

WHITE PAPER

AUTONOMIC TIERED STORAGE: ELIMINATING THE BOTTLENECK OF PERFORMANCE STARVED APPLICATIONS

Bill Mottram
VP Marketing
Atrato, Inc.

April 2010

Abstract:

Performance starved applications are top of mind in the data center. These applications are unable to reach optimal performance due to storage solutions that are unable to produce enough IOPS to satisfy read/write requests in a timely manner. Atrato has effectively closed the performance gap between storage and servers. With Atrato's hybrid solution, storage not only keeps pace with the demands of performance starved applications, but the bottleneck that previously resided on the storage system is moved back to the server. In the past servers were hindered by inefficient storage, now storage performance can transcend the needs of even the highest performance applications.

ATRATO[™]
DATA MANAGED STORAGE

www.Atrato.com

ATRATO'S AUTONOMIC TIERED STORAGE

The storage industry is very cognizant of the number of conflicting challenges that face the storage user; and Atrato is no exception. Performance is top of mind for all IT Administrators. This concern is especially present in data centers that utilize I/O intensive applications.

Performance Starved Applications

Data intensive applications are growing in number; a telltale sign of an increasingly complex business environment. These applications are comprised of extremely efficient code which effectively utilizes client initiator resources. Servers (CPUs on which client initiators run) are purpose-built to drive the maximum IPS (instructions per second) limit dictated by the processors. In fact client initiator capabilities have experienced a progressive exponential growth trend over the last 10 years. This seemingly complete picture is missing a major element. These applications are data hungry, and they must be fed. Not only must they be fed, but they must be fed at a rate such that maximum IPS can be satisfied. In other words, the storage must never keep the application waiting if peak performance is to be realized in the data center. The problem: disk-based storage systems have not been able to keep pace with the progressive nature of client initiator growth. In fact, the client initiator enjoys approximately 50% annual performance CAGR, while disk based storage systems are crawling at around 10%. An increasingly widening gap in client initiator and disk-based storage performance shows the dilemma faced by data centers today (see Figure A). For this reason these applications are referred to as “performance starved.” They demand data at rates that cannot possibly be achieved by legacy storage solutions. The performance capabilities of these legacy disk-based storage systems grow at just one fifth the pace of their client initiator counterparts. This widening performance gap creates an overall inefficiency in the data center.

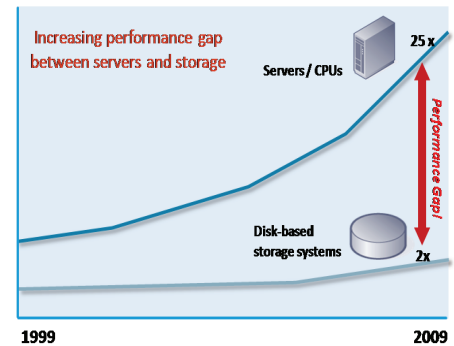


Figure A. The Performance Gap

Which Environments Are Feeling the Effects of Performance Starved Applications

This is a very widespread storage concern that touches a multitude of vertical markets. Data centers across the board are feeling the performance pinch. Some of the most prevalent include but are certainly not limited to the following.

Virtual Environments: The virtual environment consists of a multitude of virtual machines which are created and managed on each single physical server. These machines are dynamically deployed to support a variety of operating systems, databases, platforms, and applications. Virtual environments require a storage solution that eliminates the bandwidth constraints on the number of virtual machines that can exist within a virtual environment. To extend the bandwidth for a virtual environment a high number of IOPS must be available to satisfy thousands of simultaneous requests.

Virtual Desktop Infrastructure (VDI): VDI is offering new heights in data consolidation. In a VDI environment every program run, file saved, application, and process is centralized on a server instead of locally situated on an individual's desktop. In effect, employees can access all of their information from their desktops, home computers, and smart phones. Any authorized remote client (including thin clients) can access the information they need regardless of geographic location or supported device. The value in this technology is easily seen, but not as easily implemented. It is no small task to centrally store all this data and deliver it to multiple clients as rapidly as expected and required.

Transactional Databases: Transactional databases are widely used and highly accessed databases. They are also known as subject area databases. They store a multitude of data ranging from medical, financial, video, and many others. For some transactional databases fast access is paramount. Medical databases must have patient records available in an instant, financial institutions must have numbers in hand to invest correctly, and businesses must have information at the ready.



ATRATO'S AUTONOMIC TIERED STORAGE

to intelligently anticipate next steps. In any of these cases it is easy to discern the outcomes should data accessibility be compromised.

Data Analytics: Analytics are very simply put the process of pattern recognition in a given data set. Business and financial analytics dominate the data analytics space, demanding performance to support timely processing. Many important functions are carried out by data analytics. Governments use data analytics to gain intelligence, improve national security, bolster fraud detection efforts, and predict economic trends. Marketing and sales departments use analytics to profile potential customers and make profit forecasts. Geospatial applications use it to discover geographic patterns that may indicate the location of oil, coal, and other mineral deposits. These diverse applications have one unifying factor- myriads of data that must be quickly sorted, compiled, and analyzed. The software required to effectively carry out these tasks is CPU intensive and consequently performance starved in many cases. Without proper data storage I/O capabilities, companies that rely on this information in a timely manner may suffer.

Data Warehousing: Data warehousing is somewhat of a misnomer. It implies that data is kept and left in a storehouse. This is true, but it is only half the story. Data is far from stagnant in a data warehouse. In fact many data mining processes are performed on data stored in such warehouses. The sheer size of the data warehouse presents challenges in recalling data in a timely manner. Data warehouses are enormous in size, therefore, data mining queries often become performance starved.

What Solutions Are Available to Sustain These Applications

Performance starved applications are often thought of in a “one off” manner. This view may have been accurate three years ago, but the need for high performance storage to support high performance applications is burgeoning. This is not an issue confined to singular boutique applications. This is a pervasive issue whose solutions must be robust enough for adoption into the mainstream. The resurgence of high performance requirements is staggering in its intensity. Storage vendors have been forced to respond quickly. With this quick response came solutions that only offered a “work around.” These solutions include short-stroking and simplistic SSD approaches.

Short-Stroking: Short-stroking is the practice of only writing data to the disk sectors located on the outermost portion of the disk platter. This approach speeds access time by shortening the seek time required by the disk actuator. This quick fix comes with a major drawback—only 25% percent of drive is utilized. Enormous amounts of disk space and subsequently capacity is wasted (not to mention the power costs of constantly spinning multitudes of disks with 75% of unutilized capacity). In a purely performance application this waste in capacity may seem a negligible price to pay for the touted 50% decrease in access time. However, it is important to bear in mind that performance starved applications are often capacity intensive as well. When considering this angle remember that 75% of the disk's capacity remains unutilized; equating to 75% of storage purchased and wasted. To achieve a truly efficient data center, performance starved applications must be addressed with a storage solution that meets both performance and capacity requirements.

SSD: New and exciting hardware is a beautiful thing, and it is surely the crux of advancement in the data center. However, storage veterans will all agree that simply throwing hardware at a problem will undoubtedly do more damage than good. Such is the case of the famed Solid State Drive (SSD). These drives are certainly the fastest things around with the exception of DRAM. However, they are relatively new and extraordinarily expensive when compared to hard disk drives (HDDs). These drives are often mismanaged in terms of cost and expected performance gain. Effective and manageable SSD solutions must have a hearty software platform on which to run. To take full advantage of the hardware platform, the ability to understand the applications and determine where SSD is needed most is key.



ATRATO'S AUTONOMIC TIERED STORAGE

The Atrato Approach

Atrato's solution caters to performance applications. Atrato has been addressing the need for high performance storage since its inception. Its innovative Data Managed Storage (DMS) architecture addresses the varying challenges faced by those experiencing performance starved applications. Three pervasive and recurring challenges Atrato has uncovered are as follows:

- 1) the need for storage performance that eliminates I/O bottlenecks,
- 2) scalability that meets or exceeds growth demands for active data, and
- 3) an architecture that is sustainable in terms of cost (power, cooling, and space management).

Atrato's innovative portfolio of technologies uniquely addresses these challenges to alleviate the performance pain points experienced in the data center.

The V1000: Atrato's flagship product, the Velocity1000 (V1000) uses an extremely dense form factor to deliver extreme performance and capacity. The V1000 is capable of solving many performance stumbling blocks. With 160 HDDs, the V1000 delivers an industry-leading 16,000 random IOPS in a 3U enclosure. In addition Atrato also features 3 years of zero maintenance on drives. This guarantee is based on the V1000's self-healing architecture which includes preventative drive maintenance, error handling, hot spare management, virtual sparing, background drive scrubbing, and duty cycle management. (See Atrato V1000 datasheet for more information.)

The VxSSD: The VxSSD is the solid state hardware upgrade component to the V1000. The data explosion is cresting to a point where purely spinning disk storage solutions cannot address every need within the data center. For performance starved applications Atrato delivers terabytes of solid state disk for added acceleration. Each VxSSD enclosure supports 10 SSD drives to increase IOPS and bandwidth capabilities. The VxSSD and the flagship V1000 work together to create a hybrid hardware environment that, when coupled with the innovative ApplicationSmart software, caters to the performance requirement of each individual application.

ApplicationSmart: The challenge of effective data management remains the ability to match the value and requirements of the application with the value and performance of the storage. Resolving this challenge is the motivation behind storage tiering. The elimination of the administrative complexity that storage tiering can introduce is the catalyst driving advances in automation. In the case of Atrato the technology has moved beyond traditional automation to a level best described as autonomic—the ability to manage or improve operation with little or no human involvement. To keep pace with the rapid expansion of data Atrato has made available the ApplicationSmart software suite and the VxSSD 10 hardware component. This new offering is an addition to the V1000 product line, and can be used as an upgrade to the flagship V1000 or part of an initial implementation. The solution allows companies to scale in line with their performance needs. The newly expanded Velocity series is a unique approach to hybrid SSD usage and tiering.

While other storage vendors take a more policy-based approach, Atrato has integrated powerful software to provide autonomic tiered storage management. This added intelligence helps to optimize performance, minimize cost and remove the management complexity found in traditional tiered storage solutions. The ApplicationSmart software allows users to implement SSD and bring true autonomies to their tiering environment. As part of the software suite, Atrato offers the following unique management components:



ATRATO'S AUTONOMIC TIERED STORAGE

Data Access Profiler (AP): The AP eliminates much of the management complexity in creating and managing storage tiers. AP identifies highly accessed data and creates real-time access profiles. It then uses the information to determine the initial amount of SSD needed and the estimated level of performance increase from Tier 0. Once the tiered storage environment is in place the AP will continuously monitor data access patterns and make recommendations to continually promote highly accessed data to the solid state tier.

Tiered Management Engine (TME): The TME eliminates the management complexity inherent in performing manual batch processing, which is the approach most commonly used by other alternative storage solutions. The TME is responsible for replicating data stored in Tier 1 to Tier 0 based upon recommendations from the AP. It provides block level management for increased responsiveness and granularity. The TME also identifies which data residing in SSD is no longer qualified as highly accessed, and performs the write-over function; effectively replacing less frequently accessed data with higher accessed data.

Egress Accelerator (EA): The EA speeds up sequential I/O and lowers read latency. The EA is essentially a read-ahead cache located in physical RAM within the controller. When data is requested it pulls the next several blocks of data into the EA making the assumption that data that was written together may very well be read together. The EA provides another alternative to searching the general storage pool when a write request is made.

Ingest Accelerator (IA): The IA optimizes the value of the client initiator by storing write requests in SSD and declaring the request complete upon receipt. Acting as an ingest log, it eliminates the wait time the client initiators face while data is being written to the general storage pool. The IA will reform I/Os within the solid state before sending them to the spinning disk while client initiator requests are low for optimal write speeds.

Solid State LUN Manager (SLM): The SLM allows users to configure pure SSD VLUNs. The SLM is ideal for permanently allocating highly accessed data such as metadata and database indexes in solid state. This feature allows the customer to use SSD's as the general storage pool. The SLM effectively extends the performance of solid state to all data regardless of access pattern. SSD VLUNs are RAID 10 or RAID 50 protected.

Hybrid VLUNs: Hybrid VLUNs bring together the advantages of SSD and HDD, to quickly respond to changing access demands. Within a hybrid storage pool, highly active blocks are replicated to the SSD layer to effectively supercharge the data. This added performance boost satisfies workload spikes or changing peak loads, eliminating the contention created in traditional storage arrays. Autonomic software simplifies management and transparently adapts to customers' unique data access and application demands.

This unique and robust solution set mitigates the management complexity associated with deploying SSD solutions. With Atrato's Velocity and VxSSD series, companies can enjoy the low cost of HDD drives for less accessed storage, and the performance gains of SSD for their highly accessed data, all in a tightly integrated hybrid solution. The autonomic tiering provided by the system makes it far easier to implement and manage when compared to other hybrid solutions available in the market today. (See data sheet for more information on ApplicationSmart, the Velocity and the VxSSD series)