Solution Showcase

Next-level Resiliency with IBM's Dynamic Disk Pools

Date: April 2016 **Author:** Scott Sinclair, Senior Storage Analyst

Abstract: The rapid growth of data being driven by contemporary mainframe workloads has, in many cases, rendered traditional storage architectures unsustainable. As businesses continue to deploy new generations of analytics and intelligence workloads, the costs of managing the resulting capacity growth are expected to increase. IBM's Dynamic Disk Pool technology, available in its TS7700 product line, provides a differentiated and higher level of protection, resulting in increased resiliency and performance to support a new generation of workloads.

Overview

The impact of data growth has been so pervasive; it is difficult to discuss storage technology without paying some homage to the effects that growth has had on the industry. While the storyline has existed for years, many organizations have recently reached a tipping point, where the level of content has surpassed the ability for traditional data protection methods to keep pace. In other words, the cumulative impact that has resulted from years of rapid data growth has rendered traditional storage architectures unsustainable for many enterprise environments.

This effect is potentially most impactful in mainframe environments. The security, resiliency, and performance offered by mainframe technology has led mainframe infrastructure to become a bastion of an enterprise's most secure and often most valuable digital assets. In part, it is this advanced level of security and performance offered by mainframe environments that has created such an affinity between these environments and business intelligence workloads. The recent rise in adoption of these big data analytics and business intelligence workloads is only further serving to increase the cost and complexity of data storage for mainframe environments.

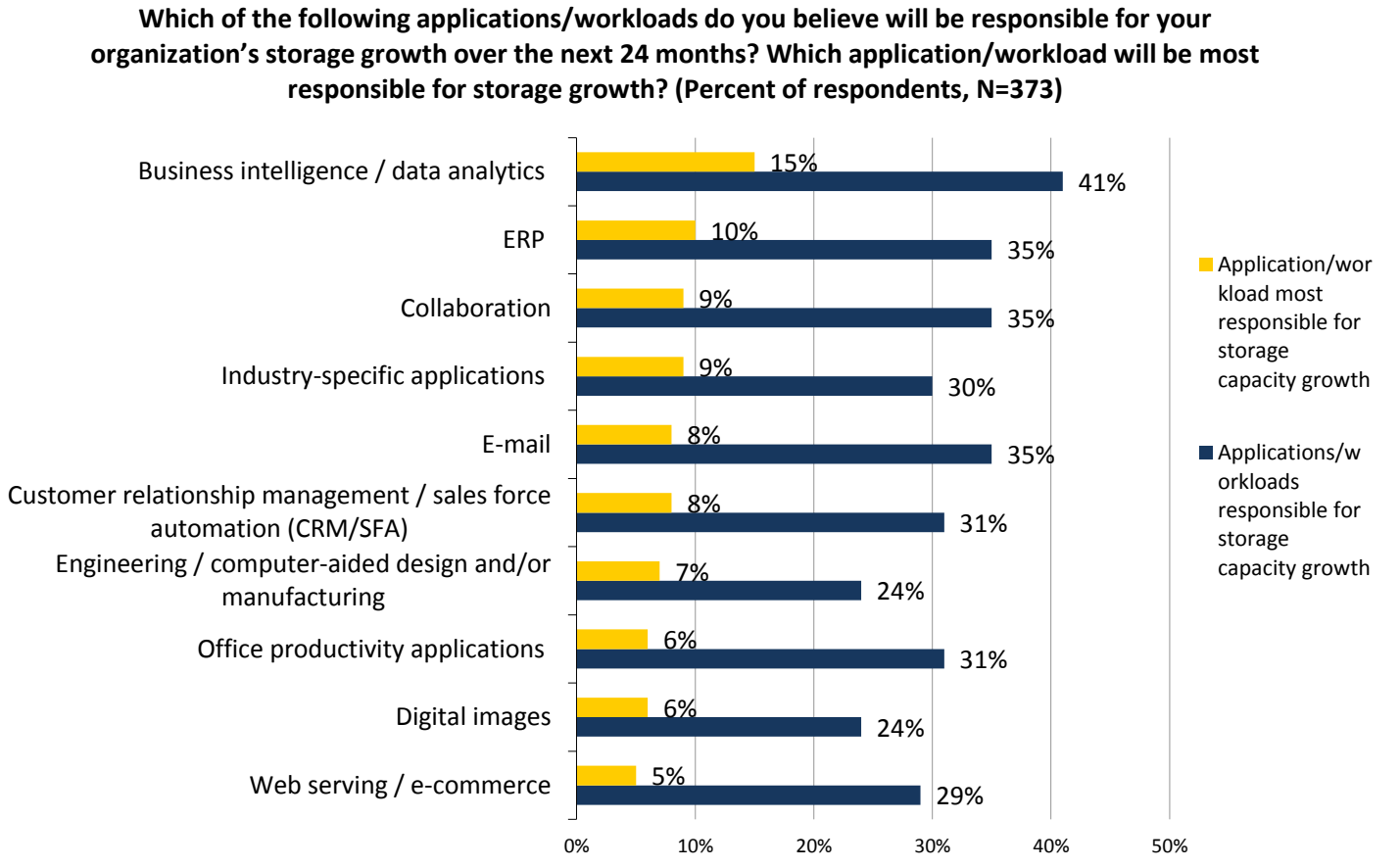
IBM, a technology innovator with a history of leadership in mainframe solutions, is delivering an advanced protection scheme called Dynamic Disk Pools (DDP) designed to address the challenges of managing ever-increasing content levels. IBM's DDP technology is integrated in its TS7700 solution to provide increased resiliency as well as performance benefits to support an emerging generation of workloads. The cost saving benefits of the DDP technology are further increased when paired with the TS7700's ability to spread information across both disk and tape. The net result simplifies data storage protection and ensures long-term survivability of critical data, while keeping costs under control.

The Impact of the Rapid Rate of Data Growth

In 2015, Enterprise Strategy Group (ESG) surveyed 373 IT decision makers responsible for their organizations' data storage environment as part of a research study investigating general storage trends.[1] This study asked participants to identify their organization's biggest challenges in terms of their storage environment. Participants were able to select multiple answers with the two most-cited responses being hardware costs (27%) and the rapid data growth rate (26%). These results were not necessarily surprising. It is important, however, that the remainder of the top ten responses, which included such challenges as data protection, staff costs, data migration, data and device management, physical space, and power and cooling, in addition to the aforementioned hardware costs, are all exacerbated by the rapid growth of data.

These impacts are even more critical in mainframe environments, where the added demands for security, resiliency, and performance increase the cost of capital and of operations. This study also asked storage leaders to identify the applications that they expected to be responsible for their organization's storage growth over the next 24 months, and business intelligence and data analytics (41%) was the most popular response. As analytical workloads increase deployment in mainframe environments, the subsequent increases in the amount of data that organizations need to store, access, and protect will serve to substantially increase the storage burden and, subsequently, the cost.

FIGURE 1. Top Ten Applications/Workloads Most Responsible for Storage Growth



Source: Enterprise Strategy Group, 2016

IBM's DDP technology was designed specifically to address the cumulative impacts that years of exponential data growth place on storage infrastructure. Traditional RAID architectures were often designed for capacities much smaller than those

[1] Source: ESG Research Report, *2015 Data Storage Market Trends*, October 2015. All other ESG research references and charts in this solution showcase have been taken from this research report.

that exist today. As a result, the scale of these contemporary storage environments significantly reduces the data protection potential of traditional RAID, increasing the risk to data.

Next-generation Resiliency Powered by IBM's Dynamic Disk Pools

Traditional storage systems and their supporting RAID architectures were designed primarily to protect against single drive failures. The advent of larger hard drive capacities as well as larger capacity storage environments, however, have increased the likelihood of multi-failure events, which leaves the organization exposed to the danger of losing data. Multi-failure scenarios can arise in a couple of different ways: additional drives could fail during a rebuild activity currently in progress, or the rebuild could encounter bad sectors on a previously designated healthy drive. The likelihood of an individual bit of data being unrecoverable is rare. In the case of near-line SAS drives, the chances of encountering an unrecoverable bit is often about 1 in 10^14 bits. While that may seem rare, a single 5TB drive has 4x10^13 bits which increases the probability of encountering an unrecoverable bit on a given disk to about one in three. With traditional RAID solutions, often the unrecoverable bits are not identified until after a drive fails and a rebuild activity is initiated. In these cases a single failure has already happened and encountering the unrecoverable bit represents the second failure. As a result, the industry has shifted toward solutions with additional (dual and even triple) parity calculations to help alleviate some of these concerns.

Depending on which study is referenced, spinning hard drives have been found to have around a 3-5% chance to fail in a given year.[2] While those percentages may seem small, multi-petabyte environments with hundreds or thousands of drives often see disk failures as a regular occurrence. For example, if you assume the chance of a drive failure is 4% in a given year, then for a 10PB storage environment with 2000 drives, there is about a one in five chance that at least one drive will fail in the environment on any given day. When that percentage is combined with the likelihood of encountering an unrecoverable bit, the chances are even higher that at least some portion of the data set will be vulnerable to an additional failure during a rebuild operation. Longer rebuild times only further increase the amount of time a RAID group is exposed to the potential of a tertiary failure. Reducing the recovery time directly reduces the amount of time that a data set is exposed to a failure, ultimately improving resiliency.

The Potential of IBM's Dynamic Disk Pools

At the heart of IBM's DDP technology are the concepts of parallelism and performance. Storage media, like all technologies, is imperfect and fallible. In large capacity environments, failures are commonplace. To ensure survivability of data, component failure must be expected and recovery must be accelerated.

In a traditional RAID architecture, when a drive fails, another—often called a hot spare—is designated to replace it. The data from the failed drive is recovered and written to this new drive. During this rebuild process, data performance is often slowed as the storage system designates a portion of its processing power to recreating and writing the data. As hard drive capacities have increased, the amount of data that needs to be rewritten during a rebuild process has also increased. The end result is a longer rebuild, increasing the amount of time that system performance is affected and that the data being rebuilt is exposed to a multi-failure event.

IBM's DDP technology abstracts the failure domain from the physical media. If a drive fails, in this scenario, the capacity represented by the failed drive is not replaced with a single drive. It is replaced by free space across a large quantity of remaining drives. According to IBM, this process can reduce the amount required for a rebuild by up to 80%, which can reduce the likelihood of a disk pool failure to one-fifth of what is seen by alternative methods. Additionally, the amount of time that system resources are burdened by rebuild activities is also reduced. The net result can be a dramatically faster

[2] http://www.cs.toronto.edu/~bianca/papers/fast07.pdf
http://static.googleusercontent.com/media/research.google.com/en//archive/disk_failures.pdf

rebuild, translating into higher system performance and higher resiliency as multiple components work in parallel to restore the data.

Modern enterprise data capacities have in many cases exceeded the protection capabilities of traditional RAID architectures. In high-capacity environments, therefore, technology that can accelerate rebuild operations, like that of DDP, is paramount. While for these large capacity environments, technologies like DDP are simply a necessity, DDP also offers benefits applicable to a broad range of storage environments, such as:

Dynamic Disk Pools

- Reduce rebuild times by 80%
- Significantly improve resiliency
- Support environments 10+ PB in capacity

- **Superior Data Availability and Resiliency:** IBM claims that Dynamic Disk Pools reduce rebuild times by up to 80%. As stated previously in the paper, this ability is paramount in large capacity environments. As such, IBM has revealed that the technology is often deployed in environments supporting 10s of petabytes of data, or more. This added resiliency can also lead to greater performance as the system wastes less resources on rebuild operations.
- **Simplified and Automatic Protection:** By abstracting data protection from the specific disk configuration, DDP simplifies storage management. There is no need to identify or manage hot spares. Reserved capacity is automatically set aside for rebuild activities. And when delivered as part of the TS7700, IBM offers a storage virtualization solution that can consolidate both tape and disk resources.
- **Improved Performance and Density:** Empowered by the resiliency of DDP, IBM is able to offer higher density storage enclosures. These enclosures allow the resulting solution to do more while taking up less space, power, and cooling. This greater density also reduces the number of storage drawers and cabinets required for performance.

While the added resiliency and automation provided by DDP can certainly help reduce TCO, these benefits are increased with the flexibility offered by the TS7700. The ability to offer a diverse set of storage media, including disk and tape, allows IBM to tailor the storage infrastructure to the specific needs of the workload. The net result from the combined TS7700 and DDP solution can present significant TCO savings, which is critical in an era of ever increasing data.

The Bigger Truth

Though critical, architectural innovations that deliver differentiated storage resiliency, such as DDP, can often be overlooked during the storage system evaluation process. Almost every enterprise storage system professes the ability to maintain data integrity and availability in the wake of component failures. And at smaller capacity levels, those claims hold up. When new storage deployments take place, data capacity is often at the smaller side of its lifecycle. It is at this point that the benefits of technologies like DDP are less pronounced. They are, however, no less important. As capacities increase, those systems with less resilient architectures create limitations. These limits to capacity growth often spawn a costly but inevitable data migration event to a platform offering a technology like that of DDP. In other words, IBM's DDP technology delivers benefits to both high capacity multi-petabyte environments as well as those environments that have yet to reach those capacity points, improving resiliency, adding performance, and reducing costs.