Next Generation Storage Area Network Design: 2Gbit/sec Performance Plus Advanced Functionality

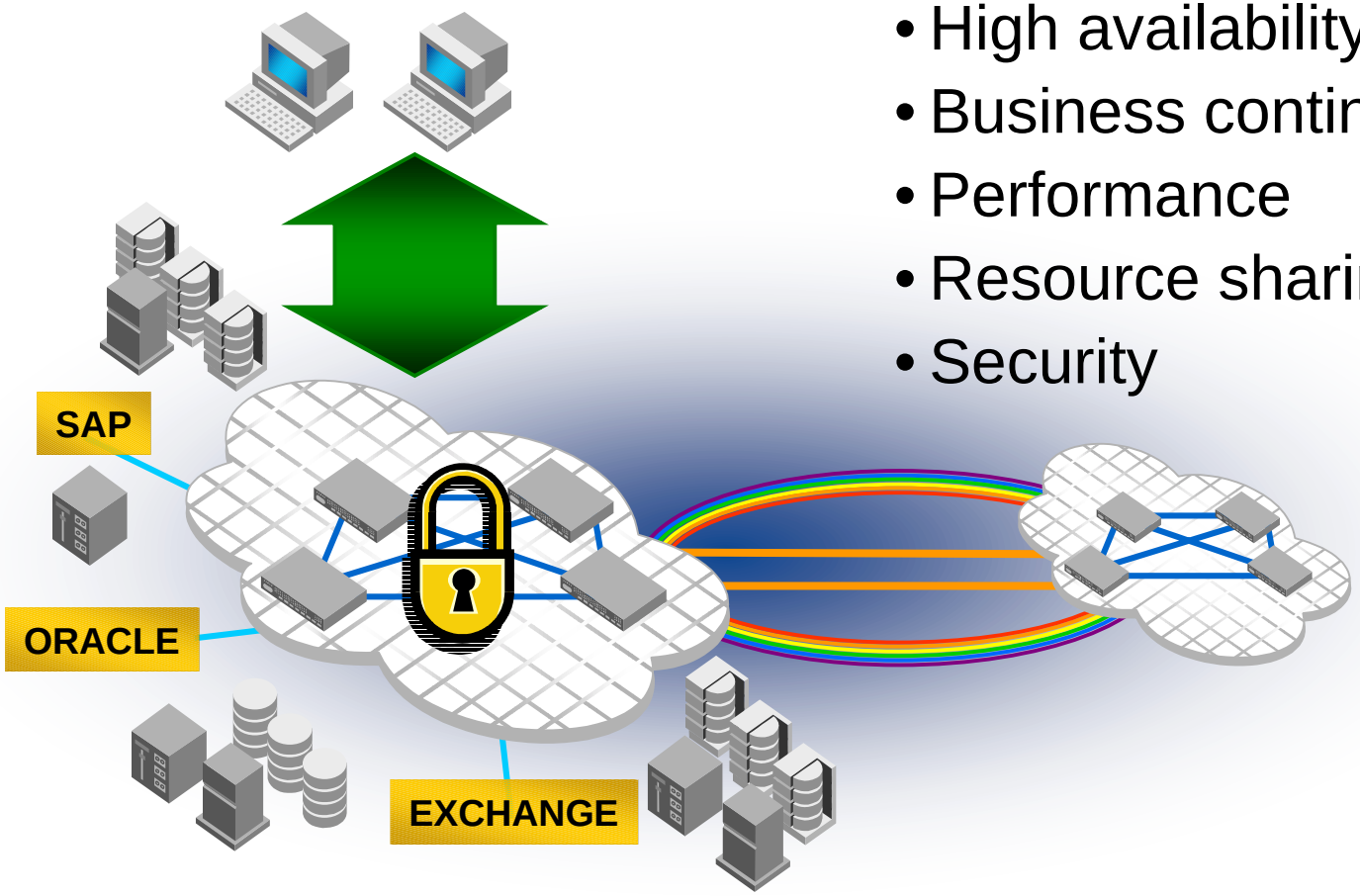
Daniel Cohen

Solutioneer

Brocade Communication Systems, Inc.

September 24, 2002

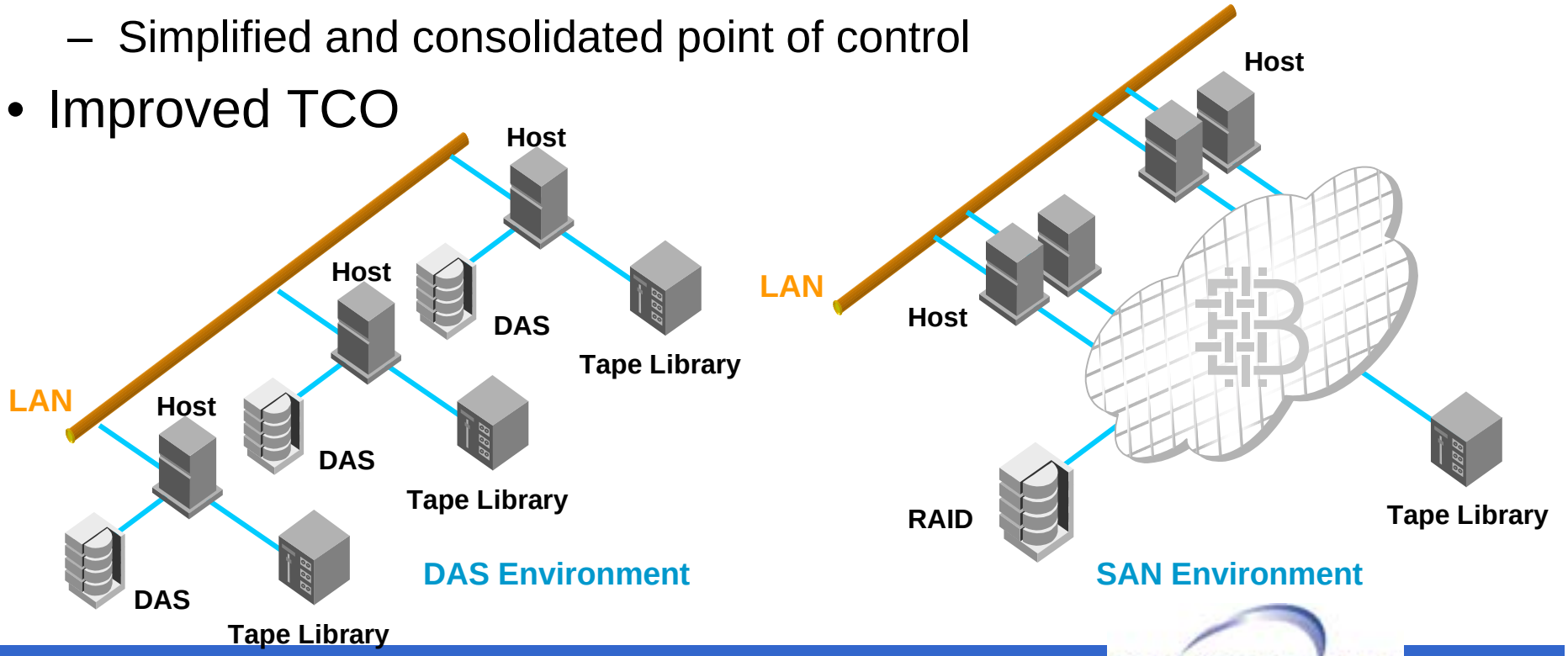
Requirements of the New Storage Environment



- Manageability
- High availability
- Business continuance
- Performance
- Resource sharing
- Security

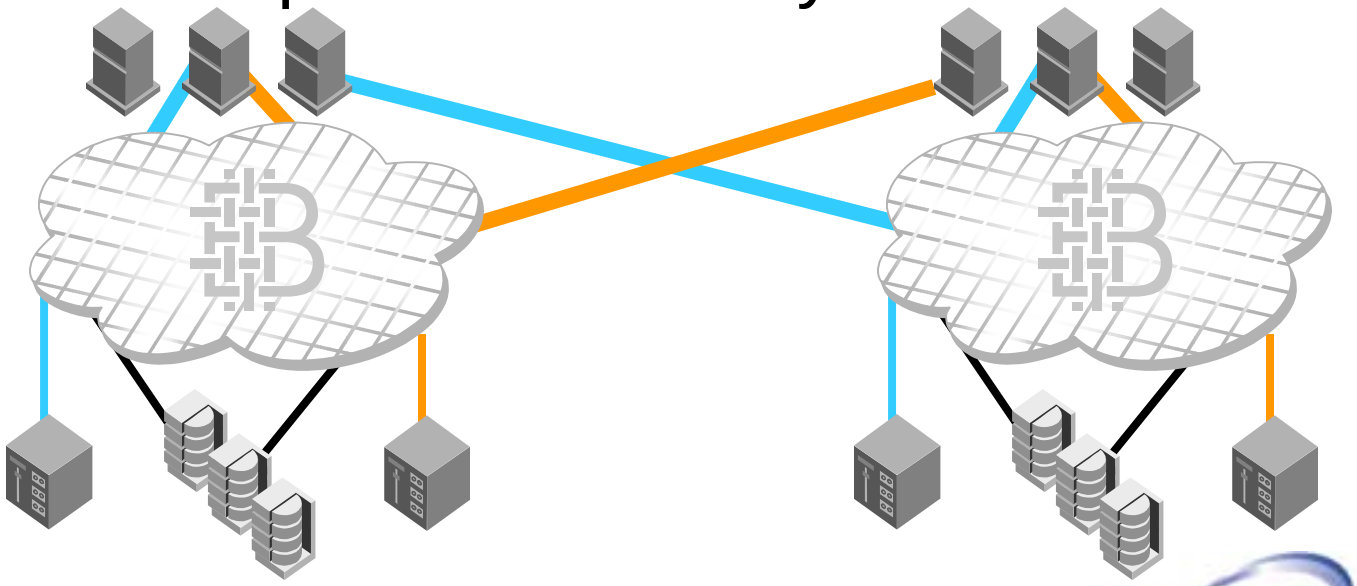
DAS to SAN Migration
New Application Deployment

- Sharing of resources
  - Centralized backup and consolidated storage
- Ease of administration
  - Simplified and consolidated point of control
- Improved TCO



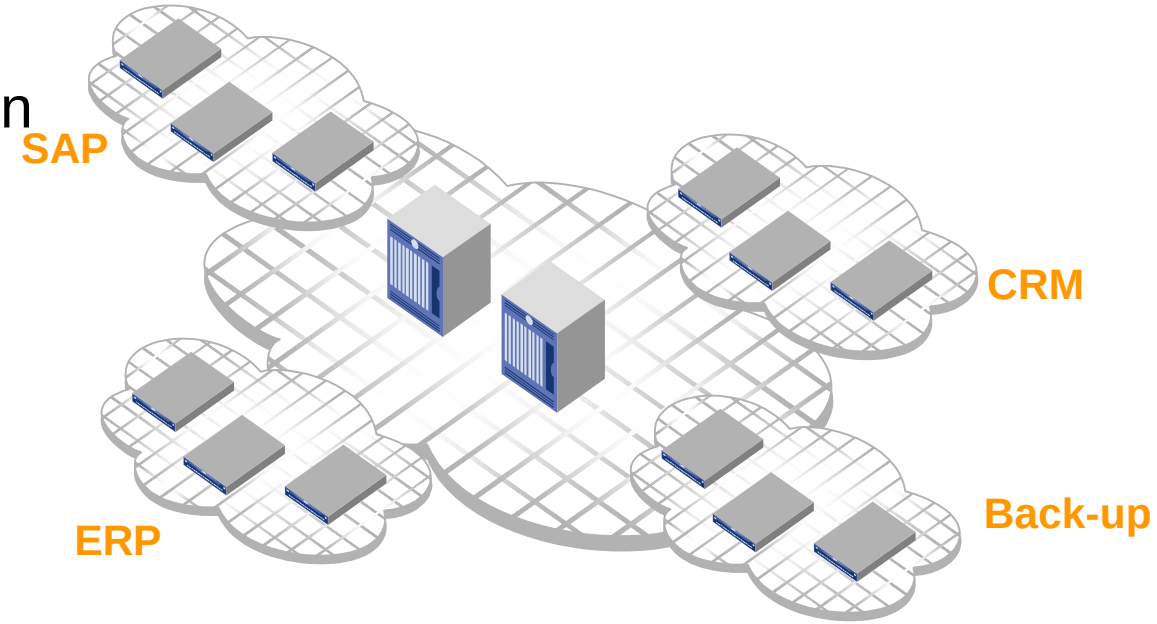
Emphasis on High Availability

- Requirement for Always On Infrastructure
  - Mission critical applications demand 24 x forever availability
  - Ideal for databases and mail applications
- Enables transparent scalability

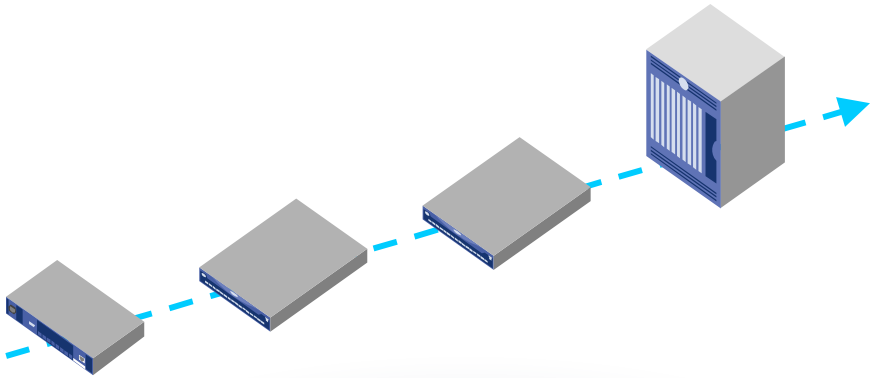


Why Larger SANs?

- Combine many small SAN islands into fewer larger SANs
- Centralized management simplifies and maximizes resource utilization
- Share server and storage resources across more applications and users
- Full Investment Protection

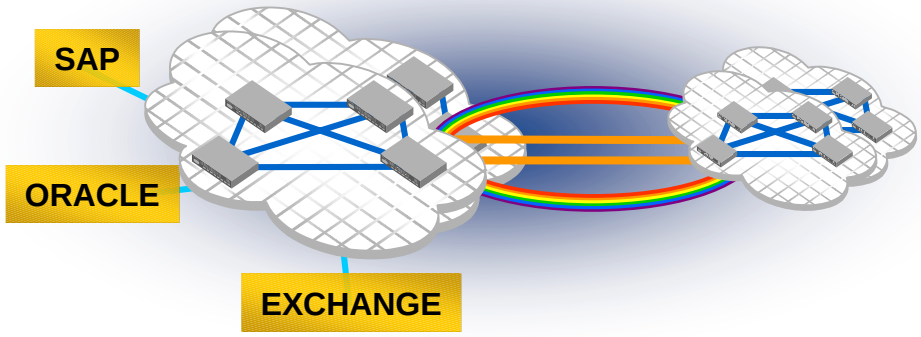


Requirements for Next Generation SAN



Highly Reliable Switches

- Redundant components
- Hot-swappable components
- Proven field reliability

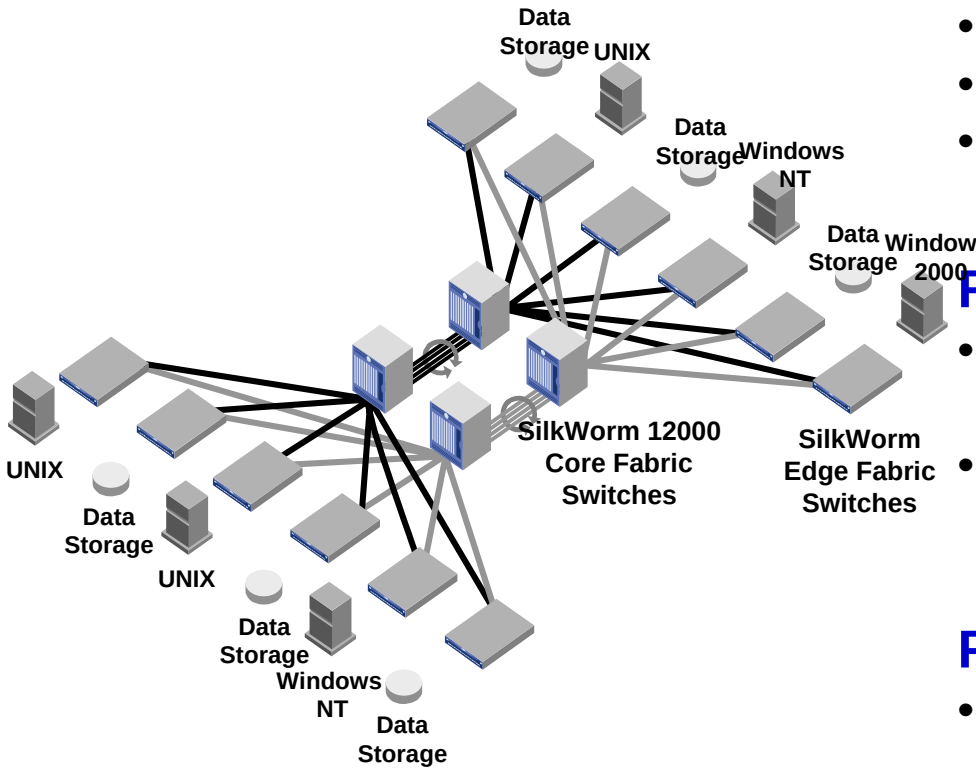


Highly Available Fabrics

- Cost effective dual fabric designs
- FSPF dynamic path re-routing
- *Extended Fabrics* enable remote data replication via optical or IP links

Ensure Application Uptime and Business Continuance
Cost Effective High Availability Solutions

Requirements for Next Generation SAN



High Performance Connectivity

- 2Gbit/sec Fibre Channel on all ports
- Non-blocking, low latency switching
- 8 Gbit/sec *Trunking* between switches
- Optimal performance under heavy load

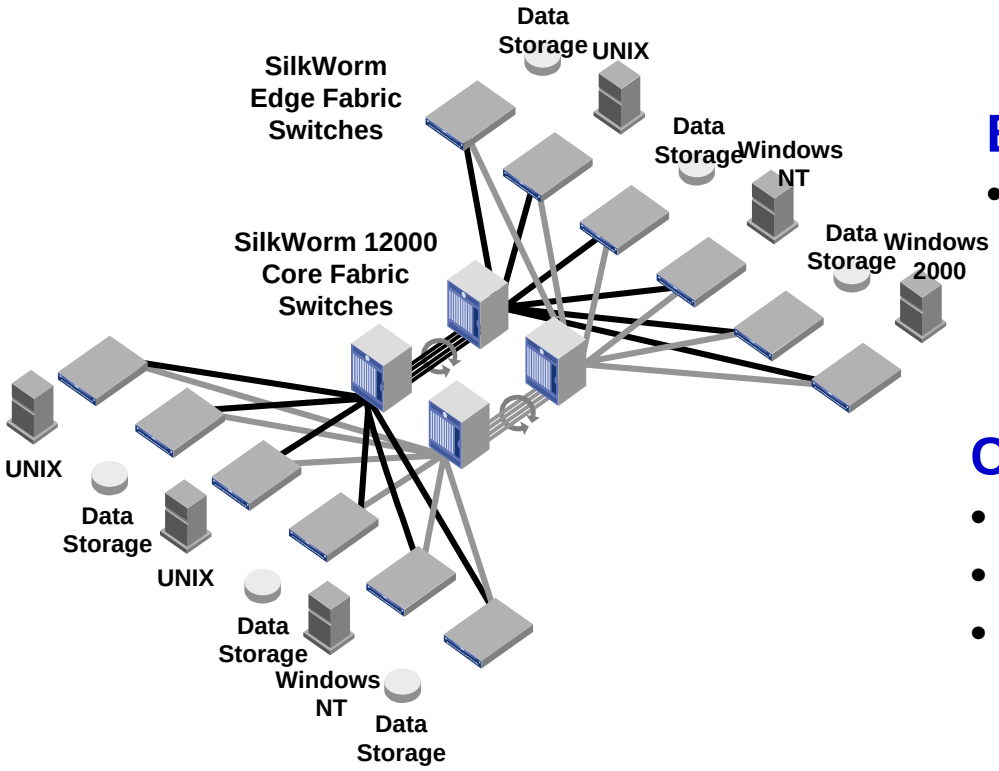
Real-Time Performance Monitoring

- Analyze performance by port, link or data path
- Foundation for capacity planning or network bill back applications

Fabric Health Monitoring

- Real-time threshold monitoring and notification of performance levels or error conditions
- Reduce the risk of unplanned downtime

Fabrics Deliver Scalability

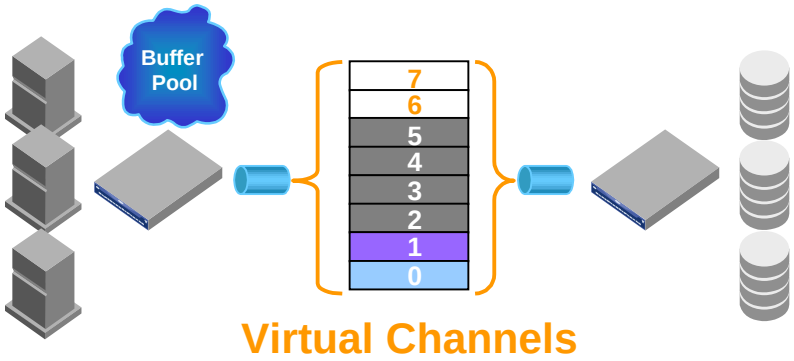


Broad Product Family

- 8-port to 128-port switching products

Core-Edge Fabric Design

- Networking model for storage
- Cost-effective, any-to-any connection
- Enables non-disruptive network expansion



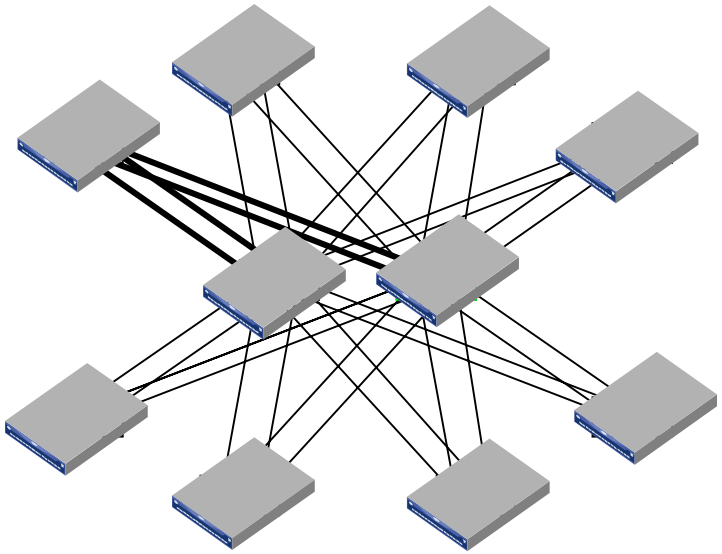
Most scalable architecture

- ASIC-based Virtual Channels ensure balanced performance under load
- Fabric OS optimized for large fabrics

Topology Examples

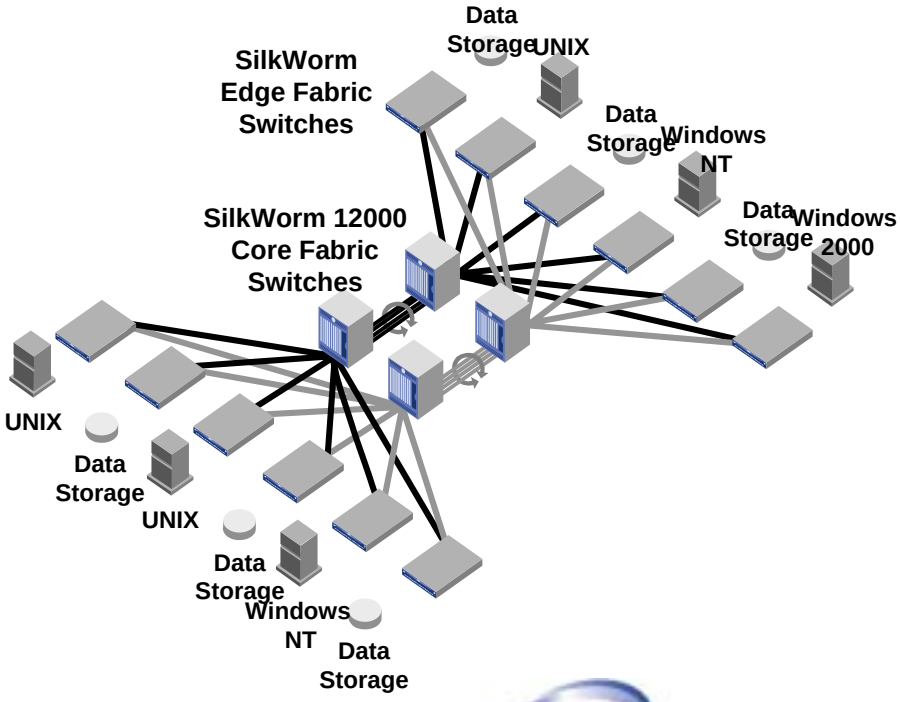
Mesh Advantages:

- Infinitely expandable
- High connectivity
- Less expensive



Core/Edge Advantages:

- High performance
- Highly expandable
- Less locality required



Intelligent Fabric Services Architecture

The Platform

HP SureStore 1Gb/2Gb Switch 8b
SilkWorm 3200

- Cost effective 8-port 2Gbit/sec fabric switch
- Complements Entry-to-Enterprise 2 Gbit/sec switch portfolio
- Supports Advanced Fabric Services
- 3rd generation Advanced Fabric Service

HP SureStore FC 1Gb/2Gb Switch 16b
SilkWorm 3800

- 16 auto-sensing 1 and 2 Gbit/sec ports
- 1U standard rack mount
- Redundant hot-swap power and cooling
- Supports Advanced Fabric Services

BEST OF COMDEX 2001
WINNER
Best Networking Hardware

DATAMATION Product of the Year 2001
IT Manager's Choice

HP StorageWorks Core Switch 2/64
SilkWorm 12000 Core Fabric Switch

- **Flexible, Modular Architecture**
  - Scalable up to 128 ports
  - 2 Gbit/sec ports; autosensing
  - 3rd Generation Brocade ASIC
  - High port density – 14U enclosure
- **99.999% Availability**
  - Redundant, hot-swap elements
  - Non-disruptive software updates
- **Advanced Fabric Services**
  - Existing Fabric OS services
  - New 3rd Generation services
- **Multi-protocol Architecture**
  - FC, IP, FICON…

NetworkWorld BLUE RIBBON

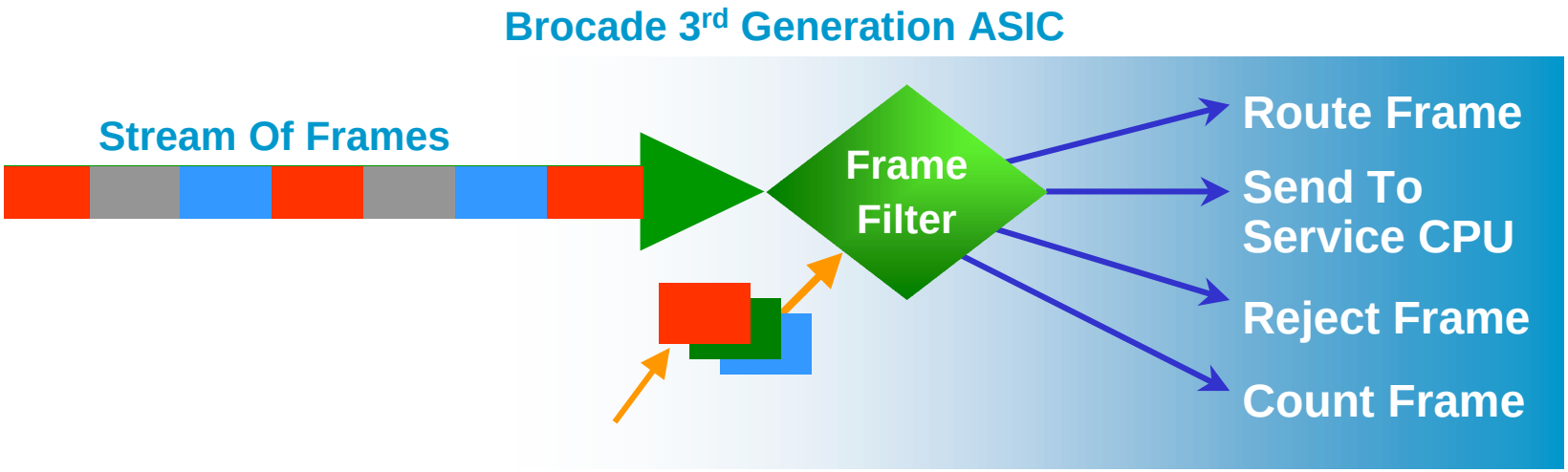
Intelligent Fabric Services Architecture

The Fabric OS

Advanced Fabric Services

- **Frame Filtering**
  - **Enabling technology based on Brocade ASIC**
- **ISL Trunking**
  - simplifies design and administration
  - enhances performance
  - higher availability
- **End-to-end Performance Analysis**
  - ensures application performance
  - enables proactive network optimization
- **Advanced Zoning**
  - Heterogeneous environments

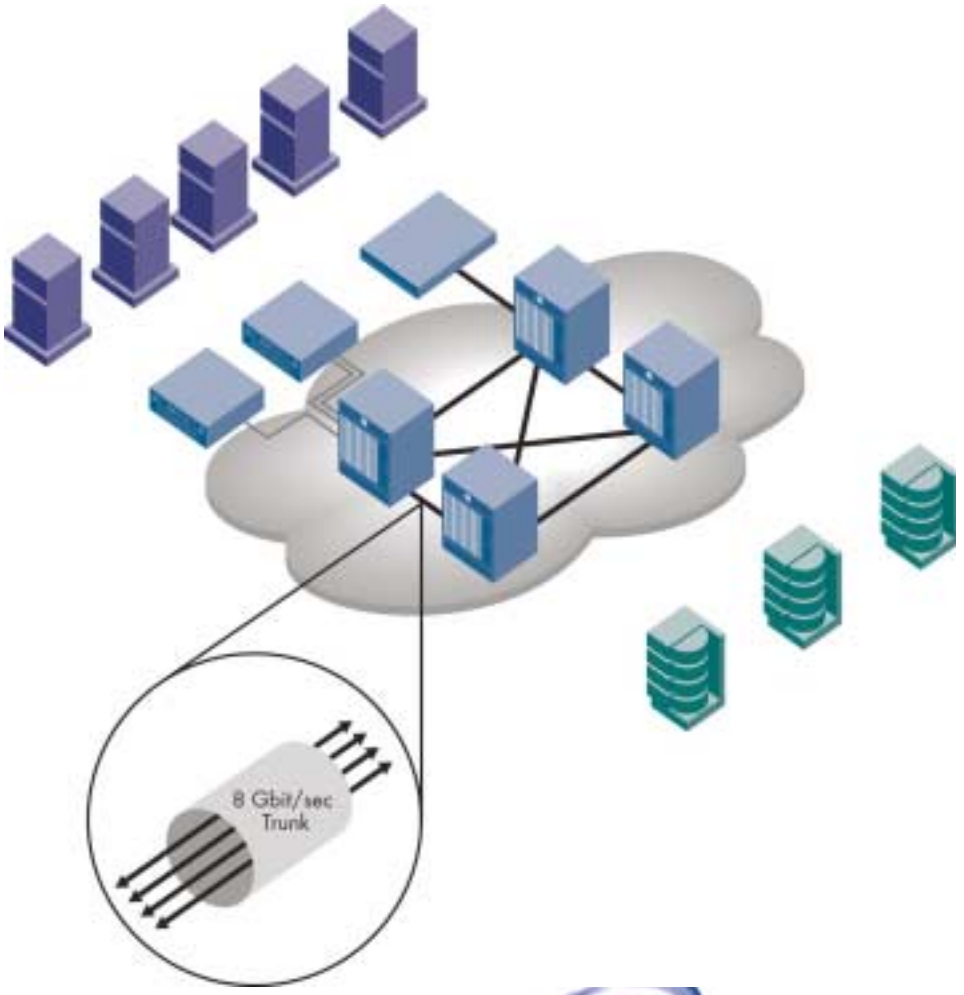
Frame Filtering



- Intelligent data routing through Frame Filtering

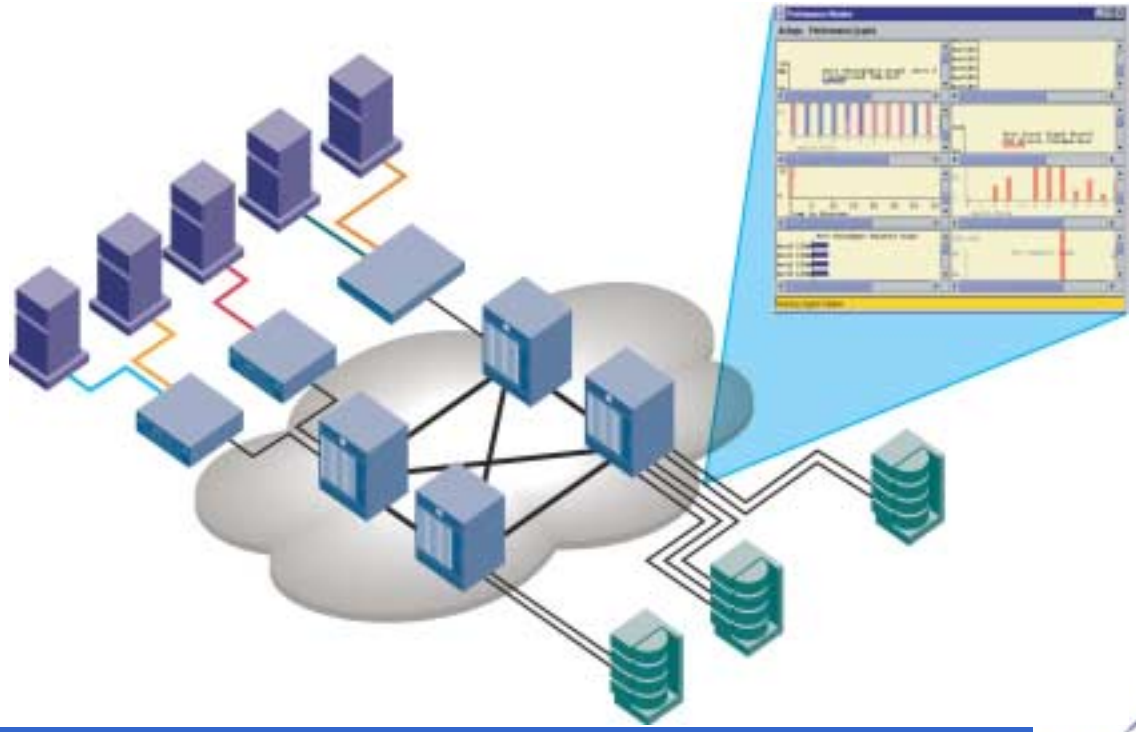
Trunking

- Available on 3000 series and 12000 only
- Enabled by default for SilkWorm 3200, 3800, and 12000 switches
- Requires a license
- A Trunk consists of four contiguous ports from the same ASIC

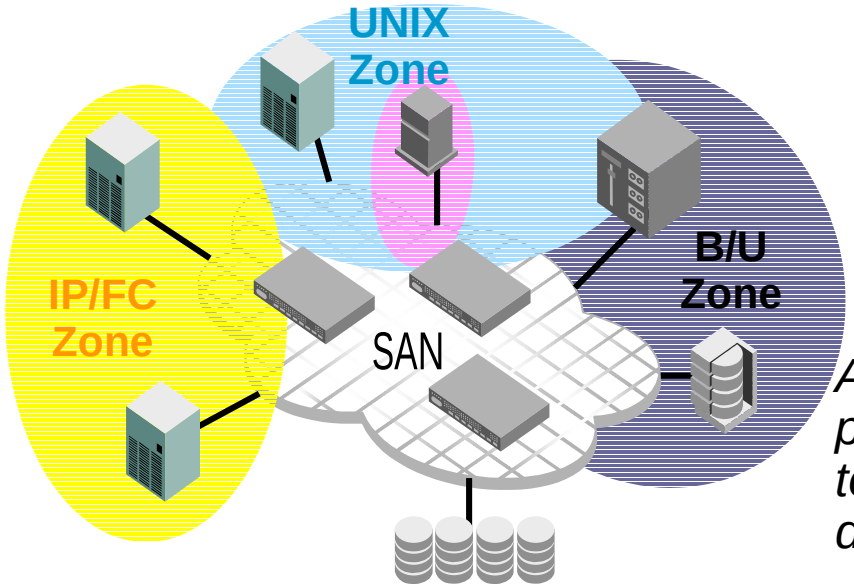


Advanced Performance Monitoring

- Simplify SAN capacity planning and network expansion
- Measure service-level-agreement compliance
- Enable bill-back applications



Advanced Zoning



Administrators can organize the physical fabric into logical groups to prevent unauthorized access by devices outside the zone.

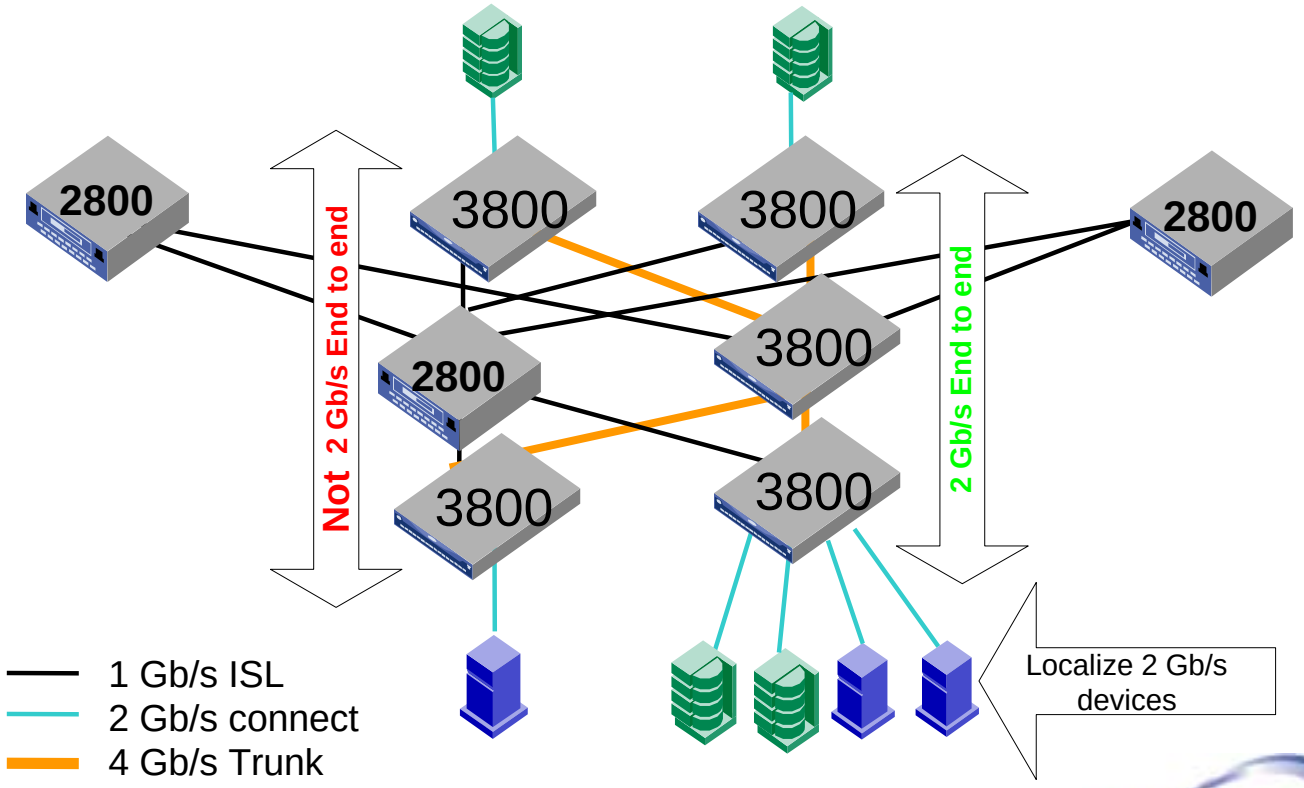
- **Built upon Frame Filtering™ technology**
- **Hardware enforced access control for heterogeneous environments**
- **Now supports WWN-based zoning in hardware – simple and safe**
- **Simplifies zoning administration**

Intelligent Fabric Services Architecture

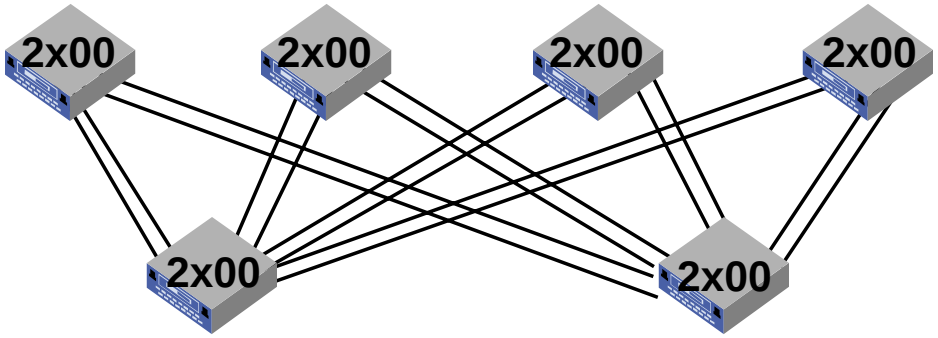
2 Gbit/sec Features

Use 2 Gbit/sec Switches as The Core

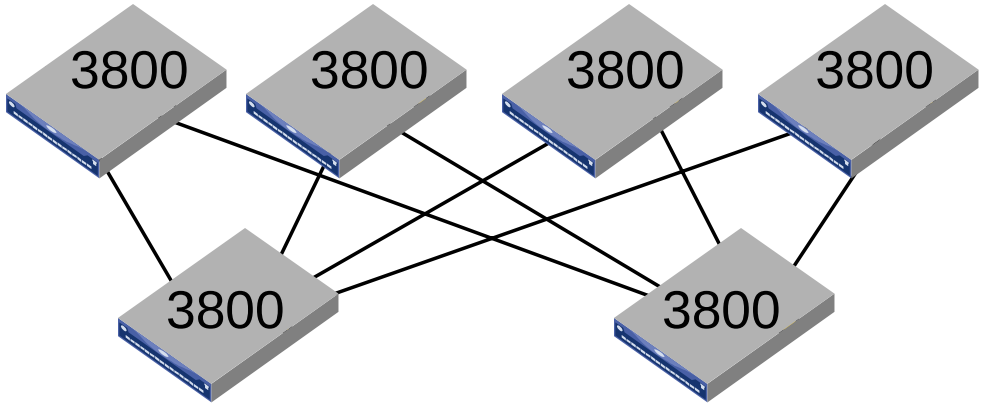
- Enables other 2 Gbit/sec switches to trunk
- Ensures end-to-end 2 Gbit/sec path if using Core/Edge
- Can use for 2 Gbit/sec device attach too



2 Gbit/sec Frees Up Ports



64 ports
3:1 over subscription ratio (worst case scenario)



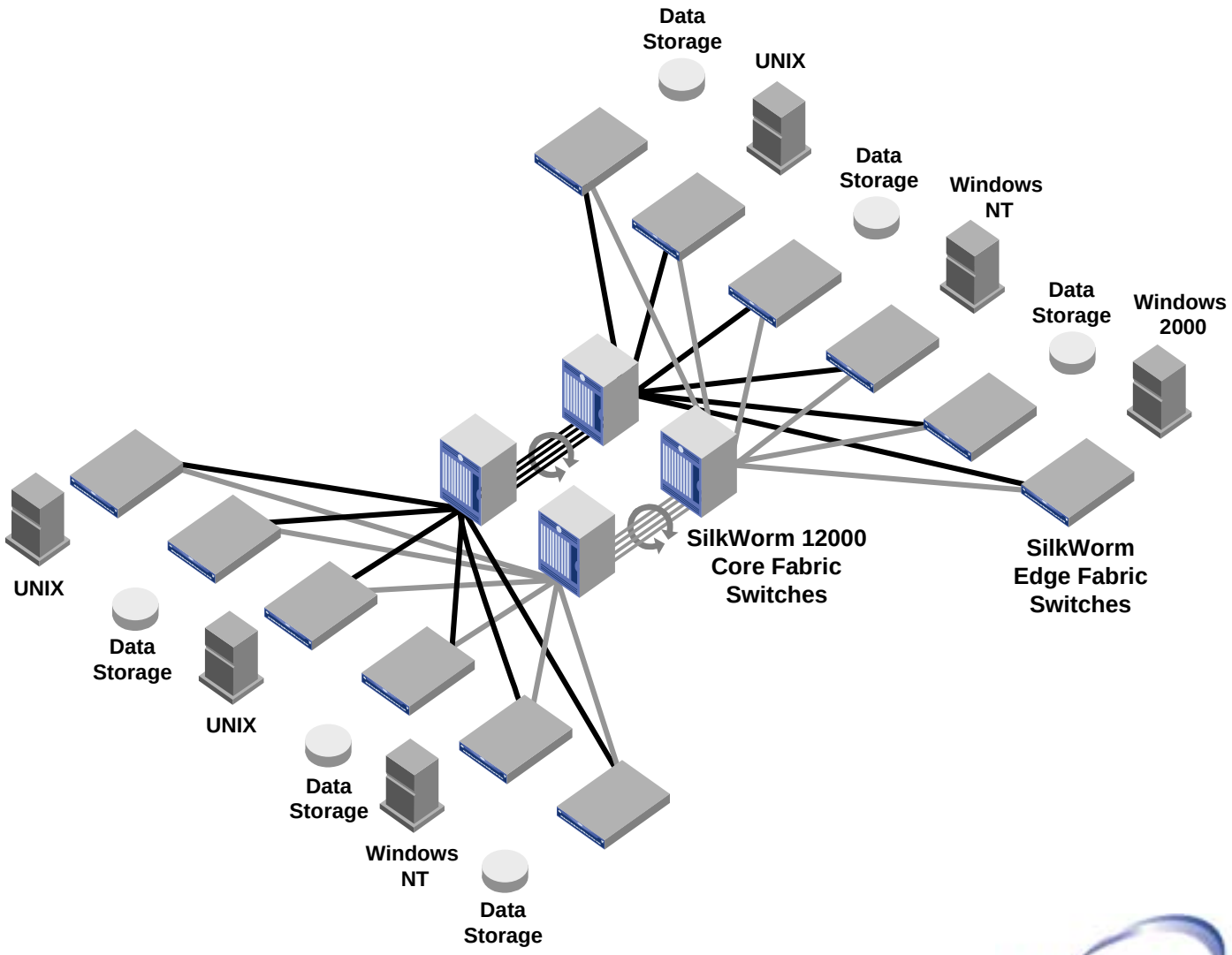
80 ports
Assumes 1 Gbit/sec Edge devices
3.5:1 over subscription ratio (worst case scenario)

Speed Negotiation

- **Almost all 2 Gbit/sec devices today work with Brocade auto-sense**

- **Some older devices have difficulty with speed negotiation**
  - Use portCfgSpeed <slot>/<port #>, 1 on switch
  - Lock down speed on that device

Trunking Optimizes Performance



SAN Design Best Practices

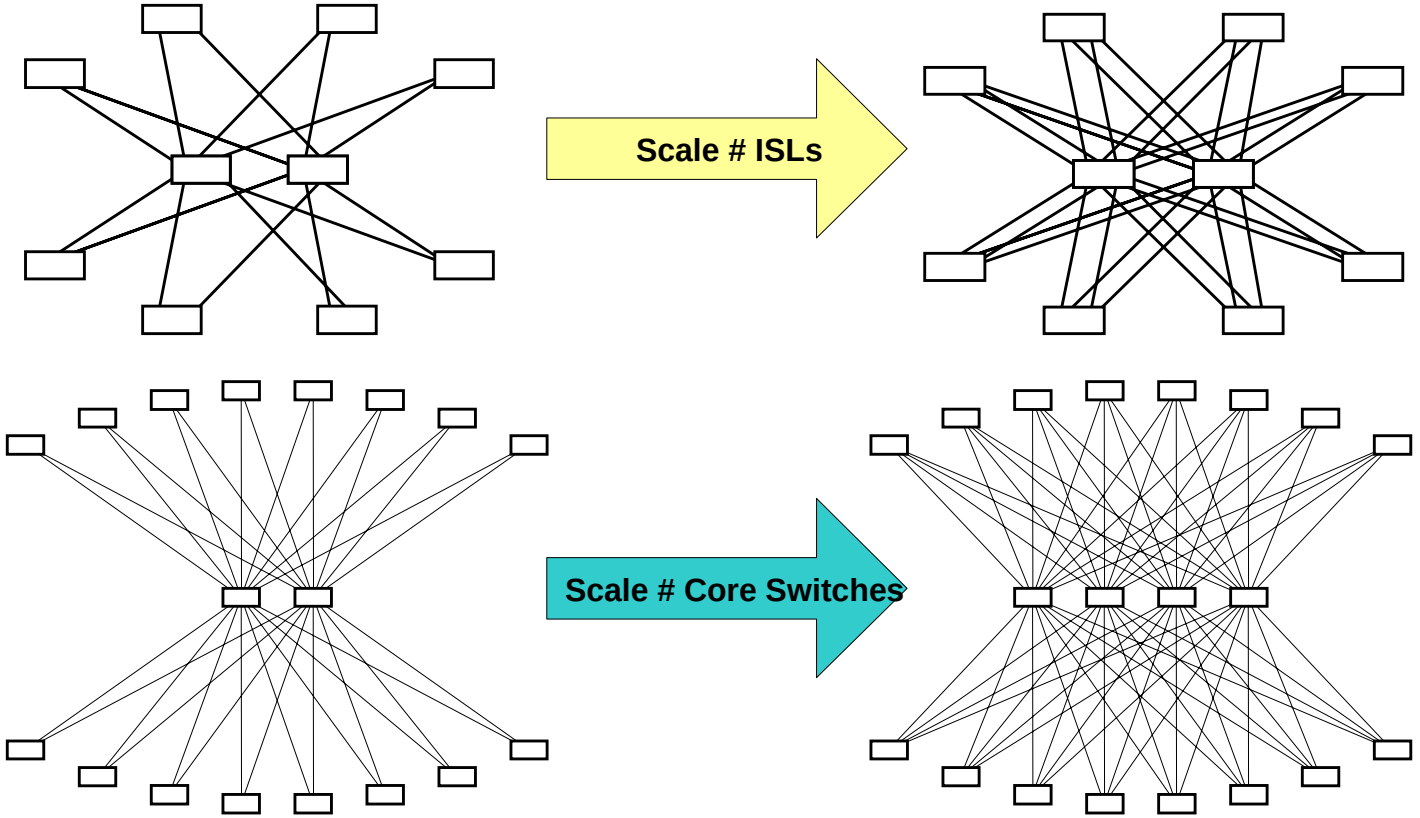
SAN Design Is Not Rocket Science

- **The freedom to build any fabric topology you desire with the following caveat:**
  - There are support limits imposed by the switch support partner
- **Core / Edge is becoming a standard architecture and straightforward to design**
  - Fits wide range of application, performance, scalability, and availability requirements
  - Extensively tested
  - Easy to specify and to understand
- **New products can and do impact SAN design**
  - Trunking
  - 2 Gbit/sec
  - SilkWorm 3200, 3800, 12000

Helpful Checklists

- A Site Survey is very helpful in identifying the information necessary to put together a Core Switch 2/64 based SAN design
- The install checklist is an important document to complete during the install
- Do a little work now or do lots of work later

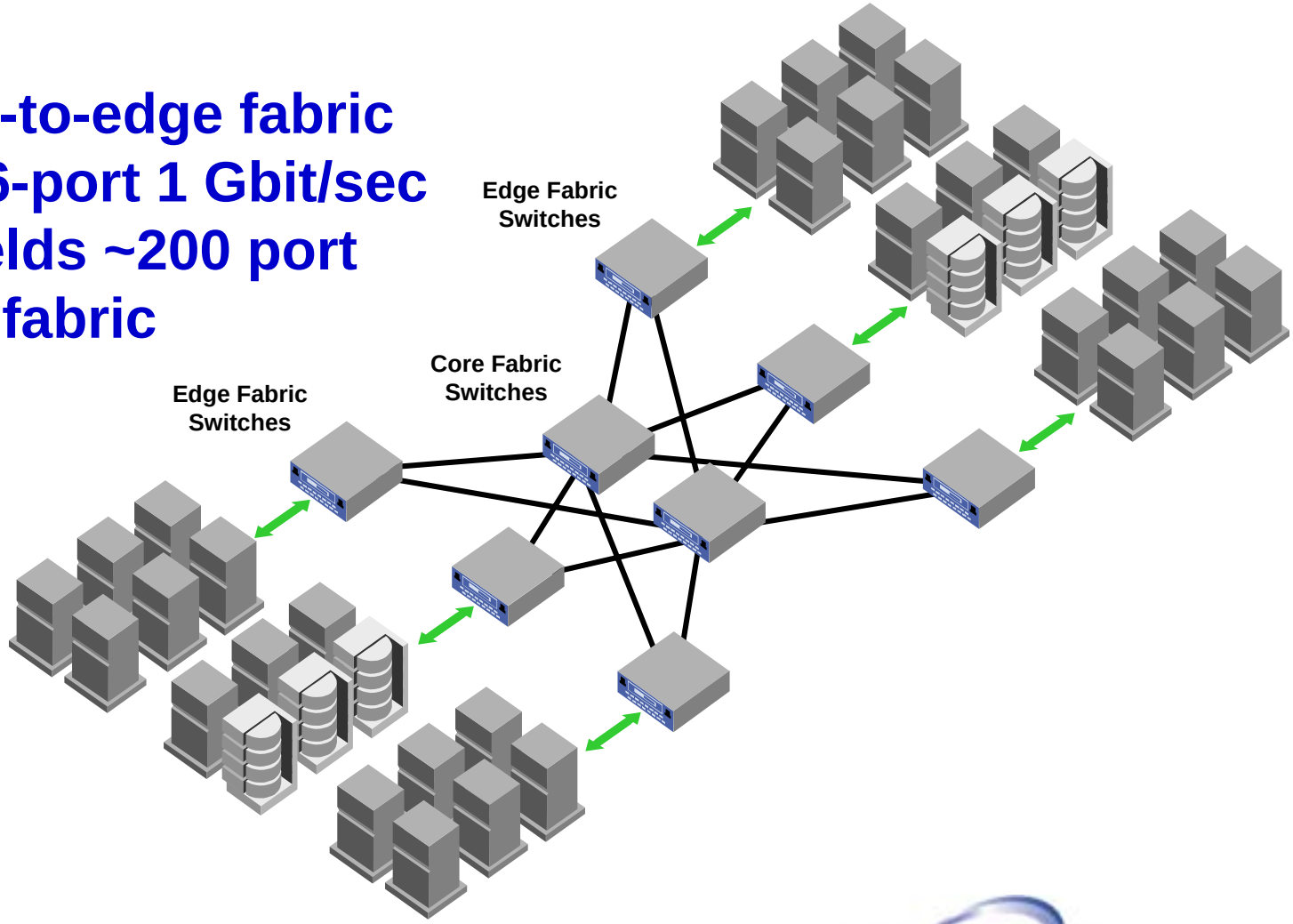
Scalability



- Scale fabric by adding switches
- Scale fabric by replacing existing core with a larger core
- Scale performance by adding ISLs or more core switches

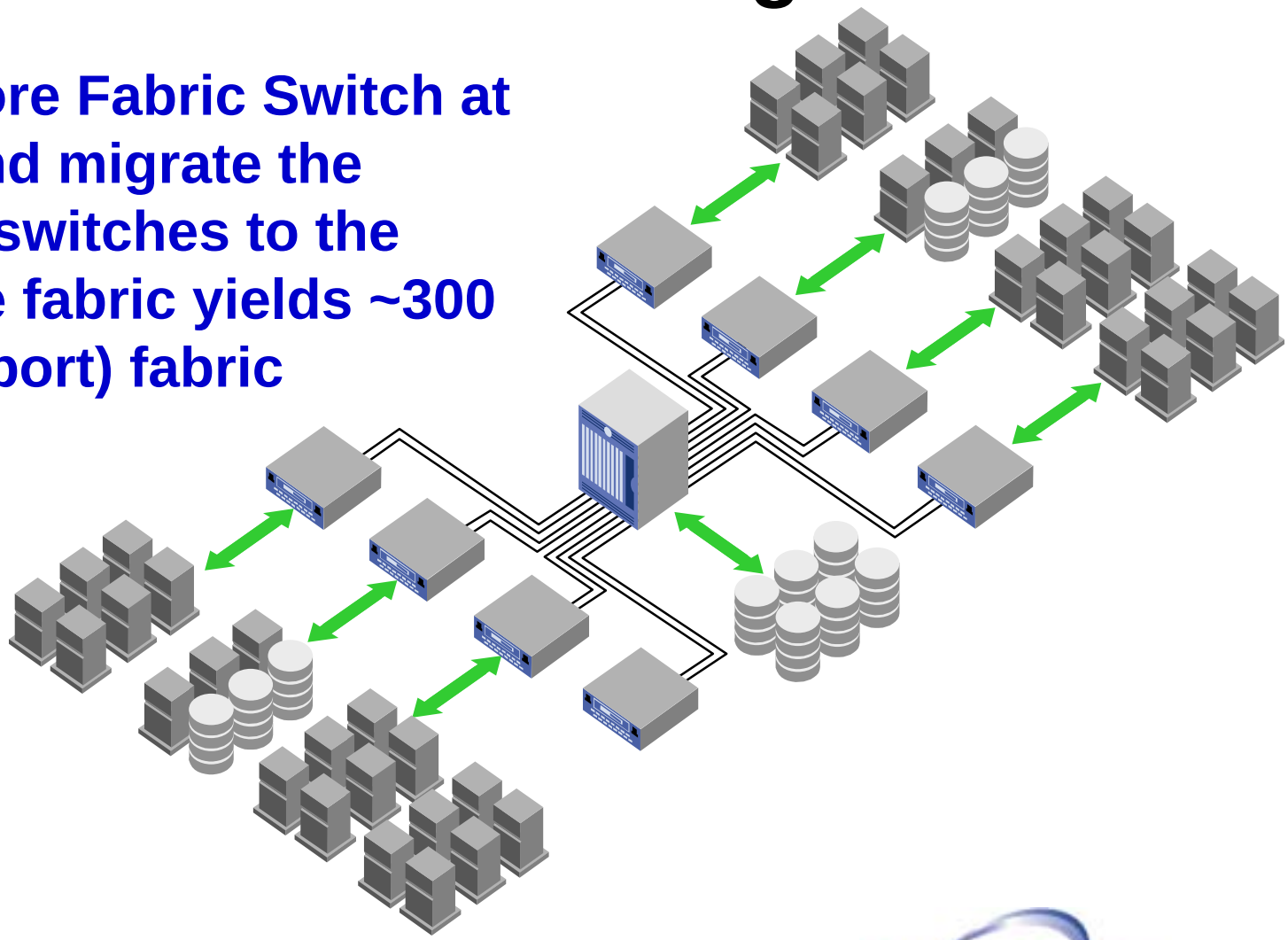
Core-to-Edge Topology:
Scalable Solutions for Large Fabrics

A large core-to-edge fabric built from 16-port 1 Gbit/sec switches yields ~200 port (user ports) fabric



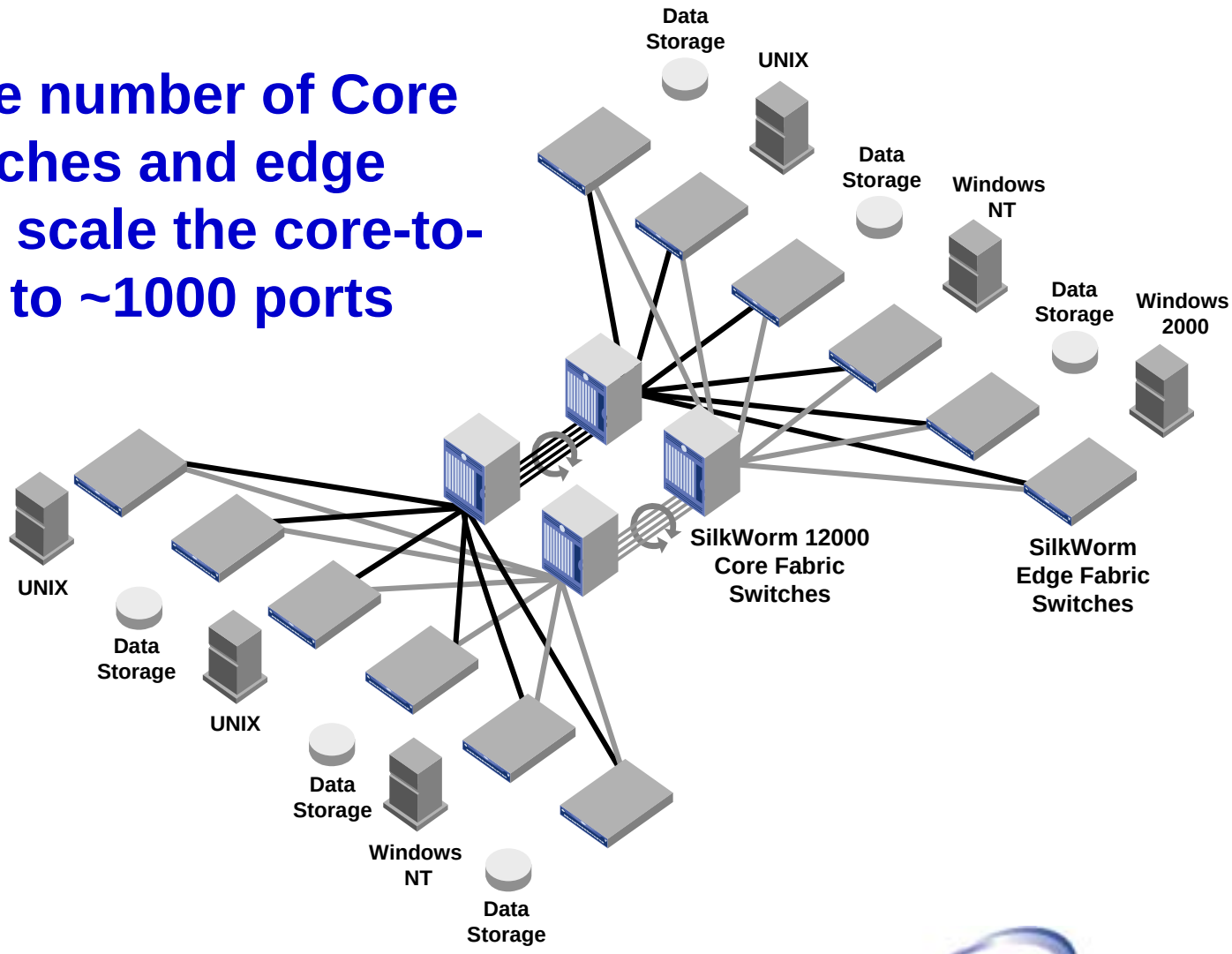
Core-to-Edge Topology: Scalable Solutions for Large Fabrics

Install a Core Fabric Switch at the core and migrate the 1 Gbit/sec switches to the edge of the fabric yields ~300 port (user port) fabric

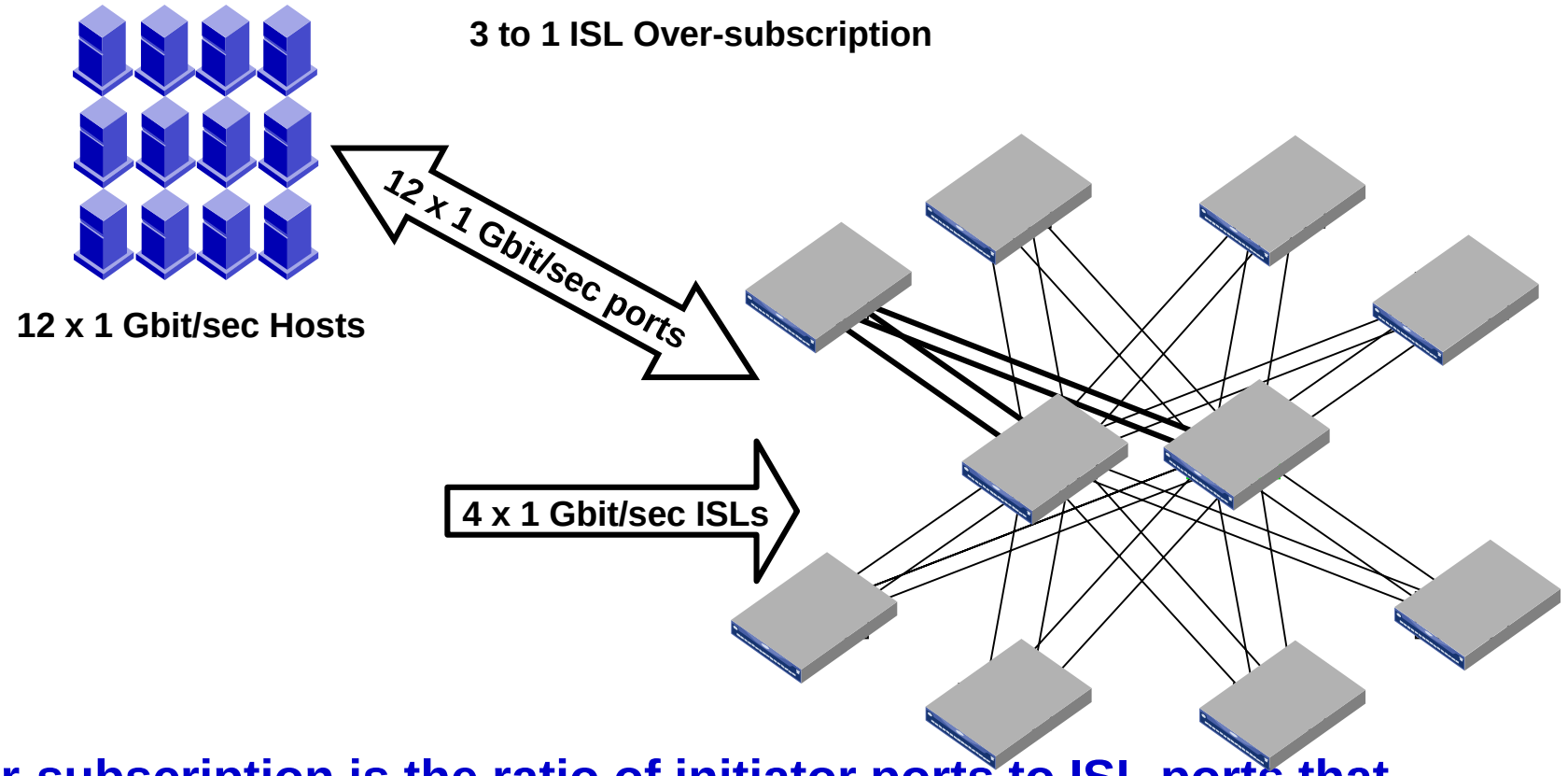


Extending the Networking Model

Increase the number of Core Fabric Switches and edge switches to scale the core-to-edge fabric to ~1000 ports

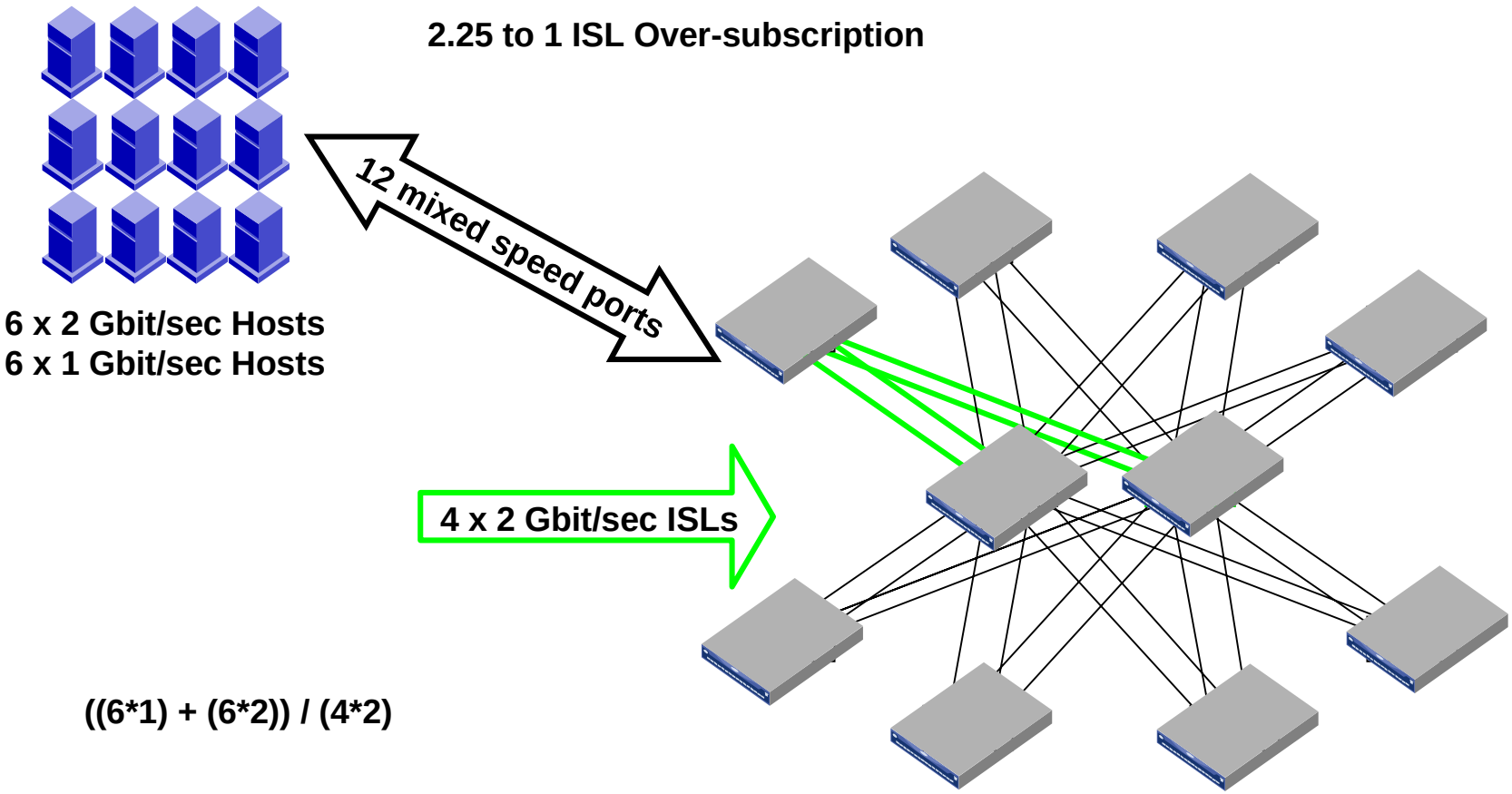


ISL Over Subscription



- Over-subscription is the ratio of initiator ports to ISL ports that traffic might traverse between contiguous switches in a fabric
- Congestion can be the product of over-subscription

ISL Over Subscription



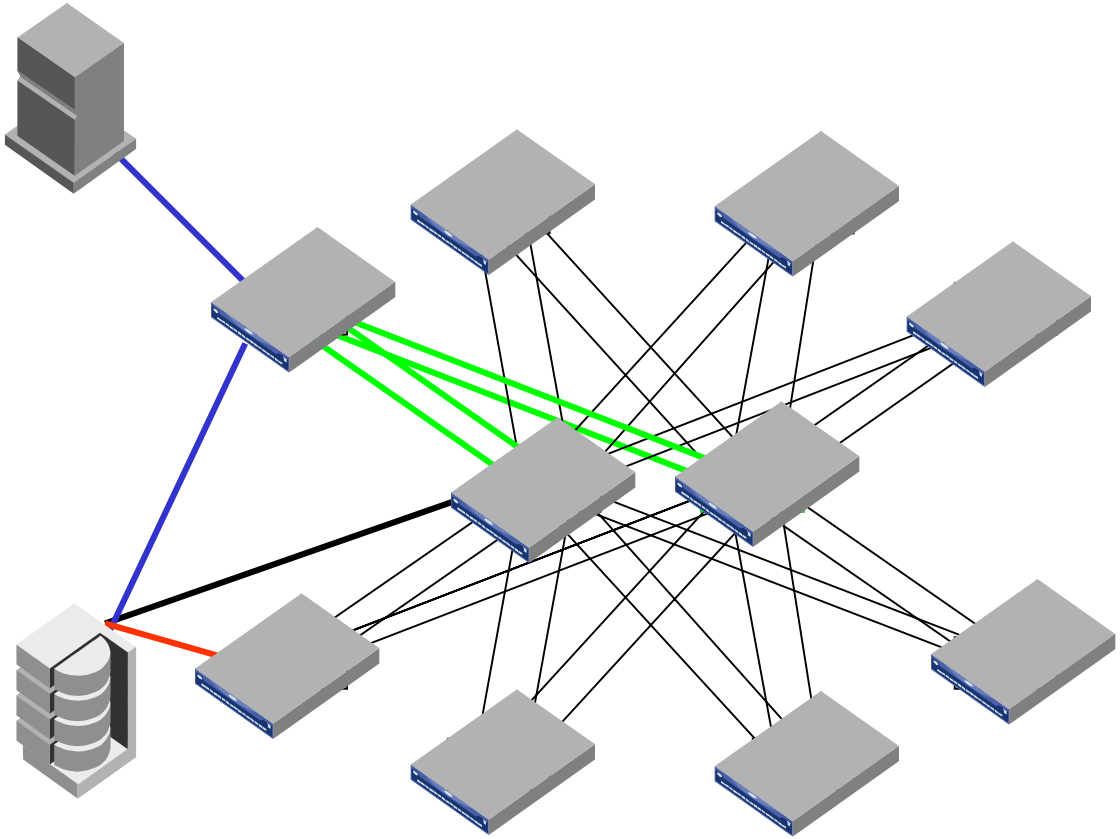
- **If devices are 1 Gbit/sec and ISLs are 2 Gbit/sec, the formula changes**
  - The sum of the speed of the input ports divided by the speed of the output ports

Device Attachment Points Do Matter

Scenario A
0 Hops

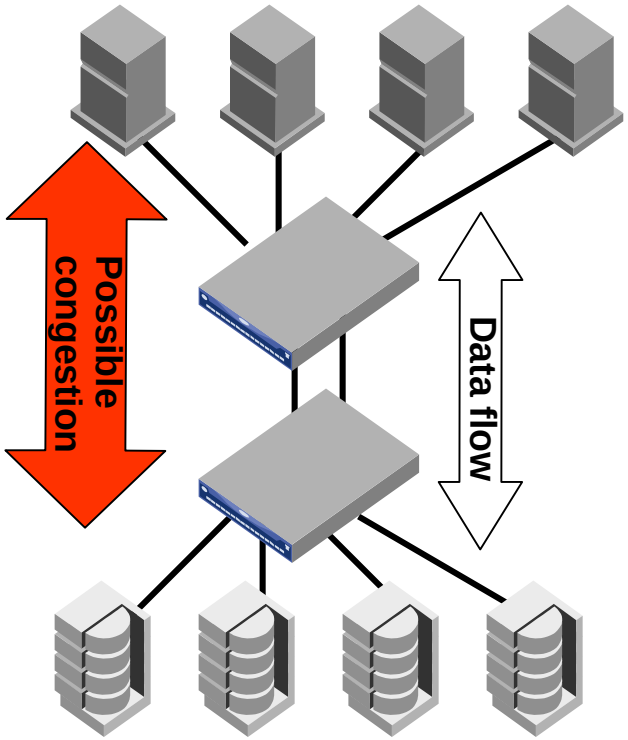
Scenario B
1 Hop

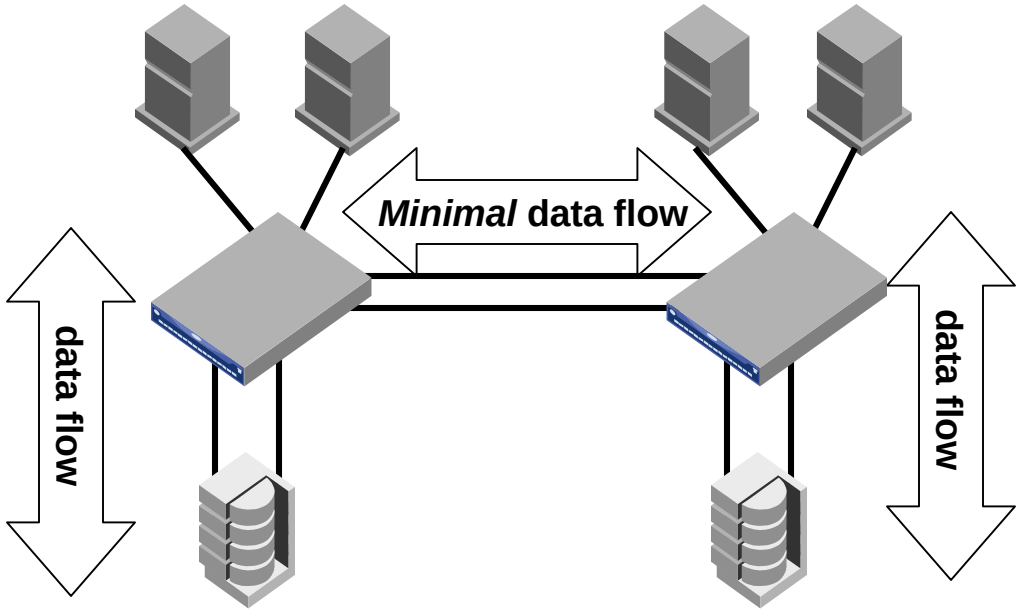
Scenario C
2 Hops



- **While device placement does not *constitute* fabric topology, it may *affect* and be affected by topology**
- **For example: Attaching a device to a core switch reduces the quantity of Core ports available for expansion**

Locality

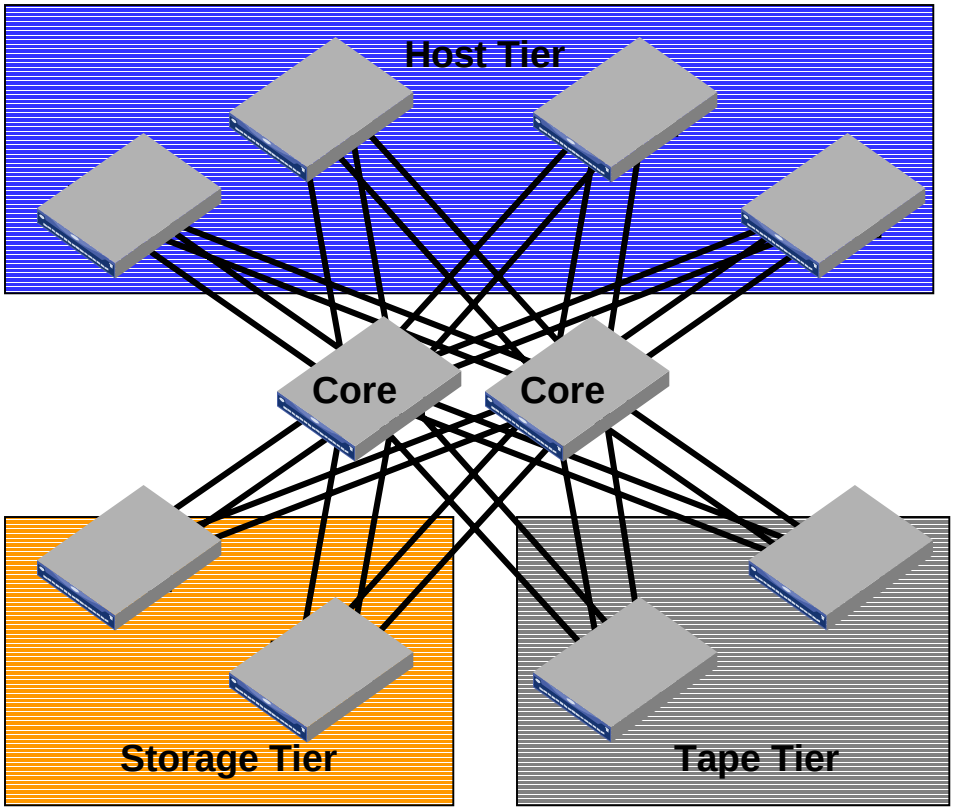




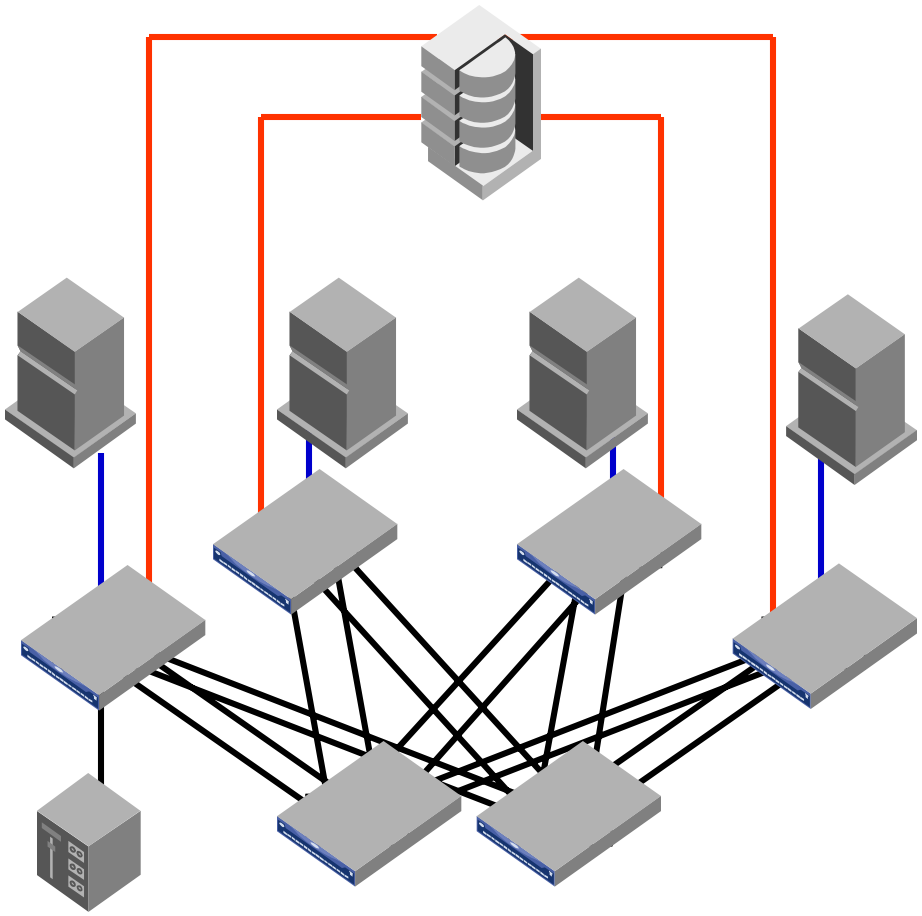
- Zero locality can be acceptable (ex. Tiered fabric)
- Good locality does improve performance
- Localize to a switch or group of switches

Tiering

- Tiering is an effective solution to SAN management
- Switches are grouped by connection (i.e. disk, tape and host)
- Increasing the number of tape, disk or host resources is done by attaching the new devices to an open port on the appropriate tier and enabling access (i.e. zoning)
- If a particular tier requires expansion, add a new switch.



Locality & Device Placement



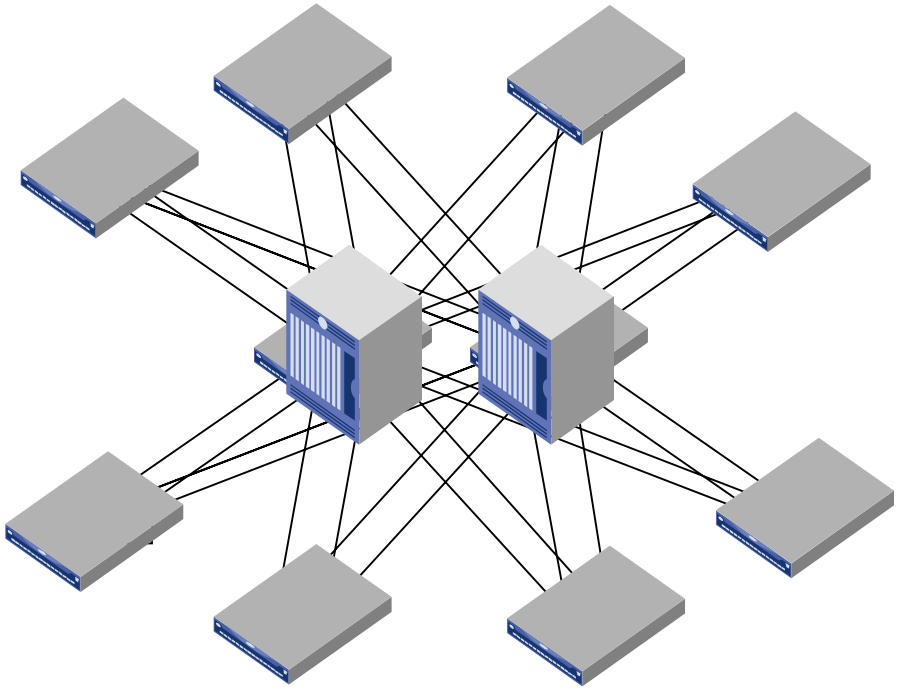
- **Locality and tiering are opposites**
- **Sprinkle storage across the edge for high locality and performance**
- **The tape is not localized, but is accessible to all devices (any-to-any connectivity)**
- **Consider a formula:**
  - 4 ports ISL
  - 2 port storage
  - 10 ports hosts

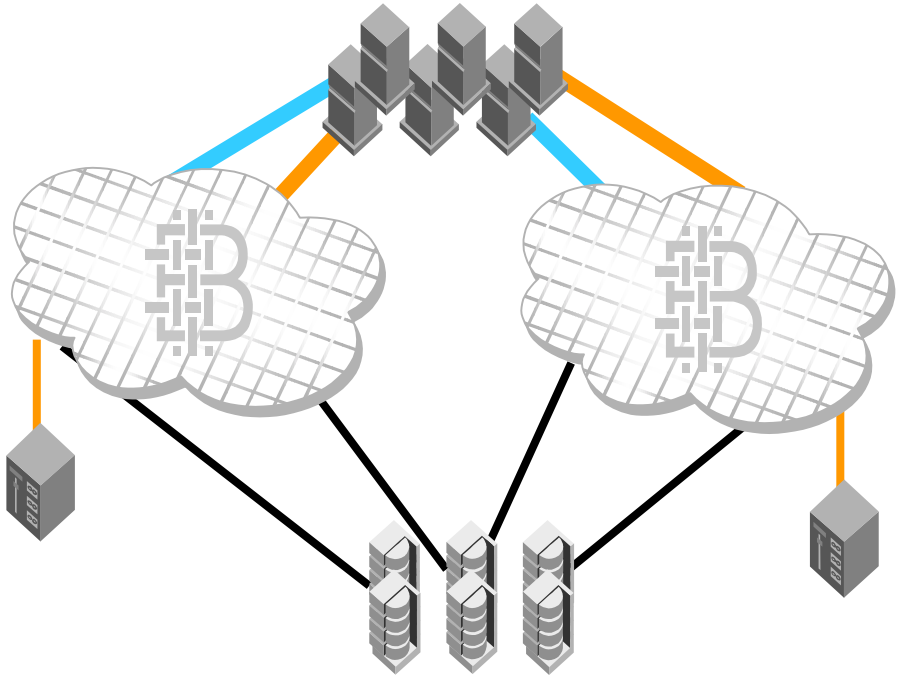
SAN Design Using The Core Switch 2/64

Core Switch 2/64 SAN Design

- **Convergence on 7:1 ISL over subscription ratio or less when thinner ISLs are sufficient**
  - No more than 4 ports per 16-port edge switch dedicated to ISLs
  - 16 ports per 64-port Edge switch dedicated to ISLs
- **Device attachment strategies increasingly important since we now have 64-ports to connect**
  - Enable scaling of performance and fabric size
  - Needed for optimal availability
  - Simplifies things too
- **Other topologies (I.e. full mesh) that used to be boring are now interesting**
  - Core-to-edge is still the de facto standard

SilkWorm 12000 Maintain Current Course & Direction



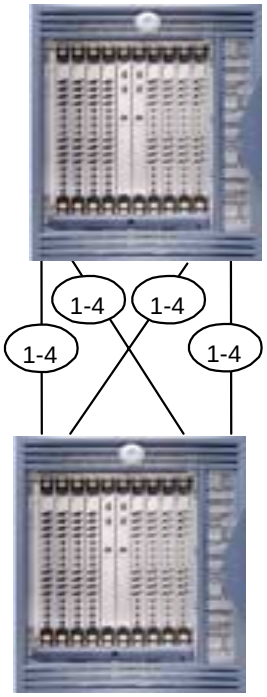


- Some solutions will use a single core when deploying a dual fabric
- A high port count core switch will more readily be used for devices

SilkWorm 12000 ONE Fabric Per Chassis

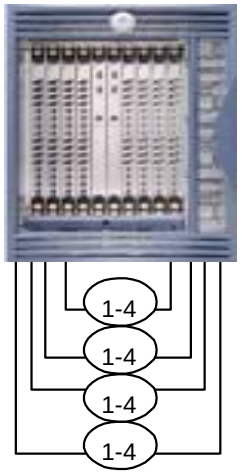
- **Don't forget – one fabric per chassis recommended for high availability**
  - Mitigates environmental catastrophes that could take out an entire chassis
  - Prevents same Fabric OS from populating two fabrics
  - Two fabrics per chassis should be considered on an exception basis only
  - Limits operator error to a single fabric

Summary Of SilkWorm 12000 Topologies

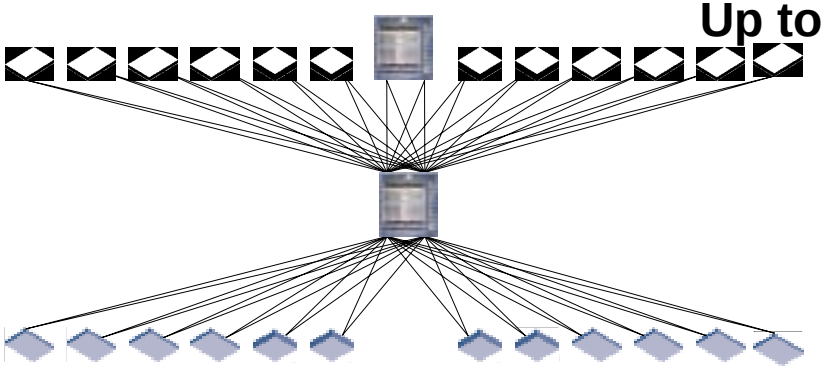


4 x 12000
Up to 248 available user ports

1 x 12000
64 available user ports

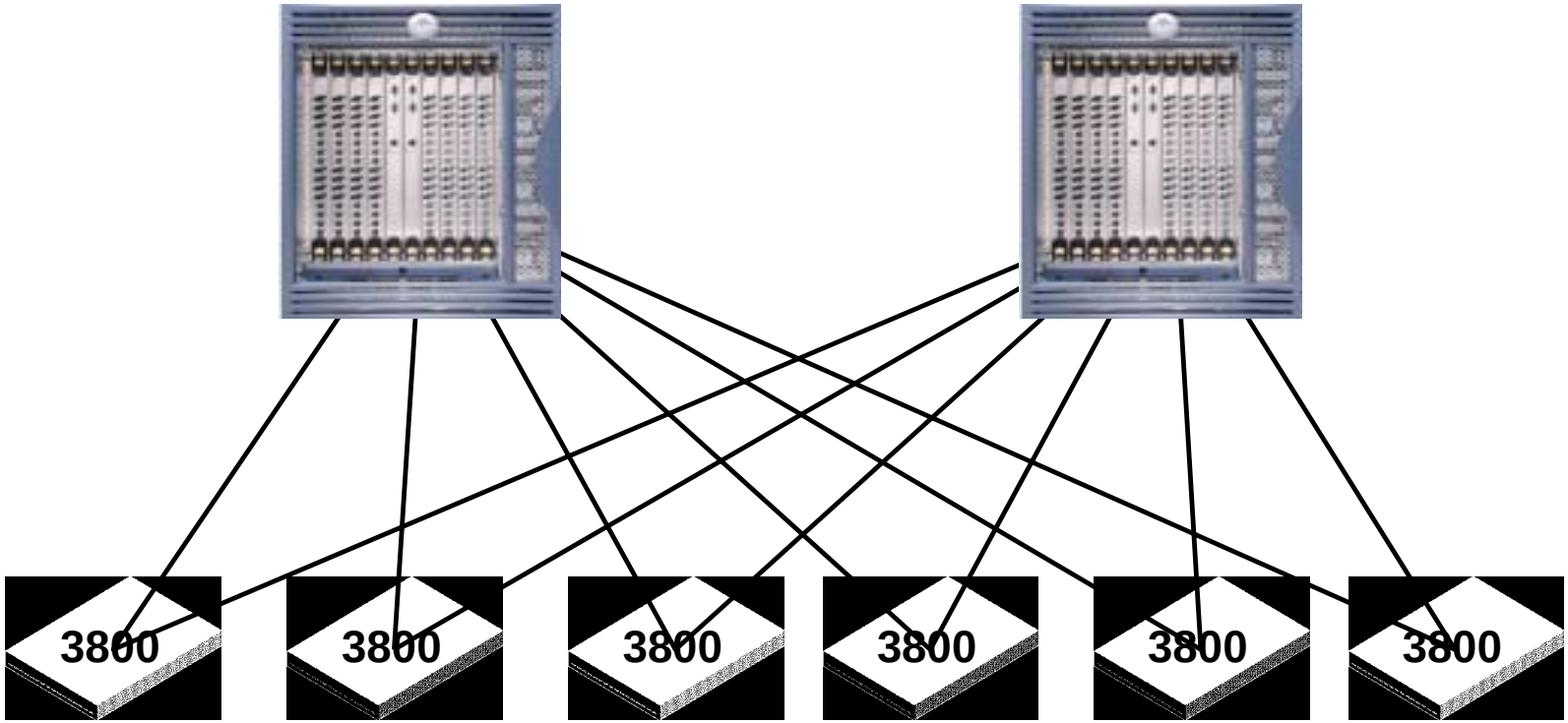


2 x 12000
Up to 124 available user ports

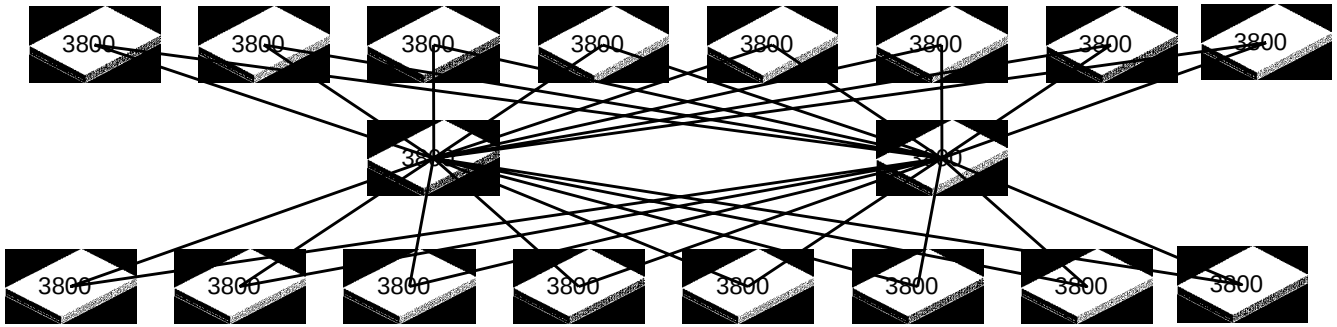


4 x 12000 , 24 x 16-port switches
Up to 500 available user ports

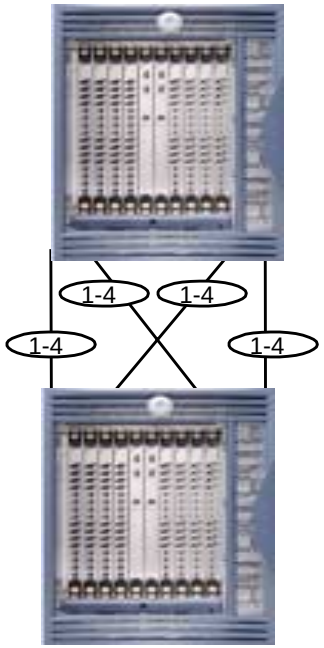
For The Highest Availability Use 2 x SilkWorm 12000 Chassis In The Core



4 x SilkWorm 12000 Switches = ~ 240 ports



18 x 3800
224 available user ports



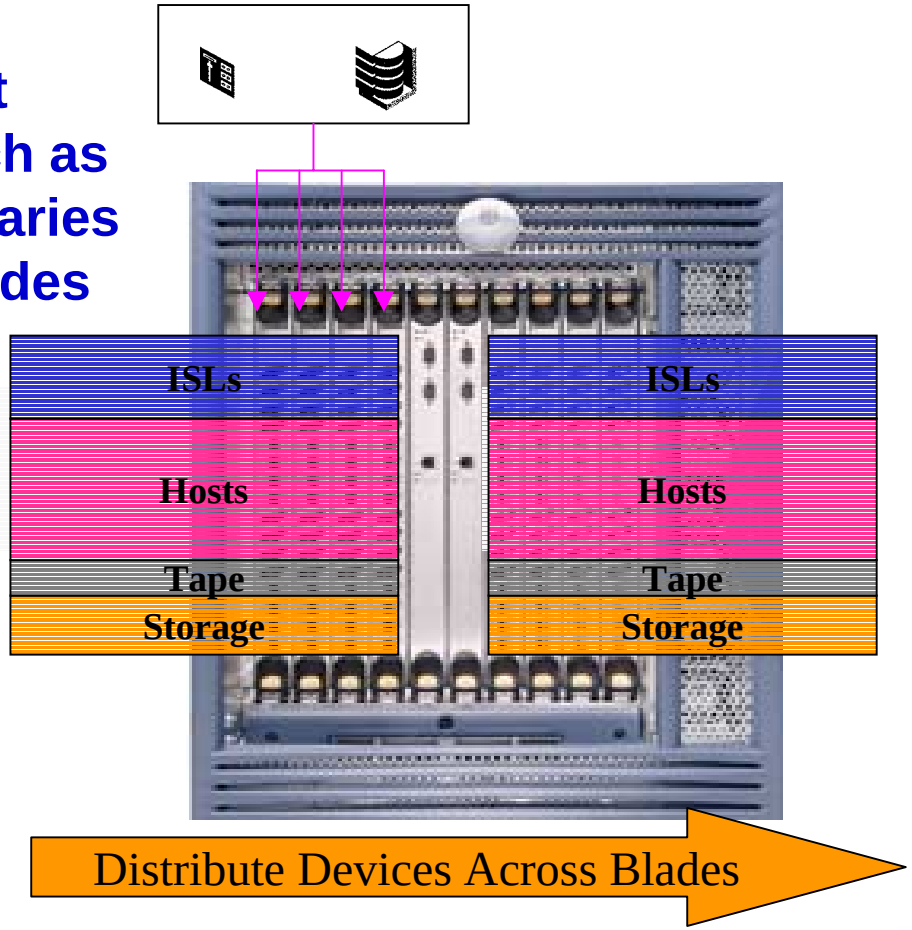
2 x 12000
Up to 248 available user ports

Note: When we say SilkWorm 12000, we mean logical switch. When we say a Silkworm 12000 chassis, it can be 0, 1, or 2 SilkWorm 12000 logical switches.

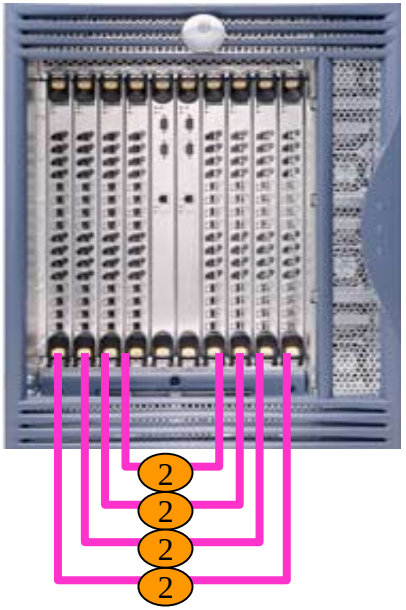
4 x SilkWorm 12000 switches is 2, 3, or 4 SilkWorm 12000 chassis

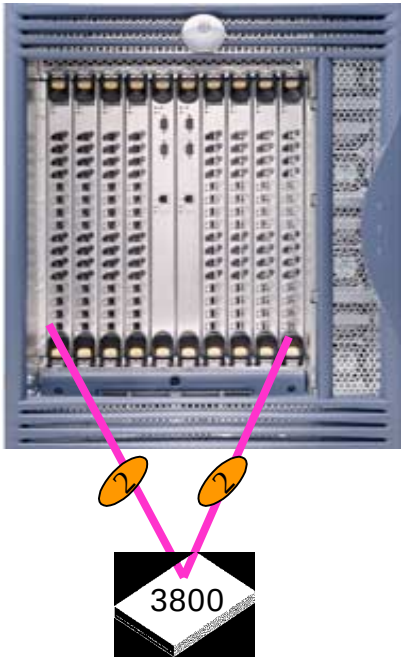
Device Placement Makes A Difference

Distribute High Port Count Devices, Such as Arrays or Tape Libraries Across Multiple Blades



Core Switch 2/64 ISL Trunk Strategies
Balance Performance & Availability





When multiple trunks connect switches, distribute the Trunks across Switch Blade Assemblies. (i.e. 4 x 2 ISL trunks vs. 2 x 4 ISL trunks to optimize performance and availability)

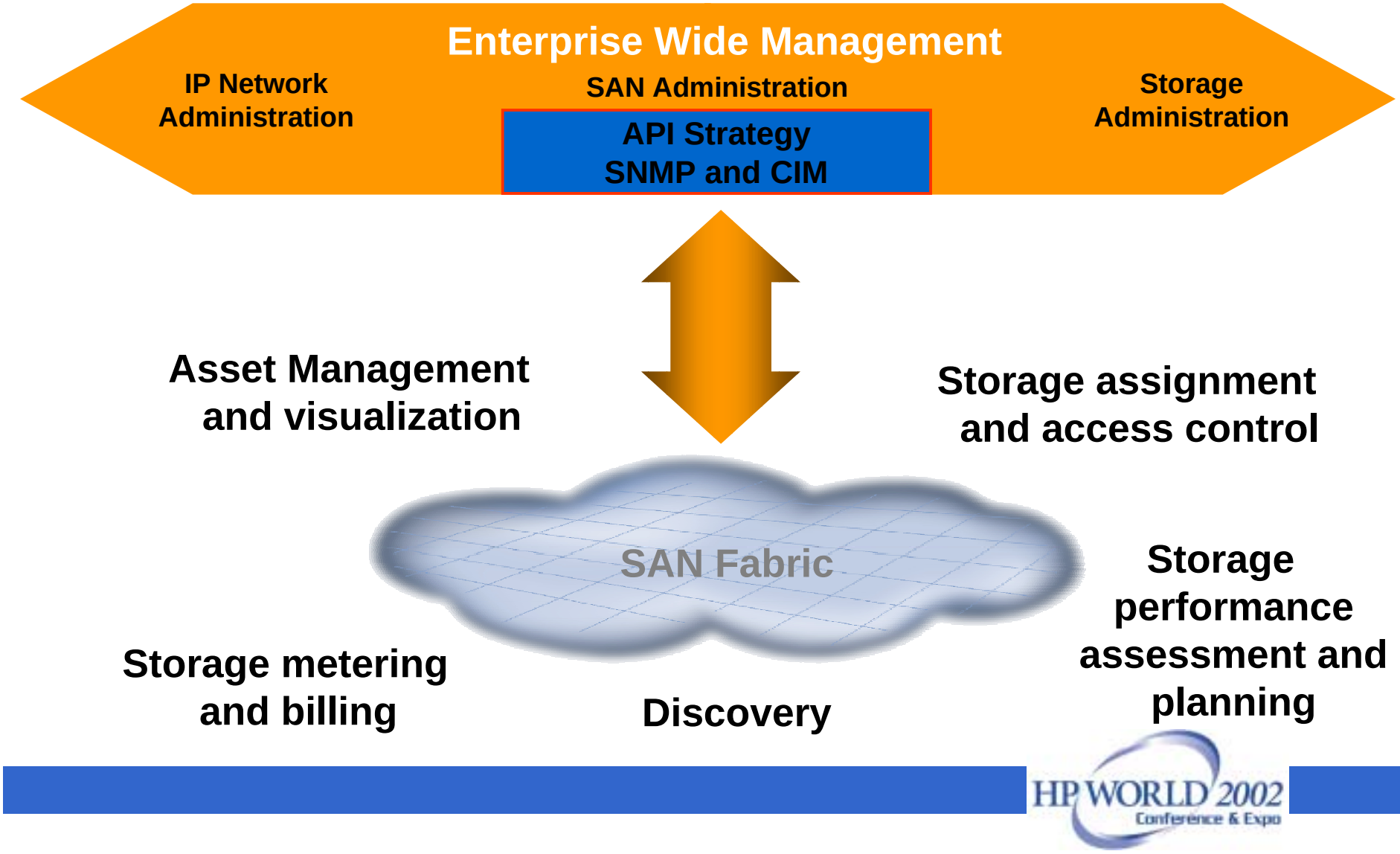
Connect a single 4 Gbit/sec trunk to the Switch Blade Assembly on each logical switch vs. spreading two ISLs across two Switch Blade Assemblies

Summary

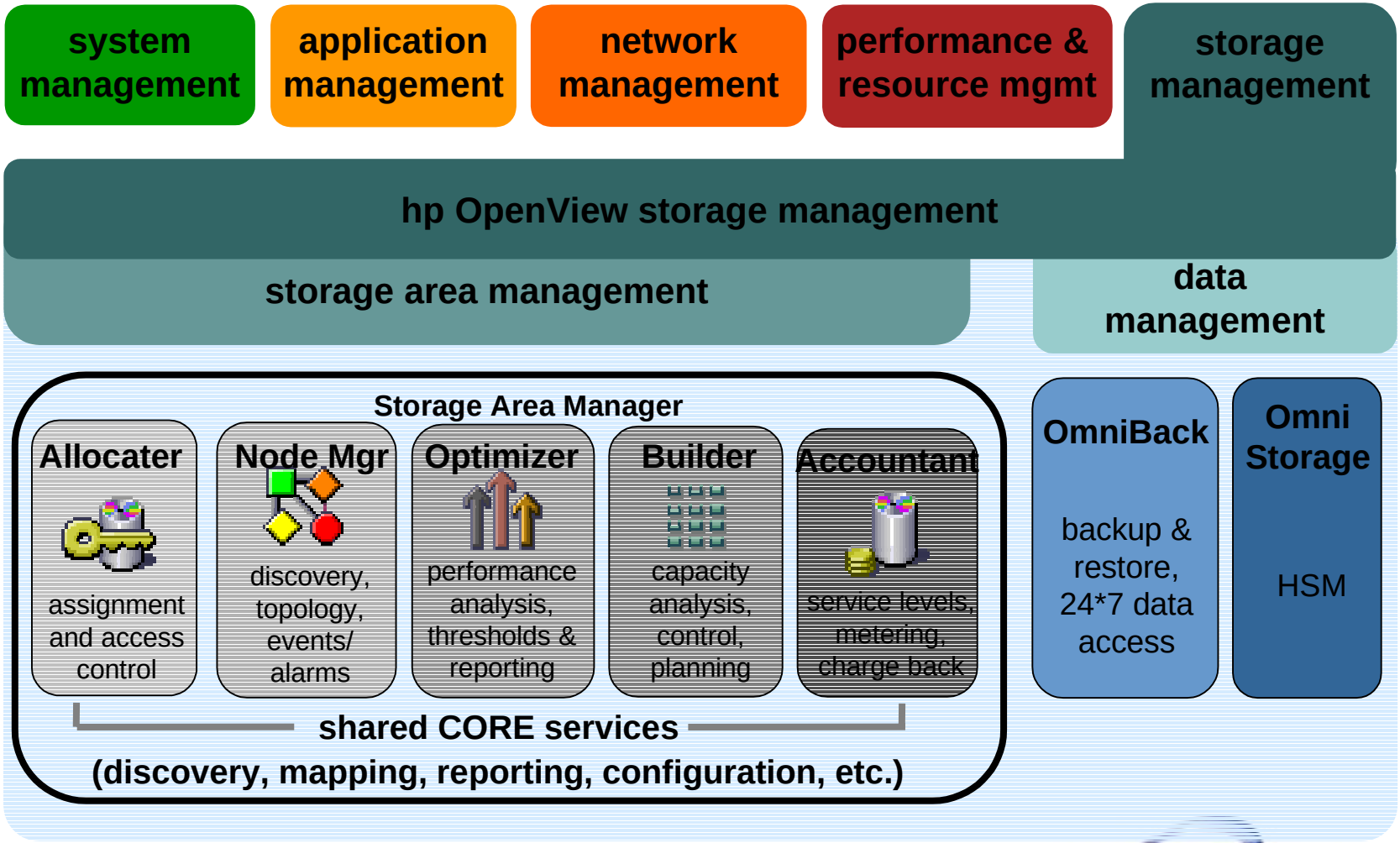
- **Quick SAN Design primer**
  - Core/Edge is a principle SAN design
  - ISL oversubscription ratios
  - Locality
- **Just like carpentry: measure twice, cut once**
  - Or…a little planning with site surveys, checklists etc. goes a long way
- **Core Switch 2/64 based SAN design**
  - 1 fabric per chassis
  - Device attachment strategies are important
  - You can build large core/edge fabrics with the Core Switch 2/64
- **Core Switch 2/64 Implementation tips & checklists**
  - Use v2.6.0c & v3.0.2 or later for 2000, 3000 series switches
  - Slot based and chassis-wide commands are new
  - Enable core PID format
  - Use site survey & install checklists

SAN Management

HP OpenView Management Software



HP Enterprise Management View



HP OpenView

Storage Node Manager

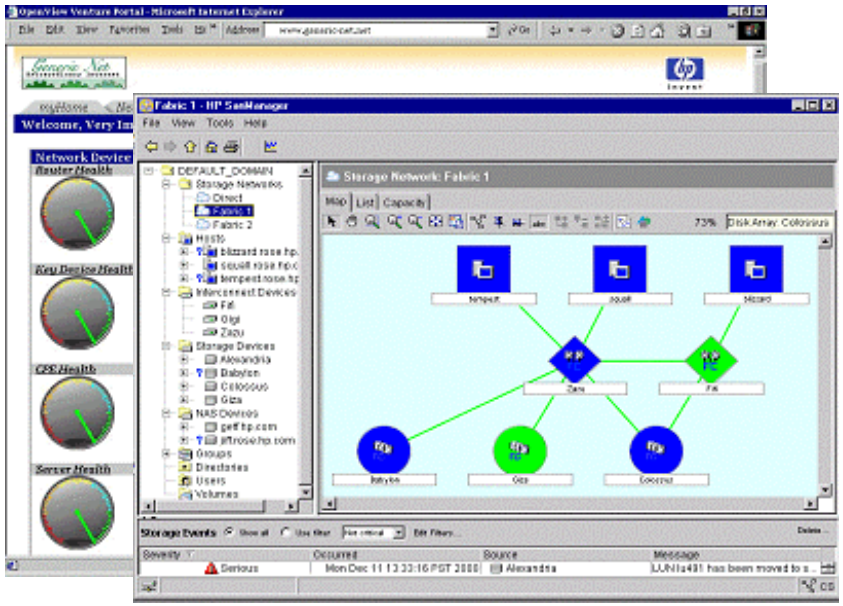
Storage asset management and visualization

Auto-discovery and topology mapping

Continuous event monitoring and status information

Centralized configuration and troubleshooting

Common launch platform for storage device applications



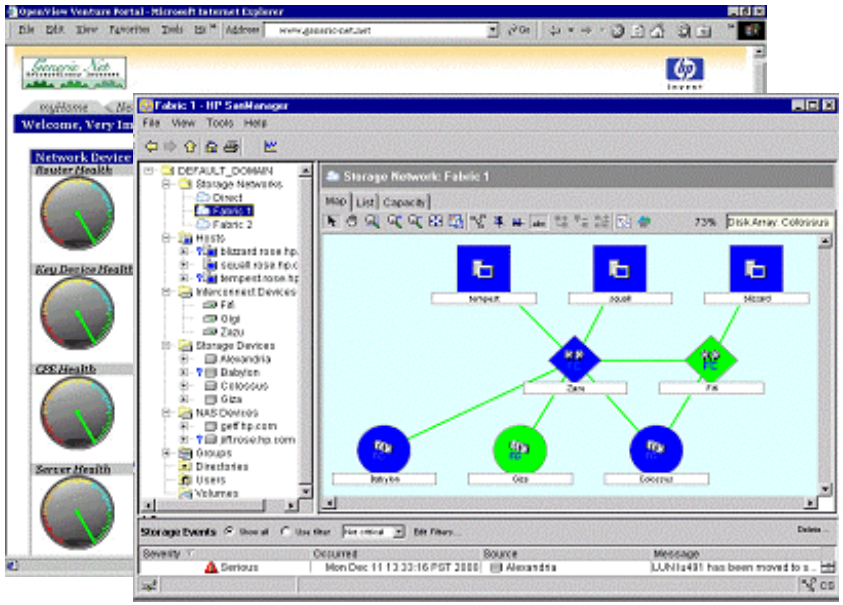
HP OpenView
Optimizer

Storage performance assessment and planning

Performance monitoring of the entire storage network

Historical trending and performance metrics

Extensive reporting



HP OpenView
Builder

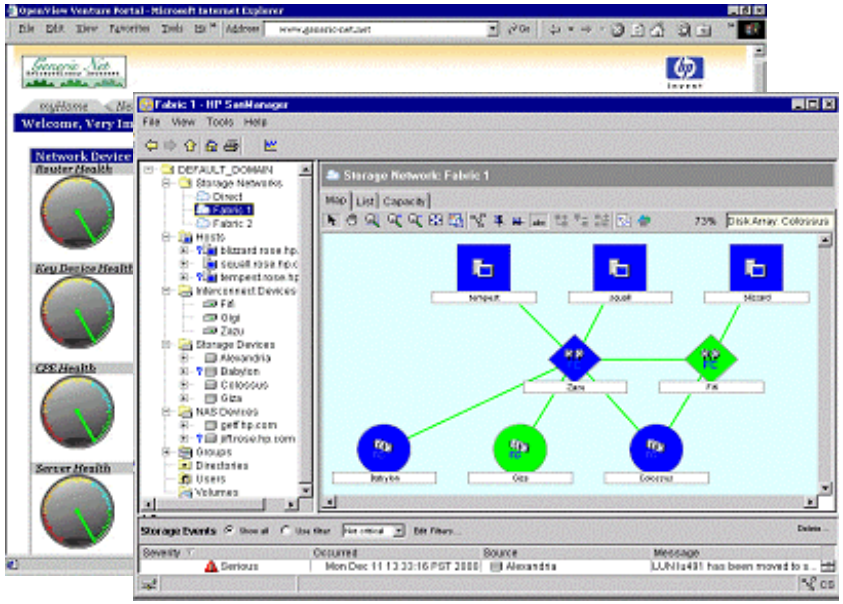
Storage capacity assessment and planning

Storage inventory and monitoring

Automated identification of junk/stale files

Historical trending and future extrapolation

Extensive reporting



HP OpenView
Allocator

Storage assignment and access control

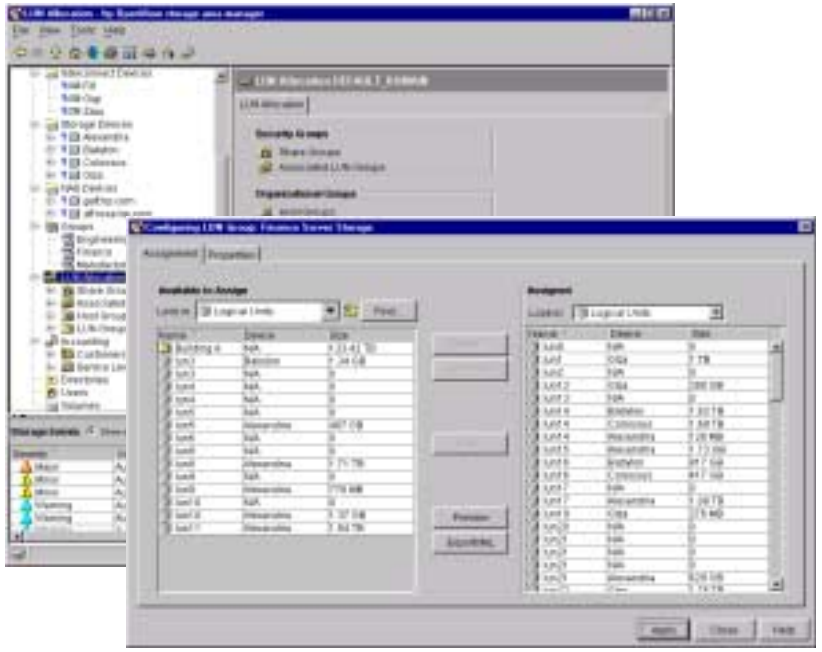
Virtualized access control

Dynamic storage assignment without reboot

Automated host/storage (LUN) discovery

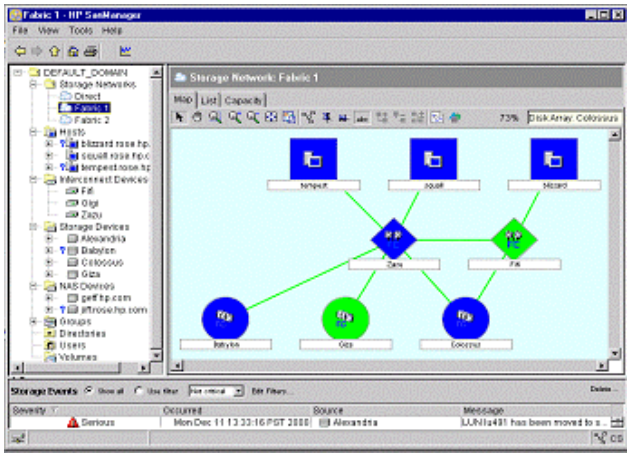
SAN topology independent

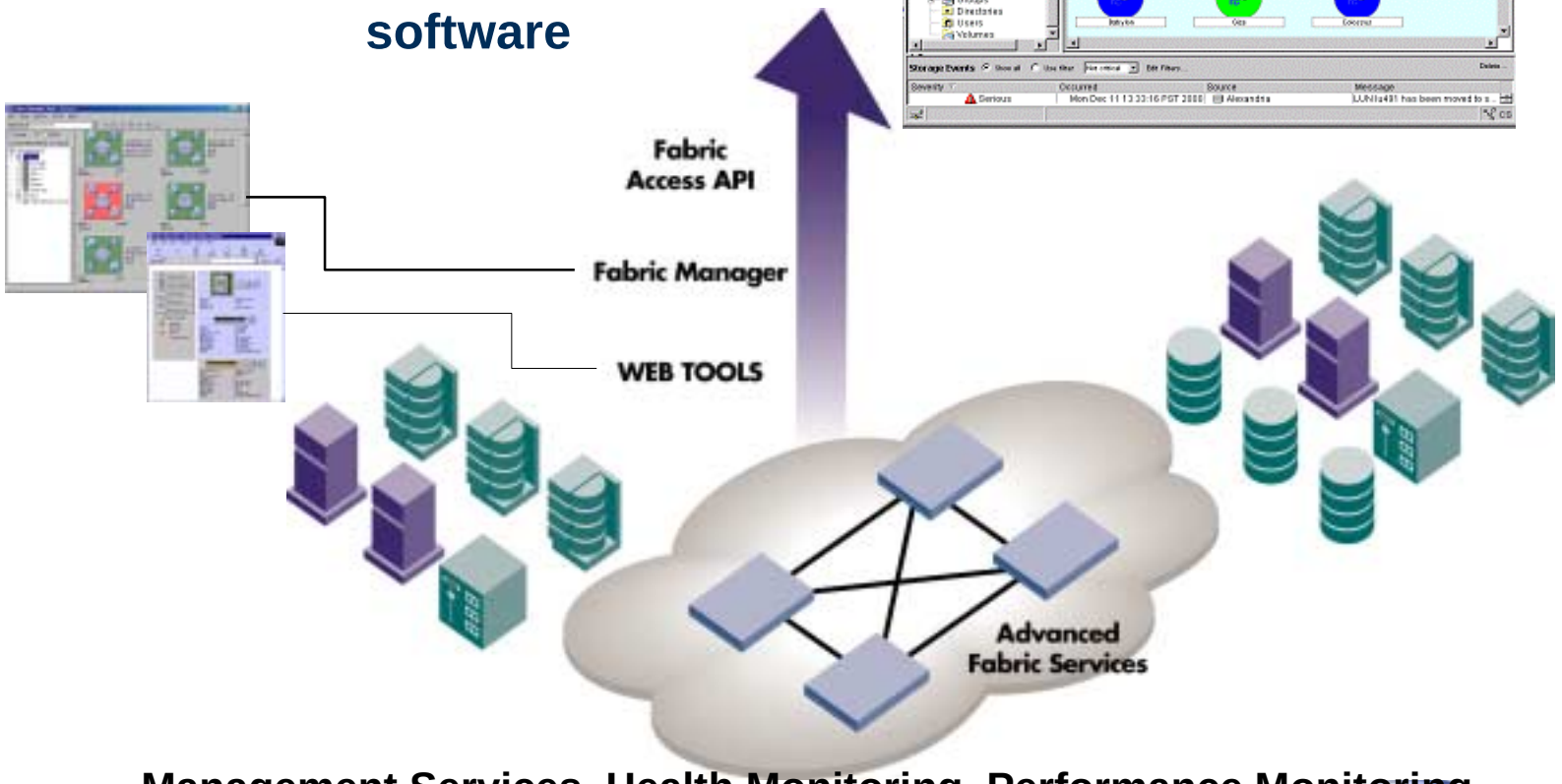
No single point of failure



Open Fabric Management

Integration with HP OpenView Storage Area Management software





Management Services, Health Monitoring, Performance Monitoring

Migration Strategies

Migration Strategies

- **Rolling migration (non-disruptive):**
  - take down one fabric in a dual fabric SAN
  - do migration
  - do the other fabric
- **Live migration (less disruptive)**
  - In a resilient fabric migrate one switch at a time
- **Examples based on Core/Edge fabric and the replacement of the 16-port cores with 64-port cores**
  - Processes are transferable to other topologies / SANs

High Level Steps

- **Prepare**
  - Setup new switches
    (IP address, switch name, non-conflicting domain ID)
  - Prepare a cabling scheme
    - Suggest translation table: from (switch/port), to (switch/port)
  - Profile the fabric
    (switchShow, nsShow, fabricShow, nsAllShow)
    - # of devices per switch, # of switches, # of devices in the fabric
  - Verify that the multipathing software is functional
- **Migrate**
- **Verify that fabric is functional – compare before and after SAN profile**

Live Migration

- If migrating a single fabric SAN, the choices are a *live* migration or to take the fabric down to perform the migration
- A resilient fabric with no SPOF is necessary
- If migrating to a Core Switch 2/64 (SW12K), you need to have Core PID format enabled
- There may be a slight disruption of I/O resulting from the fabric reconfiguration as each switch goes down and comes back up

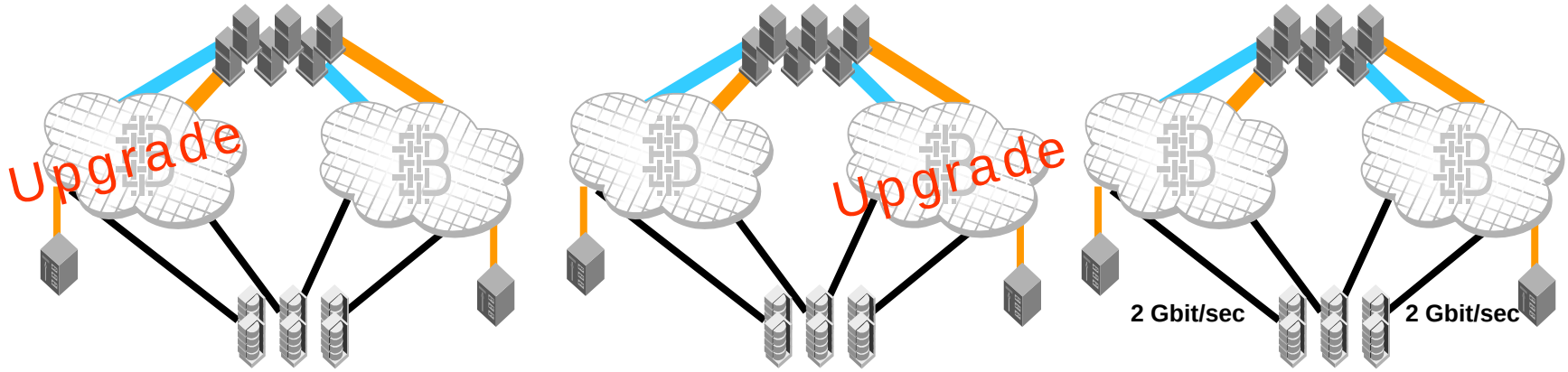
Live Migration Detailed Steps

- Take a SAN profile
- Disable one of the cores
- Verify that multipathing software is functioning and accommodates disabling of core
- Transfer cables from disabled core to new core
- Enable the new core
- Take SAN profile and verify no missing devices and that fabric is functional
- Repeat steps 2 through 6 for the remaining switches in the fabric that need upgrading

Note: Some multipathing software takes several seconds to a minute or more to detect a path failure

Rolling Migration

- Need dual fabric to do this
- Necessary if you need to set core PID format
- No disruption to I/O



Rolling Migration Steps

- **Take a SAN profile**
- **Disable both of the cores**
  - This essentially disables the whole fabric
- **Disable the edge switches**
- **Modify the Core PID if necessary**
- **Verify that multipathing software is functioning and accommodates disabling of the fabrics by sending I/O to the other fabric**
- **Transfer cables from disabled cores to new cores**
- **Enable the edge switches**
- **Enable the new cores**
- **Take SAN profile and verify no missing devices and that fabric is functional and passing I/O**
- **Repeat steps 1 through 9 for the remaining fabric**

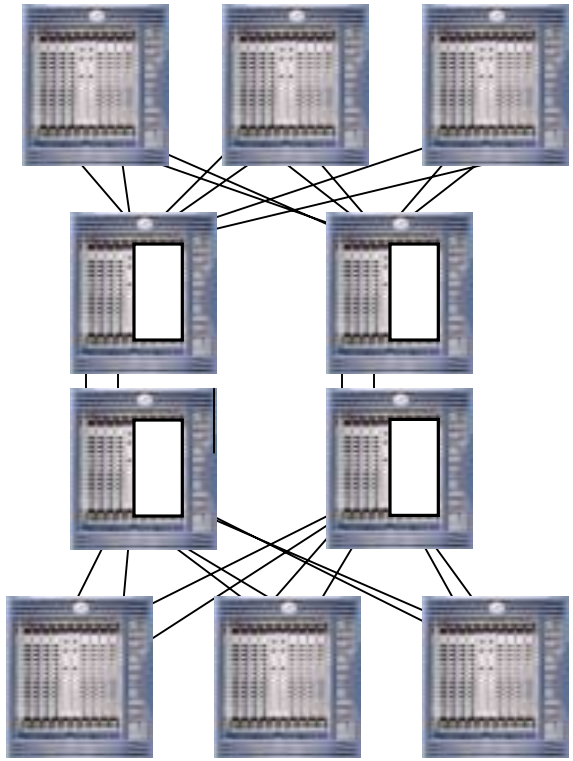
Note: Some multipathing software takes several seconds to a minute or more to detect a path failure

Summary

- **You don't need 2 Gbit/sec devices to take advantage of 2 Gbit/s Fabrics**
  - Put 2 Gbit/sec switches in the core
  - 2 Gbit/sec switches enable higher performing SANs – even if you are using 1 Gbit/sec devices
- **2 Gbit/sec details**
  - Enable Core PID format
  - Trunking minimizes the SAN administration effort and enables optimal performance
- **Migration strategies**
  - Live migration (single or dual fabric)
  - Rolling migration (dual fabric)

What's Next?

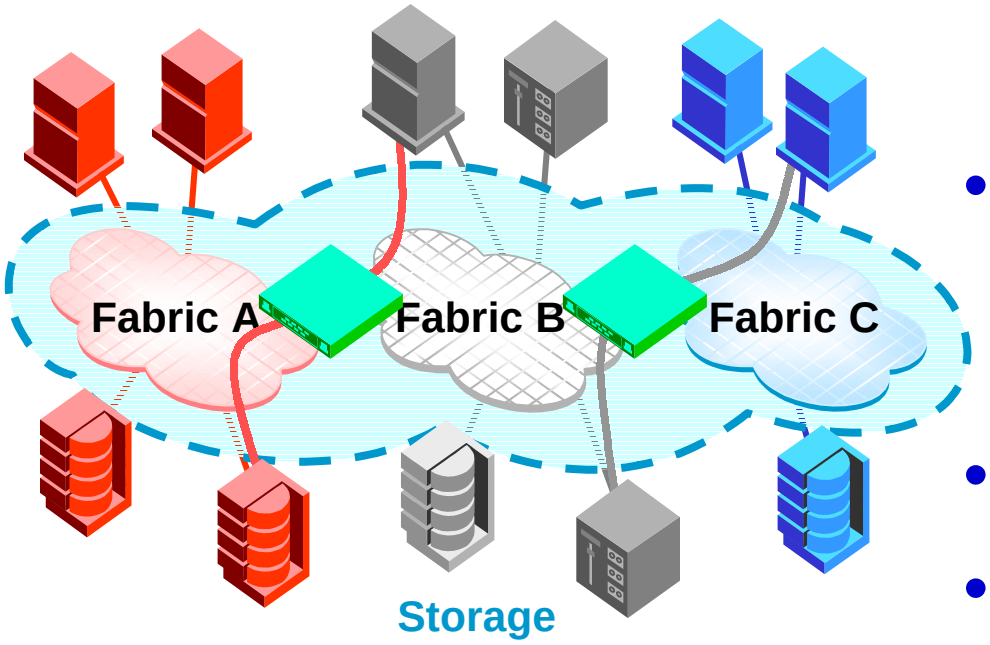
Scalability – The Next Step



- **Testing & supporting larger configurations ~ 1000 ports**

What's Next - SAN Subnets (future)

Accelerating Corporate SAN Growth



- **SANs start as small networks**
  - Project specific
  - Different storage vendors
- **Barriers to merging project SANS**
  - Cost
  - Organizational
- **Inter-Fabric Switch**
  - Simplifies SAN growth by allowing selective resource sharing while maintaining separate management domains
- **Servers see single "Virtual" Fabric**
- **Ensure seamless SAN growth while protecting existing investments**

Thank You

Brocade Communications Systems, Inc.