Innovative devices, sensors, streaming video and many other sources are accelerating the volume, variety and velocity of data that is ripe for big data analytics. Characterized by enormous data sets beyond the capabilities of conventional software tools to capture, manage and process within acceptable time periods, big data offers unprecedented insight into strategic planning and growth opportunities. As a result, IT leaders across many industries are responding to the competitive urgency for launching big data initiatives that tap into massive repositories of structured, semi-structured and unstructured data.

At the outset, many IT organizations are discovering that they do not have the computing technology or power to adequately collect, sort and categorize data mined from stores containing terabytes, petabytes and even exabytes of data. To capitalize on these rich information sources, organizations require scalable technologies that enable them to store, manage and access data designed to improve business intelligence without increasing operating expenses.

Accelerating data capture
In today’s business environment where decisions are based on dispersed, dynamic information sources, access to big data for analytics must keep pace. An in-memory approach expedites data capture by loading an entire database into main memory on the database server. Access to data stored in a database server’s RAM is substantially faster than access to data stored on hard disk drives (HDDs) — typically measured in nanoseconds, versus milliseconds.

Another highly efficient alternative is flash-based storage or solid-state drives (SSDs) that are designed to significantly increase the performance of servers, such as those used for big data tasks. And compared to HDDs, SSDs help reduce power consumption. Moreover, SSDs are significantly faster, with access time in microseconds, than traditional disk-based arrays — even high-performing arrays. They are designed to provide an optimal solution for organizations that want rapid access to frequently used data taken directly from memory — or a warm cache — and access to data used less frequently from disk drives — or a cold cache.

Dell™ PowerEdge™ servers — equipped with fast, power-efficient Samsung® Green Double Data Rate 3 (DDR3) dynamic RAM (DRAM) and Samsung Green SSDs — can run database software capable of handling large amounts of warm and cold cache data. These servers, in particular, offer scalable platforms for database-intensive applications and help deliver the bandwidth needed for moving data quickly across the enterprise.

Comparing warm and cold cache access
To test data access and analysis capabilities, Samsung and Microsoft performed proof-of-concept (POC) testing in June 2012 at the Microsoft Technology Center (MTC) in Timely analysis of information derived from big data volumes depends largely on how fast the data can be accessed. Dell, Samsung and Microsoft put access speed to the test using advanced servers with energy-efficient memory and solid-state drives.

By Peyman Blumstengel, Marc Crespell, Sylvie Kadivar, Yogev Shimony and Jessica Titus
In addition to power consumption efficiency, the testing measured the duration of queries in four scenarios that represent key data access and analysis tasks using the TCP Benchmark™H (TCP-H) decision support benchmark.

The test configuration consisted of two Dell PowerEdge R910 servers powered by Intel® Xeon® E7-4870 processors with Samsung Green DDR3 DRAM. One server was set up with 1 TB Samsung Green 20 nm–class DDR3-1066 memory modules in a high-performance, energy-efficient configuration; the other server was equipped with mainstream, 50 nm–class DDR3-1066 memory modules in a standard configuration. Each server was set up to dual boot Microsoft® Windows Server® 2008 R2 Service Pack 1 (SP1) and Windows Server 2012 Release Candidate operating systems, and each ran the Microsoft® SQL Server® 2012 database engine. The high-performance, energy-efficient configuration included Samsung Green SSDs and Samsung PM830 Serial ATA (SATA) SSDs for a total capacity of 4 TB. The standard configuration was equipped with 15,000 rpm Serial Attached SCSI (SAS) HDDs for a total capacity of 3.2 TB.

The test results demonstrated that the high-performance, energy-efficient configuration was 15 times faster than the standard configuration for data access in a cold cache environment. This configuration saved 94 percent of system power consumption (see figure).

The high-performance, energy-efficient system demonstrated 2 percent faster data access than the standard configuration in a warm cache environment in which almost all the data was located in memory. This configuration saved 30 percent of system power consumption (see figure).

Mining big data quickly and efficiently

In this test study, a Dell PowerEdge R910 server in a high-performance, energy-efficient configuration with Samsung Green DDR3 and Samsung Green SSDs running Microsoft software demonstrated a data access speed well suited for big data analytics. These test results indicate that the POC test configuration is designed to deliver rapid access, management and processing of data that can be mined for enterprise-scale big data analytics within a cost-efficient power envelope.

Authors

Peyman Blumstengel is a green evangelist at Samsung Semiconductor Europe.

Marc Crespell is a global alliances manager at Dell.

Sylvie Kadivar is a senior director of strategic marketing at Samsung Semiconductor Inc.

Yogev Shimony is a business development manager in Global Partner Development at Dell.

Jessica Titus is a senior manager of strategic marketing at Samsung Semiconductor Inc.

Learn more

Dell PowerEdge servers:
Dell.com/poweredge

Samsung Green Memory:
www.samsung.com/greenmemory

Samsung IT ecosystems:
qrs.ly/e83bu4j