z/VSE Parallel Access Volume Support

Vischer, Andreas
z/VSE I/O Lead Developer
vischer@de.ibm.com

IBM Germany Research & Development GmbH
Schoenaicherstr. 220
71032 Boeblingen
Germany

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries. For a complete list of IBM Trademarks, see www.ibm.com/legal/copytrade.shtml: e-business logo, eServer, FICON, IBM, IBM Logo, VSE/ESA, z/VSE, System z

Overview

Parallel access volumes (PAV) allow your system to access volumes in parallel when you use an IBM System z host.

PAV is a feature of IBM System Storage Disk Systems which enables more than one I/O operation to be processed for a single logical device. The use of PAVs can significantly reduce device queue delays to busy devices (all I/O operations are normally processed from a single queue). PAVs are managed by creating multiple addresses for a single logical device.
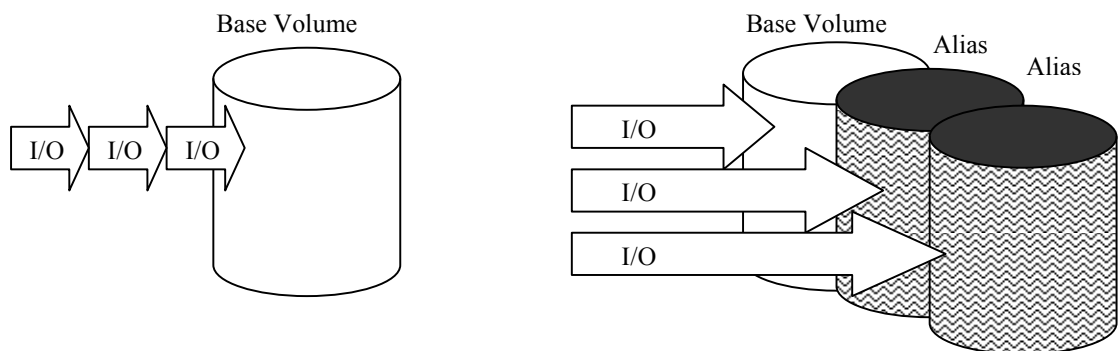
 PAV is an optional licensed feature of the IBM System Storage DS8000 and DS6000 series, available with the PAV indicator feature, and corresponding DS8000/DS6000 series function authorization. PAV also requires the purchase of the FICON/ESCON Attachment feature in the Turbo models.

You must configure both your storage unit and operating system to use PAVs. You can use the logical configuration definition to define PAV-bases, PAV-aliases, and their relationship in the storage unit hardware. This unit address relationship creates a single logical volume, allowing concurrent I/O operations.
PAV can improve the performance of large volumes. You can get better performance with one base and two aliases on a 3390 Model 9 than from three 3390 Model 3 volumes with no PAV support.

The storage unit supports concurrent or parallel data transfer operations to or from the same volume from the same system or system image on System z hosts. A System z with PAV software support enables multiple users and jobs to simultaneously access a logical volume. Read and write operations can be accessed simultaneously to different domains. (The domain of an I/O operation is the specified extents to which the I/O operation applies.)
Today, a PAV volume has one volume serial number and multiple subchannels.
When accessing a PAV volume in parallel, reads can be processed simultaneously from the subchannel point of view, as well as writes to different areas on the volume.
z/VSE will balance I/O traffic from different tasks upon the available base and alias volumes by an internal algorithm and considers of course load balancing.

Prerequisites and Activation of the z/VSE PAV Support

- PAV feature purchased, installed and enabled on the IBM System Storage Disk system.
  (DS8000 Series Feature Codes 78xx / DS6000 Series Feature Codes 51xx)

- Base and alias devices defined in the System z IOCDS.
  (Base and alias devices are transparent to the IOCDS, and therefore both fully qualified
  IODEVICE definitions in the IOCDS)

- Base device defined in the z/VSE IPL start up procedure.
  (corresponding ADD statement)

- Starting of the z/VSE PAV support via SYSDEF SYSTEM,PAV=START command (see
  following section)

For more information, examples and a more detailed description about PAV in the z/VSE context please
see *z/VSE Administration SC33-8304-02 .*

New and Extended Commands:

- SYSDEF SYSTEM
- SIR
- VOLUME
- QUERY SYSTEM

SYSDEF SYSTEM (new options)
The SYSDEF SYSTEM command has been extented to start/stop PAV processing on the z/VSE system.

```
AR 0015 1I40I   READY
sysdef system,pav=Start
AR 0028 1K01I ALIAS DEVICE 778 FOR BASE 777 ESTABLISHED
AR 0028 1K01I ALIAS DEVICE 77E FOR BASE 777 ESTABLISHED
AR 0028 1K01I ALIAS DEVICE 779 FOR BASE 777 ESTABLISHED
AR 0028 1K01I ALIAS DEVICE 77A FOR BASE 777 ESTABLISHED
AR 0028 1K01I ALIAS DEVICE 77B FOR BASE 777 ESTABLISHED
AR 0015 1I40I   READY

AR 0015 1I40I   READY
sysdef system,pav=Stop
AR 0015 1I40I   READY
```

QUERY SYSTEM (new command / output extension)
The QUERY SYSTEM output will also display the current PAV setting on the z/VSE system.

```
AR 0015 1I40I   READY
query system
AR 0015    …
AR 0015    …
AR 0015   PARALLEL ACCESS VOLUME (PAV):  ACTIVE      ←→  ( INACTIVE )
AR 0015 1I40I   READY
```

SIR   (output extension)

The SIR (SYS) output will identify if PAV has been activated on the z/VSE system.

```
sir sys
AR 0015 CPUID  VM = 000xxxxxxxxx000              VSE = FF03xxxyyyy8000
AR 0015 PROCESSOR = 2009-00                       LPAR
AR 0015 VM-SYSTEM =                (0702)  USERID =        VMCF = ON
AR 0015      CPUs = 0000                       Cap. = 00%
AR 0015 PROC-MODE = z/Arch(64-BIT)  IPL(200)   07:40:21     04/10/2008
AR 0015 SYSTEM    = z/VSE           4.2.0                   28/01/2008
AR 0015            VSE/AF           8.2.0                   02/28/2008
AR 0015            VSE/POWER        8.2.0                   01/11/2008
AR 0015 IPL-PROC  = $IPLESA         JCL-PROC  = $$JCL
AR 0015 SUPVR     = $$A$SUPI        TURBO-DISPATCHER (66) ACTIVE
AR 0015                             HARDWARE COMPRESSION ENABLED
AR 0015                             PAV SUPPORT ACTIVE
AR 0015 SEC. MGR. = BASIC           SECURITY  = ONLINE
…
```

The SIR SMF[1],VSE,cuu output will display the maximum I/O queue for a particular cuu .

The value is only given
- if SMF is activated
  (see *Hints & Tips for z/VSE on* http://www-03.ibm.com/servers/eserver/zseries/zvse/)
- if a single device is specified
- if the VSE option has been specified
- if cuu is not a PAV base device while PAV is active on the system.

```
sir smf,vse,261
AR 0015 TIMINGS FOR 261 BASED ON          589 I/O INSTRUCTION
AR 0015 MAXIMUM I/O QUEUE²   3
AR 0015
AR 0015   QUEUED      PENDING     CONNECT     DISCONN    DEV.BUSY      TOTAL
AR 0015  msec/SSCH   msec/SSCH   msec/SSCH   msec/SSCH   msec/SSCH   msec/SSCH
AR 0015    0.000       0.000       0.369       0.000       0.000       0.369
AR 0015 1I40I  READY
```

For PAV base devices the SIR SMF,VSE,cuu output will display I/O counts for the associated aliases.

The values are given only
- if PAV is active on the system
- if SMF is activated
- if a single device is specified
- if the VSE option has been specified

An I/O count of " * " indicates an overflow [> 32767 I/O INSTRUCTIONS].

```
sir smf,vse,777
AR 0015 TIMINGS FOR 777 BASED ON          111 I/O INSTRUCTION
AR 0015   BASE      777                     7 I/O INSTRUCTION
AR 0015   ALIAS     778                     0 I/O INSTRUCTION
AR 0015   ALIAS     779                     0 I/O INSTRUCTION
AR 0015   ALIAS     77A                    10 I/O INSTRUCTION
AR 0015   ALIAS     77B                    94 I/O INSTRUCTION
AR 0015
AR 0015   QUEUED      PENDING     CONNECT     DISCONN    DEV.BUSY      TOTAL
AR 0015  msec/SSCH   msec/SSCH   msec/SSCH   msec/SSCH   msec/SSCH   msec/SSCH
AR 0015    0.004       0.000       5.922       0.000       0.000       5.927
AR 0015 1I40I  READY
```

---

[1] SMF: Channel Subsystem Monitoring Facility
Facility in the channel subsystem that a program can retrieve measured values on performance (I/O count/ Connect Time….) for a designated subchannel.
[2] see Section 'z/VSE Performance Considerations with PAV' about possible usage

VOLUME   (output extension / new option)

The VOLUME has been extended to mark a base volume with PAV active by adding an ***B** in the "code" field.

```
AR 0015 1I40I  READY
volume
AR 0015 CUU  CODE DEV.-TYP   VOLID  USAGE    SHARED     STATUS    CAPACITY
....
....
AR 0015 261  6E   2107-900  VIS001  UNUSED                                  20 CYL
AR 0015 262  6E   2107-900  VIS002  UNUSED                                  20 CYL
AR 0015 777  6E*B 2105-000  FRA740  UNUSED                               10017 CYL
....
....
```

A new DETAIL option has been implemented to display possible alias devices related to the base.

```
AR 0015 1I40I  READY
volume 777,detail
AR 0015 CUU  CODE DEV.-TYP   VOLID  USAGE    SHARED     STATUS    CAPACITY
AR 0015 777  6E*B 2105-000  FRA740  UNUSED                               10017 CYL
AR 0015            BASE TO 778,77E,779,77A,77B
AR 0015 1I40I  READY
```

Alias devices put in parenthesis do indicate an ALIAS device being NOT OPERATIONAL.

```
...
AR 0015 777  6E*B 2105-000  FRA740  UNUSED                               10017 CYL
AR 0015            BASE TO 778,77E,779,(77A),(77B)
...
```

z/VSE Performance Considerations with PAV

Referring to the basic concept of PAV (see Chapter "Overview") it is obvious that most benefit and better performance can be achieved by defining alias devices for base devices that have a lot of parallel I/O ongoing .
→ speaking for z/VSE : having multiple tasks/partitions issuing I/O against the same volume.

z/VSE does load levelling for parallel I/O's by selecting a "free" device among the available base and alias devices.

To help identifying devices that are candidates to improve performance by using PAV, the output of the SIR SMF,VSE,cuu command can be used.

Example:

```
sir smf,vse,261
...
AR 0015 MAXIMUM I/O QUEUE    3
...
```

In this case a maximum of 3 I/O requests were queued to the device at one point in time. One waiting for the other to complete before the next can be processed.
From a high level perspective the configuration and definition of two aliases for this –then been called 'base'- device could improve overall system performance and throughput.

z/VSE General Concepts / Considerations / Restrictions

- Even if the hardware allows the definition of more than seven alias devices, z/VSE will only support one base and up to seven alias devices. This takes into account that a PAV enabled control unit can have a maximum of eight physically CHPID's configured.
- An alias device number must not be used in any z/VSE IPL context. If an ADD statement for a later – during PAV activation - identified alias device was specified the ADD'ed device will be boxed and taken out of any z/VSE processing.
- PAV activation requires to have one 'spare' I/O device eligible, thus restricting the z/VSE device limit to 1023. If 1024 devices are added MSG1YS2I is issued.
- During PAV activation all available alias devices for a base will be considered useable. There is no support to exclude or only use some of the aliases available.
- PAV activation on a native system or under LPAR may take some time, depending on your I/O configuration (→ALIAS recognition).
- A dynamic add of alias devices is not supported. Thus devices becoming active are just ignored. However, if a base device becomes ready while having PAV support active all corresponding aliases will be activated, too.
- Alias devices becoming not operational are taken out of PAV processing. As indication the alias cuu is put in brackets on the VOLUME cuu,DETAIL output.
- Even though an alias is considered somewhat virtual from a z/VSE administrator and operator point of view, it will make use of copy blocks, channel queue entries and other I/O related resources like a 'real' device. Therefore the allocation of such resources <u>must</u> consider alias devices.
- For former processing, transparency and due the nature that all alias devices correspond physically to the same base device any system service and message will make use of the base device (cuu) only.
- Inactivation of the PAV support via the SYSDEF command can lead to delays as the support is set inactive only after all I/O's on all alias devices have been completed.
- PAV processing will not improve performance on a 'single task' base.
  Start I/O – WAIT – Start I/O ….
  Obviously having more tasks in parallel issuing I/O's against a single device will benefit from PAV.
  NOTE: Due the extended possibility and usage of concurrent operations on the DS8K/DS6K, even more exploited by the z/VSE PAV support, it's still up to the application program to ensure that logically related requests are executed in the required sequence ! [ensure I/O completion prior starting the next I/O request; xmp. (format) writes].
- Subsystem Monitoring Facility (SMF) is enabled to collect base device information only.

z/VSE PAV support if running under z/VM (restriction)

Even though the z/VM PAV support is mostly transparent to guest operating systems following restriction apply:

- Control units having Hyper-PAV and basic PAV support enabled and having z/VM Mdisks allocated on these volumes (controlled by z/VM) do not support basic PAV emulation over Hyper-PAV and vice-versa!
  → aliases attached to z/VSE while Hyper-PAV is active on the corresponding control unit are not recognized during z/VSE PAV activation.

Commands relevant for z/VM PAV guest support

> o *QUERY PAV*
> o *QUERY VIRTUAL PAV*
> o *QUERY CU ....*
> o *SET CU ...*

*For further information please see **CP Commands and Utilities Reference SC24-6081-06** ; **CP Planning and Administration SC24-6083-06** .*