Intel memory tiering on Red Hat OpenShift

Highlights

Modernize applications and real-time database infrastructure and unify management by deploying Red Hat OpenShift.

Provide cloud-native development and enable DevOps methodologies, while deploying on-premise, in the cloud of your choosing, or with hybrid cloud.

Employ memory tiering strategies that vastly expand memory capacities on individual servers at lower cost while providing greater availability.

Automate the creation, configuration, and management of Kubernetes-native applications with Red Hat OpenShift

Modern applications and infrastructure for a cloud-native world

Many organizations face contradictory pressures as they seek to modernize their applications and infrastructure for an increasingly competitive, demanding, and cloud-oriented world. They must embrace cloud-native development while managing a profusion of new applications, services, and microservices. They need immense and growing petabyte-scale databases and high-speed, high-throughput database operations to take advantage of increasingly small time-window opportunities. They need to scale infrastructure rapidly, reliably, and in an automated fashion—not by one or two servers, but often by 10, 20, or 30 at a time. They need to manage applications along with underlying database infrastructure together. They need the flexibility to deploy added security on-premise or in the hyperscale or hybrid cloud as their needs dictate—optimizing in-house or cloud-based infrastructure as it grows for the most efficient and cost-effective use of resources.

Red Hat® OpenShift® has emerged as a platform that can deliver these abilities in a robust cloud-native environment for myriad applications—including the real-time applications that use in-memory databases. To this end, Red Hat and Intel are working together to deliver affordable real-time computing at petabyte scale on Red Hat OpenShift. With crucial support from Intel, the platform brings together tested and trusted services to reduce the friction of developing, modernizing, deploying, running, and managing applications. Built on Kubernetes, Red Hat OpenShift delivers a consistent experience across public-cloud, on-premise, hybrid-cloud, or edge architecture.

Intel Integration with Red Hat OpenShift

Diverse organizations are selecting containers and Kubernetes orchestration as their chosen application deployment environments. Containers have revolutionized the ways that applications are developed and deployed. Monolithic applications were expensive, static, and inflexible. Virtualization and private cloud dramatically improved agility and cost models. Containers have further reduced overhead and enabled fully cloud-native development and deployment. Containers implement operating system-based virtualization, allowing multiple containers to share a single operating system (OS) instance on bare metal or inside a virtual machine. Containerized applications become more agile and are less dependent on specific hardware or operating environments.

Red Hat OpenShift is an enterprise Kubernetes platform built on and optimized for Red Hat Enterprise Linux®, a proven foundation for application infrastructure. Red Hat OpenShift provides a single platform that can run on any infrastructure to give you consistency for development and operations across all clouds. As part of Red Hat's open hybrid cloud strategy, Red Hat OpenShift offers developers an agile and flexible application environment to develop, orchestrate, and run their applications while giving sysadmins and operations teams a common operating environment to manage their infrastructure. This consistent experience across environments allows organizations to deliver robust and highly automated IT infrastructure.

The path to CXL

Intel® Optane™ PMem provides a compelling memory tiering solution today as well as a path toward open standards like Compute Express Link (CXL). As an industry-supported cache-coherent interconnect for processors, memory expansion, and accelerators, CXL promises to facilitate breakthrough performance for emerging usage models while supporting an open ecosystem for datacenter accelerators and other high-speed enhancements. As a member of the CXL Board of Directors, Intel is taking an active role in these exciting new developments.

Red Hat OpenShift operators automate the creation, configuration, and management of instances of Kubernetes-native applications. Operators provide automation at every level of the stack, from managing the platform's components to applications offered as a managed service. Intel and Red Hat-certified operators allow the orchestration of their technologies through Red Hat OpenShift.

Intel Optane PMem CSI operator

Intel Optane Persistent Memory (PMem) is an advanced memory technology with performance characteristics that can break through system bottlenecks that constrain memory-intensive workloads. Each module provides 128GB, 256GB, or 512GB of affordable, high-performance, byte-addressable memory in a standard dual in-line memory module (DIMM) form factor, providing significantly more memory per server than is possible with modern 32GB or 64GB DRAM DIMMs. Intel Optane PMem supports 2 distinct modes:

- In **Memory Mode**, Intel Optane PMem operates as volatile memory for the server platform, and a small amount of DRAM acts as a cache for the most frequently used data. In essence, Intel Optane PMem expands the amount of system memory with little to no impact on memory performance.
- **App Direct Mode** takes advantage of the inherent persistence of Intel Optane PMem. The application sees traditional DRAM as fast, volatile memory and Intel Optane PMem as high-capacity persistent memory.

To make Intel Optane Persistent Memory (PMem) available to containerized applications, Intel provides a Container Storage Interface (CSI)-compliant operator known as Intel® Optane™ PMem CSI. The operator is a standardized driver implementation that simplifies the provisioning and management of shared storage by Red Hat OpenShift. The Intel Optane PMem CSI project also offers a Kubernetes operator that provides a streamlined consumption path for developers.

The Intel Optane PMem CSI operator exposes Intel Optane PMem to applications as persistent storage to containerized applications, including support for advanced storage capabilities such as cloning and snapshots. The operator supports the following types of volumes on Intel Optane PMem:

- **Persistent volumes** are created independently from the application. Data in persistent volumes are retained until the volume is deleted.
- **CSI ephemeral volumes** are created each time an application starts to run on a node. Ephemeral volumes are deleted when the application stops.
- **Raw block volumes** appear as raw block devices. Pages mapped on the raw block device go through the Linux page cache.

Memory tiering with Intel technologies

As memory-intensive applications grow, they face a new set of challenges. Even as compute performance accelerates, the evolution of DRAM density is slowing. While previous decades saw exponential growth every few years, the last decade has seen DRAM capacity double only about every four years. As data proliferates, this slower growth in DRAM density creates a widening gap—with the cost of memory as a percentage of overall server costs becoming larger. Complicating these challenges, modern applications and the organizations that depend on them require ever-faster access to data for rapid, actionable insights. Solutions must provide both scalability and increased economic viability.

Memory tiering provides a compelling approach that reserves more expensive DRAM for hot data that requires the most rapid performance. Less costly but more scalable memory technologies can serve as a capacity tier for less performance-intensive tasks that still require large-scale memory capacity. For example, Intel Optane PMem can provide cost-effective capacity of up to 8TB of total system memory per processor socket.

Intel memory tiering reserves expensive DRAM for high-speed uses while vastly expanding addressable memory capacity or providing persistent storage.

Intel memory tiering provides multiple opportunities and several options for where to place the index and data for a real-time database, including:

- Hosting the primary index and data in Intel Optane PMem
- Hosting the Primary index in Intel Optane PMem, and the data on NVMe Flash
- Hosting the primary index in DRAM and data on NVMe Flash
- Hosting the index and data in NVMe Flash

Informed by Intel testing, Table 1 provides a matrix that describes the benefits of different approaches and indications where a given approach is not recommended.

Table 1. Best practices for the primary index and data storage for a real-time database

Primary index storage	Data storage		
	NVMe Flash	**DRAM**	**Intel Optane PMem**
NVMe Flash	All Flash: Ultra large record sets	Not recommended	Not recommended
DRAM	Hybrid: Best performance	Highest performance: no persistence	Not recommended
Intel Optane PMem	Hybrid: Best price-performance, fast restart, very large data sets	Not recommended	Fast restart, high performance, scalability

Memory tiering on Red Hat OpenShift

Intel has performed cost analysis and performance testing to validate the viability of memory tiering in a Red Hat OpenShift environment.

System cost of acquisition

Table 2 illustrates the dramatic system cost reduction available with Intel memory tiering. As shown, reduced memory system cost drives significant system cost reductions. Cost savings can approach 50% even across a small three-node cluster.[1]

Table 2. Cost of acquisition for a 3-node cluster

	CPU cost	Memory subsystem cost	Per-system cost	Three-node cluster cost
DRAM-based system	$9,214	$71,440	$80,654	$241,962
Memory-tiered system	$9,214	$31,853	$41,057	$123,201

Memory tiering performance

Figure 1 shows recent Intel testing[2] utilizing a popular real-time database running on Red Hat OpenShift. The testing compared placing the database index in DRAM against a memory-tiered configuration that placed the database index in Intel Optane PMem 200 Series. In each of these tests, data storage was placed on all-flash NVMe drives to optimize the scalability of the data set. As shown, placing the database index in Intel Optane PMem incurred only a slight 5-6% performance tradeoff against the scalability and economic benefits of memory tiering.

1 *System comparison based on 2-socket servers powered by 32-core Intel Xeon Platinum 8358 CPUs. DRAM-based system configured with 1024GB DDR4 DRAM memory (16x64GB). Memory-tiered system configured with a total of 1152GB consisting of Intel Optane PMem 200 Series (8x12GB) and DRAM (8x16GB) as calculated by Intel internal cost estimation tool.*

2 *Configuration 1 - DRAM BASELINE: Test by Intel as of Mar 13. 3-nodes, 2x Intel(R) Xeon(R) Platinum 8358 CPU @ 2.60GHz, 32 cores, HT On, Turbo On, Total Memory 1024GB (32x64GB DDR4 2933 MT/s [2933 MT/s]); 2048GB, BIOS 1.2, microcode 0xd000363, 4x Ethernet Controller E810-C for QSFP, 2x Ethernet Controller 10-Gigabit X540-AT2, 1x 447.1G INTEL SSDSC2BB48, 6x 7T INTEL SSDPF2KX076TZ, 1x 745.2G INTEL SSDPF21Q800GB, Red Hat Enterprise Linux CoreOS 410.84.202208302227-0 (Ootpa), 4.18.0-305.62.1.el8_4.x86_64, Red Hat OpenShift Container Platform 4.10, Aerospike Enterprise Edition 6.1.0.1, Asbench - Aerospike Benchmark Utility Version 1.5.3, C Client Version 6.0.0, score=3,883,400 (index in DRAM) transactions per second.*

Configuration 2 – index in PMem: Test by Intel as of Mar 08. 3-nodes, 2x Intel(R) Xeon(R) Platinum 8358 CPU @ 2.60GHz, 32 cores, HT On, Turbo On, Total Memory 1024GB (16x64GB DDR4 2933 MT/s [2933 MT/s]); 2048GB (16x128GB Logical non-volatile device 3200 MT/s [2933 MT/s]), BIOS 1.2, microcode 0xd000363, 4x Ethernet Controller E810-C for QSFP, 2x Ethernet Controller 10-Gigabit X540-AT2, 1x 447.1G INTEL SSDSC2BB48, 6x 7T INTEL SSDPF2KX076TZ, 1x 745.2G INTEL SSDPF21Q800GB, Red Hat Enterprise Linux CoreOS 410.84.202208302227-0 (Ootpa), 4.18.0-305.62.1.el8_4.x86_64, Red Hat OpenShift Container Platform 4.10, Aerospike Enterprise Edition 6.1.0.1, Asbench - Aerospike Benchmark Utility Version 1.5.3, C Client Version 6.0.0, score=3,660,254 (index in PMem) transactions per second.

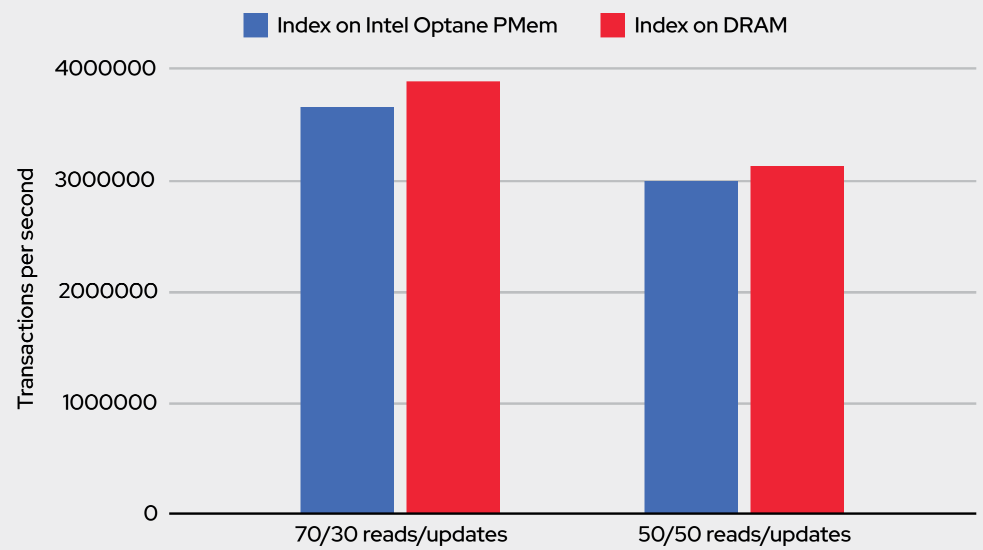

Figure 1. Memory tiering on Red Hat OpenShift

Learn more

As the industry moves towards standards like CXL, memory tiering with Intel® Optane™ PMem will continue to be a viable strategy for the cost-effective scaling of demanding memory-intensive workloads like real-time databases. Employing memory tiering while organizations continue to modernize their applications and manage their infrastructure with Red Hat OpenShift means that organizations can make the necessary capacity and performance tradeoffs they need while harnessing containers and Kubernetes orchestration. For more information on this solution, see relevant vendor documents concerning Red Hat OpenShift, or Intel memory tiering.

facebook.com/redhatinc
@RedHat
linkedin.com/company/red-hat

North America
1 888 REDHAT1
www.redhat.com

Europe, Middle East, and Africa
00800 7334 2835
europe@redhat.com

Asia Pacific
+65 6490 4200
apac@redhat.com

Latin America
+54 11 4329 7300
info-latam@redhat.com