Building No Single Point of Failure Clusters using HP Linux Technology

Eddie Williams

Senior Software Engineer
SteelEye Technology

eddie.williams@steeleye.com

HP WORLD 2003
Solutions and Technology Conference & Expo

No Single-Point-of-Failure

- What is a Single Point of Failure (SPOF)?
- What is the cost of a failure?
- What is the cost of a SPOF?
- Can you afford a SPOF?
- Methods to avoid a SPOF (storage focused)
- Building no SPOF clusters with HP and Linux

Single Point of Failure

A single element of hardware or software which, if it fails, brings down the entire computer system.

Source: In Search of Clusters, 2nd edition, Gregory F. Pfister

"entire computer system" - the computer system is not available to accomplish the task it is intended to perform.

Who cares?

The Enterprise does!

What is still needed to enable Linux acceptance in the Enterprise:

- “must provide error detection and diagnostics”
- “have the proven reliability characteristics of current Unix”
- “have good recoverability and error handling”

Source: Gartner, Spring Symposium, 2003

Cost of down time

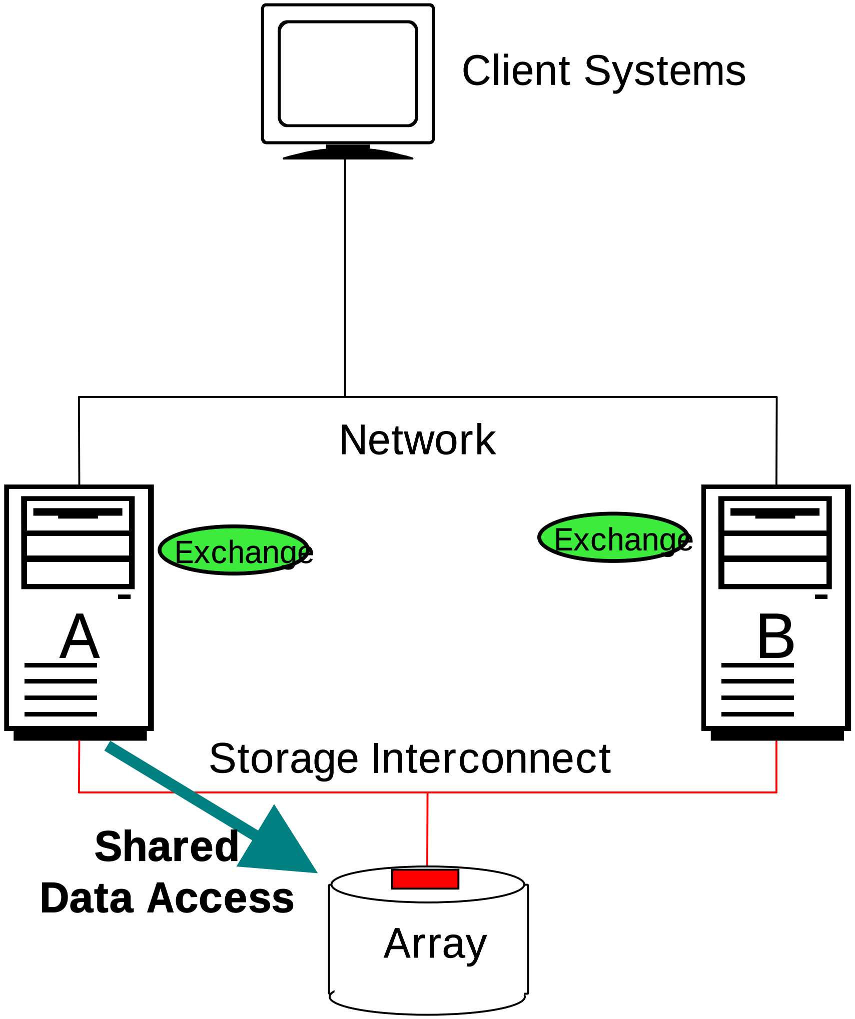
Business Operation	Average Cost per Hour of Downtime
Communications: Converged Services	> $10.0 million
Financial: Brokerage Operations	$6.45 million
Financial: Credit Card/Sales Authorization	$2.6 million
Media: Pay per view	$150,000
Retail: Merchandise Sales	$140,000
Transportation: Airline Ticketing	$89,500
Media: Event Ticket Sales	$69,000

Source: Gartner, Dataquest, Contingency Planning Research and Others

Case Study – Corporate Messaging System

- Goal: Provide protection for email system used by executives, sales, marketing.
- Shoe-string budget, used hardware “not being used” in the lab.
- Only local cluster initially implemented while high-speed WAN negotiated, configured and remote system set up.

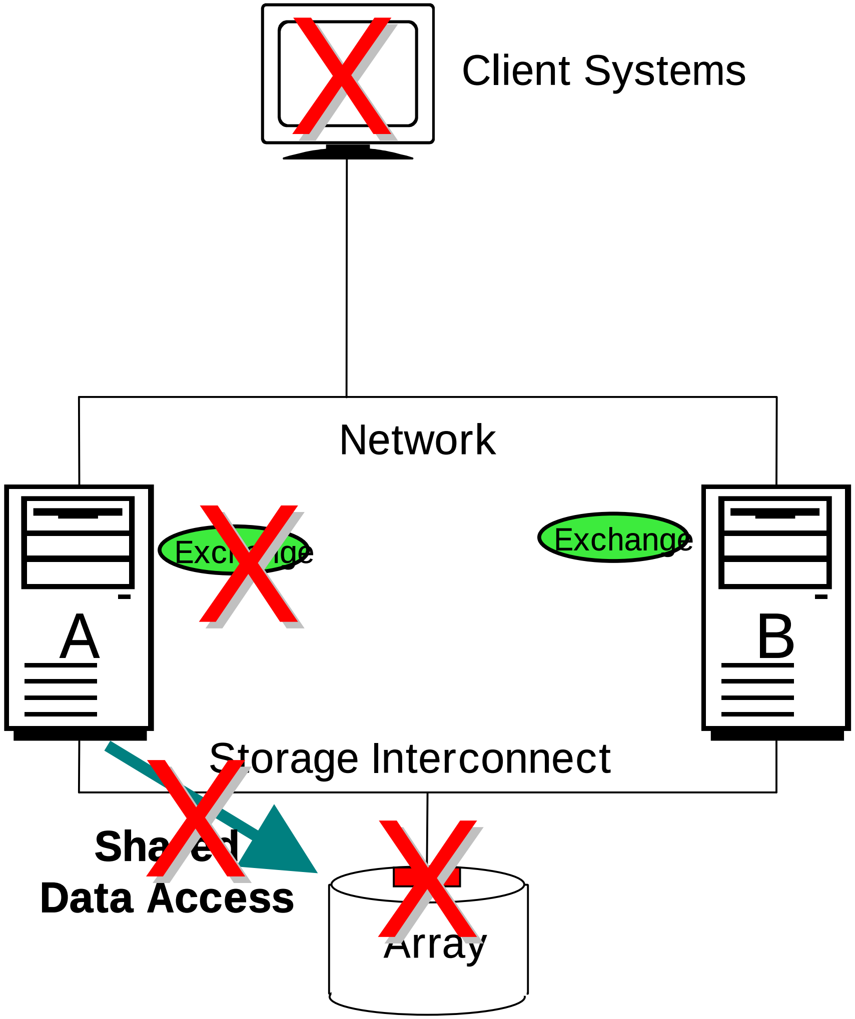
Case Study – Corporate Messaging System



SPOF

- Single array controller
- Parallel SCSI

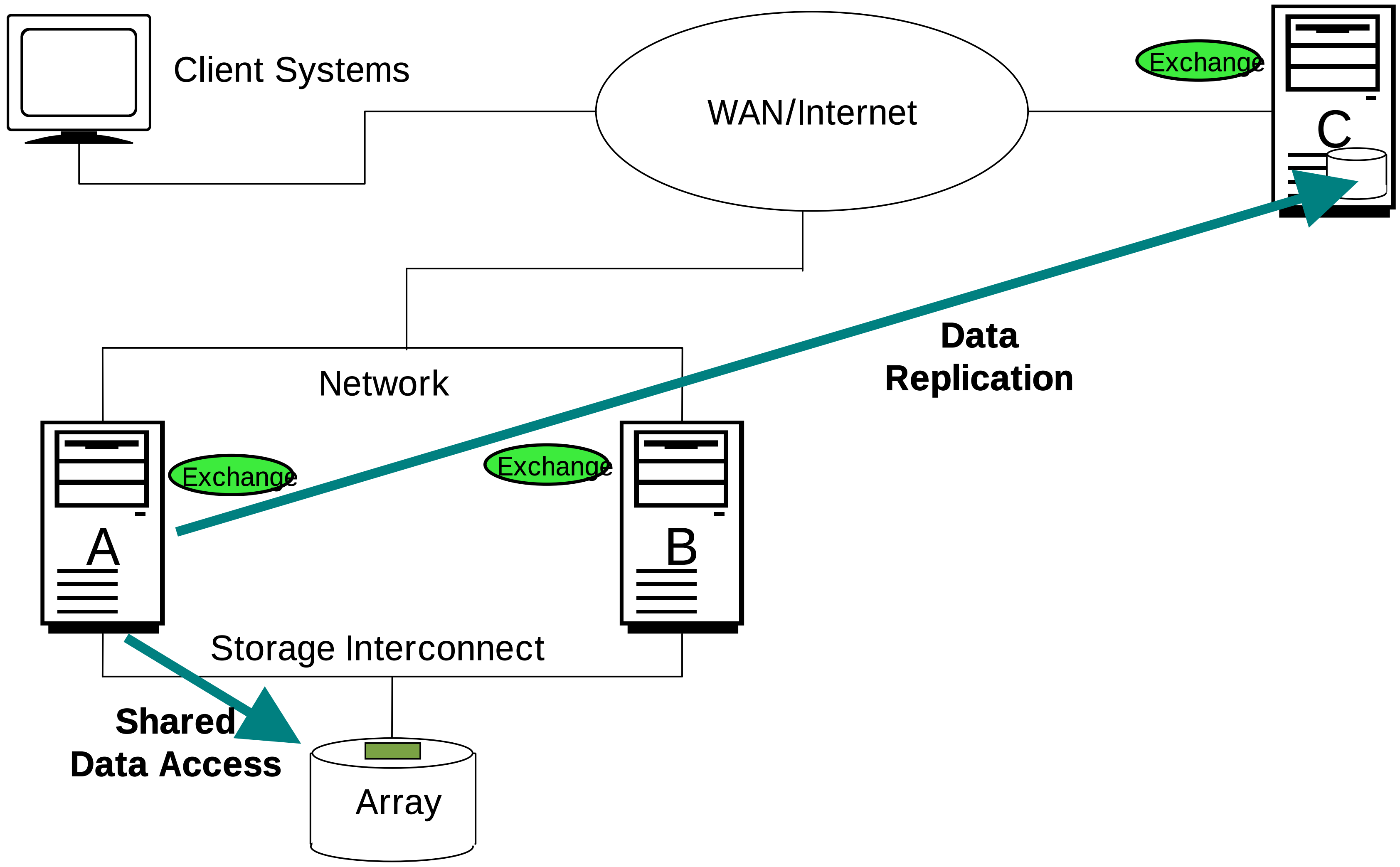
Case Study – Corporate Messaging System



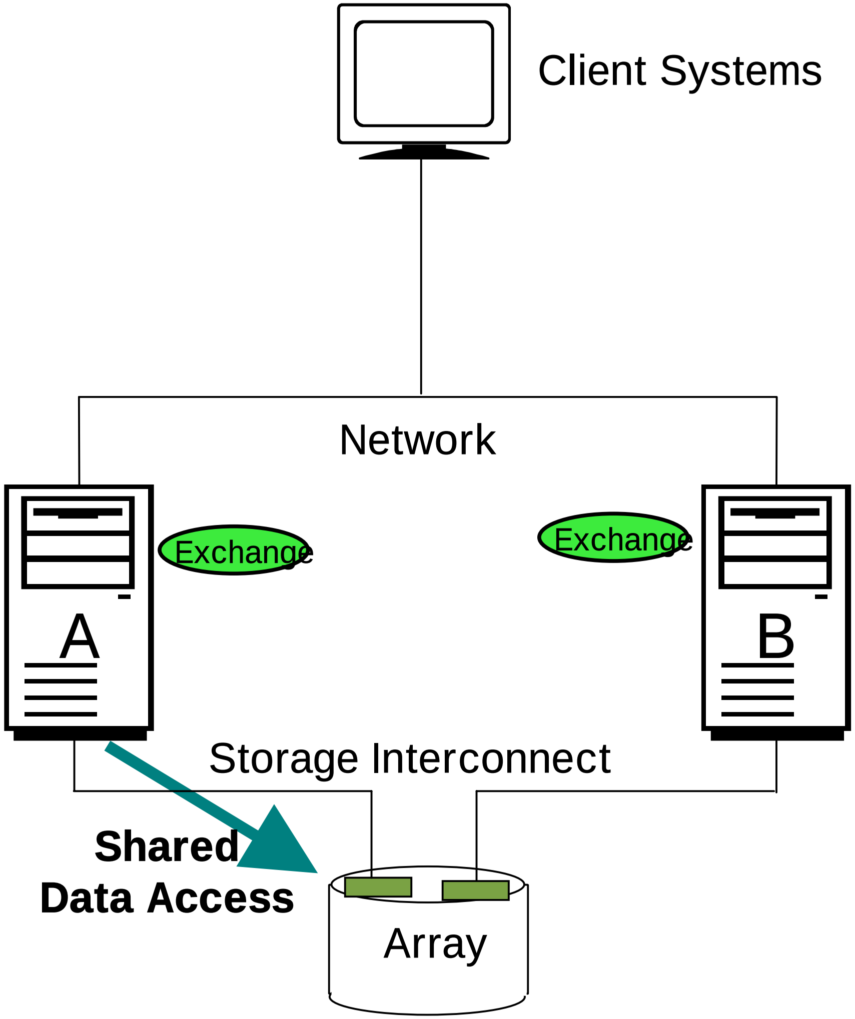
Failure analysis

- SPOF – array controller
- Cost
  - 14 hours of lost data
  - Unavailable one day
- Solutions
  - Data replication
  - redundant controllers

Case Study – Solution 1
Corporate Messaging System



Case Study – Solution 2
Corporate Messaging System



Can you afford a SPOF?

Do you know the cost of downtime?

- What is the personnel cost?
  - Administrator and support. Overtime? Lost work?
  - Personnel unable to complete tasks while down.
- What is the cost of lost goods or services?
  - Customers can not shop your web site.
  - Customers can not pay for goods at POS terminal.

If the cost of downtime is unknown, selling no SPOF will be difficult.

Methods to avoid SPOF

- Fault Tolerance
  - "A single stand-alone piece of computer hardware more or less bulletproof"
  - Expensive $$
- ✓ High Availability
  - Often defined by the number of "9s"
    - Class 4 (99.99%), about an hour of downtime **per year**
    - Class 5 (99.999%), about 5 minutes of downtime **per year**
  - The ability to recover from a single failure - "Sufficiently reliable to repair before something else breaks"
  - Less expensive but not cheap!

Source: In Search of Clusters, 2nd edition, Gregory F. Pfister

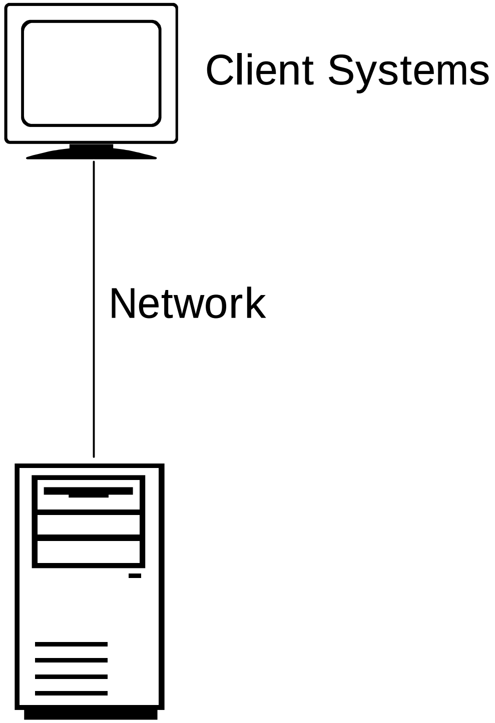
Availability definitions

- Good
  - Many single failures will result in services being unavailable for extended amount of time (> 1 hour)
  - Class of 9’s definition: ~2-3
- Better
  - Some single failures will result in services being unavailable for short amount of time (~ 1 hour)
  - Class of 9’s definition: ~4
- Best
  - Few failures will result in services being unavailable (~ 5 minutes)
  - Class of 9’s definition: ~5

Eliminating SPOF

- Additional hardware
  - Highly available hardware with redundant power supply, RAID for internal disks, etc.
  - Redundant networks
  - Specialized storage: SACS, MSA1000, and EVA
- Additional software
  - Clustering software: SteelEye LifeKeeper, Serviceguard for Linux, Red Hat Cluster Manager
  - Custom storage software: Secure Path

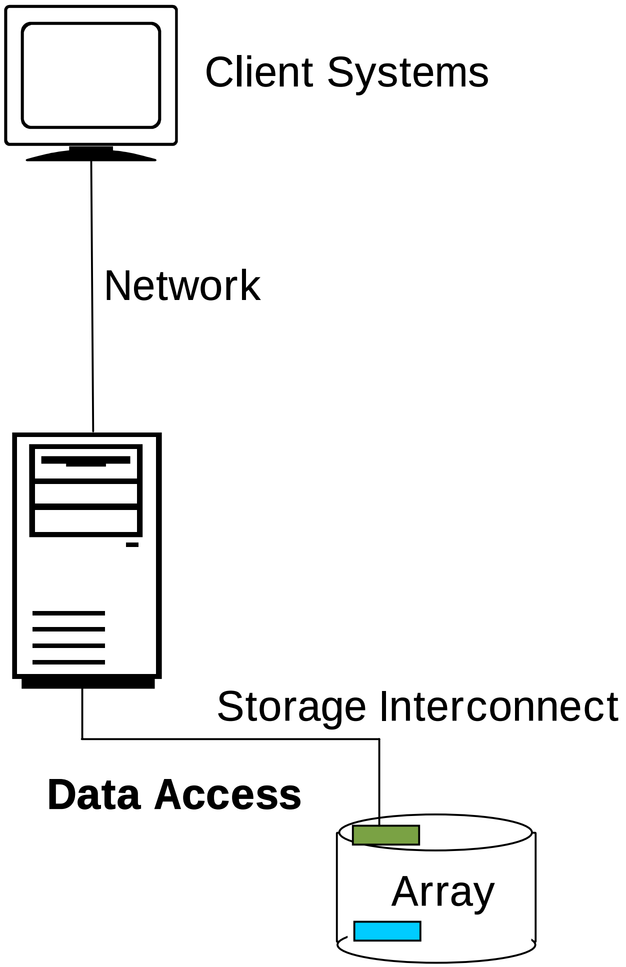
Simple configuration - server



- DL380 G3 – $4,488
  - 1 Intel Xeon 3.06GHz/1MB
  - 1 GB
  - ✓ Redundant NIC's
- DL380 G3 - $**5,095**
  - Battery back write cache
  - ✓ Redundant power supply/fans
- SPOF:
  - 5i Disk controller
  - PCI bus
  - OS
- Availability: Good

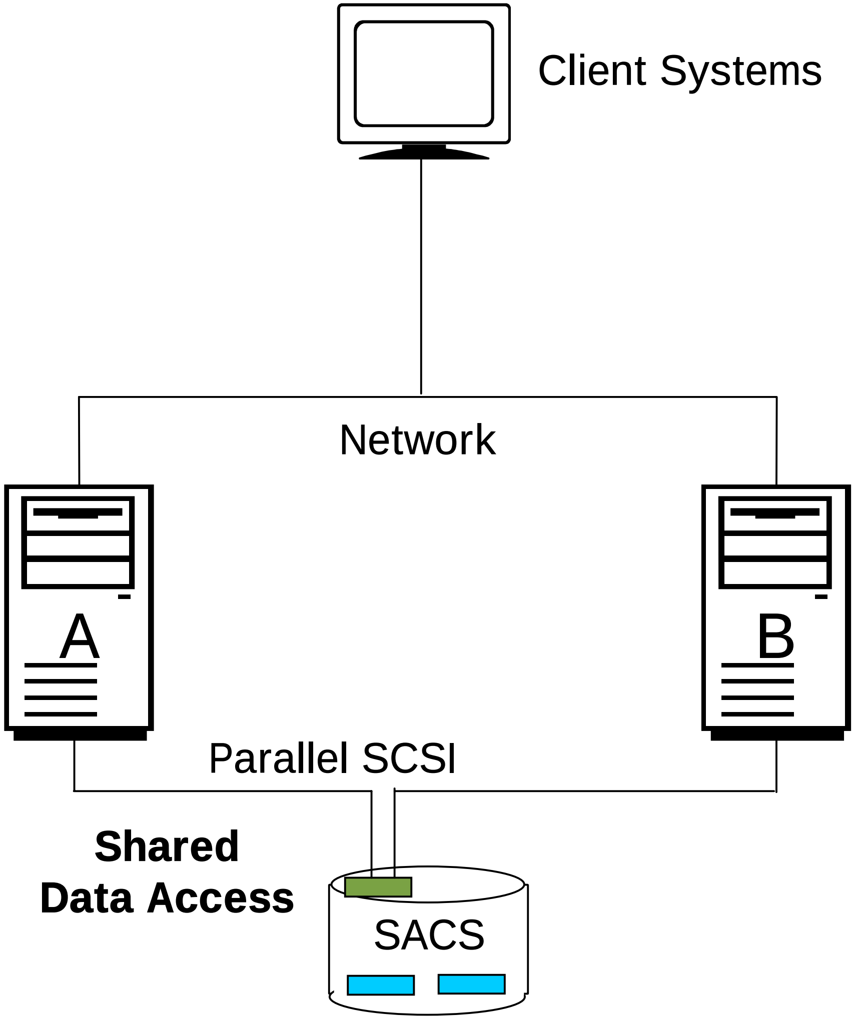
WARNING: all pricing from web, no apps, no services

Simple configuration – external storage



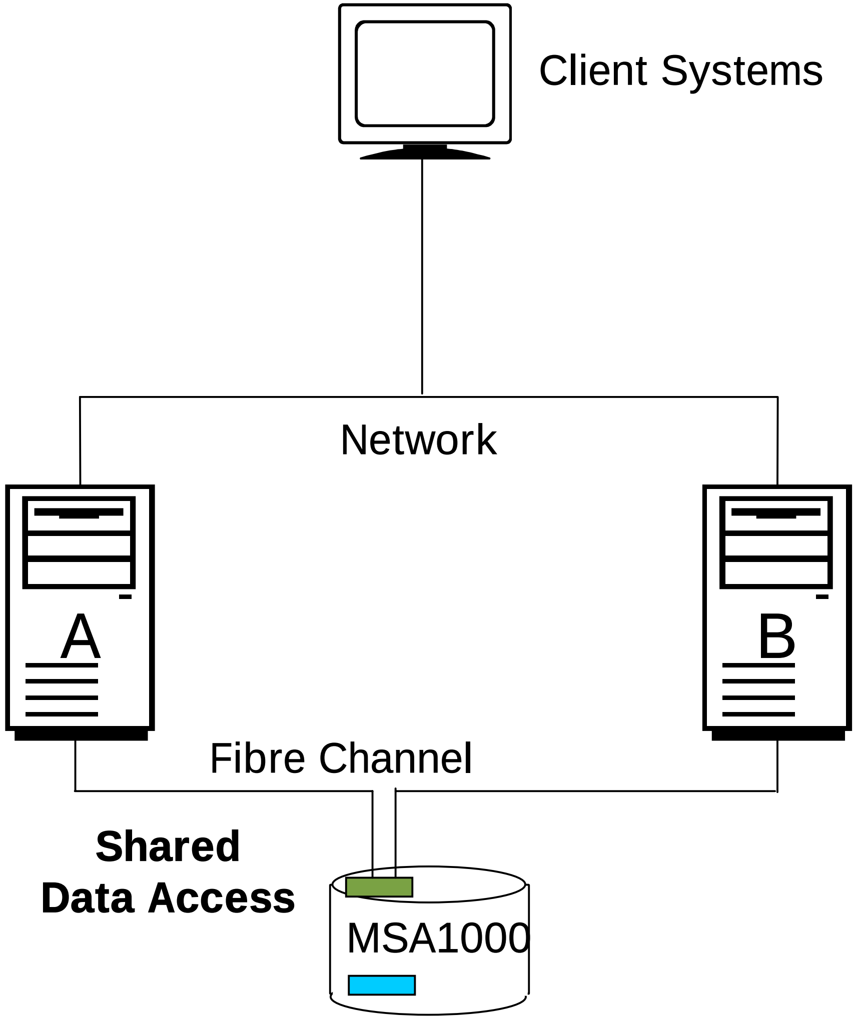
- DL380 G3/MSA – $16,680
- DL380 G3/SACS - $10,817
- SPOF:
  - 5i Disk controller/Qlogic
  - PCI bus
  - OS
  - Interconnect connection with storage
  - Array controller
- Availability: Good. Single failure causes loss of availability

Simple cluster configuration – Packaged Cluster



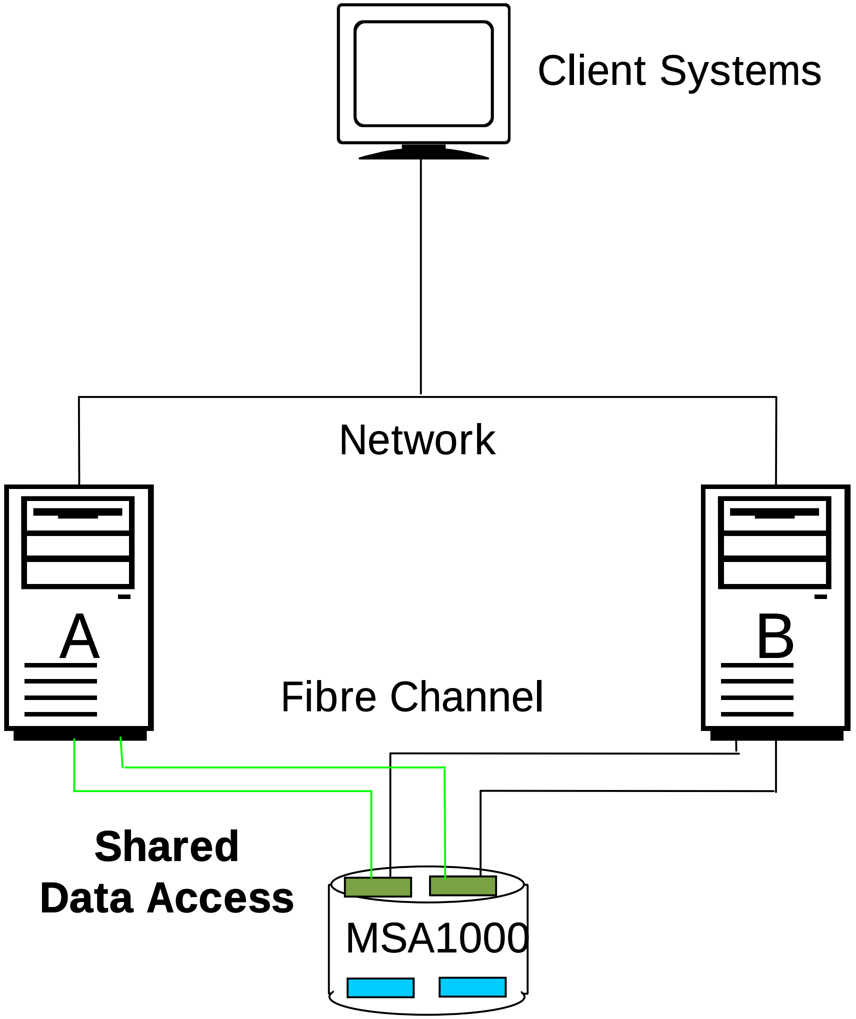
- DL380G3 PC $22,660
  - ✓ Redundant servers
  - ✓ Redundant controllers
- SPOF: EMU
- Availability: Better. Single path failure cause system failover!

Simple cluster configuration – MSA1000



- 2 DL380G3, MSA1000 $32,965
  - ✓ Redundant servers
- SPOF:
  - – Embedded switch
  - – MSA1000 controller
  - – MSA1000 backplane
- Availability: Better. Single Path failure cause system failover!

MultiPath cluster configuration – MSA1000



- 2 DL380G3, MSA1000 $52,040
  - ✓ Redundant controllers
    - Qlogic
    - Array
  - ✓ Redundant switches
  - ✓ Redundant paths
- SPOF:
  - – MSA1000 backplane
- Availability: Best. Single path failure will not cause failover.

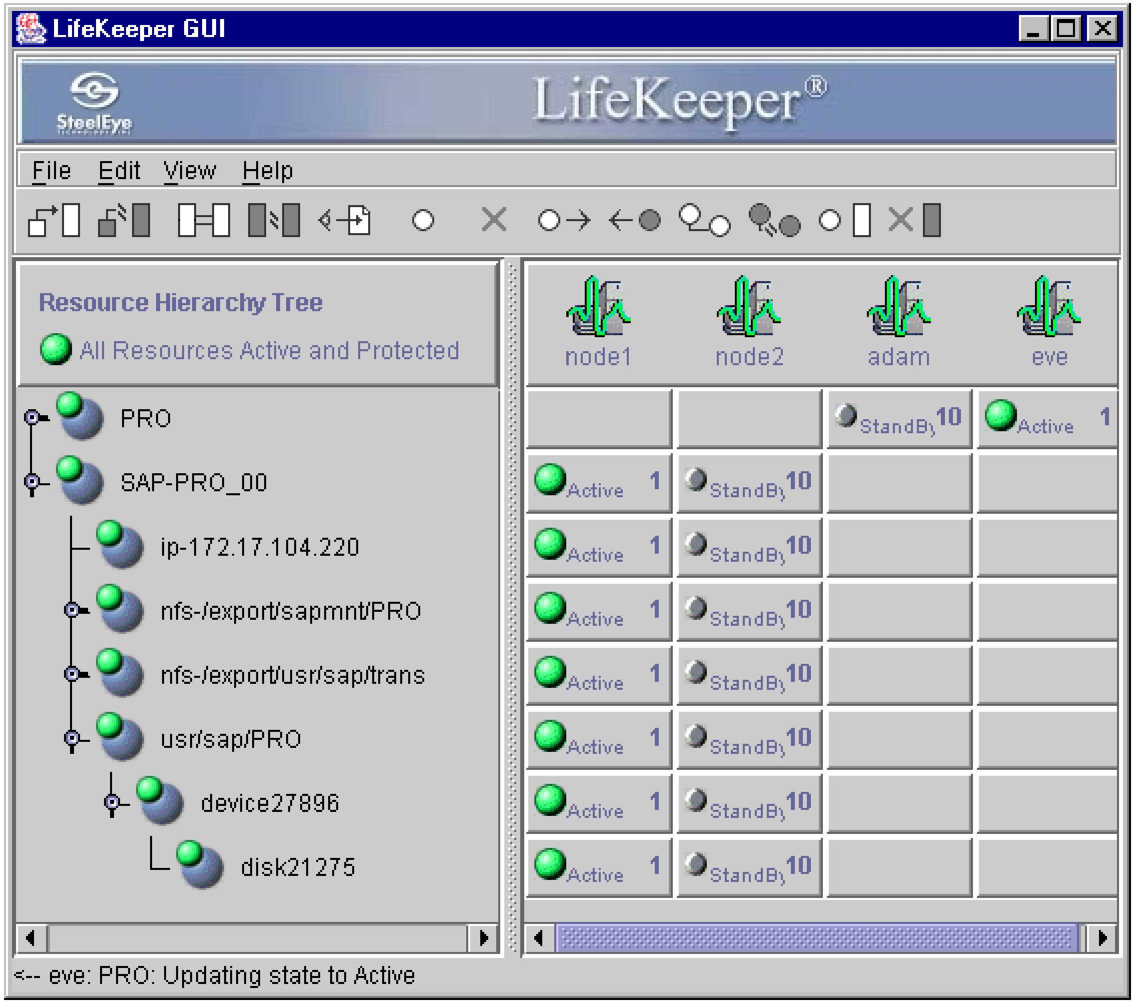
HP Solutions – Secure Path

- Allows independent Fibre Channel fabric paths
  - StorageWorks dual-controller RAID systems
  - Servers equipped with multiple HBAs
- Monitors each path
  - Reroutes I/O on failure
  - Monitors failed paths to detect restoration
  - Restores access to repaired paths, if desired
- Detects failures reliably without inducing false or unnecessary failovers
- Avoids failover/restore thrashing

HP Solutions – Secure Path plus SteelEye LifeKeeper

- Adds monitoring of entire system including
  - Network interfaces
  - Disk subsystems
  - Applications
- Integrated solution provides end-to-end HA protection
  - SteelEye and HP perform extensive integration testing assuring that cluster is installable, stable and reliable.

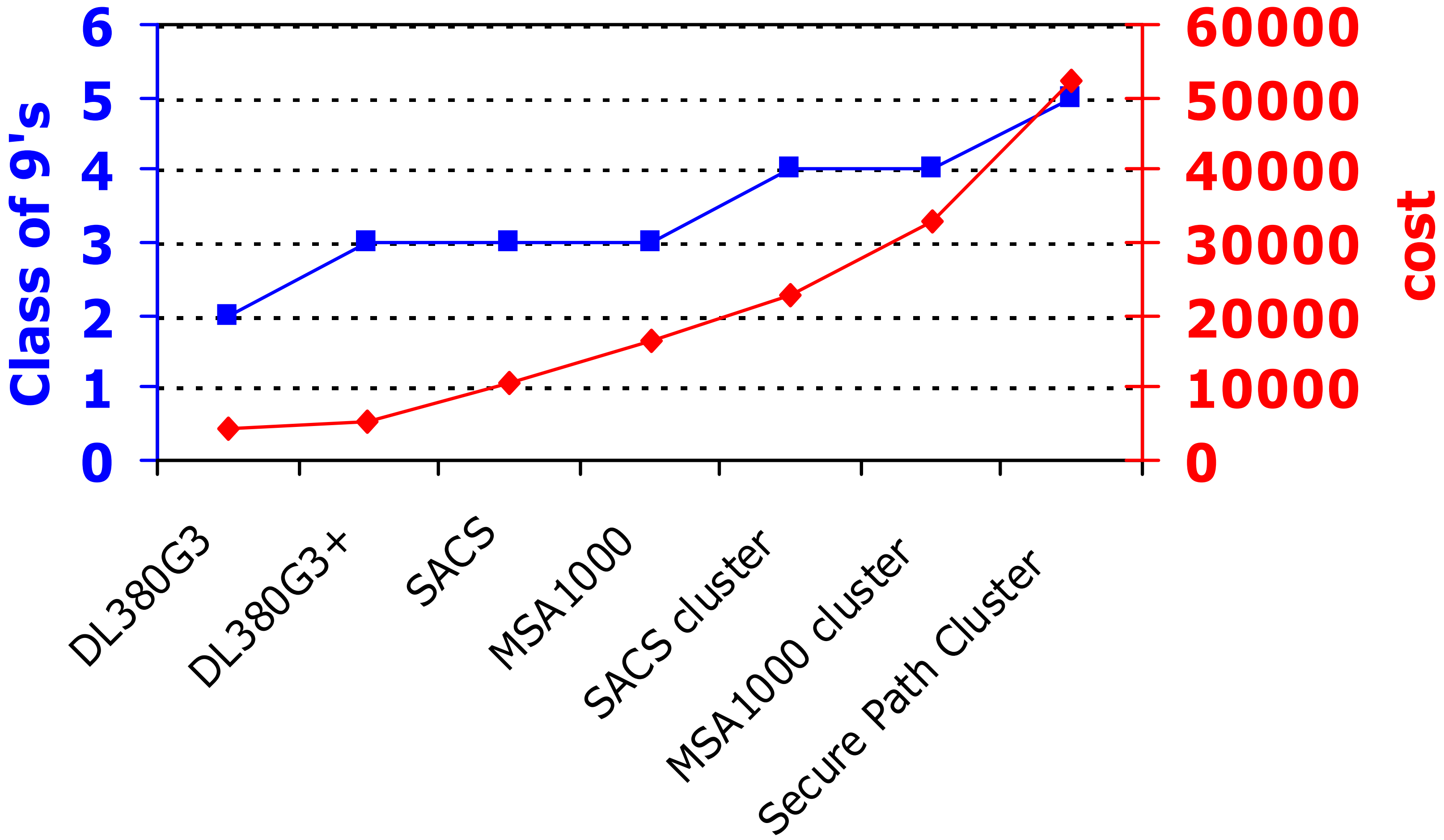
HP Solutions – Secure Path plus SteelEye LifeKeeper



Resource Hierarchy Tree All Resources Active and Protected	node1	node2	adam	eve
PRO			StandBy 10	Active 1
SAP-PRO_00	Active 1	StandBy 10		
ip-172.17.104.220	Active 1	StandBy 10		
nfs-/export/sapmnt/PRO	Active 1	StandBy 10		
nfs-/export/usr/sap/trans	Active 1	StandBy 10		
usr/sap/PRO	Active 1	StandBy 10		
device27896	Active 1	StandBy 10		
disk21275	Active 1	StandBy 10		

<-- eve: PRO: Updating state to Active

Summary


Class of 9's
6
5
4
3
2
1
0
60000
50000
40000
30000
20000
10000
0
cost
DL380G3
DL380G3+
SACS
MSA1000
SACS cluster
MSA1000 cluster
Secure Path Cluster

Conclusion

As with any business decision, deciding how to architect your critical application environment for HA requires a cost/benefit analysis.

Understanding the cost of your downtime and the range of options for deploying an HA solution are critical to performing this analysis.

HP is quickly moving into a position with Secure Path, MSA1000, EVA, ProLiant Servers, Services and SteelEye LifeKeeper to start cracking the vaunted Enterprise mission-critical market.

HP WORLD 2003
Solutions and Technology Conference & Expo

Eddie Williams

Senior Software Engineer
SteelEye Technology

eddie.williams@steeleye.com

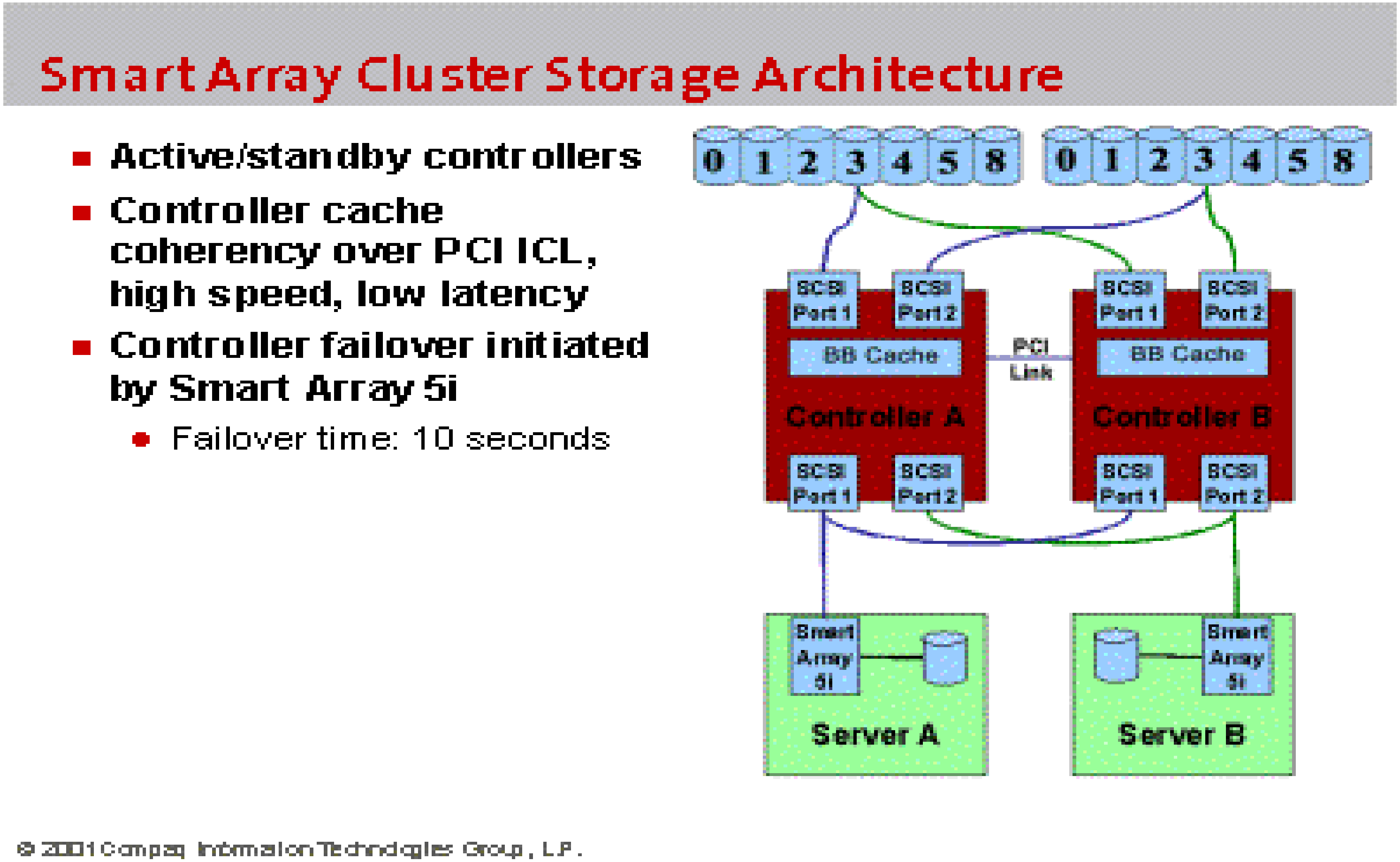
Interex, Encompass and HP bring you a powerful new HP World.

Backup slides

HP Solutions – Packaged Cluster

- Designed for small to medium-sized businesses
- 2 nodes
- 8U packaging
- Built-in hardware redundancy
  - Power supplies
  - Fans
  - Buses
  - Controllers in storage

HP Solutions – Packaged Cluster

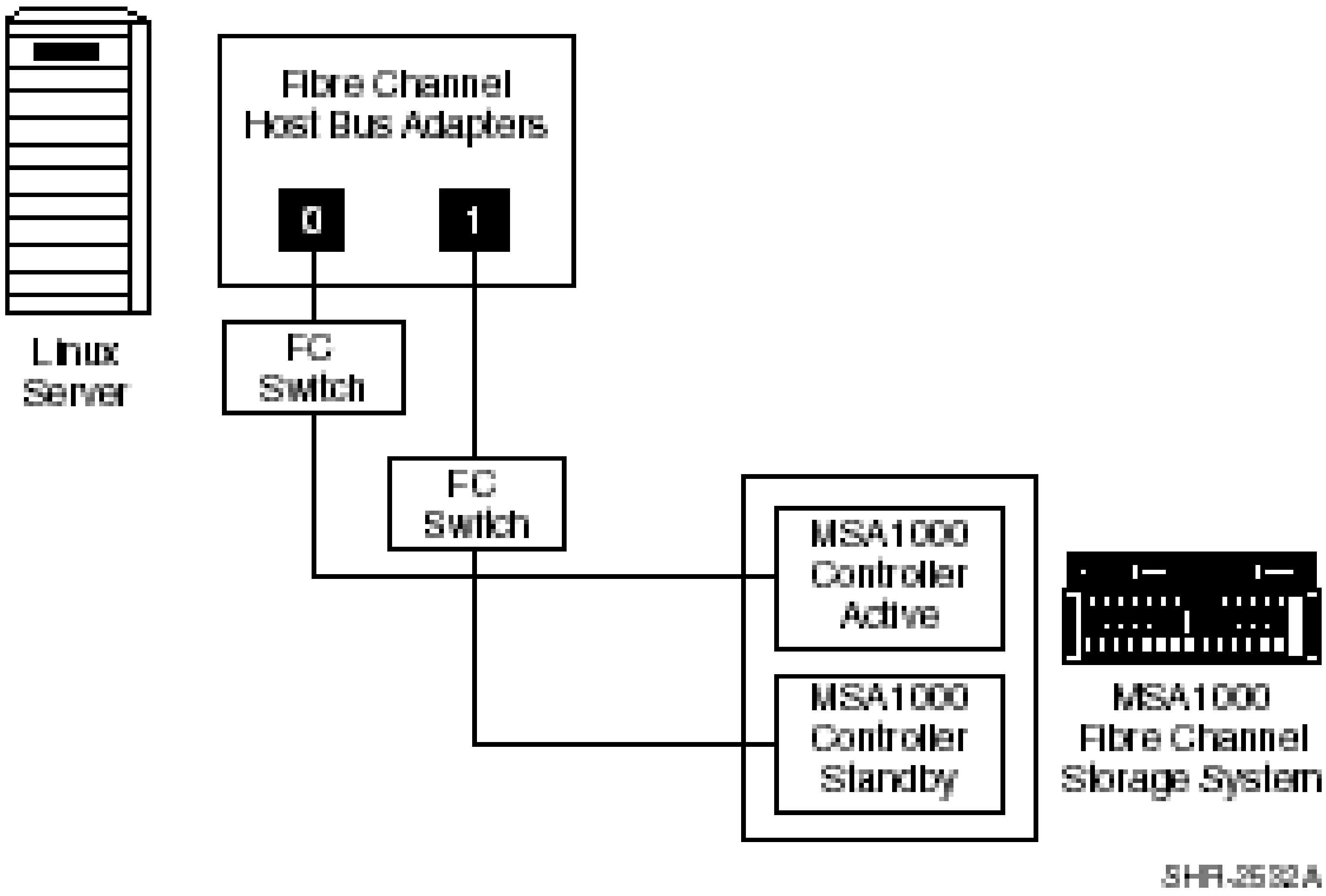


Smart Array Cluster Storage Architecture

- Active/standby controllers
- Controller cache coherency over PCI ICL, high speed, low latency
- Controller failover initiated by Smart Array 5i
  - Failover time: 10 seconds

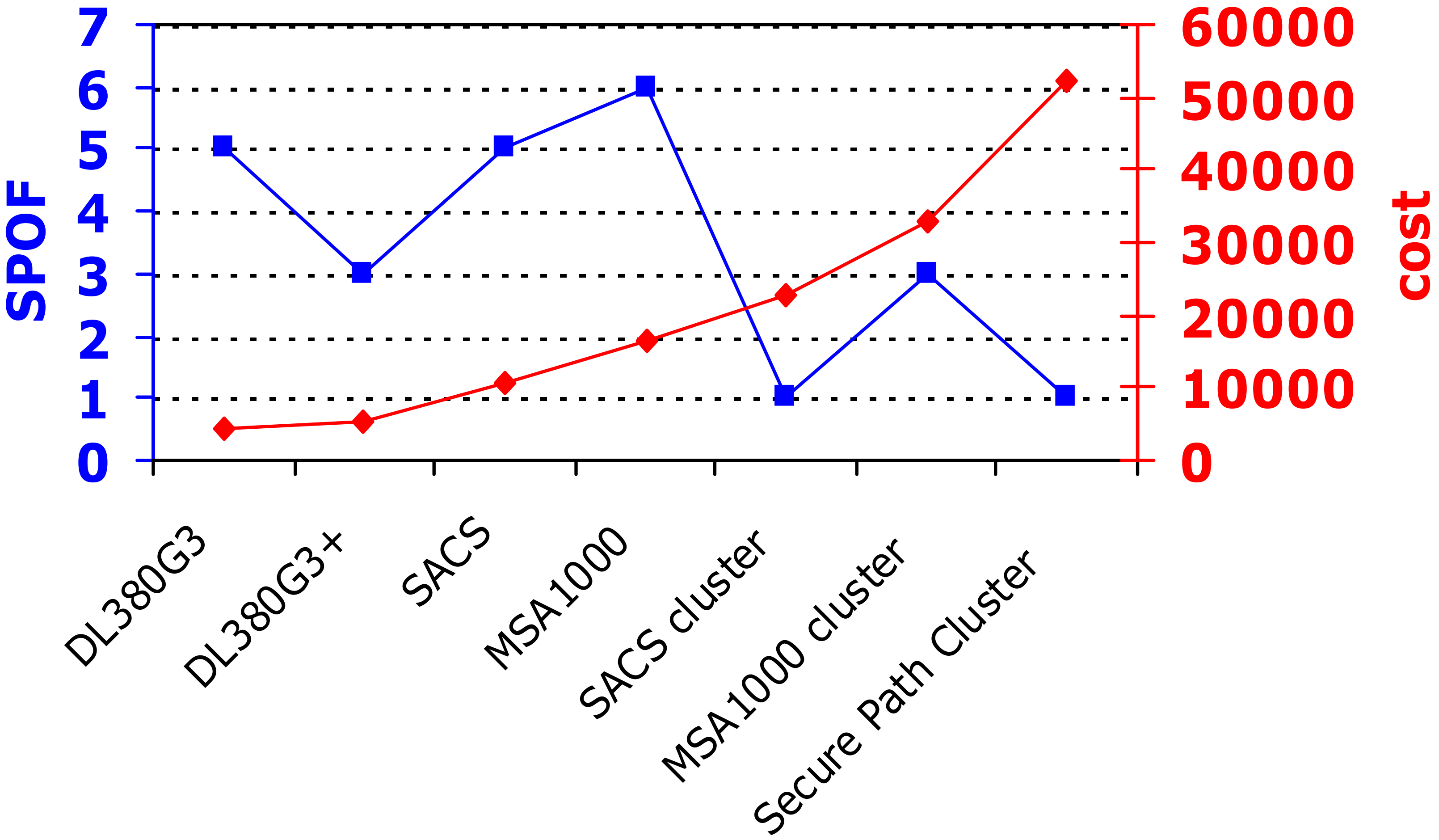
0 1 2 3 4 5 8
0 1 2 3 4 5 8
SCSI Port 1
SCSI Port 2
BB Cache
Controller A
SCSI Port 1
SCSI Port 2
PCI Link
SCSI Port 1
SCSI Port 2
BB Cache
Controller B
SCSI Port 1
SCSI Port 2
Smart Array 5i
Server A
Smart Array 5i
Server B

HP Solutions – Secure Path



Linux Server
Fibre Channel Host Bus Adapters
0
1
FC Switch
FC Switch
MSA1000 Controller Active
MSA1000 Controller Standby
MSA1000 Fibre Channel Storage System
SHR-2582A

SPOF versus cost comparison


SPOF
cost
DL380G3
DL380G3+
SACS
MSA1000
SACS cluster
MSA1000 cluster
Secure Path Cluster