Virtualisation: A case study in database administration laboratory work

Greg Cranitch and Michael Rees
School of Information Technology
Bond University, Australia

This paper discusses the issues involved in using virtual machines to teach database administration concepts and the associated issues in a university student environment. Previous work on using virtual machines in system/network administration university labs is reviewed as well as the use of virtual machines in a database development environment. A virtual machine project for a Virtual Information Technology Teaching Laboratory (VITTL) using central servers offering a potential solution is described. This solution provides a secure environment with each student isolated from others with their own virtual machines.

Keywords: Database administration, virtual machines, student labs, student mobility, virtualisation

Introduction

Data is now widely regarded as a valuable corporate asset that needs to be carefully managed and protected from unauthorized access. Rob and Coronel (2009) have suggested that the database administration function requires staff to cover planning, requirements gathering, logical transaction and database design, physical design and implementation, testing, operations, maintenance and training/support. The requirements gathering, development and testing areas can be covered using a central server with desktop development tools. This is completed in our first year database course.

The planning, physical design and implementation, and operations and maintenance are difficult to practice in an environment where machines are shared by more than one person or group. Only one person should be administering the database at a given point in time.

A typical student teaching environment at a university has a limited number of physical servers available with a limited number of licenses for a particular DBMS. It is possible to get multiple licenses for teaching for a very small charge from the Oracle Academy, the MSDNAA or the IBM Academic Initiative. However, it is usually not possible to have one physical server per student.

The shared servers that are available are accessed from on campus teaching laboratories that are used for supervised, timetabled practical sessions and unsupervised open access. Outside of these times the servers are available for unsupervised open access from the teaching laboratories. In addition, they can often be accessed from off-campus via a VPN or other secure off-campus access. University IT administrators are naturally very concerned about what students do on university equipment when granted significant administrator access. As a result the DBMS teaching environment usually suffers a number of artificial access and operational restraints that are far removed from real-world DBMS environments. Teaching scenarios are therefore restricted.

Students also have to balance their life among study, family and work commitments. Many students have off campus internet access and a variety of hardware and software for use with their university work. Consequently, students often want to spend as little time as possible in university provided labs but still want access to licensed software..

Virtualization is not a new phenomenon. Christopher Strachey from Oxford and John McCarthy from Stanford used time sharing on large fast computers as early as 1959. Each user appeared to have control of the machine and the operating system kept users from accidentally interfering with each other. To each user and their application the machine appeared to be different. IBM has used virtualization on their mainframes. Virtualization could also be considered as allowing a machine to emulate a totally different machine. It is also possible to configure a computer so that it behaves in a manner different to its standard configuration. It is possible to have a Linux machine run a Windows® machine inside while running other Linux applications concurrently.

Database administration labs

The aim of the database labs in this subject is to allow students to experience several scenarios that they may encounter when they have graduated. The student needs to be able to make mistakes and learn from these without creating havoc for other students. They are also required to solve problems for which no standard solution has been provided to them.

Problem

In our Advanced Database subject students are required to install a DMBS with a known sample schema, create users and experiment with security, backup and recovery, performance tuning and transaction concurrency control. This is very difficult in a single machine environment as there are limited workstations and only one database server. It is not possible to give one workstation to each student nor is it feasible to give the database server to each student for a set time. One possible solution is to create a virtual machine with a set configuration – disks, memory, OS and DBMS installation software. There are a range of free tools including the Xen[tm] virtual machine monitor[i] and easyvmx[ii] and the commercial ESX Server from VMware[iii].

Previous work

VMware (2005) reported in their white paper that "one of the most significant challenges facing development teams is the loss of productivity resulting from team members contending for limited hardware resources". The high hardware costs were also a concern. These are also a concern for university administrators. Virtualization has been attempted in a database development environment as well as in university labs on networking and system administration.

Virtualization for database development
Database developers need to develop and test their applications in a know environment. They need a specific version of the DBMS and operating system. They can have individual machines with all know configurations as a series of virtual machines that reflect the configurations. Students have a similar experience in a database development and management class to developers in a development environment. Students are competing for a single resource during the administration component – database administration of the installation.

It is possible to treat the server as a scarce resource and to schedule time for each student. The student performs their work and ties up the server for their allocated time slot. During this time the server is sitting idle most of the time. When the session is completed the server is then rolled back to the known initial state for the next student. During the exercise students need access to the web for resources to help with issues they have identified.

This is a significant waste of expensive resources while a student reads and tries to solve the issues associated with the lab exercises. This also means that a student must complete the lab exercise in one sitting and that no lab can depend on a previous lab. Students have different abilities and consequently they require different amounts of time to complete the same task.

Fritchey (2007) reported on virtualization for database development. The development teams had a virtual server isolated from the production environment. The team treated virtual machines like software with machines being checked in and out. Security was a major issue while trying to synchronize permissions between real and virtual machines. The developer could work in one environment but needed different permissions in the other. Also the virtual machines were usually 20% slower than the real machine and the developers were horrified at the degradation.

In a previous semester, our students were allowed to install SQL Server on a workstation dedicated for their sole use in the subject. The workstation was available to other students at other times and consequently many of the permissions that were needed where not available. This is similar to what Fritchey (2007) reported. The database students were locked into one machine for their database work.

Virtualization for network labs
There are numerous articles on virtual machines and system/network administration. Stockman, Nyland and Weed (2005) discussed three approaches to distributing the virtual machines for use in their networking/systems administration lab. Their students needed sole access to the administration features of their machine. This is also a requirement for our database administration students.

One option was for each student to download a copy of the virtual machine to a removable device. The student was then responsible for the safe keeping of the machine. For small virtual machines this was inexpensive for students and students could complete the work at their own pace and save the environment between sessions.

Another option was to download the virtual machine from the central server for each session. This created significant network traffic at the start of each class session as all students tried to download at the same time. Students already complain about the time it takes for their workstation to start at the beginning of each class.

The third was to provide access to centrally stored virtual machines located on a NAS. This did not require students to download their machines at the start of each class.

All three approaches required a virtual machine player loaded on the individual student workstations. In their situation the lab was isolated from the rest of the network so students were not able to have an impact on the rest of the network.

They also raised a concern with student mobility and virtual machines. In one of their implementations students were locked into using a fixed machine for the quarter similar to our previous attempt with a dedicated machine for each student for a given subject while the machine could be used by other students. If another student was using it then a student had to wait until the other student finished even though there may have been vacant workstations near by. One of our requirements is that a student should have complete mobility on and off campus when using their virtual machine.

All three approaches are regarded as a security risk by our central IT administration when used in a non isolated lab. With a virtual machine player available on the desktop there is a risk that any operating system could be run from a removable device and be configured to use NAT and gain access to the external Internet.

Possible solution

One possible solution adopted by our school is to use one or more virtual machines for each student running on a central server. As Wiseman, Wong, Wolf, & Gorinsky (2008) point out "Virtual laboratories are a potential replacement for standard laboratory facilities. Use of these virtual resources can reduce cost and maintenance overheads for teaching institutions while still ensuring that students have access to real equipment". While physical teaching labs containing standard workstations can be released for general use there is still the cost of the server hardware needed to support a virtual laboratory to be found.

To use a central virtual laboratory it is necessary for a virtual machine to be designed and built by the lecture concerned. To this end a virtual machine is created with VMware Workstation, or another free tool such as easyvmx or Xen. The virtual machine needs to be as small as possible as it has a single purpose. One of the major reasons to keep the disk space compact is to add some restraints as the student is granted complete control of the virtual machine. With such freedom they may be tempted to use additional disk space for private purposes not relevant to the subject for which the virtual machine is provided. From the lecturer's perspective the sequence of activities is:

1. Install VMware Workstation on a trusted staff machine.
2. Start the virtual machine with a bootable operating system installation CD/DVD in the drive and install the operating system. For the database subject the base operating system of choice was Windows Server 2003 Enterprise Edition

3. Configure and test the virtual machine making sure that it has no access to any other machines but it can be accessed from machines on the campus network.
4. Leave a copy of the DBMS installation software on the virtual machine.
5. At his point an IT technical support person will make copies on the central server, rename the virtual machines, create local accounts with the required access permissions with agreed passwords, and allocate static IP addresses.
6. The lecturer will then distribute the IP address and password for each virtual machine to the students.
7. Students may then use the Remote Desktop Connection client, present on all Windows machines, to connect to the virtual machine(s) allocated to them. The central IT administration sets agreed access permissions from the teaching labs and other on-campus workstations. Remote access from any Internet workstation can also be set up.

It is possible to create virtual machine images that will fit on a DVD when zipped. Students could then do their work on their own machines off the campus network provided there was no need to update their virtual machine images on the central server. However, this was a concern as the required software could only be installed on university owned equipment. Students were not given a copy of the virtual machine.

Central virtual machine server environment

The student virtual machines for the database subject were hosted on the Virtual Information Technology Teaching Laboratory (VITTL)[iv] a university teaching project managed by the second author. The hardware and network components are shown in Figure 1.

The central host servers, VITTL1 and VITTL2, are IBM 3950 machines with 8 Xeon processors and 40 GB memory, each interfacing with a IBM 3400 Storage Access Network (SAN) with 3 TB of storage. VITTL1 and VITTL2 run VMware ESX Server 3.5 already licensed by the university for basic infrastructure purposes. The SAN is composed of 12 drives organized as two RAID 5 clusters. Redundant network paths connect the SAN to the two host servers providing network connection resilience between the hosts and their storage.
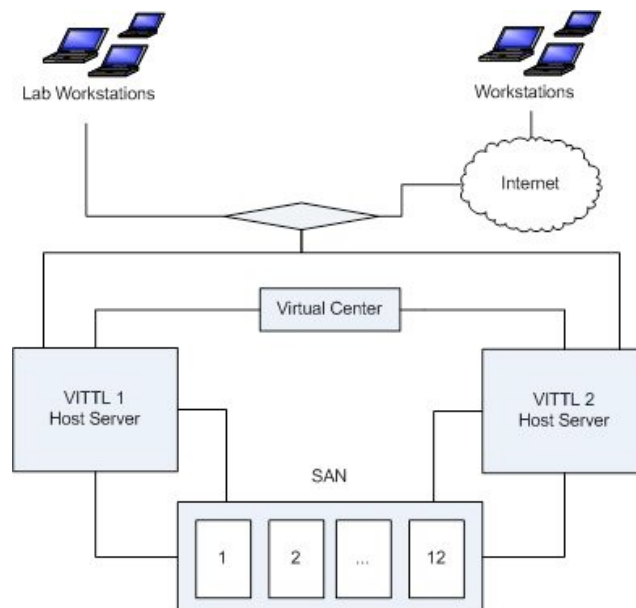

Figure 1. Virtual IT laboratory host environment

The VITTL hardware is controlled from a Virtual Center control server running Windows 2003. A Virtual Center client can be run by the VITTL administrators to control all aspects of the virtual machines being hosted such as installation, configuration of virtual devices, operation, and performance. The virtual machines run in their own VLAN under the central control of the university IT administrators.

The servers too are fully redundant in both a hardware and operational sense. One of the powerful features of the VMware ESX server controlled by Virtual Center is the ability to migrate a running virtual machine from one host to another. This can be used for load balancing, and, most importantly, when a host server is failing. In the latter event it means that a student's virtual machine will still run in the event

of a host failure. Obviously the remaining host must now support additional virtual machines and will run more slowly.

Once installed the virtual machines are available to the students on a 24/7 basis and are granted Internet access for downloading additional software. Outside scheduled classes in the teaching labs they can use their virtual machines directly from other on-campus workstations. In addition, they have virtual machine access from their own laptops via the university wireless network and from any Internet-connected workstation off campus. A Citrix server accessed via the Citrix client is used to support this off-campus access. The virtual machines run without stopping so extended downloads and long-running processes can be supported.

The virtual machines for the database subject are configured to run Windows Server 2003 with 10 GB disk space and 1 GB of RAM. Each machine had the ORACLE 10g install software on the desktop. There was no loopback adaptor installed.

Experiences

The subject is a level 3 subject so students had some previous IT experience. In semester 3 of 2008 there were 8 students enrolled. A detailed handout of the required laboratory tasks was given to each student one week before they were given access to the virtual machines. They were encouraged to read the whole document before commencing the exercise online. Each student was given their virtual machine ip address, user id and initial password at the start of a laboratory session. Each student was able to gain access to the allocated virtual machine with no apparent issues.
Of the 8 students three rendered their virtual machines inoperable in the first 30 minutes. Two installed the Microsoft loopback adaptor required by Oracle when using DHCP and removed the real virtual network adaptor. They logged off and could not get back in. The third started playing with the network from the virtual machine and tried to setup a network to bypass the security and rendered the virtual machine useless. Another student decided to shut down the virtual machine only to discover that there was no on button available. These problems were resolved by the school's system administrator who repaired the machines and restarted the fourth. Once these were sorted out they progressed through the lab.

All students installed Oracle with no major problems other than those identified above. When students accessed their virtual machine over the network there was no noticeable degradation while they executed scripts that generated significant traffic so that their individual virtual machines were running at 100% CPU capacity.

During the labs students received many error messages from the DBMS. With access to Oracle Technology Network, students were able to find solutions for their individual issues. Each student had sufficient rights to the virtual machine to complete tasks necessary without contacting central IT support to resolve the issue. Students only needed to connect to the remote desktop of their virtual machine once. They could then attach to it using a web browser, SQL Developer or jDeveloper installed on the desktop workstation. This meant that students were able to simulate a client/server setup with one server running all the virtual machines. The server hosting the virtual machines was backed up on a regular basis so that each student did not need to backup their own complete virtual machine.

It could be argued that this is an expensive solution when compared to running VMware Player on each lab workstation. The VMPlayer is free. The player and virtual machine grab their memory and the workstation uses the rest so there is no additional cost. The student could load and keep the virtual machine on their network drive which would be similar to the NAS solution proposed by Stockman, Nyland and Weed (2005). Unfortunately, students could create their own virtual machines from scratch with whatever network options and operating system they choose and then run these from within the university network. This security risk justified the additional expense of our solution.

This is the second semester using virtual machines and much experience was gained from piloting the use of virtual machines for teaching two other subjects during the previous semester in 2008. Two low capacity existing servers with only 4 GB of memory were repurposed to act as virtual machine hosts. An extra 8 GB of memory was added to one server that had to support 20 student virtual machines. The two subjects were very different:

1. Computational finance subject using specialized financial modeling software with 5 students. The lower capacity host server was used here. Within a very short time it became clear that the financial

model simulations ran more slowly than needed even though the models could be left running for a few days in order to complete. The lecturer was forced to give up his virtual machine to maximize the capacity for the students. Real-world software was able to be used but the slow execution elicited some negative comments from the students.

2. Web applications subject with 20 students. Here the disk space and memory for each virtual machine was lower. Although these virtual machines all ran together non-stop the processor and memory loads were significantly lower and the machines ran faultlessly for the whole semester. The students enjoyed the flexibility of downloading, installing and testing different web servers, web browsers and web application development tools leading to a realistic and valuable learning experience. The students' reports were positive.

The students this semester have all expressed positive comments regarding the experience. One student was very positive about being able to try different storage structures in his assignment. This would not have been possible without the administrative access rights. Another was pleased that the topics covered in theory could be fully explored in practical sessions.

As the lecturer involved with the database subject I was pleased with the outcome even though 50% had problems with their virtual machines. In early 2009 I had a first year second semester class of 18 students using virtual machines hosted in a similar way. No students reconfigured the hardware or shutdown their virtual machine.

Conclusion

There are several different ways of providing a virtual machine for student use to teach database administration. Depending on the institution's attitude to risk the NAS solution and the ability to isolate a lab used by Stockman, Nyland and Weed is a cheaper solution but does not offer the flexibility of use from student laptops or from their place of residence. Server-based virtual machines running on hardware with adequate capacity are more expensive but offer more flexibility to the students and network safety to the central IT administrators. If it is not possible to isolate a lab then the server running the virtual machines described here is a good choice.

With careful planning and testing it is possible to create a virtual machine with sufficient memory and disk space so that a student can complete the required work for a subject and not be able to load additional software that is not required.

Acknowledgments

References

Fritchey, G. (2007). Virtualization for Database Development. Retrieved April 9, 2008 from http://www.sqlservercentral.com/articles/Development+Process/61675/#.

Rob, P. & Coronel, C. (2009). Database Systems: design, implementation and management, 8th ed., Thomson.

Stockman, M., Nyland, J. & Weed, W. (2005). Centrally-stored and delivered virtual machines in the networking/system administration lab. *SIGITE Newsletter,* Vol. 2, No. 2, pp.4-6., Retrieved April 9, 2008 from http://doi.acm.org/10.1145/1072968.1072969

VMware. (2005). Accelerate Software Development, Testing and Deployment with the VMware Virtualisation Platform. Retrieved April 10, 2008 from http://www.vmware.com/pdf/dev_test.pdf

Wiseman, C., Wong, K., Wolf, T. & Gorinsky, S. (2008). Operational experience with a virtual networking laboratory, *SIGCSE '08: Proceedings of the 39th SIGCSE technical symposium on Computer Science Education*, New York, NY, USA: ACM, 2008, pp. 427–431. Retrieved May 20, 2008 from http://doi.acm.org/10.1145/1352135.1352280

Contact author: Greg Cranitch, School of Information Technology, Bond University QLD 4229.
Email: gcranitc@bond.edu.au

i http://www.cl.cam.ac.uk/research/srg/netos/xen/

ii http://www.easyvmx.com

iii http://www.vmware.com

ivhttp://vittlproject.pbwiki.com/