Syncronizing cluster copies in a TS7700 Grid (PRESTAGE and COPYRFSH)

(Last update: 2018, February 13)

To use any of the tools mentioned in this document, you will need to download and install the IBM Tape Tools files: IBMJCL.XMI, IBMLOAD.XMI, IBMCNTL.XMI, IBMPAT.XMI.
Use IBMTOOLS.TXT for installation instructions.

Note. The program COPYRFSH must be located in the library APF (Authorized Program Facility) and the DD STEPLIB in the job COPYRFSH should be updated accordingly.

A database is maintained on each individual TS7700 cluster that contains status information about the logical volumes defined to the grid.

The following discussion relates specifically to syncing volumes when a second cluster is added to make a 2-cluster grid. The process can easily be adapted to work when a third or forth cluster is added.

If this is a grid where the customer recently added a cluster, you can have them run the BVIRMES job where the MC parameter is set to a name using NR so the lvol will only reside in the CL1 cache, assuming that the new cluster is CL1. The BVIRMES job requests a BVIR VOLUME STATUS from CL1. The VESYNC job would then read the MESFILE and find all lvols with the mes_flag set meaning they need a copy.

MES Volume
This field indicates that the logical volume was created in the TS7700 Cluster prior to it being merged into a Grid configuration. Volumes that existed in a TS7700 Cluster prior to being included in a Grid configuration are not automatically copied to the other TS7700 Clusters in the configuration until they have been accessed and closed. This field could be used to determine which volumes in each TS7700 Cluster that have not been copied and used to build a set of jobs to access them and force the copy.

MES_FLAG values:
This field indicates whether the volume was part of an MES merge operation. 'Y' indicates the volume existed prior to merging the cluster into a Grid configuration. After code level 8.3 this value is no longer used and is replaced with 'W' and 'M'. 'W' indicates the volume existed prior to merging the cluster into a Grid configuration and has not been accessed since the MES merge operation. 'M' indicates the volume existed prior to merging the cluster into a Grid configuration and has been mounted/demounted without being modified. The volume will be copied to the clusters specified in the copy_mode fields. 'N' indicates that the volume was not part of the MES merge operation or if it was, has since been modified and successfully copied to the clusters specified in the copy_mode fields.

After you have the list of lvols to recall, you need a BVIR VOLUME MAP from the cluster where the recalls will be done (one of the original clusters). You optionally need a snapshot of the tape catalog so filters could be used to perhaps pick high priority applications to have copies made first and to ignore recalling scratch volumes.

You can confirm that the VOLUME MAP actually did come from CL0 by browsing the file and verifying that the cluster id in the headers is for CL0. Then, you would run PRESTAGE from CL0 and it would do all the recalls efficiently since the VOLUME MAP is from CL0. If the TS7700 Micro Code is R2.1PGA1 (8.21.0.118) or higher on all clusters, then the COPYRFSH tool can be used instead of PRESTAGE. PRESTAGE does a host mount/dismount to cause the peer copies to be made. COPYRFSH issues a command to the TS7700 so the copies can be done without host mounts. The lvols still need to be loaded into the sourse cluster's TVC for the copy to be made.

A way of running PRESTAGE would be to use the MAXDR= 1 parameter so that only one recall at a time would be occurring, but you could recall the lvols from multiple physicals and more or less be continually doing multiple recalls at a time. To be more aggressive, MAXDR=3 would allow 3 at a time to be recalled, but that would be three times the link traffic. Yes, there are ways of limiting how many recalls are submitted at one time to

minimize TVC flushing and the recalled volumes can be set to PG0 also by using the STORECLASS= parameter. If it isn't used, then lvols will use their original PGn attribute when recalled.

Both PRESTAGE and COPYRFSH use the following parameters to limit what gets recalled into a source cluster at one time.

```
*MAXGB= 1000;  LIMIT THE GIGABYTES RECALLED TO CACHE (DFLT 4000)
*ONEPHYSICAL;    LIMIT RECALLS TO JUST THOSE ON ONE PHYSICAL VOLUME.
*              THE ONE WITH THE MOST RECALLS OR THE FIRST ONE IN
*              ALPHANUMERIC SEQUENCE DEPENDS ON SORTBYNUMBER USE.
```

After all of these recalls have completed, you would need to start the process over again from the beginning because the previous list of lvols would now have a bunch of them recalled.

The customer would need to restrict when reclaim is being done on a cluster or it might affect the efficient recall of the lvols. PRESTAGE orders the lvol recalls based on where they reside on the pvols. Reclaim might change the pvol where lvols reside and cause too many pvol mounts.

1. Turn off reclaim on CL0
2. Run BVIRVTS on CL0 to get VOLUME MAP
3. Run BVIRMES on CL1 to get VOLUME STATUS
4. Run VESYNC to get current list of lvols needing copies
5. Run PRESTAGE to build IEBGENER jobs to orderly recall lvols or run COPYRFSH if at the right micro code level.
   Go back to step 3 for next set of recalls.

If turning off reclaim for an extended period presents a problem, then repeat steps 2-5 each time.

PRESTAGE will allow the user to assign a temporary MGMTCLAS value (like RDD) to influence recalls to be in the desired cluster. It also provides for changing it back to the original MC name. Both of the changes are done in the job actually doing the recalls. COPYRFSH does this by assigning a source cluster.

The actions performed by a management class can be changed so a copy will be made by the next dismount, but the NAME of the management class can only be changed via the LIBRARY LMPOLICY command which PRESTAGE will do if requested.

Each generated RECALLnn job will start by mounting a BVIR VOLUME MAP volume which will cause a device in the desired cluster to be allocated. It then concatenates all of its recalled volumes to the same device address so all recalls should occur in the desired cluster.