



Controlling Virtual Machine Sprawl

How to Better Utilize Virtual Infrastructure

WHITE PAPER
NOVEMBER 2012

Table of Contents

Overview 3
 The Roots of Sprawl3
Improve Infrastructure Utilization Efficiency..... 4
 Reduce Unauthorized and Over-provisioned Machines4
 The ROI Potential4
 Reuse Resources Automatically4
 The ROI Potential5
 Recycle Inactive and Abandoned Machines.....5
 The ROI Potential5
Management: The Key to Efficient Resource Utilization..... 6
Additional Information 6

Overview

Virtualization not only improves hardware efficiency via server consolidation, it dramatically reduces the time required to bring a new server online from weeks or months to hours or minutes. This ability to bring machines online quickly provides companies with improved business agility. As with most improvements new issues can be created, and if not addressed, can reduce a virtual infrastructure's overall efficiency over time.

Virtual machine sprawl is one of the biggest concerns facing many companies using desktop or server virtualization. The ability to quickly create virtual machines without the disciplines and controls of the physical world results in machines being provisioned unnecessarily without proper justification and approval, over-provisioned (too much CPU, memory, or disk), or consuming resources well after they no longer are required.

To address this concern, a number of reporting products have been created to help track and identify virtual machine sprawl. The problem with report-only products is they provide only a partial solution. They only help identify the problem and track it whether it is getting better or worse. They do nothing to prevent the problem or correct it once it has occurred. This white paper looks at how a cloud automation platform can help decrease virtual machine sprawl by automating the process of reducing, reusing, and recycling compute resources in a shared physical infrastructure.

The Roots of Sprawl

A look back at the IT industry shows that virtual machines are not the first computing resource with sprawl problems. In general, the easier it is to create something the more of them you get, and the harder they are to clean up. Email messages, files, databases, and storage resources are just some examples of information technologies that suffer from sprawl issues. Reporting is a key component to identifying the problem and scope. When it comes to fixing the problem, reporting is just part of the overall solution. With any technology that is susceptible to sprawl, policies and governance along with automation to help clean up and recycle are necessary components to achieving efficient resource utilization.

Email is an excellent example of sprawl with which everyone is familiar. Most people easily can find out how many email messages they have, how much space they consume, and which messages consume the most space. Those reports, however, do little to prevent email sprawl. The only way administrators control email resource consumption is to implement policies that control the amount of capacity each user can consume and how large and what type of attachments are permissible. Additional policies that automate the archiving of old email messages or the deduplication of common information further optimize email resource utilization.

Server virtualization is no different than other IT resources that caused sprawl concerns in the past. There are two primary reasons why virtual computing infrastructure results in more sprawl issues than their physical counterparts.

1. Speed

Requisitioning a physical machine is a lengthy process with many justifications and approvals required along the way. Even after the machine arrives, getting it racked and operational can be time consuming. The long process of bringing a physical asset online ensures machines are not provisioned needlessly. On the other hand, virtual machines typically can be created in a matter of hours from pre-existing physical resources.

2. Governance

For many companies, the process and tools used for provisioning virtual machines are completely different than those used to provision and manage physical infrastructure. In addition, the tools used to manage physical infrastructure are not appropriate for virtual environments. In virtual environments, hypervisor device managers provide basic management controls and do not enforce the same governance and compliance that are part of the physical management process.

Improve Infrastructure Utilization Efficiency

There are many corollaries between the environmental movement trying to eliminate waste and reduce resource consumption and IT organizations trying to drive higher utilization efficiencies in their virtual computing infrastructure. Reduce, reuse, and recycle are three basic rules that drive the overall environmental movement. These same rules apply directly to the efficient management of a virtual computing infrastructure.

Reduce Unauthorized and Over-provisioned Machines

Reduce is defined as using fewer resources in the first place. Of the three approaches to eliminating waste, reducing the number of needlessly provisioned or over-provisioned machines has the largest potential return on investment. Gaining better control of machine inflow is much easier than trying to weed out unnecessary or over-provisioned machines at a later date.

Many virtualization management tools do not provide the governance and automation necessary to ensure requests for new virtual machines receive appropriate review and business justification. Once the provisioning process starts, there are few limits on what resources can be consumed and how much of those resources each virtual machine should receive. In an effort to limit the influx of new machines, many companies set up administrators who try to manually enforce governance. This model does not scale as the environment grows without adversely impacting operational costs. The ideal solution is to delegate the provisioning process downstream to resource consumers. Yet without appropriate controls and automation, this strategy is impossible to implement without further loss of control.

A cloud automation platform is needed that provides a self-service portal that enables users to initiate provisioning requests for new virtual machines and ongoing resource reconfiguration throughout the virtual machine lifecycle, including its eventual decommission. This portal abstracts the complexities of using one or more virtualization device managers. More importantly, the software automates delivery and enforces governance relative to the resources a given user is allowed to consume. With the appropriate policies in place, administrators control not only how the machine can be built, but what and how many resources can be consumed. For additional control, policies can be established to automate optional approval workflows, further ensuring that machines are not provisioned unnecessarily or over-provisioned.

The ROI Potential

A typical company with 1,000 virtual machines—with five percent of machines created without proper business justification and another five percent over-provisioned—could easily save over USD \$100,000 to USD \$150,000 in capital expenditures through better control of the front-end provisioning process. By delivering the right sized machine at the right service level, companies can eliminate waste, improve resource utilization, and lower costs.

Reuse Resources Automatically

Reuse is defined as using an item more than once. This includes reusing a resource for the same function or a completely different function. One of the core tenets of virtualization is to improve efficiency through improved resource utilization. The degree to which resources can be reused has a high impact on the overall efficiency of virtual infrastructure.

The problem with reuse in a virtual environment is that it is largely a manual process. There are a number of use cases where machines are needed for short periods of time yet continue to exist well beyond the time required. Examples of applications that require machines for short periods include data mining, calculation farms, development, testing, month-end processing, and so on. The same is true for virtual machine archival or snapshot systems.

Cloud automation platforms should automatically reclaim and reuse resources when they no longer are required. Policies control how long a machine exists, and process automation orchestrates the decommissioning and reuse upon lease expiration of the virtual machine. Machines that need to be archived may not consume CPU and memory, yet their storage capacity can be very costly. Policies that control how long a virtual machine is archived after it no longer is needed, and automatically clean up after the archival period, are essential to driving improved efficiencies. Snapshotting is another area that, if not managed properly, has a tendency to grow and consume storage resources unnecessarily. Policies that control whether a machine can have a snapshot, how many snapshots can be created, and how long they are allowed to exist before they are automatically coalesced are essential to efficient resource management.

The ROI Potential

A typical company with 1,000 virtual machines that has five percent of its machines used for temporary applications can expect to save approximately USD \$60,000 to USD \$70,000 by automating the reuse of leased machines. Environments such as development and testing, which have a larger need for temporary machines, can expect even greater savings. The same sized company that archives 10 percent of its machines on a yearly basis can expect to save between USD \$30,000 and USD \$50,000 annually by reclaiming and reusing storage resources after the archive period ends.

Recycle Inactive and Abandoned Machines

Recycling is the process of collecting, processing, remanufacturing, and reusing materials. As with most recycling programs, identification and collection of unused resources tends to be the most manual and time-consuming portion of the process. The same is true in virtual infrastructure, where reclaiming inactive, abandoned, and stranded capacity can be so labor-intensive and time-consuming that the process is performed very infrequently, if at all. As a result, inactive machines tend to stay around longer and consume more resources unnecessarily than when an efficient reclamation process is in place.

The first problem with reclaiming inactive resources is the identification of inactive machines. For most companies, this manual process consists of data collection scripts and spreadsheets. Even companies that invested in reporting software to help identify inactive machines often encounter another problem: filtering out truly inactive machines from those that appear to be inactive. Some machines meet several of the criteria that cause them to appear in an inactive exception report, yet are known to be machines required by the business. If a large percentage of these essential machines keep appearing on an exception report, the reports become ineffective. Administrators cannot see problem machines through the clutter of information. This identification issue highlights the third recycling problem: validating that a virtual machine is truly inactive by confirming with the owner and automating the reclamation process. Report-only solutions are little help, as they were not around during machine creation. They are unable to identify the owner and do not have process automation in place to help verify the machine is not needed and initiate the process to recycle the machine's resources.

A cloud automation platform should provide exception reports that help identify stranded, inactive, and abandoned machines and automate the workflow associated with reclaiming those resources. Report-only solutions solve less than half of this problem and leave the remaining manual work to administrators, decreasing overall process operational efficiency.

The first step is to identify stranded capacity within the host or host clusters used to run the virtual machines. Stranded capacity is caused when one of the three critical resources (CPU, memory, or storage) is completely consumed, yet the remaining resources are in abundance. This capacity is effectively stranded because it cannot be used. Reports that help identify hosts with this condition enable administrators to add capacity to the bottlenecked resource or reallocate stranded capacity to other hosts, improving overall utilization of the virtual computing infrastructure.

The second step is to identify inactive and abandoned virtual machines. Since no single criterion can unequivocally identify whether a virtual machine is needed, the best that can be done is to identify machines that are potentially inactive. To effectively manage the reclamation process, a cloud automation platform needs to be able to eliminate known exceptions from these reports, validate with the owner that the machine is required, and decommission and reclaim the resources of machines that no longer are needed. The problem with many of the solutions available today is that they do little to help with the validation and reclamation process.

The ROI Potential

Even if a company automates the recycling and reuse of resources used for temporary machines, identification and recycling of resources used for permanent machines can be problematic. This is especially true for virtual desktops, which tend to have more churn than virtual servers. A typical company with 1,000 virtual machines and 10 percent of its resources consumed by inactive and abandoned virtual machines can expect to save between USD \$80,000 and USD \$100,000 on capital expenditures annually. Even if the automated process is run more frequently than the current manual process (such as quarterly versus annually), companies can expect to save an additional USD \$20,000 in operational expenses associated with identification and reclamation of inactive resources.

Management: The Key to Efficient Resource Utilization

Infrastructure virtualization has an immediate and quantifiable return on investment. It is easy to compare the cost of a physical asset to its virtual counterpart. Consolidation ratios of 10:1 up to 20:1 are not uncommon for servers, and consolidation ratios are even higher for desktops. If virtual infrastructure is not used efficiently, a large percentage of the savings can be wasted, resulting in actual savings that are far less than what was theoretically envisioned.

Hypervisor device managers provide solid management capabilities but do little to enforce the governance required to reduce the inflow of unnecessary and over-provisioned machines. Report-only solutions help identify inactive and abandoned resources, yet do little to automate the reclamation process. A cloud automation platform that automates the service delivery and ongoing management of virtual infrastructure can justify its cost through improved resource utilization alone. Additional savings result from improving operational efficiencies through process automation and helping accelerate virtualization deployments by providing the tools necessary to manage a growing virtual infrastructure more efficiently.

Additional Information

For more information, visit the [vCloud Automation Center product page](#).

For information or to purchase VMware products, call 1-877-4VMWARE (outside of North America, +1-650-427-5000), visit <http://www.vmware.com/products> or search online for an authorized reseller.



VMware, Inc. 3401 Hillview Avenue Palo Alto CA 94304 USA Tel 877-486-9273 Fax 650-427-5001 www.vmware.com

Copyright © 2012 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at <http://www.vmware.com/go/patents>. VMware is a registered trademark or trademark of VMware, Inc. in the United States and/or other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies. Item No: VMW-TWP-CONTROLLING-VM-SPRAWL-USLET-101