Micro Focus Security ArcSight Investigate

Software Version: 2.30

Deployment Guide

Document Release Date: October 30, 2018
Software Release Date: October 2018

Legal Notices

Warranty

The only warranties for products and services of Micro Focus and its affiliates and licensors (“Micro Focus”) are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Micro Focus shall not be liable for technical or editorial errors or omissions contained herein. The information contained herein is subject to change without notice.

Restricted Rights Legend

Confidential computer software. Except as specifically indicated otherwise, a valid license from Micro Focus is required for possession, use or copying. Consistent with FAR 12.211 and 12.212, Commercial Computer Software, Computer Software Documentation, and Technical Data for Commercial Items are licensed to the U.S. Government under vendor's standard commercial license.

Copyright Notice

Trademark Notices

Adobe™ is a trademark of Adobe Systems Incorporated.

Microsoft® and Windows® are U.S. registered trademarks of Microsoft Corporation.

UNIX® is a registered trademark of The Open Group.

Support

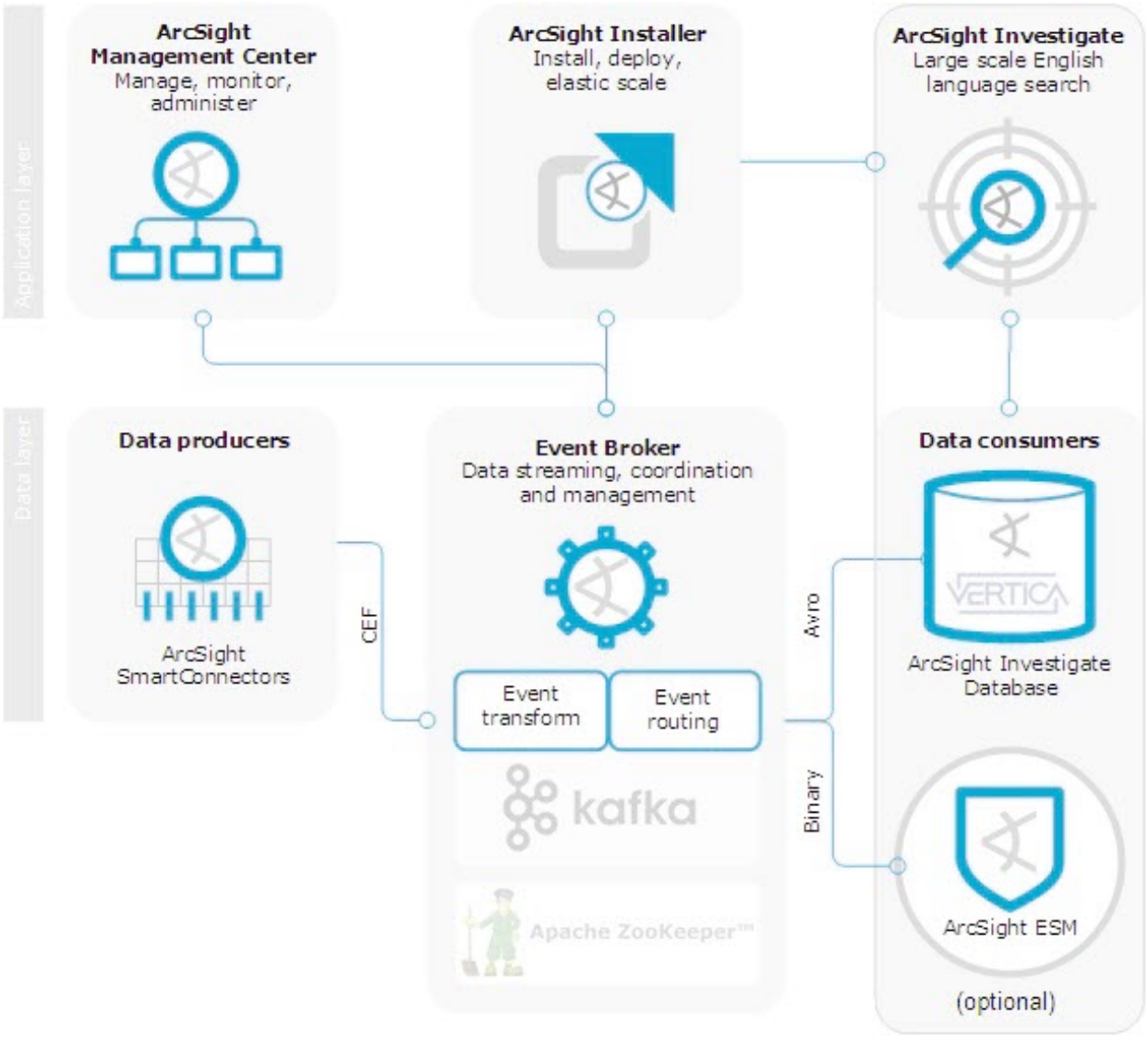
Contact Information

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: Understanding ArcSight Investigate Architecture and Deployment

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.

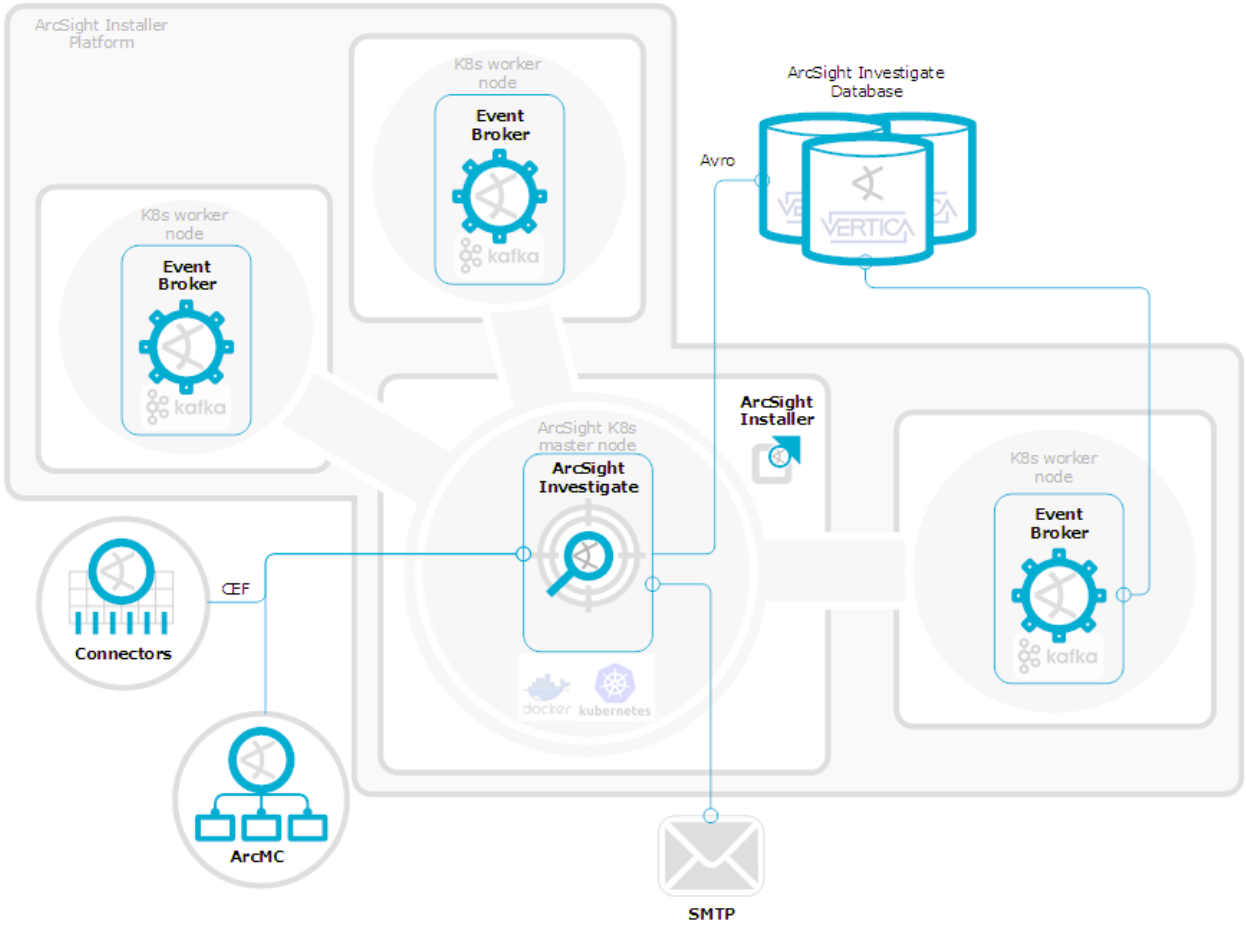


Component	Description
ArcSight Investigate	High-capacity data management, search, and analysis web application
ArcSight Investigate Vertica database	Investigate analytic database powered by Vertica Install the Vertica database separately.
ArcSight Installer	Web application for deploying and configuring Investigate components, including Event Broker A Kubernetes cluster manages the components. The master node hosts the ArcSight Installer and Investigate web applications, and the worker nodes host Event Broker.
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 Event Broker. You can subscribe to data that Event Broker manages with Investigate, ArcSight Deployment Platform (ADP), Logger, ArcSight ESM, Apache HDFS, or your own third-party consumer.
Event Broker	Centralizes event processing Event Broker 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 Event Broker topics and is sold as part of ADP.

Understanding Deployment Architecture

ArcSight Investigate runs in Docker containers that Kubernetes manages. The ArcSight Installer deploys the Kubernetes nodes and then deploys Investigate on a node within the Kubernetes cluster. The Investigate node can be a master or worker node within the cluster. For production deployments, Micro Focus recommends that you deploy three master nodes and three worker nodes. ArcSight also supports a configuration with a single master node and three worker nodes. In this case, install Event Broker on the three worker nodes and Investigate on the master node. For more information, contact your ArcSight Technical Specialist.

The following image represents a typical deployment with one master node and three worker nodes:



Component	Host	Contents/Description
ArcSight Installer	Master node and each worker node	ArcSight Installer web application
Kubernetes master node	Either one or three VMs or physical servers	• Multi-master deployment: three or more servers • Single master deployment: one server • Arcsight Installer on all servers • For single master deployment, Investigate on the master server
Kubernetes worker nodes	Three VMs or physical servers	Three Kubernetes nodes Each Event Broker instance requires a worker node. For multi-master deployments, Investigate requires one worker node.
Vertica database	Three physical servers	One Vertica database cluster with three nodes
ArcSightSmartConnectors	Stand-alone or part of ArcSight Management Center (ArcMC)	Normalizes event data from network devices and formats them in Common Event Format (CEF)
ArcSight Management Console (ArcMC)	Separate installation	Provides runtime management of Event Broker topics
SMTP server	Separate installation	Allows Investigate to send notification messages to users

Understanding the Order for Deploying Investigate

Before you deploy Investigate, you must install and configure the Vertica database and ArcSight Installer, and then use ArcSight Installer to deploy Event Broker.

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

1. Obtain the Investigate images.

   For more information, see Obtaining Investigate Images Using the Download Script and Downloading Images From an Offline Location.
2. Obtain the Event Broker images.

   For more information, see the Event Broker Deployment Guide.
3. Configure the Vertica server and install the database.

   For more information, see Configuring the Vertica Server and Installing the Database.
4. Ensure that Event Broker and Investigate each have a dedicated server.

   If other applications run on the same servers as Event Broker and Investigate, you might experience performance problems.
5. Install or upgrade to ArcSight Installer 1.50 and Event Broker 2.21, and then configure Event Broker.

   For more information, see the Event Broker Deployment Guide.
6. Deploy or upgrade to Investigate 2.30.

   For more information about deploying Investigate, see Deploying ArcSight Investigate.

   For more information about upgrading to Investigate 2.30, see Upgrading ArcSight Investigate.
7. Configure Investigate and the Investigate components.

   For more information, see Configuring ArcSight Investigate and Components.

Planning Your Deployment

Before you deploy Investigate, review the following sections:

- Understanding TLS Requirements
- Understanding Network Requirements
- Understanding Encryption Modes

Understanding 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.

Understanding 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.

Understanding Encryption Modes

Before deployment, determine the encryption mode you want to use to encrypt communications between ArcSight components. Before you connect components to Event Broker, configure them to use the desired encryption mode. Event Broker and the components that connect to it must use the same encryption mode. Changing encryption modes after you deploy Event Broker requires system down time. For information about changing the encryption mode after you deploy Event Broker, see the Event Broker Administrator's Guide.

Product	Open ports	Supported encryption modes	Guidance
ArcSight Management Center (ArcMC)	38080	• TLS • FIPS • ClientAuth	Install ArcMC before you deploy Investigate and Event Broker. For more information, see the ArcMC Administrator's Guide.
ArcSight SmartConnectors	9093	• TLS • FIPS • ClientAuth	You can install and run SmartConnectors before you deploy Investigate and Event Broker. FIPS mode is not supported between Connector version 7.5 and Event Broker. TLS and ClientAuth are the only encryption methods that are supported between SmartConnector version 7.5 and Event Broker. For more information, see the SmartConnector User Guide.

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 Event Broker. 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. For more information, see the ESM Installation Guide and the ESM Administrator's Guide.
ArcSight Logger (optional)	9093	• TLS • FIPS • ClientAuth	You can install and run Logger before you deploy Investigate and Event Broker. For more information, see the Logger Administrator's Guide.

Chapter 2: Understanding System Requirements

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

Understanding Supported Operating Systems

ArcSight Investigate supports the following operating systems:

Version	Component	Operating system
Investigate 2.30	Investigate	CentOS/RHEL 7.5 CentOS/RHEL 7.4, with kernel version 3.10.0-693.21.1.el7.x86_64 or later
	Vertica 8.1.1-3 database	RHEL or CentOS 7.3 **Note:** Investigate supports using Vertica on a host with a Linux Logical Volume Manager (LVM) formatted disk.

Understanding 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.

Understanding Supported Components

Investigate is compatible with the following ArcSight components:

Component	Version
ArcSight Installer	1.5
ArcSight Event Broker	2.21
ArcSight SmartConnectors	7.5 and later
ArcSight Management Center (ArcMC)	2.80
ArcSight Logger	6.6 and later
ArcSight ESM	6.11
DNS Analytics feature in Investigate	This feature requires MS-DNS Connector version 7.8.0.8070 or later.

Chapter 3: Preparing Your Environment for Deployment

This chapter provides information about preparing your environment to deploy ArcSight Investigate.

Before you deploy Investigate, review Understanding the Order for Deploying Investigate, Planning Your Deployment, and Understanding System Requirements.

Understanding Sizing Needs

Use the following guidelines to provision the servers (or VMs) that you are using for the deployment. The guidelines are based on a default setup.

For tailored sizing recommendations for a production environment, contact ArcSight Customer Support.

For supported platforms and operating systems, see Understanding System Requirements.

Component	Nodes	Required resources	Required ports
ArcSight Investigate + Event Broker	1 master 3 worker	• One CPU with 24 cores • 32 GB RAM • 8 TB disk space • Java (OpenJDK) 1.8.0_121 or later • Method for obtaining Docker containers, either via Internet (or proxy) or other internal method • 10 GigE network **Note:** If you choose to deploy ArcSight Investigate on a worker node, the `nginx` reverse proxy that is used to connect to Investigate is always deployed on the master node. Therefore, no matter where you deploy Investigate in a Kubernetes cluster, always access Investigate using the host/IP of the master node.	• Kubernetes: 2379, 2380, 4001, 4194, 5000, 5443, 8080, 8088, 8200, 8285, 8443, 10248-10252, 10255, 30001 • Network File System (NFS): 111, 2049, 20048, 37189 • Investigate: 5443, 21085, 30001 For required Event Broker ports, see the Event Broker Deployment Guide.
Vertica database	3	• Two CPUs with 24 cores each • 128 GB RAM • 8 TB disk space • 10 GigE network minimum (dual recommended) You must install the Vertica database on the same sub-network as the Investigate master and worker nodes. Micro Focus recommends installing Vertica on a dedicated physical server. For example, an HPE Proliant G9 or similar. Vertica performs better on a physical server than in a virtualized environment because of the overhead and resource constraints that the virtualization software imposes. For more information, see the Vertica Analytics Platform Version 8.1.x documentation.	5433
ArcSight Management Center (ArcMC)	1	• One CPU quad-core • 16 GB RAM • 50 GB of free disk space For information about deploying ArcMC, see the ArcSight Management Center Administrator's Guide.	
SmartConnectors	1	SmartConnector version 7.5 (can be standalone or managed by ArcMC) For information about deploying SmartConnectors, see the SmartConnector User Guide.	

Understanding Network Ranges

By default, ArcSight Installer uses the following network ranges:

- 172.16.0.0/16

  This sub-network contains 65,536 addresses for the operation of Kubernetes pods with containers. Each pod operates with the /24 sub-network.
- 172.30.78.0/24

  This sub-network contains 256 addresses for the operation of Kubernetes services, including the internal Kubernetes DNS service on pod 172.30.78.78.

Caution: If your network already uses the default network ranges, the deployment might fail or you might experience issues after deployment.

If these address ranges are occupied or inaccessible, utilize another address range by making corresponding changes to the `POD_CIDR`, `SERVICE_CIDR`, and `DNS_SVC_IP` parameters in the `./<Path_to_Secure_Location_on_Master_Node>/arcsight-installer-<Version>/arcsight-installer-master.sh` script.

Understanding Firewall Requirements

Ensure that the ports listed in Understanding Sizing Needs are free and available for firewall configuration.

If the `firewalld.service` is running, ArcSight Installer configures firewall settings during setup on Kubernetes nodes.

Vertica requires several ports to be open on the local network. Vertica does not recommend placing a firewall between nodes. If you must use a firewall between nodes, ensure that the following ports are available:

Port	Protocol	Service	Notes
7	TCP	Management Console	Required by Management Console to discover Vertica nodes
22	TCP	sshd	Required by Administration Tools and the Management Console Cluster Installation wizard
5433	TCP	Vertica	Vertica client (such as vsql, ODBC, JDBC) port

Port	Protocol	Service	Notes
5434	TCP	Vertica	Intra- and inter-cluster communication Vertica opens the Vertica client port +1 (5434 by default) for intra-cluster communication, such as during a plan. If the port +1 from the default client port is not available, then Vertica opens a random port for intra-cluster communication.
5433	UDP	Vertica	Vertica spread monitoring
5444	TCP	Management Console	Management Console-to-node and node-to-node (agent) communications port
5450	TCP	Management Console	Used to connect to Management Console from a web browser and allow communication from nodes to the Management Console application/web server
4803	TCP	Spread	Client connections
4803 and 4804	UDP	Spread	Daemon to daemon connections
6543	UDP	Spread	Monitor to daemon connection

Configuring Proxy Settings

If your organization uses a proxy server, comment out proxy data in the `/etc/profile.d/proxy.sh` file on all nodes, and then add your proxy data to the `~/.bashrc` file as shown in the following example:

```
export http_proxy=http://<Proxy_Server>:8080/

export https_proxy=http://<Proxy_Server>:8080/

export HTTP_PROXY=http://<Proxy_Server>:8080/

export HTTPS_PROXY=http://<Proxy_Server>:8080/

export no_proxy="<Master_IP>,<Worker-1_IP>,<Worker-2_IP>,<Worker-3_
IP>,localhost,<Domain>"

export NO_PROXY="<Master_IP>,<Worker-1_IP>,<Worker-2_IP>,<Worker-3_
IP>,localhost,<Domain>"
```

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

Before you install the Vertica database, configure the server using the procedure in this section as a guideline. The server configuration that is described here is based on an HPE ProLiant DL380 Gen9 server with 128 GB memory. For more information, see Configuring the HPE Proliant DL380 Gen9 24-SFF CTO Server as a Vertica Node.

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

To configure the Vertica server:

1. Provision the server with an ext4 partition and at least 2 GB of swap space, running on RHEL 7.3 or CentOS 7.3.

   Note: Investigate supports using Vertica on a host with a Linux Logical Volume Manager (LVM) formatted disk.

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
`dirty_ratio = 8`	Sets the number of pages at which a process that is generating disk writes will start writing out dirty data For more information, see Tuning Linux Dirty Data Parameters for Vertica.
`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.
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.

Generating the SSH Key Pair

Before you install the Vertica database, generate a key pair on node 1 and then copy the public key to all nodes in the cluster, including node 1. This enables password-less SSH access from the node 1 server to all of the other node servers in the cluster.

Note: You must repeat the authentication process for each node in the cluster.

To generate the SSH key pair:

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
   ```

Installing Vertica

After you configure the Vertica server and generate the SSH key pair, install the Vertica database.

To install Vertica:

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

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

2. Copy `arcsight-vertica-installer_2.30.0-1.tar.gz` and `arcsight-vertica-installer_2.30.0-1.tar.gz.md5` to `vertica-install-DIR`.

3. Verify that the tarball matches the MD5 checksum:

   ```
   cd vertica-install-DIR
   md5sum arcsight-vertica-installer_2.30.0-1.tar.gz
   cat arcsight-vertica-installer_2.30.0-1.md5
   ```

4. Extract the `.tar` file:

   ```
   tar xvfz arcsight-vertica-installer_2.30.0-1.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.
`db_retention_day`	Used for the data retention policy For more information, see Enabling the Data Retention Policy on the Vertica Cluster.

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.
7. Create the schema:

   ```
   ./vertica_installer create-schema
   ```
8. In order to start the Kafka scheduler after you create it, you must disable the scheduler SSL:

   ```
   ./sched_ssl_setup --disable-ssl
   ```
9. Create the Kafka scheduler:

   ```
   ./kafka_scheduler create <Event_Broker_Node_1_IP>:9092,<Event_Broker_Node_2_IP>:9092,<Event_Broker_Node_3_IP>:9092
   ```

   For a list of options that you can specify when installing the scheduler, see Kafka Scheduler Options.
10. 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 5: Deploying ArcSight Investigate

Before you deploy Investigate, install the Vertica database and ArcSight installer, and then use ArcSight Installer to deploy Event Broker. For information about installing the Vertica database, see Configuring the Vertica Server and Installing the Database. For information about installing ArcSight Installer and deploying Event Broker, see the Event Broker Deployment Guide.

After you install the database and installer and deploy Event Broker, use the installer to deploy Investigate. If the `firewalld` service is running, the installer also configures firewall settings on the Kubernetes master and worker nodes.

Before you start the Investigate deployment, use one of the following methods to download the docker images:

- Pull the images from docker.com using the `download_images` script that is packaged with ArcSight Installer.

  For more information, see Obtaining Investigate Images Using the Download Script.
- Download a `.tar` file from the Software Licenses and Downloads portal.

  For more information, see Downloading Images From an Offline Location.

If you choose a multi-master deployment for Event Broker, Micro Focus recommends that you deploy Investigate on a worker node. You can deploy Investigate on a worker node that is dedicated to Investigate or on a worker node where Event Broker pods are running.

Using ArcSight Installer to Label Nodes

A typical deployment consists of three Kubernetes master nodes and three Kubernetes worker nodes. Investigate also supports a configuration with a single Kubernetes master node and three Kubernetes worker nodes. If you choose to use a single master node, deploy Event Broker on the three worker nodes and Investigate on the master node.

You can add additional worker nodes to extend the Kafka cluster nodes. After you add the worker nodes, you can use labels to assign specific pods to them. For more information, see the Event Broker Deployment Guide.

To label nodes:

1. Log in to the ArcSight Installer web application:

   `https://<Master_FQDN>:5443` or `http://<Virtual_IP>` if the cluster supports multi-master
2. In the left navigation pane, click **Node Management**.
3. Click **+** next to the node to which you want to deploy Investigate.
4. Click **Investigate**, and then click **Add**.

Obtaining Investigate Images Using the Download Script

You can use a script that is packaged with ArcSight Installer to download the Investigate images.

To use the download script to obtain Investigate images:

1. Enter the following commands:

   ```
   cd /opt/arcsight/kubernetes/scripts
   ./downloadimages.sh --suite investigate --registry docker --org arcsightsecurity
   ```
2. Select the 2.30 version.
3. Upload the images to the local Docker registry:

   ```
   ./uploadimages.sh --suite investigate
   ```

Downloading Images From an Offline Location

If necessary for your environment, you can download the Investigate images from an offline location.

To download the images from an offline location:

1. Download `arcsight-investigate-2.30.12.tar` from the Software Licenses and Downloads portal.
2. Verify the download.

   Micro Focus provides a digital public key that allows you to verify that signed software is from Micro Focus and has not been manipulated by a third party. For more information, see the Software Licenses and Downloads portal.
3. Place `arcsight-investigate-2.30.12.tar` in `<Offline_Install_Directory>`, where `Offline_Install_Directory` is a local directory where you can access the `.tar` file.
4. Upload the Investigate images to make them available to the Installer:

   ```
   <Offline_Install_Directory>
   tar xvf arcsight-investigate-2.30.12.tar
   ```

   The `./investigate` directory contains the Investigate images.

   ```
   cd /opt/arcsight/kubernetes/scripts
   ./uploadimages.sh --suite investigate --dir <Offline_Install_
   Directory>/investigate
   ```

Deploying and Undeploying Investigate Images

Deploy the Investigate images from the **Deployment** page of the ArcSight Installer web application. You can also use the installer to undeploy Investigate.

To deploy the Investigate images:

1. Log in to the ArcSight Installer web application:

   `https://<Master_FQDN>:5443` or `http://<Virtual_IP>` if the cluster supports multi-master
2. In the the left navigation pane, click **Deployment**.
3. In the ArcSight Investigate row, click **Deploy**.

4. Select version 2.30, and then click **Deploy**.

   The installer displays a progress indicator and notification when it starts the deployment.
5. When the deployment is complete, configure the Vertica database connection.

   For more information, see Configuring the Vertica Database Connection.
6. Check the pod status using one of the following methods:
   - On the **Deployment** page, click the ellipses in the **Details** column.
   - Use SSH to connect a master Kubernetes node, and then run the following commands:

     For all pods: `kubectl get pods --all-namespaces`.

     For Investigate pods: `# kubectl get pods --all-namespaces | grep investigate`

     If the pods are healthy, they should be in Running status. It might take several minutes for all pods to start running.

To undeploy Investigate:

1. In the left navigation pane of ArcSight Installer, click **Deployment**.
2. In the ArcSight Investigate row, click **Undeploy**.
3. From the Kubernetes master node, run the following command:

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

   If the command does not return pod information, the undeployment was successful.

Chapter 6: 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.

Establishing 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://<Virtual_IP>`.
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.

Configuring the Vertica Database Connection

Use the ArcSight Installer to configure the connection to the Vertica database. If you undeploy Investigate and then redeploy, you must reconfigure the database connection. Each time you change the connection, the search 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://<Virtual_IP>:5443` if you deployed in multi-master mode
2. From the left navigation, select **Configuration** > **ArcSight Investigate** > **Vertica**, and then provide the following information:
   - Vertica host

     You can specify any Vertica node IP address, but only specify one address.
   - Vertica search user name that you defined when you installed Vertica
   - Vertica search user password that you created when you installed Vertica
   - Vertica database name

     The name is hard coded to Investigate. You cannot change the name.

Configuring Event Broker for ArcSight Investigate

After you deploy Event Broker, you must use the ArcSight Installer to configure the Event Broker data pipeline for Investigate.

Event Broker consumers need a signed certificate from Event Broker to establish secure communication with Investigate. For more information, see the Event Broker Deployment Guide.

To configure Event Broker for Investigate:

1. In the left navigation pane of ArcSight Installer, select **Configuration**.
2. Select **ArcSight Event Broker**.

3. Select **Replicas**.
4. Click **+** next to **Transforming String Processor**, and then click **Save**.
   The number changes from 0 to 1.

Configuring the SMTP Server

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

To configure the SMTP server:

1. In ArcSight Installer, select **Configuration** > **ArcSight Installer**, and then click **User Management**.
2. Specify the following information, and then click **Save**:
   - SMTP host
   - SMTP port
   - SMTP user name
   - SMTP password
   - Sender address

Configuring Session and Search Settings

You can configure the following properties in ArcSight Installer:

- Session timeout
  You can configure the amount of time that a user session runs before the user must log in again. The default session timeout is 60 minutes.
- 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. Select **Configuration** > **ArcSight Investigate**.
2. On the **General** tab, specify the following:
   - In the **Session timeout** field, specify the maximum time (in seconds) that you want a session to run.
   - In the **Search query timeout** field, specify the maximum time (in minutes) that you want a search query to run.

Configuring Vertica SSL

The ArcSight Installer contains a script, `/opt/arcsight/installer/k8s/master/cert-utils.sh`, that enables you to generate a certificate that is signed by the root certificate authority. Kubernetes and all modules use the certificate.

To configure Vertica SSL:

1. Connect to the master node and run `./cert-utils.sh generate-certificate vertica => script`.

   The script produces `vertica.key` and `vertica.crt`.
2. Copy `vertica.key` and `vertica.crt` to all Vertica nodes.
3. Copy the certificate to all Vertica nodes.

   By default, the certificate is located in `/opt/arcsight/kuberntes/ssl/ca.crt`.
4. On each node, start the Vertica administration tool:

   ```
   su - -c adminTools dbadmin => vertica
   ```
5. From the main menu in the administration tool, select **Configuration Menu**, and then click **OK**.
6. Select **Distribute Config Files**, and then click **OK**.
7. Select **SSL Keys**, and then click **OK**.
8. Select the database on which you want to distribute the files, and then click **OK**.
9. Add the file locations for the `vertica.crt`, `vertica.key`, and `ca.crt` files, and then click **OK** to distribute the files.

Configuring TLS on Vertica

The Investigate components use encrypted communication with the Transport Layer Security (TLS) cryptographic protocol. For successful communication between components, you must distribute the key and certificate files on all Vertica nodes and enable TLS on Vertica.

Before you can enable TLS, you must have the following files:

- Valid digital certificate signed by a certificate authority (CA), including a server certificate file (`server.crt`) and a root certificate file (`root.crt`)
- Private key file (`server.key`)

Note: Vertica does not need to be running when you distribute the key and certificate files.

To configure TLS:

1. Copy the `.crt` and `.key` files to one of the Vertica nodes.
2. Start the Vertica Administration Tools.

   For information about using the Administration Tools, see the Vertica documentation.
3. From the Main Menu, select **Configuration Menu** and click **OK**.
4. Select **Distribute Config Files** and click **OK**.
5. Select **SSL Keys** and click **OK**.
6. Select the database on which you want to distribute the files and click **OK**.
7. Modify the file path to the location where you copied the `server.crt`, `root.crt`, and `server.key` files and click **OK**.
8. From the Main Menu, select **Connect to Database** and click **OK**.
9. When prompted, enter the database password.
10. Run the following command, and then restart the database:

```
ALTER DATABASE mydb SET EnableSSL = 1;
```

Enabling the Data Retention Policy on the Vertica Cluster

If you enable data retention, you can purge Vertica data that is older than the retention period. The retention period can be from 1 to 366 days. The data retention policy is based on calendar days.

The default data retention period is 90 days. If you run the data retention script on 6/30/2018 and the `db_retention_days` property is set to 90, then data older than 04/01/2018 will be deleted. You can purge data in real time or by using a scheduled cron job.

Note: Vertica saves events based on the device timestamp (`deviceReceiptTime` field), which might not correlate with the current date.

If you need to retain data for more than 366 days or the Vertica node is in Recovery status, do not enable data retention.

To enable data retention:

1. Back up Vertica data.
   For more information, see Backing Up the Vertica Database.
2. Run the following commands:
   ```
   # cd <Vertica_Install_Directory>/scripts/
   # ./retention_policy_util.sh -h
   ```
3. Run the following command:
   ```
   # vi ../config/vertica_user.properties
   ```
4. Uncomment `#db_retention_days=90`.

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

   ```
   # ./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.

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

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

To purge Vertica data:

1. 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 *****
   ```

2. To use a cron job to purge data, enter the following command:

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

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

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

Chapter 7: Upgrading ArcSight Investigate

ArcSight supports an upgrade from Investigate version 2.20 to version 2.30.

Preparing to Upgrade

Before you upgrade Investigate, complete the following tasks:

- Make a configuration backup.
  For more information, see Backing Up and Restoring the Vertica Database.
- Ensure that you are running a supported operating system.
  For information about supported operating systems, see Understanding Supported Operating Systems.
- Upgrade to the latest version of ArcSight Installer.
  For more information, see the Event Broker Deployment Guide.
- Upgrade to the latest version of Event Broker.
  For more information, see the Event Broker Deployment Guide.

Backing Up and Restoring Lookup Lists

If you are using a lookup list and do not have the original CSV file, you must back up the lookup list before you upgrade in order to migrate the IP and MAC addresses from string data type to ByteArray data type. After you upgrade all Investigate components, you must restore the lookup list.

To back up and restore a lookup list:

1. Create a fieldset and select the lookup list fields.
2. To the left of the `in list` operator in the search query, remove all unused fields.
3. In the **Search** field, use the `in list` operator to specify a field, and then select the lookup list.
   For example, `Source Address in list l1_srcAddr`

4. Add an inline filter for `Source Address`.

   For example, `Source Address <hide duplicates> <empty value>`

   Note: This is the same field that is specified to the left of the `in list` operator in the search query.

5. Run the search and then export the results from the **Events** table to the CSV file.
6. After upgrading to Investigate 2.30 and before uploading the CSV file, edit the CSV file as follows:
   - Remove the field that is to the left of the query.
   - Edit the column headers so that they include only the lookup list field names.
7. After you upgrade all of the Investigate components, upload the modified CSV file to update the lookup list data type.

Upgrading the Vertica Installer

Upgrading the Vertica Installer allows you to upgrade to the latest version of Investigate.

To upgrade the Vertica installer:

1. Download the Investigate 2.30 Vertica installer `tar.gz` and `md5` files to a temporary location on the primary Vertica server.

   Note: The primary Vertica server is the server on which you ran the Vertica installation scripts to initially set up the cluster.

2. Check the `md5sum` of the `tar.gz` file and `cat` the `md5` file to ensure that they match.
3. Untar the `tar.gz` file.
4. Run the `investigate_upgrade` script as the root user.

   Note: The script assumes that the existing Vertica utilities were installed under `/root/install-vertica`. If that location does not exist, the script fails.

   The script adds the data retention scripts and updates the `vertica.properties` file with new configuration parameters, but does not enable data retention. To enable data retention, see Enabling the Data Retention Policy on the Vertica Cluster.

5. Delete the directories where you untarred the `tar.gz` file.

Upgrading Vertica

Before you upgrade Vertica, resolve any error messages in `vertica.log` and ensure that no other applications are installed on the Vertica server.

The upgrade results in the following changes:

- The upgrade creates two files from the `vertica.properties` file: `/opt/install-vertica/config/vertica_user.properties` and `/opt/install-vertica/config/vertica_credentials.properties`. The upgrade moves properties to the new files as follows:

Old file	New file	Properties
`vertica.properties`	`config/vertica_user.properties`	`hosts=localhost` `license=<Your_License_File>.dat`
`vertica.properties`	`config/vertica_credentials.properties`	
`config/sched.properties`	`config/vertica_credentials.properties`	

- After the upgrade, the string IP and MAC address columns in the `investigation.events` table do not receive new data during ingestion. Existing values in those columns are not changed.
- In the `investigation.version_metadata` table, the upgrade adds two parameters: `schemaVersion='4.4.0'` and `installerVersion='2.30.0'`.
- The upgrade replaces old files, with the exception of files in the `config` directory of `/root/install-vertica`, with new files from `/tmp/upgrade-vertica`.
- In the `target_columns` column of the `investigation_scheduler.stream_microbatches` table, the upgrade removes IP and MAC column names that do not end with "Bin."

Note: The upgrade stops the Kafka scheduler. After the upgrade is complete, you must disable SSL on the scheduler and then restart it. For more information, see the procedure that follows.

To upgrade Vertica:

1. On the Vertica cluster node 1 server, create a folder for the Vertica installer scripts:
   `mkdir new-vertica-install-DIR`
2. Copy the Vertica installer scripts to the folder that you created:
   `arcsight-vertica-installer_2.30.0-1.tar.gz` and `arcsight-vertica-installer_2.30.0-1.md5` to `new-vertica-install-DIR`

3. Verify that the tarball matches the MD5 checksum:

   ```
   cd new-vertica-install-DIR
   md5sum arcsight-vertica-installer_2.30.0-1.tar.gz
   cat arcsight-vertica-installer_2.30.0-1.md5
   ```

4. Extract the `.tar` file:

   ```
   tar xvfz arcsight-vertica-installer_2.30.0-1.tar.gz
   ```

   Note: The original directory for the Vertica installer script is `install-vertica`.

5. Upgrade Vertica:

   ```
   ./investigate_upgrade -c upgrade-investigate
   ```

   The output is similar to the following:

   ```
   Upgrade related changes cannot be rolled back, do you want to continue
   with the upgrade (Y/N):Y
   Starting upgrade...
   ****************** Start of Investigate Upgrade ***********************
   Enter previous installed location (/root/install-vertica):/opt/install-
   vertica
   Checking all Vertica nodes are UP
   All Vertica nodes are UP
   ********************* Start of Investigate Upgrade to 2.30.0 ***********
   Pre Upgrade check for 2.30.0
   Current Investigate version is: 2.20.0
   Investigate will be upgraded to 2.30.0
   Updating privileges for search user: search
   Upgrading script and config files.
   Creating backup directory: /opt/install-vertica/oldVersion
   Backing up: /opt/install-vertica/vertica.properties
   Backing up: /opt/install-vertica/data
   Backing up: /opt/install-vertica/upgrade
   Backing up: /opt/install-vertica/scripts
   Backing up: /opt/install-vertica/resources
   Backing up: /opt/install-vertica/vertica_installer
   Backing up: /opt/install-vertica/vertica_upgrade.py
   Backing up: /opt/install-vertica/investigate_upgrade
   Backing up: /opt/install-vertica/kafka_scheduler
   Upgrading: /opt/install-vertica/vertica.properties
   Upgrading: /opt/install-vertica/lib
   ```

```
Upgrading: /opt/install-vertica/data
Upgrading: /opt/install-vertica/arcsight-vertica-installer_master-107.md5
Upgrading: /opt/install-vertica/upgrade
Upgrading: /opt/install-vertica/scripts
Upgrading: /opt/install-vertica/resources
Upgrading: /opt/install-vertica/vertica_installer
Upgrading: /opt/install-vertica/arcsight-vertica-installer_master-107.tar
Upgrading: /opt/install-vertica/vertica-upgrade.log
Upgrading: /opt/install-vertica/vertica_upgrade.py
Upgrading: /opt/install-vertica/vertica_ssl_setup
Upgrading: /opt/install-vertica/investigate_upgrade
Upgrading: /opt/install-vertica/kafka_scheduler
Upgrading: /opt/install-vertica/sched_ssl_setup
************ Investigate Upgraded Complete. Version is 2.30.0
******************
```

6. In order to start the Kafka scheduler after the upgrade completes, you must first disable the scheduler SSL:

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

7. Start the Kafka scheduler:

   ```
   ./kafka_scheduler start
   ```

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

   ```
   ./kafka_scheduler status
   ./kafka_scheduler events
   ./kafka_scheduler messages
   ```

9. At the end of the `config/vertica_user.properties file`, add `#db_retention_days=90`.

Upgrading Investigate

The Investigate upgrade process supports only an offline upgrade.

To upgrade Investigate:

1. Download the Investigate offline upgrade file (`.tar.gz`) from Micro Focus Software Support.

   Note: Investigate documentation is not included in your download package. You can access documentation on the ArcSight Documentation page.

2. Verify the download.

   Micro Focus provides a digital public key that allows you to verify that signed software is from Micro Focus and has not been manipulated by a third party. For more information, see the Software Licenses and Downloads portal.
3. Extract the offline upgrade file to `/opt/arcsight/upgrade/investigate-2.30`.
4. Upload the images:

   ```
   cd /opt/arcsight/kubernetes/scripts
   ./uploadimages.sh --suite investigate --dir
   /opt/arcsight/upgrade/investigate-2.30
   ```
5. Log in to the ArcSight Installer web application:

   `https://<Master_FQDN>:5443` or `http://<Virtual_IP>` if the cluster supports multi-master
6. Click **Node Management** and ensure that worker nodes are in Ready status.
7. On the **Deployment** page, click **Upgrade** for Investigate.
8. Configure Event Broker for Investigate:
   - On the **Configuration** page, select **ArcSight Event Broker**.
   - Select **Replicas**.
   - Click **+** next to **Transforming String Processor**, and then click **Save**.
9. Use one of the following methods to check the pod status:
   - On the **Deployment** page, click the ellipses in the **Details** column.
   - Use SSH to connect a master Kubernetes node, and then run the command `kubectl get pods --all-namespaces`. If the pods are healthy, they should be in Running status.

     It might take several minutes for all pods to start running.

Migrating Investigate Search Components

After you upgrade Investigate, you must migrate lookup lists, existing searches, and dashboard charts.

To migrate Investigate search components:

1. Restore the lookup list backup that you created in Backing Up and Restoring Lookup Lists .
2. Change the time of existing searches, and then execute them again.

   If you do not change the time, existing searches will generate errors. After you change the time and re-execute the searches, you can go back to the original search time and re-execute the original search without errors.
3. Refresh or recreate dashboard charts.

   The dashboard supports two time ranges. One is fixed and the other is custom, such as the last one minute. You can only refresh widgets that use a custom time range.

   To resolve errors that existing widgets generate, do the following:

   - Refresh all charts that have a custom time range.
   - For charts with a fixed time range, click **...** > **Create search** to open the Search page and create a new search using the query from the dashboard chart.

Chapter 8: 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 8.0.1 and restore it to Vertica 8.1.0.
- 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, 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:

   ```
   # su -l $db_admin
   # /opt/vertica/bin/vbr --setupconfig
   ```

   The `vertica_backup.ini` file is created in `/home/$db_admin`.

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

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 Event Broker 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
```

Appendix A: Troubleshooting Deployment

This appendix describes issues that might occur during Investigate deployment.

Vertica Scheduler Exception: '[Vertica][VJDBC](5156) ERROR: Unavailable: initiator locks for query - Locking failure...'

This is an informational message stating that the Investigate schema is not accessible at that moment and is expected behavior. You can ignore this message.

Pod Starting Order

After you deploy Event Broker, pods should start in the following order. Otherwise, downstream pods will not start.

1. Zookeeper pods

   The total number of Zookeepers pods must be odd, starting with 3. All Zookeeper pods must be running.
2. Kafka pods
3. Schema registry pod
4. Bootstrap Web Service and Kafka Manager
5. Transformation Stream Processor and Routing Stream Processor

SSL Connection Errors

These errors occur if there is a connection issue between Kafka and a consumer or producer. Ensure that you imported certificates to Event Broker and the consumers.

kubectl Command Returns Refused or Time-out Connection

Unset the proxy, and then repeat the command.

Vertica Scheduler Unable to Read Events from Kafka

If this is an initial setup:

- Ensure that the Kafka scheduler is configured to communicate to Kafka port 9092.
- Check the network connection.
- Ensure that Event Broker is running.

If this is an existing setup:

- If the scheduler stops working and does not recognize the offset ids of messages that are in the topic, delete the scheduler and then create a new one:

  ```
  kafka_scheduler delete
  ```

  ```
  kafka_scheduler create <Broker_List:Port>
  ```

  Note: You might receive duplicate events in Vertica.

- Ensure that Event Broker is running.

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 2.30)

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!