Simplifying Networked Storage™

PeerStorage™ Architecture Whitepaper

Fundamentally Altering the Economics of Shared Storage

9 Townsend West
Nashua, NH 03063
1.603.579.9762

www.equallogic.com or www.peerstorage.com

Introduction

Consolidated storage promises many benefits: virtualization, reliability, high availability, and centralized management. Although these promises have yielded interesting technologies, they do not meet the need today for affordable, easy-to-use products and are impractical for most environments.

To solve the business problems associated with storage configuration, scaling, and management, EqualLogic™, Inc. has architected a technology that provides a cost-effective, self-managing alternative to expensive and complex offerings — bringing the benefits of consolidated storage to a broad range of markets.

The PeerStorage™ Architecture fundamentally alters the economic and management landscape for consolidated storage by combining proven operating system, storage, and networking technologies with patented, intelligent automation for a true state-of-the-art storage system.

This blend of technologies makes consolidated storage a cost-effective reality for all operating systems and applications. The PeerStorage Architecture uses page-based volume management between arrays and incorporates high-speed, standards-based technologies such as Gigabit Ethernet, iSCSI, and Serial ATA. The result is an affordable, scalable storage solution that grows easily from 100s of gigabytes to more than 100 terabytes, while management stays minimal and constant.

While the industry talks about "next-generation storage," EqualLogic is already there.

The PeerStorage Architecture was designed to meet several goals:

- Provide a single, shared storage service to hosts and applications.
- Scale seamlessly from small to very large configurations
- Automate repetitive management tasks:
  - Perform RAID configuration and sparing.
  - Provision new resources in the most efficient manner.
  - Allocate resources to meet changing workload demands.
- Comply with industry standards.
- Require no special host software for access and management.
- Deliver uncompromised availability and serviceability.
- Offer a full set of features, including capabilities that are usually available only as expensive add-ons, like snapshot and volume replication capabilities.
- Provide a cost-effective, easy-to-use replacement for direct-attached storage (DAS).

This whitepaper gives a brief overview of the technologies that make the PeerStorage Architecture uniquely able to address the issues associated with consolidated storage.

PeerStorage Architecture Fundamentals

To achieve the goal of providing an affordable, self-managing consolidated storage solution, the PeerStorage Architecture employs three essential elements: the PeerStorage array, the PeerStorage group, and volumes.

PeerStorage Array

The foundation of the PeerStorage Architecture is a PeerStorage array — a high-performance block storage device that is ideal for departmental or enterprise storage. Each array is a fully-redundant storage array containing disks, multiple high-performance network interfaces, controllers with mirrored battery-backed caches, and other advanced features. The disks in an array are automatically protected with RAID (RAID 10 or RAID 50) and hot spares.

PeerStorage Group

A PeerStorage group is comprised of a single PeerStorage array or hundreds of arrays. The group is seen as a single, shared storage service by client servers and administrators. When an array is configured as a group member, its RAID-protected disk space is added to the group's storage pool. Data and network I/O to the group is load balanced across the group members' resources.

Volumes

Administrators create volumes from the available space in the PeerStorage group storage pool. A volume can be spread across multiple disks and multiple group members. The group exports volumes as iSCSI targets protected with security, including authentication and authorization, for both discovery and access. Upon connection, hosts see volumes as local disks. Volume snapshots and volume replicas can also be created from the storage pool.

PeerStorage Groups Facilitate Storage Management

Although comprised of one or more PeerStorage arrays, a PeerStorage Group appears to client servers as a single entity that offers network storage access in block mode — a storage area network (SAN). Unlike conventional SAN products, each group member "cooperates" with other members to automate resource provisioning and performance optimization.

The disk data served by a group is dispersed across the group members, with the placement continually adjusted for optimal performance as resources are added or load changes occur. In addition, client network sessions are distributed across members to spread out the I/O requests.

The result is a storage system that scales linearly in all significant dimensions — storage capacity, network bandwidth, client request processing capacity, and performance. In contrast, administration remains stable as the group is expanded. There is no separate management of arrays and volumes — all administration is performed at the group level.

With its ability to export many targets, a PeerStorage group can be used to consolidate storage for a wide range of applications and servers, including e-mail, Web servers, and database servers, without compatibility or performance concerns and without a prohibitive increase in administrative expenses.

Distributed Page-Based Volume Management

The RAID-protected disk space that each member contributes to the PeerStorage group's storage pool is partitioned into fixed-sized chunks of data, called pages. Each volume has a page map that allocates and re-allocates pages to the members. A page map is shared by each member that contains a page of the volume. Page maps are also used in the implementation of features like volume snapshots, replication, and over-subscription.

Page movement is a transactional operation between group members that ensures data integrity. Pages can be moved transparently (that is, online and without blocking host access) across disks in a member and across group members for automatic load balancing and to ensure data availability. For example, to eliminate a frequently-accessed data "hot spot," the affected pages may be spread across different disks or group members, without the need to move all the pages for an entire volume. The built-in load balancing features use real-time and historical access trends, as well as capacity and status information, to continuously improve performance within individual arrays and the group as a whole.

A lookup and transfer mechanism is used to efficiently track and process pages located on the group members. When a PeerStorage group receives a client server request, it identifies the location of the data and transfers the request to the member that contains the data. A request may be transferred one more time if the desired data is located on another member.

The entire page lookup and transfer operation increases latency time by less than one percent. In addition, the PeerStorage Architecture's optimization technology continually adjusts storage placement and network connections to minimize routing.

Fast, Simple Provisioning and Scaling

Using the scalability model from the PeerStorage Architecture, a PeerStorage group can scale linearly in both capacity and performance — all while online. The model allows for one or hundreds of members in a single group, with a controlled and efficient addition of resources as new members are added.

Individually, each member is a fully-functional, high-performance, highly-available storage array with mirrored write-back caches and multiple storage network connections. Resources like disks, controllers, caches, and network connections can be easily added to and removed from a group with no complex administration tasks or impact on availability.

As capacity and performance requirements increase, a group can be expanded. New members "learn" configuration and performance information from the group — with no manual intervention. Data and client connection load balancing occur automatically as the group scales. Page activity is monitored, and data and network connections are adjusted as needed.

The scalability model allows for automated, online expansion in all storage dimensions, and the PeerStorage Architecture nearly eliminates downtime caused by expanding or managing a storage system. Because capacity can be added so easily, IT managers need to buy only the storage necessary for today's applications, easing budget constraints caused by excessive and superfluous purchases.

Table 1 shows how the scalability model of the PeerStorage Architecture compares to traditional DAS and SAN solutions.

Table 1: Storage Expansion Comparison

Expansion activity	Traditional DAS and SAN	PeerStorage Architecture
Add storage capacity or network bandwidth	Limited expansion capabilities are available.	Add another member to a group; data remains available.
Increase network connections	Requires repetitive manual tasks.	Minimal administrative effort is required.
Add controllers	Storage must be taken offline; data is not available.	Expansion occurs online; data remains available
Perform load balancing	Requires repetitive manual tasks; data may not be available.	Operation is fully automated; data remains available.

Increased Availability

Consolidated storage increases the need for availability in the storage infrastructure. The availability model used by the PeerStorage Architecture eliminates all single points of failure and enables an array to survive multiple, simultaneous failures:

- Each PeerStorage array is composed of redundant components — disks, controllers with mirrored write-back caches, network interfaces, power supplies, and cooling fans.
- All hardware components are hot swappable.
- Disks are automatically configured with RAID, and hot spares are reserved to ensure that no data goes unprotected.
- Hot sparing and data recovery require no user intervention.
- Volume replication can provide site-level disaster protection.
- Group member hardware upgrades can be done while a group is online; simply add the upgraded array to the group.
- A member can be decommissioned while a group is online; simply remove the array from the group.

In addition, when upgrading to the latest technology, IT managers often find themselves writing off relatively new equipment because it is not compatible with new acquisitions. The PeerStorage Architecture eliminates this problem by ensuring that previously-purchased hardware and advanced technologies will remain interoperable.

Automated Management

The PeerStorage Architecture is designed to simplify storage management in several ways. RAID configuration and hot sparing is automated, and dynamic storage and network I/O load balancing occur as resources and performance metrics change. No longer must administrators manually map application data to specific physical devices and controllers.

For example, to create a volume from a group's storage pool, an administrator supplies only the name and size. The group members handle all details of storage allocation and load balancing.

Because a group is seen as a single entity, an administrator has a centralized view of the storage; management tasks remain constant, regardless of scale. A group can be managed through several built-in mechanisms, including SNMP, serial line, telnet, and Web-based user interfaces. No external management station or management software is required, eliminating the need for additional purchases and administrator training.

In addition, a group can alert responsible individuals to management activity or problems through log files, SNMP traps, and e-mail notification methods. Finally, data and management access is protected with authorization and authentication mechanisms not found in today's SAN solutions.

Advanced Features

With traditional SANs, many advanced features are typically expensive add-ons in. For example, conventional SAN solutions require server licenses or additional fees for snapshot and replication capabilities. Solutions developed with the PeerStorage Architecture include these features at no extra charge.

In addition, the PeerStorage Architecture's snapshot capability exceeds the snapshot features available on the market today. The real-time synchronization of snapshots across multiple volumes occurs without interrupting applications. Snapshots allow administrators to perform online backups and can be scheduled at regular time intervals. If data loss occurs, archived information can be rapidly retrieved to restore data and return to normal operations.

The storage replication capabilities provided with the PeerStorage Architecture also enable administrators to replicate storage to a separate location and enhance disaster recovery procedures.

Leverages Industry Standards

Each PeerStorage array uses standard Ethernet networks and industry-standard iSCSI protocols. An array has multiple Gigabit Ethernet network interfaces, and each interface performs at 124 MB/second, full duplex. This design lowers purchase and connection costs and — unlike Fiber Channel solutions — greatly expands the number of administrators familiar with the technology that forms the basis of the PeerStorage Architecture, thus keeping the training required to a minimum.

Any Gigabit Ethernet switch can be used with an array to ensure support for most existing storage network infrastructures. By placing the intelligence in the storage arrays instead of the switch, storage and network load balancing occurs in the most efficient and cost-effective manner.

In addition, an array includes Serial ATA disks — the new generation of disk drives — which provide enhanced performance and reliability. Each drive has a dedicated 150MB/second channel into the array. The drives are hot swappable and come in a variety of models optimized for performance and capacity. The advantages of Gigabit Ethernet and Serial ATA are many, and their feature sets and performance will continue to grow.

The PeerStorage Architecture also can be used with all leading operating systems, applications, and management frameworks. Any standards-compliant iSCSI initiator can access a PeerStorage Group by using software implementations included with major operating systems or by using dedicated iSCSI host bust adapters provided by leading storage vendors. Thus, a solution is available for any environment and for every cost or performance need.

Table 2 describes how solutions based on the PeerStorage Architecture compare to traditional DAS and SAN solutions.

Table 2: PeerStorage Architecture Solves Business Problems Facing IT Managers

	Direct-Attached Storage (DAS)	Traditional Storage Area Networks (SAN)	EqualLogic PeerStorage Solution
Budget	Must plan for both initial and future growth.	Must plan for both initial and future growth.	Buy only what you need now. Can easily expand at any time.
	Lowest acquisition and management cost (for a single server).	High acquisition and management (people and tools) costs.	Low acquisition and management costs.
	Scaling results in repeated purchases and higher management costs.	Scaling results in expensive infrastructure upgrades.	Lowest scaling costs.
Resources	Each storage component separately deployed; not a shared storage solution.	Shared storage solution; separate storage network infrastructure and tools.	Shared storage solution; utilizes familiar network infrastructure and easy-to-use tools.
	Dedicated.	Many interoperability issues.	Complete Ethernet networking interoperability.
	Upgrades and maintenance cause server and application downtime.	Upgrades and maintenance cause server and application downtime.	Deploy storage on demand and easily change configuration online; upgrade and perform maintenance online.
	Poor utilization of capacity and performance.	Good capacity utilization and performance.	Automated capacity utilization and performance optimization.
People	Each storage element is managed individually.	Each storage element is managed individually on a network.	Managed as single storage system, regardless of scale.
	Repetitive, manual management tasks in dissimilar locations.	Repetitive, manual management tasks.	Automation dramatically reduces management time and costs.
	No specialized knowledge and experience required.	Highly specialized knowledge and experience required.	No specialized knowledge and experience required.
Time	Management effort grows as number of servers grows.	Management effort grows slowly as number of servers grows.	Management effort remains the same as number of servers grows.
	Short training time.	Long initial training time.	Short training time.

Conclusion – Self-Managing, Affordable Storage

By allowing arrays to work collaboratively in PeerStorage groups and by automating time-consuming manual tasks, PeerStorage Architecture offers a fundamental change in SAN implementation and operation. The PeerStorage Architecture significantly lowers the cost of consolidated storage without a loss of capabilities. The result is a full-featured, self-managing SAN that includes dynamic load balancing, hot swapping, hardware redundancy, snapshots, and volume replication.

Storage acquisitions are driven by current application needs, not a prediction of future needs. The PeerStorage Architecture represents a dramatic advancement in storage economics, from purchase and set up through operation and upgrades. By eliminating the barriers to adopting consolidated storage — that is, non-standard and expensive hardware, complex technology, high staff skill requirements — the PeerStorage Architecture provides storage solutions that can quickly grow from hundreds of gigabytes to hundreds of terabytes, using existing resources and familiar skills.

About EqualLogic

The PeerStorage Architecture is a breakthrough in the process of buying, managing, and growing SANs. Affordable, self-managing storage solutions from EqualLogic use iSCSI and Serial ATA technologies with patented PeerStorage intelligent automation to dramatically reduce storage complexity and operating costs for businesses in need of consolidated storage.