

Recalls preferred for cache removal (RECLPG0)

Normally, a volume that is recalled into cache is managed as though it were newly created or modified because it is in the TS7700 that is selected for I/O operations on the volume. A recalled volume displaces other volumes in cache.

If the remote TS7700 is used for recovery, the recovery time is minimized by having most of the needed volumes in cache. However, it is not likely that all of the volumes to restore are resident in the cache. Therefore, some number of recalls is required. Unless you can explicitly control the sequence of volumes to be restored, it is likely that recalled volumes displace cached volumes that are not yet restored from, which results in more recalls later in the recovery process.

After a restore completes from a recalled volume, that volume is no longer needed. These volumes must be removed from the cache after they are accessed so that they minimally displace other volumes in the cache.

Based on business requirements, this behavior can be modified by using the RECLPG0 setting, as shown in the following example:

```
LI REQ, <distributed-library>, SETTING, CACHE, RECLPG0, <ENABLE/DISABLE>
```

This setting features the following characteristics:

- ▶ When disabled, which is the default, virtual volumes that are recalled into cache are managed by using the actions that are defined for the Storage Class construct that is associated with the volume as defined in the local TS7700.
- ▶ When enabled, recalls are managed as PG0 volumes (prefer out of cache first by largest size), regardless of the local definition of the associated Storage Class construct.

4.2.8 High availability considerations

High availability (HA) means providing continuous access to virtual volumes through planned and unplanned outages with as little user effect or intervention as possible. It does *not* mean that all potential for user effect or action is eliminated. The following guidelines relate to establishing a grid configuration for HA:

- ▶ The production systems, which are the sysplexes and logical partitions (LPARs), have FICON channel connectivity to both clusters in the grid. The IBM Data Facility Storage Management Subsystem (DFSMS) library definitions and input/output definition file (IODF) are established, and the appropriate FICON Directors, DWDM attachments, and fiber are in place.

Virtual tape devices in both clusters in the grid configuration are varied online to the production systems. If virtual tape device addresses are not normally varied on to both clusters, the virtual tape devices to the standby cluster must be varied on in a planned or unplanned outage to enable production to continue.

- ▶ For the workload placed on the grid configuration, performance throughput must be sufficient to meet service level agreements (SLAs) when only one of the clusters is used. Assume that both clusters are normally used by the production systems (the virtual devices in both clusters are varied online to production). In the case where one of the clusters is unavailable, the available performance capacity of the grid configuration can be reduced by up to one half.
- ▶ For all data that is critical for high availability, consider the use of an MC whose Copy Consistency Point definition includes both clusters with a Copy Consistency Point of RUN (immediate copy) or SYNC (sync mode copy).