IBM SPECTRUM COPY DATA MANAGEMENT DRIVES IT MODERNIZATION

NOVEMBER 2016

THE NET NET Modern IT organizations are facing explosive data growth. This growth is driven by the multitude of data copies needed to run and protect a company, including full backups, incremental snapshots, replication for disaster recovery (DR) and copies to support business operations such as agile development and test environments, data retention for compliance, business insight and data analytics, and application sandboxes for training and product evaluations.

Yet one of the real copy data problems stems from the fact that a large percentage of data copies sit idle and add no business value because the data copies are not tracked and are not actively managed since most data copies live in separate stacks of uncoordinated infrastructure. This lack of data management often leads to duplicate data, storage silos, unnecessary administrative overhead and unauthorized data access. Add to this the fact that most business groups (application development, QA, finance and so on) have their own budgets, applications and resources, and there's a good chance most companies have rogue data copies, redundant data protection systems and overburdened IT administrators.

The other major issue is slow access to copies of data. As application teams, such as developers and QA groups become active consumers of data copies, companies need to move away from traditional "request and wait" IT processes and move towards "automated self-service" IT platforms that enable greater agility and speed.

Fortunately, today there are software and infrastructure vendors that offer copy data management (CDM) solutions designed to tackle the enterprise copy data challenge in a systematic and unified fashion. These CDM products offer the backup and recovery capability of data protection applications and offer development groups and other operational teams self-service management of application environments, which enables autonomy as well as agility and resources on demand. In addition, these solutions provide IT the centralized data management needed to provide continuous visibility into where data lives, who has access to it and how it's being used. As a result, IT achieves better business alignment and resource optimization for greater efficiency and cost savings. And importantly, IT maintains centralized oversight to ensure the data center environment delivers rock solid reliability, enterprise-grade security and regulatory compliance.

Simply put – CDM is an IT modernization technology that automates data management, delivering the self-directed data access, orchestration and visibility needed to transform business and IT operations. And now leading CDM solutions offer cost-effective DR and scalable data archival by offering integration with public clouds, such as IBM Cloud (SoftLayer and Bluemix), Amazon Web Services (AWS), Microsoft Azure and Google Cloud.

Building on their strength in storage and cloud infrastructure, IBM recently launched IBM Spectrum Copy Data Management, a powerful CDM solution that fully leverages IBM's broad range of storage solutions and the IBM Cloud. This powerful combination puts IBM Spectrum Copy Data Management

among the top CDM solutions by offering end-to-end CDM that delivers easy operation, full copy data automation, excellent storage efficiency and utilization, robust backup and recovery, cost-effective hybrid cloud DR and scalable data archival. These are all key capabilities needed to consolidate the management of all secondary use cases under one umbrella.

In this paper, we will look at the challenges of IT modernization and the key elements of CDM. We will also provide a detailed overview of IBM Spectrum Copy Data Management.

BRIDGING THE GAPS – THE CHALLENGES IMPEDING IT MODERNIZATION

A recent survey conducted by Taneja Group showed that enterprise primary storage needs are best met by virtualized storage systems that provide a single storage pool, a common set of data services and a common interface to manage storage resources. These same storage system characteristics can certainly benefit secondary data environments, but unfortunately secondary workloads are largely siloed since companies often use several applications for data backup and DR and usually have no centralized management of data copies used by various application teams. As a result, companies have major gaps highlighted below that significantly inhibit their progress towards IT modernization.

Gap #1 – the difference between data copies companies have and data copies they need – We estimate that roughly 80% of all customer data is in secondary storage, with a growth rate of 40-50% per year. Companies typically can't explain why data is growing so fast and they usually don't fully understand how much data they have, where data lives and how often data is being accessed. This means that the majority of secondary data is sitting in secondary storage, opaque and rarely, if ever, accessed, representing a significant cost with no corresponding benefit as well as creating potential security issues and putting businesses at risk of being fined for regulatory non-compliance.

Gap #2 – the difference between data protection and data management – Standalone data protection solutions offer the backup and recovery capabilities needed to ensure IT resilience, but these applications do little to deliver the data management needed to help companies accelerate business operations and extract business value from information contained within secondary data copies. This is because data protection solutions lack self-service and data lifecycle automation required for application teams to easily spin up and spin down application environments. And traditional back and recovery applications lack the discovery, indexing and analytics functionality needed to provide a complete and in-depth understanding of the secondary data environment.

Gap #3 – limitations created when storage administration inhibits application team agility – Development groups, finance departments and product teams are avid consumers of secondary data copies needed for testing new releases, performing data analytics and creating sandboxes for demos and training. Moreover, competitive pressures to go faster and the "consumer-like" craving for greater convenience drives these teams to want agility and control. Meeting this need requires self-service and resources on demand, which IT is often inclined to provide given their goal for better business alignment and their desire to get relief from repetitive tasks. Unfortunately, IT's need for control often rules the day because they must ensure rock solid reliability, enterprise security and regulatory compliance. This means application teams are usually resigned to generating numerous IT tickets over several hours and sometimes days to have IT manually provision the resources they need, not to mention that fact that IT professionals remain burdened with administrative tasks.

Gap #4 – the lack of continuity between existing IT operations and cloud environments – It is common knowledge that public cloud infrastructure can offer significant OpEx and CapEx savings and provide other benefits such as usage-based pricing, resources on demand and cost-effective off-site storage for archival and/or DR. However, for these benefits to be realized, IT must have the ability to seamlessly and intelligently integrate public clouds with existing operations to provide the security, visibility and control needed to guarantee the required service levels when secondary data or entire applications stacks are hosted in public or hybrid cloud environments.

KEY ELEMENTS OF COPY DATA MANAGEMENT

Since day one, the adoption of copy data management solutions has been driven by their ability to surpass traditional data protection applications by offering backup and recovery capabilities plus their ability to efficiently support secondary data use cases, such as DevOps, by centrally managing virtual copies that point to a single master instance. Building on this early success, leading copy data management solutions have evolved to provide greater operational efficiency and business agility by adding self-service capabilities, full workflow automation, data analytics and public cloud integration. Below is an outline of four functional categories and corresponding key elements that we believe are essential in modern copy data management offerings.

Category	Key Elements	Description
Data Protection	Backup, replication and recovery	Modern data protection requires data backup and replication and rapid recovery within physical, virtual (VMware vSphere, Microsoft Hyper-V) and public cloud (IBM Cloud, AWS, Azure) environments. Policy-based multi-site replication, application awareness and flexible restore capabilities are essential for granular file access , application consistency and DR.
Storage Efficiency & Scalability	Storage space reduction	Includes support for storage space reduction techniques such as thin provisioning, compression and deduplication. Also includes support for virtual copies that point back to a single master copy while supporting redirect-on-write, so that only changes to the original data requires additional storage space.
	Software-defined storage approach	CDM solutions should integrate with virtualized storage environments to support mixed workloads within a single storage pool that unifies heterogeneous flash, disk, tape and object storage devices and leverage common copy services that automatically create, optimize, protect and migrate copy data. An example of optimization is the ability to automatically move frequently accessed data to high-performance flash storage.
	Cloud and object storage support	Because enterprise unstructured data can easily reach hundreds of terabytes or petabytes, copy data management solutions should support the ability to backup and migrate inactive data to scale-out object storage and public cloud environments designed for data archival and cost-effective DR.
Copy Data Services	Self-service	The ability to access data copies in a self-directed way without IT administrator involvement; including finding, provisioning and managing resources, which requires extensible workflow templates, a marketplace, role-based access and integrated lifecycle controls.
	Policy-driven orchestration	The automation of the copy data lifecycle including creating and modifying data workflows (or templates) and full automation or the ability to spin-up and down of the entire application environments. Specifically, this means creating policies for provisioning VMs and data copies (selecting a data source, a copy method and a target location on premises or in the cloud), and setting network parameters, refresh frequency and the retention period and cleaning up copies and VMs as needed.

Copy Data Services (Cont'd)	Extensibility	Integration with applications, virtual environments, third party tools and public clouds. Application-consistent snapshots that guarantee a usable backup by flushing any outstanding writes are essential when working with transactional systems like SQL Server, Oracle or SAP databases. Modern data centers are virtualized, so integration with major hypervisors (vSphere and Hyper-V) is also essential. And since DevOps is an important CDM use case, plug-ins for DevOps tools such as Puppet and Chef are needed to manage application environments using native commands. In addition, cloud support is critical for companies to backup, replicate and migrate data to the public clouds, such as SoftLayer, AWS, and Azure, while maintaining control and orchestration through a centralized CDM platform.
Copy Data Analytics	Cataloging	Automatic discovery and registration of all infrastructure resources (hypervisors and storage devices) to build a complete understanding of the secondary data environment.
	Data visibility	A comprehensive catalog with metadata enables insight into all physical and virtual infrastructure resources so administrators can quickly determine where copy data lives and when data is accessed. An analytics module should provide predefined reports, dashboards with filters and the ability to create custom reports tailored to specific needs. Search is another essential capability. Search filters can help administrators quickly find objects that match certain criteria, such as VMs or volumes in a certain location or data copies greater than a certain size.

HOW IBM SPECTRUM COPY DATA MANAGEMENT DRIVES IT MODERNIZATION

As companies seek to modernize their IT environment, they strive to lower storage costs, accelerate business operations and reduce corporate risk. And since the vast majority of data resides in secondary data copies, many businesses have discovered that improving copy data management (CDM) is important to achieve these objectives. As a result, enterprises are moving from standalone data protection solutions towards CDM platforms that offer robust backup and recovery capabilities plus self-service, full copy data lifecycle automation and better data visibility and compliance.

IBM Spectrum Copy Data Management offers all these essential CDM capabilities and IBM Spectrum Copy Data Management leverages IBM's complete range of storage solutions and the IBM Cloud as well as AWS. This powerful combination puts IBM Spectrum Copy Data Management among the top CDM solutions by offering end-to-end CDM that delivers easy operation, full copy data automation, excellent storage efficiency and utilization, robust backup and recovery, cost-effective hybrid cloud DR and scalable data archival. These are all key capabilities needed to consolidate the management of all secondary use cases under one umbrella.

A core strength of IBM Spectrum Copy Data Management is copy data automation – a capability that is vital to accelerate DevOps and other critical business operations by bridging the gap between applications teams, such as DevOps and business analysts, and storage administration. We found the completeness of the copy data automation functionality impressive, including the ability to fully automate SLA compliance as well as the process of building, deploying and spinning down full application stacks and the required infrastructure (VMs, networking and storage) including accessing the system of record (data copy), both on premises and in the cloud (IBM Cloud, AWS and Azure). IBM Spectrum Copy Data Management accomplishes this through the use of a well-developed

REST API, integration with Puppet, Chef and Bluemix, a self-service portal/marketplace, IT controlled workflow templates, role-based access control and policy-based orchestration. Another key benefit of copy data automation is automated DR testing, which transforms DR testing from a disruptive full or half-day process to a one- to two-hour automated workflow, and the DR testing is done in a fenced VMware environment to ensure the testing activity doesn't impact the primary storage environment. The importance of operationalizing DR processes should be emphasized. In a recent Taneja Group survey, 40% of companies indicated that manual, error-prone DR processes and the inability to test DR plans often jeopardize a company's ability to rapidly and reliably recover from extended outages resulting from disasters or other major disruptions.

IBM Spectrum Copy Data Management's easy operation is apparent from the get-go. It installs as a VMware virtual appliance that can be deployed in a matter of minutes. And it catalogs storage and VMware environments without the need to deploy agents. The web-based management portal and an easy to navigate, step-by-step user interface, see screenshot below, make it easy for storage admins to create policy-based workflow templates for provisioning VMs and storage. Once workflow templates are published, application teams can select a workflow template that best meets their service level requirements. For example, the finance team might select a gold template, with policies that ensure a frequent data refresh rate, long term data retention and application consistency.
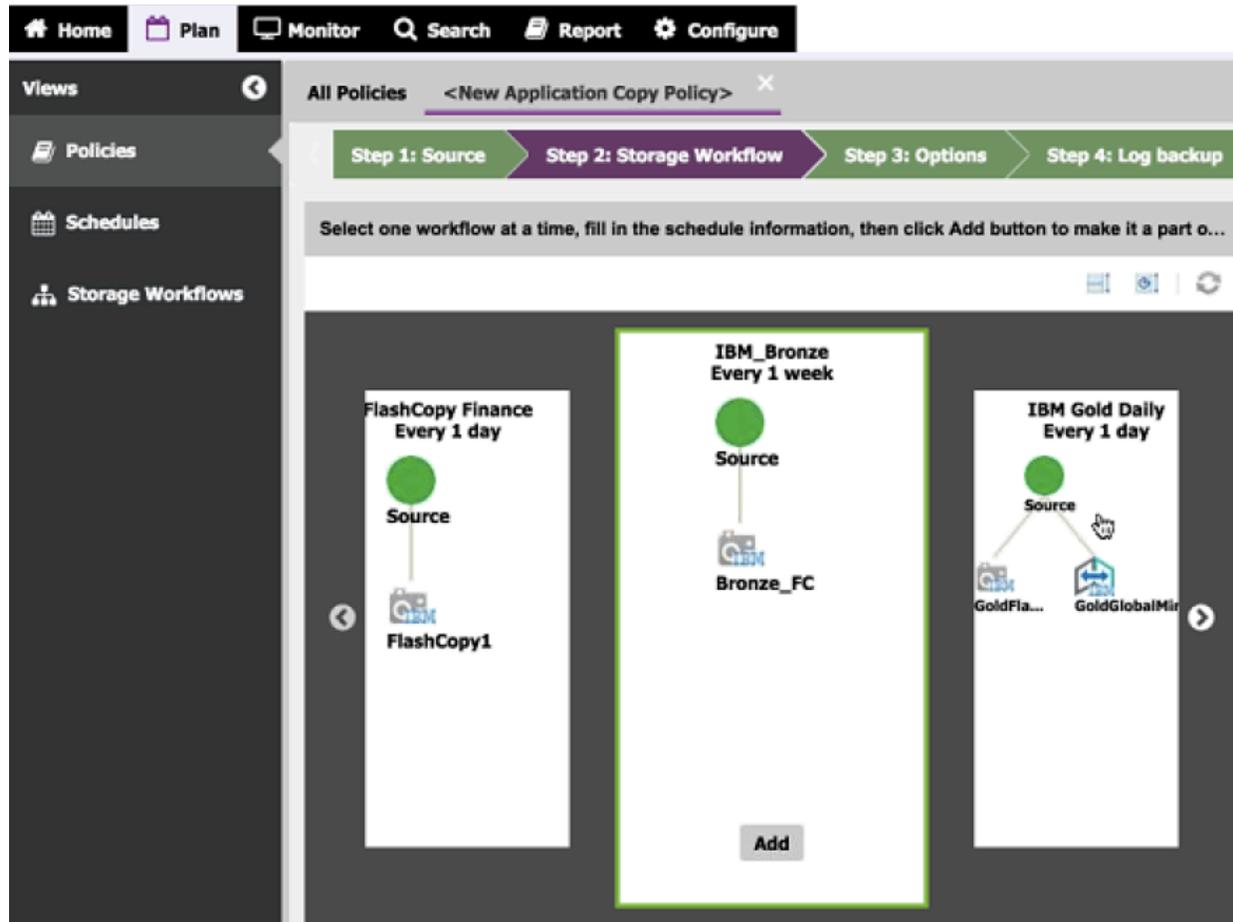
Data visibility, including cataloging, indexing, search and reporting/analytics is also a very important given the need to make all data copy objects trackable and searchable as well as enable analytics and compliance. IBM Spectrum Copy Data Management automates the discovery and registration of infrastructure resources to create a rich metadata catalog of VMs and storage devices. Each resource object in the catalog is added to an index to improve search performance. Examples of objects in the

index are volumes, files, snapshots and virtual machines. IBM Spectrum Copy Data Management also provides predefined reports and the ability to create custom reports tailored for specific needs. Examples of predefined reports include insight into storage utilization and data protection reports that identify unprotected objects to help ensure compliance and reduce risk. Search is also available to help administrators quickly find objects that match certain criteria, such as VMs or volumes in a certain location or data copies greater than a certain size or data accessed during a certain period time.

Finally, IBM Spectrum Copy Data Management leverages IBM Spectrum Protect for data backup and recovery. The ability to leverage the enterprise-grade snapshot and replication capabilities of IBM Spectrum Protect is important, since IT professionals make it clear that the benefits of centralizing all copy data management under one roof are only compelling only if they don't have to compromise on critical data protection functions. IBM Spectrum Copy Data Management's Copy Workflows allow users to create point-in-time application consistent snapshots and replicas of VMware applications or stand-alone storage volumes. If the primary storage is IBM storage, IBM FlashCopy is used to perform snapshots. FlashCopy is a snapshot method that requires minimal space for each copy because only new data writes use additional storage space. This approach helps IT provide copies of applications to multiple users quickly and efficiently. If the primary storage is not IBM storage, VMware vSphere Storage APIs - Data Protection (VADP) is used to copy the virtual machines to a target IBM storage volume. VMware VADP is a software framework that enables block level incremental backup of VMware virtual machines. Replication is performed using IBM Global Mirror Change Volume protection to maintain identical data in both the source and target volumes. Copy data policies define recovery point objectives (RPOs) that determine frequency and retention of a FlashCopy. Copy data policies also enable instantaneous application recovery by automatically locating and mounting the correct FlashCopy depending on the particular point in time that is selected. For application stacks with several VMs, policies can be used to bring up applications in the required sequence.

A POWERFUL COMBINATION – CDM, IBM STORAGE FAMILY AND IBM CLOUD

As the heart of IBM Spectrum Copy Data Management is Catalogic's proven software that utilizes a highly-effective "in-place" CDM approach. A key advantage of the in-place CDM approach is the ability to use existing storage infrastructure. For IBM storage customers, this means they can fully leverage IBM's impressive storage portfolio, including Spectrum Virtualize (formerly SAN Volume Controller), Storwize hybrid arrays, and FlashSystem V9000 as well as other storage arrays, such as NetApp for file storage. In the interest of brevity, we won't list all the advantages, but below are three options we believe highlight synergies created when IBM Spectrum Copy Data Management is used in conjunction with IBM storage and IBM Cloud.

IBM Spectrum Copy Data Management with IBM Spectrum Virtualize – Spectrum Virtualize is one of the industry's strongest storage virtualization platforms. When used in conjunction with IBM Spectrum Copy Data Management, Spectrum Virtualize enables the consolidation of mixed copy data workloads within a single storage pool that unifies heterogeneous storage devices. A key reason Spectrum Virtualize can deliver consistent performance across varied workloads is Easy Tier – IBM's auto-tiering capability. Easy Tier optimizes performance and cost-effectiveness by identifying more active data and moving it to faster storage (flash) – generating a 3X performance boost with only 5% flash, according to IBM. Multiple IBM storage systems leverage Spectrum Virtualize, including Storwize hybrid arrays, FlashSystem all-flash arrays and VersaStack for converged infrastructure.

IBM Spectrum Copy Data Management with FlashSystem V9000 – The use of CDM with all-flash enables companies to support multiple workloads with a single snapshot through the use of IBM FlashCopy. This is particularly important in the case of transactional workloads, such as the sharing of Oracle or Microsoft SQL Server database instances. Normally the I/O load generated by multiple

workloads tied to a single snapshot would degrade performance, but the FlashSystem V9000 offers the low latency required to make this next level of copy data efficiency a reality.

IBM Spectrum Copy Data Management with IBM Cloud – Hybrid cloud DR can offer significant savings through on-demand compute and the elimination of costs for a second datacenter. And importantly, IBM Spectrum Copy Data Management's automation and orchestration enables companies to leverage SoftLayer's elastic compute and storage while keeping centralized control. IBM Spectrum Copy Data Management also provides full data lifecycle control, so companies can spin workloads up and down as needed, minimizing cloud charges.

TANEJA GROUP OPINION

You can no longer tell application teams they must wait days for storage to be provisioned so they can get access to an application or database instance. The consumerization of IT and continuous application delivery initiatives have changed their mindset and teams will no longer tolerate the delay and inconvenience. You also can't tell the CIO that the company is non-compliant because all copy data is not properly accounted for and all copy data is not properly protected. Combine these two must avoid situations with a third reality, which is that amount of data and the need to access that data is exploding, and you quickly understand the tremendous motivation companies have to improve the management of their secondary or copy data.

Standalone CDM players saw this need a few years ago and started chipping away at the data protection market. These standalone CDM solutions provide a dedicated environment for copy data storage and management. However, a downside of the standalone approach is the added cost of copy technology dedicated to CDM.

As a leader in DP, IBM understands the evolution from DP to data management. Therefore, it's no surprise they are aggressively pursuing the CDM market. We knew it would be imperative that IBM's CDM solution fully leverage the power of their leading storage portfolio and the IBM Cloud. This made Catalogic's in-place CDM approach with its elegant control plane that normalizes CDM functions the ideal hub to unify their CDM offering.

It's clear to us that IBM Spectrum Copy Data Management is among the leading CDM offerings in the market and its powerful copy automation capability combined with IBM Spectrum Virtualize, one of the industry's strongest storage virtualization platforms, and IBM Cloud, an industry leading hybrid cloud, makes IBM Spectrum Copy Data Management a go-to solution for any company looking to consolidate the management of all secondary use cases under one umbrella.