

Tape Consolidation: A User's Experience

This user demonstrates one way to reduce the number of tape volumes at your site and more efficiently use each tape volume.

A review of our site's tape usage led us to reduce the number of small or poorly utilized tape volumes. Tape Mount Management (TMM) is the long-term solution, but its full implementation is also a long way off. As an interim measure we decided to do post-creation tape stacking.

There are several vendors with post-creation tape stacking tools such as TapeSaver (Mobius Management Systems, Inc., New Rochelle, N.Y.), CARTS-TS (Technologic Software, Irvine, Calif.) and Tape Reclaim (Open Tech Systems, Plano, TX), etc. Each of these products has benefits and drawbacks. We viewed our tape consolidation needs as being immediate but, with TMM coming, short lived. We elected to use Computer Associate's (Islandia, N.Y.) CA-1 CopyCat because it is free to CA-1 users, does what we need, and we can accept its limitations.

CA-1 CopyCat is basically a tape copying utility. It has two functions, a volume copy utility much like IBM's Ditto, and a file copy utility. This article examines the file copy functions.

OPERATING ENVIRONMENT

Our organization functions as an in-house service bureau consolidating more than 30 operations and data centers into three sites. Two of the three consolidated sites manage 75 images with about 30 Tape Management Catalogs (TMCs) between them. The terms "image" and "TMC" will be used interchangeably in this text. Our environment consists of:

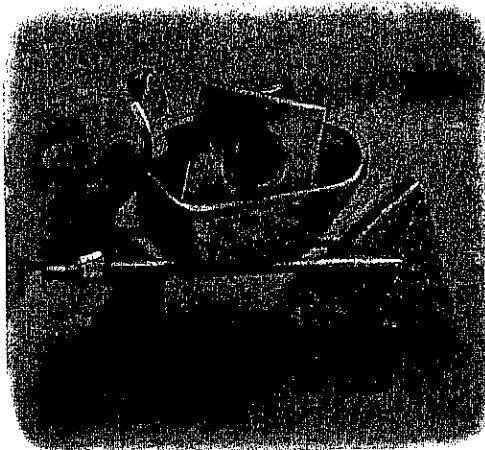
- MVS/ESA 4.22;
- SAS 6.07;
- CA-ACF2 5.2;
- SAR 6.5;
- CA-1 5.0;
- CA-1 Copycat 1.0;
- DF/HSM 2.6;
- StorageTek ACS 4400s (silos); and
- StorageTek EXLM 1.2.

A range of tape volumes for each image is assigned solely for use within the tape silo. For example, one image has the volser range of 00000Y-29999Y with 16000Y-19999Y residing within the silo complex. Data subject to being sent off-site is created only on external transports. Volumes within the (silo) volser range are not ejected from the silo.

IDENTIFYING CANDIDATES

CA-1 CopyCat will process standard label, CA-1-managed tapes but you must provide either a volume list or a data set name list (ICF catalog look-ups are done unless the DSN is accompanied by volser). The consolidation is a two-step process that identifies the volumes (candidates) to be stacked, followed by the actual stacking step.

Using a SAS program (available in the TSARTICLES LIB of NaSCOM as filename FEB95001.EX1) to read the TMC, we identified tapes within the eligible volser range (example: 16000Y-19999Y). We did not consider volumes within a multi-volume chain or multi-file volumes. We also excluded several patterns (DSN, creating job name, DDNAME); these patterns exclude DF/HSM tapes, LEGENT's SAR volumes, IMS log tapes, TMC back-ups, etc. We wanted to exclude any product that has



internal pointers to tape volsers.

We also looked at last reference and expiration dates and excluded any volume last referenced within the past 14 days or due to expire within the next 14 days. Tapes with special CA-1 expiration (permanent, catalog control, cycle control, inactive days) are eligible for consolidation.

We calculated data set size by multiplying block size by block count and dividing by one million. At present, we are entirely on 18-track tape, so we figured a capacity of 300MB per cartridge with IBM's data compression Improved Data Recording Capability (IDRC). We considered anything up to 125MB eligible for consolidation.

We then sorted the candidates by data set size (in ascending order). If necessary, we could limit the number of tapes being processed and still get the most benefit possible from any run.

CONSOLIDATION

Once the candidate list was generated, it was passed to the CA-1 CopyCat step. This step uses the control cards shown in Figure 1.

With the exception of unit addresses (creating and last used) TMC data for input volumes are duplicated in the TMC record for the output. For example, a data set created on Julian date 94050 and copied by CopyCat on 94060 will still have a TMC creation date of 94050. We deemed this important for the user community, especially for uncataloged data.

Problems and Gotchas

1 Data consolidated onto a volume may expire at different times. We set up a second CopyCat job that identifies inactive (unreferenced for six months) consolidation tapes and basically recycles them. When doing this we recommend keeping the CTLFILE the same for all CopyCat runs to prevent copying to the new volumes too. See items 7 and 8 for more on the recycle process.

2 Don't consolidate data for products with internal pointers to tape volser. Find out what products you have that can do this.

3 CopyCat's output will be chained volumes. If we are writing the tenth file on output tape one when we encounter the end of volume, we continue writing the file on a second output tape. With large amounts of data going into the consolidation run, we wind up with long volume chains, some as long as 50 volumes (more than 375 tapes have been read in). It would be nice if Computer Associates would provide the ability to cut off the data set being copied at end of volume and start with a new tape.

4 It is possible to have a single job try to reference two data sets on the same consolidation volume in the same job step. This results in an S513 abend. If we encounter this, we will copy one of the data sets to another tape and re-run the abending job. We and our clients accept the risk because we have little tape data not referenced within the 14-day period prior to the consolidation being referenced again. Some products provide additional protection.

5 Analyze your tape data. Look out for data sets created on tape volumes using referbacks. The referback may be to a data set that is now on a consolidated volume and the file sequence numbers can become a problem (S813/SA13 abends).

6 Beware of data sets likely to be opened for output or modification; they are unlikely to be the last file on the consolidated tape.

7 We needed to apply APAR G026090 for our recycle job. This APAR causes CopyCat to table the updates to expiration dates for the input files and apply the updates at successful close of volume (or volume chain). Before applying this APAR we ran a scratch job while CopyCat was processing one of our multi-volume consolidation tapes and it scratched the input volumes before all of the files were copied.

8 We encountered an S878-10 abend in our recycle job. We were passing to CopyCat all of the volumes in a multi-volume chain. CopyCat reads a passed volume, identifies all chained tapes and DSNB records for that volume, and then moves to the next volume in the selection list. If we processed a three-volume chain of TAPE01, TAPE02 and TAPE03 CopyCat would build a table of all files on the volume chain three times. When we passed only the first volume in a multi-volume chain the S878 abends stopped.

We used the same control cards in more than one CopyCat run, including a DF/HSM-like recycle which drops inactive (expired) data sets. If a data set was not cataloged prior to our consolidation run, it remained uncataloged; previously-cataloged data sets are re-cataloged. We keep the original expiration dates.

The CTLFILE is a dummy (one block) file written as the first file on all CopyCat output volumes. ACF2 rules require that a user wishing to read the nth file on a tape must have read access to the first file on the tape. In Figure 1, the CTLFILE name is COPYCAT.CTLFILE and all users are granted read access with a very limited number of people able to allocate/update/delete it.

We sorted the candidate list by data set size as part of our selection process because we sometimes limit the number of input volumes to be processed and wanted the best (smallest) candidates. Although CopyCat is passed the volume list in data set size order, it will internally sort the selected volumes in volser sequence to make pre-pulling tapes easier and issues mounts in volser order. CopyCat also groups by density so if there are IDRC and non-IDRC compressed tapes in the candidate list you should expect them to be used in density and volser sequence. If we have 10 volumes that are IDRC compressed and two volumes that are not IDRC compressed, the group of 10 compressed volumes will be copied onto one set of output volumes and the two non-compressed volumes will be copied onto another set of output volumes. Each set is mounted in volser order.

RESULTS

We saw very dramatic results on the trial image; it had about 30,000 tapes. In the four months since we began this project, we've consolidated approximately 6,000 volumes (about 20 percent) onto about 1,500 volumes. As a result, we were able to avoid purchasing 4,000 tapes at a savings of about \$16,000. We found nearly 1,000 tapes with less than 25MB of data.

We know that data has been read from the consolidated volumes and have had no reports of users unable to read data. Because we select only inactive volumes, we have not had any occurrences of multiple jobs simultaneously requesting the same tape volume.

Figure 1: CA-1 CopyCat Control Cards

FILECOPY	Copy files
INPUT=TMCDATA	DDNAME for selected volumes
VOL=2	Column position where volsers begin
FILES=ACTIVE	Copy only active (non-expired) files
SAVEINFO=YES	Keep TMC data the same
RECATALOG=PREV	Re-catalog any cataloged data sets
OUTDISP=SAME	Keep expiration dates the same
INDISP=RETPD=D	Expire the input volumes
CTLFILE=COPYCAT.CTLFILE	Name of control file

We now have additional jobs running that consolidate tapes at the high end of each TMC range. Our goal is to free several thousand tapes at the high end of each range. The tapes freed by the CopyCat consolidation process are being re-assigned to images for new clients and allow us to avoid additional tape and tape rack purchases.

For what we perceive as an immediate but short-lived requirement to reduce the number of tape volumes at our sites and to more efficiently use each tape volume, we used CA-1 CopyCat to consolidate poorly-utilized volumes. Other products were reviewed and provide additional features, but with the short life of this project, the free tool from Computer Associates was our best option. **TS**

Was this article of value to you? If so, please circle Reader Response Card No. 21.

Would you like more information on the product mentioned in this article, if so, please circle Reader Response Card No. 51.



Lloyd Christensen has worked in operations support and storage management in the public utility and financial services industries for nine years. Lloyd is also a former communications chairman for TechSys, NaSPA's Dallas/Fort Worth chapter.