Look Inside.™

White Paper
Intel®-based Microservers

Flexible, Low Power Microservers for Lightweight Scale-Out Workloads

Large numbers of small compute nodes based on Intel® architecture can address some of the emerging lightweight scale-out workload needs and achieve new levels of power efficiency and density.

Microservers have emerged as an option for specific classes of workloads that tend to scale well with large numbers of relatively lightweight nodes[1] such as low-end dedicated hosting and simple content delivery. Intel supports this topology with a broad range of processor offerings engineered to give flexibility of choice, improve node density, and reduce node power consumption. This paper introduces the design concepts behind microservers and gives cloud service providers and hosters the background to make informed decisions about whether microservers can help them address their server infrastructure needs.

The paper explains the context in which microservers have evolved and the characteristics of system designs. It offers considerations for identifying which workloads are well suited to microserver deployments. The paper also provides an overview of microservers based on the Intel® Xeon® processor E3 family and Intel® Atom™ processor C2000 product family.

1 Overview

As cloud service providers and hosters continue to evolve and grow their service offerings, new applications and software models are emerging rapidly. For example, the web tier has grown to become a large portion of the server infrastructure in many cloud and hosting data centers, handling content for millions of users. Highly parallelized workloads such as analytics and non-relational databases are becoming more widespread. And a host of new content delivery services are driving different ways to store and distribute digital content.

While the majority of emerging scale-out workloads requires high-performance systems to handle dynamic, complex tasks, a subset places lower demands on server infrastructure. These lightweight applications don't take full advantage of mainstream server platforms that are richly configured with memory, disk, and networking options typically found in enterprise data centers today.

Microservers are an emerging category of system design that has been created to address lightweight applications and simultaneously improve data center efficiency. Microservers are characterized by large numbers of server nodes configured to share infrastructure, such as power and cooling fans, in a common chassis.

The trend toward density isn't new. Innovations in server architecture over the past 15 years have helped drive improved efficiency, density, and manageability. Blade servers, for instance, have been widely adopted in the enterprise, and more recent system designs such as half-width rack-mount servers have primarily targeted cloud service provider data center implementations. As shown in Figure 1, microservers are now taking their place among other server form factors as a new density-optimized system.

What is a Microserver?

The microserver represents a new server architecture characterized by many lightweight server nodes bundled together in a shared chassis infrastructure. This topology is designed specifically for density, lower power per node, reduced costs, and increased operational efficiency. By sharing common fans, switching, power supplies, and the metal chassis, microservers streamline rack infrastructure and improve density over standard rack options.

Innovation in Server Form Factors

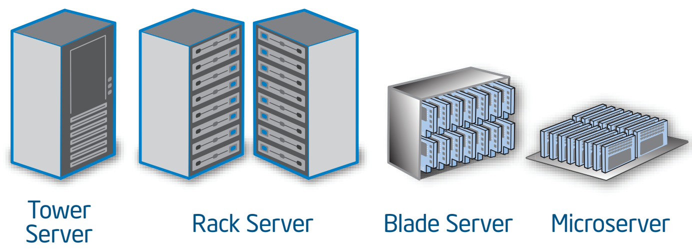

Figure 1. Microservers have emerged as a new density-optimized category that will complement tower, rack, and blade form factors.

2 Microservers for Lightweight Scale-Out Workloads

A key ongoing challenge for computing professionals is to identify where their requirements fit along a spectrum of small-scale to large-scale systems. With the microserver category, a new set of options exists for cloud service providers and hosters as they consider whether a large number of relatively modest nodes is preferable to a smaller number of more powerful nodes for a specific implementation.

Distributed, highly parallel workloads with relatively low compute requirements per node can be scaled out effectively in some cases using microservers. While this class of system is not suited to every workload, it holds particular promise for some market segments. For example, a hosting environment often must support very large numbers of computationally light tasks, such as requests for login authentication or serving small, static HTML pages.

The large number of small nodes in microserver topologies can be ideal for such jobs, as the software is highly parallel and transactions are easily distributed into smaller computational tasks. However, the more computationally intense tasks that make up the majority of jobs in the enterprise are likely to overwhelm small-scale microserver nodes' resources, such as compute power, memory footprint, networking, and storage. Traditional high performance server infrastructures are more effective in those cases.

The diagram in Figure 2 provides key considerations when evaluating the suitability of microservers for various server workloads. In this figure, each of three important server attributes is mapped to a specific axis, and server workloads can be plotted according to each axis. The three axes are:

- **Scales with core count**. Higher values on this axis (as they progress to the left) correspond to highly parallel workloads that increasingly benefit from more cores being added to an individual server, so that each unit of work shares common I/O, system memory, storage, and so on.
- **Scales with physical nodes**. Higher values on this axis (as they progress to the right) also correspond to highly parallel workloads, but they tend to scale better by adding more physical servers because of a tendency toward lower memory, I/O, and storage requirements per server.
- **Scales with brawny cores**. Higher values on this axis (as they progress upward) scale best with robust cores, in areas such as frequency, cache, and memory bandwidth. These workloads tend to be single-threaded or have limited parallelization.

The area in Figure 2 identified as "Typical Microserver Workloads" represents workloads that scale best by adding large numbers of lightweight servers. Examples of workloads in this area are low-end web serving, static content delivery, offline or batch analytics, and low-end dedicated hosting.

Typical Microserver Workloads

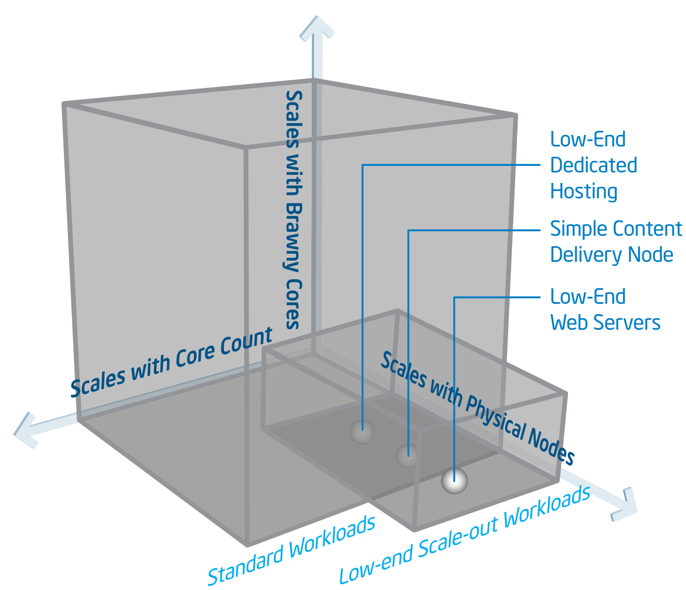

Figure 2. This 3D diagram shows that ideal workloads for microservers are those that tend to have relatively lower processing requirements and scale best by adding more interconnected physical nodes.

3 Matching Capability to the Task: A Range of Intel® Platforms

Even within the microserver category, there is no one-size-fits-all answer to system design or processor choice. Some microservers may have high-performing single-socket processors with robust memory and storage, while others may have a far higher number of miniature dense configurations with lower power and relatively lower compute capacity per node. To meet the full breadth of these requirements, Intel provides a range of processor options from the smaller to the larger end of the spectrum so companies can select what's appropriate for their workloads.

Intel® Xeon® processors are well suited to applications in many segments, such as high-performance computing, financial services, cloud computing, and databases. For microserver workloads, the Intel Xeon processor E3 family offers a choice of node performance, performance per watt, and flexibility. The Intel Atom processor C2000 product family provides extreme low power and higher density compared to Intel Xeon E3 microservers.

Because all these platforms share a common instruction set architecture, application code can run across the full spectrum. That characteristic is very valuable in the sense that workload requirements can change over time. For example, a bottleneck that exists today could change substantially over a period of months, transitioning from being processor-bound to memory-bound to I/O-bound. The flexibility of Intel architecture is extremely valuable in accommodating that variability.

Customers have consistently indicated that they want the same kinds of features and capabilities in microservers that they have come to expect from traditional rack-and-blade server infrastructure. In particular, they need support for 64-bit software, virtualization support, error-correcting code (ECC) memory, a full range of power options, and broad software compatibility. The Intel Xeon processor E3 v3 family ranging from 84W to 13W offers all those advantages for microserver platforms. The Intel Atom processor C2000 series also support these enterprise class server features in a very low power envelope of just 6W TDP.

4 Growing Industry Momentum for Intel®-based Microservers

Industry momentum is growing in the microserver segment among OEMs, with an array of product offerings from industry leaders such as Dell, Hitachi, HP, NEC, Quanta, Supermicro, and Tyan. Broad-based innovation includes microservers based on the full range of platforms, using Intel Xeon and Intel Atom processors offering choice and flexibility.

Standardization facilitates this breadth of innovation in microservers. To that end, Intel worked with other companies, such as Quanta and Tyan, in the Server System Infrastructure (SSI) Forum to create the Micro Module Server Specification. SSI, a trade group based in San Ramon, California, works with industry leaders to drive infrastructure standards. The forum's goal is to enable market growth through the standardization of computing technologies, emphasizing power efficiency and cost reduction for microserver end-customers.

The Micro Module Server Specification is a flexible node building block that is explicitly designed to be easily tailored to a wide range of power and performance design points. The specification defines the module dimensions and pinouts, but allows OEMs significant flexibility to customize system architecture including backplane, interconnect, and storage configurations. It uses broadly available, industry-proven, low-cost connectors and requires module mechanical interoperability while leaving electrical interoperability optional.

Overall, this approach strikes a balance between flexibility and standardization by facilitating product differentiation by OEMs within their product lines and relative to competitors while maximizing design reuse. As such, the specification fosters a number of OEM benefits, including the following:

- Address varied product requirements by innovating within a standards framework in areas such as system form factor, chassis, cooling, switching, and management.
- Maximize ROI by taking advantage of system design investments in multiple products that target different applications and market segments.
- Accelerate time to market by taking advantage of design efficiencies associated with the specification itself, existing design work, and Intel validation and design collateral.

End-customers are the eventual beneficiaries of this design flexibility and efficiency, which benefits them in terms of innovative product availability to meet evolving business needs with low cost and high ROI.

5 Conclusion

Microservers based on Intel architecture from the Intel Xeon processor E3 v3 to the Intel Atom processor C2000 series provide extreme low power, high-density systems for the deployment of compute solutions that are particularly well suited to the needs of lightweight scale-out workloads, such as low-end dedicated hosting and static web serving. Microservers can provide efficiency benefits by reducing system infrastructure such as fans and power supplies.

Microservers may also enable customers to increase density, providing new ways to improve the use of existing rack and power infrastructure. As the industry has started to incorporate microservers into existing and new usage models, they have the potential to improve the TCO associated with supporting this particular class of workloads.

Key Microserver Requirements

- Low power
- 64-bit processors
- Intel® Virtualization Technology support
- ECC memory
- Broad software compatibility

Processor Choice for Optimized Scaling

Matching processor capabilities to the scaling characteristics of specific data-center workloads is vital to optimizing the value of the microserver opportunity.

	Intel® Xeon® Processor E3-1200 v3 Product Family	Intel® Atom™ Processor C2000 Product Family	Intel Atom Processor S1200 Product Family
Core Count (max)	4	8	2
Processor Frequency (max)	3.6 GHz	2.6 GHz	2.0 GHz
Memory Capacity (max)	32 GB	64 GB	8 GB
PCI Express* (PCIe*) Lanes	16 (PCIe 3.0), 8 (PCIe 2.0)	16 (PCIe 2.0)	8 (PCIe 2.0)
Power (min)	13W	6W	6.1W

- **Scale with node performance.** The Intel Xeon processor E3-1200 v3 product family features highest single threaded performance and performance per node of these three options, plus the most robust PCIe connectivity and advanced graphics support.
- **Scale with high node count.** The Intel Atom processor C2000 and Atom S1200 product families provides high node density. Intel Atom processor C2000 offers larger memory and PCIe capacity to share among highly parallel workloads along with higher node performance and performance per watt as compared to Intel Atom Processor S1200.

To learn more microservers based on Intel® architecture, visit www.intel.com/microservers.

Visit us at:

[1] Node is a collection of atleast an SoC, its local system memory and ideally all required IO components based on required implementation.

[3] Yeraswork, Zewde, "SSI Forum Releases Micro Module Server Specification," CRN (2011).
http://www.crn.com/news/components-peripherals/229000520/ssi-forum-releases-micro-module-server-specification.htm;jsessionid=18NGOEkb0OTI5-cBnYSN7g**.ecappj01

0813/KH/MESH/PDF 325345-002US