



Analytics

Analytics applications are rapidly becoming the key applications for Big Data workloads. Analytics applications address the large data-sets that are generated by transactional processing to find the patterns in the data that can be leveraged to take decisive action in a fast-moving marketplace.

Analytics is an umbrella term used to describe a number of specific workloads that are widely deployed within financial services companies. These workloads are needed to cope with the data tsunami that is hitting financial services firms—generated from a variety of data sources. Customers must quickly find the patterns in the data in order to make accurate and timely business decisions.

Hadoop is a leading software architecture that allows customers to identify the key data-points in extremely large multi-terabyte (TB) datasets. SanDisk® evaluated a Hadoop six data-node cluster, using the Terasort workload, to determine the impact of running those workloads on flash-enabled servers.

With Hadoop, the master server is the one that maps the computing tasks to specific servers – making it possible for each individual server to perform well, while adding more servers to the cluster.

Testing Hadoop

SanDisk has run Terasort benchmark tests on Hadoop servers, to see how solid-state drives (SSDs) can accelerate Hadoop, as it is running in real-time on servers.

In a test of a six data-node cluster, the Hadoop instance supporting a 1TB dataset running across all six nodes achieved results 32% faster at 15% less cost when compared to traditional hard-disk drives (HDDs).

“Analytics is rapidly becoming the key application for Big Data workloads”

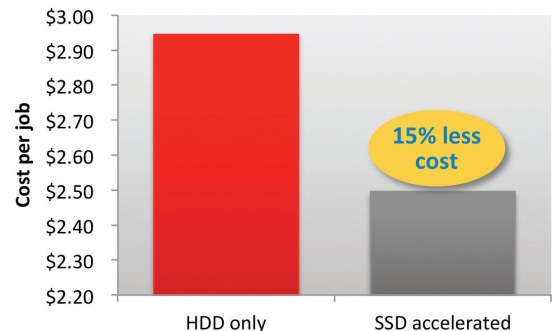
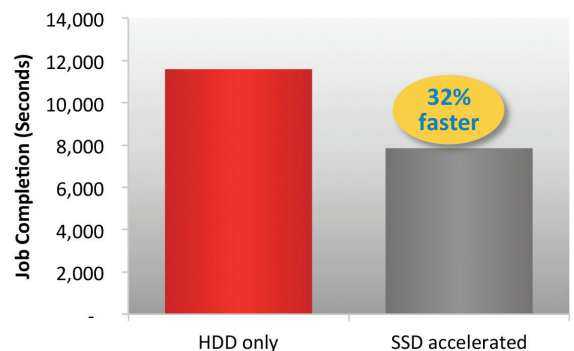
Jean S. Bozman

Hadoop

Hadoop is well-known both for its applicability to financial services applications, and for its ability to “scale up” along with the number of servers attached to a Hadoop cluster. The ability of Hadoop to work with large data-sets, and to parse out the computing tasks – mapping them to servers within the cluster – accounts for its wide adoption within the financial services world.

Parallelized workloads, like Hadoop, are ideally suited to a scale-out computing world, in which more servers can be added, as needed, as demand for computing increases. In fact, with cloud computing, customers can tap into the processing power of more than 100 for compute-capacity, if needed.

6 node Hadoop Cluster Example



These results are shown for a six data-node Hadoop cluster, but the findings can be applied to larger clusters, with more server nodes included. All of the Hadoop processes, including loading the data, sorting the data, and completing the computation, benefits from the use of flash SSDs.

Advantages of Flash

Flash technology accelerates the performance of Hadoop clusters, and its benefits are extensible, as the Hadoop cluster expands through the addition of nodes. The unique design of Hadoop software offloads the increasing data traffic from the master node to the individual nodes for processing – and then gathers the results. Customers who acquire flash-enabled servers will see performance benefits, with dramatically reduced latency for I/O – improving the time-to-results.

Using SSDs brings a number of advantages to customers in terms of CapEx and OpEx costs. First, in deployments with SSDs, fewer servers will be needed to deliver the same storage capacity as server deployments leveraging HDDs. The performance characteristics of SSDs make them much less subject to the response time issues that affect HDDs. Operational expenditures are less, because the number of servers required within the data center is less.

With fewer drives, and fewer systems required, power and cooling costs are lower than for an HDD-based server solution. SSDs save time and money,

because they reduce latency, while improving quality of service (QoS). And with no moving parts, SSDs don't experience failures due to mechanical parts wearing out. In terms of high availability for mission-critical data, SSDs' non-volatile memory preserves data, reducing time to recovery from outages.

Summary

The digital universe is expanding – creating new demands on those who must analyze it, and take actions based on the analytics results. SanDisk SSDs can be put into use immediately through simple on-site replacements of existing HDDs. Or, SSDs can be acquired as built-in devices inside OEM systems vendor products that are being acquired for new projects.

For technology refresh, SanDisk SSDs plug into standardized interfaces for SAS, SATA, and PCIe directly—so they fit into existing data center systems with no disruption of the infrastructure. New deployments bring the benefits of flash technology, as well. Flash SSDs are built into the servers being acquired from major systems vendors worldwide. SanDisk SSDs are being shipped by 6 of the top 7 server and storage OEMs worldwide.

Fast-paced financial markets value technology that allows them to analyze transactional data—and to predict where the market is heading. Solid state drives provide rapid processing, and shorter time-frames to meet customers' quick decision horizons.

Contact information

ESS.channelsales@sandisk.com
ESS.OEMsales@sandisk.com
businesspartners@sandisk.co

Western Digital Technologies, Inc.

951 SanDisk Drive
Milpitas, CA 95035-7933, USA
T: 1-866-744-2165

Western Digital Technologies, Inc. is the seller of record and licensee in the Americas of SanDisk® products.

For more information, please visit:

www.sandisk.com/enterprise

SanDisk®

a Western Digital brand

At SanDisk, we're expanding the possibilities of data storage. For more than 25 years, SanDisk's ideas have helped transform the industry, delivering next generation storage solutions for consumers and businesses around the globe.