SGI NAS
and VMware®: A True Virtualization Solution

Advantages of Using SGI NAS within a VMware Environment

Overview

When looking at virtual architecture, there is one thing in particular that stands out—it is the software that makes the cloud infrastructure flexible, fault tolerant, and scalable. The virtualized application is software, the hypervisor is software, and so is the guest operating system.

Using virtualization software results in the application itself becoming independent of the hardware. When there is a need for more computing power, scaling the virtual environment becomes a matter of just adding more standardized hardware.

As resources (hardware) and virtual machines (software) are added, the environment scales out as needed. But equally important, hardware refresh and migrations become a breeze.

SGI NAS makes it easy to virtualize standardized hardware into an enterprise-class, unified storage solution. All enterprise features, such as thin provisioning, storage snapshots, cloning, and replication, are built in and implemented without any additional license or costs.

In contrast, with a traditional storage appliance model, customers are typically required to do a controller upgrade if performance, or the number of spindles, reach a certain level.

The functionality of the SGI NAS solution allows you to scale your storage needs as you would scale your virtual environment (by adding more hardware). Also, storage migration techniques are built-in, thereby allowing for simple replacement of existing drives with larger and newer drives online and on the fly.

Moreover, legacy storage model limits dictate the overall cloud design. Such systems limit the number of snapshots and the maximum volume size, which needlessly complicates data management.

As data sets grow over time, the chance of silent data corruption increases. Traditional RAID systems do not detect silent data corruption. SGI NAS provides unique end-to-end data integrity features that can detect and repair silent data corruption.

SGI NAS provides next generation software that matches the benefits of server and desktop virtualization for primary storage, using hardware that breaks away from legacy design issues while guaranteeing data integrity.

Key Highlights

- **Hybrid Storage Pools:**
  Leverage SSDs plus disks to provide unmatched price / performance
- **Deduplication:**
  60% of enterprises view primary storage deduplication as the most important disk- related initiative
- **Increased efficiency** through:
  - Compression
  - Unlimited Snapshots
  - Unlimited Clones
- **VAAI:**
  vStorage API for Array Integration allows for certain tasks to be offloaded and performed at storage controller level

Hybrid Storage Pools

Hybrid Storage Pools (HSP) are the heart of the system and dictate overall performance. Primarily, initial capacity and the expected capacity growth, as well as IOPS and the read/write ratio between them, are the most dominant factors.

Block allocation algorithms of the Hybrid Storage Pool transform small random writes into a single, more efficient sequential write operation. Virtual machines invoke a large number of random read / writes, and the HSP provides intelligent methods of handling these unpredictable workloads. With SGI NAS, these fundamental features dramatically accelerate performance.

The Hybrid Storage Pool uses RAM, Solid State Devices (SSDs), and rotational drives all at once. As such, instead of just adding spindles, the design of the pool provides more degrees of freedom. Within the HSP, three different device types can be created: one virtual device type for cache (read operations); a log device (write operations); and a regular virtual device (capacity).

A virtual device is a collection of drives that has a redundancy scheme built in. With SGI NAS, different protection levels are available, protecting up to three concurrent device failures within one single virtual device. By adding virtual devices, capacity and performance are able to scale.

By using low latency cache and log devices, which typically are RAM- or FLASH-based drives, the entire pool performance is accelerated to the level of SSDs in terms of latency.

When using mirrors for redundancy, it is recommended to mirror the drives over two distinct JBOD chassis to maximize redundancy. For the virtual cache device, MLC, SSD, and SAS devices are used. And for the log device that serves the synchronous writes (with VMware, all NFS mounts are synchronous), a SAS RAM device is used. Also, it is highly recommended to mirror the log devices for redundancy reasons.

To fill the cache devices, a certain level of baseline performance is required. It is recommended to have at least 7 or 8 top-level virtual devices to start with. As performance requirements grow, RAM, SSD, and high-capacity drives can be added to match the required, yet growing, capacity and performance needed over time.

Although JBODs can be stacked perfectly, it is recommended to add more SAS HBAs to the system before stacking JBODs. In other words, go wide before you go deep.

As mentioned earlier, it is important to have an estimate on the required capacity during the economic lifespan of the system. As such, when configuring the commodity storage controller, it is important that enough PCI slots are available to accommodate the required storage capacity, IOPS requirement, and network port connections.

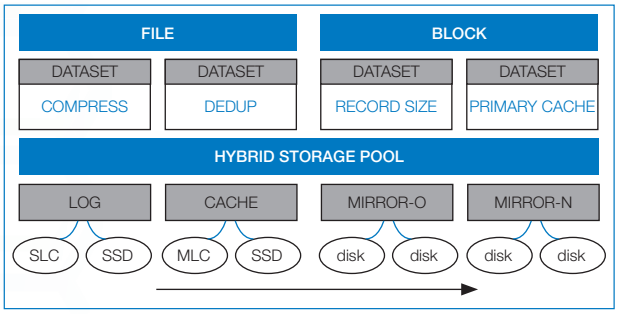

Figure 1: Scale Dynamics Stripe Using SGI NAS

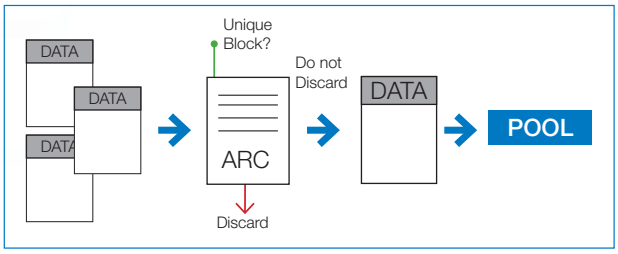

Figure 2: Inline Data Deduplication

Built-in Space Savers

Deduplication: SGI NAS is the only software storage solution that supports inline data deduplication for primary storage, which saves storage capacity and decreases the number of required I/Os when data is highly redundant.

With the SGI NAS Web interface, administrators can optionally configure data volumes for deduplication that, in virtualization environments, provide capacity savings and IO reduction where there are many systems with replicate sets of data. Data duplication only stores unique data blocks in the pool and can save up to 90% of storage capacity.

Unlimited Snapshots and Cloning: Snapshots provide an efficient point-in-time copy of the data set. Only changes to the data set are written to the pool. Cloning snapshots also enables read / write actions, only keeping track of the new written blocks to the newly created clones.

Replication is a snapshot that allows for incremental block changes to the disaster recovery site, with deduplication support over the wire. This potentially can save replication time and directly has an impact on RPO, RTO, and bandwidth costs.

SGI NAS supports an unlimited number of snapshots, and deals with snapshots as a simple increment of the reference count of the uberblock, instead of incrementing the reference count of each individual block. This leads to an unlimited number of snapshots.

Leveraging ZFS snapshot cloning for the creation of new VMs creates new images instantaneously. But, more important, having multiple VMs sourced from a base image will populate that source image into the main cache of the storage controller in SGI NAS. This increases performance and lowers the overhead of data retrieval.

VAAI: vStorage API for Array Integration provides five different benefits. In effect, the ESX hypervisor instructs the storage controller to off-load certain tasks and perform them at storage controller level, leaving I/O and CPU cycles available to the VMs.

- **SCSI write same.** Accelerates zero block writes when creating new virtual disks.
- **SCSI ATS.** Enables a specific LUN region to be locked instead of the entire LUN when cloning a VM.
- **SCSI block copy.** Avoids reading and writing of block data through the ESX host during a block copy operation.
- **SCSI unmap.** Enables freed blocks to be returned to the pool for new allocation when no longer used for VM storage.

VAAI support is applicable only with block-based protocols like iSCSI. Other SCSI commands are all performance related.

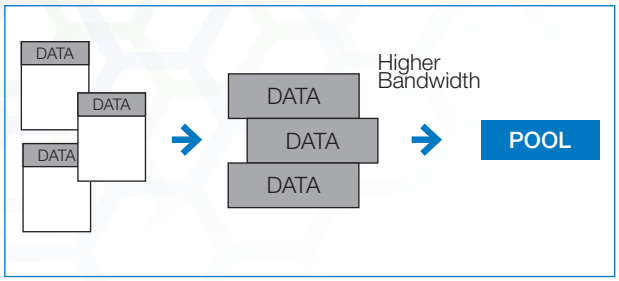

Figure 3: Unlimited Snapshots and Cloning

Compression: This is another configurable option for SGI NAS data volumes and, like deduplication, it conserves storage space and reduces the amount of I/Os. With ZFS, compression occurs at the block level and not at the file level like other file system compression implementations. By default, compression is enabled.

Compression with virtual machines can provide great savings in terms of capacity, with very small overhead on the host. In fact, performance can increase where VMs are compressed to a ratio where the bandwidth is reduced, lowering storage resource consumption.

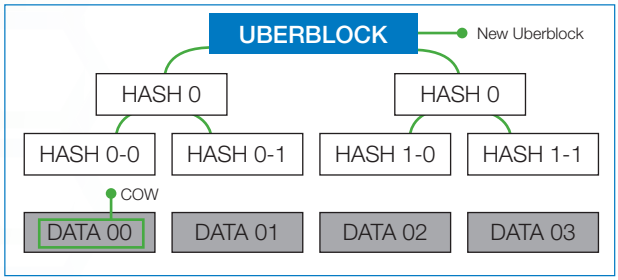

Figure 4: Copy On Write Performance

When SGI NAS is deployed for virtualization workloads, compression can be applied to enhance write speeds.

Depending on workload requirements, SGI NAS makes it easy for an administrator to apply deduplication or compression—or a combination of both to specific data volumes to enhance virtual environment performance.

Additional Functionality

SGI NAS has the ability to match the data set block size to the application I/O size. For instance, virtual machines can benefit from proper block size, again streamlining performance. SGI NAS can tune block sizes (also called "record sizes") for every individual data store.

When dealing with Raw Device Mapping (RDM), inspecting the optimal block size per LUN is highly recommended. SGI NAS provides DTrace support that allows seeing which server does what I/O at what size and location.

The block size can be changed on the fly, so when dealing with a new environment, with unknown block sizes, setting it to 4KB is a reasonable starting point. Other data set settings can include:

- Log bias set to latency
- Sync mode set to standard
- Access time disabled, therefore reducing metadata writes
- ESX swap could be consolidated into a separate low-latency SSD pool. This improves vMotion speed and reduces swap latency, if any.

Summary

Throughout this solution brief, several features have been described to show how they can help to build a cost-effective storage solution in a virtualized environment. When configuring a virtual unified storage solution, the expected workload in terms of IOPS, its read and write ratio, as well as the initial capacity requirements and growth, are essential input parameters for the initial configuration.

By choosing SGI NAS as your storage solution, you can apply the same scalability benefits that apply to server and desktop virtualization.

By adding commodity hardware, the software transforms into a cost-effective enterprise solution without the legacy design issues that can be found in appliance-based models.

SGI NAS Key Points

- Creates shared pools of storage from any combination of storage hardware, including SSDs.
- Features unlimited snapshots and unlimited file sizes.
- Delivers end-to-end data integrity, integrated search, and inline virus scanning.
- Provides full integration with virtualization approaches, even for mixed vendor environments.
- Enables efficient one-click virtual machine and other storage provisioning.

Global Sales and Support: sgi.com/global