Virtualization technology is hot again, but for the right reasons?

BY MACHE CREEGER, MODERATOR

CTO Roundtable on Virtualization

THIS MONTH WE present the second in a series of ACM CTO Roundtable forums. Overseen by the ACM Professions Board, the goal of the forums is to provide high-powered expertise to practicing IT managers to help inform their decisions when investing in new architectures and technologies.

The topic of this forum is virtualization. When investing in virtualization technologies, IT managers must know what is considered standard practice and what is considered too leading-edge and risky for near-term deployment. For this forum we've assembled five leading experts on virtualization to discuss what those best practices should be. While the participants might not always agree with each other, we hope their insights will help IT managers navigate the virtualization landscape and make informed decisions on how best to use the technology. Next month we will present Part II of this forum, discussing such topics as clouds and virtualization, using virtualization to streamline desktop delivery, and how to choose appropriate virtual machine platforms and management tools.

Participants

Mache Creeger (Moderator): Creeger is a longtime technology industry veteran based in Silicon Valley. Along with being an *ACM Queue* columnist, he is the principal of Emergent Technology Associates, marketing and business development consultants to technology companies worldwide.

Tom Bishop is CTO of BMC Software. Prior to BMC, Bishop worked at Tivoli, both before and after their initial public offering and acquisition by IBM, and also at Tandem Computers. Earlier in his career Bishop spent 12 years at Bell Labs' Naperville, IL facility and then worked for UNIX International. He graduated from Cornell University with both bachelor's and master's degrees in computer science.

Simon Crosby is the CTO of the Virtualization Management Division

at Citrix. He was one of the founders of XenSource and was on the faculty of Cambridge University, where he earned his Ph.D. in computer science. Crosby grew up in South Africa and has master degrees in applied probability and computer science.

Gustav. This is a pseudonym due to the policies of his employer, a large financial services company where he runs distributed systems. Early in his career, Gustav wrote assembler code for telephone switches and did CAD/CAM work on the NASA space station *Freedom*. He later moved to large system design while working on a government contract and subsequently worked for a messaging and security

alization, virtualization management, and application virtualization. Stewart is a Microsoft Certified Architect and is on the Board of Directors of the Microsoft Certified Architect Program.

Steve Herrod is the CTO of VMware, where he's worked for seven years. Before VMware, Herrod worked in Texas for companies such as EDS and Bell Northern Research. Earlier in his career Herrod studied with Mendel Rosenblum, the founder of VMware, at Stanford and then worked for TransMeta, a computer hardware and software emulation company.

Steve Bourne is chair of the ACM Professions Board. He is also a past president of the ACM and Editor-in-

SIMON CROSBY: The power-savings issue is a big red herring because the CPU is a small portion of the total power consumption compared to spinning all those disk drives in a storage array. I'll be the first one to say that free, ubiquitous CPU virtualization is just an emergent property of Moore's Law, just part of the box. Memory is another major power consumer and memory architectures are definitely not keeping up. When you're talking about virtualizing infrastructure, you should be talking about what bits of it you virtualize and how: CPU, storage, and/or memory. You have to look at the whole thing. As for showing lower overall power consumption, I have yet

Steve Bourne

Mache Creeger

Allen Stewart

startup company in Silicon Valley, taking it public in the mid-1990s. After starting his own consulting firm, he began working at his first large financial firm. Seven or eight years later, he landed at his current company.

Allen Stewart is a Principle Program Manager Lead in the Windows Server Division at Microsoft. He began his career working on Unix and Windows operating systems as a system programmer and then moved on to IBM, where he worked on Windows systems integration on Wall Street. After IBM, Stewart joined Microsoft, where for the first six years he worked as an architect in the newly formed Financial Services Group. He then moved into the Windows Server Division Engineering organization to work on Windows Server releases. His primary focus is virtualization technologies: hardware virtu-

Chief of the *ACM Queue* editorial advisory board. A fellow alumnus with Simon Crosby, Bourne received his Ph.D. from Trinity College, Cambridge. Bourne held management roles at Cisco, Sun, DEC, and SGI and currently is CTO at El Dorado Ventures, where he advises the firm on their technology investments.

MACHE CREEGER: Virtualization is a technology that everyone is talking about, and with the increased cost of energy the server consolidation part of the value proposition has become even more compelling. Let's take that as a given and go beyond that. How do we manage large numbers of virtualized servers and create an integral IT architecture that's extensible, scalable, and meets all the criteria that reasonable people can agree on?

to see a good calculation for that.

GUSTAV: I support virtualization for a number of reasons, but cost savings isn't one of them. What I typically see is that the server guy makes a decision to reduce his costs, but that significantly impacts storage and the network, making their costs go up.

To put eight software services on a single machine, instead of buying the \$3,000 two-socket 4GB 1U blade, I bought the four-socket, 16GB system for \$20,000. While that calculation provides an obvious savings, because I want to use VMotion I have to purchase an additional storage array that can connect to two servers. The result is that I paid more money than the traditional architecture would cost to support the same service set.

That's why you see a large interest in virtualization deployment followed

by this trough of disillusionment. Once deployed, people find out that a) it's hard and b) oddly, they are spending all this money on storage.

SIMON CROSBY: I agree with that. Solving storage is the hardest problem of all. It's the big bear in the room.

From an infrastructural cost perspective, you have to consider the human cost: The number of administrators still grows linearly with the number of VMs (virtual machines). Server-consolidation costs are very real. However, as virtualization addresses the mainstream production workload, costs are still going to be driven by the number of administrators. Unless the administrator problem is solved, IT has not re-

threaded has taken off at all.

TOM BISHOP: I think that is the elephant in the room. We as an industry haven't built the correct abstraction which can fungibly apply computing to a domain and allow dynamic apportioning of computing capacity to the problems to be solved. Storage is a piece of it, but ultimately it's more complex than just that.

If you take virtual memory as an example, we spent approximately 20 years trying to figure out the correct abstraction for memory. One can define memory consumption using one abstraction and build a set of automated mechanisms to dynamically allocate a much smaller physical resource

amount of demand, and I want to satisfy that demand with my capacity, using an affordable cost equation in a way that is moment-to-moment optimal for heterogeneous systems.

GUSTAV: The beauty of having a good working abstraction of an underlying function is that the operating system can give you less than you asked for and still provide acceptable results. Virtual memory is a good example.

MACHE CREEGER: So how does this apply to the poor guy who's struggling in a small to medium-size company? What is he supposed to take away from all this?

SIMON CROSBY: The single-server notion of virtualization—server con-

Tom Bishop

Simon Crosby

Steve Herrod

ally been transformed.

GUSTAV: Fully using all the cores on the die is one of the biggest things that drive the need for virtualization. In the not-too-distant future we are going to be forced to buy eight-core CPUs. Because our computational problem has not grown as fast as the ability to throw CPU resources at it, we will not be able to keep all those cores fully utilized.

It's not an issue of power savings, as Intel will nicely power down the unused cores. The real benefit of multi-core CPUs is to be able to address the huge legacy of single-threaded processes in parallel.

SIMON CROSBY: I think that should be the high-order bit. There are 35 million servers out there, all of which are single threaded. So let's just admit that, run one VM per core, and have a model that actually works. I don't think multi-

to a much larger virtual demand and still achieve optimal performance moment-to-moment. We have not found that same abstraction for the core IT mission of delivering application services.

SIMON CROSBY: But isn't determining the biting constraint the fundamental problem with any one of these things? I don't care about optimizing something that is not a biting constraint. It may be memory, CPU, storage, and/or power—I have no clue. Different points of the operating spectrum might have different views on what is or is not a biting constraint. The grid guys, the network guys—all have different views and needs.

TOM BISHOP: I predict we will develop a workable abstraction for representing the basic IT mission: *I've got a certain amount of capacity, a certain*

solidation—is very well established and basically free. It will be part of every server, an emergent property of Moore's Law, and multiple vendors will give it to you.

The orchestrated assignment of resources across the boundaries of multiple servers or even multiple different resources is a major problem and I don't think we really have begun to understand it as yet.

TOM BISHOP: Regarding vendor lock-in, I think there is a body of experience, certainly in the companies I speak to, that says the heck with vendor lock-in. If my chance of getting a solution I can live with is better if I limit myself to a single vendor, I'm prepared to accept that trade-off.

GUSTAV: If you have a relatively small number of servers, there is no problem that you have that virtualization

can't solve. However, there will be new things that you may choose to spend money on that in the past were not a problem, such as doing live migration between servers. But if you just want to run a shop that has up to around 20 servers, unless you're doing something really weird, you should do it. It is easy and readily available from any of the major vendors. This addresses the relatively easy problem of "Can I put three things on one server?"

If you then realize you have new problems, meaning "Now that I have three things on one server, I want that server to be more available or I want to migrate stuff if that server fails," this is a level of sophistication that the market is only beginning to address. Different vendors have different definitions.

SIMON CROSBY: Once you achieve basic virtualization, the next big issue is increasing overall availability. If I can get higher availability for some key workloads, that transforms the business.

STEVE HERROD: I agree. In fact, we currently have a large number of customers that buy one VM per box first and foremost for availability and second for provisioning.

TOM BISHOP: About two years ago, the best session at a conference I attended was "Tales from the Front, Disaster Recovery Lessons Learned from Hurricane Katrina." A large aerospace company had two data centers, one just south of New Orleans and another one about 60 miles away in Mississippi. Each center backed the other up and both ended up under 20 feet of water.

The lesson they learned was to virtualize their data center. In response to that experience they built a complete specification of their data center where it could be instantiated instantaneously and physically anywhere in the world.

MACHE CREEGER: Our target IT manager is trying to squeeze a lot out of his budget, to walk the line between what's over the edge and what's realistic. Are you saying that all this load balancing, dynamic migration, that the marketing literature from Citrix, VMware, and Microsoft defines as the next big hurdle and the vision for where virtualization is going are not what folks should be focusing on?

SIMON CROSBY: Organizations today build organizational structures around current implementations of technology, but virtualization changes all of it. The biggest problem we have right now is that we have to change the architecture of the IT organization. That's people's invested learning and their organizational structure. They're worried about their jobs. That's much harder than moving a VM between two servers.

MACHE CREEGER: A while back I did a consulting job for a well-known data-center automation company and they brought up this issue as well. When you change the architecture of the data center you blow up all the traditional boundaries that define what people do for a living—how they develop their careers, how they got promoted, and everything else. It's a big impediment for technology adoption.

SIMON CROSBY: One of the reasons I think that cloud-based IT is very interesting is none of the cloud vendors have invested in the disaster of today's typical enterprise IT infrastructure. It is horrendously expensive because none of it works together; it's unmanageable except with a lot of people. Many enterprise IT shops have bought one or more expensive proprietary subsystems that impose significant labor-intensive requirements to make it all work.

Clouds are way cheaper to operate because they build large, flat architectures that are automated from the get-go, making the cost for their infrastructure much lower than most companies' enterprise IT. If I'm Amazon Web Services and I want to offer a disaster recovery service, the numbers are in my favor. I need only provide enough additional capacity to address the expected failure rate of my combined customer set, plus a few spares and, just like an actuary, determine the risks and cost. A very simple and compelling business model.

ALLEN STEWART: The thing that challenges the cloud environment and most enterprise data centers is the heterogeneity of the shop and the types of applications they run. To take advantage of the cloud, you have to develop an application model that suits disconnected state and applications. I think that challenges enterprise IT

shops, because they look out and see a completely dissimilar range of applications without a common development framework.

GUSTAV: I just built two data centers and fully populated them. If I look at the stereotypical cloud case right now, EC2 (www.amazon.com/ec2) is about \$0.80 per hour per eight-CPU box. My cost, having bought the entire data center, is between \$0.04 and \$0.08.

Having bought the entire data center, I have the budget and scale to blow away that \$0.80 EC2 pricing. SMBs (small- and medium-size businesses) probably do not have that option. The cloud guys can produce tremendous margin for themselves by producing the scale of an entire data center and selling parts of it to SMBs.

TOM BISHOP: The model that's going to prevail is exactly the way the power companies work today. Every company that builds power-generation capacity has a certain model for their demand. They build a certain amount of capacity for some base level demand and then they have a whole set of very sophisticated provisioning contracts.

MACHE CREEGER: Or reinsurance.

TOM BISHOP: Reinsurance to basically go get electricity off the grid when they need it. So the power we get is a combination of locally generated capacity and capacity bought off the grid.

SIMON CROSBY: As a graduate student, I read a really interesting book on control theory that showed mathematically that arbitrage is fundamental to the stability of a market and the determination of true market price. Based on that statement, virtualization is just an enabler of a relatively efficient market for data center capacity; it's a provisioning unit of resource.

Virtualization allows for late binding, which is generally considered to be a good thing. Late binding means I can lazily (that is, just-in-time) compose my workload (a VM) from the OS, the applications, and the other relevant infrastructural components. I can bind them together at the last possible moment on the virtualized infrastructure, delaying the resource commitment decision as long as possible to gain flexibility and dynamism. Virtualization provides an abstraction that allows us to late bind on resources.

STEVE HERROD: The opportunity to

have a VM and to put the policy around that VM for late binding is pretty powerful. You create your application or your service, which might be a multi-machine service, and you associate with it the security level you want, the availability level you want, and the SLAs (service level agreements) that should go with it. The beauty of this bubble, which is the workload and the policy, is it can move from one data center to another, or to an offsite third party, if it satisfies the demands that you've wrapped around it.

GUSTAV: Our administrative costs generally scale in a nonlinear fashion, but the work produced is based on the number of operating-system instances more than the number of hardware instances. The number of servers may drive some capital costs, but it doesn't drive my support costs.

TOM BISHOP: What you're really managing is state. The more places you have state in its different forms, the more complex your environment is and the more complex and more expensive it is to manage.

SIMON CROSBY: I'll disagree. You're managing bindings. The more bindings that are static, the worse it is, the more they are dynamic, the better it is.

We have a large financial services customer that has 250,000 PCs that need to be replaced. They want to do it using VDI (virtual desktop infrastructure) running desktop OSes as VMs in the data center to provide a rich, remote desktop to an appliance platform.

Following the "state" argument, we would have ended up with 250,000 VMs consuming a lot of storage. By focusing on bindings, given that they only support Windows XP or Vista, we really need only two VM images for the base OS. Dynamically streaming in the applications once the user has logged in allows us to provide the user with a customized desktop, but leaves us with only two golden-image VM templates to manage through the patch-update cycle.

Steve Herrod, Mike Neil from Microsoft, and I have been working on an emerging standard called OVF to define a common abstraction to package applications into a container. Under this definition, an application is some number of template VMs, plus all the

STEVE HERROD

The opportunity to have a VM and to put the policy around that VM for late binding is pretty powerful. You create your application or your service. The beauty of this bubble, which is the workload and the policy, is it can move from one data center to another, or to an offsite third party, if it satisfies the demands that you've wrapped around it.

metadata about how much resource they need, how they're interconnected, and how they should be instantiated.

We started working on it because there was the potential for a "VHS versus Betamax" virtual-hard-disk format war and none of us wanted that to happen. It started out as a portable virtual-machine format but is now emerging into more of an application description language. The container contains one instance of every component of the application, but when you roll it out at runtime you may request multiple copies. I think that's a very important step forward in terms of standardization.

STEVE HERROD: Virtualization breaks up something that's been unnaturally tied together. However, allowing late binding introduces new problems. If you cannot be more efficient with virtualization, then you shouldn't be using it.

We do surveys every single year on the number of workloads per administrator. Our numbers are generally good but it is because we effectively treat a server like a document and apply well-known document-management procedures to gain efficiencies. This approach forces you to put processes around things that did not have them before. For smaller companies that don't have provisioning infrastructure in place, it allows you much better management control. It's not a substitute for the planning part, but rather a tool that lets you wrap these procedures in a better way.

MACHE CREEGER: So how do people decide whether to choose VMware, Citrix, or Microsoft? How are people going to architect data centers with all the varying choices? Given that the vendors are just starting to talk about standards and that no agreements on benchmarking exist, on what basis are people expected to make architectural commitments?

GUSTAV: I think this is a place where the technology is ready enough for operations, but there are enough different management/software theories out there that I fully expect to have VMware, Microsoft, and Xen in different forms in my environment. That doesn't concern me nearly as much as having both SuSE and RedHat in my environment.

TOM BISHOP: Every customer we talk

to says they'll have at least three.

MACHE CREEGER: As a large enterprise customer, aren't you worried about having isolated islands of functionality?

GUSTAV: No, I have HP and Dell. That's a desirable case.

MACHE CREEGER: But that's different. They have the x86 platform; it's relatively standardized.

SIMON CROSBY: It's not. You'll never move a VM between AMD and Intel—not unless you're foolhardy. They have different floating point resolution and a whole bunch of other architectural differences.

People tend to buy a set of servers for a particular workload, virtualize the lot, and run that workload on those newly virtualized machines. If we treated all platforms as generic, things would break. AMD and Intel cannot afford to allow themselves to become undifferentiated commodities, and moreover, they have a legitimate need to innovate below the "virtual hardware line."

MACHE CREEGER: So you are saying that I'm going to spec a data center for a specific workload—spec it at peak, which is expensive—and keep all those assets in place specifically for that load. Doesn't that fly in the face of the discussions about minimizing capital costs, flexibility, workload migration, and high-asset utilization?

TOM BISHOP: You're making an assumption that every business defines risk in the same way. Gustav defines risk in a particular way that says "The cost of excess capacity is minuscule compared to the risk of not having the service at the right time."

MACHE CREEGER: In financial services, that's true, but there are other people that can't support that kind of value proposition for their assets.

SIMON CROSBY: That's an availability argument, where the trade-off is between having the service highly available on one end of the line, and lower capital costs, higher asset utilization, and lower availability at the other end. Virtualization can enhance availability.

GUSTAV: You will tend to use the VM, because while there are differences now at the hypervisor level, those differences are converging relatively rapidly and will ultimately disappear.

If you're worried about the long-

ALLEN STEWART

The thing that challenges the cloud environment and most enterprise data centers is the heterogeneity of the shop and the types of applications they run. To take advantage of the cloud, you have to develop to an application model that suits disconnected state and applications.

term trend of hypervisors, you're worried about the wrong thing. Choose the VM that is most compatible today to the application you are going to run. If you're doing desktop virtualization, you're probably going to use CXD (Citrix Xen Desktop). If you're doing Windows server virtualization, you're going to use either Veridian or, depending on what you're trying to do regarding availability management, VMware.

The first question to ask is "What are you used to?" That's going to determine what you're likely VM is. The second question is "What is the problem you're trying to solve?" The more complex the management problem the more attractive an integrated tool suite from VMware becomes. If you are saying "I don't have complex problems now but I'm going to have complex problems in three or four years," the more attractive Microsoft becomes. If you are going to build it on your own and/or have your own toolsets to integrate, which is most of the enterprise, you're going to find the Xen/Citrix option more attractive. If you're coming from the desktop side, you're at the other side of Citrix, and that is back to Xen. Where you're coming from is going to determine your VM product selection much more than where you're going to because they're all heading to the same place.

SIMON CROSBY: Looking at Microsoft Hyper-V/System Center and VMware VSX and Virtual Center (VC), both of these are complete architectures. Neither of them has a well-established ISV ecosystem significantly limiting customer choices. That said, I think the ecosystem around VMware is now starting to emerge due to the adoption of standards-based APIs.

What worries me is whether the missing functionality in any vendor's product needs to be developed by the vendor or whether the customer is OK with a solution composed of a vendor product and ISV add-ons. Both Stratus and Marathon offer fault-tolerant virtual machine infrastructure products using Citrix XenServer as an embedded component. That's because they focus on how to build the world's best fault tolerance whereas Citrix, VMware, and Microsoft do not. We have an open architecture, and that allows

the world's best talent to look at how to extend it and build solutions beyond our core competence. This is a very powerful model.

From an architectural perspective, I am absolutely passionate about the fact that virtualization should be open because then you get this very powerful model of innovation.

I have an ongoing discussion with one of the major analyst organizations because virtualization in their brains is shaped like VMware's products are shaped today. They think of it as a thing called ESX Server. So if VMware's ESX Server is viewed as a fully integrated car, then Xen should be viewed as a single engine. I would assert that because we don't know how virtualization is going to be in five years, you do not want to bind your consumption of virtualization to a particular car right now. As technology innovation occurs, virtualization will take different shapes. For example, the storage industry is innovating rapidly in virtualization and VMware cannot take advantage of it with their (current) closed architecture. Xen is open and can adapt: Xen runs on a 4,096 CPU supercomputer from SGI and it runs on a PC. That is an engine story; it is not a car story.

It's really critical we have an architecture that allows independent innovation around the components of virtualization. Virtualization is just a technology for forcing separation as far down the stack as you can—on the server, separated by the hypervisor, in the storage system—and then let's see how things build. I'm not in favor of any architecture which precludes innovation from a huge ecosystem.

STEVE HERROD: I actually agree on several parts. Especially for the middle market, the number-one thing that people need is something easy to use. I think there's a reasonable middle road which can provide a very nice framework or a common way of doing things, but also have tie in to the partner ecosystem as well. Microsoft has done this very well for a long time.

STEVE BOURNE: These bindings may be ABIs or they may not be, but they sound like the analogue of the ABIs. ABIs are a pain in the neck. So are these bindings a pain in the neck?

SIMON CROSBY: Bindings are a very hot

area. The hottest one for us right now is the VM you run on XenServer will run on Microsoft Hyper-V. This is a virtual hardware interface, where, when you move a virtual machine from one product to the other, the VM will still think it has the same hardware underneath it.

Right now if you take a VM from VMware and try to run on Citrix you will get a blue screen. It's just the same as if you took a hard disk out of a server and put it in another server and expected the OS to boot correctly. VMware and XenSource actually had discussions on how to design a common standard hardware ABI, but we couldn't get other major vendors to play.

If we actually were able to define an industry standard virtual hardware ABI, the first guys who'd try to break it would be Intel and AMD. Neither of those companies can afford for that line to be drawn because it would render all their differentiation meaningless, making their products undifferentiated commodities. Even if you move everything into the hardware, the ABIs would still be different.

In the ABI discussion there are two things that count. There's "Will the VM just boot and run?" Then it's "If the VM is up and running can I manage it using anybody's management tool?" I think we're all in the same position on standards-based management interfaces—DMTF (Distributed Management Task Force) is doing the job.

MACHE CREEGER: Let's take a moment to summarize.

Server consolidation should not be a focus of VM deployment. One should architect their data center around the strengths of virtualization such as availability and accessibility to clouds.

An IT architect should keep the operating application environment in mind as he makes his VM choices. Each of the VM vendors has particular strengths and one should plan deployments around those strengths.

In discussing cloud computing we said the kind of expertise resident in large enterprises may not be available to the SMB. Virtualization will enable SMBs and others to outsource data center operations rather than requiring investment in large, in-house facilities. They may be more limited in the types of application services

available, but things will be a lot more cost effective with a lot more flexibility than would otherwise be available. Using their in-house expertise, large enterprises will build data centers to excess, and either sell that excess computing capacity like an independent power generator or not, depending on their own needs for access to quick capacity.

GUSTAV: The one point we talked around, that we all have agreement on, is that server administrators will have to learn a lot more about storage and a lot more about networks than they were ever required to do before. We are back to the limiting constraint problem. The limiting constraint used to be the number of servers you had and given their configuration, how they were limited in what they could do. Now with virtualized servers the limiting constraint has changed.

With cheap gigabit Ethernet switches a single box only consumes 60 to 100 megabits. Consolidate that box and three others into a single box supporting four servers and suddenly I'm well past the 100 megabit limit. If I start pushing toward the theoretical limits with my CPU load, which is 40-to-1 or an average of 2%, suddenly I've massively exceeded GigaE. There is no free lunch. Virtualization pushes the limiting constraint to either the network or to storage; it's one of those two things. When we look at places that screw up virtualization, they generally over consolidate CPUs, pushing great demands on network and/or storage.

You shouldn't tell your management that your target is 80% CPU utilization. Your target should be to utilize the box most effectively. When I have to start buying really, really, high-end storage to make this box consolidatable, I have a really big problem. Set your target right. Think of it like cycle scavenging, not achieving maximum utilization. When you start by saying "I want 100% CPU utilization," you start spending money in storage and networks to get there that you never needed to spend. That is a very bad bargain. **C**

Mache Creeger (mache@creeger.com) is the principal of Emergent Technology Associates, marketing and business development consultants.

© 2008 ACM 0001-0782/08/1100 \$5.00