panasas™

PREDICTIVE SELF-MANAGEMENT

WHITE PAPER | NOVEMBER 2006

TABLE OF CONTENTS

A new level of reliability and availability for clustered storage to enable continuous performance through constant monitoring, intelligent prediction and pre-emptive action.

INTRODUCTION

The Panasas ActiveScale™ operating environment provides an object-based architecture that enables parallel data access by dynamically distributing data across a large number of intelligent storage blades. Panasas StorageBlades™ contain the data objects, while Panasas DirectorBlades™ control and monitor data access between clients and StorageBlades. This unique ability to perform parallel I/O makes the Panasas Storage Cluster™ ideal for large High Performance Computing (HPC) Linux® clusters that are running parallel applications on a massive scale.

Panasas Predictive Self-Management encompasses a number of key technologies that work together to deliver elements of health monitoring and self-healing, ensuring that the Panasas Storage Cluster continues to operate and provide maximum performance to the users and applications running on the cluster. These technologies can be grouped into four major areas: monitor and fix; monitor and alert; load balancing; and redundancy.

All of these technologies are interlinked and support the ActiveScale operating environment in providing the highest continuous performance with a level of manageability previously unseen in clustered storage. Panasas has extended its support to include interactive applications, providing a Unified HPC Storage Fabric across the entire technical computing workflow. This makes technologies like Predictive Self-Management even more important. Unified HPC Storage systems can reach Petabyte scale making reliability critical. Unlike typical SAN-based storage systems used in HPC environments, which have three separately managed components, the Panasas ActiveStor systems are fully integrated making Predictive Self-Management possible while dramatically lowering overall Total Cost of Ownership.

MONITOR AND FIX

ActiveScan and ActiveGuard are new capabilities available in the 3.0 version of the Panasas ActiveScale Operating Environment.

ActiveScan

ActiveScan is a Panasas technology that ensures continuous performance by monitoring the StorageBlades and their contents and taking action if a problem is detected. ActiveScan continuously monitors data objects, the RAID parity, the disk media and the disk drive attributes. If a potential problem is detected with the blocks on the disk, then the data can be moved to spare blocks on the same disk. By using statistical analysis of the disk drive SMART attributes, potential future disk drive failure can be predicted, and action can be taken to protect data before failure occurs. If a disk drive failure is predicted, then the data objects can be preemptively migrated to other StorageBlades, eliminating the need for reconstruction. A replacement StorageBlade can be inserted and the DirectorBlades will migrate data objects onto the new blade as part of their normal capacity balancing routine. If a StorageBlade should happen to fail while its contents are being migrated, then only the remaining portion of the blade needs to be rebuilt. This significantly reduces the reconstruction workload and maintains performance.

ActiveGuard

Panasas ActiveGuard encompasses two key high availability technologies: DirectorBlade failover and storage network failover. In the event that a DirectorBlade fails, all volume services are au-

tomatically transferred to a backup Director Blade. All DirectorBlades work in an Active/Active configuration, meaning that there are no "spares" required and all DirectorBlades are always fully engaged in managing the Storage Cluster activities. There is no loss of service, and the system continues to operate without interruption. All applications using the DirectFLOW datapath will be able to seamlessly continue data processing.

The Panasas ActiveStor systems are based on Gigabit Ethernet and iSCSI. There are two complete datapaths for all of the blades built into every shelf. When the ActiveGuard option is employed, the network can seamlessly failover to an alternate path and ensure continuous access from all applications.

These two key capabilities, combined with the standard object-based RAID capabilities in the base ActiveScale operating environment, enhance the already highly redundant Panasas Storage Solution to provide the highest operating availability to customers running their critical applications.

MONITOR AND ALERT

There are several features that are included in the ActiveScale Operating Environment that monitor nearly all aspects of the Panasas clustered storage and alert the administrator when appropriate.

Fans, Power Supplies and Battery

The condition of all elements of the shelf are carefully monitored and compared to predicted norms. In the event of attributes approaching or exceeding certain limits, an alert is sent to the administrator so that action can be taken before failure actually occurs. Examples include: fan not working, power supply about to fail, low battery, etc.

Temperature

The thermal environment inside the storage shelf and rack system is one of the greatest contributors to long-term reliability. This is why Panasas shelves are carefully designed to maintain correct operating temperature. The system monitors temperature and the rate of change of temperature so that appropriate action can be taken before limits are exceeded. Messages are sent to the administrator at pre-determined levels. If no action is taken and it is predicted that the shelf temperature will rise to an unsafe level, then the system will execute a graceful shutdown to protect data.

PanActive Manager

The Panasas ActiveStor Storage Clusters™ have integrated Web and command line (CLI) management interfaces. This system consolidates data from all of the blades to a single interface, whether it is a single shelf or Petabytes of storage, minimizing the cost of ownership. The PanActive Manager™ also manages security that allows the system administrator to provide consistent control access and permissions across all protocols including DirectFLOW™, NFS and CIFS and even control root access to the file system at the IP address level. All monitoring and control features are available via the Web graphical user interface (GUI) or the CLI eliminating the problem of remembering which interface to use for which functions. The GUI contains a full set of storage resource management (SRM) tools and reports including historical performance reports, asset reports and historical and predictive storage utilization reports.

Hot Client Monitor

The PanActive Manger continuously monitors client and application demands on the Storage Cluster. It maintains a continuously updated table of the most active clients using the DirectFLOW or NFS protocols. This allows the system administrator to detect runaway jobs or redistribute application loads across more client nodes in the cluster.

SNMP

Alerts can also be configured to be sent to the administrator via SNMPv3. SNMP is used by the Storage Cluster to issue alerts (traps) and to allow queries for system status (gets), but for security purposes, does not allow changes to the configuration (sets). This is particularly valuable for environments where the data center has a single point of performance or failure monitoring across all systems.

LOAD BALANCING

Being able to proactively balance storage resources for maximum performance and data accessibility is one of the most important aspects of the Panasas ActiveScale Operating Environment. Below are several features that address this critical need.

Passive Object Balancing

The highest possible performance in a parallel architecture is available when the disk drives are evenly accessed. The ActiveScale operating environment manages data placement across all of the StorageBlades in a Storage Cluster. A service called the Performance Manager in ActiveScale maintains a table of the load on each StorageBlade that is a combination of the performance demands and the capacity utilization. When a client requests to write data to a system, the DirectorBlades use this table to determine the optimal set of StorageBlades to write the data. Thus, the more the system is used, the more balanced it becomes, improving performance and avoiding hot spots on the blades.

Active Object Balancing

Passive object balancing may not be able to fully balance the system in the event that new capacity is added or replaced. In this event, the ActiveScale operating environment will use the Object table created by the Performance Manager to decide which objects need to be moved to achieve better load balancing across the StorageBlades in the Storage Cluster. The DirectorBlades then move the objects to the appropriate blades under the direction of the Performance Manager.

Read Coalescing

As demands are placed on the storage system, the StorageBlades will gather read requests so that they can be executed in the most efficient order, achieving the best utilization of StorageBlade cache. This has the effect of maintaining performance so that optimal performance is achieved in workloads where many clients are requesting the same data, which often occurs when parallel applications are started.

Reconstruction Load Balancing

While ActiveScan reduces the likelihood that a RAID reconstruction will be required, in the event of an unanticipated disk drive failure, the data is secure because of the RAID protection, but the missing objects must be rebuilt. Reconstruction load balancing enables the rebuild workload to be shared among all the DirectorBlades in a Storage Cluster so that the reconstruction occurs in a substantially shorter time period than a conventional RAID rebuild. This is particularly important with the ever-increasing density of disk drives where traditional RAID controllers can take a full day to rebuild a 500 GB disk, when an ActiveStor cluster will complete the task in 90 minutes. This significantly reduces the time when the data is unprotected and effectively minimizes the performance overhead normally associated with a rebuild.

REDUNDANCY

Panasas ActiveStor Storage Clusters provide redundancy across many levels to ensure maximum data accessibility.

RAID

Panasas object-based ActiveRAID™ will intelligently and automatically select RAID 1 or RAID 5 depending on the size of the data file within the same virtual volume. System users and administrators no longer need to be concerned about optimizing RAID layouts, increasing efficiency and lowering cost as compared to traditional systems that require administrators to create separate volumes for different RAID layouts, and for users to manage their applications to take advantage of them. With ActiveRAID, small files are mirrored to provide the greatest performance; larger files use RAID 5 to be efficient with space. The data is striped across multiple blades for maximum performance. The PanFS parallel file system has a unique intelligent reconstruction capability; spare disk space is distributed across all the StorageBlades and all those blades work in parallel to accelerate any required reconstruction. This significantly reduces the time to rebuild compared to traditional RAID systems.

Fans

Three variable speed fans are built into each shelf in an ActiveStor system. In the event of a fan failure, the surviving fans will automatically be accelerated to maintain the best thermal environment. Notification will be provided to the administrator so that the failed unit can be replaced. Whenever a blade, power supply or battery unit is removed from the shelf, the shelf automatically seals the opening to maintain a strong, well-directed airflow.

Power Supply Units

Dual redundant, hot-swap power supplies ensure continued power to each storage shelf. If one power supply develops a problem, the other can take over and provide full power to the shelf. An alert is sent to the administrator with the shelf ID so that the failed unit can be replaced.

Battery Backup

Battery backup for the entire storage shelf ensures data integrity in the event of an unplanned power loss. This built-in UPS allows the Storage Cluster to aggressively use caching to increase storage

performance while guaranteeing that data is always written to stable storage. There is sufficient power in the battery module for data in all caches in the system to be written out to the disks. The shelf will then execute a graceful shutdown. The ActiveStor system will periodically check the battery to make sure that it is maintaining a sufficient charge and notify the administrator when it needs to be replaced.

ActiveScale Mirror

The ActiveScale operating environment on each DirectorBlade or StorageBlade is mirrored so that in the event of a problem accessing one copy, the alternate copy can provide the required files without the need for data reconstruction. This provides high availability for the ActiveScale operating environment on each blade, maximizing availability of the blades and the ActiveStor system.

ActiveUpgrade

All storage systems require updates to their operating environments to add new features or eliminate defects. Panasas ActiveUpgrade enables upgrades to be made to the system without disruption to applications or users. Upgrades are made available to Panasas customers on their personalized MyPanasas Service Portal. These upgrades can then be applied to all of the blades in an ActiveStor system in parallel as a background task and updated without interrupting application processing.

SUMMARY

Panasas ActiveStor Storage Clusters offer a fully integrated hardware and software storage solution, featuring the industry leading bandwidth and random I/O performance. The proven PanFS parallel file system is based on a highly scalable object-based architecture, which provides superior performance, scalability and manageability.

Panasas Predictive Self-Management is a collection of key technologies that work together to provide high availability, resilience and continuous performance for HPC Linux cluster environments. Predictive Self-Management eliminates many of the typical tasks that must be performed by system administrators in traditional storage architectures, which lowers the cost of ownership and increases the number of terabytes that can be managed by a single administrator.

Panasas offers comprehensive storage solutions for all HPC environments with highest bandwidth and I/O performance; scalable performance and capacity; and simple unified storage cluster management. The Panasas integrated hardware and software storage solution is proven in hundreds of Linux cluster installations around the globe. To find out more, contact us today at www.panasas.com or 1-888-PANASAS.

panasas™

Accelerating Time to Results™ with Clustered Storage

6520 Kaiser Drive Fremont, California 94555 Phone: 1-888-PANASAS Fax: 510-608-4798 www.panasas.com