



Industry Trends and Technology Perspective White Paper

BlueArc Titan NAS: Scalability with Stability

Addressing Storage Management Challenges

By Greg Schulz

Founder and Senior Analyst, the StorageIO Group



December 12th, 2006

This white paper is compliments of BlueArc and looks at the benefits of the BlueArc Titan 2000 series as an integrated hardware-software solution providing intelligent, transparent data and storage management capabilities. The BlueArc Titan 2000 features that address on-going tasks including data protection, data and storage management, workload, server and storage consolidation, and performance optimization, are analyzed.



Introduction

The value proposition of Network Attached Storage (NAS) is ease of use, including installation, simplified on-going management and transparent upgrades to support growth without introducing additional costs or complexity. Similarly, a NAS solution needs to support scalability with stability, meaning that as more users, storage capacity and feature functionality are added, the storage system should remain available and support predictable performance to enable increased utilization of storage capacity without a corresponding decrease in application performance and subsequent bottlenecks. An effective NAS storage solution includes robust hardware with extensible software features combining virtualization, data protection, and flexibility with ease of use.

IT storage management challenges

In general, common data and storage management challenges are centered on having to do more work and support more applications, requiring additional storage and enhancing data protection while delivering improved service to IT customers. As if these needs were not enough, most IT environments are also faced with shrinking or flat budgets while being forced to protect and manage more data with less time.

Storage management challenges facing most IT environments include:

- Performance and capacity scaling limitations resulting in proliferation of NAS storage and file servers
- Increased cost and complexity from managing and protecting data on islands of NAS and file servers
- Inefficient storage capacity utilization due to limited automated and transparent data migration tools
- Performance challenges due to bottlenecks resulting from over-consolidation of storage
- File system limitations including number of files supported, size of file system and individual file size
- Application disruptions due to data movement when adding storage or during technology replacement
- Lengthy data migration for technology replacement due to application and data interdependences
- Electrical power limitations resulting in consolidation to reduce power and cooling requirements
- Timely data protection including local and remote data backup and replication
- No insight into resource usage resulting in limited forecasts to avoid peak processing bottlenecks

For a storage management solution to be successful, it must deliver improved service in terms of availability, performance and compliance while enhancing the quality and timeliness of data protection to meet potential threats. Data governance and compliance issues include the preservation, retention, and security of data in line with current regulatory standards. Future technology acquisitions, ongoing maintenance and licensing costs are also important considerations when evaluating candidate storage solutions. Data survivability and accessibility need to be enhanced on a local and wide area basis to support server consolidation. A successful storage solution will address the proliferation of small NAS servers and the corresponding management headaches.

Leveraging NAS to address storage management challenges

Many techniques and technologies exist to eliminate IT challenges pertaining to data and storage management. NAS combines several techniques and technologies to deliver shared storage and data via a standard network interface and protocol as well as provide integrated storage management and data protection features. Scalable NAS solutions with enterprise class data protection and management features provide more than just storage and data sharing over a common IP-based network.

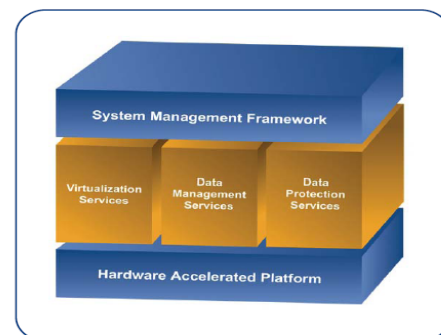
NAS should now be seen as a storage and data management vehicle to help address most storage and performance issues. Scalability means being able to increase capacity dynamically along with performance, including I/O operations per second (IOPs), bandwidth and reduced latency, without loss of availability or added complexity. For example, by using a storage system that scales performance in

conjunction with storage capacity bottlenecks that only increase storage capacity or the cost and complexity of adding additional individual file servers can be avoided.

Virtualization is another technology that can be integrated as part of a scalable NAS solution to address storage management challenges. For example, virtualization in the form of abstraction masks the complexities of underlying disk drive technology to support tiered storage, aligning the appropriate technology to application service level requirements. Abstraction, enabled via transparent data movement and migration across different storage tiers, reduces application disruptions and the impact of interdependencies that complicate and result in extended periods of time for data migration. Automated policy-based data movement can be combined with a scalable NAS storage solution to address complexities and cost associated with traditional storage management tasks.

BlueArc Titan enabling features for storage management

To address storage and data management challenges, including complete data protection, ease of use and scalability, the BlueArc Titan combines a robust hardware architected platform with a feature-rich software suite as a storage management solution. More than a clustered storage system that supports NAS (NFS and CIFS) and iSCSI block data of tiered storage, the Titan provides a platform to address and eliminate complexities associated with data and storage management. For example, by leveraging virtualization technology, applications are abstracted from the underlying physical attributes of tiered storage while relying on intelligent policy based management to automate data protection and storage management tasks.



BlueArc Software Architecture Framework

BlueArc Storage Management Framework Source: BlueArc

BlueArc Titan features that address storage and data management challenges include:

- Virtualization services – abstract physical resources and complexities from applications
- Data management services – policy -based automated data management for dynamic environments
- Data protection services – timely, robust local and remote data protection for DR and compliance
- General management features – flexibility, availability, performance, capacity and ease of use

Virtualization technology is present in four tiers (Figure-1) in the BlueArc Titan to address ease of use and mask underlying complexities from applications. For example, a virtual file system view is presented to users via a global or clustered name space (CNS), where multiple file systems can be presented in a unified view across BlueArc Titans. This means that different file systems, even different disk technology, simply appear as another folder in a unified directory structure.

To support transparent movement of file systems and workgroups, multiple virtual servers are supported on each physical Titan node. To shift the storage management paradigm toward a simpler, easy to use model, virtual servers can be used to move workloads around different physical nodes to meet specific performance and availability needs – very similar to the benefits of virtual machines in application server virtualization. To facilitate transparent data access, virtual storage pools optimize storage sharing and simplify storage allocation across tiered storage. Virtual servers and their associated file systems can be dynamically moved to different Titan nodes to meet varying peak workload or cyclical project lifecycles as needed.

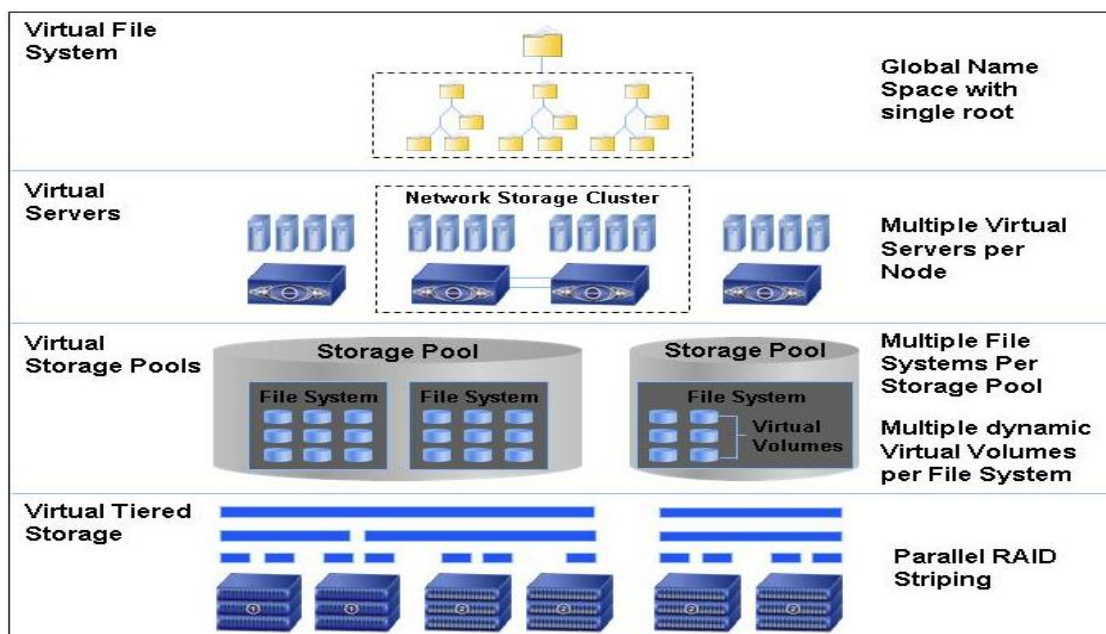


Figure-1: BlueArc Titan Storage Virtualization Features (Source: BlueArc)

Storage pools

Virtual storage combines underlying external, open storage based RAID sets into a logical storage pool of shared storage that is dynamically allocated to file systems. By leveraging a virtual storage pool, capacity is optimized while providing an abstraction layer between external hardware RAID sets and file systems, on stand-alone or clustered Titan storage systems, to simplify storage provisioning for applications and workgroups. Virtual storage is used to create virtual volumes, independent of underlying physical disk type, performance, location or capacity, to enable over-subscription (thin provisioning) of storage volumes.

File systems on the Titan can be subdivided into work groups to correspond with application specific project needs based on project or data lifecycle characteristics. Hence, BlueArc refers to these file system subgroups as “projects” as a convenient way for organizing as well as isolating user project and work files. For example, for software development and quality assurance, separate projects can be created to correspond to different release trains or builds. Another example is separate project to correspond to active files for digital video editing and another project for items destined for the “cutting room floor” to be archived for future use. Projects can also be used for email and general file sharing where current files are stored in active project and inactive files stored in an archive or compliance project, yet all are under the same file system for ease of access and management.

Using hard and soft quotas, storage volumes can project more data available for use than actually exists. For example, four separate 10TB projects or file systems could be created requiring a total of 40TB of storage capacity, however, only using 10TB of actual storage capacity. The Titan thin storage provisioning capability enables storage capacity to be dynamically expanded and an administrator notified when certain quotas are reached. This capability enables administrators the flexibility to buy the appropriate amount and type of storage when needed to support just in time storage acquisitions. File systems can start small and grow dynamically based on needs and policies instead of being over-allocated initially resulting in less than efficient storage usage.



In addition to improved storage capacity utilization, Titan storage pools also enhance performance by enabling administrators to define policies on how data is automatically load balanced across storage. By leveraging shared storage pools instead of dedicated storage associated to a specific filer or storage processing node, all storage I/O performance capabilities, including horizontal and vertical RAID striping, can be aggregated to support peak application processing demands and avoid bottlenecks. For example, a 10GByte database could be spread across storage pools independent of underlying RAID sets which, in turn, spread data across independent disk drives. Built from the ground up, first are the RAID sets for the horizontal raid striping then the vertical striping (parallel raid) via the storage pools and virtual volumes. The net effect is that such a database is spread across more disk drives (spindles) to support low latency, I/O intensive performance for OLTP applications.

The granularity of storage pools is such that an administrator can setup a large pool or many smaller pools depending on specific needs as well as assign many file systems to fan-in (many to one) to a single storage pool. In addition, administrators have the flexibility of having a single file system fan-out (one to many) to multiple storage pools, as well as many file systems access many storage pools (any to many). Basically, the Titan storage pools, virtual volumes, flexible file systems with dynamic thin provisioning, enable an administrator to tune the storage to specific application service and performance objectives.

Storage pools and virtual volumes enable:

- ✓ Easy and flexible storage allocation
- ✓ Improve storage use with shared pools
- ✓ Use your storage where and when needed
- ✓ Avoid dedicated storage on each server
- ✓ Support dynamic growth and load-balancing
- ✓ Ease of data access regardless of storage tier

Snapshots

To facilitate timely data protection and rapid recovery of data, BlueArc Titan storage systems support capacity efficient point-in-time (PIT) copy snapshots providing a cumulative history of data. To avoid performance penalties to applications, BlueArc Titan storage systems leverage robust hardware architecture combined with intelligent policy-based software for a highly efficient data copy operation. To support complete data protection, Titan snapshots are integrated with replication and backup policies to ensure data transactional and application integrity. For flexibility, snapshots can be instantly used for data recovery while providing different views of the file systems to match different recovery point objectives. As the snapshot capabilities are built into the hardware architecture, Titan can perform significantly more snapshots with less application impact than software-only based NAS filer systems. For example, Titan can support up to 1,024 policy based snapshots per file systems, and time granularity as small as every second for continuous data protection along with comprehensive data protection (CDP).

Local and remote data replication

Automated policy-based creation of one or more identical copies of data while maintaining synchronization with the source data is essential for timely business continuance (BC) and disaster recovery (DR) protection. Flexibility, data consistency and performance are critical attributes to consider in local and remote data replication technologies. BlueArc Titan storage systems support synchronous for local real-time and asynchronous for wide area data protection modes of operation along with options for full or incremental copies. A benefit of asynchronous copies is that it allows users to keep a copy of data on a lower cost tier of storage rather than the identical storage tiers required for synchronous mirroring which can be costly to implement. Additional features include preservation of NFS and CIFS file attributes, including permissions, as well as integration with Symantec/Veritas, EMC/Legato and CommVault for 3-way backup capabilities.



Accelerated Data Copy (ADC)

Data can be moved without affecting applications or users from one tier of storage to another, from one Titan file system to another, or moved to a different Titan system at a different location over a WAN or campus LAN. While Data Migrator leverages redirection technology to move data, ADC moves the entire file. ADC is also well suited for performing disk to disk (D2D) backups or implementing a policy to ensure that two discrete copies of data are maintained. Local ADC copies will leverage SAN with remote copy using standard IP LAN and WAN networks – SAN based local copies eliminate bandwidth impact on local users and applications. If local copies are done over the LAN, this can impact the user or application performance, increasing support call issues. Additionally, ADC supports incremental replication to allow for asynchronous mirroring of data, which ensures data protection while minimizing the time to copy data as only the changes are copied across the network. This can be done at the file or block level, depending on file size.

Use Accelerated Data Copy (ADC) to:

- ✓ Move files from one file system to another
- ✓ Data movement between different Titans
- ✓ Multiple copies of data for redundancy
- ✓ Mirror and replicate data for BC and DR

Supporting tiered storage and timely data protection

Transparent data movement and migration are the fundamental building blocks of automated and policy-based storage management. When it is unnecessary to migrate data except for data protection purposes, an enormous amount of day-to-day administrative tasks are eliminated, and it's a stepping stone to a new and simplified storage management paradigm. With BlueArc Titan's built-in Data Migrator, data movement and migration are non-intrusive to volumes, servers and file systems. For example, data can be migrated from one file system on one BlueArc Titan to another BlueArc Titan (local or remote), a file system moved from one tier of storage to another, a file system migrated from one virtual server to another virtual server, and a file moved from file system to file system.

This abstraction enables seamless data movement using policy-based rules to implement tiered storage, ILM and other timely data protection initiatives. Working in conjunction with integrated point-in-time (PIT) snapshot copies, local and remote replication, BlueArc Titans facilitate timely backups for rapid restoration of corrupted, infected, or accidentally deleted data along with supporting traditional data protection needs including business continuance and disaster recovery. For example, users have the ability to access snapshots to minimize the impact of administrators recovering accidentally deleted files.

Policy based data management services

Administrators can set up policies to automatically and transparently migrate data to different tiers of storage based on various rules, parameters and triggers. Based on rules triggered at specified times or events, such as at a certain percentage of storage utilization, policies invoke snapshots, remote mirroring, data migration and other functions on an individual or multiple Titans basis. BlueArc supplies a set of templates that users can fill in for their specific needs as well as create customer policies to automate storage management tasks. For example, if storage capacity utilization exceeds 75%, certain files could be migrated to a different storage tier. Alternatively, a weekly policy could be triggered to perform cleanup and migrate files not accessed in the last year to SATA based storage.

Data Migrator

Real-time, transparent data migration is used to leverage tiered storage to reduce storage costs, improve performance of primary storage, and support peak workload with better utilization of primary storage. For example, policies can be defined to transparently move data from high performance Fibre Channel disk drives to lower cost, higher capacity SATA.



The business benefit is that high performance storage capacity can be freed for those applications that need it while reducing storage costs. Unlike solutions that disrupt applications while data is moved back from SATA to Fibre Channel, BlueArc Data Migrator enables transparent, on-the-fly access to the migrated data. File header information remains on the high-speed disks with transparent redirection to data blocks stored on lower cost storage.

Use Data Migrator for:

- ✓ Maximize high performance storage capacity
- ✓ Real-time policy based data migration
- ✓ Automatic data movement for ILM initiatives
- ✓ Enable tiered storage for project files without having to change applications or notify users

This means that users and applications find their migrated files where they were originally placed; there is no re-pointing hard links or sending out emails with new file locations. Policies can be established to move data between storage tiers based on last access, last modified data, file name or type, file size and access path as well as various combinations of these. Titan enables administrators to run policy rules for “what if” scenarios for planning purposes, such as to determine how many files would be moved and space required, without actually moving any data.

Write Once Read Many (WORM) for compliance

BlueArc supports two WORM modes. One mode is non-strict and is good for staging and planning purposes because changes can still be made to data by the administrator while end users see the data as read-only. The other is a strict mode where changes to data are not allowed until the expiration of the retention period. Titans WORM capabilities can be used as part of an overall compliance solution.

Storage management challenge	Titan Feature to address storage management challenges
File system limitations	Support for large files, large number of files and large capacity file systems combined with thin provisioning dynamic allocation
Improve service delivery	Robust hardware with integrated software including virtualization capabilities and storage pools to support NFS file and iSCSI block access
Timely data protection	Snapshots, local and remote data replication using ADC, clustering, RAID and integration with popular backup solutions
Simplify management of multiple NAS appliances	Scalability (performance and capacity), tiered shared storage pools, thin provisioning, management and monitoring tools
Enhance performance and/or reduce storage costs	Data Migrator for transparent and automated intelligent data movement across storage pools to implement ILM tiered storage
Reduce power consumption	Optimized to support NAS and file server consolidation without incurring performance bottleneck or management complexity
Support data protection, BC and DR	Remote and local mirroring and snapshots along with automated policy based data movement local or remotely
Data protection	
Respond to dynamic performance needs	Virtual servers, CNS, clustering, hardware architected for performance, data migrator and storage pools
Application disruptions due when adding storage	Storage pools, virtual volumes, file systems, virtual servers, clustered name space and thin provisioning
Lengthy data migration	Data Migrator, ADC, snapshots, virtual volumes, storage pools, virtual servers, file system projects, robust hardware design

Table-1: BlueArc Titan features to address storage management challenges

Table-1 shows how various storage management challenges are addressed with the BlueArc Titan’s robust performance and advanced feature set, including integrated storage virtualization capabilities.



Conclusion

The BlueArc Titan 2000 series enables dynamic scaling of performance, capacity and data protection via robust features, including automated intelligent data migration, virtual servers to transparently shift workloads across physical servers, snapshots to protect data in a timely manner, local and remote replication to protect and isolate against system and site failures as well as to facilitate wide area data migration and consolidation for distributed data and NAS filers.

In addition to BlueArc's unique hardware capabilities architected for to support scaling of capacity and availability without compromising storage performance, the company's software offerings combine to address various storage management challenges. By leveraging storage virtualization management features, BlueArc Titan 2000 series of storage systems provide a resilient, scalable and flexible platform for supporting data infrastructures including file server and storage consolidation while addressing storage management headaches and challenges.

About the author:

Greg Schulz is founder and Sr. Analyst of the StorageIO group and author of the book "*Resilient Storage Networks - Designing Flexible Scalable Data Infrastructures*" (Elsevier / Digital Press) ISBN 1555583113.

All trademarks are the property of their respective companies and owners. The StorageIO group makes no expressed or implied warranties in this document relating to the use or operation of the products and techniques described herein. The StorageIO group in no event shall be liable for any indirect, consequential, special, incidental or other damages arising out of or associated with any aspect of this document, its use, reliance upon the information, recommendations, or inadvertent errors contained herein. Information and recommendations made by the StorageIO group are based upon public information believed to be accurate, reliable, and subject to change.