



## Ensuring Enduring Access to Big Data

Addison Snell

November 2009

*White paper sponsored by Spectra Logic*

### EXECUTIVE SUMMARY

With data demands increasing exponentially across multiple industries, tape library systems continue to play an important role in long-term data management, within and beyond high performance computing (HPC) markets. T-Finity from Spectra Logic continues to push the barriers of scalability in data archival, with 30 years of expertise in tiered storage systems. To achieve its highest levels of scalability, T-Finity incorporates a wide range of performance and reliability features, including:

- Density improvements reaching 2.5 petabytes per frame, or 72 TB per square foot;
- Upgradability throughout the T-Series product line, using the same software, in 10-slot increments;
- A choice of high-performance drives for high throughput;
- Enterprise data center features including integrated data encryption, media health monitoring, redundant failover architecture, and storage virtualization.

With industry consolidation casting other vendors' roadmaps into question, Spectra is backing T-Finity with 30 years of commitment and expanding OEM agreements. For big-data users who need to ensure enduring access to vast amounts of information, Spectra has the opportunity to provide a compelling solution with T-Finity.

### MARKET DYNAMICS

*Q: What do you give to someone who has everything?*

*A: Someplace to put it all.*

#### Ever-Increasing Data Demands

As more types of data become digitized, and the fidelity of that data increases, requirements for long-term access to massive amounts of digital information continue to increase across many computing sectors. The high performance computing industry has always had more than its share of big-data challenges, and some vertical sectors are undergoing transitions that are bringing about changes at a new order of magnitude.

Within media and entertainment, for example, there is widespread adoption of high-definition (HD) standards for television, and animated movies are being produced in digital 3D. Within the past two years, the oil and gas industry has undergone a similar transition, capturing wide-azimuth (WAZ) data that allows them to see subterranean features with much higher fidelity. And the medical service and research facilities have seen three overlapping trends that all have contributed to a data crush – the digitization of medical scans (e.g., moving from films to digital capture of data); legislation (HIPAA in the U.S.



and similar regulations abroad) that requires medical data to be kept accessible and secure; and increasing patient survivability rates and lengthening life spans, which are of course not a “problem” as such but nevertheless increase the amount of time medical data must be stored.

These examples are merely illustrations of a widespread trend that touches many industries and IT sectors, the ever-increasing need to store and access vast amounts of data. Similar trends exist in manufacturing, finance, national defense, and e-commerce. Storage is the fastest growing product spending category in traditional HPC markets, forecasted to grow by one-third from its 2008 levels to reach approximately \$4 billion in 2013, despite recessionary contraction in 2009.<sup>1</sup>

### Enduring Access to Data

One of the challenges organizations face in the onslaught of so much data is uncertainty as to which pieces of data will be important. To be sure, the majority of the data stored is in a file format colloquially known as WORN – “write once, read never.” So why save it at all? Here we have the data-equivalent corollary to Murphy’s Law: The file you delete is the one you’ll go looking for later.

As an example, consider the footage of an unspectacular, mid-season NASCAR race. It is broadcast in HD, and the next day a few seconds may wind up on sports highlights shows, but the rest of that footage is not likely to be needed again. But what would happen if two years later, a lesser-known rookie driver from that race became suddenly notorious for some reason – perhaps dying in an accident, committing a crime, or winning the lottery. There is now demand for specific pieces of that footage, and it would look foolish not to have it.

Oil and gas companies also save significant amounts of data representing their existing reservoirs. Oil extraction is not a perfectly efficient business, and these companies have significant motivation to continually reevaluate their assets in light of new technologies that may help them extract more resources from current fields. Banks, insurance companies, and even national security agencies all store historical data not knowing which pieces will be vital to future intelligence.

### The Ongoing Importance of Tape

With such a demand to hold so much data for so long, tape storage is an enduring part of many organizations’ overall storage strategies. InterSect360 Research studies have shown that HPC users within big-data vertical markets commonly use tape archival as part of their environments.<sup>2</sup>

The reasons for the persistence of tape boil down to economics. Users must contend with two budgets: a capital budget (i.e., how much money can be spent) and a facilities budget (i.e., how much equipment can be installed and supported in the data center). Tape can offer significant advantages over disk in both cases, storing more bytes per dollar and per floor tile. Mixed tape and disk strategies have gone by multiple names over the last 10 years – first *hierarchical storage management*, then *information lifecycle management*, and now *tiered storage* – but the ongoing need has not changed, except perhaps to the extent that it has intensified.

The continuing onslaught of data means that tape will continue to be part of high-volume storage environments for some time. Furthermore, tape-using organizations want to maintain access to the data they already have. Scalability and continuity are both significant features in tape deployments.

<sup>1</sup> InterSect360 Research market advisory service, “Traditional HPC Total Market Forecast: 2009 to 2013,” September 2009.

<sup>2</sup> InterSect360 Research market advisory service, 2009 HPC Site Census data.

## T-FINITY FROM SPECTRA LOGIC

T-Finity is the newest line of tape library products from Spectra Logic, in its 30<sup>th</sup> anniversary year. T-Finity brings not only a new generation of scalability and density improvements, but also a raft of reliability and performance features that are designed to address the needs of a wide range of big-data customers. With T-Finity, Spectra continues to fill a critical role in the data management ecosystem for HPC and related markets.

### Company History

Spectra was founded in 1979 and brought its first tape subsystems to market in the early 1990s. Since that time, the company has continued to focus on data archival and protection, with solutions scaling from entry-level systems to those that compete at the top of the scalability continuum. Spectra has consistently emphasized storage density as an enabler for larger tape libraries; in recent years it has also introduced reliability and availability features to extend the mindset of enterprise datacenter resiliency to archival systems. In 2007 Spectra introduced the nTier product family, disk-based backup and archive servers, to complement its tape library products.

This year Spectra is leveraging its 30<sup>th</sup> anniversary with new products designed to expand its opportunity in high-end HPC and enterprise markets. The new T-Finity product continues Spectra's tradition of scalability, and the company is marketing a multitude of T-Finity features in an attempt to displace competition throughout the tape market.

### T-Finity Features

The first thing to notice about the new T-Finity is its scalability. T-Finity is a full tape complex that scales to 100 frames with eight robots, storing hundreds of petabytes of data. For facilities that can handle the length, T-Finity can place up to 25 frames in a single line with no pass-through, meaning all robots have equal access to the tapes without pass-throughs to slow down access.

What is less evident about T-Finity's scalability is that it is truly scalable within the T-Series line; that is, it is straightforward for a user to move up the upgrade path from product to product. Spectra's TranScale architecture spans the entire line of T-Series rack-mounted, single-frame, and multi-frame tape libraries, from the entry-level, 4U T50e up through the robust T950 to the new T-Finity. Within the same architecture, users can upgrade from any product in the T-Series line to any other, in increments of 10 tape slots.

According to Spectra, this benefit comes from its design of an architecture that is "built in, not bolted on." The application servers are incorporated into the T-Series product line design, improving overall product density and lowering power and cooling costs versus solutions that separate them. Furthermore Spectra focuses on the high



**Spectra Logic T-Finity**

Source: Spectra Logic

end of its scalability range as its software design point, and all products from entry-level to ultrascale run the same BlueScale software. This software provides a single management interface up and down the product line.

One key enabler to T-Finity's scalability is density. A single T-Finity frame can store over 2.5 petabytes of data, or about 72 TB of data per square foot of floor space. This level of space efficiency not only allows scalability to hundreds of petabytes of data, but also can be an effective value proposition in replacing older tape libraries that are consuming valuable datacenter space.

Another factor empowering T-Finity's scalability is performance. Without a relatively high degree of data throughput, ultrascale environments can become impractical to manage. T-Finity employs dual robotic libraries for high tape exchange rates (multi-library complexes have two robots per library), as well as a choice of LTO 3, LTO 4, or LTO 5 drives. High-throughput, LTO 5 drives provide a total 24-drive, single-frame throughput of 24.2 TB/hour, for over 120 TB/hour in a 25-frame library. For enterprise users who want a proprietary option, Spectra also supports the IBM TS1130 drive in the T-Finity product.

In order to improve manageability and reliability at scale, T-Finity includes a number of enterprise class features that also have resonance in HPC:

- *Integrated data encryption:* T-Finity incorporates straightforward, single-platform data encryption as a standard feature. This is relevant in many HPC markets, such as national security, finance, and medical research, in which data security is paramount, even in relation to different departments within an organization.
- *Media health monitoring:* System monitoring automatically detects tape media that may be in danger of failing, allowing users to proactively replace aging tapes. This feature helps ensure enduring access to historical data.
- *High availability architecture:* Redundant control and communication paths provide failover capabilities supporting uninterrupted data access.
- *Storage virtualization:* Single physical libraries can be virtualized into as many as 32 logical partitions, creating dynamic storage space for resource-sharing groups.

Although many of these features are primarily associated with enterprise markets, there are HPC application areas that will take advantage of them as well. Furthermore these data management and resiliency features can become essential as the scale of a user's data challenges grows.

## INTERSECT360 RESEARCH ANALYSIS

For users with extreme high-volume data challenges – and there are many within HPC – tape solutions will continue to be an important part of the storage environment, and ongoing improvements and enhancements in tape technologies will be necessary in order to keep up with accelerating big-data requirements.

In these cases, the user has a significant motivation to ensure continuous access to both new and historical data. Spectra's 30-year history serves the company well in this regard, especially in comparison to murkiness in the Sun StorageTek roadmap. StorageTek, perhaps the most significant competitor to Spectra T-Finity solutions, was acquired by Sun in 2005, and now Sun is facing its own acquisition by Oracle, which has not issued any statements regarding the future roadmap of the tape libraries.



Furthermore Spectra has been successful in attracting key OEMs, such as Cray, an enduring force in the supercomputing market; SGI, an iconic brand in both HPC and graphics; and both Avid and Thomson Grass Valley, focused players in technology solutions for media and entertainment. These OEM relationships could be useful in expanding Spectra's presence and in bolstering its push to displace aging StorageTek products.

In pursuing this goal, Spectra will do well to keep in mind the multitude of budgets involved. Certainly there are the capital costs of any acquisition to be considered, but there are also operating expenses and facilities costs, such as personnel, power and cooling, and floor space. Floor space can be a significant concern even when it is not an explicitly budgeted cost, when data centers simply run out of room. In these cases, reducing the floor space of the tape archive can have the added benefit of allowing the purchase of additional computational resources.

One challenge for Spectra will be to keep its message focused. Because T-Finity offers differentiation along many dimensions, there could be a tendency for a presentation to become scattered in its delivery. Feature improvements in scalability, density, performance, manageability, and availability should all be wrapped together in support of a common goal that is aligned with the end user's interest. Spectra has an opportunity to present its solutions as a dominant choice for delivering enduring access to critical data, delivering some peace of mind to the data center manager who's looking for a place to put everything.



InterSect360  
RESEARCH