Western Digital

Storage for MongoDB® Analytics

MongoDB is a leading open source, NoSQL database tool for meeting the challenge of extracting value from Big Data. Adding Ultrastar helium-filled HelioSeal® hard disk drives (HDDs) and either Ultrastar solid-state drives (SSDs) to the solution reduces costs and accelerates business results.

What Western Digital storage can do for MongoDB

- Ease dealing with exponential growth of customer data
- Decrease time-to-results by increasing query speed with SSDs
- Balance performance needs with cost constraints
- Maximize storage capacity per server

Storage Solutions for MongoDB Analytics

Analytics is a key method of extracting value from Big Data. Analytics applications address the large data sets that are generated by transactional processing, real-world sensor data, and other high-velocity, high-volume data sources. The analytics process finds the patterns in the data that can be leveraged to take decisive action in a fast-moving marketplace. MongoDB with Western Digital's Ultrastar® drives can allow a deeper and timelier analysis of large datasets, delivering real business value.

Analytics

Analytics describes varied workloads that are widely deployed within financial services, manufacturing, Internet of Things (IoT), web development, and other industries. These workloads are needed to cope with the data tsunami that is hitting companies that generate a variety of rich, unstructured data sources. Customers must quickly find the patterns in the data in order to make accurate and timely business decisions. MongoDB allows customers to identify key data points in extremely large petabyte-scale datasets of unstructured data.

MongoDB

MongoDB is an open source, scale-out NoSQL database. Unlike traditional relational databases that store structured data, MongoDB enables unstructured or schema-less databases. As applications evolve, many types of data—including photos, videos, audio files, and other types of data—can be analyzed more efficiently with NoSQL than with traditional SQL databases.

The business benefits of this approach are clear: NoSQL databases are highly scalable for growing data sets and can, therefore, operate and manage massive amounts of data. Most importantly from an application architect perspective, they can address multiple data types, including structured, semi-structured, and unstructured data, all within the same data repository. Creating and analyzing data lakes is fully supported, which is important when managing social media files with photos, sensor data, historical behavior patterns, and other data in varying formats.

Customers use MongoDB for many aspects of their business, including risk analysis and reporting, distribution and synchronization of data across geographically separated sites, market data management, real-time sensor streams, and aggregation of customer history and behavior. "Auto-sharding" of data makes it easy to distribute the terabytes or petabytes of data under MongoDB management across many hundreds of physical servers. That power, however, creates the potential for massive server sprawl and related issues, as well as performance problems as managed data grows under each server.

Ultrastar HelioSeal Hard Drives for MongoDB

Even with the dramatic increase in flash density and the accompanying cost reductions, the most cost-effective method for storing petabytes of MongoDB data is still on hard disk, and in particular high-capacity Ultrastar helium-filled hard drives based on HelioSeal technology.

Pain Point: Exponential Data Growth

By using Ultrastar helium HDDs with the industry-standard SAS interface and high-density JBODs, an individual MongoDB server can store massive amounts of historical data while providing exceptional cost, power, and rack space performance. Multiple hard drives per server can provide exceptional performance to support many MongoDB streaming queries.

Ultrastar SSDs for MongoDB

While using HDDs is a powerful and cost-efficient method of managing MongoDB datasets in many cases, the method has inherent limits when the data has a high update frequency or there are multiple processing jobs running in parallel. For these workloads, flash-based solid state drives (SSDs) can provide a significant performance advantage while maintaining a reasonable cost.

Pain Point: MongoDB Query Bandwidth Starvation

Ultrastar SSDs can accelerate MongoDB server-tier performance, resulting in faster query turnaround and quicker business insights. Under the hood, SSDs accelerate most I/O workloads by reducing latencies of individual I/O operations and providing vastly higher bandwidth per device than hard drives. For example, a single Ultrastar DC SN630 NVMe™ SSD, with a direct link to the server CPU over PCI Express, can provide the read performance of as many as 10 standard hard drives.

Pain Point: Scaling Servers for Storage Performance

SSDs can bring capital and operating cost advantages to performance-starved customers. As described above, fewer SSDs may be needed to deliver the same or better application performance level than HDDs would. The performance characteristics of SSDs make them much less subject to the response time issues that affect HDDs in multiple-query workloads. For smaller installations, Ultrastar DC SN630 SATA SSDs can replace HDDs directly. When larger datasets are required, large quantities of Ultrastar DC SS530 SAS SSDs may be connected to an existing SAS infrastructure to provide high-performance, high-capacity storage.

Pain Point: Ongoing Costs of HDD-Based MongoDB

Operational expenditures are also reduced when SSDs are used, because fewer servers may be required to produce the same outcome, more quickly, and with a higher quality of service. This higher quality of service can provide real business value by supporting higher-grade SLAs.

Summary

The digital universe is expanding, creating new demands on those who must analyze the data and take actions based on it. Storing this expanding digital universe is a job for both hard drives and solid state drives. Ultrastar helium HDDs can provide the massive amounts of bulk storage needed for MongoDB historical analysis, while judiciously applied Ultrastar SSDs can accelerate specific performance-critical applications.

Pain Point	NVMe SSD Ultrastar DC SN630	SAS SSD Ultrastar DC SS530	SATA/SAS HelioSeal HDD Ultrastar DC HC500 Series
Exponential data growth		●	● ● ●
MongoDB query bandwidth starvation	● ● ●	● ●	
Scaling servers for storage performance	● ●	● ●	
Ongoing costs of HDD-based MongoDB	●	● ●	

Legend: ● = Good ● ● = Better ● ● ● = Best

Western Digital.