Tru64 UNIX Best Practice

Configuring the AlphaServer GS320 for Performance

March 2002

Product Version: **Tru64 UNIX Version 5.1 or higher**

This Best Practice describes how to configure and tune an AlphaServer GS320 server to improve performance.

Compaq Computer Corporation
Houston, Texas

Contents

Configuring the AlphaServer GS320 for Performance

Configuring the AlphaServer GS320 for Performance

The recommendations in this Best Practice result from testing an AlphaServer GS320 running Tru64 UNIX Version 5.1 or higher. The AlphaServer GS320 consists of a collection of tightly coupled CPUs, memory modules, and I/O controllers connected through a fast interconnect. When configured properly, the AlphaServer GS320 offers superior performance and memory access times.

See the full set of specifications and configurations at the following URL:

http://www.compaq.com/alphaserver/gs320/

This Best Practice is a synopsis of the following presentation:

http://www.tru64unix.compaq.com/docs/best_practices/BP_GSPRESENTATION/slides/gs320opt/sld001.htm

See the Tru64 UNIX Best Practices Web page for more information about Best Practices documentation:

http://www.tru64unix.compaq.com/docs/best_practices/index.html

Is This Best Practice Right for You?

Not all Best Practices apply to all configurations, so you must be sure that it is appropriate for your system and circumstances. To use this Best Practice, you must meet the requirements described in the following table:

Requirement	Description
Operating System	Tru64 UNIX Version 5.1 or higher
System Configuration	AlphaServer GS320 systems
Application	You are running an application or workload that has multiple processes and I/O patterns to make full use of the CPU, memory, and I/O path available in a GS series platform.

Before You Begin

Before you apply the Best Practice to configure AlphaServer GS320 servers, you must understand some background information and perform some preliminary tasks.

If configured properly, an AlphaServer GS320 will give you consistent performance under heavy loads. For the majority of applications, a full Quad Building Block (QBB) should contain:

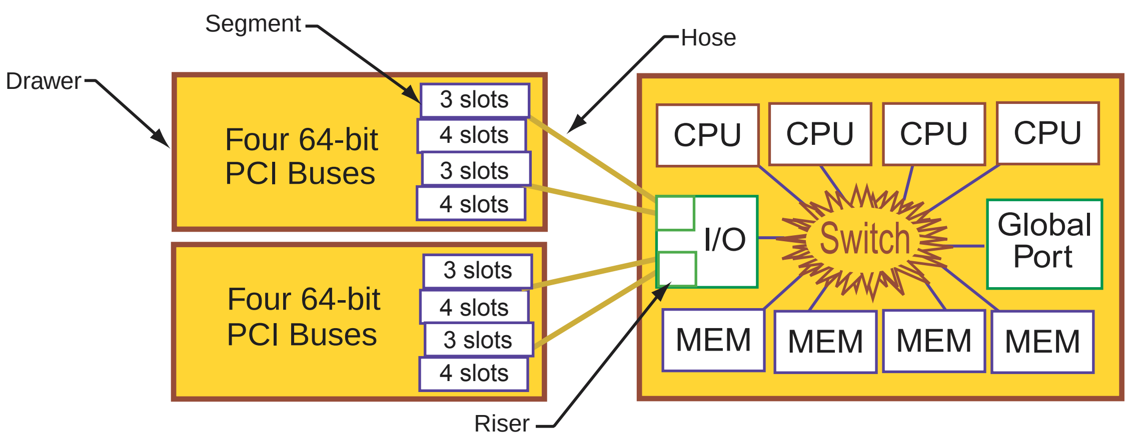
- Four CPUs that read from and write to four memory options through a high-performance switch.
- An I/O port, attached to a high-performance switch, that lets you access PCI buses.
- A global port that lets you connect to other QBBs.

Alpha Server GS320 Quad Building Block shows a full QBB. Each QBB has an I/O interface module containing two risers. Each riser has two external connections called hoses. These hoses are connected to I/O components in PCI drawers. Each riser and PCI drawer has two connections for two hoses; therefore, a QBB fully configured for I/O will be connected to two drawers with a total of four hoses.

Within a drawer, each hose is connected to two PCI buses called segments. Because of the physical layout of the drawers, a hose is connected to two segments that are asymmetrical: one segment has three slots for I/O cards and the other has four slots for I/O cards.

PCI drawers come in two types: a master drawer and an expander drawer. A master drawer is required to create a console environment and has one I/O card slot in one segment used for a standard I/O module, which contains the default CD, network, disk and console connections. This leaves only 13 usable I/O card slots within the drawer. An expander drawer cannot be used to create a console environment, does not contain the standard I/O module, so it has 14 usable I/O card slots.

Alpha Server GS320 Quad Building Block



Applying the Best Practice

Before you apply this best practice, be sure to follow the recommendations in *Before You Begin*.

Maintaining Balance and Staying Local

To maintain balance and to stay local, your system should meet these requirements:

- Each QBB should have an equal amount of CPUs.
- Every QBB should be full (have 4 CPUs).
- Each CPU should get its data from local memory.

  Use Tru64 UNIX Version 5.1 or higher to ensure you get local memory. These operating systems are Non-Uniform Memory Access (NUMA) aware and know how to make processes get memory locally within QBBs, and how to distribute the work within the system. For more information on NUMA, see

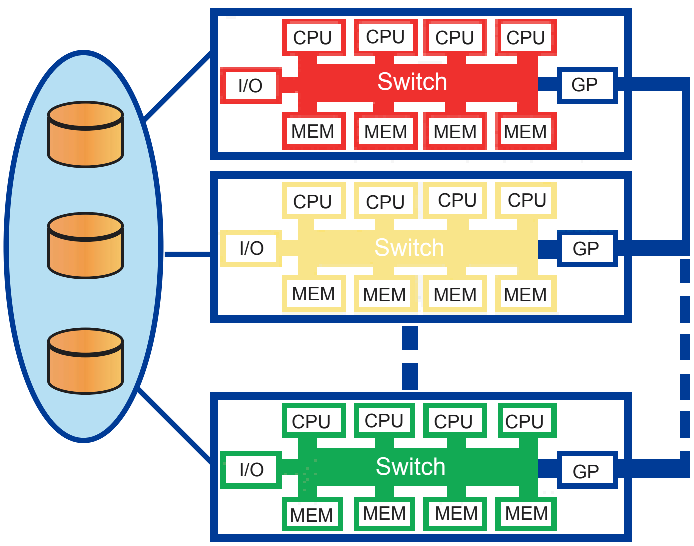
`http://www.tru64unix.compaq.com/docs/base_doc/DOCUMENTATION/V51_HTML/NUMA/TITLE.HTM`

- Each CPU should get its I/O from disks that are attached to that QBB.

  Configure your system with Fibre Channel storage. All your storage will go through a Fibre Channel fabric that is connected to every single QBB. Every QBB sees all the storage, which guarantees local access from any QBB.

Achieving High Performance shows a best case architecture where the CPUs are getting all their data from local memory, and all of their I/O from disks that are attached to that QBB.

Achieving High Performance



Increasing Your CPU Performance

To increase your CPU performance, consider the following recommendations:

- Add more CPUs through SMP (this allows more users to do the same amount of work). The AlphaServer GS320 can have up to 32 processors with four system boxes and eight QBBs.
- Upgrade your CPU (this allows the same number of users to do work faster). The fastest CPU supported by the AlphaServer GS320 is the EV68 at 1GHz.
- Select the processor count according to your application requirements and if possible:
  - Fill all the CPU slots in a utilized QBB.
  - Densely pack CPUs into the smallest number of QBBs. Fewer QBBs reduce inter-QBB accesses.
- Be symmetric, but if you cannot avoid asymmetry:
  - Take the lower QBB count as your first-level tradeoff.

For example, having 7 CPUs configured as 4/3/0/0 across QBBs is usually better than having them configured as 2/2/2/1. The 4/3/0/0 configuration leaves two QBBs empty for future growth.

  – Create balance if the QBB count does not change.

    For example, having 10 CPUs configured as 4/3/3 across QBBs is probably better than having them configured as 4/4/2.

- For extreme Very Large Memory (VLM) and I/O configurations, such as applications that prioritize on data rate, it is acceptable to spread your limited number of CPUs across QBBs.

Increasing Your Memory Performance

To improve your memory access and memory latency, consider the following recommendations:

- Use 16-way interleaving per QBB for commercial applications. For memory bandwidth-intensive applications, use 32-way interleaving per QBB (4 modules, 8 options). The memory interleaving enables the processors to access memory banks simultaneously rather than sequentially.
- Create balance and symmetry within a QBB.

  Asymmetric configurations cause the system to interleave at the lowest common denominator. For example, if you have a system with four memory options that can be configured to 16-way interleave and one memory option that can be configured to 8-way interleave, the system will be limited to an 8-way interleave.
- Create balance and symmetry between QBBs.

  Have equal amounts of memory in all QBBs. This makes it easier for Tru64 UNIX Version 5.1 or higher to keep memory local to processes.

Increasing Your I/O Performance Within a QBB

The I/O is the slowest part of the system. It is therefore important to configure it properly.

To increase your I/O performance, consider the following recommendations:

- Include at least one PCI drawer for every QBB. For high I/O use, you should have two PCI drawers.

- Use KGPSA Fibre Channel adapters to connect to the Fibre Channel fabric.
- Use HSG80 controllers.

  For each HSG80 controller:

  – Configure enough spindles.

  – Use RAID 5 or RAID 0+1.

    RAID 5 does redundant striping using a parity disk; RAID 0+1 does striping and mirroring combined. The latter is the best combination of performance and availability but it is also costly; you need twice as many disks as you have storage.

  – Divide up units between the dual controllers for load balancing.

  – Set the appropriate chunk size. A good default amount is 64 K (128 blocks) for Tru64 UNIX.

  – Set the maximum cached transfer size for caching small transactions.

  – Set in multibus failover mode.
- Do not configure more adapters (storage, cluster interconnect, or network) per PCI hose/segment pair than can be supported in terms of the bandwidth capability of the hose/segment pair.

  For example, when using KGPSA cards:

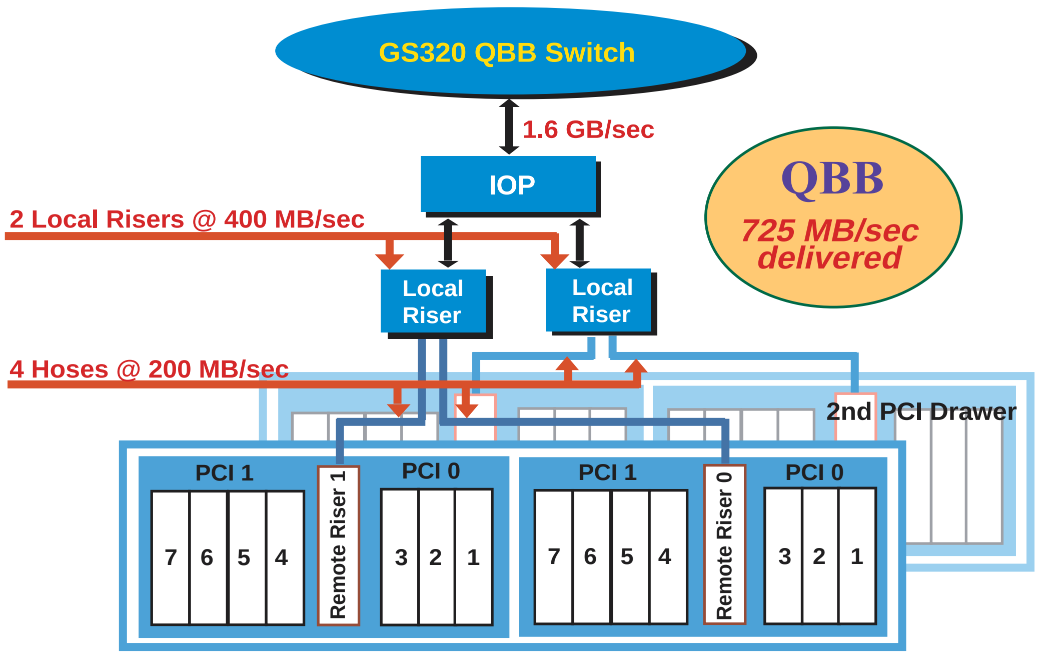
  – Each hose can currently handle up to 200 MB/sec. Therefore, insert no more than two 1-GB KGPSA-CAs or one 2-GB KGPSA-DA in the segments connected to the hose.

  – In the case of the two KGPSA-CA, insert one card in each of the segments attached to the hose.

  Follow a similar process when using other high performance and performance critical I/O cards.

GS320 QBB PCI Performance shows the performance and input of the GS320 AlphaServer between a QBB and two PCI drawers.

GS320 QBB PCI Performance

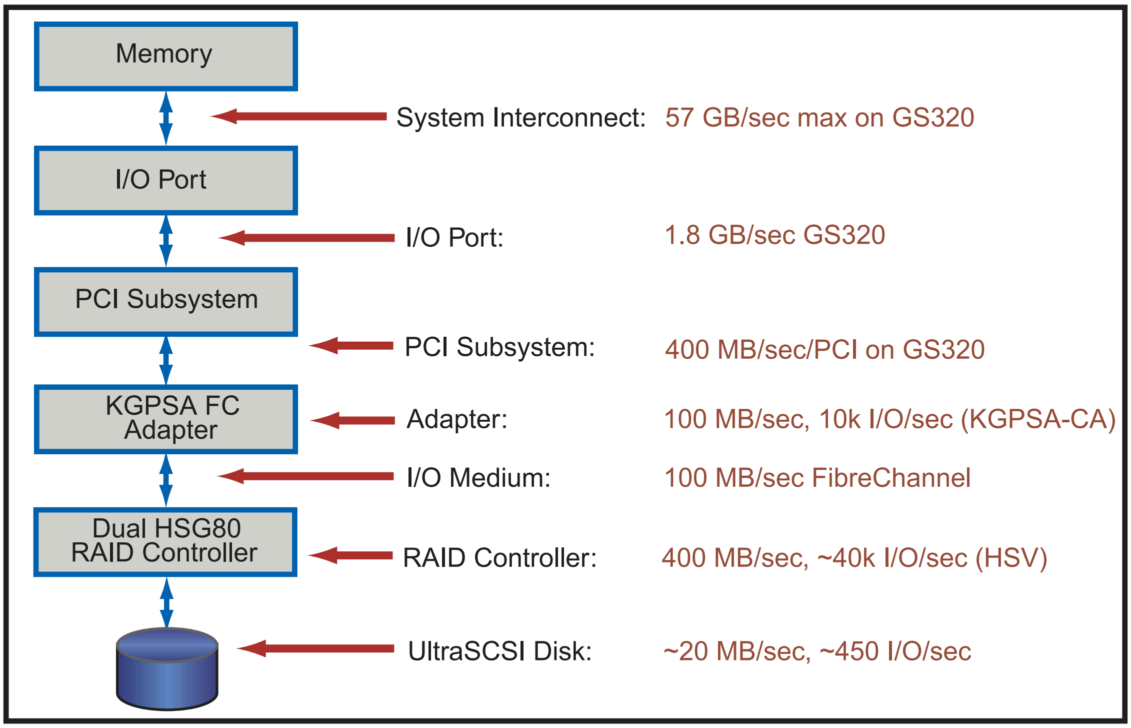


PCI Drawer / 33 MHz

Configuring to Component Peaks

After selecting and configuring your hardware components, you need to understand the chain of performance. Each component must be able to support and handle the same speed and feed as its predecessor to maximize the performance at all levels. *Configure to Component Peaks* shows the various levels of hardware and the speed and feed that they can support.

Configure to Component Peaks



Using Fibre Channel Storage Fabrics

A Fibre Channel fabric is an interconnected collection of host adapters, cables, switches, and storage. Multiple fabrics can connect a host to the same or different storage endpoints. A storage area network (SAN) is created when multiple hosts are connected to a fabric.

The GS320 achieves maximum I/O performance and availability when all its QBBs have multiple paths through multiple Fibre Channel fabrics to its connected storage.

On an AlphaServer GS320, you can achieve multipathing at two levels:

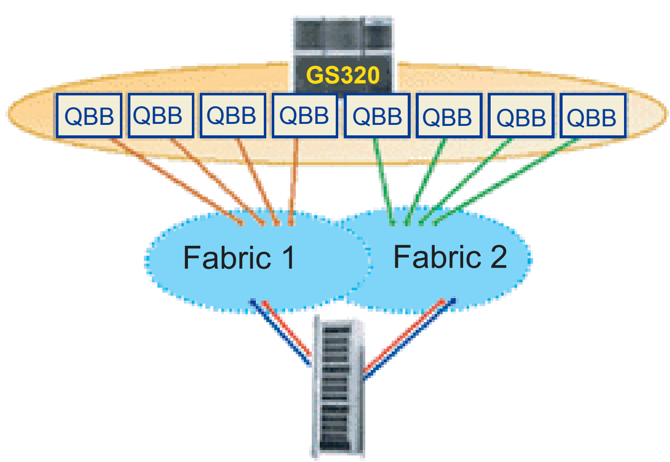
- The host level, common to other non AlphaServer GS systems, where multiple host bus adapters (HBAs) link the system through one or more fabrics to the same storage end points.
- The QBB level, which is a lower level, where not only does the GS320 host have multiple HBAs, but it has at least two in every QBB. These

HBAs link each QBB through one or more fabrics to the same storage end points.

The GS320 achieves best performance when each QBB has at least two HBAs, creating multiple paths to every storage unit.

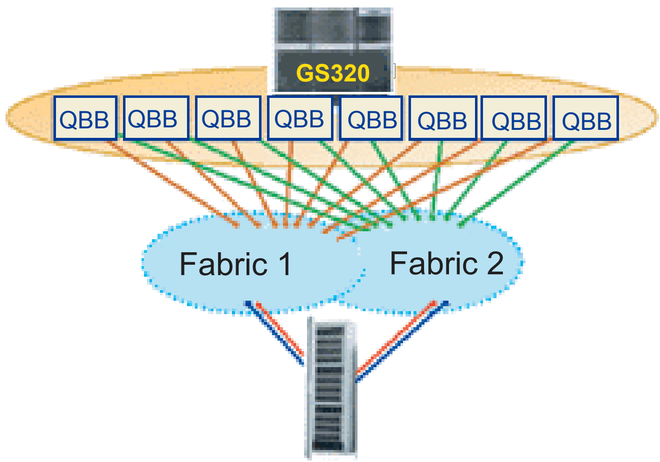
GS320 Fibre Channel Fabric: One HBA per QBB shows a GS320 where each QBB has an HBA that connects it to one of two fabrics. Both fabrics are connected to the same storage. At the host level, there are multiple paths to the storage — one per QBB. At the QBB level, there is only one path to the storage. Because the configuration includes two fabrics, the environment is highly available, which makes it a reasonable configuration for performance and availability.

GS320 Fibre Channel Fabric: One HBA per QBB



GS320 Fibre Channel Fabric: Two HBAs per QBB shows that for a GS320 Fibre Channel SAN with multipathing, each QBB has two HBAs. In this case, each QBB is connected to both fabrics, providing multiple paths. The paths are not only at the host level but also at the QBB level. This configuration is the optimal configuration for performance and availability.

GS320 Fibre Channel Fabric: Two HBAs per QBB



Using Tape Devices

You must take special considerations if you are using tape devices in order to achieve more optimal backup performance. Because tapes are single stream devices, Tru64 UNIX can only use one path at a time to access them, even if there are multiple paths available. Tru64 UNIX selects one path for each tape device and uses that path until an error occurs. If an error occurs, it tries to select a different path to ensure failover. When you have multiple HBAs and multiple tape devices connected to the same fabric, the system will most likely choose the same HBA as the default path for all tapes. When Tru64 UNIX probes for paths to devices, the paths through the first HBA will appear as the path to all tapes.

To create multiple paths for tape devices through multiple HBAs, you must use Selective Storage Presentation (SSP) on the MDR device that connects the tape device to the fabric. The SSP will force specific HBAs to see specific tape devices and will make it the only available paths. Future releases of Tru64 UNIX or future patches may eliminate this need.

Verifying Success

After you apply the Best Practice to configure the AlphaServer GS320 system for performance, you can verify whether it was successful.

You can use the following tools to verify performance of an AlphaServer GS320:

- `collect -s c`

Shows CPU statistics.

- `collect -s d`

  Shows disk statistics.

- `collect -s m`

  Shows aggregate memory statistics.

- `vmstat -R`

  Shows per QBB memory statistics.

The `collect` command can also be run in the background to be replayed later on the command line or graphically with the `collgui` graphical user interface. By using these tools, you can verify that the CPU, I/O, and memory load are evenly distributed on your system.

If the Best Practice was not successful, see *Troubleshooting* for information about identifying and solving problems.

Troubleshooting

If you determine that the Best Practice was not successful as described in *Verifying Success*, use the following table to identify and solve problems:

Problem	Possible Solutions
Hardware errors	• Look in the `/var/adm/syslog.dated` log file for system ASCII logs. • Use Compaq Analyze to view system binary logs. • Use SysMan Event Viewer for event logs.
Accidental off line CPUs	Check the processor state with `psrinfo`.
Misconfigured hardware	Use `hwmgr` to confirm that your hardware configuration matches the one described in this Best Practice.

Problem	Possible Solutions
Limited application scalability	Use `collect` to check for low memory, CPU, and I/O activity while your application is running.
General performance or stability issues	Check for the latest patch kits. Use `setld` and `dupatch` to obtain a list of patches installed on your system. Compare it to the most recent patch kit available at: http://ftp.support.compaq.com/patches/.new/unix.shtml

Comments and Questions

We value your comments and questions on the information in this document. Please mail your comments to us at this address:

`best_practices@zk3.dec.com`

Legal Notice