Fault-Resilient Computer Platforms

By CommWeb.com
Oct 3, 2000 (3:17 PM)
URL:

Both fault-resilient and fault-tolerant computer systems have "high availability" — they are more reliable than conventional PCs, allowing them to be used in "mission critical" systems (characterized by such things ashigh port count, multiple applications, use by service providers and Central Offices) or applications where revenue is generated on a minute-by-minute basis, such as prepaid calling card systems and enhanced services systems. A fault-resilient system will remain "up" about 99.99% of the time, or will suffer downtime for only about 52 minutes a year. A fault-tolerant system will be up "five nines" or 99.999% of the time, which amounts to about five minutes per year of downtime.

Both fault-resilient and fault-tolerant computers have redundant (and in many cases hot-swappable) components. A true fault-tolerant system will also have two CPUs, both of which work on the same application processes, but with some kind of polling process taking place to detect CPU failure.

If a CPU fails, then an "automatic failover" process switches control to the healthy CPU.

A fault-resilient computer has but a single CPU, which makes it a lot less expensive than a true fault-tolerant system. However, since CPUs only tend to fail if the fans fail, and if fans are monitored by an alarming board, then a fault-resilient PC can give you nearly the same reliability as a similarly sized fault-tolerant system; it'll just cost less.

Since fault-resilient computers closely resembled fault-tolerant systems but had no single name of their own, vendors were calling their fault resilient machines "industrial", "fault tolerant", "ruggedized", "heavy-duty", "highly available" and other obtuse terms. Richard "Zippy" Grigonis, Chief Technical Editor of Computer Telephony Magazine, settled on the term "fault resilient computer" and popularized it from 1995 onwards.

What to Look For:

These computers should have the following characteristics:

Passive backplane technology.

Although there are such things as "industrial" motherboards, most fault resilient computers need more slots than afforded by a conventional motherboard. Computer telephony applications use many add-in cards for voice, fax and digital switching. Passive backplanes provide up to about 25 expansion slots and are easier to upgrade and

service than motherboard-based systems.

In such systems the CPU, its memory and various I/O ports are situated on a plug-in card too, called a "single board computer" or SBC. The SBC can easily be replaced or upgraded — you just buy and plug in a new card.

The motherboard itself is replaced with a passive backplane that has nothing on it other than connectors (this is why this technology is sometimes referred to as "slot cards"). The chance of a passive backplane failing is quite low.

Passive backplanes can be "segmented", so that, for example, you can have a fault tolerant system where one SBC or set of telephony resource cards can "fail over" to another set should trouble rise. With current technology you can have up to four segments in one PC enclosure.

Most backplanes and motherboards are based on the Peripheral Component Interconnect (PCI) bus, but now the CompactPCI bus is starting to appear, which is electrically similar to a desktop PCI bus computer, but has card connectors that are far more rugged and reliable, and which also supports "hot swappable" cards.

Redundancy.

What makes fault-resilient systems genuinely fault resilient is component redundancy. For example, redundant disk drive I/O channels, fans, host controllers, and redundant power supplies that can be swapped out while the system is running ("hot swappable"). Redundant hot-swappable power supplies can also come in load sharing, (also known as "load balancing" or "current sharing") and N+1 (need+1) configurations.

In the case of two "load sharing" power supplies, each can be supplying 50% of the total system power plus or minus 20%. So you might have two 500 watt power supplies running in a system designed to consume only 500 watts. Each power supply is therefore subject to only a 50% load, so both hot-swappable power supplies run at cooler temperatures and their life is extended. When one of them finally does fail, the other power supply takes up the full load while you pull out the failed power supply and replace it with a new one.

N+1 architecture means "more than two." In an N+1 configuration load sharing still exists but the load is now balanced among more than two power modules.

RAID.

Redundant data storage comes in the form of a multiple disk drive RAID subsystem. RAID means Redundant Array of Independent Disks. It's a disk subsystem architecture that in most cases writes (or "stripes") data across multiple hard disks to achieve fault tolerance, if not continuous availability. One drive in the drive array is often referred to as the "parity" drive, which contains data that can be used to recreate data on any one of the other drives, assuming that the other drives remain operational.

For example, you might have a five-drive array where one drive is designated as the parity drive. If a data drive fails and is then replaced, the drive array controller will

rebuild the data on that drive using the parity drive and the other functioning drives.

There are various types of RAIDs, each indicated by a number or "level":

RAID-0

Uses disk striping without parity information. RAID-0 is the fastest and most efficient array type but offers no data protection and is actually subject to the opposite of fault tolerance, since the failure of any disk will bring down the system. It's the one RAID level that's not really "RAID" at all!

RAID-1

This is the array of choice for performance-critical, fault-tolerant environments. This, the most simple, secure and reliable type of RAID is also called "mirroring" or "dual copy" or "shadowing", where two hard drives are connected to the same disk controller, separate controllers, or the disk drive control is provided in software. Whenever data is written to a file, it's duplicated and written simultaneously to both disks. Thus, the data on one disk is a copy or "mirror image" of the other.

One benefit of a RAID-1 system is that if a drive goes down, your system doesn't waste time reconstructing the data. Unfortunately, mirroring is the most expensive alternative in terms of overhead (highest cost per byte), since half of the subsystem array is redundant. You always need twice as much storage as you would with a single disk drive.

RAID-2

A disk subsystem architecture that uses disk striping across multiple disks at the bit level with parity. In RAID-2, which includes error detection and correction, an array of four disks requires three parity disks of equal size. This is seldom used since error correction codes are embedded in sectors of almost all disk drives anyway. Still, some implementations exist for supercomputer storage.

RAID-3

Uses disk striping at the byte level with only one disk per array dedicated to parity information. This can be used in single-user environments and performs best when accessing long sequential records. However, RAID-3 does not allow multiple I/O operations to be overlapped and needs synchronized drives in order to prevent performance degradation involving short records.

RAID-4

Same as RAID 3 but stripes data in larger chunks (whole sectors or records). This allows multiple reads to be overlapped but not multiple writes. Like RAID-3, a dedicated disk stores parity information.

RAID-5

Same as RAID 4 but data is stripped in sector-sized blocks and parity data is also striped

across the disks interleaved with the data. It supports both overlapping reads and writes, but write performance is slightly degraded because of the need to update parity data.

RAID-6

Same as RAID 5, plus additional striping so two disks can fail simultaneously, redundant controllers, fans, power supplies, etc. An array providing striping of data across multiple drives and two parity sets for increased fault tolerance. Highly reliable but suffers from slow performance.

RAID-7

RAID 7 is not yet an industry-standard term, but rather a product name for an RAID-like approach for multiple-host, UNIX-based environments running on various hardware platforms, including those from DEC, Silicon Graphics, Sun Microsystems, Hewlett-Packard, IBM and Sequent.

Storage Area Networks (SNAs) are independent networked-attached (or fiber channel attached) clusters of RAID-systems used for the centralized storage management of huge "server farms". Since they are independent of any particular server, reminiscent of how networked peripherals like printers are independent of any particular PC, one can perform a data back up of a SNA with special high-bandwidth hardware that doesn't interfere with the normal "conventional" network.

Pressurized forced-air cooling.

Since fault resilient computers can hold more computer telephony resource cards than conventional computers, along with disk drives (if the disk drive array is internal), cooling immediately becomes a problem.

A good fault resilient PC has several fans in a plenum, circulating cooling air throughout the card cage then vented out the back of the chassis. In the case of compactPCI designs, the system is vertical and a "chimney effect" carries the heat up and out of the unit.

Whatever the design, the fault resilient PC should have a removable, washable filter so the circulating air is free of corrosive dirt.

System Monitoring and Alarming.

Even if a fault resilient computer can recover quickly from a problem, such as a blown power supply or failed disk drive, the system must somehow alert technicians that something went wrong and components must be replaced.

We once encountered a system that was running with a RAID subsystem where one drive had failed months before but there was no visual nor audio indication; no one was the wiser!

Incidents such as these indicate that a computer's monitoring / alarm subsystem is crucial to alerting maintenance technicians to system problems, enabling them to diagnose problems immediately without waiting for the system to continue "using up" its

redundant components and failing completely. Such a subsystem should be able to perform related activities such as resetting unattended systems and sending distress signals over a network, modem or pager.

Alarming boards can be PCI or older ISA cards, standalone cards or rack modules with their own power connector, or built-in features on CPU cards. The best usually take the form of an intelligent card independent of the backplane with its own microprocessor and battery backup, along with user programmability, automatic switchover to a stand-by system, multichassis support, and an LED display integrated in the chassis or module. Some have the display housed in a separate package that fits into a 5.25" half-height drive bay.

NEBS.

It stands for (Network Equipment Building System). Some companies — large carriers — feel that their applications must be run on the most fault tolerant computers available — computers that can survive a massive earthquake, the Chicago fire, a lightning strike, or other disastrous phenomena, whether they be natural or man-made. NEBS certified equipment is designed to run applications under the most physically extreme conditions.

NEBS defines a minimum generic set of spatial and environmental / safety requirements developed by Bellcore (now called Telcordia) for use in Central Offices and other telephone buildings of the Bellcore Client Companies (BCCs). NEBS compliance is recommended only and is not required by regional BCCs. The decision to actually purchase NEBS certified equipment is totally at the discretion of each BCC.

Many companies claim that they manufacture computers that are NEBS-compliant; but how many of these machines are actually NEBS-certified? NEBS certification for a PC costs over $100,000 and takes months. Among other things, the PC chassis undergoes a "shake and bake" process that would wreck stout mechanical devices, let along an electronic computer!

Benefits:

Fault-resilient and fault-tolerant computers can keep running if a component runs into trouble and can be serviced rather quickly, in most cases without even having to bring down the system. Since they tend to be 19-inch wide rackmounts, many can be stacked upon each other to build systems having tens of thousands of ports.

Web servers, enterprise communication servers, 911 systems and many other business critical systems benefit from equipment that can "take a licking" and yet continue to run.