



A n a l y t i c s R e p o r t

2011 Backup Survey: New Possibilities for Data Protection

There's no excuse for your backup strategy being stuck in a rut. Today's system administrators can take advantage of myriad new and improved data protection technologies, including disk-to-disk backup, virtual tape libraries, data deduplication and cloud-based backups. These and other methods can help overcome two big problems: continued use of backup tapes as an archiving strategy (bad idea) and a lack of data protection for branch offices and remote employees.

By Howard Marks



A n a l y t i c s R e p o r t s

T
A
B
L
E
O
F
C
O
N
T
E
N
T
S

4	Author's Bio
5	Executive Summary
6	Research Synopsis
7	Risk Takers
9	Not Dead Yet
11	Disk to the Rescue
14	Out of Sight, Out of Scope?
15	Cut Out the Copies
17	Software Deduplication
18	Deduplication for Replication
21	Freeze Frame
21	What Is a Backup, Anyway?
23	Continuous Data Protection
26	Things Get Cloudy
27	Virtually Safe
30	A Better Way?
31	Working Without a Net
34	Nasty Surprises in Store?
35	Appendix



Analytics Reports

T
A
B
L
E
O
F
C
O
N
T
E
N
T
S

- 7 Figure 1: Percent of Physical Servers Backed Up Weekly
- 8 Figure 2: Percent of Virtual Servers Backed Up Weekly
- 10 Figure 3: Software Used to Backup Primary Physical Systems
- 12 Figure 4: Length of Time of Data Kept on Disk/VTL/Appliance
- 13 Figure 5: Length of Time of Data Kept on Tape or Removable Media
- 14 Figure 6: Backup at Remote or Branch Offices
- 16 Figure 7: Amount of Total Data Protected
- 17 Figure 8: Frequency of Restore Types
- 19 Figure 9: Highest Deduplication Ratio for Backups
- 21 Figure 10: Software Used to Back Up Virtual Machine Infrastructure
- 22 Figure 11: Backup Technologies in Use
- 24 Figure 12: Backup Technologies Planned for Use
- 25 Figure 13: Number of Physical Servers Backup Protected
- 26 Figure 14: Virtual Server Backup Methods
- 27 Figure 15: Use Same Backup System for Physical and Virtual Servers
- 28 Figure 16: Number of Virtual Servers Backup Protected
- 29 Figure 17: Reasons for Not Backing Up Servers on a Weekly Basis
- 30 Figure 18: Satisfaction Level With Current Backup System
- 31 Figure 19: Percent of Encrypted Backup Tapes
- 32 Figure 20: Data Protection Challenges
- 33 Figure 21: Frequency of Test Restores of Data or Applications
- 35 Figure 22: Company Revenue
- 36 Figure 23: Company Size
- 37 Figure 24: Job Title
- 38 Figure 25: Industry



A n a l y t i c s R e p o r t s

Howard Marks
DeepStorage.net



Howard Marks is founder and chief scientist at DeepStorage.net, a storage and networking consultancy based in Hoboken, N.J. In over 25 years of consulting, Marks has designed and implemented storage systems, networks, management systems and Internet strategies at organizations including American Express, J.P. Morgan, Borden Foods, U.S. Tobacco, BBDO Worldwide, Foxwoods Resort Casino and the State University of New York at Purchase.

He has been an a frequent contributor to *Network Computing* and *InformationWeek* since 1999 and a speaker at industry conferences including Comnet, PC Expo, Interop and Microsoft's TechEd since 1990. He is the author of *Networking Windows* and co-author of *Windows NT Unleashed* (Sams).



Executive Summary

For our latest look at the state of our backups, we surveyed 420 business technology professionals to see how widely they've adopted advances, like disk-to-disk backup, deduplication and cloud-based storage services, that can help us protect data better, for less money. We also asked about how respondents are protecting their virtualized environments and whether they've (finally) seen enough breach notifications that they flip on the encryption capabilities included in their backup software.

While our respondents come predominantly from midmarket businesses, with most hailing from organizations with fewer than 500 employees, they're largely ahead of the curve implementing new technologies in their backup practices. More than half back up at least some of their servers to disk, and half use some form of data deduplication. Respondents are also generally satisfied with their backup systems; 36% call themselves very satisfied, and 44% report that they're somewhat satisfied. A mere 6% are somewhat or very dissatisfied. While it would be comforting to view this level of approval as proof that all is well in backup-ville, other responses—like the one indicating that less than half of respondents test restores regularly—bring to mind the old saying that ignorance is bliss.

This general (if possibly misplaced) satisfaction, along perhaps with ongoing limited budgets, is also reflected in the small differences between the technologies our respondents are currently using for their backups and those they plan to employ in the future. While 13% expect to add hardware deduplication, for example, and the same number plan to start using software deduplication, only 6% say they'll stop backing up directly to tape. Some also plan to increase their use of snapshots (5%), and of course, cloud backup (14%).

In this report, we'll explore how the development of new backup—and computing—technologies is causing the backup process to evolve. We'll also examine how the market has adopted these technologies and whether, in the real world, they live up to the hype.



Research Synopsis

Survey Name: *InformationWeek Analytics* 2011 Backup Technologies Survey

Survey Date: January 2011

Region: North America

Number of Respondents: 420

Purpose:

To explore the state of data backups today and interest in emerging technologies

Methodology:

InformationWeek Analytics surveyed business technology decision-makers at North American companies. The survey was conducted online, and respondents were recruited via an e-mail invitation containing an embedded link to the survey. The e-mail invitation was sent to qualified *InformationWeek* subscribers.

ABOUT US | *InformationWeek Analytics*' experienced analysts arm business technology decision-makers with real-world perspective based on a combination of qualitative and quantitative research, business and technology assessment and planning tools, and technology adoption best practices gleaned from experience.

If you'd like to contact us, write to managing director **Art Wittmann** at awittmann@techweb.com, content director **Lorna Garey** at lgarey@techweb.com and research managing editor **Heather Vallis** at hvallis@techweb.com. Find all of our reports at www.analytics.informationweek.com.



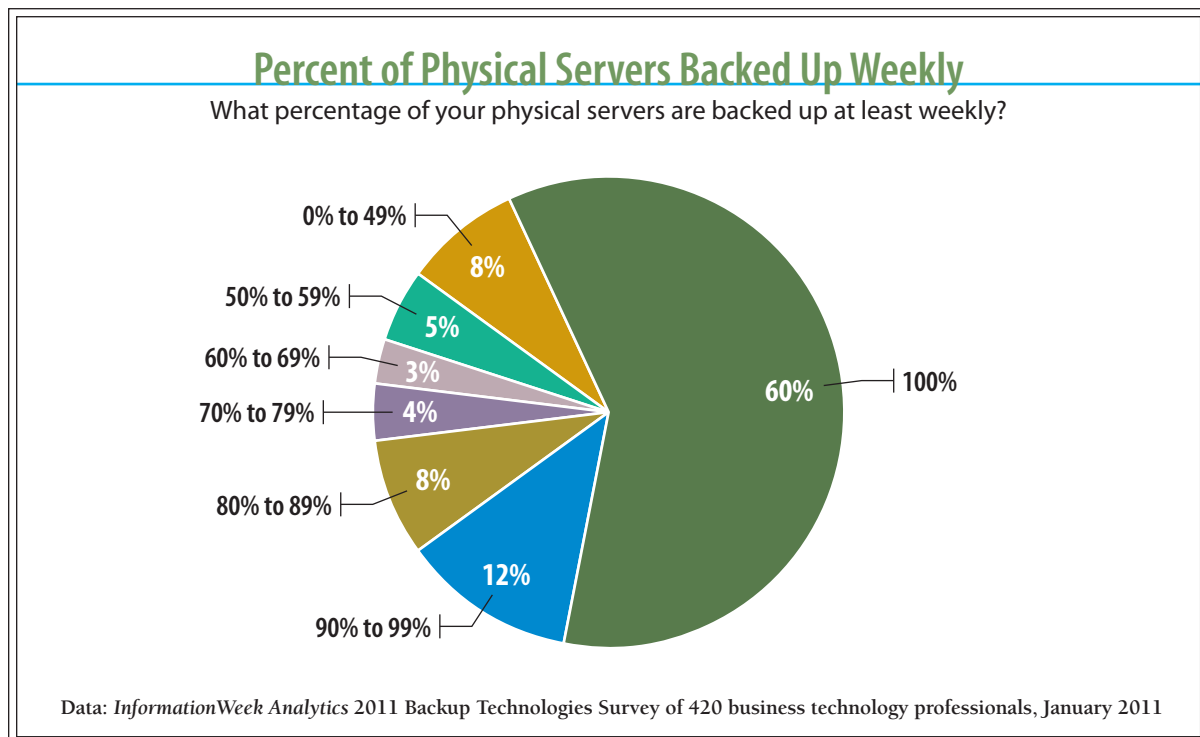
Risk Takers

You'll never hear a CIO downplay the importance of data protection. But our January *InformationWeek Analytics* 2011 Backup Technologies Survey of 420 business technology professionals shows organizations too often fail to put their money where their mouths are.

Respondents may be satisfied with their backup practices, but our survey and experience show companies regularly leave data either underprotected or outright in the wind.

For example, we asked several questions about how frequently servers are backed up. Respondents are pretty conscientious about backing up most of their physical servers at least weekly. They don't seem to be as concerned, however, about their virtual servers. Only half say they back up all their virtual servers every week. Even more alarming, 22% report that they back up fewer than *half* their virtual servers weekly. In some cases, respondents' organizations are using physical servers for mission-critical applications and those containing frequently updated data. But we still had an uncomfortably large number giving reasons like "too diffi-

Figure 1





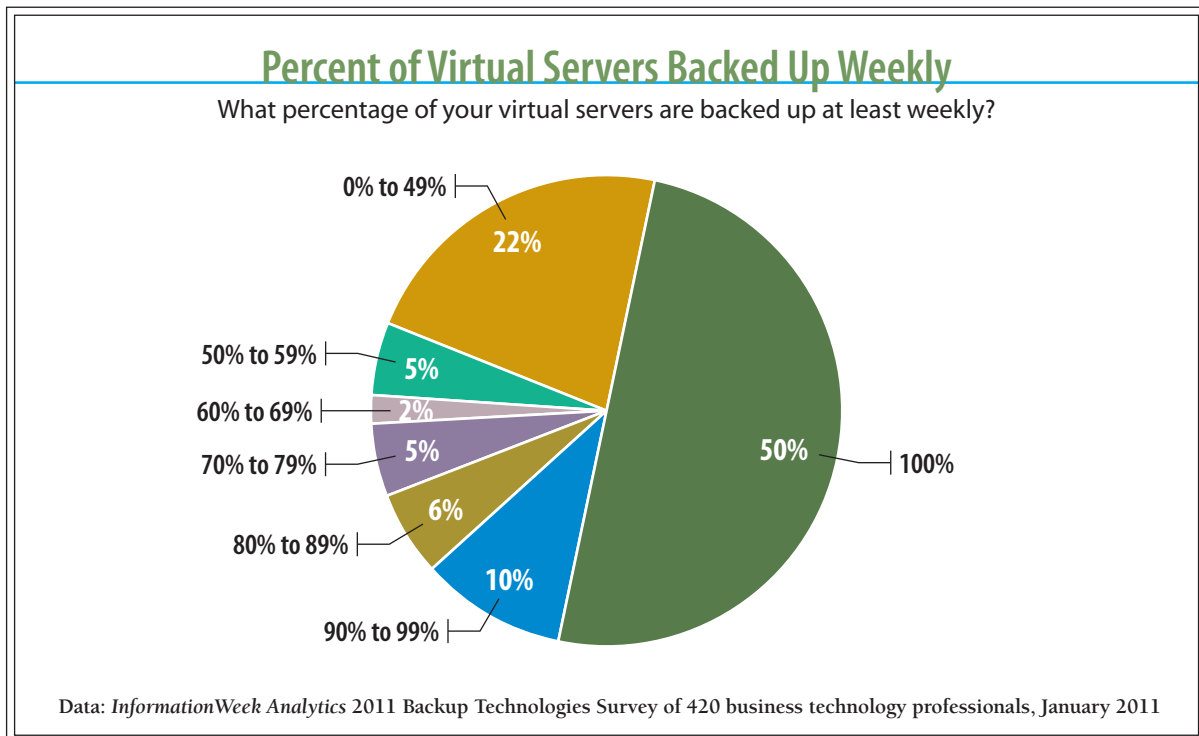
Analytics Reports

cult/not practical” (7%) or saying the server/application owner either never asked for (10%) or refused (6%) backup.

We can say from personal experience that when (not if) one of these systems fails and data is lost, the organization will suffer, and IT will be blamed—even if the application owner had convinced himself there was a good reason to skip backing up just that one server.

Fortunately, a range of new and improved technologies can make backups smooth enough that there will be no reason not to protect all your data, while saving money to boot. We’ll explain how deduplication can help you cost-effectively back up to disk and keep those backups online for fast restores while also getting that valuable data offsite to expand your protection universe, from just guarding against user error and server failures to also covering larger problems, like fires and floods. For example, organizations that don’t have multiple data centers or dedicated disaster recover sites can economically ship deduplicated data to cloud storage providers, getting it offsite faster and more reliably than the old standby of transporting tapes.

Figure 2





Not Dead Yet

Despite ongoing reports of its imminent demise, tape—either direct to or via disk—is still the backbone of most respondents' data protection systems, largely thanks to a combination of low cost and high portability. But using tape as a backup medium has some significant drawbacks. Chief among them is the handling of tapes themselves. Backing up a lot of data to tape means someone has to label, catalog and store hundreds or thousands of cartridges. Even if you have robotic autoloaders that can automate much of this process, library slots cost money, so you might have a week's backups in the library and another several hundred tapes on the shelf and/or at Iron Mountain. When a user calls the helpdesk and asks that a file last seen a month ago be restored, the process of figuring out which tape the file is backed up to, retrieving that tape from storage and mounting the tape requires significant effort. In many large organizations that back up directly to tape, restores can take more than a day—and cause a lot of aggravation for IT.

Many IT teams also have difficulty with tape's performance characteristics. Most backup systems feed a single backup job to a tape drive at a time. If the data source takes several hours to deliver its contents, other jobs must wait for a tape drive to become available. While today's drives can accept data at relatively high rates, up to 280 MBps for LTO-5, they have problems with data sources, like a mailbox-by-mailbox backup of an Exchange server, that can't keep the data streaming continuously to the drive.

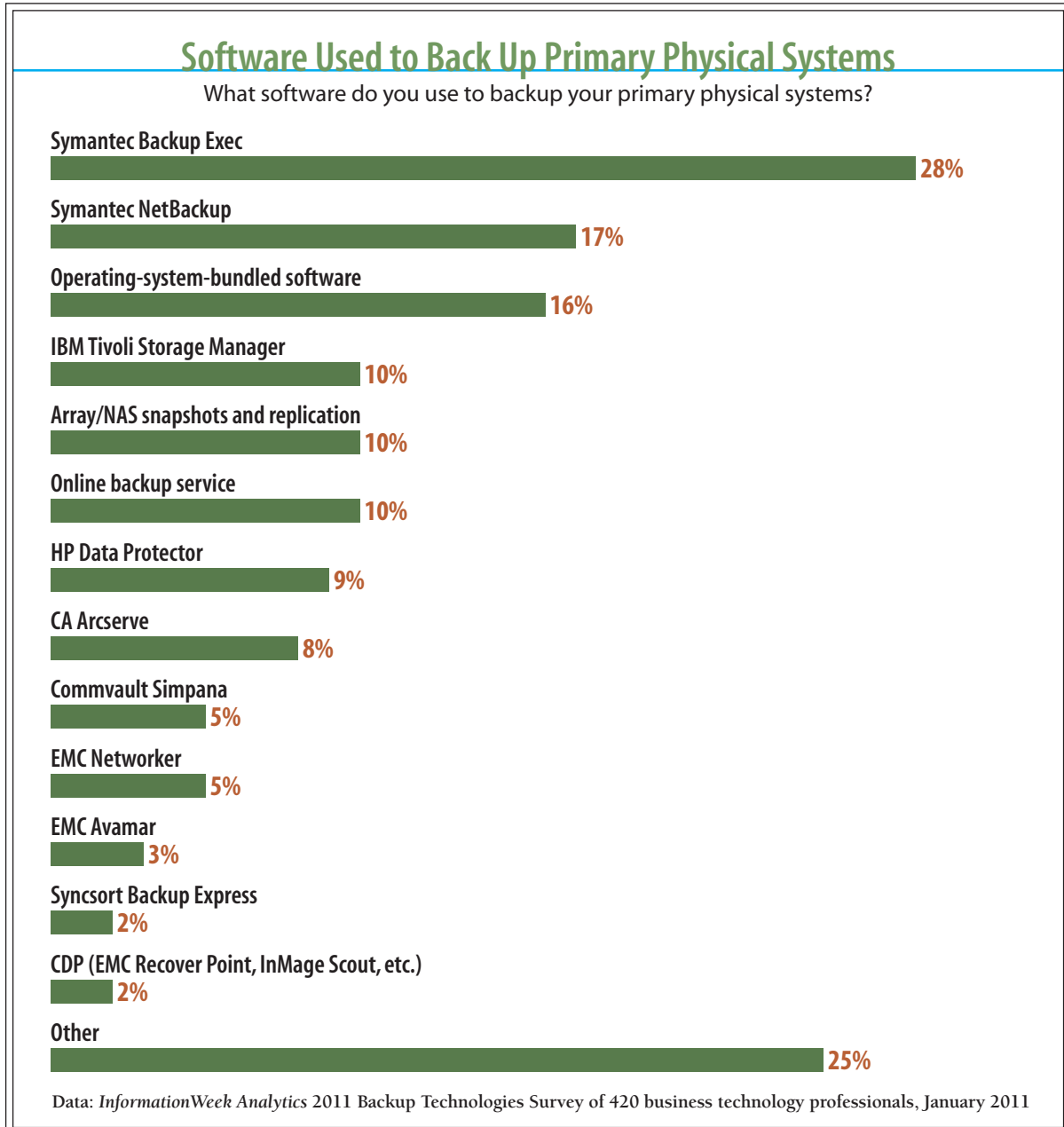
If the data source can't keep a tape drive fed, then every time the drive's buffer is emptied, it has to stop the tape, rewind to the point where it stopped recording and start the recording process again. Since today's drives move tape across the heads at 100 inches per second or faster, this is a time-consuming process since stopping on a dime would snap, or at least stretch, the tape.

As a result, moderately slow backups become very slow backups.

Some enterprise backup applications address these problems by sending the data from multiple backup jobs to a single tape drive interleaved, or multiplexed, together. While multiplexing does keep the tape drive busy, it slows the restore process because the backup app has to skip over data from other servers. Multiplexing can also end up spreading the data for a single backup job across multiple tapes, a problem when restoring even a small number of files.



Figure 3





A n a l y t i c s R e p o r t s

Backup administrators squeezed between the constantly expanding volume of data they're expected to protect and the business' growing demands that applications be available 24/7 may be left without large enough backup windows to accommodate the limitations of tape backup. If this sounds like your world:

- Use incremental forever or synthetic full-backup techniques to reduce the amount of data being backed up over the weekend;
- Convert to backup-to-disk to allow more backup jobs to run in parallel, shrinking the backup window; and
- Consider employing snapshots and CDP to provide faster recoveries and more than one restore point per day.

Disk to the Rescue

While it's still less expensive to store a unit of data on tape than disk, the gap has closed significantly, largely because the price of disk is dropping faster than that of tape. While disk is still more expensive on a per-gigabyte basis, keeping data online on disk simplifies restores and slashes the cost of handling tapes, at least partially offsetting the hit with improved service.

One barrier to the use of SATA disk arrays for backups is that the backup software—and perhaps more importantly, the processes and procedures—used by midmarket and enterprise organizations is often designed around tape. One way around that: virtual tape library (VTL) disk arrays that emulate robotic tape libraries. Even though backup applications now support some level of native backup-to-disk functionality, larger organizations continue to purchase VTLs because they can be easily inserted into existing backup processes and procedures and can connect directly to Fibre Channel SANs.

Most organizations we work with adopt disk systems to fit their backups within an available downtime window, but there are added advantages beyond the performance boost. First, a disk array, VTL or disk-based backup appliance isn't limited to running only as many simultaneous backup jobs as there are drives in the system, the way tape libraries are. Disks don't need to restart and can accept data at any rate up to their limits. The biggest advantage of a backup-to-disk architecture, however, is that since the data is online, backup administrators can satisfy



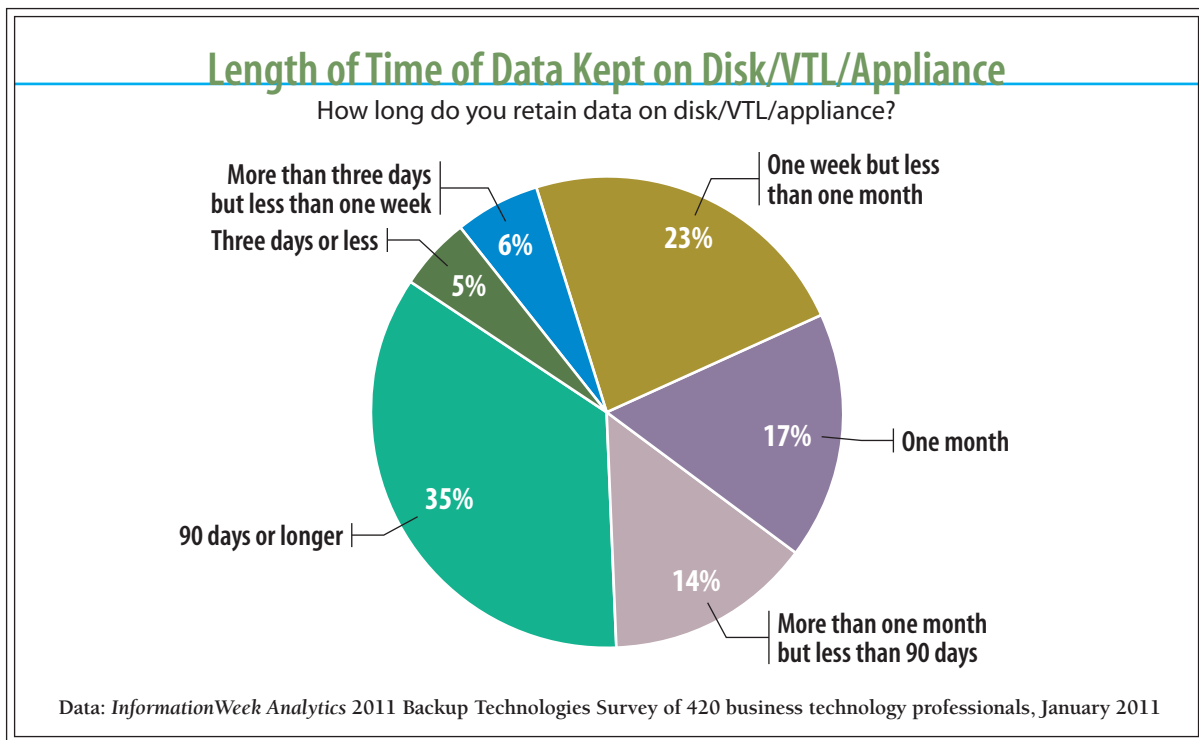
A n a l y t i c s R e p o r t s

most restore requests in a matter of minutes instead of days. Because restores of one or a few files are significantly more common than full system restores among our survey respondents, this capability may be worth the price of admission.

While faster backups and restores make disk systems attractive for day-to-day backups, they're still significantly more expensive than tape, especially when secondary costs like power are considered.

When doing a TCO comparison, consider that a tape drive draws about 20 watts when running, a disk drive about 10, but the disk runs all day and has RAID controllers also drawing power. The tape drive runs only during the backup cycle, and the robot perhaps a total of one hour a day loading and unloading tapes. Since one tape drive in a library usually has 10 or so drive slots behind it, one tape drive may run 20 watts for 10 TB, where 10 TB is 12 to 14 disk drives. Technologies like dedupe and MAID change the equation, but so does the fact that for each tape in the library, there are many more taking up space on the shelf or in offsite storage.

Figure 4



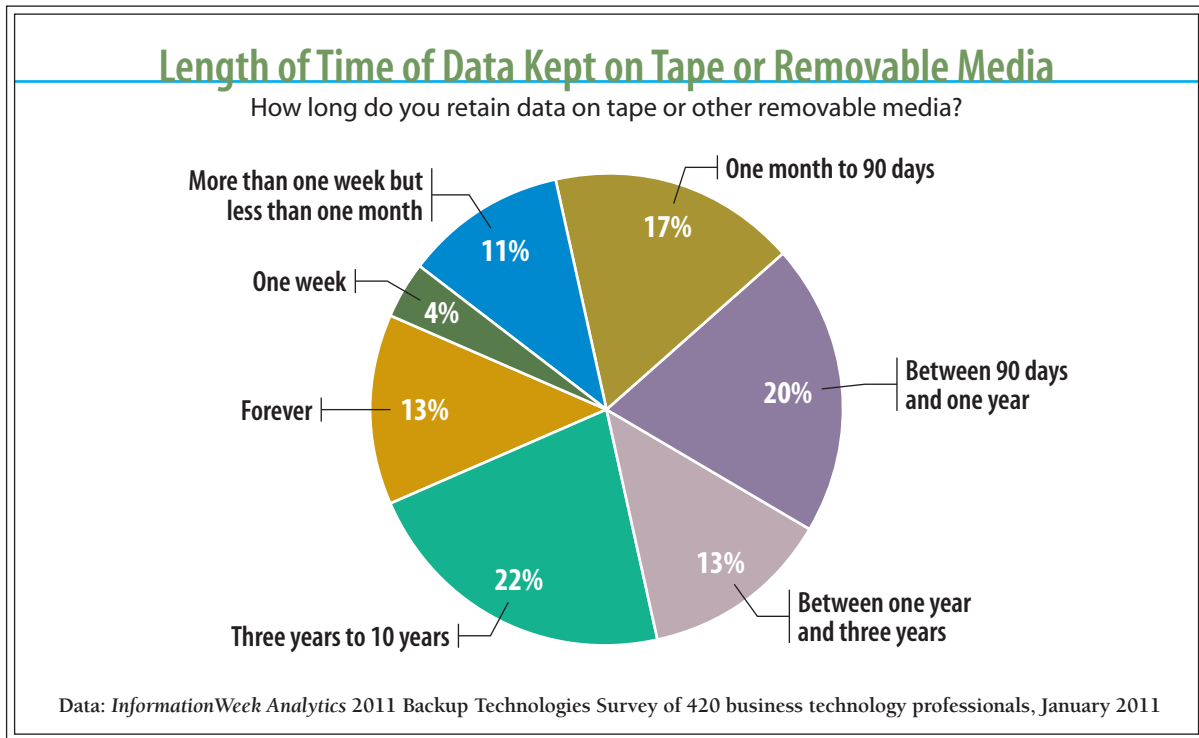


As a result, many organizations use disk as a temporary repository, spooling data off to tape for longer retention. Since most restore requests are for data backed up in the past few days, the number of restore jobs that have to be performed from tape is small. Only 35% of our respondents store data on disk for more than 90 days, while 68% report that they keep data on tape or other removable media for 90 days or more.

One fairly disturbing observation from our survey is that organizations continue to use backup tapes as the basis of their long-term archiving strategy. Thirty-five percent of our respondents report that they keep their backup tapes more than three years, with 13% keeping tapes forever. Given the difficulty of recovering data from five- or 10-year-old tapes, taking into account cataloging the data and changing tape formats, we believe that an old stack of backup tapes does not a proper archive make. While backups and archive systems both rely on copying data from the primary data store, they're designed to solve different problems.

- **Backup systems** make copies of data to restore it back to the place, and purpose, it was in

Figure 5





A n a l y t i c s R e p o r t s

before some event damaged it. To do this, the only metadata backup systems need is that indicating when the data was backed up and where it came from. Some image backup systems take the minimalist approach to metadata to the extreme, cataloging just the system and volume backed up without any information about the volume's contents.

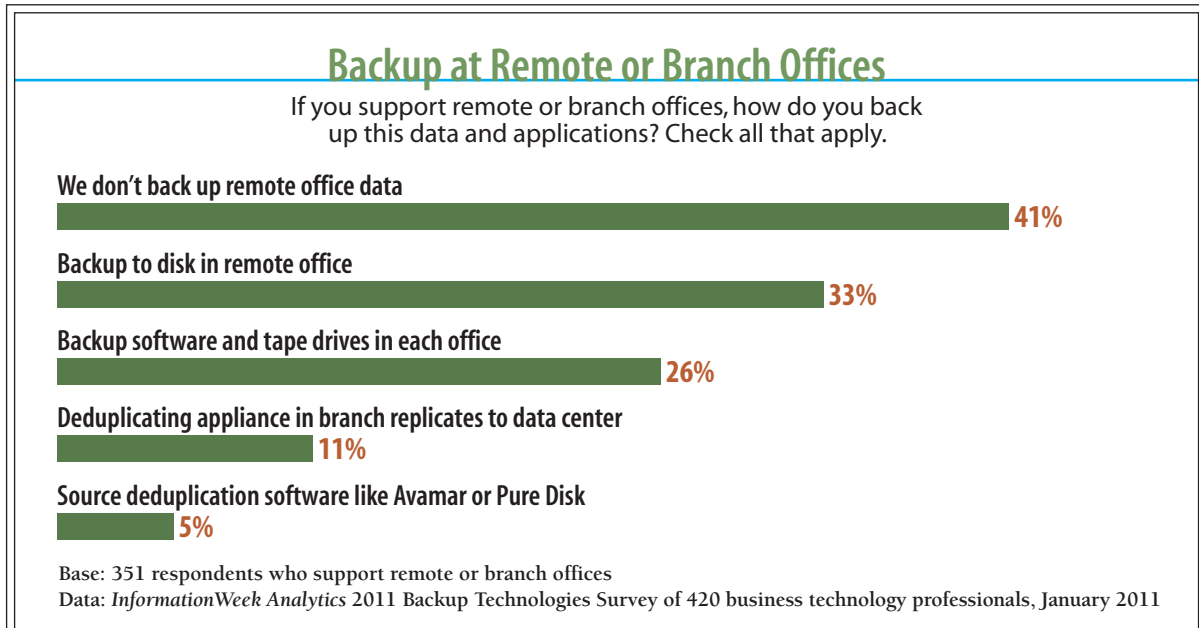
- **Archive systems**, on the other hand are designed to store data so it can be retrieved later, typically for a purpose other than just restoring it to its previous use—say, as evidence in a lawsuit. As a result, archive systems organize data not by where it came from but based on its contents, usually including full text indexes to support functions like e-discovery, where requests come for information related to Smith Inc., not files from server 3453.

Bottom line, you need separate plans for backups vs. long-term archiving.

Out of Sight, Out of Scope?

Another glaring problem: While 351 respondents say they support remote offices, a shocking 41% don't back up remote office data—leaving their organizations exposed to unacceptable levels of data loss. Just 16% deduplicate and replicate their data to the home office.

Figure 6





A n a l y t i c s R e p o r t s

Now, traditional tape backup systems have long delivered unsatisfactory results in branch offices. These sites are frequently lacking anyone with the technical expertise to change tapes and check backup status. As a result, tapes don't get changed or sent offsite as scheduled, and backup failures aren't addressed in a timely fashion. But given today's technologies, there's no excuse for not protecting the data in remote sites, including home offices, and for mobile workers. Look into compact deduplicating backup appliances or source deduplication software that replicates backups to the main data center, or consider sending offsite data to an online backup or cloud storage provider. Both strategies improve backup reliability and reduce the amount of time IT staffs spend supporting remote backups. We'll talk more about these technologies.

Cut Out the Copies

Of all the data protection technologies developed over the past 10 years, none has generated more buzz than data deduplication. Interest peaked with the summer 2009 bidding war between NetApp and EMC for deduplication pioneer and market leader Data Domain, which ended with EMC paying over \$2 billion for the company, but the technology is still garnering a lot of ink.

Data deduplication systems identify the repeats of data in a backup set—whether Windows .DLL files and a company logo on multiple reports or contents of a database backed up daily and retained for a month—and replaces the duplicates with pointers to a single stored copy. Identifying duplicate files is a relatively simple process, but data deduplication technology goes significantly further, identifying not just duplicate files but also smaller chunks of duplicate data stored repeatedly in multiple files. Backup data sets are especially rich in duplicate data as they usually include multiple copies of the same files backed up at different times. An organization making weekly full backups of its servers with daily incremental backups will, over the course of a month, generate at least four full copies of its data. Deduplication therefore enables organizations to retain their backup data on disk longer using the same amount of space—and as importantly, while consuming the same power and data center space.

While the deduplication process varies from vendor to vendor, and even among multiple products from the same vendor, most deduplication systems use basically the same technique.

First, the system breaks data into smaller blocks, typically somewhere between 4 KB and 64 KB. Each vendor has its own secret sauce for determining whether to use a fixed block size or



Analytics Reports

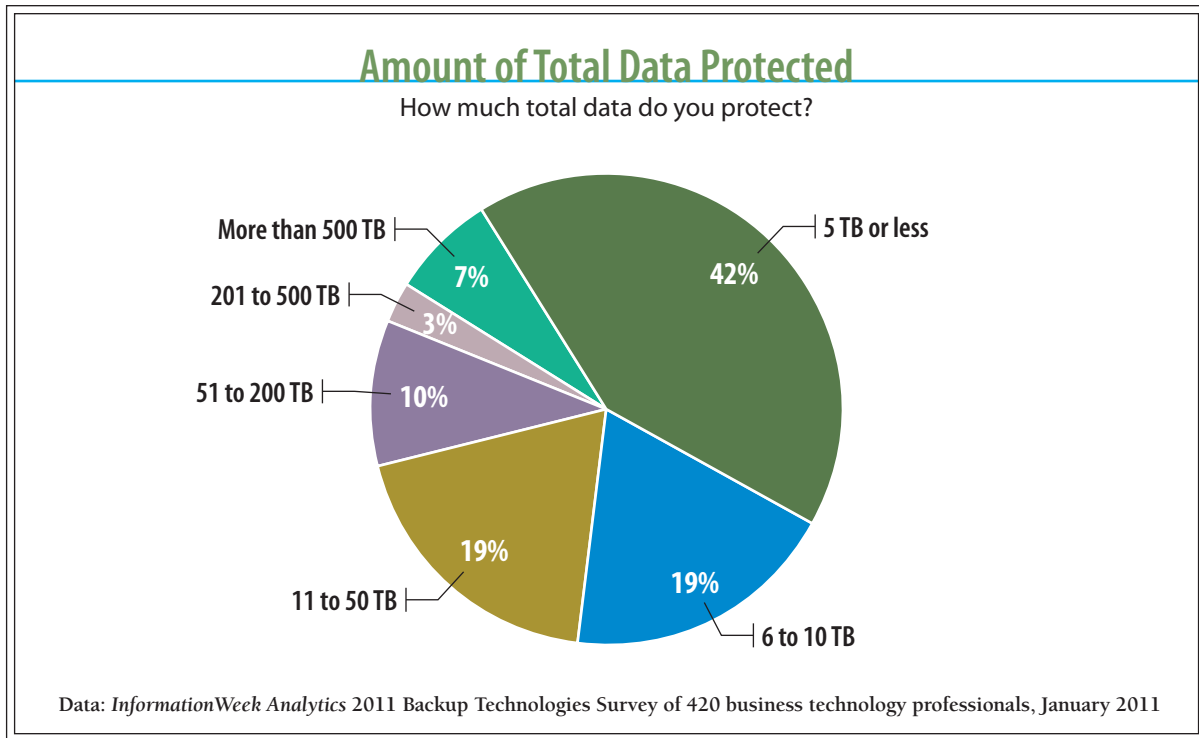
adjust the block size to match the data, and they all, of course, claim their techniques get the best data reduction with the least performance impact.

The system then determines if the new block contains data that's the same as a block it's already saved. It does this by running the data through a hash function, like MD-5 or SHA-1, that creates a much smaller hash value, 160 bits (20 bytes) for SHA-1. Blocks containing different data will generate different hash values, so the system has to store the data for a given hash value only once; it then uses pointers to keep track of where the data goes.

While the odds against a hash collision are astronomical, some vendors take the extra step of calculating a hash using a second independent hash function, or performing a byte-by-byte comparison of the data in a block, before declaring the data to be the same.

Deduplicating backup appliances, like VTLs before them, can be inserted into an existing backup architecture relatively easily and can provide high performance, up to 5 TB per hour or more

Figure 7





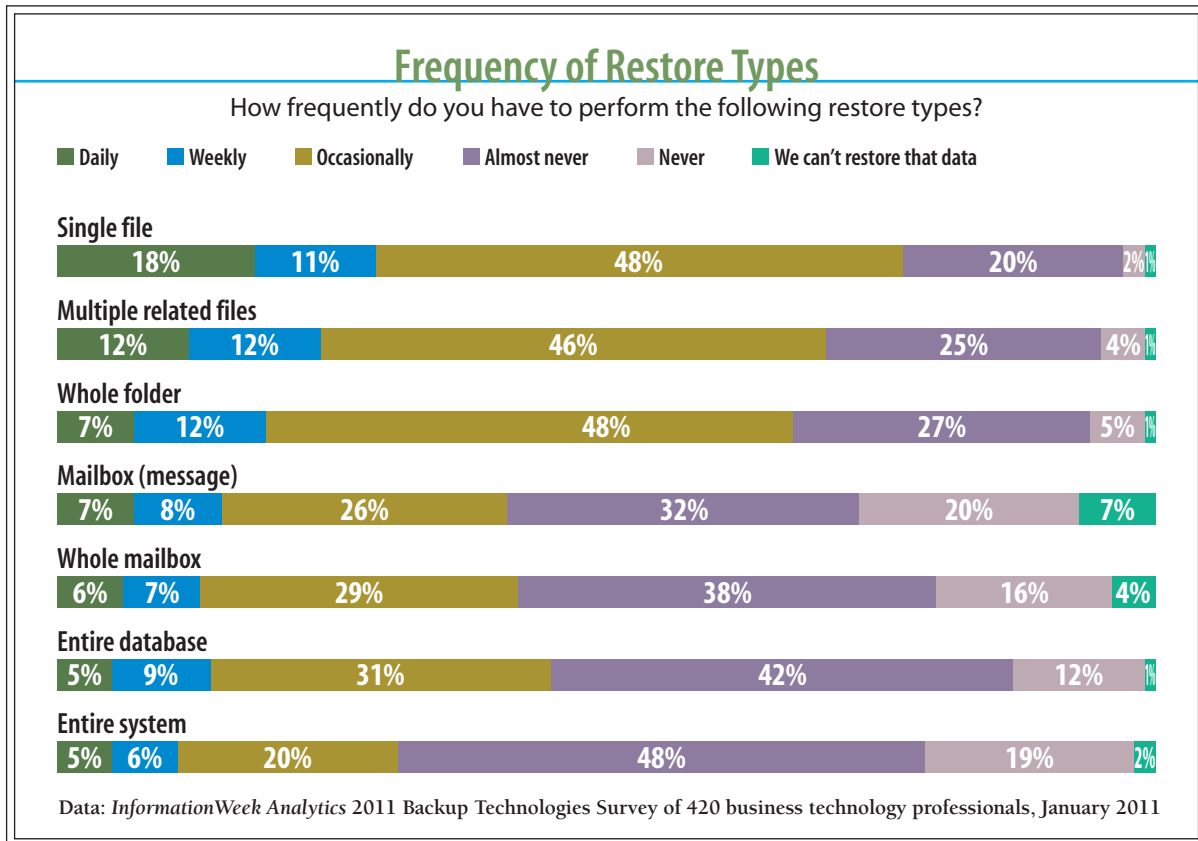
A n a l y t i c s R e p o r t s

for some models. The downside to deduplicating appliances is that they can be expensive, with some high-end models starting at over \$250,000. While dedupe as a feature is working its way into file and unified (file and block) systems, it hasn't made it to actual arrays yet. Moreover, the dedupe methods used for primary storage don't work well for backup data from a backup application. Tuning the deduplication process for backups can significantly boost data reduction.

Software Deduplication

Even though data dedupe is often delivered in the form of backup appliances, the actual deduplication is performed by software embedded in these appliances. Today, data deduplication is also a feature of many backup applications. The first form of software-based deduplication to reach the market deduplicated data at the source server, rather than at the storage system.

Figure 8





A n a l y t i c s R e p o r t s

These products, like EMC's Avamar and Symantec's PureDisk, use a relatively heavyweight agent in each system to be backed up. This agent, like a conventional backup application, scans the system for files that have changed since the last backup and then deduplicates that data before sending it to the server that acts as the repository.

Deduplicating at the source trades memory and CPU cycles in the source server for bandwidth on the connection between the source server and the data repository. This makes it a good choice for bandwidth-limited environments, like backing up remote offices or virtual servers. Inside the data center, however, administrators tend to be wary of agents, and bandwidth is more available, so most backup applications now can also deduplicate data at the backup media server before storing it to disk. Data deduplication is a compute-intensive function, so media server deduplication performance is limited when compared with the performance of backup appliances.

Deduplication for Replication

As much as data deduplication has improved the backup process in the data center, its biggest impact is in how it enables organizations to send data offsite over a WAN link instead of via tapes and a courier service. Without data reduction, most of us couldn't afford sufficient WAN bandwidth to replicate even daily changes. The data reduction rates provided by the combination of deduplication and compression (which reduces the size of data by removing small repetitions) can let IT squeeze remote office data into a much more affordable WAN connection.

Organizations that couldn't justify the WAN bandwidth replication required without data reduction can now get their backup data offsite in close to real time, and without the hassles of shipping tapes. Multiple remote sites can replicate to a single headquarters system, and organizations can create multilevel cascaded replication schemes, with branches replicating to regional offices that then replicate to HQ.

When multiple deduplicating systems replicate their data to a central repository—for example, branch or remote offices replicating to a single corporate data center—IT can achieve even greater bandwidth and disk savings through a feature known as multisite data deduplication.

With multisite deduplication, a source system sends the hash values of the blocks it has stored since the last replication. The destination system checks those hashes against the set of data



A n a l y t i c s R e p o r t s

blocks it has already stored. It then tells the source system to send the data corresponding to the hash values for which it doesn't already have data. The source system sends only the blocks it is storing that haven't been backed up across the organization and, of course, the metadata describing the new backup data.

Sounds good, and in fact, half of our respondents say they use data deduplication to some extent, although only 25% use hardware or software deduplication extensively. Our survey results also point out one area of confusion: Deduplication ratios are difficult to predict. Results are highly dependent on the composition of your data, and "deduplication" is a blanket term that covers different implementations of similar technologies. The intersection of deduplication, backup software and data all affect the reduction achieved. Even though deduplication vendors claim you could see deduplication ratios of 5:1 to 20:1, 58% of our respondents report getting less than 5:1 data reduction, with 29% reporting 2:1 or less. Yes, every bit helps, but test on your data before buying.

Figure 9

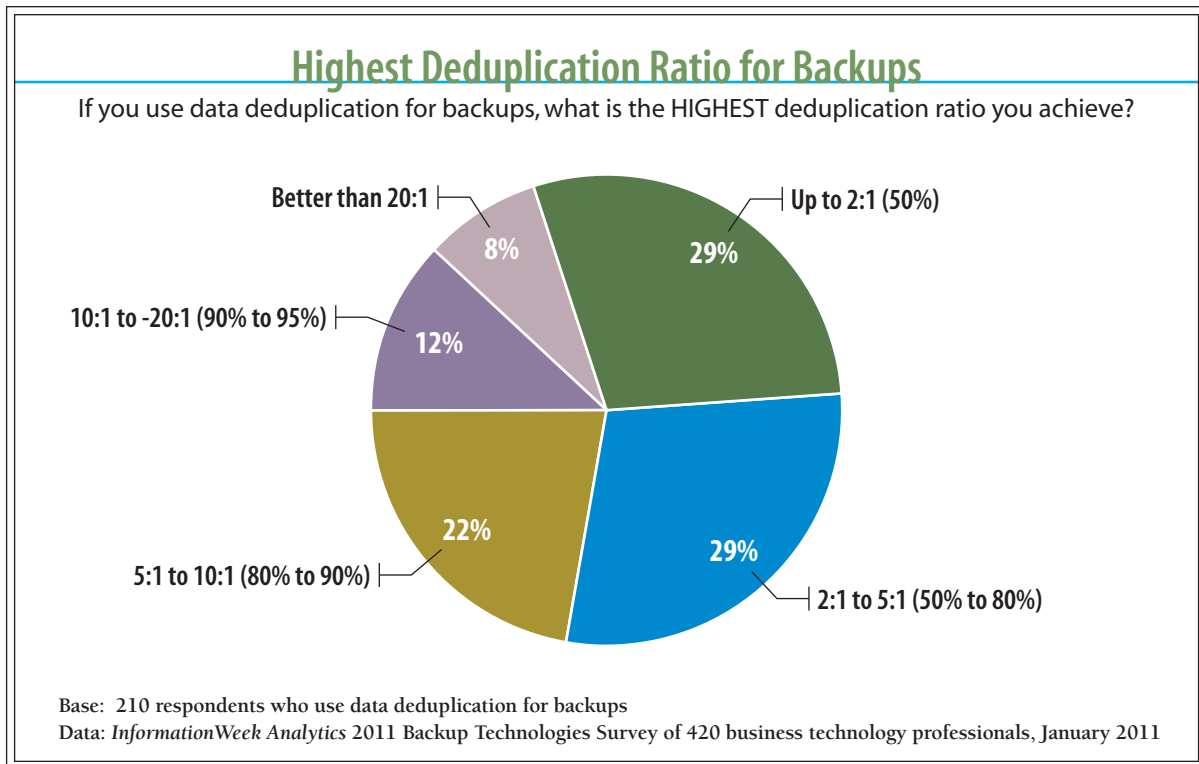
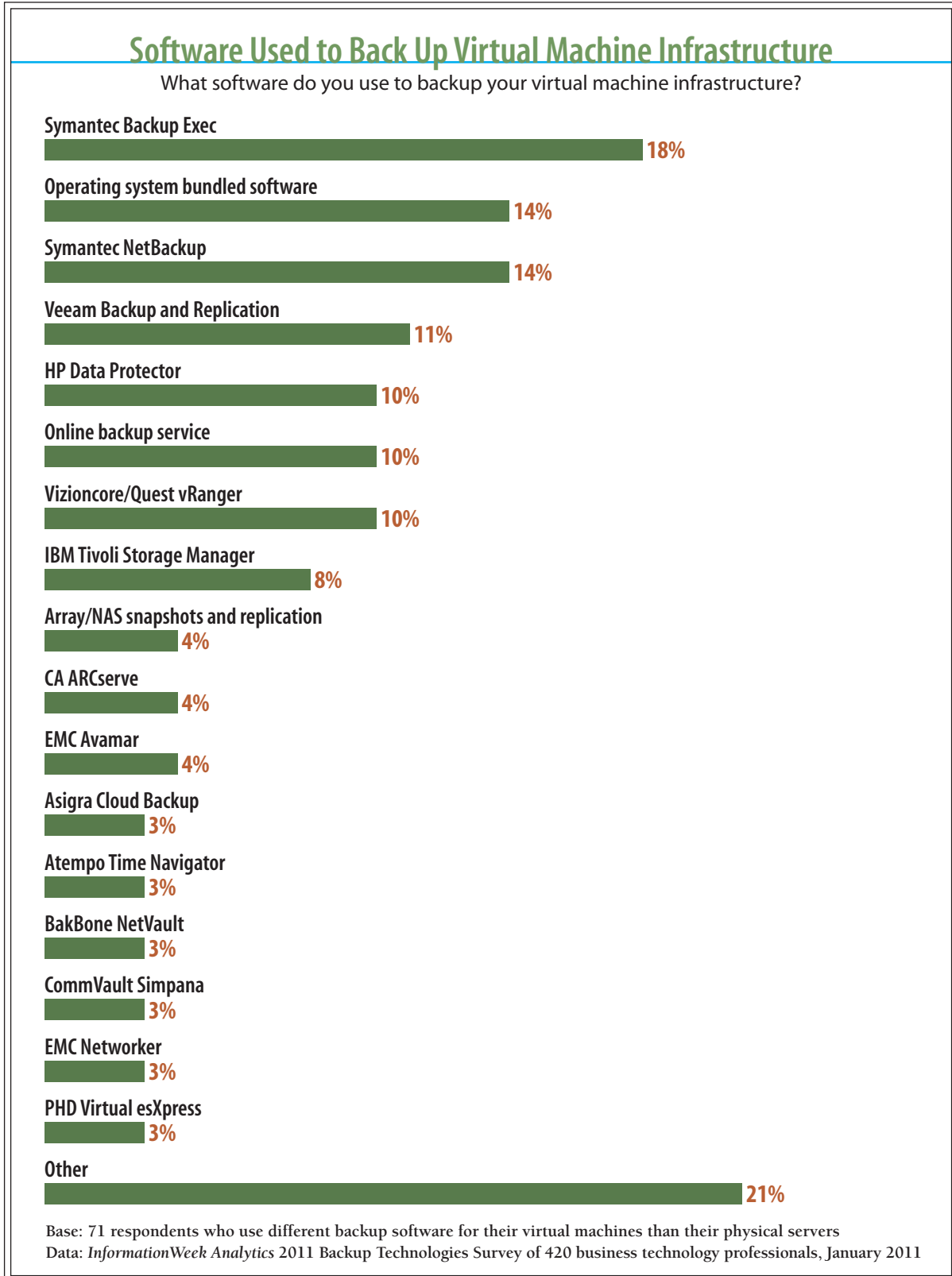




Figure 10





Freeze Frame

Snapshots are most frequently created in a storage system such as a RAID array or NAS appliance, although some operating systems and logical volume managers also have the capability to create and manage snapshots. A snapshot provides a picture of a data volume or file system at a specific point in time. Once a snapshot is created, a system administrator can mount it as a read-only volume to restore a file or to use as a source for another backup process. If a full system restore is needed, IT can revert the volume to the state it was in when a snapshot was taken. With some systems, snapshots can also be mounted as read-write volumes.

Unlike backups, snapshots aren't necessarily independent copies of a data store. Rather, most snapshot implementations are based the copy-on-write process. A system that's creating copy-, or redirect-, on-write snapshots preserves the contents of disk blocks as they're overwritten in a journal. When a snapshot is taken, the system freezes the journaled data that's changed since the last snapshot and starts saving changed blocks separately for the next snapshot.

Each copy-on-write snapshot takes up only as much space as the blocks that have changed since the last snapshot—generally much less than the total size of the volume. The smaller size of copy-on-write snapshots enables administrators to take snapshots frequently and keep more

What Is a Backup, Anyway?

As data protection technologies have evolved from copies of data on tape to the spectrum of products available today, the very definition of what constitutes a backup has come into question. We find it useful to consider the existential purpose behind backups when determining if a given product satisfies that purpose.

The purpose of backups is, of course, to provide a mechanism by which an organization can recover its systems from a failure by accessing a copy of the data that's stored outside the affected area of the failure. Therefore, the key attribute of a backup is that it is an independent copy of the data stored outside the primary system.

While high-availability systems and RAID protect data

against disk and hardware failures, they leave it exposed to software errors that will corrupt data and humans who may delete or corrupt data accidentally or maliciously. And don't forget Mother Nature. Snapshots can protect data against deletion and corruption but are still contained within the primary storage system and are dependent on the primary copy of the data, so they don't provide protection against a total failure of the storage system.

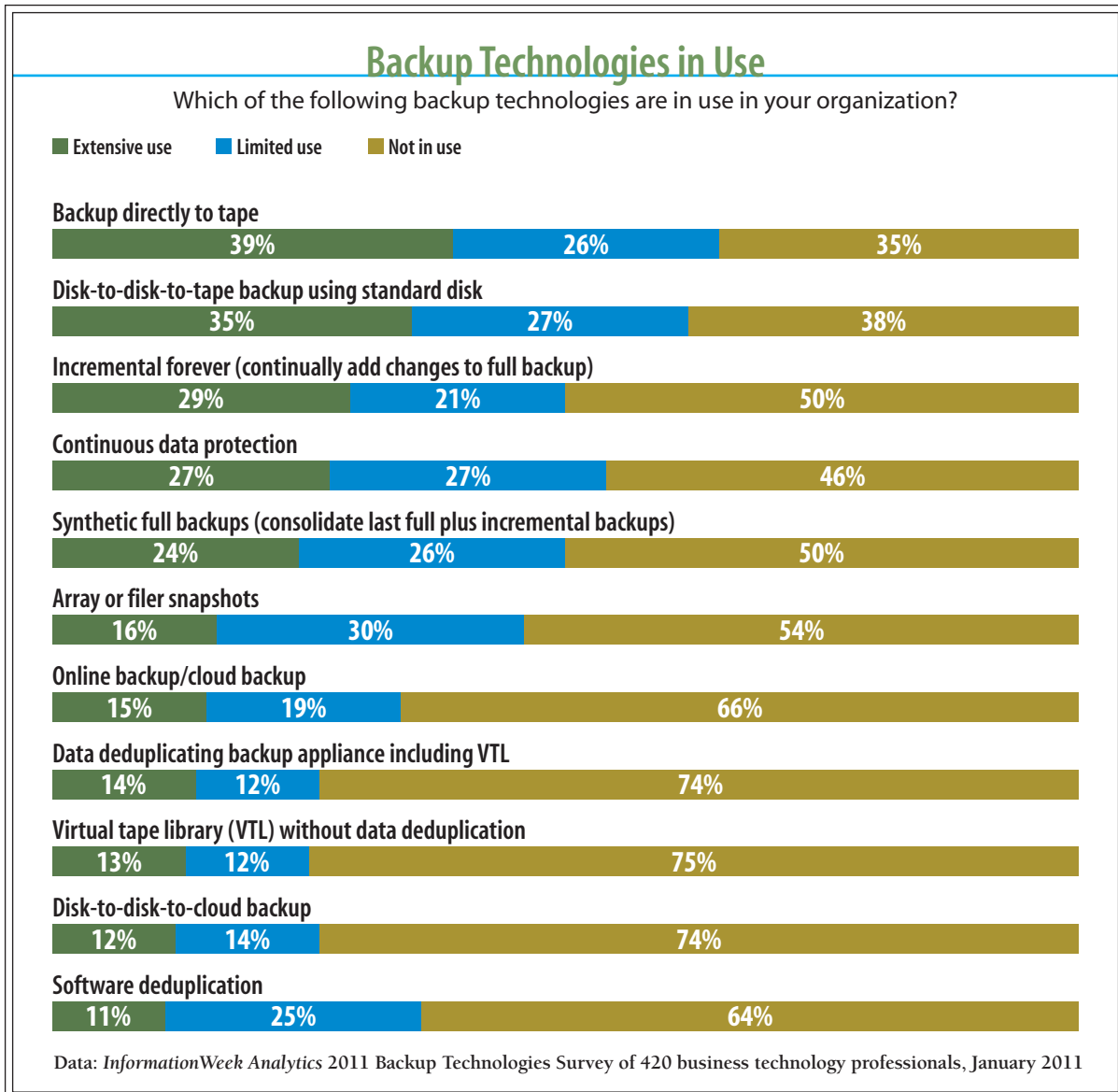
While a fully realized backup scheme should ensure that the protected systems can be recovered in the event of a larger failure via moving backup copies offsite, we would consider any set of independent copies that can recover systems from failures to be backups.



A n a l y t i c s R e p o r t s

snapshots online than they could full backups. Snapshots also serve an important function to backup applications in that they allow the application to freeze changes to a data set while it's being backed up. For databases and other demanding data sets, the backup application can ask the database engine to flush its buffers and quiese, or freeze, itself; trigger a snapshot; release

Figure 11





A n a l y t i c s R e p o r t s

the database engine; and back up the temporary snapshot. Microsoft's VSS (Volume Shadowcopy Service) acts as an interface among databases, snapshot providers and backup applications to simplify this process, which can require custom scripts on other platforms.

Since snapshots are dependent on the storage system, you can't rely on primary data snapshots alone to ensure complete backups. However, when snapshots are combined with a replication system to copy this data to a secondary storage system, preferably offsite, snapshots can be an effective backup method—albeit for those needing limited retention, since most storage systems can maintain only a small number of snapshots of any given volume.

Ten percent of our respondents use snapshots and replication to protect their servers, as opposed to conventional data copy backups. Now, we see the attraction; snapshots are efficient and provide a very fast recovery time. But anyone using snapshots and replication as a primary backup method should also carefully consider their retention requirements and implement an archiving strategy. Yes, snapshots are a great way to satisfy the most common restore requests quickly, as the snapshot can be mounted as a read-only volume and the data restored in just a few minutes. But be aware that most systems support only a limited number of snapshots and may slow down significantly when multiple snapshots are stored on them. Even more significantly, since snapshots are on the same storage system as the primary copy of the data, they provide no protection against storage system failure.

Continuous Data Protection

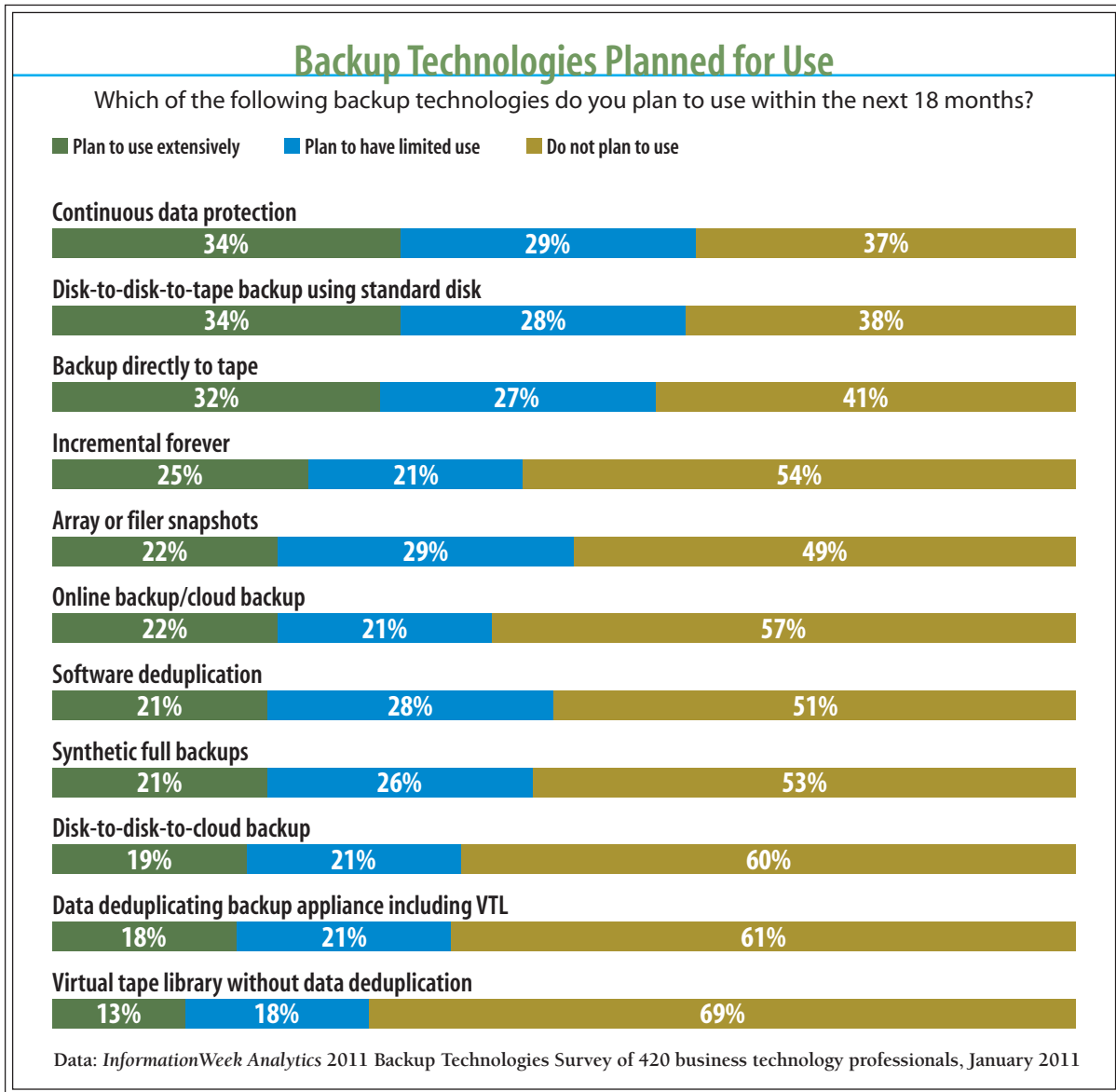
A few years ago, the industry was all abuzz about continuous data protection, with startups like Revivio and Mendacino attracting a lot of attention. Over the years we've realized that CDP isn't the right answer for most applications, and just a few CDP systems remain, like InMage's Scout and EMC's RecoverPoint. The problem, in part, is that the organizations that most needed the level of protection CDP promised weren't comfortable buying the technology from startups, and they wanted integration of CDP systems with their critical applications and conventional backup in a single management console.

Where traditional backup systems provide a single restore point a day and snapshots could provide restore points as granular as every 15 minutes, CDP systems allow data to be restored to any point in time. As with replication, CDP tools duplicate write requests made to the pri-



mary storage device to another location, in the order the requests are made. Rather than applying the changes as quickly as possible to a secondary storage location, enabling IT to recover the source server immediately, CDP systems journal disk writes to another location. This allows the system administrator to restore the system to any point in time by selecting how much of

Figure 12





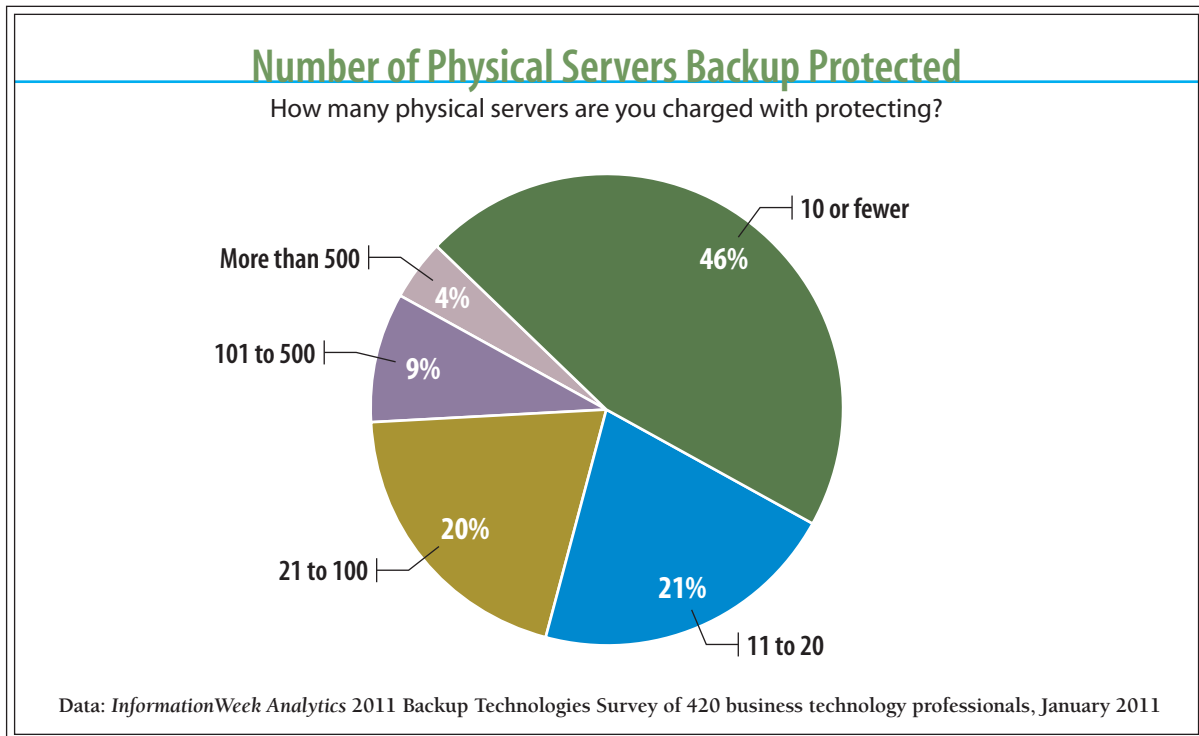
A n a l y t i c s R e p o r t s

the journal file to apply to the data. Theoretically, CDP would allow a database administrator to restore a SQL server or Oracle database to the point just before it was corrupted. In reality, however, having infinite choices of restore points just makes choosing the right restore point more difficult. Better CDP systems annotate the journal with application events, like database checkpoints, to aid an administrator's selection.

While CDP adds complexity to the primary systems, requiring filter drivers or write splitters to send the new data to the CDP system, critical applications can certainly benefit from the ability to recover to a point immediately before the system failed.

Given the size of the organizations most of our respondents work for, we were surprised that 27% say they make extensive use of CDP technology to protect their physical servers. Generally, we would expect CDP to be used predominantly in larger organizations and suspect many of these folks are actually using "real time" file backups that copy files as they're saved, rather than true CDP.

Figure 13



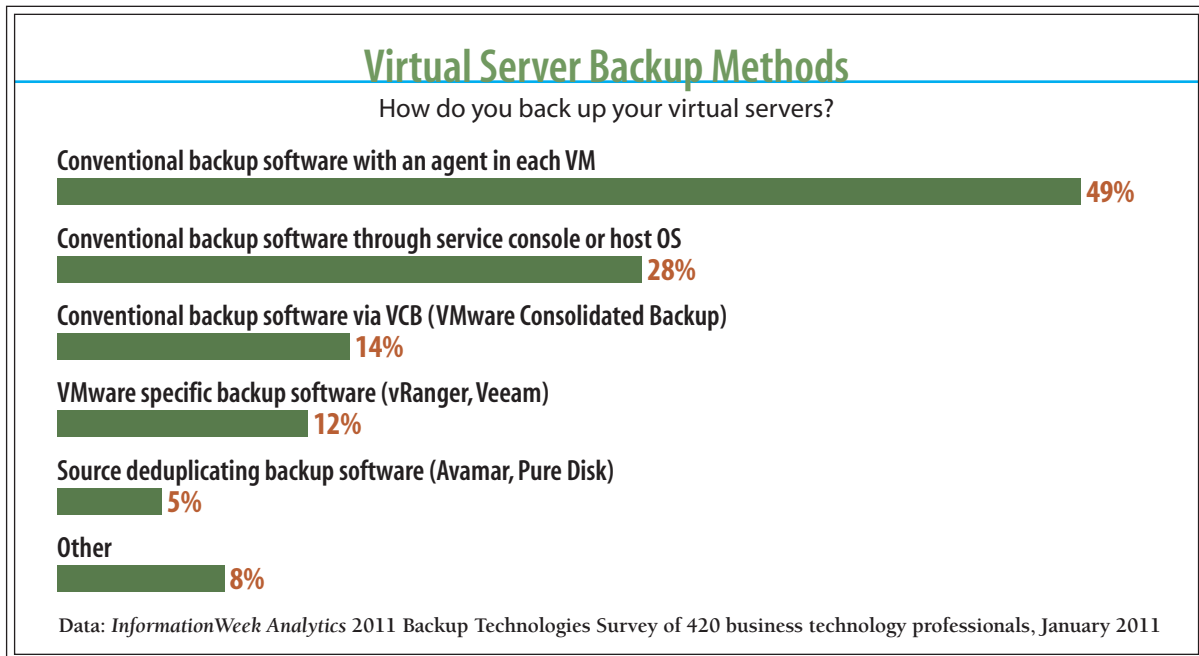


Things Get Cloudy

While online backup services have been available since the mid '90s, they've remained a niche offering due to limited application support, high monthly charges based on the amount of data being backed up and concerns about restore times across the public Internet. In addition, most system administrators we work with prefer to be in control of their backup environments.

Still, there is growing recognition that public cloud offerings can be a good choice for some use cases. The addition of data deduplication in applications like Asigra Cloud Backup will help reduce data volumes, and therefore costs, and as organizations go virtual and mobile, we believe it will increasingly be more practical to have remote employees back up to a common service than for central IT try to support the process. Over the past two years, vendors have extended backup applications like CommVault's Simpana Backup and Symantec's Backup Exec to use public cloud storage systems like Amazon's S3 or Nirvanix as just another backup medium. We find the combination of mainstream backup software and cloud storage compelling: Users get the breadth of application support and local backups for quick restores that mature applications provide plus an affordable way to get their data offsite in close to real time, while encrypting it for security.

Figure 14



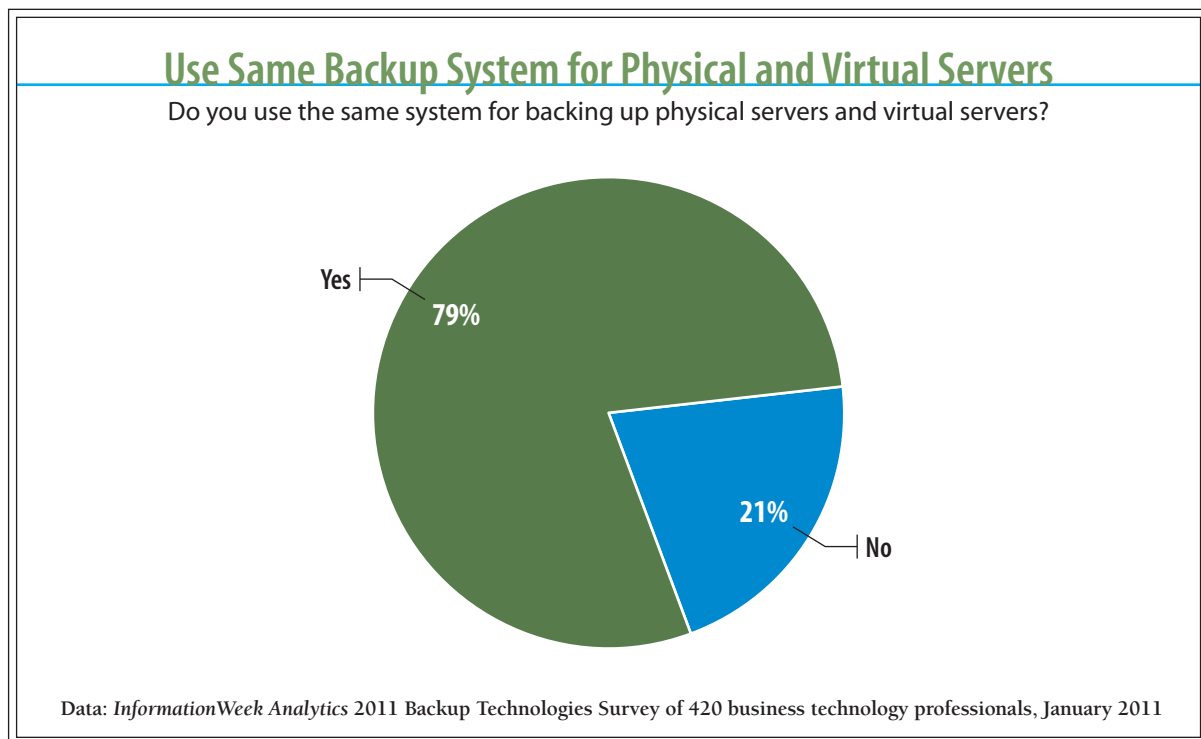


Some of our respondents also find disk-to-disk-to-cloud backup attractive, with 26% reporting they are currently using some form of online or cloud backup and 40% planning to use it in the future. Organizations satisfied with their current backup systems can further improve their data protection by signing up with a cloud storage provider that supports their software and copying backup data to the cloud. Those looking for a complete system as opposed to supplemental protection should select a provider that supports a local cache or copy of their servers' backup data on an appliance, as well as in the cloud, to speed restores, especially of databases.

Virtually Safe

Server virtualization is a clear win for the data center, but it presents a mixed bag of challenges and opportunities for backup administrators and vendors. In our survey, 49% of respondents say they treat virtual servers the same way they treat their physical servers—installing an agent for their favorite backup application in each virtual machine.

Figure 15



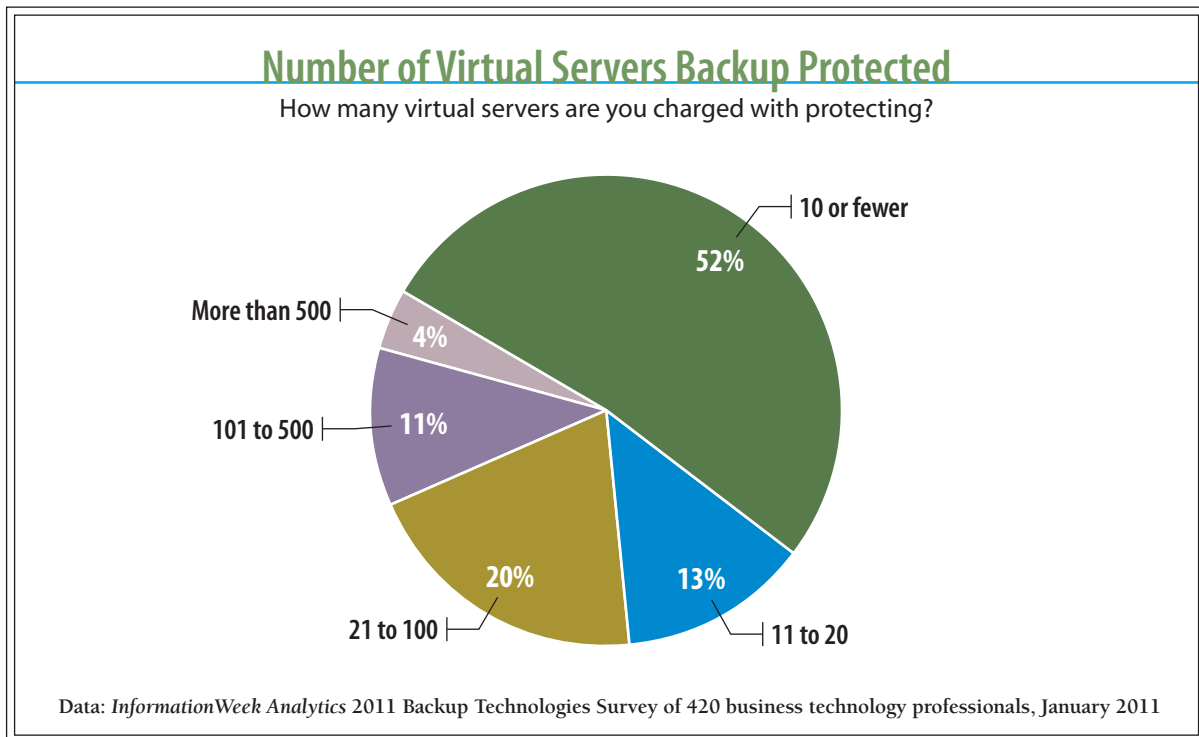


A n a l y t i c s R e p o r t s

Now, treating VMs like physical machines may yield a straightforward backup architecture and ensure application support, but the cost for that simplicity could be significant performance problems. For example, multiple virtual machines on the same virtual server host will share network and storage interfaces. When a media server requests full backups from several VMs on the same host, all the agents on those hosts will concurrently start pulling the requested data up from the storage system and shooting it out across the WAN. This will saturate the LAN and SAN connections for the virtual server host, not only reducing backup performance but also choking other VMs on the host that may be serving users, depriving them of storage and network resources.

In addition, this proliferation of agents consumes memory in each VM, and when the backup vendor comes out with an updated agent or patch, pity the poor administrator. Patching agents in powered-down VMs and VM templates can be a complex process, as these machines have to be spun up, patched and spun back down.

Figure 16

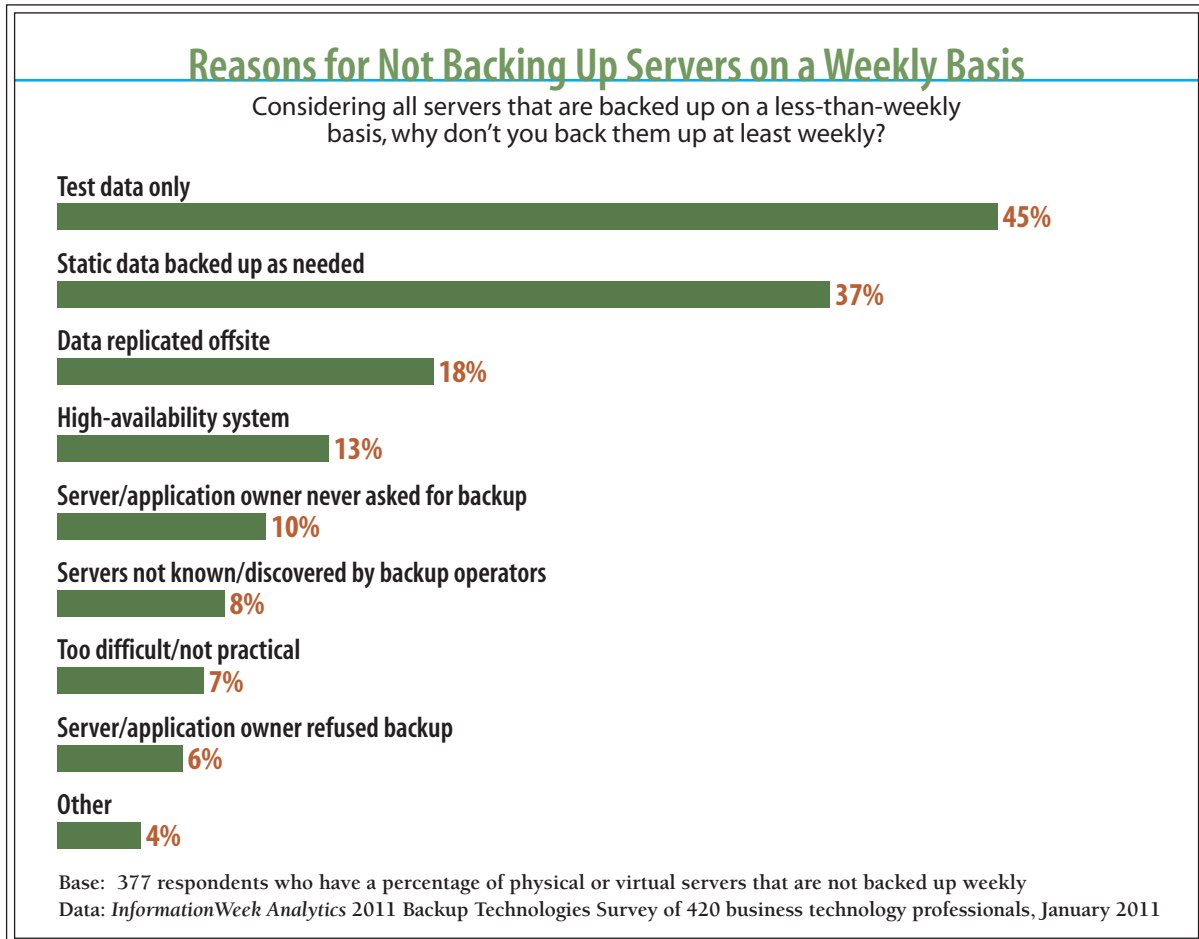




Hypervisors including Citrix Xen, Microsoft Hyper-V and VMware ESX have host operating systems or service console partitions that enable some administrators—and 28% of our respondents—to install their backup application agents on the host and back up their virtual servers as .VHD or .VMDK files. While this method does reduce the number of agents that have to be installed and maintained, it still requires that the administrator either shut down the virtual machines to back them up or back up the virtual machine files as open files, creating “crash consistent” backups.

At No. 4 on the hit list of IT euphemisms, “crash consistent” actually means “only as consistent

Figure 17





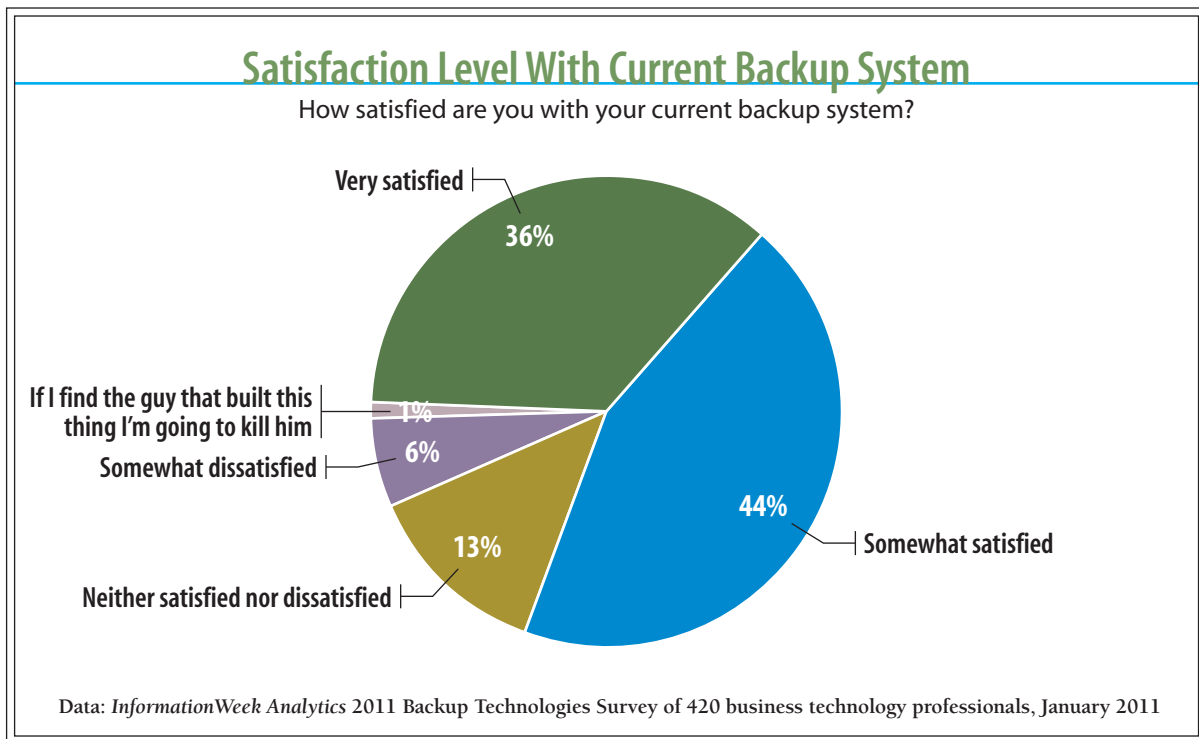
as the data on disk would be if the server crashed.” You will lose any data cached in memory by the application and/or volume manager.

A Better Way?

VMware’s first attempt to improve this process, VMware Consolidated Backup (VCB), was a rather clunky piece of software that took a VMware snapshot of the VM and provided access to the snapshot through a Windows proxy server. VCB did supply a mechanism for third-party backup applications to make disk-image backups of VMs and file-by-file backups of Window VMs, but this was still a slow and cumbersome procedure.

With vSphere 4, VMware has replaced VCB with a new vStorage backup API that eliminates the need for the Windows proxy server while still providing a mechanism for backing up applications. It can perform image-level or file-by-file backups of VMs by accessing the shared stor-

Figure 18



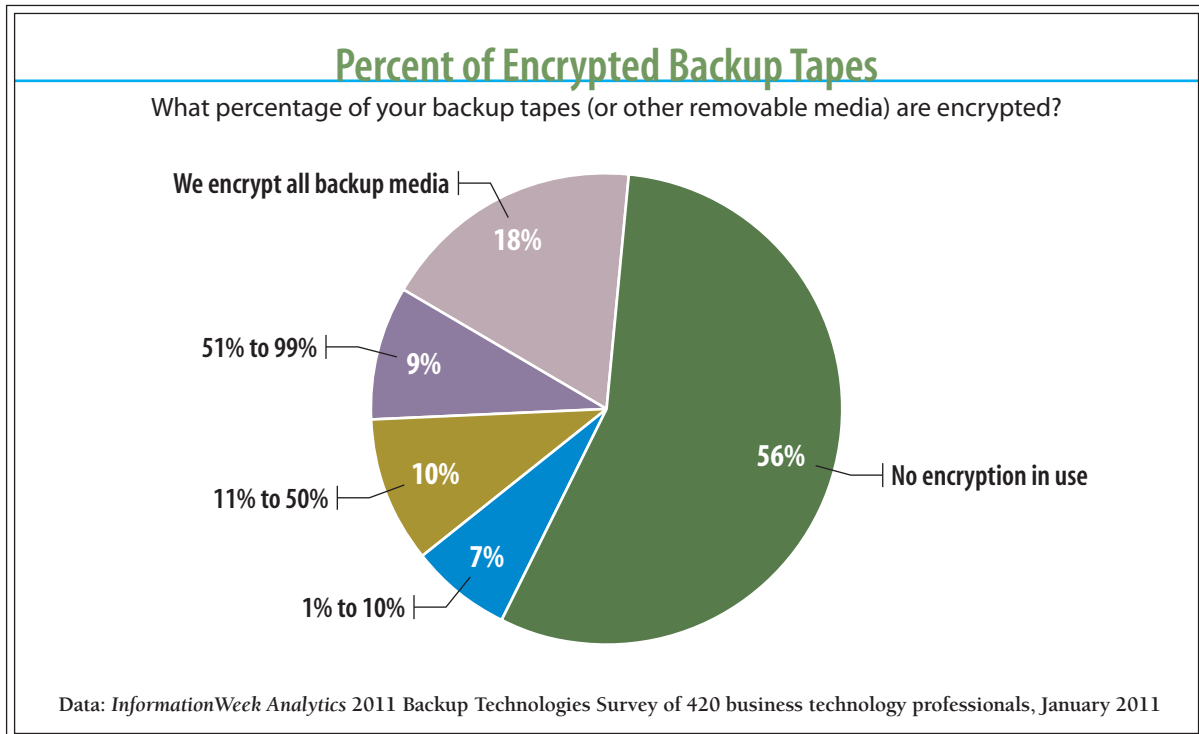


age that hosts the VMs, taking the load off the virtual server host. The vStorage APIs also do changed-block tracking, which allows backup apps to create block-level incremental backups that are faster and smaller than traditional incremental backups. Just 14% of our survey respondents say they use VCB to back up their virtual servers today, but we expect that number to increase. Meanwhile, three smaller vendors—Vizioncore, now a division of Quest Software; Veeam; and PHD Virtual—have developed backup applications specifically for VMware environments. Their applications leverage the vStorage APIs and add additional functionality, including individual-item restore for applications like Exchange and SharePoint, replication, and simple restores from a series of block incremental backups. An additional 12% of our respondents use these products.

Working Without a Net

Most states now have data breach laws similar to California’s landmark SB 1386, which requires organizations that lose personally identifiable information, like credit card or Social

Figure 19

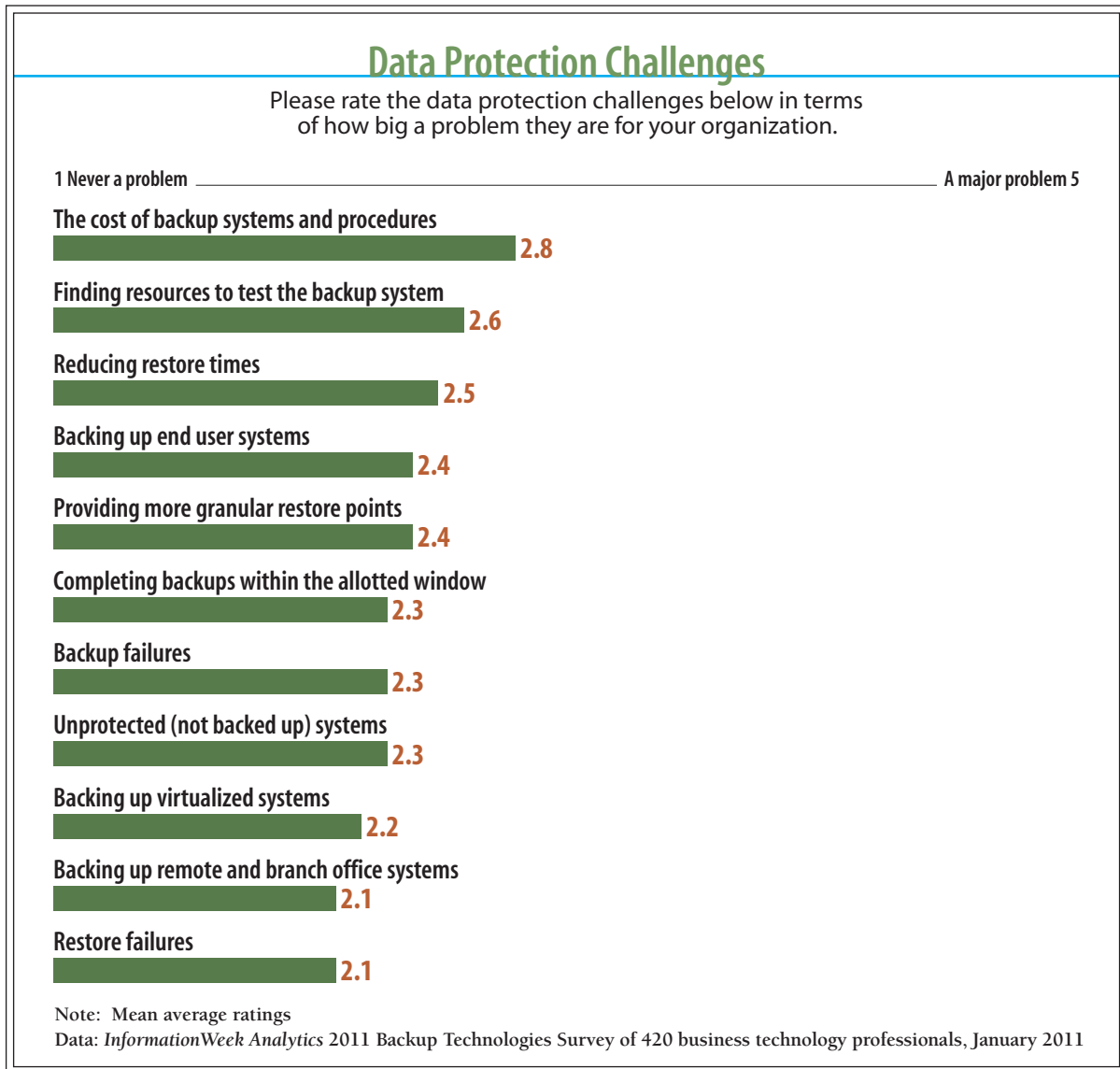




A n a l y t i c s R e p o r t s

Security numbers, to notify their customers, employees or other affected persons of the breach. As if this notification wasn't expensive, and embarrassing, enough, organizations that have had breaches generally find that they also have to purchase credit report monitoring and other identity theft protection services for the affected customers or employees.

Figure 20



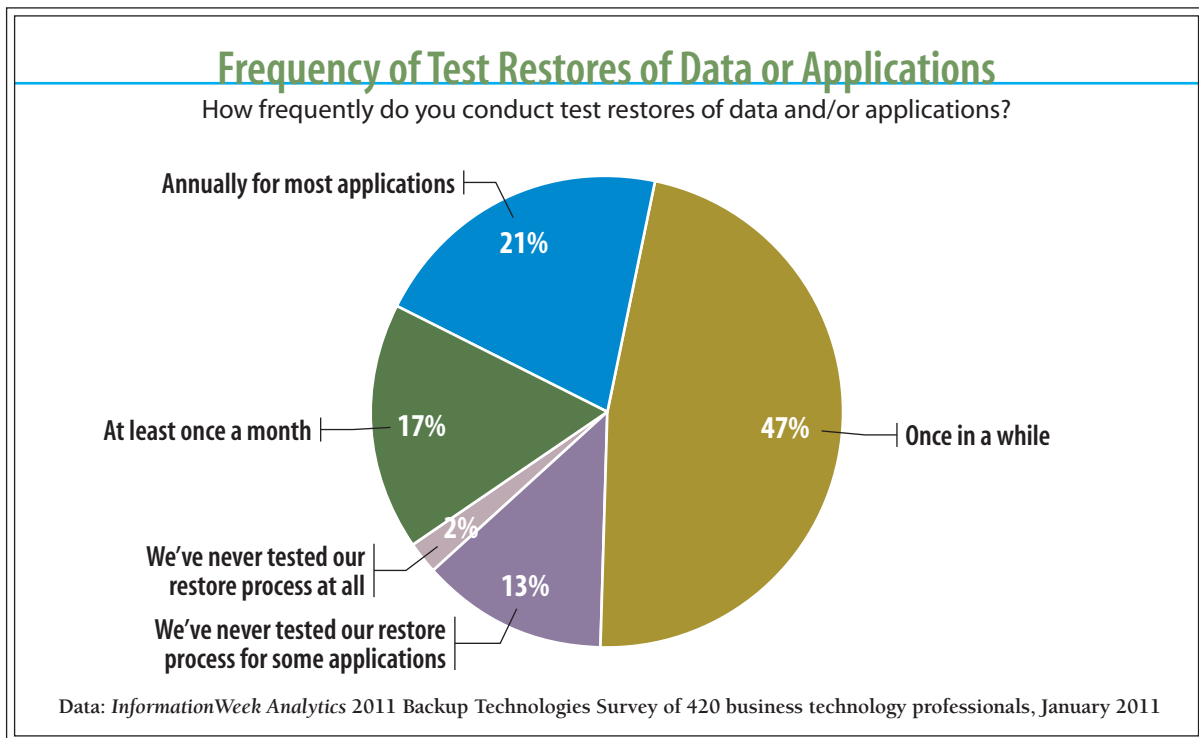


A n a l y t i c s R e p o r t s

Many of the highly publicized data breaches have been caused by lost backup tapes, including tapes lost by couriers from the most respected records management companies in the business. In one case, a courier decided to call it a night with customer tapes in the trunk of his car, intending to drop the tapes off at the warehouse in the morning. Unfortunately, however, his car was stolen from his driveway, and so the tapes were lost. Even if the probability of some evildoer actually reading the tapes is minuscule, the organization is legally required to notify the people whose data was on the tape, and if enough people are involved, must actually notify the news media as well. The good news: Since the data on *encrypted* tapes will be useless to identity thieves, most breach notification laws specifically exempt encrypted data from notification requirements. While key management remains an issue, almost all commercial backup software supports tape encryption, and the latest-generation LTO-5 tape drives enable hardware encryption with no impact on performance.

Even so, a full 63% of our respondents encrypt less than 10% of their backup tapes. Since 65% of respondents still back up directly to tape, with 39% reporting that tape is their main backup medium, we can only assume that many of these organizations are playing with fire.

Figure 21





A n a l y t i c s R e p o r t s

We believe organizations of all sizes should encrypt the data on *all* removable media—and that goes double for any media destined for offsite backup or storage. Even if all your tapes are encrypted with the a single encryption key, loss of the media will not trigger the cost and embarrassment of a breach notification. Turn on encryption in your backup software, and store the encryption key(s) in several safe places, separate from the backup media itself. USB thumb drives kept in the CIO's home and/or safe deposit boxes are an inexpensive but effective way to secure keys.

Nasty Surprises in Store?

Despite the oft-stated fact that the only reason to back systems up is so they can be restored, IT teams all too often just assume that they can restore all the data they're backing up. But even if the frequently quoted, albeit never attributed, statistic that 40% of all restore attempts from tape fail is only half right, this is still almost certainly a faulty assumption.

While 38% of our respondents report that they test their restoration processes at least once a year for most of their applications 15% say they back up applications that they have never tried to restore. At best, this inexperience will add substantially to restore times as administrators discover that their applications aren't quite as easy to restore as they thought, or the database engine takes hours to perform a consistency check if it wasn't backed up right.

At worst it will mean some applications can't be restored because there was something missing from the backup process.

Luckily, despite this lack of testing, only 7% of our respondents cite restore failures as a significant or major problem. We're impressed this figure is so low—and the cynic in us can't help thinking that it's largely predicated on an overall lack of attempted restores.

Our recommendation: All organizations should test their restore processes at least once a year, plus whenever major changes, like updates to backup software, are made to your systems. You don't want to end up like the city of New Orleans, which lost thousands of real estate records when an update to its backup system went wrong, and no one noticed until it was too late.



Figure 22

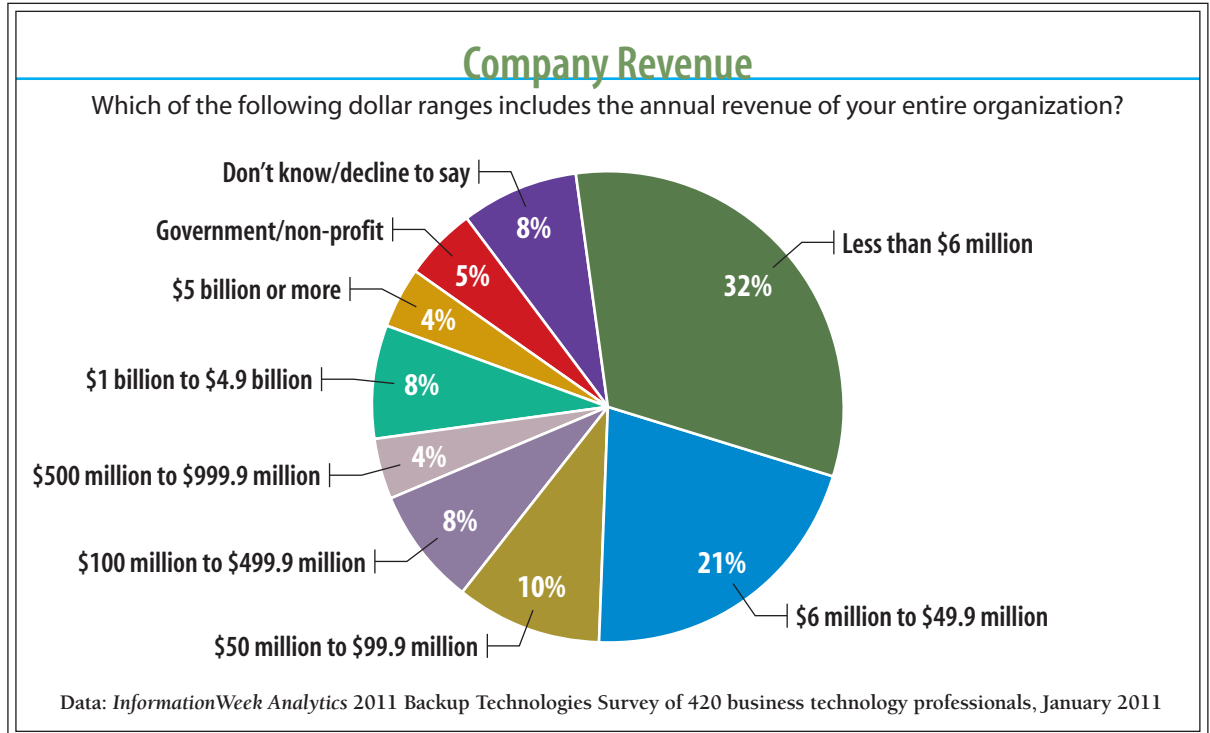




Figure 23

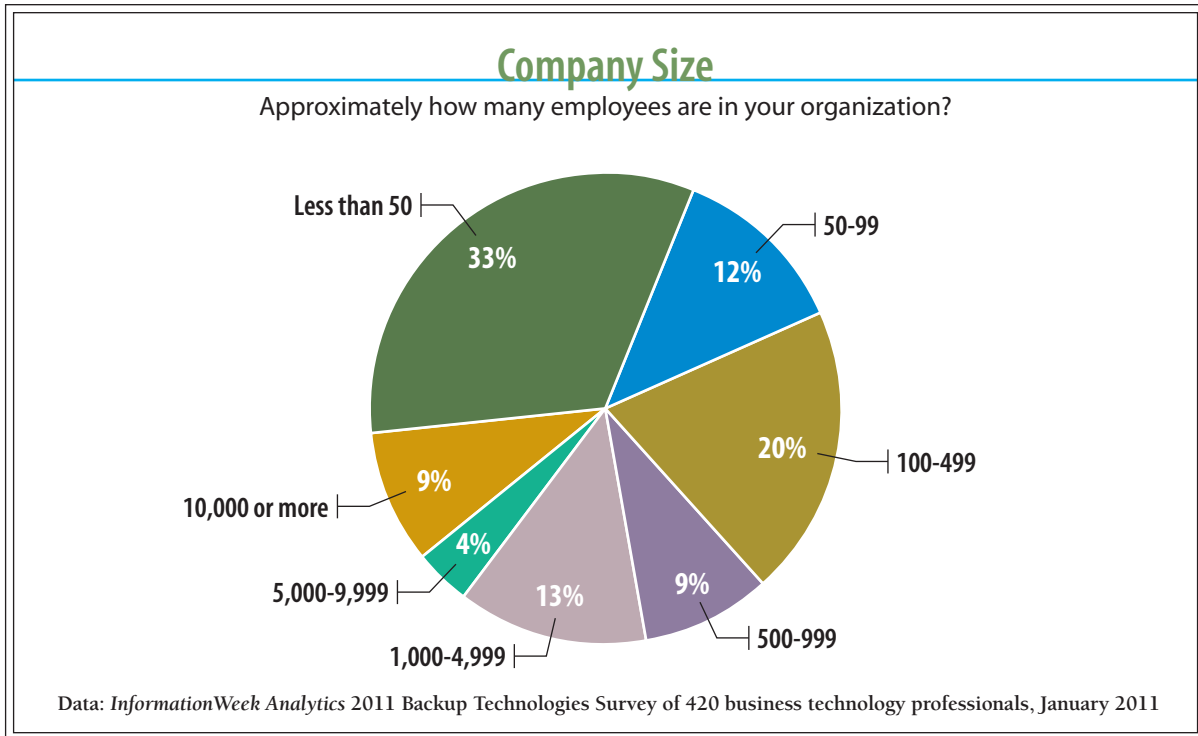
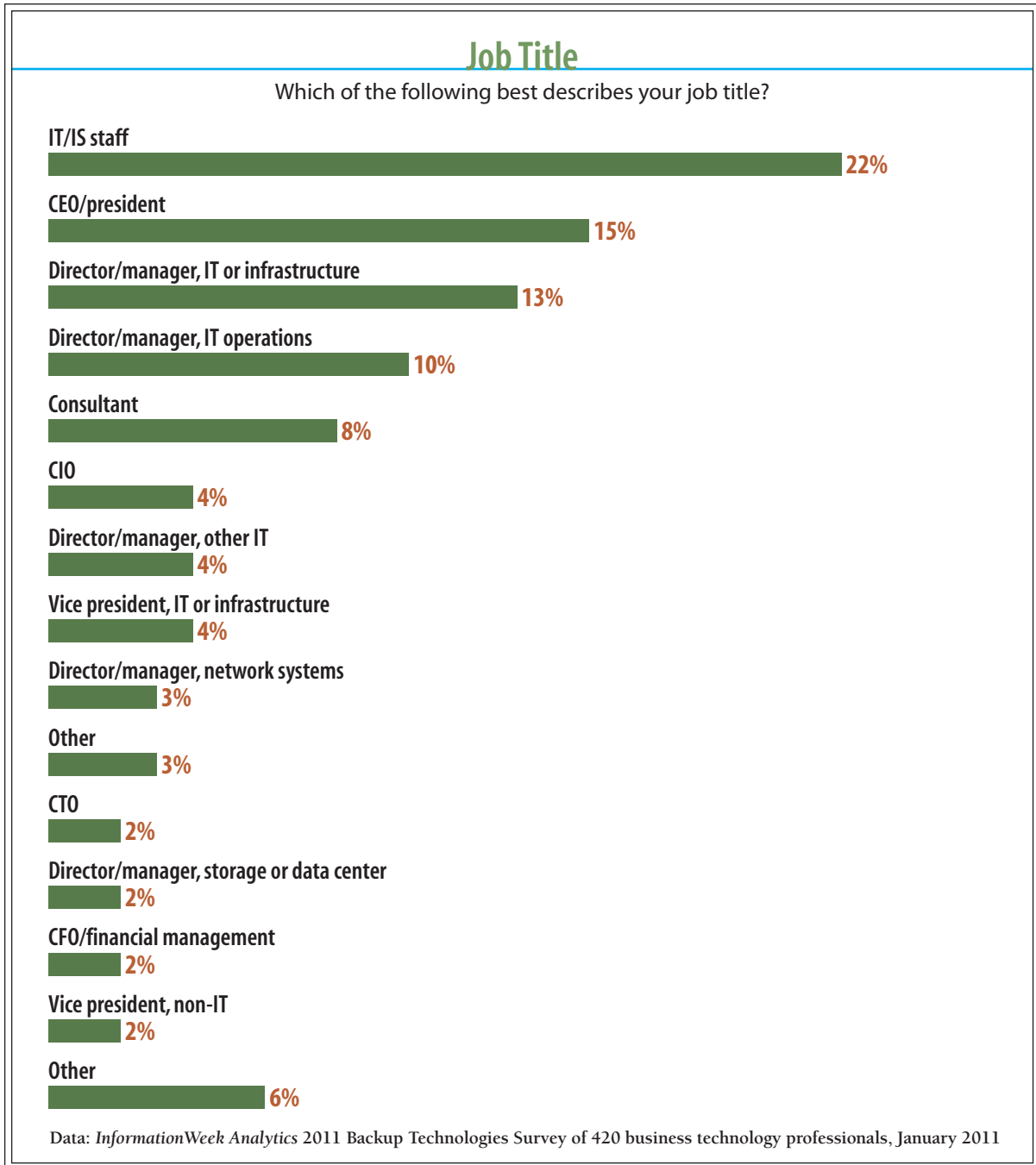




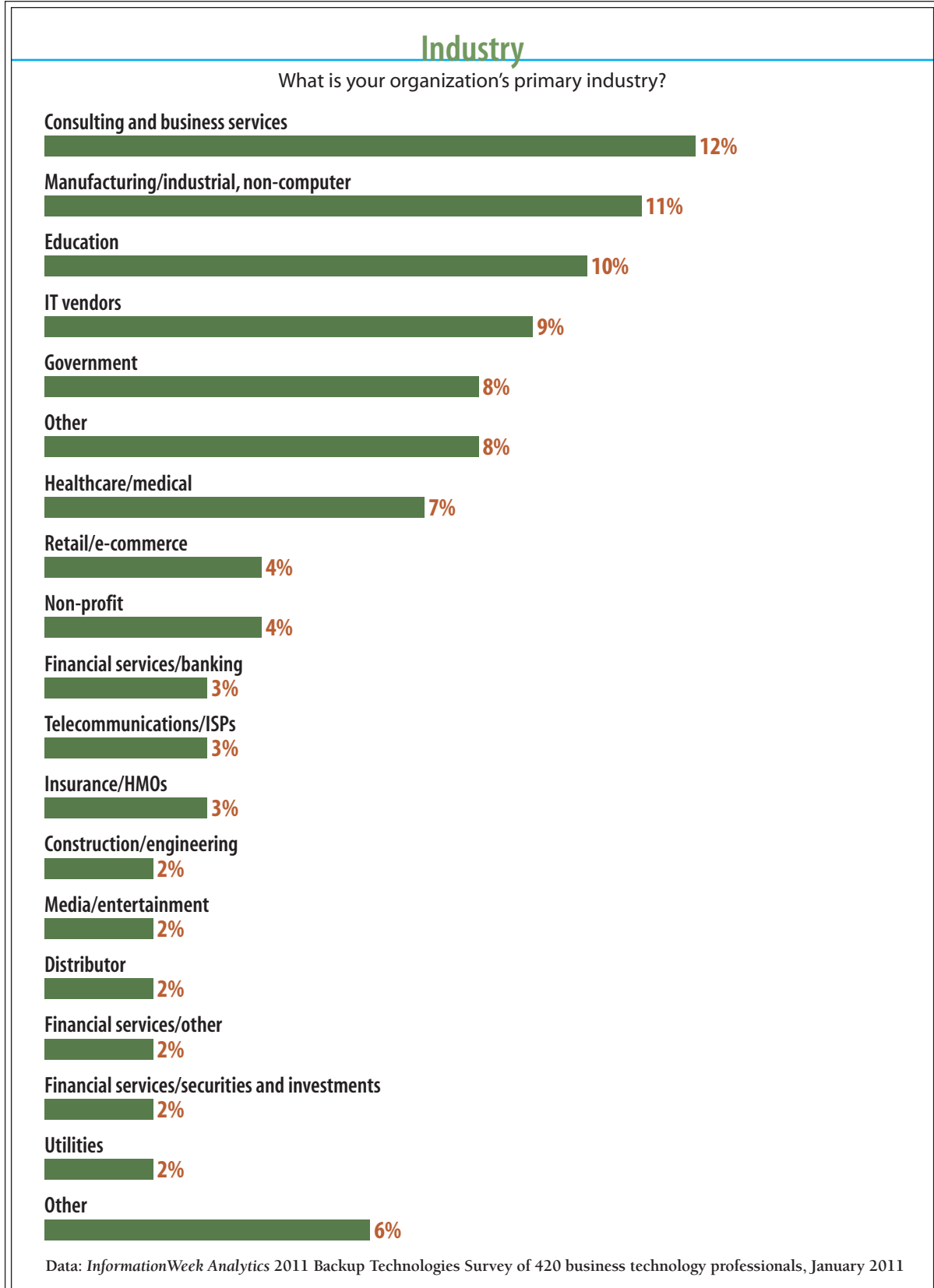
Figure 24





A n a l y t i c s R e p o r t s

Figure 25





Want More Like This?

Making the right technology choices is a challenge for IT teams everywhere. Whether it's sorting through vendor claims, justifying new projects or implementing new systems, there's no substitute for experience. And that's what *InformationWeek Analytics* provides—analysis and advice from IT professionals. Our subscription-based site houses more than 800 reports and briefs, and more than 100 new reports are slated for release in 2011. *InformationWeek Analytics* members have access to:

Research: 2011 State of Storage: The word for 2011 is “consolidation.” It's happening now in the industry writ large, through acquisitions by all the major storage vendors, and on a smaller scale in the data center. Smart CIOs will accelerate the trend.

Research: 2010 Data Deduplication: As the volume of corporate data continues to grow, IT pros keep investing in new storage usage technologies. Compression still ranks No. 1, but dedupe is gaining fast.

Strategy: Long-Term Data Preservation: IT teams are so busy making room for new data that many aren't paying enough attention to the long-term picture.

Strategy: Storage Virtualization: It's time to face the hard truth about enterprise storage: Apart from the intelligent software embedded in the controller, that Tier 1 SAN you just broke your budget to pay for is just a conglomeration of commodity components. What exactly are we paying for here?

Fundamentals: Virtual Machine Backups: Best practices for backing up VM disk files and building a resilient infrastructure that can tolerate hardware and software failures.

PLUS: Signature reports, such as the *InformationWeek Salary Survey*, *InformationWeek 500* and the annual State of Security report; full issues; and much more.

For more information on our subscription plans, please [CLICK HERE](#).