What's this "Cloud Computing" all about, really?

There's a lot of speculation (not to mention misconceptions) flying around about VMware's recent move towards the "Cloud". Critics believe it to be absolute rubbish, a step back to the age of the mainframe, warning us of the dangers of centralized data storage and the security/availability issues that might bring about. "Only your own PC is safe enough for your data." Proponents counter these arguments by quoting the numerous applications that people use that basically require no more than a dumb terminal anyway. For example, our readers wouldn't technically need a monster PC to read our articles and their emails; just a thin client with web access would suffice.

We believe these arguments, though undoubtedly relevant with regards to personal computing, largely miss the point of what the Cloud Computing movement is looking to achieve, and with the announcement of VMware's vSphere, we would like to take some time out to clarify just what Cloud Computing is all about.

VMware's press release on this hits the nail on the head quite nicely, in fact, arguing that the main problem in companies' IT infrastructure today is their high infrastructure maintenance costs. Add to this the rapidly widening gap between hardware development and the software that is actually being used in companies, and many data centers end up awfully bloated, even though only a fracture of their capacity is being used.

Virtualization, as documented in many articles over the past year, provides a pretty solid answer to this, and is actually the core technology at the base of the Cloud Infrastructure. Its evolution from consolidation tool to fully-fledged dynamic data center management enabler has paved the way for an even bigger jump forward: IT as a Service.

Virtualization has grown to the point where we can reduce a data center to one big resource pool, with the virtual machines more or less floating on top, blissfully unaware of where their actual resources come from. IT as a Service aims to make it possible to extend this internal company resource pool (conveniently nicknamed the "internal cloud" from now on) reliably with resources obtained in much the same way as electricity, or a phone line. The idea is pay as you go, pay for what you use, always available and reliable, and with a large choice of service providers (from here on referred to as the "external cloud"). This is what VMware vSphere 4 will be able to provide.

So what is VMware releasing?

While vSphere 4 is technically an upgrade to the existing VMware Infrastructure (the latest version being ESX 3.5), they are now opening up the product to allow management of both the internal and external clouds. This allows users to extend the existing company infrastructure with that of IT service providers all over the world. (Here is a list of providers that are currently ready to offer these services.)

VMware is looking to merge these two worlds with vSphere, but while still acknowledging the key differences between the two: Internal company clouds should be deployed in a non-disruptive, evolutionary way. Any data center that has already been virtualized should be able to move to the cloud without a glitch, and it should remain as trusted, reliable, and secure as expected from an internal infrastructure.

External clouds are to be put to work for extra capacity, and quite possibly high availability and disaster recovery as well. Migrating VMs from the internal to the external cloud should work seamlessly through Storage VMotion (we believe this should work, provided that the VM is on a LUN by itself). Think of this environment as a large-scale VPN, independent of your own infrastructure, which is paid for not in terms of infrastructure costs, but per unit of actual work completed. This environment can be as large or small as is required at the time, and should add a lot of flexibility to existing data centers in an efficient way.
vSphere is truly meant to be the very first "Cloud-OS", a system that can break up separate hardware platforms into what they offer in terms of resources, and use these resources as building blocks for what is, essentially, a supercomputer that encompasses the entire virtualized data center and beyond. VMware aims to make vSphere compatible with any existing or future applications and maintain strict security, even in the external cloud, while keeping the technology as non-intrusive as possible. Ideally, there should be no lock into specific service providers, or locks into irreversible decisions. As long as a company's intent is not to make their entire infrastructure irreversibly dependent on the external cloud, it should always be able to fall back on its internal infrastructure.

vSphere, a feature overview

The software is based on three basic pillars, combined by a single slogan:

"Cut capital and operational costs over 50% for all applications (Efficiency) while automating quality of service (Control) and remaining independent of hardware, operating system, application stack, and service providers (Choice)."

As the efficiency claims will provide quite a challenge for VMware to overcome, let's have a look at the actual changes they have made to their existing software to achieve this goal.

Until recently, the VMware Infrastructure Efficiency front was largely defined by three of its main features: CPU/Memory Optimization, Dynamic Resource Scheduling, and the VMFS. Each of these was able to provide a boost to performance by taking advantage of the entire virtual infrastructure. Through the combined improvements to vCompute, vStorage and vNetwork, VMware predicts a 30% increased consolidation ratio compared to the last version of VMware Infrastructure. We asked VMware about this, and heard the results were measured by using their very own VMmark, comparing ESX 3.5 and 4.0 on the same hardware platform. We believe the results to be generally realistic, but they should still be taken with a grain of salt.

Furthermore, VMware promises 50% more storage savings, thanks to vStorage Thin Provisioning and up to 20% additional power and cooling savings, thanks to Distributed Power Management that allows the data center to automatically VMotion VMs to as few physical machines as possible, allowing users to power down physical servers that are not needed at the moment.

Besides this, vSphere will offer more powerful virtual machines than before, allowing for:

- 8 vCPUs per VM (up from 4)
- 10 virtual NICs per VM
- up to 255GB RAM per VM
- 3 times more network throughput

http://www.anandtech.com/print/2753
a little over 200,000 IOPS maximum

When it comes out, a single vSphere deployment will be able to manage the following resources:

- 32 physical servers with up to 2048 cores
- 1280 virtual machines
- 32TB of RAM
- 16PB (petabytes - 1000TB) of storage
- 8000 network ports

Keep in mind that it will be possible to cluster up to 10 vCenter Servers (vSphere's main management component).

On the Control side of things, vSphere makes managing this large amount of resources easy through the use of Host Profiles and the vNetwork Distributed Switch, which allows for comprehensive networking between all these virtual machines to be configured just as easily as in a “regular” data center. In fact, the Switch can actually be plugged into by vendors such as Cisco, so it can be configured in exactly the same way as a regular Cisco switch.

As mentioned before in one of our blogs, one-click configurable Fault Tolerance should keep important VMs safe from unexpected hardware failures, and storage-maintenance related downtime should be further minimized by introducing a feature called Storage VMotion. Remember the currently documented limitations of Fault Tolerance, however. It's High Availability, but only on the hardware-level, and will not account for software failures. Also, at this point, Fault Tolerance is limited to VMs with only a single vCPU, so although it's a very interesting feature, it leaves a lot of room for improvement.

Data Recovery is another nice little tidbit, which will essentially plug into your vSphere as a virtual appliance with a single purpose: one-click backup of any virtual machine, answering the first question asked by quite a few companies getting started with virtualization.

Lastly, vShield Zones is a self-learning, self-configuring firewall service, allowing users to set security policies to overall environments, enforcing the rules on an application level, and allowing them to travel along with the VM as it VMotions across the (internal or external) cloud.

For vSphere's last main pillar, VMware likes Choice (as long as they are included in whichever choice you end up making, naturally), and for that reason they proudly hold the claim that they are able to support any kind of server, any kind of storage, any kind of OS, any application, be they on premise or off premise. In the picture below, they compare their own supported OS list to that of Hyper-V.

On top of that, VMware has partnered up with several service providers (as linked before) to properly kick off their external cloud services with a large team of reliable players.

Page 4

vSphere Pricing favors hex-cores

On to the interesting part: pricing. Once again, VMware has thoroughly scrambled their existing pricing scheme with some very interesting entry points.
VMware's own calculations are apparently off in the essentials package, as the actual minimum price is at $166/CPU.

The existing Foundation package is split up into two entry-level packages. These are limited to three servers in total, and will not be able to scale beyond that point, but otherwise roughly translate to a slightly less than full-featured Standard version (no High Availability), and a slightly less than full-featured Advanced version (including Data Recovery, but not including Fault Tolerance and VMotion). A very nice feature comparison of the different versions can be found here.

Interesting to note, however, is that in Essentials, Essentials Plus and Standard, the limit to six physical cores per CPU from VMware Infrastructure 3 is apparently not lifted. Currently, that means that hex-cores will quite possibly be the more favorable CPUs to use in this bracket, leaving the door open for Istanbul to potentially get the most out of these configurations. Though the press release reads "physical" cores quite literally, we cannot help but wonder whether Nehalem's Hyper-Threading is included in that. If that is the case, the most optimal platform might just be Intel after all, though this is something that we'd like official confirmation on.

Before the actual announcement was made, we wondered whether VMware would allow for quad-socket machines in the Essentials packages, upping the maximum amount of CPUs to 12. According to their website, however, the three physical servers running Essentials are limited to 2 sockets each. The basic Essentials package starts at $995, which, at $166 per CPU equals six CPUs total. Essentials Plus costs $2995 all told, which, following the same logic, translates to $499/CPU.

For the Advanced and Enterprise Plus packages, the maximum number of physical cores per CPU is upped to 12, putting the future 8-core Beckton and 12-core Magny-Cours (both scheduled for release in 2010) in direct competition with each other in these price brackets.