HYVE & MICRON: A Standards-based Storage Approach for Today's Peta-scale Workloads

A growing number of hyperscale businesses are realizing they may need to rebalance their cloud strategies. A recent Andreesen Horowitz article called "The Cost of Cloud, a Trillion Dollar Paradox" focused on Dropbox and how it moved some of its high-volume data from third-party cloud resources to in-house storage infrastructure. This move reflected a shifting, macro-scale paradigm: The world has ever more data, compounding at 26% annually, according to IDC. More data means bigger workloads, which in turn require more compute and memory resources, especially when those workloads need to process in real time. That reality forces enterprises to recalibrate their hybrid cloud strategies.

Concurrently, those massive datasets are often processed by analytics and artificial intelligence (AI) applications. Such software benefits from performance-oriented storage tier architectures running with mass-scale capacity resources. For optimal results and ROI, the speed and robustness of the storage infrastructure should be matched by the storage components. It's one thing to use off-the-shelf hardware based on open standards, but that shouldn't be confused with using consumer-class storage prone to low throughput and premature failure when handling enterprise datasets.

Data trends can shift cost dynamics, and now hyperscale workloads increasingly belong in the private cloud. Let's dig into a use case built around this new reality. Today's hyperscale enterprises face multiple simultaneous pressures: petabyte-scale workloads; increasing demands for real-time performance; and the need to deploy cost-effective, scalable compute and storage infrastructure that will accommodate these ever-increasing, real-time data demands. That's a tall order, but new component advances from Micron Technology and groundbreaking server infrastructure options from Hyve Solutions now make meeting these demands more than possible.

EXAMPLE USE CASES: Analytics and AI

Assume a growing enterprise with expanding data needs and multi-regional services has done the cost analysis and determined that it needs to rebalance its cloud deployment. Like most enterprises, much of its revenue either directly or indirectly hinges on database performance. It likely makes heavy use of frameworks such as Apache Kafka for real-time, many-source data integration and streaming analytics. These frameworks require lightning-fast storage able to minimize any latency between hot data in system memory and secondary data tiers.

Similarly, there's a high likelihood that such an enterprise will either currently employ artificial intelligence (AI) or have plans to do so in the near future. AI model training accuracy directly correlates to the size of the training dataset. Larger datasets yield more accurate, practical models, but they also require significant (potentially petabyte-scale) storage infrastructure and the ability to use that storage efficiently so that training operations can be parallelized. Factors such as high storage bandwidth and intelligent tiering strategies can make all the difference in cost- and time-effective AI training.

In both application types, datasets are too large for in-memory architectures. Some amount of data swapping to storage is necessary. The challenge then revolves around minimizing I/O latencies and implementing tiered caching in a manner that optimizes keeping as much "warm" data as possible within the CPU's immediate reach while simultaneously accommodating budget limitations.

In short, these use cases require performant, capacious storage tailored to hyperscale infrastructures. In turn, these infrastructures will yield the greatest results when based on proven, open systems designed to optimize around space, power, management, and hardware resource constraints.

A partnership between Micron and Hyve four years in the making now offers a unique, forward-looking platform that addresses these challenges and delivers unique value for hyperscale analytics and AI applications.

SOLUTION: Micron Storage & Hyve Modular Servers

Many hyperscale enterprises divide their infrastructure across two system types: high compute and high I/O (HIO). Compute systems emphasize balancing raw performance, power efficiency, and solution density within constrained rack space. HIO systems often target applications such as Kafka, MySQL, and caching tiers.

HYVE HARDWARE

To accommodate all manner of hyperscale deployments, Hyve Solutions developed a novel, adaptable, and broadly compatible server platform based on Open Compute Project (OCP) datacenter systems standards. Hyve specializes in mass-scale computing infrastructure that is purpose-built and customizable to fit customers' exact application needs. With decade-long roots deep in the OCP, Hyve remains a proponent of standards meant to optimize rack-level system design and power distribution. With the OCP's new Open Rack version 3 (ORv3) specification, the industry now has blueprints for how to create and deploy hyperscale infrastructure that focuses on highly efficient, 48V, shelf-format DC power distribution. Hyve adapted ORv3 to a 19" rack width and 960mm depth to better suit the needs and existing investments of next-wave hyperscalers and colocation data centers.

However, Hyve's true innovation within the ORv3 sphere was to create a product portfolio based on 1U Intel- or AMD-based compute sleds. Each 1S, 2S, or 4S node comes equipped with Micron DDR4 memory for exceptional compute performance and base-level SSD and GPU resources. Rather than try to stuff unneeded resources in each compute node, though, Hyve also created its "JBOx" (Just a Bunch of <resource type>) line of modular I/O resources. JBOx devices are 1U sleds that connect to Hyve compute nodes via PCI Express Gen4 or higher. Depending on the JBOx's resource, each sled can contain up to 16 3.5" storage drives (JBOD), up to 16 hot-pluggable E1.S or E1.L flash SSDs (JBOF), or up to four FHFLDW 350W graphics cards or eight FHFLSW 150W graphics cards (JBOG). This disaggregated resource modularity and ability to mix-and-match sleds at will within very dense rack dimensions allows organizations to add exactly the resources needed for their specific applications and workloads. No one need pay for costly overprovisioning because of SKU or vendor design limitations again.

MICRON STORAGE

Within Hyve's ORv3 family, Micron plays a key role in compute memory and especially in HIO storage. The company's 7400 SSD family, with support for PCI Express Gen4, NVMe, and E1.S form factor capacities up to 3.84TB, specifically targets low power consumption while providing maximum performance (up to 6600 MB/s on sequential reads, 3500 MB/s on sequential writes, and 800K/150K IOPS on random reads/writes) and endurance of up to three drive writes per day.

Micron optimized the 7400 for mission-critical workloads, including AI/ML training, OLTP, NoSQL databases, big data, and large object storage, with a balance of performance, security, and reduced costs. The move to PCIe Gen4 (up to 6.6 GB/s) was instrumental, as it allowed Micron to unleash much existing product development on a bus with twice the prior generation's bandwidth. This bandwidth is critical in providing the highest and most efficient I/O performance between CPU and storage.

In many applications, removing the storage bottleneck can eliminate performance impairments that often get mistakenly (and expensively) blamed on system memory. This wider bus, combined with NVMe's protocol efficiency, superior Micron flash media components, and Micron's advanced engineering, delivers exceptionally low latency results that most benefit applications with latency sensitivity. SQL and NoSQL databases, as well as databases like RocksDB and object storage platforms such as Ceph, fit this criterion. High latency also takes a visible toll on virtualized solutions, such as virtual desktop infrastructure (VDI). In general, any application with a high tail latency, wherein the slowest I/O operations determine overall performance, needs the storage characteristics of Micron's 7400 SSDs.

COMBINED BENEFITS

Hyve was one of the earliest and most committed adopters of the E1.S storage standard. The form factor addressed many hyperscale datacenter needs, including hotplug support, thermal efficiency, and highly dense capacity. As shown in Hyve's JBOF design, E1.S enables up to 16 vertically stacked flash modules in a single 1U enclosure.

The 7400 is Micron's first SSD made specifically to support OCP specifications. While also available in U.3 and M.2 form factors, the 7400's E1.S models were an ideal fit for Hyve's modular, ORv3-based platform. The drive's attributes made it an ideal tier to fit between system memory and a petabyte-scale mass storage data lake. Thus, an enterprise might deploy several JBOFs containing 7400 drives and many more JBOFs containing slower but lower-cost SSDs. Micron's use of high-capacity media in the 7400 complement's the company's already-high drive endurance. In fact, 7400 SSDs typically withstand multiple petabytes of writes throughout their lifetime, making them well-suited to the demands and durations typical of Hyve hyperscale deployments.

Combining the ORv3 specification with Micron 7400 SSDs allowed Hyve to create a JBOF with proven reliability and power efficiency and bring this design to market in a convenient, modular form factor ready for asset right-sizing and physical compatibility with existing hyperscale investments. The final product reflects the years of collaboration the two companies have shared in developing, refining, and validating memory and storage configurations.

"Large-scale solutions today are incredibly complex," notes Micron senior product marketing manager Andrew Mierau. "There has to be a fine balance between storage, memory, compute, and networking. They all need to be in symbiosis. If one of them falls out of balance with the others, your application will fail to perform as expected, and you'll probably pay too much for the solution. We worked very closely with Hyve to deliver a balanced product that will suit the needs of our hyperscale customers."

Hyve and Micron's partnership enables a deeper, broader base of support for customers. Hyve buyers are accustomed to fine tuning and optimizing solutions to suit their specific needs; Micron shares in this help by addressing memory and storage needs whenever possible. Going forward, the two companies are already working on future DDR5 implementations, Fabric-over-NVMe support, architecture support for the open Compute Express Link™ (CXL™) interface standard, and other next-generation datacenter technologies that will extend benefits in analytics and AI training applications even further.

CONCLUSION

The modularity and flexibility of Hyve's modified ORv3 portfolio promises to end to the "SKU sprawl" that results from organizations amassing random systems meant for different tasks over time. Hyperscale applications are particularly prone to these piecemeal approaches. However, while piecemeal solutions may be cheaper in the short term, their lack of efficiency and optimization often prohibits them from fulfilling the requirements necessary for repatriating data from the public cloud. Tackling petabyte-scale applications, especially analytics and AI training, requires a finely planned and balanced platform designed around specific workload needs.

Rather than trying to force disparate systems into a semblance of cohesion, Hyve now delivers a building block-like approach to infrastructure, so resources can always be compatible (thanks to open standards compliance) and exactly what's needed (thanks to disaggregated modularity). Hyve's modular server solutions are tailored in every aspect. However, server solutions can only succeed if equipped with performant, dependable, environment-optimized components. That's where Micron DRAM and 7400 SSDs come in.

Micron adds the foundational reliability, speed, and value Hyve compute sleds and JBOFs need to meet hyperscaler demands. Together, this hardware collaboration helps establish a better path for datacenters to be more power-efficient, resource-optimized, and capable of meeting tomorrow's data deluge and all the incredible opportunities it will provide.

Click HERE to learn more about the Micron 7400 SSD and HERE to learn more about the Hyve customized server solutions that incorporate them.

510-668-3877 | sales@hyvesolutions.com | 44201 Nobel Drive, Fremont, CA 94538 | hyvesolutions.com