Solution Brief

Health and Life Sciences

Transporting High-Throughput Instrument Data from Edge to Cloud

IT teams tasked with moving high data volumes can benefit from Zettar's scale-out, highly available, petabyte-scale data-transfer software solution, powered by Intel.

Instrument-generated data is exploding

Large volumes of data from high-throughput instruments are used by health and life sciences (HLS) researchers to fuel breakthroughs in genomics, medicine, and biology. For example, the latest Illumina NovaSeq 6000 next-generation sequencing (NGS) S4 flow cell system, used by geneticists to understand the structure of DNA, generates up to 3 TB of data per run.[1] A Thermo Fisher Scientific Krios Cryo-Transmission Electron Microscope (Cryo-TEM) with a Thermo Fisher Scientific Falcon 4i Direct Electron Detector generates 7–10 TB of data per day, giving structural biologists visibility into the nature of atoms.[2] Once an instrument generates valuable data, IT teams are responsible for quickly moving that data either to the cloud or to an on-premises compute cluster for processing and analysis.

IT teams face challenges moving instrument data

Today's IT teams face challenges when moving data generated by high-throughput instruments. Current solutions are costly and complex, and they can constrain IT agility. Zettar, a deep-tech innovator in the data-movement field, offers an innovative alternative for moving data at scale and speed.[3]

The company holds the record for the Supercomputing Asia Data Mover Challenge for moving data at a rate of 68 gigabits per second (Gbps).[4] Zettar has also worked with the US Department of Energy to reach a mean data-transfer speed of 77.25 Gbps over their Energy Sciences Network (ESnet) 100G network.[5]

Zettar zx data-transfer software, running on servers powered by Intel® Xeon® Scalable processors and Intel® Optane™ solid state drives (SSDs), can speed data transfers in a variety of scale-out deployments. Zettar zx software transports data up to 10x faster than current-generation solutions.[5] Also, according to Chin Fang, CEO, "The Zettar data transfer software solution can deliver petabyte-scale data transfer at up to 75 percent lower TCO than other leading data-mover platforms."

Table 1 provides an overview of two use cases, one for genomics and one for structural biology. These use cases illustrate how Zettar zx software transports high-throughput instrument data and uses Intel® technology in edge-to-cloud or compute cluster scenarios.

Table 1. Two Zettar zx high-throughput instrument data-transport use cases powered by Intel technology

	Genomics use case	Structural biology use case
Description of high-throughput instruments	NGS systems determine nucleotide sequences and detect mutations in genetic data.	Cryogenic electron microscopes (Cryo-EMs) generate movies of cells, viruses, bacteria, and molecules at nearly atomic resolution. The server converts movies to image frames.
Objective	Accelerate understanding of gene expression, genetic variations, and human diseases.	Elucidate the structure of molecules, proteins, and other biological structures for use in drug discovery.
Data transport challenge	Move NGS data from a sequencer to genomic analysis pipelines.	Move data from the Cryo-EM microscope to the cloud/compute cluster in near real time.
Data transport model	Moving bulk data transfer at maximum possible throughput.	Moving extremely time-sensitive new data.

Genomics use case

Challenge: The secondary analysis process of an NGS system generates a 100 GB Binary Alignment/Map (BAM) file and a 30 GB genomic variant call format (gVCF) file. Approximately 350 GB are needed for storage-prediction requirements. See Table 1 for use case details.

Solution: Zettar zx software sits on the edge server powered by Intel Xeon Scalable processors and Intel Optane SSD P5800X drives, the world's fastest data center SSDs.[6] The Zettar zx software implements a "store-and-forward" model to move data. Through a combination of integrated software and underlying hardware capabilities, Zettar zx software transports data at a speed of 100 Gbps or higher, depending on the infrastructure and setup.[7]

Genomics use case data model:
Bulk data transfer at the highest possible throughput

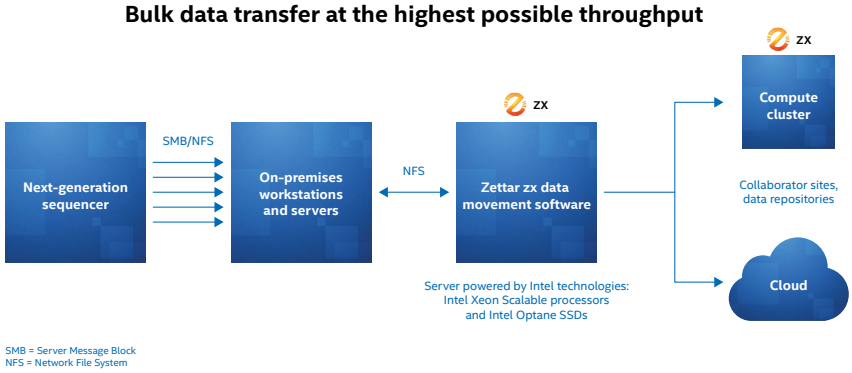

Figure 1. Zettar zx software sits on an edge server built with Intel technologies, and it forwards data to a compute cluster or the cloud

Structural biology use case

Challenge: A modern Cryo-EM can generate up to 7–10 TB of data per day. These massive datasets are needed so biologists can view interactive, animated 3D images down to the atomic level. See Table 1 for use case details.

Solution: Zettar zx software sits on an edge server and transports data to an upstream pipeline at a rate of 100 Gbps or higher.[5] The software uses the performance of the underlying infrastructure, including Intel Xeon Scalable processors and Intel Optane SSDs.

Accelerate breakthroughs with Zettar and Intel

HLS researchers on the forefront of scientific discovery use high-throughput instruments that generate high volumes of data. Zettar offers a simple, cost-effective, and versatile software solution for transporting this data to the cloud or to compute clusters where components such as Intel Xeon Scalable processors and Intel Optane SSDs are used. Zettar zx software transports data up to 10x faster than alternative data-mover solutions.[5] Faster data movement can lead to faster medical and scientific breakthroughs, improving the quality of life for all.

Learn more:

Intel health and life sciences technologies: intel.com/health

Intel Xeon Scalable processors: intel.com/xeon

Intel Optane memory and storage: intel.com/optane

Contact the Zettar sales team: zettar.com/contact-sales

[1] 6 TB over two days = 3 TB over one day. Source: Illumina. "Illumina Releases NovaSeq S4 Flow Cell and NovaSeq Xp Workflow." illumina.com/company/news-center/press-releases/press-release-details.html?newsid=dbfb9a3d-e784-41ba-ade9-145366bf650e.
[2] Based on direct communication with Thermo Fisher as of April 28, 2022 and referencing Thermo Fisher. "Falcon 4i Electron Detector." thermofisher.com/us/en/home/electron-microscopy/products/accessories-em/falcon-detector.html.
A rate of 7–10 TB per day for the Cryo-EM microscope assumes a 50–75 percent utilization rate.
[3] For more information on deep tech, see: Feedough. "What Is Deep Tech? – Use Cases, Examples, & Future." April 2021. feedough.com/what-is-deep-tech/.
[4] National Supercomputing Centre (NSCC). "Data Mover Challenge – WINNERS Announced." March 2019. sc-asia.org/2019/wp-content/uploads/2019/03/Data-Mover-Challenge-Release-13-March-2019.pdf.
[5] Testing done by ESnet in January 2021. Performance testing results source: Light Bytes. "100G DTN Experiment: Testing Technologies for Next-Generation File Transfer." January 2021. https://lightbytes.es.net/2021/01/12/100g-dtn-experiment-testing-technologies-for-next-generation-file-transfer/. For test configuration details, see: ESnet. "100G SDN Testbed." es.net/network-r-and-d/experimental-network-testbeds/100g-sdn-testbed/.
[6] Source: claim 14 at Intel. "Performance Index: Intel® Optane™ SSD P5800X Series." https://edc.intel.com/content/www/us/en/products/performance/benchmarks/intel-optane-ssd-p5800x-series/.
[7] Zettar and ESnet. "High-Performance Data Movement Services – DTNaaS: A behind the scene look." 2021. youtube.com/watch?v=LIQxIZVAxks.

Printed in USA 0722/EH/PRW/PDF Please Recycle 351062-001US