LVM — 1016 Physical Partitions per Physical Volume

Special Notices

Information in this document is correct to the best of our knowledge at the time of this writing. Please send feedback by fax to "AIXServ Information" at (512) 823-4009.

Please use this information with care. IBM will not be responsible for damages of any kind resulting from its use. The use of this information is the sole responsibility of the customer and depends on the customer's ability to evaluate and integrate this information into the customer's operational environment.

About This Document

This document discusses the limit of 1016 physical partitions, per disk, per volume group and how this limit affects the addition of disks larger than 4 GB with the mkvg or extendvg commands. This document applies to AIX 3.2 and 4.1.

Overview

In the design of LVM (Logical Volume Manager), each logical partition maps to one physical partition, and each physical partition maps to one or more physical partitions, and each physical partition maps to a number of disk sectors (only when mirroring, one LP maps to 2 or 3 physical partitions). The design of LVM limits the number of physical partitions that LVM can track per disk to 1016. In most cases, not all the possible 1016 tracking partitions are used by a disk. The default size of each physical partition during a "mkvg" command is 4 MB, which implies that individual disks up to 4 GB can be correctly tracked in a volume group.

Disks Larger Than 4 GB

If a disk larger than 4 GB is added to a volume group (based on usage of the default 4 MB size for physical partition) the disk addition will fail with a warning message that the physical partition size needs to be increased (see Note 1 at the end of this section). There are two instances where this limitation will be enforced.

The first case is when the user tries to use "mkvg" to create a volume group in which the number of physical partitions on one of the disks in the volume group would exceed 1016. In this case, the user must pick from the available physical partition ranges of:

```
1, 2, (4), 8, 16, 32, 64, 128, or 256 MB
```

and use the "-s" option with "mkvg".

The second case is when the user attempts to use "extendvg" to join the disk which violates the 1016 limitation with a pre-existing volume group. The user can either recreate the volume group with a larger physical partition size (which will allow the new disk to work with the 1016 limitation) or the user can create a standalone volume group (consisting of a larger physical partition size) for the new disk.

In AIX 4.1 and 3.2.5, if the install code detects that the rootvg drive is larger than 4 GB, it will change the "mkvg -s" value until the entire disk capacity can be mapped to the available 1016 tracks. (See Note 2 at the end of this section.) This install change also implies that all other disks added to rootvg, regardless of size, will also be defined at that new physical partitions size.

Note 1: This bug was fixed in APAR IX48926. Current AIX 3.2.5 and bos.rte.lvm 4.1.3.0 at AIX 4.1. AIX versions 3.2.5 and 4.1 which do not have this fix applied, will allow the creation of volume groups with more than 1016 partitions per disk. The implication of this bug allowing more than 1016 physical partitions is that the user may access all portions of the logical volume. However during disk mirroring, the status of partitions beyond the 1016 limit will not be tracked correctly. If mirrors beyond the 1016 range become "stale", LVM will not be aware of their condition, and data consistency may become an issue for those partitions. Additionally, the "migratepv" and "reorgvg" commands create mirrors and deletes them as a method for moving logical volumes within/between disks. If the 1016 limit is violated, then the "migratepv" or "reorgvg" commands may not behave correctly.

Note 2: This bug was fixed for AIX 3.2.5 rootvg install in APARs IX46862 and IX46863 and for AIX 4.1. with bos.rte.lvm 4.1.3.0.

Frequently Asked Questions

- What are the symptoms of this problem?

  The 1016 VGSA is used to track the "staleness of mirrors". If you are in violation of 1016, you may get a false report of a non-mirrored logical volume being "stale". You may also get a false indication that one of the mirror copies has gone stale. Next, migratepv may fail because migratepv briefly uses mirroring to move a logical volume from one disk to another. If the target logical partition is incorrectly considered "stale", then the migratepv cannot remove the source logical partition and the migratepv command will fail in the middle of migration. Another command which performs its actions by using temporary mirroring is reorgvg.

- How safe is my data? What if I never use mirroring or migratepv or reorgvg?

  The only case where data may be lost is if one is mirroring a logical volume and ALL copies go bad at the same time and LVM isn't aware of it because the copies that go bad are beyond the 1016 tracking range. However, in

this case, you would lose data even if you were within the 1016 range. If you never mirror or use migratepv, then this issue shouldn't concern you.

- Can I move this volume group between RS/6000 systems and versions of AIX?

  Yes. The enforcement of the 1016 limit is only during mkvg and extendvg. The "safeness" of the data on the volume group on AIX 3.2 is the same as it is on AIX 4.1.

- Will this limitation be changed?

  No, the ability to change the PP size assures that regardless of the size of your disk drive, you'll always be able to accommodate the drive below the 1016 limit. If a change to the limitation were made, then a huge incompatibility bug would be introduced into LVM.

- Will I have to rebuild the volume group?

  Yes, the fixes mentioned in this document only prevent the future addition/creation of disks (into volume groups) that (the disks) will violate the 1016 limitation. Disks that are already in violation of the 1016 limit must be recreated at larger PP sizes.

Reader's Comments

Please fax this form to (512) 823-4009, attention "AIXServ Information". You may also e-mail comments to: elizabet@austin.ibm.com. These comments should include the same customer information requested below.

Use this form to tell us what you think about this document. If you have found errors in it, or if you want to express your opinion about it (such as organization, subject matter, appearance) or make suggestions for improvement, this is the form to use.

If you need technical assistance, contact your local branch office, point of sale, or 1-800-CALL-AIX (for information about support offerings). These services may be billable. Faxes on a variety of subjects may be ordered free of charge from 1-800-IBM-4FAX. Outside the U.S. call 415-855-4329 using a fax machine phone.

When you send comments to IBM, you grant IBM a nonexclusive right to use or distribute your comments in any way it believes appropriate without incurring any obligation to you.

Note: If you have a problem report or item number, supplying that number may help us determine why a procedure did or did not work in your specific situation.

Problem Report or Item #: Branch Office or Customer #:

Be sure to print your name and fax number below if you would like a reply:

Name: Fax Number:

END OF DOCUMENT (lvm.1016pps.32-41.krn 4FAX 3689)