IDC ANALYZE THE FUTURE

IDC TECHNOLOGY SPOTLIGHT

Sponsored by: Lightbits Labs

For web-scale infrastructures on- and off-premises, disaggregated storage results in the most efficient deployments at scale. Lightbits Labs promises easy to deploy, software-defined, disaggregated storage solutions that leverage industry standards and NVMe.

NVMe/TCP Enables the Democratization of Disaggregated, NVMe-based Storage

November 2020

Written by: Eric Burgener, Research Vice President, Infrastructure Systems, Platforms and Technologies

Introduction

Digital transformation (DX) – the move to much more data-driven business models – is under way at most enterprises today. To better use the data they are collecting from their customers and about their products, markets, and internal processes, enterprises are developing and deploying new types of workloads and modernizing their IT infrastructure. To better support the agility needed in the digital era, many of these next-generation workloads are developed as cloud-native applications that can run very effectively on web-scale infrastructure. To better meet the performance, scalability, availability and efficiency requirements of the new era, enterprises are also modernizing their storage infrastructures, moving to more server-based, software-defined architectures that support non-volatile memory express (NVMe) and cloud storage technologies.

AT A GLANCE

KEY TAKEAWAYS

- Direct-attach storage architectures pose serious obstacles to IT organizations implementing large web-scale infrastructures
- NVMe over Fabrics using TCP as a transport protocol enables cost-effective versions of the high-performance disaggregated storage architectures like the ones hyperscalers are using
- Enterprises and service providers alike are looking for off-the-shelf solutions that replicate hyperscale infrastructure but are easier to deploy and include 7x24 commercial support

As part of this evolution, many enterprises and cloud service providers are looking to the web-scale infrastructures of the hyperscalers. The use of NVMe-based solid-state media, disaggregated storage, and scale-out software-defined architectures deliver very efficient infrastructure for the hyperscalers, but enterprises and most service providers do not typically have the same sophisticated development resources available to them to build these infrastructures. What these non-hyperscaler customers need is a more off-the-shelf solution that delivers the same set of performance, availability, scalability, and efficiency benefits. A core technology to enable solutions in this area is NVMe over TCP (NVMe/TCP), an NVMe over Fabrics (NVMe-oF) implementation that runs on industry standard ethernet and TCP/IP networks.

Server-based, software-defined designs initially depended on local rather than shared storage to meet performance and cost requirements, and often require NVMe storage to deliver needed latencies. But at scale, the use of local storage has some serious limitations. Due to traditional storage network latencies, local storage architectures can only deliver low latency to local applications (i.e., applications running on the same server that houses the storage), effectively tying the applications to specific servers. Local storage approaches deliver very poor utilization of storage performance and capacity resources, making expensive storage devices like NVMe solid-state disks (SSDs) even more expensive. Other

drawbacks to local storage approaches include limited capacity, long recovery times (due to having to rebuild the data of failed devices and/or servers across a network), and a lack of enterprise-class data services like RAID, snapshots, and replication to better manage storage. For administrators interested in building highly performance, available, scalable, and efficient web-scale infrastructure at low cost, the use of direct attach storage (DAS) poses serious obstacles.

To address these problems, hyperscalers are moving much of their multi-tenant web-scale infrastructure to disaggregated architectures. NVMe/TCP will be a critical part of this solution because it supports local storage latencies in networked storage environments based on industry-standard and ubiquitous (i.e., low cost, easy to manage) hardware and software. Customers that today run iSCSI can run NVMe/TCP using the same hardware and enjoy at least an order of magnitude lower storage latencies – and there are no specialized drivers, custom hardware, or additional expenses (over and above iSCSI). NVMe/TCP enables environments that today feel they require DAS to move to disaggregated (i.e., shared) storage without giving up performance. In doing so, they can potentially enjoy all the other benefits of shared storage.

But NVMe/TCP is only part of this solution. Customers that want to share high-performance NVMe-based storage in web-scale infrastructure environments should look for software-defined solutions that use NVMe/TCP for host connections but also provide the features to deliver performance (low latency, high throughput) at scale, along with flexible data protection, fast recovery and high availability. These features enable easily scalable capacity, enterprise-class data services, flash-optimized efficiency, and allow administrators to scale compute and storage resources independently.

Benefits

The ideal environment for this type of platform (a disaggregated, NVMe-based storage solution) is one that is currently using DAS or other software-defined storage solutions based on Ceph or iSCSI. The workload targets include high-performance database environments (both SQL and NoSQL-based), big data analytics (fraud analytics and other real-time analytics, log processing, artificial intelligence and machine learning (AI/ML), Apache Kafka streams, etc.), and web-scale infrastructure environments using either virtual or container-based architectures. Because of the performance requirements, the local storage in these configurations is generally PCIe or NVMe based.

These types of customers will benefit from disaggregated, NVMe-based storage solutions leveraging NVMe/TCP in the following ways:

- They will achieve performance at least on par with their existing solutions, and in some cases as much as an order of magnitude better, at significantly less cost and with much smaller storage infrastructures that require less energy and floorspace.
- They will experience much greater efficiencies in resource utilization, be able to scale compute and storage resources independently, have access to much greater capacity scalability, enjoy faster recovery and overall much higher availability, enjoy more flexible application portability to better meet IT agility requirements, and potentially (based on specific vendor implementation) have access to enterprise-class data services that make the storage itself easier to manage.

» They will achieve all this with an industry-standard solution that requires no custom content and can leverage their existing Ethernet networking infrastructure and network configuration practices, keeping costs low.

NVMe-based All-Flash Arrays (NAFAs) are available in the market today that are disaggregated and use NVMe-oF for host connections. But their host connections use either Fibre Channel, which is too expensive for web-scale infrastructure, or Remote Direct Memory Access (RDMA)-based protocols like RoCE that require custom storage networking components and separate switch settings (and add cost as well as impose additional management overhead).

Considering Lightbits Labs: Democratizing Hyperscale Infrastructure

Lightbits Labs offers software-defined disaggregated storage based around high-performance storage technologies like NVMe and NVMe/TCP. Founded in 2016, the San Jose, California-based vendor pioneered and standardized NVMe/TCP, shipping its first products in early 2019. Lightbits today offers disaggregated, NVMe-based, enterprise-class storage solutions for web-scale infrastructures, and customers include service providers as well as enterprises that are building on-premises infrastructure around web-scale designs (which generally also includes private cloud).

Established enterprise infrastructure vendors quickly recognized the strategic nature of what Lightbits was doing, and the company's investors include Cisco Investments, Dell Technologies Capital, Intel Capital, Micron, and others. Lightbits sells both direct and indirect, although they prefer to fulfill through the channel, working with partners like Dell, HPE, Supermicro and a number of others.

Lightbits' mission is clear: to provide high-performance, highly available hyperscale storage that is easy to deploy, manage, and scale. The company provides a disaggregated, software-defined, block-based storage cluster platform that can deliver local storage latencies, includes a variety of enterprise-class data services, requires no proprietary client side software, and scales from several hundred terabytes to almost eight petabytes of raw storage capacity. The enterprise-class data services in LightOS (the storage operating system) include host multi-pathing (asymmetric namespace access), thin provisioning, compression, RAID-based data protection, quality of service, and local replication and will be enhanced in early 2021 to include snapshots and clones. The storage OS is built specifically for solid-state storage and includes flash optimizations that promote consistent performance at scale and enhance media endurance (a particularly important capability with the solution's support for very cost-effective quad level cell flash media).

Lightbits' software-defined design maximizes deployment flexibility, making it very easy to adopt. It can run on a variety of different industry standard servers from vendors like Dell, HPE, Supermicro and others as long as they are running one of the most popular Linux distributions (Red Hat, CentOS, ubuntu, SUSE, Debian or Fedora) that include the standard NVMe/TCP driver. It supports off-the-shelf NVMe SSDs in a variety of different capacities. It uses the standard Ethernet NICs that ship with every x86 server and uses the standard TCP/IP switch settings in the storage network infrastructure (no ECN, Global Pause, PFC or VLANs required). LightOS is "failure domain aware," which means that it can replicate across local domains which are user-definable at the rack, row or power zone level, enabling extremely resilient high availability configurations.

The performance of the Lightbits Storage Cluster is reportedly impressive. A system with only two storage nodes (storage clusters can support up to 16 nodes) can deliver up to 6M 4K read IOPS and 1.6M 4K write IOPS with latencies, according to the company — well under 500 µs (average read latencies are as low as 170 µs) and far outperforming comparably configured Ceph or iSCSI configurations. The concentrated performance density of the Lightbits solution allows much

smaller systems to service a given performance requirement, lowering cost as well as energy and floorspace consumption. For many workloads, a Lightbits solution will far outperform a Ceph configuration with the same usable storage capacity with less than half of the required infrastruc-ture (representing a significant cost savings). Lightbits shared storage solutions are also half the cost of DAS-based configurations, primarily because they are so much more efficient at sharing the performance and capacity resources of NVMe-based storage.

Lightbits also provides an excellent platform for cloud-native applications. Its disaggregation makes applications quick to restart and easy to move to other servers. A cloud-native solution should have programmatic interfaces (such as Container Storage Interface (CSI)) that allow for common management of containers across both public and private clouds. Lightbits supports CSI, OpenStack Cinder, Kubernetes, Docker, and a REST API for management. And it uses the same server-based, software-defined web-scale architecture that has proven so efficient at scale in hyperscaler environments – but in an off-the-shelf solution that's much easier to deploy and enjoys 7x24 commercial support.

Challenges

NVMe/TCP is a new protocol, and prior industry experience with iSCSI will no doubt have enterprises wondering whether TCP can support high-performance networked storage. Lightbits anticipated this concern and has addressed it by releasing benchmark data to validate its performance claims. In addition to the Ceph comparison quoted earlier that showed better performance at half the cost, a Cassandra benchmark (run by Storage Solutions Engineering at Micron) against local NVMe-based configurations shows comparable or better performance, improved reliability, and access to data services like thin provisioning and compression, as well as better support for containerization and lower cost.

Intel did performance testing of NVMe/TCP with Intel NICs using Application Device Queues (ADQ) showcasing latency reductions that significantly narrow the latency gap between RoCE and NVMe/TCP. In a direct comparison against iSCSI running on the same hardware, Lightbits reportedly showed 6x the IOPS at significantly lower (4x lower) read latencies. The whole point to NVMe/TCP is to enable high-performance, disaggregated NVMe-based storage solutions on industry-standard hardware, so prospects are right to want to look into the performance differences across various workloads, but Lightbits apparently delivers the goods.

Conclusion

In recent years, hyperscalers have introduced several infrastructure innovations that have ultimately been adopted by IT organizations across all industries. What unlocks these innovations for more widespread use, however, are vendors that create off-the-shelf, easy-to-deploy solutions leveraging those innovations. The move to disaggregation by the hyperscalers for web-scale infrastructure at scale is a particularly interesting development in the storage space, and Lightbits has created a solution that promises to democratize this innovation.

Lightbits' storage cluster is a highly available, disaggregated platform that lets customers very efficiently share high performance, NVMe-based storage in a software-defined solution that is completely based on industry standards. As the pioneer in NVMe/TCP, a critical foundation technology for this type of disaggregated solution, Lightbits is in an excellent position to succeed by enabling

Lightbits Labs delivers software-defined, disaggregated storage based around cost-effective industry standards and high-performance storage technologies like NVMe and NVMe/TCP.

customers to build high-performance, highly scalable, and very flexible NVMe-based web-scale infrastructure with a compelling total cost of ownership.

About the Analyst

Eric Burgener, *Research Vice President, Infrastructure Systems, Platforms and Technologies*

[Eric Burgener is Research Vice President within IDC's Enterprise Infrastructure practice. Mr. Burgener's core research coverage includes storage systems, software and solutions, content infrastructure, quarterly trackers, and end-user research as well as advisory services and consulting programs. Based on his background covering enterprise storage, Mr. Burgener's research includes a particular emphasis on solid-state technologies in enterprise storage systems as well as software-defined infrastructure. He was awarded the Alexander Motsenigos Memorial Award for Outstanding Innovation in Market Research in 2017 by IDC and is an active participant in the IT Buyer's Research Program at IDC.

MESSAGE FROM THE SPONSOR

Lightbits Labs LightOS Hyperscale Storage for All is composable NVMe/TCP block storage that unleashes infrastructure flexibility by performing like local flash. It maximizes performance and capacity utilization, scaling storage infrastructure with ease.

LightOS provides rich data services: Logical volumes, thin provisioning, compression and selectable redundancy levels at a fraction of the cost of traditional, proprietary arrays.

- Software-defined on standard servers, SSDs and TCP/IP networking
- High IOPs and throughput, consistent low latency
- Drive, network path and storage server fault tolerance

LightOS is architected for cloud-native applications and is dynamically provisioned via CLI, RESTful API, Kubernetes (CSI) and Openstack (Cinder).

Others may add the NVMe/TCP protocol to their existing products but without architecting for NVMe end-to-end, they likely suffer the same bottlenecks they already have. We hope you'll check us out yourself: https://media-

www.micron.com/-/media/client/global/documents/products/white-paper/wp-hse_mongodb_with_lightbits.pdf?rev=65df6bccc5fc49d78dcd99c139af13db

IDC Custom Solutions

The content in this paper was adapted from existing IDC research published on www.idc.com.

IDC Research, Inc.
5 Speen Street
Framingham, MA 01701, USA
T 508.872.8200
F 508.935.4015
Twitter @IDC
idc-insights-community.com
www.idc.com