Hewlett Packard
Enterprise

HPE Security ArcSight Connectors

SmartConnector for ArcSight Common Event Format Hadoop

Configuration Guide

October 17, 2017

Configuration Guide

SmartConnector for ArcSight Common Event Format Hadoop

October 17, 2017

Warranty

Restricted Rights Legend

Revision History

Date	Description
10/17/2017	Added encryption parameters to Global Parameters.
08/15/2017	Updated link to CEF Implementation Standard.
11/30/2016	Updated installation procedure for setting preferred IP address mode.
05/16/2016	Added overview information about the CEF Implementation Standard.
02/15/2016	Clarified statement of what events are collected.
08/14/2015	First release of this Configuration Guide.

Contents

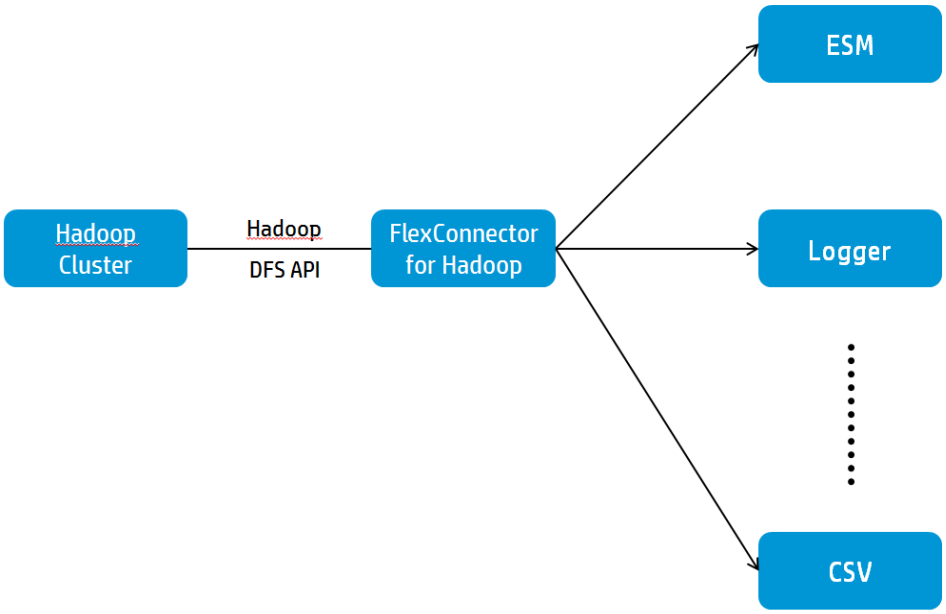
SmartConnector for ArcSight Common Event Format Hadoop

This guide provides information for installing the SmartConnector for ArcSight Common Event Format Hadoop (CEF Hadoop connector) and configuring it for event collection. Hadoop DFS version 2.5.1 is supported.

Product Overview

The Hadoop Distributed File System (HDFS) splits and stores large data files for processing across Hadoop machines in a cluster. This distributed file system provides high-throughput access to application data.

The connector provides a configurable method to collect any event (or record) stored in HDFS and forward the events to ESM or other destinations. The events collected must be in ArcSight Common Event Format (CEF).



This connector is designed to collect data from static files only. These files can be compressed (in `.gzip` or `.bz2` formats) or plain text. The folder that contains the source files can contain all three file formats. The HDFS API determines the compression type automatically (using the file extension of the compressed file) and then collects the data.

The file compression type .lzo is not currently certified for data collection with this connector.

The connector can collect data from local files or remote files. It collects data in batch mode, with no new events being written to the files. It collects files from a single folder that contains multiple files. It cannot collect from subfolders. All data in subfolders is ignored by the connector.

The default monitoring interval is `3600000 msec` to check for new files to collect. The connector checks for new files after the monitoring interval has expired or it is done processing the files from the

previous collection, whichever is later. The monitoring interval value can be changed; see "Parameter Modifications to Optimize Connector Performance" for details.

If the connector stops collecting data for any reason, it will start collecting data at the point it left off when data collection resumes. Files that are processed by the connector are moved to a processed-files folder, and the extension `.processed` is appended to the processed files. See "Hadoop DFS API Security Settings" for more information.

Common Event Format Implementation

The Common Event Format (CEF) standard format, developed by ArcSight, lets vendors and their customers quickly integrate their product information into ESM. CEF is an open log management standard that simplifies log management, letting third parties create their own device schemas that are compatible with a standard that is used industry-wide for normalizing security events. Technology companies and customers can use the standardized CEF format to facilitate data collection and aggregation, for later analysis by an enterprise management system.

The ArcSight Common Event Format (CEF) Guide, also known as "Implementing ArcSight Common Event Format (CEF)" defines the CEF protocol and provides details about how to implement the standard. It details the header and predefined extensions used within the standard as well as how to create user defined extensions. It also includes a list of CEF mappings as well as supported date formats.

To access this standard, go to https://community.saas.hpe.com/t5/ArcSight-Connectors/ArcSight-Common-Event-Format-CEF-Guide/ta-p/1589306 .

Configuration

Hadoop DFS API Security Settings

Certain default security properties for the Hadoop cluster must be changed to allow the CEF Hadoop connector to collect data for it. These changes should be made for the Name node, as it acts as a Master node. These properties are checked before other access control checks such as file permission, so verify that they are set as shown before proceeding:

Property: `hadoop.security.authorization`

This property is in the file `${HADOOP_CONF_DIR}/core-site.xml` and appears in these lines of XML:

```
<property>
<name>hadoop.security.authorization</name>
<value>true</value>
<description>Service level authorization params.</description>
</property>
```

The value should be **false** for this property.

Property: `security.client.protocol.acl`

This property is in the file `${HADOOP_CONF_DIR}/hadoop-policy.xml` and appears in these lines of XML:

```
<property>
<name>security.client.protocol.acl</name>
<value>*</value>
</property>
```

The value should be ***** (an asterisk) for this property.

Property: `dfs.permissions=false`

The value should be **false** for this property in `${HADOOP_CONF_DIR}/hdfs-site.xml.`

This property creates a processed-files folder inside the configured base folder (from which files are read) and move files into it after the files are processed by the connector. Processed files have the extension `.processed` appended to them. If a `Processed` folder does not exist, the connector creates one. If the connector cannot create this folder due to permissions or other issues, it logs an error message. In this case, files are processed, but are left in the base folder.

Parameter Modifications to Optimize Connector Performance

To optimize connector performance, you might want to modify the values for the internal parameters listed below.

After SmartConnector installation, you can access the connector's parameters as follows:

1. From the `$ARCSIGHT_HOME\current\user\agent` directory open the file `agent.properties` in a pure ASCII text editor.
2. In the `agent.properties` file, locate the parameters whose values you want to modify.
3. Enter values for these parameters as needed:

Parameter	Default Value	Description
filecheckinterval	3600000 msec	Time interval (in milliseconds) to change how frequently the connector retrieves events from the Hadoop cluster.
fileupdatewaitinterval	10000 msec	Time interval (in milliseconds) to wait before starting to process a new file. The file should not have been modified in the last 10 seconds as confirmation that the file transfer and writing is complete and the file is ready for processing.
processedfolderpath	/user/hadoop/processed	Path for the processed folder on the Hadoop cluster to contain files that are moved after processing.

4. Save and exit the `agent.properties` file.
5. Restart the connector.

Install the SmartConnector

The following sections provide instructions for installing and configuring your selected SmartConnector.

Certain security properties for the Hadoop cluster must be disabled for the SmartConnector for Common Event Format Hadoop to collect data. See "Hadoop DFS API Security Settings" for details.

Prepare to Install Connector

Before you install any SmartConnectors, make sure that the ArcSight products with which the connectors will communicate have already been installed correctly (such as ArcSight ESM or ArcSight Logger).

For complete product information, read the *Administrator's Guide* as well as the *Installation and Configuration* guide for your ArcSight product before installing a new SmartConnector. If you are adding a connector to the ArcSight Management Center, see the *ArcSight Management Center Administrator's Guide* for instructions, and start the installation procedure at "Set Global Parameters (optional)" or "Select Connector and Add Parameter Information."

Before installing the SmartConnector, be sure the following are available:

- Local access to the machine where the SmartConnector is to be installed
- Administrator passwords

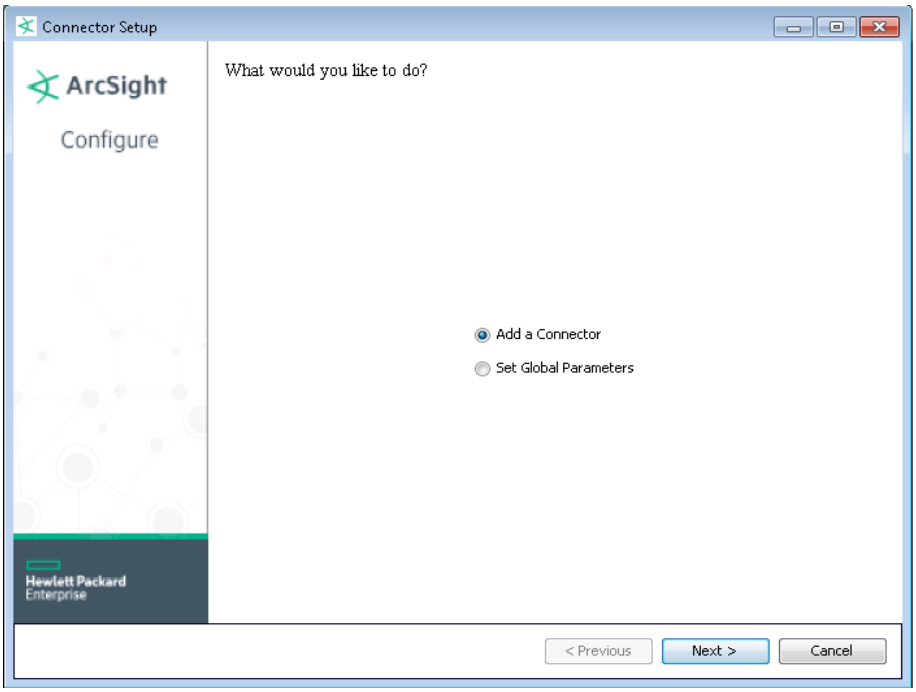
Install Core Software

Unless specified otherwise at the beginning of this guide, this SmartConnector can be installed on all ArcSight supported platforms; for the complete list, see the *SmartConnector Product and Platform Support* document, available from the HPE SSO and Protect 724 sites.

1. Download the SmartConnector executable for your operating system from the HPE SSO site.
2. Start the SmartConnector installation and configuration wizard by running the executable.

   Follow the wizard through the following folder selection tasks and installation of the core connector software:

   Introduction
   Choose Install Folder
   Choose Shortcut Folder
   Pre-Installation Summary
   Installing...
3. When the installation of SmartConnector core component software is finished, the following window is displayed:



Set Global Parameters (optional)

If you choose to perform any of the operations shown in the following table, do so before adding your connector. You can set the following parameters:

Parameter	Setting
FIPS mode	Select 'Enabled' to enable FIPS compliant mode. To enable FIPS Suite B Mode, see the SmartConnector User Guide under "Modifying Connector Parameters" for instructions. Initially, this value is set to 'Disabled'.
Remote Management	Select 'Enabled' to enable remote management from ArcSight Management Center. When queried by the remote management device, the values you specify here for enabling remote management and the port number will be used. Initially, this value is set to 'Disabled'.
Remote Management Listener Port	The remote management device will listen to the port specified in this field. The default port number is 9001.
Preferred IP Version	When both IPv4 and IPv6 IP addresses are available for the local host (the machine on which the connector is installed), you can choose which version is preferred. Otherwise, you will see only one selection. The initial setting is IPv4.

The following parameters should be configured only if you are using HPE SecureData solutions to provide encryption. See the *HPE SecureData Architecture Guide* for more information.

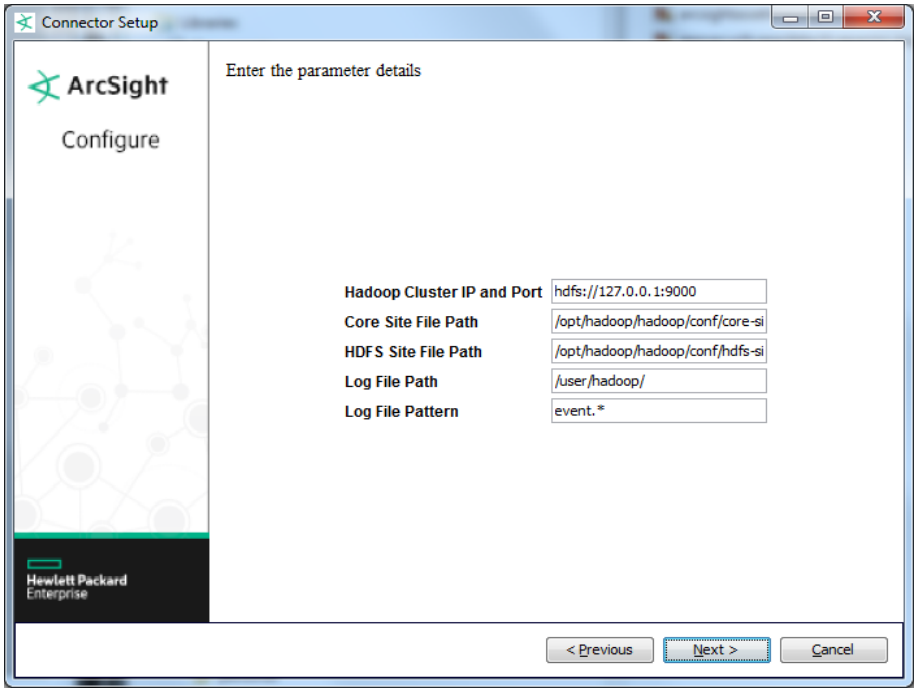
Parameter	Setting
Format Preserving Encryption	Data leaving the connector machine to a specified destination can be encrypted by selecting ‘Enabled’ to encrypt the fields identified in ‘Event Fields to Encrypt' before forwarding events. If encryption is enabled, it cannot be disabled. Changing any of the encryption parameters again will require a fresh installation of the connector.
Format Preserving Policy URL	Enter the URL where the HPE SecureData Server is installed.
Proxy Server (https)	Enter the proxy host for https connection if any proxy is enabled for this machine.
Proxy Port	Enter the proxy port for https connection if any proxy is enabled for this machine.

Parameter	Setting
Format Preserving Identity	The HPE SecureData client software allows client applications to protect and access data based on key names. This key name is referred to as the identity. Enter the user identity configured for HPE SecureData.
Format Preserving Secret	Enter the secret configured for HPE SecureData to use for encryption.
Event Fields to Encrypt	Recommended fields for encryption are listed; delete any fields you do not want encrypted and add any string or numeric fields you want encrypted. Encrypting more fields can affect performance, with 20 fields being the maximum recommended. Also, because encryption changes the value, rules or categorization could also be affected. Once encryption is enabled, the list of event fields cannot be edited.

After making your selections, click **Next**. A summary screen is displayed. Review the summary of your selections and click **Next**. Click **Continue** to return to proceed with "Add a Connector" window. Continue the installation procedure with "Select Connector and Add Parameter Information."

Select Connector and Add Parameter Information

1. Select **Add a Connector** and click **Next**. If applicable, you can enable FIPS mode and enable remote management later in the wizard after SmartConnector configuration.
2. Select **ArcSight Common Event Format Hadoop** and click **Next**.
3. Enter the required SmartConnector parameters to configure the SmartConnector, then click **Next**.



Parameter	Description
Hadoop Cluster IP and Port	Enter the IP address and port number of the Name Node (also known as the Master Node).

Parameter	Description
Core Site File Path	Enter the file path to the Hadoop Core Site.
HDFS Site File Path	Enter the file path to the Hadoop Distributed File System Site.
Log File Path	Enter the file path to the Hadoop log file.
Log File Pattern	Enter a pattern for data file names. Using the default value (event.*), the connector will look for log files starting with "event."

Select a Destination

1. The next window asks for the destination type; select a destination and click **Next**. For information about the destinations listed, see the *ArcSight SmartConnector User Guide*.

2. Enter values for the destination. For the ArcSight Manager destination, the values you enter for **User** and **Password** should be the same ArcSight user name and password you created during the ArcSight Manager installation. Click **Next**.

3. Enter a name for the SmartConnector and provide other information identifying the connector's use in your environment. Click **Next**. The connector starts the registration process.

4. If you have selected ArcSight Manager as the destination, the certificate import window for the ArcSight Manager is displayed. Select **Import the certificate to the connector from destination** and click **Next**. (If you select **Do not import the certificate to connector from destination**, the connector installation will end.) The certificate is imported and the **Add connector Summary** window is displayed.

Complete Installation and Configuration

1. Review the **Add Connector Summary** and click **Next**. If the summary is incorrect, click **Previous** to make changes.

2. The wizard now prompts you to choose whether you want to run the SmartConnector as a stand-alone process or as a service. If you choose to run the connector as a stand-alone process, select **Leave as a standalone application**, click **Next**, and continue with step 5.

3. If you chose to run the connector as a service, with **Install as a service** selected, click **Next**. The wizard prompts you to define service parameters. Enter values for **Service Internal Name** and **Service Display Name** and select **Yes** or **No** for **Start the service automatically**. The **Install Service Summary** window is displayed when you click **Next**.

4. Click **Next** on the summary window.

5. To complete the installation, choose **Exit** and Click **Next**.

For instructions about upgrading the connector or modifying parameters, see the *SmartConnector User Guide*.

Run the SmartConnector

SmartConnectors can be installed and run in stand-alone mode, on Windows platforms as a Windows service, or on UNIX platforms as a UNIX daemon, depending upon the platform supported. On Windows platforms, SmartConnectors also can be run using shortcuts and optional Start menu entries.

If the connector is installed in stand-alone mode, it must be started manually and is not automatically active when a host is restarted. If installed as a service or daemon, the connector runs automatically when the host is restarted. For information about connectors running as services or daemons, see the *ArcSight SmartConnector User Guide*.

To run all SmartConnectors installed in stand-alone mode on a particular host, open a command window, go to `$ARCSIGHT_HOME\current\bin` and run: `arcsight connectors`

To view the SmartConnector log, read the file `$ARCSIGHT_HOME\current\logs\agent.log`; to stop all SmartConnectors, enter `Ctrl+C` in the command window.

Device Event Mapping to ArcSight Data Fields

Refer to vendor CEF documentation for device mappings for that vendor's product.

Information from vendors is formatted according to the CEF standard and sent to the ArcSight SmartConnector, which translates the data into an ArcSight event.

Troubleshooting

Why am I getting a Java exception error - 'Failed to locate the winutils binary in the hadoop binary path'?

This error can sometimes happen when you are running a connector in a Windows environment.

Microsoft technical support recommends that you download the compiled `winutils.exe` program from the following link, and save it to the `C:\hadoop\winutils\bin` directory:
http://social.msdn.microsoft.com/Forums/windowsazure/en-US/28a57efb-082b-424b-8d9e-731b1fe135de/please-read-if-experiencing-job-failures?forum=hdinsight

Alternatively, you can fix the problem by editing the `agent.properties` file (which can be found at `$ARCSIGHT_HOME\current\user\agent`) and adding the `winutilpath` parameter to enter the current path to the utility; for example:

```
agents[0].winutilpath=c\:\\hadoop\\winutils\\
```

Why am I getting a Java exception error about missing permission for moving to the processed file?

You could receive this message when the connector does not have permission for renaming and moving the file from the path you identified in the parameter **Log File Path** configured during the connector installation process (or **logfilepath** parameter specified in the `agent.properties` file). Make sure the folder this path specifies has full read/write permission for the relevant account (you can change it with a command, such as `hadoop dfs -chmod a+w`). The connector then can read the file,

rename it, and move it out to the processed log file path. The same rule applies to the processed log file path.