About that ‘Unassailable Economic Argument’ for Public Cloud Computing

Lori MacVittie, 2011-25-05

Turns out that 'unassailable' economic argument for public cloud computing is very assailable

66 The economic arguments are unassailable. Economies of scale make cloud computing more cost effective than running their own servers for all but the largest organisations. Cloud computing is also a perfect fit for the smart mobile devices that are eating into PC and laptop market. 99 -- Tim Anderson, "Let the Cloud Developer Wars Begin"
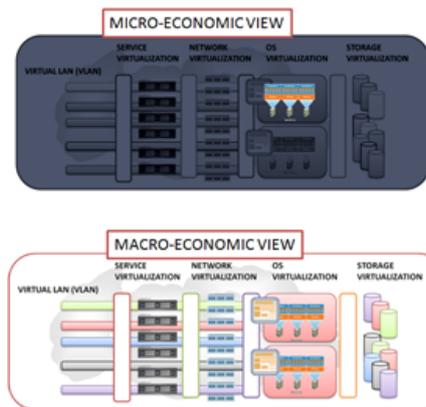
Ah, Tim. The arguments are not unassailable and, in fact, it appears you might be guilty of having tunnel vision – seeing only the list price and forgetting to factor in the associated costs that make public cloud computing not so economically attractive under many situations. Yes, on a per hour basis, per CPU cycle, per byte of RAM, public cloud computing is almost certainly cheaper than any other option. But that doesn't mean that arguments for cloud computing (which is much more than just cheap compute resources) are economically unassailable. Ignoring for a moment that it isn't as clear cut as basing a deployment strategy purely on costs, the variability in bandwidth and storage costs along with other factors that generate both hard and soft costs associated with applications must be considered .

MACRO versus MICRO ECONOMICS

The economic arguments for cloud computing almost always boil down to the competing views of micro versus macro economics. Those in favor of public cloud computing are micro-economic enthusiasts, narrowing in on the cost per cycle or hour of a given resource. But micro-economics don't work for an application because an application is not an island of functionality; it's an integrated, dependent component that is part of a larger, macro-economic environment in which other factors impact total costs.

The lack of control over resources in external environments can be problematic for IT organizations seeking to leverage cheaper, commodity resources in public cloud environments. Failing to impose constraints on auto-scaling – as well as defining processes for de-scaling – and the inability to track and manage developer instances launched and left running are certainly two of the more common causes of "cloud sprawl." Such scenarios can certainly lead to spiraling costs that, while not technically the fault of cloud computing or providers, may engender enough concern in enterprise IT to keep from pushing the "launch" button.

> 66 The touted cost savings associated with cloud services didn't pan out for Ernie Neuman, not because the savings weren't real, but because the use of the service got out of hand.
>
> When he worked in IT for the Cole & Weber advertising firm in Seattle two and a half years ago, Neuman enlisted cloud services from a provider called Tier3, but had to bail because the costs quickly overran the budget, a victim of what he calls cloud sprawl - the uncontrolled growth of virtual servers as developers set them up at will, then abandoned them to work on other servers without shutting down the servers they no longer need.
>
> Whereas he expected the developers to use up to 25 virtual servers, the actual number hit 70 or so. "The bills were out of control compared with what the business planned to spend," he says. 99
>
> -- Unchecked usage can kill cost benefits of cloud services

But these are not the only causes of cost overruns in public cloud computing environments and, in fact, uncontrolled provisioning whether due to auto-scaling or developers forgetfulness is not peculiar to public cloud but rather can be a problem in private cloud computing implementations as well. Without the proper processes and policies – and the right infrastructure and systems to enforce them – cloud sprawl will certainly impact especially those large enterprises for whom private cloud is becoming so attractive an option.

While it's vastly more difficult to implement the proper processes and procedures automatically in public as opposed to private cloud computing environments because of the lack of maturity in infrastructure services in the public arena, there are other, hotter issues in public cloud that will just as quickly burn up an IT or business budget if not recognized and addressed before deployment. And it's this that cloud computing cannot necessarily address even by offering infrastructure services, which makes private cloud all the more attractive.

TRAFFIC SPRAWL

Though not quite technically accurate, we'll use traffic sprawl to describe increasing amounts of unrelated traffic a cloud-deployed application must process. It's the extra traffic – the malicious attacks and the leftovers from the last application that occupied an IP address – that the application must field and ultimately reject. This traffic is nothing less than a money pit, burning up CPU cycles and RAM that translate directly into dollars for customers. Every request an application handles – good or bad – costs money.

The traditional answer to preventing the unnecessary consumption of resources on servers due to malicious or unwanted traffic is a web application firewall (WAF) and basic firewalling services. Both do, in fact, prevent that traffic from consuming resources on the server because they reject it, thereby preventing it from ever being seen by the application. So far so good. But in a public cloud computing environment you're going to have to pay for the resources the services consumed, too. In other words, you're paying per hour to process illegitimate and unwanted traffic *no matter what.* Even if IaaS providers were to offer WAF and more firewall services, you're going to pay for *that* and all the unwanted, malicious traffic that comes your way will cost you, burning up your budget faster than you can say "technological money pit."

This is not to say that both types of firewall services are *not* a good idea in a public cloud environment; they are a valuable resource regardless and should be part and parcel of any dynamic infrastructure. But it is true that in a public cloud environment they address only security issues, and are unlikely to redress cost overruns but instead may help you further along the path to budget burnout.

HYBRID WILL DOMINATE CLOUD COMPUTING

I've made the statement before, I'll make it again: hybrid models will dominate cloud computing in general due primarily to issues around control. Control over processes, over budgets, and over services. The inability to effectively control traffic at the network layer imposes higher processing and server consumption rates in public environments than in private, controlled environments even when public resources are leveraged in the private environment through hybrid architectures enabled by virtual private cloud computing technologies. Traffic sprawl initiated because of shared IP addresses in public cloud computing environments alone is simply not a factor in private and even hybrid style architectures where public resources are never exposed via a publicly accessible IP address. Malicious traffic is never processed by applications and servers in a well-secured and architected private environment because firewalls and application firewalls screen out such traffic and prevent them from unnecessarily increasing compute and network resource consumption, thereby expanding the capacity of existing resources. The costs of such technology and controls are shared across the organization and are fixed, leading to better forecasting in budgeting and planning and eliminating the concern that such essential services are not the cause of a budget overrun.

Control over provisioning of resources in private environments is more easily achieved through existing and emerging technology, while public cloud computing environments still struggle to offer even the most rudimentary of data center infrastructure services. Without the ability to apply enterprise-class controls and limits on public cloud computing resources, organizations are likely to find that the macro-economic costs of cloud end up negating the benefits initially realized by cheap, easy to provision resources. A clear strategy with defined boundaries and processes – both technical and people related – must be defined before making the leap lest sprawl overrun budgets and eliminate the micro-economic benefits that could be realized by public cloud computing.