Index

RAID

RAID is an acronym first defined by David A. Patterson, Garth A. Gibson and Randy Katz at the University of California, Berkeley in 1987 to describe a Redundant Array of Inexpensive Disks, a technology that allowed computer users to achieve high levels of storage reliability from low-cost and less reliable PC-class disk-drive components, via the technique of arranging the devices into arrays for redundancy.

More recently, marketers representing industry RAID manufacturers reinvented the term to describe a Redundant Array of Independent Disks as a means of disassociating a "low cost" expectation from RAID technology.

"RAID" is now used as an umbrella term for computer data storage schemes that can divide and replicate data among multiple hard disk drives. RAID's various designs all involve two key design goals: increased data reliability or increased input/output performance. When multiple physical disks are set up to use RAID technology, they are said to be in a RAID array. This array distributes data across multiple disks, but the array is seen by the computer user and operating system as one single disk. RAID can be set up to serve several different purposes.

Purpose and basics

Redundancy is achieved by either writing the same data to multiple drives (known as mirroring), or writing extra data (known as parity data) across the array, calculated such that the failure of one (or possibly more, depending on the type of RAID) disks in the array will not result in loss of data. A failed disk may be replaced by a new one, and the lost data reconstructed from the remaining data and the parity data. Organizing disks into a redundant array decreases the usable storage capacity. For instance, a 2-disk RAID 1 array loses half of the total capacity that would have otherwise been available using both disks independently, and a RAID 5 array with several disks loses the capacity of one disk. Other types of RAID arrays are arranged so that they are faster to write to and read from than a single disk.

There are various combinations of these approaches giving different trade-offs of protection against data loss, capacity, and speed. RAID levels 0, 1, and 5 are the most commonly found, and cover most requirements.

RAID 0 (striped disks) distributes data across several disks in a way that gives improved speed and no lost capacity, but all data on all disks will be lost if any one disk fails. Although such an array has no actual redundancy, it is customary to call it RAID 0.

RAID 1 (mirrored settings/disks) duplicates data across every disk in the array, providing full redundancy. Two (or more) disks each store exactly the same data, at the same time, and at all times. Data is not lost as long as one disk survives. Total capacity of the array equals the capacity of the smallest disk in the array. At any given instant, the contents of each disk in the array are identical to that of every other disk in the array.

RAID 5 (striped disks with parity) combines three or more disks in a way that protects data against loss of any one disk; the storage capacity of the array is reduced by one disk.

RAID 6 (striped disks with dual parity) (less common) can recover from the loss of two disks.

RAID 10 (or 1+0) uses both striping and mirroring. "01" or "0+1" is sometimes distinguished from "10" or "1+0": a striped set of mirrored subsets and a mirrored set of striped subsets are both valid, but distinct, configurations.

RAID can involve significant computation when reading and writing information. With traditional "real" RAID hardware, a separate controller does this computation. In other cases the operating system or simpler and less expensive controllers require the host computer's processor to do the computing, which reduces the computer's performance on processor-intensive tasks. Simpler RAID controllers may provide only levels 0 and 1, which require less processing.

RAID systems with redundancy continue working without interruption when one (or possibly more, depending on the type of RAID) disks of the array fail, although they are then vulnerable to further failures. When the bad disk is replaced by a new one the array is rebuilt while the system continues to operate normally. Some systems have to be powered down when removing or adding a drive; others support hot swapping, allowing drives to be replaced without powering down. RAID with hot-swapping is often used in high availability systems, where it is important that the system remains running as much of the time as possible.

RAID is not a good alternative to backing up data. Data may become damaged or destroyed without harm to the drive(s) on which they are stored. For example, part of the data may be overwritten by a system malfunction; a file may be damaged or deleted by user error or malice and not noticed for days or weeks; and, of course, the entire array is at risk of physical damage.

Principles

RAID combines two or more physical hard disks into a single logical unit by using either special hardware or software. Hardware solutions often are designed to present themselves to the attached system as a single hard drive, so that the operating system would be unaware of the technical workings. For example, you might configure a 1TB RAID 5 array using three 500GB hard drives in hardware RAID, the operating system would simply be presented with a "single" 1TB disk. Software solutions are typically implemented in the operating system and would present the RAID drive as a single drive to applications running upon the operating system.

There are three key concepts in RAID: mirroring, the copying of data to more than one disk; striping, the splitting of data across more than one disk; and error correction, where redundant data is stored to allow problems to be detected and possibly fixed (known as fault tolerance). Different RAID levels use one or more of these techniques, depending on the system requirements. RAID's main aim can be either to improve reliability and availability of data, ensuring that important data is available more often than not (e.g. a database of customer orders), or merely to improve the access speed to files (e.g. for a system that delivers video on demand TV programs to many viewers).

The configuration affects reliability and performance in different ways. The problem with using more disks is that it is more likely that one will fail, but by using error checking the total system can be made more reliable by being able to survive and repair the failure. Basic mirroring can speed up reading data as a system can read different data from both the disks, but it may be slow for writing if the configuration requires that both disks must confirm that the data is correctly written. Striping is often used for performance, where it allows sequences of data to be read from multiple disks at the same time. Error checking typically will slow the system down as data needs to be read from several places and compared. The design of RAID systems is therefore a compromise and understanding the requirements of a system is important. Modern disk arrays typically provide the facility to select the appropriate RAID configuration.

Index

ATA over Ethernet

ATA over Ethernet (AoE) is a network protocol developed by the Brantley Coile Company, designed for simple, high-performance access of SATA storage devices over Ethernet networks. It gives the possibility to build SANs with low-cost, standard technologies.

AoE does not rely on network layers above Ethernet, such as IP and TCP. In this regard it is more comparable to Fibre Channel over Ethernet than iSCSI. While the non-routability means AoE cannot be accessed over the Internet or other IP networks, the feature makes AoE more lightweight (with less load on the host), easier to implement, provides a layer of inherent security, and offers higher performance. The AoE specification is 8 pages compared with iSCSI's 257 pages.

Protocol Description

ATA Encapsulation

SATA (and older PATA) hard drives use the Advanced Technology Attachment (ATA) protocol to issue commands, such as read, write, and status. AoE encapsulates those commands inside Ethernet frames and lets them travel over an Ethernet network instead of a SATA or 40-pin ribbon cable. By using an AoE driver, the host operating system is able to access a remote disk as if it were directly attached.

The encapsulation of ATA provided by AoE is simple and low-level, allowing the translation to happen either at high performance or inside a small, embedded device, or both.

Routability

AoE runs directly on top of Ethernet, instead of an intermediate protocol such as TCP/IP. This reduces the significant CPU overhead of TCP/IP. However, this means that routers cannot be used to route a packet across disparate networks (such as the Internet). Instead, AoE packets can travel within a single local Ethernet storage area network (such as one created by a switch or VLAN).

Security

The non-routability of AoE is a source of inherent security (ie, an intruder can't connect through a router. He must physically plug into the local Ethernet switch). However, there are no AoE-specific mechanisms for password verification or encryption. Additional security may be implemented at the file-system level. Certain AoE targets, GGAOED for example, support access lists allowing connections only from specific MAC addresses.

Config String

The AoE protocol provides a mechanism for host-based cooperative locking. When more than one AoE initiator is using an AoE target, they must communicate. The hosts need a way to avoid interfering with one another as they use and modify the data on the shared AoE device.

One option provided by AoE is to use the storage device itself as the mechanism for determining the access of particular hosts. The AoE protocol includes a "config string" feature. The config string can record who is using the device. (It can also record any other information.) If more than one host tries to set the config string simultaneously, only one succeeds. The other host is informed of the conflict.

Index

iSCSI

In computing, iSCSI an abbreviation of Internet Small Computer System Interface, an Internet Protocol (IP)-based storage networking standard for linking data storage facilities. By carrying SCSI commands over IP networks, iSCSI is used to facilitate data transfers over intranets and to manage storage over long distances. iSCSI can be used to transmit data over local area networks (LANs), wide area networks (WANs), or the Internet and can enable location-independent data storage and retrieval. The protocol allows clients (called initiators) to send SCSI commands (CDBs) to SCSI storage devices (targets) on remote servers. It is a popular storage area network (SAN) protocol, allowing organizations to consolidate storage into data center storage arrays while providing hosts (such as database and web servers) with the illusion of locally-attached disks. Unlike traditional Fibre Channel, which requires special-purpose cabling, iSCSI can be run over long distances using existing network infrastructure.

Functionality

iSCSI uses TCP/IP (typically TCP ports 860 and 3260). In essence, iSCSI simply allows two hosts to negotiate and then exchange SCSI commands using IP networks. By doing this, iSCSI takes a popular high-performance local storage bus and emulates it over wide-area networks, creating a storage area network (SAN). Unlike some SAN protocols, iSCSI requires no dedicated cabling; it can be run over existing switching and IP infrastructure. As a result, iSCSI is often seen as a low-cost alternative to Fibre Channel, which requires dedicated infrastructure.

Although iSCSI can communicate with arbitrary types of SCSI devices, system administrators almost always use it to allow server computers (such as database servers) to access disk volumes on storage arrays. iSCSI SANs often have one of two objectives:

Storage consolidation

Organizations move disparate storage resources from servers around their network to central locations, often in data centers; this allows for more efficiency in the allocation of storage. In a SAN environment, a server can be allocated a new disk volume without any change to hardware or cabling.

Disaster recovery

Organizations mirror storage resources from one data center to a remote data center, which can serve as a hot standby in the event of a prolonged outage. In particular, iSCSI SANs allow entire disk arrays to be migrated across a WAN with minimal configuration changes, in effect making storage "routable" in the same manner as network traffic.

History

In the context of computer storage, a SAN system allows a machine to use a network protocol to connect to remote storage resources such as disks and tape drives on an IP network for block level input and output. From the point of view of the class drivers and application software, the devices appear as locally attached devices.

There is a difference between a SAN device and a network-attached storage (NAS) device, where computers access resources through a file-based interface rather than through a low-level device interface. A NAS server arbitrates access from multiple clients, thus allowing the arbitrary addition of consumers for its resources. With iSCSI, the burden of synchronizing access to shared resources generally belongs to the initiator (network client) rather than with the target (network server). Sharing low-level device interfaces is a requirement of computer clusters which use specialized cluster software to manage the use of shared resources.

Index

Network-attached storage (NAS)

Description

A NAS unit is essentially a self-contained computer connected to a network, with the sole purpose of supplying file-based data storage services to other devices on the network. The operating system and other software on the NAS unit provide the functionality of data storage, file systems, and access to files, and the management of these functionalities. The unit is not designed to carry out general-purpose computing tasks, although it may technically be possible to run other software on it. NAS units usually do not have a keyboard or display, and are controlled and configured over the network, often by connecting a browser to their network address. The alternative to NAS storage on a network is to use a computer as a file server. In its most basic form a dedicated file server is no more than a NAS unit with keyboard and display and an operating system which, while optimised for providing storage services, can run other tasks; however, file servers are increasingly used to supply other functionality, such as supplying database services, email services, and so on.

A general-purpose operating system is not needed on a NAS device, and often minimal-functionality or stripped-down operating systems are used. For example FreeNAS, which is Free / open source NAS software designed for use on standard computer hardware, is just a version of FreeBSD with all functionality not related to data storage stripped out. NASLite is a highly optimized Linux distribution running from a floppy disk for the sole purpose of a NAS. Likewise, NexentaStor is based upon the core of the NexentaOS, a Free / open source hybrid operating system with an OpenSolaris core and a GNU user environment.

NAS systems contain one or more hard disks, often arranged into logical, redundant storage containers or RAID arrays (redundant arrays of inexpensive/independent disks). NAS removes the responsibility of file serving from other servers on the network.

NAS uses file-based protocols such as NFS (popular on UNIX systems), SMB/CIFS (Server Message Block/Common Internet File System) (used with MS Windows systems), or AFP (used with Apple Macintosh Computers. NAS units rarely limit clients to a single protocol.

NAS provides both storage and filesystem. This is often contrasted with SAN (Storage Area Network), which provides only block-based storage and leaves filesystem concerns on the "client" side. SAN protocols are SCSI, Fibre Channel, iSCSI, ATA over Ethernet (AoE), or HyperSCSI.

Despite differences SAN and NAS are not exclusive and may be combined in one solution: SAN-NAS hybrid

The boundaries between NAS and SAN systems are starting to overlap, with some products making the obvious next evolution and offering both file level protocols (NAS) and block level protocols (SAN) from the same system. An example of this is Openfiler, a free software product running on Linux.

History

Network-attached storage was introduced with the early file sharing Novell's NetWare server operating system and NCP protocol in 1983. In the UNIX world, Sun Microsystems' 1984 release of NFS allowed network servers to share their storage space with networked clients. 3Com and Microsoft would develop the LAN Manager software and protocol to further this new market. 3Com's 3Server and 3+Share software was the first purpose-built servers (including proprietary hardware, software, and multiple disks) for open systems servers. Inspired by the success of file servers from Novell, IBM, and Sun, several firms developed dedicated file servers. While 3Com was among the first firms to build a dedicated NAS for desktop operating systems, Auspex Systems was one of the first to develop a dedicated NFS server for use in the UNIX market. A group of Auspex engineers split away in the early 1990s to create the integrated NetApp filer, which supported both Windows' CIFS and UNIX'es NFS, and had superior scalability and ease of deployment. This started the market for proprietary NAS devices now led by NetApp and EMC Celerra.

Starting in the early 2000s, a series of startups emerged offering alternative solutions to single filer solutions in the form of clustered NAS – Spinnaker Networks (now acquired by NetApp), Exanet, IBRIX, Isilon, PolyServe and Panasas, to name a few.

In 2009, the parallel NFS PNFS extension to the NFSv4 standard was ratified by the IETF working group.

Index

Storage area network (SAN)

Technology

A storage area network (SAN) is an architecture to attach remote computer storage devices (such as disk arrays, tape libraries, and optical jukeboxes) to servers in such a way that the devices appear as locally attached to the operating system. Although the cost and complexity of SANs are dropping, they are still uncommon outside larger enterprises.

Network attached storage (NAS), in contrast to SAN, uses file-based protocols such as NFS or SMB/CIFS where it is clear that the storage is remote, and computers request a portion of an abstract file rather than a disk block.

Index

Solid-state drive (SSD)

A solid-state drive (SSD) is a data storage device that uses solid-state memory to store persistent data. An SSD emulates a hard disk drive interface, thus easily replacing it in most applications. An SSD using SRAM or DRAM (instead of flash memory) is often called a RAM-drive.

The original usage of the term solid-state (from solid-state physics) refers to the use of semiconductor devices rather than electron tubes, but in this context, has been adopted to distinguish solid-state electronics from electromechanical devices as well. With no moving parts, solid-state drives are less fragile than hard disks and are also silent (unless a cooling fan is used); as there are no mechanical delays, they usually employ low access time and latency.

History

The first ferrite memory SSD devices, or auxiliary memory units as they were called at the time, emerged during the era of vacuum tube computers. But with the introduction of cheaper drum storage units, their use was discontinued. Later, in the 1970s and 1980s, SSDs were implemented in semiconductor memory for early supercomputers of IBM, Amdahl and Cray;[4] however, the prohibitively high price of the built-to-order SSDs made them quite seldom used.

In 1978 StorageTek developed the first modern type of solid-state drive. In the mid-1980s Santa Clara Systems introduced BatRam, an array of 1 megabit DIP RAM Chips and a custom controller card that emulated a hard disk. The package included a rechargeable battery to preserve the memory chip contents when the array was not powered. The Sharp PC-5000, introduced in 1983, used 128 kilobyte (128 KB) solid-state storage cartridges, containing bubble memory.

RAM "disks" were popular as boot media in the 1980s when hard drives were expensive, floppy drives were slow, and a few systems, such as the Amiga series, the Apple IIgs, and later the Macintosh Portable, supported such booting. Tandy MS-DOS machines were equipped with DOS and DeskMate in ROM, as well. At the cost of some main memory, the system could be soft-rebooted and be back in the operating system in mere seconds instead of minutes. Some systems were battery-backed so contents could persist when the system was shut down.

In 1995 M-Systems introduced flash-based solid-state drives. (SanDisk acquired M-Systems in November 2006). Since then, SSDs have been used successfully as hard disk drive replacements by the military and aerospace industries, as well as other mission-critical applications. These applications require the exceptional mean time between failures (MTBF) rates that solid-state drives achieve, by virtue of their ability to withstand extreme shock, vibration and temperature ranges.

Enterprise Flash drives (EFDs) are designed for applications requiring high I/O performance (Input/Output Operations Per Second), reliability and energy efficiency.

On September 25, 2007, Fusion-io announced the ioDrive to be available in Q4 2007, with capacities of 80GB, 160GB and 320GB. The ioDrive actually did not begin shipping until April 7, 2008.

OCZ has recently demoed at Cebit 2009 a 1 TB flash SSD drive utilizing PCI Express x8 interface, it achieves a minimum read speed of 654MB/s and maximum read speed of 712MB/s.

On March 2, 2009, Hewlett-Packard announced the HP StorageWorks IO Accelerator, the world's first enterprise flash drive especially designed to attach directly to the PCI fabric of a blade server. The mezzanine card, based on Fusion-io's ioDrive technology, serves over 100,000 IOPS and up to 800MB/s of bandwidth. HP provides the IO Accelerator in capacities of 80GB, 160GB and 320GB.[8]

In April 2009, Texas Memory System announced the highest capacity rack mounted flash storage unit to date, a 5TB RamSan-620. It has a throughput of 3GB/s and a sustained random read/write of

250,000 I/O's per second (IOPS). It utilizes high-speed Fibre Channel or Infiniband interface for data transfers.[9][10]

On May 5, 2009, Photofast announced the G-Monster-PROMISE PCIe SSD with capacity choices from 128GB to 1TB, with 1000MB/s of read/write speeds.

Architecture and function

Flash based drive

Most SSD manufacturers use non-volatile flash memory to create more rugged and compact devices for the consumer market. These flash memory-based SSDs, also known as flash drives, do not require batteries. They are often packaged in standard disk drive form factors (1.8-inch, 2.5-inch, and 3.5-inch). In addition, non-volatility allows flash SSDs to retain memory even during sudden power outages, ensuring data persistence. SSDs are slower than DRAM (and even traditional HDDs on big files), but may perform better than hard drives (at least with regard to reads) because of negligible seek time (flash SSDs have no moving parts, and thus eliminate spin-up time, and greatly reduce seek time, latency, and other delays inherent in conventional electro-mechanical disks).

Components:

Cache: A flash based SSD uses a small amount of DRAM as a cache, similar to the cache in Hard disk drives. A directory of block placement and wear leveling data is also kept in the cache while the drive is operating.

Energy storage: Another component in higher performing SSDs is a capacitor or some form of batteries. These are necessary to maintain data integrity such that the data in the cache can be flushed to the drive when power is dropped; some may even hold power long enough to maintain data in the cache until power is resumed.

The performance of the SSD can scale with the number of parallel NAND flash chips used in the device. A single NAND chip is relatively slow, due to narrow (8/16 bit) asynchronous IO interface, and additional high latency of basic IO operations (typical for SLC NAND - ~25 μs to fetch a 4K page from the array to the IO buffer on a read, ~250 μs to commit a 4K page from the IO buffer to the array on a write, ~2 ms to erase a 256 KB block). When multiple NAND devices operate in parallel inside an SSD, the bandwidth scales, and the high latencies can be hidden, as long as enough outstanding operations are pending and the load is evenly distributed between devices.

Micron/Intel SSD made faster flash drives by implementing data striping (similar to RAID0) and interleaving. This allowed creation of ultra-fast SSDs with 250 MB/s effective read/write, the maximum the SATA interface can manage.

In 2008-03-31, Fusion-IO announced ioDrive, the company's first product, would begin shipping on 2008-04-07. It uses a PCI Express slot to overcome the bandwidth limitation of SATA and SAS interfaces, and the 80GB unit has a read speed of over 800MB/s and write speed of 694MB/s.

SLC versus MLC

Lower priced drives usually use multi-level cell (MLC) flash memory, which is slower and less reliable than single-level cell (SLC) flash memory.

DRAM based drive

SSDs based on volatile memory such as DRAM are characterized by ultra fast data access, generally less than 0.01 milliseconds, and are used primarily to accelerate applications that would otherwise be held back by the latency of Flash SSDs or traditional HDDs. DRAM-based SSDs usually incorporate internal battery and backup storage systems to ensure data persistence while no power is being supplied to the drive from external sources. If power is lost, the battery provides power while all information is copied from random access memory (RAM) to back-up storage. When the power is restored, the information is copied back to the RAM from the back-up storage, and the SSD resumes normal operation. (Similar to the hibernate function used in modern operating systems.)

These types of SSD are usually fitted with the same type of DRAM modules used in regular PCs and servers, allowing them to be swapped out and replaced with larger modules.

A secondary computer with a fast network or (direct) Infiniband connection can be used as a RAM-based SSD.

Open casing of 2.5” traditional hard disk drive (left) and solid-state drive (center).

DRAM based solid-state drives are especially useful on computers that already have the maximum amount of supported RAM. For example, some computer systems built on the x86-32 architecture can effectively be extended beyond the 4 GB limit by putting the paging file or swap file on an SSD. Owing to the bandwidth bottleneck of the bus they connect to, DRAM SSDs cannot read and write data as fast as main RAM can, but they are far faster than any mechanical hard drive. Placing the swap/scratch files on a RAM SSD, as opposed to a traditional hard drive, therefore can increase performance significantly.

Commercialization

Cost and capacity

Until recently, flash based solid-state drives were too costly for widespread use in mobile computing.[citation needed] As flash manufacturers transition from NOR flash to single-level cell (SLC) NAND flash and most recently to multi-level cell (MLC) NAND flash to maximize silicon die usage and reduce associated costs, "solid-state disks" are now being more accurately renamed "solid-state drives" – they have no disks but function as drives – for mobile computing in the enterprise and consumer electronics space. This technological trend is accompanied by an annual 50% decline in raw flash material costs, while capacities continue to double at the same rate. As a result, flash-based solid-state drives are becoming increasingly popular in markets such as notebook PCs and sub-notebooks for enterprises, Ultra-Mobile PCs (UMPC), and Tablet PCs for the healthcare and consumer electronics sectors. Major PC companies have now started to offer such technology.

Availability

Solid-state drive (SSD) technology has been marketed to the military and niche industrial markets since the mid-1990s

CompactFlash card used as SSD

Along with the emerging enterprise market, SSDs have been appearing in ultra-mobile PCs and a few lightweight laptop systems, adding significantly to the price of the laptop, depending on the capacity, form factor and transfer speeds. As of 2008 some manufacturers have begun shipping affordable, fast, energy-efficient drives priced at $350 to computer manufacturers. For low-end applications, a USB flash drive may be obtained for $10 to $100 or so, depending on capacity, or a CompactFlash card may be paired with a CF-to-IDE or CF-to-SATA converter at a similar cost. Either of these requires that write-cycle endurance issues be managed, either by not storing frequently written files on the drive, or by using a Flash file system. Standard CompactFlash cards usually have write speeds of 7 to 15 megabytes per second while the more expensive upmarket cards claim speeds of up to 40 MB/s.

One of the first mainstream releases of SSD was the XO Laptop, built as part of the 'One Laptop Per Child' project. Mass production of these computers, built for children in developing countries, began in December 2007. These machines use 1024 MiB SLC NAND flash as primary storage which is considered more suitable for the harsher than normal conditions in which they are expected to be used. Dell began shipping ultra-portable laptops with SanDisk SSDs on April 26, 2007.[1] Asus released the Eee PC subnotebook on October 16 2007, and after a successful commercial start in 2007, expects to ship several million PCs in 2008, with 2, 4 or 8 gigabytes of flash memory.[41] On January 31, 2008, Apple Inc. released the MacBook Air, a thin laptop with optional 64 GB SSD. The Apple store cost was $999 more for this option, as compared to that of an 80 GB 4200 rpm Hard Disk Drive. Another option - Lenovo ThinkPad X300 with a 64Gbyte SSD - was announced by Lenovo in February 2008, and is, as of 2008, available to consumers in some countries. On August 26, 2008, Lenovo released ThinkPad X301 with 128GB SSD option which adds approximately $200 US.

The Mtron SSD

As of October 14, 2008, Apple's MacBook and MacBook Pro lines carry optional solid state hard drives at an additional cost. Dell began to offer optional 256 GB solid state drives on select notebook models in January 2009.

In late 2008, Sun released the Sun Storage 7000 Unified Storage Systems (codenamed Amber Road), which use both solid state drives and conventional hard drives to take advantage of the speed offered by SSDs and the economy and capacity offered by conventional hard disks.

Quality and performance

SSD is still currently a developing technology. A January 2009 review of the market by technology reviewer Tom's Hardware concluded that comparatively few of the tested devices showed acceptable I/O performance, with several disappointments, and that Intel (who make their own SSD chipset) still produce the best performing SSD drive as at this time, a view also echoed by Anandtech. In particular, operations that require many small writes, such as log files, are particularly badly affected on some devices, potentially causing the entire host system to freeze for periods of up to one second at a time.

According to Anandtech, this is due to controller chip design issues with a widely used set of components, and at least partly arises because most manufacturers are memory manufacturers only, rather than full microchip design and fabrication businesses - they often rebrand others' products,[47] inadvertantly replicating their problems. Of the other manufacturers in the market, Memoright, Mtron, OCZ, Samsung and Soliware were also named positively for at least some areas of testing.

The overall conclusion by Tom's Hardware however, was that "none of the [non-Intel] drives was really impressive. They all have significant weaknesses: usually either low I/O performance, poor write throughput or unacceptable power consumption".

Index

Serial ATA

The Serial ATA computer bus is a storage-interface for connecting host bus adapters (most commonly integrated into laptop computers and desktop motherboards) to mass storage devices (such as hard disk drives and optical drives).

Conceptually, SATA (Serial Advanced Technology Attachment) is a 'wire replacement' for the older AT Attachment standard (ATA). Serial ATA host-adapters and devices communicate via a high-speed serial cable.

SATA offers several compelling advantages over the older parallel ATA/"EIDE" interface: reduced cable-bulk and cost (7 pins vs 40 pins), faster and more efficient data transfer, and the ability to remove or add devices while operating (hot swapping).

As of 2009, SATA has all but replaced the legacy ATA (retroactively renamed Parallel ATA or PATA) in all shipping consumer PCs. PATA remains dominant in industrial and embedded applications dependent on CompactFlash storage though the new CFast storage standard will be based on SATA.

SATA specification bodies

There are at least four bodies with possible responsibility for providing SATA specifications: the trade organisation, SATA-IO; the INCITS T10 subcommittee (SCSI); a subgroup of T10 responsible for SAS; and the INCITS T13 subcommittee (ATA). This has caused confusion as the ATA/ATAPI-7 specification from T13 incorporated an early, incomplete SATA rev. 1 specification from SATA-IO. The remainder of this article will try to use the terminology and specifications of SATA-IO.

Advanced Host Controller Interface

As their de facto standard interface, SATA controllers use the Advanced Host Controller Interface (AHCI), which allows advanced features of SATA such as hot plug and native command queuing (NCQ). If AHCI is not enabled by the motherboard and chipset, SATA controllers typically operate in "IDE emulation" mode which does not allow features of devices to be accessed if the ATA/IDE standard does not support them. Windows device drivers that are labelled as SATA are usually running in IDE emulation mode unless they explicitly state that they are AHCI. While the drivers included with Windows XP do not support AHCI, AHCI has been implemented by proprietary device drivers. Windows Vista and Linux with kernel version 2.6.19 onward have native support for AHCI.

Features

The current SATA rev. 2.x specifications detail data transfer rates as high as 3.0 Gbit/s per device. SATA uses only 4 signal lines; cables are more compact and cheaper than PATA. SATA supports hot-swapping and NCQ. There is a special connector (eSATA) specified for external devices, and an optionally implemented provision for clips to hold internal connectors firmly in place. SATA drives may be plugged into SAS controllers and communicate on the same physical cable as native SAS disks, but SATA controllers cannot handle SAS disks.

Throughput

SATA 1.5 Gbit/s

First-generation SATA interfaces, also known as SATA 1.5 Gbit/s or unofficially as SATA 1, communicate at a rate of 1.5 Gbit/s. Taking 8b/10b encoding overhead into account, they have an actual uncoded transfer rate of 1.2 Gbit/s, or 1,200 Mbit/s. The theoretical burst throughput of SATA 1.5 Gbit/s is similar to that of PATA/133, but newer SATA devices offer enhancements such as NCQ which improve performance in a multitasking environment.

Today's mechanical hard disk drives can transfer data at up to 127 MB/s, which is within the capabilities of the older PATA/133 specification. However, high-performance flash drives can transfer data at up to 201 MB/s. SATA 1.5 Gbit/s does not provide sufficient throughput for these drives.

During the initial period after SATA 1.5 Gbit/s finalization, adapter and drive manufacturers used a "bridge chip" to convert existing PATA designs for use with the SATA interface. Bridged drives have a SATA connector, may include either or both kinds of power connectors, and generally perform identically to their PATA equivalents. Most lack support for some SATA-specific features such as NCQ. Bridged products gradually gave way to native SATA products.

SATA 3 Gbit/s

Soon after the introduction of SATA 1.5 Gbit/s, a number of shortcomings emerged. At the application level SATA could handle only one pending transaction at a time-like PATA. The SCSI interface has long been able to accept multiple outstanding requests and service them in the order which minimizes response time. This feature, native command queuing (NCQ), was adopted as an optional supported feature for SATA 1.5 Gbit/s and SATA 3 Gbit/s devices.

First-generation SATA devices operated at best a little faster than parallel ATA/133 devices. Subsequently, a 3 Gbit/s signaling rate was added to the physical layer (PHY layer), effectively doubling maximum data throughput from 150 MB/s to 300 MB/s.

For mechanical hard drives, SATA 3 Gbit/s transfer rate is expected to satisfy drive throughput requirements for some time, as the fastest mechanical drives barely saturate a SATA 1.5 Gbit/s link. A SATA data cable rated for 1.5 Gbit/s will handle current mechanical drives without any loss of sustained and burst data transfer performance. However, high-performance flash drives are approaching SATA 3 Gbit/s transfer rate.

Given the importance of backward compatibility between SATA 1.5 Gbit/s controllers and SATA 3 Gbit/s devices, SATA 3 Gbit/s autonegotiation sequence is designed to fall back to SATA 1.5 Gbit/s speed when in communication with such devices. In practice, some older SATA controllers do not properly implement SATA speed negotiation. Affected systems require the user to set the SATA 3 Gbit/s peripherals to 1.5 Gbit/s mode, generally through the use of a jumper, however some drives lack this jumper. Chipsets known to have this fault include the VIA VT8237 and VT8237R southbridges, and the VIA VT6420 and VT6421L standalone SATA controllers. SiS's 760 and 964 chipsets also initially exhibited this problem, though it can be rectified with an updated SATA controller ROM.

This table shows the real speed of SATA 1.5 Gbit/s and SATA 3 Gbit/s; note the bottom row shows megabytes per second (MB/s, not Mbit/s):

	SATA 1.5 Gbit/s	**SATA 3 Gbit/s**
Frequency	1.5 GHz	3 GHz
Bits/clock	1	1
8b/10b encoding	80%	80%
bits/Byte	8	8
Real speed	150 MB/s	300 MB/s

SATA II misnomer

Popular usage may refer to the SATA 3 Gbit/s specification as "Serial ATA II" ("SATA II" or "SATA2"), contrary to the wishes of the Serial ATA International Organization (SATA-IO) which defines the standard. SATA II was originally the name of a committee defining updated SATA standards, of which the 3 Gbit/s standard was just one. However since it was among the most prominent features defined by the former SATA II committee, and, more critically, the term "II" is commonly used for successors, the name SATA II became synonymous with the 3 Gbit/s standard, so the group has since changed names to the Serial ATA International Organization, or SATA-IO, to avoid further confusion.

SATA 6 Gbit/s

Serial ATA International Organization presented the draft specification of SATA 6 Gbit/s physical layer in July 2008 and ratified its physical layer specification on August 18, 2008. The full 3.0 standard is expected to be available in 2009. While even the fastest conventional hard disk drives can barely saturate the original SATA 1.5 Gbit/s bandwidth, Intel's Solid State Disk drives are close to saturating the SATA 3 Gbit/s limit at 250 MB/s net read speed, and other new drives including Super Talent, Memoright and Samsung are close to that as well. Ten channels of fast flash can actually reach well over 500 MB/s with new ONFI drives, so a move from SATA 3 Gbit/s to SATA 6 Gbit/s would benefit the flash read speeds. As for the standard hard disks, the reads from their built-in DRAM cache will end up faster across the new interface.

The new specification will include a handful of extensions to the command set, especially in the area of data and command queuing. The enhancements are generally aimed at improving quality of service for video streaming and high priority interrupts. In addition, the standard will continue to support distances up to a meter. The new speeds may require higher power consumption for supporting chips, factors that new process technologies and power management techniques are expected to mitigate. The new specification can use existing SATA cables and connectors, although some OEMs are expected to upgrade host connectors for the higher speeds. Also, the new standard is backwards compatible with SATA 3 Gbit/s.

In order to avoid parallels to the common "SATA II" misnomer, the SATA-IO has compiled a set of marketing guidelines for the new specification. The specification should be called "Serial ATA International Organization: Serial ATA Revision 3.0", and the technology itself is to be referred to as "SATA 6 Gbit/s". A product using this standard should be called the "SATA 6 Gbit/s [product name]". The terms "SATA III" or "SATA 3.0", which are considered to cause confusion among consumers, should not be used.

Cables and connectors

Connectors and cables present the most visible differences between SATA and Parallel ATA drives. Unlike PATA, the same connectors are used on 3.5 in (89 mm) SATA hard disks for desktop and server computers and 2.5 in (64 mm) disks for portable or small computers; this allows 2.5 in drives to be used in desktop computers without the need for wiring adapters (a mounting adaptor is still likely to be needed to securely mount the drive).

SATA connectors are not as robust as PATA connectors. For example, the motherboard connector on SATA includes a plastic tab which can be broken if the connector is bent. This might happen if the cable is pulled to one side. Because such a broken connector is on the motherboard rather than the cable, it is not easy to replace.

Data

The SATA standard defines a data cable with seven conductors (3 grounds and 4 active data lines in two pairs) and 8 mm wide wafer connectors on each end. SATA cables can have lengths up to 1 metre (3.3 ft), and connect one motherboard socket to one hard drive. PATA ribbon cables, in comparison, connect one motherboard socket to up to two hard drives, carry either 40 or 80 wires, and are limited to 45 centimetres (18 in) in length by the PATA specification (however, cables up to 90 centimetres (35 in) are readily available). Thus, SATA connectors and cables are easier to fit in closed spaces and reduce obstructions to air cooling. They are more susceptible to accidental unplugging and breakage than PATA, but cables can be purchased that have a 'locking' feature, whereby a small (usually metal) spring holds the plug in the socket.

One of the problems associated with the transmission of data at high speed over electrical connections is loosely described as 'noise'. Despite attempts to avoid it, some electrical coupling will exist both between data circuits and between them and other circuits. As a result, the data circuits can both affect other circuits, whether they are within the same piece of equipment or not, and can be affected by them. Designers use a number of techniques to reduce the undesirable effects of such unintentional coupling. One such technique used in SATA links is differential signalling. This is an enhancement over PATA, which uses single-ended signaling. Twisted pair cabling also gives superior performance in this regard.

Data:

Pin #	Function
1	Ground
2	A+ (Transmit)
3	A− (Transmit)
4	Ground
5	B− (Receive)
6	B+ (Receive)
7	Ground
8	coding notch

A 7-pin Serial ATA right-angle data cable.

Index