

February 4, 2015



**IBM® Virtualization Engine TS7700 Series
Best Practices**

Synchronous Mode Copy Version 1.1

Author : Takeshi Nohta

nohta@jp.ibm.com

RMSS/SSD-VTS – Japan

Co-Author : Katja Deneleh

Katja.Deneleh@de.ibm.com

Consulting IT Specialist, IBM systems



- 1 Introduction..... 3
 - 1.1 Change History 4
- 2 Synchronous Mode Copy..... 5
 - 2.1 Synchronous Mode Copy Requirements and Recommendations 7
 - 2.2 Who Will Use Sync Mode Copy?..... 7
 - 2.3 Sync Mode Copy Options..... 7
 - 2.3.1 Synchronous Deferred on Write Failure Option 7
 - 2.3.2 On Private Mount Option..... 8
 - 2.3.3 Mount Results When One or Both "S" Clusters are Unavailable 9
 - 2.4 Primary "S" Selection 10
 - 2.5 What is the Synchronous-Deferred State? 11
 - 2.6 Sync Mode Copy's Affect on Other Functions..... 11
 - 2.6.1 Scratch Allocation Assist (SAA) 12
 - 2.6.2 Copy Policy Overrides 12
 - 2.7 Synchronous Mode Copy and the TS7700 Management Interface 13
 - 2.7.1 Defining Copy Policy on the Management Class Panel 13
 - 2.7.2 Sync Mode Copy Indications on the TS7700 MI 14
 - 2.8 Host APAR OA37267..... 17
 - 2.8.1 Synchronous Mode Copy..... 17
 - 2.8.2 Other Tape Library Support Enhancements 18
 - 2.9 Host APAR OA40572..... 19
 - 2.9.1 New and Changed Host Messages 19
- 3 Host Applications and Sync Events..... 21
 - 3.1 Explicit and Implicit Tape Synchronous Operations 21
- 4 Synchronous Mode Copy Performance 22
- 5 Disclaimers: 23

1 Introduction

The IBM TS7700 Virtualization Engine (TS7700) is the latest in the line of tape virtualization products that has revolutionized the way mainframe customers utilize their tape resources. Tape Virtualization subsystems have become an essential part of most mainframe customer's operations. The ability to connect two or more clusters into a grid provides support for high availability and disaster recovery configurations. The customer can choose which volumes to copy to other clusters, what clusters to copy them to, the number of copies to be made, and how the copies are completed.

Prior to Release 2.1 the customer could choose the copy to be immediate or deferred. An immediate copy results in a logical volume copy being made at Rewind/Unload (RUN) time. The completion of the RUN is held until the copy is completed. A deferred copy results in the copy being placed into a queue for subsequent copy to other clusters. The RUN response is not held for a deferred copy.

Release 2.1 adds a second flavor of immediate copy called Synchronous Mode Copy (SMC). With SMC, two copies of a logical volume are written at the same time to two clusters in a grid. When an explicit or implicit tape sync operation occurs (see section 3.1 - Explicit and Implicit Tape Synchronous Operations) the TS7700 grid ensures that the data written into the device buffer so far has also been written to each TS7700's disk cache before the response to the tape operation is returned.

Synchronous mode copy is a form of replication similar to host-initiated duplexing that can provide a zero recovery point objective (RPO) for sets of records (datasets) or byte-stream data (objects) written to virtual tape.

System z applications like DFSMSHsm, DFSMSdfp OAM Object Support, and other workloads can achieve a zero RPO when using this method of replication. When zero RPO is achieved, any exposure to delayed replication is eliminated and host-initiated duplexing (e.g., DFSMSHsm's duplex tape support) is not required. SMC duplexes all compressed host writes simultaneously to two TS7700 clusters using the grid links.

Before any host-initiated explicit or implicit synchronize tape operation can succeed, all host data sent to a virtual device since the last synchronizing event has been written on virtual tape up to that point is written to persistent disk cache at both TS7700 clusters. If a cluster fails, the host application or job can access the secondary copy with consistent access up to the last completed synchronization point.

Additional SMC options enable a user to enforce strict synchronization for a given workload, or permit a write workload to enter a synchronous-deferred state when full synchronization is not possible. These are discussed later in the document.

1.1 Change History

Version 1.0 – April 2, 2012

- Original Release

Version 1.1 – November 4, 2014

- Add the new Synchronous Mode Copy Private Mount (Dual open for MOD) option supported in Release 3.2.
- Add the Time Delayed Copy Policy supported in Release 3.1.

2 Synchronous Mode Copy

The figure below illustrates duplex copying without Sync Mode Copy (SMC). The **blue example** shows the Host performing two writes, one to a TS7700 and another to another TS7700 or to physical tape. The **purple example** shows two separate writes to a TS7700 grid which uses copy policies to write the volumes to separate clusters in the grid. The **red example** shows a single write to one cluster in a grid with the second copy as a RUN or deferred copy. In the first two examples, the host performs double duty because it has to write the same data twice. In the third example the host has to write the data just once, but the second copy doesn't occur until after the RUN is issued.

- **Without Sync Mode Copy the Host must write the same data twice.**
 - To separate device/library
 - To other cluster via remote mount
 - To other cluster via immediate copy



Figure 1 - Without Synchronous Mode Copy

With Sync Mode Copy the Host only writes data once. The TS7700 performs the duplexing. Just one logical volume volser to manage.

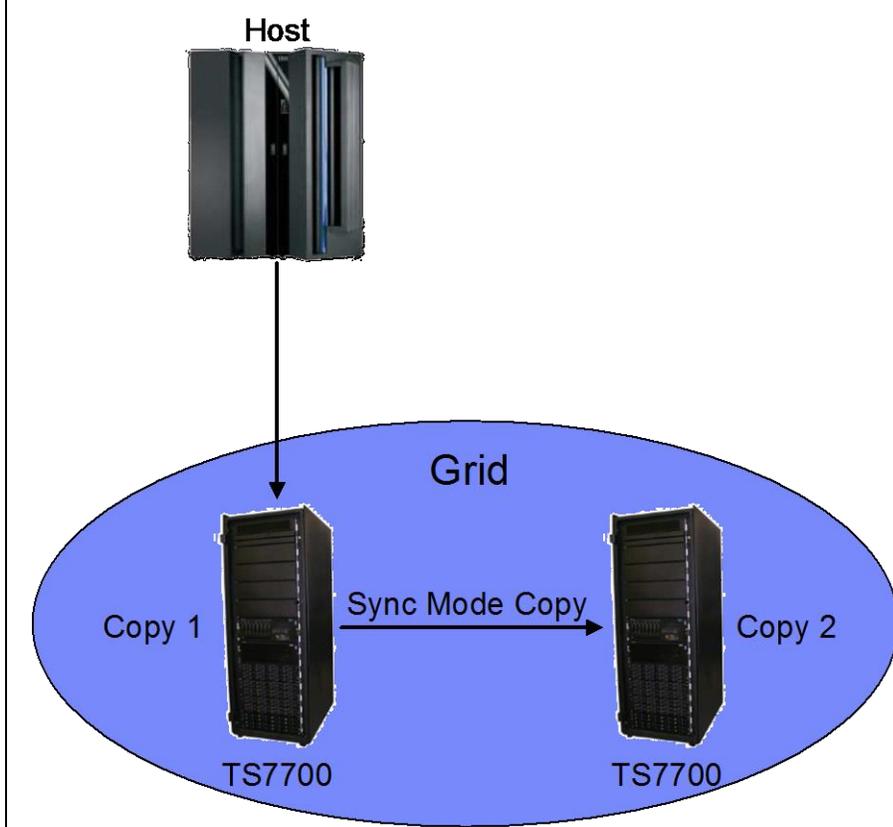


Figure 2 - With Synchronous Mode Copy

The figure below illustrates that with SMC the TS7700 writes to two TS7700 at the same time. When an explicit or implicit sync event occurs, the event is not responded to until the data at both clusters is written to the disk cache.

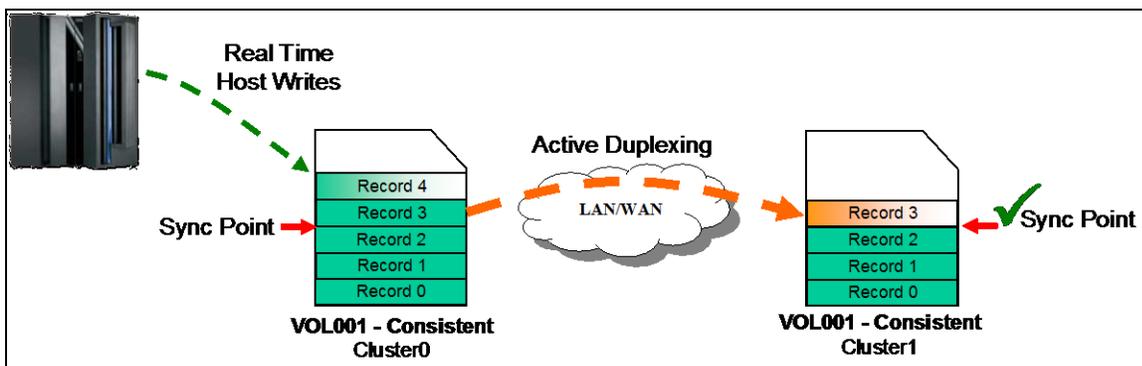


Figure 3 - Synchronous Mode Copy Sync Point

2.1 Synchronous Mode Copy Requirements and Recommendations

- All clusters in a grid must be at a microcode level of 8.21.0.xx or higher to use synchronous mode copy.
- APAR OA37267 is recommended to support new host messaging. The APAR is not a requirement for SMC to work.
- To enable synchronous mode copy, create a management class that specifies exactly two grid clusters with the synchronized ("S") copy mode. Synchronous mode copy is configured using construct options on the Management Class panel of the TS7700 Virtualization Engine Management Interface. **Other copies to additional clusters can be made, but only RUN, Deferred or Time Delay replication can be selected for these clusters.**

2.2 Who Will Use Sync Mode Copy?

Applications such as DFSMSHsm and DFSMSdfp OAM Object Support which often store datasets or objects on logical tape could benefit from the use of Synchronous Mode Copy (SMC). After writing some data to the TS7700 the host application issues a synchronize operation that forces all previously written records to tape volume cache (TVC). The host application (e.g. DFSMSHsm during ML2 processing) can then classify the dataset as written to tape and may discard the primary host source copy before the volume is closed.

- Applications that use dataset-style stacking and migration are the expected use case for synchronous mode copy.
- However, any host application that requires near-zero RPOs can benefit from the synchronous mode copy feature.
- There is no host software dependency other than the ability to issue an implicit or explicit synchronize command after critical data has been written.
- If no implicit or explicit sync operations occur, volume close processing will result in an implicit sync operation resulting in a zero RPO once the RUN command is issued.
- The synchronous mode copy occurs external to the System z server, within the TS7700 Virtualization Engine, and relies entirely on the TS7700 Virtualization Engine Grid network.
- No FICON channel activity occurs to the secondary duplexed location.
- Performance is similar to a remote mount for write.

2.3 Sync Mode Copy Options

The following sections describe the two options available with Synchronous Mode Copy (SMC). The two options are "Synchronous Deferred on Write Failure" and "Open Both Copies on Private Mount".

2.3.1 Synchronous Deferred on Write Failure Option

The default behavior of SMC is to fail a write operation if both clusters with the "S" copy policy are not available or become unavailable during write operations. You can enable this option to permit update operations to continue to any valid consistency point in the grid. If there is a write failure, the failed "S" locations are set to a state of "synchronous-deferred". After the volume is

February 4, 2015

closed, any synchronous-deferred locations are updated to an equivalent consistency point through asynchronous replication. If the Synchronous Deferred on Write Failure option is not checked (default) and a write failure occurs at either of the "S" locations, then host operations fail and only content up to the last successful sync point should be viewed as valid.

Note: An "R", "D", or "T" **site as defined in the management class** is chosen as the primary consistency point only when both "S" locations are unavailable.

For example, imagine a three cluster grid and a copy policy of SSD, Sync Copy to clusters 0 and 1 and a deferred copy to Cluster 2. The host is connected to clusters 0 and 1. With this option disabled both clusters 0 and 1 must be available for write operations. If either one becomes unavailable, write operations will fail. With the option enabled, if either cluster 0 or 1 became unavailable, write operations would continue. The second "S" copy would become a synchronous deferred copy.

In the above example, if the host was attached to cluster 2 only and the option was enabled, the write operations would continue even if both clusters 0 and 1 became unavailable. The "S" copies would become synchronous deferred copies.

The default SMC behavior will fail tape operations if two copies of a volume cannot be maintained during an update (synchronous failure). When the synchronous failure setting is used, a zero RPO is provided for the target workload, independent of failures. Consider the following circumstances when using the default strict synchronization behavior:

- If the failure to synchronize is detected after the mount has already occurred, then tape operations fail to the targeted volume until a RUN occurs and a demount command is issued.
- If content was written prior to the synchronization failure, then previous content on the emulated volume up to the last successful tape synchronization operation point is considered persistently synchronized and can be accessed later from either "S" consistency point.
- If either "S" consistency point is unavailable, then scratch mount operations fail.

2.3.2 On Private Mount Option

The name of the option "Open Both Copies on Private Mount" prior to Release 3.2 is changed to "On Private Mount" in Release 3.2 and a user needs to select either of three options (Always open single copy, Always open both copies and Open both copies on z/OS implied update) within the menu. This is to provide a better solution in regard to **the** state for a data append case with "S" **mode**.

Some applications, like DFSMSHsm may append data to a logical volume. This results in entering the synchronous deferred state **for these volumes** if "Open Both Copies on Private Mount" option is not selected. Since the vast majority of the private mounts don't append, changing the option introduces an excessive amount of overhead, especially when the peer location is a TS7740 which may require a recall.



February 4, 2015

A new option “Open both copies on z/OS implied update” is provided. With the option, both “S” locations are opened only when the host (z/OS) indicates to append data via a flag in a private mount processing.

Select “Always open both copies” to ensure that always both copies are opened regardless if a read or a write will occur (selecting “Always open single copy” or “Always open both copies works in the same way with disabling or enabling “Open Both Copies on Private Mount” option prior to Release 3.2).

If one or both “S” locations are on back end tape, the tape volumes are first recalled into disk cache within those locations. Private mounts can be affected by cache misses when this option is used meaning the mount complete will be delayed until the copy of the logical volume is recalled into both “S” cluster caches.

Other circumstances to consider include:

- If a private mount on both locations is successfully opened, then all read operations use the primary location. If any read fails, then the host read also fails and no failover to the secondary source occurs unless a z/OS DDR swap is initiated by the host.
- If a write operation occurs, both locations receive write data and must synchronize it to TVC disk during each implicit or explicit synchronization command.
- If either location fails to synchronize, the host job either fails or enters the synchronous-deferred state, depending on whether the Synchronous Deferred on Write Failure option is enabled.
- If either or both “S” location are inconsistent or unavailable at the time of the mount, the mount operation will still succeed and read operations will be permitted so long as at least one consistent copy is available at the time of TVC selection. This provides a means to access previously written data even under this condition. Any write operation will then continue or fail depending on whether the Synchronous Deferred on Write Failure option is enabled.

The default setting for this option is “Always open single copy” (don’t open both copies). The default behavior for SMC is to open the logical volume only on one TVC during a private mount. The best TVC choice is used to satisfy the mount. The best TVC choice selection is made with location preferences in this order: synchronized [“S”], RUN [“R”], deferred [“D”], and time delayed [“T”]. If a write operation occurs, the job enters the synchronous-deferred state regardless of whether the Synchronous Deferred on Write Failure option is enabled or disabled.

2.3.3 Mount Results When One or Both “S” Clusters are Unavailable

The table below describes the outcome of a logical mount and subsequent write based on the settings of the two SMC options. For scratch mounts the “On Private Mount” option does not apply.

| Scenario | Mount Type | Sync Deferred on Write Failure | On Private Mount | | | Dual opened? | Mount delay if both "S" clusters are paused or out of physical scratch ² | Results when one or both "S" locations are unavailable at mount time ¹ | | | |
|----------|------------|--------------------------------|-------------------------|-------------------------|---|-------------------|---|---|----------------------------------|--------------------------|--------------------------------------|
| | | | Always open single copy | Always open both copies | Open both copies on z/OS implied update | | | Mount result | Enter sync defer on mount finish | Fail when write received | Enter sync defer when write received |
| 1 | Scratch | Disabled | N/A | N/A | N/A | Both | Yes | Failure | No | N/A | N/A |
| 2 | Scratch | Enabled | N/A | N/A | N/A | Both | Yes | Success ³ | No | No | Yes |
| 3 | Private | Disabled | N/A | Set | N/A | Both | Yes | Success ³ | No | Yes | No |
| 4 | Private | Enabled | N/A | Set | N/A | Both | Yes | Success ³ | No | No | Yes |
| 5 | Private | Disabled | N/A | N/A | Set | Both ⁴ | Yes ⁵ | Success ³ | No | Yes ⁵ | No ⁶ |
| 6 | Private | Enabled | N/A | N/A | Set | Both ⁴ | Yes ⁵ | Success ³ | No | No | Yes |
| 7 | Private | Disabled | Set | N/A | N/A | Single | No | Success ³ | No | No ⁷ | Yes |
| 8 | Private | Enabled | Set | N/A | N/A | Single | No | Success ³ | No | No | Yes |

Table 1 - Mount and Write Results When One or Both "S" Clusters are Unavailable

1. Assumes at least one healthy (and consistent for private mount) cluster when both 'S' locations are unavailable. If no other clusters with a consistent copy of the logical volume are available the mount will fail.
2. Any delay due to a paused state is only applicable if a recall is required.
3. The best TVC choice (preferring locations in this order: "S", "R", "D", and "T") is used to satisfy the mount before it continues.
4. With "Open both copies on z/OS implied update" specified: Both when requested by the host; Single when not requested by the host.
5. With "Open both copies on z/OS implied update" specified: Yes when requested by the host; No when not requested by the host.
6. With "Open both copies on z/OS implied update" specified: No when requested by the host; Yes when not requested by the host.
7. The library always enters the synchronous-deferred state independent of whether the "Synchronous Deferred on Write Failure" option is specified

2.4 Primary "S" Selection

The mounting TS7700 determines which cluster will be designated the primary "S" location based on the criteria below. The Primary location is the cluster used for read operations. Write and sync operations commands are sent to both "S" clusters in parallel.

- One location is identified as the primary when both "S" locations are available.
- The local cluster (mounting cluster) always takes precedence over remote clusters for primary identification.
- If both "S" locations are remote, the "S" location within the same cluster family takes precedence.

- If both "S" locations are within the same family or are external to the mount-point family, then the existing performance-based and latency-based selection criteria is used.
- If only one "S" location is available, it takes precedence over all "R", "D", and "T" locations.
- If neither "S" location is available, the "R" then "D" locations take precedence.
- All read operations are always sent to the primary location.
- If a read fails at any time when two instances of a volume are open, the read does not failover to the secondary location and the host command fails.

2.5 What is the Synchronous-Deferred State?

The synchronous-deferred state means there are "S" copies that have been placed on the deferred copy queue. If one or both "S" locations are not in an active, synchronized state and a write operation is permitted to continue (refer to Table 1 above), the distributed library enters the synchronous-deferred state. The distributed library remains in this state until all "S" copies managed by the distributed library are replicated using an alternative method.

The "S" copies managed by the distributed library are those that were initiated through mounts targeting the distributed library. If one or more distributed libraries enter the synchronous-deferred state, the composite library also enters this state. A distributed library exits the synchronous deferred state when it no longer has deferred "S" copies to be received at any distributed library which resulted from a mount with write targeting the library in the synchronous deferred state. The composite library exits this state when no distributed libraries are in the synchronous-deferred state.

The priority of a synchronous-deferred copy is one level above immediate-deferred. The priority order of the deferred copy queue is as follows:

1. Synchronous-deferred
2. Immediate-deferred
3. Deferred
 - a. Deferred
 - b. COPYRFSH Deferred
 - c. Family Deferred

Note: COPYRFSH and Family Deferred copy queues are handled in the same Deferred priority. COPYRFSH Deferred has the lower priority than the regular Deferred copy queue. Family Deferred has the lowest priority in the Deferred queue priority.

2.6 Sync Mode Copy's Affect on Other Functions

This section describes the affect of "S" copies on existing functions in the TS7700.

2.6.1 Scratch Allocation Assist (SAA)

There is no direct association between the scratch mount candidates and clusters that are identified as needing synchronous mode copies. If an "S" cluster is not identified as a scratch mount candidate, then it is not included in the cluster candidate list that is returned to the host.

2.6.2 Copy Policy Overrides

The synchronous copy mode takes precedence over any copy override settings. The "Force volumes mounted on this cluster to be copied to the local cache" (Force Local Override) is affected by a mount specifying a synchronous mode copy.

If the Force Local Override is selected on a cluster with the "No Copy" copy mode, the override is ignored and up to two "S" clusters are selected as the Tape Volume Cache. For example, imagine a three cluster grid with a copy policy of NSS and the Force Local Override is set on Cluster 0. When cluster 0 receives a host mount request using the Management Class with the NSS copy policy, a local copy is not forced.

2.7 Synchronous Mode Copy and the TS7700 Management Interface

These sections describe the portions of the TS7700 Management Interface (MI) related to Synchronous Mode Copy (SMC).

2.7.1 Defining Copy Policy on the Management Class Panel

An additional copy policy type and two new SMC related options are added to the Management Class panel to support the SMC feature. A fourth copy policy of “Synchronous Copy” is added to the existing three copy policies. If SMC is selected as a copy policy the two SMC related options become available. The two options are “Synchronous Deferred on Write Failure” and “On Private Mount”. You can choose to select none of the options, one of the options, or both. Refer to sections [2.3.1 - Synchronous Deferred on Write Failure Option](#) and [2.3.2 - On Private Mount Option](#) for more details concerning these options.

Note: If “Synchronous Copy” is selected for one cluster, you must select it for two clusters, no more, no less.

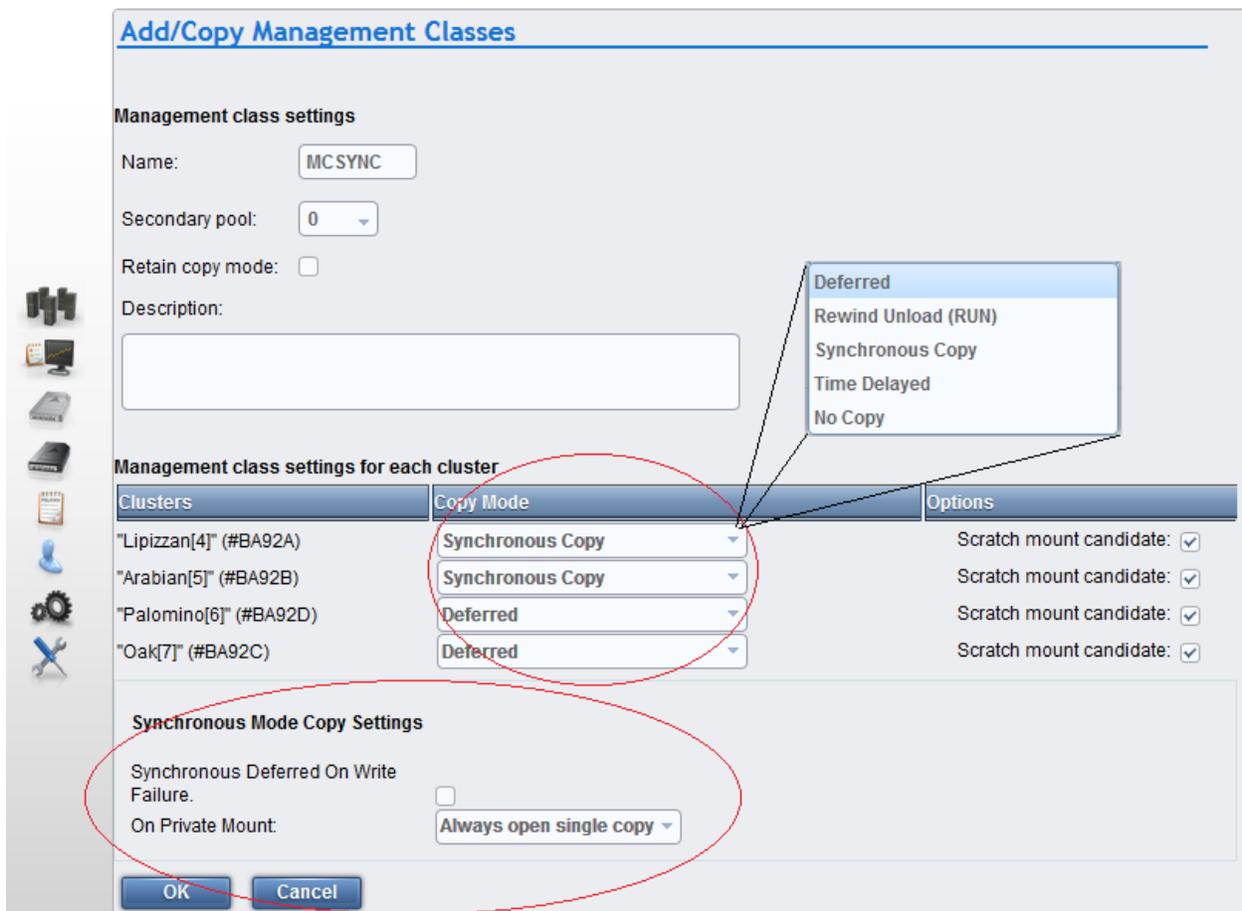


Figure 4 - Add/Copy Management Class Panel

2.7.2 Sync Mode Copy Indications on the TS7700 MI

The following MI panels indicate where synchronous copies reside and which of the two clusters the primary mount cluster is and which is the secondary mount cluster.

The “Virtual Tape Drives” panel shows the Cached Mount Cluster for the currently mounted logical volumes. For a mount with synchronous copies, the column will contain the primary and secondary TVC cluster. If the mount does not have any synchronous copies, the column will contain the cluster whose Tape Volume Cache (TVC) is being used for the mount.

| Address | Mounted ... | Time On Drive | Cache Mount Cluster |
|---------|-------------|--------------------------------|----------------------------|
| vtd000 | Z00001 | 0 hours, 3 minutes, 35 seconds | Lipizzan (4), Palomino (6) |
| vtd001 | Z00002 | 0 hours, 3 minutes, 27 seconds | Lipizzan (4), Palomino (6) |
| vtd002 | Z00003 | 0 hours, 2 minutes, 27 seconds | Lipizzan (4) |
| vtd003 | Z00004 | 0 hours, 2 minutes, 19 seconds | Lipizzan (4) |
| vtd004 | | | |
| vtd005 | | | |
| vtd006 | | | |
| vtd007 | | | |
| vtd008 | | | |

Figure 5 - Virtual Tape Drives Panel

The Virtual Tape Drive Properties panel shows the same Cache Mount Cluster information.

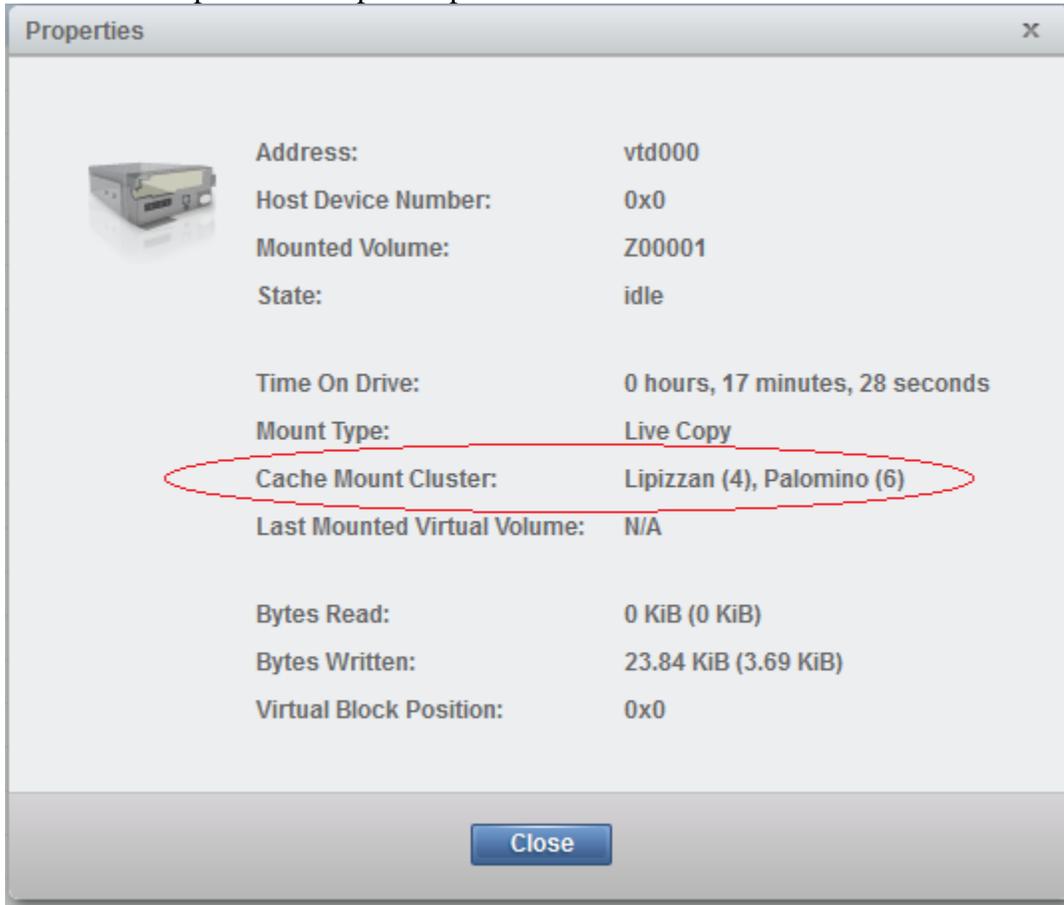


Figure 6 - Virtual Tape Drive Details Panel

February 4, 2015

The Virtual Volume Details page shows the Queue Type and Copy Mode rows. The Queue Type cells will contain a queue type if the logical volume is currently in a queue to be copied. Notice that “Sync” is not a possible value, but “Sync Deferred” is. This is because a normal Synchronous Copy is never in a queue since it is written directly to the cluster. Only when a Synchronous Copy is deferred is it in a queue to be copied.

The Copy Mode indicates what method was used to copy a volume to a cluster.

Virtual volume details:

| | |
|-----------------------------------|---|
| Volser | Z00001 |
| Media Type | Enhanced Capacity Cartridge System Tape |
| Current Volume Size (Device) | 14.9 MiB |
| Maximum Volume Capacity (Device) | 800 MiB |
| Current Owner | "Lipizzan[4]" (#BA92A) |
| Currently Mounted | No |
| vNode | - |
| Virtual Drive | - |
| Cached Copy Used for Mount | "Lipizzan[4]" (#BA92A) |
| Mount State | - |
| Cache Management Preference Group | 1 |
| Last Accessed by a Host | Oct 15, 2014, 1:40:16 PM |
| Last Modified | Oct 15, 2014, 1:39:48 PM |
| Category | 000F |
| Storage Group | ----- |
| Management Class | NNNRSRST |
| Storage Class | CP2IART |
| Data Class | ----- |
| Volume Data State | Active |
| Flash Copy | Not Active |
| Earliest Deletion On | - |
| Logical WORM | No |

- **Queue Type**
 - Rewind Unload (RUN)
 - Deferred
 - Sync Deferred
 - Immediate Deferred
 - Time Delayed
- **Copy Mode**
 - Rewind Unload (RUN)
 - Deferred
 - No Copy
 - Sync
 - Time Delayed
 - Exist

Cluster-specific Virtual Volume Properties:

| | "Lipizzan[4]" (#BA92A) | "Arabian[5]" (#BA92B) | "Palomino[6]" (#BA92D) | "Oak[7]" (#BA92C) |
|---------------------------|------------------------|-----------------------|------------------------|--|
| In Cache | Yes | unknown | - | - |
| Device Bytes Stored | 14.9 MiB (Device) | 0 MiB (Device) | 14.9 MiB (Device) | 0 MiB (Device) |
| Primary Physical Volume | None | None | - | - |
| Secondary Physical Volume | None | None | - | - |
| Copy Activity | Complete | - | Required | Time delayed until Thu, Oct 16, 2014, 1:39:48 PM GMT+09:00 |
| Queue Type | - | - | Sync Deferred | Time Delayed |
| Copy Mode | Synchronous Copy | - | Synchronous Copy | Time Delayed |
| Deleted | - | - | - | - |
| Removal Residency | - | - | - | - |
| Removal Time | - | - | - | - |
| Partition Number | 2 | - | - | - |
| Premigration Delay Time | - | - | - | - |
| Storage Preference | - | - | Prefer Keep | Prefer Keep |

Figure 7 - Logical Volume Details Panel

2.8 Host APAR OA37267

This z/OS Host APAR is part of the full function support for Release 2.1 of the TS7700 Virtualization Engine. This includes enhancements for the 3957 Model V06 and the V07 (TS7740 Virtualization Engines) and for the 3957 VEA and VEB (TS7720 Virtualization Engines).

For full support, install this APAR OA37267 (z/OS V1R11, V1R12 and V1R13). This APAR brings in the needed support from the other component: AOM (OA37268).

Although not required to use the new functionality provided by Release 2.1, the recommendation is to install this host software when running with Release 2.1 of the TS7700 Virtualization Engine.

2.8.1 Synchronous Mode Copy

Release 2.1 of the TS7700 Virtualization Engine introduces a new copy mode option that provides tape copy capabilities up to synchronous level granularity across two clusters within a multi-cluster grid configuration. Copies to other clusters in the grid can be made using immediate or deferred mode. When data is written to a logical volume in the TS7700 Virtualization Engine, a copy policy is applied through the assigned Management Class and policy settings established at the library. The policy settings at the library indicate in which distributed libraries (or clusters) a copy of the data (logical volume) should exist and how those copies should be made (through immediate, deferred or synchronous replication policies). The synchronous copy mode option provides tape copy capabilities up to synchronous level granularity across two clusters within a multi-cluster grid configuration. As jobs (or applications) write data to tape, with the synchronous copy mode option, the TS7700 Virtualization Engine will fork (dual write) the writes to two clusters simultaneously which provides dual copy consistency for all data written prior to the last implicit or explicit synchronize operation.

When the library encounters an issue and cannot maintain the two synchronous copies of the logical volume, there are two failure options that can be used. By default, the library will "Fail the Host Operation" if it cannot maintain two synchronous copies of a volume. Once a failure occurs both copies of the volume will remain consistent up to the last successful synchronize operation. If instead, the "Synchronous Deferred On Write Failure" option is enabled through the library's management interface (MI), host tape commands will be allowed to continue and once the volume is demounted, asynchronous replication will be used to provide a copy of the volume to the failed clusters.

When the ability to synchronize is broken, and the "Synchronous Deferred on Write Failure" option is enabled, the distributed library associated with the mount and its composite library will also enter the "synchronous-deferred" state and will remain in this state until all failed synchronous mode copy requests are complete (for that distributed library). The distributed library that enters this state will be the cluster in which the mount request was issued (since it is

this cluster that is managing the synchronize operations for that mount). The following host messages will be issued:

- CBR3730E One or more synchronous mode copy operations deferred in library library-name.
- CBR3731I All deferred synchronous mode copies completed in library library-name.

The LIBRARY REQUEST, composite-library, STATUS, GRID command can be used to determine which distributed libraries have a synchronous-deferred copy queue with a count greater than zero.

2.8.2 Other Tape Library Support Enhancements

1. Enhancement to the DISPLAY SMS, LIBRARY command - New status message "Synchronous mode copy operations deferred" can be displayed. This status message indicates at least one synchronous mode copy operation in the library was unable to complete. Distributed libraries will remain in this state until all of the synchronous copies that it is managing have been successfully replicated. Composite libraries enter this state when one or more of its distributed libraries enter this state. Composite libraries exit this state when all of its distributed libraries are not in this state.
2. Enhancement to the DISPLAY SMS, VOLUME command - New status line "LM CATEGORY: xxxx" will be displayed for volumes which are library resident in an automated (or virtual) tape library. This line displays the actual 4 digit hexadecimal volume category. If the library manager category value cannot be obtained, '0000' will be displayed for the category value. This status line will not be displayed for volumes which are not library resident or volumes which reside in a manual tape library.
3. Enhancement to CBRXLCS FUNC=QVR - New fields have been added to CBRTVI that indicate whether or not a cluster in a multi-cluster grid has a validated consistent copy of a volume. Each bit in new field TVICCMAP maps to a cluster (left to right starting with cluster 0). The new copy cluster bit mapping (TVICCMAP) is returned starting at TS7700 Release 2.1. TVICCMOK indicates whether or not TVICCMAP is valid to use, and even if TVICCMAP is valid to use, the actual validity of TVICCMAP is based on the availability of each cluster in a TS7700 multi-cluster grid so if one or more of the clusters are not available (and cannot be queried), the bit indicators in TVICCMAP may not be appropriately reflected.
4. Copy export enhancements - For the copy export acceleration enhancement there are changes to the copy export list file sample JCL (CBRSPX03) documenting the library's new LMTDBPVL option. To accelerate the copy export process, this new option limits the number of copy exported stacked volumes which have the TS7740 database backup written to them.

Then with the copy export merge enhancements and the ability to place copy exported stacked volumes into an existing TS7740, manual updates will need to made (as appropriate) to the host data bases (tape configuration data base (TCDB) and tape

management system data base) to show in which composite library the copy exported logical volumes now reside.

5. Enhancement to delete expire hold - Starting with Release 2.1 of the TS7700, the library now allows the host to move restricted volumes to a non-fast ready (private) category in case they were accidentally returned to scratch. Prior to this, restricted volumes could not be used to satisfy a mount request and could not be moved to a different category until their expire time had elapsed.

2.9 Host APAR OA40572

When Sync Mode Copy “Open both copies on z/OS implied update” option is used, both “S” locations are opened only when the host (z/OS) specifies a flag (append indicator) in the mount processing. The flag is set when this z/OS APAR which has been released for Release 3.1 of the TS7700 Virtualization Engine is applied and DISP=NEW is specified in the job, then this APAR is required at a minimum.

2.9.1 New and Changed Host Messages

2.9.1.1 CBR1110I

This is a changed message. Displays new status line: "Synchronous mode copy operations deferred".

2.9.1.2 CBR1180I

This is a changed message. Displays new status line: "LM CATEGORY: xxxx".

2.9.1.3 CBR3730E

This is a new message. CBR3730E One or more synchronous mode copy operations deferred in library library-name.

Explanation: At least one synchronous mode copy operation in library library-name was unable to complete. The distributed-library will remain in this state until all of the synchronous copies that it is managing have been successfully replicated. When one or more distributed libraries enter this state, the composite library enters this state. The composite library exits this state when all of its distributed libraries are not in this state.

System Action: The synchronous mode copy operation is deferred. This message is retained until all of the synchronous mode copy operations that were deferred have completed.

Message CBR3731I is also issued after all of the deferred synchronous mode copies have completed.

February 4, 2015



Operator Response: Determine why the synchronous mode copies have been deferred, then search problem reporting databases for a fix for the problem. If no fix exists, contact the IBM Support Center.

2.9.1.4 CBR3731I

This is a new message. CBR3731I All deferred synchronous mode copies completed in library library-name.

Explanation: All synchronous mode copy operations that had been deferred in library library-name have been completed. One or more synchronous mode copy operations had previously been deferred. Message CBR3730E indicates when the library had initially entered this state.

2.9.1.5 CBR4171I

This is an existing message. Mount failed. LVOL=logical-volser, LIB=library-name, PVOL=physical-volser, RSN=reason-code. New reason code X'27' - Unable to meet the synchronous copy requirements.

3 Host Applications and Sync Events

The data is always being written to both clusters at the same time independent of implicit or explicit sync points. Though copy is in the name, it is actually a real time duplex or dual write. The sync point only flushes any data in buffers up to that point on tape, approximately 1MB or 2MB. A read from one cluster's disk cache is not needed to get the data to the second "S" cluster. Therefore, a job that has no sync points (implicit or explicit) would be at most a couple megabytes behind (compressed data). Any close processing will include a tape mark which will act as an implicit sync. Thus, all applications can use this form of replication independent of when or if they issue implicit or explicit sync points. The worst case lag is around 2MB and that would be flushed to persistent disk cache during close processing before indicating to the host that the close is complete.

3.1 Explicit and Implicit Tape Synchronous Operations

This section lists the explicit and implicit tape sync operations.

Explicit synchronization operations are commands that host applications issue intentionally in order to synchronize data to tape.

- CCW(SYNC) – Synchronize

Implicit sync operations are those commands which the host application may issue which will result in synchronization outboard, but synchronization may not have been the intended reason for the operation.

- Write Tape Mark – Often viewed as an explicit sync operation
- Rewind
- Rewind Unload
- Synchronize Buffer
- Forward Space Block
- Forward Space File
- Backward Space Block
- Backward Space File
- Erase Gap
- Data Security Erase
- Locate Block
- Read Forward
- Read Previous
- Read Backwards (and there is buffered write data)
- A write command is issued in a command chain following either Mode Set command which specifies the tape write immediate mode or a Set Tape Write Immediate command
- Logical end of volume point is reached

4 Synchronous Mode Copy Performance

Refer to the “IBM® Virtualization Engine™ TS7720 and TS7740 Releases 1.5, 1.6 through R2.1 PGA2, 3.0 and 3.1 Performance White Paper on Techdocs for detailed performance information on the TS7700 including Sync Mode Copy performance.

<http://http://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102421>

Performance testing has shown the sustained performance with a two cluster grid using synchronous mode copy (SMC) performs better than immediate mode (RUN) copy but slower than deferred mode copy. The SMC method performs better than the RUN method over the long run because it has less work to do.

With a RUN copy the logical volume is written into one cluster’s disk cache then, at RUN time, is read from the disk cache and copied across the grid links. With SMC, the read from disk cache isn’t necessary to replicate the logical volume. This is because the logical volume is written to two clusters at the same time. This reduces the overhead on the disk cache repository. However, the copy to a second cluster uses the TS7700’s CPU and grid bandwidth at the time of the write instead of at RUN time.

Looking at the big picture, SMC uses fewer resources as compared to a RUN copy. The two methods use TS7700 resources at different times during the logical volume mount. The comparison below is for a Sync Mode Copy where one cluster is the local cluster and the other cluster is remote.

- RUN Immediate Copy
 - CPU and memory to process write from host
 - CPU and disk cache bandwidth to write logical volume to disk cache
 - CPU and disk cache bandwidth to read logical volume from disk cache
 - CPU and grid bandwidth to copy logical volume across the grid
 - CPU, grid bandwidth, and disk cache bandwidth to copy volume into second cluster’s disk cache
- SMC Immediate Copy
 - CPU and memory to process write from host
 - CPU and disk cache bandwidth to write logical volume to disk cache
 - CPU and grid bandwidth to copy logical volume across the grid (at the same time as write to first cluster)
 - CPU, grid bandwidth, and disk cache bandwidth to copy volume into second cluster’s disk cache

5 Disclaimers:

Copyright © 2014 by International Business Machines Corporation.

No part of this document may be reproduced or transmitted in any form without written permission from IBM Corporation.

Product data has been reviewed for accuracy as of the date of initial publication. Product data is subject to change without notice. This information could include technical inaccuracies or typographical errors. IBM may make improvements and/or changes in the product(s) and/or programs(s) at any time without notice.

References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business. Any reference to an IBM Program Product in this document is not intended to state or imply that only that program product may be used. Any functionally equivalent program, that does not infringe IBM's intellectually property rights, may be used instead. It is the user's responsibility to evaluate and verify the operation of any non-IBM product, program or service.

THE INFORMATION PROVIDED IN THIS DOCUMENT IS DISTRIBUTED "AS IS" WITHOUT ANY WARRANTY, EITHER EXPRESS OR IMPLIED. IBM EXPRESSLY DISCLAIMS ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NON INFRINGEMENT. IBM shall have no responsibility to update this information. IBM products are warranted according to the terms and conditions of the agreements (*e.g.*, IBM Customer Agreement, Statement of Limited Warranty, International Program License Agreement, etc.) under which they are provided. IBM is not responsible for the performance or interoperability of any non-IBM products discussed herein. The customer is responsible for the implementation of these techniques in its environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. Unless otherwise noted, IBM has not tested those products in connection with this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents or copyrights. Inquiries regarding patent or copyright licenses should be made, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

Trademarks

The following are trademarks or registered trademarks of International Business Machines in the United States, other countries, or both.

IBM, TotalStorage, DFSMS/MVS, S/390, z/OS, System z, and zSeries.

Other company, product, or service names may be the trademarks or service marks of others.

End of Document