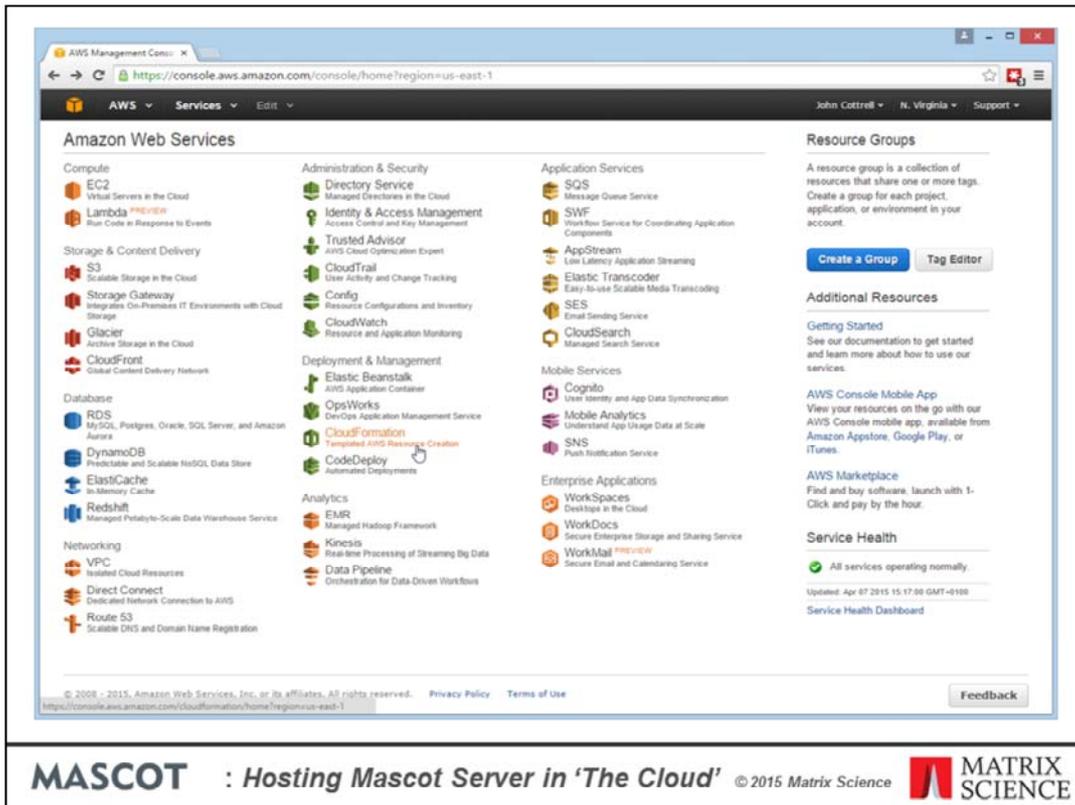


Hosting Mascot Server in 'The Cloud'

John Cottrell
Matrix Science

MASCOT : *Hosting Mascot Server in 'The Cloud'* © 2015 Matrix Science  **MATRIX
SCIENCE**

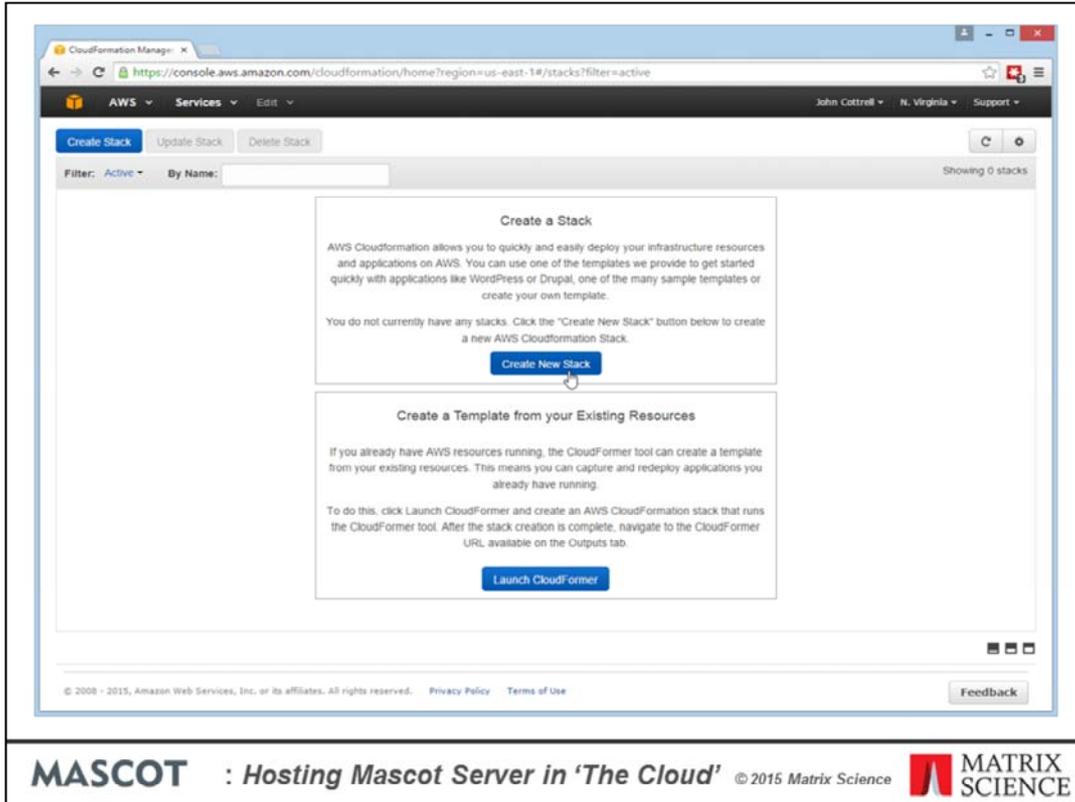
Cloud computing is a marketing term that means different things to different people. In this presentation, we look at the pros and cons of using Amazon Web Services rather than setting up a physical server on your LAN



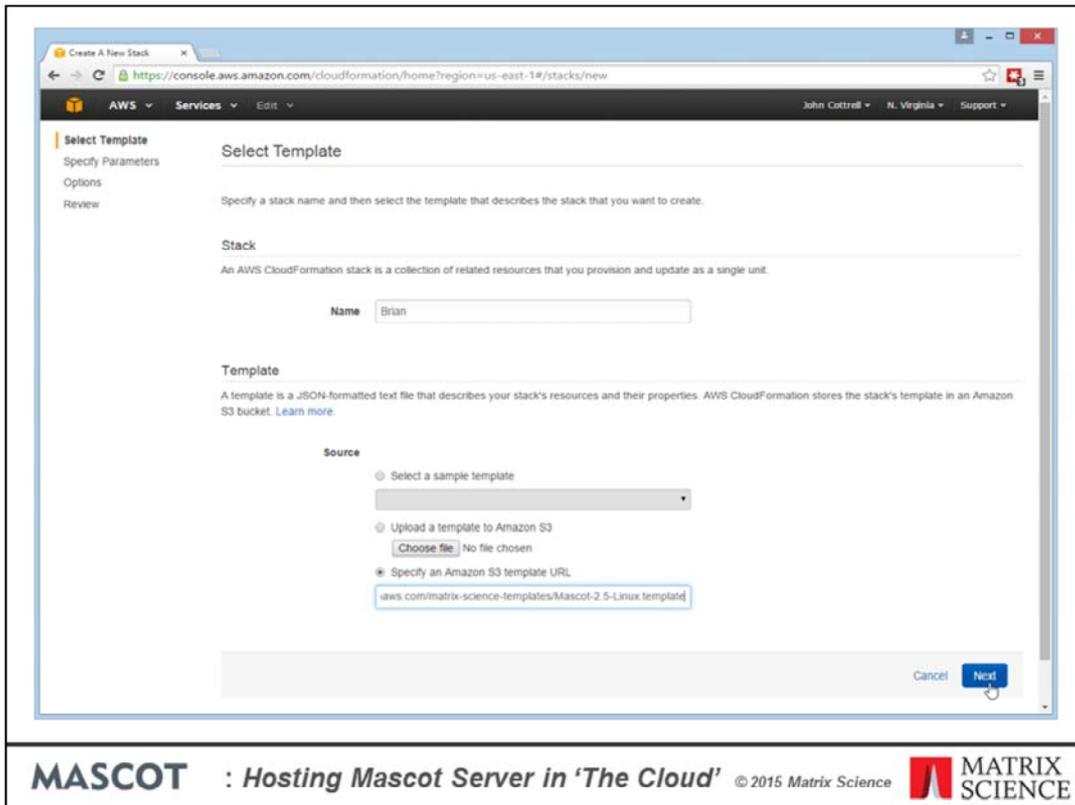
MASCOT : *Hosting Mascot Server in 'The Cloud'* © 2015 Matrix Science 

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.

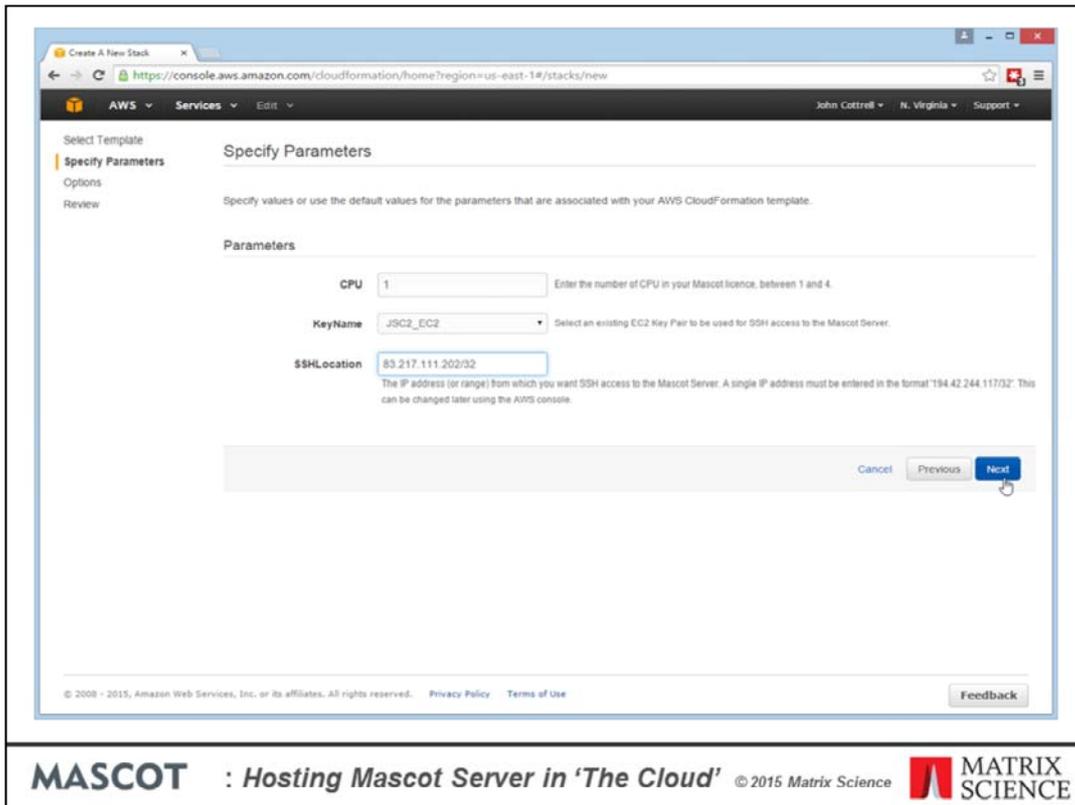
Before you can do this, you need to sign up to Amazon Web Services and give them a credit card number. Amazon will bill you direct for usage of their services; it has nothing to do with Matrix Science.



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.

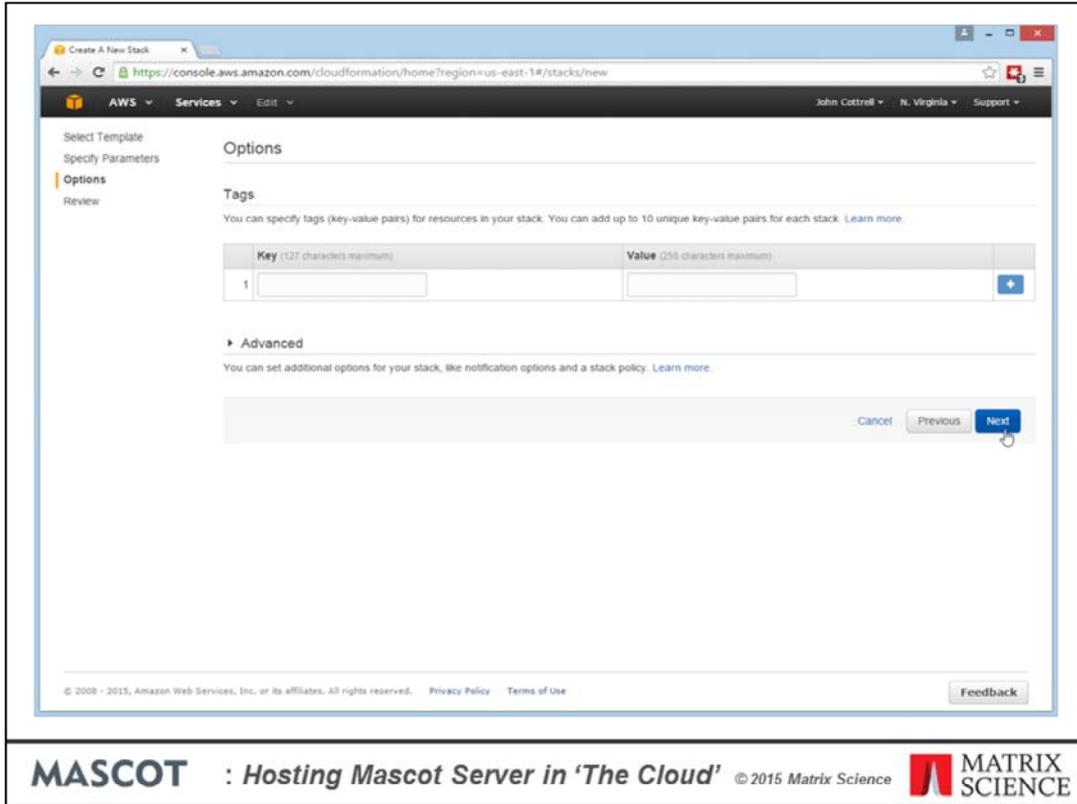


MASCOT : *Hosting Mascot Server in 'The Cloud'* © 2015 Matrix Science  **MATRIX SCIENCE**

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.

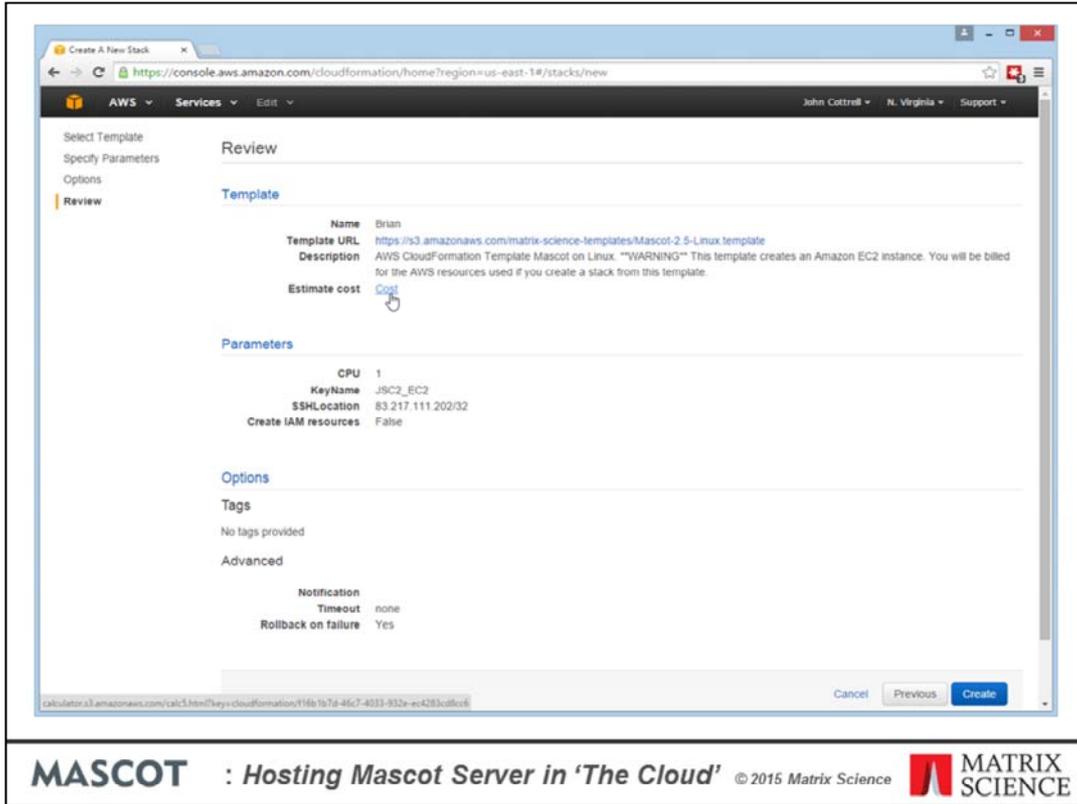
Console access using SSH under Linux or Remote Desktop under Windows is secured using public key cryptography. You create a key pair elsewhere in AWS and select it here.

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.



MASCOT : *Hosting Mascot Server in 'The Cloud'* © 2015 Matrix Science  **MATRIX SCIENCE**

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.

The screenshot shows the AWS CloudFormation console interface. At the top, there are navigation tabs for 'Create Stack', 'Update Stack', and 'Delete Stack'. Below this, a table lists the stack 'Brian' with a status of 'CREATE_IN_PROGRESS'. The description for the stack is 'AWS CloudFormation Template Mascot on Linux. **WARNING** This template creates an Amazon Linux instance profile.' Below the table, there are tabs for 'Overview', 'Outputs', 'Resources', 'Events', 'Template', 'Parameters', 'Tags', and 'Stack Policy'. The 'Events' tab is selected, showing a log of events. The first event is dated '2015-04-07' at '15:27:54 UTC+0100' with a status of 'CREATE_IN_PROGRESS', type 'AWS::CloudFormation::Stack', logical ID 'Brian', and status reason 'User Initiated'. At the bottom of the console, there is a footer with the text '© 2008 - 2015, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use' and a 'Feedback' button.

MASCOT : *Hosting Mascot Server in 'The Cloud'* © 2015 Matrix Science 

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

The screenshot displays the AWS CloudFormation console interface. At the top, there are navigation buttons for 'Create Stack', 'Update Stack', and 'Delete Stack'. Below these, a filter is set to 'Active' and the stack name 'Brian' is entered. The main table shows one stack with the following details:

Stack Name	Created Time	Status	Description
Brian	2015-04-07 15:27:54 UTC+0100	CREATE_COMPLETE	AWS CloudFormation Template Mascot on Linux. WARNING This template creates an Amaz

Below the stack list, the 'Events' tab is selected, showing a detailed log of the stack's creation process:

Time	Status	Type	Logical ID	Status Reason
2015-04-07 15:30:12 UTC+0100	CREATE_COMPLETE	AWS::CloudFormation::Stack	Brian	
2015-04-07 15:30:09 UTC+0100	CREATE_COMPLETE	AWS::EC2::EIPAssociation	MyEIPAssociation	
2015-04-07 15:29:53 UTC+0100	CREATE_IN_PROGRESS	AWS::EC2::EIPAssociation	MyEIPAssociation	Resource creation initiated
2015-04-07 15:29:53 UTC+0100	CREATE_IN_PROGRESS	AWS::EC2::EIPAssociation	MyEIPAssociation	
2015-04-07 15:29:49 UTC+0100	CREATE_COMPLETE	AWS::EC2::Instance	MascotServerInstance	
2015-04-07 15:29:02 UTC+0100	CREATE_IN_PROGRESS	AWS::EC2::Instance	MascotServerInstance	Resource creation initiated
2015-04-07 15:29:01 UTC+0100	CREATE_COMPLETE	AWS::EC2::SubnetRouteTableAssociation	MySubnetRouteTableAssociati on	
2015-04-07 15:29:00 UTC+0100	CREATE_IN_PROGRESS	AWS::EC2::Instance	MascotServerInstance	
2015-04-07 15:29:00 UTC+0100	CREATE_COMPLETE	AWS::EC2::Route	MyRoute	
2015-04-07 15:28:58 UTC+0100	CREATE_COMPLETE	AWS::EC2::EIP	MyEIP	

At the bottom of the console, there is a footer with copyright information: © 2008 - 2015, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use. A 'Feedback' button is also visible.

MASCOT : *Hosting Mascot Server in 'The Cloud'* © 2015 Matrix Science 

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

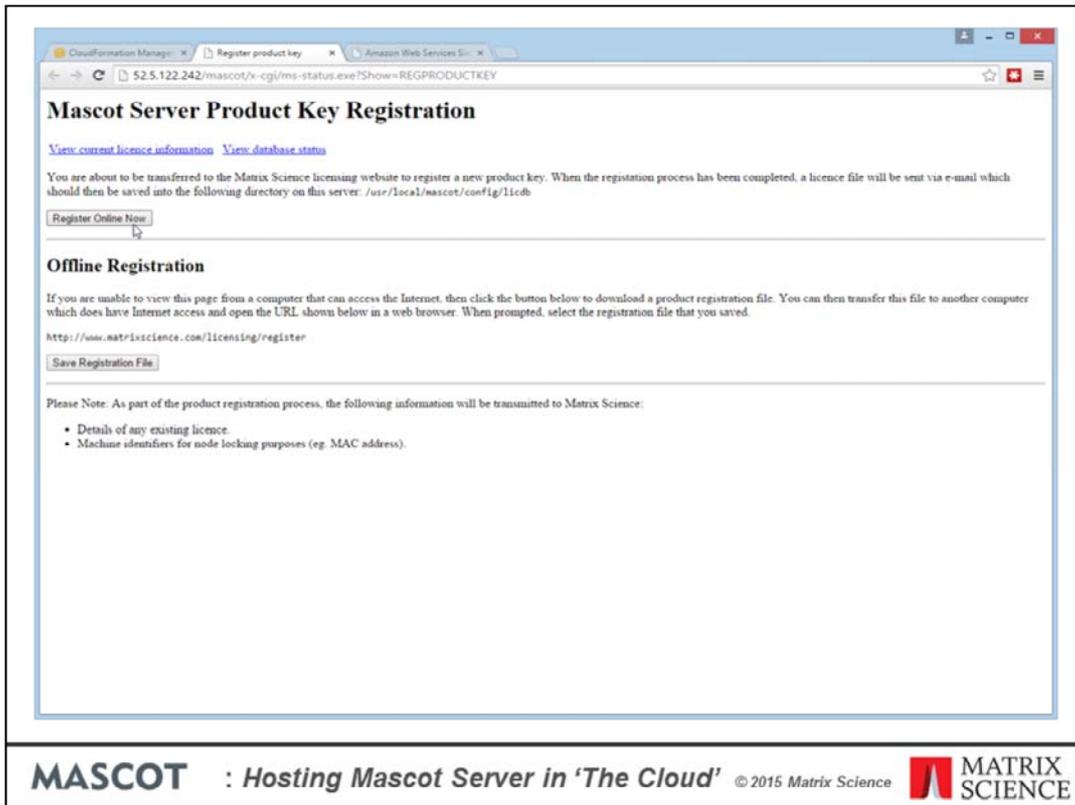
The screenshot shows the AWS CloudFormation console interface. At the top, there are buttons for 'Create Stack', 'Update Stack', and 'Delete Stack'. Below these, a table lists the stack 'Brian' with a status of 'CREATE_COMPLETE'. The 'Outputs' tab is selected, displaying a table with the following data:

Key	Value	Description
Home	http://52.5.122.242/mascot	Mascot Server home page
Status	http://52.5.122.242/mascot/x-cgims-status.exe	Database Status
Register	http://52.5.122.242/mascot/x-cgims-status.exe?Show=REGPRO DUCTKEY	Register product key
Upload	http://52.5.122.242/mascot/x-cgi/upload_lic.pl	Upload licence file

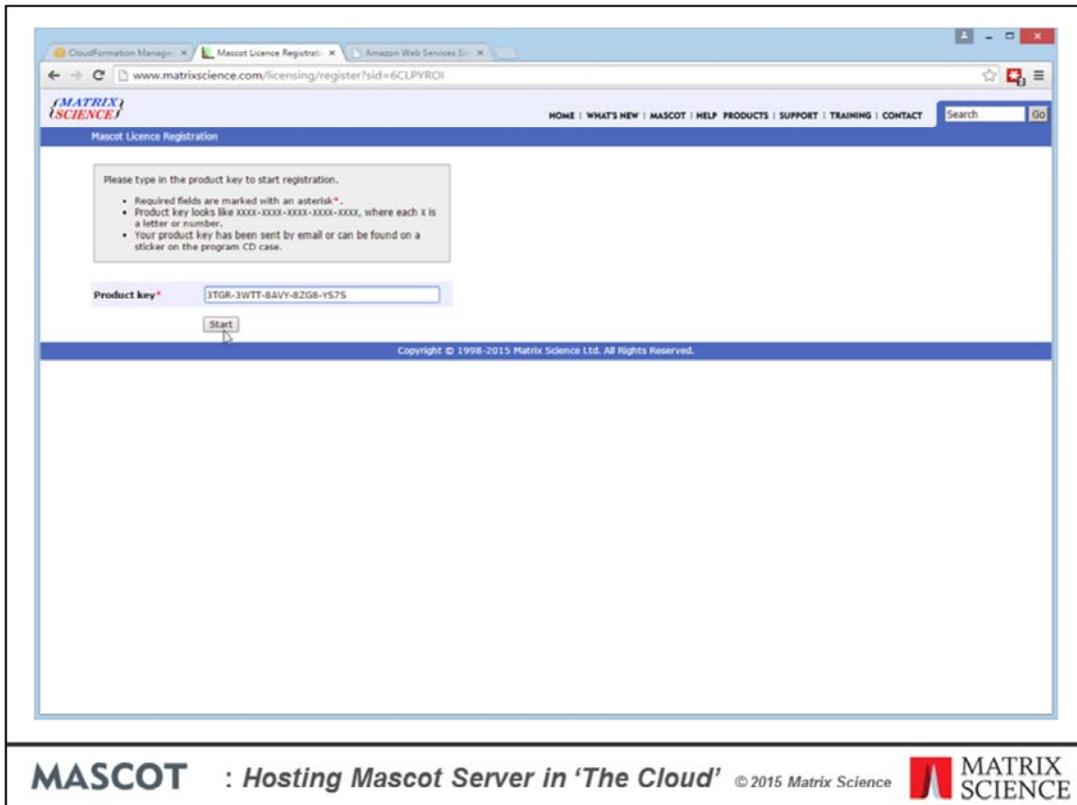
A tooltip for the 'Register' output shows a button labeled 'Register product key'. At the bottom of the console, there is a footer with copyright information and a 'Feedback' button.

MASCOT : *Hosting Mascot Server in 'The Cloud'* © 2015 Matrix Science 

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

Please type in your contact details to finish creating the licence.

- If any of the product and licence details shown below are incorrect, please contact Matrix Science support.
- Required fields are marked with an asterisk*.

The licence file will be sent to the email address you supply below. Make sure your spam filter does not reject this message by adding support@matrixscience.com to your "whitelist" or address book.

Product key 3T0R-3UTT-8AVY-8208-Y5T5
Product family Mascot Server
Licence type Temporary; expiry date is 2015-05-07.
Licensed features Version 2.5
 1 CPUs
 Top down searches

Email address* jcottrell@matrixscience.com
Verify email address* jcottrell@matrixscience.com
Full name* John Cottrell
Organisation* Matrix Science
Address line 1* 64 Baker Street
Address line 2
City* London
Zip/postcode W1U 7GB
State/region Greater London
Country* United Kingdom
Phone number 442074861050
Fax number 442072241344

If you are installing the software on behalf of the

MASCOT : Hosting Mascot Server in 'The Cloud' © 2015 Matrix Science **MATRIX SCIENCE**

Fill in the usual stuff.

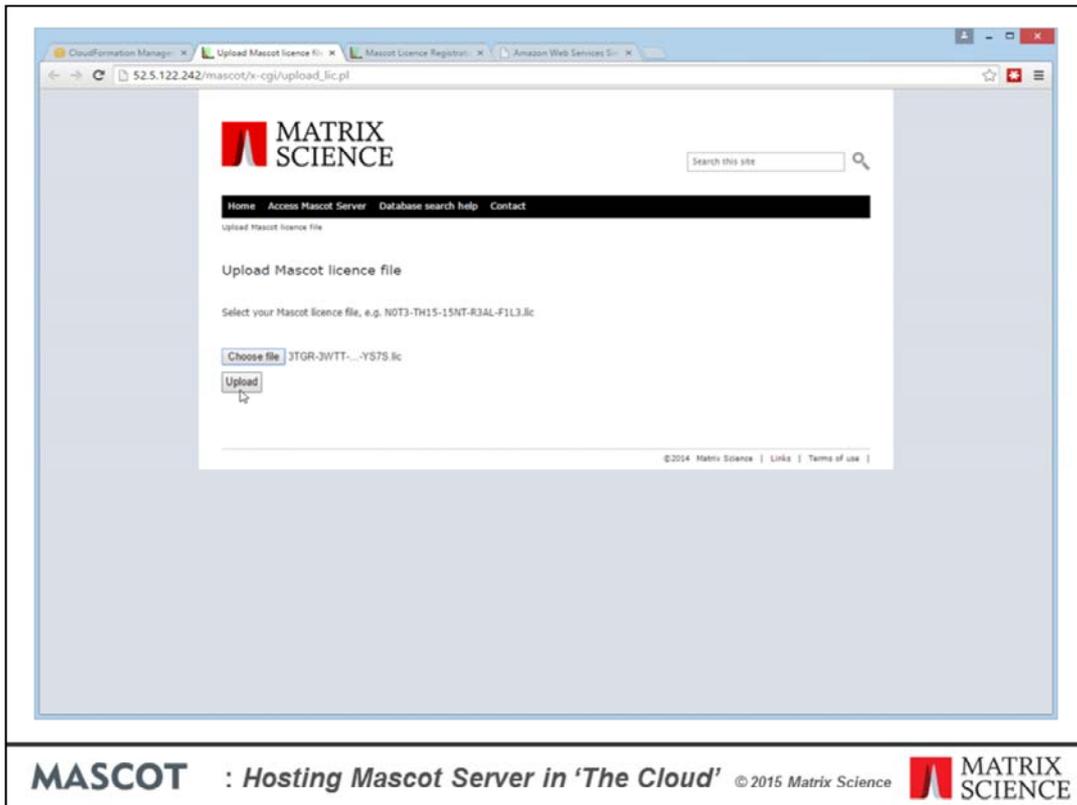
The screenshot shows the AWS CloudFormation console interface. At the top, there are navigation tabs for 'Create Stack', 'Update Stack', and 'Delete Stack'. Below this, a table lists the stack 'Brian' with a status of 'CREATE_COMPLETE'. The 'Outputs' tab is selected, displaying a table with the following data:

Key	Value	Description
Home	http://52.5.122.242/mascot	Mascot Server home page
Status	http://52.5.122.242/mascot/x-cgims-status.exe	Database Status
Register	http://52.5.122.242/mascot/x-cgims-status.exe?Show=REGPRO DUCTKEY	Register product key
Upload	http://52.5.122.242/mascot/x-cgi/upload_lc.pl	Upload licence file

At the bottom of the console, there is a footer with copyright information: '© 2008 - 2015, Amazon Web Services, Inc. or its affiliates. All rights reserved.' and a 'Feedback' button.

MASCOT : *Hosting Mascot Server in 'The Cloud'* © 2015 Matrix Science 

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



And upload. Once this is done

The screenshot shows the AWS CloudFormation console interface. At the top, there are buttons for 'Create Stack', 'Update Stack', and 'Delete Stack'. Below these, a table lists the stack 'Brian' with a status of 'CREATE_COMPLETE'. The 'Outputs' tab is selected, displaying a table of stack outputs:

Key	Value	Description
Home	http://52.5.122.242/mascot	Mascot Server home page
Status	http://52.5.122.242/mascot/x-cgims-status.exe	Database Status
Register	http://52.5.122.242/mascot/x-cgims-status.exe?Show=REGPRO DUCTKEY	Register product key
Upload	http://52.5.122.242/mascot/x-cgi/upload_lic.pl	Upload licence file

A tooltip is visible over the 'Status' output, showing the URL and a 'Database Status' link. The footer of the console contains copyright information and a 'Feedback' button.

MASCOT : *Hosting Mascot Server in 'The Cloud'* © 2015 Matrix Science 

We can follow the link to Mascot Database Status

MASCOT search status page

Version: 2.5.1 - Matrix Science (3TGR-3WTT-SAVY-SZG8-YS7S) [Licence info](#)
 8 logical, 1 physical Intel processors (hyper-threading enabled, quad core). CPUs: 0 1 2 3 4 5 6 7 available, using: 0 1 2 3 4 5 6 7. [0 searches running]

[Search log](#) [monitor log](#) [error log](#) [Error message descriptions](#) [Do not auto refresh this page](#)

```

Name = contaminants Family = /usr/local/mascot/sequence/contaminants/current/contaminants_*.fasta
Filename = contaminants_20120713.fasta Pathname = /usr/local/mascot/sequence/contaminants/current/contaminants_20120713.fasta
Status = In use Statistics
State Time = Tue Apr 7 14:36:56 # searches = 0
Mem mapped = YES Request to mem map = YES Request unmap = NO Mem locked = NO
Number of threads = -1 Current = YES

Name = cRAP Family = /usr/local/mascot/sequence/cRAP/current/cRAP_*.fasta
Filename = cRAP_20150130.fasta Pathname = /usr/local/mascot/sequence/cRAP/current/cRAP_20150130.fasta
Status = In use Statistics
State Time = Tue Apr 7 14:36:55 # searches = 0
Mem mapped = YES Request to mem map = YES Request unmap = NO Mem locked = NO
Number of threads = -1 Current = YES

Name = Fungi_EST Family = /usr/local/mascot/sequence/Fungi_EST/current/Fungi_EST_*.fasta
Filename = Fungi_EST_123.fasta Pathname = /usr/local/mascot/sequence/Fungi_EST/current/Fungi_EST_123.fasta
Status = In use Statistics Unidentified taxonomy
State Time = Tue Apr 7 14:36:56 # searches = 0
Mem mapped = YES Request to mem map = YES Request unmap = NO Mem locked = NO
Number of threads = -1 Current = YES

Name = Human_EST Family = /usr/local/mascot/sequence/Human_EST/current/Human_EST_*.fasta
Filename = Human_EST_123.fasta Pathname = /usr/local/mascot/sequence/Human_EST/current/Human_EST_123.fasta
Status = In use Statistics Unidentified taxonomy
State Time = Tue Apr 7 14:36:56 # searches = 0
Mem mapped = YES Request to mem map = YES Request unmap = NO Mem locked = NO
Number of threads = -1 Current = YES

Name = Hus_EST Family = /usr/local/mascot/sequence/Hus_EST/current/Hus_EST_*.fasta
Filename = Hus_EST_123.fasta Pathname = /usr/local/mascot/sequence/Hus_EST/current/Hus_EST_123.fasta
Status = In use Statistics Unidentified taxonomy
State Time = Tue Apr 7 14:36:56 # searches = 0
Mem mapped = YES Request to mem map = YES Request unmap = NO Mem locked = NO
Number of threads = -1 Current = YES

Name = NCBItr Family = /usr/local/mascot/sequence/NCBItr/current/NCBItr_*.fasta
Filename = NCBItr_20150401.fasta Pathname = /usr/local/mascot/sequence/NCBItr/current/NCBItr_20150401.fasta
Status = In use Statistics Unidentified taxonomy
State Time = Tue Apr 7 14:36:56 # searches = 0
  
```

MASCOT : *Hosting Mascot Server in 'The Cloud'* © 2015 Matrix Science 

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 threads of popular databases, such as NCBItr and SwissProt.

Mascot Server in Amazon EC2

- Convenience
- Cost
- Performance
- Security

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.

Cost - Instance types

- **On demand instance**
 - Highest hourly rate
 - Need only pay for actual running hours
- **Reserved instance**
 - Discount up to 60% if commit to 3 years and pay up front
- **Spot instance**
 - The market rate
 - Availability not guaranteed
 - Cannot be stopped and started

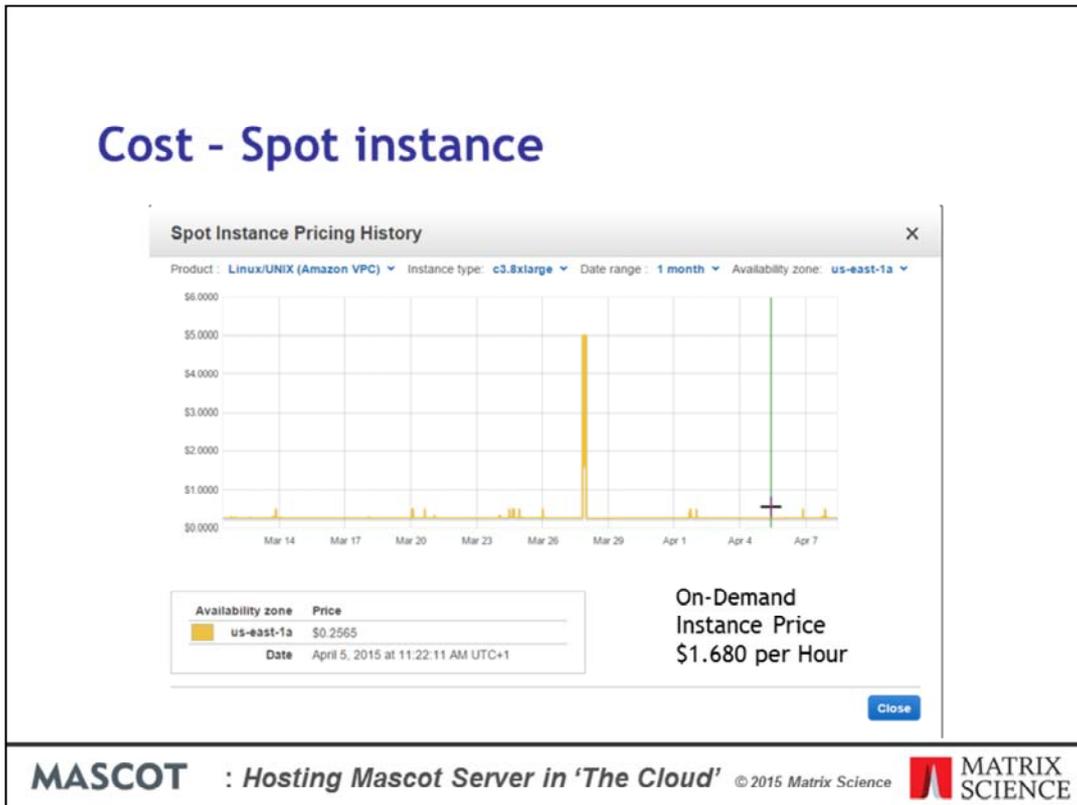
This is where it gets slightly complicated. Rather like buying a cell phone or cable TV, there are so many choices that it becomes difficult to make a meaningful comparison. There are three different ways of buying time on Amazon's virtual servers:

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.

Cost - Spot instance



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.

Cost - Comparison over 3 year term

- **US East, Linux, prices as of April 2015**
 - Windows significantly more expensive
 - Storage charges relatively small
 - Inbound data transfer is free

Mascot CPU	EC2 instance	On demand 24/7	On demand 9/5	Reserved	Spot (approx)
1	m3.2xlarge	14,717	3,931	5,493	2,000
2	c3.4xlarge	22,075	5,897	8,265	3,500
4	c3.8xlarge	44,150	11,794	16,506	7,000

MASCOT : *Hosting Mascot Server in 'The Cloud'* © 2015 Matrix Science 

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 NCBIInr 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



MASCOT : Hosting Mascot Server in 'The Cloud' © 2015 Matrix Science 

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 - I/O speed

Database Manager tasks

Database Manager processes scheduled tasks one at a time in the background as long as Mascot (ms-monitor.exe) is running.

Task processing is currently **enabled**. Task queue and currently running tasks are shown below. **▶ Task process control**

No tasks in queue.

Title	Started	Progress	
Download files for NCBIInr <small>dbman_download.pl NCBIInr</small>	Fri Apr 10 09:30:01 2015	[9.4%] Downloading <ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nr.gz> (try 1 out of 5): 38.01 MB/s (956 MB/14.6 GB, 6m 7s remaining)	<input type="button" value="Cancel"/>

Task log (most recent 25 entries)

Process end time	Command	Last message
[Tue Dec 2 12:19:31 2014]	dbman_download.pl Trembl2	(success) 'Trembl2' successfully updated.
[Tue Dec 2 12:03:01 2014]	dbman_download.pl Trembl2	(success) Downloaded files and decompressed version file; user must now select version
[Tue Dec 2 09:46:12 2014]	dbman_download.pl Trembl1	(success) 'Trembl1' successfully updated.
[Thu Nov 13 14:23:33 2014]	dbman_download.pl SwissProt	(success) 'SwissProt' successfully updated.

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 NCBIInr Fasta file

Security

- Physical security

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.

Security

- Physical security
- **Hacking**

Amazon Web Services • Overview of Security Processes June 2014

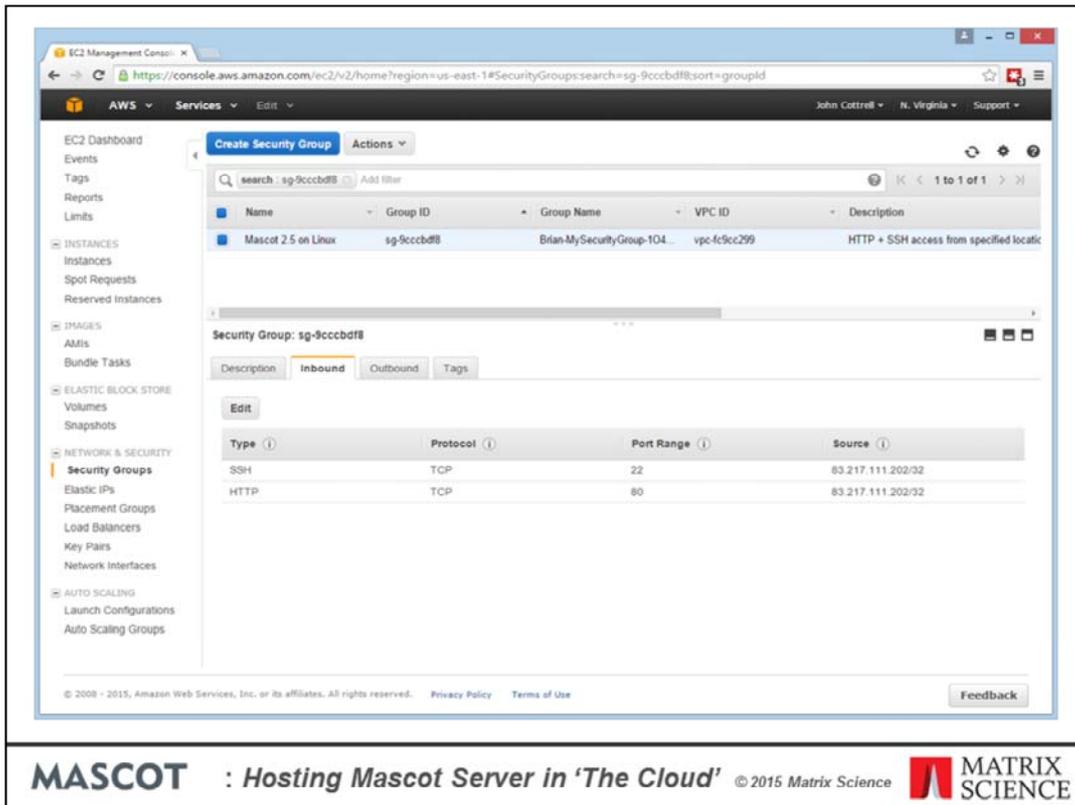


Amazon Web Services: Overview of Security Processes
June 2014

(Please consult <http://aws.amazon.com/security/> for the latest version of this paper)

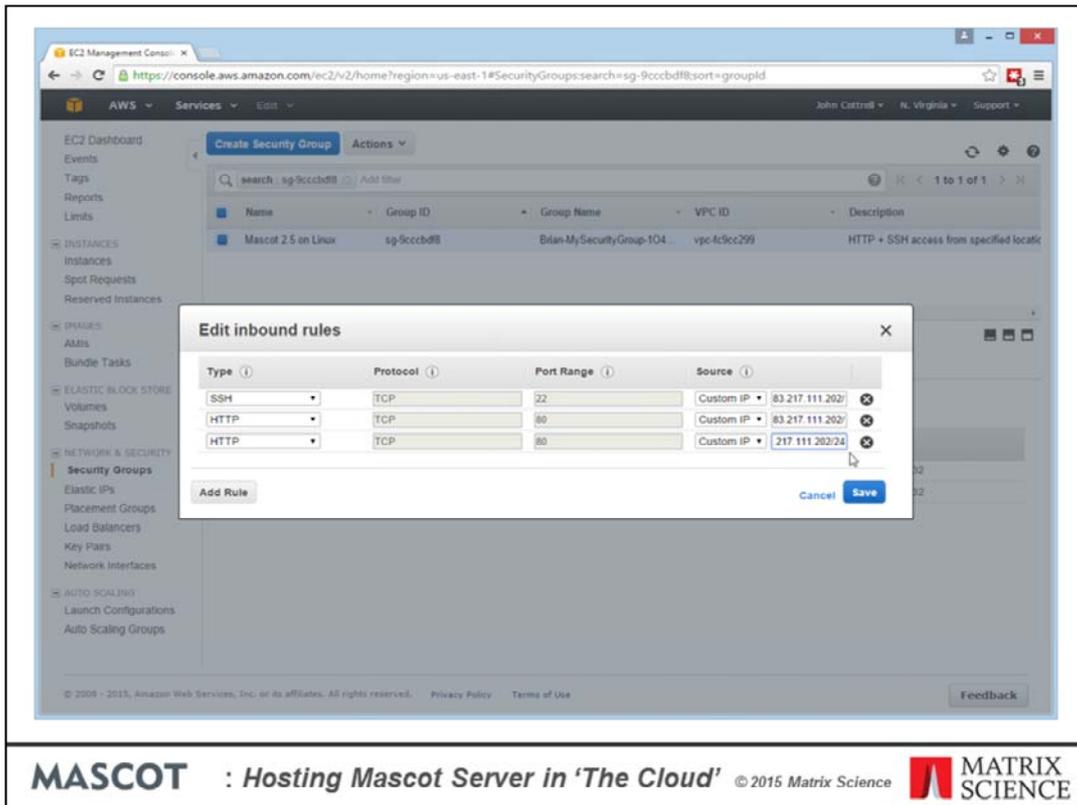
MASCOT : *Hosting Mascot Server in 'The Cloud'* © 2015 Matrix Science  **MATRIX SCIENCE**

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



MASCOT : Hosting Mascot Server in 'The Cloud' © 2015 Matrix Science 

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'* © 2015 Matrix Science  **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

Mascot Server in Amazon EC2

- **Convenience**
 - Excellent for On demand and reserved instances
 - Good for spot instances
- **Cost**
 - May be higher or lower than in-house hardware, (depends on your overheads for power, rack space, etc.)
- **Performance**
 - Very good
- **Security**
 - Usually better than in-house

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

Give it a try

- Sign up to Amazon web services
- Email support@matrixscience.com
- We'll email instructions plus a 30 day product key
- If you like it, buy a permanent licence or request a transfer of an existing licence

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.