

How VMware virtual machines are used at RUBRIC

1 Introduction

The RUBRIC project utilises virtual infrastructure to provide services to the project partners, as well as a development and testing environment for the technical team. This document will briefly outline what virtualisation is, what virtual machines are, and how they are used by the RUBRIC project. This document is aimed at RUBRIC project partners, for a more in-depth evaluation of the topic please see the websites listed in the further reading section, or contact the RUBRIC technical team.

2 What is a Virtual Machine?

The traditional way of utilising a computer, and in particular a server, is to create a specialised server that has a single operating system installed and use the server to run a small number of carefully selected applications. The operating system is tightly coupled to the underlying computer hardware. This means the operating system, applications, and data stored on the server are not very portable. The number of applications that are run on the server is very small and they are carefully chosen to reduce the risk of conflicts during the normal operation of the server, for example not running more than one instance of a web server. Using this type of architecture it is possible that if an error occurs in one application the other applications running on the server are also adversely affected. A typical error is one application consuming all available memory and CPU resources and causing the other applications to fail. This type of monolithic architecture can be seen in figure 2.1 below.

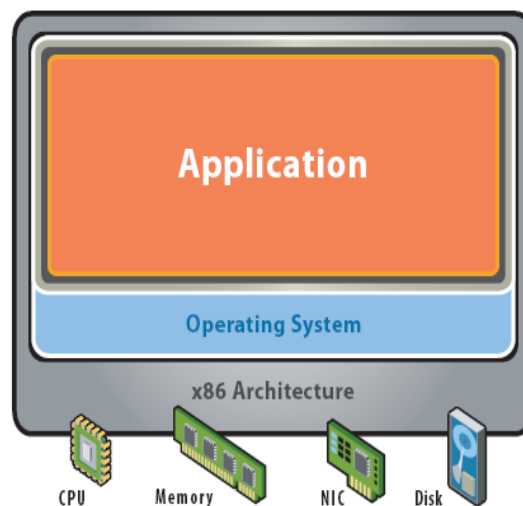


Figure 2.1 Traditional computing architecture
(Source: VMware Virtualisation Overview)

An alternate architecture that is gaining acceptance is the use of virtual infrastructure. Virtual infrastructure provides a layer of abstraction between the underlying computer hardware, the operating system and applications that run on it. This layer of abstraction is achieved through the use of virtual machines. When using virtual machines it is possible to run multiple

operating systems on a single physical system. Each operating system, and associated applications, are installed in their own virtual machine. A virtual machine shares access to the underlying physical system in a controlled manner via the virtualisation layer. Each virtual machine is known as a “guest” with the underlying system known as the “master”. The architecture can be seen more clearly in figure 2.2.

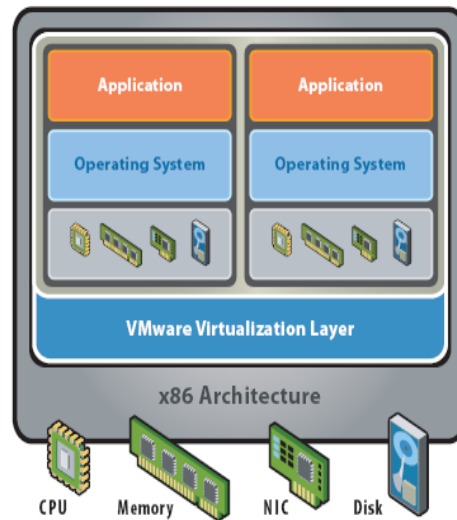


Figure 2.2 Virtual infrastructure architecture (Source: VMware Virtualisation Overview)

The virtual machine presents to the guest operating system a defined set of virtual hardware, this virtual hardware is specific to the virtual machine and not to the underlying physical computer system. As the hardware presented to the operating system is virtual the operating system is no longer tightly coupled to the underlying computer system. It is therefore possible to migrate a virtual machine from one physical system to another without the need to reconfigure the operating system or applications. The only constraint in such a move is that the same virtualisation layer be present, it is not possible to migrate a virtual machine created using VMware products to a server running Microsoft Virtual Server for example.

Another benefit of using virtual infrastructure is that each virtual machine is completely independent of any other virtual machine running on the same system. The same physical system can support virtual machines running Linux, Microsoft Windows, or other operating systems at the same time. It also means that if one virtual machine has its security compromised, the effects are contained to the one virtual machine with the others not being affected. Lastly the virtualisation layer controls access to the underlying physical system and shares the resources among all of the virtual machines within specified parameters. This means that one virtual machine, and therefore the operating system or application within it, can not consume all of the available resources on the physical system and adversely affect the other virtual machines.

3 How are Virtual Machines used at RUBRIC

There are two different ways in which virtual machines are used at RUBRIC. The first are virtual machines used for development and testing on the technical team laptops. The second are virtual machines that are used as servers providing services to the project partners.

3.1 Technical team virtual machines

Each technical team member is provided with a laptop that is used for day-to-day tasks. The base operating system of these laptops is Microsoft Windows XP as this is part of the standard operating environment at USQ. Using Windows presents a challenge in that all of the software used to provide institutional repositories that RUBRIC is investigating run on the Linux operating system. This presented two different scenarios to the technical team. The first scenario is to use a dedicated Linux server for testing and development work. The second was to use virtual machines. By using the VMware Workstation product the technical team are able to have multiple virtual machines specifically dedicated to the development and testing of the Linux based software.

The technical team maintains a “master” virtual machine. This virtual machine has a basic installation of Linux appropriately configured as a basic server. Typically the process of installing a basic Linux installation, with some customisations, can be achieved in approximately an hour and a half. Before testing a new software system, such as a new version of DSpace or the Shibboleth IdP software a copy of the virtual machine is made. This copy, or clone, is ready to use in approximately 10 minutes, significantly reducing the amount of time it takes to create a Linux installation ready for testing purposes.

Another benefit of using VMware workstation is the ability to take “snapshots”. A snapshot preserves the state of the virtual machine at a specific moment in time. It is then possible to revert to a previous snapshot should the need arise. This enhances the work flow of the technical team considerably. The snapshot feature can provide a work flow similar to the following when testing new software, including the upgrade of the software to a new version.

1. Clone the master virtual machine
2. Install the software to be tested, e.g. DSpace
3. Take a snapshot of the virtual machine once the system is installed
4. Upgrade the software to the latest version
5. Test the installation of the software

In the event that the upgrade procedure fails it is possible to debug the issue, document what occurred, fall back to the snapshot taken in step 3 and try the process again. By utilising virtual machines it is possible to have a complete testing environment hosted on one physical laptop.

3.2 Virtual machines as servers

The RUBRIC project has only three physical servers hosted in the USQ data centre. The RUBRIC project uses these servers to provide services to the RUBRIC project partners such as communal instances of the various institutional repository solutions and, where necessary, a dedicated instance for use by the project partner alone. It is also necessary to provide servers for use by the RUBRIC project team such as an instance of the Trac software for management of the project.

The primary server is running VMware ESX server and is connected to one of the main USQ storage area networks (SAN). The VMware ESX server allows the RUBRIC project to host all of the servers that we require as individual virtual machines. These servers include:

6. A master installation of Linux
7. Instances of DSpace, Fez and Vital for use by all of our project partners
8. A test instance of one of the institutional repository software to facilitate more in-depth testing by each project partner institution
9. The rubric-central server used to manage the RUBRIC project
10. The rubric-idp server used for participation in the MAMS Shibboleth federation

The files that make up the virtual machines are all stored on one of the main USQ storage area networks. As each virtual machine presents virtual hardware to the guest operating system, and as it is not tightly coupled to the underlying hardware it is possible to migrate these virtual machines between servers with a minimum of downtime. Just recently the RUBRIC virtual machines were migrated from one server to another without the need to reconfigure the operating system, or the various applications in each virtual machine.

A second, smaller, server is used to run VMware Virtual Centre and the backup software esxRanger from Vizioncore. The Virtual Centre software provides the functionality necessary to manage all aspects of the virtual machines running on the ESX server. This includes such things as the amount of disk space each virtual machine has, the amount of memory available, and how many CPU's are available. It is also possible to make clones of virtual machines very quickly which allows the team to provision new servers very quickly. The esxRanger software allows the technical team to create "hot" backups of each virtual machine. The backups are centrally managed on the server and a backup can be taken while the virtual machine is still running.

The third server is being used to test various new software and systems, such as VMware Server which is currently in beta. The testing is being carried out on a server in the USQ data centre, as opposed to a server in the technical team offices, to allow the technical team to conduct testing in as close to a real world scenario as possible. A possible outcome of this testing will be the ability to create virtual machines at RUBRIC that can be deployed to our project partners. This testing is still in its very early stages and it will be some time before any outcomes are realised.

4 Other Virtualisation Solutions

It should be noted that the VMware software used by the RUBRIC project is not the only virtualisation software available. Other notable examples are Microsoft Virtual Server 2005 and Xen 3.0 from XenSource. Unfortunately at the time of writing the Microsoft software did not allow Linux to be installed in virtual machines, and the Xen software only works with Linux. Through thorough testing and evaluation the VMware suite of products has been determined to provide the appropriate level of functionality required by the RUBRIC project. It should be noted that any virtualisation should be evaluated by organisations independently and must take into account the needs of the individual organisation.

5 Further Reading

VMware Virtualization Overview

<http://www.vmware.com/pdf/virtualization.pdf>

VMware Workstation

<http://www.vmware.com/products/ws/>

VMware ESX Server

<http://www.vmware.com/products/esx/>

VMware Virtual Centre

<http://www.vmware.com/products/vc/>

VMware Virtual Server

<http://www.vmware.com/products/server/>

Vizioncore esxRanger

<http://www.vizioncore.com/esxrangerpro.html>

Microsoft Virtual Server 2005

<http://www.microsoft.com/windowsserversystem/virtualserver/default.msp>

Xen 3.0 from XenSource

<http://www.xensource.com/>

DSpace Institutional Repository

<http://www.dspace.org/>

Shibboleth distributed authentication system

<http://shibboleth.internet2.edu/>

Trac Project Management

<http://www.edgewall.com/trac/>

Virtualisation and IRs – Part 1

<http://sophiaca.wordpress.com/?p=5>