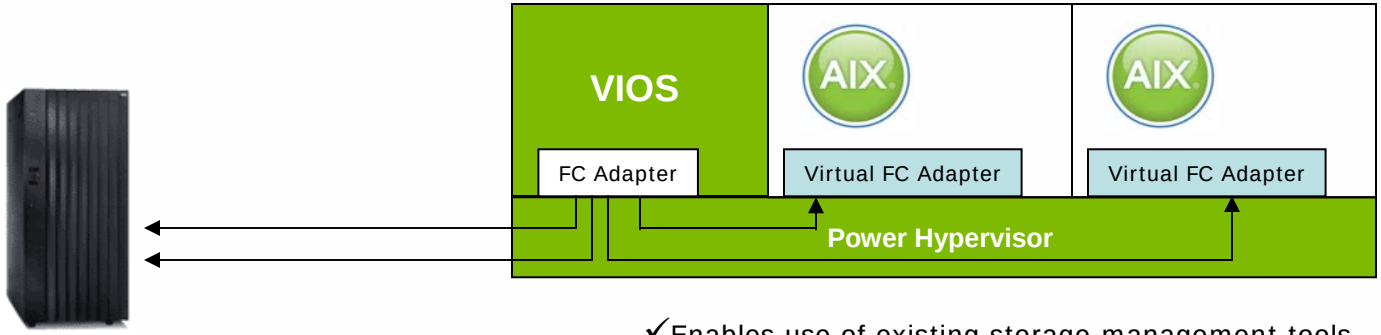
VIO 2.1 and NPIV – An Update

Dec. 10, 2008

Michel Marion
mmarion@ca.ibm.com

NPIV

- **N_Port ID Virtualization (NPIV) provides direct Fibre Channel connections from client partitions to SAN resources , simplifying SAN management**
  - Fibre Channel Host Bus Adapter is owned by VIOS partition
  - Supported with PowerVM Express, Standard, and Enterprise Edition
  - Supports AIX 5.3 and AIX 6.1 partitions
  - Power 520, 550, 560, and 570, with an 8 GB PCIe Fibre Channel Adapter



✓Enables use of existing storage management tools
✓Simplifies storage provisioning (i.e. zoning, LUN masking)
✓Enables access to SAN devices including tape libraries

- **Statement of Direction**
  - *IBM intends to support N_Port ID Virtualization (NPIV) on the POWER6 processor-based Power 595, BladeCenter JS12, and BladeCenter JS22 in 2009.*
  - *IBM intends to support NPIV with IBM i and Linux environments in 2009.*

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

The New Power Equation

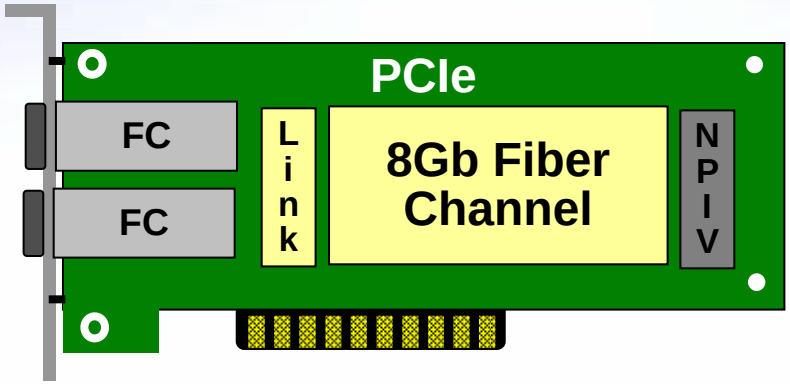
NPIV Overview

- N_Port ID Virtualization (NPIV) is a fibre channel industry standard method for virtualizing a physical fibre channel port.
- NPIV allows one F_Port (switch port) to be associated with multiple N_Port IDs, so a physical fibre channel HBA can be shared across multiple guest operating systems in a virtual environment.
- On POWER, NPIV allows logical LPARs to have dedicated N_Port IDs, giving the OS a unique identity to the SAN, just as if it had a dedicated physical HBA(s).

NPIV specifics

- December 2008 target date
- VIOS V2.1 (PowerVM Express, Standard, and Enterprise)
- client OS support: AIX(5.3 and 6.1), Linux(2009), and IBM i (2009)
- POWER6 only (Emulex), Blade support in 2009 (Qlogic)
- 8Gb PCIe HBA
- unique WWPN generation (allocated in pairs)
- Each virtual FC HBA has a unique and persistent identity
- Compatible with LPM (live partition mobility)
- VIOS can support NPIV and vSCSI simultaneously
- Each physical NPIV capable FC HBA will support 64 virtual ports
- HMC-managed and IVM-managed servers
- header strip/merge
- Direct connect storage (no switch) NOT SUPPORTED

8Gb Fiber Channel



Description:

- Dual Ported
- PCIe x4 Architecture
- Multiple speeds: 2Gb, 4Gb, or 8Gb
- Supported Systems: Power 520 / 550 / 560 / 570 / 575
- Supported OS: AIX and Linux
- NPIV enabled
- Feature Code: 5735

Technical Specs....

- **Cable support: Multimode fiber optic cables with Short-Wave lasers that adhere to the following specifications:**
  - –OM3 - Multimode 50/125 micron fiber, 2000 MHz*km bandwidth
  - –OM2 - Multimode 50/125 micron fiber, 500 MHz*km bandwidth

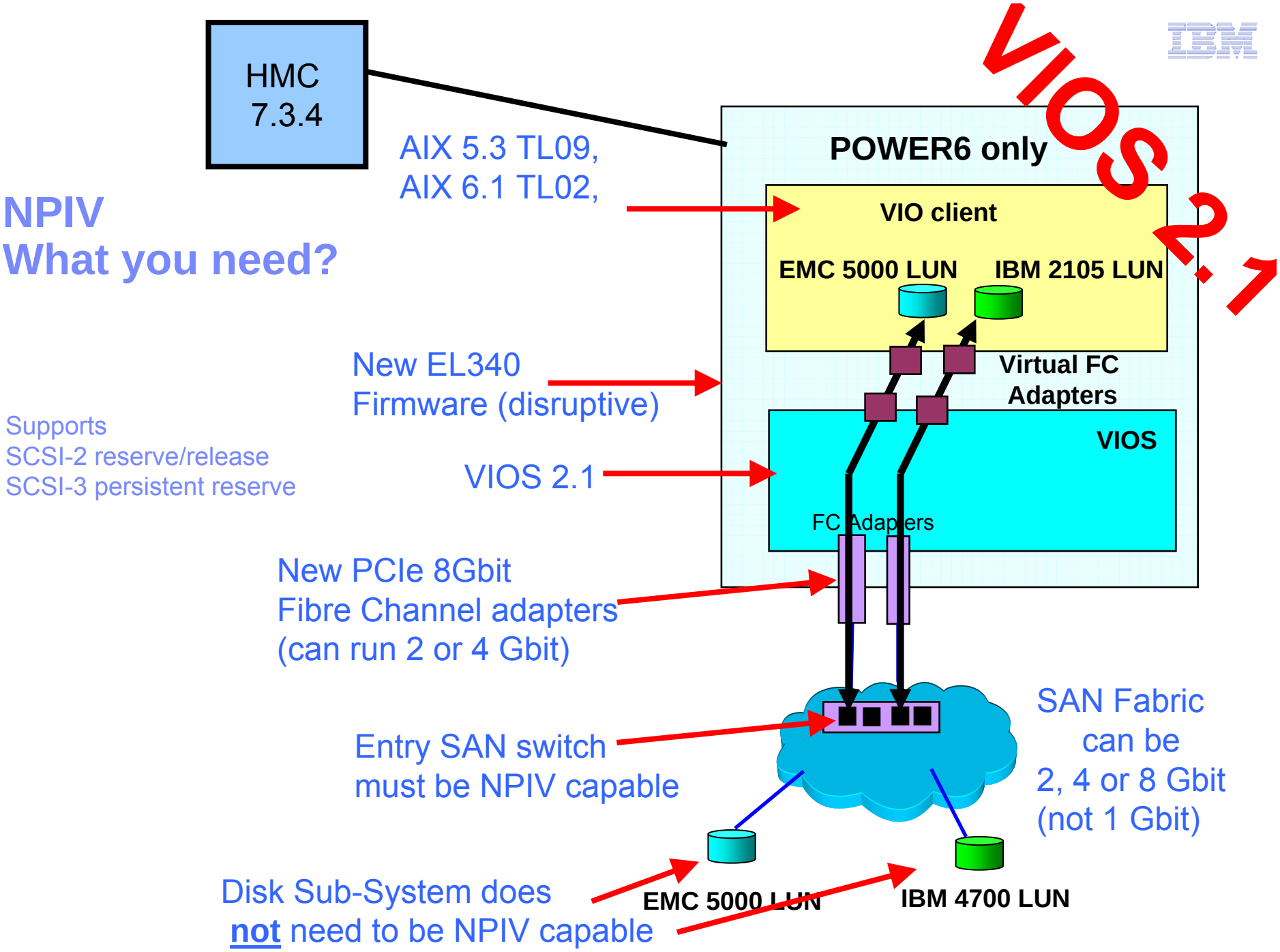
- **OS level required:**
  - –IBM i 6.1.
  - –AIX 5L for POWER version 5.3 with the 5300-09 Technology Level
  - –AIX Version 6.1 with the 6100-02 Technology Level
  - –SUSE Linux Enterprise Server 10 SP2 for POWER Systems or later.
  - –Red Hat Enterprise LInux for POWER version 4.7 or later.
  - –Red Hat Enterprise Linux for POWER version 5.2 or later.

NPIV Requirements – take note!!!

Software	**VIOS 2.1 Fixpack 20.1** **AIX 5.3 TL9 SP2** **AIX 6.1 TL2 SP2**
HMC	**7.3.4**
Firmware	**EL_340_036 (L = low end; 520, 550)** **EM_340_036 (M = midrange; 560, 570)** **(disruptive)**
Hardware	**FC 5735 PCIe 8Gb DP fibre channel adapter** **NPIV capable switches in first layer of SAN fabric.** **Power 6 520, 550, 570** *** 140 MB memory per virtual client adapter**

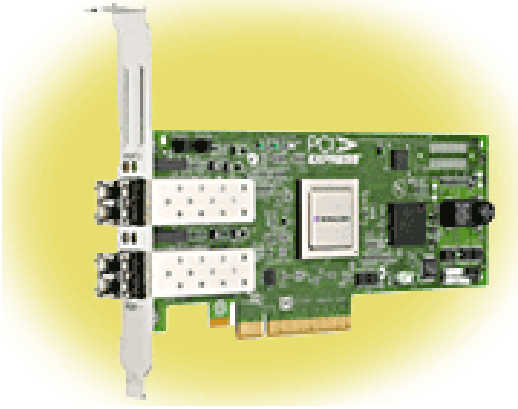
NPIV Support GA Date: Dec. 19, 2008

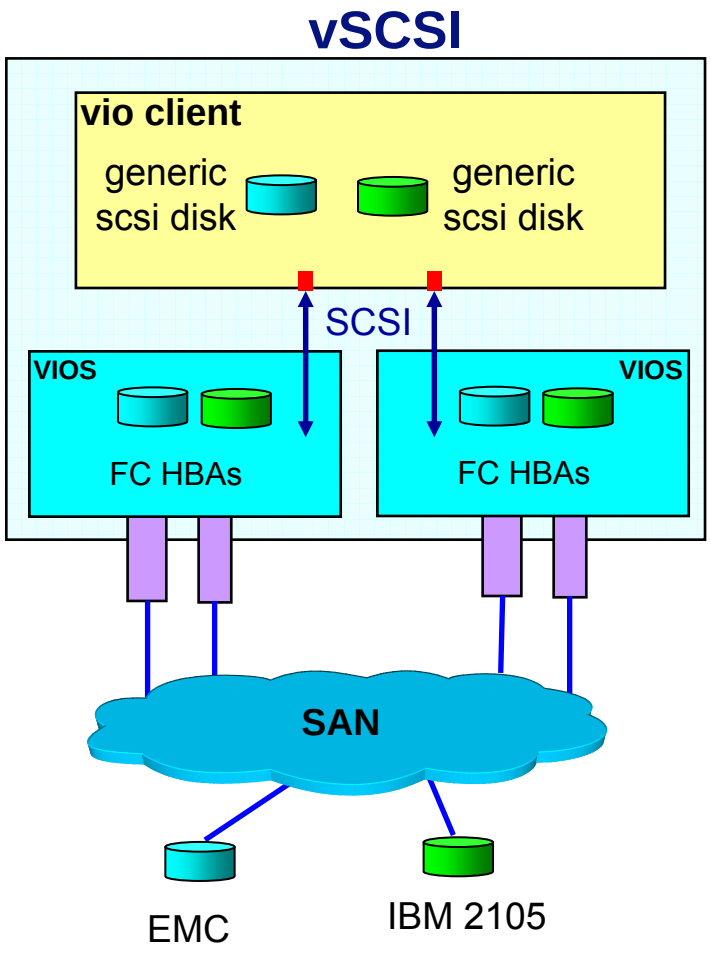
NPIV
What you need?

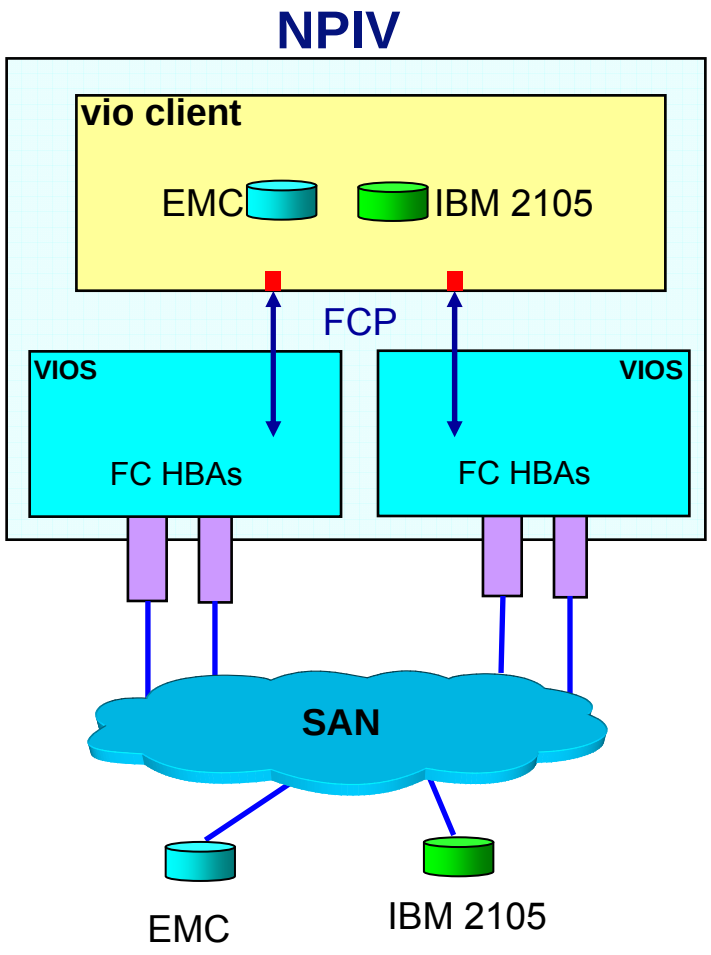


#5735 PCIe 8Gb Fibre Channel Adapter

- Supported on 520, 550, 560, 570, 575
- Dual port adapter - each port provides single initiator
  - Automatically adjusts to SAN fabric 8 Gbps, 4 Gbps, 2 Gbps
  - LED on card indicates link speed
  - PCIe not available on p595
- Ports have LC type connectors
  - Cables are the responsibility of the customer.
  - Use multimode fibre optic cables with short-wave lasers:
    - OM3 - multimode 50/125 micron fibre, 2000 MHz*km bandwidth
      - 2Gb (.5 – 500m) 4Gb (.5 – 380m) 8Gb (,5 – 150m)
    - OM2 - multimode 50/125 micron fibre, 500 MHz*km bandwidth
      - 2Gb (.5 – 150m) 4Gb (.5 – 70m) 8Gb (,5 – 21m)
    - OM1 - multimode 62.5/125 micron fibre, 200 MHz*km bandwidth
      - 2Gb (.5 – 300m) 4Gb (.5 – 150m) 8Gb (,5 – 50m)







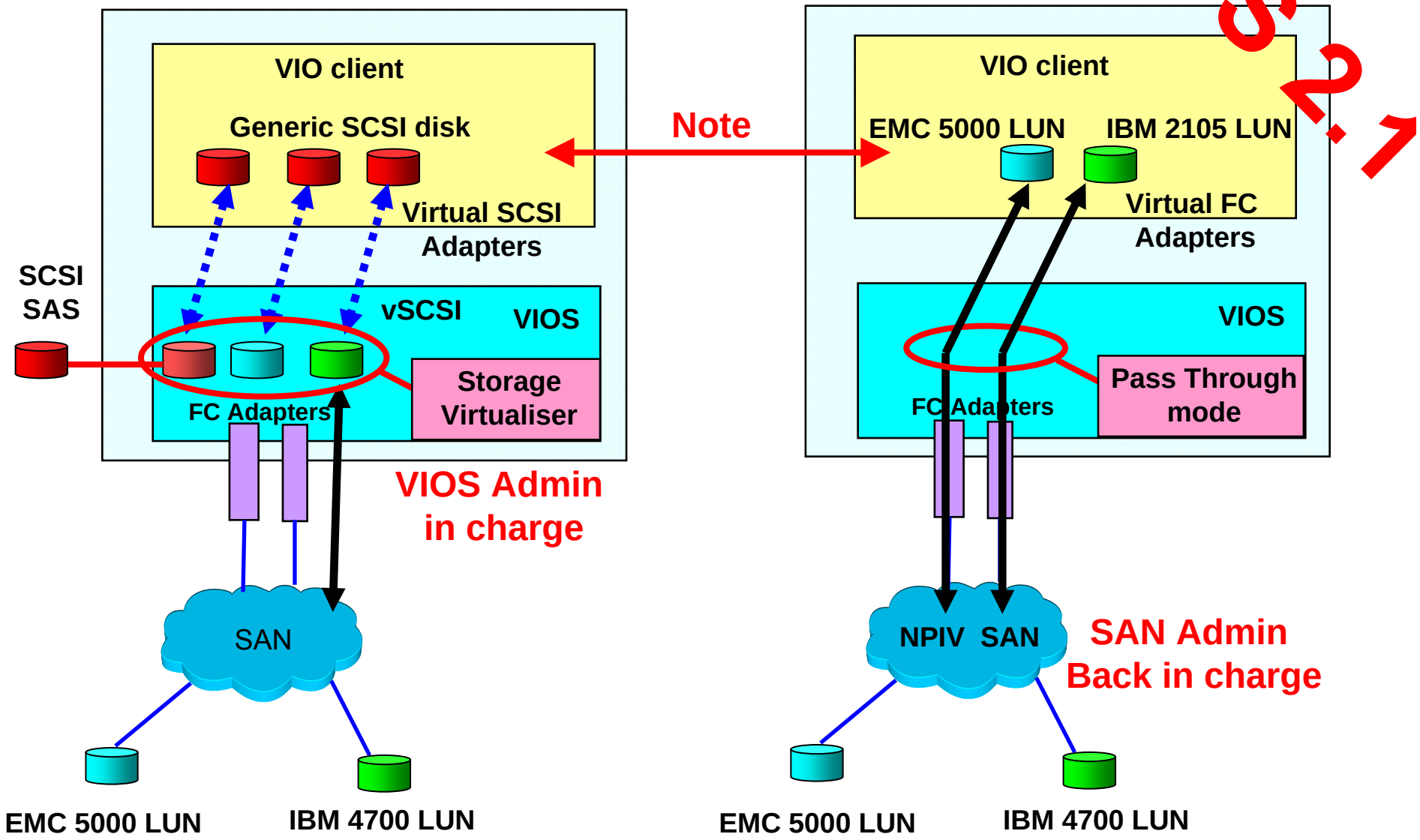
The vSCSI model for sharing storage resources is storage virtualizer. Heterogeneous storage is pooled by the VIOS into a homogeneous pool of block storage and then allocated to client LPARs in the form of generic SCSI LUNs. The VIOS performs SCSI emulation and acts as the SCSI Target.

With NPIV, the VIOS's role is fundamentally different. The VIOS facilitates adapter sharing only, there is no device level abstraction or emulation. Rather than a storage virtualizer, the VIOS serving NPIV is a passthru, providing an FCP connection from the client to the SAN.

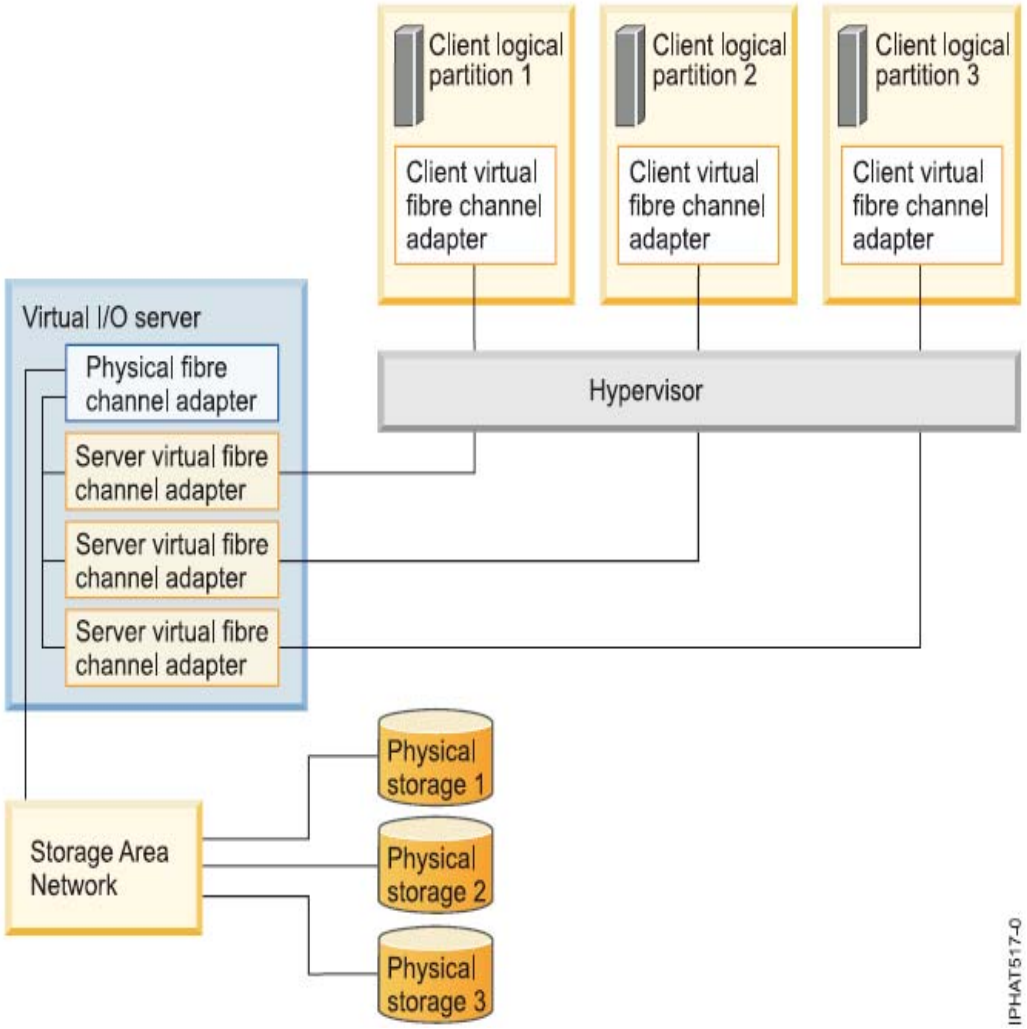
Storage Virtualisation

With NPIV

VIOS 2.1



NPIV – The basics



- physical FC adapter connects to three virtual FC adapters; all three virtual adapters connect to same physical port
- each virtual FC adapter on the VIOS connects to one client virtual FC adapter
- each virtual client FC adapter gets a unique pair of WWPNs; one to log into SAN, the other is for LPM
- clients can now discover & manage their physical storage on the SAN
- VIOS can't access or emulate storage; just provides clients with connection to adapter
- **'vfcmap'** on VIOS to connect virtual adapters on VIOS to physical adapters

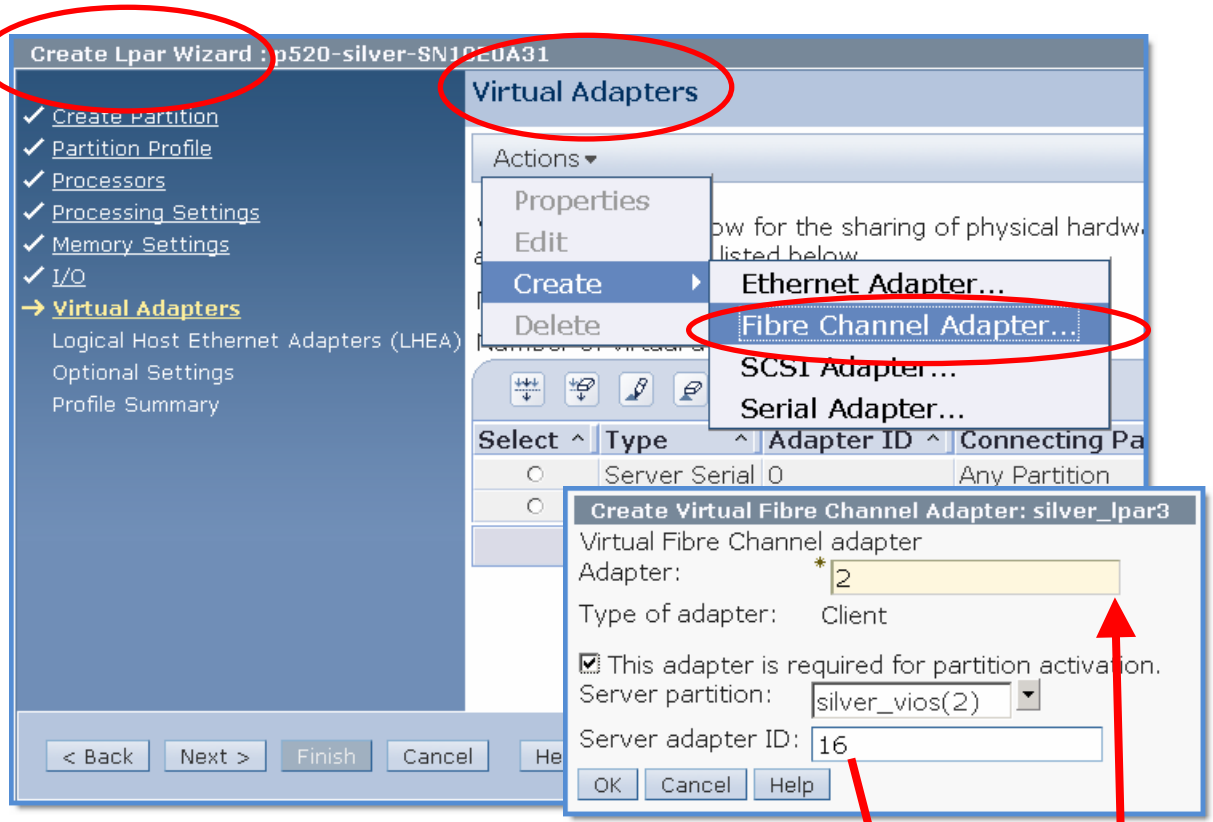
The Basics …

- **One to one relationship between VIOS vFC adapter and client vFC adapter**
- **Use SAN tools to zone and mask LUNs that include WWPNs that are assigned to vFC adapters on client partitions the same as if they were assigned to physical ports**
- **To avoid SPOF, do not connect 2 vFC adapters from the same client to the same physical adapter**
- **HMC generates WWPNs based on prefix in vpd of managed system; includes 64K (32K pairs of WWPNs)**
- **If you run out have to purchase activation codes (i.e.don't do DLPAR operations)**

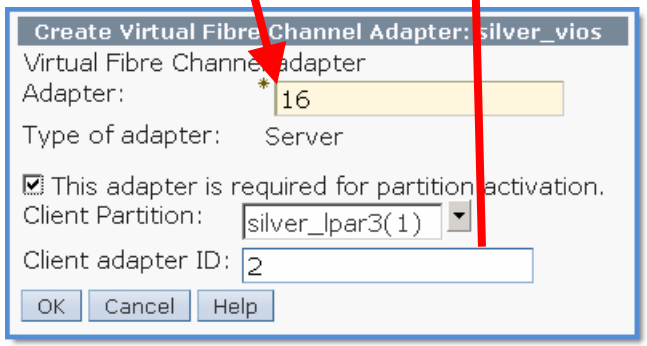
What you do?

1. Check SAN switch NPIV compatibility
   - Upgrade firmware if necessary
   - Run any commands on the switch needed to make the SAN switch ports NPIV capable.

NPIV
What you do?

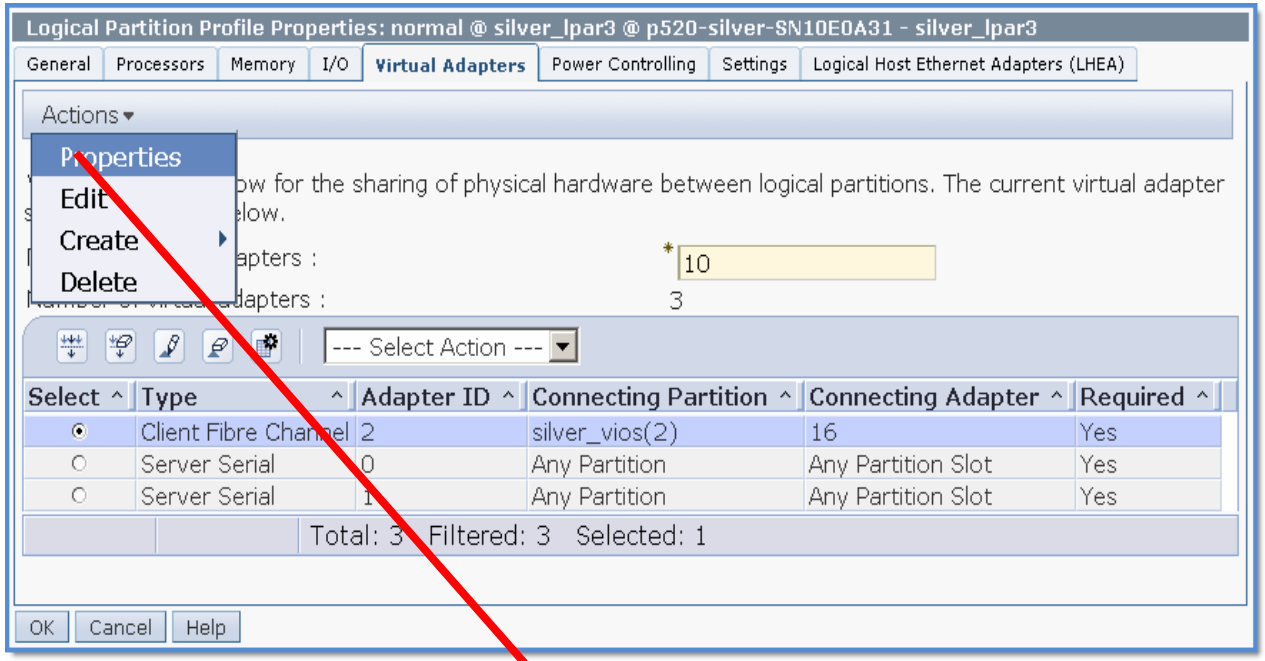


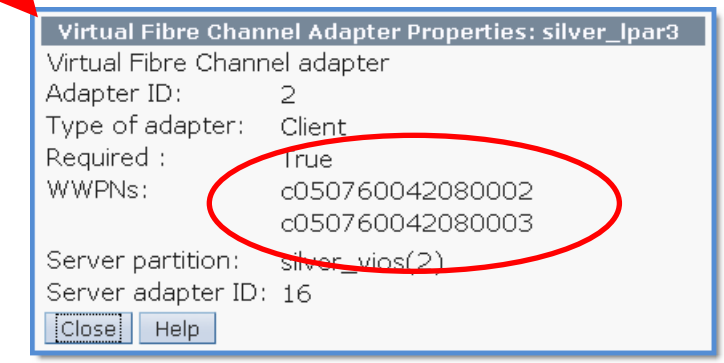
2. ON HMC (7.3.4) Configure virtual FC adapters on VIOS and client (just like vSCSI); then run **cfgdev** on VIOS to configure the vFC adapters on VIOS



Virtual I/O Server

NPIV
What you do?





3. Once Created:
   LPAR Config
   →Manage Profiles
   →Edit click FC Adapter
   → Properties
   and the WWPNs are available

NPIV – What you do?

- **`lsdev -dev vfchost*`**

  lists all available virtual FC adapters on VIOS

```
$ lsdev -dev vfchost*
name             status      description
vfchost0         Available   Virtual FC Server Adapter
```

- **`lsdev -dev fcs*`**

  lists all available physical FC adapters in VIOS

```
$ lsdev -dev fcs*
name             status      description
fcs0             Available   4Gb FC PCI Express Adapter (df1000fe)
fcs1             Available   4Gb FC PCI Express Adapter (df1000fe)
fcs2             Available   8Gb PCI Express Dual Port FC Adapter (df1000f114108a03)
fcs3             Available   8Gb PCI Express Dual Port FC Adapter (df1000f114108a03)
```

NPIV – What you do?

- **`lsnports`**

  check FC adapter and SAN switch NPIV readiness

```
$ lsnports
name             physloc                        fabric tports aports swwpns  awwpns
fcs3             U789D.001.DQDYKYW-P1-C6-T2          1     64     63   2048    2046
```

- **`vfcmap –vadapter vfchost0 –fcp fcs3`**

  maps virtual adapter vfchost0 to physical adapter fcs3

```
$ vfcmap -vadapter vfchost0 -fcp fcs3
vfchost0 changed
```

NPIV – What you do?

- **`lsmap -npiv -all`**
  to list mappings

```
$ lsmap -npiv -all
Name           Physloc                            ClntID ClntName       ClntOS
============== ================================== ====== ============== =======
vfchost0       U9117.MMA.101F170-V1-C31                3 AIX61          AIX

Status:LOGGED_IN
FC name:fcs3                    FC loc code:U789D.001.DQDYKYW-P1-C6-T2
Ports logged in:1
Flags:a<LOGGED_IN,STRIP_MERGE>

VFC client name:fcs0            VFC client DRC:U9117.MMA.101F170-V3-C31-T1
```

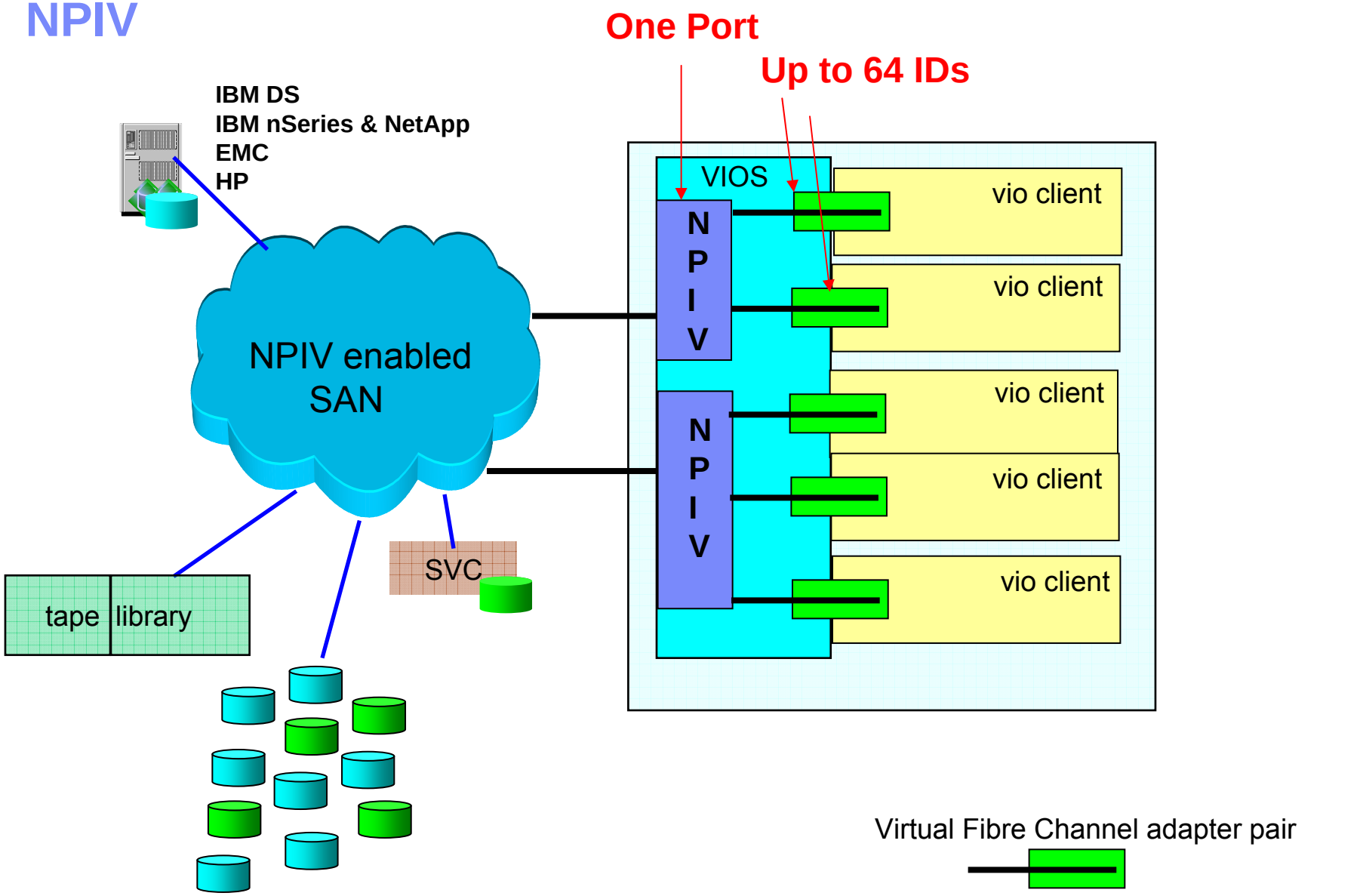
NPIV - What you do?

```
$ ioslevel
2.1.0.0
$ lsdev | grep FC
fcs0             Available   FC Adapter
fscsi0           Available   FC SCSI I/O Controller Protocol Device
vfchost0         Available   Virtual FC Server Adapter
$ vfcmap -vadapter vfchost0 -fcp fcs0
vfchost0 changed
$
```

4. VIOS connect the virtual FC adapter to the physical FC adapter
   - With **vfcmap**
   - **lsmap –all –npiv**
   - **lsnports → shows physical ports supporting NPIV**
5. SAN Zoning
   - To allow the LPAR access to the LUN via the new WWPN
   - Allow both WWPN and on any Partition Mobility target.

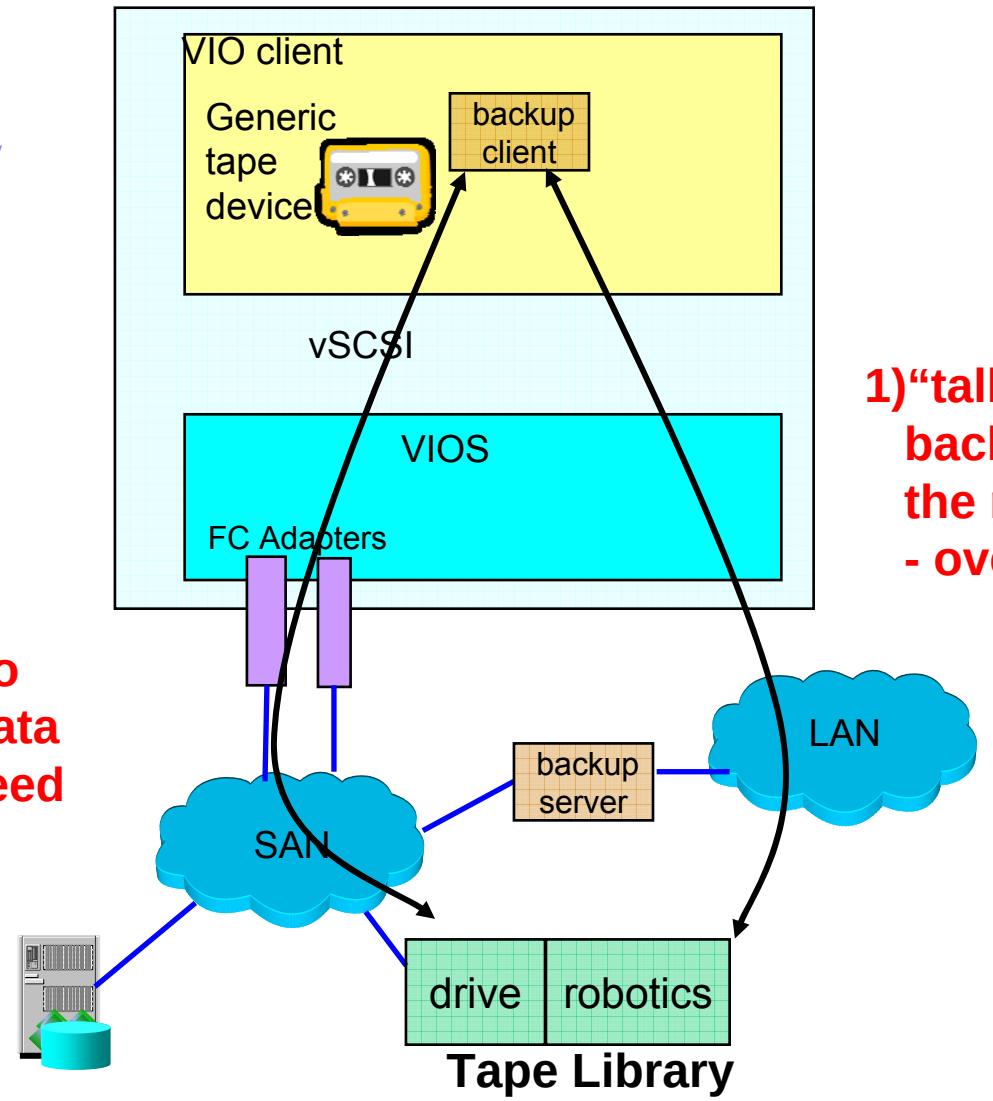
Advanced NPIV Side Effects

NPIV



Max NPIV of 255 but 64 max recommended

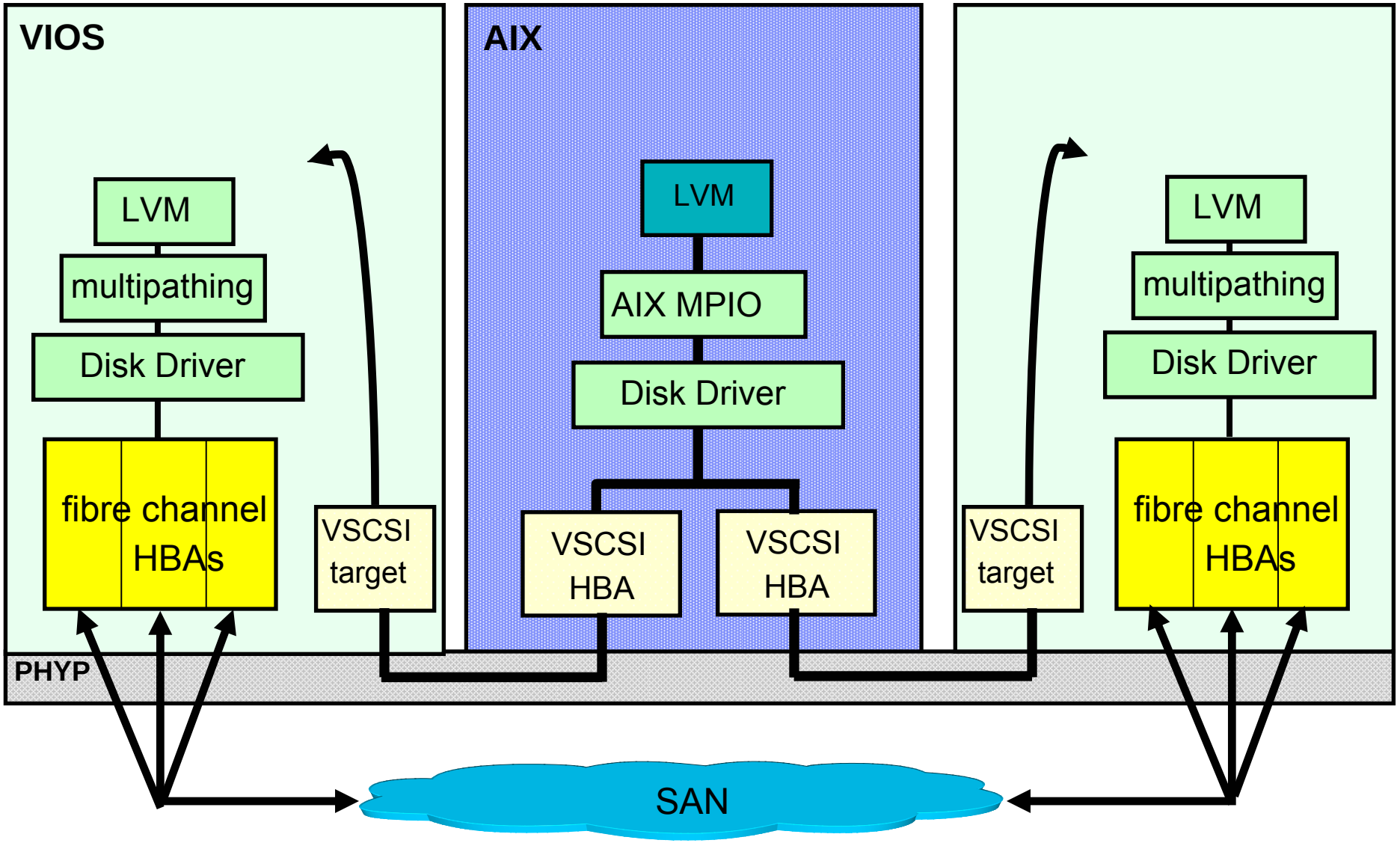
Virtual Tape With NPIV Conceptually



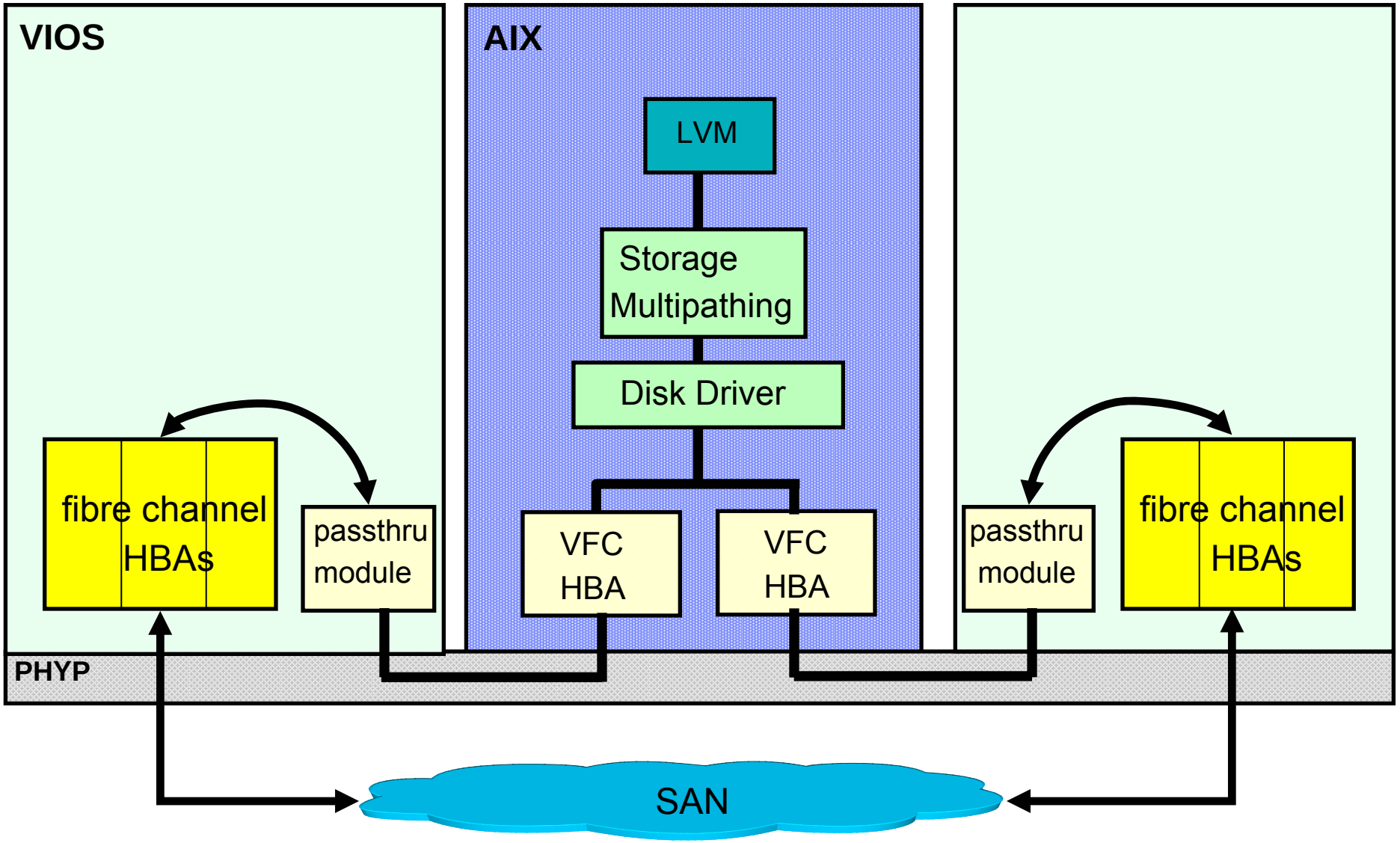
1)"talk" to robot or backup server to get the right tape loaded - over the network

2) Use SAN to save the data at high speed

Previous Virtual SCSI to Fibre Channel

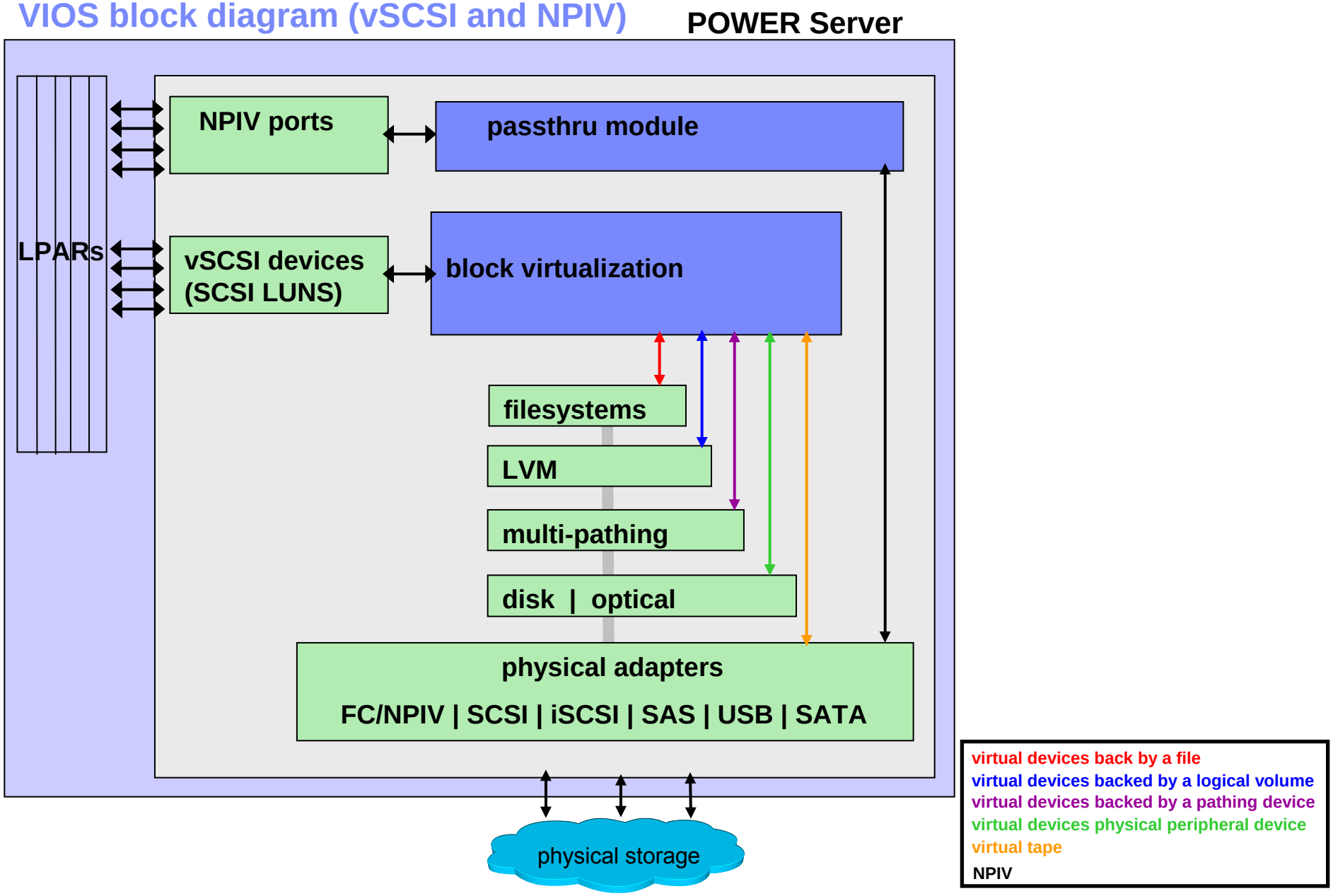


New NPIV pure Fibre Channel

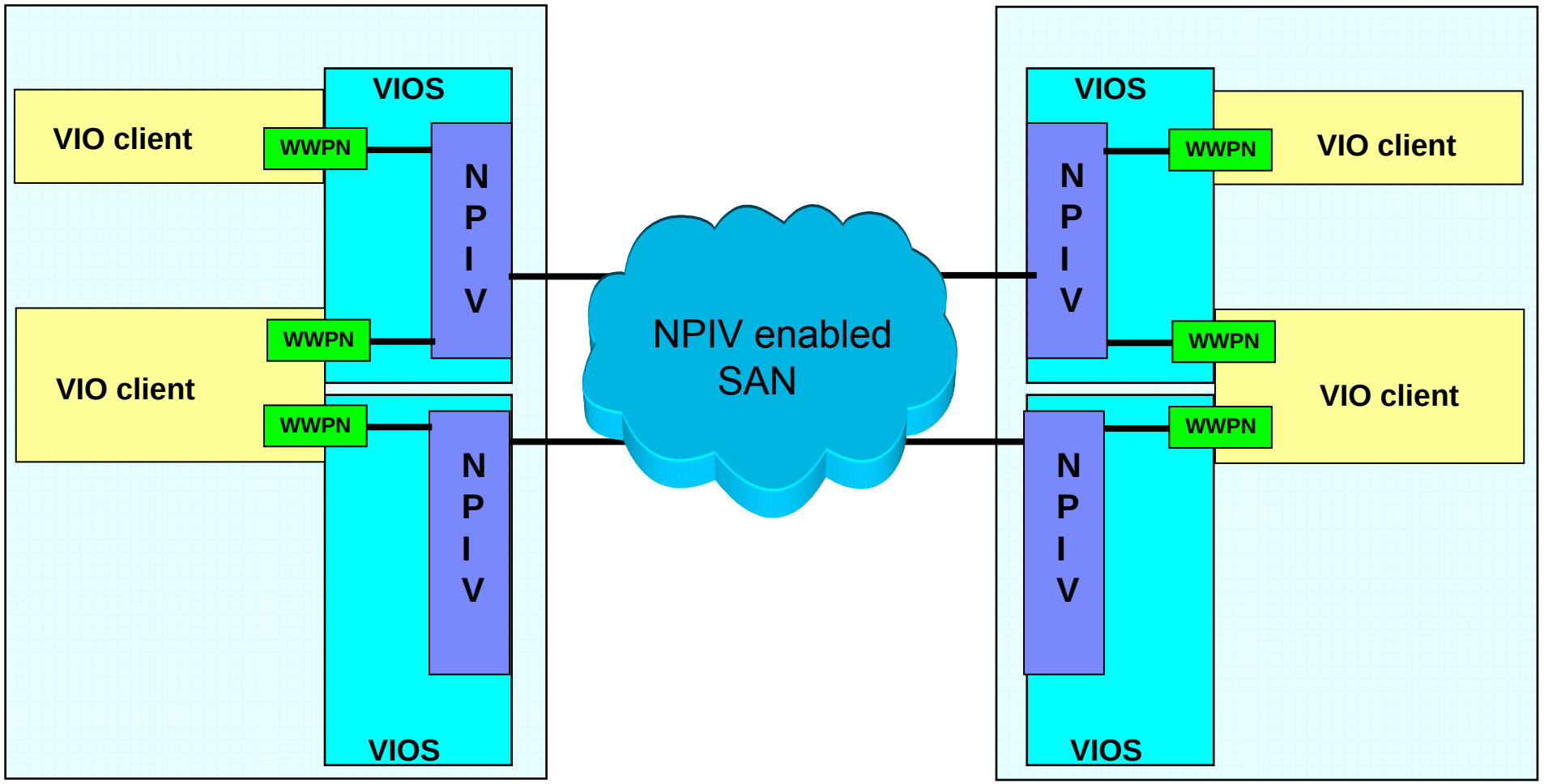


No VIOS side multipath, more client setup per LPAR but Thinner Stack

VIOS block diagram (vSCSI and NPIV)



Live Partition Mobility with NPIV

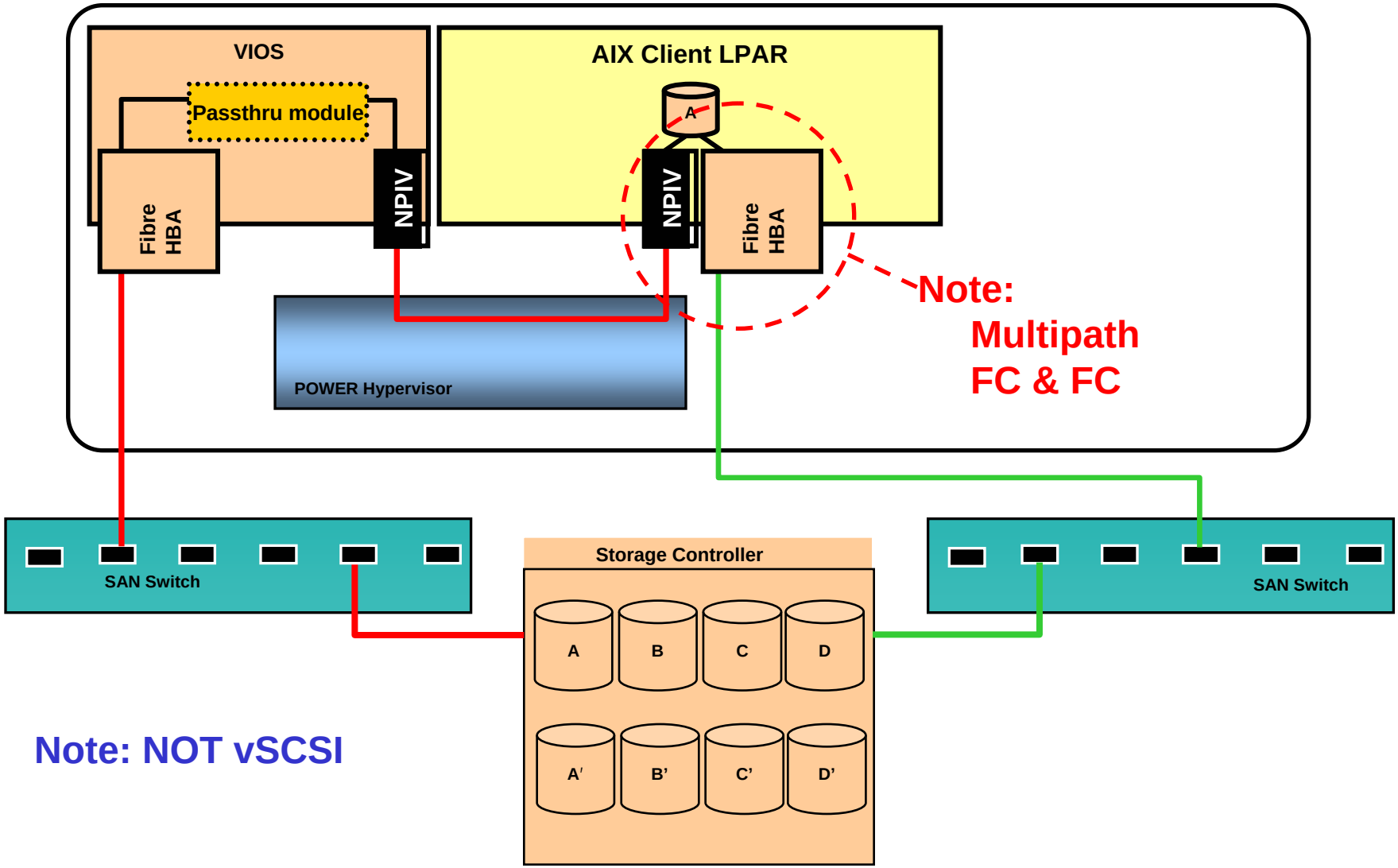


1. WWPNs are allocated in pairs and used in PM
2. Target VIOS uses 2nd WWPN to check access
3. Target LPAR uses 2nd WWPN for disk access

Live Partition Mobility

- **Client Partition with more that one VFC adapter should have the corresponding VFC server adapter mapped to different Fibre Channel Adapter Port**
- **Zoning on source and destination switch for access to the same targets**
- **Destination VIOS partition/s should have Fibre Channel Adapters to have same configuration as on the source**
- **Once moved over uses second WWPN; if partition comes back it uses original WWPN**

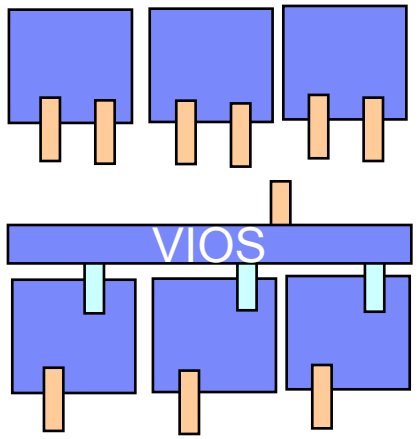
Heterogeneous Multipathing



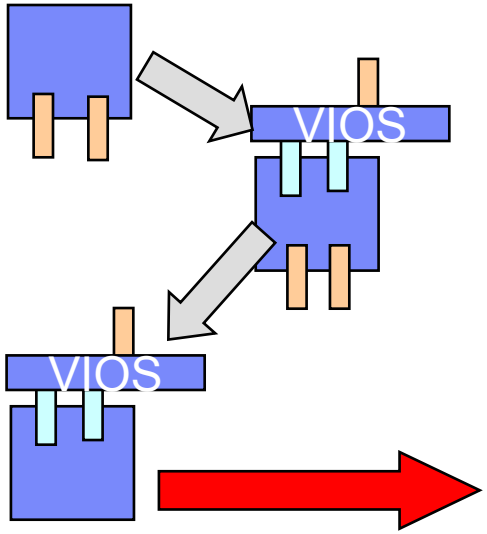
Note: NOT vSCSI

Heterogeneous Multipathing - 2 uses

- 1) Reduced Adapter for small LPARs
  - Small dedicated adapter LPAR
  - 2 adapters for redundancy but overkill !!
  - Use 1 direct & 1 backup via “cheap” VIOS



- 2) Partition Mobility for Dedicated Adapters
  - Temporarily add virtual adapters
  - Move I/O to virtual adapters
  - Remove physical adapters
  - Now do Partition Mobility
  - Move back to physical adapters



Dynamic Physical/Virtual Multipathing (2 of 2)

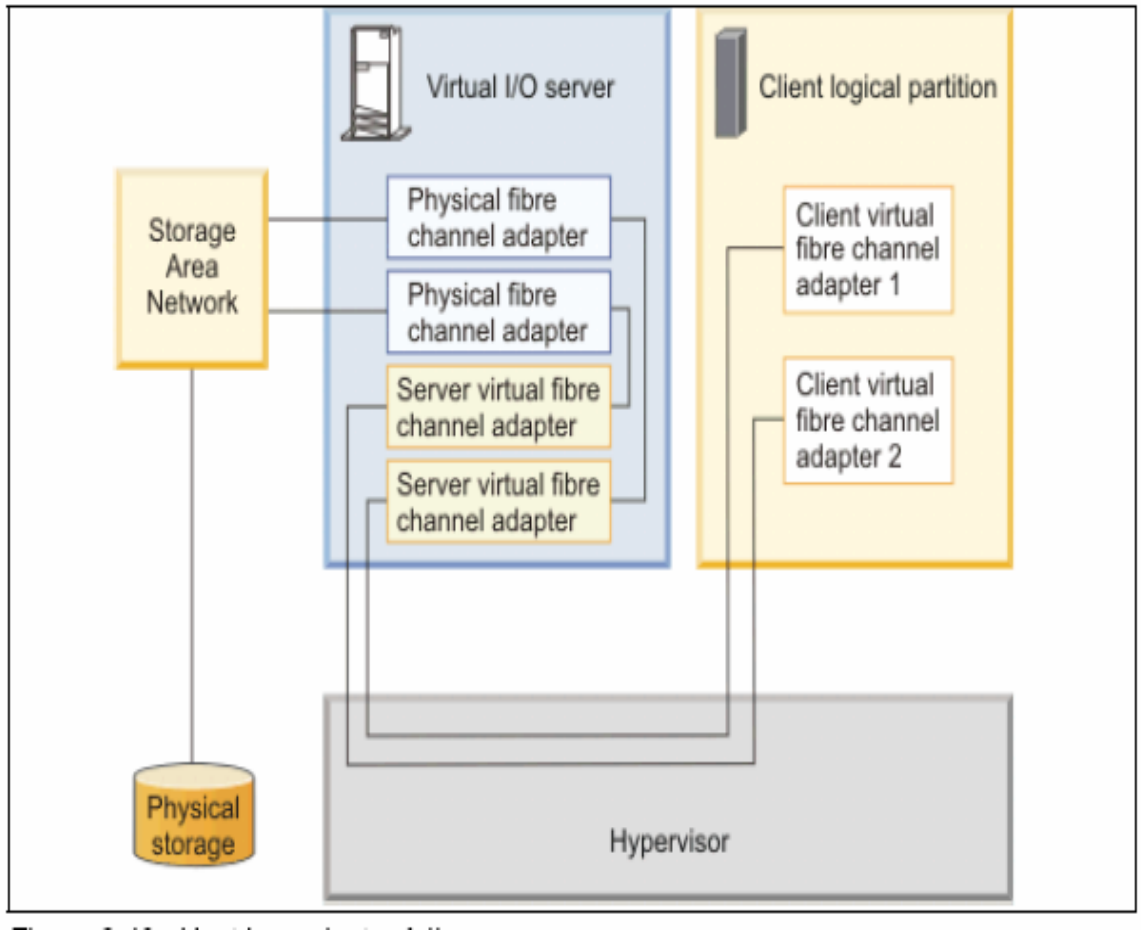
- Physical-to-Virtual Failover or Vise Versa
- LPM Enabler
- Full MPIO Capabilities
  - Round Robin
  - Load Balancing
  - Path Failover

NPIV benefits to Client

- NPIV allows storage administrators to used existing tools and techniques for storage management
- solutions such as SAN managers, Copy Services, backup / restore, should work right out of the box
- storage provisioning / ease-of-use
- Zoning / LUN masking to client
- physical <-> virtual device compatibility
- tape libraries
- SCSI-2 Reserve/Release and SCSI3 Persistent Reserve
- clustered/distributed solutions
- Load balancing (active/active)
- Storage, multipathing, apps, monitoring…..

Redundant Configurations

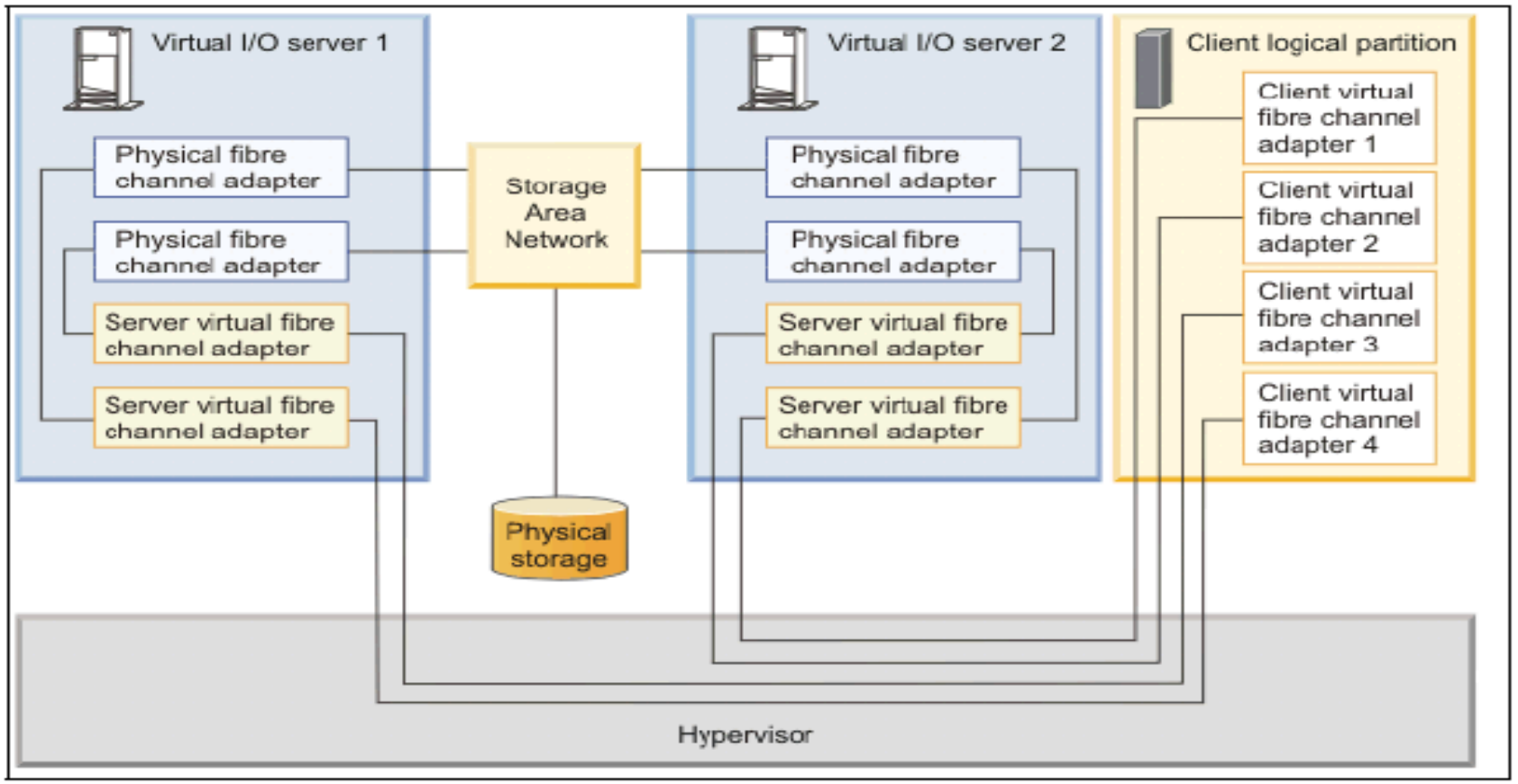
- **HBA Failover**



- each client vFC adapter connected to server vFC adapter mapped to physical port on different HBA
- client uses MPIO; if physical HBA is used, path exists through the other one

Redundant Configurations

- **HBA and VIOS Failover**



➢ same as above but with VIO redundancy

Commands …

VIOS Server Commands

- **vfcmap – binding the VFC Server to the Fibre Channel Port**
  - **vfcmap -help**

    Usage: vfcmap -vadapter VFCServerAdapter -fcp FCPName

    Maps the Virtual Fibre Channel Adapter to the physical Fibre Channel Port

    -vadapter Specifies the virtual server adapter.

    -fcp Specifies the physical Fibre Channel Port

    vfcmap –vadapter vfchost1 –fcp fcs0

VIOS Server Commands

- lsnports

$ lsnports –help

Usage: lsnports [-fmt delimiter] [-field FieldName]

Lists available NPIV capable ports and related information.

-fmt Divides output by a user-specified delimiter.

-field Specifies a list of fields to be displayed.

$ lsnports

name	physloc	fabric	tports	aports	swwpns	awwpns
fcs0	U789D.001.DQDMLMP-P1-C1-T1	1	64	64	2048	2047
fcs1	U789D.001.DQDMLMP-P1-C1-T2	1	64	64	2048	2047

VFC Server Commands

- lsmap – enhanced to list VFC binding information

```
$ lsmap -all -npiv
Name             Physloc                            ClntID       ClntName        ClntOS
============= ================================== ====== ============== =======
vfchost2         U8203.E4A.10D4461-V2-C13               15      AIX_vFC_PM2_53Q    AIX

Status:LOGGED_IN
FC name:fcs0                    FC loc code:U789C.001.DQD1760-P1-C2-T1
Ports logged in:2
Flags:a<LOGGED_IN,STRIP_MERGE>
VFC client name:fcs1            VFC client DRC:U8203.E4A.10D4461-V15-C4-T1
```

Virtual FC Adapter Information (1 of 4)

```
# lsdev -Cc adapter -s vdevice -t IBM,vfc-client
fcs0 Available C3-T1 Virtual Fibre Channel Client Adapter
fcs1 Available C4-T1 Virtual Fibre Channel Client Adapter
```

Virtual FC Adapter Information (2 of 4)

```
# lsattr -El fcs0
intr_priority 3         N/A False
lg_term_dma   0x800000 N/A True
max_xfer_size 0x100000 N/A True
num_cmd_elems 200       N/A True
sw_fc_class   2         N/A True
```

Virtual FC Adapter Information (3 of 4)

```
# lscfg -vpsl fcs0
  fcs0             U9117.MMA.107086C-V15-C3-T1
        Virtual Fibre Channel Client Adapter

        Network Address.............C05076003D40009C
        ROS Level and ID............
        Device Specific.(Z0)........
        Device Specific.(Z1)........
        Device Specific.(Z2)........
        Device Specific.(Z3)........
        Device Specific.(Z4)........
        Device Specific.(Z5)........
        Device Specific.(Z6)........
        Device Specific.(Z7)........
        Device Specific.(Z8)........C05076003D40009C
        Device Specific.(Z9)........
        Hardware Location Code......U9117.MMA.107086C-V15-C3-T1
```

Output omitted for brevity

Virtual FC Adapter Information (4 of 4)

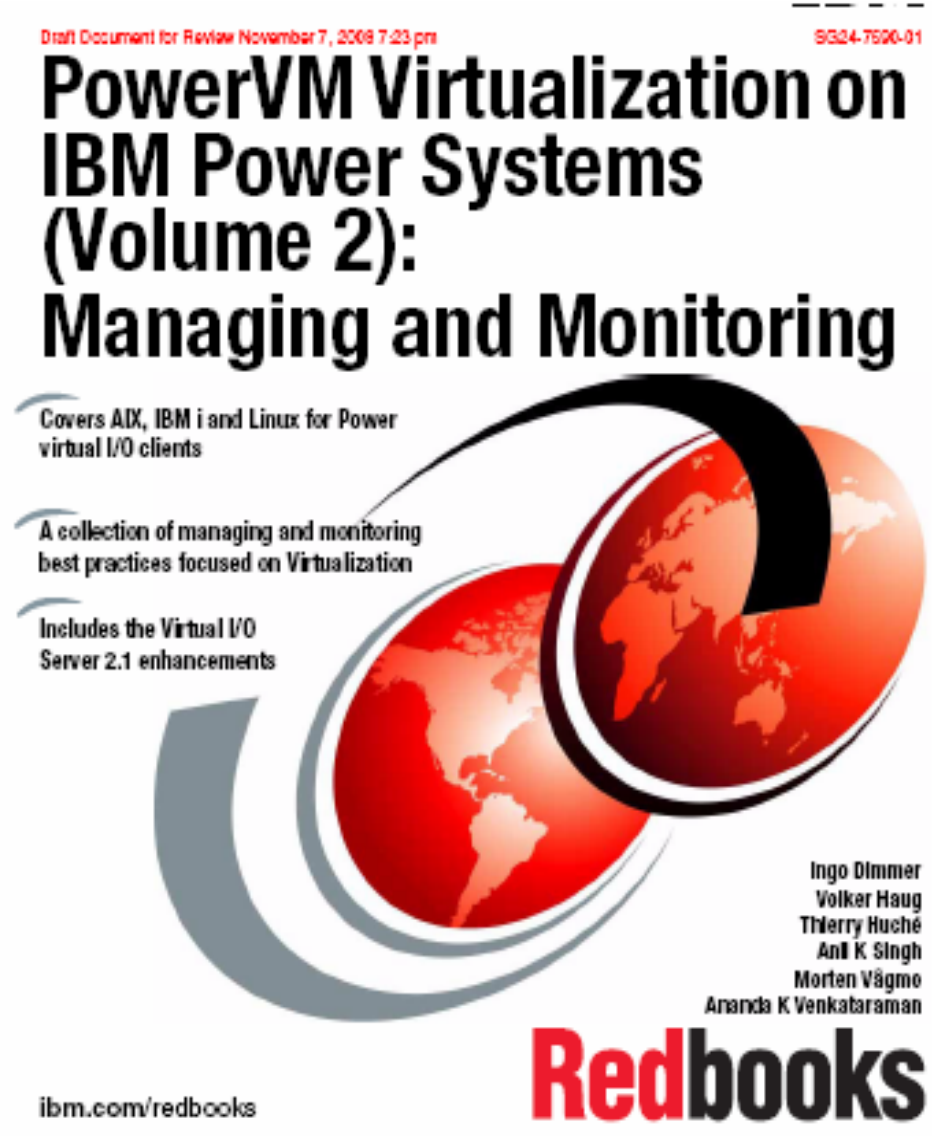
```
# fcstat fcs0
FIBRE CHANNEL STATISTICS REPORT: fcs0

Device Type: FC Adapter (adapter/vdevice/IBM,vfc-client)
Serial Number: UNKNOWN
Option ROM Version: UNKNOWN
Firmware Version: UNKNOWN
World Wide Node Name: 0xC05076003D40009D
World Wide Port Name: 0xC05076003D40009D

FC-4 TYPES:
  Supported: 0x0000010000000000000000000000000000000000000000000000000000000000
  Active:    0x0000010000000000000000000000000000000000000000000000000000000000
Class of Service: 3
Port Speed (supported): UNKNOWN
Port Speed (running):   0 GBIT
Port FC ID: 0x020802
Port Type: Fabric
```

Output omitted for brevity

Redbook Updated to include NPIV …



VIOS 2.1

IBM

What is new with VIOS 2.1

1. N_Port ID Virtualisation (NPIV)
   - Simplifies Fibre Channel SAN LUN Administration
   - Enables access to other SAN devices like Tape
2. Virtual Tape
   - Simplifies backup & restore with shared devices
   - Both virtual internal and via NPIV
3. Dynamic Heterogeneous Multi-Path I/O
   - LPAR with Direct attached SAN to temporarily go virtual for Partition Mobility
   - Inexpensive alternative paths = virtual standby backup path
4. Partition Mobility between HMCs
   - Improved flexibility
5. IVM Support for i on 520(8203) and 550(8204)
   - Provides an easier to use, lower cost of entry virtualization solution
6. Active Memory Sharing – statement of direction
   - Like Shared CPU but for memory
7. PowerVM Lx86 1.3
   - New higher performance

VIOS 2.1 - Upgrade Process

VIOS 2.1

- Major upgrade (AIX 5 → AIX 6 based)
- Boot from VIOS Migrate DVD [Not the Install DVD]
  - Like AIX 5.3 to 6.1 upgrade – approximately an hour
  - Can also boot from NIM to do the upgrade
  - Will preserve VIOS configuration & resources
  - From then on regular VIOS fix packs
- OK for Dual VIOS = Online Non-disruptive Update
  - VIOS can be at different levels (recommend 1.5 and 2.1) as SEA & vSCSI interfaces are not changed
  - Take one VIOS down, upgrade, start up, take over, upgrade the other
  - Sensibly to keep Dual VIOS at the same level
  - Recommend to update both VIOS in say the same day/weekend

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States, other countries, or both.

Not all common law marks used by IBM are listed on this page. Failure of a mark to appear does not mean that IBM does not use the mark nor does it mean that the product is not actively marketed or is not significant within its relevant market.

Those trademarks followed by ® are registered trademarks of IBM in the United States; all others are trademarks or common law marks of IBM in the United States.

- **Some, all or none of this presentation might, may or will be true or not, as applicable.**

For a complete list of IBM Trademarks, see www.ibm.com/legal/copytrade.shtml:

*, AS/400®, e business(logo)®, DBE, ESCO, eServer, FICON, IBM®, IBM (logo)®, iSeries®, MVS, OS/390®, pSeries®, RS/6000®, S/30, VM/ESA®, VSE/ESA, WebSphere®, xSeries®, z/OS®, zSeries®, z/VM®, System i, System i5, System p, System p5, System x, System z, System z9®, BladeCenter®

The following are trademarks or registered trademarks of other companies.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.
Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.
Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.
Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.
Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.
UNIX is a registered trademark of The Open Group in the United States and other countries.
Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.
ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.
IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency, which is now part of the Office of Government Commerce.

* All other products may be trademarks or registered trademarks of their respective companies.

Notes:
Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.
IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.
All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.
This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.
All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.
Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.
Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.