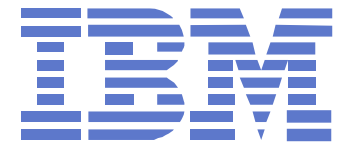


WLM and ESS: New Ways of Managing I/O Resources



OS/390 Expo
Session P7

Brent
Beardsley
IBM Tucson

Bob Rogers
IBM
Poughkeepsie

**"The Storage Server standard for the
new millennium"**

IBM Storage Solutions



Copyrights and Trademarks

(c) Copyright IBM Corporation 1999

MVS/ESA, OS/390, Database 2, Enterprise Storage Server, and RMF are trademarks of the International Business Machines Corporation.

IBM, RAMAC, ESCON, and DB2 are registered trademarks of the International Business Machines Corporation.

Information in this presentation is not intended to be an assertion of future action by IBM.

Permission is granted to OS/390 Expo to copy, reproduce or republish this document in whole or in part for related activities only.

IBM Storage Solutions



Disclaimer

- Performance values shown for ESS are a combination of current laboratory measurements on pre-GA product, and IBM's best projections of product performance at time of general availability. Actual performance may change over time.
- Performance values shown for non-IBM products are based on some measurement data, engineering projections, and input from industry experts. Actual performance may be different than shown.
- All performance data shown is based on specific workload and configuration assumptions. Performance will change under different workloads or configurations.

Agenda

- **Part I:** New functions improve I/O performance and management of mixed workloads:
 - Multiple allegiance
 - Parallel access volumes (PAVs)
 - I/O priority queuing
 - Performance CCWs

- **Part II:** WLM provides automatic management of PAVs
 - How it works
 - Externals

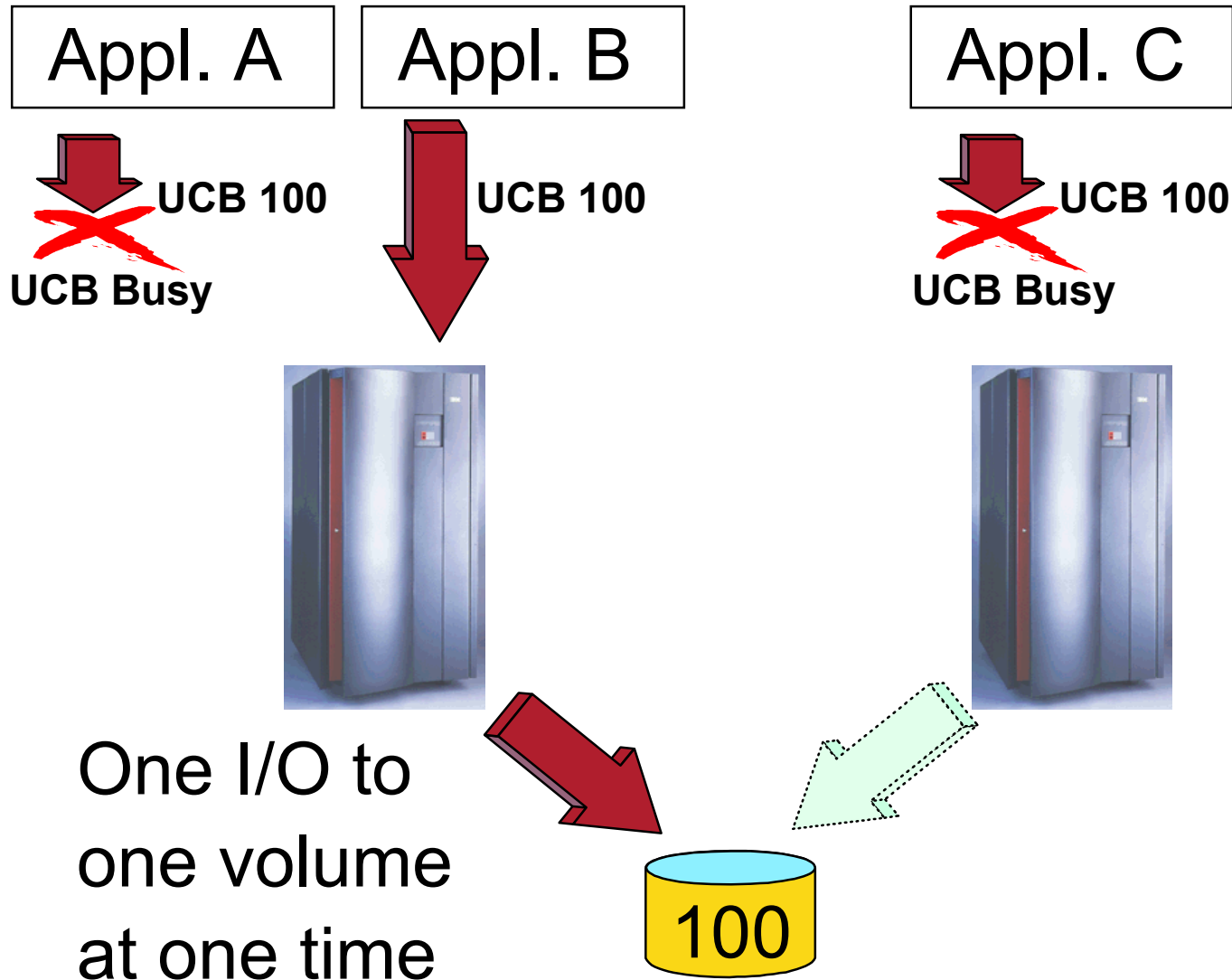
Managing UCB Contention

- UCB contention has been managed over the years in a variety of ways
 - Move the hot datasets to lightly loaded volumes
 - This has been a mostly manual process, although some vendors are offering products to help identify and move the data
 - Isolate the hot datasets on their own volumes
 - Hyper-volumes seem to fit this requirement nicely
 - Use high performance (cached) controllers
- However wouldn't it be nice if the problem could be handled automatically, without special volumes, without moving the data?
 - Multiple Allegiance
 - Allows different host systems to access the same volume at the same time
 - Parallel Access Volumes
 - Allows the same host system to drive multiple parallel I/Os to the same volume

IBM Storage Solutions

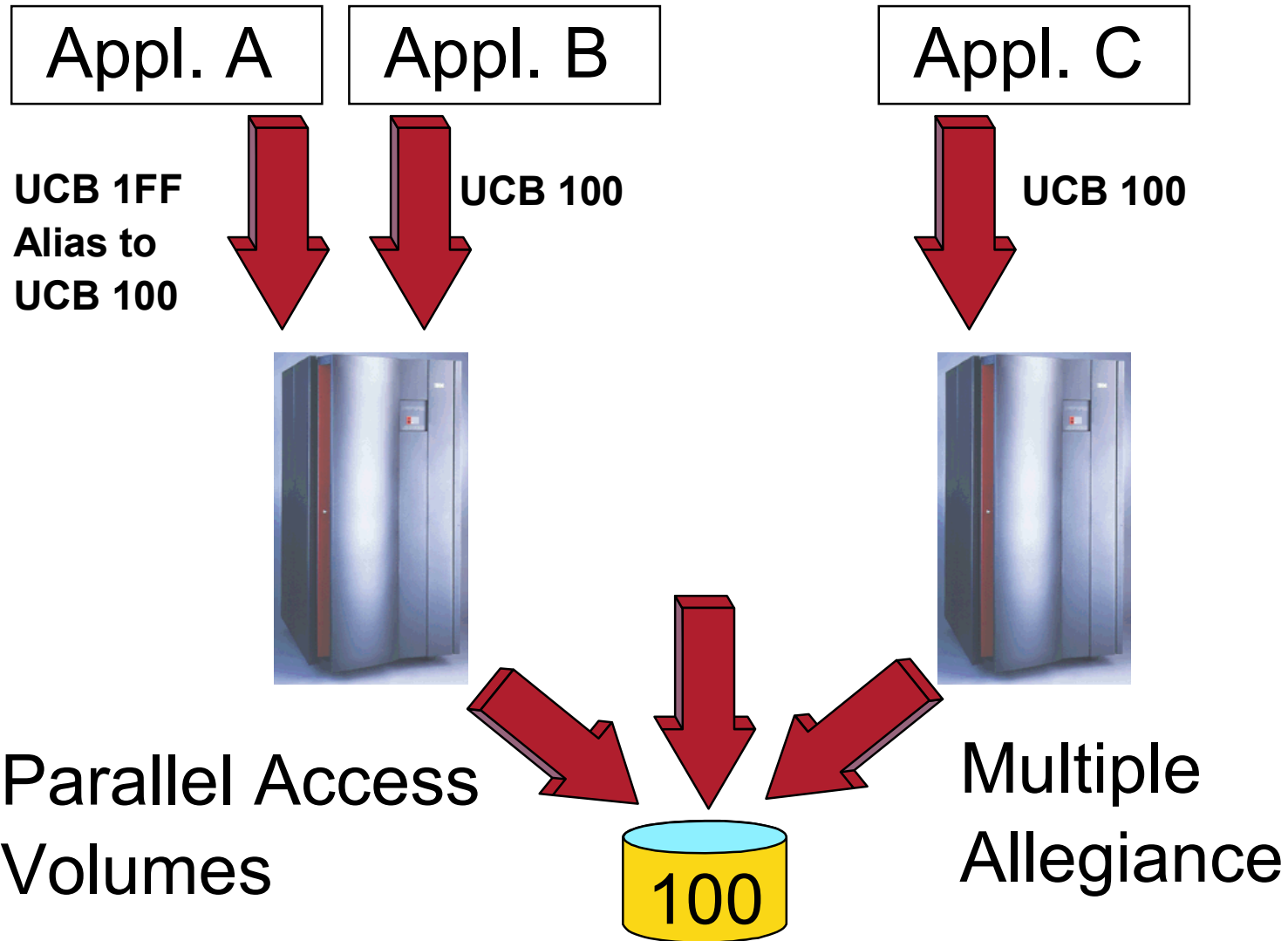


Traditional MVS Behavior



IBM Storage Solutions

Simultaneous I/O on ESS

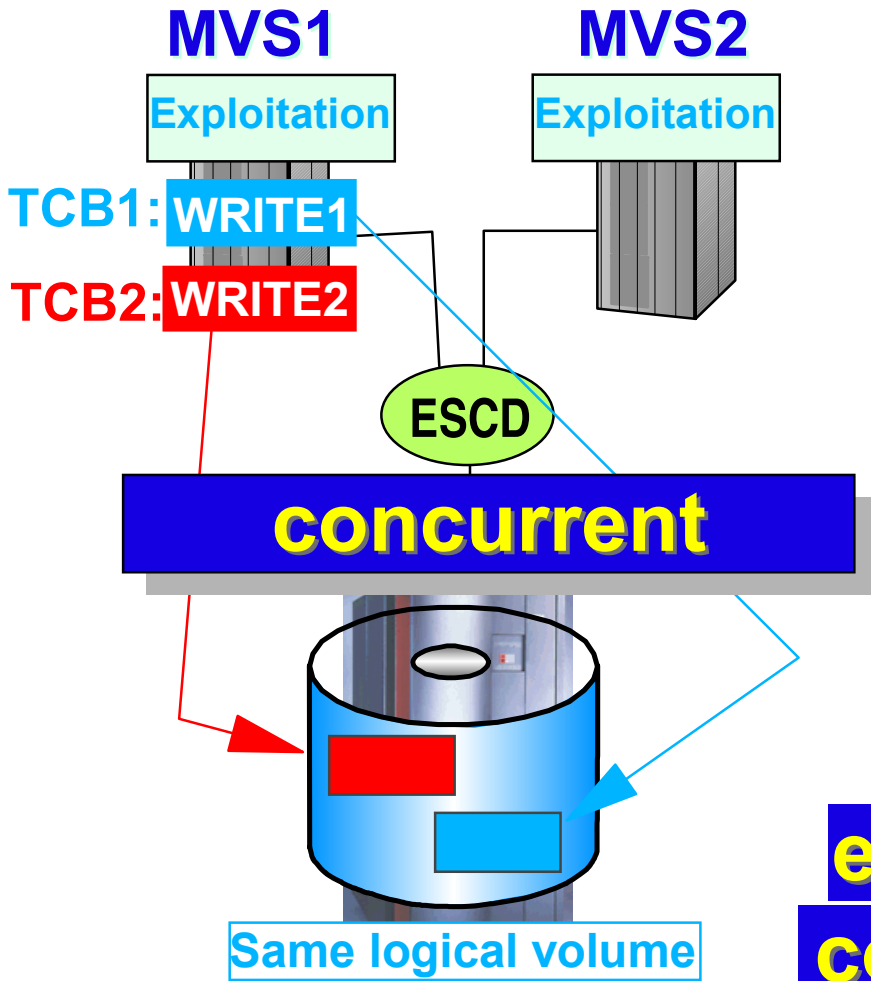


IBM Storage Solutions

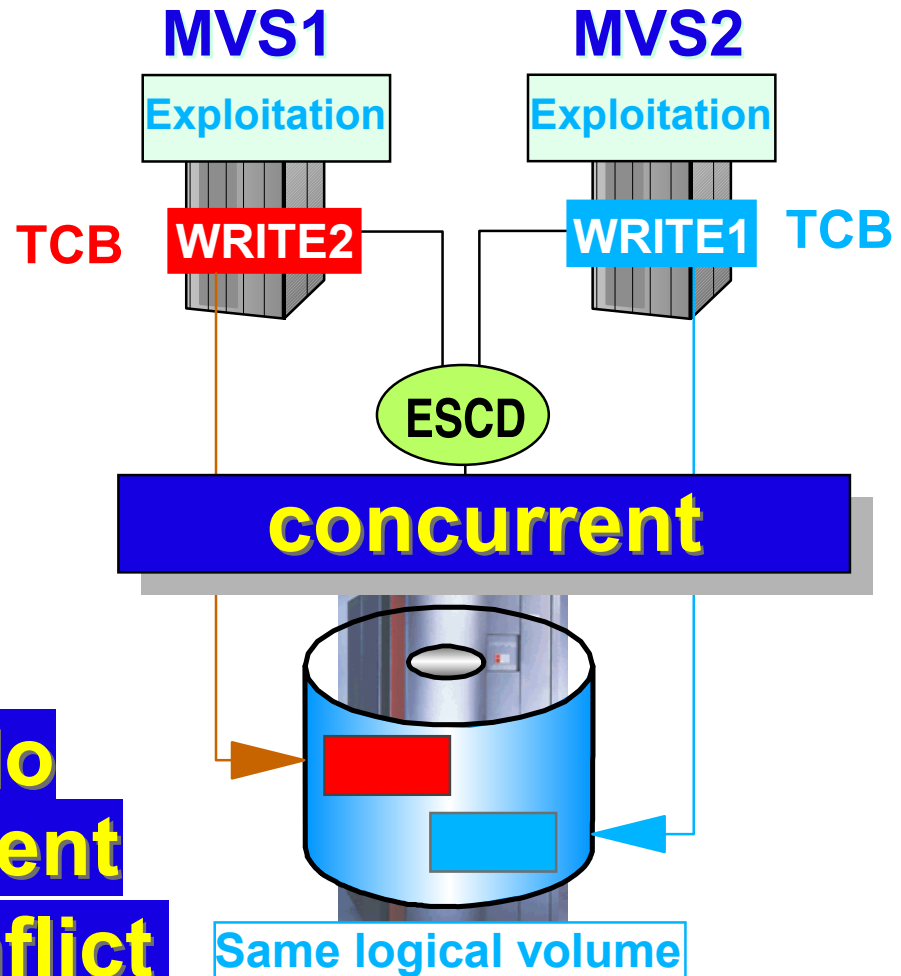


Concurrent Writes on ESS

Parallel Access Volumes



Multiple Allegiance



IBM Storage Solutions



Multiple Allegiance

- ★ Previously available for airlines RPQ/TPF
- ★ No special host software required
- ★ Allows simultaneous access of logical volume by different hosts simultaneously
- ★ Eliminates or sharply reduces PEND Time
 - Device Busy Delay
- ★ Simultaneous reads allowed to same Define Extent Range
 - Writes to same Def-Ext-Range cause I/Os to serialize
 - ESS guarantees data integrity
- ★ Measurement simulates DB2 data-mining application

IBM Storage Solutions



Benefits of Multiple Allegiance

	Host 1 (4K read hits)	Host 2 (32 record 4K read chains)
Max ops/sec - Isolated	767 SIOs/SEC	55.1 SIOs/sec
Max ops/sec - 100% Extent Conflicts	59.3 SIOs/SEC	54.5 SIOs/sec
Max ops/sec - full Multiple Allegiance	756 SIOs/SEC	54.5 SIOs/sec

IBM Storage Solutions



- ★ Allows multiple UCBs per logical volume
- ★ Volume previously seen as single resource serially reused
 - ▶ One I/O operation at a time was permitted to a volume
 - ▶ Problematic as devices became larger
- ★ SCSI allows command queuing
- ★ Provides multiple concurrent data transfers to/from a DASD volume
- ★ PAVs allow simultaneous access to logical volume by multiple users or jobs



Parallel Access Volume (*continued*)



- ★ Cache and RAID allow simultaneous accesses
 - Multiple I/Os to same volume can be serviced from cache or different DDM
- ★ Reads and Writes are simultaneous
 - Writes cause serialization to the same Define-Extent-Range
 - Eliminates or sharply reduces IOSQ
- ★ Chargeable feature

IBM Storage Solutions

A horizontal bar composed of four colored segments: purple, yellow, red, and green.

Parallel Access Volume (continued)

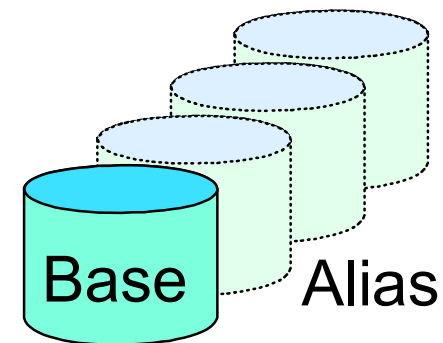


★ Multiple unit addresses (and therefore UCBs) per volume

- ▶ Trading off UCBs and subchannels for performance
 - Subchannels for aliases consume HSA like any other device
- ▶ Larger volume sizes should be considered
 - PAVs attack UCB contention
 - Volumes "carved" from RANK of > 100 GB

★ Base address

- ▶ Actual unit address of the volume
- ▶ One base address per volume
- ▶ Space associated with base



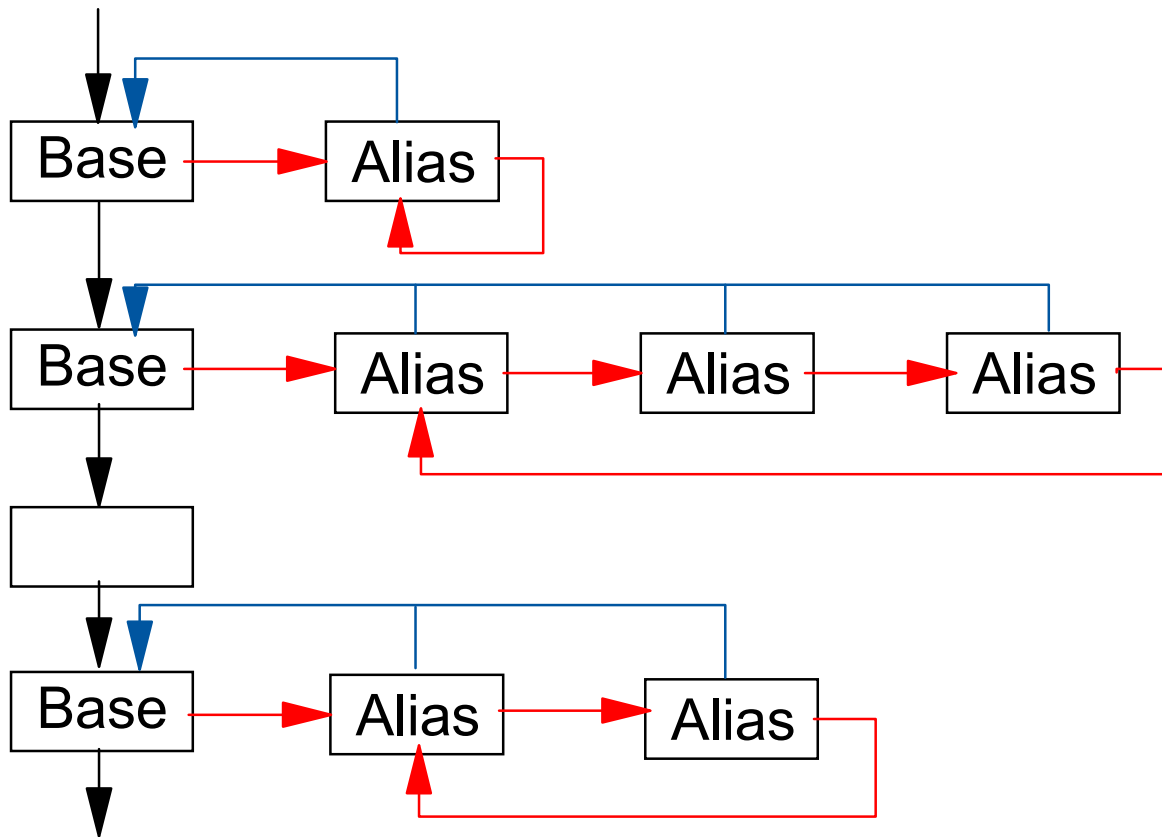
★ Alias address

- ▶ Maps to base address. I/O to an alias address runs against the base
- ▶ No physical HDD space associated with alias
- ▶ Aliases are visible only to IOS
- ▶ Alias UCBs are above the 16MB line

IBM Storage Solutions



UCB Chain with Base and Aliases



To IOS, each base and alias is a separate device. Thus, a base UCB could have an active request (be "busy") even when an alias for that same base is idle.

IBM Storage Solutions



PAV Types

★ Static PAVs

- ▶ Association between PAV-base and its PAV-aliases is predefined through ESS Specialist
- ▶ PAV assignments can be changed as needed without IML
- ▶ Static PAVs are supported by OS/390 V1 Release 3 and DFSMS/MVS 1.3 with PTFs

★ Dynamic PAVs

- ▶ Association between PAV-base and its PAV-aliases is dynamic
- ▶ WLM in goal mode manages the assignment of alias addresses
- ▶ WLM instructs IOS when to reassign an alias
- ▶ Dynamic PAVs are supported by OS/390 V2 Release 7 and DFSMS/MVS 1.5

Verifying PAVs

DEVSERV QPAVS

- ▶ DS QPAVS,D200,VOLUME

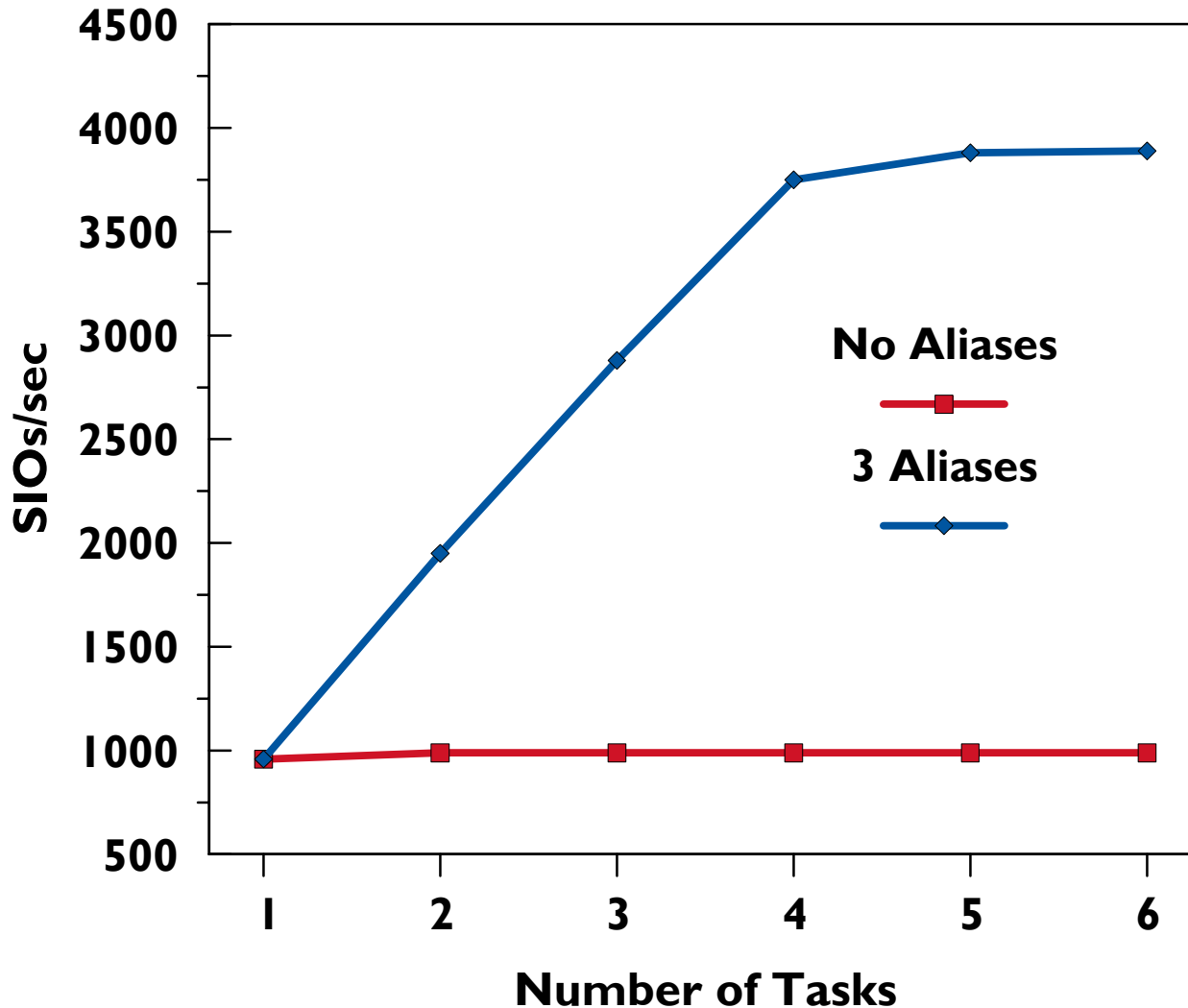
```

IEE459I 08.20.32 DEVSERV QPATHS 591
      Host                               Subsystem
      Configuration                       Configuration
      -----
UNIT                                     UNIT   UA
NUM. UA  TYPE          STATUS          SSID  ADDR.  TYPE
----- --  -----  -
D200 00  BASE
D2FE FE  ALIAS-D200
D2FF FF  ALIAS-D200
***          3 DEVICE(S) MET THE SELECTION CRITERIA
  
```

IBM Storage Solutions



Effects of PAVs on read hit rates for a single volume

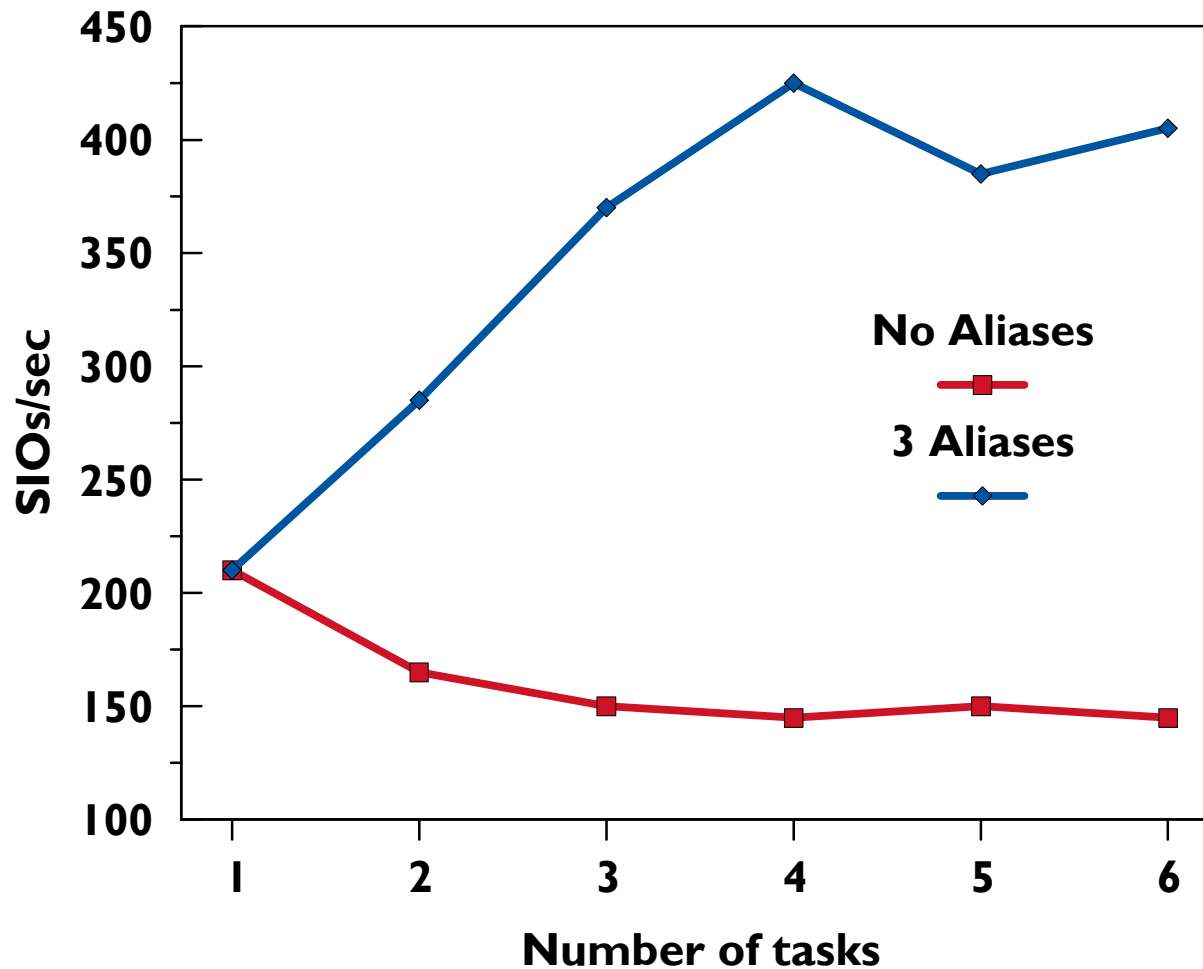


- 100% Cache Read Hit
- No extent conflicts in this test
- This shows a channel limit at about 1000 ops/sec

IBM Storage Solutions



Effects of PAVs on read miss rates for a single volume

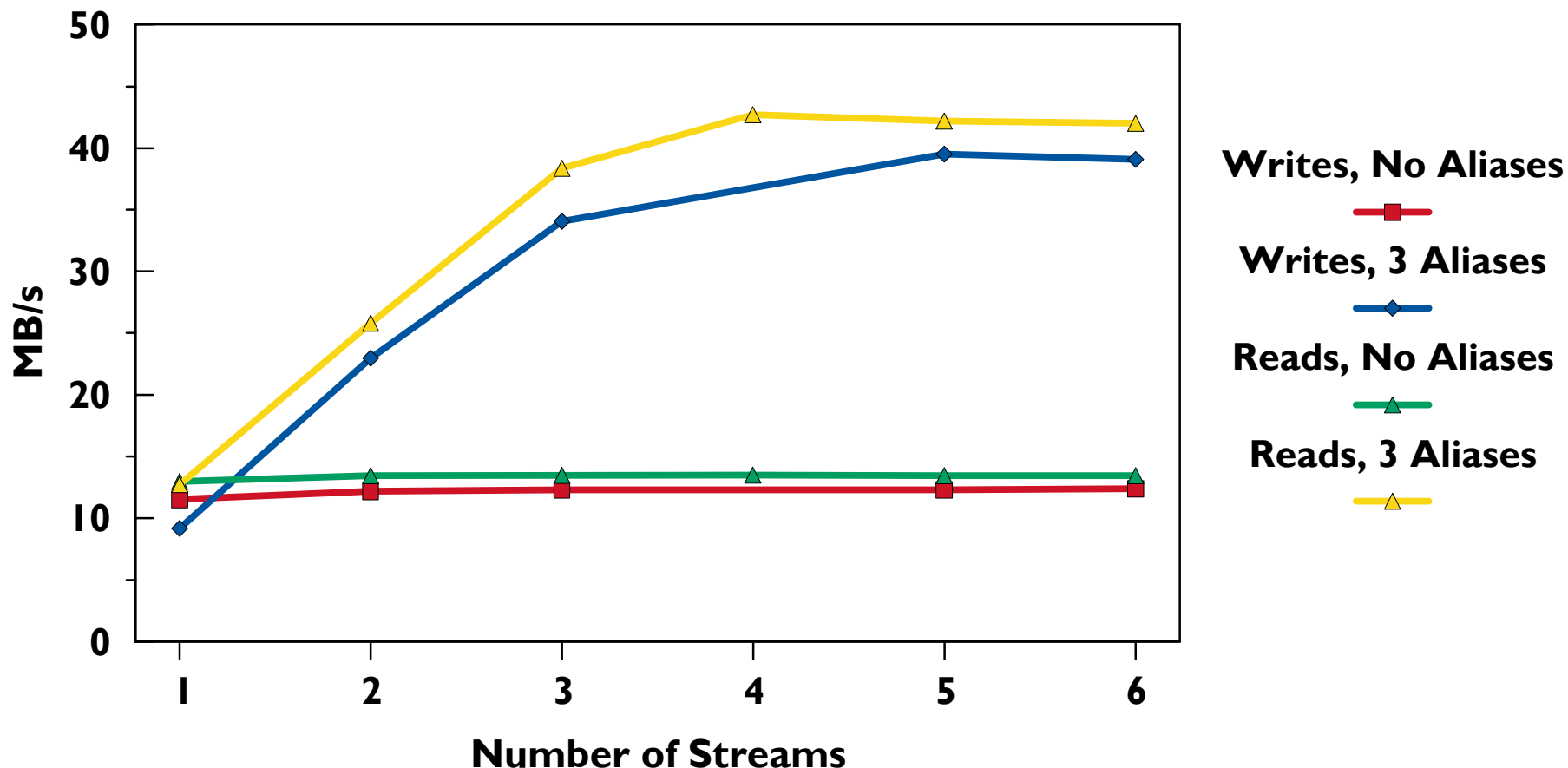


- 100% Cache Read Miss
 - ▶ Note: Seek distance increases as # Tasks increase due to test design
- Volumes are striped across all HDDs in Rank, so multiple arms can be moving for same volume
- Multiple I/Os can be queued against same HDD, improves HDD operations per second
- No extent conflicts in this test

IBM Storage Solutions



Effect of PAVs on QSAM job streams against a single volume



Notes:

With PAVs - maximum rate is approximately device speed for a RAID5 array

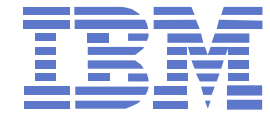
Without PAVs - maximum rate is limited to ESCON channel speed

Workload was 27K QSAM with bufno=5, no extent conflicts

IBM Storage Solutions

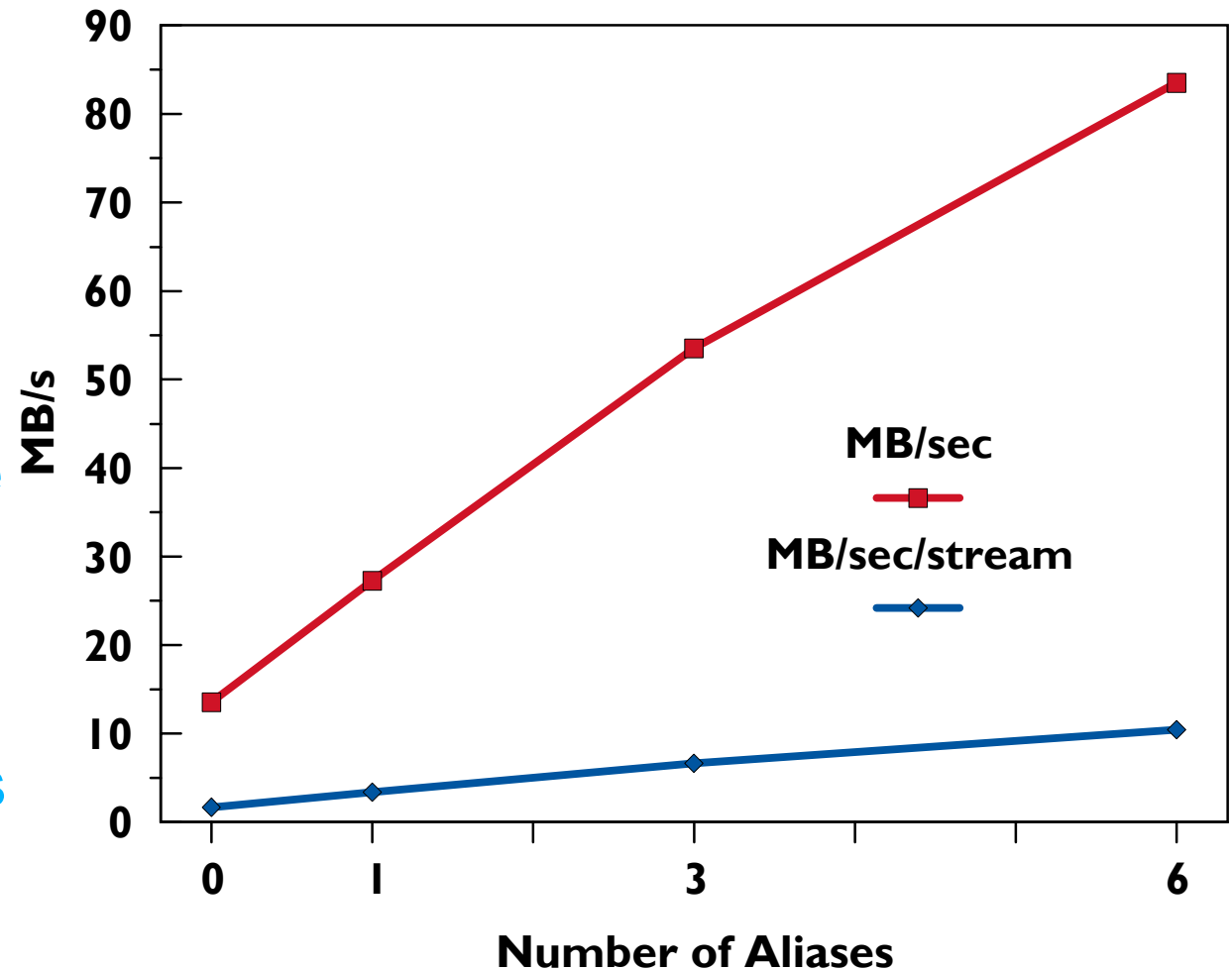


PAV, Reads against a single dataset



Experiment:

- 8 simultaneous reads against a single data set on a volume
- QSAM 27,648 byte blocks, bufno=5
- In all cases, cache hit ratio was 100%
- 8 available channels



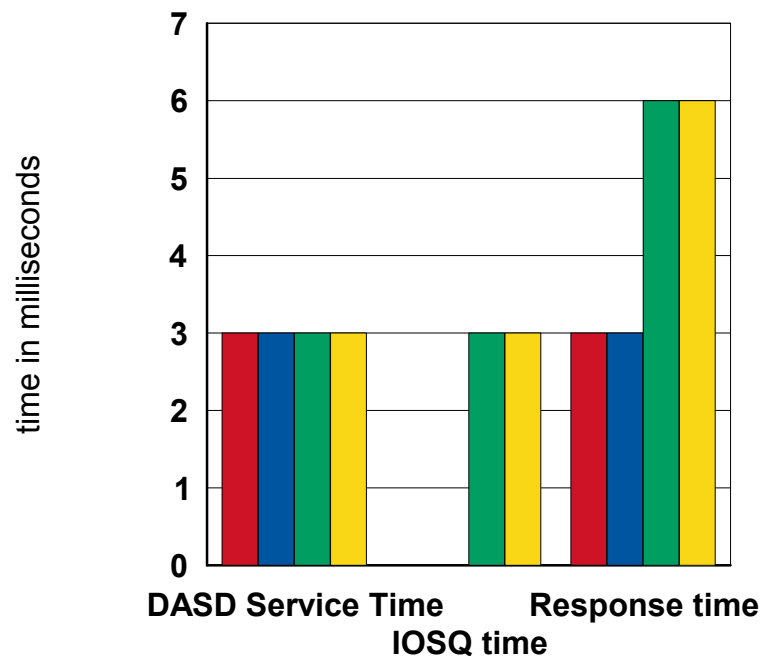
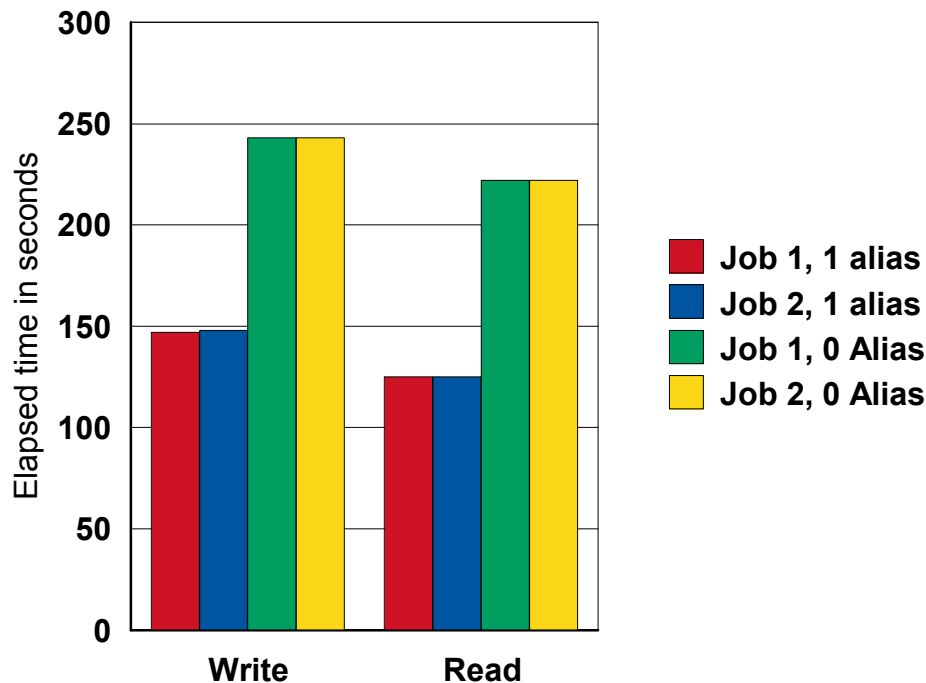
IBM Storage Solutions



ESP PAV Test

Test scenario:

2 Jobs with run in parallel against a volume with PAV (1 Alias) and without PAV. The first jobstep creates a seq. dataset with IEBDG (9,000,000 records, recln=80, Size of the ds: app. 1665 cyl), the second step reads this ds with ICKDSF-REPRO. Only these 2 jobs are running against the ESS.

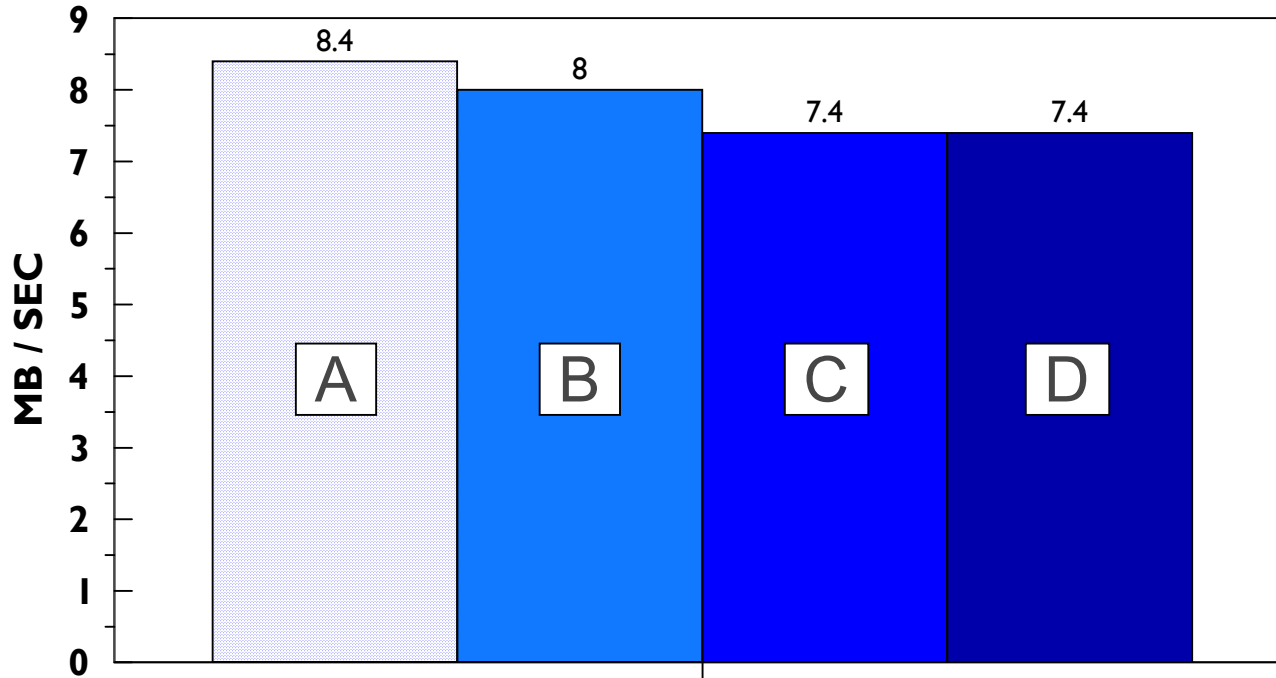


IBM Storage Solutions



DB2 Logging Rates on ESS

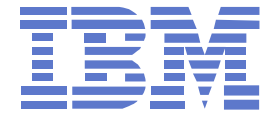
- SHARK - SINGLE
- SHARK - DUAL - 1 LCU / 2 VOL
- SHARK - DUAL - 2 LCU / 2 VOL
- SHARK - DUAL - 1 LCU / 1 VOL



- A - Single logging. Log rate was 8.4 MB/sec.
- B - Dual logging. Logs were on separate LCU's, rate was 8.0 MB/sec. Best layout for dual logging. By LCU I mean CUADDR=XX.
- C - Dual logging. Logs on the same LCU but different volumes. Logging rate went down about 7.5% compared to case B. Not recommended.
- D - Dual logging. Logs on the same volume. This is REALLY not recommended, but notice that there is no degradation over case C. PAV really works, IOSQ remains zero. RAMAC3 experience--log rate was 3.4 MB/sec, went down to 2.2 MB/sec with logs on the same volume, and IOSQ increased dramatically.

IBM Storage Solutions





One ESP Test with MA & PAV

An ESP customer recently conducted a test comparing Shark to a competitor using SYNC GENER WRITE jobs.

When running 2 jobs, both datasets reside on the same volume.

Based on this test results, The customer used the following words regarding the Shark box: "promising" and "impressed" !

They even rated performance as "1" on the weekly ESP rating.

The customer defined 2 aliases for all base addresses.

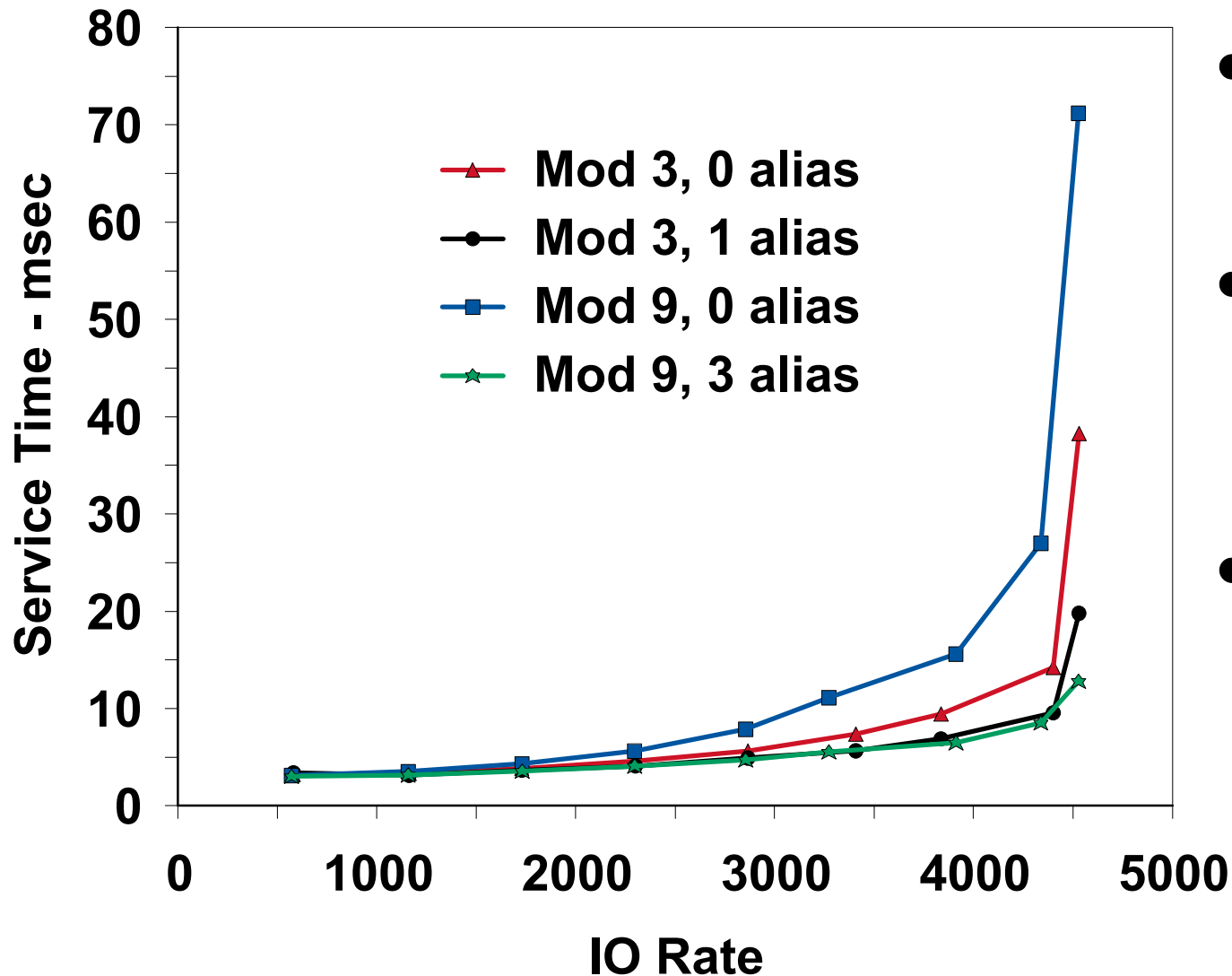
This is the result:

	Average I/O R/T on Shark	Average I/O R/T on Competitor
1 job, doing multitrack I/O	57ms	78ms
2 jobs running on 1 MVS system	56ms	166ms
2 jobs running on 2 MVS systems	56ms / 56ms	216ms / 128ms

IBM Storage Solutions



Mod 3 vs Mod 9 with aliases - Cache Standard

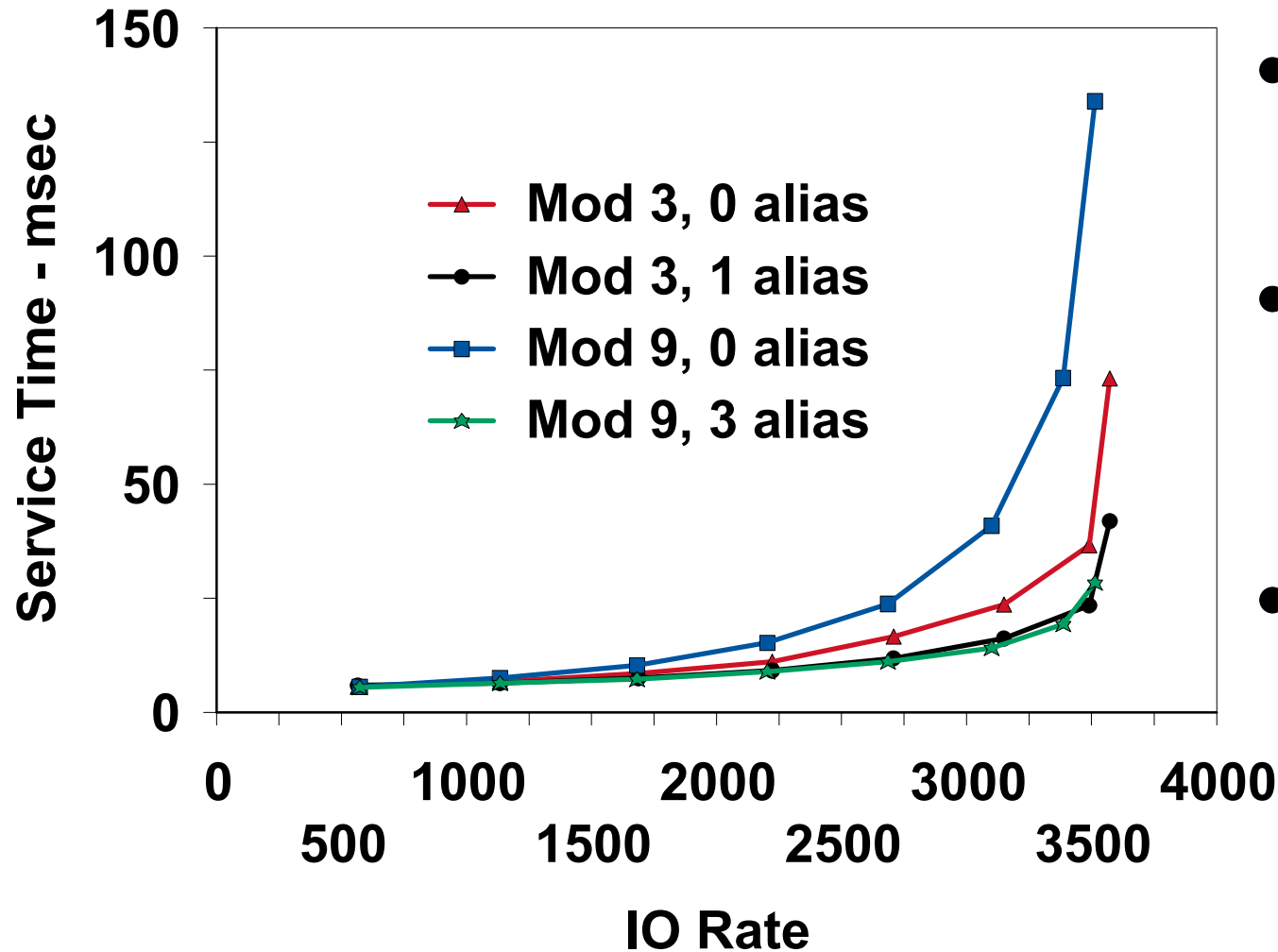
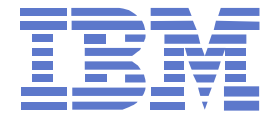


- Run was against a single ESS cluster
- Mod 9 volumes had 3 times the access data of Mod 3 volumes
- Workload is a IMS transaction workload

IBM Storage Solutions



Mod 3 vs Mod 9 with aliases - Cache Hostile



- Run was against a single ESS cluster
- Mod 9 volumes had 3 times the access data of Mod 3 volumes
- Workload is a DB2 workload with a low cache hit ratio

IBM Storage Solutions



Performance Enhanced CCWs

★ Track Commands

- ▶ Read Track Data and Write Track Data CCWs
- ▶ Will be used by media manager to reduce ESCON protocol for multiple record transfer chains
- ▶ Measurements on 4K records using an EXCP channel program showed a 15% reduction in channel overhead for the Read Track Data CCW
- ▶ Used by IBM software at ESS exploitation levels
- ▶ Disclosed to ISVs
- ▶ VM/ESA will allow guest machines to use it

★ Extent Limiting

- ▶ Benefits PAV & Multiple Allegiance
- ▶ Limit the Define-Extent-Range in the Define Extent command
 - It now usually contains the dataset extent
 - Limit the extent to a minimum
 - Will allow more concurrent I/Os when writes are executed
- ▶ Used by IBM software at ESS exploitation levels
- ▶ Disclosed to ISVs

IBM Storage Solutions



I/O Priority Queuing

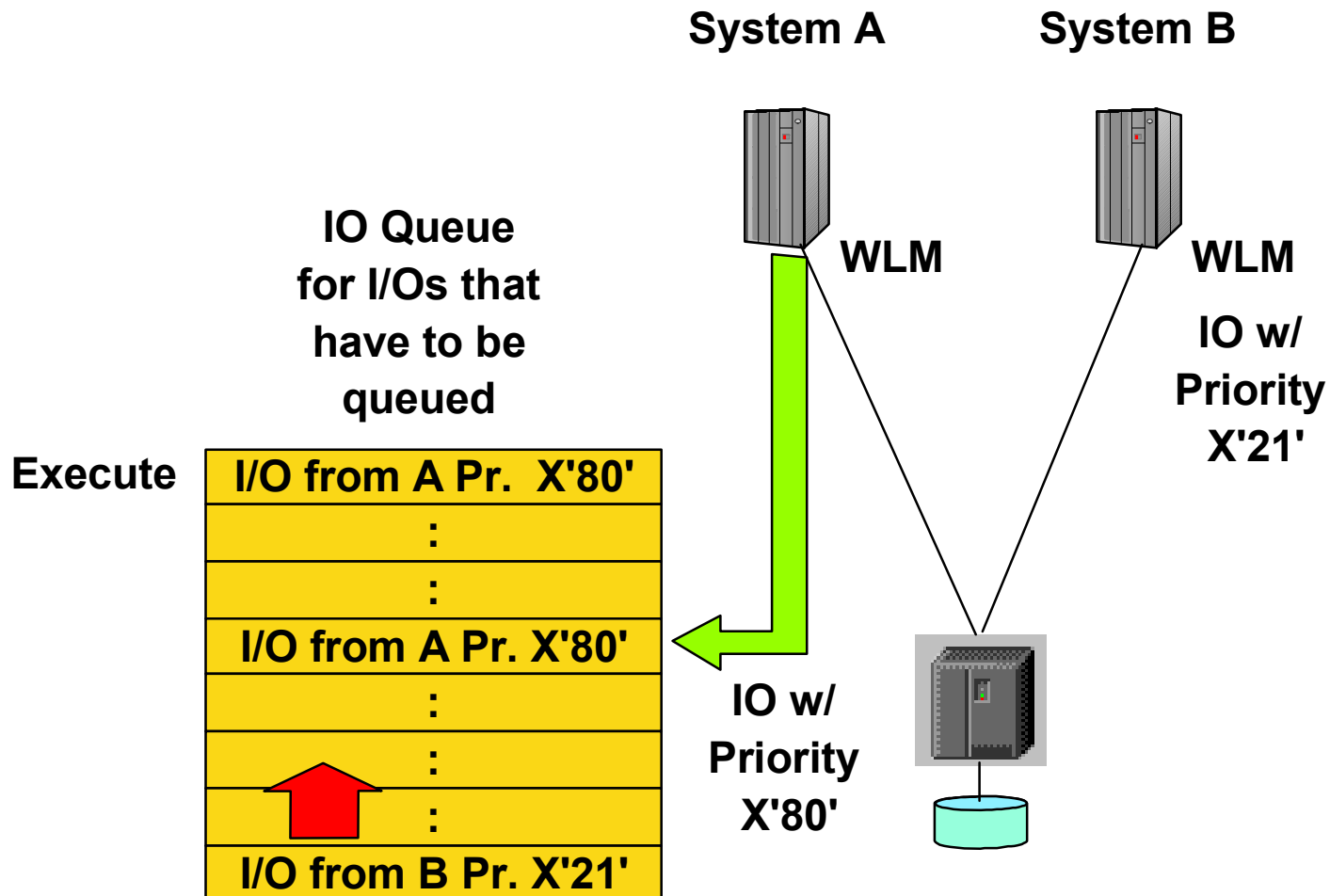
- ★ If I/Os cannot run in parallel, they are queued
 - ▶ ESS can internally queue I/Os
 - ▶ Avoids overhead associated with posting "device busy" status and redriving the channel program
 - ▶ Eliminates race conditions when one system is faster than the other

- ★ Priority queuing
 - ▶ I/Os can be queued in a priority order
 - ▶ A priority byte can be set in the CCW to prioritize I/Os
 - ▶ OS/390's Workload Manager sets the I/O priority byte when running in goal mode if WLM I/O priority management is enabled; otherwise no priority is used
 - ▶ I/O priority for systems in a sysplex
 - ▶ Fair share for each system
 - ▶ Requires OS/390 V1R3 or higher

IBM Storage Solutions



I/O Priority Queuing



IBM Storage Solutions



Part II:

WLM Management of PAVs

IBM Storage Solutions



WLM Management of PAVs

- ★ Topics
 - ▶ Overview of WLM function
 - ▶ New externals
 - ▶ RMF support

WLM Management of PAVs

- ★ PAVs must be defined to MVS and the Storage Server
 - ▶ The initial association of aliases to a base is done in the Storage Server using ESS Specialist
 - ▶ MVS Hardware Configuration Definition (HCD) must be used to define PAV base and alias addresses:
 - for example, 3390B for base and 3390A for alias
 - **WLMPAV = YES | NO** enables / disables WLM management
 - The default is WLMPAV = YES, enabling WLM management
 - ▶ HCD definitions must match ESS definitions

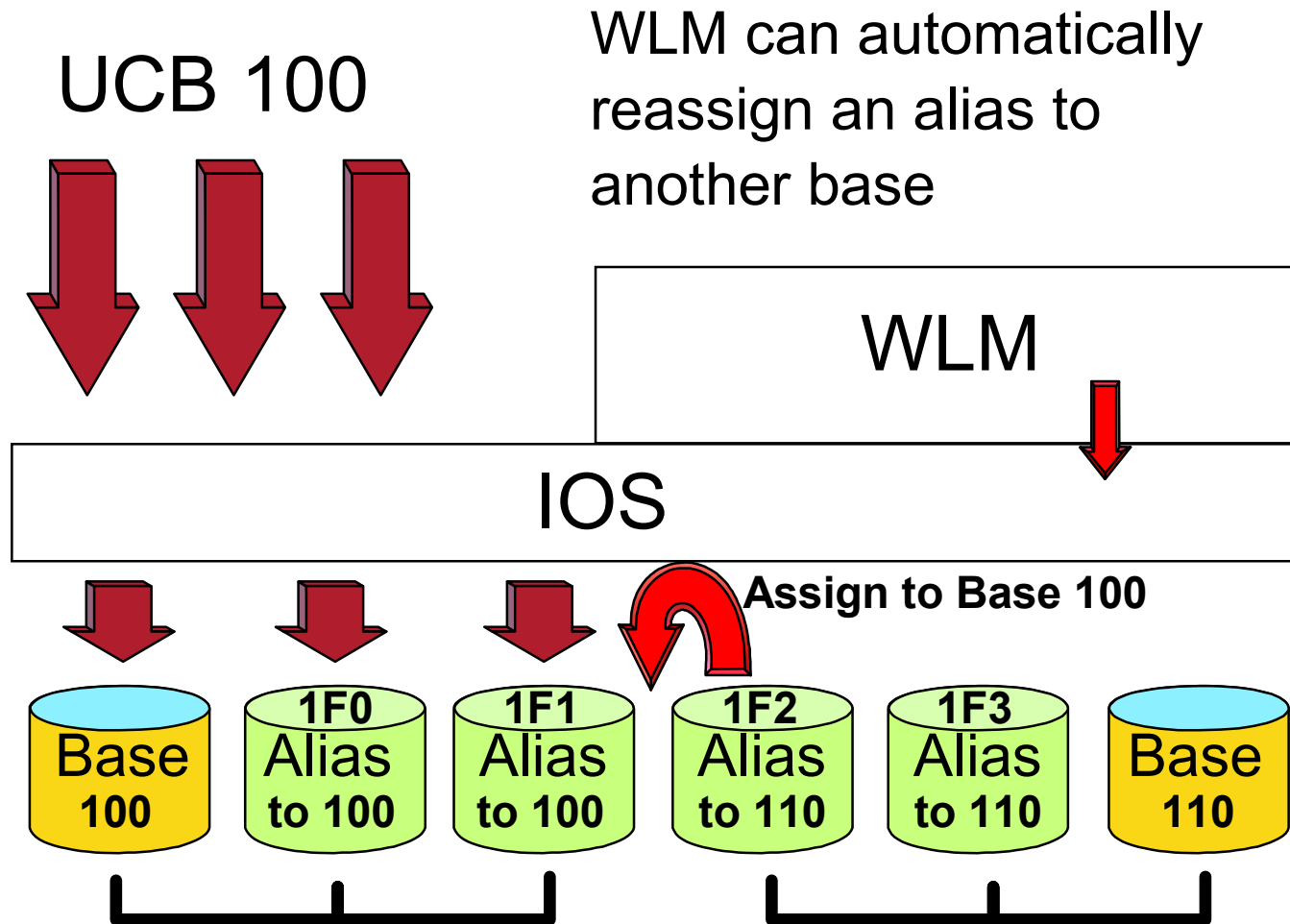
- ★ WLM senses the aliases initially assigned to a base

- ★ WLM can then reassign aliases to another base

IBM Storage Solutions



Dynamic Reassignment of Aliases

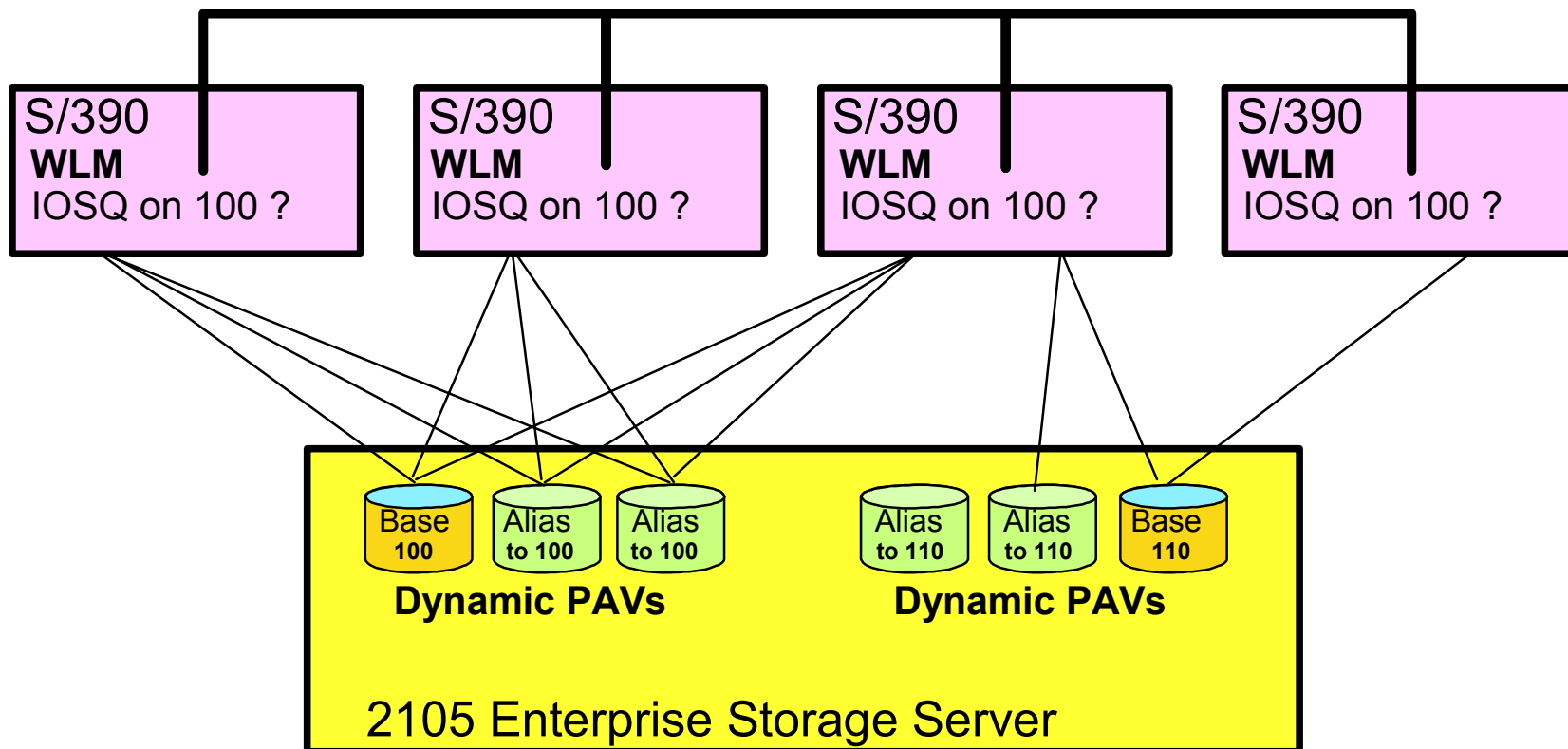


IBM Storage Solutions



Alias Management in a Sysplex

- ✓ WLMs exchange performance information
- ✓ Goals not met because of IOSQ delay?
- ✓ If goals met, can aliases be used more efficiently?
- ✓ Who can donate an alias?

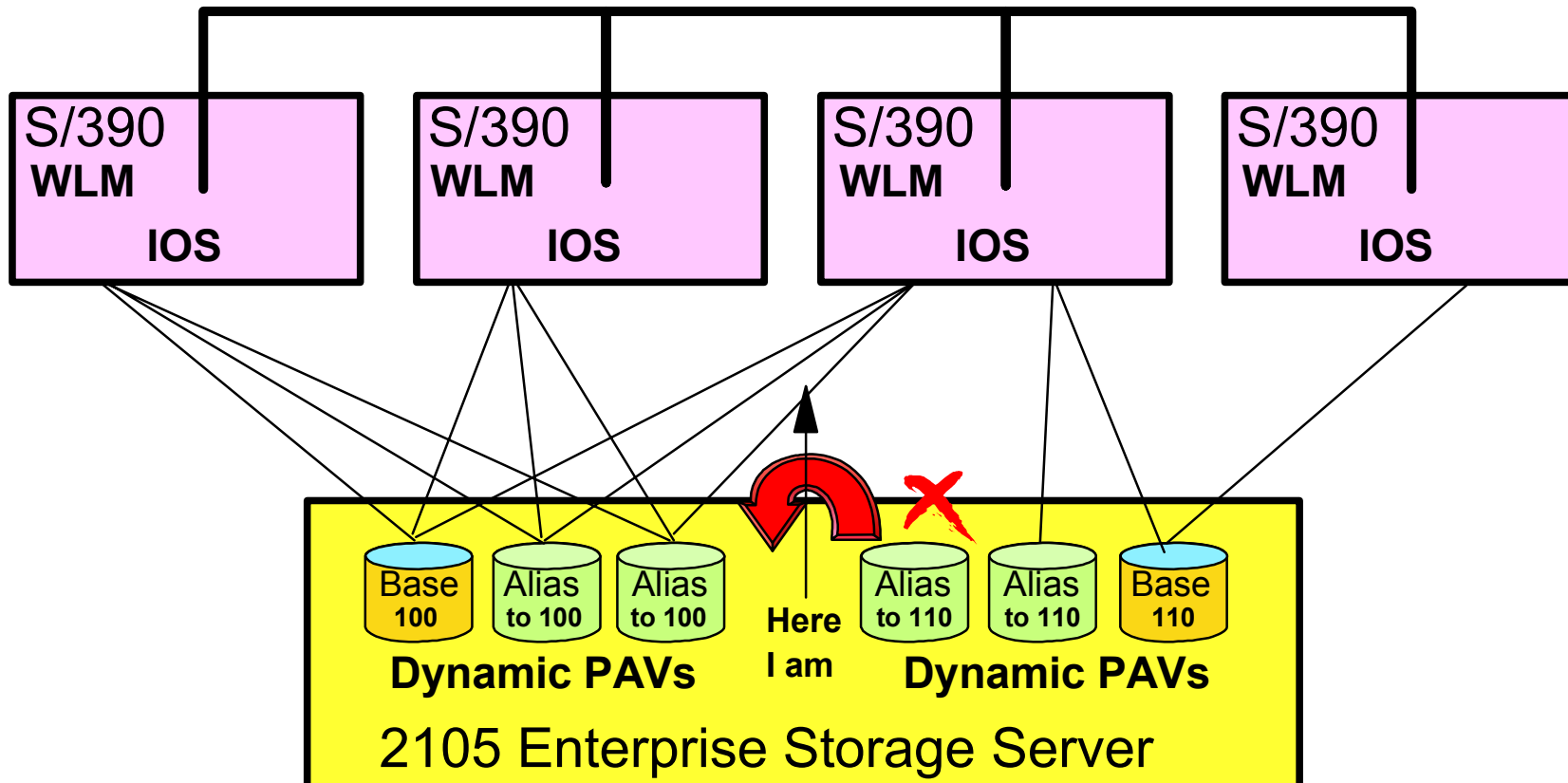


IBM Storage Solutions



Alias Reassignment in a Sysplex

- ✓ To reassign an alias, a token is used by IOS to serialize the alias move across the sysplex.
- ✓ Systems informed when move complete.



IBM Storage Solutions



WLM Algorithms for Dynamic PAVs

- ★ Goal algorithm - Help service class period meet its goal
 - ▶ Service class period missing goal
 - ▶ Significant IOS queuing on PAV device used by period
 - ▶ Not overly constrained in control unit
 - ▶ Alias available from idle device
 - or one used by work of lesser or equal importance
 - ▶ Twice per minute; multiple aliases can be moved at once
 - ▶ Takes precedence over efficiency algorithm

- ★ Efficiency algorithm - Reduce overall IOS queuing
 - ▶ Significant IOS queuing on a PAV device
 - ▶ Alias available from another device
 - ▶ But don't increase queuing on donor device
 - ▶ Once per minute
 - ▶ Conservative
 - ▶ Goal mode only

IBM Storage Solutions

A horizontal bar composed of four colored segments: purple, yellow, red, and green.

PAV Management Switches

	WLM I/O mgmt	WLM Alias Tuning	HCD WLMPAV
Scope	sysplex	sysplex	one device within one system
Default	Disabled	Disabled	Enabled
Value that disables WLM management	No	No	No
Most important gotchas	Velocity goals require recalibration when value is changed	Changing value to Yes upgrades functionality level of service definition, which causes operational problems if any system in the sysplex does not support new f-level	If <u>any</u> system in the sysplex specifies Yes, alias is dynamic



Allowing WLM to Manage PAVs

- ★ WLM PAV Management is called Dynamic Alias Mgmt
- ★ To enable Dynamic Alias Mgmt you need *all* of these:
 - ▶ WLM in Goal Mode
 - ▶ WLM I/O Priority Management = YES (OS/390 R3 option)
 - ▶ WLM Dynamic Alias Management = YES (OS/390 R7 option)
 - ▶ HCD WLMPAV = YES on at least one system for every ESS device
 - ▶ HCD definitions must match ESS definitions
- ★ WLM options are located in the WLM administrative application, on the "Service Coefficients/Options" panel
 - ▶ Changing WLM options affects functionality levels!

IBM Storage Solutions



Mixing PAV Types is Not Recommended

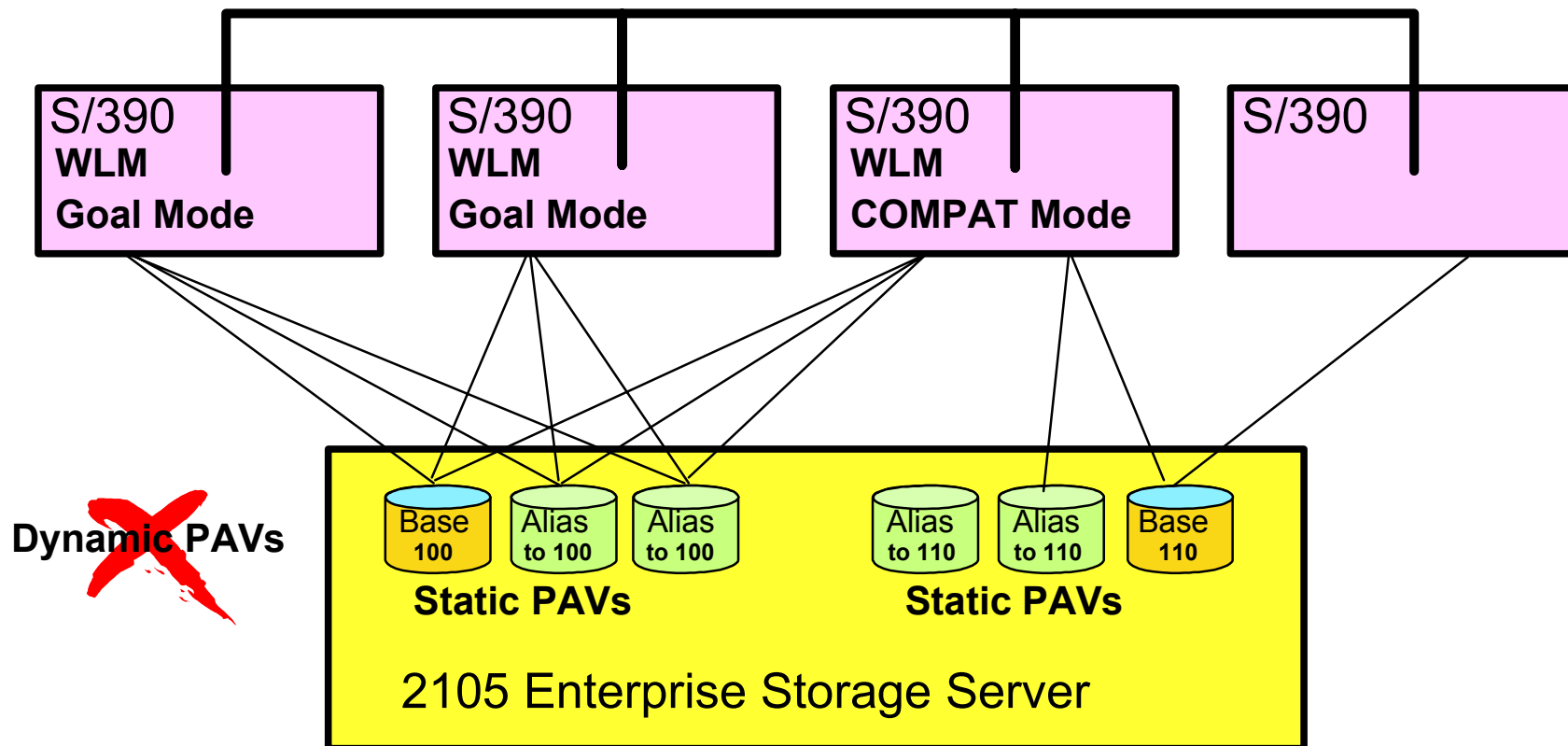


- ★ For the same volumes, do not use dynamic (WLM-managed) PAVs on one system unless all sharing systems use dynamic PAVs!
 - ▶ Systems not supporting dynamic PAVs still recognize and tolerate alias changes
 - pre-V2R7, WLM compatibility mode, or non-OS/390
 - ▶ But WLM decisions will not take into account the I/O activity of these systems
 - ▶ Customer options:
 - Upgrade your systems to OS/390 V2 Release 7
 - Switch to goal mode, if you already have V2 Release 7
 - Set HCD WLMPAV option to OFF for devices used by mixed systems



Mixing of PAV Types

- ✓ HCD option to enable / disable WLM alias management by base device
- ✓ WLM option to disable dynamic alias management for sysplex
- ✓ Do not mix PAV types



IBM Storage Solutions



WLM panel with PAV option

```

Coefficients/Options  Notes  Options  Help
-----
                Service Coefficient/Service Definition Options
Command ===> _____

Enter or change the Service Coefficients:

CPU . . . . . 1.0      (0.0-99.9)
IOC . . . . . 0.5      (0.0-99.9)
MSO . . . . . 0.0000   (0.0000-99.9999)
SRB . . . . . 1.0      (0.0-99.9)

Enter or change the service definition options:

I/O priority management . . . . . YES   (Yes or No)
Dynamic alias management . . . . . NO    (Yes or No)

```

- ✓ Sysplex-wide default is "NO".
- ✓ Specifying "YES" changes functionality level to LEVEL008.



ESS Specialist: Add Alias

Configure LCU 3

Device Adapter Pair: 2, Cluster 2

LCU Settings	3990-6	Logical Control Unit Emulation Mode
	AA03	Subsystem Identifier (SSID)
	enabled	Parallel Access Volumes (PAV)
	63	Starting PAV Address

LCU Devices

Device ID	Base/Alias	Storage Type	Volume Type	Cylinders	Capacity	Location
00	base (1 aliases exist)	RAID Array	3390-3	3339	2.87 GB	RAID Array: 02, Vol: 000
01	base (1 aliases exist)	RAID Array	3390-3	3339	2.87 GB	RAID Array: 02, Vol: 001
02	base (1 aliases exist)	RAID Array	3390-3	3339	2.87 GB	RAID Array: 02, Vol: 002
03	base (2 aliases exist)	RAID Array	3390-3	3339	2.87 GB	RAID Array: 02, Vol: 003
04	base (2 aliases exist)	RAID Array	3390-3	3339	2.87 GB	RAID Array: 02, Vol: 004
05	base	RAID Array	3390-3	3339	2.87 GB	RAID Array: 02, Vol: 005
06	base	RAID Array	3390-3	3339	2.87 GB	RAID Array: 02, Vol: 006
07	base	RAID Array	3390-3	3339	2.87 GB	RAID Array: 02, Vol: 007
08	base	RAID Array	3390-3	3339	2.87 GB	RAID Array: 02, Vol: 008

- Add Parallel Access Volume aliases to each selected volume
 Delete selected Parallel Access Volume alias(es)

HCD panel with PAV option

```

----- View Device Parameter / Feature Definition -----
                                                                    Row 1 of 6
Command ==> _____ Scroll ==> CSR

Configuration ID . . : B710
Device number . . . : 2300          Device type . . . : 3390B
Generic / VM device type . . . . : 3390

ENTER to continue.

Parameter/
Feature      Value      Req.  Description
OFFLINE      No          Device considered online or offline at IPL
DYNAMIC      Yes         Device supports dynamic configuration
LOCANY       Yes         UCB can reside in 31 bit storage
WLMPAV       Yes         Device supports work load manager
SHARED       Yes         Device shared with other systems
SHAREDUP     No          Shared when system physically partitioned
***** Bottom of data *****

```

WLMPAV default is "YES"

Other PAV Support

- ★ D M (display matrix)
- ★ D U (display units)
- ★ CONFIGxx
 - PAV keyword causes d m=config(xx) command to report unbound aliases
- ★ GTF I/O trace
 - When tracing a PAV base, also traces aliases bound to it
- ★ Programming interfaces:
 - UCINFO PAVINFO Reports on aliases for a PAV base
 - UCINFO DEVINFO Identifies PAV devices and # aliases
 - ENF Signal 33 Signals unbind and bind of alias

Displaying PAVs

D MATRIX

```

S607  d m=dev(411)
S607  IEE174I 17.09.28 DISPLAY M 506
DEVICE 0411      STATUS=ONLINE
CHP                60 70 80 90
PATH ONLINE        Y  Y  Y  Y
CHP PHYSICALLY ONLINE Y  Y  Y  Y
PATH OPERATIONAL   Y  Y  Y  Y
PAV BASE AND ALIASES 4

S607  d m=dev(41f)
S607  IEE174I 17.11.48 DISPLAY M 512
DEVICE 041F      STATUS=ALIAS OF BASE 0413

```



Summary

- ★ New functions improve I/O performance and management of mixed workloads:
 - ▶ Multiple allegiance
 - ▶ Parallel access volumes (PAVs)
 - ▶ I/O priority queuing

- ★ WLM provides easier management of PAVs
 - ▶ By default, WLM management of PAVs is enabled
 - ▶ Sysplex-level toggle in WLM
 - ▶ Device-level toggle in HCD
 - ▶ Device definitions must match between HCD and ESS Specialist
 - Device address
 - Volume type
 - Base vs. alias

Want to know more?

- ★ Website: www.ibm.com/storage/ess
- ★ Product book: *ESS Introduction and Planning* (GC26-7294)
- ★ Redbooks: www.redbooks.ibm.com
 - ▶ *IBM Enterprise Storage Server* (SG24-5465)
 - ▶ *Implementing the IBM ESS in Your Environment* (SG24-5420)
- ★ IBM Service Link has list of supporting PTFs:
 - ▶ PSP bucket upgrade ID: 2105device
 - ▶ Subset ID: 2105MVS/ESA