SILICON
MECHANICS
SERVERS • STORAGE • HPC

KEEPING PACE WITH
EVOLVING STORAGE
CHALLENGES

Across organizations of all sizes, two factors are placing new demands on storage. Data volumes are growing and the time to convert that data into actionable information is getting shorter. Whether it's a small business working with terabytes of customer data, or a research lab with petabytes of experimental data, the value comes from searching, interpreting, visualizing and analyzing ever-growing quantities of data.

Unfortunately, given the amounts of data routinely used in organizations, many computational workloads are impacted by performance and latency problems. Traditional approaches to address such issues resort to expensive over-provisioning of resources, greatly increasing compute and memory capacity, and adding large amounts of storage. However, in many cases even these steps do not optimize computational workflows and often the added equipment strains budgets by increasing operational costs.

To accommodate and effectively utilize the growing volumes of data, organizations need optimized storage solutions that leverage leading-edge technologies. The solutions must easily scale in capacity and performance, integrate built-in data management and protection, and offer a variety of price and performance options to meet an organization's strategic and budgetary requirements. Additionally, the storage solutions must be backed by a technology partner that brings the expertise needed to help select, deploy and optimize the right systems without overspending.

More Data, More Challenges

While much attention has always been focused on boosting compute performance to keep pace with growing demands for faster results, addressing storage performance is increasingly becoming just as critical.

Working with today's datasets requires storage solutions that can easily scale, while often times the traditional approach has been to use systems that scale-up, allowing the addition of more drives and arrays. While this approach can accommodate the growing volumes of data if the solution is architected properly, performance can be enhanced by spreading the data across many drives or arrays. However, as data volumes grow and performance requirements become more stringent, organizations must look for solutions that also scale-out. In such architectures, scale-out storage can harness the extra storage added across arrays while adding performance.

Simply put, market conditions are creating a new set of storage challenges. Staying with the status quo is not an option. Organizations need solutions that can support today's explosive data growth and match the variable input/output (I/O) data transfer rates of their workloads. To meet these requirements necessitates a solution that integrates a mix of storage technologies and devices, each offering different price/performance characteristics to match the goals of the organization.

A closer look at the common storage issues organizations are dealing with helps illustrate why different solutions are in demand today:

Data Growth

Virtually every field is making use of new and more refined data. Examples abound:

- Electric utilities are moving from analog meters that were read once a month to smart meters that provide usage and other information every second. They need storage solutions that allow for fast analysis of streaming data.

- Healthcare and life sciences organizations have moved from paper medical records and diagnostic equipment that produced film, to electronic health records and X-ray, MRI, and CAT scans consisting of high resolution images that can have files hundreds of GBs per image. These organizations need inexpensive, high-capacity storage.

- Retail and eCommerce organizations now have data from many more touch points (in-store, web, and mobile) with customers and from multiple channels of communications and interactions (POS terminals, email, call centers, websites, and social media). They need storage solutions that can handle the variety of data and make it easily accessible to processing servers for fast analysis.

- Educational, government and research organizations are running visualization, analysis, and modeling routines against large datasets in a variety of applications. They need highly-scalable and high-performance storage to support these efforts.

In all cases, it is important to match various types of data to an appropriate storage tier and media based on performance, availability and reliability requirements. This matching is needed to ensure the IT infrastructure can quickly move the large volumes of data required for computations between storage devices and servers.

Performance Scalability

During the honeymoon phase a new storage solution might initially seem capable of meeting the growing data volume demands, but fall short when actually trying to keep CPU utilization high. To help accelerate workflows, a storage solution must scale in performance, supporting the higher I/O and throughput required to get faster results. For example, an analytics routine running on a high-performance server would be able to finish a run faster if ingestion of the large dataset involved is sped up with a higher performance connection to the data.

Data Variability

A storage solution must be capable of running workloads that include both large and small data files, as well as different types of data. If all data were in the same format – a structured database, for example – or of approximately the same file size, a solution could be implemented and optimized to handle the similar data. Working with the mixed data used today requires a storage solution that optimizes workflow performance for each data type.

Workload Variability

If an organization runs just one application on a single set of data, it can pick a storage solution optimized for that particular combination. However, that scenario is not realistic. Most organizations use a variety of applications and different data types in the course of normal operations. Some applications may require that a single large data file be read, others may make continuous searches and queries of a large database, and still others might need to work with many small files. For the applications to run without delays, data needs to be moved onto and off of the storage systems at a rate that matches the needs of the applications. As such, a storage solution needs to be capable of handling variable workloads. To match the variability, a storage solution must offer an easy to manage way to migrate data from the different storage tiers, and the ability to intelligently stage data on the highest performance drives when needed to sustain or accelerate a workflow.

Storage Solutions For Today's Challenges

In some cases a single solution can meet the needs of an entire organization. More often it's a combination of solutions, based on specific workloads or use cases that are deployed to handle storage needs.

Big Data analytics: In the past, almost all the data organizations had to deal with was structured data. Today, many organizations are undertaking massive parallel analysis of big unstructured data which can include email, documents and social media content. This requires highly-scalable storage that supports tiering of data to match the performance of the analytics routines.

Virtualization: Running multiple virtual machines (VMs) on a single physical server places great demands on a storage solution. Each VM may need simultaneous access to the same storage device and each VM will have different I/O requirements. A suitable storage solution will need to provide the I/O and throughput to improve scalability and allow an organization to improve VM density.

Transaction processing/Database applications: Online Transaction Processing (OLTP) and DB applications require fast access to small blocks of data. Flash storage can deliver consistent, low latency, high bandwidth performance to accelerate the performance of these applications.

High-Performance Computing: Parallel distributed file systems such as Lustre and General Parallel File System (GPFS) offer a high-performance clustered file system to boost performance in HPC environments. Parallel file systems handle concurrent reads and writes of data; some solutions offer additional features that help in the event of a node or drive loss. Unfortunately, configuring, deploying and managing Lustre, GPFS, and other parallel distributed file systems frequently require special skills that are lacking in most organizations. The systems have mainly been used in national labs, but as the technology is needed in commercial, educational, research and government settings, organizations are looking for help from their storage provider.

Private cloud/Hybrid cloud: Organizations are increasingly adopting private clouds as a way to more efficiently utilize IT resources, automate processes, and provide end users with services in a timely manner. Open source solutions based on OpenStack for object storage, or commercially supported alternatives like Microsoft Azure, can provide software-defined infrastructures and convergence on-premises and when connecting to data in HPC clouds, enabling organizations to run and scale up applications, as well as store and provide access to data.

Acceleration using flash: Some applications and workloads simply need storage that offers a performance boost above what can be delivered by spinning disks. Selectively using flash drives can help deliver the requisite performance. Flash storage typically offers I/O benefits that can improve application performance by an order of magnitude over disk.

Partnering With the Right Systems Integrator

Determining the right storage solution to use in these scenarios requires expertise about new storage technologies and how to make them work together efficiently.

Silicon Mechanics helps organizations take the headache out of deploying storage solutions to meet the demands of the business. Using its **Expert included.** approach, Silicon Mechanics collaborates with customers to architect, build and deploy a customized system, followed by ongoing support and warranty services.

Knowing that every organization has unique needs and requirements, Silicon Mechanics offers an extensive portfolio of flexible and customizable storage solutions.

NAS/SAN Storage

Products include open solutions with different price/performance characteristics to meet specific business requirements. Silicon Mechanics offers highly scalable solutions that enable organizations to scale their storage to ensure critical business applications can work with large datasets and run fast. Additionally, Silicon Mechanics offers high-density, flexible and powerful storage solutions to backup, archive and protect important data.

Parallel File Systems

Handling the massive growth of digital data and the HPC workloads that make use of that data requires high-performance storage. Silicon Mechanics offers storage solutions that scale-out to meet the most demanding performance requirements.

Object-Based Storage

In many organizations today, there is a need for highly-scalable storage that allows shared access to, and analysis of, data while offering easier data management of large datasets. Silicon Mechanics offers object-based storage to meet these and other requirements.

As a sole-source provider for deployments of all sizes, Silicon Mechanics will ensure your customized, cost-effective storage solution is right-sized for your storage needs.

For more information about how to select the right storage solutions for your computing efforts, visit http://www.siliconmechanics.com/storage.