Multi-tier storage saves energy, space

BY MATTHEW BRISSE

Combining energy-efficient tape, disk, and data de-duplication technology is the answer to many storage-related energy consumption concerns.

Power, cooling, and space limitations are key issues faced by IT managers. Demand-side contributors include decreased hardware acquisition costs, sharp increases in server density, and an exponential rise in the volume of data being stored and managed.

Many companies are growing their data centers at an exponential rate. The demand for increases in real estate and floor space is at a premium. Many IT organizations are faced with consolidating resources through virtualization and optimization of equipment and processes or be faced with up to a 200% increase in infrastructure cost to build or expand into a new facility.

On the supply side, electric utilities already struggle to keep up with demand. Brown-out restrictions can be commonplace during peak consumption periods, and many IT directors are discovering they cannot increase the amount of power they source from the local power grids. IT analysts believe that by the end of the decade, the world's data centers will have run out of power.

The EPA has predicted that data-center power usage will double over the next five years. In fact, a paper by Stanford University professor Dr. Jonathan G. Koomey stated that 1.2% of all power purchased in the US was consumed in the operation and cooling of data-center equipment. With that figure rising, the need to focus on "green" computing is clear.

Multiple tiers of storage

One outcome of increasing resource pressure is a trend toward multi-tier storage architectures that use both tape and disk technology. Primary storage volumes are decreasing as companies begin to segregate their data according to business requirements for data access and retention versus operational constraints on power, space, and cooling.

The logic of this is clear: Tape cartridges in a tape library consume power at lower rates than disk systems, and tape cartridges stored in a vault consume the least of all, as well as providing the lowest aggregate cost per gigabyte. So, for data that must be retained for quarters and years and do not need as much fast recovery, a tape-based, tertiary storage tier conserves energy and space resources. This class of storage is typically used for data retention beyond three to five years.

The other factor at work in the evolution toward multi-tier

architectures is data de-duplication, which dramatically reduces disk requirements through the elimination of redundant data. The SNIA Data Management Forum explains data de-duplication as "the process of examining a data set or I/O stream at the sub-file level and storing and/or sending only unique data. The definition of 'what is a duplicate' is predicated upon the method used to evaluate, identify, track, and avoid duplication. The de-duplication process includes updating tracking information, data that is new and unique, and disregarding any data that is a duplicate."

Disk solutions with data de-duplication strategically combined with tape storage enable IT managers to effectively manage data growth, data protection, and energy usage by creating significant operational efficiencies. Through the use of data de-duplication, savings in space, power, and cooling can be significant.

Optimizing tiers with de-dupe

Data de-duplication reduces the amount of disk required to protect a given amount of primary data by detecting and eliminating redundant blocks within files and, in some cases, between different files and file types. Data de-duplication enables users to exploit disk performance while dramatically reducing capital and operating expenses, including power, cooling, and space.

Some implementations of data de-duplication also include enhanced disaster-recovery protection by increasing the replica frequency to improve the recovery-point objective (RPO). Using the same technology that identifies duplicate segments within data sets, some data de-duplication systems can also reduce the bandwidth needed to transmit backup sets over a network. Once systems are synchronized, whole backup sets can be replicated while only changed blocks are actually moved. For example, if a new backup is only 5% different from a previous one at a block level, bandwidth needed for transmission can be reduced by up to 95%.

Conventional disk systems without data de-duplication make sense only for a subset of data where recovery-time objectives (RTOs) override other considerations such as performance, long-term retention, and total cost. Data sets with normal access requirements can use disk with data de-duplication technology, both as a first site for backup data and for medium-term retention. Data de-duplication technology

---

Matthew Brisse is a business development director for the Quantum Corp. and is also a member of the board of directors for the Storage Networking Industry Association (SNIA) Data Management Forum.

reduces space, power, and cooling requirements enough to make disk economical as a retention medium for weeks or months, at which point tape becomes an ideal medium.

Easing the pressure

Finding power-aware solutions that keep data accessible and protected will likely be an IT priority well into the future. But a tiered architecture that combines the efficiencies of tape with de-duplicated disk storage is an approach businesses can employ today to ease operational pressures without compromising service levels. These techniques will provide much-needed breathing room as additional green solutions continue to develop.

Finding power-aware solutions that keep data accessible and protected will likely be an IT priority well into the future.

Storage vendors are on board, offering the technologies and capabilities required to bring all these elements together. The SNIA Data Duplication and Space Reduction (DDSR) Special Interest Group (SIG) in collaboration with the SNIA Green Storage Initiative (GSI) is promoting common metrics, features, and functionalities that could help the storage industry reduce the storage footprint through data de-duplication.

Composed of IT professionals, integrators, and storage vendors, the SNIA DDSR SIG focuses on advancing space reduction in all network storage technologies. By defining and promoting efficient network storage solutions and common implementations, the DDSR SIG is enabling sustainable data storage operations that reduce both storage cost and the environmental impact of data-center infrastructures.

You can find out more about the SNIA DDSR SIG and other SNIA forums by visiting www.snia.com. ◉