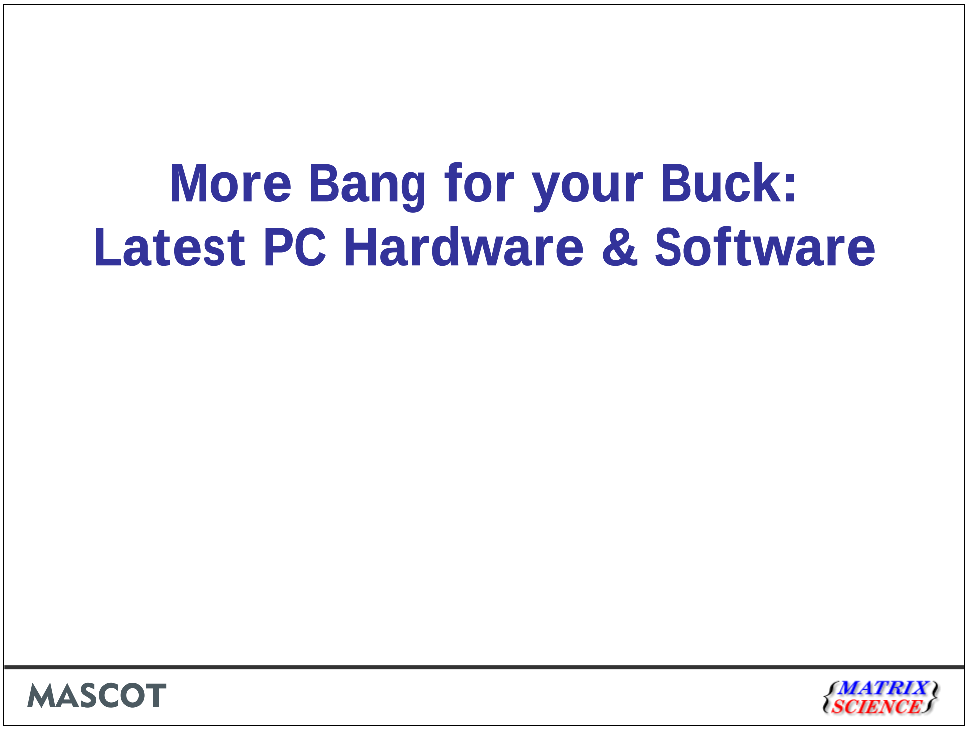
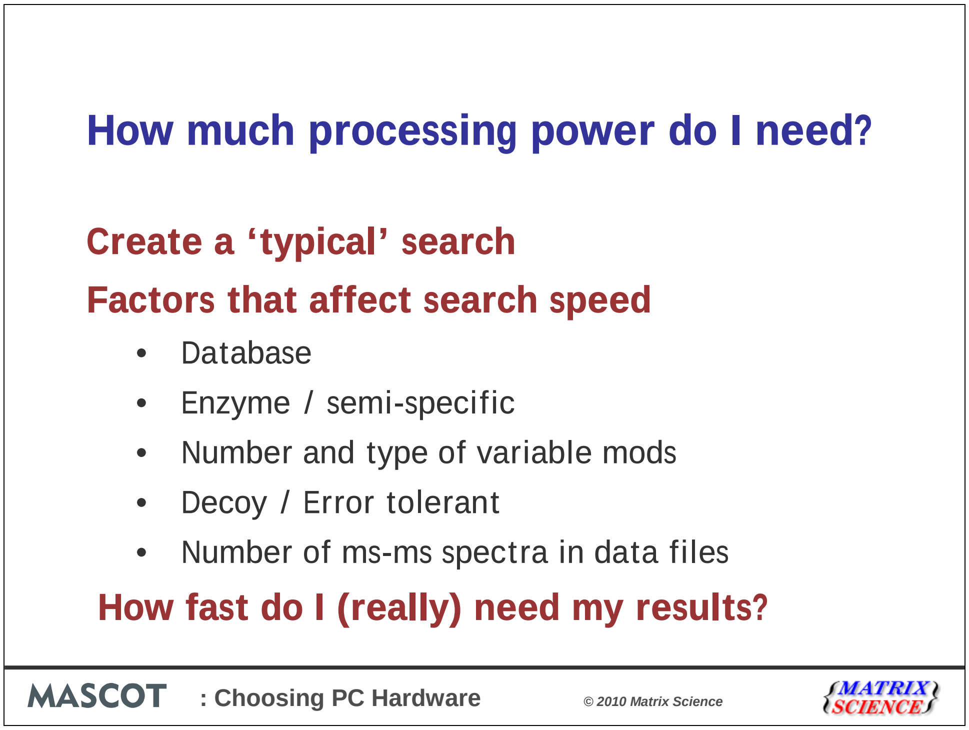


More Bang for your Buck: Latest PC Hardware & Software



How much processing power do I need?

Create a ‘typical’ search

Factors that affect search speed

- Database
- Enzyme / semi-specific
- Number and type of variable mods
- Decoy / Error tolerant
- Number of ms-ms spectra in data files

How fast do I (really) need my results?

MASCOT : Choosing PC Hardware © 2010 Matrix Science MATRIX SCIENCE

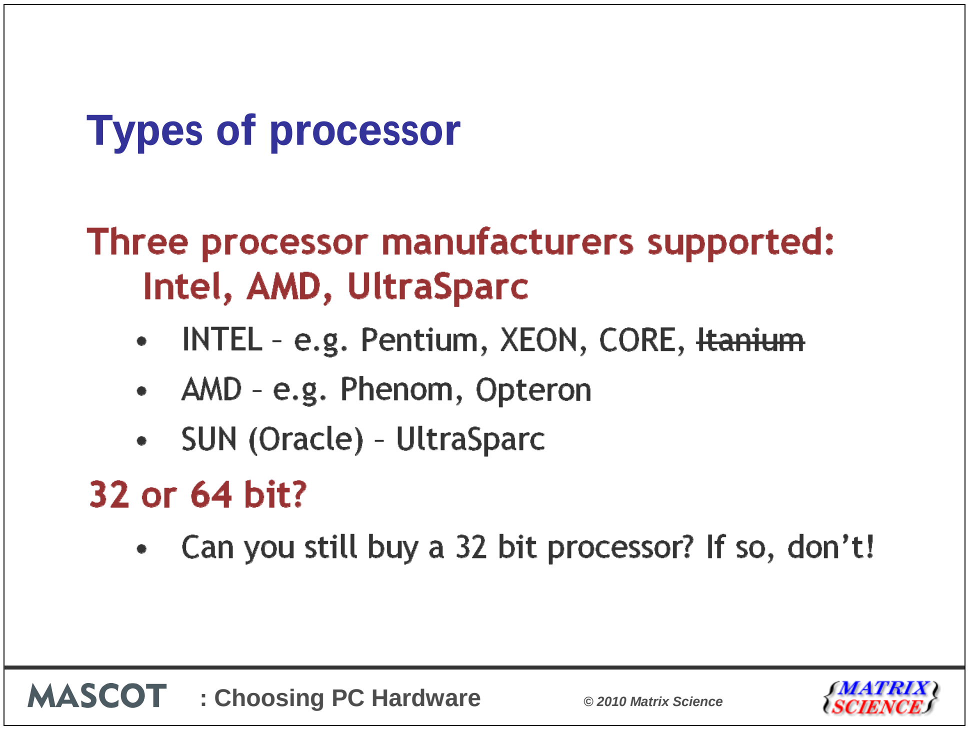
We often get asked for advice about what hardware to buy. The question comes equally from people buying Mascot for the first time and from existing customers. When people initially contact us, they often say that they need a large cluster and want help deciding how large it needs to be. It doesn’t have to be a guessing game, and I want to show you how to make a pretty accurate estimate of what processing power you need. Buying something too powerful is a waste of money, and getting something too slow is also be expensive in terms of wasted time.

The first key step is try and create an ‘typical search’ and either run a test yourself or ask us to test it. Once you’ve seen how long it takes to run, you can extrapolate to determine what hardware you need. Remember database size may double during the lifetime of your hardware.

The following factors affect the speed of a search:

The second key step is to think carefully and realistically about how fast you need your results. Suppose you typically do an LC run that finishes after you leave the lab for the evening and then a search is run automatically when the run is finished. It doesn’t matter whether a search takes 10 minutes or 8 hours unless you intend to return to work in the middle of the night. Obviously, that’s an extreme case, but most people don’t need the results from a 5 hour run in 5 minutes.

So, if you determine that a Mascot search on your current system takes 8 hours and you need it done in 2, you need a factor of 4. I’ll be covering how to get this factor of 4 improvement in the most cost efficient way.



Types of processor

Three processor manufacturers supported: Intel, AMD, UltraSparc

- INTEL – e.g. Pentium, XEON, CORE, ~~Itanium~~
- AMD – e.g. Phenom, Opteron
- SUN (Oracle) – UltraSparc

32 or 64 bit?

- Can you still buy a 32 bit processor? If so, don't!

MASCOT : Choosing PC Hardware © 2010 Matrix Science MATRIX SCIENCE

Mascot is processor bound and the most important decision to make is what type of processor.

We support processors from Intel, AMD and Sun (Oracle).

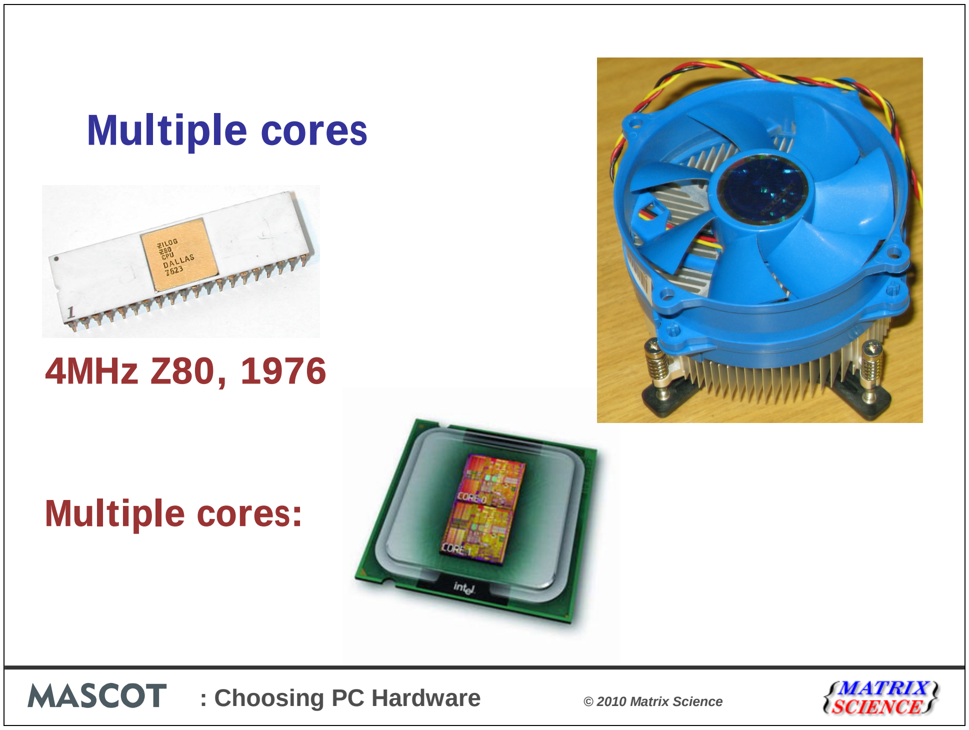
Note that we don't support Itanium and have no intention to do so.

UltraSparc is outside the scope of this presentation. You need to be a real Solaris 'fan' to want to use these.

Mascot 2.2 was the last version with support for Power chips in IBM AIX systems.

A few years ago, AMD processors suddenly became faster and better value for money than Intel processors. This was a 'wake up' call for the huge Intel corporation who have responded with faster processors using less power.

The question of 32 or 64 bit processors is now 'dead' – just choose 64 bit.



Multiple cores

4MHz Z80, 1976

Multiple cores:

MASCOT : Choosing PC Hardware

© 2010 Matrix Science

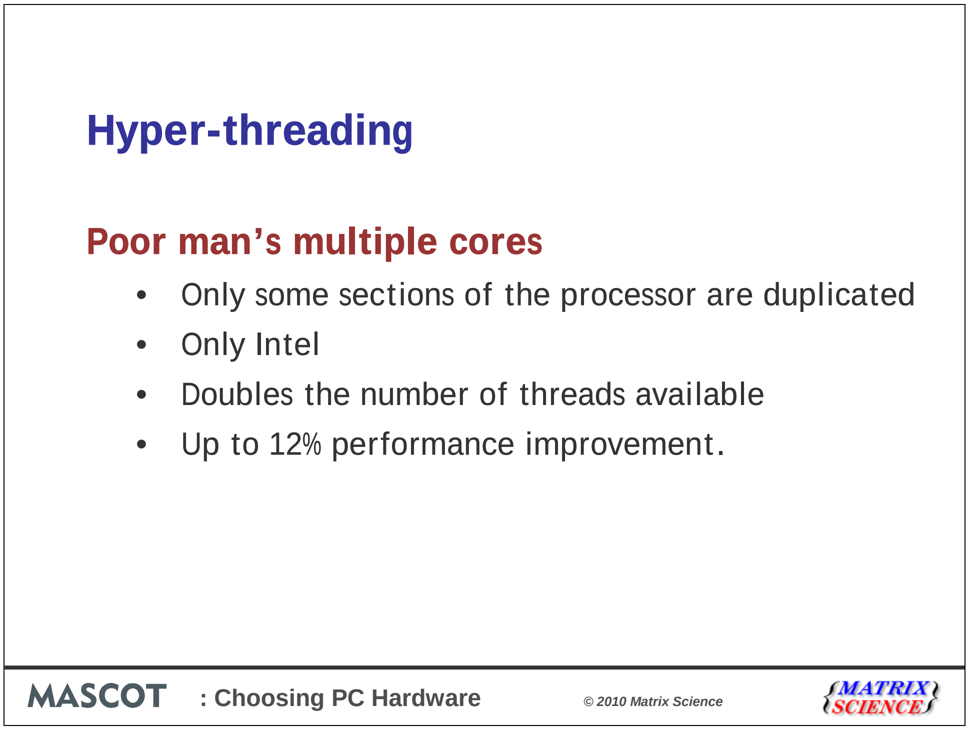
MATRIX SCIENCE

Those of us who are of a certain age will remember the 4MHz Zilog Z80 processors from back in the late 1970s.

Processor speeds increased regularly until about 2006. On the right is, I believe, the Intel PC processor with the fastest official clock speed ever. It’s the Presler Extreme edition 3.73GHz introduced in March 2006. The image is rather small, because I wanted to show the next image roughly to scale – this red hot processor required a huge fan and heat sink to keep it cool.

Heat was one of the major reasons why processors couldn’t carry on getting faster. One option was obviously to put more processors on the same motherboard, but this increases cost and physical limits become an issue, particularly for laptops. So, the solution is to put more processors in the same package. It’s not quite the whole processor that is duplicated, but ‘just’ the ‘core’ of the processor. Some things like internal cache memory are shared between the cores. For Intel and AMD processors, the current limit is 8 cores.

Thanks to Gennadiy Shvets for the photograph of the Z80 processor. Taken from http://en.wikipedia.org/wiki/File:Zilog_Z80.jpg



Hyper-threading

Poor man’s multiple cores

- Only some sections of the processor are duplicated
- Only Intel
- Doubles the number of threads available
- Up to 12% performance improvement.

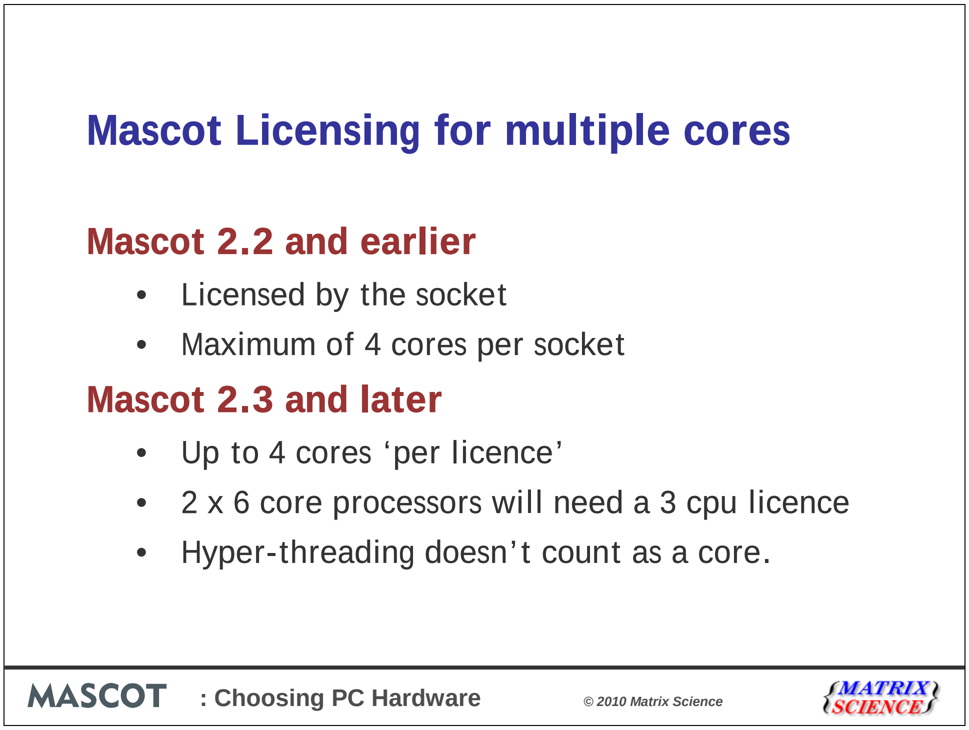
MASCOT : Choosing PC Hardware © 2010 Matrix Science MATRIX SCIENCE

At this point, I should briefly mention hyper-threading, even though it’s not something to get excited about.

This is available on Intel processors, but not AMD. Hyper-Threading works by duplicating certain sections of the processor - those that store the architectural state - but not duplicating the main execution resources. This allows a Hyper-Threading equipped processor to pretend to be two "logical" processors to the host operating system, allowing the operating system to schedule two threads or processes simultaneously.

When HT is enabled, 2 logical "processors" per physical processor will be visible to Mascot. So, for example, a single physical Xeon 5000 processor with dual cores and HT will appear to have 4 cpus. In this case, the number of threads should be set to 4 using the database maintenance utility.

Hyper-threading can give up to a 12% performance increase. It is not equivalent to a true multi-core processor.



Mascot Licensing for multiple cores

Mascot 2.2 and earlier

- Licensed by the socket
- Maximum of 4 cores per socket

Mascot 2.3 and later

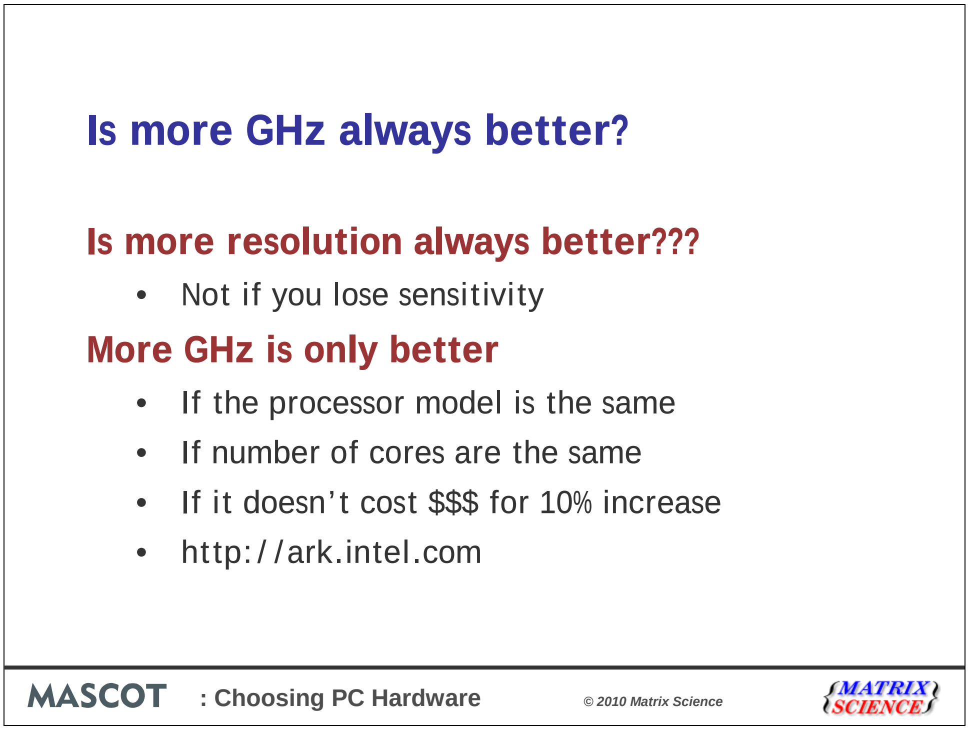
- Up to 4 cores ‘per licence’
- 2 x 6 core processors will need a 3 cpu licence
- Hyper-threading doesn’t count as a core.

Mascot has always been licensed by the number of cpus because we believe it’s fairer if heavier uses with large amounts of data to process pay more than someone just performing say, some PMF searches. This worked fine when processor speeds pretty much kept up with increase in size of the databases. The equivalent increase in performance has been met with 2 and 4 core processors, but it now looks as though we may be heading towards very expensive processors with a very large number of cores. So, in Mascot 2.2, we decided to put a cap of 4 cores per socket. If you buy a new system with 6, 8 or more cores, you simply can’t run Mascot 2.2 on this system. Hyper-threading does not count towards the number of cores.

Btw, you will need to apply the latest patch for Mascot 2.2 if you want to use the latest Nehalem or Westmere quad core processors.

For Mascot 2.3 and later, we have changed the licensing so that you can have 4 cores per license. For example, a system with 2 six core processors has 12 cores, so needs a 3 cpu license. A system with 2 dual core processors just needs a single cpu license.

For earlier versions of Mascot, please see the help page on PC specs which states which versions support dual or quad core processors.



Is more GHz always better?

Is more resolution always better???

- Not if you lose sensitivity

More GHz is only better

- If the processor model is the same
- If number of cores are the same
- If it doesn’t cost $$$ for 10% increase
- http://ark.intel.com

MASCOT : Choosing PC Hardware

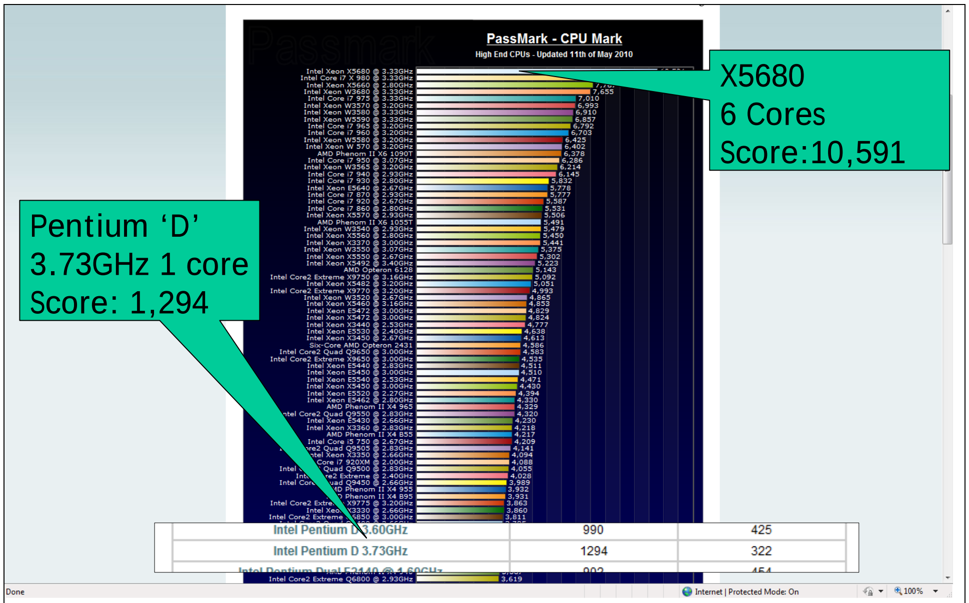
© 2010 Matrix Science

MATRIX SCIENCE

This is so similar to the question: Is more resolution always better…. and the answer is clearly no.

Hopefully, you’ll have understood that a single core Presler Extreme edition running at 3.73GHz is not better than a recent quad core processor running at 2 GHz.

Comparing like with like is quite hard. If you want to compare two Intel processors, type your processor numbers into the search box in the ark.intel.com web site which incidentally also has the estimated price.


Passmark
PassMark - CPU Mark
High End CPUs - Updated 11th of May 2010
Intel Xeon X5680 @ 3.33GHz
Intel Core i7 X 980 @ 3.33GHz
Intel Xeon X5660 @ 2.80GHz 7,767
Intel Xeon W3680 @ 3.33GHz 7,655
Intel Core i7 975 @ 3.33GHz 7,010
Intel Xeon W3570 @ 3.20GHz 6,993
Intel Xeon W3580 @ 3.33GHz 6,910
Intel Xeon W5590 @ 3.33GHz 6,857
Intel Core i7 965 @ 3.20GHz 6,792
Intel Core i7 960 @ 3.20GHz 6,703
Intel Xeon W5580 @ 3.20GHz 6,425
Intel Xeon W 570 @ 3.20GHz 6,402
AMD Phenom II X6 1090T 6,378
Intel Core i7 950 @ 3.07GHz 6,286
Intel Xeon W3565 @ 3.20GHz 6,214
Intel Core i7 940 @ 2.93GHz 6,145
Intel Core i7 930 @ 2.80GHz 5,832
Intel Xeon E5640 @ 2.67GHz 5,778
Intel Core i7 870 @ 2.93GHz 5,777
Intel Core i7 920 @ 2.67GHz 5,587
Intel Core i7 860 @ 2.80GHz 5,531
Intel Xeon X5570 @ 2.93GHz 5,506
AMD Phenom II X6 1055T 5,491
Intel Xeon W3540 @ 2.93GHz 5,479
Intel Xeon X5560 @ 2.80GHz 5,450
Intel Xeon X3370 @ 3.00GHz 5,441
Intel Xeon W3550 @ 3.07GHz 5,375
Intel Xeon X5550 @ 2.67GHz 5,302
Intel Xeon X5492 @ 3.40GHz 5,223
AMD Opteron 6128 5,143
Intel Core2 Extreme X9750 @ 3.16GHz 5,092
Intel Xeon X5482 @ 3.20GHz 5,051
Intel Core2 Extreme X9770 @ 3.20GHz 4,993
Intel Xeon W3520 @ 2.67GHz 4,865
Intel Xeon X5460 @ 3.16GHz 4,853
Intel Xeon E5472 @ 3.00GHz 4,829
Intel Xeon X5472 @ 3.00GHz 4,824
Intel Xeon X3440 @ 2.53GHz 4,777
Intel Xeon E5530 @ 2.40GHz 4,638
Intel Xeon X3450 @ 2.67GHz 4,613
Six-Core AMD Opteron 2431 4,586
Intel Core2 Quad Q9650 @ 3.00GHz 4,583
Intel Core2 Extreme X9650 @ 3.00GHz 4,535
Intel Xeon E5440 @ 2.83GHz 4,511
Intel Xeon E5450 @ 3.00GHz 4,510
Intel Xeon E5540 @ 2.53GHz 4,471
Intel Xeon X5450 @ 3.00GHz 4,430
Intel Xeon E5520 @ 2.27GHz 4,394
Intel Xeon E5462 @ 2.80GHz 4,330
AMD Phenom II X4 965 4,329
ntel Core2 Quad Q9550 @ 2.83GHz 4,320
Intel Xeon E5430 @ 2.66GHz 4,230
Intel Xeon X3360 @ 2.83GHz 4,218
AMD Phenom II X4 B55 4,217
Intel Core i5 750 @ 2.67GHz 4,209
ore2 Quad Q9505 @ 2.83GHz 4,141
ntel Xeon X3350 @ 2.66GHz 4,094
Core i7 920XM @ 2.00GHz 4,088
Intel Quad Q9500 @ 2.83GHz 4,055
Int re2 Extreme @ 2.40GHz 4,028
Intel Core uad Q9450 @ 2.66GHz 3,989
MD Phenom II X4 955 3,932
D Phenom II X4 B95 3,931
Intel Core2 Extre X9775 @ 3.20GHz 3,863
Intel Xeo X3330 @ 2.66GHz 3,860
Intel Core2 Extreme 6850 @ 3.00GHz 3,811
Intel Core2 Extreme Q6800 @ 2.93GHz 3,619
Done
Internet | Protected Mode: On
100%

X5680
6 Cores
Score:10,591

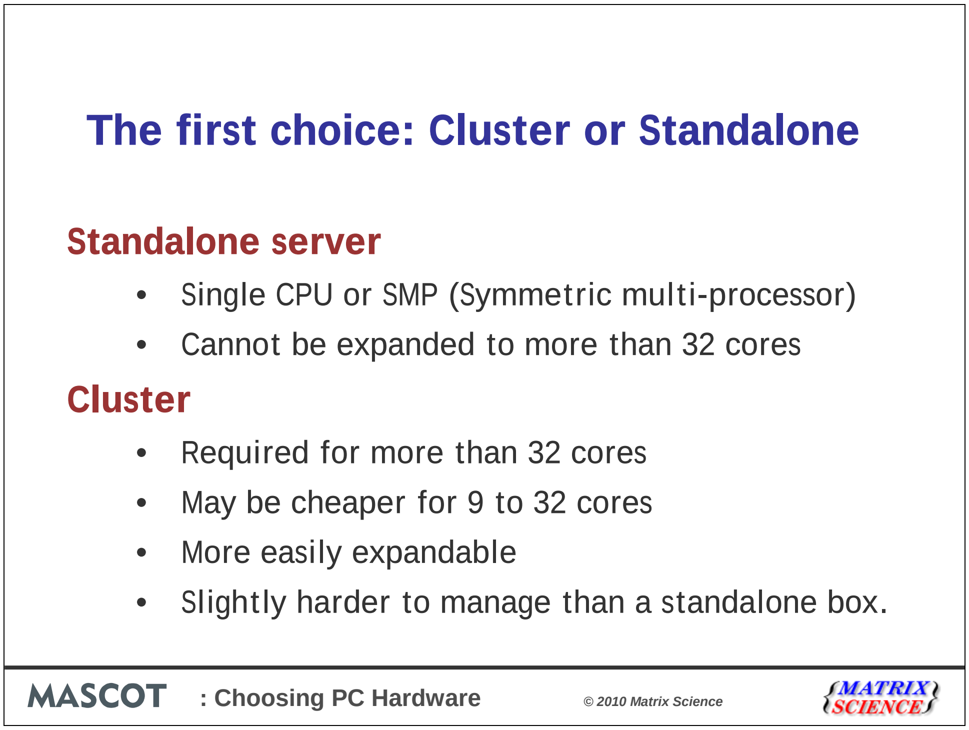
Pentium 'D'
3.73GHz 1 core
Score: 1,294

Intel Pentium D 3.60GHz	990	425
Intel Pentium D 3.73GHz	1294	322
Intel Pentium Dual E2140 @ 1.60GHz	902	454

http://www.cpubenchmark.net/high_end_cpus.html

The 'Pass mark' web site is excellent and gives a good indication of comparative performance for processors for Mascot

Suppose that you are currently running a system with one of those red hot Pentium D 'Presler' processors running at 3.73 GHz and that you have calculated that you need a factor of 4 improvement. Not all processors are listed on the home page and you may find it easier to find your processor with the search facility . The 'Presler' has a score of 1,294 and 1294 * 4 is approximately 5000. So, you can achieve double this requirement with a system with an X5680 processor at the top of the chart, or with many of the other current processors.



The first choice: Cluster or Standalone

Standalone server

- Single CPU or SMP (Symmetric multi-processor)
- Cannot be expanded to more than 32 cores

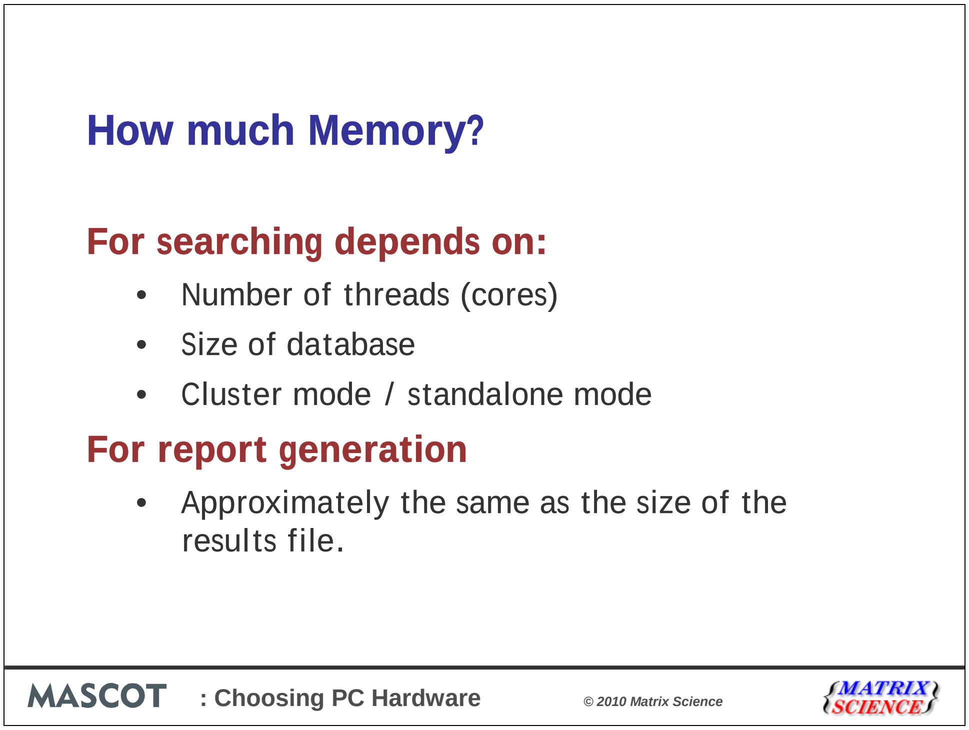
Cluster

- Required for more than 32 cores
- May be cheaper for 9 to 32 cores
- More easily expandable
- Slightly harder to manage than a standalone box.

MASCOT : Choosing PC Hardware © 2010 Matrix Science

In the previous example, we saw that the required performance could easily be met by a single quad core cpu. If we needed double that performance, then a system with the fastest currently available six core processor would be suitable. If we needed double that, then a system with 2 x 6 core processors would be sufficient. It will soon be possible to buy a system with 4 eight core processors, but if you need to go faster than that, then you will need to install a cluster.

As we’ll see soon, a cluster doesn’t have to be something grand. It can be just 2 or 3 standard PCs connected together on the LAN. One advantage of the cluster configuration is that to go faster at a later date, you simply add more nodes and there is no extra software to install apart from the operating system. For an SMP box that is full, there is no upgrade path except possibly replacing the processors with slightly faster ones.



How much Memory?

For searching depends on:

- Number of threads (cores)
- Size of database
- Cluster mode / standalone mode

For report generation

- Approximately the same as the size of the results file.

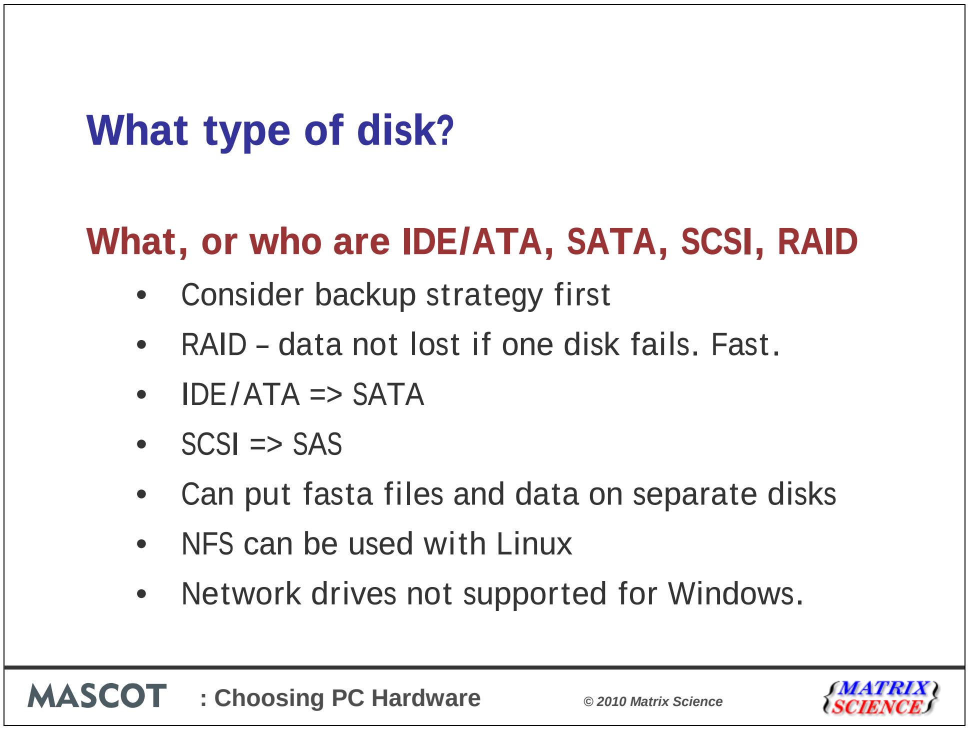
MASCOT : Choosing PC Hardware © 2010 Matrix Science

This is unfortunately a very difficult question! The executive summary at the top of our help page says you should have “at least 4GB of RAM (preferably 8Gb)”.

The amount of memory required for each search depends on a number of factors, the main one being the number of threads which, for a standalone system relates to the number of cores used for searching.

For report generation, the amount of memory required increases with the size of the results file and you will need approximately the same amount of memory as the size of the results file. If you want to be loading reports at the same time as running searches, you need to take this into account. So, if you are creating 6 Gb results files, allow 6 Gb of memory.

The good news is that memory is cheap, so in general the more the better. For a system with a single quad core proccessor, I’d recommend at least 8 Gb. For more than that, 16 Gb is recommended.



What type of disk?

What, or who are IDE/ATA, SATA, SCSI, RAID

- Consider backup strategy first
- RAID – data not lost if one disk fails. Fast.
- IDE/ATA => SATA
- SCSI => SAS
- Can put fasta files and data on separate disks
- NFS can be used with Linux
- Network drives not supported for Windows.

MASCOT : Choosing PC Hardware © 2010 Matrix Science {MATRIX} {SCIENCE}

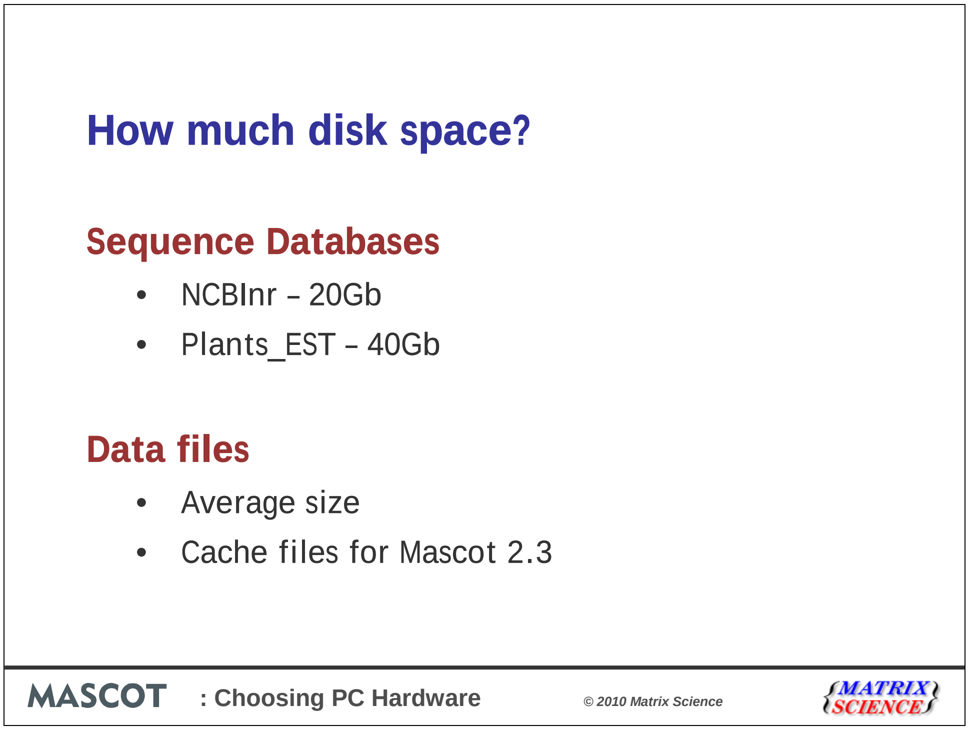
Before Mascot 2.3, we didn't care too much about disk performance and recommended that you spend more on memory and less on disks. For Mascot 2.3 and later, we use the disk more heavily for caching the results and for splitting up large searches and percolation, so faster disks are worth paying extra for.

One decision that you need to make early on is how upset you will be if your hard disk crashes. If, for example you always print out your results and you don't ever go back to old results or if you take a full backup every night, disk failure is not likely to cause a nervous breakdown. If on the other hand you only backup every six months and losing all your data is likely to induce you to jump off the top of a tall building, then you might like to consider the safer alternative of RAID. RAID stands for a redundant array of independent disks. One advantage is that it can be faster than a single disk because each chunk of data comes from multiple disks and the other is that in most configurations, if one disk fails all is not lost and you just need to arrange to replace the faulty disk as quickly as possible. However, a decent RAID card is expensive, and you will need at least three disks, so this may not be an option. Your computer supplier should be able to advise on the best type of RAID within your budget if you explain that you need good performance and don't want to lose your data when one disk fails.

If you can't afford or don't need RAID then you want to find the fastest disks possible. In the old days, there were two main alternatives: IDE (also known as ATA) or SCSI. Typically, Apple Macs and expensive servers had SCSI, while cheaper systems had IDE. These were parallel devices that have now been superseded by serial devices: serial ATA and serially attached SCSI. In general, you will find that the SAS drives are about twice as fast as the SATA drives and cost twice as much. If you can afford SAS, it is worth buying them, but not if it means having less memory or much slower processors.

For all Mascot installations, it's easy to put the fasta files wherever you like, so you can move these to a separate disk. For Windows, it's slightly harder to move the data directory to a separate drive, but it is possible. With linux, of course it's easy to use soft links and put files wherever you like – even on an NFS drive.

We don't support using network drives with Windows.



How much disk space?

Sequence Databases

- NCBInr – 20Gb
- Plants_EST – 40Gb

Data files

- Average size
- Cache files for Mascot 2.3

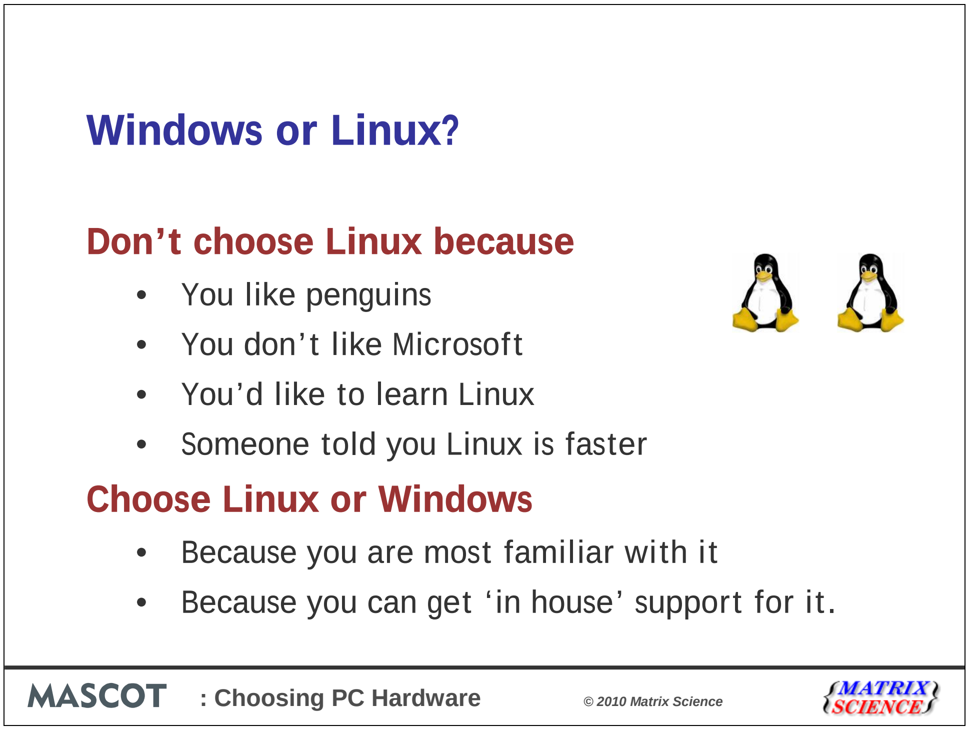
MASCOT : Choosing PC Hardware © 2010 Matrix Science MATRIX SCIENCE

Sequence databases take up a lot of disk space. For example, you’ll need 20 Gb for NCBInr today assuming that you want to use the database maintenance utility to update the database regularly. The Plants_EST database needs twice this amount.

For the datafiles, it’s best to run an example search as mentioned earlier, see how large it is and then multiply by the number you are likely to run in a year.

Mascot 2.3 and later will create cache files for viewing reports more rapidly and these can also take up a large amount of space. Add approximately 20% to the size of the .dat files. These can obviously be removed if you run out of space and they are just re-created when the search is viewed again.

It can be a surprisingly large amount of disk space. This number is also useful for your backup strategy.



Windows or Linux?

Don’t choose Linux because

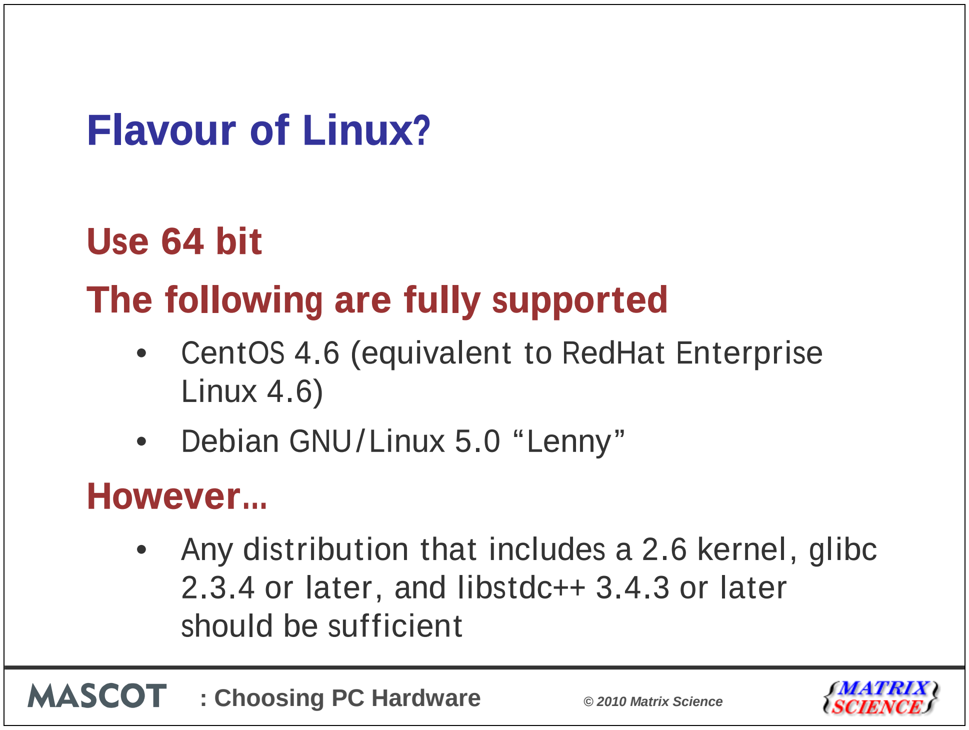
- You like penguins
- You don’t like Microsoft
- You’d like to learn Linux
- Someone told you Linux is faster

Choose Linux or Windows

- Because you are most familiar with it
- Because you can get ‘in house’ support for it.

Free to swap platform during first year or at any time when under warranty

We don’t support OSX or OpenSolaris or any other OS on Intel/AMD processors.



Flavour of Linux?

Use 64 bit

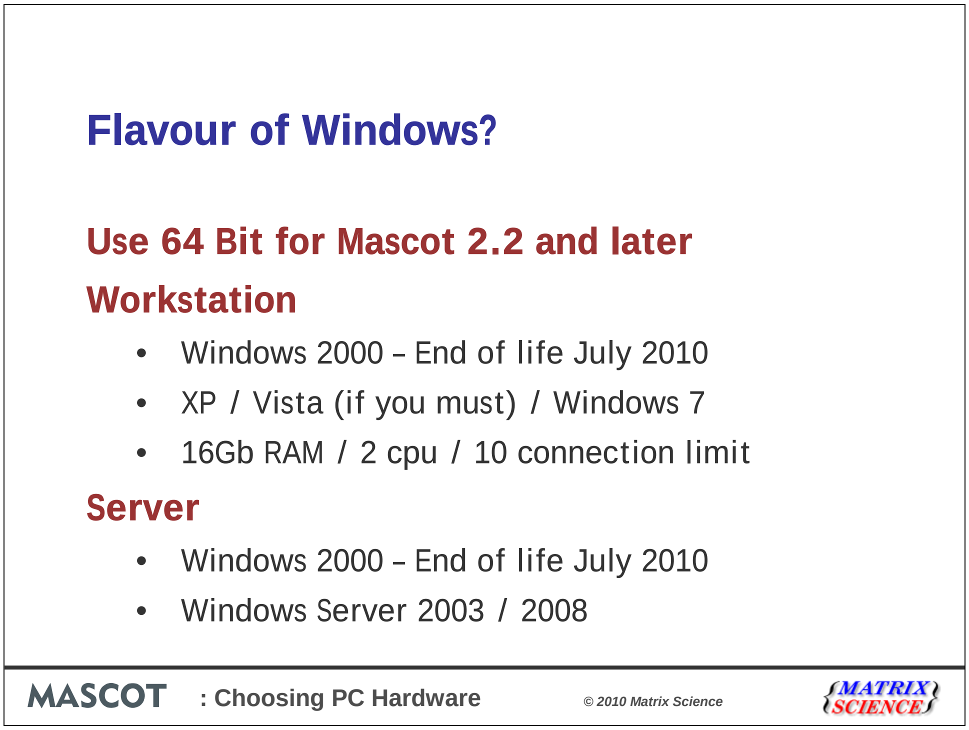
The following are fully supported

- CentOS 4.6 (equivalent to RedHat Enterprise Linux 4.6)
- Debian GNU/Linux 5.0 “Lenny”

However…

- Any distribution that includes a 2.6 kernel, glibc 2.3.4 or later, and libstdc++ 3.4.3 or later should be sufficient

We still support 32 bit Linux, but if you are installing new server, use 64 bit.



Flavour of Windows?

Use 64 Bit for Mascot 2.2 and later

Workstation

- Windows 2000 – End of life July 2010
- XP / Vista (if you must) / Windows 7
- 16Gb RAM / 2 cpu / 10 connection limit

Server

- Windows 2000 – End of life July 2010
- Windows Server 2003 / 2008

MASCOT : Choosing PC Hardware

MATRIX SCIENCE

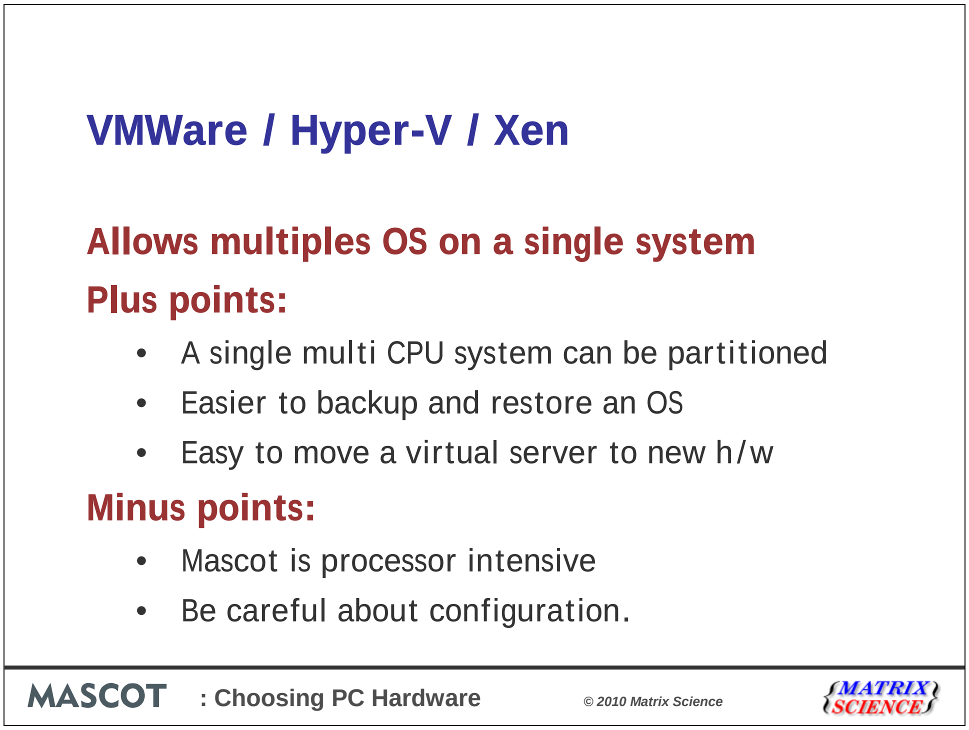
We still support 32 bit Windows, but if you are installing on a new system, it’s much better to use 64 bit.

For an existing setup, there is no advantage in changing from say XP to Windows 7 unless you also need to move from a 32bit to 64bit version.

For the 64 bit workstation versions like XP, Vista and Windows 7 there are some limitations. For example some versions limit you to 16Gb RAM and most to 2 processors. There is also a limit to the number of connections through IIS for some of these, so you may not be able to submit more than 5 searches at a time, which shouldn’t be an issue.

If these limitations persuade you to go to a server edition, this is going to cost more. Either Windows Server 2003 or 2008 are just fine. There is no advantage to purchasing the more expensive editions which can cost over $1000. If you can get the 64 bit 2008 Web edition, then this is probably the cheapest option and has a limit of 4 cpu and 32 Gb RAM.

Mascot 2.3 includes 64 bit binaries, so just make sure you run setup64.exe rather than setup32.exe



VMWare / Hyper-V / Xen

Allows multiples OS on a single system

Plus points:

- A single multi CPU system can be partitioned
- Easier to backup and restore an OS
- Easy to move a virtual server to new h/w

Minus points:

- Mascot is processor intensive
- Be careful about configuration.

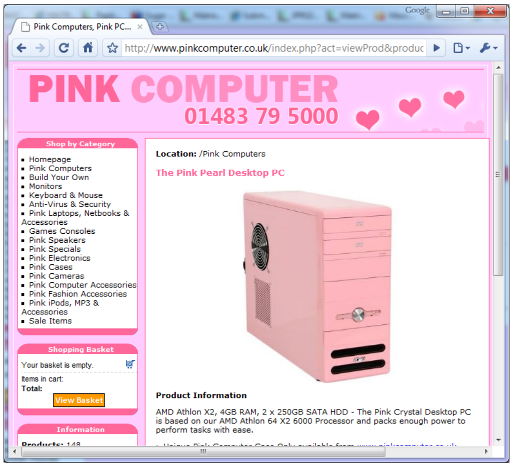
VMWare. Microsoft’s Hyper-V and Xen are virtualisation software products that enable you to install multiple operating systems on a single system. If the system has sufficient resources, then several different Operating systems can run concurrently. For each virtual server that you set up, you can configure how much memory, disk space and processor resource is available. The ‘real’ resources on the system obviously have to be shared between the virtual servers. IT guys love VMWare because it is easy to backup and restore an operating system. Also, if there is a hardware problem, it’s mostly very easy to move the virtual server to new hardware. For many pieces of software that don’t fully use the CPU, these systems can really save money.

Mascot however, is very processor intensive, and the benefits aren’t so obvious. You also need to be careful how you configure the CPUs

How important is the colour of the box?

Not very? Consider

- Rack mounted
- Noise issues
- Space issues
- Power requirements



Pink Computers, Pink PC...

http://www.pinkcomputer.co.uk/index.php?act=viewProd&produc

PINK COMPUTER

01483 79 5000

Shop by Category

- Homepage
- Pink Computers
- Build Your Own
- Monitors
- Keyboard & Mouse
- Anti-Virus & Security
- Pink Laptops, Netbooks & Accessories
- Games Consoles
- Pink Speakers
- Pink Specials
- Pink Electronics
- Pink Cases
- Pink Cameras
- Pink Computer Accessories
- Pink Fashion Accessories
- Pink iPods, MP3 & Accessories
- Sale Items

Shopping Basket

Your basket is empty.

Items in cart:
Total:

View Basket

Information

Products: 148

Location: /Pink Computers

The Pink Pearl Desktop PC

Product Information

AMD Athlon X2, 4GB RAM, 2 x 250GB SATA HDD - The Pink Crystal Desktop PC is based on our AMD Athlon 64 X2 6000 Processor and packs enough power to perform tasks with ease.

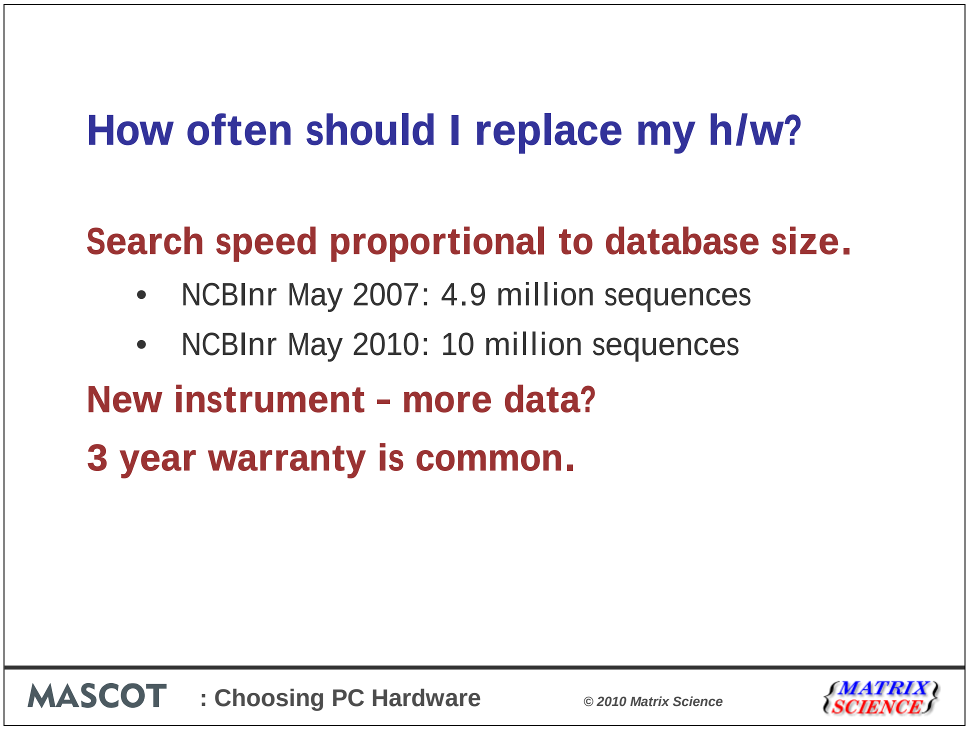
MASCOT : Choosing PC Hardware © 2010 Matrix Science MATRIX SCIENCE

Colour is surprisingly important to some people. If you want pink, these are the girls to go to! They do sell the cases separately if you need more powerful system inside.

On a more serious note, you’ll need to decide on rack mounted or standalone. If you don’t have a rack, best not to get rack mounted!

Noise and space can be a serious issue. We recently had a slim rack mounted unit that made so much noise in our computer room we had to return it. If you are putting a server system in an office, check the noise.

The other thing to watch out for is power requirements. The larger IBM cluster that we sell cannot be plugged into a standard power outlet.



How often should I replace my h/w?

Search speed proportional to database size.

- NCBInr May 2007: 4.9 million sequences
- NCBInr May 2010: 10 million sequences

New instrument – more data?

3 year warranty is common.

MASCOT : Choosing PC Hardware © 2010 Matrix Science MATRIX SCIENCE

You should decide how often you intend to replace your hardware. Search speed is directly proportional to the size of the database, and you can see that NCBInr has almost exactly doubled in the last 3 years. If you want to be running your searches at your defined acceptable speed in three years, then you need a system that runs twice as fast as it needs to at the moment.

It's also quite common to get a new instrument every few years and each new instrument seems capable of generating more data.

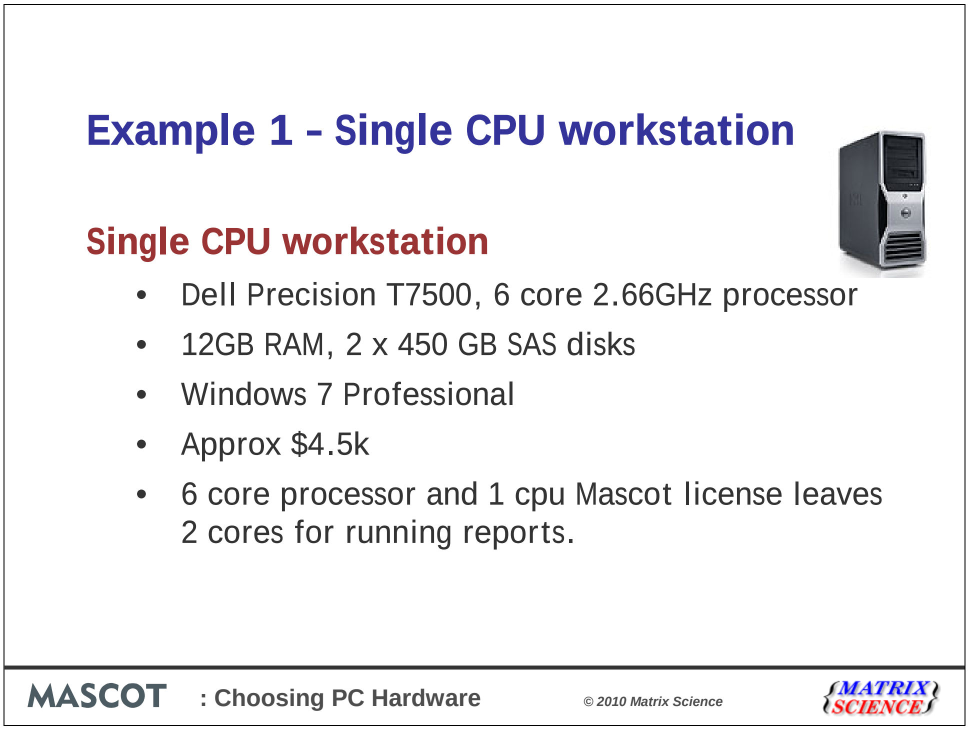
Most systems come with a three year warranty, and this seems like a reasonable period to aim to keep a system. We just had an email from someone saying that there Mascot server was 10 years old. Now that's really get excellent value for money. I'd now like to go through a few example configurations and put some numbers on possible systems.

May 12 2010, sequences=10949590, residues=3727672036

May 12 2009, sequences=8795322, residues=3010574027

May 12 2008, sequences=6507231, residues=2219987828

May 12 2007, sequences=4900652, residues=1692193060



Example 1 - Single CPU workstation

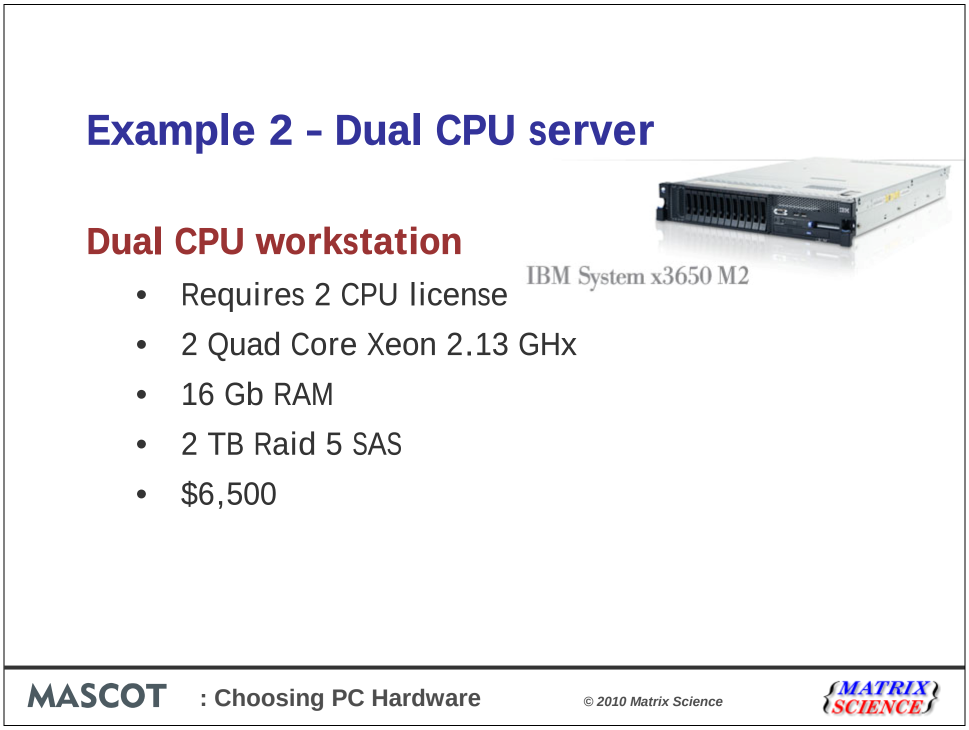
Single CPU workstation

- Dell Precision T7500, 6 core 2.66GHz processor
- 12GB RAM, 2 x 450 GB SAS disks
- Windows 7 Professional
- Approx $4.5k
- 6 core processor and 1 cpu Mascot license leaves 2 cores for running reports.

If you decide that a single CPU Mascot license is sufficient, and if you have approximately $5000 dollars to spend, this top end Dell Workstation may be suitable.

I’ve selected a 6 core processor even though the Mascot license will only use 4 cores because the other two cores can be used for viewing reports which can be quite processor intensive. If you typically don’t view old reports while searches are running, a quad core processor would save you nearly $1000. On the other hand going to a 3.3GHz 6 core processor would cost nearly an additional $1000. I suspect that 6 core processors will come down in price soon. Remember also that you need Mascot 2.3 for the 6 core processors.

I’ve chosen two of the faster SAS disks, but haven’t gone for RAID. Could save a few hundred dollars by going for SATA disks.



Example 2 - Dual CPU server

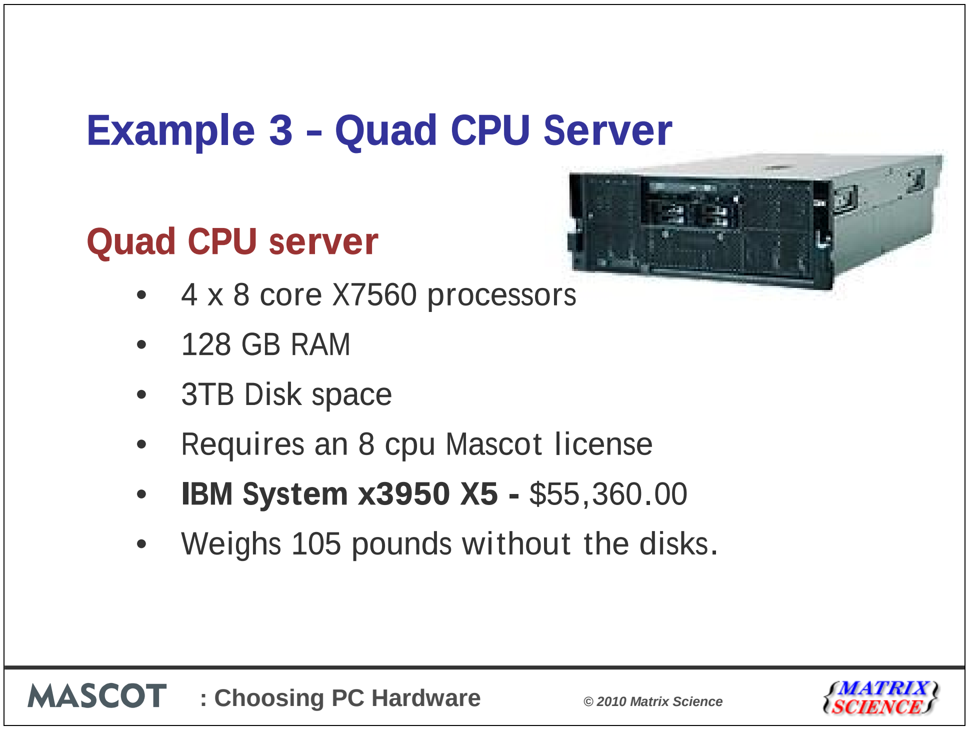
Dual CPU workstation

- Requires 2 CPU license
- 2 Quad Core Xeon 2.13 GHx
- 16 Gb RAM
- 2 TB Raid 5 SAS
- $6,500

IBM System x3650 M2

MASCOT : Choosing PC Hardware © 2010 Matrix Science MATRIX SCIENCE

This is an example of a dual CPU rack mounted server. For a 2 cpu Macsot license, it’s probably cheaper and better at the moment to get a system with 2 quad core processors rather than a system with a single 8 core processor. This advice is obviously going to change over time when 8 core processors start to get cheaper. In the end, it’s mainly a cost issue.



Example 3 - Quad CPU Server

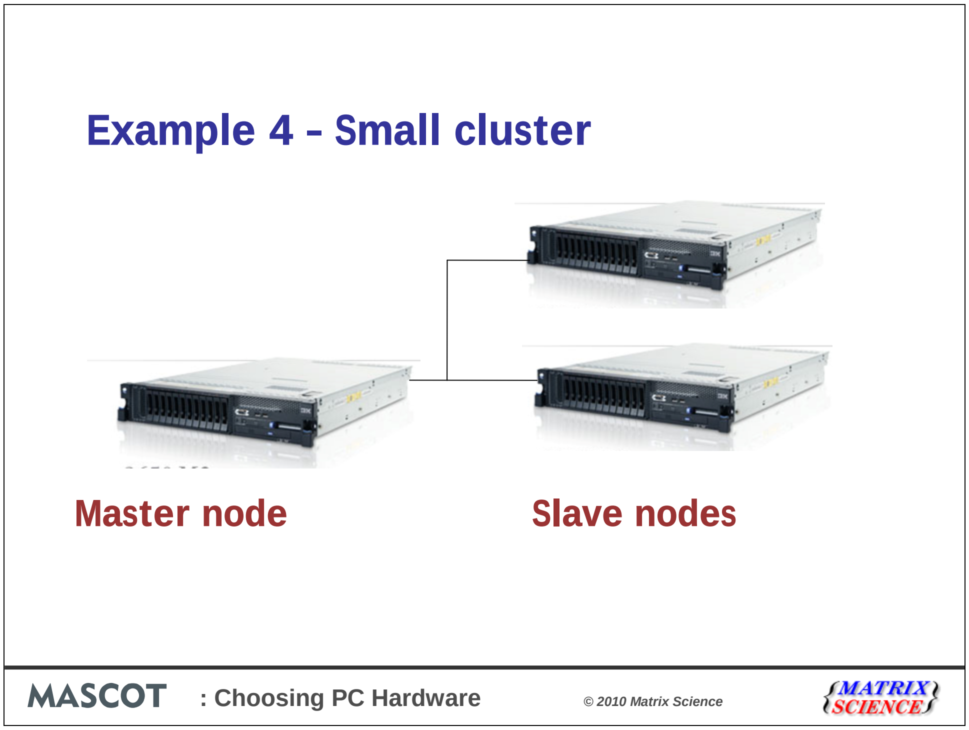
Quad CPU server

- 4 x 8 core X7560 processors
- 128 GB RAM
- 3TB Disk space
- Requires an 8 cpu Mascot license
- **IBM System x3950 X5 -** $55,360.00
- Weighs 105 pounds without the disks.

MASCOT : Choosing PC Hardware © 2010 Matrix Science MATRIX SCIENCE

This is a bit of a ‘beast’ and about the top limit for a single SMP box. 128Gb of RAM might be a little excessive, but I thought I’d fill it up! The level 5 RAID uses 10 x 300Gb disks, but you don’t really get 3TB in this configuration.

If you need 8 quad core processors, it may well be cheaper to go to a small cluster



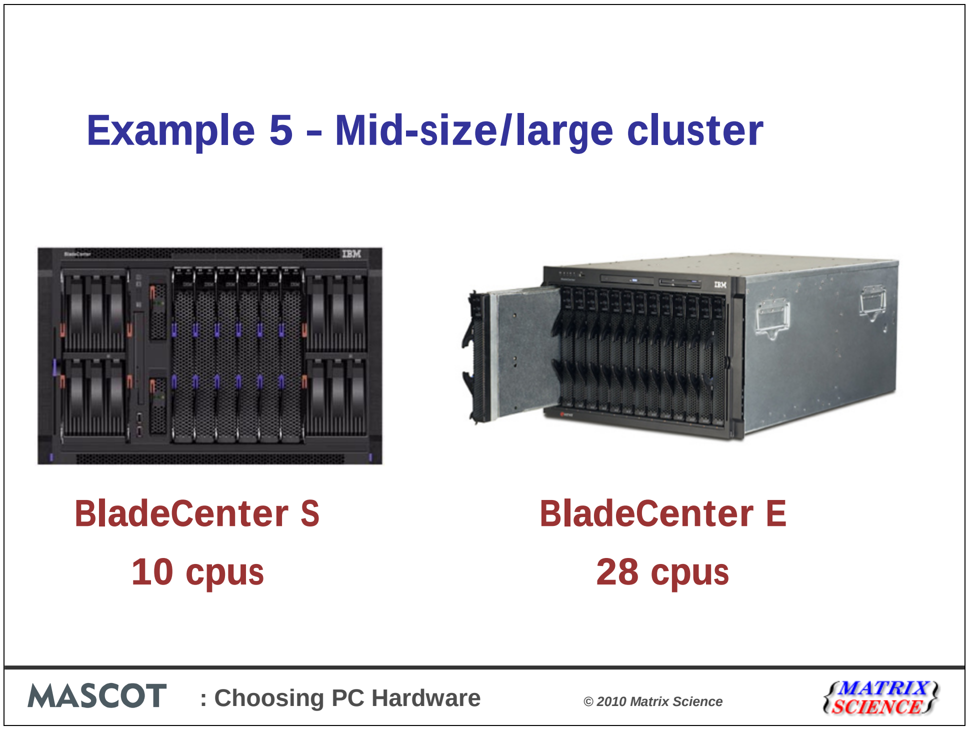
Example 4 - Small cluster

Master node

Slave nodes

What if you buy, say the dual processor system and then find you need more processing power. All is not lost. You would need to purchase two more systems, possibly the same as the original one and upgrade your license to a 4 cpu license. You should then get twice the throughput. Configuration and setup is easy and most people manage this without having to contact us for help.

To get a system for an 8 cpu Mascot license and performance similar to the quad cpu system I’ve just shown, you would need 4 nodes and one master node. At $6500 per system, this would work out at $33k, which is cheaper than the $55k for the ‘beast’. This system is also more easily extended if you need more processing power.



Example 5 - Mid-size/large cluster

BladeCenter S

10 cpus

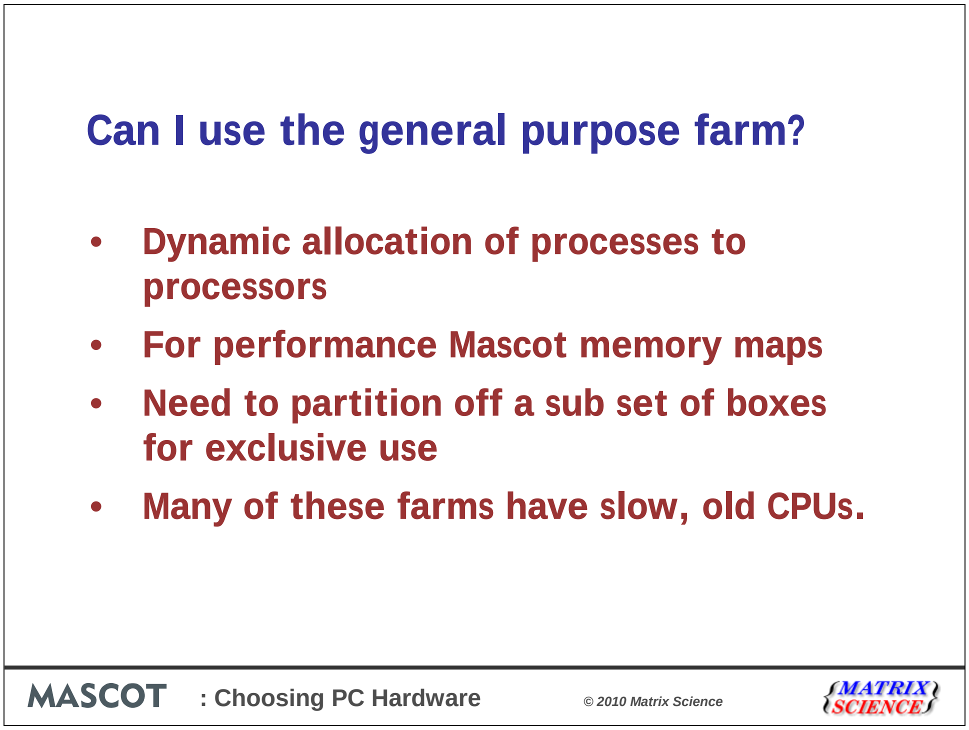
BladeCenter E

28 cpus

MASCOT : Choosing PC Hardware © 2010 Matrix Science

If that’s still not enough, and you are interested in a BladeCenter there are details on our web site of two possible configurations.

The BladeCenter S can utilise standard office power outlets, while the ‘E’ would need to be installed in a computer room.



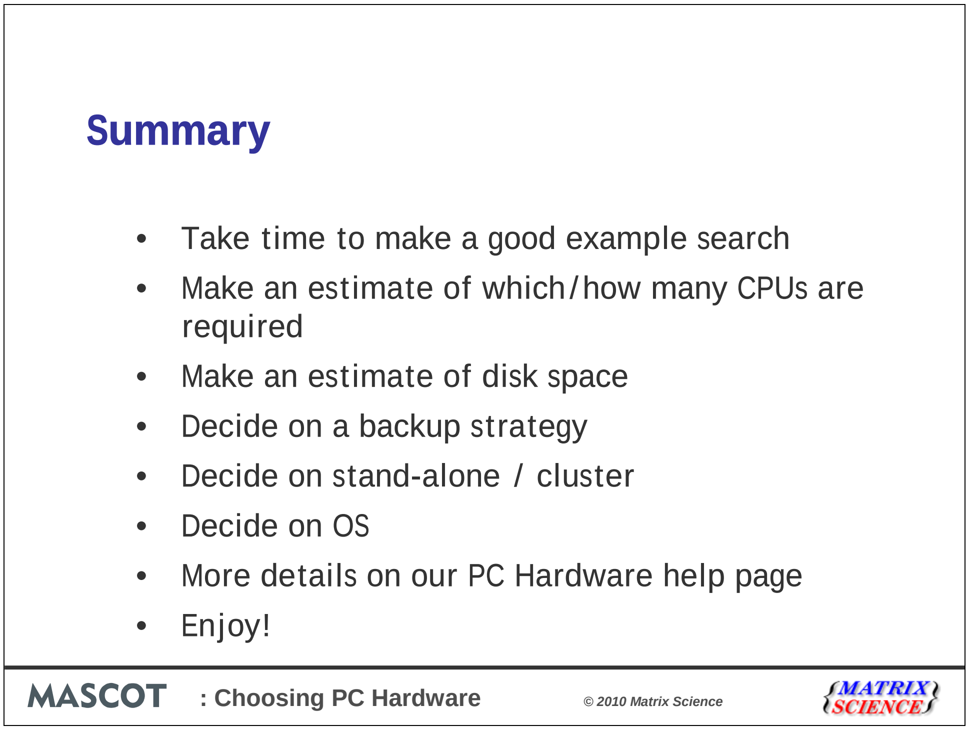
Can I use the general purpose farm?

- Dynamic allocation of processes to processors
- For performance Mascot memory maps
- Need to partition off a sub set of boxes for exclusive use
- Many of these farms have slow, old CPUs.

MASCOT : Choosing PC Hardware © 2010 Matrix Science MATRIX SCIENCE

Many organisations have server farms that are often under utilised and it may seem like a good idea to use existing hardware. However, many of these server farms work by dynamically allocating processes to different processors. This is great for some applications that are processor intensive but don’t require searching of huge databases. One of the keys to speed in a Mascot search is memory mapping the Fasta database files and minimising processor cache misses. Hence, Mascot cluster mode isn't currently compatible with a scheduler that dynamically allocates processors. During a search, each processor is streaming through its allocated segment of the sequence databases, which may total hundreds of Mb. Having to move the databases in and out of memory for each search or each time slice would be a severe overhead.

In the case of a server farm, a sub-set of boxes must be partitioned off, either for Mascot exclusive use or for Mascot "first call". If this is acceptable, you should check the specification of the CPUs on the system because many of these rather idle farms are now very old.



Summary

- Take time to make a good example search
- Make an estimate of which/how many CPUs are required
- Make an estimate of disk space
- Decide on a backup strategy
- Decide on stand-alone / cluster
- Decide on OS
- More details on our PC Hardware help page
- Enjoy!