2016 Workshop on Genomics

Starting your Amazon virtual machine

Instructors

- Sophie Shaw S.Shaw@exeter.ac.uk
- Guy Leonard G.Leonard@exeter.ac.uk
- Jan Provazník provaz@entu.cas.cz

Objectives

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

- Log into the Amazon Web Services (AWS) Console and start your instance of the appropriate workshop.
- Log in to the Amazon EC2 instance from your own computer.
- Be able to continue with the tutorials at your own pace.

Table of Contents

Introduction

For this workshop we will provide an overview of Amazon's EC2 and how, as researchers, we can use this flexible resource to get work done quickly and relatively inexpensively.

Firstly some terminology we use throughout the document might be confusing so here are some definitions.

Amazon Machine Image (AMI): This is the starting point or template for the course - it contains all the programs and data that you require to follow the course. An AMI is analogous to powering down your computer, and pulling out the hard drive -- the hard drive is an "image" of your computer.

An Instance: Almost the first thing you will do is create your own copy of the AMI - we call this an instance. It contains everything that was in the AMI plus any files you create during the course. One way you can think about an instance, and how it differs to an AMI, is that an instance is analogous to putting a hard drive into a physical computer and powering it on.

We will dive right in by logging into the Amazon management console; starting up your own copy (an Instance) of the pre-prepared Amazon Machine Image (AMI) for this workshop. We will give you a whirlwind tour of the features of Amazon's cloud and then log-in to your private instance via the X2go-client.

For this tutorial I borrowed documentation from the following site:

- 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 also has 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!

Task 2 - Connecting to Your Personal Instance

The Rules

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

- Please only launch a **single** instance of the type specified by the instructor at the beginning of the workshop.
- Please **stop** instances at the end of the day so we can avoid being billed for resources that are not actively being used.
- Please name your instance. Including your name will make it easy to find your resources in the list of class resources.
- Do not delete EBS volumes that do not belong to you.
- Do not **terminate** instances that don't belong to you.

Logging Into the Console

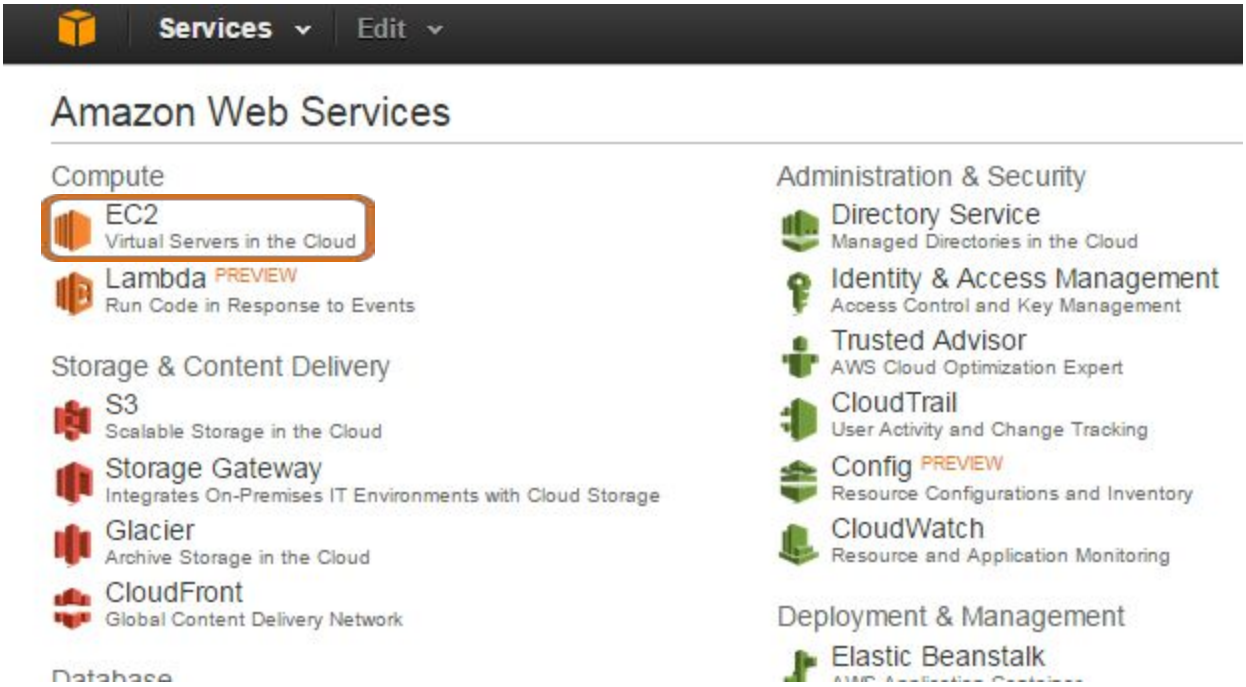
In addition to being extremely comprehensive, the Amazon cloud has a very easy-to-use interface for interacting with all their cloud offerings. All you have to do is log into a web application and most of the functionality of the Amazon tools is available for you and very easy to use.

This workshop has its own Amazon account and, and we have created a sub-account for students to use. (If you're wondering, we created the sub-account by using the Identity & Access Management tool). 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 given at the workshop.

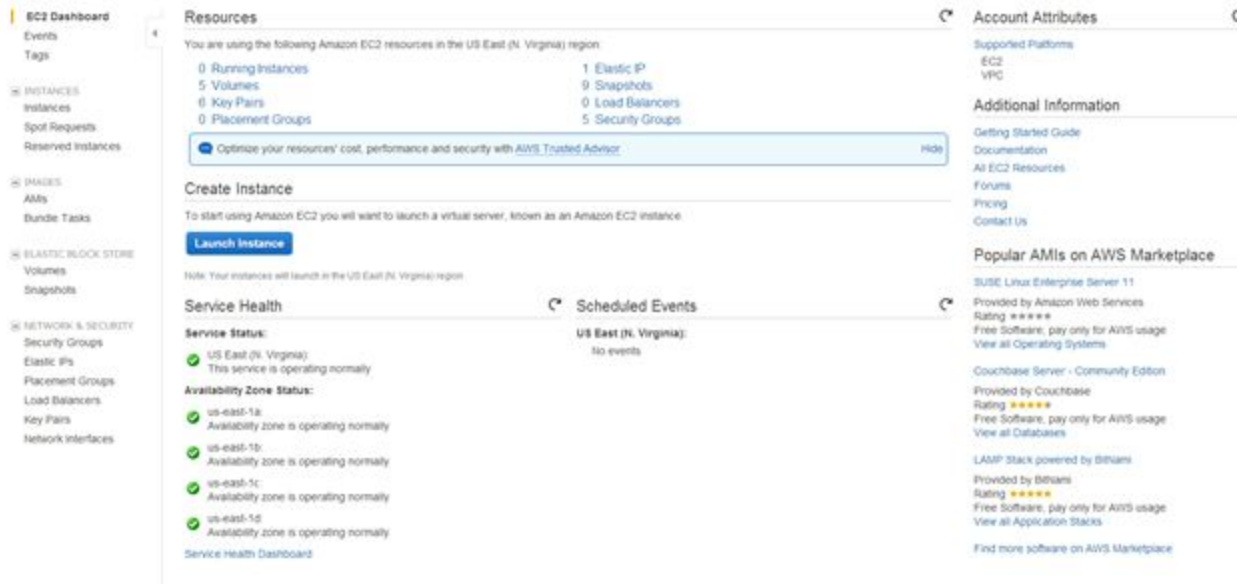
https://evomics.signin.aws.amazon.com/console

After logging in you'll be presented with a wide range of options.



The AWS Dashboard. Click on "EC2" in the top left under "Compute".

On this page you'll get a summary of the EC2 state for your account (EC2 Management Console). You can see mine below:



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 charges for each gigabyte stored every month and for each hour a machine is run. This can vary from a few cents per hour to a few dollars. It is very convenient if you are only doing analyses occasionally, though at the moment it is still cheaper to have your own computer system if you'll be using it frequently, and assuming that you do not need a very large amount of compute. On the other hand, if you have your own resources then you need to administer those resources.

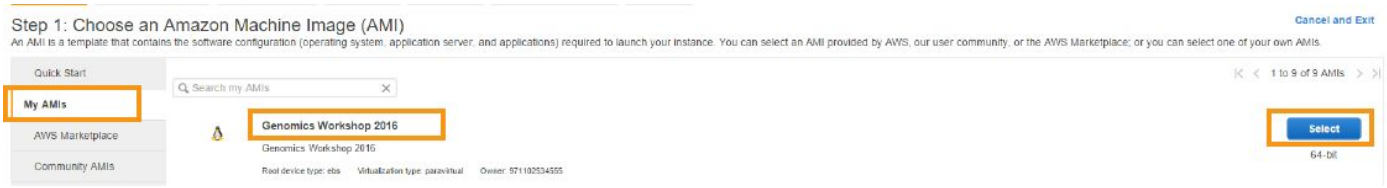
The reason we are using the cloud here is that it is the easiest way for us to provide individual systems that are setup for the tutorials, and because it is a great way to do some computer work! In the case of high-throughput sequencing data (e.g. Illumina) you will find that your desktop PC may not 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 instance from home and continue to work through the tutorial from there.

Creating an instance

Once logged into the console we can "launch" an instance of a virtual computer.

*The create instance section of the EC2 Management Console. Click on the "**Launch Instance**" button in the centre in order to create an instance.*

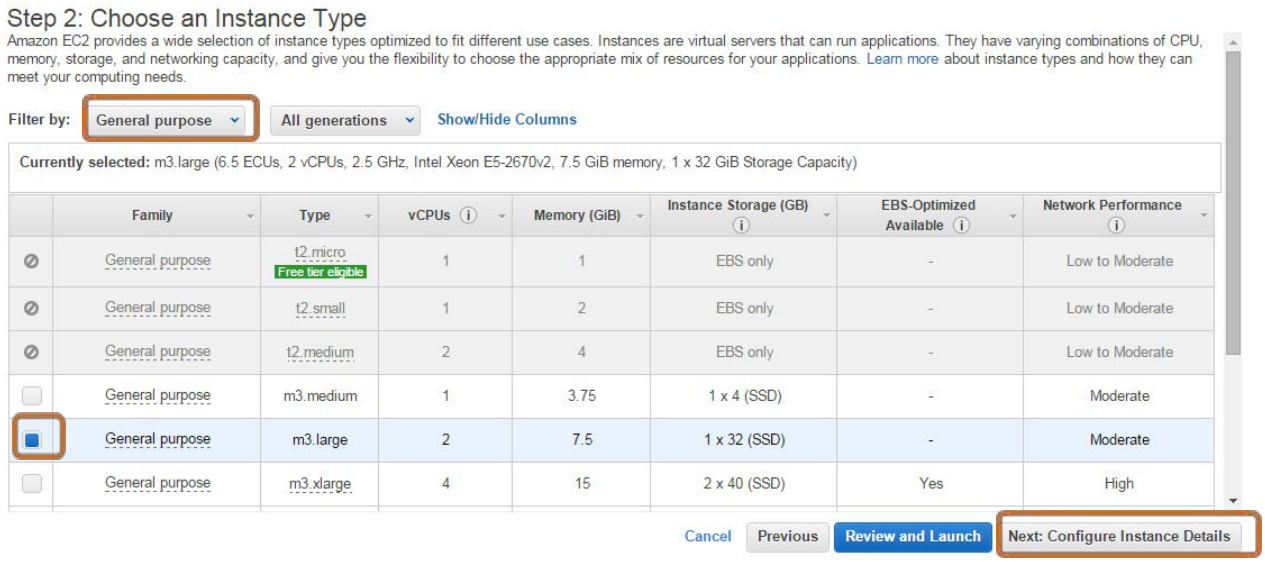
Choose an AMI



On the AMI selection screen, select "**My AMIs**". Please choose the "**Genomics Workshop 2016** ami-2f6f3445" and click "**Select**".

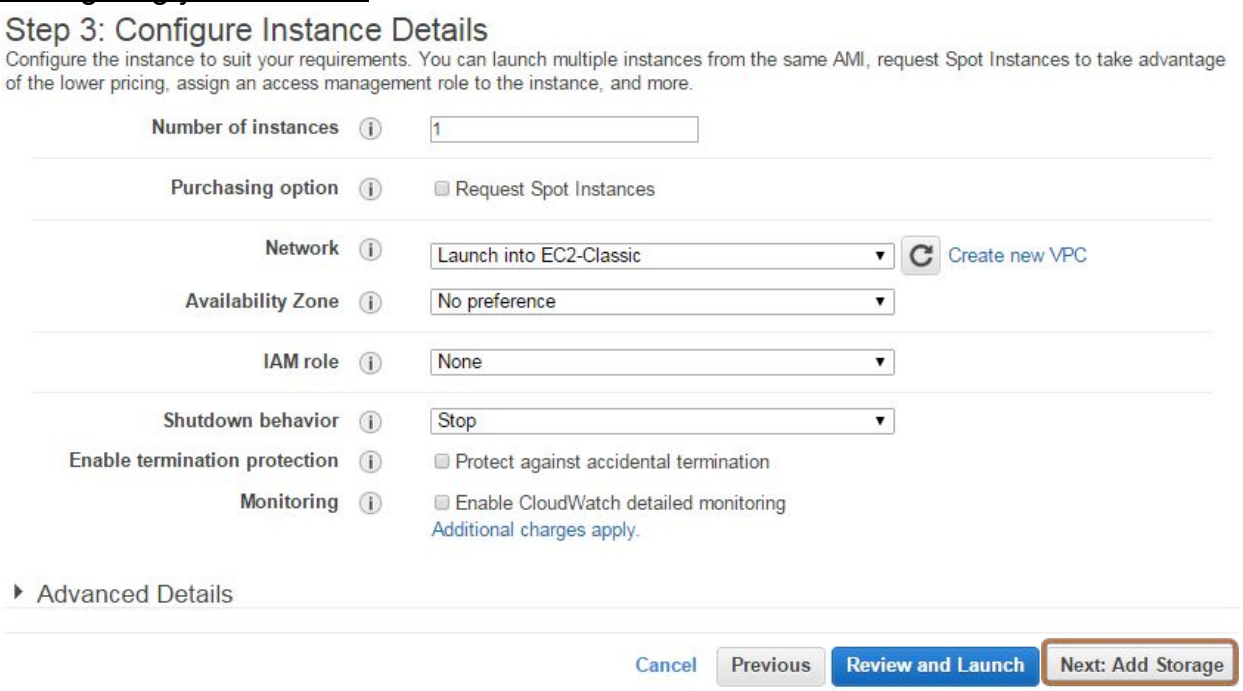
Please make sure you select the correct AMI for this workshop! The correct AMI is named "Genomics Workshop 2016" This may not be the AMI at the top of the list.

Choose an instance type



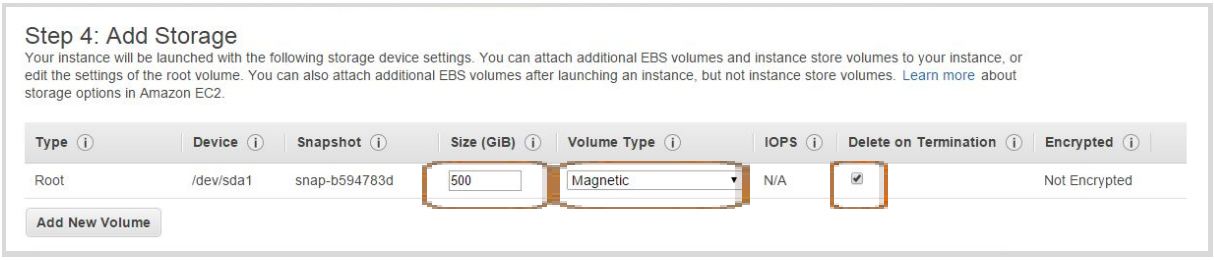
The instance type page. The instance type page allows you to define the "physical" (e.g., the number of processors, amount of RAM, etc) nature of your instance. At the top change "**Filter by**" to "**General purpose**". Please select "**m3.large**" and then click on "**Next: Configure Instance Details**."

Configuring your instance



The instance configuration page. Nothing needs to be done here, please click "**Next Add Storage**."

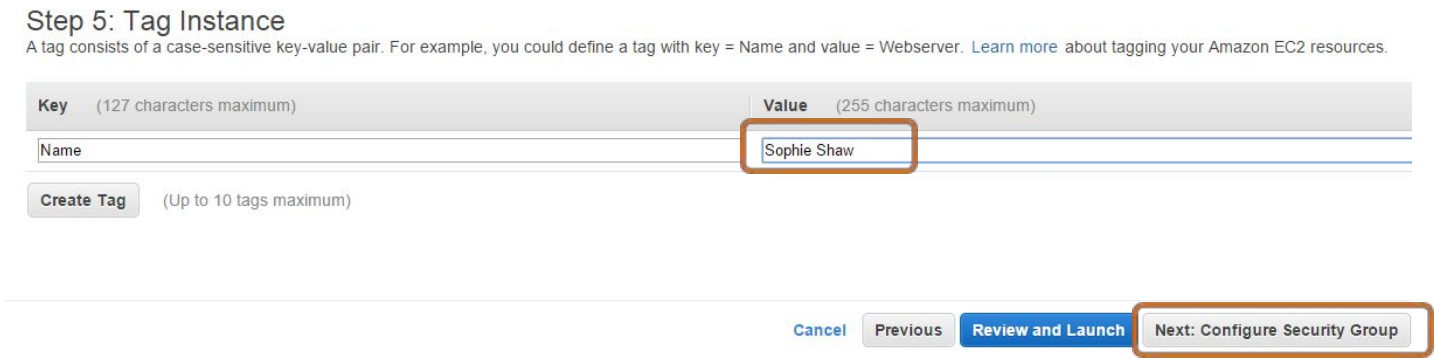
Add storage



On this page we can select how much storage we want to add to our instance. Please leave the default at 500GB, and make sure the volume type is **"Magnetic"** and the "**Delete on Termination**" checkbox is ticked. Following this, click "**Next: Tag Instance**."

The 'Delete on Termination' box 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.

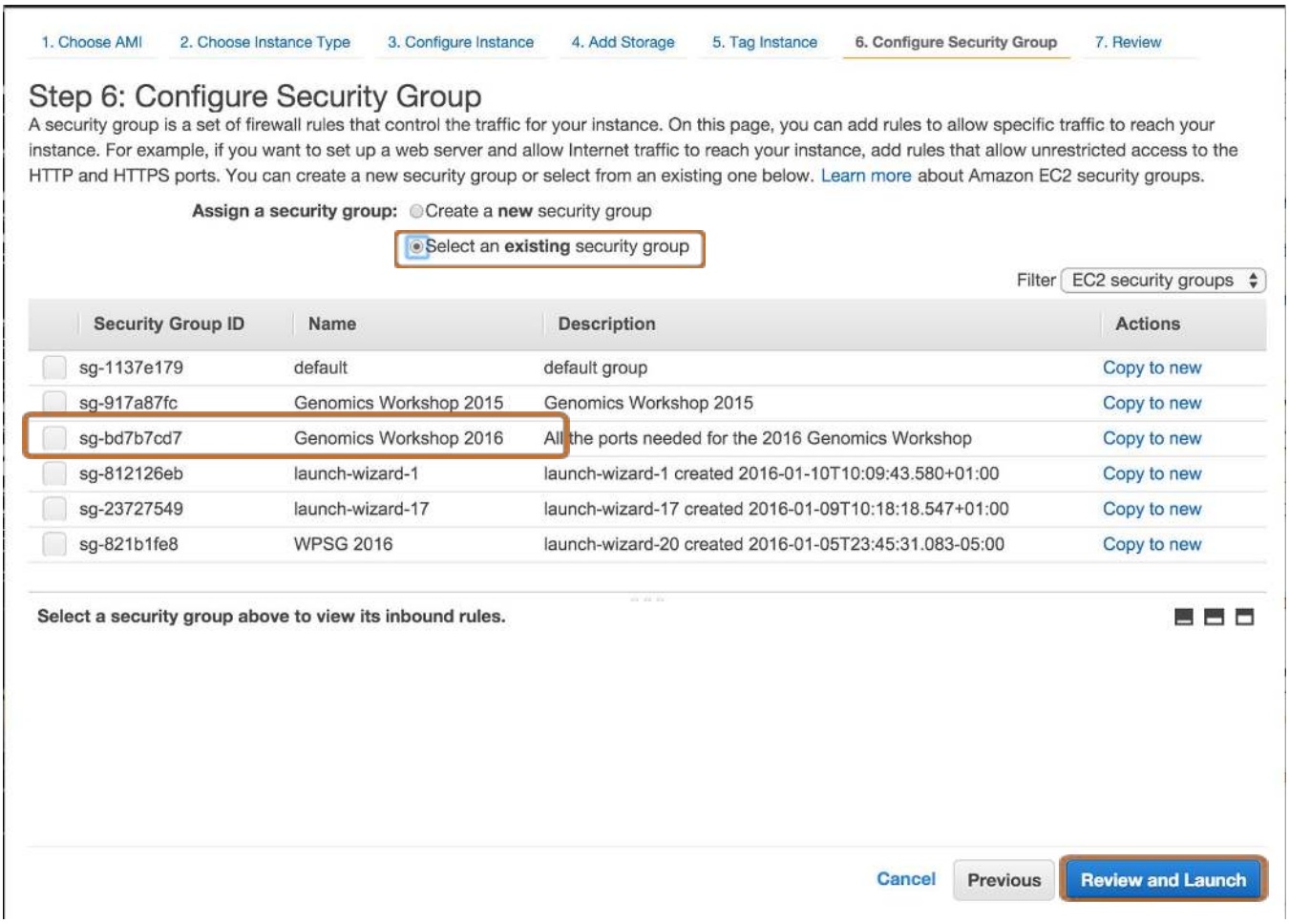
Tagging an instance



The instance tagging page. This page allows you to give your instance a name, which makes them easier to identify in the management console. 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**."

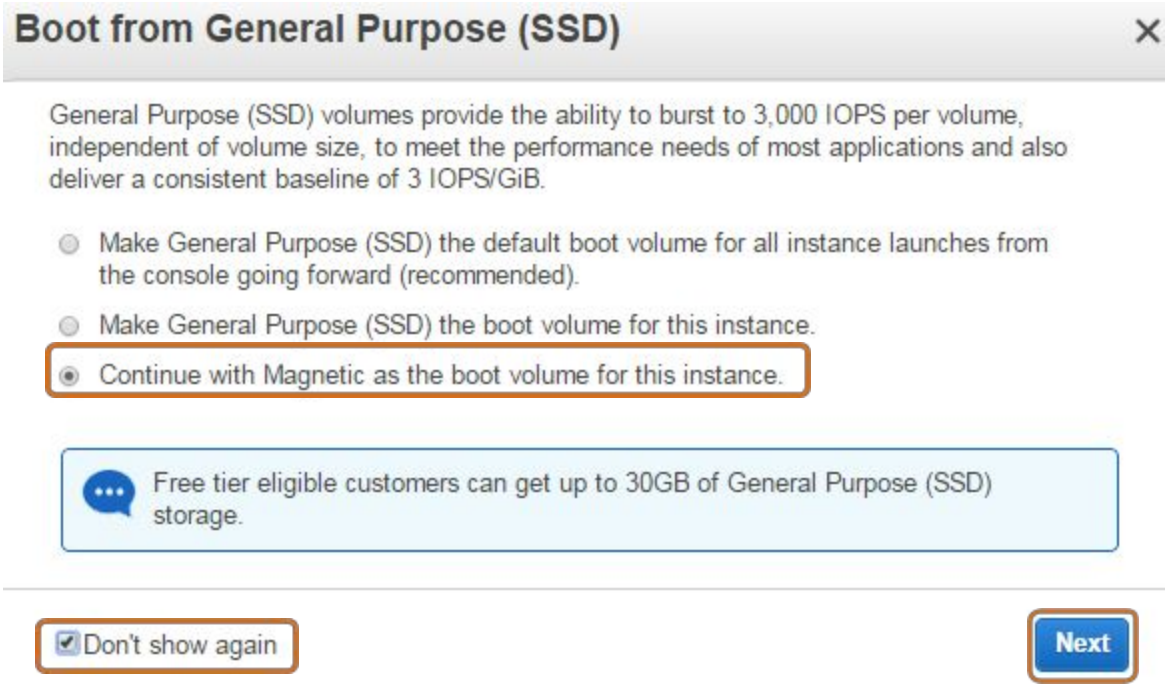
The idea of a 'tag' is that if you have multiple instances you can create tags to identify them. As we are all using a single account, it is important to be able to identify your instance.

Choose a security group



The security group selection page. First, click on "**Select an existing security group**." Next, select the "**Genomics Workshop 2016**" group. Finally, click on "**Review and Launch**."

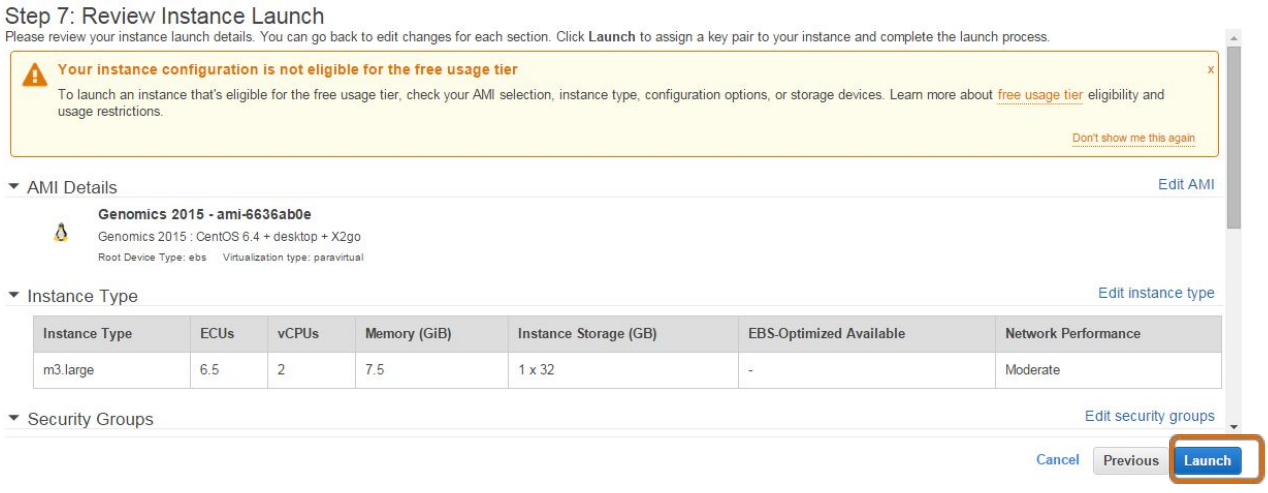
You may get the following prompt (don't worry if you don't)



Sometimes EC2 presents this screen. If this screen shows up, please make sure "**Continue with Magnetic as the boot volume for this instance**" is selected, check the "**Don't show again**" and finally click "**Next**".

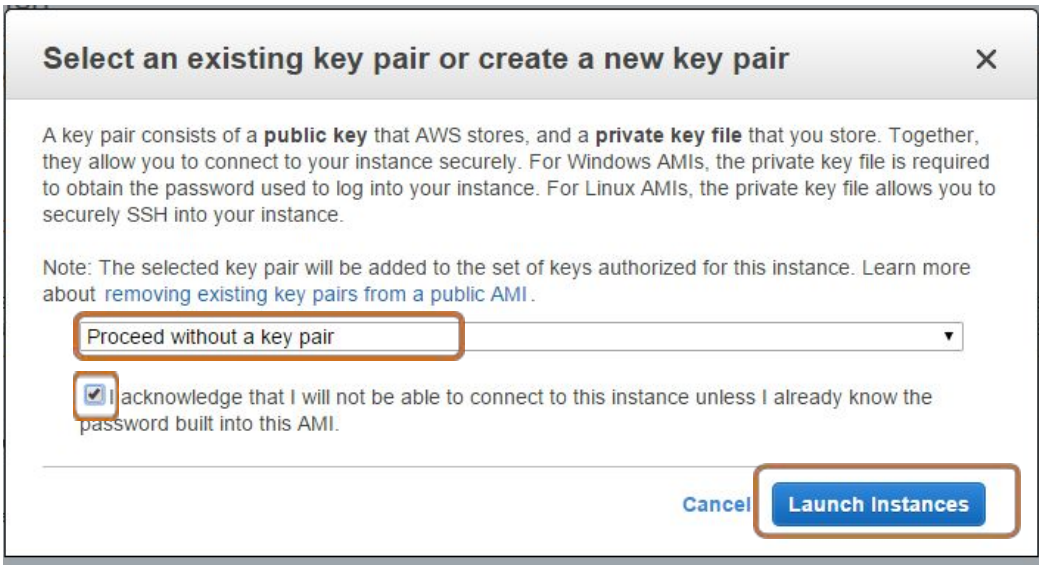
Review and Launch the Instance

The next step is to review and launch the instance

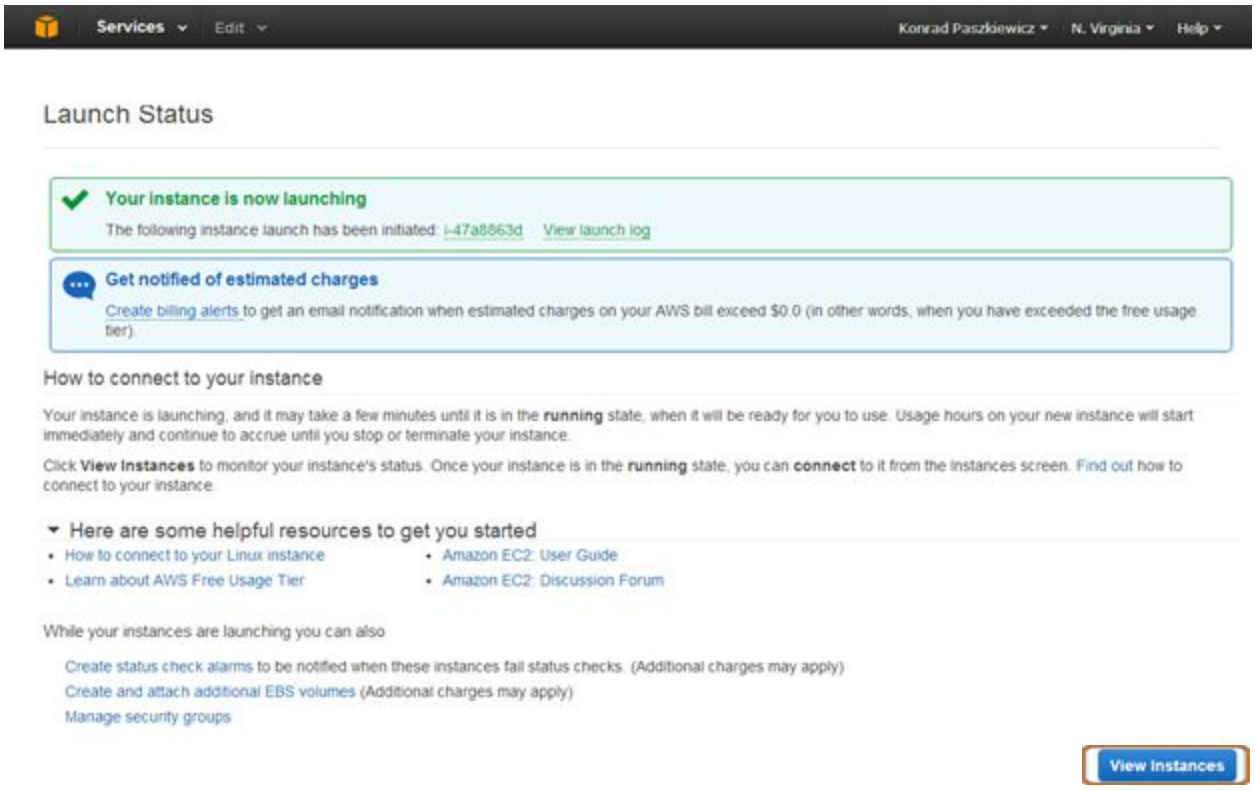


The instance summary page. Please just click "**Launch.**" The screen may show a warning about not being in the "free tier" and this warning is safe to ignore.

The final step is to select the “key pair” used to let you log into this machine. This key pair is a file that allows access without a password - (in case you forget it !)



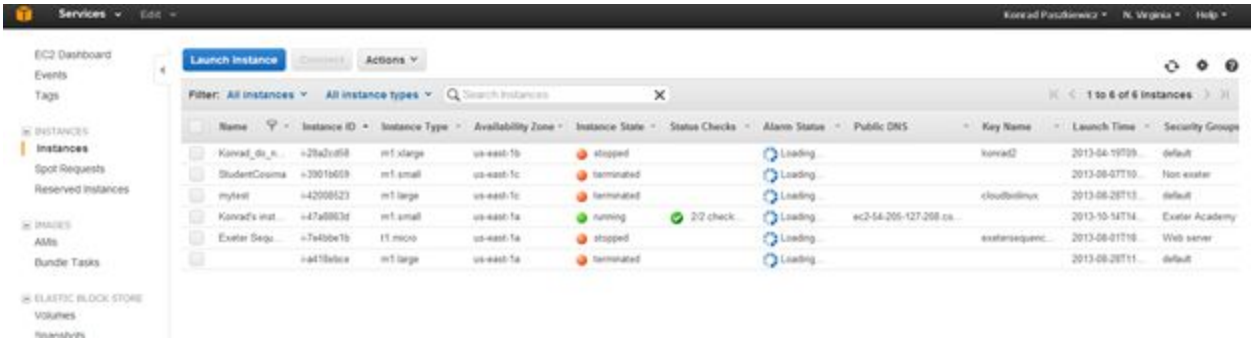
The key pair selection screen. Please select “**Proceed without a key pair**”. Next, ensure the acknowledgement box is ticked. Finally, click on “**Launch Instances**.”



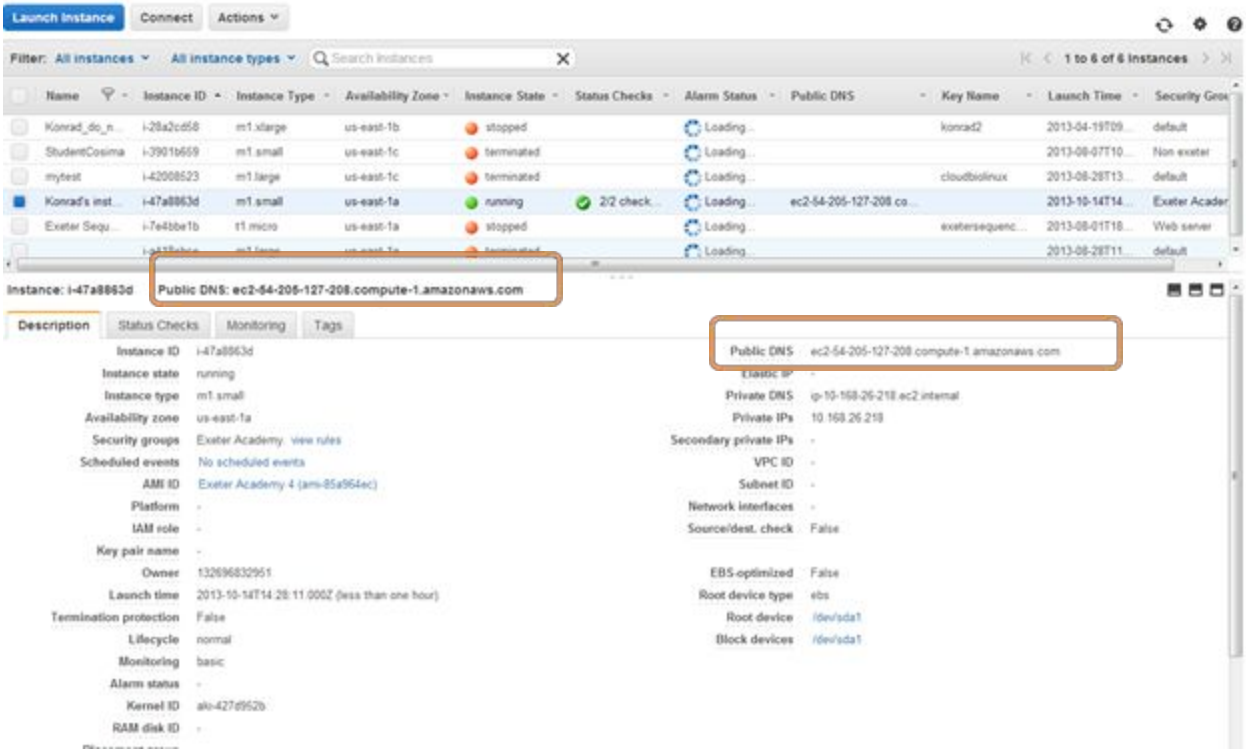
The Launch Status page. This lets you know the instance is currently starting. Please click the “**View Instances**” button to view the running instances. This will let you monitor your instance as it starts up.

Instance monitoring

At this point we need to wait just a couple minutes for the AMI instance to come online. Below 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! You'll know it's finished when "Status checks" says "2/2 checks passed".



The instance monitoring page. Once your instance turns **green** and says **2/2 checks passed**, you should **click on your AMI**.



Instance details. The instance details provides the specific details for how to actually connect to your instance. The **Public DNS** is the address that you will be using on subsequent steps to connect to your system. This is the system's unique "address / phone number" on the Internet.

Log into the Running Instance's Desktop with X2Go

Whilst your instance is initializing - please note it may take some time (~15 minutes) - take this opportunity to install the X2Go client software which you will need to connect to the instance.

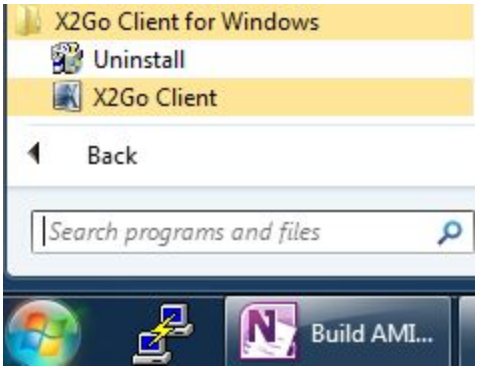
This will allow you to see a windowing environment (like your Desktop) rather than just a terminal! It is a great option if you want to use a GUI application (Graphical User Interface like IGV). It's very cool to see a remote desktop with Firefox and every other GUI application rendered quick and snappy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

Here are the steps to get remote X2Go login working… *Note that these instructions will only work for this workshops' particular AMI. Many AMIs will not have the X2Go server installed and therefore you will not be able to connect using the X2Go 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.*

First determine whether *X2Go Client* is installed on your computer. If it is skip this section. It is not something that is installed by default, so it probably is not there.

There maybe a link on your desktop:

Or look for it in your start menu



X2Go on Windows

Installing X2Go client in Windows.

If you need to install X2Go and you have a windows computer, install it from this link (you will need admin rights on your computer). These instructions are specific to people running Windows. If you have a Mac, please scroll down to the section on **Installing X2Go client on MacOSX**. If you have a Linux machine, please scroll down to the section on **Installing X2Go client on Ubuntu**.

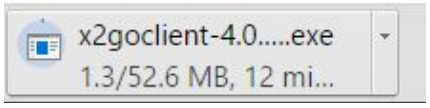
The link to install for Windows is here:
http://code.x2go.org/releases/binary-win32/x2goclient/releases/

(If you do not have admin rights, alternative instructions can be found here
http://wiki.x2go.org/doku.php/doc:installation:x2goclient)

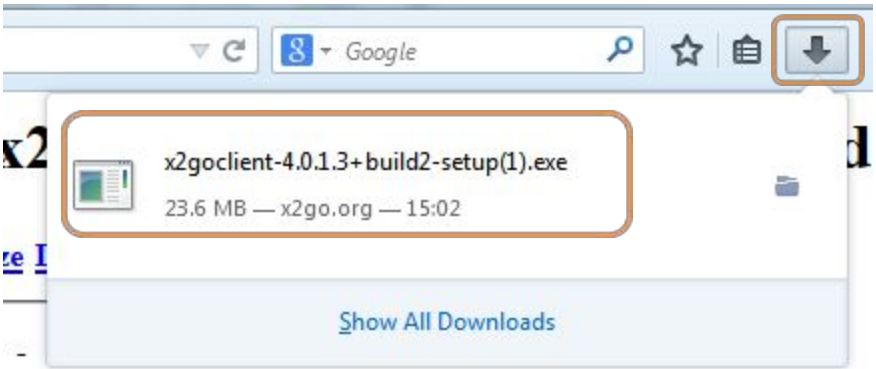
click on the latest link:

Then select the setup program.

This will download the setup program to your computer.
Find the file and click on it: Depending on what browser you are using

In Chrome: 

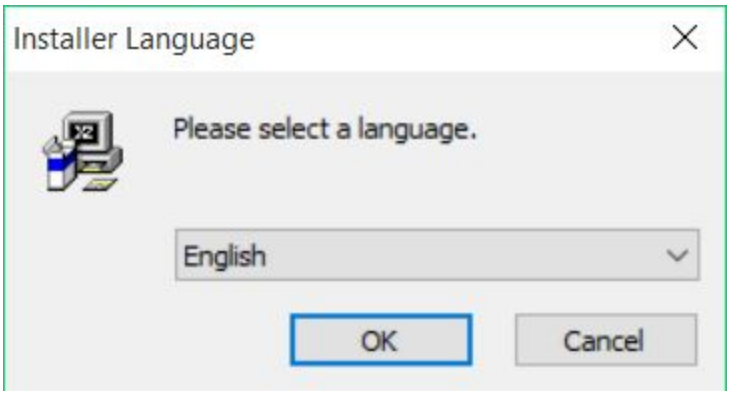
Or FireFox (version number will vary from this screenshot, do not panic :)):



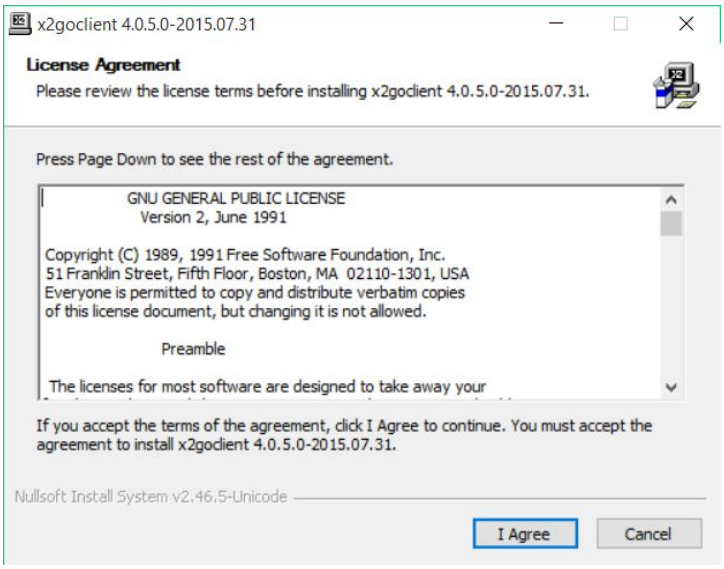
Internet Explorer (Yet again, the version number will differ, this is no cause for concern :)):

You will get a screen asking for permission to continue - select 'yes'.

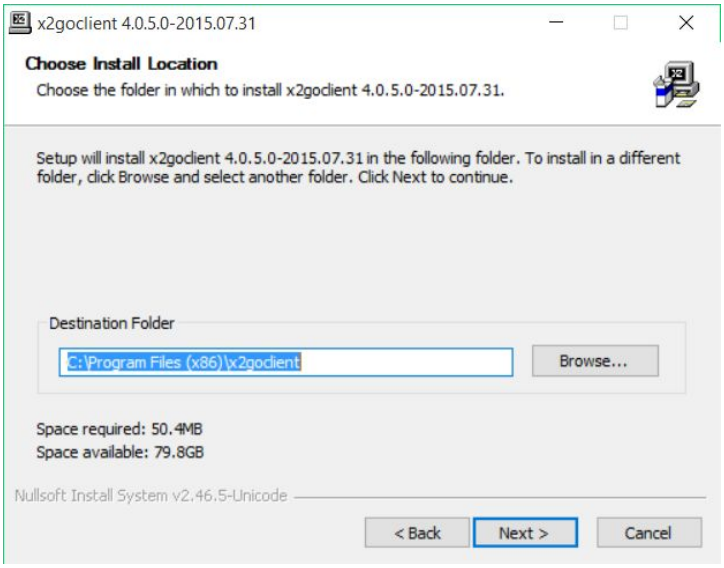
The installer will prompt you to select the language of installation. English might be fine.



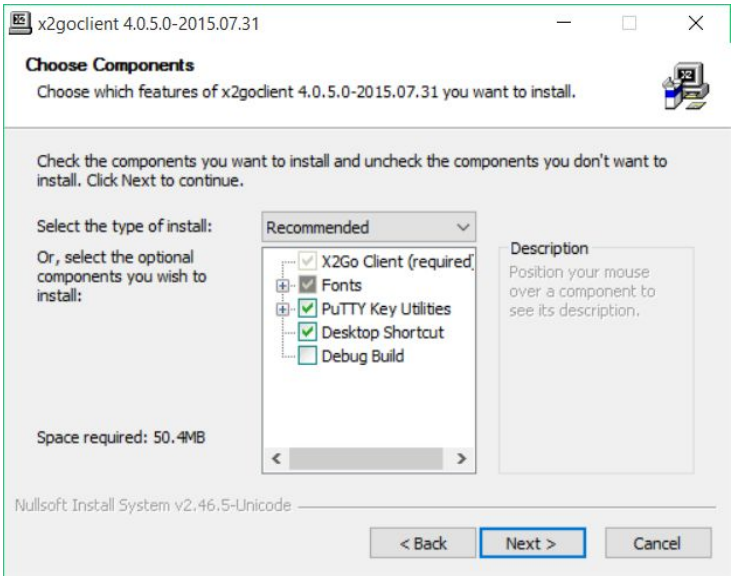
The license agreement. Click “**I agree**”.



The install location. The default location is acceptable, please just click “**Next**”.

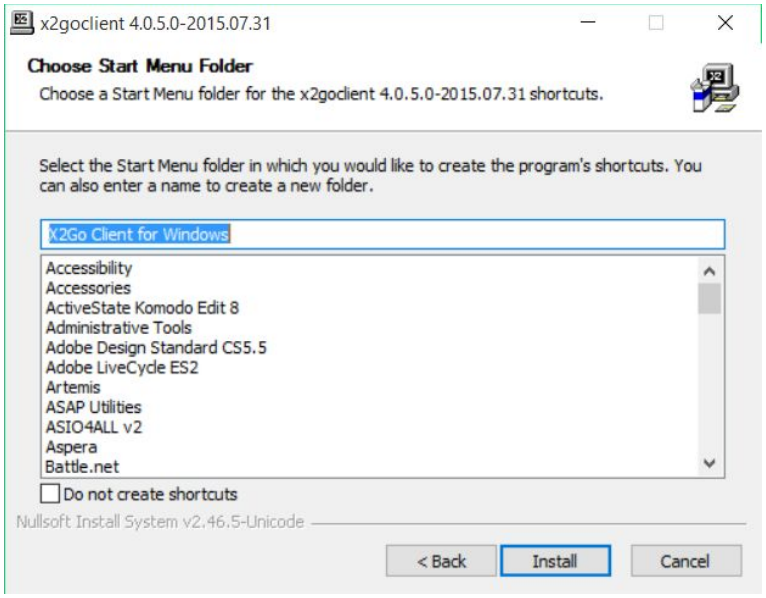


Components. In this step you can choose components to install. Default is OK. Click “**Next**”.



.

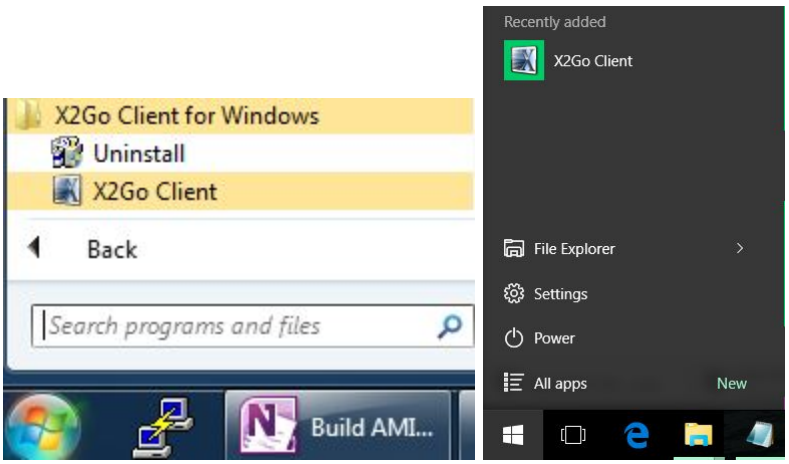
The start menu location. The defaults are acceptable, please just click “**Install”**.



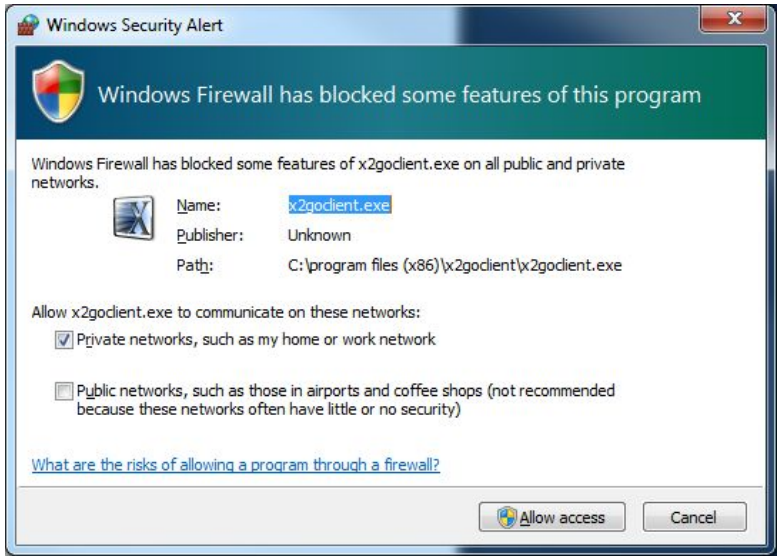
Start X2Go client in Windows.

Once installed, you should have an **X2go Client** icon on your desktop which you can double click. The icon looks like:

Alternatively, you can look for it in your Start Menu :



When you first run the **X2go Client**, you may get a message about changes to your firewall. These changes are fine.



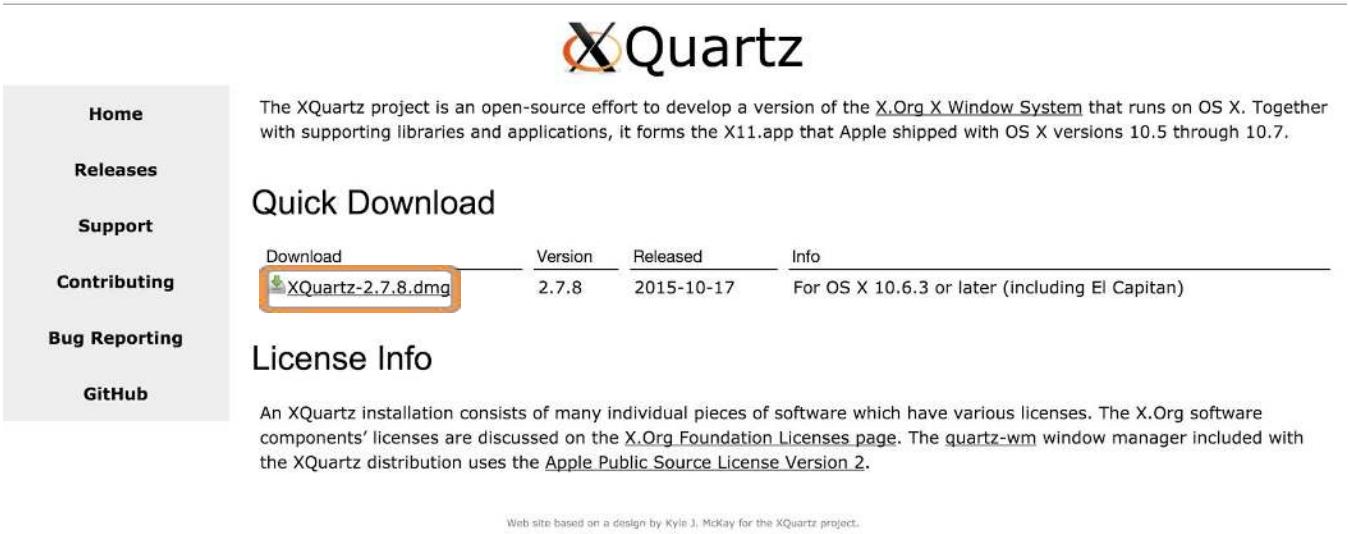
Windows firewall changes. These changes are not necessary, so please just click “**Cancel**.”

X2Go on Mac OSX

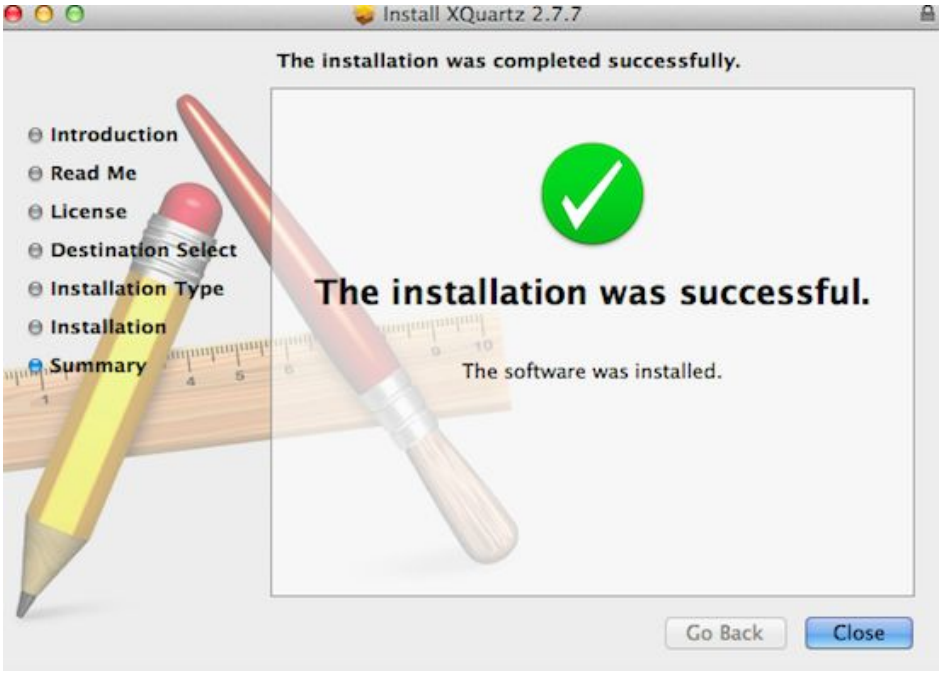
Installing X2Go client on Mac OSX

Prior to the installation of *X2Go*, you will have to install *XQuartz* (if you have not done it at home already).

Download the dmg file at http://xquartz.macosforge.org/landing/

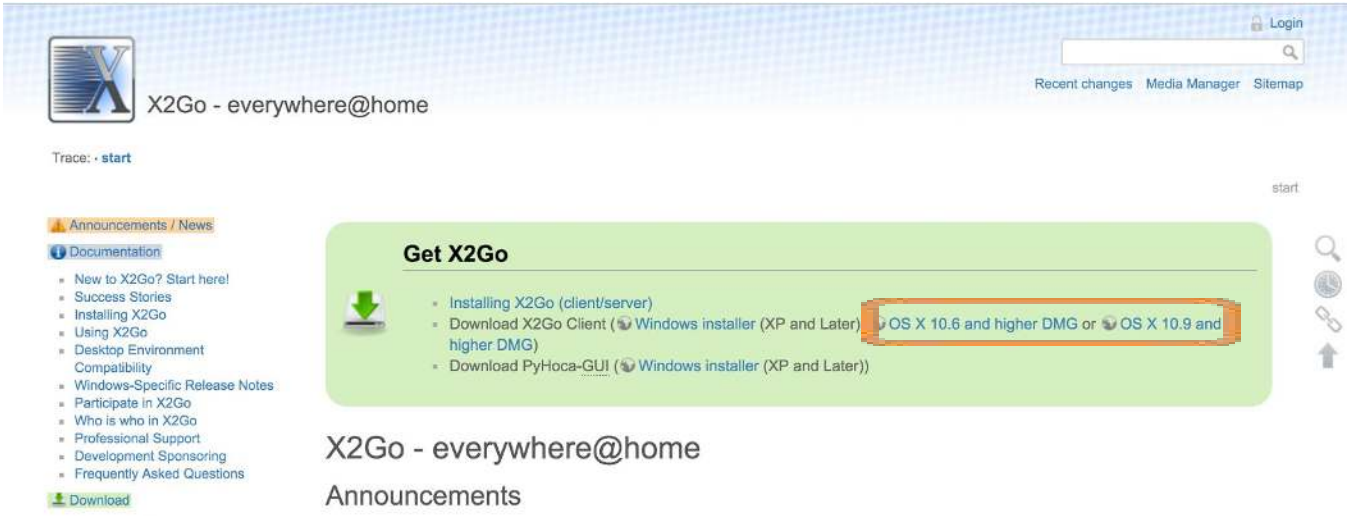


Once it is downloaded, just click on the **XQuartz-2.7.8.dmg** and then open the **XQuartz.pkg**. Follow the standard installation procedures until you reach the following screen :



Once XQuartz is installed, you'll then be able to install X2Go.

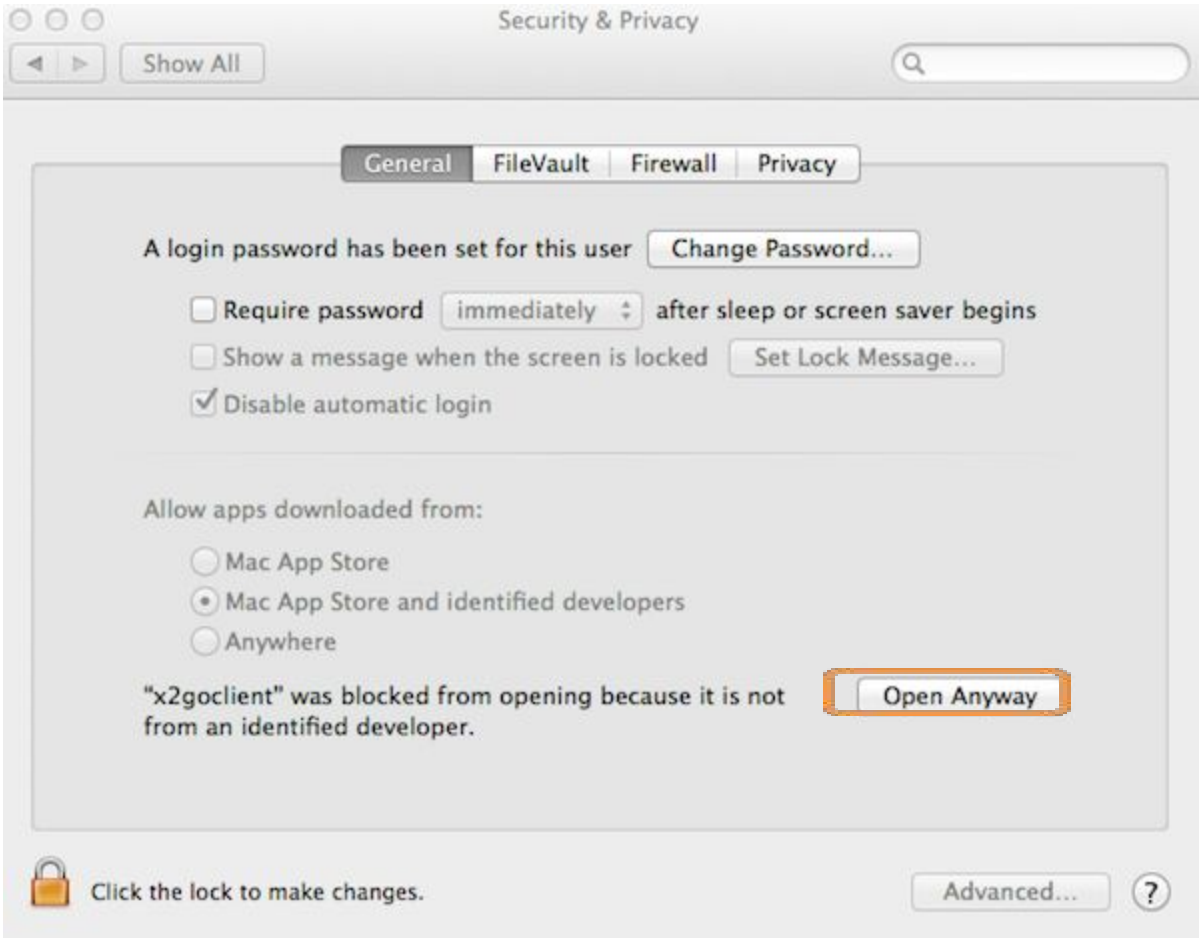
Go to http://wiki.x2go.org/doku.php and click on the appropriate MacOS dmg for your computer to download.



The X2Go download page. Please click on the "**OS X 10.6 and higher DMG**" or "**OS X 10.9 and higher DMG"** link.

Once downloaded, try to open the **X2GoClient_latest_macosx_10_?.dmg** file. You may get the following warning:

If you see this message, to open the file, please navigate to your Security and Privacy settings. You can do this by going to your System Preferences (under the Apple menu in the upper left of your screen) and then clicking on "**Security and Privacy**."

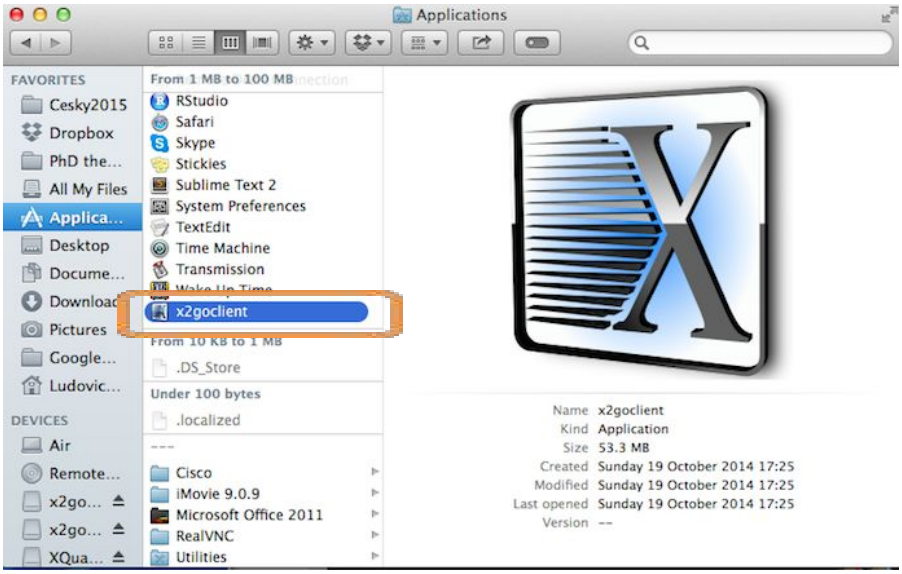


OSX's Security and Privacy settings. Please "**Open Anyway**" for the x2goclient.

You can now go back and open the **X2GoClient_latest_macosx.dmg** file. Once open, please move **x2go** into your **Applications** folder.

Starting the X2Go client on Mac OSX

From Applications, open **x2goclient**.



X2Go on Linux

Installing X2Go client on Ubuntu (Linux).

X2Go Client is part of Ubuntu 12.04 & later, as well as Debian Wheezy & Jessie. In Ubuntu, to install it you will probably need admin rights (sudo, root, etc.):

- Open a terminal (Ctrl + Alt + T)
- In the terminal, type "`sudo apt-get install x2goclient`"

```
honza@jan-msi:~$ sudo apt-get install x2goclient
```

Detailed instructions for other Linux flavours can be found in this link:
http://wiki.x2go.org/doku.php/doc:installation:x2goclient

Start X2Go client in Ubuntu (Linux).

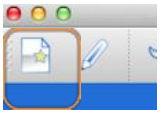
In a terminal, type "`x2goclient`".

```
honza@jan-msi:~$ x2goclient
```

Create a session with X2Go client (all OS).

After launching X2Go in your OS (see above), you should see the main screen. **Note: on some versions of Windows you might get a security message. If so, please select "keep blocking."**

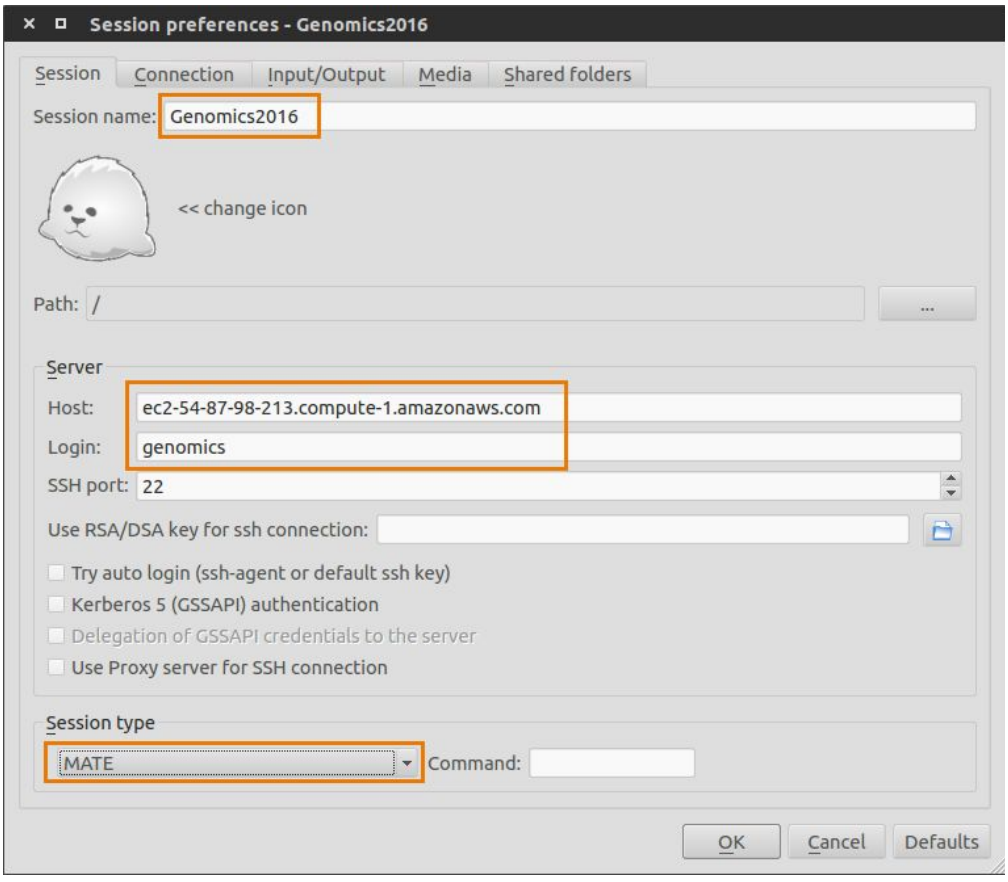
Now you need to tell your computer where to connect to. If this is the first time you've opened X2Go, a new dialog will automatically pop up. If this is not the first time, then you'll need to click on the "New session" icon.



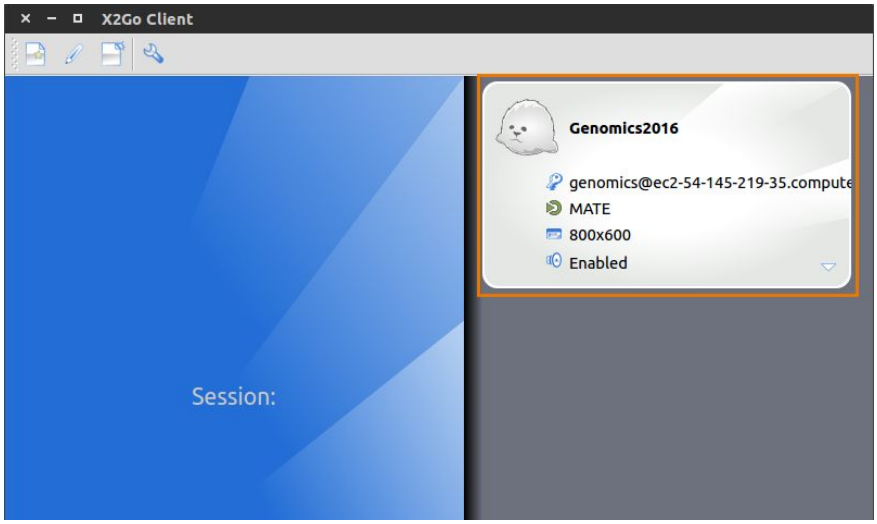
Within the new session dialog box, you'll need to specify:

- *A session name*. We recommend "**Genomics 2016**".
- *A host*. This is your instance. Please enter the **Public DNS** of your EC2 instance (copied from the Amazon console in your web browser).

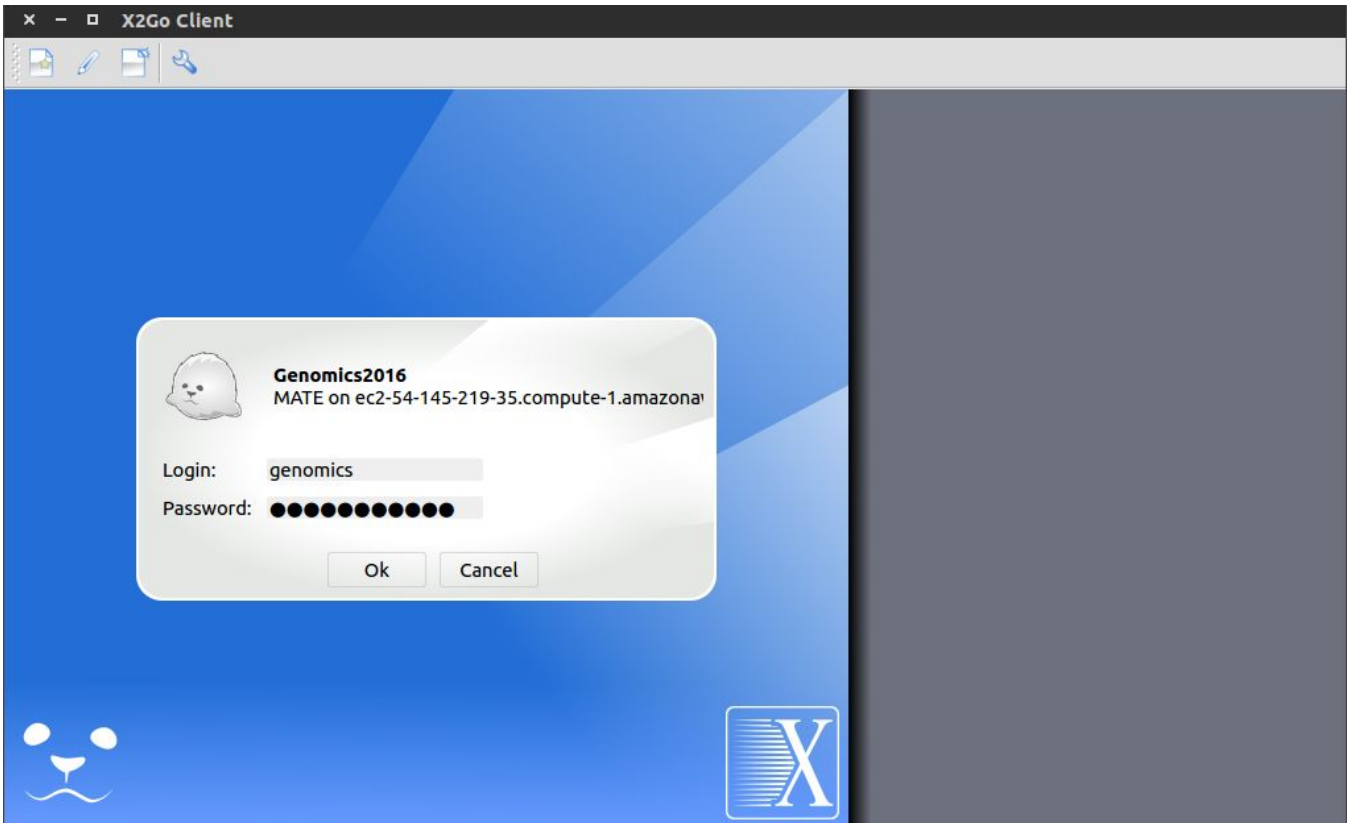
- *A login*. This is the username. Please enter "**genomics**".
- *The session type*. Please select "**MATE**".



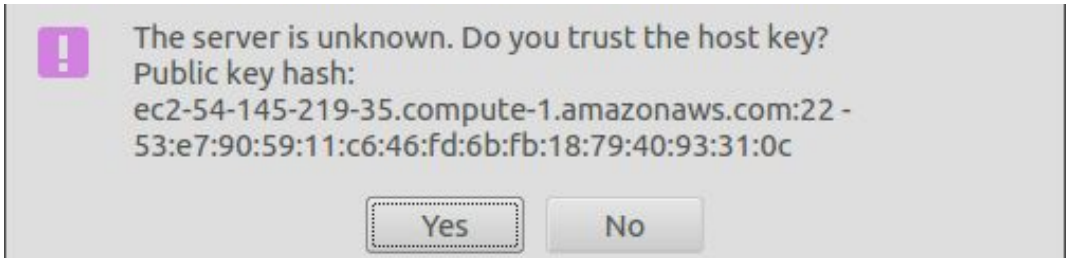
When you click on OK, you should be taken to a new page that looks like the following.



Click anywhere on the white area. This will bring up a new prompt that will allow you to enter a password. Please enter "**evomics2016**" as the password.



The first time you connect to your instance (or if the public DNS changes) you will see a message that looks something like this (look of this error is dependent on your system, so it does not have to be exactly same as here) :

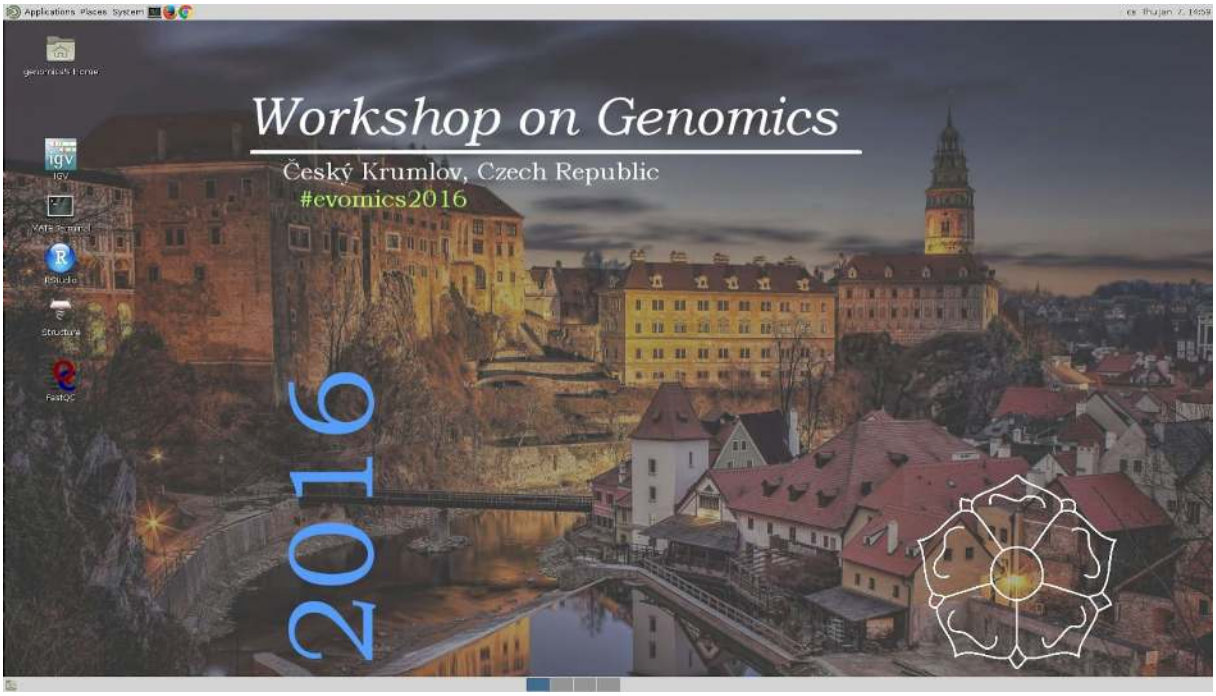


Simply click next/yes to continue.

Note : If you are using Mac, you will see two error messages one after the other, just ignore them.

After approximately 30 seconds, you should see the connection open as below.

Congratulations!



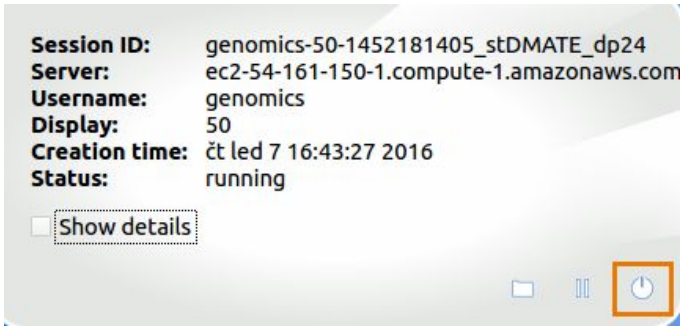
Connection Management

At the end of a working session, first we will need to suspend the connection from X2Go to the Amazon Cloud, then stop the Amazon Cloud instance.

Suspending and reconnecting connection in X2Go

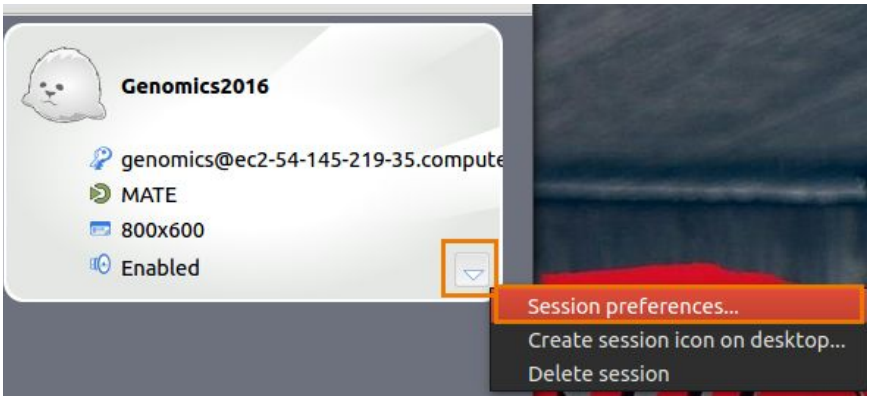
Disconnecting

When you disconnect from X2Go, this will close all your windows and log you off the instance. **The Instance will still be running so make sure you stop in the AWS console – see below.** To disconnect, please go to your X2Go terminal, and click the button on the bottom right.



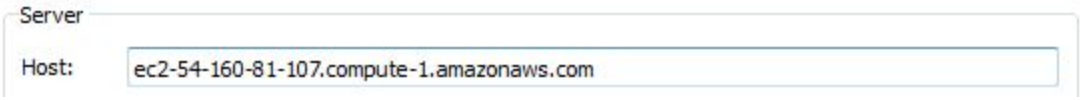
Reconnecting

Whenever you restart your AMI on amazon (see later), your public DNS will change and you will need to update it in the X2Go client. To update, please navigate to your session preferences.



An example of navigating to your X2Go session preferences.

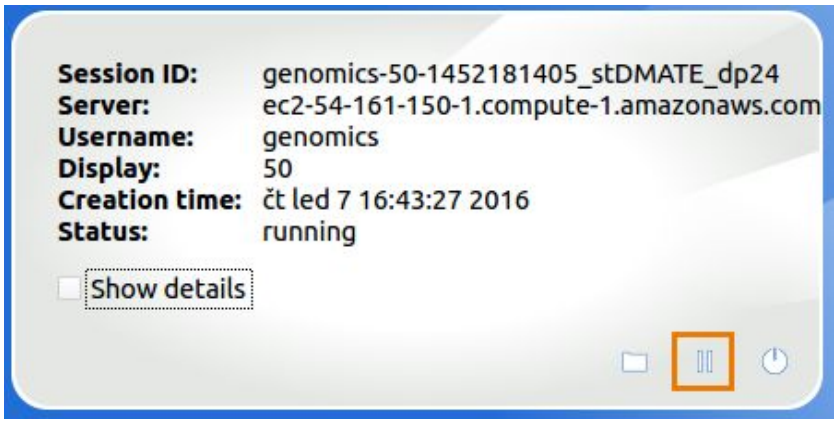
Within the session preferences, please go to the Session tab. You can modify the "**host**", which is where you need to specify your new **Public DNS**.



*The host field in the Session tab of your Session preferences. The value in this field needs to be the same as your **Public DNS** of your EC2 instance.*

Suspending

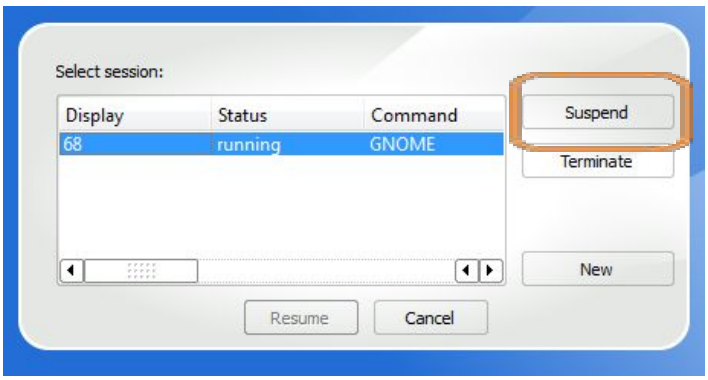
If you want to close X2Go but leave windows open and running, instead of disconnecting, you can alternatively suspend the session.



Within your session details, if you'd like to suspend the session, please click the pause button.

You can now resume your session later on from exactly where you left off.

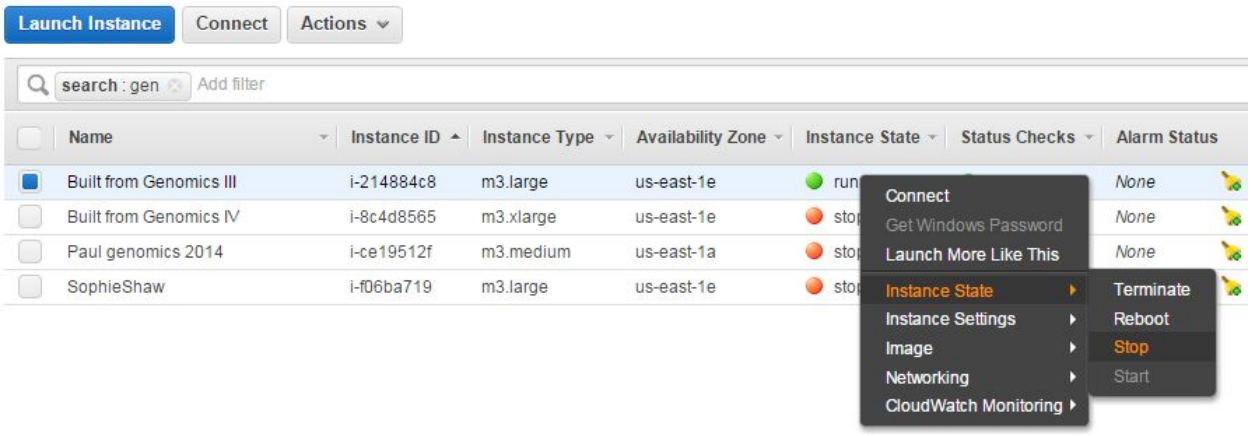
If your desktop computer crashes or disconnects for any reason - your session should still be running. When you try to log on you will see a this window:



Note that 'Resume' is greyed out…
Click on suspend first and then you can click on resume.

Stopping and starting the Instance in Amazon

When you're not working on the course it's very important to turn off your instance to avoid unnecessary charges. Log back onto the AWS console and find your instance, right click the instance to get the menu.



The AWS EC2 instance viewer. The above image shows the navigation menu that appears when you “right-click” on an instance.

If you wish to keep your data, use the 'Stop' option. In this state you will not be charged for computing time, but will still be charged for storage.

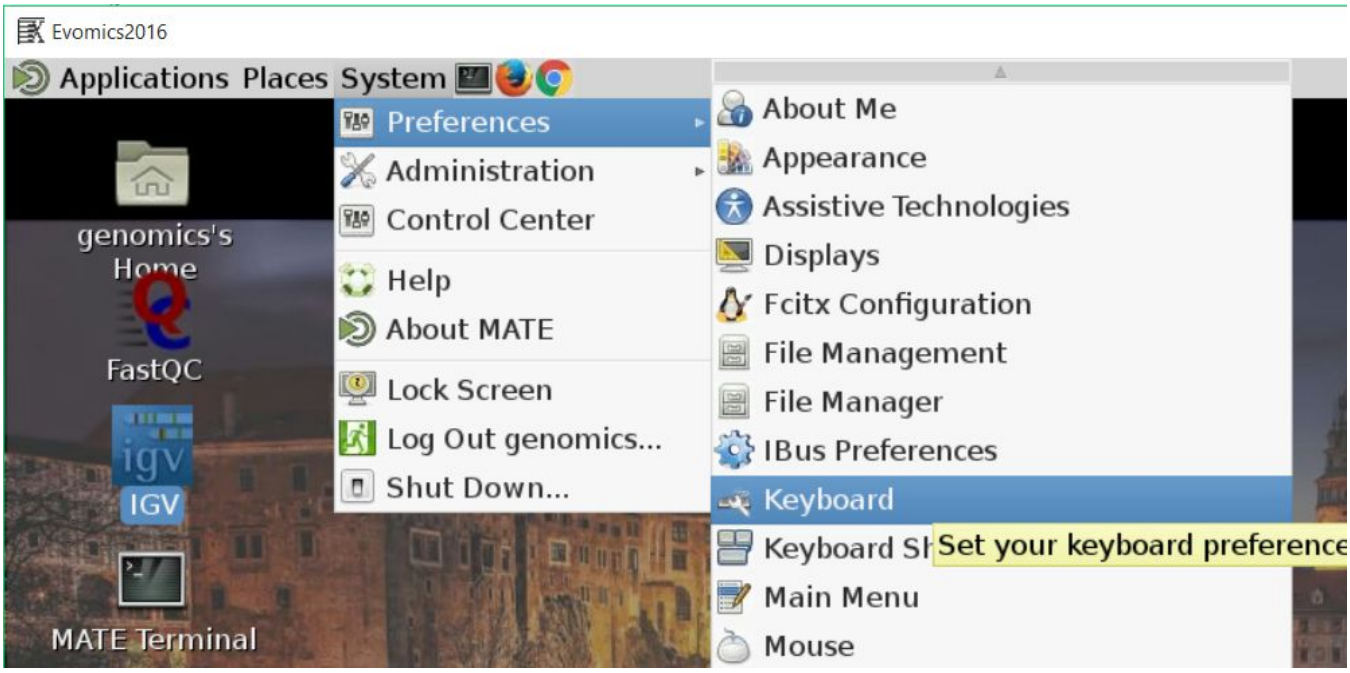
When you're completely finished with the workshop, right click the instance and select “Terminate”. It will ask you to confirm. You can then watch the status change from “shutting down” to “terminated”. Note – this will destroy all work done to date.

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 X2Go.

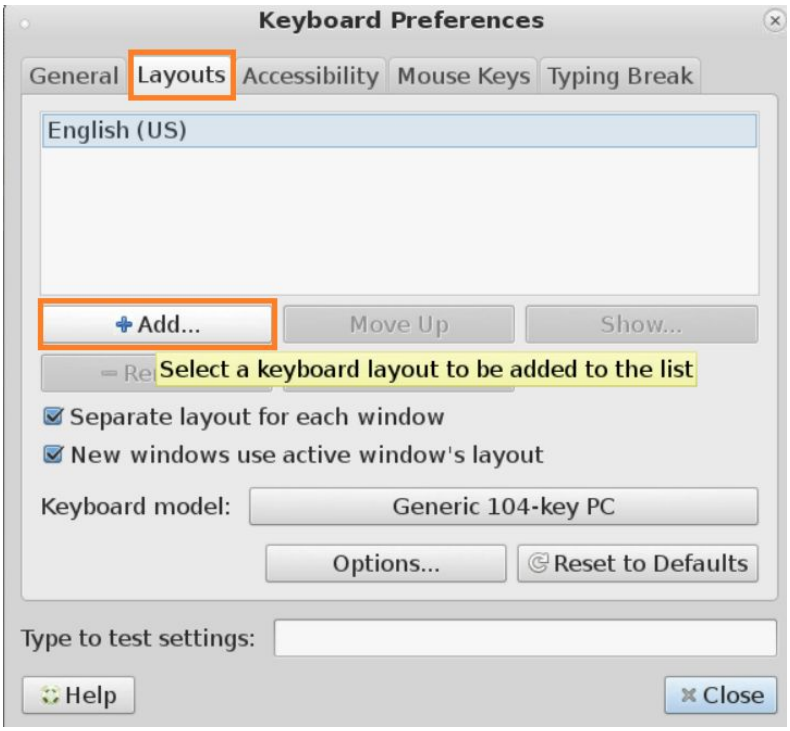
Although much of what we have just done may not make much sense yet, most of you will feel totally comfortable and confident working on an EC2 node running Linux within a few hours. It's really amazing how quickly the fact that this is a remote computer will fade away. It may be hundreds of miles away but it will act just like a local computer, especially if you connect via X2go.

Troubleshooting the keyboard layout.

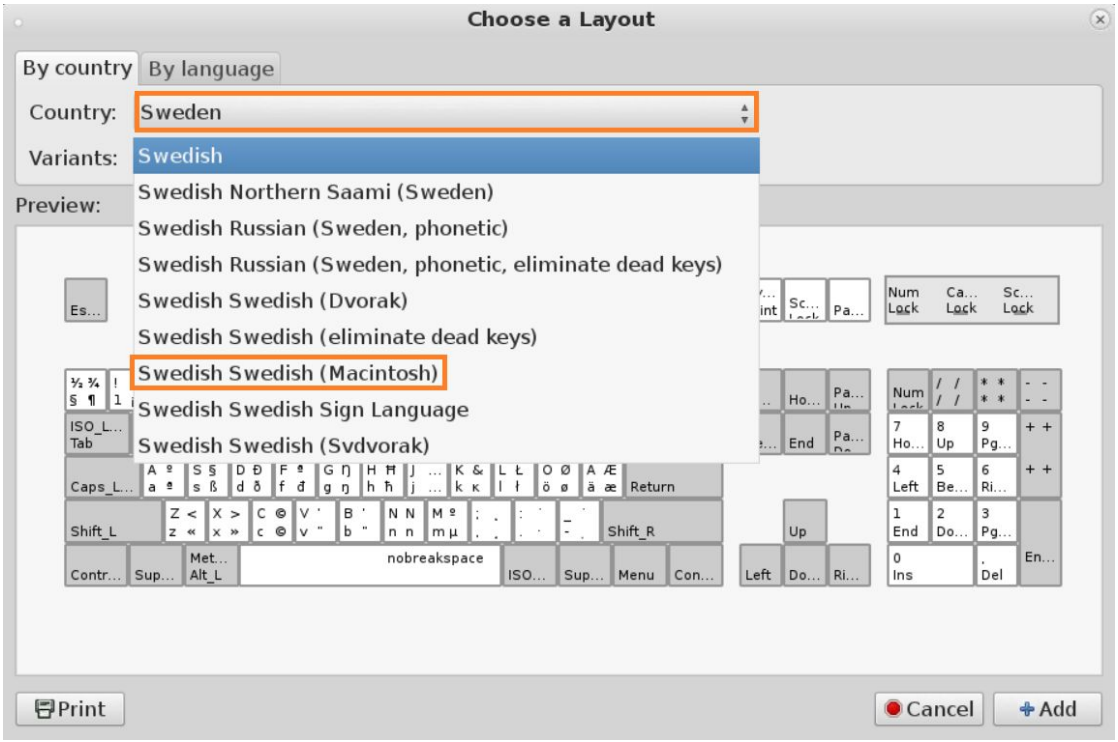
In most cases X2Go should transfer the layout from your computer to the virtual instance.
If for some reason this does not work for you, there is an option to change the layout in the instance itself.
Click on “**System**” on the top bar on the instance desktop. Then go “**Preferences**” and “**Keyboard**”.



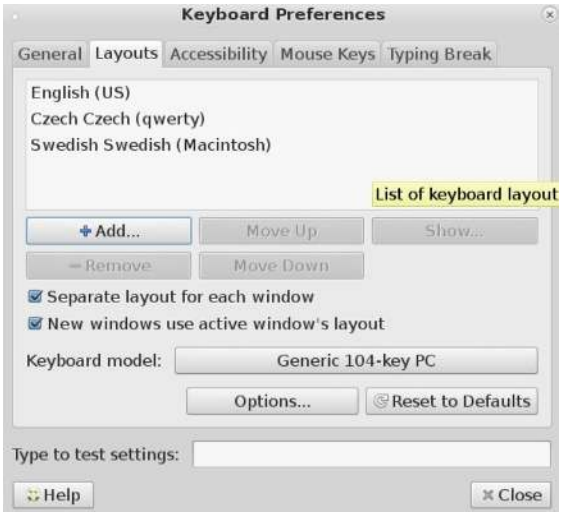
A keyboard preferences menu will appear. Click the “**Layout**” tab and then the “Add” button.



Then you can choose the desired layout by country and variant. Choose the one you like and you can preview the layout on a scheme just below the selection. When you are happy click the “**Add**” button in the bottom.



After this a new options for layout should appear in the main keyboard menu like so:



You can then switch the keyboard by clicking a button on the top left of your instance desktop:

Here is and short list of most used keys that you should get familiar with:

sign	name	czech keyboard	swedish keyboard
/	forward slash	shift + ú	shift +7
\	backslash	rAlt + Q	alt + shift +7
|	pipe	rAlt + W	alt + 7
*	asterisk	rAlt + -	shift + ‘
$	dollar sign	rAlt + ů	alt + 4
#	hash	rAlt + x	shift + 3
>	greater than sign	rAlt + .	separate key
<	lesser than sign	rAlt + ,	separate key
~	tilde	rAlt + 1	alt + ¨/^ (the key with two dots)

Optional – Log into the Running instance via SSH

This is intended 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 genomics@public-dns-numbers.<amazonawsDNS>.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 genomics@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)

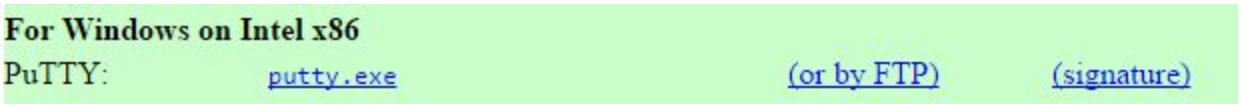
Windows Tip:

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. Instructions for both of these programmes can be found below.

Note, again if you're using a instance 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.

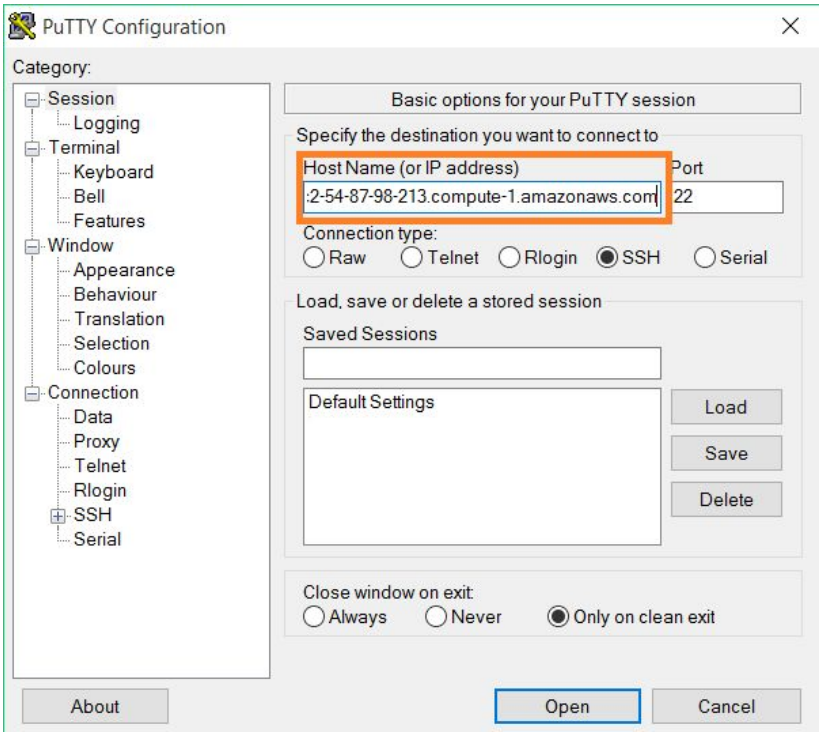
Using PuTTY (Windows Only)

PuTTY is a SSH terminal for Windows. It can be used to access our instance as a terminal. To download putty go to http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html and select the version you want. In this case it should be putty.exe



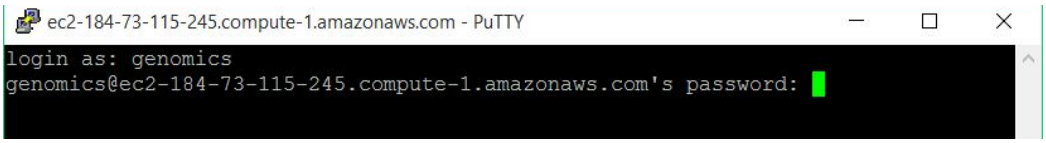
After downloading the file, run it.

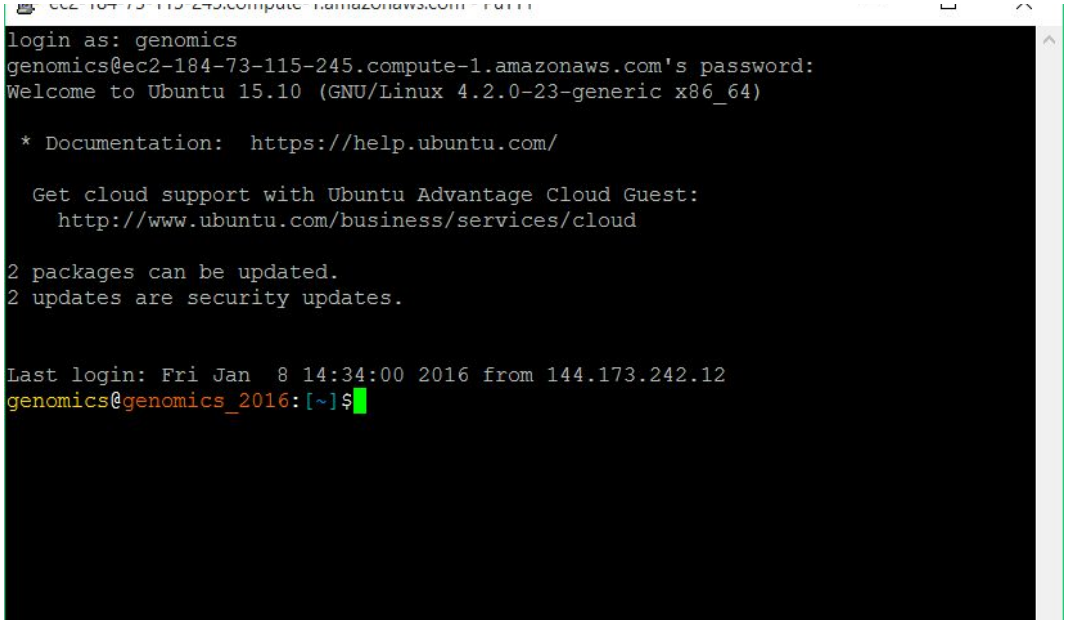
In the Host Name bar insert the Public DNS number for the Amazon Instance.



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

Enter the username “genomics” and the password “evomics2016”.





You are now accessing the terminal of your instance. Yeei!\

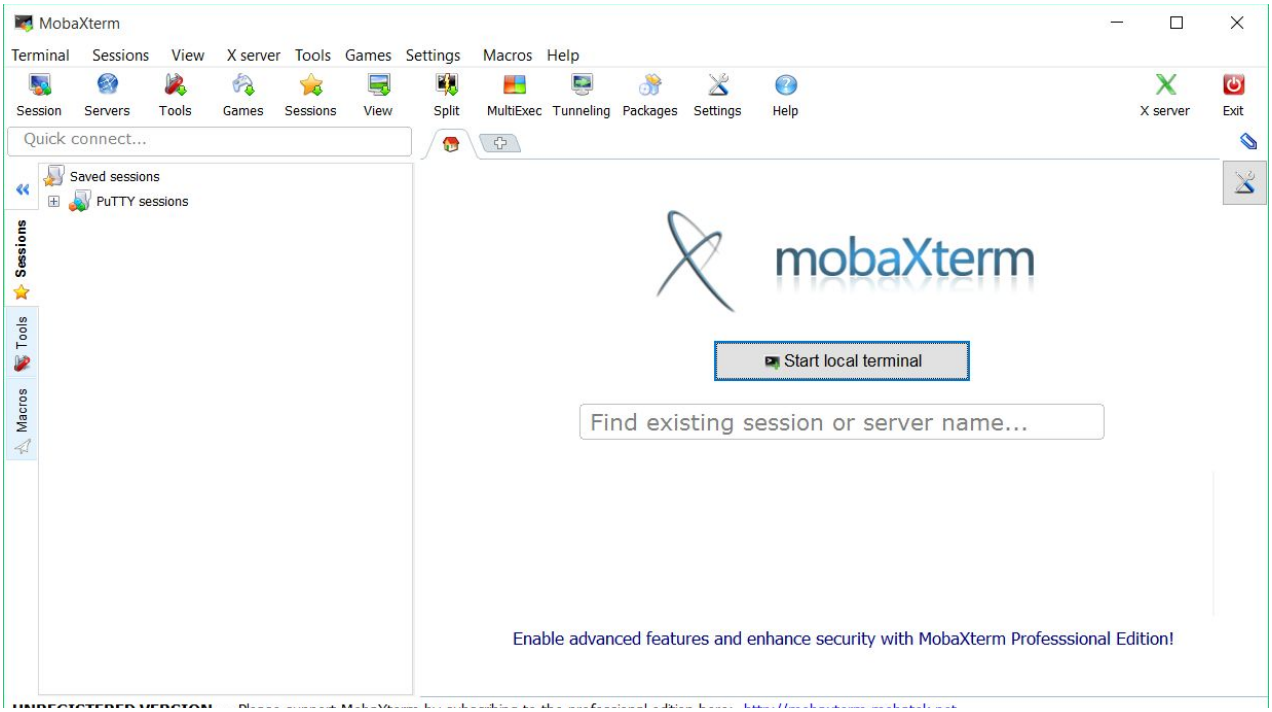
Using MobaXTerm (Windows Only)

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.

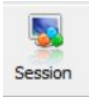
Download here - http://mobaxterm.mobatek.net/MobaXterm_v8.5.zip

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….

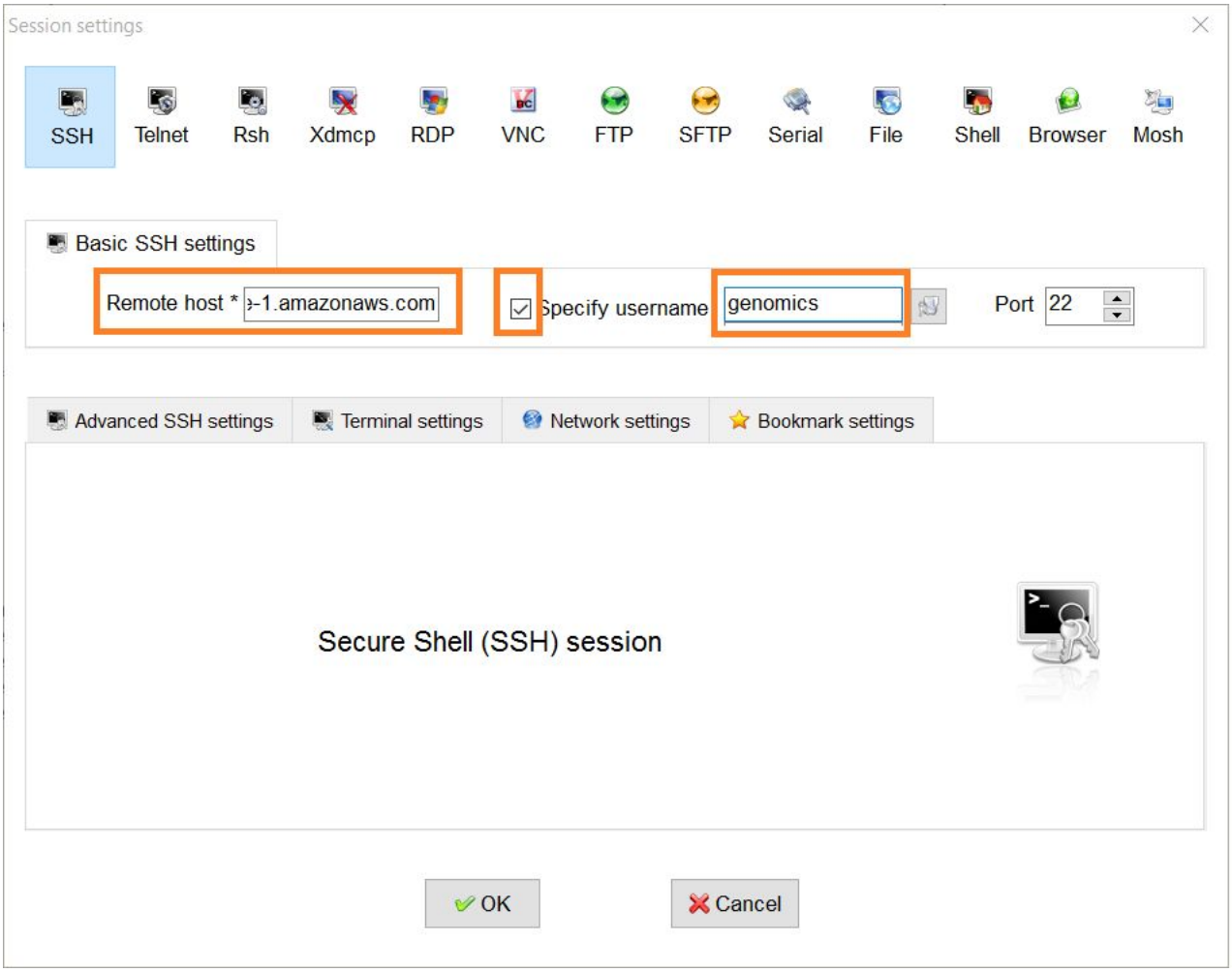
When the program has started you will be shown a screen like below:



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 “genomics”. Click OK and you will then be asked for your password, “evomics2016”, 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.

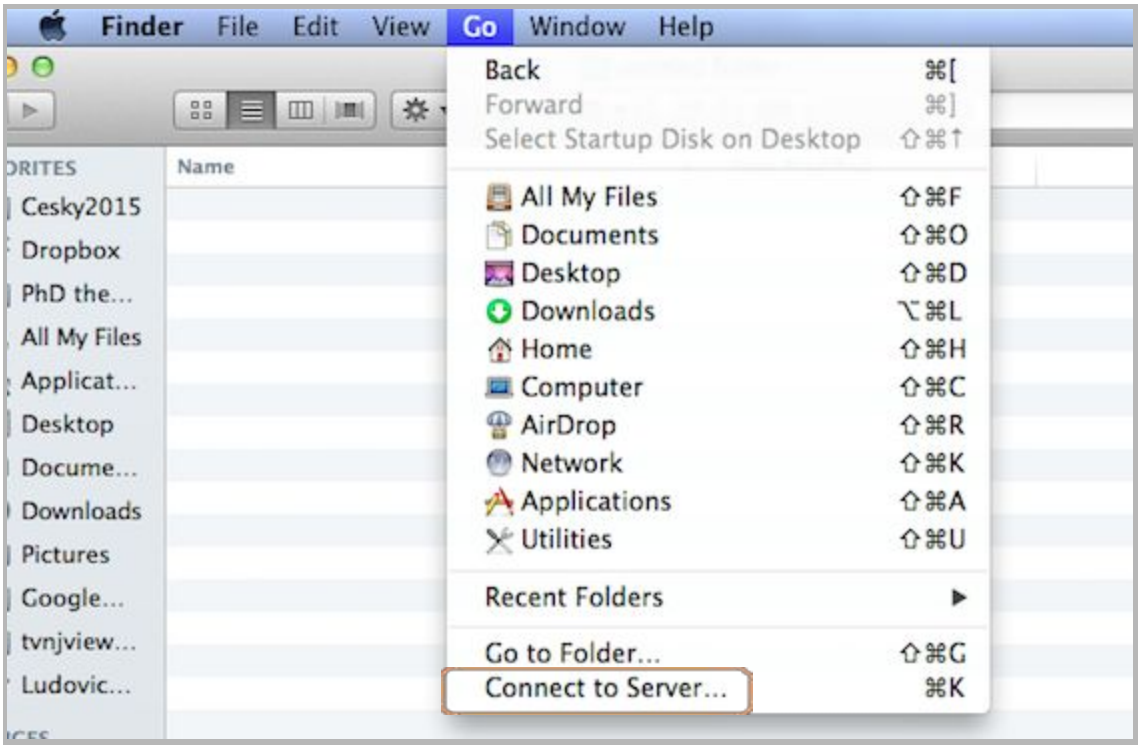
MobaXTerm should also save all your session details, including passwords and private keys between sessions of using it. Your saved sessions will appear on the left hand side of your program screen.

Full installation is also possible if you do not want the portable version. The link for download is here: http://mobaxterm.mobatek.net/MobaXterm_Setup_8.5.msi

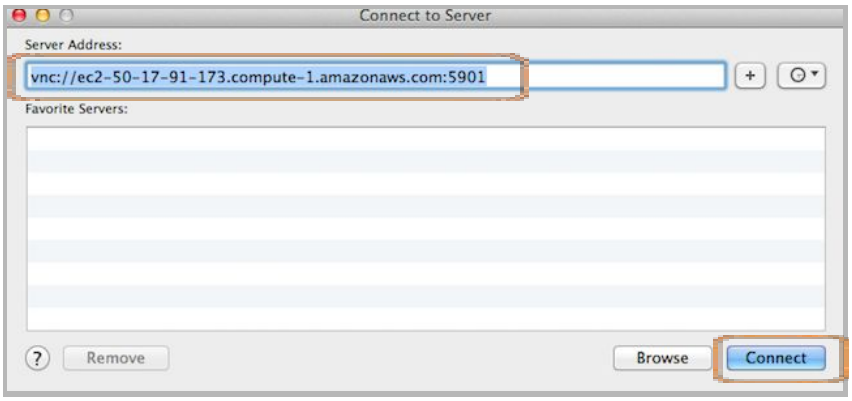
VNC Connection from OS X:

In the event that you **cannot** get x2go running from your OSX system, you can try the following VNC connection. **You very likely do not need to go through this section!**

Go in the finder, Select *Go> Connect to server*



In the Finder, click the "**Go**" pull down menu, and select "**Connect to Server…**".



Enter “vnc://**Public DNS**:5901” in the server address section. Please make sure to replace **Public DNS** with the public DNS of your instance. Then click “**Connect**.”

When prompted for a password, enter “**evomics2016**”.