IBM® Virtualization Engine™ TS7720 and TS7740
Releases 1.6, 1.7, 2.0, 2.1 and 2.1 PGA2
Performance White Paper
Version 2.0

By Khanh Ly
Tape Performance
IBM Tucson

© Copyright IBM Corporation
# Table of Contents

Table of Contents ................................................................................................. 2  
Introduction ........................................................................................................... 3  
TS7700 Overview .................................................................................................. 5  
Hardware Configuration ....................................................................................... 9  
TS7700 Performance Overview ............................................................................ 11  
TS7700 Basic Performance .................................................................................. 13  
Additional Performance Metrics ......................................................................... 26  
Performance at Distance ..................................................................................... 37  
Performance Tunings ......................................................................................... 40  
Performance Tools .............................................................................................. 44  
Conclusions .......................................................................................................... 45  
Acknowledgements .............................................................................................. 46
Introduction

This paper provides performance information for the IBM Virtualization Engine™ TS7720 and TS7740, which are the two current products in the TS7700 Virtualization Engine family of products. The paper is intended for use by IBM field personnel and their customers in designing virtual tape solutions for their applications. This is an update to the previous TS7700 paper dated March 10, 2012 and reflects changes for release 2.1 PGA2, which introduces the new disk cache model 3956-CS9 for TS7720. Cache throughput for the CS9 TS7720 has improved over the CS8 and thus significantly increases performance runs which are limited by disk cache throughput. For example, the RR and SS dual active runs. This improvement is most relevant for small drawer configurations when comparing CS8 to CS9. Currently, the CS9 is available only via RPQ and limited to sales in the U.S.

The following are performance related changes for the following releases:

**Release 1.6**

- Four-way grid
- TS7720/TS7740 hybrid
- Cluster families
- New tuning setting
  - DCT smoothing
  - Number of reclaim tasks

**Release 1.7**

- New disk cache controller models 3956-CC8 and 3956-CS8
- External storage expansion frame for TS7720
  - Supports up to two additional SATA strings of 6-drawers each (including controllers).
  - Two 3956-CS8 cache controllers
  - Zero to ten 3956 XS7 expansion drawers
- Single drawer increments on cache MES (base and expansion)
- Large DDM support
  - SATA 2TB drive
    - Up to 162 TB capacity in base frame
    - Up to 280 TB additional capacity in expansion frame
  - Fibre channel 600GB drive
    - Up to 28 TB maximum configuration
- Doubling the premigration default setting
  - PMPIOR = 1600
  - PMTHLVL = 2000

**Release 2.0**

- Model V07 and VEB Server Platforms
- Four 1Gb Grid Ethernet links
- Two 10Gb Grid Ethernet Links
- 8 Gb Fibre Channel Adapter
- 8 Gb Fibre Channel Switch
- Scratch Allocation Assist
Release 2.1

- Sync Mode Copy
- TS1140 JAG4 Support
- Increasing default premigration drives from 6 to 10.

Release 2.1 PGA2

- New disk cache controller model 3956-CS9.
- Support up to nine XS9 expansion drawers in the base frame.
- Higher performance and more capacity.
- Support data encryption at rest.
TS7700 Overview

Notation Convention

The following convention is used in this paper:

<table>
<thead>
<tr>
<th>Orientation</th>
<th>Binary</th>
<th>Decimal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Symbol Values</td>
<td>Symbol Name</td>
</tr>
<tr>
<td>kibibyte KiB</td>
<td>$2^{10}$</td>
<td>kilobyte KB</td>
</tr>
<tr>
<td>mebibyte MiB</td>
<td>$2^{20}$</td>
<td>megabyte MB</td>
</tr>
<tr>
<td>gibibyte GiB</td>
<td>$2^{30}$</td>
<td>gigabyte GB</td>
</tr>
<tr>
<td>tebibyte TiB</td>
<td>$2^{40}$</td>
<td>terabyte TB</td>
</tr>
<tr>
<td>pebibyte PiB</td>
<td>$2^{50}$</td>
<td>petabyte PB</td>
</tr>
</tbody>
</table>

Product Evolution

The TS7700 has its roots in the IBM TotalStorage® Virtual Tape Server (VTS) products, but features new hardware components and redesigned product architecture. Figure 1 below shows the evolution of the TS7740 from the monolithic VTS architecture to a scalable architecture consisting of functional nodes.

TS7740 - Architectural Evolution

- Break up the monolithic VTS into scalable pieces (nodes)

![Diagram of TS7740 Architectural Evolution]

Figure 1. TS7740 Architectural Evolution

© Copyright IBM Corporation
Product Architecture

The TS7700 is built upon a unit of scalability called a “cluster”. The TS7700 is available with a single cluster, or as two, three, or four clusters linked together via a TCP/IP interface, termed a “Grid”. Five or six clusters are possible via iRPQ. Each cluster supports up to four FICON® host channels and 256 virtual tape drives. The TS7740 supports up to sixteen physical tape drives, while the TS7720 is a disk-only cluster. The basic building blocks of a cluster are nodes, the TS7700 having two, a “Vnode” and an “Hnode”. A Vnode is a code stack that performs all of the actions needed to present a library image and drive images to a host. An Hnode is a code stack that performs all of the actions needed to coordinate the contents of the disk cache with the data on backend physical tape. It also includes the logic for managing changes and replication of the data across different clusters.

Figure 2 shows an architectural view of the Vnode and Hnode, as well as a three-cluster Grid. A Vnode and Hnode running in the same physical controller can be referred to collectively as a “Gnode”.

© Copyright IBM Corporation
Performance Evolution

The TS7700 modular architecture continues to provide a base for product growth in both performance and functionality. Figures 3 and 4 shows the write and read performance improvement histories.

Figure 3. VTS/TS7700 Standalone Write Bandwidth. All runs were with 128 concurrent jobs, each job writing 800 MiB (300 MiB volumes @ 2.66:1 compression) using 32KiB blocks, and QSAM BUFNO = 20.

Notes: (n-DRs - 4x4Gb – z10) refers to (number of cache drawers – four 4Gb FICON channels – z10 host)
Figure 4. VTS/TS7700 Standalone Read Bandwidth. All runs were with 128 concurrent jobs, each job writing 800 MiB (300 MiB volumes @ 2.66:1 compression) using 32KiB blocks, and QSAM BUFNO = 20. See definition for Read Hit in the section “TS7700 Performance Overview”.

Notes: (n-DRs - 4x4Gb – z10) refers to (number of cache drawers – four 4Gb FICON channels – z10 host)
Hardware Configuration

The following hardware was used in performance measurements. Except stated otherwise, all measurements were made using the maximum resources available in term of cache size.

## R 2.0 Hardware Setup

<table>
<thead>
<tr>
<th>TS7700</th>
<th>Model</th>
<th>Cache size/ number of drawers</th>
<th>Tape Drives</th>
<th>IBM System z™ Host</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS7740 V07</td>
<td>3956</td>
<td>28 TB / 4 DRs*</td>
<td>12 TS1130</td>
<td>z10 with four 4Gb FICON</td>
</tr>
<tr>
<td>(Standalone)</td>
<td>CC8/CX7</td>
<td></td>
<td></td>
<td>channels</td>
</tr>
<tr>
<td>TS7720 VEB</td>
<td>3956</td>
<td>162 TB / 7 DRs or 442 TB / 19 DRs</td>
<td>N/A</td>
<td>z10 with four 4Gb FICON</td>
</tr>
<tr>
<td>(Standalone)</td>
<td>CS8/XS7</td>
<td></td>
<td></td>
<td>channels</td>
</tr>
<tr>
<td>TS7740 V07</td>
<td>3956</td>
<td>28 TB / 4 DRs</td>
<td>12 TS1130</td>
<td>z10 with four 4Gb FICON</td>
</tr>
<tr>
<td>(Grid)</td>
<td>CC8/CX7</td>
<td></td>
<td></td>
<td>channels</td>
</tr>
<tr>
<td>TS7720 VEB</td>
<td>3956</td>
<td>162 TB / 7 DRs or 442 TB / 19 DRs</td>
<td>N/A</td>
<td>z10 with four 4Gb FICON</td>
</tr>
<tr>
<td>(Grid)</td>
<td>CS8/XS7</td>
<td></td>
<td></td>
<td>channels</td>
</tr>
</tbody>
</table>

* DRs refers to Drawers

## R 2.1 Hardware Setup

<table>
<thead>
<tr>
<th>TS7700</th>
<th>Model</th>
<th>Cache size/ number of drawers</th>
<th>Tape Drives</th>
<th>IBM System z™ Host</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS7740 V07</td>
<td>3956</td>
<td>28 TB / 4 DRs*</td>
<td>12 TS1140</td>
<td>z10 with four 4Gb FICON</td>
</tr>
<tr>
<td>(Standalone)</td>
<td>CC8/CX7</td>
<td></td>
<td></td>
<td>channels</td>
</tr>
<tr>
<td>TS7720 VEB</td>
<td>3956</td>
<td>162 TB / 7 DRs</td>
<td>N/A</td>
<td>z10 with four 4Gb FICON</td>
</tr>
<tr>
<td>(Standalone)</td>
<td>CS8/XS7</td>
<td></td>
<td></td>
<td>channels</td>
</tr>
<tr>
<td>TS7740 V07</td>
<td>3956</td>
<td>28 TB / 4 DRs</td>
<td>12 TS1140</td>
<td>z10 with four 4Gb FICON</td>
</tr>
<tr>
<td>(Grid)</td>
<td>CC8/CX7</td>
<td></td>
<td></td>
<td>channels</td>
</tr>
<tr>
<td>TS7720 VEB</td>
<td>3956</td>
<td>162 TB / 7 DRs</td>
<td>N/A</td>
<td>z10 with four 4Gb FICON</td>
</tr>
<tr>
<td>(Grid)</td>
<td>CS8/XS7</td>
<td></td>
<td></td>
<td>channels</td>
</tr>
</tbody>
</table>

* DRs refers to Drawers
### R 2.1 PGA2 Hardware Setup

<table>
<thead>
<tr>
<th>TS7700</th>
<th>Model</th>
<th>Cache size/ number of drawers</th>
<th>Tape Drives</th>
<th>IBM System z™ Host</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS7720 VEB (Standalone)</td>
<td>3956 CS9/XS9</td>
<td>119 TB / 5 DRs or 239 TB / 10 DRs</td>
<td>N/A</td>
<td>z10 with four 4Gb FICON channels</td>
</tr>
<tr>
<td>TS7720 VEB (Grid)</td>
<td>3956 CS9/XS9</td>
<td>239 TB / 10 DRs</td>
<td>N/A</td>
<td>z10 with four 4Gb FICON channels</td>
</tr>
<tr>
<td>TS7740 V07 (Standalone)</td>
<td>3956 CC8/CX7</td>
<td>28 TB / 4 DRs</td>
<td>12 TS1140</td>
<td>z10 with four 4Gb FICON channels</td>
</tr>
</tbody>
</table>

* DRs refers to Drawers
**TS7700 Performance Overview**

**Performance Workloads and Metrics**

Performance shown in this paper has been derived from measurements that generally attempt to simulate common user environments, namely a large number of jobs writing and/or reading multiple tape volumes simultaneously. Unless otherwise noted, all of the measurements were made with 128 simultaneously active virtual tape jobs per active cluster. Each tape job was writing or reading 800 MiB of uncompressed data using 32 KiB blocks and QSAM BUFNO=20, using data that compresses at the TS7700 at 2.66:1. Measurements were made with four 4-gigabit (Gb) FICON channels on a z10 host. All runs are begun with the virtual tape subsystem inactive.

Unless otherwise stated, all runs were made with default tuning values (DCOPYT=125, DCTAVGTD=100, PMPIOR=1600, PMTHLVL=2000, ICOPYT=ENABLED, Reclaim disabled, Number of premigration drives per pool is 6 for R 2.0 and is 10 for R 2.1). Refer to the IBM® Virtualization Engine TS7700 Series Best Practices - Understanding, Monitoring and Tuning the TS7700 Performance white paper for detailed description of different tuning settings.

**Types of Throughput**

Because the TS7720 is a disk-cache only cluster. The read and write data rates have been found to be fairly consistent throughout a given workload. Because the TS7740 contains physical tapes to which the cache data will be periodically written, the TS7740 has been found to exhibit four basic throughput rates: peak write, sustained write, read-hit, and recall. These four rates are described below.

**Peak and Sustained Write Throughput.**

For the TS7740 a measurement is not begun until all previously written data have been copied, or premigrated, from the disk cache to physical tape. Starting with this initial condition, data from the host is first written into the TS7740 disk cache with little if any premigration activity taking place. This allows for a higher initial data rate, and is termed the "peak" data rate. Once a pre-established threshold is reached of non-premigrated data, the amount of premigration is increased, which can reduce the host write data rate. This threshold is called the premigration priority threshold, and has default value of 1600 gigabytes (GB). When a further threshold of non-premigrated data is reached, the incoming host activity is actively throttled to allow for increased premigration activity. This throttling mechanism operates to achieve a balance between the amount of data coming in from the host and the amount of data being copied to physical tape. The resulting data rate for this mode of behavior is called the "sustained" data rate, and could theoretically continue on forever, given a constant supply of logical and physical scratch tapes. This second
threshold is called the premigration throttling threshold, and has a default value of 2000 gigabytes (GB). These two thresholds can be used in conjunction with the peak data rate to project the duration of the peak period. Note that both the priority and throttling thresholds can be increased via a host command line request.

Read-hit and Recall Throughput

Similar to write activity, there are two types of TS7740 read performance: “read-hit” (also referred to as “peak”) and “recall” (also referred to as “read-miss”). A read hit occurs when the data requested by the host is currently in the local disk cache. A recall occurs when the data requested is no longer in the disk cache and must be first read in from physical tape. Read-hit data rates are typically higher than recall data rates.

These two read performance metrics, along with peak and sustained write performance are sometimes referred to as the “four corners” of virtual tape performance. The following charts in the paper show three of these corners:

1. peak write
2. sustained write
3. read hit

Recall performance is dependent on several factors that can vary greatly from installation to installation, such as number of physical tape drives, spread of requested logical volumes over physical volumes, location of the logical volumes on the physical volumes, and length of the physical media. Because these factors are hard to control in the laboratory environment, recall is not part of lab measurement.

Grid Considerations

Up to four TS7700 clusters can be linked together to form a Grid configuration. Five- or six-way grid configuration is available via iRPQ. The connection between these clusters is provided by two 1-Gb TCP/IP links (default). Four 1-Gb links or two 10-Gb links options are also available. Data written to one TS7700 cluster can be optionally copied to the other cluster(s).

Data can be copied between the clusters in either deferred or RUN (also known as “Immediate”) copy mode. In RUN copy mode the rewind-unload response at job end is held up until the received data is copied to the other cluster(s). In deferred copy mode data is queued for copying, but the copy does not have to occur prior to job end. Deferred copy mode allows for a temporarily higher host data rate than RUN copy mode because copies to the other cluster(s) are delayed until end of the host I/O, and can be useful for meeting peak workload demands. Care must be taken, however, to be certain that there is sufficient recovery time for deferred copy mode so that the deferred copies can be completed prior to the next peak demand.
Deferred copies are controlled during heavy host I/O with a default setting of 125 for the Deferred Copy Throttling (DCT). More priority can be given to deferred copying by lowering the DCT value. Refer to the IBM® Virtualization Engine TS7700 Series Best Practices - Understanding, Monitoring and Tuning the TS7700 Performance white paper for details.

**TS7700 Basic Performance**

The following sets of graphs show basic TS7720 and TS7740 bandwidths. The graphs in Figures 5 and 6 show single cluster, or standalone, performance for the two offerings. The performance metric shown in these and all other data rate charts in this paper is host-view (uncompressed) MB/sec.

**TS7720 Standalone Performance**

![TS7720 Standalone Bandwidth](image)

Figure 5. TS7720 Standalone Write and/or Read Bandwidth. All runs were made with 128 concurrent jobs, each job writing 800 MiB (uncompressed) using 32KiB blocks, QSAM BUFNO = 20, using four 4Gb FICON channels from a z10 LPAR.

Notes: Mixed workload refers to 50% write plus 50% read hit workload driven from the host. The resulting data rate seen from TS7720 is not 50-50 read/write mixed; normally it is more read than write.
Figure 6. TS7740 Standalone Write and/or Read Bandwidth. All runs were made with 128 concurrent jobs, each job writing 800 MiB (uncompressed) using 32KiB blocks, QSAM BUFNO = 20, using four 4Gb FICON channels from a z10 LPAR.

Notes: Mixed workload refers to 50% write plus 50% read hit workload driven from the host. The resulting data rate seen from TS7740 is not 50-50 read/write mixed; normally it is more read than write.

All the runs for R 2.0 were made with the default setting of 6 premigration drives per pool. Higher sustained write rate were observed with more drives reserved for premigration (see Premigration Drive Tuning on pages 37-39).

In R2.1, the default setting for premigration drives per pool for V07 based configurations changed from 6 to 10, which can result in a significant performance improvement for sustained data rates. For V07 based configurations, each additional drive enabled above 6 has proven to improve sustained data rates.
TS7720 - TS7740 Grid Performance

Figures 7, 8, 10 through 12, 14, 16, 18, 20, 22 and 24 show Grid bandwidth. Only figures 10 and 14 have Release 1.7 grid performance data.

For these charts "D" stands for deferred copy mode, and "R" stands for RUN (immediate) copy mode. For example, in Figure 6, RR represents RUN for cluster 0, and RUN for cluster 1. All measurements for these graphs were made at zero or near-zero distance between clusters.

Two-way TS7700 Grid with Single Active Cluster Performance

Figure 7. Two-way TS7720 Grid Maximum Bandwidth for Single Active Cluster. Unless otherwise stated, all runs were made with 128 concurrent jobs. Each job writing or reading 800 MiB (300MiB volumes @ 2.66:1 compression) using 32 KiB block size, QSAM BUFFNO = 20, using four 4Gb FICON channels from a z10 LPAR. Clusters are located at zero or near zero distance to each other in laboratory setup. DCT=125.

* 64 jobs – 64 jobs per active cluster.
Figure 8. Two-way TS7740 Grid Maximum Bandwidth for Single Active Cluster. All runs were made with 128 concurrent jobs. Each job writing or reading 800 MiB (300MiB volumes @ 2.66:1 compression) using 32 KiB block size, QSAM BUFFNO = 20, using four 4Gb FICON channels from a z10 LPAR. Clusters are located at zero or near zero distance to each other in laboratory setup. DCT=125.

In R2.1, the default setting for premigration drives per pool for V07 based configurations changed from 6 to 10, which can result in a significant performance improvement for sustained data rates. For V07 based configurations, each additional drive enabled above 6 has proven to improve sustained data rates.
Figure 9. Two-way TS7700 Hybrid Grid H1

![Diagram of Two-way TS7700 Hybrid Grid H1]

Figure 10. Two-way TS7700 Hybrid Grid H1 Maximum Bandwidth for Single Active Cluster. All runs were made with 128 concurrent jobs. Each job writing or reading 800 MiB (300 MiB volumes @ 2.66:1 compression) using 32 KiB block size, QSAM BUFFNO = 20, using four 4Gb FICON channels from a z10 LPAR. Clusters are located at zero or near zero distance to each other in laboratory setup. DCT=125.

© Copyright IBM Corporation
Two-way TS7700 Grid with Dual Active Cluster Performance

*64 jobs – 128 jobs (64 jobs per active cluster)*

Figure 11. Two-way TS7720 Grid Maximum Bandwidth for Dual Active Clusters. Unless otherwise stated, all runs were made with 256 concurrent jobs (128 jobs per active cluster). Each job writing or reading 800 MiB (300 MiB volumes @ 2.66:1 compression) using 32 KiB block size, QSAM BUFFNO = 20, using eight 4Gb FICON channels from a z10 LPAR (four 4Gb channels per active cluster). Clusters are located at zero or near zero distance to each other in laboratory setup. DCT=125.

For 2-way TS7720 dual active clusters SS run, reducing the current jobs per active cluster from 128 to 64 give about 10% performance improvement.
Two-way TS7740 Grid Maximum Bandwidth for Dual Active Clusters

Figure 12. Two-way TS7740 Grid Maximum Bandwidth for Dual Active Clusters. Unless otherwise stated, all runs were made with 256 concurrent jobs (128 jobs per active cluster). Each job writing or reading 800 MiB (300 MiB volumes @ 2.66:1 compression) using 32 KiB block size, QSAM BUFFNO = 20, using eight 4Gb FICON channels from a z10 LPAR (four 4Gb channels per active cluster). Clusters are located at zero or near zero distance to each other in laboratory setup. DCT=125.

For 2-way TS7740 dual active clusters SS run, reducing the current jobs per active cluster from 128 to 64 give over 27% performance improvement in sustained rate.
Figure 13. Two-way TS7700 Hybrid Grid H2

Figure 14. Two-way TS7700 Hybrid Grid H2 Maximum Bandwidth for Dual Active Clusters. Unless otherwise stated, all runs were made with 256 concurrent jobs (128 jobs per active cluster). Each job writing or reading 800 MiB (300 MiB volumes @ 2.66:1 compression) using 32 KiB block size, QSAM BUFFNO = 20, using eight 4Gb FICON channels from a z10 LPAR (four 4Gb channels per active cluster). Clusters are located at zero or near zero distance to each other in laboratory setup. DCT=125.

* 64 jobs – 128 jobs (64 jobs per active cluster)
Three-way TS7700 Grid with Dual Active Cluster Performance

Figure 15. Three-way TS7700 Hybrid Grid H3

Three-way TS7700 Hybrid Grid H3 R 1.6 - R 1.7 - R 2.0 - R 2.1 - R 2.1 PGA2 Performance

(Dual Active Clusters)

* 64 jobs – 128 jobs (64 jobs per active cluster)

Figure 16. Three-way TS7700 Hybrid Grid H3 Maximum Bandwidth for Dual Active Clusters. Unless otherwise stated, all runs were made with 256 concurrent jobs (128 jobs per active cluster). Each job writing or reading 800 MiB (300 MiB volumes @ 2.66:1 compression) using 32 KiB block size, QSAM BUFFNO = 20, using eight 4Gb FICON channels from a z10 LPAR (four 4Gb channels per active cluster). Clusters are located at zero or near zero distance to each other in laboratory setup. DCT=125.
Figure 17. Three-way TS7700 Hybrid Grid H4

Figure 18. Three-way TS7700 Hybrid Grid H4 Maximum Bandwidth for Dual Active Clusters. Unless otherwise stated, all runs were made with 256 concurrent jobs (128 jobs per active cluster). Each job writing or reading 800 MiB (300 MiB volumes @ 2.66:1 compression) using 32 KiB block size, QSAM BUFFNO = 20, using eight 4Gb FICON channels from a z10 LPAR (four 4Gb channels per active cluster). Clusters are located at zero or near zero distance to each other in laboratory setup. DCT=125.

Four-way TS7700 Grid with Dual Active Cluster Performance

© Copyright IBM Corporation
Figure 19. Four-way TS7700 Hybrid Grid H5

Four-way TS7700 Hybrid Grid with Dual Active Clusters Bandwidth

Figure 20. Four-way TS7700 Hybrid Grid H5 Maximum Bandwidth for Dual Active Clusters. All runs were made with 256 concurrent jobs (128 jobs per active cluster). Each job writing or reading 800 MiB (300 MiB volumes @ 2.66:1 compression) using 32 KiB block size, QSAM BUFFNO = 20, using eight 4Gb FICON channels from a z10 LPAR (four 4Gb channels per active cluster). Clusters are located at zero or near zero distance to each other in laboratory setup. DCT=125.
Figure 21. Four-way TS7700 Hybrid Grid H6

Figure 22. Four-way TS7700 Hybrid Grid H6 Maximum Bandwidth for Dual Active Clusters. Unless otherwise stated, all runs were made with 256 concurrent jobs (128 jobs per active cluster). Each job writing or reading 800 MiB (300 MiB volumes @ 2.66:1 compression) using 32 KiB block size, QSAM BUFFNO = 20, using eight 4Gb FICON channels from a z10 LPAR (four 4Gb channels per active cluster). Clusters are located at zero or near zero distance to each other in laboratory setup. DCT=125.

* 64 jobs – 128 jobs (64 jobs per active cluster).
Four-way TS7700 Grid with Quadruple Active Cluster Performance

Figure 23. Four-way TS7700 Hybrid Grid H7

Four-way TS7700 Hybrid Grid with Quadruple Active Clusters Bandwidth

Figure 24. Four-way TS7700 Hybrid Grid H7 Maximum Bandwidth for Quadruple Active Clusters. Unless otherwise stated, all runs were made with 256 concurrent jobs per z10 LPAR (each LPAR drives two clusters). Each job writing or reading 800 MiB (300 MiB volumes @ 2.66:1 compression) using 32 KiB block size, QSAM BUFFNO = 20, using eight 4Gb FICON channels from a z10 LPAR (four 4Gb channels per active cluster). Clusters are located at zero or near zero distance to each other in laboratory setup. DCT=125.

* 64 jobs – 256 jobs (64 jobs per active cluster)
Additional Performance Metrics

Deferred Copy Rates (Compressed)

Figures 25 through 29 show deferred copy rates. Data rate over the grid links are compressed. Host I/O with deferred copy mode was driven from host to the active cluster(s). After the copy queue(s) accumulated several terabyte (TB) of data, host I/O was terminated. In the subsequent hours, copies took place from the source cluster to the target cluster. There was no other TS7700 activity during the deferred copy except for premigration if the target cluster was a TS7740.

Peak and Sustained Copy Throughput (Compressed)

When the target cluster is a TS7740, copy and premigration run concurrently after host I/O stops. Initially, when the non-premigrated data in the queue has not reached the premigration threshold, copy activity has higher priority than premigration, and this allows for a higher initial copy rate, called “peak” copy rate. As more data is copied to the target cluster disk cache than is premigrated out to the physical tape, the premigration queue grows until a pre-established threshold is reached. At that time, incoming copy stream is throttled to allow for increased premigration activity. This throttling mechanism operates to achieve a balance between the amount of data coming in from the source cluster and the amount of data being copied to physical tape. The resulting data rate for this mode of behavior is called the “sustained” copy rate, and could theoretically continue on forever, given a constant supply of logical and physical scratch.

The premigration thresholds can be tuned to elongate the peak copy duration via host console request and are the same settings used to prolong host peak write period (refer to “Peak and Sustained Write Throughput” for details).

In R2.1, the default setting for premigration drives per pool for V07 configurations was changed from 6 to 10, resulting in an increased amount of background premigration activity during replication. As a result, the peak copy rates dropped due to the additional system activity introduced by the added drives. When the premigration priority threshold is reached and the V07 enters the sustained period of operation, the additional premigration activity results in a higher sustained replication rate as the target cluster is able to off load the inbound replicated data faster. Since the peak duration is usually short, the benefit of a faster sustained replication rate should provide an overall more attractive recovery point objective.
One-way Deferred Copy Rates

Figure 25. Two-way TS7720 DD Single-direction Copy (source and target clusters have the same number of drawers). Clusters are located at zero or near zero distance to each other in laboratory setup. DCT=125.

Two-way TS7720 R 1.6 - R 2.0 - R 2.1 - R 2.1 PGA2 Copy Performance (Single-direction)

Figure 26. Two-way TS7740 DD Single-direction Copy (source and target clusters have the same number of drawers). Clusters are located at zero or near zero distance to each other in laboratory setup. DCT=125.

TS7700 Single-direction Maximum Copy Rate
Two-way Deferred Copy Rates

Figure 27. Two-way TS7700 Hybrid H1 DD Single-direction Copy. Clusters are located at zero or near zero distance to each other in laboratory setup. DCT=125.

Figure 28. Two-way TS7720 DD Dual-direction Accumulative Copy Rate (source and target clusters have the same number of drawers). Clusters are located at zero or near zero distance to each other in laboratory setup. DCT=125.
Figure 29. TS7740 Two-way TS7740 DD Dual-direction Accumulative Copy Rate (source and target clusters have the same number of drawers). Clusters are located at zero or near zero distance to each other in laboratory setup. DCT=125.
Sustained Rate vs. Number of Premigration Drives

The following chart shows sustained write data rates achieved in the laboratory for a standalone TS7740 with different numbers of TS1130 backend physical tape drives reserved for premigration (the default is 6 drives). The data for the chart was measured with no TS7740 activity other than the sustained writes and premigration (host write balanced with premigration to tape).

Figure 31. Standalone TS7740 sustained write rate vs. the number of TS1130 drives reserved for premigration. All runs were made with 128 concurrent jobs. Each job wrote 800 MB (300MB volumes @ 2.66:1 compression) using 32KB blocks, QSAM BUFNO = 20, and four 4Gb FICON channels from a z10 LPAR.
Performance vs. Data Compression

Incoming data is compressed by the TS7720/TS7740 channel adapters prior to being written to the disk cache. Likewise, outgoing data is decompressed by the channel adapters prior to sending to the host. TS7720/TS7740 data compression can be turned off, but it should be left on except in those exceptional situations where the data has already been compressed prior to sending it to the TS7700. Note that data compression does not reduce host throughput.

The following charts show data rates for compression rates varying from 1:1 (no compression) to 4:1. While data compression greater than 4:1 is possible, it is uncommon. Three types of workloads are depicted in the charts:

♦ all jobs writing
♦ all jobs reading
♦ half the jobs writing / half the jobs reading

Charts showing data rates for these workloads are shown in figures 32 and 33 on the following pages.

Figure 32. TS7720 VEB 3956 CS8/SX7 standalone read hit and write bandwidth vs. data compression. All runs were made with 128 concurrent jobs, each job writing 800 MB (uncompressed) using 32KB blocks, QSAM BUFNO = 20, and four 4Gb FICON channels from a z10 LPAR.

© Copyright IBM Corporation
Figure 33. TS7740 V07 3956 CC8/CX7 standalone read hit and write bandwidth vs. data compression. All runs were made with 128 concurrent jobs, each job writing 800 MB (uncompressed) using 32KB blocks, QSAM BUFNO = 20, and four 4Gb FICON channels from a z10 LPAR.
Data Rates vs. Volume Size

Maximum data rates vary according to volume sizes. The following charts show examples of measured data rates vs. volume size.

**Figure 34.** TS7720 VEB 3956 CS8/SX7 standalone data rate vs. volume size. All runs were made with 128 concurrent jobs using 32KB blocks, compression ratio 2.66 to 1, QSAM BUFNO = 20, and four 4Gb FICON channels from a z10 LPAR.

**Figure 35.** TS7740 V07 3956 CC8/CX7 standalone data rate vs. volume size. All runs were made with 128 concurrent jobs using 32KB blocks, compression ratio 2.66 to 1, QSAM BUFNO = 20, and four 4Gb FICON channels from a z10 LPAR.
Data Rates vs. Number of Concurrent Jobs

Read and write measurements were made on a TS7720 with different number of concurrent tape jobs.

Figure 36. TS7720 VEB 3956 CS8/XS7 standalone data rate vs. the number of concurrent jobs. All runs were made with 800 MB (300MB volumes @ 2.66:1 compression) using 32KB blocks, QSAM BUFNO = 20, and four 4Gb FICON channels from a z10 LPAR.

Figure 37. TS7740 V07 3956 CC8/CX7 standalone data rate vs. the number of concurrent jobs. All runs were made with 800 MB (300MB volumes @ 2.66:1 compression) using 32KB blocks, QSAM BUFNO = 20, and four 4Gb FICON channels from a z10 LPAR.
Job Times

Job time includes the time to mount the volumes, read/write, then unmount the volume. It depends on the type of I/O operation (read hit or write) and the volume size.

Figures 38 through 41 show the read/write job time distribution with different sizes of volume. The volume sizes shown were host view (uncompressed). Compression at the TS7700 was 2.66 to 1.

**Figure 38.** TS7720 VEB 3956 CS8/SX7 read hit job time distribution. All runs were made with 128 concurrent jobs, using 32KB blocks, compression ratio 2.66 to 1, QSAM BUFNO = 20, and four 4Gb FICON channels from a z10 LPAR.

**Figure 39.** TS7740 V07 3956 CC8/CX7 standalone read hit job time distribution. All runs were made with 128 concurrent jobs, using 32KB blocks, compression ratio 2.66 to 1, QSAM BUFNO = 20, and four 4Gb FICON channels from a z10 LPAR.
Figure 40. TS7720 VEB 3956 CS8/SX7 standalone write job time distribution. All runs were made with 128 concurrent jobs, using 32KB blocks, compression ratio 2.66 to 1, QSAM BUFNO = 20, and four 4Gb FICON channels from a z10 LPAR.

Figure 41. TS7740 V07 3956 CC8/CX7 standalone peak (PK) write and sustained (SUST.) write job time distribution. All runs were made with 128 concurrent jobs, using 32KB blocks, compression ratio 2.66 to 1, QSAM BUFNO = 20, and four 4Gb FICON channels from a z10 LPAR.
Performance at Distance

Grid performance has been seen to vary depending on:

1. distance,
2. quality of links,
3. speeds of link (1Gb or 10 Gb),
4. number of utilized tcp/ip links (up to 4 for 1Gb link and up to 2 for 10 Gb link),
5. percentage of link bandwidth available,
6. type of I/O activity (write, read, or copy).

The following setup was used to assess the remote read/remote write/deferred copy performance vs. different inter-cluster distances and packet loss percentages in the laboratory:

- Two TS7720 were direct-connected via an Anue Ethernet Network Emulator.
- Delay times were used to simulate various distances.
- Different percentages of packet loss were applied.
- Total link bandwidth (four 1Gb TCP/IP links) was available for these tests.

Note that the effective link distance between two clusters can be greater than the physical distance.

Remote Read

Figure 42 shows the remote read rate via four 1Gb links over various distances with different packet loss rates.

Figure 42. Single-direction remote read performance at simulated distances and packet loss rates. All runs were made with 128 concurrent read jobs via 4x1Gb links from the remote TS7720. Each job reading 800 MiB (300MiB volumes @ 2.66:1 compression) using 32 KIB block size, QSAM BUFFNO = 20, using four 4Gb FICON channels from a z10 LPAR.
Remote Write

Figure 43 shows the remote write rate via four 1Gb links over various distances with packet loss rates.

Figure 43. Single-direction remote write performance at simulated distances and percents of packet loss. All runs were made with 128 concurrent write jobs via 4x1Gb links to the remote TS7720. Each job writing 800 MiB (300MiB volumes @ 2.66:1 compression) using 32 KiB block size, QSAM BUFFNO = 20, using four 4Gb FICON channels from a z10 LPAR.
Remote Copy

Figure 44 shows the single-direction deferred copy rate between two TS7720 3956 VEB CS8 and six XS7 drawers. The two TS7720 are connected via 4x1Gb links. The number of deferred copy tasks is 40 (default). This provides an example of how latency and packet loss can affect replication rates. Configurations with additional drawers, 2x10Gb links, or copy tasks would provide improved results. Please refer to Figures 45 for the improved deferred copy rates with 128 deferred copy tasks.

In the run, host I/O was driven to one TS7720 (active cluster) in deferred copy mode. After tens of terabytes (TB) of uncopied data were queued at the active cluster, the workload was stopped and the only activity on the active cluster was the copying to the other TS7720 (DR cluster). Different one-way time delay values and packet loss percentages were applied to each direction.

Figure 44. Single-direction copy performance at simulated distances and percents of packet loss with 40 deferred copy tasks (default).
Performance Tunings

This section details two important tunings that can provide significant performance improvement:

1. Deferred copy task tuning -- improves deferred copy.
2. Premigration drive tuning -- improves premigration rate and sustained write rate.

Deferred Copy Tuning

In Release 2.0, a new tuning knob is made available for improving the deferred copy performance, especially if the network is “dirty” and over a great distance. The knob provides a method to change the number of concurrent replication tasks used for deferred copies. The default is 20 or 40 tasks depending on the hardware configuration. The maximum number of deferred copy tasks is 128. See the IBM® Virtualization Engine TS7700 Series z/OS Host Command Line Request User's Guide white paper on how to modify this setting.

Figure 45 shows the single-direction deferred copy rate between two TS7720 via four 1Gb links. These are the same runs as what are shown in Figure 44 except that at the end of host I/O, the number of deferred copy tasks was tuned to 128.

Notes: Having 128 deferred copy tasks during host I/O may impact on host rate. When using large task counts for replication, the deferred copy throttling settings can be tuned to reduce or eliminate any impact this may have on your workload.

Figure 45. Single-direction copy performance at simulated distances and percents of packet loss with 128 deferred copy tasks.
Figure 46 summarizes the deferred copy performance improvement by increasing the deferred copy tasks from 40 to 128.

Figure 46. Single-direction copy performance improvement with 128 copy tasks.

Premigration Drive Tuning

TS7740 premigration rates, i.e. the rates at which cache-resident data is copied to physical tapes, depend on the number of TS1130 drives reserved for premigration. By default, the number of drives reserved for premigration is six per pool in Release 2.0. With the introduction of the Power 7 based V07 server, there are additional resources made available, allowing further improvement of premigration data rates by increasing premigration drive counts. Figures 47, 48, 49, and 50 show how increasing the TS1130 drive counts used for premigration can improve:

- premigration rate (up to 31%).
- Standalone sustained rate (up to 15%).
- Grid host DD sustained rate (up to 21%) using 11 premigration drives
- Grid RR sustained rate (16% for dual active clusters) using 11 premigration drivers.
- Grid RR sustained rate (7% for single active cluster) using 11 premigration drivers.
Standalone TS7740 (V07/CC8/CX7/4 DRs)  
Premigration Performance

Figure 47. Standalone TS7740 premigration rate tuning. All runs were made with 128 concurrent jobs. Each job wrote 800 MB (300MB volumes @ 2.66:1 compression) using 32KB blocks, QSAM BUFNO = 20, and four 4Gb FICON channels from a z10 LPAR.

Sustained Write Performance

Figure 48. Standalone TS7740 sustained rate tuning. All runs were made with 128 concurrent jobs. Each job wrote 800 MB (300MB volumes @ 2.66:1 compression) using 32KB blocks, QSAM BUFNO = 20, and four 4Gb FICON channels from a z10 LPAR.
Figure 49. Two-way TS7740 Single Active Cluster Sustained Rate Tuning. All runs were made with 128 concurrent jobs. Each job wrote 800 MiB (300 MiB volumes @ 2.66:1 compression) using 32 KiB block size, QSAM BUFFNO = 20, using eight 4Gb FICON channels from a z10. Clusters are located at zero or near zero distance to each other in laboratory setup. DCT was set to 125ms.

Figure 50. Two-way TS7740 Dual Active Cluster Sustained Rate Tuning. All runs were made with 256 concurrent jobs (128 jobs per active cluster). Each job wrote 800 MiB (300 MiB volumes @ 2.66:1 compression) using 32 KiB block size, QSAM BUFFNO = 20, using eight 4Gb FICON channels from a z10 LPAR (four 4Gb channels per active cluster). Clusters are located at zero or near zero distance to each other in laboratory setup. DCT was set to 125ms.
Performance Tools

Batch Magic
This tool is available to IBM representatives and Business Partners to analyze SMF data for an existing configuration and workload, and project a suitable TS7700 configuration.

BVIRHIST plus VEHSTATS
BVIRHIST requests historical statistics from a TS7700, and VEHSTATS produces the reports. The TS7700 keeps the last 90 days of statistics. BVIRHIST provides a means for users to save statistics for periods longer than 90 days.

BVIRPIT plus VEPSTATS
BVIRPIT requests point-in-time statistics from a TS7700, and VEPSTATS produces the reports. Point-in-time statistics cover the last 15 seconds of activity and give a snapshot of the current status of drives and volumes.

The above tools are available at one of the following web sites:

Business Partners tape tools: http://testcase-yellow.boulder.ibm.com
Conclusions

The TS7700 Virtualization Engine provides significant performance capacity. Release 1.5 provides for larger cache storage capacity plus enhancements in scratch tape processing. It also introduces the disk-only model with even higher cache capacities. Release 1.6 supports 4-way grid, Hybrid configuration, and cluster families. Release R 1.7 provides larger disk cache capacity and better disk cache performance. Release 2.0 provides new Power 7 server and options to use four 1Gb Ethernet links or two 10Gb links, resulting in dramatic TS7700 performance improvement. The TS7700’s modular architecture will continue to provide a base for product growth in both performance and functionality.
Acknowledgements

The author would like to thank Joseph Swingler, Ralph Beeston, Jim Fisher, and Luis Fernando Lopez Gonzalez for their review comments and insight, and also to Paul Scheuer and Jim Fisher for distributing the paper.

Finally, the author would like to thank Luis Fernando Lopez Gonzalez, Dennis Martinez, James F. Tucker, Harold Koeppel, Gene McGaha, Douglas Clem, Michael Frick, Jim Lewis, Craig Miller, and William Hengen for hardware/MVS/network support.
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 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.