2014 Workshop on Genomics

Starting your Amazon virtual machine

Objectives:

By the end of this section you will be expected to:

- Log into the Amazon Console and start your instance of the Evomicss workshop AWS Image
- Log in to an Amazon EC2 Linux instance using your own computer
- Continue with the Short read genomics tutorials at your own pace

Introduction

For this workshop we will provide an overview of the cloud as described by Amazon and how, as researchers, we can use this flexible resource to get work done quickly and relatively inexpensively. We will dive right into starting up the pre-prepared Virtual Machine (VM) and logging into the Amazon management console. We will give you a whirlwind tour of the features of Amazon's cloud and then get you to start your AMI and log-in via the NX-client.

For this tutorial I borrowed documentation from the following sites:

· http://ged.msu.edu/angus/tutorials/unix-and-ssh-and-scp.html
· http://aws.amazon.com/documentation

Task 1 – Tour of Amazon's Cloud

In this section of the workshop we will log into Amazon's cloud (referred to as Amazon Web Services or AWS) and take a look at the various services offered by Amazon. These include:

- Elastic Cloud Compute (EC2): the service AWS is known for. It enables you to rent Linux and Windows machines by the hour. Amazon now has also special High Performance Computing nodes (HPC) and Graphical Computing nodes (GPU nodes)

- Simple Storage Service (S3): a storage service, not particularly fast but great for storing large “buckets” of data for long term storage, sharing, or temporary storage for use between instances

- Elastic Block Storage (EBS): similar to S3 but limited in size (max 1TB), these are virtual hard drives that you can attach and detach very quickly to and from your running instances. Think of these as the USB flash drive of the cloud computing world

- A ton of other services that are geared towards building highly scalable and fault-tolerant web-based services. Many can be co-opted for use in research!

We have set up user accounts for each student in the class, take a look at the form given to you in your welcome packet. You need to know the console URL, your username, and your password. You will also find the access key and secret access key, both are used by various tools to log in and use the Amazon cloud programmatically.

Task 2 - Connecting to Your Personal VM

The Rules

We ask that each student adhere to the following guidelines to ensure we have enough resources for the duration of the workshop:

1. Please only launch a **single** VM instance of the type specified by the instructor at the beginning of the workshop and “**stop**” the instance at the end of the workshop. Do not leave the image running overnight or over a whole week – you will quickly run out of funding

2. Please name your VM instance. Including your name will make it easy to find your resources in the list of class resources

3. Please do not delete EBS volumes or terminate VM instances that don't belong to you

Step 1 – Logging Into the Console

In addition to being extremely comprehensive, the Amazon cloud has a very easy-to-use interface for interacting with all their cloud offerings. All you have to do is log into a web application and most of the functionality of the Amazon tools are available for you and very easy to use.

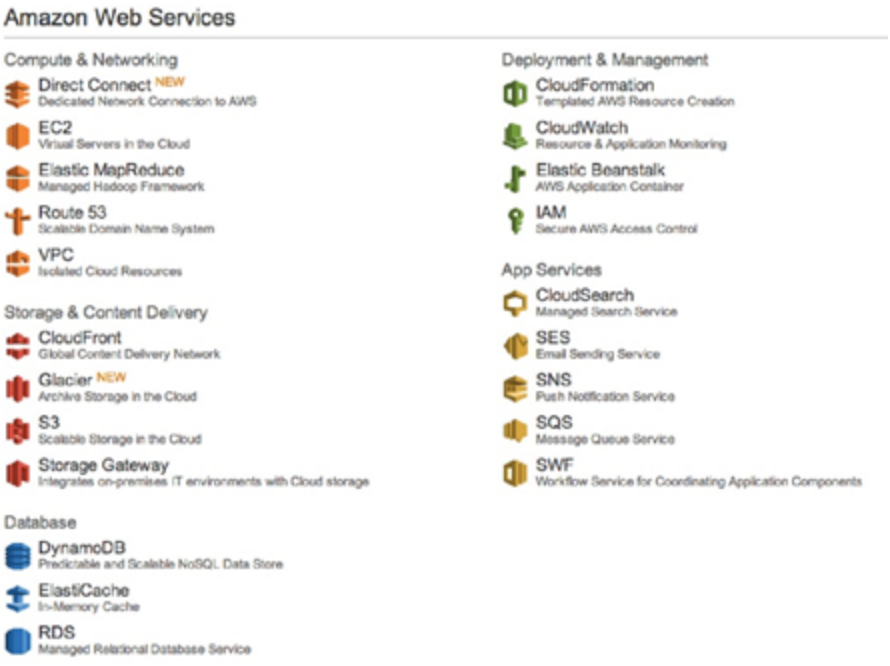
This workshop has its own Amazon account and each of you have been setup as a sub-account using something called Identity Access Manager. The nice thing about this is you have pretty much free access to the console and we can have very fine grain control on what your sub-accounts can and cannot do.

To get started **go to the following URL** and **login** with the username and password below.

https://evomics.signin.aws.amazon.com/console
or
http://bit.ly/evomics2014 (this is the same URL, it's just a little easier to remember :))

Username: student
Password: evomics

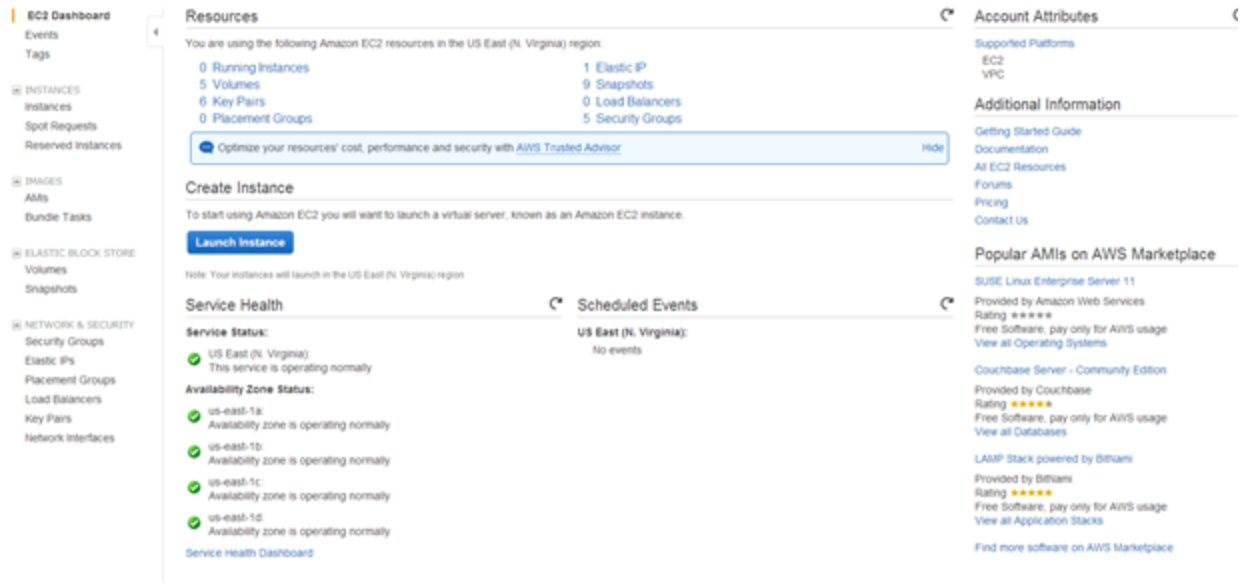
After logging in you'll be presented with a wide range of options. Click on "EC2" under "Compute & Networking".



On this page you'll get a summary of the EC2 state for your account (EC2 Management

Console). You can see mine below:

Step 2 – Clicking 'Launch Instance'



From here we can create computers on Amazon's 'cloud'. What this means is that we can create as many computers as we like, start them, log-in to them, do some work, transfer data to/from them or destroy them altogether. Amazon worry about the hardware, power, cooling and maintenance – all we need to do is specify how powerful a computer we want (micro, small, large or extra-large).

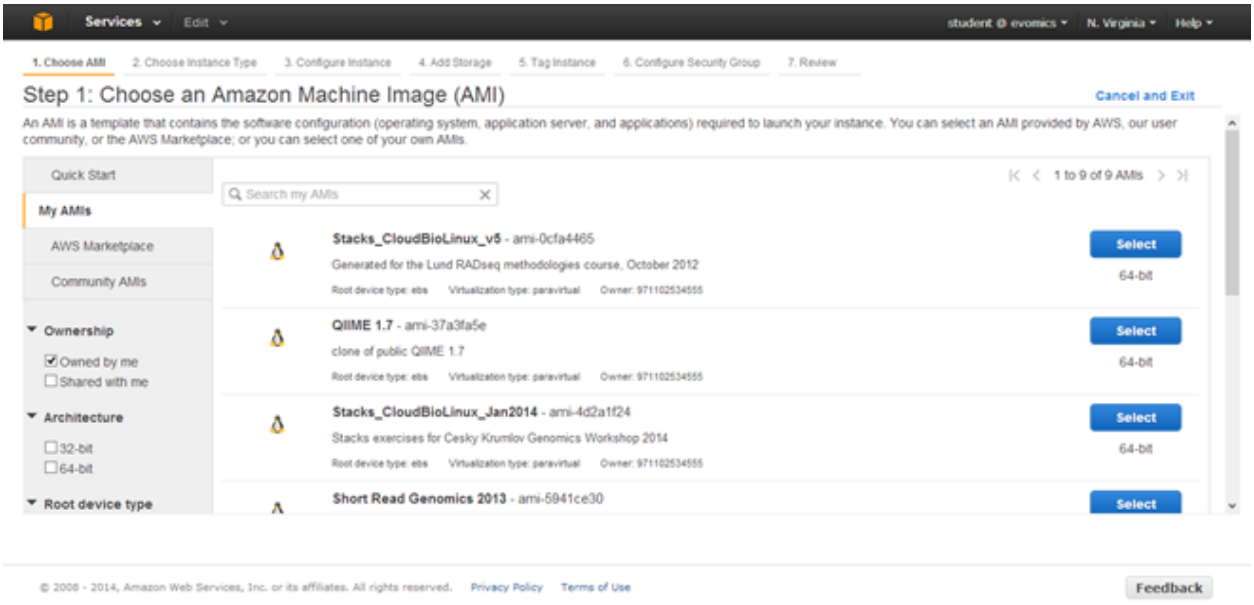
Amazon charge for each Gb stored every month and for each hour a machine is run. This can vary from a few cents per hour to a few dollars. Whilst it is very convenient if you are only doing analyses occasionally, at the moment it is still cheaper to have your own computing (although you then have the headache of maintaining it!).

The reason we are using the cloud here is that most of you will only be doing occasional analyses. In the case of high-throughput sequencing data (e.g. Illumina) you will find that your desktop PC may not be powerful enough to cope with the data. As such Amazon can offer a good alternative. It also means that you can start and stop your AMI from home and continue to work through the tutorial from there.

Once logged into the console we can select an Amazon Machine Image (AMI) Virtual Machine and launch it.

Click on the "Launch Instance" button in the centre.

Then **click/select ‘My AMIs’ to continue**.



If you don’t see the above AMI’s, make sure the 'My AMIs' tab is selected.

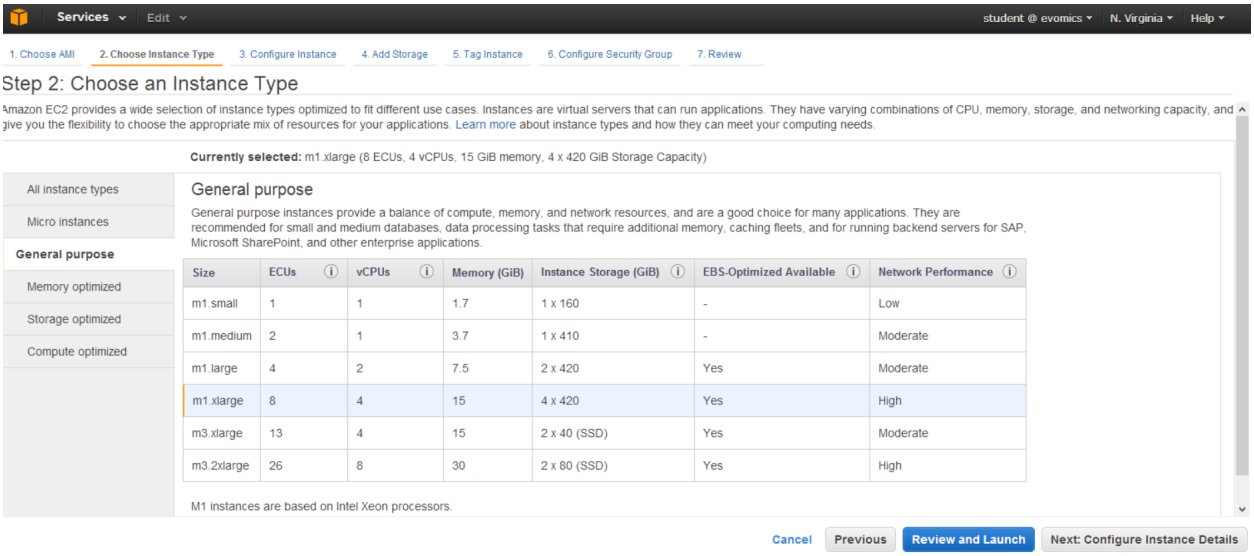
The next step in this process is to select an AMI. The AMI is a “snapshot” of the VM we prepared for you for this class. To select the VM we have prepared for this workshop –

Click “Select” next to “WORKSHOP ON GENOMICS 2014 V1”

Please do not use any other AMI for this workshop!

Once you have selected the instance you will need to make sure the ‘General Purpose’ tab on the left is selected. This will allow you to see the different Instance Types available. For this class **your choice should be a large instance (m1.xlarge)**. Depending on how many CPUs and how much memory you need you could choose a different instance if you were running this outside the class.

See http://aws.amazon.com/ec2/instance-types/ for information about the instance types and http://aws.amazon.com/ec2/pricing/ for information about pricing per hour.

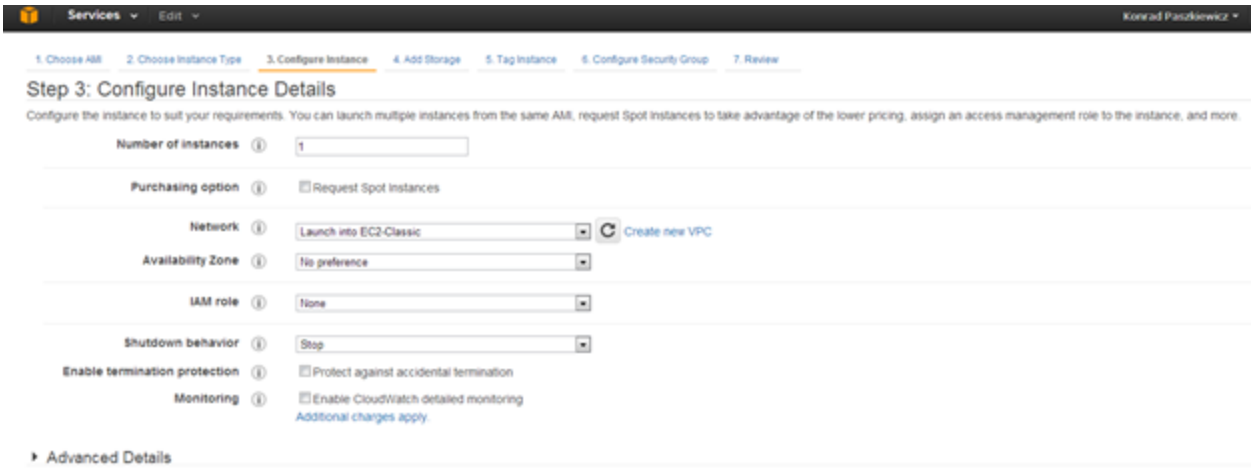


Choose the "General Purpose" tab to left.

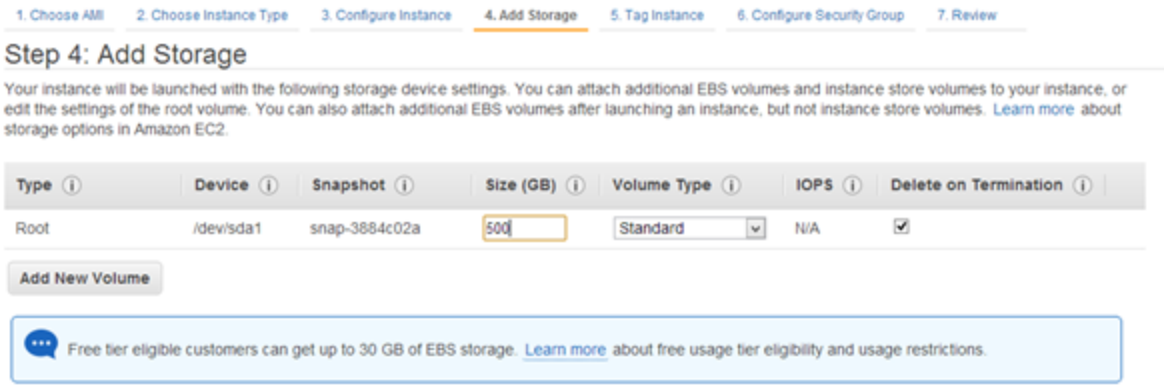
Select the m1.xlarge instance type - that's the *extra* large instance...

Once you have selected the m1.xlarge instance type, click "Next: Configure Instance Details"

We are now on a page which enables us to customise the instance if we so wish. We don't need to change anything so…



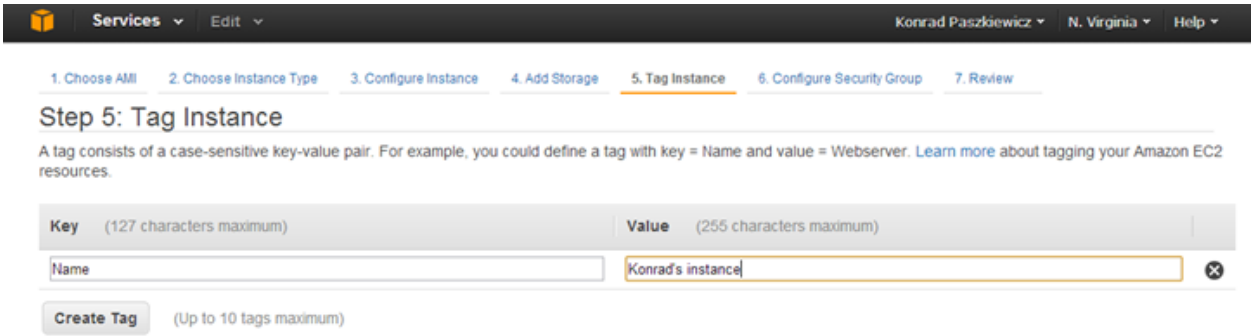
Click on "Next: Add Storage"



On this page we can select how much storage we want to add to our instance. Here we've selected the default of 500 Gb. Note that we have also ticked the 'Delete on Termination' box. This deletes the virtual hard drive once the instance is terminated. In real life I would recommend against this as you could easily lose valuable data. However for the purposes of the workshop, it makes management easier so we'll select it.

Select the 'Delete on Termination' check-box and then

Click on "Next: Tag Instance"

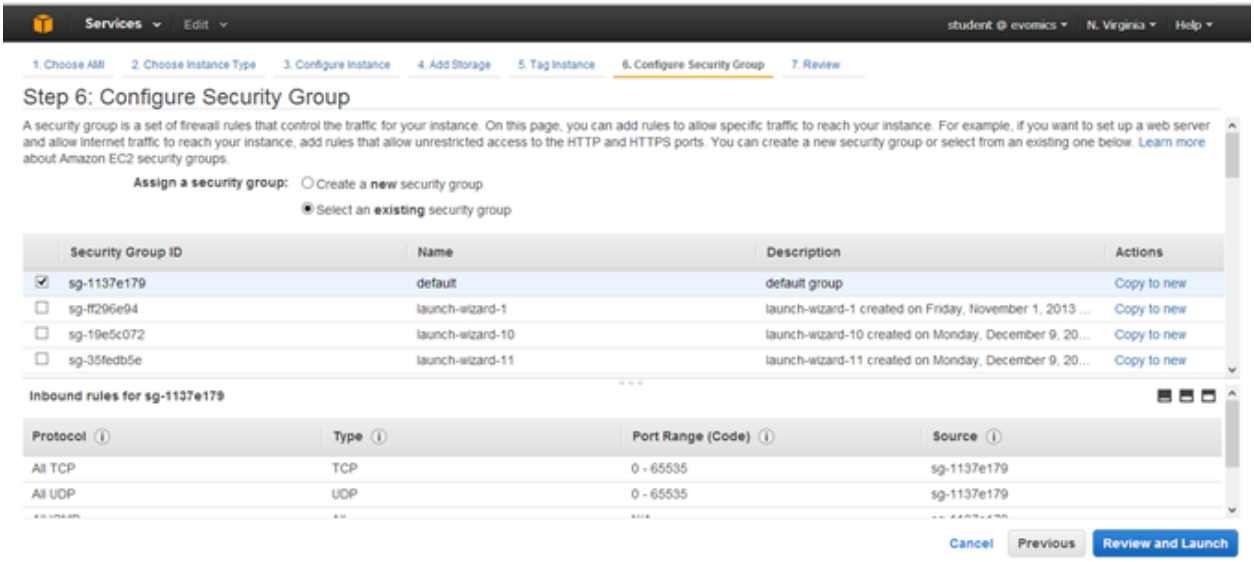


The idea of a 'tag' is that if you have multiple instances you can create tags to identify them. As we are all using a single account, it is important to be able to identify your instance.

In the 'Value' column next to 'Name' make sure you give the instance a name which includes your name so that you can identify it

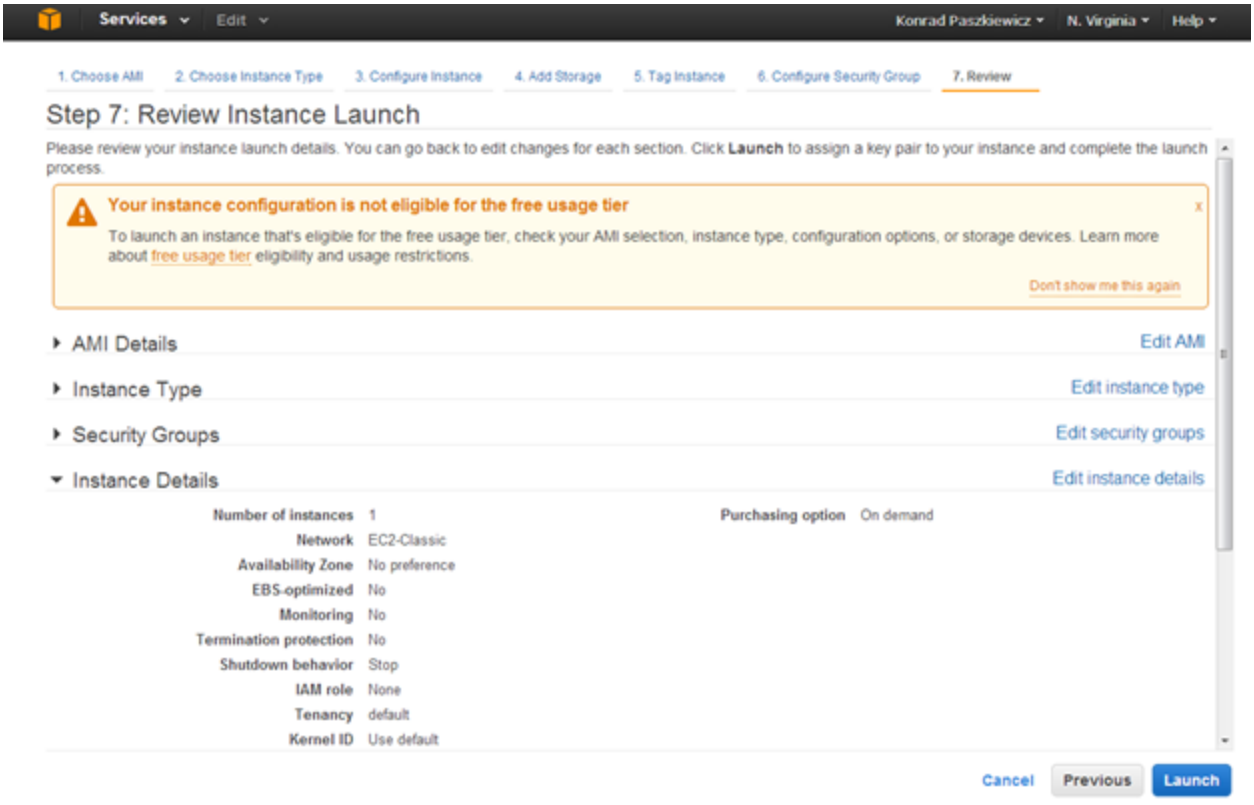
Then click on 'Next: Configure Security Group'

Click on 'Select an existing security group' and select 'default'.



Then click on ‘Review and Launch'

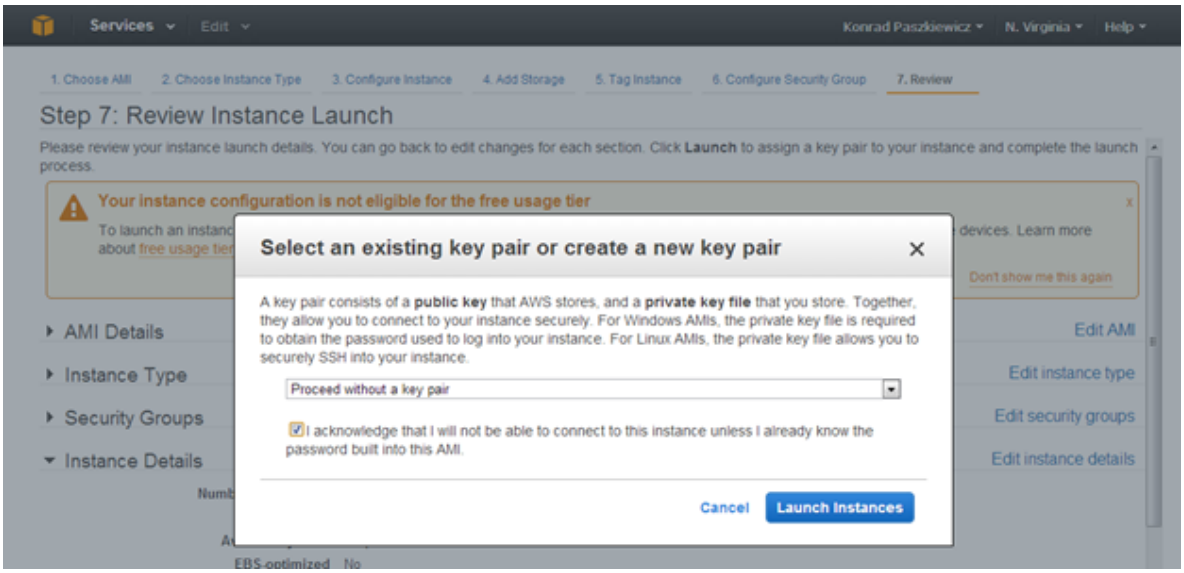
The next step is to review and launch the instance and set up any access keys

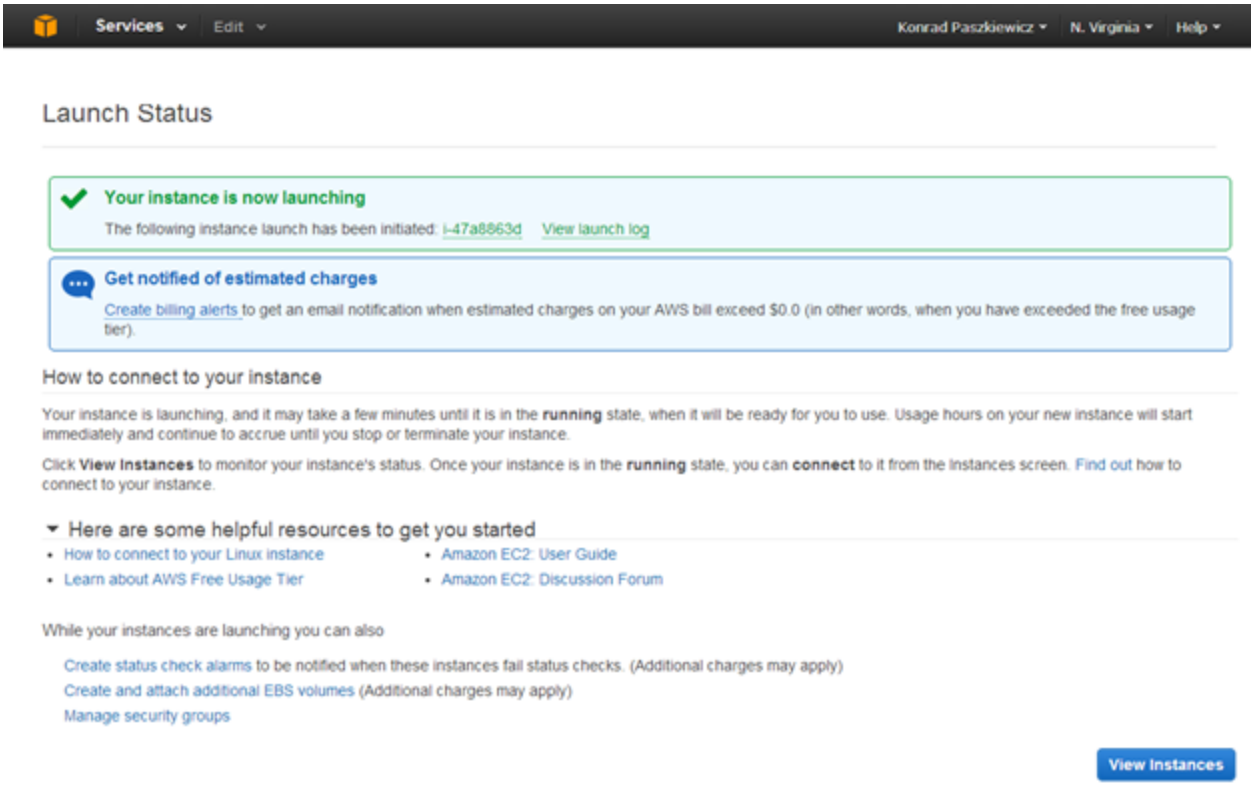


Then click on ‘Launch’

The final step is to select the “keypair” used to let you log into this machine. This keypair is a file you download and supply to your SSH client in order to connect to the running server. Our instance does not need this so,

Select “Proceed without a Key Pair” and tick the check box. Click ‘Launch Instances’

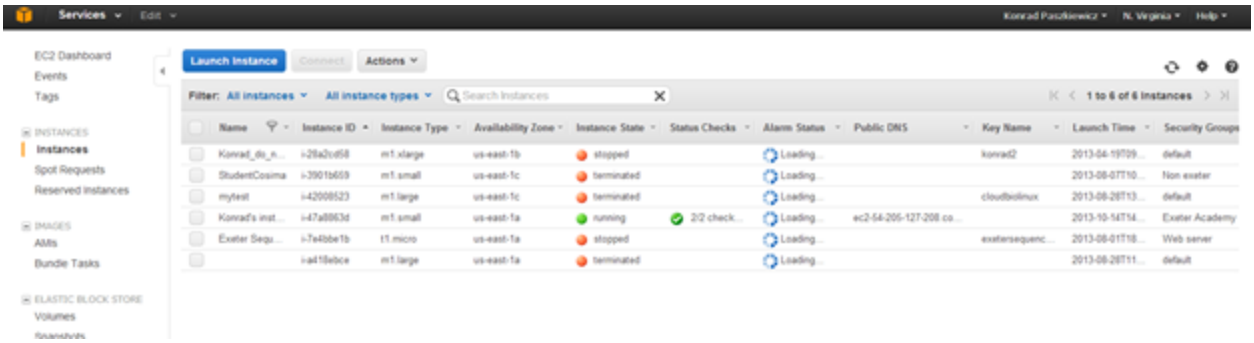




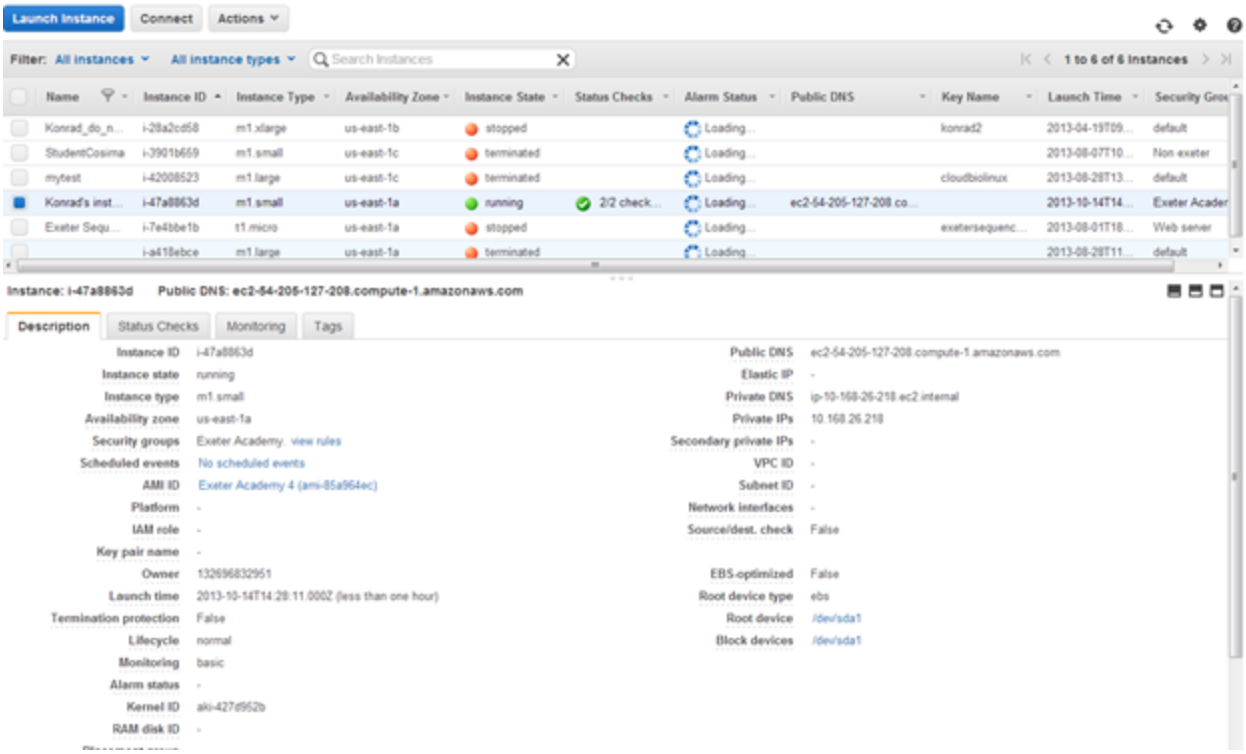
At this point, click **“View Instance”** and you can watch your instance boot.

At this point you wait just a couple minutes for the AMI instance to come online.

Above you can see the instance is running, give it a couple minutes to finish its boot cycle. It's booting somewhere on a virtualized cluster node in Virginia!



Once it turns **green**, you should **click on the AMI** and **copy the Public DNS address to the clipboard** – (shown below is an example where the Public DNS is ec2-54-205-127-208.compute-1.amazonaws.com). We will need it in a moment.



Step 3 – Log into the Running VM's Remote Desktop with NoMachine

Whilst your AMI is initializing - please note it may take some time (~15 minutes) - take this opportunity to install the NoMachine (sometimes referred to as NX) client software you will need to connect to the AMI.

This will allow you to see a windowing environment (like your Desktop) rather than just an terminal! This is a great option if you want to use a GUI application (Graphical User Interface like Apollo or Artemis). It's very cool to see a remote desktop with Firefox and every other GUI application rendered quick and snappy over the Internet!

Here are the steps to get remote NoMachine login working… *Note that these instructions will only work for this workshops' particular VM AMI. Many AMIs will not have the NX server installed and therefore you will not be able to connect using the NX-client. In these cases you will have to look at Step 4 and use SSH. But you should not need to do that for these tutorials.*

1. Download an NX Client

Use the following NX clients depending on the operating system:

Mac OSX: https://www.nomachine.com/download/download&id=15

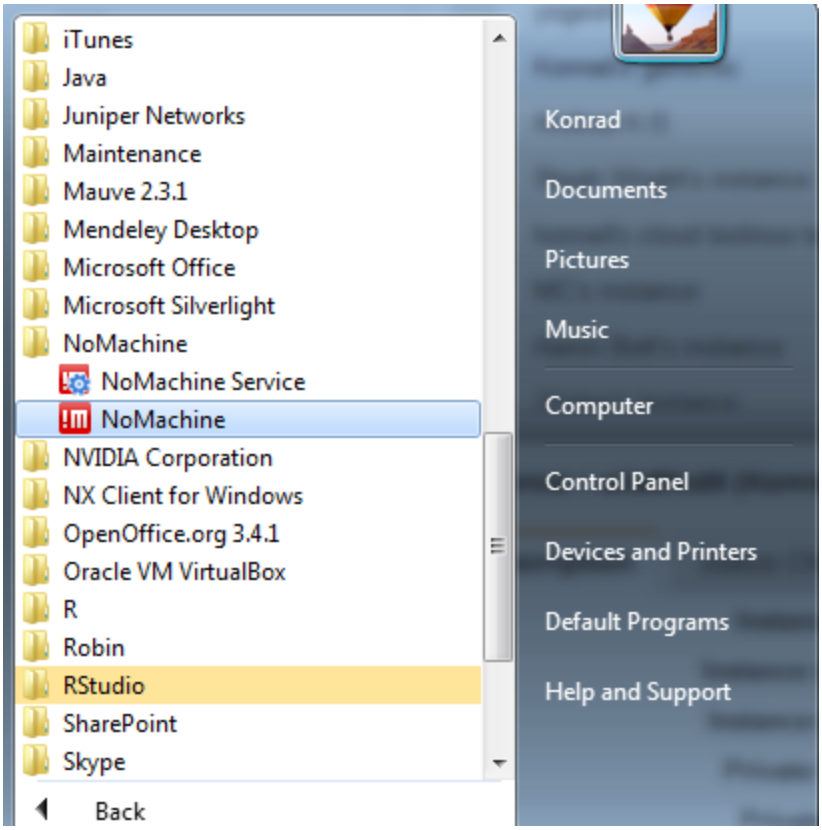
Windows 7 & Windows 8: https://www.nomachine.com/download/download&id=16

Windows XP: You may need an old version called NX 3.5 http://nx-client-for-windows.software.informer.com/3.5/ - please try the Win7 version first, if you cannot get it to work, please ask an instructor to help with the settings of the old version as it looks somewhat different to the instructions below….

Linux: https://www.nomachine.com/download/linux&id=4 - Debian (Ubuntu etc) should install this via 'sudo dpkg -i nomachine.deb' and not from the source install.

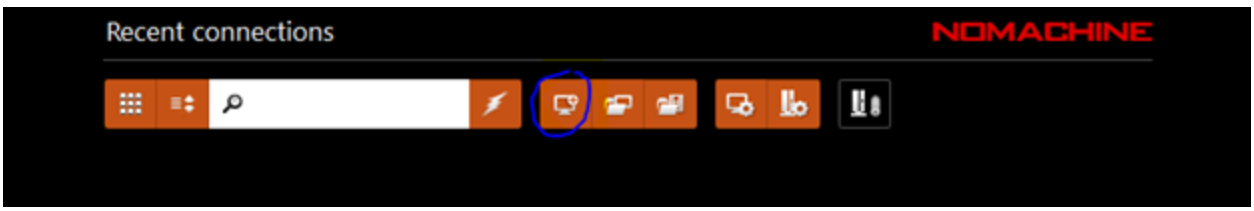
2. Launch the Client and Connect

Once it is installed, start the NoMachine on your laptop. On Windows the start menu the icon will look like:



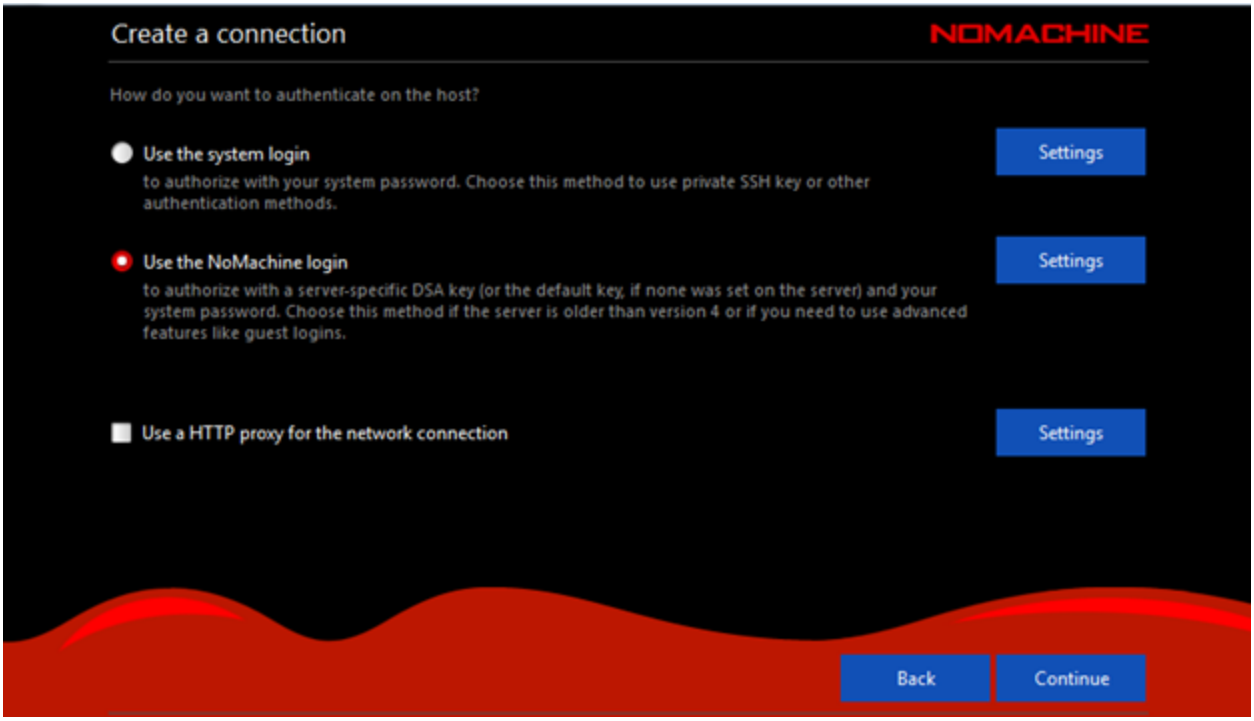
On Mac OSX it should be in your applications folder and for Ubuntu the NoMachine client will be in your menu system (e.g. Unity)

At the next screen you may get some introductory screens if this is the first time you have run this application. Click through these until you see the main screen and make sure to check the box so that you don't see them again...:

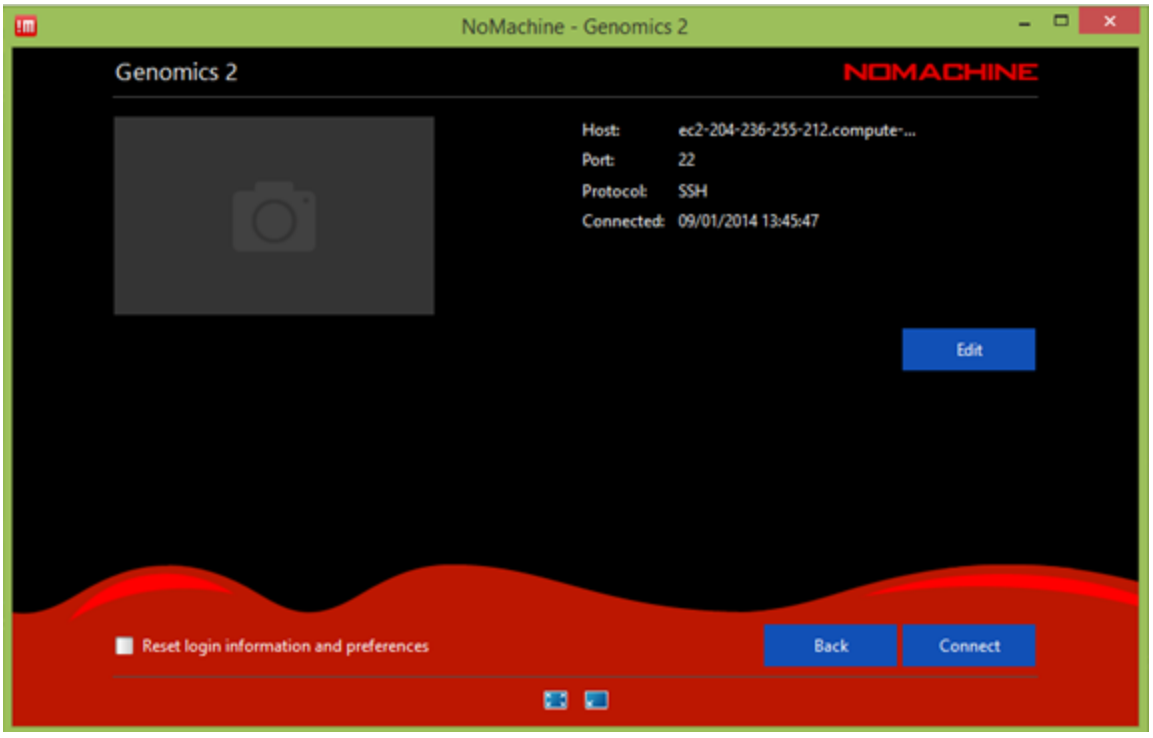


Click on the highlighted icon:

You should name the connection Genomics Workshop and set the host-name to be the public DNS address of your Amazon AMI (copied from earlier). Also, ensure that the Protocol is set to SSH and that the port is set to 22. Now Click on 'Advanced'.

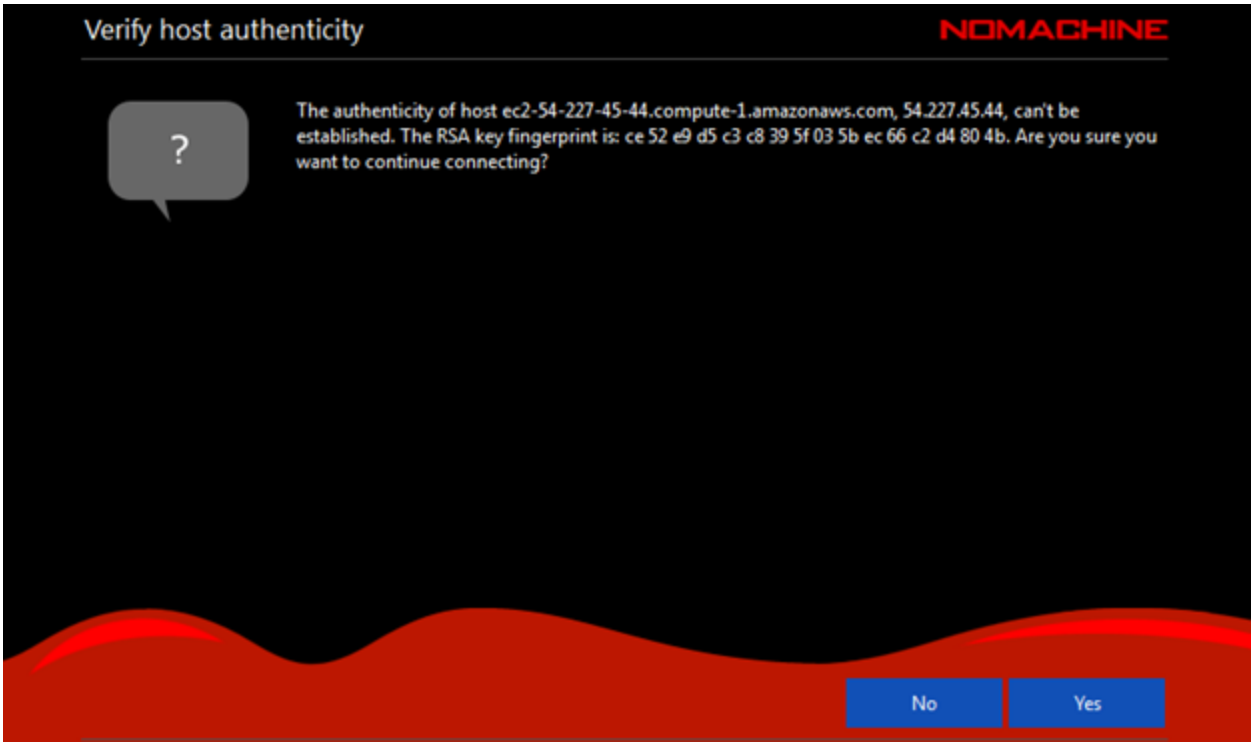


Set the connection to use the ‘NoMachine login’ and ‘Click Continue’.

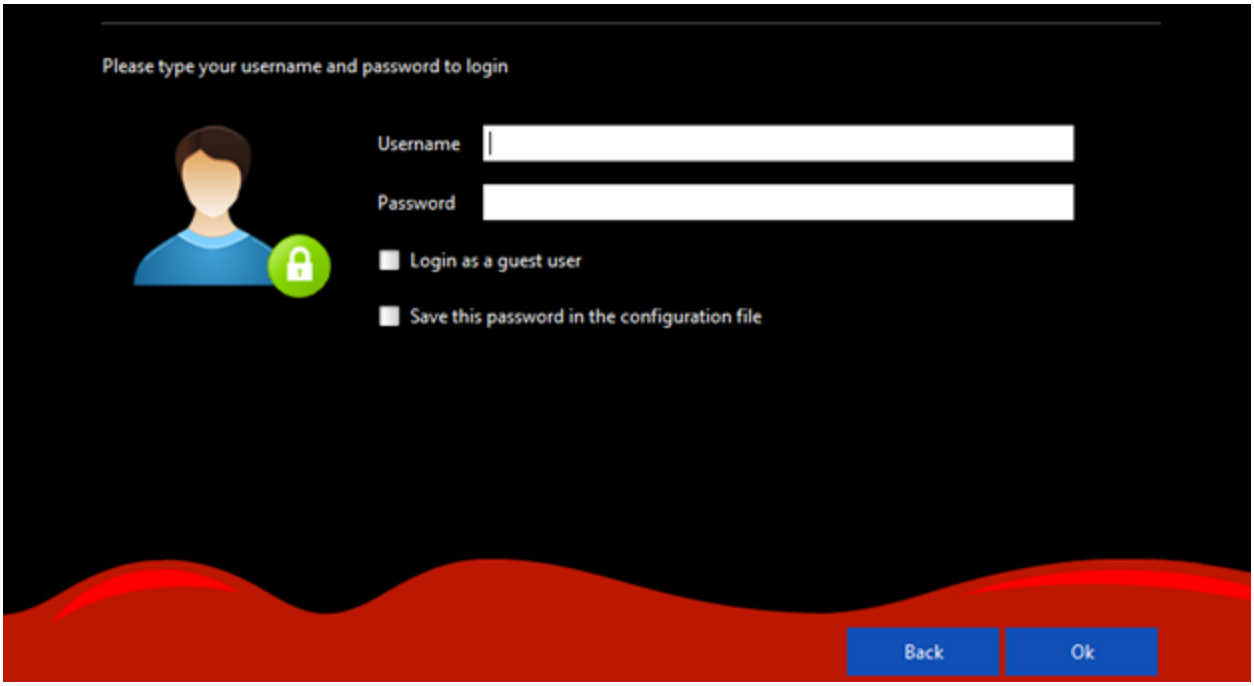


Click “Connect”.

You may find that you get a screen like this:

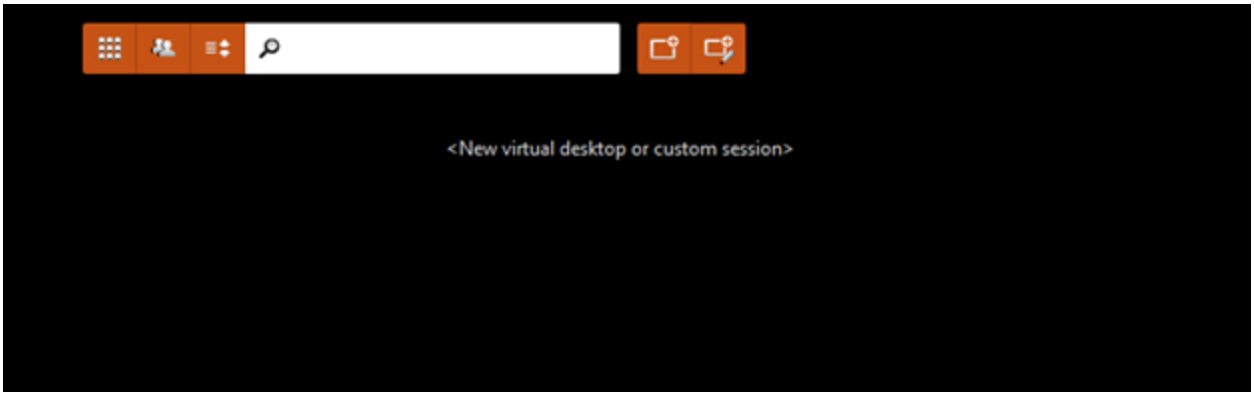


This is just SSH telling you that it hasn't connected to this server before and cannot be sure that someone isn't pretending to be your server. Click "Yes".



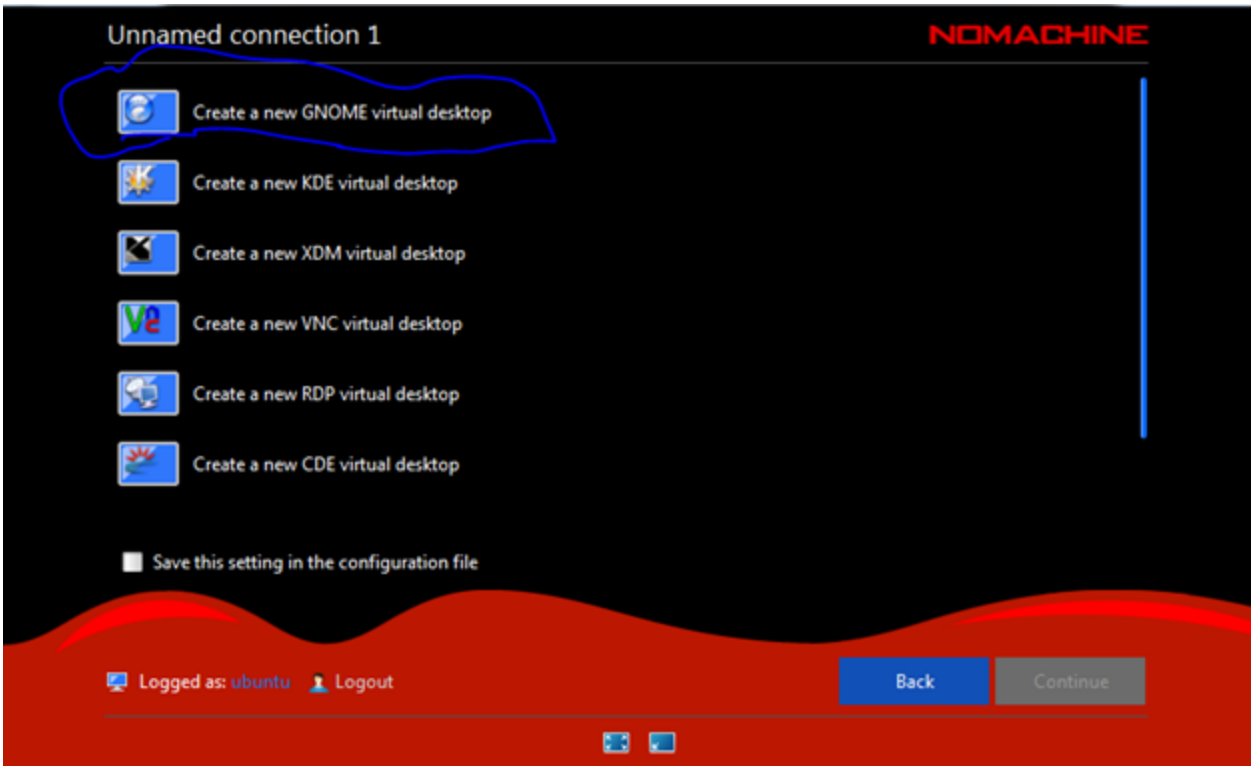
Enter the login username "ubuntu" and password "evomics" (without the quotation marks)

At the next screen you should see the option to start a ‘New virtual desktop or custom session’. Click on this.



If you do not get to this screen, wait a little time as the AMI might not have initialized just yet, double check your settings in the meantime and try again. If you are still having issues then please alert an instructor… Don’t panic!

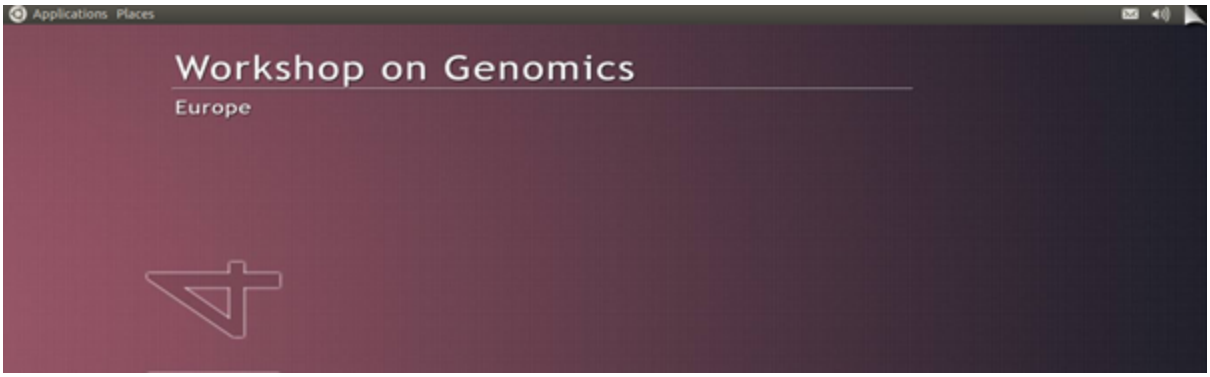
Select “Create a new GNOME virtual desktop”:



Click ‘Next >’ and ‘Next >’ again on the next screen. Double click on the NX shortcut which will be on your Desktop.

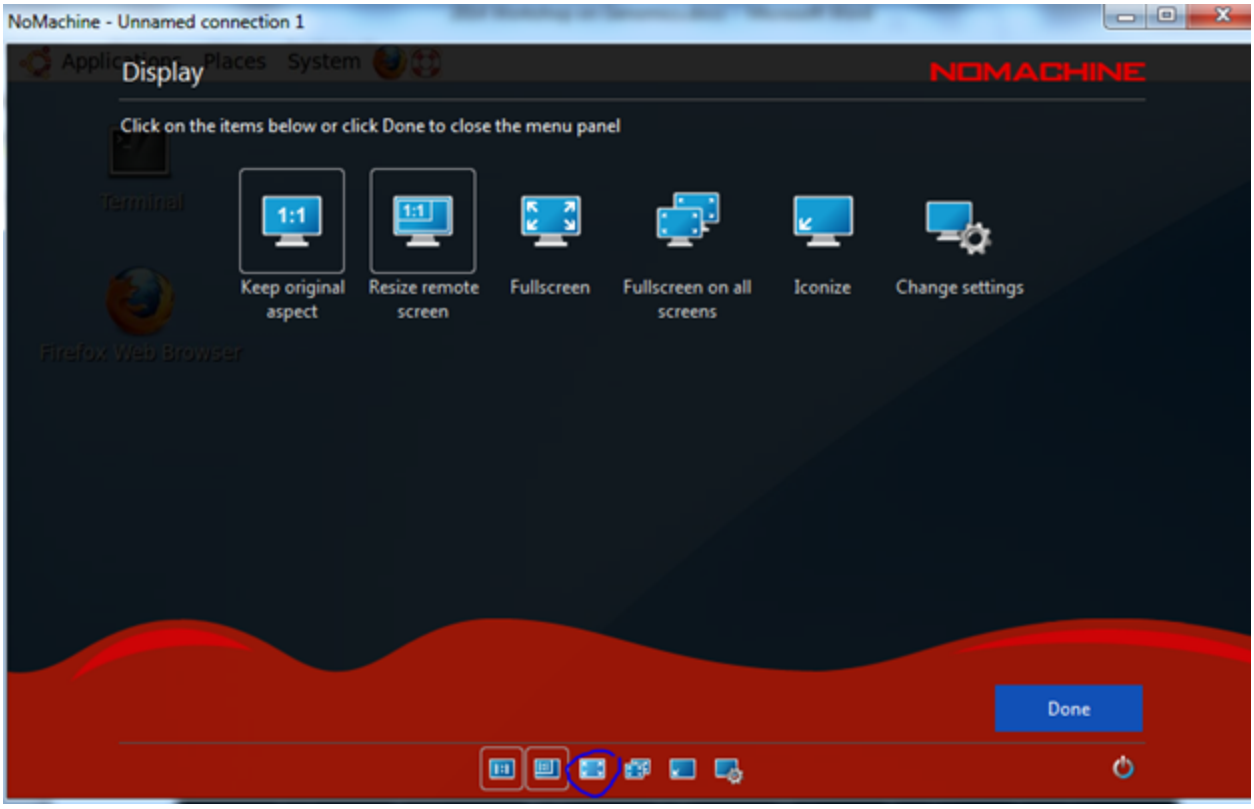
Once connected you should get a nice desktop with links to the Terminal and to the Firefox web-browser. *If a window pops up asking if you would like to run any updates, do not apply any updates – it will take hours.*

Once you have logged in and dismissed any update windows, you'll notice that the window is somewhat small and low-resolution. We can improve this, by moving the mouse cursor over the top-right-hand corner of the screen and clicking on the 'page-curl' illustrated below. If you cannot find the "page-curl" then you can press Ctrl+Alt+0:



This should bring up a window similar to this. Click on the 'Display icon'.

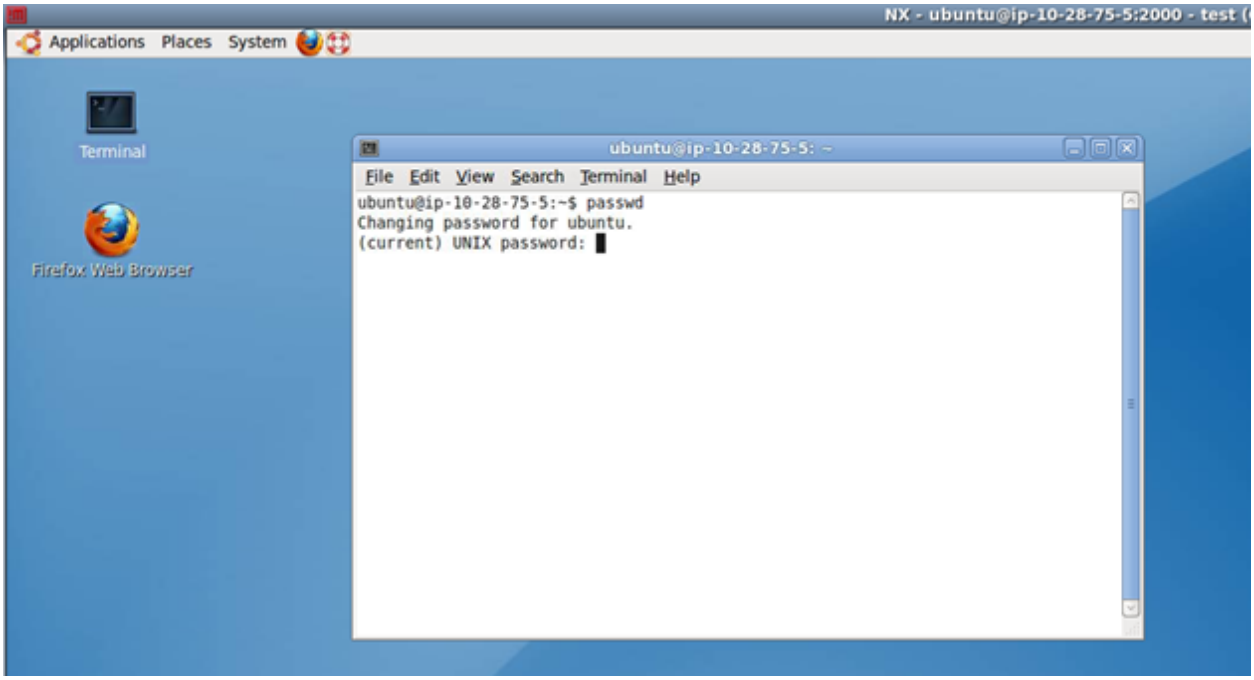
This will give you some more options:



Click on the ‘Resize remote screen’ and also the small icon circled in blue in the figure above. Then click on ‘Done’. This should give you a nice full size desktop to work on. If you need to move out of it, you can again return to the top right hand corner ‘page-curl’ icon and deselect the options you’ve just used.

One final task is to change your password from ubuntu to something less easy to guess. Double-click on the 'Terminal' icon and type the command 'passwd' (without the quote marks). Enter 'evomics' as the current password (again without the quotes) and then type in a new password of your choice. Note that for security, no characters will appear as you type. Please make it difficult for someone to guess your new password! You will then need to re-enter it to confirm.

Every time you log in via NX after this you will need to enter 'ubuntu' as the username and the password you have just entered. *Please do not forget it!*



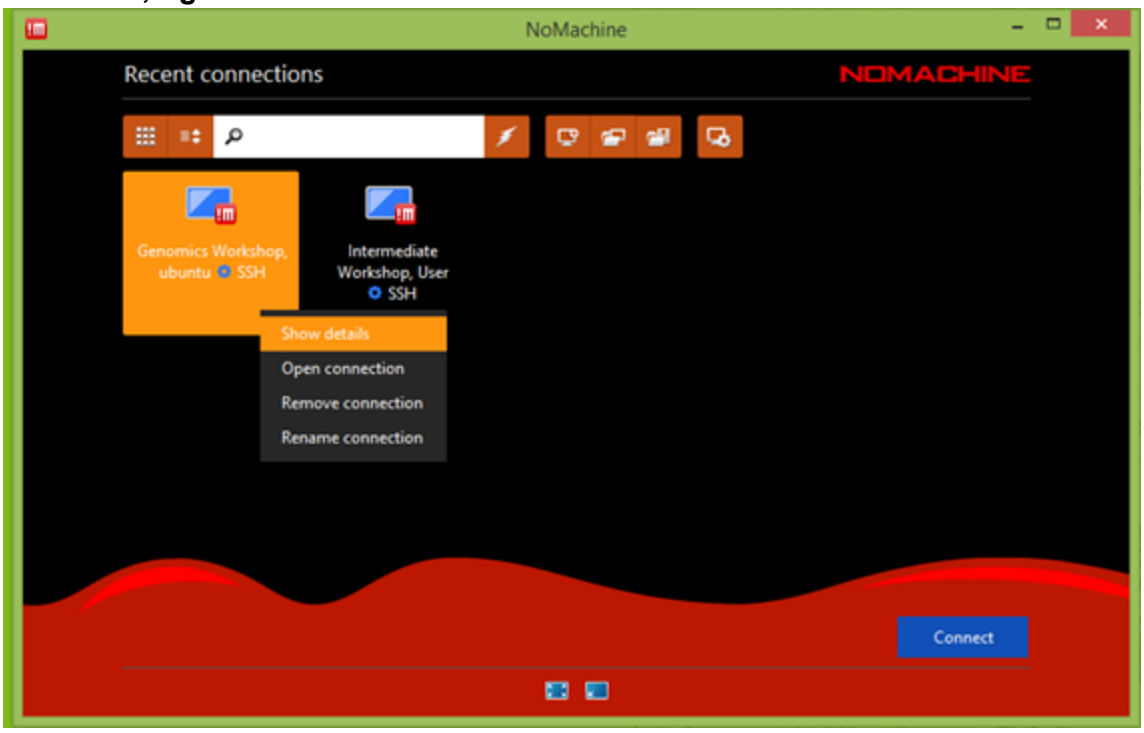
To close the connection just close the window, you can choose “Disconnect” if you want to log back in later and pick up where you left off. However, if you stop or restart the AMI via the console, any unsaved information will be lost.

At this point you've logged into the machine.

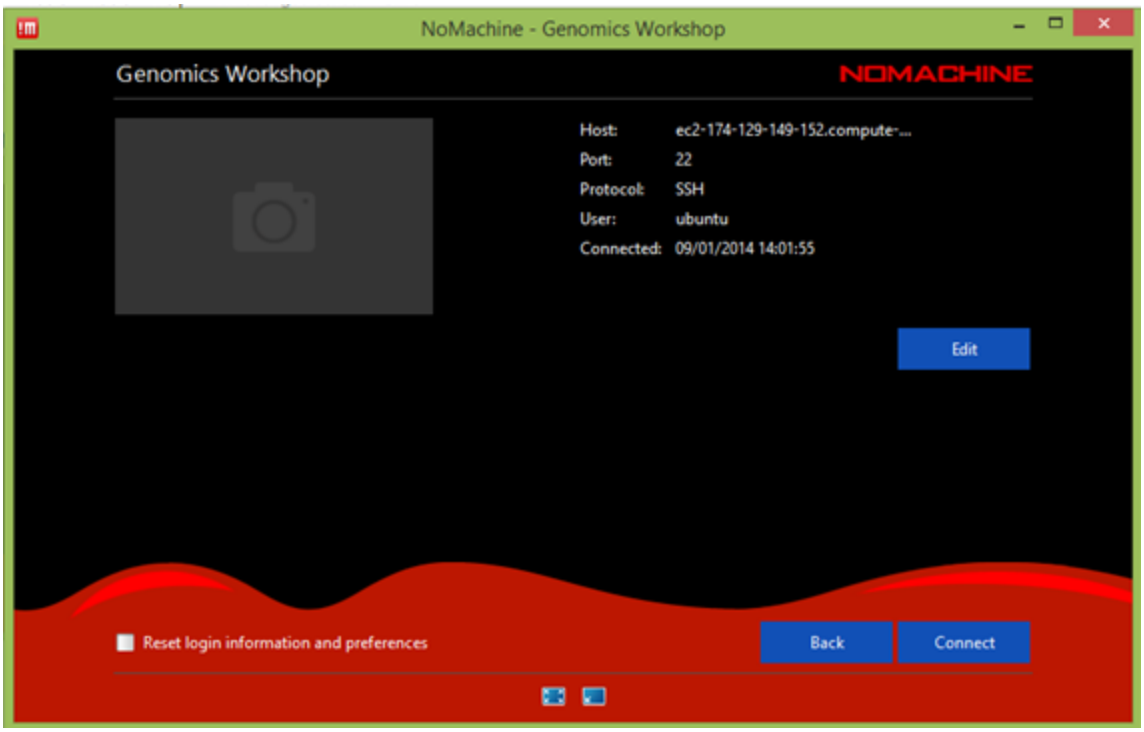
If you do stop or restart, remember to check and change the Public DNS address using the AWS console and change the DNS address in NX.

Please see the next page...

To do this, right-click on the Connection and Select “Show Details”



Select Edit and update the Public DNS number:



Step 4 – Stopping and starting the Instance

When you're done with the workshop it's very important to turn off the VM to avoid unnecessary charges. Just log back into the console, right click the instance and select “Terminate”. It will ask you to confirm. You can then watch the status change from “shutting down” to “terminated”. **Note – this will destroy all work done to date.** If you wish to keep your data, use the 'Stop' option instead. In this state you will not be charged for computing time, but will still be charged for storage.

Very important!

If you stop and then start your instance your Public DNS address may change. If this happens you will need to use the new DNS address with the NoMachine.

Although much of what we have just done may not make much sense yet, most of you will feel totally comfortable and confident working on an EC2 node running Linux within a few hours. It's really amazing how quickly the fact that this is a remote computer will fade away. It may be hundreds of miles away but it will act just like a local computer, especially if you connect via NX.

Optional Step 5 – Log into the Running VM via SSH

Note – this is for advanced users who may want to access the server via SSH. Do NOT try this during these tutorials.

To connect over SSH you need to get the public DNS address, as above, and type:

$> ssh ubuntu@public-dns-numbers.amazonaws.com

You will then be asked to enter your password, you may also have to accept the encryption key.

Note, if you were working on another AMI which requires a key pair, you should have the key file you downloaded present in the same directory that you execute the command from. For example, the command might look like the following and this assumes key-StudentKonrad.pem is in the same directory:

$> ssh -i key-StudentKonrad.pem ubuntu@ec2-174-129-70-43.compute-1.amazonaws.com

Linux/Mac Tip:
When you do the above command it may complain and say “permissions are too loose on the .pem file”. If this happens use chmod to make the file read/write only to you (it's supposed to be private):

chmod a-rwx
key-StudentKonrad.pem chmod
u+rw key-StudentKonrad.pem

And try the SSH command again.
(You'll learn exactly what these commands do during the Unix tutorial)

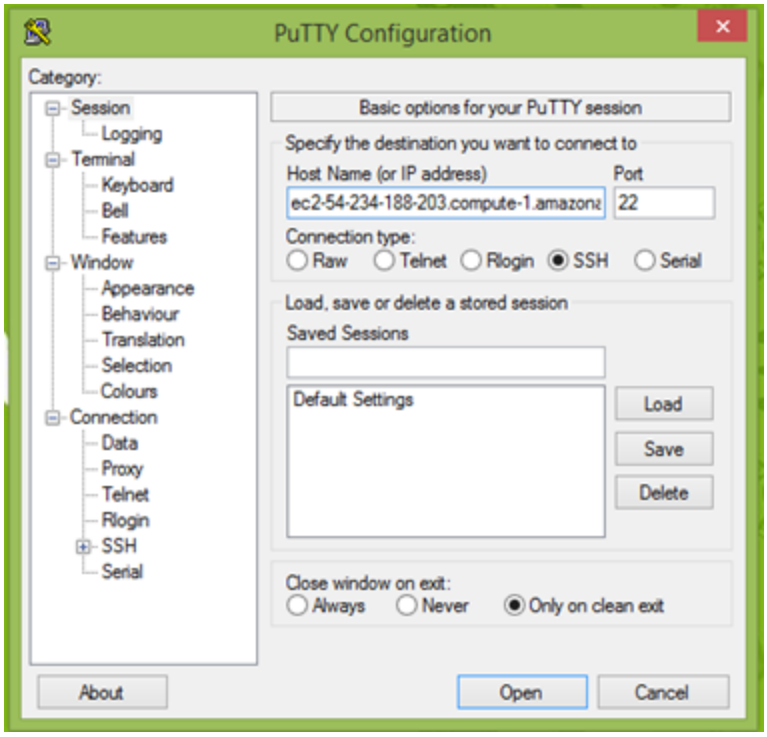
Windows Tip:

If you ever use a different AMI from the one used in this workshop, the chances are you will need an "SSH" client to connect to the instance. Mac and Linux have this built in, just open a terminal and you're ready to execute the command above. For Windows you should download the Putty program (http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html) or MobaXTerm (http://mobaxterm.mobatek.net/MobaXterm_v6.6.zip) which gives you a very easy-to-use SSH program for Windows. The username above is "ubuntu" and the server is " ec2-174-129-70-43.compute-1.amazonaws.com". You'll use the server name that's assigned to the one you launched. Instructions for both of these programmes can be found below.

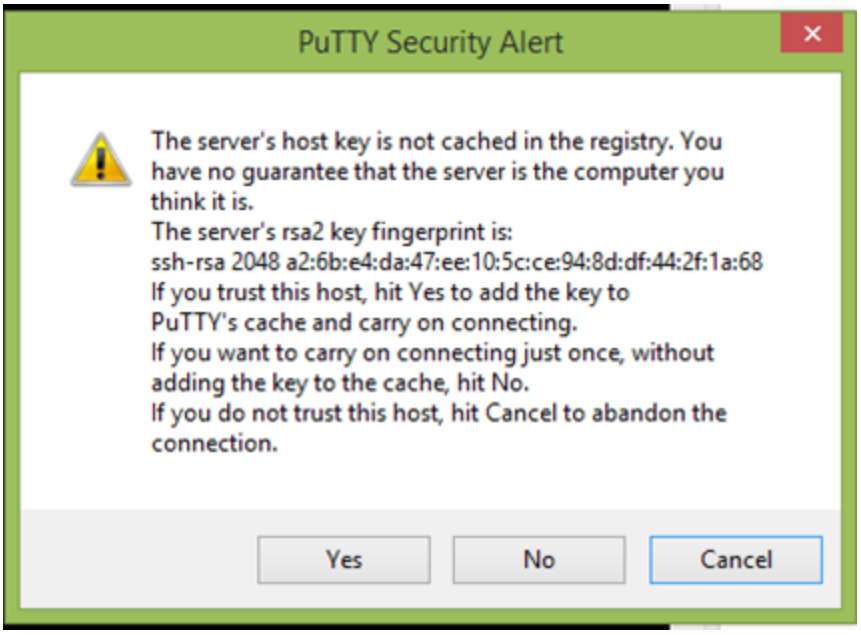
Note, again if you're using a VM that requires a key pair, when you launch this program look for the following setting, you'll need to provide the program with the path to your .pem file that you downloaded when launching your cluster node. See the "Private key file for authentication" option in the screenshots below.

Using PuTTY (Windows Only)

PuTTY is a SSH terminal for Windows. It can be used to access our VM as a terminal. To download putty go to http://www.chiark.greenend.org.uk/~sgtatham/putty/.

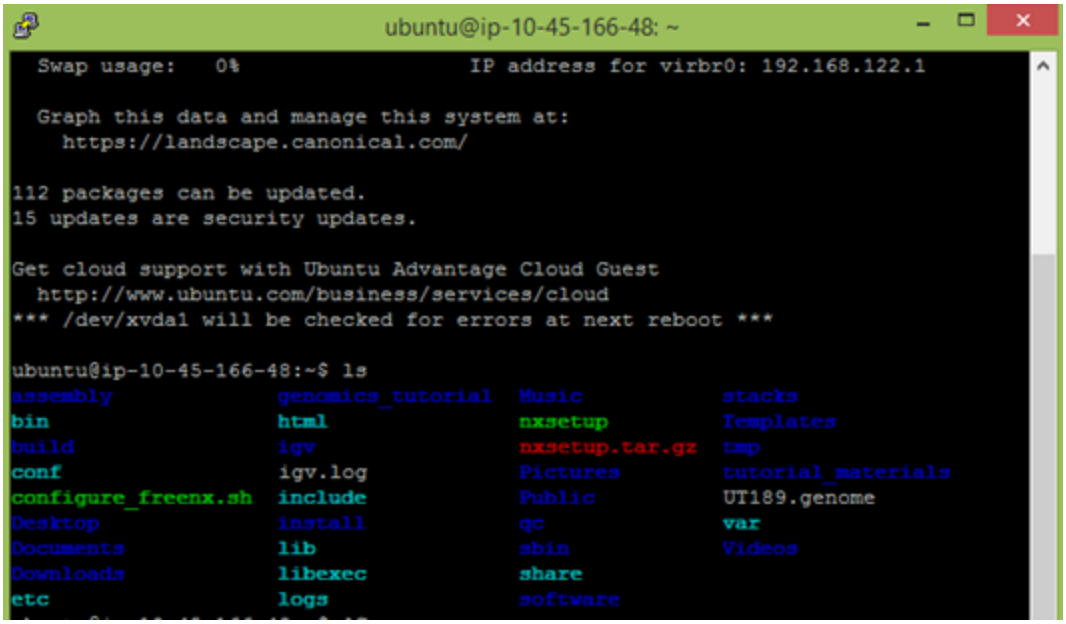


In the Host Name bar insert the Public DNS number for the Amazon Instance.



If this warning message appears, click “Yes”. This is a check that you trust the computer you are connecting to.

Enter the username “ubuntu” and the password “evomics”.



You are now accessing the terminal of your VM. Here we can see all the files are listed.

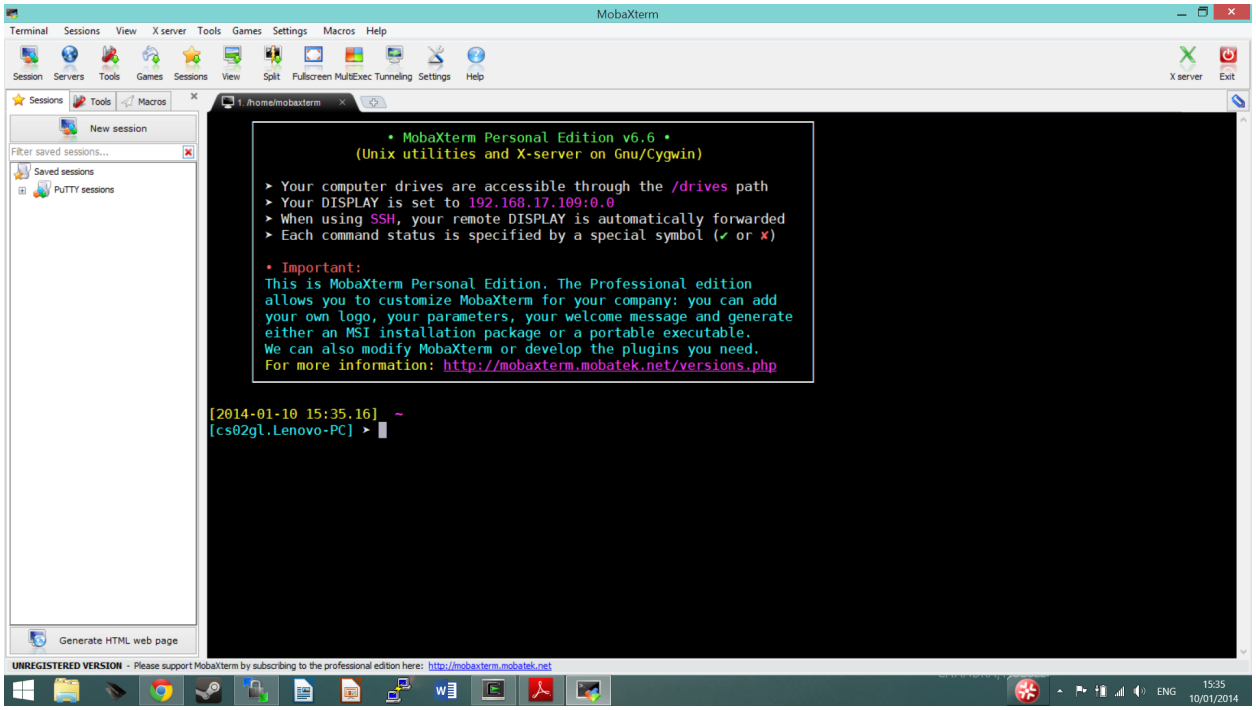
Using MobaXTerm (Windows Only)

MobaXTerm is another terminal for use in Windows environments. It has more features and options than puTTy, some of which you will need to pay for to use. However, the majority of the options/features you will use in these sessions are available in the free portable version! This means you do not need to be an administrator to use/install the program.

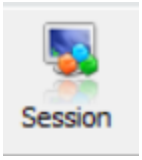
Download here - http://mobaxterm.mobatek.net/MobaXterm_v6.6.zip

Use your favourite unzip manager (e.g. 7-Zip) to unzip the archive and place the executable file somewhere you can find it (perhaps in your “Program Files” folder under MobaXTerm, not your Desktop if you can help it!). Double click the file to run the program….

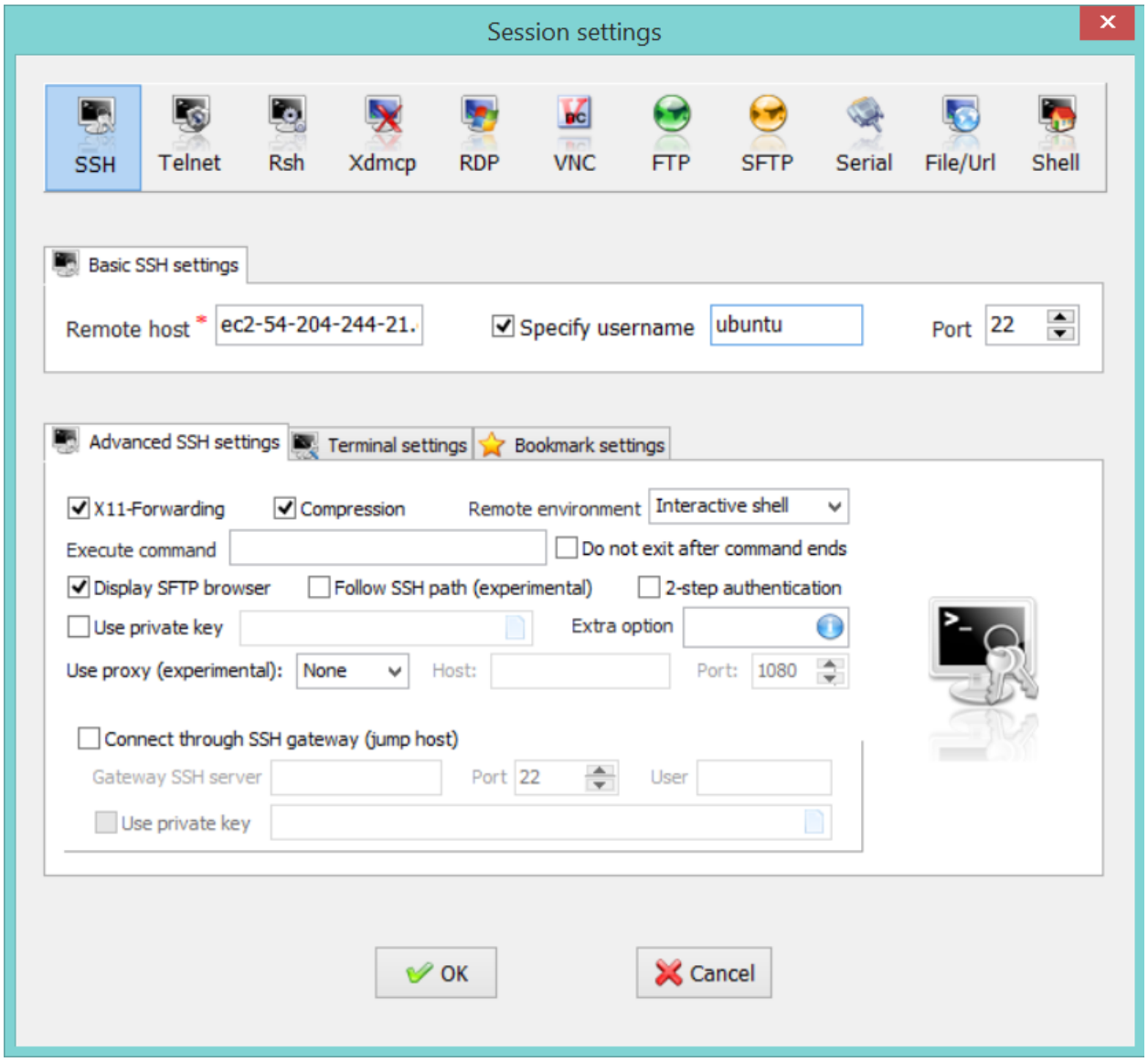
When the program has started you will be shown a screen like below:



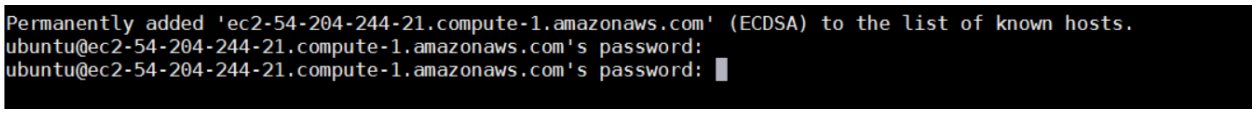
The black screen - terminal - gives you access to your local computer file system with many of the UNIX commands built in (e.g. ls, cat, head).. You may also see saved PuTTY sessions already loaded on the left side of the screen, if you have used that program before and saved them.

However, if you do not you should click the “Session” button on the top left. 

You will then be shown a screen with many options of session type (e.g. SSH, Telnet, RDP, FTP). You will want to select “SSH”.



Enter your Public DNS in the “Remote host” box and specify your username as “ubuntu”. You will then be asked for your password in the terminal as below...



Please leave all settings as their defaults. You may also notice a checkbox that says “Use private key”, this is where you would specify your private key if you were using one with a different instance of an AMI, i.e. not for this workshop but your own instance.

Once you are logged in, one of the nice features of MobaXTerm is that you can easily transfer files with an inbuilt browser (via sFTP) on the left hand side of the program window in the Sftp toolbar… You can also detach your tabbed window terminal session (much like you can in Firefox or Chrome with a website tab) and should try and auto-reconnect if you lose your connection.

MobaXTerm should also save all your session details, including passwords and private keys between sessions of using it. Your saved sessions will appear on the left hand side of your program screen.