

ParaScale Develops TCO Calculator for Cloud Storage

Date: May, 2009

Author: Terri McClure, Analyst

Abstract: The underlying infrastructure cloud storage offerings are built upon will make or break both cloud storage service providers and enterprises deploying a private cloud, who must understand the trade-offs when choosing foundational technology. Much of this infrastructure will consist of commodity-based scale-out storage solutions and there are a number of offerings on the market today. ParaScale has built a total cost of ownership (TCO) model to help organizations quantify some of the trade-offs that need to be considered when deploying an Internet-scale scale-out storage solution. ESG has reviewed the model and believes it will go a long way towards helping organizations understand the issues that need to be considered when storing data at massive scale, but users should not use the tool in a vacuum—there are more factors to consider than those in the ParaScale model.

Cloud Storage Basics

ESG defines cloud storage as a combination of hardware, software, and processes designed to efficiently deliver storage services to the business. Delivered as a service via a subscriber model, the cloud service provider can be a company's internal IT group, a third party that delivers storage services, or a combination of both. The economics enable both the service provider and the subscriber to benefit: service providers gain economies of scale via multi-tenant infrastructure and a predictable, recurring revenue stream. Subscribers benefit from shifting storage costs to an operating expense—they pay only for what they use; they transfer power, cooling, and the carbon footprint burdens to the cloud service provider; and they gain business flexibility with subscriber-controlled on-demand capacity and performance.

There are really two components to the cloud that need to be considered, but they are often confused: the cloud service provider and the cloud enabler. Cloud service providers deliver application, platform, and/or infrastructure services via the Internet. Cloud enablers sell products and technologies that allow clouds to be built and cloud services to be delivered efficiently. ParaScale is the latter, providing the underlying software infrastructure required to make cloud storage a reality.

A number of characteristics must be included in the cloud storage offering if the service provider benefits are to be fully realized; cloud storage needs to scale quickly and to tremendous capacities. At a high-level, cloud storage must be elastic to quickly adapt the underlying infrastructure to changing subscriber demands and automated so that policies can be leveraged to make underlying infrastructure changes quickly without human intervention. Of course, the devil is in the details. The storage cloud must be:

- SLA-driven, automated, and integrated to provide quick response times to user demands
- Policy-based, with deep levels of automation to move data to the appropriate tier based on content and SLAs
- Secure, reliable, and scalable to multi-PB capacity with a unified management view for both block and file storage

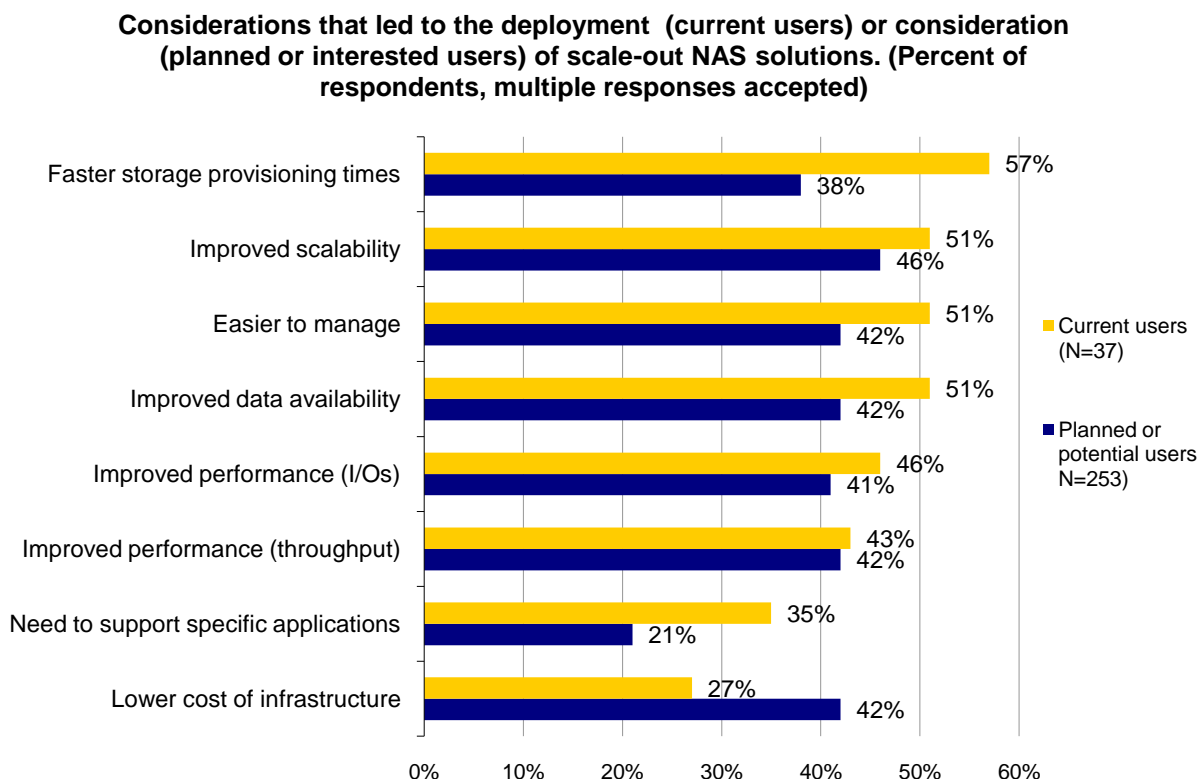
The last pieces of the cloud storage puzzle are federation and flexibility. Enterprise users have varying degrees of comfort when it comes to storing corporate data offsite, no matter how secure the storage service provider is or claims to be. The need to meet data security concerns and gain a level of comfort with the cloud storage model is driving three deployment models: public, private, and hybrid. In a private cloud, all components live within the firewall or in a physically isolated offsite storage infrastructure accessed via VPN. In a public cloud, all components sit outside the firewall in a shared infrastructure that is logically partitioned (multi-tenant) for security

and accessed over secured IP. In a hybrid cloud, some components sit within the firewall—such as a local gateway, storage cache, or bulk storage capacity—while other components are outside the firewall on multi-tenant infrastructure (storage capacity).

Considerations for Cloud Storage Infrastructure

Efficient management of the underlying cloud storage infrastructure and high utilization rates are key to the success of cloud storage service providers that will need to manage multi-petabyte storage instances. Traditional scale-up storage solutions fall short when it comes to petabyte scale; users would need to manage tens or hundreds of scale-up systems to meet such performance and capacity requirements, and would be stuck with more of the storage stovepipes so many data centers are struggling to deal with today. Internet-scale capacities require a new approach to keeping storage management costs in check and creating an elastic and automated storage environment. Scale-out or cloud storage architectures offer that approach.

FIGURE 1. USERS ADOPT SCALE-OUT NAS FOR EASE OF MANAGEMENT AND AGILITY



Source: ESG Research Brief, *Scale-Out NAS Adoption and Market Drivers*, February, 2009

Not all scale-out architectures are the same—each has unique strengths and weaknesses. Many are integrated solutions that come bundled as hardware & software, limiting the flexibility of the underlying storage infrastructure, and almost all only support either block or file access. Some require agents to run on all clients, adding additional management overhead. Others have very low utilization rates thanks to reserve space and RAID protection schemes. ParaScale takes these aspects into consideration with its TCO model.

ParaScale's Cloud Storage TCO Model

With management costs far exceeding hardware costs over the life of a storage system, the ParaScale model rightfully takes a balanced view of acquisition cost as well as the management cost per terabyte of data stored. Because ParaScale software allows lots of heterogeneous storage to be managed as a single system image, the scope of management for a single administrator is in the hundreds of terabytes, compared to managing typical scale-up systems where it is in the low tens of terabytes. Management and training savings are also gained by standardizing on one management platform, thanks to the ParaScale software's ability to manage a

heterogeneous storage environment. But ParaScale takes a conservative approach and defaults the TB per administrator for competitive systems at 100:1.

Users also need to be aware of actual storage utilization rates. ParaScale software is capacity efficient. Though it does make two copies of every file stored, both are usable copies and can be distributed geographically. This provides performance improvements and availability advantages versus in-the-box mirrors because there is little metadata penalty in using the copies and geographic dispersion has inherent high availability. Note also that, since ParaScale does not necessarily require RAID, customers can avoid lengthy rebuild times that can be as much as 18-24 hours with today's largest drives. The ParaScale model is conservative in its approach here, too, to competitive offerings; the default values of 70% utilization associated with competitors are more than reasonable (ESG rarely sees utilization rates at 70%; a rate of 30-50% is far more common, especially in scale-up systems).

Of course, the ParaScale TCO model focuses on areas that are easiest to measure and where ParaScale holds the biggest advantage: utilization and management. Looking at potential improvements in utilization and management will help define clear opportunities for savings, but other considerations must be taken into account.

What the Model Doesn't Tell You

ParaScale's model is thorough enough to give users insight into many factors that should be considered, but every IT environment has issues that an off-the-shelf model cannot possibly take into consideration. Fortunately, ParaScale realizes this and does not try to mask the issues or play tricks with numbers. It provides enough information for users to make an educated decision and reflect a reasonable amount of potential cost savings. But areas that users need to pay closer attention to because they are likely unique to their environments include:

- **Delayed new hardware purchases by reusing/repurposing existing assets.** Leveraging and repurposing legacy storage and servers as nodes within the ParaScale cluster enables IT shops considering a shift to deploying a private cloud to leverage and consolidate heterogeneous arrays into tiered storage pools managed in the ParaScale system image.
- **Power and cooling savings** based on better utilization of individual nodes. Higher utilization rates mean fewer physical disk drives need to be used, racked, powered, and cooled.
- **Abstraction of the hardware layer**, removing physical connection points between clients and storage. This insulates users from downtime during data migrations, hardware refresh cycles, or failures of storage or metadata nodes.
- **No more forklift upgrades.** With ParaScale's software abstraction layer, new servers can be added on a rolling basis, even as older hardware is removed from the environment. Users considering multi-year or multi-decade data storage environments benefit from avoiding a major data migration project every 4 or so years.
- **Global namespace management efficiencies.** Though management is somewhat covered in the model, customers evaluating foundational NAS infrastructure for cloud services need to consider the efficiency and end-user ease-of-use gains from having a single mount point versus multiple mount points as well the associated management savings. Spreadsheets don't work for managing file system mount points at very large scale.
- **Automatic performance balancing**, facilitating the day-to-day maintenance tasks associated with managing a large-scale storage implementation in the multi-petabyte range. Again, this issue is touched upon in the model a bit, but when scaling to very large capacities, special attention needs to be paid to how much administrative time could be spent on manual load balancing—across drives and across storage nodes. The last thing organizations need is administrators remapping hot spots and manually migrating data.

The Bottom Line

When a vendor creates a TCO model, there is an obvious bias: it is in the vendor's best interest to ensure its solutions gets the most positive results. You should not use these tools solely for the end result; rather, they should be used to advance your thinking and for data to input into your own business plan. The ability to buy off-the-shelf commodity hardware, scale to massive capacity, and manage as a single system image will save a bundle on capital expenses and ongoing administrative costs. Operational costs will also be positively impacted by higher utilization rates: ESG recently conducted an in-depth survey of 492 senior IT professionals concerning

their organizations' IT spending plans and priorities over the next 12-24 months. For both midmarket and enterprise IT professionals, reducing operational costs outweighed any other criteria for justifying IT purchases over the next 12-24 months by a factor of almost two-to-one.¹ Reducing operating costs by shifting to a cloud model is one of the key drivers behind the current buzz around cloud storage.

A lot of lessons were learned early in the decade when a number of vendors tried to make a go of storage as a service. The market was ready—the technology wasn't. The market drivers for cloud storage today are pretty much the same as those that existed in 2002! Data is growing like mad. Just like death and taxes, growth is a sure thing, even with the economy falling off a cliff. And the economics work; in a 2008 research survey of 516 IT executives at midsize companies, 30% cited a lack of physical space in the data center as a top challenge.² In another survey of 504 large enterprise storage buyers, 28% cited running out of power and cooling capacity as a major challenge.³ Buying capacity as a service or deploying an internal cloud storage infrastructure that leverages new technologies for higher utilization rates is a lot less expensive than building a new data center to accommodate space, power, and cooling demands.

One of the biggest reasons the cloud storage service provider market did not thrive then is that the service providers were trying to use platforms designed for big, I/O-intensive data center applications and scale-up platforms with bounded capacity limits. Today, there are technologies available to overcome the physical boundaries of the storage array—they are designed for efficiency at scale, and designed to be automated and elastic. Both public and private cloud storage providers can gain from leveraging these types of architectures.

Selecting the underlying infrastructure is one of the most important decisions a cloud storage service provider will make, whether as a public cloud provider or as an IT shop transforming its storage infrastructure into a service-based model. Leveraging a tool like the ParaScale TCO calculator, in conjunction with the other considerations outlined in this paper, users can make a much more educated decision regarding what that infrastructure should be.

All trademark names are property of their respective companies. Information contained in this publication has been obtained by sources The Enterprise Strategy Group (ESG) considers to be reliable but is not warranted by ESG. This publication may contain opinions of ESG, which are subject to change from time to time. This publication is copyrighted by The Enterprise Strategy Group, Inc. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of the Enterprise Strategy Group, Inc., is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact ESG Client Relations at (508) 482-0188.

¹ Source: ESG Research Report, *2009 Data Center Spending Intentions Survey*, March 2009.

² Source: ESG Research Report, *Medium-Size Business Server & Storage Priorities*, June, 2008.

³ Source: ESG Research Report, *2008 Enterprise Storage Systems Survey*, November 2008.