EXECUTIVE SERIES

THE GORILLA GUIDE TO...®

EXPRESS EDITION

Cloud-First Backup

Trevor Pott & James Green

INSIDE THE GUIDE:

- The hidden costs of traditional backup software
- Why cloud-first backup is changing the data protection landscape
- How security and compliance impact modern backup strategies
- Requirements for trusting a cloud provider with your sensitive data

TAKE A QUICK WALK THROUGH THE IT JUNGLE!

Compliments of solarwinds

THE GORILLA GUIDE TO...

Cloud-First Backup

Express Edition

AUTHORS
Trevor Pott & James Green

COMPLIMENTS OF SOLARWINDS

Printed in the United States of America

ACTUALTECH MEDIA
Okatie Village Ste 103-157
Bluffton, SC 29909
www.actualtechmedia.com

TABLE OF CONTENTS

CHAPTER 1

Modern Backup Is Still Too Complicated

The Data Protection Revolution

The Royal Library of Alexandria was one of the greatest repositories of human knowledge in the ancient world. Generations of effort and expense went into the Great Library, but the knowledge collected allowed Egypt to prosper for hundreds of years. But the library burned, and Egypt began its decline.

According to legend, this knowledge was amassed through a number of means, including offering trade discounts to ships that brought scrolls to the library to be copied. This represented a significant investment on the part of Egypt, but the upside was that it made Alexandria a center of scholarly excellence. Technology and academia both progressed at an accelerated rate, and the library of Alexandria was a contributing factor to a golden age for Egypt.

With its destruction, much – but not all – of the knowledge of the ancient world was lost. Some of the contents of the library were copied to other great institutions of the day, and this preservation and sharing of knowledge ultimately helped give rise to the great universities of Europe, and modern academia as we know it.

Whether in the form of papyrus scrolls or bits stored on a hard disk, data has value, and data protection has been a concern since at least the days of the library of Alexandria. Much as Egypt benefited from the data that the Great Library collected, so too do today's organizations benefit from the data they hold; that makes the protection of that data as much a must-have today as it ever has been.

While we're fortunate that today's data protection methods don't require rooms full of scribes endlessly copying scrolls by hand, data protection in the information age has traditionally been time-consuming, complicated and expensive. Thankfully, we don't need to wait another 2,000 years to see significant improvements in our approach to data protection. That revolution is happening today.

Data Protection Is Boring... Until Something Goes Wrong

Today's data protection challenges share many similarities with those facing ancient libraries. Fires and other disasters are, of course, a concern regardless of the data to be protected. Data must be constantly migrated from one storage medium to the next to keep it accessible; paper crumbles, hard drives seize up, and SSDs degrade.

The data has to be transmitted from one location to another. Whether that transmission uses horses, boats, a FedEx truck or the internet, errors in transmission can occur, and they must be planned for. Ensuring that data can be read and understood is important; libraries focus on translating languages, whereas IT teams worry about virtualization platforms and physical host compatibility.

Librarians and archivists had to be – and still are – scrupulous about the details of their craft. The acidity and quality of inks and paper, humidity, temperature, and dozens of other considerations affected the lifespan of documents under their care. Modern industrial processes have solved many of these problems, but care and attention are still required.

Systems administration is no different, and there are many moving parts in the modern data center to be concerned with. Backup administrators need to be able to navigate all the technologies that underlie the workloads for which they're responsible. In addition, backup administrators must concern themselves with privacy, security and data sovereignty issues, especially as they relate to regulatory compliance.

As with librarians and archivists before them, today's backup administrators are turning to automation to assist them. Data protection may seem simultaneously boring and complicated by an overwhelming amount of nagging details, but it's still a vital consideration. No organization wants to end up making headlines for IT failures, but a lack of adequate investment in data protection is an increasingly common way to make the news.

CHAPTER 2

Hidden Backup Software Costs

Adding to the woes of the modern backup administrator is the fact that very few organizations get to tear their entire IT apparatus down and take a completely modern, greenfield approach. In the real world, data centers are the result of incremental growth. Some workloads may have existed, largely unchanged, for decades, while entirely new workloads are added all the time.

Individual workloads in a data center each have their own requirements. These workloads are also frequently “owned” by different individuals or groups. Combined with the different ages of various workloads, the result is that, in practice, many organizations are running multiple data protection solutions, each covering different groups of workloads.

Multiple data protection solutions often means higher data protection costs. Each backup solution has to be licensed, and this licensing is often both an initial expense as well as an ongoing subscription for support and/or cloud services.

Most data protection solutions have storage considerations that also impact costs. Many data protection solutions require some form of local, on-premises storage, even those that use a cloud provider as the ultimate backup destination. Storage may be in the form of hardware appliances provided by the data protection vendor, or require dedicated storage to be designed and provided by the organization deploying the data protection solution.

Off-premises storage is a necessary part of data protection design; you don't want to keep all copies of your scrolls in one library. Not so long ago, it was common for organizations to accomplish this by backing up to tape, and shipping tapes offsite. Some organizations still use this approach.

When backing up to tape, backups are streamed to a Virtual Tape Library (VTL). This data is then copied to physical tapes, put on a truck and sent to a secure offsite storage facility. This creates three copies of the data in the form of the production copy, the onsite VTL backup and the offsite tapes.

The 3-2-1 backup strategy states that organizations should keep 3 copies of their data on at least 2 different types of storage media, with at least 1 copy being offsite. Offsite backups are an absolutely critical component of any data protection strategy. If your data doesn't exist in at least two places, then it doesn't exist.

Tapes aren't the greatest transmission medium, and as internet access became more affordable, tapes largely went away. Today, most organizations back up their data to a local disk-based storage array and then send a copy to their disaster recovery site over the internet. But this approach has its own problems.

Maintaining a disaster recovery site is expensive; done properly, a disaster recovery site is an exact duplicate of the production environment that never gets used except in case of emergency. While it's true that the disaster recovery site will pay for itself in a hurry when disaster strikes, it isn't uncommon for organizations to seek to find a way of sharing the cost of maintaining this backup compute capacity with others.

This is where cloud data protection comes in.

Cloud Data Protection

Modern data protection solutions involve sending a copy of data to a cloud storage location instead of an organization-owned disaster recovery location. Clouds are shared infrastructure, used by multiple organizations. This allows organizations to pay only for what they actually use.

More importantly, cloud infrastructure is managed by the cloud provider. Systems administrators don’t have to occupy themselves with storage management, server management or any of the other tasks associated with ensuring that the data protection destination is operational and ready to serve its purpose. Instead, they can focus on tasks that are more interesting and satisfying.

Job satisfaction has important implications for the business. Dissatisfied employees leave. Replacing them is costly. Organizations shedding administrators because of a refusal to modernize data protection could also be facing the problem of losing administrators with critical organization-specific knowledge. If one’s data protection approach consists of a patchwork of solutions, losing the only people who understand how it’s all held together can be dangerous.

Modern cloud backup solutions can help. They solve a lot of very messy, very complicated, and very time-intensive IT infrastructure problems. But cloud backup solutions are significantly different from their predecessors. Those differences should be understood before implementing cloud backups.

CHAPTER 3

Cloud-First Backup Has the Answers

Cloud backups place an organization's data in the care of the cloud provider. This is an act of trust. The organization is trusting that the cloud provider will not lose, exploit, or sell that data. Trust is hard to earn, and once lost, it's nearly impossible to regain. So how can a cloud provider gain an organization's trust, and how does an organization set about learning to trust a cloud provider?

Backing up data to the cloud vies with using Dropbox-like sync-and-share solutions as the first step taken into the cloud by organizations of all sizes. At first glance, this may seem somewhat odd; backups contain all of an organization's data. If one wanted to compromise an organization, backups seem like an ideal starting point.

Unlike many other cloud-based workloads, however, backups can easily be encrypted. Depending on the encryption solution used, and how it's implemented, this doesn't require trusting the cloud provider. The cloud provider can't decrypt the backups to see what's inside.

The current standard for backup encryption is AES 256-bit, which is generally considered acceptable even for military use. This encryption has 1.1 x10^{77} possible combinations, and with the most advanced known techniques – ones specific to AES, and not restricted to optimal theories about key space size – it would take longer than the universe itself has existed to crack.

How encryption keys are handled is also a part of determining whether or not backups are secure. Cloud backup solutions should support both private key and centralized key management solutions. With private keys, the organization handles its own key management. Centralized key management is typically for those organizations looking for more ease of use, and may require a third-party key management solution, or be integrated directly into the data protection software itself.

Encryption in-flight is just as important as encryption at-rest. Communication between an organization's on-premises data center and the data protection solution running on the cloud side should occur via TLS or VPN tunnels. This will minimize the chances of data being compromised by malicious third parties that have access to one of the intermediate network links between the organization and the cloud provider.

Perhaps most importantly, the cloud provider should be committed to meeting as many compliance standards as possible. Important compliance standards to look out for are the use of ISO 27001 certified data centers, compliance with the SSAE 18 auditing standard, and the Service Organization Control 1 (SOC-1) Type 2 requirements covering the data center's internal control over financial reporting.

Certifications are important; not because of the ability to check boxes in a compliance effort, but because of the underlying effort they represent. From a systems administrator point of view, the goal is to select data center operators that have tight controls over who can access the data stored in these data centers, and who ensure that every access is logged and audited.

This approach to data center design, operation and management is important to ensuring that sensitive Personally Identifiable Information (PII) is protected. Protecting PII is at the core of a number of regulatory regimes, including the Sarbanes–Oxley Act (SOX), the Health Insurance Portability and Accountability Act

(HIPPA) and PCI-DSS, all of which data centers can – and should – achieve certification for.

Data center certification efforts are also part of, but not the whole of, achieving compliance for more demanding regulatory standards, such as the European Union's General Data Protection Regulation (GDPR). While the GDPR requires efforts that transcend IT and place demands on business practices and organizational structure, IT teams are required to exercise care over where the data is stored, who can access that data, when, and under what circumstances.

Privacy and data sovereignty concerns may require that organizations store their backup data not only in data centers meeting a high standard of security compliance, but also that the data be physically stored within specific jurisdictions. A global footprint is increasingly important, especially for any organization seeking to handle the data of EU citizens.

This global footprint is ever-evolving. Currently, Europe recognizes 12 countries or territories as having adequate privacy laws, a concept important for the handling of EU citizen data. Canada, once considered to have adequate privacy laws, is increasingly considered to be only partially adequate. The U.S., meanwhile, is questionably adequate, and only for as long as the Privacy Shield legislation remains undefeated in EU courts, something that may happen sooner rather than later.

Because the intricacies of politics are so fluid, the details don’t matter quite as much. What matters is that when choosing a cloud provider to act as a data protection destination, a diverse global footprint – preferably one that includes multiple EU nations – and a commitment to expanding that footprint should be considerations as important as encryption or compliance standards.

Checkbox compliance isn’t enough, either; cloud data protection providers must demonstrate an ongoing commitment to meeting ever-evolving privacy, security and data sovereignty concerns. Ideally, they should also have a track record of confronting these challenges in a proactive manner, so that organizations are never left in regulatory limbo.

CHAPTER 4

Deploying Cloud-First Backup Is Easy

Cloud backup solutions can be broadly divided into two categories: those that require some form of Cloud Storage Gateway (CSG), and those that don't. While there is certainly a great deal of variation among the various offerings, solutions that require a local repository or staging area tend to have notable differences from those that don't, and these differences influence all aspects of data protection design within an organization.

Virtually all cloud data protection solutions employ some form of data efficiency. This is a real-world requirement of cloud data protection; internet bandwidth is expensive, and many organizations have an upper limit on how much upstream bandwidth their local internet service providers are willing to provide. Deduplication, compression and sometimes WAN acceleration are all technologies commonly used to achieve this goal.

Solutions that employ a CSG serve as a buffer for backups. Backups are sent to the CSG, which then performs data efficiency activities upon the data; that data is then unspooled to the cloud data protection destination.

CSGs have several advantages. Because they're a dedicated device, they can be configured to contain a full manifest of all data blocks in the data protection regime. This allows for more efficient deduplication, and thus may place a lower demand on internet throughput than cloud data protection solutions that don't use CSGs.

CSGs also usually have dedicated hardware. This means that data protection activities don't impact the CPU of running workloads. If the CSGs serve a dual role as a local repository, they may also contain a copy of some or all of the backups that have been sent to the cloud.

The disadvantages of CSGs is that they're typically both a significant capital expense and an ongoing operational and management expense. In addition, because CSGs buffer data to be sent to the cloud and unspool it over time, the backup arrival at the data protection destination can be delayed for several hours when compared a direct backup approach. This lengthens Recovery Point Objectives (RPOs), so in the event of a disaster, more data could be lost than would be the case with a direct cloud backup solution.

Both direct-to-cloud data protection solutions and CSG-based solutions may use Changed Block Tracking (CBT) or CBT-like technologies as part of their data efficiency. All CBT solutions function in a similar manner: an initial replication is performed, which includes a full copy of the backup set. From that point forward, every write that the protected workload makes to its storage is sent to the data protection destination.

Direct-to-cloud data protection solutions typically use an agent that's installed. This has the advantage of shorter RPOs when compared to CSGs, but they generally must perform their data efficiency operations in the protected workload itself. This can mean an increased CPU overhead for each protected workload, but in practice this overhead typically proves to be minor.

CHAPTER 5

Restores Are A Pleasure

While all the concepts discussed here are important considerations, the single most important concept in any data protection discussion is the ability to restore your data. If you can't recover your data, then it doesn't exist.

There are two primary restore concerns: the Recovery Time Objective (RTO), and restore verification. Restore verification is the most important component of data protection. Tragically, it's also often the most overlooked. Restore verification ensures that data can be extracted from backups, restored as expected, and used as intended.

In the case of individual file restoration, this means the ability to extract the correct file version and then restore it to the location desired by the end user. In the case of a complete workload, restore verification would mean the ability to return that workload to operation, typically through the use of virtualization.

RTOs are the measure of how long it takes to restore requested data or workloads to their intended state. RTOs are significantly impacted by the technologies in use. Local backup repositories, for example, typically offer the fastest RTOs for restoring a workload back to its original on-premises infrastructure. These are frequently used for workloads where low RTOs are a necessity, as it can take several minutes or even hours to pull a complete workload down from an offsite data protection location, such as a cloud provider.

Data protection solutions that allow for instantiation of workloads directly from a backup set are faster still. These solutions can restore a workload to functionality in seconds. This functionality may be offered by solutions that integrate temporary workload capacity into a local data repository or CSG.

Advanced cloud data protection solutions offer even more choices. Those solutions with block-level restore granularity can roll back workloads to a previous version quickly, as only change blocks need to be fetched from the cloud.

Many solutions also offer the ability to "pre-warm" a data recovery site, downloading a copy of protected workloads and keeping them up to date until a planned failover occurs. This same technology can be extended to offer continuous recovery, where a disaster recovery site is kept perpetually warm and ready to take over should the primary location fail.

CHAPTER 6

Cloud-First Backup Keeps Your Business Available

Regardless of which data protection solution is chosen, the important thing is to ensure that your data is being protected. In a recent survey that included more than 2,200 respondents, only 18% believe that their current data protection solution will meet all future business challenges. The same survey reports that the average cost of data loss is as high as $900,000, with the average cost of downtime being $555,000. Data protection is not an option. It's a business necessity.

Data protection is also complicated, and as a result it's easy to get wrong. No one wants to be in the headlines for getting it wrong. Because of this, cloud data protection is quickly becoming the first choice for data protection. Cloud providers take care of a lot of the complexity involved with data protection. They also reduce the expense of having an additional site for disaster recovery purposes.

The biggest reason to fear cloud data protection is that trusting others is hard. This has been solved.

Encryption gives organizations control over their backups, and data protection providers have upped their game by investing in certification efforts that aim to enable trust.

With the right cloud provider, not only are backups safe in the cloud, but running workloads – as required during disaster recovery events – can be reasonably considered as safe as they would be on-premises. “Cloud first” is the new normal for data protection, and these solutions are increasingly easy to try out.

Take the time to sign up to a cloud data protection provider and protect your workloads today.

After all, you never know what might happen to your library tomorrow.