IBM 3494 B18 HPO Virtual Tape Server, A Paper Describing the Relationship Between Attributes of Virtual Volumes and Physical Storage Requirements

Version 1.2
Dated October 18, 1999

Carl Bauske, WSC

This updated paper begins with a long and complex title, which might be said of the contents as well. However. it is my intention that this paper will serve to clear up some of the mystery surrounding the relationship between the managed objects (virtual volumes) and the number of physical volumes and library slots required to store these objects, within the IBM Virtual Tape Server (VTS). For additional information about the 3494 and the VTS, please see the ITSO redbooks : ***IBM Magstar Tape Libraries: A Practical Guide, SG24-4632*** **and** ***Enhanced IBM Magstar Virtual Tape Server: Implementation Guide SG24-2229*****.**

The VTS is designed to be a fully automated storage subsystem and may be allowed to function with a minimum of applied management. Detailed monitoring of this subsystem is optional, and is pursued by some customers and not others. This is especially true of the B18 model, and even more so of future models.

This paper was written to explain some of the observable details of its operation, but by no means am I recommending that the user attempt to micro-manage this complex machine. Some users are content to let the device take care of itself, and indeed, this is why it is designed the way it is. Human curiosity is a wonderful part of all of us and many have an unbending desire to understand what they observe. This paper is for those of us with the time and the need to satisfy our curiosity.

The fact is that there is no set relationship between the number of virtual volumes defined to a VTS and the number of physical volumes required to stack and store them, and this has led to many questions from many VTS operators who are puzzling as to why the inventory of physical volumes is not static and, in many cases, seems unpredictable. This seemingly uncertain behavior may be explained once the user understands the factors that determine the relationship.

Many users would like a simple rule of thumb, but you will not find that here. When I have employed some arithmetic to the problem, I will "show my work".

I begin with the definitions of the terms used in this paper. A consistent lexicon for those who use the Virtual Tape Server is key to succinct and meaningful discussion.

Virtual Volume: A volume inserted into (defined to) and managed and stored by the Virtual Tape Server, on behalf of the owning system(s). I will call these Vvols for the sake of brevity. These are of either 400 or 800 megabyte maximum capacity as defined by the user and may contain compressed or uncompressed data. Partially filled Vvols are only as large as their contents, a characteristic that is unique to Vvols. It is also characteristic of

the 3490E architecture, which the VTS emulates, that all the data on a volume may be either compressed or uncompressed. It is important to understand that these volumes may be of three basic types:
1. Never mounted, typically also a scratch volume,
2. Mounted at least once and returned to scratch,
3. Mounted at least once and currently unexpired (private).

Physical Volume: A 3590 High Performance Tape Cartridge of 10 gigabyte native capacity that is under the control of the VTS. I will call these Pvols.
These volumes may be:
1. Scratch (empty),
2. Filling (one of several volumes that are candidates for storage of additional Vvols),
3. Full (a transitory state),
4. Somewhat less than full (due to expiration or invalidation of Vvols they once contained).
The Pvols described by 2, 3 and 4, are all in the VTS private volume category (FF04). Pvols in a VTS always contain compressed data.

Reconciliation: An internal VTS process that occurs once every 18-24 hours that, among other things, marks Pvols as being eligible for reclaim. This determination is made on the basis of the then-current reclaim threshold (percentage) set by the user in the Library Manager. If a Pvol has invalid and expired space such that the percentage of valid space falls below this threshold, the Pvol is marked for reclaim.

Reclaim: The process of tape-to-tape copy that moves the valid objects (Vvols) from an eligible Pvol to a filling Pvol open for output. The end of this process leaves the eligible Pvol in the scratch category of the VTS, making it a candidate to be open for output again. Until reclaim is complete, a Pvol that was once full and now has some invalid space cannot be used for output. Reclaim may run at any time, controlled by demand, availability of resources, and inhibited by a daily schedule set by the user. Later versions of the VTS may preemptively start reclaim by temporarily raising the reclaim threshold in order to provide scratch volumes, rather than shut down the VTS should the Pvol scratch pool become completely exhausted.

Freespace: The available storage in gigabytes provided by the Pvols in the scratch (empty) category. These Pvols are the only volumes counted in this calculation. It does not include invalid space on once-full volumes that have not yet been reclaimed. In the B16, this number (found in the SMF 94 record) is the number of scratch Pvols times 20 gigabytes. In the B18, this number is calculated using a histogram of accomplished compression ratio and expressed in gigabytes of host data (uncompressed).

Freespace Threshold: A value expressed in gigabytes, which the user sets in the library manager, that determines at what level of freespace the library will signal the host that the VTS is low on physical storage (Pvols).

Library Scratch Threshold: This value is set in the ISMF Library Define Panel in DFSMS and is expressed as a number of volumes (Vvols in a VTS library). The user sets a threshold for each media type in each of the libraries in the installation. For a VTS library, the media types are limited to media 1 (400 megabytes) and media 2 (800 megabytes). When the level of scratch volumes drops below this threshold, the operators are alerted with a CBR message that remains active until volumes are inserted to bring the scratch count to double the threshold.

Premigrate, or Copy: The process of copying a Vvol from Tape Volume Cache (TVC) to a Pvol open for output. This process must complete before a Vvol may be removed from cache to make room for a more recently used Vvol. Officially called Copy, I will use Premigrate in this paper.

SMF Type 94 Record: This is sent to all attached MVS hosts every hour on the hour by the library manager. The record contains many fields which may be examined, analyzed and graphed in order to monitor the VTS for capacity planning purposes. For a description of the fields, please see the above mentioned redbook, *Implementation Guide,* appendix C. To acquire formatting tools, go to ftp://index.storsys.ibm.com and download the Tapetool package.

Zoe: A storage administrator, with a strong dose of curiosity, chosen to install and monitor the Virtual Tape Server in my examples that follow.

For clarity, I will use Zoe's VTS and walk through the life of this machine from its installation as a B16 (first generation) through upgrade to a B18 with the EHPO compression feature, to the present day. As a window into this machine, I will use charts at each snapshot, with the pertinent information displayed.

We start at

June 1997 -

The VTS is installed as a new B16 with six 3590 B1A drives and 72 gigabytes of TVC. Zoe defines (inserts) 15,000 Vvols, resisting the urge to define 50,000, which was the maximum allowed by the B16 at that time. She also defined and inserted 300 Pvols, although she had 800 on hand, labeled and intialized. This is what our table looks like:

Vvols Inserted	15,000
Vvols Never Mounted	15,000
Vvols in Scratch	15,000
Vvols in Private (unexpired)	0
Average Size of Vvols	0
Pvols inserted	300
Pvols in Scratch (freespace volumes)	298
Pvols in use (Private)	2 (mounted automatically)

December 1997 -

Zoe has had the B16 in production for six months and has gradually added selected applications to the workload, taking appropriate care to monitor the device, checking for any indications of overload. Zoe has noticed that there is some reclaim activity in the SMF 94 statistics. Now the library has alerted the operator that the freespace threshold is crossed. Zoe is puzzled by this and thinks: "I have not added Vvols since the VTS was installed. All the applications that I run in the VTS expire the data written within 30 days. Why have I not reached equilibrium? Why is the freespace continuing to shrink?" Here is the data at this time:

Vvols Inserted	**15,000**
Vvols Never Mounted	**2,512**
Vvols in Scratch	**7,155**
Vvols in Private (unexpired)	**7,845**
Average Size of Vvols	**224 MB**
Pvols inserted	**300**
Pvols in Scratch (freespace volumes)	**20**
Pvols in use (private)	**280**

At this point the B16 is still consuming freespace because all of the Vvols originally inserted have not been mounted for the first use yet. When a Vvol is mounted the first time, the Vvol will occupy space on a Pvol for the first time (after premigrate) and will occupy space forever after, whether expired (scratch) or not (private), unless it is ejected or exported from the VTS.

Although the 15,000 Vvols inserted is far less than the allowable, it is still more than Zoe needed. The result of this is storing many scratch Vvols on Pvols. With a retention period of less than 31 days, and 7,845 private volumes, Zoe only needed to insert about 8,500 Vvols, by this point in time. The VTS would have reached equilibrium with only about 190 Pvols being required.

Reasoning:
With 7,845 private volumes, and a 30 day retention, then the average volumes scratched and rewritten per day is 261. I doubled this to provide ample Vvols in the scratch pool so 7,845 plus 522 = 8,367, which I then rounded up to 8,500. 8,500 X 224 = 1,904,000 megabytes. For the B16, assuming 2:1 compression achieved in the drive drawer, and Pvols at 50% average fullness, means 10,000 megabytes per Pvol. Divide 1.904,000 by 10,000 and you get 190 Pvols required.

Zoe could have started with a smaller number of Vvols and inserted them gradually, using the scratch threshold in the library definition (one for each media type) as a trigger to alert the operators that the VTS was running low on scratch Vvols. As we shall see, it didn't really matter because the VTS will continue to grow beyond the 15,000 Vvols anyway.

Here is a formula which may be used to determine the number of Vvols that a user should insert at the beginning of VTS operations. Some of the values in the formula may have to be estimated.

Vv = T + (Ax + Si)N

Where Vv = number of Vvols to be inserted
T = scratch threshold for media 2 set in the library definition in ISMF
Ax = the average expiration period in days for the workload going into the VTS or the number of days that the VTS was designed and sized to operate, whichever is smaller.
Si = the interval in days between scratch processing by the tape management system
N = the number of scratch volumes mounted per day

The number of Vvols in the VTS inventory that have never been mounted is not reported in the SMF 94 record. It is generally possible to use the Tape Management System to determine the number of Vvols in a given volser range, that have never been mounted. As long as there are some, demand for Pvol capacity will climb. Obviously, inserting new Vvols will also increase the demand for Pvols, as these too, are mounted for the first time.

So, Zoe defined and then inserted another 100 Pvols, reasoning that this is a 33% increase, and there are 17% of the Vvols that have never been mounted.

June 1998 -

The VTS again signals that the freespace threshold has been crossed. Zoe investigates and finds this data:

Vvols Inserted	**15,000**
Vvols Never Mounted	**411**
Vvols in Scratch	**6,995**
Vvols in Private (unexpired)	**8,005**
Average Size of Vvols	**260 MB**
Pvols inserted	**400**
Pvols in Scratch (freespace volumes)	**20**
Pvols in use (private)	**380**

By comparing with earlier data, Zoe finds that another 2,101 Vvols were mounted for the first time and the Vvol average size has increased from 224 MB to 260 MB. This is a function of many applications, that the data required by the application grows due to a growing business. Zoe adds another 200 Pvols to the VTS inventory. Please note that there are still some Vvols never mounted. It takes a long to mount all the Vvols within a certain range because some number of Vvols are skipped each time through the scratch pool. They will not have a chance to be mounted again until the next time through, and may even be skipped again. The scratch selection algorithm is described in detail in the redbook mentioned above, *A Practical Guide.*

September, 1998 -

This month was marked by the library signaling that it needed more scratch Vvols. This had not occurred for over a year because of the generous amount inserted at install time. Zoe inserted an additional 2,000 Vvols. She learned that the I/S management had decided on an upgrade to the new B18 model VTS due to attractive pricing and scheduled this for November. Zoe's business was experiencing rapid growth at this time.

However, there remained sufficient freespace and the upgrade date was reached without adding any Pvols.

November 7, 1998 -

Just before the upgrade, Zoe takes a record of the stats:

Vvols Inserted	**17,000**
Vvols Never Mounted	**1,522**
Vvols in Scratch	**3,002**
Vvols in Private (unexpired)	**13,998**
Average Size of Vvols	**304 MB**
Pvols inserted	**600**
Pvols in Scratch (freespace volumes)	**130**
Pvols in use (private)	**470**

The VTS is upgraded to a B18 with compression (EHPO). Initial testing shows no big difference in performance and the Service Director indicates that no compression is occurring. Zoe realizes that the Dataclass construct used by the B16 had compact = no set and that the B18 was honoring this by not compressing the data. While she is fixing this, she checks the global parameter in the DEVSUP member of Parmlib as well.

November 8, 1998 -
The Dataclass is changed and the compression is active, and Zoe is monitoring the statistics coming out of the new B18.

November 22, 1998 -
After two weeks as a B18, she noticed, first, a dramatic improvement in performance and virtual mount time, and second, that the average Vvol size began to fall. Zoe reasoned that the EHPO feature would be compressing the Vvols, so Pvol capacity would be freed up over time. However, freespace began to fall even faster.

Vvols Inserted	**17,000**
Vvols Never Mounted	**614**
Vvols in Scratch	**2,120**
Vvols in Private (unexpired)	**14,880**
Average Size of Vvols	**256 MB**
Pvols inserted	**600**
Pvols in Scratch (freespace volumes)	**40**
Pvols in use (private)	**560**

Then she remembered that compression had been occurring in the B16 also, just at a different point in the stacking process. In the B16, the compression was done in the 3590 B1A drive drawer, while in the B18 it is done in the channel attachment cards (Input/Ouput Processors), as well as in the drive. While compressing data twice may add additional compression, it is not much, and sometimes, the data even expands slightly. Also, application blocksize can affect compression ratio in the B18, because the compressors operate better on large blocksizes than on small, to a greater extent than the compressors in the 3590 B1A drive drawers.

==> Note:
The compression algorithm and chip set used in the ESCON cards are very different from those used in the drive drawer. This compression engine is optimized for maximum throughput, and not maximum compression ratio. For some data patterns the ratio achieved by the cards is better than the drive, but for most it is not. It is also seen that once

the data has been compressed by the cards at a relatively low ratio, the compressibility realized at the drive may also be low, such that the overall ratio is less than it would be if the drive was doing all the compression.

The effect of this is usually negligible, but, in some cases it is more than nominal. The benefit of front-end compression is throughput, since the data handled by the processor and the TVC is reduced by this compression. Any future increase in native cartridge capacity due to drive and cartridge technology advances will serve to obviate this difference.

Zoe also realized that the Vvol size reported in the SMF 94 record was now a compressed volume. This would explain the decrease. In fact, the actual Vvol size as stored on the Pvols had increased. The application owners were taking advantage of improved performance by adding workload to the VTS. One characteristic of this added workload was larger dataset sizes than the workload that had been so carefully scrutinized in the B16. Jobs were completing sooner, allowing more work to get done in a day, so the whole business was benefiting by doing more work in the same amount of time. So Zoe added another 200 Pvols.

December 1998 -

Zoe has been thinking about what is happening in the VTS at this point. There are some Vvols that are uncompressed and some that are compressed. Where are they? Are they all on the same Pvols, or are they getting mixed up together? What is the effect of this? What happens when a B16 Vvol is scratched and then reused?

Zoe makes two conclusions based on this careful thought.

1. Her Vvol scratch pool is growing in terms of total capacity. As B16 Vvols are scratched and re-used, they get larger in terms of the host data written to them, at least for any application that writes full volumes. Instead of being "full" at 800 MB, the Vvols hold 800 X (Compression Ratio). This means that they are occupying more space on the Pvols as well. Therefore, she could allow the number of scratch Vvols to drop over time, since more data may be written to each scratch, reducing the number required to satisfy the demands of a week's processing.

2. The B16 Vvols that were stored on Pvols that have been reclaimed since the conversion make the new Pvols seem more full than they really are, which tends to suppress reclaim. The reason for this is that ADSM calculates the fullness of a Pvol by comparing the maximum amount that was written to the Pvol the last time it was closed for output, (the high water mark), against the total size of all the Vvols which are valid (not scratched and reused) that are still stored on it. When some of these are B16 Vvols (because the Pvol was filling when reclaim was running). and some are B18 Vvols, ADSM may see the percentage of valid data stored on the tape as higher than it really is. This is because the B16 Vvols are much more compressible by the drive drawer than B18 Vvols that have passed through the ESCON card compressors before being "measured" by ADSM.

The effect of this is heightened when a mix of long-term B16 data and short term B18 data is present in the machine. As the percentage of B16 data drops, the effect is diminished, so

this is also a transitory state. It should be pointed out that the best way to approach this is to adjust the threshold such that the same number of Pvols are being reclaimed each week after the conversion as were being reclaimed before the conversion.

So, Zoe decides to measure the compression ratio since November and allow the scratch pool to go down by some amount, based on her understanding of the workload that fills Vvols to EOT. This means adjusting the scratch threshold count in the LIBRARY DEFINITION panel in ISMF.

She also decides NOT to raise the reclaim threshold on her VTS because her B16 Vvols have a short expiration period and will be purged quickly from the VTS. If she had a mix of long term volumes that would remain in the VTS for years, it may have been wise to raise the reclaim threshold until the long-term Vvols became such a small percentage of stored volumes that the effect on reclaim was negligible.

January 1999 -
The SMF 94 data indicated that the B18 was not nearly operating at peak throughput with plenty of horsepower for additional workload. An application that had previously used optical storage for long term archive was moved to the VTS. The amount of data stored by this application was relatively small, but the retention period for the Vvols was virtually infinite. The freespace began to drop faster than Zoe had predicted and the number of available slots in the library was now becoming an issue. She realized that the present rate of consumption would demand additional storage slots and media in six weeks or less. Note that the reported Vvol size has continued to drop, and now reflects all compressed volumes.

Vvols Inserted	**17,000**
Vvols Never Mounted	**204**
Vvols in Scratch	**1,867**
Vvols in Private (unexpired)	**15,133**
Average Size of Vvols	**198 MB**
Pvols inserted	**800**
Pvols in Scratch (freespace volumes)	**81**
Pvols in use (private)	**719**

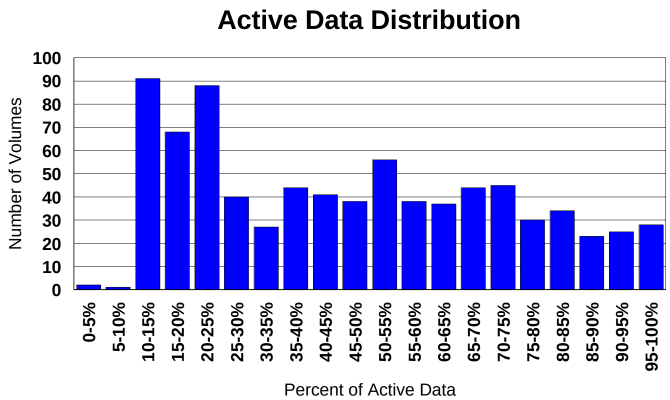
Zoe decided on a two pronged approach to this problem

1. Go to her management with a request for funding to grow the slots and media
2. Try to determine why the freespace was being consumed more quickly than she predicted.

1. Zoe went in to see the I/S manager with cap in hand, afraid that the manager would be upset that this subsystem needed upgrading, and that they should have been able to "make do" with what they had originally designed. However, after reviewing a thorough report on the history of the VTS, the manager said "This is excellent news. This report tells a very good story, and I am appreciative that you have collected and kept this data. You are taking good advantage of this efficient technology. Growing this platform makes perfect sense for our growing business. This device has freed up floor space, reduced operational

workload, improved performance, so of course, further investment is a sound business decision. I will order the additional frames and cartridges today!"

2. Combing through the SMF 94 data did not clearly indicate why the utilization of the inventory of Pvols seemed to be dropping. Zoe reread the operator's guide for the 3494 to see if there was something within her control that could improve the operation. She discovered the Active Data Distribution graph available from the library manger console. A quick glance at this chart told the tale. In this VTS, with the present workload, the percent of "unexpirable" data had grown beyond the reclaim threshold percent set in the machine, which was 10% (the default). The data distribution graph looked like this:



A large number of Pvols were "stuck" just above the reclaim threshold because they contained large proportions of unexpirable Vvols. Zoe adjusted the reclaim threshold to be somewhat higher than the percentage of long-retention data being stored, allowing more thorough reclamation of the Pvols.

Reasoning:
With 719 Volumes in use, assuming a 1:1 compression in the drive, and assuming that the volumes continue to be 50% full on average, 719 X 5,000 MB = 3,595,000 MB. Average volume size is 198 so 3,595,000 divided by 198 = 18,157 Vvols that should be stored, but there are only 17,000 - 204 = 16,796 stored. Therefore, the Pvols are less than 50% full, on the average. The same observation made in the other direction would be: 16,796 Vvols stored at 198 MB each = 3,325,608 MB. Divide this by 719 Pvols and you get 4,625 MB per Pvol. This is 46% full on average.

Today -
Zoe continues to track the growth of workload in the machine. She has certainly seen that there is no simple rule of thumb based on the number of virtual volumes inserted. Instead, this storage is a complex function of many variables. The values of these variables are likely to change as the workload changes. The variables that I have pointed to in this example are:

1. Number of Vvols defined (and mounted at least once)
2. Size of the Vvols stored (reported as uncompressed in the B16 and compressed in the B18 EHPO)
3. Retention period of Vvols
4. Reclaim Threshold (which may be affected by the presence of long-term B16 Vvols)
5. Differences in compression ratio achieved by different hardware compressors.