Embracing tiered storage

Digital systems capable of capturing, storing, analyzing and presenting data about almost every facet of our organizations have created a data tsunami. Decision making tools based on Big Data are not a novelty any more; but they are becoming a competitive necessity. It is not surprising, then, that the datacenter storage capacity needs of all organizations, large and small, have seen phenomenal growth in the past decade. This growth is expected to continue into the foreseable future. However, all data is not created equal and thus all data need not be treated the same by the storage system. IT executives must deploy innovative solutions to make the most relevant data available to applications and decision makers in the most expedient way while containing overall storage costs.

> **We live in a data driven world.**

Tiering strategies have emerged as an indispensable tool to cost- effectively harness the massive data growth. Conceptually, tiering enables an organization to move less frequently accessed (warm) data to high capacity (and hence less expensive) sections of a storage solution while keeping the more frequently accessed (hot) data in high performance (and hence more expensive) sections of the storage solution. Most tiered storage solutions in the market today provide simple data-aging based tiering that store data on multiple storage tiers based on drive performance and capacity, including flash drives. However, enterprise storage solutions that deliver low TCO and high ROI must come with intelligent software that can optimize the placement of the right data at the right time on the right storage tier.

Need for Automated and Intelligent Tiered Storage

Lately tiered storage solutions have gained momentum due to increased adoption of Information Lifecycle Management (ILM) paradigm for storage cost containment. The ILM paradigm asserts that:

- The time value of data changes significantly over its lifespan. With time, the need to access a piece of data diminishes exponentially.
- At any given point in time, only a small portion of data is being accessed
- Even though only a small portion of data is actively accessed, old data has to be kept for a long time for many business reasons, like regulatory compliance.

The duality of data types – hot and warm – creates a need for storage solutions that can save costs by storing data on different tiers of storage based on performance. While tiered storage solutions that can categorize data off-line with human intervention have been around for a long time, to truly bring the ILM vision to life, and thus provide the storage cost savings, requires a tiered storage solution to automate the process of categorizing data based on real-time access patterns and move the data intelligently to the right storage tier at the appropriate time.

Automated and intelligent tiering solutions typically categorize the data (hot or warm) based on multiple attributes (also known as metadata), for example, how often is the data accessed, how important is the data, and when was the data last accessed. Such tiering solutions primarily tier data based on the following important factors:

1. **Data Activity:** Data activity, also known as data temperature, is determined by taking into account the attribute values of the data established during the application workload execution. How active is the

data? What is the data temperature? Hot data moves towards high performance tiers and warm data stays on high capacity tiers.

2. **Data Sensitivity**: Data sensitivity is determined by granularity of data movement and frequency of data movement. What is the minimum data set that can be tiered or moved around? At a file level granularity can be either whole file or entire root directory. At a block level granularity can be either a few blocks or the entire volume. How often is data tiered or moved around? Frequency can be once a day or once every hour or once every minute.

3. **Data Location**: Data location consists of various tiers of storage in a storage solution. A storage tier is based not only on its drive type, size and speed, but also on its RAID characteristics.

Data attribute values and how they change over time are based on a workload characteristics. The workload intelligence gathered during its execution is used by the tiered storage solution to place each data on an appropriate tier. Thus, workload characteristics play a key role in determining the cost effectiveness of a tiered storage solution.

For example, tiered storage required by a light activity task-worker NAS file-share workload is very different than that required by a high volume OLTP workload. The NAS workload may be cost- effectively tiered on RAID50 based SAS and NL-SAS storage tiers, whereas RAID10 based SAS and flash storage tiers may be needed to optimize storage costs for the OLTP workload. Similarly, the NAS workload may be cost-effectively optimized by using a tiering sensitivity of file granularity once a day, whereas a tiering sensitivity of 4KB block granularity once every 4 hours may be needed to optimize storage costs for the OLTP workload. Tiered storage with built-in intelligence and automation to dynamically adapt the data sensitivity and location based on the workload can eliminate manual involvement and lower TCO.

Figure 1: Primary factors driving the TCO of a tiered storage solution: While data activity, data sensitivity and data location drive the cost effectiveness of a tiered storage solution, its TCO is also determined by the ease of provisioning storage capacity in each tier.

Moreover, the ease of reconfiguring tier capacities in a tiered storage solution also impacts its cost effectiveness. As the workload storage needs change over time, the ease of adding or removing storage capacity in each tier affects staff productivity. For example, a storage solution where the storage capacity in each tier can be provisioned on-line by a general system administrator can achieve a lower TCO than the storage solution which requires a scheduled down time and a vendor representative or a highly skilled storage administrator to provision storage capacity in each tier.

Business Benefits of Tiered Storage

Tiered storage solutions can deliver real and measurable benefits that directly impact the datacenter TCO. Some of the more important benefits of tiered storage solutions are:

- **Cost reduction while meeting SLA**: Tiered storage can cut the overall storage costs by up to 80 percent. By providing the flexibility to provision the right mix of high performance and high capacity storage, tiering can reduce the total cost while meeting service levels.

- **Improved Utilization Efficiency**: Tiered storage places the right data at the right time on the right tier from a performance and cost perspective. As a result, such solutions can improve ROI by optimizing the utilization of each storage tier - higher tiers of storage are optimized for performance utilization and lower tiers of storage are optimized for capacity utilization.
- **Operational Simplification**: Tiered Storage solutions that have built-in usage- gathering intelligence and policy based tiering automation can eliminate the need for manual intervention to maximize performance and utilization. Also, the flexibility of performance, capacity expansion or evacuation within each tier also simplify planning and management.

Adoption Considerations

In today's complex data centers, implementing policies that leverage diverse storage tiers is not an option but a necessity. Hence developing an understanding of key adoption considerations is important to differentiate the available options and identify the right solution.

- **Workload Characteristics**: In order for a tiered storage solution to be cost effective, it must take into account workload characteristics and how they change over time. Virtually all production workloads can benefit from simple tiering schemes. However, storage solution that can dynamically and automatically tier data based on workload characteristics will be able to minimize the storage TCO.
- **Storage Virtualization**: Storage virtualization, with data services such as automated small block data movement, creates a robust foundation for an advanced tiered storage solution. As a result a tiered storage solution with a built-in virtualization framework is much more effective in providing tiering benefits than a legacy storage solution with a bolted on virtualization framework.
- **Advanced, Automated and Real-time Intelligence gathering:** Given the changing attribute values of application data and the large size of storage capacity requirements today, it is impractical and inefficient to perform manual or off-line data classification. Tiered storage solutions with built-in automation and native intelligence gathering for a comprehensive data classification at a granular level can lower TCO by eliminating the management intervention required to realize effective tiering on an ongoing basis.
- **Tiering Flexibility:** Consolidated tiered storage solutions are shared among a variety of workloads with differing access patterns and performance needs. As a result, the ability to adapt the granularity, data movement policy and frequency can have a huge impact on its cost-effectiveness.
- **Tiering and Flash**: Implementing tiering in flash or solid state drive based hybrid storage solutions warrants that additional considerations, like frequently accessed read-only data or frequently accessed write-intensive data, be taken into account than tiering solutions for traditional hard disk based storage solutions. Tiering solutions that enable customers to intelligently, efficiently and flexibly integrate and protect different types of flash and RAID implementations across server and storage domains are highly desirable in modern performance hungry datacenters.

Conclusion

Tiered storage solutions are an indispensable foundation for today's data intensive world where IT executives must deploy information management best practices, such as ILM, to optimize datacenter TCO. Flash adoption in the server and storage domains further accentuates the need for robust tiering solutions to achieve lower cost and higher performance. Intelligent tiered storage solutions that extend the scope of tiering to include server and storage domains are vital to realizing the full potential of tiering in the modern datacenter.

Dell is a diversified information technology partner with a broad portfolio of products that meets every need in the modern data center. Dell's Fluid Data Architecture is designed to automatically and intelligently optimize the data center by integrating best-of-breed technologies across server, storage and networking to put the right data in the right place at the right time – for the right cost.

Dell can optimize your storage investments today, contact us to find out more about Dell's automated and intelligent tiered storage solution.