

CIO Challenge: Performance vs. Cost

The Data Temperature Spectrum

By: Dan Graham
General Manager
Enterprise Systems
Teradata Corporation

TERADATA®

THE BEST
DECISION
POSSIBLE™

CIO Challenge: Performance vs. Cost

Table of Contents

<i>Balancing Performance and Costs</i>	3
<i>Data Temperatures</i>	3
<i>Data Storage Grows Faster than Moore's Law</i>	4
<i>The Data Temperature Spectrum</i>	5
<i>Blazing In-Memory Tier</i>	5
<i>Hot, Warm, and Cold Tiers</i>	7
<i>Arctic Archival Tier</i>	7
<i>Data Movement Granularity</i>	8
<i>Results</i>	9
<i>Summary</i>	9

Preface

This white paper examines the importance of data temperature, storage virtualization, big data and in-memory concepts, and a vision for the future of integrated data warehouses.

I would like to thank the following contributors for their insights and editorial guidance: Todd Walter, Jim Dietz, John Catozzi, and Martin Willcox.

CIO Challenge: Performance vs. Cost

Balancing Performance and Costs

CIOs now have much-needed options in the struggle to provide system performance balanced against costs. Server virtualization is one such option, matching the performance needs to the server costs. Server virtualization enables the consolidation of applications onto one powerful server or the reuse of low-cost servers for low-use applications.

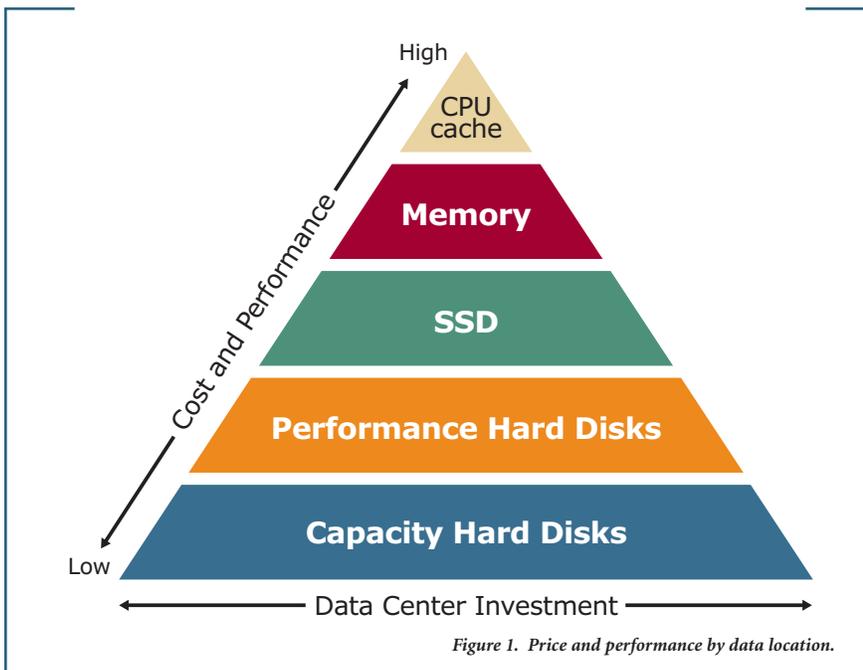
Storage virtualization is needed for the same reasons. Ideally, seldom-used data should be kept on low-cost, high-capacity disk storage, while high-use data should be placed on high-performance solid-state drives (SSDs). On some systems, this can be done with significant manual labor or

complex policy-based software. A better approach is storage virtualization.

This pyramid is based on some simple facts of computer physics. At the top are semiconductors which become smaller and faster every year. This includes memory within the processor itself (CPU cache) and, on the next layer, main memory within a server. These semiconductor devices are extremely expensive to manufacture, hence their higher price per gigabyte. In the middle are SSDs that use “flash” memory so they act like regular memory. SSDs and the tiers below them are persistent: the data is retained even after power is turned off, unlike CPUs and main memory. At the bottom are mechanical storage devices that are inexpensive to manufacture. Magnetic disks hold dramatically more

data each year, but have not improved much in speed over the last decade. The difference in price and performance spanning these hardware tiers is why storage virtualization is needed.

The storage market uses “temperature” as a metaphor for frequency of access. Teradata Virtual Storage uses data temperature to measure and automatically move data throughout the spectrum of storage media based on its performance needs. This innovation is profound in that it automatically migrates data based on actual use, responding directly to business users’ behavior. No policies, no labor – simply watching actual data usage. Teradata Virtual Storage knows the difference between data loading, user queries, and system tasks so it can optimize data placement for analytic workloads.



Data Temperatures

The most recently created data – whether in files or database tables – is the most heavily used. For business users, this means the most recent sales results, the most recent inventory on hand, the last 90 days of financial activity, customer responses to promotions, and more. But in analytic databases, data temperature can change rapidly and unexpectedly. Many business events can cause temperature changes on year-old “cold” data such as a corporate reorganization of sales regions, a new supplier of products displacing older SKUs, or catastrophic events like hurricanes. In such cases, data that was cold (or dormant) leaps in importance for a few days or weeks.

CIO Challenge: Performance vs. Cost

Teradata Virtual Storage tracks the data access patterns of the business users and batch jobs. This reveals that data warehouse queries are highly weighted to a small number of disk cylinders. Consequently, we describe the most popular data as “hot,” medium popularity is “warm,” and low-popularity data is called “cold.” In the Figure 2 example, 94 percent of the read/write activity by users is fulfilled by 20 percent of the storage capacity (in this actual analysis, roughly 10 terabytes). Measurements of multiple Teradata customers reveal some have high-capacity workloads (less than 10 percent of the data is hot) while others have higher performance needs (35 percent of the data is hot).

Seeing this, an astute CFO quickly asks, “Why not delete the cold data and save some money?” It turns out the cold data is often last year’s financials, government risk and compliance data, or consumer behaviors the marketing department uses for loyalty or churn analysis. The CFO needs that cold data for year-over-year analysis. Cold data may not be the most popular, but it is used daily to support vital strategic analytics and decision-making.

Teradata Virtual Storage increases a temperature indicator on the disk cylinders¹ each time a business user or batch application accesses the data. When the cylinder “warms up” enough, the data automatically moves from the device it resides on to faster storage in the hierarchy. The result is higher performance for

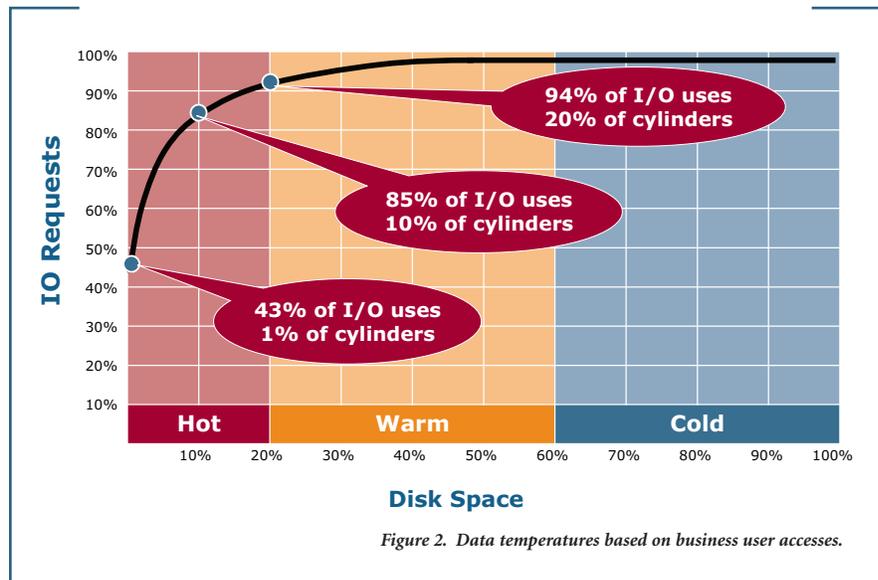


Figure 2. Data temperatures based on business user accesses.

data the business actually uses. In contrast, data cools off if it is not accessed for a few days and automatically migrates to slower disks in the hierarchy. The purpose of this temperature-based migration is to get higher performance on the most popular data while blending high and low storage costs. *InformationWeek* called Teradata Virtual Storage a “hybrid Porsche.”

“Just as a hybrid car doesn’t require the driver to switch from electric to gas and back again, the 6680 storage virtualization layer automatically determines when data is hot and when it’s not. It moves the information that’s in demand (typically the freshest stuff) onto SSDs and memory and then moves it back to less-expensive, hard-drive storage when it’s no longer the subject of intensive analysis.”

Data Storage Grows Faster than Moore’s Law

In 1965, Intel® co-founder Gordon Moore predicted that the number of transistors in an integrated circuit would double every two years. Moore’s Law has proven true for the last 50 years and is on track to remain so for the next 20. Silicon wafer advances provide dramatic increases in processing speed and memory capacity.

Less well known, is Kryder’s Law.ⁱⁱ Dr. Mark Kryder, Carnegie-Mellon Professor and CTO of Seagate, observed in 2005 that magnetic disk areal storage density doubles approximately every 18 months. On this trajectory, by 2020 there will be 14 terabyte hard disk drives (HDDs) that cost as little as \$100. Notice that Moore’s law doubles transistors every 24 months while Kryder’s law

1 Historically, cylinders referred to all the circular tracks on all disk platters where the disk read/write heads are currently sitting. Today Teradata uses this term for the allocated extent of contiguous storage regardless of the disk’s physical characteristics. These extents are typically 2-12 megabytes of data.

CIO Challenge: Performance vs. Cost

predicts disk density doubling every 18 months.² This means disk capacity grows faster than transistors. But while disk capacity is growing fast, HDD performance has not changed significantly in years. Meanwhile CPU's have grown exponentially faster. This is what makes ultra-fast SSDs an ideal solution to feed modern CPUs.

Kryder's Law and its effect on storage cost is just one of many trends driving the gathering of more and more data storage. Data scientists, actuaries, and marketing applications all need years of detailed data to improve the accuracy of their analytic models. Government regulations often dictate five to seven years of data retention for tax, liability, or consumer evidence. Even the basic concept of a transaction record has expanded from invoice line-items to include promotion codes, text comments, mouse clicks, GPS location, and more. Consequently, analytics are stepping up in sophistication from transactions to interactions as corporations collect larger amounts of data.

Big data is yet another trend pushing more data collections into the data warehouse. Smartphones, sensors, social media, weblogs, medical records, and the "Internet of Things"² are all emitting data at enormous rates. Converting these new data types into structured data and then into competitive advantage is the current visionary trend, with mainstream buyers not far behind.

The Data Temperature Spectrum

The data temperature spectrum provides a foundation for optimizing data placement across both performance and price. Optimized data placement needs to encompass server memory, disk storage, and external systems, as well. At the pinnacle of the hierarchy, we find the fastest, most expensive storage media: main memory.

In the middle, virtual data moves throughout the hierarchy on persistent disk storage driven by actual usage of the data. In contrast, data that has reached a "sub-zero" temperature may need to be moved off of the system to an archive such as the Teradata Extreme Data Appliance.

Ideally, the temperature of a data element would direct it to residing in the corresponding part of the hierarchy. Data found in the hot, cold, or warm tier can at any time be fetched and placed in memory to support a user query. In the future, "arctic" temperature data might even migrate automatically to and from the archive system.

Blazing In-Memory Tier

Teradata has been optimizing the in-memory hot cache in every hardware system release for more than 30 years. Each major expansion of memory size may require minor or substantial behind-the-scenes enhancements to exploit larger memories. The result is a highly mature, full exploitation of available memory in each release.

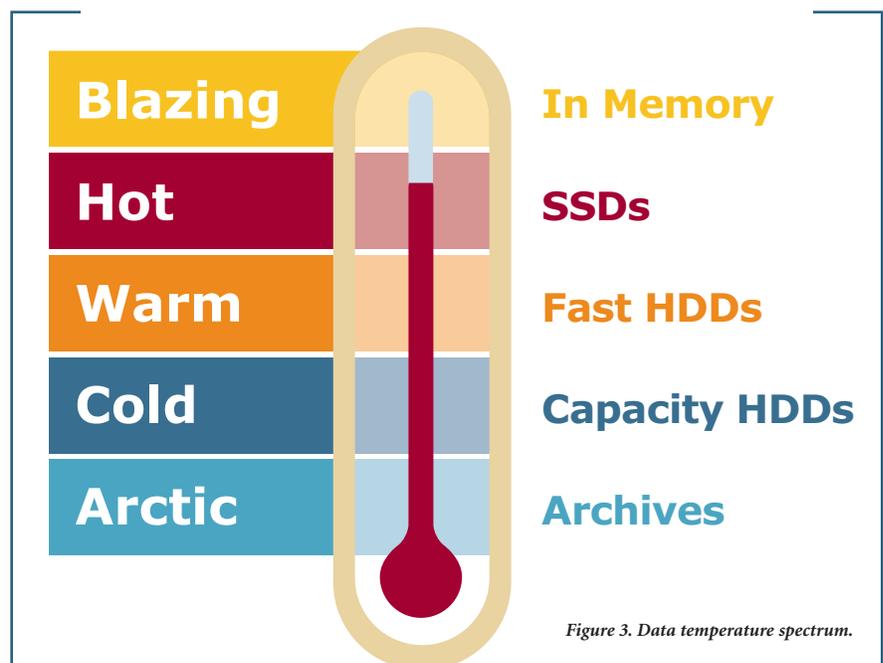


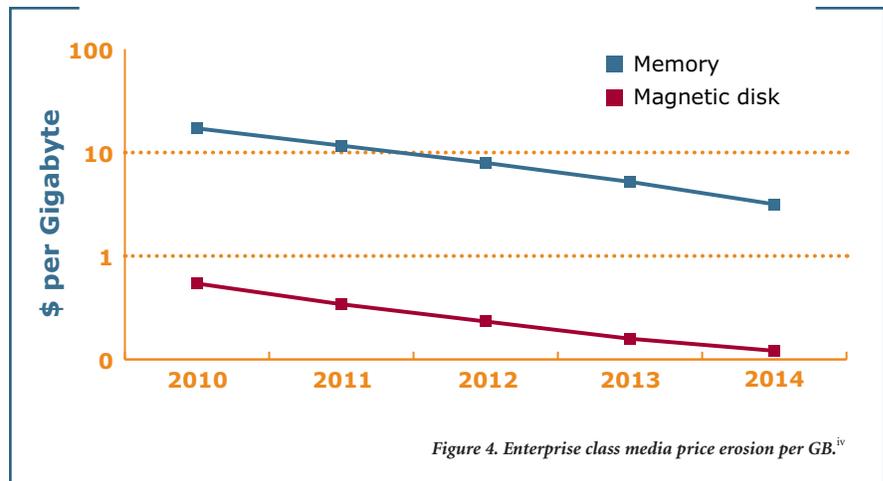
Figure 3. Data temperature spectrum.

2 Billions of smart phones, RFID tags, and sensors of all kinds connected to the internet. <http://www.economist.com/node/17388356>

CIO Challenge: Performance vs. Cost

At a high level, there are three kinds of memory to be managed: OS/database software, “hot memory cache,” and “working space cache.” Hot cache is optimized to hold blazing hot data for long periods of time. For example, small dimension tables may get into the hot cache and stick for days because they are constantly used. Account types, highly popular product IDs or SKUs, sales region parameters, and top suppliers lists get used in relational joins constantly, ensuring they stick in the hot memory cache. This kind of data is noted as blazing hot.

In contrast, working space cache holds data that is passing through memory for a short time to satisfy a user query. Often, the working space cache has dramatic turnover as one big table scan floods memory with data, followed by another user query that does the same thing. The key is to know which data is truly blazing hot and should stick in the hot memory cache versus which data is just passing through and can be discarded as soon as the query finishes. This is dramatically different from online transaction systems which simply age the oldest unused data out of main memory. Online transaction systems deal with small amounts of data that stick in memory for one to 10 minutes, so managing memory is relatively simple. This is different from analytic databases that manage hundreds of small, medium, and large answer sets running concurrently with some queries running for hours. What should be kept in memory then?



Teradata’s Optimizer and file system does not allow a big table scan to disrupt the hot memory cache. This one-time-use query data flows thru memory without being cached. Therefore, analytic databases must be more efficient by separating the blazing hot data from the other data.

In-memory processing is the current over-hyped technology that prescribes stuffing all the data into server memory to achieve faster performance while using compression techniques to cope with data volumes. The avoidance of disk access does indeed speed up performance with the trade-off of higher memory costs and substantial compression-decompression by the micro-processors. (Note that the system resources burned doing compress-decompress are resources that are not doing business tasks.) The application designer must weigh the trade-offs to determine when compression is effective or a drain on performance.

In 2011, memory prices hovered over \$10 per GB. They should drop to \$2 to \$3 per GB by 2015. Hard disk costs in 2011 were less than a dollar per gigabyte, but should dip into the pennies-per-GB range over that same period.

The slope of these two metrics has been consistent for many years. The clear implication is that for the foreseeable future, the cost of storage will always be more than a magnitude less than memory. Consequently, in-memory storage will not financially displace magnetic disk for many years, if ever.

Second, we caution buyers to beware of vendor claims that compression or columnar techniques produce the 50-to-1 or 100-to-1 reduction in storage use. Occasionally such dramatic results are possible. But more often, good compression techniques will deliver 4-to-1 up to 7-to-1 results.

CIO Challenge: Performance vs. Cost

This still leads to the feasibility of stuffing data marts completely into memory because of their normally smaller sizes. A data mart of 500GB compressed data should fit within a 512GB server memory.

But a data warehouse of 10 terabytes is not feasible in-memory today. If you could compress 10TB to one, it would fit in a TB-sized memory. But the operating system and database software would not. While TB-sized memories will become mainstream in a few years, 128 to 256GB memory sizes are the norm in 2012. Since demand for analytic data grows faster than memory sizes, the feasibility of putting all data into memory is limited to smaller workloads for the foreseeable future.

In-memory systems must carry cold data in the server memory. This means that cold data is stored at memory prices, even though it's only used once a day, once a week, or once a month. It is a non-trivial business decision whether cold data should be stored on low cost HDD or in main memory which costs 10-20 times more per gigabyte. Teradata recommends placing data on the most cost-effective location in the storage hierarchy.

Further memory price drops will fuel the in-memory debate. With three out of ten organizations collecting 100GB of new data daily, the thirst for data outruns Moore's law in regards to memory capacity. However, increased data demand from users and even cheaper disk capacities will ensure the need for temperature-based data management.

Hot, Warm, and Cold Tiers

SSDs have invigorated disk performance by providing an order-of-magnitude performance benefit in regards to input-outputs per second (IOPS) and data transfer speeds. Neatly wedged between traditional HDDs and main memory, SSDs provide the missing link in the performance hierarchy. In general, an SSD can achieve the same IOPS as twenty HDDs. Access time to any block of data is measured in hundreds of microseconds compared to HDDs, which are accessed in milliseconds. The SSD value proposition is simple: reduced latency and higher CPU utilization.

However the longer-lasting value of SSD will be found in its total cost of ownership (TCO). Performance delivers efficiency and the ability to provide corporate services faster. Keeping employees and business processes moving without delays is a benefit of faster hardware. Equally important is the labor cost of managing and tuning the performance provided by SSDs. Choosing the wrong SSD subsystem can lead to substantial labor per month by is cutting and pasting files into the SSD, they cannot be working on higher-value business-user support. Thus, SSDs should be simple to manage and use. Be careful: "automated storage" from some SSD vendors means it's automated after a long learning curve, substantial policy test and development, and ongoing policy maintenance labor.

Enterprise-class HDDs are found in primarily two types. High-capacity HDDs provide economical storage, but are slower because the platters spin at 7,500 revolutions per minute (RPM). High-capacity HDDs are an ideal location to store cold data. As an added benefit, Teradata Virtual Storage automatically compresses cold data to get even greater cost savings. High-performance HDDs have smaller storage capacity and spin at 10,000 or 15,000 RPMs. These devices are ideal for warm data temperatures, but can also be used for cold data. HDDs are an order-of-magnitude cheaper than SSDs in 2012.

Arctic Archival Tier

One purpose of the archival tier is to capture huge volumes of data economically for specific workloads. One use of the Teradata Extreme Data Appliance is as an easy-access archive system. Because a low cost per terabyte of storage is coupled with the Teradata Database functionality, it is an ideal location to do predictive analytics as well as providing government-mandated data retention. The archival tier neatly supports the data scientist, actuary, data miner, and auditing department. Numerous Teradata customers deploy the Teradata Extreme Data Appliance to hold thee to seven years of history while carrying one to three years on their Active Enterprise Data Warehouse system. By loading current data to both systems and deleting the oldest data on each, they are able to provide data for different purposes at different price-performance levels.

CIO Challenge: Performance vs. Cost

For some organizations “disk is the new tape,” meaning they are replacing magnetic tape units with random access HDDs. The economics are simple: with 2TB-capacity disks, it is feasible to replace numerous tape cartridges and also have the benefit of immediate access to the data. Magnetic tapes deteriorate – HDDs do not, and HDDs have RAID protection. Furthermore, floor space is reduced with the small 2.5 inch HDD form factor over tape storage. Nevertheless, not all tape drives need to be thrown out just yet – it will be at least a decade before magnetic tape is obsolete.

Data Movement Granularity

Exactly what moves across the temperature spectrum makes a big difference in both performance and cost. With Teradata Virtual Storage, granularity is set at disk cylinder size (roughly 2 to 12MB). With SSD capacities in the 400GB size, this allows the packing of up to 200,000 hot cylinders into an SSD to accelerate query performance. As a result, Teradata Virtual Storage migrates only hot data into the SSD with very little wasted space.

“Disk-tiering” software has existed for years to manage coarse-grained data movement between slow and fast disk storage. With labor-based policies for data temperature, the unit of movement is the “sub-LUN” (a Logical Disk Unit partition). Sub-LUN granularity at best is 1GB in size, and at worst 10GB or more. Unfortunately,

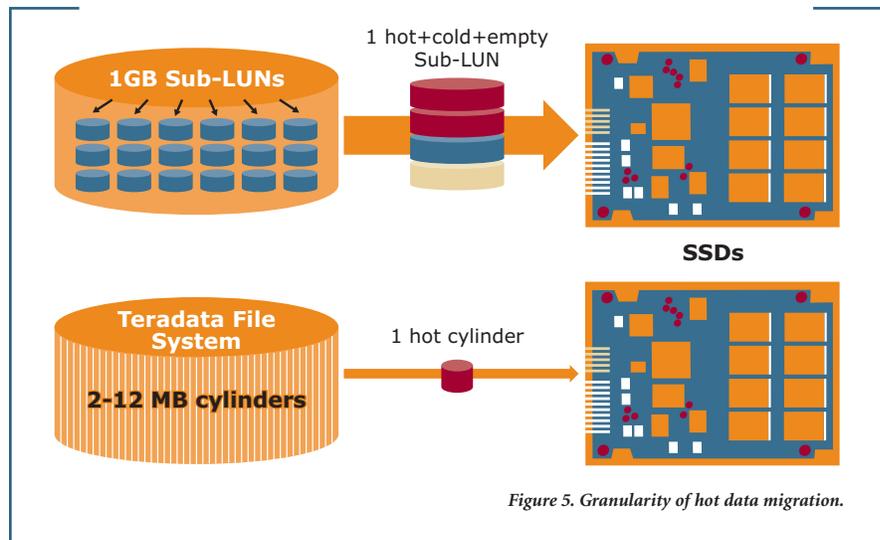


Figure 5. Granularity of hot data migration.

sub-LUN migrations are slow, plus they copy hot, cold and empty cylinders into the SSD, thereby wasting the valuable SSD space. This is a poor choice for exploiting the temperature spectrum, especially in real-time migrations.

Some disk-tiering software can move data blocks through the tiers in 4MB to 8MB sizes, similar to Teradata Virtual Storage. However, some of these only make decisions once a day regarding what data to migrate. Others have a moving average of one month for managing temperature. Teradata Virtual Storage keeps track of user activity by the week and migrates small amounts of data every five minutes. This means it is hundreds of times more responsive to user activity in adjusting temperatures. Furthermore, other disk-tiering subsystems do not know the internal use of the data, whether it’s a blazing dimension table, spool³ space,

data loading, popular query data, or table scans. Lacking knowledge of actual usage of the data blocks prevents full optimization of the data movement into and out of an SSD. For example, disk-tiering storage tools don’t know that spool data is temporary, yet vital to query performance. Spool files can exist for milliseconds, seconds, or minutes. Disk-tiering subsystems do not understand the difference between these many uses of spool files and a batch file that won’t be used again for days. Similarly, data that isn’t the most recently used could still be very hot though it’s not used for several minutes. Similarly, dimension tables are an analytic database concept unknown to disk-tiering subsystems. What is needed is an end-to-end usage monitoring and storage virtualization that clearly understands how the data is actually used at all levels in the storage hierarchy. Teradata Virtual Storage fills this need.

³ Spool is the interim result-set generated as the database optimizer joins tables together. Each spool is fed into the next query step, and then discarded as a new spool is produced.

CIO Challenge: Performance vs. Cost

Results

Customers using Teradata's hybrid systems are enjoying clear benefits from Teradata Virtual Storage. One customer found 77 percent of all queries were satisfied from the SSD while their cold data queries are using the hard disks. In this customer's workload, end-to-end service time for an IO is 360 percent faster than hard disks. In the table shown, why isn't the SSD 20 times faster than the HDD? It's because there is heavy contention (89 percent) for access to each SSD. This specific customer could benefit from adding a few more SSDs to their system, although they are very happy with the current configuration.

Another Teradata customer bought a new hybrid system configured to have 1.5 times higher system performance than the system being replaced. Once installed, they ran proof of concept stress tests – thousands of SQL statements – simulating more than 50 distinct users on the SSD/HDD hybrid system. In testing, the existing system ran the tests in 14.5 hours, while the hybrid SSD system ran the same workload in just under 6 hours – a 2.5 times improvement.

Average	SSD	HDD
Service time	.78 ms	2.81ms
Utilization	89%	34%

Several customers have also found that SSD performance is reducing disk IO queues (pending IOs for a given disk) for both SSDs and hard disks. Typically with Teradata Virtual Storage, more than 80 percent of all disk IO requests are routed to SSDs. This shrinks the hard disk IO request queues dramatically. Consequently, cold data queries get access to the HDDs quicker because there is less waiting behind other requests. Moreover, the SSDs are so fast that they empty their request queues quickly and are ready for more work. The result is better throughput overall.

Of course, these results apply to workloads that are IOPS-constrained, which is where Teradata Virtual Storage provides the greatest benefits. CPU-constrained queries still require faster microprocessors and a smart database query optimizer.

Summary

Teradata believes that every organization possesses a wide range of value in the data they use for analytics. Consequently, we are pursuing a strategy to match the cost of storage to the value of data in the organization – using all classes of storage simultaneously (in-memory, SSD, performance disk, capacity disk, and archives). This hybrid storage model allows data of all levels of value to be fully accessible in a single analytics environment – with no boundaries to the users, no labor costs for data placement, and a hybrid cost to match the hybrid value of the data. Like server virtualization, virtual storage is automatically moved to the right place in the data temperature spectrum for the best price and performance combination.

Teradata is now ready to take your data temperature.

CIO Challenge: Performance vs. Cost

About Teradata

Teradata is the world's largest company solely focused on creating enterprise agility through database software, enterprise data warehousing, data warehouse appliances, and analytics. Teradata provides the best database for analytics with the architectural flexibility to address any technology and business need for companies of all sizes. Supported by active technology for unmatched performance and scalability, Teradata's experienced professionals and analytic solutions empower leaders and innovators to create visibility, cutting through the complexities of business to make smarter, faster decisions. Simply put, Teradata solutions give companies the agility to outperform and outmaneuver for the competitive edge.

- i Doug Henschen, "Teradata Boosts Data Warehousing Performance With SSDs," InformationWeek, April 13, 2011
- ii C.Walter, "Kryder's Law," Scientific American, July 2005
- iii IDC, "Extracting Value from Chaos," June 2011
- iv IDC, "The Datacenter of the Future," May 2011; Gartner, "Forecast: Memory, Worldwide, 2005-2015 4Q11 Update," Dec 13, 2011; "Flash Storage Gets Cheaper, Disk Storage Gets More Expensive," IT Jungle, Jan 16, 2012, <http://www.itjungle.com/tfh/tfh011612-story07.html>
- v Ventana Research, "The Challenge of Big Data," Dec 2011

The Best Decision Possible is a trademark, and Teradata and the Teradata logo are registered trademarks of Teradata Corporation and/or its affiliates in the U.S. or worldwide. Teradata continually improves products as new technologies and components become available. Teradata, therefore, reserves the right to change specifications without prior notice. All features, functions, and operations described herein may not be marketed in all parts of the world. Consult your Teradata representative or Teradata.com for more information.

Copyright © 2012 by Teradata Corporation All Rights Reserved. Produced in U.S.A.