



File names supported on the archive systems



Introduction

The HCPMP is in the process of implementing a new archival storage service as part of the Storage Lifecycle Management (SLM) project. In order to accommodate this new archival storage service, certain object types, directory and file names will only be supported on archive systems within a regular file container such as a tar file. Prior to the new archival storage service no specific direction was given across the HPCMP for object types or directory and file name support on archive systems.

Summary

Archive systems directory and file names will not support ASCII control characters, some special characters or leading and trailing SPACE characters. In addition only regular files and directories will be supported. Also unsupported are large object-count directories. Up to 250K files or directories per directory will be supported. Characters that are supported in object names on the archive systems of HPCMP include all alphanumeric characters and most printable ASCII characters. The supported characters are the subset of printable ASCII characters, ASCII 32 to ASCII 126, hex 20 to 7e, excluding the special characters SLASH (ASCII 47, h2f, /), BACKSLASH (ASCII 92, h5c, \), and ASTERISK (ASCII 42, h2a, *) and the SPACE (ASCII 32, h20, ' ') as leading or trailing characters.

Unsupported object types, directory and file names will still be supported within a regular file container such as a tar file.

We have provided a script at `${SAMPLES_HOME}/slm/bad-files` that will traverse your directory structure and provide output containing the unsupported objects and object names. The script can be used to review your archive holdings proactively or to confirm that issues you received notification of have been addressed. We request that you examine your archive data for unsupported objects, delete anything other than regular directories and files, and re-name unsupported object names.

Users will be notified if they have files with bad names, unsupported file types, or unsupported large object-count directories.

Please update your scripts accordingly so that newly created files do not contain these unsupported attributes. For more information see the longer version below.

Thank you for your continued support of this initiative and please contact CCAC with questions/concerns.

Details

The HCPMP is in the process of implementing a new archival storage service as part of the Storage Lifecycle Management (SLM) project. The ARL, AFRL, ERDC, Navy, ORS, and MHPCC DSRCs are preparing the archive filesystems in order to accommodate this upcoming solution for both good practice and technical reasons.

We will need your help with the good practice remediation aspect.

An example of good practice is cleaning up non-printing characters, such as ASCII control character codes for BACKSPACE, TAB, RETURN, DELETE and others, that are occasionally introduced by accident into directory and file names.

They present a low-level risk to system security and integrity. They cause problems retrieving files from system backups, and with scripted verification. Almost invariably, they cause confusion for the owner and group owners of the file, who have trouble renaming, deleting, or using the files because the file names simply aren't as they appear to be.

Users will not be able to see files on the archival storage systems with bad names from the new archival storage interface. Once we switch to the new archival storage interface users will no longer be able to archive files with bad names or of unsupported types except within a regular file container such as a tar file.

Container files

Unsupported object types, directory and file names will still be supported on archive systems within a regular file container such as a tar file.

Unsupported file types

As a part of the SLM implementation, the following lists what file attributes and objects will and will not be supported.

Filesystem object types - only regular files and directories will be supported. The output from the 'ls -l' command shows the filesystem object type code as the first character.

Supported:

- d The entry is a directory.
- The entry is an ordinary file.

Not supported:

- D The entry is a door.
- l The entry is a symbolic link.
- b The entry is a block special file.
- c The entry is a character special file.
- p The entry is a FIFO (or "named pipe") special file.
- P The entry is an event port.
- s The entry is an AF_UNIX address family socket.

These unsupported object types will have no meaning in the new archive service and existing objects of these types in your archive files will not be visible through the new service.

Unsupported characters in object names

- All the ASCII control characters, ASCII 00 to ASCII, 31 hex 00 to 1f, including TAB (ASCII 9, h09), RETURN (ASCII 13, h0d), DELETE (ASCII 127, h7f), and BACKSPACE (ASCII 8, h08)
- SPACE (ASCII 32, h20, ' ') as leading or trailing characters
- SLASH (ASCII 47, h2f, /)
- BACKSLASH (ASCII 92, h5c \)
- ASTERISK (ASCII 42, h2a, *)
- Extended characters, ASCII 128 to ASCII 255
- Unicode characters

In SLM names with unsupported characters will be disallowed when you try to create a new object and existing objects having these names in your archive files will not be visible through SLM.

Supported character notes

- The SPACE (ASCII 32, h20, ' ') character is supported as an embedded character but use of

names containing the SPACE character require special handling in the command line, scripts, network protocols and other places. If you use spaces in names you will find cases where this special handling has to be added. A very common case is a simple script that you receive from someone that you have to 'fix' to include the special handling for spaces in file-names. In addition some of the special handling may confuse a user when a name encoded by special handling is viewed eg: a%20name%20encoded.

- The HYPHEN (ASCII 45, h2d, '-') character is supported but it's use as the first character of a name will require special handling in a shell command line and other places.

Unsupported large object-count directories

Also unsupported are large object-count directories. 'Large' means more than 250,000 directories and/or files within a single directory. An 'ls | wc -l' within a directory will produce a count of the number of objects.

Users will be notified of files with bad names or of unsupported file types

We have provided a script at \${SAMPLES_HOME}/slm/bad-files that will traverse your directory structure and provide output containing the unsupported objects and object names. The script can be used to review your archive holdings proactively or to confirm that issues you received notification of have been addressed. We request that you examine your archive data for unsupported objects, delete anything other than regular directories and files, and re-name unsupported object names.

Users will be notified if they have files with bad names, of unsupported file types, or unsupported large object-count directories.

Please update your scripts accordingly so that newly created files do not contain these attributes.

Thank you for your continued support of this initiative and please contact CCAC with questions/concerns.

Unsupported names/objects CCAC email boiler plate

Dear <USER>,

Files you own will not be supported by the new archival storage service as part of the Storage Lifecycle Management (SLM) project. Archive systems directory and file names will no longer support ASCII control characters, some special characters or leading and trailing SPACE characters. In addition only regular files and directories will be supported. Also unsupported are large object-count directories.

We have provided a script at \${SAMPLES_HOME}/slm/bad-files that will traverse your directory structure and provide output containing the unsupported objects and object names. The script can be used to review your archive holdings proactively or to confirm that issues you received notification of have been addressed.

CCAC URL: <http://centers.hpc.mil>

DSN: 785-0679

Toll Free: 1-877-CCAC-039 (1-877-222-2039)

Commercial: 1-937-255-0679

FAX: 1-937-656-9538

Emails: help@ccac.hpc.mil (CCAC Service Center)