IDC
Analyze the Future

IDC TECHNOLOGY SPOTLIGHT

Storage Tiering

September 2009

Adapted from *Achieving Storage Efficiency Through Storage Tiers* by Noemi Greyzdorf, IDC #217734

Sponsored by BlueArc

In the age of tighter budgets and continuing data growth, storage administrators are facing an ever increasing challenge of aligning storage resources with the demands of business applications. IDC predicts that data growth will reach a 50% CAGR through 2012. The drivers for such growth are:

- More content is being created and stored in digital format than ever before. Additionally, rich media content is increasing in size due to improvements in imaging technology.
- Continuing globalization of local economies is driving the need to communicate and stay connected with colleagues in dispersed locations as well as with partners, vendors, and customers, causing more data to be created, replicated, and shared.
- Regulatory compliance, governance, litigation support, and best practices in support of business operations and product development are driving longer retention periods for data.
- Data protection and disaster recovery initiatives are creating multiple copies of data to ensure business continuity.

In most datacenters, over 70% of the data being created, stored, and retained is considered inactive. Not all inactive data is old and useless, but the cost of keeping all this data on high-end storage is expensive, and organizations large and small are seeking more intelligent ways to store data over time without adding management overhead or complexity.

Four Steps to Achieving Storage Efficiency

Storage efficiency is not a widget one can buy; it is an iterative process by which an organization examines how resources are consumed and where there are inefficiencies that can be addressed either through a process change or through a more intelligent use of the technology. There are four key steps to this process that promise the optimal return on time and money invested:

- It is difficult to make a decision about data without knowing what value this data represents to the organization. Business units, legal, and storage management need to work together to understand what information exists in the organization, where it is stored, how it is being used, and the business and legal requirements for retaining it. Setting up policies to determine retention, protection, and access policies can identify and eliminate data that holds no value and properly store data that does.

- Align application storage demands for performance, capacity, reliability, and availability with the actual storage resource. Don't overprovision or underprovision storage or it may cost you in productivity. Additionally, implement intelligent storage tiers to accommodate changes in data access patterns and data roles over time.
- Implement capacity optimization technologies to help stave off acquisition of additional storage and to more effectively utilize the storage resources already on hand. These technologies include thin provisioning, single instancing, deduplication, and compression.
- Design and implement data protection and disaster recovery policies to reflect the business demands on the data being protected. Ensure proper execution of the policies and timely allocation and reclamation of storage resources.

Intelligent Storage Tiers

There are two types of storage tiers based on different parameters. The first tier is based on the demands of the application and is designed to respond to applications as data is being created and written to storage media or as data is being read off storage media for the purpose of processing it by the application. The second type of storage tier is based on multiple parameters of the data such as the changing access patterns and value of data over time. Data is moved into the second type of tier transparently by the storage system without any impact to the application or end user. Often, the second type of storage tier, sometimes referred to as a tier within a tier, serves as a repository of more inactive data or data that must remain online and available but is not accessed as frequently. The decisions for the first type of tier and the second type of tier are independent of each other though they may be represented by similar types of storage components.

Application Demand–Based Storage Tier

This tier is primarily defined by the performance parameters required by a given application. If the application requires performance, allocating slower storage would negatively impact productivity and may result in revenue loss. Conversely, allocating high-end storage to an application that doesn't have the stringent performance demands may result in higher than necessary costs for the resource. In the end, each application's storage tier may vary based on performance and access (FC, iSCSI, NFS, or CIFS), drive type (FC, SAS, SATA), reliability (RAID 1, 10, 5, 6), and availability (multipathing, redundant controllers).

Storage Tier Within a Tier

Once data has been created by an application, it begins its life cycle. Over time the value of the data may diminish, access patterns to it may change, or performance requirements may be relaxed. Keeping such data on the same storage tier as active data may consume unnecessary resources and costs. This less active data, at some point in time, may be better served by a lower tier of storage. The challenge is identifying when and how to migrate this less active data to a different storage tier automatically, seamlessly, and transparently to the application and the end users. There are four main ways this can be achieved:

- Some applications have a built-in tiering capability, allowing storage administrators to take advantage of such built-in functionalities.
- HSM is third-party software that moves files to a different storage tier based on file characteristics such as age, size, and type. In order to do this without impacting the end user or an application, a stub file is left behind to redirect the requester to the file's new location.

- Storage tier within a tier can be achieved within a volume through block-level storage virtualization. In this scenario, a single volume consists of multiple storage tiers and blocks are migrated across those tiers based on access patterns.
- A single file system may have the ability to manage multiple pools of storage, each representing a different tier. The file system migrates the file across tiers without impacting the end user or application experience. No stub files are required.

In all four cases, a tier within a tier, or moving data to a lower tier in the background by the storage system, can be created for more granular alignment between the application, data, and storage resources.

Benefits of Storage Tiers

There are two main wins with storage tiers. First, appropriate storage keeps applications humming. Second, appropriate storage reduces costs. A seamless implementation of storage tiers simplifies data and storage management and more closely aligns the resources with business demands.

Product Profile

BlueArc SiliconFS has been designed to deliver high-performance, highly reliable file services to applications and end users. The file system allows users to define how best to use this technology. Key areas where BlueArc's SiliconFS provides significant value and differentiation are:

- The system can be configured in a number of ways. Multiple file systems can coexist on a single server, each representing a different class of storage allowing business units and application owners to define more directly required storage characteristics. At the same time, multiple file systems can be abstracted with a global name space that can span multiple servers, allowing for ease of use, scalability, and data persistence.
- Each file system on a single server can represent a different pool of storage, allowing the administrators to define policies for data placement. Using BlueArc Data Migrator™ software allows files to be migrated across storage tiers seamlessly, without affecting the end user or application experience. Using Cross Volume Links™ together with Data Migrator allows administrators to take advantage of older file-based storage systems as a secondary tier.
- Data must be protected in a way that represents its value to the organization. SiliconFS supports a variety of data protection options that give managers the desired level of flexibility and recoverability.

One way that users can store data more efficiently and with greater utilization of their storage resources is through storage tiering. BlueArc's SiliconFS has designed a way for managers to automate the process of migrating data to secondary storage without impacting the application or end-user experience and yet not create unnecessary complexity or management overhead.

SiliconFS delivers a tiered architecture using two key software features: Data Migrator and Cross Volume Links. Storage tiers are configured first with SiliconFS.

Each tier of storage, be it FC, SAS, or SATA connected to the filer gateway, can be configured into one or more storage pools. Within each storage pool, one or more file systems can be created to serve as source or target for data migration. A tier 1 file system is typically designated as the active, or source, file system, and a tier 2 file system is typically designated as a target file system.

When an application creates data, the data is written to the active file system. Storage managers can define policies for migrating data to the second tier automatically. The parameters that drive the

policy can be broad, including the most common file characteristics such as age, size, file type, owner, and last time accessed. Once a file complies with all the rules of the policy, the Data Migrator software moves the file to the second tier, leaving behind an intelligent stub file. Once created, policies can be activated by triggers (e.g., the file system is 95% full), run at scheduled times, or run on demand by the system administrator.

A stub is a file that redirects the user to the new location of the file that has been moved. In SiliconFS, the stub file contains additional information about the file, making metadata searches and views of the directory simple and fast. The stub file contains enough metadata about the actual file to provide necessary information during searches without retrieving the actual file. This is very efficient, making the application and end-user experience more transparent.

When the Data Migrator moves the file from tier to tier, the migration occurs over the SAN interconnect. The file systems representing different tiers reside inside one server and are connected to the various storage pools via SAN. This is especially useful when the LAN is congested or needed for more active operations. BlueArc doesn't limit storage tiering to file systems within a gateway. Other file systems or file-based storage systems can be used as a tier. Cross Volume Links allows the system to access foreign file-based storage systems via NFS and use them as a secondary tier. This allows older storage to be repurposed and delivers a level of investment protection that is easy to justify. The application doesn't know that the data has been moved to another NFS mount because there is no impact on it from the usability perspective.

Moving data across storage tiers is a challenge, but making sure you protect all data, active and secondary, can be an even bigger challenge. In many traditional HSM implementations, the backup system would pick up all files in the active file system, regardless of whether they were actual data or stub files. When the restore is performed, it could be complicated. First, the stub file has to be restored and checked that it still represents the location of the actual data. Only then could the data itself be accessed. In these scenarios, the secondary tier would also have to be backed up to ensure data recoverability and availability. This creates additional complexity in both the backup phase and the restore phase of data protection. SiliconFS has addressed this issue in three ways:

- The administrator has the option to back up only the active file system. This allows for a faster, more efficient backup. The files that have been moved to the secondary tier of storage could be backed up separately through the active file system on a different schedule, potentially less frequent in line with a migration schedule.
- The complete data set can be backed up, including data on the secondary storage. This represents the complete image of the directories used by applications and end users.
- On the restore, the system is aware of the files that have been migrated. When a file is restored, it is always restored to the active file system, but knowing that it belongs on the second tier, it automatically migrates there in order not to fill up the active file system or run out of space. This is a key differentiator that allows the system to provide the advantage of storage tiers and data protection without added complexity or management overhead.

The result of the implementation of SiliconFS is an efficient use of storage based on application demands as well as changing characteristics of data as it goes through its life cycle. Using storage tiers translates into savings in storage costs, datacenter floor space, and power consumption, all of which are key challenges managers are seeking to address.

Challenges

The file-based storage market continues to evolve as demands on the solutions continue to change in response to application requirements. The challenge many file-based storage solution vendors are facing is identifying the market where they can differentiate and compete successfully over an extended period of time. It is hard to decide not to be all things to all people, but it is important in order to be successful. Storage tiering is a feature that is valued in many market segments, but alone it is not enough. The whole package has to be positioned to address as many of the requirements of the market segment as possible, but it is a valuable functionality that will become a must-have as the market continues to evolve.

Conclusion

Achieving storage efficiency, unfortunately, is not as easy as buying an efficiency widget, but it is an attainable goal. Organizations need to make storage a priority and start from the beginning, learning about the information being stored, how it is being stored and where, and what value it represents over time to the organization. With this knowledge in hand, storage administrators can start making decisions regarding technologies that help them achieve desired results. Some of these technologies, such as thin provisioning, are readily available on many storage systems; others, such as intelligent storage tiering, are less prevalent. These technologies, as well as proper execution of established policies, can help organizations save significant dollars in storage and operations costs. Intelligent storage tiering reduces the acquisition cost of storage, the datacenter footprint consumed by storage, and the energy and personnel required to power and manage the environment. In the end, remember that attaining efficiency is the journey, not the summit.
