

Sockets, cores and threads: Choosing PC hardware for Mascot

MASCOT

{MATRIX}
{SCIENCE}

Sockets, cores and threads: Choosing PC hardware for Mascot

Multiple choice 'quiz' - it's not so easy. . .

Desktop, server, Workstation or laptop?

Processors (type, speed, HT, cores)

Random Access Memory (RAM)

Disks (type, size)

OS (Windows, Linux, 64bit vs 32 bit)

Cluster mode

MASCOT Choosing PC Hardware for Mascot

© 2006 Matrix Science



This session is aimed at people who have some knowledge of computers, but quite probably don't really understand what sockets, cores and threads are. My aim is that at the end of this session, you will have enough information to enable you to select a suitable PC for Mascot that will give you best value for money. For those of you who have to go through an IT group for purchasing hardware, I hope that you will have learnt enough to talk the same language as your IT folks.

We'll start with a simple multiple choice quiz - I'm going to show you a selection of systems and ask you to make the best choice for a single CPU Mascot license. I'll give the answer at the end of session, although hopefully you will all know the answer by then.

So, we will cover what type of computer is best - a desktop, server, Workstation or laptop.

A big part of the session will be about processors

I'll also cover, RAM, Disks, choice of OS and, briefly, cluster mode.

Quiz! You have a 1 cpu Mascot license Select the most suitable computer:

1. 64bit UltraSparc (Sun) Laptop



- 64bit UltraSparc IIIi
1.2GHz processor.
- 2Gb RAM
- 60 GB IDE disk
- \$6445.

So, you have just raised a purchase order for a 1 cpu Mascot license, you have your credit card ready and want to buy a computer for Mascot. I'm going to give you 5 choices of computer - your task is to decide which is best.

Now, to make things simple, just imagine that it isn't your credit card - so you don't need to worry about the cost. You just want to run Mascot searches as fast as possible and have enough disk space to last a year. The type of search that you want to run fast is about 500 spectra from a typical lcms-ms run, searching using Trypsin against NCBIr with a couple of variable modifications.

First on the block is a bit of an odd ball. It's a top of the range laptop, 64 bit ultra sparc processor, 2Gb RAM and quite a reasonable size disk for a laptop. Bit of a bargain really at six and a half thousand dollars

Choosing h/w for Mascot - Quiz!

2. 64bit, IBM Express series 260



- 4x 3.66GHz Xeon.
- 52Gb RAM
- 12 x 300 GB SCSI disk
- Warranty etc.
- Die-cast alum. wheels
- \$69,920.75

Subtotal (update) \$69,920.75 *

[Check out](#)

[Save cart for later](#)

[Call me now](#) [Chat online](#)

MASCOT Choosing PC Hardware for Mascot

© 2006 Matrix Science

MATRIX
SCIENCE

Second on the list is a 64 bit system from IBM.

It's rather a large beast, but it comes on a very nice set of wheels.

It has 4 3.66 GHz processors, a massive 52Gb of RAM and 12 very fast 300 GB SCSI disks configured in a top notch raid system. This all comes with the best warranty (2 hour service) that money can buy. And you'll need quite a bit of cash as you aren't going to get much change from 70 thousand dollars.

Oh, and I must confess that I haven't actually tried this model, but if someone has got a credit card handy, I'll order one right away to do a proper test...

Choosing h/w for Mascot - Quiz!

3. Dell PowerEdge 830



- 1x 3.4 GHz Dual Core Pentium D
- 4Gb RAM
- 2 x 300 GB SCSI disk
- RedHat Linux ES4
- No monitor
- \$2,895.75

The third system, from Dell is a Power Edge 830.

This comes with a single processor, slightly slower than the processors on the previous system, and a mere 4Gb RAM. It also has a couple of 300Gb SCSI disks, Linux but no monitor. Rather more reasonably priced at just less than 3k

It can't be upgraded to have more than one processor.

Choosing h/w for Mascot - Quiz!

4. IBM IntelliStation A Pro



- 1x 2.4 GHz Dual Core Opteron
- 4Gb RAM
- 650 GB on 2 SATA disks
- Windows XP
- No monitor
- Can be upgraded to 2 cpu
- \$6,083.00

MASCOT Choosing PC Hardware for Mascot

© 2006 Matrix Science



Our fourth system is an IBM IntelliStation A Pro, with a single dual core Opteron processor - even fewer GHz than the previous model from Dell.

4Gb RAM as the previous one, and slightly more disk space except these ones are SATA whatever that might be.

We've got Windows XP Professional on this, also no monitor.

This one is a 2 cpu system, but we only have 1 cpu for this price.

It's quite pricey at 6000 dollars, but it's not your money is it?

Choosing h/w for Mascot - Quiz!

5. Dell Precision 670



- 1x 2.8 GHz Dual Core Xeon
- 4Gb RAM
- 2 x 300 GB SCSI disks
- Windows XP Pro
- No monitor
- Can be upgraded to 2 cpu
- \$4,725.00

MASCOT Choosing PC Hardware for Mascot

© 2006 Matrix Science



Our fifth and final system is another DELL - a Dell Precision 670 workstation. It has one dual core Xeon processor - slightly more GHz than the previous model from IBM.

4Gb RAM as the previous one, and 600 Gb SCSI disks.

We've got Windows XP Professional on this, also no monitor.

This one is a 2 cpu system, but we only have 1 cpu for this price.

It's cheaper than the dual processor system from IBM at just under 5000 dollars

If you are feeling brave, you might like to make a guess at which of these four systems is going to run a Mascot search fastest. I'm not going to give you the answer now, but at the end of this session we will return to a summary of these systems and the answer. By that time, I hope that you will all be able to work out the answer.

Desktop, workstation, server, laptop?

Desktop - cheaper

Laptop - uses less power

Workstations

- can have slightly faster (or dual) processors
- expensive graphics system and SCSI drives (\$\$\$)

Servers

- multiple processors
- often have room and power for more disks (\$\$\$)
- redundant power supplies (\$\$\$)
- server management systems (\$\$\$).

MASCOT

Choosing PC Hardware for Mascot

© 2006 Matrix Science

MATRIX
SCIENCE

If you are looking on the IBM or DELL web site for a PC, the first choice you have to make is whether you want a desktop, a laptop, a workstation or a server. These terms are primarily "marketing" terms, but there are some general differences.

Desktops are mostly cheaper and smaller than workstations.

Laptops are now not totally unsuited to running Mascot. They can have at least 2Gb RAM, and the processors are often pretty fast.

Workstations are often touted as top of the range desktops - indeed they sometimes, but not always have slightly faster processors. In some cases, there will be a dual processor option that won't be available on a desktop. Most importantly, they will have a classy graphics system - which is of course totally useless for Mascot and most mass spec software. They will also often have SCSI drives, that are also not really required for Mascot. So, we can see that if you are only looking for a single processor system, it's best to avoid the workstations

Servers can also have dual processors. They tend to be larger with plenty of room and power for multiple hard disks. They may also have two or more power supplies in case one fails - called redundant power supplies. Finally, they often have server management software that will only be of real use to you if you have the time and skills to learn how to use it. The server management software will not make Mascot run faster.

In summary, it is normally best to choose a desktop if you want a single CPU system. If you want a dual processor system, then you will need to look through the server and workstation options for the best system

Processors - terminology

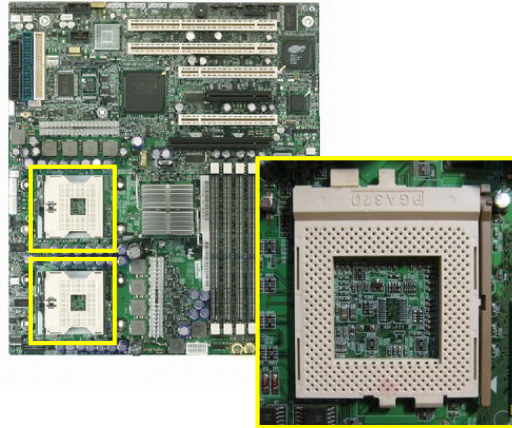
“Mascot uses socket based licensing. A single cpu Mascot license on a PC with one dual core processor and hyper-threading enabled in the bios can fully utilise the processor if configured to use 4 threads”

(if you understood that the next few minutes may be boring...)

Read from slide

Mascot uses 'Socket Based' licensing

What's a socket?



Dual sockets.. but
only one in use.

MASCOT Choosing PC Hardware for Mascot

© 2006 Matrix Science

MATRIX
SCIENCE

I'd like to tackle the phrase "Socket Based licensing" first of all.

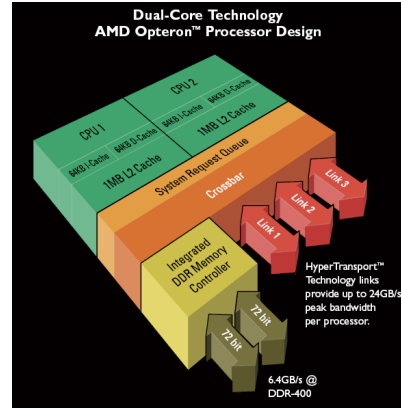
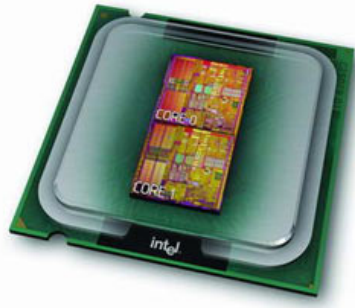
In the UK, we call these 'Sockets' - although I think you often call these outlets over this side of the pond. This type of socket is a little simpler than one on a computer mother board - although you really shouldn't try and stick your fingers in either.

If we look closely at this computer motherboard, we can see that this is a dual processor mother board, because it has two sockets. And this is just a close-up of a typical socket for a processor.

Clearly, when we refer to 'socket based licensing', these sockets are pretty useless unless something is plugged into them - what we should really be saying is "socket with something plugged into it based licensing" but that is a bit unwieldy. The only people I know of that charge for the case when you don't even have something plugged in are Oracle.

So, what we are saying by 'Socket based licensing' is that we don't care what sort of processor you put in this socket - it just counts as one Mascot license. If the processor that you put into the socket can pretend it is really two processors, then that is just fine -we still count it as one.

... Multiple cores?



MASCOT Choosing PC Hardware for Mascot

© 2006 Matrix Science



So why do the manufacturers want to make a single physical processor act like it is two processors? Mainly because they have just about reached the limit of how fast a processor can go, so the only way to speed things up further is to put more of the same into the one package.

On a dual core Intel processor, nearly all of the same circuitry is repeated twice.

If you give a single core processor 2 sums, it will do one followed by the other. If you give a dual core processor the same 2 sums, it can do them at the same time - so you get the results twice as fast.

Almost.

On this diagram here for a dual core AMD processor, we can see two quite separate central processing units, 2 quite separate level 1 caches (that's really fast local memory), but finally, they both share the same main memory. Obviously, if they are both competing for this, then everything will run more slowly.

. . . Multiple cores

- **How good is dual core?**
 - Can be twice as fast as single core, but
 - Depends on the type of search
- **Sparc T1/T2 8 core systems**
- **Future licensing plans**
 - For Intel/AMD we will limit to max of 2 cores per license.

A dual core system can run a search in half the time that a similar single core processor can run the same search. However, this is only true for a computationally intensive search - that is a no enzyme search or a search with several modifications. For other searches, remember that it is streaming through the database and both cores will try and access different parts of the main memory, so there will be some slow down.

Currently, AMD and Intel only offer dual core processors, although they both expect to release packages with a higher numbers of cores by the end of next year.

However, Sun have released T1 and T2 processors which have up to 8 cores. We've not tested these yet, but our current estimate is that an 8 core T1 is about the same as a dual core Opteron.

A significant amount of our income as a company comes from selling multiple processor licenses. If Intel and AMD were to release an 8 core processors, and we were to only require a single cpu license, then we would probably have to just increase the entry price for Mascot. This doesn't seem fair or reasonable to people who do only need a single processor, so we have decided to cap the number of cores per license to 2 for Intel and AMD processors.

. . . Hyper-threading

- Intel only (HTT on AMD is different)
- Improves performance by 10%
- Works by duplicating some sections of the processor - but not the main execution resources.
- A single cpu with HT enabled appears to the operating system as 2 cpus.

Which brings us onto our next topic - 'Hyper-threading'.

This was first made available by Intel on their processors a couple of years ago, and suffered from 'hype' - their claims were over-exaggerated.

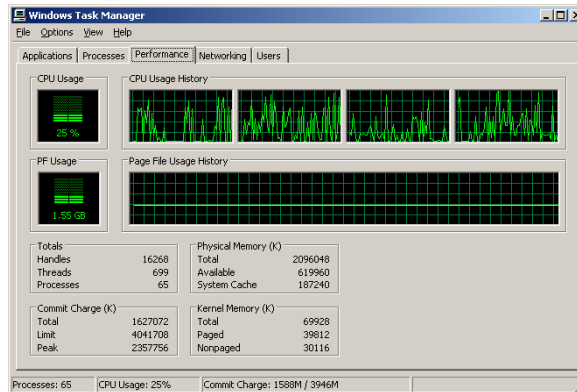
We only see a 10 to 15% increase in performance with hyper-threading enabled - and you will need a later version of Windows or Linux to see any benefit at all.

Hyper-threading works by duplicating some sections of the processor - but not the main execution resources. Think of it as poor man's version of dual core.

One thing to be aware of is that a single cpu with HT enabled appears to the operating system as 2 cpus

. . . Hyper-threading

- A dual CPU Xeon system with HT enabled:



MASCOT Choosing PC Hardware for Mascot

© 2006 Matrix Science



In your Windows task manager on the performance tab, you can see 4 small graphs - each one is the usage for a separate ‘processor’.

This is from a dual processor (dual socket) system -so each processor appears as two because each one has hyper threading.

On a dual processor system with each processor having dual cores and hyper-threading, we would expect to see 8 ‘cpus’:

. . . Hyper-threading and dual core

- **Socket 1**

- Core 1 1
 - Hyper threaded 2
- Core 2 3
 - Hyper threaded 4

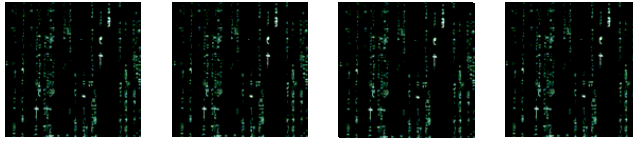
- **Socket 2**

- Core 1 5
 - Hyper threaded 6
- Core 2 7
 - Hyper threaded 8

Like this.

So, your Mascot status page should report that you have 2 physical processors with cpu numbers 1 to 8

. . . Multiple threads?



A single threaded application doesn't benefit from multiple CPUs.

Mascot is multi-threaded

MASCOT Choosing PC Hardware for Mascot

© 2006 Matrix Science

MATRIX
SCIENCE

When I described multiple cores, I said “If you give a dual core processor the same 2 sums, it can do them at the same time - so you get the results twice as fast”

Unfortunately, most software is just not capable of giving a dual core processor 2 sums at once. To do this, the software needs to be designed to use multiple ‘threads’. What this means in simple terms is that the program says: Start a thread to do X and let me know when you are done; start a thread to do Y and let me know when you are done. The program can then just wait until both X and Y are finished. On a single core, single processor, this program would still work, but it would run very slightly more slowly than if it just had one thread because both threads would be sharing the same cpu.

In the Mascot Database Maintenance utility, you are required to enter the number of threads to be used for each database.

Making an old program support multiple and this is often a big task for software developers which is why a lot of software won't use hyper-threading or dual cores. However, Mascot has supported multiple threads from day one.

Processors - AMD or Intel

- **AMD currently (May 2006) have a slight edge over Intel - especially for dual CPU**
- **Cannot compare AMD GHz and Intel GHz.**
 - For example, 2.4GHz Opteron == 3.45GHz Intel
- **Cache size can be important for some searches**
- **Faster is always better.**

For a dual CPU system in particular, AMD currently have a slight edge over Intel in terms of performance and in particular power consumption.

However, comparison is not easy. For Mascot, a 2.4GHz Opteron will perform similarly to a 3.45GHz Intel processor.

Similarly, the comparison between mobile Intel 'M' processors and standard processors is non-trivial.

The other choice for processors is cache size. Mascot has been designed to work well with a 1Mb cache - so you may not see much benefit from a 2Mb cache. Avoid Intel Celeron processors with only 256k cache

The faster the processor, the better - except of course if your search is disk bound.

Intel processors

Name	GHz	#CPUs	Dual core	64 bit	HT	L2 cache	FSB MHz
Extreme Edition	3.46	1	Y	Y	Y	2Mb	1066
Pentium D	3.4	1	Y	Y	N	2Mb	800
Pentium 4 Extreme	3.73	1	N	Y	Y	2Mb	1066
Pentium 4 with HT	3.8	1	N	Y	Y	1Mb	800
Xeon 5000	3.8	2	N	Y	Y	2Mb	800
Xeon 5000 dual core	2.8	2	Y	Y	Y	2Mb	800
Xeon 7000	3.0	8	Y	Y	Y	2Mb	800

MASCOT Choosing PC Hardware for Mascot

© 2006 Matrix Science



This is a simple table of current processors from Intel. Of course, it is already out of date. This shows the fastest processors available for each type.

The obvious thing to point out here is that if you want dual core processors for a system with more than one socket, then they are going to be slower. They have now got up to 3.0GHz with the Xeon 7000 processors, but these are really expensive.

The screenshot shows a Mozilla Firefox browser window displaying an article on The Register website. The article title is "Intel's Paxville: too slow, too hot, too dumb". The author is Ashlee Vance, and it was published on Thursday, 20th October 2005 at 17:58 GMT. The article text begins with "When we nicknamed Intel's new dual-core Xeon processor we didn't know how prophetic this would be. A fresh start comparing the Xeon with AMD's dual-core Opteron convinced us the product to be a power hungry demon that doesn't perform well." A photograph of Pat Gelsinger, Intel's server chip chief, is shown with a large, glowing, fiery explosion behind his head. The caption below the photo reads: "Intel server chip chief Pat Gelsinger. Behind him, Paxville is caught during a critical phase of its boot sequence." The website footer includes "MASCOT Choosing PC Hardware for Mascot", "© 2006 Matrix Science", and the "MATRIX SCIENCE" logo.

The problem with Intel dual core Xeons has been overheating as this article last year in our favourite on-line magazine shows. The Register was not particularly impressed with the dual core Xeons and further down the article shows a picture of Intel server chip chief Pat Gelsinger with the caption:

Behind him, Paxville (the code name for the dual core Xeon) is caught during a critical phase of its boot sequence.

One thing to bear in mind is the huge power consumption of these processors. There's also the noise from the fans, extra costs, both financial and environmental of cooling the room.

Unfortunately, Dell have for a long time refused to supply systems with AMD processors. I believe that this is finally about to change although I can't see any official announcement yet.

AMD Opteron

Name	GHz	#CPUs	Dual core	64 bit	HT	L2 cache	FSB MHz
Opteron dual core - e.g. 880	2.4	8	Y	Y	N/a	1Mb	2400
Opteron single core - e.g. 852	2.8	8	Y	Y	N/a	1Mb	2800
Athalon 64 X2	2.4	1	Y	Y	N/a	1Mb	2400

MASCOT Choosing PC Hardware for Mascot

© 2006 Matrix Science



Here's the details for the current Opteron processors.

Again, you will notice that dual core processors don't run quite as fast as their single core counterparts. However, the power consumption for the dual core processors is significantly less than the Intel Processors

2.8GHz Xeon, Idle: 393 Full power: 428

Opteron 285: Idle: 235 Full power: 301

Processors - Intel Itanium

- Matrix Science does not support Itanium processors
- A test port to native Itanium processors gave comparatively poor performance compared with Pentium processors.

As on slide

Processors - how many?

Scale linearly

Embarrassingly parallel

MASCOT Choosing PC Hardware for Mascot

© 2006 Matrix Science



If you have a single processor Mascot license and need searches to run twice as fast, the obvious answer is to add a second processor. Searches will then run twice as fast because Mascot is embarrassingly parallel.

Random Access Memory (RAM)

- **Memory at least 100x faster than a hard disk.**
- **'Short' searches much faster if databases in RAM**
- **Minimum 2Gb, recommend 4Gb**
- **Cost is ~ \$250 per Gb**
- **Often considerably cheaper from Crucial**
- **Always use www.memtest86.com**

MASCOT Choosing PC Hardware for Mascot

© 2006 Matrix Science



In a Mascot search against the NCBI non-redundant database, the search engine needs to read through 1.6Gb of sequence data. On a very recent SATA disk, it takes about 25 seconds to stream through this data. It takes a fraction of a second to stream through this in memory, so each search would take nearly 25 seconds longer from disk than if the database was held in memory.

A far worse situation is if two searches are running at the same time, because the disk heads need to keep jumping from one spot to another. In this case, it takes an amazing 7 minutes to stream through the database - quite a long time to wait for a peptide mass fingerprint that should maybe take 30 seconds.

We recommend a bare minimum of 2Gb RAM - this is really a small cost compared with the cost of the instrument or the Mascot software. It is now possible to put more than 2Gb in a PC, so we would recommend 4Gb now.

Memory is often more expensive from the computer manufacturers - can be half the price if you are prepared to install it yourself.

We have seen a handful of problems over the last few years that are a direct result of faulty RAM. These problems are often really hard to diagnose. Our strong recommendation, based on bitter experience is to always test a new system using the excellent free software from this address. The memory testing software from Dell or IBM is not as thorough as this. If you do detect faulty memory, then the only safe thing is to wipe the PC and start again - remember any installed programs or the OS itself may be corrupt.

Disks

RAID (Redundant array of independent disks)

- Raid can be faster and fault tolerant

SCSI - 300Gb, 10000 rpm - \$750

IDE - 300Gb, 7200 rpm - \$100

SATA - 250Gb, 7200 rpm - \$110

IDE/ATA is slow and old...

Disks space requirements for searches:

There are so many different types of disk, and within each broad type of disk, there are many options.

One of the first choices is whether to use RAID or just independent disks. Disks are inherently slow because they are mechanical devices. Just like with processors, one way of getting better throughput is to have lots of them. So, if you need 1Gb of data, it will be faster to use 4 disks and get a quarter of the data from each one - as long as you have a suitable piece of hardware, called a RAID controller to do this. Software RAID is not sensible for Mascot because the processors are already very busy. A decent RAID controller is going to cost several hundred dollars. More memory may be a cheaper and better alternative. The other advantage is that with some types of RAID, if one disk fails then all is not lost and the disk can just be replaced without any loss of data - that's the 'redundant' bit.

Beyond that there are several different types of disk.

Serial ATA - Up to 300 Mb/s

IDE (Parallel ATA) - up to 100Mb/s

I'd strongly recommend SATA disks as they are fast and cheap. Also, they can be used in RAID if required.

Disk - space requirement for fasta files

Database	Total size of files (GB)	Max Disk Space (GB)
Swiss-Prot	1	3
NCBIInr	3	9
MSDB	5	15
EST_human	7	21
EST_mouse	4	12
EST_others	21	63

For the sequence databases, you will need to maintain free disk space of the order of 3 times the largest database. This is because, during a database update, there may be the current FASTA file, reference file and its associated compressed files plus the equivalent for the incoming database. Mascot also keeps a copy of one previous database. Current (January 2006) disk requirements for the common databases are shown in this table.

It would not be unreasonable to allow 120GB for databases, and this could grow to 200GB within 2 years. However, it is unusual to require all three EST databases.

Disk space requirements for searches

PMF - 50k

MS-MS - 1 spectrum, 23k

MS-MS - 400 spectra, 2Mb

MS-MS - 40,000 spectra, 200Mb

**Assume 20 x 4000 spectra per day, 200 days
per annum. Need 80Gb per year**

Mascot keeps all the search results on the disk. A typical peptide mass fingerprint search might take 50k.

(read from slide)

OS - Windows or Linux

Performance is almost identical

Use what you have most experience with

- Do you understand ‘tar’, ‘chown’ ‘chmod’?
No - use windows
Yes - use either Windows or Linux

MASCOT Choosing PC Hardware for Mascot

© 2006 Matrix Science



Despite what some over zealous advocates of Linux may say, a Mascot search will at about the same speed on Windows or Linux. We know, because we have several systems that run both OS and the difference is never more than about 1%.

The key question is “What do you have most experience with”

This is our simple test:

If you can’t answer this, please don’t decide that it’s a good time to learn and I’ll start with Mascot. We don’t have time to teach Linux.

Windows Versions

Operating system	Max CPU	Max RAM
2000 Professional	2	4
2000 Server	4	4
2000 Advanced Server	8	8
2000 Data Center	32	32
XP Professional	2	4
2003 Web Edition	2	2
2003 Standard	4	4
2003 Enterprise Edition	8	32
2003 Data Center Edition	32	64

MASCOT Choosing PC Hardware for Mascot

© 2006 Matrix Science



There are a bewildering number of choices of versions of Windows.

We will cease all support for NT4 in 2007, so that isn't on my list here. Apart from cost, the major consideration is the number of CPUs that the OS can support and the maximum amount of RAM.

For the number of CPUs, note that Microsoft have sensibly decided to follow our lead and decided use socket based licensing!

You can see from this chart that XP Professional is probably the best bet for most systems.

We would recommend the 2003 Web edition - it really is rather cheap - except for the limit of 2Gb RAM

We don't currently support 64 bit Windows - there will be support for this in Mascot 2.2

OS - Linux

With multiple cores or HT use 2.6 kernel

64 bit supported with Mascot 2.1.03

We build on RH8

We test on EL3, EL4, Debian 3.1 and FC4/5

Use your 'normal' distribution

MASCOT Choosing PC Hardware for Mascot

© 2006 Matrix Science



With Linux, the main requirement is to use something with a 2.6 kernel if your processors have multiple cores or hyper-threading.

64 bit support is included with Mascot 2.1.03 or later. I should point out however, that we are still using 32 bit binaries for the applications - it is only the parser library that is supplied as 64 bit code.

Mascot 2.2 will, almost certainly come with 32bit and 64bit executables. Our tests show a slight performance increase for 64 bit mode - a little less than 10%. However, I should point out that this may well be due to compiler differences rather than anything specific about 64 bit in itself.

We currently build on RH8 and this is the oldest version of Linux that we support. However, we also test on a number of different distributions.

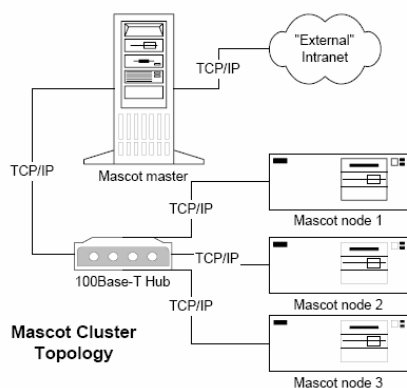
Lifetime of a PC?

- As database sizes increase, searches start to run more slowly
- Don't upgrade unless you need faster throughput!
- A system purchased today will be roughly twice as fast as one purchased 2 to 3 years ago
- 3 year warranty may be good value.

Read from slide.

Still not fast enough?

- 'Cluster mode'
- License cost decreases with number of cpus



MASCOT Choosing PC Hardware for Mascot

© 2006 Matrix Science



Suppose you already have a dual processor Mascot license, you've followed all the advice in this session and have pretty much the best PC you can buy, but, you've just purchased 3 new fast tof-tof instruments and your searches need to run at least 4 times faster to keep up. What should you do.

One choice is to purchase an 8 cpu system. However, this will be very, very expensive. The other choice is to run Mascot in cluster mode.

This is available for Windows or Linux, or indeed any of the other OS that we support. Cluster mode is available for 4 cpus or more.

The basic setup is to use a 'master' computer which could be your original 2 cpu system - in fact that would be the least amount of work. The only extra requirement for the master is that it has 2 ethernet connections, one to your lab and the other to a new switch or hub that will cost about 50 bucks. You then have any number of slave computers. These need to be a similar specification to the master.

Software installation is easy - you don't need to install anything on the nodes. Just edit a configuration file on the master to say where the nodes are, and when you start the Mascot service it will install all the required software and databases on the nodes. When you run a search, just one part of the database on each node gets searched. It's as easy as that.

If you want 10 or more nodes, it may make sense to purchase an IBM BladeCenter, pre-configured from Matrix Science.

Quiz - the answer!

1	Sparc Laptop	1.2GHz	\$6445	Single	Search duration in seconds
2	IBM 620	3.7GHz	\$69,920	Single	
3	Dell 830	3.4GHz	\$2895	Dual	
4	IBM Intellistation	2.4GHz	\$6080	Dual	
5	Dell Precision	2.8GHz	\$4725	Dual	

MASCOT Choosing PC Hardware for Mascot

© 2006 Matrix Science



So, back to our quiz!

These were the five systems. Remember that you only have a 1 cpu Mascot license. Also, remember that we are searching against NCBI_{nr} and this will all fit comfortably in the 4Gb RAM - except for on the laptop. Since laptops have slower hard disks, we will have to rule this one out straight away. Also, the more astute will have realised that this talk was supposed to be about PC hardware and this isn't strictly PC hardware..

If we were to go by price or processor speed then we would have to select this fancy IBM box!

Your last chance to make up your mind before I reveal the final column which has the timings of our standard benchmark test search. Ready!

So, the fastest by a tiny amount is the cheapest box. However, I wouldn't rule out this dual Opteron box from IBM. Although it costs almost double, you can just add another processor to get twice the throughput and that's something that you can't do in the Dell 830. Also, it's worth pointing out that IBM will have a single cpu system very similar to the the Dell 830.

Summary - Help page on web site

http://www.matrixscience.com/help/pc_specs.html

Or, on Matrix Science public web site:

- Help menu
 - Miscellaneous
 - PC hardware for Mascot Server

MASCOT Choosing PC Hardware for Mascot

© 2006 Matrix Science



And finally, if you've not managed to remember all this, there is a help page with most of this information on our web site.

Thank-you for your attention.