CONTENT PROVIDED BY

Redmond

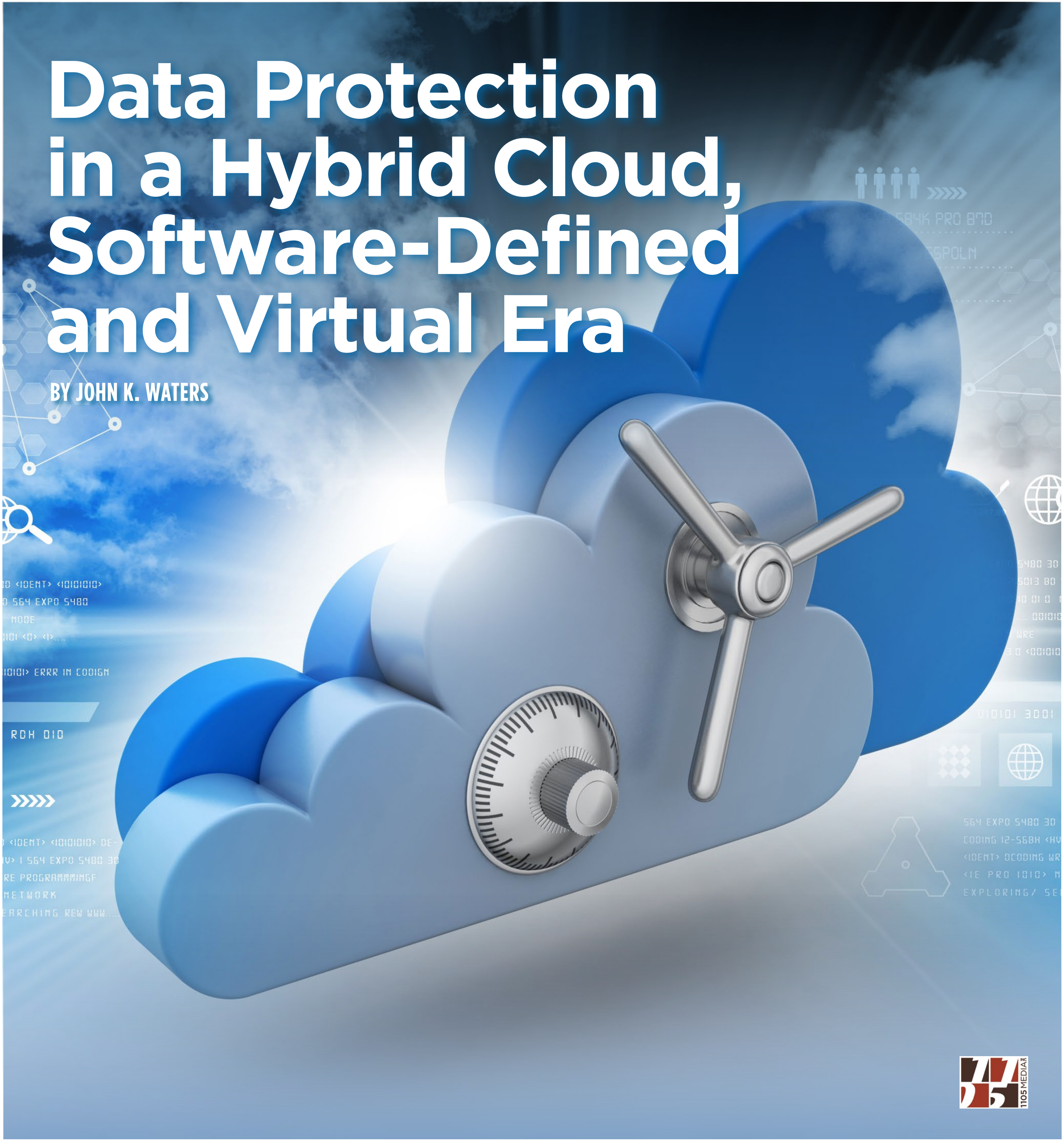
THE INDEPENDENT VOICE OF THE MICROSOFT IT COMMUNITY

SPONSORED BY

DELL Software

Data Protection in a Hybrid Cloud, Software-Defined and Virtual Era

BY JOHN K. WATERS



1105 MEDIA

Introduction

The enterprise data protection challenge has never been greater. Not only are IT organizations being asked to safeguard more data than ever before from a range of sources most of us couldn't have imagined just a decade ago, they're being asked to do it in distributed environments over which they have less direct control, and within the constraints of budgets that seem to be shrinking as their burdens expand. At the same time, competitive pressures are increasing the need for extremely rapid recovery across physical, virtual, and cloud environments.

The days when an organization's data could be tucked safely behind a local firewall and preserved with a nightly backup routine are long gone. This paper looks at current trends emerging around enterprise data protection in this era of virtualized infrastructures, hybrid clouds, and software-defined everything, and lays out some of the ways in which IT organizations are coping with, and deriving value from, these developments.

The days when an organization's data could be tucked safely behind a local firewall and preserved with a nightly backup routine are long gone.

Data Protection Defined

For the purposes of this paper, the term "data protection" refers to backup, operational recovery (OR), disaster recovery (DR), and business continuity (BC). A "backup," of course, is a data-protection copy of the official working copy of the data. OR uses backups to restore individual and small groups of systems and data. DR refers to the process of recovering an entire IT infrastructure from a remote site after a primary site has gone down. Solution vendors commonly present DR and business continuity (BC/DR) as a combined set of processes for recovering from a disaster while maintaining routine business operations.

Current Trends

Industry analysts at IDC expect (**http://www.idc.com/getdoc.jsp?containerId=EMEA40856515**) the market for increasingly sophisticated, cloud-based data protection solutions to approach $8.1 billion by 2019. Here are some of the current trends driving that growth.

Expanding Multi-Cloud Universe

The number of clouds in the average enterprise is growing and will likely continue to grow. Analysts at Gartner have estimated (**http://www.gartner.com/newsroom/id/2599315**) that 50% of large enterprises will be using hybrid clouds by 2017. Organizations with significant infrastructure investments are moving to the hybrid model to take advantage of the benefits of a public cloud (on-demand pricing, automation, scalability, flexibility), while maintaining some of the service predictability and security of a private cloud. By some estimates (**https://www.rightscale.com/lp/state-of-the-cloud?-campaign=701700000015euh**), the average number of active clouds in an enterprise has already grown to half a dozen, and Gartner expects that number to reach 17 sometime next year.

The latest catalog of data protection solutions targeting hybrid cloud environments is varied, and more products and services are emerging every day. Solutions range from on-site-appliance-based systems to pure cloud offerings that integrate identity management, data protection and monitoring, and security analytics and threat monitoring.

The latest catalog of data protection solutions targeting hybrid cloud environments is varied, and more products and services are emerging every day.

A truly comprehensive data protection solution addresses the risk to all three states of information: Data at Rest (archived data on disk or tape), Data in Use (active data currently being manipulated by an application), and Data in Motion (in transit).

It also addresses an organization's Recovery Time Objective (RTO), which is the amount of time within which a business process must be restored from backup before the downtime breaks continuity, and its Recovery Point Objective (RPO), which establishes the maximum amount of data loss an organization can tolerate during a disaster. The RPO is derived from the business continuity plan. A growing number of solutions in this market are taking advantage of the cloud's ability to shorten recovery times; some are promising zero RTO and "aggressive" RPOs.

Blurring Lines Between Insiders and Outsiders

As the adoption of hybrid cloud architectures grows, so does the number of third-party relationships that involve data protection issues. Enterprises increasingly rely on cloud services providers to handle core business functions. That means more outside entities have some level of access to the enterprise network and sensitive data. Analysts at 451 Research have concluded (**http://enterprise-encryption.vormetric.com/rs/480-LWA-970/images/Vormetric_2016_Data_Threat_Report_Global_WEB.pdf**) that the number of third-party data relationships can "easily number in the tens of thousands" in some large global firms today.

Information in a hybrid cloud is, by definition, under the control of both the enterprise and the cloud services provider at various points in the data's lifecycle. Before trusting a provider with your data, you'll want understand its data protection capabilities as well as your own. Ask specific questions about how the provider handles replication, backup, and disaster recovery. Consider how the provider protects data in motion, in process and at rest.

A hybrid architecture essentially extends the private cloud into a public cloud, so it makes sense to do everything possible to replicate your organization's in-house security policies and best practices across both environments. Security controls such as authentication, authorization, and identity management must work in both clouds. Also consider things like maintaining company ownership of server keys and encrypting data in motion, in process and at rest. Two approaches are recommended here: 1) replicate controls in both clouds and keep security data synchronized, or 2) establish a single identity management system for both clouds.

Thanks in no small part to the advent of the Internet of Things (IoT), petabytes of structured and unstructured data are flowing into public and private clouds at astounding rates.

Big Data Explodes the Cloud

Thanks in no small part to the advent of the Internet of Things (IoT), petabytes of structured and unstructured data are flowing into public and private clouds at astounding rates from a dizzying array of devices. That large and exponentially growing volume of data flowing into your organization known as Big Data is creating an unprecedented data protection challenge for IT organizations. More data inside the backup system means longer recovery times, which means established backup windows are being cracked out of their frames. Adding to the challenge are government regulations such as data localization laws are forcing organizations to retain more data for longer periods of time.

The volume of enterprise data continues to grow, unabated, but infrastructure has its limits. Although a certain amount of data redundancy across datacenters is essential to mitigate the impact of an outage in a single datacenter, most organizations are going to need to employ some type of data reduction strategy to maintain valid backup windows.

One of the most often recommended solutions here: data deduplication. Sometimes called intelligence compression or single-instance storage, data deduplication is the process of eliminating redundant data in a backup environment by retaining a single, unique instance of the data in storage. The redundant data is replaced with a pointer to the unique data copy.

Deduping technology is not new, and a wide range of vendors provide solutions with powerful dedupe capabilities. These tools commonly perform the deduping on the source-side data (software running on the server), but any strategy for reducing the source-side data should include a coordinated reduction of target-side data (data in storage), analysts recommend.

Compliance Gets Complicated

A multi-cloud universe presents some unique challenges when it comes to regulatory compliance. When an organization's data lives on both public and private clouds, it can be difficult to demonstrate compliance with security regulations mandating the protection of sensitive financial, healthcare, or personally identifiable information (PII). The enterprise controls the internal systems, but has much less control over the cloud provider. And how do you prove that the data-in-motion between the two clouds is protected? Meanwhile, the European Union has just approved the new General Data Protection Regulation (GDPR), which will impact any organization dealing with the personal data of EU citizens, and new legislation is under consideration in the U.S.

A multi-cloud universe presents some unique challenges when it comes to regulatory compliance.

Fortunately, a new generation of data protection solutions is providing tools and technologies for controlling encryption keys, managing encryption policies, and dealing with access and authentication in hybrid clouds. Current solutions address many compliance and privacy requirements around encrypting data in motion between private and public clouds.

Software-Defined Datacenter

VMware threw a spotlight on (**https://blogs.vmware.com/tribalknowledge/2012/08/the-software-defined-datacenter-meets-vm-world.html**) the concept of a datacenter in which "all infrastructure is virtualized and delivered as a service, and the control of this data-center is entirely automated by software" back in 2012. The so-called software-defined datacenter (SDDC) helps to eliminate traditional data center silos by consolidating infrastructure components and services. The problem is that traditional data protection models are not optimized to work in this new software-defined environment.

By 2020, analysts at Wikibon predict (**http://siliconangle.com/blog/2016/02/08/will-software-defined-revolutionize-data-protection-cubeconversations/**), most IT organizations will need to re-architect their backup, primarily because they will have failed to adapt their system to the changing environment. One oft-cited reason for this failure is the misconception among organizations considering many of the trends cited in this paper that they need to make wholesale changes. But many analysts advocate a baby-steps approach. Implementing an SDDC, for example, could start with a small-scale, non-mission-critical project that addresses a single aspect of the environment—compute, storage, or networking, but not all three.

Looking Ahead

They're not industry drivers yet, but several cutting-edge trends in the data protection space are well worth tracking.

They're not industry drivers yet, but several cutting-edge trends in the data protection space are well worth tracking.

Software-Defined Data Protection

The newish software-defined data protection (SDDP) category is being called the next generation of backup. Services providing SDDP allow IT organizations to manage backup data in an automated, policy driven, and transparent way. A growing number of vendors are offering solutions that support this approach to reduce bulk and improve usability.

This software-defined approach to data protection can also show up as a capability within Infrastructure-as-a-Service offerings, where the organization has the ability to manage DR in the cloud, and in Backup-as-a-Service solutions, in which the service provider manages VM backup from the production datacenter into the cloud.

Disaster-Recovery-as-a-Service

According to Gartner, by 2017, 50% of large enterprises will use IT services failover between multiple data center sites as their primary disaster recovery strategy. Disaster-Recovery-as-a-Service (DRaaS) is emerging as something of a successor to traditional DR. It is most often defined as the replication and hosting of physical and virtual servers by a third-party to provide failover in the event of a disaster. The definition is a little squishier among vendors, many of whom are pushing DRaaS as the next big thing, but Gartner added it as a Magic Quadrant last year. The analyst firm noted (**http://www.datacenterknowledge.com/archives/2015/04/24/gartner-names-sungard-leader-first-draas-magic-quadrant/**) at the time that companies with “increasingly aggressive recovery-time objectives” and “blended” datacenter operations were “forcing DRaaS to evolve.”

Flat Backups

So-called flat backups, which use snapshots to perform backups directly to the storage system without the use of traditional backup software, are gaining traction in the enterprise. The concept isn’t new, but because snapshots reside on the same storage array as the data being protected, they have not traditionally been considered true backups. Solutions performing flat backups today replicate snapshots in a different location, eliminating that vulnerability. Because a flat backup doesn’t require a backup server or a media server, it’s much less expensive than traditional backups. As of this writing, the number of data protection vendors supporting flat backups is small.

Hyper-convergence

The converged datacenter, which integrates physical servers, storage systems, and network devices, is being superseded by the hyper-converged datacenter as more and more enterprises recognize the potential value of a software-centric datacenter architecture that tightly integrates compute, storage, networking, and virtualization. Eliminating barriers among these resources in this way promises faster deployment, simpler management and operations, cost savings, and greater elasticity and scalability.

The converged datacenter, which integrates physical servers, storage systems, and network devices, is being superseded by the hyper-converged datacenter.

Microsegmentation

High-profile security breaches of the past few years have made it clear that traditional perimeter security technologies focused on “north-south” traffic—firewalls, intrusion detection and prevention solutions, etc.—can’t protect a datacenter buzzing with internal “east-west” traffic. By some estimates, three-quarters of datacenter traffic today is east-west. Enter microsegmentation, a security architecture that divides the datacenter into smaller, protected zone. Essentially, every VM gets its own firewall, creating a zero-trust network within the datacenter. In the coming year, expect the market to move beyond early-generation solutions thanks to lively competition among vendors, which could turn all the talk about this security strategy into reality.

Conclusion

Data protection in a virtualized, multi-cloud, and increasingly software-defined world is an evolving and challenging process. By staying abreast of the industry trends and competitive imperatives that are driving enterprises into these environments, IT organizations can take advantage of new tools and emerging practices to keep their companies' data safe and accessible.

John K. Waters has been covering the information technology beat from Silicon Valley and the San Francisco Bay Area for more than two decades. He serves as Editor-at-Large for Application Development Trends *and contributes to a range of online and print publications. He's also the author of more than a dozen books, including* The Everything Guide to Social Media, The Everything Computer Book, Blobitecture, Waveform Architecture and Digital Design, John Chambers and the Cisco Way, *and* Diablo: The Official Strategy Guide.