Cloud File Data Storage Consolidation and Economic Comparison Model

Server StorageIO® Industry Trends Perspectives Report

Distributed File Server and Consolidated Cloud Storage Economic Comparison

INTRODUCTION

This Server StorageIO® Industry Trends Perspective white paper report looks at industry and customer trend challenges about remote (on-prem) distributed file-servers and cloud storage consolidation decision making.

IT data infrastructure[1] resource (servers, storage, I/O network, hardware, software, services) decision making involves evaluating and comparing technical attributes (speeds, feeds, features) of a solution or service. Another aspect of data infrastructure resource decision making involves assessing how a solution or service will support and enable a given application workload from a Performance, Availability, Capacity, and Economic (PACE[2]) perspective.

Who This Applies To

Information Technology (IT) organizations, Cloud Service Providers (CSP), Managed Service Providers (MSP), remote office branch office (ROBO) and workgroups with:

- Distributed file servers and shared data
- Budget constraints needing to reduce costs
- Growing application and data demands
- Shrinking staff or external headcount

Keep in mind that all application workloads have some amount of PACE resource requirements that may be high, low or various permutations. Performance, Availability (including data protection along with security) as well as Capacity are addressed via technical speeds, feeds, functionality along with workload suitability analysis. The ***E*** in PACE resource decision making is about the Economic analysis of various costs associated with different solution approaches.

Industry and Customer Trend Challenges

Common demand drivers and industry trends causing challenges for IT organizations include:

- New (and existing) applications that are data intensive with video, image, rich content
- Increase in distributed, edge and fog enabled data infrastructure (remote sites) with data
- Reliance on information along with collaborative 7 x 24 workforces
- Expanding data footprint due to growth as well as associated overhead
- New and emerging data protection along with security threat risks
- Focus on reducing costs while maintaining or boosting customer service
- Navigating the hyperbole of various technology solution deployment options
- Need to comply with regulations such as the Global Data Protection Regulation (GDPR)

[1] Data Infrastructures exist inside physical datacenters (legacy and public cloud) and include servers, storage, I/O networks, hardware, software, service, processes and policies. Learn more at https://storageioblog.com/data-infrastructure-primer-overview/ .

[2] All applications have some amount of PACE requirement attributes. Learn more about PACE characteristics and related topics in "Software Defined Data Infrastructure Essentials" (CRC Press) https://tinyurl.com/yblnkuwu

Informed Data Infrastructure Decision Making

Making informed decisions for data infrastructure resources including cloud storage consolidation and distributed file servers involves technical, application workload as well as business economic analysis. Which of the three (technical, application workload, financial) is more important for enabling a business benefit will depend on your perspective, as well as area of focus. However, all the above need to be considered in the balance as part of making an informed data infrastructure resource decision.

TECHNICAL COMPARISON (SPEEDS, FEEDS, FEATURES)

Technical capabilities including speeds feed, feature functionality capabilities of a solution or service are essential. Server StorageIO sees some environments make decisions based primarily on the technical features of technology without regard to economic or business enablement. While technology features of a solution are important, ask yourself do you need the functionality, or want it. In other words, does a solution meet or exceed your must-have list while adding value with what you would like to have, or would it be a nice to have the capability? Also, keep in perspective how likely are you to use a feature function capability or merely use it as part of a decision-making process.

APPLICATION WORKLOAD (HOW DATA USED)

A common challenge is customers making decisions about how a solution will support or enable a given workload without an analysis. A simple analysis can be using a vendor or service provider published information on how a solution supports a given application workload similar to yours. Another approach is to install a solution or try a service as a proof of concept (POC) or another test. While it may seem common sense, look beyond the necessary available usable storage space capacity in a solution, also focus on the performance, as well as availability attributes to support the needs of your environment. If you are not sure what to do, test, try or how to evaluate, consult with others including independent third-parties such as Server StorageIO among others. Learn more about workload (and data) PACE and related decision-making topics at https://storageioblog.com/application-data-value-characteristics-part1/

FINANCIAL COMPARISON

Moving beyond technical speeds feeds, feature functionality comparisons, along with evaluating how a given solution aligns to support your application workload data activity, let's look at an economic-based financial comparison mode. Pro-forma forward-looking predictive model or forecasts compare estimated costs on an economic basis. The model described below combines your known costs and configuration information along with prevailing industry standards, industry recommended general rules of thumb (RUT), and those found with a Google search.

The more you know about your data infrastructure environment (central, cloud as well as ROBO or distributed) the better off you will be. This includes identifying the application workloads along with their associated PACE requirement characteristic, as well as costs. By not understanding your costs for supporting different workloads, you are at a disadvantage of not being able to make informed decisions. Worse, you could perhaps miss out on being able to capitalize on potential economic benefits.

An Economic Example – Cloud Storage Consolidation

The following example of a fundamental financial model comparing remote distributed on-prem file server storage to centralized, consolidated cloud storage file sharing can be done on the back of a napkin, or with a spreadsheet. Table-1 shows various parameters that reflect remote distributed file server costs.

	Description and Comments
CTY	Storage Cost per TB per Year - Annual cost for enterprise data storage on a per TB basis, which we will pro-forma at $2,113 for enterprise-class storage. Given the numerous permutations of configuration (RAID, Replication, HA), along with types of storage solutions, your costs, as well as those you will find in the industry will vary.
DER	Data Efficiency Ratio - This metric is similar however different from more common storage efficiency ratios. Storage efficiency ratios tend to focus on capacity space savings (or avoidance) using data footprint reduction (DFR[3]) technologies and techniques. DER, on the other hand, is the indicator of how much common data exists across multiple sites that add to data as well as storage space capacity overhead.[4]
EMC	Extra Management Cost – Any additional management cost per remote location to support distributed file servers including server, storage, networking, troubleshooting, diagnostics.
TMA	TB Managed per Admin – The average number of TB managed per admin. Note that some admins can handle a more significant amount of TB compared to others depending on the type of enterprise storage among other factors.
SAE	Storage/server Admin Efficiency – How efficient an admin is as they have to share their time managing across multiple remote sites. With more locations, the efficiency goes down.
AMC	Average Management Cost – Cost of storage/server/system admin x SAE. Annual admin cost includes salary and other benefits. AMC is adjusted based on TB Managed per Admin (TMA).
GSR	Gross Storage Required - Total amount of usable storage capacity needed to be adjusted for data efficiency ratio (DER). E.g., Total TB of Data / DER.
ARC	Additional Remote Cost – Any additional cost for remote file server management including software tools, staff costs, facilities (power, cooling, floor space).
RDC	Remote Data Protection Cost - Cost of remote data protection and backup software fees.
DSO	Distributed Storage Overhead – Is the overhead of primary along with additional admin cost (EMC) adjusted for the number of sites and TMA, plus RDC and ARC.

Table-1 – Distributed File Server Remote Storage (e.g., on-prem) cost and parameters

Using table-1 parameters, we can do an analysis of 20 sites each with 6TB of data. Assuming each site has one file server with 6TB utilized storage, the combined monthly cost (AMC) of $40,635 (20 sites).

[3] Data Footprint Reduction (DFR) are tools, technologies and techniques to reduce or counter impacts of expanding data footprint along with their associated overhead (space capacity, data protection and storage management, costs). DFR technology and techniques include archiving, copy data management, compression, compaction, consolidation, data management, data deletion, dedupe, thin provisioning, space saving snapshots or point in time copies and storage tiering among others. Learn more about DFR at https://storageio.com/dfr and in Software Defined Data Infrastructure Essentials (CRC Press) https://tinyurl.com/ydeartvr .

[4] A higher DER, leveraging data and storage consolidation results in less required storage capacity reducing management costs. Centralized data infrastructures tend to have less overhead (more efficiency managed) duplicate data resulting in a higher DER ratio (for example 85%). Distributed sites including remote file servers tend to have more duplicate data a lower DER ratio (for example 65%). Note leveraging DFR can result in further storage savings.

Note that actual usable storage capacity could be 9TB or higher per location with a DER of 65%. The file servers are being managed to a target of 65% utilization (or better), also assume two (2) hours of extra admin time per month (networking, server, other activity) and an SAE of 80% (e.g., overhead of managing 20 sites) or about $2650 per month. Additional Remote Cost (ARC) for facilities, power, cooling and other for this example is 3.5% of annual storage cost or about $1,138 per month.

Also assume annual data protection, backup and associated management software costs of $995 per site. Assume yearly admin cost (salary and benefits) of $140,000 and the ability to manage 350TB per person. Another assumption is using an annual CTY (see table-1) of $2,113 per TB of enterprise-class distributed file server storage. Also, assume that 15% of data stored across remote sites is redundant that can benefit from further consolidation (e.g., DER benefit).

For data protection in this example, assume no additional RAID, mirroring, parity, erasure code or HA. Note however that RAID or other data protection for availability would increase remote site costs, as well as should enhance any storage consolidation benefits due to lower overhead.

Total estimated usable storage capacity across the 20 sites is 184.6TB with a monthly cost of about USD 46,081, or $1,658,915 for three years (36 months). To contrast the above example in a consolidated cloud storage scenario, in addition to remote location (on-prem) costs and parameters discussed above, table-2 outlines parameters that reflect consolidated cloud storage costs.

	Description and Comments
CSS	Cloud Storage Software – Any software that is installed on to cloud compute server instances aka virtual machines (VM) to support file sharing, along with appropriate remote client software. The software may also include additional data management, data protection, cache, and general storage management functions.
CVM	Cloud VM – Cloud virtual machine (cloud instance) cost. Note the CVM(s) should be sized to support the appropriate number of guest, client or remote sites to a given level of service with adequate CPU, Memory, I/O, networking, high-performance block storage as well as redundancy. Another consideration is how you purchase cloud compute resources such as pre-pay reserved instances (RI), on-demand as well as bring your host operating system license.
CSC	Cloud Storage Capacity – Amount of durable high-performance persistent (e.g., non-ephemeral) block storage such as AWS EBS or similar for storing shared consolidated data.
ACE	Average Cloud Egress – Cost of accessing or retrieving data from cloud storage. Note most clouds provide free Ingress (entry, storing of data, incoming) however have fees or out-going.
CER	Cloud Egress Ratio – Ratio of how much storage or data access impacted by reads and subsequent ACE costs. Note that improved remote cache capabilities can reduce the amount of network traffic and subsequent associated cloud ACE.

Table-2 – Cloud consolidated shared file storage cost and parameters

As an example, let's assume CSS of $9,200 (per month) that handles a consolidated shared file serving along with host or client-side caching along with associated management tools. Monthly Cloud compute VM or CVM costs for supporting workload of 20 sites is $279.74 and CSC of $6,505.41 assuming an 85% DER or consolidation benefit. Data access by remote clients of consolidated shared storage assuming a 25% CER would result in about $3,176 per month in ACE for cloud network egress fees (assuming that edge caching is used). Also note that efficient client or edge cache software could enable a CER of 50% or higher depending on application data access patterns, and amount of data sharing.

Note with improved client-side cache, CER and ACE could decrease. However, those savings are subject to the dynamics of different application workload profiles. The resulting cloud consolidated storage cost per month for this example is about $19,162.48 or about $689,849 for 36 months (three years).

The net result of a fundamental financial pro-forma analysis model for enabling decision making or project justification of a cloud-based storage consolidation project comparisons is shown in figure-1. Using the examples outline previously in this white paper, the three years (36 months) estimated cloud consolidated storage file sharing costs would be $689,849 (a savings of $969,066 or just under a Million dollars) compared to $1,658,915 for distributed on-prem file servers. Note that this example analysis does not reflect original enterprise-class remote file server acquisitions, only those of an on-going basis.

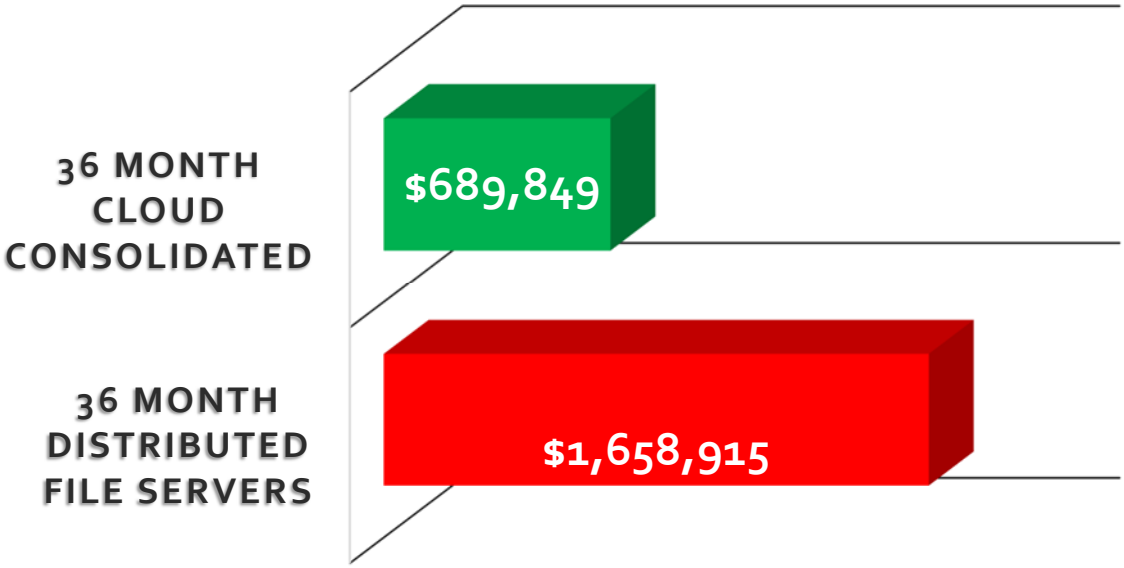

Figure-1 – Estimated consolidated cloud storage cost savings compared with distributed file servers

Note that "your mileage will vary," or your costs will vary depending on many different factors outlined above. Your actual results may differ based on the parameters and values used, as with any model or forecast, garbage data in and result in garbage info out along with resulting decision making. While the above is a fundamental analysis model that can be done on the back of a napkin, or in a spreadsheet, it can also be extended to reflect additional detail.

Server StorageIO recommends starting with a fundamental model like that presented, use your known costs, or select those available from the industry that are reflective of your environment. Likewise, cloud compute, storage and networking charges vary depending on the types of services used. Exercise caution in selecting lowest cost classes of compute instances or VMs, storage and network resources that may not be reflective of what is needed to support your workload.

Another tip recommendation is that if you find yourself spending too much time trying to determine what your variable costs are, or values to use, reach out and ask others. Likewise, to conduct a fundamental economic business model comparison or justification should not have to result in an extensive multi-month professional service consulting model for most environments.

The Server StorageIO view: What this all means

When comparing and making data infrastructure resource decisions, consider the application workload PACE characteristics. Also keep in mind that PACE means Performance (productivity), Availability (data protection), Capacity and Economics. This includes making decisions from a technical feature, functionality (speeds and feeds) capacity as well as how the solution supports your application workload.

Some applications need more performance (server computer, or storage and network I/O), while others need space capacity (storage, memory, network, or I/O connectivity). Likewise, some applications have different availability needs (data protection, durability, security, resiliency, backup, business continuity, disaster recovery) that determine the tools, technologies, and techniques to use.

Key Points and Takeaway Items

Key take away points include among others:

- ✓ Make informed decisions that include:
  - o Technical capabilities (feature/functions)
  - o Workload capability (productivity)
  - o Economic comparisons (business benefit)
- ✓ Know your cost, start simple and expand
- ✓ Economic models should be relevant
- ✓ Fundamental models should be easy to use
- ✓ Balance overly simple, or complex analysis

The E in PACE is for Economic considerations as part of a technology comparison. Budgets and economics are also always a concern, which for some applications means enabling more performance per cost while others are focused on maximizing space capacity and protection level per cost.

When it comes to making cloud-related data infrastructure and associated resource decisions, avoid flying blind. Have good insight that includes technical, workload and user productivity performance as well as economic situational awareness. Leverage a fundamental economic analysis model for comparing cloud storage consolidation costs to distributed file serving for making informed decisions.

ABOUT THE AUTHOR

Greg Schulz is Founder and Sr. Consulting Analyst of independent IT advisory consultancy firm Server StorageIO (e.g., StorageIO®). He has worked in IT for an electrical utility, financial services, and transportation companies in roles ranging from business applications development to systems management, architecture, strategy, performance, and capacity planning. Mr. Schulz is the author of the books "Software Defined Data Infrastructures Essentials" (CRC Press 2017), as well as Intel, Recommended Reading List books "Cloud and Virtual Data Storage Networking" and "The Green and Virtual Data Center" via CRC Press, along with "Resilient Storage Networks" (Elsevier). Greg is a Microsoft MVP (Cloud Data Center Management) and VMware vExpert. Learn more at storageio.com and storageioblog.com on Twitter @StorageIO.