



Education

VM-Aware SAN: How a SAN can Support Mobility, Security and Manageability

Fabrizio Corno, Cisco Systems

- The material contained in this tutorial is copyrighted by the SNIA.
- Member companies and individual members may use this material in presentations and literature under the following conditions:
 - ◆ Any slide or slides used must be reproduced in their entirety without modification
 - ◆ The SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of the SNIA Education Committee.
- Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.
- The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.

➤ VM-Aware SAN

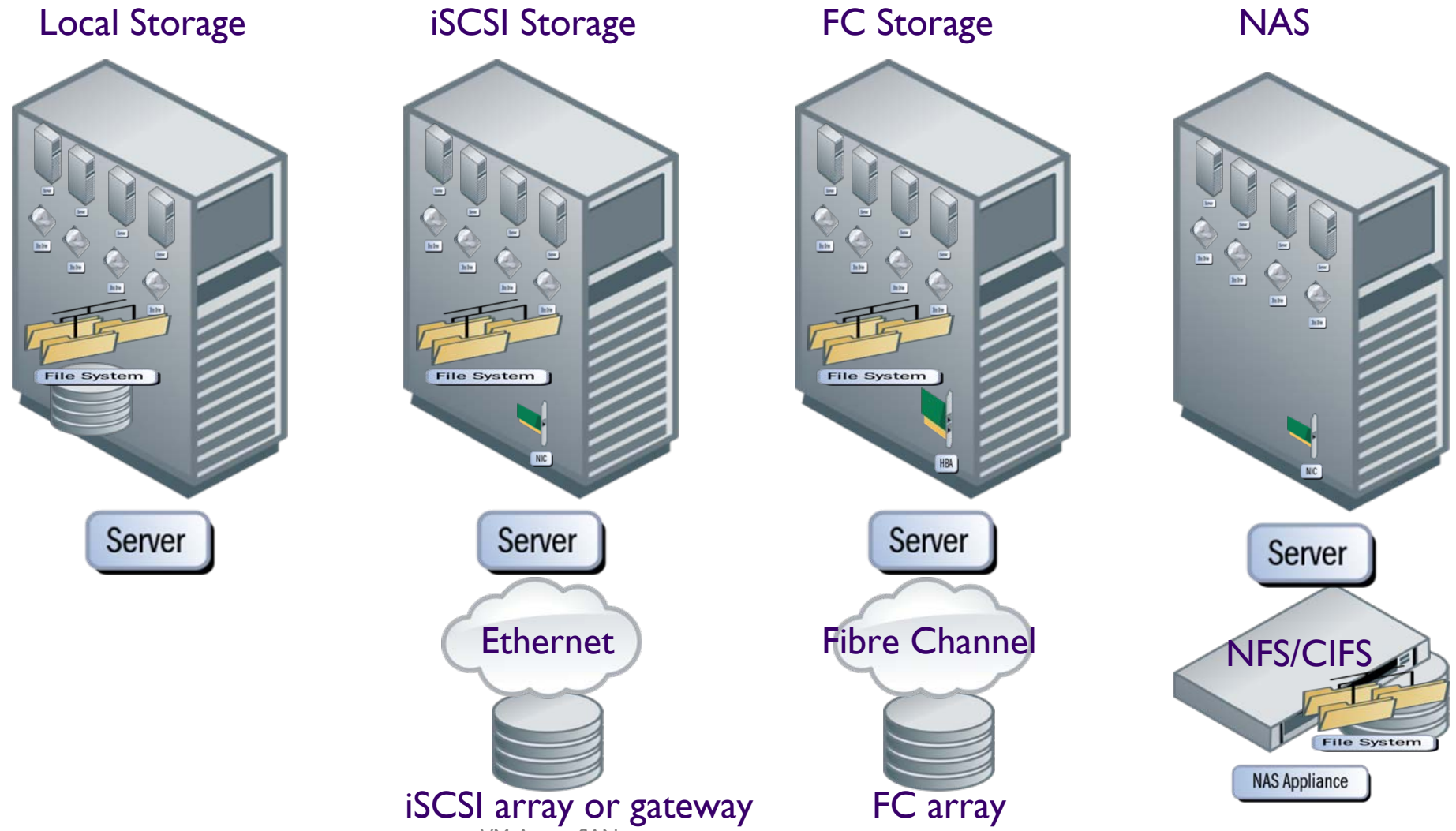
- ◆ This session will appeal to Data Center Professionals, SAN and Storage Architects, and those that are planning the deployment of VM (Virtual Machines).
- ◆ The audience will receive the fundamental grounding in the optional approach an Hypervisor can offer to provide storage to VMs
- ◆ The session will describe the challenges that the adoption of VMs poses on the SAN infrastructure and the features and best practices to overcome them.
- ◆ Focus is on the SAN infrastructure, management and security
- ◆ The terms VM and Hypervisor are used in their generic meaning, without reference to a specific technology

- Hypervisor and VMs (Virtual Machines) and storage networking
 - ◆ Storage connectivity options
 - ◆ Requirements for VMs mobility
 - ◆ NPIV
- Challenges VM deployment poses onto the:
 - ◆ SAN infrastructure
 - ◆ SAN management
 - ◆ SAN Security
- Backup and BC/DR

- Virtual Machine storage connectivity options:
 - ◆ Direct attach
 - ◆ NAS
 - ◆ FC
 - ◆ iSCSI
- SAN storage mapping option
 - ◆ Using Clustered FS
 - ◆ Using SCSI Mapping
- Hypervisor support for NPIV
- Blade server considerations
 - ◆ NPIV/NPV architecture

Virtual Machines Storage Connectivity Options

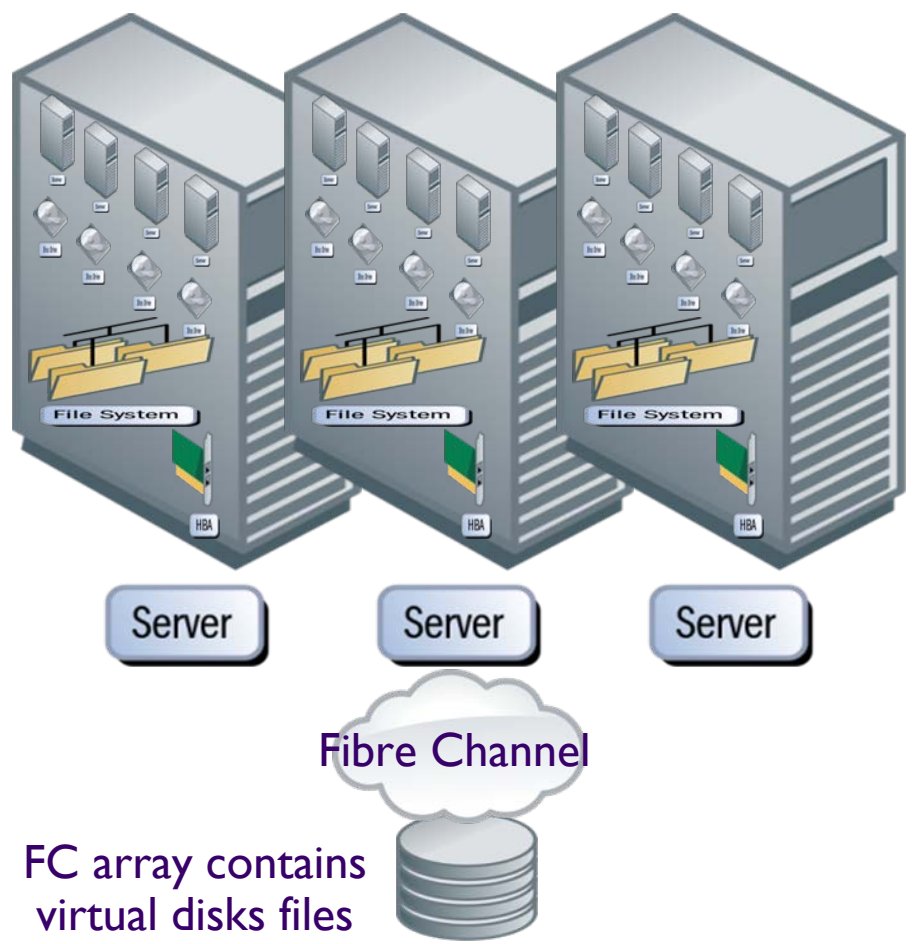
To allow VM mobility the VM data must be accessible at the same time by all servers



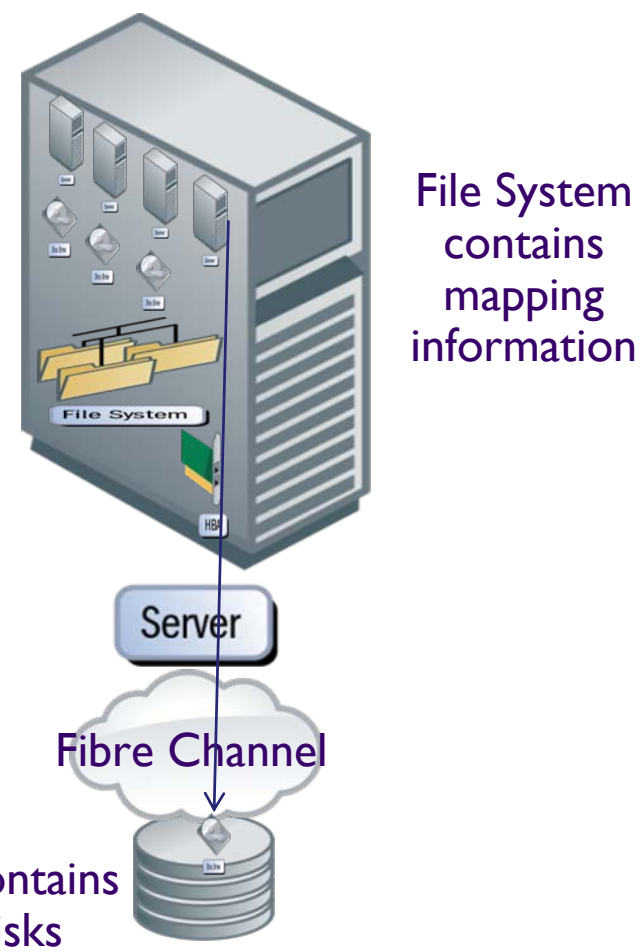
Virtual Machine Storage Mapping Options

To allow VM mobility the file system must be a distributed file system accessible by all servers

Distributed File System



SCSI Mapping

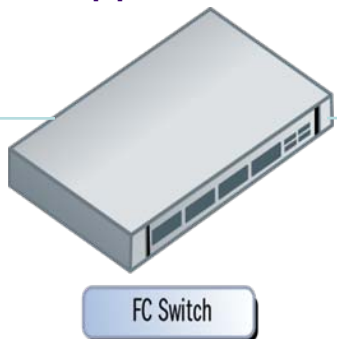


N-Port Device Virtualization (NPIV)

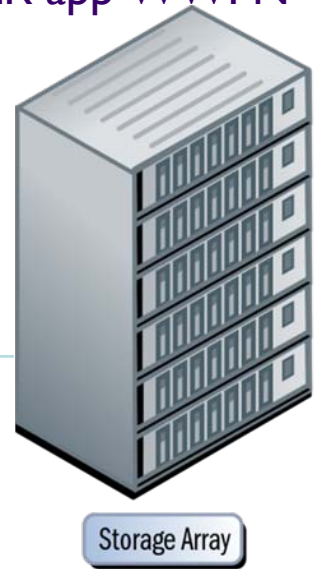
- N port identifier virtualization (NPIV) provides a standard protocol to assign multiple WWPN and FCIDs to a single N port.
- This feature was intended to allow multiple applications to share the same HBA
- The use of different WWPN was intended to allow access control, zoning, and port security to be implemented at the application level.



FC DNS, Zoning:
WEB app WWPN/FCID
EMAIL app WWPN/FCID
FILER app WWPN/FCID



LUN mapping, masking
WEB app WWPN
EMAIL app WWPN
FILER app WWPN



Application of NPIV: Not as Expected! SNIA

The model one server running multiple application did not find much support, so NPIV was not widely deployed as a way to give personality to applications.

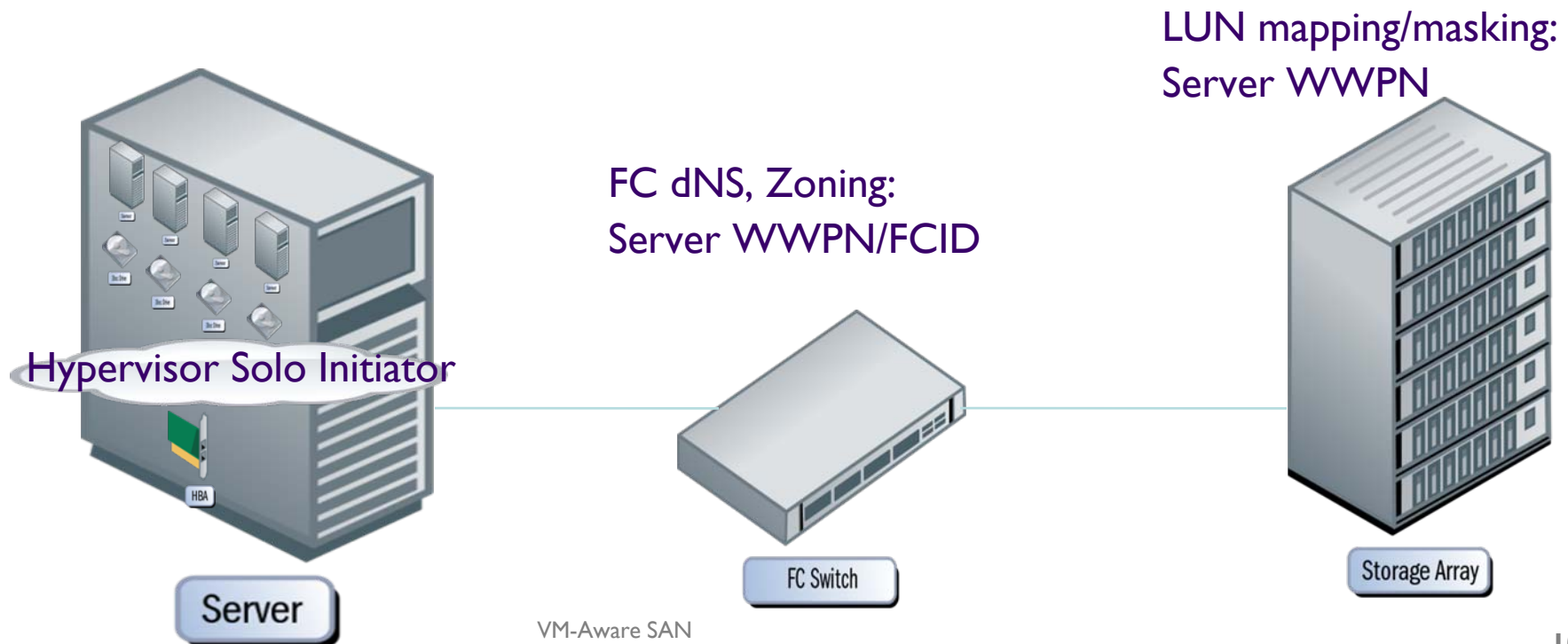
NPIV has a new life to:

- Allow the Hypervisor to associate a virtual HBA to a VM (with limitation, like support for SCSI mapping only). This option has still relatively limited diffusion.
- Implement top of rack and blade FC switches that can provide connectivity to servers without the need of join the fabric. This option is widely adopted.

Hypervisor sharing the physical HBA across VMs

The Hypervisor accesses storage on behalf of all VMs using the same physical HBA WWPN

- Pro: It is extremely easy to configure, the SAN administrator and Storage administrator are not involved in the provisioning of a new VM
- Cons: Storage access control is solely managed by the Hypervisor administrator, a single configuration error can put at risk the data



How the Hypervisor uses NPIV

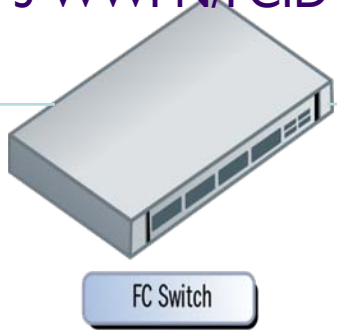
The Hypervisor creates multiple virtual N-Ports (virtual HBAs) to be exclusively assigned to a specific VM

- Pro: storage access control is managed by the Hypervisor and Hypervisor administrator, by the SAN administrator and by the Storage administrator, only multiple configuration errors can put at risk the data. (Other advantages in the following pages)
- Cons: multiple entities are involved in provisioning a VM, it is time consuming and complex as provisioning a physical machine



FC dNS, Zoning:
VM-1 WWPN/FCID
VM-2 WWPN/FCID
VM-3 WWPN/FCID
VM-3 WWPN/FCID

LUN mapping/masking:
VM-1 WWPN
VM-2 WWPN
VM-3 WWPN
VM-3 WWPN



Virtual HBA for the Virtual Machine (I) SNIA

NPIV value propositions:

1. Enables using fabric QoS per individual VM
2. Enables collecting traffic statistics per individual VM
3. Shows the bi-directional association of storage to the individual VM
4. Enables using fabric zoning for individual VM

Virtual HBA for the Virtual Machine (II) SNIA

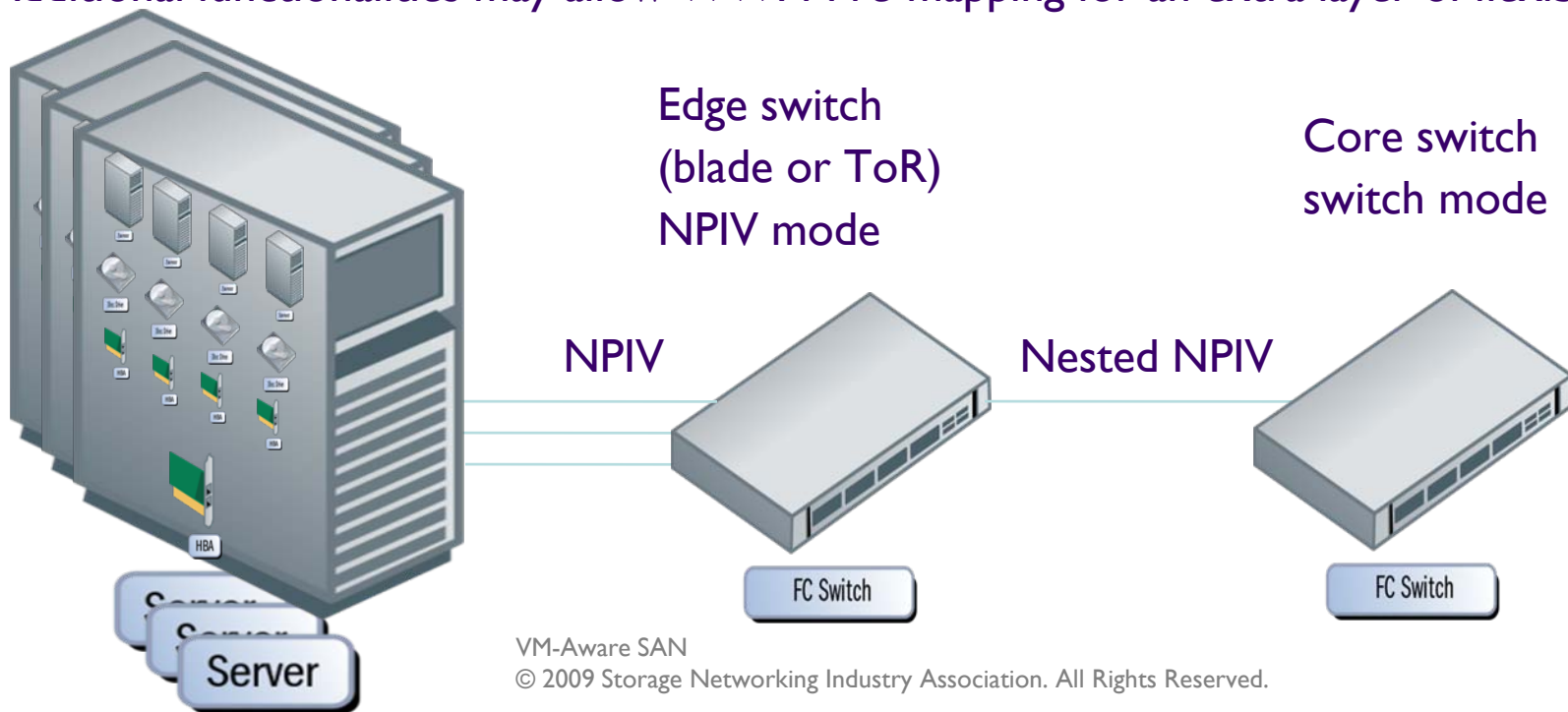
NPIV value propositions (cont.):

5. Enables routing individual VM across virtual fabrics
6. Enables using storage array LUN mapping and masking for individual VM
7. Enables better performance from storage array cache by eliminating “I/O blending”
8. Troubleshooting tools as fcping/fctraceroute can pinpoint the individual VM

Blade Server FC Switch in NPIV Mode

To simplify management and increase scalability (number of Domain IDs), most top of rack and blade switches are configured in NPIV mode

- The edge switch logs in as an NPIV enabled server into the core, no ISL
- It is used to deploy top of rack and blade FC switches that can provide connectivity to servers, without the need of a new switch joining the fabric
- The deployment of VM with virtual HBAs requires support for nested NPIV at the edge switch
- Additional functionalities may allow WWPN re-mapping for an extra layer of flexibility



Blade Switch

Virtualizing HBA names

➤ The Problem:

- ◆ SANs often identify hosts by the hardware ID (Port World Wide Name) of the Host Bus Adapter for purposes such as Zoning (security)
- ◆ Hardware moves and replacements require the admin to change Zoning and other aspects of the SAN configuration

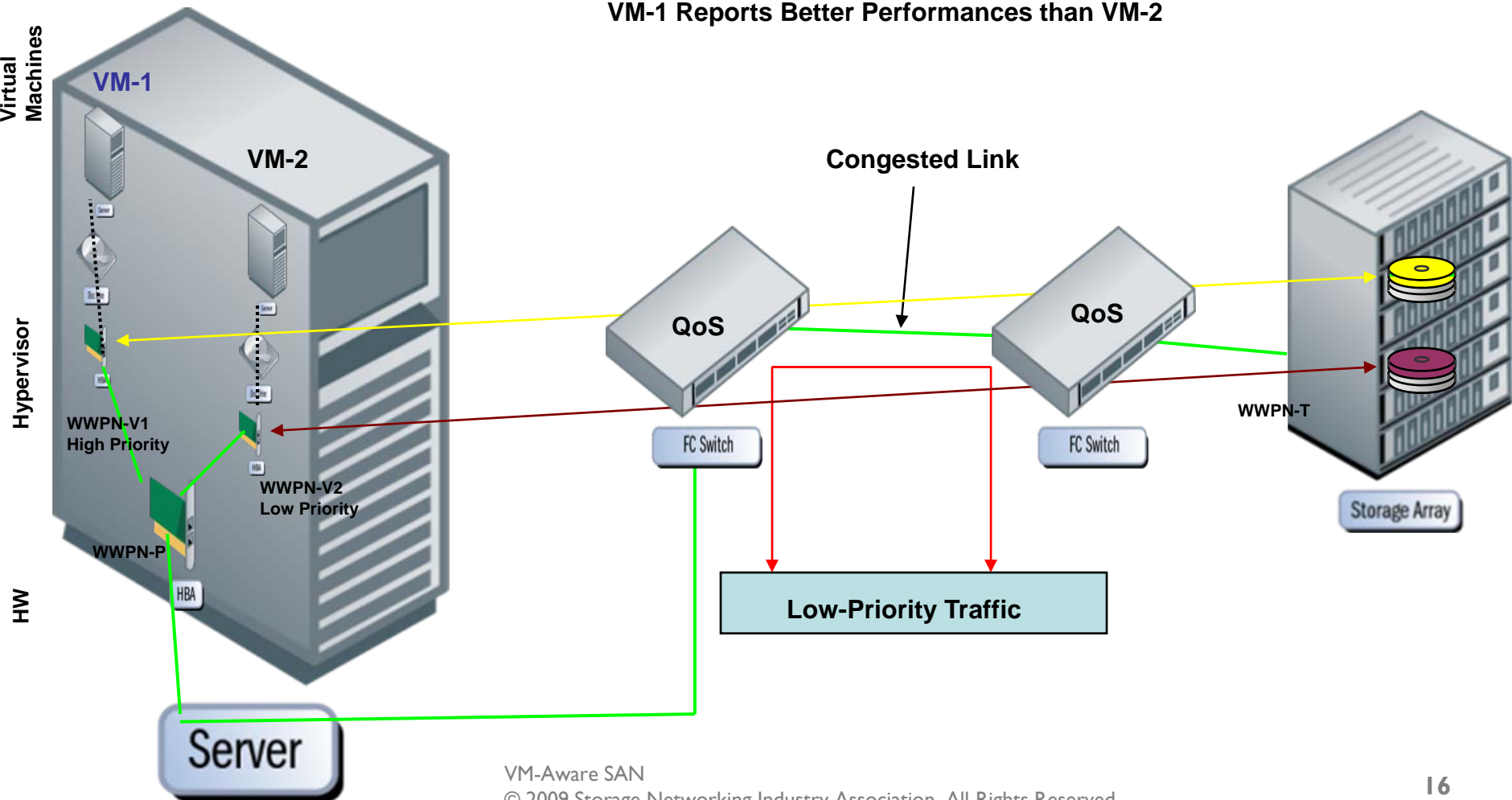
➤ The Solution:

- ◆ Mapping of physical WWPN
 - 1) I want my application to stay here even if I change the hardware (replace the HBA or blade server)
 - ➔ Map virtual WWPN to port on switch
 - ➔ Whatever HBA appears on that port will get that virtual WWPNReduces effort needed for server upgrade or maintenance
 - 2) I want my application to follow the hardware (e.g, if I move the blade server)
 - ➔ Map virtual WWPN to HBA's physical WWPN
 - ➔ Virtual WWPN will follow the HBA or blade if it moves to any other blade switch in the fabric

Example: Zone-Based QoS:

VM-1 has Priority; VM-2 and any Additional Traffic has Lower Priority

VM-1 Reports Better Performances than VM-2



Virtual Machines Pose New Requirements on the Infrastructure

The SAN infrastructure foundation is given by predictable switching performances

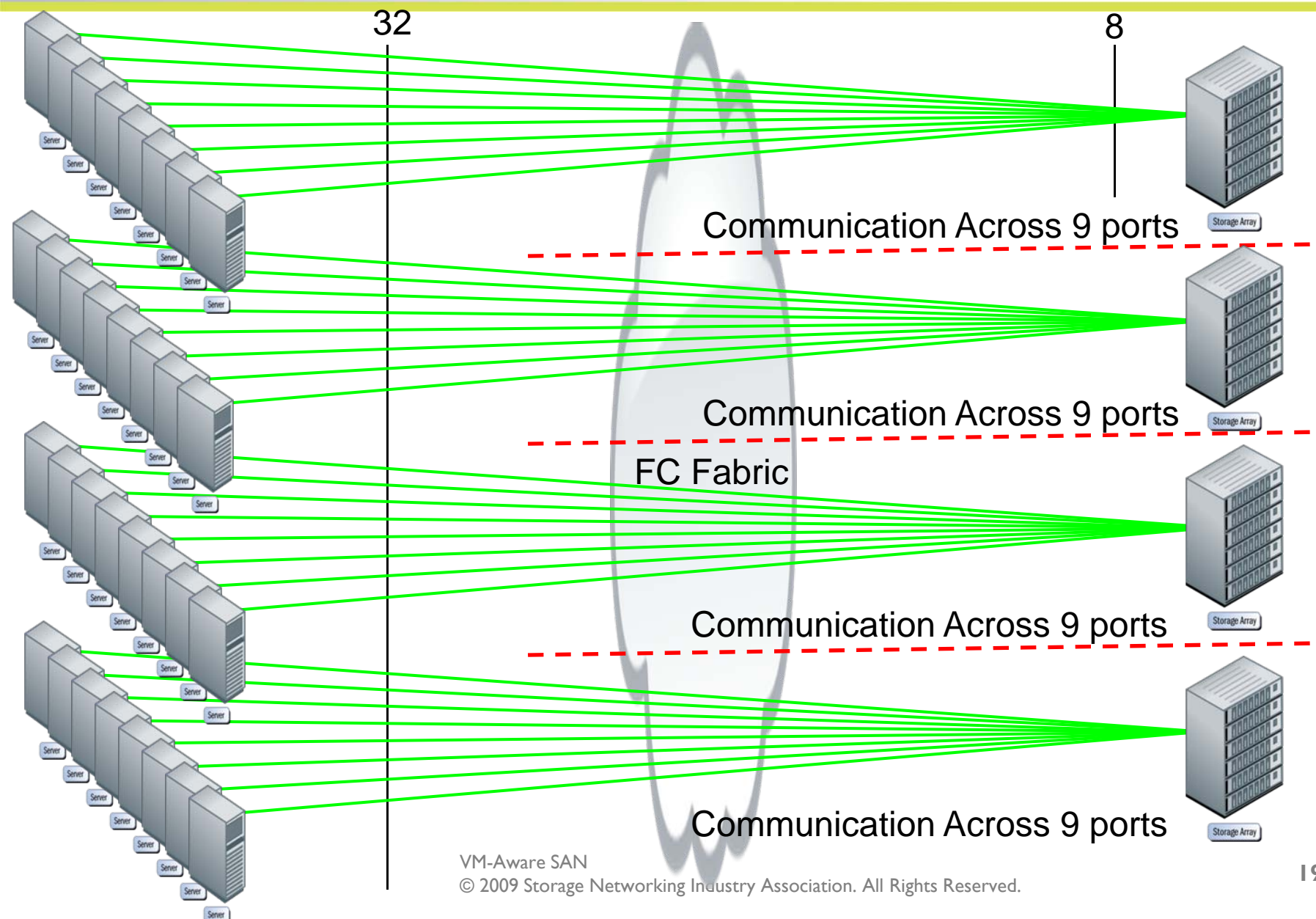
- Support complex, unpredictable, dynamically changing traffic patterns without impact on performances
- Provide traffic scalability for higher workload
- Differentiate QoS on a per VM basis

New traffic patterns challenges the SAN infrastructure

Traditional Data Center	Deployment of VM
Each application server accesses a storage port	Each application server mapped to a VM in a cluster of physical servers
Many-to-one Traffic Many application servers share the same storage port	Many-to-few or Partial Mesh Traffic Each physical server accesses all the available storage ports
Predictable Traffic Patterns Each application server is statically connected to a specific fabric port	Unpredictable Traffic Patterns VMs move across physical servers

Prevalent Traffic Pattern: Many-to-One

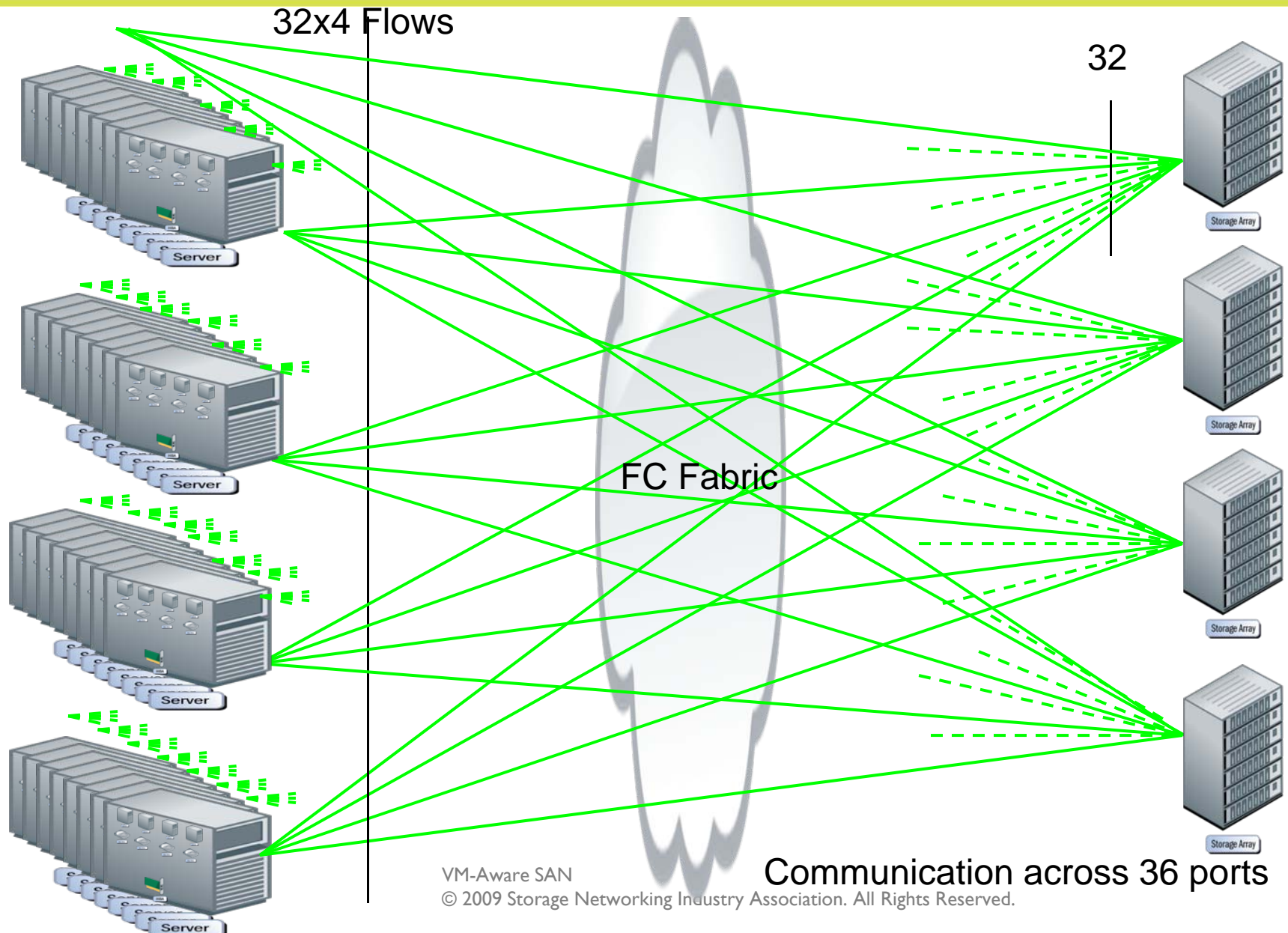
32 Physical Application Servers



Local-Switching-Friendly-Many-to-One Traffic

Prevalent Traffic Pattern: Many-to-Many

32-Nodes Hypervisor Cluster



Local-Switching-Unfriendly Partial-Mesh Traffic

SAN Design to support VM traffic

The VM cluster traffic distribution requires that the switches and SAN architecture provide the best performance in variable traffic conditions:

➤ SAN

- ◆ placement of devices to optimize availability rather than locality

➤ Switches

- ◆ Prevention of head-of-line blocking
- ◆ Even and predictable throughput and latency for many-to-one and many-to-few traffic conditions
- ◆ 100% wirespeed for both large and small frames
- ◆ Fair load-balancing for both large and small frames
- ◆ QoS granularity up to the individual VM (possible with NPIV)

Virtual Machines Pose New Requirements on the Management

➤ Deployment, Management, Security

- ◆ Create flexible and isolated SAN segments, support management Access Control
- ◆ Support performance monitoring, trending and capacity planning up to each VM
- ◆ Allow VM mobility without compromising security

The Virtual Fabric technology complements the VLAN technology and the VM technology to completely virtualize the Data Center

FC-LS (Link Services) rev 1.62

“The Virtual Fabric Tagging Header (VFT_Header, see FC-FS-2) allows Fibre Channel frames to be tagged with the Virtual Fabric Identifier (VF_ID) of the Virtual Fabric (VF) to which they belong.

Tagged frames (i.e., frames with a VFT_Header) belonging to different Virtual Fabrics may be transmitted over the same physical link. The VFT_Header may be supported by N_Ports, F_Ports and E_Ports (see FC-FS-2).”

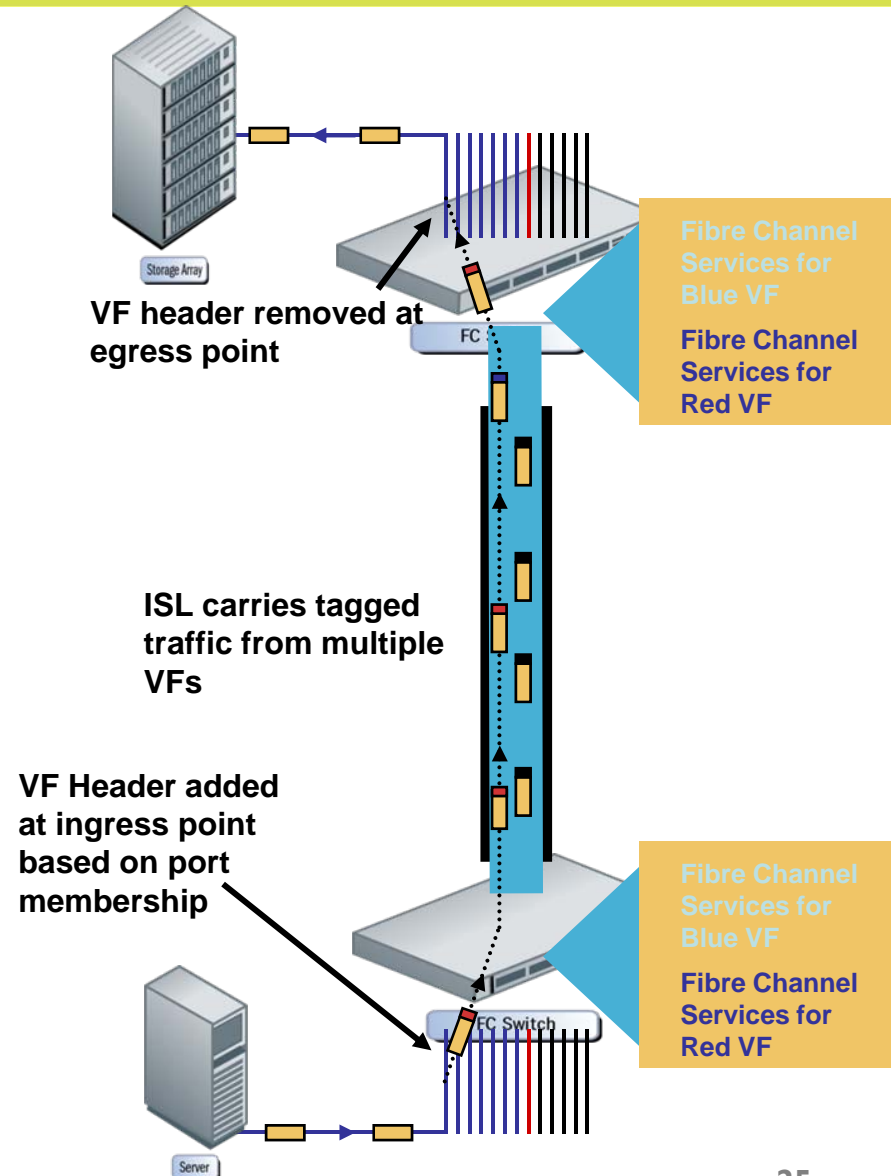
A physical HBA uses a different unique WWPN per VF

➤ Hardware-based isolation of tagged traffic belonging to different VFs

- ◆ Traffic tagged at Fx_Port ingress and carried across links between switches
- ◆ Any switch interface in the fabric can be placed in any VF

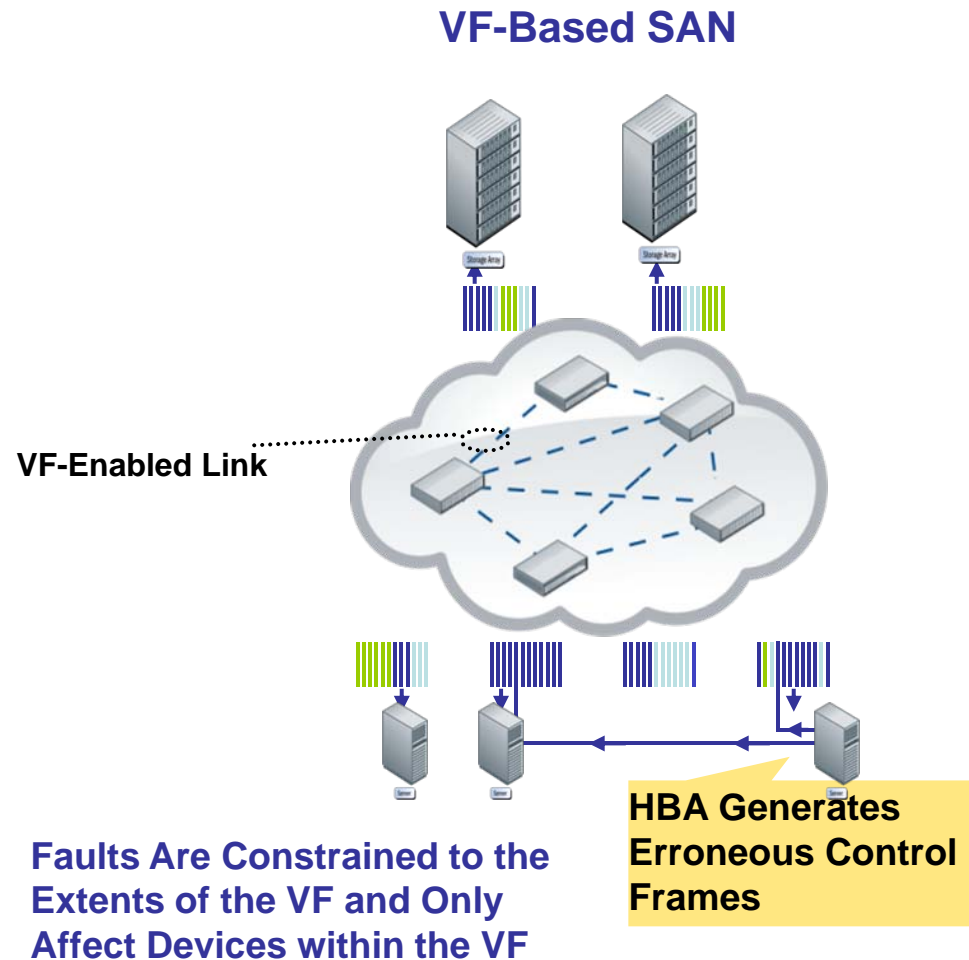
➤ Independent instance of Fibre Channel services for each newly created VF

- ◆ Zone server, name server, management server, principle switch election, etc.
- ◆ Each service runs independently and is managed/configured independently



Virtual Fabrics Sectionalize Fabric and Contain Fault Impacts

- All fabric services replicated and maintained per VF (name services, zoning services, etc.)
- Fabric events isolated per VF—maintains isolation for HA
 - ◆ Misbehaving HBA or controller
 - ◆ Fabric rebuild event
 - ◆ Zone set change
 - ◆ RSCNs—all forms
 - ◆ etc.
- Fabric recovery from a disruptive event ‘per VF’ results in faster reconvergence—smaller scope
- Management segregation per VF



Synergistic and complementary with the goals for adopting Virtual Machines

➤ Consolidation

- ◆ Drive SAN utilization and reduce over provisioning

➤ High Availability

- ◆ Isolation prevents any undesired access across VFs
- ◆ Isolation improves Data Security Standards and Regulation compliance
- ◆ Faults and mis-configurations contained to a given VF

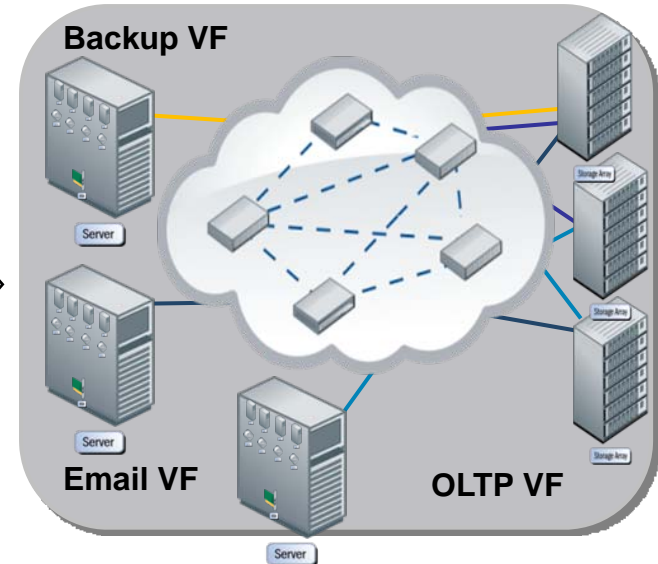
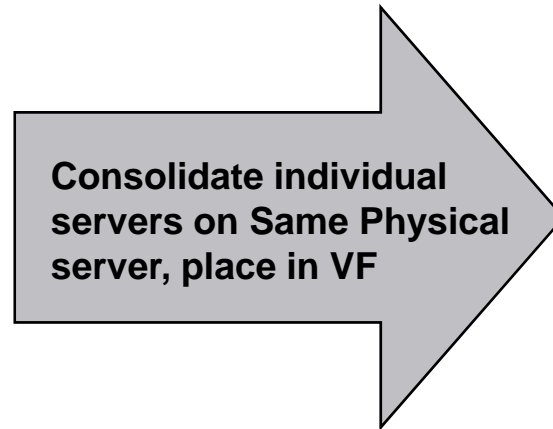
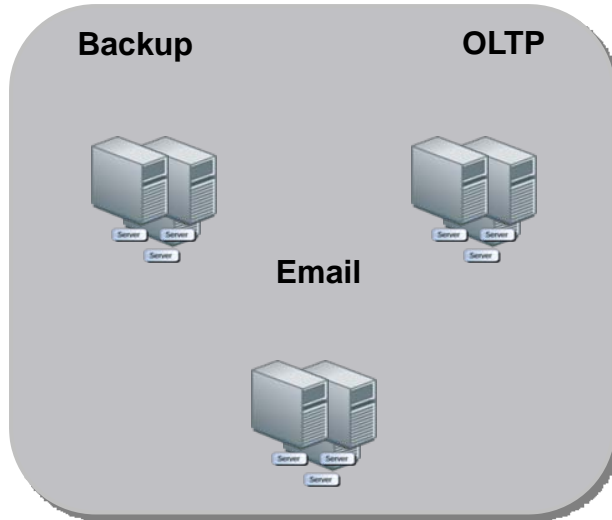
➤ Performance

- ◆ QoS, SLA provided at the desired level independently for each VF

➤ Easy management

- ◆ No wiring reconfiguration
- ◆ Individual administration for each VF using RBAC (Role Based Access Control)

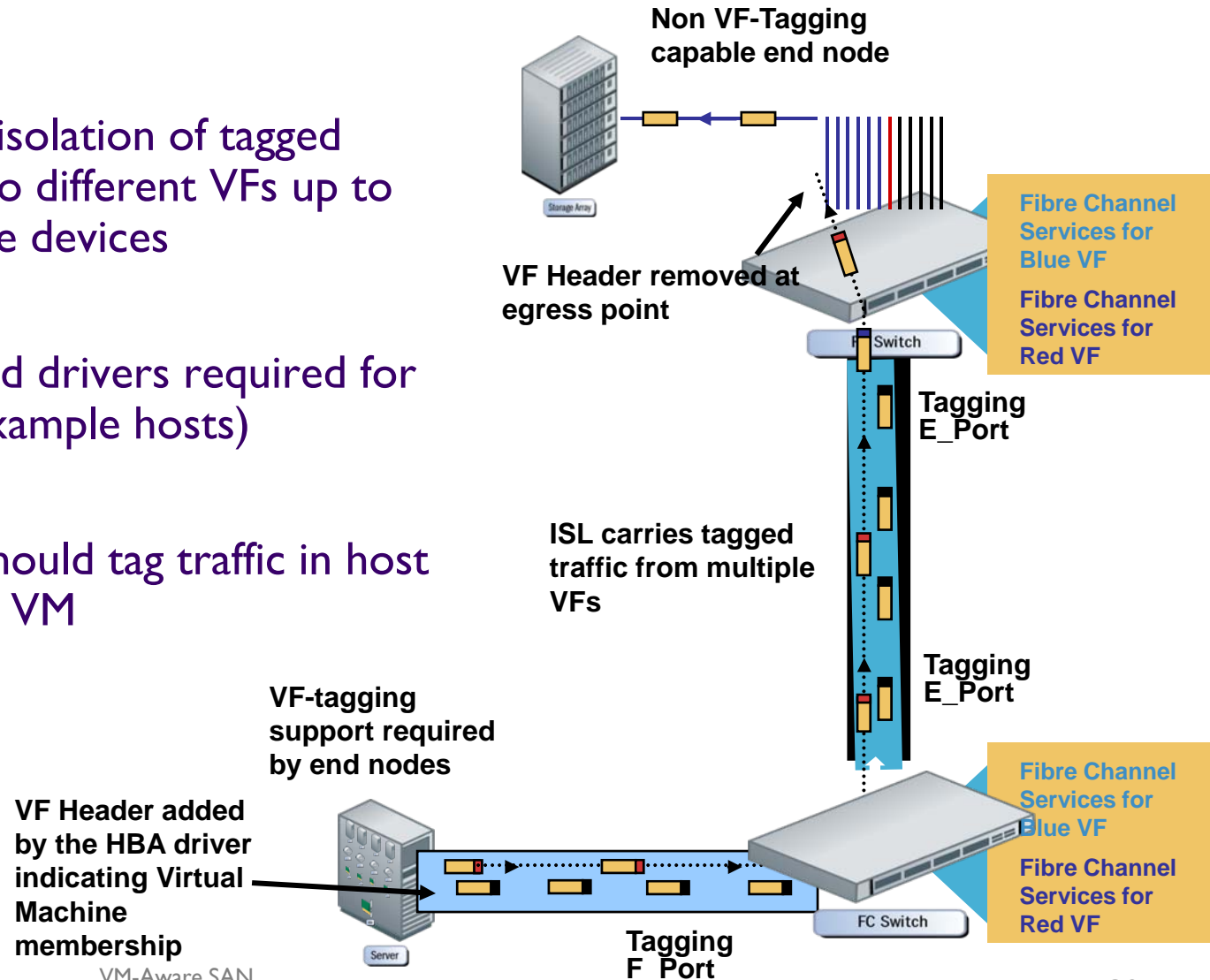
Consolidation using Virtual Machines and VFs



SAN Islands	Attribute	Consolidated VFs
More	Number of SAN Switches and Servers	Fewer
No	Share Disk/Tape	Yes
No	Share DR Facilities	Yes
Complex	Server Management	Simple
High	Overall TCO	Low

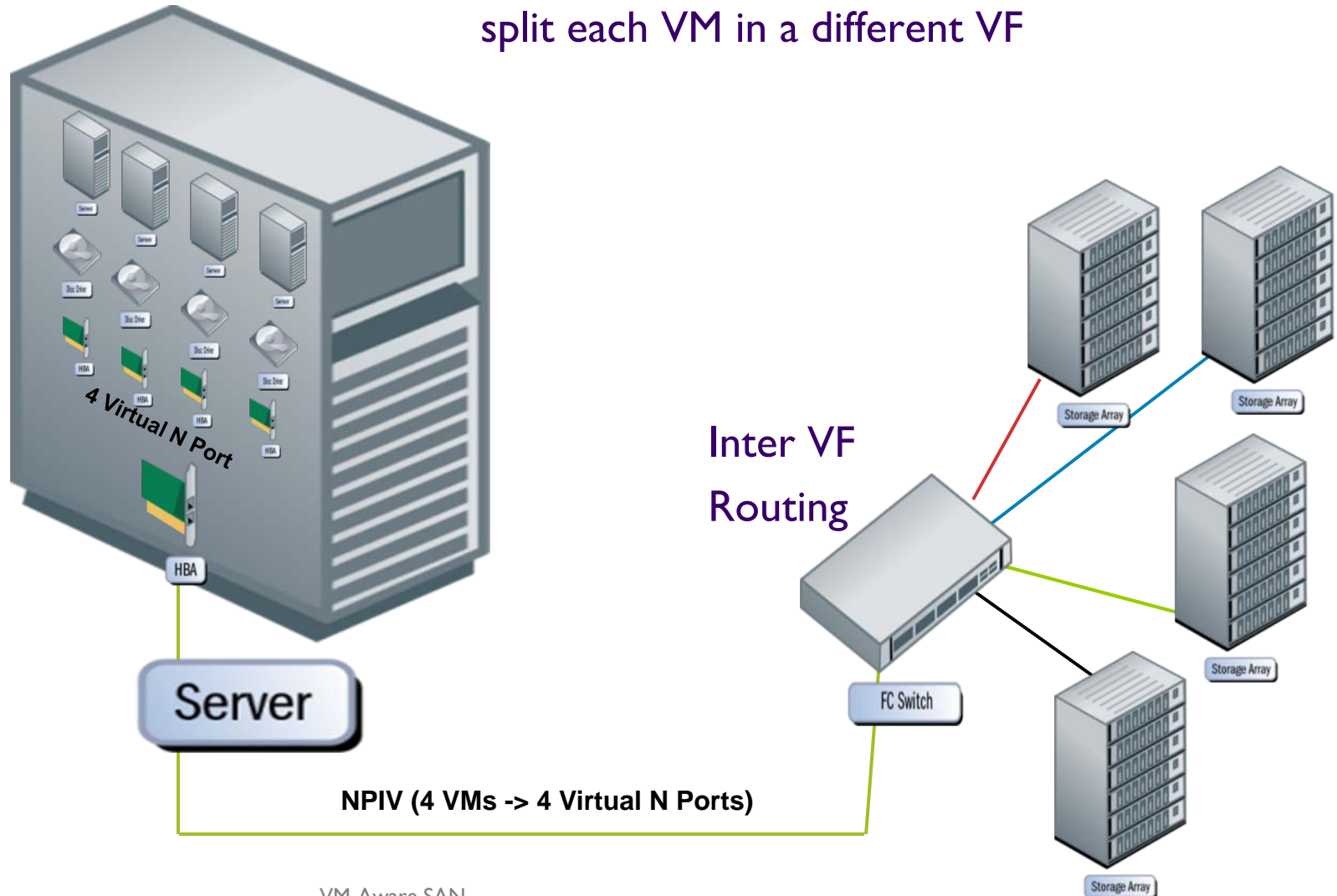
Extend VF tagging to the N_Port-F_Port Connection

- Hardware-based isolation of tagged traffic belonging to different VFs up to servers or storage devices
- VF-tagging-enabled drivers required for end nodes (for example hosts)
- The hypervisor should tag traffic in host depending on the VM

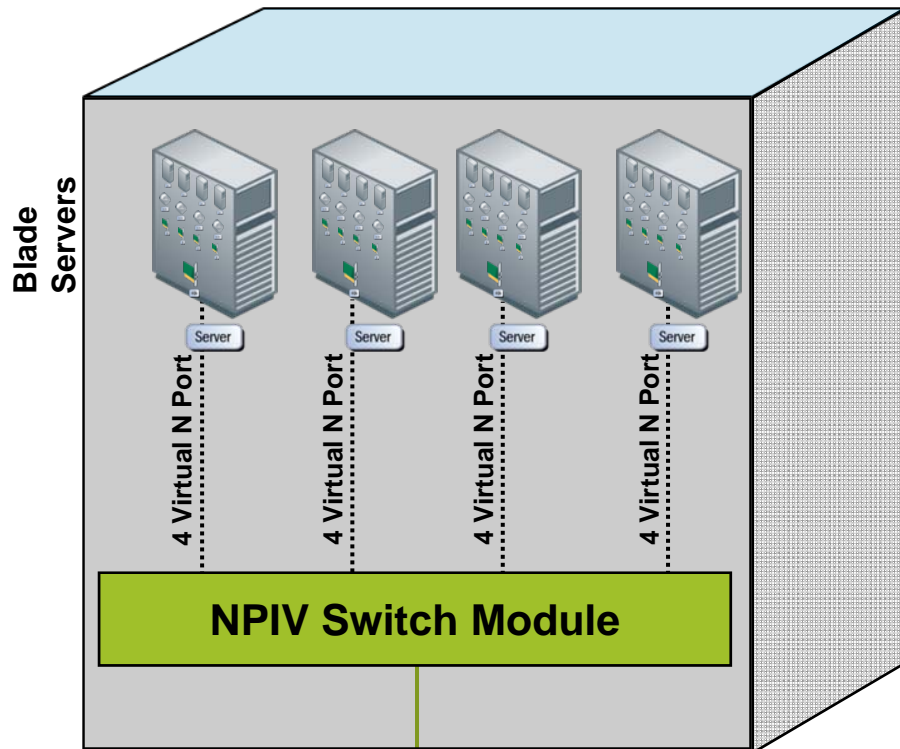


Routing Virtual Machines across VFs Using NPIV and Inter VF Routing

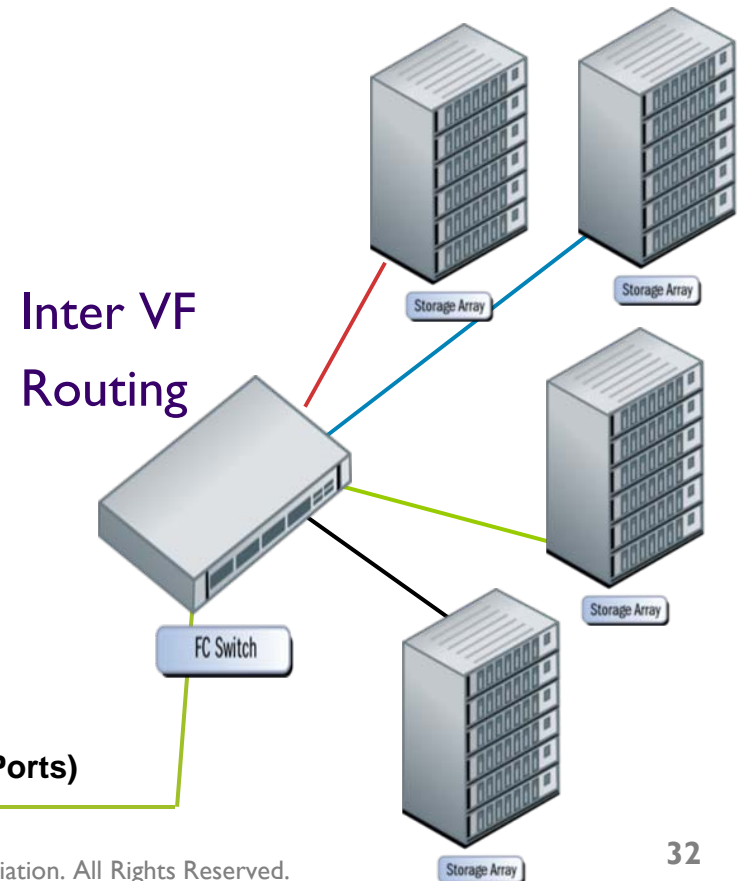
In combination with NPIV, Inter VF Routing can split each VM in a different VF



Using VMs in Blade Servers with NPIV Mode Switch and Inter VF Routing



In combination with NPIV, Inter VF Routing can split each VM in a different VF



NPIV (16 VMs -> 16 Virtual N Ports)

Clusters of servers running virtual machines becomes isolated “Virtual Data Center” by deploying VF and RBAC

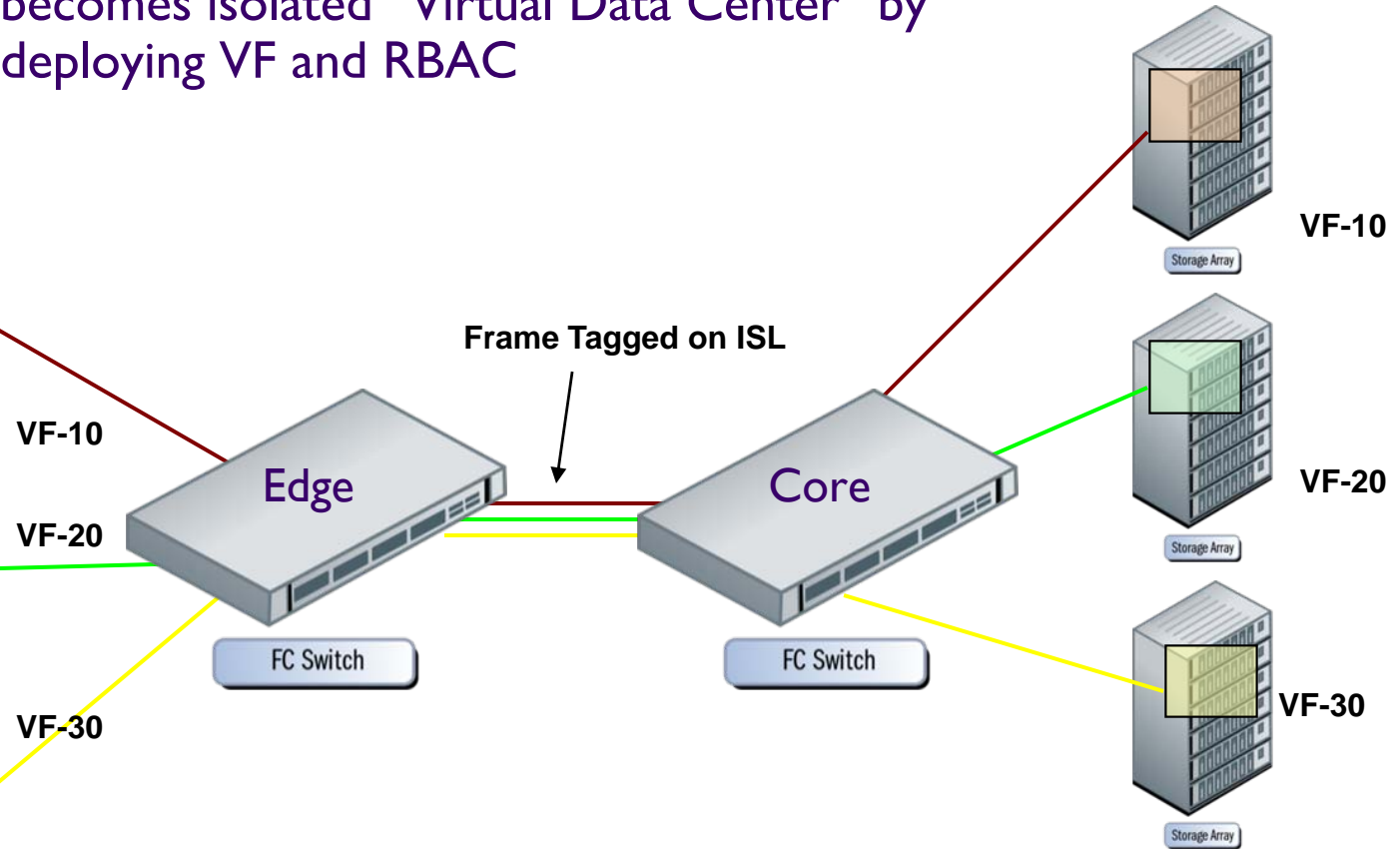
Virtual Data Center
Red Cluster



Virtual Data Center
Green Cluster



Virtual Data Center
Yellow Cluster



- A medium organization can easily deploy over twenty VFs
- Server virtualization allows additional flexibility, leading to a growing number of VF
- Fabric support for a large number of VFs is a key factor for a flexible deployment of VMs
- VF configuration and management must not introduce overhead

Application	Number of VFs (Example)
OLTP	1
Email	1
Accounting	1
HR	1
Filers	2
Back Up	2
FICON	1
Data replication	2
Test	4
Virtual infrastructure	4
Virtual desktop	1

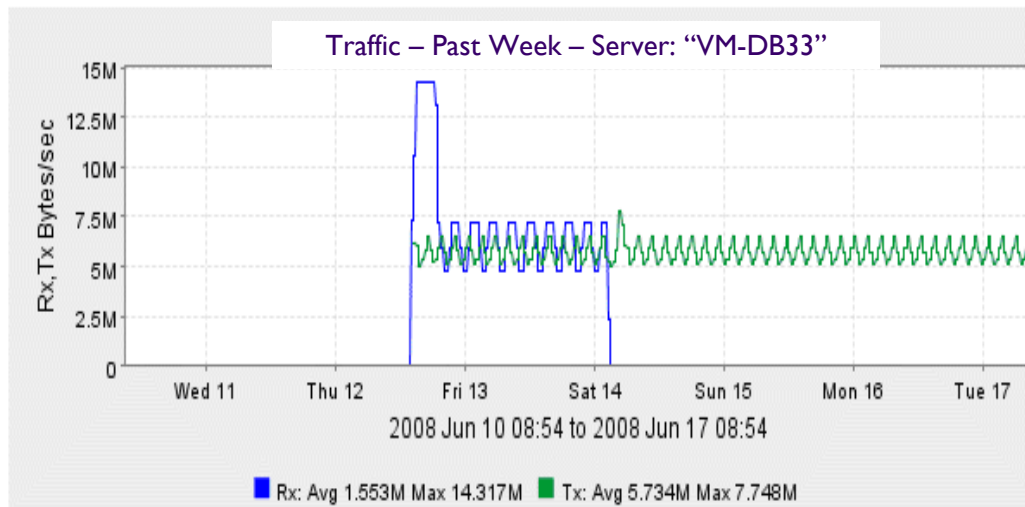
Management Isolation via RBAC (Role Based Access Control)

- The virtual infrastructure management usually provides a very granular RBAC, with the ability of creating multiple custom roles
- An equally flexible RBAC provided by the fabric administration allow, for instance, the creation of multiple custom roles to manage virtual isolated data centers, on the top of the VF infrastructure

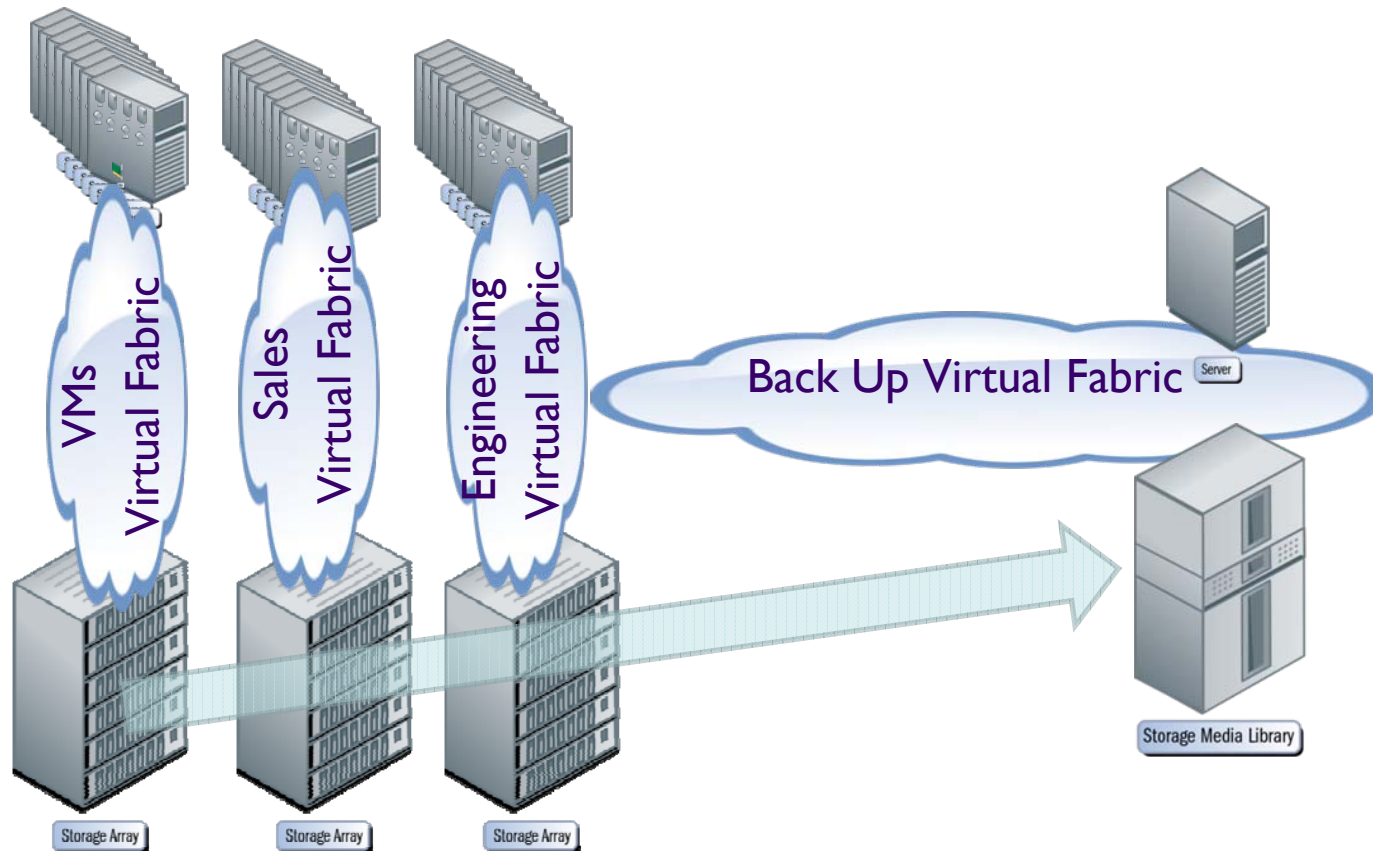
Administrator Privileges			
<i>Administrative Team</i>	<i>Virtual Machines</i>	<i>Storage Network</i>	<i>Storage</i>
Red	Cluster Red	VF-10	Array Red
Green	Cluster Green	VF-20	Array Green
Yellow	Cluster Yellow	VF-30	Array Yellow

Performance Monitoring of an individual Virtual Machine

- The same performance monitoring capabilities available for the physical devices are available for the individual NPIV-enabled virtual machines

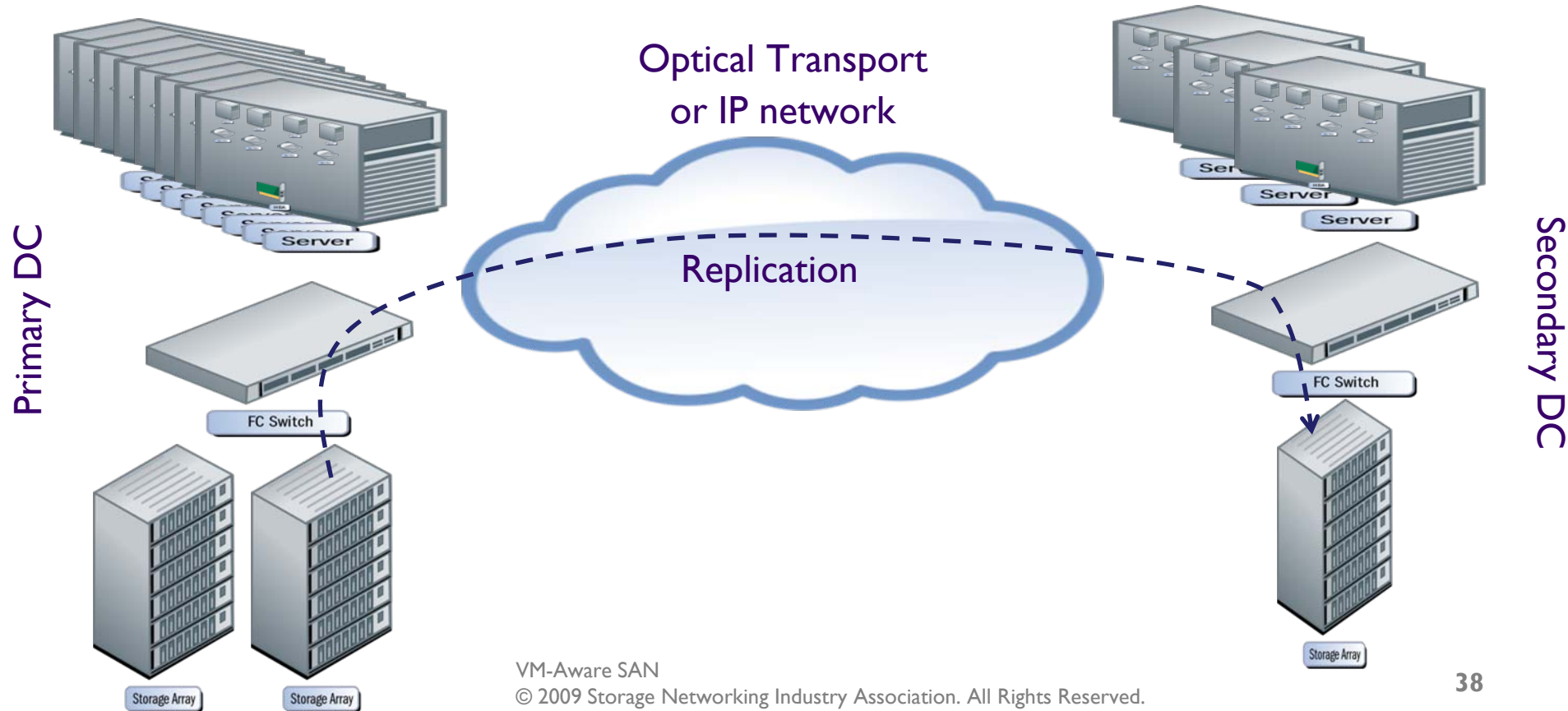


- The back up infrastructure for VMs should make use of shared resource
- Isolation between user groups sharing a common back up infrastructure is possible using multiple Virtual Fabrics and integrated routing



BC/DR for the Virtual Data Center

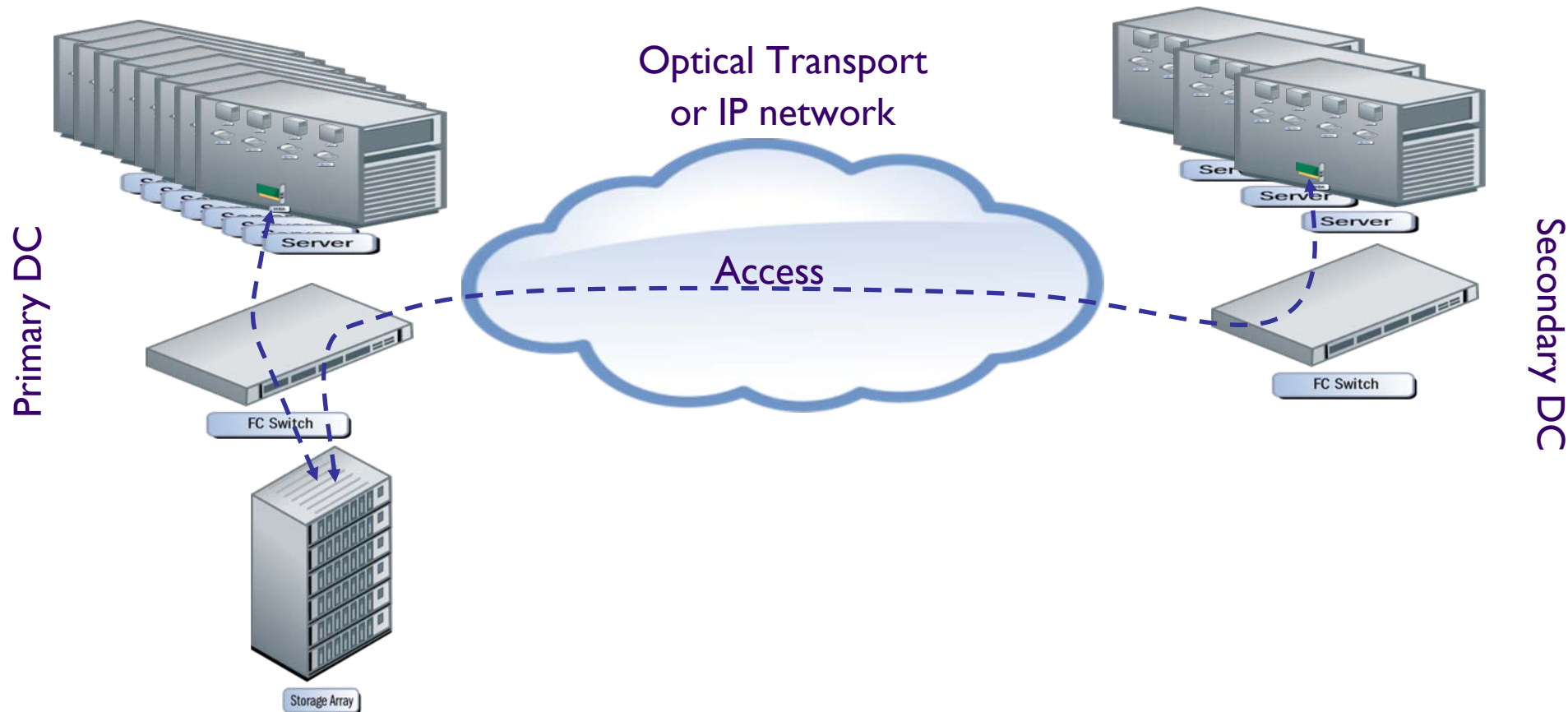
- The VM technology offers tools to automate recovery after a disaster
- A replica of data must be available after the disaster in the secondary DC
- Recovery tools rely upon traditional data replication technologies



- VM mobility is a scheduled procedure, not suitable for disaster recovery
- Use Cases
 - ◆ Disaster avoidance
 - ◆ Workload relocation
 - ◆ DC evacuation
- Data must be accessible in the primary and secondary data center at the same time
- Three modes:
 - ◆ Shared storage
 - ◆ Data mobility followed by VM mobility
 - ◆ Active-active storage

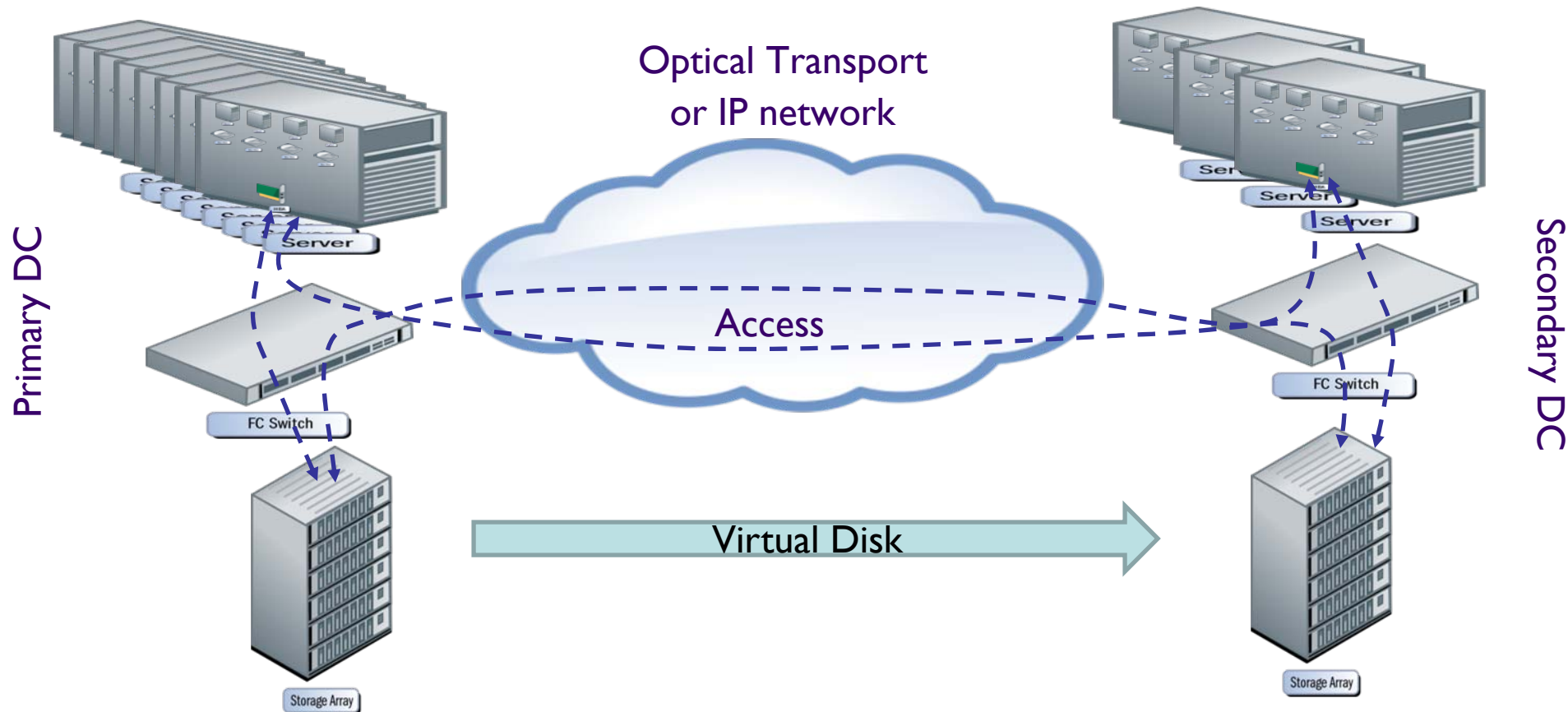
Long Distance VMs Mobility: Shared Storage

Storage resides in the local DC and it is accessed by servers in both the local and remote DC



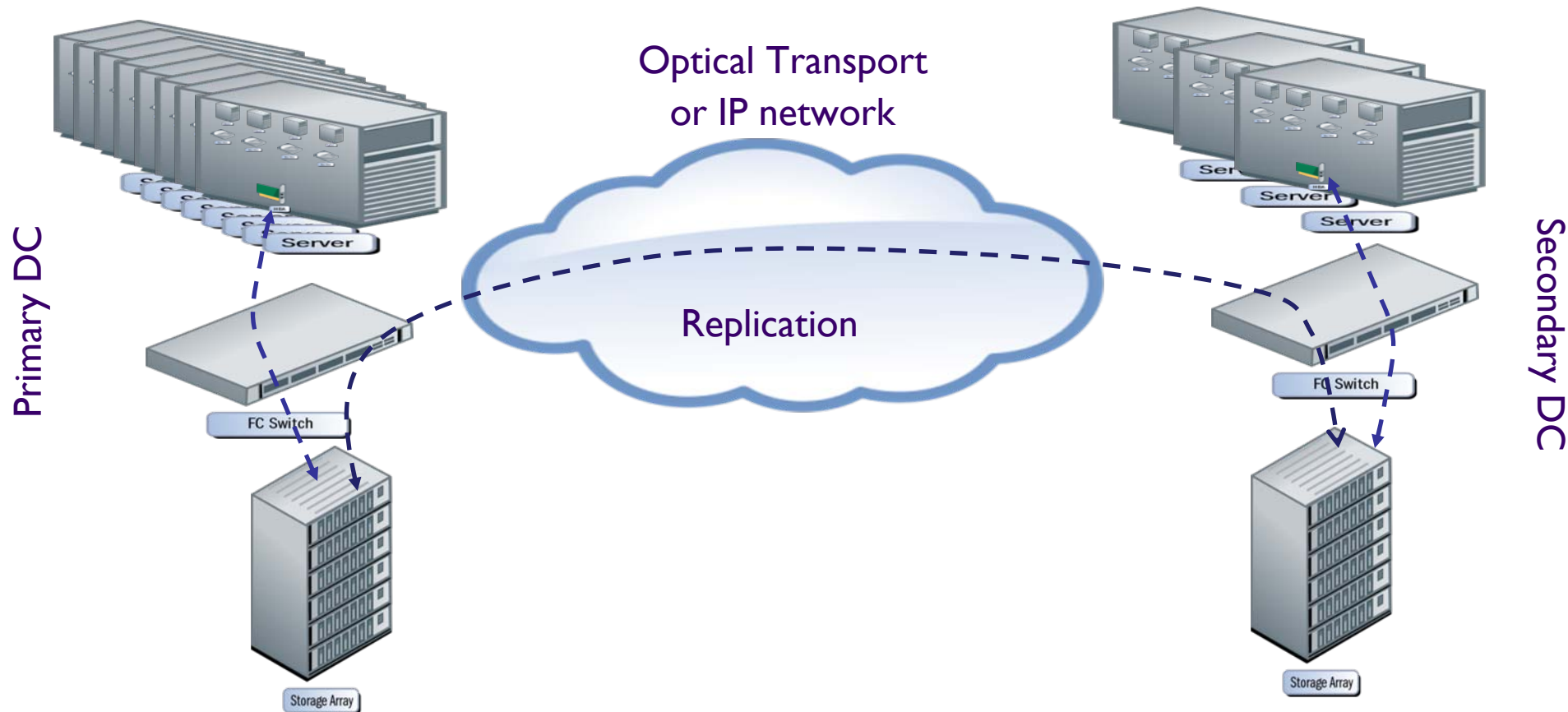
Long Distance VMs Mobility: Data motion followed by VM

Virtual disks are moved to the remote DC, then the VM is sent over.
Both storages are still accessed by servers in the local and remote DC



Long Distance VMs Mobility: Active-active storage

Virtual disks are replicated in real time and available in both DC.
Servers access local storage.



The SAN infrastructure must provide a solid and seamless SAN extension solution to enable replication and mobility

- High performance, not to affect applications
 - ◆ Extended Buffer-to-buffer credits or TCP optimization
 - ◆ Compression
 - ◆ Acceleration services efficient across redundant links
- Secure to meet compliance requirements
 - ◆ Link encryption for extension based on native FC
 - ◆ IPsec for extension based on FCIP
 - ◆ Data center isolation (Virtual Fabrics)

- **Simple deployment and management**
 - ◆ Integrated solution with multiple options (optical, WDM, FCIP)
 - ◆ Management isolation across data centers (RBAC based on Virtual Fabrics)
- **Support of replication over heterogeneous hardware**
 - ◆ With server virtualization the secondary DC is no more requested to have the same servers HW of the primary
 - ◆ The replication technology should allow a degree of flexibility in the choice of the storage arrays
- **Support for long distance VMs mobility**
 - ◆ Long distance VMs mobility enables disaster avoidance, work load relocation, DC evacuation
 - ◆ End to end solution design to meet clients connectivity and storage requirements

- Please send any questions or comments on this presentation to SNIA: tracknetworking@snia.org

**Many thanks to the following individuals
for their contributions to this tutorial.**

- SNIA Education Committee

Fabrizio Corno