DMI
Data Management Institute

AUTOMATED, INTEGRATED AND PROTECTED: THE DATA STEWARDSHIP IMPERATIVE

BY

JON TOIGO

CHAIRMAN, DATA MANAGEMENT INSTITUTE

SUMMARY

If there is one fact that has become indisputable in 2018, it is that data protection is the centerpiece of strategies for both disaster preparedness and business continuity, and also data security and information privacy. The increasing frequency of both natural and man-made disasters, and of security breaches and unauthorized data disclosures, have contributed to making data protection a front-of-mind concern for organizations of all sizes and across all industries.

Whether you use mainframes, distributed computing systems, or a mix of both, whether you rely on on-premises data centers solely or have adopted cloud computing as part of a hybrid IT model, data protection remains a priority in IT management, planning and operations. When building business-savvy technology infrastructure and next generation applications, planners must consider the risks in the milieu in which the business operates. They must provide, in every aspect of the compute, network and storage platform, and in the design of applications themselves, the means to safeguard data, to avoid preventable interruptions, and to recover quickly from interruption events that cannot be prevented.

In short, data protection needs to be pervasive -- and effective -- as part of the data stewardship imperative. This paper examines the latest techniques for reducing the likelihood of physical failures and security breaches that can compromise data and expose the organization to large financial losses. Pervasive data protection will not only reduce the risk of financial and reputational loss, but also of legal or regulatory non-compliance.

INTRODUCTION

The challenges confronting the contemporary IT practice are many. IT planners are deluged with fast paced changes to

- Application and operating system software,
- Infrastructure models and components, and to
- Business processes and their support and service level requirements.

At the same time, planners must constantly evaluate new technologies, deploy those that make sense, all while maintaining the infrastructure and administering the complexity that inevitably results. And, all of this must be done while the necessary IT staff skills and knowledge are cultivated through training and/or hired from a shrinking pool of qualified workers.

These challenges are not new, but they are confronting planners against a backdrop of greater risk and uncertainty. Recent years have seen an increase in the number and severity of natural disasters. Fifteen separate weather and climate disasters in the US caused at least $1 billion in damages in the U.S. in 2017, according to the National Oceanic and Atmospheric Administration[1]. Initial estimates from AccuWeather[2] estimate the damage from the northern California wildfires to be more than $1 billion. If so, 2017 will tie 2011 for the most billion-dollar disasters. And that doesn't even count man-made calamities ranging from infrastructure failures to terrorist attacks, which are also on the uptick.

In the realm of information security, hardly a week goes by without another account of a data breach or malware or ransomware attack. In the first half of 2017, over 1.9 billion electronic records had been breached by bad actors[3], continuing a pattern of attack that has security analysts placing the likelihood of a data breach within an organization over the next 24 months at 27.7%[4]. Hackers have accelerated their activities and turned to revenue-generating strategies such as leveraging ransomware and other attacks to monetize, rather than simply vandalize. The effect, according to analysts, is a surge of attacks on data hosted on all computer platforms that exceeds all prior records[5].

Governments and industry associations are responding to the situation by tightening up regulatory and legal mandates that govern how organizations host private and sensitive data. This May, the European Union will implement a sweeping set of rules, called the General Data Protection Regulation (GDPR), that may well presage a wave of similar mandates seeking to

[1] https://www.ncdc.noaa.gov/billions/events/US/1980-2017

[2] https://www.accuweather.com/en/weather-news/devastating-california-wildfires-predicted-to-cost-us-economy-85-billion-containment-may-take-weeks/70003000

[3] Breach Level Index -- http://breachlevelindex.com/

[4] 2017 Ponemon Cost of Data Breach Study https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=SEL03130WWEN&

[5] https://betanews.com/2018/01/25/2017-ransomware-volume/

enforce information security discipline on business and governmental entities[6]. IT planners in companies doing business in Europe are already scrambling to find ways to comply.

So, the prioritization of data protection on the to-do lists of business IT planners should come as no surprise. Even in mainframe shops, which were once thought to have a more resilient attack surface than other compute platforms, IT planners are scrambling to find ways to reinforce existing data protection and recovery processes and to mitigate risks as best as they can.

Standing in the way is the challenge of complexity. Infrastructure tends to be heterogeneous (provided by different vendors) and its management tends to be fragmented. New technology for software-defined and cloud architecture is increasing the complexity, rather than simplifying, workload processes and maintenance models. It is nearly impossible to find a one-size-fits-most solution that can wrap around the existing IT environment and deliver the improved security and resiliency that are sought.

The solution must begin at the layer of data itself. Data is the one common denominator in the calculus of disaster recovery and information security. Companies can recover from a temporary infrastructure outage, network failure, or facility loss. However, if data is lost, recovery is impossible.

THE PRIMACY OF DATA PROTECTION

The primacy of data in successful recovery strategies is well understood by business continuity and information security practitioners. From a disaster recovery standpoint, a copy of data must be made on an on-going basis and stored at a distance from the original data sufficient to prevent both the original and the copy from being consumed by the same disaster.

Tape technology remains a stalwart in data copy and offsite storage, and interest in the technology is growing as higher capacity tape media and ease-of-use improvements in tape backup and archive come on line. IBM®'s innovations in tape technology, both in library and tape drive hardware and in data storage methodology, and their support for the latest high capacity tape media formats – both IBM 3592 Enterprise and Linear Tape Open (LTO) – have helped to provide a renaissance in tape in 2017 that is likely to persist into the coming years.

Even cloud service providers, some of whom previously disparaged tape technology, have begun to embrace the medium. Large purchases of IBM libraries, featuring the latest TS1155 enterprise drives were made in 2017 by leading cloud service providers in expectation of zettabyte sized storage requirements by 2020. Tape capacity, which is showing greater capacity growth than any other media, will be required to host the 60+ zettabytes of new data cloud providers expect to see by 2020, and the 160 zettabytes expected by 2024.

[6] http://www.itpro.co.uk/general-data-protection-regulation-gdpr/30107/get-gdpr-ready

IBM tape offers not only capacity (15TB uncompressed for the latest enterprise tape drive TS1155 and 12TB uncompressed for LTO-8 media), but also ease of use. Innovations such as the Linear Tape File System (LTFS) have made tape as simple to use for data storage as a USB thumb drive. Plus, innovations in media substrates and coatings by industry leaders have created media with 30 year lifespans and capacities that rival every other media type.

As a practical matter, tape enables the movement of large quantities of backup data to off-site or cloud-based destinations. "Cloud seeding" using LTFS tape is much more efficient than using wide-area or metropolitan-area network links for data backup transfers, especially in the case of large quantities of data. And cloud/off-site storage providers find the handling and administration of LTFS media to be far easier than prior tape formats.

Of course, storing all copies of data off-site is not a complete solution to the challenge of data protection. For many companies, using a Virtual Tape Library (VTL), to host a copy of data locally (in addition to and not as a replacement for safe off-site storage) makes sense, too.

Localized storage to a VTL of critical backup data provides a means to perform faster local restores in response to disasters or security breaches that do not necessitate relocation to a different operating facility or to a hot site. In a strategy some call disk-to-disk-to-tape or disk-to-disk-to-cloud, the second instantiation of backup data to a local disk or flash-based VTL provides an expedient means to restore data to a usable form within seconds of an application or primary hardware failure or malware/ransomware attack.

IBM's VTL platform, the TS7700 system family, exemplify the best of VTL technology. The platform supports the latest links and interconnects, including 16 Gb FICON fabrics, to expedite data movement and to reduce the time required to protect customer and corporate data. TS7700 also leverages the latest data reduction technology, if desired, to compress backup data so that users get the most storage in the smallest possible footprint for their fast-growing critical datasets.

The latest TS7700 features performance that rivals most tier one storage arrays and supports up to 4 million logical volumes (equivalent to tape cartridges), doubling the capacity of the previous model. For horizontal scaling, the TS7700 can be connected in a grid of up to eight systems.

The VTL also complements IBM Z's "pervasive encryption" architecture, a transparent and consumable approach to enabling data encryption of data in-flight and at-rest that simplifies and reduces the cost associated with information security and regulatory compliance. (See below.) This encryption can be used when writing data to each volume and when data is sent between grid members.

These days, VTLs provide an effective means to host advanced data protection services. They can also provide a bridge to cloud services, or to local tape resources, and can support both the IT planner's recovery strategies for limited (application or hardware error- or malware attack-related) interruptions and more wide ranging (facility or milieu layer) catastrophes.

Of course, it helps if the rest of the infrastructure has data protection features built in so that protection can become truly pervasive. IBM is certainly approaching data protection this way.

The latest z14 mainframe features cryptographic functionality embedded on every processor core, called Central Processor Assist for Cryptographic Function (CPACF). CPACF makes applying encryption to sensitive data seven times faster than unassisted encryption methods, and the application transparency of CPACF makes it easy to use with all workloads. CPACF is where IBM's pervasive encryption approach begins, ensuring that data will be safer when traversing networks and fabrics (e.g. in-flight)[7].

At its storage destination, which in IBM shops tends to be a DS8800 family storage system, data is encrypted at-rest by leveraging special functionality embedded on the DS array controller. The story doesn't end there: the DS8880 continues the pervasive encryption of data in-flight through its system-to-system and system-to-cloud data transfers, as well.

Only the IBM DS8880 delivers "six nines" of availability by enabling the automatic replication of data between 2, 3 or 4 sites with all paths – including FCP, FICON and z High-Performance FICON (zHPF) channels -- encrypted. This technology supports data moves between DS systems and between the DS8880 and the IBM VTL, the TS7700. The IBM DS8880 also supports encryption of data in-flight for hybrid cloud environments by providing encryption protection for client and corporate data moving to multi-cloud environments using IBM's Transparent Cloud Tiering.

COMPREHENSIVE DATA PROTECTION IS WITHIN REACH

Taken together, IBM's pervasive encryption raises the bar for data security by encrypting all data movement and data storage platforms in a unified way. For companies that are not ready to upgrade to the latest capabilities across their entire infrastructure, there are many ways that the enhanced data protection capabilities enabled by IBM can be added to existing infrastructure. However, when the time comes to refresh older infrastructure, IBM's pervasive encryption technology should be a key consideration in platform choices.

The simple truth is that IT managers no longer have the time and resources to operate multiple disparate technologies in the hopes of achieving resiliency and data protection in their environments. With IBM's pervasive encryption approach and their best-of-breed tape, VTL, storage and processor technologies, a much simplified and highly effective data protection capability can be brought to bear that will help IT managers solve their data protection challenges and meet increasingly demanding regulatory and legal mandates around data protection and data privacy.

[7] Source: Pervasive Encryption, Achieving Security and Business Gains. IDC, October 2017 https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=ZSL03487USEN

ABOUT THE DATA MANAGEMENT INSTITUTE

This is a paper from the Data Management Institute LLC, a membership-driven organization for those who create, store, protect, preserve and secure digital information assets. Membership in the Data Management Institute is free at www.datainstitute.org.