Accelerating Genomics Data Processing with Persistent Memory and Big Memory Software

How MemVerge Memory Machine™ and Intel® Optane™ Persistent Memory Helped TGen Achieve a 36% Increase in their RNA Sequencing Pipeline

TOPIC OVERVIEW

As the size and complexity of datasets expands and the rapidity with which they need to be processed increases, organizations are facing the challenge of having to optimize their infrastructure to handle a simultaneous growth in both volume and velocity. The rise of data-intensive workloads such as real-time analytics, AI/ML, animation, and bioinformatics is driving demand for memory-optimized hardware which will be able to process large datasets in a minimum amount of time. However, present compute infrastructure is not always capable of supporting the complex configurations that these growing demands for memory technology require.

AT•A•GLANCE

> **ISSUES**
  Insufficient memory for large sample pipelines and no fault-tolerance.

> **SOLUTION**
  - phoenixNAP's Hardware-as-a-Service
  - (HaaS) 2nd Generation Intel® Xeon® Scalable Processors
  - Intel® Optane™ Persistent Memory
  - MemVerge Memory Machine

> **RESULT**
  - 36% increase in single cell RNA-seq performance
  - Improvement in SplAdder* analysis from 2 months to 13 days

This white paper considers current memory deployment trends and practices, providing a brief market analysis to identify the challenges that modern companies are facing. It also covers the recent technology developments that aim to address these challenges and gives a best practice example of memory optimization for the purposes of genomics data processing. Through the use case of The Translational Genomics Research Institute (TGen), an affiliate of City of Hope, the paper examines the benefits of using next generation computation technology developed by MemVerge on phoenixNAP's Hardware-as-a-Service platform to virtualize Intel® Optane™ Persistent Memory.

THE STATE OF BIG MEMORY

In the mature phase of the digital transformation, organizations are generating massive amounts of digital data that needs to be processed and delivered in real time. This type of data is often referred to as "big and fast" data and it requires advanced compute to be handled efficiently. Multimedia streaming, e-commerce purchases, e-banking transactions, real-time advertising, and analytical biosciences provide some examples of datasets that need to be loaded into memory for accelerated processing time. This increases the need for memory, forcing organizations to constantly expand their memory capacity. As a result, the memory technology market is booming and this is likely to remain a trend in the years to come.

According to a Research and Markets report from November 2021, the in-memory market is expected to grow at a CAGR of around 25% during 2021-2026. Memory optimizations will continue to be vital for the support of accelerated data growth, but this does not mean that organizations will have to keep adding more capacity. Instead, new technologies will help them bridge the gap between "capacity" and "need," which has been growing in recent years due to limitations of traditional memory and storage technologies.

COMPUTE DEMAND ACCELERATING[1]

CPU Core Growth Projection Over Time

Traditional DRAM is fast and can meet the most sophisticated performance needs. However, it is both significantly more costly than 3D NAND SSDs and limited in its scaling potential as most servers are not built to support advanced memory deployments. For the most common business purposes, scaling DRAM is too expensive to justify the cost-to-performance ratio. NAND SSDs, on the other hand, are too slow to support large and complex data processing.

CAPACITY VS NEED IS GROWING[2]

Memory need growth

DRAM scaling phase 3 2x/4 years

These technical issues, translate into the following business challenges:

- **Increased memory costs** as more capacity is needed to support constantly growing datasets.
- **Hardware limitations** by memory slots and RAM size.
- **IT infrastructure inefficiency** due to performance issues and limited ability to scale
- **High cost-to-performance ratio** for traditional DRAM deployments
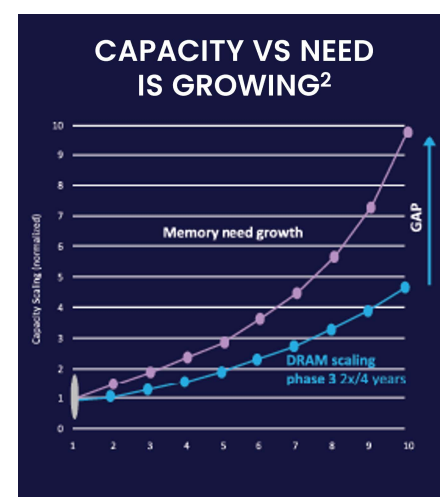
To help overcome the capacity vs. need gap and provide a cost-effective solution for big memory processing, Intel Optane technology uses a unique architecture which provides near-DRAM speeds with the persistence of traditional SSDs. Combining the best features of both DRAM and NAND drives, this technology can be adjusted to each specific use case.

Another important solution to the big memory processing challenge is offered by MemVerge Memory Machine, the first software that virtualizes DRAM and persistent memory to make it possible for data to be accessed, tiered, scaled, and protected in-memory. In the case of workloads that need to analyze large amounts of data and perform operations promptly, this software allows for quick memory scaling while keeping costs under control.

The combination of these two technologies on the 2nd Gen Intel Xeon Scalable processor platform deployed by phoenixNAP has helped TGen accelerate their genomics processing pipelines.

THE MEMORY CHALLENGE: GENOMICS DATA PROCESSING

The Translational Genomics Research Institute (TGen), an affiliate of City of Hope, is a non-profit organization based in Phoenix, Arizona. Dedicated to "conducting groundbreaking research with life-changing results", the organization analyzes the genetic components of neurological disorders, infectious diseases, and rare childhood illnesses to develop earlier diagnostics and smarter treatments for various common and complex diseases. Founded in 2002, the Institute strives to bring more rationality and precision into global healthcare by personalizing treatments on the basis of the underlying genetic components of a specific disease.

Working on projects that rely on massive amounts of unstructured data, TGen needs advanced computational power in order to derive value from such data. TGen has been analyzing the expression of ~30K genes in ~114K individual cells, a 3.42 billion element matrix, with the purpose of identifying relevant features and patterns in some of their single-cell RNA-Seq experiments. Besides the sheer number of genes and cells that needed to be analyzed, the other major challenge for TGen was that the number of cells in an experiment kept constantly growing. They were struggling to maximize throughput of long-running RNA-Seq pipelines and provide fault tolerance for faster results.

To address these issues, TGen worked with phoenixNAP, Intel, and MemVerge to deploy a customized infrastructure that would help them optimize their memory use.

"We needed to increase our memory capacity without adding more compute cores. Intel Optane Persistent Memory provided us with a unique solution to address that challenge. MemVerge Memory Machine software allowed us to virtually pool together and manage Optane PMEM and DRAM to provide the memory capacity and performance we needed."

Glen Otero, PhD
VP of Scientific Computing

BIG MEMORY SOFTWARE BY MEMVERGE

MemVerge is an innovative startup company founded in 2017 with the aim of opening the door to Big Memory Computing via its Memory Machine Software. Recognizing the potential of persistent memory technology to change how organizations perceive and deploy memory and storage, MemVerge created a way to facilitate the transition.

Its Memory Machine is the first in a new class of Big Memory software solutions built to transform memory infrastructure for maximized processing and scaling efficiency. The software virtualizes DRAM and Persistent Memory, making it possible for organizations to access DRAM-like capabilities without the need to modify multiple applications. This approach ensures optimized memory usage and increased performance of demanding workloads and applications such as real-time analytics and AI/ML apps.

The solution provides the following key capabilities:

- A Memory Virtualization Platform
- Software-Defined Memory Service
- Memory tiering for maximized use
- Low-latency memory replication
- Quick database recovery

"The potential of MemVerge Memory Machine is huge and our collaboration with Intel, TGen, and phoenixNAP has helped us demonstrate it. TGen was facing a challenge that is quite common and our solution helped them transform their existing infrastructure to quickly derive value from their data. We are honored to have had a chance to contribute to their research and we hope our technology will find many more applications in the field of life sciences."

Charles Fan,
CEO of MemVerge

INTEL® OPTANE™ PERSISTENT MEMORY

Recognizing the need for a completely new memory/storage hierarchy, Intel has built a unique memory architecture that provides advanced capabilities at an affordable price point. Intel Optane Persistent Memory (PMem) is an innovative memory technology solution that addresses the emerging need for fast, scalable, and cost-efficient data processing.

Offsetting the limitations of DRAM to feasibly extend memory capacity and NAND's inefficiency to serve as memory, PMem combines features of the two to allow customized application based on a specific use case. The architectural design allows for memory cells to be individually managed, which helps eliminate garbage collection and maximize the use of available resources. This way, PMem enables workload-optimized deployments that provide near-DRAM speeds while ensuring persistence like traditional SSDs.

"Intel Optane Persistent Memory disrupts the existing storage/memory hierarchy, transforming the ways organizations can derive value from complex data. Thanks to its flexibility, the solution can have a wide variety of applications in the Life Sciences sector and numerous other industries that are rapidly adopting AI/ML and HPC. TGen's research demonstrates the potential of this technology and we are honored to be able to support their much-valued work ."

Michael J. McManus, PhD
Director, Precision Medicine & Principal Engineer, Intel

BUILDING A CUSTOMIZABLE INFRASTRUCTURE

The diversity and complexity of TGen's projects requires a customized infrastructure solution that can quickly adapt to the organization's changing needs. phoenixNAP's Hardware as a Service (HaaS) platform has been powering TGen workloads for years, providing stability, capacity, and security for its services and applications. TGen is also using phoenixNAP's API-driven dedicated server solution Bare Metal Cloud for workloads that need to be scaled quickly and securely across the globe. As an automation-driven platform, Bare Metal Cloud provides nearly instant access to burst resources and advanced flexibility for dynamic workloads.

All phoenixNAP infrastructure solutions are powered by Intel Xeon technologies, ensuring consistent performance for TGen's data-intensive workloads. For its most complex projects, TGen is using customized implementations of Intel technologies to ensure maximum efficiency and deliver results faster. Its current infrastructure runs on 2nd Generation Intel Xeon Scalable processors, which provide support for the latest generation PMem solution, giving TGen the tools and resources necessary to handle its data more efficiently.

With Intel Optane Persistent Memory deployed through MemVerge Memory Machine, TGen accelerated its single-cell RNA-seq pipeline by 36%. This increase means the organization is able to deliver results faster, while lowering its overall IT spending. The organization also reported a reduction in the time required to process a typical SplAdder analysis of 2000 samples from 2 months to 13 days, a ~79% reduction.

"Every second we gain in performing these complex analyses translates into added value for our overall research. We appreciate the support of Intel, MemVerge, and phoenixNAP on all our projects, as their technologies make an essential difference in how we work."

Glen Otero, Ph.D.
VP of Scientific Computing

CONCLUSIONS

The innovative approach to memory technology by Intel provided TGen with a critical resource to accelerate its groundbreaking research. Combined with MemVerge Memory Machine, this technology enabled TGen to optimize its memory use and ensure streamlined scaling as their datasets grow. This implementation demonstrates the potential of new memory technologies and opens a window of opportunity for all organizations working with data-intensive applications.

Notes

1. SplAdder is a splicing toolbox that uses RNA-Seq alignment data to analyze alternative splicing. The tool helps identify, qualify, and test alternative splicing events from RNA-Seq data for the purposes of differential analyses.

2. Intel Optane Persistent Memory can be run in two modes: Memory and App Direct. In the Memory Mode, PMem can be used as an expanded volatile system memory and enabled without any system modifications. The App Direct mode needs OS updates to leverage memory persistence, as well as byte addressability and cache coherence. Through these two modes, PMem can be seamlessly adjusted to a wide variety of use cases focusing on in-memory databases and in-memory analytics frameworks.

Sources:

1. Research and Markets: In-Memory Computing Market: Global Industry Trends, Share, Size, Growth, Opportunity and Forecast 2021-2026, available at: https://www.researchandmarkets.com/reports/5441938/in-memory-computing-market-global-industry

2. Source for "Compute Demand Accelerating" Intel Market Intelligence Group - DRAM Market Monitor Q1-20" by Yole Development

3. Source - Intel. Results may vary.

4. Source - Intel. Results have been estimated or simulated using internal Intel analysis or architecture simulation or modeling, and provided to you for informational purposes. Results may vary.

ABOUT phoenixNAP

phoenixNAP is a global IT services provider offering progressive Infrastructure-as-a-Service solutions from locations worldwide. Our bare metal server, cloud, hardware leasing and colocation options are built to meet the evolving technology demands businesses require without sacrificing performance. Scalable OpEx solutions to support with the systems and staff to assist; phoenixNAP global IT services.

Contact phoenixNAP at: **sales@phoenixnap.com**
or **855.330.1508 | www.phoenixnap.com**