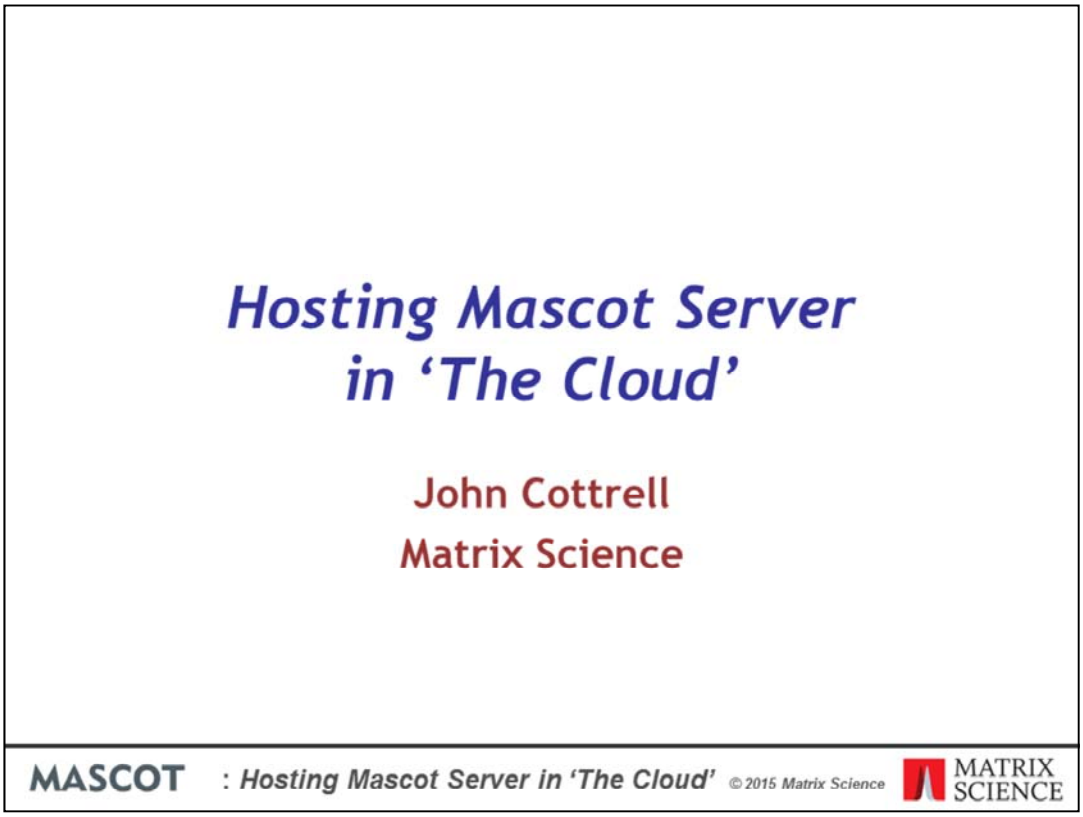
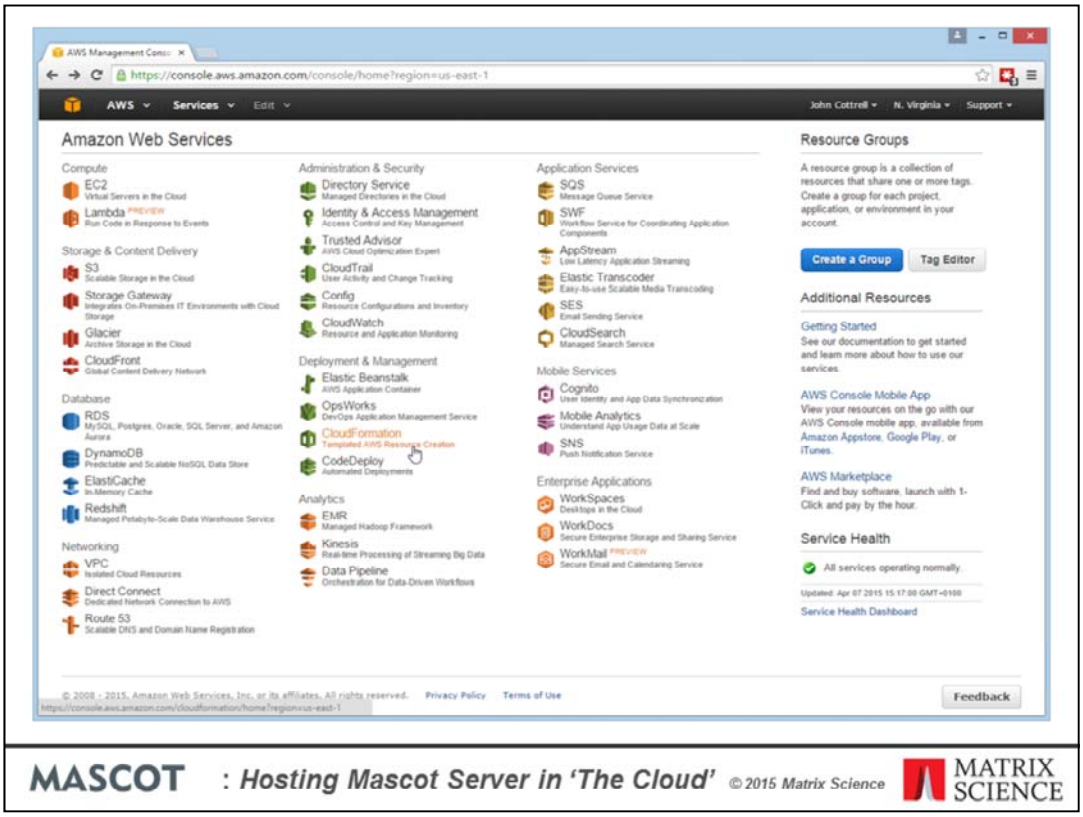
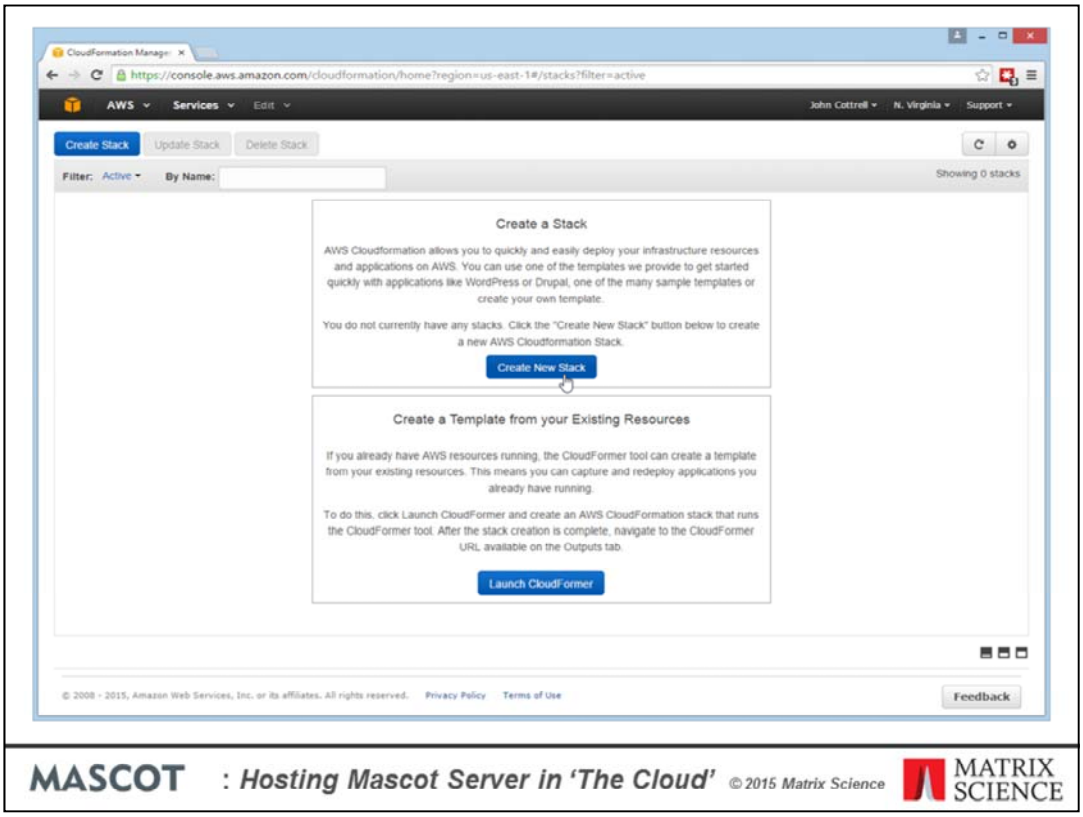


Cloud computing is a marketing term that means different things to different people. In this presentation, we look at the pros and cons of using Amazon Web Services rather than setting up a physical server on your LAN

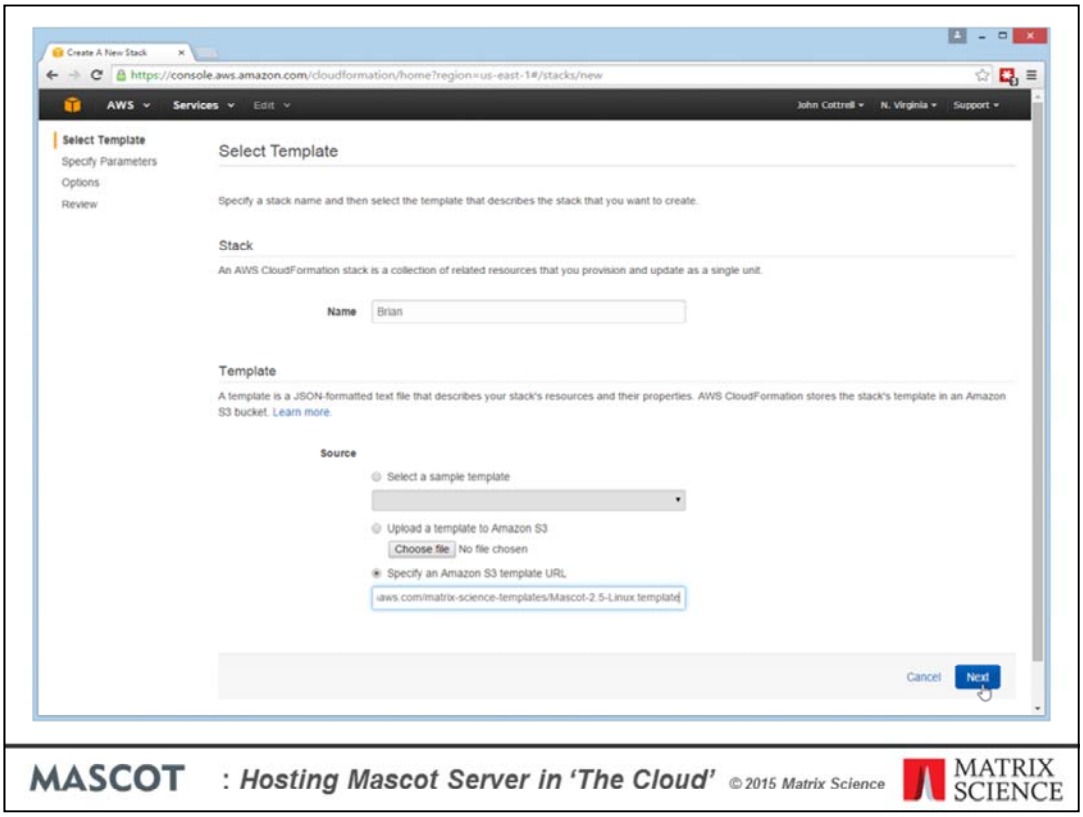


Amazon are not the only company offering such facilities, but they have a well established and comprehensive set of services. At first sight, the sheer number of different services may seem off-putting, and there is a learning curve to some. What I hope to show you is how simple it is to use CloudFormation to create a Mascot Server on Amazon's cloud; much faster and easier than setting up a physical server.

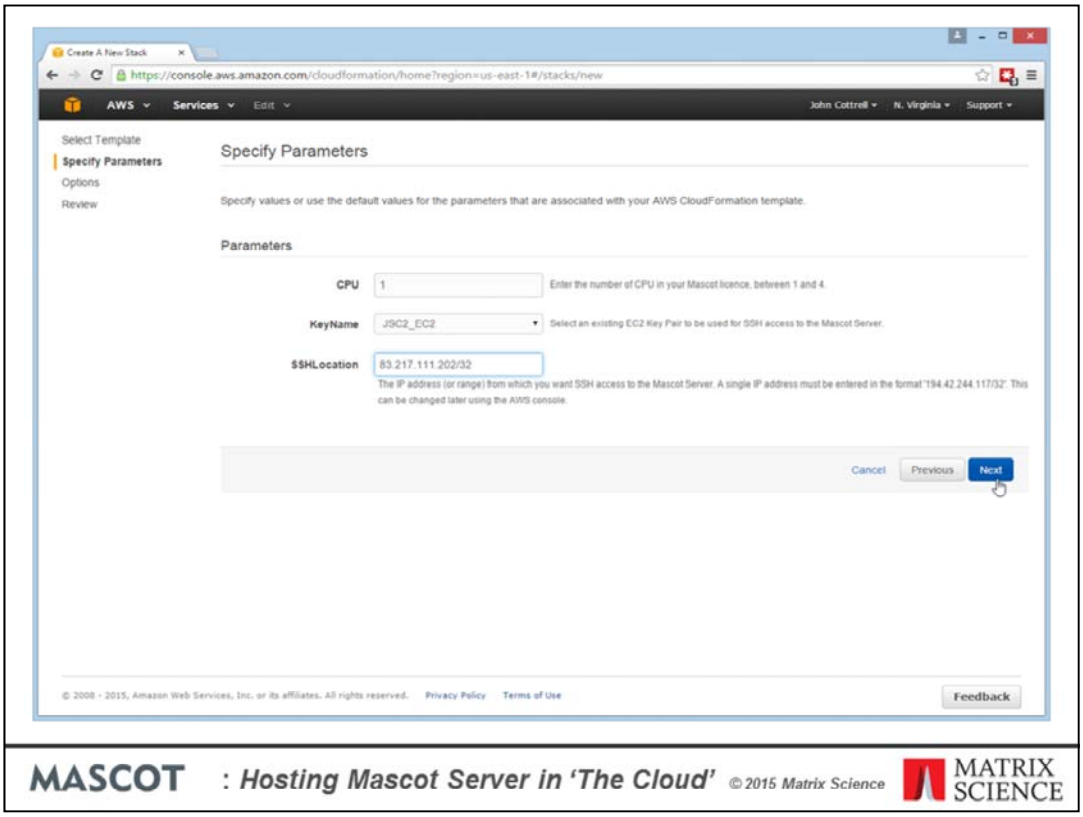
Before you can do this, you need to sign up to Amazon Web Services and give them a credit card number. Amazon will bill you direct for usage of their services; it has nothing to do with Matrix Science.



Once you are signed up, choose CloudFormation and you will be presented with this screen. Amazon use the term Stack to cover all the resources associated with a virtual server, such as the storage, network interface and security settings. We have written templates that define and build a virtual Mascot Server; all you have to do is choose the template. Choose 'Create a New Stack'



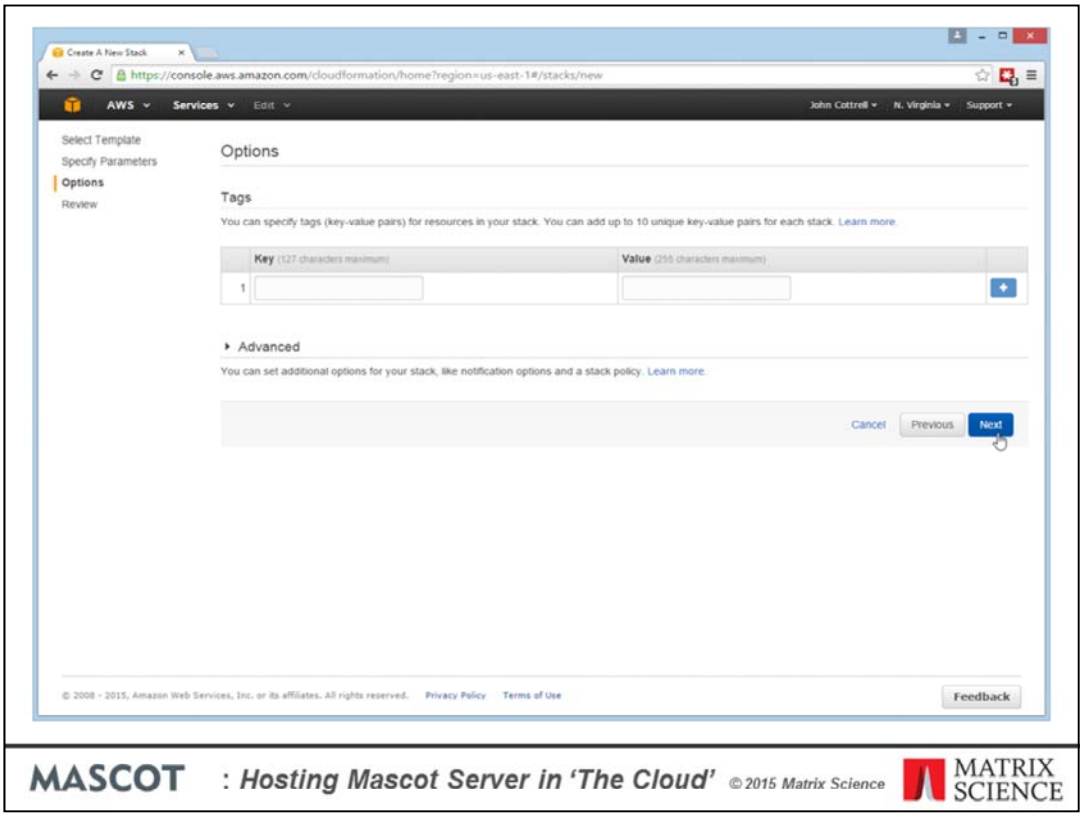
Give the stack a name. My example will be called Brian. Choose from two templates, one for a Linux Server and one for a Windows Server.



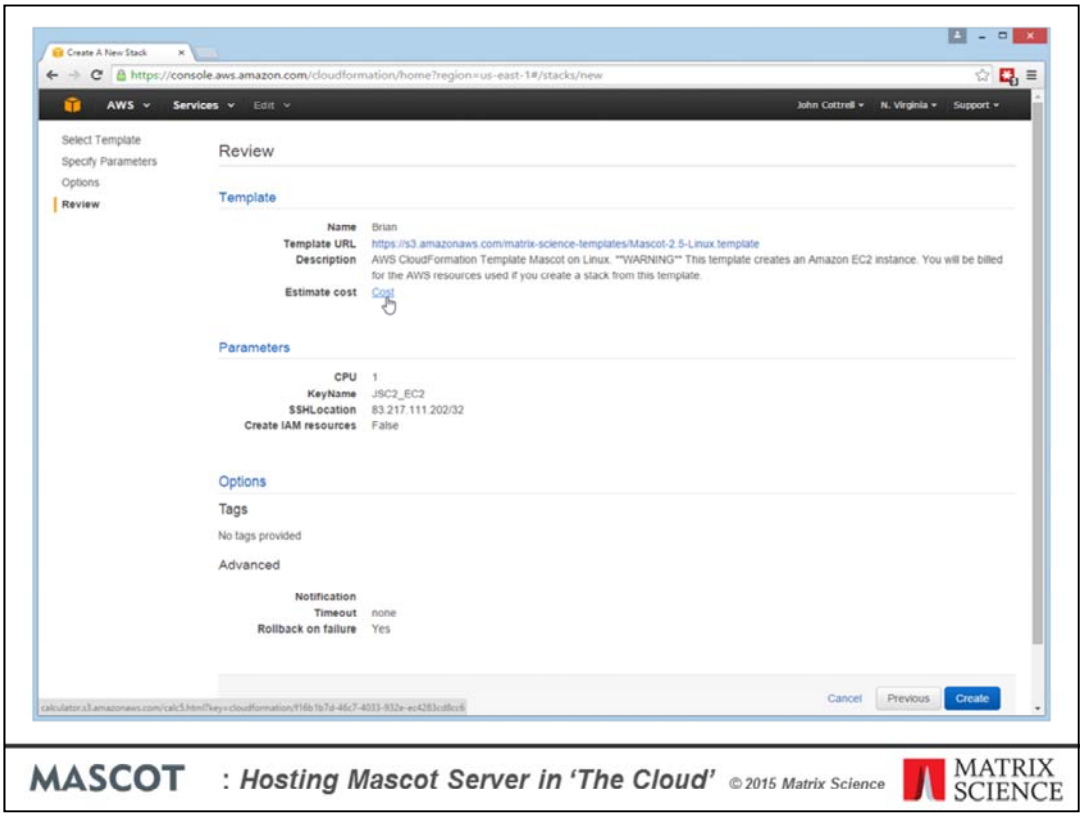
Initially, we have only written templates for single machine configurations, which means a maximum of 4 processors or 16 cores. It is perfectly possible to configure clusters to support larger licences and we will write templates for a cluster if there is the demand. Right now, you would have to set up a 1 cpu system as the head node and then configure the search nodes manually.

Console access using SSH under Linux or Remote Desktop under Windows is secured using public key cryptography. You create a key pair elsewhere in AWS and select it here.

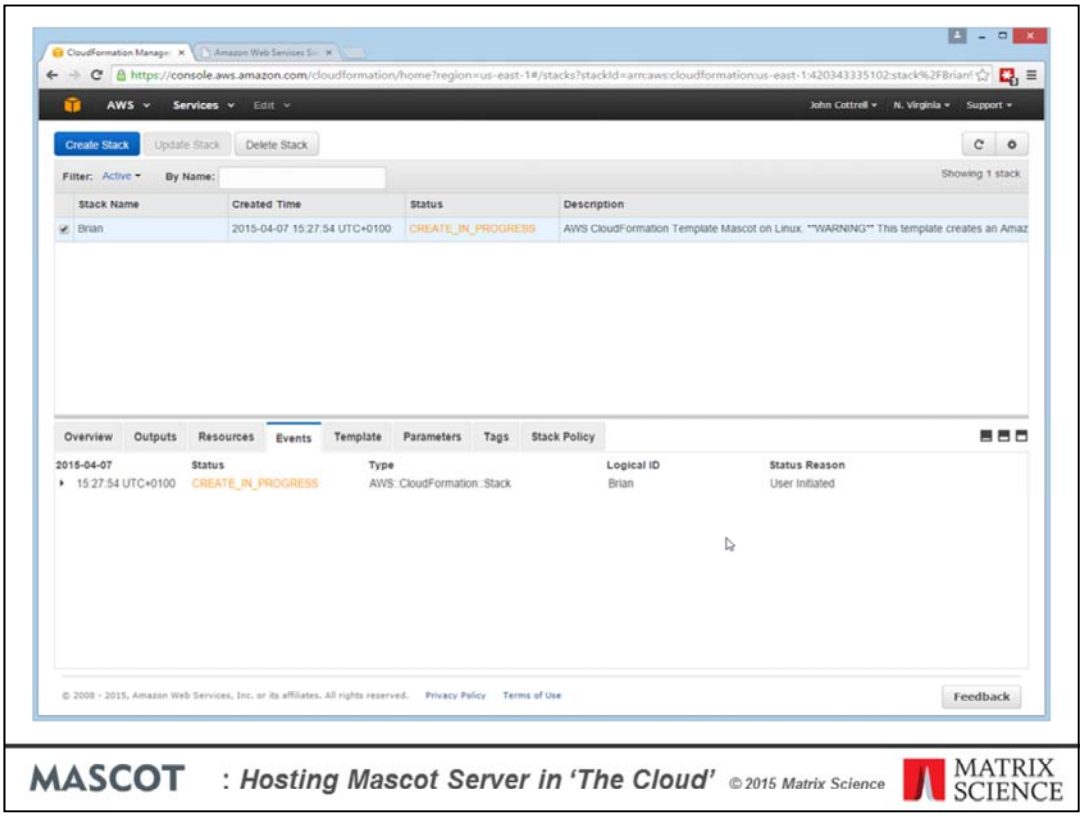
The final parameter is the IP address from which you will access the system, initially. Using IP addresses to restrict who can access the server is a very important aspect of security. Most likely, you will want to restrict console access to a very limited number of addresses, but allow HTTP access from web browsers from a wider range of addresses. I'll come back to this later.



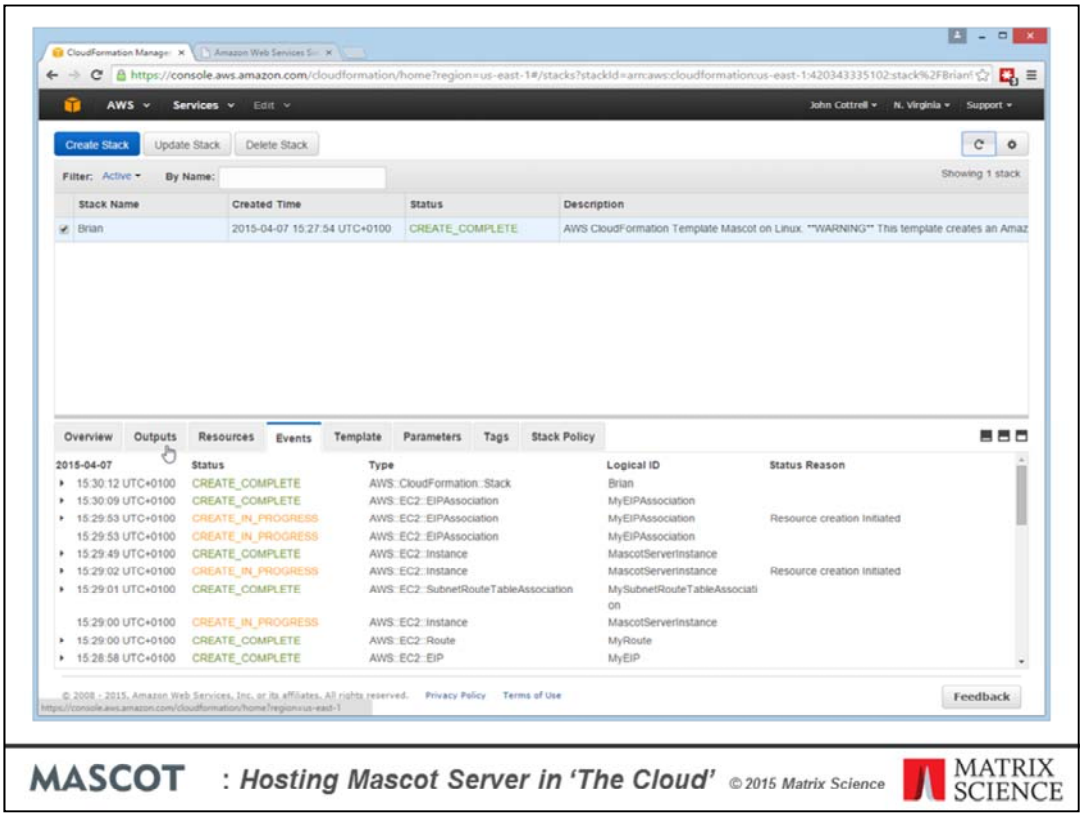
Major users of AWS use tags to organise their resources. If Brian is our only server, we can skip this page.



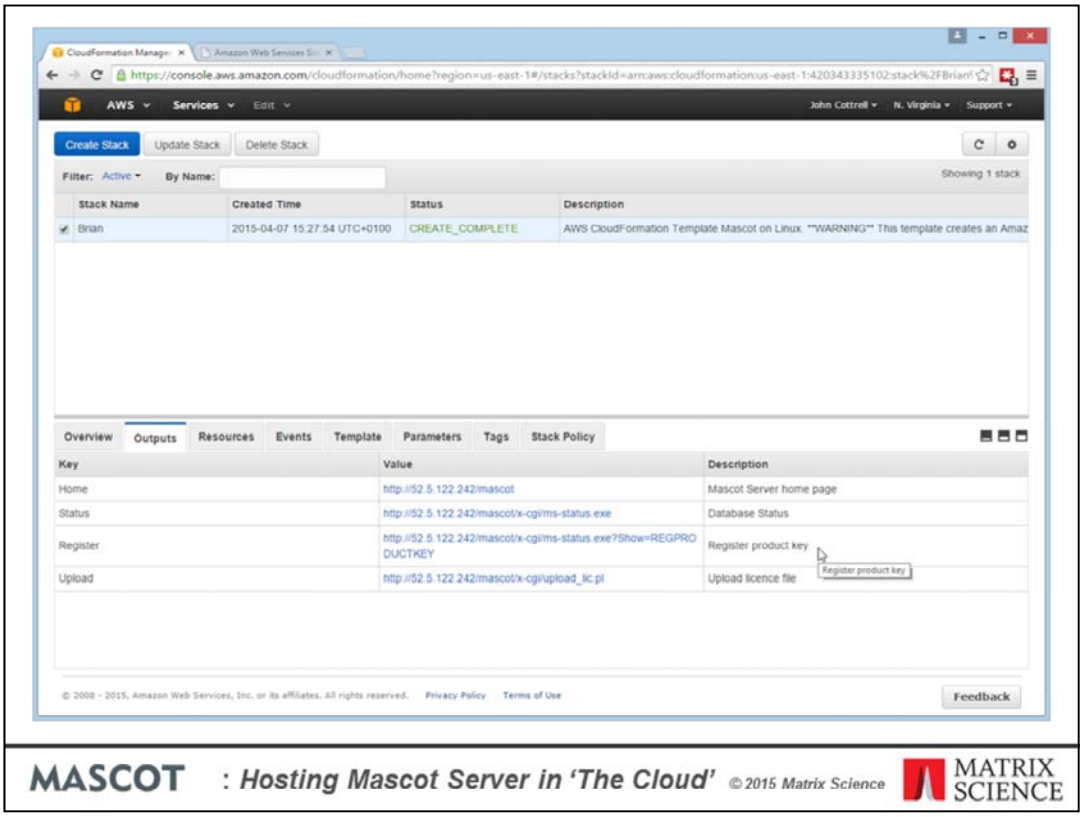
If you want a cost estimate, you can get it here. Choose Create to proceed.



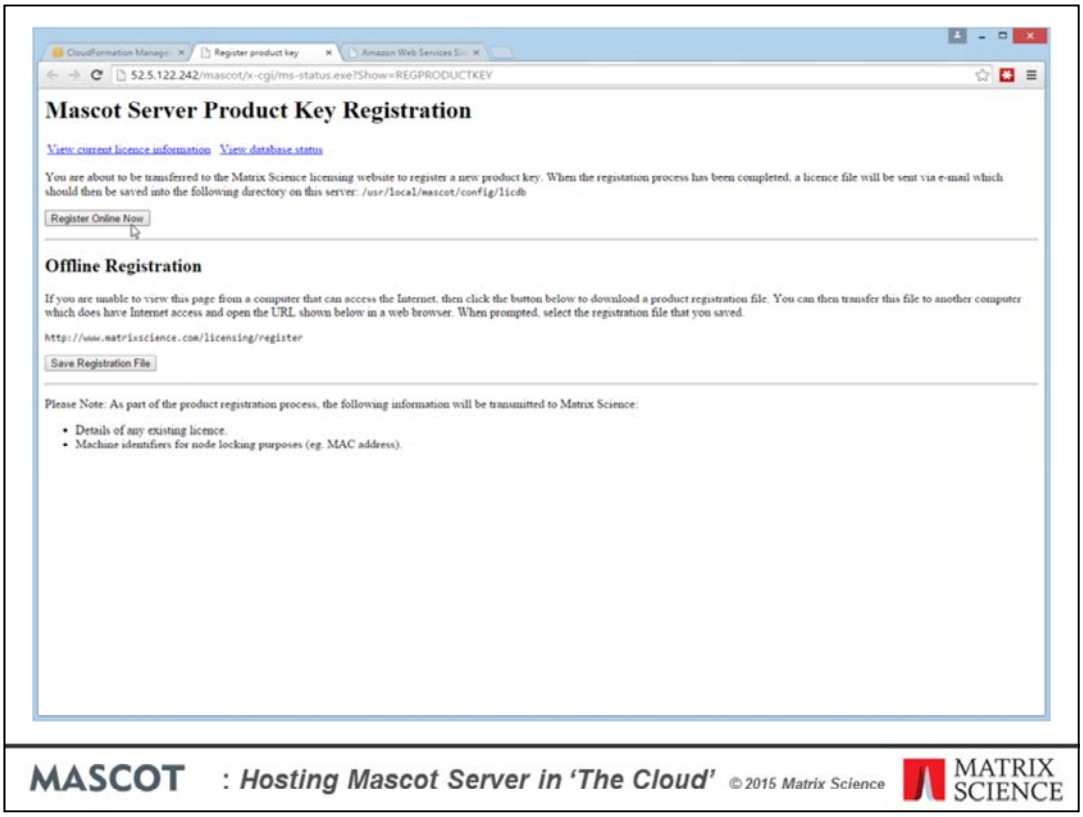
This is the progress page showing the steps in the creation of the stack. Notice that we started at just before 15:28



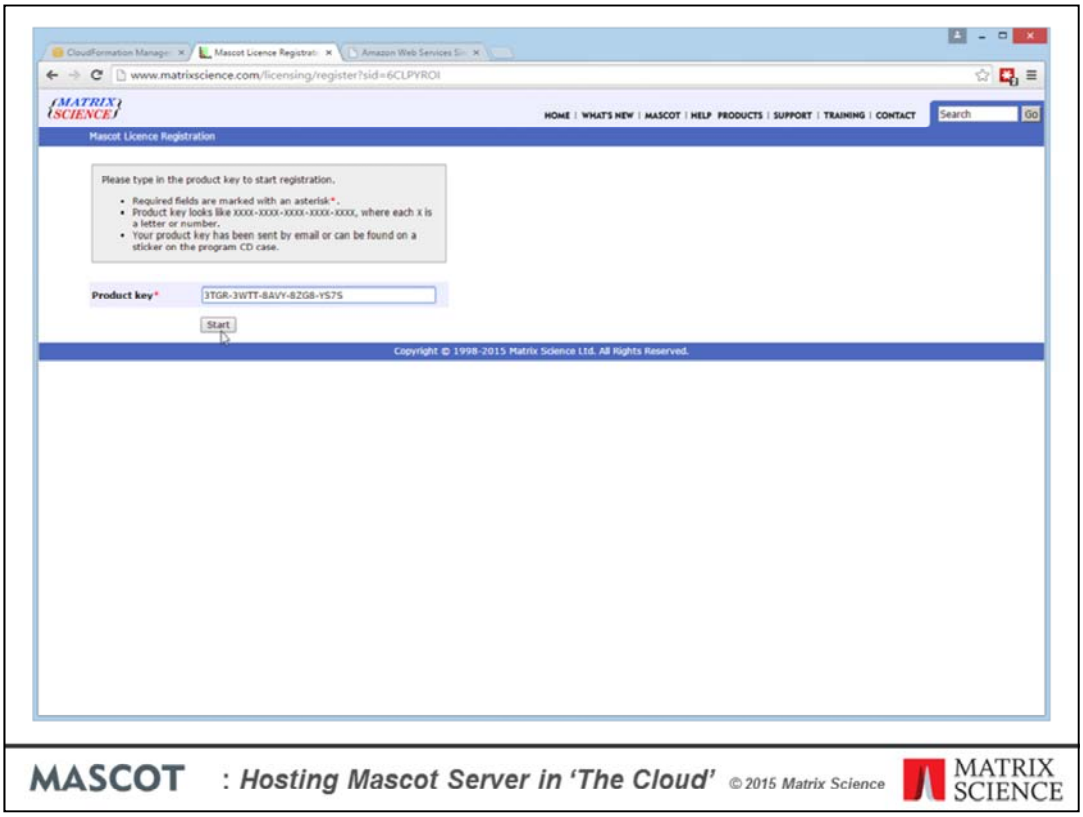
And, we're finished shortly after 15:30. Some two and a half minutes. If you switch to the Outputs tab …



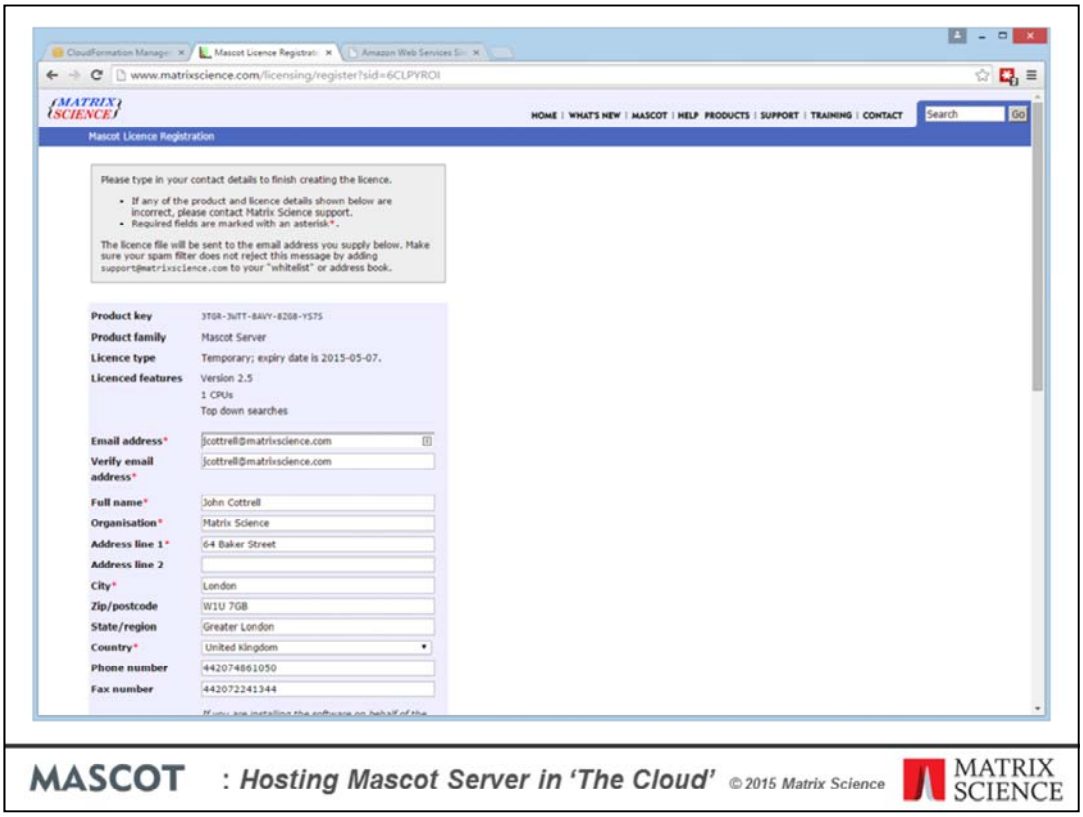
This lists some important URLs for accessing your new Mascot Server. First you have to register your product key. Initially, this will be your 30 day evaluation key.



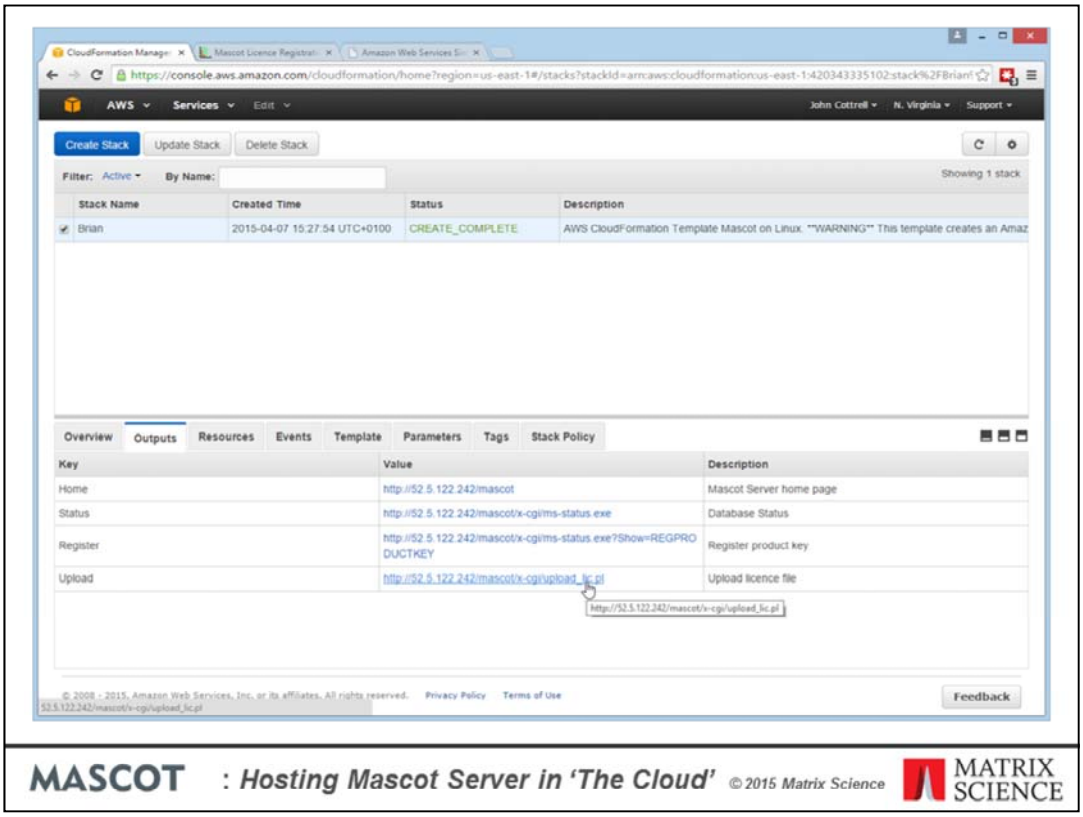
By following the link, we connect direct to the new Mascot Server. Choose 'Register Online Now'



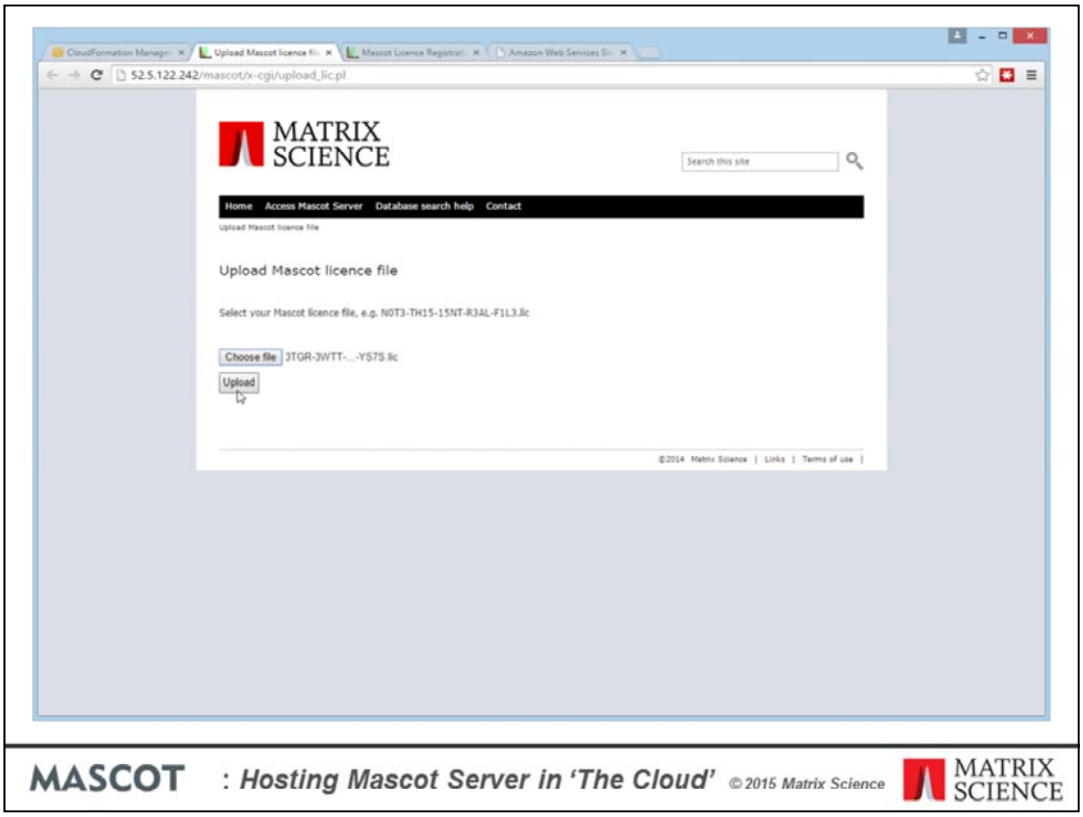
Paste in your product key



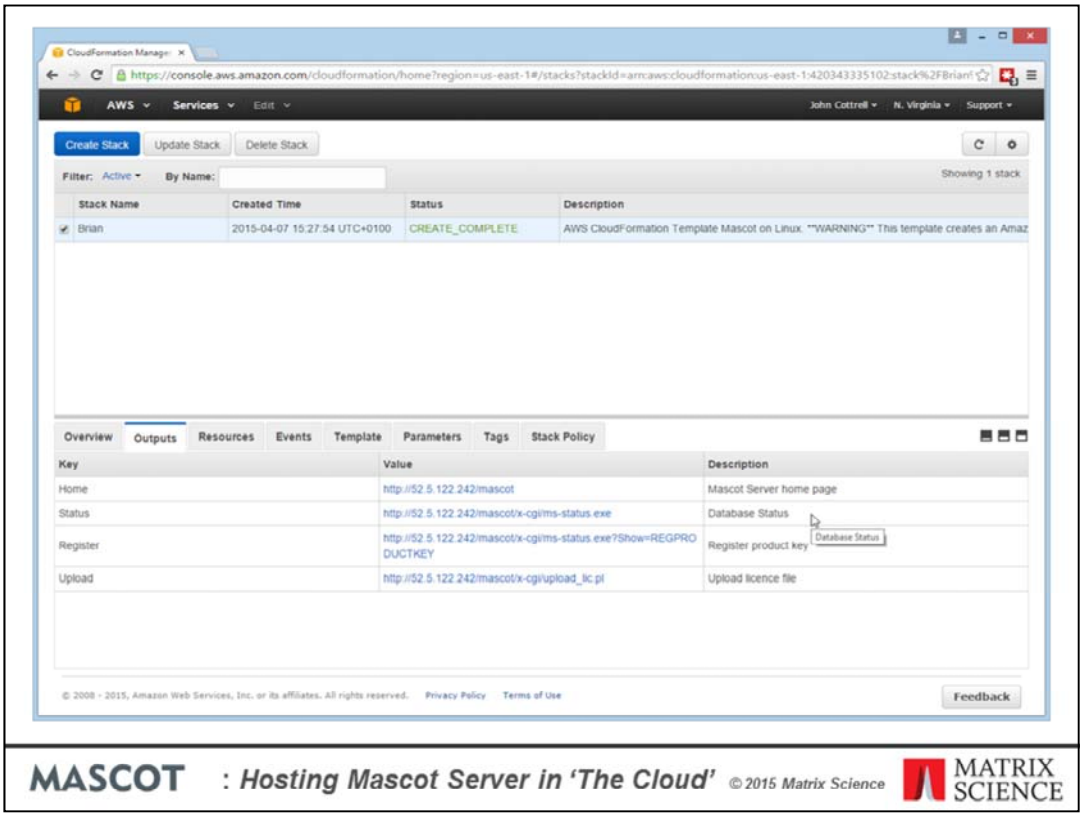
Fill in the usual stuff.



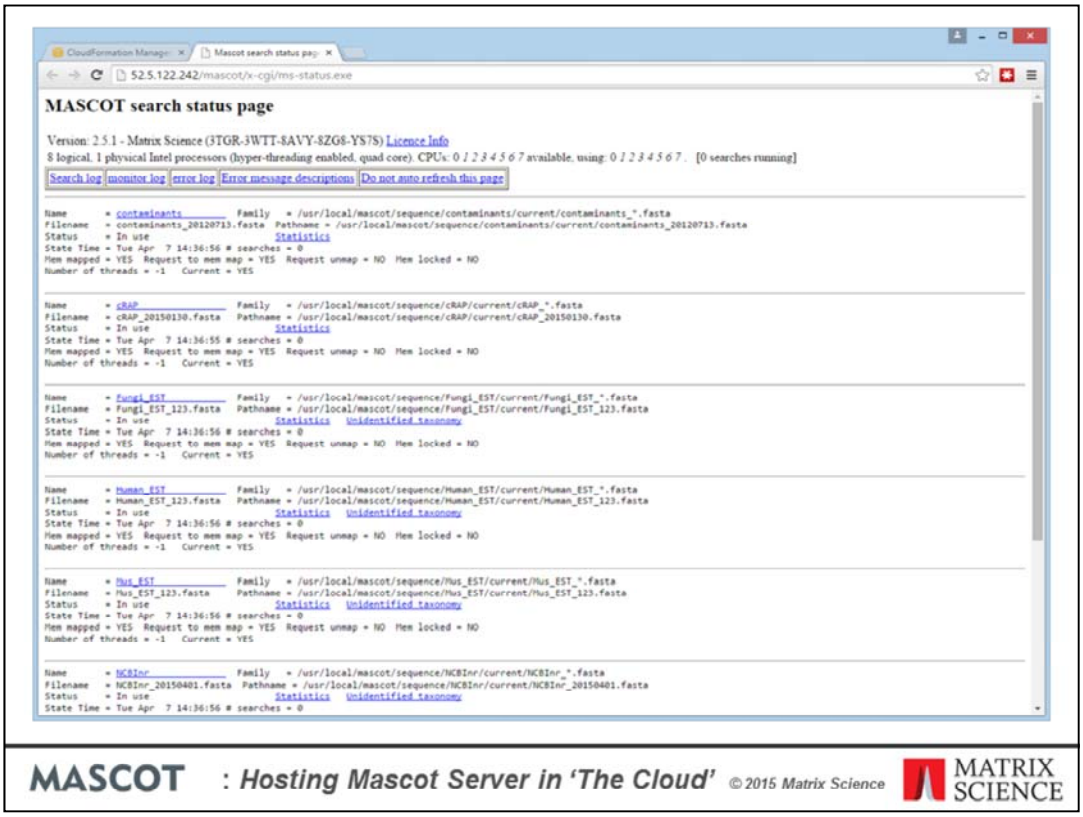
Your licence file will be returned by email. To upload it, follow one of the other links



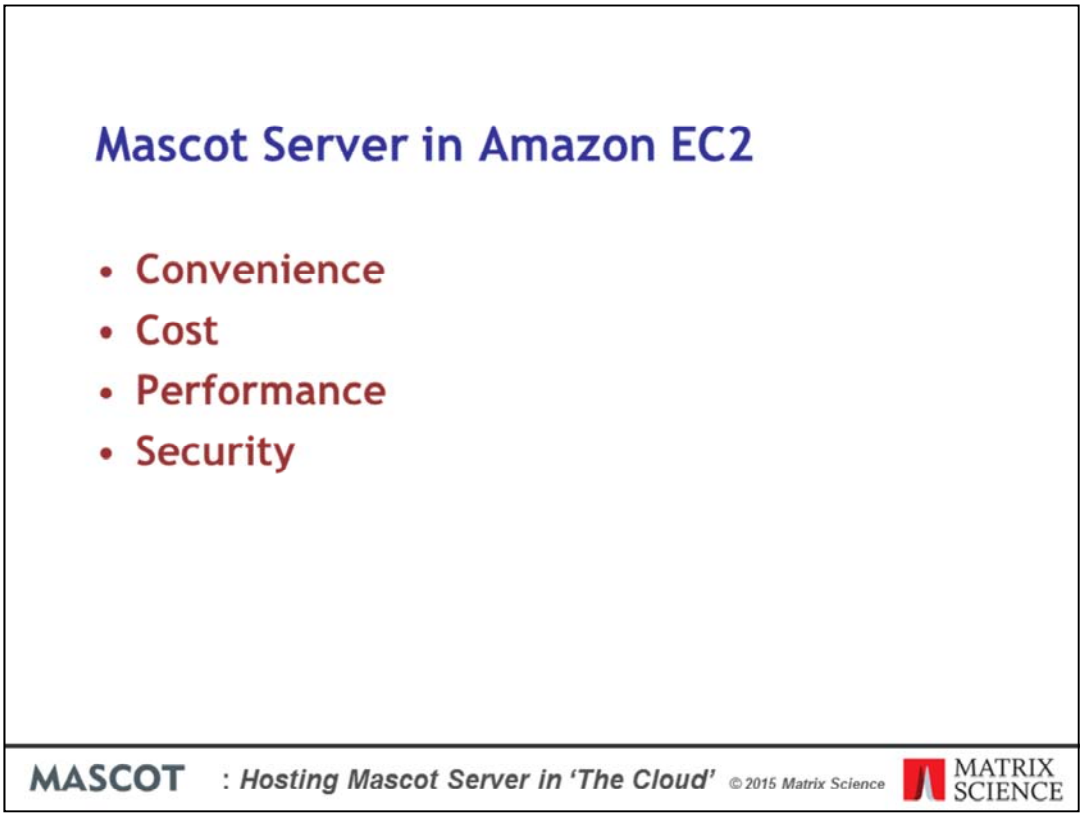
And upload. Once this is done



We can follow the link to Mascot Database Status



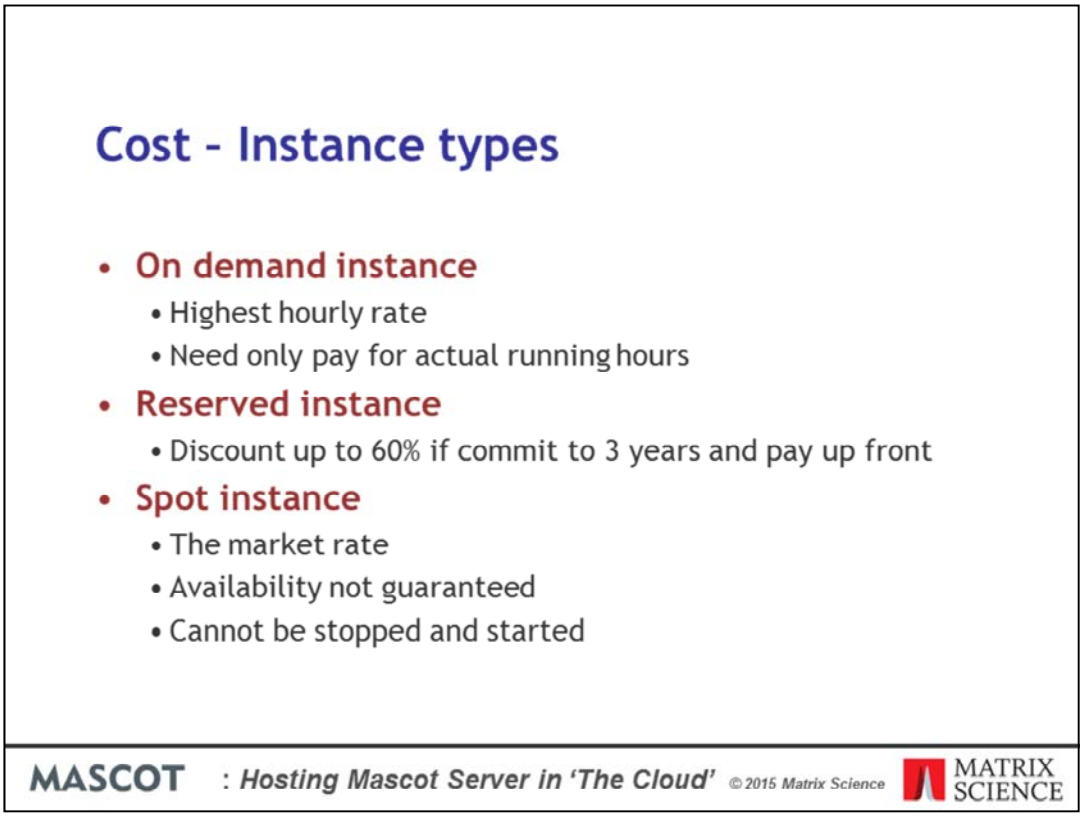
Now, I hope it becomes clear how streamlined the setup has been made. Everything is installed and configured exactly according to the manual. The server is ready for use with a selection of popular databases, such as NCBInr and SwissProt.



I think we need to look at the pros and cons of using the cloud under these four aspects.

In terms of convenience, I hope the preceding slides have convinced you. The only things I didn't show were signing up to Amazon and creating your security key pair. Everything else was shown and - literally - took a few minutes.

What about cost? The cloud is often promoted as being cost-effective because you rent shared resources rather than make capital purchases that may not be fully utilised. On the other hand, Amazon is not a charity.

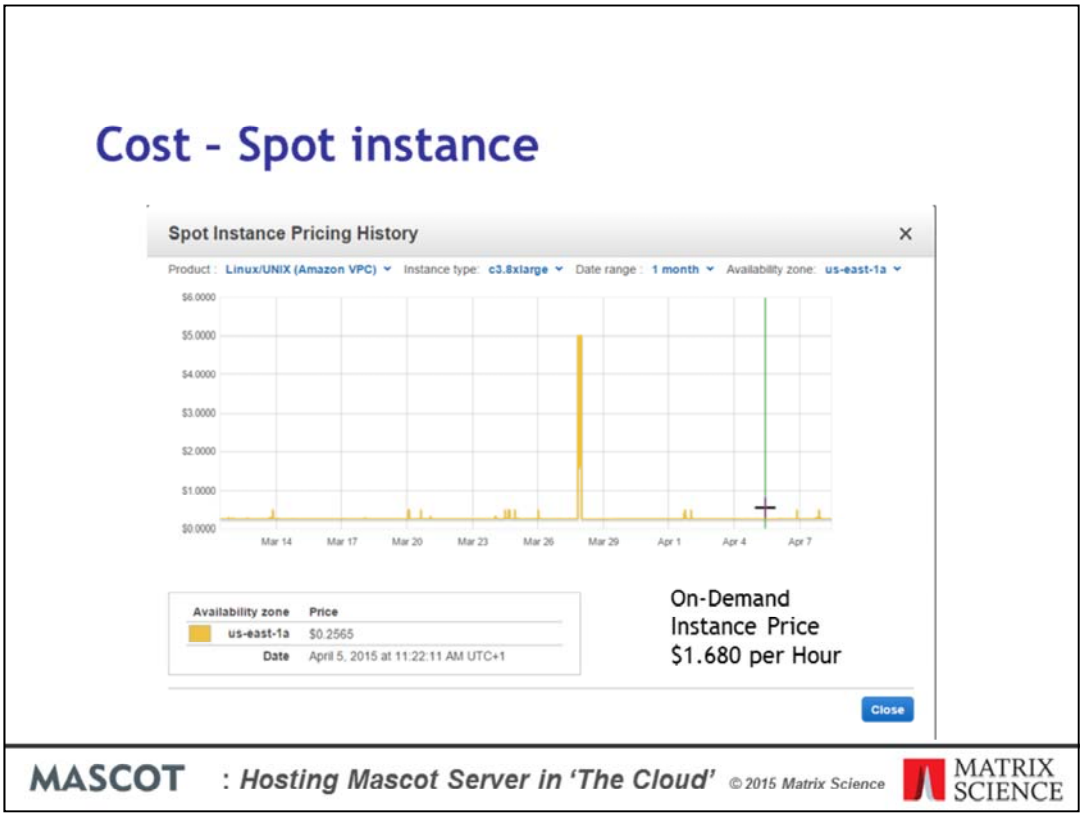


This is where it gets slightly complicated. Rather like buying a cell phone or cable TV, there are so many choices that it becomes difficult to make a meaningful comparison. There are three different ways of buying time on Amazon's virtual servers:

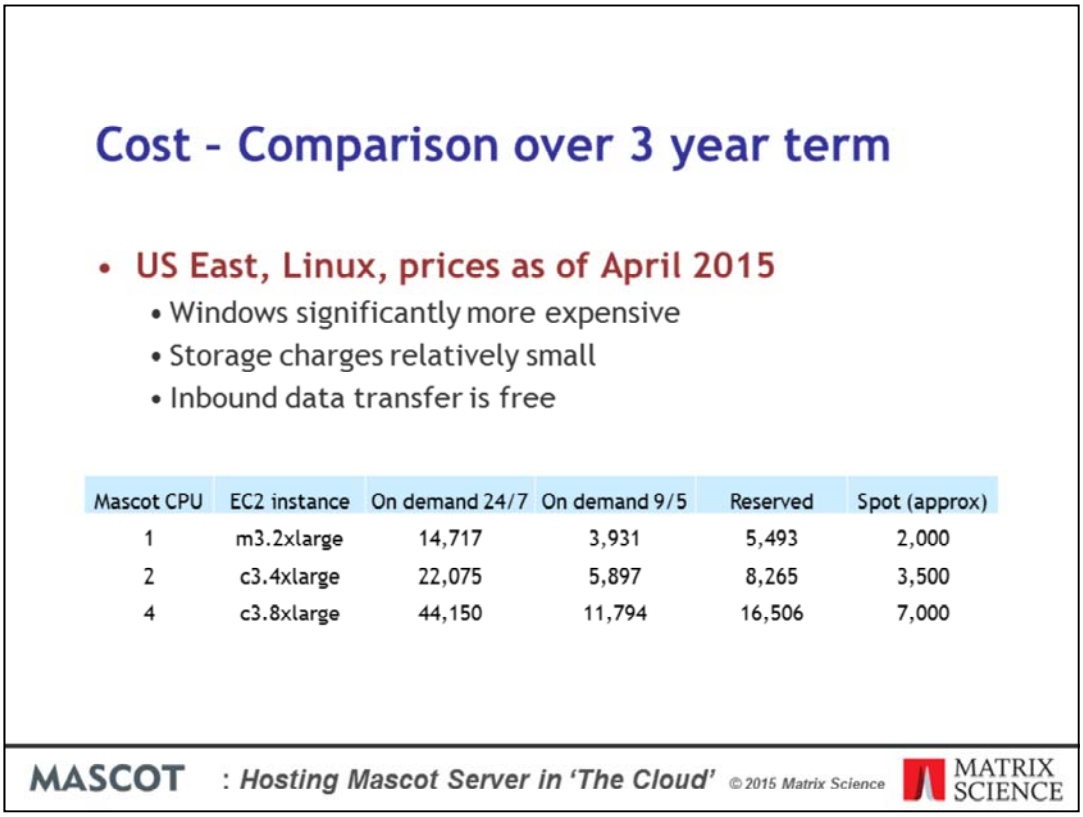
On demand instances require no commitment. You can rent one for as little as 1 hour. This carries the highest hourly rate but, because you only pay for the hours the server is running, it can be the most attractive option for a system that is not used continuously.

If you think the server will be heavily used, and are prepared to pay up front, you can get a discount on the hourly rate of up to 60% for a 3 year term. This is known as a reserved instance

A spot instance is where you make buy spare capacity as long as the market rate is below your specified maximum price. The spot rate can be a small fraction of the on-demand rate. But, availability not guaranteed and a spot instance cannot be stopped and started.



Here is an example of the market price for an instance with an on-demand hourly rate of $1.68. Most of the time, the spot rate is only 25c. But, it blips up when there is a shortage of capacity. If you had set your maximum to less than $5 per hour, you would have lost your server on March 27th. This doesn't necessarily mean you lose your data. As long as the server is correctly configured, the disk images are preserved and you can start up a new instance and re-attach the disks. But, this is potentially disruptive. You could enter a very high spot price, but this can be risky, and Amazon strongly discourage this practice. Sometimes, the spot price rises to very high levels and stays there.



So, at risk of slight over-simplification, this is what the costs look like over a period of 3 years. Prices for a Windows instance are higher because of the cost of the Windows licence. The prices in the table are the 3 year total for the instance, and don't include charges for image storage and data I/O, but these extras will be negligible for typical Mascot usage. In particular, note that inbound data transfer is free, which covers uploading of large peak lists or raw files or huge database files such as NCBInr or Trembl. You pay for outbound data transfer, but this is mostly HTML reports, which are relatively small.

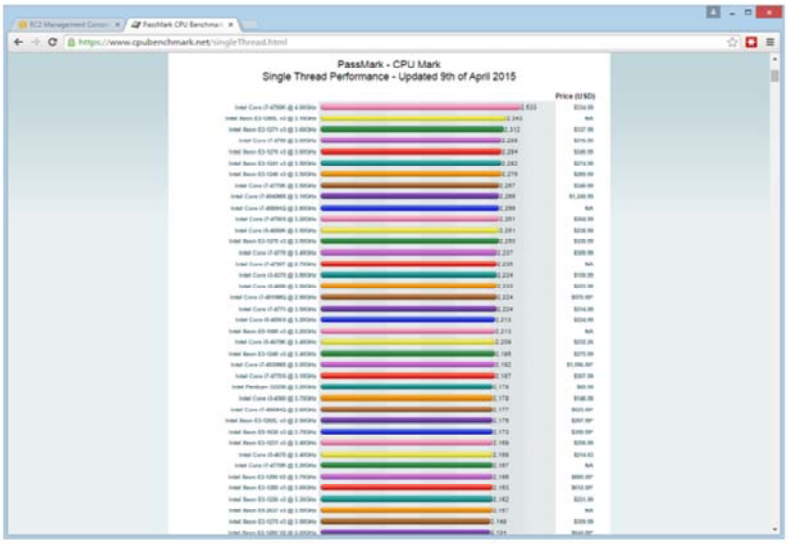
The spot price is approximate, for the reasons given earlier.

In general terms, if you anticipate continuous usage, and want guaranteed availability, choose a reserved instance. For a server that is only used during 'office hours', on-demand can be a better deal, but you have to remember to stop and start the instance each day. (This can be automated.) If you can live with occasional interruptions, and are prepared to re-create your server when it is terminated, a spot instance is the cheapest option.

Performance - processor speed

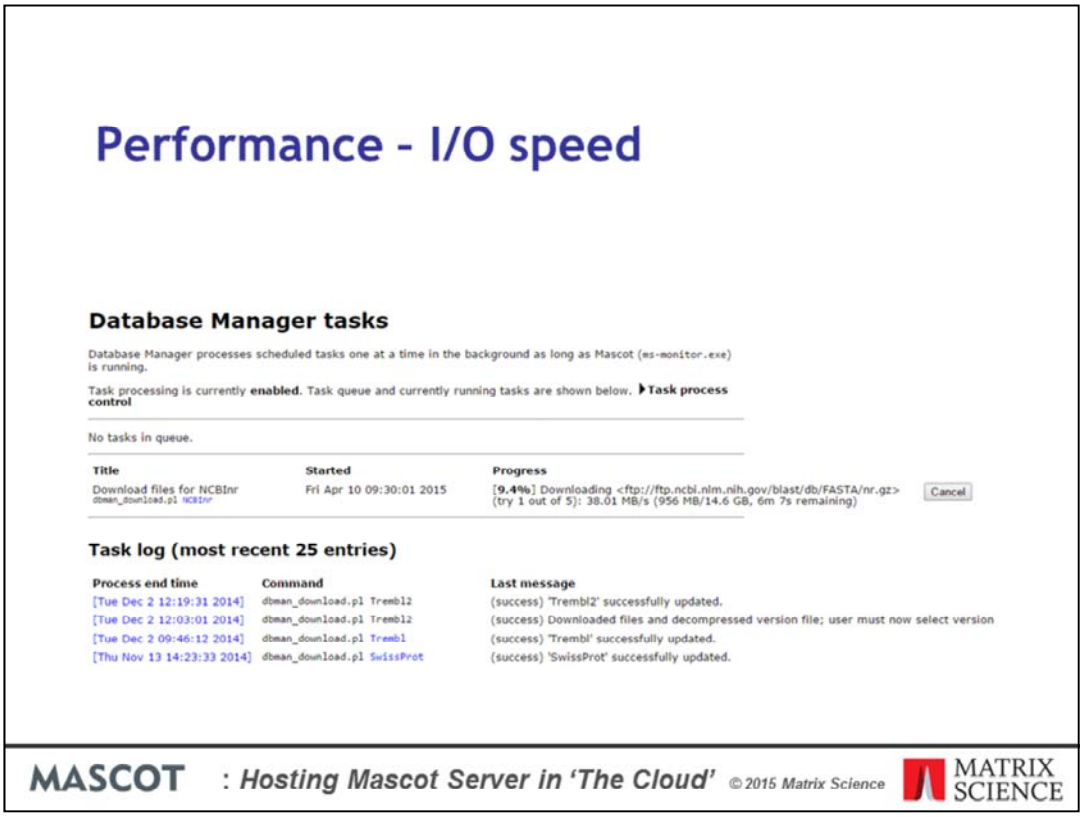
What about performance? In a word, excellent. We're looking at the 'per thread' benchmark, which is the important one because Mascot is licenced according to the number of threads

Performance - processor speed

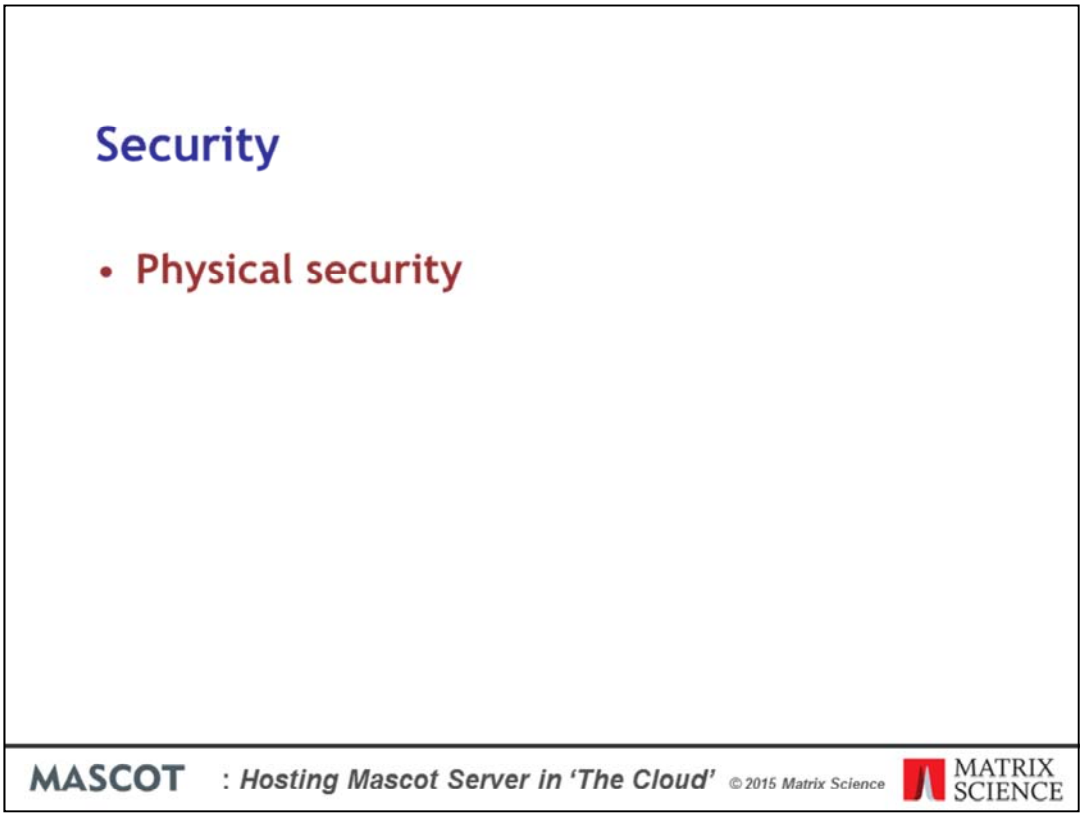


MASCOT : *Hosting Mascot Server in 'The Cloud'* MATRIX SCIENCE

Processor speed compares well with fast current generation processors. We find that the CPU Mark is a good guide as to Mascot performance



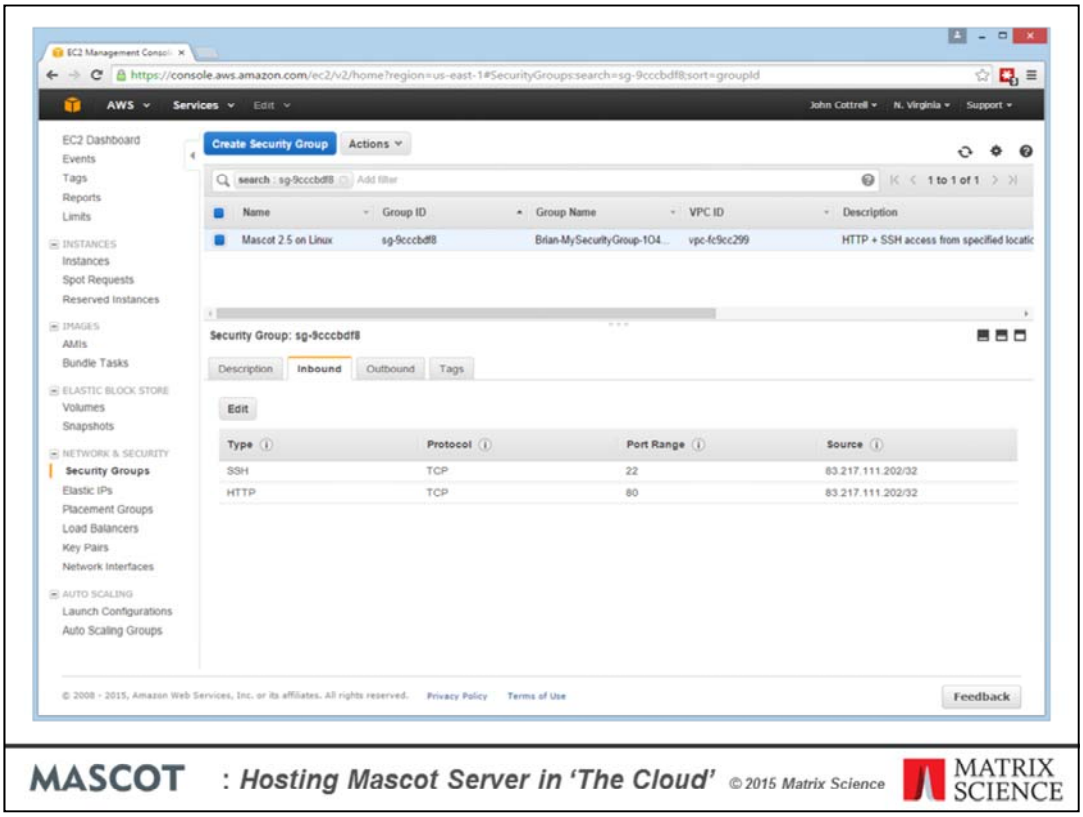
I/O speed is very fast. You can see this database download from NCBI is getting 38 MB/s … not Mb / s. That's less than 7 minutes for the compressed NCBInr Fasta file



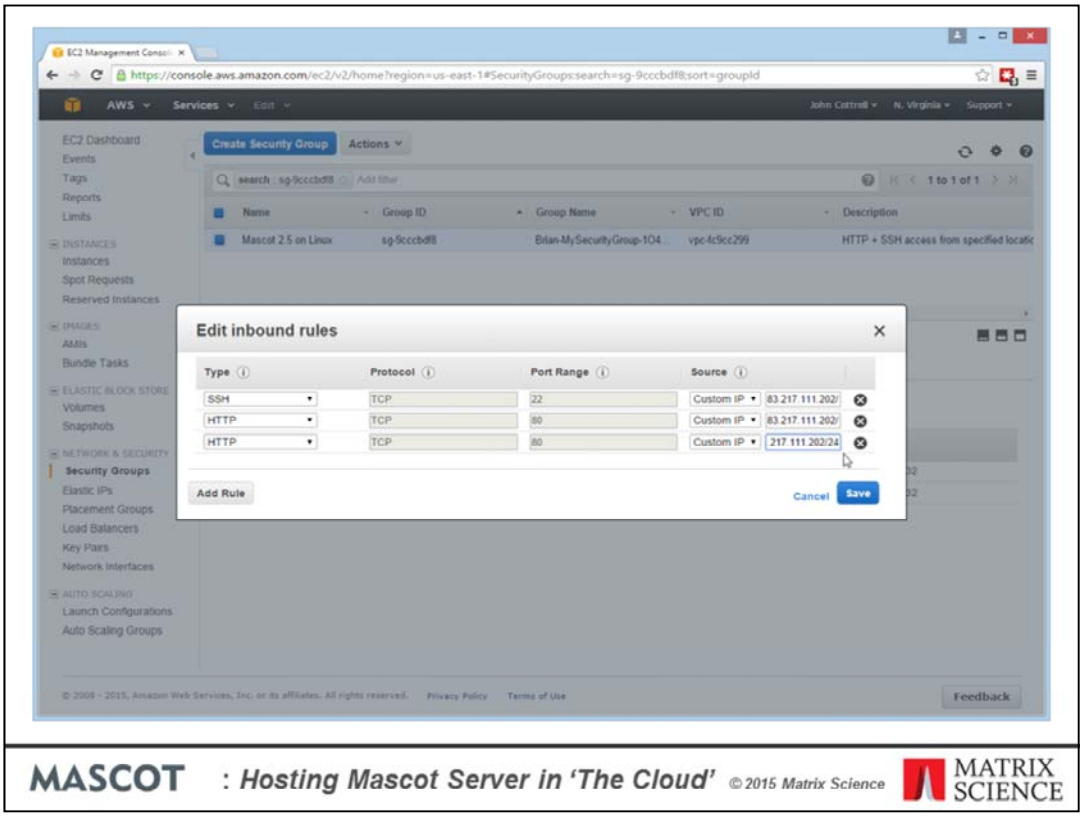
Next, security. I think there are three aspects to this.

First, physical security. Someone walking into the lab and walking out with your PC under their arm. Or, the lab burning down or being flooded. I would guess that an Amazon data centre is very much more secure than the average lab.

Second, electronic access. As long as you keep your key pair safe and only open up the firewall for specific ports and specific IP addresses, I think this is going to be at least as secure as having a server sitting on your LAN. Amazon takes security very seriously and are fairly open about their measures



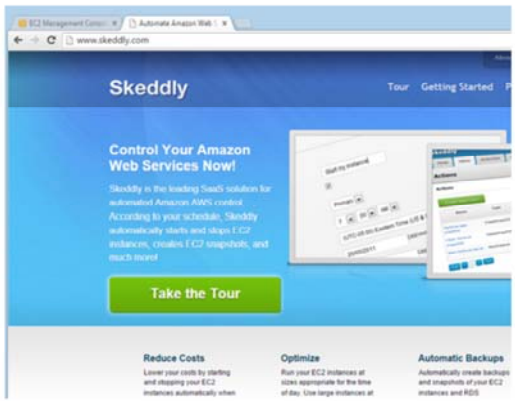
On the subject of the firewall, you change the security settings using the Amazon EC2 console. After a stack is created, everything is blocked apart from the IP addresses you entered into the template. The /32 at the end of the IP address defines the range as a single address



Here, we are opening up HTTP access to a range of /24 which means all addresses from 83.217.111.0 to 83.217.111.255. The user interface is very intuitive.

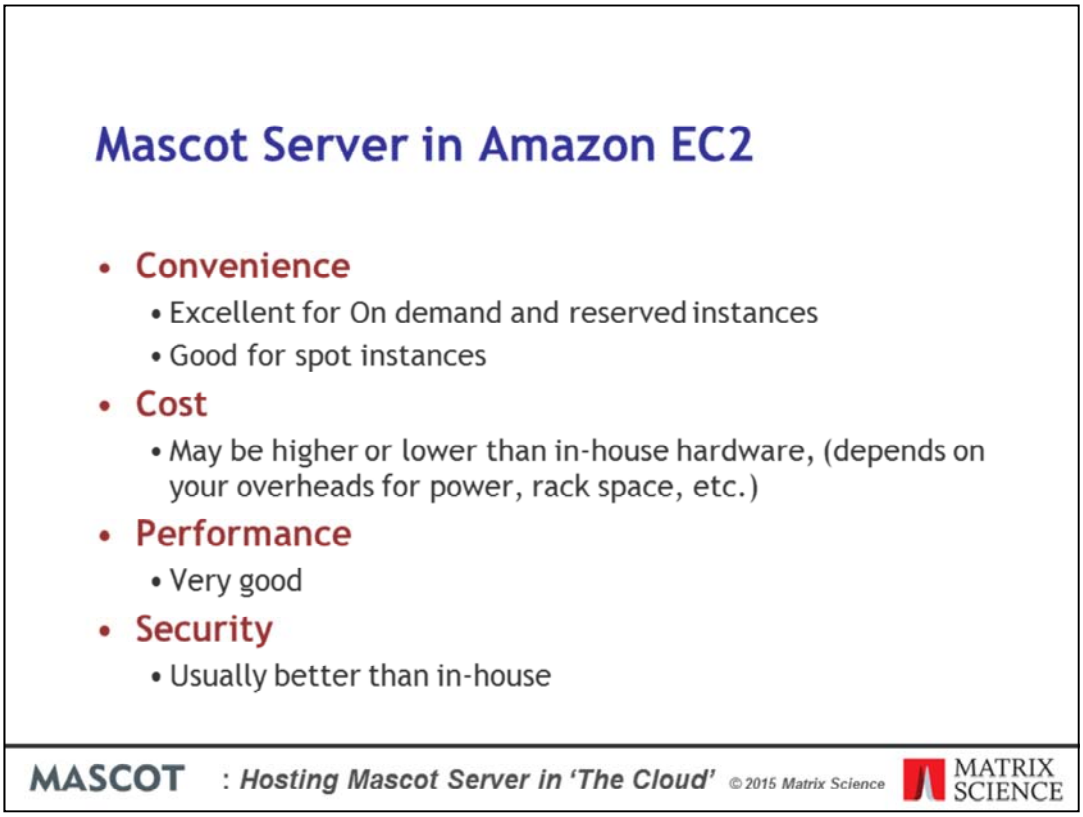
Security

- Physical security
- Hacking
- Backup



MASCOT : Hosting Mascot Server in 'The Cloud' MATRIX SCIENCE

Third, data backup. This is up to you. Disks can and do fail. You can backup selected files to Amazon S3 (Simple Storage Service) or Glacier (very low cost storage), or make snapshots of complete disks. As you might imagine, there are companies who offer automated backup as a service, such as Skeddly



So, to summarise

I would rate convenience as excellent for on demand and reserved instances. I reduce this to good for spot instances because of the need to recreate the instance when it is terminated

Cost may be higher or lower than in-house hardware. It depends on whether you are charged overheads for power, rack space, etc.

Performance is very good compared with commodity PC hardware

Security will be better than hosting a server in-house, in most cases

If this looks interesting and you want to give it a try, what do you have to do?

First of all, you have to sign up to Amazon web services.

Then, email support and we'll send detailed instructions plus a 30 day product key.

If you decide to continue, and you didn't have a licence already, you'll need to buy one. There is no difference in pricing between a licence on Amazon and a licence on local hardware

If you had an existing licence and it was an older version, you'll need to buy an update. Once on the current version, just request that your licence be transferred.