

COMPUTERWORLD

Maximizing Server Uptime: Best practices

By John Edwards

Computerworld - In an IT world full of elusive goals, there's probably no target as slippery and generally elusive as server uptime. Keeping servers alive and awake, or at least ready to instantly spring into action whenever needed, is an ambition close to the heart of virtually all data center leaders.

Six steps to maximizing server uptime

1. **Plan carefully.** Aggressively enforce life-cycle management, and double-check the work, including system configurations and maintenance schedules. Server acquisitions and upgrades should be scheduled and coordinated with an eye toward system availability as well as performance.
2. **Practice routine preventive maintenance.** This is perhaps the easiest and least painful way of bolstering server reliability. As the old car-repair commercial warned, "You can pay now or pay later."
3. **Use management and monitoring tools.** Without adequate oversight, you can't get to the root of uptime-robbing server problems or measure downtime's impact on critical business services.
4. **Bolster security.** Don't let attackers interfere with your uptime goals. Anti-malware products, firewalls and independent audits are among the many security tools and practices that have a positive influence on server uptime.
5. **Acquire quality hardware.** The road to downtime is paved with trashy servers.
6. **Use common sense.** Don't waste time, energy and money trying to squeeze the last drop of life out of an aging or problem-prone server.

Yet few managers can honestly say that they are doing absolutely everything to squeeze the most uptime out of their systems. Indeed, many managers needlessly lavish time and funds on technologies and practices that have little or no positive impact on uptime, experts say.

Achieving server uptime excellence is both a science and a management art, says Walter Beddoe, vice president of IT and logistics at Six Telekurs USA, a financial data provider in Stamford, Conn. "It's a combination of many different things, including having a competent staff, using fault-tolerant hardware, adopting dynamic security practices, and embracing good maintenance and change management practices," he says. "Most of all, you must have a commitment to doing your very best."

Alan Howard, IT director at Princeton Radiology, a diagnostic medical imaging firm in Princeton, N.J., urges managers not to waste time and resources on activities and tools that don't directly contribute to uptime enhancement. The effort put into clustering, for example, can be "pretty wasteful," he says, noting that redundancy is better achieved with a tool that provides full automation.

Clustering that is not automated -- where the synchronization is done manually -- can cause more problems than it's worth, Howard says. "A failure of a primary node can cause havoc; we'd have been better off simply recovering from the primary-node failure than failing to the standby node," he says.

For instance, his shop had a Windows Server cluster that, upon failover, would cause the application to crash because a change to an application configuration file had not been applied to the stand-by server. "The effort to

fix the cause of the application crash tended to be much more than the effort to fix the cause of the cluster-node failure," Howard says.

His shop no longer provisions clustered servers in the traditional sense. Instead, he has a "cluster" of stand-alone servers -- all mapped to a dual-controller Compellent Storage Center SAN -- "among which we can migrate virtual machines on demand quite seamlessly."

Getting organized

Most managers agree that carefully planning all server-related work, from acquisition to management to replacement, is a key step in guaranteeing system reliability.

Raoul Gabiam, IT operations and engineering manager at George Washington University, says life-cycle management is an integral part of [server uptime planning](#) in his shop. "Knowing when and how to replace hardware and upgrade software is important, as it affects performance, sustainability and overall uptime," he says.

For example, if you have to perform a software upgrade, understanding the hardware requirements and the state of your current hardware is critical. You may want to buy the hardware as part of the software upgrade to ensure that requirements are met and to avoid further outages, or perform one before the other to minimize the number of changes, Gabiam explains.

Gabiam is also a strong believer in standardization and coordination as a way of ensuring reliable server operation. "Before anybody installs anything or makes a change, there has to be a change management process," he says.

Change management means knowing "how everything is configured and stood up, and [evaluating] the changes before they're implemented," Gabiam says. "That way, you'll always know how things are supposed to be and how things will interact."

He says the discipline of change management makes it possible to predict how servers will react when configured in certain ways or if placed into a new environment.

Paul Franko, chief technology officer at Online Resources, a Chantilly, Va.-based company that provides transaction services to financial institutions, says attitude plays a big role, too. He says he makes an extra effort to ensure that routine yet critical server-related tasks are taken seriously and addressed promptly.

"We've put in a system of checks and balances to make sure that our policies are being followed," he says. According to Franko, having managers routinely examine staff members' administrative work, along with double-checking in other ways, helps minimize the impact of human error. "People make mistakes, and if you don't have multiple points of verification, then things are going to slip through the cracks," he explains.

Practice preventive maintenance

Routine preventive maintenance is perhaps the easiest and least painful way of bolstering server reliability. "Your uptime is only as high as the weakest component in the delivery chain," Beddoe says. Performing a variety of essential tasks -- updating system software, providing conditioned power and ensuring adequate cooling -- can go a long way toward creating a data center full of happy servers without breaking the budget or distracting staff members from other vital tasks.

To ensure that all necessary work is performed when required, server maintenance tasks should be identified and organized into a schedule, says Franko. "There are certain things that need to go [into place] straight away --

like security updates -- and there are other things that make sense to batch up and apply at regular intervals." This second category includes software updates with non-critical functionality improvements, for example.

Franko adds that maintenance work should be handled in such a way that the practice itself doesn't steal server uptime. "We don't take the system down for doing certain types of maintenance activities -- we strive for that, anyway," he says.

When it's essential to pull down a server for maintenance, Franko's team schedules the work for an overnight or weekend time frame when user demand is low. The only legitimate reason for pulling a functional server down during regular business hours would be the installation of a critical software update, such as the application of a zero-day security patch.

Automate essential server-management tasks

It's no secret that [server management](#) has become significantly more complex over the past several years, mostly due to the arrival of virtualization and related technologies and practices designed to increase server efficiency and utilization.

Virtualization itself helps protect data centers from the effects of server downtime. By consolidating servers and connecting them into a shared environment, virtualization allows multiple virtual machines to run on different hosts. Failure of any one host will result in the workload being redistributed across the remaining hosts. "You may get a server failure, but that doesn't mean it has to impact the service," Gabiam observes.

To manage this increasingly virtualized environment, vendors such as Xenos Software, [Uptime Software](#), Nimsoft and Nagios Enterprises offer tools that are designed to help data center staff keep an eye on server performance, spot emerging problems and capitalize on performance improvement opportunities.

Beddoe feels that such tools are essential. "You need to have some reassurance that all of your servers are doing what they're supposed to be doing at all times," he says.

Make sure your tools trigger alerts

Beddoe, who uses uptime management software from [Uptime Software](#), says it's important to look for a tool that can trigger an alert whenever a server condition crosses a specific threshold, such as when memory overload or excessive CPU utilization occurs.

While most tools come with built-in alert functions, Beddoe stresses the need to look for a product with configurable warnings -- thresholds that trigger e-mails or SMS messages. "You need meaningful information so that you can take the steps necessary to correct the situation -- whatever it is that works for your environment, including alerting on the big screen monitor for your operations staff," he says.

Jerry Gregg, operations manager at Carfax -- a Centreville, Va.-based company that generates vehicle ownership reports -- says it's important to understand that the uptime rates calculated by many performance measurement tools are only approximations. "They're rough guidelines, at best," he notes.

Gregg observes that some basic uptime measurement tools can actually be deceptive, because they can't adequately differentiate between an hour-long server outage that occurred on a sleepy Sunday morning and a 10-minute failure that hit on a Thursday afternoon when scores of business-critical processes were running. That's why it's a good idea to invest in measurement tools that provide full time- and event-based analytical capabilities, he suggests.

To make uptime analysis more meaningful, Gregg relies on measurements that show the impact of server failures on key business services. Gregg uses BMC Software's ProactiveNet Performance Management software to directly correlate server downtime with sales transactions and other kinds of service-oriented business data. "It allows us to quantify the impact of an outage not just in time, but in dollars," he says.

Information generated by the application helps him determine whether a pattern of failures is threatening to make a significant dent in the company's bottom line, justifying the expense of new servers, better network gear or other reliability-enhancing technologies and services. "Without this information, you're making cost-benefit decisions without really knowing the cost," Gregg says.

Don't let hackers steal your uptime

Security also plays an important role in ensuring server uptime. Not surprisingly, servers that are compromised by malware or unsecured network paths are more likely to go down than their well-protected counterparts. "You start off with physical security -- your data center building -- and making sure that it's physically secure," Beddoe says.

Next, it's important to have server-access rules that are known and enforced, secure shelves, antivirus programs, firewalls and disciplined administrators, he says. "They all play an equally important role in server security and promoting uptime."

John Luludis, who supervises server operations for Superior Technology Solutions, an IT consulting firm and custom software developer in Pearl River, N.Y., says that to really ensure maximum server uptime, it's important to move beyond basic security practices. Luludis is a strong believer in regular independent security audits. "I have my network go through penetration tests on a regular basis, and I do that because as much as I may think that my network is secure, it's also important to have an outside point of view," he says.

Protect your data

While Princeton Radiology's Howard is also a strong believer in regular server maintenance, he notes that some amount of failure is inevitable despite the best efforts of both managers and employees. To guard against any data losses caused by server failure, Howard recommends developing a data protection plan that's tied into the enterprise's comprehensive business continuity strategy.

Princeton uses an off-site storage solution from Compellent Technologies to replicate all of its stored data. "Even though it's a disaster recovery data center, we actually run some servers primarily from that site, so we replicate in both directions," Howard says.

Gabiam, meanwhile, relies on the load-balancing technology built into his network infrastructure to protect against sudden server failure. "If one server crashes or one application becomes unresponsive, that traffic is redirected to other, similar servers that can handle the load," he says.

Unlike Princeton's Howard, Gabiam is a fan of clustering and uses Novell Cluster Services to provide an additional layer of redundancy, Gabiam says. If one of the cluster nodes fails, or needs down time for maintenance, the clustered application or component of a service running on that node can run seamlessly on another node in the cluster, he explains.

This migration process can be configured to be manual or automatic fail over. "Usually, you would want the application to automatically fail over to the next preferred node in the event of a hardware or software failure," Gabiam says, but administrators could initiate a migration to another node if they needed to perform maintenance on a specific node.

Look at hardware quality

Acquiring quality servers rather than cut-rate boxes or blades is an obvious way to enhance long-term server reliability. "There's a decided difference in the longevity of hardware as you move to midgrade or high-grade servers," says Jeffrey Driscoll, director of operations at E-N Computers, an IT services provider based in Fishersville, Va.

Yet in the real world, budget-strapped managers often face a painful choice between meeting their server needs with low-cost products or acquiring better, more reliable systems that meet established performance criteria. What to do?

Driscoll advises shopping intelligently, looking for bargains and, whenever possible, working with management to get a budget that reflects real-world operational needs. It's also not a bad idea to show management the financial damage that can be caused by unreliable servers. "It's a point that can be easily proved with simple figures and projections," Driscoll says.

Know when it's time to cut your losses

Simple common sense may be the best way of ensuring maximum server uptime without breaking the budget. "Hardware is hardware. At some point, something will break," Gabiam says. "It's important to learn from whatever happened and to be ready with a plan if it ever happens again."

Using common sense also means knowing when it's time to cut your losses and move on to something new, regardless of your replacement cycle's current stage. "If your IT staff is spending 25% of its time fighting fires and supporting out-of-date systems, who wouldn't see that as a huge waste of time?" Beddoe asks.

While maximizing server uptime creates some extra work, most managers feel that the final rewards far outweigh the added exertion. "It's hard to say that any effort is wasted when it applies to uptime," Luludis says. "Anything you do can help."

Beddoe feels that striving for the most uptime almost guarantees the creation of a more reliable data center. He contends that an "active environment" -- one that continually encourages staff members to identify and squelch potential problems before they can cause any damage -- is key to [maximizing uptime](#). "In 17 years, we have not had a major outage that has impacted our clients."

John Edwards is a technology writer in the Phoenix area. Contact him at jedwards@gojohndwards.com.