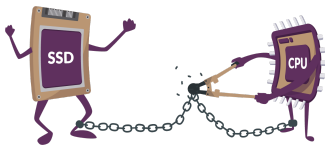
Disaggregation of Cassandra nodes with No Drama Lightbits™ LightOS® SDS



Modern cloud-scale architectures running data-hungry applications have resorted to installing high-performance NVMe Flash storage directly into their Cassandra compute nodes. This DAS model introduces expensive trade-offs: Performance is increased but compute and storage are locked together, data services are limited, and the storage capacity is underutilized.

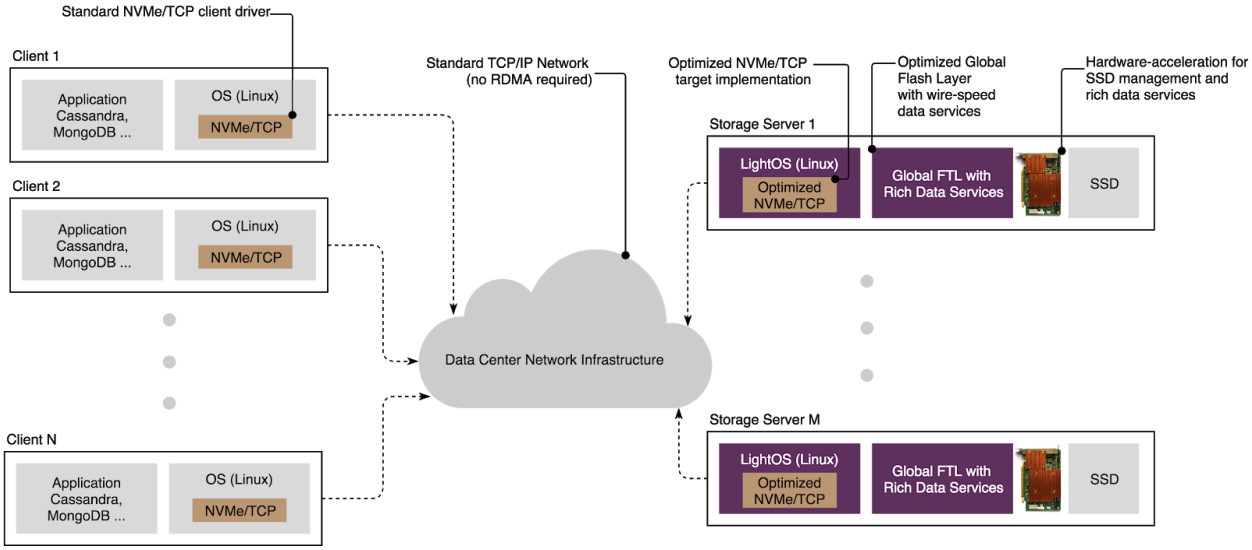
Disaggregation of compute and storage using the Lightbits™ LightOS® Software Defined Storage (SDS) solution enables Cassandra compute nodes to achieve high performance without stranding and wasting storage within each compute node.

Lightbits Cassandra benchmarking results show that LightOS delivers similar average latency and even better tail latency to applications while serving the same transactional performance as directly attached NVMe SSDs. Plus, LightOS offers data services such as Data Reduction, Thin Provisioning, and Erasure Coding through the LightOS Global FTL.

Lightbits LightOS Solution

Installed on standard servers in large scale data centers, LightOS is Lightbits SDS running on a standard Linux distribution. LightOS is optimized to deliver high performance and consistently low latency for I/O intensive compute clusters, such as Cassandra, MySQL, MongoDB, FDB and time series databases.

LightOS is a patent pending multi-NVMe SSD management system and NVMe over TCP (NVMe/TCP) storage target that requires no changes to networking infrastructure or application server software.



LightOS uses a pool of storage provisioned into volumes as a unique NVMe namespace. Clients connect with NVMe/TCP client-side software that is freely available from operating system software repositories.

LightOS is designed to disaggregate storage from compute that provides:

- **Increased Application Performance:** When Cassandra updates millions of records, the LightOS Global FTL operating on a pool of flash with inline data reduction dramatically reduces tail latency. LightOS can work in software-only mode or in concert with the optional LightField hardware accelerator which reduces latency further.
- **Standard, Simple, Secure, Storage Access:** Full support for standard NVMe-oF using standard NVMe connection protocols and Access Control List functionality. Using standard NVMe-oF mechanisms, storage administrators can set Access Control Lists for each NVMe-oF volume exported by LightOS. LightOS will only let NVMe/TCP connected clients access volumes that the client has access permissions for.
- **Software-Defined Storage (SDS) Solution:** Install LightOS on your own storage servers.
- **Reduced Total Cost of Ownership:** Utilize all of your storage using thin provisioning and only buy more storage when you need it. Avoid stranded capacity and limits to scalability.
- **Separate Storage and Compute:** Only upgrade the storage servers or the application servers when you need to. No need to bundle and waste storage or CPU capacity.
- **Simple, Safe Data Management:** Data is protected from SSD failures using a highly optimized Erasure Coding (EC) algorithm managed in concert with the high-performance LightOS Global FTL, streamlining data management across the pool of SSDs, increasing performance and ensuring data is safe.

Cassandra Benchmarking Environment

First, Cassandra benchmarks were run using a typical cloud-based workload where each Cassandra node contained NVMe SSDs. A dedicated stress client machine benchmarked each Cassandra compute node. Then, the benchmark was run on the same stress clients and Cassandra nodes using disaggregated storage in a Lightbits LightOS storage server over NVMe/TCP.

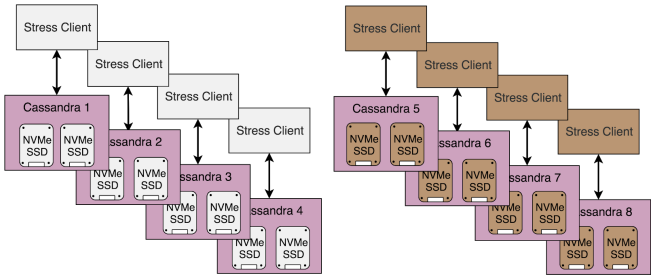
Stress Client and Cassandra Node Hardware Details

Component	Stress Client	Cassandra Node
Description	A compute node with Yahoo! Cloud Serving Benchmark (YCSB) installed.	A server with the Cassandra application running on it and with direct attached storage (in DAS benchmark runs)
Memory	64GB	64GB
CPU	Intel ® Xeon ® E5-2603 v4 @ 1.70GHz	Intel® Xeon® Gold 5120 CPU @ 2.20GHz
Network Card	Mellanox ® ConnectX-4™ LX 25GbE	Mellanox® ConnectX®-5
Storage	N/A	2x Intel 4510 2TB NVMe SSDs (4 PCIE gen 3 lanes / SSD)

LightOS Storage Server

Component	Description
Memory	512GB
CPU(x2)	Xeon(R) Gold 5120 CPU @ 2.20GHz (56HT) (SkyLake)
Network Card(x2)	Mellanox® ConnectX®-5
Storage	16x Intel 4510 2TB NVMe SSD (4 PCIE gen 3 lanes / SSD)

Benchmarking Methodology

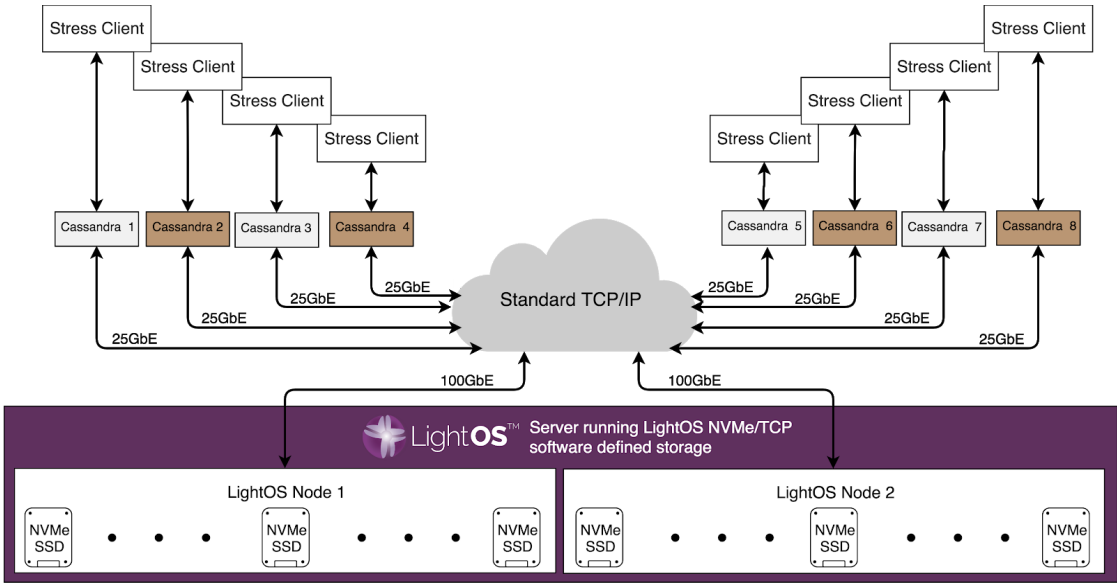


Cassandra 3.0.9 was installed on bare metal servers running Linux (Ubuntu 16.04). Cassandra DAS nodes contained two NVMe SSDs configured in RAID-0 to utilize the full capacity and IOPS available on both devices.

The Yahoo! Cloud Serving Benchmark (YCSB) tool was executed on the Stress Clients, connected to the Cassandra nodes via 25GbE over a standard TCP/IP network infrastructure.

When benchmarking Cassandra using LightOS, there were two LightOS nodes configured on the Lightbits storage server. This configuration enabled testing a scaled up Cassandra configuration. LightOS data services were also enabled such as the efficient and accelerated Erasure Coding (EC) mechanism to protect data from a drive failure at line rate performance.

Multiple Cassandra servers disaggregated to LightOS SDS Platform



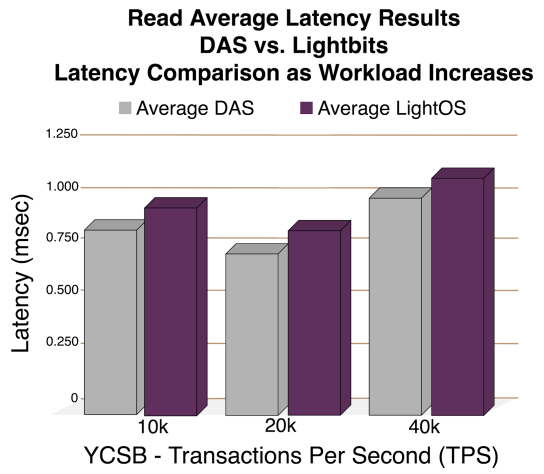
In both DAS and LightOS based configurations, the YCSB benchmarking software was configured to run on a Cassandra database where the database size was calculated to consume 65% of the usable capacity. The YCSB workload performed 50% read operations and 50% update operations, with 4K record sizes using a uniform distribution.

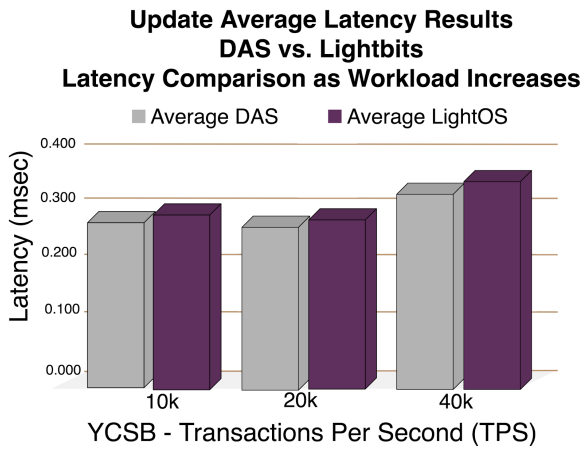
Benchmarking Results

We compared the performance of a Cassandra DAS based compute node to a Cassandra compute node using storage on Lightbits LightOS.

As can be seen in the benchmark results, when looking at the average latency, LightOS delivers the same transactional performance as DAS at very similar latency, despite performing all I/O operations over a TCP/IP network.

Read and Update ***Average*** Latency Benchmark

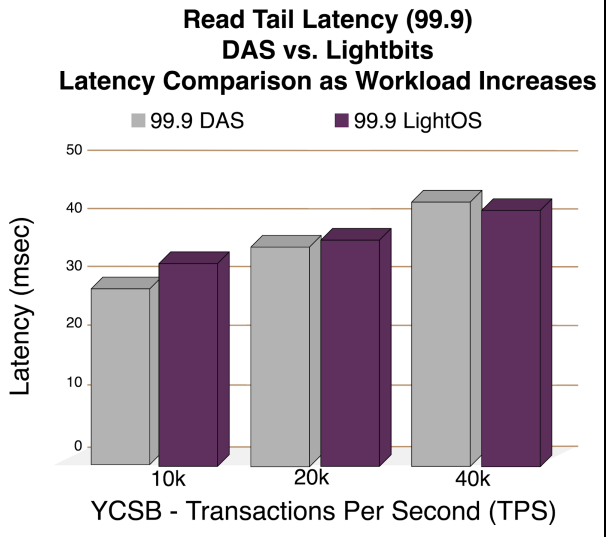


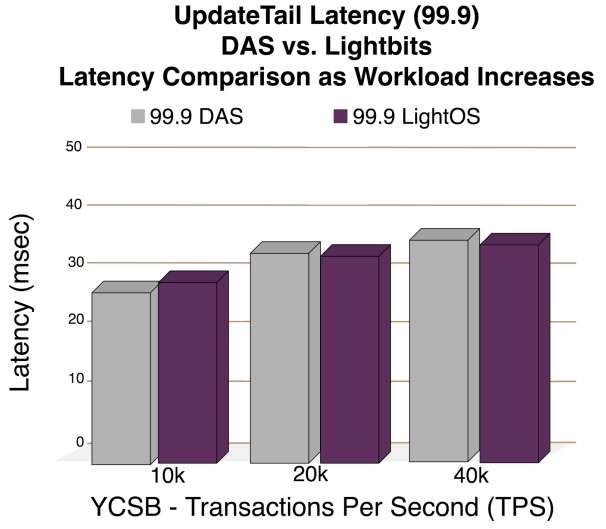


Conclusion: LightOS and DAS have similar read and update average latency.

An important metric for large scale cloud deployments is the tail latency, defined as the time it took to complete the I/O requests that required the most time to finish. When looking at tail latency with increasing workload intensity, LightOS provides lower (better) latencies.

Read and Update ***99.9%*** Tail Latency Benchmark





Conclusion: LightOS read and update *99.9%* latency is lower than DAS.

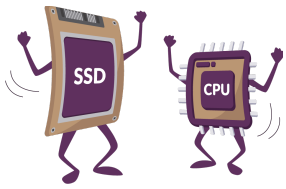
Delivering consistently low latencies is what the LightOS Global FTL was designed to deliver, versus the simple DAS-based configurations.

Conclusion

The Lightbits LightOS SDS solution enables disaggregation of storage and compute for Cassandra workloads. It can reach the same transactional performance as DAS, despite all I/Os going over your standard data center network, without hurting average latencies, and often improving tail latencies. It can supercharge your cloud database applications by providing:

- Similar performance to DAS with greater utilization of your storage investment
- Improved service levels and a better user experience with reduced tail latency
- Standard, simple, secure storage access to your application servers
- No changes to your TCP/IP network with no changes to application servers

No more managing a growing number of Cassandra servers with their own underutilized silos of direct attached storage inhibiting simple and efficient cloud-scale architecture.



Finally, you can separate storage from compute without all the drama.

For more information about the Lightbits solution, contact info@lightbitslabs.com.

About Lightbits Labs™

The Lightbits NVMe/TCP solution is the latest innovation from the Lightbits team, who were key contributors to the NVMe standard and among the originators of NVMe over Fabrics (NVMe-oF). Unlike other NVMe-oF approaches, the Lightbits NVMe/TCP solution separates storage and compute without touching the network infrastructure or data center clients. With NVMe/TCP, Lightbits delivers the same IOPS as direct-attached NVMe SSDs and up to a 50% reduction in tail latency. The transition is so smooth your applications teams won't even notice the change. Once they switch to NVMe/TCP, they can scale storage and compute independently and with consistently better user experience.

www.lightbitslabs.com info@lightbitslabs.com

US Office
1830 The Alameda,
San Jose, CA 95126, USA

Israel Offices, Kfar Saba
17 Atir Yeda Street,
Kfar Saba, Israel 4464313

Haifa
3 Habankim Street,
Haifa, Israel 3326115