Today's Data Protection Business Dilemma and Associated Challenges

By Ashar Baig and Christine Taylor

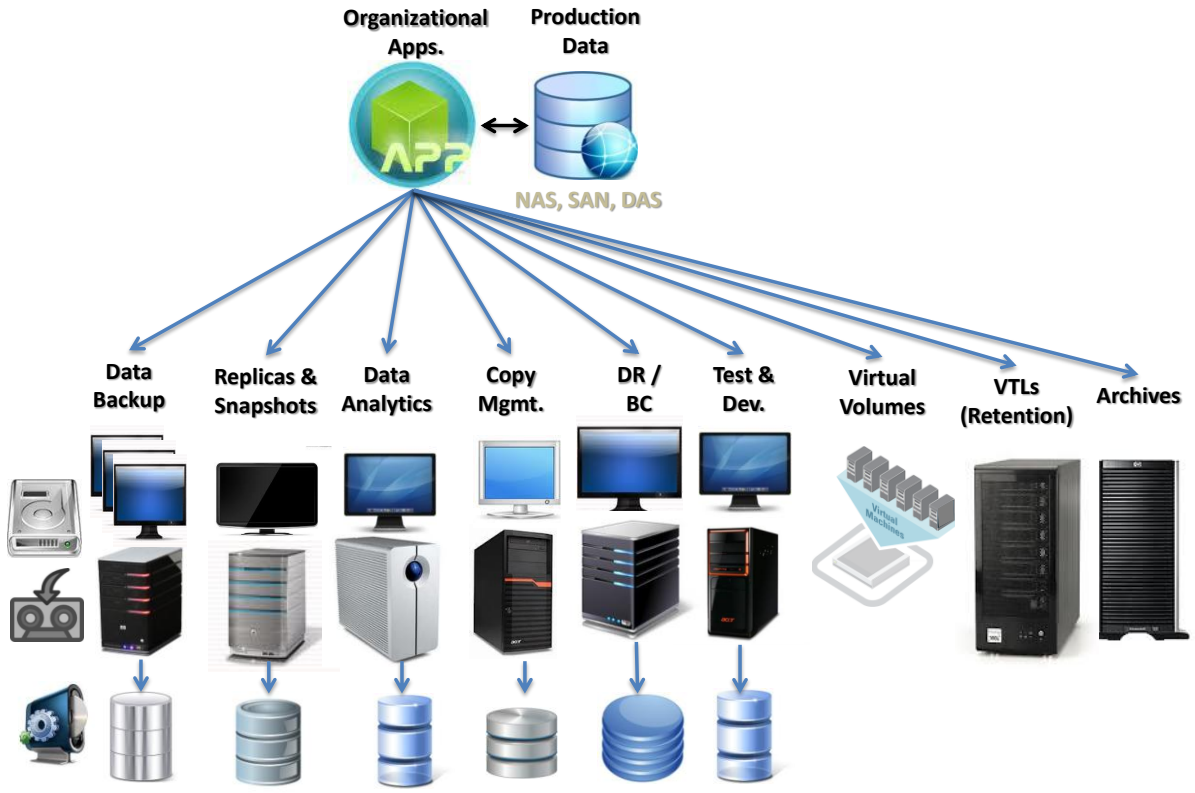
In our 2006 research paper, Taneja Group presented a vision to the industry introducing the concept of Continuous Data Technologies (CDT) (Taneja Group TECHNOLOGY IN BRIEF Continuous Data Technologies: A New Paradigm April, 2006.) The underlying issue was that data protection and the associated data management methods available to the market for the last three decades were broken and, in our opinion, required a complete overhaul.

Traditional data protection and data management involves making multiple copies of the production data for backup, replication, snapshots, mirroring, Continuous Data Protection (CDP), cloning, and more. These operations generate anywhere from 25 to over 100 mish mashed copies of the production data that are cumbersome to manage, resulting in storage silos that get littered across backup disks and tapes, Virtual Tape Libraries (VTLs), storage vaults, archive systems, snapshot repositories, virtual volumes, data analytics platforms, data clones and now even cloud storage environments.

This data deluge, also considered big data, causes organizations to purchase more storage hardware and related products to protect and manage the exponentially growing volumes of data copies across one or many physical locations. This untenable situation increases the secondary storage spend by a factor of five, or more, relative to the cost of high priced production storage.

Complications that arise from data sprawl include:

- Inability to backup data
- Network congestion
- Multiple expensive software licenses
- Cloud-lock-in



Traditional Data Management

CDT methodology can transform how businesses handle virtually any act of "copy creation." It conducts very granular data capture of production data creating a single data source, essentially *a fountain of life.* This allows administrators to create unlimited Any Point in Time (APiT) images to feed backup, recovery, replication, testing, development, support, archiving, migrations, disaster recovery, and point-in-time approaches to copy creation such as snapshots.

Five years ago, we predicted that the concept of CDT is radically more efficient, less complex, more seamless, and flexible while maintaining data consistency, increasing data availability, and providing longer data retention. It has the potential to radically transform the data center over the next decade.

A number of small, innovative vendors were trying to bring the CDT concept to life but for a variety of reasons they all failed to deliver. Mendocino Software was one. Revivio, whose assets were later bought by Symantec, was another. Perhaps a few fundamental technologies were missing at that point in time, causing their efforts to go a portion of the way to completion.

So what has changed over the last five years?

1. **Deduplication has made a real difference in the amount of data stored or transmitted off site.** It has improved backup and restore performance and

reliability. But the process surrounding data protection has remained unchanged. It has simply made an ugly child less ugly. While traditional deduplication appliances make data backups and restores faster and more reliable they still rely on traditional backup software which is built on fundamentally flawed backup and restore principles. The appliances are non-disruptive to the existing data protection infrastructure, reduce the backup storage footprint, but don't eliminate backup windows or restore times.

2. **Data reduction technologies have enabled IT to consolidate all their data copies that used to be on disparate media.** Disk became the single media where all copies are kept.
3. **Disk systems have become commoditized.** Just like the focus of CPUs have shifted from clock speed to energy consumption, the focus of storage systems have shifted to "$/TB" from "RAID level".
4. **Server virtualization has gained traction and has been deployed across companies of all size, across all industries.** This has played havoc on data protection, even while it has simplified DR in many ways. Special products have been created to protect VMs and the data associated with them. But once again the fundamentals of data protection have remained the same. These specially designed data protection products have simply eliminated the major issues related to using traditional products and have created yet another point solution creating data copies.
5. **Storage virtualization technologies have been incorporated across many storage products.** Virtualization is not limited to the storage array.
6. **Virtualization as a technology has become widely adopted and accepted.** IT has demonstrated a strong drive to virtualize IT operations and infrastructure.
7. **Applications and operating systems have become more manageable.** Modern interfaces allow for better information about correlation between data and metadata.
8. **IT is moving towards the OPEX-based service model.** Service providers may offer savings by replacing traditional capital expenditures and local maintenance.

Of course, alongside these technical developments data growth has continued unabated at an average 60% a year. Storage capacities have followed the standard Moore's law and doubled capacities every 18 to 24 months, while improvements in storage performance has pretty much remained constant.

So for all practical purposes, these advances have made life only incrementally better. Few fundamental changes have occurred to radically change data protection at the root level.

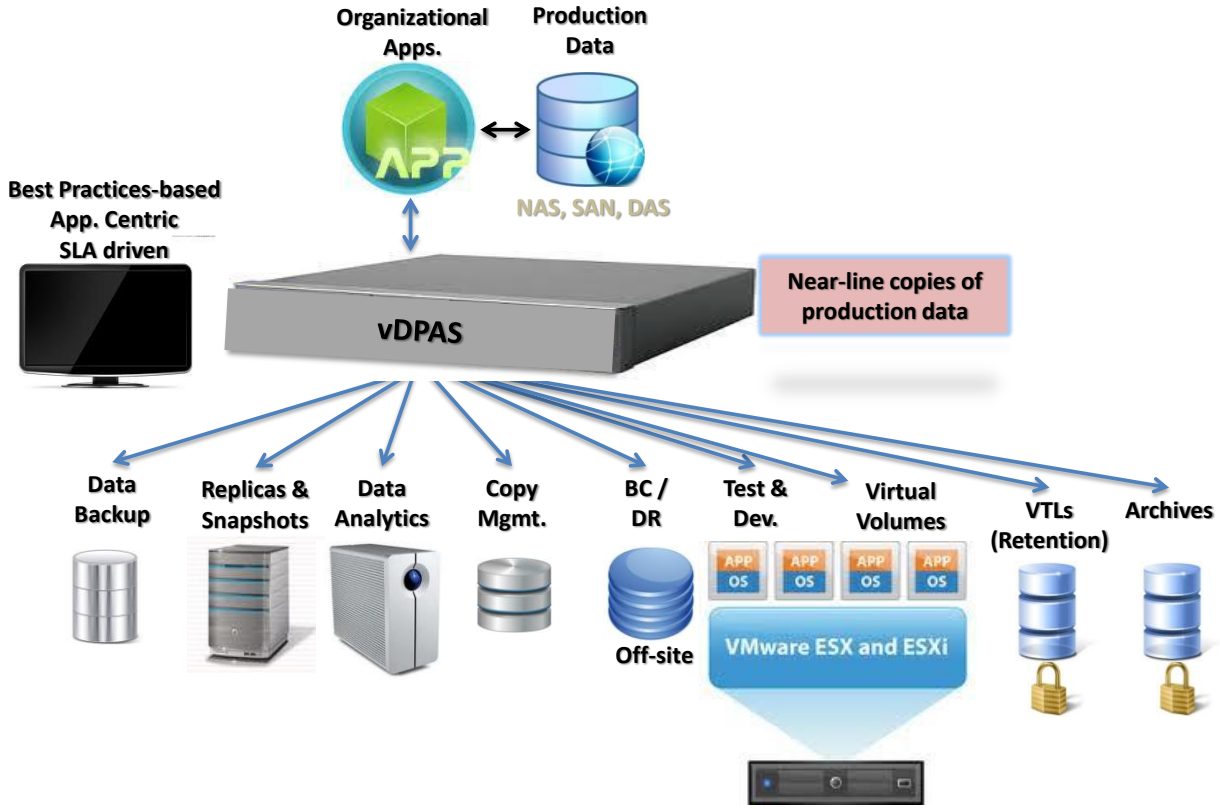
Enter vDPAS, pronounced "v-DPAS."

A New Category in Data Protection and Data Availability

The traditional data protection approach is challenged by a new breed of vendors with a bold new approach to data protection and availability. It introduces the concept of protecting multiple systems, environments (physical, virtual and cloud) and applications

in a single, highly efficient, virtualized storage pool – the *fountain of life* conceptualized by Taneja Group over five years ago.

This new approach to data protection and data availability is so radically different from the choices available in the marketplace today that it warrants its own category. Taneja Group calls the new category Virtualized Data Protection and Availability Storage (vDPAS), which administers SLA-driven data protection and data management across the entire production environment – regardless of application, system or deployment model.



vDPAS-Based Data Management

vDPAS-based products radically simplify data protection and associated functions and reduce storage-related costs by up to 90%. Cost savings apply across all environments whether they are located on virtual, physical or cloud data infrastructures. This is an application-centric, SLA-driven data management paradigm that is independent of server, storage, or network infrastructure.

How vDPAS Works

vDPAS-based products use a variety of APIs to access applications to make efficient copies of data. Using Change Block Tracking (CBT) to identify changes in the copy, relative to a previous copy, vDPAS creates a snapshot of changed blocks creating a historical time machine of application data. Current changes are kept in raw format for instant point-in-time recovery. Historical changes are deduplicated and compressed. vDPAS stores the unique objects and removes the snapshot from production storage. Users define their own

flexible SLAs per application, including options like asynchronous replication to a remote DR site. vDPAS' virtualized storage pool decouples storage from physical location; hence, storage can be local, remote over the WAN, cloud-based or a hybrid. Since, the changed blocks are generally very few, bandwidth usage is negligible.

Application-level, block-level, object-level or file-level restores range from very fast to instantaneous. vDPAS can immediately mount point-in-time copies in raw format for instantaneous recovery, or quickly rehydrate deduplicated and compressed data on demand. Recovery settings are policy-driven and based on specific application or system service level agreements (SLAs). By allowing applications to access their point-in-time copies from a single virtualized storage pool, all protected systems can restore immediately with minimal IT oversight. vDPAS also enables IT to instantaneously restore a primary environment from a DR site, and lets users efficiently create virtual volumes and clones for test and /development or support purposes.

Let's see how this would work in a data center that contains, for example, a SAP application using an Oracle database, virtualized servers using VMware, Microsoft Exchange servers, and several back-office file servers. Normally, this environment would have a variety of data protection applied – incremental backup for Exchange, VMware and file servers, daily full backups for SAP to a VTL, weekly full backups for the rest, frequent snapshot-based replication for SAP, Exchange archiving software that migrates 90-day-aged emails and attachments to the Storage Area Network (SAN). Storage includes an iSCSI SAN, the VTL and several Network Attached Storage (NAS) filers. The VTL replicates nightly to a remote DR site.

Challenging and confusing? You bet.
Dozens to over a hundred copies of each file? Impossible to meet your RTOs/RPOs?
Achieving SLA's a dream? That's right.
And it's quite common in a traditional data protection environment.

Now let's take a look at a data center that uses a vDPAS-based platform. It replaces the many separate data protection and availability operations that would normally flow to many different storage silos. The vDPAS solution unifies all of these operations into a single straightforward process. After the initialization of the data that captures the current state of the volumes, the following process applies:

1. IT schedules hourly snapshots of the critical SAP database.
2. vDPAS tracks changed blocks, takes a snapshot, and stores only the changes in raw format for immediate point-in-time recovery.
    a. Each hour it creates another scheduled snap and counts off the number of snaps.
    b. The 24th copy is considered the daily backup and is deduplicated, compressed and moved to vDPAS virtualized storage.
    c. In a week's time the 7th of these daily copies becomes the weekly backup.
3. Following this successful pilot run, IT then extends vDPAS to Microsoft Exchange and virtual servers.

4. IT chooses optimal SLAs for each application or system and launches the vDPAS process, which houses the tiny unique objects to the same virtualized storage pool as the database objects.

The results are exceptional. The platform, which simply slots into the existing infrastructure, eliminates redundant data and replaces it with unique de-duped copies and instant recovery. The resulting savings easily reach 90% by deferring primary and secondary storage purchases, replacing budget for multiple data protection products, and radically lowering ongoing maintenance related OPEX.

vDPAS Benefits

vDPAS enables the following benefits:

1. Replaces a slew of disparate point products while co-existing with current infrastructure and operations.
2. Extremely small footprint saves up to 95% over traditional storage footprints.
3. Enables near-zero backup windows.
4. Provides instantaneous restores (local and DR site) with a miniscule amount of data movement from primary storage to vDPAS.
5. Slashes bandwidth and network usage: WAN bandwidth reduces by a typical 80% or more, and LAN network usage typically lowers by 70% and higher.
6. Offers TCO reductions of 80%+ relative to legacy products.
7. Application-centric, SLA-driven.
8. Protects and restores any physical or virtual application granular (object level) backups and recoveries. Backups and restores can be initiated by end users or IT generalists.

vDPAS – Real Benefits for Real People

The benefits of vDPAS for businesses are powerful and real. Let's see how vDPAS plays out with some workgroups who are hardest hit by traditional data protection and management: backup administrators, test/development professionals, and MSPs providing cloud-based backup and storage.

- **Backup Administrators.**
  The Issue: Due to 60% year-over-year data growth, preferred backup windows are increasingly inadequate. IT spends big money on fast backup and restore solutions for Tier 1 applications, but Tier 2 applications can take weeks to restore. IT administrators spend a vast majority of their time managing backup environments with multiple data protection and disaster recovery products, complex operations, and numerous storage silos. Additionally, huge volumes of data and the ever increasing number of production data copies forces IT administrators to spend even more time provisioning, managing, and protecting than they have ever done before.

The growing cost of this copy data increases the cost of site replication for business continuity and consumes an increasingly larger amount of bandwidth each year.
The Solution: vDPAS eliminates this complex and expensive environment with near zero backup time and instantaneous recovery from a single virtualized storage pool across all applications and environments. The process is simple to administer, hence, IT generalists can manage the entire data protection lifecycle including SLAs, resulting in huge OPEX savings.

- **Test and Development.**
  The Issue: Production clones consume massive storage capacity and demand significant server, network, and IT resources. Test and development professionals do their best to keep the data demand down, but outdated or partial clones can yield less-than-stellar results. With vDPAS, test and development professionals can reuse the same stored data that is also used for data protection.
  The Solution: With vDPAS, the ability to clone large data sets without extra storage – and with considerably less network, server and IT support -- dramatically drives down the overall cost of the test and development environment. It also improves results since test and development can freely update production clones as needed.

- **Managed Service Provider (MSP) providing Cloud Backup or Cloud-based storage.**
  The Issue: MSPs must protect customer data to adhere to the SLAs stipulated in customer contracts. However, traditional data protection methodologies and the immense competition cut deep into MSP's profits. Customers also suffer because their ability to move large data sets to their MSP depends on the bandwidth the customer can afford.
  The Solution: With vDPAS, data copied at customer site becomes a mere trickle, which slashes their bandwidth requirements. The MSP dramatically lowers its cost of protecting customer data, while simultaneously improving their ability to protect and restore. This improves storage MSP's profits and enables them to offer premium services with stronger SLAs at an attractive price point. Incremental costs of offering new services beyond backup on-demand do not increase cost basis significantly for the service provider.

vDPAS is the Future – Now

For the past three decades businesses have tried to unsuccessfully tackle the issue of data protection and data availability, which is littered with storage silos and legacy methodologies. This data sprawl problem has been exasperated by continued data growth, now into the big data area. Organizations have tried to tackle this problem by throwing more hardware at it but that only resulted in increased CAPEX spend. In today's environment of economic uncertainty, this CAPEX-based model is unsustainable and can be deemed fiscally irresponsible. These data protection related capital expenditures can be best spent towards driving business innovation and competitive advantage.

vDPAS is a revolutionary new category of storage that is fundamentally different from the traditional approaches to data protection and is engineered to fulfill a dire market need – ushering in a new era in data protection – and it is available today.

Some legacy storage vendors already have pieces of the puzzle such as the ability to move only changed blocks, or to virtualize storage pools, manage data lifecycles of data on multiple tiers of physical storage or deduplicated data for faster backup and more reliable recovery. But the vDPAS category is much more than these point solutions. Some vendors have even subscribed to a common repository of data concept but these solutions still only take you part way to what vDPAS is designed to deliver.

vDPAS is a game-changing technological approach to tackle dizzying data growth and the resulting data sprawl, storage silos and inability to effectively manage, protect and share large amounts of data. It requires brand new thinking and a fresh new approach to start from a clean slate without any legacy baggage.

A few years back, Mendocino Software and Revivio tried and failed at their attempt to create such a product due to the lack of maturity in storage virtualization technologies, amongst other factors, at the time. Today, the storage virtualization technologies are mature and the conditions are ripe for vDPAS solutions in the marketplace. Besides a few stealth companies, we expect that the legacy players have ongoing developments targeted at this space. In our view, a few companies, exemplified by NetApp, already have the underpinnings to meet vDPAS requirements and will likely do so in the near future. What NetApp needs to do to better adhere to vDPAS is an enhanced de-duplication ratio, ability to hold more snapshots, adeptness to manage data at finer granularity than a file, and have a comprehensive SLA to manage all the copies.

Unquestionably, in our view, one new entrant symbolizes the vDPAS vision more so than anyone else in the market today. That company is Actifio. Actifio is a true visionary in the data protection and data availability space. It attacked the data protection and availability problem with a clean slate and has developed a single product, PAS, or Protection and Availability Storage, to enable a radically transformed storage story enabling a new level of agility, nimbleness, flexibility, and efficiency that is unfathomable with traditional siloed data protection and data availability approaches.

Actifio's solution represents a paradigm shift in data protection by attacking the data storage problem with a modular approach to managing the content under the control of its software. PAS uses a distributed object file system to virtualize data management and provide a single-solution for data protection and restoration instantly providing thin or full copies on-demand to all stakeholder applications. Specific SLAs utilize application-aware priority settings and can be configured in the interface utilizing best practices, based on system needs. This single solution is easily deployed into an existing infrastructure to replace disparate data protection products and test and development toolsets. PAS can also be used for unified search across indexed data and as an analytics platform for improved compliance and business value. Actifio solution does not require backup windows yet

providing instantaneous restores without data movement – allowing organizations to always meet Recovery Time Objectives (RTOs) and Recovery Point Objectives (RPOs).

We expect additional storage vendors to enter the market for vDPAS. We are also convinced that all major legacy data protection vendors are feverishly working to develop and introduce products that will logically fall in this category. And we consider a few existing players to have a partial foot in this category already. But bold, fresh moves like Actifio's PAS only happen when there is no legacy baggage, or revenue streams, and no existing customers to support with legacy products.

In today's economic turbulent environment where IT budgets are constantly under the microscope, businesses are eager to graduate to exponentially more attractive products that can give them fundamental advantages in IT, resulting in increased competitiveness in their markets. We believe in the world of data protection products that meet the vDPAS criteria will provide such an advantage in the coming years.

Ashar Baig and Christine Taylor are Taneja Group analysts specializing in data protection and information management. www.tanejagroup.com.