security/interset/sslcert/all-truststore.jks**.
   - **Truststore Password**: The password to the trust store.
   - **Keystore Path**: The path to the keystore file. The default location is **/etc/security/interset/sslcert/localhost-keystore.jks**.
   - **Keystore Password**: The password to the keystore.
8. Click **Apply** to close the processor properties and return to the main flow.

Configure the PublishKafkaRecord_0_10 Processor

> **Important**: If you are not using Workflow in your Interset installation, delete the connection leading to the **PublishKafkaRecord_0_10** processor. Otherwise, as the data ingest progresses, the processor queue will reach its maximum capacity and then apply back-pressure to the upstream processors, which will cause the entire ingest to stall.

1. Right-click the **PublishKafkaRecord_0_10** processor and select **Configure**.
2. Select the **Properties** tab.
3. In text box next to the **Kafka Brokers** property, enter the name(s) and port(s) of your Kafka Broker (s).

   ✓
   - The default location in Ambari is **<*stream_node_fqdn*>:6667**
   - In a secured cluster in Cloudera, the default location is **<*stream_node_fqdn*>:9093**
   - In an unsecured cluster in Cloudera, the default location is **<*stream_node_fqdn*>:9092**
   - In case of multiple stream nodes, each stream and port needs to be specified and separated with a coma. **< *stream_ node_fqdn*>:port,<*stream_node_fqdn*>:port...etc**

4. Edit the Topic Name property to include the correct tenant ID and data source ID. For example, for the default tenant ID (0) and data source ID (0), **Topic Name** would be set to `interset_ad_events_0_0`
5. If you are running Interset on a secure cluster, set the properties required for Kerberos as follows:
   - **Security Protocol**: Select SSL protocol to communicate with brokers. Corresponds to Kafka's `security.protocol` property.
   - **Kerberos Credentials Service**: No value required.
   - **Kerberos Service Name**: No value required.
   - **Kerberos Principal**: No value required.
   - **Kerberos Keytab**: No value required.
   - **SSL Context Service**: Select the SSL context service to use for communicating with Kafka (the service that you configured in the Configure an SSL Context Service section).
6. Click **Apply** to close the processor properties and return to the main flow.

Start the Data Flow

To start the data flow:

1. In a web browser, return to the NiFi user interface.
2. If you haven't already done so, enable the controller services in your top-level process group.

   > **Tip:**
   >
   > To enable controller services:
   >
   > 1. Right-click the process group, and then click **Configure**.
   > 2. Select the **Controller Services** tab.
   > 3. For each controller service displayed, ensure that **Enabled** is displayed in the **State** column.
   > 4. For any disabled controller services, click the right-arrow →, and then click Enable ⚡. The **Enable Controller Service** dialog is displayed.
   > 5. Click **Enable**, and then click **Close**.
   > 6. Repeat for all disabled controller services, and then close the configuration dialog.

3. Right-click the **AD Loading (Generic CSV) Configuration** group and select **Start**.

Your process flow is now running, and you should see the byte counts incrementing at each stage of the flow. You might need to refresh your view periodically by right-clicking the canvas and selecting **Refresh**.

Configure Violations Loading

1. If you haven't already, open a browser and navigate to **<*nifi_node_fqdn*>:8085/nifi/**.

   > In a secure cluster, the NiFi URL is **https://<*nifi_node_fqdn*>:8085/nifi/**

2. In NiFi, double-click the **Interset Marketplace** template to open it.
3. In the **Additional Deliverables - For ingesting data into HBase and Elasticsearch** (yellow) canvas group, right-click the **Violations Loading Configuration** process group and click **Copy**.
4. Right-click the **Interset Marketplace** canvas and select **Leave Group**.
5. Right-click the **NiFi Flow** canvas and select **Paste**. A copy of the **Violations Loading Configuration** process group appears on your canvas next to **Interset Marketplace** template.
6. Right-click the **Violation Loading** processor and select **Enter group**.
7. To configure **Violation to Hbase**, right-click on the process and click on configure.
8. Select the properties tab

Run Analytics

Run Analytics to confirm that the cluster configuration is correct and to periodically update the analytics. This can be done with a cron job, or manually.

> If you're analyzing NetFlow data and you do not have a unique entity identifier (such as a machine name or static IP addresses), run the following script before you run Analytics. This script disables models that compare new data to entities' historical baselines. You must run this script as the Spark user:
>
> ```
> sudo su spark
> /opt/interset/analytics/bin/sql.sh --action initialize --tenantID <tenantID> --dbServer <HBase_server_name>
> /opt/interset/bin/sysprep/scripts/nf_dynamic_ips_weights.sh <HBase_server_name> <tenantID>
> ```

To run Analytics, run the following command on the **Master node** as the **Spark** user:

```
/opt/interset/analytics/bin/analytics.sh /opt/interset/analytics/conf/interset.conf
```

Note: This command is already configured to run via cron in the **interset** user's crontab.

Running Analytics in a secure environment

> **Note:**Before running Analytics, you need to run kinit for i_spark.
>
> ```
> sudo -u i_spark kinit i_spark@<realm>.com -kt /etc/security/interset/keytab/i_
> spark.headless.keytab
> ```

To run Analytics in a secure environment, run the following command on the **Master node** as the **i_spark** user:

```
sudo runuser i_spark -- /opt/interset/analytics/bin/analytics.sh /opt/interset/analytics/conf/interset.conf
```

To ensure the validity of the ticket while analytics is still running the following can be used. This will prevent the ticket to expire during an active analytics run:

```
sudo -u i_spark kinit -R i_spark@<realm>.com && /opt/interset/analytics/bin/analytics.sh
/opt/interset/analytics/conf/t01.conf
```

While running Analytics through a scheduled cronjob, Interset recommends enabling **logging** for any errors while renewing the ticket and ensuring its validity. Use the following command for this:

```
0 0 * * * sudo -u i_spark sh -c \"kinit -R i_spark@<realm> /opt/interset/analytics/bin/analytics.sh
/opt/interset/analytics/conf/interset.conf > /opt/interset/log/analytics.log 2>&1\"
```

Configure the Search Indexes

> **Important:** You cannot configure Search until data has been ingested and Analytics has been run.

After running Analytics for a tenant for the first time, the search indexes must be set up. This step needs to be done only once per tenant; however, it must be done after the first Analytics run.

Setting up the Kibana indexes will enable the Search feature for the different data types.

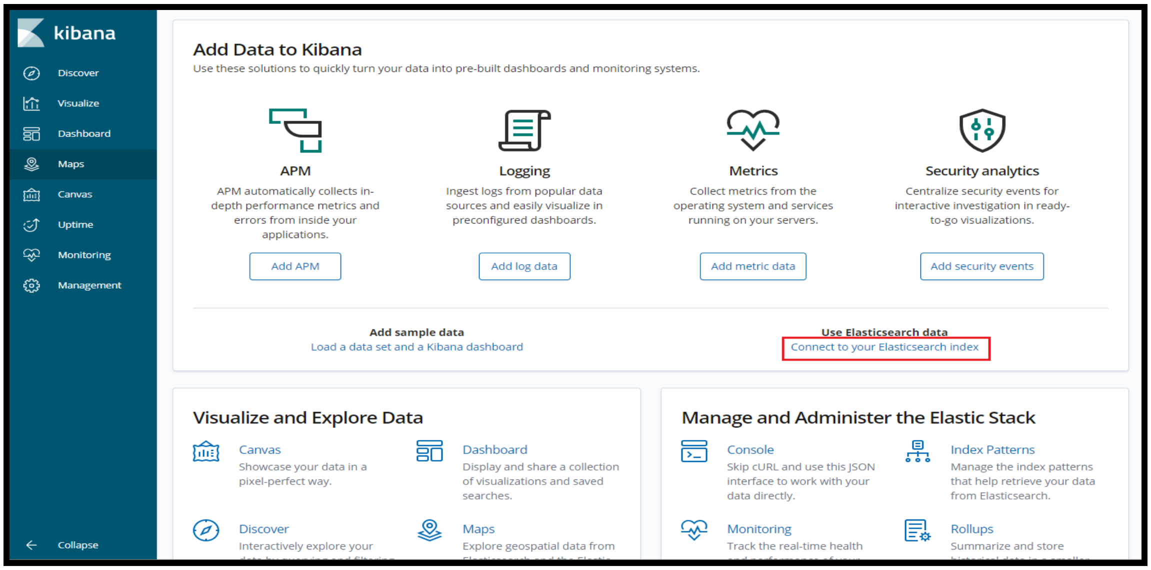
For this section, you need to know:

- the tenant ID (TID, the default is **0**)
- the data sources being used

  The data sources will always include violations (from Workflow), and one or more of Active Directory event logs, repository data, printer logs, and so on.

Steps

1. Open a Web browser, and go to **https://<*reporting_fqdn*>/search/<*tid*>**
2. In the Kibana home page, click **Connect to your Elasticsearch index**.



Use these solutions to quickly turn your data into pre-built dashboards and monitoring systems.

APM automatically collects in-depth performance metrics and errors from inside your applications.

Ingest logs from popular data sources and easily visualize in preconfigured dashboards.

Collect metrics from the operating system and services running on your servers.

Centralize security events for interactive investigation in ready-to-go visualizations.

Showcase your data in a pixel-perfect way.

Display and share a collection of visualizations and saved searches.

Interactively explore your

Explore geospatial data from

Skip cURL and use this JSON interface to work with your data directly.

Manage the index patterns that help retrieve your data from Elasticsearch.

Track the real-time health

Summarize and store

3. On the **Create index pattern** page, under **Step 1 of 2: Define index pattern** in the **Index pattern** box, enter the index name from the table below, replacing *<tid>* with your tenant ID.

   Although earlier releases allowed uppercase characters in the <*tid*>, the <*tid*> must be lower-case characters.

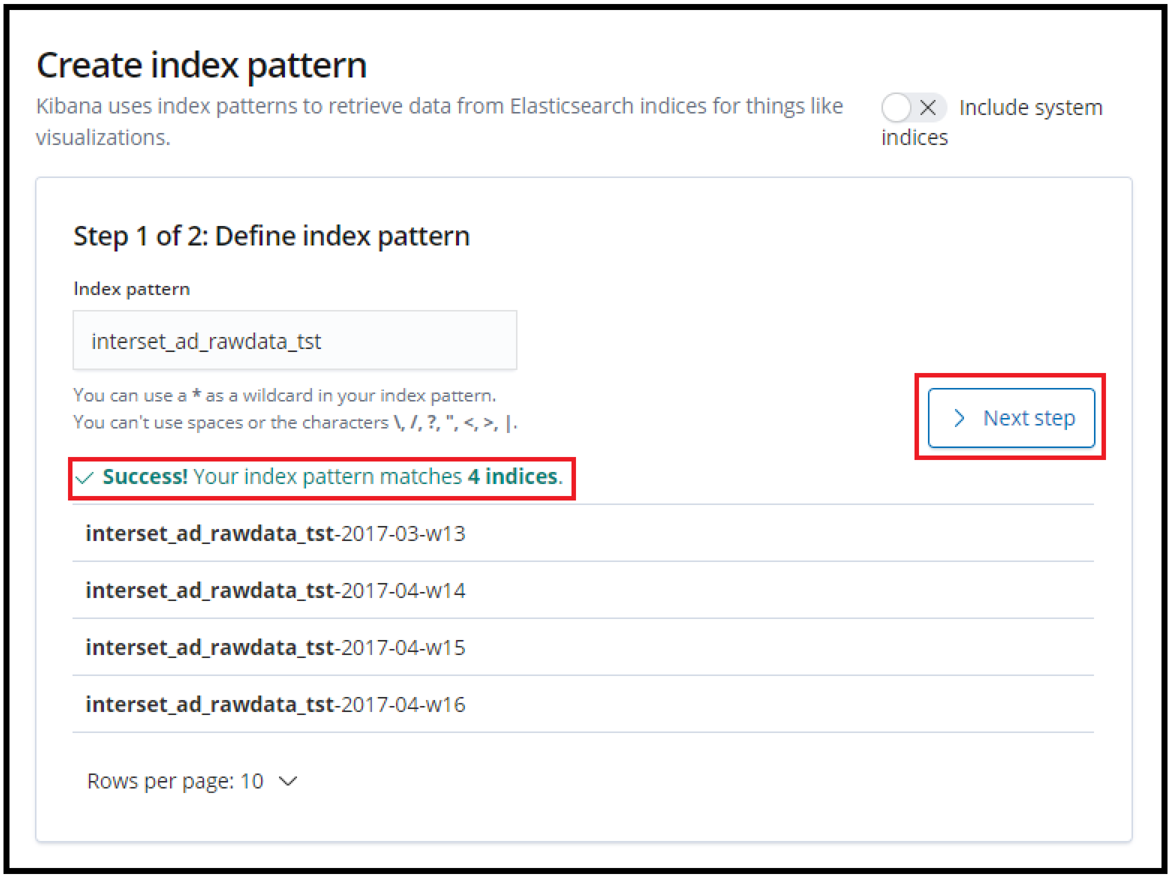
Data Type	Index Name	Timestamp Field
All	interset_*_rawdata_<tid>	timestamp
Access	interset_access_rawdata_<tid>	timestamp
Access, Active Directory, Authentication auditD, VPN	interset_auth_rawdata_<tid>	timestamp
auditD	interset_auditd_rawdata_<tid>	timestamp
Authentication	Interset_authentication_rawdata_<tid>	timestamp
Email	interset_email_rawdata_<tid>	timestamp
Finance Expense	interset_expense_rawdata_<tid>	timestamp
Interset EDR Server	interset_sensor_rawdata_<tid>	timestamp
NetFlow	interset_netflow_rawdata_<tid>	timestamp
Printer	interset_printer_rawdata_<tid>	timestamp
Repository	interset_repo_rawdata_<tid>	timestamp
Universal Alert	interset_violations_<tid>	timestamp
VPN	Interset_vpn_rawdata_<tid>	timestamp
Web Proxy	interset_webproxy_rawdata_<tid>	timestamp
Workflow Violations	interset_violations_<tid>	timestamp

Notes:

- The Workflow Violations **interset_violations_<*tid*>** index will not exist unless a violation was triggered by a Workflow.
- You will always need to add the Workflow index, and one or more of the other data types.
- If Kibana does not recognize the index pattern, do not create it at this time.
- When you configure the **All** data type **interset_*_rawdata_<*tid*>** index, you are creating a search index for queries that run against all ingested data types. This type of search index is useful when, for example, you want to see all the Analytics data for one specific entity (user, machine, controller, and so on) across all data types.

✓ **Tip:** When you enter your index pattern name, Kibana provides an indication whether your index name matches existing indexes.

4. Click **Next step**.



Create index pattern

Kibana uses index patterns to retrieve data from Elasticsearch indices for things like visualizations.

Include system indices

Step 1 of 2: Define index pattern

Index pattern

interset_ad_rawdata_tst

You can use a * as a wildcard in your index pattern.
You can't use spaces or the characters \, /, ?, ", <, >, |.

✓ Success! Your index pattern matches 4 indices.

interset_ad_rawdata_tst-2017-03-w13

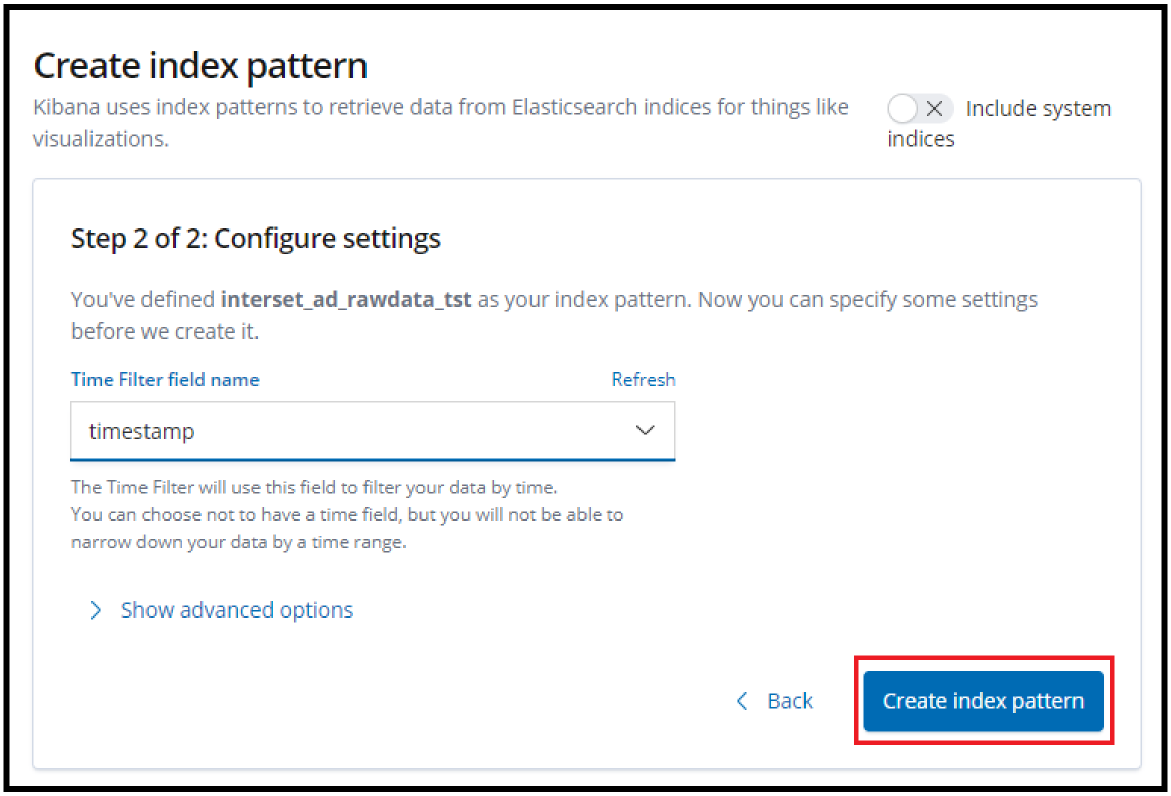
interset_ad_rawdata_tst-2017-04-w14

interset_ad_rawdata_tst-2017-04-w15

interset_ad_rawdata_tst-2017-04-w16

5. Under **Step 2 of 2: Configure settings**, in the **Time Filter field name** box, click the dropdown arrow, and then select the appropriate **Timestamp Field** value from the table above.

6. Click **Create index pattern**.



Create index pattern

Kibana uses index patterns to retrieve data from Elasticsearch indices for things like visualizations.

Step 2 of 2: Configure settings

You've defined **interset_ad_rawdata_tst** as your index pattern. Now you can specify some settings before we create it.

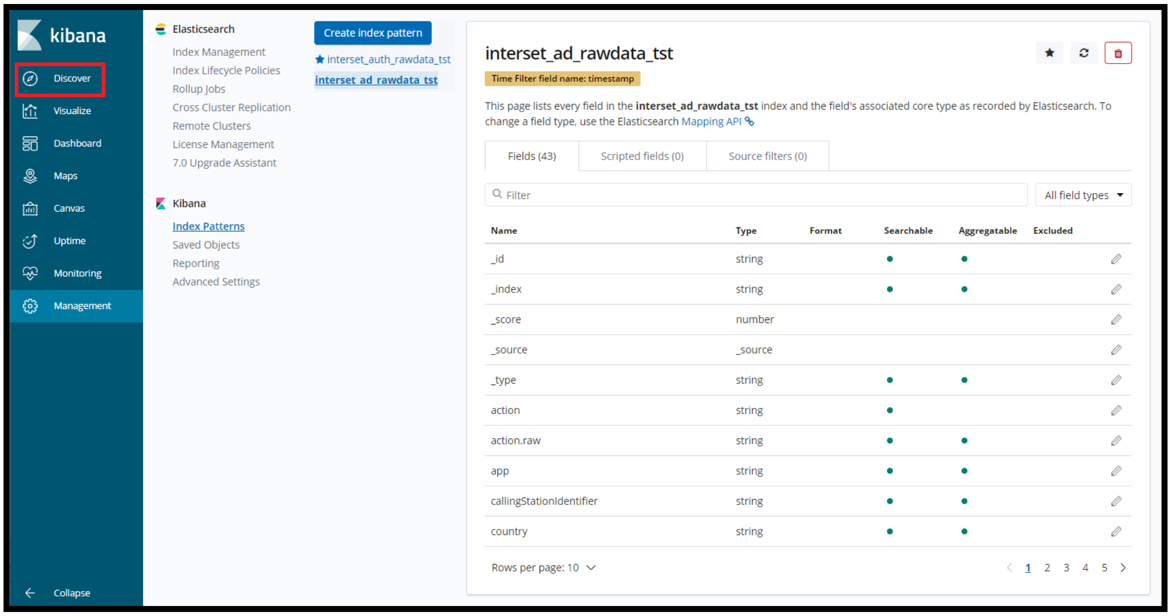
Time Filter field name: timestamp

The Time Filter will use this field to filter your data by time.
You can choose not to have a time field, but you will not be able to narrow down your data by a time range.

Kibana displays a summary of the new index pattern, including a list of all of the data fields, each field's core type, and whether the field is searchable and aggregatable. To the left of the summary table, Kibana provides a list of all configured index patterns.

7. For each subsequent data type, click **+Create Index Pattern** to the left of the index pattern summary table, and then repeat Steps 3 through 7.

8. After creating index patterns and thus enabling Search for one or more data types, click **Discover** in the Kibana sidebar menu to begin searching.



interset_ad_rawdata_tst

Time Filter field name: timestamp

This page lists every field in the **interset_ad_rawdata_tst** index and the field's associated core type as recorded by Elasticsearch. To change a field type, use the Elasticsearch Mapping API

Name	Type	Format	Searchable	Aggregatable	Excluded
_id	string		●	●	
_index	string		●	●	
_score	number				
_source	_source				
_type	string		●	●	
action	string		●		
action.raw	string		●	●	
app	string		●	●	
callingStationIdentifier	string		●	●	
country	string		●	●	

Enable Windowed Analytics

After your Interset cluster is installed and configured, your historical data has been ingested and analyzed, and data is appearing in the Interset user interface, we recommend that you optimize the ongoing processing of data by enabling **Windowed Analytics**.

By default, Interset is configured in batch mode, which means that when new data is ingested, Analytics is run on both the new and the existing data. Although this process is beneficial when you first install and configure your Interset cluster (for testing and validation purposes), running Analytics on the entirety of your data on an ongoing basis unnecessarily uses system resources.

When you enable Windowed Analytics, you configure Interset to run Analytics only on newly ingested data, as determined by the date of the last Analytics run and the timestamp of the data. Interset identifies the data it has already analyzed, and then runs Analytics only on the new data. These results are then aggregated with the existing results to produce updated, current Analytics results for the entire data set.

Windowed Analytics has a positive impact on performance and stability because it allows the system to analyze and aggregate smaller, more consistently sized quantities of data than batch mode, particularly as the total amount of data in your system continues to grow.

> **Note**: After you have validated the initial data ingest and Analytics run for your Interset cluster, you might need to ingest and analyze historical data. In this scenario, you must continue to run Analytics in batch mode to ensure all the data is included.

Steps

1. Navigate to **/opt/interset/analytics/conf/** and open the **interset.conf** file.
2. In the **Analytics configuration** section, edit the **batchProcessing** parameter as follows:

   Change

   ```
   batchProcessing = true
   ```

   to

   ```
   batchProcessing = false
   ```

Windowed Analytics is now enabled, and will begin with the next Analytics run.

> The first Windowed Analytics run for a tenant performs a full batch run to establish the baseline for the system going forward. The second and subsequent runs occur as Windowed Analytics.

Configure the 'Peek-Back' Window for Windowed Analytics

The 'peek-back' window is a best-effort buffer that ensures that delayed or out-of-order data is not missed between Windowed Analytics runs.

The value for the peek-back window is specified in milliseconds. The default value for the peek-back window is 24 hours (86400000.0 milliseconds).

This means that when a Windowed Analytics job runs, it loads data starting from the end of the last run, minus 24 hours. For example, if the previous windowed run started at midnight and completed successfully at noon today, the next run loads data beginning from noon yesterday. The intent of this overlap is to include data that was ingested after midnight, but that has event timestamps between noon and midnight of the previous day.

The peek-back window can be adjusted as follows:

1. As the **interset** user, log in to the **Master** node where Interset is installed.
2. At the command console, run the following command:

   ```
   /opt/interset/analytics/bin/sql.sh --action console
   ```

   > Note: If you are using Kerberos, you must run `kinit` and specify a keytab/principal with appropriate permissions in HBase before you can successfully run this command.

3. In the Phoenix console, enter the following commands, replacing `VALUE_IN_MS_HERE` with the desired value of the peek-back window, specified in milliseconds (for example, 86400000.0):

   ```
   UPSERT INTO PARAMETERS(TID, NAME, VAL) VALUES('0', 'SCORE WINDOW BUFFER', VALUE_IN_MS_HERE);
   UPSERT INTO PARAMETERS(TID, NAME, VAL) VALUES('0', 'AGGREGATE WINDOW BUFFER', VALUE_IN_MS_HERE);
   ```

4. Type `!quit` to exit the console.

Appendix A: Configure the Sample Data

This appendix provides an end-to-end introduction to Interset, using sample datasets. These tasks should only be performed after Interset has been installed and configured in your environment.

When you configure the sample data, you will:

- create a dedicated tenant for the sample data;
- configure NiFi to ingest sample data from universal Active Directory logs in .csv format;
- run Interset Analytics; and
- review the top risky users and their associated most significant anomalies and underlying events.

The sample dataset includes four different data types (authentication, repository, Web proxy, and NetFlow) to demonstrate that Interset provides an integrated view of risk across multiple datasets. As Interset Analytics is designed to highlight the top risky entities, IT security team members can therefore prioritize their efforts and investigations.

The sample datasets are small, to reduce any impact on the performance or storage requirements of your Interset production installation.

For the purposes of this example, we will only be ingesting sample authentication data.

Create the Samples Tenant

By creating a separate Samples tenant, you reduce any impact to your deployment. This Guide uses **int** as the tenant ID (TID) but you may choose an alternate tenant ID.

1. Log in to Interset as **root** (the default password is **root**).
2. Click **Settings**⚙ and then, in the dropdown list, select **Tenants.**
3. On the **Tenants** page, click **New**.
4. In the **Create a new Tenant** dialog, enter a new **Tenant ID** (in this case, **int**) and **Tenant Name**.
5. Click **Create Tenant.** The new tenant appears in the tenant list.

Create an Administrator User for the Samples Tenant

For users to log in to Interset, they must have a username and password.

1. Log in to Interset as **root** (the default password is **root**).
2. Click **Settings**⚙ and then, in the dropdown list, select **Tenants.**
3. On the **Tenants** page, select the tenant you just created from the list on the left, and then click **New User**.

4. In the **Create a User for <tenant name>** dialog, enter the **Name**, **Username**, **Role**, and **Password** for the new user.

   For the purposes of this exercise, select the **admin** role. The **admin** role can perform tasks such as configuring data sources, creating Workflows, and accessing the REST API, while the **user** role cannot.
5. Click **Create User**. The new user appears in the user list.

Copy the Interset Sample Datasets

The sample datasets are in the **sampledata** directory of the Interset repository.

The sample dataset includes three data types, each of which requires a separate data source configuration. In the Interset repository, locate the three sample datasets and copy them to a location that is accessible to the stream node(s). This location should have three directories, one for each dataset:

```
/opt/interset/sampledata/authentication
```

```
/opt/interset/sampledata/repository
```

```
/opt/interset/sampledata/webproxy
```

If you cannot access the repository, please contact Interset support at interset.support@microfocus.com.

Configure the Sample Authentication Data Source

Follow the instructions in "Configure a New Data Source" on page 97 to configure NiFi and ingest the sample data.

Run Analytics

To run Analytics, run the following command on the **Master node** as the **Spark** user:

```
/opt/interset/analytics/bin/analytics.sh /opt/interset/analytics/conf/interset.conf
```

When Analytics has completed its run, log into the Interset user interface to view the data, using the new tenant and user credentials.

Appendix B: Run the Enable Kerberos Wizard on Cloudera

The Interset installer automatically configures Kerberos on Cloudera. Use the steps below to run the wizard if you need to manually change the configuration.

To start the Kerberos wizard:

1. Go to the Cloudera Manager Admin Console and click ▼ to the right of the cluster for which you want to enable Kerberos authentication.
2. Select **Enable Kerberos**.

The steps below will guide you through the wizard to secure your cluster.

1. Getting Started
2. Setup KDC
3. Manage krb5.conf
4. Setup KDC Account
5. Configure Kerberos
6. Summary

Getting Started

The first page lists steps you should have completed before starting the wizard.

- Set up a working KDC. Cloudera Manager supports authentication with MIT KDC, Active Directory, and Red Hat Identity Management/FreeIPA.
- Configure the KDC to allow renewable tickets with non-zero ticket lifetimes.

  Active Directory allows renewable tickets with non-zero lifetimes by default. You can verify this by checking **Domain Security Settings** > **Account Policies** > **Kerberos Policy** in Active Directory.

  For MIT KDC, make sure you have the following lines in the **kdc.conf**.

  ```
  max_life = 1d
  max_renewable_life = 7d
  ```

- If you are using Active Directory, make sure LDAP over TLS/SSL (LDAPS) is enabled for the Domain Controllers.
- Hostnames *must* be in lowercase. If you use uppercase letters in any hostname, the cluster services will not start after enabling Kerberos.
- Install the OS-specific packages for your cluster listed in the table:

OS	Packages Required
RHEL 7 Compatible **RHEL 6 Compatible**	• `openldap-clients` on the Cloudera Manager Server host

OS	Packages Required
	• `krb5-workstation`, `krb5-libs` on ALL hosts • **(Red Hat IdM/FreeIPA only)** `freeipa-client` on all cluster hosts
SLES	• `openldap2-client` on the Cloudera Manager Server host • `krb5-client` on ALL hosts • **(Red Hat IdM/FreeIPA only)** `freeipa-client` on all cluster hosts
Ubuntu	• `ldap-utils` on the Cloudera Manager Server host • `krb5-user` on ALL hosts • **(Red Hat IdM/FreeIPA only)** `freeipa-client` on all cluster hosts
Windows	• `krb5-workstation`, `krb5-libs` on ALL hosts

- Create an account for Cloudera Manager that has the permissions to create other accounts in the KDC.

Important:

If YARN Resource Manager HA has been enabled in a non-secure cluster, before enabling Kerberos you must clear the StateStore znode in ZooKeeper, as follows:

1. Go to the Cloudera Manager Admin Console home page, click to the right of the YARN service and select **Stop**.
2. When you see a **Finished** status, the service has stopped.
3. Go to the YARN service and select **Actions** > **Format State Store**.
4. When the command completes, click **Close**.

Once you are able to check all the items on this list, click **Continue**.

Setup KDC

On this page, select the KDC type you are using: **MIT KDC**, **Active Directory**, or **Red Hat IPA**. Complete the fields as applicable to enable Cloudera Manager to generate principals/accounts for the CDH services running on the cluster.

Note:

- If you are using AD and have multiple Domain Controllers behind a Load Balancer, enter the name of the Load Balancer in the **KDC Server Host** field and any *one* of the Domain Controllers in **Active Directory Domain Controller Override**. Hadoop daemons will use the Load Balancer for authentication, but Cloudera Manager will use the override for creating accounts.
- If you have multiple Domain Controllers (in case of AD) or MIT KDC servers, only enter the name of any *one* of them in the **KDC Server Host** field. Cloudera Manager will use that server only for creating accounts. If you choose to use Cloudera Manager to manage `krb5.conf`, you can specify the rest of the Domain Controllers using Safety Valve as explained below.
- Make sure the entries for the **Kerberos Encryption Types** field matches what your KDC supports.
- If you are using an Active Directory KDC, you can configure Active Directory account properties such as `objectClass` and `accountExpires` directly from the Cloudera Manager UI. You can also enable Cloudera Manager to delete existing AD accounts so that new ones can be created when Kerberos credentials are being regenerated. See the Cloudera article "Viewing or Regenerating Kerberos Credentials Using Cloudera Manager".

Click **Continue** to proceed.

Manage krb5.conf

Note:

If you are using Red Hat IdM/FreeIPA, by default the `krb5.conf` file contains a line similar to the following:

```
default_ccache_name = KEYRING:persistent:%{uid}
```

CDH does not support the keyring credential cache. Comment out this line on every cluster host by adding a hash mark (`#`) at the beginning, like this:

```
#default_ccache_name = KEYRING:persistent:%{uid}
```

If you configure Cloudera Manager to manage the `krb5.conf` file, you do not need to do anything.

Choose whether Cloudera Manager should deploy and manage the `krb5.conf` on your cluster or not. If left unchecked, you must ensure that the `krb5.conf` is deployed on all hosts in the cluster, including the Cloudera Manager Server's host.

If you check **Manage krb5.conf through Cloudera Manager**, this page will let you configure the properties that will be emitted in it. In particular, the safety valves on this page can be used to configure cross-realm authentication. More information can be found in the following Cloudera help article: "Configuring a Dedicated MIT KDC for Cross-Realm Trust".

Click **Continue** to proceed.

Setup KDC Account

> **Note**: Enter the REALM portion of the principal in upper-case only to conform to Kerberos convention.

> **Note**:
>
> Many enterprises employ policies that require all passwords to be changed after a particular number of days. If you must change the password in Cloudera Manager for the Account Manager, then:
>
> 1. In the Cloudera Manager Admin Console, select **Administration** > **Security**.
> 2. Go to the **Kerberos Credentials** tab and click **Import Kerberos Account Manager Credentials**.
> 3. In the **Import Kerberos Account Manager Credentials** dialog box, enter the username and password for the user that can create principals for CDH cluster in the KDC.

Enter the username and password for the user that can create principals for CDH cluster in the KDC. Cloudera Manager encrypts the username and password into a keytab and uses it as needed to create new principals.

If you are using Red Hat IdM/FreeIPA, enter the IPA admin credentials here. These admin credentials are not stored, and are used only to create a new user and role (named `cmadin-`*`<random_id>`* and `cmadminrole`, respectively) and retrieve its keytab. Cloudera Manager stores this keytab for future Kerberos operations, such as regenerating the credentials of the CDH service accounts.

Click **Continue** to proceed.

The **Command Details** page displays the outcome of the **Import KDC Account Manager Credentials** command. After it successfully completes, click **Continue**.

Configure Kerberos

If you have not already done so, run the provided commands on each cluster host to install the Kerberos libraries.

Then, specify the privileged ports needed by the DataNode Transceiver Protocol and the HTTP Web UI in a secure cluster.

You can configure custom service principals for CDH services. Before you begin making configuration changes, see the Cloudera article "Customizing Kerberos Principals" for some additional configuration changes required and limitations. If you want to use custom service principals, uncheck the box labeled **Use Default Kerberos Principals**, and then specify a custom principal for each service.

Click **Continue** to proceed.

The **Command Details** page displays the outcome of the **Enable Kerberos** command. After it successfully completes, click **Continue**.

Summary

The final page lists the cluster(s) for which Kerberos has been successfully enabled. Click **Finish** to return to the Cloudera Manager Admin Console home page.

Appendix C: Optional Installations and Configurations

This Appendix provides instructions for installing and configuring optional installation features.

Configure Redundancy & High Availability (Optional)

The following sections discuss how to configure redundancy and high availability across the Intersetcluster; this includes all services where available in HDP/Ambari, Elasticsearch.

> **Tip:** We recommend setting up High Availability for only those service which involve **data**.

For any further questions regarding HA and redundancy, contact Micro Focus Interset Support at interset.support@microfocus.com.

Enable HDP Services (e.g. HBase, HDFS, Storm, YARN, ZooKeeper) High Availability

This configuration requires a minimum of three **Master** nodes in your environment. We recommend three (3) nodes - one of which has lower resource requirements as it will only run ZooKeeper and an HDFS Journal Node.

To configure additional Master nodes:

1. On each new **Master** node you wish to add to the cluster, run the following command:

   ```
   bash <(curl -ks http://<mirror_fqdn>/<release>/interset/deploy.sh)
   ```

2. On the **Ambari** node, where the installation was originally executed, go to **/tmp/interset_installer/etc** and edit the **config** file.
3. Modify the MASTER="hostname1" line to read as MASTER="hostname1 hostname2 hostname3" where each hostname refers to a different server in your environment.

   This can be any number of hostnames, and must be space delimited.

4. On the **Ambari** node, run the following commands as the **interset** user:

   ```
   cd /tmp/interset_installer
   ```

   ```
   ./install.sh
   ```

5. From the available installation options, select:

   ```
   Installer option: 1 for initial server config (set up SSH, offline_setup script, etc...)
   ```

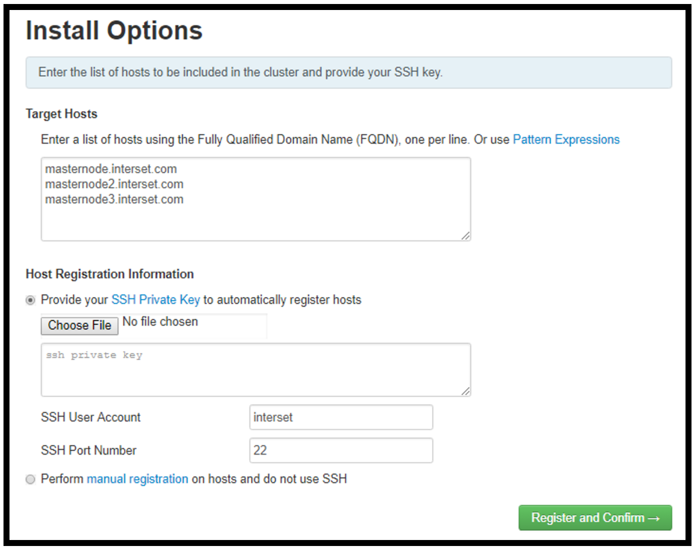
To add the new hosts to Ambari:

1. Open the **Ambari** web interface (**http://<*ambari_fqdn*>:8080**) and log in.
2. At the top menu, click on **Hosts** tab.
3. Click **Actions**, from the dropdown select **Add New Hosts** to open the **Add Host Wizard**.
4. In the **Install Options** page:
   - In the **Target Host** box, enter the hostname(s) of the new nodes being added (one per line).
   - Under **Host Registration Information**, select **Provide your SSH Private Key** to automatically register hosts, and then copy and paste your SSH private key from into the SSH private key text box.

     ✓ **Tip:** To retrieve your RSA key at any time, on your original **Ambari** node, run the following command:

     ```
     cat /home/interset/.ssh/id_rsa
     ```

   - In the **SSH User Account Name** text box, type **interset**.
   - Click **Register and Confirm**.



Install Options

Enter the list of hosts to be included in the cluster and provide your SSH key.

Target Hosts

Enter a list of hosts using the Fully Qualified Domain Name (FQDN), one per line. Or use Pattern Expressions

```
masternode.interset.com
masternode2.interset.com
masternode3.interset.com
```

Host Registration Information

● Provide your SSH Private Key to automatically register hosts

Choose File No file chosen

ssh private key

SSH User Account: interset

SSH Port Number: 22

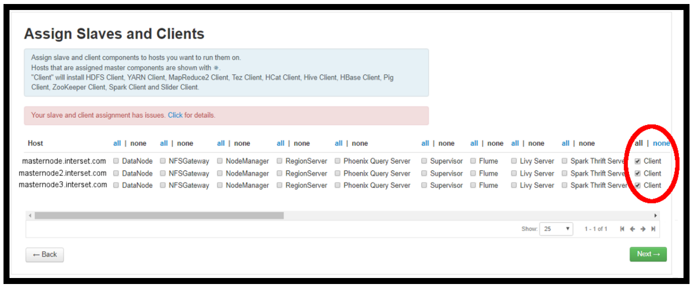
○ Perform manual registration on hosts and do not use SSH

5. Wait for the tests to complete on the **Confirm Hosts** page, and then click **Next**.

   Note: Warnings about Snappy packages can be ignored, as they are expected.

   Important: If any of these nodes will be used as **Compute** nodes, select **Data Node**, **Node Manager**, and **Region Server** in addition to **Client**.

6. On the **Assign Slaves and Clients** page, select only **Client**, and then click **Next**.



Assign Slaves and Clients

Assign slave and client components to hosts you want to run them on.
Hosts that are assigned master components are shown with ✱.
"Client" will install HDFS Client, YARN Client, MapReduce2 Client, Tez Client, HCat Client, Hive Client, HBase Client, Pig Client, ZooKeeper Client, Spark Client and Slider Client.

Your slave and client assignment has issues. Click for details.

Host	all | none	all | none	all | none	all | none	all | none	all | none	all | none	all | none	all | none	all | none
masternode.interset.com	□ DataNode	□ NFSGateway	□ NodeManager	□ RegionServer	□ Phoenix Query Server	□ Supervisor	□ Flume	□ Livy Server	□ Spark Thrift Serve	☑ Client
masternode2.interset.com	□ DataNode	□ NFSGateway	□ NodeManager	□ RegionServer	□ Phoenix Query Server	□ Supervisor	□ Flume	□ Livy Server	□ Spark Thrift Serve	☑ Client
masternode3.interset.com	□ DataNode	□ NFSGateway	□ NodeManager	□ RegionServer	□ Phoenix Query Server	□ Supervisor	□ Flume	□ Livy Server	□ Spark Thrift Server	☑ Client

7. On the **Configurations** page, click **Next**.
8. Review the information on the **Review** page, and then click **Deploy**.

   The deployment will take about 15-45 minutes to complete.
9. When the installation is complete, click **Next**.
10. On the **Summary** page, click **Complete**.
11. When the deployment wizard is finished, click the **Services** tab, and then click **ZooKeeper**.
12. On the ZooKeeper settings tab, click **Service Actions**, and then select **Add ZooKeeper Server**.
13. From the drop-down list, select the new Master node(s), and then click **Add**.
14. Once ZooKeeper is installed on the new Master node(s), the ZooKeeper services will need to be restarted:
    - Click the **Services** tab, and then click **ZooKeeper**.
    - On the ZooKeeper settings tab, click **Service Actions**, and then select **Restart All**.

    After the ZooKeeper server restarts, any affected services will also need to be restarted.

    Services needing to be restarted will be flagged with a yellow Restart Required icon.
15. Navigate to **/opt/interset/rules/conf**, open the **rules.conf** file, and edit the file as follows:
    - Change the following line to include the new Master node(s):

      ```
      PhoenixZk = <first_master_node_fqdn>,<second_master_node_fqdn>,<third_master_node_fqdn>:2181:/hbase-unsecure
      ```

      ✓ **Tip:** Only the last Master node in the list requires the port and HBase information.
    - Kill and re-deploy Workflow topology, using the following commands:

      ```
      /opt/interset/rules/bin/workflow.sh --kill /opt/interset/rules/conf/rules.conf
      ```

      ```
      /opt/interset/rules/bin/workflow.sh --deploy /opt/interset/rules/conf/rules.conf
      ```
16. Update any existing Flume configurations to include the new ZooKeeper instances.
17. Update the Analytics **interset.conf** configuration file (**/opt/interset/analytics/conf/interset.conf**) to include the additional ZooKeeper hosts.

    Hosts are added to the **zkPhoenix** variable. The syntax example for doing this is as follows:

```
zkPhoenix = master-1.interset.com,master-2.interset.com,master-3.interset.com:2181:/hbase-unsecure
```

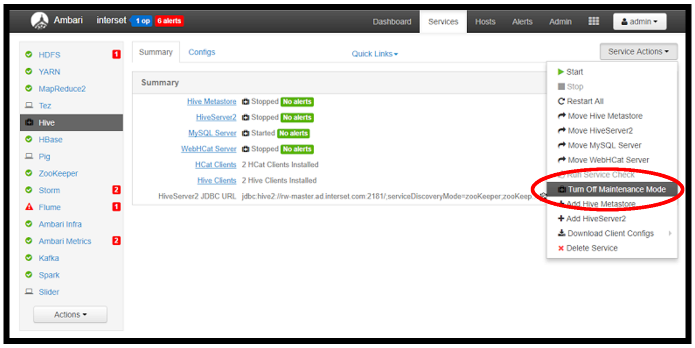
After the new node(s) is added and you have a total of three (3) Master nodes, with each one running a ZooKeeper server instance, proceed to the next section, Enable HDFS High Availability (HA) to enable HA.

Enable HDFS High Availability (HA)

This section applies only if you have chosen to dedicate three servers as Master nodes. The steps below will ensure that your Interset system experiences minimal down time in the event a Master node is lost. Should a Master node go down, some non-essential components may still need to be recovered; however, the critical server components should continue to function normally.

The Ambari wizard will guide you through the process of enabling HDFS HA. You will be required to SSH into the **Master** nodes to execute commands manually. All commands are provided in the wizard, and indicate the machines on which the commands are to be executed. To enable HDFS HA, do the following:

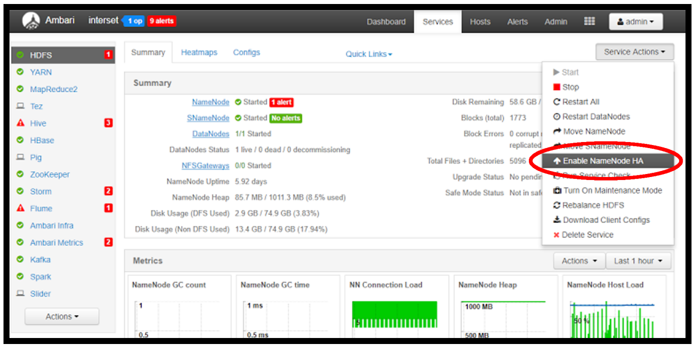
1. In a Web browser, navigate to the **Ambari** web interface (**http://<*ambari_fqdn*>:8080**) and log in.
2. On the sidebar, select the **Hive** service.
3. On the **Hive** settings tab, click **Service Actions**.
4. From the dropdown menu, select **Turn Off Maintenance Mode**.



Summary

Hive Metastore Stopped No alerts
HiveServer2 Stopped No alerts
MySQL Server Started No alerts
WebHCat Server Stopped No alerts
HCat Clients 2 HCat Clients Installed
Hive Clients 2 Hive Clients Installed
HiveServer2 JDBC URL jdbc:hive2://rw-master.ad.interset.com:2181/;serviceDiscoveryMode=zooKeeper;zooKeep...

5. On the sidebar, select the **HDFS** service.
6. On the **HDFS** settings tab, click **Service Actions**.
7. From the dropdown menu, select **Enable NameNode HA**.



Summary

NameNode Started 1 alert
SNameNode Started No alerts
DataNodes 1/1 Started
DataNodes Status 1 live / 0 dead / 0 decommissioning
NFSGateways 0/0 Started
NameNode Uptime 5.92 days
NameNode Heap 85.7 MB / 1011.3 MB (8.5% used)
Disk Usage (DFS Used) 2.9 GB / 74.9 GB (3.83%)
Disk Usage (Non DFS Used) 13.4 GB / 74.9 GB (17.94%)
Disk Remaining 58.6 GB /
Blocks (total) 1773
Block Errors 0 corrupt replicated
Total Files + Directories 5096
Upgrade Status No pendi
Safe Mode Status Not in saf

8. Follow the instructions in the HDFS HA setup wizard, ensuring that you shut down **HBase** before proceeding.

9. When selecting hosts for the additional HA components, spread the three **JournalNodes** over each of the three **Master nodes**.

   Note: If you are only using two Master nodes, you may need another node to provide double-duty on this. The journal node is relatively light weight, so a Compute node is a reasonable option in that scenario.

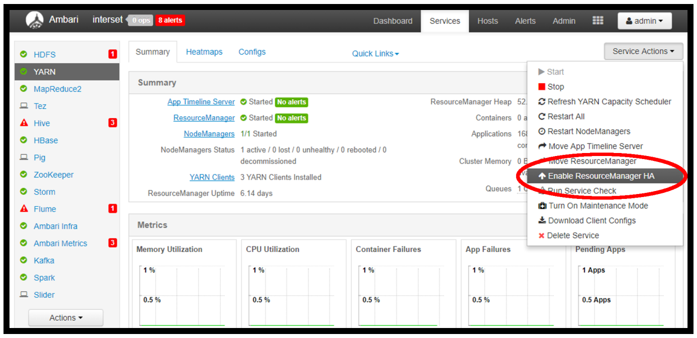
10. Select an alternate **Master** node for the additional NameNode.

11. When the wizard is complete, restart **HBase**, and then **exit maintenance mode**.

    Note: When you finalize the HA setup, the last step of the wizard, **Start All Services**, may fail. If this occurs, wait 60 seconds and then click **Retry**. The services should start after one or two attempts.

Enable YARN High Availability

1. In a Web browser, navigate to the **Ambari** Web interface (**http://<*ambari_fqdn*>:8080**), and log in.
2. On the sidebar, select the **YARN** service.
3. On the **YARN** settings tab, click **Service Actions**. From the dropdown menu, select **Enable ResourceManager HA**.



App Timeline Server: Started No alerts
ResourceManager: Started No alerts
NodeManagers: 1/1 Started
NodeManagers Status: 1 active / 0 lost / 0 unhealthy / 0 rebooted / 0 decommissioned
YARN Clients: 3 YARN Clients Installed
ResourceManager Uptime: 6.14 days
ResourceManager Heap: 52
Containers: 0 a
Applications: 16
Cluster Memory: 0 E
Queues: 1

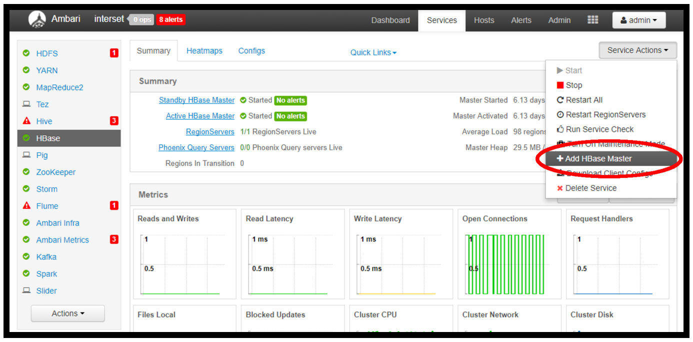
4. Select one of the Master nodes for the additional Resource Manager, and then click **Next**.
5. Click **Next** until the wizard is finished.

   Note: When finalizing the HA setup, the last step of the wizard, Start All Services, may fail. Should this occur, wait 60 seconds and then click Retry. The services should start after one or two attempts.

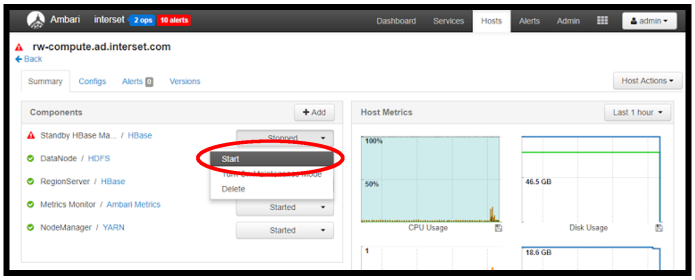
Enable HBase High Availability

1. In a Web browser, navigate to the **Ambari** web interface (**http://<*ambari_fqdn*>:8080**), and log in.
2. On the sidebar, select the **HBase** service.
3. On the **HBase** settings tab, click **Service Actions**.
4. From the dropdown menu, select **Add HBase Master**.

   A dialog box opens, providing a list of hosts available to run the new **HBase Master**.

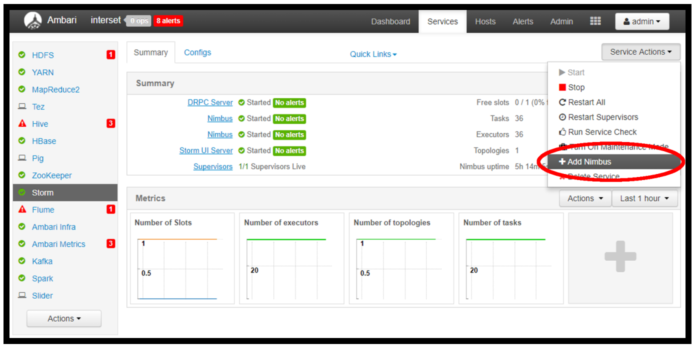


5. Choose an available host to run the new **HBase Master**, and then click **Confirm Add**. Ambari will install the new service on the node selected.
6. When the installation is complete, start the service by completing the following commands:
   a. In the **Summary** section, select the node on which the new **HBase Master** was installed.
   b. From the **HBase Master** dropdown list, select **Start**.



Enable Storm High Availability (HA)

1. In a Web browser, navigate to the **Ambari** web interface (**http://<*ambari_fqdn*>:8080**, and log in.
2. On the sidebar, select the **Storm** service.
3. On the **Storm** settings tab, click **Service Actions**.
4. From the dropdown menu, select **Add Nimbus**.

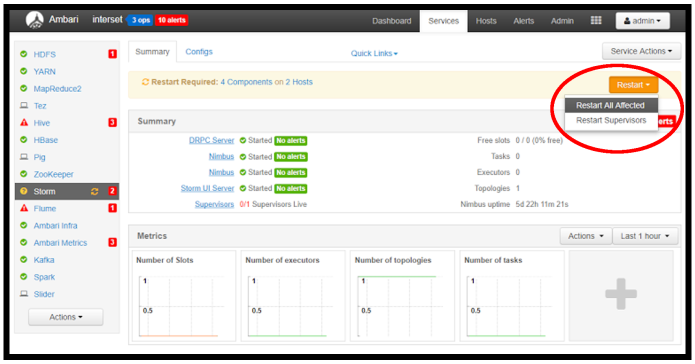


5. A dialog box opens, providing a list of hosts available to run the new Nimbus service.
6. Select an available host to run the new Nimbus service, and then click **Confirm Add**.

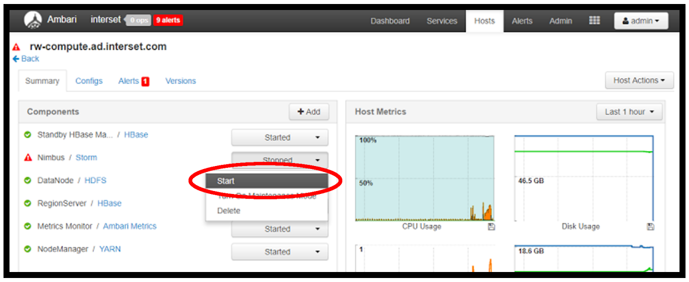
   Ambari will install the new service on the node selected.

7. When the installation is complete, start the service by completing the following steps:
   a. On the **Storm** settings tab, click **Restart**, and then click **Restart All Affected**.

      This restarts the **Storm** services.



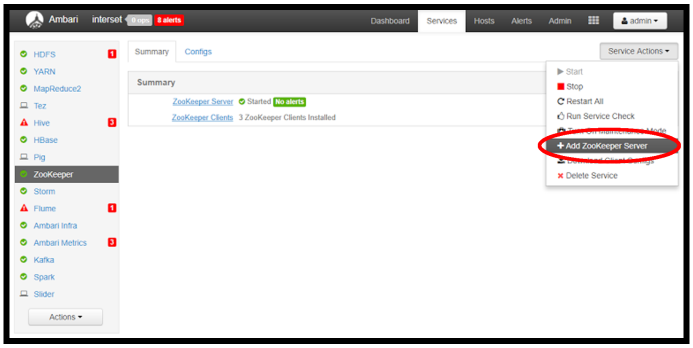
   b. In the **Summary** section, click the node on which the new Nimbus service was installed.
   c. Click the dropdown list, and then select **Start**.



Enable ZooKeeper High Availability

ZooKeeper operates in a quorum to determine the validity of information. With this in mind, an odd number of ZooKeeper services is always recommended in a cluster (e.g. 1,3,5). For Interset purposes, three (3) is sufficient. To enable ZooKeeper High Availability:

1. In a Web browser, navigate to the **Ambari** web interface (**http://<*ambari_fqdn*>:8080**), and log in.
2. On the sidebar, select the **ZooKeeper** service.
3. Click **Service Actions** and then, from the dropdown menu, select **Add ZooKeeper Server**.



4. Select the target host on the **Confirmation** page, and click **Confirm Add**.
5. Repeat the above steps for each additional server on which to add ZooKeeper.

Note: The new ZooKeeper servers will not start automatically, and may require the restart of other services. A full restart of the HDP services, when plausible, is recommended.

Enable Elasticsearch High Availability

Adding additional Elasticsearch nodes can be done using the Interset Installer:

1. On each new **Stream** node you wish to add to the cluster, run the following command:

   ```
   bash <(curl -ks http://<mirror_fqdn>/<release>/interset/deploy.sh)
   ```

2. On the **Ambari node**, where the install was originally executed, go to **/tmp/interset_installer/etc** and edit the **config** file.
3. Modify the SEARCH="hostname1" line to read as SEARCH="hostname1 hostname2 hostname3" where each hostname refers to a different server in your environment.

   This can be any number of hostnames, and must be space delimited.

4. On the **Ambari** node, run the following commands as the **interset** user:

   ```
   cd /tmp/interset_installer
   ```

   ```
   ./install.sh
   ```

5. From the available installation options, select:

   ```
   Installer option: 1 for initial server config (set up SSH, offline_setup script, etc...)
   ```

6. When the **option 1** install is complete, from the available installation options, select:

   ```
   Installer option: 6 for Search node(s) installation
   ```

> **Note:** It may take a considerable amount of time for Elasticsearch to rebalance the cluster with the addition of the new nodes, and a performance impact should be expected while this is taking place.

Configure DXL (Optional)

To configure DXL messages as Workflow outcomes (notifications), you must set up DXL authentication. This involves

- generating a signed client certificate
- uploading that certificate to McAfee ePolicy Orchestrator (ePO)
- exporting broker certificates from ePO
- generating the Java Keystore
- exporting the broker list from ePO

- configuring the Workflow engine
- setting up the McAfee ESM parser

Note: The McAfee ePO build used for this section was ePO Build: ePolicy Orchestrator 5.3.1 (Build 188).

Important: When DXL messages are configured as Workflow outcomes, the Workflow **rules.conf** configuration file is edited to enable DXL output and include connection information. On startup, the Workflow engine attempts to establish the DXL broker(s) connection. If the Workflow engine is unable to connect to the DXL broker (s), the Workflow topology in Storm will fail to start, preventing any Workflows from running. To avoid this situation, ensure that the DXL connection information you provide in the **rules.conf** file is valid.

Generate a Signed Client Certificate

1. Using openssl, create the certificate authority (CA) and fill out the prompts using the following command:

   ```
   openssl req -new -x509 -days 3650 -extensions v3_ca -keyout ca.key -out ca.crt
   ```

2. Run the following command to generate the private key for the DXL client:

   ```
   openssl genrsa -out client.key 2048
   ```

3. Create CSR for the client key:

   ```
   openssl req -out client.csr -key client.key -new
   ```

   Tip: The value entered for **Organizational Unit Name** must not be the same as the one entered when creating the CA.

4. Sign the certificate request, using the following command:

   ```
   openssl x509 -req -in client.csr -CA ca.crt -CAkey ca.key -CAcreateserial -out client.crt -days 3650
   ```

   Tip: When prompted for a PEM password, use the same one you used when creating the CA.

Upload the CA to McAfee ePolicy Orchestrator (ePO)

1. In the ePO Web user interface, click **Server Settings**, **DXL Certificates**, and then click **Edit**.
2. In the **Client Certificates** section, select **Import**.
3. Select the file **ca.crt** that was created in the last section.
4. Click **OK**, and then click **Save**.

Export the Broker Certificates from McAfee ePolicy Orchestrator (ePO)

1. In the ePO Web user interface, click **Server Settings**, **DXL Certificates**, and then click **Edit**.
2. In the **Broker Certificates** section, click **Export All**.

   The exported file, **brokercerts.crt**, will be saved locally.

Generate the Java Keystore

Because private keys created with openssl are not easily imported into java keystore files, Micro Focus Interset Support at interset.support@microfocus.com recommends using the PKCS12 format as a work-around.

1. Run the following command to create a PKCS12 bundle using the client's private key and the CA certificate:

   ```
   openssl pkcs12 -export -in client.crt -inkey client.key -chain -CAfile ca.crt -name dxlClientKey -out
   dxlClient.p12
   ```

2. Convert the PKCS12 bundle into a java keystore file, using the following command:

   ```
   keytool -importkeystore -deststorepass <YOUR_KEYSTORE_PASS> -destkeystore dxlClient.jks -srckeystore
   dxlClient.p12 -srcstoretype PKCS12
   ```

3. Open the **brokercerts.crt** file in a text editor, and copy the last certificate in the file.

   Tip: Copy from **---BEGIN CERTIFICATE---** to **---END CERTIFICATE---** inclusively.

4. Paste the copied string into a new text file, and save the file as **brokerca.crt**
5. Import the broker certificate file into the keystore using the following command:

   ```
   keytool -import -trustcacerts -alias brokerCertCa -file brokerca.crt -keystore dxlClient.jks
   ```

   Note: The keystore file must be copied onto the node where **Storm** is installed.

   When prompted whether to trust this certificate enter **yes**.

   ```
   Trust this certificate? [no]: yes
   ```

Important: The **dxlClient.jks** java keystore file is a globally-writable file. If your organization requires that third-party application files not be writable, you must manually change the **dxlClient.jks** file permissions.

Export Broker List from EPO

1. Navigate to server settings -> **DXL Certificates**
2. Select **"Edit"**

3. In the **Broker List** section, select **"Export All"**
4. The exported file **"brokerlist.properties**" will be saved locally.

Configure the Workflow Engine

1. Open the **/opt/interset/rules/conf/rules.conf** file, and edit the following parameters:

   ```
   DxlOutputEnabled = true
   DxlEnableTLS = true
   DxlBrokerList = <contents of brokerlist.properties>
   DxlKeyStoreFile = <path/to/dxlClient.jks>
   DxlKeyStoreAlias = brokerCertCa
   DxlKeyStorePass = <your keystore password>
   ```

   > **Important:** The DXL Keystore (dxlClient.jks file) must be present on all of the nodes in the Interset cluster where the Storm Supervisor process runs. In a typical cluster configuration, Storm is located on the Compute node(s).

Set Up the McAfee ESM Parser

To send Workflow notifications to McAfee ESM, you will need to configure the ESM Parser.

> **Tip:** If you plan to simply send Workflow notifications to other DXL recipients, you can ignore this section.

1. In the ESM console, select **Policy Editor**.
2. Under **Rule Type**, select **Receiver**, **Advanced Syslog Parser**.
3. Select **New**, **Advanced Syslog Parser Rule**.
4. On the **General** tab, enter a name for the Workflow.
5. Because the DXL output from Workflow is in JSON format, on the **Parsing** tab, enter **{** for the context string that the Workflow will try to match.
6. On the **Field Assignment** tab, enter the following suggested mappings:

   Tip: Use the **+** button to add any ESM parameters that are not present by default.

   ```
   Application -> "Interset"
   External_SubEventID -> tenantId
   First Time -> timestampEpoch
   Message Text -> dxlMessage
   Rule Name -> ruleName
   sourceUser -> sourceUser
   UUID -> ruleUUID
   ```

   Note: After saving the configuration, Application field will be retitled as AppId.
7. On the **Mapping** tab, add an entry for a custom format for the **First Time** field.

   The time format string should be **%s**
8. Click **Finish**, and then roll the policy out to all applicable devices.

Configure Splunk (Optional)

To configure Splunk notifications as Workflow outcomes, you must configure the Workflow engine to enable this feature. This involves:

- editing the **rules.conf** configuration file
- restarting the Workflow engine
- configuring the Splunk KV_MODE to enable automatic extraction fields from the JSON data

Tips:

- Interset recommends that you configure a Splunk user specifically for Workflow access.

  Alternatively, an existing Splunk username and password with permission to write events can be used.
- When configuring the Workflow to send notifications to Splunk, you have the option to create a Splunk index and message. If no index is configured, a new index for Workflow notifications will be created automatically in Splunk.

Edit the rules.conf Configuration File

1. On the **Master node** where Analytics is installed, as the **interset** user, open the **/opt/interset/rules/conf/rules.conf** file, remove the hash (#) symbols to enable the following parameters, and update the values with the appropriate information for your Splunk instance:

```
cd /opt/interset/rules/conf
#SplunkHost = <SPLUNK_HOST>
#SplunkPort = <SPLUNK_PORT>
#SplunkUsername = <USERNAME>
#SplunkPassword = <PASSWORD>
#SplunkSourceAppName = interset_workflow
```

2. Save the **rules.conf** file.

Restart the Workflow Engine

1. On the **Master node** where Analytics is installed, as the **interset** user, run the following commands to stop and then restart the Workflow engine:

```
/opt/interset/rules/bin/workflow.sh --kill /opt/interset/rules/conf/rules.conf
```

```
/opt/interset/rules/bin/workflow.sh --deploy /opt/interset/rules/conf/rules.conf
```

   Workflow is now able to export notifications to Splunk.

Configure the Splunk KV_MODE

Configuring the KV_MODE in Splunk enables the automatic extraction of JSON data fields. This automatic extraction will facilitate subsequent data searches.

1. Log in to Splunk as an administrator.
2. In the top right corner of the page, click Settings, and then select Source types.
3. Select the _json source type.
4. Expand the Advanced section, and then change the KV_MODE setting from none to json.

Configure Phantom (Optional)

If you wish to configure Phantom Workflow outcomes (notifications), you must set up Phantom integration. These instructions assume that Phantom is already configured and functional in your production environment.

Integrating Phantom with Interset for Workflow outcomes involves:

- configuring a Phantom API
- updating the Interset **rules.conf** configuration file
- validating the Phantom integration

Configure the Phantom API

1. Configure a Phantom API for Interset as a custom data source. To enable this API, please see the Phantom documentation:

   https://my.phantom.us/2.1/docs/rest/custom_script
2. After you've created the custom API, do the following:
   - on the **Asset Settings** tab, click **Edit** to ensure the API includes a custom python handler
   - record the following information from the Phantom API:
     - on the **Asset Info** tab, in the **Asset Name** box, the custom **API name**
     - on the **Asset Settings** tab, in the **POST incoming for REST Data Source to this location** box, the **URL**
     - on the **Ingest Settings** tab, in the **Authentication Configuration for REST API** box, the **ph-auth-token**

   You will require this information when editing the Workflow **rules.conf** configuration file.

Edit the rules.conf Configuration File

1. On the **Master node** where Analytics is installed, as the **interset** user, edit the **/opt/interset/rules/conf/rules.conf** file, remove the hash (#) symbols to enable the following parameters, and update the values with the appropriate information for your Phantom instance:

```
#PhantomPostUrl = <URL>
#PhantomAuthSource = <API_name>
#PhantomAuthToken = <ph_auth_token>
#PhantomContainerExpirationInDays = <number>
```

✓ **Tip:** If you leave the **PhantomContainerExpirationInDays** parameter commented out, the value will be set to one (1) day.

2. Save the **rules.conf** file.

Validate the Phantom Integration

1. Run the following script to verify the Phantom settings:

   ```
   /opt/interset/rules/bin/workflow.sh  --validate
   ```

   The validation script outputs the configured Phantom settings, and creates a new container and artifact in Phantom.

2. Do the following to view the new Phantom container and artifact:

   - Log in to the Phantom user interface.
   - Click **Home**, then **Sources**, and finally **Data Source**.
   - Select the newly-created container, and view the artifact.

   Workflow is now able to export notifications to Phantom.

Appendix D: Add New Nodes to an Existing Cloudera Cluster

This section provides information about adding a new node to an existing CDH cluster. For further questions about node configuration and sizing, contact Micro Focus Interset Support (interset.support@microfocus.com).

1. Ensure that the new node is connected on the same network as the existing cluster and communicates with cluster nodes without any network interruption.
2. Navigate to the **/opt/interset/installer/etc** directory, and then locate and open the **config** file for editing. This file has a line for each role, or node in the Interset cluster (for example, Monitoring or Stream).
3. For the role that has the new node, enter the FQDN of the server(s) you want to add. If the role has more than one physical node specified, enter multiple FQDNs separated by a space.
4. On the new node, run the following command as a user with sudo access, substituting <mirror_fqdn> with the FQDN for your installation repository.

   ```
   bash <(curl -ks http://<mirror_fqdn>/5.9.2/interset/deploy.sh)
   ```

   > **Tip:** In an online installation, you should also include your access credentials. For example,
   >
   > ```
   > username:password@<mirror_fqdn>/5.9.2/interset/deploy.sh
   > ```

   This script copies the Interset installer package, untars contents to the **/opt/interset/installer** directory, creates the **interset** user, and gives ownership of the **interset_installer** folder to **interset** user.

   > **Tip:** The default password for the **interset** user is **interset**. You will be prompted to change this password at this stage.

5. In a command console on the Monitoring node, run the following commands to navigate to the **/opt/interset/installer** directory and launch the Interset installation menu.

   ```
   cd /opt/interset/installer
   ```

   ```
   ./install.sh
   ```

6. From the available installation options, type **1** to select installation **[1] Initial server config (set up SSH, offline_setup script, database ...)**, and then press **Enter**.

   ```
   Please select an installation option from the list above: 1
   ```

   Enter the required information, as prompted by the installer. This process is same as Step 1 of "Perform the Base Configuration of All Nodes" on page 31.

7. From the available installation options, type **4** to select installation **[4] Generate TLS Certificate**, and then press **Enter**.

   ```
   Please select an installation option from the list above: 4
   ```

   This step will create new certificates, as done previously during the initial cluster set up.

8. From the available installation options, type **8** to select installation **[8] Schema Registry Setup**, and then press Enter.

   ```
   Please select an installation option from the list above: 8
   ```

   Follow the steps for Schema Registry set up.

9. From the available installation options, type **11** to select installation **[11] Search node(s) installation**, and then press **Enter**.

   ```
   Please select an installation option from the list above: 11
   ```

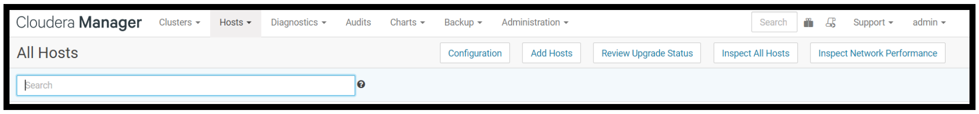
   When the Search node installation is complete, the expected result is:

   ```
   Search Node Configuration Complete!
   ```

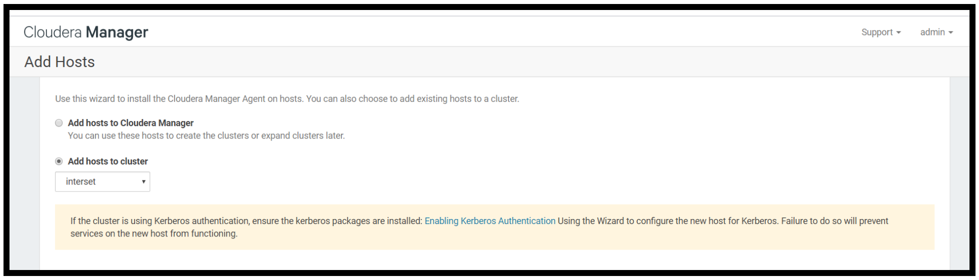
10. After installating specific services, exit the installer.

11. In a Web browser, navigate to https://<monitoring_server_fqdn>:7183. Log in to the Cloudera Manager Admin Console.

12. Click **Hosts** > **All Hosts** > **Add Hosts**.



13. Click **Add hosts to cluster**.



Cloudera Manager — Support — admin

Add Hosts

Use this wizard to install the Cloudera Manager Agent on hosts. You can also choose to add existing hosts to a cluster.

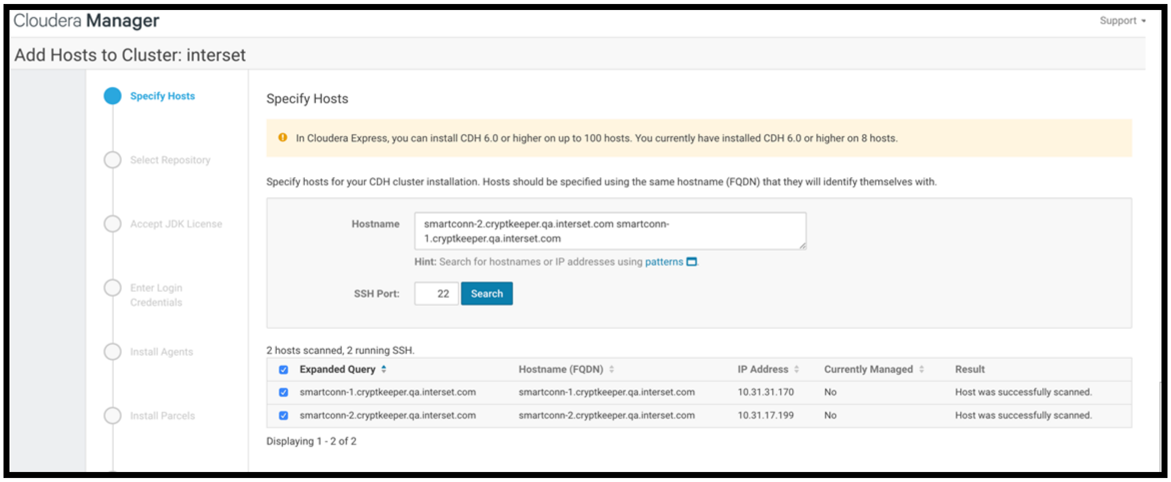
○ **Add hosts to Cloudera Manager**
You can use these hosts to create the clusters or expand clusters later.

● **Add hosts to cluster**

interset ▾

If the cluster is using Kerberos authentication, ensure the kerberos packages are installed: Enabling Kerberos Authentication Using the Wizard to configure the new host for Kerberos. Failure to do so will prevent services on the new host from functioning.

14. Specify the name of host(s) you want to add to the cluster.



Add Hosts to Cluster: interset

Specify Hosts

In Cloudera Express, you can install CDH 6.0 or higher on up to 100 hosts. You currently have installed CDH 6.0 or higher on 8 hosts.

Specify hosts for your CDH cluster installation. Hosts should be specified using the same hostname (FQDN) that they will identify themselves with.

Hostname: smartconn-2.cryptkeeper.qa.interset.com smartconn-1.cryptkeeper.qa.interset.com

Hint: Search for hostnames or IP addresses using patterns

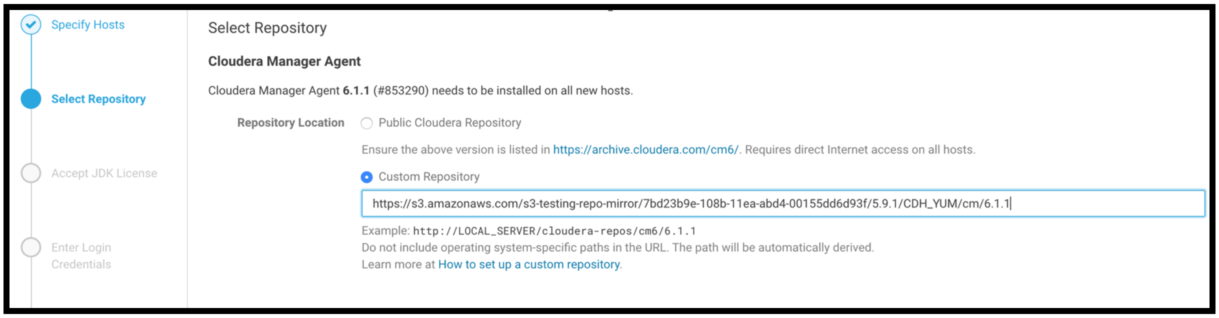
SSH Port: 22

2 hosts scanned, 2 running SSH.

☑	Expanded Query	Hostname (FQDN)	IP Address	Currently Managed	Result
☑	smartconn-1.cryptkeeper.qa.interset.com	smartconn-1.cryptkeeper.qa.interset.com	10.31.31.170	No	Host was successfully scanned.
☑	smartconn-2.cryptkeeper.qa.interset.com	smartconn-2.cryptkeeper.qa.interset.com	10.31.17.199	No	Host was successfully scanned.

Displaying 1 - 2 of 2

15. Add the Cloudera repository location. For example.



Select Repository

Cloudera Manager Agent

Cloudera Manager Agent **6.1.1** (#853290) needs to be installed on all new hosts.

Repository Location: ○ Public Cloudera Repository

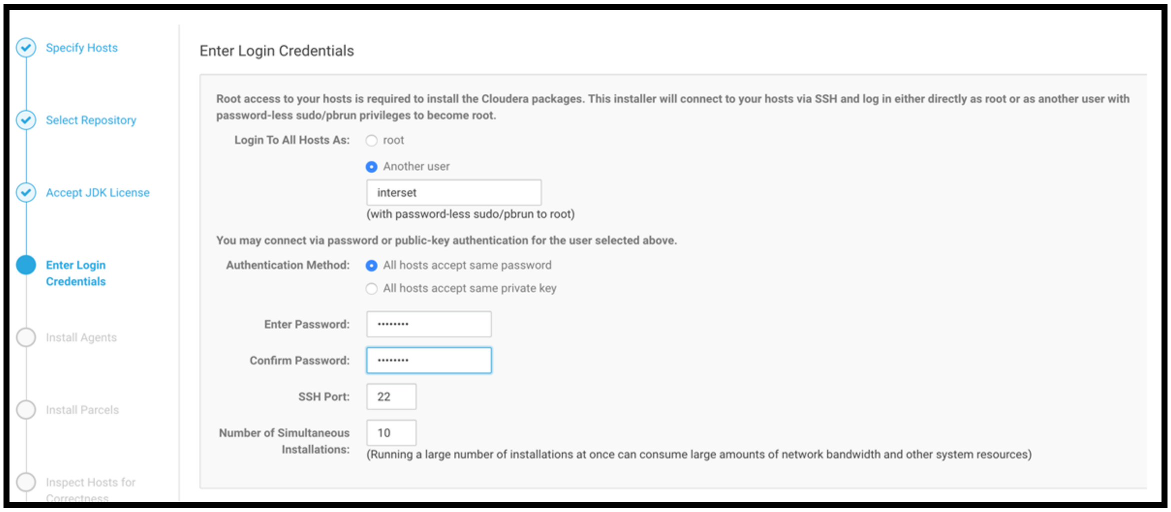
Ensure the above version is listed in https://archive.cloudera.com/cm6/. Requires direct Internet access on all hosts.

● Custom Repository

https://s3.amazonaws.com/s3-testing-repo-mirror/7bd23b9e-108b-11ea-abd4-00155dd6d93f/5.9.1/CDH_YUM/cm/6.1.1

Example: `http://LOCAL_SERVER/cloudera-repos/cm6/6.1.1`
Do not include operating system-specific paths in the URL. The path will be automatically derived.
Learn more at How to set up a custom repository.

16. Specify the log in details for the **interset user**.



Enter Login Credentials

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
interset
(with password-less sudo/pbrun to root)

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

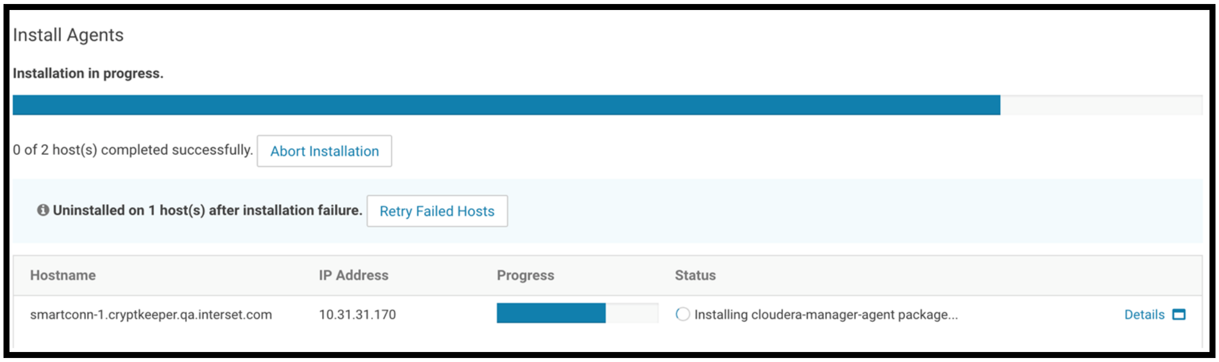
Enter Password: ••••••••

Confirm Password: ••••••••

SSH Port: 22

Number of Simultaneous Installations: 10
(Running a large number of installations at once can consume large amounts of network bandwidth and other system resources)

The installation for the host begins and you can see the installation progress at the top.



Install Agents

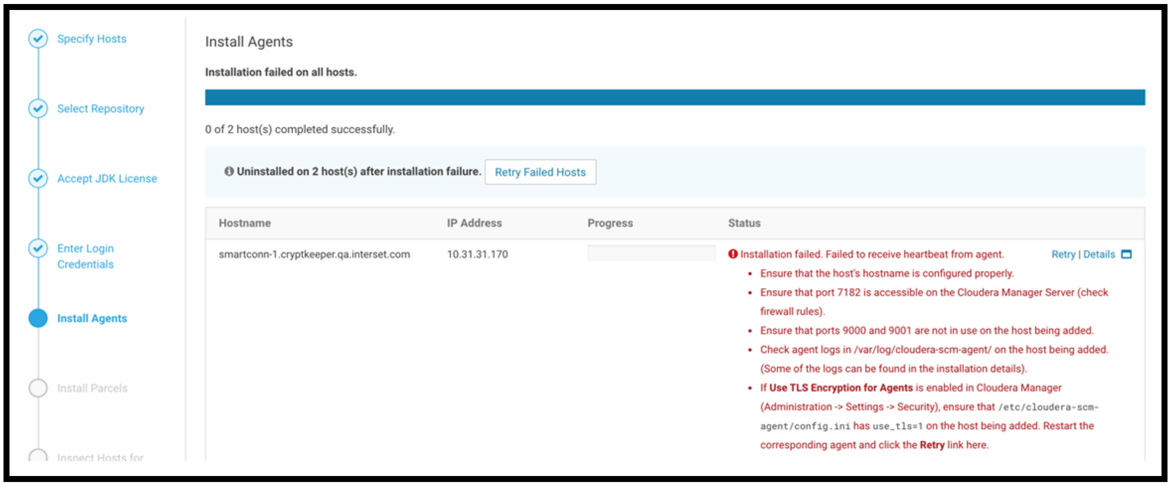
Installation in progress.

0 of 2 host(s) completed successfully.

Uninstalled on 1 host(s) after installation failure.

Hostname	IP Address	Progress	Status	
smartconn-1.cryptkeeper.qa.interset.com	10.31.31.170		Installing cloudera-manager-agent package...	Details

It is expected to see a failure message at the end of the agent installation.



Install Agents

Installation failed on all hosts.

0 of 2 host(s) completed successfully.

Uninstalled on 2 host(s) after installation failure.

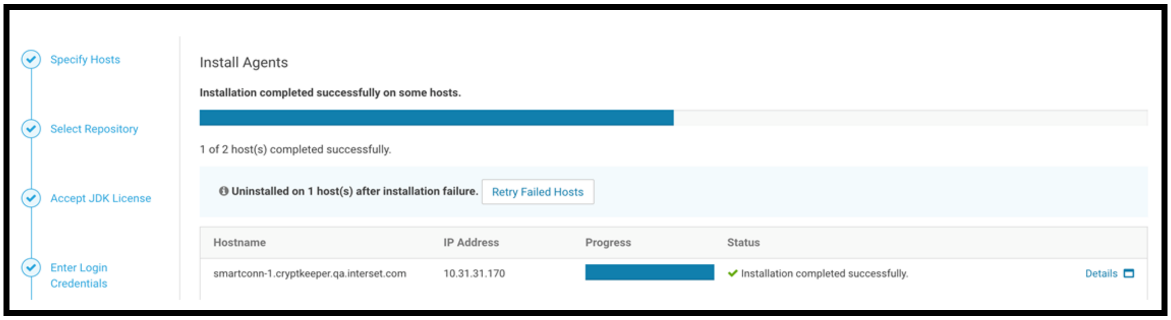
Hostname	IP Address	Progress	Status
smartconn-1.cryptkeeper.qa.interset.com	10.31.31.170		Installation failed. Failed to receive heartbeat from agent. Retry | Details • Ensure that the host's hostname is configured properly. • Ensure that port 7182 is accessible on the Cloudera Manager Server (check firewall rules). • Ensure that ports 9000 and 9001 are not in use on the host being added. • Check agent logs in /var/log/cloudera-scm-agent/ on the host being added. (Some of the logs can be found in the installation details). • If **Use TLS Encryption for Agents** is enabled in Cloudera Manager (Administration -> Settings -> Security), ensure that `/etc/cloudera-scm-agent/config.ini` has `use_tls=1` on the host being added. Restart the corresponding agent and click the **Retry** link here.

17. To resolve the error, follow the steps below:
    1. On the node where Cloudera Manager is installed, navigate to the **/cloudera-scm-agent/-config.ini** file.
    2. Search for the **use_tls** parameter within the file and set its value as 1. **user_tls =1**
    3. Restart the Cloudera agent using the following command.

       ```
       sudo service cloudera-scm-agent restart
       ```

    4. Retry the failed hosts.

    The agents will now install successfully.



Install Agents

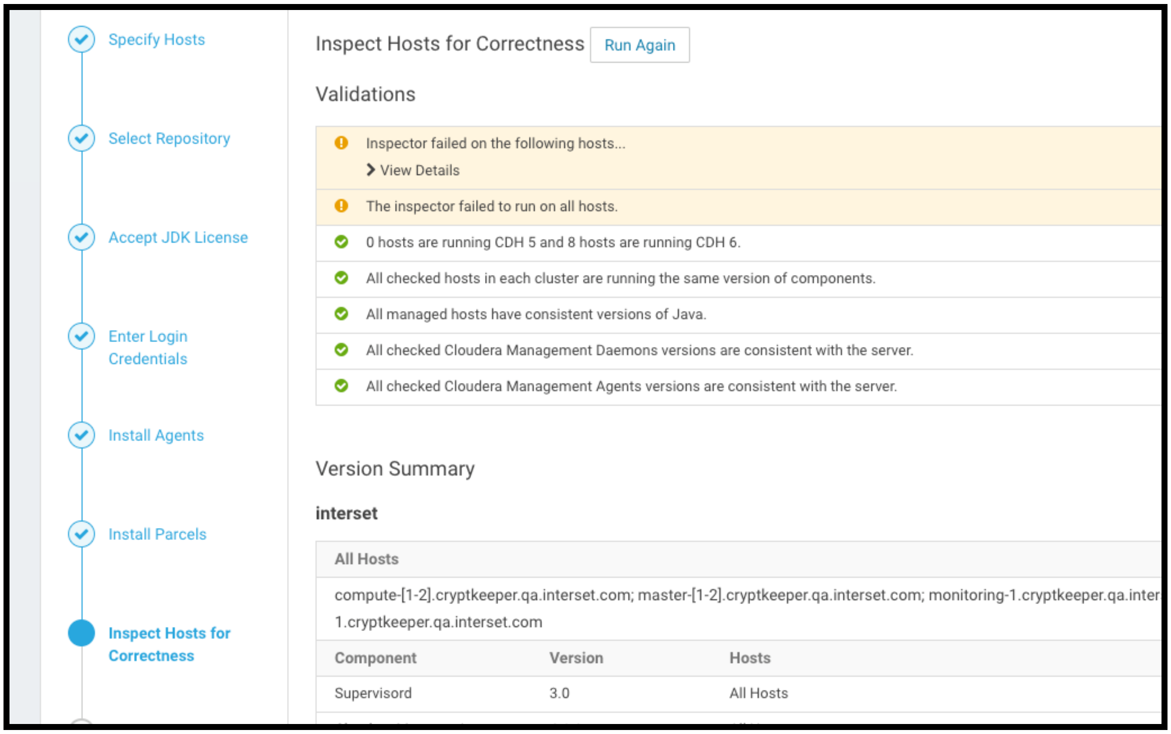
Installation completed successfully on some hosts.

1 of 2 host(s) completed successfully.

Uninstalled on 1 host(s) after installation failure.

Hostname	IP Address	Progress	Status	
smartconn-1.cryptkeeper.qa.interset.com	10.31.31.170		✓ Installation completed successfully.	Details

18. Once the installation is complete, check the correctness and validate any failed components.



Inspect Hosts for Correctness

Validations

- Inspector failed on the following hosts...
- The inspector failed to run on all hosts.
- 0 hosts are running CDH 5 and 8 hosts are running CDH 6.
- All checked hosts in each cluster are running the same version of components.
- All managed hosts have consistent versions of Java.
- All checked Cloudera Management Daemons versions are consistent with the server.
- All checked Cloudera Management Agents versions are consistent with the server.

Version Summary

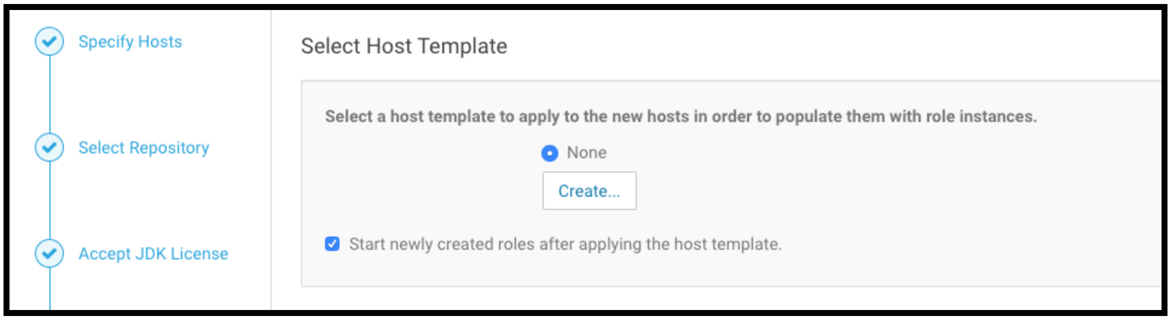
interset

All Hosts

compute-[1-2].cryptkeeper.qa.interset.com; master-[1-2].cryptkeeper.qa.interset.com; monitoring-1.cryptkeeper.qa.inter
1.cryptkeeper.qa.interset.com

Component	Version	Hosts
Supervisord	3.0	All Hosts

19. Select a host template for the new node.



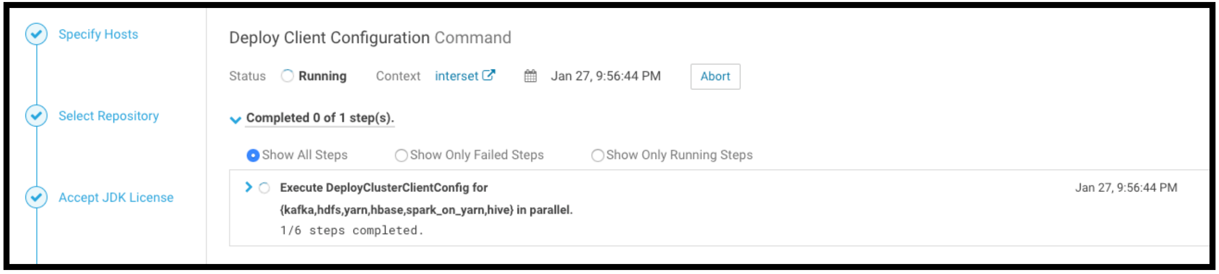
Select Host Template

Select a host template to apply to the new hosts in order to populate them with role instances.

● None

☑ Start newly created roles after applying the host template.

20. Deploy the client configuration command for the new node. This will start setting up various components such as Kafka, HDFS, and HBase on the new node.



Deploy Client Configuration Command

Status Running Context interset Jan 27, 9:56:44 PM

Completed 0 of 1 step(s).

● Show All Steps ○ Show Only Failed Steps ○ Show Only Running Steps

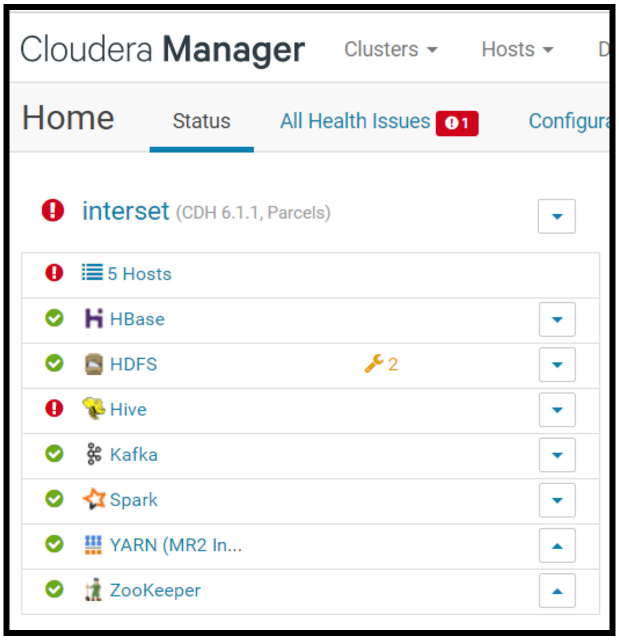
Execute DeployClusterClientConfig for {kafka,hdfs,yarn,hbase,spark_on_yarn,hive} in parallel. Jan 27, 9:56:44 PM

```
1/6 steps completed.
```

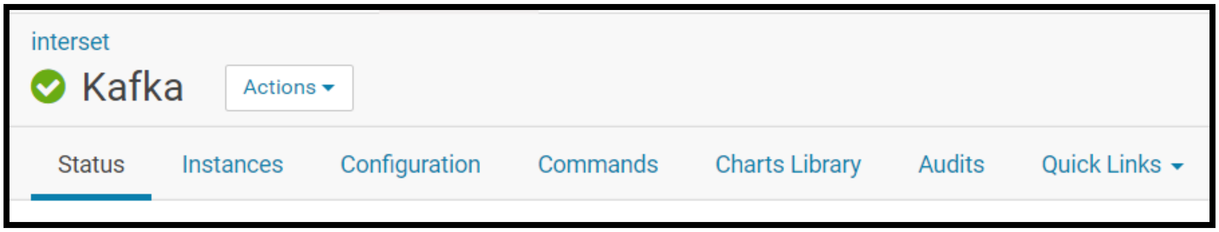
Appendix E: Assigning services to a New node in Cloudera

This section details about assigning gateway services such as Kafka gateway or HBase gateway on to a node.

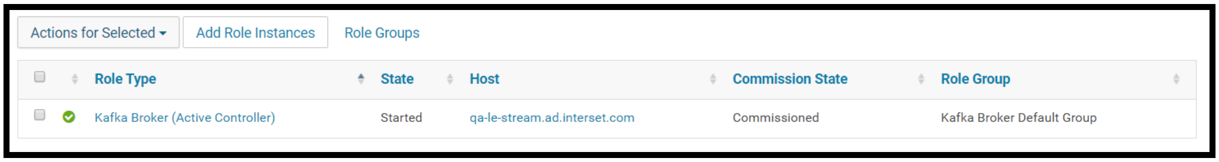
1. In a Web browser, navigate to https://<monitoring_server_fqdn>:7183. Log in to the Cloudera Manager Admin Console.
2. Click the service for which you want to assign a host.



3. Once on the service page, click **Instances** from the menu bar. For example, the following image displays adding an instance for Kafka.

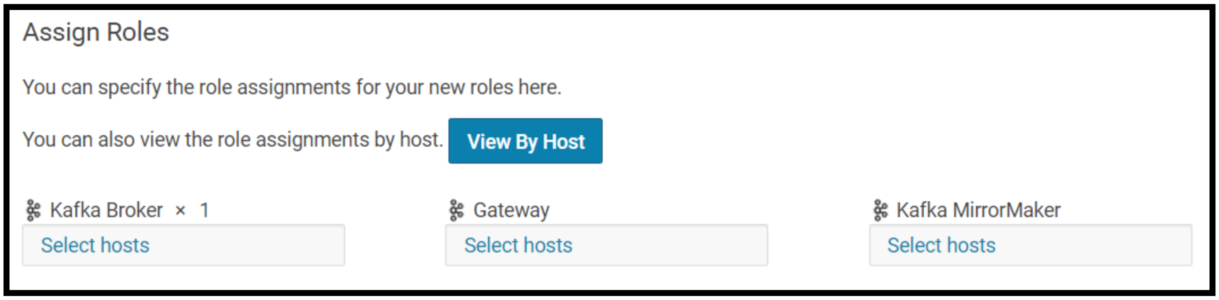


4. Select **Add Role Instances**.


Actions for Selected ▾ | Add Role Instances | Role Groups

□	Role Type	State	Host	Commission State	Role Group
□	✓ Kafka Broker (Active Controller)	Started	qa-le-stream.ad.interset.com	Commissioned	Kafka Broker Default Group

5. Based on your requirement, add the appropriate host against the gateway role specified below.

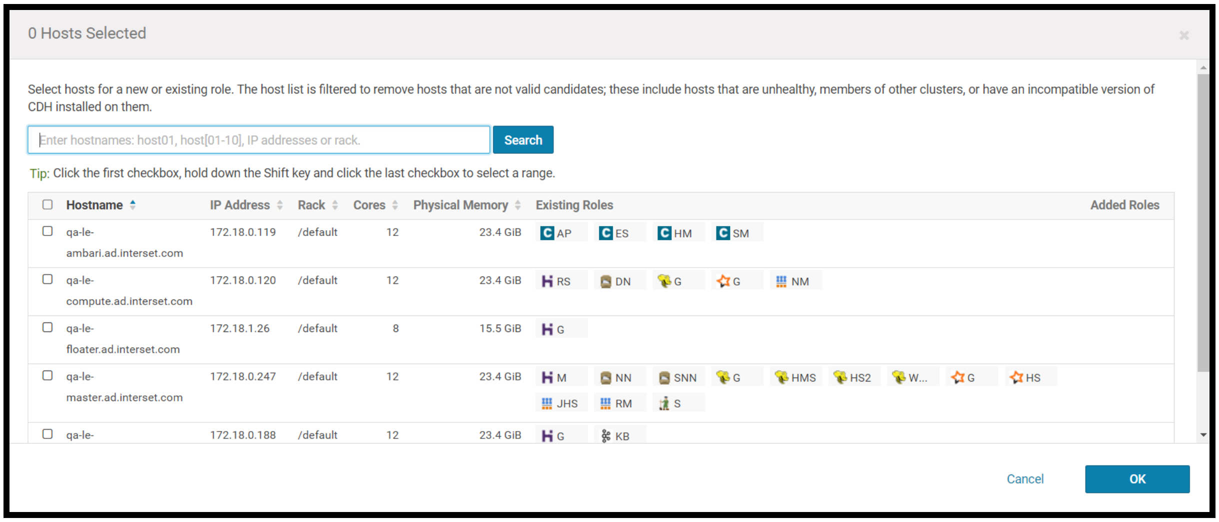


Assign Roles

You can specify the role assignments for your new roles here.

You can also view the role assignments by host.

For example, to add a Gateway role, select **Select hosts** under the Gateway option. You can now click the host you want to add for the Gateway role.



0 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 CDH installed on them.

Enter hostnames: host01, host[01-10], IP addresses or rack.

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

□	Hostname	IP Address	Rack	Cores	Physical Memory	Existing Roles	Added Roles
□	qa-le-ambari.ad.interset.com	172.18.0.119	/default	12	23.4 GiB	AP ES HM SM	
□	qa-le-compute.ad.interset.com	172.18.0.120	/default	12	23.4 GiB	RS DN G G NM	
□	qa-le-floater.ad.interset.com	172.18.1.26	/default	8	15.5 GiB	G	
□	qa-le-master.ad.interset.com	172.18.0.247	/default	12	23.4 GiB	M NN SNN G HMS HS2 W... G HS JHS RM S	
□	qa-le-	172.18.0.188	/default	12	23.4 GiB	G KB	

6. Click **OK** and continue with the assignment.

Note: :For more details refer to Cloudera documentation at: https://-docs.cloudera.com/documentation/enterprise/latest/topics/cm_ mc_ role_ instances.html

Index