VIO Options

John Banchy
Systems Architect

Version 1.0

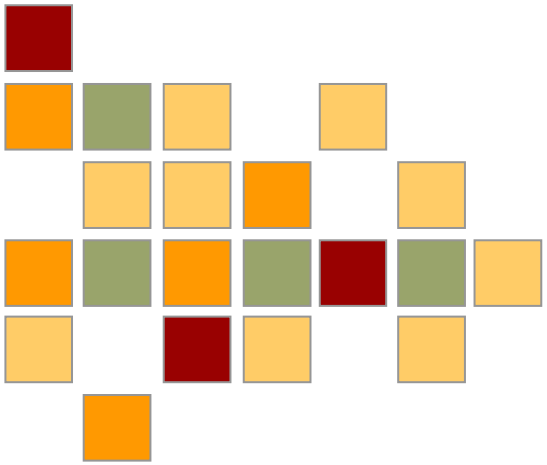
I would like to thank the following people for their help on this document:

- Bob G Kovacs - Austin
- Jorge R Nogueras - Austin
- Dan Braden - ATS
- John Tesch - ATS
- Ron Barker - ATS
- Greg R Lee - Markham
- Tenley Jackson - Southfield
- Armin M. Warda - Postbank, Germany
- Nigel Griffiths - UK
- Thomas Prokop – Minneapolis
- George Potter – Milwaukee
- Indulis Bernsteins - Australia
- Bill Armstrong – Rochester
- Bret Olszewski – Austin
- Jaya Srikrishnan – Poughkeepsie
- Kyle Lucke – Rochester
- Rakesh Sharma – Austin
- Ron Young – Rochester
- Satya Sharma – Austin
- Timothy Marchini – Poughkeepsie
- Keith Zblewski - Rochester

Disclaimers

IBM has not formally reviewed this document. While effort has been made to verify the information, this document may contain errors. IBM makes no warranties or representations with respect to the content hereof and specifically disclaim any implied warranties of merchantability or fitness for any particular purpose. IBM assumes no responsibility for any errors that may appear in this document. The information contained in this document is subject to change without any notice. IBM reserves the right to make any such changes without obligation to notify any person of such revision or changes. IBM makes no commitment to keep the information contained herein up to date.

If you have any suggestions or corrections please send comments to:
jbanchy@us.ibm.com



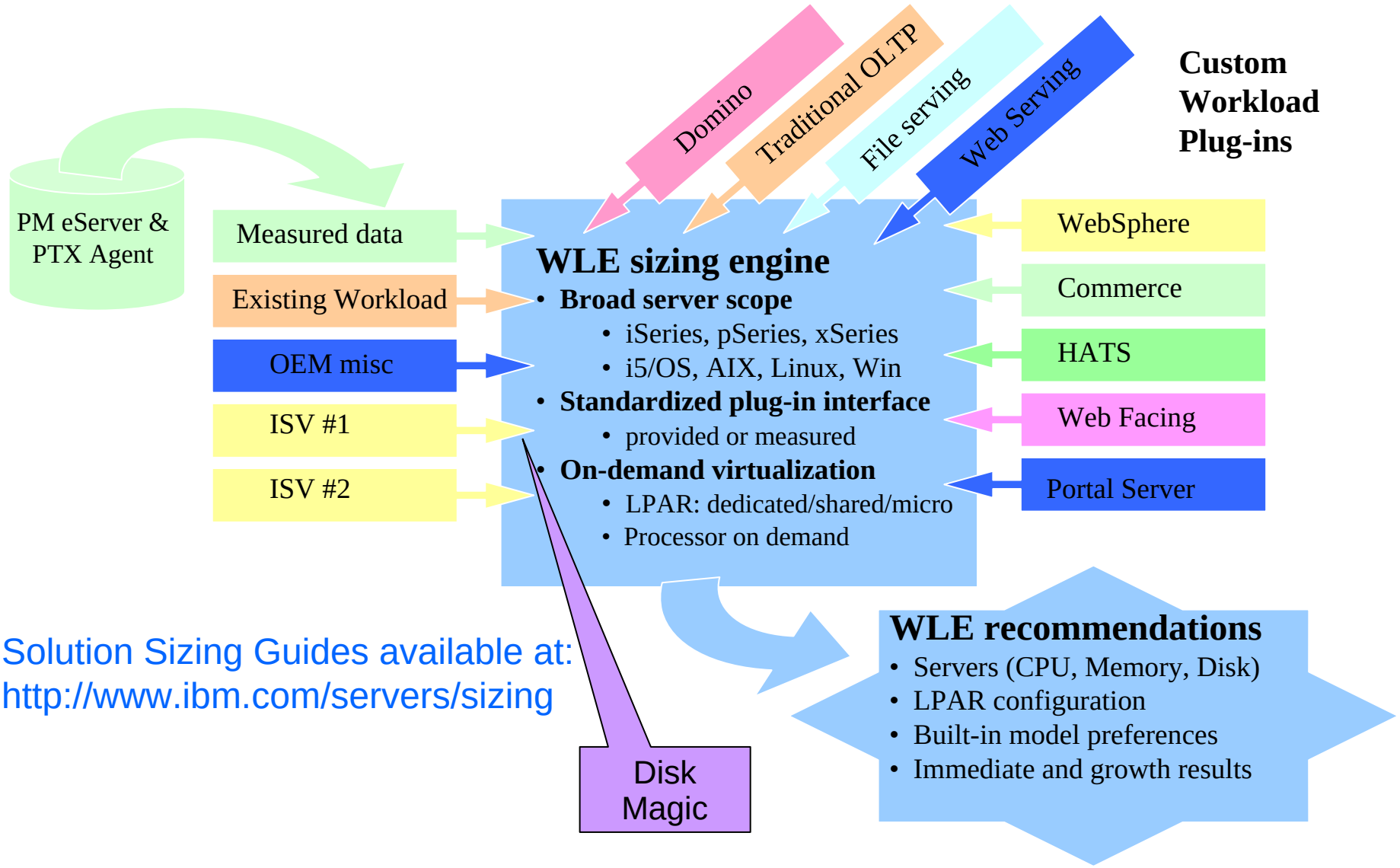
Workload Estimator

What is the Workload Estimator?

- **IBM Systems Workload Estimator (WLE) is a web-based sizing tool designed to assist customers, partners, and the IBM sales/support team when sizing**
  - Upgrades
  - Migrations
  - Server Consolidations
  - New Systems
- **WLE helps estimate the processor, disk, and memory required to run various workloads**
  - Uses a mature, proven sizing engine available since 1999
  - Based on detailed performance measurements taken by IBM Systems Performance experts
  - Introduced support for Virtualization (LPAR) in 2002 and VIO Server in 2005
  - Supports Linux workloads since 2003 and AIX since 2004
  - Focused on OS and middleware workload sizing with support for ISV workloads

Workload Estimator

http://www-912.ibm.com/wle/EstimatorServlet

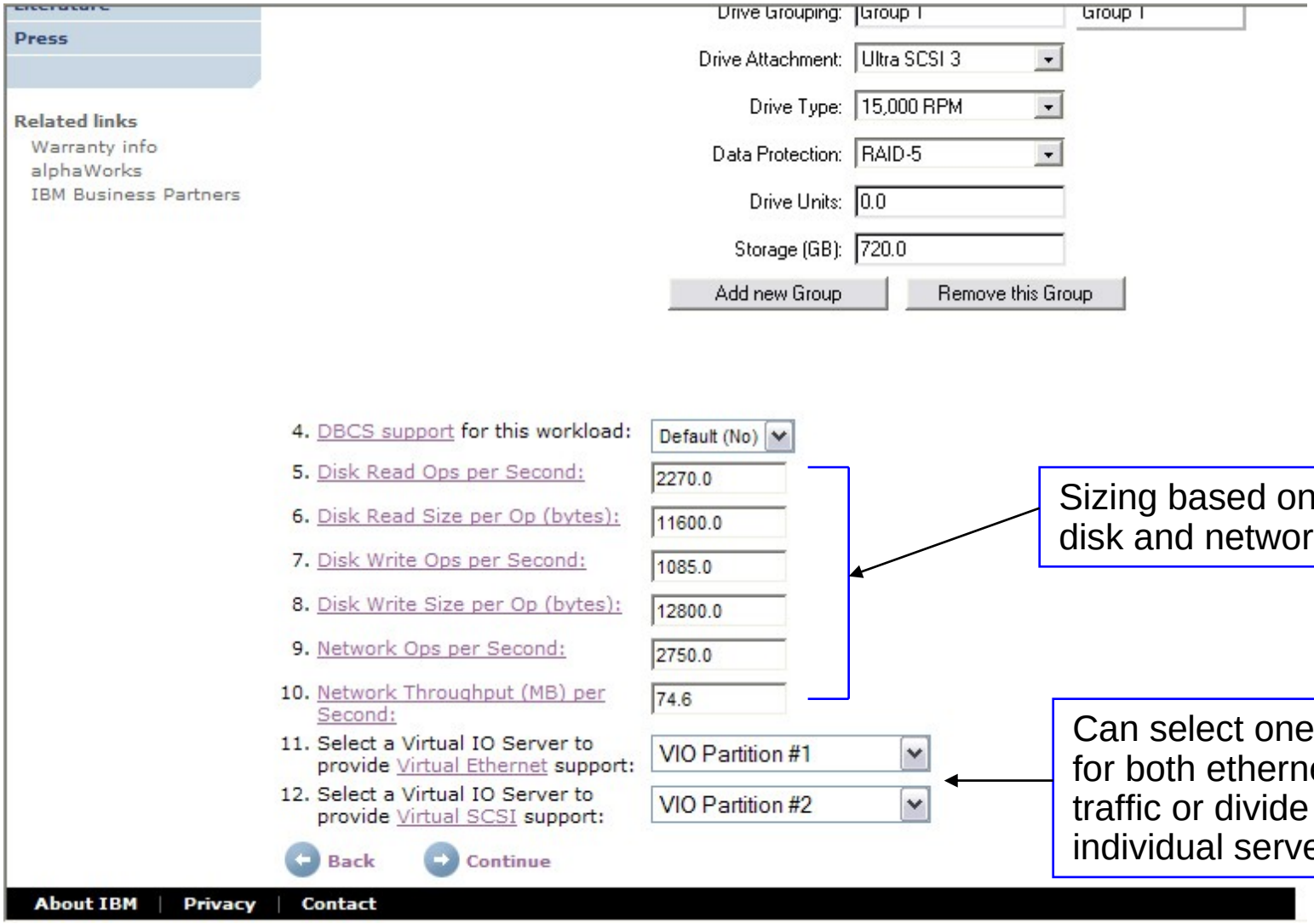


Solution Sizing Guides available at:
http://www.ibm.com/servers/sizing

Workloads Supported in WLE

- **i5/OS (OS/400)**
  - Existing
    - Manual input using performance data from existing system or LPAR
  - Generic
    - Based on estimated CPW required
  - Measured Customer Data (PM iSeries)
  - Domino
  - HATS
  - Traditional
    - OLTP Interactive/Batch Users
  - Web Serving (HTTP Server)
  - WebFacing
  - WebSphere
  - WebSphere Business Integration
  - WebSphere Commerce
  - WebSphere Portal Server
  - Workplace
  - Virtual I/O
- **AIX**
  - Existing
  - Generic
    - Based on estimated rPerf required
  - Measured Customer Data
    - PTX Agent and PM pSeries
  - File Serving
  - Web Serving
  - WebSphere Application Server
  - Virtual I/O Server (pSeries)
  - Domino (via eServer Sizing Guides)
- **Linux**
  - Generic
  - File Serving (Samba)
  - Web Serving (Apache)
  - WebSphere Application Server
  - HATS
  - Network Infrastructure
    - Firewall, DNS, DHCP servers
  - Virtual I/O Server (pSeries)

Sizing Virtual I/O Server Resources with WLE

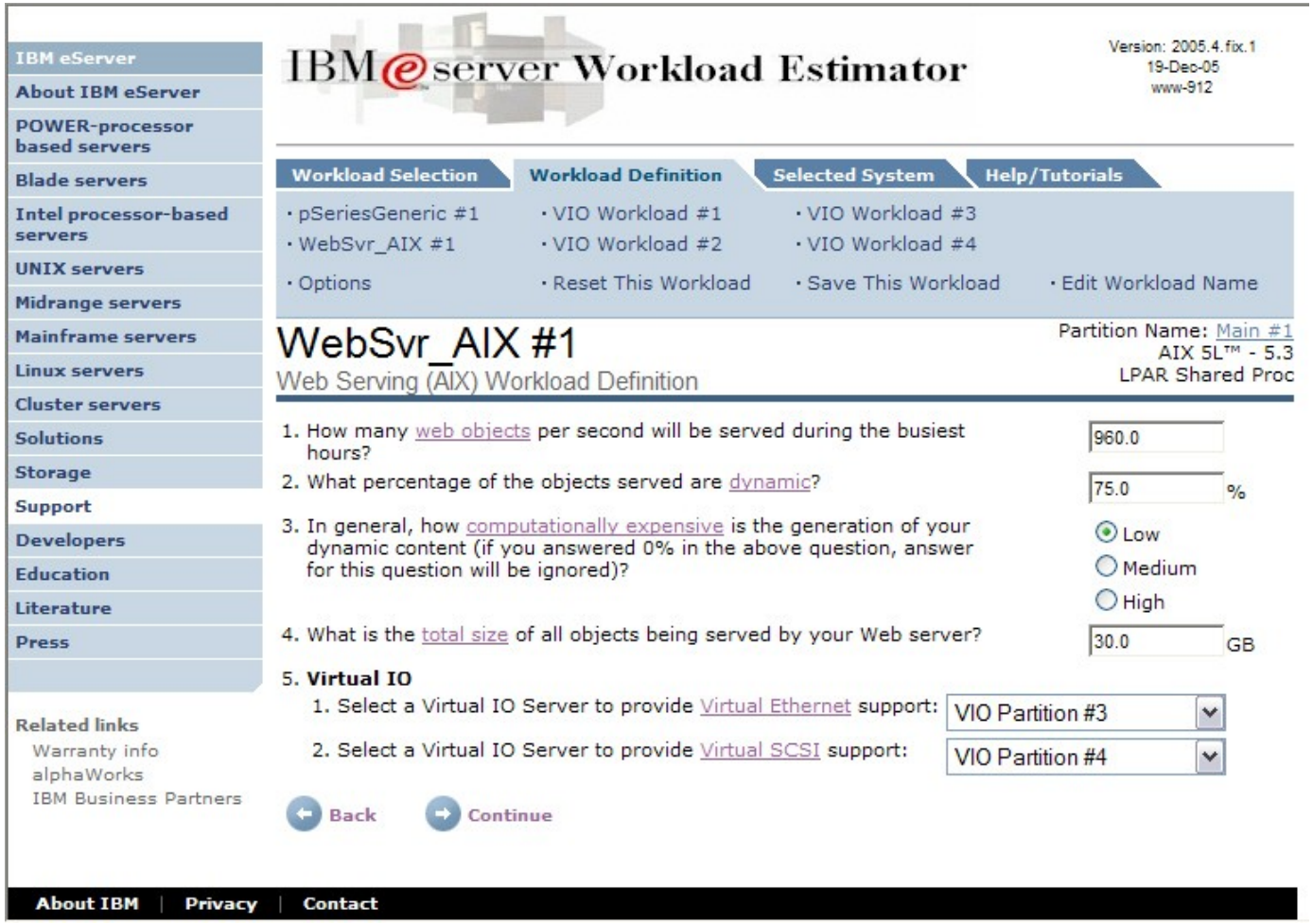


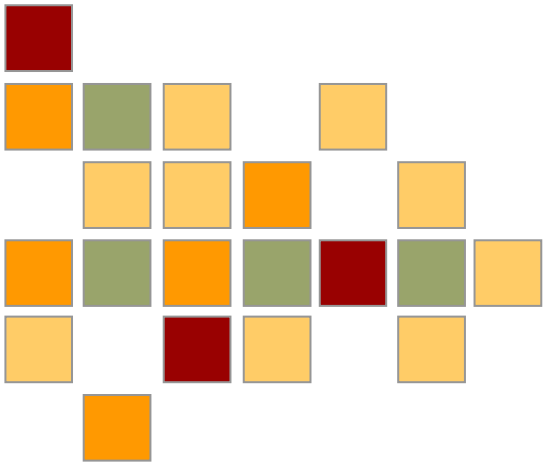
Sizing based on specified disk and network I/O rates

Can select one VIO server for both ethernet and SCSI traffic or divide traffic on individual servers.

Sizing Virtual I/O Servers with WLE

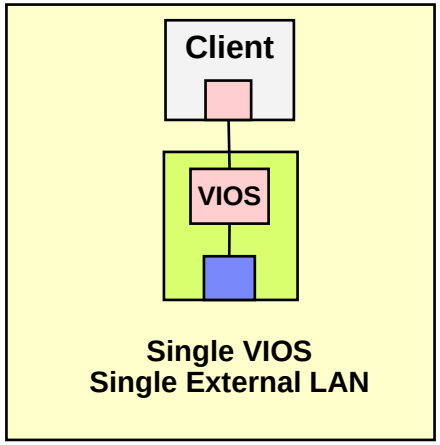
Sizing virtual I/O for pre-defined workloads like web serving is based on an estimate of I/O traffic that the specified inputs would generate.



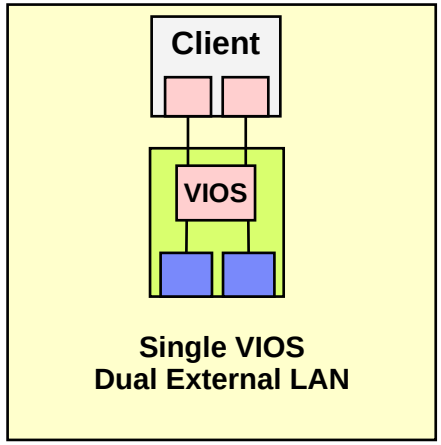


Virtual Ethernet

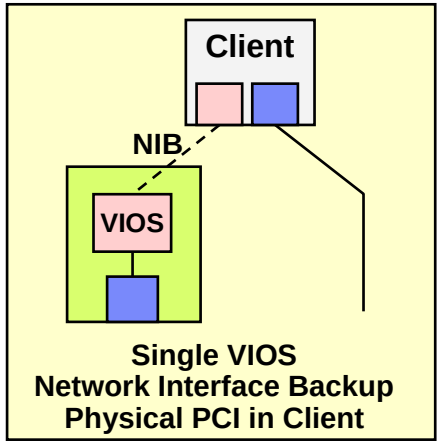
Ethernet Options in this Document



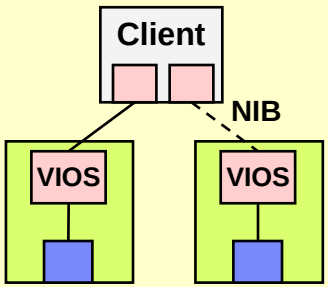
Single VIOS
Single External LAN



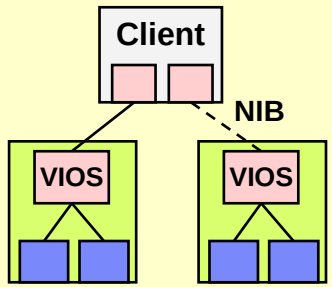
Single VIOS
Dual External LAN



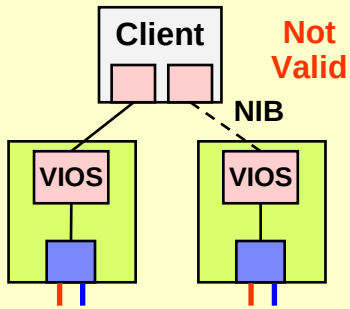
Single VIOS
Network Interface Backup
Physical PCI in Client



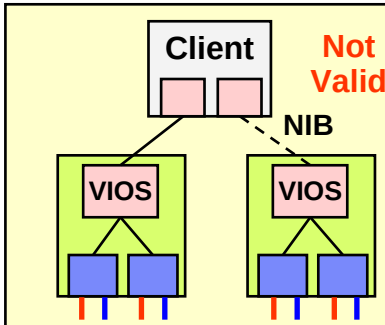
Dual VIOS
Network Interface Backup



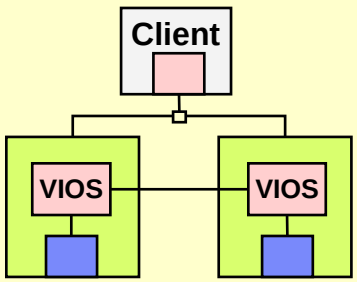
Dual VIOS
Network Interface Backup
Link Aggregation



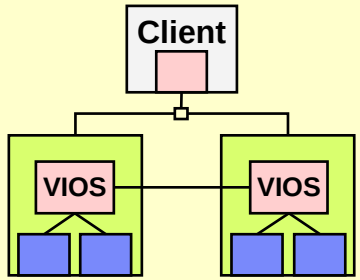
Dual VIOS
Network Interface Backup
VLANs



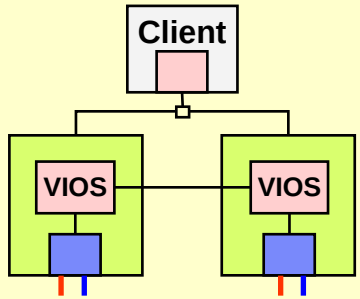
Dual VIOS
Network Interface Backup
Link Aggregation & VLANs



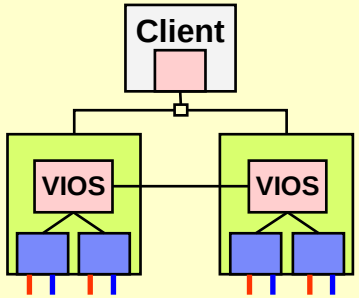
Dual VIOS
Shared Ethernet Failover



Dual VIOS
Shared Ethernet Failover
Link Aggregation



Dual VIOS
Shared Ethernet Failover
VLANs



Dual VIOS
Shared Ethernet Failover
Link Aggregation & VLANs

Virtual Ethernet Performance (not Shared Ethernet Switch)

- **Rules of Thumb**
  - Performance will depend on CPU entitlement and TCP/IP tuning
  - Choose the largest MTU size that makes sense for the traffic on the virtual network
  - Keep the attribute tcp_pmtu_discover set to “active discovery”
  - Use SMT unless your application required it to be turned off.
  - Performance scales with entitlement, not the number of virtual processors

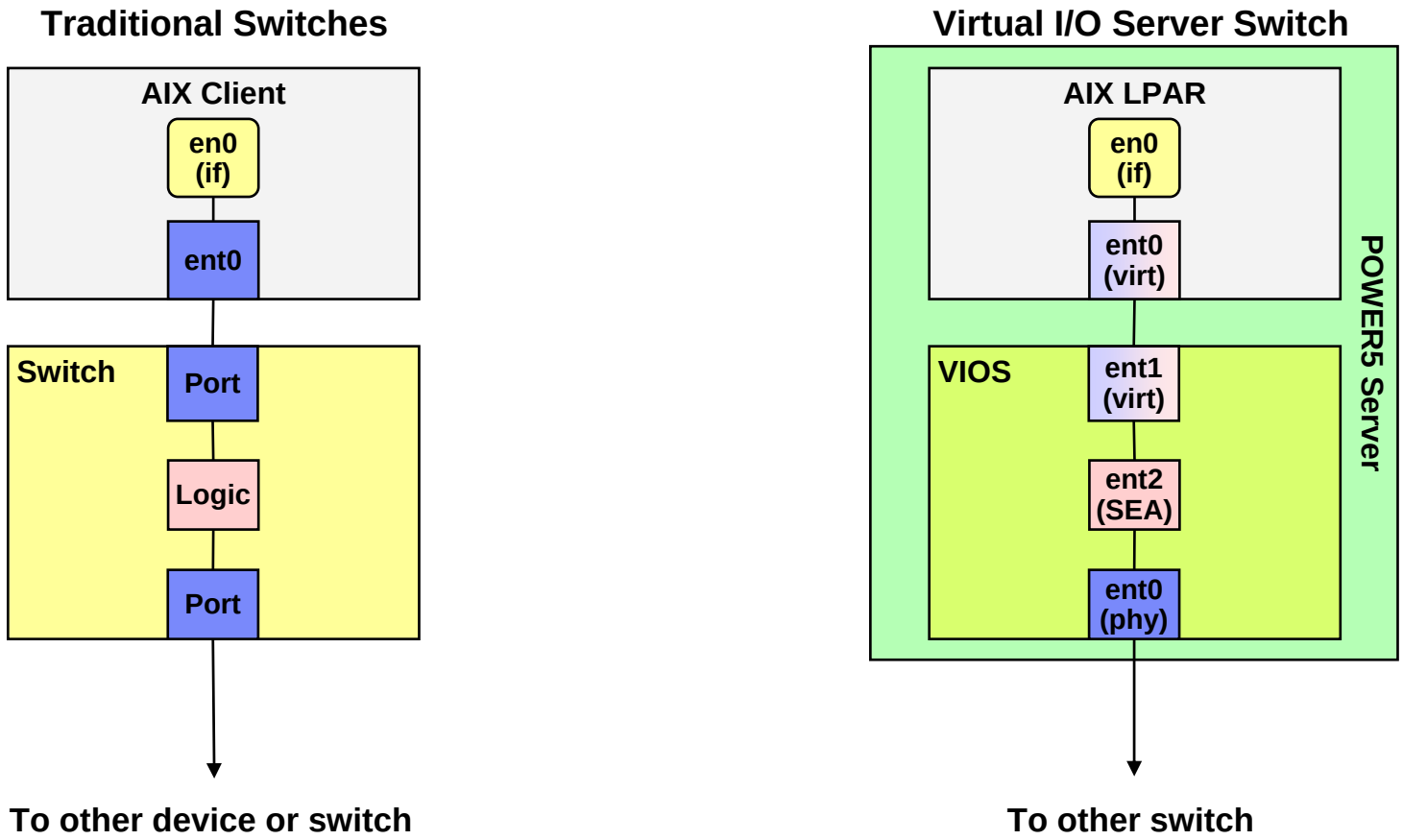
Shared Ethernet Performance

- **Rules of thumb for performance concerns**
  - Use dedicated adapters for demanding workloads
  - Choose the highest MTU size that makes sense
    - 1500 MTU or many small transactions can take 100% of a CPU (Gigabit Ethernet/1.65 GHz CPU)
    - One Gigabit Ethernet is the equivalent to 10 100 Mbps Ethernet adapters
    - Jumbo frames (9000 MTU) cut CPU utilization in half, but many small transactions can dilute the effect of a larger MTU size
  - Size the Virtual I/O Server for the combined VSCSI and Shared Ethernet workloads or use a separate VIO Server partition for Shared Ethernet adapter-only
  - Consider using dedicated processors versus micro-partitions, or make micro-partitions uncapped

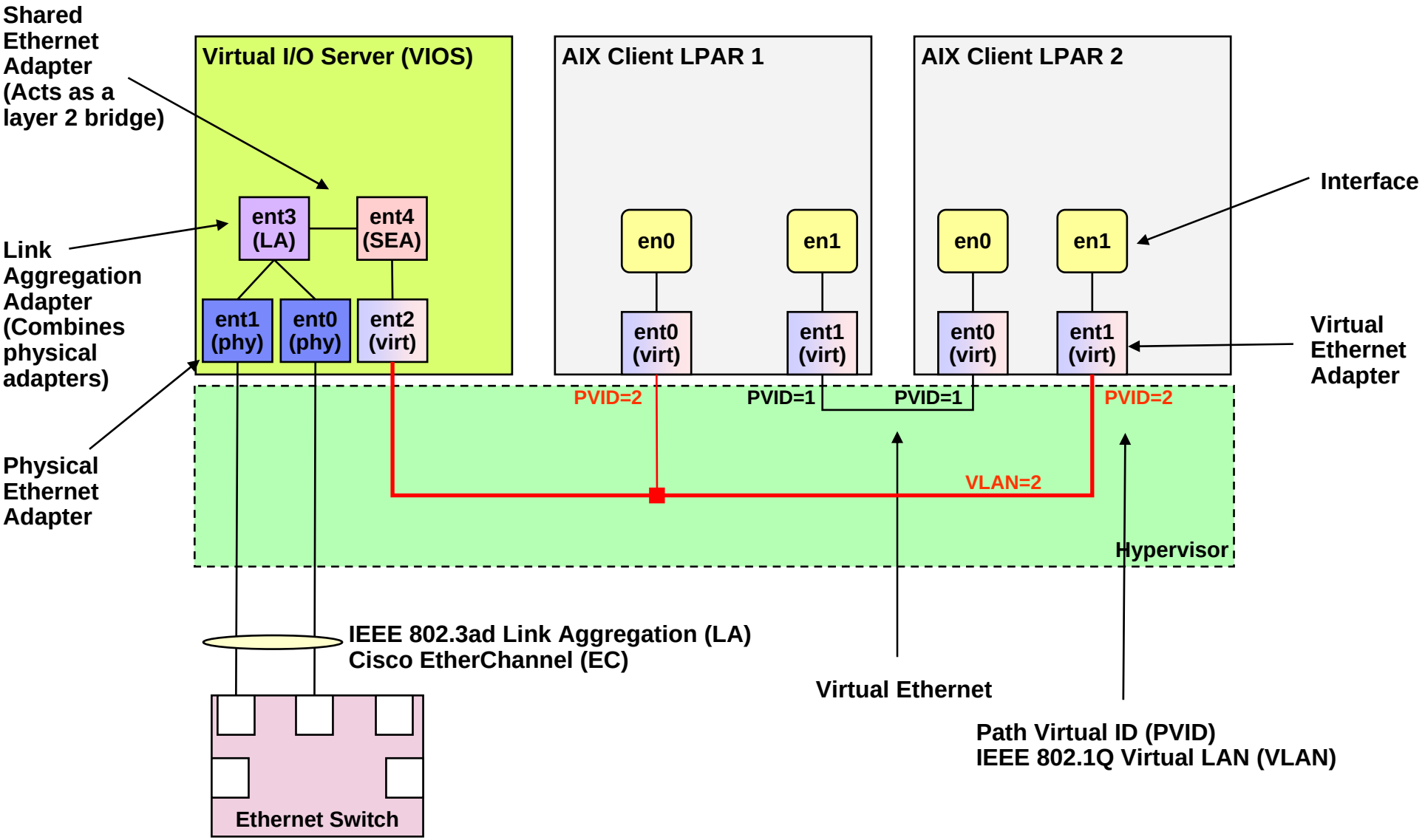
Shared Ethernet Performance

- **Rules of thumb (continued)**
  - Only use Shared Ethernet adapter threading if the Virtual I/O Server is also serving VSCSI
    - On the VIO Server, you can run these commands as padmin:
    - $ lsdev –dev ent3 –attr (show if threading is enabled; default is true)
    - $ chdev –dev ent3 –attr thread=0 (to turn it off; no vSCSI)
  - Always use SMT in AIX unless the application requires that it be turned off
  - Simplex, full and half-duplex jobs have different performance characteristics
    - Full duplex will perform better, if the media supports it
    - Full duplex will NOT be 2 times simplex, though, because of the ACK packets that are sent; about 1.5x simplex (Gigabit)
    - Some workloads require simplex or half-duplex

POWER5 VIO Server Switch Concepts



Virtual I/O Network Terms



Cisco EtherChannel and IEEE 802.3ad Link Aggregation

MODE	HASH MODE	OUTGOING TRAFFIC DISTRIBUTION (across adapter ports within the EtherChannel)
standard or 8023ad	default	The traditional AIX behavior. The adapter selection algorithm uses the last byte of the destination IP address (for TCP/IP traffic) or MAC address (for ARP and other non-IP traffic). This mode is typically the best initial choice for a server with a large number of clients.
standard or 8023ad	src_dst_port	The outgoing adapter path is selected via algorithm using the combined source and destination TCP or UDP port values. Average the TCP/IP address suffix values in the "Local" and "Foreign" columns shown by **netstat -an** command. Since each connection has a unique TCP or UDP port, the three port-based hash modes provide additional adapter distribution flexibility when there are several, separate TCP or UDP connections between an IP address pair.
standard or 8023ad	src_port	The adapter selection algorithm uses the source TCP or UDP port value. In **netstat -an** command output, the port is the TCP/IP address suffix value in the "Local" column.
standard or 8023ad	dst_port	The outgoing adapter path is selected via algorithm using the destination system port value. In **netstat -an** command output, the TCP/IP address suffix in the "Foreign" column is the TCP or UDP destination port value.
round robin	default	Outgoing traffic is spread evenly across all the adapter ports in the EtherChannel. This mode is the typical choice for two hosts connected back-to-back (i.e. without an intervening switch).

Notes

- All of the table except round robin (EtherChannel only) applies to both EtherChannel and IEEE 802.3ad.
- IBM recommends using source destination port (src_dst_port) as the mechanism where possible.

Virtual Ethernet Legend

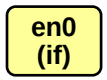
Failover

Active

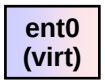
VIOS VIO Server

VIOC VIO Client

Physical Ethernet Adapter



Network Interface



Virtual Ethernet Adapter

Shared Ethernet Adapter

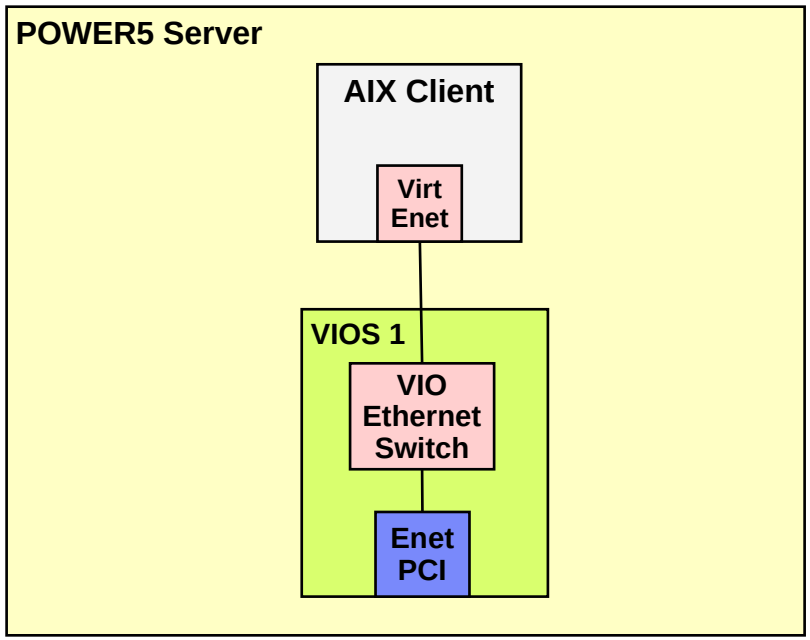
Cisco EtherChannel or
IEE 802.3ad Link Aggregation

Link Adapter

Virtual Ethernet Options

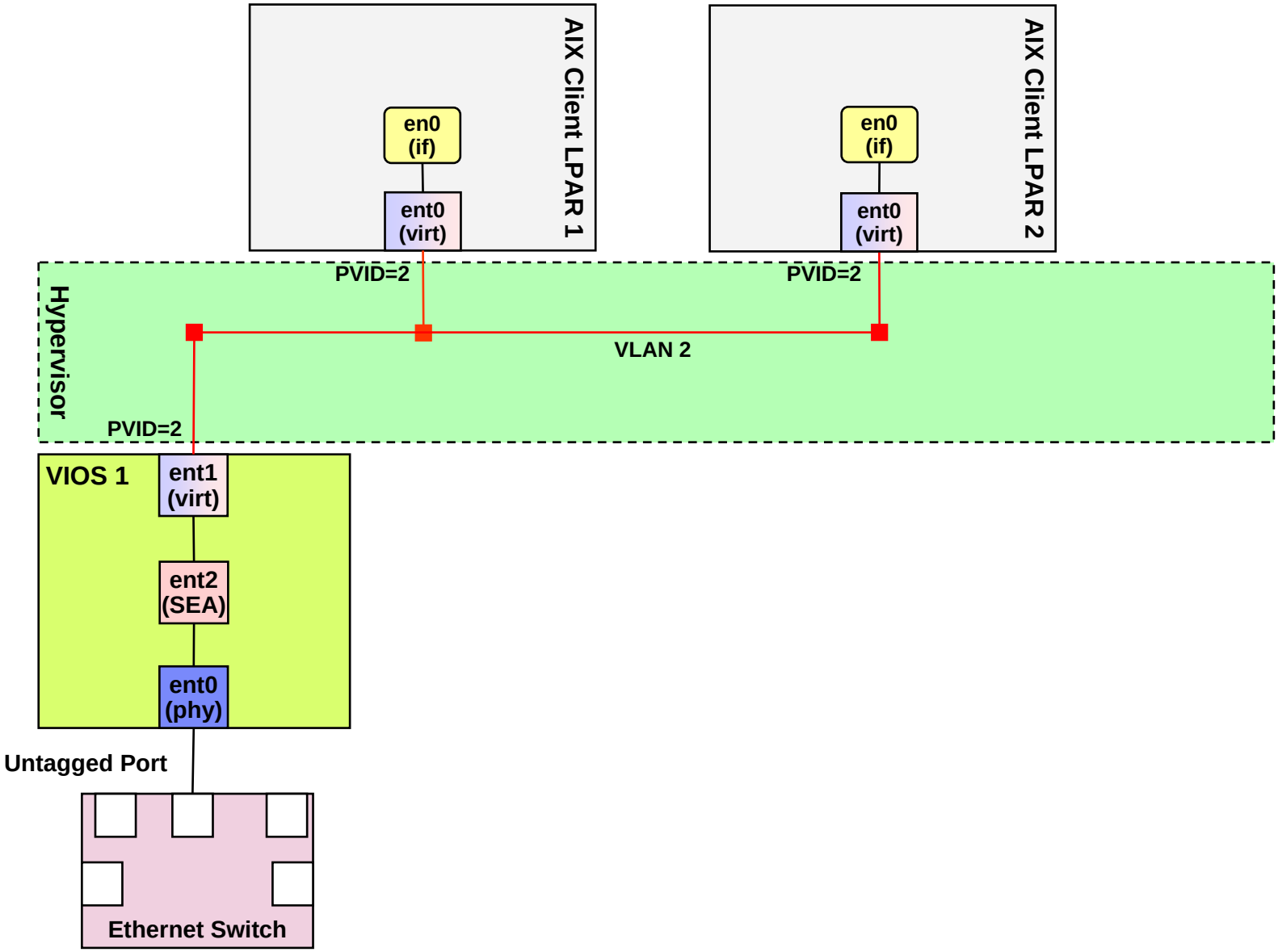
Single VIOS – Single LAN

- **Complexity**
  - Easy to setup and manage
  - No specialized configuration on switch
  - No specialized configuration on client
- **Resilience**
  - VIOS Ethernet adapter, switch port and switch are single points of failure
- **Throughput / Scalability**
  - Performance limited to a single Ethernet adapter



Virtual Ethernet Options - Details

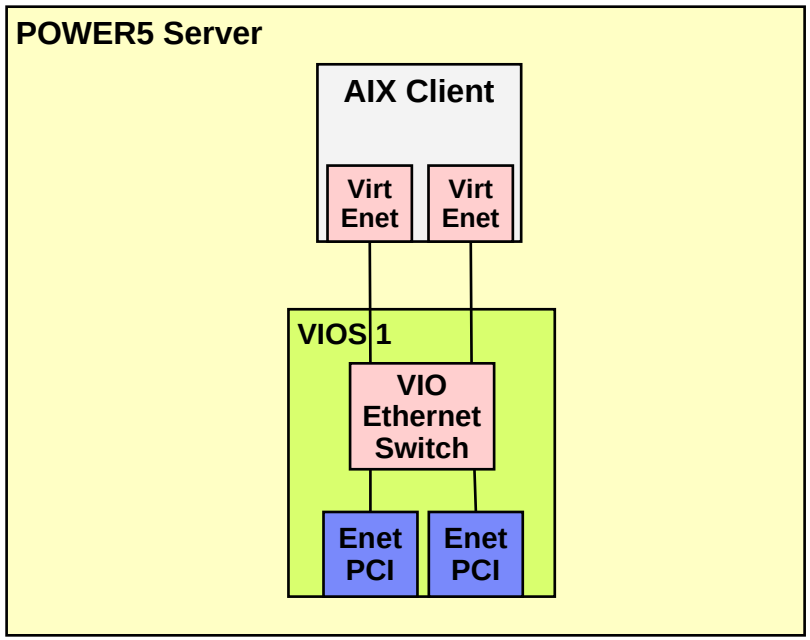
Single VIOS – Single LAN Segment



Virtual Ethernet Options

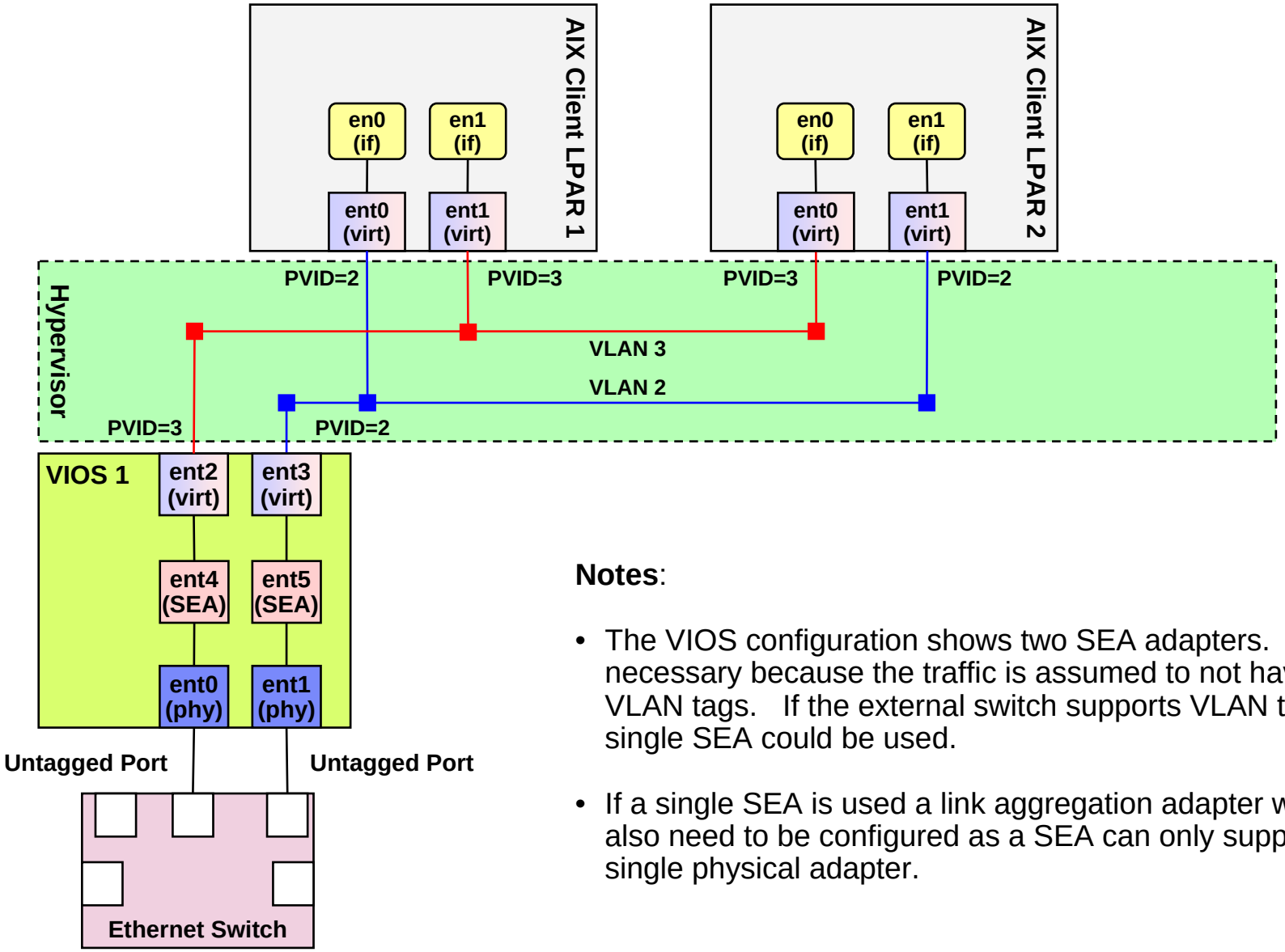
Single VIOS – Multiple LAN Segments

- **Complexity**
  - Easy to setup and manage
  - No specialized configuration on switch
  - No specialized configuration on client
- **Resilience**
  - VIOS Ethernet adapter, switch port, and switch are single points of failure
- **Throughput / Scalability**
  - Performance limited to a single Ethernet adapter for each LAN segment



Virtual Ethernet Options - Details

Single VIOS – Multiple LAN Segments



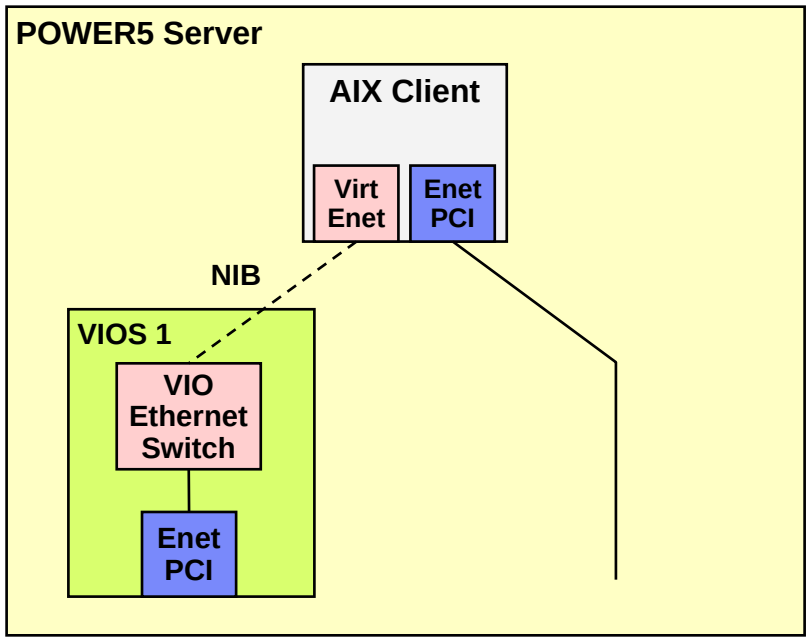
Notes:

- The VIOS configuration shows two SEA adapters. This is necessary because the traffic is assumed to not have VLAN tags. If the external switch supports VLAN tags, a single SEA could be used.
- If a single SEA is used a link aggregation adapter would also need to be configured as a SEA can only support a single physical adapter.

Virtual Ethernet Options

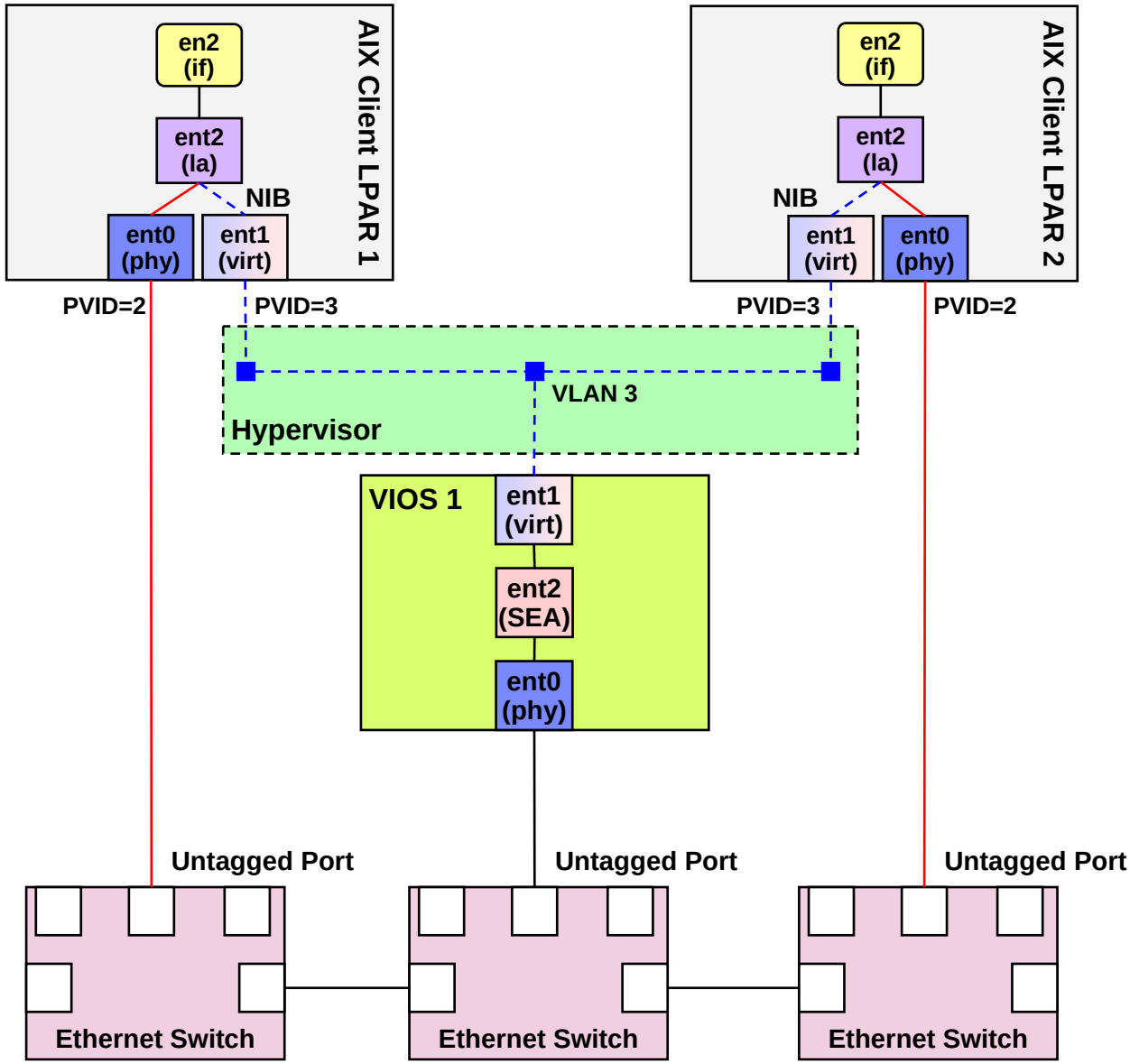
AIX Network Interface Backup (NIB), Single VIOS, PCI Adapter in Client

- **Complexity**
  - Requires specialized setup on client (NIB)
  - Needs to ping outside host from the client to initiate NIB failover Resilience
  - Protects against single switch port / switch / Ethernet adapter failure
- **Throughput / Scalability**
  - Backup performance limited to a single Ethernet adapter
- **Notes**
  - NIB does not support tagged VLANs on physical LAN.
  - Must use external switches not hubs



Virtual Ethernet Options – Details

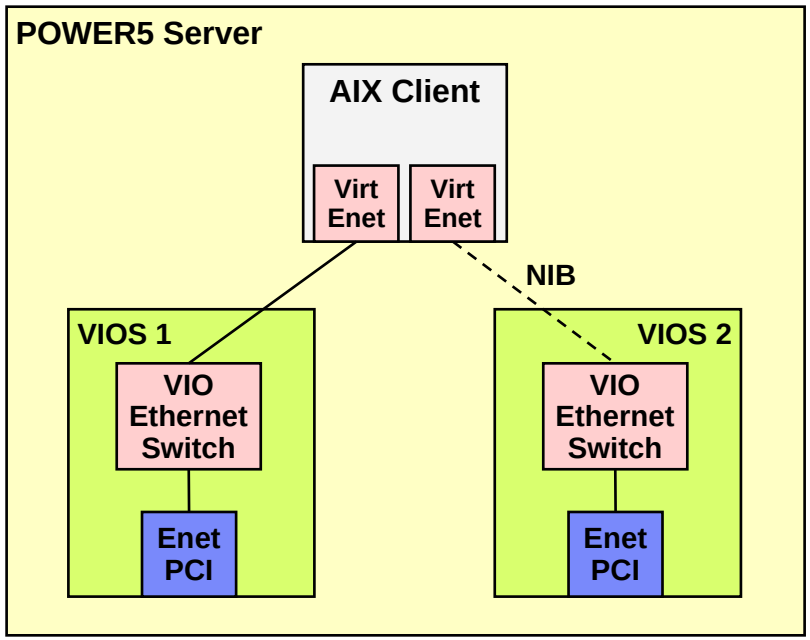
AIX Network Interface Backup (NIB), Single VIOS, PCI Adapter in Client



Virtual Ethernet Options

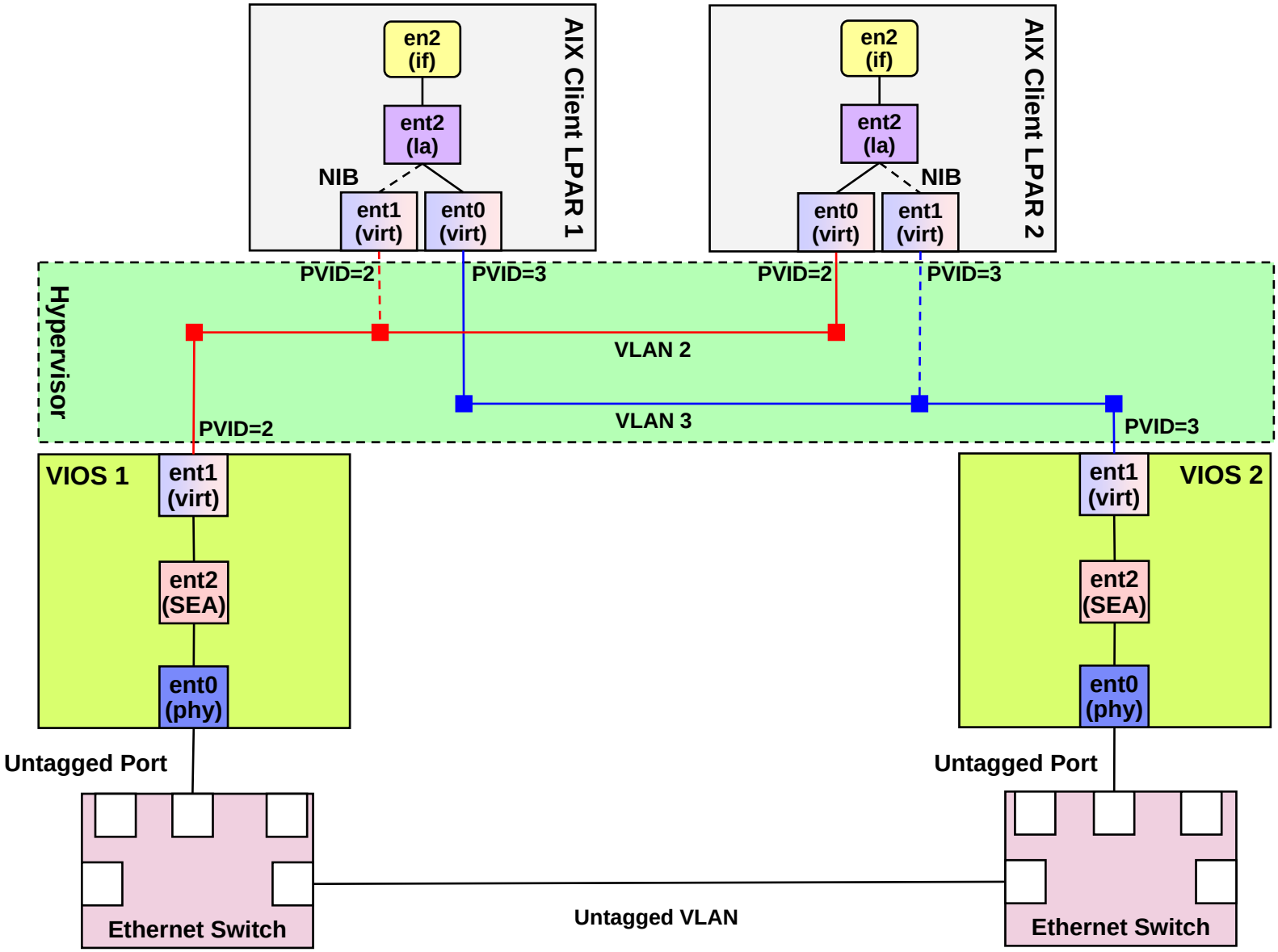
AIX Network Interface Backup (NIB), Dual VIOS

- **Complexity**
  - Requires specialized setup on client (NIB)
  - Needs to ping outside host from the client to initiate NIB failover
- **Resilience**
  - Protects against single VIOS / switch port / switch / Ethernet adapter failures
- **Throughput / Scalability**
  - Allows load-sharing between VIOS's
- **Notes**
  - NIB does not support tagged VLANs on physical LAN
  - Must use external switches not hubs



Virtual Ethernet Options - Details

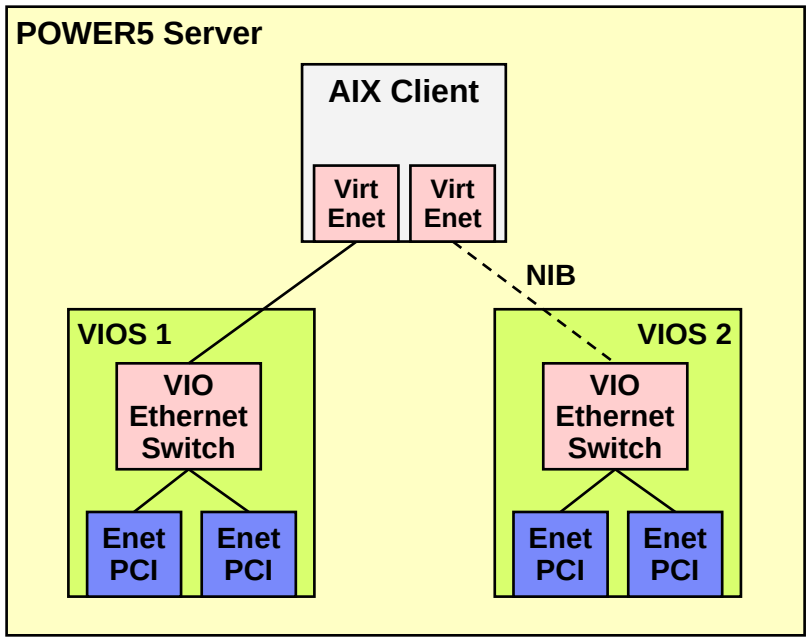
AIX Network Interface Backup (NIB), Dual VIOS



Virtual Ethernet Options

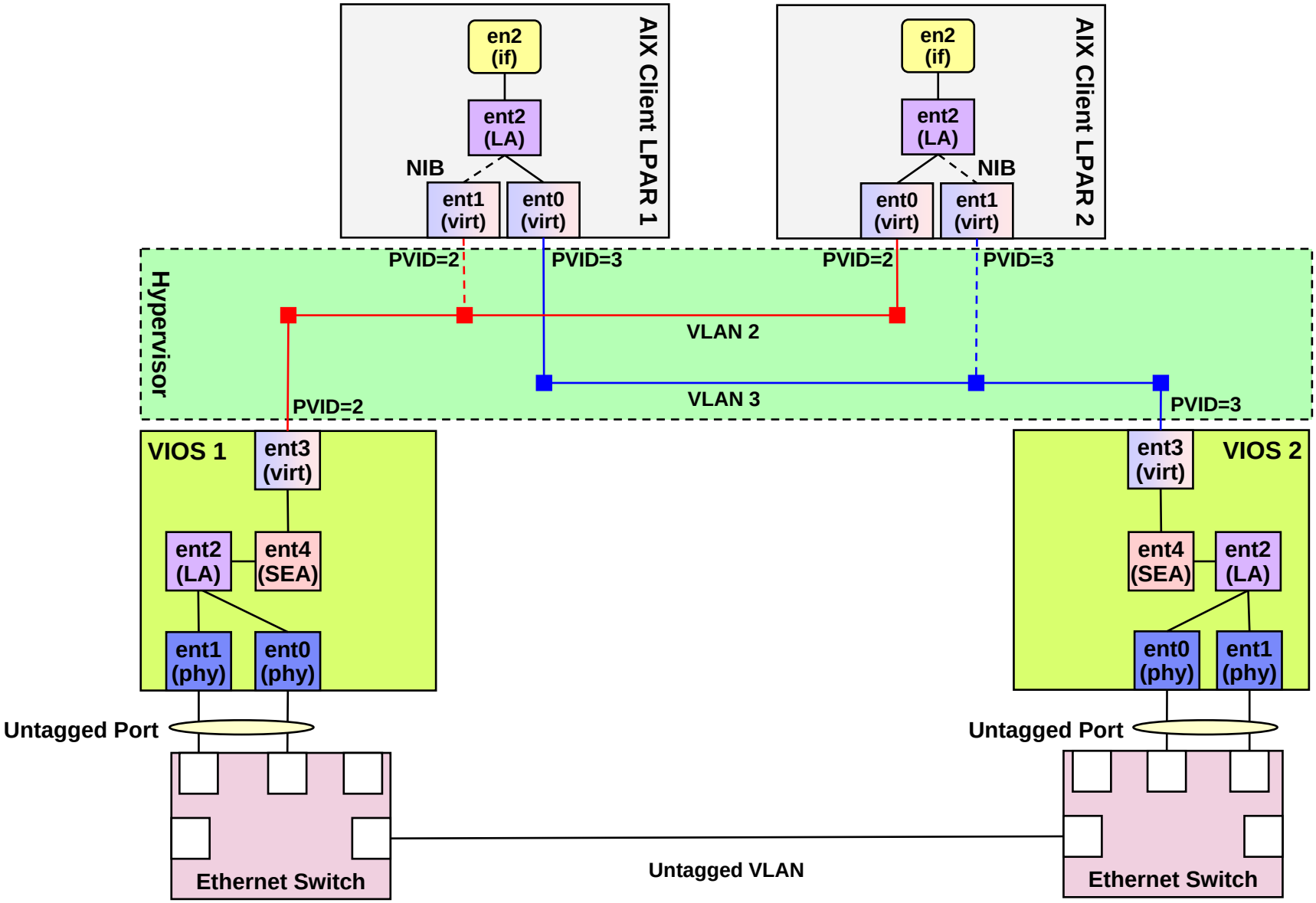
AIX Network Interface Backup (NIB) , Dual VIOS with Link Aggregation (LA)

- **Complexity**
  - Requires specialized setup on client (NIB)
  - Requires link aggregation setup on external switches
  - Needs to ping outside host from the client to initiate NIB failover.
- **Resilience**
  - Protection against single VIOS/ switch port / switch / Ethernet adapter failure
  - Protection against adapter failures within VIOS
- **Throughput / Scalability**
  - Allows each client to use a different primary VIOS sharing network load across multiple VIOS's.
  - Potential for increased bandwidth with LA
- **Notes**
  - NIB does not support tagged VLANs on physical LAN
  - Must use external switches not hubs



Virtual Ethernet Options - Details

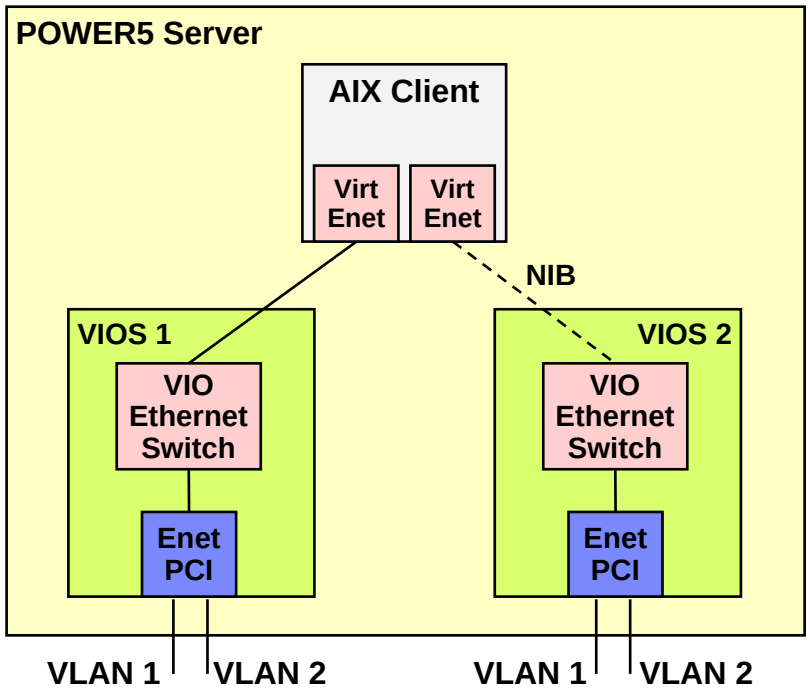
AIX Network Interface Backup (NIB), Dual VIOS with Link Aggregation



Virtual Ethernet Options

AIX Network Interface Backup (NIB), Dual VIOS with VLANs

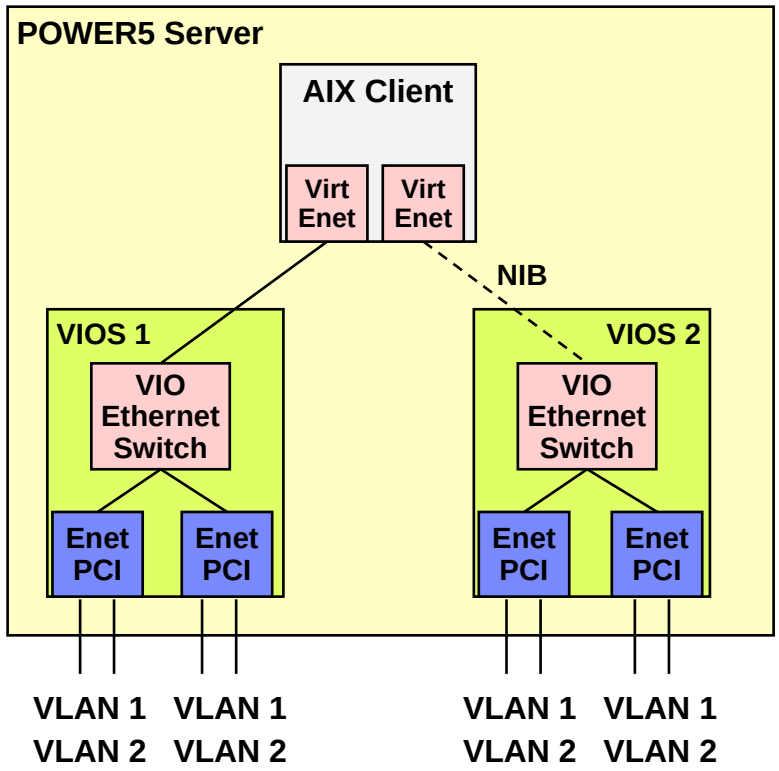
- **Notes**
  - This configuration is **not supported** as all outbound traffic from the VIO server will need to be untagged.



Virtual Ethernet Options

AIX Network Interface Backup, Dual VIOS with VLANs & Link Aggregation

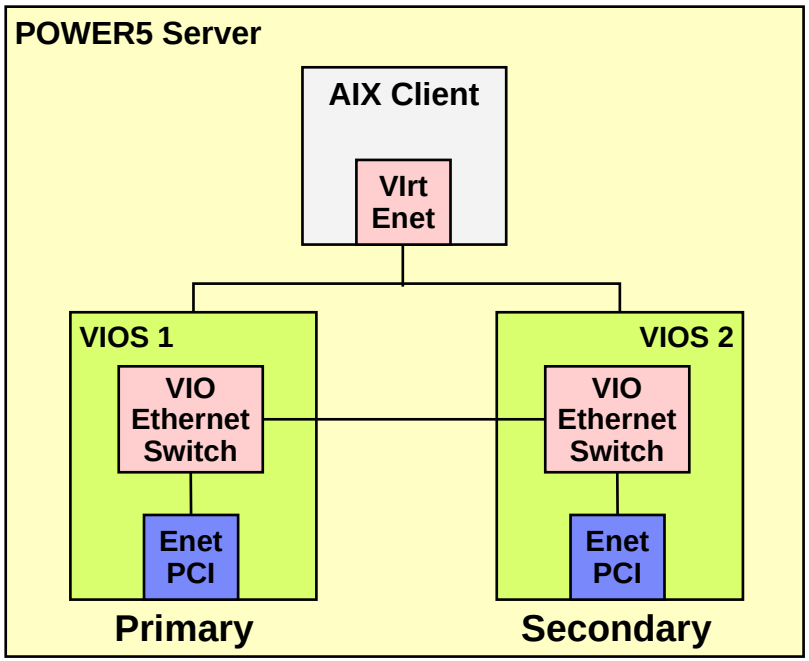
- **Notes**
  - This configuration is **not supported** as all outbound traffic from the VIO server will need to be untagged.



Virtual Ethernet Options

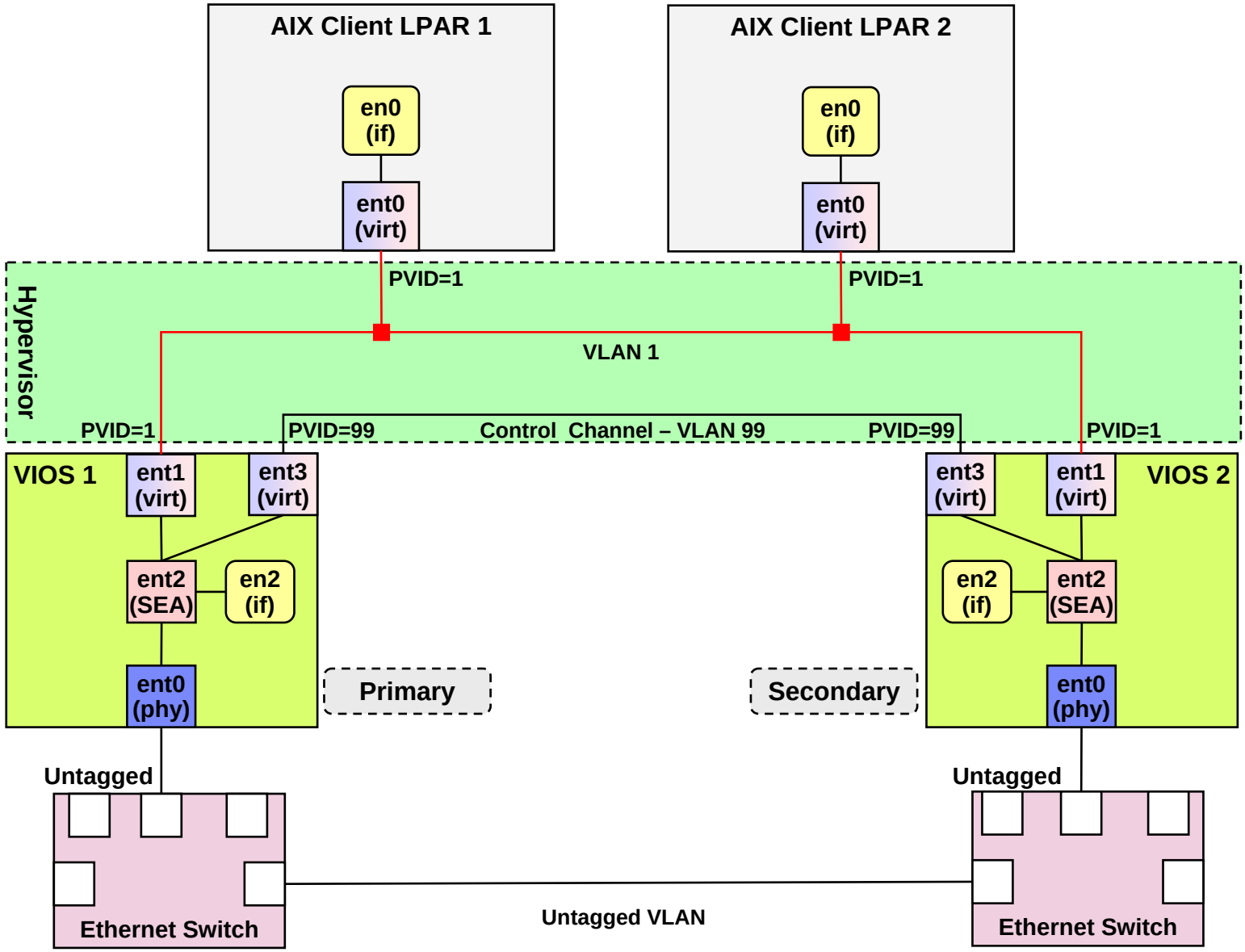
Shared Ethernet Adapter Failover, Dual VIOS

- **Complexity**
  - Specialized setup confined to VIOS
- **Resilience**
  - Protection against single VIOS / switch port / switch / Ethernet adapter failure
- **Throughput / Scalability**
  - Cannot do load-sharing between VIOS's (stand-by SE is idle until needed.
  - SEA failure initiated by:
    - Standby SEA detects the active SEA has failed.
    - Active SEA detects a loss of the physical link
    - Manual failover by putting SEA in standby mode
    - Active SEA cannot ping a given IP address.
- **Notes**
  - Requires VIOS V1.2 and SF235 platform firmware
  - Can be used on any type of client (AIX, Linux)
  - Outside traffic may be tagged



Virtual Ethernet Options - Details

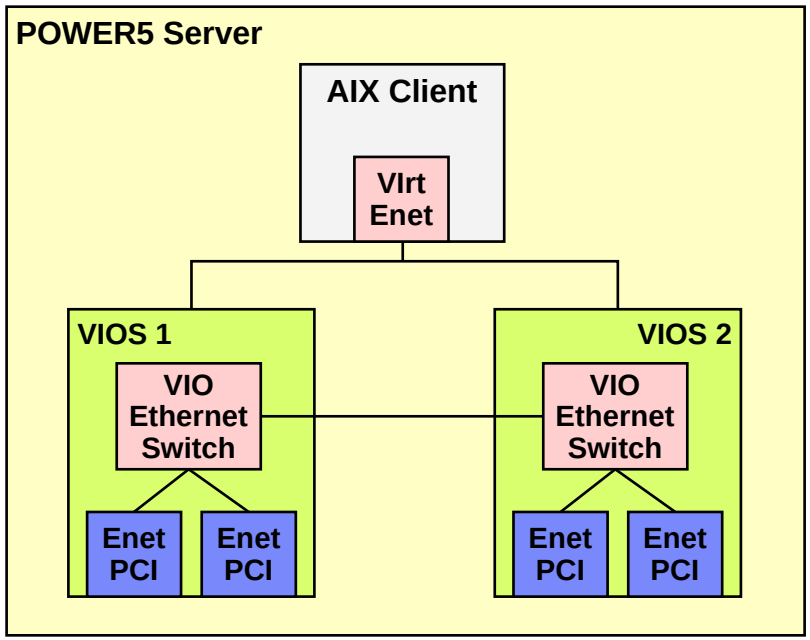
Shared Ethernet Adapter Failover, Dual VIOS



Virtual Ethernet Options

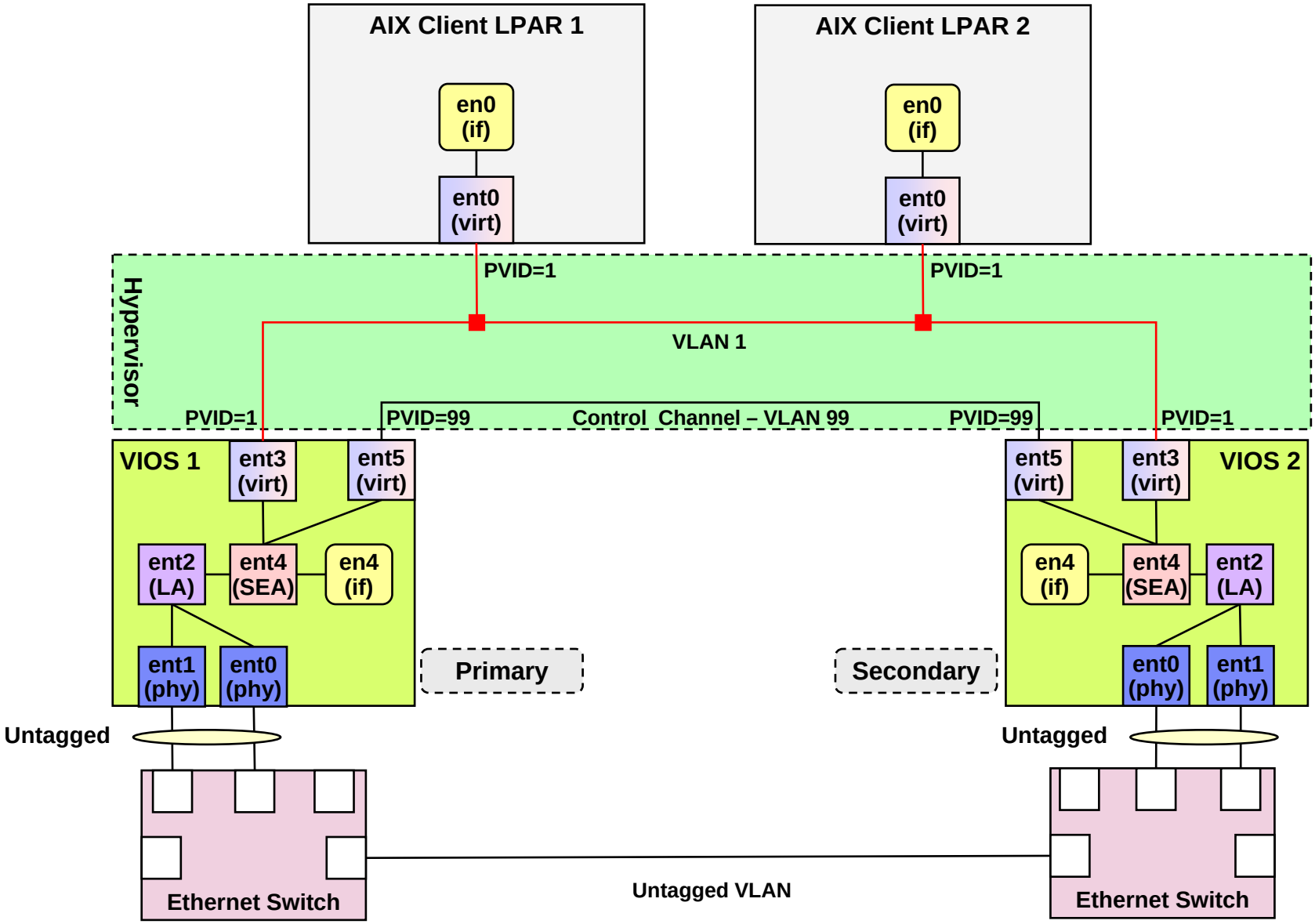
Shared Ethernet Adapter Failover, Dual VIOS with Link Aggregation (LA)

- **Complexity**
  - Specialized setup configured to VIOS's
  - Requires link aggregation setup on external switches
- **Resilience**
  - Protection against single VIOS / switch port / switch / Ethernet adapter failure
  - Protection against Ethernet adapter failures within VIOS
- **Throughput / Scalability**
  - Cannot do load-sharing between VIOS's (stand-by SEA is idle until needed)
  - Potential for increased bandwidth due to LA
- **Notes**
  - Outside traffic may be tagged
  - Requires VIOS V1.2 and SF235 platform firmware



Virtual Ethernet Options - Details

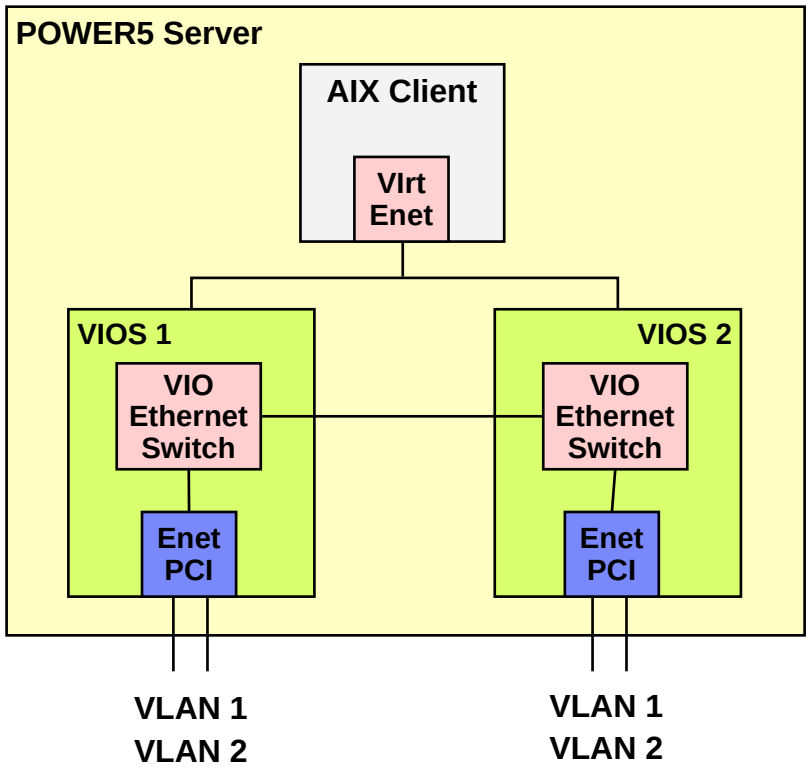
Shared Ethernet Adapter Failover, Dual VIOS with Link Aggregation



Virtual Ethernet Options

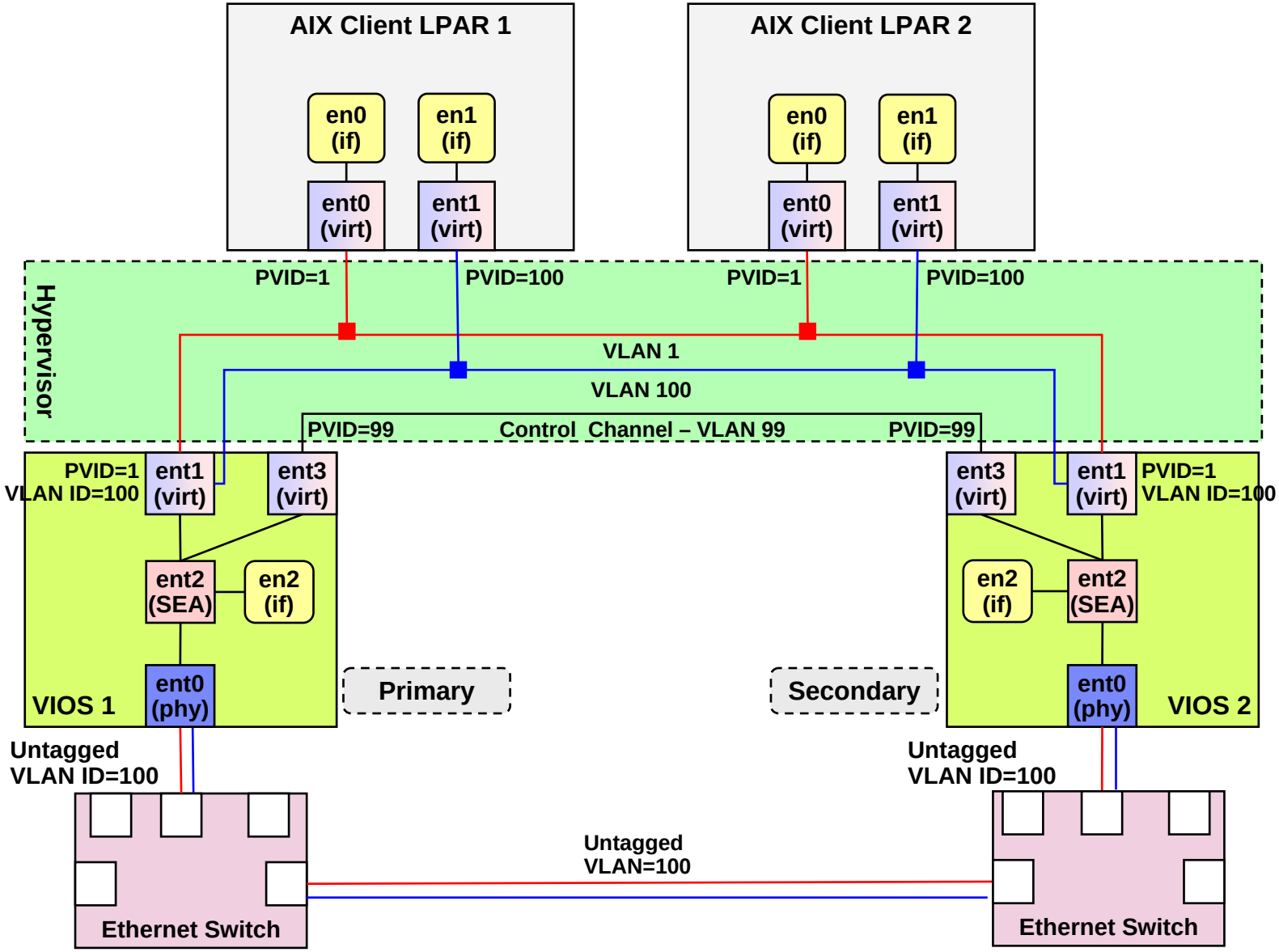
Shared Ethernet Adapter Failover, Dual VIOS with VLANs

- **Complexity**
  - Specialized setup confined to VIOS's
  - Requires VLAN setup of appropriate switch ports
- **Resilience**
  - Protection against single VIOS /switch port / switch / Ethernet adapter failure
- **Throughput / Scalability**
  - Cannot do load-sharing between VIOS's (stand-by SEA is idle until needed)
- **Notes**
  - Tagging allows different VLANs to coexist within same CEC
  - Requires VIOS V1.2 and SF235 platform firmware



Virtual Ethernet Options - Details

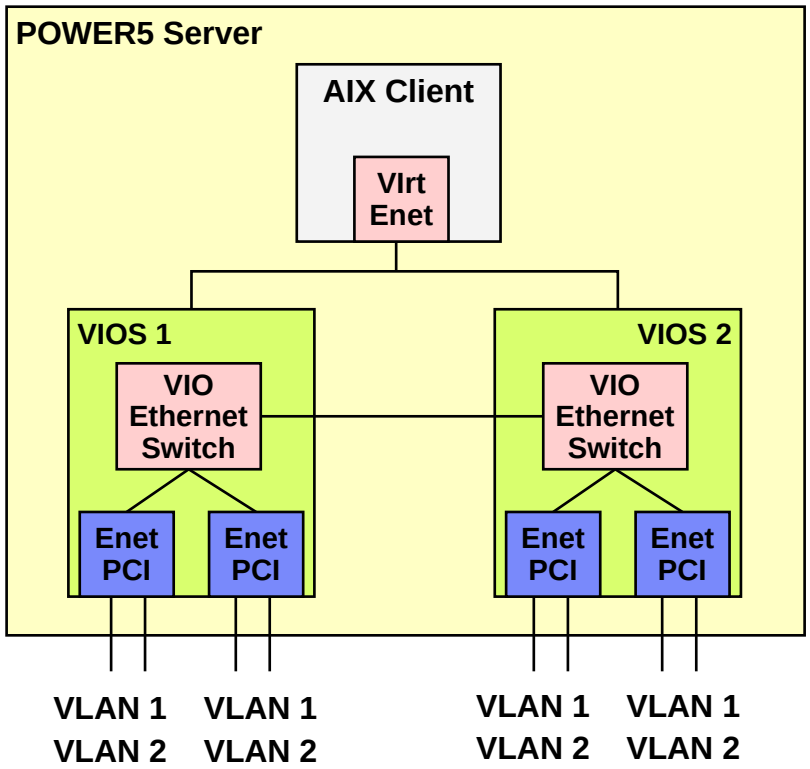
Shared Ethernet Adapter Failover, Dual VIOS with VLANs



Virtual Ethernet Options

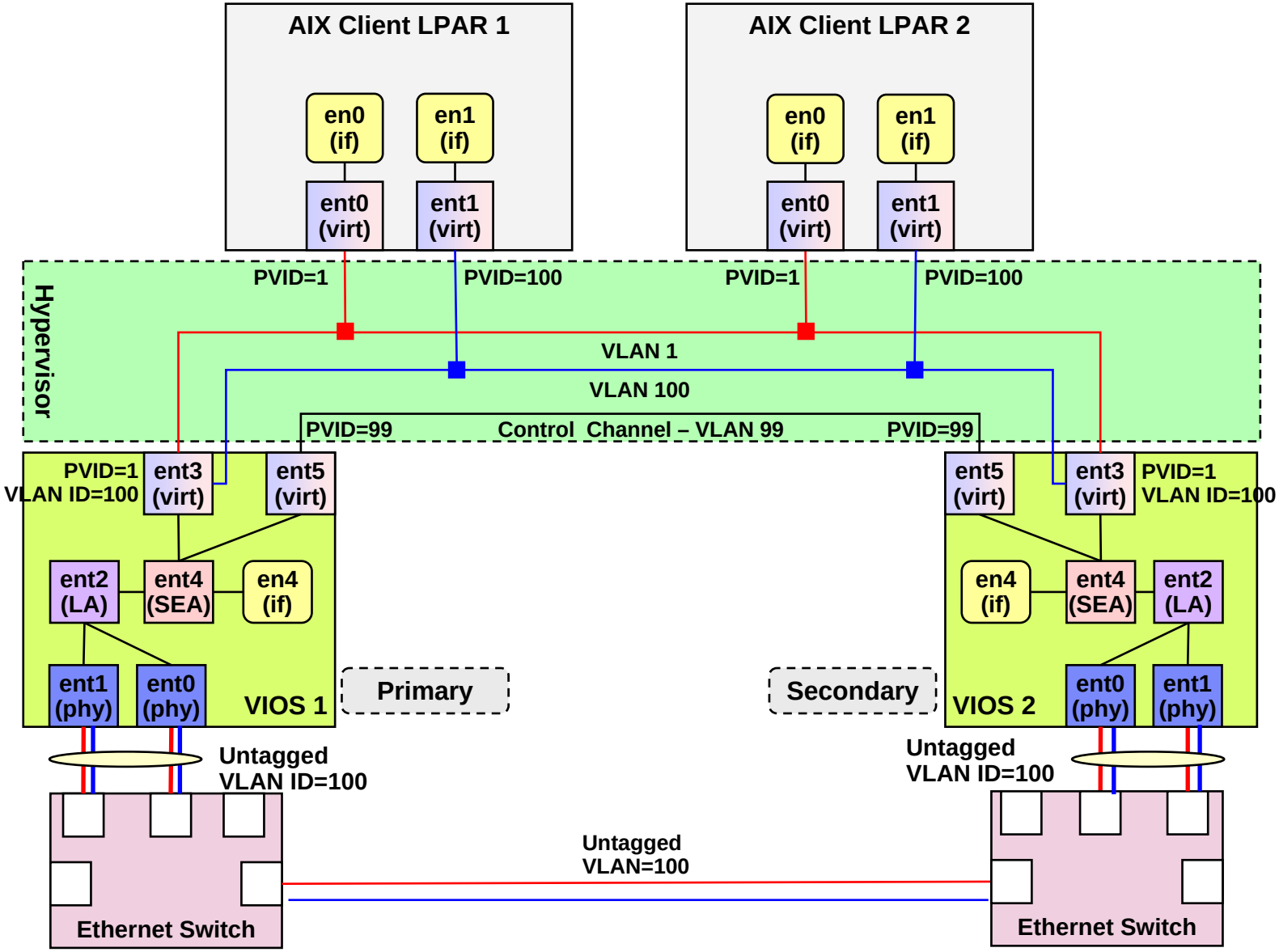
Shared Ethernet Adapter Failover, Dual VIOS with VLANs and Link Aggregation

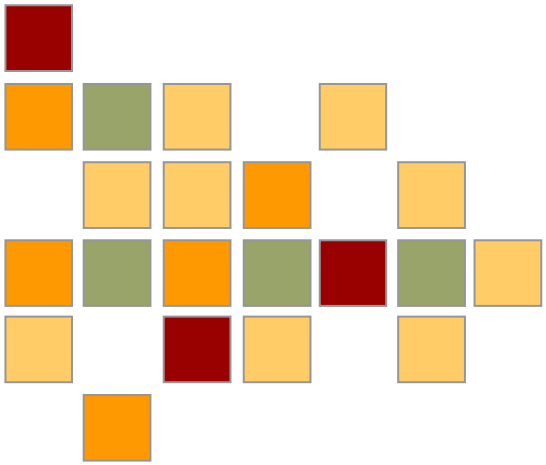
- **Complexity**
  - Specialized setup confined to VIOS's
  - Requests Link Aggregation setup of appropriate switch ports.
  - Requires VLAN setup of appropriate switch ports
- **Resilience**
  - Protection against single VIOS / switch port / switch / Ethernet adapter failure
  - Protection against adapter failures with VIOS
- **Throughput / Scalability**
  - Potential for increased bandwidth due to Link Aggregation
  - Cannot do load-sharing between VIOS's (stand-by SEA is idle until needed)
- **Notes**
  - Tagging allows different VLANs to coexist within same CEC
  - Requires VIOS V1.2 and SF235 platform firmware



Virtual Ethernet Options - Details

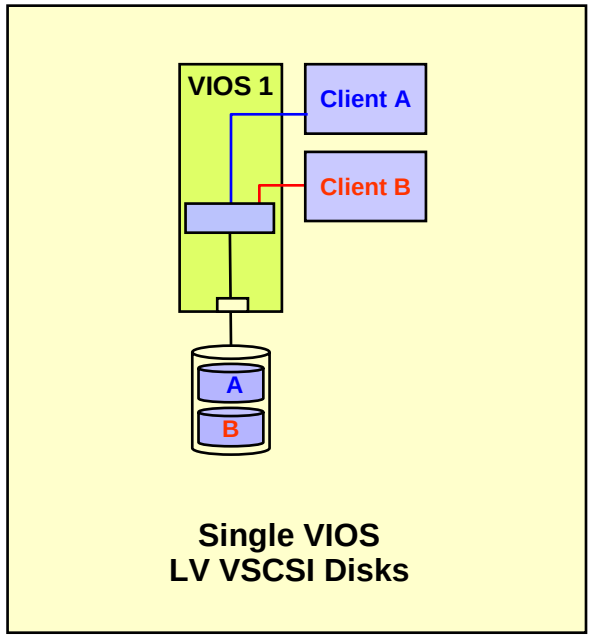
Shared Ethernet Adapter Failover, Dual VIOS with VLANs and Link Aggregation



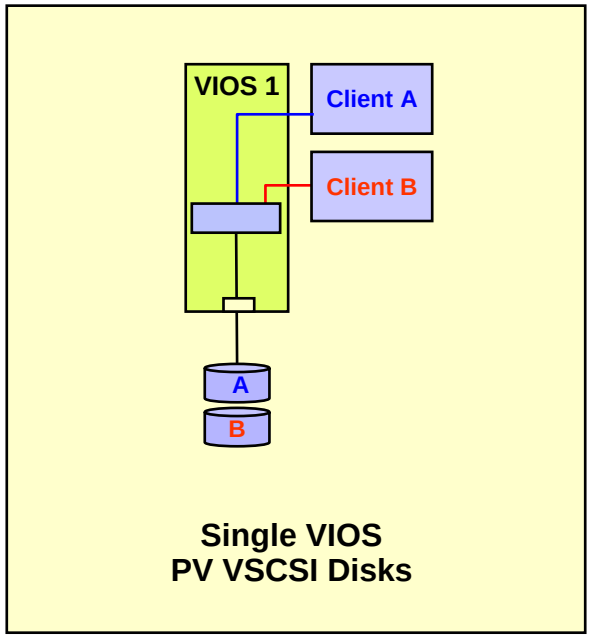


Virtual SCSI

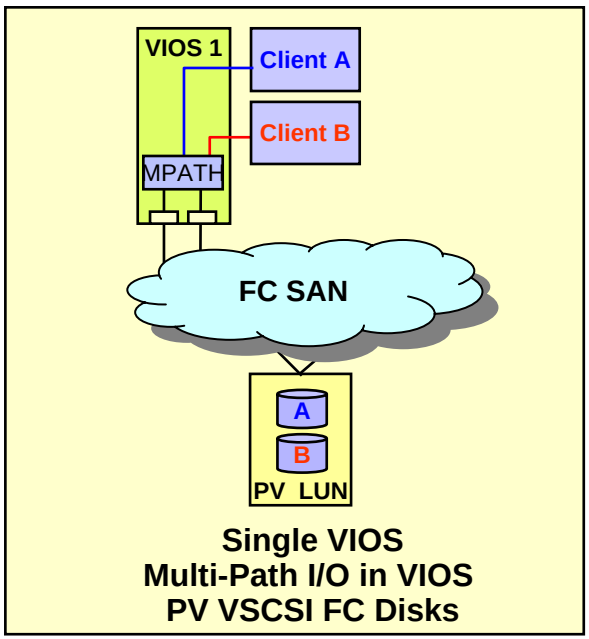
Virtual SCSI Single VIOS Options in this Document



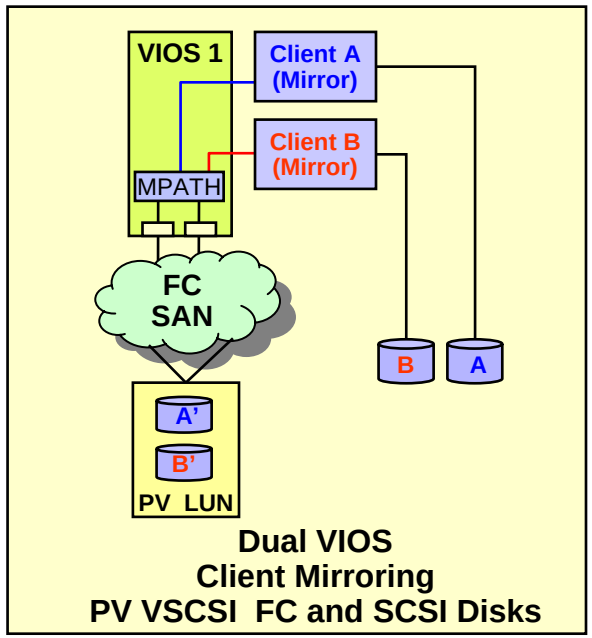
Single VIOS
LV VSCSI Disks



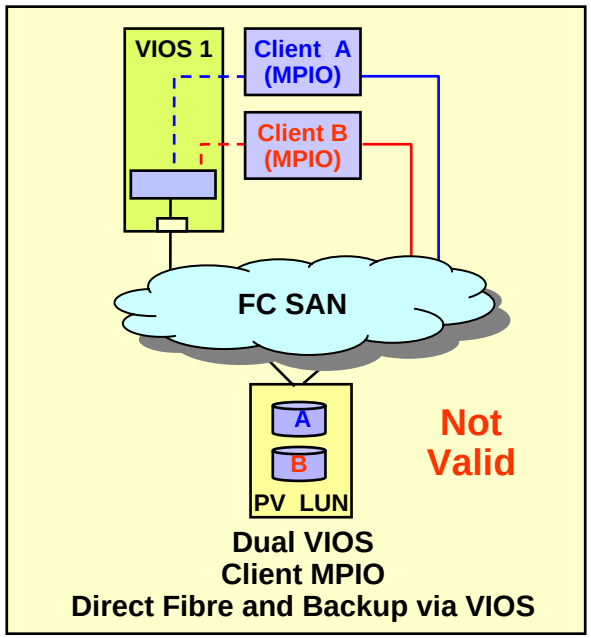
Single VIOS
PV VSCSI Disks



Single VIOS
Multi-Path I/O in VIOS
PV VSCSI FC Disks

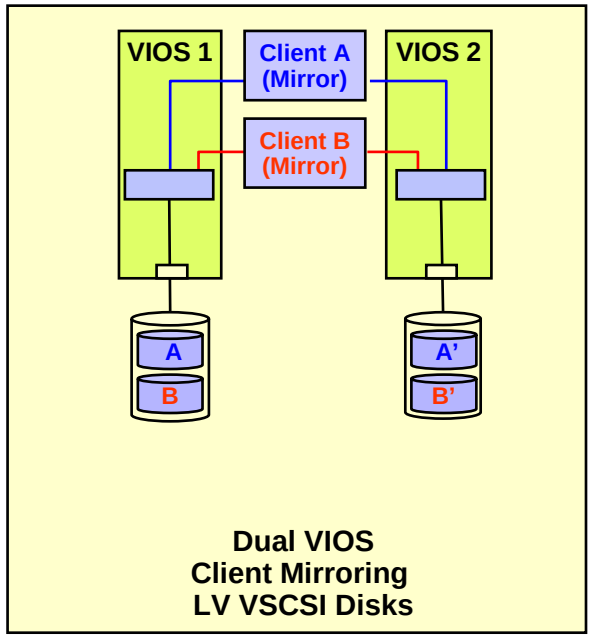


Dual VIOS
Client Mirroring
PV VSCSI FC and SCSI Disks

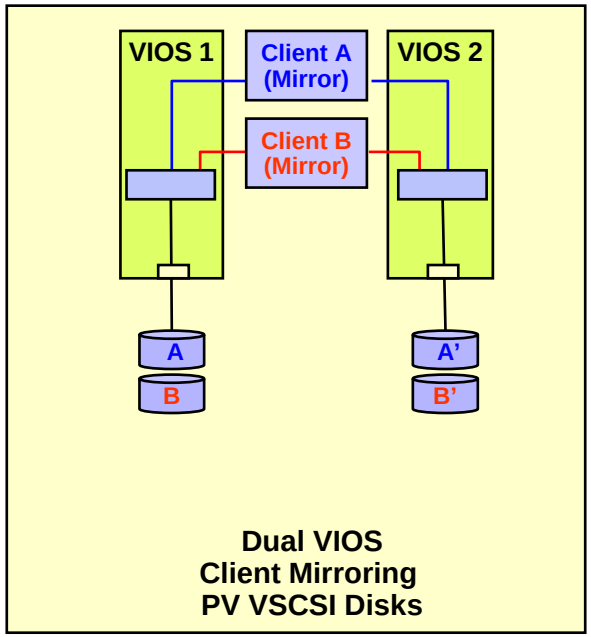


Dual VIOS
Client MPIO
Direct Fibre and Backup via VIOS

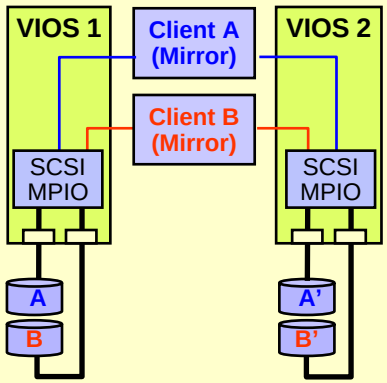
Virtual SCSI Dual VIOS Client Mirror Options in this Document



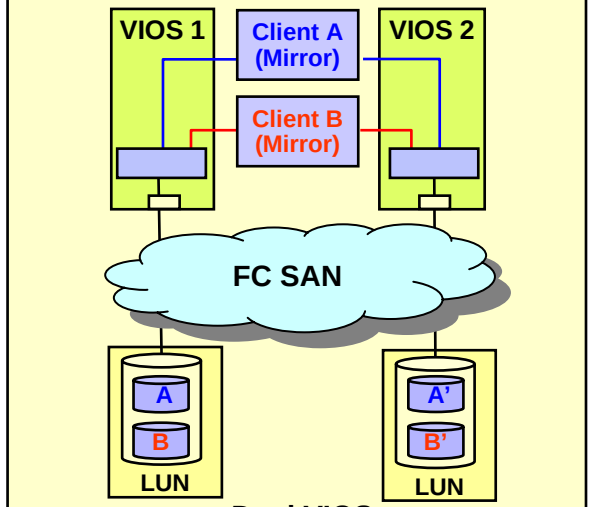
Dual VIOS
Client Mirroring
LV VSCSI Disks



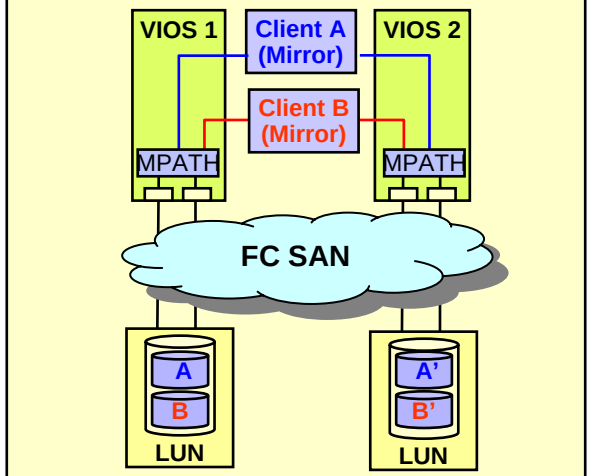
Dual VIOS
Client Mirroring
PV VSCSI Disks



Dual VIOS
Client Mirroring , SCSI MPIO
PV VSCSI Disks

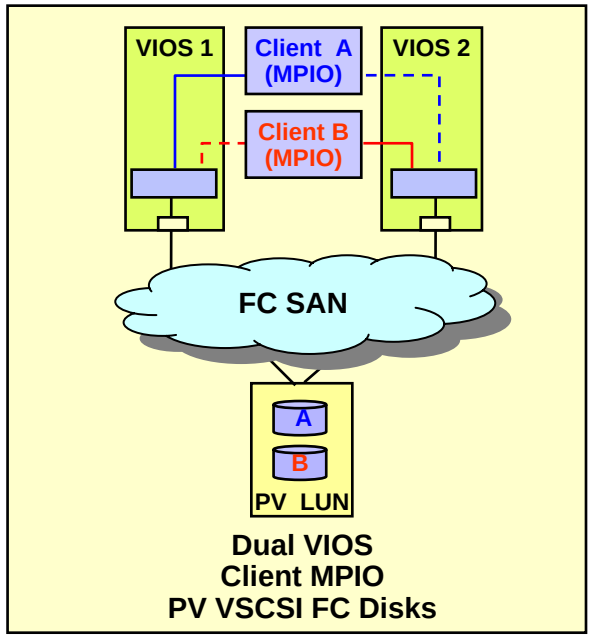


Dual VIOS
Client Mirroring
LV or PV VSCSI FC Disks

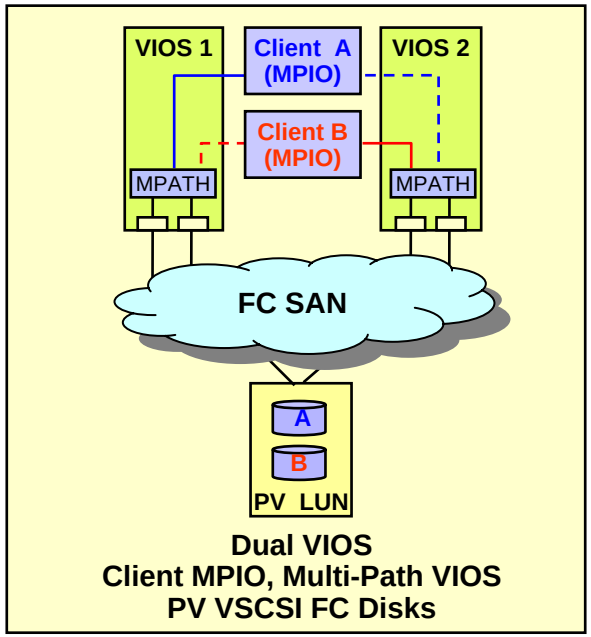


Dual VIOS
Client Mirror, Multi-Path VIOS
LV or PV VSCSI FC Disks

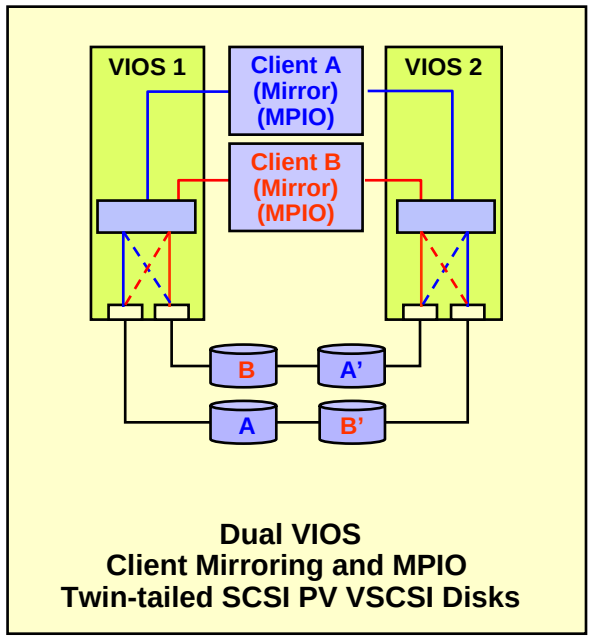
Virtual SCSI Dual VIOS Client MPIO Options in this Document



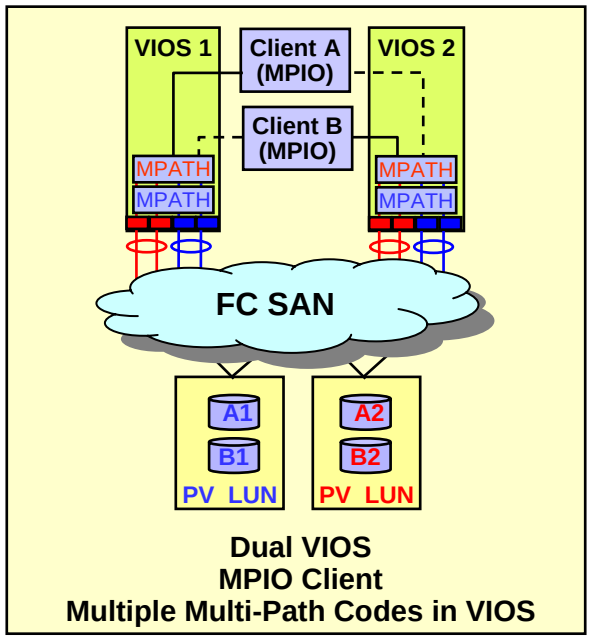
Dual VIOS
Client MPIO
PV VSCSI FC Disks



Dual VIOS
Client MPIO, Multi-Path VIOS
PV VSCSI FC Disks



Dual VIOS
Client Mirroring and MPIO
Twin-tailed SCSI PV VSCSI Disks



Dual VIOS
MPIO Client
Multiple Multi-Path Codes in VIOS

Cycles per I/O Operation Comparison by Block Size for a 1.65 GHz POWER5 Processor

Number of CPU cycles per I/O operation for	4 KB	8 KB	32 KB	64 KB	128 KB
Phys disk	45000	47000	58000	81000	120000
LVM	49000	51000	59000	74000	105000

These numbers assume a 1.65 GHz CPU. To adjust for different processor speeds, determine the ratio of the 1.65 GHz processor to the one you're going to use and multiple the number of cycles by that number. For example, using a 1.5 GHz processor, you would divide 1.65 by 1.5 for a ratio of 1.1. Therefore, a 32 KB operation on physical disk would require 58,000 x 1.1, or 63,800 cycles on a 1.5 GHz processor.

That number times the number of peak I/O operations per second, divided by the processor speed, would give you the amount of entitlement you need. For example, 63,800 x 7000 I/O operations = .297, or .30 processors. For a dedicated partition, your entitlement would be 1.0. For a micro-partition, you'd only need an entitlement of 0.3 for this particular workload.

VSCSI Sizing and Performance Notes

- **Sizing based on the I/O configuration**
  - Number of disks x I/O operations per second x CPU cycles per operation (adjusted for CPU speed) , divided by the speed of the processor
  - A reasonable rule of thumb would assume 100-200 I/Os per second per disk. Use 150 IOPS for boot disks and 8K block size if disk numbers are unknown
  - Example Assume 8K Block Size and PV Disks: 47,000 Cycles/Operation
    1.9 GHz Power5
    2 Disks @ 150 IO/s each

    2 Disks * 150 IO/s * 47,000 Cycles/IO * 1.65/1.9
    (CPU Speed Adjustment) / 1,900,000,000 (Speed) = 0.01 CPUs
- **VSCSI Performance**
  - I/O latency will depend upon a system's utilization and configuration.
  - VSCSI CPU overhead is small and relatively linear as throughout increases
  - Since there is no data caching on a Virtual I/O Server, 1 GB is often enough memory for the VIO Server partition
  - Use SMT unless your application requires it be turned off

Virtual SCSI General Notes

- **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
  - Logical volumes within the VIOS that are exported as virtual SCSI devices may 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 Notes….

- **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)

Virtual SCSI General Notes….

- **There are different methods to identify uniquely a disk for use in Virtual I/O Server (VIOS), such as:**
  - Unique device identifier (UDID)
  - IEEE volume identifier
  - Physical volume identifier (PVID)
- **Discussion**
  - Each of these methods may result in different data formats on the disk. The preferred disk identification method for virtual disks is the use of UDIDs. MPIO uses the UDID method.
  - Most non-MPIO disk storage multi-pathing software products use the PVID method instead of the UDID method. Because of the different data format associated with the PVID method, customers with non-MPIO environments should be aware that certain future actions performed in the VIOS LPAR may require data migration, that is, some type of backup and restore of the attached disks.

Virtual SCSI Legend

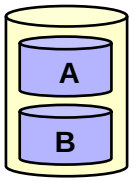
Physical Adapter

Failover

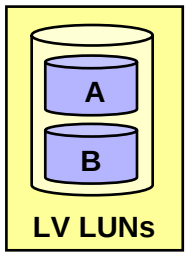
Active

VIOS VIO Server

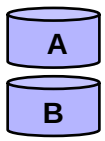
VIOC VIO Client



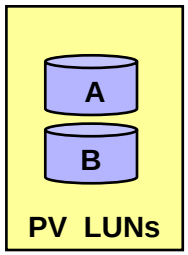
LV VSCSI Disks
Use logical SCSI volumes on the VIOS for VIOC disk



LV VSCSI FC Disks
Use logical volume FC LUNS on the VIOS for VIOC disk



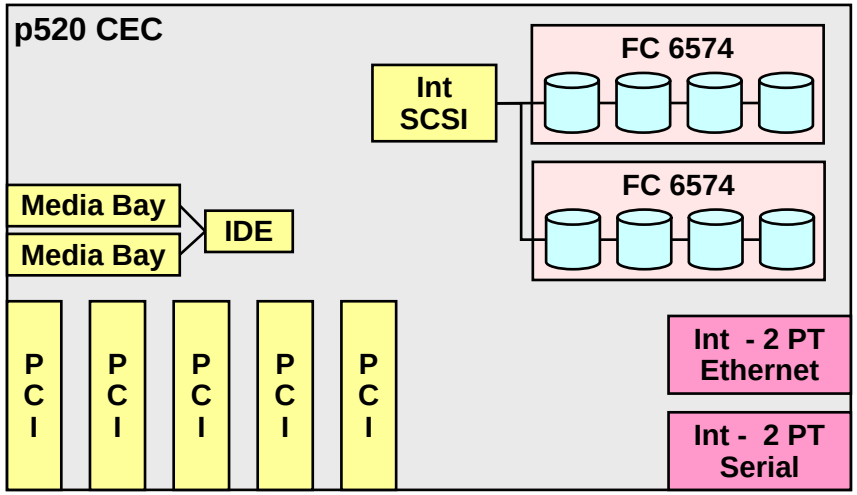
PV VSCSI Disks
Use physical SCSI volumes on the VIOS for VIOC disk

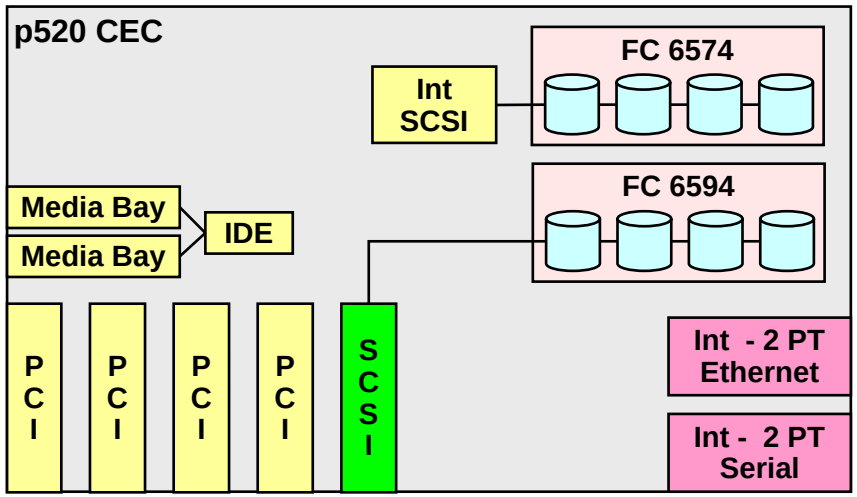


PV VSCSI FC Disks
Use physical volume FC LUNS on the VIOS for VIOC disk

p520 / System p5 p520 SCSI Options

- **p520 SCSI Disk Options**
  - p520 SCSI adapter can be ordered in a RAID or non-RAID configuration
  - The single integrated SCSI controller can attach one or two four packs of disk in a single assignable string using two four pack feature code 6574s.
  - A second independently assignable string of disk (RAID or non-RAID) can be attached if a second four pack is ordered as a feature code 6594

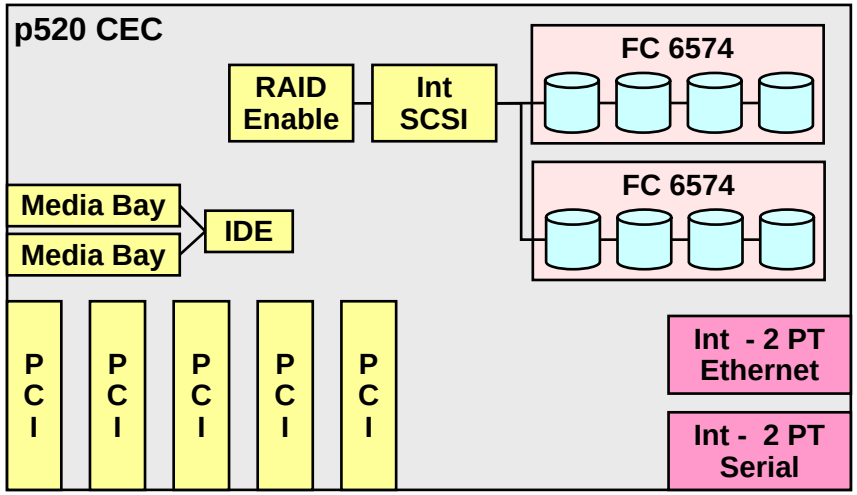


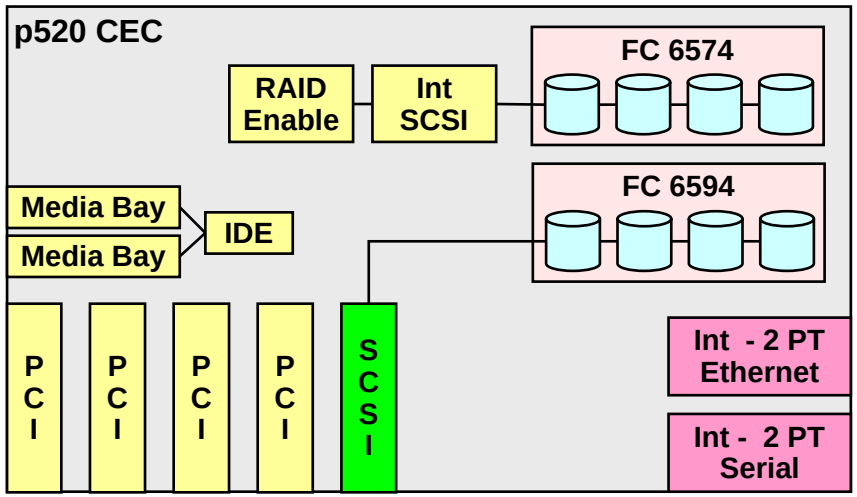


Note: The PCI SCSI adapter can be either RAID or non- RAID

p520 / System p5 p520 SCSI Options

- **p520 SCSI Disk Options…**
  - With the addition of a RAID enablement feature, the two internal four-packs can become a single RAID array
  - Optionally a second RAID (or non-RAID) adapter can be added to form a second RAID array (or JBOD string)

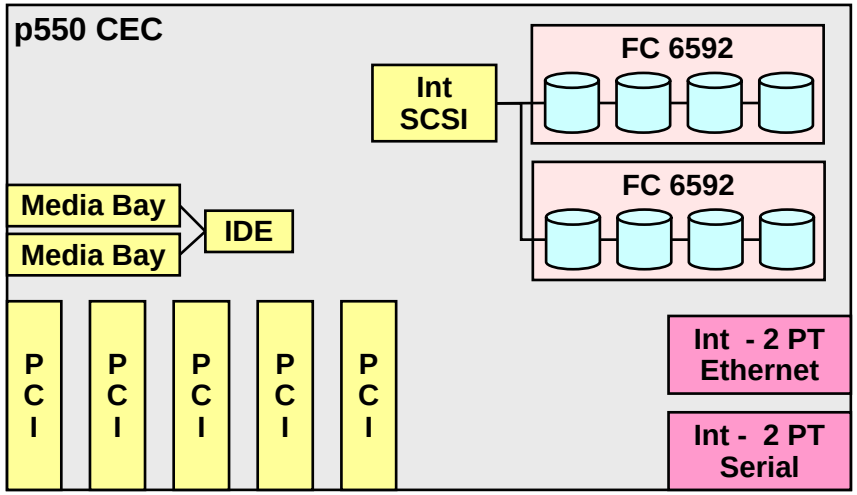


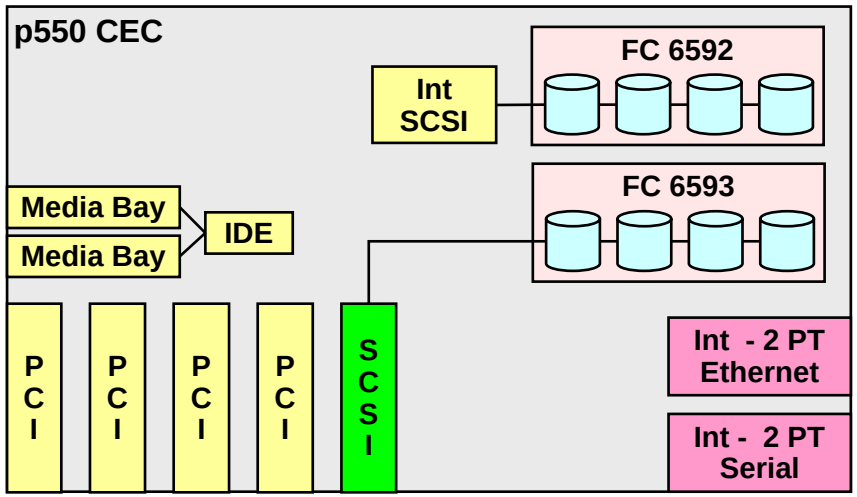


Note: The PCI SCSI adapter can be either RAID or non- RAID

p550 / System p5 550 / System p5 550Q SCSI Options

- **p550 SCSI Disk Options**
  - The p550 SCSI adapter can be ordered in a RAID or non-RAID configuration
  - The single integrated SCSI controller can attach to one or two independent four packs of disk (FC 6592).
  - A second independent string of disk (FC 6593) can be attached via a RAID or non-RAID adapter

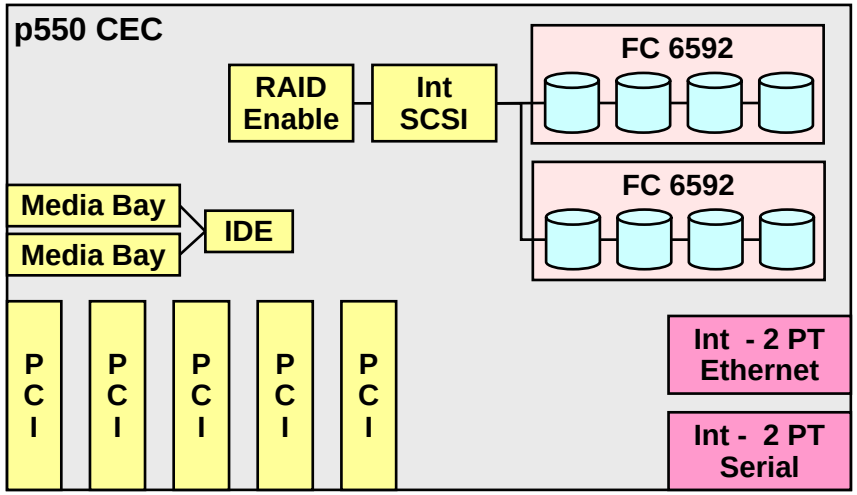


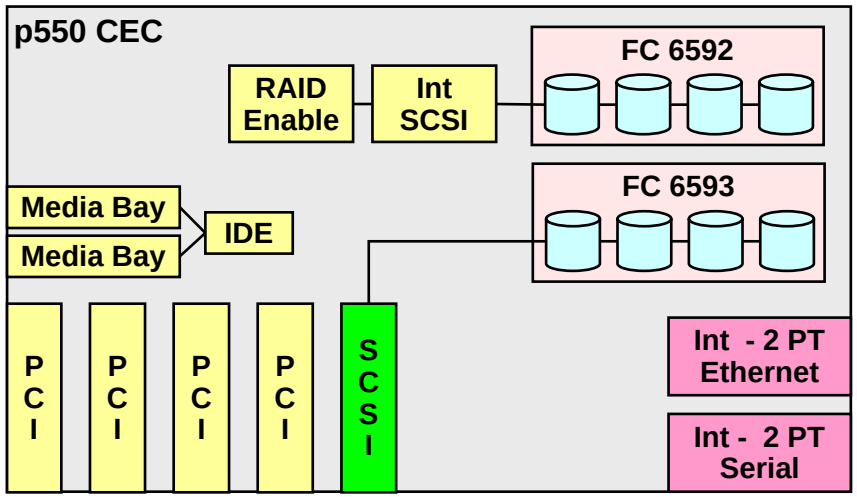


Note: The PCI SCSI adapter can be either RAID or non- RAID

p550 / System p5 550 / System p5 550Q SCSI Options

- **p550 SCSI Disk Options…**
  - With the addition of a RAID enablement feature, the two internal four-packs (FC 6592) can become a single RAID array
  - Optionally, a second RAID or non-RAID adapter can be added to form a second RAID array (or JBOD string) using FC 6593 four packs

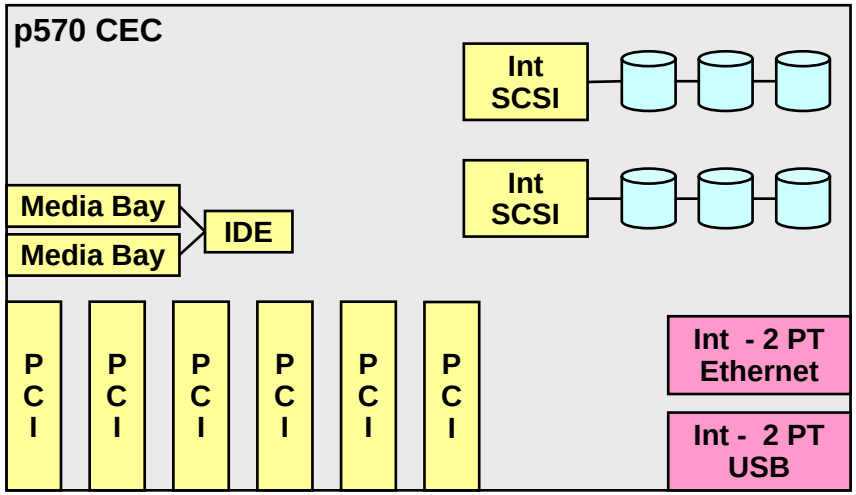


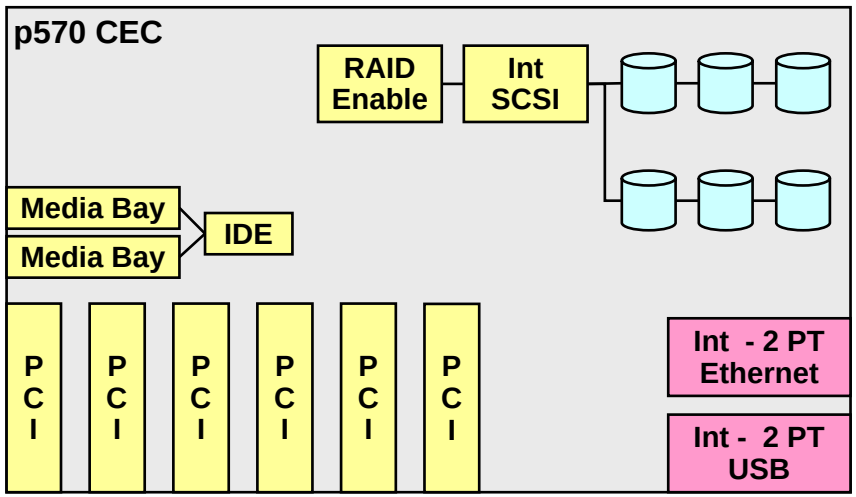


Note: The PCI SCSI adapter can be either RAID or non- RAID

p570 SCSI Options

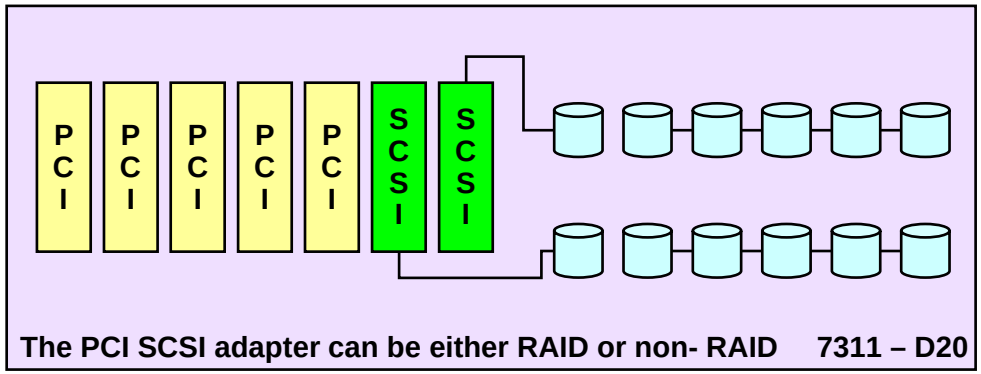
- **p570 SCSI Disk Options**
  - p570 internal disk can be configured as two separate strings of three disks or one string of RAID disk

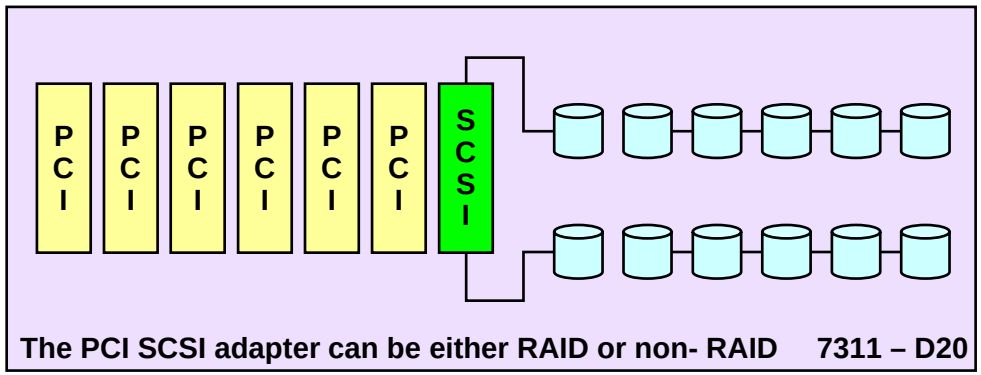




7311-D20 I/O Drawer

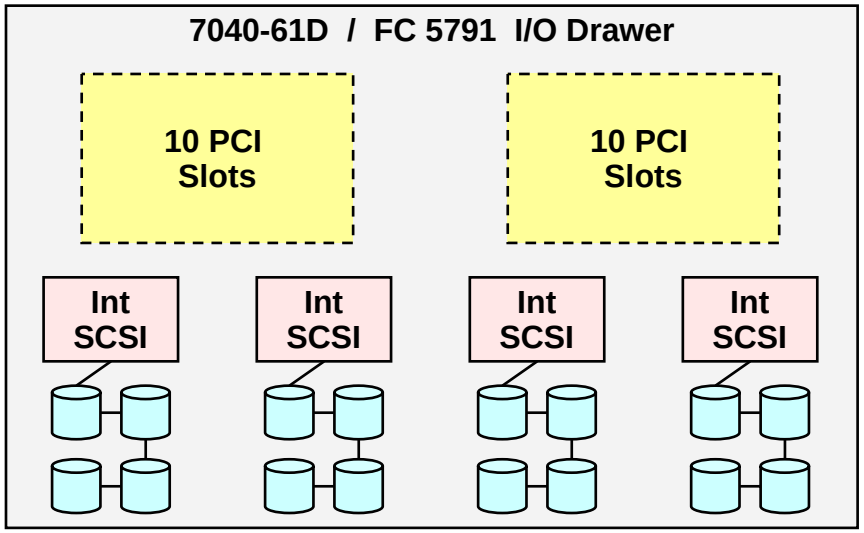
- **7311-D20 I/O Drawer**
  - The 7311-D20 drawer can be used with the p570, p550, and p520
  - The drawer can be configured with one or two independent strings of JBOD or RAID disks
  - The two six packs can be individually assigned to an LPAR or combined to form one logical JBOD or RAID string.
  - The SCSI adapter require a PCI slot

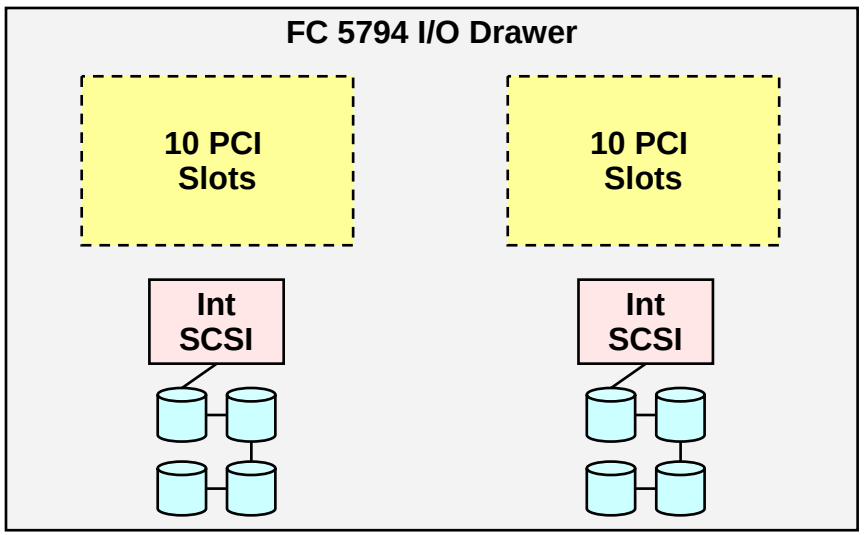




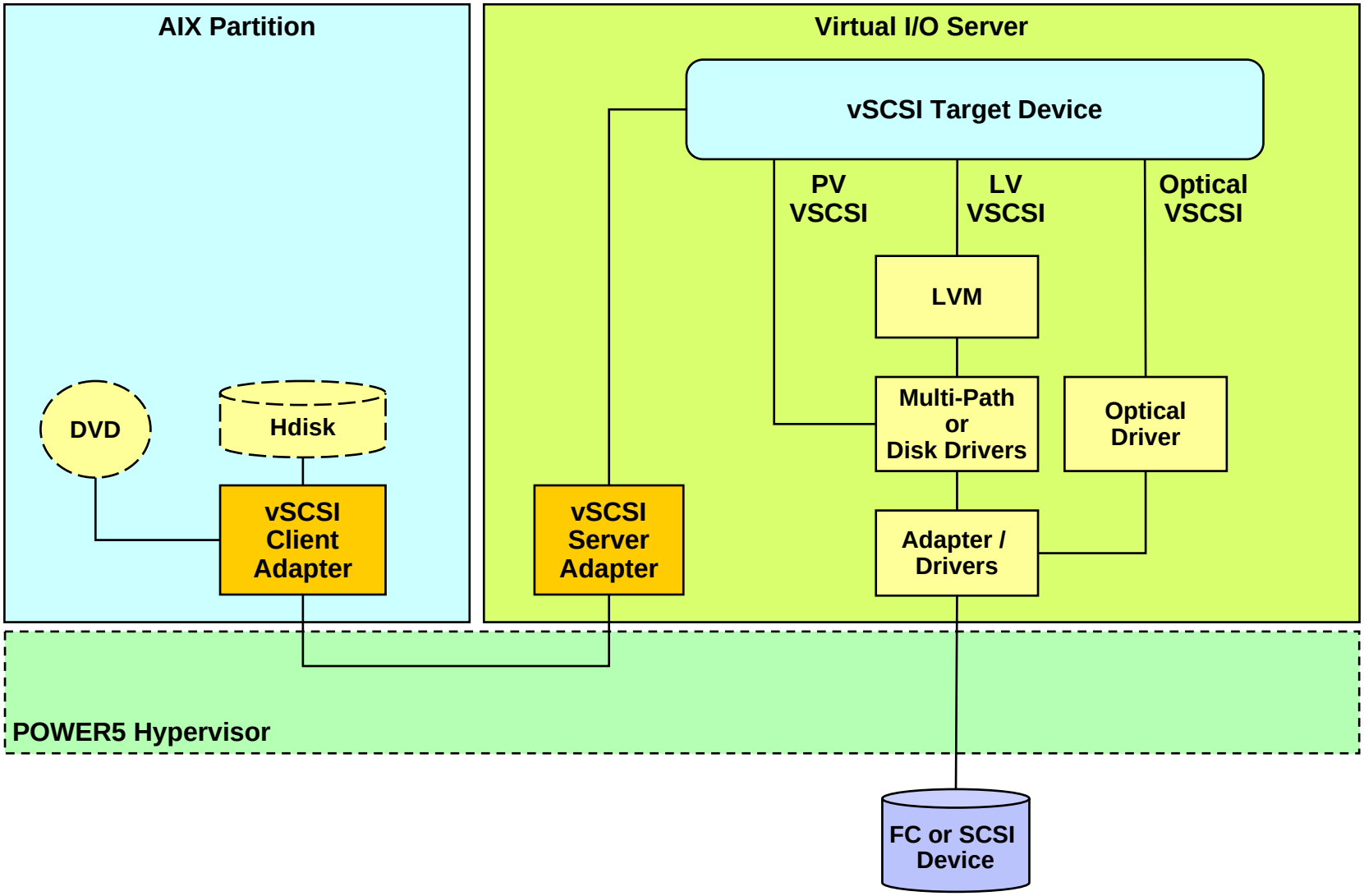
p590/p595/p575 I/O Drawers

- **7040-61D / FC 5791 I/O Drawer**
  - The drawer can be used with the p575, p590, or p595 servers
  - The drawer contains four integrated non-RAID SCSI controllers each controlling up to four disks.
- **FC 5794 I/O Drawer**
  - The drawer can be used with the p575, p590, or p595 servers
  - The drawer contains two integrated non-RAID SCSI controllers each controlling up to four disks.





Virtual SCSI Basic Architecture



VIOS Multi-Path Options

VIOS Multi-Path Driver	VIO Configuration	Storage System Family (Note 1)	Multi-Path Algorithm
MPIO Default PCM	Single or Dual	EMC (Not CLARiiON)	Failover, Round-Robin
MPIO Default PCM	Single or Dual	HDS	Failover, Round-Robin
MPIO Default PCM	Single or Dual	HP (Not EVA, some members of XP family)	Failover, Round-Robin
MPIO Default PCM	Single or Dual	Virtual SCSI Devices	Failover Only
MPIO SDDPCM	Single or Dual	IBM ESS, DS800, DS6000	Failover, Round Robin, Load Balancing
RDAC	Single or Dual	IBM DS4000, IBM FastT	Failover Only
PowerPath	Single or Dual	EMC	Basic Failover, Round Robin, Least I/Os, Least Blocks, No Redirect, Adaptive, Request, SymOpt, ClarOpt
AutoPath	Single or Dual	HP (Not EVA, some members of XP family	Failover, Round-Robin, exRound-Robin
HDLM	Single or Dual	HDS	Failover, Round-Robin, exRound-Robin
SDD	Single	IBM ESS, DS800, DS6000, SVC	Failover, Round Robin, Load Balancing

Notes:

1. See vendor documentation for specific supported models, microcode requirements, AIX levels, etc.
2. Not all multi-path codes are compatible with one another. SDD and RDAC are not supported with PowerPath. MPIO *compliant* code is not supported with non MPIO *compliant* code for similar disk subsystems (e.g., one cannot use MPIO for one EMC disk subsystem and PowerPath for another on the same *VIOS*, *nor* can one use SDD and SDDPCM on the same VIOS).

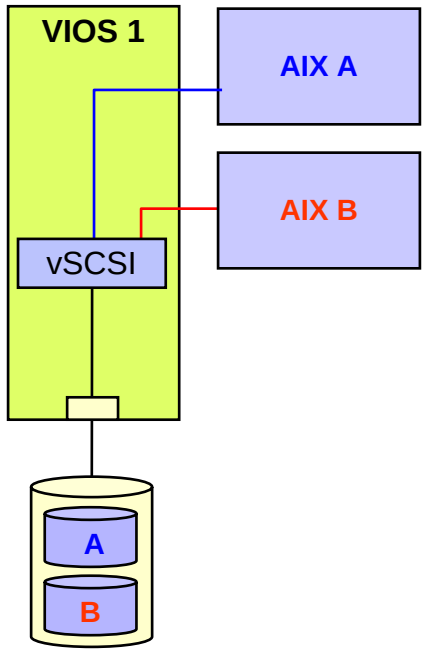
   Separate sets of FC adapters are required when using different multi-path codes on a VIO. If incompatible multi-path codes are required, then one should use separate VIOSs for the incompatible multi-path codes. In general, multi-path codes that adhere to the MPIO architecture are compatible.

3. IBM VIO support options: **http://techsupport.services.ibm.com/server/vios/documentation/datasheet.html**

Virtual SCSI Options

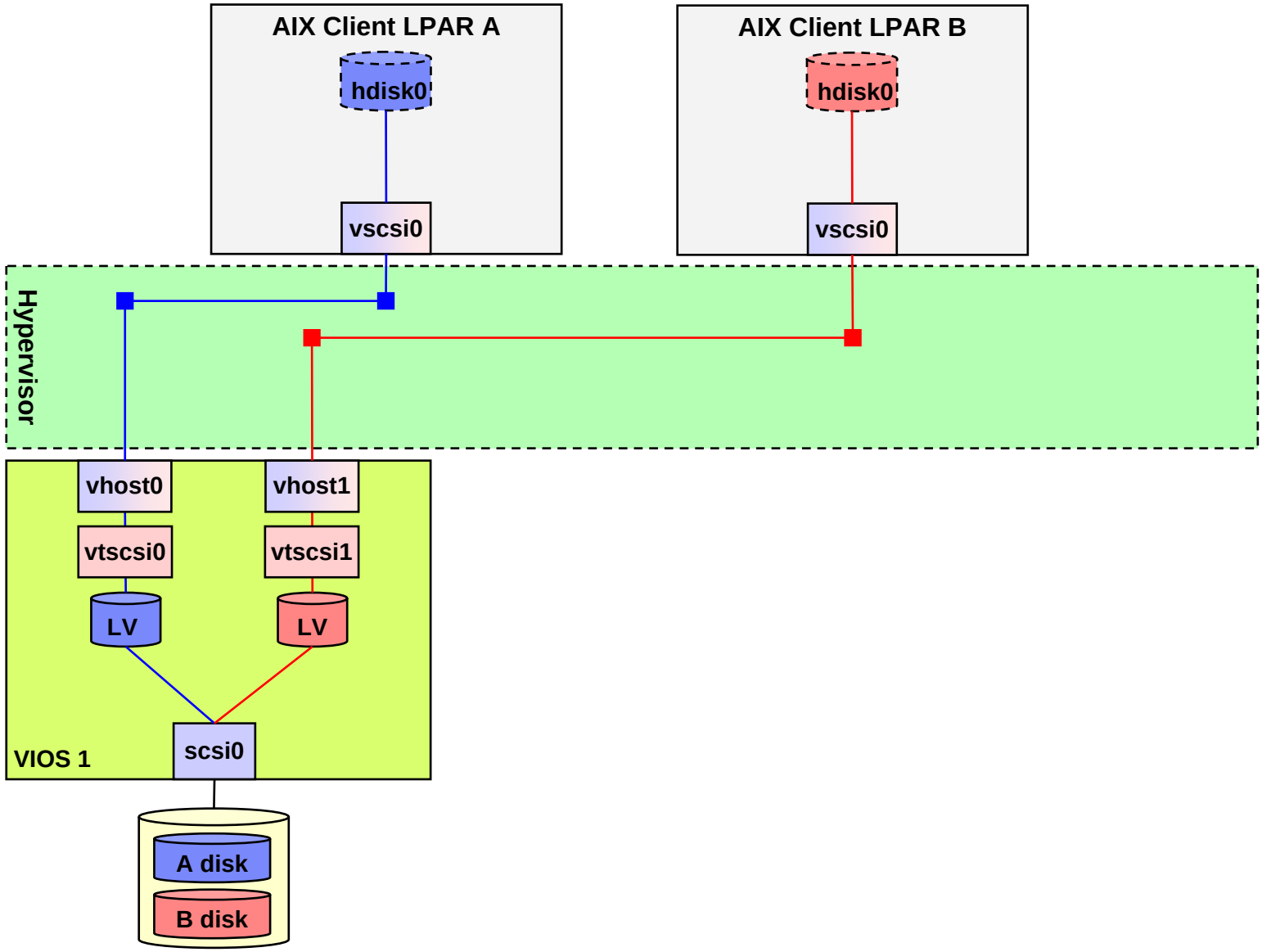
Single VIOS, LV VSCSI Disks

- **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 physical disk may impact more than one client
- **Throughput / Scalability**
  - Performance limited by single SCSI adapter and internal SCSI disks.
- **Notes**
  - Low cost disk alternative



Virtual SCSI Options - Details

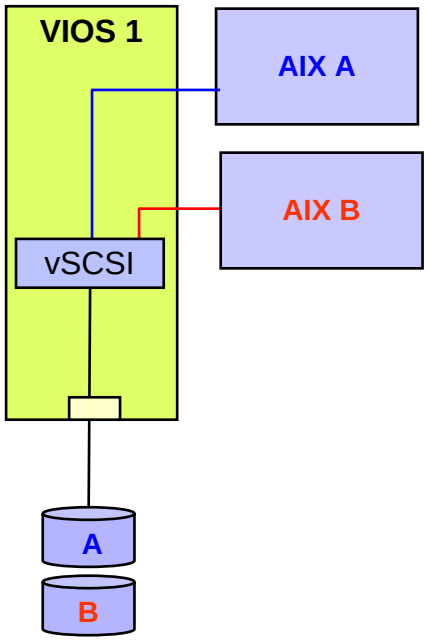
Single VIOS, LV VSCSI Disks



Virtual SCSI Options

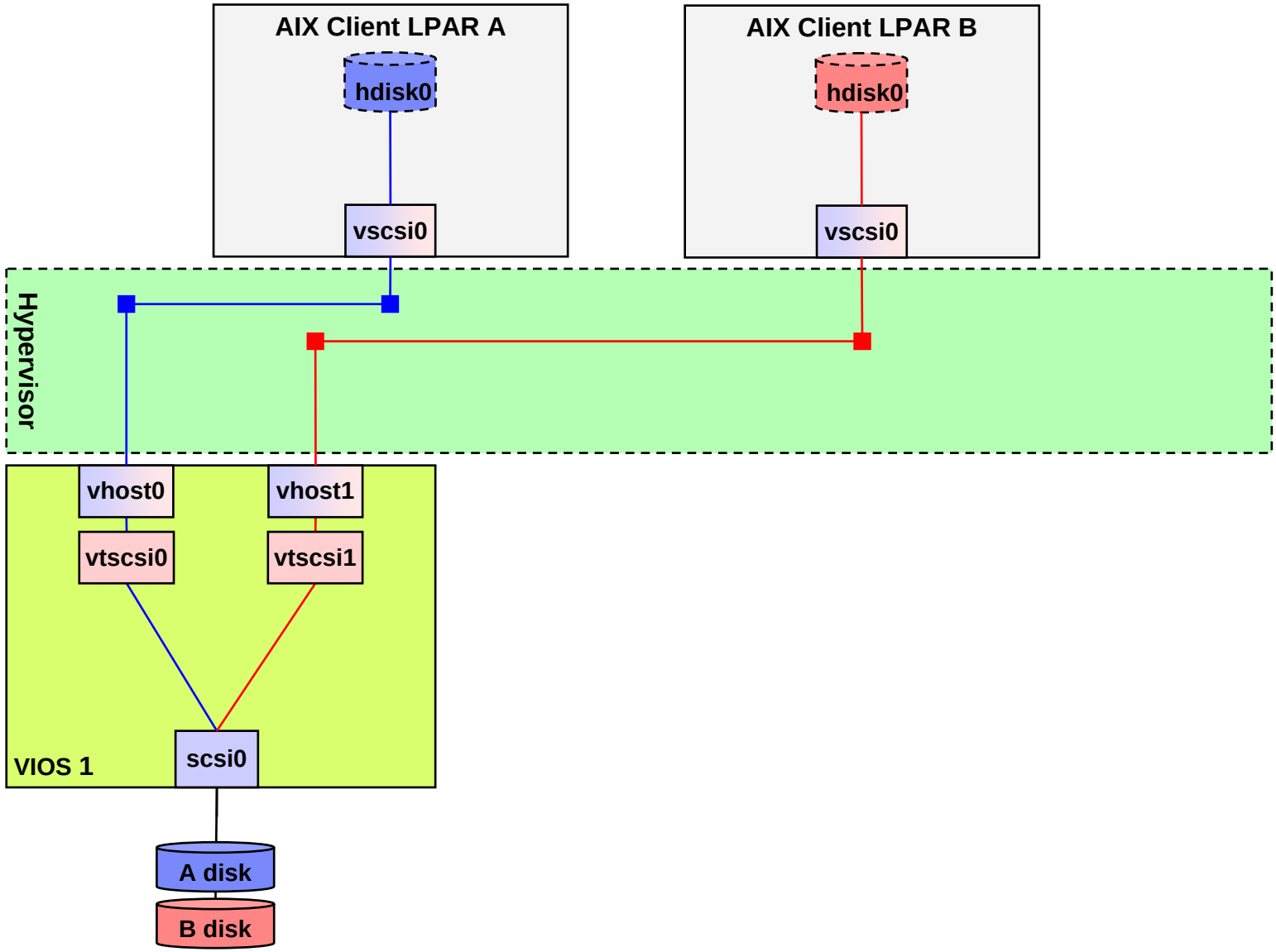
Single VIOS, PV VSCSI Disks

- **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.



Virtual SCSI Options – Details

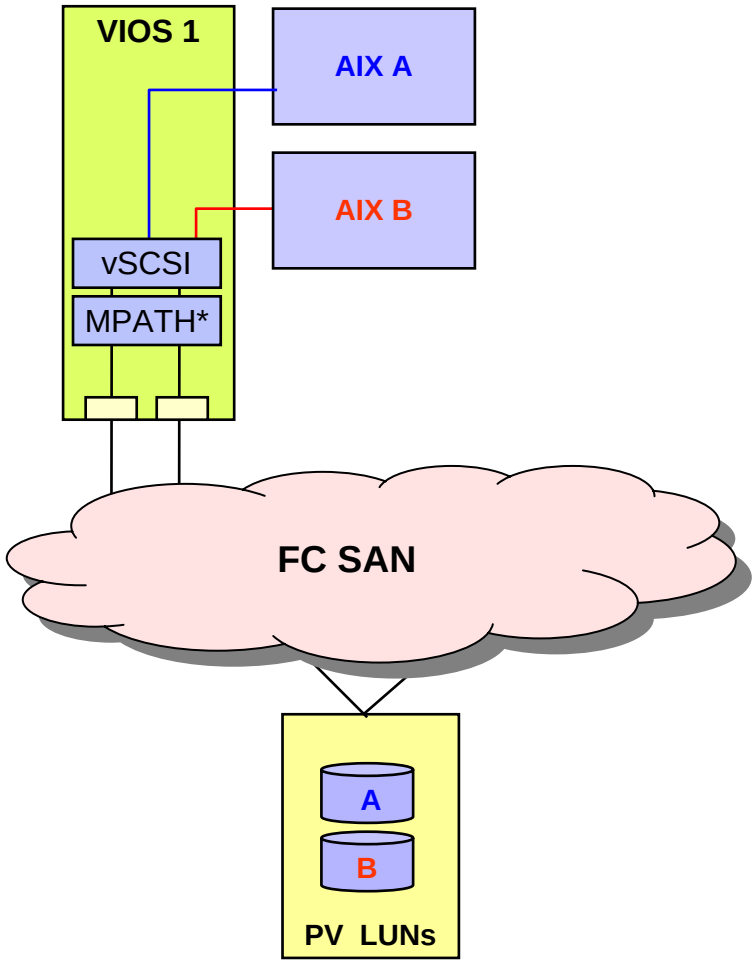
Single VIOS, PV VSCSI Disks



Virtual SCSI Options

Single VIOS with Multi-Path I/O

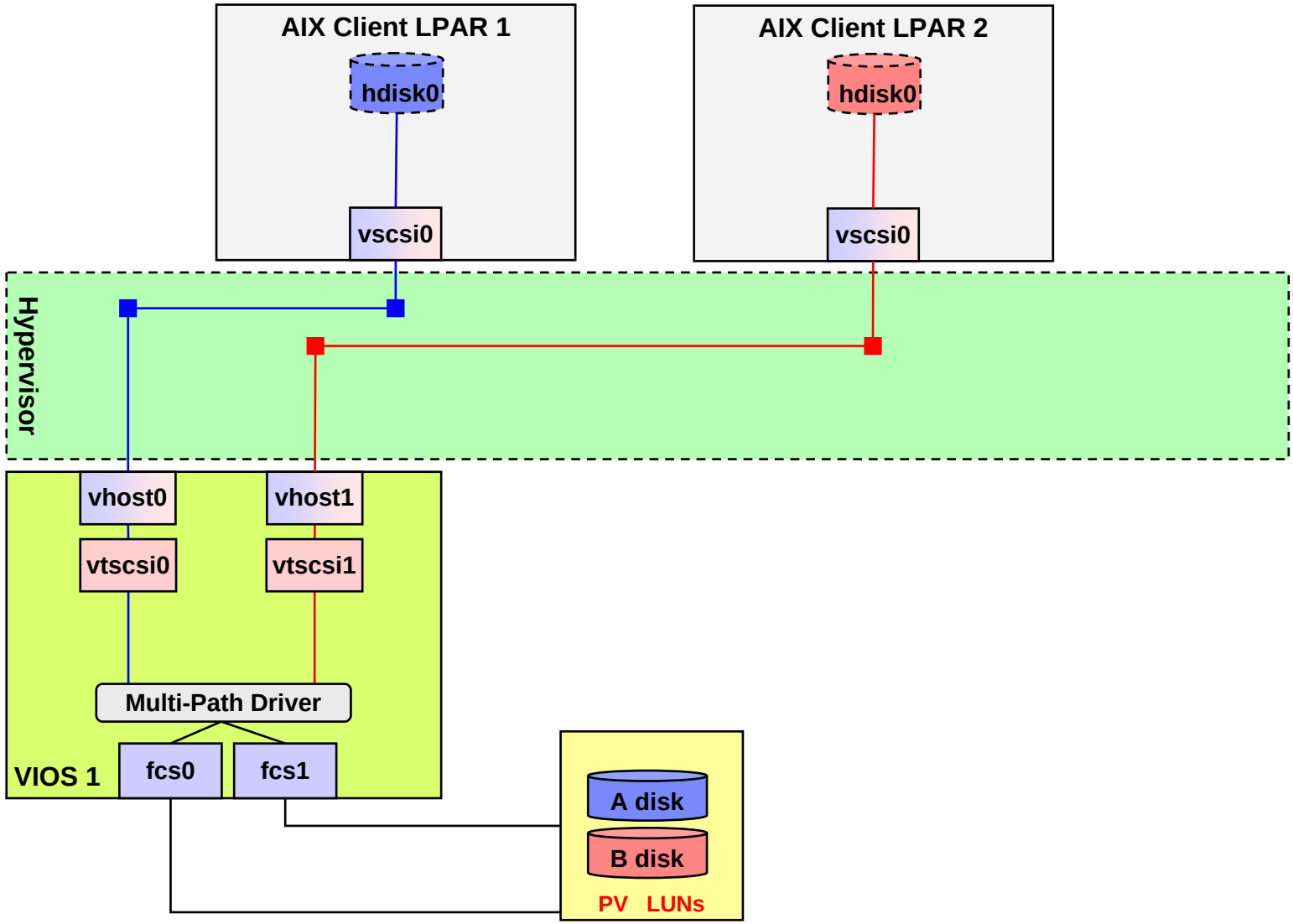
- **Complexity**
  - Simpler to setup and manage than dual VIO servers
  - Requires Multi-Path I/O setup on the VIOS
  - No specialized setup on the client
- **Resilience**
  - VIOS is a single point of failure
- **Throughput / Scalability**
  - Potential for increased bandwidth due to multi-path I/O.
  - Could divide clients across independent VIOS allowing more VIOS adapter bandwidth.



* Note: See the slide labeled VIOS Multi-Path Options for a high level overview of MPATH options.

Virtual SCSI Options - Details

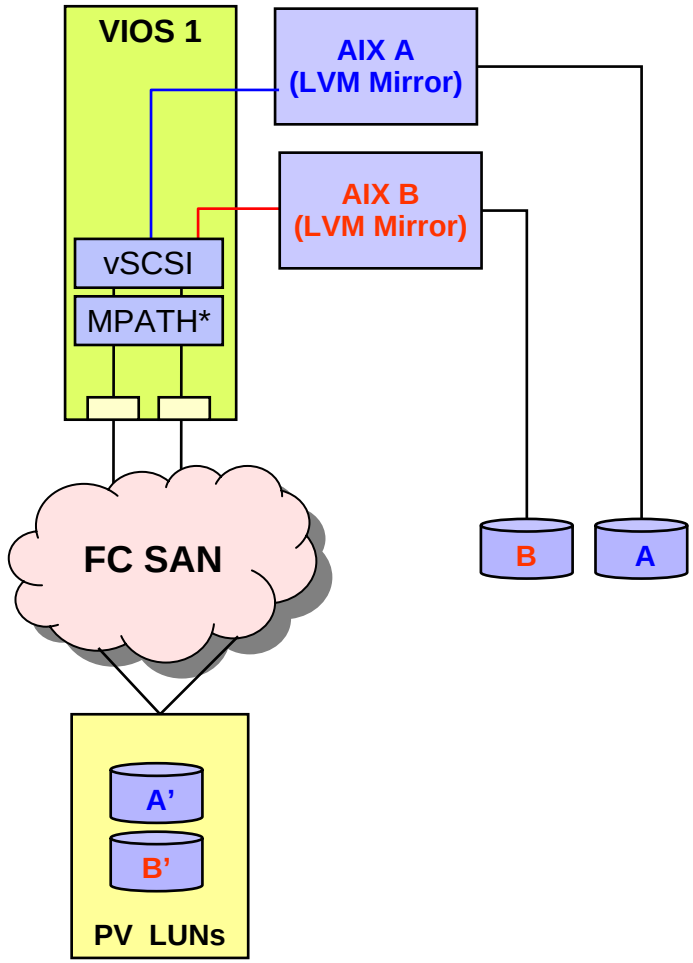
Single VIOS with Multi-Path I/O



Virtual SCSI Options

AIX Client Mirroring with direct attach SCSI and VIOS PV VSCSI disks

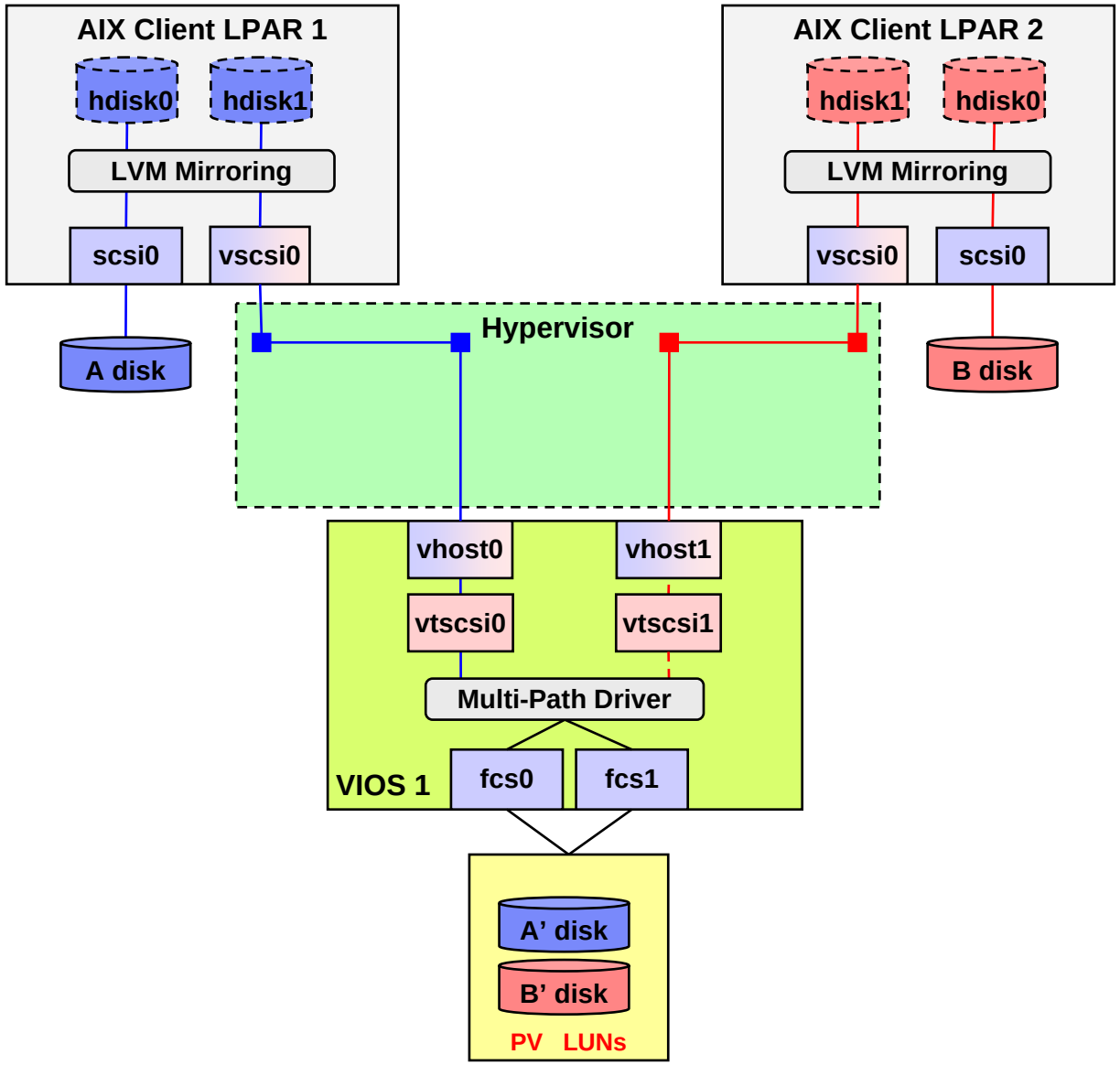
- **Complexity**
  - Requires LVM mirroring to be setup on the VIOC
  - Multi-Path I/O setup on the VIOS
  - If a VIOS is rebooted, the mirrored disks will need to be resynchronized via a varyonvg on the VIOC.
  - Additional complexity due to multiple disk types, Multi-Path I/O setup, and client mirroring.
- **Resilience**
  - Protection against failure of single adapter failure (or path) or disk
  - Potential protection against FC adapter failures within VIOS (if Multi-Path I/O is configured)



* Note: See the slide labeled VIOS Multi-Path Options for a high level overview of MPATH options.

Virtual SCSI Options

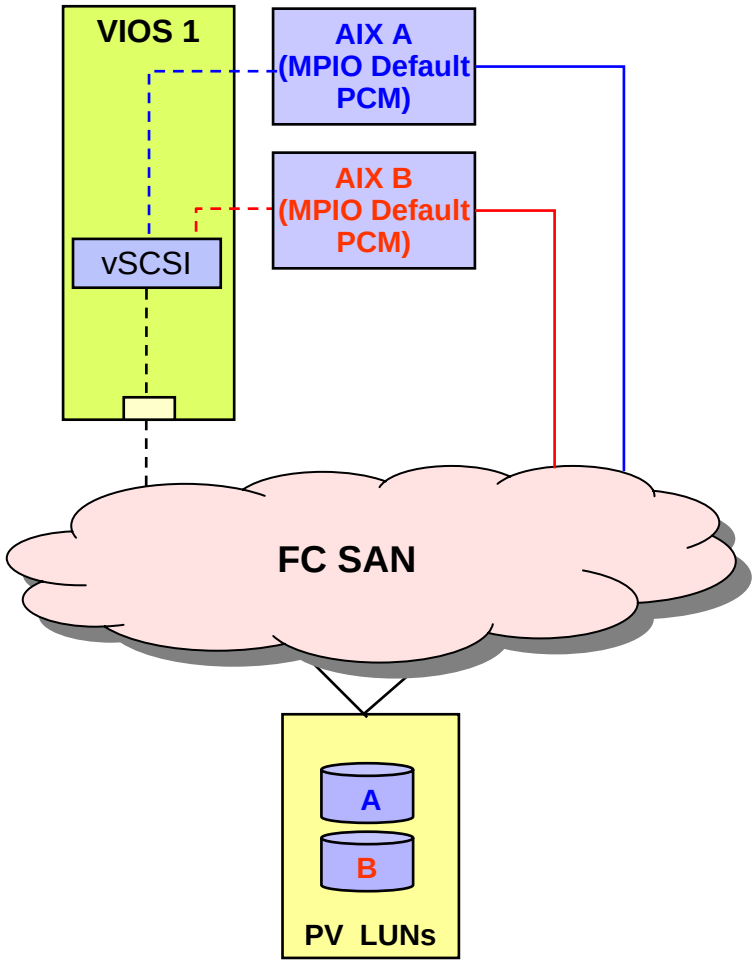
AIX Client Mirroring with direct attach SCSI and VIOS PV VSCSI disks



Virtual SCSI Options

AIX Default MPIO PCM Driver in Client, Direct Fibre and Backup via VIOS

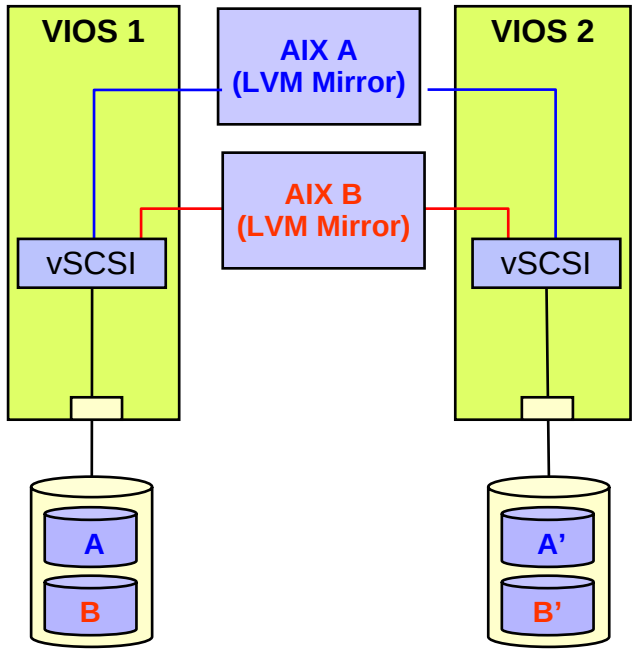
- **Notes**
  - This configuration is **not supported.**



Virtual SCSI Options

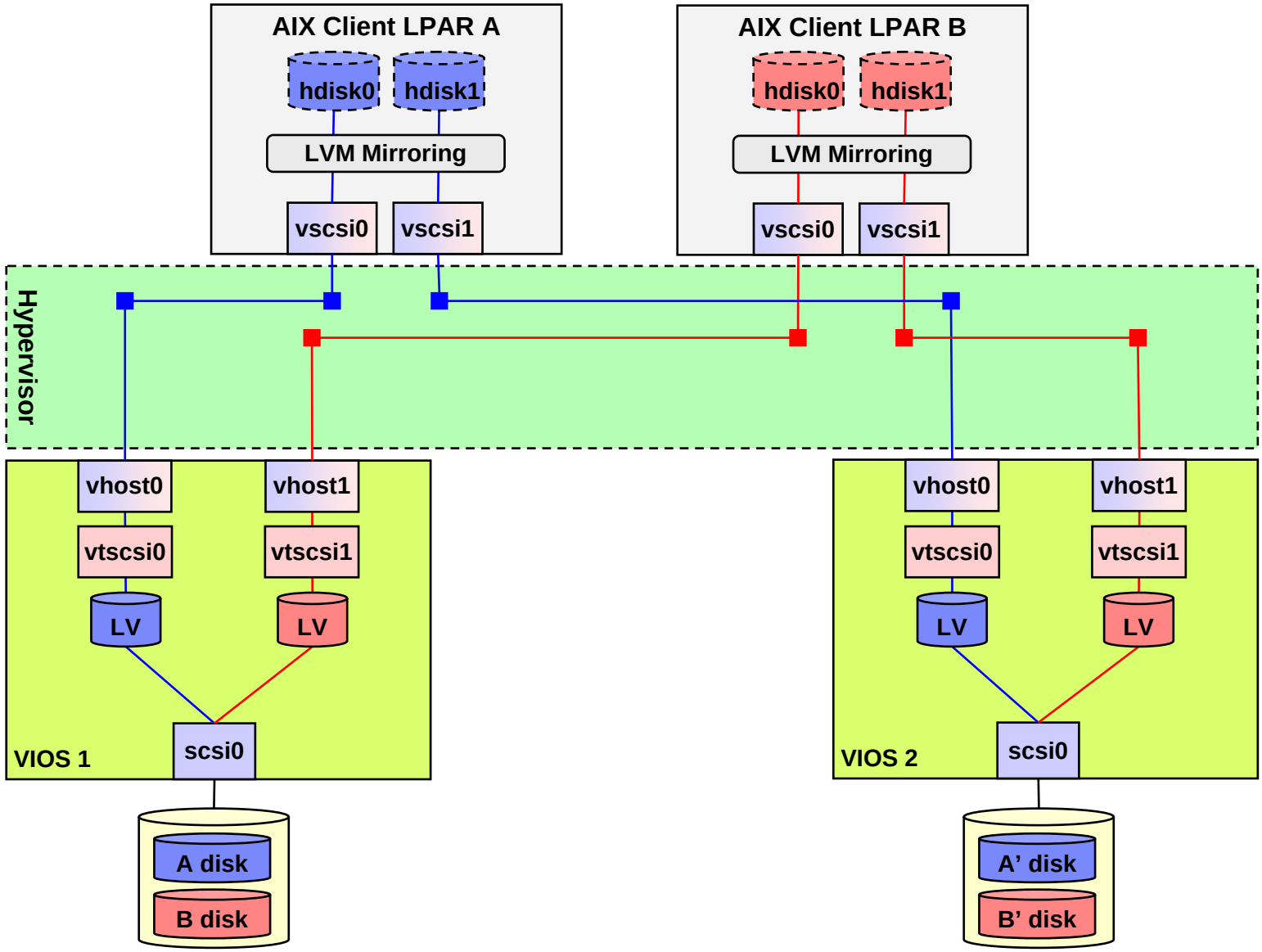
AIX Client Mirroring, Single Path in VIOS, LV VSCSI Disks

- **Complexity**
  - More complicated than single VIO server but does not require SAN ports or setup
  - Requires LVM mirroring to be setup on the client
  - If a VIOS is rebooted, the mirrored disks will need to be resynchronized via a varyonvg on the VIOC.
- **Resilience**
  - Protection against failure of single VIOS / SCSI disk / SCSI controller.
- **Throughput / Scalability**
  - VIOS performance limited by single SCSI adapter and internal SCSI disks.



Virtual SCSI Options - Details

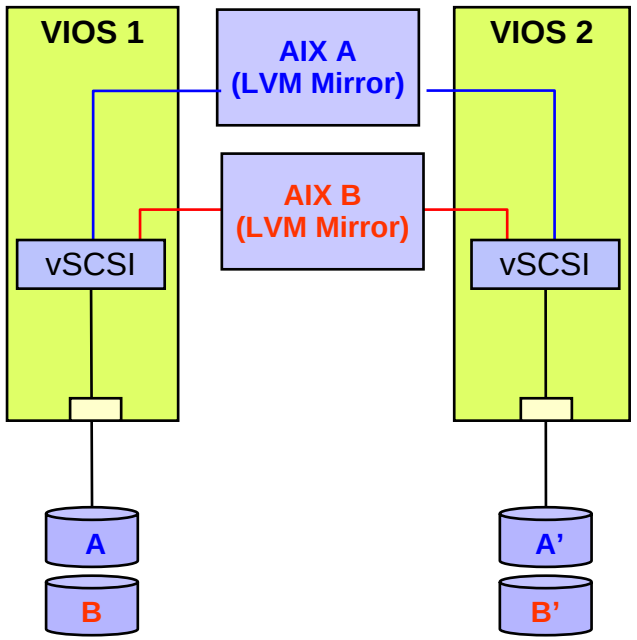
AIX Client Mirroring, Single Path in VIOS, LV VSCSI Disks



Virtual SCSI Options

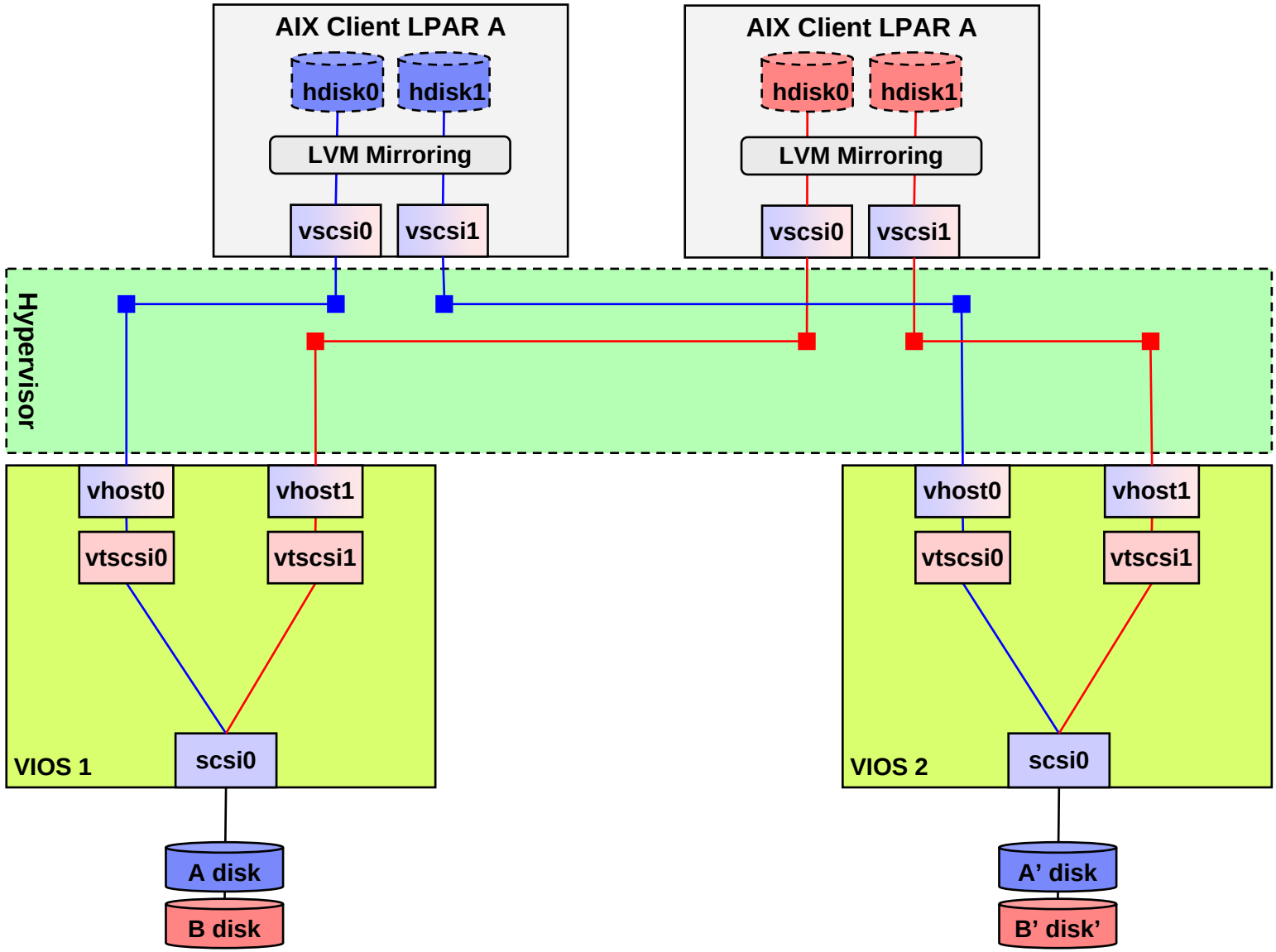
AIX Client Mirroring, Single Path in VIOS, PV VSCSI Disks

- **Complexity**
  - More complicated than single VIO server but does not require SAN ports or setup
  - Requires LVM mirroring to be setup on the client
  - If a VIOS is rebooted, the mirrored disks will need to be resynchronized via a varyonvg on the VIOC
- **Resilience**
  - Protection against single VIOS / SCSI disk / SCSI controller
  - The loss of a single physical disk would affect only one client
- **Throughput / Scalability**
  - VIOS performance limited by single SCSI adapter and internal SCSI disks.



Virtual SCSI Options - Details

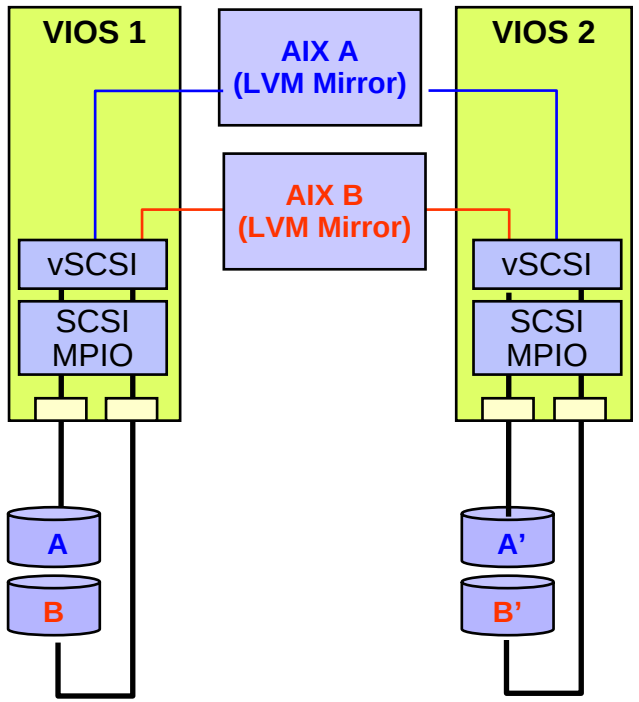
AIX Client Mirroring, Single Path in VIOS, PV VSCSI Disks



Virtual SCSI Options

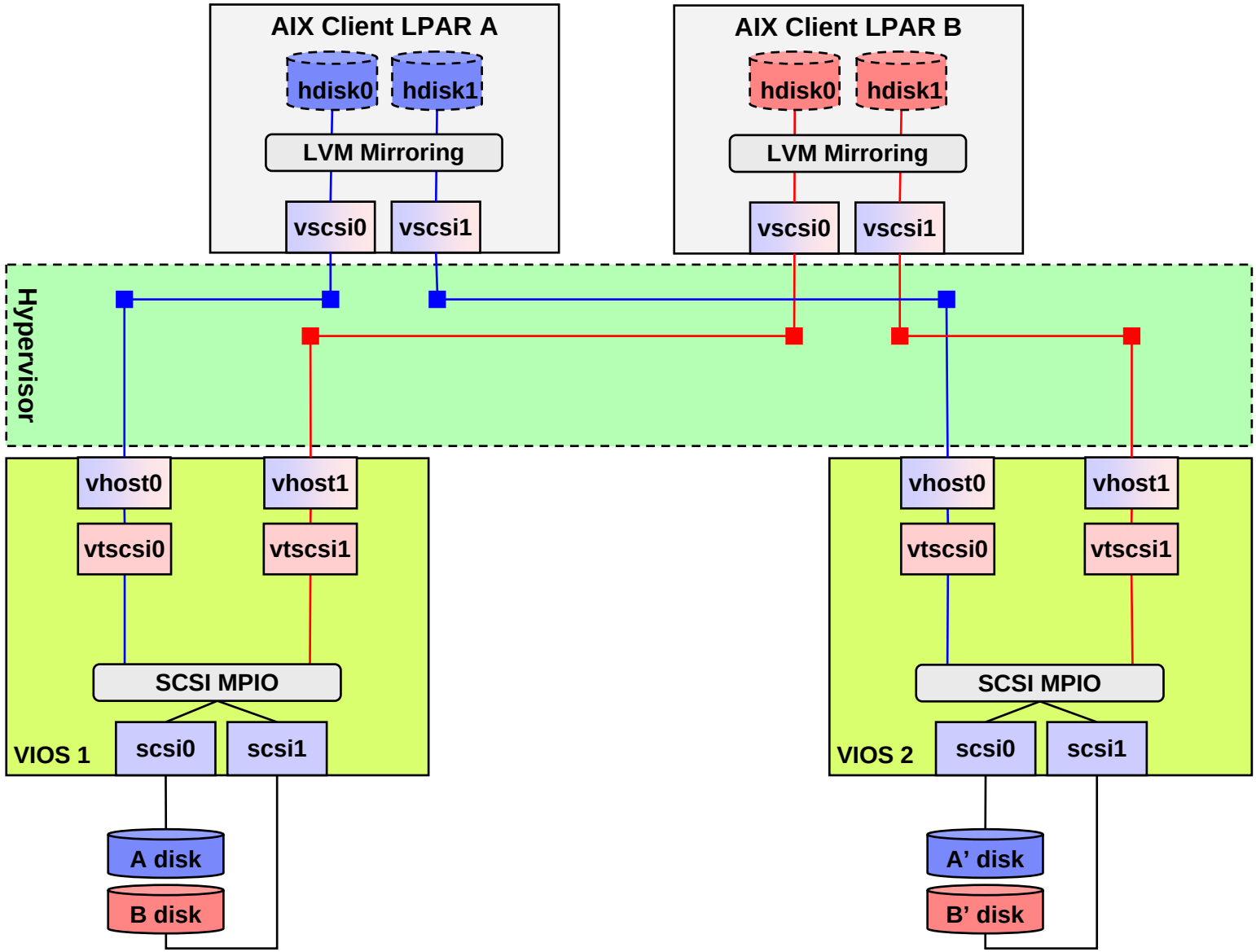
AIX Client Mirroring, with SCSI MPIO in VIO Server, PV VSCSI Disks

- **Complexity**
  - Requires LVM mirroring to be setup on the client
  - If a VIOS is rebooted, the mirrored disks will need to be resynchronized via a varyonvg on the VIOC
- **Resilience**
  - Protection against failure of single VIOS / SCSI disk / SCSI controller
- **Notes**
  - RAID not supported
  - Requires multiple adapters on each VIOS



Virtual SCSI Options - Details

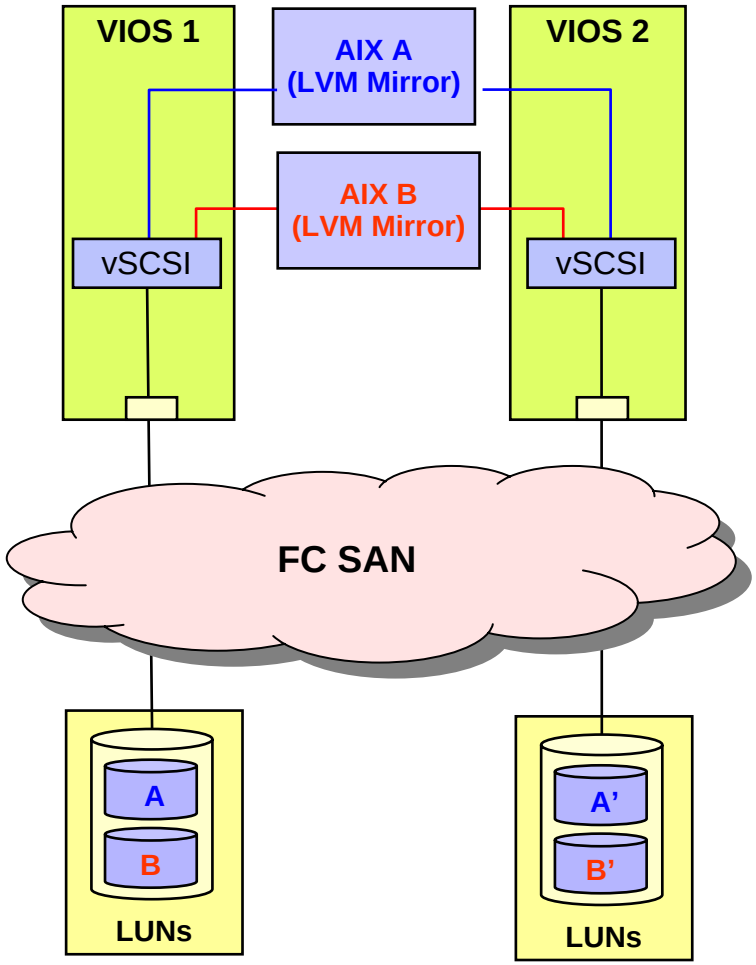
AIX Client Mirroring, with SCSI MPIO in VIOS, PV VSCSI Disks



Virtual SCSI Options

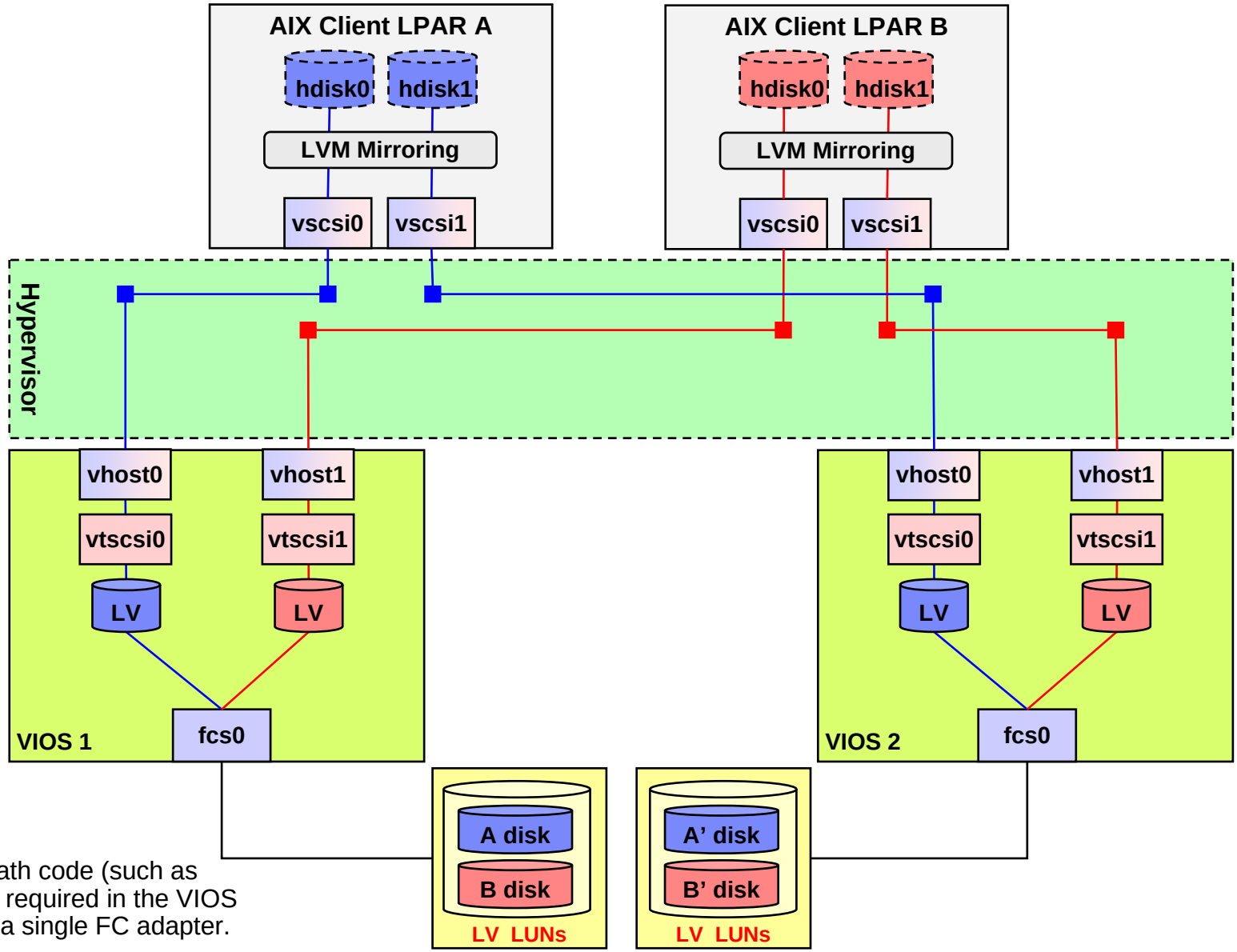
AIX Client Mirroring, Single Path in VIOS, LV or PV VSCSI FC Disks

- **Complexity**
  - Requires SAN configuration
  - Requires LVM mirroring to be setup on the client
- **Resilience**
  - Protection against failure of single VIOS / FC adapter (or path)
  - No protection against FC adapter failures within VIOS
- **Throughput / Scalability**
  - Performance limited by a single FC adapter.
- **Notes**
  - Disks must only be seen by one VIO server.
  - LV VSCSI LUNs could also be PV VSCSI LUNs.



Virtual SCSI Options - Details

AIX Client Mirroring, Single Path in VIOS, LV or PV VSCSI FC Disks

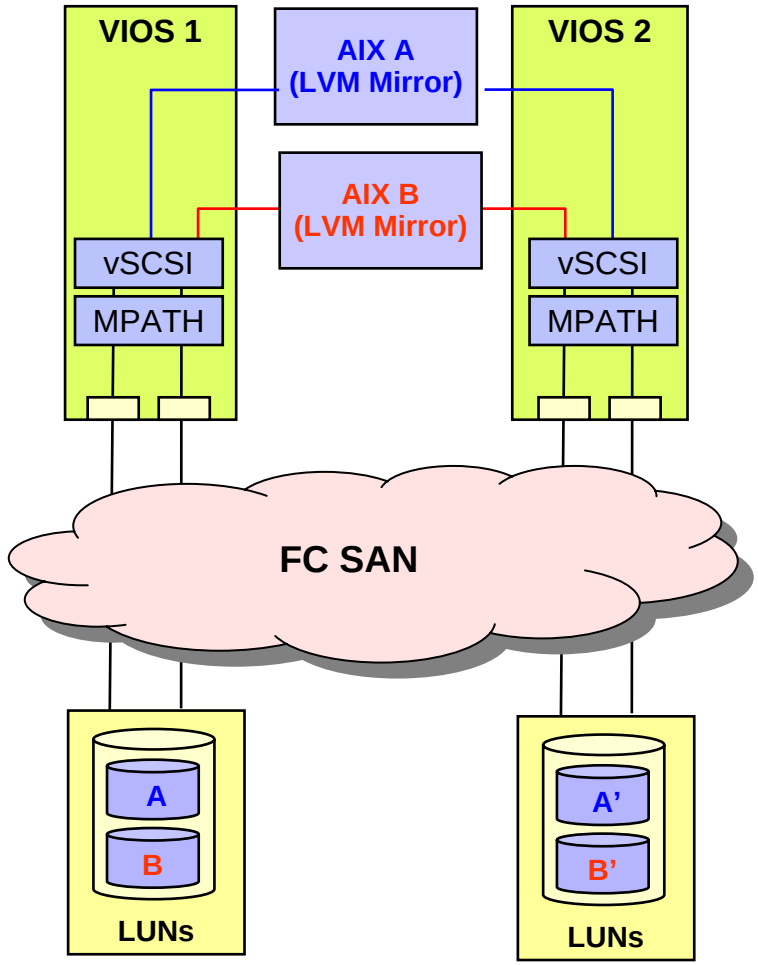


Note: A multi-path code (such as RDAC) may be required in the VIOS even if there is a single FC adapter.

Virtual SCSI Options

AIX Client Mirroring, Multi-Path I/O in VIOS, LV or PV VSCSI FC Disks

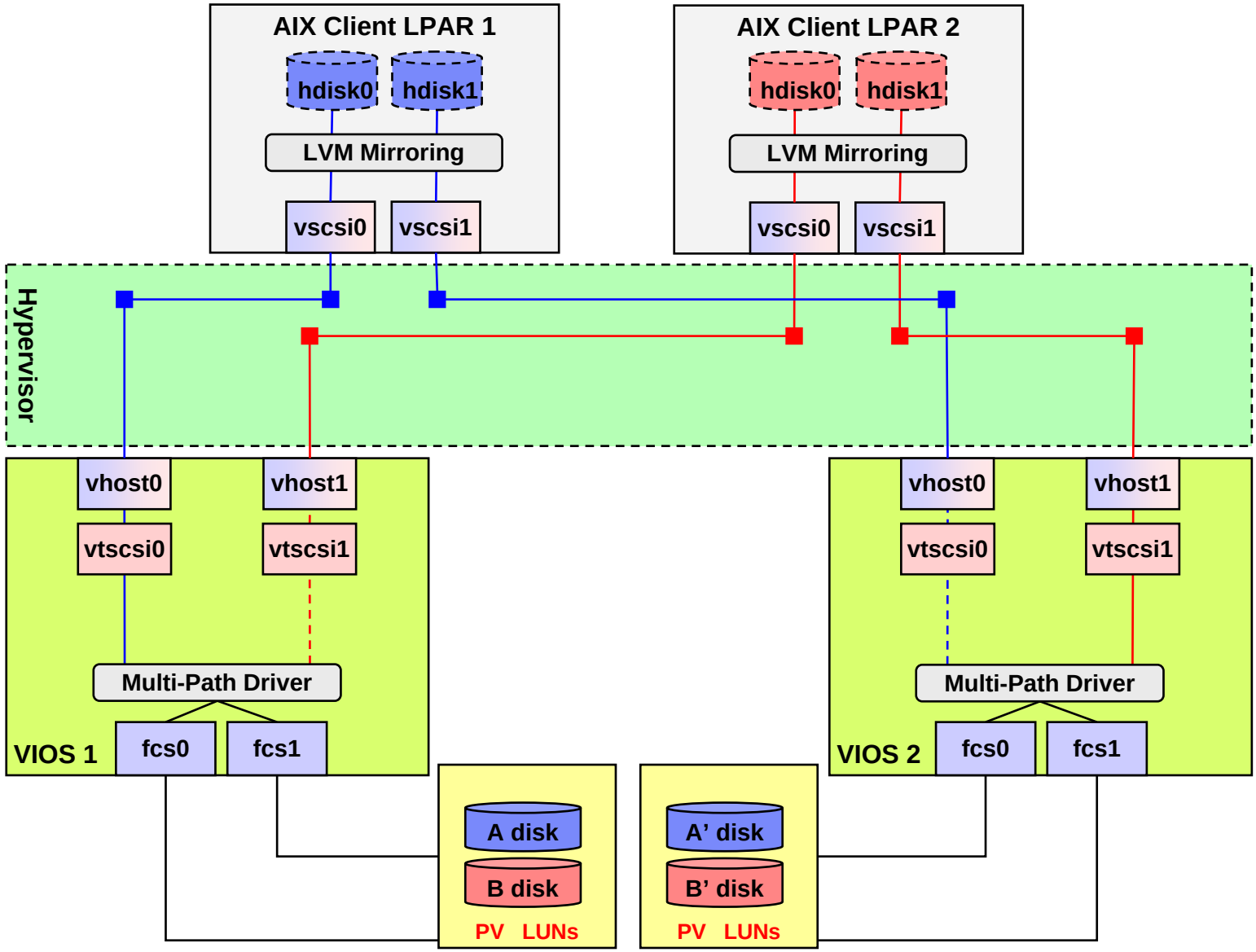
- **Complexity**
  - Requires LVM mirroring to be setup on the VIOC
  - Requires Multi-Path I/O setup on the VIOS
  - If a VIOS is rebooted, the mirrored disks will need to be resynchronized via a varyonvg on the VIOC
- **Resilience**
  - Protection against failure of single VIOS / FC adapter failure (or path)
  - Protection against FC adapter failures within VIOS
- **Throughput / Scalability**
  - Potential for increased bandwidth due to multi-path I/O
- **Notes**
  - LUNs used for this purpose can only be assigned to a single VIOS
  - LV VSCSI LUNs could also be PV VSCSI LUNs.



* Note: See the slide labeled VIOS Multi-Path Options for a high level overview of MPATH options.

Virtual SCSI Options - Details

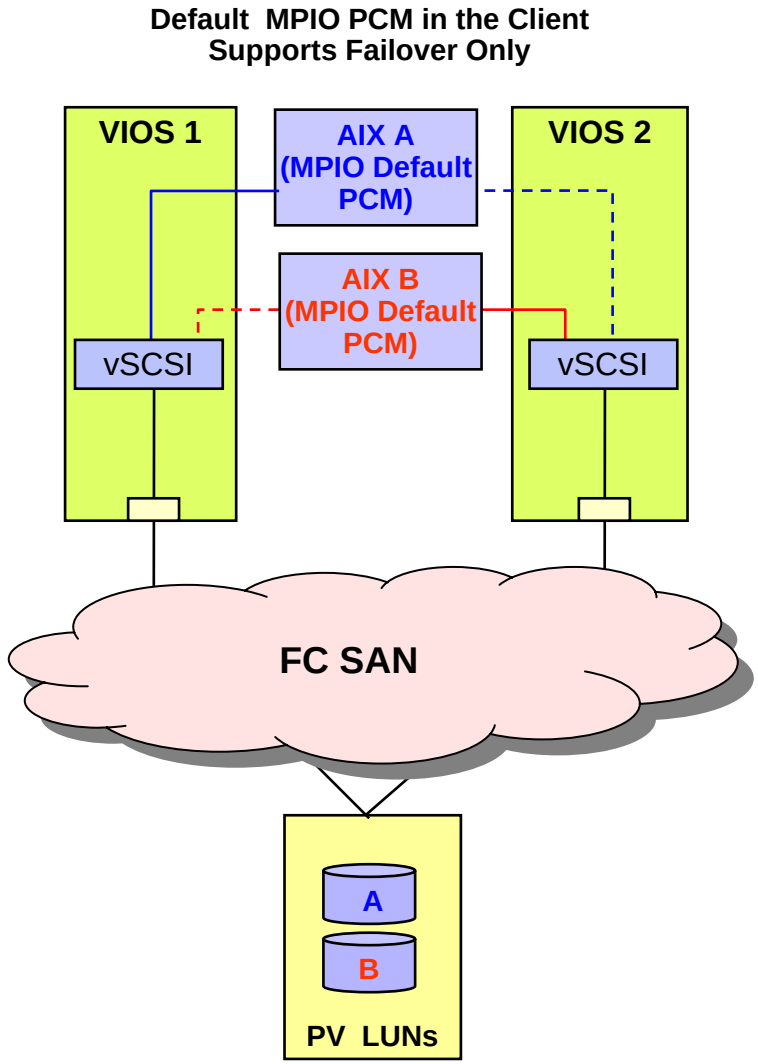
AIX Client Mirroring, Multi-Path I/O in VIOS, LV VSCSI FC Disks



Virtual SCSI Options

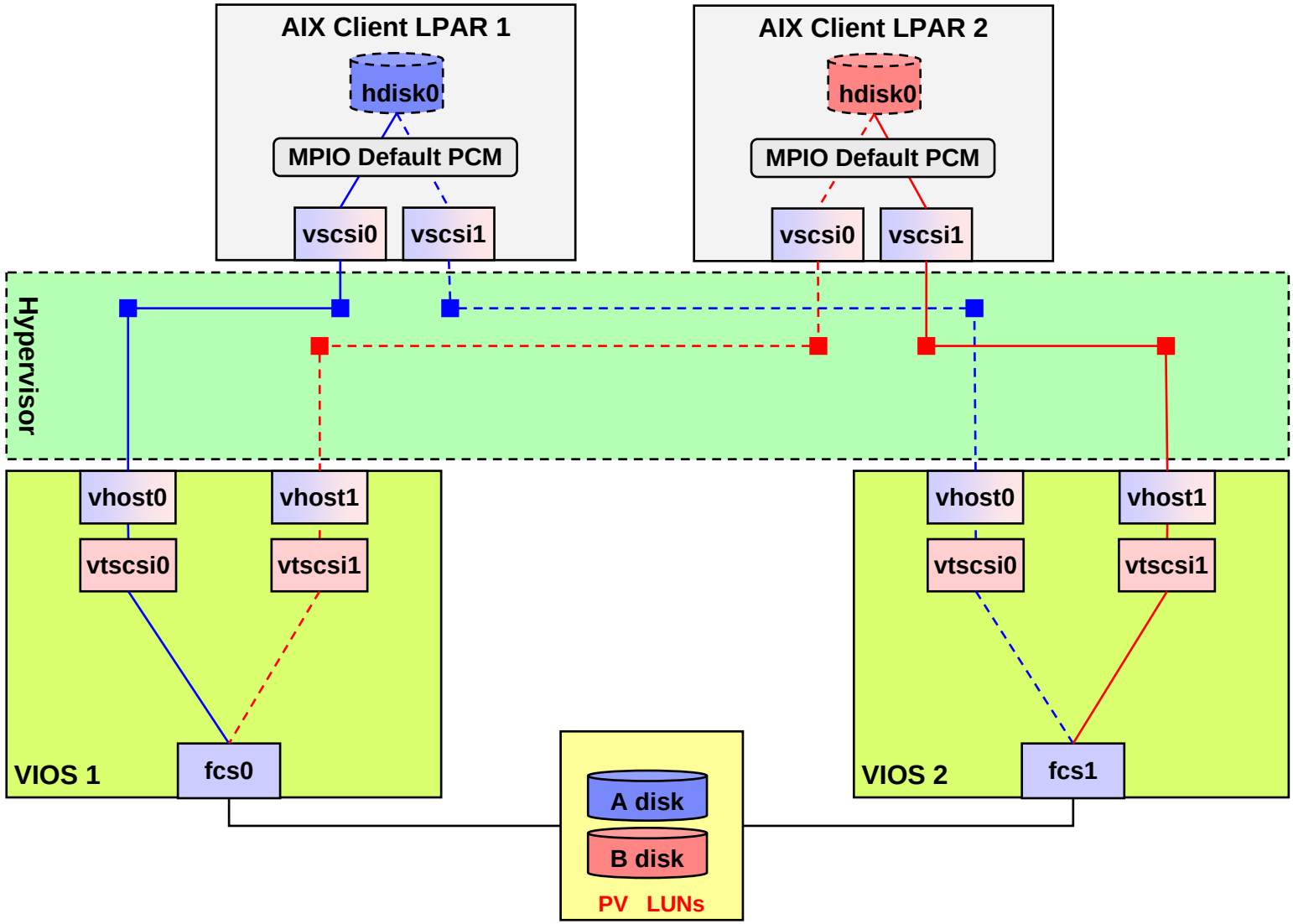
AIX MPIO Default PCM Driver in Client, Single Path in VIOS

- **Complexity**
  - Simplest dual VIOS FC option
  - Requires MPIO to be setup on the client
- **Resilience**
  - Protection against failure of a single VIOS / FC adapter (or path)
- **Throughput / Scalability**
  - Primary LUNs can be split across multiple VIOS to help balance the I/O load.
- **Notes**
  - Lowest port costs of dual VIOS FC options
  - Must be PV VSCSI disks.



Virtual SCSI Options - Details

AIX MPIO Default PCM Driver in Client, Single Path in VIOS

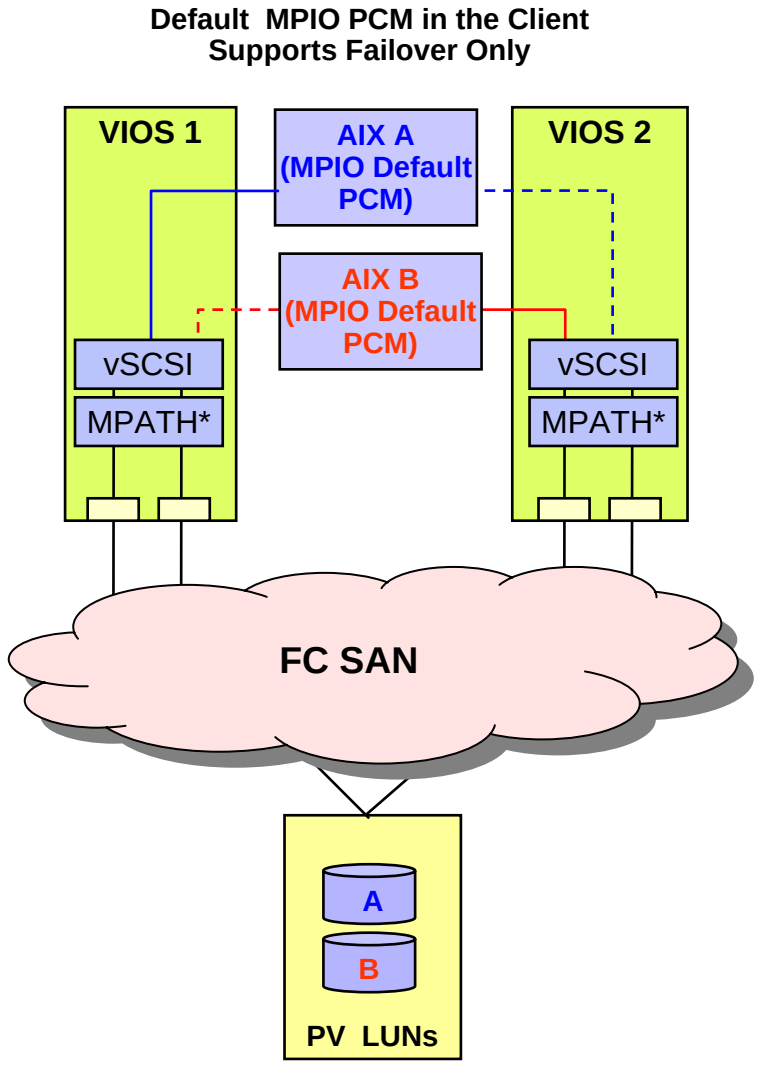


Note: A multi-path code (such as RDAC) may be required in the VIOS even if there is a single FC adapter.

Virtual SCSI Options

AIX MPIO Default PCM Driver in Client, Multi-Path I/O in VIOS

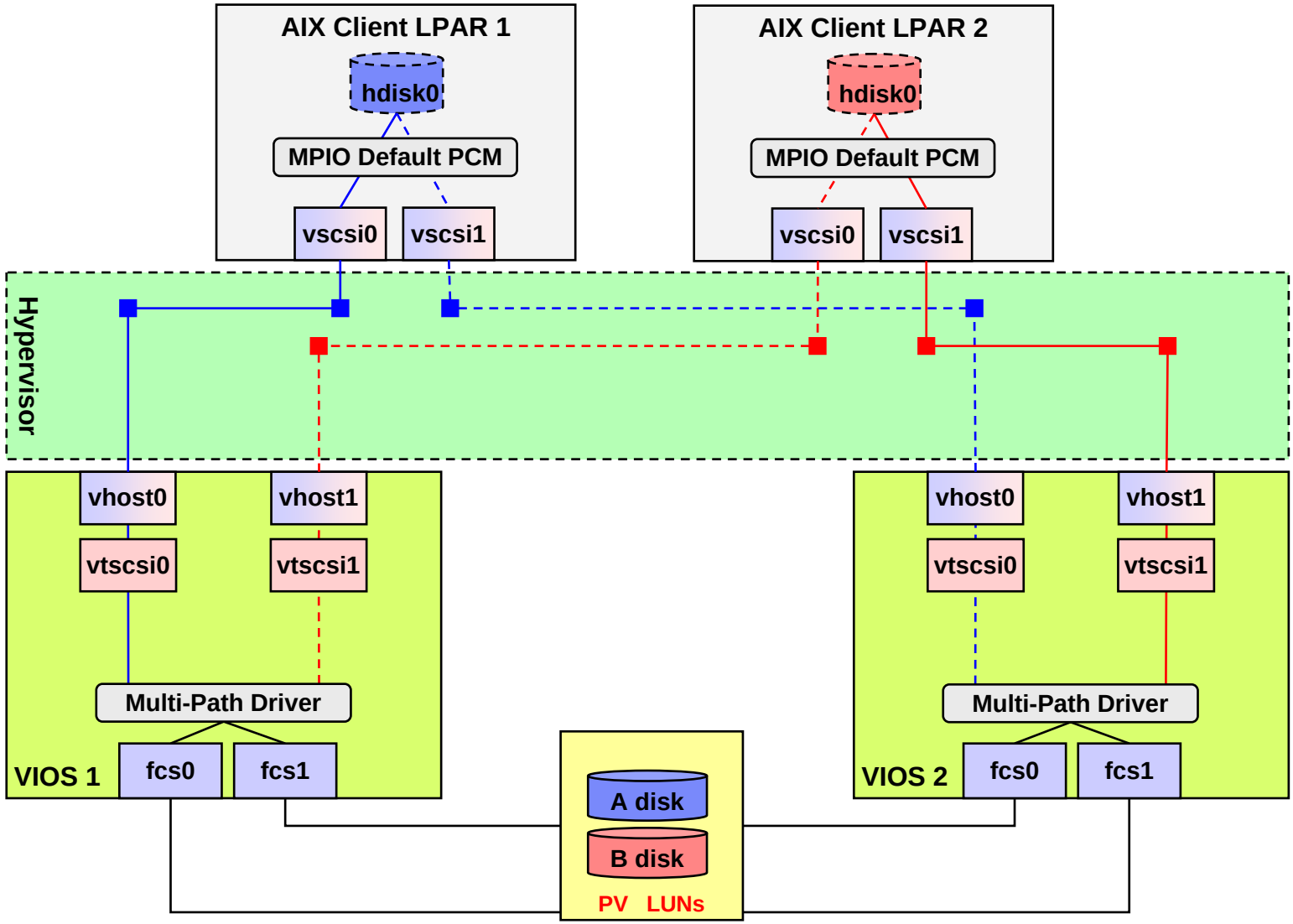
- **Complexity**
  - Requires MPIO to be setup on the client
  - Requires Multi-Path I/O setup on the VIOS
- **Resilience**
  - Protection against failure of a single VIOS, FC adapter, or path.
  - Protection against FC adapter failures within VIOS
- **Throughput / Scalability**
  - Potential for increased bandwidth due to Multi-Path I/O
  - Primary LUNs can be split across multiple VIOS to help balance the I/O load.
- **Notes**
  - Must be PV VSCSI disks.



* Note: See the slide labeled VIOS Multi-Path Options for a high level overview of MPATH options.

Virtual SCSI Options - Details

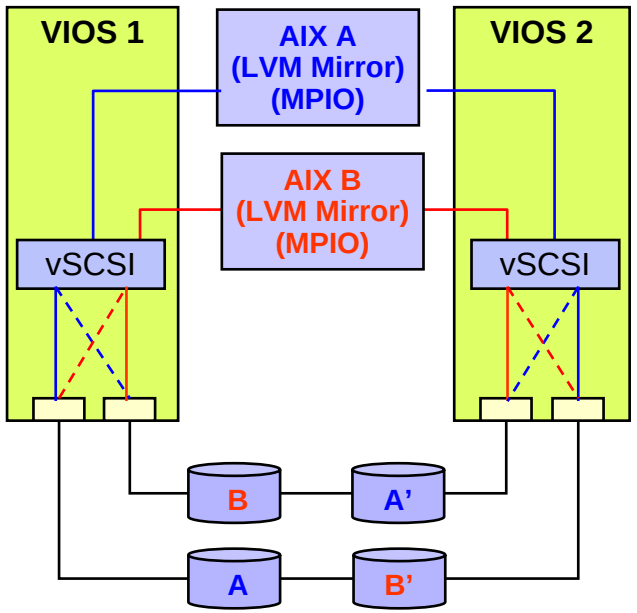
AIX MPIO Default PCM Driver in Client, Multi-Path I/O in VIOS



Virtual SCSI Options

AIX Client Mirroring, with twin-tailed SCSI in VIOS, PV VSCSI Disks

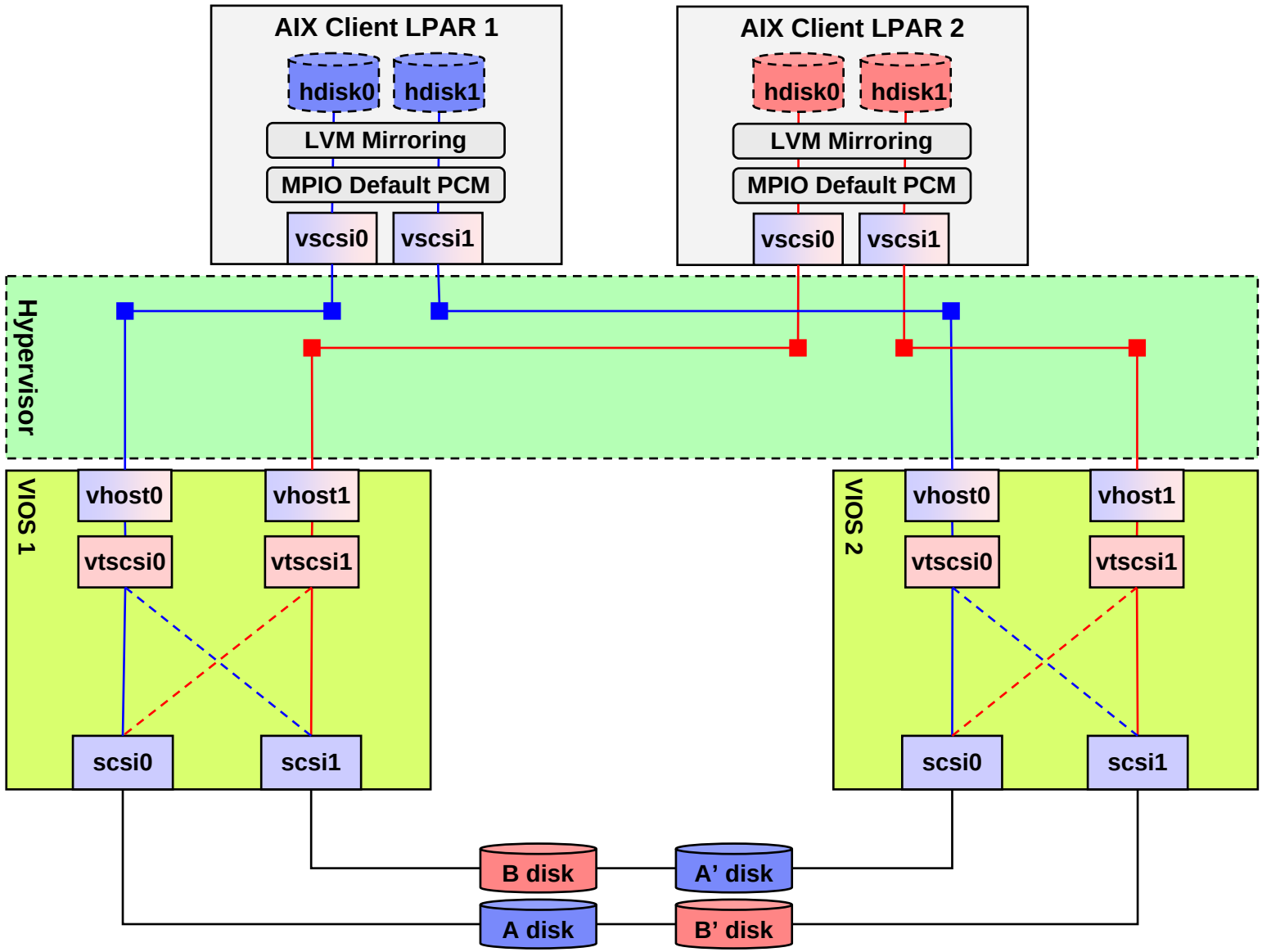
- **Complexity**
  - Requires LVM mirroring on the VIOC to protect against a disk or SCSI bus failure
  - Requires MPIO on the VIOC to protect against a VIOS failure
  - Failure of a VIOS does not result in a loss of a LVM mirror
- **Resilience**
  - Protects against failure of single VIOS
  - Reboot of VIOS or failure of a SCSI adapter (assuming it doesn't cause the SCSI bus to fail) on a VIOS should be transparent to the VIOC, as disks are accessed via the other VIOS
  - RAID not supported
- **Notes**
  - Disk subsystem cache isn't available with this option. Thus, this is only appropriate for low I/O requirements.
  - This is supported using 2104-DU3/TU3 (no longer available) and 2104-DS4/TS4 using any Ultra2 and above non-RAID SCSI adapter. See current support list.



External SCSI Disk

Virtual SCSI Options - Details

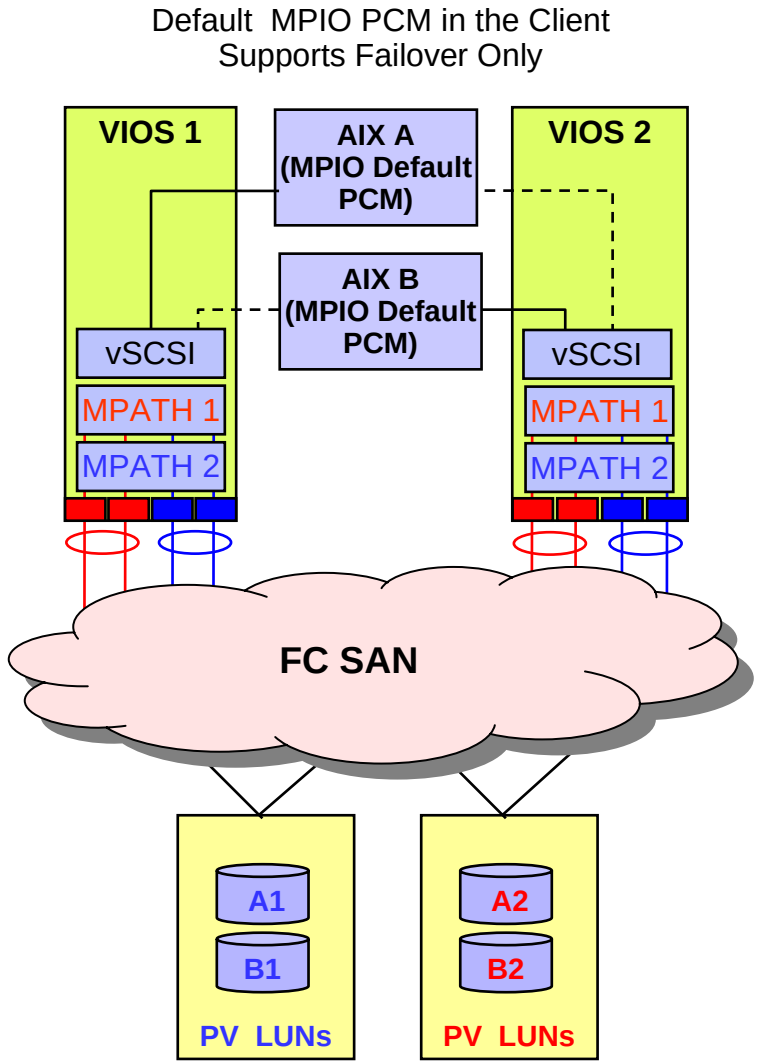
AIX Client Mirroring, with twin-tailed SCSI in VIOS, PV VSCSI Disks



Virtual SCSI Options

AIX Default MPIO PCM Driver in Client, Multiple Multi-path Codes in VIOS

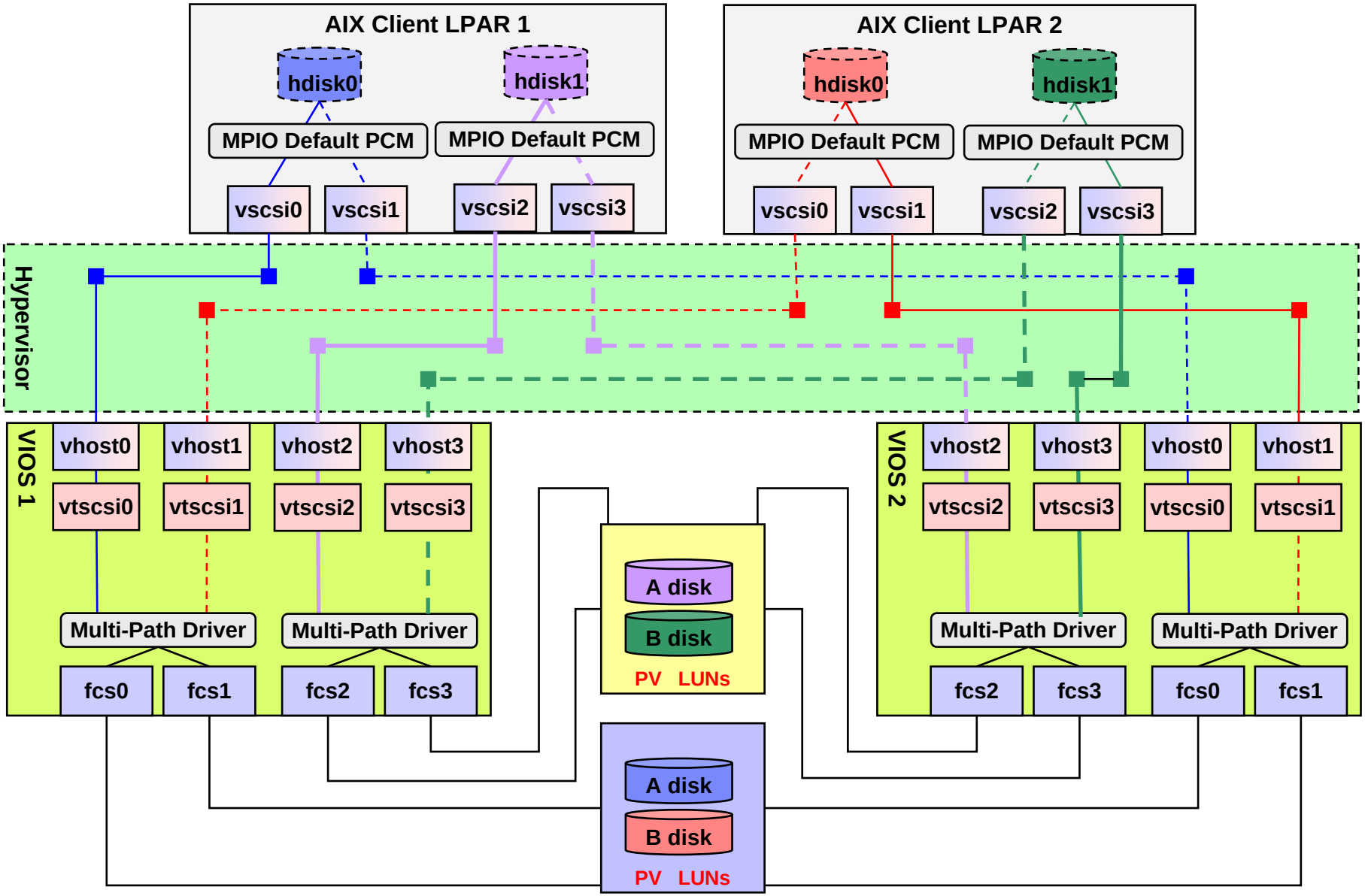
- **Complexity**
  - Complex to setup and manage
  - Requires MPIO to be setup on the client
  - Requires multiple Multi-Path I/O codes setup on the VIOS
  - A simpler option would be to add additional VIOS each supporting different MPATH drivers
- **Resilience**
  - Protection against failure of a single VIOS, FC adapter, or path.
  - Protection against FC adapter failures within VIOS
- **Throughput / Scalability**
  - Potential for increased bandwidth due to multi-path I/O.
  - Primary LUNs can be split across multiple VIOS to help balance the I/O load.
- **Notes**
  - Must be PV VSCSI disks.
  - Requires separate HBAs for each Multi-Path driver.

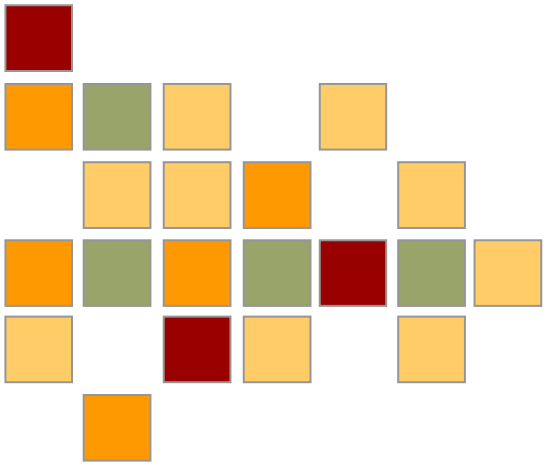


* Note: See the slide labeled VIOS Multi-Path Options for a high level overview of MPATH options.

Virtual SCSI Options - Details

AIX Default MPIO PCM Driver in Client, Multiple Multi-path Codes in VIOS





Backup Material

Steps to Setup Virtual SCSI

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.

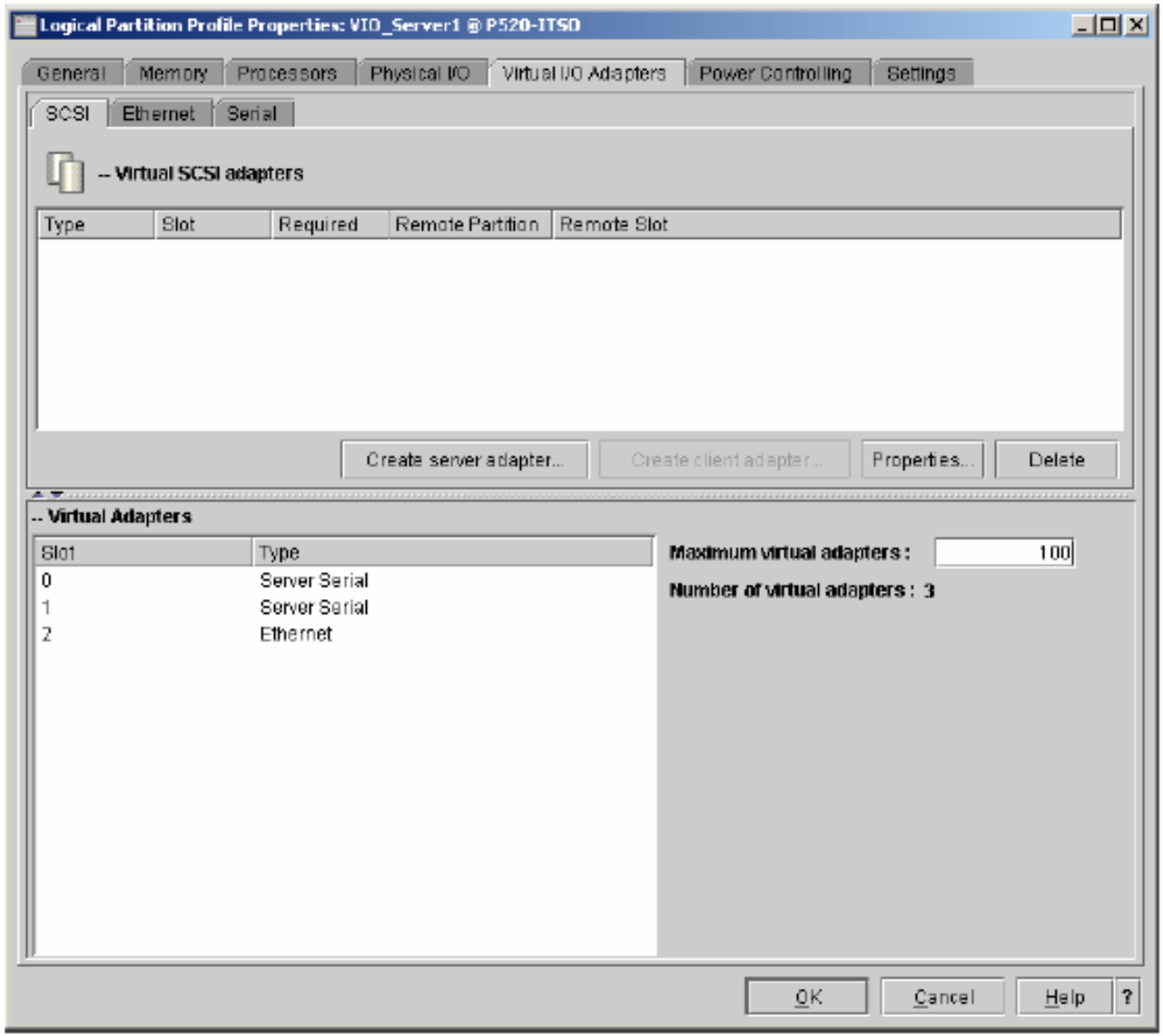
2. **Define the vSCSI client adapter**
   - This is also done on HMC and creates a Virtual SCSI Client Adapter (for example vscsi0) with a selectable slot number. When creating the Virtual SCSI Client Adapter you have to choose the desired I/O Server partition and the slot number of the Virtual SCSI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

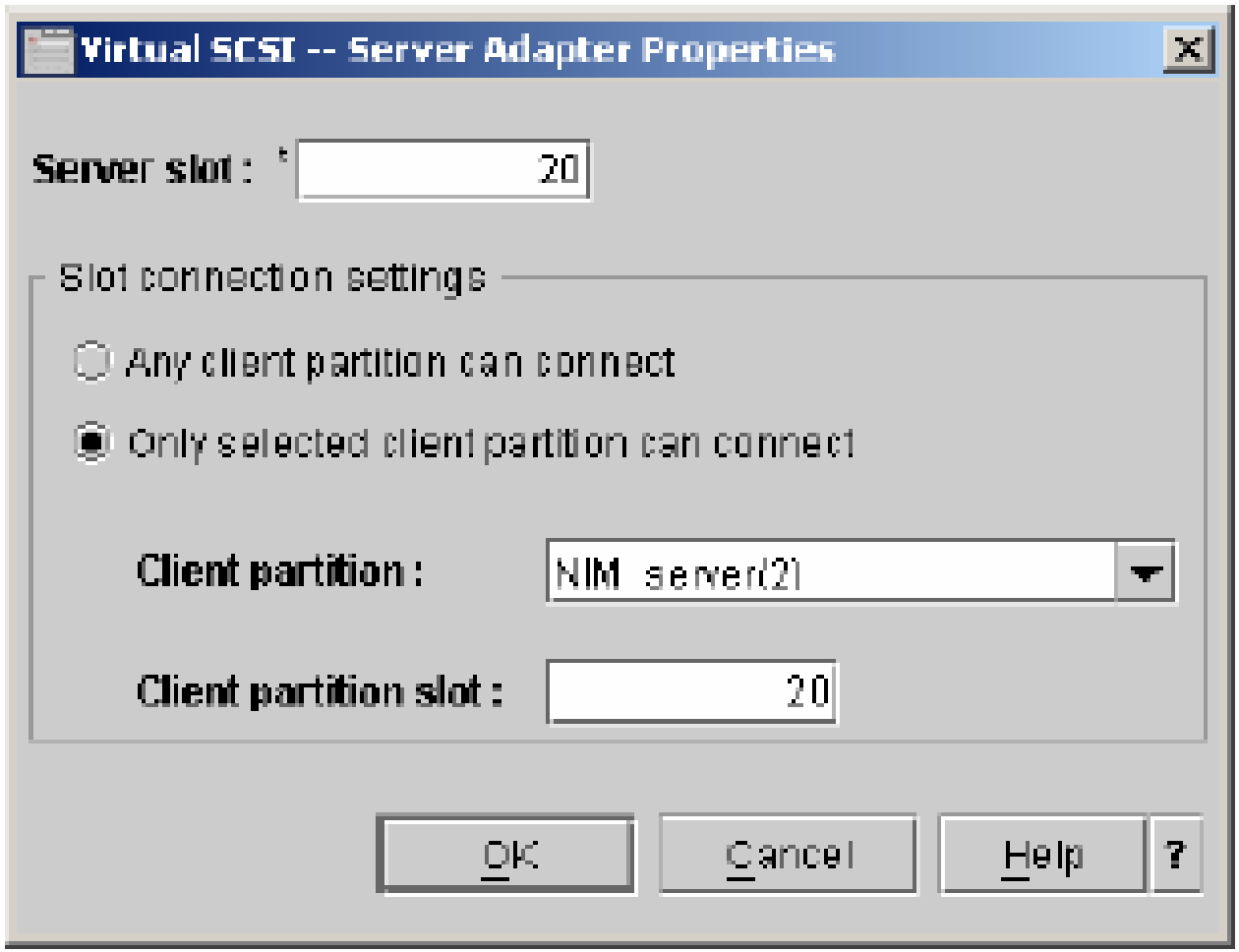
     The mapped volume now appears on the AIX partition as an hdisk device.

Step 1 - Define the Virtual SCSI Server Adapter
On the HMC..

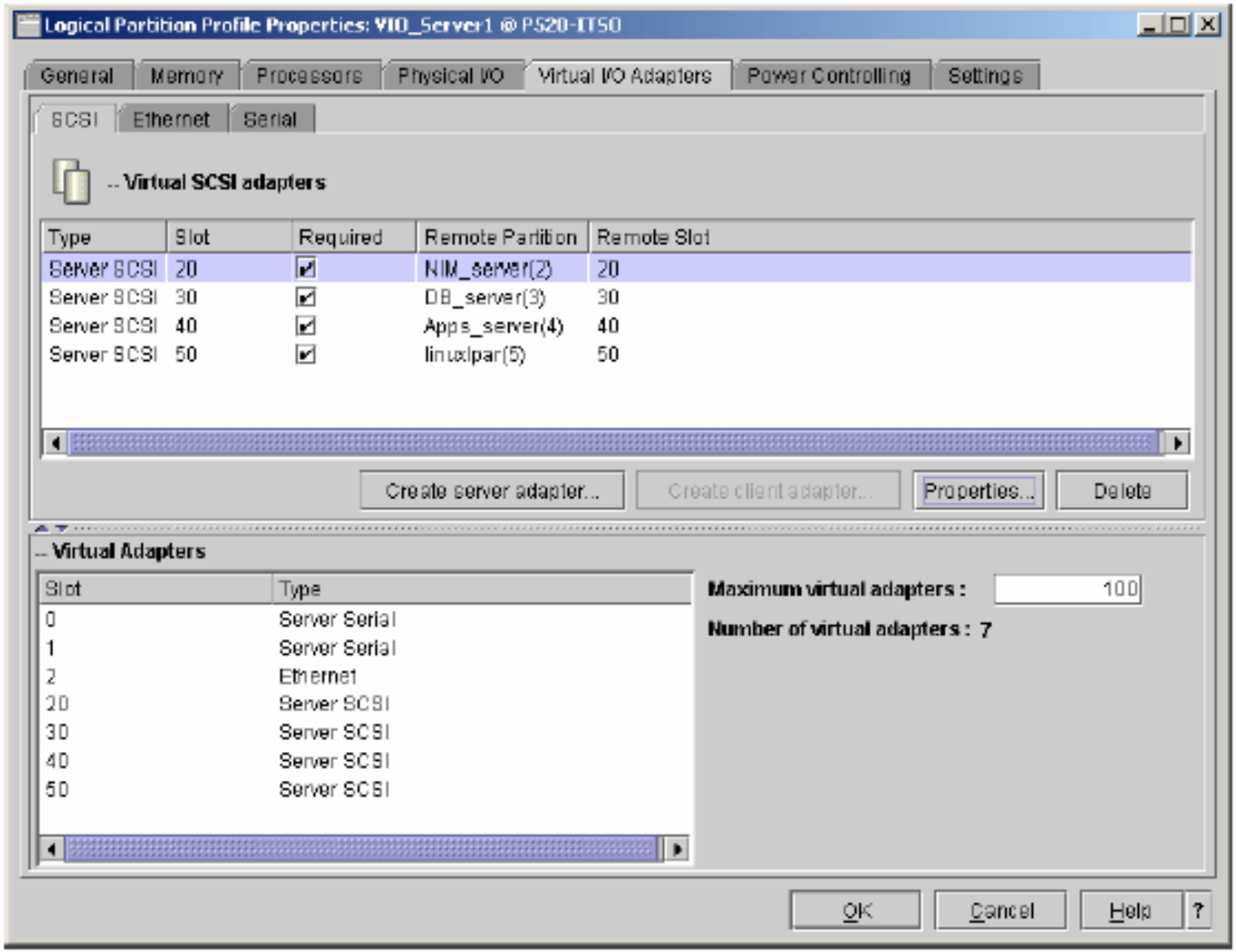


Step 1 - Define the Virtual SCSI Server Adapter

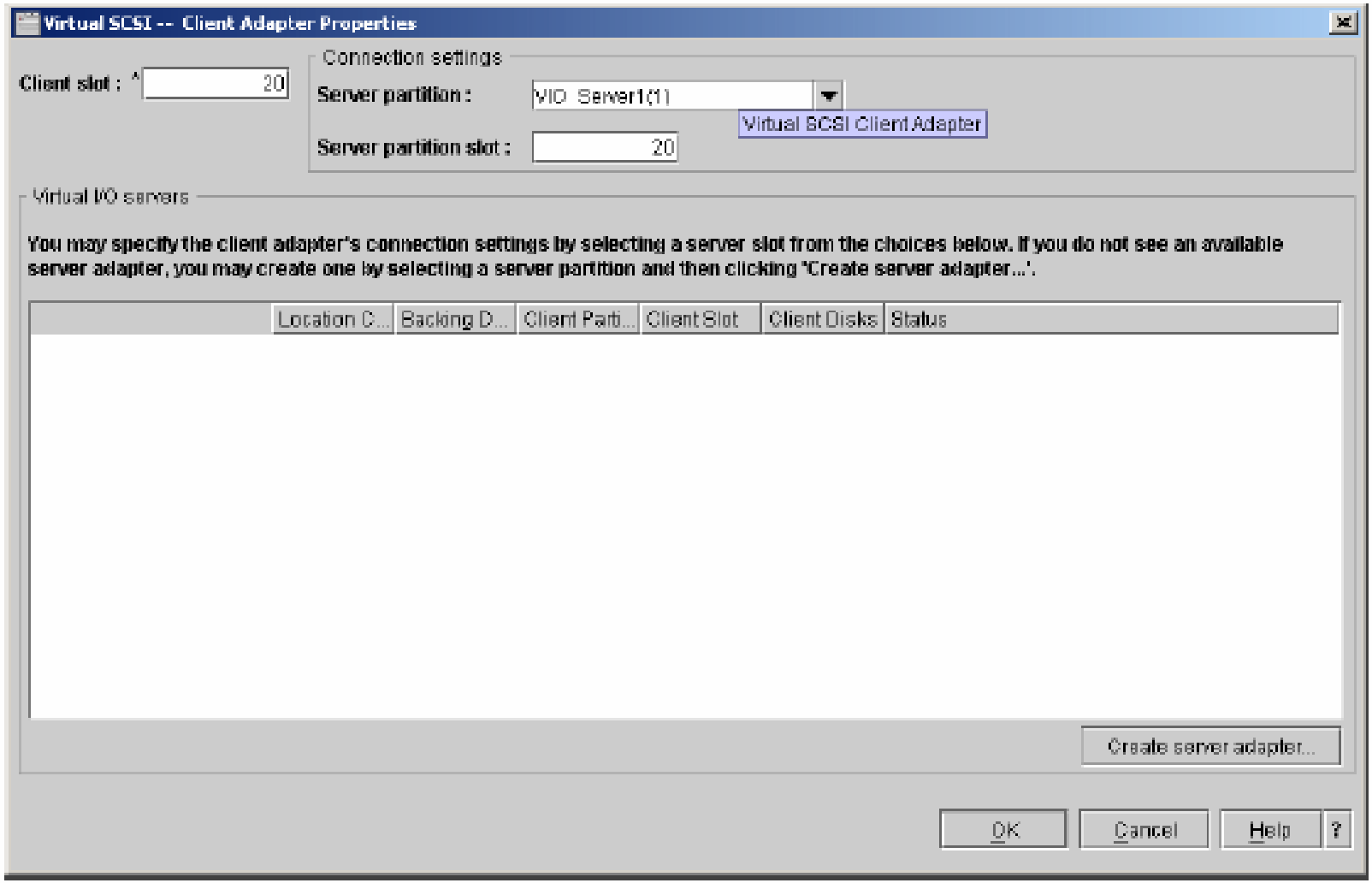
Define Virtual Slot, Server Adapter, Remote Information



Step 1 - Define the Virtual SCSI Server Adapter
Display results



Step 2 - Define the Virtual SCSI Client Adapter
On the HMC…



Step 3 - Create the physical VGs and LVs
On the VIO Server

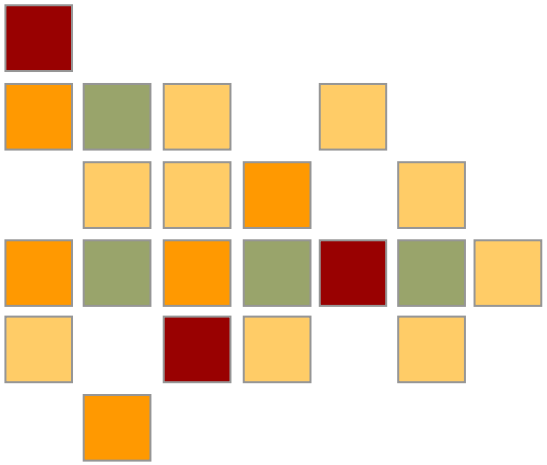
- **Create a volume group and assign disk to this volume group using the mkvg command as follows:**
  - mkvg -f -vg rootvg_clients hdisk2
- **Define the logical volume which will be visible as a disk to the client partition.**
  - mklv -lv rootvg_dbsrv rootvg_clients 2G

Step 4 - Creating virtual SCSI mapping
On the VIO Server…

- **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.**

URLs

- **Fibre Channel marketing bulletins: IBM eServer pSeries and RS/6000**
  - http://w3-1.ibm.com/sales/systems/portal/_s.155/254?navID=f220s240&geoID=AM&prodID=IBM%20eServer%20And%20TotalStorage%20Products&docID=rsfcsk.skit&docType=SalesKit&skCat=DocumentType
- **Workload Estimator**
  - http://www-912.ibm.com/wle/EstimatorServlet



END