The need for capacity optimisation

Examining storage efficiency in the data centre.

By Larry Freeman, SNIA Data Protection and Capacity Optimization (DPCO) committee, NetApp.

In the past decade, system vendors have responded to the ever-growing demand for applications by producing application servers with more processing power and smaller footprints. However, these servers, while providing far greater computational ability than their predecessors, began consuming alarming amounts of data centre costs, space, and energy. A worldwide phenomenon, “server sprawl” was highlighted in a study presented to the United States Congress in 2007 by the Environmental Protection Agency.

In that study, the EPA categorized data centre power consumption into 6 segments, as shown below:

The results from this study made it abundantly clear that servers were a large area of concern in data centre operations, consuming 40% of all data centre power – and 80% of overall IT equipment load. In response to this dilemma, IT organizations began turning to server virtualization as a way to maximize server efficiency and reduce the operational cost of housing vast numbers of servers in the data centre.

Server virtualization provided an immediate payback by allowing the removal of large numbers of physical servers, and replacing them with Virtual Machines, or Guest OS’s, running on a smaller population of servers. This, however, has lead to a new phenomenon, “VM sprawl”.

In conjunction with the growth of virtual servers, the thirst for larger and faster data storage systems became equally demanding. In some respects, the illusive quest to improve data centre efficiency was simply moving from one category to another. Servers were shrinking, but storage was becoming bloated. Increased use of virtual servers and the steady creep of overall data growth were bringing the need for capacity optimization to new heights.

Capacity Optimization According To SNIA

The Storage Networking Industry Association (SNIA) was well aware of the trend taking place, and the pain being felt by Users. As a result, several groups formed to address the problem of growing storage, and to provide guidance and consistency to Users struggling with data growth. Specifically, three cross-vendor groups were formed:

Data Deduplication and Space Reduction (DDSR) Special Interest Group
Educated the user community on the nuances of data deduplication, data compression, and other forms of space reduction.

Data Protection and Capacity Optimization (DPCO) Committee
The latest incarnation of the DDSR, merging data protection into the capacity optimization discussion

Green Storage Initiative Technical Work Group (GSI TWG)
Focused on defining metrics and providing tools to quantify capacity optimization hardware and software technologies These three groups also worked with the SNIA Technical Council to provide standardized terms and definitions, including:

Raw capacity
The sum total amount of addressable capacity of the storage devices in a storage system.

Effective capacity
The amount of data stored on a storage system, plus the amount of unused formatted capacity in that system.

Storage efficiency
The ratio of a storage system's effective capacity to its raw capacity.

These terms (and many others) are published in the 2010 SNIA dictionary and form a basis of consistency that allows Users to make uniform comparisons of the efficiencies and optimization provided by storage vendors.

Capacity Optimization Methods

Referring back to the SNIA dictionary, capacity optimization is defined as follows:

Capacity optimization methods
Methods which reduce the consumption of space required to store a data set, such as compression, data deduplication, thin provisioning, and delta snapshots. RAID 5 and 6 may also be considered as capacity optimizing methods, as they use less space than ordinary mirroring to perform a necessary function: protecting data from storage device failure. Lets take a closer look at the five capacity optimization methods described in the above definition:

Data compression
Data compression algorithms reduce stored data by identifying numerical patterns within data streams and replacing those patterns with smaller data objects which can then be uncompressed by reversing the algorithm.

Data deduplication
Unlike data compression, data deduplication does not "shrink" but rather removes redundant data and replaces this data with a marker that references an identical, stored data object.

Thin provisioning
Thin provisioning has a two-fold effect on capacity optimization.

Data Center Power Consumption By Category

Equipment Type	% of Total
High End Server	2
Midrange Server	4
Volume Server	34
Storage	5
Network Equipment	5
Site Infrustructure	50
Total Power Requirement	**100**

1) Thin provisioned storage systems in effect provide "just-in-time" storage capacity. When a data container (usually a filesystem volume or a block-based LUN) is provisioned, the storage system grants the capacity but does not pre-allocate the capacity. Only once data is written by the application does the container consume any capacity.

2) Interestingly, another effect of thin provisioning is the ability to oversubscribe the storage system. For instance, a 10 Terabyte storage system could present itself as a 100 Terabyte storage system by oversubscribing all of its data containers. Although this is not an endorsed capacity optimization method (the physical capacity limit of the system will prevent any container from growing very large), this practice has been used by storage administrators to appease application developers who insist on having large data containers but typically only use a small portion of the container.

Delta snapshots
As the name implies, snapshots are point-in-time images of data containers. Typically, a snapshot is accomplished through the creation of a 2nd set of markers (i.e. pointers) that 'freeze' the container view as a snapshot image. An interesting application of delta snapshots are writeable snapshots, sometimes call virtual clones. With virtual clones, all common data is shared by the parent and the clone. The end result is that data copies can be made using very little capacity.

RAID
To reduce the risk associated with drive failure, the traditional response has been disk mirroring, often referred to as RAID 1, RAID 1+0, or RAID 10. RAID mirroring does indeed provide a higher level of protection, but carries with it a high capacity penalty. In response to this, single and dual parity RAID levels (often referred to as RAID 5 or RAID 6) offer adequate protection with a much lower capacity penalty.

Summary

The cost, space and energy consumption of data storage systems continues to be a significant portion of total expenditures made by data centre managers. Finding ways to increase storage optimization is of critical importance to IT organizations, and has even become a significant public policy issue. SNIA has been a leader in providing a vehicle for vendor collaboration and to help Users manage and monitor storage systems for optimal space and power efficiency.

SNIA's DPCO and GSI TWG are dedicated to addressing these growing business and societal concerns by continually educating User on methods of 1) increasing capacity optimization and 2) reducing data storage power consumption.