Promise of Low-Latency Stable Storage for Enterprise Solutions

Janet Wu

Principal Software Engineer

Oracle

janet.wu@oracle.com

Latency Sensitive Applications

- Sample Real-Time Use Cases
  - Real-time quote service, algorithmic trading, fraud detection (financial analytics)
  - Online travel agent (access to reservation data)
  - Cell call usage metering (>100,000 concurrent calls to monitor & centralize)
- Industries
  - Telecom/networking, trading & financial services, travel and logistics, bidding, real-time enterprises
- Financial Impact
  - Amazon loses 1% of sale for every 100ms it takes for the site to load*
  - Shopzilla reduced loading time from 7 seconds to 2, resulting in 25% pageview increase and 9.5% revenue increase**

Need for Stable Storage

- A subset of performance applications have mission-critical Service Level Agreements
  - Must be highly-available, fault-tolerant while delivering minimal latency during high transaction volumes
- *Stable Storage* – a storage abstraction that survives failures, including transaction failures, system crashes, or disk failures
- Approximated by Replication & Redundancy: maintaining multiple copies on distinct media
- Software robustness against system, hardware, or power failures

Latency Reduction Techniques

- Traditional disk-based short-stroking, high RPM disks, hybrid SSD/HDD storage
- Software painstakingly written to avoid disk latency
  - Bloom filters, caching, batching
- Much of existing database & file system designs assume disk-based data and optimize around disk mechanisms
  - optimization algorithms, buffer pool management, indexed retrieval techniques
- Use lots of DRAM – ex. Oracle TimesTen
  - In-memory DB, can be directly linked into an application's address space to eliminate IPC overheads
  - Memory optimized algorithms & data structures
  - Without NVRAM still requires disk-based commits for recovery (logging, checkpointing) which can impact transaction response time and system throughput

Data Accuracy & Availability

- Databases have numerous checks to ensure data integrity & data consistency – write-ahead logging, referential integrity checks, block checksum option, data structure consistency within blocks
- Mirroring or other redundancy such as remote replication can be part of or outside the database itself
- Backup or mirror used to recover corrupt or partial blocks
- Component hot-plug and component monitoring (e.g. wear) required for system availability
- Not all IOs are created equal

Requirements Summary

- Enterprise applications (financial in particular) require high performance with low and predictable latency
- Mission-critical applications require maximum availability and have zero tolerance for service outages, data loss, or data inaccuracies
- Applications must meet reliability & performance SLAs despite hw failures, system errors, or varying system loads across global time zones
- Financial application SLAs guarantee max latencies of several milliseconds
- ROI - standardized solutions preferred over closed & proprietary frameworks for easy integration with existing/new services (e.g. backup, snapshot)

Flash for Enterprise Applications

- Oracle Open World "*This is your Database on Flash: Insights from Oracle Development*"
- Great IOPS/$ but….
- Unpredictable and sometimes very slow outliers cause IO storms and impacts striped performance, very high thread count required for peak IOPS
- FMS 2011 Oracle Keynote "*Mythbusting Flash Performance*"
- Lack of standardization results in a plethora of proprietary implementations; high adoption cost and poor ROI

Flash for Enterprise Applications

- No standardized way to enforce QoS around complex Flash Translation Layers (FTL)
- Flash not always fast enough; memory-based solutions still dominant in the most demanding enterprise applications

NVMe Advantages

- Efficient, low-overhead, low-latency protocol
- Deep queues & parallelism enables extremely high transaction rates per device
- Enable application SLAs with dedicated queue pairs and interrupts
- Standard *de jure* atomicity
- Multi-pathing support
- End-to-end data integrity support & compatibility with T10 DIF
- Optional metadata for use by applications

NVMe Advantages Cont'd

- Consistent management interface for inherent memory characteristics such as wear, temperature, health/usage statistics
- ROI protection - standard-based and compatible with future non-volatile memory technologies

Promises of NVMe

- With NVMe devices, enterprise applications will be able to
  - Continue to reduce latency for business-critical & mission-critical transactions
  - Continue to meet reliability SLAs with end-to-end integrity and multi-pathing support
  - Match application parallelism /priority-level/access-granularity with device parallelism/ priority-level/access-granularity
  - Employ a single driver instance compatible with all devices thus simplifying testing and application integration
  - Preserve ROI using an interface compatible with future NVM technologies

References

- *Greg Linden, Amazon, “Make Data Useful”
- **Phil Dixon, Shopzilla, Velocity 2009