

SOLUTION BRIEF TRANSITIONING TO A “TRUE” DIGITAL LIBRARY



Data is the New Oil, and it has to be Online to Extract the Value

OPPORTUNITY / CHALLENGE

Big Data is playing an increasingly larger role in how businesses of all sizes across all industries grow revenue and improve operational efficiency. Users are demanding more access to big data sources and IT organizations need a plan to manage access to these information sources.

Managing the exploding volume of data is a huge challenge for enterprise IT. CIOs and IT Managers are experiencing sticker shock with escalating cost and complexity to manage it. Unstructured data is growing at >40% per year, yet overall IT budgets are increasing at only 5%, leaving a substantial gap in needed resources.

Top five storage pain points for most organizations according to Gartner^{*1} are outlined below. While these points generally focus on operational efficiency, business units want to keep the ever-growing data online and accessible to analytics programs in order to grow the business.

- Managing Growth
- Capacity Forecasting
- Backup Administration & Management
- Managing Cost
- Managing Complexity



Protecting this valuable data is critical too. RAID has served the enterprise well for many years, but it can no longer scale to meet Big Data requirements. Drive sizes are 3/4TB today and according to Seagate^{*} could be 60TB by 2016. With the Petabyte fast becoming “the new” Terabyte the industry needs a new approach.

Luckily there is a new storage solution built from the ground up by Amplidata that addresses these challenges. A solution that protects data with >fifteen 9s of durability, is modular, scalable, flexible, and highly automated to significantly improve CapEx and OpEx. It is also a system that is safe and secure in your datacenter at fraction of the cost of traditional storage systems.

^{*}Other Brands and Names are the property of their respective owners.

¹Gartner 2012 CIO survey.

The Cost and Complexity of Traditional Storage Continues to Spiral Higher

MOVING TO A “TRUE” DIGITAL LIBRARY

Content repositories are the new digital library; however, if we compare them to a “true” library where we put a book (data) on a shelf, never have to move it or make backup copies, and it remains accessible for all to use – they fall short.

A common strategy to reduce cost is to move data from expensive high performance systems to lower cost systems. There are local backup copies either on disk or tape and there is another copy off-site for disaster prevention/archive. Another scenario is to have multiple copies on disk-based systems at different locations for backup and disaster prevention. Both scenarios require additional and costly software to perform things like backups snapshots, replication, archive, and disaster recovery functions.

If this sounds expensive and complex, it is. That book on the shelf (data) is being rewritten multiple times on different systems and media types. What if you could consolidate multiple storage tiers into a primary tier for the highest performance needs and a single unified tier (“True” Digital Library) for everything else?

RAID No Longer Scales to Meet Current Data Protection Requirements

RAID AND BIG DATA, NOT THE BEST OF FRIENDS

Let’s quickly address the fact that RAID no longer scales to meet current data protection requirements. Prior to multi-terabyte size drives, RAID was a viable solution. With drive capacity growth projected to reach 60TB by 2016, and throughput gains flat, RAID rebuild times become unimaginable.

To rebuild today’s 3/4TB drives can take 10s of hours or even weeks depending on the assigned priority. With RAID 6 only protecting against two simultaneous failures there’s a significant risk of a cascaded drive failure, or an unrecoverable error during the long rebuild leading to the loss of all data in the RAID set. IT managers could also find their systems in perpetual rebuild, grinding their business to a halt.

If rebuilding data on a large drive in a RAID set seems challenging, restoring data from disk backups or tape borders on sadistic. Every IT Manager or CIO who has ever had do it will tell you how difficult it can be. Amplistor renders backups and restores obsolete. It does this with unique and powerful capabilities that other enterprise-level storage system find difficult to match.

Amplidata has Redefined Storage with AmpliStor

THE SOLUTION – UNIFIED STORAGE WITH AMPLISTOR

Amplidata spent five years developing a solution that addresses these challenges. Businesses of all sizes can now implement a “True” Digital Library with AmpliStor, a software defined storage platform that delivers unbreakable durability, infinite scalability, and extreme efficiency. IT Managers and CIOs can reduce overall cost and complexity by offloading expensive primary storage and/or consolidating near-line and archive tiers into a single unified tier – simplifying their hardware and software infrastructure.



SYSTEM ARCHITECTURE

You might be asking how is that possible? AmpliStor's has a unique modular architecture with a fully abstracted software stack optimized for commercial-off-the-shelf (COTS) Intel-based hardware. This means you get flexibility and choice and the ability to easily take advantage of the latest CPUs and drive densities while never having to move data off the system. AmpliStor is ideal for deployments from 100s of Terabytes to Exabytes with literally no limit.

The two-tier intelligent multi-threaded grid architecture connected with a 10GbE/1GbE fabric, allows performance and capacity to be scaled independently and automatically by simply adding more controller and/or storage nodes. Patented BitSpread and BitDynamics, extract maximum performance from the latest CPUs scaling linearly with each new node. System configuration and management is through a single pane that provides easy access to all key system resources and processes. All of these breakthroughs add up to extreme efficiency and longevity with ROIs possible over five years vs. three years.

Key Benefits

- Easy Integration
- REST, NFS, CIFS, Other
- Runs on COTS Hardware
- > Fifteen 9s Durability
- Limitless Scalability
- Fast Self Healing
- Single Pane Management

AmpliStor Unified Storage Services Software Stack

The heart of the system is the amazing fully abstracted software stack that can run on a variety of hardware platforms. It consists of four services layers including; data protection, data integrity and repair, disaster prevention, and unified global access. Let's explore each of these now.

BitSpread – Data Protection and Scalability

Patented Bitspread® technology is the core of the data protection and scalability services layer that allows users to store and protect data according to a chosen durability policy. Data stored in an AmpliStor namespace are referred to as objects. BitSpread delivers unbreakable durability at greater than fifteen 9s and can tolerate up to 19 simultaneous hardware failures (two for RAID 6) with no data loss. That's pretty amazing compared to RAID and a must have for big data now.

Incoming data is sliced, encoded with hash checks and automatically spread per the chosen policy as wide as possible across nodes, drives, and racks within the datacenter to minimize the impact of a hardware failure. BitSpread also stores three copies of the metadata across three controllers as well as on storage nodes to protect against a major disaster. Different buckets of data can be created within AmpliStor, each with its own durability policy to further maximize capacity and data efficiency. Durability policies can be changed on-the-fly with the data being automatically "upgraded" to the new policy. Try doing that with RAID!

BitDynamics – Data Integrity And Repair

Moving on to BitDynamics, Amplidata's other key patented technology that enables continuous data integrity audit and repair services. BitDynamics is incredibly smart as it protects against disk and bit errors by predicting where failures may occur and takes proactive steps to prevent problems before they occur. BitDynamics automatically self-heals in the background to bring the entire system back to the chosen durability level.



*Other Brands and Names are the property of their respective owners.

AmpliStor's multi-threaded grid architecture brings a lot of parallelism to the system. If a data repair is required, all affected nodes participate in parallel taking only minutes to hours to repair the data vs. 10s to days or even weeks. BitDynamics does its job automatically, and out-of-band, never consuming more than 10% of the system's performance. No more support calls about slow systems.

GeoSpread – Disaster Prevention and Multi-Site Accessibility

If a little data dispersion is good, then more is better. That's exactly what AmpliStor provides with GeoSpread. BitSpread offers superior data protection within a datacenter and GeoSpread extends the BitSpread hierarchy to spread data across multiple datacenter/sites. Metadata is stored at three different datacenter locations to protect against a datacenter outage. With GeoSpread, an entire data-center can become unavailable without losing data or access to the data.

From an efficiency perspective, a typical RAID6 replicated solution requires ~60% more raw capacity than a three-site AmpliStor solution. As a result, CapEx and OpEx are significantly lower because fewer systems are needed, large SATA drives can be used with no fear of data loss and disaster recovery software is eliminated.

Unified Global Access and System Manageability

Last but not least, AmpliStor is easy to integrate into existing environments by the unified access services layer. All major interfaces are supported including, native REST, S3* and NFS* and CIFS. Data is accessed via a locality aware, unlimited global namespace. A number of namespaces can be created with different attributes and policies, essentially making tiers within the same system.

To further accelerate performance, intelligent local caching is possible at each location and the system is fully secure with end-to-end security, authentication, and encryption. To track usage by customer, multi-tenancy is supported and can track capacity, transactions, and bandwidth usage per namespace/per user. The system is easily managed through a single pane of glass with a rich GUI console.

CONCLUSION

There is no question the volume and velocity of data is only increasing and that traditional RAID-based systems can no longer meet current needs without crushing IT budgets with escalating cost and complexity. AmpliStor paves the way to implement a "True" Digital Library with unbreakable durability, infinite scalability, and extreme efficiency. AmpliStor's efficiency means you get near-line storage at an archive price. That's pretty amazing!

To find out more about how Amplidata and AmpliStor can help you implement a "True" Digital Library, send us an email or give us call. We are ready to help.

North America: 408-433-1600
EMEA: +32 9 324 25 90
Email: info@amplidata.com

**AmpliStor Uses 60% Less
Capacity than a RAID6
Replicated Solution**

**Near-line Storage at an
Archive Price**

