How Internet Content Providers Store Data:

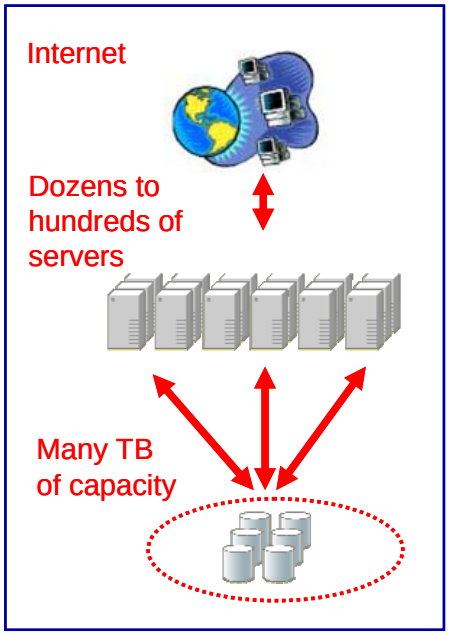
Lessons from the world's largest data sharing applications

June, 2006

ONStor, Inc. 254 East Hacienda Ave., Campbell CA, 95008 408-963-2400

The Need for Data Sharing

Imagine adding *200,000 users every day* to your storage environment. How would you keep up? Amazing as that sounds, that's the demand that IT mangers now face at some fast-growing Internet businesses. For them, storage hurdles can include a million new users per week, along with perhaps 10 or 20 terabytes of new capacity. What have these companies learned in managing their storage, and how can it help you to better manage yours?



This group of businesses, sometimes called "Internet Content Providers," have quickly become some of the biggest players in shared storage. Essentially huge data sharing services, they distribute various forms of information to millions of users. Their data sets that can extend to petabytes of capacity. Examples of large-scale data sharing applications include:

- **Online communities:** Fast growing sites such as Facebook.com and Myspace.com now add millions of users each month and even greater numbers of images.
- **Shopping sites:** E-Commerce sites and shopping comparison sites (such as Shopzilla.com and CNET) host hundreds of thousands of product images.
- **Imaging and video services:** Map imaging, and photo and video sharing services, such as Myphotoalbum.com and Flickr.com now grow their combined storage by petabytes per month.

Why are these applications relevant to IT managers in non-Internet businesses? Two reasons. First, the basic issues faced by Content Providers are similar to those seen in typical corporate environments. They need to manage fast-growing storag and maintain a strict budget, all while delivering performance, availability, and data protection.

Second, their infrastructure has a lot in common with a typical corporate datacenter. Their use of technologies such as SAN, NAS, general purpose servers, and storage management apps would be familiar to most IT managers.

The difference is the scale of the application. Because they contend with such large scale, they have gone to lengths to determine which architecture gives them the best bang for the buck. This makes their learning valuable to all.

The Storage Challenge

The storage challenges faced by Internet Content Providers include:

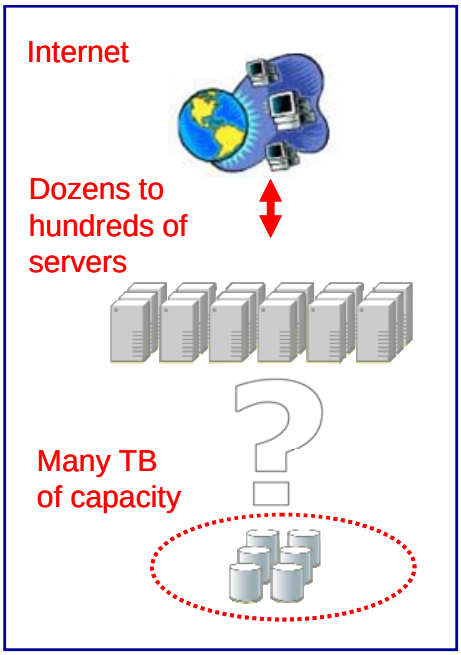
- **Explosive growth:** Some sites now add over a million users and tens of TB per week. Capacity and throughput demands may double monthly.
- **Mission critical performance:** Impatient web users expect immediate response. Slow response during periods of peak traffic is unacceptable. At a shopping comparison site for example, slow response can mean a lost referral. That referral is revenue, so time literally is money.
- **"Bursty" load:** When dealing with content that varies minute to minute, traffic may be impossible to predict. One hot new video can drive millions of accesses in hours.
- **24x7 uptime:** Users expect continuous data access at all hours. One matchmaking site estimates that an outage would cost the company $9,000 per minute as potential subscribers click away to find love elsewhere.

- **Cost, cost cost:** "Do more with less" was never more true. For these companies, IT is often the largest piece of the overall cost structure. And costs are critical; the drive for profitability has fully displaced the free-spending mentality of the 1990's.

Deciding on the Storage Infrastructure

Determining the optimal data delivery infrastructure for these applications involves hundreds of decisions, but two points are clear. The first is the need for multiple web servers. A few dozen to over a thousand web servers may be required to meet bandwidth, response time, and uptime objectives.

The second is the need for shared data. To optimally balance load across the servers, each server needs access to a common data pool. Achieving this shared data access -- and doing it as simply and inexpensively as possible -- is the challenge.



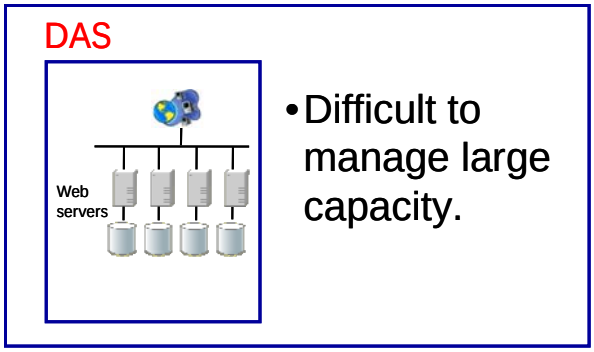
The Infrastructure Options

What is the best way to achieve shared data access? Whether you're servering data to 50 users in a workgroup, or to 1,000,000 simultaneous users in an online community, the basic options are pretty much the same.

DAS

The simplest approach is direct attached storage, or DAS. In this architecture, each server is connected to its own disk. This works fine in small-scale applications where a single device can provide all the bandwidth and capacity you'll ever need.

The problem is growth: as soon as the device reaches its limits, either in capacity or performance, additional servers are needed. If the application requires shared storage, each of those servers will require its own replicated data copy. As servers and storage proliferate, growth quickly causes the environment to become costly and unmanageable.

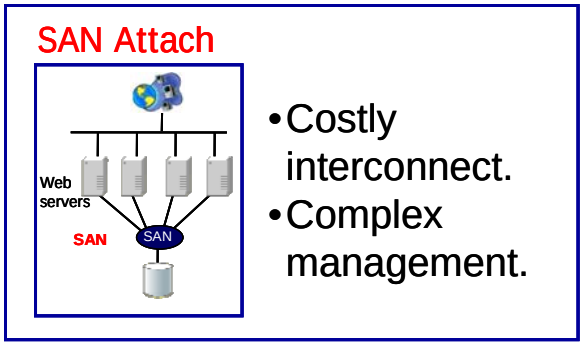


- **Pros:**
  - Simple.
  - Lowest cost for small applications.
- **Cons:**
  - Does not scale on either capacity or performance.

FC SAN

A Fibre Channel storage area network (SAN) may appear ideal for data sharing across servers. After all, SANs were designed for large-scale server-to-storage connectivity. As it turns out, there are two significant issues with the approach.

The first issue is complexity. A SAN enables shared *connectivity,* not shared *data.* What's the difference? The SAN is a switched interconnect that allows any-to-any connectivity from servers to



disk. But data is written in blocks that are meaningless by themselves. To reconstruct the blocks into useful data, you have to know how the data was laid out when written. That's the province of the "file system," which is software that controls the disk I/O. The file system normally operates in a one-to-one manner, with one server accessing the data at a time. For simultaneous data sharing (multiple servers directly accessing data from the same disk), you will need special software known as a "clustered file system" that allows the servers to share information about how data is recorded and retrieved. This allows a single copy of data to be shared, but at a cost: each server needs to host an instance of the clustered file system. This drives up management and capital expense.

Clustered file systems extract a performance penalty as well. To remain in synch, the servers must maintain constant communication with each other, a burden that becomes significant as the server count grows. As a result, performance growth when adding servers ultimately reaches a limit. This is a significant issue. At some point, adding servers will actually cause overall performance to decline.

Whether the practical limit is 16 nodes, 32 nodes, or 50 nodes, the bottom line is the same: this is not an easy way to share data and certainly is not a practical option for server farms that number in the hundreds or thousands of devices.

The second issue with SAN connectivity is cost. Each server requires an FC HBA, plus an FC port at the switch. In a 100 server environment, the total cost just for the SAN and the clustered file system software can easily exceed $700,000. When you couple this with the complexity of managing drivers across a large pool (each of those HBAs requires drivers certified with your exact storage device), most Content Providers conclude for their application, large-scale FC SAN is not the way to go. .

- **Pros:**
  - Scales well.
- **Cons:**
  - Expensive.
  - Clustered file system adds management complexity and limits performance scalability.
  - HBAs and drivers on each server.

iSCSI SAN

iSCSI is another form of SAN. Although the hype often suggests otherwise, iSCSI does bear significant resemblence to FC SAN. The full name of iSCSI is "SCSI over IP", which means "the SCSI command set communicated over Internet Protocol." The full name of FC SAN is "SCSI over Fibre Channel," which is the same information, just moving over a different media.

The advantge of iSCSI is cost. Because it travels over Ethernet, you can use the Ethernet port that's built into your server's motherboard. Switch costs are lower too, and some of the driver issues are simplified (no need for that extra HBA). Consequently, iSCSI addresses the hardware cost issue. But it does not fix the data sharing issue. You still need a clustered file system to share data across multiple servers.

- **Pros:**
  - Scales well.
- **Cons:**
  - Data sharing requires a clustered file system that adds management complexity and limits performance scalability.
  - iSCSI drivers on each server.

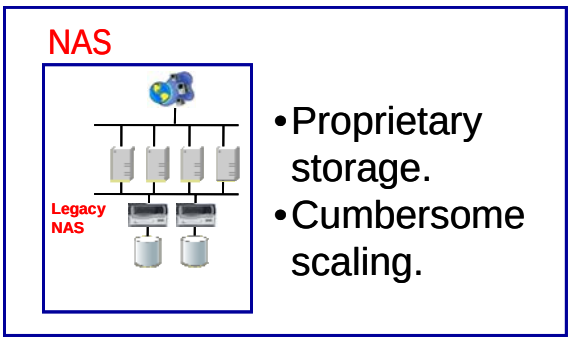
Conventional NAS

By simplifying data sharing, Network attached storage (NAS) resolves the big issues of FC SAN and iSCSI. With NAS, data sharing is built in. When a NAS device sends data to a client or server, it sends the entire file. No reconstruction is needed and the recipient needs no knowledge of the file system. Whether the file is a document, an image, a song, or an excel sheet, NAS allows any number of devices to

transparently share access. There is no need for a clustered file system and, hence, none of the associated performance, cost, and management issues.

With NAS, the interconnect is low-cost as well. Ethernet cabling simplifies connectivity, and industry standard protocols handle communication.

Conventional NAS limitations
The limitation of conventional NAS is scalability. A conventional NAS device has capacity and performance limits. When your workload reaches these limits the only recourse is to add another device. While this sounds easy (just buy one and connect it to your Ethernet, right?), it's not so straightforward. To take advantage of the new device you need to move files to it and re-map the servers that access those files. Both are labor intensive processes.



Furthermore, when the environment grows to include *multiple NAS devices*, resource utilization becomes an issue. Workloads and capacity utilization can vary dramatically across devices, and if you want to optimize response times by load balancing across devices you're again faced with re-mapping.

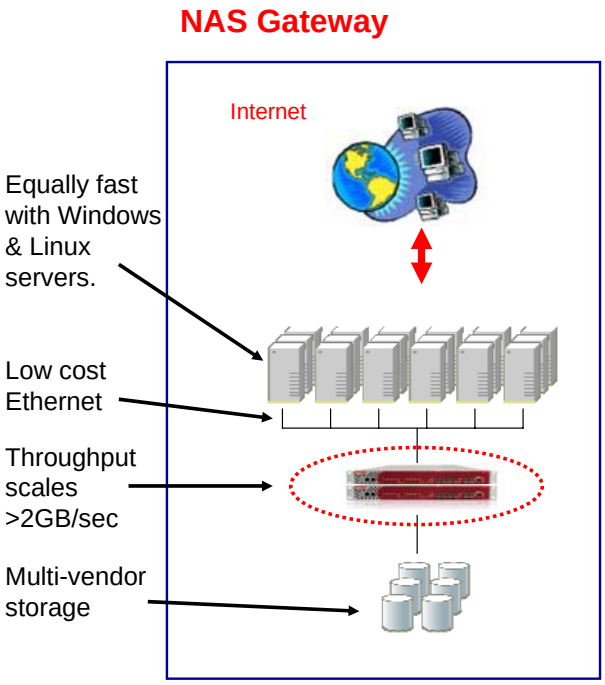
When additional capacity is needed, disk is added to one device at a time. If your needs change and you find the disk would be better off deployed on another device, you have a physical move on your hands.

A final limitation of NAS is proprietary storage. With most NAS, the disk arrays within the NAS are proprietary. That is, you may have a data center full of arrays from EMC, with lots of available capacity, but if you buy NetApp you will have to buy new disk. Not only is this wasteful, you're also adding another disk vendor to manage, another set of spares to maintain, and another volume management environment to train on. Analysts recommend keeping disk vendors to a minimum, so if you're happy where you are, it may not make sense to add another.

- **Pros:**
  - Easy to share information across servers.
  - Low cost interconnect.
  - No clustered file system, no special drivers
- **Cons:**
  - Does not scale easily.
  - Requires proprietary disk.

ONStor NAS Gateways

ONStor NAS Gateways are similar to NAS in many respects. They share data at the file level, so there is no need for special file systems, drivers, or agents. The interconnect to the servers is Ethernet, so that cost is low as well.

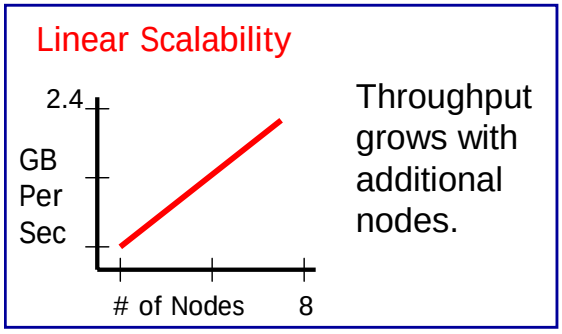


Two attributes separate ONStor NAS Gateways from conventional NAS. The first is scalability. Unlike NAS that has fixed limits, ONStor NAS Gateways scale far beyond what possible with conventional NAS, and usually beyond what's required in a single environment, thus eliminating the need to ever add new devices and re-map servers.

Performance Scalability
The architecture employs clustered nodes each of which moves up to 300MB per second. A cluster of eight nodes delivers 2.4GB per second. Because workload can be transparently balanced among the nodes without any impact to ongoing operations, bandwidth can be added on demand simply by adding nodes.

Capacity Scalability
Capacity is easily scaled as well. All disk is maintained within a single storage pool, up to 40,000TB of capacity, and is directly accessible to all nodes. This greatly simplifies your storage management tasks.

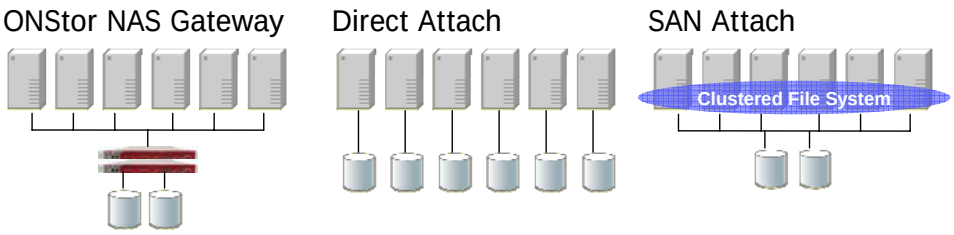
Open Storage
The second attribute is open storage. ONStor NAS Gateways do not include disk. Instead, you use the disk storage of your choice. Certified interoperable with just about every disk solution on the market, NAS Gateways let you leverage your current investment and maintain fewer disk vendor relationships and associated spares pools.



A third differentiation is value. ONStor NAS Gateways deliver 2X to 3X the price/ performance of conventional NAS. That, combined with the inherent economy of open storage, ensures superior ROI.

- **Pros:**
  - Easy to share information across servers.
  - Low cost interconnect.
  - No clustered file system, no special drivers.
  - Scales easily.
  - Open storage.
  - Linear performance scaling. Every node adds the full performance increment.
  - Space and power efficient: 1U height and 160 watts power.

Solution Comparison Chart



Scaling	**Simple.** Add performance or capacity **on demand**. No disruption.	**Complex.** Capacity and performance and capacity grow together. Disruptive.	**Moderate.** Each server requires instance of **clustering software.**
Cost	**Low cost** Ethernet interconnect. **Open storage.** High resource utilization.	**Expensive.** Capacity must usually be replicated on each server. **Low resource utilization.**	**Moderate to Expensive. Cost of SAN** attach adds up.
Performance	**Fast.** Up to 2.4GB/sec, 200K+ operations per second. Real-time **load balancing.**	**Moderate. Difficult to scale** streaming throughput.	**Good** streaming performance. Less scalable on random. **Non-linear performance growth** due to cache coherency.
Management	**Simple. Single pane of glass.** Single pool of capacity.	**Moderate.** Simple for small capacity. **Complex** for large, due to replication.	**Complex.** Must manage **clustered file system** + **large-scale SAN** interconnect.

Summary

Internet Content Providers face extreme storage requirements, but their basic concerns are familiar to all of us. Many of them have found that ONStor NAS Gateways provides a cost-effective data sharing solution for these reasons:

ONStor Content Provider Customers

- **High performance:** Up to 300MB/sec of continuous streaming bandwidth from a single node. A cluster of four will deliver 1.2GB/sec, sufficient bandwidth for tens of thousands of simultaneous broadband-quality video streams.

- **Linear scaling:** Each NAS Gateway added to a cluster delivers all the performance you paid for.

- **Space and power efficient:** At only 1U height and just 160 watts of power consumption, Bobcat packs performance into tight spaces.

- **Heterogeneous:** Windows, UNIX, and Linux application servers can share and simultaneously access information. ONStor's native protocol implementation assures high performance for all types of access.

- **Low capital cost:** NAS Gateways drive down capital cost in three ways.
  - Low upfront acquisition cost:
    - 3X the price performance of legacy NAS solutions.
    - Buy what you need now rather than anticipating demand.
  - Reduced management expense:
    - Scales on demand in a single management environment.
    - Fewer devices to manage and fewer backup software licenses.
  - Open storage:
    - Re-use the storage you have.
    - Buy the exact storage that meets your needs.
    - The *only* NAS Gateway that lets you deploy the storage of your choice.

Proven in datacenters worldwide, ONStor Clustered NAS Gateways deliver the industry's most scalable enterprise NAS solution.

Compatible with disk arrays from most vendors, ONStor Clustered NAS Gateways can either complement existing SAN implementations or provide the most cost effective path for adding NAS. Contact ONStor today and learn how over 140 enterprises saved 50% to 80% on their storage deployment with ONStor Enterprise NAS.

ONStor, Inc. 254 East Hacienda Ave., Campbell CA, 95008 408-963-2400