

Server Virtualization: One Path That Leads to Cloud Computing

Gartner RAS Core Research Note G00171730, Thomas J. Bittman, 29 October 2009, RA3 12012010

Virtualization is much more than a tactical cost-cutting trend. Virtualization must be seen strategically as a catalyst of overall IT change and modernization. These changes position IT more as a service provider for the business, helps IT behave more like a cloud provider, and prepares the business to be a better consumer of cloud computing.

Key Findings

- Virtualization should be approached strategically, as a modernization catalyst with far-reaching technology, process, people and business ramifications.
- Virtualization enables four of the five attributes of a cloud computing provider: service-based, scalable and elastic, shared, and metered by use.
- Virtualization is not the only path to cloud computing, but it will be one of the most important and common on-ramps to the cloud computing style (both private and public cloud computing).

Recommendations

- Develop service-level descriptions for virtual resources that are relatively simple, and benchmark what others (especially cloud providers) are doing.
- Make the necessary changes in management tools, operational processes and procedures before virtualization takes root.
- Overcome customer objections to sharing resources by building successful references and proof points in the enterprise, and by using metrics like lower costs, increased speed, or new offerings like disaster recovery.
- Measure and track usage details, but keep chargeback simple and less dynamic until the enterprise is ready for a full usage pricing model.

ANALYSIS

Server virtualization is a major priority for large enterprises, and during the next few years, for smaller enterprises as well. Initially, organizations see virtualization as an effective technology for consolidation – reducing server sprawl, increasing utilization rates, and reducing capital and power/cooling expense. Done well, this is all true. But this is also very tactical thinking.

Strategically, server virtualization is an IT modernization catalyst that will change how IT is acquired, consumed, managed, sourced and paid for. Virtualization will even change how businesses innovate and grow. Done well, server virtualization makes fundamental changes that can lead an organization down the path of private and public cloud computing.

Server virtualization is absolutely not the only path to cloud computing. Cloud services will be offered at many levels of the IT stack, including applications (also known as software as a service) and application infrastructure service (application platform as a service) that certainly don't require or even use virtual servers. However, concurrently with internal IT modernization, server virtualization will be a natural path for existing workloads to evolve in a stepwise manner toward the cloud-computing style.

These changes are not all inevitable success stories. Organizations that focus on the tactical deployments and benefits of server virtualization will find themselves constantly in reaction mode due to continuous change being forced on them. There is danger in the changes caused by virtualization (process breakage, virtualization sprawl, uncontrolled deployments, etc.). The key is to understand and either mitigate or avoid these issues proactively.

Gartner has identified five key attributes of cloud computing. They are:

- **Service-based:** Consumer concerns are abstracted from provider concerns through service interfaces.
- **Scalable and elastic:** Services scale on demand to add or remove resources as needed.
- **Shared:** Services share a pool of resources to build economies of scale.
- **Metered by use:** Services are tracked with usage metrics to enable multiple payment models.
- **Internet technologies:** Services are delivered through use of Internet identifiers, formats, and protocols.

Virtualization is a major enabler of the first four. Here's how:

Service-Based. A successful IT organization is treated within the enterprise as a business that offers services and delivers those services according to service-level agreements. Unfortunately, this is extremely rare. Most IT organizations do not have service catalogs or service-level agreements, and their customers are intimately involved in the details of physical IT implementations. Virtualization is a forcing function that abstracts the physical implementation from the user, and requires the discussion between the business and IT organization to focus on service

requirements, rather than on implementation details. Because most IT organizations do not have an explicit understanding of their customers' service level requirements, and because they don't have fine-grained tools to tune IT delivery, they tend to overprovision. Virtualization makes fine-tuned resource delivery possible, but it requires a more explicit understanding of service-level requirements. Essentially, this shifts the relationship between the IT organization and its customers from being a technology host to being a service provider. Note that virtualization, by itself, does not create a rich, end-to-end, service-management environment – but it is a start.

Challenges and Recommendations: Determining specific service-level requirements is difficult, and finding an appropriate language to describe service choices is difficult, too. The best approach is to keep it as simple as possible, and to benchmark what is done elsewhere (for example, standard instances of virtual servers delivered by Amazon Elastic Compute Cloud [EC2] are “small,” “large” and “extra large”). Also, consider how these service levels might be delivered through a self-service interface in the future – which would enable a private cloud service.

Scalable and Elastic. Early adopters of cloud computing talk about how quickly they can get new servers online – as quickly as 20 minutes (compared to a traditional enterprise IT delay of a month or more). Virtual machines make it possible to separate hardware acquisition and deployment from software deployment, and can improve delivery within an enterprise to 10, 20, or even 30 times faster. Best-of-breed enterprises report that they, too, can deliver new virtual instances in 20 minutes using virtual machines. All of this requires fundamental changes to the provisioning process and capacity planning.

Faster delivery fundamentally changes business behavior. If IT delivers faster, that lowers the barrier to entry, and IT's customers generally ask for more resource. A lower barrier to entry (and exit) will spur an enterprise to experiment more often, and fundamentally changes business behavior with respect to IT usage. Speed can help a business grow; however, business processes will also need to change in order to take full advantage of this speed. Like the shift to becoming more service-based, speed and elasticity drive a significant culture change in IT and the business.

However, as IT delivers faster, business expectations of IT's responsiveness normalize yet again; and if IT does not meet these new expectations, the attractiveness of public cloud providers will increase accordingly, putting IT at risk of losing customers to faster (albeit less-robust) service providers.

Challenges and Recommendations: The IT processes that an organization has developed for years are generally not set up for “speed” and rapid change. Asset management, configuration management, performance management and capacity planning all must be modified and must embrace a much more dynamic environment. Gartner has encountered many organizations that

have virtualized too far, too fast. Their management tools and processes did not keep up, and the result was increased, rather than decreased, complexity – speed led to virtualization sprawl. Organizations undertaking virtualization must examine the impact to their management tools and operational processes before virtualization takes root.

Shared. While server consolidation is typically about statically combining small servers together into much larger servers, virtualization is much more about enabling economies of scale – sharing unused capacity within a physical server, within a farm, across a distributed enterprise – and doing it dynamically based on current requirements. Virtualization increases the importance of having a varied workload mix, either through many different workloads, or through workloads with different usage characteristics. The larger and more varied the enterprise and enterprise workload mix, the more efficiently they can utilize their capacity. Virtualization is a mechanism used by many cloud computing providers (such as Amazon [EC2], Microsoft [Windows Azure Platform], Terremark Worldwide, Savvis and many others) to significantly reduce their costs through economies of scale.

Challenges and Recommendations: IT customers don't like to share. As a part of the move to service-level agreements, customers need to be convinced that sharing is in their best interests. This is best proved by lower costs (if costs are charged back), increased speed, or new benefits, such as disaster recovery. Proof points elsewhere in the enterprise always help.

Virtualization “unlevels” the playing field, making it difficult for small organizations to use their capital equipment as efficiently as larger organizations (which is why public cloud computing will be very appealing to smaller organizations). Likewise, larger organizations cannot provide competitive IT costs, unless they proactively and effectively leverage virtualization technologies. Resource sharing, holistic capacity planning, dynamic workload analysis and placement – these changes are not optional. Virtualization enables but does not automatically create economies of scale – it requires effort, and can be done badly. Also, overanalysis and overmanagement of virtualization can lead to increased operational costs. Successful organizations will be proactive in modifying their processes to optimize their economies of scale, without dramatically increasing the human effort required. Failure to compete in this unlevel playing field will lead many small and some large organizations to outsourcing and cloud computing.

Metered by Use. As virtualization becomes a more common industry default, software vendors will be forced to change their pricing models. Today's models are often based on the physical capacity of a server and specific serial numbers. In a virtual-server architecture, virtual machines use a small portion of a physical server, can change dynamically, and can even move to other physical servers in the cluster. Many vendors are slow to change to accommodate virtualization, but as virtualization expands from 16% penetration today to near 50% by 2012, software vendors will be forced to create some kind of usage-based pricing model by customer pressure, competition, and software-as-a-service alternatives.

In addition, when servers can be delivered in minutes or days rather than weeks, IT's customers ask for more. Many Gartner clients report that their demand has doubled postvirtualization. A lower barrier to entry means that there is much less friction in the system. Many Gartner clients are introducing chargeback as a way to ensure that good business decisions are being made when servers are requested, and tools are emerging to help IT measure virtual resource usage. Furthermore, they are getting better at capacity planning.

Challenges and Recommendations: While software vendors will change their pricing models to embrace dynamic virtualization, this does not mean they will build models that will reduce their revenue. Enterprises should be cautious about early experiments in usage-based pricing by software vendors – they could very likely cost more than traditional pricing models. Software pricing and licensing will take years to adjust to the new normal of virtualization and cloud computing.

Chargeback may not be politically viable for some IT organizations (which may delay them from focusing on optimizing IT costs, and may make them uncompetitive). In those cases, IT should still build the capability to estimate costs for new projects, and measure real resource usage in order to report business unit usage of IT resources. If IT costs are increasing, then at least it will be possible to identify which businesses are driving the cost increases.

For organizations that are deploying chargeback for the first time, the best rule of thumb is to capture and measure the detail, but to keep it simple and relatively static for their customers. Go slowly. Gartner clients that tried to introduce a fully dynamic, “by-the-drink” chargeback model found that they had to spend a lot of time arguing with their customers about costing details – and many businesses are not ready for IT to become a fully variable (and unpredictable) expense. A better strategy is to introduce pricing tiers and typical service bundles that may change every few months (or annually) – with monitoring to watch for unusual usage spikes. More variability and detailed pricing can be introduced over time, as the enterprise gets used to yet another culture change.

Bottom Line

Virtualization, done well, will shift IT to become more service-based, delivering scalable and elastic resources with much greater speed, driving economies of scale with shared resources, and measuring and perhaps charging back based on dynamic usage. These are major steps toward the cloud-computing style, and will help the IT organization behave much more like a cloud-computing provider. As external cloud services mature, there will be a much more level playing field to more accurately compare internal IT services with external IT services, and the organization will have also gone through fundamental cultural, political, and funding changes that will make external cloud-computing services much easier to consume when (and if) they are ready.