



A White Paper by Stratus Technologies
June 2010

**Server Virtualization and Cloud Computing:
Four Hidden Impacts on
Uptime and Availability**

Abstract

As x86 virtualization has become established and cloud computing emerges, a robust IT infrastructure actually matters more than ever. Software aspects get most of the attention partly because they are new. Despite the hype, not every technical challenge is better solved by software alone because of the tradeoffs involved. As more IT organizations look to move beyond simple server consolidation to virtualize tier-1 applications and include cloud infrastructure designs in their IT strategies, the need for complete, bullet-proof availability increases dramatically. For high-performance, high-density or mission-critical services and applications, using fault-tolerant server hardware hardens against new vulnerabilities and complements the resilience you can achieve with virtualization and cloud computing.

Virtualization and the Cloud: Where Are We Now?

Early adopters of x86 virtualization focused on server consolidation, and they have benefited from cost savings and increased flexibility. Now they look forward to using virtualization to help them enable cloud computing, the next big trend on the horizon.

The cloud concept takes virtualization a step further by enabling users of IT resources to avoid investing in dedicated infrastructure. IT costs become a variable operational expense for business users because capacity is shared. Adopting this mode allows capacity to scale up and down dynamically and immediately in a manner that advances how virtualization is used today. The cloud model is designed to let companies use IT resources as a service, taking advantage of shared applications, processing and storage managed within the cloud – either inside a private cloud at an internal data center, or in an external cloud at a service provider.

Cloud Computing Defined

As in any rapidly developing area, there are multiple definitions of cloud computing and what it constitutes. The type of cloud computing discussed in this paper includes both internal and external clouds. VMware, whose vSphere™ 4 is positioned as the industry's first operating system for building the internal cloud, uses this definition:

“Cloud computing is the use of networked infrastructure software and capacity to provide resources to users in an on-demand environment. Sometimes known as utility computing, clouds provide a set of typically virtualized computers which can provide users with the ability to start and stop servers or use compute cycles only when needed, often paying only upon usage.”¹

At a fundamental level, the value proposition promised by cloud computing remains much the same as for virtualization: a means to reduce expenses and improved agility to meet changing business needs. Using virtualization to enable disaster recovery (DR) and business continuity has become another significant selling point as adoption of virtualization technology has grown.

¹ “VMware Cloud Computing FAQs,” retrieved May 8, 2009 from <http://www.vmware.com/solutions/cloud-computing/faqs.html>

While virtualization solutions are a known quantity and deployed in real-world IT environments, cloud computing is still taking shape. In reports published in early 2009, Forrester Research estimated that 54% of enterprises had implemented x86 server virtualization or planned to do so within the next 12 months. Some 53% of small and medium businesses had already implemented x86 server virtualization or planned to do so within the next 12 months. In the cloud computing area, Forrester found 5% of enterprises had already implemented pay-per-use-hosting of virtual servers, with 3% more implementing within the next 12 months. Among SMBs, 2% had already implemented pay-per-use-hosting of virtual servers.²

Figure 1: From Server Virtualization to Cloud Computing

Today	Tomorrow
<ul style="list-style-type: none"> • Server virtualization is the major trend • Virtualization initially driven by hardware consolidation and the resulting savings in capex, power, space • Business agility, disaster recovery and business continuity became additional drivers • Virtualization solutions are established; cloud computing solutions are emerging • Overall, the requirement for IT expertise is still significant for virtualization and internal clouds • Management and monitoring capabilities are still developing <ul style="list-style-type: none"> - Ongoing IT administration is required - Policy making is not automated • Vendors have not implemented virtualization and cloud computing standards • Security, privacy and compliance issues still to be addressed for sensitive applications/services • Most external cloud service level uptime agreements (SLAs) are below mission-critical levels³; for example, Amazon EC2 is at 99.95%⁴ 	<ul style="list-style-type: none"> • Cloud computing is the major trend • Driven by desire to benefit from IT resources provisioned as a service • Pay-as-you-go, “utility computing” models allow business users to avoid fixed costs • Highly elastic, instant access to computing resources and IT services • Virtualization and cloud solutions become mature • Simpler deployment, use will make benefits broadly achievable • Management and monitoring come of age to reduce requirements for specialized IT expertise <ul style="list-style-type: none"> - Self-management - Automated policy-making • Virtualization and cloud standards will increase ease of deployment, reduce risk • Appropriate best practices for security, privacy and compliance will develop • Enterprise users of utility computing will expect resilient cloud infrastructure and specify uptime as high as 99.999% for mission-critical applications and services

² Forrester Research, Inc. press release, [“Forrester: Server Virtualization Reaches A Majority Of Firms, Spurring Initial Adoption Of Cloud Computing.”](#) March 4, 2009

³ ZDNet.com, [“The race to 99.999 percent uptime: 3Tera ups the cloud SLA ante.”](#) March 19, 2009

⁴ “Amazon EC2 Service Level Agreement,” retrieved May 11, 2009 from <http://aws.amazon.com/ec2-sla/>

The abstraction of services and applications away from the physical IT infrastructure that delivers them is an important characteristic shared by virtualization and cloud computing. One obvious example: Provisioning a virtual machine is almost effortless compared with setting up a new physical server. The relative ease of adding and moving virtual machines leads to a dynamic environment — one that unintentionally obscures how service levels are affected by interdependencies between the application software, server hardware, storage configuration and network resources.

Hidden Impacts on IT Infrastructure Availability

Under the physical-to-virtual abstraction introduced by x86 virtualization and cloud computing, a number of not so obvious impacts affect the uptime and availability the IT infrastructure can deliver. Because cloud computing is enabled by virtualization technology, many of the potential issues are common to both. These impacts, in turn, influence how well the virtualized and cloud environments are able to meet the SLAs required by your enterprise users. The implications can be profound in high-performance, high-density or mission-critical environments.

1. Single Points of Failure

Virtualization and cloud computing can be used quite successfully to improve the resilience of an IT environment because they provide the means to recover quickly from component or system malfunctions using failover, and to back up essential applications and data. Virtual machines can be migrated from one physical server to another in a live migration; virtual machine images can be restarted in a different location to provide for disaster recovery.

Downtime exposure grows with the number of virtual machines on any single physical server. It's worth noting that individually, none of the virtual machines that run on a physical server may be deemed mission-critical or even particularly essential. Yet collectively, the virtual machines on that server take on a critical status if an outage or slowdown would have a serious negative impact on your business processes. In addition, in a cloud computing scenario, the number of virtual machines (and types of applications) running on any given physical server may be out of your company's direct control.

Easy provisioning of virtual machines also keeps virtualized and cloud environments in perpetual motion. That very flexibility can introduce single points of failure when too many of the virtual machines that support a particular service or application are concentrated on the same physical server. Close attention must be paid to how many virtual machines reside on a physical server, and which virtual machines are placed on which physical server(s).

2. I/O and Scalability Limitations

Virtualized and cloud environments use software emulation to abstract IT resources away from the physical hardware, which comes at a cost. Overhead is incurred because virtual resources have to be mapped to physical resources, including the physical server's buses, I/O adapters and disks. Applications that are I/O-intensive often experience latency issues in a virtualized environment for that reason.

I/O and scalability concerns increase when a virtualization or cloud solution uses software lockstepping to provide high availability or software-based fault tolerance. Although software lockstep increases reliability, the technique imposes additional overhead on the CPU, network and I/O.

Furthermore, using software lockstepping to achieve high availability or fault tolerance adds complexity and runs counter to the goal of server consolidation. That is because software lockstep depends on having primary and secondary physical systems in place: at least one duplicate server, a duplicate copy of any software and the planning to ensure that failover will work properly. In this respect, software lockstep resembles the server clusters that have offered a high availability alternative for years.

Software lockstep in addition may compromise SLA performance when application response time degrades because of the replication and heartbeat processing necessary to keep the primary and secondary virtual machines in sync.

Another limitation of software lockstep is that it restricts a virtual machine to a single processor core. This precludes a virtual machine from using the symmetric multiprocessing (SMP) or multicore capacity of a CPU, which prevents the application from using all available processor cores to scale up in performance.

3. Failover and Fault Isolation

As mentioned earlier, virtualization software enables failover: the ability to transfer processing from one virtual machine to another virtual machine on a different server in the event of a hardware or application failure (and during scheduled maintenance). Storage capacity may be virtualized as well. Depending on the capabilities of the virtualization or cloud solution you choose, failover may be neither automatic nor instantaneous.

Virtualization and cloud software solutions on the market today cannot isolate and then determine the root cause of a hardware or software failure. Even though failover is provided for, the cause of the failure may be propagated to the secondary virtual machine. That allows the same failure to repeat along with the same threats of downtime and data corruption.

In addition, finding and resolving the cause of a failure within a virtualization software stack poses a challenge due to the increased software complexity and a lack of self-diagnostics.

Even virtualization and cloud solutions that enable high availability or fault tolerance in software are not designed to deal with transient (temporary) hardware errors that can lead to downtime and data corruption when these errors are left unaddressed.

4. Data Integrity

The need for data integrity is a trait of application environments that depend on high performance, have a high density of virtual machines or are mission-critical in nature. As just observed, error-handling and problem resolution have a meaningful impact on data integrity. And virtualization or cloud software alone do not provide the complete protection required.

Avoiding a “split-brain” condition is crucial to preventing data corruption as well. Split brain occurs when a server or servers go down; they reboot or restart; and each virtual machine believes itself to be the primary VM. Selecting a virtualization solution that prevents split brain is good practice anytime you plan to provide redundancy by using multiple physical servers and network paths in a virtualized or cloud setup.

Hardware Assistance from Fault-Tolerant Servers

When used together with software-based x86 virtualization or cloud computing, fault-tolerant server hardware makes a good thing better by alleviating the new concerns that arise around single points of failure, I/O constraints, scalability limitations, failover, fault isolation, problem resolution and data integrity.

While earlier generations of fault-tolerant servers were proprietary and expensive, today's versions are vastly more cost-effective Intel® architecture servers that run Windows® or Linux® operating systems. These servers offer an effective means of hardening virtualized and cloud environments without adding complexity. With an optimal fault-tolerant architecture, software including virtual machines need not be modified in any way to benefit from the server's reliability and availability features.

Eliminating single points of failure is a signature attribute of these systems, which are said to have continuous availability. If one of the server's components fails, its duplicate keeps the server processing without interruption or degradation in performance. Failover is eliminated rather than minimized. For these reasons, such fault-tolerant servers have been proven to sustain 99.999% — “five nines” — or better availability.

Because their fault tolerance is based in hardware, such servers eliminate the overhead of software emulation along with the I/O limitations and scalability constraints that result. True symmetric multiprocessing (SMP) allows applications to scale across multiple CPU cores while in a virtual machine.

Component and functional redundancy within the footprint of a single server also simplifies deployment and support. Automatic fault isolation and diagnostic tools can be built into the server architecture, as well. The server's internal redundancy also protects the integrity of committed (completed) transactions and preserves in-flight data despite a component failure or transient error, such as a driver malfunction. To the extent all this can be done on a single physical server rather than requiring multiple servers, hardware cost of ownership and software license fees are far less complicated and less expensive.

Deploying fault-tolerant servers as part of a virtual resource pool therefore hardens the environment at otherwise vulnerable points. You are also positioned to support tiered levels of service, with the agility to upgrade virtual machines to a higher level of availability protection when needed. This same principle will hold true as cloud computing evolves.

Conclusion

Server virtualization and cloud computing are no exceptions to the rule that every new paradigm in computing brings new benefits (for example, client/server versus peer-to-peer), though not without raising concerns and issues. It's the familiar situation of solving one set of problems only to reveal several new ones.

While software-enabled virtualization and cloud computing allow IT services and applications to appear independent of the underlying resources, in several respects these computing models introduce stress points that demand *greater* robustness of the IT infrastructure. Using fault-tolerant server hardware for selected key roles within virtualized and cloud environments builds in reliability and reduces management complexity, helping to sustain the high service levels that enterprises can count on.

About Stratus Technologies

Stratus Technologies focuses exclusively on helping customers keep critical business operations online without interruption. Business continuity requires resiliency and superior availability throughout the IT infrastructure, including virtual environments. Stratus delivers a range of solutions that includes software-based high availability, fault-tolerant servers, availability consulting and assessment, and remote systems management services. Based on more than 30 years of expertise in product and services technology for total availability, Stratus is a trusted solutions provider to customers in manufacturing, health care, financial services, public safety, transportation & logistics, and other industries. **For more information, visit www.stratus.com.**

Stratus and the Stratus Technologies logo are trademarks of Stratus Technologies Bermuda Ltd. VMware is a registered trademark and vSphere is a trademark of VMware, Inc., in the United States and/or other countries. Intel is a registered trademark of Intel Corporation or its subsidiaries in the United States and/or other countries/regions. Windows is a registered trademark of Microsoft Corporation in the United States and/or other countries/regions. The registered trademark Linux is used pursuant to a sublicense from the Linux Mark Institute, the exclusive licensee of Linus Torvalds, owner of the mark on a worldwide basis. All other trademarks and registered trademarks are the property of their respective holders.

Copyright © 2010 Stratus Technologies Bermuda Ltd. All rights reserved.
X992-A