CIM·DATA LTD.

Executive Summary

A brokerage's main server is usually their single biggest technology investment. A fast, responsive server can improve user productivity and customer service.

Newer servers, especially virtualization servers, are being configured in new ways.

This information sheet explains, in fairly basic language, some of the design decisions that can remarkably influence the performance yielded when investing in new server technologies.

Choosing a Fast, Reliable Server

5155 Spectrum Way, Unit #18
Mississauga, Ontario L4W 5A1
www.cim-data.com
905 624-8300
800 668-6041

Server Components

There are nine major components to consider when specifying a server. Each affects perceived server speed and total server reliability.

No matter where users are working, whether workstations on a LAN, or thin-clients in a branch office, when they hit a key, or move their mouse, you want them to see instant results. Considering that a Remote Desktop Server or Virtual Desktop Server can have many users all vying for server resources, if even one component is mismatched, user performance can be reduced.

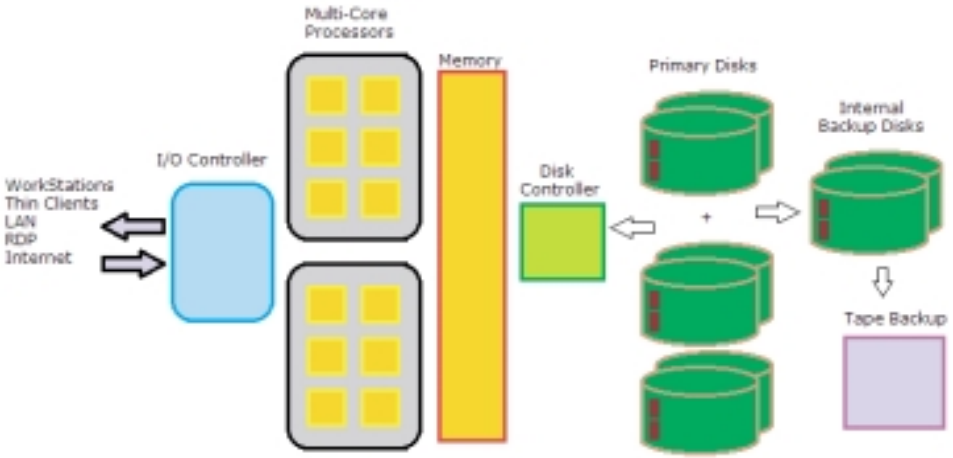

Figure 1 - Server Components

I/O Controller

The heart of any server is its motherboard and the heart of a mother board is its I/O Controller. The I/O Controller is responsible for getting data "In" and "Out" of the motherboard and it's architecture determines the type of processors, memory, and other controllers that can be used.

Processor and Chassis

A Processor's performance depends upon it's design speed and ability to dissipate heat. During moments of high utilization, they get hot, and when they get too hot, they slow themselves down to prevent damage. A server chassis with good air flow can dissipate heat, thereby allowing the processors to run fast for longer periods of time.

Of course, there are many other, new design factors that go into each generation of processor to allow it to do more while generating less heat. For example, the cores in today's multi-core processors are actually what we would have considered to be individual Pentium or Xeon CPUs of a few years ago. So a dual Xeon machine with two six core processors can actually devote the equivalent of twelve separate CPUs to twelve users at the same moment.

Memory

To keep even six cores working at their capabilities of perhaps fifteen to thirty times higher speed than processors of only a few years ago, a lot of data has to be fed to them. There is a wide choice in types, sizes, and configurations of memory to choose from. When the right, optimized choices are made, memory performance can reach levels up to 50% higher than un-optimized configurations, and in some cases, a higher performance gain can be realized by investing in better memory than faster processors.

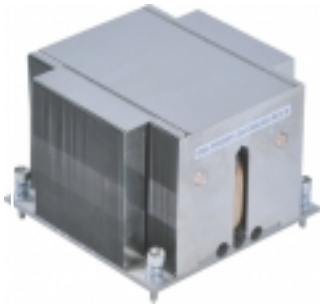

Figure 2 - Xeon Multi-Core Heat Sink (note cooling fins)

There is also a trend to larger software programs and bigger data files so more memory is often required than just a few years ago. 64bit servers automatically require from 50% to 100% more memory than 32bit systems.

Adding extra, "unused" memory also plays a particularly advantageous role by supplying disk caching capacity. When files are loaded into memory to be used by running programs, they are first loaded into a disk cache - a portion of extra, unused memory. If a program needs to access data from the disk, the operating

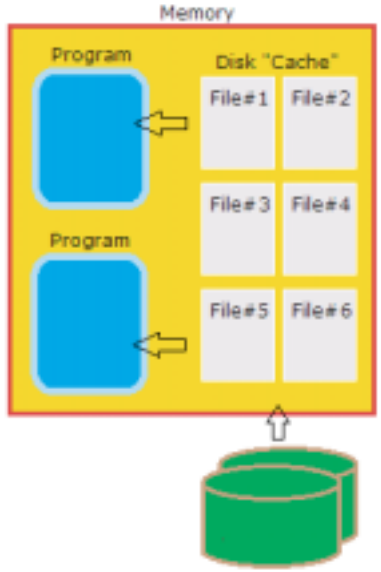

Figure 3 - Extra memory reduces repeated disk access by "caching" data

system first checks to see if it still has the file (or the portion that we want) in memory. If it does, it doesn't have to wait and read it in again from the disk. By purposely adding more, "extra" memory, the speed of any system can be increased significantly.

Disk Controller

Of course, to feed the memory that's feeding the processors at high speed, the data has to be loaded from disk storage quickly too. Disk Controllers are responsible for reading and writing disk data, and high-bandwidth controllers with onboard, battery backed up memory buffers have a significant speed advantage. Both a cheap and an expensive controller can read and write equally well, but a high performance controller can pump way more data per second.

Disk Storage

There are currently three types of disk technologies available. Low cost SATA drives compete in the consumer market and offer reasonable speed and really high capacity. In fact a single SATA drive can have much more capacity than the average business server would ever need. Problem is, a single SATA drive is basically just quick enough for a single workstation, and as a consumer product, is designed for low price, not high reliability.

Figure 4 - A 16 hard drive, 2U, rack mount server

Enterprise SAS drives, on the other hand, compete in a quality market. They are often faster, more costly, and may have lower capacity than their SATA counterparts, but most importantly for servers storing critical business data, have lower error rates, and lower failure rates.

And then there are Solid State drives that are basically really big, fast USB drives in a hard drive case. These devices can be incredibly fast, but are also the most expensive. They suffer from high failure rates although that is improving. Their most serious drawback though, is that they simply "wear out". And when they do, they go bad, badly.

RAID Arrays

There is a trick to getting high performance and reliability out of hard drives, and that is to use several of them. By spreading the data across them in parallel, the combined speed of several disks working in parallel for a single moment can retrieve a file faster than one drive reading the file serially.

Also, by storing the same data more than once across multiple disks, data is stored redundantly, so that if one drive fails, the data is still available on another drive. By installing spare drives (called "hot spares"), the Disk Controller can disable a failed drive and automatically replace it with a hot spare, then automatically rebuild the data from the redundant copies.

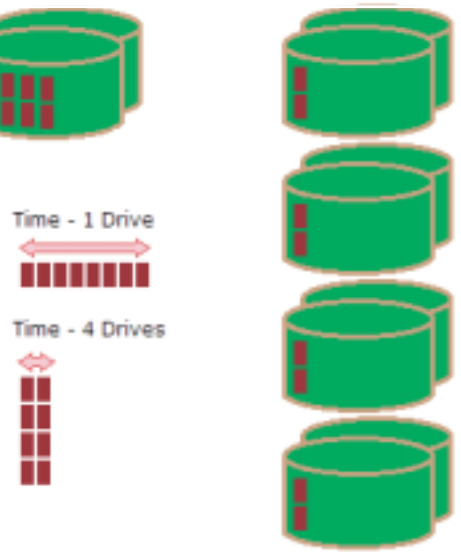

Figure 5 - Parallel file I/O with RAID

Consider a file such as an email or a program you want to run. All data is stored on disks in files which are themselves stored as blocks of data. To load the file, all the file's blocks have to be read from disk. Now consider storing the blocks of a file on a single disk. A disk drive can only read one block at a time, in a serial fashion. Now consider what would happen if the blocks for that same file could all be read at nearly the same moment simply be spreading them across multiple disks. Suddenly, you now have the speed of several disks working for you.

That's what server RAID Arrays do - store data redundantly and in parallel. They take inherently unreliable disks, improve their overall reliability substantially, and improve their performance. There are several types of RAID, and some are faster, and some are more reliable than others.

With "mirrored" RAID, files are actually written in full twice. Parallelism can be improved even more with mirrors since two separate files (requested by two different users) can be read in the same

moment - one from one copy, and one from the other copy.

Primary Data vs. Backup Drives

While RAID can significantly reduce the likelihood of lost data, it is not a guarantee, and cannot protect against all causes of file damage, especially inadvertent deletions by users. For those reasons, modern servers more and more are carrying additional, internal backup drives. Backup drives are typically higher capacity but slower drives compared to the main data drives. Some people may choose cheaper, single SATA drives for backup, but in the event, of a meltdown of the primary data drives, it is worth the investment in having reliable, redundant, SAS backup drives.

Disaster Protection - Tape, Cloud, and Disk Backup

Redundant data drives, and internal backup drives cannot protect against theft of the server, or damage by fire, or earthquake damage to disks. Not having protection against these forms of disaster will nearly guarantee the failure of your business in the event a calamity occurs. For this reason, daily backups of your server to a tape media or removable hard drive that can be removed offsite; a backup to a remote computer (for example in a branch office); or backing up to an Internet "cloud" backup service, will provide a replacement for your server's data in worst case scenarios.

Power

Servers run on electricity which must be transformed and conditioned for the internal components to use it. A typical option for servers is to contain two separate Power Supply Units so that if one fails, the other can carry on. Each power supply has its own power cord. Each power cord is plugged into a separate UPS. A Battery backed up UPS helps prevent damage to the server from electrical disturbances from the power lines (Never plug a server directly into the wall).

Summary

It should now be apparent that servers can be built with varying degrees of performance and very different cost and reliability levels.

Figure 6 - 3U rear view - dual Power Supply Units + dual cooling fans

Choosing the best design for your server is an art based on a good fundamental understanding of how servers work. We have only touched on the basics here, but important ones.

Making appropriate server choices ultimately affects user productivity and the protection of your data.