Big Data Analytics and Flash-Based Storage:
Smaller, Simpler, Faster and Better

CONTENTS:

Data has become the lifeblood of business and its exponential growth offers opportunities as well as risks for a significant number of enterprises. In today's technology-rich business and consumer environments, vast amounts of data are generated through personal (social networks, mobile devices, Web commerce), societal (geolocation, images, media) and industrial (digital sensors, automation) interactions. As a result, companies are increasingly challenged when it comes to processing and managing the massive amounts of digital information flowing in and out of their data centers.

Moreover, a considerable portion of an organization's OpEx is dedicated to IT personnel and resources needed to run today's data center. As support costs skyrocket, these companies are identifying flash-based storage as a compelling alternative. In addition to a smaller data center footprint, they're gaining measurable benefits via simplified management, reduced application response times, and decreased power consumption at the system, rack, and data center levels.

In this white paper, we look at the specific advantages of flash adoption, the high performance level of SSDs, and how this technology is providing businesses with a competitive edge.

TechTarget Custom Media

Micron

Flash-Based Storage: Simpler Approach, Long-Term Benefits

Currently, there are divergent definitions of big data. On the one hand, the standard approach consists of a conventional Hadoop-distributed file system where MapReduce query results flow into a Cassandra-managed database. Analysts then mine those results using Cassandra's storage location.

But there's another, broader concept. An alternative view focuses beyond Hadoop. It derives actionable results from any extraordinarily large, unstructured data set through implementing flash-based storage to deliver new levels of speed, fault tolerance, and cost efficiency.

Such capabilities are necessary for scaling on demand and flexibly meeting the processing requirements of big data. This is particularly true as it relates to internal process logging as well as results logging and analytics. Especially if there's a need for real-time, decision-oriented processing or critical business-oriented outcomes stemming from that data.

Over time, it's become increasingly clear that traditional rotating media is less effective for meeting the needs of these big data and analytics applications, and for providing the optimal level of response: low latency, high capacity, and extreme resiliency.

In the current Hadoop framework, most data mining consists of real-time read/write requests. When employing Cassandra-based data analysis, it's necessary to reach multiple tables in numerous places across several storage locations to reassemble the record. This often leads to extremely random, fragmented, and highly distributed results with inaccuracies and inconsistencies that can be time-consuming to resolve.

In addition, employing traditional rotating media for big data and analytics applications often results in latencies that adversely impact the net value of the mining results—too little information too late. These can become a key concern when performing data mining under strict time constraints. In contrast, cost-efficient, enterprise-ready flash excels at supporting these types of operations, offering near-instantaneous scalability and high throughput.

How SSDs Reduce OpEx

As companies move beyond traditional, legacy hard disk drives (HDD) and migrate to a higher performance solid state infrastructure, they're achieving important gains. In addition to reaching new performance levels, these organizations are using fewer nodes to perform the same work in the same amount of time.

This is especially important in the case of data centers that have limited physical capacity. With the traditional Hadoop ecosystem, administrators simply add more actual servers to reach the desired level of performance. But what's the value of an extra unit of rackspace to an organization in a space-constrained data center?

The issue is further complicated when the number of watts allocated per rack is capped. Even if there's more work to do inside that rack, an organization cannot exceed a certain amount. Flash storage solutions help alleviate such concerns.

In fact, with fewer nodes to accomplish the same amount of work, organizations can realize a decrease in power consumption. Such reductions are significant, especially for enterprises located in regions and metropolitan centers where energy expenses are considerable.

The greater thermal efficiency of flash means that data center managers can reduce OpEx in terms of cooling costs as well. Finally, since flash-based drives process workloads faster, they transition to a low-power state more quickly.

Companies are achieving reduced OpEx because these smaller flash-based clusters are also easier to manage. High performance, enterprise-ready flash adoption means not only dedicating fewer personnel hours for maintaining big data initiatives, but also the reduced expense of having to stock fewer spare server components.

Advantages of the Micron Portfolio and the XPERT Feature Set

Micron Technology offers an SSD portfolio that covers every aspect of today's data center. The Micron portfolio ranges from very resilient, high-read performance, boot-centric devices for general purpose servers through high endurance, extremely fault tolerant SSDs for both internal process logging, results logging and later analytics. The storage architecture enhancement extends through to PCIe-attached devices with extraordinary low latency for customers' main data sets.

As a single provider offering a comprehensive selection, Micron excels at meeting the needs of a diverse customer base. In addition to clients who must quickly process large data sets, Micron provides resources to specialized big data companies who use flash to drive real-world business innovations. These companies require the processing of petabytes of unstructured data as they perform an array of services, such as full-motion video advertising, highly secure user identification for the financial industry, targeted marketing, and instant online ad exchanges.

Cost-efficient flash storage media, such as PCIe SSDs, enables these companies to perform immediate real-time data analysis and instant decision-making, then seamlessly respond, all in a matter of milliseconds. Attaining these levels of speed and resiliency with traditional HDDs would be nearly impossible, or else so expensive it would be impractical for any competitive enterprise.

Partly in response to increased data consumption and growth, Micron has introduced the eXtended Performance and Enhanced Reliability Technology (XPERT) suite of features for its enterprise-class SSDs to broaden performance and reliability. XPERT intelligently integrates the storage media and controller into a single, comprehensive architecture, thus extending drive life, protecting data during power failures, and ensuring overall data integrity. Storage media are designed with XPERT architecture enhancements to precisely meet the big data and analytics application requirements of each customer.

Benefits of Redundant Array of Independent NAND (RAIN)

Principle among the XPERT modalities, the Redundant Array of Independent NAND (RAIN) provides data protection well beyond common error correction code (ECC) use cases. RAIN implementations are design-specific and they embed protected data with user data. Operating in real time as a parity-protecting mechanism, RAIN ensures the following safeguards:

- Data-to-Parity Ratio: The X:Y ratio or stripe size is optimized for intended drive, workload, or performance.
- Parity Storage: The location for parity may be fixed, relative, or rotating.
- Protection Level: Helps avoid catastrophic media failures.
- Hardware Acceleration: RAIN is self-managed, transparent to the host platform, and accelerated inside the SSD controller hardware resulting in zero impact on performance.

For additional information on Micron's XPERT feature set, see the **Technical Brief**

Conclusion

Increasingly, enterprises are seeing the value of solid state solutions. They're realizing positive net business value per dollar investment in flash-based storage. And they're finding that adoption offers both a simpler approach and long-term benefits.

It's no secret that traditional storage media are increasingly inadequate for meeting the current big data and analytics goals of enterprises. As projects that generate large, random unstructured data sets become more common, companies are searching for ways to ensure low latencies and achieve more with less. Moreover, new innovative enterprises are pushing the limits of big data possibilities by mining petabytes of data, retrieving near-instantaneous query results, and responding to their customers within nanoseconds.

Flash-based storage can deliver results for both clustered applications typically found in conventional big data ecosystems as well as for these more complex, time-sensitive and decision-oriented processing requirements. As a single supplier, Micron offers clients high-level expertise, advice, and support, with an understanding of both client workloads and design needs.

In addition to meeting current big data requirements, your company can achieve reductions in power consumption and operating expenses as well as minimize the personnel hours spent managing workloads. It's easy to begin a conversation. Contact Micron at **SSD@micron.com**, connect on **Twitter at @MicronStorage** and read the Micron Storage blog at **www.micron.com/storageblog**.