



PROFESSIONAL SERVICES TECH NOTE #041

Unicode

The Concept

Computers encode alphanumeric characters, symbols and punctuation with a unique number. In the 8-bit processor days an encoding system like ASCII could only represent 256 characters. To handle all characters and languages, multiple encoding systems were created. Problems arose when trying to share data between computers using different encoding systems. The Unicode Standard was created as a universal encoding scheme (according to Wikipedia it now covers 107,000 characters). The Unicode format is also often referred to as UTF-8 or UTF-16.

The NFS protocol was created before Unicode (1984 vs. 1991) and was designed to be localized so it does not encode in Unicode. The CIFS protocol encodes with Unicode by default. Therefore, CIFS directories are directly accessible to NFS clients, but NFS directories are not directly accessible to CIFS clients. To provide a CIFS client with access to an NFS directory, your storage system must first convert the NFS directory to Unicode format. This is done on the fly when ONTAP receives the file access request. If a large amount of data is requested, the Unicode conversion can take time and consume storage system resources. To reduce the overhead, Data ONTAP provides volume options to turn Unicode encoding and automatic conversion for new files on or off.

Examples

```
fas1> vol options User_Homes create_unicode on - all new files will be created in Unicode format
```

```
fas1> vol options Groups_vol convert_unicode on - all current non-Unicode files and any files later copied/moved into the volume will be converted to Unicode format.
```

What It Means To You

Unicode support is required on systems licensed for both CIFS and NFS but only on NFS volumes and qtrees which will also be accessed by CIFS clients. Volumes with FCP/iSCSI LUNs always require `create_unicode` be on. NetApp recommends enabling Unicode support at volume/qtree creation time to alleviate the resource consuming conversion later. If you enable these options on the root volume then all new volumes will inherit the settings (Data ONTAP automatically enables `create_unicode` on the root volume when configuring new systems with a CIFS license so CIFS clients can access Snapshots). Essentially, enable the options unless your system only uses NFS.

Note: Do not enable the `convert_unicode` option on unconverted directories which contain more than 50,000 files.

More Information

http://now.netapp.com/NOW/knowledge/docs/ontap/rel727/html/ontap/filesag/accessing/task/t_oc_accs_optimizing_NFS_directory_access_for_CIFS_clients.html#t_oc_accs_optimizing_NFS_directory_access_for_CIFS_clients

Brought to you by NetApp Americas Professional Services

Send comments, corrections & suggestion to Michael Cope at mcope@netapp.com
Tech Note archive: http://communities.netapp.com/community/products_and_solutions?view=documents