PARTNER'S GUIDE TO THE

Future of Availability

SHUTTERSTOCK

Existing high-availability and business-continuity technologies continue to be effective and to evolve. But the emergence of cloud and the ubiquity of virtualization begs a question: How will availability in the near future be defined? **By Scott Bekker**

Availability is a critical part of computing, especially for enterprise applications. Order-processing systems and other mission-critical programs can be incredibly costly when they go down. The infrastructure for these types of applications is largely standing in place in the datacenter while everything else moves at varying speeds into the cloud.

The cloud is many things, but a mission-critical platform for essential business applications isn't yet one of them. That state of affairs won't last forever. In fact, some forward-thinkers are already working out how to make the cloud more amenable to high-availability (HA) applications. When those technologies become generally available, several opportunities could open up for solution providers and other partners.

COST OF DOWNTIME

One thing that cloud won't change is the hard dollar cost of IT downtime for organizations. In a report titled, "Datacenter Downtime: How Much Does It Really Cost?" researchers at Boston-based Aberdeen Group analyzed the frequency and associated costs of downtime for 134 organizations in February 2012.

Aberdeen came up with a cost-per-hour figure of $138,000. Cloud may not change the cost, but other factors are actually making downtime more expensive. In a similar research report two years earlier, Aberdeen researchers pegged the hourly cost of downtime at nearly $98,000. Aberdeen attributed the 38 percent jump in costs in large part to increasing automation, making system failures that much more expensive.

Keep that hourly figure in mind when thinking about standard uptime guarantees. A conventional unmanaged system hits about 99 percent uptime. That sounds good until you do the calculation that 1 percent downtime amounts to 87.6 hours over the course of a year. The substantially more stringent 99.9 percent uptime allows for 8.7 hours of downtime.

The seriousness with which end users approach availability leads to widely variant downtime costs over the course of a year. Aberdeen split the organizations it studied in 2012 into three groups. A best-in-class group consisting of the top 20 percent of the organizations averaged 0.3 business interruption events for a total annual disruption of .03 hours and a cost of downtime of $3,000. A laggard group of the bottom 30 percent averaged 4.4 interruptions, 39.6 hours of downtime and $3.9 million in downtime costs. The middle 50 percent averaged more than $400,000 in downtime costs per year.

Tallying those downtime costs is important because availability can be a tricky concept to sell.

"When such solutions are implemented correctly, the end result is that *nothing happens* [emphasis original]." Aberdeen's research looked at lost revenue as a key measure to determine the impact of downtime, and discovered that "even above-average organizations were absorbing staggering costs each year," the report stated.

SERVER-SPECIFIC AVAILABILITY

Certainly, the applications have to be worth something to make availability worth the effort. Aberdeen approached companies with quantifiable downtime costs. Other organizations may have downtime costs that run only into the hundreds of dollars a year despite lengthy outages, not a great case for an intensive availability project.

Those who have HA needs know who they are, and what their precise business case is. To date, HA applications are often associated with a particular server. That server tends to be lovingly administered and carefully maintained. Perhaps the server is clustered, using standard Windows or Linux tools or a more specialized kit. The server cluster might have HA links to a remote datacenter for geographic redundancy. More recently, those servers might have added robust virtualization schemes that allow the application to pop back up immediately if it winks out of service for some unforeseen reason.

> In the public cloud, a three-nine guarantee is the standard.

THE TROUBLE WITH CLOUD

All of that differs substantially from cloud environments, be they public, private or hybrid. The design parameters for cloud environments tend to call for lots of commodity servers and storage that can be swapped out easily in the event that a component fails. When it comes to the abilities, the priorities are scalability, flexibility and affordability. HA isn't generally in the running.

A case in point is the availability guarantees of the most popular cloud services. Traditional HA customer conversations tended to focus on levels such as five nines (99.999 percent), four nines (99.99 percent) and, at the low end (almost out of the realm of serious conversation), was three nines (99.9 percent).

In the public cloud, a three-nine guarantee is the standard. That's the uptime Microsoft guarantees for most of its Azure services. Amazon Web Services EC2, meanwhile, offers a 10 percent service credit when uptime falls between 99.95 percent and 99.0 percent and a 30 percent service credit if uptime falls below 99.0 percent. Both companies have caveats written into the agreements that carefully define which kinds of outages count as downtime and which outages don't.

Also, in traditional HA environments, HA policies tend to be set at the infrastructure layer, independently of the application. In most clouds, however, the application must often include its own HA logic.

THE PROMISE OF CLOUD

Problems with cloud aside, more and more infrastructure is moving there. As expertise improves and more of the core infrastructure of organizations takes up residence in a private cloud, HA applications will occupy a lonely place in the datacenter. Advantages of cloud generally also apply to HA applications. The elasticity of cloud implementations can have an especially big upside for mission-critical applications that are subject to peaks in demand.

While there's little in the way of products yet, some core concepts are emerging that show a possible path for bringing mission-critical applications to the cloud.

One is the concept of Software-Defined Availability (SDA), which is a variation on the "software-defined" terminology on areas of infrastructure from Software-Defined Networking (SDN) to Software-Defined Datacenters (SDDC) to Software-Defined Storage (SDS). In most of those cases, the idea is to separate infrastructure into planes, with administrators interacting with a software-based control plane, where they can set policies and make changes that are executed in an operations plane.

Pioneering SDA is Stratus Technologies, which is using decades of experience in HA and fault tolerance to create a beta program for an implementation on the OpenStack open source software cloud computing platform.

"Stratus is applying its unparalleled availability expertise to cloud environments with a new suite of software solutions that will provide organizations with all the benefits that cloud computing offers with less of the risks," Nigel Dessau, chief marketing officer at Stratus, said in a May statement. "Building availability into the cloud infrastructure itself will significantly improve the quality of cloud SLAs while removing the availability worries and cost barriers that prevent legacy applications from moving to the cloud."

Although Stratus builds fault-tolerant servers for Windows and Linux, the company's cloud availability solution will draw on its software heritage. The Stratus package will consist of three parts: an Availability Engine, an Orchestrator and a

Hourly Cost of Downtime =

$138,000

(ABERDEEN GROUP, 2012 ESTIMATE)

"When [availability] solutions are implemented correctly, the end result is that *nothing happens* ... [but] even above-average organization were absorbing staggering costs each year."

ABERDEEN GROUP REPORT, "DATACENTER DOWNTIME: HOW MUCH DOES IT REALLY COST?"

Portal/Service Catalog. The Availability Engine will sit at the cloud infrastructure level and include a fault-tolerant KVM hypervisor for use in the cloud. The Orchestrator uses smart tags to define an application by availability requirements, resource needs, performance characteristics and compliance requirements. The Orchestrator uses that information to automatically deploy workloads to the appropriate locations in the infrastructure. Finally, the Portal and Service Catalog is a dashboard for streamlining application provisioning and for monitoring application health.

The Stratus thinking is that users would be able to set up a catalog of services that would allow IT administrators to pick the level of availability they need for specific applications. For example, e-mail might be "Standard," accounting might be "Business Critical" and an OLTP system would be "Mission Critical." In that same scenario, at the end of a quarter, an administrator would be able to switch accounting to the more robust and available (and therefore expensive to use) part of the private cloud known as "Mission Critical." Once the financial reporting is done, accounting could be switched back to a less resource-intensive portion of the cloud.

SHUTTERSTOCK

PARTNER OPPORTUNITIES

The initial Stratus beta program is primarily for enterprise customers, and only time will tell if the brand-new approach to bringing HA to clouds will catch on. In any case, the idea of highly available components of cloud infrastructure could lead to a number of new opportunities for solution providers, be it on the Stratus infrastructure or other implementations.

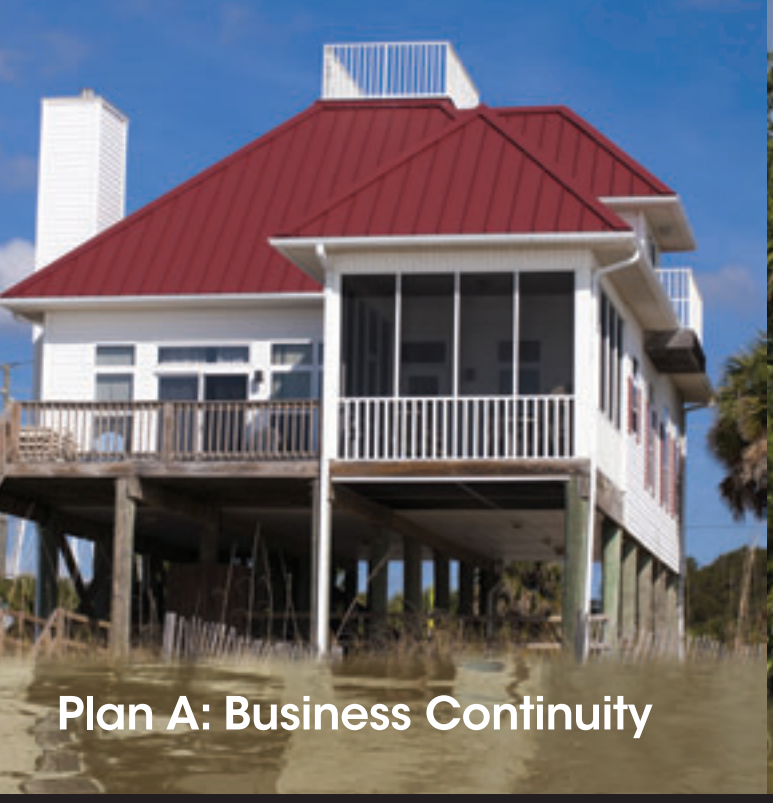
* **Systems Integrators:** One of the largest opportunities in HA would be for systems integrators capable of setting up private clouds for enterprise customers. Hybrid and private clouds are complicated to set up in the first place, and systems integrators are already in demand for their ability to help with configurations and deployments. Expertise in migrating mission-critical applications to the private cloud in a highly available way is another way systems integrators will be able to bring value to customers.

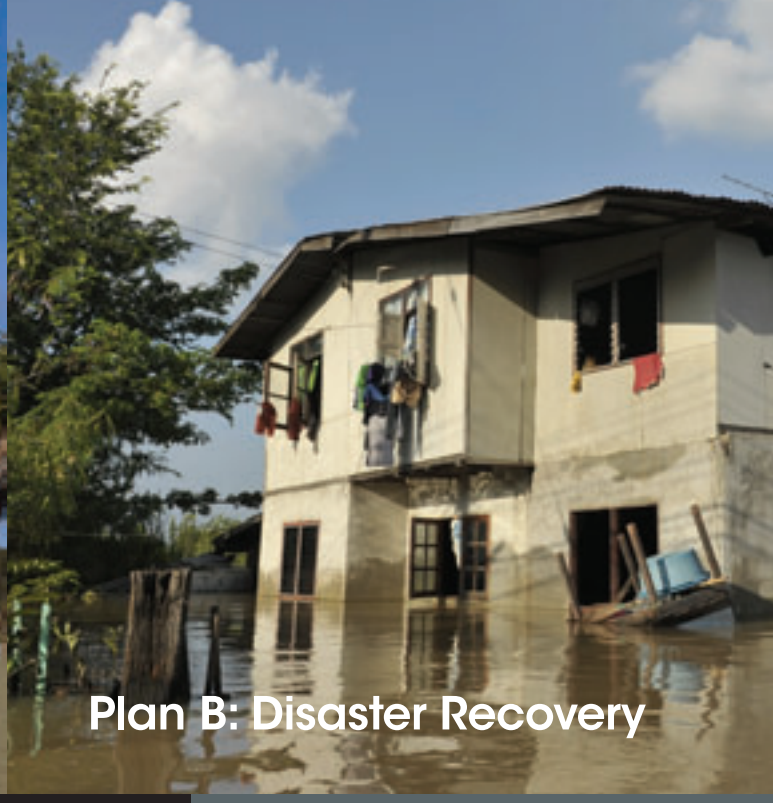
* **Hosting Providers:** Many private and hybrid clouds depend on hosting providers for all or part of their infrastructure. Offering availability services, especially a menu of several different tiers, would be a way for hosters to differentiate themselves. There's also an opportunity in availability for hosters with vertical specializations. Some of the verticals especially sensitive to downtime include manufacturing, retail, public safety, financial services, health care and transportation.

* **Custom Application Developers:** Ideally, HA infrastructures in private clouds eliminate the need for any changes to applications to make them highly available. In cases where the need doesn't rise to the level of requiring a highly available subcloud within a private cloud, organizations will need help rewriting their applications to improve availability within a cloud infrastructure. Such rewrites can be time-consuming and risky, making a custom-application development partner with experience across a range of such projects attractive to a customer that hasn't ported applications before.

* **Consultants:** Many enterprise organizations will have the expertise to configure their own private clouds, but may look to consulting partners for expertise in architecting highly available subclouds within their private cloud infrastructures. Financial services, telecommunications and hosting providers may be among the organizations looking for availability expertise even as they configure and deploy their own clouds. Consultants with availability expertise are also a logical partner-to-partner fit with systems integrators building more standard private clouds. •

Scott Bekker is editor in chief of Redmond Channel Partner.





Which IT Operational Plan Best Suits Your Customers?

Prevention Trumps Recovery

While you certainly want your customers to be in a position to recover quickly from a disaster, wouldn't it be great if they could be prevented in the first place? Everyday server faults, drive crashes and human errors account for 73% of downtime in the typical organization. Invoking a disaster recovery plan for these frequent and relatively short-term events just doesn't make business sense in terms of cost or responsiveness. Instead, just like the forward-thinking owner of the home depicted in "Plan A" above - who built his house on stilts, your customers need a business continuity strategy that enables them to work through system faults or site-level problems to keep disruptions from occurring in the first place.

Learn More

Download Stratus Technologies' informative white paper, "Focus on Business Continuity Instead of Disaster Recovery," at go.stratus.com/continuity to learn why preventing everyday "disasters" is more practical than trying to recover from them after the damage is done.

Stratus® Technologies™ For an **Always-On** World