

CiB installation manual

REVISION HISTORY

NUMBER	DATE	DESCRIPTION	NAME
0.3	2 Nov 2012	Edit the last 4 chapters (nada server, database server, data manager desktop, sysadmin desktop)	CK
0.2	19 Oct 2012	Draft after merging two previous documents	CK

Contents

1	Introduction	1
2	Preparing for the installation	3
2.1	Check the hardware pre-requisites	3
2.2	Check the software pre-requisites	3
2.3	Check the network setting pre-requisites	4
3	Hypervisor host installation	5
3.1	Check that the mac mini is working fine with the pre-installed OSX	5
3.2	Start the installation	5
3.3	Enable ESXi shell	10
3.4	Update the Broadcom NIC driver	12
3.5	Configure and test the network settings	13
3.6	Install the vSphere Client	14
4	Virtual machine installation	15
4.1	Prepare for the virtual machine installation	15
4.2	Create a virtual machine	15
4.3	Install Windows 7 as a guest operating system	21
4.3.1	Post-installations	22
4.3.2	Miscellaneous	22
4.4	Install Linux Ubuntu as a guest operating system	22
4.4.1	Setup static IP address	30
4.5	Install VMware Tools	31
5	Archive Server installation	35
5.1	Role	35
5.2	Setup the Virtual machine	35
5.3	Samba installation	35
5.3.1	Install the Samba package	35
5.3.2	Allow only authenticated users	36

5.3.3	Create the shared directory for all authenticated users	36
5.3.4	Create a private directory for home	36
5.3.5	Add and manage users	36
5.3.6	Access the shares from Windows client	37
5.3.7	Access the shares from linux clients	37
5.4	Git installation	37
5.4.1	Installation	37
5.4.2	Configuration	37
6	Nada server installation	38
6.1	Role	38
6.2	Setup the virtual machine	38
6.3	Install the Apache web server	38
6.4	Install PHP and required extensions	39
6.5	Install MySQL database server	40
6.6	Install the NADA application files	41
6.7	Configure the NADA database connection settings	41
6.8	Run the web installer	41
6.9	Launch Nada	43
7	Database server installation	45
7.1	Role	45
7.2	Install MS SQL Express 2008 on Windows 7	45
7.2.1	Setup the VM	45
7.2.2	Install MS SQL 2008 R2 Express edition with tools	45
7.2.3	Setup remote access	54
7.2.4	Install the analytical database	54
7.3	MySql on Ubuntu 12.08	55
7.3.1	Setup the virtual machine	56
7.3.2	Install mysql	56
7.3.3	Configure	56
7.3.4	Resources	56
8	Data Manager desktop installation	57
8.1	Role	57
8.2	Setup the virtual machine	57
8.3	Pentaho Kettle	57
8.3.1	Install Java Runtime Environment	57
8.3.2	Install Pentaho Kettle Community edition	59
8.3.3	Install a native JDBC driver for MS SQL server	61

8.3.3.1	Setup a connection to the analytical database	62
8.3.4	Build your first transformation	63
8.3.4.1	Retrieve data from a flat file	64
8.3.4.2	Save your transformation	67
8.3.4.3	Filter records with missing postal codes	68
8.3.4.4	Loading your data into a relational database	70
8.3.4.5	Retrieve data from the Lookup file	72
8.3.4.6	Resolving missing Zip Code information	72
8.3.4.7	Completing your Transformation	73
8.3.4.8	Run Your Transformation	74
8.3.5	Build your first job	75
8.3.6	Resources	77
8.4	Nesstar Publisher	77
8.4.1	Install Nesstar Publisher	78
8.5	Zotero	81
8.5.1	Install Zotero for Firefox	81
8.5.2	Install Zotero Standalone	83
8.5.3	Resources	83
9	SysAdmin desktop installation	84
9.1	Role	84
9.2	Setup the virtual machine	84
9.3	Install vSphere Client	84
9.3.1	GUI tools	84
9.3.2	CLI tools	85
9.4	Install WireShark	85
9.4.1	Resources	91
9.5	Install guettoVCB backup script	91
9.5.1	Setup	92
9.5.2	Configuration	92
9.5.3	Usage	92
9.5.4	Setup a cron job	93
9.5.5	Resources	93
A	Annexures	94
A.1	Site code	94
A.2	Naming convention	94
A.3	Example Glossary	95

Chapter 1

Introduction

This manual describes how to install and configure a *Centre-in-a-box* or *CiB* on a Mac Mini server.

At a high level of abstraction, the CiB solution consists of several application servers running on Windows or Linux virtual machines. The virtual machines are hosted by a VMware ESXi hypervisor. The hypervisor sits on top of the physical hardware as shown in the figure below.

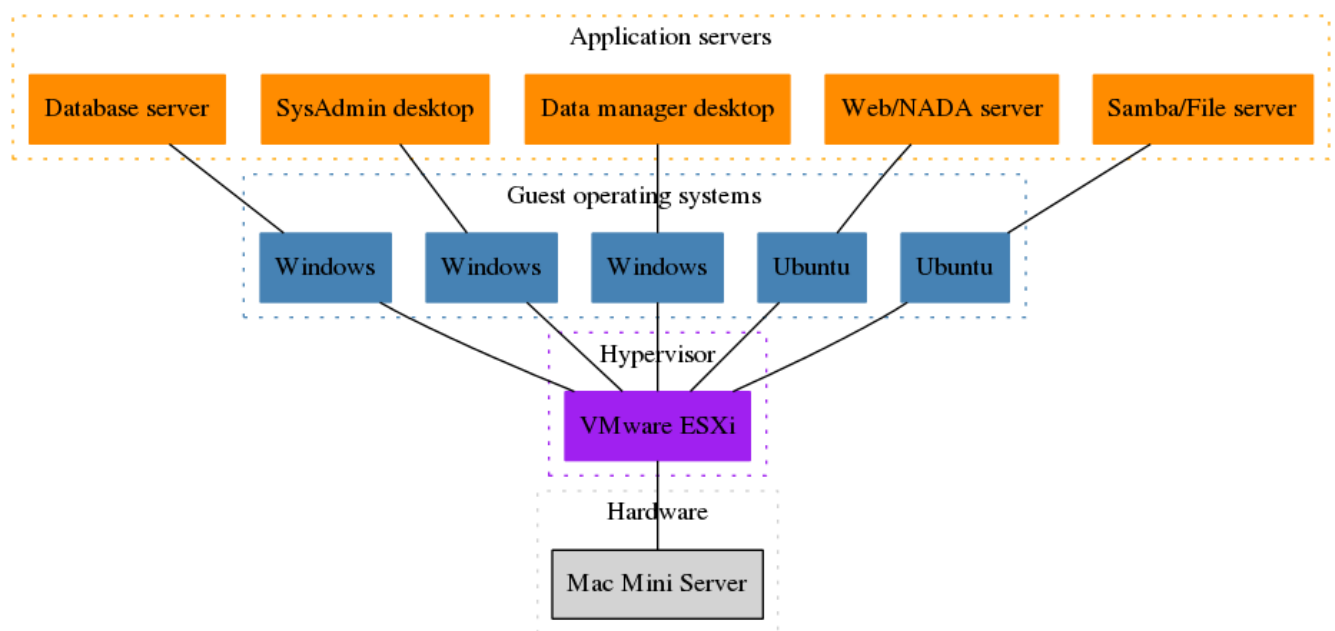


Figure 1.1: CiB components

Our objective is to setup those application servers (Samba file, ETL, NADA, SystemAdmin,) by preparing the Mac Mini Server hardware, then installing the VMWare ESXi hypervisor followed by the setup of the guest operating systems (Windows 7 or Linux Ubuntu).

The manual includes the following topics:

1. Preparing for the installation
2. Installing the hypervisor host
3. Install the virtual machines as guests.

4. Configuring the Archive server
5. Configuring the NADA server
6. Configuring the Database server
7. Configuring the Data Manager desktop
8. Configuring the System admin desktop

The intended audience is anyone who needs to install and setup a CiB. The information is written for Windows or Linux system administrators who have a basic knowledge of virtual machine technology, operating systems and database administration.

Chapter 2

Preparing for the installation

Before starting the installation, ensure that you have the correct hardware and software pre-requisites.

2.1 Check the hardware pre-requisites

Ensure that you have the following equipment:

- Apple Mac Mini Server 2011, 2.0GHz quad-core, Intel Core i7, Hard drive 500GB/700GB
- 2 x 8G RAM
- Apple Super Drive
- Standard Windows USB keyboard
- Standard USB 2-button mouse
- HDMI to DVI cable
- Monitor with DVI interface
- Ethernet Network cable
- Laptop or workstation with Windows XP/7 Operating systems

2.2 Check the software pre-requisites

You should have access to a local or network folder with the following softwares:

- For the hypervisor installation:
 - VMware ESXi 5.0.0-469512 ISO burned on CD-ROM
 - Broadcom NIC driver DD for Broadcom NetXtreme | Gigabit Ethernet Driver version 3.120h.v50.2 build 547149
 - VMware vSphere client version 5.0
 - Account to register the VMware host
 - For the guest operating system installation:
 - Windows 7 Professional 64-bit edition ISO + valid licenses
 - Ubuntu 2012-04 LTS 4-bit edition ISO
-

- For the archive server installation:
 - Samba
 - Git
 - ftp
- For the NADA server installation:
 - NADA
 - MySql version 5.5
 - PHP
 - Apache
- For the database server installation:
 - Microsoft SQL Express 2008 R2 with tools
 - MySQL version 5.5
- For the Data Manager desktop installation:
 - Pentaho Kettle Community edition version 4.3.0-stable
 - Nesstar Publisher
 - Firefox
- For the system administration desktop installation:
 - vSphere Client version 5.0
 - WireShark 64-bit version 1.80
 - GuettoVCB

2.3 Check the network setting pre-requisites

Your local network administrator should provide the following information:

- CiB IP address :
 - Subnet net address:
 - gateway IP address:
 - Primary name servers:
 - Secondary name servers:
 - IP address for each virtual machine
 - Http proxy if needed to access Internet
 - Domain administrator rights to join machines to the domain
-

Chapter 3

Hypervisor host installation

In this step, you are going to uninstall the built-in Apple Lion server operating systems and replace it by the VMware ESXi 5 operating system.

The instruction below are adapted from the expert advices found on [Paraguin consulting blog](#).

This is a first version with a combination of GUI mouse clicks and CLI keystrokes. The entire process should take about 1 hour on the first attempt but less than 20 minutes when one becomes more familiar with the process.

3.1 Check that the mac mini is working fine with the pre-installed OSX

If you receive a brand new Mac Mini server, boot it to check that the new box is working fine.

- Connect the keyboard, the mouse, the network cable and power cable.
- Press the power button.
- Follow the instructions.
- Check the network settings.
- Setup the proxy settings.
- Check your connection to internet.

3.2 Start the installation

- Plug the Super Drive into the Mac Mini
- Load the VMware ESXi installation CD into the SuperDrive
- Press <Alt> while powering on the Mac Mini
- Choose the EFI option

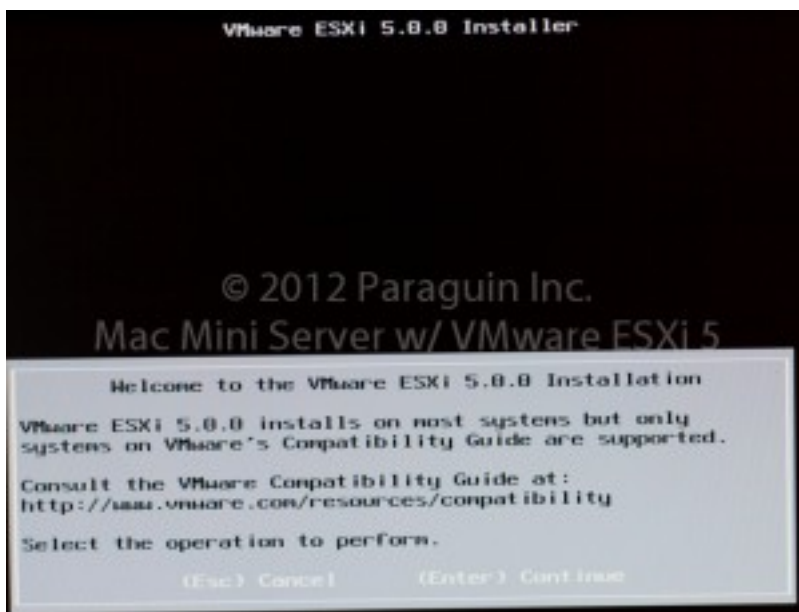
The system will load the CD and start the vmware setup process.



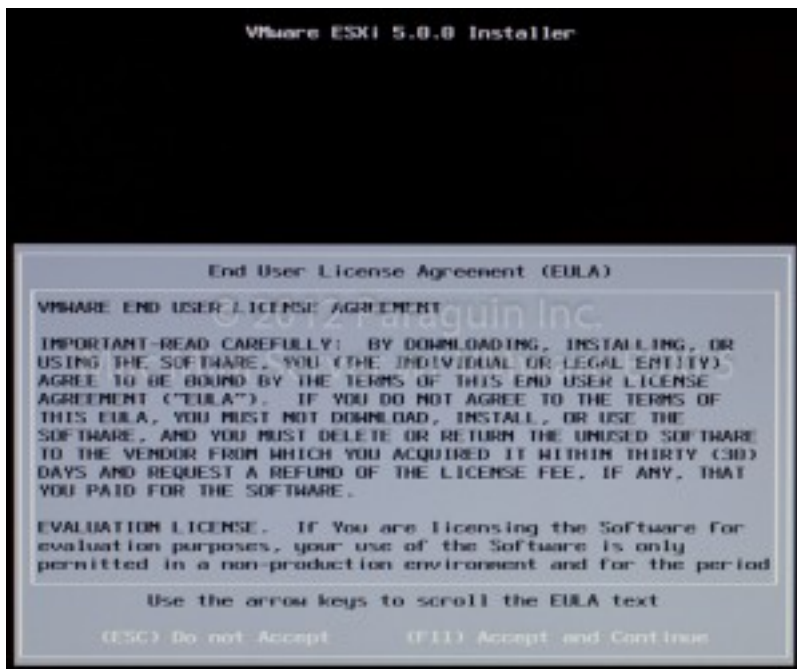
The VMware setup will start



After a moment, you should see the "Welcome to the VMWare ESXi 5 installation dialog.

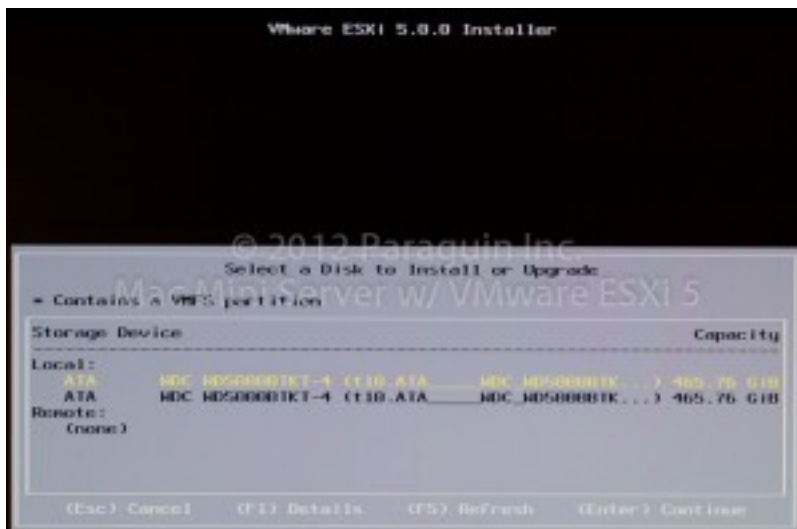


- Press <Enter> to continue

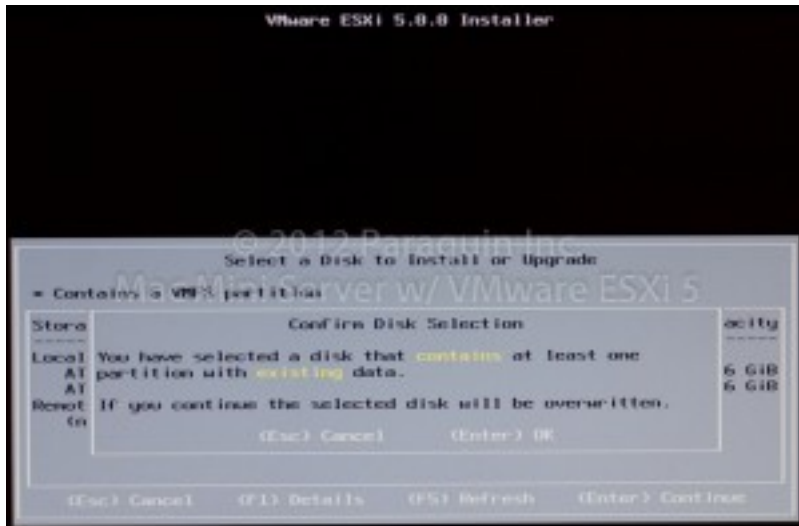


- Press <F11> to accept the EULA

The drive selection dialog will appear.



- Press <Enter> to Select the first drive in the array to install ESXi

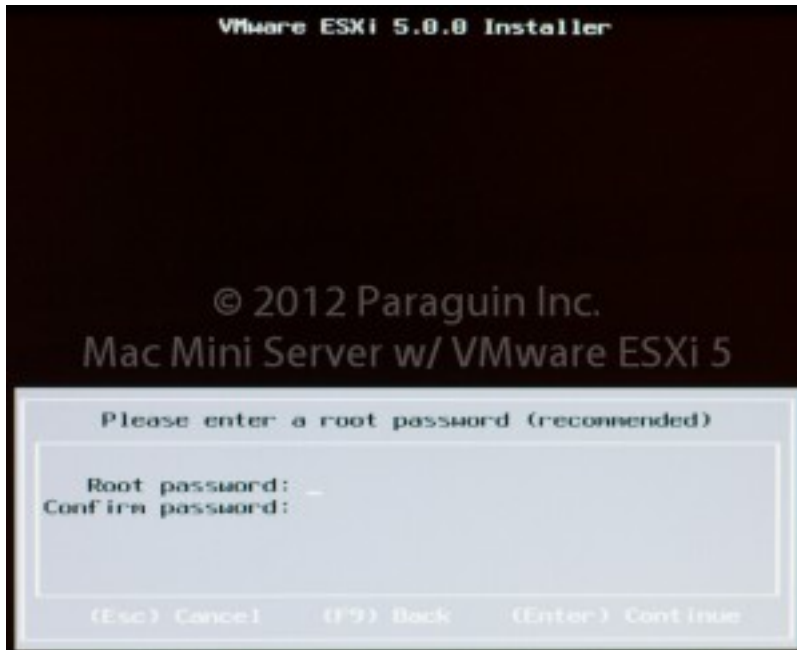


- Press <Enter> to confirm the disk selection



- Select the US default keyboard layout

You can now setup the password



- Setup the root password to *T3amw0rk*
- Confirm the root password

Confirm the installation as the disks will get repartitioned.

- Hit the <F11> key to start the installation

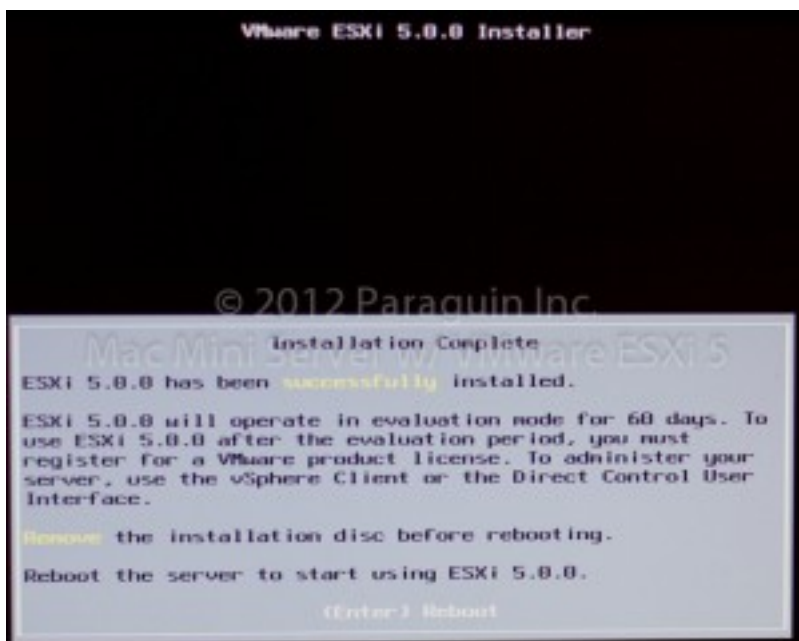


Wait while the installation is progressing



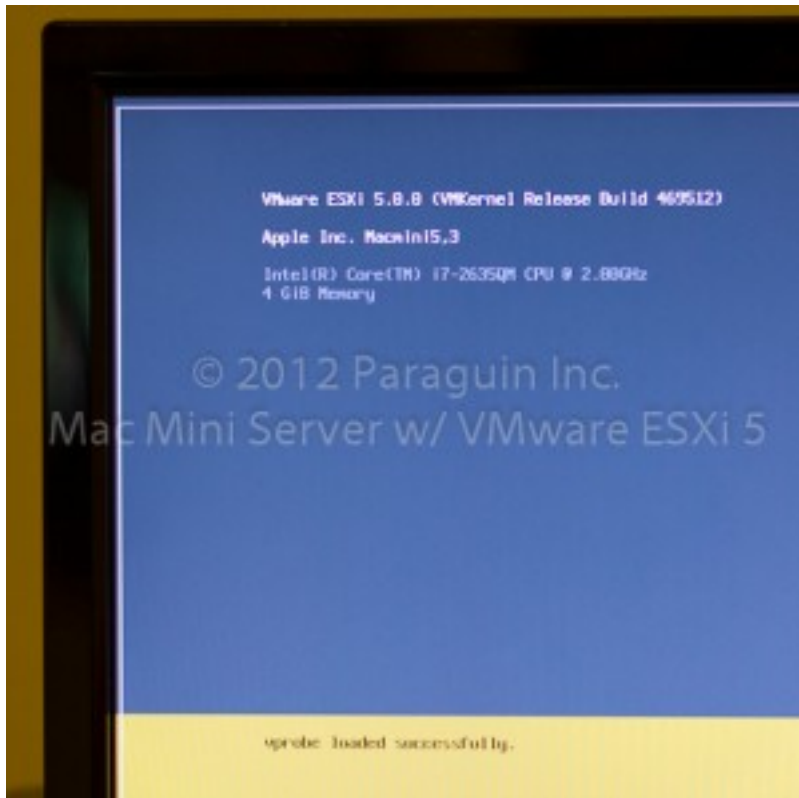
Once the installation is complete,

- Unplug the Super Drive before rebooting
- Hit the <Enter> key to perform the required reboot

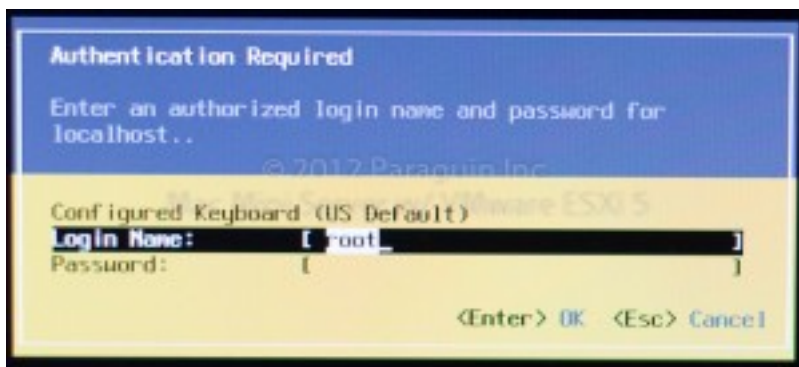


3.3 Enable ESXi shell

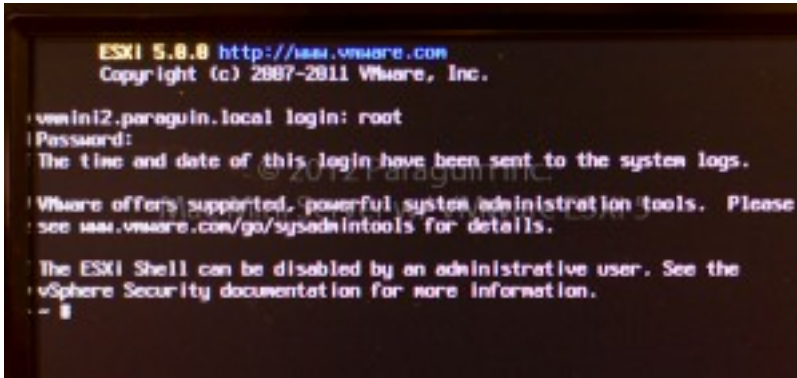
The system has booted up at this point. You should be able to see the VMware screen.



- Hit <F2> to go to the "system customization"
- Enter the root password (e.g. T3amw0rk)



- Press <enter> to Select "Troubleshooting options"
- Press <enter> to enable ESXi shell
- Press <esc> to apply the changes
- press <alt> + <F1> to go to the shell
- login as root



3.4 Update the Broadcom NIC driver

- Plug the SuperDrive to your Windows workstation
- Eject the VMware ESXi setup CD-ROM
- Insert the Broadcom NIC driver CD-ROM

The following are the commands of how to mount the CD from the SuperDrive, copy the driver over to the local system, update the driver on the system, and then reboot.

- load the VMKernel module

```
# vmkload_mod iso9660
```

You should get the message *Module iso9660 loaded successfully*

- Mount the CD

```
# vsish -e set /vmkModules/iso9660/mount $(esxcli config mp -b | grep "CD-ROM" | awk '{print $1 ← }')
```

- Unzip the TG3*.zip file to a temp directory

```
# mkdir /tmp-brcm-driver
# cp /vmfs/volumes/mpx.vmhba32XXXXX/TG3312~1.ZIP /tmp-brcm-driver
# cd /tmp-brcm-driver
# unzip TG3312~1.ZIP
```

- Copy the offline bundle zip to /var/log/vmware

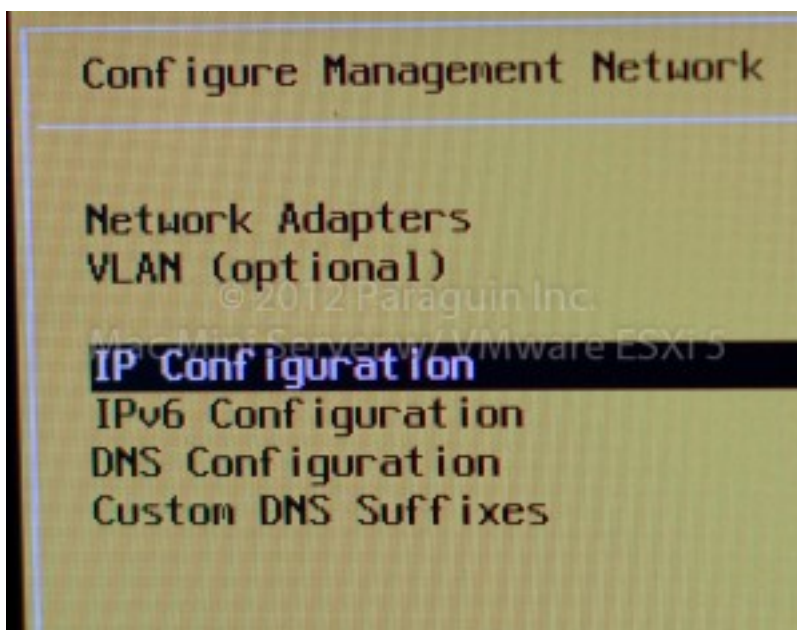
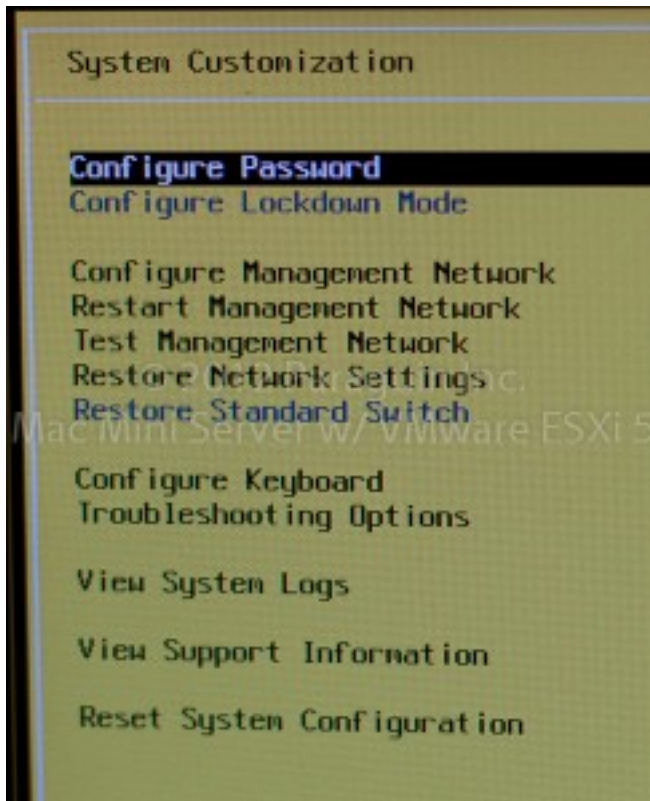
```
# cp tg3-3.120h.v50.2-offline_bundle-547149.zip /var/log/vmware
```

- Perform the driver upgrade and reboot

```
# esxcli software vib install --no-sig-check --maintenance-mode -d tg3-3.120.h.v50.2- ←
offline-bundle-547149.zip
# reboot
```

3.5 Configure and test the network settings

- Hit <F2> in order to go into the configuration
- Set the root password to T3amw0rk
- Select "Configure management network"



- Set a static IP address with the information given by the network administrator.

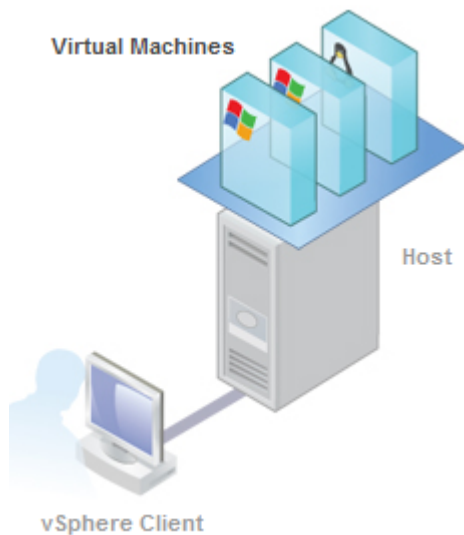
In this instance, I used the following data:

```
address          : 172.18.21.251
netmask          : {netmask}
gateway          : {gateway-ip-address}
Primary name server : {primary-name-server-ip-address}
Secondary name server : {secondary-name-server-ip_address}
hostname         : acs-cib.africacentre.local
```

- Press <ESC> to go back to the "system customization screen"
- Press <Y> to apply the changes and restart management network
- Select "Test configuration".

3.6 Install the vSphere Client

To remotely access the newly installed ESXi host, you are going to install vSphere Client on your workstation.



- Go the shared repository
- Select vSphere client
- Run the installer
- Start the client
- Log in to the CiB host
- Register the host with the license key 5N2CQ-02H4L-08C49-0K1RP-A13J1

Congratulations! You are now ready to start the CIB configuration.

Chapter 4

Virtual machine installation

In this step, you are going to create virtual machines running Linux Ubuntu or Microsoft 7 Professional operating systems.

The process can take between 2 to 4 hours depending of the experience and familiarity with the tools. It is at this stage a manual process involving a combination of GUI mouse clicks and CLI keystrokes.

To complete this task, you should have some experience in installing Windows 7, Ubuntu, databases and other applications.

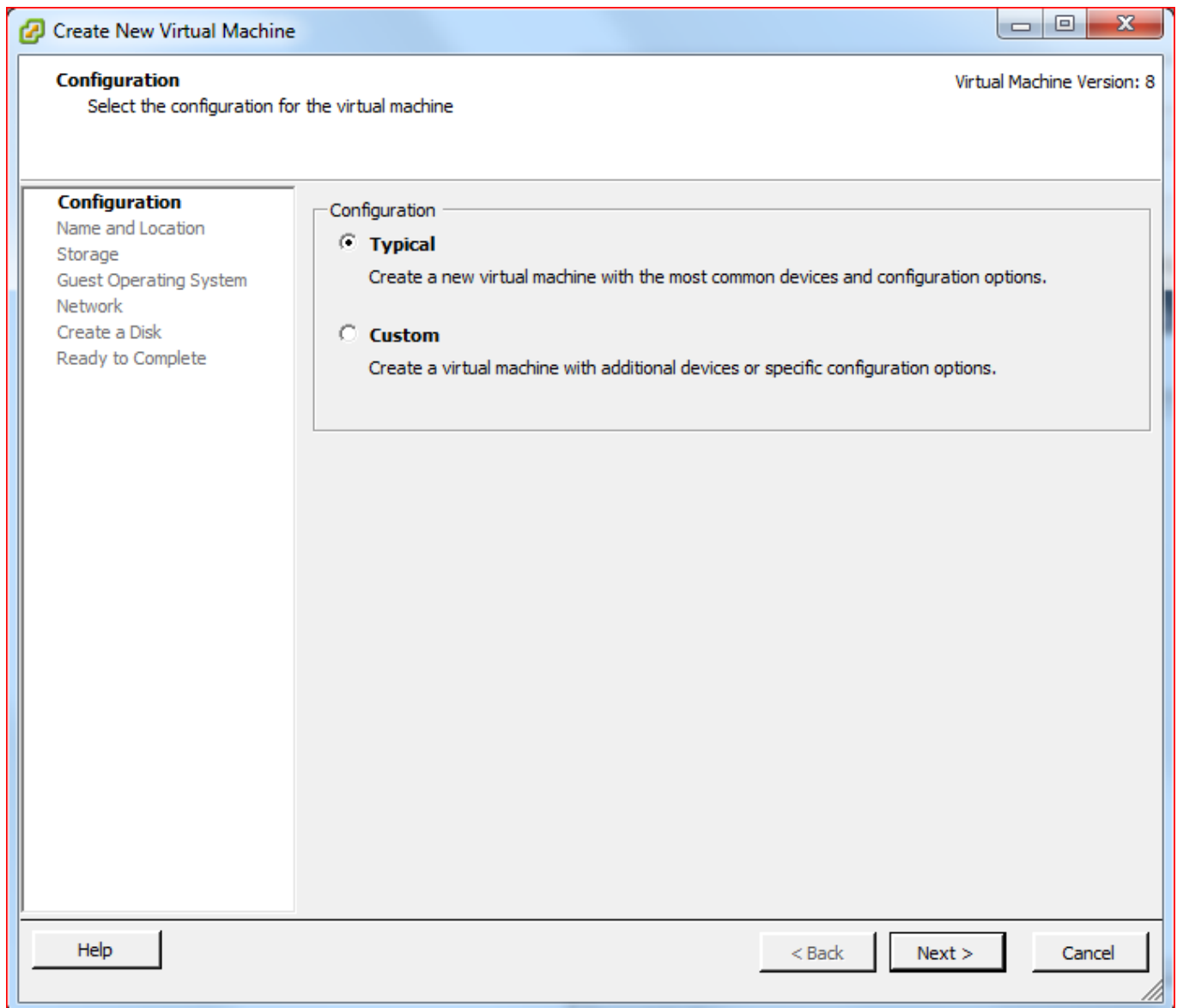
4.1 Prepare for the virtual machine installation

- Create a datastore
- Create a folder
- Transfer the Linux 12.04 LTS 64-bit ISO images from the repository to the datastore
- Transfer the Windows 7 Professional 64-bit ISO images from the repository to the datastore

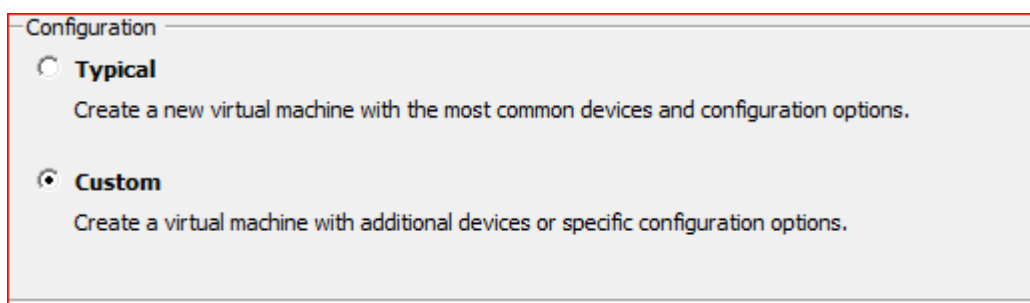
4.2 Create a virtual machine

- Open your vSphere Client
- Go to **File** → **New** → **Virtual Machine**

The new virtual machine wizard opens





- Select **Custom** to create a virtual machine with additional devices or specific configuration options.



- Name your virtual machine
- Select the datastore1

Select a destination storage for the virtual machine files:

Name	Drive Type	Capacity	Provisioned	Free	Type	Thin Pr
 datastore1	Non-SSD	460.75 GB	288.42 GB	177.30 GB	VMFS5	Support
 VMBackup	Non-SSD	465.50 GB	62.55 GB	402.95 GB	VMFS5	Support

- Select the virtual machine version: 8

Virtual Machine Version

This host or cluster supports more than one VMware virtual machine version. Specify the virtual machine version to use.

☐ **Virtual Machine Version: 7**

This version will run on VMware ESX/ESXi 4.0 and later. This version is recommended when sharing storage or virtual machines with ESX/ESXi up to 4.1.

☒ **Virtual Machine Version: 8**

This version will run on VMware ESXi 5.0 and later. Choose this version if you need the latest virtual machine features and do not need to migrate to ESX/ESXi 4.

- Select the guest operating system
 - For Windows

Guest Operating System:

☒ Windows

☐ Linux

☐ Other

Version:

Microsoft Windows 7 (64-bit)

Identifying the guest operating system here allows the wizard to provide the appropriate defaults for the operating system installation.

- For Linux

Guest Operating System:

☐ Windows

☒ Linux

☐ Other

Version:

Ubuntu Linux (64-bit)

Identifying the guest operating system here allows the wizard to provide the appropriate defaults for the operating system installation.

- Choose the number of cores

Number of virtual sockets:

Number of cores per virtual socket:

Total number of cores: 1

The number of virtual CPUs that you can add to a VM depends on the number of CPUs on the host and the number of CPUs supported by the guest OS.

- Choose the memory

Memory Configuration

Memory Size:

1011 GB
512 GB
256 GB
128 GB
64 GB
32 GB
16 GB
8 GB
4 GB
2 GB
1 GB
512 MB
256 MB
128 MB
64 MB
32 MB
16 MB
8 MB
4 MB

Maximum recommended for this guest OS: 128 GB.
Maximum recommended for best performance: 16296 MB.
Default recommended for this guest OS: 2 GB.
Minimum recommended for this guest OS: 512 MB.

- Choose the number of network connections

Create Network Connections

How many NICs do you want to connect?

Network	Adapter	Connect at Power On
NIC 1: VM Network	E1000	<input checked="" type="checkbox"/>

- Choose the SCSI controller

SCSI controller

☐ BusLogic Parallel (not recommended for this guest OS)

☐ LSI Logic Parallel

☒ LSI Logic SAS

☐ VMware Paravirtual

- Create a new virtual disk

Disk

☒ Create a new virtual disk

☐ Use an existing virtual disk
Reuse a previously configured virtual disk.

☐ Raw Device Mappings
Give your virtual machine direct access to SAN. This option allows you to use existing SAN commands to manage the storage and continue to access it using a datastore.

☐ Do not create disk

Capacity

Disk Size:

Disk Provisioning

☒ Thick Provision Lazy Zeroed

☐ Thick Provision Eager Zeroed

☐ Thin Provision

Location

☒ Store with the virtual machine

☐ Specify a datastore or datastore cluster:

Advanced option for the virtual disk

Specify the advanced options for this virtual disk. These options do not normally need to be changed.

Virtual Device Node

☒ SCSI (0:0)

☐ IDE (0:0)

Mode

☐ Independent
Independent disks are not affected by snapshots.

☐ Persistent
Changes are immediately and permanently written to the disk.

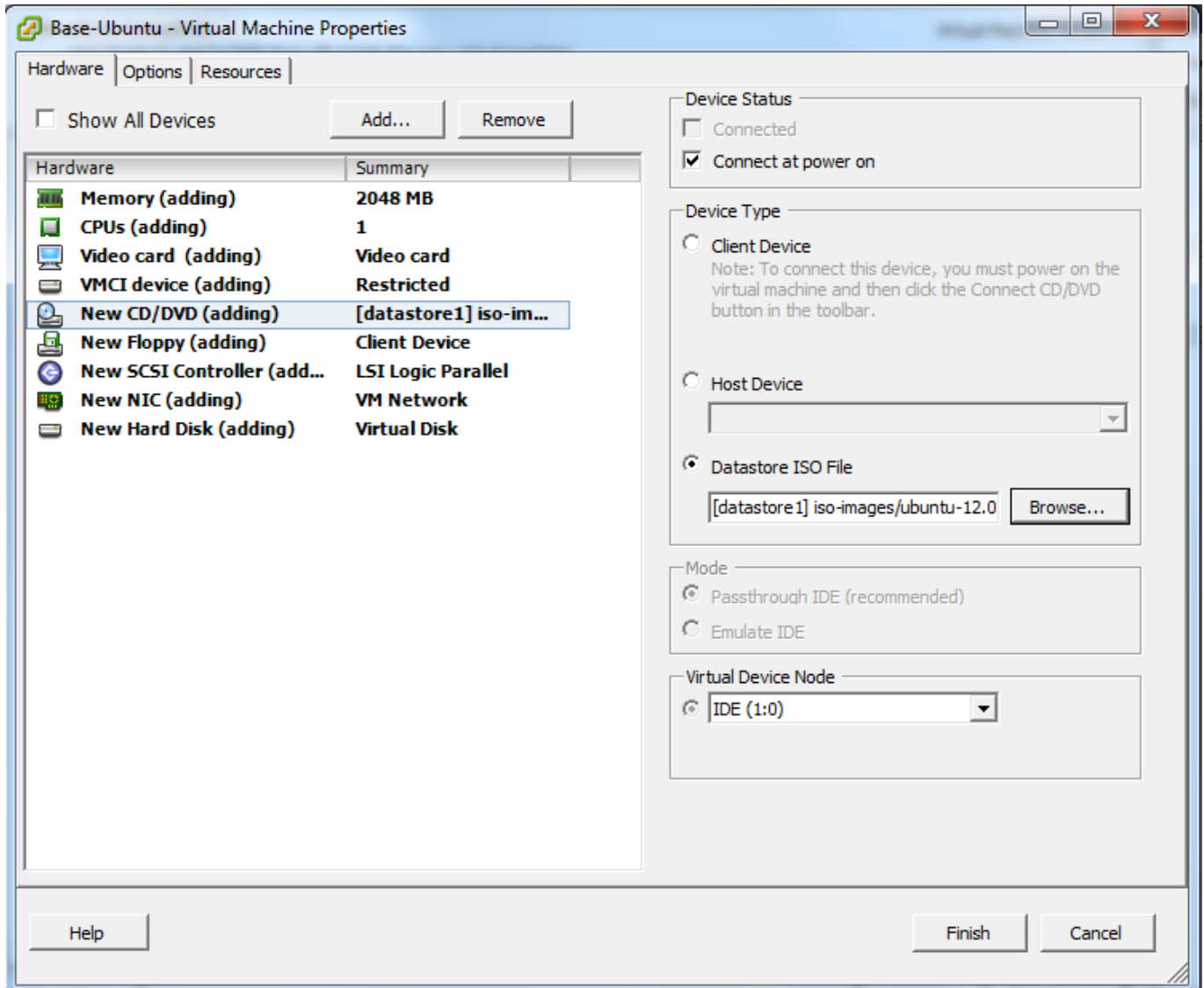
☐ Nonpersistent
Changes to this disk are discarded when you power off or revert to the snapshot.

- Edit the virtual machine settings before completion

You can see the virtual machine properties

On the hardware tabsheet,

- Select new CD/DVD,
- Check the *connect at power on*
- Select *Datastore ISO File*



At this point, the new VM has been created. You still need to install the Operating System. The next two steps shows how to install Windows 7 or Linux Ubuntu 12.04 on the newly created virtual machine.

4.3 Install Windows 7 as a guest operating system

- Open vSphere client
- Select the newly created virtual machine

During the installation of Windows 7 machines,

- Select English as your language.
- Install Windows in the Disk 0 unallocated space.
- Select time and currency format : English (United States).
- Select keyboard or input method: US.
- Accept the license terms.

- Type a user name : see login naming conventions .
- Type a computer name : see computer naming conventions.
- Type a password.
- Retype a password.
- Type a password hint.
- Install important updates only
- Select your time and date settings: Time zone,
 - Unselect **Automatically adjust clock for day light saving time**
- Select **Work network**

4.3.1 Post-installations

- Join the domain if you are part of a domain
- Setup private IP addresses
- Activate the Windows license

4.3.2 Miscellaneous

- Install Acrobat reader

4.4 Install Linux Ubuntu as a guest operating system

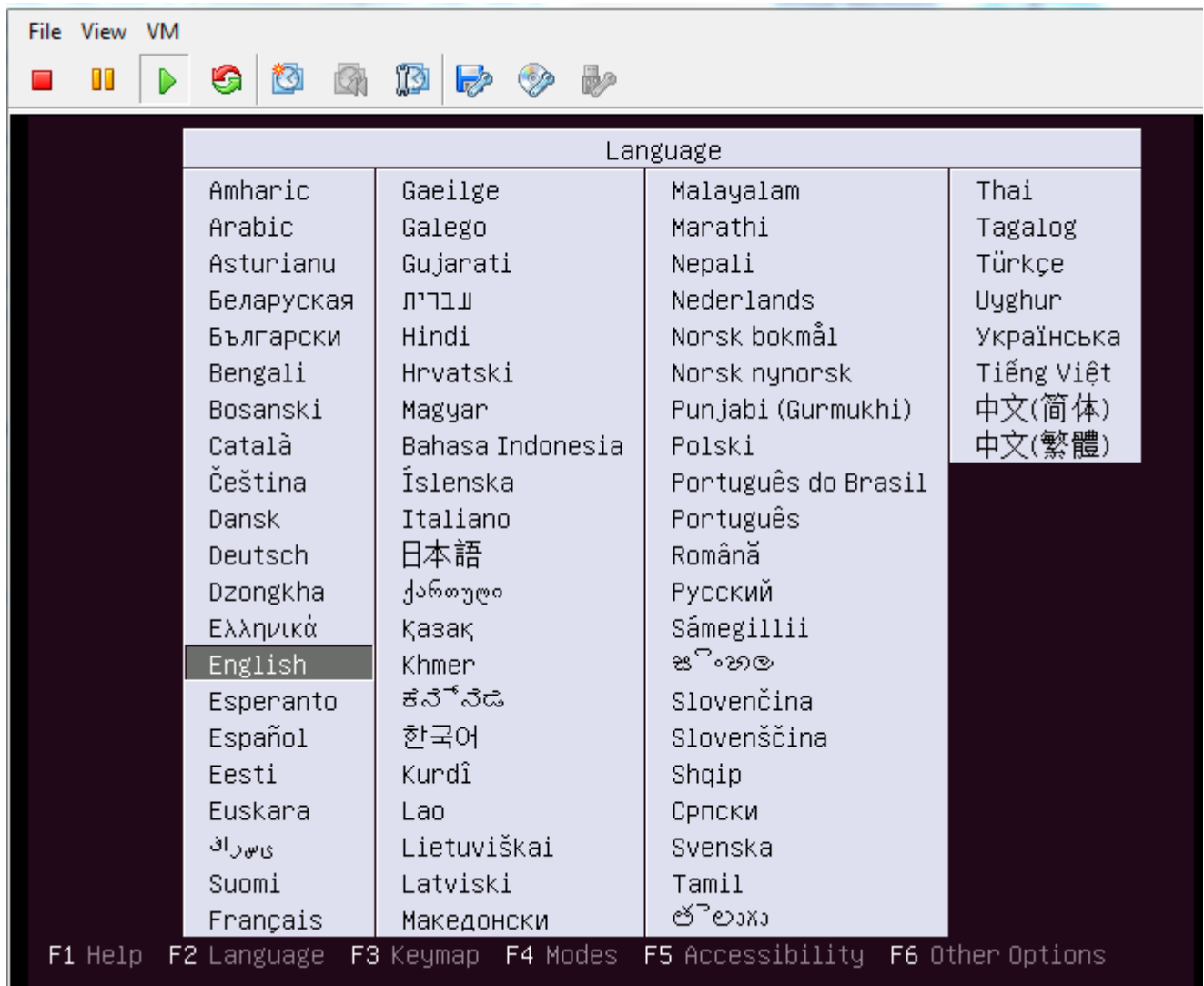
In this step, you are going to install Linux operating system on the virtual machine.

For experienced system administrators

During the installation of Ubuntu servers, choose

- 64-bit edition
- Guided disk with LVM
- OpenSSH server
- Install the GRUB boot loader to the master boot record
- Use **apt** package management to install applications

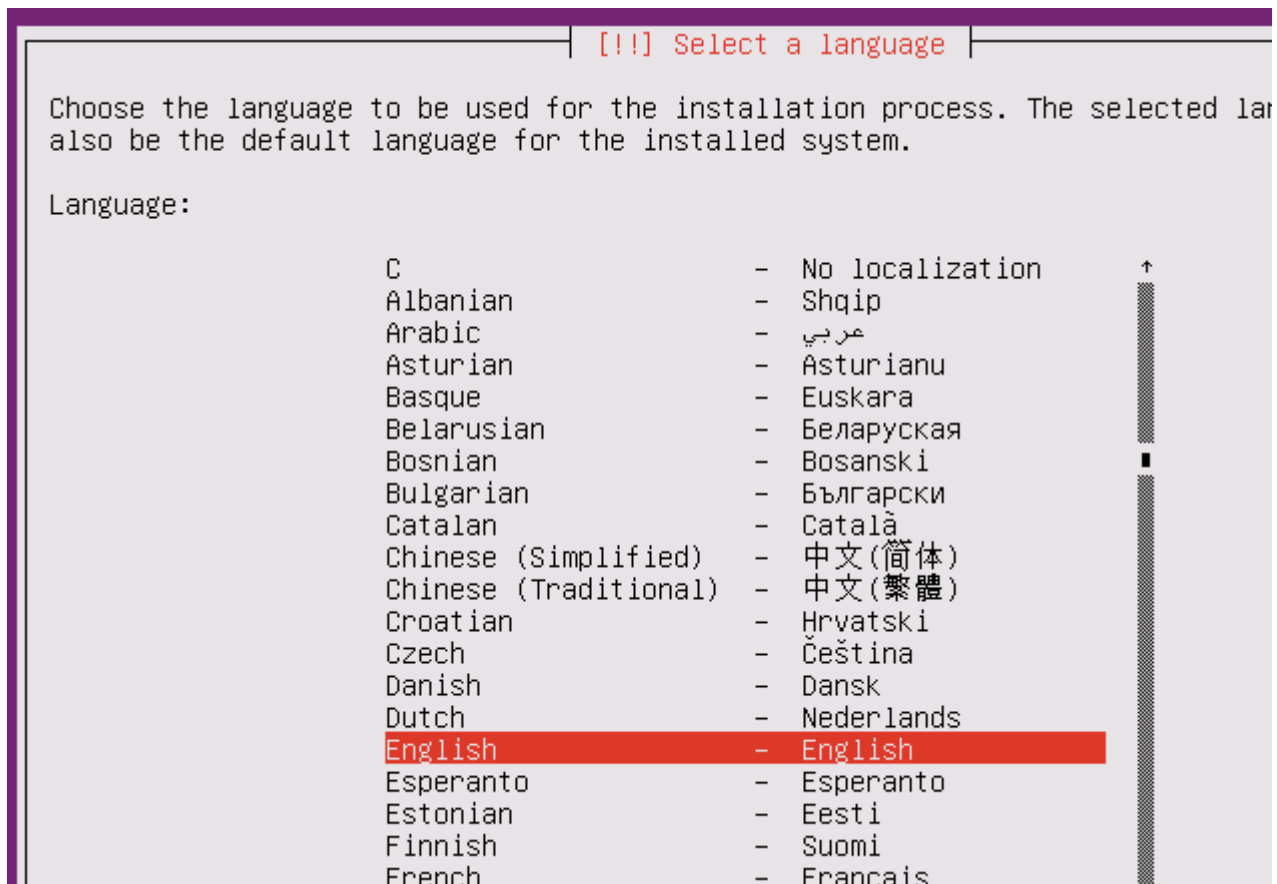
- Select English as the default language



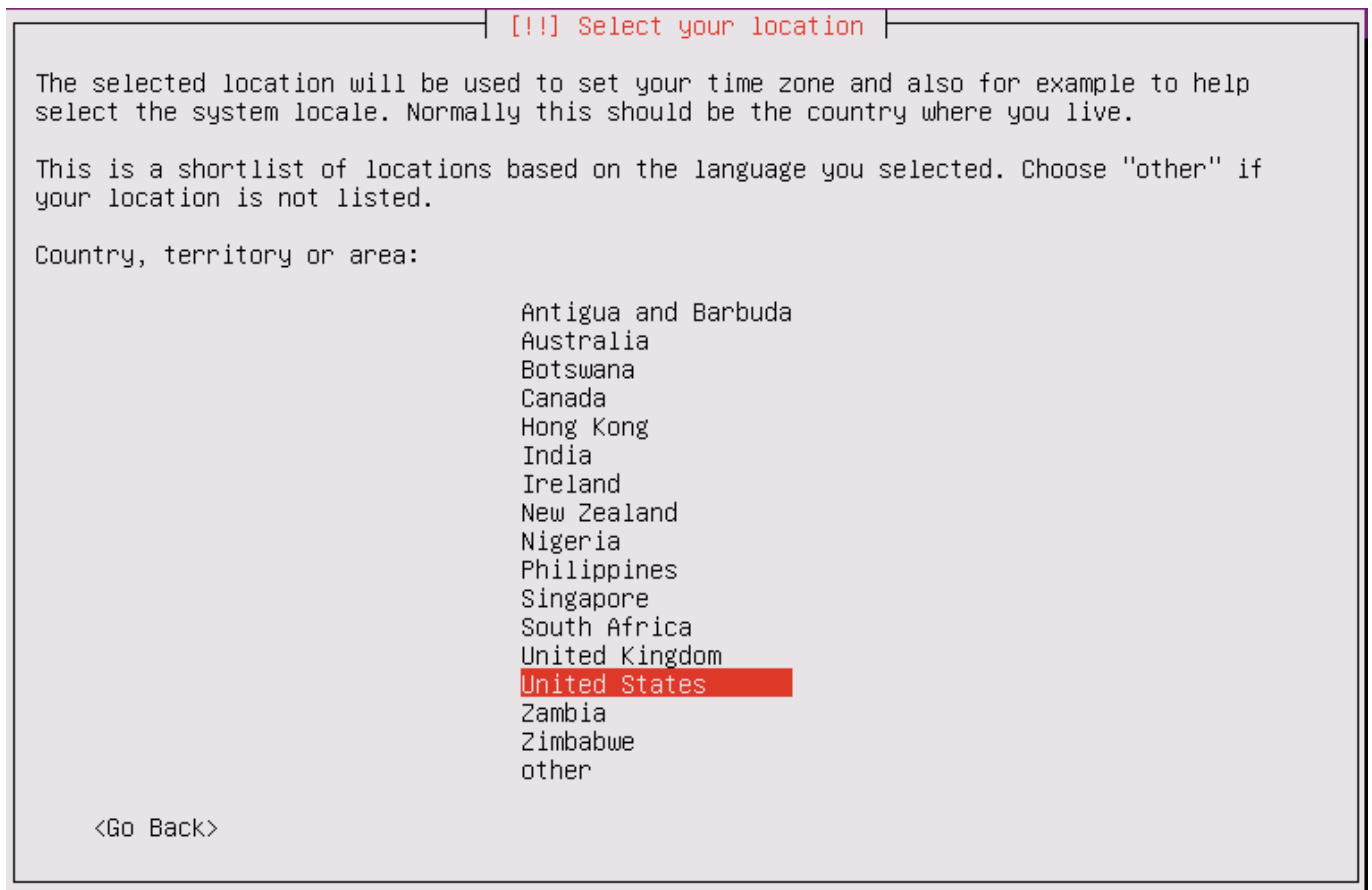
- Choose to install Ubuntu server

image:images/ubuntu-Install-server.PNG

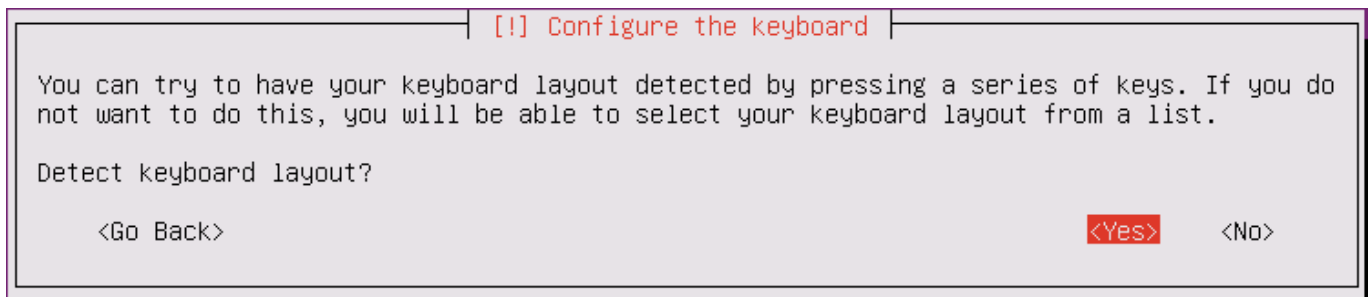
- Choose **English** as the default language



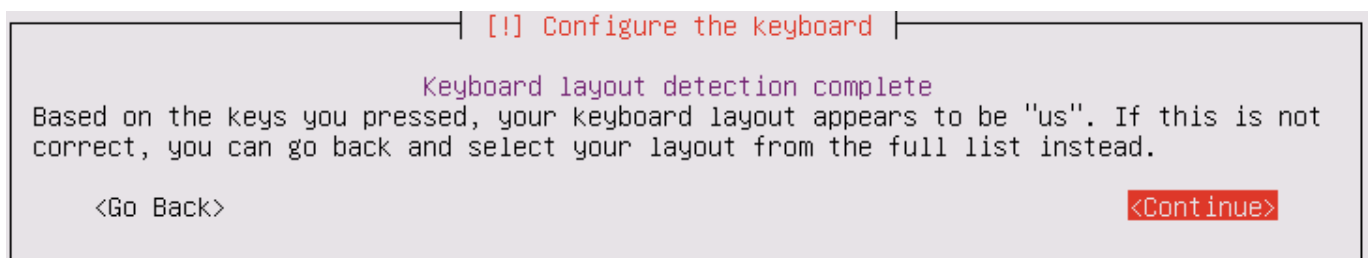
- Select your location based on the language you select. Choose other If your location is not listed.



- Select **Detect keyboard layout** to configure the keyboard



Follow the prompt until your keyboard is detected by the installer.



Enter the hostname

!!! Configure the network

Please enter the hostname for this system.

The hostname is a single word that identifies your system to the network. If you don't know what your hostname should be, consult your network administrator. If you are setting up your own home network, you can make something up here.

Hostname:

base-ubuntu

<Go Back> <Continue>

Enter your full name

!!! Set up users and passwords

A user account will be created for you to use instead of the root account for non-administrative activities.

Please enter the real name of this user. This information will be used for instance as default origin for emails sent by this user as well as any program which displays or uses the user's real name. Your full name is a reasonable choice.

Full name for the new user:

Christian Kyony

<Go Back> <Continue>

Enter your user name (see naming conventions in appendix)

!!! Set up users and passwords

Select a username for the new account. Your first name is a reasonable choice. The username should start with a lower-case letter, which can be followed by any combination of numbers and more lower-case letters.

Username for your account:

ckyony

<Go Back> <Continue>

Choose a good password (see naming conventions in appendix)

!!! Set up users and passwords

A good password will contain a mixture of letters, numbers and punctuation and should be changed at regular intervals.

Choose a password for the new user:

<Go Back> <Continue>

[[!]] Set up users and passwords

Please enter the same user password again to verify you have typed it correctly.

Re-enter password to verify:

<Go Back> <Continue>

Do not encrypt your home directory

[[!]] Set up users and passwords

You may configure your home directory for encryption, such that any files stored there remain private even if your computer is stolen.

The system will seamlessly mount your encrypted home directory each time you login and automatically unmount when you log out of all active sessions.

Encrypt your home directory?

<Go Back> <Yes> <No>

Select **Guided disk with LVM** as your partitioning method

[[!]] Partition disks

The installer can guide you through partitioning a disk (using different standard schemes) or, if you prefer, you can do it manually. With guided partitioning you will still have a chance later to review and customise the results.

If you choose guided partitioning for an entire disk, you will next be asked which disk should be used.

Partitioning method:

- Guided - use entire disk
- Guided - use entire disk and set up LVM**
- Guided - use entire disk and set up encrypted LVM
- Manual

<Go Back>

select disk to partition

!!! Partition disks

Note that all data on the disk you select will be erased, but not before you have confirmed that you really want to make the changes.

Select disk to partition:

SCSI3 (0,0,0) (sda) - 17.2 GB VMware Virtual disk

<Go Back>

To write the changes to disks and configure LVM, select **Yes**

!!! Partition disks

Before the Logical Volume Manager can be configured, the current partitioning scheme has to be written to disk. These changes cannot be undone.

After the Logical Volume Manager is configured, no additional changes to the partitioning scheme of disks containing physical volumes are allowed during the installation. Please decide if you are satisfied with the current partitioning scheme before continuing.

The partition tables of the following devices are changed:
SCSI3 (0,0,0) (sda)

Write the changes to disks and configure LVM?

<Yes>

<No>

Use the whole volume group for guided partitioning

!!! Partition disks

If you continue, the changes listed below will be written to the disks. Otherwise, you will be able to make further changes manually.

The following partitions are going to be formatted:
LVM VG base-ubuntu, LV root as ext4
LVM VG base-ubuntu, LV swap_1 as swap
partition #1 of SCSI3 (0,0,0) (sda) as ext2

Write the changes to disks?

<Yes>

<No>

Enter the proxy information if needed otherwise, leave this blank.

[!] Configure the package manager

If you need to use a HTTP proxy to access the outside world, enter the proxy information here. Otherwise, leave this blank.

The proxy information should be given in the standard form of "http://[[user][:pass]@]host[:port]/".

HTTP proxy information (blank for none):

http://172.18.20.14:8080

<Go Back>

<Continue>

No automatic updates

[!] Configuring tasksel

Applying updates on a frequent basis is an important part of keeping your system secure.

By default, updates need to be applied manually using package management tools. Alternatively, you can choose to have this system automatically download and install security updates, or you can choose to manage this system over the web as part of a group of systems using Canonical's Landscape service.

How do you want to manage upgrades on this system?

No automatic updates

Install security updates automatically

Manage system with Landscape

Select OpenSSH

Install GRUB boot loader on the hard disk

[!] Install the GRUB boot loader on a hard disk

It seems that this new installation is the only operating system on this computer. If so, it should be safe to install the GRUB boot loader to the master boot record of your first hard drive.

Warning: If the installer failed to detect another operating system that is present on your computer, modifying the master boot record will make that operating system temporarily unbootable, though GRUB can be manually configured later to boot it.

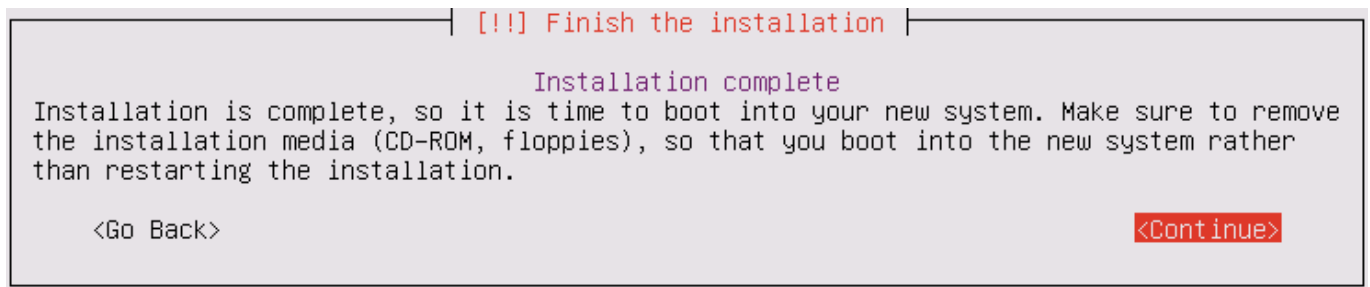
Install the GRUB boot loader to the master boot record?

<Go Back>

<Yes>

<No>

Installation is complete



The system is rebooting and display the prompt below when ready. Enter the username and password

```
Ubuntu 12.04 LTS base-ubuntu tty1

base-ubuntu login: ckyony
Password: _
```

```
Welcome to Ubuntu 12.04 LTS (GNU/Linux 3.2.0-23-generic x86_64)

* Documentation:  https://help.ubuntu.com/

System information as of Thu Oct 18 10:57:17 SAST 2012

System load:  0.06               Processes:            66
Usage of /:   8.0% of 13.69GB    Users logged in:     0
Memory usage: 1%                IP address for eth0: 172.18.21.69
Swap usage:   0%

Graph this data and manage this system at https://landscape.canonical.com/

94 packages can be updated.
46 updates are security updates.

The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

ckyony@base-ubuntu:~$ _
```

4.4.1 Setup static IP address

To set a static IP address in Ubuntu, you are going to edit **/etc/network/interfaces** file.

Initially, the file only contains information about the local loopback address:

```
auto lo
iface lo inet loopback
```

To assign a static IP address, you will need to make some changes to this file.

Let us say you want to assign a static IP of 172.18.20.191 to your *eth0* network connection with a subnet mask of 255.255.255.0 and a local gateway of 172.18.20.1.

First, make a backup copy of the interfaces file in your home directory in case something goes amiss during the editing process.

```
sudo cp /etc/network/interfaces ~
```

Add the following lines in **/etc/network/interfaces**

```
iface eth0 inet static
address 172.18.20.191
netmask 255.255.254.0
gateway 172.18.20.1
```

After saving the updated file to disk, use the command below to force Ubuntu to re-read the configuration files:

```
sudo ifup eth0
```

To test if the change is successfully, ping the host

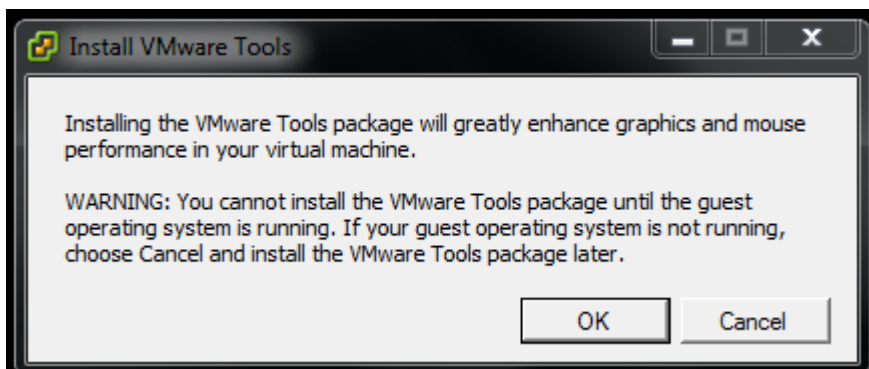
```
ping 172.18.20.191
```

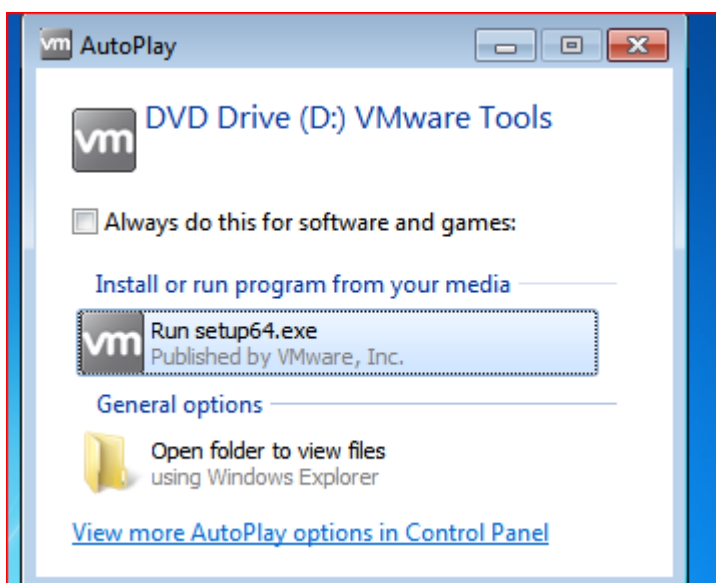
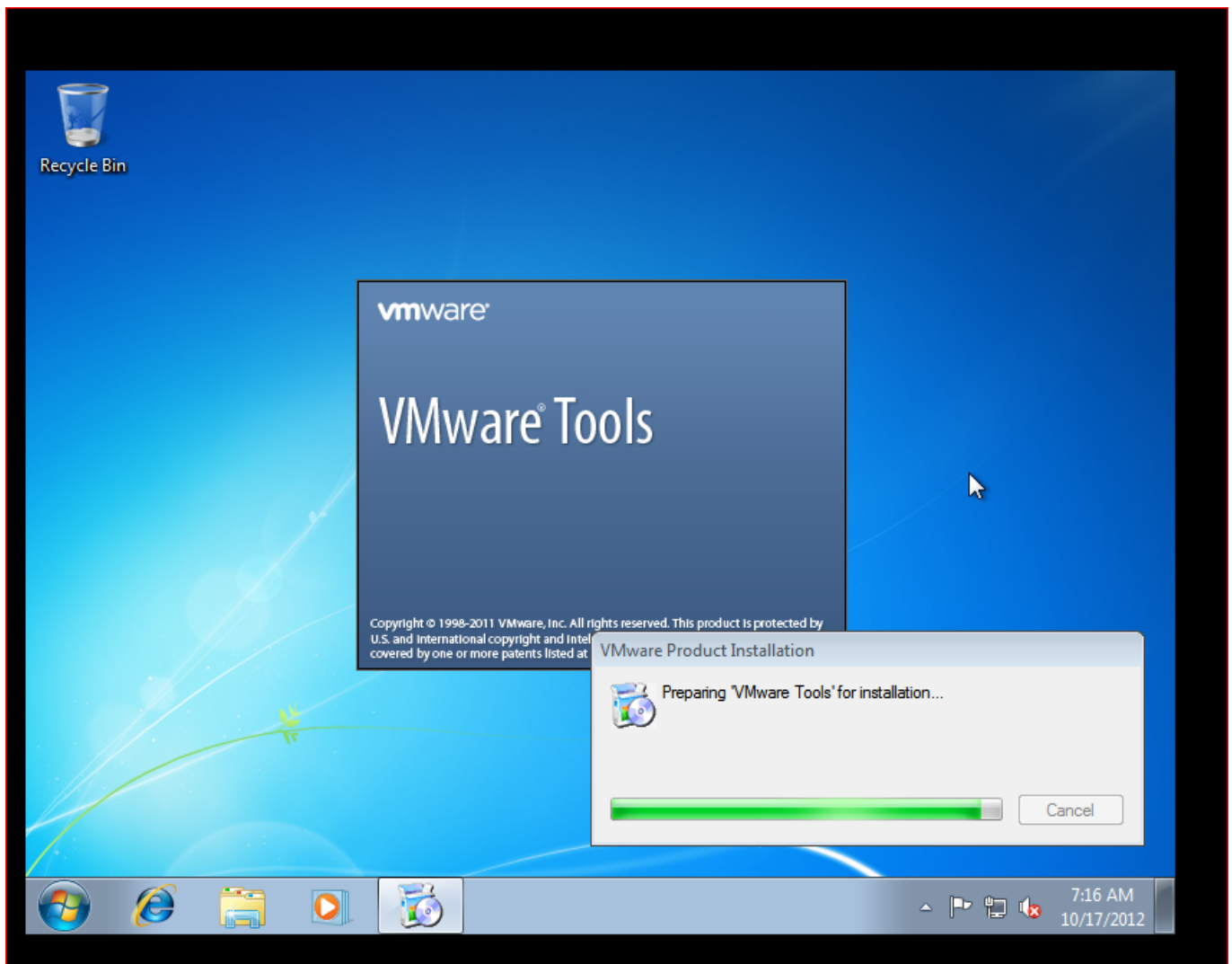
Congratulations! You have successfully installed Ubuntu on your VM

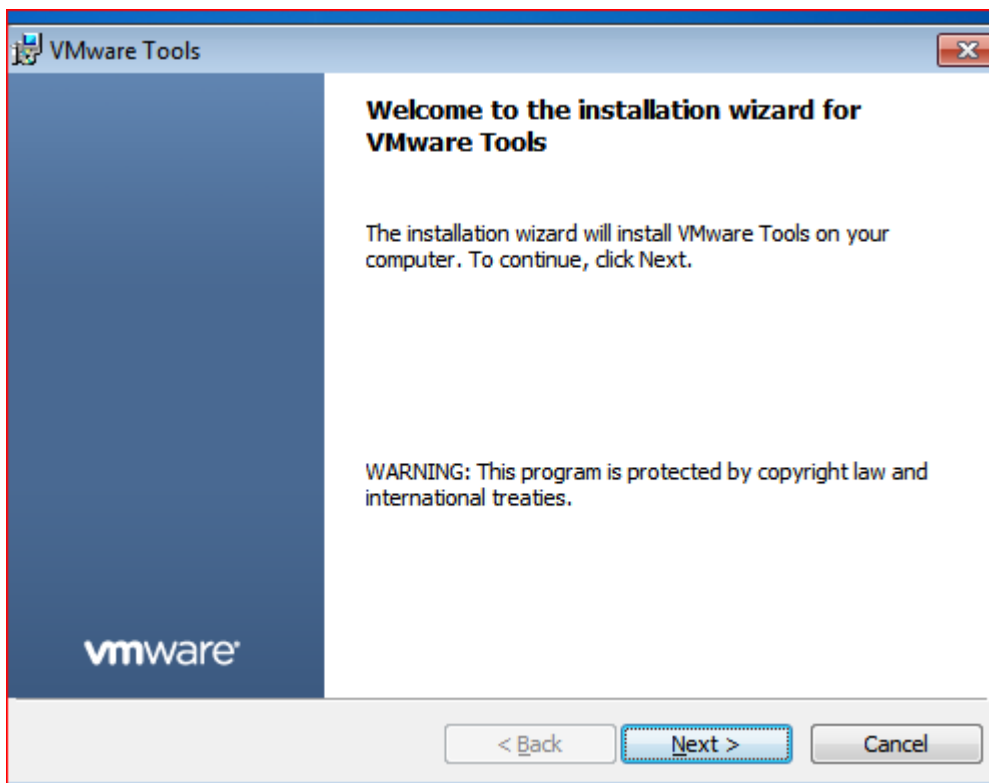
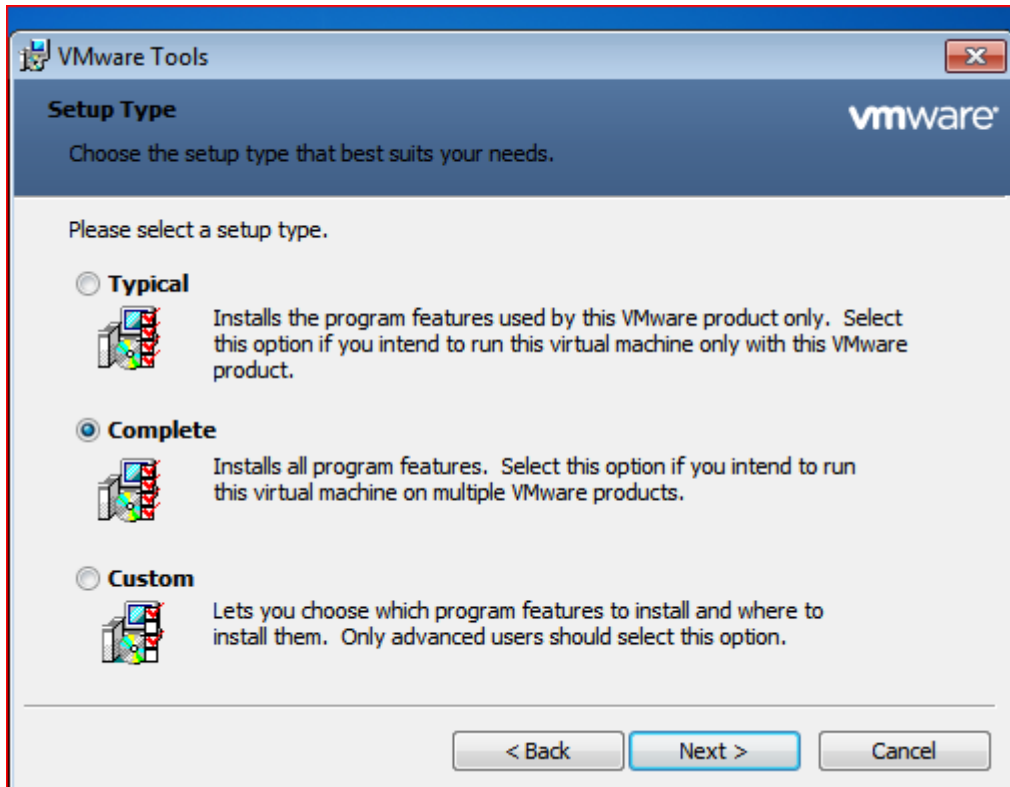
4.5 Install VMware Tools

In this step, you are going to install VMware Tools to enhance graphics and mouse performance in the newly installed virtual machine.

- Open vSphere Client
- Select the virtual machine
- Open the console
- Go to **VM > Guest > Install/Upgrade VMware Tools**







- Run setup64.exe
- Welcome wizard
- Choose **Complete** as the Setup type
- Restart the virtual machine for the configuration changes to take effect.

At this point, you should know how to create a virtual machine how to install Windows 7 or Ubuntu 12.04 How to install VMware tools to improve GUI performance

In the next steps, you are going to install the application servers on the virtual machines.

Chapter 5

Archive Server installation

5.1 Role

The main role of the archive server is to provide a shared file storage to Windows and Linux clients. It is essentially a file server running the SAMBA package on a Linux based machine. It is also setup as a FTP /GIT server to facilitate file sharing with iSHARE2 technical support.

The secondary role is to provide authentication services for Linux clients in a Active Directory environment.

5.2 Setup the Virtual machine

Using the instructions in Section 4.2, “[Create a virtual machine](#)” and Section 4.4, “[Install Linux Ubuntu as a guest operating system](#)”, setup a new VM with the following specifications:

- OS : Ubuntu Server 12.04 LTS 64-bit
- CPU : 1 core
- RAM : 2 GB
- Storage : 100 GB
- IP Address: 172.18.20.191
- Host Name: archive.cib.<code site>

5.3 Samba installation

5.3.1 Install the Samba package

In this step, we are going to install the Samba packages.

```
sudo apt-get install libcups2 samba samba-common
```

5.3.2 Allow only authenticated users

Edit the `/etc/samba/smb.conf`

In the global section, remove the `#` at the beginning of the line `security = user` so that it looks like this:

```
# "security = user" is always a good idea. This will require a Unix account
# in this server for every user accessing the server. See
# /usr/share/doc/samba-doc/htmldocs/Samba3-HOWTO/ServerType.html
# in the samba-doc package for details.
security = user
```

This enables Linux users to log in to the Samba server. Close the file and restart Samba.

```
sudo service smbd restart
```

5.3.3 Create the shared directory for all authenticated users

Create the directory for sharing the files and change the group to the *users* group:

```
sudo mkdir /home/public/
sudo chown -R root:users /home/public/
sudo chmod -R ug+rw,ox+rx-w /home/public/
```

At the end of the `/etc/samba/smb.conf` file add the following lines:

```
[public]
comment = public folder accessible by authenticated users
path = /home/public
valid users = @users
force group = users
create mask = 0660
directory mask = 0771
writable = yes
```

5.3.4 Create a private directory for home

if you want all users to be able to read and write to their home directories via Samba, add the following lines to the `/etc/samba/smb.conf` file:

```
[homes]
comment = Home directories
browseable = no
valid users = %S
writable = yes
create mask = 0700
directory mask = 0700
```

Now restart Samba

```
sudo service smbd restart
```

5.3.5 Add and manage users

In this example, you will add a user named **mandela**. You can add as many users as you need in the same way, just replace username **mandela** with the desired username in the commands.

```
sudo passwd mandela
```

→ Enter the password for the new user.

Now add the user to the Samba user database:

```
sudo smbpasswd -a mandela
```

→ Enter the password of the new user.

5.3.6 Access the shares from Windows client

Now you should be able to log in from your Windows workstation with the file explorer (address is \\172.18.20.191) using the username *mandela* and the chosen password

5.3.7 Access the shares from linux clients

Create a mount point

```
sudo mkdir -p /mnt/cib/archive
```

Mount the share using the cifs protocol (instead of the deprecated smbfs)

```
sudo mount -t cifs //172.18.20.191/public/ /mnt/cib/archive -o user=ckyonny
```

If you intend to access your share frequently, edit **/etc/fstab** file and add a line that looks like this:

```
//172.18.20.191/public /mnt/cib/public cifs noauto,credentials=/home/ckyonny/.cib- ↵  
credentials 0 0
```

Create a **~/.cib-credentials/** file in your home directory to store your credentials

```
user=ckyonny  
password=mysecret
```

Protect the file so only you can access the credentials file

```
chmod 600 ~/.cib-credentials
```

To access the share, mount it with the simple command

```
sudo mount /mnt/cib/public
```

5.4 Git installation

It is important to keep configuration under strict version control. **Git** is a distributed revision control system with a command set that provides both high-level operations and full access to internals.

5.4.1 Installation

5.4.2 Configuration

Chapter 6

Nada server installation

6.1 Role

The main role of this server is to host the NADA data repository.

6.2 Setup the virtual machine

Using the instructions in Section 4.2, “[Create a virtual machine](#)” and Section 4.4, “[Install Linux Ubuntu as a guest operating system](#)”, setup a new VM with the following specifications:

- OS : Ubuntu Server 12.04 LTS
- CPU : 1
- RAM : 2
- Storage : 40 GB
- IP Address : 172.18.21.232
- Host Name : Nada.cib.<code site>

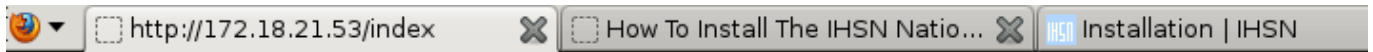
In the next steps, you are going to install the software required by NADA: Apache 2+2+ , PHP 5.2+ and MySql 5.0+

6.3 Install the Apache web server

Open a terminal and type the following command:

```
sudo apt-get install apache2
```

To test if Apache is installed, open your browser with the Nada IP address e.g. 172.18.21.232. You should see the message below:



It works!

This is the default web page for this server.

The web server software is running but no content has been added, yet.

6.4 Install PHP and required extensions

Open a terminal

```
sudo apt-get install php5
```

To test if PHP is correctly installed, create a text file called `/var/www/info.php`.

Enter the following text into the file

```
<?php echo phpinfo(); ?>
```

After saving the file, browse to <http://localhost/info.php> from the local machine. If you are accessing the server from another machine, enter the IP

If PHP is correctly installed, you will see a screen like this:



PHP Version 5.3.10-1ubuntu3.4



System	Linux base-ubuntu 3.2.0-23-generic #36-Ubuntu SMP Tue Apr 10 20:39:51 UTC 2012 x86_64
Build Date	Sep 12 2012 18:42:53
Server API	Apache 2.0 Handler
Virtual Directory Support	disabled
Configuration File (php.ini) Path	/etc/php5/apache2
Loaded Configuration File	/etc/php5/apache2/php.ini
Scan this dir for additional .ini files	/etc/php5/apache2/conf.d
Additional .ini files parsed	/etc/php5/apache2/conf.d/pdo.ini
PHP API	20090626

The NADA application requires the following PHP extensions: xml, xsl, mysql, sqlite, simplexml, zip, mbstring, openssl
Install them with the command line shown below

```
sudo apt-get install libapache2-mod-php5 php5-mysql php5-xsl
```

Restart the Apache server

```
sudo service apache2 restart
```

Increase the **upload_max_filesize** and **post_max_size** settings in the **/etc/php5/apache2/php.ini** configuration file.

```
post_max_size = 20M  
upload_max_filesize = 16M
```

- Set your time zone in the ***/etc/php5/config/apache2/conf**

```
date.timezone = 'Africa/Pretoria'
```

6.5 Install MySQL database server

Open a terminal, type the command line shown below:

```
sudo apt-get install mysql-server
```



Important

Make a note of the password you set during the mysql installation.

Create a database for the NADA installation

```
sudo mysql -u root -p
```

- Enter the root password you set when you installed the LAMP server
- Create a database called **nada**

```
mysql> create database nada4_ddp;
```

- Create a MySQL user called nada who has permission to access the nada database created above.

```
mysql> grant select, insert, update, delete, create, drop, index, alter, create temporary  
      tables,  
lock tables on nada.* to 'nada'@'localhost' identified by 'yourpassword';
```

yourpassword can be anything you choose. nada is the name of the database the user gets access to. localhost is the location which gets access to your database. Note as before remember this password you will need it later!

- Exit the MySql shell

```
mysql> exit
```

6.6 Install the NADA application files

Make a directory for the nada software in the **/var/www** folder and copy

```
cd /var/www/  
sudo wget http://{archive-ip-address}/public/nada/nada-4.0.zip
```

- Unzip the file:

```
sudo unzip nada-4.0.zip
```

- Change the permissions to the following folders:
 - **/var/www/nada/datafiles**: stores the DDI's documentation and data.
 - **/var/www/nada/cache**: stores cached web pages.
 - **/var/www/nada/logs**: stores log files.

```
sudo chmod -R a+rwX {nada-install-dir}datafiles  
sudo chmod -R a+rwX {nada-install-dir}application/cache  
sudo chmod -R a+rwX {nada-install-dir}application/logs
```

6.7 Configure the NADA database connection settings

To enter the correct database connection settings, edit the **/var/www/nada/application/config/database.php**


Go to line number 45 to 48

```
$db['default']['hostname'] = "localhost";  
$db['default']['username'] = "nada"  
$db['default']['password'] = "T3amw0rk"  
$db['default']['database'] = "nada4_ddp"
```

6.8 Run the web installer

You are now all set to run the Web installer and get your site up. To run the installer, open a web browser and navigate to the location for your new website. (e.g **http://172.18.21.232/nada/index.php/install**)

The following screen should be displayed:


NADA 3

[ARABIC](#)
[ENGLISH](#)
[ENGLISH_DDP](#)
[FRENCH](#)
[MONGOLIAN](#)
[MONGOLIAN - COPY](#)
[RUSSIAN](#)
[SPANISH](#)

NADA Installer

Server information

PHP version:	5.3.10-1ubuntu3.4
DB version:	5.5.24-0ubuntu0.12.04.1 - connection was successful!
Web server:	Apache/2.2.22 (Ubuntu)

Required PHP Extensions

Extensions	Enabled
xsl	✓
xml	✓
simplexml	✓
xmlreader	✓
zip (optional)	✓
mbstring (optional)	✓
mysql	✓

Other PHPRINI Settings

Setting	Value	Recommended
file_uploads	Enabled	Enabled
post_max_size	20M	20M
upload_max_filesize	16M	15M
date.timezone	Africa/Pretoria	See how to configure and select the right timezone

Folder READ/WRITE/DELETE permissions

Folder	Read/Write	Delete
Catalog (datafiles)	✓	✓
Cache (./cache/)	✓	✓
Log (./logs/)	✓	✓

Install Database

<http://172.18.21.53/nada/index.php/install> [-]
 [1/4]Top

The installer checks a number of settings on the server and provides a summary like the above.

- It provides information on the version of PHP running on your server (NADA requires PHP 5.2+ to be installed) If you do not have the correct version to meet the minimum requirements for the NADA, either upgrade your PHP version to the required version or request your hosting company (ISP) to install PHP 5.2 or later for your account.
- Check that the MySQL you are using is at least version 5.0, and, if not, install a newer version or request your hosting company to install a newer version for you. The checklist also shows if you have successfully connected to the database or not. If you have not, then go back to the instructions on the database connection settings and correct the problem.
- Check that there are green ticks next to all the required PHP extensions listed. If there is a red cross next to any of them then you will need to go back and install and configure the missing PHP extensions on your server (or request that your ISP enables the missing extensions for you). If you do not know how to do this then Windows users go here. For Linux users consult the User Guides on how to install PHP extensions for your distribution.
- Review the section on php.ini settings. These settings determine how big a file your PHP settings will allow to be uploaded to the server. The recommended value is given as a guideline for you. To change the settings edit your php.ini file as described in the PHP installation guide for your operating system\distribution. This process usually involves locating your php.ini file on your server, opening it with a text editor, looking for the variables file_uploads, post_max_size and upload_max_filesize and changing the values to the recommended values (up to the maximum size of the files you will upload). The next steps are saving the file and, lastly, restarting your web server.
- The last section deals with file permissions for folders which are required to have READ/WRITE/DELETE permissions set for them. If any of these folders have red crosses next to them then go back and 1) check that the folder exists on the server –if not create it in the location shown in the grey text next to each item. 2) change the permissions to read/write/delete.

- Once all the settings above have been completed as per [?sidebar], click on the **Install Database** button at the bottom left of the screen.
- Create the administrator account

image:images/nada-create-administrator-account.PNG

- Click the **Create account** button.

... Missing figure of Nada successful installation

6.9 Launch Nada

Open browser and type the URL to launch the NADA.

Firefox | iSHARE2 | Home | localhost / localhost | phpMyAdmin ... | INDEPTH iSHARE2 - Home

localhost/nada/index.php/home


Y! Yahoo! Search SEARCH Google

INDEPTH Network
Better Health Information for Better Health Policy

HOME WHAT IS THIS? HOW TO USE IT? HISTORY DATA CITATIONS NEWS CONTACT US


INDEPTH Network Data Repository

INDEPTH iSHARE2





Dikgale HDSS, South Africa
The HDSS was established in 1995, which includes eight villages. The Dikgale HDSS site is located in the central region of Mankweng District, Northern Province, South Africa, about 40 km from Polokwane, the capital of the Limpopo Province, and 15 km from the University of the North. The site covers an area of 71 km² and is 6 km x 10.8 km. The villages are typical rural settlements where the dwelling units consist of a mixture of shacks, traditional mud huts and conventional brick houses. Each village has a central residential area comprising demarcated housing stands, with communal grazing land some distance away. [Read More »](#)

Central Data Catalog
The Central Data Catalog is a portal for all surveys and datasets held in catalogs maintained by the INDEPTH Network and a number of contributing external catalogs.

Search the Central Microdata Catalog  [View all Surveys »](#)

INDEPTH iSHARE2 REPOSITORIES

 **Climate Change**


As of October 30, 2012 the Library contains
3 surveys
0 citations
151 variables

FAQ'S
• [How can I contribute to improving the catalog?](#)
• [Can you help with analyzing the data?](#)
• [Can I get help in implementing a survey catalog in my agency?](#)
[Click here for more...](#)

LATEST ADDITIONS

Chapter 7

Database server installation

7.1 Role

The main role of this server is to host the member site database.

There are a variety of options: MS SQL Server, Access, FoxPro, MySQL, Oracle, dBase. In this manual, I will provide two examples:

- MS SQL Express 2008 R2 with Tools on Windows 7
- MySql 5.5 on Ubuntu 12.04

7.2 Install MS SQL Express 2008 on Windows 7

7.2.1 Setup the VM

Following the instructions in Section 4.2, “[Create a virtual machine](#)” and Section 4.3, “[Install Windows 7 as a guest operating system](#)”, setup a new virtual machine with the following specifications:

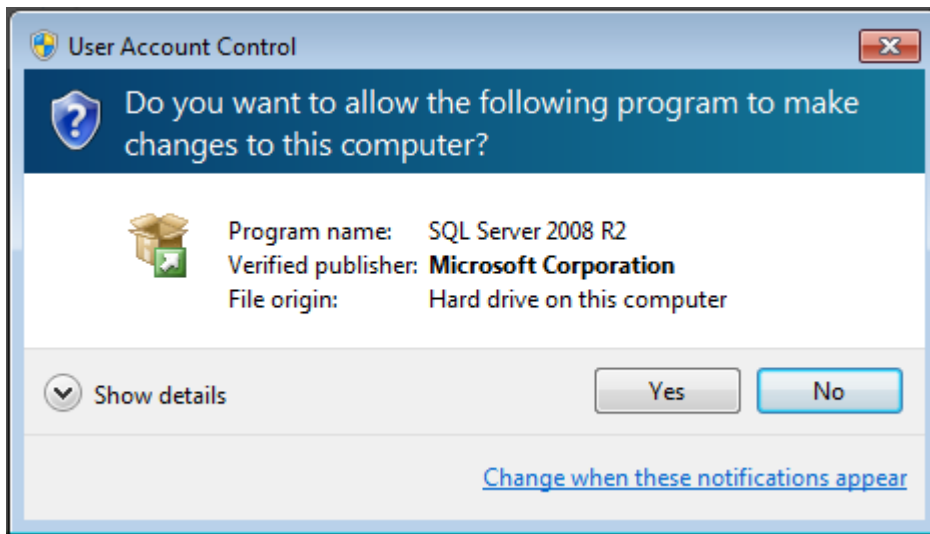
- OS : Windows 7 Pro 64-bit
- CPU : 2-core
- RAM : 4 GB
- Storage : 40 GB
- IP Address : 172.18.20.xxx
- Host Name : db.cib.<code site>

7.2.2 Install MS SQL 2008 R2 Express edition with tools

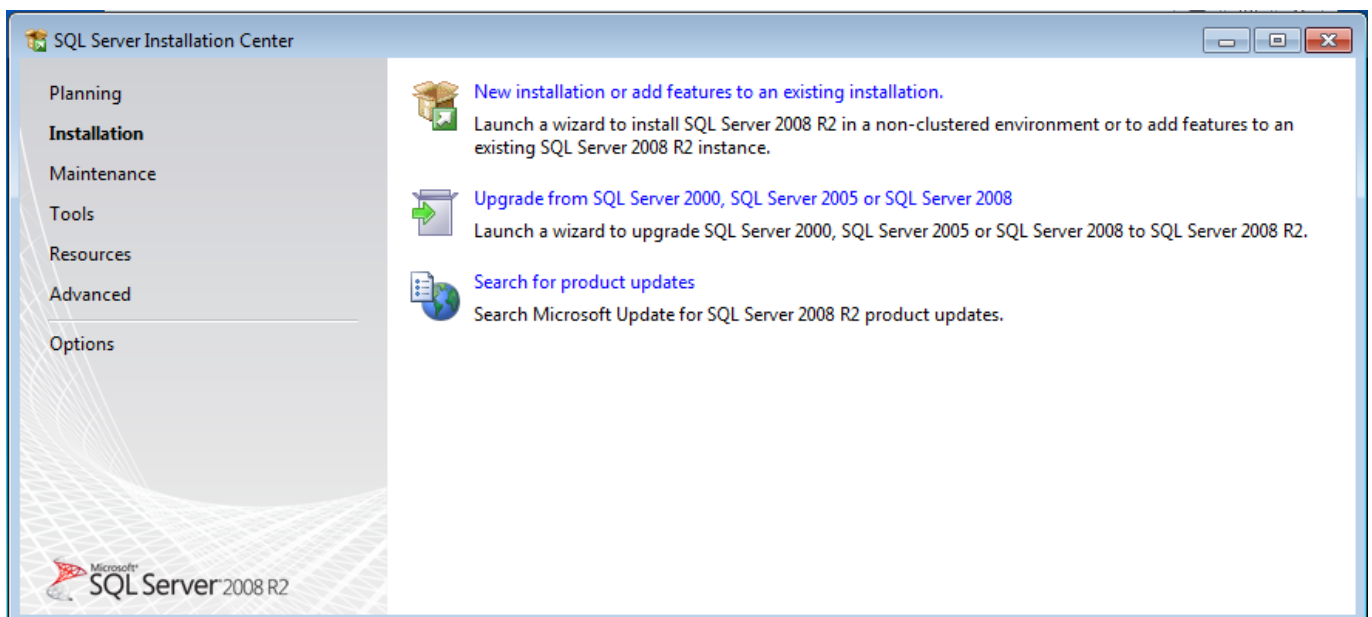
In this step, you are going to install Microsoft SQL 2008 R2 Express edition with Advanced Tools.

- Use the File Explorer
 - Go to Samba → Software library → SQL
 - Run **SQLEXPRT_x64_ENU_R2.exe**
-

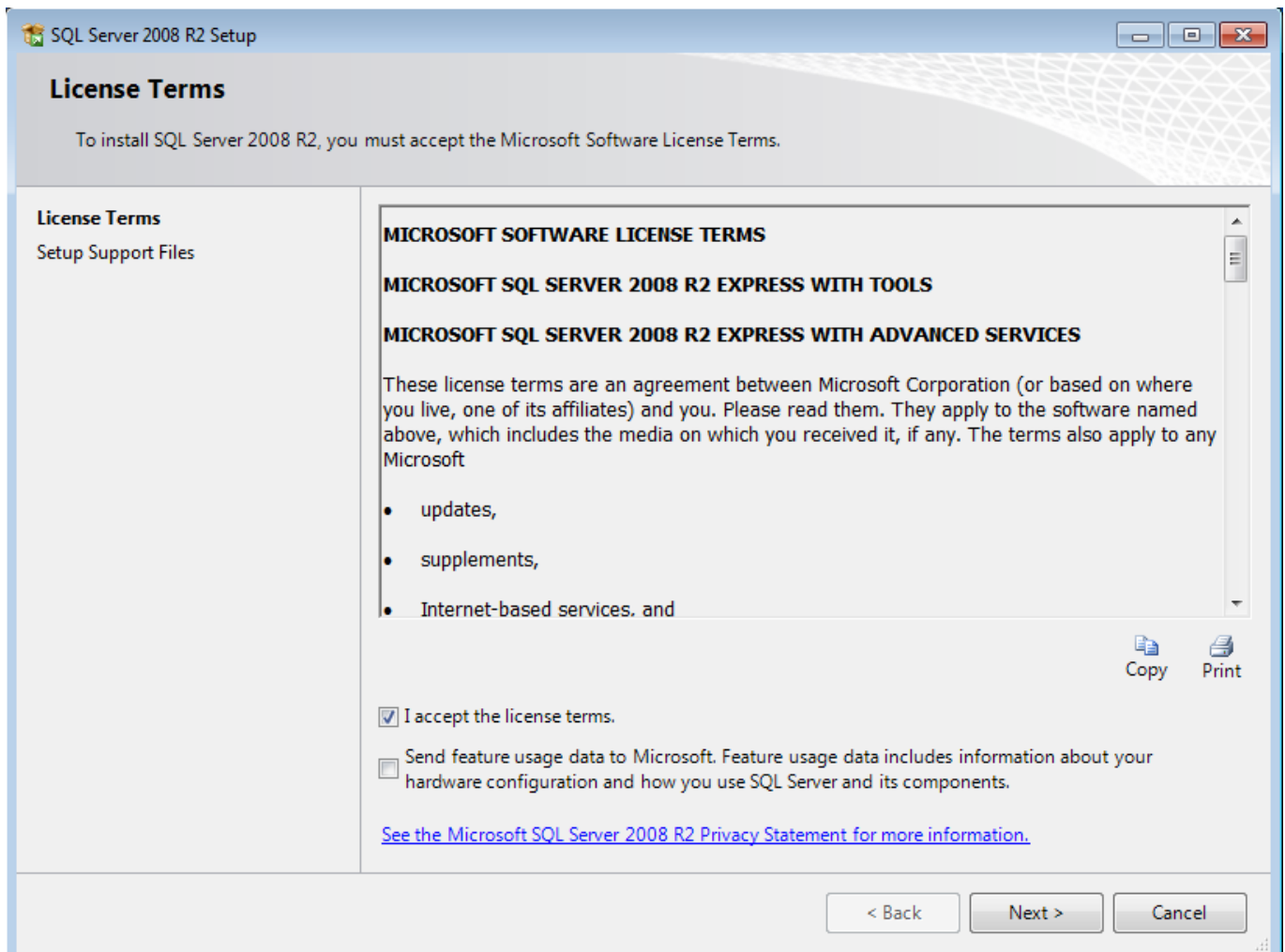
- Allow the setup program to make changes to your computer



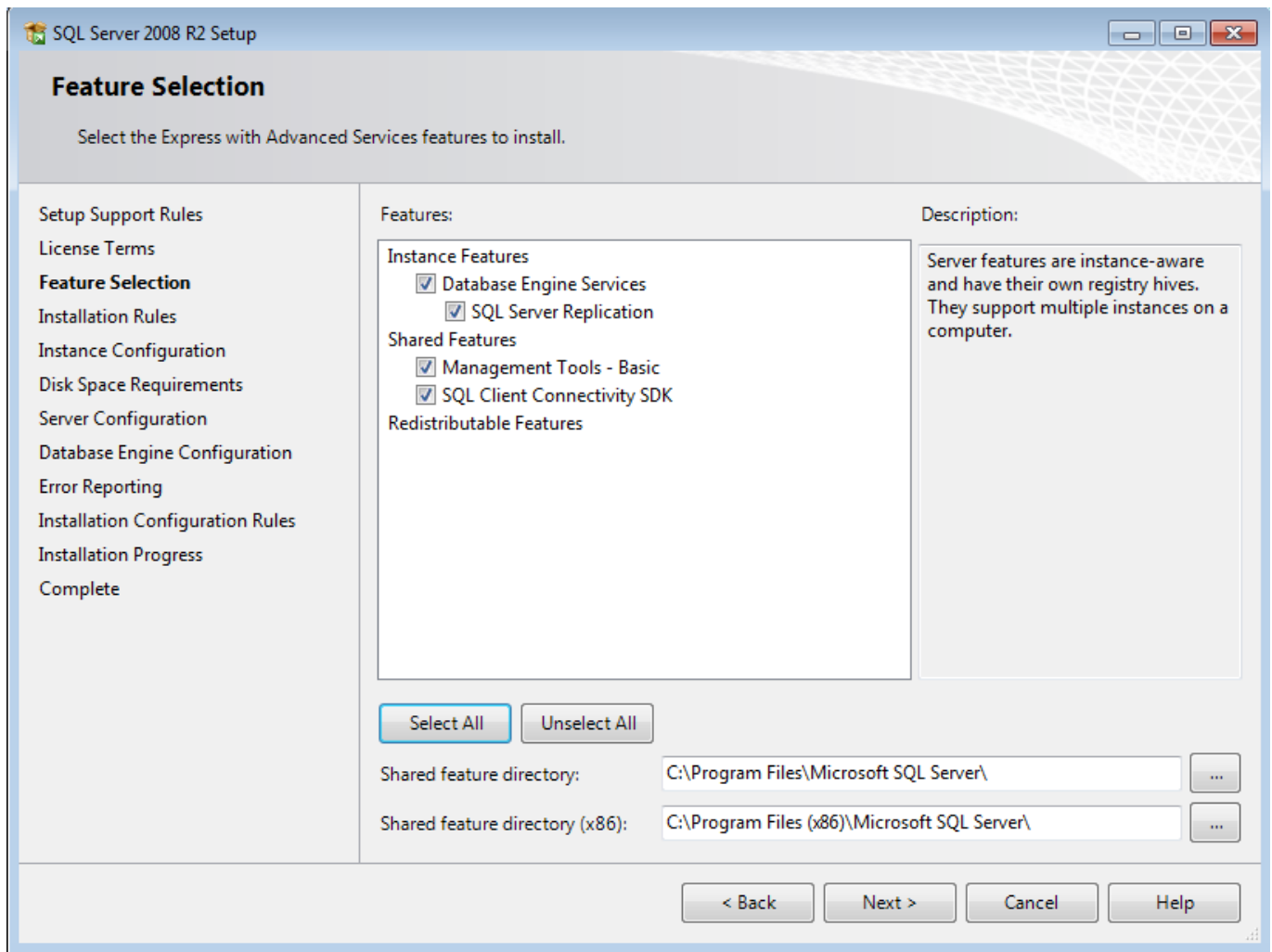
- Choose new installation



- Accept the license terms



- Select *All* features in the default shared feature directory



- Install MS SQL server as the default instance

The screenshot shows the 'Instance Configuration' window of the SQL Server 2008 R2 Setup. The window title is 'SQL Server 2008 R2 Setup'. The main heading is 'Instance Configuration'. Below the heading, it says 'Specify the name and instance ID for the instance of SQL Server. Instance ID becomes part of the installation path.'

On the left side, there is a list of setup steps: Setup Support Rules, License Terms, Feature Selection, Installation Rules, **Instance Configuration** (highlighted), Disk Space Requirements, Server Configuration, Database Engine Configuration, Error Reporting, Installation Configuration Rules, Installation Progress, and Complete.

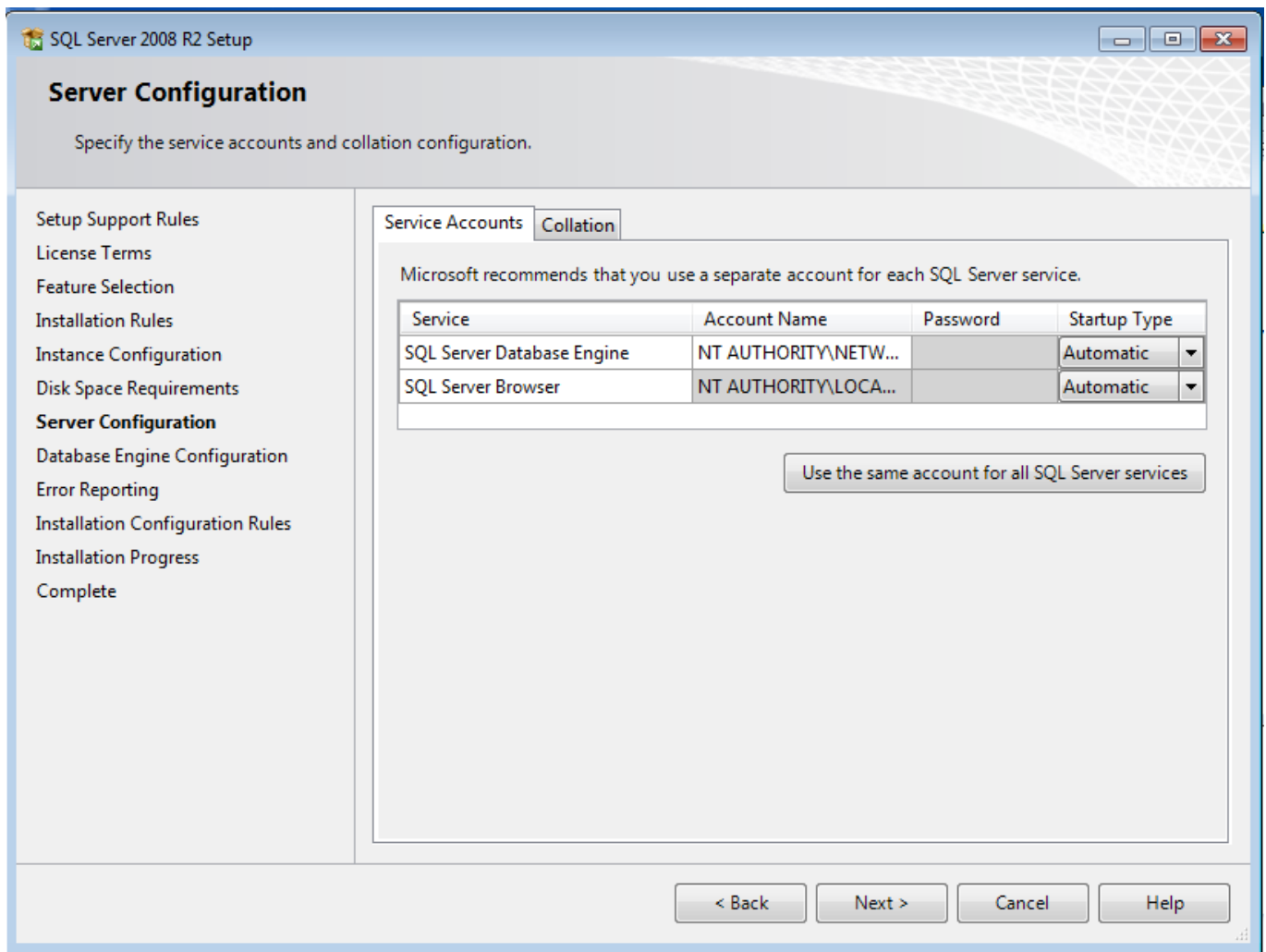
The main area contains the following configuration options:

- ☒ Default instance
- ☐ Named instance:
- Instance ID:
- Instance root directory: ...
- SQL Server directory:
- Installed instances:

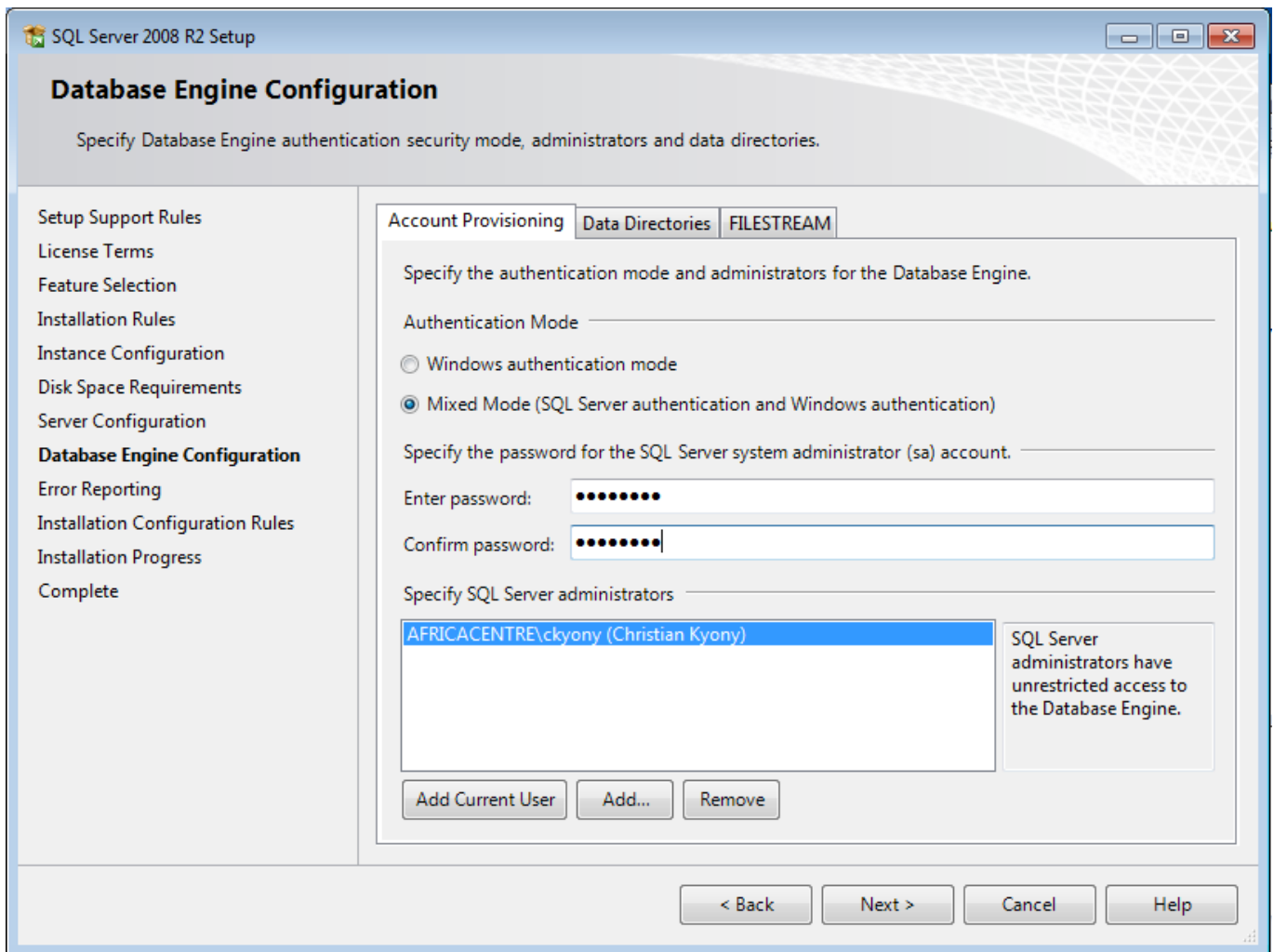
Below the 'Installed instances' label is a table with the following columns: Instance Name, Instance ID, Features, Edition, and Version. The table is currently empty.

At the bottom right, there are four buttons: '< Back', 'Next >', 'Cancel', and 'Help'.

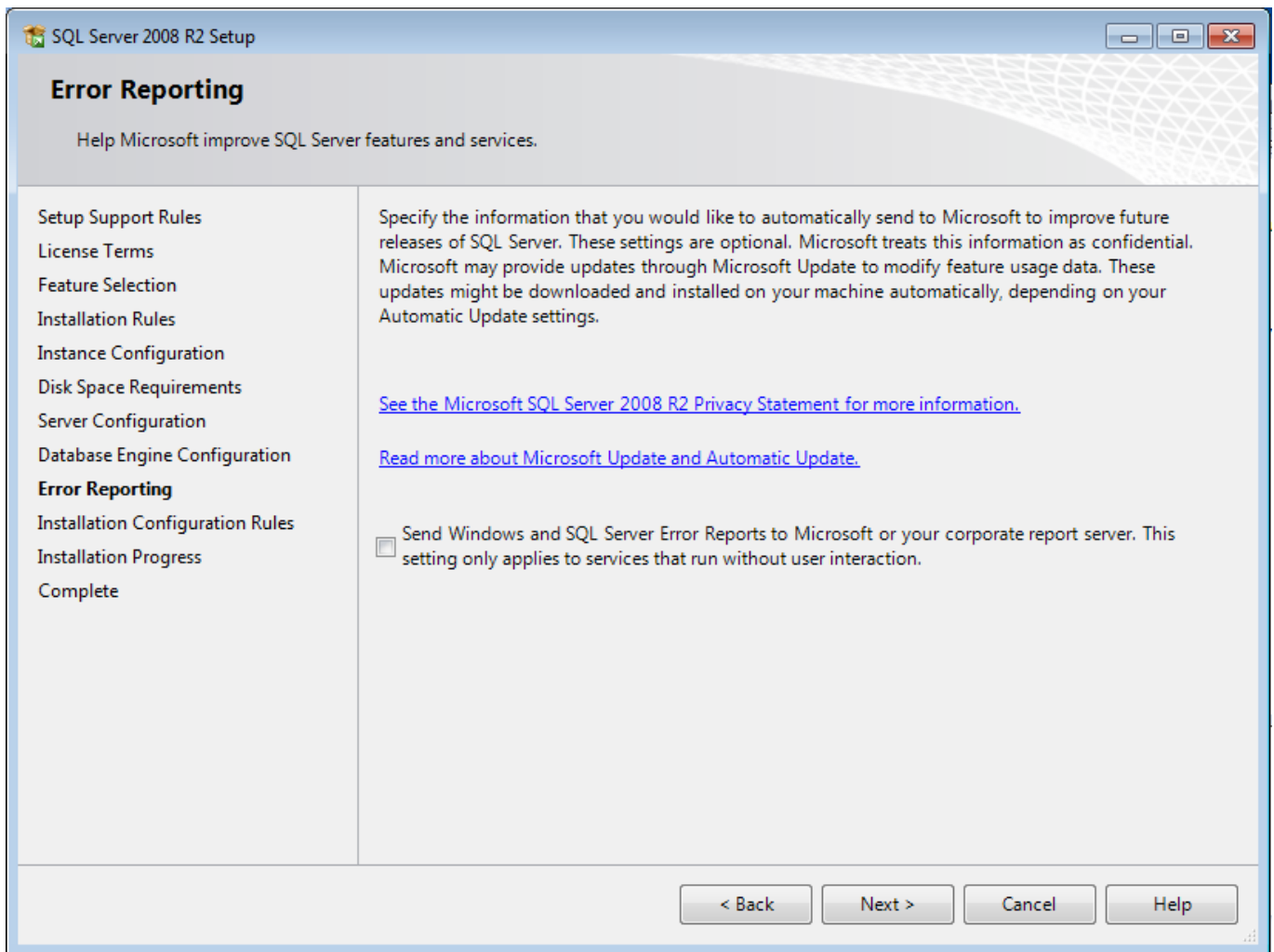
Use separate account for each SQL server service. Change the startup type to *automatic*



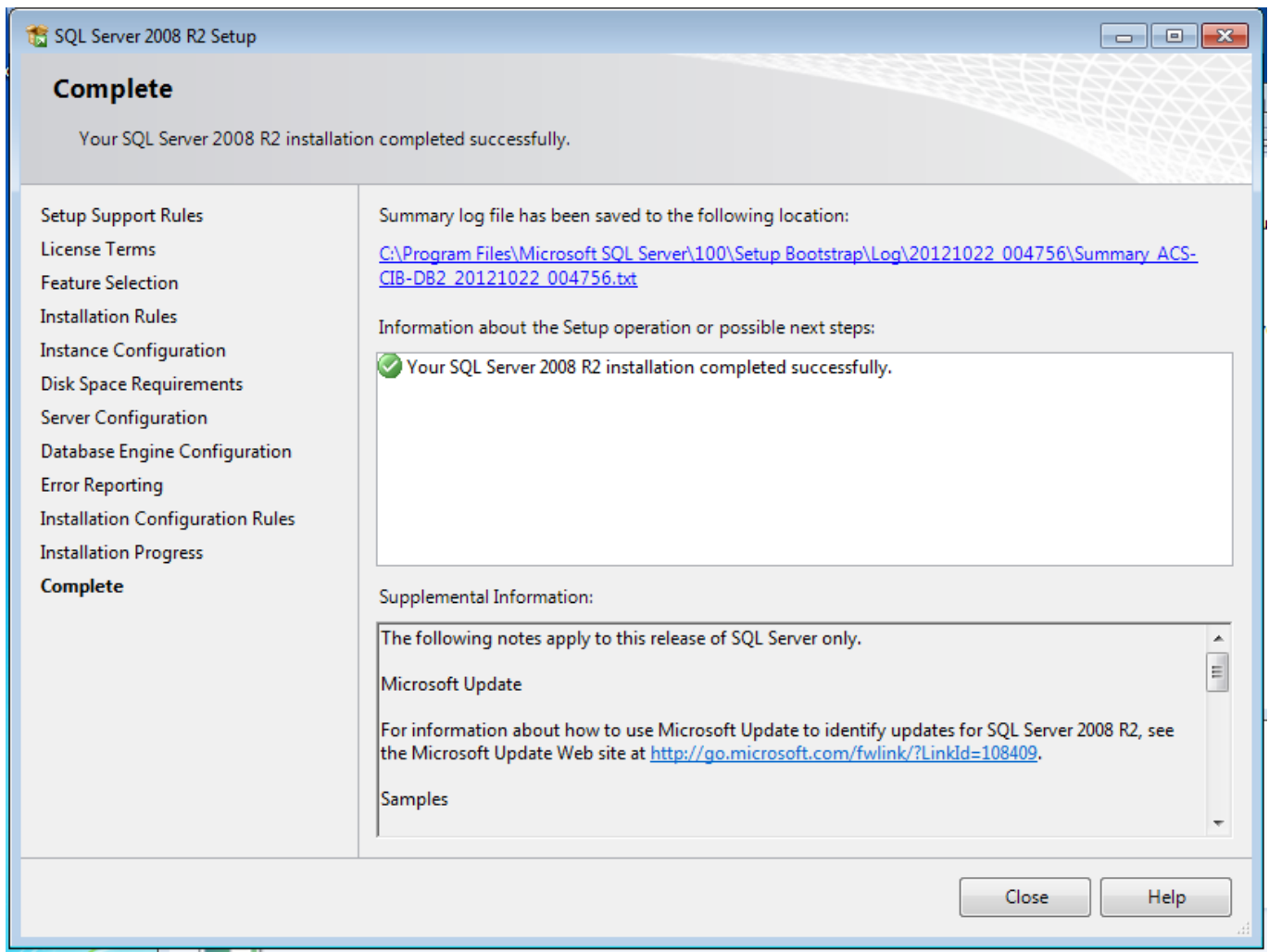
Choose *Mixed mode (SQL server authentication and Windows authentication)*. Specify the password for SQL Server system administrator. Add yourself as SQL server administrator.



gg Do not send Windows and SQL server Error Reports to Microsoft

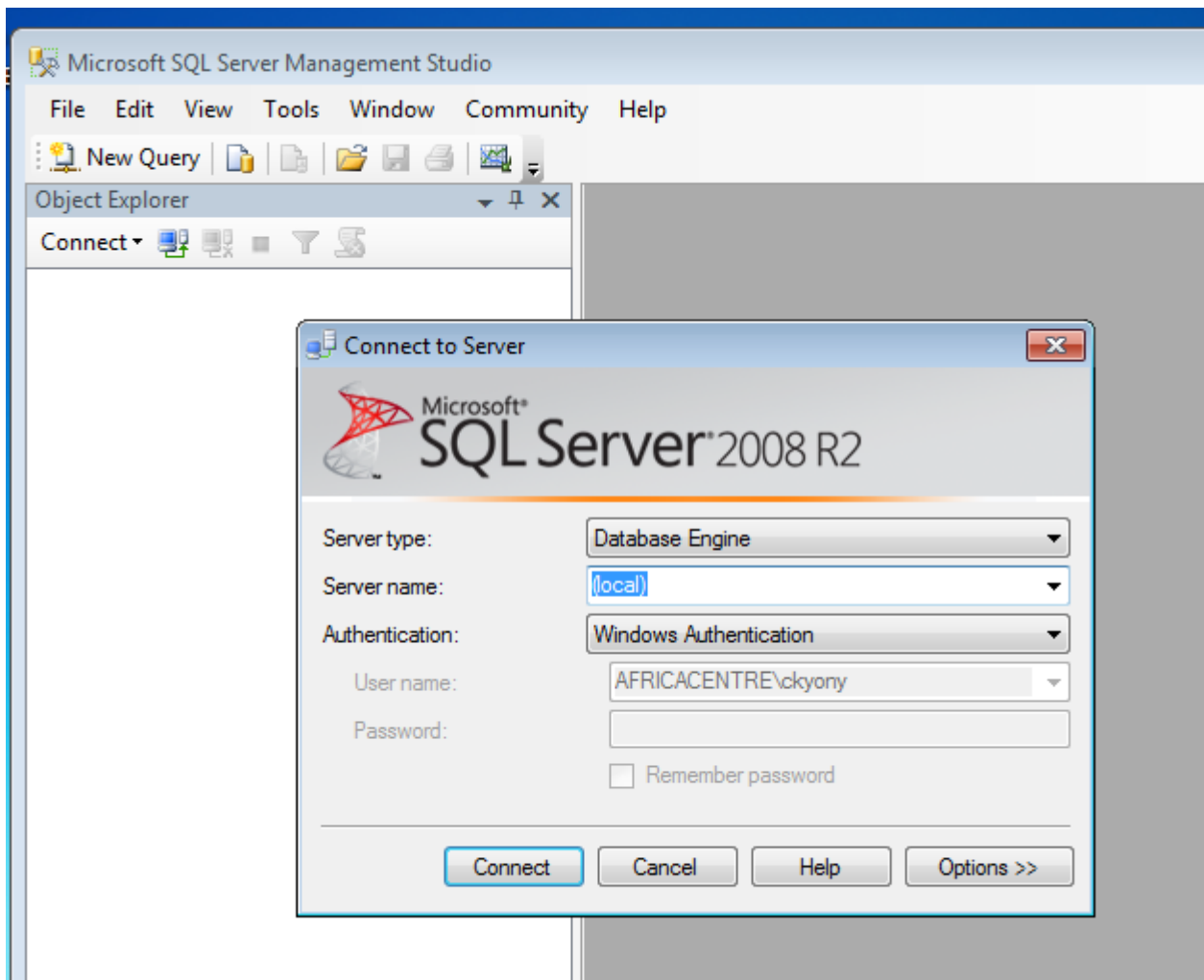


Click *Close* when the installation



To test that SQL server has been correctly installed, open Microsoft SQL server management studio for the first time

- Go to Start → All Applications → Microsoft SQL Server 2008 R2 → SQL server Management Studio



Enter your username and password if you are using the mixed authentication mode.

Click **Connect**.

7.2.3 Setup remote access

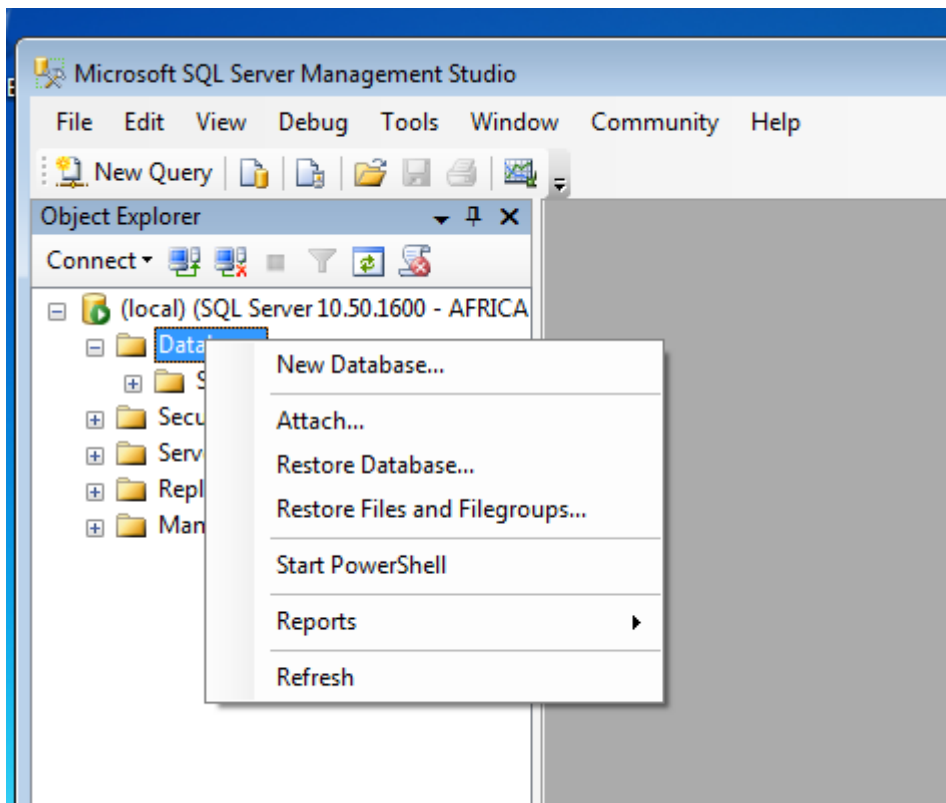
In this step, you are going to setup MS SQL 2008 to accept remote connections

7.2.4 Install the analytical database

In this step, you are going to install your **analytical** database from a backup of the operational database.

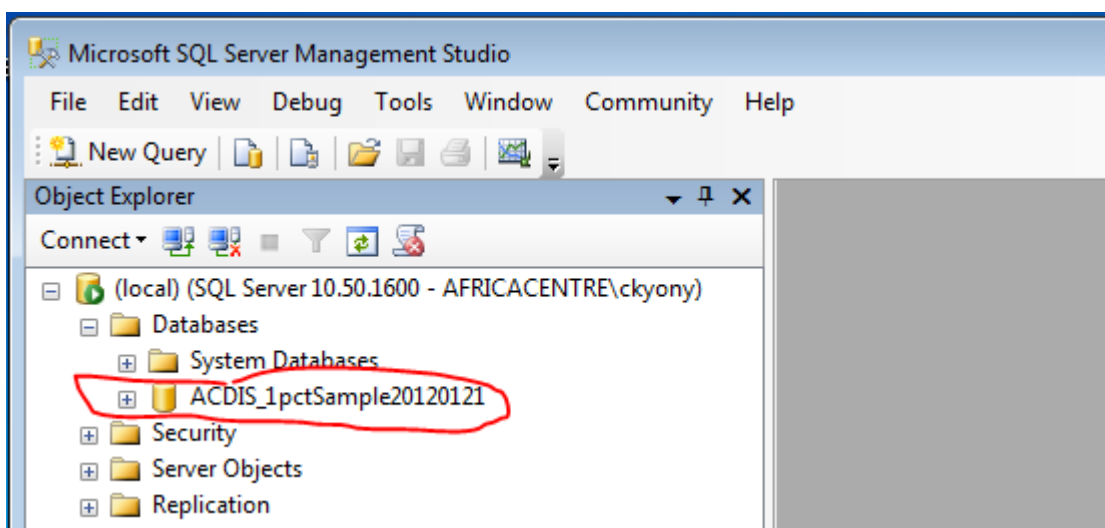
Here goes the definition of the analytical database versus operational database.

- Open Microsoft SQL Server Management Studio
- In the object Explorer, select **Database**.
- Right-click
- Choose **Restore database ...**



- Type the name of the new analytical database
- Choose **From device** as the source of the backup set
- Browse to the folder
- Select the backup sets to restore. Don't forget to check the box in the **Restore** column

After , you should see the screen shot below



You see the new analytical database in Microsoft SQL Server Management Studio

7.3 MySql on Ubuntu 12.08

In this step, you are going to setup a mysql database on Ubuntu server

7.3.1 Setup the virtual machine

Using the instructions in Section 4.2, “[Create a virtual machine](#)” and Section 4.4, “[Install Linux Ubuntu as a guest operating system](#)”, setup a new VM with the following specifications:

- OS : Ubuntu 12.04 LTS
- CPU : 2-core
- RAM : 4 GB
- IP Address : 172.18.21.229
- Host Name : database.cib.<code site>

7.3.2 Install mysql

To install MySql 5.5

```
sudo apt-get install mysql-server
```

During the installation process you will be prompted to enter a password for the MySQL root user.

Once the installation is complete, the MySQL server should be started automatically.

You can run the following command from a terminal prompt to check whether the MySQL server is running:

```
sudo netstat -tap | grep mysql
```

When you run this command, you should see the following line or something similar:

```
tcp          0      0 localhost:mysql    *:*              LISTEN       2556/mysqld
```

If the server is not running correctly, you can type the following command to start it:

```
sudo service mysql restart
```

7.3.3 Configure

You can edit the `/etc/mysql/my.cnf` file to configure the basic settings — log file, port number, etc. For example, to configure MySQL to listen for connections from network hosts, change the `bind-address` directive to the server’s IP address:

```
bind-address            = 192.168.0.5
```

Replace 192.168.0.5 with the appropriate address.

After making a change to `/etc/mysql/my.cnf`, restart the MySQL daemon

```
sudo service mysql restart
```

If you would like to change the MySQL root password, in a terminal enter:

```
sudo dpkg-reconfigure mysql-server-5.5
```

The MySQL daemon will be stopped, and you will be prompted to enter a new password.

7.3.4 Resources

For more information on the configuration and administration of the MySQL database, consult the following resources:

1. Go to [MySQL Home Page](#)
2. Consult the full documentation both online and offline format from the [MySQL Developers portal](#)

Chapter 8

Data Manager desktop installation

8.1 Role

The primary role of this desktop is to host the tools needed by the Data Manager:

- Pentaho Data Integration also called *Kettle* for ETL
- Nesstar Publisher for DDI documentation
- Zotero for citation management

8.2 Setup the virtual machine

Using the instructions in Section 4.2, “[Create a virtual machine](#)” and Section 4.3, “[Install Windows 7 as a guest operating system](#)”, setup a new VM with the following specifications:

- OS : Windows 7 Professional
- CPU : 2
- RAM : 4 GB
- Storage : 30 GB
- IP Address : 172.18.20.xxx
- Host Name: dm.cib.<code site>

8.3 Pentaho Kettle

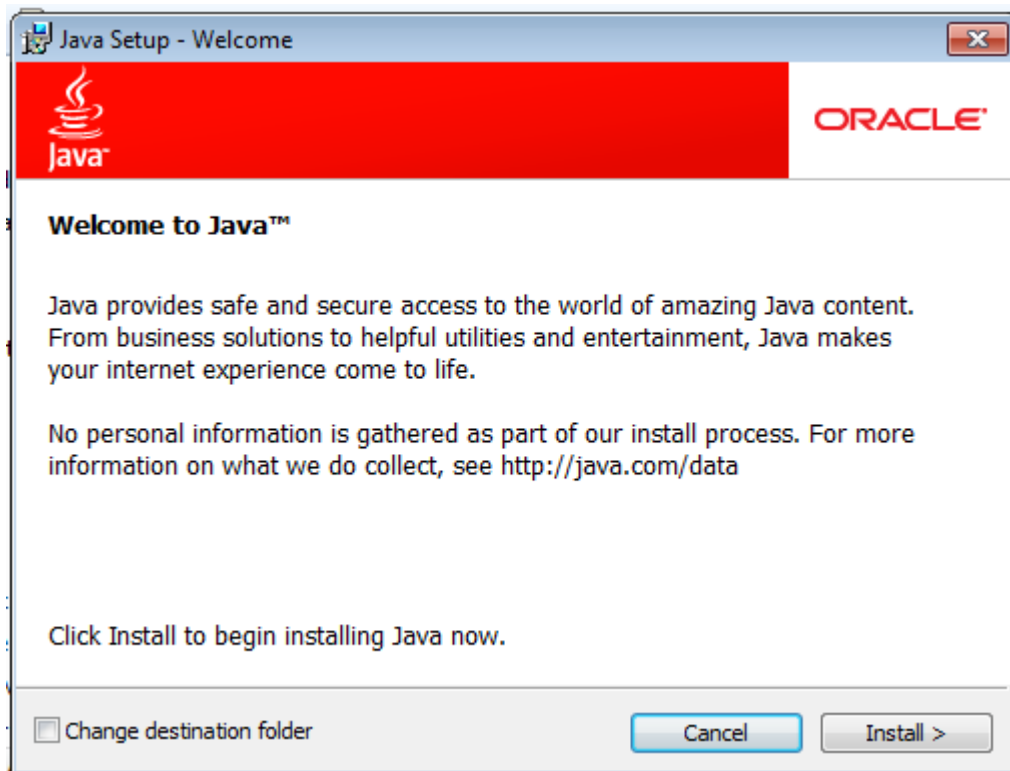
8.3.1 Install Java Runtime Environment

Kettle requires the Oracle Java Runtime Environment version 1.6.

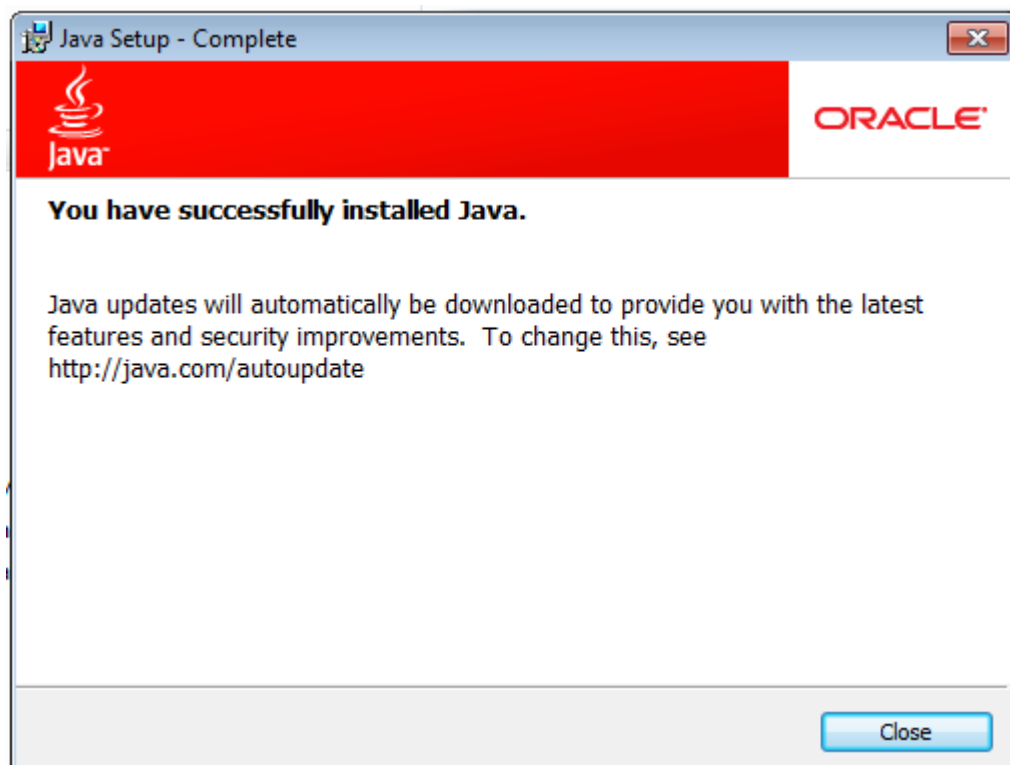
**Warning**

Kettle does not support Java 1.7.

To start the installation, double-click on the **jre-6u37-windows-x64.exe** executable. You should see the Welcome screen below



- Do not change the default installation folder.
- Click **Install**.
- Wait until you see the screen shot below



- Click **Close** to complete the installation.

8.3.2 Install Pentaho Kettle Community edition

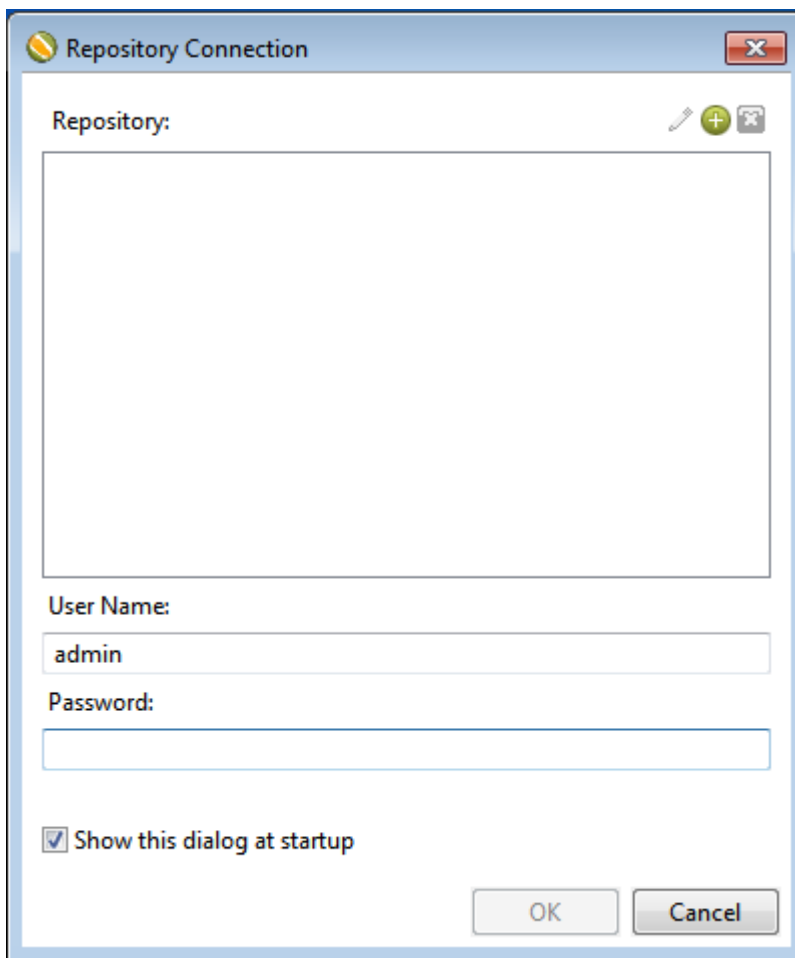
In this step, you are going to install the Pentaho Kettle application, create your first repository, connect to the database server and create your first job and transformation

- Create **C:\Pentaho** folder .
- Download **pdi-ce-4.3.0-stable.zip** from the Archive server into the **C:\Pentaho** folder.
- Unzip the zip file in the **C:\Pentaho\pdi-ce-4.3.0-stable.zip*** folder.
- To start the application, run the **Spoon.bat** batch file in the **C:\Pentaho\data-integration** folder.

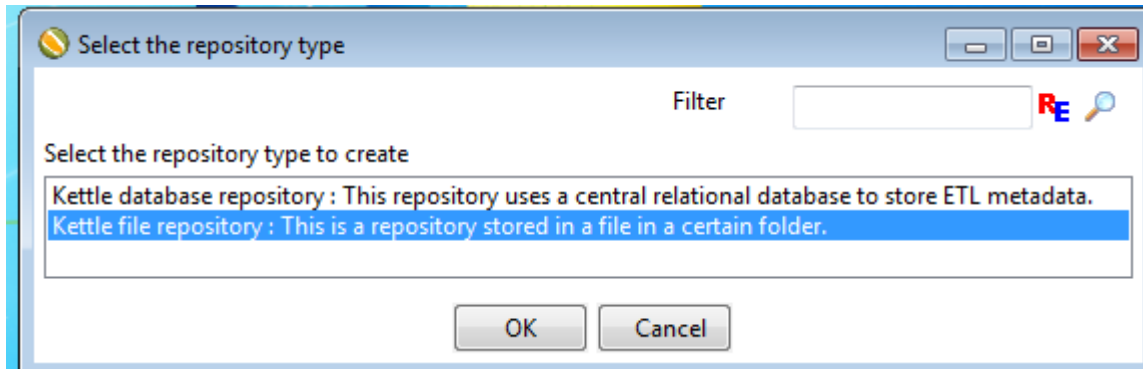
Tip

Create a shortcut on your desktop for the **Spoon client**. It will be easy to launch the application from there.

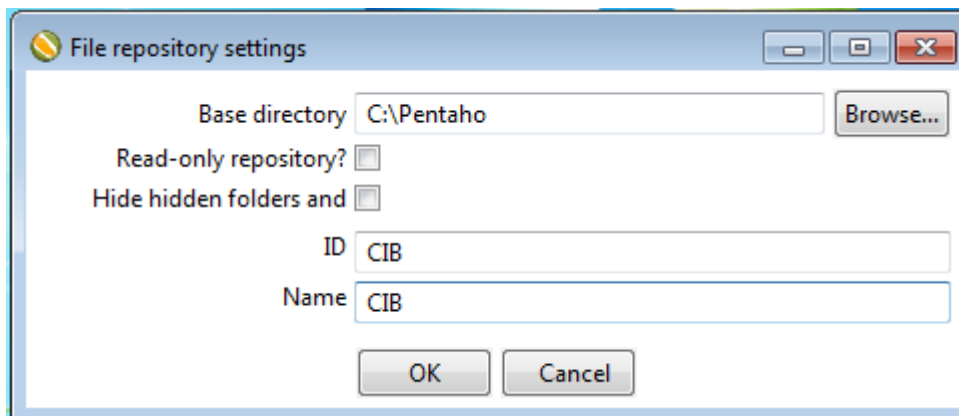
- The first time you launch, you will be ask to create your first repository.



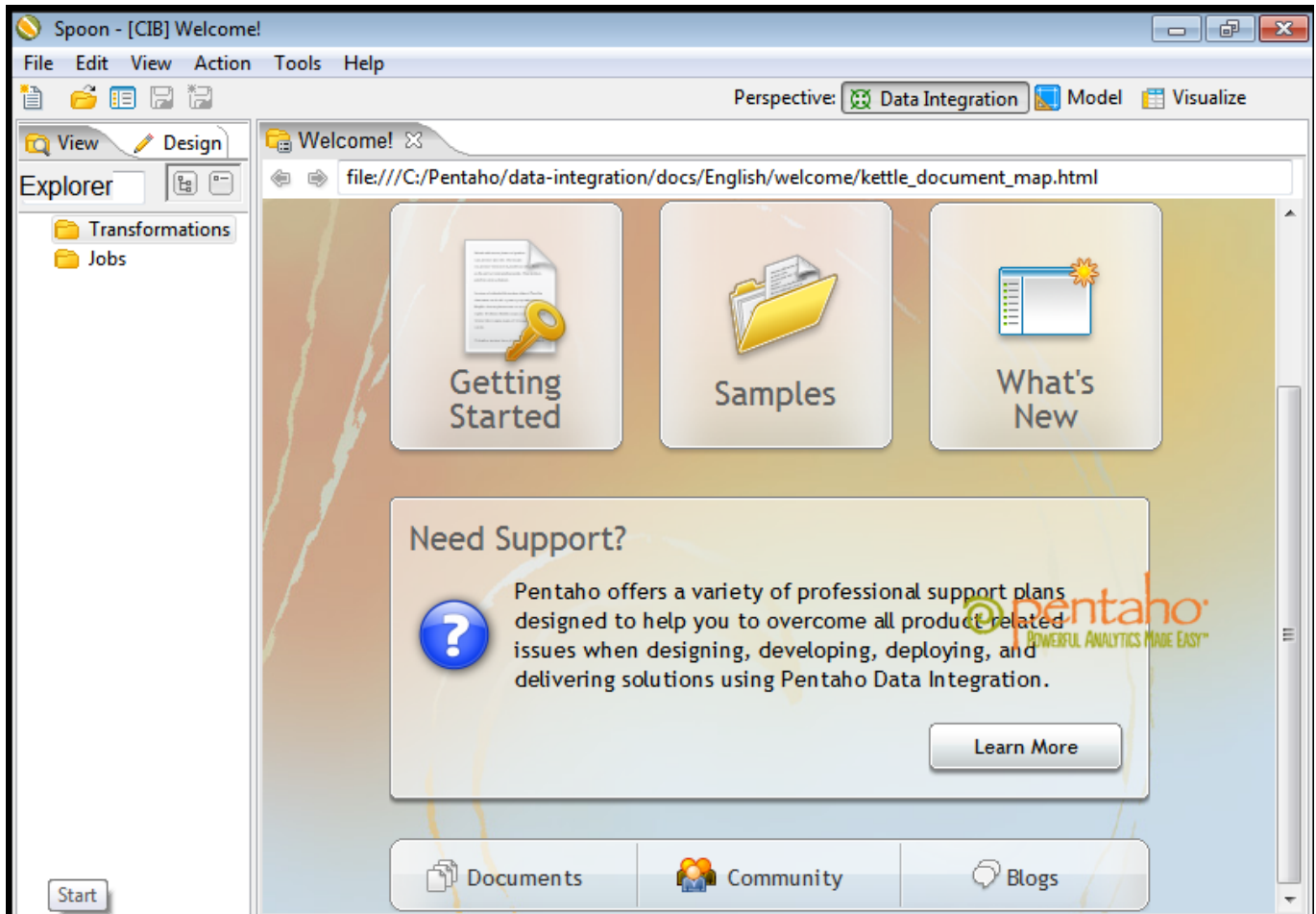
- To add a new repository, click on the green button on the top right of the screen.
 - Select **Kettle file repository** as per the dialog box below.
-



- To enter the base directory, browse to **C:\Pentaho**. Enter **CIB** for the **ID** and the **Name** of the repository.



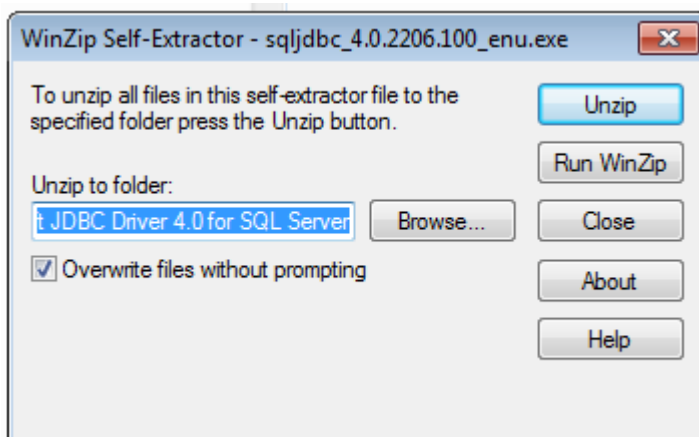
- Click **OK** and you should see the Spoon environment.



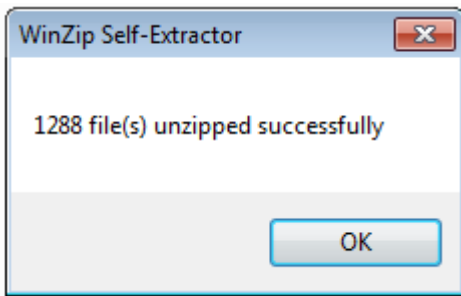
8.3.3 Install a native JDBC driver for MS SQL server

If you are using a Microsoft SQL server 2012, 2008 R2, 2008, 2005 or Azure, you need to install a native JDBC driver in the correct directory.

- Download the **sqljdbc_4.0.2206.100_enu.exe** driver from the archive server to the temporary folder **C:\\temp**.
- Run **sqljdbc_4.0.2206.100_enu.exe**.
- When prompted for an installation directory, enter **C:\\Pentaho\\Microsoft Driver 4.0 for SQL server**.



You will see the following screen:



The next step is to copy the driver in the Pentaho installation directory.

- Go to **C:\Pentaho\Microsoft JDBC Driver 4.0 for SQL Server\sqljdbc_4.0\enu**
- Copy the **sqljdbc4.jar** file to the **{pentaho-install-dir}\data-integration\libext\JDBC**

8.3.3.1 Setup a connection to the analytical database

The final step is to create a database connection to your analytical database.

- Go to **File** → **New** → **Database Connection**.

Tip

You may also click on the (New) icon below the **File** menu.

- Enter the following information:
 - Connection Name: express
 - Host name:
 - Database name:
 - Instance Name:
 - Port Number:
 - User name: kettle
 - Password: T3amw0rk

Tip

You must choose **SQL authentication** for the **JDBC(native)** driver on MS SQL Server.

Database Connection

General
Advanced
Options
Pooling
Clustering

Connection Name:
cib-db

Connection Type:
Intersystems Cache
KingbaseES
LucidDB
MS Access
MS SQL Server
MS SQL Server (Native)
MaxDB (SAP DB)
MonetDB
MySQL
Neoview
Nettezza
Oracle
Oracle RDB
Palo MOLAP Server
PostgreSQL
Remedy Action Request System

Access:
Native (JDBC)
ODBC
JNDI

Settings
Host Name: db-cib
Database Name: tutorial
Instance Name:
Port Number: 1433
User Name: kettle
Password:
☐ Use .. to Separate Schema and Table

Test Feature List Explore

OK Cancel

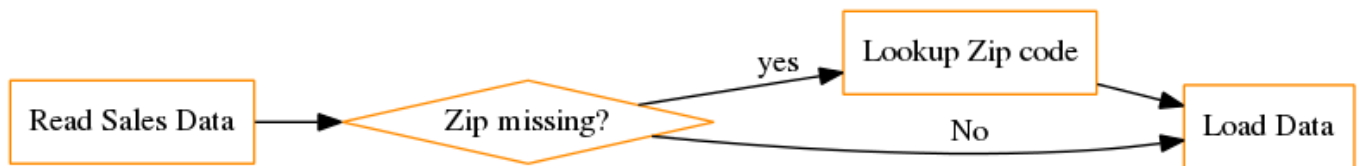
- Click **Test** to make sure your entries are correct. A success message appears.
- Click **OK**, to exit the Database Connections dialog box.

8.3.4 Build your first transformation

In this step, you are going to build your first transformation. Transformations are used to describe the data flows for ETL such as reading from a source, transforming data and loading into a target location.

In this scenario, you are going to load a flat file (CSV) of sales data into a database so that mailing lists can be generated. Several of the customer records are missing postal codes that must be resolved before loading into the database.

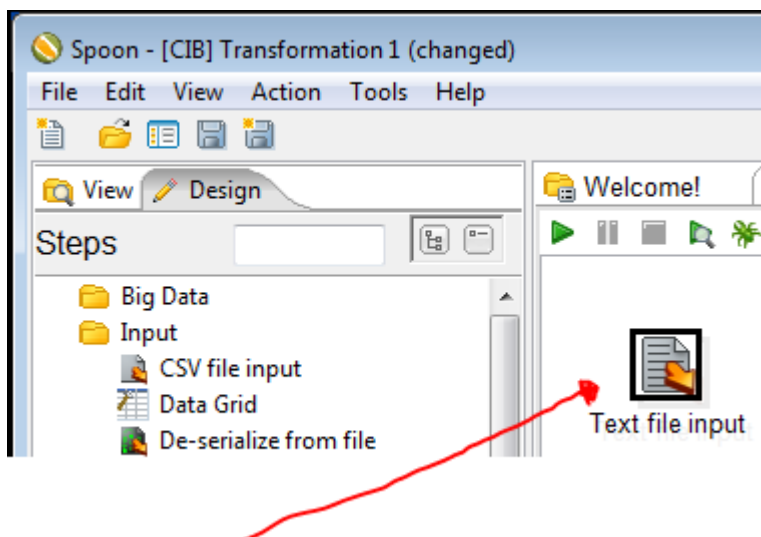
The logic looks like this:



8.3.4.1 Retrieve data from a flat file

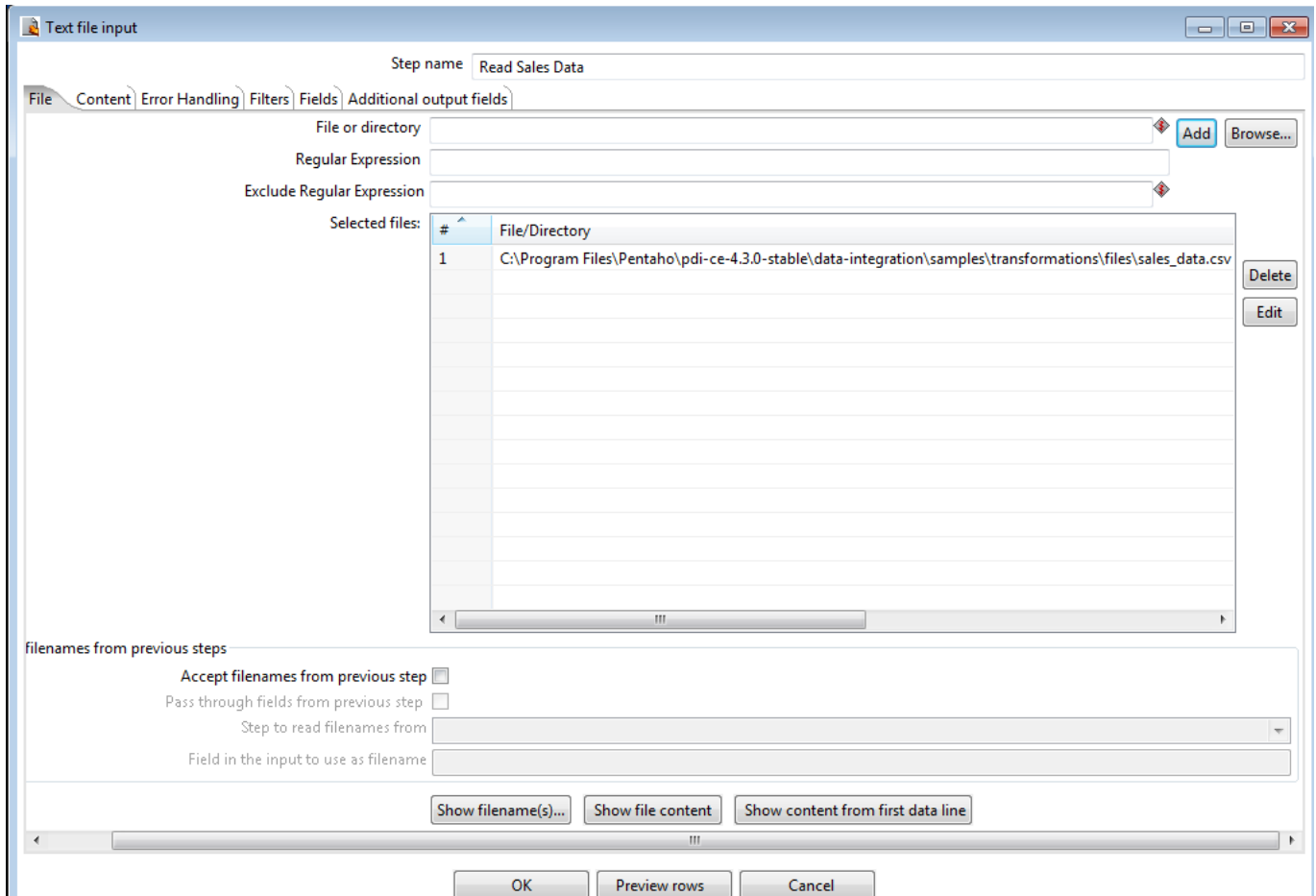
Follow the instructions below to retrieve data from a flat file.

- Click **New** in the upper left corner of the Spoon graphical interface.
- Select **Transformation** from the list
- Under the **Design** tab, expand the **Input** node; then, select and drag a **Text File Input** step onto the canvas on the right.



- Double-click on the **Text File input** step.

The edit properties dialog box associated with the Text File input step appears. In this dialog box, you specify the properties related to a particular step.



- In the **Step Name** field, type **Read Sales Data**.
- Click **Browse** to locate the source file, **sales_data.csv**, available at **... \design-tools\data-integration\samples\transformations\files**.

The path to the source file appears in the File or directory field.

Click **Add**.

The path to the file appears under **Selected Files**. You can look at the contents of the file by clicking the **Show file content** content to determine things such as how the input file is delimited, what enclosure character is used, and whether or not a header row is present.

In the example, the input file is comma (,) delimited, the enclosure character being a quotation mark (") and it contains a single header row containing field names.

- Click the **Content** tab. The fields under the **Content** tab allow you to define how your data is formatted.
- Make sure that the **Separator** is set to comma (,) and that the **Enclosure** is set to quotation mark ("). Enable **Header** because there is one line of header rows in the file.

Text file input

Step name: Read Sales Data

File | Content | Error Handling | Filters | Fields | Additional output fields

Filetype: CSV

Separator: ,

Enclosure: "

Allow breaks in enclosed fields? ☐

Escape:

Header ☒ Number of header lines: 1

Footer ☐ Number of footer lines: 1

Wrapped lines? ☐ Number of times wrapped: 1

Paged layout (printout)? ☐ Number lines per page: 80

- Click the **Fields** tab and click **Get Fields** to retrieve the input fields from your source file.

Text file input

Step name: Read Sales Data

File | Content | Error Handling | Filters | Fields | Additional output fields

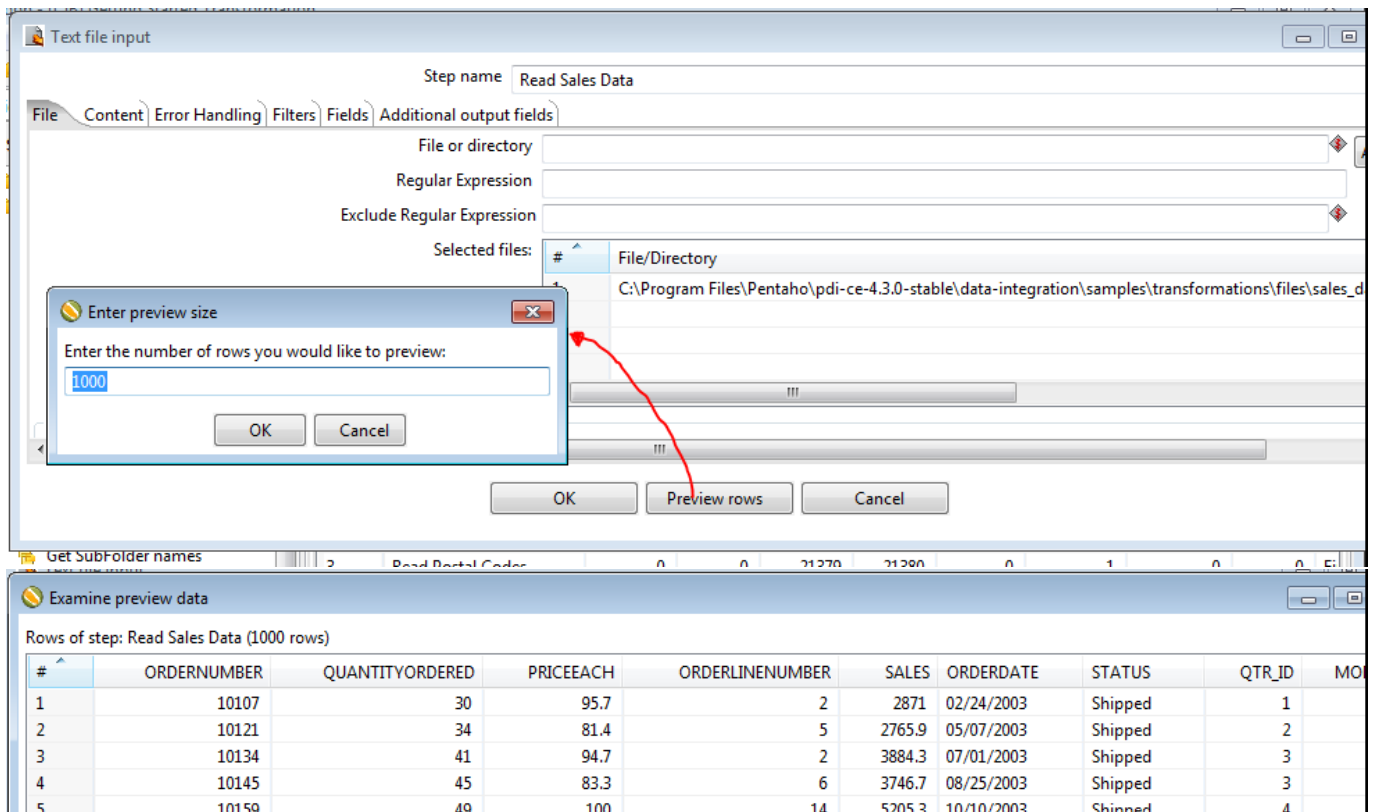
#	Name	Type	Format	Position	Length	Precision	Currency	Decimal
1	ORDERNUMBER	Integer	#		15		\$.
2	QUANTITYORDERED	Integer	#		15		\$.
3	PRICEEACH	Number	#,.		15		\$.
4	ORDERLINENUMBER	Integer	#		15		\$.
5	SALES	Number	#,.		15		\$.
6	ORDERDATE	Date	MM/dd/yyyy				\$.

Get Fields

OK Preview rows Cancel

A dialog box appears requesting that you to specify the number of lines to scan, allowing you to determine default settings for the fields such as their format, length, and precision. Type **0** (zero) in the **Number of Sample Lines** text box to scan all lines. By scanning all lines, you ensure that Pentaho Data Integration has read the entire contents of the file and you reduce the possibility of errors that may cause a transformation not to run. Click **OK** and the summary of the scan results appears. Once you are done examining the scan results, click Close to return to the step properties editor.

- Click **Preview Rows** to verify that your file is being read correctly. You can change the number of rows to preview. Click **OK** to exit the step properties dialog box.



- Save your transformation as shown in Section 8.3.4.2, “[Save your transformation](#)”.

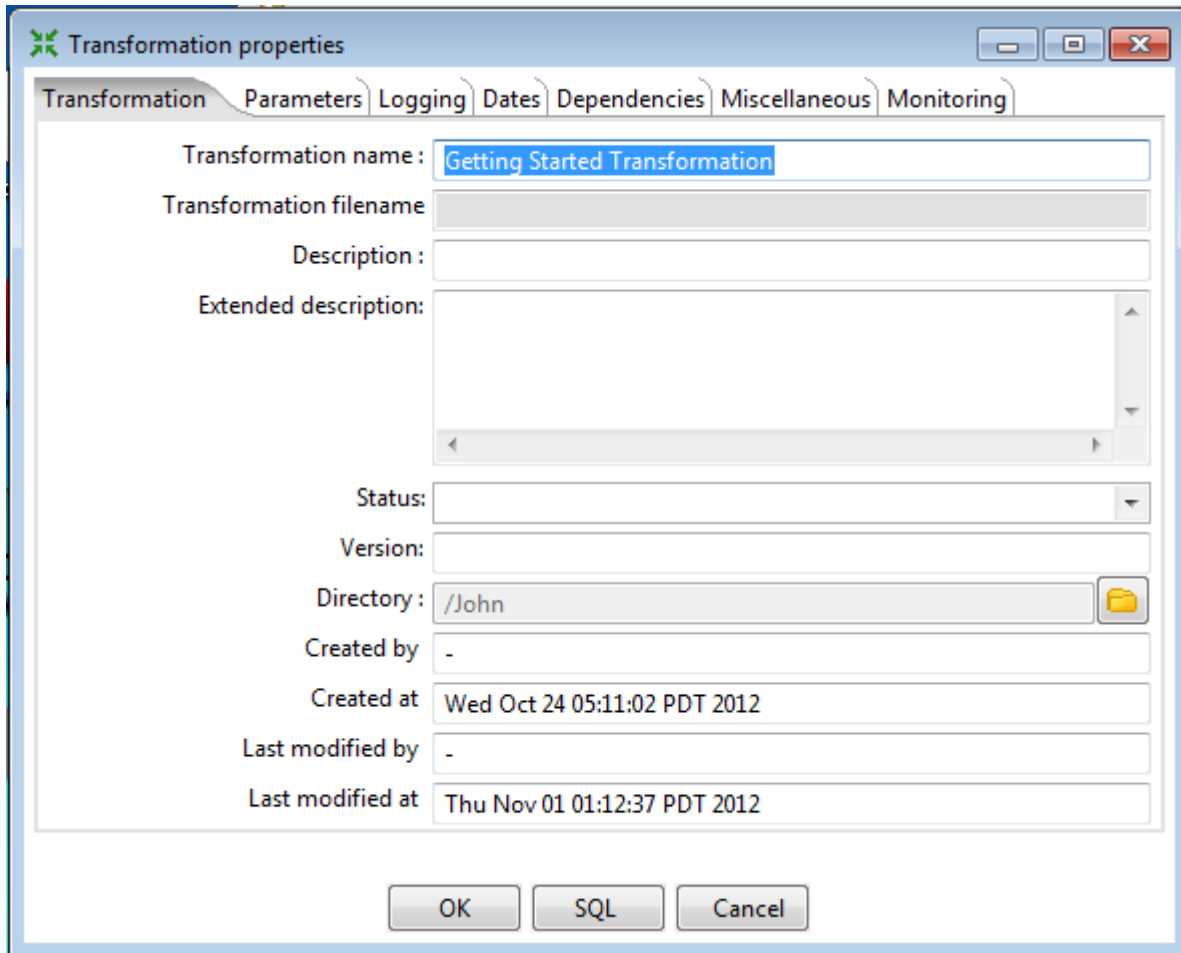
8.3.4.2 Save your transformation

Follow the instructions below to save your transformation.

- In the Spoon designer, click **File** → **Save As**.

The Transformation Properties dialog box appears.

- In the **Transformation Name** field, type **Getting Started Transformation**.



The image shows a 'Transformation properties' dialog box with several tabs: Transformation, Parameters, Logging, Dates, Dependencies, Miscellaneous, and Monitoring. The 'Transformation' tab is active. It contains the following fields:

- Transformation name:** Getting Started Transformation
- Transformation filename:** (empty)
- Description:** (empty)
- Extended description:** (empty text area)
- Status:** (dropdown menu)
- Version:** (empty)
- Directory:** /John (with a folder icon button)
- Created by:** -
- Created at:** Wed Oct 24 05:11:02 PDT 2012
- Last modified by:** -
- Last modified at:** Thu Nov 01 01:12:37 PDT 2012

At the bottom are three buttons: OK, SQL, and Cancel.

- In the **Directory** field, click the folder icon to select a repository folder where you will save your transformation.
- Expand the Home directory and double-click the joe folder.

Your transformation will be stored in the **tutorial** folder in the CIB Repository.

- Click **OK** to exit the **Transformation Properties** dialog box.

The **Enter Comment** dialog box appears.

- Click in the **Enter Comment** dialog box and press **Delete** to remove the default text string. Type a meaningful comment about your transformation.

The comment and your transformation are tracked for version control purposes in the Repository.

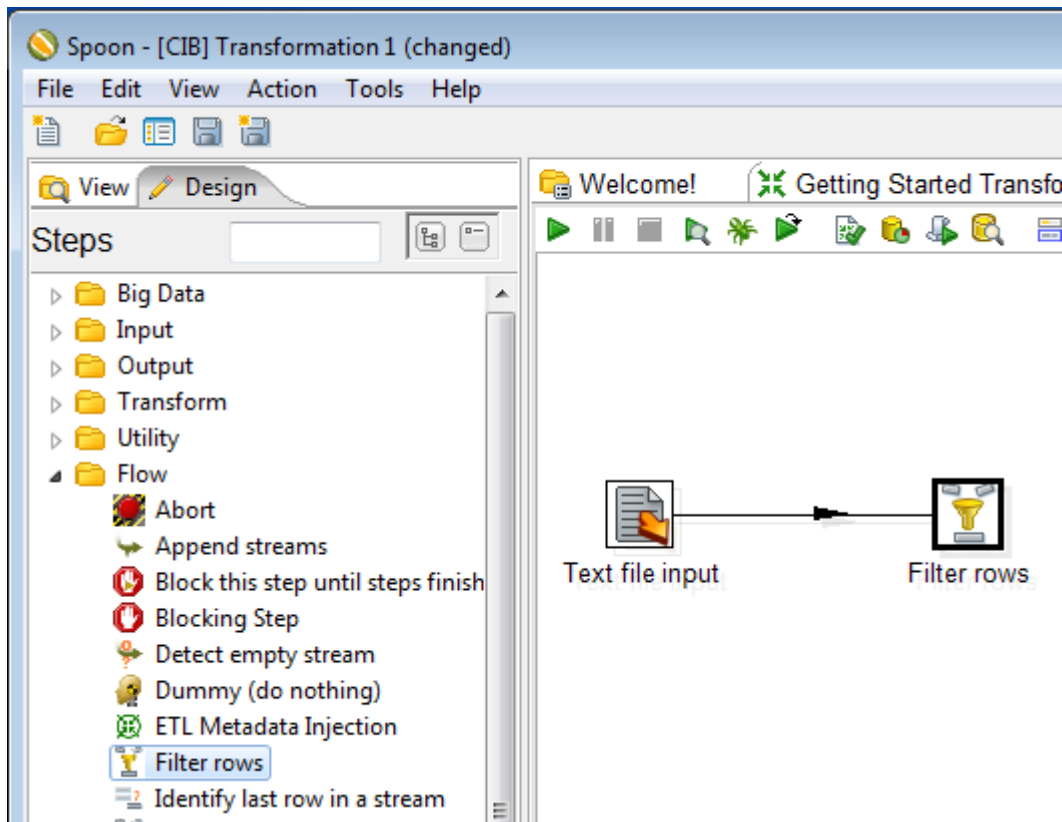
- Click **OK** to exit the Enter Comment dialog box.

8.3.4.3 Filter records with missing postal codes

The source file contains several records that are missing postal codes. You will now use the Filter Rows transformation step to separate out those records so that you can resolve them in a later step.

- Add a **Filter Rows** step to your transformation. Under the **Design** tab, go to **Flow** → **Filter Rows**.

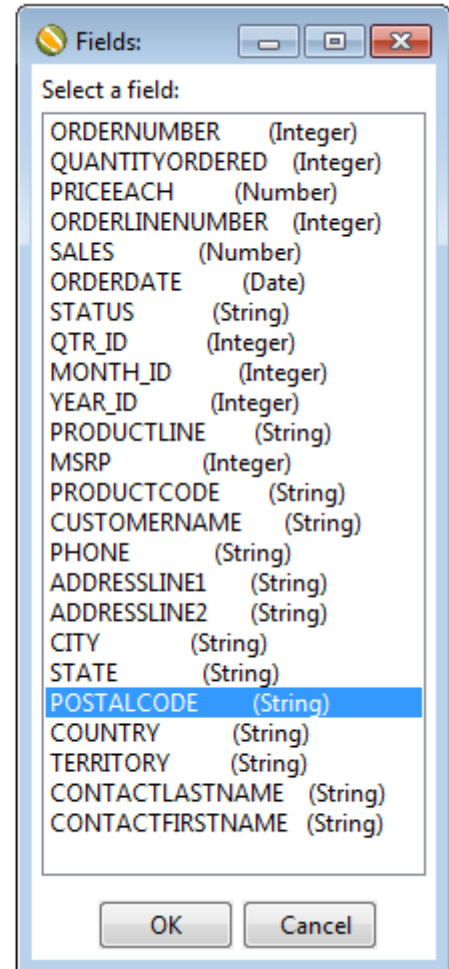
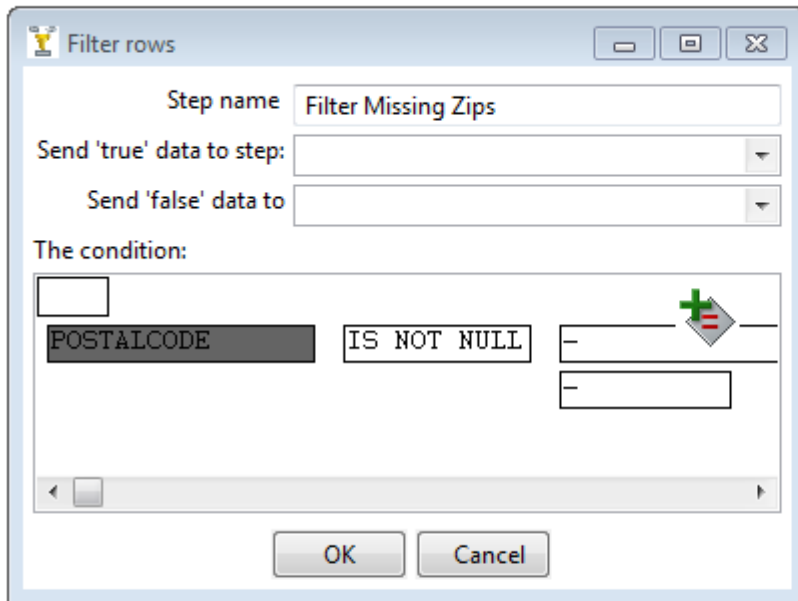
- Create a "hop" between the **Read Sales Data** (Text File Input) step and the **Filter Rows** step. Hops are used to describe the flow of data in your transformation. To create the hop, click the **Read Sales Data** (Text File input) step, then press the <SHIFT> key down and draw a line to the Filter Rows step.



- Double-click the **Filter Rows** step.

The Filter Rows edit properties dialog box appears.

- In the **Step Name** field type, **Filter Missing Zips**.
- Under **The condition**, click <field>. A dialog box that contains the fields you can use to create your condition appears.

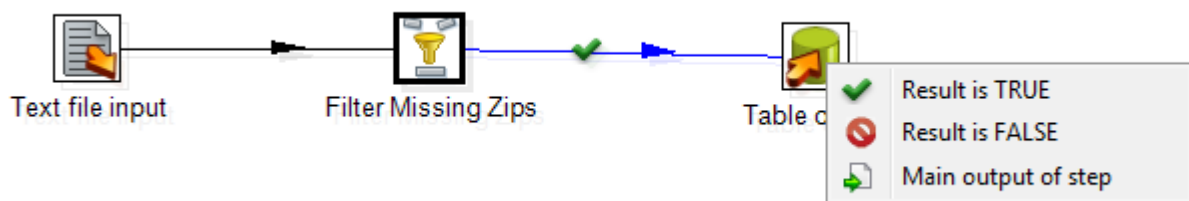


- In the **Fields:** dialog box select **POSTALCODE** and click **OK**.
- Click on the comparison operator (set to = by default) and select the **IS NOT NULL** function and click OK. Click OK to exit the Filter Rows properties dialog box.
- Save your transformation as per instructions in Section 8.3.4.2, “[Save your transformation](#)”.

8.3.4.4 Loading your data into a relational database

In this step, you will take all records exiting the Filter rows step where the POSTAL CODE was not null (the **true** condition) and load them into a database table.

- Under the Design tab, expand the contents of the **Output** node.
- Click and drag a **Table Output** step into your transformation; create a hop between the **Filter Missing Zips** (Filter Rows) and **Table Output** steps. Select **Result is TRUE**.



- Double-click the **Table Output** step to open its edit properties dialog box.

Table output

Step name: Write to Database

Connection: express [Edit...] [New...]

Target schema: dbo [Browse...]

Target table: Sales_data [Browse...]

Commit size: 1000

Truncate table: ☒

Ignore insert errors: ☐

Specify database fields: ☐

Main options | Database fields

Partition data over tables: ☐

Partitioning field: []

Partition data per month: ☒

Partition data per day: ☐

Use batch update for inserts: ☒

Is the name of the table defined in a field?: ☐

Field that contains name of table: []

Store the tablename field: ☒

Return auto-generated key: ☐

Name of auto-generated key field: []

[OK] [Cancel] [SQL]

- Rename your Table Output Step to **Write to Database**.
- In the **Connection**, select the connection configured in Section 8.3.3.1, “[Setup a connection to the analytical database](#)”.
- Type **SALES_DATA** in the **Target Table** text field.

This table does not exist in the target database. In the next steps you will generate the Data Definition Language (DDL) to create the table and execute it. DDL are the SQL commands that define the different structures in a database such as CREATE TABLE.

- In the **Write to Database** edit properties dialog box, enable the **Truncate Table** property.
- Click **SQL** to generate the DDL for creating your target table.
- Click **Execute** to run the SQL.

A results dialog box appears indicating that one SQL statement was executed. Click **OK** close the execution dialog box. Click **Close** to close the Simple SQL editor dialog box. Click **OK** to close the Table Output edit properties dialog box.

- Save your transformation.

8.3.4.5 Retrieve data from the Lookup file

You have been provided a second text file containing a list of cities, states, and postal codes that you will now use to look up the postal codes for all of the records where they were missing (the 'false' branch of your Filter rows step). First, you will use a Text file input step to read from the source file, next you will use a Stream lookup step to bring the resolved Postal Codes into the stream.

- Add a new **Text File Input** step to your transformation. In this step you will retrieve the records from your lookup file.
- Rename your **Text File input** step to, **Read Postal Codes**.
- Click **Browse** to locate the source file, **Zipssortedbycitystate.csv**, located at ... \design-tools\data-integration\samples\transformat
- Click **Add**.

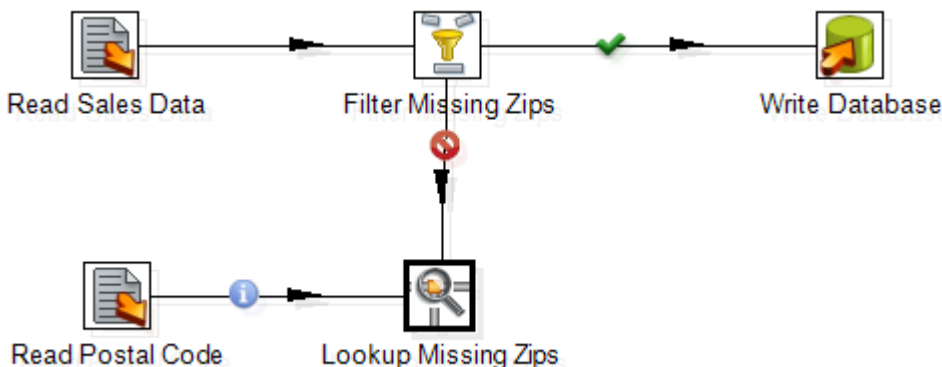
The path to the file appears under **Selected Files**.

- Under the **Content** tab, enable the **Header** option. Change the separator character to a comma (,). and confirm that the enclosure setting is correct.
- Under the **Fields** tab, click **Get Fields** to retrieve the data from your .csv file.
- Click **Preview Rows** to make sure your entries are correct and click **OK** to exit the Text File input properties dialog box.
- Save your transformation.

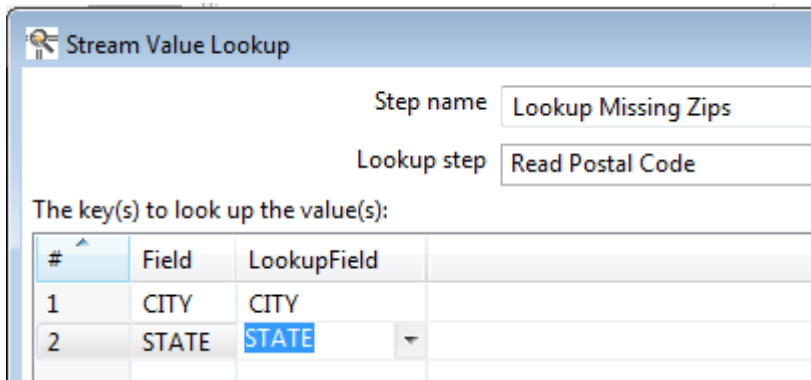
8.3.4.6 Resolving missing Zip Code information

In this exercise, you will begin to resolve the missing zip codes.

- Add a **Stream Lookup** step to your transformation. Under the Design tab, expand the **Lookup** folder and choose **Stream Lookup**.
- Draw a hop between the **Filter Missing Zips** (Filter rows) and **Stream Lookup** steps. Select the **Result is FALSE**.
- Create a hop from the **Read Postal Codes** step (Text File input) to the Stream lookup step.
- Double-click on the **Stream lookup** step to open its edit properties dialog box.
- Rename **Stream Lookup** to **Lookup Missing Zips**.



- Select the **Read Postal Codes** (Text File input) as the Lookup step.



Stream Value Lookup

Step name: Lookup Missing Zips

Lookup step: Read Postal Code

The key(s) to look up the value(s):

#	Field	LookupField
1	CITY	CITY
2	STATE	STATE

- Define the **CITY** and **STATE** fields in the keys to look up the values table. Click the drop down in the **Field** column and select **CITY**. Then, click in the **LookupField** column and select **CITY**. Perform the same actions to define the second key based on the **STATE** fields coming in on the source and lookup streams:
- Click **Get Lookup Fields**. **POSTALCODE** is the only field you want to retrieve. (To delete the extra **CITY** and **STATE** lines, right-click in the line and select **Delete Selected Line**.) Give **POSTALCODE** a new name of **ZIP_RESOLVED** and make sure the **Type** is set to **String**. Click **OK** to close the **Stream Lookup edit** properties dialog box.

Specify the fields to retrieve :

#	Field	New name	Default	Type
1	POSTALCODE	ZIP_RESOLVED		String
2				
3				

Preserve memory (costs CPU) ☒

Key and value are exactly one integer field ☐

Use sorted list (i.s.o. hashtable) ☐

OK Cancel Get Fields Get lookup fields

- Save your transformation.

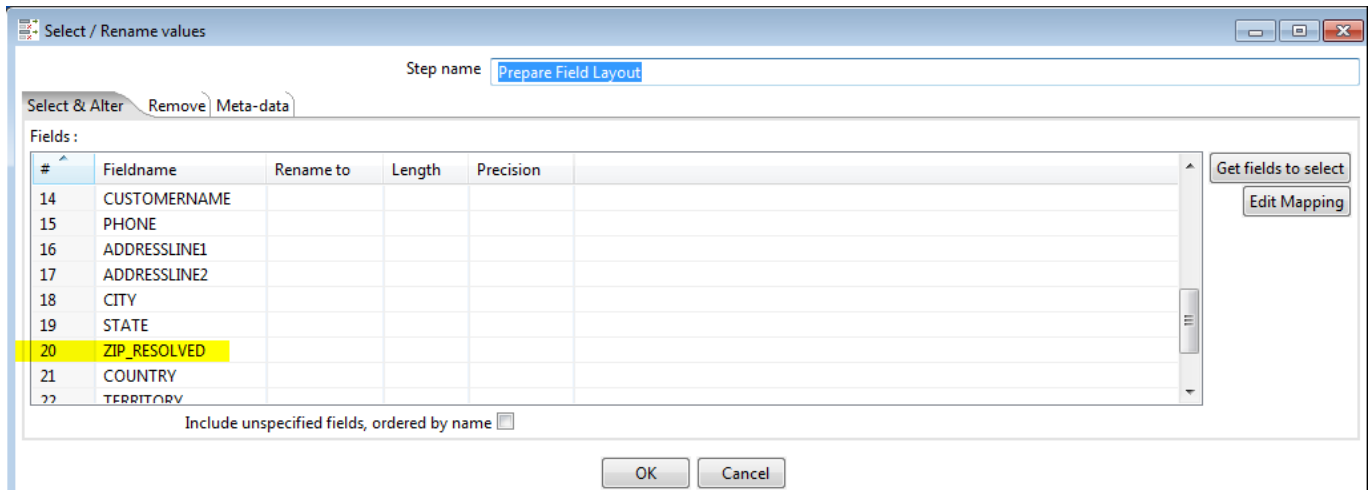
You can now select the **Lookup Missing Zips** step (Stream lookup) in the graphical workspace. Right-click and select **Preview** to display the preview/debugger dialog box. Click **Quick Launch** to preview the data flowing through this step. Notice that the new field, **ZIP_RESOLVED**, has been added to the stream containing your resolved postal codes.

8.3.4.7 Completing your Transformation

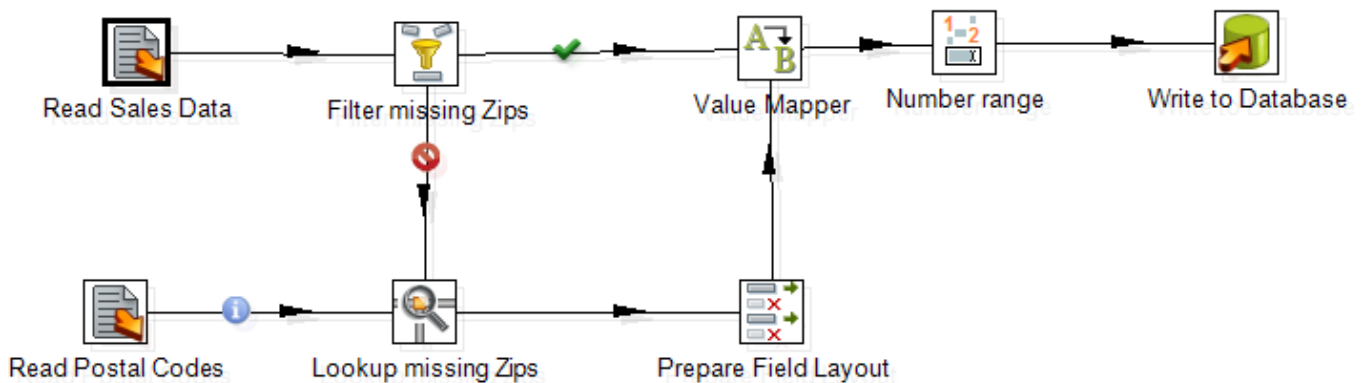
The last task is to clean up the field layout on your lookup stream so that it matches the format and layout of your other stream going to the **Write to Database (Table output)** step. You will create a **Select values** step. This is a very useful step for renaming fields on the stream, removing unnecessary fields, and more.

- Add a **Select Values** step to your transformation. Expand the **Transform** folder and choose **Select Values**.
- Create a hop between the **Lookup Missing Zips** and **Select Values** steps.
- Double-click the **Select Values** step to open its properties dialog box.

- Rename the **Select Values** step to, **Prepare Field Layout**.
- Click **Get fields** to select to retrieve all fields and begin modifying the stream layout.
- Select the **ZIP_RESOLVED** field in the Fields list and use <CTRL><UP> to move it just below the **POSTALCODE** (The one that still contains null values).
- Select the old **POSTALCODE** field in the list (line 20) and delete it.



- The original **POSTALCODE** field was formatted as an 9-character string. You must modify your new field to match the form. Click the **Meta-Data** tab.
- In the first row of the **Fields to alter table**, click in the **Fieldname** column and select **ZIP_RESOLVED**.
- Type **POSTALCODE** in the **Rename** to column; select **String** in the Type column, and type **9** in the **Length** column. Click **OK** to exit the edit properties dialog box.
- Draw a hop from the **Prepare Field Layout** (Select values) step to the **Write to Database** (Table output) step.



- Save your transformation.

8.3.4.8 Run Your Transformation

In the Spoon graphical interface, click (Run this Transformation or Job).

The **Execute a Transformation** dialog box appears. You can run a transformation locally, remotely, or in a clustered environment. For the purposes of this exercise, keep the default **Local Execution**.

- Click **Launch**.

The transformation executes. Upon running the transformation, the **Execution Results** panel opens below the graphical workspace.

Execution Results										
Execution History Logging Step Metrics Performance Graph										
#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Av
1	Read Sales Data	0	0	2823	2824	0	1	0	0	Fi
2	Filter missing Zips	0	2823	2823	0	0	0	0	0	Fi
3	Read Postal Codes	0	0	21379	21380	0	1	0	0	Fi
4	Lookup missing Zips	0	21455	76	0	0	0	0	0	Fi
5	Prepare Field Layout	0	76	76	0	0	0	0	0	Fi
6	Value Mapper	0	2823	2823	0	0	0	0	0	Fi
7	Number range	0	2823	2823	0	0	0	0	0	Fi
8	Write to Database	0	2823	2823	0	2823	0	0	0	Fi

The **Step Metrics** tab provides statistics for each step in your transformation including how many records were read, written, caused an error, processing speed (rows per second) and more. If any of the steps caused the transformation to fail, they would be highlighted in red as shown below.

Execution Results										
Execution History Logging Step Metrics Performance Graph										
#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Acti
1	Read Sales Data	0	0	2323	2325	0	0	0	0	Stop
2	Filter missing Zips	0	2223	2223	0	0	0	0	0	Stop
3	Read Postal Codes	0	0	9901	9902	0	0	0	0	Stop
4	Lookup missing Zips	0	1	0	0	0	0	0	1	Stop
5	Prepare Field Layout	0	0	0	0	0	0	0	0	Stop
6	Value Mapper	0	2053	2053	0	0	0	0	0	Stop
7	Number range	0	1953	1953	0	0	0	0	0	Stop
8	Write to Database	0	1	0	0	1	0	0	0	Stop

The **Logging** tab displays the logging details for the most recent execution of the transformation. Error lines are highlighted in red.

Kettle text file input

You can see that in this case the **Lookup Missing Zips** step caused an error because it attempted to lookup values on a field called `POSTALCODE2`, which did not exist in the lookup stream.

The **Execution History** tab provides you access to the Step Metrics and log information from previous executions of the transformation. This feature works only if you have configured your transformation to log to a database through the Logging tab of the Transformation Settings dialog.

The **Performance Graph** allows you to analyze the performance of steps based on a variety of metrics including how many records were read, written, caused an error, processing speed (rows per second) and more.

Kettle text file input

8.3.5 Build your first job

Jobs are used to coordinate ETL activities such as:

- Defining the flow and dependencies for what order transformations should be run
- Preparing for execution by checking conditions such as, "Is my source file available?," or "Does a table exist?"
- Performing bulk load database operations
- File Management such as posting or retrieving files using FTP, copying files and deleting files
- Sending success or failure notifications through email

For this step, imagine that an external system is responsible for placing your **sales_data.csv** input in its source location every Saturday night at 9 p.m. You want to create a job that will check to see that the file has arrived and run your transformation to load the records into the database. In a subsequent exercise, you will schedule the job to be run every Sunday morning at 9 a.m.

- Click (**New**) in the upper left corner of the Spoon graphical interface.
- Select **Job** from the list.
- Expand the **General** folder and drag a **Start** job entry onto the graphical workspace..

Kettle text file input

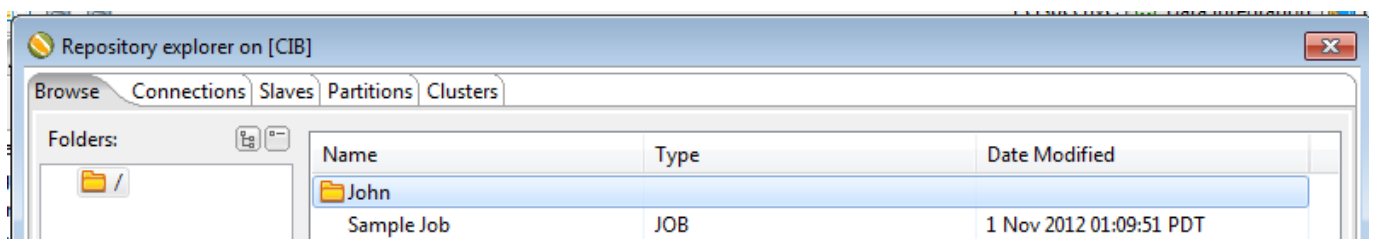
The Start job entry defines where the execution will begin.

- Expand the **Conditions** folder and add a **File Exists** job entry.
- Draw a hop from the **Start** job entry to the **File Exists** job entry.
- Double-click the **File Exists** job entry to open its edit properties dialog box. Click Browse and select the **sales_data.csv** from the following location: **... \design-tools\data-integration\samples \transformations\files**.

Be sure to set the filter to CSV files to see the file.

Kettle

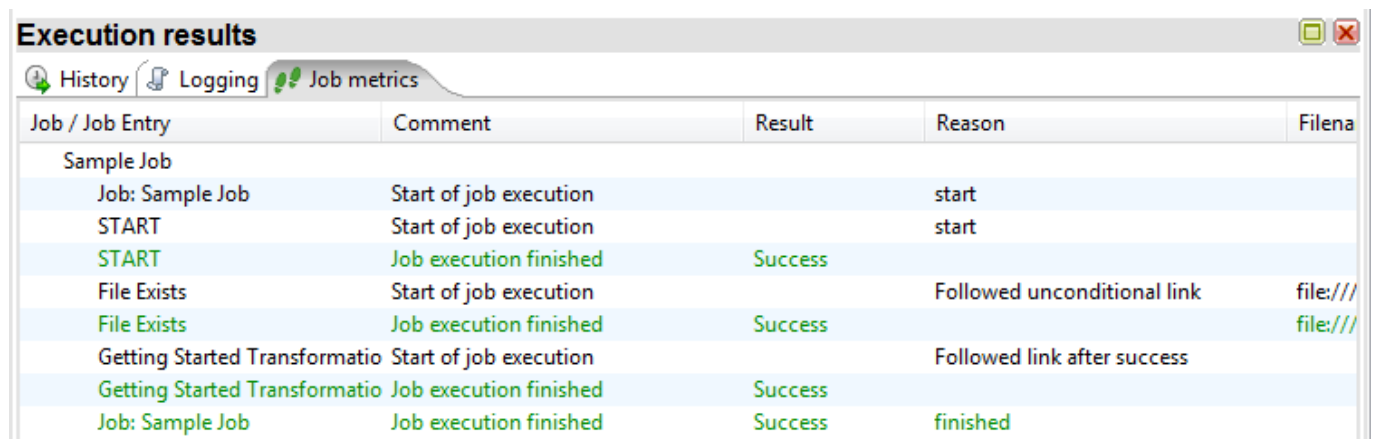
- Expand the **General** folder and add a **Transformation** job entry.
- Draw a hop between the **File Exists** and the **Transformation** job entries.
- Double-click the **Transformation** job entry to open its edit properties dialog box.
- Select the **Specify by name** and directory option. Click (Browse).
- Expand the repository tree to find your sample transformation. Select it and click **OK**. You likely have your transformation stored under the "tutorial," (not public), folder.



- Save your transformation as **Sample Job**.



- Click **Run Job**. When the **Execute a Job** dialog box appears, choose **Local Execution** and click **Launch**.



The screenshot shows a window titled "Execution results" with three tabs: "History", "Logging", and "Job metrics". The "Job metrics" tab is selected, displaying a table with the following data:

Job / Job Entry	Comment	Result	Reason	File name
Sample Job				
Job: Sample Job	Start of job execution		start	
START	Start of job execution		start	
START	Job execution finished	Success		
File Exists	Start of job execution		Followed unconditional link	file:///
File Exists	Job execution finished	Success		file:///
Getting Started Transformatio	Start of job execution		Followed link after success	
Getting Started Transformatio	Job execution finished	Success		
Job: Sample Job	Job execution finished	Success	finished	

The **Execution Results** panel should open showing you the job metrics and log information for the job execution.

8.3.6 Resources

For more information on the configuration and administration of the Pentaho Data Integration ,

1. Read the *CiB end user manual*
2. Visit the [Pentaho Community Home Page](#)

8.4 Nesstar Publisher

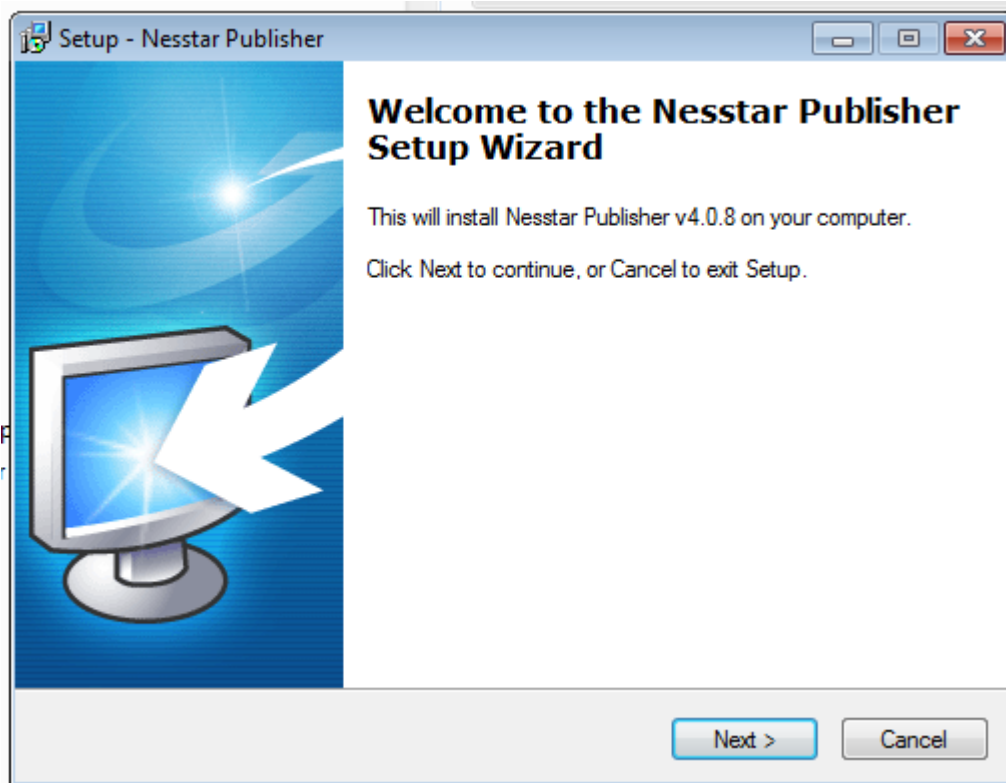
In this step, you are going to install the Nesstar Publisher application.

Nesstar Publisher have the following main features:

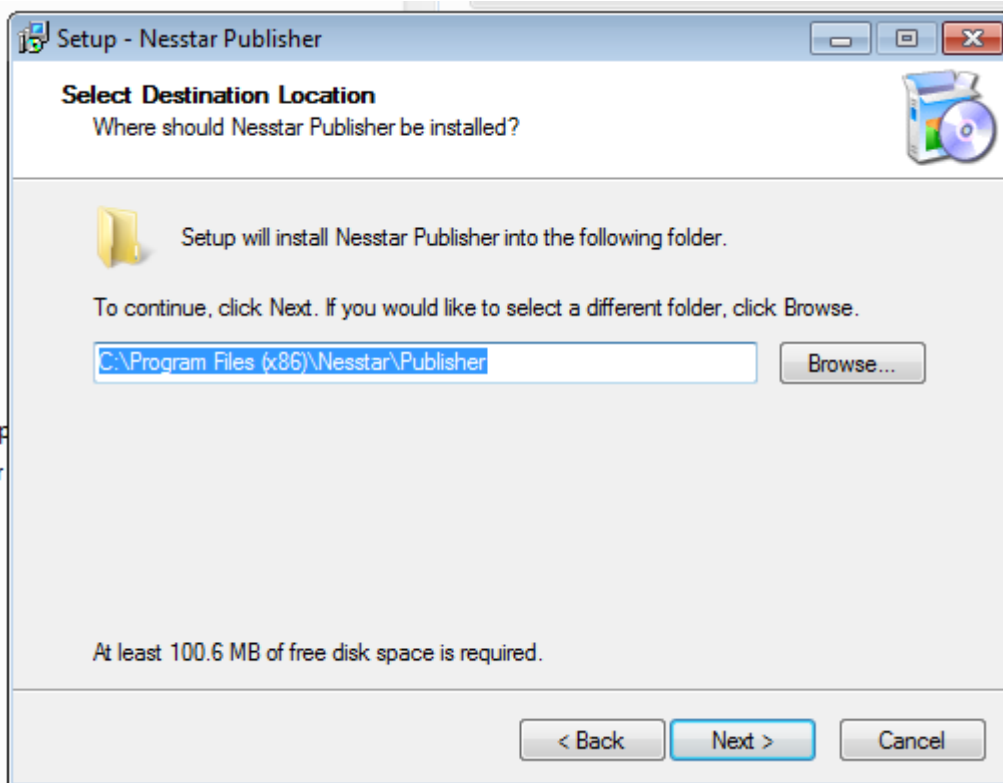
- The easy editing/creation of DDI documented datasets with no need to know XML
- The ability to import and export single DDI file/language studies
- The ability to create templates enabling your organization to standardise its use of the DDI
- Adding default text and controlled vocabularies to templates
- A Variable Repository to allow the sharing of variable categories within a dataset and between datasets
- The ability to create variable groups
- The easy setting of survey weights
- The inclusion of frequency and summary statistics options for each variable
- The ability to import and export data to the most common statistical formats
- The ability to insert links to relevant websites or documents which can then be viewed within Nesstar WebView
- Functionality to manage resources published to a Nesstar Server
- The ability to link hierarchically related survey datasets

8.4.1 Install Nesstar Publisher

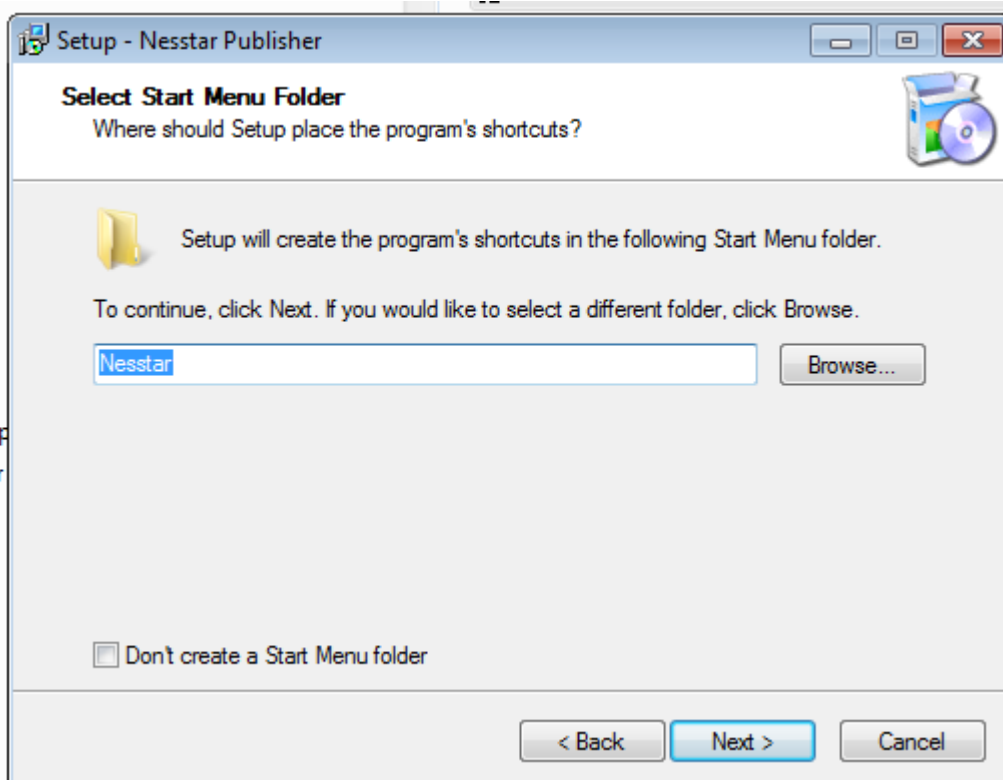
- Download the **NesstarPublisher_v4.0.8.exe** file from the Archive server.
- Double-click on the file to start the installation wizard.



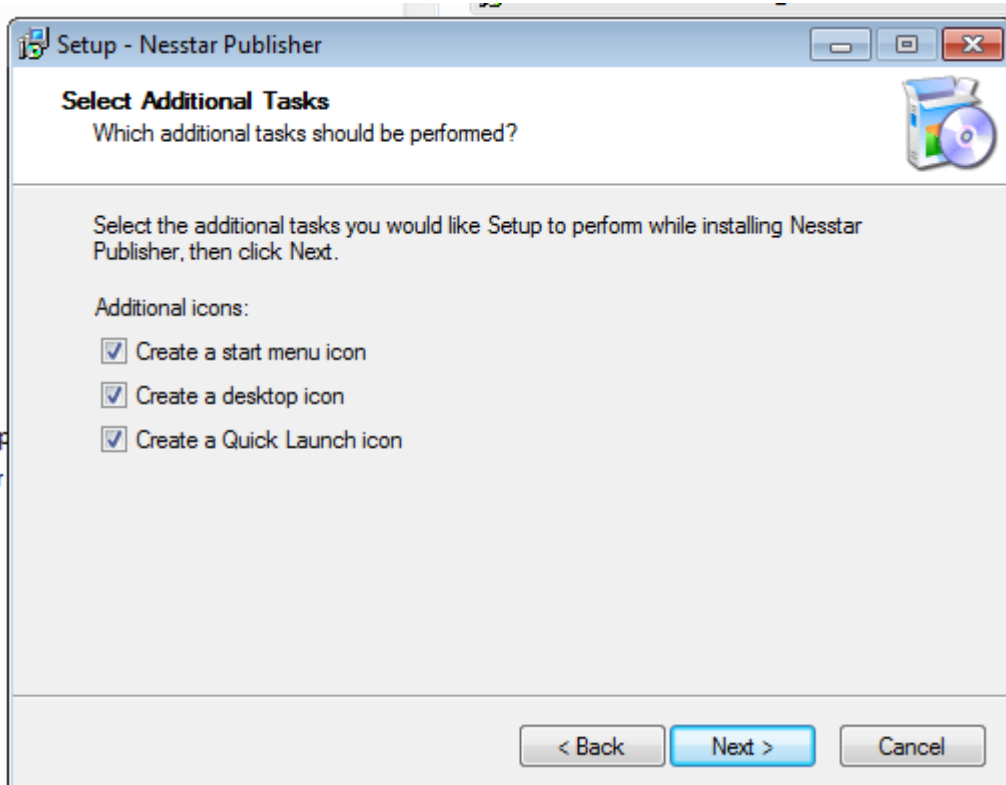
- Accept the license agreement.
- Accept the default destination folder: **C:\Program Files (x86)\Nesstar\Publisher**.



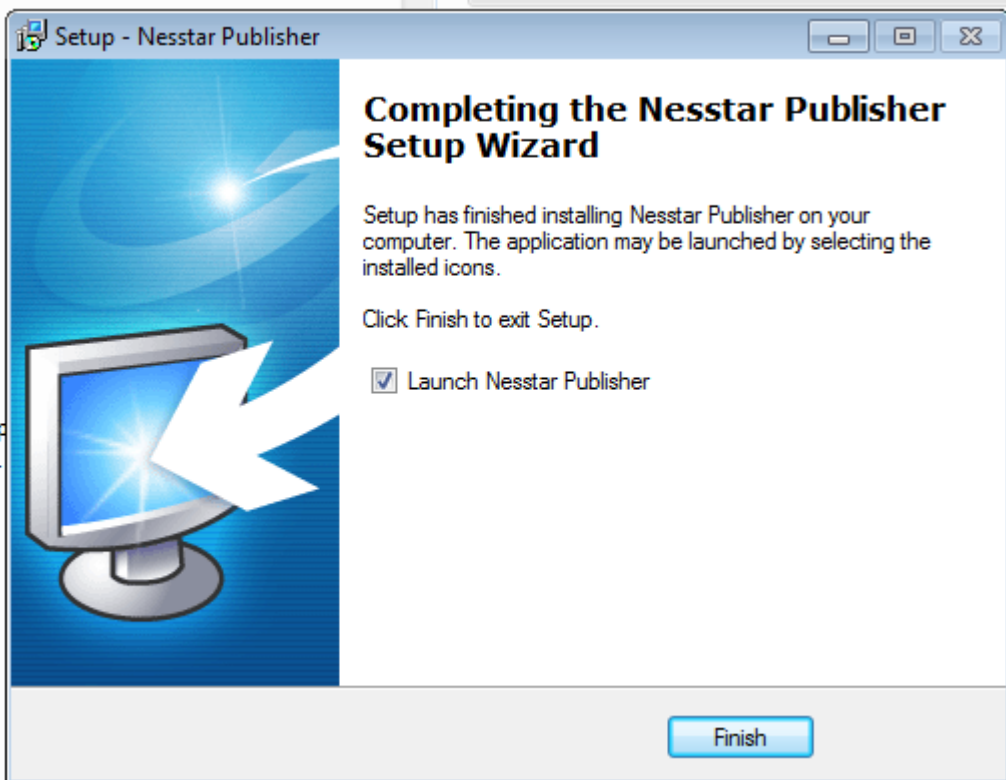
- Create the program's shortcut in the **Start** → **Nesstar** menu folder.



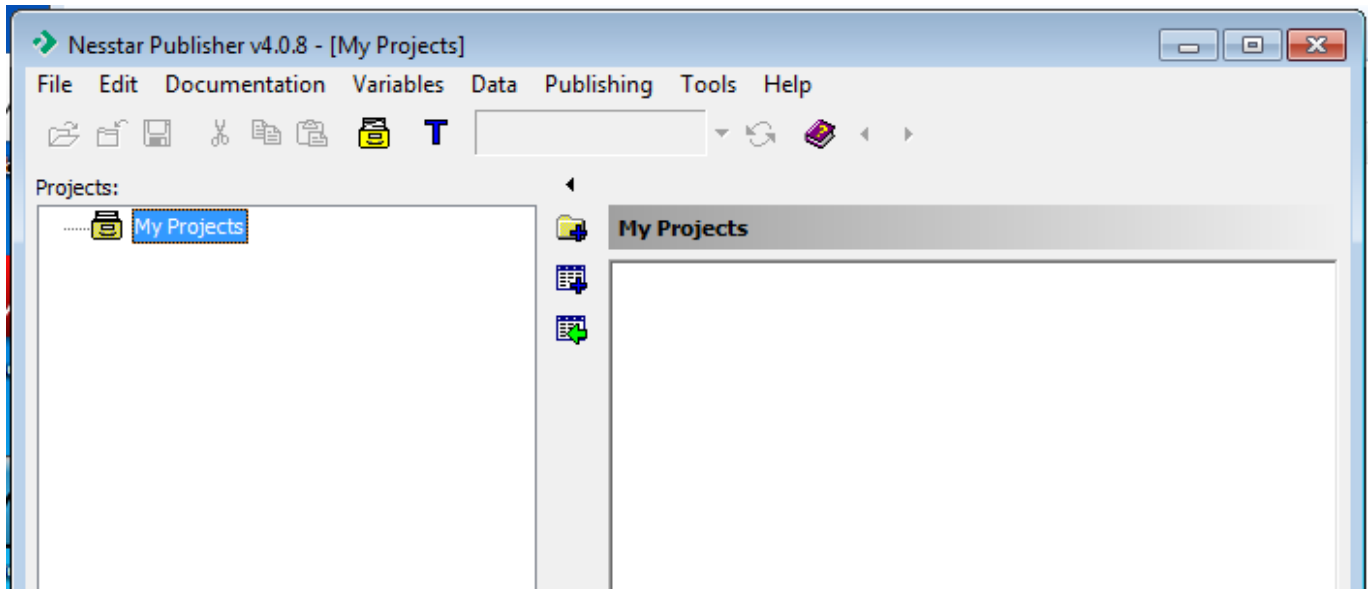
- Create a start menu icon, a desktop icon and a quick launch icon.



- Complete the installation and launch Nesstar for the first time.



You should see the screen shoot below.



- Download the user guide.

8.5 Zotero

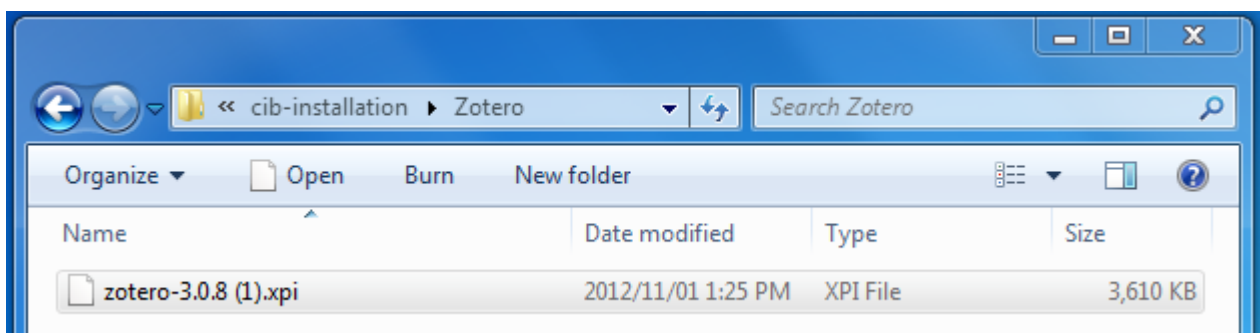
In this step, you are going to install Zotero. Zotero is a free, open-source tool to help you collect, organize, cite and share your research sources.

Zotero is available in two configurations: Zotero for Firefox and Zotero Standalone.

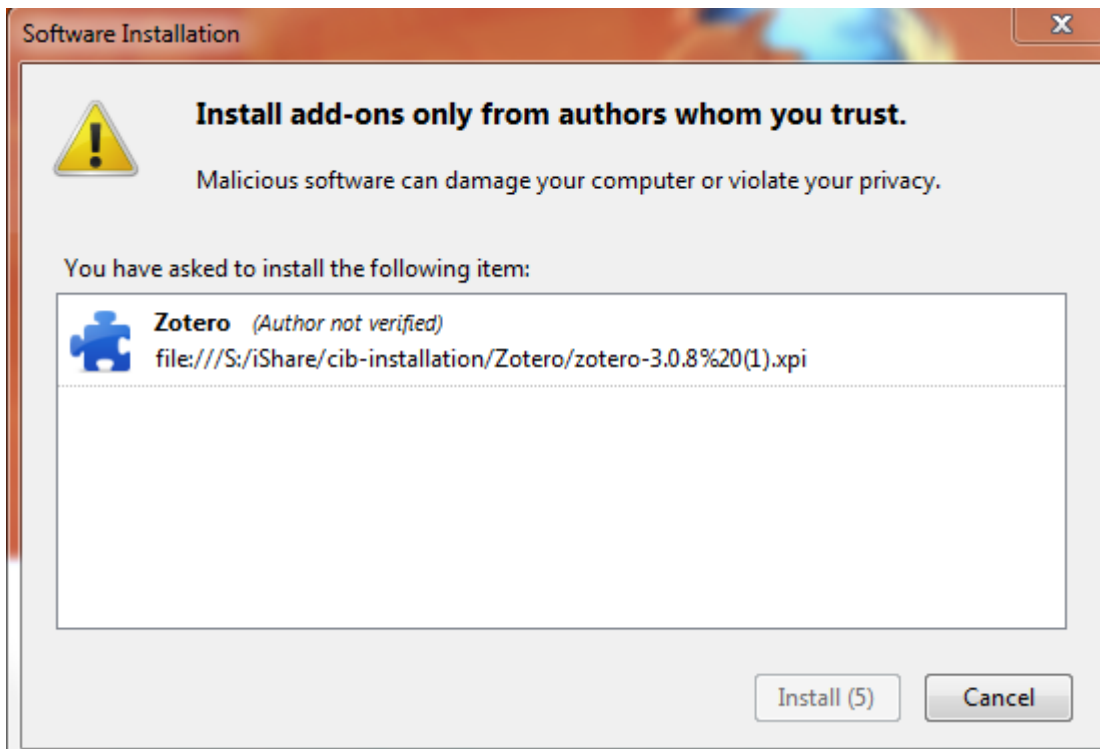
8.5.1 Install Zotero for Firefox

To install the Zotero Firefox extension,

- Locate the Zotero XPI file (e.g `zotero-3.0.8(1).xpi`) on the Archive server.



- Drag the **zotero-3.0.8(1).xpi** file onto a Firefox window.

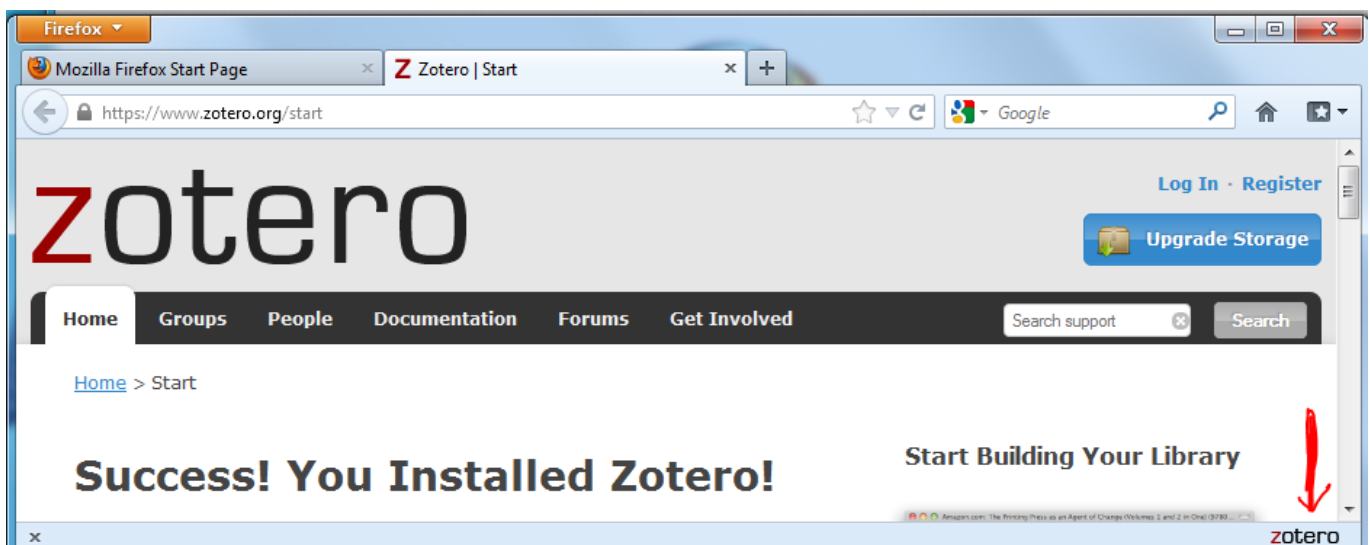


- Select the file then click **Install**.
- Restart Firefox

Tip

If a box appears above the web page with "Firefox prevented this site (www.zotero.org) from asking you to install software on your computer.", click the "Allow" button. Then click the "Install Now" button in the "Software Installation" pop up window (the button may be greyed out for a few seconds while Firefox downloads Zotero), and restart Firefox after the installation has completed.

You should now see the Zotero logo in the status bar in the right bottom corner of your Firefox browser window.



8.5.2 Install Zotero Standalone

To install Zotero Standalone,

- Run **Zotero-3.0.8_setup (1).exe** if you are running Windows 7
- Run **Zotero-3.0.8_linux-x86_64.tar.bz2.tar.bz2.tar.bz2** if you are running Ubuntu

Note

Word processor plugins for Microsoft Word and LibreOffice/OpenOffice are bundled with the Standalone Zotero. You don't need a separate installation.

- Install the plug-in for Word.
- Register to take full advantage of Zotero on <http://www.zotero.org/start>
- email validation .
- Read the quick start guide
- Watch the video tour

8.5.3 Resources

For more information on the configuration and administration of the Zotero application,

1. Visit [Zotero Home](#)

Chapter 9

SysAdmin desktop installation

9.1 Role

- Routine maintenance
- Backup

9.2 Setup the virtual machine

Using the instructions in Section 4.2, “[Create a virtual machine](#)” and Section 4.4, “[Install Linux Ubuntu as a guest operating system](#)”, setup a new VM with the following specifications:

- OS : Ubuntu Server 12.04 LTS
- CPU : 1
- RAM : 2
- Storage : 40 GB
- IP Address: 172.18.21.231
- Host Name: sysadmin.africacentre.local

9.3 Install vSphere Client

In this step, you are going to install vSphere Client version 5.0 on the SysAdmin desktop. If you use another workstation when not logged to the CiB, you may use the same procedure to install vSphere on that workstation.

9.3.1 GUI tools

You can refer to the Section 3.6, “[Install the vSphere Client](#)”.

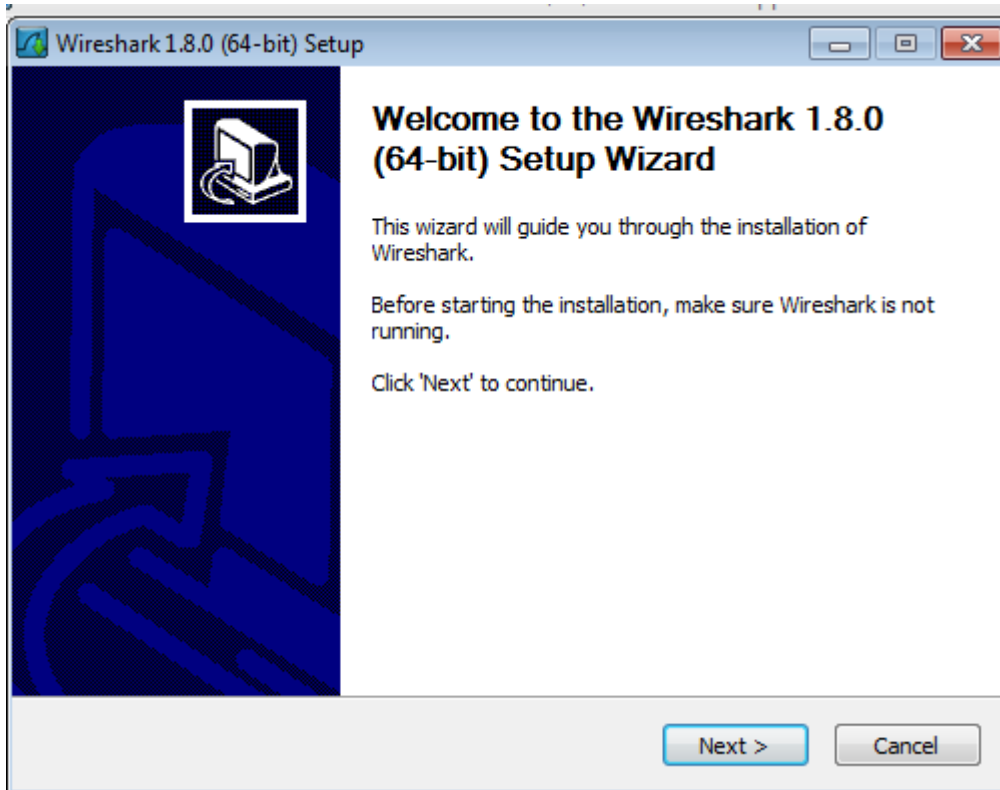
9.3.2 CLI tools

9.4 Install Wireshark

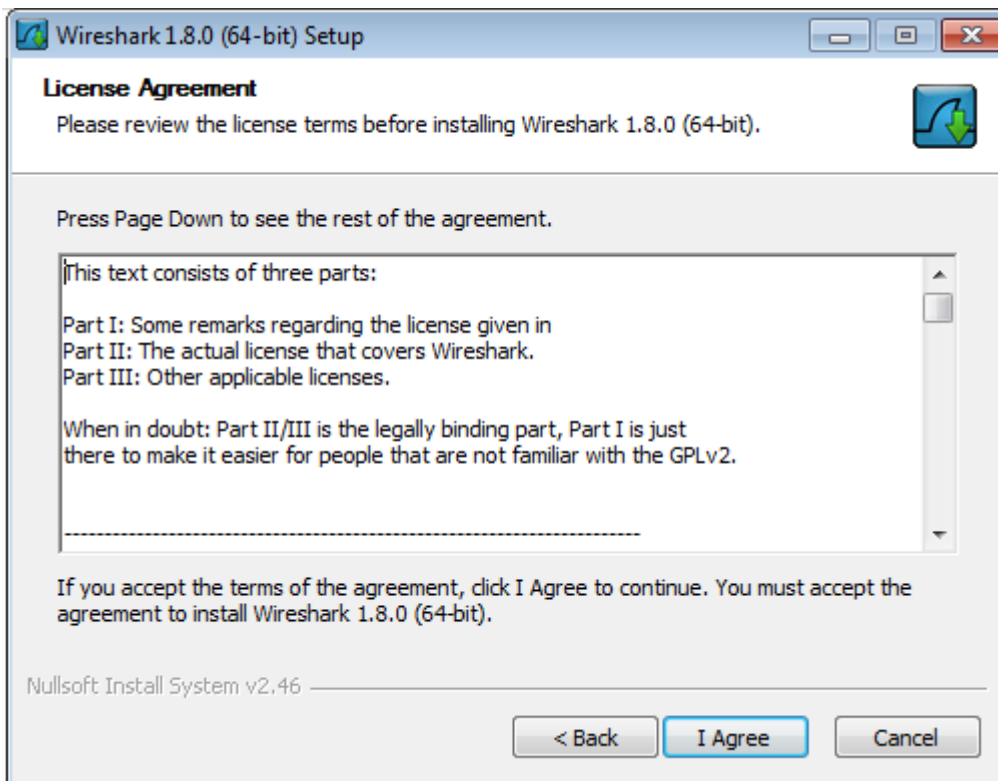
Download the Wireshark installer from the File server and Execute it.

Locate the Wireshark folder in the repository

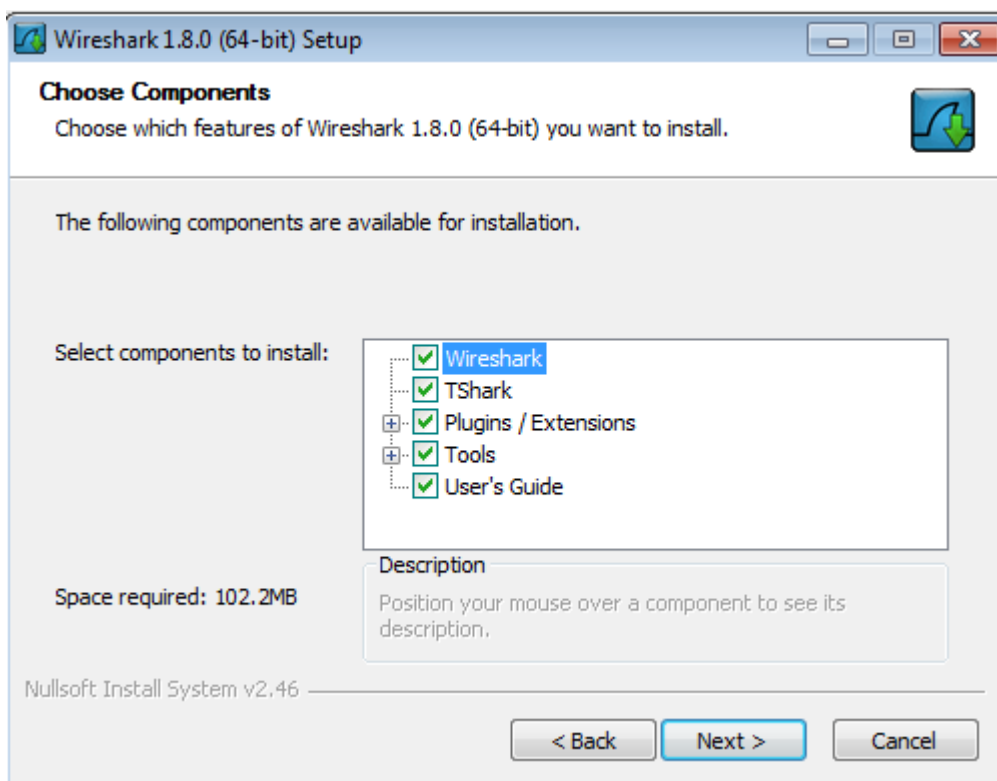
- Run the **Wireshark-win64-1.8.0.exe**.



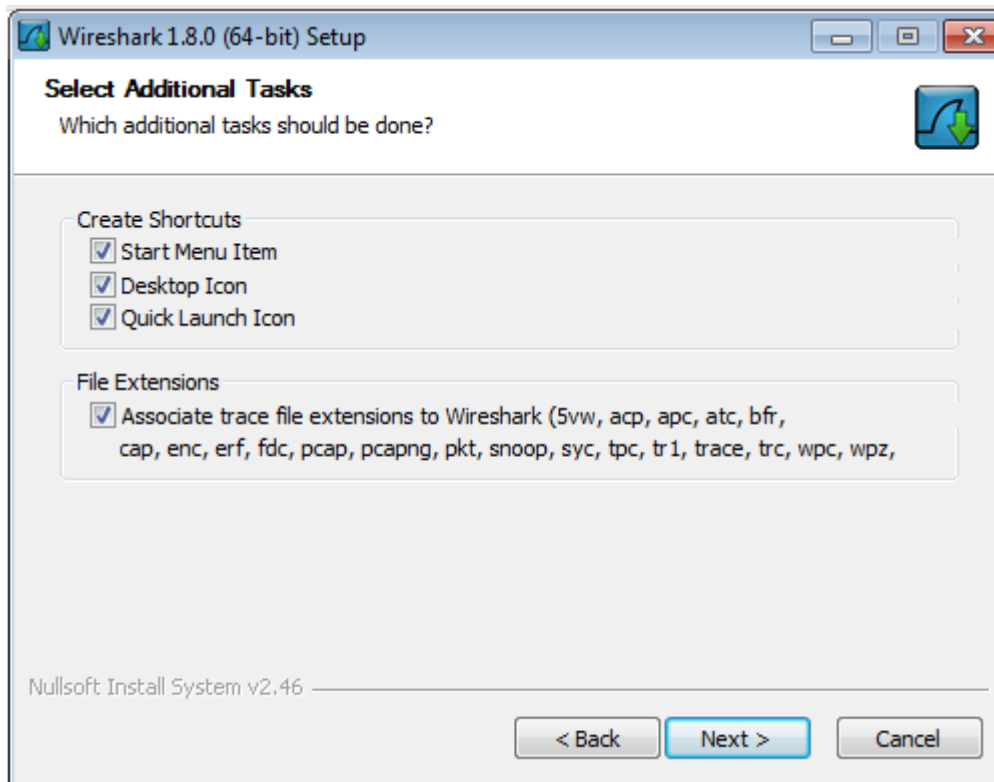
- Accept the license terms.



- Choose **All** components.

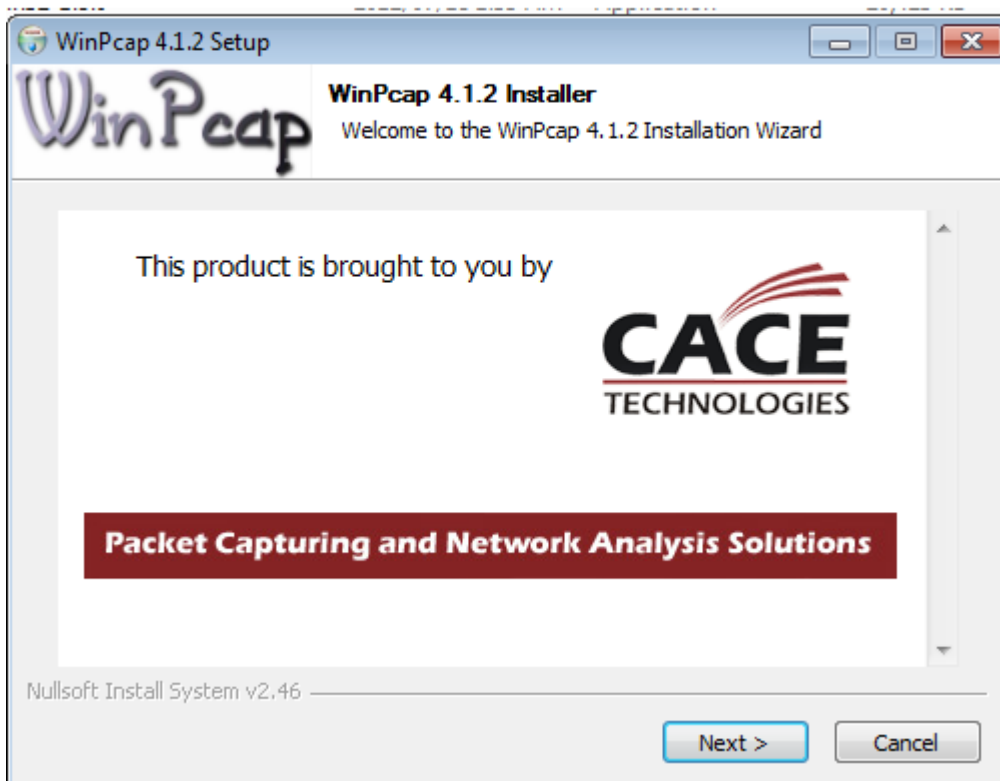


- Select **All** additional tasks .

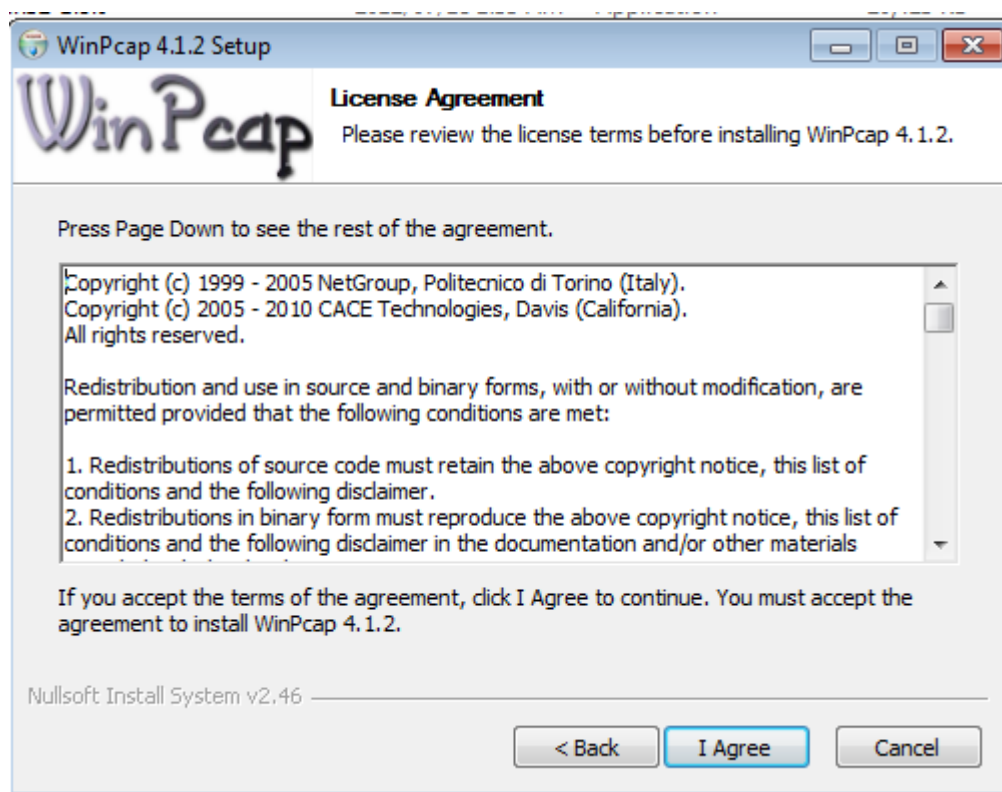
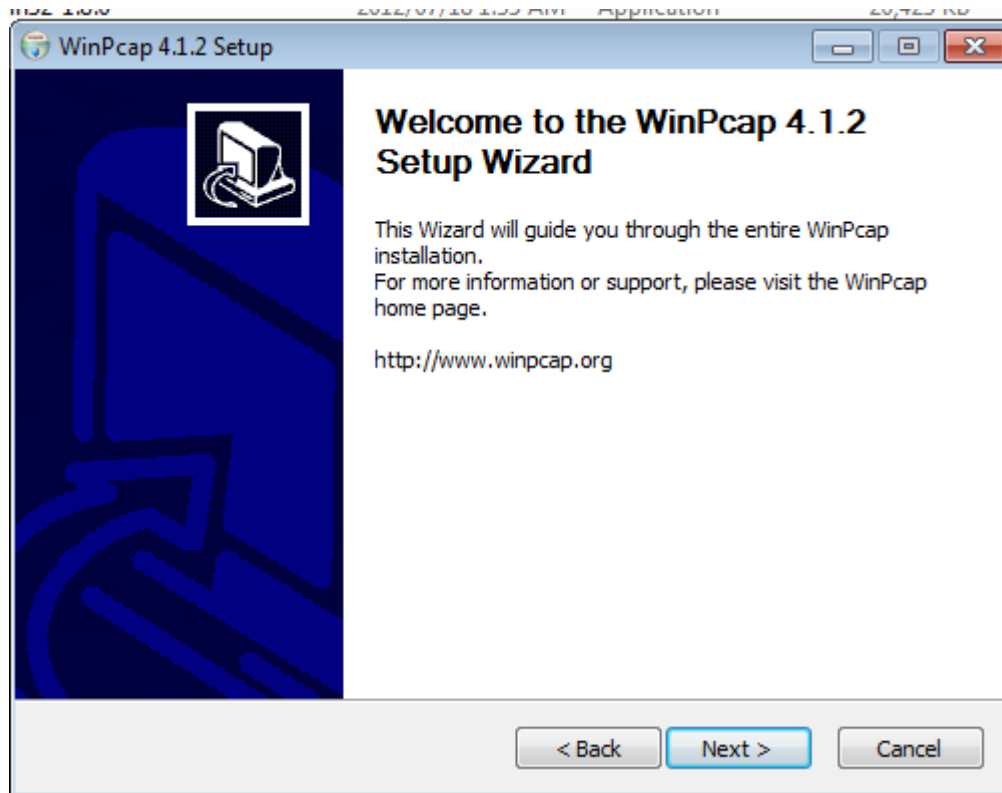


- Install **WinPcap**.

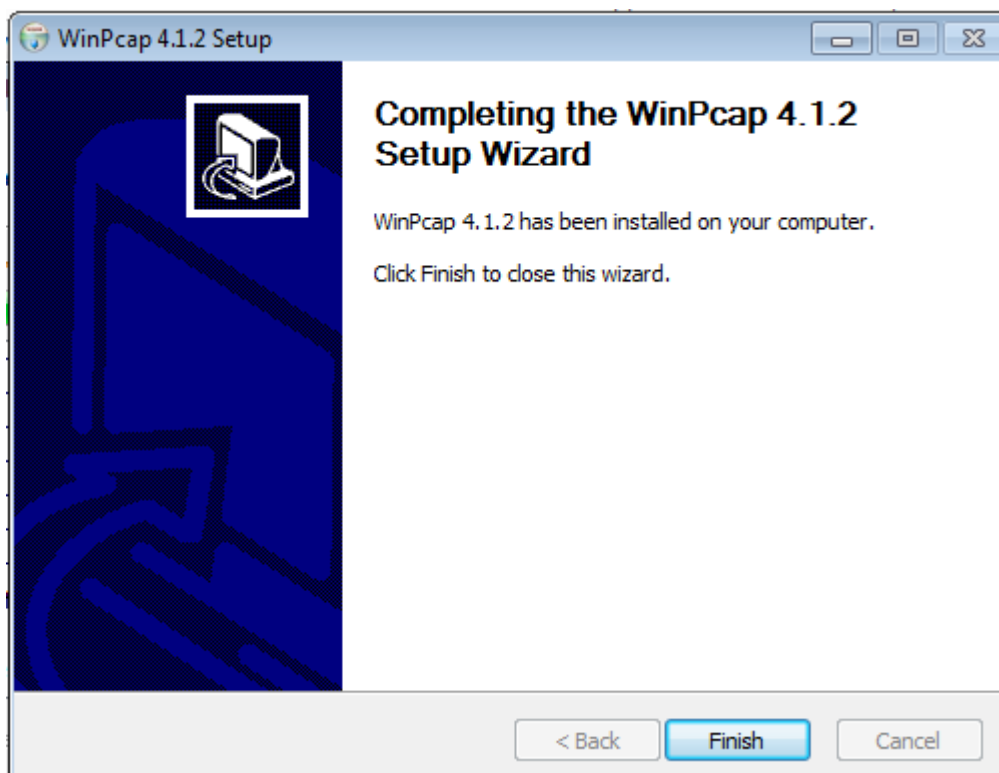
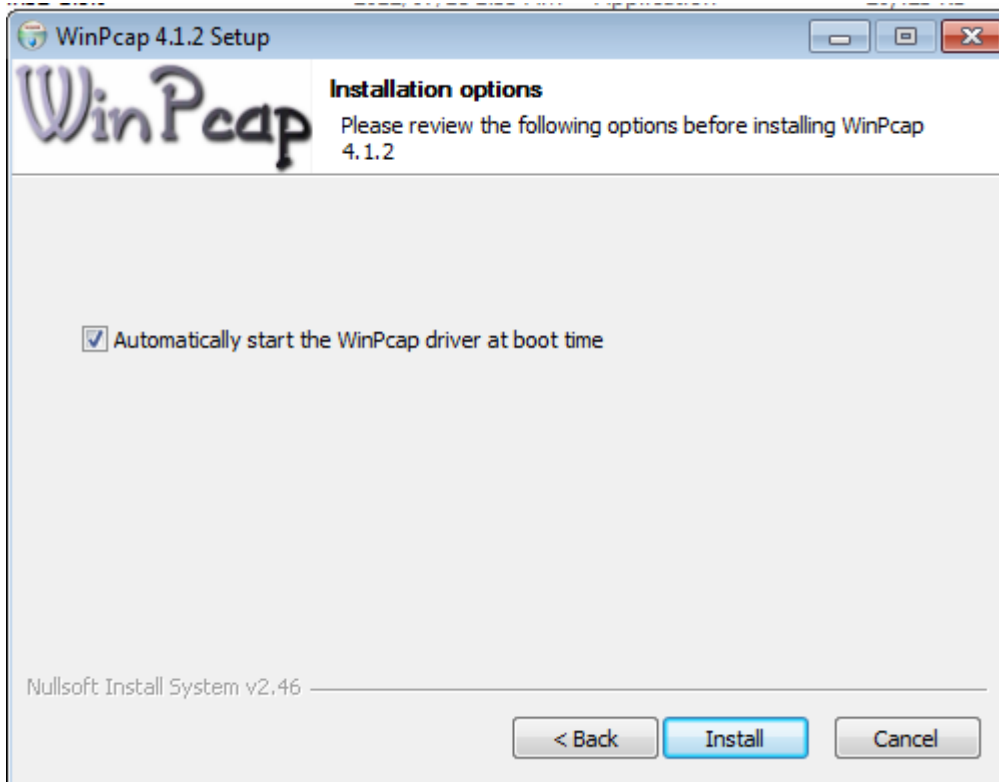
This will launch the **WinPcap** installer.

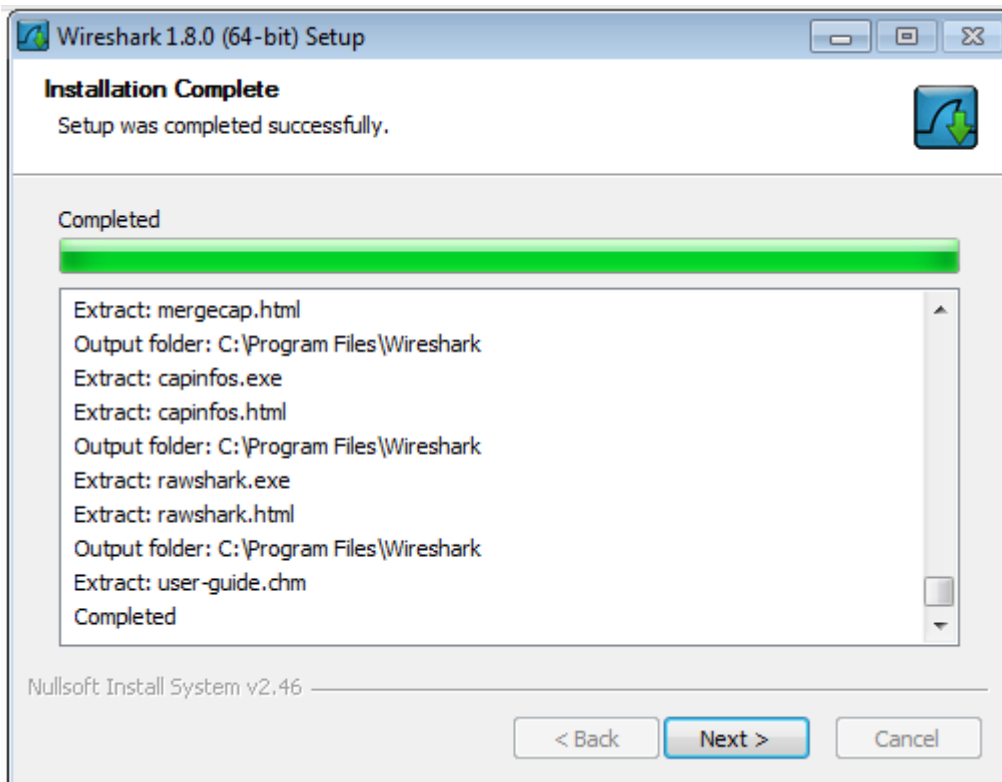


Follow the prompts and accept the default options.

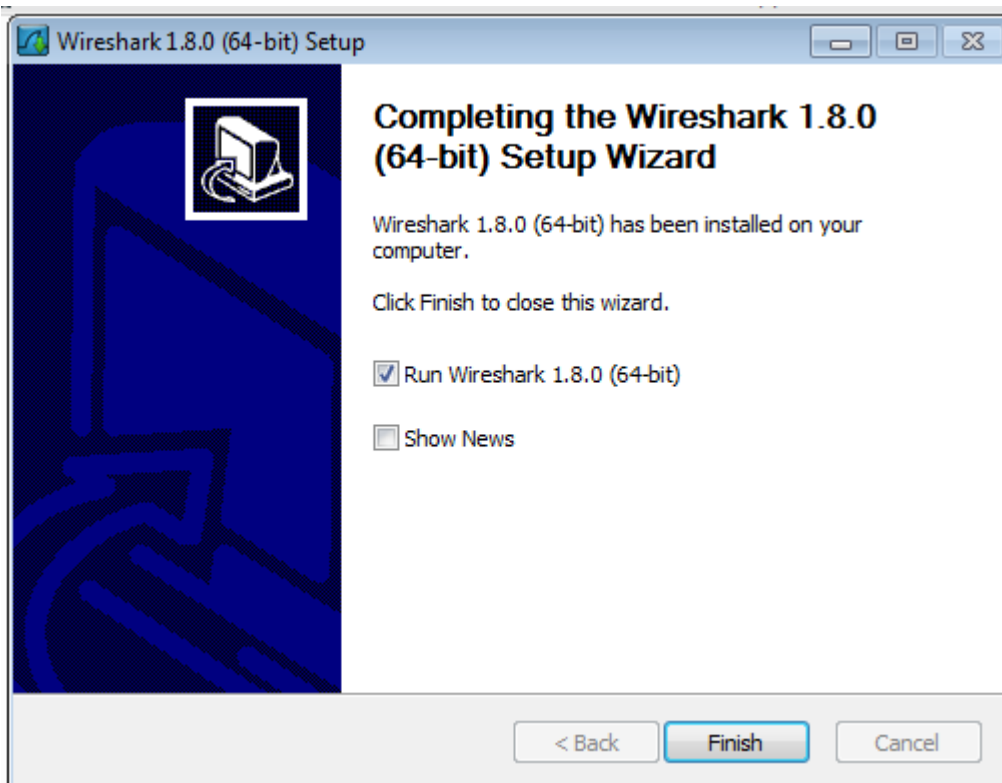


Select **Start the WinPcap driver at boot time**

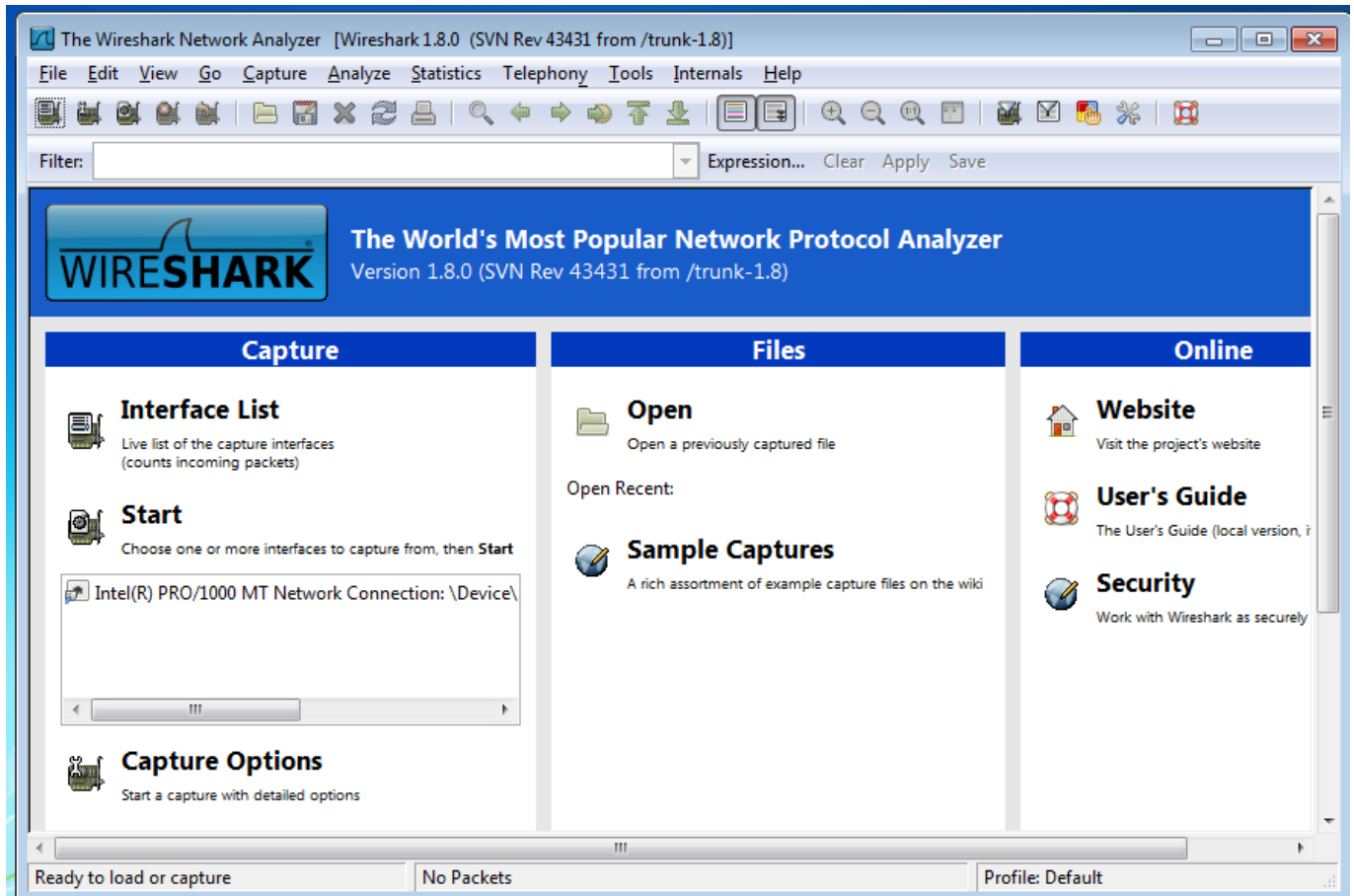




Run wireshark for the first time to check if the installation is successful.



You should see the screen below.



9.4.1 Resources

For more information on the configuration and administration of the Wireshark application, consult the following resources:

1. Read the **CiB administration manual**.
2. Go to [MySQL Home Page](#).
3. Consult the full documentation both online and offline format from the [MySQL Developers portal](#).

9.5 Install guettoVCB backup script

In this step, you are going to install and configure **guettoVCB**. The **guettoVCB** script takes snapshots of live running virtual machines, backs up the master VMDKs and upon completion, deletes the snapshot until the next backup.

It is a free alternative to the proprietary VMware's VCB tools. The only caveat is that it utilizes resources available to the Busybox console of the ESXi server running the backups as opposed to following the traditional method of offloading virtual machine backups through a VCB proxy.

The script has been tested on ESXi 5.x and support backup on local storage. The script is non-iterative and will be setup to run via **cron**. The script accepts a text file that lists the display name of virtual machines that are to be backed up. Additionally, you can specify a folder containing the configuration files on a per VM basis for granular control over backup policies.

Additionally, for CiB host that don't have persistent NFS datastores designated for backups, the script offers the ability to automatically connect the ESXi server to a NFS exported folder and then upon backup completion, disconnect it. The connection is established by creating an NFS datastore link which enables monolithic (or thick) VMDK backups as opposed to using the usual UNIX mount command which necessitates breaking VMDK files into the 2Gbsparse format for backup.

In its current configuration, the script will allow up to 3 unique backups of the VM before it will overwrite the previous backups;

- guettoVCB version x.xx

9.5.1 Setup

- Connect to the CIB host with SSH
- Download **guettoVCB** from the file server
- Extract the contents of the tarball

```
cd
tar -zxvf lamw-ghettoVCB-19e0d4b.tar.gz
```

The script is now ready to be used and is located in a directory name **guettoVCB**

9.5.2 Configuration

```
VM_BACKUP_VOLUME=/vmfs/volumes/backups/cib
DISK_BACKUP_FORMAT=thin
VM_BACKUP_ROTATION_COUNT=3
POWER_VM_DOWN_BEFORE_BACKUP=0
ENABLE_HARD_POWER_OFF=0
ITER_TO_WAIT_SHUTDOWN=3
POWER_DOWN_TIMEOUT=5

ENABLE_COMPRESSION=0
ADAPTER_FORMAT=buslogic
VM_SNAPSHOT_MEMORY=0
VM_SNAPSHOT_QUIESCE=0
ENABLE_NON_PERSISTENT_NFS=0
UNMOUNT_NFS=0
NFS_SERVER=172.30.0.195
NFS_MOUNT=/nfsshare
NFS_LOCAL_NAME=nfs_storage_backup
NFS_VM_BACKUP_DIR=daily
SNAPSHOT_TIMEOUT=15
EMAIL_LOG=0
EMAIL_DEBUG=0
EMAIL_SERVER={archive-ip-server}
EMAIL_SERVER_PORT=25
EMAIL_DELAY_INTERVAL=1
EMAIL_TO= {help-email}
EMAIL_FROM= {site-code}@indepth-ishare.org
```

9.5.3 Usage

- Create a **vms-to-backup** file with the names of the VM to be backup

```
archive
db
dm
nada
sysadm
```

- Test the script in dry mode

```
.guettoVCB.sh -f vms_to_backup -d dryrun
```

- Backup VMS stored in a list

```
./guettoVCB -f vms_to_backup
```

9.5.4 Setup a cron job

In this step, you are going to use a **cron** job to schedule backup every night at midnight evreryday and send the output to a unique log file.

- Setup the cronjob bu appending the following line to /var/spool/cron/crontabs/root

```
0 0 * * 1-7 /vmfs/volumes/simple-jack-local-storage/guettoVCB.sh -f
/vmfs/volumes/simplejack-local-storage/backuplist >
/vmfs/voulmes/simplejack-local-storage/cib-backup-$(date +%s).log
```

- Verify that the crontab entry has been updated by using "cat" utility

```
cat /var/spool/cron/crontabs/root
```

- Restart the current cron daemon to apply the changes

```
~ # kill $(cat /var/run/crond.pid)
~ # busybox crond
```

To ensure that this cronjob will persist through a reboot, add the following two lines to **/etc/rc.local/**.

```
/bin/kill $(cat /var/run/crond.pid)
/bin/echo "0 0 * * 1-5 /vmfs/volumes/simplejack-local-storage/ghettoVCB.sh -f /vmfs/volumes ↵
/simplejack-local-storage/backuplist > /vmfs/volumes/simplejack-local-storage/ghettoVCB- ↵
backup-$(date +%s).log" >> /var/spool/cron/crontabs/root
/bin/busybox crond
```

- To test your configuration, initiate manually an ESXi backup by running **auto-backup.sh**

```
~ # /sbin/auto-backup.sh
```

9.5.5 Resources

For more information on the configuration and administration of the **guettoVCB** application, consult the following resources:

1. Read the **CiB administration manual**
2. Go to [GhettoVCB home page](#)

Appendix A

Annexures

A.1 Site code

BD011 ICDDR-B: Matlab	BD012 ICDDR-B: Bandarban
BD013 ICDDR-B: Chakaria	BF011 Kaya
BF021 Nanoro	BF031 Nouna
BF041 Ouagadougou	BF051 Sapone
CI011 Taabo	ET011 Butajira
ET021 Gilgel Gibe	ET031 Kilite Awlaelo
GH001 INDEPTH Network	GH011 Navrongo
GH021 Kintampo	GH031 Dodowa
GM011 Farafenni	GM021 West Kiang
GW011 Bandim	ID011 Purworejo
IN011 Ballabgarh	IN021 Vadu
KE011 Kilifi	KE021 Kisumu
KE031 Nairobi	KE041 Mbita
KH011 Mekong	MW011 Karonga
MZ011 Manhica	PG011 Wosera
SN011 IRD: Bandafassi	SN012 IRD: Mlomp
SN013 IRD: Niakhar	TH011 Kanchanaburi
TZ011 Ifakara Health Institute: Ifakara	TZ012 Ifakara Health Institute: Rufiji
TZ021 Magu	UG011 Iganga/Mayuge
UG021 Rakai	VN011 Hanoi Medical University: Dodalab
VN012 Hanoi Medical University: Filabavi	VN021 Chililab
ZA011 Agincourt	ZA021 Dikgale
ZA031 Africa Centre	

Table A.1: Site code sorted by country, precedence

A.2 Naming convention

- Username : first letter of the first name followed by the surname, all in lower case
- Hostname : CIB-<code site> : Example CIB-ZA031
- analytical database :
- Virtual machine's display name: do not use space in the name

A.3 Example Glossary

Glossaries are optional. Glossaries entries are an example of a style of AsciiDoc labeled lists.

hypervisor

The corresponding (indented) definition.

VM

Virtual Machine The corresponding (indented) definition.

VMDK

Virtual Machine DK
