Jonoke Software Dev. Inc.

Database Server Hardware

3450 - 78 Ave Edmonton, AB T6B2X9 T (800)254-0739 F (780) 448-3741 info@jonoke.com www.jonoke.com

Table of Contents

Database Server Hardware

Introduction

This paper is not intended to be read by the general purchaser of JonokeMed™. Most people that purchase JonokeMed™ get Jonoke to install the servers that they need. They leave this critical decision to us to advise them based on our experience, and that our reputation is on the line. There are always those people that like to know more. These people want to know more detail and be more involved in the decision process (I'm one of those). There there are people that like to compare prices, and in some ways this is who this document is aimed at. Often these people don't know about a lot of the considerations that go into the decision on which server. These people will often purchase a much less expensive server and then complain about performance, and reliability.

I recently was told by a person that they could purchase a certain brand and model for $700 and it was 1/2 price of another brand and model of a computer as a workstation. I didn't believe that there was this much difference in price. I did some research on the two computers. Though the computers both had the latest processors, the size of the L3 cache was my first clue of their differences. The one used a near top end processor, and the other used a near bottom end processor. Then there were differences in the bus speed, and types of hard drives used. The difference in speed between the two would be very noticeable even though their CPU clock speed was essentially the same. The inexpensive computer was the low end of the one manufacture, while the other was at the upper end of the other manufacture. The sales person didn't differentiate the two pieces of hardware - instead comparing them as though they were the same - just claiming that the one brand is always more expensive. The sales person also made the statement that the less expensive computer would do everything this person needed, and would work for the software they were planning to use it for. It makes me wonder why would this sales person make these statements, and not inform this shopper of the differences. Could it be that the one manufacture often gives sales people bonuses to the sales people for their total sales? I don't know the reason, but this kind of misinformation leads people to have substandard systems. With a workstation this will certainly effect them, but with a Data Base server this will dramatically effect every user on the system.

It was that experience that lead me to write this paper for people looking to purchase computes, but more specifically database servers.

1.0 Database Servers - Not a File Server

When you look at running an application like JonokeMed™ you are looking to have something that is going to run your whole business. This means that almost everything you do will be effected by JonokeMed™ and underneath that the Database server hardware you purchase. The last you want is for the system to go down because the hardware has failed, or to be waiting on the system on every step of your work. It is for this reason that paying attention to the server hardware that you run JonokeMed™ on is very important.

Yes, JonokeMed™ is unique, but not so unique in the database world. A lot of the concepts presented here apply to the purchase of a server for all databases. DO NOT let the hardware sales person sell you something that is a file

server system - for a database server. These are vastly different needs, and therefore have vastly different needs in the hardware.

Unlike what the sales person at the electronics store will tell you - there is a difference with computer hardware. Looking at the specifications on the outside of the box is not good enough. Just because the computer has a large drive, the latest processor, lots of RAM, does not mean it is a good as you can get. There are far more things to look at when looking for hardware for your server. Frankly - there is a reason that the computer is several hundreds or even thousands of dollars less than Jonoke quoted you for the server.

When you purchase a vehicle, your intended purpose will direct you to the type of vehicle you need. Do you need a small car to park easily in the core of the city, wanting good gas milage, and excellent pollution prevention features? Or do you need a 4 wheel drive 1 ton truck to haul trailers, and all kinds of materials from around the farm in all seasons. Obviously one car does not fit all situations. Just like when buying a vehicle, the intended purpose of your computer will help direct you to the correct hardware. The electronics store sales person that tells you the computer he is recommending will be great as a JonokeMed™ server - well you should run away from him. How does he know? Does he know JonokeMed™ or what it even is? Does he know what it does, and what a medical office needs? Very like the answer to all the above is - no.

This document covers some of the things that we look at when we purchase a server to run JonokeMed™. Our reputation is on the line when we install JonokeMed™. Software with JonokeMed™ powered software is our only business. We want your system to be stable, reliable, and fast. We also guarantee in writing that your server is not going to run out of room to store your data in one year. When the server is slow, or unreliable, or looses data, or goes down, people do not say the hardware is bad. What they say is JonokeMed™ is slow, unreliable, looses data, or is down. Therefore it is to our advantage to ensure that the JonokeMed™ server hardware fits your needs the best. This document along with hardware recommendations are to help guide you in understanding your hardware purchase. This document is mainly about servers, but we will touch base on workstations as well.

Lets start with some base set of assumptions when looking for the data base server computer. We are looking for a server that will provide the following:

- reliable storage of the data (not loosing data)

- permit us to back the data up to take off site

- It will run trouble free (except for planned maintenance) for several years (3 to 5 years)

- It will have enough room for the data you will need to store and backup for its lifetime.

- It will be fast enough so that everyone using the system are not constantly waiting on the system to give them information, save information, etc. Time is money - you don't want to be wasting it.

- It needs to run the JonokeMed™ system.

So with these assumptions lets dive right in to understanding what database server you should be looking at.

1.1 The Components to Consider

There are many components in a server that must fit together to create a great server for you. This section introduces most of these components that you should be considering. In the discussion here I am presenting more detail than people get from looking at the box specification, or they are likely to get from the person selling them the computer. Of course I could always delve deeper into the topic and provide more detail as these details do effect performance. The goal of this paper though is to get people thinking more about what they buy, and to not get talked into buying something that does not fit the application they are using the system for. If you want to know more, there is lots of information in text books (electrical engineering), and on the web i.e. Wikipedia).

Components to Consider

Component	Description
CPU - Central Processing Unit	The CPU is the 'brain' of the computer. All the data must go through this processor and be handled in some fashion. Speed of the processor is just one of the measures of this component to consider.
Level 3 Cache	This is memory that is directly in or attached to the CPU. The faster data and instructions can be fed to the CPU the faster the application will run.
RAM (Random Access Memory)	This is the memory that is gone when you shut your computer off. This is its disadvantage. Its advantage is that is is thousands of times faster than a hard drive.
Disks / SDD	This is memory that stores your data, Operating System, and programs when the computer is turned off. They are far larger than RAM.
Attached Storage (SAN & NAS)	This really is the same as the disks (and is made up of them), but are not directly connected to the server computer.
Disk / RAID Controller	The computer talks to this controller and the controller send the information to the disk drives to write the data, or retrieve the data.
RAID	This is a configuration of Disks / SSD that give more speed and/or backup capability while running the system live.
BUS Speed	This is the speed that data moves between the various components of the computer.
OS	Operating System. This is not hardware, but it is the first step above the hardware and firmware. It will critically effect how fast your server is.

1.2 Specifications don't tell the whole story - Critical Tested Package

As in everything else we buy, there is difference in quality. Generally the more expensive the higher the quality. This is no different in computers. If you look at the major computer manufactures like DELL and HP, they have different lines of computers. The different lines come at different prices, but if you look at the general specifications they look the same - they are not. The software used to control the various parts of the computer (i.e. BIOS) can be very different between models. The actual way items are designed for heat, and for MTF (mean time between failure) can be all very different. There are computers that are sold for home (occasional) use. These are not engineered to work in the same environment as the business or enterprise versions. As well, the warranty and support that comes with them is likely very different. All of these things go together to set the price of the hardware.

The specifications on the box (just like with food packages), do not tell the whole story. The specifications describe the computer in general. A part that is never described by any vendor though is how all the parts work together. Buying a bunch of high performance parts and building a server far from guarantees a fast and reliable server.

Early in Jonoke's move to expand we were convinced to accept that a local supplier that built computers would be supplying all the networking, workstations, printers, and servers. We really are not in the hardware business, and having a local person look after all of this was good from our and the client's point of view.

We looked over the specifications. They were using good parts from reputable companies. We arrived and installed JonokeMed™ and trained the client. All was well... for a few weeks. The server was fast, the workstations were fast, the network fast. Then the problems started to be exposed. Software that was working on thousands of computers at over 100 sites was suddenly having dramatic speed problems. We were confident that this was not in the software, but I like to be certain of the facts. We spent many hours looking at the slowdowns and could not find a pattern - thus not a consistent place in our code causing the slow down. Then we started to look at the data file. Wow - it was corrupted. We would repair and check again. The corruption would be cleaned up, but within a couple of days the same thing was happening. This process went on for weeks. We were being beat up by the client and spending huge amounts of senior staff time examining the problem. We even flew to the site to grab their data file and bring it back. This was before the internet speeds permitted copying large files in reasonable time.

We repaired the file on our server and the file cleaned up wonderfully and our torturing of the files never made it get unhealthy. We were more and more positive that it was the server. As usual we had a hard time to convince them that the problem was the server. Finally the hardware vendor replaced the server with a temporary server, and did various tests with the server. Ahhh... the RAID controller was corrupting the data. The version of the RAID controller that he used did not work properly with the version of the motherboard he had installed. Both were good brands, both were good quality, but because they are sold as parts for companies like this to use to build computers, they are not tested for thousands of hours together before they are released. The company replaced the RAID controller with a different version and the problems went away!

This caused everyone involved a great amount of grief. The client suffered from slow performance, and down time. The hardware vendor spent a lot of time researching their end, and in the end replacing a piece of hardware (the least expensive part). Jonoke spent hundreds of hours, and received lots of angry words from the client about this issue and likely suffered lots of angry statements to non-clients - thus loosing lots of potential sales.

The moral of this story (that I have learned once again since this event), is that having all the components tested together is critical to ensure a solid product. I am talking about serious thousands of hours of testing that validates the reads and writes, temperature control etc. There is NO local hardware shop that builds computers that can afford to run their hardware through these tests. Therefore I will not come anywhere near server hardware that it not built by a tier one supplier.

It is simply not worth the reputation destruction that we will receive, the loss of hundreds of hours of proving where the problem is - because we are always blamed.

2.0 CPUs

The CPU is the 'brain' of the computer. The Central Processing Unit touches all the data and processing. For this reason the performance of the CPU is critical to a fast Server (or workstation). CPUs are far more complex than they used to be. CPUs now come with at least 2 cores (processors), and can come with as many as eight. With servers you can then get multiple CPUs. Then there is processor speed, turbo speed, parallel processing, cache for each processor, included instruction set, 32 or 64 bit processor, graphics processor, just to mention a few items that will dramatically effect performance of anything that uses this CPU (or group of CPUs).

Item	Description
CPU	This is more complex than it used to be. Now it can include components that only a decade ago were a totally different component in the computer. Each brand and model of CPU will include more or less components.
Cores	The core is really what used to be called the CPU. The core is the processor that is the component that reads and executes program instructions and acts upon the data. Today a CPU can have multiple Cores. Most CPUs come with a minimum of 2 cores. This the major suppliers of computers (desktop / server / laptop) CPUs will also have 2, 4, 6, or even 8 cores per CPU. Having multiple cores on the same CPU permit processing to occur faster. A big reason for this is that the data does not need to travel as far (distance is a factor now). There are other advantages and some disadvantages. Overall though having multiple cores is an advantage.

Item	Description
Speed	The speed of the CPU (and other components) are measured in the number of cycles per second. The CPU speed is not a direct indicator of how many instructions it processes. For example some instructions many take 5 or more cycles to execute, while multiple instructions can be executed within one cycle in other instances. The make/model of CPU has an effect on this. The CPU speed within a family of processors is a way of comparing their performance. Again - this is just an indicator. Therefore when comparing computer don't just rely upon the speed of the processor as a statement of speed. Even within the same family there are different configuration that effect performance. For example there are many configurations of the Intel i7 processor that depending on the configuration a slower CPU can actually perform more work than a faster CPU.
L3 cache	With a computer the faster that the required program instructions, or data can get to the CPU/Core the faster the computer will get your work done. L3 cache is memory that is on the CPU. It is faster thus more expensive. It is also closer to the Cores thus the data gets to the Core faster than from main memory. The CPU determines (through software) which information to store in the L3 cache. The goal is to have the used information in the L3 cache all the time. Therefore the more L3 cache a CPU has, and the faster the speed of the bus the better performance the computer will have. There are L1 and L2 caches too that can be considered in this equation.
Turbo Boosting	Turbo charging a CPU has been around for many years. The speed of a CPU is not determined until after it is manufactured. They test the CPU to determine what speed they want to say that CPU should run at. Therefore there are ways to make a CPU run faster. It is not that easy though. There are problems with doing this than can cause the system to be slower, or even fail. Modern CPUs will come from the manufacture (i.e. Intel) with engineered ways to have the CPU run faster at specific times. They have engineered into the chip ways for it to know when it can run faster, and when not to. Therefore building in turbo boosting of the speed, but putting in safe guards to prevent failures. I much prefer the chip to do this so it is engineered this way. In the CPUs available today this can be very beneficial.
Simultaneous multithreading	This is a architecture that lets each core have multiple threads running at the same time within the core. This lets the processor make more efficient use of the resources to process more in less time. The ability to do this is greatly effected by the technology in the CPU as well as in the Operating System, and the Application.

Item	Description
Task Specific Instruction Sets	With each CPU it comes with specific kinds of programming that are optimized right in the hardware. This makes these processes much faster than they can be in a software application. Some general purpose instruction sets are now being included in some processors that make JonokeMed™ faster when run on those processors.
Heat	Heat is a by product of running the electricity through the processor. This is the data and instructions moving through the CPU. Heat is very destructive to the computer components. Therefore how the CPU handles the heat it will produce is critical to preventing failures.

When looking at a server get the model number of the CPU. Just because it is an i3/i5/i7 etc does not mean that all the i7s are the same. There are other identifiers. Once you have this information you can go onto the web and do a search for the processor model. You will get a lot of information from AMD, or intel on what that specific model has for components. You can then do comparisons to see what you are really getting. This is when you learn why there is a price difference.

2.1 More Cores or Faster Cores

Prior to version 5.5 of JonokeMed™ this answer was easy - Faster Cores. With JonokeMed 5.5 and beyond we are lucky because of the way the database engine is designed. We can have fast cores, and multiple processors/cores. JonokeMed™ server will use them all. JonokeMed™ will use up to 32,000 processors/cores. JonokeMed™ uses OS threading to span the tasks over the multiple cores. You can see then that JonokeMed™ can scale up to a very big system.

An advantage of these servers with multiple CPUs and Cores is that they can now handle more tasks. For example for a small site you can have one computer provide services for the JonokeMed™ server, Terminal Server, and ImageButler on the same computer. Previously you should have divided this up into three different computers. This is if you wanted good performance.

† So why did Jonoke recommend 2 core processors for JonokeMed™ if JonokeMed™ prior to version 5.5 only took advantage of one processor. Good question - you have been paying attention if you have been thinking this. The reason is that JonokeMed™ is not the only thing running on the server. There is that very critical item called the Operating System (OS). It needs to perform many tasks too. Then there are all the tag along applications that will also be running on the Server. So while JonokeMed™ is utilizing one processor, the OS and other applications can be using the other processor.

When the JonokeMed™ server flushes the cache to disk, it actually tells the OS to perform this task. Therefore the sooner the JonokeMed™ server can send this task to the server, and then not have to wait on the OS to perform the task then the sooner the JonokeMed™ server can get back to doing what you are asking it. So flushing of the cache will occur quicker with a dual core processor than a single core processor.

2.2 Task Specific Instruction Sets

AES (Advanced Encryption Standard) this improves the speed of applications doing encryption and decryption using the AEStandard. Encryption is becoming more and more common in systems. To protect the privacy of the

data from someone that steals your system - this is important. Therefore this involves having the data on the storage media (say hard drives) encrypted and having the CPU or RAID controller decrypt on the fly is more and more common. Therefore having these instructions built right into the hardware is much faster than having software (i.e. JonokeMed™ or the OS) perform this task.

Video is becoming very common data type on computers. Therefore Intel has made the decision in its latest processors to include instructions that process video. This is not such an important task on a server (currently), but is valuable on a workstation. I say currently, because facial recognition and other such tasks can be kept on the server and processed faster than dragging the data across the network to be performed on the client.

2.3 Heat

Computers work very well at a temperature that we like. Therefore well designed desktops and laptops rarely have problem with heat because we are sitting beside them. The computers with the biggest heat problem are the servers. The servers are put into a room where people don't work (which is the way it should be, because they are noisey). We then tend to forget about them. The problem comes when that room gets hot.

Heat kills computer components. Excessive heat will reduce the life span of your computer. Heat also slows down the computer. The electrons move slower through your system. Well designed systems have been engineered to dissipate the heat they produce. They also know how much heat each component will produce at various workloads. This way they know how to build in the airflow for the computer internally, where to put the fans, and the external vents (both incoming and outgoing). Therefore ensure to put all the covers for your computer on when you are using it. Read the documentation that came with your computer. Ensure to leave enough room around the computer to give it good airflow. Also pay attention to the temperature in the room your server is in. Just because it will fit in a closet, or desk cabinet does not mean it should go there - it should not. Keep a thermometer in the room where your server is. If it is getting over 85 degrees you likely need to look at various strategies to keep the server space cooler. This could mean vents in the doors (low & high for air flow), fans to suck the hot air out (ensure to have a vent to let cool air in), or actual cooling systems that blow the hot air out, and put cool air back into the room.

Knowledge of what is happening with the production computers out in the wild (no matter how much engineering they do) is critical to improving future models. There are various sensors in computes that tell the temperature of various components. Some manufactures (like Apple) put a lot of sensors in their computers to know what has been happening with their computers. For example, Apple laptops have sensors on the bottom, around the battery, on the palm rests, at the CPU, on the motherboard, and several more. You can get applications that will display what Apple is capturing with these sensors. This lets them build the next generation of laptop (computer / server) better by having this knowledge.

Someday we will see computers run by light, instead of electrons. These will be much faster, use less electricity, and produce less heat. Until then, we must be aware of the temperature of the environment we are running our servers in.

3.0 RAM - Random Access Memory

RAM is critical to the performance of your system. Next to the L3 cache it is the fastest place to have data reside to get great performance. There once was a data base system that was know for being fast. It was for small systems only though. Why was it fast - because it had to have ALL data in memory.

Getting a piece of data from RAM is hundreds of thousands of times faster than getting it off a fast RAID hard drive system. This is a fact of physics. Remember with the CPU, the distance between components effects speed, so having to wait for the spin of a hard drive platter to get to the location of the data, as well as move the head of the hard drive (think of a vinyl record player - hard drives work the same way) is why it takes so very long.

An aspect of RAM is that it comes in different speed. The faster the RAM, the faster CPU can write to it and get on to other tasks, and the faster the CPUD can get data from the RAM and get on to other tasks. So the lest time the CPU is sitting around tapping its fingers waiting on RAM the faster your computer will be. You want to get a server with 1333 Mhz speed (as of the writing of this document).

How Much RAM

With systems the size of JonokeMed™ it is very unlikely that you will spend the money to have all your data in RAM. If you do, you will have a system that is extremely fast. The issue is that RAM is expensive - relative to hard drive space. So this is the typical cost / benefit equation.

So the question then gets down to how much RAM should you have to give you reasonable performance. Prior to version 5.5 of JonokeMed™ this was rather easy - 4 Gigs. You would want 250 Mb for the cache, and 1.5 Gig for the application. Then you want enough for the Windows Server OS to run effectively, and all the little add ons that will come along for the ride in such a system.

Prior to version 5.5 of JonokeMed™, JonokeMed™ sever was a 32-bit application only. It could only utilize that much RAM. Therefore buying more was just a waste of money. This was ok until the last few years as the Server Operating system was a 32 bit application so it was limited to 4 Gigs too.

64 Bit Application

Starting with version 5.5 JonokeMed™ server can be installed as a 64 bit application. Therefore it can utilize Gigabytes of RAM, fully utilizing the 64 bit Server Application. So now this question is more complex. So lets delve into the questions and answers.

Utilize Multiple CPUs / Cores

Another advantage starting with JonokeMed™ version 5.5 is that it will fully utilize the multi-threaded architecture of the Windows OS to take advantage of multiple processors. This is a huge performance enhancement. To feed these processors though you need to ensure that you have enough RAM to keep the data coming. So the more processors you will be putting into your sever you had better add the RAM or you will be ‘starving’ the CPUs for data and thus slowing them down and therefore wasting your money on those CPUs.

How large is your Data Set

When you first get JonokeMed™ and you have your brand new server it is always fast. There is a reason for this. All the data likely fits in the database cache which is primarily located in RAM. Therefore all the components of the data (we will discuss this soon), can be loaded by the data base server into RAM as it needs it. Because it has enough room for it while it loads in other data it can keep it all in RAM. This equates to speed! Thus beware of the sales demo which looks extremely fast - this is likely the reason. That is why you must see a site with the system that is mature and in use. As questions like how much data they have.

As your data grows with use, less and less of the data can fit into RAM. All the serious RDBS (Relational Data Base Systems) recognize this and have built in their own special ways of handling this. The data is actually made up of parts that are not just the data you look at, or entered. This is the meta-data that is used by the RDBS to find the data quickly. Two of these are the Address Table and the Index Tables. So as the data grows JonokeMed™ server will stop loading data into RAM, and will limit itself to the Address Tables and Index Tables. Having all of this in RAM will still provide exceptional performance but it certainly is a step down from all data in memory.

When the JonokeMed™ server needs data it can use the Index Tables / Address Tables to find where the data will be located on the hard drive and then direct (through the OS / RAID controller) to get this data.

As your data continues to grow, the size of the Index tables and Address Tables gets to be such that they can't be held in the cache, along with the data that the system is currently using. Therefore the Server uses sophisticated algorithms to determine which ones it should keep in the cache, and which ones it should get rid of so that more that the Address tables / Index Tables / Data it currently needs can be brought into the cache.

Even the process of making this decision, getting rid of information takes time. Not only that, but the information in the cache needs to be rearranged to make room for the data so that it can be kept ordered and together in an efficient manner. This last piece of information is critical to understand that too much cache can also be bad as it can take to long to organize and rearrange the items in the cache. That is why in versions prior to version JonokeMed™ version 5.5 and with the hardware / OS of the day we limited the cache size to 250 Mb.

There will also likely come a time when not even all of one index will fit into the cache you have because there will be data, and other partial index tables and partial address tables in the cache. You need to have enough RAM so that the cache can hold at least enough of the index table / address table to be fast in its search of it. Having an often used table's index in the cache will make the system faster (i.e. patient table indexes). If a high percentage of these table's indexes do not fit in the cache (RAM) then the system will be slow.

Another aspect that must be considered. Everything in RAM dies when the power goes out. That is another reason that JonokeMed™ is not set up to have everything in RAM. At some point it must be written to disk so that if the power goes out your data is save. This is called flushing the buffers. When this is occurring nothing else can happen. If you have gigs of data that has changed, this pause will take several minutes. Therefore to keep this forced pause time to the minimum we flush the buffers every 6 minutes. In this way users generally do not see the pause as all the changed data is written from RAM to the hard drives. This is another reason to keep the cache smaller, and thus the less need for large amounts of RAM.

We will let you have another peek under the hood by bringing your attention to some 'gauges' on the JonokeMed™ server. The gauges from version 2 through 5.1 of JonokeMed™ stayed the same. These gauges let you see what percentage of time that the server went looking for data that it found the data in RAM. The JonokeMed™ server also displayed a dialog on screen when it was flushing the data in the cache to the hard drive. If this was flushing more than every 6 minutes it meant that a lot of data was getting changed and the cache was getting full.

With JonokeMed™ version 5.5 the server provides this information and more. You can watch this and learn what YOUR server is doing.

Jonoke has spent many long hours testing various memory settings, and torturing the system to test what the best settings are. Certainly we have much more experience with versions before 5.5 as the architecture of the server was mostly the same. With version 5.5 we are learning and testing again. So we expect that we will be modifying our recommendations as we learn more in the next few years.

The Bottom Line

- For sites with 10 workstations or more - Go with a 64 bit Server OS

- Get multiple CPUs (max 1 core / 25 users) Ideally get it down to between 5 and 10. Why - you want speed.

- You should have a minimum of 2 Gigs of RAM per CPU. If you have less you can be starving the CPU of data.

- i.e. 100 user site - 3 - 4 core CPUs with 120Gigs of RAM.

- For sites with 10 or less workstations go with a 32 OS - thus 4 Gigs is the maximum RAM you can have, OR

- For smaller sites, get a multi-core server, 64bit OS and use it for the JonokeMed™ server, ImageButler, and Terminal Server.

Wasting Money

If you have a 32 bit version of the JonokeMed™ server (all versions prior to version 5.5) and the 32 bit version of the JonokeMed™ 5.5 server then JonokeMed™ cannot access more than 4 Gigs of RAM. That is the limit period. Therefore unless you are planning on moving to JonokeMed™ 5.5 and the 64 bit server don't purchase more than 4 Gigs of RAM for the JonokeMed™ on this server.

† Going to a 64 bit Server OS with say 6 Gigs will give you better performance because the JonokeMed™ server can utilize the full 4 Gigs of RAM, while the OS and other tag alongs can use the remaining 2 gigs.

4.0 Secondary On-Line Storage

Most people do not think of hard drives as this, but before the time when hard drives came on every computer, there were a full range of options to consider. For the last 20 years there really has been only one technology to consider for this. That is fast hard drives configured with a RAID controller. This was the fastest, reasonable priced solution for Secondary On-line storage that included redundancy so that if one died (which it will), you can keep operating without a loss of data. Now there are several different configurations, and technologies to consider. Price, data size, and performance requirements will help you determine what you want / need.

4.1 Hard Drives

Not long ago (at least for my memory) there were two basic choices for hard drives. These were the ATA (IDE) drives aimed at workstations and the more serious (performance) SCSI drives aimed at servers. If your vendor

quoted you IDE drives for your server, they were just trying to get your business based on price. SCSI drives are faster, and there are reasons more than just the speed of the platter spinning as to why a hard drive can be faster.

Now days there are the same choices, but the technology for both has advanced. There are the SATA (Serial ATA) drives and SAS (Serial Attached SCSI) drives. Both offer better performance than their predecessor. The performance gap between them has narrowed, but not enough that you should choose SATA over SAS for any but the smaller implementations.

Just like the outside of the computer box, it is unlikely that all the details of the drive(s) are stated. Typically SATA drives have a lower RPM (Revolutions per minute) than the SAS drives. Typical SATA drives are between 5400 - 7500 RPM. There are some 10,000 RPM SATA drives though. SATA drives will come in larger sizes than SATA drives. OK, those are the basics that you likely find out from the box. Just from this the SATA drives will be faster at serving the data to your CPU. There are other factors though that will effect performance.

The spinning of the platters is not the only item that effects performance. If you remember (or have seen) a vinyl music record player that is essentially a simple hard drive. There is a platter (the music record), and the arm that reads the record, is the head that reads the music. A big difference with hard drives is that the head is not touching the surface of the hard drive. Therefore the head can be pushed toward the centre or outside of the platter so that it will find the data it needs. Each time data is needed, the RAID controller (or disk controller) will tell the hard drive where to find the data on the platter. The hard drive will determine where in the spin cycle this is, and move the head in or out so that it can read the data. This is a mechanical operation along with the spinning of the platter. This takes decades of time when measured by computers. Thus it is very slow. For us humans we will even notice the difference in the speed. SAS drives will have faster performing mechanisms to get the head of the heard drive to the spot where the data is.

† This description is critical in thinking about why your server's hard drives and data files need to be maintained. They need to have the data file compacted, and the hard drives defragmented. Not doing so will make any fast combination of hard drive and RAID controller be slow in time. This has nothing to do with the database engine - it has everything to do with the way the hard drives work. I could go on more about this, but this is not the scope of this paper.

In addition the hard drive itself with have 'RAM' on board it to store data for writing / reading. Therefore faster drives will have more of this on board RAM within the hard drive. As well, the interface between the computer and the hard drive will be faster. For servers now, you should be looking at the 6gig interface. These are only available on the SAS drives.

Now we get to the OS / and RAID controllers. Since the SCSI / SAS drives have supported advanced speed operations for longer, it is more likely that the RAID controller and OS will have support for these advanced speed operations built in. Without this, the computer cannot take advantage of these features. If the SATA drives support these features it may be wasted.

Another place where physics gets involved is in the size of the platter of the hard drive.

Bottom Line:

SAS drives, that have a high RMP, and lots of on board RAM, and high speed interface.

4.2 Solid State Drives (SSD)

Most everyone is familiar with these devices in their inexpensive format. They are know by many names; thumb drives, flash drives...

These are not hard drives at all. They are memory chips that do not loose what is stored on them when the power is taken away from them. The consumer drives that we generally use are not know for their speed. Part of this is that they are sold to be cheap - thus the technology used is inexpensive. Another reason is that the interface (USB) that they are used with is a limiting factor. So there is no use making these faster than the USB interface between the flash drive and the computer.

So why would we consider this technology. Well, like with hard drives, there is a superior technology for the flash drives that is called SSDs. Solid State drives have many advantages over a hard drive. The first is that there is nothing mechanical. Therefore they can simply be faster. There is no head to move in or out to get to the place where the data is stored. This is a huge performance increase right there. The SSDs don't have to spin the data around and wait until it gets into a certain place to read it. So this is another huge increase in performance. It is all about the physics of getting to the data. SSDs are very fast mainly because of this.

They are only recently becoming available though because of the cost to manufacture them, and some other technology issues that needed to be over come. Older SSDs had a limited number of read write operations before they would just not work. Though hard drives are the most common failure point in a computer, the life span of earlier SSDs was much less than a hard drives. There is also the issue of being able to clear out the 'memory' in the SSD so that artifacts of data did not remain. This was another technology hurdle. These hurdles have been mainly over come and manufactures are offering SSDs for servers and even workstations (mostly laptops). If anyone has used a laptop with a SSD in it, they will know how much faster it starts up. This is the most disk intensive operation that people will regularly notice. With a hard drive it may take a few minutes to startup (go get a cup of coffee). With an SSD instead of a hard drive, well you can try and get a sip of that coffee before your computer is ready to go. This is all because of the SSD technology.

Like with hard drives, there are different technologies in SSD. Buyer beware is critical. I will only purchase from a tier one manufacturer. With servers, I will only purchase the SSDs and controller from the tier one supplier. Their reputation is on the line with the server, and they have tested the components together. They will have installed the controller software (critical) that works best with the SSD and will be the latest version to provide the best performance and data management.

Bottom Line

- If you have more than 25 workstations - I would recommend using SSDrives in your system. They are certainly more expensive - but your time is much more expensive. The time they will give you each and every operation will more than make up for their cost.

4.3 Battery Backed Up RAM Drives

Remember that a few pages ago I stated that when you first get your JonokeMed™ system it is always fast because all your data is in RAM The problem was that your data outgrew what you could get in RAM. The reason is that motherboard manufactures do not permit you to add as much RAM as you might want to have all this RAM. As well, previous to 64 bit Operating Systems, the OS just could not access that much RAM.

So to give the brand new server (actually brand new data file) speed, some manufactures put in the effort (and investment) to design and build battery backed up RAM modules that looked to the OS as a normal hard drive. Therefore when the OS asks the 'drive controller' for some data from the disk it is delivered much quicker because it is running as fast as the IO within the computer. The RAM is battery backed up just in case the power is turned off. Of course it is a battery backup so the battery will die soon. Therefore you will likely have enough battery for an hour or so. So these system are always paired with hard drives so that the data can be written to them at least once per day, and if the power ever goes out.

Obviously this is the fastest solution for Secondary On Line storage. The downside is cost, and that you must have good secondary power backup.

Bottom Line

- If you can afford it - this is the fastest solution. It will raise your eyebrows for sure (both on its performance and its price tag).

4.3 SAN or NAS Drives

SANs (Storage Array Networks) and NAS (Network Attached Storage) are becoming more popular in enterprise networks. They have distinct advantages, thus why they are becoming more popular in enterprise systems. SANs and NANs are typically more expensive than DAS (Directly Attached Storage) which we have been talking about in the previous sections.

Some of the advantages of SAN and NAS.

It is easier to use many drives. You can have huge storage needs, and these devices can be configured to provide that storage.

They may have 'very large' cache so large amounts of data can be dumped to them. Memory is easy to keep running on batteries then hard drives. So the data in this cache can live longer while the power is out.

These are easier to make redundant across multiple servers. Typically you will have two fibre-channel ports on a SAN, letting you have two systems connected to the SAN.

The management interface for these will usually include fancier rebuild possibilities, and features such as snapshots that can make a backup and mirroring operations easier.

For the reasons mentioned above, SAN and NAS are an excellent choice for network drives when you are storing many small documents that users will need access to. For example if your office has many word processing files, spreadsheets, presentation files, graphic files, videos etc.

Notice I did not say database files.

Yes, all databases can use SAN and NAS. The reason is that to the Operating System these look just like a regular disk drive. The database engine does not look at volumes as such. It just requests data from the OS, and tells the OS to save the changes. SAN and NAS are much more complex to setup than DAS. When they are configured wrong, their performance is severely effected, and you can loose data. The technicians setting these up must know what they are doing.

Everything else being equal, going through the SAN/NAS interface no matter the type always adds latency no matter what. This is latency on top of the latency that is already in the storage devices used (i.e. hard drives). Remember our discussion in comparing hard drives and SSDs. Latency is why there is a huge difference between the SSDs and Hard Drives.

The cost for a SAN/NAS is very much higher in terms of performance per dollar compared to DAS.

There are reasons to choose SAN/NAS but it will have nothing to do with the database system at all. In fact the choice of SAN/NAS is made at the detriment of the performance of the data base system. Understand this is NOT a JonokeMed™ thing. This is ALL database systems.

5.0 Interface between Storage and Computer

5.1 Thunderbolt

As I was writing this paper, intel announced the release of their new interface for connecting devices such as hard drives. Frankly this is a game changing interface that 'blows the doors off' of all the competing technologies (USB, Firewire, and even internal controllers). The performance is 10GB/s bidirectional. A good high end RAID controller configured correctly will get over 1 GB/s unidirectional. As more devices come out to support this interface, and as the manufactures built their systems to support this new interface we will see the performance of database servers go up a magnitude again.

Apple was the first to implement this interface of all the computer manufactures. They implemented this in their MacBook Pro line. With the release of these MacBook Pros with Thunderbolt the only device that could use the port was Apple's Cinema displays. We are all having to wait 6 months for companies like Lacie to release their hard drives that will support Thunderbolt. As Lacie showed, only the fastest SSDrives configured in a Stripped Array can come close to keeping up with this interface, all the while, it is streaming data to other devices all over the same interface.

This interface release at the time of the writing of this paper, just goes to show that technology changes and can change dramatically. Therefore the need to keep up what is happening in the world of technology.

5.2 Drive Controller - RAID Controller

RAID - Redundant Array of Inexpensive Devices. Essentially it is linking together several hard drives (SSDs) in such a way that the OS sees the multiple devices as one device. There are different levels of RAID that can be implemented, and each has its advantages and thus its purpose.

Hardware vs Software RAID

There are two ways to implement RAID. One is using software that is run by the OS, and another is having a hardware card that sits between the motherboard and the hard drives. All the major OSes have the ability built in to have software RAID. Just because it is there though, does not mean you should use it. Software RAID has an advantage - it is inexpensive. It came with the OS purchase. The problem is that the whole reason you want to go

with RAID is typically got something to do with speed. Using software RAID is not fast, so do not use it. Hardware RAID is much faster than software RAID.

Software RAID has some disadvantages beyond its poor speed. It is much harder to configure. You are required to do that, and because you can do it, you or someone else with access to the server could play with the settings and at least cause even worse performance, or at worse loose all your data. Software RAID can loose its settings. When this happens you loose your performance at worse and can loose all your data at the worst. Remember too, when these bad things are happening you are not running your business. So the cost of the technician to fix it is the least of these costs.

The advantages of the hardware RAID controller is of course the opposite of the disadvantages of the software RAID controller. Hardware RAID advantages include: ease of setup, speed, reliability, and on some onboard encryption.

There are some inexpensive RAID controller cards that are sub $300. These are commonly referred to as Fake RAID. These are cards that don't actually include a dedicated storage processor on them, which is one of the parts that jumps the price up on the 'real' hardware RAID cards. These cards leave the processing to the OS. I am of the opinion that these cards are worse than software RAID. They may have better performance than software RAID, but they introduce another level of complexity, and typically have poor drivers.

Memory on the controller is another way to look at the performance that the RAID controller will give. The more RAM on the controller the faster it can give control back to the OS --> JonokeMed™ when it is working on saving data to the RAID.

A good RAID controller should also have battery backed up memory right on the card. This means that if there are brown outs to the computer, the sensitive data on the RAID controller will be protected better.

The disadvantage of the hardware RAID controller is cost. They cost anywhere from $300 to $1,200 for a server. For the performance they provide you will quickly save more than you will have paid out.

As with all things, all hardware RAID controllers are not equal. For example the DELL PERC6, H700 and H800 RAID controllers are very good performers. The DELL PERC5 and earlier models tended to be slow. HP P400 and P600 are as their number suggests lower performance controllers. The P800 is a respected RAID controller with good performance. The same is for other brands of computers / RAID controllers. Do your research - we have.

Encrypted Data

In some jurisdictions they require that all the data be encrypted on the hard drives. Though again you could use the OS to perform this task you will sacrifice a lot of speed if you use the OS to perform this task. Various RAID controllers will encrypt the data on the hard drives. In this case there are special codes that you must write down and store in a safe place (Other than on this server). These codes are the encryption keys. If the RAID controller goes bad and needs to be replaced you need to have these encryption keys to put into the replacement RAID controller. Without these keys you have just lots all your data in the highly organized encrypted data on the hard drives which may be in perfect shape. This cannot be over stated. You MUST store these keys safely. When you purchase a server from Jonoke we store these keys for you in JERMS, and on paper with your contract in our office. We also provide you with these keys too for you to store.

5.3 RAID Levels

RAID 0

RAID 0 is purely for speed. It is also called striping. The name striping describes what it does. It takes a stream of data and writes the data across all the drives in the RAID at the same time. Up to a limit the more drives in the the RAID the faster it will be in a linear fashion. Therefore a 3 disk in the RAID will be three times faster than 1 disk.

The downside is that when one of the drives fails you have essentially lost your data. So though the speed is tempting to jump for, the sobering loss of your data because a hard drive failed will change that jump to RAID 0.

RAID 1

RAID 1 is purely for backup of your data on the fly. With some sophisticated RAID 1 systems the reading of data can be faster because it will read from multiple drives at the same time. Though this is the case, the writes are no faster, and a database like JonokeMed™ does a lot of writing of data.

Do to the lack of speed increase (across both operations), RAID 1 is typically not used in database servers. The advantage is that if one of the drives fail the system continues to operate as the data is always written in more than one place.

RAID 5

RAID 5 has been used in Jonoke provided JonokeMed™ servers for almost 20 years. It is also called Striped with Parity. RAID 5 sits between RAID 1 and RAID 0. It strips data across multiple drives similar to RAID 0 to give better performance. Data is also written in a way that the data resides in more than one place. This is done so that if one drive fails no data is lost and the system continues. Write performance suffers with RAID 5 in comparison to RAID 0. Though this is the case, it is faster than RAID 1.

RAID 6

RAID 6 is similar to RAID 5, except that there is more parity information saved. This means that up to 2 drives can fail and the system does not loose data and can continue to operate. Another advantage of RAID 6 is that it is faster at rebuilding the new drive that was placed in for a failed one. RAID 6 is a newer RAID style that we have used from time to time.

RAID 10

RAID 10 (or 1+0) provides the speed of RAID 0, with the redundancy of RAID 5. It does this by both stripping data across 2 drives (RAID 0), and also writing parity data at the same time to another drive (RAID 5). RAID 10 is particularly good in an environment where there is a lot of writing of data. Thus RAID 10 is the best for high performance database systems like JonokeMed™. Jonoke typically uses RAID 10 in our larger installs.

Drive Error Handling

What most people do not recognize is that there are failures on hard drives often. There gets to be more and more the older the drive is, and the more the system is abused (heat, power outages). This is why database maintenance is required. Hard drive manufactures know this and deal with this in a few ways.

When a hard drive finds an error (can't read a sector of the drive correctly) it will perform different tasks. SATA drives (read consumer drives) are configured to be aggressive in retrying to correct the error automatically. What this means is that the drive will spend 'alot' of time trying read the data and correct the problem. They are designed this way to try and preserve the space on the drive, as this is considered more important than the people's time waiting for the data.

With SAS drives the assumption of the configuration of the system, and the use being made of it is different. With SAS drives the drive will not spend much time trying to resolve the problem. It will assume there is another copy of the data (i.e. RAID 5/6/10). It will quickly report that it cannot get the data. The system will flag that sector as bad so that no new data is written to it. The system will also get the data from the alternate location to feed the system, and then write another copy of the data to continue having redundancy of the data.

5.4 Hard Drive Reliability Reports

Hard drives are the component of the computer that is the most likely to fail. The reason is multi-fold. First is that it has mechanical parts, and mechanical parts are more prone to failure. As well, the disk platters that the data is written to degrade with time, and abuse. In fact your data is degrading each day on the hard drives - even without use. It is for this fact that Jonoke always places drives in a redundant array.

There are public hard drive reliability studies. Large cloud storage providers have used many hard drives and thus have lots of stats on this. It is this data that we have come to some of the conclusions in this paper.

When this knowledge is applied to your servers, you can begin to understand a major reason why your server slows down with age. More and more drive errors are occurring. This slows down the operation of the server.

5.5 Controller Firmware (software)

As discussed earlier it is important that you ensure that the RAID controller, and the firmware installed on it matches well with the motherboard and the disk drives you are using. This was the cause of Jonoke's first Server nightmare from a client's locally provided server. The expected result of the mismatch was reliability. It was corrupting all the data.

SATA drives are oriented at consumers. Consumer products will often see large changes in the firmware. What is worse is that the change will not show up in a model number change. To most consumers this will not effect any issues for them. In something as critical as a database server where a replacement drive is installed in a RAID configuration a change in the firmware can have catastrophic results (slow the system down, and corrupt your data). SAS drives tend to have a much more stable firmware and thus longer lifespan. This firmware is part of what determines the size of the drives that can be offered. Therefore SAS drives tend to lag behind SATA drives in their size.

5.6 Using External hard drive(s) over Firewire / USB

Using drives over USB or Firewire has a distinct disadvantage other than the slow interface. These interfaces typically do not sent the error reports back to the system and thus not have the advantages to handle these error conditions. They also may not properly handle write caching for the same reason. Therefore they will be slower in writing than the interface speed may indicate.

Using eSATA is much better in this regard. eSATA is designed to operate just like the DAS drives. The write caching is handled properly and the error reports are also handled.

5.7 Write Back caches

As discussed elsewhere hard drives are very slow compared to the CPU of the computer. Therefore make the application wait for all the data to be written to the hard drive before it performs the next operation would dramatically effect the performance of the system. For this reason modern computer systems have caches built in at several levels. JonokeMed™ has it's own cache as well - adding another level of caching of data. This is very typical of enterprise database systems.

A modern computer system will have different levels of caches.

Database cache: JonokeMed™ server has it's own cache. JonokeMed™ will store changes to the data in this cache so that it can carry on operations quickly. It also stores data that is not changing but needs regularly here as well. Previous to version 5.5 JonokeMed™ stored changes to the data to a maximum of 6 minutes before passing the changes through the 'chain of control' to be written to the hard drives. With JonokeMed™ version 5.5 due to the size of the cache, and other factors this has been reduced to every 20 seconds for maximum performance.

Operating System Cache: This cache will easily be gigabytes in size. This cache will often be filled by JonokeMed™ server when it tells the OS to save all changed data in the JonokeMed™ cache. This is good in that as far as JonokeMed™ knows the data has changed, and in logic operations it has been changed. Having this cache helps to reduce the time that the 'flushing of the cache' icon will be seen on the server. We want this to be as short as possible because when JonokeMed™ is flushing the cache all data base operations must stop. Having this be so short that users never notice it is what we are targeting for.

RAID Controller Cache: As we described earlier the RAID controller has it's own cache too. This is typically in the sub Gigabyte size. For most usages of JonokeMed™ this size is perfect. What occurs here is that when all the writes to the hard drive are passed from the RAID controller and can be contained in memory of the RAID controller, as far as the OS is concerned the data has been saved. Then the RAID controller can feed this data to be written to the hard drives as fast as they can take the data.

Disk Controller write Cache: All SAS and SATA drives have their own cache too. These are typically in the 8 to 32MB size.

To avoid going into a lot of detail, to ensure that all this data sitting in these caches is not lost, your system's components must all be set up to work well together. This is another peak at why working with a reputable server manufacture is very important.

6.0 Virtual Server

Virtual environments have become very popular in the last few years. There are some very good reasons to implement virtual servers in some environments. Here are some of the reasons:

- easy to backup all the various virtual environments and then easy to restore them should the need arise.

- when managing many different OS systems it takes less training to get an IT person competent at restoring these multitude of systems.

- when you have many different pieces of software that require different OSes, and different configurations for the software a virtual environment makes sense when all these systems can reside on the same server hardware.

- When different applications can run on the same server, having a virtual environment is easy to help keep them separate, backup, restore.

So this is all about ease of maintenance by the IT staff, and sharing the resources of the server hardware with other applications. Does JonokeMed™ work in a virtual environment - yes. There are times in large organization where their rules dictate that the server must be in their server farm where everything is on a virtual server. In these cases the users have no choice but to have a virtual server environment.

In 99% of the other cases JonokeMed™ is running in an environment where it is the only, or one of only a few servers in the organization.

There is only one server computer. Therefore when you put multiple OSes and multiple applications on the server hardware, it must now serve its resources out to all those extra demands on it. It cannot be anything but slower. Any argument by hardware people is absolutely wrong. I often find that the hardware maintenance people do not care about the performance of the system and the wasted time that the users have waiting for the server to deliver other services. They care only that their hardware is easy to maintain, and fits within their training of system management.

The IT company / staff has a far different perspective than Jonoke has. We know that our users see our name all day long as they perform each task. If it is not fast then they think about JonokeMed™ being slow - not the IT company that they never see being the cause of the problem. Jonoke is very concerned about the performance of the system because it is our reputation on the line.

So only put JonokeMed™ on your server computer. You do not want it sharing resources so that you can have maximum performance. Once you have only one OS and one application running on a server then it is a huge waste of time and effort to put JonokeMed™ on a virtual environment.

Bottom line - DO NOT USE A VIRTUAL ENVIRONMENT.