

WA2592 Applied Data Science and Big Data Analytics

Classroom Setup Guide

Web Age Solutions Inc.

Table of Contents

Part 1 - Class Setup.....	3
Part 2 - Minimum Software Requirements for the Client Component.....	3
Part 3 - Software Provided.....	3
Part 4 - Instructions.....	3
Part 5 - Organize Windows Explorer Folder Views	5
Part 6 - Installing R Programming Language 3.3.1 on Windows	8
Part 7 - Verify Installation.....	12
Part 8 - Installing RStudio Desktop v0.99 on Windows.....	13
Part 9 - Verify Installation.....	13
Part 10 - Nginx.....	15
Part 11 - Minimum Hardware Requirements for the Lab Server	16
Part 12 - Minimum Software Requirements	16
Part 13 - Software Provided.....	16
Part 14 - Preparation.....	16
Part 15 - Installing the VMWare image.....	17
Part 16 - Running the VM	20
Part 17 - Setting Up the IP Address of the Lab Server VM.....	21

Part 1 - Class Setup

This class requires two components to be installed:

1. **The Client**
2. **The Lab Server**

The Client and the Lab Server must be installed on different machines; the Lab Server must be accessible by the Client.

Both machines must have access to the Internet.

The minimum software requirements for the Client and the Lab Server machines are different and are listed in different sections of this document.

Also, depending on the class software packaging option, you may have either one or two ZIP files.

Part 2 - Minimum Software Requirements for the Client Component

- Windows OS: Windows Vista / 7.
- Latest Google Chrome browser

Part 3 - Software Provided

You will receive the following file:

- **WA2592.ZIP**

All other software listed under Minimum Software Requirements is either commercially licensed software that you must provide or software that is freely available off the Internet.

Part 4 - Instructions

__1. Make sure the account that you are using to install the software has administrative privileges and the student using this machine will have the same rights.

__2. Extract the **WA2592.ZIP** file to C:\

__3. Review that the following folders were created:

- **C:\LabFiles**
- **C:\Software**
- **C:\Software\nginx-1.4.3**

__4. Verify the following files were created:

- **C:\Software\RStudio\R\R-3.3.1-win.exe**
- **C:\Software\RStudio\R\RStudio-0.99.903.exe**

__5. Download and install the latest Google Chrome browser from:

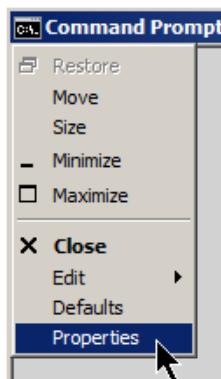
<https://www.google.com/intl/en/chrome/browser>

__6. Create a shortcut to the Windows Command Prompt onto the desktop.



__7. Double click the Command Prompt shortcut to open the Command Prompt window.

__8. In the Command Prompt window, click the black icon in the top left-hand corner and select **Properties** from the context menu.



The Properties dialog opens.

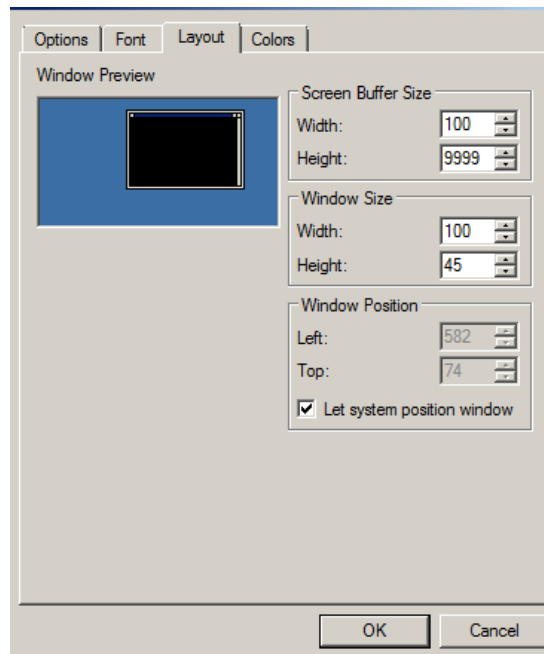
__9. In the Properties dialog, check the **Quick Edit Mode** check box.

Note: This option allows a user to copy and paste text in the command prompt using mouse actions

instead of an edit menu.

__10. Click the **Layout** tab.

__11. In the *Layout* tab, enter **100** for *Width* property (for both *Width* text windows in the *Layout* tab window), **9999** for the *Height* of the *Screen Buffer Size* property, and **45** for the *Height* property of the *Window Size* property.



__12. Click **OK** to close the Properties dialog.

__13. If an *Apply Properties to Shortcut* dialog appears, select **Modify shortcut that started this window** and click **OK**.



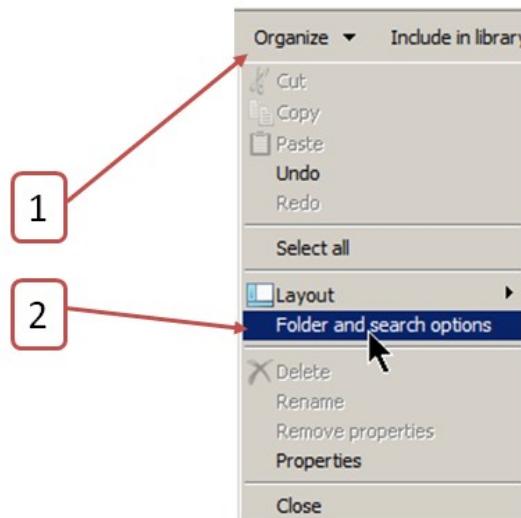
Part 5 - Organize Windows Explorer Folder Views

__1. Start Windows Explorer

The steps below may slightly vary depending on the Windows version (the steps below are shown for

Windows 7). The main purpose of these steps is to enable the system-wide display of user files extensions and hidden files in Windows Explorer.

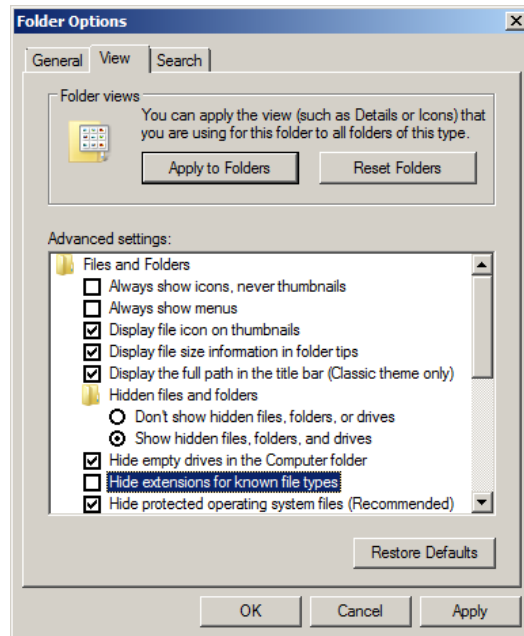
___2. In the menu bar of Windows Explorer, click the **Organize** drop-down menu and select **Folder and search options**



___3. In the *Folder Options* dialog that opens, click the **View** tab

___4. In the **View** tab:

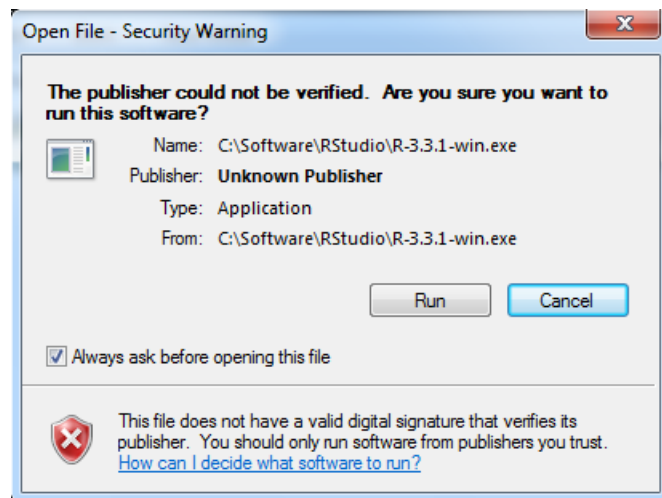
- Check the **Display the full path in the title bar ...** check box
- Select the **Show hidden files, folders, and drives** radio button
- Uncheck the **Hide extensions for known file types** check box



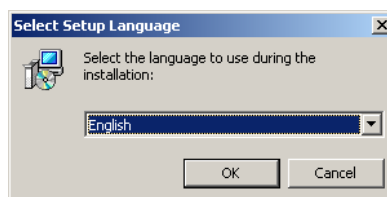
__5. Click the **OK** button

Part 6 - Installing R Programming Language 3.3.1 on Windows

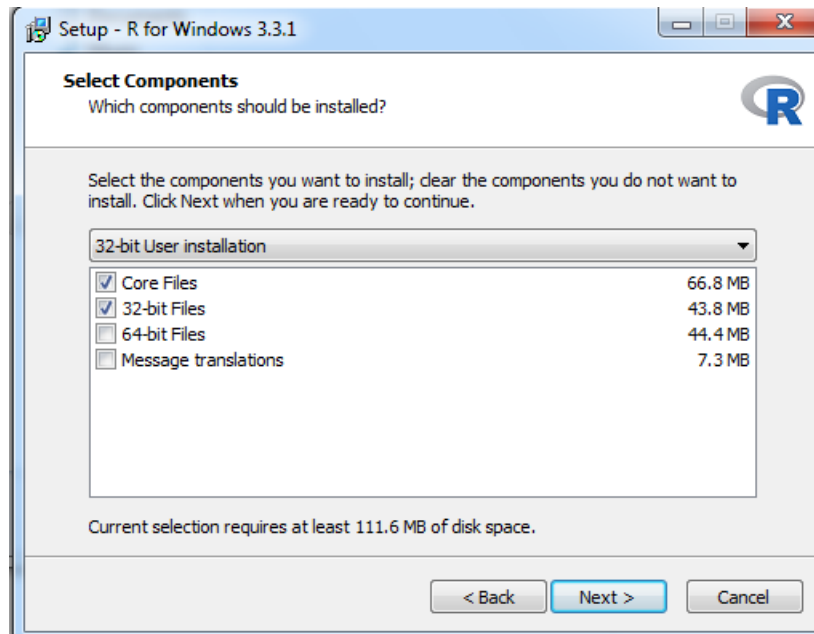
- __1. (Skip this step, if not applicable) If you are not yet logged in on the Student computer, log in as the user who will be using this software in the class.
- __2. From the **C:\Software\RStudio\R** directory, run **R-3.3.1-win.exe**
- __3. If prompted with the *Windows Security Warning*, click **Run**.



- __4. If prompted with the Windows system *User Account Control* dialog, click **Yes**.
- __5. Accept **English** as the Setup Language and click **OK**.



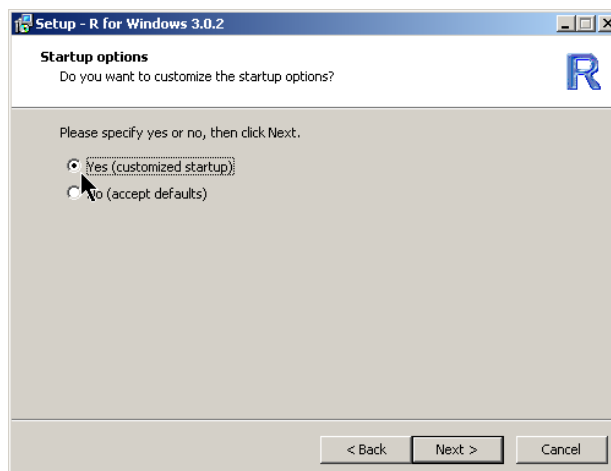
- __6. In the *Welcome* screen that opens, click **Next >**
- __7. In the License Dialog, click **Next >**
- __8. In the *Destination Location* dialog, enter **c:\Software\R** for the folder location and click **Next >**
- __9. In the *Select Components* dialog, select the **32-bit User Installation** option from the drop-down box.



__10. Accept default preselected options for *Core Files* and *32-bit Files* checkboxes.

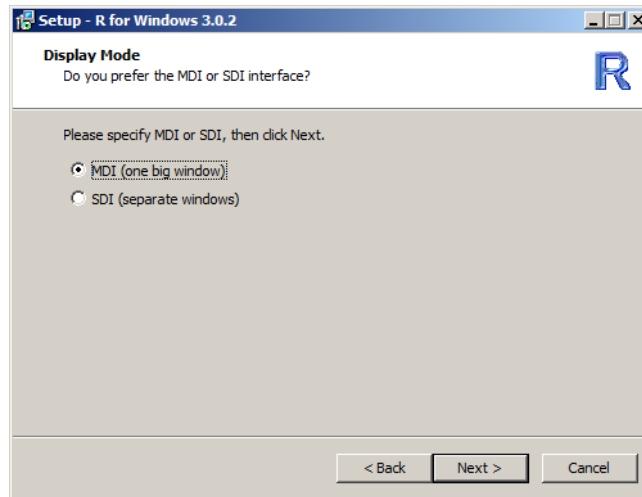
__11. Click **Next >**

__12. In the *Startup options* dialog, select **Yes** (customized startup).



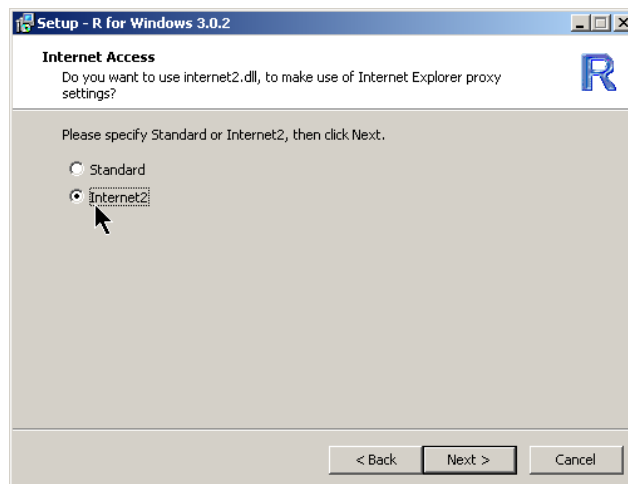
__13. Click **Next >**

__14. In the *Display Mode* dialog, accept the default *MDI* option and click **Next >**



__15. In the *Help Style* dialog, accept the default *HTML help* option and click **Next >**

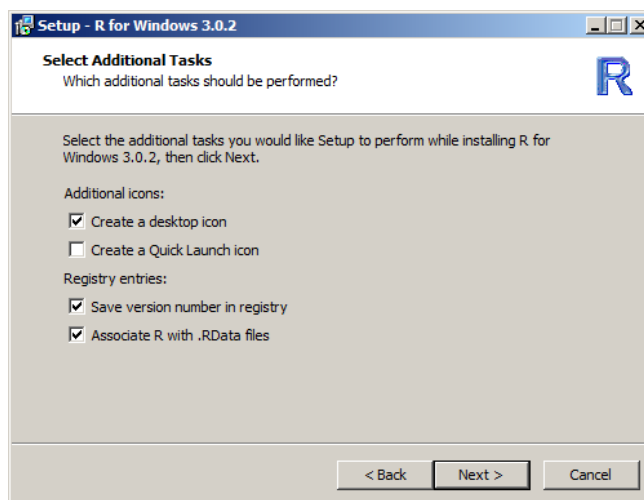
__16. If prompted, in the *Internet Access* dialog, select **Internet2** option.



__17. Click **Next >**

__18. In the *Select Start Menu Folder* dialog, accept the default **R** name and click **Next >**

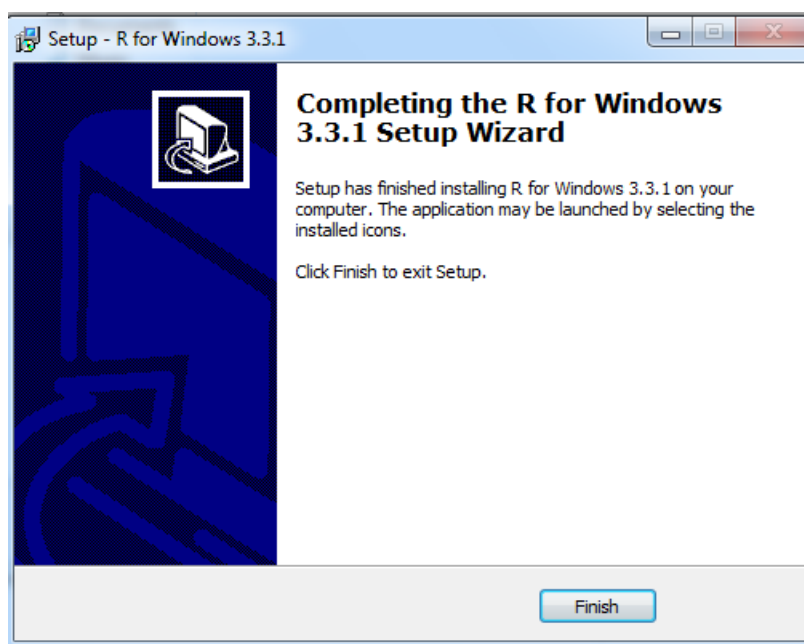
__19. In the *Select Additional Tasks* dialog, accept defaults and click **Next >**



Installation process begins.

Wait for the installation process to complete.

When installation is complete, you will be presented with the confirmation dialog.



__20. Click **Finish**.

Part 7 - Verify Installation

__1. Find the R short-cut created on the desktop and double click it.

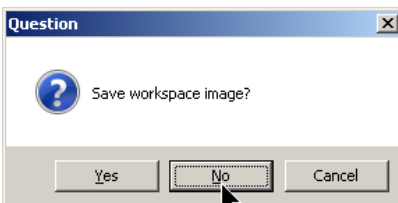


The *R GUI* console should open.

__2. Type **q()** in console and press **Enter**.

```
> q()
```

__3. In the *Question* dialog that opens, click **No**.



Installation and verification steps are complete.

Part 8 - Installing RStudio Desktop v0.99 on Windows

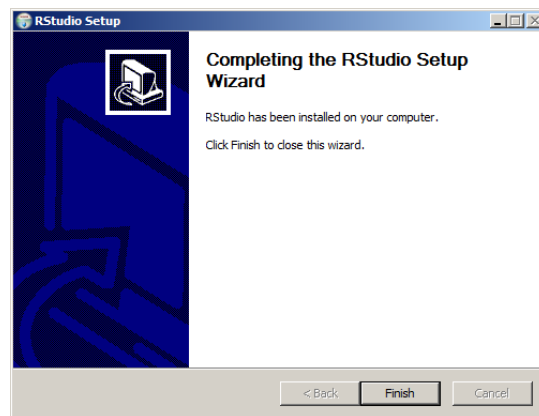
Note: The prerequisite for installing this package is the presence of the R package version 2.11.1 (or higher) on the target system (as per *R-3.3.1-win-32bit.odt* document).

- __ 1. On the Student computer, log in as the user who will be using this software in the class
- __ 2. From the **C:\Software\RStudio\R** directory, run **RStudio-0.99.903.exe**
- __ 3. If prompted with the Windows system *User Account Control* dialog, click **Yes**
- __ 4. On the *RStudio Setup Welcome* Screen, click **Next >**
- __ 5. In the *Choose Install Location* dialog, enter **c:\Software\R** for the destination folder and click **Next >**
- __ 6. In the *Choose Start Menu Folder* dialog, accept defaults and click **Install**

Installation process begins.

Wait for the installation process to complete.

When installation is complete, you will be presented with the confirmation dialog.



- __ 7. Click **Finish**

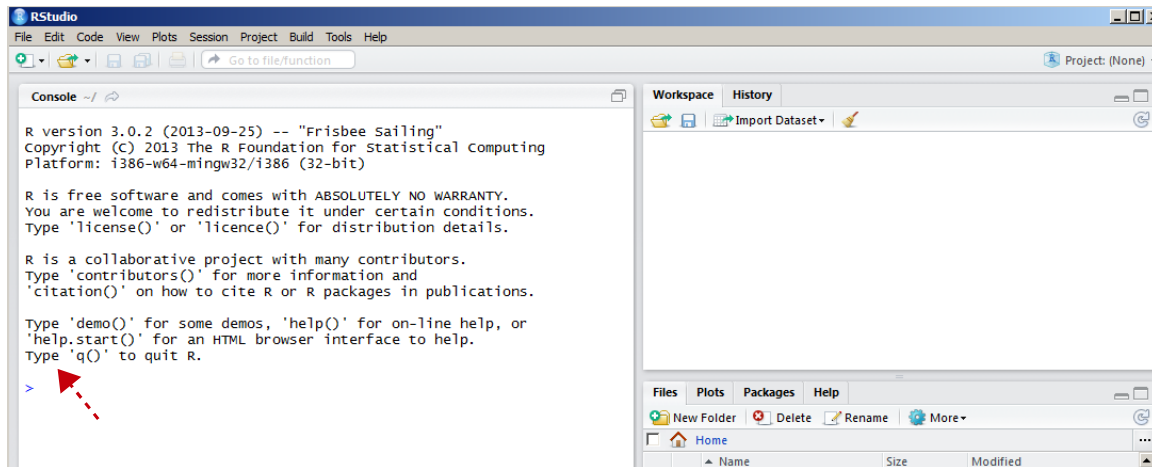
Part 9 - Verify Installation

- __ 1. Create a Desktop shortcut pointing to the **C:\Software\R\bin\rstudio.exe** folder



2. Click the **RStudio** Desktop short-cut

The RStudio IDE opens.



3. In the *Console* window on the left hand side, type in **q()** and press **Enter**

RStudio closes.

Installation and verification steps are complete.

Part 10 - Nginx

__1. Disable any service using Port 80 to be able to run Nginx, if you have IIS or other service running, stop and disable them.

__2. Switch to the user that the students will use during the course.

__3. Open the Command Prompt window and type in the following command at the prompt and press **ENTER** (execute the command):

```
cd C:\Software\nginx-1.4.3
```

This command will change directory to where the *nginx* web server resides (represented by the *nginx.exe* file).

__4. Start the nginx web server by executing the following command:

```
start nginx
```

This command will launch the *nginx* web server that starts listening on port 80. Allow access if the Firewall window appear.

__5. If you are prompted for the admin password, enter it to allow the software to run.

__6. Open Google Chrome browser and navigate to **http://localhost**

You should see the *nginx* welcome page.

Welcome to nginx!

If you see this page, the nginx web server is successfully installed and working. Further configuration is required.

For online documentation and support please refer to nginx.org.
Commercial support is available at nginx.com.

Thank you for using nginx.

__7. Close Chrome browser.

__8. In the Command Prompt window where you started the *nginx* web server, type in the following command at the prompt and press **ENTER**:

```
nginx -s stop
```

This command will stop the web server.

__9. Switch back to the admin user.

Nginx is installed.

Part 11 - Minimum Hardware Requirements for the Lab Server

The Lab server is a 64-bit VM that requires a 64-bit host OS and a virtualization product that can support a 64-bit guest OS. **This VM uses 6 GB of total RAM and 2 vCPU.** The total system memory required varies depending on the size of data sets used in labs and on the other processes that are running in the VM.

- 8 GB RAM
- 80 GB Hard Disk

Part 12 - Minimum Software Requirements

- Windows XP / Vista / 7 - 64 bit
- VMware player 6.x or higher
- Chrome

Part 13 - Software Provided

You will receive the following file (further referred to as the **VM ZIP** file) containing the VMware player-compatible virtual machine:

- **VM_WA2341_CDH5-REL_2_2-Sep-2016.zip**

Part 14 - Preparation

__1. Extract the **VM ZIP file** to C:\

Note: Every student in the class will need a dedicated Lab Server. So the class setup will require as many Lab Servers as there are students in the class. In other words, you will need to perform this setup as many times.

It is recommended to have each Lab Server VM installed on a separate physical machine, although they can be collocated as long as their *Network Connectivity* is setup with the *Bridged* option (see

details further in the document).

Part 15 - Installing the VMWare image

__1. Open a file browser and navigate inside the unzipped **VM ZIP** folder. Locate the VMware player executable file **vmplayer.exe**.

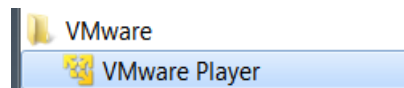
Note. If you don't find the VMware player executable file in this folder, download the VMware player **6.x** or higher from the VMware website using the following link:

<http://www.vmware.com>

__2. Install the VMware player accepting all the defaults during the installation.

__3. Restart the computer.

__1. From the Start menu, select **All Programs > VMware > VMware Player**.



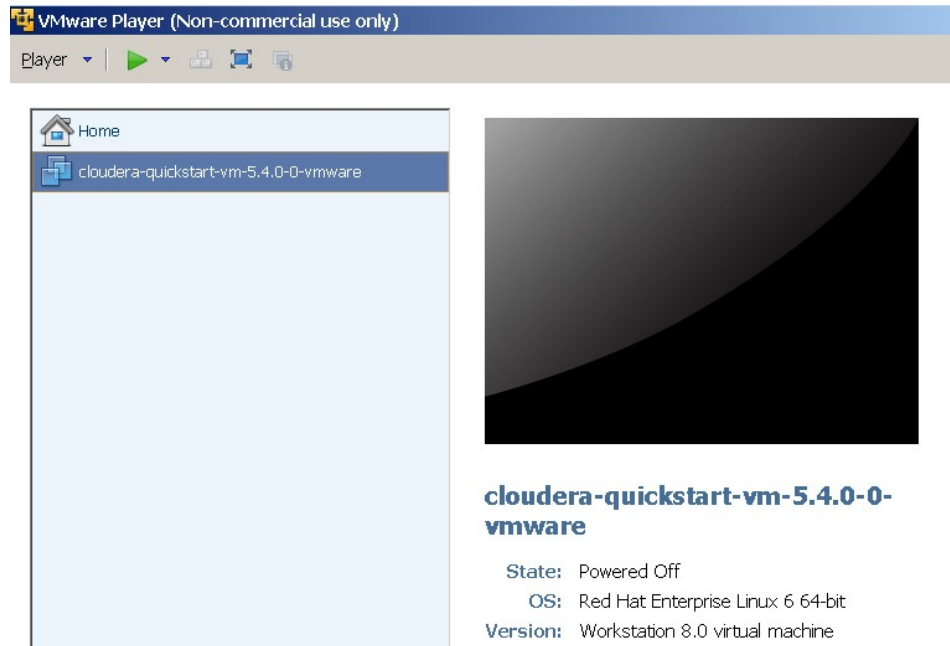
__2. If prompted to download a new version of VMware player decline the update.

__3. Press **Ctrl-O**.

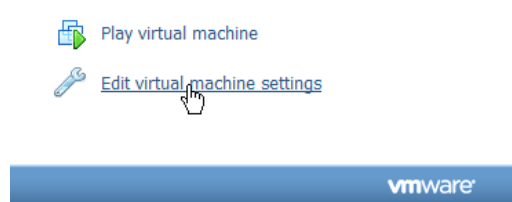
The *Open Virtual Machine* dialog opens

__4. Locate and select the **cloudera-quickstart-vm-5.4.0-0-vmware.vmx** file located under the unzipped **VM ZIP** folder and click **Open**.

The **cloudera-quickstart-vm-5.4.0-0-vmware** menu option will appear on the list of available virtual machines.



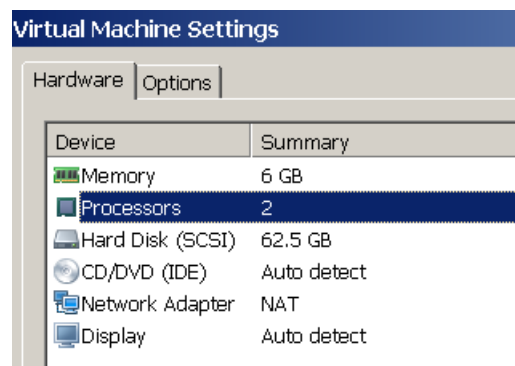
__5. Click **Edit virtual machine settings** at the bottom of the VMWare Player



The *Virtual Machine Settings* screen opens with the **Hardware** tab opened by default.

__6. Change the Memory VM size attribute to **6 G (6144 M)**

__7. Change the Number of Processors VM attribute from 1 to **2**

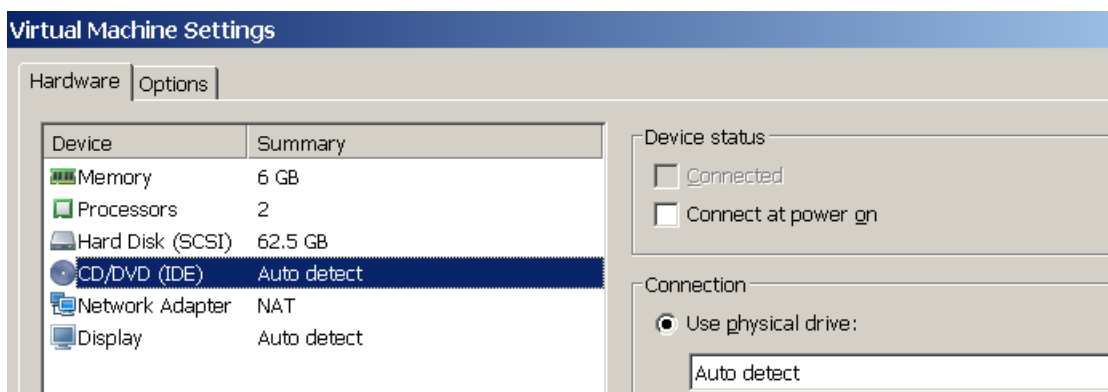


__8. The **Network Adapter** VM attribute can be configured for the **Bridged** or **NAT** connection options.

- ✓ As a rule of thumb, use **NAT** for the **VM** being installed locally on the physical student machine, use **Bridged** on remote machines. If these suggestions do not work, use the options that best suite your environment.
- ✓ The NAT options is preselected by default; for the Bridged option, see the **Setting Up the IP Address of the Lab Server VM** lab step at the end of the document

__9. Click **CD/DVD (IDE) Device**

__10. Uncheck the **Connect at power on** (or keep it clear if it is already unchecked) *Device status*



__11. If **Floppy** is present, click **Floppy** (if this option is **not** present, skip to the next numbered step)

- ✓ Uncheck the **Connect at power on** (or keep it clear if it is already unchecked)

__12. If **Sound Card** is present, click **Sound Card** (if this option is **not** present, skip to the next to the next numbered step)

- ✓ Uncheck the **Connect at power on** (or keep it clear if it is already unchecked)

__13. If **Printer** is present, click **Printer** (if this option is **not** present, skip to the next to the next numbered step)

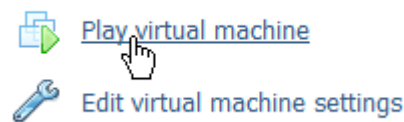
- ✓ Uncheck the **Connect at power on** (or keep it clear if it is already unchecked)

__14. Click **OK** at the bottom of the *Virtual Machine Settings Screen* to close it.

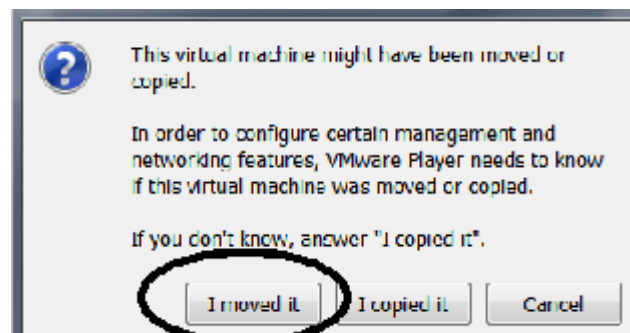


Part 16 - Running the VM

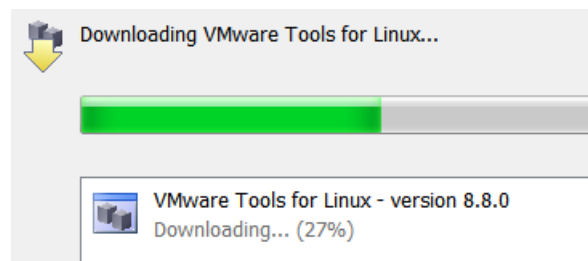
__1. Select the **cloudera-quickstart-vm-5.4.0-0-vmware** virtual machine (it should already be pre-selected) and click **Play virtual machine**.



__2. Click "**I moved it**", if prompted.



__3. If you are prompted to download and install the VMware Tools for Linux, accepted the option.



Accept reasonable options if and when they appear.

VM bootstrapping may take some time, and when it completes, you should be automatically logged in the Lab Server VM as the *cloudera* user and presented with the Cloudera Desktop.



The installation of the Lab Server virtual machine is completed.

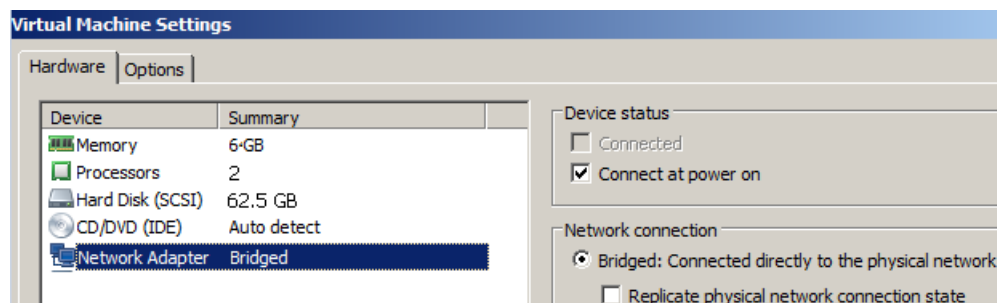
The last Lab setup step is required if you want to set up the student VMs with the Bridged network configuration option.

Note: The remote (SSH) access to the Lab Server VM is done under the *cloudera* username with *cloudera* password.

The *cloudera* account has sudo privileges in the Lab Server. The root account password is **cloudera**

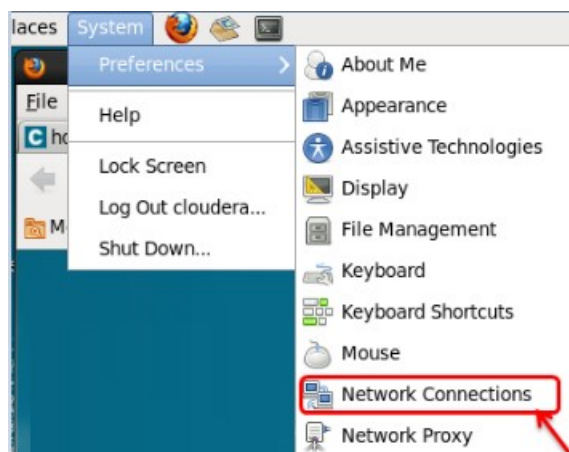
Part 17 - Setting Up the IP Address of the Lab Server VM

If you setup your VM Network Adapter with the *Bridged* option as shown in the screen-shoot below, by default, you will have a DHCP leased IP address assigned to the Lab Server. It may be a convenient feature from the administration perspective, but will affect student SSH connections during the class as they will always be required to change the Lab Server IP address whenever the IP address of the Lab Server changes (IP lease may be configured to expire every day and the class runs for four days).



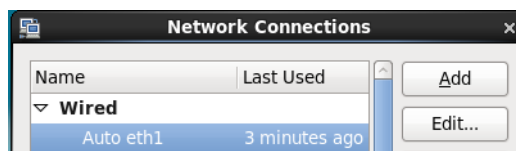
Considering the inconvenience to the students, it may be worthwhile to assign each Lab Server a unique IP address.

__1. From the Lab Server toolbar, select *System > Preferences > Network Connections*.



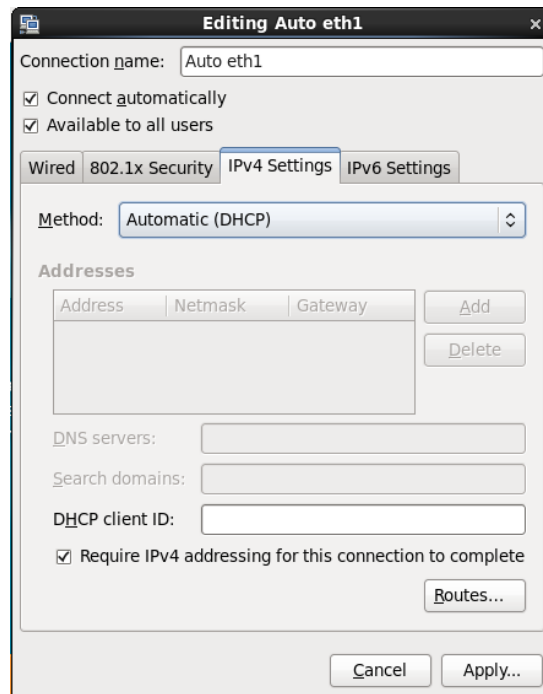
The *Network Connections* Dialog opens.

__2. Select **Wired / Auto eth1** and click **Edit ...**



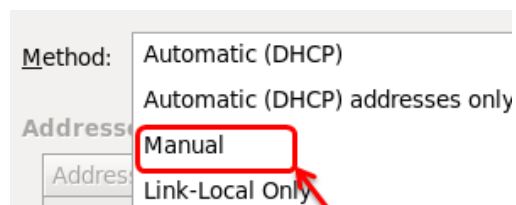
The **Editing Auto eth1** Dialog opens.

__3. Select the **IPv4 Settings** Tab.



In the screen-shoot above the network adapter is configured to receive IP address from the DHCP server.

__4. For setting up the static IP address, select **Manual** from the *Method:* drop-down.



__5. Click **Add** to add an *IP Address*, *Netmask*, *Gateway* and *DNS* as per your network settings.

Sample input screen is shown below.

Editing Auto eth1

Connection name: Auto eth1

☒ Connect automatically

☒ Available to all users

Wired 802.1x Security IPv4 Settings IPv6 Settings

Method: Manual

Addresses

Address	Netmask	Gateway
192.168.0.101	24	192.168.0.1

Add Delete

DNS servers: 192.168.0.1

Search domains:

DHCP client ID:

☒ Require IPv4 addressing for this connection to complete

Routes...

Cancel Apply...

__6. Click **Apply ...**

The *Authenticate* Dialog opens up.

__7. In the *Password for root:* text window, enter **cloudera** and click **Authenticate**.

You should be returned to the *Network Connections* Dialog.

__8. Click **Close**

This is the final step of the Lab Server setup.

You have successfully installed the software for this course.