



Hadoop on Paladian ADS1000

Performance, Scalability and Simplification
Accelerating Real-Time Big Data

Wednesday, March 25, 2020

Overview



- ▶ Problem Statement
 - ▶ Growing clusters cause datacenter sprawl
 - ▶ Most clusters stop growing the performance because of size
- ▶ Cloudera Benchmark with Paladian ADS1000
 - ▶ 60% – 90% Cluster Size Reduction
- ▶ Hortonworks Benchmark with Paladian ADS1000
 - ▶ 50% – 80% Cluster Size Reduction

Problem Statement

- ▶ Growing Hadoop clusters causes datacenter sprawl
- ▶ Reasons why Hadoop clusters grow:
 - ▶ Needing more storage capacity
 - ▶ Needing more throughput, which can be throttled by
 - ▶ IO bandwidth limitations
 - ▶ CPU limitations
 - ▶ Memory limitations (Spark will increase system utilization)
- ▶ Users have to buy servers to increase storage

Most Hadoop clusters do not grow beyond 50-75 servers due to management and TCO concerns



Paladian Data

Cloudera HDFS Paladian ADS1000 Benchmarks

The following slides are a snapshot of how ADS1000 compared when using Cloudera CDH 5.10.0

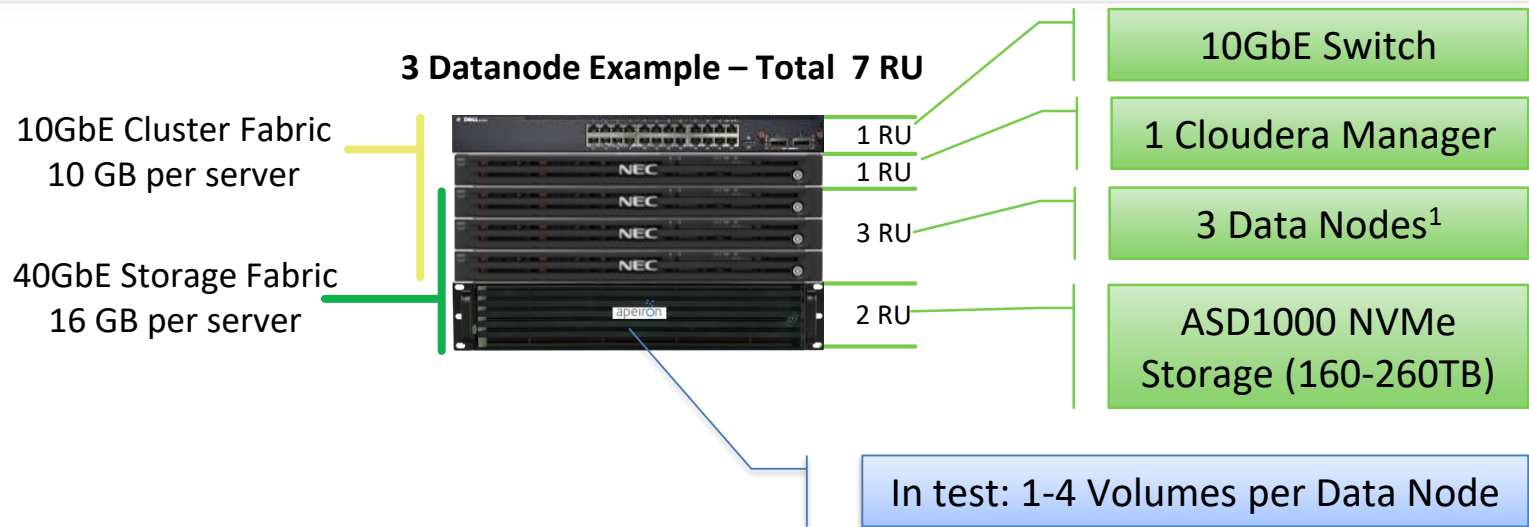
Paladian Cloudera Hadoop Appliance Test Configuration (SPC TSC)



Hosts	Count	Existing Roles
d30.hq.apeirondata.com	1	B, NN, SNN, G, HMS, HS2, HS, AP, ES, HM, RM, SM, OS, JHS, RM, S
d[31-33].hq.apeirondata.com	3	DN, G, NM

1 Cloudera Manager. CDH 5.10.0,
Java version: 1.7.0_67

3 Data Nodes



¹ Up to 24 Data Nodes can be supported for each ADS1000 Storage Array (depending on storage & performance requirements)

Paladian TestDFSIO 1TB I/O Performance

- ▶ For all the comparisons in these slides we have used “the best HDFS benchmark on the web”.
- ▶ The Paladian ADS1000 shows a performance which is up to an order of magnitude better.

	Unit	HDD	SATA SSD ¹	ADS1000 NAND SSD	Comments
Servers		7	7	4	
Datanodes		6	6	3	Half number of datanodes
Disks		12	12	12	
Runtime Write	s	5131	1319	686	2x faster
Runtime Read	s	1756	485	75.3	6x faster
Throughput Write	MB/s	218	982	1,718	1.75x higher TP
Throughput Read	MB/s	722	4,087	39,320	9.6x higher TP

¹ “CloudSpeed™ SATA SSDs Support Faster Hadoop Performance and TCO Savings”

The Paladian ADS1000 shows a 2x - 10x better performance in a cluster with half the datanodes



Cluster Size Consolidation

- ▶ The Paladian Cloudera Hadoop Appliance can deliver a 60% cluster size reduction based on write performance requirements
- ▶ **>80% cluster consolidation can be reached for read intense clusters**

	Unit	SATA SSD	ADS1000 NAND SSD	Comments
Total Size	RU/Racks	82/2	31/1	>60% size reduction
Switch	RU	2	1	
Datanodes	RU	78	21	75% fewer nodes
Shared Storage	RU	-	8	
Throughput ¹ Write	GB/s	10.0	10.4	Same Write I/O Perf.
Throughput ¹ Read	GB/s	27.4	95.2	Read >3x faster

The Paladian ADS1000 will allow you to reduce the cluster size:

60% in Write heavy cases

Up to 80% or more in Read heavy cases

¹ In this case the Throughput numbers are based on TestDFSIO runtimes (which is more conservative than the reported TP numbers).

The Paladian ADS1000 Performance

- ▶ NVMe Storage is so fast that “zlib.compress.level” can be set to no_compression rather than default_compression without lowering the I/O performance.
- ▶ CPU cycles are saved by turning off the extra cpu-cycles for compression

Parameter	Unit	defaultCompress	noCompress	Notes
Write Throughput	MB/s	1,458	1,464	~0.4% better
Read Throughput	MB/s	18,160	17,970	~1% slower

The Paladian ADS1000 I/O performance competes with the default compression methodology for a Cloudera cluster.

Compression will save you disk space, but not on performance nor CPU resources

Paladian Cloudera Performance Summary

- ▶ The Paladian ADS1000 Hadoop Appliance for Cloudera shows 4x – 20x I/O Performance improvement when compared to earlier test platforms
- ▶ The performance can be used to reduce the size of a Hadoop HDFS cluster 60% – 90%
- ▶ The Paladian ADS1000 Hadoop Appliance delivers better performance than seen with captive storage, with the TCO benefits of shared storage
- ▶ The Paladian ADS1000 Hadoop Appliance opens new doors for future Real Time Applications using Hadoop/HDFS



Paladian Data

Hortonworks HDFS & Optane Paladian ADS1000 Benchmarks

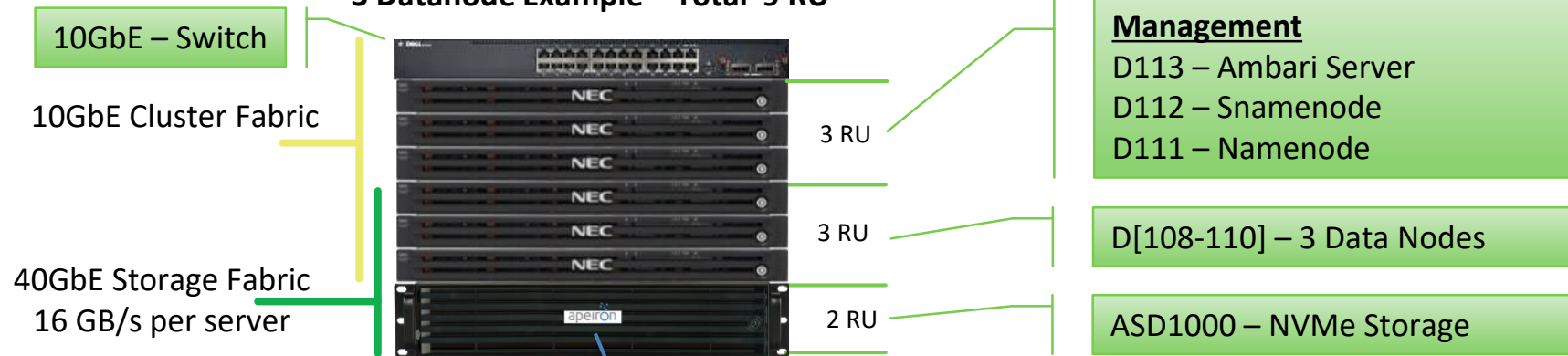
The following slides are a snapshot of how ADS1000 compared when using Hortonworks 2.6.2.0-205

Paladian Hortonworks Hadoop Test Appliance



Hortonworks rev: 2.6.2.0-205
Ambari rev: 0.1.0
Spark: 2.1

3 Datanode Example – Total 9 RU



Servers: R120f-1M Intel Haswell-EP
Dual Xeon 14 Core E5-2695 V3, 2.3 GHz
128GB (DDR4-2133Mhz)

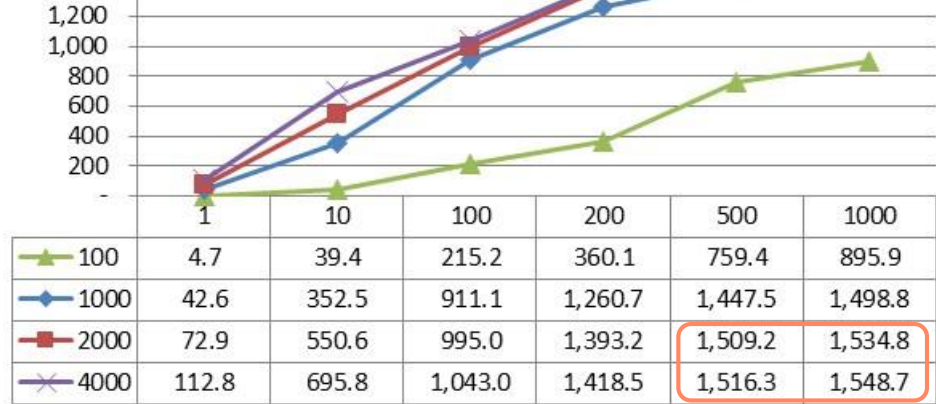
4 Volumes per datanode (2.5" NVMe SSDs)

Paladian Hortonworks TestDFSIO Results

3 Datanodes – 4 NVMe SSD drives per node

TestDFSIO Write [MB/s] Calc. from Exec Time

File Sizes: 100,1000,2000,4000 MB



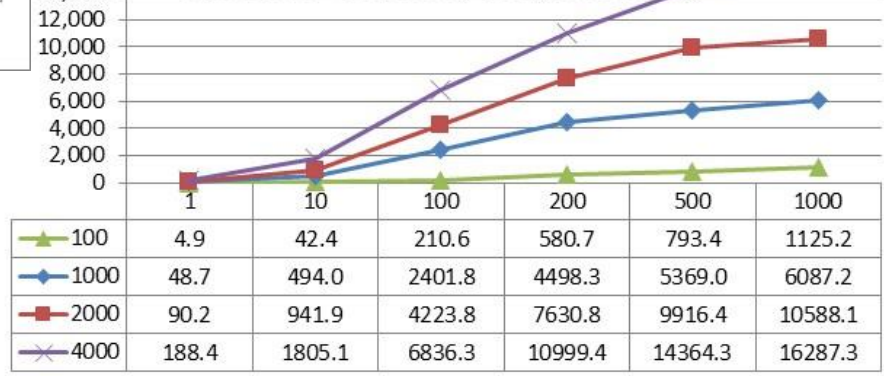
nrFiles

Read throughput reaches 16 GB/s (~1.3 PB/day)

Write throughput reaches 4.5 GB/s of total writes (replication 3)
This means ~130TB/day of real data

TestDFSIO Read [MB/s] Calc. from Exec Time

File Sizes: 100,1000,2000,4000 MB

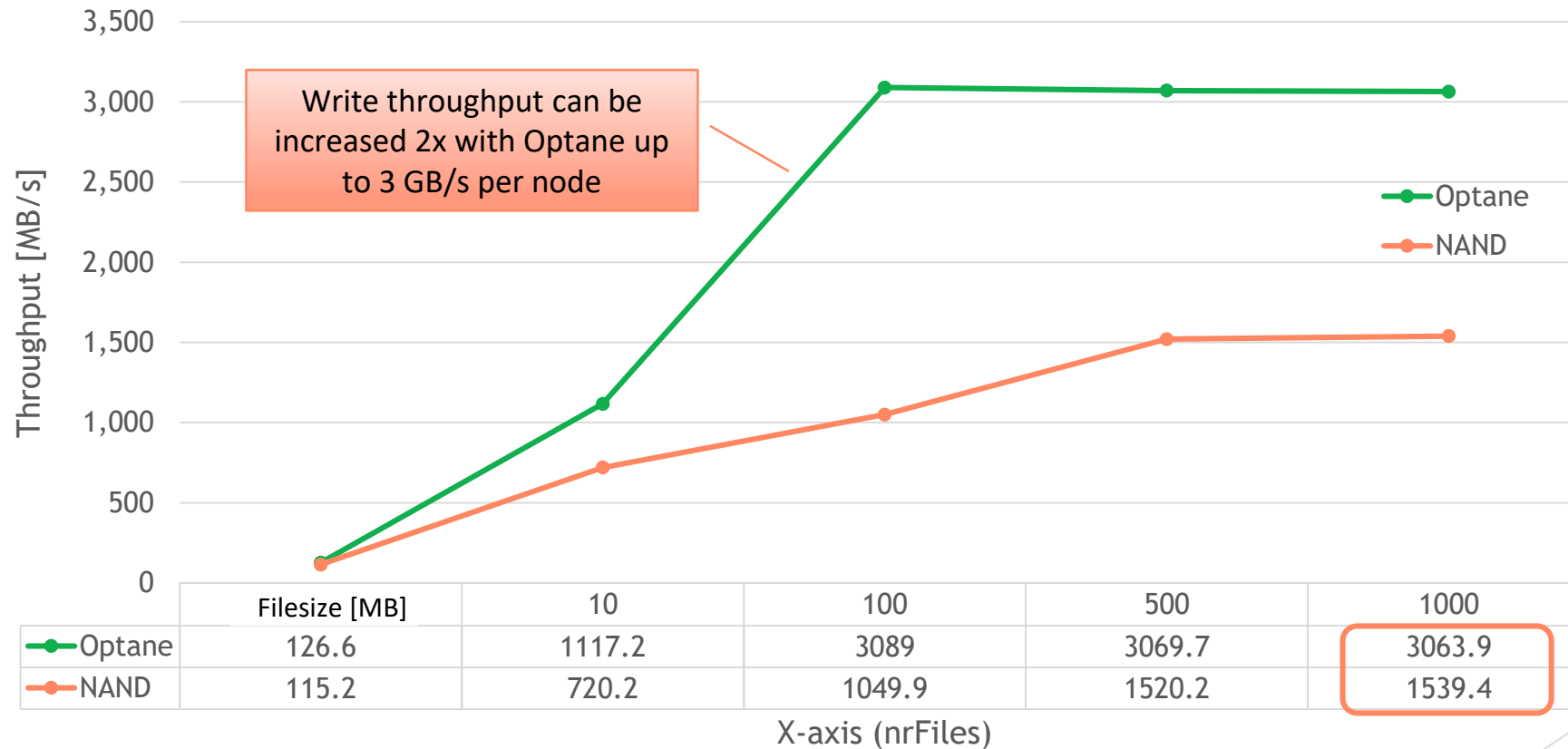


nrFiles

Paladian TestDFSIO NAND vs. Optane

Write Throughput, one datanode

Throughput [MB/s] vs. Number of Files



Paladian TestDFSIO Industry Comparison

- ▶ ADS1000 removes the datanode IO bottlenecks in a cluster
- ▶ More CPU cores can be added to each datanode for a better scale out performance factor

TestDFSIO 1TB per datanode	Sandisk HDD	Sandisk SATA SSD	Paladian w/ NAND	Paladian w/ Optane	Paladian Advantage
Read [MB/s]	722	4,087	21,647	16,863	30x better than HDD 5x better than captive SSD
Write ¹ [MB/s]	218	952	1,539	3,089	14x better than HDD 3.2x better than captive SSD

¹ Write performance is more affected by any selected compression format. Paladian NAND performance has been shown to make the compression irrelevant, and not needed, for the datanode IO performance.

Paladian HiBench Benchmark Summary

Benchmark	[MB]	Duration [s]		Throughput [MB/s]		Improved
Name	Datasize	Hadoop	Spark	Hadoop	Spark	In Spark
Terasort BD	600,000.00	8198.528	3848.645	73.18	155.90	2.1x
Wordcount BD	1,642,457	29395.69	702.247	55.87	2,338.86	41.8x
ML-Bayes G	3,763.18	4415.009	53.382	0.852	70.50	82.7x
ML-Kmeans H	20,081.83	944.707	80.162	21.26	250.52	11.8x
SQL-Aggregation H	372.37	97.802	40.807	3.81	9.13	2.4x
SQL-Join G	19,194.93	399.241	87.378	48.08	219.67	4.6x
SQL-Scan H	2,009.51	57.350	35.968	35.04	55.87	1.6x

- Spark shows 2x to 80x improvement over MapReduce
- Spark is RAM hungry
- Without enough RAM Spark will not finish

Paladian Hortonworks Cluster Size Consolidation

(Based on a Read/Write mix of 50%/50%)

- ▶ The table shows different cluster sizes based on resource requirements
- ▶ Paladian's I/O performance allows for massive consolidation
- ▶ Fewer servers enables more cost effective RAM increases for Spark
- ▶ Next step: Pooled Optane enables even larger Spark datasets

Cluster Requirements		Rack Units			Paladian Savings
Parameter	Requirement	Horton	Paladian w/ NAND	Paladian w/ Optane	
Physical cores [#]	1000	100	48	48	52%
Throughput for storage [GB/s]	360	100	41	25	59%-75%
Storage capacity [TB]	5000	105	49	-	53%
Total memory [TB]	25	102	67	67	34%



Note: Paladian solutions assume 28 core x86 nodes, 512G RAM, 15TB Toshiba SSDs (or Optane)

Paladian Hortonworks Summary

- ▶ Local Storage is preferred for scale out solutions
 - ▶ Paladian ADS1000 delivers
 - ▶ Better performance than local storage with ease of management
 - ▶ Unlimited capacity
- ▶ The ADS1000's industry leading performance enables a remarkable cluster size reduction
 - ▶ The Paladian ADS1000 IO performance allows for 50% - 80% server reduction
 - ▶ In a smaller cluster you can efficiently increase cores and RAM to achieve optimal resource utilization
- ▶ Spark is very fast and enables real-time performance in Hadoop
 - ▶ The ADS1000 enables cost effective RAM increases in fewer servers and paves the way for pooled RAM extension on Optane

Problem Statement Summary

Classic Solutions

1. Growing Hadoop clusters causes datacenter sprawl
2. Reasons why Hadoop clusters grow:
 1. Needing more storage capacity
 2. Needing more throughput, which can be throttled by
 1. IO bandwidth limitations
 2. CPU limitations
 3. Memory limitations (Spark will increase system utilization)
3. Users have to buy servers to increase storage

Paladian Solutions

1. Paladian will reduce the cluster size 50% – 80%
2. Paladian ADS1000 changes the paradigm
 1. Up to PB on a single server
 2. Better throughput than any captive storage
 1. No shared storage bottlenecks
 2. More efficient use of the CPU resources
 3. More efficient use of memory
3. “Unlimited” storage on each server

Paladian scales linearly to PBs with the best performance per datanode