The Tod Point Group

High-End Virtual Tape:

IBM Leaps Ahead with a New Grid Architecture

Summary

IBM has invented virtual tape again and far for the better.

Introduction

IBM invented virtual tape in 1996 beating StorageTek, to the market by two years. (Note that Sun acquired StorageTek and all subsequent references will be to Sun in this document).

IBM and Sun dominate the high-end virtual tape market, which has previously been mostly a game of performance leapfrog.

However, IBM recently introduced a dramatically improved architecture and products, while Sun continues to plod along with an old architecture.

Points to Remember

- IBM invented virtual tape.
- IBM's new virtual tape grid architecture lets virtual tape data reside anywhere on a virtual tape grid.
- New gigabit IP links are easier to use and cost less.
- IBM's new policy management features make it easy to meet recovery objectives.
- IBM's new hardware outperforms Sun's in most cases.
- IBM has fulfilled on its goal to facilitate migration from its older architecture products.
- Sun has lagged behind IBM in replication and FICON support
- IBM is much more open than Sun.

Architectures

IBM

Named the IBM TS7700 Virtualization Engine, this new system is a significant change from the previous architecture. IBM calls this new architecture its “Virtual Tape Grid” and, while it continues to be based on outboard subsystems, the way nodes operate and are managed is totally changed. At its heart, this grid is a move away from a monolithic approach to a more distributed one. Unlike IBM’s old peer-to-peer architecture, the individual TS7700 nodes are managed as an integrated whole. Moreover, each node adds capacity to the local grid and sports new gigabit IP links for inter-node traffic.

Another important architectural difference is that IBM now virtualizes the location and number of virtual volumes. This permits access and recovery from any node and was a key design goal. The idea is to completely divorce the user from having to know where tape data actually resides and how many copies there are. Instead, users can specify recovery policies that will ultimately dictate these. As we discuss below, this added virtualization dimension is not only a big step in supporting business continuity with advanced policy management; it is also a necessary step in content-based access for tape data.

User interest in the new gigabit IP links has been strong as they look for simpler and less expensive ways to replicate tape data. IBM led the market with synchronous replication and has always had asynchronous replication.

Although IBM’s tape grid currently supports three nodes, it has been designed to support at least eight nodes as well as to support future enhancements in cache capacity, performance, content-based access and data de-duplication. It also leverages IBM’s vast array of technologies for performance, encryption and futures.

Playing catch-up, IBM has finally added the capability to export a copy of the logical volumes stored to the node for disaster recovery purposes. Also, IBM’s tape grid is not compatible with its previous architecture, but a wealth of migration tools are available.

Sun

Sun continues to employ a hybrid approach where most of the virtual tape logic sits on mainframe-based software while tape I/O and replication services are handled by a relatively dumb external hardware subsystem based on Sun’s SVA array which is no longer being actively marketed.

In addition, availability features are limited to two-node clusters with no more than four nodes total (one local cluster and one remote cluster) often referred to as a “quadplex”. Worse yet, only one set of channels must handle all I/O (host, tape drives, and remote communication) making it difficult to balance changing workloads. In addition, without gigabit IP links, expensive channel extenders or routers must be used to reach remote sites.

A key missing feature is synchronous replication. When and if Sun delivers this feature, we expect its use must be carefully managed and doubt the VSM subsystem has enough power to handle many cases.

Unlike IBM, a tape volume's location is not virtualized. Although VSM can manage a limited number of copies of a tape volume, it is up to the user to specify and track.

System Comparison

See Table 1 at the end of this report for a comparison of features. For comparison purposes, we have used the top-end systems. Below we discuss what we see as the key differences:

IBM 7700

As discussed earlier, IBM's Tape Grid architecture is by far the biggest differentiator and we will see more of its advantages as IBM rolls out future feature and function. The switch to IP links for inter-nodal traffic has been heartily embraced by users who find it much easier and less expensive than using FICON or ESCON channels, particularly for wide-area applications.

IBM's specifications and design reveal more processing power, more processors and more channels resulting in a 20+% faster throughput than VSM5.

Sun VSM5:

Sun has struggled for years with timely FICON support and has not yet made 4-Gbps FICON publicly available. Also, VSM5 lacks sufficient channels especially in clustered configurations.

Performance

Leapfrog is the best way to characterize vendor-specified performance. Sun's VSM hardware is based on a disk subsystem design that goes back to its original 1992 Iceberg product. Though revolutionary in design, that product and its many successors consistently failed to deliver competitive performance. As a consequence, Sun either sold on price or would deliver two systems in place of one -- true for both the disk and virtual tape versions. Indeed, Sun is no longer actively marketing this disk subsystem. In addition, customers often complained that when improvements were made to the disk subsystem version, those improvements took too long to show up in VSM.

Nonetheless, late last year, Sun finally released software and hardware upgrades that bring VSM5's specifications up to and in some cases exceeding IBM's. Not exactly a huge leap, but worth noting provided it proves out in real world environments.

However, real-world experiences rarely match paper specifications. Workloads, block sizes, compression ratios, read/write ratios and architectural differences greatly impact the performance experience. IBM's TS7700 has a traditional front-end, controller, back-end design where hosts/servers talk to the front end and real tape drives talk to the back end. VSM does not use this design. Instead, there is only one set of channels that must be divided and dedicated to hosts, real tape drives, local cluster links and remote links. This can result in both under and over provisioning. Under provisioning hurts performance and over provisioning wastes money. Moreover, when workloads vary over time, users cannot dynamically shift these dedicated resources.

Performance in multi-node configurations is interesting. In IBM's case, three nodes can behave like one big node offering three times the number of virtual tape drives and roughly three times the performance. Of course, Sun also scales when more nodes are added, but each node is treated and managed as an individual node.

Data De-Duplication

Data de-duplication has emerged as one of the hottest technologies in the market today and virtual tape is a prime opportunity. However, neither IBM nor Sun offers data de-duplication for their mainframe virtual tape solutions. Sun recently announced a deal with Diligent Technologies, but this is just a reseller agreement and not an OEM agreement. Thus, as things stand today, Sun cannot integrate Diligent's software into its VSM.

IBM, on the other hand, has had various forms of data reduction running for years in its Tivoli Storage Manager (TSM) and although TSM formed the software basis for the previous IBM virtual tape servers, the TS7700 has purpose built firmware. Moreover, we expect IBM to introduce native data de-duplication for TSM within six months or less. We believe IBM is developing its own data de-duplication core technologies and is expected to deliver this function in offerings across its portfolio including, of course, the TS7700.

Green Storage

If we compare a maximally configured virtual tape system, IBM's TS7700 uses roughly 16% less energy than Sun's VSM5. What's more, when comparing a total solution including virtual tape server, tape drives and tape robots, then IBM has half of the energy costs of Sun.

Content Management and Tape File Systems

A most subtle jewel lies at the center of IBM's new Tape Grid architecture and it begins with virtualizing the location of a virtual volume.

Today, we could say a real-tape file system comprises the tape catalog which relates tape file names to a particular real-tape volume, a real-tape management system which tracks the location of real tapes and an I/O subsystem to read and write tape data. All of this software runs on the mainframe and is not very integrated. Moreover, such a file system never knows that a file has been directed to a virtual volume in a virtual tape subsystem and is actually living on disk or on a different real-tape volume than the catalog indicates.

Enter content management where users want to index the contents of existing tape files. It makes little sense to index existing tape data on the mainframe where often a huge number of tapes would have to be located, brought to the mainframe(s), mounted, read and indexed. Instead, why not distribute that function to an outboard platform such as a virtual tape server where the work can be spread over multiple servers that, in IBM's case, know the real physical location of every tape volume. If the indexing function is added to the TS7700, for example, tapes do not have to be moved. A tape at a remote location can be indexed offline by a TS7700 in the remote location, but with the nature if IBM's tape grid, the derived metadata can be shared by all the nodes on the grid. What's more, the shared knowledge of a volumes location could be used to direct the distribution of the indexing workload(s) thereby minimizing tape movement. And that is the hidden architectural jewel in IBM's tape grid.

We believe IBM will add tape data indexing to its content management offerings and by virtualizing the location of a virtual volume IBM has taken a small, but fundamental step towards a tape content file system. Moreover, IBM has immense and diverse resources it can apply to this issue and is well positioned to become the leader.

Sun is of course not standing still in content management, but we have not seen any indications of a well-developed competitive strategy for content management of tapes.

Openness and Accuracy

Overall, IBM is much more open than Sun about all its storage products, directions and strategies. IBM engages in a robust customer advocate program that provides input to the development process. Comprehensive documentation is available to all and announcements are public and well promoted. On the other hand, Sun's approach is best characterized as stealth marketing. Most documentation is not available to the public, product improvements are not usually publicly announced, and Sun's virtual tape strategy and directions are arcane at best – particularly the future of VSM underlying hardware platform which no longer benefits from improvements to the disk array version. As such we have only relied on information that is publicly available. We also note that Sun's web site and collateral materials often contain errors regarding VSM.

Bottom Line

IBM ranks highly on a vision and execution basis. It has listened well to the market and responded with an up-to-date solution. IBM is well positioned today and for the future. Sun continues to play catch-up and needs to become more open about its products and strategy. Both vendors need to add data de-duplication capabilities.

Nick Allen, Founder
March, 2008

Regarding the information in this report:
The Tod Point Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Tod Point Group cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.
This report was developed by The Tod Point Group with IBM and funding. This report may utilize information including publicly available data, provided by various companies and sources, including IBM. The opinions are those of the report's author, and do not necessarily represent IBM's position on these issues

Table 1. Virtual Tape Subsystem Comparison

Feature	IBM TS7740	Sun VSM 5	IBM Score[1]	Sun Score[1]
Maximum Number of Virtual Tape Drives per Single Node/Multi-Node	**256/768**	256	1	
Maximum Number of Nodes	**3 (arch. for 8)**	2		
Maximum Peak Throughput per Node[2]	**800 MB/s**	650 MB/s	1	
Maximum Sustained Write Throughput per Node[3]	550 MB/s	613 MB/s		**1**
Data Compression Ratio – System z	4:1	4:1		
Maximum Disk Cache Raw Capacity	6 TB	**7 TB**		1
Maximum Effective Disk Cache Capacity (4:1 Compression)	24 TB	**28 TB**		1
General Availability	2H06	2H06		
System z Attachment	**1, 2, and 4 Gbps FICON (auto negotiate)**	2 Gbps FICON	1	
FICON Concurrent I/O Support per port	**32**	16	1	
Open Systems Attachment	None	None		
Maximum Number of FICON Channels per node	4 (4Gbps)	16	See text on Perfor-mance	See text on Perfor-mance
Maximum Number of FC links to real tape drives per node	**2 - 4Gbps paths to all real tape drives**	None – channels must be dedicated to host, tape drive and cluster links	1	
Gigabit TCP/IP links	**2**	0	1	
Maximum Number of Real Tape Drives per node	16	**32**		1
Maximum Number of Virtual Tape Volumes	1,000,000	**Not limited**		1
Synchronous Replication	**Yes**	None publicly announced	1	
Asynchronous Replication	Yes	Yes		
On-demand Disk Cache upgrades	**Yes**	No	1	
On-demand Performance upgrades	**Yes**	No	1	
Maximum Power Consumption/dissipation	**3.2 KVA; 11.0 KBTU/hour**	3.8 KVA;12.4 KBTU/hour	1	
		Score Totals	**10**	5

Notes:
1 - This score is a simple binary -- which vendor has better specifications
2 - Depends on lots of things including compression ratio, read/write ratio, block size, etc.
3 - Depends on features and updates installed

Sources: IBM, Sun, and Google