



Spectra Logic BlueScale11 Delivers Data Integrity Verification to SME, Mid-Range and Enterprise Users

Analyst: David Reine

Management Summary

We all have expectations. When we walk into a room and turn on the light switch, we *expect* that electricity is going to go to the light bulb and illuminate the room. When we turn on the TV set, we *expect* the screen to light up (as long as the remote has fresh batteries). When we get up in the morning and get in the shower, we *expect* to be awash in hot water. If any of these events do not occur, we are not only disappointed, we are surprised. However, any of these failures are quickly evident, as the room stays dark, the television remains silent, and we are left to shiver in a cold shower. Moreover, we have a right to be angry; utilities, such as power and water, have attained a position in our lives where we feel that we are entitled to them.

The same can be said for our IT expectations in the data center. As long as we do not overload our circuits with an excessive demand, the processing and storage components in the infrastructure are *expected* to be there *in working order!* When any of these components do fail, systems crash and the enterprise is at risk for the loss of business, and therefore, revenue (or, maybe, reduced service delivery). Fortunately, these failures are quickly evident, as all processing comes to a grinding halt. Unfortunately, in the data center we also run the risk of another type of failure, a silent failure, one where we have a failure that we do not know has occurred. We are especially susceptible to this type of failure when we are writing data out to disk or tape. The data may appear to be written successfully, but in fact is not. This is especially crucial when we are doing a backup or archiving data to tape. As data stores continue to grow, more and more data has to be protected to comply with industry standards and government regulations. In order to reduce the TCO of data protection, many companies are expanding their library capabilities, or, in some cases, returning to tape from a Disk-to-Disk (D2D) environment. Every data center has great concern for the reliability of the recovered data, in the event of an on-line data loss, and for the retrieval of archived information, five, ten, or even thirty years after the archive has been created.

One company that is paying close attention to the reliability of backups and archived data is Spectra Logic. With their *BlueScale* management software, Spectra Logic has invested heavily in terms of personnel and resources to ensure that the data that you read back, *whenever* you read it back, is *exactly* the same as it was on the day that you wrote it. Spectra Logic has invested in innovative practices to provide data integrity assurance through the latest release of BlueScale, *BlueScale11.3*, to deliver *Data Integrity Verification* to the data center. BlueScale11.3 will now pre-scan the data, validate it and then re-verify it whenever migration to new media is required. These enterprise features are now available to the small-to-medium enterprise (SME), as well as the largest enterprises. To learn more about how Spectra Logic continues to innovate, and is investing to protect important data, please read on.

IN THIS ISSUE

➤ The Enterprise Data Center.....	2
➤ Data Protection Strategies	3
➤ Spectra Logics' Platform Solutions	3
➤ Conclusion	4

The Enterprise Data Center

The Enterprise data center is faced with a serious dilemma. How does it preserve mission-critical data at the lowest possible cost while still maintaining the integrity and security of that data? With primary storage requirements doubling every 12-to-18 months, this challenge is becoming more difficult with each passing year. Improvements in technology are being made regularly to facilitate the long-term preservation of data, but not necessarily reducing the total cost of ownership (TCO)¹ on a per PB basis.

What are the components that make up the TCO of a long-term storage system? Clearly the first, and most obvious cost encountered, is the acquisition cost (CapEx). Acquiring *and* deploying a long-term storage system in a mission-critical data center is significant. However, the operational costs cannot be overlooked, especially if you have opted for cloud storage. The data center must plan and budget for:

- Annual maintenance charges (after warranty);
- Administrative costs to manage that data, usually on a 7x24x365 basis;
- The amount of energy required to drive the hardware and cool the data center, and the electrical costs associated with it;
- The floor space necessary to house the data, in an era of ever-expanding storage requirements; and, most significantly,
- The cost of the media required to hold that data.

However, while the CIO of every enterprise has to search out a solution which produces the lowest cost per TB that delivers an acceptable Quality of Service (QoS), including accessing cloud storage, it cannot be acquired in a vacuum. That CIO must still ensure the availability of the platform and the integrity of the data for as long as the enterprise may need it. This requires the highest reliability characteristics for the data stored and the components needed to preserve and retrieve it.

The data center staff is responsible for both the backup and archiving of this most valuable asset – the enterprise's data. They need to weigh the backup and recovery times in order to provide continuous operation, completing the backup in a timely fashion and meeting corporate recovery

time objectives (RTO), in order to satisfy data center SLAs and keep the business on the air and engaged.

In addition to backup and recovery, an enterprise's long-term storage platforms must also support the archival of very large volumes of information, perhaps forever. Quite clearly, the retrieval of data from an archive does not face the same stringent time requirements as does the recovery of current data from a backup. Data is placed in the archive because there is less likelihood that it will be needed quickly, but still needs to be preserved, no matter what the likelihood is for retrieval. For example, if a banking client needs to verify his balance, the recovery of that data has more urgency than the need to look at an image of a check more than a year old. Even if that takes a couple of minutes, it's acceptable, especially for a service being provided typically without an additional fee. However, that does imply that the data center has the necessary applications to find your specific data quickly. The storage platform requires an integrated file system to facilitate the retrieval of both structured and unstructured data, in a timely fashion, perhaps forever. In addition, the metadata for unstructured data must be preserved and managed to facilitate any search.

The length of time that it takes to search for your data and retrieve it is less urgent than an instance of data recovery; however, a fast streaming capability is a requirement. **The length of time it takes to retrieve the data is irrelevant, however, if the data that has been preserved cannot be read, or if the data that is read is invalid or incorrect. The data center needs a long-term data protection system that can ensure the integrity and security of the data.** This includes features such as WORM² and encryption to ensure that no one has intentionally altered the data or gained improper access to it. Any enterprise data center needs a storage system that can validate that what is written can be retrieved *exactly*, now and in the future. Furthermore, that validation needs to be integrated into the data protection process so that the validation is done automatically.

It would also be nice if the CIO could find a system that will protect not only the data, but also the investment that is being made in it today. The data center staff needs to find a solution with the lowest cost/PB that has historical credibility and a long-term roadmap to continue to deliver and

¹ See the issue of *Clipper Notes* dated December 20, 2010, entitled *In Search of the Long-Term Archiving Solution – Tape Delivers Significant TCO Advantages over Disk*, and available at <http://www.clipper.com/research/TCG2010054.pdf>.

² Write Once, Read Many.

Exhibit 1 — Device Reliability

Media/Device	Hard Error Rate in Bits	Equivalent in Bytes	PB Equivalent
Consumer Disk - SATA	1.00E+15	1.25E+14	0.89
Enterprise Disk - SATA	1.00E+16	1.25E+15	8.88
Enterprise Disk - SAS/FC	1.00E+17	1.25E+16	88.82
LTO Tape	1.00E+18	1.25E+17	888.18
Enterprise Tape	1.00E+20	1.25E+19	88,817.84

Source: Spectra Logic

enhance data preservation, so that the solution will not have to be replaced in the foreseeable future. It also is mandatory to find a solution that will preserve the environment as well, i.e., one with the high density and power efficiency required to promote a green environment.

Data Protection Strategies

There is a significant amount of debate today on what is the best strategy to preserve your data. Some argue that a traditional disk-to-tape (D2T) is satisfactory. Clearly, it has the lowest cost, but, unfortunately, it does not provide the fastest recovery time in case of a data failure. In order to achieve the near instantaneous recovery of data, some pundits recommend a purely D2D environment. While this will provide you with the prerequisite millisecond response, it does so at a significant cost, not only in CapEx, but in OpEx as well. We have seen that energy costs, floor space costs, and administrative costs could overrun your budget. Furthermore, how reliable is a D2D environment when you have to deploy your secondary storage as a RAID array, because it is not a question of *if* the high-density disk drives will fail, but *when* and *how many*. One study by CERN³ indicates that only 10% of all disk errors are single bit errors, while another 10% are 512 byte sector errors. The remaining 80% of disk errors are in 64KB regions. CERN also ran the verify command on 492 RAID systems each week for four weeks. They discovered 300 read/write errors covering 2.4PB of data. When you consider that the rebuild time for 1TB and 2TB disk drives can be prohibitively long, this strategy could compromise your RTO and SLA requirements. Spectra Logic has also published the hard error rate that they have reported, with both open

systems LTO-5⁴ devices and enterprise tape drives from IBM, the *TS1130*⁵, exceeding the ratings for SATA and SAS/FC disks (See Exhibit 1, above).

Recently, we have seen an increase in the number of data centers that have migrated to a disk-to-disk-to-tape (D2D2T) scenario, achieving the best of both worlds. The transmission speed of disk (to comply with the SLAs established for the backup window) and the short-term recovery of data, and the documented lower cost of tape (to enable the data center to preserve the multi-PBs of information of an archive system within the always-being-squeezed IT budget).

All of these data preservation strategies address the security, TCO and RTO of the archiving of data but they do not specifically address the data validation issue. Neither disk vendors nor tape vendors have stepped up to the plate, until now. Spectra Logic has introduced the newest member of their tape hardware family, the *Spectra Library Server Gen 3*, or *Spectra LS*. (See Exhibit 2, on next page, for Spectra LS features.) Combined with the latest revision to Spectra's *Blue Scale11* management software, Spectra Logic is now delivering data integrity verification for their entire family of *T-Series* tape libraries.

Spectra Logic's Platform Solutions

Less than one year ago, Spectra Logic upgraded its T-Series Tape Libraries with LTO-5 drives⁶. This upgrade enabled the data center

³ See "Data integrity" (Draft 1.3) by Bernd Panzer-Steindel (April 8, 2007) and available at <http://indico.cern.ch/getFile.py/access?contribId=3&sessionId=0&resId=1&materialId=paper&confId=13797>.

⁴ See [The Clipper Group Navigator](#) dated January 29, 2010, entitled *LTO Program Announces Next Gen Tape – LTO-5 Raises the Bar for Tier-3 Storage*, available at <http://www.clipper.com/research/TCG2010002.pdf>.

⁵ While this bulletin is aimed at the SME space, Spectra also has qualified the IBM TS1130 Enterprise Drive in the Spectra T-Finity Library for deployment in 2Q11. See [The Clipper Group Navigator](#) dated July 16, 2008, entitled *Lowering the TCO of the Data Center – IBM Innovates Tape Architecture ... Again*, available at <http://www.clipper.com/research/TCG2008037.pdf>.

⁶ See [The Clipper Group Navigator](#) dated June 24, 2010, entitled *Spectra Logic Upgrades T-Series with LTO-5 –*

Exhibit 2 — Spectra LS Gen 3 Features

- Single pass, fast verification, for high duty cycle environments;
- Faster BlueScale application speed and processing;
- Quicker response times on BlueScale front panel and remote management interfaces;
- Supports hyper-threading;
- Increased capacity to track a larger number of cartridges;
- Faster Inventory times;
- More memory capacity for processing and larger code base; and
- Increased log storage for long-term reference.

Source: Spectra Logic

with improved capacity and performance, along with multi-partitions in support of the Active Archive Alliance. Now, Spectra Logic is upgrading the entire T-Series line with the latest version of Spectra LS with *BlueScale11.3 Data Integrity Verification*, designed to meet the challenges of rapidly growing digital archiving, long-term data retention, and cloud storage of both structure and unstructured data. With the rapid growth in data volumes, along with the extended timeframes for their retention, the most significant of these challenges is answering the question: *Will I be able to retrieve my data as it was written?* Data Integrity Verification has been designed to answer that question, and to answer it in the affirmative, at no extra charge, to T-Series Library clients.

Starting at a point where a data center with a Spectra LS and BlueScale11.3 installed can ensure that the media is certified, i.e., pre-labeled and Media Lifecycle Management (MLM)-ready, there are three levels of data integrity assurance: *import, validate, and re-verify*:

- *Import phase* – performs a *PreScan* on the tape before writing data to ensure that data is written to healthy media. The verification software checks for a broken leader, a mismatched encryption key, write protection, and Media Lifecycle Management tape health. The Spectra Library Server will not allow data to be written to a tape that fails a PreScan.

- *Validate phase* – performs a *QuickScan* after writing the data to verify that the written data can be read. This is a single pass, fast verification for high-duty cycle environments. It supplements, but does not replace the *PostScan* process. This phase will notify the administrator immediately after data has been written to tape, but is found to be unreadable.
- *Re-Verify phase* – performs a *PostScan* to check the archive and notify the data center, when a migration to media may be required. It performs a complete checksum verification of all data on the tape, notifying the administrator when a migration to new media is required.

With data verification now available on all Spectra libraries, the CIO of every data center, regardless of size, can now take advantage, at no additional charge, of Spectra's Data Integrity Verification. From the *T50e* to *T-Finity*, Spectra Logic provides the data center with confidence in their data. BlueScale 11.3 with Data Integrity Verification features will be available for all new T-Series shipments, as well as field upgrades, by March 30, 2011.

Conclusion

With an ever-expanding data store to be preserved for longer periods of time, the urgency to ensure the accuracy of the data on tape is great. It is clear that any successful data protection strategy will include both disk and tape in a D2D2T architecture. It is necessary to use them in conjunction with each other in order to ensure the lowest cost/PB along with meeting the enterprise SLAs for recovery time. Deploying a solution with a data verification capability is also mandatory in view of the fact that you may not discover a write error for years, when it is way too late to correct.

With their T-Series hardware and BlueScale11.3 software with Data Integrity Verification, Spectra Logic is providing the data center with all of the tools that are required by the expanding SME to protect and secure their data, and the price is definitely right! If your enterprise is looking for a data protection system that will support your data center for decades to come, you need to look to Spectra Logic for the solution that might just be what you seek.



Provides Ideal Engine for Active Archive, available at <http://www.clipper.com/research/TCG2010030.pdf>.

About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

- ***The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.***

About the Author

David Reine is a Senior Contributing Analyst for The Clipper Group. Mr. Reine specializes in enterprise servers, storage, and software, strategic business solutions, and trends in open systems architectures. In 2002, he joined The Clipper Group after three decades in server and storage product marketing and program management for Groupe Bull, Zenith Data Systems, and Honeywell Information Systems. Mr. Reine earned a Bachelor of Arts degree from Tufts University, and an MBA from Northeastern University.

- ***Reach David Reine via e-mail at dave.reine@clipper.com or at 781-235-0085 Ext. 123. (Please dial “123” when you hear the automated attendant.)***

Regarding Trademarks and Service Marks

The Clipper Group Navigator, The Clipper Group Explorer, The Clipper Group Observer, The Clipper Group Captain's Log, The Clipper Group Voyager, Clipper Notes, and “*clipper.com*” are trademarks of The Clipper Group, Inc., and the clipper ship drawings, “*Navigating Information Technology Horizons*”, and “*teraproductivity*” are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosures

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

After publication of a bulletin on *clipper.com*, The Clipper Group offers all vendors and users the opportunity to license its publications for a fee, since linking to Clipper's web pages, posting of Clipper documents on other's websites, and printing of hard-copy reprints is not allowed without payment of related fee(s). Less than half of our publications are licensed in this way. In addition, analysts regularly receive briefings from many vendors. Occasionally, Clipper analysts' travel and/or lodging expenses and/or conference fees have been subsidized by a vendor, in order to participate in briefings. The Clipper Group does not charge any professional fees to participate in these information-gathering events. In addition, some vendors sometime provide binders, USB drives containing presentations, and other conference-related paraphernalia to Clipper's analysts.

Regarding the Information in this Issue

The Clipper 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 Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.