Intel® RAID Controllers

Best Practices White Paper

Revision 1.0

April, 2008

Enterprise Platforms and Services Division - Marketing

Revision History

Date	Revision Number	Modifications
April, 2008	1.0	Initial release.

Disclaimers

Information in this document is provided in connection with Intel® products. No license, express or implied, by estoppel or otherwise, to any intellectual property rights is granted by this document. Except as provided in Intel's Terms and Conditions of Sale for such products, Intel assumes no liability whatsoever, and Intel disclaims any express or implied warranty, relating to sale and/or use of Intel products including liability or warranties relating to fitness for a particular purpose, merchantability, or infringement of any patent, copyright or other intellectual property right. Intel products are not intended for use in medical, life saving, or life sustaining applications. Intel may make changes to specifications and product descriptions at any time, without notice.

Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined." Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them.

The Intel® RAID Controllers may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Intel, Pentium, Celeron, and Xeon are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

*Other brands and names may be claimed as the property of others.

Table of Contents

1. Overview

RAID (Redundant Array of Independent Disks) technology has been commonly implemented in server usage models as an option to provide additional data protection. RAID solutions are now also found in other computer environments such as desktops, workstations, and external storage devices which support a large number of hard drives.

Intel is committed to providing customers with stable, high-performance, and high-reliability RAID products. This white paper is intended to describe the best practices that should be used when deploying Intel® Server RAID solutions. The focus of this document is to review the implementation, configuration, and troubleshooting of RAID solutions for Intel® server products.

2. Installing and Configuring Intel® RAID Controllers

2.1 Installing Intel® RAID Controllers

Please refer to the Intel® RAID Controller Configuration Guide for information about selecting the appropriate RAID controllers and accessories for your RAID solution. For more information on how to install RAID controllers into the server solution refer to the Intel® RAID Controller Hardware User Guide. These guides are available at the Intel Support Web site: http://support.intel.com/support/motherboards/server/.

2.2 Configuring a RAID Array

Data redundancy and system performance can be enhanced by correctly installing the RAID controller and by choosing the right RAID level. The RAID level and the system usage model should match. Other configuration options include matching the virtual drive stripe size to the application. The default RAID controller configuration options are set to protect data, but they are not optimized for performance. Please refer to the Intel® RAID Software User Guide for information on how to setup RAID array solutions using Intel® RAID Controllers, and the RAID Controller Performance Tuning Guide at http://support.intel.com/support/motherboards/server/ for addition detail regarding these configuration options.

2.3 Getting the Latest RAID Software

The most current RAID drivers, firmware, and management software for Intel® RAID Controllers can be downloaded from the Intel Support Web site: http://support.intel.com/support/motherboards/server/.

3. Importance of a Backup Solution

Do not rely on a RAID subsystem as a single disaster protection plan; always keep an independent verified backup of your data in a separate physical location.

A backup plan should be an integral part of a healthy data security system. Computers and computer components can and do fail. Multiple disk failures in RAID configurations, data center catastrophes (no matter how small) and virus infections can take down a system and corrupt critical data. Often there is no warning before a failure, and then it is too late.

Data loss can be costly and may impact productivity both in terms of lost opportunity when there is no access to data, and in terms of effort and expense to recover missing data. In some situations, data may not be recoverable.

Implementing a backup plan will often turn several days of lost productivity and weeks of reorganizing information into an hour of restoring a disk image. A good backup strategy will include maintaining multiple copies of data that have been made over time so that you can recover the latest backup or step back to data files that were copied days or weeks earlier as needed. Depending on the criticality of your data and to guard against a catastrophic event, it may be a good idea to store a regular backup securely off site.

Backup copies of data can be made to a local drive or tape drive, or to a remote data backup service provider.

It is especially important to keep a copy of data that is held on storage systems or RAID arrays. Due to the large capacity of some drives and the time required to rebuild data, there is an increased risk that additional drives may fail during the rebuild, resulting in possible data loss. Even though there is redundancy built into a RAID or Storage system, a backup should be performed on a regular basis.

4. Tuning Controller Performance

Optimizing the overall performance of a RAID subsystem requires careful consideration of several factors that can affect performance, including the controller and disk drive cache settings and the interaction of these settings with system applications. The sections below provide a limited discussion of some of these factors. For a full review of performance tuning please refer to the Intel® RAID Controller Performance Optimization whitepaper, available at http://support.intel.com/support/motherboards/server/ .

Note: There are a variety of factors that can affect the performance of the RAID subsystem including PCI bus bandwidth, logical drive cache settings, stripe size, hard disk drive cache settings, RAID level, ratio of read versus write operations, ratio of sequential versus random operations, and the number of disks in an array.

4.1 Tuning Controller Cache Options

Tuning cache memory options on the RAID controller can improve performance. There are three settings available in the controller cache to allow fine tuning:

- Read Ahead Option
- Write Back Option
- Cached I/O Option

The following table provides a quick reference for RAID settings. The information is simplified and may not be accurate with some applications or tests. For detailed performance tuning information please refer to the Intel® RAID Controller Performance Optimization whitepaper, available at http://support.intel.com/support/motherboards/server/.

Read Cache Policy	Direct I/O
Read Ahead Policy	Adaptive Read Ahead
Write Cache Policy	Write Back*

* A RAID controller battery should be used whenever virtual drive write-back cache is enabled and data is mission critical.

4.2 Hard Disk Cache

Disk drive cache can be enabled in the virtual drive properties page of the RAID configuration utility. There is a risk of data loss using a hard drive cache; an overview is provided below.

Hard disk drive cache is located within the logic of the hard drive. Cache provides enhanced performance for sequential read access by retrieving adjacent data on the drive into the data buffer in case the host computer requests it. This process allows the data to be directly transferred from the drive's memory when it is requested rather than waiting for a disk access, which results in lower latency. Enabling the hard drive cache can also improve write performance by providing additional memory space for queued data. Write data can be queued in the disk cache and reported as written even though the data will not move from memory to the disk until disk access is available. This reduces the delay during disk I/O operations.

There is inherent risk in holding data in the drive cache when a write has been acknowledged as complete but hasn't been written to the disk. If the drive loses power the data in the cache will be lost before it is written to the disk that can cause a "hole" in a data file, which makes the file unusable. Using a UPS will mitigate this risk but not eliminate it.

Note: A soft or hard reset (<Ctrl> + <Alt> + <Del> or the reset button) does not affect the completion of a disk write operation because the disk cache will be flushed as long as drive power is maintained.

5. When to Use a RAID Controller Battery

A RAID controller battery should be used whenever virtual drive write-back cache is enabled and data is mission critical.

Cache-to-cache I/O is much faster than any other type of I/O operation occurring on the data bus. It is faster to write data to the RAID adapter's cache memory than it is to write it directly to a storage device because the time required to spin target data under a read or write head is longer than the time required to perform the read or write to a memory device.

If the RAID Controller's write-back cache option is enabled, data is first written to the cache memory and the write is acknowledged, and then the RAID controller writes the cached data to the storage device when it is available to service the I/O request. However, this method of writing data first to cache memory, acknowledging the write as complete, and then completing the write when the drive is available carries inherent risk. Cached data on the RAID controller can be lost if the AC power fails before the cached data is written to the storage device. The Smart Battery mitigates this risk by providing battery power to the RAID controller memory and holding the data in the RAID cache memory until power is restored. The battery can hold data in the RAID controller's memory for up to 72 hours.

The Smart Battery accomplishes all of this by monitoring the voltage level of the DRAM modules on the RAID controller. If the voltage drops below a defined level, the Smart Battery switches the memory power source from the RAID controller to the battery pack. The battery pack provides power for the memory until the voltage returns to an acceptable level, at which time the Smart Battery circuit board switches the power source back to the RAID controller. Cached data is then written to the storage device just as though the power loss had never occurred. The Smart Battery provides additional fault tolerance even when used with a UPS, which does not prevent a system power supply failure or other system internal power failure.

6. Why Use a UPS?

An uninterruptible power supply (UPS) is a battery-based system power supply that helps protect electronic equipment from an unexpected loss of power. A UPS is highly recommended to protect data in mission critical configurations. Computers and accessories can suffer damage during a power outage or experience a loss of data that is in transit during the power outage.

There is no way to provide a battery backup of data that is temporarily stored in the hard disk cache but has not been written to disk. A power outage could corrupt the data on a server or make data unavailable to users. A UPS can reduce the chance that a power outage could corrupt data on a server. Although the addition of UPS is not a guarantee that data cannot be lost, it does add additional security.

7. Enterprise-class versus Desktop-class Drives

Enterprise class hard drives should always be used on an enterprise class system. Use of a desktop class drive is not recommended due to I/O timeout incompatibilities, lower tolerances for vibration, and a lack of end-to-end data error detection and correction.

Hard drive manufacturers develop drives to meet specific customer requirements for reliability, capacity, performance and power consumption. Using drives in the application for which they were designed ensures your data is available when and how you need it. Using drives outside of their intended application can negatively impact server productivity.

7.1 Drive Vibration

A hard drive is a non-volatile storage device which typically stores data on rapidly rotating magnetic platters. Data is usually read and written by a device which is nanometers away from the surface of the platters. Vibration can significantly affect hard drive reliability and performance.

There are a number of factors that can cause drives to vibrate, including: vibrations from other drives or the drive itself, spindle imbalance or torque, and vibration from other system components such as system fans. These vibrations can cause the read/write heads to misalign with the data track. When this happens, a retry is required to ensure integrity of the read/write data. A retry requires milliseconds of time but because drive and storage subsystem electronics are operating in micro or nanoseconds, a wait of milliseconds can significantly reduce the overall performance of the storage solution.

Two general categories of drives have evolved to meet customer needs:

- **Desktop Drives**: These drives perform at an acceptable level when rotational vibration does not exceed 10 radians per second. They are typically used in single or dual-drive environments where rotational vibration is limited. A desktop drive is built to have a low read and write workload over an 8 hour period, 5 days a week. Desktop drives are designed to work in environments that do not exceed 25 degrees Celsius. As heat increases, the mean time between failures (MTBF) decreases and the drives are more likely to fail.
- **Enterprise Drives**: These drives perform at an acceptable level when rotational vibration does not exceed 21 radians per second. Enterprise drives are typically used in multi-drive environments where rotational vibration is normally above 10 radians per second. The drives are built using advanced technology and components to meet the performance, workload, and reliability requirements to perform thousands of read/writes per second, 24 hours a day, 7 days a week. Enterprise drives are also designed for higher temperatures. The low end of an enterprise drive temperature specification is at the high end of the desktop specification, making enterprise drives more reliable in applications where desktop drives fail due to high temperatures. Enterprise drives usually have a longer warranty period than desktop drives. They often have advanced power management options. These modes reduce power consumption along with server cooling requirements, which equate to lower operational costs and less thermal impact on surrounding system components. Enterprise drives with error recovery control have the ability to quickly respond with data or to return an error to the host controller. With a quick response the RAID set is preserved by rebuilding the requested data. Full

RAID set rebuilds are executed for true drive failures. Error recovery is usually not offered on desktop drives due to the cost. A desktop drive often responds too slowly to preserve the RAID and the drive is flagged for replacement. The RAID set then operates in a slower degraded mode until the drive is replaced. Complete loss of data is possible if another drive times out while the drive is in a degraded mode. With higher MTBF and error recovery, enterprise drives offer greater reliability than desktop drives in the server environment.

Hard drives last longer when used in the application for which they were designed. Desktop drives often lack workload management to lower thermal stresses, have a lower tolerance for the normal rotational vibration found in a server environment, are not designed to run 24 hours a day, 7 days a week; and may fail prematurely when installed in a server.

To get the best performance and avoid drive failures, Intel recommends using enterprise drives for server applications. SAS and SATA drives should not be mixed in the same enclosure. Please refer to the Enterprise-class versus Desktop-class Hard Drives white paper available on the http://support.intel.com/support/motherboards/server/ Web site.

8. Basic Troubleshooting

Some basic troubleshooting information is provided below for your reference.

Note: Before attempting to diagnosis RAID failures or make any changes to the RAID configuration, please confirm that a complete and verified backup of critical data is available. A verified backup exists when the backed up data has been compared against the original data.

Note: If you encounter a drive failure or an offline drive, do not remove any drives from the system (hot plug) or shut the system down until you have verified the cause of the failure. Contact Intel Customer Support if you have any questions.

8.1 Drive State Definition

The SAS Software Stack firmware defines the following states for physical disks connected to the controller:

- Unconfigured Good – A disk accessible to the RAID controller but not configured as a part of a virtual disk. For example, a new drive inserted into a system.
- Online – A disk accessible to the RAID controller and configured as part of a virtual disk.
- Failed – A disk drive that is part of a virtual disk, but has failed and is no longer usable.
- Rebuild – A disk drive to which data is being written to restore full redundancy to a virtual disk.
- Unconfigured Bad – A disk drive that is no longer part of an array and is known to be bad. This state is typically assigned to a drive that has failed, but is no longer part of a configured virtual disk because it has been replaced by a Hot-Spare drive.
- Foreign – When a disk has configuration information on the drive (metadata) that is not in the NVRAM of the controller, it is considered “Foreign”, When importing disks from a different RAID controller (foreign metadata), the physical disk is marked as foreign until user action is taken to add the configuration on the disks to the existing configuration in the NVRAM on the controller. Foreign is not actually a drive state, but rather it indicates that a drive is from another configuration. Foreign drives are typically in an unconfigured good state until they are imported into the current configuration. For example, when a system is powered on with drives from another system that contain the RAID configuration information, they are considered “foreign” until they have been accepted or declined as part of the current configuration.
- Hot spare – A disk drive that is defined as a hot spare. A hot spare is used to automatically come online and replace the first failed drive in a virtual disk. A hot spare will only come online if it is the same size or larger than the failing drive, and if a drive has been marked as failed.
- Offline – A disk drive that is still part of a configured Virtual Disk Drive, but which is not active. This state is used to represent a configured drive for which the data is not valid. This state can occur as a transition state or due to a user action.

8.2 Virtual Disk State Description

- Optimal – A virtual disk with member drives that are online.

- Partially Degraded – A virtual disk with a RAID level capable of sustaining more than one member drive failure experiences a member failure but the virtual disk is not degraded or offline.
- Degraded – A virtual disk that already has one or more member drive failures and cannot sustain a subsequent drive failure.
- Offline – A virtual disk with one or more member drive failures that cause the data to not be accessible.

8.3 Tips and Tricks

8.3.1 Setup Tips

- Check cables for proper connection.
- Verify that all the cable ends are properly seated and that none of the pins are bent.
- Verify that an approved cable is being used. Cables must be speed compatible and meet signal integrity specifications.
- SATA cables are designed to connect directly from the RAID controller to the hard drive or drive enclosure.

8.3.2 Debug Tips

- Improvements in RAID controller and hard drive communication and control are frequently incorporated into updated versions of RAID controller and hard drive firmware. It is generally recommended to review the release notes for firmware updates and apply the updates as warranted.
- Review firmware updates for the server board and intelligent backplane and complete updates as necessary.
- Grown defects may not indicate that the drive is failing. However, if the number of grown defects is large or increasing, the drive may be in the process of failing. It is recommended that you replace the drive.
- Bad block redirections may not indicate that a drive is failing. However, if the number of redirections is large or the number is increasing, the drive may in the process of failing. It is recommended that you replace the drive.
- Parity errors in a log may indicate a failing controller, failing drive, or a memory issue on the RAID controller. Replace the controller and / or hard drive. Some RAID controllers include a DIMM site; verify that the memory used is listed on the RAID controller's tested memory list. If errors persist, try changing the memory module and / or RAID controller.
- Write compare errors indicate a failing controller or failing drive. Replace the controller and / or hard drive.
- Do not reinstall a drive that has failed.

9. Summary

Intel is committed to providing customers with a stable product that offers both high-performance and high-reliability. In this document we have provided guidance on cache options, using a RAID controller battery, using a UPS, and other important options that can enhance data safety and controller performance.