EMC® VNX® Series
Release 8.1

Using International Character Sets on VNX® for File
P/N 300-015-112 Rev 01

Contents

As part of an effort to improve and enhance the performance and capabilities of its product lines, EMC periodically releases revisions of its hardware and software. Therefore, some functions described in this document may not be supported by all versions of the software or hardware currently in use. For the most up-to-date information on product features, refer to your product release notes.

If a product does not function properly or does not function as described in this document, please contact your EMC representative.

Special notice conventions

EMC uses the following conventions for special notices:

Note: Emphasizes content that is of exceptional importance or interest but does not relate to personal injury or business/data loss.

NOTICE Identifies content that warns of potential business or data loss.

⚠ CAUTION Indicates a hazardous situation which, if not avoided, could result in minor or moderate injury.

⚠ WARNING Indicates a hazardous situation which, if not avoided, could result in death or serious injury.

⚠ DANGER Indicates a hazardous situation which, if not avoided, will result in death or serious injury.

Where to get help

EMC support, product, and licensing information can be obtained as follows:

Product information—For documentation, release notes, software updates, or for information about EMC products, licensing, and service, go to EMC Online Support (registration required) at http://Support.EMC.com.

Troubleshooting—Go to EMC Online Support at http://Support.EMC.com. After logging in, locate the applicable Support by Product page.

Technical support—For technical support and service requests, go to EMC Customer Service on EMC Online Support at http://Support.EMC.com. After logging in, locate the applicable Support by Product page, and choose either **Live Chat** or **Create a service request**. To open a service request through EMC Online Support, you must have a valid support agreement. Contact your EMC sales representative for details about obtaining a valid support agreement or with questions about your account.

Note: Do not request a specific support representative unless one has already been assigned to your particular system problem.

Your comments

Your suggestions will help us continue to improve the accuracy, organization, and overall quality of the user publications.

Please send your opinion of this document to:

techpubcomments@EMC.com

Introduction

The EMC VNX supports the use of international character sets. To provide support for world languages, VNX has adopted the Unicode language character standards and the UTF-8 multibyte character-encoding format. VNX support of Unicode and multibyte character encoding allows you to create files and directories as well as VNX data element names in your native language.

This document provides an overview of VNX internationalization support, also referred to as I18N. It describes how to enable Unicode. It also helps you determine whether you need to customize the VNX to support an international character set used by your operating environment. It also explains how to convert existing file or directory names to Unicode.

This document is part of the VNX documentation set and is intended for use by administrative users who need to internationalize their VNX.

Topics included are:

- System requirements on page 8
- Restrictions on page 8
- Cautions on page 8
- User interface choices on page 8
- Related information on page 9

System requirements

Table 1 on page 8 describes the EMC® VNX® software, hardware, network, and storage configurations.

Table 1. System requirements

Software	VNX version 8.1.
Hardware	No specific requirements.
Network	No specific requirements.
Storage	No specific requirements.

Restrictions

Using Unicode and international character sets with VNX has the following restrictions:

◆ Celerra® Monitor does not support Unicode and international character sets.

◆ If Unicode is enabled, VNX does not support SHARE or UNIX-based user authentication of Common Internet File System (CIFS) clients.

Cautions

If any of this information is unclear, contact your EMC Customer Support Representative for assistance:

◆ The nas_server -ConvertI18N option for a Virtual Data Mover (VDM) does not trigger Unicode conversion for a file system mounted on the VDM. You must convert file and directory names by using the uc_config -convert option as described in Converting user files and directories to Unicode on page 30.

◆ The procedures in this document are irreversible. EMC strongly advises that you back up your system to tape before performing any of the procedures documented here. Contact EMC Customer Support for help in performing this procedure.

User interface choices

This document describes how to enable Unicode and customize the use of international character sets by using the command line interface (CLI). You can also enable Unicode by using other VNX management applications, such as EMC Unisphere® software, VNX wizards, and VNX Installation Assistant for File/Unified.

Related information

Specific information related to the features and functionality described in this document is included in:

- *EMC VNX Command Line Interface Reference for File*
- VNX for File man pages
- *Parameters Guide for VNX for File*
- *Configuring NDMP Backups on VNX*
- *Configuring and Managing CIFS on VNX*
- *Configuring NFS on VNX*
- *Using FTP, TFTP, and SFTP on VNX*

The official Unicode standards are available at www.unicode.org/unicode/standard/versions.

EMC VNX documentation on EMC Online Support

The complete set of EMC VNX series customer publications is available on EMC Online Support. To search for technical documentation, go to http://Support.EMC.com. After logging in to the website, click **Support by Product** and type **VNX series** in the Find a Product text box. Then search for the specific feature required.

VNX wizards

Unisphere software provides wizards for performing setup and configuration tasks. The Unisphere online help provides more details on the wizards.

2

Concepts

The concepts to understand VNX internationalization support, also referred to as I18N are:

- VNX Unicode support on page 12
- VNX translation support for user files and directories on page 13
- Translation directory structure and files on page 14
- Planning considerations on page 18
- Enable Unicode on page 23
- Multibyte support for VNX data elements on page 23

VNX Unicode support

VNX has adopted the language character standards specified by the Unicode Consortium. The standard for digital representation of the characters used in writing all of the world's languages, Unicode provides a uniform means for storing, searching, and interchanging text in any language. It is used by all modern computers and is the foundation for processing text on the Internet. Unicode is developed and maintained by the Unicode Consortium. Currently, the Unicode standard contains distinct coded characters derived from 24 supported language scripts. These characters cover the principal written languages of the world including:

- European Latin-based and Slavic languages

- Semitic languages, such as Arabic and Hebrew

- East Asia languages, such as Chinese, Korean, and Japanese in all their orthographic versions

Each character stored on the computer is represented by using an encoding scheme in a Unicode Transformation Format (UTF). The most commonly used schemes are UTF-8 and UTF-16.

The VNX default character-encoding scheme is ASCII. VNX Unicode character-encoding support must be enabled. Once Unicode is enabled, VNX can recognize and process any characters defined by the Unicode 3.0 standard.

Important: EMC strongly recommends that you enable Unicode. Enable Unicode on page 23 describes the different methods that you can use to enable Unicode on your VNX.

Unicode support for VNX data element names

If your operating system and terminal software support the language being used, and are set to expect received data in UTF-8 character encoding format, the VNX management interfaces and tools allow you to create and display VNX data element names in any language characters defined by the Unicode 3.0 standard after Unicode is enabled on a system. Interfaces and tools include:

- Unisphere software

- VNX wizards

- CLI

- XML API

- Unisphere MMC snap-ins

For example, if you are the storage administrator for a Japanese company you can create a new CIFS share name and comment in Japanese by using the server_export command

or the CIFS Share Wizard and view this CIFS share name and comment in Unisphere software.

VNX support for languages that use multibyte character sets is being implemented incrementally. The *EMC VNX Operating Environment for File Release Notes* provide a list of the VNX data element names that currently support multibyte characters.

Unicode support for user file and directory names

After Unicode is enabled on VNX system, and if needed, the necessary character translation files have been installed, users on client systems can create file and directory names in any language characters defined by the Unicode 3.0 standard and have them accurately stored on disk.

Note: To view language characters on a client workstation, ensure that the correct fonts have been installed. The workstation or operating system's user documentation provides more information for specific installation and configuration instructions.

EMC strongly recommends enabling Unicode on a newly installed VNX before users begin creating files and directories. User files and directories that are created before the system is configured are considered existing files. If there are any non-ASCII file or directory names that were created on client systems that use character encodings such as SJIS, Extended UNIX Code (EUC), or Big5, they must be converted to Unicode for those names to be shared. Otherwise, a filename that looks fine on one computer may be garbled on another.

For example, if your Data Mover is in ASCII mode, the Data Mover saves the non-ASCII file and directory names as they are without any translation. When you enable Unicode on an ASCII Data Mover, the non-ASCII file and directory names cannot be shared with different local NFS and CIFS clients until you perform conversion to UTF-8 by using the uc_config -convert command. When you move a file system from an ASCII Data Mover to a Unicode Data Mover, you might also need to run the uc_config -convert command after you move the file system onto the Unicode Data Mover.

VNX translation support for user files and directories

The file access protocols (NFS, FTP, and CIFS) and character sets used by client systems determine whether you need to customize your VNX to translate international characters when Unicode is enabled.

For NFS and FTP clients who represent data that use characters sets that support UTF-8 format and for CIFS clients:

◆ VNX stores file and directory names as received for NFS or FTP clients who represent data that use character sets that support UTF-8 format. No translation is necessary between the data received from the client and the VNX.

◆ VNX stores file and directory names as received for CIFS clients who represent data in UTF-16 format (UCS-2). VNX stores these names as Microsoft Windows names, and

simultaneously translates them to UTF-8 format and stores them as NFS names. No translation is necessary between the data received from the client and the VNX.

Note: Although you can communicate with VNX from a Windows client by using the NFS or FTP protocols (for example, using a third-party software package such as Hummingbird NFS or FTP), the Windows client is then considered an NFS or FTP client, and requires character translation support. If you must support Windows clients by using the NFS and FTP protocols, contact the third-party software vendor and their user documentation for assistance in determining the encodings for your environment.

For all clients that require character translation, EMC Customer Support will determine the specific languages or encoding that your environment supports. This information is necessary so you can supply the correct character translation files (if not already available from EMC) and configure VNX to support the languages you require. Contact EMC Customer Support for help in determining your encoding requirements.

Translation directory structure and files

During VNX software installation, the following directories and files are created:

*   /nas/site/locale directory
*   /.etc_common/xlt common file system and translation directory

Note: If these directories do not yet exist, it is likely that the system is not configured correctly. Check with an EMC Customer Support Representative for assistance.

The directory /nas/site/locale is the source for the xlt.cfg file, unidata2.txt file, and all the required translation files as determined by you in conjunction with EMC. Character translation files on page 15 describes these files. The /.etc_common/xlt common file system and translation directory are used to store these files and make them available to a Data Mover. Any changes to these files must be first made to the copies in /nas/site/locale, and then copied to the /.etc_common/xlt directory by using the uc_config -update command.

Note: The name of the mount point for the common file system is /.etc_common. The actual name of the common file system is root_fs_common.

xlt.cfg file

The xlt.cfg file defines how VNX should translate file and directory names for NFS and FTP clients into Unicode. The file also specifies the character translation file that should be used and allows you to configure all Data Movers to use either the same character encoding or different character encoding formats for different segments of your network. If you have only CIFS clients, you do not need to edit the xlt.cfg file.

The overall structure of the xlt.cfg file and the format of text strings and variables are described in Appendix A.

By default, the xlt.cfg file assumes that all clients of all Data Movers are configured to use the same character encoding (Latin-1):

◆ If your environment supports Latin-1 encoding for all clients of all Data Movers, you do not need to edit xlt.cfg.

◆ If you use a character translation file other than Latin-1, or if you configure different character encoding formats for different segments of your network, you will need to customize the VNX to identify the client's native character encoding to the Data Mover as described in Customize translation support for user files and directories on page 28.

unidata2.txt

The unidata2.txt specifies attributes in the Unicode character database, including the uppercase or lowercase mappings of characters. The unidata2.txt is used to generate CIFS filenames.

Character translation files

Depending on your environment, VNX might use some or all of the available character translation files to interpret, translate, and store international characters in Unicode format.

Table 2 on page 16 provides a description of each file, its purpose, and the instances in which it is required to complete translations.

Note: The character translation files described in Table 2 on page 16 are those encodings that EMC has verified at the time this publication went to print, including French, German, Japanese, Korean, and Chinese. While Unicode-compliant implementations should be capable of supporting all languages, EMC does not provide any assurance that a specific language is supported by VNX until it is fully tested by EMC. Contact your EMC Customer Support Representative for a current list of verified languages or encodings, to obtain different translation files, or with questions about when to use the supplied translation files.

Table 2. Character translation files

File	Definition	Use	Details
8859-1.txt (default)	Latin-1 character translation file	Used by NFS or FTP clients in North America and parts of Western Europe who require Latin-1 character support. A superset of ASCII.	The 8859-1.txt character translation file does not require modification. Content is fixed. If your environment supports 8859-1.txt, you do not need to edit xlt.cfg to specify this translation file. This translation file should already be configured as the system default.
8859-15.txt	Latin-9 character translation file	Used by NFS or FTP clients in North America and parts of Western Europe who require Latin-9 character support. Includes additional characters, such as the euro symbol.	The 8859-15.txt character translation file does not require modification. Content is fixed.
cp437.txt	Microsoft character translation file	Contains encodings used by CIFS clients in the United States. Also referred to as Extended ASCII.	The cp437.txt character translation file does not require modification. Content is fixed.
jp-euc.txt	Japanese character translation file for EUC	Used by NFS or FTP clients in Japan. This file has additional characters, such as circle digit one, which are frequently used in Japan.	The jpeuc.txt character translation file does not require modification. Content is fixed. If your environment requires jp-euc.txt, you need to edit xlt.cfg to specify this translation file.
jp-euc1.txt	Japanese character translation file for EUC	Used by NFS or FTP clients in Japan. This file is a superset of jpeuc.txt. Duplicate translations have been commented out.	The jp-euc1.txt character translation file does not require modification. Content is fixed. If your environment requires jpeuc1.txt, you need to edit xlt.cfg to specify this translation file.

Table 2. Character translation files *(continued)*

File	Definition	Use	Details
jp-pck.txt	Japanese character translation file for SJIS	Used by NFS or FTP clients in Japan.	The jp-pck.txt character translation file does not require modification. Content is fixed. If your environment requires jp-pck.txt, you need to edit xlt.cfg to specify this translation file.
jp-pck1.txt	Japanese character translation file for SJIS	Used by NFS or FTP clients in Japan. This file is a superset of jppck1.txt. Duplicate translations have been commented out.	The jp-pck1.txt character translation file does not require modification. Content is fixed. If your environment requires jp-pck1.txt, you need to edit xlt.cfg to specify this translation file.
sjis.txt	Japanese character translation file for SJIS	Used by NFS or FTP clients in East Asia and Japan.	The sjis.txt character translation file does not require modification. Content is fixed. If your environment requires sjis.txt, you need to edit xlt.cfg to specify this translation file.
big5.txt	Chinese character translation file for Big5	Used by NFS or FTP clients in East Asia.	The big5.txt character translation file does not require modification. Content is fixed. If your environment requires big5.txt, you need to edit xlt.cfg to specify this translation file.
GB2312.txt	Chinese translation file for Simplified Chinese	Used by NFS or FTP clients in East Asia. This file is for Simplified Chinese and is based on GB2312-80.	The GB2312.txt character translation file does not require modification. Content is fixed. If your environment requires GB2312.txt, you need to edit xlt.cfg to specify this translation file.
kr-uhc.txt	Korean Hangul translation file	Used by NFS or FTP clients in Korea. This is a superset of KSC5601 and also includes extended Hangul codes designed by Microsoft.	The kr-uhc.txt character translation file does not require modification. Content is fixed. If your environment requires kr-uhc.txt, you need to edit xlt.cfg to specify this translation file.

Multiple client encodings

VNX can be configured to support more than one NFS or FTP client encoding. VNX can interpret and store names in Latin-1 and SJIS character encodings at the same time.

This is possible because character translation files include a flexible mechanism for associating character encodings with clients.

Different character encodings can be supported per Data Mover, IP subnet, IP address or hostname, or network protocol. This choice is defined in the xlt.cfg file in a hierarchical manner. Customize translation support for user files and directories on page 28 and Appendix A provide more information.

Note: When upgrading VNX software to support international characters, you can convert directories or files on a Data Mover in one local character encoding only to Unicode UTF-8. The encoding (character translation file) is configured when using the uc_config -convert command (described in Appendix B). If you need to convert more than one character encoding per Data Mover, contact your EMC Customer Support Representative for assistance.

Planning considerations

Before using Unicode and international character sets, consider the following restrictions and limitations.

Byte limitations on data element fields

The VNX data element fields that support multibyte character values are limited to a maximum number of bytes. Unicode supports character sets that use multiple bytes (one to four bytes) to represent a character while ASCII characters use only one byte. Consequently, text entered in a language that uses multibyte characters, such as Japanese, may use more of the available bytes than text entered in a language based on ASCII, such as American English. If you enter text that exceeds the maximum byte size, an error indicates the text is too long.

Discrepancies between character translation files

When configured to support Unicode, VNX can interpret, translate, and store international characters for virtually any language. However, VNX cannot map a character to a different encoding when a character does not exist in a particular encoding.

Making files and directories visible to all clients

If your VNX is configured to support more than one NFS or FTP client encoding, you may need to edit the translation files to ensure that character translation is done consistently.

Some characters in the EUC or SJIS encoding schemes are not translated on a one-to-one basis to Unicode. This is due mainly to individual vendors that implement different translation standards. For example, the Japanese character 0xA1C1 can be translated by Unicode as U+301C (Wave Dash) and U+FF5E (FullWidth Tilde).

Consequently, several of the translation files provided by VNX contain duplicate entries.

jp-euc.txt has the following duplicate entries:

```
-----------------------------------------------
0xA1C1 0xFF5E # FULLWIDTH TILDE
0xA1C1 0x301C # WAVE DASH
-----------------------------------------------
0xA1C2 0x2225 # PARALLEL TO
0xA1C2 0x2016 # DOUBLE VERTICAL LINE
-----------------------------------------------
0xA1DD 0xFF0D # FULLWIDTH HYPHEN-MINUS
0xA1DD 0x2212 # MINUS SIGN
-----------------------------------------------
0xA1F1 0xFFE0 # FULLWIDTH CENT SIGN
0xA1F1 0x00A2 # CENT SIGN
-----------------------------------------------
0xA1F2 0xFFE1 # FULLWIDTH POUND SIGN
0xA1F2 0x00A3 # POUND SIGN
-----------------------------------------------
0xA2CC 0xFFE2 # FULLWIDTH NOT SIGN
0xA2CC 0x00AC # NOT SIGN
-----------------------------------------------
```

jp-pck.txt has the following duplicate entries:

```
-----------------------------------------------
0x8160 0xFF5E # FULLWIDTH TILDE
0x8160 0x301C # FULLWIDTH TILDE (added Dec. 2001)
-----------------------------------------------
0x8161 0x2225 # PARALLEL TO
0x8161 0x2016 # PARALLEL TO (added Dec. 2001)
-----------------------------------------------
0x817C 0xFF0D # FULLWIDTH HYPHEN-MINUS
0x817C 0x2212 # FULLWIDTH HYPHEN-MINUS (added Dec. 2001)
-----------------------------------------------
0x8191 0xFFE0 # FULLWIDTH CENT SIGN
0x8191 0x00A2 # FULLWIDTH CENT SIGN (added Dec. 2001)
-----------------------------------------------
0x8192 0xFFE1 # FULLWIDTH POUND SIGN
0x8192 0x00A3 # FULLWIDTH POUND SIGN (added Dec. 2001)
-----------------------------------------------
0x81CA 0xFFE2 # FULLWIDTH NOT SIGN
0x81CA 0x00AC # FULLWIDTH NOT SIGN (added Dec. 2001)
-----------------------------------------------
```

If you have NFS clients from vendors that map characters in their filenames differently, files on one client are not visible on the other client. For example, Vendor A maps 0xA1C1 to U+301C (Wave Dash) and Vendor B maps 0xA1C1 to U+FF5E (FullWidth Tilde). Files on Vendor A that contain the Wave Dash character in the filename are not visible on Vendor B's NFS client, and files on Vendor B that contain the FullWidth Tilde character in the filename are not visible on Vendor A's NFS client.

Furthermore, NFS and FTP clients can access only the first entry for each character translation. This means that if a file is created on a CIFS client by using any entry except the first, this file cannot be accessed or managed by an NFS client. NFS clients will always use the first entry in a Unicode character translation file to map a specific character set, for example EUC to UTF-8.

To resolve inconsistent character translation, edit the translation file to best suit your environment, deleting or commenting out any additional entries for individual character translations. EMC strongly recommends that you use only one character translation per NFS and FTP client. CIFS clients are not affected. For example, if all of the Japanese clients use 0xA1C1 as Wave Dash, delete or comment out the FullWidth Tilde line.

Maximum character length for user file and directory names

The maximum length of UTF-8 strings for directory or filenames is currently limited to 255 bytes per name. System generated file system names such as <pfs>_replica1 can be more than 240 bytes but are limited to 255 bytes.

This maximum length applies to the UTF-8 version of the name, not the name as it is represented in the client encoding. This is because UTF-8 strings often take up more space when they are converted from a client encoding to UTF-8 format.

Consider the following examples:

◆ If a client encoding is Latin-1, and an ASCII name "abc" is created, the UTF-8 will look exactly the same and take the same number of bytes: 5.

◆ If a client encoding is Latin-1, and the user tries to create a name that is not entirely ASCII (such as "Äbc") the A with an umlaut will use 2 bytes when translated to UTF-8, and the entire string will require 6 bytes, one more than its Latin-1 equivalent.

If the client encoding in this example was EUC, many of the Asian characters require 2 bytes, but when translated to UTF-8 they require 3 bytes.

Set the correct locale

If the NFS client locale and the VNX encoding configuration for the NFS client do not match, VNX saves filenames as they are, without translating them. You can see the filename from the NFS client, but the filename cannot be shared with NFS clients in different locales or with CIFS clients because the filename is not saved as UTF-8. On the Unicode Data Mover, the file and directory names that are not UTF-8 that were created from NFS clients with the wrong locale appear as _cs2ConversionError~x on a CIFS client, where x is the number of the file in the directory. On an ASCII Data Mover, the non-ASCII file and directory names that are created from NFS clients appear as _____~x on a CIFS client, where x is some digit chosen to keep the name unique. To remedy this, you must ensure that you have Unicode enabled, and have the proper encoding configuration in the correct locale. Then, convert the filenames to UTF-8.

Enable Unicode on all Data Movers

When accessing file systems on VNX with Unicode enabled only on some Data Movers, you will not see the expected Unicode directory or filenames if you mount a file system by using a Data Mover that does not have Unicode enabled.

Consider the following example:

VNX has two Data Movers: server_2 has Unicode enabled and server_3 has Unicode disabled. The xlt.cfg file on server_2 is configured for a particular client to use Latin-1 encoding (character translation file 8859-1.txt).

The user on this client creates a file named ä. When the client sends this string to the server, it is encoded in Latin-1 and has the value 0xe4. The UCS-2 representation is 0x00e4, which uses two bytes. The server converts this string to UTF-8 and stores it on disk. The UTF-8 version of this name has two bytes: 0xC3 0xA4.

When the client executes the ls command on a directory, VNX reads the name from disk, and it converts the utf8 back to latin 0*e4, a single byte. Unicode cannot be passed to an NFS client. All of these translations or conversions happen because Unicode is enabled on server_2.

Alternatively, if the system administrator decides to unmount this file system from server_2 and mount it on server_3, when the client executes an ls command on the directory on server_2, VNX reads the name from disk and gets 0xC3 0xA4.

Because Unicode is not enabled, VNX will not convert or translate the name, but will send these bytes directly to the client where it will display the characters A˜$.

To correct this problem you must either enable Unicode on server_3, or move the file system back to server_2.

When you move the file system from an ASCII Data Mover to a Unicode Data Mover, you might need to run the uc_config -convert command to convert the file system, unless you are sure that there are no non-ASCII file or directory names in the file system.

Multiple non-ASCII encodings limitation

You should not have multiple non-ASCII encodings in the same file system. Conversion problems can occur if you have multiple non-ASCII encodings. For example, in the EUC encoding, there is a character A4A2. In the SJIS encoding, there is an A4 and an A2 character. If you convert SJIS first, the single EUC A4A2 character might be converted into the separate SJIS characters of A4 and A2. If this happens, the result cannot be reversed. You must fix the file manually from a CIFS or NFS client. If there is a separate EUC directory and SJIS directory on the same file system, you can convert the file systems one by one by running the EUC conversion first and ensuring a backup occurs before performing any conversion.

Note: As a best practice, translate the larger encoding first. Perform the EUC encoding before SJIS encoding and all Asian sets before encoding Latin. This will translate everything remaining if a full Latin set (like cp437) is used.

There are other code conflicts besides this example. This limitation should be considered before performing any operation relating to code translation and conversion.

Using share-level security on a Data Mover

Share-level security is not supported by VNX when Unicode is enabled. If you want to reject non-ASCII file and directory names, set the shadow asciifilter parameter to 1 on the Data Mover that uses SHARE or UNIX security.

Compatibility issue with Macintosh

The Macintosh operating system (Mac OS X and later for NFS) might have an issue when Unicode file or directory names are shared with other operating system clients, such as Sun Solaris and Microsoft Windows. The Macintosh operating system tries to decompose all decomposable characters. However, some other operating systems, like Solaris and Windows, do not. VNX can handle decomposed characters and nondecomposed characters without a problem. This is a compatibility issue between the Macintosh operating system and other operating systems.

Unicode and replication

Because the files on the destination site are read-only, you cannot touch them directly with the conversion process. Conversion touches every name (not the actual file) in the source site and then, as replication is taking place, the names that have been touched are updated to the destination site.

If your source site and destination site are both currently ASCII and you want to enable Unicode, you must perform the following steps:

1. Set up the translation files for the source and destination sites. Edit the xlt.cfg file on page 28 provides more information on setting up translation files.

2. Start Unicode conversion at the source site. Start Unicode conversion on page 30 provides more information on starting Unicode conversion.

3. Find and convert files on the source site. Find and convert files on page 32 provides more information about converting files.

4. Stop conversion on the source site when all the files have been converted. Also, ensure that clients who connect to the destination site can access the files.

5. Check the destination site to ensure that the changes have been transferred.

Unicode and FileMover

If you have a FileMover-enabled file system mounted on a Data Mover, you must have Unicode enabled. Even if Unicode is enabled, any non-ASCII filename that has not been converted to UTF-8 cannot be migrated by using the policy and migration software. You must convert the non-ASCII filenames first before using FileMover. This applies to non-ASCII filenames created from a NFS client with a locale that does not match the

VNX encoding information for the client. The primary VNX file server, secondary file server (VNX or UNIX), and policy and migration software all must use the same locale.

Enable Unicode

EMC strongly recommends that you enable Unicode on every Data Mover in your VNX system. You should do this before creating any file systems. You can enable Unicode by using one of the following methods:

- Through the VNX Installation Assistant for File/Unified which automatically enables Unicode on all Data Movers

- Through the Setup Wizard for File available in Unisphere software

- On the Data Mover Properties page available in Unisphere software

- Issuing the uc_config -on command in the VNX for file CLI

Note: The uc_config -on command does not convert existing file or directory names to Unicode UTF-8 format when upgrading VNX that supports non-ASCII characters. To convert existing files and directories, complete the steps as described in Converting user files and directories to Unicode.

⚠ **CAUTION** After enabling Unicode, you cannot disable it. In other words, you cannot return to ASCII mode from Unicode mode.

Multibyte support for VNX data elements

In the VNX management interfaces, multibyte support allows users to input, store, and display multibyte characters. It enables the users to input and view VNX object fields in their native language. The multibyte feature is supported since Celerra NAS 5.6 and later. The following data elements are enabled with multibyte support and also support the Unicode 3.0 standard:

- CIFS Share name

- CIFS Share comment

- CIFS export path

- Tree quota path

- Tree quota comment

- Quota user name

- CIFS server netbios name

- CIFS server computer name

- CIFS server alias

- CIFS server comment

- CIFS domain name
- CIFS workgroup
- Organizational Unit (OU) name
- DFS root name
- CDMS source share name
- CDMS local path
- CDMS source export path
- Mount point name
- File system name
- Checkpoint name
- NFS export name
- NFS export path

Configuring Language Support

To create and display the VNX data element names in the VNX management interfaces in a language that uses multibyte character sets, you must configure your operating system and terminal software to use the appropriate locale, language, font, and character encoding scheme.

Topics included are:

- Configure the VNX for File GUI on page 26
- Configure the VNX for File CLI on page 26

Configure the VNX for File GUI

By default, Windows installs the files for most input languages supported by Windows. To set the locale and language appropriate for your environment, go to **Control Panel ➤ Regional and Language Options**.

You may need to install these language files if you enter or display text in any of these languages:

* East Asian languages, such as Chinese, Japanese, or Korean. To use East Asian languages, select **Install files for East Asian languages** option.

* Complex script and right-to-left languages, such as Arabic, Armenian, Georgian, Hebrew, the Indic languages, Thai, or Vietnamese.

Note: You may see some problems with the display of numbers when using a right-to-left language. Currently, numerical content gets reversed, displaying an incorrect value. If you are using a right-to-left language, you may need to temporarily set your locale to English to view numerical content in the correct order. After you have noted the correct value, you can reset your locale as appropriate for your environment.

Configure the VNX for File CLI

Your terminal software or UNIX OS terminal must be installed with the appropriate language files and set to expect received data in UTF-8 character encoding format.

For example, to specify UTF-8, use the terminal emulator PuTTY and go to **Window ➤ Translation** and set the Character set translation on received data field to UTF-8. In addition, select the font appropriate for your language.

If your terminal software or UNIX OS terminal is not set to UTF-8, you will receive an error "Input must consist of ASCII or UTF-8 characters", if you enter a multibyte character.

Managing

Topics included are:

- Customize translation support for user files and directories on page 28
- Convert user files and directories to Unicode on page 30
- Implement international character support with NDMP on page 33

Customize translation support for user files and directories

The file access protocols (NFS, FTP, and CIFS) and character sets used by the client systems determine whether you need to customize your VNX to translate international characters when Unicode is enabled:

1. Determine if your environment requires character translation support. Chapter 2 provides the information you need to consider.

2. Identify and supply any needed character translation files and define the required files in the xlt.cfg file. Edit the xlt.cfg file on page 28 provides more details.

3. Copy the edited xlt.cfg file and character translation files to /.etc_common/xlt so the files can be seen by Data Movers. Copy the configuration and character translation files on page 28 provides more details.

4. Verify the configuration of the xlt.cfg file. Verify translation configuration settings on page 29 provides more details.

5. Enable Unicode encoding as described in Enable Unicode on page 23.

Edit the xlt.cfg file

Before beginning, familiarize yourself with the overall structure of the xlt.cfg file and the format of text strings and variables as described in Appendix A. To edit the xlt.cfg file, you must perform the following steps:

1. Using a UNIX text editor, open the xlt.cfg file located in /nas/site/locale.

   Note: The xlt.cfg file is automatically installed in the /nas/site/locale directory during VNX software installation.

2. Add or modify the existing text strings for your environment.

   Note: Table 2 on page 16 lists the character translation files supplied by EMC.

3. Save and close the file.

Copy the configuration and character translation files

Action
To copy the configuration and character translation files from /nas/site/locale to /.etc_common/xlt so Data Movers can access these files to complete the requested translations, use this command syntax:
$ **/nas/sbin/uc_config -update** *<filename>*

Action
where:
<filename> = name of the specific file to copy
Examples:
To update all files in /.etc_common/xlt, type:
$ /nas/sbin/uc_config -update
To update a specific file (for example, xlt.cfg) in /.etc_common/xlt, type:
$ /nas/sbin/uc_config -update xlt.cfg

Output
| ```
operation in progress (not interruptible)...
done
``` |

Note: If no filenames are specified, all files in /nas/site/locale are copied to the /.etc_common/xlt directory.

Verify translation configuration settings

If you edited the xlt.cfg file, confirm that the configuration of xlt.cfg is correct for your environment. If you have several clients or subnets that use different encodings, EMC strongly recommends testing each client or subnet individually to ensure xlt.cfg is configured correctly.

Action
To verify the translation settings for a specific client or subnet, use this command syntax:
$ **/nas/sbin/uc_config -verify** {*<IP_address>*|*<hostname>*} **-mover** *<movername>*
where:
{*<IP_address>*|*<hostname>*} = IP address or hostname of a particular client whose configuration that you want to verify
<movername> = name of the Data Mover whose configuration that you want to verify
Example:
To confirm that the configuration of xlt.cfg is correct for a client with an IP address of 168.159.30.77 for server_3, type:
$ /nas/sbin/uc_config -verify 168.159.30.77 -mover server_3

Output	Note
server_3: 168.159.30.77 is UTF-8	This option verifies the configuration of xlt.cfg for a particular client's IP address or hostname. The xlt.cfg file is scanned to locate the text string that applies to the specified client. Once found, the system informs you whether the search was successful and displays the IP address of the client, and the character encoding the client is configured to support.

Convert user files and directories to Unicode

The steps to upgrade VNX to support the use of language characters defined by the Unicode 3.0 standard for user file and directory names require converting existing file and directory names on disk to Unicode format. The size of your file systems and directories determines the length of time the conversion process takes. After enabling Unicode conversion, each file and directory is converted as accessed by users. Unicode conversion remains enabled until you disable it:

Note: If during conversion a character in a file or directory name cannot be found in the conversion file, the entire file or directory name is not converted.

1. Start Unicode conversion as described in Start Unicode conversion on page 30.

   Note: When the conversion process starts, Unicode is automatically enabled.

2. Conversion occurs when files are accessed by users although you can manipulate when this occurs. Find and convert files on page 32 provides details.

3. Stop Unicode conversion as described in Stop Unicode conversion on page 33.

4. If necessary, rerun the Unicode conversion. Rerun the conversion on page 33 provides details.

Start Unicode conversion

Before you begin

Before you begin the conversion, perform the following tasks. Failure to do so might jeopardize the upgrade process:

- Back up your file system.
- Verify that the translation directories were created and that the required files are installed.

- Determine if your environment requires character translation support and, if so, complete the steps as described in Customize translation support for user files and directories on page 28.

Keep the following guidelines in mind:

- For a relatively small number of file systems and directories, the conversion should not take long and you should not notice any impact to system performance.

- When converting a large number of file systems and directories, you might experience some delay as file or directory names are converted.

- If you have several very large directories, you might want to perform the conversion during a period of low network activity.

- You can convert directories or files to Unicode from one local character encoding only. The desired character translation file is configured in the uc_config -convert command, for example, big5.txt. This indicates that the Chinese character encoding, Big5, should be used to translate the directories or files on the specified Data Mover. If your environment supports more than one character encoding per Data Mover, contact EMC Customer Support for assistance in converting your system.

Procedure

EMC recommends leaving Unicode conversion enabled to allow users to access (and convert) the majority of files or directories. The length of time this takes depends on the size of your file system and directories.

Action
To start Unicode conversion, use this command syntax:
$ **/nas/sbin/uc_config -convert start** *<filename>* **-mover** {*<movername>*|**ALL**}
where:
<filename> = name of the character translation file that should be used to convert data to Unicode. Note that only one character encoding is supported per Data Mover.
<movername> = name of the Data Mover for which you want to start Unicode conversion. This option causes Unicode conversion to start for all file systems on the specified Data Mover.
If ALL is used, Unicode conversion starts for all Data Movers in VNX cabinet.
Example:
To start a Unicode conversion that uses Big5 encoding for all file systems on server_3, type:
$ **/nas/sbin/uc_config -convert start big5.txt -mover server_3**

Output
| ```
You are using the translation file (big5.txt), which will translate
BIG5 to Unicode
It is recommended that you back up your file system before running
this command. Have you backed up your file system? Please make sure
that you are using the right translation file, otherwise data on the
disk will be corrupted and unrecoverable!
Do you wish to continue [yes or no]: yes
``` |

Note
The start option enables Unicode conversion only. The actual conversion of a file system does not start until the file system is accessed by a user.
Unicode conversion remains enabled until you disable it.

Find and convert files

You can find and convert files at any time after enabling Unicode. Running find or search commands immediately after starting the Unicode conversion causes each file or directory name to be enumerated and converted without waiting for users to specifically access them. This is recommended if you want to convert everything at once.

To convert files gradually, run the find or search commands after most or all of the directories have been accessed. The purpose of running the find or search commands is to catch a small number of directories or files that have not been accessed yet. Determine the best timing for your system before deciding when to run these commands.

Type the find command from any UNIX or Linux client, or perform a search in a Windows environment so that every directory on each effected Data Mover is found. This enumerates and converts to Unicode files in each directory (if not already done). For example, find <mountpath> - name randomNonexistentFile.

Windows search guidelines

You can perform the following steps for an effective search:

◆ Ensure that all subfolders are searched. It does not matter what name you search for as long as you wait for the search to complete. EMC recommends using a name that does not exist so no matches are found. All directories and files will be found and converted without interruption.

◆ Ensure that all directories on a Data Mover are found and not just those in a particular share.

◆ If directories or files are not accessible to CIFS clients, perform the search from a UNIX or Linux client instead.

Stop Unicode conversion

Action
To stop Unicode conversion after converting all files and directories, use this command syntax:
$ **/nas/sbin/uc_config -convert stop** *<filename>* **-mover** {*<movername>*|**ALL**}
where:
<filename> = name of the character translation file that is being used to convert data to Unicode.
<movername> = name of the Data Mover for which you want to stop Unicode conversion. This option causes Unicode conversion to stop for all file systems on the specified Data Mover.
If ALL is used, Unicode conversion stops for all Data Movers in VNX cabinet.
Example:
To stop Unicode conversion (Big5 encoding) for all file systems on a specified Data Mover, type:
$ /nas/sbin/uc_config -convert stop big5.txt -mover server_3
Output
server_3: done

Rerun the conversion

When you run the uc_config -convert start command, VNX creates a timestamp. If the timestamp of a directory is older than the conversion timestamp, VNX converts the files in the directory. After conversion, the timestamp of the directory is later than the timestamp of the directory before conversion. The next user access of the directory does not trigger conversion.

However, if you need to run the uc_config -convert start command again, for example in a case where the NFS client had the wrong locale, you can simply reissue the uc_config -convert start command to touch all the directories again. Only unconverted files are updated. Files that are already converted are skipped.

Implement international character support with NDMP

Ensure that the following issues are addressed when performing NDMP backups with Unicode enabled:

◆ Verify vendor support on page 34

◆ Data backup behavior on page 34

◆ Set parameters for backups on page 35

◆

Verify vendor support

Verify that the EMC-qualified NDMP backup vendor you choose supports international character sets. The backup vendor's software support for international characters with NDMP will determine which of the following two VNX implementations you need to follow.

Note: If international character sets are not supported by your vendor, review the information in this document for details about choosing the best character translation file for your environment.

Vendor supports international characters

If your vendor supports international character sets, follow the vendor instructions for implementation and follow up with your backup software vendor regarding any issues you encounter.

You must set the parameters for Data Movers for NDMP and for international character sets. *Configuring NDMP Backups on VNX* provides more details about setting these parameters.

⚠ **CAUTION** The EMC E-Lab™ Interoperability Navigator and your vendor's documentation provide information to determine if the vendor is qualified and supports the use of international character sets.

Vendor does not support international characters

If Unicode is enabled on your VNX, but your backup vendor does not support international character sets, you might set up your environment to work with the best character translation file for your environment. The default character encoding format is UTF-8. VNX translation support for user files and directories on page 13 provides details about character translation files and choosing the appropriate file for your configuration. You must also set the parameters as specified in *Configuring NDMP Backups on VNX*.

Data backup behavior

Data is backed up onto tape by using either the tar, dump, or VBB backup type. If you have chosen a character translation file, the filename is converted to the client character encoding and sent to the client as the file history. During an NDMP tar or dump backup, if a file is found that the character translation file cannot translate, the catalog information for that file will not be sent to the backup client. An error log is created, the data is backed up without the catalog information, and the rest of the NDMP backup continues normally. The file with the catalog information that could not be backed up can be recovered only through a full restore of the directory.

Set parameters for backups

If your backup software does not support Unicode, and a backup is done with Unicode enabled on the VNX, the filename must be converted.

In this case, you need to set the dialect (also known as the character encoding) by using the following procedure:

Note: During a backup, if a file is found that the character translation file cannot translate, an error log is created. For example, if a file was created by using a character that does not appear in the code page's character set, or if an NDMP dialect does not match the dialect used to create the file, then an error log is created.

1. To set the dialect (character encoding), use this command syntax:

   $ **server_param** *<movername>* **-facility NDMP -modify dialect -value** *<dialect string>*

   where:

   <movername> = name of the Data Mover

   <dialect string> = dialect used on the ASCII mode Data Mover. The default value is 8859-1 (Latin-1).

2. To reboot the Data Mover, use this command syntax:

   $ **server_cpu** *<movername>* **-reboot -monitor now**

   where:

   <movername> = name of the Data Mover

Specify a conversion dialect

A conversion dialect must be specified while in internationalization mode when restoring a backup image of an ASCII mode Data Mover where a non-Latin dialect was used. Perform the following steps:

1. To specify a conversion dialect (character encoding), use this command syntax:

   $ **server_param** *<movername>* **-facility NDMP -modify convDialect -value** *<di alect string>*

   where:

   <movername> = name of the Data Mover

   <dialect string> = dialect used on the ASCII mode Data Mover. The default value is 8859-1 (Latin-1).

2. To reboot the Data Mover, use this command syntax:

```
$ server_cpu <movername> -reboot -monitor now
```

where:

\<movername\> = name of the Data Mover

Troubleshooting

As part of an effort to continuously improve and enhance the performance and capabilities of its product lines, EMC periodically releases new versions of its hardware and software. Therefore, some functions described in this document may not be supported by all versions of the software or hardware currently in use. For the most up-to-date information on product features, refer to your product release notes.

If a product does not function properly or does not function as described in this document, contact your EMC Customer Support Representative.

Problem Resolution Roadmap for VNX contains additional information about using EMC Online Support and resolving problems.

Topics included in this chapter are:

EMC E-Lab Interoperability Navigator

The EMC E-Lab™ Interoperability Navigator is a searchable, web-based application that provides access to EMC interoperability support matrices. It is available on EMC Online Support at http://Support.EMC.com. After logging in, in the right pane under **Product and Support Tools**, click **E-Lab Navigator**.

VNX user customized documentation

EMC provides the ability to create step-by-step planning, installation, and maintenance instructions tailored to your environment. To create VNX user customized documentation, go to: https://mydocs.emc.com/VNX.

Using network trace

If a Unicode file or directory name problem is seen, get a network trace on the NFS or CIFS client. The network trace shows where the filename is incorrect. If the network trace has the expected filename codes, the problem might be locale configuration on the NFS client.

Error messages

All event, alert, and status messages provide detailed information and recommended actions to help you troubleshoot the situation.

To view message details, use any of these methods:

◆ Unisphere software:

   • Right-click an event, alert, or status message and select to view Event Details, Alert Details, or Status Details.

◆ CLI:

   • Type nas_message -info <MessageID>, where <MessageID> is the message identification number.

◆ *Celerra Error Messages Guide*:

   • Use this guide to locate information about messages that are in the earlier-release message format.

◆ EMC Online Support:

- Use the text from the error message's brief description or the message's ID to search the Knowledgebase on EMC Online Support. After logging in to EMC Online Support, locate the applicable **Support by Product** page, and search for the error message.

EMC Training and Professional Services

EMC Customer Education courses help you learn how EMC storage products work together within your environment to maximize your entire infrastructure investment. EMC Customer Education features online and hands-on training in state-of-the-art labs conveniently located throughout the world. EMC customer training courses are developed and delivered by EMC experts. Go to EMC Online Support at http://Support.EMC.com for course and registration information.

EMC Professional Services can help you implement your system efficiently. Consultants evaluate your business, IT processes, and technology, and recommend ways that you can leverage your information for the most benefit. From business plan to implementation, you get the experience and expertise that you need without straining your IT staff or hiring and training new personnel. Contact your EMC Customer Support Representative for more information.

Appendix A

The xlt.cfg File

EMC strongly recommends that you familiarize yourself with the following information to learn the format of the xlt.cfg file and how to edit the xlt.cfg file:

xlt.cfg file format example

The xlt.cfg file is an ASCII text file. It contains a series of text strings, each consisting of up to six fields, that define the character translation files to be used by a Data Mover to translate a client's file or directory names into Unicode format. You can configure all Data Movers to support the same character translation, or you can configure different character encoding formats for different segments of your network, including:

◆ Data Mover

◆ IP subnet

◆ IP address or hostname

◆ Network protocol

```
#The line below identifies the meaning of each of the fields in this file
#All of these fields can have data in them except for the reserved field
#which must be left empty (no characters, not even whitespace).
#
#<server>:<protocol>:<address>:<reserved>:<translate>:<comment>
#
#Each of the lines below are an example of how to use some of the features
#a typical installation will use. These lines are ordered from most specific
#to least specific, since the first line that matches a client's request
#will be the one that actually applies. To adapt one of the lines below
#for a particular installation, simply remove the leading '#' and change
the
#address field and the translation file to be correct for the particular
network.
#
#The line below shows how to identify a particular client which uses an
#encoding that is different from other clients.
#
#::168.159.6.0::8859-1.txt:Clients on the 168.159.6.0 use latin-1
#
#the default is that all clients that don't match any of the earlier lines
#will be assumed to use latin-1
::::8859-1.txt:Anything that didn't match above will be assumed to be latin-1
```

Edit the xlt.cfg

You need to edit xlt.cfg in the following cases:

◆ If your environment supports UNIX, Linux, or CIFS clients that use the NFS or FTP protocol, and you require a character encoding format other than the default, Latin-1 (character translation file 8859-1.txt).

◆ If you need to specify different character encoding formats for different segments of your network, that is, for UNIX, Linux, or CIFS clients that use the NFS or FTP protocol.

When editing the xlt.cfg file:

◆ Specify only the number of lines required by your operating environment; performance could be impacted in systems that specify large numbers of translation strings. Also, be sure to remove translation strings that are no longer being used.

◆ Arrange text strings in a hierarchical order, that is, from the most specific to the least specific. When the file is parsed, the first line that matches a client request is used.

◆ The file is parsed from left to right.

◆ No translations occur in the following situations:

  • If xlt.cfg is not stored in the common translation directory, all clients are treated as Unicode clients.

  • If no lines in xlf.cfg apply, the character string is treated as UTF-8.

  • If no character translation files are defined in a particular line, the system does not translate any characters. For example:

    ```
    ::168.159.30.77:::no translations occur for this particular client
    ```

Special characters

describes the special characters that are supported when editing xlt.cfg.

Table 3. Special characters

Special character or delimiter	Description	Function
#	Comment symbol	Indicates that a line should be skipped when parsing the configuration file. If you want a line to be read, remove the #.
:	Delimiter	Separates fields in the text string. If a field is skipped, an entry is considered to be NULL.

Text string format

The format of text strings in xlt.cfg is as follows:

```
<server>:<protocol>:<address>:<reserved>:<translation>:<comment>
```

provides information for a definition of each field and sample text strings. The examples in the table assume that the default character translation file, Latin-1 (8859-1.txt), is used.

Note: The maximum length of each line in the xlt.cfg file is 255 characters.

Table 4. Text string formats

Field	Description	Example text strings
`<server>`	Identifies the default translation to be used for a particular server (Data Mover). The value can be the default server name for the Data Mover (for example server_3) or a user-specified name (for example beta_3). If this field is set, the protocol and address fields should be left empty. If you change the name of a Data Mover, you must edit xlt.cfg to add the correct name.	`server_3:::::8859-1.txt:this Data Mover defaults to latin-1`
`<protocol>`	Allows you to specify translations by protocol. The following values are supported: ◆ All ◆ NFS ◆ FTP For example, if set to NFS, all translation requests received from NFS clients will be processed as specified in this text string.	`:nfs:::8859-1.txt:all nfs clients use latin-1`

`<address>`
or
`<host name>`
	Specifies the address or hostname of a particular client or subnet. The address can be the ASCII name, IP address of a client, or a subnet address for all clients on that subnet.
	
	The syntax for a subnet address is as follows:
	
	subnet address, subnet mask.
	
	Wildcards are not supported when specifying a subnet address or mask for Unicode.
	
	Hostnames are supported if one of the following is enabled for a Data Mover:
	
	◆ NIS
	
	◆ DNS
	
	◆ /.etc/hosts on the Data Mover
	
	The hostname can be the DNS or NIS name assigned for a Data Mover.
		Example using an IP address:
		`::168.159.30.77::8859-1.txt:this client uses latin-1`
		
		Example using a subnet address:
		`::10.15.20.0,255.255.255.0::big5.txt:Clients on the 10.15.20.0 subnet use big5 encoding`
		
		Example using an ASCII client name:
		`::lss0163::8859-1.txt:this client uses latin-1`
		
		Example using a DNS hostname:
		`::lns0077.emc.com::8859-1.txt:this host uses latin-1`

`<reserved>`	Reserved for future development. Leave empty.	Not available

Table 4. Text string formats *(continued)*

Field	Description	Example text strings
`<transla-tion>`	Specifies the name of the file to be used for the translation. Specify the name of each translation file your environment requires in a separate line in xlt.cfg. EMC supplies a default translation file, Latin-1 8859-1.txt, as shown in the example. The translation file you specify in xlt.cfg must be located in /.etc_common/xlt. If it is not there, this field is ignored. **REVIEWERS: Are there any limitations on the length of characters in the <translations> field. Are spaces supported?**	`::168.159.30.77::8859-1.txt:this client uses latin-1`
`<comment>`	Reserved for the system administrator's use. It can contain descriptive comments regarding the structure of this file, or it might be left blank.	`::168.159.30.77::8859-1.txt:this client uses latin-1`

Appendix B

The uc_config Command

This section describes the usage of the uc_config command with examples:

◆

uc_config

Allows you to manage the character translation files needed for VNX international character support.

Synopsis

```
uc_config
 -setup | -info | -list | -help
 -update [<filename> <filename>..]
 -convert {start <filename> | stop} -mover {<movername> | ALL}
 -verify {<IP_address> | <hostname>} -mover <movername>
 -on -mover <movername>
```

Description

The uc_config command is used for manipulating the character translation files needed for VNX international character support. These files are stored in a translation directory on a common file system, /.etc_common/xlt, that is available read-only to all Data Movers.

The translation directory is created when the VNX software is installed on the Control Station.

Options

-setup

Checks for the existence of the common file system and translation subdirectory and creates the subdirectory if needed.

This option is used by EMC Customer Support engineers to determine if the common file system and translation directory are created. They should have been created when the VNX software was installed. If the common file system does not yet exist, it is likely that the system is not configured correctly. Contact an EMC Customer Support Representative for assistance.

-info

Runs a quick test on the common file system configuration and prints out the observed status. The status includes whether the common file system exists, and a list of the servers on which the common file system is mounted read-only.

Note: The output for the uc_config -info command is listed in the order in which server entries were initially created, not necessarily in sequential order.

-list

Lists the set of translation files for conversion or update.

-help

Displays a more verbose usage explanation.

-update[<filename>.. <filename>..]

Copies the current contents of /nas/site/locale to /.etc_common/xlt on the common file system.

If any filenames are given as options, only those files are copied to the common file system. Otherwise, if no filenames are specified, all files found in /nas/site/locale are copied.

If any updated files already exist on the common file system, they are overwritten with the content from /nas/site/locale.

-convert{**start** <filename> | **stop**} **-mover** {<movername> | **ALL**}

Starts or stops Unicode UTF-8 conversion for all file systems on a specified Data Mover or for all Data Movers in VNX cabinet.

The start option enables Unicode UTF-8 conversion only. The actual conversion of a file system does not start until the file system is accessed by a user.

<filename> is the name of the character translation file that should be used to convert data to UTF-8. This field is mandatory. VNX cannot complete the conversions without a specified translation file.

If a <movername> is specified, UTF-8 conversion starts or stops for all file systems on the specified Data Mover. Otherwise, if ALL is used, UTF-8 conversion starts or stops for all Data Movers in VNX cabinet.

Note: Currently, you can upgrade a Data Mover in one character encoding only to UTF-8. If your environment supports more than one character encoding, contact EMC's Customer Support for assistance in converting your system.

-verify {<IP_address> | <hostname>} **-mover** <movername>

Allows you to confirm that the configuration of xlt.cfg is correct for your environment.

When invoked, this option verifies the configuration of xlt.cfg for a particular client's IP address or hostname with regard to a Data Mover.

The xlt.cfg file is scanned to locate the text string that applies to the specified client. Once found, the system informs you if the search was successful and displays the IP address of the client, and the character encoding it is configured to support.

If you have several clients or subnets that use different character encodings, EMC strongly recommends testing each individually to ensure xlt.cfg is configured correctly.

-on -mover <movername>

Allows you to enable Unicode UTF-8 encoding for one or more Data Movers.

Note: Do not attempt to use /nas/sbin/uc_config -on when upgrading VNX that supports non ASCII characters. Instead, complete the steps described in . Unicode is automatically enabled when the conversion process starts.

Example 1

To check for the existence of the common file system and translation subdirectory and create them if necessary, type:

```
$ /nas/sbin/uc_config -setup
```

If you receive the following message, your setup is correct:

```
Common Unicode translation subdirectory already exists.
```

If you do not receive this message, contact EMC Customer Support for assistance.

Example 2

To run a quick test on the common file system configuration and print the observed status, type:

```
$ /nas/sbin/uc_config -info
```

Output:

```
common filesystem [root_fs_common] exists
common filesystem is presently mounted read-only on:
server_2
server_3
common filesystem is not mounted read-write anywhere
```

Note: The output for the uc_config -info command is listed in the order in which server entries were initially created, not necessarily in sequential order.

Example 3

To list the set of translation files available for conversion or update, type:

```
$ /nas/sbin/uc_config -list
```

Output from Control Station /nas/site/locale:

```
8859-15.txt cp437.txt jp-euc1.txt jp-pck.txt sjis.txt xlt.cfg
8859-1.txt eucjp1.txt jp-euc.txt kr-uhc.txt unidata2-old.txt xlt.crc
big5.txt GB2312.txt jp-pck1.txt SCCS unidata2.txt
```

Output from Data Mover /.etc_common/xlt:

```
8859-15.txt 8859-1.txt big5.txt cp437.txt eucjp1.txt GB2312.txt
jp-euc1.txt jp-euc.txt jp-pck1.txt jp-pck.txt kr-uhc.txt sjis.txt
unidata2-old.txt unidata2.txt xlt.cfg xlt.crc
```

If the two lists do not match, with the exception of SCCS, run uc_config -update.

Note: unidata2-old.txt, unidata2.txt, xlt.cfg, xlt.crc, and SCCS are not translation files.

Example 4

To update all configuration files in /.etc_common/xlt, type:

```
$ /nas/sbin/uc_config -update
```

Output:

```
operation in progress (not interruptible)...
done
```

Example 5

To update a specific configuration file (for this example, xlt.cfg) in /.etc_common/xlt, type:

```
$ /nas/sbin/uc_config -update xlt.cfg
```

Output:

```
operation in progress (not interruptible)...
done
```

Example 6

To start a Unicode conversion that uses Big5 encoding for all file systems on server_3, type:

```
$ /nas/sbin/uc_config -convert start big5.txt -mover server_3
```

Output:

```
You are using the translation file (big5.txt), which will translate
BIG5 to Unicode
It is recommended that you back up your file system before running this
command. Have you backed up your file system? Please make sure that you
are using the right translation file, otherwise data on the disk will
be corrupted and unrecoverable!
Do you wish to continue [yes or no]: yes
server_10 : done
```

Currently, you can upgrade a Data Mover in one local character encoding only to UTF-8. If your environment supports more than one character encoding, contact EMC Customer Support for assistance in converting your system.

Example 7

To stop Unicode conversion (Big5 encoding) for all file systems on a specified Data Mover, type:

```
$ /nas/sbin/uc_config -convert stop big5.txt -mover server_3
```

Output:

```
server_3 : done
```

Example 8

To confirm that the configuration of xlt.cfg is correct for a client with an IP address of 168.159.30.77 for server_3, type:

```
$ /nas/sbin/uc_config -verify 168.159.30.77 -mover server_3
```

Output:

```
server_3:
168.159.30.77 is UTF-8
```

Example 9

To enable Unicode for one or more Data Movers, type:

```
$ /nas/sbin/uc_config -on -mover server_2 server_3
```

Output:

```
server_2 : done
server_3 : done
```

Glossary

A

ASCII (American Standard Code for Information Interchange)
7-bit (one byte) character-encoding scheme for representing text and controlling printers and computer communication based on the ordering of the English alphabet. US-ASCII was the most commonly used character encoding on the Internet until it was recently surpassed by UTF-8. ASCII was incorporated into the Unicode character set as the first 128 symbols, so the ASCII characters have the same numeric codes in both sets.

C

character encoding
The mapping from a coded character set to a sequence of bits.

character set
Set of characters used by a language (ANSI/ASCII or Unicode). The character set affects file format (how the software stores and transmits data) and string processing (the logic with which text is manipulated).

CIFS
See Common Internet File System.

D

data element
An attribute in VNX for file derived from a string.

F

File Transfer Protocol (FTP)
High-level protocol for transferring files from one machine to another. Implemented as an application-level program based on the OSI model, FTP uses the TCP protocol.

G

globalization (G11N)
The process of internationalizing and localizing software.

I

I18N
An abbreviation for internationalization (I + 18 letters + N).

See internationalization.

J

JIS (Japanese Industrial Standards)
Multibyte character encoding system used to represent Japanese. Includes Shift JIS (SJIS) encoding. The main alternatives to JIS encoding are EUC (used on UNIX systems where the JIS encodings are incompatible with POSIX standards) and more recently Unicode, particularly in the form of UTF-8.

L

locale
A set of user preferences for a particular country or region including language (for example, en-US for American English) and associated settings such as the formatting and parsing of dates, times, numbers, and currencies, units of measure, sort-order (collation), and translated names for time zones, languages, and countries.

localization (L10N)
The process of adapting software for a specific culture or language by adding locale-specific components and translating text.

M

multibyte support
Generic term for encoding schemes for characters that may require more than one byte.

N

network file system (NFS)
Network file system (NFS) is a network file system protocol that allows a user on a client computer to access files over a network as easily as if the network devices were attached to its local disks.

U

UCS-2 (Universal Character Set-2)
An obsolete character encoding which is a predecessor to UTF-16.

Unicode
Family of universal character encoding standards used for representation of text for computer processing.

Unicode Consortium

Organization responsible for defining the behavior of and relationships among Unicode characters and for providing technical information to implementors (www.unicode.org).

UTF-16 (16-bit Unicode Transformation Format)

Multibyte encoding form that supports the 16-bit Unicode or "wide" characters used by Windows. UTF-16 is an extension of UCS-2.

UTF-8 (8-bit Unicode Transformation Format)

Multibyte encoding form that uses an algorithmic mapping scheme to convert every Unicode value to a unique 1- to 4-byte sequence with no embedded null characters.

See *Unicode* or *UCS Transformation Format-8*.

Glossary

Index

User interface choices 8

V

vendor support 34

X

xlt.cfg file format 42