Servers Get Multiple Personalities

By Dennis Barker, GRIDtoday, 06/23/2008

As a former IT architect turned CEO, Tony Bishop is in an enviable situation: He knows what technology his company needs, and he is in a very good position to make sure the company gets it. In his last job as chief architect in Wachovia's Corporate Investment Banking Technology Group, he led a team that designed and built an infrastructure for utility computing. In his current role as chief executive for Adaptivity, he leads a company that designs and helps other companies implement service-oriented systems that are, above all, adaptive.

"Adaptivity's core expertise is integrating the often disparate arenas of service-oriented architecture (SOA), virtualization, infrastructure automation and Web 2.0 technologies," Bishop says. "We help large end-users, systems integrators and product manufacturers understand the ecosystem, the integration points, their individual weaknesses, and put together actionable projects in the context of a multi-year strategic roadmap."

Because Adaptivity and its customers are in the service business, that ecosystem has to be able to respond to sudden changes in demand. Adaptivity designs real-time infrastructure systems, and for the supremely critical task of rapidly repurposing servers, Bishop chose technology from Scalent Systems. Scalent's Virtual Operating Environment (V/OE) software is a key building block for Adaptivity's real-time solutions and frequently is an early part of the roadmaps they design for customers, he says.

"If you think about IT in economic terms," Bishop says, "a technology that can instantly determine the relative importance or priority of a workload based on a business policy and reallocate capacity accordingly is in the demand-side category. [Examples of that are] DataSynapse's GridServer and FabricServer products. On the other hand, a technology that within minutes can turn an application down on a server and make that server available to boot up as a completely different application or system falls into the category of supply-side. And in a nutshell, that is the capability that Scalent enables." (Scalent, by the way, has a partnership deal with DataSynapse.)

"Some people classify us as automation management or as virtualization, and we do those things," says Scalent CEO Ben Linder, "but I think we tend to fit into the vision of real-time infrastructure, the notion that through automation management you can achieve higher levels of agility and higher levels of utilization. The central thesis behind what we do is that resources in the datacenter, including networking and storage, should be agile and malleable. Resources should be repurposable in real-time to do whatever workloads the business needs."

V/OE installs an agent on every physical server that inventories physical capabilities, takes a snapshot of the server in action, and stores it centrally on the SAN. Any machine, or an entire farm of servers, then can be repurposed on the fly, including all associated network and storage access topologies, the company says. Scalent virtualizes the underlying physical machine operating system, network connectivity, and storage access.

V/OE lets datacenter managers quickly change which servers are running, what software stacks servers are running, and how those servers are connected to network and storage, Linder explains. There's no need for IT to make changes to physical machines, cables, LAN connections or SANs.

"VO/E adds a layer of management that lets you control and automate servers, storage, and network switches so that under a single umbrella you can provision and orchestrate those resources. You can put a new OS, a new application on a server, then connect that server to a new network, all in a matter of minutes," says Linder. "Our motto is 'rack once, table once, repurpose infinitely.'"
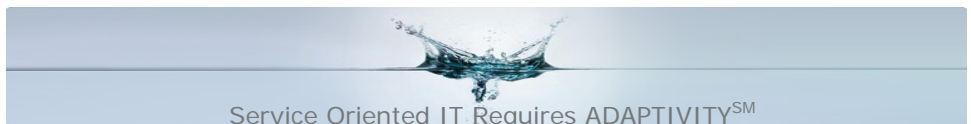
Datacenter Reliability

Scalent's customers tend to be large datacenters with a few hundred to a few thousand servers, "with a few thousand more common now as datacenters are just growing out of control," Linder says. "Customers use our software to actually manage resources allocated through the grid, so grids can grow and shrink dynamically depending on the needs. One customer, for example, borrows resources out of their test-and-development labs and repurposes those servers into the grid at the end of the month when they need to add another thousand nodes to the grid for processing calculations."

One of those big datacenter customers is Blackboard Inc., a leading provider of online learning software and educational scheduling tools used in post-secondary schools. Blackboard Managed Hosting is a unit within Blackboard that provides fully managed hosting services for the company's software to more than 700 clients worldwide. "We currently have six datacenter facilities on three different continents, over 1,300 Internet-facing Web application servers, and nearly a thousand Oracle databases," says Pete Waterman, senior manager of technology and innovation for Blackboard's hosting unit. He leads a small group chartered with the sounds-familiar task of identifying and implementing emerging technologies that will increase efficiency, performance and availability.

"We're using the Scalent V/OE as a core component in our new datacenter environment design," Waterman says. "This design has the primary goals of removing any link or dependency between a customer's installation and a physical server or network segment while significantly decreasing backup, recovery and provisioning cycles. We are using Scalent primarily to control our dynamic data assignment between virtual and physical servers, as well as to provide logical network segmentation as appropriate via network controls.

"For us, Scalent provides a framework that makes it simple to address our current problems of slow recovery from hardware failures or data loss. Six hours is simply too long for a client to be offline in the event of a catastrophic hardware failure. However, with our new design we can recover within minutes with minimal intervention. While this doesn't directly impact most of our clients due to the scarcity of hardware failures, it will make a huge impact in those rare situations as we continue to roll this technology out."

"Scalent does everything 'out of band,'" Bishop says. "If there was ever a catastrophic failure with Scalent, it would only affect your ability to do new server repur-

posing from that point forward, all systems already running under Scalent would be completely unaffected."

Another core benefit for Blackboard, Waterman says, is the "enterprisewide control and manipulation of the Xen hypervisor combined with excellent management for Linux. We had actually sat down and built out a list of requirements for a set of software to build in-house for our new datacenter design, and Scalent came in with 75% percent of this capability out-of-the-box. For us, the focus on Linux, Xen, and simple technology such as snmp, ipmi, and so on shows that Scalent understands how a large-scale Linux shop runs."

Taking Disaster Recovery to the Bank

As a banking systems architect, Bishop got to see Scalent from the user point of view. When his team first deployed Scalent's technology at Wachovia, they were focusing on mid-range disaster recovery to a remote datacenter that was hundreds of miles away --"something every experienced IT executive will groan about." But with Scalent, he says, it is "almost a trivial exercise to failover a server, or hundreds of servers, to a remote datacenter (as long as you have SAN replication between the sites). It removes a tremendous amount of complexity -- and therefore risk of human error -- from the process. And when you bring each server persona back online in the remote site, there's no worry about 'Does the server have the latest security patches?' or 'Are they running that new release of the application that was just promoted to product last week?' You bring online the same system image that was running in the primary production datacenter minutes ago."

Bishop says he's careful to make clients aware that Scalent's system "is not a provisioning or patch management tool. In a manner of speaking, you can use Scalent to 'clone' a basic server image and, yes, that saves you time. But you still need to build that basic server image, as well as the complete application image. You also still need to apply patches to those personas -- including the ones stored on the SAN but not running. You could do that manually, but if you start having hundreds or thousands of personas, the ROI of a tool to perform that work automatically is fairly easy to justify."

Improving Utilization

Bishop says one benefit that "really cements" Scalent's place in Adaptivity's real-time infrastructure ecosystem is the way it can boost utilization of resources. "One of the things we discovered as we built out the real-time infrastructure capabilities at Wachovia was that while having that … the vast majority of the time the workloads are actually very predictable and follow time-of-day/day-of-week patterns. What Scalent gave us was the ability to proactively shift personas for hundreds of servers based on those simple patterns, which in turn enabled us to run our infrastructure at much higher 24-hour utilization percentages than we could have achieved otherwise."

Linder says that through effective automation and management of infrastructure, Scalent allows a doubling of resource utilization in most datacenters. "Typically, server utilization hovers at 5 to 10 percent," he says, "and by moving to a dynamically managed infrastructure, datacenters can get twice the utilization."

New Version Faces Virtual Reality

V/OE 2.5 (announced June 10) introduces a new feature with an appropriately metaphysical name: infinite virtual transition. "IVT is in response to the fact that so many people have adopted VMware and Xen as hypervisors and now not only have a lot of physical machines, but a lot of virtual machines to manage," Linder says. "And what we've found is that for the life of an application, there are many times an app needs to live as a virtual machine, like in the test and development phase. But when it goes into production, an application generally wants to be deployed on bare-metal hardware, on the server itself. Very few datacenters run their production apps on virtual machines because you tend to get much better performance, more predictable performance, on hardware."

"What IVT enables," he says, "is seamless, real-time movement of images at boot between physical hardware, virtual hardware and different hypervisors, without any conversion needed."

What IVT does that's unique, he says, is "let people choose whether to run their servers as physical or virtual without any conversion, so at any point you can run a server as virtual or physical and go back and forth depending on what your needs are at the time. I believe we're the only one to provide this capability."

"What is happening out there is datacenters are becoming hybrids, split between physical and virtual," Linder says, "and the more the use of virtual machines spreads, the more customers want to move between physical and virtual for various reasons, including power savings. Let's say on the weekend, you want to keep all your physical servers accessible but don't want to use all that electricity. You could save 15 to 25 percent on electric costs by simply shutting those servers down, but you know demand could change and there are applications that might need those servers. So you migrate your servers to virtual servers for the weekend, turn off the unused servers, then, when the weekend's over, our software automatically powers up all the physical servers and migrates the virtual servers back to their physical machines." The end-of-weekend rebooting of physical machines takes about 5 to 10 minutes, Linder says.

"Organizations can no longer allow expensive resources to be left idle," says Dan Kusnetzky, president and principal analyst of Kusnetzky Group, an IT consultancy with specialty in virtualization and Web-based applications. "Scalent's V/OE allows IT administrators to treat all of their resources as a pool of resources that can be allocated on a moment-to-moment basis. Applications that need only a little processing power would get what they needed and no more. Other applications would be assigned the resources they need. The resources would be carefully managed to meet service-level objectives, respond to outages, and still live up to the preset policies the IT staff had set.

In the end, Kuznetsky said, the organization gets what it paid for: the computing power they purchased to actually do something for the organization. Companies do not have to over-purchase or over-provision to assure stable, responsive operations.

"Management of virtual and physical resources can be very challenging. It truly is rocket science," Kusnetzky says. "Scalent is one of a group of suppliers who have found a way to package that expertise and put it into a product."