

WHAT'S NEXT FOR THE FINTECH UPRISING?

FOR A NUMBER OF YEARS NOW, real-time trading infrastructures have been designed and deployed assuming that storage technology is too slow, and that no individual component can be trusted to provide a reliable service. Failure management is generally built in to the design above the infrastructure layer, combined with double or triple infrastructure resilience at every level, with all the associated costs. While this cost model may be appropriate for real-time trading, that only represents a tiny fraction of the applications and processes in the FinTech landscape.

Storage is the slowest part of any infrastructure. Initially reliant on disk technology and more recently augmented with SSD technology to try and boost performance, in reality storage is still orders of magnitude slower than the networks and processors that complete the infrastructure.

The performance of storage directly impacts the performance of the business process. When new data is written and absolutely must be committed, storage write performance is key. And when applications randomly interrogate large datasets, most of the data will be accessed from storage. Whilst most database models attempt to keep as much data as possible – and certainly the most frequently accessed data - in server DRAM, there are many instances where that simply is not possible.

This is where flash technology comes in. The service time or latency is tiny compared to disk technology. Enterprise class flash solutions can deliver hundreds of thousands of transactions with a latency lower than 500 microseconds – an almost real time response – and crucially this technology is now at a price point where all T1 and T2 applications could be moved to an 'all flash infrastructure' today.

Alongside the obvious benefit of faster application performance, the total infrastructure cost can also be dramatically improved. In the traditional model, slow storage service times have led to storage being over-provisioned to deliver lower contention at the storage layer, and servers being over-provisioned because much of their useful processing cycles are idle, waiting for storage. In an all-flash infrastructure, storage is provisioned on a 'capacity needed' basis rather than a 'performance-needed' basis, cutting



Steve Willson
VP Technology Services
EMEA, Violin Memory



“ENTERPRISE CLASS FLASH SOLUTIONS CAN DELIVER HUNDREDS OF THOUSANDS OF TRANSACTIONS WITH A LATENCY LOWER THAN 500 MICROSECONDS – AN ALMOST REAL TIME RESPONSE – AND CRUCIALLY THIS TECHNOLOGY IS NOW AT A PRICE POINT WHERE ALL T1 AND T2 APPLICATIONS COULD BE MOVED TO AN ‘ALL FLASH INFRASTRUCTURE’ TODAY.”

real deployed storage capacity by up to 50%.

Removing the bottleneck at the storage layer also allows the hyper-consolidation of databases on far fewer instances and servers. This often drives server utilization (and associated licensing) up from typically 20% to 80%, providing the opportunity to cut total deployed cores, again leading to potential cost savings.

Finally, the remaining unvirtualised databases that have relied on a proprietary dedicated compute farm can now be migrated safely to a virtual environment. Tests show that databases (random read, sub 20% write load) that are moved from physical to virtual environments suffer a 15% to 20% drop in performance when no other optimisations are made. The performance benefits that flash delivers to the same application set far offsets this penalty, enabling the move to a fully virtualized datacenter to become a reality, even for the mission-critical high performance applications that have been ring-fenced.

Flash memory will become the standard platform for active data in FinTech over the next couple of years, and we will start to critique flash storage just as we did disk-based storage arrays over the last 15 years. Not all technology is equal, even where the underlying components are similar. Performance is important, but cannot be achieved at the cost of availability, manageability or application integration. Data will need to be replicated, data and snapshots will have to be consistent. The environment will need to integrate with hypervisor functions, and be manageable from existing toolsets. And it needs to do this while delivering sustained low-latency performance.

FinTech Companies are starting to exploit all-flash arrays in areas as diverse as back testing, value at risk, compliance and news/sentiment analysis. Some believe this is the tipping point to a world where data will live on flash memory unless it can be proved to be adequately served by disk. A polarized world of flash and ‘trash’ will exist, where inactive data that must be kept but often not accessed will live on the cheapest densest object stores, with all other data living in an all-flash environment.

As with most technologies, what looked niche yesterday becomes standard tomorrow. And FinTech companies are usually at the forefront of innovation.



_Steve Willson
VP Technology Services EMEA,
Violin Memory

