Micro Focus Security ArcSight Investigate

Software Version: 2.4.0

Deployment Guide

Document Release Date: July, 2019
Software Release Date: July, 2019

Legal Notices

Copyright Notice

Trademark Notices

Support

Contact Information

Phone	A list of phone numbers is available on the Technical Support Page: https://softwaresupport.softwaregrp.com/support-contact-information
Support Web Site	https://softwaresupport.softwaregrp.com/
ArcSight Product Documentation	https://community.softwaregrp.com/t5/ArcSight-Product-Documentation/ct-p/productdocs

Contents

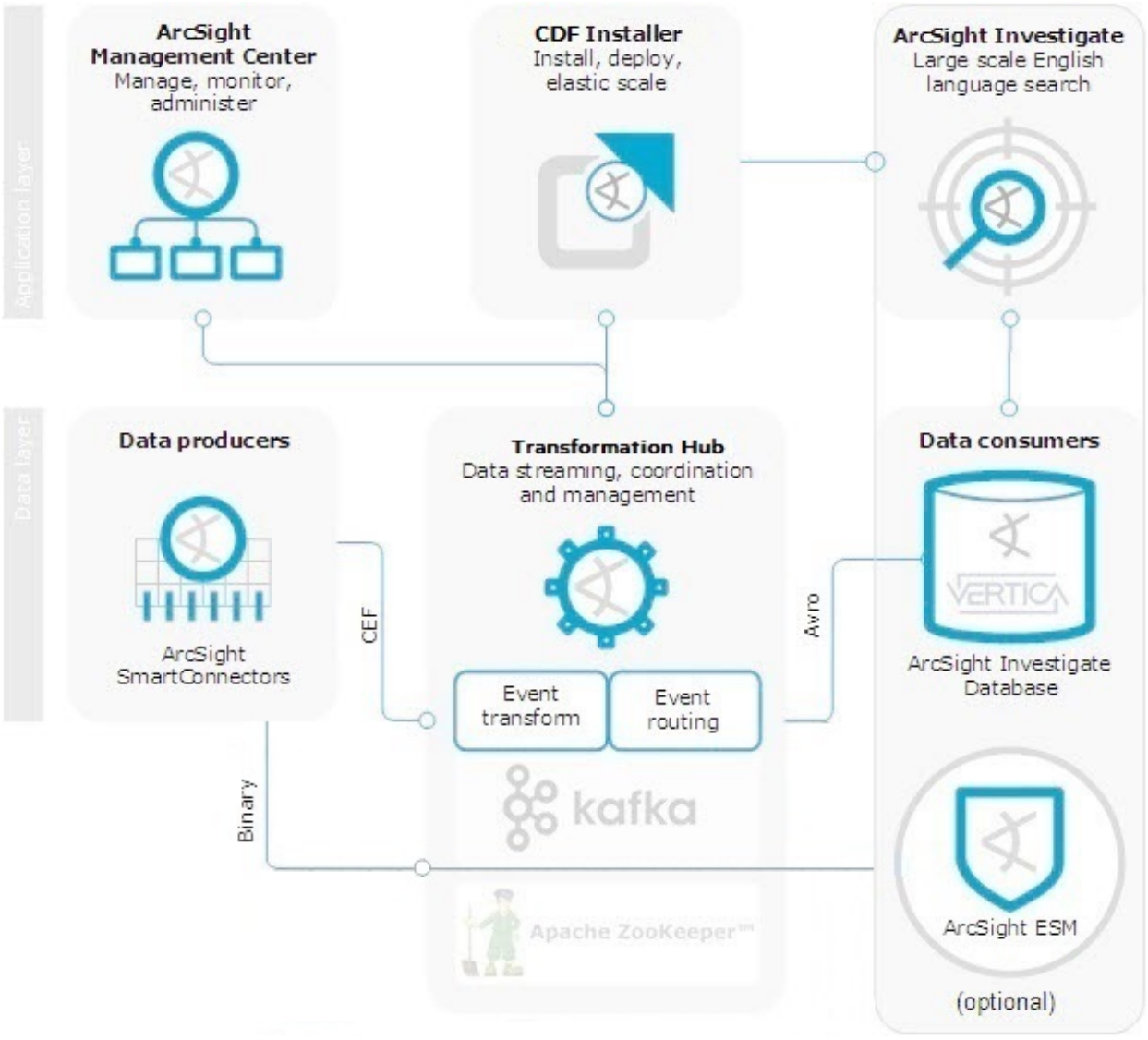
Chapter 1: ArcSight Investigate Architecture 1
ArcSight Investigate 3
Transformation Hub 3
Open Data Platform (ODP) 3
Identity Intelligence 4
Logger 4
SmartConnectors 4
Management Center (ArcMC) 5
Chapter 2: System Requirements 6
Supported Operating Systems 6
Supported Browsers 6
TLS Requirements 6
Network Requirements 7
Encryption Modes 7
Chapter 3: Deployment Planning and Preparation 9
Deployment Overview 9
Download Installation Packages 10
Gather Required Information 11
Calculating volume storage usage for each Transformation Hub (TH) worker node 12
Chapter 4: Configuring the Vertica Server and Installing the Database 14
Configuring the Vertica Server 14
Enabling Password-less SSH Access 17
To Install Vertica 18
Chapter 5: Installation and Deployment 19
Configure and Install the CDF Installer 19
Configure and Deploy the Kubernetes Cluster 20
Download Transformation Hub, Investigate and Core Images to the Local Docker Registry 26
Uploading Images 26
Verify Prerequisite and Installation Images 27
Deploy Node Infrastructure and Services 27
Preparation Complete 29
Configure and Deploy Transformation Hub 29
Configure and Deploy Investigate 31
Label Worker Nodes 34
Check Deployment Status 36

Restart the keepalived Process 37
Check Cluster Status 37
Post-Deployment Configuration 37
Additional Steps 38
Updating Topic Partition Number 38
Adjusting Flannel Pod Memory 39
Updating CDF Hard Eviction Policy 39
Management Center: Configuring Transformation Hub 41
Reminder: Install Your License Key 41
Chapter 6: Complete Vertica Setup 42
Vertica Installer Options 42
Kafka Scheduler Options 43
Chapter 7: Setting FIPS on Vertica 44
To enable FIPS in the OS 44
To disable FIPS 44
Enabling FIPS in Nginx 45
Chapter 8: Configuring Vertica SSL 46
Enabling Vertica SSL 48
Enabling SSL in Scheduler 49
Creating Scheduler with SSL Enabled 49
Setting up Investigate with SSL Enabled 49
Chapter 9: Configuring ArcSight Investigate and Components 51
Creating the System Administrator 51
Updating the Vertica Database Connection 52
Updating the SMTP Server 52
Configuring Search Settings 53
Chapter 10: Enabling the Data Retention Policy on the Vertica Cluster 54
Chapter 11: Backing Up and Restoring the Vertica Database 57
Preparing the Backup Host 57
Backing Up the Vertica Database 59
Backing Up Vertica Incrementally 61
Verifying the Integrity of the Backup 62
Managing Backups 63
Restoring Vertica Data 63
Restoring the Vertica Database 63
Backing Up Investigate Management and Search Datastores 65
Restoring Investigate Management and Search Datastores 65
Troubleshooting 66
Chapter 12: Integrating Transformation Hub Into Your ArcSight Environment 69

Default Topics 69
Configuring ArcMC to Manage Transformation Hub 70
Configuring a SmartConnector as Transformation Hub Producer 71
Configuring Logger as a Transformation Hub Producer 73
Configuring Logger as a Transformation Hub Consumer 75
Appendix A: CDF Installer Script install.sh Command Line Arguments 77
Send Documentation Feedback 80

Chapter 1: ArcSight Investigate Architecture

ArcSight Investigate is a high-capacity data management and analysis engine that enables you to search, analyze, and visualize machine-generated data gathered from web sites, applications, sensors, and devices that comprise your monitored network. Investigate indexes the events from your data source so that you can view and search them. The intuitive search language makes it easy to formulate queries and then create reports and visualizations based on the search results.



Component	Description
ArcSight Investigate	High-capacity data management, search, and analysis web application
ArcSight Investigate Vertica database	Investigate analytic database powered by Vertica Install the Vertica database separately.
CDF Installer	Web application for deploying and configuring Investigate components, including Transformation Hub A Kubernetes cluster manages the components. The master node hosts the CDF Installer, and the worker nodes host Transformation Hub and Investigate web applications.
ArcSight SmartConnectors	Collect and normalize event data from nodes on your network Connectors normalize values (such as severity, priority, and time zone) into a common format and normalize the data structure into a common schema. The connectors then filter and aggregate events to reduce the volume of events sent to the system. Install and maintain connectors separately. Connectors are producers that publish data to Transformation Hub. You can subscribe to data that Transformation Hub manages with Investigate, ArcSight Deployment Platform (ADP), Logger, ArcSight ESM, Apache HDFS, or your own third-party consumer.
Transformation Hub	Centralizes event processing Transformation Hub enables you to take advantage of scalable, high-throughput, multi-broker clusters for publishing and subscribing to event data. It coordinates and manages data streams, which enables your ArcSight environment to scale, and opens ArcSight events to third-party data solutions.
ArcSight Management Center (ArcMC)	Centralized management tool that simplifies security policy configuration, deployment maintenance, and monitoring ArcMC provides run-time management of Transformation Hub topics and is sold as part of ADP.

ArcSight Investigate

ArcSight Investigate is a high-capacity data management and analysis engine that enables you to search, analyze, and visualize machine-generated data gathered from web sites, applications, sensors, and devices that comprise your monitored network. Investigate indexes the events from your data source so you can view and search on them. You can use the English-like search language to generate results from which to create reports and visualizations.

Transformation Hub

Transformation Hub is the high-performance message bus for ArcSight security, network, flows, application, and other events. It can queue, transform, and route security events to other ArcSight or third party software. This Kafka-based platform allows ArcSight components like Logger, ESM, and Investigate to receive the event stream, while smoothing event spikes, and functioning as an extended cache.

Transformation Hub ingests, enriches, normalizes, and then routes Open Data Platform data from data producers to connections between existing data lakes, analytics platforms, and other security technologies and the multiple systems within the Security Operations Center (SOC). Transformation Hub can seamlessly broker data from any source and to any destination. Its architecture is based on Apache Kafka and it supports native Hadoop Distributed File System (HDFS) capabilities, enabling both the ArcSight Logger and ArcSight Investigate technologies to push to HDFS for long-term, low-cost storage.

The latest releases of ArcSight Investigate are integrated with the Transformation Hub for raw events, as well as integrated with ESM to receive alerts and start the investigation process.

ArcSight ESM receives events through the Common Event Format (CEF) interface for dashboarding and further correlation.

This architecture reduces the overall ArcSight infrastructure footprint, scales event ingestion using built-in capabilities and greatly simplifies upgrades to newer Transformation Hub releases. It also positions the platform to support an analytics streaming plug-in framework, supporting automated machine learning and artificial intelligence engines for data source onboarding, event enrichment, and entities and actors detection and attribution.

Open Data Platform (ODP)

ODP centralizes management, monitoring and configuration of the entire data-centric ecosystem using an open architecture. It is configured and monitored through the ArcSight Management Center (ArcMC) user interface. ODP comprises the following ArcSight products:

- Transformation Hub (TH)
- Management Center (ArcMC)
- Smart Connectors (SC)

Identity Intelligence

Micro Focus Identity Intelligence provides interactive and reporting capabilities for identity governance data so you can evaluate requests and approval process activities, support audits of identity governance processes, and review the status of users and access rights. Identity Intelligence gathers data from Micro Focus Identity Manager and Micro Focus Identity Governance, then pushes it to the provided Transformation Hub for processing and Vertica for storage.

Logger

ArcSight Logger provides proven cost-effective and highly-scalable log data management and retention capabilities for the SIEM, expandable to hundreds of nodes and supporting parallel searches. Notable features of Logger include:

- Immutable storage
- High compression
- Archiving mechanism and management
- Transformation Hub integration
- Advanced reporting wizard
- Deployed as an appliance, software or cloud infrastructure
- Regulatory compliance packages

SmartConnectors

SmartConnectors serve to collect, parse, normalize and categorize log data. Connectors are available for forwarding events between and from Micro Focus ArcSight systems like Transformation Hub and ESM, enabling the creation of multi-tier monitoring and logging architectures for large organizations and for Managed Service Providers.

The connector framework on which all SmartConnectors are built offers advanced features that ensures the reliability, completeness, and security of log collection, as well as optimization of network usage. Those features include: throttling, bandwidth management, caching, state persistence, filtering, encryption and event enrichment. The granular normalization of log data allows for the deterministic correlation that detects the latest threats including Advanced Persistent Threats and prepares data to be fed into machine learning models.

SmartConnector technology supports over 400 different device types, leveraging ArcSight's industry-standard Common Event Format (CEF) for both Micro Focus and certified device vendors. This partner ecosystem keeps growing not only with the number of supported devices but also with the level of native adoption of CEF from device vendors.

Management Center (ArcMC)

ArcMC is a central administrative user interface for managing ODP. This management console administers ODP infrastructure, including:

- User management
- Configuration management
- Backup, update and health monitoring to connectors and storage instances

ArcMC's Topology view shows administrators event flow through the entire environment, including a specific focus on monitoring endpoint device log delivery.

Chapter 2: System Requirements

This chapter provides information about supported operating systems, browsers, and compatibility between ArcSight components.

Supported Operating Systems

ArcSight Investigate supports the following operating systems:

Version	Component	Operating system
Investigate 2.4.0	Investigate	CentOS/RHEL 7.5 and 7.6
	Vertica 9.2.1-0 database	CentOS/RHEL 7.5 and 7.6

Supported Browsers

You can use the following browsers with Investigate:

- Microsoft Edge
- Google Chrome
- Mozilla Firefox

Investigate supports the browser version that is available at the time of the Investigate release.

TLS Requirements

The Investigate components interact using encrypted communication with the Transport Layer Security (TLS) 1.2 protocol. TLS implementation requires digital certificates. Before you deploy the components, decide which type of certificate to use. You cannot reconfigure the certificates after deployment.

You can use the following types of certificates:

- Self-signed certificate

  Kubernetes includes the capability to generate self-signed certificates. By default, the Kubernetes deployment process generates certificates for the Kubernetes cluster, but you can choose to generate a self-signed certificate instead. You can also generate Kubernetes certificates for other components.
- Certificate signed by a certificate authority (CA)

  Depending on your organization's security policy, you might be required to use a certificate from a trusted CA. In this case, ensure that you have a root certificate file and a private key file. Copy these files to the Kubernetes master node.

Network Requirements

Before you deploy, complete the following network configuration tasks:

- Configure each node with a fully-qualified domain name.
- Configure DNS across all systems, including correct forward and reverse DNS lookups.
- Enable internet access in order to download the product container images.
- If your organization's network has a proxy, define the proxy environment variable on all servers. Define these variables per user, and not system-wide. For more information, see Configuring Proxy Settings.

Encryption Modes

Before deployment, determine the encryption mode you want to use to encrypt communications between ArcSight components. Before you connect components to Investigate, configure them to use the desired encryption mode. Investigate and the components that connect to it must use the same encryption mode. Changing encryption modes after you deploy Investigate requires system down time.

Product	Open ports	Supported encryption modes	Guidance
ArcSight Management Center (ArcMC)	38080	• TLS • FIPS • ClientAuth	Install ArcMC before you deploy Investigate and Transformation Hub.
ArcSight SmartConnectors	9093	• TLS • FIPS • ClientAuth	You can install and run SmartConnectors before you deploy Investigate and Transformation Hub. FIPS mode is not supported between Connector version 7.5 and Transformation Hub. TLS and ClientAuth are the only encryption methods that are supported between SmartConnector version 7.5 and Transformation Hub.

Product	Open ports	Supported encryption modes	Guidance
ArcSight ESM (optional)	9093	• TLS • FIPS • ClientAuth	You can install and run ESM before you deploy Investigate and Transformation Hub. ESM ingests events more quickly than Investigate. You can leave the ingestion rate asynchronous, or you can equalize the ingestion rates by setting ESM to a lower rate at the connector to reduce the likelihood of a lag in search results when you start Investigate from ESM.
ArcSight Logger (optional)	9093	• TLS • FIPS • ClientAuth	You can install and run Logger before you deploy Investigate and Transformation Hub.

Chapter 3: Deployment Planning and Preparation

Before proceeding with the installation process described in this document, it is assumed that you have already planned and provisioned your network, storage and the cluster of host systems based on requirements described in the CDF Planning Guide requirements. ***You must plan and configure set up a valid environment for deployment, as described in the CDF Planning Guide, before deploying Transformation Hub and Investigate.***

The complete process of deploying Investigate comprises the following high-level steps:

1. **Configure and Install the CDF Installer:** The CDF Installer installs the container management infrastructure. Containerized applications, such as Transformation Hub and Investigate, run in this environment. Depending on your environment, you may need to adjust the default installation parameter values.
2. **Configure and Deploy the Kubernetes Cluster:** Configure and deploy the Master and Worker Nodes, NFS storage, network connectivity, and other infrastructure requirements.
3. **Configure and Deploy Transformation Hub and Investigate:** Using the CDF Installer wizard, configure and deploy Transformation Hub and Investigate to run in the CDF-managed Kubernetes cluster.
4. **Manage Transformation Hub from the Management Center:** Configure the Management Center (ArcMC) to recognize and manage the Transformation Hub cluster.
5. **Integrate Transformation Hub with Other ArcSight Products:** Configure your SmartConnectors and Collectors as producers of events into Transformation Hub and add destinations, as well as configure event Consumers such as Logger and ESM.

Note: The deployment process will validate the infrastructure environment of both, Transformation Hub and Investigate before and after deployment.

Deployment Overview

Before you deploy Investigate, you must install and configure the Vertica database and CDF Installer, and then use ArcSight Installer to deploy Transformation Hub.

Note: Micro Focus recommends that you install these components in a test environment before you put them into production.

1. Obtain the Investigate image and Vertica Installer.
2. Obtain the Transformation Hub images.
3. Configure the Vertica server and install the database.
4. Ensure that Transformation Hub and Investigate each have a dedicated server.

If other applications run on the same servers as Transformation Hub and Investigate, you might experience performance problems.

5. Install the CDF Installer
6. Deploy both Transformation Hub 3.0 and Investigate 2.4.0.
7. Configure both Transformation Hub 3.0 and Investigate 2.4.0.

Download Installation Packages

Now download the installation packages for both the CDF Installer and the Transformation Hub to your Initial Master Node from the Micro Focus Entitlement Portal. After download, validate the digital signature of each file, and then unarchive them.

For a complete list of files required for download, see the Transformation Hub 3.0.0 Release Notes.

To access the ArcSight software in the Micro Focus ArcSight Entitlement Portal, use your Micro Focus credentials which will be authenticated before allowing the download.

Navigate to the version of Transformation Hub you wish to install and download the installation packages for the CDF Installer, the Transformation Hub, and all supporting scripts and wizards that help automate these installs to the directory `{RootFolder}/download/`. The recommended value for the RootFolder is `/opt/arcsight`/.

About the Micro Focus Entitlement Portal

The Micro Focus Entitlement Portal contains ArcSight installation and other product-related materials. This is the only location where you can download the full set of materials needed for Transformation Hub installation.

Some downloaded software will be in compressed format, and in addition, and will have associated signature files (`.md5 or .sig`) that are used to ensure that the downloaded software is authentic.

Validating Downloaded File Signatures

Micro Focus provides a digital public key that is used to verify the software you downloaded from the Micro Focus software entitlement site is indeed from Micro Focus and has not been tampered with by a third party. Visit the Micro Focus Code Signing site for information and instructions on validating the downloaded software.

To verify that the downloaded files are authentic by comparing MD5 file signatures, perform the following steps as the root user on the Initial Master Node for downloaded files.

```
cd {RootFolder}/download/

md5sum {File Name}

cat {File Name}.md5
```

Outputs from each set of compressed installation packages should match their corresponding MD5 signatures. If they do not match, contact Micro Focus Customer Support.

Unarchive Installation Packages

Run the following commands to unarchive your installation packages.

```
cd {RootFolder}/download

unzip cdf-2019.05.xxxx.zip

tar -xvf cdf-core-images-2019.05.xxxx.tgz

tar -xvf transformationhub-3.0.0.xxx.tgz

tar -xvf investigate-2.4.0.xxx.tgz

tar -xvf analytics-2.4.0.xxx.tgz
```

Resulting Directories

After the successful validation and decompression of the installation packages, the following directories and files will be located on your Initial Master Node and contain the installation materials:

```
{RootFolder}/download/cdf-2019.05.xxxx

{RootFolder}/download/cdf-core-images-2019.05.xxxx

{RootFolder}/download/transformationhub-3.0.0.xxx

{RootFolder}/download/investigate-2.4.0.xxx

{RootFolder}/download/analytics-2.4.0.xxx
```

Gather Required Information

During the process described in CDF Planning Guide, you made configuration decisions about your environment, platforms, network, and storage. You will need this information handy now in order to complete the installation of CDF and Transformation Hub.

- **Master and Worker Node Info:** Ensure you have relevant configuration information of the Master and Worker Nodes, including:
  - Credentials for the root or `sudo` user used to run the deployment
  - IP Address and FQDN for every host system in the cluster
  - NFS Server IP Address and FQDN
  - Virtual IP (only required if Master Nodes are configured for high-availability)
- **License Keys:** Ensure you have all required Micro Focus License Keys for the software being installed.

- **Security Mode:** Determine a security mode (FIPS, TLS, or Client Authentication) for communication between ArcSight components.
- **Infrastructure:** Validate and, if necessary, remediate Transformation Hub infrastructure prerequisites.
  - Review, analyze and adjust your Transformation Hub infrastructure configuration properties to meet throughput expectations (for example, Events per Second processing rates).
  - Copy the Deployment Size Calculator spreadsheet (available on the beta FTP site) and edit its contents to determine your disk storage requirements and apply these during the pre-deployment configuration process.
- **Download Access:** Finally, ensure you have access to the Micro Focus software download location. You will download installation packages to the Initial Master Node.

Calculating volume storage usage for each Transformation Hub (TH) worker node

The **volume storage** is where kubernetes and all its related product image and data reside.

The Deployment Size Calculator spreadsheet will be used to calculate the disk storage used for th-cef topic and th-arcsight-avro topic only.

The th-cef topic partition size will be applied to all other predefined topics, i.e. th-binary_esm, th-cef-other, and th-syslog,

Only th-cef and th-arcsight-avro topic partition number will be changed.

- **Calculate the topic size:**

  Follow the worksheet work flow to calculate the total storage needed for both th-cef topic and th-arcsight-avro topic

- **Calculate th-cef and th-arcsight-avro topic partition size on each worker node:**

Partition number = 24 * number of Vertica nodes

Partition number on each TH worker node = (Total planned partition number * 2) / TH worker node number

  - th-cef topic partition size = Total needed th-cef topic size / Partition number on each TH worker node
  - th-arcsight-avro topic partition size = Total needed th-cef topic size / Partition number on each TH worker node

- **Calculate other topics size on each worker node:**
  - Other topics size = th-cef topic partition size * 7 (other default topics) * 6 (default partition number) * 2 (replication) / TH worker node number
- **Partition overhead size on each Transformation Hub worker node:**

  The overhead for each partition segment is around 0.19%, if the default kafka partition segment size, 1G, is applied.

- Total th-cef topic partition overhead = th-cef topic partition size * 0.19% * Partition number on each TH worker node
- Total th-arcsight-avro topic partition overhead = th-arcsight-avro topic partition size * 0.19% * Partition number on each TH worker node

- **Determine CDF Hard Eviction Policy on Worker Node:**

Container Deployment Foundation (CDF) uses a hard eviction policy for worker node. When a hard eviction policy threshold is met, Kubernetes stops all pods immediately.

The default CDF eviction policy is 15%, which means that 15% of the volume disk storage on the worker node can't be used.

Please determine your CDF eviction policy here. To modify the hard eviction policy please see "Updating CDF Hard Eviction Policy" on page 39.

Note: Perform the hard eviction policy after a successful deployment.

- **Total volume disk storage needed:**

Total volume disk storage =

(CDF hard eviction policy) + (th-cef topic partition size * Partition number on each TH worker node) +

(th-arcsight-avro topic partition size * Partition number on each TH worker node) +

(other topics size) + (Total th-cef topic partition overhead) +

(Total th-arcsight-avro topic partition overhead) + (Storage for upgrade 37 GB * 2 = 74 GB) + (some buffer storage)

Chapter 4: Configuring the Vertica Server and Installing the Database

This chapter provides information about configuring the Vertica server and installing the database.

Configuring the Vertica Server

To configure the Vertica server details, please see the Vertica Hardware Guide, and the Vertica System Configuration Task Overview.

The procedure described in this section is a guideline for reference only.

The server configuration is based on an HPE ProLiant DL380 Gen9 server with 48 cores and 128 GB memory.

To avoid performance issues, the Vertica server should be a dedicated server.

To configure the Vertica server:

1. Provision the server with at least 2 GB of swap space, running on CentOS 7.5 and 7.6 or RHEL 7.5 and 7.6.

   Note: Vertica 9.2.1 supports ext3, ext4, and XFS file system.

2. Add the following parameters to `/etc/sysctl.conf`. You must reboot the server for the changes to take effect.

Parameter	Description
`net.core.somaxconn = 1024`	Increases the number of incoming connections
`net.core.wmem_max = 16777216`	Sets the send socket buffer maximum size in bytes
`net.core.rmem_max = 16777216`	Sets the receive socket buffer maximum size in bytes
`net.core.wmem_default = 262144`	Sets the receive socket buffer default size in bytes
`net.core.rmem_default = 262144`	Controls the default size of receive buffers used by sockets

`net.core.netdev_max_backlog = 100000`	Increase the length of the processor input queue
`net.ipv4.tcp_mem = 16777216 16777216 16777216`	
`net.ipv4.tcp_wmem = 8192 262144 8388608`	
`net.ipv4.tcp_rmem = 8192 262144 8388608`	
`net.ipv4.udp_mem = 16777216 16777216 16777216`	
`net.ipv4.udp_rmem_min = 16384`	
`net.ipv4.udp_wmem_min = 16384`	
`net.ipv4.tcp_max_syn_backlog = 4096`	Increases the number of outstanding syn requests allowed
`vm.swappiness = 1`	Defines the amount and frequency at which the kernel copies RAM contents to a swap space For more information, see Check for Swappiness.

3. Add the following parameters to `/etc/rc.local`. You must reboot the server for the changes to take effect.

Parameter	Description
`echo 'echo deadline > /sys/block/sda/queue/scheduler' >> /etc/rc.local`	Changes I/O scheduling to a supported scheduler For more information, see I/O Scheduling.
`echo '/sbin/blockdev --setra 2048 /dev/sda' >> /etc/rc.local`	`/dev/sda` is where Vertica (`/opt`) resides. Sets the disk readahead value For more information, see Disk Readahead.
`echo 'cpupower frequency-set --governor performance' >> /etc/rc.local`	Sets the CPU frequency scaling method This parameter only applies for CentOS. For more information, see CPU Frequency Scaling.
`chmod +x /etc/rc.local`	

4. To increase the process limit, add the following to `/etc/security/limits.d/20-nproc.conf`:

```
* soft nproc 10240
* hard nproc 10240
* soft nofile 65536
* hard nofile 65536
* soft core unlimited
* hard core unlimited
```

5. In `/etc/default/grub`, append line `GRUB_CMDLINE_LINUX` with `intel_idle.max_cstate=0 processor.max_cstate=1`. For example:

   ```
   GRUB_CMDLINE_LINUX="vconsole.keymap=us crashkernel=auto
   vconsole.font=latarcyrheb-sun16 rhgb quiet intel_idle.max_cstate=0
   processor.max_cstate=1"
   grub2-mkconfig -o /boot/grub2/grub.cfg
   ```

6. Use `iptables` to disable the firewall:

   ```
   iptables -F
   iptables -t nat -F
   iptables -t mangle -F
   iptables -X
   systemctl mask firewalld
   systemctl disable firewalld
   systemctl stop firewalld
   ```

   For more information, see Firewall Considerations.

Firewall Requirements

Vertica requires several ports to be open on the local network. It is not recommended to place a firewall between nodes (all nodes should be behind a firewall), but if you must use a firewall between nodes, ensure the following ports are available:

Port	Protocol	Service	Notes
22	TCP	sshd	Required by Administration Tools and the Management Console Cluster Installation wizard.
5433	TCP	Vertica	Vertica client (vsql, ODBC, JDBC, etc) port.
5434	TCP	Vertica	Intra- and inter-cluster communication.
5433	UDP	Vertica	Vertica spread monitoring.
5444	TCP	Vertica Management Console	MC-to-node and node-to-node (agent) communications port. See Changing MC or Agent Ports.

Port	Protocol	Service	Notes
5450	TCP	Vertica Management Console	Port used to connect to MC from a web browser and allows communication from nodes to the MC application/web server.
4803	TCP	Spread	Client connections.
4803	UDP	Spread	Daemon to daemon connections.
4804	UDP	Spread	Daemon to daemon connections.
6543	UDP	Spread	Monitor to daemon connection.

7. Set SELinux to permissive mode:

   ```
   vi /etc/selinux/config
   SELINUX=permissive
   ```

   For more information, see SELinux Configuration.

8. Configure the BIOS for maximum performance:

   System Configuration > **BIOS/Platform Configuration (RBSU)** > **Power Management** > **HPE Power Profile** > **Maximum Performance**

9. Reboot the system, and then use the `ulimit -a` command to verify that the limits were increased.To

Enabling Password-less SSH Access

This section describes how to enable password-less SSH access from the node 1 server to all of the node servers in the cluster.

Note: You must repeat the authentication process for all nodes in the cluster.

To enable password-less SSH access:

1. On the node 1 server, run the `ssh-keygen` command:

   ```
   ssh-keygen -q -t rsa
   ```

2. Copy the key from node 1 to all of the nodes, including node 1, using the node IP address:

   ```
   ssh-copy-id -i ~/.ssh/id_rsa.pub root@11.111.111.111
   ```

   The system displays the key fingerprint and requests to authenticate with the node server.

3. Enter the required credentials for the node.

   The operation is successful when the system displays the following message:

   ```
   Number of key(s) added: 1
   ```

4. To verify successful key installation, run the following command from node 1 to the target node to verify that node 1 can successfully log in:

   ```
   ssh root@11.111.111.111
   ```

To Install Vertica

After you configured the Vertica server and enabled password-less SSH access, install the Vertica database.

1. On the Vertica cluster node 1 server, create a folder for the Investigate Vertica database installer script:

   ```
   mkdir $vertica-install-DIR
   ```

   > **Note:** `$vertica-install-DIR` should not be under `/root`.

2. Copy `arcsight-vertica-installer_2.4.0-7.tar.gz` and `arcsight-vertica-installer_2.4.0-7.tar.gz.md5` to `$vertica-install-DIR`.
3. Verify that the tarball matches the MD5 checksum:

   ```
   cd $vertica-install-DIR
   md5sum arcsight-vertica-installer_2.4.0-7.tar.gz
   cat arcsight-vertica-installer_2.4.0-7.md5
   ```

4. Extract the `.tar` file:

   ```
   tar xvfz arcsight-vertica-installer_2.4.0-7.tar.gz
   ```

5. Edit the `config/vertica_user.properties` file. The `hosts` and `license` properties are required.

Property	Description
`hosts`	A comma separated list of the Investigate Vertica database servers in IPv4 format (for example, 1.1.1.1, 1.1.1.2, 1.1.1.3) If it is necessary to construct the cluster, avoid using local loopback (localhost, 127.0.0.1, etc.).
`license`	Download the license file from the Software Licenses and Downloads portal, and then edit this parameter to point to the license file. **Note:** Without a valid license, an instant-on license will be applied to build a 3 node Vertica cluster only.
`db_retention_day`	Used for the data retention policy.

6. Install Vertica:

   ```
   ./vertica_installer install
   ```

   When prompted, create the database administrator user and the Investigate search user.

   You will need the database administrator credentials to access the Vertica database host. You will need the search user credentials when you configure Vertica from the ArcSight Installer.

   For a list of options that you can specify when installing Vertica, see Vertica Installer Options.

Chapter 5: Installation and Deployment

Once the installation packages have been downloaded, validated, and uncompressed, you are ready to proceed with installation and deployment. In outline, the complete installation and deployment of Transformation Hub consists of these steps, which must be performed in order:

1. Configure and Deploy the CDF Installer

2. Configure and Deploy Kubernetes (k8s)

3. Upload Core Images to the Docker Registry

4. Configure and Deploy Transformation Hub and Investigate

Each of these steps is explained in detail in this chapter.

Configure and Install the CDF Installer

Once the installation packages have already been downloaded, validated and uncompressed in the download folder, you are ready to configure and install the CDF Installer.

Note: You can install the CDF Installer as a root user, or, optionally, as a `sudo` user. However, if you choose to install as a `sudo` user, you must first configure installation permissions from the root user. For more information on providing permissions for the `sudo` user, see Appendix B of the CDF Planning Guide.

To configure and install the CDF Installer:

1. Log in to one of the local Master Nodes where you downloaded and extracted the installation files as the root user. (In this document, the selected Master Node will be referred to as the Initial Master Node. Installations will be initiated from the Initial Master Node.)
2. Install the CDF Installer on the Initial Master Node with the following commands.

Note: For NFS parameter definitions, refer to the CDF Planning Guide section "Configure an NFS Server environment".

```
cd  {RootFolder}/download/{unzipped_CDF_directory}
```

```
./install -m {path_to_a_metadata_file} --k8s-home {path_to_installation_
directory} --docker-http-proxy {your_docker_http_proxy_value} --docker-https-
proxy {your_docker_https_proxy_value} --docker-no-proxy {your_docker_no_
proxy_value} --nfs-server {your_nfs_server_FQDN or IP Address} --nfs-folder
{itom_volume_folder} --ha-virtual-ip {your_HA_ip}
```

You will be prompted for your Admin password, which will inherently meet your password strength requirements. Alternatively, users can include the optional `--password` parameter to supply the password in the installation command.

Example:

```
cd /opt/arcsight/download/cdf-2019.05.xxxx
```

```
./install -m /tmp/arcsight-installer-metadata-2.0.0.xxx.tar.gz --k8s-home
/opt/arcsight/kubernetes --docker-http-proxy "http://web-
proxy.example.com:8080" --docker-https-proxy "http://web-
proxy.example.com:8080" --docker-no-proxy "localhost,127.0.0.1,my-vmenv-
node1,my-vmenv-node1.infra.net,infra.net,15.78.235.235" --nfs-server pueas-
vmenv-nfs.swinfra.net --nfs-folder /opt/nfs/volumes/itom/itom_vol --ha-
virtual-ip 216.3.128.12
```

You may need to configure some additional parameters, depending on your organization's OS, network, and storage configurations.

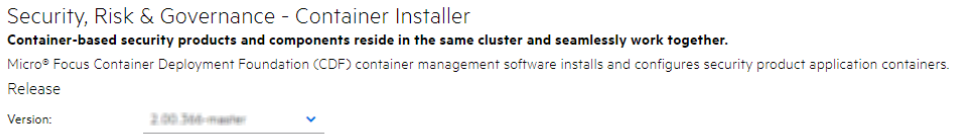
Note: For a description of valid CDF Installer command line arguments, see Installer CLI Commands.

Once the CDF Installer is configured and installed, you can use it to deploy one or more products or components into the cluster.

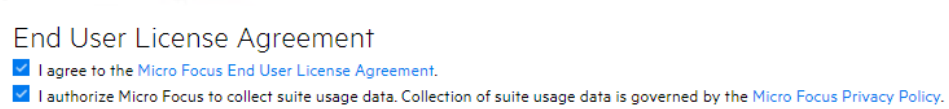
Configure and Deploy the Kubernetes Cluster

After you install the CDF Installer, complete the following steps to deploy your Kubernetes cluster.

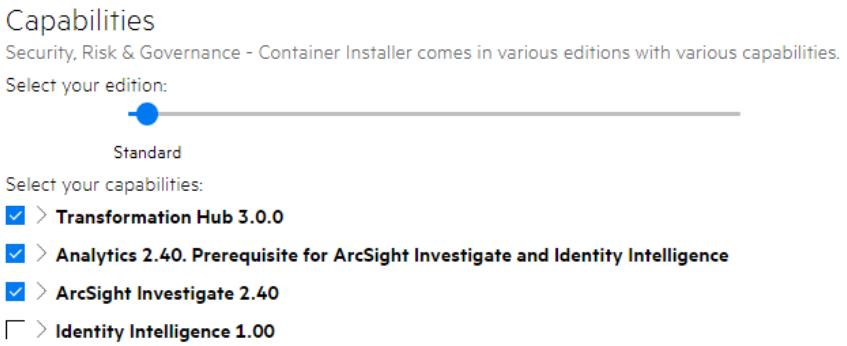
1. Browse to the Initial Master Node at `https://{master_FQDN}:3000`. Log in using admin USERID and the password you specified during the platform installation in the command line argument. (This URL is displayed at the successful completion of the CDF installation shown earlier.)
2. On the **Security Risk and Governance - Container Installer** page, choose the CDF base product metadata version. Then, click **Next**.



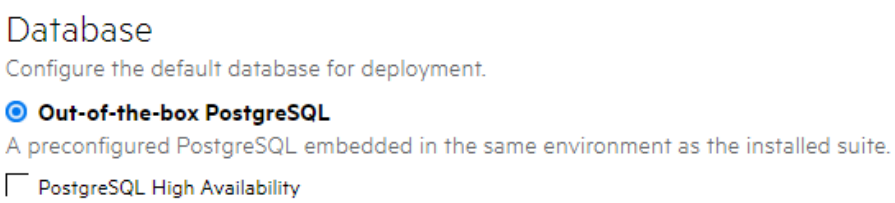
3. On the **End User License Agreement** page, review the EULA and select the *'I agree...'* checkbox. You may optionally choose to have suite utilization information passed to Micro Focus. Then, click **Next**.



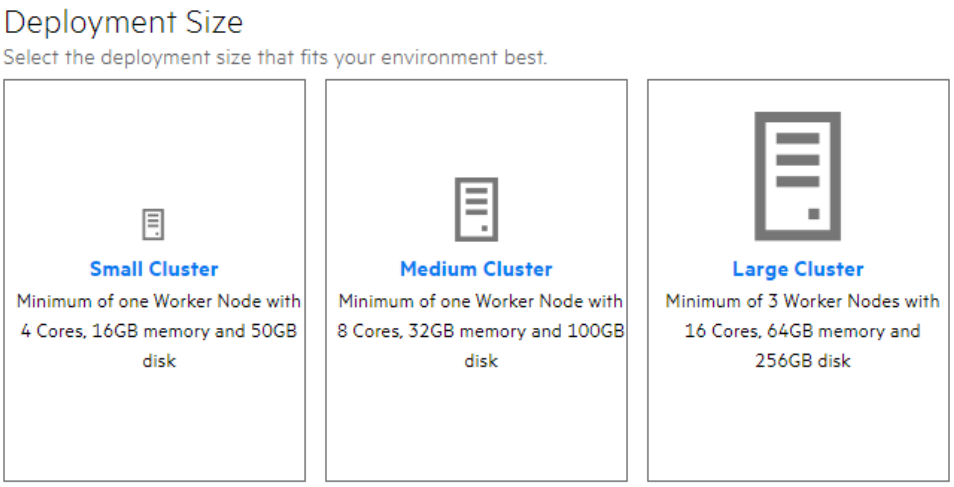
4. On the **Capabilities** page, choose the capabilities and/or products to be installed. To install Transformation Hub as a standalone install, select it. (Note that other products may require Transformation Hub or other capabilities as prerequisites. Such requirements will be noted in the pull-down text associated with the capability.) To show additional infomation associated with the product, click the **>** (greater than) arrow. Then, click **Next**.



5. On the **Database** page, make sure the **PostgreSQL High Availability** box is *deselected*. This database is not used by capabilities in Open Data Platform (ODP).



6. Click **Next.**
7. On the **Deployment Size** page, choose a size for your deployment based on your planned implementation.



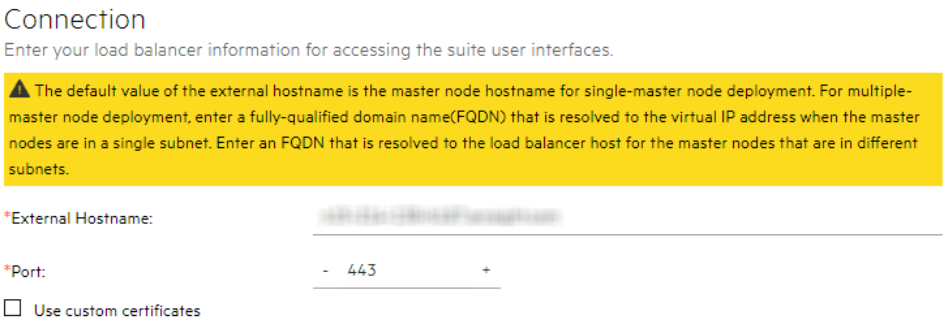
- **Small Cluster:** Minimum of one Worker Node deployed (each node with 4 cores, 16 GB memory, 50 GB disk)
- **Medium Cluster:** Minimum of 1 Worker Node deployed (each node with 8 cores, 32 GB memory, 100 GB disk)
- **Large Cluster:** Minimum of 3 Worker Nodes deployed (each node with 16 cores,64 GB memory, 256 GB disk)

Note: The installation will not proceed if minimal hardware requirements for the deployment are not met.

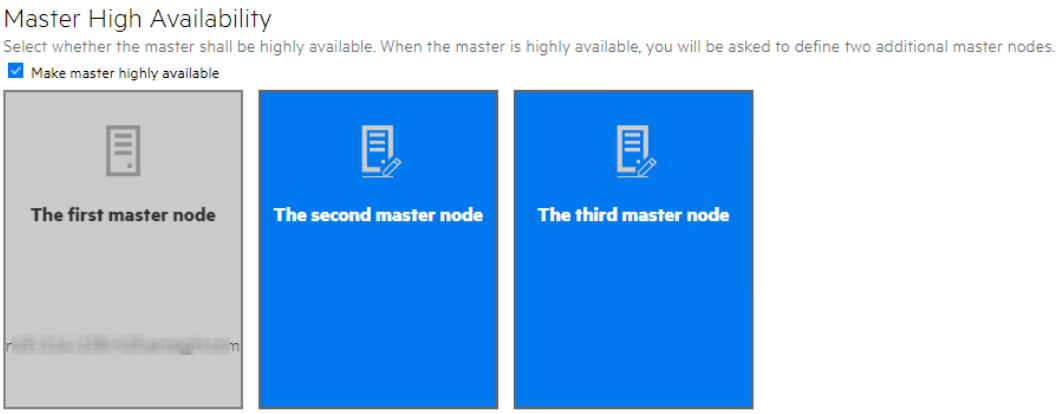
Additional Worker Nodes, with each running on their own host system, can be configured in subsequent steps.

Select your appropriate deployment size, and then click **Next**.

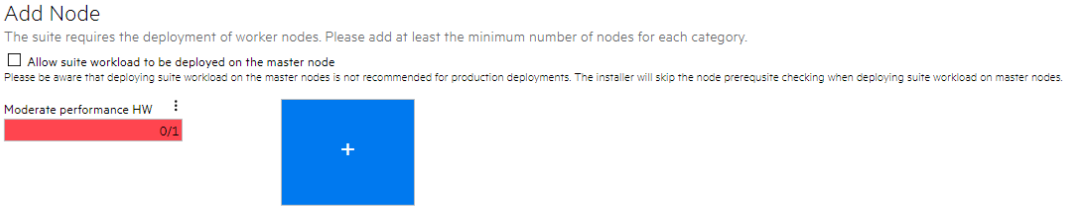
8. On the **Connection** page, an external hostname is automatically populated. This is resolved from the Virtual IP (VIP) specified earlier during the install of CDF (`--ha-virtual-ip parameter`), or the Master Node hostname if the `--ha-virtual-ip` parameter was not specified during CDF installation. Confirm the VIP is correct and then click **Next**.



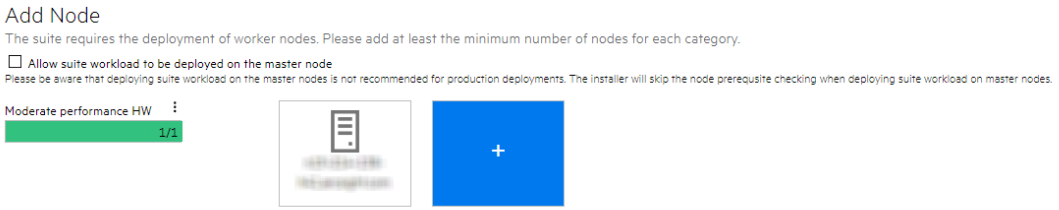
9. On the **Master High Availabilty** page , if high availabilty (HA) is desired, select **Make master highly available** and add 2 additional Master nodes. (CDF requires 3 Master nodes to support high availability.) When complete, or if HA is not desired, click **Next**.



10. On the **Add Node** page, add the first Worker Node as required for your deployment by clicking on the **+** (Add) symbol in the box to the right. The current number of nodes is initially shown in red.

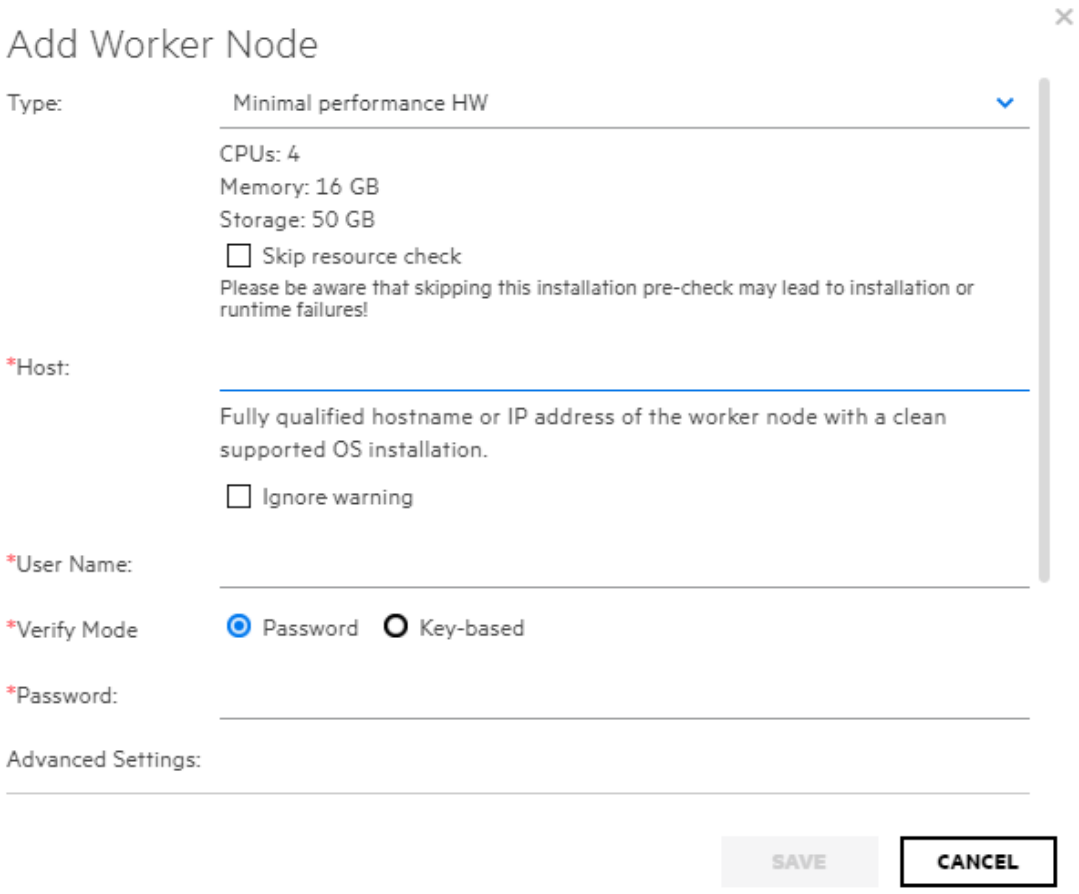


As you add Worker Nodes, each Node is then verified for system requirements. The node count progress bar on the **Add Node** page will progressively show the current number of verified Worker Nodes you have added. This progress will continue until the necessary count is met so the bar will turn from red to green, meaning you have reached the minimum number of Worker Nodes as shown selected in Step 7, above. You may add more Nodes than the minimum number.



Note: Check the **Allow suite workload to be deployed on the master node** to combine master/worker functionality on the same node (Not recommended for production).

On the **Add Worker Node** dialog, enter the required configuration information for the Worker Node, and then click **Save**. Repeat this process for each of the Worker Nodes you wish to add.



Worker Node parameters include:

- **Type:** Default is based on the deployment size you selected earlier, and shows minimum system requirements in terms of CPU, memory, and storage.
- **Skip Resource Check:** If your Worker Node does not meet minimum requirements, select **Skip resource check** to bypass minimum node requirement verification. (The progress bar on the **Add Node** page will still show the total of added Worker Nodes in green, but reflects that the resources of one or more of these have not been verified for minumum requirements.)
- **Host:** FQDN (only) of Node you are adding.

Warning: When adding any Worker Node for Transformation Hub workload, on the **Add Node** page, **always** use the FQDN to specify the Node. **Do not use the IP address.**

- **Ignore Warnings:** If selected, the installer will ignore any warnings that occur during the pre-checks on the server. If deselected, the add node process will stop and a window will display any warning messages. You may wish to start with this deselected in order to view any warnings displayed. You may then evaluate whether to ignore or rectify any warnings, and then run the deployment again with the box selected to avoid stopping.

- **User Name:** User credential to login to the Node.
- **Verify Mode:** Select a verification credential type: Password or Key-based. Then enter the actual credential.

Note: Only one worker node can be added for Investigate. Investigate and Transformation Hub should not reside on the same worker node.

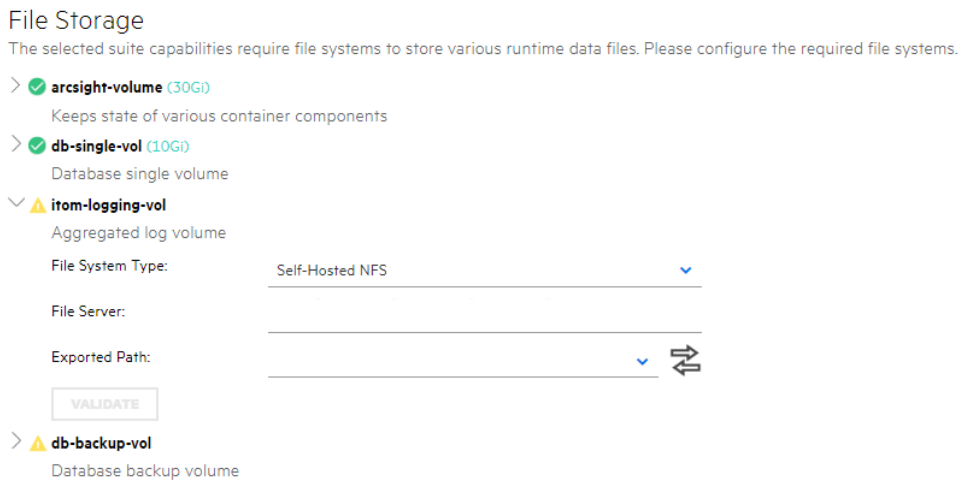
Once all the required Worker Nodes have been added, click **Next**.

12. On the **File Storage** page, configure your NFS volumes.

(For NFS parameter definitions, refer to the CDF Planning Guide section "Configure an NFS Server environment".) For each NFS volume, do the following:

- In **File Server,** enter the IP address or FQDN for the NFS server.
- On the **Exported Path** drop-down, select the appropriate volume.
- Click **Validate**.

Note: All volumes must validate successfully to continue with the installation.



Note: A *Self-hosted NFS* refers to the external NFS that you prepared during when you configured an NFS server environment, as outlined in the CDF Planning Guide. Always choose this value for **File System Type.**

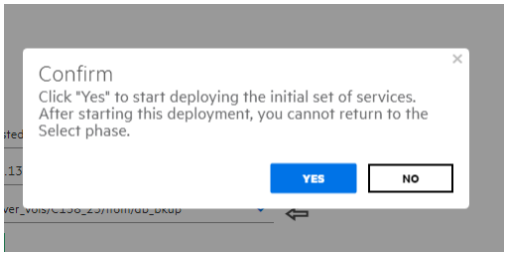
CDF NFS Volume claim	Your NFS volume
arcsight-volume	{NFS_ROOT_FOLDER}/arcsight
itom-vol-claim	{NFS_ROOT_FOLDER}/itom_vol

CDF NFS Volume claim	Your NFS volume
db-single-vol	{NFS_ROOT_FOLDER}/itom/db
itom-logging-vol	{NFS_ROOT_FOLDER}/itom/logging
db-backup-vol	{NFS_ROOT_FOLDER}/itom/db_backup

13. Click **Next**.

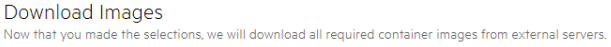
Warning: After you click **Next**, the infrastructure implementation will be deployed. ***Please ensure that your infrastructure choices are adequate to your needs.*** An incorrect or insufficient configuration may require a reinstall of all capabilities.

14. On the **Confirm** dialog, click **Yes** to start deploying Master and Worker Nodes.



Download Transformation Hub, Investigate and Core Images to the Local Docker Registry

By this point, the Transformation Hub, Investigate, and Analytics packages to be installed have already been downloaded from the Micro Focus software site, validated and uncompressed.



No files require download at this point, so on the **Download Images** page, click **Next** to skip this step.

Uploading Images

The **Check Image Availability** page lists the images which have currently been loaded into the local Docker Registry from the originally-downloaded set of images. For a first install, it is expected that no images have already been loaded yet. You will upload the images at this step.

To upload the images to the local Docker Registry:

1. Log on to the Initial Master Node in a terminal session as the root or sudo user
2. Run the following commands to upload the core images to the Local Docker Registry:

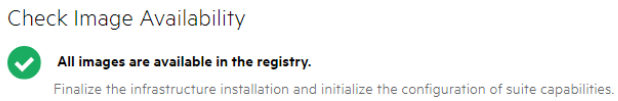
```
cd $k8s-home/scripts
```

```
./uploadimages.sh -u registry-admin -d {RootFolder}/download/analytics-2.4.0.xxx

./uploadimages.sh -u registry-admin -d {RootFolder}/download/transformationhub-3.0.0.xxx

./uploadimages.sh -u registry-admin -d {RootFolder}/download/cdf-core-images-2019.05.xxxx

./uploadimages.sh -u registry-admin -d {RootFolder}/download/investigate-2.4.0.xxx
```

Note: Prior running the image upload process by script, you will be prompted for the administrator password previously specified in the topic "Configure and Install the CDF Installer" on page 19.

Verify Prerequisite and Installation Images

The pre-deployment validation process will verify that all environment prerequisites have been met prior to installing the Transformation Hub.



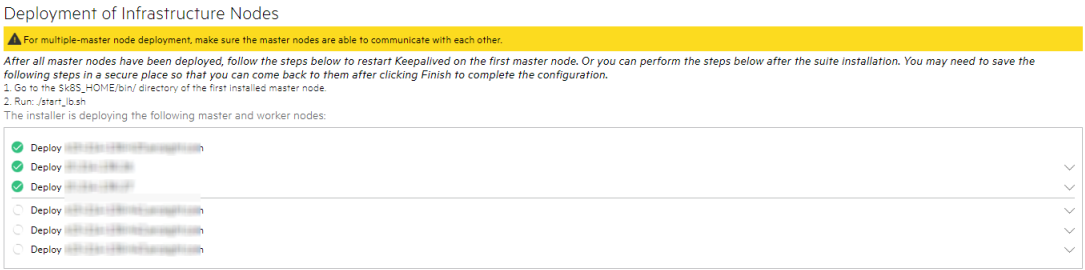
To verify completion of the upload of all images, return to the CDF Management Portal's Check Availability page and click **Check Image Availability Again**. All required component uploads are complete when the message displayed is: ***All images are available in the registry.***

Once verified, click **Next**.

Deploy Node Infrastructure and Services

Node Infrastructure

After the images are verified and you click **Next**, the node infrastructure is deployed. The **Deployment of Infrastructure Nodes** page will display progress.



Please be patient. Wait for all Master and Worker Nodes to be properly deployed (showing a green check icon). Depending on the speed of your network and node servers, this can take up to 15 minutes to complete. Should any node show a red icon, then this process may have timed out. If this occurs, click the drop-down arrow to view the logs and rectify any issues. Then click the **Retry** icon to retry the deployment for that node.

Note: Clicking the **Retry** button will trigger additional communication with the problematic node, until the button converts to a spinning progress wheel indicating that the node deployment process is being started again. Until this occurs, refrain from additional clicking of **Retry**.

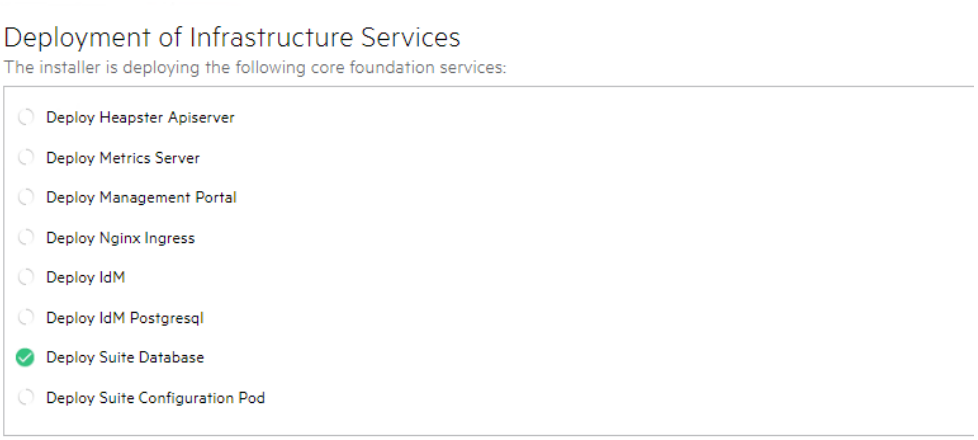
Monitoring Progress: You can monitor deployment progress on a node in the following ways:

- During installation, check the log on the node of interest, in `/tmp/install<timestamp>.log.` Run the command:
  `tail - <logfilename>`
  - After installation has finished, the logs are copied to `$k8s-home/log/scripts/install`
- You can watch the status of deployment pods with the command:
  `kubectl get pods --namespace core -o wide | grep -i cdf-add-node`

Note: the Initial Master Node is not reflected by its own `cdf-add-node` pod.

Infrastructure Services

Infrastructure services are then deployed. The **Deployment of Infrastructure Services** page shows progress.



Please be patient. Wait for all services to be properly deployed (showing a green check icon). This can take up to 15 minutes to complete. Should any service show a red icon, then this process may have timed out. If this occurs, click the **Retry** icon to retry the deployment for that service.

To monitor progress as pods are being deployed, on the Initial Master Node, run the command:

```
watch 'kubectl get pods --all-namespaces'
```

Note: If you try to access the CDF Management Portal Web UI (port 3000) too quickly after this part of the install has finished, you might receive 'Bad Gateway' error. Allow more time for the Web UI to start before retrying your login.

After all services show a green check mark, click **Next**.

Preparation Complete

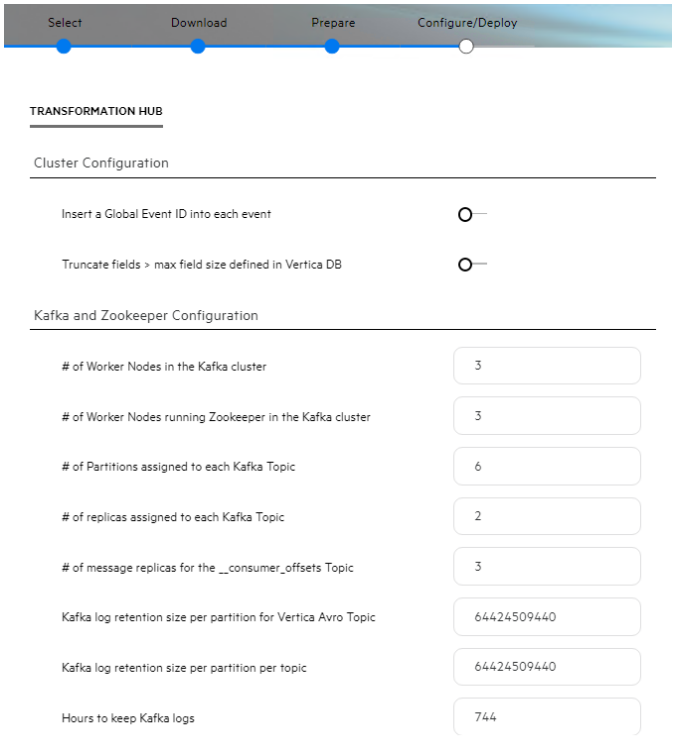
Once all Nodes have been configured, and all services have been started on all nodes, the **Preparation Complete** page will be shown, meaning that the installation process is now ready to configure product-specific installation attributes.

Click **Next** to configure the products and components of the deployment.

Configure and Deploy Transformation Hub

The Transformation Hub is now ready to be configured. The Transformation Hub Pre-Deployment Configuration page is displayed to configure the products and capabilities chosen at the start of the installation process.

The pre-deployment configuration page allows tuning of the initial installation properties. Click the **Transformation Hub** tab and modify the configuration properties as required, based on the size of your cluster and its throughput requirements. Refer to the Deployment Sizing Calculator spreadsheet for guidance on setting some of these properties. Hover over any value to see a detailed description associated with the configuration property.



Worker Node Properties: You must adjust several of these properties with the number of Worker Nodes installed earlier in this installation process.

Minimally, synchronize the following properties to the Worker Nodes.

- # of Worker Nodes in the Kafka cluster
  - Input the worker nodes number used to calculate topic partition size
- # of Worker Nodes running Zookeeper in the Kafka cluster
- # of Schema Registry nodes in the Kafka cluster
- # of Kafka nodes required to run Schema Registry
- Kafka replication factor (this must be set to '1' for a Single Worker deployment)

For example, if you chose a Single Worker installation, you would set the values of all of these properties to 1.

Note: Do not change **# of partitions assigned to each kafka topic.**
# of partitions= 24* Number of Vertica.
# of partitions must be changed after deployment has been completed and successfully. For more information, please see Updating Topic Partition Number.

Log Properties: It is highly likely the following configuration properties should also be adjusted from their default values. Note that proper log sizes are critical. Should logs run out of space, messages (events) will be dropped and are not recoverable.

- Retention log size for each partition of a Vertica Topic
  - Input the calculated th-arcsight-avro topic partition size
- Retention log size per Kafka Topic
  - Input the calculated th-def topic partition size
- Hours to keep Kafka logs
  - Input the hours used for calculating th-def topic partition size
- Kafka partition segment size

ArcMC Properties: For managing your cluster with ArcMC, you can add your Management Center FQDN: {port}. Note that this can also be configured on the post-deployment configuration page.

After updating configuration property values, click **Next** to deploy Transformation Hub. After a few minutes, the CDF Management Portal URL will be displayed. Select this URL to finish Transformation Hub deployment.

Configure and Deploy Investigate

Investigate is now ready to be configured. Investigate Pre-Deployment Configuration page is displayed to configure the products and capabilities chosen at the start of the installation process.

The pre-deployment configuration page allows tuning of the initial installation properties.

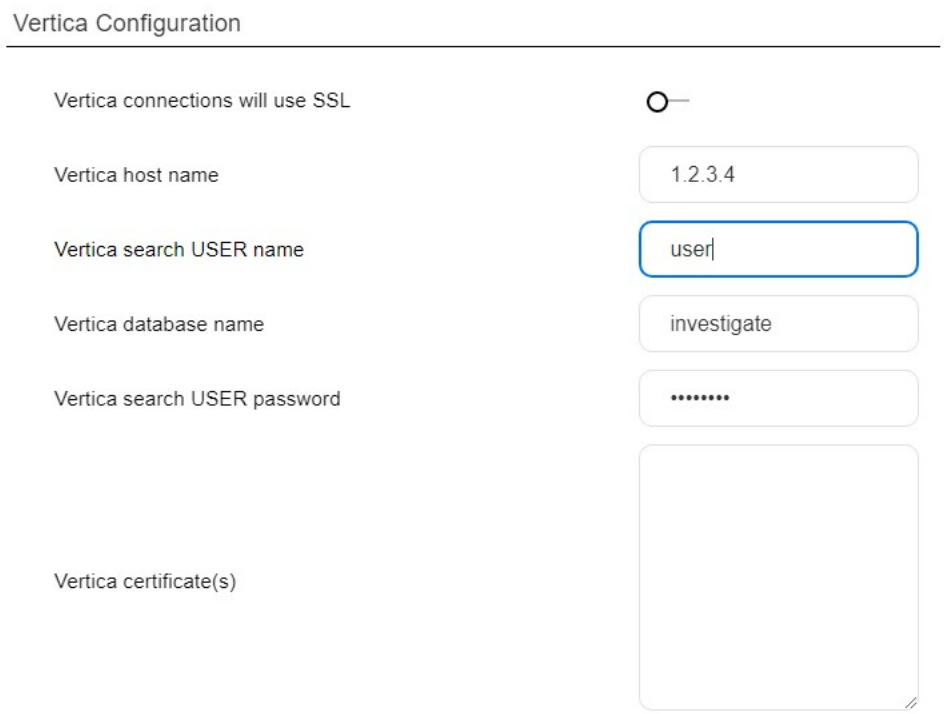
In order to Configure and Deploy investigate, perform the following procedures:

1. Setup of Vertica Database Connection (Mandatory step)
2. Setup SMTP Server (Optional)

Setting Up Vertica Database Connection

Click the **ANALYTICS** tab and modify the configuration properties as required.

In order to setup the set up Vertica database connection, scroll down to **Vertica Configuration**



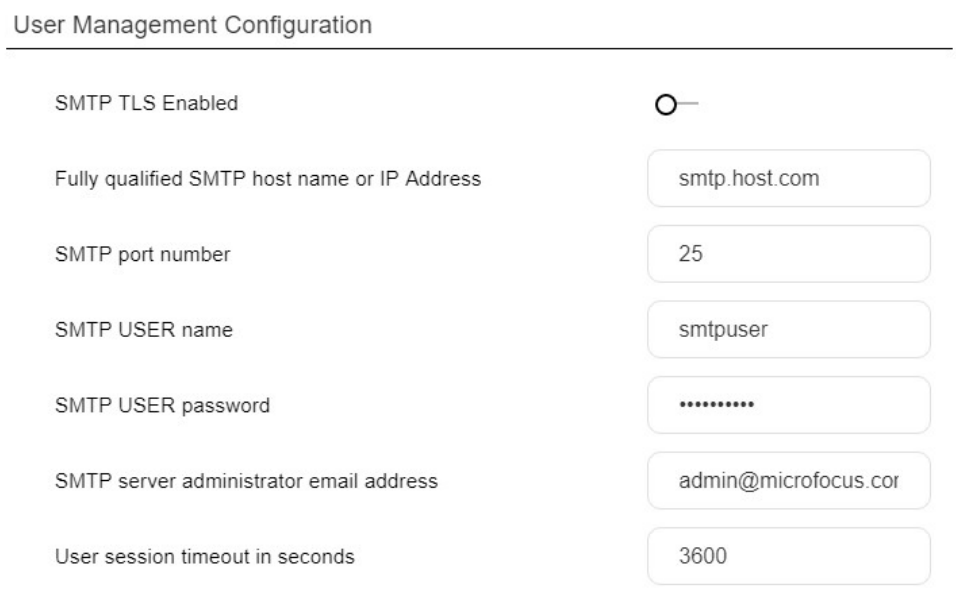
Under Vertica Configuration, provide the following information to update the Vertica connection parameters:

- **Vertica host name:** You can specify any Vertica node IP address, but only specify one address (Use IP address only).
- **Vertica search USER name:** The search user name that you defined when you installed Vertica.
- **Vertica database name:** Investigate.
- **Vertica search USER password:** The search user password that you created when you installed Vertica

Setup SMTP Server

Click the **ANALYTICS** tab and modify the configuration properties as required.

In order to setup the SMTP Server, scroll down to **User Management Configuration**

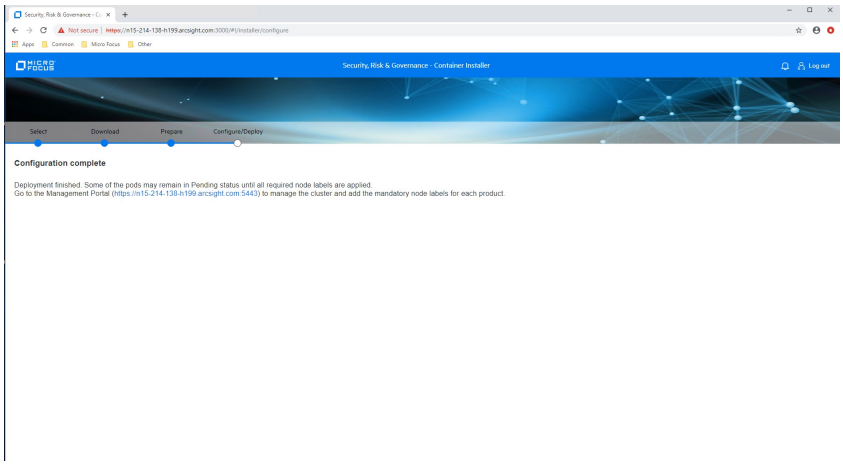


Input the following information, and then click **SAVE:**

- SMTP TLS Enabled
- Fully qualified SMTP host name or IP Address
- SMTP port number
- SMTP USER name
- SMTP USER password
- SMTP server administrator email address
- User session timeout in seconds

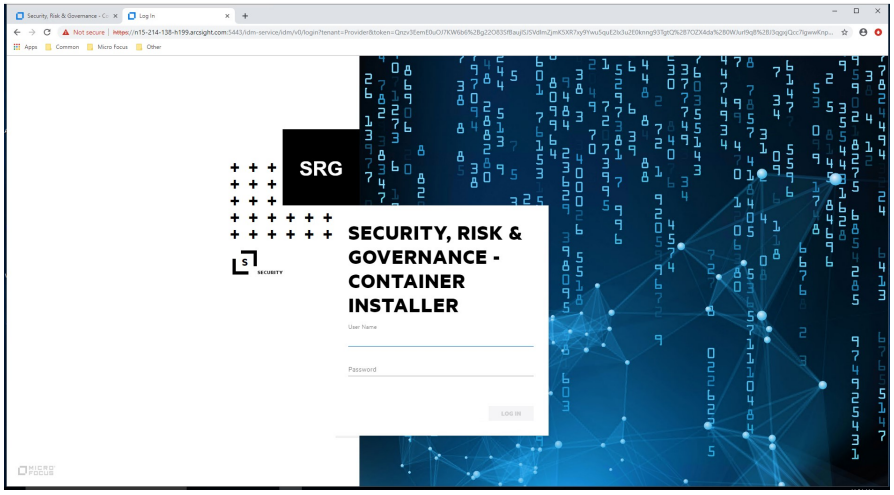
Pre-deployment Configuration Completion

Configuration complete page will be displayed, once pre-deployment has been successfully completed.



To Continue Setup from Management Portal

Go to Management Portal by either clicking the Management Portal link displayed on the Configuration complete page, or browse to `https://<Master_FQDN>:5443` or `https://<Virtualhost_FQDN>:5443` if you deployed in multi-master mode.



Input the following information, and then click **LOG IN**

- User Name: admin
- Password: Password provided during installation

Continue to "Label Worker Nodes" below section.

Label Worker Nodes

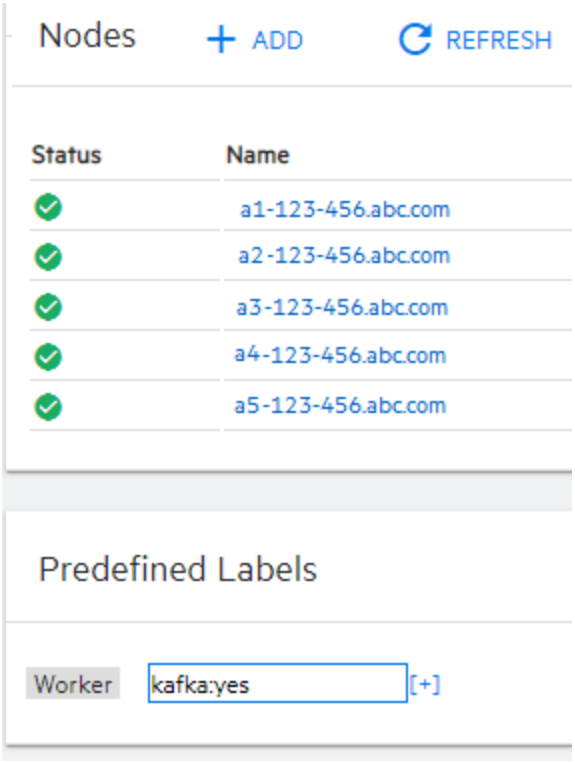
Labeling a node tells Kubernetes what types of workloads can run on a specific host system. Labeling is a means for identifying application processing and qualifying the application as a candidate to run on a specific host system.

Many services will remain in a ***Pending*** state awaiting the labeling process to be completed. Once labeling is completed, Kubernetes will immediately schedule and start the label-dependent containers on the labeled nodes. (Note that starting of services may take 15 minutes or more to complete.)

To label your worker nodes:

1. Login to Management Portal
2. Go to **Administration -> Nodes.**
3. Click **Add**.
4. Add the following labels to the set of predefined labels below the nodes list. Enter the label name and then click the **+** icon. Note that these must be typed exactly as shown below, and they are case-sensitive.

```
analytics:yes

zk:yes

kafka:yes

th-processing:yes

th-platform:yes
```

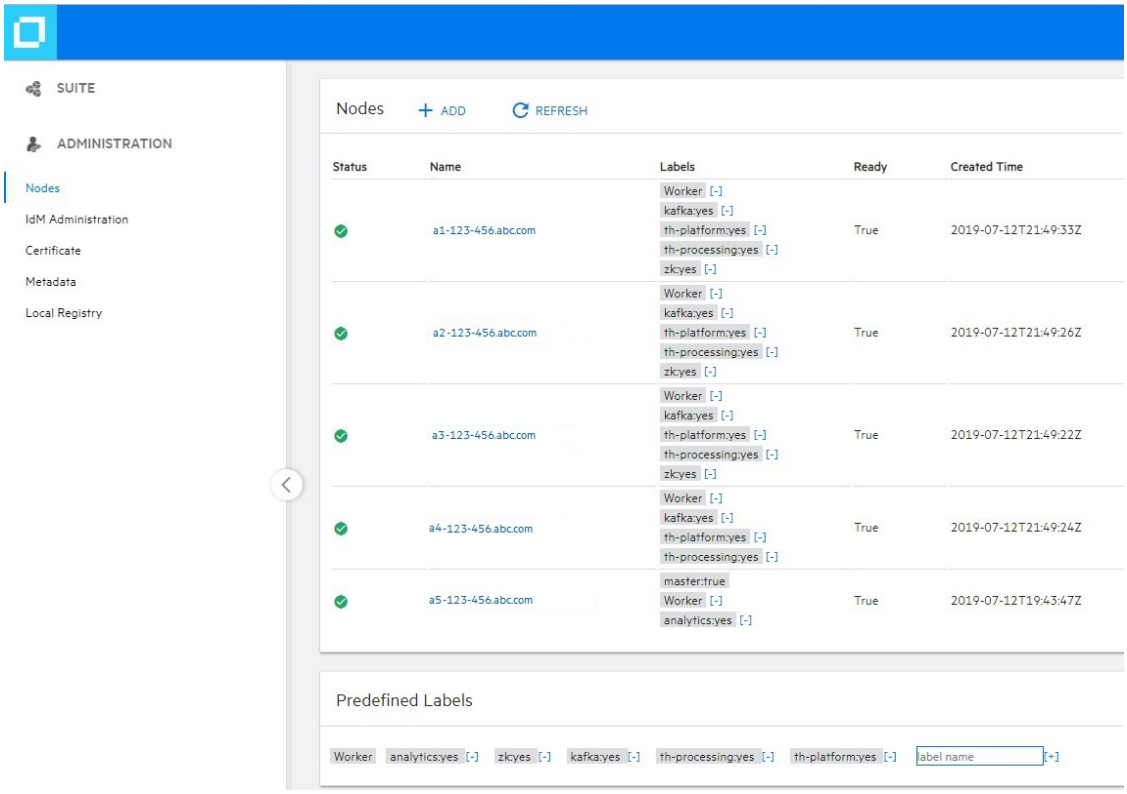


4. Drag and drop a new label from the **Predefined Labels** area to each of the Worker Nodes, based on your workload sharing configuration. This will apply the selected label to the Node.

Note: You must click **Refresh** to see any labels that you have already applied to Nodes.

For Kafka and ZooKeeper, make sure that the number of the nodes you labeled correspond to the number of Worker Nodes in the Kafka cluster and the number of Worker Nodes running Zookeeper in the Kafka cluster properties from the pre-deployment configuration page. The default number is 3 for a Multiple Worker deployment.

For the Investigate node, drag the `analytics: yes` label to the Investigate node.



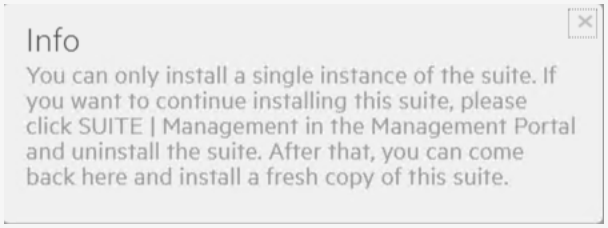
Once the Nodes have been properly labeled, the Transformation Hub services status will change from a ***Pending*** to a ***Running*** state. You can monitor the process by running the following command on the Initial Master Node:

```
kubectl get pods --all-namespaces -o wide
```

Check Deployment Status

When the Configuration Complete page displays, the pod deployment is finished. (Pods that have not been labeled will remain in a ***Pending*** state until labeled.)

Note: If the following error is shown when attempting to log in to the CDF Management Portal on port 3000, this typically means that the CDF installation process has completed, port 3000 is no longer required, and has been closed. Instead of port 3000, log in to the Management Portal on port 5443.



Restart the `keepalived` Process

Note: This step only applies to clusters with multiple Master Nodes.

After deployment has completed, manually restart the keepalive process, which is shut off automatically during the deployment.

1. On the Initial Master Node, go to the `$k8s-home/bin/` directory.
2. Run the following script:

```
./start_lb.sh
```

Check Cluster Status

To verify the success of the deployment, check the cluster status and make sure all pods are running.

Note: You may need to wait 10 minutes or more for all pods to be in a *Running* or *Completed* state.

1. Log into the Initial Master Node.
2. Run the command:

```
kubectl get pods --all-namespaces
```

Review the output to determine the status of all pods.

Post-Deployment Configuration

Depending on your architecture, after deployment, you may need to adjust some of the post-deployment configuration properties in order for Transformation Hub to function correctly.

If you plan to manage Transformation Hub with ArcMC, then you will need to adjust some settings in the post-deployment stage with your ArcMC details. Whether you need to adjust other properties during post-configuration will depend on the specifics of your implementation.

For a more detailed discussion of post-deployment configuration settings, see the Transformation Hub Administrator's Guide.

To configure post-deployment settings:

1. Browse to the Management Portal at `https://<master_FQDN>:5443` or : `https://<Virtualhost_FQDN>:5443` if you deployed in multi-master mode
   - *User Name:* admin
   - *Password:* Password provided during installation

2. Navigate to suite options: **Suite > Management** .
3. Click the **...** (Browse) icon to the right of the main window.
4. From the drop-down, click **Reconfigure**. A new tab will be opened.
5. Select **ANALYTICS**, and scroll down to **Stream Processors and Routers**
6. Under **Stream Processors and Routers**, input the appropriate value for **# of CEF-to-Avro Stream Processors instances to start.**

   Note: 15 was tested as the appropriate value for 120 partitions.

7. For more information on configuration management of Transformation Hub with ArcMC, see Configuring ArcMC Management of Transformation Hub
8. Click **SAVE.**

Web services in the cluster will be restarted (in a rolling manner) across the cluster nodes.

Note: In order to enable ArcMC management, some configuration of ArcMC is also necessary. For more information, see Configuring ArcMC Management of Transformation Hub

Additional Steps

Updating Topic Partition Number

Adjust the partition number for th-cef topic and th-arcsight-avro topic, from default (6) to the number we used to calculate the partition size.

Perform the following steps to update the topic partition number from the master node 1:

1. Run the following commands :

- Find the server ($server), running th-kafka-0:

```
kubectl get pods --all-namespaces -o wide | grep th-kafka-0 | awk '{print $8}'
```

- Find NAMESPACE ($NAMESPACE), for th-kafka-0:

```
kubectl get pods --all-namespaces | grep th-kafka-0 | awk '{print $1}'
```

- Update th-arcsight-avro topic partition number:

```
kubectl exec -n $NAMESPACES th-kafka-0 -- /usr/bin/kafka-topics --zookeeper $server:32181 --alter --topic th-arcsight-avro --partitions $number
```

Note: `$number` is the number used to calculate the partition size.

- Update th-cef topic partition number:

```
kubectl exec -n $NAMESPACES th-kafka-0 -- /usr/bin/kafka-topics --zookeeper $server:32181 --alter --topic th-cef --partitions $number
```

- Use the kafka manager to verify the partition number of th-cef topic and th-arcsight-avro topic have been updated to **$number.**

Adjusting Flannel Pod Memory

In some cases, after flannel pods have been running continuously for some time, the Kafka Manager pod (and others) may terminate abruptly.

To prevent this issue, perform the following steps from the master node 1, after deployment has been successfully completed :

1. Backup existing flannel yaml:

```
cp ${k8s-home}/objectdefs/flannel.yaml ${k8s-
home}/objectdefs/flannel.yaml.orig
```

2. Modify the flannel yaml:

```
sed -i s/50Mi/250Mi/g ${k8s-home}/objectdefs/flannel.yaml
```

3. Delete the flannel file still in use:

```
kubectl delete -f ${k8s-home}/objectdefs/flannel.yaml
```

4. Create a new flannel yaml file:

```
kubectl create -f ${k8s-home}/objectdefs/flannel.yaml
```

To verify the change was made correctly. For each flannel pod, run the following command and make sure the memory value is set to 250Mi.

1. Find all flannel pods,

```
kubectl get pods -n kube-system | grep flannel | awk '{print $1}'
```

2. Check each flannel pod memory limits:

```
kubectl get pods -n kube-system $flannel_pod -o yaml | grep -A6 resources |
grep memory
```

Expected results:

```
memory: 250Mi
memory: 250Mi
```

Updating CDF Hard Eviction Policy

To update the CDF Hard Eviction Policy, perform the following steps on each worker node, after deployment has been successfully completed.

Note: Please verify the operation is successfully executed on one work node first, then proceed on the next worker node.

Note: `eviction-hard` can either be defined as a percentage or a specific amount. The percentage or the specific amount will be determined by the volume storage.

- Run: `cp /usr/lib/systemd/system/kubelet.service /usr/lib/systemd/system/kubelet.service.orig`

```
vim /usr/lib/systemd/system/kubelet.service
```

behind the line

```
ExecStart=/usr/bin/kubelet \
```

add line

```
--eviction-
hard=memory.available<100Mi,nodefs.available<100Gi,imagefs.available<2Gi \
```

- Run: `systemctl daemon-reload and systemctl restart kubelet`

To verify, run: `systemctl status kubelet`

No error should be reported.

Management Center: Configuring Transformation Hub

The Management Center (ArcMC) is the centralized console for managing Micro Focus products.

Connectivity between Transformation Hub and ArcMC is configured in ArcMC when you add Transformation Hub as a managed host into ArcMC. For details on adding your Transformation Hub to ArcMC, see here.

Reminder: Install Your License Key

Transformation Hub ships with a 90-day instant-on evaluation license, which will enable functionality for 90 days after installation. In order for Transformation Hub to continue working past the initial evaluation period, you will need to apply a valid license key to Transformation Hub. A Transformation Hub license key, as well as a legacy ArcMC ADP license key, can be used for licensing Transformation Hub.

For details on how to apply a your license key to Transformation Hub, see the Licensing chapter of the Transformation Hub Administrator's Guide.

IMPORTANT: To ensure continuity of functionality and event flow, make sure you apply your product license **before** the evaluation license has expired.

Chapter 6: Complete Vertica Setup

Follow the steps below to complete the Vertica Setup.

1. Create the schema:

   ```
   ./vertica_installer create-schema
   ```

2. In order to create the Kafka scheduler, run the below commands:

   - If SSL is disabled:

     ```
     ./sched_ssl_setup --disable-ssl
     ```

   - If SSL is enabled, see Configuring Vertica SSL .

3. Create the Kafka scheduler:

   ```
   ./kafka_scheduler create <Transformation_Hub_Node_1_
   IP>:9092,<Transformation_Hub_Node_2_IP>:9092,<Transformation_Hub_Node_3_
   IP>:9092,...
   ```

   For a list of options that you can specify when installing the scheduler, see Kafka Scheduler Options.

4. Check the Vertica status:

   ```
   ./vertica_installer status
   ```

5. Check the scheduler status, event-copy progress, and messages:

   ```
   ./kafka_scheduler status
   ```

   ```
   ./kafka_scheduler events
   ```

   ```
   ./kafka_scheduler messages
   ```

Vertica Installer Options

You can specify the following options when installing Vertica. To specify an option, type `./vertica_installer <Option_Name>`.

Option	Description
install	Installs the Vertica database
uninstall	Uninstalls the Vertica database and deletes data and users
create-schema	Creates the database schema for Investigate
delete-schema	Deletes the Investigate database schema

Option	Description
start-db	Starts the Vertica database with the `dba_password` specified in `vertica_credentials.properties`
stop-db	Stops the Vertica database
status	Prints the Vertica cluster status

Kafka Scheduler Options

You can specify the following options when installing the Kafka scheduler. To specify an option, type `./kafka_scheduler <Option_Name>`.

Option	Description
update	Updates the scheduler
start	Starts the scheduler and begins copying data from all registered Kafka brokers
stop	Stops the scheduler and ends copying data from all registered Kafka brokers
delete	Deletes all registered Kafka instances from the scheduler
status	Prints the following information and log status for a running or stopped scheduler: • Current Kafka cluster assigned to the scheduler • Name and Vertica host where the active scheduler is running • Name, Vertica host, and process ID of every running scheduler (active or backup)
events	Prints event copy progress for the scheduler
messages	Prints scheduler messages

Chapter 7: Setting FIPS on Vertica

In order to enable FIPS mode in Investigate we have to set the OS in FIPS mode.

To enable FIPS in the OS

1. Run the below commands:

```
yum install dracut-fips
```

```
yum install dracut-fips-aesni
```

```
rpm -q prelink && sed -i '/^PRELINKING/s,yes,no,' /etc/sysconfig/prelink
```
- prelink may not be installed, ignore error

```
mv -v /boot/initramfs-$(uname -r).img{,.bak}
```

```
dracut
```

```
grubby --update-kernel=$(grubby --default-kernel) --args=fips=1
```

```
uuid=$(findmnt -no uuid /boot)
```

```
[[ -n $uuid ]] && grubby --update-kernel=$(grubby --default-kernel) --
args=boot=UUID=${uuid}
```

```
reboot
```

2. To verify if FIPS has been enabled, run the following command:

```
sysctl crypto.fips_enabled
```

Expected Result: `crypto.fips_enabled = 1`

To disable FIPS

1. Run the below commands:

```
yum remove dracut-fips
```

```
dracut --force
```

```
grubby --update-kernel=$(grubby --default-kernel) --remove-args=fips=1
```

```
reboot
```

2. To verify if FIPS has been disabled, run the following command:

```
sysctl crypto.fips_enabled
```

Expected Result: crypto.fips_enabled = 0

Enabling FIPS in Nginx

No user action is required to enable FIPS for Nginx. The Nginx docker container is FIPS enabled by default. The FIPS enabled Nginx server will accept TLS 1.2 connections using FIPS compliant Cipher Suites.

Chapter 8: Configuring Vertica SSL

To Acquire the Transformation Hub Certificate:

1. Run the following command from the master node, to obtain the search engine pod:

```
kubectl get pods --all-namespaces | grep hercules-search-engine
```

Output Example:

```
arcsight-installer-9tmsc hercules-search-engine-c97657f9999xpx 2/2 Running 0
17m
```

Note: The search engine pod is reflected as `hercules-search-engine-c97657f9999xpx` in the example above.

2. Run the following command from the master node, to obtain the search engine certificates:

```
kubectl cp <namespace/><pod>:/vault-crt/RE <path to copy> -c <container>
```

Command Example:

```
kubectl cp arcsight-installer-9tmsc/hercules-search-engine-c97657f99-
99xpx:/vault-crt/RE /root -c hercules-search-engine
```

Note: Three certificates will be generated `issue_ca.crt`, `vertica.crt` and `vertica.key`.

3. Copy `issue_ca.crt`, `vertica.crt` and `vertica.key` certificates to <vertica-cluster-node-1>/root. These certificates will be used to configure Vertica to enable SSL between Vertica/scheduler an Vertica/search engine.

To Generate the Certificate for Vertica

From Vertica cluster node 1, perform the following operations:

1. Generate Root CA by running the following command:

```
openssl req -new -x509 -sha256 -newkey rsa:2048 -nodes -keyout ca.key -days
365 -out ca.crt
```

Enter the required information that will be incorparated into your certificate request.

Country Name (2 letter code) []:US

State or Province Name (full name) []:California

Locality Name (eg, city) []:Sunnyvale

Organization Name (eg, company) []:Microfocus

Organizational Unit Name (eg, section) []:Arcsight

Common Name (eg, fully qualified host name) []:Root_CA

Email Address []:admin@arcsight.com

2. Generate Vertica server certificate by running the following command:

```
openssl req -newkey rsa:2048 -new -nodes -keyout server.key -out server.csr
```

Enter the required information that will be incorporated into your certificate request.

Country Name (2 letter code) []:US

State or Province Name (full name) []:California

Locality Name (eg, city) []:Sunnyvale

Organization Name (eg, company) []:Microfocus

Organizational Unit Name (eg, section) []:Arcsight

Common Name (eg, fully qualified host name) []:vertica.domain.com

Email Address []:admin@arcsight.com

Please enter the following 'extra' attributes to be sent with your certificate request

A challenge password []:

3. After entering the requested information, run the following command:

```
openssl x509 -req -in server.csr -days 3650 -sha1 -CAcreateserial -CA ca.crt
-CAkey ca.key -out server.crt
```

Output Example:

```
Signature ok
subject=/C=US/ST=California/L=Sunnyvale/O=Microfocus/OU=Arcsight/CN=vertica.d
omain.com/emailAddress=admin@arcsight.com Getting CA Private Key
```

4. To verify the generated certificate, run the following three comands:

```
openssl x509 -noout -purpose -in server.crt | grep "SSL server"
```

Expected output:

```
SSL server : Yes

SSL server CA : No

Netscape SSL server : Yes

Netscape SSL server CA : No
```

```
openssl x509 -noout -purpose -in ca.crt | grep "SSL server CA : Yes"
```

Expected output:

```
SSL server CA : Yes (WARNING code=3)
```

```
Netscape SSL server CA : Yes (WARNING code=3)
```

```
openssl verify -CAfile ca.crt server.crt
```

Expected output:

```
server.crt: OK
```

Enabling Vertica SSL

In order to enable Vertica SSL run the following command:

1. Login to vertica cluster node 1 as root user

```
cp vertica.crt vertica.key issue_ca.crt /tmp
```

```
 $vertica-install-DIR/vertica_ssl_setup --enable-ssl --vertica-cert-path
server.crt --vertica-key-path server.key --client-ca-path /tmp/issue_ca.crt
```

Verification:

2. Login to vertica server as dbadmin user

```
mkdir ~/.vsql
```

```
cp /tmp/vertica.crt ~/.vsql/client.crt
```

```
cp /tmp/vertica.key ~/.vsql/client.key
```

```
cp /tmp/issue_ca.crt ~/.vsql/root.crt
```

```
chmod 600 ~/.vsql/client.key
```

3. Login to vertica cluster node 1 as root user:

```
rm -rf /tmp/vertica.crt /tmp/vertica.key /tmp/issue_ca.crt
```

4. Check the Vertica connection:

```
vsql -m require
```

Password:

Expected result:

```
 SSL connection (cipher: DHE-RSA-AES256-GCM-SHA384, bits: 256, protocol:
TLSv1.2)
```

Run the following command:

```
dbadmin=> select user,authentication_method, ssl_state from sessions where
session_id = current_session();
```

Expected result:

```
current_user | authentication_method | ssl_state

-------------+----------------------+-----------

dbadmin | Password | Mutual

(1 row)
```

Enabling SSL in Scheduler

To enable SSL in scheduler, run the following command:

```
$vertica-install-DIR/sched_ssl_setup --enable-ssl --sched-cert-path
vertica.crt --sched-key-path vertica.key --vertica-ca-path ca.crt --kafka-ca-
path issue_ca.crt
```

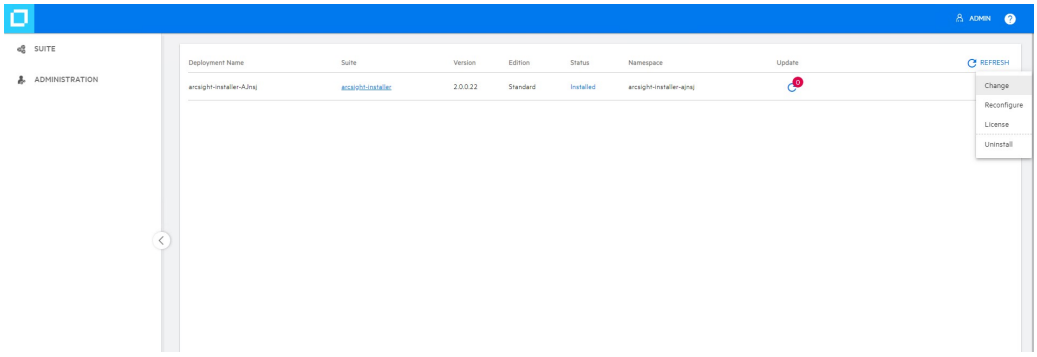
Creating Scheduler with SSL Enabled

To create Scheduler with SSL enabled, run the following command:

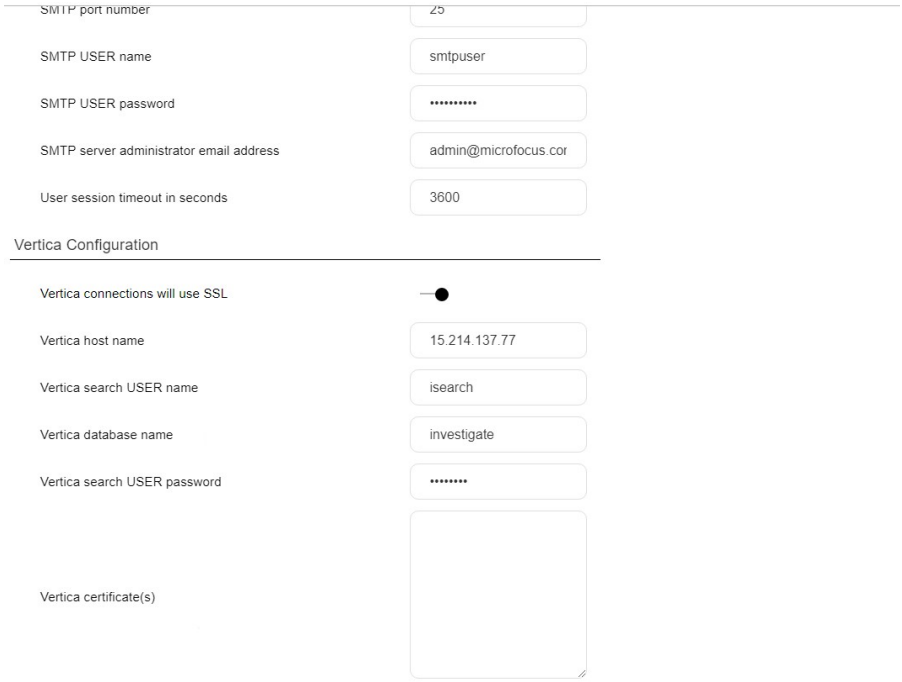
```
$vertica-install-DIR/kafka_scheduler create
<WorkerNode1>:9093,<WorkerNode2>:9093 <WorkerNode3>:9093,…
```

Setting up Investigate with SSL Enabled

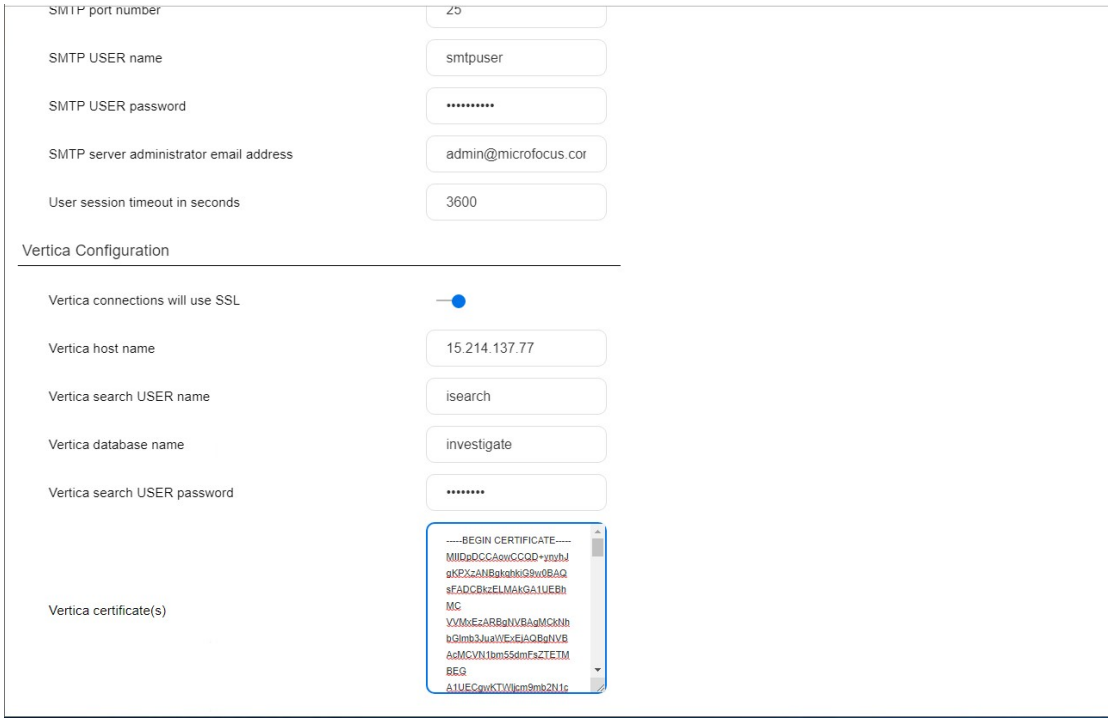
1. Browse to https://<virtual-server-FQDN>:5443, if it is a multiple master, or https://<master-FQDN>:5443, if it is a single master.
2. Navigate to suite options: **Suite > Management**
3. Click the **...** icon under **REFRESH** and Select **Reconfigure.** A new tab will be opened.



4. Select **ANALYTICS**, and scroll down to **Vertica Configuration**
5. Under **Vertica Configuration**, enable **Vertica connections will use SSL**



6. Copy the Vertica ca certificate into the **Vertica Certificate(s)** field, make sure not to include any blank spaces or missing line breaks to prevent a handshake authentication failure.



7. Click **SAVE**. This will restart the search engine pod for the SSL changes to take effect

Chapter 9: Configuring ArcSight Investigate and Components

After you deploy Investigate, use the **Configuration** page of the ArcSight Installer to configure the product. After you change a product setting, Investigate restarts.

Creating the System Administrator

When you log in to Investigate for the first time, you must create the system administrator account. Investigate assigns the `system admin` role to this account.

To create the system administrator account:

1. If you deployed Investigate in single-master mode, open `https://<Master_FQDN>`.

   If you deployed Investigate in multi-master mode, open `https://<Virtualhost_FQDN>`.
2. On the welcome page, enter the name, email, and password information for the system administrator account and then click **Create System Admin**.
3. On the login page, enter the email and password for the system administrator account.

Updating the Vertica Database Connection

Use the ArcSight Installer to update the connection to the Vertica database. Each time you change the connection, the search engine container restarts.

Note: The Vertica database name was defined when you created the schema. You cannot change the name.

To configure the Vertica database connection:

1. Log in to the ArcSight Installer:

   `https://<Master_FQDN>:5443` or `https://<Virtualhost_FQDN>:5443` if you deployed in multi-master mode.
2. Navigate to suite options: **Suite > Management**
3. Click the 3 dots at the end of the selected investigate suite and Select **Reconfigure.**
4. Select **ANALYTICS**, and scroll down to **Vertica Configuration**
5. Under Vertica Configuration, provide the following information to update the Vertica connection parameters:
   - **Vertica host name:** You can specify any Vertica node IP address, but only specify one address.
   - **Vertica search USER name:** The search user name that you defined when you installed Vertica.
   - **Vertica database name:** The name is hard coded to Investigate. You should not change it.
   - **Vertica search USER password:** The search user password that you created when you installed Vertica

Updating the SMTP Server

Update access to your SMTP server to enable users that you create in Investigate to receive notification emails.

To update the SMTP server:

1. Log in to the ArcSight Installer:
   `https://<Master_FQDN>:5443` or `https://<Virtualhost_FQDN>:5443` if you deployed in multi-master mode.
2. Navigate to suite options: **Suite > Management**
3. Click the **...** icon under **REFRESH** and Select **Reconfigure.** A new tab will be opened.
4. Select **ANALYTICS**, and scroll down to **User Management Configuration**
5. Input the following information, and then click **SAVE**:

- Fully qualified SMTP host name or IP Address
- SMTP port number
- SMTP USER name
- SMTP USER password
- SMTP server administrator email address
- User session timeout in seconds

Configuring Search Settings

You can configure the following properties in ArcSight Installer:

- Search query timeout

  Search queries might take a long time and impact performance. You can limit the amount of time that a search query runs. The default search query timeout is 60 minutes.

To configure session and search settings:

1. Log in to the ArcSight Installer: `https://<Master_FQDN>:5443` or `https://<Virtualhost_FQDN>:5443` if you deployed in multi-master mode.
2. Navigate to suite options: **Suite > Management**
3. Click the **...** icon under **REFRESH** and Select **Reconfigure.** A new tab will be opened.
4. Select **ANALYTICS**, and under **Cluster Configuration**, input the appropriate value for **Search Query Timeout in minutes**.

Chapter 10: Enabling the Data Retention Policy on the Vertica Cluster

When Vertica storage approaches usage limits, storage needs to be cleaned up for new events. Data retention script purges old data to reclaim storage.

Note: Storage usage limits are defined by the User.

The retention period can range from 1 to 366 days. The data retention policy is based on calendar days. Calendar day is based on event's deviceReceiptTime

Note: deviceReceiptTime may not correlate to the current date.

The default data retention period is 90 days. If you run the data retention script on 6/30/2019 and the `db_retention_days property` is set to 90, then data older than 04/01/2019 will be deleted. You can purge data in real time or by using a scheduled cron job. Confirmation is needed when retention period is set to less than 30 days.

Note: Vertica data needs to be backed-up routinely. The backup policy is defined by the user. Always evaluate (-e option) retention policy before purging data.

To enable data retention:

1. Run the following command to check disk usage:
   ```
   cd <Vertica_Install_Directory>
   ./vertica_installer status
   Check the disk_space_free_percent
   ```
2. Back up Vertica data.
   For more information, see Backing Up the Vertica Database.
3. Run the following commands:
   ```
   cd <Vertica_Install_Directory>/config
   vi vertica_user.properties
   ```
   Uncomment `#db_retention_days=90`

4. Verify the number of days of data in the Vertica database:

   ```
   cd <Vertica_Install_Directory>/script
   ./retention_policy_util.sh -t
   ```

   The result should be similar to the following:

   ```
   ------------------------------------------------------------------------
   Investigate has 100 day(s) with time-range: [2017-10-26 - 2018-02-06].
   ------------------------------------------------------------------------
   ```

   Note: There are more than 100 calendar days between 2017-10-26 to 2018-02-06. The results above show that there are only 100 event days, meaning that 100 days have incoming events. Certain calendar days did not have incoming events.

5. To change the default retention period, enter the following command:

   ```
   ./retention_policy_util.sh -u <Number_of_Days>
   ```

To purge Vertica data:

1. To create the purge process, enter the following command:

   ```
   ./retention_policy_util.sh -s
   ```

   Note: A cron job is scheduled to purge data daily.

2. To verify the created cron job, enter the following command:

   ```
   ./retention_policy_util.sh -l
   ```

   Expected results:

   ```
   ------------------------------------------------------------------------------------
   Current retention value is set to: 90 day(s)
   ------------------------------------------------------------------------------------
   Current cronjob is running:
   (59 23 * * * /opt/installer/scripts/retention_policy_util.sh -p &>>
   /opt/installer/vertica-installer.log)
   ------------------------------------------------------------------------------------
   ```

3. To preview the purge results, enter the following command:

   ```
   ./retention_policy_util.sh -e
   ```

   The results should be similar to the following:

   ```
   ************************************************************************
   No data will be purged. This is only evaluation for your retention policy
   ************************************************************************
   Will purge time range : [ 2017-10-26 - 2017-10-31 ].
   ```

```
Will purge day 1, (2017-10-26)
Will purge day 2, (2017-10-27)
Will purge day 3, (2017-10-28)
Will purge day 4, (2017-10-29)
Will purge day 5, (2017-10-31)
***** done *****
```

4. To purge data in real time, enter the following command:

   ```
   ./retention_policy_util.sh -p
   ```

5. To disable the purge cron job, enter the following command:

   ```
   ./retention_policy_util.sh -d
   ```

6. To verify the disabled cron job, enter the following command:

   ```
   ./retention_policy_util.sh -l
   ```

   Expected results:

   ```
   ------------------------------------------------------------------------------
   Current retention value is set to: 90 day(s)
   ------------------------------------------------------------------------------
   ```

Chapter 11: Backing Up and Restoring the Vertica Database

You should back up and restore the Vertica database before you upgrade Vertica or before you add or remove a Vertica node.

Consider the following when backing up and restoring the database:

- The backup process can consume additional storage. The amount of space that the backup consumes depends on the size of your catalog and any objects that you drop during the backup. The backup process releases this storage after the backup is complete.
- You can only restore backups to the same version of Vertica. For example, you cannot back up Vertica 9.1.0 and restore it to Vertica 9.2.1.
- Ingesting events into the database during backup might exclude the most recently ingested events from the backup. To ensure that all events are backed up, stop ingestion before you start the backup.
- For optimal network performance, each Vertica node should have its own backup host.
- Use one directory on each Vertica node to store successive backups.
- You can save backups to the local folder on the Vertica node or to a remote server.
- You can perform backups on ext3, ext4, XFS, and NFS file systems.

Preparing the Backup Host

Micro Focus recommends that each backup host have space for at least twice the database node footprint size. Consider your long-term backup storage needs.

If you are using a single backup location, you can use the following Vertica operation to estimate the required storage space for the Vertica cluster:

```
dbadmin=> select sum(used_bytes) as total_used_bytes from v_monitor.storage_
containers;

total_used_bytes

------------------

 5717700329

(1 row)
```

If you are using multiple backup locations, one per node, use the following Vertica operation to estimate the required storage space:

```
dbadmin=> select node_name, sum(used_bytes) as total_used_bytes from v_
monitor.storage_containers group by node_name;
```

```
 node_name | total_used_bytes
-----------------------+--------------------
 v_investigate_node0002 | 1906279083
 v_investigate_node0003 | 1905384292
 v_investigate_node0001 | 1906036954
(3 rows)
```

Remote backup hosts must have SSH access, and you must configure password-less SSH from Vertica node 1 in order for the database administrator to access the hosts.

If one host is the backup destination for multiple Vertica nodes, increase the maximum SSH connections on the backup host by increasing the `MaxStartups` parameter in `/etc/ssh/sshd_config`. The `MaxStartups` number should be greater than the number of nodes in the Vertica cluster.

To set up password-less SSH:

1. Log in to the backup server.
2. Create user `$db_admin`.

   `$db_admin` is the administrator for the Vertica cluster.
3. Ensure that `$db_admin` has write permission to the dedicated directory where you will store the backup.
4. Log in to Vertica node 1 as `root`.
5. Become the Vertica database administrator:

   ```
   # su -l $db_admin
   ```
6. Setup password-less SSH for all backup servers:

   ```
   # ssh-keygen -t rsa
   # ssh-copy-id -i ~/.ssh/id_rsa.pub $db_admin@$back_up_server_ip
   ```

Backing Up the Vertica Database

The `$db_admin` user must perform the backup.

The following options are available for the backup configuration file:

- The default for the number of restore points is 52, assuming a weekly backup for one year. Using multiple restore points gives you the option to recover from one of several backups. For example, if you specify 3, you have 1 current backup and 3 backup archives. Vertica stores the value you enter as the `restorePointLimit` parameter in the `vbr` configuration file.
- To avoid prompting in the future, the backup configuration can save the `$db_admin` password.
- Advanced options allow additional security measures, but Micro Focus recommends using the default options.

To back up the database:

1. Log in to Vertica cluster node 1 as `root`.
2. Generate a backup configuration file.

   Note: The configuration file is required for all future backup and restore operations.

   For each Vertica node, you must specify the backup host. The host can be either the local computer or a remote host.

   For each backup host, you must specify the directory where you want to store the backup. Following is an example configuration:

```
# su -l $db_admin
# /opt/vertica/bin/vbr --setupconfig
Number of restore points: 52
Specify objects:
Object restore mode (coexist, createOrReplace or create): createOrReplace
Vertica user name: $db_admin
Save password to avoid runtime prompt? [y/n]: n
Node v_investigate_node0001
Backup host name: <Backup_Host_1_IP>
Backup directory: /opt/vertica/backup1
Node v_investigate_node0002
Backup host name: <Backup_Host_2_IP>
Backup directory: /opt/vertica/backup2
Node v_investigate_node0003
Backup host name: <Backup_Host_3_IP>
```

```
Backup directory: /opt/vertica/backup3
Change advanced settings? [y/n]: n
Config file name: vertica_backup.ini
Saved vbr config to vertica_backup.ini.
The vertica_backup.ini file is created in /home/$db_admin.
# cat ./vertica_backup.ini
[Misc]
snapshotName = vertica_backup
restorePointLimit = 52
objectRestoreMode = createOrReplace
[Database]
dbName = investigate
dbUser = analyst
dbPromptForPassword = True
[Transmission]
[Mapping]
v_investigate_node0001 = <Backup_Host_1_IP>:/opt/vertica/backup1
v_investigate_node0002 = <Backup_Host_2_IP>:/opt/vertica/backup2
v_investigate_node0003 = <Backup_Host_3_IP>:/opt/vertica/backup3
```

3. Initialize the backup locations:

```
# /opt/vertica/bin/vbr --task init --config-file vertica_backup.ini
```

4. To ensure that you do not lose events during the backup, stop the Kafka scheduler:

```
# exit
# cd /root/install-vertica
./kafka_scheduler stop
```

5. Back up Vertica data:

```
# su -l $db_admin
# /opt/vertica/bin/vbr --task backup --config-file vertica_backup.ini
Starting backup of database investigate.
Participating nodes: v_investigate_node0001.
Enter vertica password:
Snapshotting database.
Snapshot complete.
Approximate bytes to copy: 270383427 of 270383427 total.
[==================================================] 100%
Copying backup metadata.
Finalizing backup.
```

```
Backup complete!
```

6. Verify that the backup files were written to the backup locations:

```
# ssh [BACKUP HOST 1 IP] ls /opt/vertica/backup1
backup_manifest
Objects
Snapshots
# ssh [BACKUP HOST 2 IP] ls /opt/vertica/backup2
backup_manifest
Objects
Snapshots
# ssh [BACKUP HOST 3 IP] ls /opt/vertica/backup3
backup_manifest
Objects
Snapshots
```

Backing Up Vertica Incrementally

Incremental backups use the same setup as a full backup and only back up what changed from the previous full backup. When you perform a full backup using the same configuration file, subsequent backups are incremental. When you start an incremental backup, the `vbr` tool displays a backup size that is a portion of the total backup size. This portion represents the delta changes that will be backed up during the incremental backup.

Run the following command to perform an incremental backup:

```
# /opt/vertica/bin/vbr --task backup --config-file vertica_backup.ini
```

Verifying the Integrity of the Backup

Use the `full-check` option to verify the integrity of the Vertica database backup. The option reports the following:

- Incomplete restore points
- Damaged restore points
- Missing backup files
- Unreferenced files

To verify the backup integrity, run the following command:

```
# /opt/vertica/bin/vbr --task full-check --config-file vertica_backup.ini
```

The output is similar to the following:

```
Checking backup consistency.

List all snapshots in backup location:

Snapshot name and restore point: backup_snapshot_20180116_172347, nodes:['v_
investigate_node0001'].

Snapshot name and restore point: backup_snapshot_20180116_172253, nodes:['v_
investigate_node0001'].

Snapshot name and restore point: backup_snapshot_20180116_172236, nodes:['v_
investigate_node0001'].

Snapshot name and restore point: backup_snapshot_20180116_172310, nodes:['v_
investigate_node0001'].

Snapshot name and restore point: backup_snapshot_20180116_172158, nodes:['v_
investigate_node0001'].

Snapshots that have missing objects(hint: use 'vbr --task remove' to delete
these snapshots):

Backup locations have 0 unreferenced objects

Backup locations have 0 missing objects

Backup consistency check complete.
```

Managing Backups

This section describes how to view and delete backups.

To view available backups, run the following command:

```
# /opt/vertica/bin/vbr --task listbackup --config-file vertica_backup.ini
```

The output is similar to the following:

```
backup backup_type epoch objects nodes(hosts) file_system_type

vertica_backup_20180104_142326 full 29 v_investigate_node0001(10.12.57.27)
[Linux]
```

The backup name includes the backup timestamp.

To delete a backup, run the following command:

```
# /opt/vertica/bin/vbr --task remove --config-file /backup/vertica_backup.ini
--archive 20180104_142326
```

The output is similar to the following:

```
# 20180104_142326 is the backup timestamp

Removing restore points: 20180104_142326

Remove complete!
```

Restoring Vertica Data

Before you restore Vertica data, ensure that your environment meets the following requirements:

- You can only restore backups to the same version of Vertica from which you made the backup. For example, you cannot backup Vertica 8.0.1 and restore it to Vertica 8.1.0.
- You must restore to a cluster that is identical to the cluster from which you made the backup. Ensure that the cluster meets the following requirements:
  - The target database is created and empty.
  - The target database name matches the backup database name.
  - The target database is stopped.
  - All Vertica nodes in the target cluster are running.
  - All Vertica node names in the target cluster match the names from the backup.

Restoring the Vertica Database

The **`$db_admin`** user must perform the restore.

To restore the database:

1. Build a Vertica cluster that is identical to the original cluster.
2. Log in to Vertica node 1 and stop the database:

   ```
   # cd <Vertica_Installation_Directory>
   # ./vertica_installer stop-db
   ```

3. Become the `$db_admin` user:

   ```
   # su -l $db_admin
   ```

4. Copy `vertica_backup.ini` to `/home/$db_admin`.
5. Restore the backup data:

   ```
   # /opt/vertica/bin/vbr --task restore --config-file vertica_backup.ini
   ```

   The output should be similar to the following:

   ```
   Starting full restore of database investigate.
   Participating nodes: v_investigate_node0001, v_investigate_node0002, v_
   investigate_node0003.
   Restoring from restore point: investigate_backup_20180110_010826
   Determining what data to restore from backup.
   [==================================================] 100%
   Approximate bytes to copy: 2246248425 of 2246250258 total.
   Syncing data from backup to cluster nodes.
   [==================================================] 100%
   Restoring catalog.
   Restore complete!
   ```

6. Start the database:

   ```
   # exit
   # ./vertica_installer start-db
   ```

   The output should be similar to the following:

   ```
   Starting nodes:
   v_investigate_node0001 (127.0.0.1)
   Starting Vertica on all nodes. Please wait, databases with a large catalog
   may take a while to initialize.
   Node Status: v_investigate_node0001: (DOWN)
   Node Status: v_investigate_node0001: (DOWN)
   Node Status: v_investigate_node0001: (DOWN)
   Node Status: v_investigate_node0001: (DOWN)
   Node Status: v_investigate_node0001: (UP)
   Database investigate started successfully
   ```

7. Start the Kafka scheduler:

```
# cd /root/install-vertica
# ./kafka_scheduler start
```

Backing Up Investigate Management and Search Datastores

Micro Focus recommends that you use a backup location that is not under the `/opt/arcsight` directory. Use a local folder on the system or a remote location.

This procedure uses the `/opt/investigate/backup` directory as an example.

To back up the data stores:

1. To prohibit database access, undeploy Investigate.

   For information about undeploying Investigate, see .
2. SSH to the Kubernetes cluster master node 1.
3. Run the following commands:

```
# cd /opt/arcsight/volumes/investigate/
# mkdir -p /opt/investigate/backup
# cp –R * /opt/investigate/backup
# diff -r -s /opt/investigate/backup/mgmt
/opt/arcsight/volumes/investigate/mgmt
# diff -r -s /opt/investigate/backup/search
/opt/arcsight/volumes/investigate/search
```

   If you do not receive a message that states that the files are identical, repeat the commands.
4. Redeploy Investigate to resume operations.

   For information about deploying Investigate, see Deploying and Undeploying Investigate Images.
5. Before you resume Investigate operations, ensure that the pods are in Running status:

```
# kubectl get pods --all-namespaces | grep investigate
```

Restoring Investigate Management and Search Datastores

When restoring the Investigate management and search datastores, retain the original directory structure under `/opt/arcsight/volumes/investigate/`.

The management datastore will be restored to the `/opt/arcsight/volumes/investigate/mgmt/db/` directory. The search datastore will be restored to the `/opt/arcsight/volumes/investigate/search` directory.

To restore the datastores:

1. Ensure that you have a valid backup of the datastores.

   For more information, see Backing Up Investigate Management and Search Datastores.
2. To prohibit access to the database, undeploy Investigate.

   For information about undeploying Investigate, see .
3. SSH to the Kubernetes master node, and then run the following commands:

   ```
   # cd /opt/investigate/backup
   # cp –R search/* /opt/arcsight/volumes/investigate/search
   ```

   Reply `yes` to overwrite files and folders.

   ```
   # cd /opt/arcsight/volumes/investigate/mgmt/db/
   # rm - rf h2.lock.db
   # cp /opt/investigate/backup/mgmt/db/h2.mv.db .
   ```

   Reply `yes` to overwrite files and folders.

   ```
   # diff -r -s /opt/arcsight/volumes/investigate/mgmt/db/h2.mv.db
   /opt/investigate/backup/mgmt/db/h2.mv.db
   # diff -r -s /opt/investigate/backup/search
   /opt/arcsight/volumes/investigate/search
   ```

   You should receive a message stating that all files are identical. If they are not identical, repeat the procedure.
4. Change the permission of the Investigate directory:

   ```
   # chown 1999:1999 -R /opt/arcsight/volumes/investigate/
   ```
5. Redeploy Investigate to resume operations.

   For information about deploying Investigate, see Deploying and Undeploying Investigate Images.
6. Before you resume Investigate operations, ensure that the pods are in Running status:

   ```
   # kubectl get pods --all-namespaces | grep investigate
   ```

Troubleshooting

If the Vertica cluster downtime exceeds the retention time for the Kafka cluster, the Vertica-stored Kafka offset might not be present in the Transformation Hub cluster. In this case, the scheduler will not be able to consume new data. This section describes how to resolve the issue.

You can confirm whether the scheduler is copying data by checking the status and examining the last copied offset in the microbatch status. If the offset number is not increasing, then the scheduler can no longer find the valid offset and must be reset.

To check the scheduler offsets, run the following command in the Vertica installation directory:

```
./kafka_scheduler events
```

```
…
Event Copy Status for (eb-internal-avro) topic:
frame_start | partition | start_offset | end_offset | end_reason | copied
bytes | copied messages
-------------------------+-----------+--------------+------------+-----------
-+--------------+-----------------
2018-06-09 16:57:40.599 | 1 | 6672721851 | 6672743683 | END_OF_STREAM | 0 | 0
2018-06-09 16:57:40.599 | 2 | 6693800372 | 6693818421 | END_OF_STREAM | 0 | 0
2018-06-09 16:57:40.599 | 0 | 6710608899 | 6710626273 | END_OF_STREAM | 0 | 0
2018-06-09 16:57:40.599 | 4 | 6684909292 | 6684928573 | END_OF_STREAM | 0 | 0
2018-06-09 16:57:40.599 | 5 | 6690363437 | 6690385300 | END_OF_STREAM | 0 | 0
2018-06-09 16:57:40.599 | 3 | 6703797344 | 6703813421 | END_OF_STREAM | 0 | 0
2018-06-09 16:57:15.573 | 2 | 6693782400 | 6693800372 | END_OF_STREAM | 0 | 0
2018-06-09 16:57:15.573 | 1 | 6672702552 | 6672721851 | END_OF_STREAM | 0 | 0
2018-06-09 16:57:15.573 | 3 | 6703785764 | 6703797344 | END_OF_STREAM | 0 | 0
2018-06-09 16:57:15.573 | 4 | 6684890676 | 6684909292 | END_OF_STREAM | 0 | 0
2018-06-09 16:57:15.573 | 5 | 6690346763 | 6690363437 | END_OF_STREAM | 0 | 0
2018-06-09 16:57:15.573 | 0 | 6710597067 | 6710608899 | END_OF_STREAM | 0 | 0
```

If the scheduler is not consuming data, recreate the scheduler:

```
# ./kafka_scheduler delete
Are you sure that you want to DELETE scheduler metadata (y/n)?y
Terminating all running scheduler processes for schema: [investigation_
scheduler]
scheduler instance(s) deleted for 192.214.138.94
bash: /root/install-vertica/kafka_scheduler.log: No such file or directory
scheduler instance(s) deleted for 192.214.138.95
bash: /root/install-vertica/kafka_scheduler.log: No such file or directory
scheduler instance(s) deleted for 192.214.138.96
db cleanup: delete scheduler metadata
# ./kafka_scheduler create
192.214.137.72:9092,192.214.137.71:9092,192.214.136.7:9092
create scheduler under: investigation_scheduler
scheduler: create target topic
```

```
scheduler: create cluster for
192.214.137.72:9092,192.214.137.71:9092,192.214.136.7:9092

scheduler: create source topic for
192.214.137.72:9092,192.214.137.71:9092,192.214.136.7:9092

scheduler: create microbatch for
192.214.137.72:9092,192.214.137.71:9092,192.214.136.7:9092

scheduler instance(s) added for 192.214.138.94

scheduler instance(s) added for 192.214.138.95

scheduler instance(s) added for 192.214.138.96
```

Chapter 12: Integrating Transformation Hub Into Your ArcSight Environment

Transformation Hub centralizes event processing and enables event routing, which helps you to scale your ArcSight environment and opens event data to ArcSight and third-party solutions. Transformation Hub takes advantage of scalable and highly-available clusters for publishing and subscribing to event data. Transformation Hub integrates with ArcSight SmartConnectors and Collectors, Logger, ESM, and ArcSight Investigate. It is managed and monitored by ArcSight Management Center.

After you install and configure Transformation Hub you can use SmartConnectors and Collectors to produce and publish data to the Transformation Hub, and to subscribe to and consume that data with Logger, ESM, ArcSight Investigate, Apache Hadoop, or your own custom consumer.

Transformation Hub supports both Common Event Format (CEF) versions, 0.1 and 1.0.

- CEF 0.1 is the legacy ArcSight CEF version that supports IPv4 addresses available with SmartConnector version 7.4 and earlier.
- CEF 1.0, available with SmartConnector version 7.5 and later and Collectors version 7.8 and later, supports IPv4 and IPv6 addresses.

Transformation Hub third-party integration and other product features are explained in detail in the Transformation Hub Administrator's Guide, available from the ArcSight support community.

Default Topics

Transformation Hub manages the distribution of events to topics, to which consumers can subscribe and receive events from.

Transformation Hub includes the following default topics:

Topic Name	Event Type	Notes
th-cef	CEF event data.	Can be configured as a SmartConnector or Collector/CTH destination.
th-binary_esm	Binary security events, which is the format consumed by ArcSight ESM.	Can be configured as a SmartConnector or Collector/CTH destination.
th-syslog	The Connector in Transformation Hub (CTH) feature sends raw syslog data to this topic using a Collector.	Can be configured as a SmartConnector or Collector/CTH destination.
th-cef-other	CEF event data destined for a non-ArcSight subscriber	

Topic Name	Event Type	Notes
th-arcsight-avro-sp_metrics	For ArcSight product use only. Stream processor operational metrics data.	
th-arcsight-avro	For ArcSight product use only. Event data in Avro format for use by ArcSight Investigate.	
th-arcsight-json-datastore	For ArcSight product use only. Event data in JSON format for use by ArcSight infrastructure management.	

In addition, using ArcSight Management Center, you can create new custom topics to which your SmartConnectors can connect and send events.

Configuring ArcMC to Manage Transformation Hub

ArcMC serves as the management UI for Transformation Hub. In order for ArcMC to manage Transformation Hub, Transformation Hub must be added as a managed host to ArcMC. This process will include these steps, explained below:

- Retrieve the ArcMC certificate from your ArcMC
- Configure the CDF cluster with ArcMC details
- Retrieve the CDF certificate
- Configure ArcMC

Retrieve the ArcMC certificate:

1. Log into ArcMC.
2. Click **Administration > System Admin > SSL Server Certificate > Generate Certificate.**
3. On the **Enter Certificate Settings** dialog, enter the required settings. In **Hostname**, your certificate settings must match the FQDN of your ArcMC.
4. Click **Generate Certificate.**
5. Once the certificate is generated, click **View Certificate** and copy the full content from `--BEGIN cert to END cert--` to the clipboard.

Configure the CDF cluster:

1. Log in to the CDF management portal.
2. Navigate to suite options: **Suite > Management.**
3. Click the **...** icon under **REFRESH** and Select **Reconfigure.** A new tab will be opened.
4. Scroll down to the Management Center Configuration section and enter values as described for the following:
   - Enter the ArcMC hostname and port 443 (for example, `arcmc.example.com:443`). If ArcMC was installed as a non-root user, enter port 9000 instead.

- **Username/Password:** Enter the username and password of the Transformation Hub.
- **ArcMC certificates:** Paste the text of the generated server certificates you copied to the clipboard as described above.

5. Click **Save**. Web services pods in the cluster will be restarted

Retrieve the CDF certificate:

1. On the initial master node of the cluster, run the following: `<root_installation_folder>/kubernetes/scripts/cdf-updateRE.sh`
2. Copy the contents of this certificate from `--BEGIN cert to END cert--` to the clipboard.

Configure ArcMC:

1. Log in to the ArcMC.
2. Click **Node Management > View All Nodes.**
3. In the navigation bar, click Default, then click **Add Host,** and enter the following values:
   - **Hostname/IP:** FQDN of the Transformation Hub.
   - **Type:** Transformation Hub Containerized (or, if using THNC, select *Non-containerized* instead)
   - **Port**: 38080
   - **Cluster Port:** 443
   - **Cluster Username:** Transformation Hub username
   - **Cluster Password:** Transformation Hub Password
   - **Cluster Certificate:** Paste the contents of this file the CDF certificate you copied earlier.
4. Click **Add**. The Transformation Hub is added as a managed host.

Configuring a SmartConnector as Transformation Hub Producer

The procedure for configuring a SmartConnector as a Transformation Hub producer will depend on whether the connector will be using SSL/TLS.

To configure a SmartConnector as a Transformation Hub producer (not using SSL/TLS):

1. Launch the SmartConnector configuration wizard and proceed through the wizard.
2. On the dialog **Enter the destination parameters**, enter or select appropriate values:

- **Initial Host:**{Kafka broker Host IP 1}:9092, {Kafka broker Host IP 2}:9092, {Kafka broker Host IP 3}:9092
- **Content type:** Select the content type the Connector will be forwarding from the drop-down list.

- **Topic:** Enter one of the following topics, based on the selection in **Content type.** (**Note:** You can mouse over to view topic recommendations.) You may only set 1 topic for the connector.

Content Type	Recommended Topic
Logger/Investigate/Hadoop/3rd parties	th-cef (CEF 0.1 format)
ESM	th-esm
Logger 6.4 or higher/IPv6/investigate	th-other (CEF 1.0 format)

- **Acknowledgement mode:** Default is leader, other options are all and none
- **Use SSL/TLS:** false
- **Use SSL/TLS Authentication:**False

To configure a SmartConnector as a Transformation Hub producer (using SSL/TLS):

1. Configure an SSL/TLS trust store file that contains the Transformation Hub certificate.
2. Download the Transformation Hub certificate to a convenient location, such as `C:\Temp\cluster.crt`.
3. Run the following commands:

```
cd C:\ArcSightSmartConnectors\current\bin
```

```
arcsight agent keytool -importcert -alias eventbrokercaroot -file
C:\Temp\cluster.crt -keystore
C:\ArcSightSmartConnectors\current\jre\lib\security\cacerts -storepass
changeit -store agentcerts
```

Answer `yes` when asked if you want to trust the certificate.

4. Launch the SmartConnector configuration wizard and proceed through the wizard.
5. On the dialog **Enter the destination parameters**, enter or select the following values:

- **Initial Host:**Ports(s): {Kafka broker Host IP 1}:9092, {Kafka broker Host IP 2}:9092, {Kafka broker Host IP 3}:9092
- **Content type:** Select the content type the Connector will be forwarding from the drop-down list.
- **Topic:** Enter one of the following topics, based on the selection in **Content type.** (**Note:** You can mouse over to view topic recommendations.) You may only set 1 topic for the connector.

Content Type	Recommended Topic
Logger/Investigate/Hadoop/3rd parties	th-cef (CEF 0.1 format)
ESM	th-esm
Logger 6.4 or higher/IPv6/investigate	th-other (CEF 1.0 format)

- **Acknowledgement mode:** Default is leader, other options are all and none
- **Use SSL/TLS:** true

- **SSL/TLS Trust Store file:** C:\ArcSightSmartConnectors\current\jre\lib\security\cacerts
- **SSL/TLS Trust Store password:** changeit (or SC password)
- **Use SSL/TLS Authentication:** True
- **SSL/TLS Trust Store file:** C:\ArcSightSmartConnectors\current\jre\lib\security\cacerts
- **SSL/TLS Trust Store password:** changeit (or SC password)
- **SSL/TLS Key password:** Leave blank

Viewing the Transformation Hub certificate

The Transformation Hub certificate can be viewed as follows:

1. Run: `arcsight keytoolgui`
2. Open `C:\ArcSightSmartConnectors\current\jre\lib\security\cacerts`.

Troubleshooting

The following troubleshooting tips may be useful in diagnosing SmartConnector integration issues.

Error Message	Issue
Unable to test connection to Kafka server: [Failed to construct kafka producer]	SmartConnector can't resolve the short or full hostname of the Transformation Hub node(s).
Unable to test connection to Kafka server: [Failed to update metadata after 30000 ms.]	SmartConnector can resolve the short or full hostname of the Transformation Hub node(s) but can't communicate with them because of routing or network issues.
Unable to test connection to Kafka server: [Failed to update metadata after 40 ms.]	You have mistyped the topic name. (Note the lower value in ms than in other messages.)
Destination parameters did not pass the verification with error [; nested exception is: java.net.SocketException: Connection reset]. Do you still want to continue?	If using SSL/TLS, you did not configure the SSL/TLS parameters correctly.

Configuring Logger as a Transformation Hub Producer

The procedure for configuring a Logger as a Transformation Hub producer will depend on whether the Logger will be using SSL/TLS.

To configure a Logger as a Transformation Hub producer (not using SSL/TLS):

1. Log in to Logger.
2. Select **Configuration > Receivers > Add.**

3. In the **Add Receiver** dialog, enter the following:
   - **Name:** Transformation Hub Receiver
   - **Type:** Transformation Hub Receiver
4. Select and edit the Transformation Hub receiver and enter the following parameters:
   - **Transformation Hub host(s) and port:** 192.168.0.4:9092,192.168.0.5:9092,192.168.0.6:9092
   - **Event Topic List:** th-cef (You can enter multiple topics with a comma-separated list.)
   - **Retrieve event from earliest offset:** true
   - **Consumer Group (Logger Pool):** Logger Pool
   - **Use SSL/TLS:** false
   - **Use Client Authentication:** false
   - **Enable:** Checked

To configure a Logger as a Transformation Hub producer (using SSL/TLS):

1. Log in to Logger.
2. Select **Configuration > Receivers > Add.**
3. In the **Add Receiver** dialog, enter the following:
   - **Name:** Transformation Hub Receiver
   - **Type:** Transformation Hub Receiver
4. Select and edit the Transformation Hub receiver and enter the following parameters:
   - **Transformation Hub host(s) and port:** 192.168.0.4:9093,192.168.0.5:9093,192.168.0.6:9093
   - **Event Topic List:** th-cef (You can enter multiple topics with a comma-separated list.)
   - **Retrieve event from earliest offset:** true
   - **Consumer Group (Logger Pool):** Logger Pool
   - **Use SSL/TLS:** true
   - **Use Client Authentication:** true
   - **Enable:** Checked

Troubleshooting

The following troubleshooting tips may be useful in diagnosing Logger integration issues.

Error Message	Issue
IP Address th1.example.com is not a valid address	Use UP addresses in Receiver configuration, not host names.
There was a problem contacting Transformation Hub: Timeout expired while fetching topic metadata, please check the receiver configuration	Logger can't communicate with Transformation Hub because of routing or network issues.
The specified Event Topic (th-other123) is not valid	You have mistyped the topic name.

Note: This process is explained in more detail in the Logger Administrator's Guide, available from the ArcSight support community.

Configuring Logger as a Transformation Hub Consumer

The procedure for configuring a Logger as a Transformation Hub producer will depend on whether the Logger will be using SSL/TLS.

To configure a Logger as a Transformation Hub consumer (not using SSL/TLS):

1. Log in to Logger.
2. Select **Configuration > Receivers > Add.**
3. In the **Add Receiver** dialog, enter the following:
   - **Name:** Enter a unique name for the new receiver.
   - **Type:** Transformation Hub Receiver
4. Select and edit the Transformation Hub receiver and enter the following parameters:
   - **Transformation Hub host(s) and port:** {Kafka broker Host IP 1}:9092, {Kafka broker Host IP 2}:9092, {Kafka broker Host IP 3}:9092
   - **Event Topic List:** th-cef (If needed, enter multiple topics with a comma-separated list.)
   - **Retrieve event from earliest offset:** true
   - **Consumer Group (Logger Pool):** Logger Pool
   - **Use SSL/TLS:** false
   - **Use Client Authentication:** false
   - **Enable:** Checked

To configure a Logger as a Transformation Hub consumer (using SSL/TLS):

1. Log in to Logger.
2. Select **Configuration > Receivers > Add.**

3. In the **Add Receiver** dialog, enter the following:
   - **Name:** Transformation Hub Receiver
   - **Type:** Transformation Hub Receiver
4. Select and edit the Transformation Hub receiver and enter the following parameters:
   - **Transformation Hub host(s) and port:** {Kafka broker Host IP 1}:9092, {Kafka broker Host IP 2}:9092, {Kafka broker Host IP 3}:9092
   - **Event Topic List:** th-cef (You can enter multiple topics with a comma-separated list.)
   - **Retrieve event from earliest offset:** true
   - **Consumer Group (Logger Pool):** Logger Pool
   - **Use SSL/TLS:** true
   - **Use Client Authentication:** true
   - **Enable:** Checked

Troubleshooting

The following troubleshooting tips may be useful in diagnosing Logger integration issues.

Error Message	Issue
IP Address th1.example.com is not a valid address	Use UP addresses in Receiver configuration, not host names.
There was a problem contacting Transformation Hub: Timeout expired while fetching topic metadata, please check the receiver configuration	Logger can't communicate with Transformation Hub because of routing or network issues.
The specified Event Topic (th-<topicname>) is not valid	You have mistyped the topic name.

Note: This process is explained in more detail in the Logger Administrator's Guide, available from the ArcSight support community.

Appendix A: CDF Installer Script `install.sh` Command Line Arguments

Argument	Description
`--auto-configure-firewall`	Flag to indicate whether to auto configure the firewall rules during node deployment. The allowable value is true or false. The default is true.
`--aws-eip`	If the deployment is to AWS, specifies the AWS Elastic IP address. It must be a static IPv4 address.
`--aws-region`	If the deployment is to AWS, specifies the AWS region to use when AWS is the cloud provider.
`--cloud_provider`	Set to a cloud provider type. Allowed values: AWS,AZURE
`--cluster-name`	Specifies the logical name of the cluster.
`--deployment-log-location`	Specifies the absolute path of the folder for placing the log files from deployments.
`--docker-http-proxy`	Proxy settings for Docker. Specify if accessing the Docker hub or Docker registry requires a proxy. By default, the value will be configured from the `http_proxy` environment variable on your system.
`--docker-https-proxy`	Proxy settings for Docker. Specify if accessing the Docker hub or Docker registry requires a proxy. By default, the value will be configured from `https_proxy` environment variable on your system
`--docker-no-proxy`	Specifies the IPv4 addresses or FQDs that do not require proxy settings for Docker. By default, the value will be configured from the `no_proxy` environment variable on your system.
`--enable_fips`	This parameter enables suites to enable and disable FIPS. The expected values are true or false. The default is ***false***.
`--fail-swap-on`	If 'swapping' is enabled, specifies whether to make the kubelet fail to start. Set to true or false. The default is ***true***.
`--flannel-backend-type`	Specifies flannel backend type. Supported values are vxlan and host-gw. The default is host-gw.
`-h, --help`	Lists a help message explaining proper parameter usage

Argument	Description
`--ha-virtual-ip`	A Virtual IP (VIP) is an IP address that is shared by all Master Nodes. The VIP is used for the connection redundancy by providing failover for one machine. Should a Master Node fail, another Master Node takes over the VIP address and responds to requests sent to the VIP. Mandatory for a Multi-Master cluster; not applicable to a single-master cluster The VIP must be resolved (forward and reverse) to the VIP Fully Qualified Domain Name (FQDN)
`--k8s-home`	Specifies the absolute path of the directory for the installation binaries. By default, the Kubernetes installation directory is /opt/arcsight/kubernetes.
`--keepalived-nopreempt`	Specifies whether to enable nopreempt mode for KeepAlived. The allowable value of this parameter is true or false. The default is true and KeepAlived is started in nopreempt mode.
`--keepalived-virtual-router-id`	Specifies the virtual router ID for KEEPALIVED. This virtual router ID is unique for each cluster under the same network segment. All nodes in the same cluster should use the same value, between 0 and 255. The default is 51.
`--kube-dns-hosts`	Specifies the absolute path of the hosts file which used for host name resolution in a non-DNS environment. **Note:** Although this option is supported by the CDF Installer, its use is strongly discouraged to avoid using DNS resolutuion in production environments due to hostname resolution issues and nuances involved in their mitigations.
`--load-balancer-host`	IP address or host name of load balancer used for communication between the Master Nodes. It is required to provide –load-balancer-host or –ha-virtual-ip arguments for a multiple Master Node cluster.
`-m, --metadata`	Specifies the absolute path of the tar.gz suite metadata packages.
`--master-api-ssl-port`	Specifies the https port for the Kubernetes (K8S) API server. The default is 8443.
`--nfs-folder`	Specifies the path to the ITOM core volume.
`--nfs-server`	Address of the NFS host.
`--pod-cidr`	Specifies the private network address range for the Kubernetes pods. Default is 172.16.0.0/16. The minimum useful network prefix is /24. The maximum useful network prefix is /8. This must not overlap with any IP ranges assigned to services (see –service-cidr parameter below) in Kubernetes. The default is 172.16.0.0/16. For the default and allowable values see the CDF Planning Guide.
`--pod-cidr-subnetlen`	Specifies the size of the subnet allocated to each host for pod network addresses. For the default and the allowable values see the CDF Planning Guide.

Argument	Description
`--registry_orgname`	The organization inside the public Docker registry name where suite images are located. Not mandatory. Choose one of the following: • Specify your own organization name (such as your company name). For example: `--registry-orgname=Mycompany.` • Skip this parameter. A default internal registry will be created under the default name HPESWITOM.
`--runtime-home`	Specifies the absolute path for placing Kubernetes runtime data. By default, the runtime data directory is $K8S_HOME/data.
`--service-cidr`	Kubernetes service IP range. Default is 172.30.78.0/24. Must not overlap the POD_CIDR range. Specifies the network address for the Kubernetes services. The minimum useful network prefix is /27 and the maximum network prefix is /12. If SERVICE_CIDR is not specified, then the default value is 172.17.17.0/24. This must not overlap with any IP ranges assigned to nodes for pods. See `--pod-cidr.`
`--skip-check-on-node-lost`	Option used to skip the time synchronization check if the node is lost. The default is true.
`--skip-warning`	Option used to skip the warnings in precheck when installing the Initial master Node. Set to true or false. The default is false.
`--system-group-id`	The group ID exposed on server, default is 1999.
`--system-user-id`	The user ID exposed on server, default is 1999.
`--thinpool-device`	Specifies the path to the Docker devicemapper, which must be in the `/dev/mapper/` directory. For example: `/dev/mapper/docker-thinpool`
`--tmp-folder`	Specifies the absolute path of the temporary folder for placing temporary files. The default temporary folder is /tmp.

Send Documentation Feedback

If you have comments about this document, you can contact the documentation team by email. If an email client is configured on this computer, click the link above and an email window opens with the following information in the subject line:

Feedback on Deployment Guide (Investigate 2.4.0)

Just add your feedback to the email and click send.

If no email client is available, copy the information above to a new message in a web mail client, and send your feedback to arcsight_doc@microfocus.com.

We appreciate your feedback!