Hewlett Packard
Enterprise

HPE Security ArcSight ESM Event Data Transfer Tool

Software Version: 1.2

User's Guide

June 30, 2016

Legal Notices

Warranty

The only warranties for Hewlett Packard Enterprise products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Hewlett Packard Enterprise shall not be liable for technical or editorial errors or omissions contained herein.

The information contained herein is subject to change without notice.

The network information used in the examples in this document (including IP addresses and hostnames) is for illustration purposes only.

HPE Security ArcSight products are highly flexible and function as you configure them. The accessibility, integrity, and confidentiality of your data is your responsibility. Implement a comprehensive security strategy and follow good security practices.

This document is confidential.

Restricted Rights Legend

Confidential computer software. Valid license from Hewlett Packard Enterprise 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

Follow this link to see a complete statement of copyrights and acknowledgements:
https://www.protect724.hpe.com/docs/DOC-13026

Support

Contact Information

Phone	A list of phone numbers is available on the HPE Security ArcSight Technical Support Page: https://softwaresupport.hpe.com/documents/10180/14684/esp-support-contact-list
Support Web Site	https://softwaresupport.hpe.com
Protect 724 Community	https://www.protect724.hpe.com

Contents

Objectives

Exporting ArcSight ESM events to Hadoop brings the opportunity for using Hadoop ecosystem technologies. The Event Data Transfer Tool exports ESM events in three formats, cef, csv, and key-value pairs. Getting event data from ESM allows more flexibility to combine analysis with unstructured data in addition to the structured CEF data. Exported events can benefit from different Hadoop technologies in the following scenarios:

- Build a security data warehouse using the Hadoop Distributed File System (HDFS), a Hadoop capability that enables inexpensive, long-term storage for Petabytes of data.
- The Event Data Transfer Tool can export ESM events in multiple formats, which enable the usage of Apache Hive to query and manage the security data warehouse. Hive provides HiveQL, a SQL-like language to query and manipulate large events stored on HDFS. See https://hive.apache.org/.
- Once the event data is in Hadoop, you can still run searches quickly. There are powerful search engines such as Elasticsearch or Solr open-source technologies. These search engines can provide results in less than a second, which is good for forensics analysis on the data archives. These platforms are built on top of Apache Lucene, the high-performance text engine library. Both technologies also provide visualization platforms.
  For more information, refer to the following links:
  https://lucene.apache.org/core/
  https://www.elastic.co/products/hadoop
  http://lucene.apache.org/solr/
- You can detect novel attacks by adopting Machine-learning approaches on the security data. For example, you can identify new botnet activities in your network by using machine-learning algorithms on large security events data. You can run clustering algorithms such as K-means that Apache Mahout provides to discover attack patterns in your big data. Apache Mahout is a suite of machine-learning libraries designed for big data analysis on Hadoop. You can also send your analysis results as alerts back to ESM by using the SmartConnector for ArcSight Common Event Format Hadoop. Refer to the configuration guide for that SmartConnector. This improves the enterprise's attack-detection capabilities. For more information, refer to the following link:
  http://mahout.apache.org/
- Run your own, homegrown analytics scripts to process and analyze large security data by using Pig Latin scripting language that Apache Pig platform provides. You can send your analysis results as alerts back to ESM by using the SmartConnector for ArcSight Common Event Format Hadoop. Refer to the configuration guide for that SmartConnector. For more information, refer to the following link:
  https://pig.apache.org/

The Event Data Transfer Tool

This tool applies to ESM with CORR-Engine and the Hadoop Distributed File System (HDFS). Read all of this before you start.

Use this Event Data Transfer Tool to send event data from ESM to a Hadoop cluster for additional processing.

Hadoop is a large batch-processing system that can scale to many computers, thus distributing work across them all. HDFS splits large data files for processing on different Hadoop machines. In this way it can process more data than can fit on a single computer's hard disk drive by separating the input across the cluster and processing the data in parallel. File blocks are copied to more than one node to mitigate against individual machine failures.

For information on what versions of ESM and Hadoop are supported, refer to the *HPE ArcSight ESM Support Matrix*, which is available on the Protect 724 Community at https://www.protect724.hpe.com.

Installing the Event Data Transfer Tool

The only prerequisites to running this installation are that you have the Hadoop `core-site.xml` file and that you install this tool on a machine running ESM with CORR-Engine. For the version and build number and supported relationships, refer to the ESM Support Matrix at https://www.protect724.hpe.com.

Before downloading the installer package from Hewlett Packard Enterprise (HPE):

HPE provides a digital public key to enable you to verify that the signed software you received is indeed from HPE and has not been manipulated in any way by a third party.

Visit the following site for information and instructions:

https://h20392.www2.hpe.com/portal/swdepot/displayProductInfo.do?productNumber=HPLinuxCodeSigning

1. Download the installer package and extract this file:
   `ArcSight-CORRHadoop-<version.build Number>-Linux.bin`
2. From the directory where this file is, run the following (We recommend that you do not do this as user *root*.):
   `./ArcSight-CORRHadoop-<version.build Number>-Linux.bin`
3. Click **Next** at the Introduction screen.
4. Read the license agreement. After you scroll to the bottom, you can select the option to agree to the terms. Click **Next**.

5. To browse to a folder where you want to install this tool, select **Choose**, navigate to the folder, and click **OK**. Do not install it in the **/opt/ESMComponents** folder.

   To accept the default folder (`/opt/arcsight/corrhadoop`), you do not have to do anything.

   This path/folder is referred to later as <CORREHADOOP_HOME>.

   Click **Next**.
6. Choose a link location, if you want to create links. Click **Next**.
7. Choose the location of the `core-site.xml` file and click **Next**.
8. Review your selections and select **Install** when satisfied.
9. When the installation is complete, click **Done**.

In case of an installation failure:

- Make sure all the prerequisites are met.
- Use the uninstaller and try reinstalling.
- If the uninstall folder was not installed, clear the installation folder and try again.

Using the Event Data Transfer Tool Command

Once installed, you use the following `arcsight` command to manage event data transfers:

Syntax: `arcsight event_transfer [parameters]`

Required Parameters:

- `-dpath <dpath>`
  Specify the path and file name to the destination.
  - If the format is CSV, the extension must be `csv`, `gz`, or `bz2`.
  - If the format is keyvalue of CEF, the extension must be be `txt`, `gz`, or `bz2`.
  - When the extension is gz or bz2 the file is compressed.
  - When the extension is csv or txt, the file is not compressed.

- `-dtype <dtype>`
  Specify the type of destination. File means a local path and Hadoop means a path to a Hadoop system.

Optional Parameters:

- `-format <format>`
  Specify the format as `cef` (common event format), `csv` (comma-separated values), or `keyvalue` (key-value pairs).
  - The default is `keyvalue`.
  - Use all lower case letters.
  - If you use the CSV format, the file name extension in `-dpath` must be `csv`, `gz`, or `bz2`.
  - If you use keyvalue or CEF formats, the file extension must be `txt`, `gz`, or `bz2`.

- `-columns <columns>`
  List the CEF column names to include in the transfer. Separate column names with spaces.The default is all columns.
- `-start <start>`
  Specify the start of the range of events to transfer as a time (mm/dd/yyyy hh:mm:ss) or by event ID. The default is yesterday at this time ($NOW-1d). The time format is for a 24-hour clock. That is, hh is 00 - 24.
- `-end <end>`
  Specify the end of the range of events to transfer as a time (mm/dd/yyyy hh:mm:ss) or by event ID. The default is the time specified by $NOW. The time format is for a 24-hour clock. That is, hh is 00 - 24.
- `-qtype <qtype>`
  Specify the type of entries you used in `-start` and `-end`. For times, the parameters can be EndTime or ManagerReceiptTime (the default). For event IDs use EventId.
- `-sg “<storageGroup>”`
  Specify one or more storage groups, in double quotes, and separated by a space. If omitted, events in all storage groups are transferred.
- `-threads <threads>`
  Specify the number of threads to use for the transfer. The default is 5. See "Threads" on the next page.
- `--h`
  Help

`Examples:`

```
arcsight event_transfer -dpath <***path***> -dtype Hadoop -sg “storage group 1” “storage group 2”
```

```
arcsight event_transfer -dtype Hadoop -dpath <***path***> -format cef -start "05/04/2016 15:45:00" -end "05/04/2016 16:45:00”
```

Note:

- The `-start`, `-end`, and `-qtype` parameters must be of the same type: either event ID or time. If you mix them up, the tool cannot tell and you get unexpected results.
- The command parameters are case sensitive; use them as shown.

Usage Notes

File Names

When the data is transferred to Hadoop, the filename you specify in `-dpath`, such as `abc.gz`, has the start and end appended in front of the extension in the form `abc_<start>_<end>.gz.`

More threads generate more files.

The file extension (`.gz`, in this example), specifies the compression used. If you do not want any compression,use a file extension such as `.txt` or `.csv`.

Threads

The number of threads selected for the transfer affects the rate of transfer in events per second (EPS). Tests have shown that increasing the number of threads increases throughput, but at some point, more threads actually reduce EPS due to resource limitations. The point at which this occurs depends on the number of processors and the amount of other work on that machine. The default of five threads should be satisfactory in most cases.

If you go to 10 threads you need 3 GB of memory available and another 3 GB for each additional 10 threads. The memory used for this process is controlled by a line in a script used by the `Event_Transfer` command. You can edit the `DirectMemorySize` value in the file, `config/corrhadoop-config.sh`.

To change the memory used:

1. Comment this line:

   `#JVM_HEAP_SIZE="-Xms3g -Xmx3g"`
2. Un-comment this line, which sets direct memory size to 6 GB by default:

   `DIRECT_MEMORY_SIZE="-XX:MaxDirectMemorySize=6g`
3. Optionally change the `DirectMemorySize` value as applicable.

   The configuration file itself provides suggestions for size based on threads.

Data Compression

You can specify a Hadoop compression codec by using the file name extension or suffix that corresponds to that codec. Compression occurs before the data is transferred. The compression codecs supported by this transfer tool are:

Suffix	Codec
.bz2	Bzip2Codec
.gz	GzipCodec

The Bzip2Codec appears to have better compression, but the GzipCodec appears to provide a higher EPS value for transfer to the Hadoop system. For more information on these compression codecs, refer to your Hadoop documentation. The `event_transfer` command removes the CORR-Engine compression and then applies the Hadoop compression before, but as part of, the transfer.

If you do not specify a codec-specific file extension, the data is not compressed. Without Hadoop compression, the data in Hadoop is larger than the archive size in the CORREngine. This is because the CORR-Engine data is uncompressed when it is transferred to the Hadoop cluster and the Hadoop file format is larger.

This event migration tool does not transfer Binary Large Object (BLOB) or Character Large Object (CLOB) data.

Transfer Failures

If the transfer fails, you must delete all the data that was transferred in the attempt, before you retry the operation.

The number of files transferred depends on the number of threads used.

File names can be identified by their timestamp.

Transfer Performance

Whether transferring data to Hadoop impacts normal ESM performance depends on how many events you transfer, which event columns you opt to transfer, how often you transfer data, and the number of threads used for data transfer.

However, there is no way to recommend settings that will work in all environments. Try various settings until you settle on the ones that work best for you.

Size of Transferred Files

You may notice a difference in the files' sizes between Hadoop Distributed File System (HDFS) and Linux local file system. This difference appears only when you use the command `ls -h` (list files with human readable format). Instead, verify the files' sizes using the basic `ls` command.

Column Names

The column (field) names assigned in Hadoop are the Common Event Format (CEF) names. For a description of the CEF field names, refer to the document entitled *Implementing ArcSight Common Event Format (CEF)*, which is available on Protect 724 at https://www.protect724.hpe.com.

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 system, click the link above and an email window opens with the following information in the subject line:

Feedback on User's Guide (ESM Event Data Transfer Tool 1.2)

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 arc-doc@hpe.com.

We appreciate your feedback!