



Dig a little deeper, and you will probably find that most web hosting providers do not have a viable backup solution in place. Why? Well, that's the main reason I'm writing this article. But to take it a step further, I'm going to discuss some awesome new software that is being developed to help hosting companies quickly and easily perform backups on their servers. I recently took some time to interview a fairly new company in the industry called Righteous Software ([www.r1soft.com](http://www.r1soft.com)). The crew over at Righteous Software is developing some of the most innovative technology solutions specifically designed for backing up servers, primarily targeting the web hosting industry. My interview started with David Wartell, founder of Righteous Software Inc., and was truly an experience worth discussing. David and his team are developing a software platform called Continuous Data Protection, or CDP – software for Linux and Windows operating systems that provides anything from bare-metal

# Server

recovery to snapshots to data integrity backup solutions. I began by asking Mr. Wartell a series of questions regarding the hosting industry and how backup solutions are being used in today's dedicated server market. You will definitely want to take a few minutes and read what he had to say about past and present backup solutions.

## The Current State of Backup and Restore

The first question I asked is "What is the current state of backup and restoration for the typical hosting provider including larger, more prominent hosting providers?" Once a Senior Network Engineer for EV1Servers (EV1), David's responsibilities included giving tours of the EV1 expansive datacenter facility. While giving tours, some of the more IT-savvy customers often asked about backup solutions that were implemented for the thousands of servers racked in the EV1 datacenter. David's response was the same as usual – "We don't back up these servers, it's the

customer's responsibility to back up their servers and provide their own backup solution." Even servers that did have a backup solution relied on a mixture of FTP, tar, rsync, and homegrown shell scripts. After all, in the dedicated server business ordering a "restore" doesn't mean the customer gets any of their data back; instead, it means the server gets wiped clean and the operating system is reinstalled. Is this an inconvenience to the customer? Absolutely. Most customers usually are not aware that a "restore" simply means a "wiped system with a fresh operating system installed." This concept often leads to frustrated business owners who just want their web sites and servers to be back online after a failure.

### What Are Other Industries Using for Backups?

So far, I have only discussed how the hosting industry handles backups for customers. But, what are the other industries using for backup solutions? What are the Enterprise-level and the multinational Fortune 500 companies using? Walk into their data centers, and you are likely to see a completely different picture. There you are going to see server data stored centrally on large Network-Attached Storage (NAS) systems from vendors like EMC and NetApp. You will see Storage Area Networks (SANs), Fibre Channel, robotic tape drive systems, and a team of specialized engineers to make it all work. Having a server failure in one of these environments is usually not a problem – just bring

to why hosting providers do not have access to this technology. And, don't expect these vendors to start lowering their pricing just to cater to the web hosting industry – after all, why should they, when most of their existing customers apparently have no problem paying the \$400 to \$2,000 per server price tag?

### Disaster Recovery

What are data retention policies like for web hosting providers? Is it fast or easy to restore servers after a serious disaster? Since most web hosting providers use tar and rsync, they don't really have an efficient way to store incremental backups. For example, if a provider wanted to keep the last ten backups they have made, it might require as much as five times their total disk usage in storage space – and that's assuming they use compression.

Many web hosting providers may only have one backup copy of their data. "I have seen some hosts have a server failure, only to find out their backups were failing part way through the process," said David. "Since they did not know they had a problem until it came time to restore, having only one backup copy meant they lost a lot of their customer's data. Keeping one copy might also land you into trouble if you have a failure during your backup." So, let's assume everything has gone right with the backups. You have a solid backup copy, and now its time to restore your server. This is where the best plans can go wrong. Does the

# Backups, Are You Serious?

By Dave Young

a new server online (or utilize the instant fail-over) – as all of the data is stored centrally.

These same companies are using disk based backup software from vendors like EMC, Veritas, and Computer Associates. They can take incremental disk-based backups, back up open files, and take consistent system wide snapshots. Backups are often done on a secondary NAS device, so production systems and storage are not impacted during the "backup to disk and tape" process.

If it sounds like cool technology that every web host should utilize, well, that's because it is cool – even very cool. But, implementing such complex solutions is often not practical for hosting companies in terms of cost and convenience. These advanced technologies are not suited to a market where customers tend to be small or medium-sized businesses – each one demanding their own "sandbox" to play in, and give their new toys a spin.

In addition to the hardware and storage networks, these aforementioned large companies are also paying a license fee, typically ranging from \$400 to \$1,000 per server, for backup software. If they want open file backups and snapshots, they are most likely paying an additional \$500 to \$1,000 per server for products like Open Transaction Manager (OTM).

With the cost of commercial backup software solutions sometimes requiring an entire year's revenue, it explains a lot as

operating system need to be installed first, before the restore can start? Maybe you need to order one of those "restores" from your data center to even get remote access to the host? Did you record how the disk was partitioned? Have there been changes to the boot loader?

These are all great questions for an experienced systems administrator. But even the experts don't want to deal with these hassles under the pressures of a server being down while at the same time, angry customers are demanding their services be up and running without any further delays. So what should you do? Is there an ideal solution out there?

### An Ideal Solution

When Mr. Wartell started in the hosting industry back in 1997, there was not a single viable solution for solid disk-based backup software. Another six years went by, and there was still no sign of an affordable backup solution on the market. In fact, the problem had become a lot worse than ever. So, what did David do? He formed a company called Righteous Software Inc. ([www.r1soft.com](http://www.r1soft.com)) and enlisted the help of a couple other software engineers. David said, "Our mission was to solve this problem once and for all. First we sat down and made a wish list. What would the ideal backup software for the hosting industry look like." He and his team created a list with all the features they would want to see in a backup solution including bare-metal disaster recovery,

[continued]

incremental backups in a few minutes, open file backups, encryption, compression, minimal performance impact, multi-user, web interface, quotas, and so on. "We tried to think of everything we could and put it on the list, even if it didn't sound realistic," he said.

The Righteous Software team of engineers looked at every solution available on the market – with limited success as solutions were few and far between. They examined their strengths, their weaknesses, and left nothing out. They looked at open source solutions and commercial solutions, and they tested everything they could get their hands on, finding very little in the form of features and functionality. None of the solutions they investigated had more than one or two of the features that were scribbled on the team's list.

Initially focusing on Linux, since at the time it was the most dominant operating system in the hosting industry, the team wrote a lot of C code. Many days and nights were filled writing lines and lines of C code. But after nine months of extensive research, programming, and a lot of C code thrown in the trash, the only lesson learned was they had to approach building a backup solution completely different than any other product on the market. Did the nine months of research and C coding pay off? Well, it taught them several things. One being there are fundamental flaws with backups performed at the file system level.

"The greatest efficiency and scalability was to be had backing up at the hard disk sector level. No file lists in memory, no limitations on the number of files, and the ability to backup only changed disk sectors," said David. "It took another nine months to figure out how to make it work. We learned the backup process had to bypass the file system, providing point-in-time snapshots of the hard disk in a completely consistent manner. This method also proved to be highly efficient [in] reducing CPU usage and I/O wait times when compared to other backup methods."

The second thing they learned is that in order to meet their goal of "incremental backups in a few minutes," the software had to passively track which sectors were changing between backups. This also had to be done in a way that had virtually no impact on performance, and in a way that was guaranteed to never miss a disk change, even through reboots or crashes. A full three years later, in the spring of 2006, Righteous Software finally reached a point where they had successfully invented a system for what they now call Continuous Data Protection (CDP).

"It delivers what we barely thought possible when we started," said David. "Open file backups, system wide snapshots, and the ability to reduce backup windows down to the time it takes to read changed sectors off of the hard disk. This fundamentally changes the way people should look at backups." The industry standard indicates that usually no more than 10% of a server's data changes on a daily basis. In practice, it's often much less. "With CDP, we have reduced the backup window down to the time it takes to send that changed data across the network. And, the more frequently you back up your data, the less data there is to send."

Customers are able to use the CDP software product to backup servers every 10 or 15 minutes, even on servers that are extremely busy. And because the method for backups is so efficient, there's no reason not to run backups frequently. Backups can now be performed even during peak periods, which is a significant advantage for safeguarding data when servers are running at maximum level.

The third thing they learned was how to efficiently store a large number of incremental sector-level backups. "We looked around for existing solutions and, unfortunately, we couldn't find anything," said David. "We next invented what we call Disk Safes. Disk Safes first store a seed image of a server. This includes the partition table, MBR, and even the file system

formatting. It excludes unused parts of the hard disk."

When incremental backups are performed, only disk sectors that have been changed are stored. Coupled with the nearly continuous backup technology offered by CDP, complete disk images can now be stored incrementally and reconstructed on a "backup server," which is where all of the disk-based backup data is stored for many Linux servers. "The process is so efficient it's nearly as fast as accessing the files as if they were on the live system. It supports compression, and end-to-end encryption," said David. "The software also has a built-in backup policy manager capable of retaining copies based on a variety of schedules."

David continued to discuss more of the unique features offered by the CDP software. "We often recommend to customers they should be able to store backups dating back to at least one year. Incremental backups can be taken and rotated out based on minutely [sic], hourly, daily, weekly, and monthly schedules." Another advantage to backing up at the sector level is that you can restore directly to a new disk. There is no need to first install the operating system and partition drives. David continued by stating, "Our customers simply boot servers into disaster recovery boot media using a CD or network PXE boot. And, with several mouse clicks, customers can watch the server's hard drive image stream across the network. Our customers are gratified by the capabilities of the CDP software, its efficiency, and ability to provide a nearly continuous backup for all their customers' important data."

## Was It Worth It

Based on Mr. Wartell's thorough explanation of where they started, what they have researched and coded, and where they are today, the CDP software appears capable of reducing disaster recovery down to four easy steps. The first is to resolve any hardware problems or deploy a new server. The second is to boot that server into disaster recovery mode. The third is to select the backup you want to use to restore the server. And the fourth is to sit down, relax, and watch your server's backup solution give you a complete disaster recovery, without paying the big bucks to other software vendors.

"Our software removes all barriers to an efficient restoration of a server," said David. "The only limitation left is hard disk speed and network bandwidth. That said, anyone can use it, and it's all controlled via a web interface." Moreover, there is now a solid backup solution for web hosts using Linux, and in the near future, Windows, that meets and in some cases far exceeds what the enterprises are using. And, it's also affordable. Still further, unlike other backup software vendors, Righteous Software owns all of its technology and pays no third party licensee fees for any components of its technology. "This allows us to reduce the price for backup software," said David. "For most customers we have been able to get it down to between \$50 and \$100 per server. Compared with \$400 to \$2,000 for other solutions, that do not perform well."

So there it is: a complete backup solution specifically designed for the hosting industry. Righteous Software has been able to achieve features and functionality of a truly continuous backup solution, at costs much lower than any competitor in any other industry. This technology is very cool stuff, and is well worth the asking price.

Consider this the next time you tell your customers it's their responsibility to backup their servers: Is it worth having an advanced, yet simple to use software program available to backup your servers without placing the burden on your customers? Absolutely. It's one more thing you can add to your array of features that could, in turn, generate more revenue. And, who doesn't want more revenue? 📣

Send your questions or comments to [writers@pingzine.com](mailto:writers@pingzine.com)