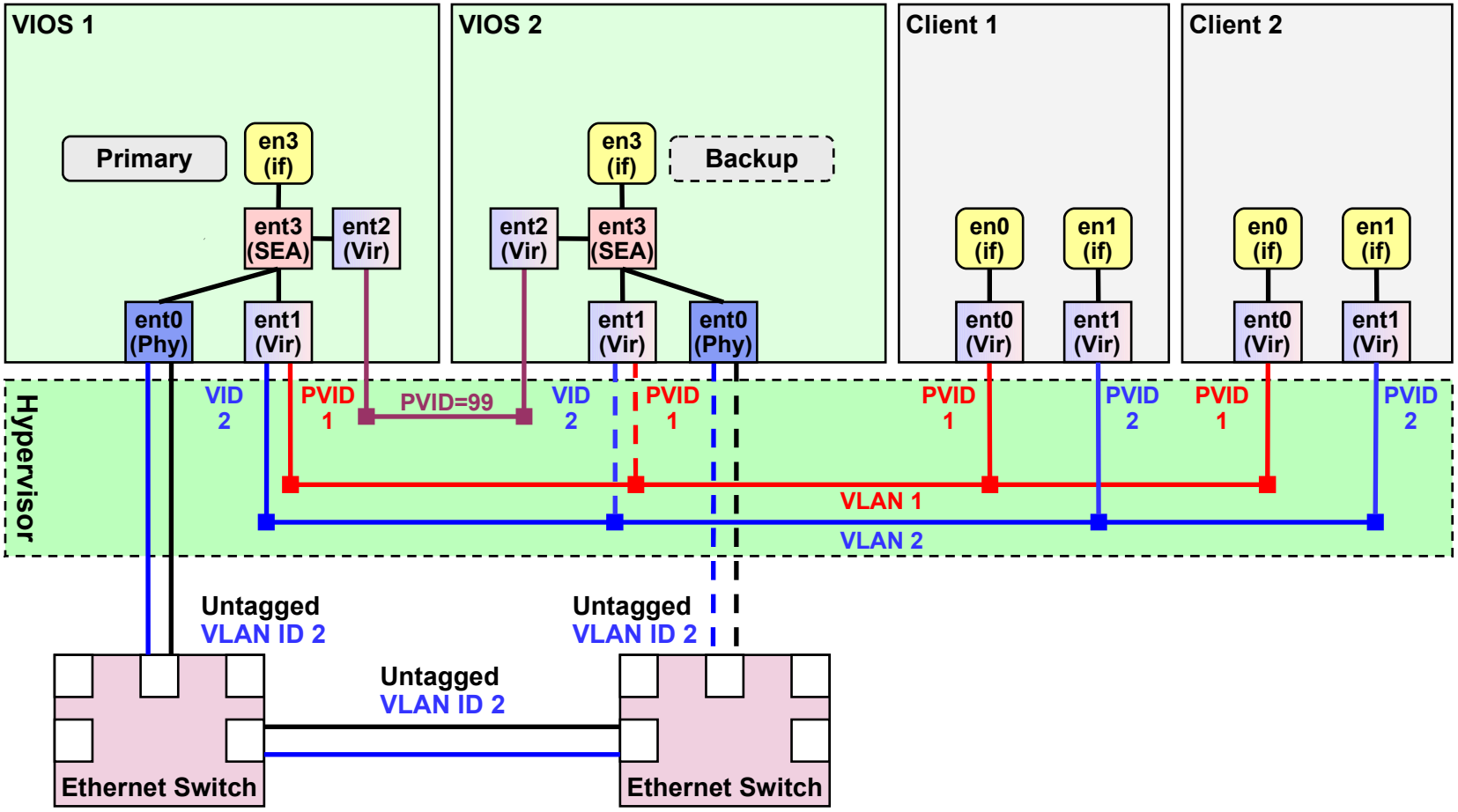
June 2012 @Dubai
IBM Power Academy

IBM PowerVM
disk virtualization

Luca Comparini
STG Lab Services Europe
IBM FR

June,13th 2012
@IBM Dubai

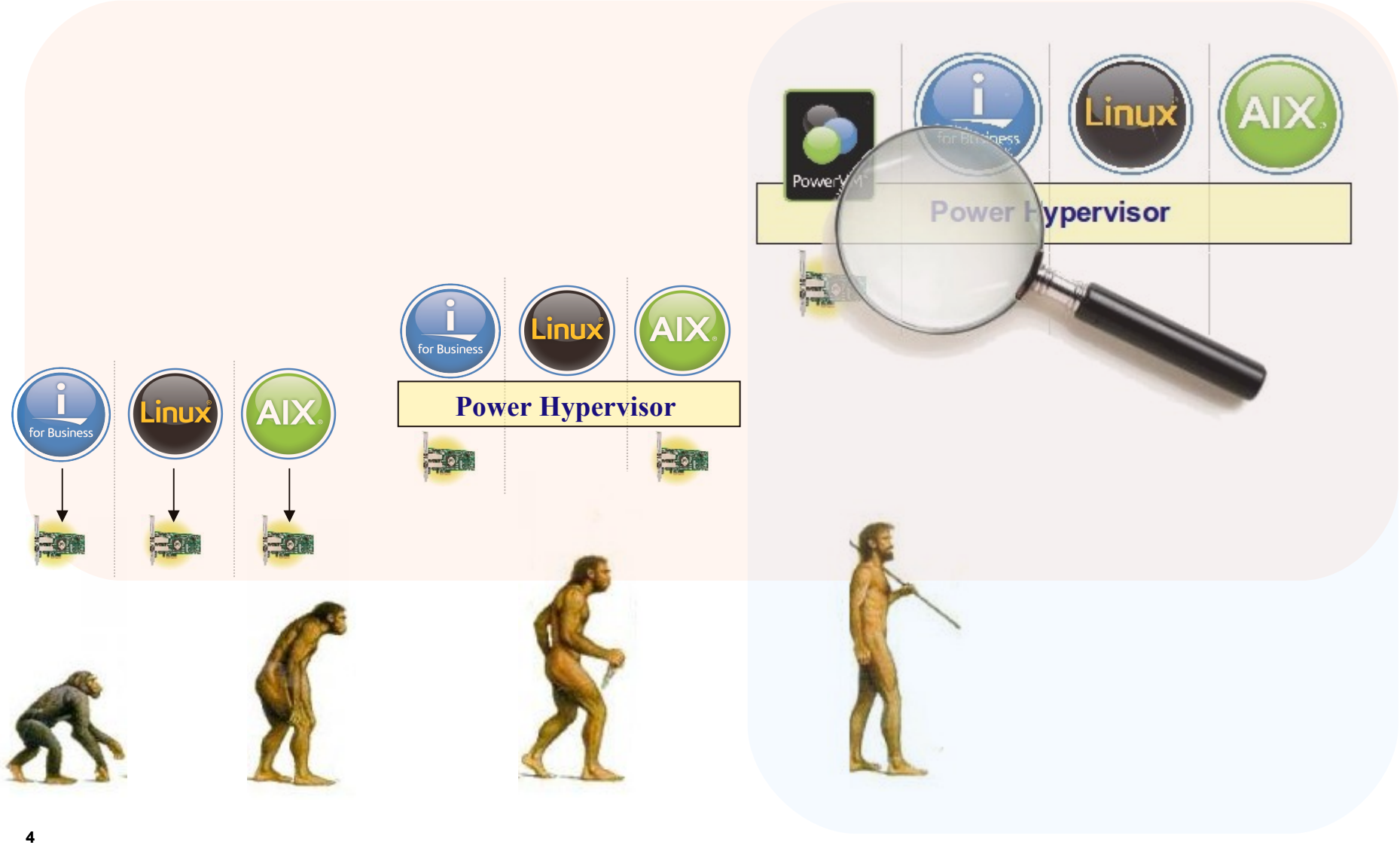
Objective of the session: understand this chart



Agenda

- PowerVM – VTSCSI Server / Client
- Single VIOS – Logical Volume as rootvg of the client LPAR
- Dual VIOS – Physical Volume as rootvg of the client LPAR
- Dual VIOS – NPIV Physical Volume as rootvg of the client LPAR

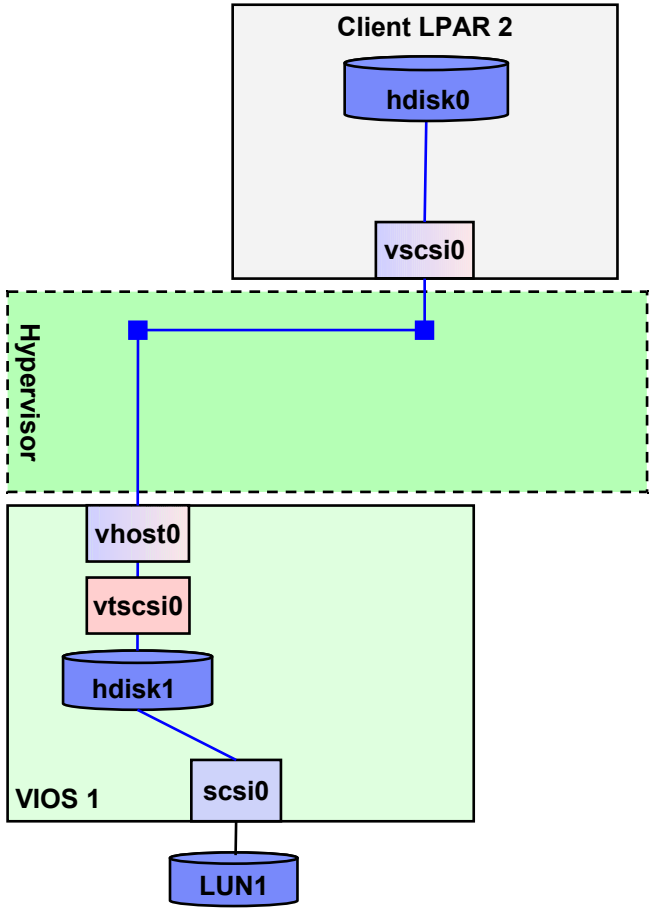
Introduction on VIOS – concepts of Virtual I/O Server - Client



The concept of VTD

Virtual SCSI (Small Computer Systems Interface) adapters provide one logical partition with the ability to use storage I/O (disk, CD, and tape) that is owned by another logical partition.

A virtual SCSI client adapter in one logical partition can communicate with a virtual SCSI server adapter in another logical partition. The virtual SCSI client adapter allows a logical partition to access a storage device being made available by the other logical partition. The logical partition owning the hardware is the server logical partition, and the logical partition that uses the virtualized hardware is the client logical partition. With this arrangement, the system can have many server logical partitions.




```
SVSA     Physloc                    Client PartitionID
--------------------------------------------------------
vhost0   U9117.570.10C8BCE-V6-C2    0x00000002

VTD              vtscsi0
Status           Available
LUN              0x8100000000000000
Backing device   hdisk1
Physloc
```

How to create virtual SCSI server / client

1. Define the vSCSI server adapter
   - This is done on the Hardware Management Console and creates a Virtual SCSI Server Adapter (for example vhost1) with a selectable slot number.

2. Define the vSCSI client adapter
   - This is also done on HMC and creates a Virtual SCSI Client Adapter (for example vscsi0) with a selectable slot number. When creating the Virtual SCSI Client Adapter you have to choose the desired I/O Server partition and the slot number of the Virtual SCSI Server Adapter defined during step 1.

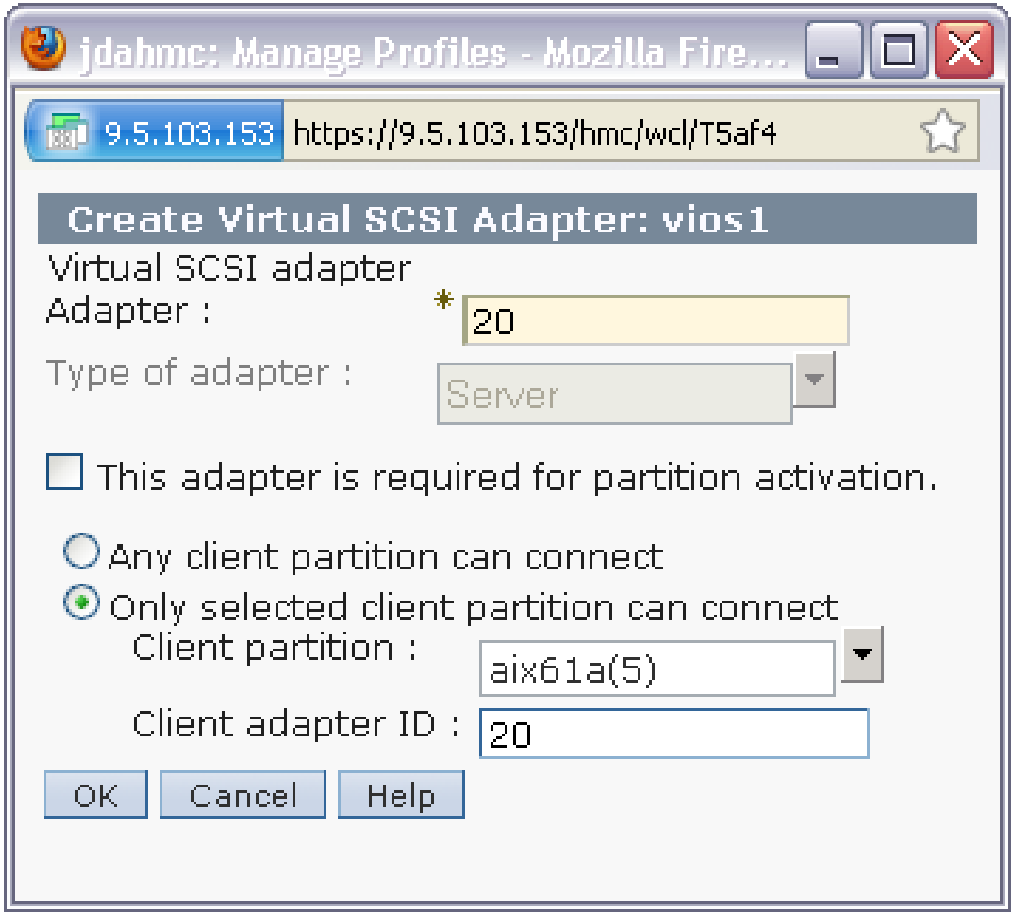
3. Create the required underlying logical volumes / volume groups / etc
   - This is done on VIO Server

4. Map the virtual SCSI Server adapter to the underlying SCSI resources.
   - On the I/O Server you have to map either a physical volume or a logical volume to the defined Virtual SCSI Server Adapter. This creates a Virtual Target Device (for example vtscsi2) that provides the connection between the I/O Server and the AIX partition through the POWER Hypervisor.

     The mapped volume now appears on the AIX partition as an hdisk device.

How to create virtual SCSI server / client

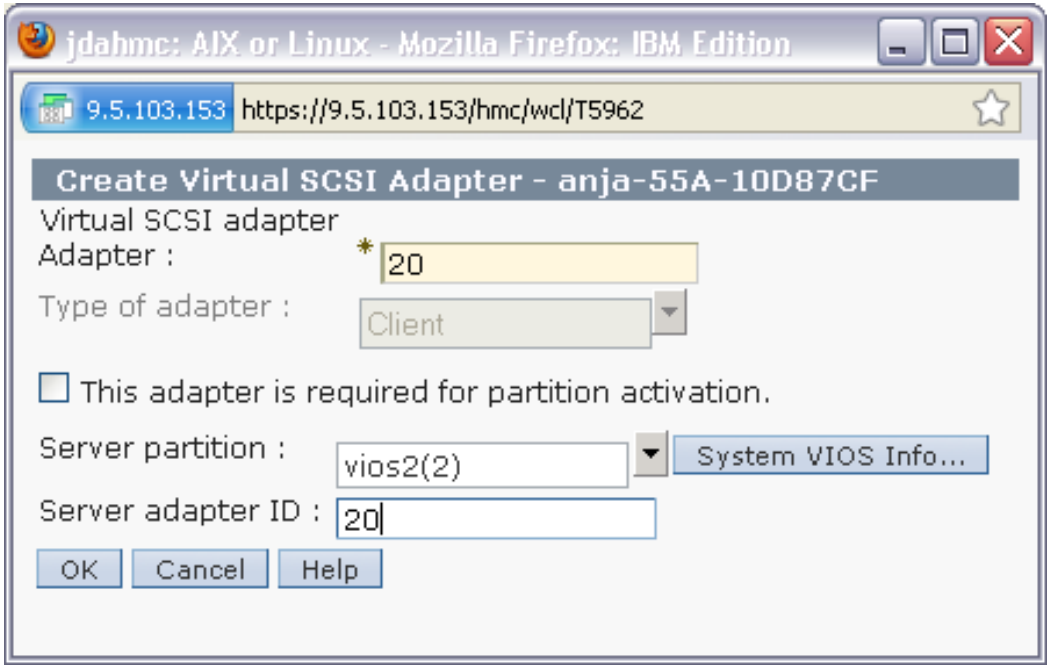
1. Define the vSCSI server adapter
   - This is done on the Hardware Management Console and creates a Virtual SCSI Server Adapter (for example vhost1) with a selectable slot number.



How to create virtual SCSI server / client

Define the vSCSI client adapter

This is also done on HMC and creates a Virtual SCSI Client Adapter (for example vscsi0) with a selectable slot number. When creating the Virtual SCSI Client Adapter you have to choose the desired I/O Server partition and the slot number of the Virtual SCSI Server Adapter defined during step 1.



How to create virtual SCSI server / client

- Create a volume group and assign disk to this volume group using the mkvg command as follows:
  - mkvg -f -vg rootvg_clients hdisk2

- Define the logical volume which will be visible as a disk to the client partition.
  - mklv -lv rootvg_dbsrv rootvg_clients 2G

How to create virtual SCSI server / client

- List the virtual Server Adapters
  - lsdev -vpd | grep vhost

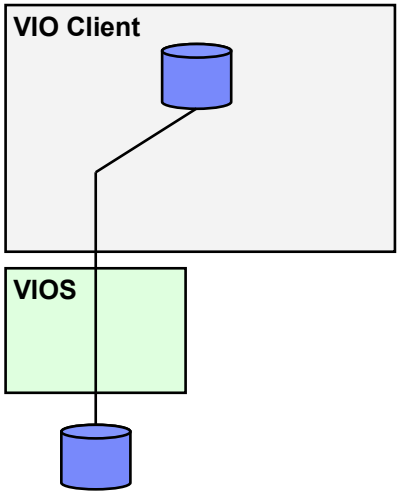
  - vhost2 U9111.520.10DDEEC-V2-C40 Virtual SCSI Server Adapter
    vhost1 U9111.520.10DDEEC-V2-C30 Virtual SCSI Server Adapter
    vhost0 U9111.520.10DDEEC-V2-C20 Virtual SCSI Server Adapter

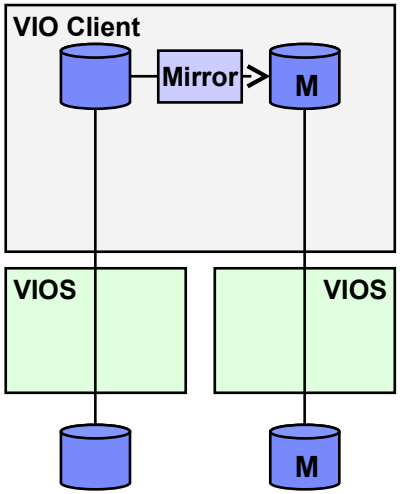
- Create a virtual target device, which maps the newly created virtual SCSI server adapters to a logical volume, by running the mkvdev command

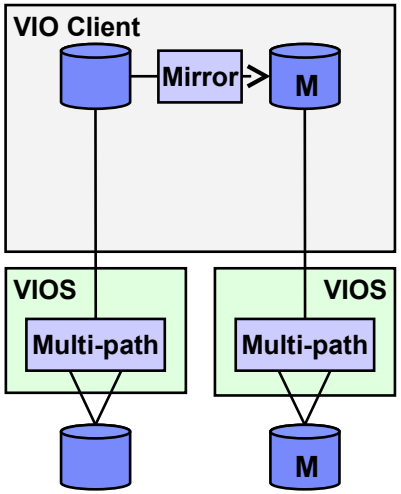
  - mkvdev -vdev rootvg_dbsrv -vadapter vhost0 -dev vdbsrv

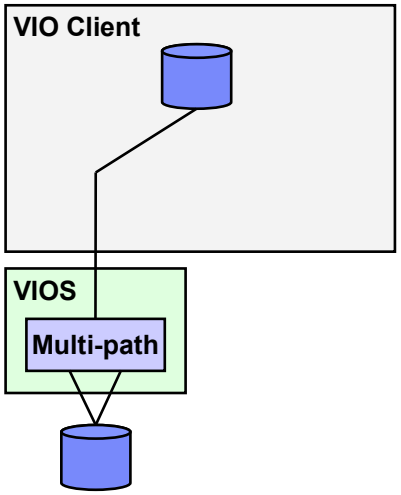
- Note: rootvg_dbsrv is a logical volume you have created before, vhost0 is your new virtual SCSI adapter and vdbsrv is the name of the new target device which will be available to the client partition.

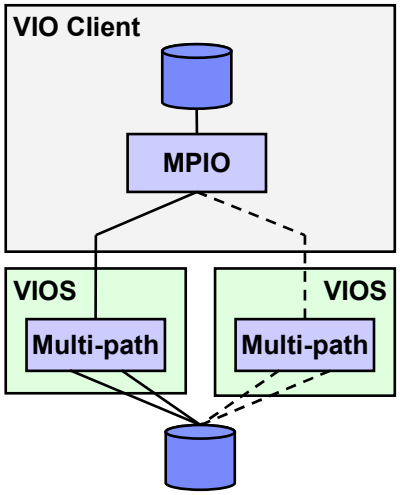
Multi path and mirroring high level options

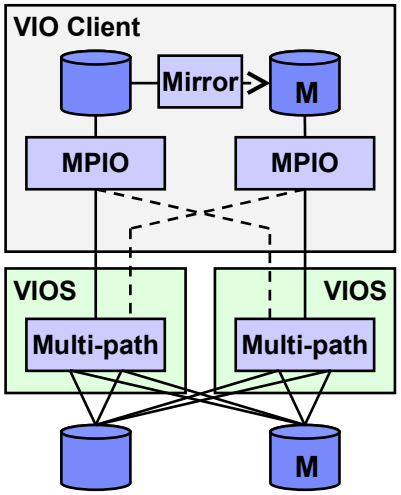




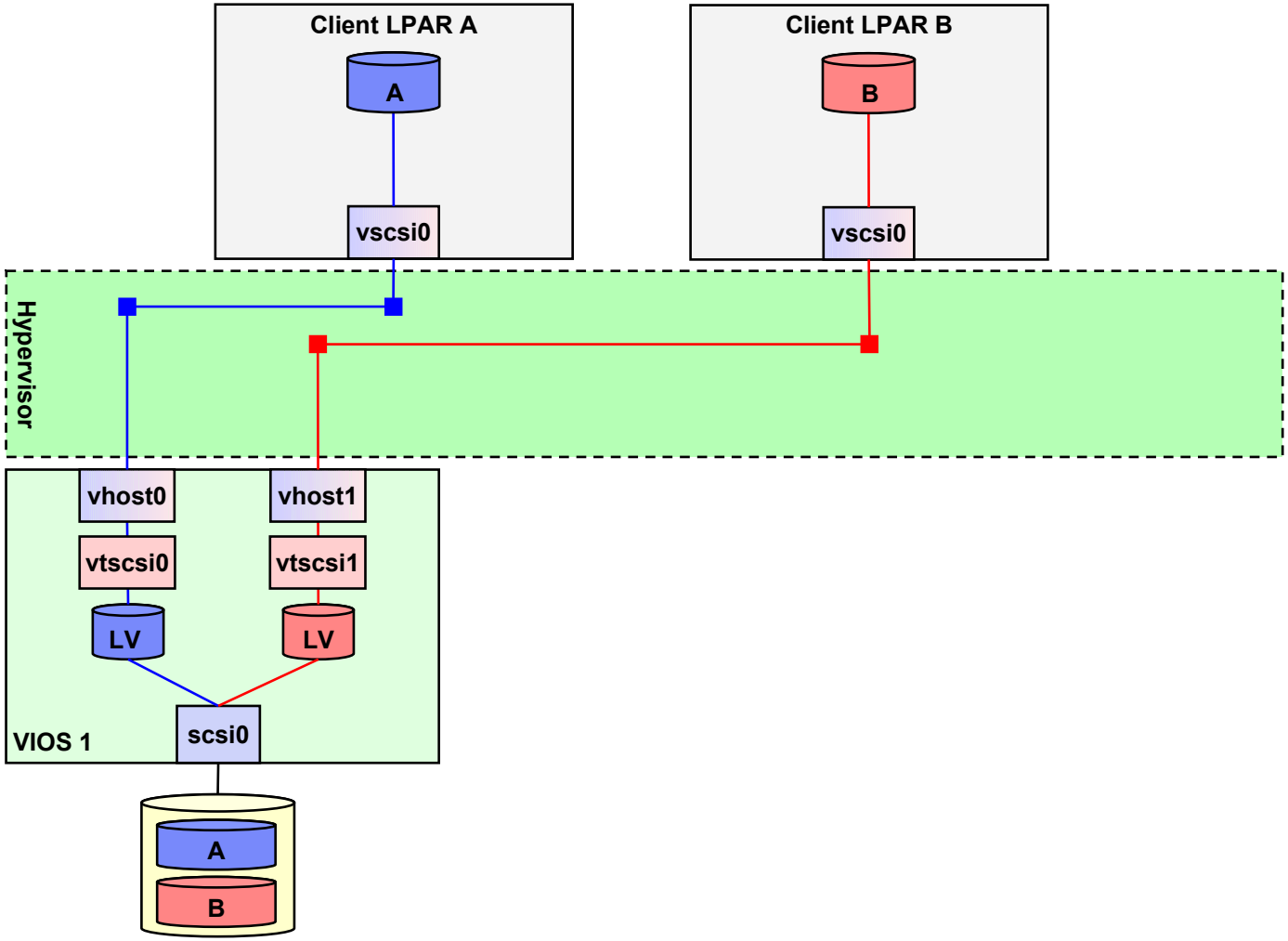






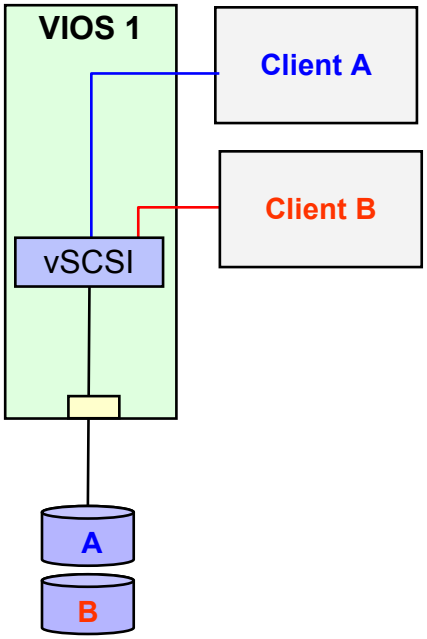


LV VTSCSI disk – Single VIOS

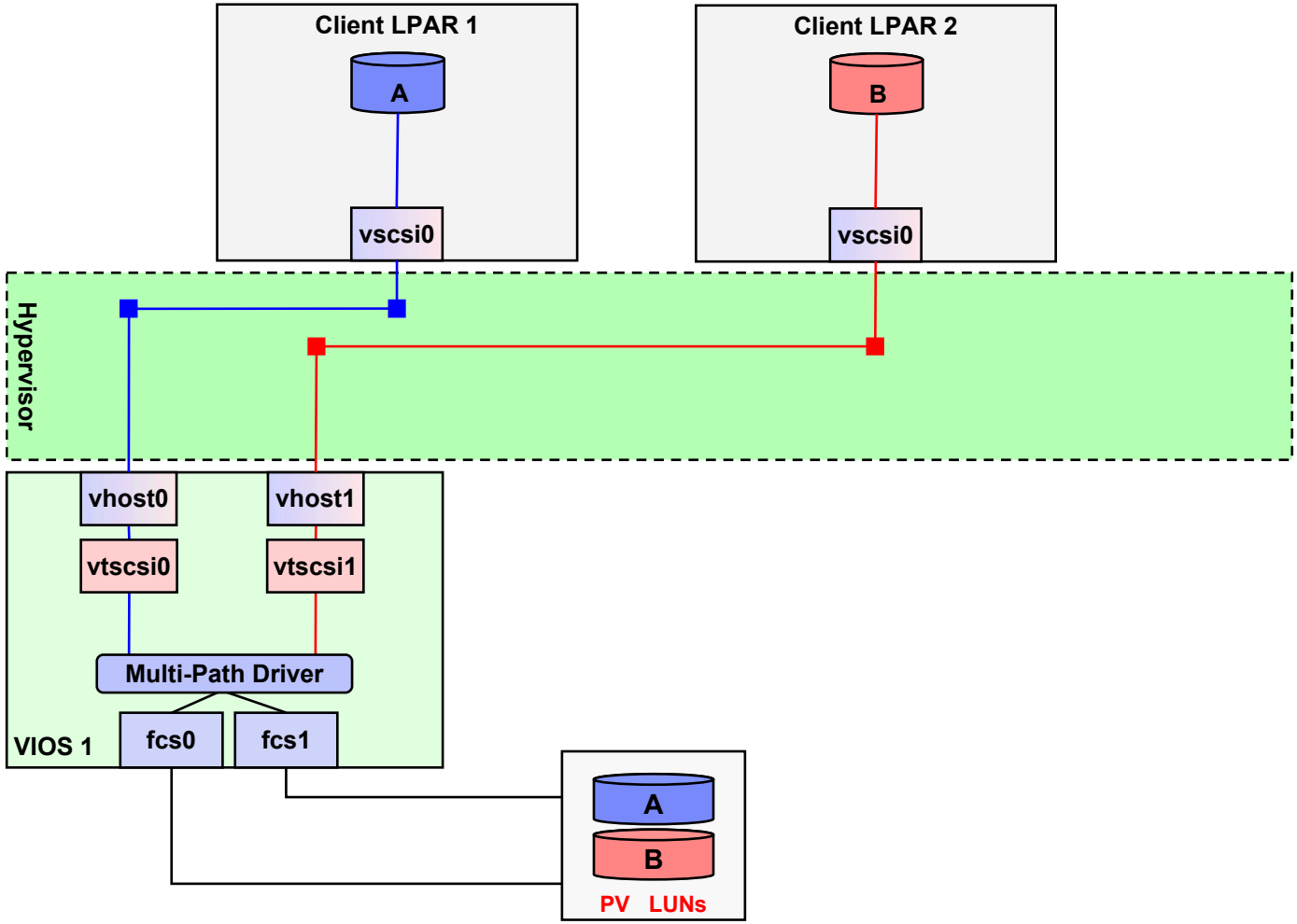


LV VTSCSI disk – Single VIOS

- Complexity
  - Simpler to setup and manage than dual VIOS
  - No specialized setup on the client
- Resilience
  - VIOS, SCSI adapter, SCSI disk are potential single points of failure
  - The loss of a single physical client disk will affect only that client
- Throughput / Scalability
  - Performance limited by single SCSI adapter and internal SCSI disks.

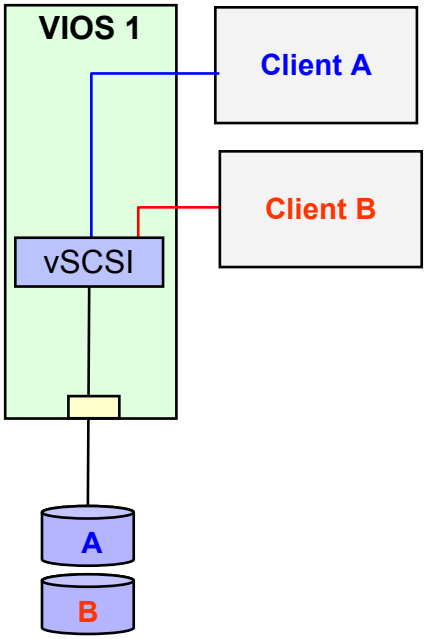


PV VTSCSI disk – Single VIOS

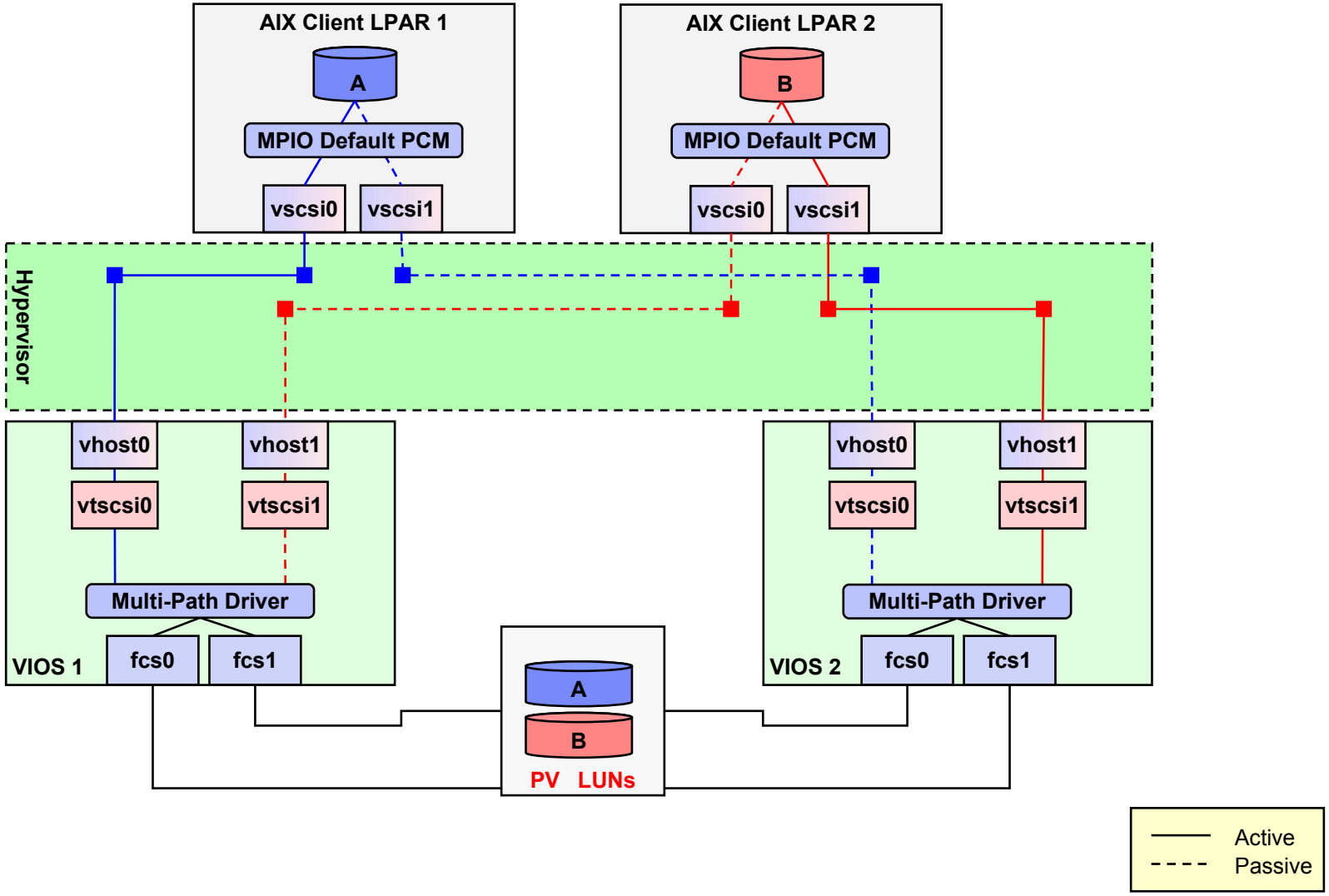


PV VTSCSI disk – Single VIOS

- Complexity
  - Simpler to setup and manage than dual VIOS
  - No specialized setup on the client
- Resilience
  - VIOS, SCSI adapter, SCSI disk are potential single points of failure
  - The loss of a single physical client disk will affect only that client
- Throughput / Scalability
  - Performance limited by single SCSI adapter and internal SCSI disks.



AIX MPIO driver in Client, Multi-Path I/O in VIOS



Virtual SCSI General Considerations

- Notes
  - Make sure you size the VIOS to handle the capacity for normal production and peak times such as backup.
  - Consider separating VIO servers that contain disk and network as the tuning issues are different
  - LVM mirroring is supported for the VIOS's own boot disk
  - A RAID card can be used by either (or both) the VIOS and VIOC disk
  - For performance reasons, logical volumes within the VIOS that are exported as virtual SCSI devices should not be striped, mirrored, span multiple physical drives, or have bad block relocation enabled..
  - SCSI reserves have to be turned off whenever we share disks across 2 VIOS. This is done by running the following command on each VIOS:

    # chdev -l <hdisk#> -a reserve_policy=no_reserve

Virtual SCSI General Considerations

- Notes
  - Make sure you size the VIOS to handle the capacity for normal production and peak times such as backup.
  - Consider separating VIO servers that contain disk and network as the tuning issues are different
  - LVM mirroring is supported for the VIOS's own boot disk
  - A RAID card can be used by either (or both) the VIOS and VIOC disk
  - For performance reasons, logical volumes within the VIOS that are exported as virtual SCSI devices should not be striped, mirrored, span multiple physical drives, or have bad block relocation enabled..
  - SCSI reserves have to be turned off whenever we share disks across 2 VIOS. This is done by running the following command on each VIOS:

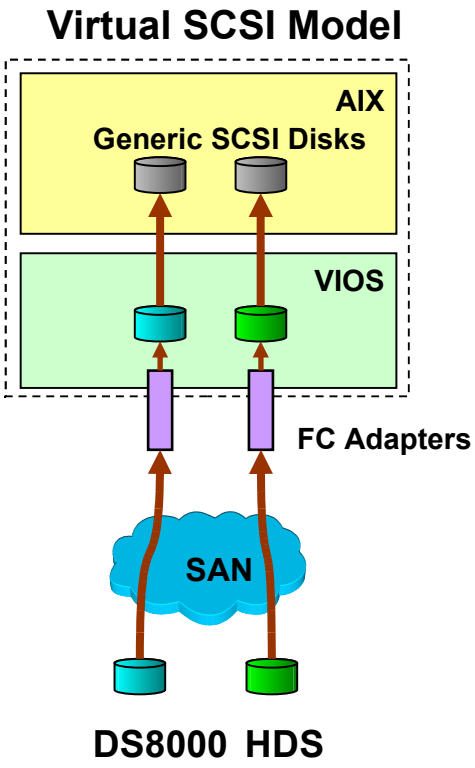
    # chdev -l <hdisk#> -a reserve_policy=no_reserve

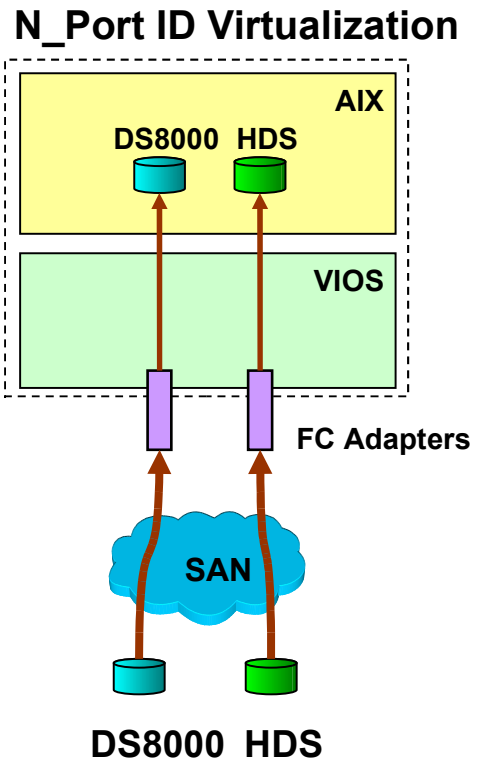
Virtual SCSI General Considerations

- Notes
  - If you are using FC Multi-Path I/O on the VIOS, set the following fscsi device values (requires switch attach):
    - dyntrk=yes (Dynamic Tracking of FC Devices)
    - fc_err_recov= fast_fail (FC Fabric Event Error Recovery Policy)
      (must be supported by switch)
  - If you are using MPIO on the VIOC, set the following hdisk device values:
    - hcheck_interval=60 (Health Check Interval)
  - If you are using MPIO on the VIOC set the following hdisk device values on the VIOS:
    - reserve_policy=no_reserve (Reserve Policy)

VTSCSI vs NPIV

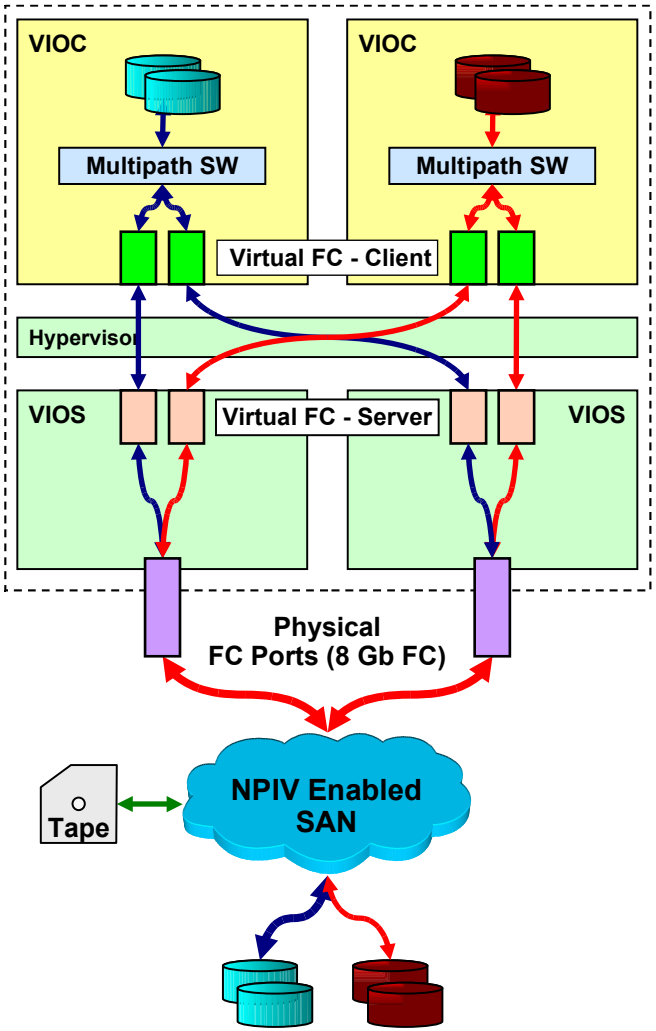
- N_Port ID Virtualization
  - Multiple Virtual World Wide Port Names per FC port – PCIe 8 Gb adapter
  - LPARs have direct visibility on SAN (Zoning/Masking)
  - I/O Virtualization configuration effort is reduced





NPIV

- N_Port ID Virtualization
  - Virtualizes FC adapters
  - Virtual WWPNs are attributes of the client virtual FC adapters not physical adapters
  - 64 WWPNs per FC port (128 per dual port HBA)
- Customer Value
  - Can use existing storage management tools and techniques
  - Allows common SAN managers, copy services, backup/restore, zoning, tape libraries, etc
  - Transparent use of storage functions such as SCSI-2 reserve/release and SCSI3 persistent reserve
  - Load balancing across VIOS
  - Allows mobility without manual management intervention



NPIV – things to consider

- WWPN pair is generated EACH time you create a VFC. NEVER is re-created or re-used. Just like a real HBA.
- If you create a new VFC, you get a NEW pair of WWPNs.
- Save the partition profile with VFCs in it. Make a copy, don't delete a profile with a VFCin it.
- Make sure the partition profile is backed up for local and disaster recovery! Otherwise you'll have to create new VFCs and map to them during a recovery.
- Target Storage SUBSYSTEM must be zoned and visible from source and destination systems for LPM to work.
- Active/Passive storage controllers must BOTH be in the SAN zone for LPM to work
- Do NOT include the VIOS physical 8G adapter WWPNs in the zone
- You should NOT see any NPIV LUNs in the VIOS
- Load multi-path code in the client LPAR, NOT in the VIOS
- No 'passthru' tunables in VIOS

Questions?

CREDITS to

John Banchy
System Architect
IBM US

Luca Comparini
STG Lab Services Europe
IBM FR

THANKS