# Dell PowerEdge R730xd Servers with Samsung SM1715 NVMe Drives Powers the Aerospike Fraud Prevention Benchmark

***Testing validation report prepared under contract with Dell***

## Introduction

Credit card fraud prevention is among the most time-sensitive and high-value of IT tasks. The databases are large, the queries are complex, with many millions of transactions happening at any given moment and every transaction must complete in a very limited amount of time. Until recently, applications of this complexity and scale required multiple copies of data across several databases, with lots of compute and storage resources backing them. Real-time transactions are often not possible, and businesses are forced to make tradeoffs between performance and reliability of their fraud prevention services.

New technologies are coming to the rescue. NoSQL databases like Aerospike are designed from the ground up to take advantage of the newest server and storage offering from Dell and Samsung. By harnessing the extremely fast performance of Non-Volatile Memory Express (NVMe) SSDs, data can be delivered to fraud prevention engines at speeds that were previously impossible. By standardizing the interface for PCIe solid state drives with performance enhancing features like optimized I/O queuing and support for parallel operations within each I/O queue, NVMe devices enable extremely low I/O request latency. When coupled with large bandwidths and throughput for each individual SSD, NVMe SSDs significantly reduce response times for complex workloads and real-time operations on large datasets.

Dell, a leading vendor of enterprise servers, commissioned Demartek to validate the performance of a complex fraud prevention analysis workload powered by the Aerospike NoSQL database running on a cluster of Dell PowerEdge R730xd servers with Samsung SM1715[1] NVMe SSDs. The Aerospike distributed NoSQL database is optimized for extremely fast transactions on flash storage, making the PowerEdge R730xd server and SM1715 NVMe drives an ideal platform to deliver exceptional transactional performance for a robust application experience.

[1] Dell certifies the Samsung SM1715 NVMe SSD as compatible hardware for deployment in PowerEdge Servers

We saw two million Aerospike NoSQL database read operations executed with 95% of response times under 7 milliseconds while nearly five hundred thousand write operations were occurring simultaneously.

## Products Included in this Analysis

### Dell R730xd server

The Dell PowerEdge R730xd is a 13th generation Dell PowerEdge Server. The PowerEdge R730xd supports the Intel Xeon E5-2600 v3 (Haswell) family of processors and provides NVMe support for up to four 2.5"form factor NVMe drives. The server also provides options for external SAS and SATA drive bays in 1.8", 2.5" and 3.5" form factors for maximum flexibility of on-host storage. With up to 768 GB of fast DDR4 RAM and 10 Gb Ethernet available on the motherboard, this server delivers a lot of processing power and I/O potential in two rack units.

### Samsung SM1715 NVMe Drive


The Samsung SM1715 is an NVMe SSD, available in PCIe card and small form factor (2.5") drive formats. This performance analysis made use of the latter. The SM1715 is rated with the following performance specifications:

- High density 3D V-NAND Flash
- Data Transfer Rates of large I/Os (128KB)
  - 3,000 MB/s sequential read
  - 2,200 MB/s sequential write
- Data I/O Speed of small I/Os (4KB)
  - 750,000 IOPS random read
  - 180,000 IOPS random write
- Latency of sustained random I/O
  - 85 us (microseconds) read I/O
  - 20 us write I/O

The SM1715 drive is available in 800 GB, 1.6T B, and 3.2 TB capacities, all of which are certified by Dell as compatible with current generation PowerEdge Servers.

### Aerospike NoSQL Database


Aerospike is an open source distributed NoSQL database optimized for flash storage to deliver speed and scalability to database applications. The database is architected to provide very low latency on read requests while under heavy load, recognizing modern business requirements for real-time data during times of intense I/O and processing activity. Aerospike directly manages local storage on its cluster nodes by bypassing filesystems to store data on the raw media, be that DRAM, flash, or traditional hard disk drives. This allows critical data to be stored on the fastest media, while the distributed nature of the Aerospike database provides a high degree of reliability in the event of cluster node failures.

## Why NVMe?

The strongest arguments for deploying NVMe in the enterprise come down to three simple statements:

- Low latency
- Power savings
- Bandwidth

The amount of data stored globally is increasing exponentially. This massive growth strains traditional HDD storage's ability to deliver information to applications fast enough to satisfy modern processing requirements. Flash helps—a lot—but flash performance is capped by the SAS and SATA bus speeds. PCIe flash drives are faster still, but proprietary drivers and the lack of a standard interface limits overall performance. NVMe (Non-Volatile Memory Express) eliminates these performance gating factors, providing a standard interface for non-volatile storage on the PCIe bus. This simplifies the deployment of high-end PCIe drives while improving performance and response times by up to six times that of SATA and SAS. The NVMe interface is hardware agnostic, intended to support any non-volatile storage medium on the PCIe bus, future-proofing it for the successor to today's flash storage.

By using the existing PCIe bus, NVMe technology is deployable to servers already installed in most datacenters. The PowerEdge R730xd server extends the PCIe bus to the chassis drive bays providing support for hot pluggable, small form factor drives, allowing use of NVMe technology without having to power down and open the server chassis.

Samsung shipped the first NVMe drive in 2014, bringing individual drive speeds of 3000 MB/s and response times as low as 25 us to the datacenter for the first time. The SM1715 drive specifications show even better response times and deliver NVMe performance in an easy-to-use pluggable SSD drive format as well as a PCIe card design. To top it off, the SM1715 draws a mere 25 Watts or less of power, a tiny fraction compared to legacy storage media.

The Aerospike NoSQL database is designed specifically for low latency reads during heavy write workloads. The database is already optimized for the low latencies of flash and capitalizes on the extremely fast response times of NVMe SSDs to decrease transaction times even more while increasing transaction counts. This allows processes that once required multiple relational database instances, in a mixture of online and offline modes, to be combined in a single online database.

## The Challenge Faced by Fraud Prevention Workloads

Determining the risk of fraudulent activity in financial transactions such as bank transfers or credit card transactions involves complex analytics in near real-time. These analytics are naturally data driven, and there is a lot of data to be processed, to create scores that determine the likelihood a given transaction is fraudulent. Multiple database transactions may be needed to satisfy a single analysis. The faster data is retrieved from a backend database, the more time the analytic system's logic has to process that data. A challenge arises when a lot of unstructured data is involved. Relational databases cannot efficiently manage unstructured data, and if backed by slow hard disk drive arrays, performance problems are inevitable. This causes two big problems, poor application response times and data lag.

A common service level agreement (SLA) for the return of fraud prevention results is less than a second. This includes time to retrieve data, run fraud prevention analytics on that data, and return a score to the application. In terms of compute time, this is a long time, and yet traditional relational databases backed with legacy HDD storage are often not able to satisfy reads quickly enough to allow analysis programs time to complete processing within the SLA window. Consistently missing the SLA may demand that simpler, less accurate fraud prevention rules be implemented, reducing the quality of results and potentially missing fraudulent transactions, which must then be covered by the financial institution, vendor, or consumer—a very real cost to the parties involved.

Data lag is the time between writing data and when that data is available for use by the fraud prevention engine. A traditional way of managing large, complex database systems involves separating reads and writes by creating two or more databases. One database is used to read data for processing, while another handles write transactions through batch loads. From time to time, these databases are synced and/or their functions switched to make newly written data available for processing. This results in a data lag, possibly many hours long, where newer records are not available for processing.

Customers need a two-phased solution: first, a database solution designed to efficiently retrieve (i.e., read) records without impacting (or being impacted by) database writes and secondly, a hardware platform robust enough not to hold back the performance of that database.

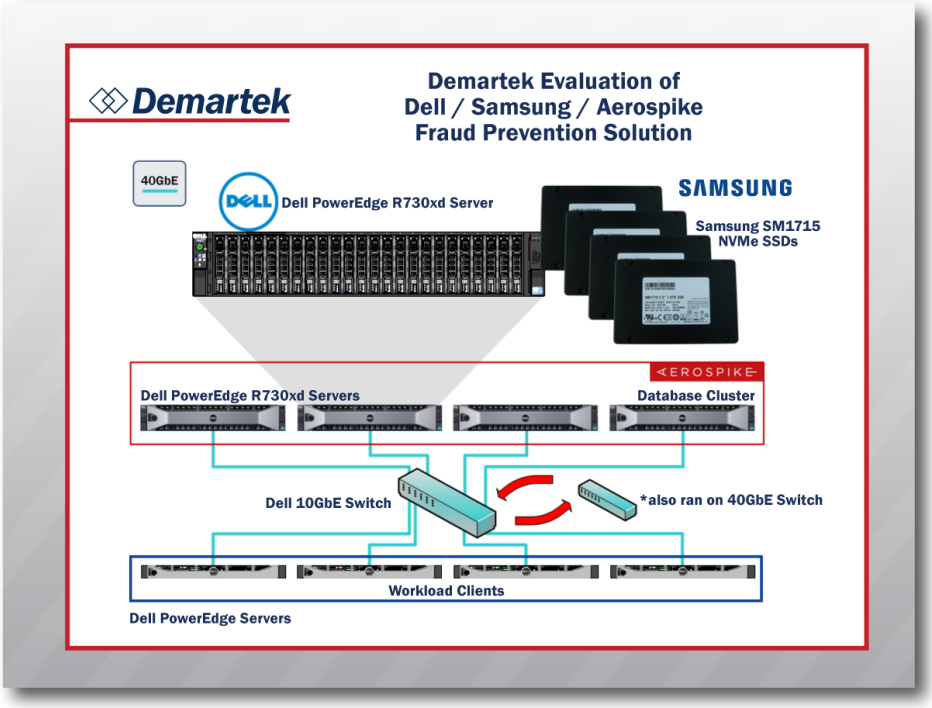
## Test Description and Environment

Aerospike developed an open source fraud prevention benchmark[2] to evaluate a compute environment's capacity to support a complex NoSQL database workload. This benchmark runs on a full installation of the Aerospike database and client software, simulating the data and operations required to evaluate the likelihood of individual financial transactions being fraudulent within a very large dataset. Up to 250 database transactions may be required for a single fraud prevention application transaction, making fast database response times a key component of this workload.

Dell PowerEdge R730xd servers were selected for the database cluster in this testing for their processing power, fast DDR4 SDRAM, and NVMe support in the drive bays. Low latency drives are essential to achieve the performance Aerospike can deliver for this type of workload. Therefore, Samsung SM1715 2.5inch NVMe drives were chosen as primary storage for the database specifically because of their extremely fast response times and ease of deployment in PowerEdge R730xd servers.

[2] Aerospike, Aerospike Java Benchmark, https://github.com/tfaulkes/aerospike-client-java

Four servers were configured as Aerospike database cluster nodes, with Ubuntu Linux 14.0.4. Benchmark testing validated two network speeds, 10 Gb Ethernet (10 GbE) and 40 Gb Ethernet (40 GbE). Intel X540 10 GbE adapters and Mellanox ConnectX-3 Pro 40 GbE adapters were installed on cluster nodes and client servers as interconnects and client-server communication. It was not expected that a 10G network would be a performance limiting factor, but with 40 GbE networks are expected to deliver lower latencies and higher throughput.




Four cluster servers were chosen for this testing strictly for ease of deployment. Aerospike easily scales to support larger datasets by simply adding additional cluster nodes. A reasonable comparison to larger environments can be made by multiplying the performance results to the scale factor of larger clusters.

Aerospike is a distributed database that is spread across the server nodes for redundancy, reliability, and scalability. Like massive-scale file and object storage frameworks, Aerospike creates multiple copies of data to protect against the loss or degradation of any node within the cluster, regardless of whether data is held in RAM or on non-volatile storage. Already optimized for flash storage, the expectation was the database would take advantage of the low latency of the Samsung SM1715 NVMe drives to deliver very low

transaction response times, even under heavy load. Bear in mind that fast database transaction responses are contingent on device latency being even lower.

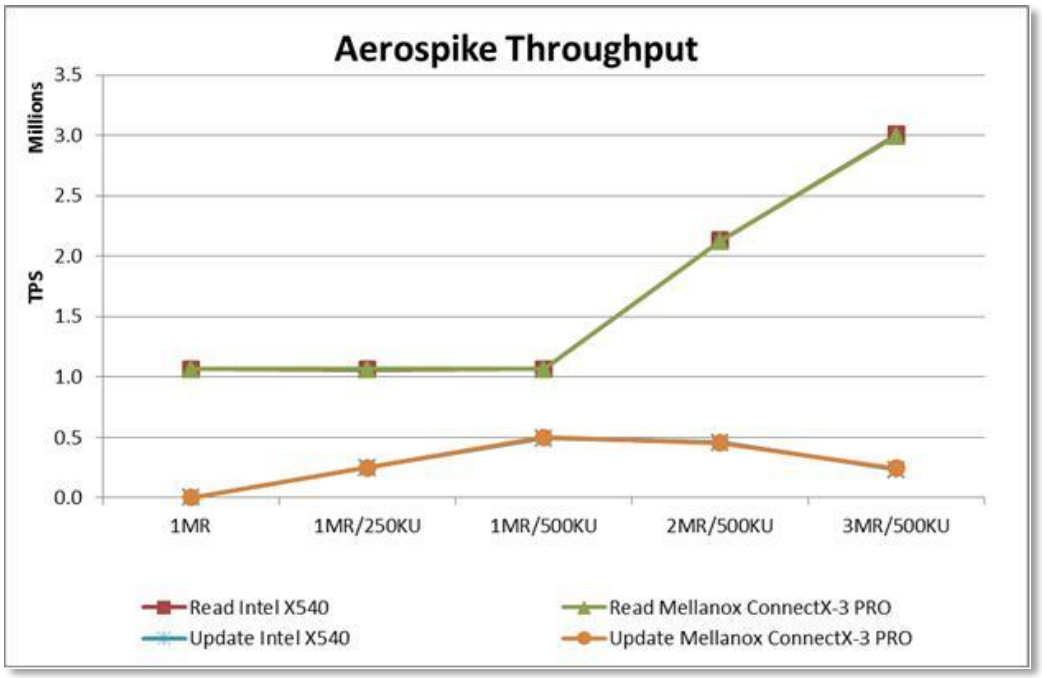
In a real-world use case, a transactional database application spends most of its time responding to read requests. To conduct fraud prevention calculations, fast reads are essential for delivering data to the analysis engine quickly enough to meet SLAs. However, during heavy write operations, traditional databases will see reads slowing down. Aerospike is designed to prevent this slowdown from affecting the application experience, or even from occurring at all.

The fraud prevention benchmark exercises Aerospike and the Dell/Samsung hardware platform in several phases. The first phase is a data load in which two billion objects are written to the database. The second phase is a read load of one million read operations per second. Subsequent phases add database write operations of 250,000 database updates per second and then 500,000 updates per second while continuing to perform the one million read operations. As if that wasn't challenging enough, the final two test points execute two million reads per second, and then three million reads, while performing 500,000 writes at the same time. The goal is to keep database transactional latency as low as possible, providing the fraud prevention engine the bulk of the SLA window to apply its rules and return a score.

Performance metrics were collected at each phase after a 10 minute ramp up to a steady-state condition. Metrics were collected from the workload client machines and represent round-trip I/O latency, from the client to the server and back again, recording the actual user experience, not just artificial best case server-only numbers.

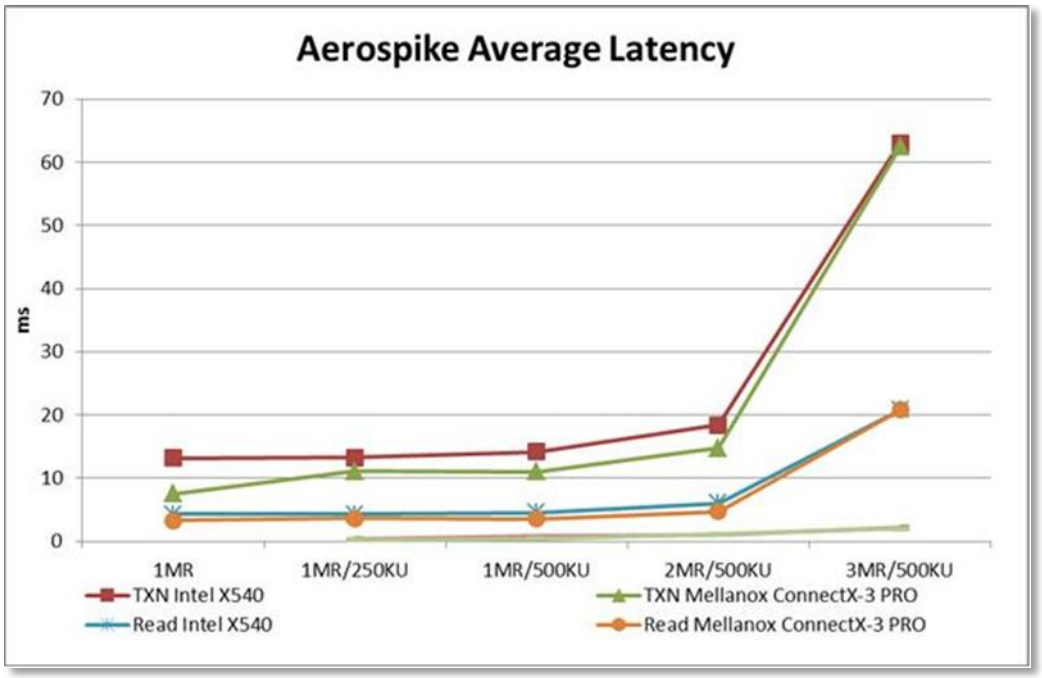
## Test Results and Analysis

The environment under test had no issue hitting the read operation rates defined in the test phases. However, at 2 million reads and above the system began to struggle to maintain 500,000 simultaneous updates. At the 3 million reads level roughly 250,000 write transactions were supportable. This is still an impressive number of transactions and demonstrates an operational fraud prevention workload running on a single Aerospike database, which is nearly impossible with legacy technology and relational databases.

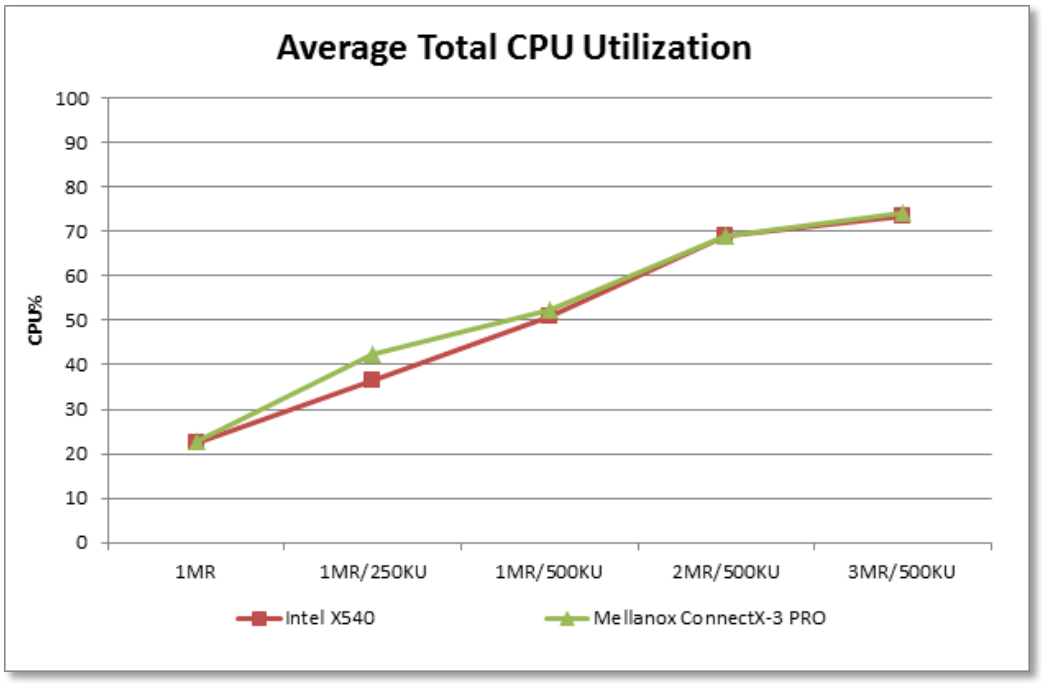



Throughput is interesting only insomuch as it demonstrates a transactional capacity. As we pointed out previously, transactional latency—specifically read transaction latency—is the critical metric for this benchmark. 95% or better of all database read operations, up to two million reads per second, while also performing 500,000 updates, were satisfied within five milliseconds. However, multiple reads are required to complete a single Aerospike database transaction. The total transactional latency was less than 20 milliseconds in four of the five scenarios. Only when trying to perform three million reads per seconds, alongside 500,000 writes, did latency begin to rise.

Even though we saw that maintaining 500,000 writes was not possible at the scale of the test system, read transactions still completed in fewer than 65 milliseconds. From the application's perspective, the database, the servers, and the drives were managing bi-directional I/O and returning results in near real-time. The application engine still had the majority of the SLA window in which to process data and compute the fraud prevention score, which was exactly what the benchmark was attempting to validate.




The sharp increase in latency at the 3 million reads level suggests that we are beginning to see the limit of the hardware's I/O capacity, but are still within acceptable levels of performance. The flattening of the server CPU usage curve at around 70% would also seem to support this hypothesis; the CPUs do not appear to be performing much more processing in spite of the heavier load, yet there is still capacity in reserve. Additional NVMe drives at this point may solve the problem, but it may also be prudent to consider 70% CPU usage a reasonable level with a reserve for periodic processing bursts and instead think about expanding the database cluster. Aerospike scales linearly, therefore it is our belief that the addition of Aerospike cluster nodes would result in latency dropping back down to the levels experienced by the prior benchmark phases.




Another takeaway from this data is that the choice of the network speed and network adapters, 10 Gb Ethernet with Intel x540 or 40 Gb Ethernet and Mellanox ConnectX-3 PRO, had no significant impact on the rest of the platform's ability to meet the intended number of transactions or maximum latency requirements in any test phase. Whichever modern network configuration this solution is deployed within, it will be adequate for the purposes of supporting this workload. 40 Gb Ethernet is a great future-proofing option at additional cost and it delivers latency benefits until the system becomes I/O limited.

## Summary and Conclusion

Regrettably, credit card fraud is a reality that must be dealt with. Businesses need to keep up with the continually evolving strategies employed by the bad guys. Without the proper tools, companies are always a step behind, and it costs them. Traditional relational databases are good at performing transactions and maintaining consistency on structured data, but they don't scale well to extremely large datasets (aka "Big Data") and struggle with unstructured data. NoSQL databases are designed to deal with these issues through improved scalability and better search logic for data that doesn't fit nicely into a relational table—just the type of date required by fraud protection engines.

13th Generation Dell servers, such as the PowerEdge R730xd, with the latest Intel processors and fast DDR4 SDRAM, provide raw compute power Aerospike can harness to manage very large and complex datasets. With the added feature of NVMe support to external drive bays, small form factor NVMe SSDs like the Samsung SM1715 can be easily deployed in individual server nodes for extremely high bandwidth and throughput with very low latency. We saw Dell, Samsung, and Aerospike sustain three million reads per second while supporting a heavy write load with database response times low enough to allow a near eternity of compute time—the better part of a second—for complex analytics to process that data within the fraud prevention SLA.

This is the performance that is demanded by modern financial transactions. Consumers are no longer content to wait hours or days for credit card or payment service transfers to go through. Businesses realize legacy architecture involving cache-fronted relational databases may work well for static data, or small-scale database applications, but it cannot deliver real-time access to large, complex, unstructured datasets. A new model is required to keep up with the demands of Big Data analytics.

The Aerospike NoSQL database is purpose built to solve this problem. Combined with a hardware platform that can deliver extremely fast device response times, achievable by NVMe flash on modern servers, very low latency database transactions are possible on very large databases. This, in turn, gives complex applications like fraud prevention the time needed to execute and return quality results before the end user gives up waiting. The Dell PowerEdge R730xd server, supporting easy to deploy Samsung SM1715 NVMe drives, gives Aerospike the hardware advantage to make this happen.

## Appendix A – Detailed test environment and configuration

The testing environment deployed four Dell PowerEdge R730xd servers in the database cluster and 4 Dell PowerEdge R620 servers as Aerospike client servers. Server specifications were as follows:

Dell PowerEdge R730xd

- Dual Intel Xeon E5-2670 v3, 2.3 GHz, 24 total cores, 48 total threads
- 256 GB DDR4 SDRAM
- Intel X540 10 Gb NIC
- Mellanox MT27520 (ConnectX-3Pro) 40Gb NIC
- 1 TB 7.2RPM 6 Gb/s SAS 2.5inch HDD Boot Drive
- 4 1.6 TB Samsung SM1715 2.5inch NVMe SSDs
- Ubuntu 14.04 (3.19)

Dell PowerEdge R620

- Dual Intel Xeon E5-2680 0, 2.7 GHz, 16 total cores, 32 total threads
- 64 GB DDR3 SDRAM
- Intel X540 10 Gb NIC
- Mellanox MT27520 (ConnectX-3Pro) 40Gb NIC
- 1 TB 7.2RPM 6 Gb/s SAS 2.5inch HDD Boot Drive
- Ubuntu 12.04.5 LTS (3.18.2-031802-generic)

Client and database server were connected by a Mellanox SX1036 40 Gb network switch.

The database test environment included the following parameters:

| | |
|---|---|
| Number of clusters | 1 |
| Number of nodes per cluster | 4 |
| Number of objects in the database | 2 billion |
| Object size | 900 bytes |
| Raw data volume (per cluster, un-replicated) | 1.8 TB |
| SSD space (Per cluster, replicated) | 25.6 TB (3.6 TB used) |
| RAM allocated (Per cluster) | 1 TB (256 GB used) |
| Write speed | 250,000 per second<br>500,000 per second |
| Tested application transactional rates (at 250 reads per financial transaction) | 4,000 per second |
| Database read rate | 1,000,000 per second<br>2,000,000 per second<br>3,000,000 per second |

Aerospike NoSQL database configuration file (similar on each host)

```
# Aerospike database configuration file for deployments using raw storage.

service {
        user root
        group root
        paxos-single-replica-limit 1 # Number of nodes where the replica count is
automatically reduced to 1.
        pidfile /var/run/aerospike/asd.pid
        service-threads 48   # This should match the number of cores on the server
        transaction-queues 48   # This should match the number of cores on the server
        transaction-threads-per-queue 32
        proto-fd-max 15000
}

logging {
        # Log file must be an absolute path.
        file /var/log/aerospike/aerospike.log {
                context any info
        }
}

network {
        service {
                address any
           access-address 10.10.10.1    # Local ip address to use for queries
                port 3000
        }

        heartbeat {
                mode mesh
                mesh-seed-address-port 10.10.10.1 3002
                mesh-seed-address-port 10.10.10.2 3002
                mesh-seed-address-port 10.10.10.3 3002
                mesh-seed-address-port 10.10.10.4 3002
                address 10.10.10.1     # Local ip to use to listen for heartbeats
                port 3002

         interface-address 10.10.10.1    # Local address to use for to transmit heartbeats
                # To use unicast-mesh heartbeats, remove the 3 lines above, and see
                # aerospike_mesh.conf for alternative.
                interval 150
                timeout 10
        }

        fabric {
                port 3001
        }

        info {
                port 3003
        }
}

namespace test {
        replication-factor 2
        memory-size 200G   # Total amount of RAM to use
        default-ttl 30d # 30 days, use 0 to never expire/evict.
        single-bin true

        # Warning - legacy data in defined raw partition devices will be erased.
        # These partitions must not be mounted by the file system.
        storage-engine device {
                # Use one or more lines like those below with actual device paths.
                device /dev/nvme0n1
                device /dev/nvme1n1
                device /dev/nvme2n1
                device /dev/nvme3n1
```

```
                # The 2 lines below optimize for SSD.
                scheduler-mode noop
                write-block-size 128K

                # Use the line below to store data in memory in addition to devices.
#               data-in-memory true
                post-write-queue 0  # Set to 0 to turn off caching. Use during benchmarking
to show worst case scenario
        }
  }
```

The most current version of this report is available at http://www.demartek.com/Demartek_Dell_Samsung_Aerospike_Fraud_Prevention_Evaluation_2015-12.html on the Demartek website.