HPE Security ArcSight SmartConnectors

ArcSight Locales and Encodings

August 15, 2017

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

© Copyright 2009, 2017 Hewlett Packard Enterprise Development, LP

Follow this link to see a complete statement of copyrights and acknowledgements:
https://community.saas.hpe.com/t5/Discussions/Third-Party-Copyright-Notices-and-License-Terms/td-p/1589228

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://community.saas.hpe.com/t5/ArcSight/ct-p/arcsight

Revision History

Date	Description
08/15/2017	Updated content.
05/15/2017	Updated document format.
01/05/2009	First release of this white paper.

Contents

About ArcSight Locales and Encodings

The information in this document applies to ArcSight ESM v4.0 SP1 and later and ArcSight Logger v2.0 and later.

ArcSight products can be configured to display in a number of languages, including Japanese, French, traditional Chinese, simplified Chinese, and Korean. However, to enable your ArcSight products to parse and display log events generated by devices localized for non-English languages, some customization of the parser script may be necessary.

Setting the Locale for any of these languages ensures that you get the appropriate environment in terms of language settings, number format, date/time format, time zone settings, and Daylight Saving Time setting for that country or language.

This document describes the updates to be taken into consideration when configuring ArcSight ESM for a supported language. Examples from key-value and regex parsers are provided.

ArcSight Logger can log events in some non-English and some non-Western languages, but has not yet been localized to a non-English language. Logger should support every encoding, but has only been certified with a representative sample.

Some of the common terms used in this document are described in the following table.

Term	Description
Character Set	A character set is a collection of characters that have been grouped together for a particular purpose. An example of a character set is the English alphabet.
Code Point	Each character value within a code set is assigned a unique value, referred to as a code point.
Code Set	Collectively, code point values are known as a code set.
Encoding	Encoding specifies how each code point is stored in memory or disk files.
Locale	Locale refers to the language specific to the region where you are running ArcSight SmartConnectors. A locale can include language, number format, date-time format, and other settings.
Localization	Localization, or Internationalization, is the process of adding language-specific files to an internationalized application so that the application supports the language. In this context, localization refers to configuring a connector to display events in the local language when the default system encoding for the connector is English.
Reverse Localization	In this context, reverse localization refers to the process of converting localized key properties back into English, when the default system encoding for the connector is not English. Reverse localization enables the connector to parse and standardize the event data correctly before sending it to the application, where it will display in the language selected at installation.

Localized ArcSight ESM

The ArcSight Manager, Database, and Console components should all be configured with the same locale.

By default, all communication between ArcSight components is done using UTF-8 character encoding. Even though ArcSight ESM supports only UTF-8 internally, if your connector receives events in UTF-16, for example, the events are still stored correctly since these events get converted to UTF-8 by the connector before they are passed on to the Manager. The Manager then passes these events to the database where they are converted to the language-specific encoding you selected while installing the database before being persisted.

ArcSight Database

Before you install ArcSight Database, select an encoding scheme.

> **Note:** You cannot make changes to the encoding after you have installed the database. Any change will require reinstallation.

You can choose between UTF-8 and pre-defined language-specific encodings during database installation. The advantage of using UTF-8 is that it supports all major languages in the world, so no data is lost when it is saved in the database. On the other hand, UTF-8 requires more space to store certain characters than the character's language-specific encoding. For example, if a certain Japanese character can be stored in two bytes using JA16SJIS encoding, the same character might take 3 bytes if stored in UTF-8.

The following table lists the available languages.

Language	Encoding
English	WE8MSWIN1252
French	WE8ISO8859P1
Japanese	JA16SJIS
Chinese_Simplified	ZHS16CGB231280
Chinese_Traditional	ZHT16BIG5
Korean	KO16KSC5601
Unified_UTF8	UTF8

In the table, "English" represents all western European languages. If you need to use a character set not shown in the table, see the *ArcSight Installation and Configuration Guide* for instructions on how to set it correctly.

If you anticipate that you will be storing events in multiple languages, choose a character set (encoding) that is compatible with ALL languages you intend to use.

For more than one non-English language, you should choose Unified_UTF8.

If you select Unified_UTF8, you must select the language in which you want the standard content to be installed on the database.

If you already have the database installed with an encoding other than Unified_UTF8, but would like to change the encoding to Unified_UTF8, you must re-install the database and select Unified_UTF-8 Database Character Set when prompted during the installation.

ArcSight Manager

Install the ArcSight Manager on an operating system that is of the same language as the language you selected while installing the database. During startup, ArcSight Manager automatically detects and uses the locale from the operating system.

ArcSight Console

Install the ArcSight Console on an operating system that is of the same language as the language you selected while installing the database. During startup, ArcSight Console automatically detects and uses the locale from the operating system.

ArcSight SmartConnectors

If a device is configured to use a language-specific encoding, the connector receiving events from this device should be configured to use the same encoding as the device.

Setting the Encoding for Selected SmartConnectors

You can set the encoding to a character set corresponding to your Locale for the following SmartConnectors only. See the SmartConnector Configuration Guides for each of these connectors for instructions about how to configure the encoding for each connector.

- SAP Real-Time Security Audit Multi-Folder Connector.
- IBM DB2 Audit File Connector.
- Oracle SYSDBA Audit Multi-Folder Connectors

These SmartConnectors support all character sets supported by Java. For a list of the character sets supported by Java see http://java.sun.com/j2se/1.5.0/docs/guide/intl/encoding.doc.html.

> **Note:** You must change the encoding to match the log files' encoding only if the log files use an encoding other than the default one.

Connectors not mentioned above use the default encoding of the operating system on which they reside. Each operating system comes with default encodings for various languages of the world. So, the encoding used in a connector is either based on the character set that you selected when installing the ArcSight Database or the operating system you are using.

Localization of Date Formats in Tokens and Operations

If your connector receives logs that contain timestamps in a non-English language or a date format that is customarily used by a non-English locale (for example, "mai 24, 2006 12:56:07.615" where "mai" is German for May) that your connector is set to, configure the `agent.parser.locale.name` property in the `agent.properties` file. This file is located in `ARCSIGHT_HOME/current/user/agent` directory.

Set the `agent.parser.locale.name` property to the value that corresponds to the connector's locale. By default, this property is set to en_US. Refer to ""List of possible values for the agent.parser.locale.name property" on page 11" for possible values for this property.

Key-Value Parsers for Localized Devices

Some localized devices not only send localized values but also localized keys in event messages. In such a case, additional processing may be needed to translate the keys to English for the event messages to be properly parsed. For example, assume that the content of a key-value parser is:

event.destinationUserName=User

And the received event message is:

User= 김 where 김 is Korean for KIM.

In this case, the parser works fine as-is, since double-byte is already supported.

If the received event message is: 우새르 = 김 where 우새르 is Korean for User, then additional mapping is needed to translate 우새르 to User.

If you encounter a need for a localized device, contact ArcSight Support.

Windows Event Log Connector supports the following locales to parse the non-English language Keys in the Windows Event Log description:

- ja (Japanese)
- de (German)
- zh_CN (Simplified Chinese)
- zh_TW (Traditional Chinese)

Contact ArcSight Support for assistance with other non-supported languages.

Examples

The following examples cover two different scenarios.

Scenario 1 - Events received in a single language only

This scenario describes what to do when your connector(s) receive data in a single language only, such as Japanese. In ESM , the default encoding for Japanese is JA16SJIS.

Database

While installing the database, in the ArcSight Oracle Installation Wizard, select one of the following from the **Database Character Set** drop down menu:

    Japanese

    Unified_UTF8

If you select **Japanese**, the database uses JA16SJIS encoding when saving the data into the database.

If you select **Unified_UTF8**, you must also select **Japanese** in the ArcSight Database Schema Initialization screen to ensure that the default system resources get installed in Japanese.

Keep in mind, some characters might take 3 bytes when stored in UTF-8 but might take only 2 bytes when stored in JA16SJIS.

ArcSight Manager, Console, and Web

You must install the ArcSight Manager, Console, and Web on a Japanese operating system. On startup, these components automatically pick up and use the locale from the operating system.

Scenario 2 - Events received in multiple languages

This scenario is an example of what to do when you are dealing with multiple connectors that receive data in different languages.

Database

When you install the database, in the ArcSight Oracle Installation Wizard:

1. Select **Unified_UTF8** from the **Database Character Set** dropdown menu. This ensures that no data is lost in translation when persisted in the database.

2. In the ArcSight Database Schema Initialization screen, select the language in which you want the standard content resources to be installed.

ArcSight Manager, Console, and Web

When you installed the ArcSight Database you selected a language in which to install the system resources. You should install the ArcSight Manager, Console, and Web on an operating system of that same language. On startup, these components automatically pick up the locale from the operating system.

Preparing to Install the Language Update

If you are currently running ESM v4.0 GA and would like to switch to a localized version, you must upgrade your ESM installation (ArcSight Database, Manager, Console, and Web server) to v4.0 SP1.

> **Note:** While upgrading your database to v4.0 SP1, make sure that the character set you select during the upgrade is compatible with the one that you selected when installing your existing database.

Once your system is running ESM v4.0 SP1, you need to install the language update.

Verifying the Character Set used on your Database

If you currently use ESM v4.0 SP1, your database already has a character set specified. Follow this procedure to validate the character set that was selected when the v4.0 SP1 database was installed:

1. Run the following command from the `ARCSIGHT_HOME/bin` directory:

   ```
   arcdbutil sql
   ```

2. When prompted for user-name, enter:

   ```
   / as sysdba
   ```

3. Run the following SQL statement:

   ```
   SQL>select "PARAMETER", "VALUE" from SYS.GV_$NLS_PARAMETERS where
   PARAMETER='NLS_CHARACTERSET';
   ```

> **Note:** You can set the encoding only during database installation. To change the encoding after installation, you must reinstall ArcSight Database.

The following character sets (encodings) are supported for ArcSight Database:

Language	Character Set
English	WE8MSWIN1252
French	WE8ISO8859P1
Japanese	JA16SJIS
Chinese Simplified	ZHS16CGB231280
Chinese Traditional	ZHT16BIG5
Korean	KO16KSC5601
Unified UTF8	UTF-8

Installing the Language Update

By now, your database should be set to the encoding of your choice. If you have not already done so, follow the instructions in "Verifying the Character Set used on your Database" to verify the database encoding, before you proceed.

You must install the language update on ArcSight Manager, Console and Web. Refer to the Release Notes for the Language Update for installation instructions.

Possible Values the Locale Property

The following table lists the possible values for this property.

Values	Language	Country	Variant
ar	Arabic		
ar_AE	Arabic	United Arab Emirates	
ar_BH	Arabic	Bahrain	
ar_DZ	Arabic	Algeria	
ar_EG	Arabic	Egypt	
ar_IQ	Arabic	Iraq	
ar_JO	Arabic	Jordan	
ar_KW	Arabic	Kuwait	
ar_LB	Arabic	Lebanon	
ar_LY	Arabic	Libya	
ar_MA	Arabic	Morocco	
ar_OM	Arabic	Oman	
ar_QA	Arabic	Qatar	
ar_SA	Arabic	Saudi Arabia	
ar_SD	Arabic	Sudan	
ar_SY	Arabic	Syria	
ar_TN	Arabic	Tunisia	

Values	Language	Country	Variant
ar_YE	Arabic	Yemen	
be	Belarusian		
be_BY	Belarusian	Belarus	
bg	Bulgarian		
bg_BG	Bulgarian	Bulgaria	
ca	Catalan		
ca_ES	Catalan	Spain	
cs	Czech		
cs_CZ	Czech	Czech Republic	
da	Danish		
da_DK	Danish	Denmark	
de	German		
de_AT	German	Austria	
de_CH	German	Switzerland	
de_DE	German	Germany	
de_LU	German	Luxembourg	
el	Greek		
el_GR	Greek	Greece	
en	English		
en_AU	English	Australia	
en_CA	English	Canada	
en_GB	English	United Kingdom	
en_IE	English	Ireland	
en_IN	English	India	
en_NZ	English	New Zealand	
en_US	English	United States	
en_ZA	English	South Africa	
es	Spanish		
es_AR	Spanish	Argentina	
es_BO	Spanish	Bolivia	

Values	Language	Country	Variant
es_CL	Spanish	Chile	
es_CO	Spanish	Columbia	
es_CR	Spanish	Costa Rica	
es_DO	Spanish	Dominican Republic	
es_EC	Spanish	Ecuador	
es_ES	Spanish	Spain	
es_GT	Spanish	Guatemala	
es_HN	Spanish	Honduras	
es_MX	Spanish	Mexico	
es_NI	Spanish	Nicaragua	
es_PA	Spanish	Panama	
es_PE	Spanish	Peru	
es_PR	Spanish	Puerto Rico	
es_PY	Spanish	Paraguay	
es_SV	Spanish	El Salvador	
es_UY	Spanish	Uruguay	
es_VE	Spanish	Venezuela	
et	Estonian		
et_EE	Estonian	Estonia	
fi	Finnish		
fi_FI	Finnish	Finland	
fr	French		
fr_BE	French	Belgium	
fr_CA	French	Canada	
fr_CH	French	Switzerland	
fr_FR	French	France	
fr_LU	French	Luxembourg	
hi_IN	Hindi	India	
hr	Croatian		
hr_HR	Croatian	Croatia	

Values	Language	Country	Variant
hu	Hungarian		
hu_HU	Hungarian	Hungary	
is	Icelandic		
is_IS	Icelandic	Iceland	
it	Italian		
it_CH	Italian	Switzerland	
it_IT	Italian	Italy	
iw	Hebrew		
iw_IL	Hebrew	Israel	
ja	Japanese		
ja_JP	Japanese	Japan	
ko	Korean		
ko_KR	Korean	Korea	
lt	Lithuanian		
lt_LT	Lithuanian	Lithuania	
lv	Latvian		
lv_LV	Latvian	Latvia	
mk	Macedonian		
mk_MK	Macedonian	Macedonia	
nl	Dutch		
nl_BE	Dutch	Belgium	
nl_NL	Dutch	Netherlands	
no	Norwegian		
no_NO	Norwegian	Norway	
no_NO_NY	Norwegian	Norway	Nynorsk
pl	Polish		
pl_PL	Polish	Poland	
pt	Portuguese		
pt_BR	Portuguese	Brazil	
pt_PT	Portuguese	Portugal	

Values	Language	Country	Variant
ro	Romanian		
ro_RO	Romanian	Romania	
ru	Russian		
ru_RU	Russian	Russia	
sk	Slovak		
sk_SK	Slovak	Slovakia	
sl	Slovanian		
sl_SI	Slovanian	Slovania	
sq	Albanian		
sq_AL	Albanian	Albania	
sv	Swedish		
sv_SE	Swedish	Sweden	
th	Thai		
th_TH	Thai	Thailand	
th_TH_TH	Thai	Thailand	TH
tr	Turkish		
tr_TR	Turkish	Turkey	
uk	Ukranian		
uk_UA	Ukranian	Ukraine	
vi	Vietnamese		
vi_VN	Vietnamese	Vietnam	
zh	Chinese		
zh_CN	Chinese	China	
zh_HK	Chinese	Hong Kong	
zh_TW	Chinese	Taiwan	