CONTENT PROVIDED BY

Redmond

THE INDEPENDENT VOICE OF THE MICROSOFT IT COMMUNITY

SPONSORED BY

DELL Software

Data Protection Trends—Evolving from Data Protection to Data Resiliency

BY GREG SCHULZ

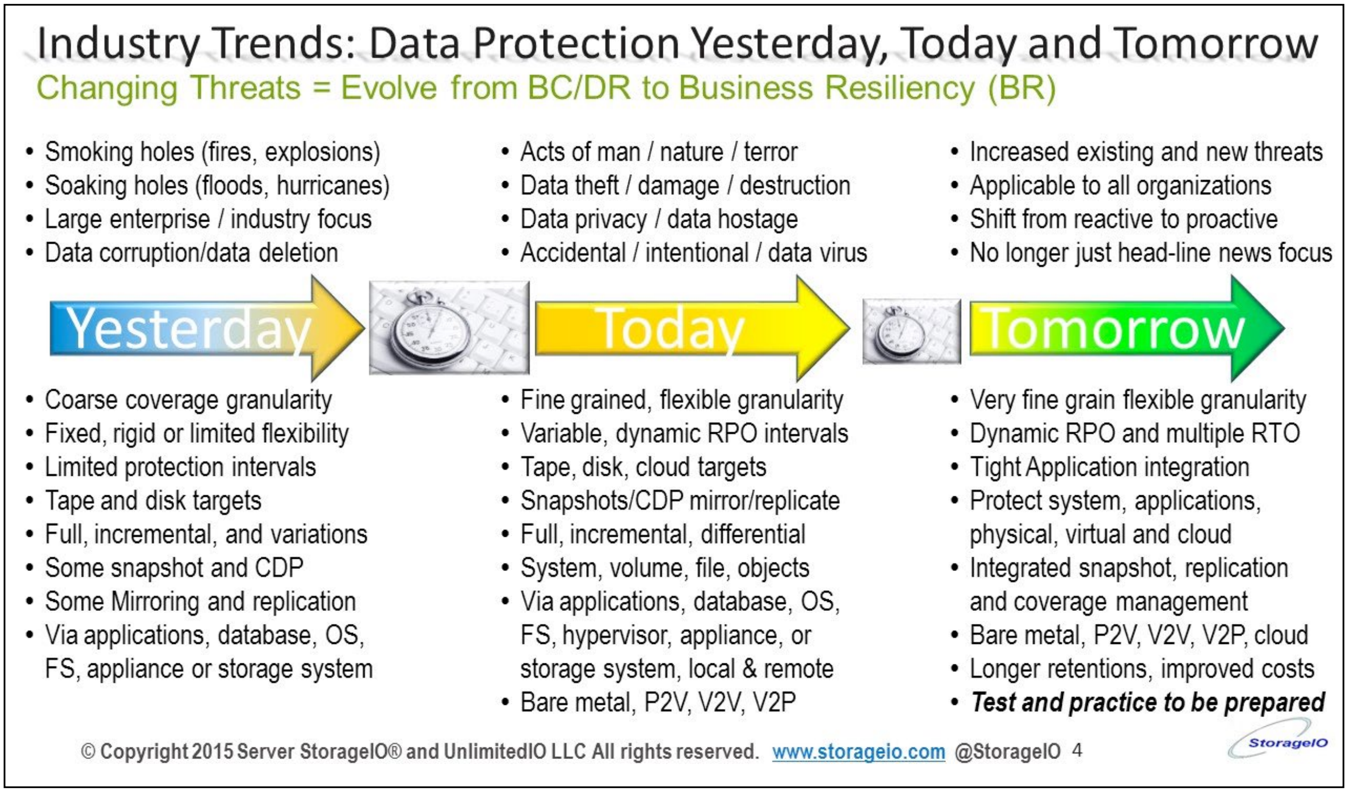
1105 MEDIA

When it comes to data protection, there are several trends that tend to get mentioned on an annual basis. Recurring trends include data protection evolving and converging from:

- Silo focus of archive or backup/restore to broader coverage and application centric
- Backup/restore to disaster recovery (DR) to business continuance (BC)
- DR and BC to business resiliency (BR) and always on accessibility
- Server, storage, device, volume or system to application aware focus

This includes expanding to proactive, upfront when data is stored versus reactive, after the fact data protection as part of enabling business resilience.

To put this in perspective, particularly when looking forward, it helps to look back. **Figure 1** shows various data protection industry trends, issues, concerns, threat risks, technology and techniques from yesterday, today and tomorrow.



Industry Trends: Data Protection Yesterday, Today and Tomorrow

Changing Threats = Evolve from BC/DR to Business Resiliency (BR)

- Smoking holes (fires, explosions)
- Soaking holes (floods, hurricanes)
- Large enterprise / industry focus
- Data corruption/data deletion

Yesterday

- Coarse coverage granularity
- Fixed, rigid or limited flexibility
- Limited protection intervals
- Tape and disk targets
- Full, incremental, and variations
- Some snapshot and CDP
- Some Mirroring and replication
- Via applications, database, OS, FS, appliance or storage system

- Acts of man / nature / terror
- Data theft / damage / destruction
- Data privacy / data hostage
- Accidental / intentional / data virus

Today

- Fine grained, flexible granularity
- Variable, dynamic RPO intervals
- Tape, disk, cloud targets
- Snapshots/CDP mirror/replicate
- Full, incremental, differential
- System, volume, file, objects
- Via applications, database, OS, FS, hypervisor, appliance, or storage system, local & remote
- Bare metal, P2V, V2V, V2P

- Increased existing and new threats
- Applicable to all organizations
- Shift from reactive to proactive
- No longer just head-line news focus

Tomorrow

- Very fine grain flexible granularity
- Dynamic RPO and multiple RTO
- Tight Application integration
- Protect system, applications, physical, virtual and cloud
- Integrated snapshot, replication and coverage management
- Bare metal, P2V, V2V, V2P, cloud
- Longer retentions, improved costs
- ***Test and practice to be prepared***

Figure 1. *Data Protection Yesterday, Today and Tomorrow*

There's a growing awareness that data protection means more than just security (logical and physical) or archiving (regulatory and non-regulatory). Physical security means barriers and locks on doors or other ways of gaining access to equipment and application interfaces. This includes keyed locks, digital access devices (ID cards, biometric eye and fingerprint readers), ande associated, identity-access management (IAM) and logging. In addition to physical security, logical security includes encryption of data at rest as well as in-flight or when being moved (onsite and offsite), two-stage authentication and IAM controls that enable productivity without weakening data protection.

Existing and increasing threat risks are driving the rethinking and expanded awareness of data protection.

Existing and increasing threat risks (those things that that put your data, information and applications at risk) are driving the rethinking and expanded awareness of data protection. Traditional threat risks include:

- Acts of nature acts of man, accidental or intentional
- Smoking hole (e.g. fire, explosion, earthquake, accidental or intentional)
- Soaking hole (e.g. flood, hurricane, tsunami, ice storm, tornado, acts of nature)
- Power outages or other environment related incidents
- Theft of data or assets, damage or destruction or other loss

Evolving threat risks and other drivers to enhance data protection include:

- Increased use and dependence on mobile, BYOD[1] and IoT[2] devices
- Reliance on cloud (public and private) services
- Data and security breach including theft, exposure, and damage
- Data sovereignty and privacy across different geopolitical borders
- Regularity compliance and other mandates
- Internal and external threats, intentional or accidental

There's plenty of discussion in the industry about hardware, software, server, storage, I/O networking, virtualization and other convergence topics. Data protection is also in those conversations and will become a more popular topic (trend) from technology, people and

[1] BYOD = Bring Your Own Device such as laptop or mobile phone platform)

[2] IoT = Internet of Things various devices that can generate, process or consume data

processes perspectives. Conversations will also center on convergence protecting converged and hyper-converged, as well as hybrid-cloud and software-defined virtual data infrastructures. Other conversations will focus on how to use converged solutions as a platform or enabling technologies to protect non-converged environments. These will range from purpose-built backup and data protection appliances to the cloud and virtual editions among other technology tools.

A barrier, however, to convergence is not necessarily the technologies, rather organizational boundaries. Part of enabling business resilience involves cross-organizational convergence spanning people, processes, practices, policies and how products or services are leveraged to protect applications and data. Often, the focus of data protection is on the data. However, applications, their configuration settings, meta-data and other resources also need to be protected and preserved.

Protection copies, as well as protection systems, also need to be safeguarded against attack.

This means that a resilient business or organization needs to have application and transactional consistency, as well as crash consistency to protect against different threat risks. For example, crash consistency protects at a broader level in case a system, server, storage appliance or drive fails enabling restoration to the last snapshot, copy, sync or backup. Transactional consistency is more fine-grained providing consistency and data integrity to the point and context of what they application was doing at a given point. Application and transaction consistency enables restoration or resumption up to the point of where a file, record, transaction, or other work being done was performed such as a database consistency point.

In the past, a primary data protection focus has been to enable BC or DR, or restore a system, application, database, volume, table, file or another object when needed. Threat risks included software, network, hardware device, system or data center failure, data corruption and deletion. Other threats have included virus, malware, and theft of data, among other intrusion risks. Today, those and other threats include damage to not only data, but to applications and their software-defined data centers and networks to cause destruction, chaos, and mayhem. This means that not only primary data can be a target. Protection copies, as well as protection systems, also need to be safeguarded against attack.

With more data (the volume and size, as well as quantity) being generated, processed and stored that requires protection for a longer time, new techniques are needed to leverage existing and emerging technologies more effectively. With the traditional approach of scanning files, objects, folders, directories, shares and other items to see what changed (needing protection), or detecting errors, verifying consistency and performing virus checks is becoming more difficult. Faster servers, I/O networks and flash SSD storage along with more efficient software algorithms (tools) help to speed things up for now.

However, at some point, more time will be spent scanning or checking to see what changed. Change block (or object or file) tracking helps, as does data footprint reduction (DFR) including compression, dedupe, and single instance storage. Change tracking along with DFR implemented as close to the source of where data is created or ingested is part of the solution to reduce what gets sent and stored downstream. Likewise, the current focus of downstream change tracking and DFR at the target destination will continue to be leveraged, working with source-side optimization. The net result should be more effective data being stored and protected but with less overhead and complexity, meaning cost savings while enabling growth.

Another trend is to implement protection based on policies when the data is created or stored.

Non-volatile memory (NVM), including flash SSD, is in your future, if not already here, both for primary storage and as part of your data protection environment. Even with more adoption of all-SSD arrays, all-flash arrays and hybrid storage, there is still a place for hard disk drives, given their increasing capacity, reliability and low cost for dormant data. When paired with NVM or flash SSD, along with some software-defined storage management tools, you can protect more data in less time at lower cost. Also, keep in mind that a bit of NVM or flash SSD storage in the right place can have a big impact; for example, as part of implementing snapshots and rapid clones, holding meta-data or indices and journals.

Part of enabling up-front data protection involves tighter application integration, meaning coordination with how and when copies are made for transactional and crash consistency, as well as for monitoring. Lower-level tools need to gain insight and awareness higher up in altitude (that is, in the software stack layers) to coordinate with applications on what to protect, when and why. This also means an increase in metadata, that is, data about the data, applications, and systems, along with policies, configurations and history information, all of which needs to be protected. Speaking of which, when was the last time you verified that your protection environment itself is protected and can be recovered? Now might be a good time to test that, however, be careful not to cause a disaster in the course of   trying to prevent one.

Another trend is to implement protection based on policies when the data is created or stored. For example, when a file is saved, email received, or another activity occurs based on policies, secure copies are made (and kept track of) in various locations for accessibility and durability. Also, where needed based on policies, archive copies are made, reducing the amount of scanning required to see what changed and needed protection. Protection policies also come into play when policy management tools that leverage insight, awareness and analytics can process secure disposition or deletion of data, including protection copies. This is where copy management ties in with data protection management along with data protection analysis tools.

The keys to clouds are leveraging them as another resource tier, applying best practices and managing controls.

Clouds, virtual- and software-defined technologies, tools and data infrastructure, along with the applications they support, need to be protected against various threat risks. Likewise, clouds, virtual and software-defined technologies, along with physical hardware are resources for protecting existing, as well as emerging environments.

Having insight and awareness or information about your environment, including what is being protected, how often and how well, the number of copies, versions, retention, disposition and other information is important. The importance of having insight is to enable making manual or automated decisions, implementing and enforcing policies. Without good insight and information, you risk flying blind or operating in a less-than-effective way. Think of it this way: a well-run factory has good processes, procedures, policies, plans, people, equipment and metrics that indicate the quality of products being produced. If a datacenter is an information factory, shouldn't you have good information on how well the product (information) is being produced, including the associated services being used or provided (such as data protection)?

The keys to clouds are leveraging them as another resource tier, applying best practices and managing controls. If you can identify your concerns, then they can be addressed or worked around. Nobody should be able to force you or your organization to go to a cloud, though there might be a peer or other types of pressure to do so. What works for one organization might not apply to yours or others. Keep in mind that everything is not the same in different datacenters, even across common industries or focus areas. However, there are similarities that can be leveraged.

Cost is a common concern with data protection, and the past trend has been to look for areas to cut capital expenditures (that is, Capex) or operating expenditures (that is, Opex). The challenge with looking for areas to cut costs is the risk of compromising the quality of service or protection layers or ending up with accidental data protection architecture that add complexity.

Trust, yet verify data protection processes, procedures, policies, people and products.

Keep in mind that complexity adds costs, particularly on an on-going or operating basis. While there is a focus on up-front cost cutting, sometimes the solution can be an upfront investment that ends up saving more over time when a return on investment, return on innovation or total cost of ownership analysis is done.

What this means is that by removing complexity without compromising service or data protection, you can end up saving costs, even if there is an up-front cost. Simply put, sometimes you have to invest a little to save a lot over time. Note that investment can be a combination of the budget as well as personal time and skills development.

Some trends and tips to keep in mind for 2016 (and beyond) include:

- Trust, yet verify data protection processes, procedures, policies, people and products.
- Invest in your data protection strategy, plan and implementation.
- Everything is not the same in the datacenter though there are similarities.
- If everything is not the same, do you need to treat and protect everything the same?
- Look for and remove complexity and you will likely also reduce costs.
- A strong defense has a good offense to counter, detect and deter threat risks.

What this all means is be prepared; it's not if, rather when, something will happen to disrupt your organization productivity. Whatever it happens to be, chances are it may not be a headline-news making event. However, if left uncontained and not isolated or protected, (e.g. the it or what is a threat to your applications and data) can cascade or snowball into a rolling disaster that could cause your organization to be in the headline news or the talk of social media. R

Greg Schulz is Founder and Sr. Advisory Analyst of independent IT advisory consultancy firm Server StorageIO and UnlimitedIO LLC (StorageIO). He has worked in IT for electrical utility, financial services, and transportation firms in roles ranging from business applications development to systems management, architecture, strategy and capacity planning. Schulz is the author of the Intel Recommended Reading List books, "Cloud and Virtual Data Storage Networking" and "The Green and Virtual Data Center" (CRC Press) and "Resilient Storage Networks" (Elsevier). Greg is a Microsoft MVP and six-time VMware vExpert. Learn more at storageio.com, on his blog at storageioblog.com and on Twitter @StorageIO.