Delivering high availability and disaster tolerance in a multi-operating-system HP Integrity server environment

Disaster tolerant reference architectures

Table of contents

1. Executive summary

Business continuity is essential to profitability and competitiveness. High availability (HA) and disaster tolerance (DT) are the IT elements that contribute to business continuity. The harm caused by only a few seconds of system downtime can be gauged on a scale ranging from inconvenience to bankruptcy to the loss of human life. Table 1 displays the cost of downtime per minute across a range of environments.

Application Type	Cost/Minute
Trading (Securities)	$73,000
Home Location Register (HLR)	$29,300
Enterprise Resource Planning (ERP)	$14,800
Order Processing	$13,300
E-Commerce	$12,600
Supply Chain	$11,500
Electronic Funds Transfer (EFT)	$6,200
Point of Sales (POS)	$4,700
Automatic Teller Machines (ATM)	$3,600
E-mail	$1,900

Source: The Standish Group 2005

Table 1. High costs of downtime

Most enterprises have a number of IT systems, each doing one or more business processes at particular priorities with applications and technology best suited for that work. Moreover, each of those environments has different requirements for HA and DT. When considering all of the possible options, configurations, management requirements and so forth, it can be a daunting task to ensure that each application serving its particular business process has precisely what it needs in terms of HA and DT.

As organizations design and budget for those systems they realize that not all parts of the organization require the same levels of protection. Most organizations select applications that are best suited to their business processes. Because not all applications will run on the same operating systems, organizations are compelled to have a variety of IT environments. This can lead to a multiplicity of technologies that may add considerable complexity to the task of deploying HA and DT solutions.

It is also clear that adaptability and agility are keys to success. Effective solutions require simple, standardized, modular, and integrated building blocks. These building blocks provide an enterprise with the ability to select applications that meet its requirements with minimal concern for the number of platforms needed to support those choices. At the same time, the enterprise must ensure that the solutions can achieve the levels of HA and DT required for each business process. The inability to do so can create overly complex, confusing, contradictory, and costly infrastructures. HP helps enterprises avoid those negatives by delivering adaptability and agility with the HP Adaptive Infrastructure, dynamically meeting business demands by supplying IT resources as required.

The purpose of this document is to show how the HP Adaptive Infrastructure is able to help enterprises achieve their HA and DT objectives in multi-operating system environments with HP Integrity servers. HP has years of experience collaborating with its customers and partners in delivering IT infrastructures that meet a wide range of business continuity requirements.

To help illustrate these HP capabilities, this paper presents reference architectures for a variety of scenarios that most enterprises encounter. The configurations are based on the four operating systems that are fully supported on HP Integrity servers (HP-UX 11i, HP OpenVMS, Linux, and Microsoft Windows Server® 2003), as well as the new HP Integrity NonStop server environment.

2. Defining the basics

HP strategic operating systems (HP-UX 11i, HP OpenVMS, Linux, Windows Server 2003, and HP NonStop) enable enterprises to support the appropriate level of high availability and disaster tolerance in a well-integrated, manageable HP Integrity server-based infrastructure. We begin our discussion by defining high availability and disaster tolerance and their modifying objectives – recovery point and recovery time. (See Appendix C for additional terminology related to topics in this paper.)

High availability (HA)

In its most general definition, HA is the assurance that the computing environment and data are available to those who need them and to the degree they are needed. From a technology perspective, HA is the capability of a system to continue to provide service in the event of the failure of one or more components, or to provide continuing service during planned downtime. This is typically considered to be a function of the server, operating system, and appropriate layered software (e.g., clustering software).

Disaster tolerance (DT)

DT is the capability of a computing environment and data to withstand a disaster (such as fire, loss of power, loss of communication components, or a natural catastrophe) and to continue operating, or to be returned to operation in a relatively short period of time. DT is achieved by building a distributed system in which redundant elements are physically separated. Distances of separation may range from adjacent buildings to thousands of miles, including intercontinental distances. DT may be fully outsourced for the most complex environments. It may also be out-managed or handled completely in-house.

Recovery point objective (RPO)

RPO is a measure of the maximum acceptable amount of data loss prior to a disaster. A major concern for data-oriented environments, RPO is typically determined by the point in time when the last copy was made and safely stored in a retrievable manner off-site. The ultimate goal for RPO is zero data loss.

When developing systems to achieve the required RPO, the following are key considerations:

- Data preservation and integrity
- Synchronous and asynchronous replication
- Database access and update

Recovery time objective (RTO)

RTO is the maximum length of time the business process can tolerate without its IT systems functioning. This is of particular concern for transaction-oriented environments. The ultimate goal for RTO is zero downtime. While some customers must recover as quickly as possible throughout the day, others (for example, stock exchanges during off-hours) have some time when an outage does not dramatically interrupt user activity.

Planning for the attainment of the RTO requires the consideration of factors such as:

- Full, single-system cluster functionality
- Active-active cluster functionality
- Hot standby
- Failover functionality

HP Integrity servers and multiple operating systems

Many organizations run a range of applications across a variety of operating systems across a range of servers. Because they are chosen to achieve specific business process goals, applications usually drive the choice of operating systems.

This multiplicity of applications, operating systems, and servers creates a complex management task in fielding HA and DT environments to meet business continuity goals. The HP Integrity server line goes a long way to mitigating this complexity though its multi-OS capabilities and full range of sizes from blades to the largest HP Integrity Superdome servers.

HP customers can take advantage of four industry-leading operating systems—HP-UX 11i, HP OpenVMS, Linux, and Microsoft Windows Server 2003 – across the entire HP Integrity server family. In addition, the HP Integrity NonStop server platform, with its multiple open interfaces, can play a key role by itself, or in conjunction with other HP Integrity server systems.

With the HP Integrity server multi-OS capability, customers can run any of these operating systems on a single server or combinations of operating systems simultaneously on mid-range and high-end servers.

New solutions are deployed quickly, reliably, and with a significantly lower total cost of ownership (TCO) while workloads are consolidated across multiple operating environments. As business requirements change, Integrity servers can be redeployed to run a different IT solution on a different operating system. And, most importantly, HP has designed HA and DT capabilities for each operating system on the HP Integrity server platform. Table 2 summarizes the HP Integrity server multi-OS HA and DT environment.

	HP-UX 11i	OpenVMS	Linux	Windows Server 2003	NonStop Integrity
Cluster	Serviceguard	Built-in	Serviceguard for Linux	MSCS built into the OS	Built-in
Quorum	Yes	Yes	Yes	Yes (XP); MNS required for EVA	Built-in
Max nodes	16 64 for Continentalclusters	96	8 (CLX) 2 (XDC)	8	255
Storage	EVA, XP, EMC	EVA, XP, EMC	EVA, XP (CLX) any SGLX-supported Fibre Channel storage (XDC)	EVA, XP, EMC	Internal, JBOD, XP
Data replication	HP Continuous Access Veritas Mirroring MirrorDisk EMC SRDF	HP Continuous Access Volume Shadowing	HP Continuous Access (CLX) MD driver (XDC)	HP Continuous Access Veritas Storage Foundation Products EMC SRDF	Remote Database Facility – Zero Lost Transactions (optional) GoldenGate Gravic's ShadowBase NTI's DRI Net
S/W fault tolerance	RTR* (planned)	RTR*	RTR*	RTR*	Built-in, with patented process pairs
Services	DTCS	DTCS	DTCS	DTCS	NonStop Solution Development and Integration (SDI)
DT variants	Extended Distance Cluster (XDC) Metrocluster Continentalclusters	Continuous from 0 to 60,000 miles	Extended Distance Cluster (XDC) Metrocluster Continentalclusters Cluster Extension (CLX)	XP – unlimited EVA – 310 mi. (500 km) max	Extended Distance Cluster (XDC) Metrocluster Continentalclusters
Management	HP OpenView System Insight Manager gWLM	HP OpenView System Insight Manager gWLM	HP OpenView System Insight Manager gWLM (Integrity)	HP OpenView System Insight Manager gWLM (Integrity)	HP OpenView System Insight Manager ASAP IR Prognosis
Interconnect	Ethernet FDDI Token ring WAN	Any	Ethernet FDDI WAN	Ethernet FDDI WAN	ServerNet Ethernet WAN
Bi-directional	Yes	Yes	Yes	Yes	Yes, using partners

*See Appendix C for a description of RTR – Reliable Transaction Router.

Table 2. HP Integrity server HA and DT operating system summary

3. HP-UX 11i

Serviceguard

Serviceguard cluster configurations deliver high availability using redundant hardware to eliminate single points of failure. This architecture – typically implemented on one site in a single data center – is sometimes called a “local cluster” (see figure1).

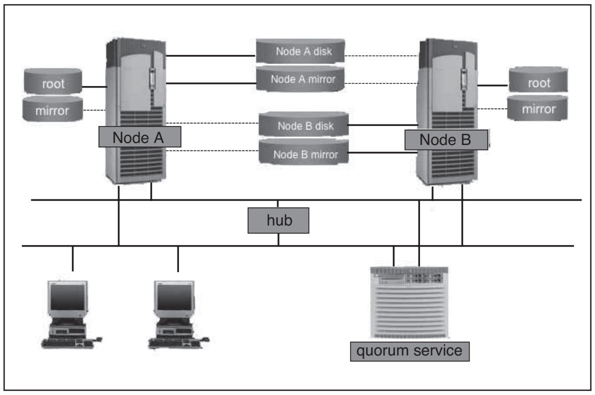

Figure 1. HP-UX 11i Serviceguard basic cluster

For some installations, the level of protection provided by a local cluster is insufficient. Consider an order-processing center where power outages are common during harsh weather, or the systems running a stock market where multiple system failures, for any reason, would have a significant financial impact. For these installations, and many more like them, it is important to guard not only against single points of failure, but also against multiple points of failure (MPOF) or single massive failures that cause many components to fail (such as the failure of a data center, an entire site, or a small area). Configurations of this type are considered disaster tolerant.

An organization may also use high-availability solutions such as Serviceguard to continue operations during planned downtime. Even though an organization may not have an unplanned outage for long periods of time, maintenance such as upgrades and patches can take a server out of use for hours. Using Serviceguard, the application can be handled easily and quickly by another server in the cluster and operations can continue as usual.

Many decisions have to be made when designing a high availability or disaster tolerant solution. These decisions can have a tremendous impact on the solution availability, data consistency, and overall solution. HP offers a range of HA and DT solutions in the HP-UX 11i environment based on the Serviceguard cluster concept:

- Extended Distance Cluster
- Metrocluster
- Continentalclusters

HP-UX 11i configurations

Extended Distance Cluster

Extended Distance Cluster (figure 2) is a normal Serviceguard cluster with nodes spread over two data centers. All nodes are on the same Internet Protocol (IP) subnet. An application runs on one node in the cluster with other nodes configured to take over in the event of a failure in an active/standby configuration.

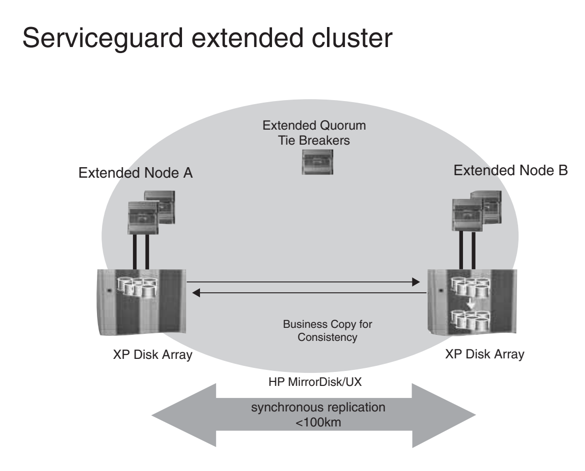

Figure 2. HP-UX 11i Extended Distance Cluster

Either HP-UX MirrorDisk/UX or Symantec Veritas VxVM mirroring is used to replicate application package data between the two data centers in Extended Distance Cluster, even if the data is stored on a RAID array. In Extended Distance Cluster architecture, each clustered server is directly connected to the storage in both data centers. This affords the capabilities listed in table 3.

Cluster topology	Single cluster up to 4 nodes across 2 datacenters (up to 16 nodes with third data center – arbitrator or quorum server)
Geography	Distance -- up to 100 km
Network subnets	Single IP subnet
Network types	Ethernet, FDDI, or Token Ring
Cluster quorum	Quorum server, arbitrator nodes, or duaL cluster lock disks
Failover type	Automatic
Failover direction	Bi-directional
Data replication	MirrorDisk/UX or Symantec Veritas Mirroring

Table 3. HP-UX 11i Extended Distance Cluster summary.

Metrocluster

Similar to the Extended Cluster, the Metrocluster (figure 3) is a normal Serviceguard cluster that has clustered nodes and storage devices located in different data centers separated by some distance. Unlike the Extended Cluster with a maximum range of 100 km, the Metrocluster operates at distances of up to 300 km and verifies the currency of data.

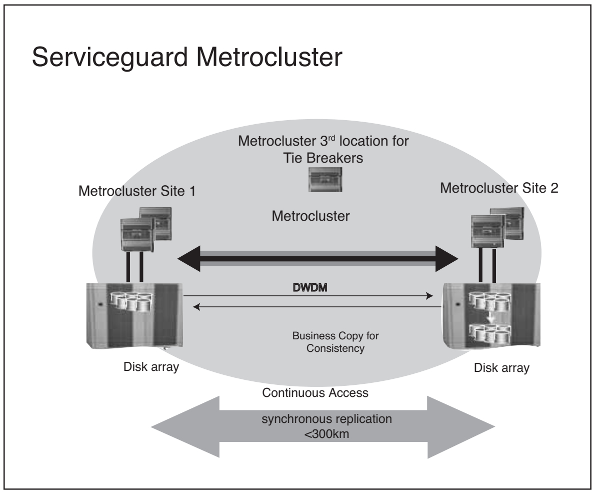
Serviceguard Metrocluster
Metrocluster 3rd location for Tie Breakers
Metrocluster Site 1
Metrocluster Site 2
Metrocluster
DWDM
Business Copy for Consistency
Disk array
Disk array
Continuous Access
synchronous replication <300km

Figure 3. HP-UX 11i Metrocluster

Applications run in an active/standby mode (application resources are only available to one node at a time). The distinct characteristic of Metrocluster is its integration with array-based data replication. Currently, Metrocluster implements three different solutions:

- Metrocluster/CA XP – HP StorageWorks Continuous Access XP
- Metrocluster CA EVA – HP StorageWorks Continuous Access EVA
- Metrocluster/SRDF -- EMC Symmetrix arrays

Table 4 summarizes the characteristics of Metrocluster.

Cluster topology	Single cluster up to 16 nodes across two main data centers and a third location (for arbitrator node(s) or quorum server)
Geography	Metropolitan -- up to 300 km
Network subnets	Single IP subnet
Network types	Ethernet or FDDI
Cluster quorum	Quorum server or arbitrator nodes
Failover type	Automatic
Failover direction	Bi-directional
Data replication	Physical data replication – Continuous Access XP, Continuous Access EVA, and EMC SRDF

Table 4. HP-UX 11i Metrocluster summary.

Continentalclusters

As shown in figure 4, Continentalclusters provides a disaster tolerant solution in which extremely long distances can separate distinct Serviceguard clusters. At these long distances, a wide area network (WAN) is the typical communication vehicle between the clusters.

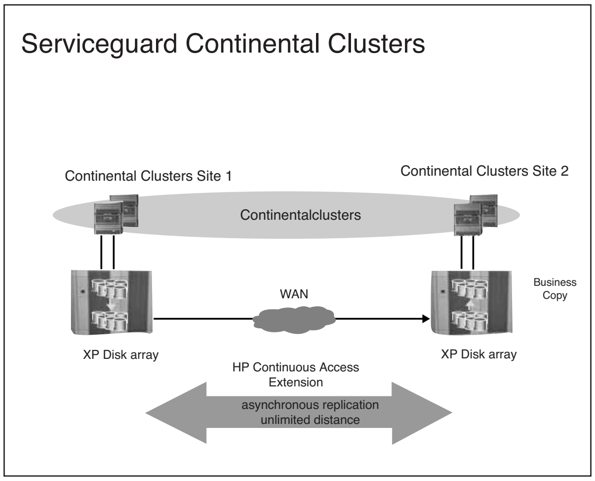
Serviceguard Continental Clusters
Continental Clusters Site 1
Continental Clusters Site 2
Continentalclusters
Business Copy
WAN
XP Disk array
XP Disk array
HP Continuous Access Extension
asynchronous replication unlimited distance

Figure 4. HP-UX 11i Continentalclusters

Unlike Extended Distance Cluster and Metrocluster with their single-cluster architectures, Continentalclusters uses multiple clusters to provide application recovery. Applications run in the active/standby mode with application data replicated between data centers by either storage array-based data replication products (such as Continuous Access XP or EMC SRDF), or software-based data replication (such as Oracle 8*i* Standby DBMS and Oracle 9*i* Data Guard).

Two types of connections are needed between the two Serviceguard clusters in this architecture: one for intercluster communication and another for data replication. Depending on the distance between the two sites, either LAN (i.e., single IP subnet) or WAN connections may be used for cluster network communication.

Continentalclusters provides the ability to monitor a Serviceguard cluster and failover mission-critical applications to another cluster if the monitored cluster should become unavailable. In addition, Continentalclusters supports mutual recovery, which allows for mission-critical applications to run on both clusters, with each cluster configured to recover the mission-critical applications of the other.

Continentalclusters supports ServiceGuard Extension for Real Application Cluster (SGeRAC) in addition to Serviceguard. In an SGeRAC configuration, Oracle RAC database instances are simultaneously accessible by nodes in the same cluster (i.e., the database is only accessible to one site at a time). The Oracle database and data are replicated to the second data center. The RAC instances are configured for recoverability so that the second data center stands by, ready to begin processing in the event of a site failure at the first data center (i.e., across sites; this is an active/standby configuration such that the database is only accessible to one site at a time).

Table 5 summarizes the characteristics of Continentalclusters

Cluster topology	Multiple clusters (2-4), each up to 16 nodes
Geography	Distance range from extended distance to continental or intercontinental
Network subnets	Single IP subnet for multiple clusters or separate IP subnet per cluster
Network types	Ethernet within each data center, LAN or WAN between data centers
Cluster quorum	Quorum server or cluster lock disk for each cluster
Failover type	Semi-automatic
Failover direction	Bi-directional
Data replication	Continuous Access XP, Continuous Access EVA, EMC SRDF, Oracle 8*i* Standby Database, "allow model" for other data. Replication methods to be integrated. Oracle 9*i* DataGuard (contributed scripts are available for single instance).

Table 5. HP-UX 11i Continentalclusters summary

Figure 5 presents an overview of HP-UX 11i inter-site distances.

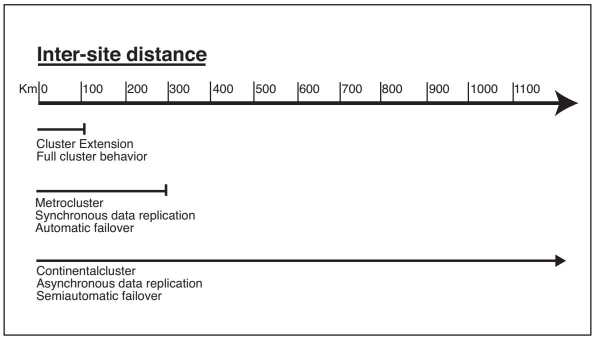

Figure 5. HP-UX 11i inter-site distance

For data replication with both Metrocluster and Continentalclusters, depending on the type of connection [Enterprise Systems Connection (ESCON) or Fibre Channel (FC)] that is supported by the data replication software, the data can be replicated over Dense Wavelength Division Multiplexing (DWDM); 100Base-T and Gigabit Ethernet using Internet Protocol (IP); Asynchronous Transfer Mode (ATM); and T1 or T3/E3 leased lines or switched lines. The Ethernet links and ATM can be implemented over multiple T1 or T3/E3 leased lines.

4. OpenVMS

With shared resources, OpenVMS Cluster systems offer higher availability than standalone systems. Properly configured, OpenVMS Cluster systems can withstand the shutdown or failure of various components. Systems in an OpenVMS Cluster system can share processing, mass storage (including system disks), and other resources under a single OpenVMS security and management domain. Within this highly integrated environment, systems retain their independence because they use local, memory-resident copies of the OpenVMS operating system. Thus, OpenVMS Cluster systems can boot and shut down independently while benefiting from common resources. This results in up to 100% application availability when properly configured.

Because OpenVMS clusters can provide full capabilities even when physically separated by up to 500 miles, they are extremely well suited to DT environments. In addition to clusters, the HP Volume Shadowing for OpenVMS technology (see description in Appendix C) provides exceedingly efficient and reliable data replication capabilities. For distances beyond 500 miles, OpenVMS is also one of the most reliable, efficient, and effective environments for ensuring business continuity because it can protect data virtually seamlessly and ensure maximum uptime.

The HP Disaster Tolerant Continuity Solution (DTCS) is an HP Services offering. It provides consulting, configuration, tools and procedures that ensure best-in-class functioning of long-distance, disaster-tolerant environments. These services are required for implementation of an OpenVMS DT environment with inter-site distances greater than 150 miles. In addition, DTCS is highly recommended any time very low RTO or RPO are desired, regardless of inter-site distance.

Moreover, due simply to light-speed-based latency, certain trade-offs may be encountered at particularly long inter-site distances. Beyond 500 miles, OpenVMS shared-everything cluster functionality may diminish. Regardless of the distance, though, it will continue to afford system failover and data replication and has been tested over distances of up to 60,000 miles.

OpenVMS cluster configurations

This section presents an overview of five different OpenVMS HA-DT configurations:

- Basic high availability
- Basic high availability with data replication
- Full high availability and data recovery
- Data replication using multi-protocol routers
- Data replication and full cluster communication

Basic high availability

As shown in figure 6, two independent high-availability clusters, each with two nodes connected to a Fibre Channel SAN, provide basic high availability[1]. Cluster functionality and communication are managed by System Communication Services (SCS) (see description in Appendix C) over the interconnect.

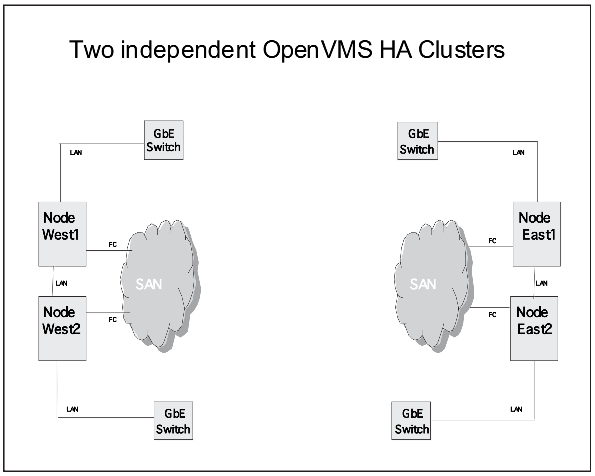

Figure 6. OpenVMS basic high availability

The system will continue to operate through the loss of any component, including the loss of a complete system. Application downtime can be virtually eliminated by proper configuration of the cluster with a cluster-aware[2] application.

Each of these clusters can incorporate any size HP Integrity server and HP AlphaServer system, up to 96 nodes in any combination of the two. Because they are independent clusters, they afford local high availability only.

Because this architecture consists of two fully independent local high-availability clusters, any disaster recovery (data preservation) will require full manual intervention and process. Typically, this would entail regular back up and transmission (manually or electronically) of the save sets to the other site. Recovery would also be manual process driven. Save sets would have to be mounted as needed.

Table 6 summarizes the characteristics of an OpenVMS basic high-availablilty configuration.

Cluster nodes max	96
High availability	Independent at each location
Data replication	None
Key features	Fully manual model capabilities determined by manual processes
Recovery process	Fully manual
Restore process (return to normal)	Fully manual
Workload flexibility	Workload typically locked into the site
Incremental enhancement	None, basic
Services	No special services
RTO	As low as zero
RPO	n/a

Table 6. OpenVMS basic high availability summary.

Basic high availability with data replication

This model builds on basic high availability by enhancing it with automatic data replication. The two clusters are still independent with data replicated between the SANs over a dedicated Fibre Channel link, as illustrated in figure 7.

[1] One of the big issues we find these days is customers not wanting to open up their WANs to non-IP protocols. In some parts of the world, it is difficult to buy a service from a telco that is not compatible or running on IP. OpenVMS cluster communications can accommodate this requirement, generally with the use of encapsulation.

[2] While all applications can take advantage if OpenVMS Clusters, those that are cluster-aware may have additional functionality that fosters enhanced RPO and RTO.

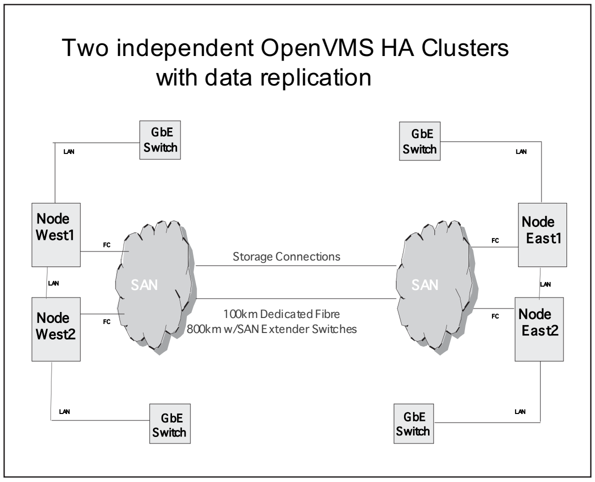

Figure 7. OpenVMS basic high availability with data replication

Using HP Continuous Access or HP Volume Shadowing for OpenVMS, data is replicated across the SANs in figure 7. East data is replicated to the West SAN and West data is replicated to the East SAN. Data is not available to the clusters at the replication sites unless and until the storage units containing the replicated data are mounted to the cluster at the replication site.

In the event of an outage:

- The storage devices holding the replicated data at the replication site are mounted on the replication site cluster
- The applications for that data are launched at the replication site
- The replicated site becomes the primary, local, and only site for both environments

When the original primary site is backed up:

- Data is transferred back to the primary site
- The database is resynchronized
- Applications are relaunched
- The replication site reverts to its original state

Unlike the first example in which save sets are transmitted but not immediately available at the remote site, backup data in this model is available at both sites. It is essential, however, to ensure that only one site has the data mounted at a time. If care is not taken to keep the data mounted at only one site at a time, data corruption is almost certain.

Table 7 summarizes the characteristics of OpenVMS basic high availability with data replication configuration.

Full high availability and data recovery

This model builds on the previous examples above by adding full cluster interconnect between the two sites resulting in a four-node, fully integrated OpenVMS Cluster depicted in figure 8.

Cluster nodes max	96
High availability	Independent at each location
Data replication	Dedicated FC SAN to SAN using HP Continuous Access or Volume Shadowing
Key features	Data replicated at remote site but can't be made available at remote site unless definitive action taken Data corruption risk if storage mounted on both sides at the same time
Recovery process	1. Change storage presentation 2. Mount replicated storage devices 3. Load applications
Restore process (return to normal)	1. Mount recovered storage to currently running site 2. Wait for re-sync of data to complete (shadow catch-up or Continuous Access Normalize) 3. Disable users 4. Shut-down applications 5. Dismount storage 6. Change storage presentation to other site 7. Reboot machines on original site 8. Re-mount storage devices at original site 9. Load applications
Workload flexibility	Workload may be moveable between sites if storage design and deployment permits
Incremental enhancement	Dedicated Fibre Channel between SANs If inter-site distance is >240 km, policy and scripting set up by DTCS
Services	DTCS
RTO	As low as zero within each site. Hours or more between sites
RPO	Approaches 100%

Table 7. OpenVMS basic high availability with data replication summary

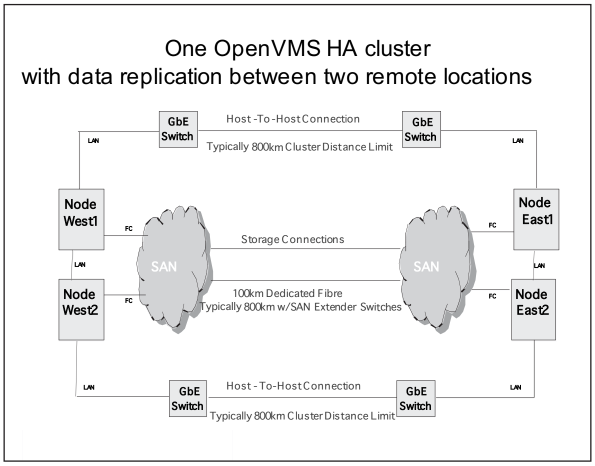

Figure 8. OpenVMS full high availability and data recovery

By adding the host-to-host links, this environment now provides the most advanced and robust high availability and data replication. It is considered fully disaster tolerant. Host-to-host interconnect ensures RTO approaching zero regardless of whether one node or an entire data center (East or West) is lost.

This model provides fully automatic replication of data to the second site and straightforward recovery procedures. Performance needs to be a consideration, though, because interconnects are used more extensively for communication than in the first two examples.

In a two-site architecture such as this, the risk of "split-brain" syndrome is very high (see Appendix C for a discussion of split-brain syndrome). To avoid the headaches that this syndrome may cause, a quorum site or establishment of a quorum disk at one site is highly recommended.

Table 8 summarizes the characteristics of an OpenVMS full high availability and data recovery configuration.

Cluster nodes max	96
High availability	Fully functional cluster with nodes up to 500 miles (800 km) apart or longer depending on conditions
Data replication	Host-based Volume Shadowing over SAN
Key features	Full cluster -- storage available everywhere all of the time with full locking coherency
Recovery process	If application not cluster-wide 1. Start application
Restore process (return to normal)	1. Reboot machines at original site 2. Mount storage to currently running site (no need to wait for shadow catch-up to complete)
Workload flexibility	Workload can typically move between sites as required
Incremental enhancement	Host-to-host communication over standard IP network If inter-site distance is >240 km (policy and scripting set up by DTCS)
Services	DTCS
RTO	As low as zero
RPO	Approaches 100%

Table 8. OpenVMS full high availability and data recovery summary

Data replication using multi-protocol routers

Data replication using multi-protocol routers model (see figure 9) starts with the basic high-availability architecture and adds multi-protocol routers between SANs and their local Gigabit Ethernet switches. In addition, the storage interconnect over the IP network provides the path for either Volume Shadowing or HP Continuous Access. Unless cluster communication is implemented over the inter-site network cluster, functionality remains local.

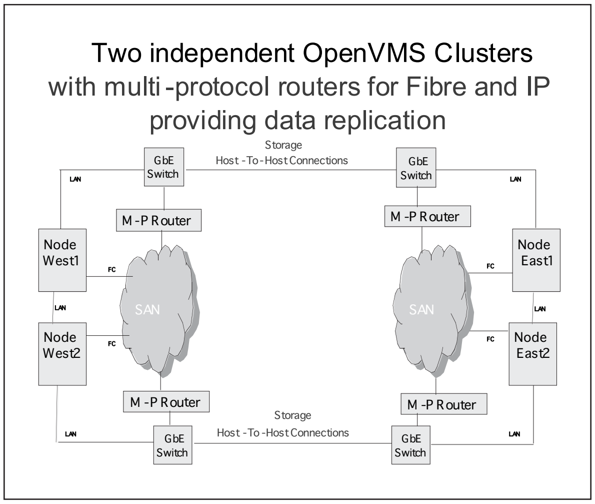

Figure 9. OpenVMS data replication using multi-protocol routers sites

In the first three models, the communications links for data replication required dedicated Fibre Channel lines between the SANs. By employing a multi-protocol router for Fibre and IP, flexibility is increased and full high availability and data replication can be achieved without a dedicated communication link between the SANs. Rather, all communication will be over the IP network. This yields a more cost-effective solution.

Unlike basic high availability in which save sets are transmitted but not immediately available at the remote site, back-up data in this model is available at both sites. It is essential, though, to ensure that only one site at a time has the data mounted. If that is not done, data corruption is almost certain.

Table 9 summarizes the characteristics of OpenVMS data replication using multi-protocol routers.

Cluster nodes max	96
High availability	Independent at each location
Data replication	Host-based Volume Shadowing over SAN carried on IP network
Key features	Data replicated at remote site but can't be made available at remote site unless definitive action taken Risk of data corruption if storage mounted on both sides at the same time
Recovery process	1. Change storage presentation 2. Mount replicated storage devices 3. Load applications
Restore process (return to normal)	1. Mount recovered storage to currently running site 2. Wait for re-sync of data to complete (shadow catch-up or Continuous Access Normalize) 3. Disable users 4. Shut-down applications 5. Dismount storage 6. Change storage presentation to other site 7. Reboot machines on original site 8. Re-mount storage devices at original site 9. Load applications
Workload flexibility	Workload may be moveable between sites if storage design and deployment permits
Incremental enhancement	All inter-site communications over standard IP network If inter-site distance is >240 km, policy and scripting set up by DTCS
Services	DTCS
RTO	As low as zero within each site Hours or more between sites
RPO	Approaches 100%

Table 9. OpenVMS data replication using multi-protocol routers summary

Data replication and full cluster communication

By implementing System Communication Services (SCS) across the communication network in the previous model, full HA and DR is achieved over the same communications links, as shown in figure 10.

Because the same network is used for both cluster traffic and data replication, performance may be diminished resulting in less than 100% RPO.

The significant advantages of this model include the lower cost of using a single, standard network for both cluster and DR traffic, as well as enhanced flexibility and consistency with HP multi-OS environments.

This model provides fully automatic replication of data to the second site with very simple recovery procedures.

In a two-site architecture such as this, it is still impossible to continue automatically from a failure of either site. If required, fully automatic behavior can be achieved by implementing a quorum site as described above.

If performance is critical, then SAN-to-SAN Fibre Channel can be added as it was in the full high-availability and data recovery approach.

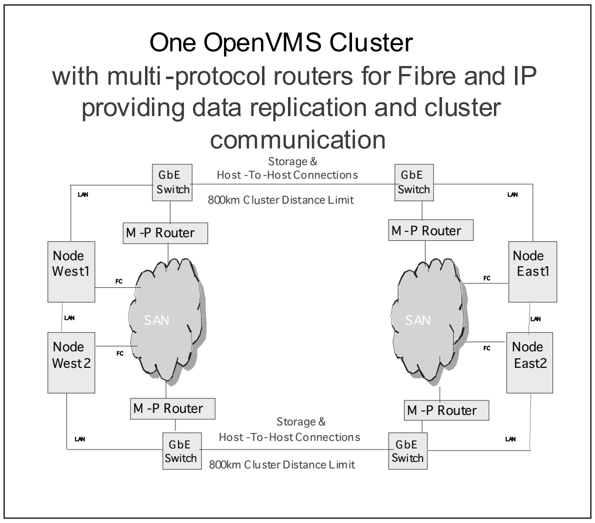

Figure 10. OpenVMS data replication and full cluster communication

Table 10 summarizes the characteristics of OpenVMS data replication and full cluster communication.

Cluster nodes max	96
High availability	Fully functional cluster with nodes up to 800 km apart or longer depending on conditions
Data replication	Host-based Volume Shadowing over SAN carried on IP network
Key features	Full cluster – storage available everywhere all of the time with full locking coherency
Recovery process	If application not cluster-wide 1. Start application
Restore process (return to normal)	1. Reboot machines at original site 2. Mount storage to currently running site (no need to wait for shadow catch-up to complete)
Workload flexibility	Workload can typically move between sites as required
Incremental enhancement	Cluster communication between sites If inter-site distance is >240 km, policy and scripting set up by DTCS
Services	DTCS
RTO	As low as zero
RPO	Approaches 100%

Table 10. OpenVMS data replication and full cluster communication summary

Figure 11 presents an overview of when DTCS should be used in building OpenVMS DT implementations.

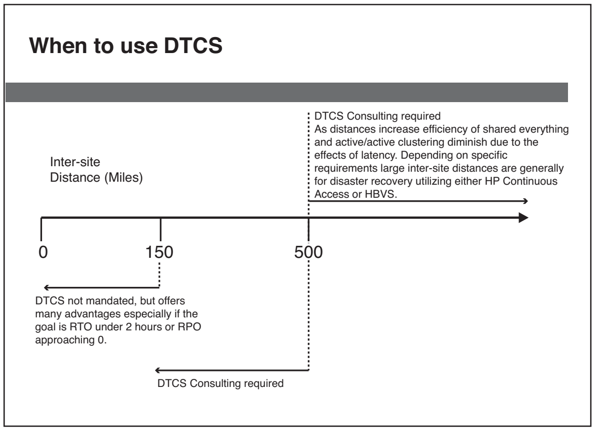
When to use DTCS
Inter-site
Distance (Miles)
DTCS Consulting required
As distances increase efficiency of shared everything and active/active clustering diminish due to the effects of latency. Depending on specific requirements large inter-site distances are generally for disaster recovery utilizing either HP Continuous Access or HBVS.
0
150
500
DTCS not mandated, but offers many advantages especially if the goal is RTO under 2 hours or RPO approaching 0.
DTCS Consulting required

Figure 11. When to use DTCS for OpenVMS DT implementations.

5. Linux

Enterprises recognize the benefits of Linux but they are concerned with securing the appropriate services and adequate support to ensure successful Linux implementations. And they want the assurance of a healthy return on their investments in open source software and industry-standard solutions.

To help solve these problems, HP has developed Linux Reference Architectures (LRA). These tested and compiled solution building blocks are designed to maximize the return on Linux investments. Linux Reference Architectures provide complete and tested Linux platform configurations and middleware stacks on industry-standard hardware including HP Integrity servers, backed by a full array of HP consulting, integration, and support services.

LRA offer a choice of SUSE Linux or Red Hat Enterprise Linux distributions, associated system software such as drivers, optional value-added components such as management systems, and workload-specific software including databases and application servers.

HP offers additional software components to enhance and support the capabilities provided within LRA. For example, HP Serviceguard for Linux is a cluster kit that provides industry-leading high-availability capabilities to Linux environments.

HP Serviceguard for Linux

HP Serviceguard for Linux is a high-availability cluster solution that leverages the strength of HP experience in the HA business, bringing the best-in-class, mission-critical HP-UX 11i technologies to the Linux environment. The local cluster and DT concepts described in the "HP-UX 11i" section of this paper also apply to the HP Serviceguard for Linux environment.

HP offers DT solutions in the Linux environment based on the Serviceguard cluster concept:

- HP Serviceguard Extended Distance Cluster (XDC) for Linux
- HP StorageWorks Cluster Extension (CLX)

HP Serviceguard for Linux, coupled with one of the DT solutions, provides a cost-effective disaster recovery solution with automatic site failover, allowing for rapid site recovery times.

Linux configurations

HP Serviceguard XDC for Linux

HP Serviceguard XDC for Linux is a normal Serviceguard for Linux cluster with nodes spread over two data centers up to 100 km. With Serviceguard XDC for Linux, software-based mirroring via the Linux Multi-Device (MD) driver is "cluster enabled" with technology added to the MD driver so that it can be safely used in a clustered environment. The cluster enablement provided by Serviceguard XDC for Linux protects against potential data corruption caused by the MD driver. An automated Recovery Point Objective timer is also provided to better handle special failure scenarios, in essence, allowing customers to define those cases when not to automatically restart.

HP Cluster Extension and Serviceguard for Linux

HP Cluster Extension and HP Serviceguard for Linux work in concert with XP or EVA mass storage resources to protect against system downtime from fault, failure or disaster by extending a single cluster over metropolitan-wide distances.

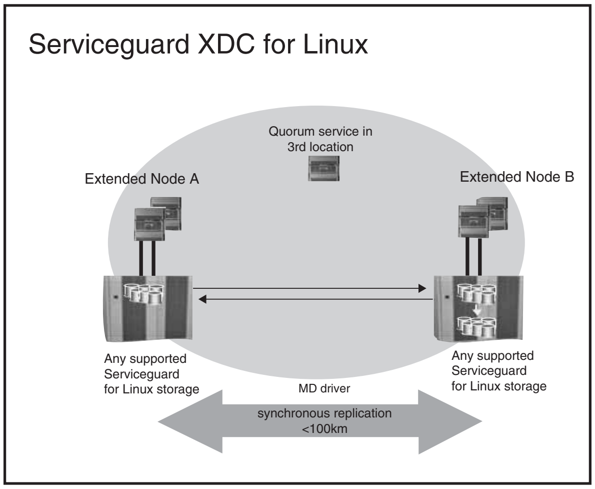

Figure 11A. Serviceguard XDC for Linux.

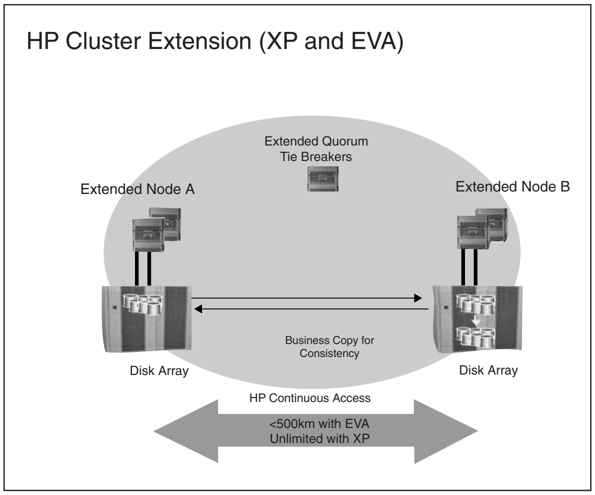

Figure 12. HP Cluster Extention (XP and EVA)

Cluster topology	Single cluster up to 8 nodes across two main data centers and a third location (for arbitrator node(s) or quorum server)
Geography	XP: distance range from extended distance to continental or intercontinental EVA: metropolitan – up to 500 km (20ms roundtrip delay)
Network subnets	Single IP subnet
Network types	Ethernet or FDDI
Cluster quorum	Quorum server or arbitrator nodes
Failover type	Automatic
Failover direction	Bi-directional
Data replication	Physical data replication – Continuous Access XP, Continuous Access EVA

Table 11. Cluster Extension for Linux summary

Cluster topology	A 2-node cluster split across two main data centers and a third location for a quorum service
Geography	Up to 100 km
Network subnets	Single IP subnet
Network types	Ethernet or FDDI
Cluster quorum	Quorum service
Failover type	Automatic
Failover direction	Bi-directional
Data replication	Software data replication - MD driver

Table 12. HP SErviceguard XDC for Linux summary

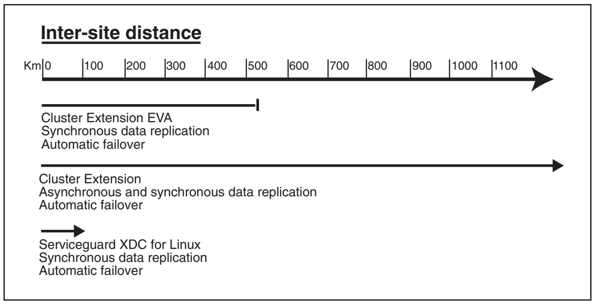

Figure 13. Overview of Linux inter-site distances.

6. Microsoft Windows Server 2003

HP and Microsoft continue to develop and deliver high-value computing solutions with data center power and reliability that help organizations create and maintain a more productive, cost-effective, and secure enterprise. HP offers an automatic path-failover product for the Windows environment. High-availability and disaster-tolerant solutions for the Microsoft Windows environment are based on Microsoft Cluster Service clustering, HP Continuous Access for data replication, and HP Cluster Extension software for geographically dispersed cluster support.

The HP solution offerings start from a single system, whether the EVA or XP disk arrays, which are fully redundant with no single point of failure. For higher levels of availability, the single system disk arrays are scalable for higher levels of availability.

Single system high availability with Windows Server 2003

HP Integrity servers running Windows Server 2003 offer an extremely reliable, industry-standard alternative to proprietary RISC systems. Reliability is built into every server with features such as enhanced Machine Check Architecture (MCA), single-bit error checking and correction (ECC) throughout the system, double chip-spare memory, and Dynamic Memory Resiliency (DMR) providing such features as dynamic page de-allocation.

Microsoft Windows Server 2003 for 64-bit Intel® Itanium®-based systems is the only Windows server OS enabling such high-availability features as Machine Check Architecture. MCA allows the processor, chipset, firmware, and operating system to cooperate in providing advanced reliability and predictive analysis. In conjunction with the Itanium® MCA, HP firmware provides the capability to shield the operating system from some failures that would otherwise cause the system to crash. The diagram in figure 14 shows an example of MCA error flow.

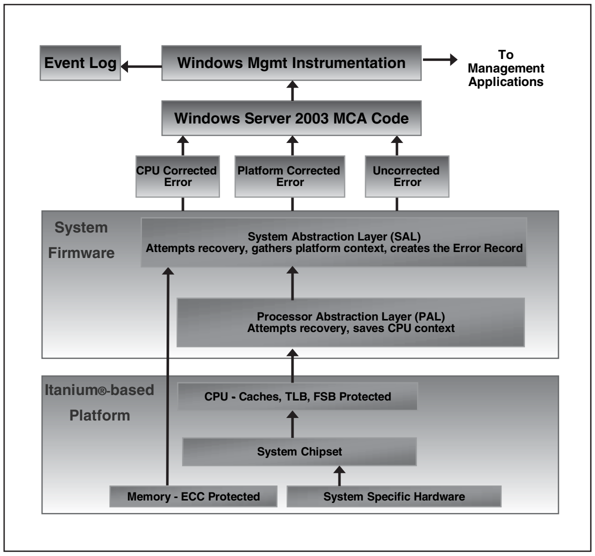
Event Log
Windows Mgmt Instrumentation
To Management Applications
Windows Server 2003 MCA Code
CPU Corrected Error
Platform Corrected Error
Uncorrected Error
System Firmware
System Abstraction Layer (SAL)
Attempts recovery, gathers platform context, creates the Error Record
Processor Abstraction Layer (PAL)
Attempts recovery, saves CPU context
Itanium®-based Platform
CPU - Caches, TLB, FSB Protected
System Chipset
Memory - ECC Protected
System Specific Hardware

Figure 14. Windows Machine Check Architecture error flow

Microsoft Windows Server 2003 configurations

Microsoft Cluster Service

The Microsoft Cluster Service is an integrated component of Windows Server 2003 and installed by default on every system with Datacenter and Enterprise Editions of the operating system. Setup and configuration is through the Cluster Manager, a simple, easy-to-use GUI-based interface.

When nodes join a cluster, the process can be done remotely and does not require a restart of any system. Multiple nodes can join a cluster simultaneously and the cluster will automatically determine the quorum disk. In addition, a new analysis phase is incorporated into the Cluster Service that will detect problems or potential problems with hardware and software configurations and advise the best course of corrective action. Cluster diagnostics and a verification tool are provided as downloadable software components.

Figure 15 is an example of a Windows Server 2003 cluster configuration with HP Integrity server systems, HP StorageWorks disk arrays, and multi-path software for redundancy.

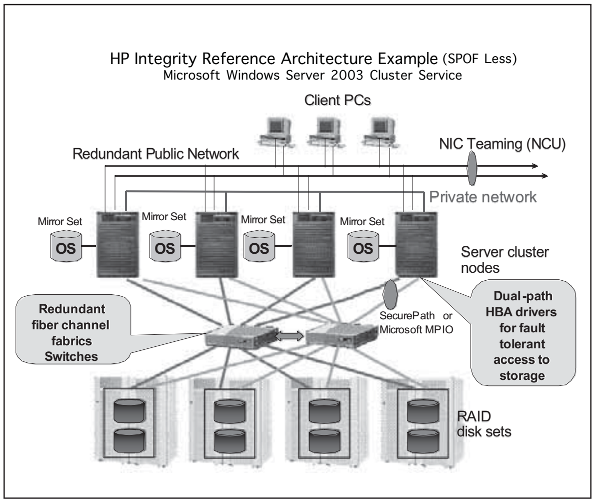
HP Integrity Reference Architecture Example (SPOF Less)
Microsoft Windows Server 2003 Cluster Service
Client PCs
Redundant Public Network
NIC Teaming (NCU)
Private network
Mirror Set
OS
Mirror Set
OS
Mirror Set
OS
Mirror Set
OS
Server cluster nodes
Redundant fiber channel fabrics Switches
SecurePath or Microsoft MPIO
Dual-path HBA drivers for fault tolerant access to storage
RAID disk sets

Figure 15. HP Integrity server Windows Server 2003 cluster

Scalable business-continuity solutions

Many of these business-continuity solutions build from and leverage the capabilities of one another. This means that as organizations consider higher levels of data availability to meet their growing business needs, HP is able to offer easy deployment of the right solution.

Scaling up to more highly available solutions, disk arrays can be integrated into cluster solutions where the EVA and XP disk arrays have been certified to work with the leading cluster software. From here, if a business requires higher availability levels, HP recommends that it consider the XP disk arrays, which come with a broader and more robust selection of business continuity solutions.

Figures 16 depicst scalable business continuity and Cluster Extension solutions with Windows.

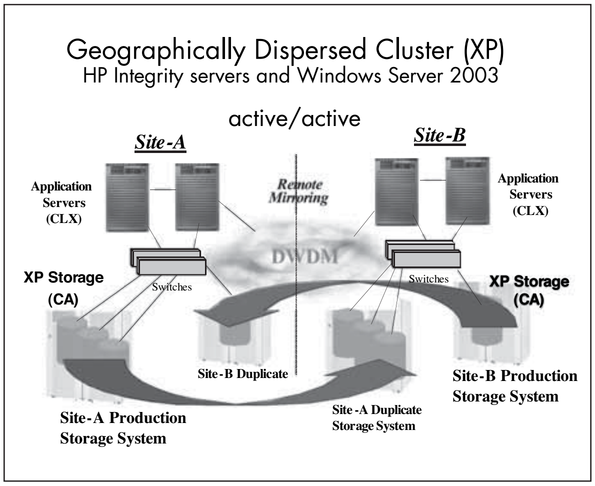
Geographically Dispersed Cluster (XP)
HP Integrity servers and Windows Server 2003
active/active
Site-A
Site-B
Application Servers (CLX)
Remote Mirroring
Application Servers (CLX)
DWDM
XP Storage (CA)
Switches
Switches
XP Storage (CA)
Site-B Duplicate
Site-B Production Storage System
Site-A Production Storage System
Site-A Duplicate Storage System

Figure 16. Cluster Extension for Window

Table 13 summarizes the Windows Cluster Extension configuration.

Cluster topology	Cluster Extension supports up to 8 nodes across two main data centers and a third location. For XP: an additional 9th server running the CLX Arbitrator Service may reside in the third location
Geography	Metropolitan – up to 500 km (20 ms roundtrip delay) Geographic – unlimited distances with XP storage; dependent on synchronous or asynchronous communication
Network subnets	Single IP subnet with Windows Server 2003
Network types	Ethernet or FDDI
Cluster quorum	For XP: Either majority node set or CLX Quorum Filter Service with an optional server running the CLX Arbitrator Service. For EVA: Majority node set only.
Failover type	Automatic
Failover direction	Bi-directional
Data replication	Physical data replication – Continuous Access XP

Table 13. HP Cluster Extension for Windows summary

Figure 17 presents an overview of Windows inter-site distances.

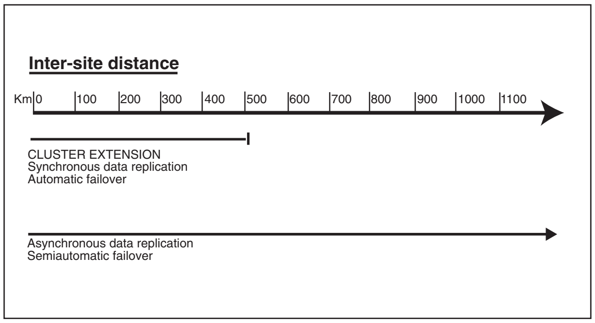

Figure 17. Windows inter-site distances

7. HP Integrity NonStop

HP NonStop servers deliver the highest levels of fault tolerance and availability possible from a single server – out of the box. In clustered configurations they deliver well beyond that, with seven nines of availability, as well as full disaster tolerance.

In 2005, the HP Integrity NonStop server was introduced with either Double or Triple Modular Redundancy (DMR or TMR) in every system. These new systems are based on the HP Integrity server and provide the best hardware fault tolerance on the market, combined with unique software error checking and system redundancies (see table 13). These servers offer a range of disaster tolerance support, from basic tape backup to Continentalclusters. More recently, in June of 2006, the HP Integrity NonStop NS 1000 became available, offering customers a lower price point entry for complete disaster tolerant computing.

Cluster capability	Up to 255 systems with Ethernet; up to 96 systems with ServerNet
Geographic limits	Metroclusters – up to 15 km Continentalclusters – unlimited distances
Location/sites	Single site or multi-site (including transactional consistency over a network)
Network types	ServerNet, Ethernet, or WAN
Cluster quorum	Built-in
Failover type	Manual
Failover direction	Bi-directional with partitioned database, or using partner products for non-partitioned database
Data replication	Asynchronous logical redo from transaction log
RTO	<25 seconds with NonStop Remote Database Facility software
RPO	100% with Zero Lost Transactions configuration

Table 14. HP Integrity NonStop server DT summary

HP Integrity NonStop servers are designed with the entire software stack tuned for transactional consistency. The HP-branded replication software (NonStop Transaction Management Facility and NonStop Remote Database Facility) provides low-level system integration with very high performance.

"Live-standby" configurations are available, allowing for faster takeover because the target system already has the application running. Moreover, cleanly partitioned databases may take advantage of a "live-live" (active-active) configuration with two or more systems sharing transaction processing. In this design, the backup database may even be initialized and loaded while the primary remains online handling all transactions temporarily.

In all configurations, NonStop RDF Software understands the state of every transaction being replicated, whether that transaction is wholly contained on one system or spans multiple NonStop servers. Replication occurs before transactions have been flushed from the source database, further minimizing data loss, overhead, and latency. In a takeover, NonStop RDF Software backs out any transaction whose final state is unknown, ensuring consistency on one target system, or a network of target systems. In addition, HP partners (GoldenGate Software, Gravic, and ITI) offer additional features for off-platform data transformation and active-active processing.

HP Integrity NonStop configuration for ZLT

With the use of StorageWorks XP technology, ZLT – Zero Lost Transactions (an optional feature) – ensures undisturbed transaction processing. As shown in figure 18, a mirrored audit trail disk, located remotely from the source system, can transfer all activity committed on the failed source, instantaneously, to a target system at the ready.

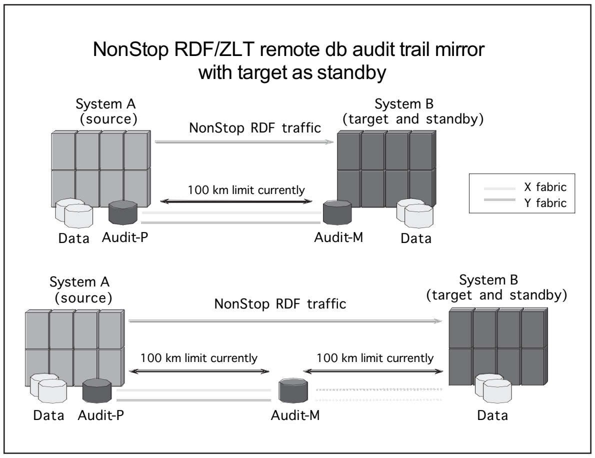

Figure 18. NonStop RDF/ZLT remote db audit trail mirror with target as standby

8. Services

Many of the reference configurations described in this document stand easily on their own, right out of the box. However, in an environment that is sufficiently critical to demand business continuity technology, we highly recommended that the appropriate services be employed, as well. The particular services required will, like the technology, depend a great deal on the specific needs and situation.

In support of business continuity, HP offers a broad and strong portfolio designed to help protect and minimize disruption to IT infrastructure and associated business processes against site outages, disasters and environmental events. This enables enterprises to focus on core competencies and the business of doing business. The HP portfolio provides the processes, technical expertise, and infrastructure and assurance at a cost in line with the risk.

Many of these services, such as the Disaster Tolerant Continuity Solution (DTCS) services, are designed to ensure that HA and DT environments function as planned.

Disaster Tolerant Services are the most stringent level of HP business continuity offerings, intended for business processes that require limited or no interruption for users and essentially no data loss.

- DT design and implementation – For your most critical business processes, HP Services can design and implement disaster tolerant infrastructures. We can split data – and potentially clusters – across two geographically separate data centers, for maximum resilience. The clusters can share production workloads and, if one site fails, the failover process is completed within minutes.
- Managed DT service – HP Services experts can also manage your disaster tolerant solution and physically house your disaster tolerant environment. With proven HP expertise, this can provide both cost and quality benefits. As part of your solution, we can also more broadly manage portions or your entire storage environment. Most importantly, we can help you develop and manage a solution your business can depend on – not just disparate, non-integrated technologies that can fail to deliver when your business needs them the most.
- Disaster Tolerant Solution Center – HP DT management services are delivered by HP Services Disaster Tolerant Solution Center, whose consultants have more than 15 years experience in engineering, delivering, and supporting DT configurations. HP has unparalleled expertise in disaster tolerance, built on more than 20 years of experience in distributed/clustered computing environments.

Beyond this portfolio, HP offers an extensive and flexible set of services to help ensure that your environment functions in the way and at the level of availability that is correct. For more information on HP Services, see www.hp.com/go/services.

9. Summary

Business continuity is essential to profitability and competitiveness. High availability and disaster tolerance are the IT elements that contribute to business continuity. The harm caused by only a few seconds of system downtime can be gauged on a scale ranging from inconvenience to bankruptcy to the loss of human life.

Most enterprises have a number of IT systems each doing one or more business processes, at particular priorities, with applications and technology best suited for that work. Moreover, each of those environments has different requirements for HA and DT. When considering all of the possible options, configurations, management requirements, and so forth, it can be a daunting task to ensure that each application serving its particular business process has precisely what it needs in terms of HA and DT.

When employing the HP Integrity server ecosystem, enterprises are able to configure the precise DT environment for each business process in an efficient, effective, adaptable, and cost-effective manner. This white paper presented a variety of disaster tolerant reference architectures for various operating systems, all with the HP Integrity server as the foundation.

In each case, whether for HP-UX 11i, HP OpenVMS, Linux, or Windows, the architectures built out from a basic high-availability environment all the way to an extreme wide-area disaster tolerant configuration. The HP Integrity NonStop environment builds out similarly.

Any specific process will probably require variations on these reference architectures. Nevertheless, they provide a solid basis for developing what will be the ideal solution to meet the unique needs of any enterprise.

Appendix A: The HP Adaptive Infrastructure

The strategic focus of HP for the past several years has been -- and continues to be -- the realization and fulfillment of its Adaptive Infrastructure approach and delivery of the most capable virtual server environment in the industry. Keeping this in mind, successful HA and DT deployments rely to a large extent on the attributes that stand behind the Adaptive Infrastructure.

At the heart of this approach is the understanding that business and IT must be synchronized. For the enterprise to capitalize on change, business and IT must become aligned and stay aligned. Because a business decision may trigger a series of IT changes, it is critical to ensure that every decision receives precisely the IT services and support that the decision demands. The HP Adaptive Infrastructure architecture helps an organization align business and IT based on four foundation principles:

- Simplification: Simplify complex IT environments through the consolidation of applications and infrastructures, the automation and orchestration of processes, and the virtualization of resources.
- Standardization: Reduce costs and simplify the management of change by establishing corporate standards. Standardize business processes using established and emerging industry-standard operational reference models. Enable collaboration across the enterprises in a value-chain by standardizing on how information and services will interoperate across organizational boundaries.
- Modularity: Improve resource sharing and cost effectiveness by modularizing monolithic capabilities into reusable business services that allow independent work on either side of the module boundary. The interface to a module should be well defined, stable and coarse-grained, enabling a loose coupling between service consumers and suppliers distributed across a network.
- Integration: Improve agility and reduce costs by dynamically linking business processes and heterogeneous, reusable IT resources both within and beyond the enterprise. Of primary importance is the ease of configuring component services to meet new needs.

This set of principles, in turn, helps determine the best HA and DT environment for a particular job because IT infrastructures built on an Adaptive infrastructure model will have the versatility to deliver the level of HA and DT (no more, no less) that is appropriate for each business process. Moreover, such an infrastructure will have the ability to change as the needs for a particular business process change with minimal disruption and maximum investment protection.

Virtual Server Environment (VSE) overview
Key to the HP Adaptive Infrastructure is the HP Virtual Server Environment for HP Integrity and HP 9000 servers. HP VSE helps organizations achieve a greater return on IT investments by optimizing server resource utilization in real time based on business priorities. HP VSE provides a pool of virtual servers that can grow or shrink based on service-level objectives and business needs.

Through tight integration with partitioning, high availability, and utility pricing, HP VSE allows organizations to maintain service levels in the event of downtime and to pay for spare capacity on an as-needed basis.

HP VSE provides intelligent control of virtualized environments through integrated planning, management, and automation. This allows organizations to consolidate multiple applications on a single server and manage clusters as one entity without compromising performance. HP VSE maximizes the use of all available server capacity while providing the highest priority applications with additional resources during peak times[3].

With the topic of this paper in mind, VSE facilitates the deployment of HA and DT capabilities in a multi-operating-system environment.

For more detail on VSE, please refer to www.hp.com/go/vse.

Appendix B: Reliable Transaction Router

Implementing HP Reliable Transaction Router (RTR) can enhance many of the configurations described in the body of this paper. HP RTR provides application disaster tolerance with transaction integrity and real-time transaction completion for HP OpenVMS, Tru64 UNIX®, Microsoft Windows, Linux Frontend systems, and Sun Solaris systems. HP-UX 11i support is planned for 2007.

RTR provides real-time automatic failover and recovery of in-flight business transactions. It is based on a proven technology that offers reliable transaction integrity with a two-phase commit process, scalability without changing the application code, and interoperability with a flexible infrastructure for easy application development and deployment.

RTR guarantees delivery of transactions despite node, database, site, or network failures.

It extends fault tolerance from single systems to fully distributed client/server wide area networks and provides protection against planned and unplanned system, software, and site disruptions. RTR also allows components of the system, such as the database or operating system, to be upgraded while the application continues to run.

RTR is designed for high-volume performance. Applications and data can be partitioned across multiple servers, putting the processing power where it is needed. Consolidated system management enables a single system manager to manage an entire distributed environment.

RTR is based on a three-layer architecture (figure B.1) consisting of front-end (FE) roles, back-end (BE) roles, and transaction router (TR) roles. Client processes run on nodes defined to have the front-end role. This layer allows computing power to be provided locally at the end-user site for transaction acquisition and presentation. Server processes run on nodes defined to have the back-end role. This layer:

- Allows the database to be distributed geographically
- Permits replication of servers to cope with network, node, or site failures
- Allows computer resources to be added to meet performance requirements
- Allows performance expansion while protecting the investments made in existing hardware and application software

The router layer contains no application software but acts as a traffic cop between front-end and back-ends, routing transactions to the appropriate destinations.

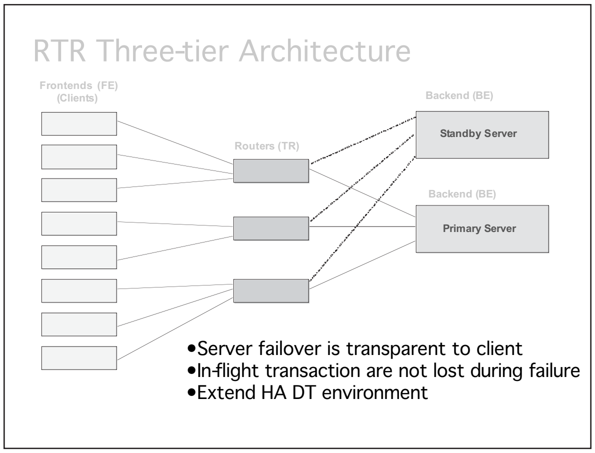
RTR Three-tier Architecture
Frontends (FE)
(Clients)
Routers (TR)
Backend (BE)
Standby Server
Backend (BE)
Primary Server
•Server failover is transparent to client
•In-flight transaction are not lost during failure
•Extend HA DT environment

Figure B.1. Reliable Transaction Router three-tier architecture

[3] In one study focused on HP-UX 11i, VSE achieved a CPU utilization rate of close to 70% in an environment that had had a maximum utilization of only 30%.

Within the three-tier architecture:

- The router is a software component that provides the failover intelligence and manages connections to the backend (not a network router).
- Servers are protected against clients:

  – Clients access the server through an interface, not directly

  – RTR serves as a buffer
- Added layer (routers) avoids resource depletion at the backend.
- Routers condense network connections and make it easier to add systems.
- The tiers are virtual entities; all three functions could be on one node, perhaps for development and debugging purposes.
- Moving the reliability and connectivity parts of the software out of the applications leads to simpler applications. Having simpler applications translates to enhanced quality in the application code and, in turn, contributes to reliability.

Appendix C: Technical terminology

Throughout this paper a number of technical terms are used. While most readers know what they mean, the goal is to ensure that any nuances are clear thus avoiding confusion.

data set: A collection of virtual units that comprise the critical volumes necessary for an application to do productive work.

direct storage interconnect: (unique to HP OpenVMS): An OpenVMS instance uses a direct storage interconnect to access storage without going through an OpenVMS storage server. The direct storage interconnect can be private, or it can allow shared access to the storage by multiple OpenVMS instances. Fibre Channel is the direct storage interconnect that allows shared access.

Disaster Tolerant Continuity Solution (DTCS): DTCS is one of HP Service's packaged offerings. It provides consulting, configuration, tools and procedures that ensure the proper functioning of long-distance disaster tolerant environments.

HP Volume Shadowing for OpenVMS [unique to HP OpenVMS and formerly known as Host-based Volume Shadowing (HBVS)]: HP Volume Shadowing for OpenVMS is the HP implementation of RAID 1 technology under the control of an OpenVMS host system. Another implementation of RAID 1 would be to use controller-based mirroring where the controller for the storage subsystem manages the duplication of data for a host system. Both Volume Shadowing and controller-based mirroring provide data redundancy, which increases the availability of data. A shadow set member can be composed of a controller-based RAID 1 device. Each shadow set member can be connected to a different controller type. Thus controller-based mirroring can be used in conjunction with Volume Shadowing to provide greater resilience and improve read performance. Data written by an application is directed to the Volume Shadowing virtual unit. Volume Shadowing manages the duplication of application data to the members of a virtual unit on a per I/O basis. The data is duplicated or written on multiple shadow-set members before status is returned to the application. Up to three devices can form a single virtual unit or shadow set. Volume Shadowing supports having 500 multiple member virtual units or 10,000 single member shadow sets. The shadow set members can be located at different physical sites for disaster tolerance purposes.

multi-protocol router: A multi-protocol router is a router that is able to use a variety of data formats and transports rather than just one. For example, a multi-protocol router may be capable of using both Fibre Channel protocol and IP. This is the case for the multi-protocol routers used in some of the reference architectures in this document.

quorum site: In any split-site solution there are a number of potential failure modes that are introduced simply by the fact that processing is now possible from more than one location. The problem is that some of the possible failures are difficult for the technology to differentiate, and can therefore cause the technology to make inappropriate decisions regarding how to continue, if they are configured for automatic continuation. For example, see "split-brain" syndrome below. One of the mechanisms used to mitigate these problems is called a "quorum site." This is a third location or virtual location, independent of the two primary processing data centers, that houses a single system that is a member of the cluster and is primarily used by the operating system to allow it to distinguish between two important classes of failure – the loss of a datacenter or of the inter-site communications between them. The fundamental idea of the quorum site is to provide an external reference point for the cluster members. In the event of a failure of the inter-site links between the primary data centers, the systems will attempt communication with the system at the quorum site. If only one site succeeds in this communication, then it is assumed to be the best site for continuation, and it will therefore resume processing. The site that fails to communicate will hang – ensuring processing continues only where coordination is possible. If both of the production sites can communicate with the quorum site then the nodes at one of the production sites will shut down – once again ensuring that only coordinated processing is possible. The quorum site is therefore a critical component of any split-site cluster where it is required to continue automatically after a site failure.

router and switch: A router is a computer-networking device that forwards data packets across an internetwork toward their destinations, through a process known as routing. Routing is the means of discovering paths in computer networks along which information (split up into packets) can be sent. In non-technical terms, a router acts as a junction between two networks to transfer data packets among them. A router is essentially different from a switch. A switch connects devices to a local network. One easy illustration for the different functions of routers and switches is to think of switches as roads connecting all homes in a city and routers as highways connecting the cities in a country. A router creates and/or maintains a table, called a "routing table" that stores the best routes to certain network destinations and the "routing metrics" associated with those routes. Routing is usually directed by routing tables, which maintain a record of the best routes to various network locations in order to keep up with the packet arrival rate.

save set: Save set is a term that represents the full complement of data that is backed up at any one time from a single site to another one site.

SCS interconnect (unique to HP OpenVMS): The OpenVMS instances that form an OpenVMS Cluster communicate with each other using System Communications Services (SCS). SCS communication is used for connection management (maintenance of the cluster membership list), locking, and other cluster communication services. When the OpenVMS instances are in the same hard partition, the SCS communication can occur through shared memory. When the OpenVMS instances are in separate hard partitions, one or more different SCS interconnects are required. The SCS interconnects available today are CI, DSSI, MEMORY CHANNEL, and network-based – the most versatile and most commonly used interconnect.

split-brain syndrome: This term describes the situation where two halves of a DT cluster each form a distinct instance of the same cluster. This frequently occurs when communication between DT sites is interrupted, causing each site to operate independently with the result that two different versions of the database are created. On resuming the communication between the sites, each database is either corrupted, out of synchronization, or comprised of the incorrect updates. The problem is generally avoided with the implementation of a quorum element (see definition above) in the configuration.
For more complex implementations, particularly over longer distances, DTCS provides resources to help avoid this problem.

Appendix D: Additional information

Industry organizations that focus on DT

Disaster Recovery Journal: www.drj.com

Contingency Planning and Management group:
www.contingencyplanning.com

Continuity Insights: www.continuityinsights.com

The IT Services Management Forum (itSMF):
http://www.itsmf.com/

Tabb Research report

Crisis in Continuity: Financial Markets Firms Tackle the 100 km Question
https://h30046.www3.hp.com/campaigns/2005/promo/wwfsi/index.php?mcc=landing_page&jumpid=ex_R2548_promo/fsipaper_mcc%7Clanding_page

US Federal Government

Interagency Paper on Sound Practices to Strengthen the Resilience of the U.S. Financial System
http://www.sec.gov/news/studies/34-47638.htm

HP-UX 11i

HP-UX 11i home page:
www.hp.com/go/unix

High availability solutions for HP-UX 11i:
www.hp.com/go/serviceguard

Disaster tolerant solutions for HP-UX 11i:
www.hp.com/go/dt

OpenVMS

HP OpenVMS home page:
www.hp.com/go/openvms

HP OpenVMS systems HA/DT:
www.hp.com/go/openvms/availability

HP OpenVMS software product description (SPD):
www.hp.com/go/spd
OpenVMS operating system #82.35.04
OpenVMS clusters #29.78.26

HP OpenVMS Cluster SPD 29.78.26:
http://h18000.www1.hp.com/info/SP2978/SP2978PF.PDF

"Fibre Channel in a disaster-tolerant OpenVMS Cluster System" white paper (Part Number 5982-9662EN):
http://h71028.www7.hp.com/ERC/downloads/5982-9662EN.pdf

"HP OpenVMS approach to high availability computing" white paper (Part Number 5983-0191EN):
http://h71028.www7.hp.com/ERC/downloads/5983-0191EN.pdf

Linux
Open Source and Linux from HP home page:
www.hp.com/go/linux

HP Serviceguard for Linux: www.hp.com/go/sglx

Cluster Extension XP: www.hp.com/go/clxxp

Windows 2003
HP and Microsoft home page: www.hp.com/go/windows
www.hp.com/go/clxeva

HP Integrity NonStop
HP Integrity NonStop computing home page:
www.hp.com/go/nonstop

HP Integrity NonStop business continuity and disaster recovery software: www.hp.com/go/nonstopcontinuity

HP Integrity servers
HP Integrity server family overview:
www.hp.com/go/integrity

Reliable Transaction Router
(RTR) home page: www.hp.com/go/rtr

Reliable Transaction Router documentation:
http://h71000.www7.hp.com/doc/rtr.html

HP Virtual Server Environment
HP VSE home page: www.hp.com/go/vse

HP Services
Business Continuity Services home page:
www.hp.com/go/businesscontinuity

To learn more, visit **www.hp.com**.

4AA0-6737ENW 9/2006

invent