

Case Study

Intel® Xeon® Scalable Processors
Intel® Optane™ SSDs
Intel® QLC 3D NAND SSDs



DUG Technology Modernizes with Exascale Flash Storage and Seizes New Opportunities Beyond Oil and Gas

DUG Technology switched from hard disk drives (HDDs) to petabytes of flash storage with VAST Data Universal Storage, powered by Intel® Technologies, to boost performance and reliability.



Seismic analysis is a high-performance computing (HPC) discipline that pieces together what lies under the surface of the earth from nothing more than the reflection of sound. To come up with useful 3D analyses requires petabytes (PB) of data and thousands of powerful computers. Not even major oil companies possess all of the computational resources necessary to conduct all of this analysis in-house, so they turn to companies like DUG Technology to tease out details from their mountains of data.

DUG refers to this capability as HPC-as-a-service (HPCaaS): specialized, full-stack exascale computation available on demand. Traditionally, DUG's compute-as-a-service technology was available only to specific customers, such as major oil and gas companies. As the market took notice of its capabilities, DUG expanded its offering to other industry verticals that use this same service to tackle a diverse set of extreme computational needs.

DUG decided to bring the same “bring-nothing-but-your-data” ease of service to businesses outside of the energy sector. DUG knew that it could serve these new industry verticals economically because of the specialized DUG McCloud service for HPC. VAST Data Universal Storage, powered by Intel® Technologies, undergirds DUG McCloud and enabled DUG to successfully break into new verticals, including academia, astrophysics, medicine and genomics, wildfire modeling, and COVID-19 research. However, getting to this point required a sea change in how DUG dealt with its storage.

About DUG Technology

With more than 17 years of experience and data centers in Perth, Houston, London, and Kuala Lumpur, DUG Technology is at the forefront of HPC. It combines innovative hardware and software solutions that enable clients to make use of large and complex datasets. DUG Technology's industry experience and strong grounding in applied physics has equipped it to provide state-of-the-art HPCaaS delivered either direct-to-client or via its DUG McCloud platform.

Challenge

For its first decade of operation, DUG had been deploying and managing HDD-based storage to deliver the scale and cost economy that its seismic workloads required. During that time, DUG thoroughly optimized its applications to make use of the capabilities, and avoid the limits, of its Lustre HDD-based infrastructure. Here, DUG had to make many compromises. For example, when Lustre file system clients would hit peak throughput for a given workflow, other users sharing the same file system would suffer. From a resilience perspective, although DUG designed its software to protect against HDD failures, the need to swap out failed drives on a weekly basis was a constant thorn in DUG's side.

Finally, while DUG's applications were well optimized for Lustre and HDD storage, the new applications that DUG was evolving to support all handled storage input/output (I/O) differently. Storage versatility and multitenancy became vitally important to DUG; any new solution would need to support a broad set of requirements and to support them at exascale. DUG also needed storage that could handle the multiplicity of throughput requirements for different applications. DUG looked to solid state drive (SSD)-based storage to provide higher performance



Immersion-cooled servers at a DUG data center

and reliability. However, moving to SSDs on Lustre would have been prohibitively expensive, and affordability was paramount for DUG.

In order to build a resilient and adaptive storage environment that enabled expansion into new markets, DUG required a new approach to storage.

Solution: VAST Data Universal Storage

DUG chose VAST Data Universal Storage to expand its business and support the needs of a wide diversity of new markets and customers. The Universal Storage offering combines the speed and scale of a parallel file system with a new level of flash affordability and multitenancy to deliver a complete technological leap forward for DUG. VAST Data's disaggregated shared everything (DASE) architecture also provides consistent performance by isolating non-optimized I/O so as not to impact other tenants. With the DASE approach, VAST Data eliminates the concurrency challenges of parallel storage to deliver high performance for specific workloads that does not come at the expense of other workloads.

Beyond significantly improving the customer performance experience, VAST Data provides a combination of reliability, management, and support that is not otherwise found with legacy HPC storage technologies. VAST Data's DASE architecture supplies exascale scalability, enabling DUG to grow to tens of petabytes of flash storage with no single points of failure in an architecture that can quickly recover from failure. The reliability of the DASE architecture comes "for free": it is a direct result of VAST Data's data-protection efficiency and the architecture's statelessness. Beyond resilience, VAST Data Universal Storage also simplifies the deployment and management experience for DUG by providing an integrated scale-out appliance that consistently pushes out new features that are automatically applied while

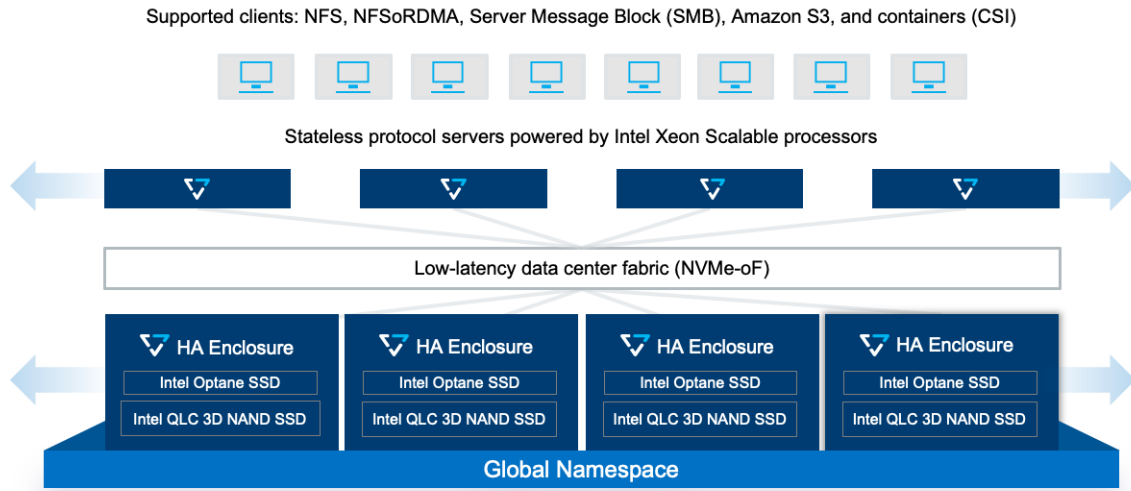
the system is online, so there's no downtime for DUG.

Overview of VAST Data Universal Storage with Intel storage technologies

VAST Data Universal Storage provides a single, global namespace so that each application has access to all of the associated data for that workload. The VAST Data solution combines all-flash drive performance, massive scalability, the economics of archive storage, and the simplicity of plug-and-play network-attached storage (NAS) connectivity.

Intel® SSDs provide the hardware basis for the cost-efficiency and reliability of VAST Data Universal Storage. Intel's pairing of vertical floating-gate technology and complementary metal-oxide-semiconductor (CMOS) under-array architecture delivers the highest areal density (gigabytes of storage per square millimeter) in the industry for the same bits per cell.¹ This means that Intel® QLC 3D NAND SSDs provide not only greater areal density than previous-generation triple-level cell (TLC) media, but greater areal density and higher reliability than competing quad-level cell (QLC) designs based on charge-trap technology.¹ The architectural innovations from Intel enable the VAST Data solution to economically store all data on flash drives. The cost effectiveness and high reliability of Intel QLC 3D NAND SSDs provides the foundation for VAST Data's architecture to reduce costs by up to 85 percent compared to HDDs, providing a dollar-per-gigabyte (GB) cost similar to that of HDD-based systems over 10 years.^{2,3}

Intel® Optane™ SSDs further accelerate write performance for workloads running on VAST Data Universal Storage. Crucially, Intel Optane SSDs buffer writes to storage, which enables full QLC erase-block writes. The low latency, high endurance, and high 4K random-write performance of Intel Optane SSDs help ensure that long-term and short-term data are not co-located in large QLC blocks. Intel Optane SSDs shield Intel



Logical diagram of the VAST Data Universal Storage solution

QLC 3D NAND SSDs from inefficient write behavior, which is one reason VAST Data can offer a 10-year SSD endurance guarantee while also delivering the economic benefit of cost-effective QLC NAND.^{2,3}

Storage capacity, cost, and capability are only part of the VAST Data Universal Storage story, however. The VAST Data solution is also quite sophisticated in the implementation of new algorithms that pioneer all-new levels of data-reduction and data-protection efficiency.⁴ VAST Data Universal Storage brings all of these architectural aspects together with 2nd Gen Intel® Xeon® Scalable processors to implement a new class of global algorithms in a DASE cluster.⁴ These processors provide the computation power underlying VAST Data Universal Storage and vital acceleration libraries. The storage performance development kit (SPDK) serves as an accelerant for VAST Data Universal Storage to deliver low-latency access from every CPU to every QLC and Intel Optane SSD. The SPDK thereby eliminates the need for complex and volatile cache-coherency operations that can otherwise inhibit scale in legacy shared-nothing storage architectures.

VAST Data Universal Storage interconnects CPUs with NVM Express (NVMe) devices using the NVMe over fabrics (NVMe-oF) protocol to provide distributed scale with the performance and latency of direct-attached storage (DAS).⁵ NVMe-oF runs over standard Ethernet or InfiniBand networks to enable the disaggregation of resources and a shared-everything architecture over commodity data center fabrics. The VAST Data connection exposes the system via ubiquitous protocols such as network file system (NFS), server message block (SMB), and an Amazon S3-compatible API, so that applications that consume universal storage do not require specialized adapters, formats, or protocols.

VAST Data changed how DUG handles data

DUG has been fully in production with VAST Data since December 2019 at DUG's data centers in Houston, Texas, and Perth, Australia, with plans for further expansion. In fact, DUG plans to double its compute capabilities in Houston and more than double those capabilities in Perth during 2020 and 2021. Fortunately, the VAST Data solution becomes more reliable, not less so, as it grows.

DUG's data-storage needs have always been large. Seismic processing projects arrive at DUG with more than 1 PB of data, and they experience a 6–8x expansion in the course of processing. During a single seismic-processing project, DUG will copy and write that data up to 50 times—and DUG typically has more than 100 projects running simultaneously at any given time. VAST Data Universal Storage is perfect for this type of data growth, and it helps DUG ensure that competing applications all experience performance fairness on a shared HPC computing resource.

Beyond efficiently handling huge quantities of data, VAST Data's data reduction is another attraction. For DUG, this is a cost-reducer. Even with seismic data, which is notoriously difficult to reduce, VAST Data's powerful data reduction capabilities can save significant amounts of money. DUG sees greater savings through data reduction with different workloads using VAST Data's new similarity-based approach to global data compression.

Another advantage for DUG is that VAST Data remotely manages the storage for DUG 24/7. This is the first time that DUG has benefited from having a vendor provide remote appliance management for its storage. DUG experiences zero downtime for updates, and its IT admins can feel confident knowing that VAST Data is closely monitoring the performance and availability of their environment. Because of this, DUG can expand storage capacity without having to grow its storage team.

Storage as a strategic asset

DUG's successful move into new markets was made possible by VAST Data Universal Storage, powered by Intel technologies. The VAST Data storage solution provided DUG with the capacity, performance, and reliability to get rid of HDDs, move beyond complex HPC file-storage technology, and provide a leadership-class customer experience for customers within and beyond the oil and gas industry. An all-silicon storage offering provides the consistency and diversity of high performance that makes it possible for DUG to efficiently build out its multitenant cloud environment for its next wave of growth. The storage, reliability, and ease of management afforded by VAST Data has turned storage into a strategic asset for DUG, and has enabled it to better achieve its broader business goals.

Learn more

DUG Technology

VAST Data

VAST Data Exascale NAS White Paper

Intel® Xeon® Scalable processors

Intel® Optane™ SSDs

Intel® QLC 3D NAND SSDs



¹ Comparing areal density, Intel measured data on a 512 GB Intel 3D NAND SSD and on representative competitors based on 2017 IEEE International Solid-State Circuits Conference papers citing Samsung Electronics and Western Digital/Toshiba die sizes for a 64-stacked 3D NAND component. Source: ISSCC 2018; H. Maejima; ISSCC 2019 C. Siau.

² VAST Data. "Redefining Storage Economics." vastdata.com/economics/.

³ VAST Data. "Zero Compromises Guarantee." April 2019. vastdata.com/wp-content/uploads/2019/06/VAST_Data-Zero-Compromises-Guarantee.pdf.

⁴ VAST Data. "Universal Storage: Innovation to Break Decades of Tradeoffs." February 2020. vastdata.com/wp-content/uploads/2019/04/VAST-Data-Overview.pdf.

⁵ VAST Data. "Architecture." <https://vastdata.com/architecture/>.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors.

Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit www.intel.com/benchmarks.

Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates. See backup for configuration details. **No product or component can be absolutely secure.**

Your costs and results may vary.

Intel technologies may require enabled hardware, software or service activation.

Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice.

© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.