Snapshots, Replication and iSCSI: A Practical Guide
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction/Overview</td>
<td>3</td>
</tr>
<tr>
<td>The New Reality -- Operations Must Be Online</td>
<td>4</td>
</tr>
<tr>
<td>What Constitutes a Disaster?</td>
<td>4</td>
</tr>
<tr>
<td>Data Protection - One Part of the Insurance Policy</td>
<td>5</td>
</tr>
<tr>
<td>A Continuum of Disaster Preparations</td>
<td>5</td>
</tr>
<tr>
<td>Do IT Over</td>
<td>5</td>
</tr>
<tr>
<td>Tape Backup/Restore</td>
<td>5</td>
</tr>
<tr>
<td>Tape-Based Backup/Restore with Replication</td>
<td>6</td>
</tr>
<tr>
<td>Recent Backup Improvements</td>
<td>7</td>
</tr>
<tr>
<td>Backups Suffer from Lack of Application Coordination</td>
<td>7</td>
</tr>
<tr>
<td>Improve Backup by Leveraging Disk Image Copies (Replication)</td>
<td>8</td>
</tr>
<tr>
<td>Snapshots</td>
<td>8</td>
</tr>
<tr>
<td>Snapshots Cover Many Failures</td>
<td>8</td>
</tr>
<tr>
<td>Continuous Replication – Powerful Yet Limited</td>
<td>9</td>
</tr>
<tr>
<td>Auto-Replication Covers All the Bases</td>
<td>10</td>
</tr>
<tr>
<td>Example -- EqualLogic Auto-Replication Implementation: Easy, Fast, and Efficient</td>
<td>11</td>
</tr>
<tr>
<td>Network and Storage Efficiency</td>
<td>11</td>
</tr>
<tr>
<td>Multiple Recovery Points</td>
<td>11</td>
</tr>
<tr>
<td>Instant Recovery</td>
<td>11</td>
</tr>
<tr>
<td>Standard Feature</td>
<td>12</td>
</tr>
<tr>
<td>Standard Networking</td>
<td>12</td>
</tr>
<tr>
<td>Delegated Management Model</td>
<td>12</td>
</tr>
<tr>
<td>No Host Software</td>
<td>12</td>
</tr>
<tr>
<td>Simple Management</td>
<td>12</td>
</tr>
<tr>
<td>Improves Backup</td>
<td>12</td>
</tr>
<tr>
<td>Automatic or On Demand</td>
<td>12</td>
</tr>
<tr>
<td>No Disk Reconfiguration</td>
<td>13</td>
</tr>
<tr>
<td>Flexible Site Configuration</td>
<td>13</td>
</tr>
<tr>
<td>What Can You Do with Auto-Replication?</td>
<td>13</td>
</tr>
<tr>
<td>Traditional Disaster Recovery</td>
<td>13</td>
</tr>
<tr>
<td>Centralized Replication &amp; Backup</td>
<td>13</td>
</tr>
<tr>
<td>Shared Resources</td>
<td>14</td>
</tr>
<tr>
<td>Moving Data Online</td>
<td>14</td>
</tr>
<tr>
<td>Summary</td>
<td>15</td>
</tr>
<tr>
<td>About Mosaic Technology</td>
<td>15</td>
</tr>
</tbody>
</table>
Introduction/Overview

Data protection has always been a fundamental responsibility of IT. Now more than ever business is demanding accessible, recoverable data. Pressures from recent disasters such as Hurricane Katrina and the Tsunami combined with a proliferation of regulatory demands makes offsite protected copies and readily accessible backups high priority.

On the surface these are mutually exclusive – or at the least – very expensive to implement. Which is what makes iSCSI so interesting and appealing. Ethernet connectivity is everywhere. It’s easy to use and there’s a common knowledge base in most companies. Using Ethernet as a backup transfer method makes sense – if you do it right.

That’s where snapshots and replication come in. Large financials institutions already employ some form of continuous replication. They cannot afford to be off line at all. However, continuous replication is not a substitute for backups. File of data corruption from any source impacts both copies in continuous replication.

iSCSI data arrays that include snapshot technology and auto-replication take advantage of low cost Ethernet to provide robust, flexible backup alternatives to local tape drives and fibre channel replication.

Better business continuity solutions are needed - to improve backup, speed restore, and reduce lost work. Only remotely replicated, disk-based image copies offer complete data protection and rapid recovery, but historically this function has only been available as an expensive software add-on. Today, iSCSI makes automatic remote replication fast, easy, and affordable.
The New Reality -- Operations Must Be Online

No matter what your organization does or what size it is, business continuity matters. Operating 24/7 is now the expected norm as companies do more and more business over the Web, around the world.

In the not so distant past, backup windows existed at night when customer and supplier interactions ceased. Today, your customers and suppliers may be anywhere in the world - awake and transacting business in a different time zone. Your business needs to operate 24 hours a day, whether serving customers or maintaining internal systems. The luxury of nightly downtime is long gone, replaced with demands for high service levels at all hours.

Consequently, business continuity is a major strategic concern for most organizations. You must have a clear "disaster recovery" plan in place to restart business within an acceptable timeframe. The plan should include key categories - buildings, equipment/infrastructure, people, and data.

What Constitutes a Disaster?

When most of us hear "disaster," we think of a devastating fire, flood, earthquake, or other dramatic event. While infrequent, these events do occur and you must be prepared for them. In terms of data protection, you're safe from these events as long as a recent copy of your data is stored a safe distance away.

But in the business world, a disaster can be any kind of interruption - a software malfunction, virus attack, database corruption, accidental file deletion, a network outage, or hardware failure. These occur more easily and frequently than catastrophic physical events and require you to implement disaster recovery plans such as replacing equipment, recovering data, restarting computer systems, and redeploying staff. For these events, current data is not enough, since it is likely to be corrupted and useless. You need multiple data copies at several different points in time to ensure clean data for business restart.

The challenge facing most organizations is how to minimize the lost revenue and operating expenses disasters cause. Even without losing buildings and equipment to hurricanes or floods, organizations suffer in many ways, including:

- Loss of employee productivity
- Loss of customer business during downtime
- Loss of customers who, when unable to do business with you, find another vendor - permanently
Financial penalties from violation of legal requirements

Failure to recover from a disaster has lead to the demise of many organizations, and threatened the existence of others. That's why companies take business continuity so seriously - and why a disaster recovery plan is critical.

**Data Protection - One Part of the Insurance Policy**

Without business data, disaster recovery is impossible. A complete data "insurance policy" is essential. Insurance starts with a regular process for data backup - creating and safely storing copies of data - which is the core of business continuity.

Everyone does backup. And until recently backup was painful - time-consuming, complex, and awkward - but a requirement. Data protection also means finding a way to recover data and restart business if a fire destroys your storage equipment, or a software virus infects your systems, or an application upgrade corrupts data. Improved backup and disaster recovery systems - a better insurance policy - would take significant burden from your IT staff. But it would need to be affordable and automatic to really make a difference, and guarantee protection from all kinds of physical and digital disasters.

**A Continuum of Disaster Preparations**

Data recovery and business restart depend completely on what you do before the disaster. Do you have the right technologies to deliver the results you need? Do you have the proper plans and processes in place?

For most organizations, key metrics of recovery solutions are "How long does it take to get running again?" and "How much completed work do I lose?" Most companies develop goals for acceptable amounts of lost time and data, known as recovery time objectives (RTO) and recovery point objectives (RPO). The next section reviews common recovery methods, and their advantages and disadvantages.

**Do IT Over**

If you have no backup capability, recovering from a failure requires doing the work over. Lost files, databases, and transaction records need to be re-created. You have no data protection "insurance," so it costs you nothing prior to the loss. However, it may be weeks or even months – if ever -- before you are up and running again and you lose an awful lot of work. You don't pay upfront - but you pay dramatically after a failure. This is unacceptable for business today.

**Tape Backup/Restore**

The most common data protection approach is tape backup. Disaster protection/recovery is accomplished by creating additional copies of data, saving them to tapes, and transporting
these copies away from the primary data regularly.

In the event of a disaster, tapes are brought to functional equipment, and data restored by backup software. By creating physical distance between the primary data and the tape copy, you protect against physical disasters. This dramatically reduces lost work compared to "do IT over." Operations can be restored much more quickly - typically in hours to days, depending on your procedures. You now have the operating expense of running backup operations regularly; and safely storing the backup copies away from the primary systems. Your cost is higher, but you have improved your data protection insurance.

**Tape-Based Backup/Restore with Replication**

This method involves making disk-based replicas of your data at various intervals. If your system fails you have data available to restart in minutes to hours, instead of hours to days required with tape. You can also use replicas as a source to create your backups, freeing up production systems from backup processing. Your upfront costs are higher, but you have better insurance and take much less time to recover.

In selecting a recovery method, you trade off cost of insurance vs. cost after failure. You may get better insurance if you pay more. However, replication technologies vary in costs and benefits, and not all offer the same results. Paying more does not necessarily mean you are getting the best protection.

**Backup - Core of Business Continuity**

Today tape backup remains the basic foundation of business continuity. Backup accomplishes several things:

- Creates copies of data
- Makes copies in "backup format" on a transportable tape medium, so it can be moved off site
- Catalogs tape contents, so when you need to recover from the tapes, you can see what data was saved, and when
- Delivers a method for restoring data

Backup offers critically important benefits. As long as backups are done correctly and consistently they provide:

- Disaster recovery - backup operations provide multiple recovery points such as "yesterday," "last week," "last month," "last year." This means that data corruptions or other unexpected changes can be recovered by using backups from a time prior to the data loss.
- A historical record, or archive, or corporate data
- Proof of regulatory compliance
- A legal record of business operations

Backup is the most affordable data protection solution, and covers all types of data loss - the common types (hardware failure, software corruption, virus, human error) as well as the uncommon types (natural disaster/physical destruction of site or equipment). It may not be the most glamorous part if IT - but organizations cannot survive without it.
Recent Backup Improvements

Backups are commonly done locally, across a LAN, or using a SAN. With local backup, each server backs up to its own tape drive. With network backups, a tape library connected to a single server replaces individual tape drives. Data is backed up over the LAN to that server. This consolidates backup operations. It requires fewer staff and usually fewer tape drives. It is more complex, but for IT it’s easier to manage and less intrusive.

SAN backups improve operations even more. By moving tape drives onto the SAN, each server sends data directly to the tape drive. The backup server manages the catalog and tape library – it does not move data. This lightens network load, offering better performance and faster backup.

Another recent backup improvement is backup to disk. Instead of writing directly to tape, the backup server writes to disk, and later uses the disk backup to send to tape. This improves the initial backup performance bottleneck that sometimes occurs when writing directly to tape, and lets you complete backups faster.

Backup Challenges
Backup is neither easy nor convenient. There are many operational and logistical challenges.

- Data is saved in backup format, and is not usable until converted back to its original format. Consequently, recovery is slow. Software must restore data and reprocess it on new storage to convert it into usable application format. This can be a lengthy process - hours or even days.

- Backups are an ongoing operational burden. They must be run regularly to be useful, occupying staff and impacting production systems. Running backups is like a large application workload running on your infrastructure daily’ It puts demands on servers, networks, and storage and must be factored into planning for each system. Performance degrades during backups, and application disruptions often occur.

- You must ensure compatibility of backup applications with heterogeneous infrastructure components. Backup is not easy to execute properly, and errors can corrupt your data. For live data, if applications are not properly prepared before backup begins, you cannot guarantee successful data recovery.

- After a failure there can be significant loss of work, because you will lose any data since the last backup. Most organizations aim for a nightly backup, which could result in 12 or more hours of lost work if a disaster strikes.

Backups Suffer from Lack of Application Coordination

The most common problem for executing backups is coordination with running applications. With open applications, open files, and data spread cross multiple disks, it is easy to make
errors. For instance, imagine you begin backing up Disk 1, then Disk 2. You have a database with a number of files that span Disk 1 and Disk 2. If one of those records is being changed while backup is executed, chances are good that the backup will not save the correct data. But you won't know for certain.

How do you solve this problem? First, you can stop applications while you run backups. This means downtime, impacts service levels, and is antithetical to business continuity. What about setting up your backup software to skip any files that are open? With databases running around the clock, that would mean they never get backed up at all. But these are the only options if you don't have replication capability.

**Improve Backup by Leveraging Disk Image Copies (Replication)**

Replication produces tremendous benefits, but various replication technologies offer different features, protections, and costs. There are three key replication methods: **Snapshots**, which are local image-based copies; **Continuous Replication**, including two kinds of remote mirrors, and **Remote Point in Time Replication**, which is essentially a Snapshot transported to a remote location.

**Snapshots**

A Snapshot is an instant copy of application data taken at a particular point in time. The copy is typically located with the primary data and shares storage with it. Snapshots can include complete copies (generated by breaking off a disk from mirrored disk sets), or just data that changed since the last Snapshot (using "copy on write" technology).

Some advantages of Snapshots:

1. They are created during normal operation of production volumes, and when well implemented, do not impact performance - they require no downtime, and do not hinder operations.
2. You can easily maintain multiple Snapshots, offering you multiple recovery points - for example, an hour ago, 4 hours ago, yesterday, or last week.
3. Snapshots can be instantly restored back to the volumes from which they were created. This image of disks is in application-usable format, so no data conversion is necessary, resulting in fast restores.
4. Snapshots can be "assigned" to another server and used for backup, testing, and other operations done in parallel, offloading processing from application servers.

**Snapshots Cover Many Failures**

Snapshots give you a "known good" copy of your data. They protect you against any outage.
or failure. You may want to take a Snapshot copy once a day or once an hour, depending on your needs, and use it to run backups from another server. By using a Snapshot backup times are irrelevant - your production systems are not involved or impacted. You still need to coordinate Snapshots with applications, but Snapshots can be created quickly - typically in under a few seconds. In addition, many applications are Snapshot aware, so this can be done online.

For example, take the case where you make Snapshots every hour. Now you’ve limited your risk/exposure - if you accidentally delete a file, you have a copy from no more than an hour ago. You can mount the Snapshot and recover that file immediately.

The same holds true for a software corruption or virus. When you discover that your data was corrupted four hours ago, you can go back to the Snapshot from five or 12 hours ago and know that you have a good copy of data. You've lost some data, but you can restore quickly.

Another advantage of this technology is to make a Snapshot of your data prior to installing a new system disk or software patch. Then perform the installation - if any problem arises, you can simply revert to the Snapshot and be back online instantaneously.

Using Snapshots together with tape backup reduces data you lose in a failure, dramatically reduces recovery time, and offloads application servers.

Snapshots also give you history. Although you are not likely to keep a year's worth of Snapshots as with backup tapes, you can easily keep multiple copies for a week or more.

The one vulnerability of Snapshots is that they are collocated with the primary data - any physical disaster that destroys the primary data will destroy the Snapshots as well.

**Continuous Replication – Powerful Yet Limited**

Continuous Replication is basically RAID, or disk mirroring, extended over distance. There are two kinds of Continuous Replication, **synchronous** and **asynchronous**. Both execute "continuous writes" to remote storage. With Continuous Replication, every time a write is done to the production disk, a write is done to the remote site. Therefore, the copy is identical to production data at every moment.

Continuous Replication minimizes data loss in the event of a physical disaster. However, they cannot be used for protection from common disasters since they provide no previous "known good" copy of data. If your production database is corrupted, you can't look to the replica because it is corrupted as well. To give yourself a historical copy and truly protect data you need backups. However, these replicas are not coordinated with applications -- they cannot be used to run backups.

Other weaknesses of synchronous or asynchronous Continuous Replication:
• They require low latency network links to support continuous writes. Your remote site must be within about 100 miles/150km of the production site to meet the latency and application performance requirements.
• Continuous Replication requires high bandwidth network links to keep both sides up to date. For both methods, if you write over the same block 100 times in an hour, you send 100 writes over the network to the remote storage.
• They offer no history or recovery points - if you have corrupted data at the production site you have corrupted data at the remote site.
• Application recovery takes longer than with Snapshots, because there is no coordination with the application.

These weaknesses combine to create an expensive solution that only protects you from some of your business continuity threats (physical disasters).

Continuous Replication is complicated to set up and complex to manage. Some large firms use Continuous Replication for a few applications, for example financial transactions that can never be even moments out of date. These firms also count on Snapshots, Remote Point in Time Replication, and tape backups because they know that Continuous Replication serves only part of their data protection needs.

**Auto-Replication Covers All the Bases**

Snapshots improve backup/restore and protect you from common disasters. Using network and storage technology, you can create Remote Point in Time copies, which are remote Snapshots. Snapshot technology gives you a replica from which to backup/restore and protects you from common corruption and human error failures, and the addition of physical distance protects from physical disasters. Usable data can be restored at the remote site, without having to convert backup data. This form of data protection is slightly more expensive than tape backups alone, because you need network links and storage at the remote site. But the advantage is clear - total protection and faster recovery. This method has all the benefits of Snapshots, without the weakness of collocation with primary data.

Automatic Remote Replication in iSCSI Disk Array
Most organizations need a business continuity solution that delivers better, faster, more storage- and network-efficient backup and disaster recovery, regardless of the kind of disaster.

iSCSI Disk Arrays combined with Snapshot technology offer just that level of better, faster, more efficient backup model.

The better iSCSI Disk Arrays include automatic Remote Point in Time Replication as a standard feature. Auto-Replication provides rapid, reliable backup and restore that...
ensures business continuity.

Remote Point in Time Replication delivers advantages of Snapshots with the additional protection of physical distance. Auto-replication can be performed in both directions - Site A can replicate to Site B, Site B can replicate to Site A, and they can replicate to each other. However, there is no need for a one-to-one correspondence - replication is between groups, not specifically between arrays. As a result, Site A may have five arrays and Site B may have three, and each can grow independently without impacting on the other.

Auto-replication uses "copy on write" technology to replicate data between storage groups and provide efficient operation over any type of network. With Auto-Replication, you create an image copy quickly at a remote location, and gain a stable data copy to use for backup or any other activities. These copies are network and storage efficient, don't hinder operations or performance, and offer multiple recovery points and quick restore. They are a practical, affordable, easy to use answer to the need for data protection that covers all kinds of functional failures.

**Example -- EqualLogic Auto-Replication Implementation: Easy, Fast, and Efficient**

EqualLogic is a fast growing iSCSI manufacturer. Their implementation of Auto-Replication is just one example of plug and play capability built into their PeerStorage (PS) iSCSI disk arrays. A number of features make Auto-Replication extremely useful to any organization seeking backup and disaster recovery improvements. These features focus on three underlying benefits: making replication easy, fast, and efficient.

**Network and Storage Efficiency**

Because Auto-Replication makes Remote Point in Time copies, replications after the first one can include only changed data, making the process both network- and storage-efficient. You won't be clogging the network with multiple writes or needing massive amounts of storage capacity in order to retain data copies.

**Multiple Recovery Points**

As with Snapshots, you can keep many differential copies for multiple recovery points. If you replicate a volume called Database (DB) at noon, 3 pm, and 6 pm, then DB(noon), DB(3pm), and DB(6pm) will store the differences from each other. They will use very little disk space, while looking to the administrator like three full copies.

**Instant Recovery**

Auto-Replication creates image copies on disk that are ready to use immediately no lengthy conversions from backup format. Like a Snapshot, the replica can be brought back online instantly - just at a different location.
Standard Feature
Because Auto-Replication is a standard, easy-to-use feature of an affordable iSCSI storage solution, you can start slowly and increase usage over time. No additional licenses are required.

Standard Networking
Auto-Replication uses standard TCP/IP and iSCSI so you can replicate over any IP network, over any distance, including T1, T3, DS3, or OCxx networks. No network devices or special equipment are required, just standard TCP/IP networking. As a result, with Auto-Replication you can continue to manage your network traffic in the same way you always have - no new management skills or tools are necessary. The flexibility, use of standard networks and procedures, and familiar network management make Auto-Replication simple to use.

Delegated Management Model
This solution is easy to set up, and uses a delegated management model that enables cooperation between administrators at different sites, while ensuring complete independence and security. Divisions of one organization that don’t normally work together can share storage capacity for data protection; similarly, an ISP that serves as a disaster recovery site for many organizations can more easily and efficiently delegate capacity with assurance of data protection.

No Host Software
Auto-Replication works in the storage array, leaving the servers to handle their own jobs.

Simple Management
Auto-Replication is managed and monitored through a single interface, preserving the architecture's ability to operate as a single unit.

Improves Backup
You can keep multiple replications for as long as you like, so they can act like backups - only much easier to access. If you use them like a backup, you make backup more efficient, offload backups from a server at the production site, and make restore much faster and simpler.

Automatic or On Demand
Replication can be done on demand or automatically according to a customer defined schedule, whether that is every 10 minutes, every hour, three times a day, weekly, or whatever suits your IT environment.
No Disk Reconfiguration
Administrators can turn replication on or off at any time without reconfiguring disks. You may decide to replicate some volumes that have contained data for two years you need only turn the feature on and select the volumes to replicate.

Flexible Site Configuration
Each location can configure its PeerStorage Group as needed by that location. There is no requirement that the locations be the same size, configuration or RAID level. Auto-Replication supports both RAID 10 and RAID 50 RAID levels.

What Can You Do with Auto-Replication?
Auto-Replication makes remote replication accessible and affordable for most organizations, whether their remote locations are across campus, across the country, or around the globe. There are many uses of this functionality, but a few are described below.

Traditional Disaster Recovery
In addition to improving backup, Auto-Replication delivers traditional disaster recovery. With PeerStorage groups at your production site and your remote site configured as replication partners, you can schedule regular replication at your disaster recovery site. This provides you with multiple recovery points that protect you from site failures as well as corruptions and viruses. If a virus strikes, you have several recovery points from which to restore data and quickly restart business operations.

Centralized Replication & Backup
Organizations with several locations can centralize disaster replication and backup. Instead of building complete backup/restore infrastructures at each location (including backup servers, tape drives, backup software, and staff), you can use one location for backup and replication.

For example, if Sites A, B, and C are production sites, you eliminate backup equipment at all sites and build it at Site D. Site A schedules replication to Site D at its convenience, as do Sites B and C. Tape backups are done at Site D from the Remote Point in Time Replicas, which are also available for quick restore should a disaster occur.
Auto-Replication also improves backups by enabling multiple recovery point replicas, offloading backup from production operations, and running backups from replicas.

Shared Resources
PeerStorage's delegated management model lets independent divisions of an organization share storage capacity for replication while remaining autonomous.

For example, a large manufacturing company has two distinct divisions - one builds helicopters, the other builds elevators. These divisions function completely independently - they have their own headquarters, data centers, staff, etc. With Auto-Replication between PeerStorage groups, the divisions agree up front to designate 500 GB of their storage capacity to the other division for replication. Once passwords are exchanged, each replicates freely, using the other site's data center as its remote site. The Helicopter Division storage administrator has complete control of his/her replicas in the Elevator Division's data center, but no access or privileges on the Elevator division's arrays. After the initial agreement, the staffs at each site do not need to communicate or work together - they function with complete independence and total security. Each administrator continues to manage his/her own SAN without interruption, and each remotely manages the replicas.

This approach offers greater security and cost efficiency. Now, organizations avoid building complete data centers for disaster recovery - instead, they leverage infrastructure at other locations within the company. They avoid large expenditures for buildings and equipment, data center infrastructure, staffing, and initial and ongoing expenses of facility management. Auto-Replication's security and delegated management model make this possible, allowing storage sharing without requiring interaction between storage administrators.

Moving Data Online
IT departments also use Auto-Replication to move data or business applications between data centers more safely and efficiently. By including Auto-Replication as a standard feature, even occasional use of replication is both easy and affordable.
**Summary**

Every organization needs a holistic business continuity solution for data protection. Auto-Replication protects you from all disasters, as well as improving backup. You cannot afford to skimp on this critical business insurance - but you don't have to buy and manage add-on software that only solves niche problems, either. With PeerStorage arrays from EqualLogic, you have a complete replication solution that is efficient, affordable, easy to manage, and delivers total protection.

**About Mosaic Technology**

For over 10 years Mosaic Technology has been bringing an objective and unbiased view of technology to our customers. In the areas of storage, backup, data management, and server technology we offer a cross section of leading technologies. We know that one size does not fit all. We have long term relationships with companies such as EqualLogic, Dell, EMC, Quantum, and Hitachi that enable us to give our customers solutions based on their specific needs rather than trying to force fit a product.