# Cloudmark Case Study

## Cloudmark Slays Spam with Fusion-io

*Security company improves performance in several areas by an order of magnitude and cuts its server footprint more than half.*

## The Challenge

Cloudmark, a messaging security provider, is a veteran in the arms race against spam, phishing, and viruses. Spammers are constantly looking for ways to game or beat the system, and Cloudmark's reaction time to new threats is of paramount importance to fighting new attacks. It invested in a high-performance disk solution, but data growth ate up every performance upgrade within a few months.

Always on the lookout for innovative ways to improve service at a lower cost, Ryan White, director of operations, began looking for a cost-effective and scalable performance solution that did not require continual outlays on disks.

Cloudmark had a couple of systems that required perpetual upgrading. The first system hosted its messaging security application and required database replication. As its data load grew, so did its performance requirements and disk investment. In order to get the top-notch performance Cloudmark's customers required, continuous investments in capital were needed. The second system was an analysis server for which they wanted a more scalable solution to analyzing ever-growing amounts of data.

After speaking with Fusion-io at LinuxWorld 2007, Ryan decided to give some ioDrives a try.

CLOUDMARK®

**SUMMARY OF BENEFITS**

- *5x improvement to database replication performance*
- *5x improvement to data-intensive queries*
- *10x improvement to analysis server performance that enabled 5x improvement to analysis routines*
- *Simplified maintenance by eliminating at least 210 disk node failure points from its system*
- *Improved reliability and availability with full system redundancy*
- *Recovered 27U of rack space*
- *Greatly lowered power and cooling expenses*

---

*"After about a year and a half of using Fusion-io, when anyone asks me for a disk array, my first question is 'Why?' There is no reason for a disk array here. It's just a box that I have to support with 25 individual failure points."*

*Ryan White,*
*Director of Operations*

FUSION-IO®

# The Solution

## BENCHMARK TESTING

Before implementing Fusion-io, Ryan ran benchmarks for various load patterns. Following are the results of a benchmark with 8KB records[1] and a MySQL simulation.[2]

### 8KB Record Benchmark

| | 2x 160GB ioDrives | 4x10k SAS (512MB write cache) | 24x15k SAS (512MB write cache) | Improvement over 10k SAS | Improvement over 15k SAS |
|---|---|---|---|---|---|
| **MBytes/sec** | | | | | |
| Read | 823.70 | 68.00 | 151.00 | 1,211.32% | 545.50% |
| Random Read | 811.80 | 10.00 | 57.00 | 8,118.00% | 1,424.21% |
| Random Write | 523.50 | 7.00 | 34.00 | 7,478.57% | 1,539.71% |
| **IOPS** | | | | | |
| Read | 100,549.32 | 8,300.78 | 18,432.62 | 1,211.32% | 545.50% |
| Random Read | 99,096.68 | 1,220.70 | 6,958.01 | 8,118.00% | 1,424.21% |
| Random Write | 63,903.81 | 854.49 | 4,150.39 | 7,478.57% | 1,539.71% |

### MySQL Simulation Benchmark

| | 2x 160GB ioDrives | 4x10k SAS (512MB write cache) | Improvement over 10k SAS |
|---|---|---|---|
| **MBytes/sec** | | | |
| Read | 308.5 | 121.7 | 253% |
| Random Read | 218.8 | 2.6 | 8,415% |
| Random Write | 211.4 | 14.7 | 1,438% |
| **IOPS** | | | |
| Read | 18,829.3 | 7,428.0 | 253% |
| Random Read | 13,354.5 | 158.7 | 8,415% |
| Random Write | 12,902.8 | 897.2 | 1,438% |

*1. The test compared two 160GB ioDrives against a 4-disk 10,000 RPM SAS array and a 24-disk 15,000 RPM SAS array. It is worth noting that bandwidth improved at exactly the same percentage as IOPS. This is expected in a simulation, where the threads and records are of equal size.*

*Also noteworthy: Ryan began a benchmark with 16KB records, but the test was taking so long on the disk array systems that he gave up.*

*2. This run consisted of 1 thread at 150GB per thread and 16KB records. The benchmark compared two 160GB write-optimized ioDrives to a 4-disk 10,000 RPM SAS array.*

Ryan also ran some benchmark tests against Cloudmark's analysis server.[3]

**Analysis System Benchmark**

| | 1x 320GB MLC ioDrive | 4x10k SAS (512MB write cache) | 24x15k SAS (512MB write cache) | Improvement to 10k SAS | Improvement to 15k SAS |
|---|---|---|---|---|---|
| **MBytes/sec** | | | | | |
| Read | 236.7 | 44.00 | 87.00 | 538% | 272% |
| Random Read | 211.2 | 6.00 | 41.70 | 3,520% | 506% |
| Random Write | 259.2 | 4.00 | 29.00 | 6,480% | 894% |
| **IOPS** | | | | | |
| Read | 57,788.10 | 10,742.19 | 21,240.23 | 538% | 272% |
| Random Read | 51,562.50 | 1,464.84 | 10,180.66 | 3,520% | 506% |
| Random Write | 63,281.30 | 976.56 | 7,080.08 | 6,480% | 894% |

## SUPERCHARGING DATABASE REPLICATION

The system supporting Cloudmark's back-end consisted of a MySQL master server that replicated to slave (or mirrored) databases at three sites. The slaves at these sites were supported by massive disk arrays taking up a whopping 49U of rack space (see page 6 for detailed system information).

Just two ioDrives in each slave server completely eliminated the replication I/O bottleneck, improving Cloudmark's average system reads and writes *over five times*. It also eliminated 210 spinning disks supporting the slave servers and recovered 27U of rack space.

Needless to say, Ryan was thrilled."Fusion-io has given us the edge in the security/spammer arms race," he said. "The ioDrive performance improvements helped our replication times immeasurably, effectively shortening the window spammers have to work in before we block them."

*3. The test compares one 320GB ioDrive to a 4-disk 10,000 RPM SAS array and a 24-disk 15,000 RPM SAS array. This run consisted of 32 threads at 256MB per thread and 8KB records.*

## CREATING OPPORTUNITIES TO IMPROVE APPLICATION PERFORMANCE

The ioDrives performance improvements had a side-effect—it allowed Cloudmark's messaging security application to fully utilize the CPUs. After adding the ioDrives, the application maxed the system's processors. "Once we installed Fusion-io, we found that we ran into CPU walls before we ran into I/O walls," Ryan said. "I've never had a database system that hit the CPU wall."

The Cloudmark team began digging into ways to more efficiently use its system's processors and found the application was making some inefficient queries whose impact wasn't felt when disk I/O was the bottleneck. For example, one of its main applications had been issuing inserts and updates one query at a time. With a small amount of effort, fixing this problem produced amazing results. Ryan told us, "Without the I/O bottleneck, we could now update 1,000 or more records per query. This work improved the performance of our data-intensive queries by about five times. We went from our disk-based slaves having constant trouble keeping up, with constant 90-100% disk I/O saturation on the P800 controller with MSA70's, to the Fusion-based slaves always keeping up at 5% or less I/O utilization."

## IMPROVING ANALYST PRODUCTIVITY AND PROBLEM TURNAROUND TIME

In addition to improving its data replication performance, Cloudmark was also interested in finding a more scalable solution for problem data analysis. This analysis consisted of investigating spam messages that should have been caught as well as authentic messages getting blocked. The investigation often involved searching massive amounts of data, including many large files (500MB-1GB in size), and could take hours.

The ideal solution would improve analyst productivity by reducing the time the system took to read in the data from its sources, analyze it, manipulate it, and update the results in the master database to be propagated to customers with the next update.

Fusion-io answered the call. The new system consisted of a server with two RAID0-striped 320GB ioDrives that provided sufficient capacity to hold the last three to four days worth of data, which covered well over 90% of the problem data cases. Cloudmark relegated its legacy 4TB system to searches of older data.

"People love it because they can do whatever they want, just bash away and it never slows down," Ryan told us. "Our analysts are much more productive and we can get customer fixes for problem data much more quickly. This freed our analysts to improve analysis routines to leverage ioDrives to crunch through data nearly five times faster."

## ENHANCING THE DATACENTER

In addition to achieving Cloudmark's primary goal of improving data replication and analysis performance while curbing future outlay on disks, it also improved the reliability and availability of the system, reduced its server footprint, and lowered its operational costs.

### *Improving Availability and Reliability*

For less money than a disk-based solution, Cloudmark implemented full system redundancy, improving both data reliability and availability.

"We RAID0 striped the ioDrives to get the capacity we needed, which meant that if we lost the data on one card, we would lose it on both." Ryan told us, "However, we saved so much money on disks that we were able to purchase additional Fusion-io servers to give us system-level clustering and redundancy."

Ryan also found Fusion-io to be more reliable than other solid-state solutions. A couple of companies he talked to thought it would be cheaper to RAID four Intel SSDs with an off-the-shelf hardware RAID controller. "But when they tried, on multiple occasions one of the drives would randomly drop out when under load and they would lose the whole array," he said. "Also, you don't get the ability to optimize these drives for write performance. They have since learned their lesson that you get what you pay for." Ryan ruled out disk-form SSDs for Cloudmark because it is a Hewlett Packard shop and he didn't want to void his support warranty by replacing hard drives with unsupported and untested configurations—not a problem for Fusion-io. Other solid-state form factors he investigated, including Violin and BitMicro, were far too expensive.

### *Lowering Operational Costs*

As if these benefits weren't enough, Cloudmark lowered future operating costs on several fronts

- First, it reclaimed valuable real estate. Even with the addition of full system redundancy, Cloudmark reduced the size of its system by over half, from 49U to 22U.
- Second, Cloudmark implemented a green solution that cut its power consumption by up to 40%[4] and over double that saving in cooling costs.[5]
- Finally, by replacing the spinning disks with ioDrives, Cloudmark eliminated 210 mechanical failure points in its system.

*4. According to the Information Business Technology Network, at least 40% of the power for a server with a performance disk array is consumed by the disks.*

*5. Cooling costs for servers double a server's power consumption. See Estimating Total Power Consumption by Servers in the U.S. and the World.*

FUSION-IO®

## APPLICATION AND DATABASE SYSTEM BEFORE

### *Master at Headquarters (9U)*

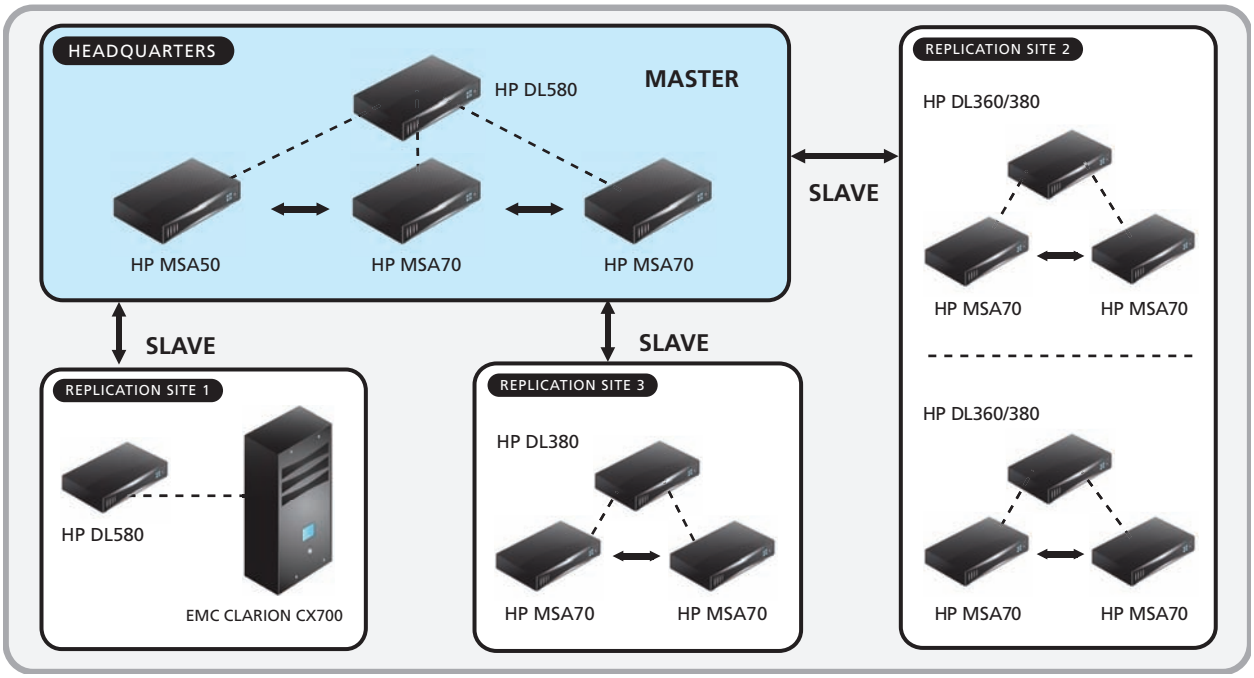
HP DL580 G5, 4x Quad Core Xeon, 128GB RAM
– OS: RHEL5 x64
– Database Software: MySQL 5.0.x
– Disks:
- 8x 73GB 10K RPM SFF internal drives
- 10x 146GB 10K RPM SFF drives, inside and HP MSA 50, connected via P600 controller
- 50x 73GB 15K RPM SFF drives, inside 2x HP MSA 70s connected via P800 controller

### *1 Slave at Replication Site 1 (22U)*

HP DL580 G5, 4x Quad Core Xeon, 128GB RAM
– OS: RHEL5 x64
– Database Software: MySQL 5.0.x
– Disks:
- 8x 73GB, 10K RPM SFF internal drives
- EMC Clariion CX700, connected via QLogic FC HBAs with 60x 36/73GB, 15K RPM FC drives

### *2 Slaves at Replication Site 2 & 1 at Site 3 (18U total)*

HP DL360/380 G5, 2x Quad Core Xeon, 32GB RAM
– OS: RHEL5 x64
– Database Software: MySQL 5.0.x
– Disks:
- 6-8x 73GB 15K RPM SFF internal drives
- Two MSA70s connected via P800 controller with 50x 73GB 15k RPM SFF drives




FUSION-IO®

## APPLICATION AND DATABASE SYSTEM AFTER

### *Master at Headquarters (9U)*

HP DL580 G5, 4x Quad Core Xeon, 128GB RAM
– OS: RHEL5 x64
– Database Software: MySQL 5.0.x
– Disks:
- 8x 73GB, 10K RPM SFF internal drives
- 10x 146GB, 10K RPM SFF drives, inside an HP MSA50 connected via P600 controller
- 50x 73GB, 15k RPM SFF drives, inside 2x HP MSA70s, connected via P800 controller

### *1 Slave at Replication Site 1 (4U)*

HP DL580 G5, 4x Quad Core Xeon, 128GB RAM
– OS: RHEL5 x64
– Database Software: MySQL 5.0.x
– Disks:
- 8x 73GB, 10K RPM SFF internal drives
- 2x 160GB ioDrive cards in RAID0

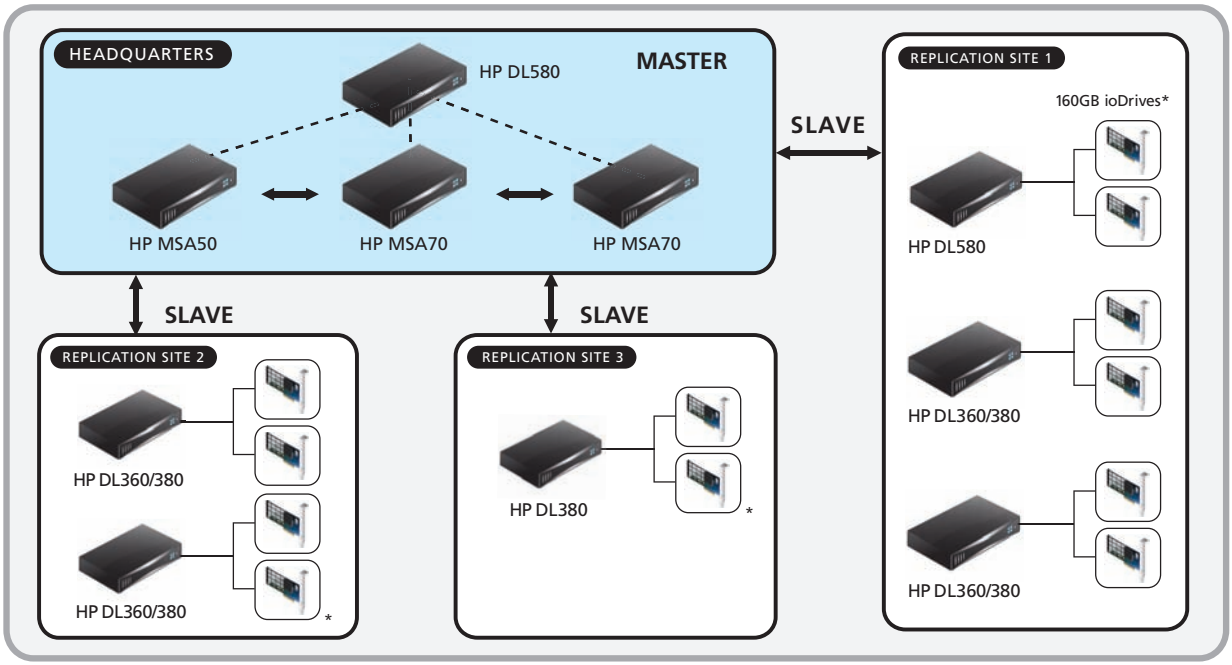
### *2 Slaves each at Replication Sites 1 & 2 (1–2U ea.)*

HP DL360/380 G5, 2x Quad Core Xeon, 32GB RAM
– OS: RHEL5 x64
– Database Software: MySQL 5.0.x
– Disks:
- 6-8x 73GB 15K, RPM SFF internal drives
- 2x 160GB ioDrives in RAID0

### *1 Slave at Replication Site 3[6] (2U)*

HP DL380 G5, 2x Quad Core Xeon, 32GB RAM
– OS: RHEL5 x64
– Database Software: MySQL 5.0.x
– Disks:
- 6-8x 73GB 15k RPM SFF internal drives
- 2x 160GB ioDrives in RAID0

*6. A second, redundant slave at Replication Site 3 was being implemented at the time this study was created.*




FUSION-IO®

# Cloudmark Case Study

## ANALYSIS SYSTEM BEFORE

Supermicro Whitebox, 1x Quad Core Xeon E5345, 4GB RAM (3U), 3Ware 9650SE RAID Controller with battery-backed write cache

– Disks:
- 16x 750GB, 7200K RPM SATA drives

## ANALYSIS SYSTEM AFTER

***Primary Search System***

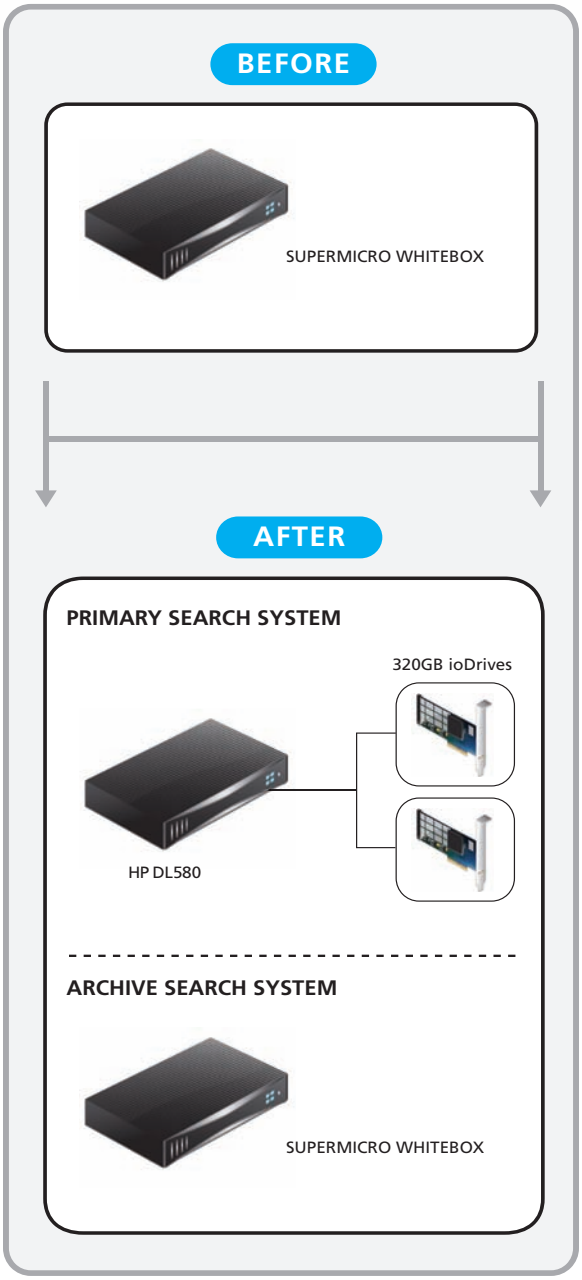
HP DL380 G5, 1x Quad Core Xeon E5430, 14GB RAM (2U)

– Disks:
- 8x 146GB, 10K RPM SFF internal drives
- 2x 320GB MLC ioDrives in RAID0

***Archive Search System***

Supermicro Whitebox, 1x Quad Core Xeon E5345, 4GB RAM, using a 3Ware 9650SE RAID Controller with battery-backed write cache

– Disks:
- 16x 750GB, 7200K RPM SATA drives




FUSION-IO®

## Summary

Fusion-io helped Cloudmark stop the continual outlay on disks previously necessary to provide its customers with industry leading performance. This investment improved its data center as follows:

- 5x improvement to database replication performance
- 5x improvement to data intensive queries
- 10x improvement to analysis server performance
- Enabled 5x improvement to analysis routines
- Simplified maintenance by eliminating at least 210 disk node failure points from its system
- Improved reliability and availability by implementing full system redundancy
- Greatly lowered power and cooling expenses
- Recovered 27U of rack space

Ryan has the following to say about Fusion-io: "After about a year and a half of using Fusion-io, when anyone asks me for a disk array, my first question is 'Why?' There is no reason for a disk array here. It's just a box that I have to support with 25 individual failure points."

While the previous statement is a testimony to Fusion-io's superiority to disk, the following is the most ringing endorsement: "We have to spend a lot less time now worrying about disk I/O and our databases keeping up, and we can focus on other things in the business, because that part of this business is no longer a constant focus for us."

## About the Customer

Cloudmark provides comprehensive messaging security solutions that protect subscribers and messaging infrastructure from spam, phishing, viruses and other harmful content. Founded by pioneers in messaging anti-abuse, Cloudmark's solutions were architected from the ground up to stop current and emerging threat vectors with high efficiency and scalability.

As a result of its unrivaled accuracy and performance, Cloudmark has emerged as the most widely deployed messaging security solution in the world today. Cloudmark's customers include over 100 of the world's largest service providers, including the majority of tier-1 operators in North America, Japan and Europe.