

Three Tips for Cutting Server Downtime by 50%: *The Essential Reasons Why Servers & Services Go Down.*

Read on to discover reasons why many large, and even sophisticated, enterprise IT environments fail. Know these pitfalls and take proactive actions against them to cut downtime by up to 50% in your organization.

Authored by: Alex Bewley, Product Manager, uptime software

Date: April, 2007

How do we get mainframe reliability in today's distributed environments?

In today's world we need to keep our organizations up and running 24/7, with limited unscheduled downtime, or face the wrath of IT directors and CEOs. It wasn't always like this, as a few years ago big "mainframes" dotted the infrastructure landscape. These behemoths were very reliable, but required so much TLC that they gave a new meaning to the phrase 'high maintenance,' requiring four hours of downtime every third month for scheduled maintenance.

So, why aren't we still on mainframes? Well, most IT professionals have a host of answers to this question, including flexibility, cost, and ease of use to name a few. In fact, there are very few System Administrators or IT managers who long for the days of mainframes. They are near extinct, and for good reason.

Today's infrastructure environment is much more flexible, with easier 'swap in/out' features with a more distributed architecture and dynamic environments.

So, why can't we have the best of both worlds? The reliability of a mainframe with the flexibility, cost, and usability of a distributed environment?

Well, we can.

The key is to apply the principles of effective server monitoring and to understand the pitfalls that continue to take down systems today. Understanding how to manage CPU and Resource overload, Disk Space, and IT Services in your environment can help cut your unscheduled downtime by 50% or more. To find out how, read on.

Next to sudden disk space shortfalls, the most frequent cause of unscheduled downtime is a critical service stopping, or even stalling.

Tip #1: IT Services

One of the most frequent causes of unscheduled downtime for critical applications is a critical service stopping or stalling. In fact, a stalled service can be more disruptive due to the difficulty in detecting it.

In this example, let's look at an environment with Windows. There are three critical applications in Windows that need their services monitored, SQL Server, IIS, and Exchange.

While all three are typically used for mission-critical applications, Exchange seems to be most susceptible to one of its services stopping or stalling. The results can be catastrophic, and happens far too often. A four hour time span of unavailability for a major enterprises e-mail system can be extremely expensive, and cause executive management to start questioning the competency of the entire IT department. Today everyone in an organization, from the CEO to the warehouse staff, rely on e-mail to be effective in their role. People instantly notice when their e-mail is not functioning at 100%.



It is important to note two Windows services properties. Windows services are usually critical and the services are hidden. Let's look at the spooler service, for example. The Service Control Manager's applet interface (the icon is the two inter-meshed straight-cut gears) can be used to stop the spooler service. However, if you then try to print you will receive an error message about printers not being installed. Although this is just one small microcosm and doesn't sound as mission critical as losing Exchange, imagine the impact of something similar across an infrastructure and business units including HR, Operations, and Marketing.

In a few extreme cases, the production server will need to be rebooted. Fortunately, this is very rare. The bad news is that under **15%** of production servers have their services monitored.

Tip: Ensure your services are monitored and that you can generate ad-hoc or automated, scheduled reports that can be sent to your team or management on a regular (weekly or monthly) basis.

Real Life Example #1 - A Large US Credit Union with over 50 Years in business and more than \$1 Billion in assets.

Originally focused on purchasing a simple event log monitoring product, this Credit Union quickly saw the huge benefit in monitoring all aspects of their Linux and Windows infrastructure. The key was the ability to manage these systems from one solution, the up.time solution. up.time is used in this environment to monitor various systems including Linux and Windows 2000/2003 servers for Performance Metrics, Exchange, and SQL Server.

In addition, due to management pressure and soon to be in place 'Service Level Agreements,' up.time is being used to create reports documenting server uptime, CPU utilization, and more. In their search for a Service Level Management tool, this client considered purchasing multiple products to meet their service level requirements. However, after trialing up.time, they found that up.time could address all their monitoring, alerting, and reporting needs in a single product.

Today this client monitors over 450 servers and network devices and raves about up.time's ease of deployment and usability compared to some of the bigger frameworks, like HP Openview, IBM Tivoli, and BMC Patrol.

Tip #2: CPU and Resource Overload

E-mail just takes longer and longer to be delivered. Next thing you know, the CEO is knocking on your door.

Of the three major causes of critical application failure, this can seem to be the least severe, as an application can continue to limp along during times of CPU and resource overload. However, this is what can make it dangerous.

Why? It can fester over months without being noticed: nothing is broken, so nothing is fixed. What happens? The e-mail systems do not simply stop delivering e-mail, as often happens with service(s) stalling. The danger is that e-mail just takes longer and longer to be delivered. What do you think is the first application that end users tend to complain about? Exactly, e-mail. Once again, managers start to blame the IT department and escalate those complaints. Eventually the complaints filter upward to the CEO, who has noticed that his e-mail has been slower as well. Guess who gets the blame?



The underlying truth is that CPU and resource overload can have serious adverse impact on application efficiency, and especially on mission-critical applications.

Let's look at an example with SQL Server. Let's assume that SQL queries are taking increasing longer to complete, and the result is lower end-user productivity. If there are 500 end users and the typical query takes 15% longer, the lost productivity quickly adds up. It is easy to see that the price of an upgraded server is often justified through one day's lost productivity.

So now you know that more hardware is needed to solve the problem. However, how do you justify this need to management? Management will want to see hard proof of the problem before they loosen the budget reins.

There are two solutions that need to be implemented by pro-active IT management to prevent and solve this type of critical bottleneck. The first is quite simple, requiring active monitoring of CPU processing, and real-time alerting when thresholds are breached. The second includes leveraging historical performance data and running a professional report

Now you can just sit back and wait for the promotion.

that graphs this CPU growth. With this trend-line based graph and report, it becomes very clear to management that new hardware is needed to solve the problem. If your software can't produce this type of historical performance report with the detailed metrics you need, then you can't justify the answer to management. However, in this case you can justify it and you did. Now you can just sit back and wait for the promotion.

Tip: Ensure your alert thresholds are set strategically. Also, make sure you have access to the historical performance data of your key metrics, so that you have the information needed to create trend-line analysis graphs and reports that easily show key metric growth over day, week, month, or year. Nothing has quite the impact as a trend-line growth graph that clearly shows management how close you are to running out of CPU and Resources, along with the consequences of that scenario.

Real Life Example #2 - A New York based Global Investment Bank with over \$1.4 trillion in assets and operations in more than 50 countries.

Automated performance reports on over 700 Unix servers, across multiple locations, that are automatically sent out for you...now that's easy.

At this large, global investment bank, a key requirement was to accurately monitor and report on mission-critical server resources, such as CPU, I/O, and file system capacity. It was critical that these, and other key metrics, remain within the thresholds specified by IT management. If these thresholds were exceeded, it required immediate action on the part of the IT team in charge. During their short list process for finding a solution, up.time was one of the enterprise solutions considered. After a grueling head-to-head comparison, up.time was chosen.

The client now has performance reports automatically emailed and posted to web sites, allowing more than 50 IT analysts to track over 700 UNIX servers in New York and London on a daily, weekly, and monthly basis. This constant and in-depth communication ensured that the CPU, I/O, and file system capacity remain inside the right thresholds and run at optimal capacity. These reports highlighted server resource performance over the past day, week, and month, so IT management had the opportunity to take pro-active measures before any potential crisis were reached.

Real Life Example #3 - A Leading Wireless Phone Company with over 20 Million customers and \$8 Billion in Revenue

One of Canada's most respected wireless operators monitors 680 mission critical Solaris servers. The IT department wanted Service Level Management reporting to show the availability of key production servers, many running very large Oracle databases. At the same time, they were also interested in pro-active alerting to help maintain these high service levels yet didn't want multiple products.

The provider turned to up.time for its unmatched alerting reliability for critical resource thresholds – including CPU, Memory, Network I/O and File System Capacity. up.time also provided this company with the in-depth management and service level reporting that ensured in-house capacity planners had the right information to make strategic capacity related decisions.

Tip #3: Disk Space and the need to plan for it.

Disk Space should be on every IT Manager or System Administrator's expertise list, but it continues to be the major cause of outages.

We saved the number one reason for unscheduled downtime for last, and we know you might be thinking, "Are you kidding? Of course I already know that." Yet it still happens far too often, even in good IT departments. Why? Read on and find out.

Even though it might be obvious that disk space needs to be managed, it is essential to plan for disk space in light of the many other tasks in your infrastructure. When your environment is status quo, disk space is usually very easy to manage, but when unexpected events happen, you don't want to be caught with your pants down.



It is important to make certain your servers have alerting functions set to notify the right people. Too much alerting can cause a similar effect to 'crying wolf,' so alerts should only be sent to those on a 'must know' list. If it needs to be escalated, it should be escalated from there. Alerting best practices include times when disk space thresholds are approaching, not at, critical.

From here, your alerting functions should send its notifications out via email, PDA, pager, phone, or SMS, doing pretty much everything except tapping you on the shoulder. False or non-existence alerts are par for the course on some of the bigger, more complex frameworks, so make sure your alerting solution is true to form.

Remember to leave you and your teams enough time to take action against these potentially critical issues. If the alert goes unanswered by the first recipient, ensure additional alerts are scheduled automatically and sent as a failsafe.

Tip: Disk Space should be on every IT manager or system administrator's expertise list, but it continues to be a major cause of outages. Make sure your monitoring software has a robust alert system that is smart enough to keep you and your team in control, no matter what.

Real Life Example #4 - A Leading California Food Company with over 250 warehouse stores and \$500 Million in annual revenue.

An early detection was estimated to have saved over \$50,000 in just one instance.

Originally purchasing up.time to consolidate AIX servers, increase server performance and availability, and capacity plan for additional resources, this company found the extra benefits of up.time for its alerting and notifications where other products had failed.

An early detection of an unscheduled outage by up.time saved this client over \$50,000 in one instance. The scary fact is that this exact same outage was completely missed by well known large, expensive, system management framework product.

The client goes on to say, "If something goes wrong that I'm responsible for (servers, infrastructure, applications), I now know about it before it becomes an issue and can be more proactive in pinning down the problem. The historical trending shows me whether it happened quickly or slowly over time."

"Using up.time, we were also able to identify servers that were sub-optimally configured for the past two years. up.time helped us correct them to perform at peak efficiency. The new configuration resulted in improved system performance and we proved that positive change to management with up.time's tracking and reports."

For the full case study of this company, please visit here: [case studies](#)

About uptime software



“After easily deploying up.time to over 125 servers, we are seeing an immediate and significant cost savings, especially in consulting fees. In fact, time spent on monitoring and planning has dropped dramatically. This year, we’ll realize a 510% ROI from using up.time.”

- Wally Beddoe,
VP of
Technology,
Telekurs
Financial

uptime software has been providing powerful, easy-to-use, and affordable server monitoring, IT service availability reporting, and capacity planning software since 2000. Our all-in-one solution includes IT performance dashboards that help organizations eliminate unnecessary IT outages, increase service availability, and reduce the costs of server management. uptime provides all this and more in one versatile software solution called ‘up.time 4’.

up.time 4 is self-deployable with no consulting needed, and IT departments can start seeing real data in a few as 15 minutes from clicking on the free 14-day trial below. up.time 4 also leads the industry in virtualization and consolidation monitoring and reporting, including the “Server Virtualization™” report that automatically identifies and flags opportunities for consolidation and virtualization across an entire infrastructure.

Come and see why uptime software is making waves in IT operations management with ‘up.time 4,’ enterprise IT performance, availability and capacity management software that just works. To learn more about, please contact phil.didaskalou@uptimesoftware.com or call 416-868-0152

For more information, please visit:

- ✓ [Systems Management ROI Calculator](#)
- ✓ [Case Studies and White Papers](#)
- ✓ [up.time Quick Solution Tour](#)
- ✓ [See for yourself](#) – 14-day Free Enterprise Trial of up.time