

2014 Harvard University Center for AIDS Research

Workshop on Metagenomics and Transcriptomics

Starting your Amazon Virtual Machine

Created by Scott A. Handley and Konrad Paszkiewicz

Last updated: September 14, 2014

OBJECTIVES

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

- Log into the Amazon Console and start your instance
- Log into an Amazon EC2 Linux instance using your own computer using NoMachine
- Explore Ubuntu and NoMachine

INTRODUCTION

For this Workshop we will provide an overview of the cloud as described by Amazon, and demonstrate how researchers can use this flexible infrastructure to get work done quickly and relatively inexpensively using Amazon's computing resources.

We will dive right in by starting a prepared Virtual Machine (VM) and logging into the Amazon Management Console and give you a whirlwind tour of Amazon's cloud features.

For this tutorial, documentation was borrowed from the following sites:

- <http://ged.msu.edu/angus/tutorials/>
- <http://aws.amazon.com/documentation/>
- Many thanks to Konrad Paszkiewicz from the University of Exeter for generating the primary version of this tutorial!

TOUR OF AMAZON'S CLOUD INFRASTRUCTURE

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, but are not limited to:

- **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) as well
- **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 of these tools may be valuable for your research!

The Rules

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

1. Only launch a **single VM** instance of the type specified by the instructor
2. **Do not** leave the image running overnight
3. Provide a **unique name** to your instance. Instructions for doing so are provided in the tutorial below
4. Do not **delete or terminate** VM instances that do not belong to you

Rules are subject to change, but this basic guideline is important for distributing resources fairly and ensuring consistent access for everyone in the course.

Why use Amazon's cloud infrastructure?

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, login to them, do some work, transfer data to/from them or destroy them altogether. Amazon will take care of 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 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 computer for long-term, data intensive analysis, although owing your own hardware come with 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.

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 to use.

This workshop has its own Amazon account and each of you have been setup as a subaccount using something called Identity Access Manager (IAM). 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 subaccounts can and cannot do.

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

<https://hu-cfar-2014.signin.aws.amazon.com/console>

username: student

password: stopaids

STEP 2 - Creating and managing your instance

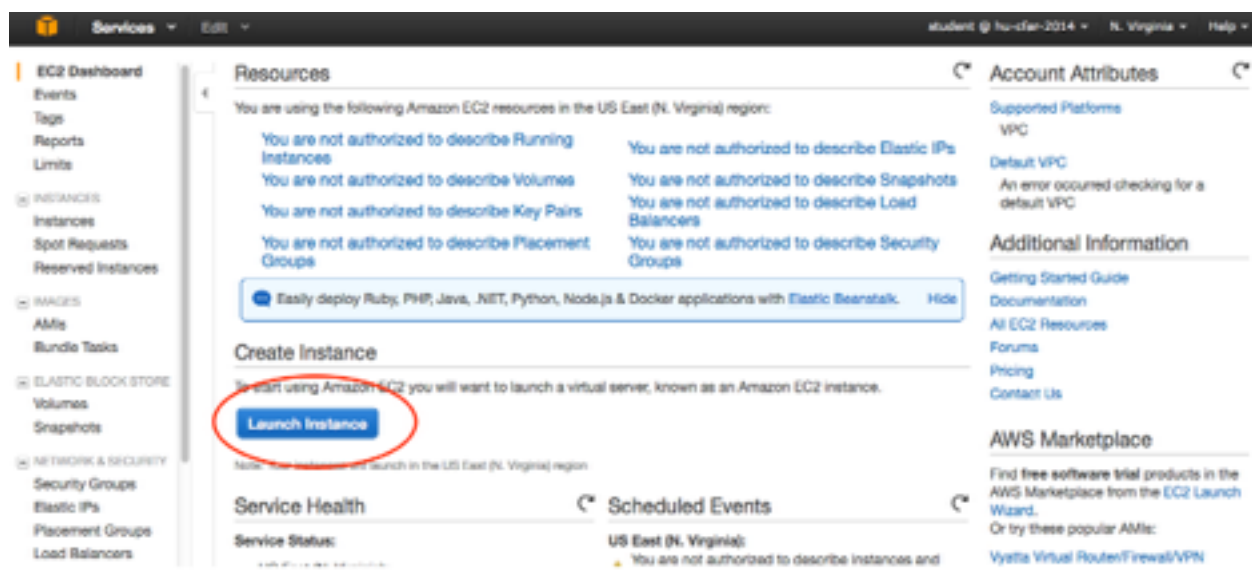
After logging in, you will be presented with a wide-array of options.

- Select **EC2** under **Compute and Networking**

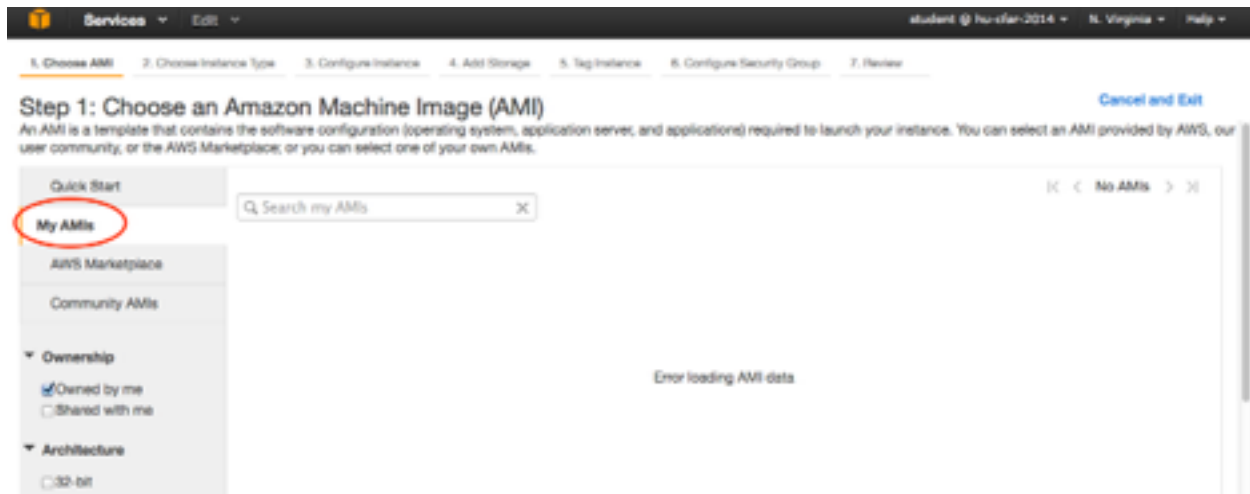


You will be taken to a page summarizing your instances and services.

- Click on the **Launch Instance** button towards the middle of the page



Now click **My AMIs** to continue.



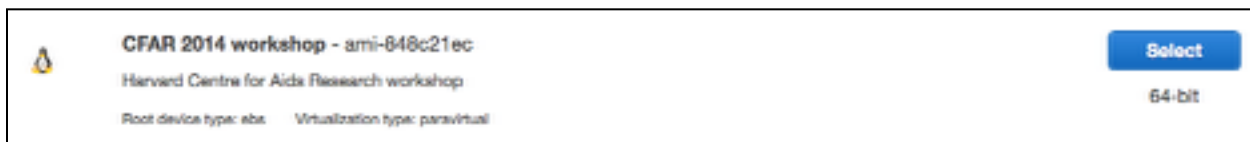
Ensure that you are logged into Amazon's **N. Virginia** servers (displayed in the upper-right corner of the page). If you are not, you can change this by clicking on the current region and selecting **US East (N. Virginia)** from the drop-down box.

It is possible that the Workshop instance will automatically appear in instance selection box. If you see "Error loading AMI data" as in the screenshot above you will need to search for our custom prepared AMI.

- Select **Community AMIs** from the list on the left of the page
- Enter the following code into the **Search my AMIs** search box.

ami-848c21ec

Once you have located the AMI, click "Select" to begin the log-in process.



Please do not use any other AMI for this Workshop!

Once you have selected the instance, change the **Filter by:** box to **General Purpose**.

For this course we will be using General Purpose **m1.xlarge** instance which you will find by scrolling to the bottom of the General Purpose list. 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/instancetype/> for information about the instance types and <http://aws.amazon.com/ec2/pricing/> for information about pricing per hour.

Step 2: Choose an Instance Type

Amazon EC2 provides a wide selection of instance types optimized to fit different use cases. Instances are virtual servers that can run applications. They have varying combinations of CPU, memory, storage, and networking capacity, and give you the flexibility to choose the appropriate mix of resources for your applications. [Learn more](#) about instance types and how they can meet your computing needs.

Filter by: **General purpose** All generations Show/Hide Columns

Currently selected: t1.micro (Variable ECUs, 1 vCPUs, 0.613 GiB memory, EBS only)

	Family	Type	vCPUs	Memory (GiB)	Instance Storage (GiB)	EBS-Optimized Available	Network Performance
<input checked="" type="radio"/>	General purpose	t2.micro <i>Free tier eligible</i>	1	1	EBS only	-	Low to Moderate
<input type="radio"/>	General purpose	t2.small	1	2	EBS only	-	Low to Moderate
<input type="radio"/>	General purpose	t2.medium	2	4	EBS only	-	Low to Moderate
<input type="radio"/>	General purpose	m3.medium	1	3.75	1 x 4 (SSD)	-	Moderate

Cancel Previous **Review and Launch** Next: Configure Instance Details

Once you have selected **m1.xlarge** instance click **Next: Configure Instance Details**.

You will be taken to a page that looks like the screenshot below. We will just be using the default settings, so for now just go ahead and click: **Next: Add Storage**.

Services Edit student @ hu-cfar-2014 N. Virginia Help

1. Choose AMI 2. Choose Instance Type 3. Configure Instance 4. Add Storage 5. Tag Instance 6. Configure Security Group 7. Review

Step 3: Configure Instance Details

Configure the instance to suit your requirements. You can launch multiple instances from the same AMI, request Spot instances to take advantage of the lower pricing, assign an access management role to the instance, and more.

Number of instances (i) 1

Purchasing option (i) ☐ Request Spot Instances

Network (i) vpc-94fe34ff (172.31.0.0/16) (default) Create new VPC

Subnet (i) subnet-6294c492 (172.31.0.0/20) | Default in vpc-94fe34ff | 4091 IP Addresses available Create new subnet

Auto-assign Public IP (i) ☐ Use subnet setting (Enable)

IAM role (i) None

Shutdown behavior (i) Stop

Cancel Previous **Review and Launch** Next: Add Storage

On the **Configure Instance Details** page you are able to select how much storage you will need for your jobs. For this Workshop, we will just use the default 500 GB. Also ensure that the **Delete on Termination** box is selected. In real life we would recommend against this as you could loose valuable data, but for this course it will be fine.

Type	Device	Snapshot	Size (GiB)	Volume Type	IOPS	Delete on Termination	Encrypted
Root	/dev/sda1	snap-0c3404af	500	General Purpose (SSD)	1500 / 3000	<input checked="" type="checkbox"/>	Not Encrypted

[Add New Volume](#)

Free tier eligible customers can get up to 30 GiB of EBS General Purpose (SSD) or Magnetic storage. [Learn more](#) about free usage tier eligibility and usage restrictions.

[Cancel](#) [Previous](#) [Review and Launch](#) [Next: Tag Instance](#)

Select the **Next: Tag Instance** button to be taken to a page where you can name your instance. This is a very important stage for our management of the system, so please complete and choose an easily identifiable name. For example, your first and last name.

Key (127 characters maximum) **Value** (255 characters maximum)

Name

[Create Tag](#) (Up to 10 tags maximum)

[Cancel](#) [Previous](#) [Review and Launch](#) [Next: Configure Security Group](#)

Once you have named (tagged) your instance, select **Next: Configure Security Group**.

On the **Configure Security Group** page, select radio button for **Create a new security group**.

Click the **Review and Launch** button at the bottom of the page.

Step 6: Configure Security Group

A security group is a set of firewall rules that control the traffic for your instance. On this page, you can add rules to allow specific traffic to reach your instance. For example, if you want to set up a web server and allow Internet traffic to reach your instance, add rules that allow unrestricted access to the HTTP and HTTPS ports. You can create a new security group or select from an existing one below. [Learn more](#) about Amazon EC2 security groups.

Assign a security group: ☒ Create a new security group
☐ Select an existing security group

Security group name:

Description:

Type	Protocol	Port Range	Source
SSH	TCP	22	Anywhere 0.0.0.0/0

[Add Rule](#)

Warning

Rules with source of 0.0.0.0/0 allow all IP addresses to access your instance. We recommend setting security group rules to allow access from known IP addresses only.

The next step will allow you to review your setting and then Launch the instance which will establish it on one of Amazon's servers in N. Virginia. Go ahead and click on **Launch** at the bottom of the page.

You may receive an additional pop-up window asking about **General Purpose (SSD)**. Just click **Next** and move on.

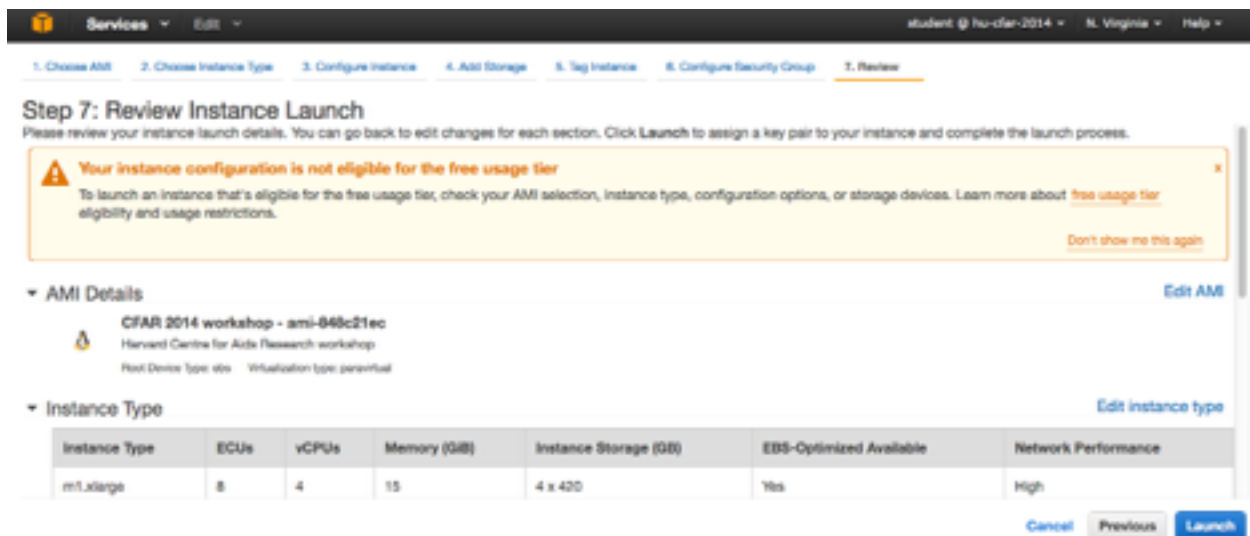
Boot from General Purpose (SSD)

General Purpose (SSD) volumes provide the ability to burst to 3,000 IOPS per volume, independent of volume size, to meet the performance needs of most applications and also deliver a consistent baseline of 3 IOPS/GiB.

- ☒ Make General Purpose (SSD) the default boot volume for all instance launches from the console going forward (recommended).
- ☐ Make General Purpose (SSD) the boot volume for this instance.
- ☐ Continue with Magnetic as the boot volume for this instance.

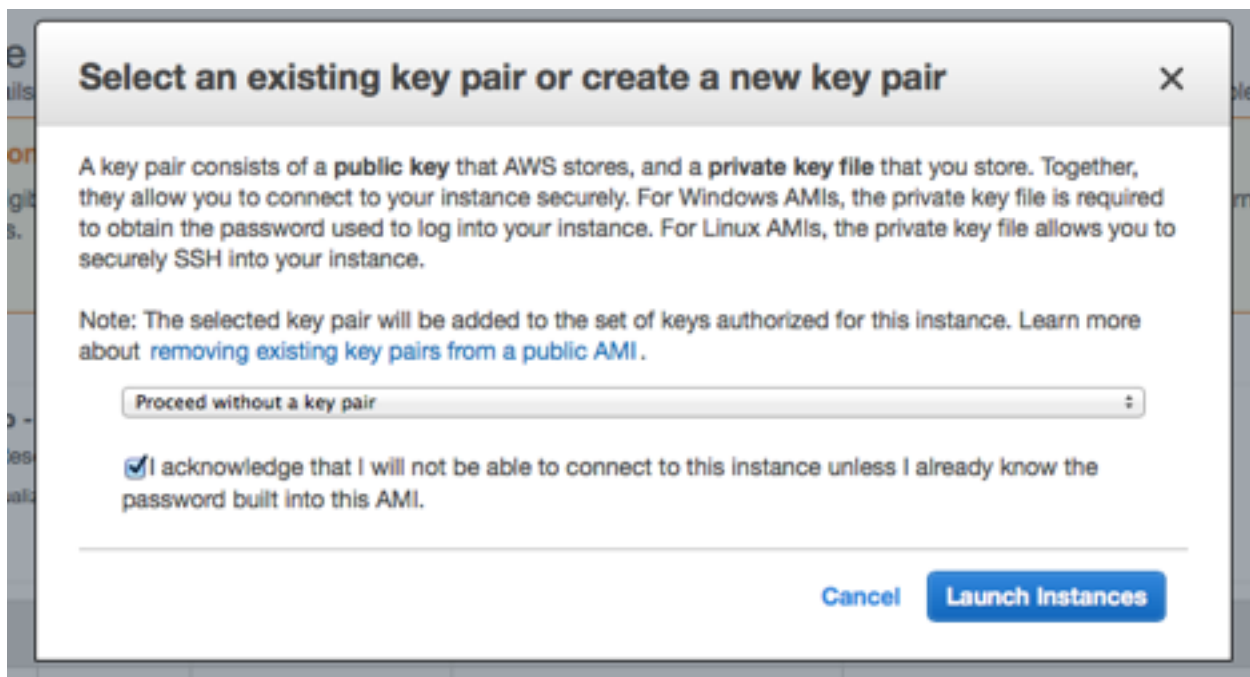
Free tier eligible customers can get up to 30GB of General Purpose (SSD) storage.

[Next](#)



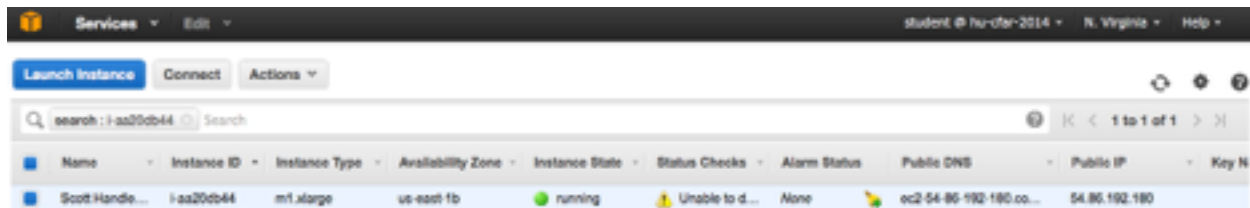
You will be presented with a window that asks about creating key pairs. We will not be using key pairs during this Workshop, so change the top drop down box to **Proceed without a key pair** to continue.

Select the check box to acknowledge that you know what you are doing and continue by clicking **Launch Instances**.



You will be taken to a page that looks like the screenshot below. Click on **View Instances** at the bottom of the page.

At this stage it may take some time for the AMI to come on-line. Give it a few minutes to finish. When it has fully initiated you will see a green "light" next to the word running and you will have just virtualized an instance on Amazon's servers in Virginia!



Once you have the green light, select the instance and **copy the Public DNS address to your clipboard**. You will need this number in a moment.

STEP 3 - Log into the Running VM Remote Desktop with NoMachine

Whilst your AMI is initializing (please note it may take some time) take this opportunity to install the NoMachine (sometimes referred to as NX) client software which we will be using to connect to your newly launched AMI.

NoMachine will allow you to see a windowing environment (like your Desktop) rather than just a terminal! This is a great option if you want to use a GUI application (Graphical User Interface like IGV, RStudio, MEGAN).

Note: Many AMIs will not have the NX server installed and therefore you will not be able to connect using the NX client. However, we have installed the NX Server on the Workshop AMI so it is usable for the duration of the Workshop, but is not guaranteed to be on other AMI's.

1. Download an NX Client

Use the following NX clients depending on your operating system:

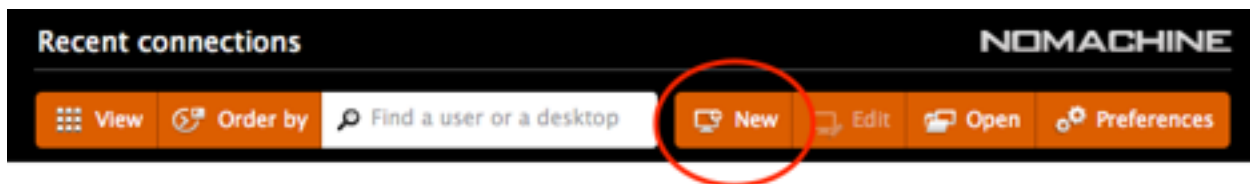
- Max OSX: <https://www.nomachine.com/download/download&id=7>
- Windows 7 or 8: <https://www.nomachine.com/download/download&id=8>
- Windows XP: Please try the Windows 7 version first. If this doesn't work, you can try a legacy version which can be found here: <http://nx-client-for-windows.software.informer.com/3.5/>. The screenshots for the remainder of the tutorial will not look the same. Ask a TA for assistance if you need it.
- 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

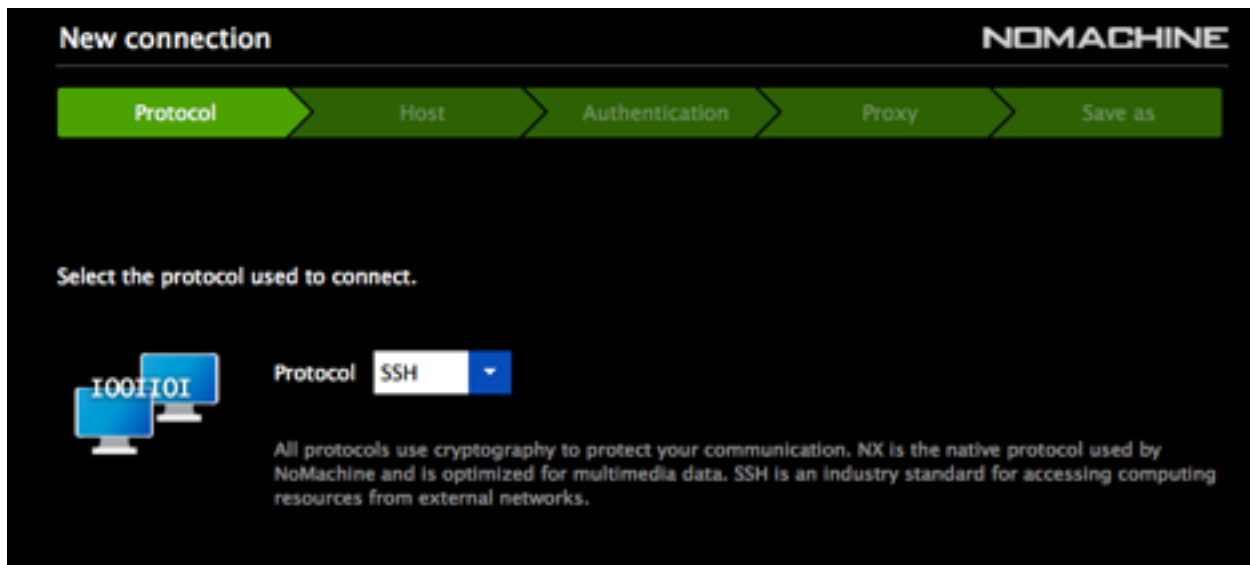
Once installed, start the NoMachine client on your laptop.

The first time you launch the program you will see some introductory help screens. Review these and click **Next/Continue** until they are finished.

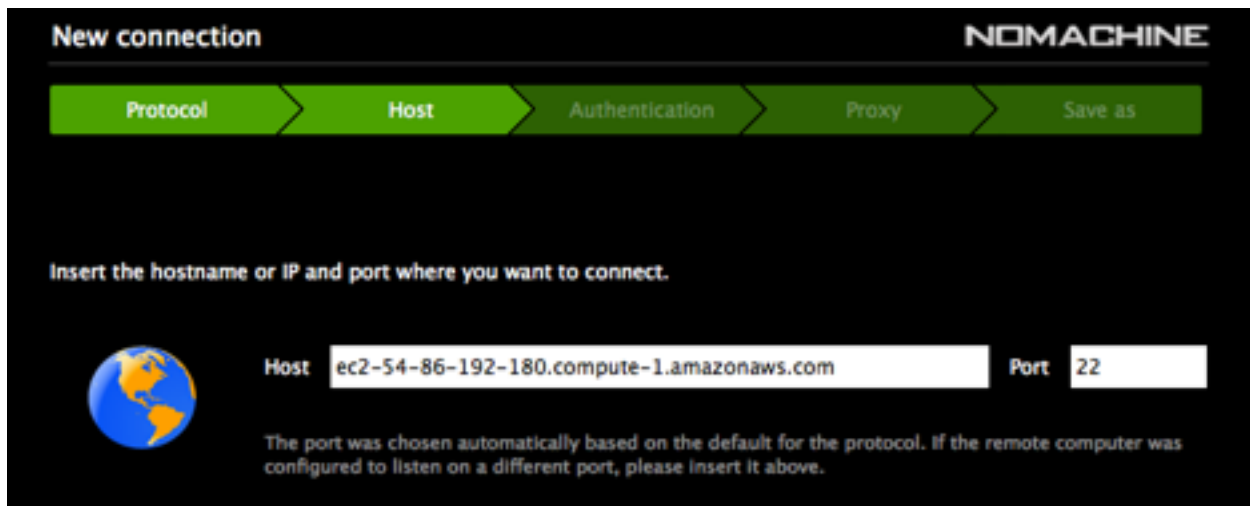
You can establish a new connection by selecting the icon indicated in the screenshot below:



The first step will be to select the connection protocol. Change the protocol box to **SSH**.

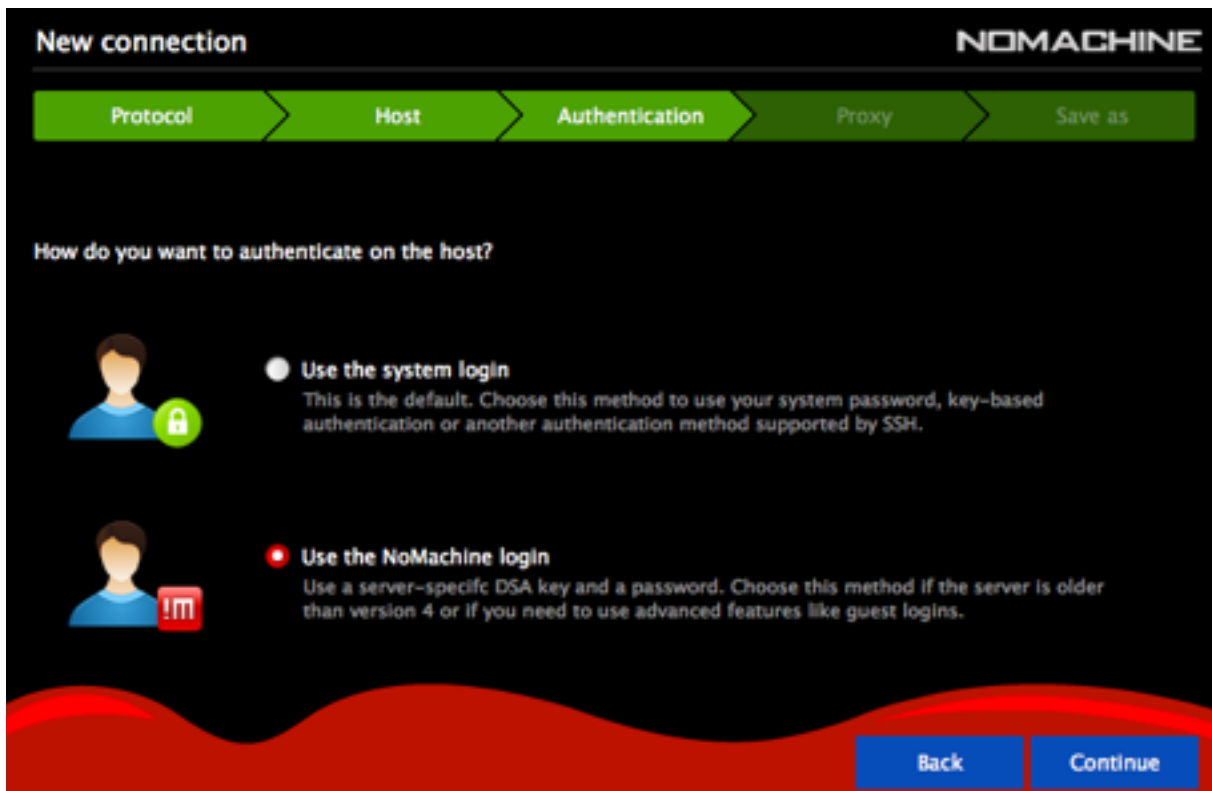


The second step will be to provide a host address. This will be the **DNS** entry you copied above. Past it into the box labelled **HOST**. The port should be set to **22**.



The screenshot shows the 'New connection' dialog in the NoMachine application. At the top, there's a progress bar with five steps: Protocol, Host, Authentication, Proxy, and Save as. The 'Host' step is currently active. Below the progress bar, a text prompt says 'Insert the hostname or IP and port where you want to connect.' To the left of the input fields is a small globe icon. There are two input fields: 'Host' containing 'ec2-54-86-192-180.compute-1.amazonaws.com' and 'Port' containing '22'. Below these fields, a note states: 'The port was chosen automatically based on the default for the protocol. If the remote computer was configured to listen on a different port, please insert it above.'

The third step asks how you would like to log-in. Use the default which is **Use the NoMachine login**

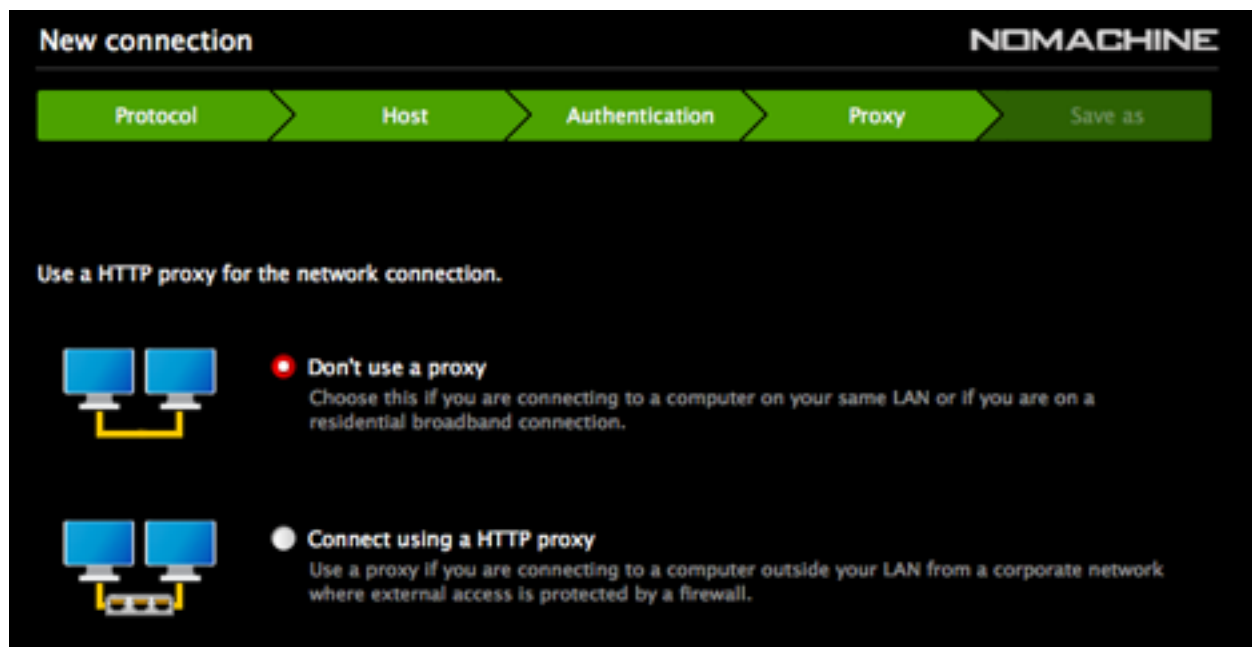


The screenshot shows the 'New connection' dialog in the NoMachine application, now at the 'Authentication' step. The progress bar at the top shows 'Authentication' as the active step. The question 'How do you want to authenticate on the host?' is displayed. There are two options, each with a user icon: 1. 'Use the system login' (selected with a white radio button), described as the default method using system password, key-based authentication, or other SSH methods. 2. 'Use the NoMachine login' (unselected with a red radio button), described as using a server-specific DSA key and password, recommended for older servers or advanced features. At the bottom right, there are 'Back' and 'Continue' buttons.

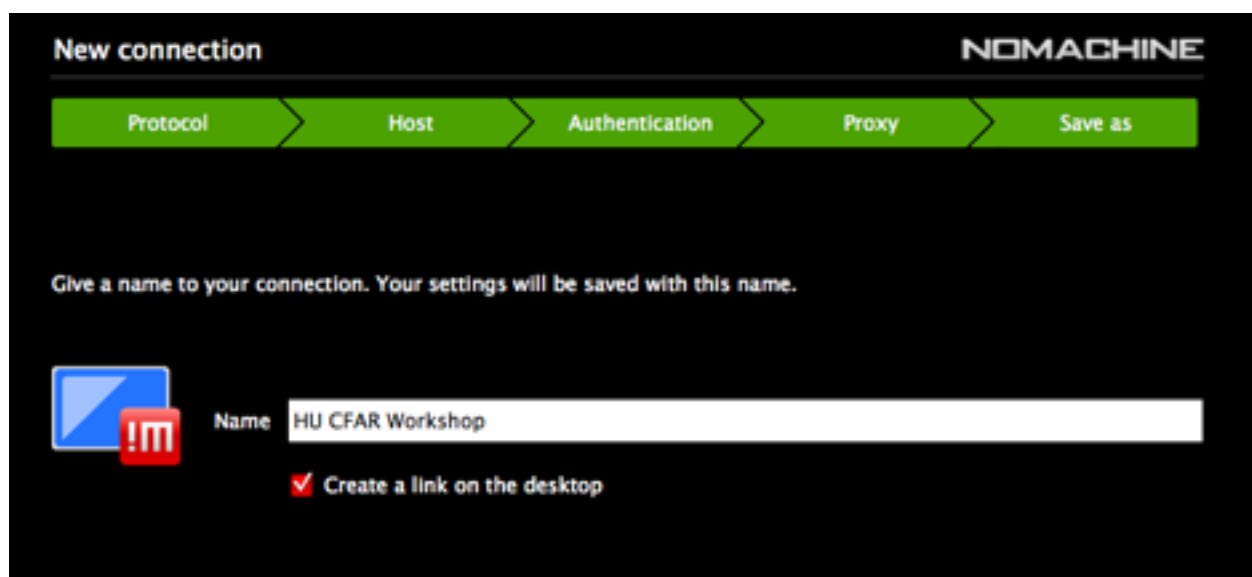
You will be given an option to **Use an alternate server key**. Just skip this and select **Continue**.

The fourth step we will tell the software how we want to authenticate the connection. For the Workshop we will be using **Password** authentication which should be your default.

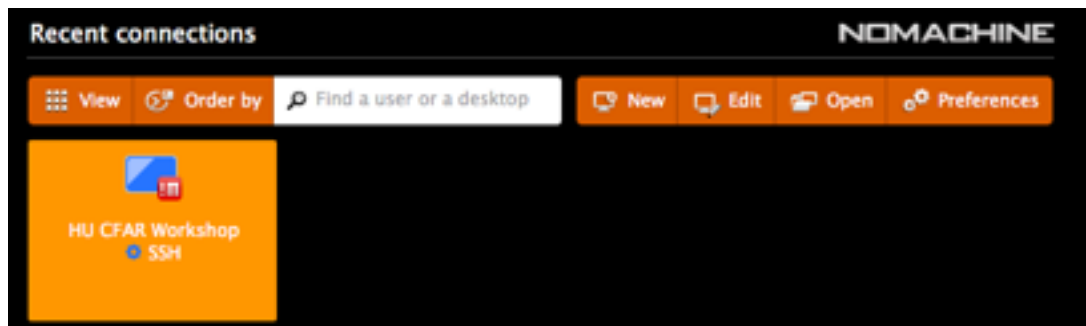
The fifth step asks if you would like to use a proxy. We don't want to, so make sure that **Don't use a proxy** is selected and continue.



The sixth step is to name the connection. Name the connection **HU CFAR Workshop** and check the box to **Create a link on the desktop**. It helps the teaching assistants if everyone has this set-up the same way, so please use these settings for consistency. But normally you can name it whatever you would like and choose to not have it create the desktop shortcut.

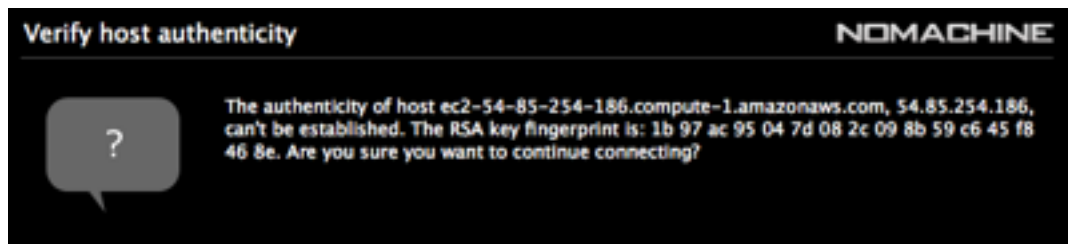


Your Recent Connections should now look like this:



Double-click the HU-CFAR Connection.

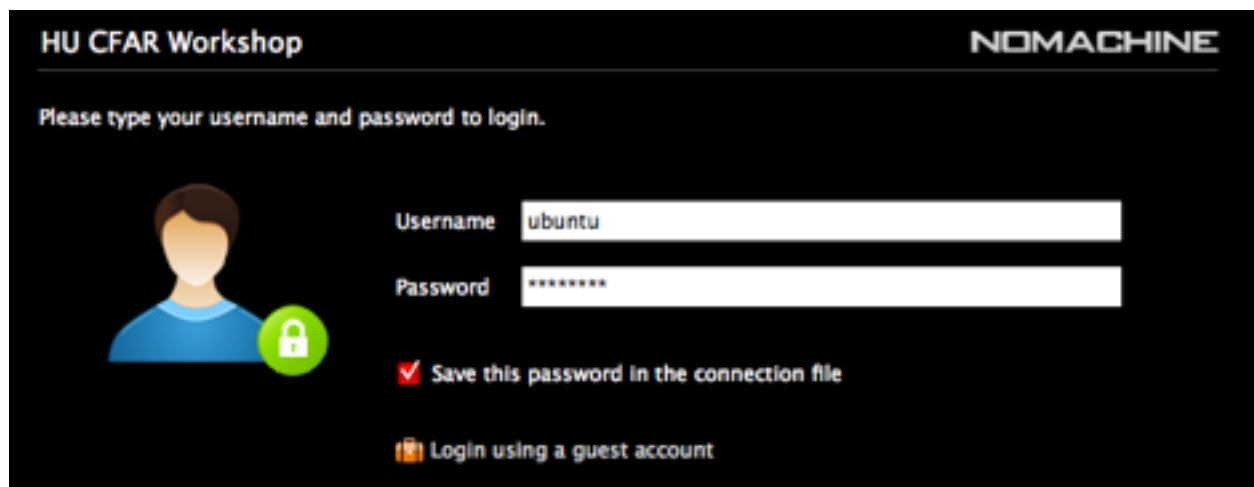
The first time you log-in you will be asked to **Verify host authenticity**. Just click **Yes** to continue.



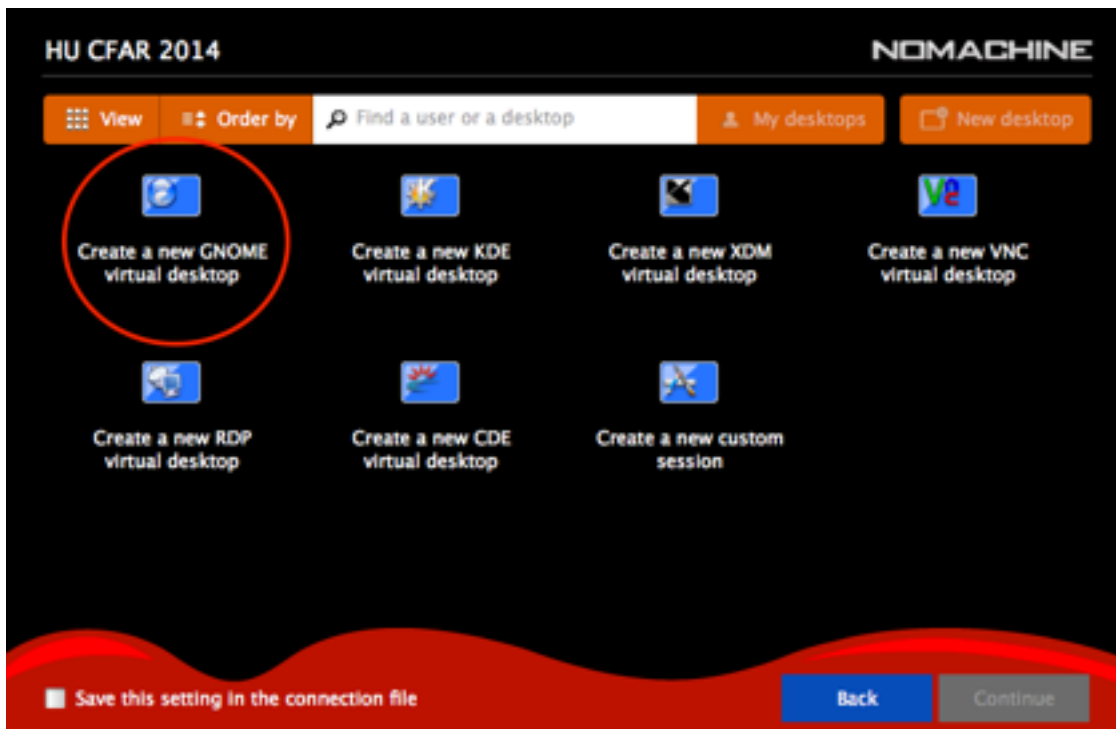
Next, you will be prompted to enter your username and password. Enter the following credentials:

Username: ubuntu

Password: cfar2014



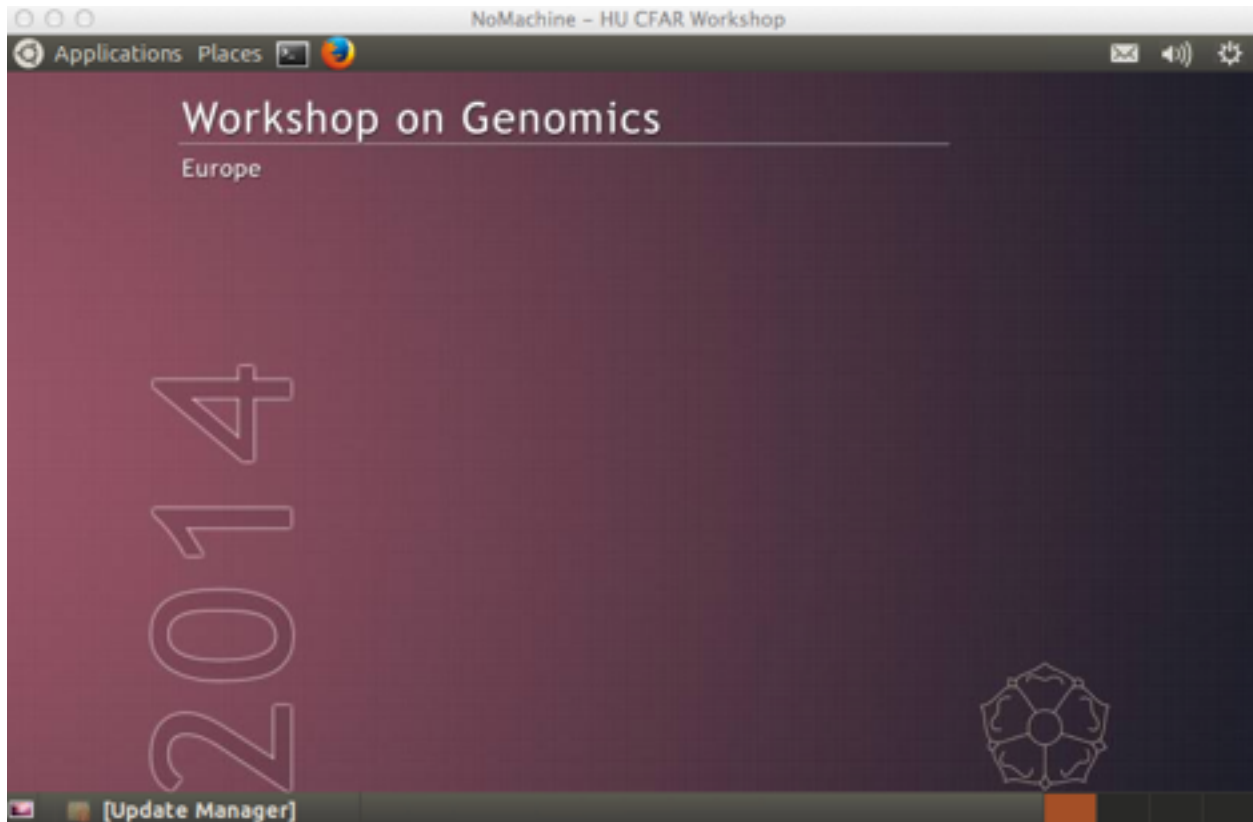
You will be asked to choose the style of desktop you would like to use. Select **Create a new GNOME virtual desktop**.



You will be presented with a series of informational screens which you should read though to get familiar with how NoMachine works. Read these and continue to hit **OK**.



At this stage you should see a screen like the one shown below. You will be asked to complete software updates. This is not necessary, so just close all of the windows that pop-up.



Note: The desktop background is what was used for the Workshop on Genomics held in Cesky Krumlov, Czech Republic in January 2014. Just ignore it! Or read more about the European Workshops at <http://evomics.org>.