High Availability vs. Fault Tolerant in a Storage Environment

Leslie Thomson

Mission Critical Solutions Center Gold Technical Acct Mgr.

Greg Tinker

Business Recovery Specialist Hetro O.S. SAN

Introduction

- ## Leslie Thomson
  - Currently work in the Mission Critical Solution Center
  - Worked at Digital/Compaq/HP for 12 years
  - Regional Support Person, Consultant in the field and Support Center
  - Background in Statistics and Computer Science

- ## Greg Tinker
  - Currently work as Business Recover Specialist (BRS)
  - Worked at HP for 5 years
  - Background, Math, Physics/Dynamics, and UNIX internals.

" I paid for a fault tolerant environment, but I don't feel I have one.".''

Mission Critical Customer

Availability in "nines"

- **Availability**   **Annual Downtime**

Availability	Annual Downtime
99%	87.6 hours
99.9%	8.76 hours
99.99%	52.5 minutes
99.999%	5.25 minutes

Importance of Availability

- **Industry Sector   Hourly Cost of Downtime**
- Manufacturing $28,000
- Transportation $90,000
- Retail, Catalog Sales $90,000
- Retail, Home Shopping $113,000
- Media, Pay Per View $1,100,000
- Banking datacenter $2,500,000
- Financial, Credit Card Processing $2,600,000
- Brokerage $6,500,000

Statistics Don't Lie People Do

- Using $28,000/hour cost of downtime

- Assuming 250 operating days per year at 8 hours per day (2000 hours per year)

- Cost of downtime at 99.9% availability:

  2000 hours/year x .1% x $28,000/hour = $56,000

- Cost of downtime at 99.999% availability: 2000 hours/year x .001% x 28,000/hour = $560

- Total savings = ($56,000/year - $560/year) x 5 years = $277,200
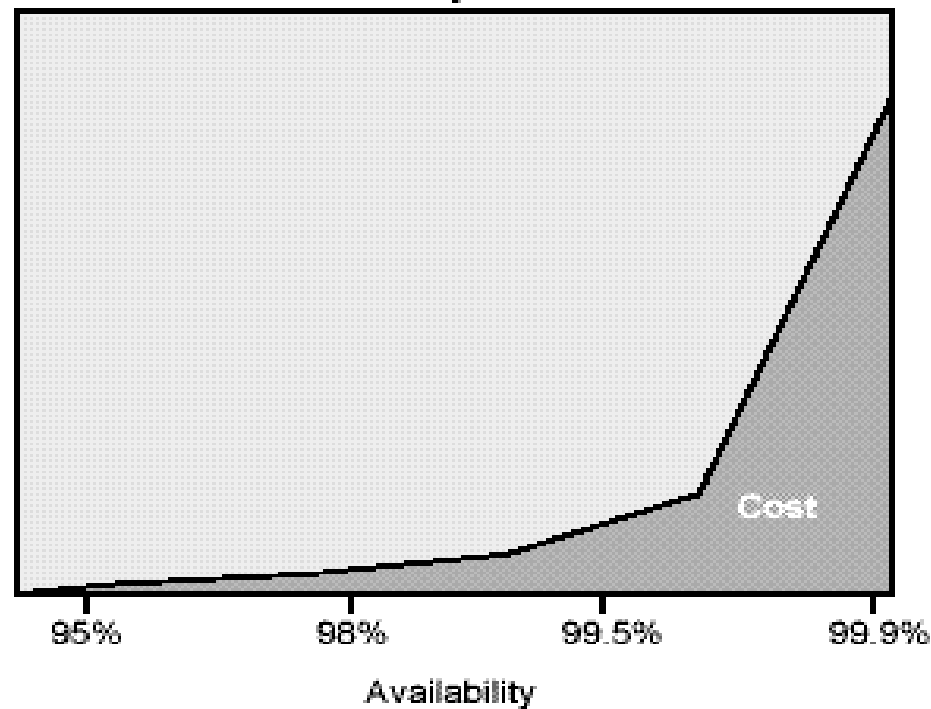
Fault Tolerance Defined

- Complete redundancy across *all* hardware components

- Fault-tolerant systems "operate through" a component failure

- No single point of failure can compromise system availability

Why isn't everyone Fault Tolerant?

- **High hardware costs**
- **Complexity and expense of writing software**

Cost of Fault Tolerance

Availability Cost Curve

Source: Source: Gartner Research

Do you need Fault Tolerance or High Availability?

- "When you really, truly have to have your computers up all the time, what you need is not high availability but fault-tolerant computing. "

- Federal Express slogan: "When it absolutely, positively has to be there overnight?"
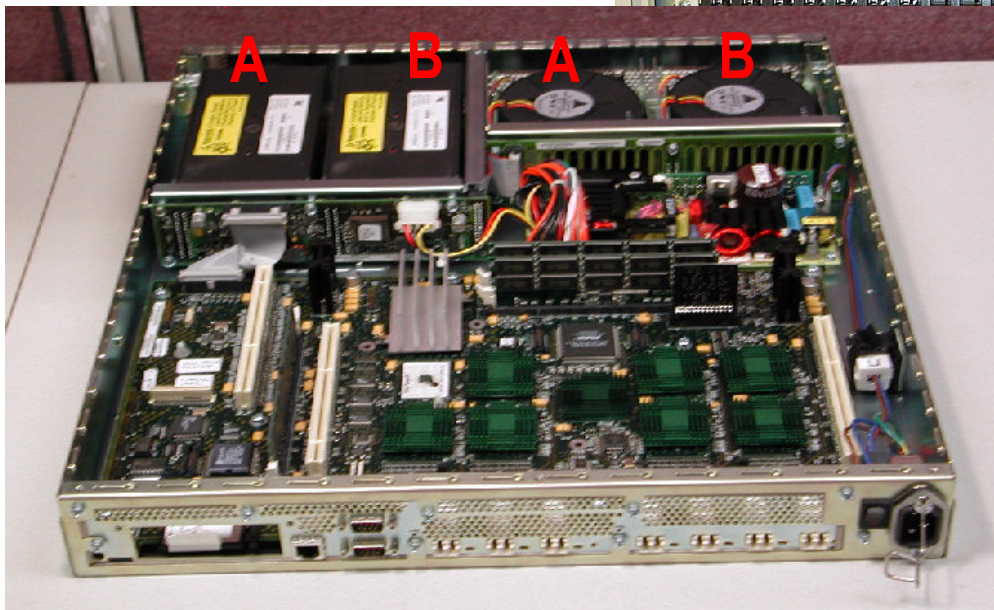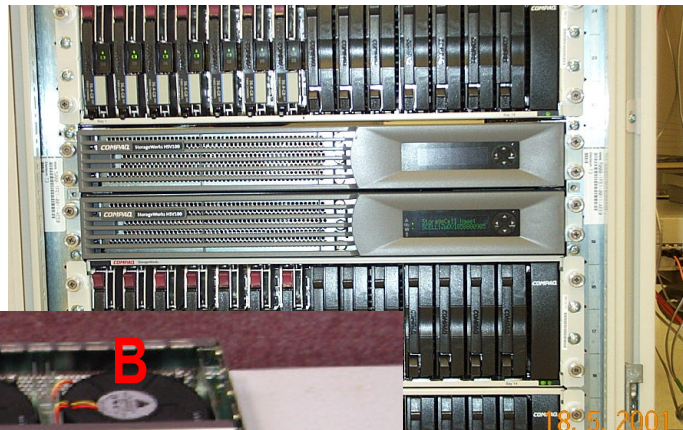
High Availability Solutions
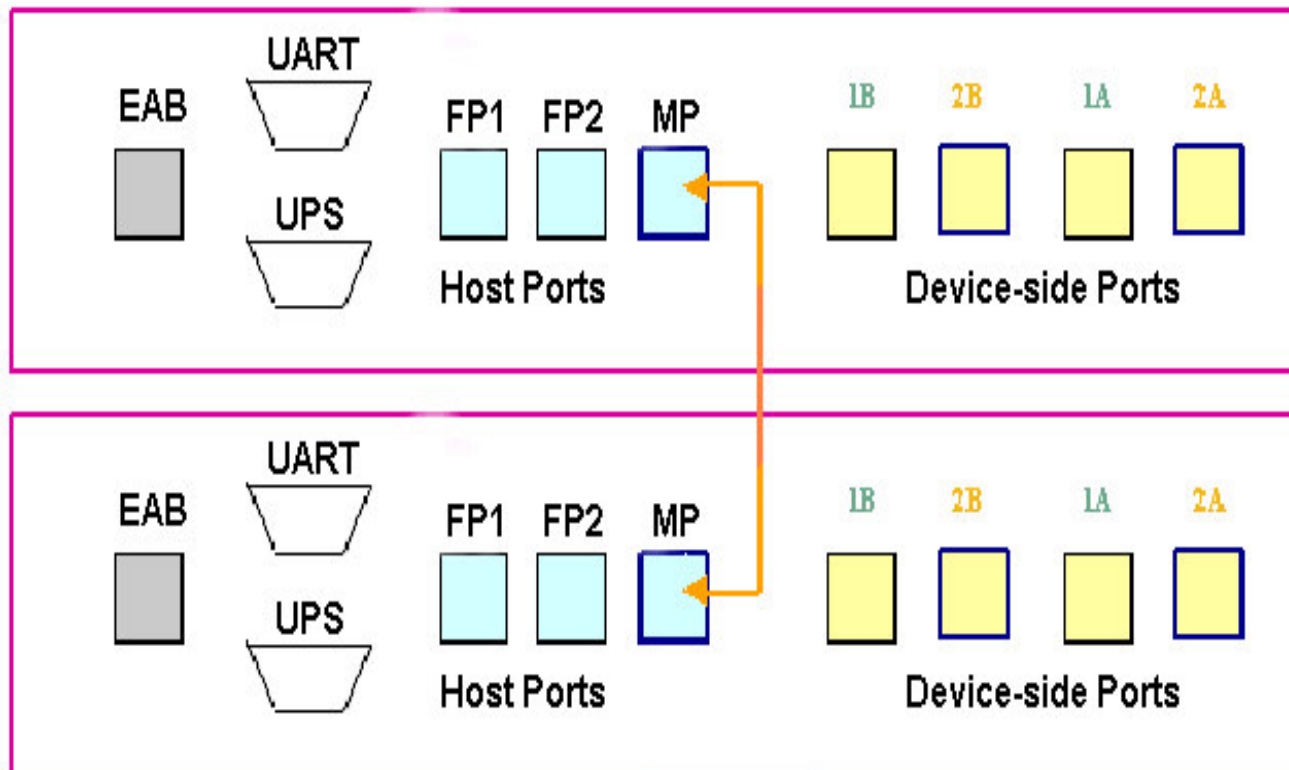
- Dual Fabrics
- Choose carefully fabric topology
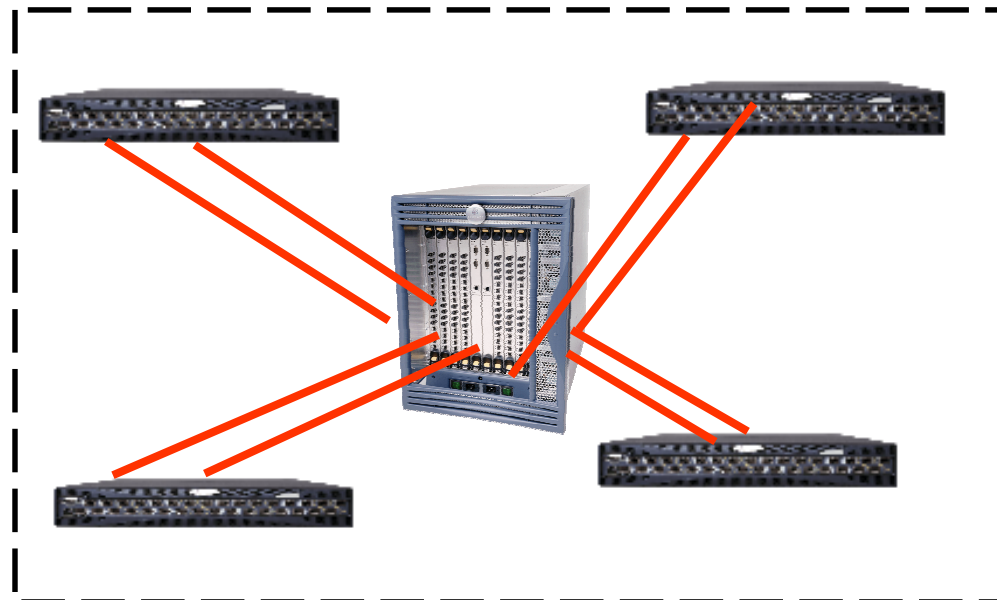
High Availability Features of the EVA

– Dual Controllers
  • Dual Blowers
  • Dual Cache Battery Modules

Mirror Port Connection between Controllers

FC Loop Switches

FC loop switches establish a fault-tolerant physical loop topology (two loops) by connecting between controllers and I/O modules A and B of drive enclosures
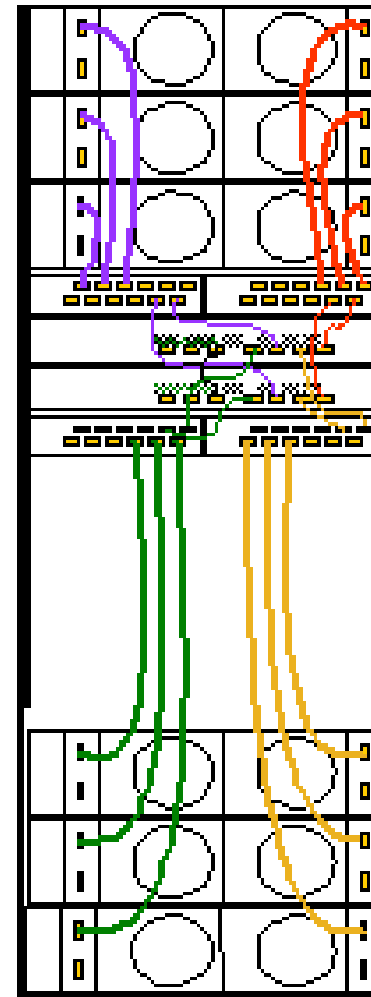
- **Four 1U, 1/2 rack wide 12-port switches**

  - Cables are no longer daisy chained though drive enclosures

  - Improved back end performance for random I/Os on large configurations, which will be more important with 2 Gb/sec host connections

  - Easier to troubleshoot - isolate to disk enclosure & disk level

  - Improved shelf replacement/addition
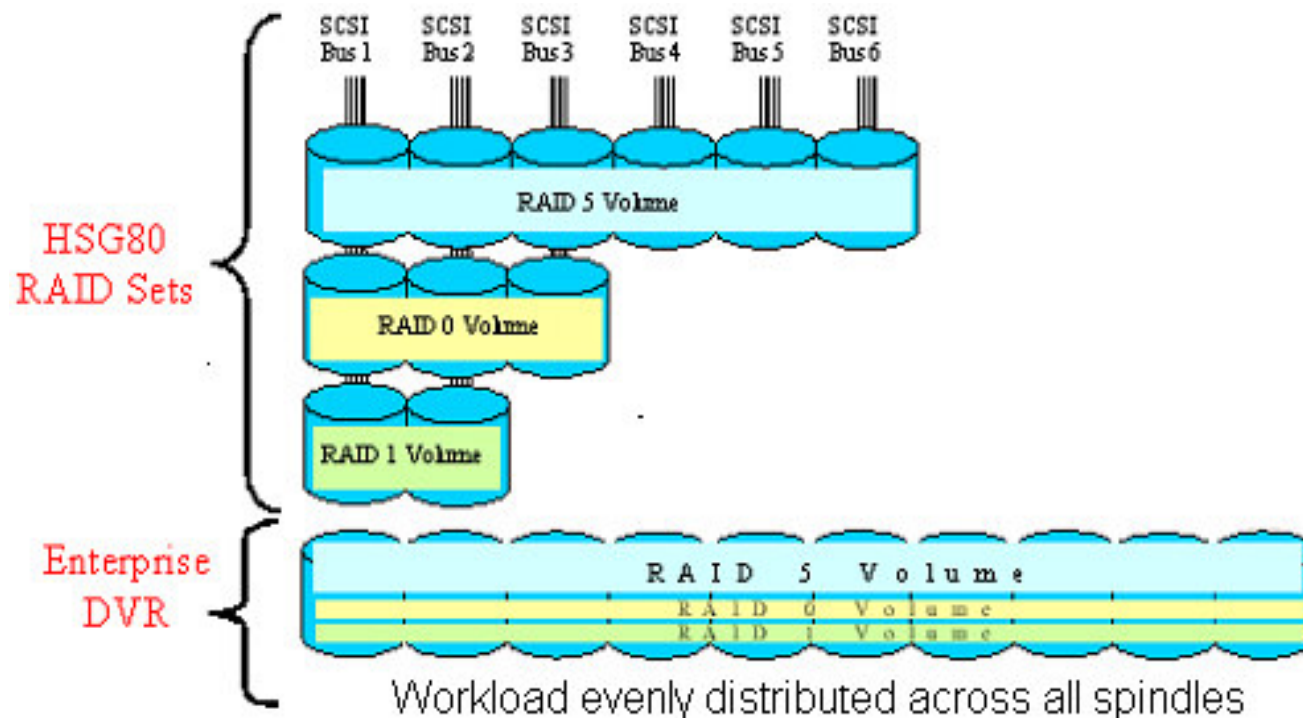
Loop 2B

Loop 2A

Loop 1B

Loop 1A

Virtual Disk Protection

- None
  - VRAID0 (striping)
- Moderate
  - VRAID5 (striping with parity)
  - Always five (4+1) physical disks per stripe are used
- High
- VRAID1 (mirroring)
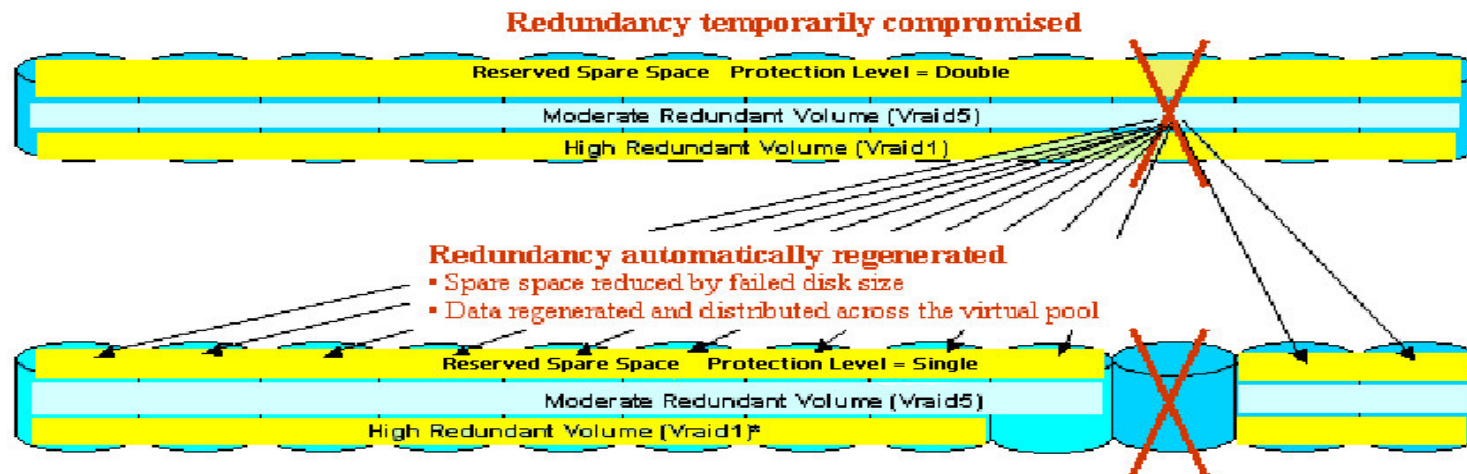- Two physical disks per mirror are used

Disk Sparing

– Spare space is allocated across all members of a Disk Group

– **You assign the amount of protection capacity you want to reserve:**

  • When you create the Disk Group

  • After the Disk Group has been created

– **Disk Failure protection level:**

  • None

  • Single (default)

    – Requires the space of TWO disk drives per Disk Group

  • Double (In most instances, the two disk failures cannot be concurrent failures)

    – Requires the space of FOUR disk drives per Disk Group

– **Check requested space versus actual space**

  • Note: If you have a Disk Group with Single disk protection that's getting full, and then request Double, requested will show Double and actual will still be Single

Redundant Storage Sets (RSS's)

RSS's make it possible within larger Disk Groups to lose multiple disks concurrently without losing data

- VRAID5 - one per RSS (two in rare situations)
- VRAID1 – one or more per RSS
- VRAID0 – No Disks

Redundancy temporarily compromised

Reserved Spare Space   Protection Level = Double
Moderate Redundant Volume (Vraid5)
High Redundant Volume (Vraid1)

Redundancy automatically regenerated
- Spare space reduced by failed disk size
- Data regenerated and distributed across the virtual pool

Reserved Spare Space   Protection Level = Single
Moderate Redundant Volume (Vraid5)
High Redundant Volume (Vraid1)*

Snapclone Implementation

- Traditional snapshot at creation time

  - The cloning process continues in the background

- Full background copy from the original virtual disk

- Virtual disk becomes its own entity

- Clone can be set to a different "Preferred" controller after the cloning process has completed

- VCS V2.x or higher will allow you to clone a Virtual Disk from one Disk Group to a different Disk Group

- You will need the appropriate snapshot license loaded to enable Snapcloning

Enterprise Metadata – Quorum Disks

Cell or System Level Metadata (Quorum Disks) contains:

- Controller Information
- WWN & Cell name
- Character map of Disk Groups and Virtual Disk members

– **When a cell is created, there will be 5 quorum disks in the default Disk Group**

– **Maximum of 16 quorum disks (1 per Disk Group)**

– **When a new Disk Group is created:**

- One quorum disk will be created on it
- One quorum disk will be removed from the Default Disk Group – until there is just one quorum disk remaining

– **Pre-allocated space (.03%) on each disk in all Disk Groups**

Continuous Access EVA

- This storage-based HP StorageWorks product consists of two or more EVA storage systems performing disk to disk replication, along with the Continuous Access management user interface that facilitates configuring, monitoring and maintaining the replicating capabilities of the storage systems.

- With this you are approaching a Fault Tolerant Environment.

Hewlett-Packard Surestore XP SAN

Concepts of HA approaching Fault Tolerant SANs.

Greg Tinker

High Availability Approaching Fault tolerant Solutions for SAN

- Duplex Write cache

- Cluster concepts for online code exchanges, and high availability.

- Multi-pathing software ensures maximum uptime and performance.

- All critical components are redundant and hot-replaceable to provide extreme reliability and availability, including: processors, I/O interfaces, power supplies, batteries and control processors on HP's XP storage and Director class switches.

High Availability Approaching Fault tolerant Solutions for SAN

- Fabric Director class, vs Edge switches.

- Dual Path fabrics, Zoning deliver Fault tolerant or HA environments?

- Disk Array Mirroring: Array to Array, and Internal array using array groups.
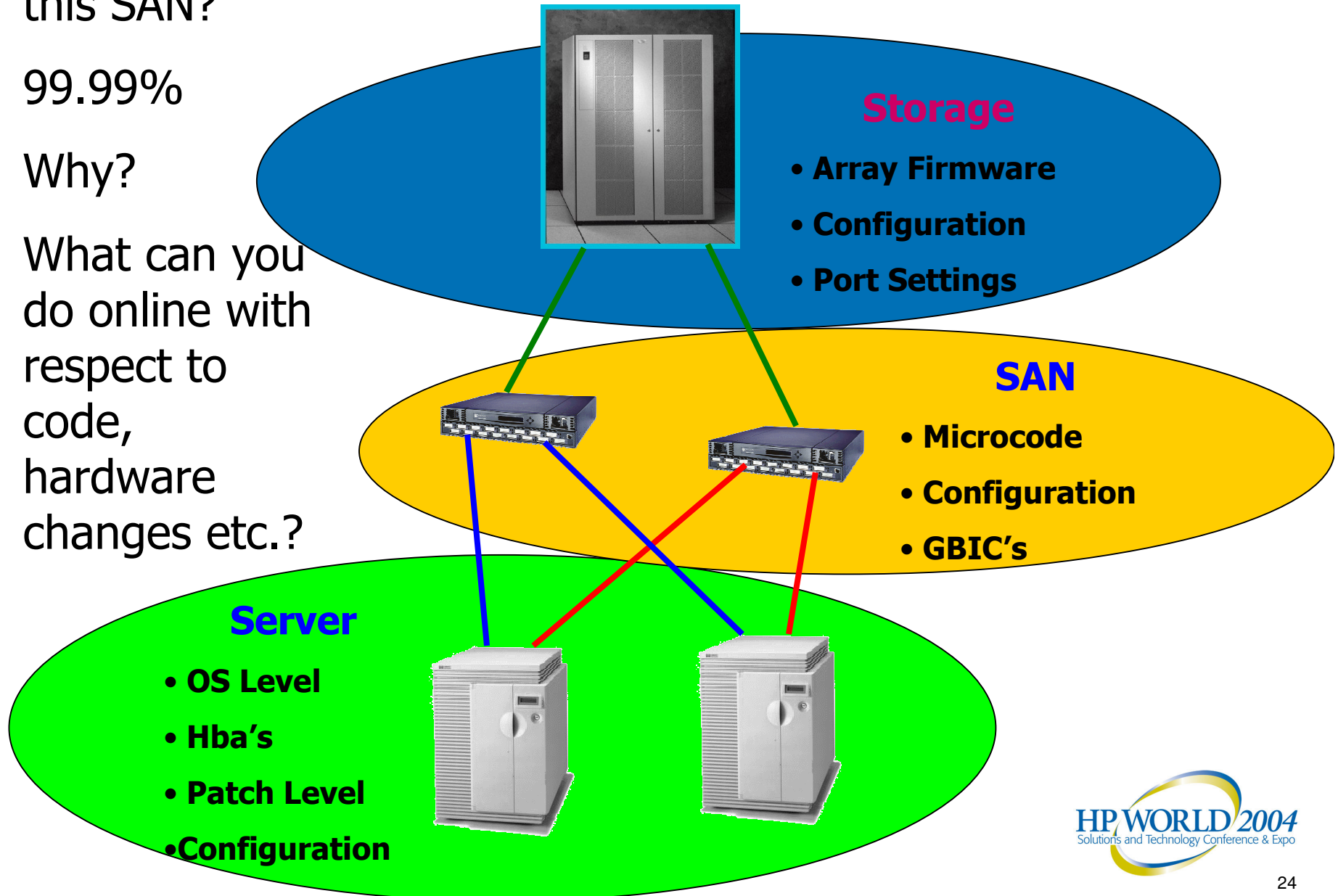
- Overview

Cluster Concepts

What's the availability of this SAN?

99.99%

Why?

What can you do online with respect to code, hardware changes etc.?

Storage

- **Array Firmware**
- **Configuration**
- **Port Settings**

SAN

- **Microcode**
- **Configuration**
- **GBIC's**

Server

- **OS Level**
- **Hba's**
- **Patch Level**
- **Configuration**

Multi Path Concepts: IO products

- Autopath, SecPath, PowerPath, DMP, MD, and others exist.

Does this allow Fault Tolerant or just HA?

- Most of the aforementioned allow round robin, others include Shortest Service Time (SST), Shortest Queue Time (SQT), Extended Round Robin (ERR)

The load balancing policies allow IO's (The completion of a framing sequence) to be distributed across multiple paths. Allowing high performance W/HA. Not Fault Tolerance.

Multi Path Concept

- coordinates access to disks that are reachable over redundant paths

- for the disk class driver, there is only one disk visible (whereas without AutoPath, two or more disks would be visible)
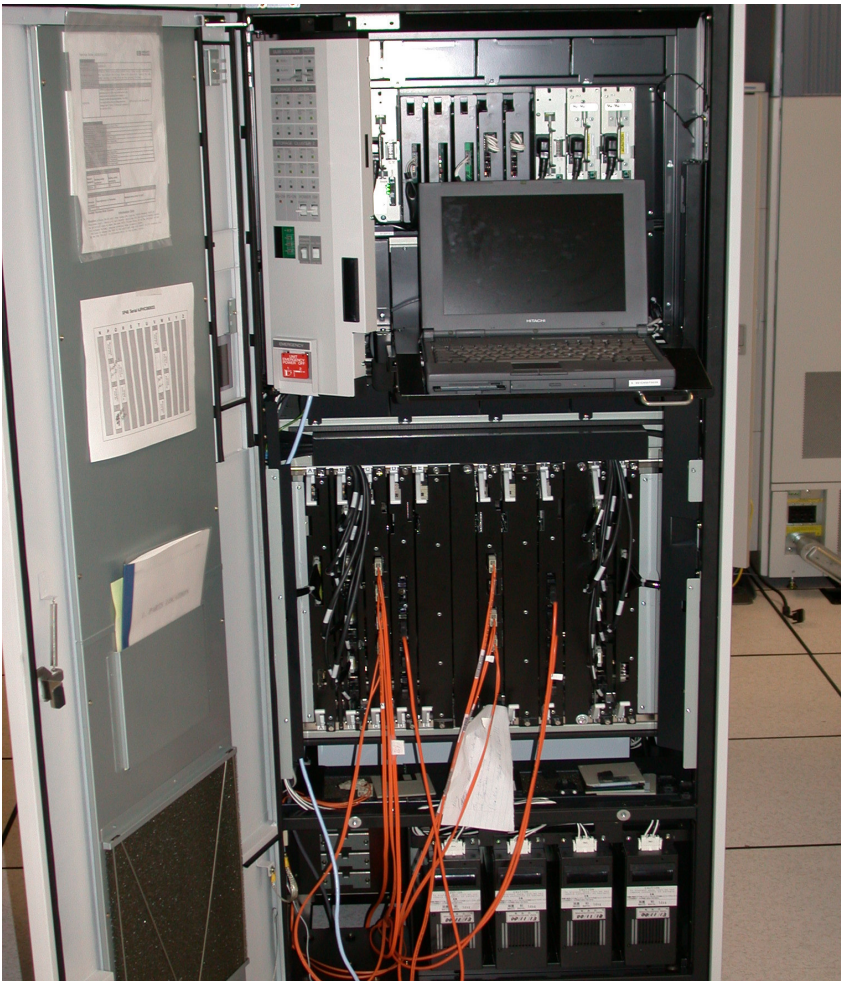
Node
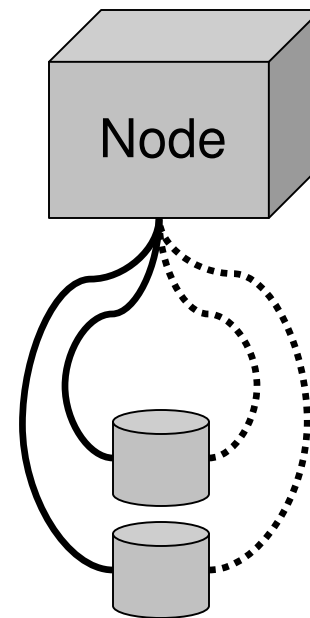
Critical Components are Redundant

- Array's
  - Mirrored write cache
  - Duplicate hardware
  - Cluster concepts

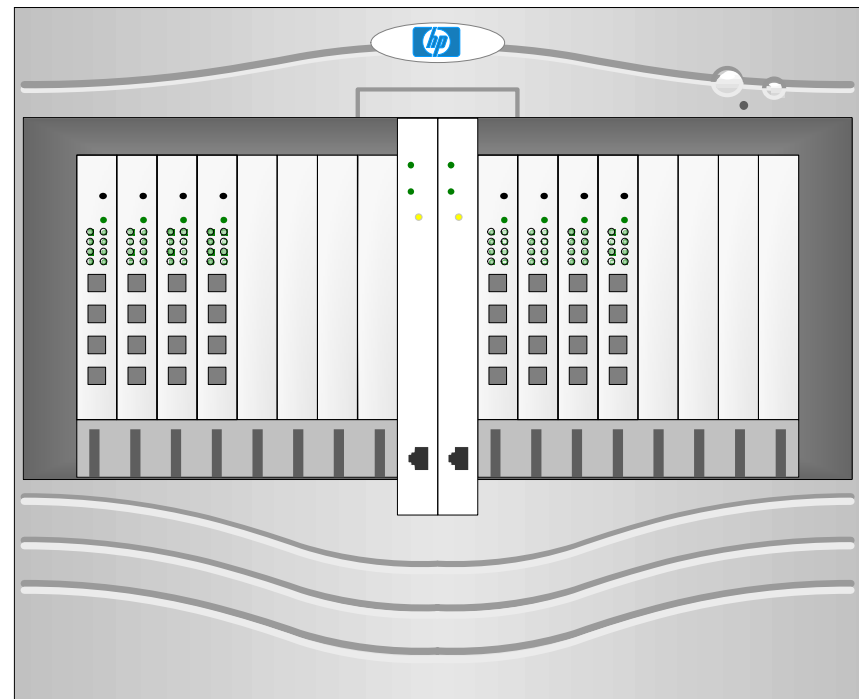- Director Class Switches
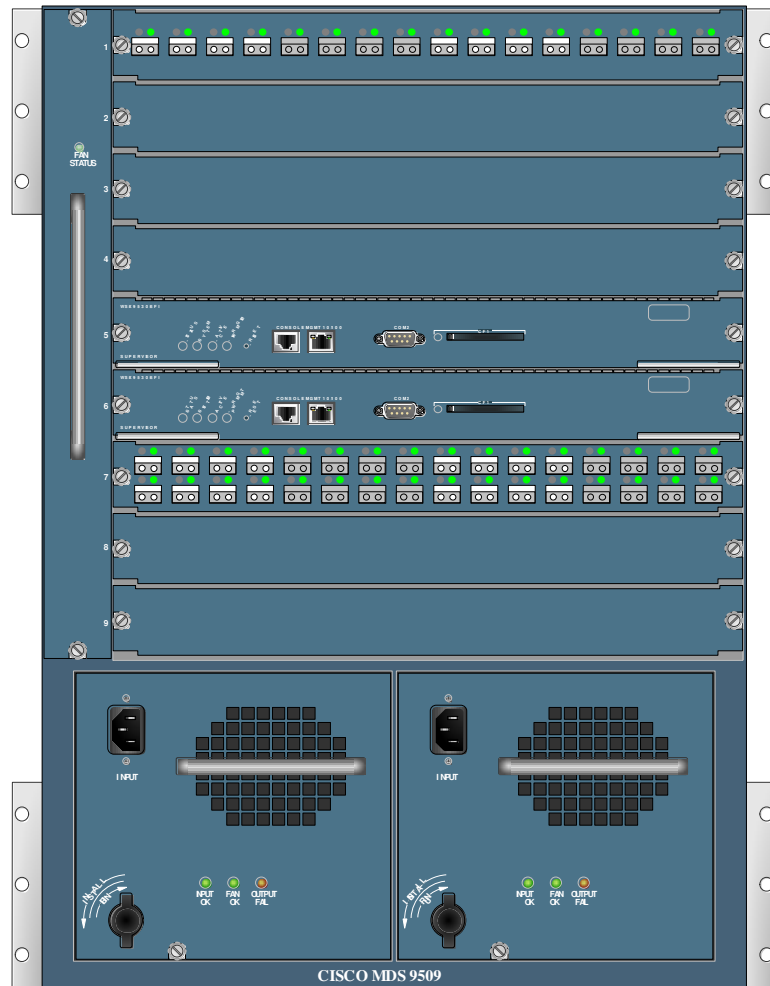 - Xbar concepts, with mirrored IO's

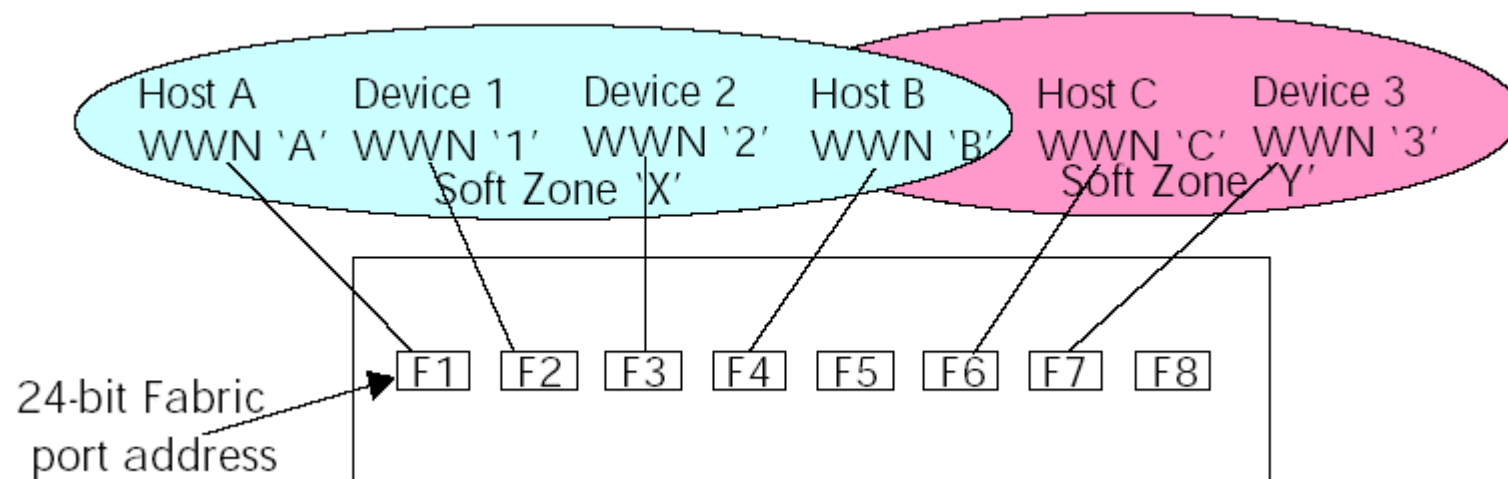# Critical Components are Redundant

Array's

Critical Components are Redundant
Fabric Director's
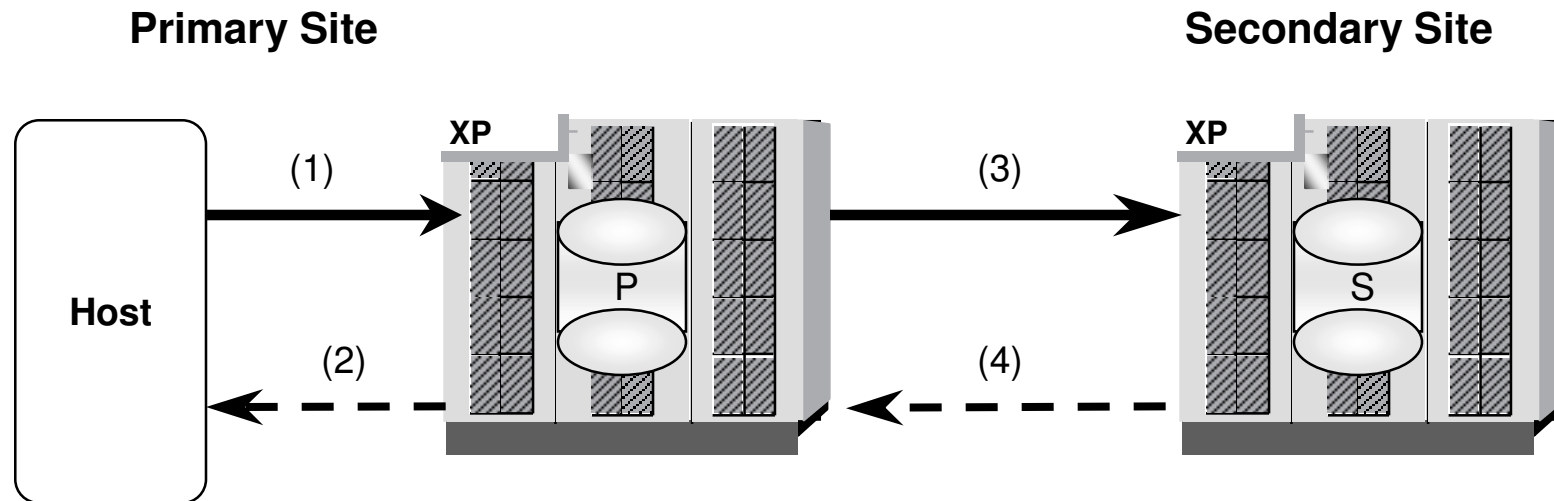

CISCO MDS 9509

Dual Path fabrics, Zoning

- Dual pathing for performance or mirroring W/ lock step IO?

- Zoning: Separating Node/Target Paths for Performance, Redundancy/HA or FT?

Continuous Access: HA or FT?
(async or sync mode)
Basic Operation

Primary Site   **Secondary Site**

Remote Copy Sequence:

(1) Application Initiates Write I/O to Primary Array

(2) Primary array confirms write to Application, writes to Primary Volume, and sends remote transmission, for sync both arrays most awk the IO.

(3) In background Primary Array is ready with "write copy" and sends it to Secondary array (async) While host is on to next IO.

(4) Secondary XP confirms remote copy receipt with Primary XP and writes to Secondary Volume

Questions and Answers

- Greg Tinker
  - Greg.tinker@hp.com
  - Hetro O.S., Enterprise SAN

- Leslie Thomson
  - Mission Critical Support
  - Leslie.thomson@hp.com
  - 770-993-0068

  Are you looking to assess your environment for fault tolerance?

  Kimberly.keeton@hp.com

  http://www.hp1.hp.com/ssp

HP WORLD 2004
Solutions and Technology Conference & Expo

Co-produced by:

interex
shared knowledge • shared power

encompass
AN HP USER GROUP

RECOMMENDED TRAINING VENUE FOR THE
HP Certified Professional