Application performance testing in VMware environments

Identify and control performance and capacity risks

Table of contents

Introduction

VMware Infrastructure allows IT organizations to consolidate underutilized servers, reduce power consumption, improve agility in introducing new business services, and improve disaster recovery and business continuity planning. It does so by introducing a level of abstraction between the underlying hardware and the operating system that run the business applications. This abstraction breaks the tight coupling between the software that provides the business service and the hardware that it resides on, making it possible to host multiple operating system images on a single hardware box as well as migrate live operating systems from one box to another.

However, this abstraction alters the direct and straightforward relationship between application performance characteristics such as response time and throughput and the hardware resource and capacity utilization. It is easy to see that adding more memory to a memory-bound server running a single multi-user business application will support more users. It becomes difficult to manage when the server is running multiple business applications, each within the confines of its operating system. Similar challenges apply to CPU, I/O and networking bottlenecks. In addition, virtualization introduces configuration options that must be tuned for optimal performance: allocation of CPU, memory, disk, and network resources to virtual machines (VMs), which storage to use, which server, and priority given for those resources.

The purpose of this white paper is to review available techniques to ensure that application performance and availability service level agreements (SLAs) continue to be met when business applications move to a virtualized environment. The paper will also explore how performance testing of virtualized environments with HP Performance Center software can supplement these efforts and avoid unpleasant surprises during production.

Performance and capacity planning techniques

There are a number of performance and capacity planning tools and techniques, of varying degrees of complexity, to choose from when consolidating existing applications to a virtualized environment.

Rough sizing estimates

This technique relies on collecting and analyzing information about the existing systems, their average and peak utilization, to come up with a rough size estimate for the consolidated system. The analysis can be automated using the free, online tool: *HP VMware Solution Sizer* [Ref. 1]. To use the tool, the user supplies information on:

- existing servers (type and number of CPUs, server model, memory)
- operating systems
- current and peak CPU utilization
- memory utilization
- disk and network utilization
- desired configuration of the virtualized environment including desired consolidation strategy
- number of VMs planned and their disk size
- NIC redundancy
- Vmotion
- HP platform of interest
- target utilizations

The Sizer utility then reports back a recommended configuration. For more details on this tool and other considerations regarding migration to a virtualized environment, refer to the *VMware Infrastructure 3, planning* white paper [Ref. 2] from HP.

The challenge in effective use of this tool is to get the correct and up-to-date information on existing servers and their utilization levels. Also, it requires the user to identify consolidation strategy and work out details of the target systems early in the consolidation lifecycle.

Capacity planning assessment services

Capacity planning assessment services automate most steps in a sizing exercise. Typically offered through a professional services engagement, these services offer a detailed view of the current server environment and recommend changes for consolidation through virtualization. This service typically includes:

- agentless discovery of candidate IT infrastructure being considered for consolidation
- collection and analysis of performance and capacity metrics through agentless monitoring software
- identification of target servers for consolidation
- preparation of a roadmap for server consolidation opportunities

VMware has a Capacity Planner utility and associated service which automates many of the aforementioned steps and is offered through authorized consulting partners. The same service is also available from HP [Ref. 3].

Through automation in discovering the existing infrastructure, collection of performance data on current utilization levels, use of sophisticated analysis software, and professional expertise, a capacity planning assessment service often provides much better results than rough sizing estimates.

Benchmark studies

Though not strictly a decision-making tool for consolidation tasks, benchmark studies comparing performance with and without virtualization provide valuable insight into expected performance under virtualization. These studies often also expose factors that may impact performance.

- *Microsoft Exchange Server Performance on VMware ESX Server 3* [Ref. 4] analyzes Microsoft Exchange Server capacity with and without VMware Infrastructure and reports response time and system capacity (number of users supported with acceptable response time) under various configurations. The results are of interest to anyone planning to move their Exchange servers to a virtualized environment. An interesting point observed in this study is that a single Exchange 2003 instance running in a physical machine is limited to using 4GB of physical memory whereas with virtualization it is possible to deploy multiple VMs on a powerful server and support larger numbers of users overall.
- *Using VMware ESX Server with IBM WebSphere Application Server* [Ref. 5] reports best WebSphere Application Server performance for IBM Trade6 benchmark application, both in terms of throughput and response time, when virtual CPUs map 1-to-1 with real CPUs. It reports better throughput with UniProcessor (UP) VMs compared to Symmetric MultiProcessor (SMP) VMs. The virtualization overhead with UP VMs is reported to be negligible for 8-way systems and within 8-13% for 4-way systems.

However, a naive extrapolation of these results into your specific scenario can be dangerous. These studies publish results for specific applications using top-of-the-line hardware, professionally tuned configurations, somewhat artificial work load, and many other favorable conditions. The same conditions and assumptions are unlikely to hold in your specific scenario. Also, too much focus on raw performance without regard to meeting service level agreements and overall cost of ownership could be counter-productive.

The case for performance testing

As mentioned earlier, sizing estimates and capacity planning assessments are useful tools in planning migration of existing servers to a virtualized environment. However, there is no guarantee that when you switch your user base to the virtualized environment everything will work according to the initial estimates and plans. Even with best of planning, one or more of the following could happen:

One or more of the assumptions turn out to be invalid. Capacity assessment exercise works on a number of assumptions and some of these assumptions may not hold for your application and work load. Or the data collected had errors and the situation changed from the time the assessment was done.

The recommendations of the assessment are not fully implemented. Perhaps the actual deployment setup differs from the recommendation of the assessment. Or perhaps you didn't go through a full assessment, relying on existing expertise and gut-feel, but would now like to validate your setup before production rollout.

The configuration of hardware resources and mapping to virtualization needs tuning. There are many different ways to package existing applications into UP or SMP VMs and even more ways to allocate virtual resources to them, such as virtual CPUs, RAM, storage and networking resources. Bottlenecks in the configuration can have significant impact on overall performance and capacity.

You underestimated the impact of live migration (i.e., VMware vMotion). VMware Infrastructure 3 (VI3) supports advanced features such as Distributed Resource Scheduling (DRS), High-Availability (HA), and Distributed Power Management (DPM) using operator or policy driven VMotion. But is your application and infrastructure tested for that?

All these considerations introduce uncertainty and could be a hindrance to adopting virtualization.

The good news is that you can address these potential issues by load testing the virtualized environment running your applications before going live. Issues identified during the load testing can then be addressed without inconveniencing the users.

Load testing with HP Performance Center software

HP Performance Center is an enterprise class performance testing solution built on HP LoadRunner, industry-leading software for load testing. Using custom-generated virtual users, HP LoadRunner allows you to drive actual workloads and capture end-user response times of key business processes and transactions to determine whether service-level agreements (SLAs) are met at expected load conditions with acceptable resource utilizations. Non-intrusive, real-time performance monitors obtain and display performance data from every application tier, server and system component, and diagnostic probes gather code-level data to isolate bottlenecks down to the SQL or method level. This combination of end-user, system-level and code-level visibility dramatically reduces time-to-problem resolution. [Ref. 6]

Capabilities of HP LoadRunner software

Performance testing with HP LoadRunner will address the following issues in virtualized environments:

Application scalability assessment: Will the applications within the virtualized environment scale to the actual number of users and load? It is often much easier validate this before going live than to face the ire of irate users and tweak the production system.

Business process performance: Do business processes supported by the applications meet the accepted response time, availability and throughput under expected load conditions?

Capacity/hardware utilization: Is the hardware (compute, memory, networking and storage resources) getting utilized properly? More importantly, are you getting the savings promised by virtualization under similar load and usage conditions?

Problem identification and assessment: Can you identify performance and/or scalability bottlenecks?

Virtualization environment feature testing: Do the VMware Infrastructure features such as Distributed Resource Scheduling and High Availability work under load conditions for your setup?

How does HP LoadRunner work?

Here is a very high level overview of how HP LoadRunner works:

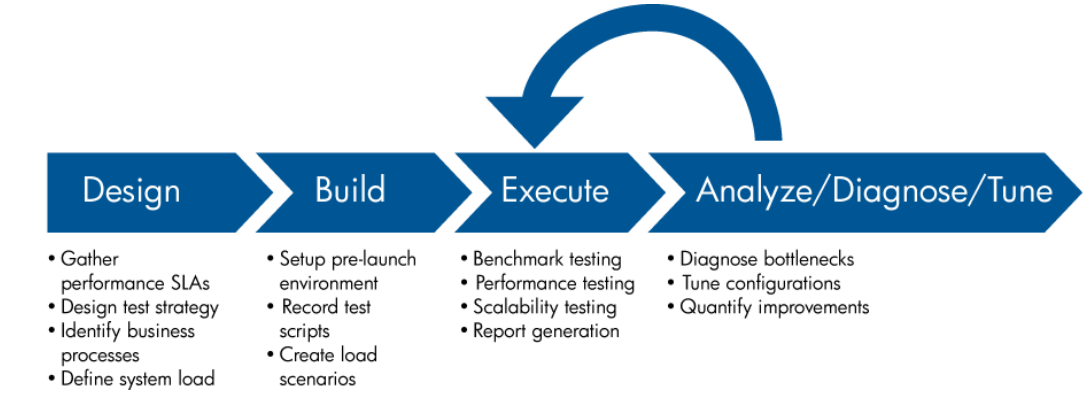
1. User records actions against the application and saves them in scripts
2. LoadRunner connects to load generators to begin test run
3. Virtual users inside load generators perform work as per the scripts
4. LoadRunner collects monitor data agentlessly
5. Diagnostics drill down into why performance is slow
6. Controller correlates end-user response times to system monitor data and creates charts and reports for inspection

Out-of-the-box support for a large number of protocols and performance monitors make this whole process fast and simple. Protocols are application specific mechanisms that clients, or virtual users, use to interact with the application. These protocols range from as simple and generic as HTTP/HTML for web applications to application specific ones such as SAP specific interaction between SAP GUI and SAP backend systems. The performance monitors represent client side performance metrics and server side hardware utilization and related activities as reported by the operating system and the application itself.

Performance testing lifecycle

The following diagram illustrates a typical performance testing lifecycle. An important point about virtualization led server consolidation projects is that much of information about the use of the system is already known and could easily be fed into the design and build phase.

Figure 1. The HP performance testing lifecycle.



What value does HP Performance Center add?

There are other components within HP Performance Center that enable either the global standardization and formation of a performance testing Center of Excellence or a performance testing IT shared service. It gives everyone, from LOB managers to testers, project-level visibility and a collaborative working environment. Efficiencies that arise from centralization and standardization, such as common skills sets, common hardware and licenses, and around-the-clock follow-the-sun usage help IT organizations test effectively at lower cost.

Use of HP Performance Center is particularly helpful in coordinating tests and sharing results in virtualization led server consolidation projects where many applications are moved en-mass and need to be performance tested on the same virtualized environment by different application owners.

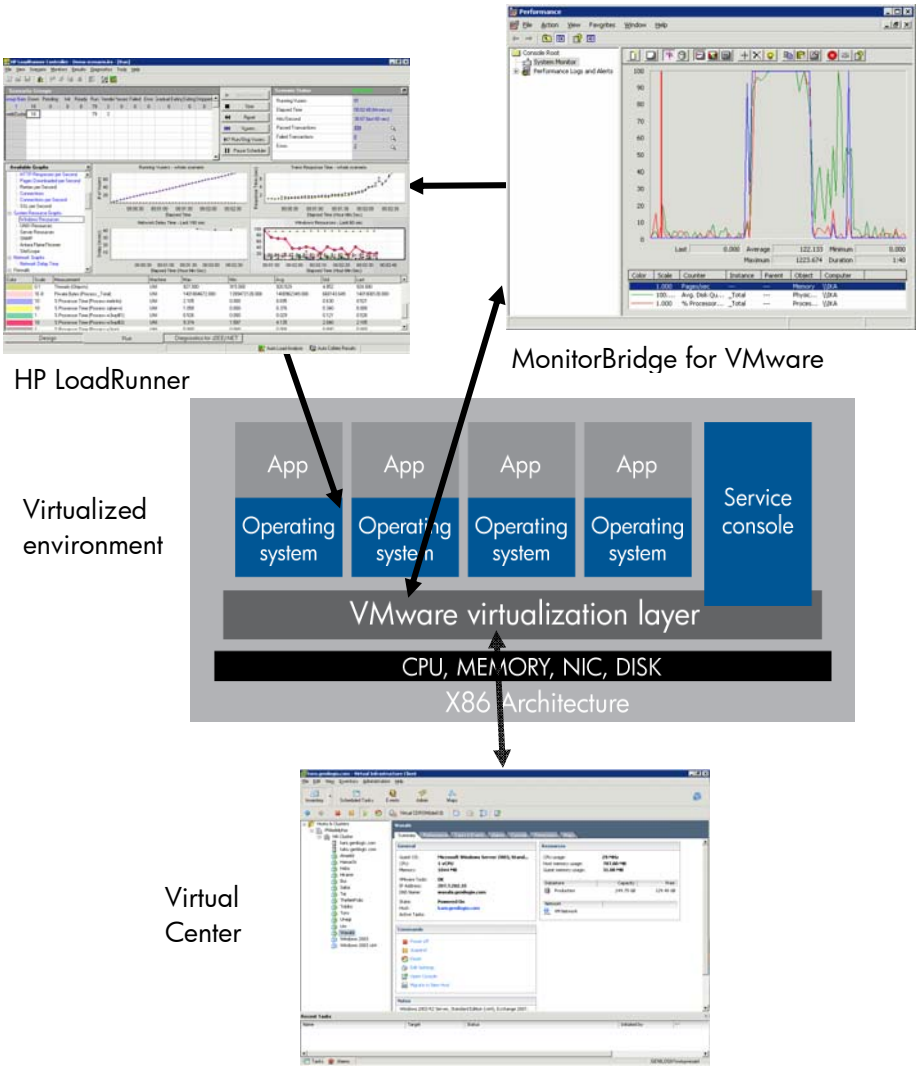
Genilogix MonitorBridge™ for VMware

HP LoadRunner gathers operating system and application performance monitor data such as CPU, memory, storage and networking utilization and correlates those to application performance characteristics such as response time and throughput. In a virtualized environment, the operating system monitors, such as CPU and memory utilization, disk and network activity, are reported by the guest operating system running on the virtualized environment and tell only half the story. The missing half is the physical system monitors as seen by the VMware ESX Server. You need both, the virtual and physical monitors, to get a complete picture of overall resource utilization and identify bottlenecks. There are times when a guest operating system reports 100% CPU or memory utilization but the underlying machine may still have idle capacity. This often points to misallocation of resources and, once identified, can easily be resolved.

Genilogix MonitorBridge for VMware [Ref. 7] is an add-in for HP LoadRunner that allows VMware ESX server statistics from multiple hosts to be agentlessly collected and seamlessly incorporated into LoadRunner's online monitoring and analysis charts and reports. The combined virtual and physical resource utilization view is much more helpful in identifying resource bottleneck issues due to configuration or setup errors.

The following diagram shows how MonitorBridge for VMware fits-in with VMware Infrastructure and HP LoadRunner.

Figure 2. Genilogix MonitorBridge provides a link between HP LoadRunner and the virtualized environment.



One customer reported that after performance testing their CRM system they noticed a high number of timed-out requests in certain conditions. An analysis of performance monitors recorded by HP LoadRunner revealed that the CPU utilization reported by a particular VM was noticeably high. At the same time, the monitors collected by Genilogix MonitorBridge for VMware, and included by HP LoadRunner in final analysis, indicated that the host machine had fairly low CPU usage. It became clear that this VM was not adequately provisioned with compute resources. Once identified, the fix was simple and the customer was able to put this VM into full use.

Continual performance monitoring

Pre-production performance testing and tuning makes you better prepared for real usage in production. However, this one time exercise provides no guarantees against future changes in load. For mission critical applications we recommend continual monitoring of application performance and resource utilization to make sure that business applications continue to perform as expected and additional capacity is planned as system usage increase. HP Business Availability Center software,

which includes HP SiteScope software, can be used to monitor business application performance and availability. HP Operations Center software, which includes HP Performance Agent software, can be used to monitor resource utilization at both virtual machines and ESX server level.

Conclusions

A large number of organizations hesitate to embrace virtualization despite the flexibility and cost reductions it offers due to uncertainty around performance, availability and capacity management of a virtualized environment. The good news is that these uncertainties can be addressed by proper capacity assessments, pre-production performance testing and post-production monitoring using the same tools currently used in physical environments. Additional benefit of doing performance testing is and early and objective assessment of cost savings by actually measuring the response time, throughput, and capacity utilization under a controlled environment.

References

[Ref. 1] HP Proliant server sizer for VMware ESX Server.
http://h71019.www7.hp.com/activeanswers/cache/120132-0-0-0-121.html

[Ref. 2] VMware Infrastructure 3, planning white paper.
http://h71019.www7.hp.com/ActiveAnswers/downloads/VMware_Infrastructure_3_planning.pdf

[Ref. 3] IT Consolidation using VMware Capacity Planner.
http://h71019.www7.hp.com/ActiveAnswers/cache/70314-0-0-225-121.html

[Ref. 4] Microsoft Exchange Server 2003 Performance on VMware ESX Server 3.
http://www.vmware.com/pdf/Virtualizing_Exchange2003.pdf

[Ref. 5] Using VMware ESX Server with IBM WebSphere Application Server.
http://www.vmware.com/partners/vmware/ESX_WAS_WP.pdf

[Ref. 6] HP Performance Center software, datasheet.
http://h71028.www7.hp.com/ERC/downloads/4AA1-2119ENW.pdf

[Ref. 7] Genilogix MonitorBridge for VMware.
http://genilogix.com/pdfs/monitorbridgeforvmware.pdf

For more information

www.hp.com/go/vmware/management

www.hp.com/go/vmware

www.vmware.com/hp

www.genilogix.com/vmware

4AA1-5799ENW, October 2007