GPFS V3.1 and V2.3 Questions and Answers

Overview

General Parallel File System (GPFS) is a high performance shared-disk file system that can provide data access from nodes in a cluster environment. Parallel and serial applications can readily access shared files using standard UNIX® file system interfaces, and the same file can be accessed concurrently from multiple nodes. GPFS is designed to provide high availability through logging and replication, and can be configured for failover from both disk and server malfunctions.

GPFS for AIX 5L™ has been widely accepted on the IBM® RS/6000® SP™ system to scale file system I/O which can help meet the requirements of a wide range of applications including seismic data processing, digital library file serving, and data mining in business intelligence. GPFS for AIX 5L supports the IBM eServer™ Cluster 1600 as well as clusters of IBM System p5™, IBM eServer p5, IBM BladeCenter® JS20, IBM BladeCenter JS21, or select IBM System i5™ servers.

GPFS for Linux® provides similar capabilities to IBM Linux clusters as its AIX 5L counterpart provides to System p™ clusters. The GPFS for Linux on Multiplatform product runs on the IBM System Cluster 1350 as well as Linux clusters based on selected IBM x86 xSeries® rack-optimized servers, select IBM BladeCenter servers, or select IBM for AMD processor-based servers. The GPFS for Linux on POWER™ product runs on select System p5, eServer p5, BladeCenter and IBM eServer OpenPower™ servers.

For further information regarding the use of GPFS in your clusters, see the *GPFS: Concepts, Planning, and Installation Guide*.

Updates to this FAQ include:

Table 1.

December 2006	
	2.2 What are the latest distributions and fix or kernel levels that GPFS has been tested with?
	3.5 What interconnects are supported for GPFS daemon-to-daemon communication in my GPFS cluster?
	7.2 What is the current service information for GPFS?

Questions & Answers

1. General questions:
   - 1.1 How do I order GPFS
   - 1.2 How is GPFS priced?
   - 1.3 Where can I find the documentation for GPFS?
   - 1.4 What resources beyond the standard documentation can help me learn and use GPFS?
   - 1.5 Does GPFS participate in the IBM Scholars Program?
   - 1.6 How can I ask a more specific question about GPFS?

2. Software questions:
   - 2.1 What levels of the AIX O/S are supported by GPFS V3.1?
   - 2.2 What are the latest distributions and fix or kernel levels that GPFS has been tested with?
   - 2.3 Can different GPFS maintenance levels coexist?

3. Machine questions:
   - 3.1 What are the minimum hardware requirements for a GPFS cluster?
   - 3.2 Is GPFS for AIX 5L supported on IBM System i5 servers?
   - 3.3 What machine models has GPFS for Linux been tested with?
   - 3.4 Is GPFS for Linux supported on all IBM ServerProven® servers?
   - 3.5 What interconnects are supported for GPFS daemon-to-daemon communication in my GPFS cluster?
   - 3.6 Does GPFS support exploitation of the Virtual I/O Server (VIOS) features of POWER5 processors?

4. Disk questions:
   - 4.1 What disk hardware has GPFS been tested with?
   - 4.2 What Fibre Channel (FC) Switches are qualified for GPFS usage and is there a FC Switch support chart available?
   - 4.3 Can I concurrently access disks from both AIX 5L and Linux nodes in my GPFS cluster?
   - 4.4 What disk support failover models does GPFS support for the IBM TotalStorage® DS4000 family of storage controllers with the Linux operating system?
   - 4.5 Are there any special considerations when my cluster consists of two nodes?

5. Scaling questions:
   - 5.1 What are the GPFS cluster size limits?
   - 5.2 What are the current file system size limits?
   - 5.3 What is the current limit on the number of mounted file systems in a GPFS cluster?
   - 5.4 What is the architectural limit of the number of files in a file system?
   - 5.5 What is the current limit on the number of nodes that may concurrently access a file system?
   - 5.6 What are the limitations on GPFS disk size?

6. Configuration and tuning questions:
   - 6.1 What specific configuration and performance tuning suggestions are there?
   - 6.2 What configuration and performance tuning suggestions are there for GPFS when used primarily for Oracle databases?
   - 6.3 What Linux configuration settings are required when NFS exporting a GPFS filesystem?

6.4 Sometimes GPFS appears to be handling a heavy I/O load, for no apparent reason. What could be causing this?

6.5 What considerations are there when using IBM Tivoli® Storage Manager with GPFS?

6.6 Are there any special considerations for using automount in my GPFS cluster?

6.7 Are there any special considerations for the use of OpenSSL?

6.8 When I allow other clusters to mount my file systems, is there a way to restrict access permissions for the root user?

7. Service questions:

7.1 What support services are available for GPFS?

7.2 What is the current service information for GPFS?

7.3 How do I download fixes for GPFS?

7.4 What are the current GPFS advisories?

7.5 What Linux kernel patches are provided for clustered file systems such as GPFS?

General questions

Q1.1: How do I order GPFS?

A1.1: To order GPFS:

- GPFS for AIX 5L and GPFS for Linux on POWER are ordered through the Technical Support organization. Login to the Technical resources and support Web page and click on **Contact techline**. Select your region and preferred method of contact.
- GPFS for Linux on Multiplatform is ordered through the Passport Advantage® program.
  - Additional information on Passport Advantage is available on the IBM Passport Advantage page.
  - For ordering information, go to the Passport Advantage program page and click on *How to Buy* in the left-hand navigation.

Q1.2: How is GPFS priced?

A1.2: The prices for GPFS for AIX 5L, GPFS for Linux on POWER, and GPFS for Linux on Multiplatform are based on the number of processors active on the server where GPFS is installed. For further information, please contact your IBM Marketing Representative. For a directory of worldwide contact, see www.ibm.com/planetwide/index.html

Q1.3: Where can I find the documentation for GPFS?

A1.3: The GPFS documentation is available in both PDF and HTML format on the Cluster Information Center at publib.boulder.ibm.com/infocenter/clresctr/vxrx/index.jsp?topic=/com.ibm.cluster.gpfs.doc/gpfsbooks.html.

Q1.4: What resources beyond the standard documentation can help me learn about and use GPFS?

A1.4: For additional information regarding GPFS see:

- The Clusters Literature site for AIX 5L at www.ibm.com/servers/eserver/clusters/library/wp_aix_lit.html
- The Clusters Literature site for Linux at www.ibm.com/servers/eserver/clusters/library/wp_linux_lit.html
- The IBM Redbooks™ and Redpapers site at www.redbooks.ibm.com
- The IBM Almaden Research GPFS web page at www.almaden.ibm.com/StorageSystems/file_systems/GPFS/index.shtml
- For the latest news on GPFS in Digital Media solutions:
  1. Go to www.ibm.com/solutions/digitalmedia/
  2. Enter **gpfs** in the box in the upper right-hand corner and click on **Search**
- Go to the @server Magazine at eservercomputing.com/ME2/Default.asp and search on GPFS.

Updated in October 2005, the GPFS primer is located at http://www.ibm.com/servers/eserver/pseries/software/whitepapers/gpfsprimer.html

Q1.5: Does GPFS participate in the IBM Scholars Program?

A1.5: GPFS is available through the IBM Scholars Program by completing a software request form. It is *not* in the download catalog. To request GPFS, an IBM Scholars Program member should:

1. Go to www.developer.ibm.com/university/scholars/
2. From the left-hand navigation bar, click **Downloads and CDs**.
3. Click **Request Software**.
4. Enter your IBM Scholars Program user name and password.
5. On the software page, scroll down to **Cluster Software**.
6. Click **Software Request Form**.
7. Complete the request form.
8. Click **submit**.
9. The software will be shipped to the address entered in the members registration form for that user name.

Q1.6: How can I ask a more specific question about GPFS?

A1.6: Depending upon the nature of your question, you may ask it in one of several ways.

- If you want to correspond with IBM regarding GPFS:
  - If your question concerns a potential software error in GPFS and you have an IBM software maintenance contract, please contact 1-800-IBM-SERV in the United States or your local IBM Service Center in other countries. IBM Scholars Program users should notify the GPFS development team of potential software bugs through gpfs@us.ibm.com.

  - If you have a question that can benefit other GPFS users, you may post it to the GPFS technical discussion forum at www.ibm.com/developerworks/forums/dw_forum.jsp?forum=479&cat=13
  - This FAQ is continually being enhanced. To contribute possible questions or answers, please send them to gpfs@us.ibm.com
- If you want to interact with other GPFS users, the San Diego Supercomputer Center maintains a GPFS user mailing list. The list is gpfs-general@sdsc.edu and those interested can subscribe to the list at lists.sdsc.edu/mailman/listinfo/gpfs-general

If your question does not fall into the above categories, you can send a note directly to the GPFS development team at gpfs@us.ibm.com. However, this mailing list is informally monitored as time permits and should not be used for priority messages to the GPFS team.

Software questions

Q2.1: What levels of the AIX O/S are supported by GPFS V3.1?

A2.1: GPFS V3.1 supports both AIX 5L V5.3 and V5.2 nodes in a homogenous or heterogenous cluster running either the AIX 5L or the Linux operating system.

Notes:

1. Enhancements to the support of Network File System (NFS) V4 in GPFS V3.1 are only available on AIX 5L V5.3 systems with the minimum technology level of 5300-04 applied.
2. Customers should consider the support plans for AIX 5L V5.2 in their operating system decision.

Q2.2: What are the latest distributions and fix or kernel levels that GPFS has been tested with?

A2.2: While GPFS runs with many different AIX 5L fixes and Linux kernel levels, it is highly suggested that customers apply the latest fix levels and kernel service updates for their operating system.

Table 2. GPFS for Linux V3.1.0-8

Linux Distribution	Kernel Level
POWER	
Red Hat EL 4.0 (Update 4)	2.6.9-42.0.3.EL[1]
SUSE Linux ES 10	2.6.16.21-0.25[3]
SUSE Linux ES 9 (SP3)	2.6.5-7.282
x86_64	
Red Hat EL 4.0 (Update 4)	2.6.9-42.0.3.EL[1]
Red Hat EL 3.0 (Update 8)	2.4.21-47.0.1.EL[1]
SUSE Linux ES 10	2.6.16.21-0.25[3]
SUSE Linux ES 9 (SP3)	2.6.5-7.282
SUSE Linux ES 8 (SP4)	2.4.21-309
IA32	
Red Hat EL 4.0 (Update 4)	2.6.9-42.0.3.EL[1,2]
Red Hat EL 3.0 (Update 8)	2.4.21-47.0.1.EL[2]
SUSE Linux ES 10	2.6.16.21-0.25[3]
SUSE Linux ES 9 (SP3)	2.6.5-7.282
SUSE Linux ES 8 (SP4)	2.4.21-309
1. Currently GPFS does not support the Red Hat EL 4.0 uniprocessor (UP) kernel. 2. Currently GPFS does not support the RHEL 3.0 and RHEL 4.0 hugemem kernel. 3. There is required service for GPFS V3.1 support of SLES 10. Please see question 7.2 *What is the current service information for GPFS*?	

Table 3. GPFS for AIX 5L on POWER

GPFS	AIX 5L
GPFS V3.1.0.8	AIX 5L V5.3
	AIX 5L V5.2
GPFS V2.3.0.18	AIX 5L V5.3
	AIX 5L V5.2

Table 4. GPFS for Linux V2.3.0-18

Linux Distribution	Kernel Level
POWER	
Red Hat EL 4.0 (Update 3)	2.6.9-34.0.2.EL[1]
SUSE Linux ES 9 (SP3)	2.6.5-7.267
SUSE Linux ES 8 (SP4)	2.4.21-309
x86_64	
Red Hat EL 4.0 (Update 3)	2.6.9-34.0.2.EL[1]
Red Hat EL 3.0 (Update 7)	2.4.21-40.EL[1]
SUSE Linux ES 9 (SP3)	2.6.5-7.267
SUSE Linux ES 8 (SP4)	2.4.21-309
IA32	
Red Hat EL 4.0 (Update 3)	2.6.9-34.0.2.EL[1,2]
Red Hat EL 3.0 (Update 7)	2.4.21-40.EL[2]
SUSE Linux ES 9 (SP3)	2.6.5-7.267
SUSE Linux ES 8 (SP4)	2.4.21-309
1. Currently GPFS does not support the Red Hat EL 4.0 uniprocessor (UP) kernel. 2. Currently GPFS does not support the RHEL 3.0 and RHEL 4.0 hugemem kernel.	

Q2.3: Can different GPFS maintenance levels coexist?

A2.3: Certain levels of GPFS can coexist, that is, be active in the same cluster and simultaneously access the same file system. This allows upgrading GPFS within a cluster without shutting down GPFS on all nodes first, and also mounting GPFS file systems from other GPFS clusters that may be running a different maintenance level of GPFS. The current maintenance level coexistence rules are:

- Starting with GPFS 2.3.0.6, different maintenance levels of GPFS 2.3 can coexist. Prior to GPFS 2.3.0.6 all nodes must be at the same maintenance level.
- All GPFS 3.1 maintenance levels can coexist with each other, unless otherwise stated in this FAQ.
- GPFS 2.3 and 3.1 are completely incompatible and can not coexist.

Machine questions

Q3.1: What are the minimum hardware requirements for a GPFS cluster?

A3.1: The minimum hardware requirements are:

- GPFS for AIX 5L on POWER: IBM POWER3™ or newer processor, 1 GB of memory
- GPFS for Linux on Multiplatform:
  - Intel® Pentium® 3 or newer processor, with 512 MB of memory
  - AMD Opteron™ processors, with 1 GB of memory
- GPFS for Linux on POWER: IBM POWER3 or newer processor, 1 GB of memory

Additionally, it is highly suggested that a sufficiently large amount of swap space is configured. While the actual configuration decisions should be made taking into account the memory requirements of other applications, it is suggested to configure at least as much swap space as there is physical memory on a given node.

GPFS is tested, and supported, on hardware as specified in the *Hardware requirements* section of the Sales Manual for GPFS. If you are running GPFS on hardware that is not listed in the *Hardware Requirements*, should problems arise, recreation of the problem by the customer on a configuration conforming to the *Hardware Requirements* may be required by IBM.

To access the Sales Manual for GPFS:

1. Go to http://www-306.ibm.com/common/ssi/OIX.wss
2. From **A specific type** menu, choose **HW&SW Desc (Sales Manual,RPQ)**.
3. To view a GPFS sales manual, choose the corresponding product number to enter in the **keyword** field then click on **Go**
   - For General Parallel File System V3.1 for AIX 5L, enter **5765-G66**
   - For General Parallel File System V3.1 for Linux on POWER, enter **5765-G67**
   - For General Parallel File System V3.1 for Linux on Multiplatform, enter **5724-N94**
   - For General Parallel File System V2.3 for AIX 5L, enter **5765-F64**
   - For General Parallel File System V2.3 for Linux on POWER, enter **5765-G20**
   - For General Parallel File System V2.3 for Linux on Multiplatform, enter **5765-G23**

Q3.2: Is GPFS for AIX 5L supported on IBM System i5 servers?

A3.2: GPFS for AIX 5L, V2.3 and V3.1 extends all features, function, and restrictions (such as operating system and scaling support) to IBM System i5 servers to match their IBM System p5 counterparts:

Table 5.

IBM System i5	IBM System p5
i5-595	p5-595
i5-570	p5-570
i5-550	p5-550
i5-520	p5-520

No service updates are required for this additional support.

Q3.3: What machine models has GPFS for Linux been tested with?

A3.3: GPFS has been tested with:

- IBM x86 xSeries machine models:
  - 330
  - 335
  - 336
  - 340
  - 342
  - 345
  - 346
  - 360
  - 365
  - 440

- IBM BladeCenter:
  - HS20
  - HS40
  - JS20
  - LS20
- IBM AMD processor-based servers:
  - 325
  - 326
- IBM eServer p5 :

  For both the p5-590 and the p5-595: See the question *What is the current service information for GPFS?*
  - 510
  - 520
  - 550
  - 570
  - 575
  - 590
  - 595
- IBM eServer OpenPower servers:
  - 710
  - 720
- IBM eServer pSeries® machines models that support Linux
- The IBM eServer Cluster 1300
- The IBM System Cluster 1350

For hardware and software certification, please see the IBM ServerProven site at www.pc.ibm.com/us/compat/eserver.html

Q3.4: Is GPFS for Linux supported on all IBM ServerProven servers?

A3.4: GPFS for Linux is supported on all IBM ServerProven servers:

1. with the distributions and kernel levels as listed in the question *What are the latest distributions and kernel levels that GPFS has been tested with?*
2. that meet the minimum hardware model requirements as listed in the question *What are the minimum hardware requirements for a GPFS cluster?*

Please see the IBM ServerProven site at http://www.ibm.com/servers/eserver/serverproven/compat/us/

Q3.5: What interconnects are supported for GPFS daemon-to-daemon communication in a GPFS cluster?

A3.5: The interconnect for GPFS daemon-to-daemon communication depends upon the types of nodes in your cluster.

Table 6. GPFS daemon -to-daemon communication interconnects

Nodes in your cluster	Supported interconnect	Supported environments
Linux	Ethernet	all supported GPFS environments
	10-Gigabit Ethernet	IBM eServer pSeries and System p5 servers running SLES 9 SP3 or SLES 10
	Myrinet	IP only
	InfiniBand	IP only SLES 9 or Red Hat EL 4.0, on System x servers SLES 9 SP 3 on System p5 servers with GPFS V3.1.0-4 or later
Linux and AIX 5L	Ethernet	all supported GPFS environments
	10-Gigabit Ethernet	IBM eServer pSeries and System p5 servers running SLES9 SP3, SLES 10, or AIX 5L V5.3

Table 6. GPFS daemon -to-daemon communication interconnects (continued)

Nodes in your cluster	Supported interconnect	Supported environments
AIX 5L	Ethernet	all supported GPFS environments
	10-Gigabit Ethernet	AIX 5L V5.3
	Myrinet	AIX 5L V5.2 and V5.3 64-bit kernel BladeCenter JS20 and p5 POWER5 servers IP only
	InfiniBand	AIX 5L V5.3 GPFS V3.1 IP only
	eServer HPS	homogenous clusters of either AIX 5L V5.2 or V5.3
	SP Switch2	GPFS V2.3

Q3.6: Does GPFS support exploitation of the Virtual I/O Server (VIOS) features of POWER5 processors?

A3.6: Yes, GPFS V3.1 and V2.3 allow exploitation of POWER5 VIOS configurations. Both the virtual SCSI (VSCSI) and the shared Ethernet adapter (SEA) are supported in single and multiple central electronics complex (CEC) configurations. This support is limited to GPFS nodes that are using the AIX 5L V5.3 operating system.

All VIO client logical partitions (LPAR) in a CEC that are GPFS cluster members should have the VIO disks mapped to each LPAR from each VIOS server (if more then one VIO server in a CEC) using virtual SCSI. This creates a virtual SAN environment where each GPFS LPAR has access to disk on a local path(s) without requiring network access. All of the Network Shared Disks (NSD) in these configurations should not be coded with a primary or secondary NSD server associated with them.

Minimum required code levels:
- VIOS Release 1.3.0.0 Fix Pack 8
- AIX 5L V5.3 Service Pack 5300-05-01

There is no GPFS fix level requirement for this support, but it is recommended that you be at the latest GPFS level available. For information on the latest levels of GPFS go to www14.software.ibm.com/webapp/set2/sas/f/gpfs/download/home.html

For further information on POWER5 VIOS go to techsupport.services.ibm.com/server/vios/documentation/faq.html

For VIOS documentation, go to techsupport.services.ibm.com/server/vios/documentation/home.html

Disk questions

Q4.1: What disk hardware has GPFS been tested with?

A4.1: This set of tables displays the set of disk hardware which has been tested by IBM and known to work with GPFS. GPFS is not limited to only using this set of disk devices. Other disk devices may work with GPFS but they have not been tested by IBM. The GPFS support team will help customers who are using devices outside of this list of tested devices, to solve problems directly related to GPFS, but not problems deemed to be issues with the underlying device's behavior including any performance issues exhibited on untested hardware.

It is important to note that:

- Each individual disk subsystem requires a specific set of device drivers for proper operation while attached to a host running GPFS or IBM Recoverable Virtual Shared Disk. The prerequisite levels of device drivers are not documented in this GPFS-specific FAQ. Refer to the disk subsystem's web page to determine the currency of the device driver stack for the host's operating system level and attachment configuration.

  For information on IBM disk storage subsystems and their related device drivers levels and Operating System support guidelines, go to www.ibm.com/servers/storage/support/disk/index.html
- Microcode levels should be at the latest levels available for your specific disk drive.

  For the IBM TotalStorage, go to www.ibm.com/servers/storage/support/allproducts/downloading.html

FAStT customers: Please also see the IBM TotalStorage FAStT Best Practices Redbook at publib-b.boulder.ibm.com/abstracts/sg246363.html?Open

Customers migrating to GPFS V2.3 or V3.1 previously using single-node quorum: Beginning with the availability of GPFS V2.3, the single-node quorum function is replaced with *node quorum with tiebreaker disks*. The new node quorum with tiebreaker disks support does not depend upon the availability of SCSI-3 persistent reserve. All disks tested with GPFS can now utilize node quorum with tiebreaker disks as opposed to GPFS node quorum (one plus half of the explicitly defined quorum nodes in the GPFS cluster). For further information, see the *GPFS: Concepts, Planning, and Installation Guide*.

Table 7. Disk hardware tested with GPFS for AIX 5L on POWER

GPFS for AIX 5L on POWER:	
	IBM TotalStorage DS6000 using either Subsystem Device Driver (SDD) or Subsystem Device Driver Path Control Module (SDDPCM) **Configuration considerations:** GPFS clusters up to 32 nodes are supported and require a firmware level of R9a.5b050318a or greater. See further requirements below.
	IBM TotalStorage DS8000 using either SDD or SDDPCM **Configuration considerations:** GPFS clusters up to 32 nodes are supported and require a firmware level of R10k.9b050406 or greater. See further requirements below.

Table 7. Disk hardware tested with GPFS for AIX 5L on POWER (continued)

	DS6000 and DS8000 service requirements: • AIX 5L V5.2 maintenance level 05 (5200-05) - APAR # IY68906, APAR # IY70905 • AIX 5L V5.3 maintenance level 02 (5300-02) - APAR # IY68966, APAR # IY71085 • GPFS for AIX 5L V2.3 - APAR # IY66584, APAR # IY70396, APAR # IY71901 **DS6000 and DS8000 SDD requirements for the Disk Leasing model:** • AIX 5L V5.2 SDD fileset **devices.sdd.52.rte** (1.6.0.2) • AIX 5L V5.3 SDD fileset **devices.sdd.53.rte** (1.6.0.2) **DS6000 and DS8000 Multi-Path I/O requirements for the Persistent Reserve model:** • AIX 5L V5.2 SDDPCM fileset **devices.sddpcm.52.rte** (2.1.0.1) • AIX 5L V5.3 SDDPCM fileset **devices.sddpcm.53.rte** (2.1.0.1)
	IBM TotalStorage DS4100 (Formerly FAStT 100) with DS4000 EXP100 Storage Expansion Unit with Serial Advanced Technology Attachment (SATA) drives
	IBM TotalStorage FAStT500
	IBM TotalStorage DS4200 Express all supported expansion drawer and disk types
	IBM TotalStorage DS4300 (Formerly FAStT 600) with DS4000 EXP710 Fibre Channel (FC) Storage Expansion Unit, DS4000 EXP700 FC Storage Expansion Unit, or EXP100
	IBM TotalStorage DS4300 Turbo with EXP710, EXP700, or EXP100
	IBM TotalStorage DS4400 (Formerly FAStT 700) with EXP710 or EXP700
	IBM TotalStorage DS4500 (Formerly FAStT 900) with EXP710, EXP700, or EXP100
	IBM TotalStorage DS4700 Express all supported expansion drawer and disk types
	IBM TotalStorage DS4800 with EXP710, or EXP100
	IBM TotalStorage ESS (2105-F20 or 2105-800 with SDD)
	IBM TotalStorage ESS (2105-F20 or 2105-800 using AIX 5L Multi-Path I/O (MPIO) and SDDPCM))
	IBM TotalStorage Storage Area Network (SAN) Volume Controller (SVC) V1.1, V1.2, and V2.1 The following APAR numbers are suggested: • IY64709 - Applies to all GPFS clusters • IY64259 - Applies only when running GPFS in an AIX V5.2 or V5.3 environment with RVSD 4.1 • IY42355 - Applies only when running GPFS in a PSSP V3.5 environment • SVC V2.1.0.1 is supported with AIX 5L V5.2 (Maintenance Level 05) and AIX 5L V5.3 (Maintenance Level 01). See www.ibm.com/support/docview.wss?rs=591&uid=ssg1S1002471 for specific advice on SAN Volume Controller recommended software levels.
	IBM 7133 Serial Disk System (all disk sizes)
	Hitachi Lightning 9900™(9910, 9960, 9970V and 9980V (Hitachi Dynamic Link Manager™) (HDLM) required)
	EMC Symmetrix DMX Storage Subsystems (FC attach only)

Table 8. Disk hardware tested with GPFS for Linux on x86 xSeries servers

GPFS for Linux on xSeries servers:	
	IBM TotalStorage FAStT 200 Storage Server
	IBM TotalStorage FAStT 500
	IBM TotalStorage DS4100 (Formerly FAStT 100) with EXP100
	IBM TotalStorage DS4200 Express all supported expansion drawer and disk types
	IBM TotalStorage DS4300 (Formerly FAStT 600) with EXP710, EXP700, or EXP100
	IBM TotalStorage DS4300 Turbo with EXP710, EXP700, or EXP100
	IBM TotalStorage DS4400 (Formerly FAStT 700) with EXP710 or EXP700
	IBM TotalStorage DS4500 (Formerly FAStT 900) with EXP710, EXP700, or EXP100
	IBM TotalStorage DS4700 Express all supported expansion drawer and disk types
	IBM TotalStorage DS4800 with EXP710, or EXP100
	IBM TotalStorage Enterprise Storage Server® (ESS) models 2105-F20 and 2105-800, with Subsystem Device Driver (SDD)
	EMC Symmetrix Direct Matrix Architecture (DMX) Storage Subsystems 1000 with PowerPath v 3.06 and v 3.07
Restrictions:	IBM ServeRAID™ adapters are not supported.

Table 9. Disk hardware tested with GPFS for Linux on POWER

GPFS for Linux on POWER:	
	IBM TotalStorage DS4200 Express all supported expansion drawer and disk types
	IBM TotalStorage DS4300 (Formerly FAStT 600) all supported drawer and disk types
	IBM TotalStorage DS4500 (Formerly FAStT 900) all supported expansion drawer and disk types
	IBM TotalStorage DS4700 Express all supported expansion drawer and disk types
	IBM TotalStorage DS8000 using SDD

Table 10. Disk hardware tested with GPFS for Linux on AMD processor-based servers

GPFS for Linux on eServer AMD processor-based servers:	No devices tested specially in this environment.

Q4.2: What Fibre Channel Switches are qualified for GPFS usage and is there a FC Switch support chart available?

A4.2: There are no special requirements for FC switches used by GPFS other than the switch must be supported by AIX 5L or Linux. For further information see www.storage.ibm.com/ibmsan/index.html

Q4.3: Can I concurrently access SAN-attached disks from both AIX 5L and Linux nodes in my GPFS cluster?

A4.3: The architecture of GPFS allows both AIX and Linux hosts to concurrently access the same set of LUNs. However, before this is implemented in a GPFS cluster you must ensure that the disk subsystem being used supports both AIX and Linux concurrently accessing LUNs. While the

GPFS architecture allows this, the underlying disk subsystem may not, and in that case, a configuration attempting it would not be supported.

Q4.4: What disk support failover models does GPFS support for the IBM TotalStorage DS4000 family of storage controllers with the Linux operating system?

A4.4: GPFS has been tested with both the Host Bus Adapter Failover and Redundant Dual Active Controller (RDAC) device drivers.

To download the current device drivers for your disk subsystem, please go to http://www.ibm.com/servers/storage/support/

1. In the **Select your product** frame, select **Disk storage systems** for your **Product family**.
2. In the **Product** selection, select your disk system.
3. Click on **Go** to find the product-specific page.

On the product-specific page:

1. Click on the **Download** tab.
2. Choose the appropriate selection under **Fixes and drivers**.

Q4.5: Are there any special considerations when my cluster consists of two nodes?

A4.5: Customers migrating to GPFS V2.3 or V3.1, who previously used single-node quorum, must be aware that the single-node quorum function has been replaced with *node quorum with tiebreaker disks*. The new node quorum with tiebreaker disks support does *not* depend upon the availability of SCSI-3 persistent reserve. All disks tested with GPFS can now utilize node quorum with tiebreaker disks as opposed to GPFS node quorum (one plus half of the explicitly defined quorum nodes in the GPFS cluster). For further information, see the *GPFS: Concepts, Planning, and Installation Guide* for your level of GPFS.

Scaling questions

Q5.1: What are the GPFS cluster size limits?

A5.1: The current GPFS cluster size limits are:

Table 11. GPFS maximum tested cluster sizes

GPFS for Linux on Multiplatform	2441 nodes
GPFS for AIX 5L on POWER	1530 nodes
Contact gpfs@us.ibm.com if you intend to exceed: 1. Linux configurations larger than 512 nodes 2. AIX configurations larger than 128 nodes	

Q5.2: What are the current file system size limits?

A5.2: The current file system size limits are:

Table 12. Current file system size limits

GPFS 2.3 or later, file system architectural limit	2^99 bytes
GPFS 2.2 file system architectural limit	2^51 bytes (2 Petabytes)
Current tested limit	approximately 2 PB
Contact gpfs@us.ibm.com if you intend to exceed 200 Terabytes	

Q5.3: What is the current limit on the number of mounted file systems in a GPFS cluster?

A5.3: The total number of mounted file systems within a GPFS cluster is currently limited to 32.

Q5.4: What is the architectural limit of the number of files in a file system?

A5.4: The architectural limit of the number of files in a file system is determined by the file system format. For file systems created prior to GPFS V2.3, the limit is 268,435,456. For file systems created with GPFS V2.3 or later, the limit is 2,147,483,648. Please note that the effective limit on the number of files in a file system is usually lower than the architectural limit, and could be adjusted using the **-F** option of the mmchfs command.

Q5.5: What is the current limit on the number of nodes that may concurrently join a cluster?

A5.5: The total number of nodes that may concurrently join a cluster is currently limited to a maximum of 4096 nodes. A node joins a given cluster if it is:

- A member of the local GPFS cluster (the **mmlscluster** command output displays the local cluster nodes) and GPFS is up and running on the node.
- A node in a different GPFS cluster that is mounting a file system from the local cluster.

For example:

- GPFS **clusterA** has 2100 member nodes as listed in the **mmlscluster** command, and GPFS is up and running on 2000 of those nodes.
- 500 nodes from **clusterB** are mounting a file system owned by **clusterA**.

clusterA therefore has 2500 concurrent nodes.

Q5.6: What are the limitations on GPFS disk size?

A5.6: The maximum disk size supported by GPFS depends on file system format and the underlying device support. For file systems created prior to GPFS version 2.3, the maximum disk size is 1 TB due to disk format limitations. For file systems created with GPFS 2.3 or later, these limitations have been removed, and the maximum disk size is only limited by the device driver support. On AIX 5L with 64-bit kernel, GPFS supports disks larger than 2 TB (provided the disk device supports it), up to the operating system limit. On other supported platforms, GPFS supports disks up to 2 TB in size.

Configuration and tuning questions

Please also see the question *What is the current service information for GPFS?*

Q6.1: What specific configuration and performance tuning suggestions are there?

A6.1: In addition to the configuration and performance tuning suggestions in the *GPFS: Concepts, Planning, and Installation Guide* for your version of GPFS:

- If your system consists of the eServer pSeries High Performance Switch, it is suggested that you configure GPFS over the **ml0** IP network interface.
- On systems running with the Linux 2.6 kernel, it is recommended you adjust the **vm.min_free_kbytes** kernel tunable. This tunable controls the amount of free memory that Linux kernel keeps available (i.e. not used in any kernel caches). When **vm.min_free_kbytes** is set to its default value, on some configurations it is possible to encounter memory exhaustion symptoms when free memory should in fact be available. Setting **vm.min_free_kbytes** to a higher value (Linux **sysctl** utility could be used for this purpose), on the order of magnitude of 5-6% of the total amount of physical memory, should help to avoid such a situation.

  Also, please see the GPFS Redpapers:
  - *GPFS Sequential Input/Output Performance on IBM pSeries 690* at www.redbooks.ibm.com/redpapers/pdfs/redp3945.pdf
  - *Native GPFS Benchmarks in an Integrated p690/AIX and x335/Linux Environment* at www.redbooks.ibm.com/redpapers/pdfs/redp3962.pdf

Q6.2: What configuration and performance tuning suggestions are there for GPFS when used primarily for Oracle databases?

A6.2: In addition to the performance tuning suggestions in the *GPFS: Concepts, Planning, and Installation Guide* for your version of GPFS, when running Oracle RAC 10g, it is suggested you increase the value for **OPROCD_DEFAULT_MARGIN** to at least 500 to avoid possible random reboots of nodes.

In the control script for the Oracle CSS daemon, located in **/etc/init.cssd** the value for **OPROCD_DEFAULT_MARGIN** is set to 500 (milliseconds) on all UNIX derivatives except for AIX. For AIX this value is set to 100. From a GPFS perspective, even 500 milliseconds maybe too low in situations where node failover may take up to a minute or two to resolve. However, if during node failure the surviving node is already doing direct IO to the **oprocd** control file, it should have the necessary tokens and indirect block cached and should therefore not have to wait during failover.

Q6.3: What Linux configuration settings are required when NFS exporting a GPFS filesystem?

A6.3: If you are running at SLES 9 SP 1, the kernel defines the sysctl variable **fs.nfs.use_underlying_lock_ops** that determines if the NFS lockd is to consult the file system when granting advisory byte-range locks. For distributed file systems like GPFS, this must be set to **true** (the default is **false**).

You can query the current setting by issuing the command:

```
sysctl fs.nfs.use_underlying_lock_ops
```

Alternatively, the record **fs.nfs.use_underlying_lock_ops = 1** may be added to **/etc/sysctl.conf**. This record must be applied after initially booting the node and after each reboot by issuing the command:

```
sysctl -p
```

As the **fs.nfs.use_underlying_lock_ops** variable is currently not available in SLES 9 SP2 or later, when NFS exporting a GPFS file system ensure your NFS server nodes are at the SP1 level (until such time the variable is made available in later service packs).

For additional considerations when NFS exporting your GPFS file system, see the:

- *GPFS: Administration Guide* for

  - NFS and GPFS
  - Managing GPFS access control lists and NFS export
- *GPFS: Concepts, Planning, and Installation Guide* for File system creation considerations

Q6.4: Sometimes GPFS appears to be handling a heavy I/O load, for no apparent reason. What could be causing this?

A6.4: On some Linux distributions, for example, Red Hat EL 3.0, the system is configured by default to run the file system indexing utility **updatedb** through the **cron** daemon on a periodic basis (usually daily). This utility traverses the file hierarchy and generates a rather extensive amount of I/O load. For this reason, it is configured by default to skip certain file system types and nonessential file systems. However, the default configuration does not prevent **updatedb** from traversing GPFS file systems.

In a cluster this results in multiple instances of **updatedb** traversing the same GPFS file system simultaneously. This causes general file system activity and lock contention in proportion to the number of nodes in the cluster. On smaller clusters, this may result in a relatively short-lived spike of activity, while on larger clusters, depending on the overall system throughput capability, the period of heavy load may last longer. Usually the file system manager node will be the busiest, and GPFS would appear sluggish on all nodes. Re-configuring the system to either make **updatedb** skip all GPFS file systems or only index GPFS files on one node in the cluster is necessary to avoid this problem.

Q6.5: What considerations are there when using IBM Tivoli Storage Manager with GPFS?

A6.5: When using Tivoli Storage Manager with GPFS, please verify the supported environments:

- IBM Tivoli Storage Manager Requirements for IBM AIX Client at www.ibm.com/support/docview.wss?rs=663&context=SSGSG7&uid=swg21052226&loc=en_US&cs=utf-8&lang=en
- IBM Tivoli Storage Manager Linux x86 Client Requirements at www.ibm.com/support/docview.wss?rs=663&context=SSGSG7&uid=swg21052223&loc=en_US&cs=utf-8&lang=en
- To search TSM support information go to www.ibm.com/software/sysmgmt/products/support/IBMTivoliStorageManager.html and enter **GPFS** as the search term

Q6.6: Are there any special considerations for using the automounter feature in my GPFS cluster?

A6.6: GPFS supports the automounter feature only on AIX 5L, SUSE Linux ES 8, and SUSE Linux ES 9 systems.

Q6.7: Are there any special considerations for the use of OpenSSL?

A6.7: To help enhance the security of mounts using Secure Sockets Layer (SSL) a working version of OpenSSL must be installed. This version must be compiled with support for the Secure Hash Algorithm (SHA).

- On AIX 5L, OpenSSL 0.9.7d-2 or later, as distributed by IBM, is required. Other publicly available versions of OpenSSL may not work. The required version of OpenSSL is included in the AIX Toolbox for Linux Applications:
  1. Go to www.ibm.com/servers/aix/products/aixos/linux/download.html
  2. Under **Sorted download**, click on **AIX Toolbox Cryptographic Content**.
  3. Either register for an IBM ID or sign-in.
  4. To view the license agreement, click on **View license**.
  5. By clicking **I agree** you agree that you have had the opportunity to review the terms and conditions and that such terms and conditions govern this transaction. If you disagree, click **I disagree**.
  6. Scroll down to **OpenSSL -- SSL Cryptographic Libraries**
  7. Ensure you download 0.9.7d-2 or later.
- For the supported versions of Linux:

- For the Red Hat EL 3 and SUSE Linux ES 9 distribution, this is currently OpenSSL 0.9.7, as included with your distribution.
- For the SUSE Linux ES 8 distribution on x86, this is currently OpenSSL 0.9.6, as included with your distribution.
- For SUSE Linux ES 8 for PowerPC64 you must compile and install OpenSSL version 0.9.7f, according to these directions, before mounting any GPFS file systems that belong to other GPFS clusters (If you are running GPFS V2.3, ensure you are at least at the minimum service level. See the question *What is the current service information for GPFS?*):

  1. Download the file **openssl-0.9.7f.tar.gz**, or later, from www.openssl.org.
  2. Unpack the file **openssl-0.9.7f.tar.gz**:

     ```
     tar xfz openssl-0.9.7f.tar.gz
     cd openssl-0.9.7f
     ```

  3. Edit the script **Configure** , changing **gcc** to **/opt/cross/bin/powerpc64-linux-gcc**:

     ```
     398c398
     < "linux-ppc64", "gcc:-bpowerpc64-linux -DB_ENDIAN -DTERMIO -O3 -fomit-frame-pointer
     -Wall::-D_REENTRANT::-ldl:SIXTY_FOUR_BIT_LONG RC4_CHAR RC4_CHUNK DES_RISC1
     DES_UNROLL:asm/linux_ppc64.o:::::::::dlfcn:linux-shared:-fPIC:-bpowerpc64-linux:.so.
     \$(SHLIB_MAJOR).\$(SHLIB_MINOR)",
     ---
     > "linux-ppc64", "/opt/cross/bin/powerpc64-linux-gcc:-bpowerpc64-linux -DB_ENDIAN
     -DTERMIO -O3 -fomit-frame-pointer -Wall::-D_REENTRANT::-ldl:SIXTY_FOUR_BIT_LONG
     RC4_CHAR RC4_CHUNK DES_RISC1
     DES_UNROLL:asm/linux_ppc64.o:::::::::dlfcn:linux-shared:-fPIC:-bpowerpc64-linux:.so.
     \$(SHLIB_MAJOR).\$(SHLIB_MINOR)",
     ```

  4. Run this script:

     ```
     ./Configure --prefix=/usr/local/ linux-ppc64
     ```

  5. Build and install the OpenSSL library:

     ```
     make
     make install
     ```

  6. Update the library cache:

     ```
     ldconfig
     ```

  7. Configure all of the PowerPC64 nodes in the GPFS cluster, listed in the file **PPC64nodes**, to use the edited library:

     ```
     mmchconfig openssllibname=/usr/local/lib/libssl.so.0.9.7 -n PPC64nodes
     ```

Q6.8: When I allow other clusters to mount my file systems, is there a way to restrict access permissions for the root user?

A6.8: Yes. Starting with GPFS 2.3.0.4, a *root squash* option is available when making a file system available for mounting by other clusters using the **mmauth** command. This option is similar in spirit to the NFS root squash option. When enabled, it causes GPFS to *squash* superuser authority on accesses to the affected file system on nodes in remote clusters.

This is accomplished by remapping the credentials: user id (UID) and group id (GID) of the root user, to a UID and GID specified by the system administrator on the home cluster, for example, the UID and GID of the user **nobody**. In effect, root squashing makes the root user on remote nodes access the file system as a non-privileged user.

Although enabling root squash is similar in spirit to setting up UID remapping (see www.ibm.com/servers/eserver/clusters/whitepapers/uid_gpfs.html), there are two important differences:

1. While enabling UID remapping on remote nodes is an option available to the remote system administrator, root squashing need only be enabled on the local cluster, and it will be enforced on remote nodes.
2. While UID remapping requires having an external infrastructure for mapping between local names and globally unique names, no such infrastructure is necessary for enabling root squashing.

When both UID remapping and root squashing are enabled, root squashing overrides the normal UID remapping mechanism for the root user. See the **mmauth** command man page for further details.

Service questions

Q7.1: What support services are available for GPFS?

A7.1: Support services for GPFS include:

- GPFS support page at www14.software.ibm.com/webapp/set2/sas/f/gpfs/home.html
- IBM Global Services - Support Line for Linux

  A 24x7 enterprise-level remote support for problem resolution and defect support for major distributions of the Linux operating system. Go to www.ibm.com/services/us/index.wss/so/its/a1000030.
- IBM Client Technology Center (CTC) for Lab Services

  A services group that assists clients worldwide in developing and deploying server and storage solutions. Go to www.ibm.com/servers/eserver/services/.
- Subscription service for pSeries, p5, and OpenPower

  This service provides technical information for IT professionals who maintain pSeries, p5 and OpenPower servers. Subscribe at www14.software.ibm.com/webapp/set2/subscriptions/pqvcmjd
- GPFS software maintenance

  GPFS defect resolution for current holders of IBM software maintenance contracts:
  - In the United States contact us toll free at 1-800-IBM-SERV (1-800-426-7378)
  - In other countries, contact your local IBM Service Center
- GPFS technical discussion forum

  General questions and IBM Scholars Program users can use the GPFS technical discussion forum at www.ibm.com/developerworks/forums/dw_forum.jsp?forum=479&cat=13.

Contact gpfs@us.ibm.com for all other services or consultation on what service is best for your situation.

Q7.2: What is the current service information for GPFS?

A7.2: The current GPFS service information includes:

- When installing or migrating GPFS , the minimum levels of service you must have applied are:
  - GPFS V3.1 you must apply APAR IY82778
  - GPFS V2.3 you must apply APAR IY63969

  If you do not apply these levels of service and you attempt to create a cluster or add a new node to an existing cluster, you will receive an error message similar to:

```
mmcrcluster -n nodeNames -p vat
mmcrcluster:  Required service not applied. Install GPFS 2.3.0.1 or later.
mmcrcluster:  The level of GPFS on node vat.ibm.com does not support the requested action.
```

- Required service for GPFS V3.1 support of SLES 10 includes:
  1. GPFS Service update 3.1.0-8 available at https://www14.software.ibm.com/webapp/set2/sas/f/gpfs/download/home.html
  2. The GPFS required level of korn shell for SLES 10 support is version ksh-93r-12.16 and can be obtained using one of these architecture-specific links:

     x86 at
     https://you.novell.com/update/i386/update/SUSE-SLES/10/PTF/43ed798d45b1ce66790327fe89fb3ca6/20061201

     POWER at
     https://you.novell.com/update/ppc/update/SUSE-SLES/10/PTF/43ed798d45b1ce66790327fe89fb3ca6/20061201

     x86_64 at
     https://you.novell.com/update/x86_64/update/SUSE-SLES/10/PTF/43ed798d45b1ce66790327fe89fb3ca6/20061201
  3. For SLES 10 on POWER:
     a. The gpfs.base 3.1.0-0 rpm must be installed using the rpm **--nopre** flag BEFORE any updates can be applied.

b. **/etc/init.d/running-kernel** shipped prior to the availability of the SLES 10 SP1 kernel source rpm contains a bug that results in the wrong set of files being copied to the kernel source tree. Until SP1 is generally available, the following change should also address the problem:

```
--- running-kernel.orig 2006-10-06 14:54:36.000000000 -0500
+++ /etc/init.d/running-kernel  2006-10-06 14:59:58.000000000 -0500
@@ -53,6 +53,7 @@
 arm*|sa110)    arch=arm ;;
 s390x)  arch=s390 ;;
 parisc64)  arch=parisc ;;
+ ppc64)     arch=powerpc ;;
 esac
 # FIXME: How to handle uml?
```

- GPFS V2.3 file systems may mount file systems that belong to other GPFS V2.3 clusters. This function is available with APARs IY64709 and IY66584 (see the question, *How do I download fixes for GPFS?*):

  Please see the updated HTML files at publib.boulder.ibm.com/infocenter/clresctr/index.jsp?topic=/com.ibm.cluster.gpfs.doc/ gpfsbooks.html or the documentation errata file at publib.boulder.ibm.com/infocenter/clresctr/topic/com.ibm.cluster.gpfs.doc/gpfs23_doc_updates/ docerrata.html for:
  - A replacement of chapter 3 *Accessing GPFS file systems from other GPFS clusters* from the *GPFS: Administration and Programming Reference*.
  - Updates to the **mmauth** and **mmremotecluster** commands.
- Included with GPFS 2.3.0-6 is a performance feature for very large files. After this maintenance level is applied, issue the **mmchfs** *device* **-V** command to enable this feature for existing file systems.

  Prior to issuing the **mmchfs -V** command on existing file systems or creating new files systems, it is suggested that all nodes in the cluster are at this maintenance level. If for any reason there are nodes in the cluster which are not at the 2.3.0-6 maintenance level, they should not be designated as nodes available to become the File System Manager. Should a node at a level lower than 2.3.0-6 become the File System Manager, conflicting messages will be sent to nodes at the 2.3.0-6 level causing them to go down.
- If you applied GPFS fix level 2.3.0-4 to your Linux system, please move to 2.3.0-5. See the question *How do I download fixes for GPFS?*
- When running GPFS on either a p5-590 or a p5-595:
  - The minimum GFW (system firmware) level required is SF222_081 (GA3 SP2), or later.

    For the latest firmware versions, see the IBM Technical Support at www14.software.ibm.com/webapp/set2/firmware/gjsn
  - The supported Linux distribution is SUSE Linux ES 9.
  - Scaling is limited to 16 total processors.
- IBM testing has revealed that some customers using the Gigabit Ethernet PCI-X adapters with the **jumbo frames** option enabled may be exposed to a potential data error. While receiving packet data, the Gigabit Ethernet PCI-X adapter may generate an erroneous DMA address when crossing a 64 KB boundary, causing a portion of the current packet and the previously received packet to be corrupted.

  These Gigabit Ethernet PCI-X adapters and integrated Gigabit Ethernet PCI-X controllers could potentially experience this issue:
  - Type 5700, Gigabit Ethernet-SX PCI-X adapter (Feature Code 5700)
  - Type 5701, 10/100/1000 Base-TX Ethernet PCI-X Adapter (Feature code 5701)
  - Type 5706, Dual Port 10/100/1000 Base-TX Ethernet PCI-X Adapter (Feature code 5706)
  - Type 5707, Dual Port Gigabit Ethernet-SX PCI-X Adapter (Feature code 5707)
  - Integrated 10/100/1000 Base-TX Ethernet PCI-X controller on machine type 7029-6C3 and 6E3 (p615)
  - Integrated Dual Port 10/100/1000 Base-TX Ethernet PCI-X controller on machine type 9111-520 (p520)

– Integrated Dual Port 10/100/1000 Base-TX Ethernet PCI-X controller on machine type 9113-550 (p550)
– Integrated Dual Port 10/100/1000 Base-TX Ethernet PCI-X controller on machine type 9117-570 (p570)

This problem is fixed with:
– For AIX 5L 5.2, APAR IY64531
– For AIX 5L 5.3, APAR IY64393

- IBM testing has revealed that some customers with the General Parallel File System who install AIX 5L Version 5.2 with the 5200-04 Recommended Maintenance package (**bos.mp64** at the 5.2.0.40 or 5.2.0.41 levels) and execute programs which reside in GPFS storage may experience a system wide hang due to a change in the AIX 5L loader. This hang is characterized by an inability to login to the system and an inability to complete some GPFS operations on other nodes. This problem is fixed with the AIX 5L APAR IY60609. It is suggested that all customers installing the **bos.mp64** fileset at the 5.2.0.40 or 5.2.0.41 level, who run GPFS, immediately install this APAR.
- Service bulletins for pSeries, p5, and OpenPower servers at www14.software.ibm.com/webapp/set2/subscriptions/pqvcmjd
  1. Sign in with your IBM ID.
  2. Under the **Bulletins** tab:
     – For the **Select a heading** option, choose **Cluster on POWER**.
     – For the **Select a topic** option, choose **General Parallel File System**.
     – For the **Select a month** option, select a particular month or choose to **All months**.

Q7.3: How do I download fixes for GPFS?

A7.3: To download fixes for GPFS, go to www14.software.ibm.com/webapp/set2/sas/f/gpfs/download/home.html

Q7.4: What are the current GPFS advisories?

A7.4: The current GPFS advisories are:

- We have recently discovered and fixed a subtle GPFS issue in truncate, where if multiple nodes are accessing the same file against which a truncate is issued on one of the nodes, a time window existed during which incorrect size information could be communicated to some nodes, which may cause GPFS to mishandle the last fragment of the file. This could lead to various failed internal consistency checks, manifested by the GPFS daemon shutting down abnormally.

  The affected GPFS PTF levels are:
  – GPFS 3.1.0-6
  – GPFS 3.1.0-5
  – GPFS 2.3.0-17
  – GPFS at 2.3.0-16
  – GPFS at 2.3.0-15

  Recommended action:
  – For customers running GPFS 3.1.0.*x* PTF 7 contains a fix and is available at www14.software.ibm.com/webapp/set2/sas/f/gpfs/download/home.html
  – For all other versions, please contact support.
- Customers running IBM Virtual Shared Disk V4.1 using a communications adapter other than the IBM eServer pSeries High Performance Switch, who have configured IBM Virtual Shared Disk with an IP packet size greater then the Max Transfer Unit (MTU) of the network, may experience packet corruption.

  IP must fragment packets that are greater than the MTU size of the network. On faster interconnects such as Gigabit Ethernet, the IP fragmentation buffer can be overrun and end up incorrectly assembling the fragments. This is an inherent limitation of the IP protocol, which can occur when the number of packets transferred exceeds the counter size, which then rolls over, potentially resulting in a duplicate packet number.

  If a duplicate packet number occurs, and the checksum matches that of the expected packet, corruption of the IBM Virtual Shared Disk packets can result in GPFS file system corruption.

IBM Virtual Shared Disk will attempt to validate the incoming packets and discard misformed packets, but it can not identify them every time (since checksums for different data patterns may be the same).

The level of IBM Virtual Shared Disk affected (shipped in AIX 5.2.*x* and later releases) has been available since October 2003, and the problem has only been confirmed as having occurred in an internal IBM test environment.

IP fragmentation can be prevented by configuring the IBM Virtual Shared Disk IP packet size less than or equal to the MTU size of the network. This will move the fragmentation into the IBM Virtual Shared Disk layer, which can correctly process the fragmentation.

The current IBM Virtual Shared Disk infrastructure allows for 160 packets per request which will limit the maximum buddy buffer size that can be used. For example:

- for an MTU of 1500, you need to set the IBM Virtual Shared Disk IP packet size to 1024 effectively limiting the maximum buddy buffer size to 128 KB.
- for an MTU of 9000, you need to set the IBM Virtual Shared Disk IP packet size to 8192 effectively limiting the maximum buddy buffer size to 1 MB.

You can check the IBM Virtual Shared Disk IP packet size with these two commands:

vsdatalst -n
: Shows you the value that will take affect at the next reboot.

statvsd
: Show you the current value that the IBM Virtual Shared Disk device driver is using.

Here is an example of how to set the IP packet size when using jumbo Ethernet frames (MTU = 9000):

```
updatevsdnode -n  ALL  -M  8192
dsh -a ctlvsd  -M  8192
```

For more information see the *RSCT for AIX 5L Managing Shared Disks* manual at publib.boulder.ibm.com/ infocenter/clresctr/index.jsp?topic=/com.ibm.cluster.rsct.doc/ rsctbooks.html and search on the commands **vsdnode**, **updatevsdnode**, and **ctlvsd**.

APAR IY66940 will completely prevent IP fragmentation and will enforce the IBM Virtual Shared Disk IP packet size being less than the MTU size. This will also remove the restrictions relating to the maximum IBM Virtual Shared Disk buddy buffer size.

Anyone who cannot take the preventive action, for whatever reason, or is unsure whether their environment may be affected, should contact IBM service to discuss their situation:

- In the United States contact us toll free at 1-800-IBM-SERV (1-800-426-7378)
- In other countries, contact your local IBM Service Center

Q7.5: What Linux kernel patches are provided for clustered file systems such as GPFS?

A7.5: The Linux kernel patches provided for clustered file systems are expected to correct problems that may be encountered when using GPFS with the Linux operating system. The supplied patches are currently being submitted to the Linux development community but may not be available in particular kernels. It is therefore suggested that they be appropriately applied based on your kernel version and distribution.

A listing of the latest patches, along with a more complete description of these patches, can be found at the General Parallel File System project on SourceForge®.net at sourceforge.net/tracker/ ?atid=719124&group_id=130828&func=browse:

1. Click on the **Summary** description for the desired patch.
2. Scroll down to the **Summary** section on the patch page for a description of and the status of the patch.
3. To download a patch:
   a. Scroll down to the **Attached Files** section.
   b. Click on the **Download** link for your distribution and kernel level.

site.mcr consideration: Patches listing a **site.mcr** define have additional steps to perform:

1. Apply the patch to the Linux kernel, recompile, and install this kernel.

2. In **site.mcr** either **#define** the option or uncomment the option if already present. Consult **/usr/lpp/mmfs/src/README** for more information.
3. Recompile and reinstall the GPFS portability layer.

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only IBM's product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any of IBM's intellectual property rights may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10594-1785
USA

For license inquiries regarding double-byte (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

IBM World Trade Asia Corporation
Licensing
2-31 Roppongi 3-chome, Minato-ku
Tokyo 106-0032, Japan

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law:

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Corporation
Intellectual Property Law
2455 South Road,P386
Poughkeepsie, NY 12601-5400
USA

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some

measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrates programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

If you are viewing this information softcopy, the photographs and color illustrations may not appear.

Trademarks

The following terms are trademarks of the International Business Machines Corporation in the United States or other countries or both:

- AIX 5L
- BladeCenter
- Enterprise Storage Server
- eServer
- IBM
- OpenPower
- POWER
- POWER5
- pSeries
- Passport Advantage
- RS/6000
- Redbooks
- SP
- ServerProven
- ServeRAID
- System p5
- Tivoli
- TotalStorage
- xSeries

Intel, Intel Inside (logos), MMX and Pentium are trademarks of Intel Corporation in the United States, other countries, or both.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Red Hat, the Red Hat "Shadow Man" logo, and all Red Hat-based trademarks and logos are trademarks or registered trademarks of Red Hat, Inc., in the United States and other countries.

UNIX is a registered trademark of the Open Group in the United States and other countries.

Other company, product, and service names may be the trademarks or service marks of others.

December 2006