STORAGE SWITZERLAND

USING BACKUP VIRTUALIZATION TO MAXIMIZE POWER EFFICIENCY

Storage Switzerland, LLC

George Crump, Senior Analyst

Improving power efficiency is of top importance to data center managers today, but so is the goal of improving the reliability and performance of the backup process. Initiatives like disk to disk backup help performance and reliability but they do so at the cost of significant power consumption. Solutions like deduplication and MAID (massive array of idle disks) bring power efficiency to disk backup but do so at the cost of performance. Data center managers are stuck between having to choose fast backups or power efficiency. By using backup virtualization performance can be maintained while maximizing power efficiency.

The challenge with traditional disk to disk solutions

Disk to disk backup has the potential to greatly improve performance, especially on slow networks that can't provide tape devices with adequate sustainable data streams. The problem with traditional disk to disk solutions is that they're not power efficient. Storing primary data a second time, actually multiple times over, leads to backup disk capacities that can be 5x to 10x of what the primary storage is. Plus, all of this disk needs to be racked, powered and cooled. Compared to the tape systems that it was originally designed to replace or at least augment, disk backup can be a massive waste of power.

The challenges with power efficient solutions
There are two basic approaches to optimizing storage and tackling the power consumption problem of disk to disk backup. The first is to have the system store less data in the same space through technologies like deduplication and compression. The second is to have the drives, and in some cases the entire system, power down while not in use. But the ultimate solution is to combine deduplication with power down for maximum efficiency.

In the space optimization category these solutions have two challenges that may impact performance. The first is the processing power to analyze inbound data as it's stored on the system. While compression calculations are relatively straight forward, deduplication requires a series of calculations and look-ups against those calculations, which can result in degraded backup performance. While some of the performance impact can be overcome through the use of higher-end processors, faster processors also consume more power and cooling. Additionally, because of all of these lookup functions, deduplication systems make terrible candidates for spin down technology since in most cases all the drives need to be active when a backup is occurring. Finally, there is often a recovery performance problem since the data needs to be 'undeduplicated' through the same series of lookups.

In the space optimization category these solutions have two challenges that may impact performance. The first is the processing power to analyze inbound data as it's stored on the system. While compression calculations are relatively straight forward, deduplication requires a series of calculations and look-ups against those calculations, which can result in degraded backup performance. While some of the performance impact can be overcome through the use of higher-end processors, faster processors also consume more power and cooling. Additionally, because of all of these lookup functions, deduplication systems make terrible candidates for spin down technology since in most cases all the drives need to be active when a backup is occurring. Finally, there is often a recovery performance problem since the data needs to be 'undeduplicated' through the same series of lookups.

MAID or Spin Down technology can come closer to the power consumption rates of tape systems by actually powering down drives, and in some cases entire systems, when not in use. As backups age entire sections of disk can be in a powered-off state for significant periods of time. However, for the reason mentioned above, MAID or Spin Down drive technologies cannot benefit from deduplication, so more capacity needs to be purchased. Even deduplication systems with MAID capabilities only offer a fraction of the power efficiency that a pure MAID system does because the lookup functions triggered by deduplication require that a significant percentage of drives be powered on.

On the recovery side, MAID systems also have a slight performance problem because Spin Down drives need to be 'spun up' before they can be read. Many IT managers are surprised to find out that when at maximum power efficiency it takes almost as long to power up a series of MAID drives as to mount a tape into a current generation tape drive and scan to a recovery position. Modern tape systems have become very efficient at mount and positioning times.

Of course deduplication and power efficient storage systems still provide significant value to the enterprise. They can make the backup process more forgiving of slow clients and slow networks and make recovery, especially of small groups of files, easier and faster than tape based systems. Ideally though, data center managers would like to balance the positives of disk based backup with the power efficiencies of legacy tape systems.

How Backup Virtualization maximizes power efficient backup

Backup virtualization may be the ideal solution for data centers looking to strike this balance as it abstracts the backup devices from the backup applications. In these systems, the backup applications communicate with the backup virtualization hypervisor which then negotiates with the specific devices. In addition to helping manage power efficient or deduplicated disk systems it also provides a common storage area for heterogeneous backup applications.

The number one goal of most backup upgrades is to improve performance and lower recovery windows without breaking the budget on additional hardware. In a backup virtualization environment a small, high speed disk cache is utilized so that inbound backup streams can be accepted as fast as the environment can transfer them. This cache only needs to be sized large enough to be a buffer as data is sent to the various backup storage devices. But it can also be made large enough to store entire backup jobs and provide a disk-based initial recovery point, if desired.

Once this data is on the high speed cache it can be directed to a variety of tape devices. For example if there is deduplication in the environment, data can be sent to those for efficient, highly compacted storage. In this scenario, however, performance of the backup and recovery is not as impacted since those processes will typically use the high speed cache. There is also less pressure on the deduplication device to perform at high throughput levels, plus, a lower power, lower cost CPU can be used.

If a MAID system is used, the caching function provided by backup virtualization allows data to be sent in one operation, not sporadically throughout the backup process. This enables the MAID system to get back to a restful, power efficient mode sooner. Also, the high speed cache means that it is less likely that the MAID system will need to be 'awakened' during a recovery.

MAID and deduplication systems can even be combined. The backups can be sent to the deduplication system from high speed cache every night, which can act as a medium term repository. Then, on a weekly basis, those backups can be sent to the MAID system. This controls the growth of the deduplicated systems and keeps the MAID system in a restful state for an extended period of time.

For systems with MAID and deduplication, backup virtualization can be the ultimate enabler. These systems could be bought in phases, with a deduplicated MAID system implemented first and used to store a quarter or two worth of backup data. Then, a new MAID system could be brought in and data from the next few quarters be written to it, allowing the older system to enter full, power-down mode. Adding a new device every half year or so would be a major undertaking with traditional backup applications writing directly to disk devices. Backup virtualization requires no change to the software, and all backups would be re-directed to the new device automatically by the backup virtualization appliance.

Backup Virtualization re-enables tape for the ultimate in power efficiency

Backup virtualization's high speed cache and ability to move data between backup devices without involving the backup application re-enables tape as a viable target for long term data storage. In this scenario data can be backed up to the high speed cache for maximum backup and recovery performance. Then it can be migrated to deduplicated or MAID-enabled disk for intermediate storage of backup data and finally moved to tape for long term storage. Tape, when used in this manner, is still undeniably the most power- and space-efficient form of data retention available to the enterprise.

Space and power optimized systems have made disk backup more feasible for many data centers. However, using these systems has also meant living with a performance compromise. Backup virtualization removes the problem allowing for high performance, no compromise backups, while still achieving excellent levels of power efficiency.

About Storage Switzerland

Storage Switzerland is an analyst firm focused on the virtualization and storage marketplaces. For more information please visit our web site: http://www.storage-switzerland.com