Optimizing Infrastructure for Workloads in OpenStack*-Based Public Cloud Services

How Intel® Architecture-based Servers help Cloud Service Providers match systems to SLAs without the disruption caused by deploying multiple silicon architectures

Executive Summary

As an OpenStack* software user, contributor and service provider, Intel investigates and shares best practices with numerous organizations, helping them deliver Infrastructure as a Service (IaaS) in a safe, reliable, and affordable manner. In those engagements we've learned that optimizing infrastructure for various customer needs requires a wider range of systems than the single, high-end, virtualized servers typically deployed. One size does not fit all. This is particularly true when it comes to the differing business requirements and Service Level Agreements (SLAs) of Cloud Service Providers (CSPs) running open source OpenStack software.

This paper examines how CSP business needs translate to infrastructure considerations for IaaS when building out or enhancing an OpenStack cloud environment. The paper examines these requirements and the foundational platform technologies that can support such a wide range of SLA requirements.

CSP Conundrums

Right-sizing the Infrastructure

The server needs of Cloud Service Providers (CSPs) vary with the requirements of their Service Level Agreements (SLAs). Much of a growing CSP business revolves around enabling Infrastructure as a Service (IaaS) for lightweight, scale-out workloads (e.g., web sites) with SLAs that differ greatly from the demands of more stringent, enterprise-class workloads. Microservers have enabled growing deployments of these smaller IaaS offerings at savings of space and power—both big concerns in CSP operations.

But as CSPs compete more aggressively for enterprise-class workloads of private, traditionally on-premise cloud deployments, CSPs will need to deploy servers that offer greater capabilities. These systems must provide more compute and I/O performance, support more virtual machines (VMs), and be more robust, secure, and manageable. Such requirements are clearly out of the scope of today's microservers. And they impact the CSP with greater equipment cost, more power and cooling demand, and larger space requirements. The choice of processor architecture can also impact the code-base for which CSPs already have built their applications, infrastructure, and operations.

Optimizing Costs

Balanced against server needs is a CSP's ongoing battle to remain cost-competitive in the market by minimizing operational expenses while delivering on SLAs. Operationally, power related costs (power, power distribution, and cooling) comprise about 31 percent of data center costs.[1] Thus, CSPs are motivated to continue to keep these costs down. Power/performance, along with price/performance, of deployed systems is important when determining what will be deployed to meet the customer's need.

Cost is another reason so many CSPs build their infrastructure on OpenStack software, the leading open source cloud operating environment. It is widely deployed and used across industry-standard x86-based machines for virtualized environments. New projects for future OpenStack releases will enable deployments of OpenStack on bare-metal—opening whole new types of deployments for CSPs that potentially will increase efficiency of operations, while offering new levels of IaaS on lower cost, more energy efficient microservers.

Intel is committed to the OpenStack project and community. Intel IT uses OpenStack in private cloud deployments and contributes code upstream, enabling cloud enhancements on Intel® technologies.

With OpenStack as the core of their operating environment, a key challenge for CSPs remains. CSPs will need to identify where their server requirements fit along the spectrum of microservers to multi-socket, high-end systems that will meet various SLAs, take advantage of optimized code, and keep operating costs under control.

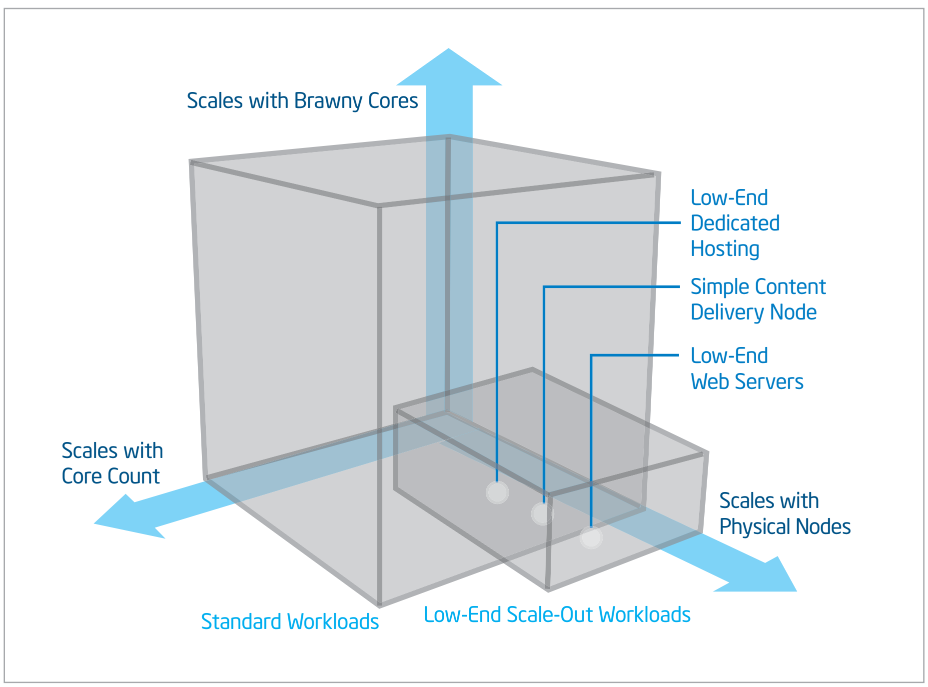

Figure 1. Typical Microserver Workloads.

Server Solutions in an OpenStack Environment

Microservers Enable Hyperscale Workloads

Microservers are relatively new to the data center. But they have had an enormous impact on CSPs. Microservers in high-density systems fit well for serving many of the lightweight, highly scaled-out (or "hyperscale") workloads running in CSP data centers.

Distributed, highly parallel workloads with relatively low compute requirements per node tend to scale better by adding more physical servers, because of their lower memory, I/O, and storage requirements per server. A CSP customer's hosting environment, where there are fully independent tasks, or a need to execute very large numbers of independent and computationally light tasks, such as serving up many static web pages, could be easily scaled out. Other examples include:

- Offline or batch analytics that can scale out to many systems without critical latency demands.
- Low-end dedicated hosting for low cost services.
- OpenStack Swift object storage services.

A large number of microserver nodes can be ideal for such jobs, as the software is highly parallel and large numbers of fully independent transactions are easily distributed into smaller computational tasks.

Because of a microserver's small size and energy-efficient design, its infrastructure (including the fan and power supply) can be shared by tens or even hundreds of physical server nodes, eliminating the space and power consumption demands of duplicate infrastructure components. The Intel reference system board area footprints and peak active power of an Intel® Atom™ processor C-2750-based microserver node are less than a tenth that of a contemporary Intel® Xeon® processor E5-265x-class node. Thus, microservers are very attractive to CSPs for their energy cost and space savings.

Traditional Virtualized Computing is Still Needed

Where workloads require greater compute capabilities, support for more virtual machines, and more robust security and management, CSPs will look beyond microservers. They will need the capabilities of the traditional multi-socket servers that enterprise IT data centers have employed to meet the SLAs of more mission-critical applications.

Based on Intel research, an Intel® Xeon® processor E5-2660 v2-based node could support ten virtual machine instances at four virtual CPUs each, or four virtual machine instances at ten virtual CPUs each. While an Intel Atom processor C-2750-based node could only support two of the four virtual CPU instances, but it could not support a ten virtual CPU instance at all. It is also fair to note that the relative performance of a four virtual CPU virtual machine instance would be considerably higher on the Intel Xeon processor E5-2660 v2-based node, particularly if the virtual machine workload utilized the superior floating point or Intel® Advanced Vector Extensions (Intel® AVX) hardware of that node class.

Such systems also support trusted compute pools, using Intel® Trusted Execution technology (Intel® TXT) integrated in Intel® Xeon® processors, in conjunction with Open Attestation. Trusted compute pools on Intel Xeon processor-based platforms, verified using Open Attestation, help ensure secure trusted launches of VMs and safe migration of workloads to them.

Optimizing the Enterprise Data Center Infrastructure with Intel Servers

Enterprise data center workloads are typically reversed from the needs of today's CSP customers. In the enterprise, the more computationally intense tasks make up the majority of jobs. They are likely to overwhelm small-scale microserver nodes' resources, such as compute power, memory footprint, networking, and storage. Intel® Xeon® processor E5 v2 family-based high-performance servers are more effective in these cases, and thus, we see these higher end servers deployed in the enterprise.

But, as data center workloads expand and become more varied, enterprise IT can benefit from microservers to support more lightweight, scale-out workloads. Here, the continuum of Intel server processor-based platforms offers the same benefits and capabilities, but essentially in reverse. Data centers traditionally deploying larger systems can add Intel® Xeon® processor E3 v3 family- and Intel Atom processor C2000 family-based microservers to meet their corporate SLAs driven by lighter workloads, without sacrificing the code benefits offered on Intel® architecture.

Intel® Server Processor Continuum

Customers have consistently indicated they need the same kinds of features and capabilities in microservers that they have come to expect from traditional rack-and-blade servers. In particular, they need support for 64-bit software, virtualization support, error-correcting code (ECC) memory, a full range of power options, encryption acceleration, and broad software compatibility. These capabilities are available to CSPs in the full range of Intel processor-based servers.

To meet the full breadth of CSP SLA requirements from lightweight, scale-out to enterprise-class workloads, Intel provides a range of server processor options, so CSPs can select what is appropriate to meet their customers' needs.

- The Intel® Atom™ processor C2000 product family provides extreme low power and higher density compared to Intel Xeon processor E3 family-based microservers.
- The Intel® Xeon® processor E3 v3 family offers a choice of node performance, performance per watt, and flexibility to meet more demanding SLAs.
- The Intel® Xeon® processor E5 v2 family delivers a high level of compute and I/O performance, with hardware-enhanced technologies to meet stringent enterprise-class workloads and demanding SLAs.
- The Intel® Xeon® processor E7 family offers the highest computing capabilities for the most demanding SLAs and mission-critical applications.
- Intel's many source code contributions to the OpenStack project help enable Intel hardware-enhanced technologies in the OpenStack environment.

A Common Foundation of Features

Intel server processors deliver the capabilities to meet technology requirements for a wide range of CSP SLAs. Furthermore, because all these Intel-based platforms share a common x86 instruction set architecture (ISA), optimized application code can run across the full range.

In addition to a common ISA, the Intel Atom processor C2000 family and Intel Xeon processors support the following, helping maintain low TCO and high flexibility of solutions within the CSP data center:

- In-common power management capabilities (e.g., thermal-trip).
- In-common RAS capabilities (e.g., machine-check architecture).
- In-common security features (e.g., encryption acceleration with AES-NI).
- In-common virtualization features (e.g., Intel® Virtualization Technology (VTx)).

Microserver Processors

Intel® Atom™ processor C2000 family—extreme low power for high-density microservers

The Intel Atom processor C2000 product family allows CSPs to maximize rack space utilization and reduce energy costs when running certain lightweight scale-out workloads.[2] The Intel Atom processor C2000 family is Intel's second generation 8-core, 64-bit Intel® Atom SoC. It provides up to 7x higher performance[2,3,4] and 8x more memory capacity, plus up to 6x[2,5,6] better power efficiency, enabling greater density, compared to the Intel® Atom™ processor S1200 family.

Intel® Xeon® processor E3 1200 v3 family—optimized performance per watt for microservers[2,7]

With up to 52 percent[2,8] better performance per watt and up to 24 percent[2,9] lower power than the previous generation, the Intel® Xeon® processor E3-1200 v3 family for microservers lets CSPs meet more demanding SLAs, while keeping power consumption low. At just 13W, the Intel Xeon processor E3-1220L v3 is ideal for microservers in the CSP data center. In addition to the data center-class capabilities listed above, this family of processors includes:

- Up to 38 percent[2,10] increase in graphics performance, hardware-accelerated media encode and decode for server graphics workloads.
- Improved I/O performance with additional USB3.0 and SATA 6G ports.
- Integrated security delivers built-in encryption acceleration and better protects against malware and denial of service (DoS) attacks.[11]

Right-sizing microservers in the CSP data center

When it comes to Intel processor-based microservers, CSPs have several choices. Which is most applicable depends on many factors, including the usage model, the data center power envelope, and the system type. For example, Intel Xeon processor E3 v3 family-based platforms may be better suited for web-scale workloads, where a highly-dense infrastructure that doesn't compromise on performance is needed. These platforms may also better suit usages that require professional graphics performance, such as media, online gaming, and desktop virtualization. Alternatively, Intel Atom processor C2000 product family-based platforms could be well suited for lightweight, scale-out workloads that require extreme high density and energy efficiency, such as entry-level, dedicated hosting, low-end front-end web hosting, offline or batch analytics, and OpenStack Swift object storage.

Intel® Xeon® Processor E5 v2 Family—high-performance capabilities for stringent SLAs

As CSPs support more enterprise-class workloads, they face the same concerns and issues in-house IT departments are looking at with respect to the cloud. Some include the following:

- Supporting security technologies that protect data/server infrastructure and allow a trusted computing environment.
- Improved virtualization performance, including live migration.
- Improved I/O performance.

The Intel Xeon processors E5 v2 family was designed for large workloads and to address these and other concerns in data centers. Along with higher levels of processor scalability and memory capacity, Intel Xeon processors offer more features designed to provide extreme reliability, availability, and serviceability (RAS).

Intel® Xeon® Processor E7—ultimate computing for the most demanding SLAs

Intel® Xeon® processor E7 family expands on the reliability features IT managers trust in Intel Xeon processors and expect in mission-critical solutions, such as Machine Check Architecture Recovery (MCA-R), with more advanced enhancements.

- Double Data Device Correction (DDDC)—extends reliability by recovering from two DRAM device failures, instead of a single failure of SDDC, helping maximize uptime.
- Partial Memory Mirroring—enables more flexible, effective, and cost-efficient memory mirroring of critical areas instead of all memory, reducing server energy demands while protecting your most important data.

Beyond Virtualization with Bare-Metal Cloud

Virtualization is often assumed to be synonymous with the cloud. But, virtualization is merely a method to achieve the objectives of the cloud. A CSP does not have to virtualize all servers. And, with the resources that virtualization demands from each server, virtualization in lower-end servers might not help meet particular SLAs.

With the introduction of high-density microservers, the performance of a single virtual machine in a virtualized high-end server can be available in a single physical microserver—without the performance overhead of a hypervisor, or the resulting dependence on orchestration through VMM-supported interfaces. Microservers introduce a cost- and power-effective vehicle for CSPs to offer dedicated independent server node instances down to the granularity of 2 or 4 Intel Atom processor-class cores. In a homogeneous Intel Xeon processor E5-class server deployment, the minimum cost and power for a dedicated hardware instance is a full Intel Xeon processor-based system footprint.

By deploying OpenStack directly on bare-metal microservers, CSPs have not lost their ability to quickly deploy new services using the OpenStack cloud environment, because that capability lies with OpenStack, not the hardware. If the server is available, OpenStack can deploy to it.

What CSPs have gained is greater control over right-sizing their infrastructure by not having to deploy high-end systems that are under-utilized, even with multiple virtual machines. They also can offer IaaS on systems physically independent from one another, rather than on being isolated using silicon and software technologies. For customers with sensitivities about hosting on non-physically isolated systems, this immediately removes a barrier to a new market.

Intel® Atom™ processor C2000 family- and Intel® Xeon® processor E3 v3 family-based microservers are ideal for bare-metal cloud deployments. They offer security features, such as encryption acceleration, needed to maintain high performance with encrypted data and other enterprise-class capabilities many customers seek.

Conclusion

CSPs have a wide range of systems they can draw on to support their expanding workloads and SLAs. From different capabilities of microservers to traditional multi-socket data center systems, they have more choices to populate their OpenStack-based infrastructure. But choosing a variety of processor architectures can mean loss of some of the benefits and optimizations designed into the software. That loss could result in an inability to meet particular SLAs, as well as increased operating costs.

By taking advantage of the range of Intel processor-based servers, CSPs can right-size their infrastructures without sacrificing capabilities or optimizations. With the Intel Atom processor C2000 family, with as little as 6W TDP, and the Intel Xeon processor E3 v3 family, ranging from 13 to 84W TDP, CSPs can benefit from low power and a range of performance options, plus enterprise-class features that some workloads and customers demand. For more demanding requirements, Intel Xeon processor E5 v2 family-based servers enable support for greater virtualization and more intensive workloads, while Intel Xeon E7 processor-based servers offer the ultimate computing platform. These capabilities come with the added benefits of consistencies from platform to platform, allowing CSPs to continue to take advantage of the code they work hard to deploy.

Intel's range of server processors and Intel's commitment to OpenStack create a strong foundation for Cloud Service Providers today and into the future.

FOR MORE INFORMATION VISIT

Intel Developer Zone/OpenStack

Microservers Powered by Intel

Open Source at Intel

1 James Hamilton http://mvdirona.com/jrh/talksAndPapers/JamesHamilton_IntelDCSGg.pdf

2 No computer system can provide absolute security. Requires an enabled Intel® processor, enabled chipset, firmware, software and may require a subscription with a capable service provider (may not be available in all countries). Intel assumes no liability for lost or stolen data and/or systems or any other damages resulting thereof. Consult your Service Provider for availability and functionality. For more information, visit http://www.intel.com/go/anti-theft. Consult your system manufacturer and/or software vendor for more information.

3 Performance based on Dynamic Web Benchmark Performance: Atom S1260 (8G B, SSD,1 GbE), Score=1522. Atom C2750 (32 GB, SSD,10 GbE), Score=11351.

4 Based on Dynamic Web Benchmark performance. 'Previous Generation' configuration – Intel® Atom™ processor S1260 (8GB, SSD, 1GbE), Score=1522. 'New Generation' configuration – Intel Atom processor C2750 (32GB, SSD, 10GbE), Score=11351. Source: Intel internal measurements as of August 2013.

5 Results have been estimated based on internal Intel analysis and are provided for informational purposes only. Any difference in system hardware or software design or configuration may affect actual performance.

6 Performance per Watt based on Dynamic Web Benchmark: Atom S1260 (8 GB, SSD,1 GbE), Score=1522, est node power=20W, PPW=76.1 Atom C2730 (32 GB, SSD,10 GbE), Score=8778, est node power=19W, PPW=462. Source: Intel Internal measurements as of August 2013.

7 Intel does not control or audit the design or implementation of third party benchmark data or Web sites referenced in this document. Intel encourages all of its customers to visit the referenced Web sites or others where similar performance benchmark data are reported and confirm whether the referenced benchmark data are accurate and reflect performance of systems available for purchase.

8 Baseline configuration: Fujitsu* TX140 S1p with one Intel® Xeon® processor E3-1265L v2 (8M Cache, 2.50 GHz), 16GB (2x8GB 2Rx8 PC3-10600E-11, ECC), 1 x 500 GB SATA, 7200 RPM, Intel® Hyper-Threading Technology-enabled, Intel® Turbo Boost Technology-enabled, Red Hat* Enterprise Linux Server Release 6.2, Kernel 2.6.32-220.el6.x86_64, compiler version 12.1.0.293 of Intel® C++ Compiler XE. Represents the best published results as of April 2013. Score: SPECint*_rate_base2006=169. http://www.spec.org/cpu2006/results/res2012q2/cpu2006-20120522-22364.html. Processor TDP = 45W, Perf/W=3.76 New configuration: Intel® C226 chipset-based Intel® Xeon® server platform with one Intel® Xeon® processor E3-1230L v3 (8M Cache, 1.8 GHz), 16GB (2x8GB dual-rank DDR3-1600 ECC UDIMM), 250 GB SATA 6Gb/s HDD, Intel® Hyper-Threading Technology-enabled, Intel® Turbo Boost Technology-enabled, Red Hat* Enterprise Linux Server 6.3 for x86_64, compiler version 13.0.0.133 of Intel® C++ Studio XE, and Intel® Fortran. Source: Intel internal estimated measurements, April 2013. Score: SPECint*_rate_base2006=143. Processor TDP = 25W, Perf/W=5.72.

9 Power reduction from Intel® Xeon® processor E3-1220L v2 with 17W TDP to Intel® Xeon® processor E3-1220L v3 with 13W TDP.

10 Baseline configuration: Intel® C206 chipset-based Intel® Xeon® workstation platform with one Intel Xeon processor E3-1275 (quad-core, 3.4GHz, 8M cache), ASNBCPT1.86C.0085.P00 July 5, 2012, Intel® Hyper Threading Technology (Intel® HT Technology) best configuration, 8GB memory (2x4GB DDR3-1333 ECC UDIMM), Intel® HD Graphics P3000 with driver 2455, 2TB 7200RPM SATAIII HDD (WD2000FYYZ), Microsoft Windows 7* Service Pack 1. Source: Intel internal testing as of April 2013, SPECviewperf* 11 , geomean of 7 workloads (ensight-04, lightwave-01, maya-03, proe-05, sw-02, tcvis-02, snx-01) Previous configuration: Intel® C226 chipset-based Intel Xeon workstation platform with one Intel Xeon processor E3-1275 v2 (quad-core, 3.5GHz, 8M cache), ACRVMBY1.86C.0096.P00 September 9, 2012, Intel HT Technology best configuration, 8GB memory (2x4GB DDR3-1600 ECC UDIMM), Intel HD Graphics P4000 with driver 2712, 2TB 7200RPM SATAIII HDD (WD2000FYYZ), Microsoft Windows 7 Service Pack 1. Source: Intel internal testing as of April 2013, SPECviewperf 11 , geomean of 7 workloads (ensight-04, lightwave-01, maya-03, proe-05, sw-02, tcvis-02, snx-01) New configuration: Intel C226 chipset-based Intel Xeon workstation platform with one Intel Xeon processor E3-1275 v3 (quad-core, 3.5GHz, 8M cache), HSWLPTU1.86C.0116.R00 March 3, 2013, Intel HT Technology best configuration, 8GB memory (2x4GB DDR3-1600 ECC UDIMM), Intel HD Graphics P4600/4700 with driver 2989, 2TB 7200RPM SATAIII HDD (WD2000FYYZ), Microsoft Windows 7 Service Pack 1. Source: Intel internal testing as of April 2013., SPECviewperf 11 , geomean of 7 workloads (ensight-04, lightwave-01, maya-03, proe-05, sw-02, tcvis-02, snx-01).

11 No computer system can provide absolute security. Requires an enabled Intel® processor and software optimized for use of the technology. Consult your system manufacturer and/or software vendor for more information.