JAXA Research and Development Memorandum

HPSS Tape Catalog Tool
-Overview, Specifications, Installation and Use-

KIMOTO Kazuhiro

August 2021

Japan Aerospace Exploration Agency

Contents

HPSS Tape Catalog Tool
-Overview, Specifications, Installation and Use-

KIMOTO Kazuhiro[*1]

ABSTRACT

The Japan Aerospace Exploration Agency (JAXA) is running the data archiving system which is called as "J-SPACE". HPSS (High Performance Storage System) is used in J-SPACE. Magnetic tape is the primary data storage medium, and its management is critical.

In order to gather information such as status and statistics of tapes in a handy manner, we developed a tool called "HPSS Tape Catalog Tool". This document is to introduce it.

1. Introduction

Supercomputer Division of The Japan Aerospace Exploration Agency (hereafter, JAXA) is operating the JSS3 (JAXA Supercomputer System Generation 3) supercomputer system. One of its major components is data archiving system which is called as J-SPACE. In J-SPACE, HPSS[1] (High Performance Storage System) is used. HPSS stores huge volume of data files of the users and manages tape volumes and disk storage devices.

Tape is used as the primary storage medium and its management is critical. HPSS offers commands and GUI based administration to view the status and statistics of tapes, although they are not handy.

In order to address such shortages, Supercomputer Division developed a tool which is named as "HPSS Tape Catalog Tool". This tool gathers information from HPSS and generates Excel reports with a simple operation manner. You could examine various status and statistical information of individual tape, PVR statistics, usage of disks, history of tape addition/removal and so on with the Excel file. Furthermore, exporting data to other software could be achieved since the format is standard Excel.

HPSS Tape Catalog tool is made public to help other HPSS installation sites. In this memorandum, overview and specifications of the tool are introduced. Then, installation procedure and usage guides are also described.

2. Overview of HPSS Tape Catalog Tool

HPSS Tape Catalog Tool gathers information retrieved via HPSS commands and generates an Excel file. Example of the screen of the Excel reports are shown at Figure 1 Example of TapeCatalog sheet .

There are about 10 Excel sheets which are generated. Major sheets and their contents are as follows.

(1) TapeCatalog: Main sheet of the tape catalog, which refers to contents of other sheets and aggregates them.
(2) lshpss -sc TapeLIST: List of Tape SC (Storage Class). Number of tapes and total used size and so on are shown here.
(3) PVR LIST: List of PVR (Physical Volume Repository). Number of tapes and total of Space Used are accumulated.
(4) lshpss -sc DiskLIST: List of Disk SC. Number of Disks are counted here.
(5) volume add_remove: History of Tape volume addition/removal. History is updated and recorded when the data is gathered.

Columns in the TapeCatalog sheets are listed in Table 1 List of TapeCatalog sheet columns .

2.1 Benefits of HPSS Tape Catalog Tool
Benefits of HPSS Tape Catalog Tool are as follows.
(1) The tool integrates various information by individual tape basis from multiple HPSS

---

command outputs.

(2) Information is gathered into an Excel file so that you could filter or sort with standard Excel operation.

(3) Execution of HPSS commands, gathering of data and generation of Excel tables are processed with scripts and Excel macros. Detail knowledge of HPSS is not necessary.

2.2   Configuration and Placement of HPSS Tape Catalog Tool

Configuration and Placement of HPSS Tape Catalog Tool are shown in Figure 2 Configuration and Placement of HPSS Tape Catalog Tool .

HPSS Tape Catalog Tool consists of scripts and an Excel file.

Scripts are placed in HPSS core server.

With the execution of the scripts, commands are issued to HPSS, the results are processed, and output files are generated.

An Excel file is run on Windows. Output files of the scripts would be transferred to Windows. Execution of Excel VBA macro reads the output files.

3.   Specifications of HPSS Tape Catalog Tool

Major specifications of HPSS Tape Catalog Tool are as follows.

3.1   Prerequisite Software
Prerequisites of HPSS core server are:
· HPSS (tested with version 7.5.3.0.0)
· bash
· Standard tools for Unix environment such as 'awk'

Prerequisites of Windows[1] are:

· Microsoft Windows (tested with Windows 10)
· Microsoft Excel (tested with Microsoft 365 ProPlus)

3.2   HPSS commands issued in the script
Here is the list of HPSS commands which are issued in the script. These are to get information. HPSS configuration and settings are not modified with them.
· lshpss -pvr
· dumppv_pvr -l *PVR name*
· dumppv_pvl
· lshpss -sc
· dump_sspvs -i 1
· dump_sspvs -i 1 --lstapes-format

3.3   Considerations of HPSS Tape Catalog Tool
The followings are considerations of HPSS Tape Catalog Tool.
· Language: Macro of the Excel file determines language environment. It displays messages in Japanese in Japanese environment, otherwise English messages would appear. Headings and values are also displayed accordingly. Script messages are in English only.
· Dependency of the software versions: Scripts process the HPSS command outputs. Detail specification of the output format of the commands are not defined. If the output format is to be changed for any reason such as version changes, modification of the scripts may be necessary.
· Subsystem of HPSS: HPSS can handle two or more subsystems in the configuration. But current version of HPSS Tape Catalog Tool assumes that there is only one subsystem and its number is 1.
· Availability of HPSS: HPSS must be active and accept commands, when the scripts are performed.
· Identification of Tape Volumes: HPSS uses 2 types of tape identifications. Different types are used in the output of HPSS commands. They are distinguished conceptually in HPSS design. But this tool concatenate Cartridge IDs with trailing "00" and treats them as identical to Volume IDs.
  · Cartridge ID: Cartridge Label of 6 characters. Example: output of

---

[1] Combinations of the following were also tested.
· Windows10 - Microsoft Office Home and Business 2019, Excel Version 2019
· Windows7 - Microsoft Office Home and Business 2013, Excel version 15.0.5189.1000
If paste method of the macro is in error, it is likely completed normally to restart Excel and run the macro again.

"dumppv_pvr".

· Volume ID: Volume Label of 8 characters. Example: output of "dumppv_pvl"

4. Installation of HPSS Tape Catalog Tool

This section will describe installation and setup of HPSS Tape Catalog Tool.

4.1 Obtaining HPSS Tape Catalog Tool

HPSS Tape Catalog Tool could be obtained at JSS public relations site, in a ZIP archived file.

ZIP file contains script and Excel files.

· https://www.jss.jaxa.jp/en/hpsscatalogtool/ (English)

· https://www.jss.jaxa.jp/hpsscatalogtool/ (Japanese)

4.2 Placement of HPSS Tape Catalog Tool

Figure 3 Directories and Files of HPSS Tape Catalog Tool shows the recommended directory layout.

Following is the installation and placement of the tool on HPSS Core Server.

(1) Make "Tool Top Directory" for HPSS Tape Catalog Tool. Use this directory as the current directory when HPSS Tape Catalog Tool is invoked.

(2) Make "Scripts Directory" under "Tool Top Directory" (you could make it elsewhere if there is a reason to do so from the management point of view).

(3) Extract the script files and an Excel file from the ZIP file obtained above into "Scripts Directory". The Excel file is a template Excel file. It would be copied to the Output directories by the script.

(4) HPSS commands must be executed by root. Permissions of the directories and files are properly set to make it possible, i.e., they must be read, written and executed by root.

(5) "Output Directories" are made by the scripts automatically. They are named as "hpssTapeInfo_YYYYMMDDHHMM" (YYYYMMDDHHMM indicates the date and time of the execution.) Output directories are located side by side under "Tool Top Directory" so that the scripts could search the previous

results for tape addition/removal history.

On Windows PC, directories are independent each other. There is no restriction for placement. But as shown right half of Figure 3, it is convenient to make "Tool Top Directory" and place the output directories underneath.

5. Operation of HPSS Tape Catalog Tool

The following is the operation procedure of HPSS Tape Catalog Tool.

5.1 HPSS Core Server

Execute the following steps.

(1) Log in to HPSS Core Server and switch to root (su command).

(2) Add "Scripts Directory" to PATH environmental variable.

(3) Go to "Tool Top Directory" and make it as a current directory.

(4) Run "get_hpss_tapeinfo.sh" which is to invoke HPSS Tape Catalog Tool. Make sure the messages appear on the screen. Output Directory name would be shown in the messages. Example is as follows.

```
get_hpss_tapeinfo.sh (version 2.0
2020/10/01) Started
Output will be stored in
/home/k/k919/hpss_info_Eng/hpssTapeInf
o_202010011644
```

(5) Wait for the completion. Elapsed Time would depend on the environment, it would be less than 1 minute for our installation.

(6) After completion of the command, make sure the output of log file, as necessary. It is "Logfile_get_hpss_tapeinfo.txt" in the Output Directory.

Zip files in the Output Directory and transfer them to Windows PC. It is recommended to include directory name so that it helps to identify executed date and time of the tool.

5.2 Windows PC

Execute the following steps.

(1) Receive the ZIP file of the "Output Directory" from HPSS Core Server.
(2) Extract the ZIP file at "Tool Top Directory".
(3) Open the Tape Catalog Excel file in the Output Directory. Its name is "HPSS_TapeCatalog_YYYYMMDDHMM.xlsm".
(4) If the message "SECURITY WARNING Macros have been disabled" appears, click "Enable Content".
(5) On the Menu of Excel, select "Developer" and click "Macros", select "automatic_process" in "Macro Name" and click "Run". (If "Developer" is not in the menu, enable it with "Customize Ribbon" in "Options".)
(6) A Dialog box would appear to tell that if required files are in the same directory and the processing will start. Click "Yes".
(7) Wait for the completion of the macro. Elapsed time depends on the performance of the PC and the amounts of data, it would be 1 to 5 minutes.
(8) A Dialog box would appear to tell the processing is completed.

The macro determines the language setting of Excel (Microsoft Office) and switches messages, heading of the tables, and string values such as "unlabeled" in the tables. If the language setting is Japanese, texts are in Japanese. Otherwise, English texts are selected. ("Yes" and "No" in the dialog box depends on the language setting of the operating system.)

The macro also checks if the macro was executed already or not. If it finds that it was executed, it would stop and displays the message to tell so.

5.3 Format of Tape Catalog Excel file and the contents which could be changed

Format of Tape Catalog Excel file are introduced here. Notes and the contents which could be changed are also discussed.

Upmost rows of the sheets in the Excel tables contain the following contents. They are not to be changed or the macro would not work properly.

(1) Row 1: description of the table contents and date/time of the data gathering

(2) Row 2: copy of the worksheet formulae to be set in the data rows. This row could be used to recover the data rows if the functions of data rows are accidentally broken.
(3) Row 3-4: blank
(4) Row 5: heading of the table, filters are applied.
(5) Row 6- : Data and formulae rows

There are some notes to modify the Excel file.
(1) Sheet names should not be changed before execution of the macro, since the macro refers to their names. Names can be changed after the completion of the macro.
(2) Insertion and other changes of rows/columns in "outfilelist" is not allowed. The macro refers and modifies this sheet. String data which could be changed are indicated in the sheet.
(3) "Cover" sheets contain brief usage description in Japanese and English. Contents could be changed according to the needs of the installation site.
(4) Sequence of the sheets does not affect to the behavior of the macro. Sequences could be changed. For example, you could move the frequently referred sheet left.
(5) Adding sheets does not affect to the functions.
(6) If you want to use the modified version of the Excel file in the future executions, modify the Excel file before the macro execution and place the file in the Scripts Directory as the template Excel file. HPSS Tape Catalog Tool script would copy it to the Output Directory.

6. Conclusion

This memorandum introduced overview and specifications of HPSS Tape Catalog Tool. Installation procedure and usage guides are also described.

We would be glad if the tool is useful and helps the management of HPSS.

We welcome your comments and information about multiple HPSS subsystems and so on, function changes and enhancements.

I would like to thank Fujimoto Yasushi of IBM Japan, for useful information and questions-answers support

on HPSS while in the development of the tool.

Reference

1)　High Performance Storage System, http://hpss-collaboration.org/ (last visited November 27, 2020)

Table 1 List of TapeCatalog sheet columns

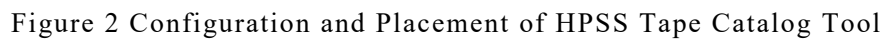| Col.	Heading	Source	Example of values (other values may be shown in the actual report)	"unallocated" or "unlabeled" is shown according to the value of SC ID column (Note)
A	Volume	dumppv_pvl	"A0021600"	
B	SC ID	dumppv_pvl	SCnumber, " unallocated", "unlabeled"	
C	PVL State	dumppv_pvl	"idle", "busy"	
D	Type	dumppv_pvl	"3592 E08 XL Tape"	
E	SC Name	dumppv_pvl	"SC_T_ALOS1"。	Yes
F	PVR Name	dumppv_pvr	"SCSI PVR (TS3500 3592)-NEW" "PVR unassigned"	
G	Total Mounts	dumppv_pvr	1,046	
H	Mount Status	dumppv_pvr	"DISMOUNTED", "MOUNTED", "CHECKED OUT"	
I	VV Cond	dump_sspvs -i	"DOWN", "EMPTY", "EOM", "RO", "RWC"	Yes
J	Space Used	dump_sspvs -i	9,699,292,801,021	Yes
K	Max Volume Size	dump_sspvs -i	12,063,461,165,268	Yes
L	Other Information	dump_sspvs -i	blank, "Read Error"	Yes
M	Last Write	dump_sspvs --lstapes-format	2019/10/10 14:47, blank	Yes
N	Last Read	dump_sspvs --lstapes-format	2019/9/28 8:59, blank	Yes
O	State	dump_sspvs --lstapes-format	"ALLOC", "EMPTY", "EOM"	Yes
P	Files	dump_sspvs --lstapes-format	9,276	Yes

Note: "Yes" in this field indicates the following conditions.

When SC ID (Column B) is either "unallocated" or "unlabeled", there is no data to refer to, and Excel shows "#N/A" unless the error condition is handled properly. In order to avoid showing "#N/A", the values of "unallocated" or "unlabeled" are displayed.

For further information, please consult HPSS documents and man pages of HPSS commands.
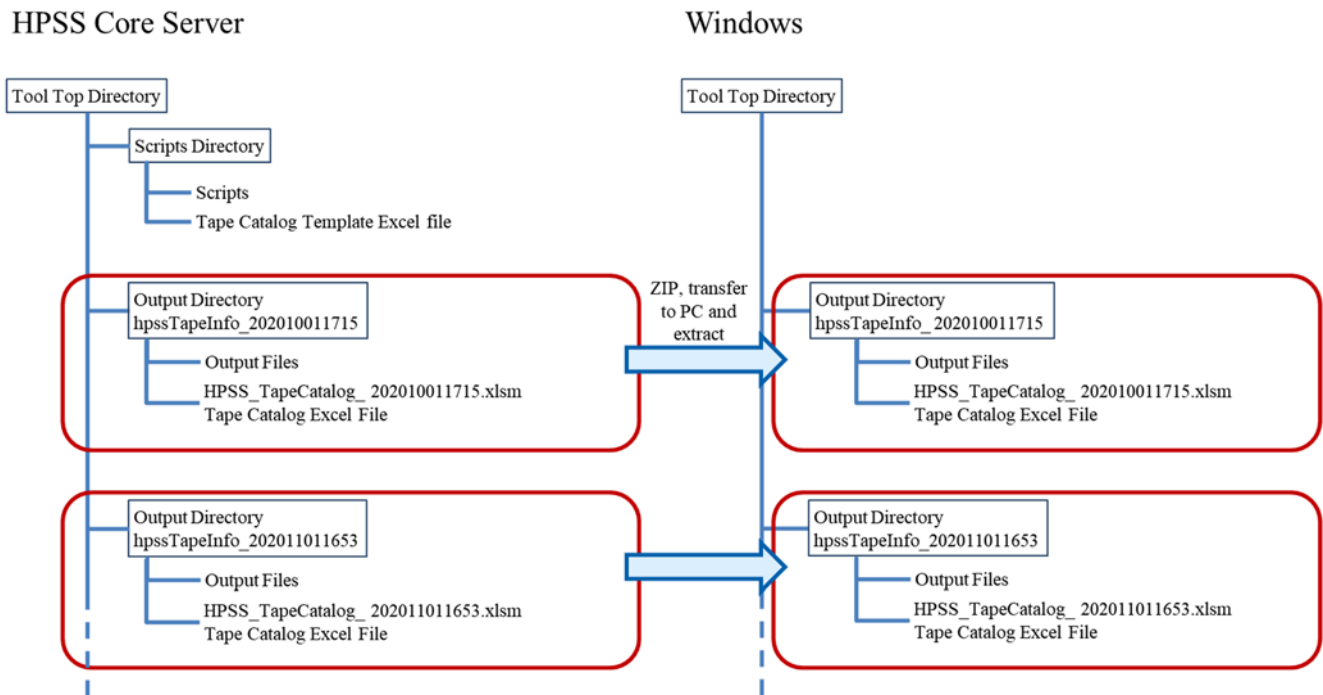
Figure 1 Example of TapeCatalog sheet

Figure 2 Configuration and Placement of HPSS Tape Catalog Tool

Figure 3 Directories and Files of HPSS Tape Catalog Tool