

Building a disaster-proof data center with HP Serviceguard for Linux



Introduction.....	2
Evaluating the need for disaster tolerance.....	3
What is a disaster tolerant architecture?	3
Understanding types of disaster-tolerant clusters	5
Extended distance clusters.....	6
HP StorageWorks Cluster Extension Software	9
Benefits of Cluster Extension.....	11
Differences Between Extended Distance Cluster and Cluster Extension	12
Conclusion.....	14
For more information.....	15

Introduction

Linux remains the fastest-growing operating system in the world for a good reason. It has enabled countless companies worldwide to transform their IT strategy, capitalize on business-critical functionality, reduce IT implementation and maintenance expenditures, and achieve more flexibility for competitive preparedness over the long term.

Today, many IT environments use Linux for an increasing number of disparate workloads—HP included. Not only has HP implemented the technology in its own enterprise, but HP is also leading the industry by adding value to Linux, providing a faster and better return on your IT investment. HP offers end-to-end open source and Linux solutions that are proven components of an Adaptive Infrastructure, in which business and IT are synchronized to capitalize on change. And with Linux running on standards based platforms, backed by services and support from HP, you can safely and confidently implement a UNIX®-like operating environment.

And, as you have come to expect from all of HP innovative products and services, HP enables Linux to perform as a core infrastructure component that will help your enterprise adapt to whatever the future might bring. Linux has become an important component in all areas of the data center, including those that are core to the business. CIOs cite business continuity as a major issue whenever they are surveyed about their main needs and concerns. Mitigating risk and reducing the business impact and costs of outages are ongoing concerns for business and IT executives. Significant technology and social and business changes over the past decade have heightened these concerns. During this time, more business and customer services have moved online. Back office computing no longer dominates; today business technology is on the front lines. In the past four years, consolidation initiatives have concentrated more IT services into fewer systems. Business technology trends such as virtualization accentuate those pressures even further. Moreover, business pressures have driven more business to deliver their services all the time, 24x365.

For many enterprises, the need to deliver constant IT services has become a requirement in many industries. These needs are driven by changes such as globalization, by online employee and supply chain IT services, and by online customer and customer support services. Taken together, these all drive the need for full-time 24x365 access to IT services. This specifically means that long recovery times jeopardize the health of the business. In these circumstances, enterprises must move from a recovery-orientation to a disaster-tolerant orientation. They make this move when the impact on the business is great enough to impact the top line, to impact customer loyalty, and impact brand reputation. Rather than just recovering their IT services, the enterprises must keep their business in operation during, and in spite of, all outages, whether caused by natural disasters, technology failures, or human error.

HP has been delivering disaster tolerant solutions to its customers for more than 20 years and introduced one of the first solutions on Linux. HP is the best vendor able to deliver a total and complete solution that embraces the full range of server and storage solutions available in the marketplace today. HP is not limited to just servers storage, one server technology, or one storage technology. No matter how stringent your technology needs and preferences are, HP can design and deliver complete disaster tolerant solutions to meet your requirements.

HP recently demonstrated these solutions and recorded a video that not only provides a clear description of the breadth, depth, and capabilities of HP solutions, but also shows a real to life simulation of how these solutions protect the business when a disastrous event takes place. To demonstrate the HP disaster tolerant capabilities on Linux, a streaming video application paused only briefly when the primary data center blew up which was protected by HP Serviceguard for Linux and HP StorageWorks XP Cluster Extension Software.

This white paper describes HP disaster tolerant solutions implemented with HP Serviceguard for Linux working for you.

Tip:

You can view the Disaster Proof video at www.hp.com/go/DisasterProof

Evaluating the need for disaster tolerance

Disaster tolerance is the ability to restore applications and data within a reasonable period of time after a disaster. Fire, flood, and earthquake are most common disasters, but a **disaster** can be any event that unexpectedly interrupts service or corrupts data in an entire data center, such as a backhoe that digs too deep and severs a network connection or an act of sabotage. Disaster tolerant architectures protect against unplanned down time due to disasters by geographically distributing the nodes in a cluster so that a disaster at one site does not disable the entire cluster. To evaluate your need for a disaster tolerant solution, weigh:

- Risk of disaster. Areas prone to tornadoes, floods, or earthquakes might require a disaster recovery solution. Some industries need to consider risks other than natural disasters or accidents, such as terrorist activity or sabotage.

The type of disaster to which your business is prone, whether due to geographical location or the nature of the business, determines the type of disaster recovery you choose. For example, if you live in a region prone to massive earthquakes, you are not likely to put your alternate or backup nodes in the same city as your primary nodes, because that type of disaster affects a large area.

The frequency of the disaster also is important in determining whether to invest in a rapid disaster recovery solution. For example, you would be more likely to protect business critical applications and data from hurricanes that happen every season, rather than protecting them from a dormant volcano.

- Vulnerability of the business. How long can your business afford to be down? Some parts of a business might be able to endure 1 or 2 days for recovery, while others need to recover in minutes.

Some parts of a business only need local protection from single outages such a node failure. Other parts of a business might need both local protection and protection in case of site failure.

It is important to consider the role the data servers play in your business. For example, you might target the assembly line production servers as most in need of quick recovery. But if the most likely disaster in your area is an earthquake, it would render the assembly line inoperable, as well as the computers. In this case disaster recovery would be moot, and local failover is probably the more appropriate level of protection.

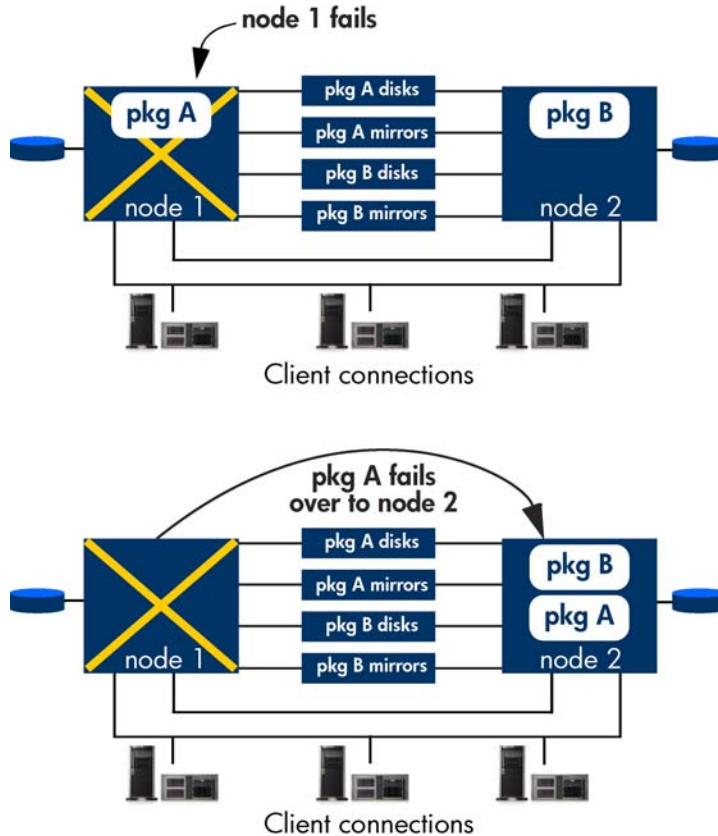
However, you might have an order-processing center that is prone to floods in the winter. The business loses thousands of dollars a minute while the order processing servers are down. A disaster tolerant architecture is appropriate protection in this situation.

Deciding to implement a disaster recovery solution depends on the balance between risk of disaster and the vulnerability of your business if a disaster occurs. The following sections give a high-level view of a variety of disaster tolerant solutions and sketch the general guidelines that you should follow in developing a disaster tolerant computing environment.

What is a disaster tolerant architecture?

In a Serviceguard cluster configuration, high availability is achieved by using redundant hardware to eliminate single points of failure. This protects the cluster against hardware faults such as the node failure in Figure 1.

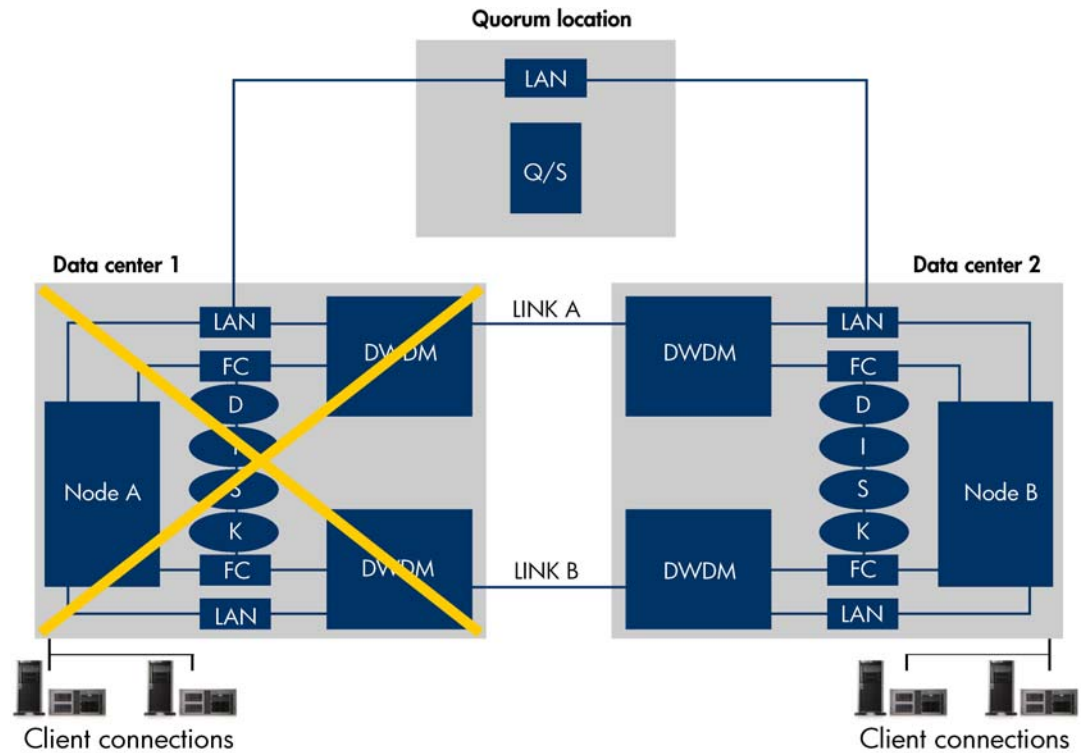
Figure 1. High availability architecture



This architecture, which is typically implemented on one site in a single data center, is sometimes called a **local cluster**. For some installations, the level of protection given by a local cluster is insufficient. Consider the order-processing center where power outages are common during harsh weather. Or, consider the systems running the stock market, where multiple system failures, for any reason, have a significant financial importance to guard not only against single points of failure, but also against **multiple points of failure (MPOF)** or against single massive failures that cause many components to fail, such as the failure of a data center, of an entire site, or of a small area. A **data center**, in the context of disaster recovery, is a physically proximate collection of nodes and disks, usually all in one room.

Creating clusters that are resistant to multiple points of failure or single massive failures requires a different type of cluster architecture called a **disaster tolerant architecture**. This architecture provides the ability to fail over automatically to another part of the cluster after certain disasters. Specifically, the disaster tolerant cluster provides appropriate failover in the case where a disaster causes an entire data center to fail, as illustrated in Figure 2.

Figure 2. Disaster tolerant architecture



Understanding types of disaster-tolerant clusters

To protect against multiple points of failure, cluster components must be geographically dispersed: nodes can be put in different rooms, on different floors of a building, or even in separate buildings or cities. The distance between the nodes is dependent on the types of disaster from which you need protection and on the technology used to replicate data. Two types of disaster-tolerant clusters are described in this white paper:

- HP Serviceguard Extended Distance Cluster for Linux
- HP StorageWorks XP Cluster Extension Software

These types of disaster-tolerant clusters differ from a simple local cluster in many ways. Extended distance clusters and **metropolitan clusters** often require right-of-way from local governments or utilities to lay network and data replication cables or connect to DWDMs. This can complicate the design and implementation. These clusters also require a different kind of control mechanism such as a quorum server, to prevent data integrity issues. Typically, extended distance and metropolitan clusters use an arbitrator site containing a computer running a quorum application.

Extended distance clusters

Note:

Extended distance clusters were formerly known as **campus clusters**, but that term is not always appropriate because the supported distances have increased beyond the typical size of a single corporate campus.

An **extended distance cluster** (also known as an **extended campus cluster**) is a normal Serviceguard cluster that has alternate nodes located in different data centers, separated by some distance with a third location supporting the quorum service. Extended distance clusters are connected using a high-speed cable that guarantees network access between the nodes as long as all guidelines for disaster tolerant architecture are followed. The maximum distance between nodes in an extended distance cluster is set by the limits of the data replication technology and networking limits. An extended distance cluster is shown in Figure 3.

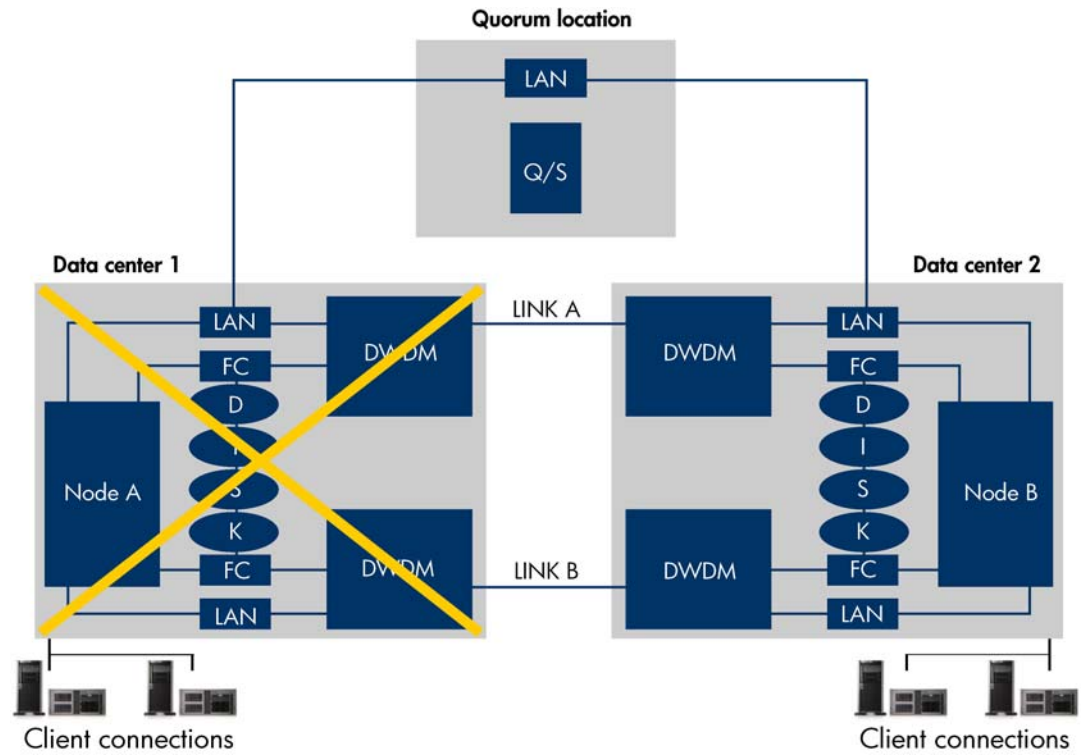
Note:

There are no guidelines or recommendations on how far the third location must be from the two main data centers. The third location can be as close as the room next door with its own power source or can be as far as in a site across town. The distance among all three locations dictates the level of disaster tolerance an extended distance cluster can provide.

In an extended distance cluster for data replication the Multiple Disk (MD) driver is used. Using the MD kernel driver, you can configure RAID 1 (mirroring) in your cluster. In a dual data center setup, to configure RAID 1, one LUN from a storage device in data center 1 is coupled with a LUN from a storage device in data center 2. As a result, the data that is written to this MD device is simultaneously written to both devices. A package that is running on one node in one data center has access data from both storage devices.

The two recommended configurations for the extended distance cluster are both described in Figure 3.

Figure 3. Extended distance cluster



In the previous configuration the network and Fibre Channel links between the data centers are combined and sent over common DWDM links. Two DWDM links provide redundancy. When one of them fails, the other can still be active and can keep the two data centers connected. Using the DWDM link, clusters can now be extended to greater distances, which was not possible earlier due to limits imposed by the Fibre Channel link for storage and Ethernet for networks. Storage in both data centers is connected to both the nodes with two Fibre Channel switches in order to provide multiple paths. This configuration supports a distance up to 100 km between data center 1 and data center 2.

Figure 4. Two data center setup

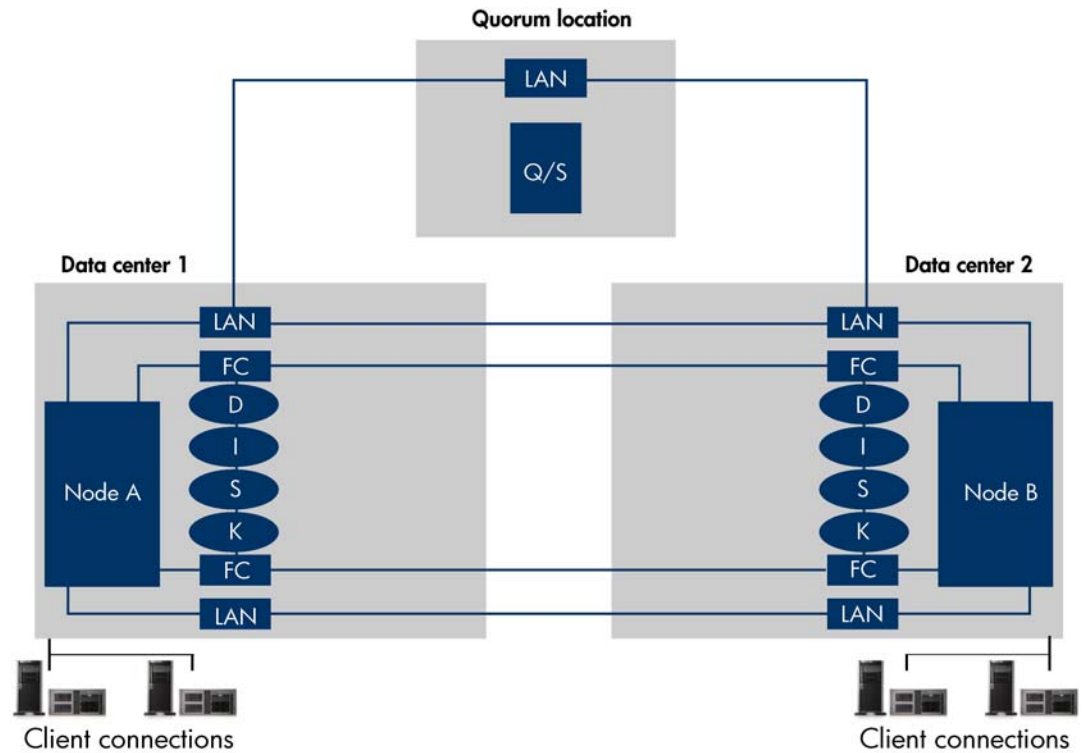


Figure 4 shows a configuration that is supported with separate network and Fibre Channel links between the data centers. In this configuration, the Fibre Channel links and the Ethernet networks are not carried over DWDM links. But each of these links is duplicated between the two data centers, for redundancy. The disadvantage of having the network and the Fibre Channel links separate is that if there is a link failure between sites, the ability to exchange heartbeats and the ability to write mirrored data are not lost at the same time. This configuration is supported to a distance of 10 km between data centers.

All the nodes in the extended distance cluster must be configured with the multipath feature of the QLogic driver to provide redundancy in connectivity to the storage devices. Mirroring for the storage is configured such that each half of the mirror (disk set) is physically present at one data center. Further, from each of the nodes, there are multiple paths to both of these mirror halves.

Benefits of Extended Distance Cluster

The following table discusses the benefits of Extended Distance Cluster.

Benefits of Extended Distance Cluster
This configuration implements a single Serviceguard cluster across two data centers, and uses Multiple Device (MD) driver for data replication.
You can choose any mix of Fibre Channel-based storage supported by Serviceguard that also supports the QLogic driver multipath feature.
This configuration might be the easiest to understand because it is similar in many ways to a standard Serviceguard cluster.
Application failover is minimized. All disks are available to all nodes, so that if a primary disk fails but the node stays up and the replica is available, there is no failover. (The application continues to run on the same node while accessing the replica.)
Data copies are peers, so there is no issue with reconfiguring a replica to function as a primary disk after failover.
Writes are synchronous, so data remains current between the primary disk and its replica, unless the link or disk is down.

Tip:

More info on HP Serviceguard Extended Distance Cluster for Linux can be obtained from <http://www.hp.com/go/xdclinux>.

HP StorageWorks Cluster Extension Software

Cluster Extension for Linux is similar to an HP-UX metropolitan cluster and is a cluster that has alternate nodes located in different parts of a city or in nearby cities. Locating nodes further apart increases the likelihood that alternate nodes are available for failover in the event of a disaster. The architectural requirements are the same as for an extended distance cluster, with the additional constraint of a third location for arbitrator nodes or a quorum server. Additionally, as with an extended distance cluster, the distance separating the nodes in a metropolitan cluster is limited by the data replication and network technology available.

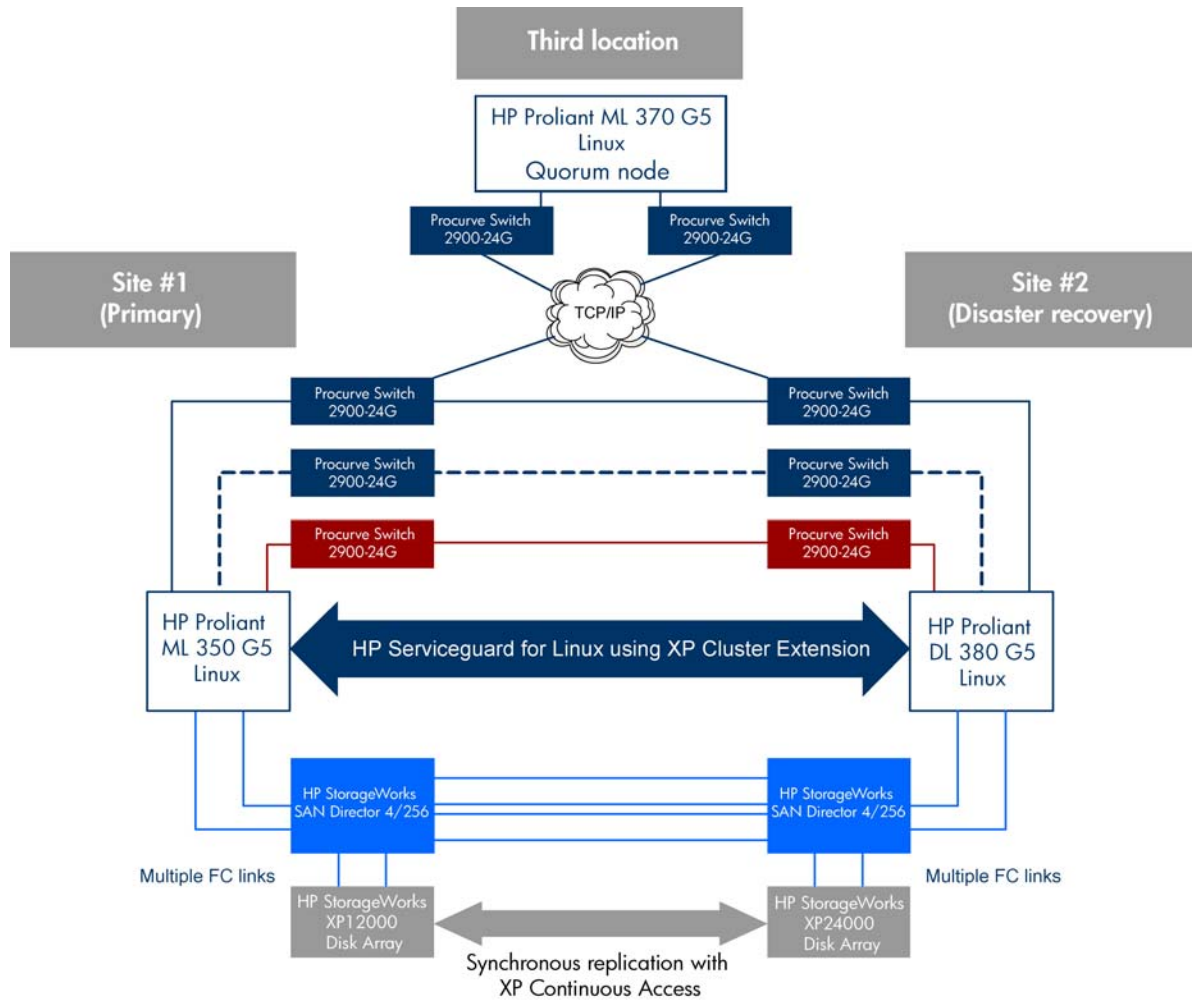
In addition, there is no hard requirement on how far the third location has to be from the two main data centers. The third location can be as close as the room next door with its own power source or can be as far as in a site across town. The distance between all three locations dictates the level of disaster tolerance a metropolitan cluster can provide.

On Linux, the metropolitan cluster is implemented using Cluster Extension:

- Cluster Extension for XP
- Cluster Extension for EVA

Figure 5 shows the Serviceguard for Linux and Cluster Extension architecture and product suite that were specifically used for the Disaster Proof video.

Figure 5. Serviceguard for Linux and Cluster Extension disaster proof demonstration architecture



HP StorageWorks XP Cluster Extension Software offers protection against system downtime to critical applications for enterprise customers using the HP StorageWorks XP Disk Array family. HP Cluster Extension enables hands-free failover and failback decision-making as it detects failures and automatically manages recovery without human intervention. Offering comprehensive disaster tolerance against application downtime from fault, failure, or site disaster by extending clusters between data centers, HP Cluster Extension works seamlessly with HP Serviceguard for Linux, HP StorageWorks XP Continuous Access software, and your XP Disk Array storage system to automate failover and failback between sites. Using HP XP Continuous Access software remote mirroring enables HP StorageWorks XP Cluster Extension Software to verify the status of the storage as well as the status of the

server cluster. The correct failover and failback decisions are made automatically, minimizing downtime and accelerating recovery.

Benefits of Cluster Extension

Benefits of Cluster Extension

Cluster Extension offers a more resilient solution than Extended Distance Cluster, as it provides complete integration between the Serviceguard application package and the data replication subsystem. The storage subsystem is queried to determine the state of the data on the arrays.

Cluster Extension knows that application package data is replicated between two data centers. It takes advantage of this knowledge to evaluate the status of the local and remote copies of the data, including whether the local site holds the primary copy or the secondary copy of data, whether the local data is consistent and whether the local data is current. Depending on the result of this evaluation, Cluster Extension decides if it is safe to start the application package, whether a resynchronization of data is needed before the package can start or whether manual intervention is required to determine the state of the data before the application package is started.

Cluster Extension allows for customization of the startup behavior for application packages depending on your requirements such as data currency or application availability. This means that by default, Cluster Extension always prioritizes data consistency and data currency over application availability. If, however, you choose to prioritize availability over currency, you can configure Cluster Extension to start up even when the state of the data cannot be determined to be fully current, However, the data is consistent.

Cluster Extension XP supports synchronous and asynchronous replication modes, allowing you to prioritize performance over data currency between the data centers.

Because data replication and resynchronization are performed by the storage subsystem, Cluster Extension can provide significantly better performance than Extended Distance Cluster during recovery. Unlike Extended Distance Cluster, Cluster Extension does not require any additional CPU time for data replication, which minimizes the impact on the host.

There is little or no lag time writing to the replica, so the data remains current.

Data can be copied in both directions, so that if the primary site fails and the replica takes over, data can be copied back to the primary site when it comes back up.

Disk resynchronization is independent of CPU failure. If the hosts at the primary site fail but the disk remains up, the disk knows it does not have to be resynchronized.

Tip:

More info on HP StorageWorks XP Cluster Extension software can be obtained from <http://www.hp.com/go/clxvp>.

More info on HP StorageWorks EVA Cluster Extension software can be obtained from <http://www.hp.com/go/clxeva>.

Differences Between Extended Distance Cluster and Cluster Extension

The major differences between an Extended Distance Cluster and Cluster Extension are:

- A key difference between extended distance clusters and HP Cluster Extension is the data replication technology used. The two basic methods available for replicating data between the data centers for Linux clusters are either host-based or storage array-based. Extended Distance Cluster always uses host-based replication (MD software mirroring on Linux). Any combination of Serviceguard supported Fibre Channel storage can be implemented in an Extended Distance Cluster. Cluster Extension always uses extremely robust array-based replication and mirroring, and requires storage from the same vendor in both data centers (that is, a pair of HP StorageWorks XP disk arrays with Continuous Access, or a pair of HP StorageWorks EVA disk arrays with Continuous Access).
- Data centers in an Extended Distance Cluster can span up to 100 km, whereas the distance limitations between data centers in a cluster with Cluster Extension are defined based on the storage array.

Cluster Extension EVA distance supported is defined by the shortest of the following distances:

- Maximum distance that guarantees a network latency of less than 20 minutes
- Maximum distance of 500 km distance
- Maximum distance supported by the data replication link
- Maximum supported distance for DWDM as stated by the provider

Cluster Extension XP distance supported is defined by the shortest of the following distances (In contrast to Cluster Extension EVA, there is no Cluster Extension software limiting factor.):

- Maximum distance supported by the data replication link
- Maximum supported distance for DWDM as stated by the provider

- In an Extended Distance Cluster, there is no built-in mechanism for determining the state of the data being replicated. When an application fails over from one data center to another, the package is allowed to start up if the volume groups can be activated. With a Cluster Extension implementation, an application is only allowed to start up based on the state of the data and the disk arrays.

Data might be updated on the disk system local to a server running a package without remote data being updated. This happens if the data link between sites is lost, usually as a precursor to a site going down. If this occurs and the site with the latest data then goes down, that data is lost. This time from losing the link to the site going down is known as the recovery point. An objective can be set for the recovery point so that if data is updated for a period less than the objective, automated failover occurs and a package starts. If the time is longer than the objective, then the package does not start. In a Linux environment, this is a user configurable parameter: RPO_TARGET.

- Extended Distance Cluster disk reads can outperform Cluster Extension in normal operations. However, Cluster Extension data resynchronization and recovery performance are better than Extended Distance Cluster.

Attributes	Extended Distance Cluster	CLX
Key Benefit	Excellent in “normal” operations, and partial failure. Since all hosts have access to both disks, in a failure where the node is running and the application is up, but the disk becomes unavailable, no failover occurs. The node will access the remote disk to continue processing.	Two significant benefits: • Provides maximum data protection. State of the data is determined before application is started. If necessary, data resynchronization is performed before application is brought up. • Better performance than Extended Distance Cluster for resynchronization, as replication is done by storage subsystem (no impact to host).
Key Limitation	No ability to check the state of the data before starting up the application. If the volume group (vg) can be activated, the application will be started. If mirrors are split or multiple paths to storage are down, as long as the vg can be activated, the application will be started. Data resynchronization does not have a big impact on system performance. However, the performance varies depending on the number of times data resynchronization occurs. In the case of MD, data resynchronization is done one disk at a time, using about 10% of the available CPU time and taking longer to resynchronize multiple LUNs. The amount of CPU time used is a configurable MD parameter.	Specialized storage required. Currently, XP with Continuous Access, and EVA with Continuous Access are supported.
Maximum Distance	100 Kilometers	Shortest of the distances between: • Cluster network latency (not to exceed 200 ms). • Data Replication Max Distance. • DWDM provider max distance.
Data Replication mechanism	Host-based, through MD. Replication can affect performance (writes are synchronous). Resynchronization can impact performance. (Complete resynchronization is required in many scenarios that have multiple failures.)	Array-based, through Continuous Access XP or Continuous Access EVA. Replication and resynchronization performed by the storage subsystem, so the host does not experience a performance hit. Incremental resynchronizations are done, based on bitmap, minimizing the need for full re-syncs.
Application Failover type	Automatic (no manual intervention required).	Automatic (no manual intervention required).
Access Mode for a package	Active/Standby	Active/Standby
Client Transparency	Client detects the lost connection. You must reconnect once the application is recovered at second site.	Client detects the lost connection. You must reconnect once the application is recovered at second site.
Maximum Cluster Size Allowed	4 nodes	2 to 16 nodes
Storage	Identical storage is not required (replication is host-based with MD mirroring).	Identical Storage is required.
Data Replication Link	Dark Fiber	Dark Fiber Continuous Access over IP Continuous Access over ATM
Cluster Network	Single or multiple IP subnet	Single or multiple IP subnet
DTS Software/ Licenses Required	Serviceguard for Linux + XDC	Serviceguard for Linux + CLX XP or CLX EVA

Conclusion

The business-continuity and availability solutions of HP empower customers to improve their business performance and protect their corporate reputations with resilient IT.

The health of your business depends on access to critical IT services and information. Virtually any amount of IT downtime can mean lost productivity, lost revenue, lost customers, lost opportunities. That means you need to be prepared for the full range of threats to the availability and stability of your core infrastructure. Threats ranging from communications disruptions, application problems, and unexpected peaks in customer traffic to full-scale disasters.

For example, The National Emergency Operations Center (NEOC) of Switzerland oversees the country's preparation, monitoring, and response to a wide range of natural and manmade crises—including floods; earthquakes; and nuclear, biological, and chemical threats. The agency must be able to physically accommodate frequent and rapid increases in staff during training and crises. In addition, NEOC must be able to stay in constant communication with each of Switzerland's 26 cantons and several technical services such as the Swiss Seismological Institute. The organization also links to numerous environmental monitoring stations that measure weather and radiological data across the country and in neighboring countries. The ability of the NEOC to fulfill its mission depends directly on the availability, responsiveness, and flexibility of its IT infrastructure. HP Cluster Extension and HP Serviceguard for Linux are the critical Linux portion of NEOCs overall multi-OS disaster tolerant solution that ensures the center can meet its objectives during times of crisis.

HP provides proven strategies, services, and technologies to reduce your exposure and vulnerability, help protect your mission-critical operations against diverse downtime threats, and ease your recovery if an unforeseeable catastrophe strikes. HP offers a comprehensive portfolio of disaster-tolerant solutions on Linux to protect your data in the event of planned or unplanned outages.

For more information

Main website: <http://www.hp.com/go/DisasterProof>

Business Continuity & Availability: <http://www.hp.com/go/continuityandavailability>

Disaster tolerance: <http://www.hp.com/go/disastertolerant>

Bulletproof XP demonstration: <http://www.hp.com/go/storageworks/bulletproofxp>

HP Serviceguard for high availability and disaster tolerance: <http://www.hp.com/go/ha>

Serviceguard for Linux: <http://www.hp.com/go/sqlx>

HP Open Source and Linux: <http://www.hp.com/go/linux>

HP Storage Essentials, Enterprise Edition: <http://www.hp.com/go/storageessentials>

HP Integrity: <http://www.hp.com/go/integrity>

HP Integrity Superdome Server: <http://www.hp.com/go/integritysuperdome>

Windows on Integrity: <http://www.hp.com/go/integrity/windows>

HP Integrity NonStop Servers: <http://www.hp.com/go/nonstop>

HP OpenVMS for Integrity servers: <http://www.hp.com/go/openvms>

HP StorageWorks XP Cluster Extension: <http://www.hp.com/go/clxxp>

HP StorageWorks 4/256 SAN Director: <http://www.hp.com/go/SAN>

HP StorageWorks XP Disk Array Family: <http://www.hp.com/go/storageworks/xp>

ESG Lab Validation Report: HP XP 'Bulletproof Software' for High Availability

<http://h71028.www7.hp.com/ERC/downloads/4AA0-8484ENW.pdf>

ProCurve Networking by HP: <http://www.procurve.com/DisasterProof>

HP ProLiant Servers: <http://www.hp.com/go/ProLiant>

HP Serviceguard Extension for RAC

http://h71028.www7.hp.com/enterprise/cache/257273-0-0-121.aspx?jumpid=reg_R1002_USEN

To learn more about National Technical Systems, go to the following websites:

NTS Corporate website: <http://www.ntscorp.com/>

NTS Camden, AR History Channel Television Documentary:

<http://www.ntscorp.com/news/historychannel.html>