New Advances by Intel, Amazon Web Services, Drive Major Cloud Savings

One advantage of cloud is the ability to capitalize on the massive buying power of major cloud providers to run workloads on the latest, most advanced server hardware. When running workloads on Amazon Web Services (AWS), for example, organizations should expect state-of-the-art AWS servers to optimize costs and performance, as compared to running the same workloads on older hardware on-premise.

Given constant improvements in server performance and efficiency, newer server hardware should be able to support workloads that used to require larger, more expensive processors, while delivering the same or better throughput and performance. Most organizations, however, have never actually quantified that price/performance comparison. Now, TSO Logic has done just that.

Analyzing anonymized data from across our North American customer base, TSO Logic has found that organizations can **reduce per-instance infrastructure costs by as much as 57%[1]** just by moving to smaller, newer instance sizes. When factoring software license savings tied to core count reduction, the **savings can reach tens, even hundreds of thousands of dollars annually**.

Quantifying Server Price/Performance

In recent months, AWS has introduced three new instance sizes, C5, M5 and T3, all based on the latest generation of Intel® Xeon Scalable processors. Each represents a newer version of an existing AWS instance (C4, M4 and T2, respectively), built on an innovative mesh architecture that delivers **consistent performance, high bandwidth, low latency and the ability to scale globally** from two sockets to eight sockets and beyond. With Intel Xeon Scalable processors, new instances provide companies **up to 2x faster reliable performance[2]** for analytics applications and database processing.

C Instances: C instances are compute-optimized and offer the lowest price per vCPU in the Amazon EC2 family, making them ideal for running advanced compute-intensive workloads. This includes workloads such as high-performance web servers, high-performance computing (HPC), batch processing, ad serving, highly scalable multiplayer gaming, video encoding, scientific modelling, distributed analytics and machine/deep learning inference.

M Instances: M instances provide a balance of compute, memory and network resources, making them a good choice for many applications.

T Instances: T instances are low-cost, general purpose instance types. They provide a baseline level of CPU performance with the ability to burst above the baseline with On-Demand instance prices.

Methodology

Just how much does next-generation Intel hardware affect price/performance for AWS cloud instances? To find out, TSO Logic analyzed millions of data points from across a subset of our 100,000-instance repository of anonymized AWS customer data. By examining enterprise compute, provisioning levels, and usage patterns, TSO Logic identified workloads currently running on older/larger AWS instance sizes that could be supported by newer and smaller instance sizes. Using up-to-date pricing information from the AWS pricing catalog, they then calculated the resulting savings in real dollars.

Driving Down the Costs of Cloud Infrastructure

Figure 1 details some examples of the savings that can be realized by moving to later-generation instance types on AWS. As the figure shows, customers can **cut per-instance costs up to 57%**[1], without sacrificing throughput or performance, just by taking advantage of the latest Intel and AWS innovations.

Figure 1

Instance Change	Cost Savings per Instance	Percentage Saved per Instance
T2.XLarge to T3.Large	$454.00	47%
T2.Large to T3.Medium	$248.14	44%
T2.Large to T3.Small	$317.96	57%
C4.8XLarge to C5.4XLarge	$3,270.33	50%

So what does that mean in real dollars? For the customer in row one of the table, the workloads analyzed encompassed 374 instances of T2.XLarge. By moving those instances to T3.Large (again, without losing any of the required throughput or performance), **the customer can save up to $169,796 annually** in AWS cloud instance costs. The customer in the second row moving from T2.Large to T3.Medium (in this case, 130 instances), **will save up to $32,258 annually**.

The examples highlighted in this paper are not rare cases. Analyzing the larger anonymized TSO Logic customer base, for the typical customer, **19% of current instances**[1] could benefit from newer, smaller AWS instance types. As the numbers show, organizations of all types could benefit from this type of processor upgrade analysis—to the tune of tens, potentially hundreds of thousands of dollars in annual cloud infrastructure savings.

Reducing Software Costs

Clearly, running workloads on smaller, newer, more powerful instances can significantly reduce an organization's cloud infrastructure costs. However, moving to newer hardware can drive even greater savings when factoring in core count reductions for applications licensed on a per-core basis.

In the fourth customer scenario detailed above, for example, moving from C4.8XLarge instances to C5.4XLarge reduces the core count for that workload by 40. For each such C4 instance the customer migrates to a C5, they will realize up to $3,270 infrastructure savings. But, if the customer is running Microsoft SQL Server on those instances, which is licensed at a cost of $1,800 per core annually, the core count reduction lowers their licensing costs by $72,000.

Intel Architecture Advantages

Clearly, shifting workloads to AWS servers with newer-generation Intel processors can have a significant impact on the bottom line as shown in Figure 1. However, ongoing Intel innovations are delivering additional benefits well beyond the ones quantified in this paper, such as Independent Software Vendor (ISV) certifications (SAP, Microsoft and others) and optimizations.

The family of new Xeon scalable processors, such as those at the heart of the new EC2 C5 instance, feature a new micro-architecture, new hardware acceleration capabilities and support for Intel® AVX-512 instruction set. Together, these capabilities allow the latest-generation EC2 instances to run more (and more complex) jobs in a fraction of the time—especially for compute-intensive workloads that characterize HPC, machine learning/ artificial intelligence, life sciences, financial services and other demanding applications.

New AWS EC2 z1d instances running on Intel Xeon Scalable processors with Intel® Turbo Boost deliver a sustained all-core frequency up to 4.0 GHz—the fastest of any cloud instance. The combination of high compute performance and high memory that these instances provide make them ideal for electronic design automation (EDA), gaming and certain relational database workloads with high per-core licensing costs.

While TSO Logic did not quantify the financial impact of these innovations, it is likely significant—especially for EDA and relational database workloads with high per-core software licensing costs. Semiconductor firms using zd1 instances, for example, can run more EDA jobs per core, amortizing their annual license subscription over more jobs and reducing their design and verification time.

---

[1] Savings figures and percentages are derived from analysis conducted on TSO Logic's 100,000-instance repository of anonymized data. For details, see the section "About the Data" in this paper.

[2] Up to 4.2x more VMs based on server virtualization consolidation workload: Based on Intel internal estimates 1-Node, 2x Intel Xeon Processor E5-2690 on Romley-EP with 256 GB Total Memory on VMware ESXi* 6.0 GA using Guest OS RHEL 6.4, glassfish3.1.2.2, postgresql9.2. Data Source: Request Number: 1718, Benchmark: server utilization consolidation, Score: 377.6 @ 21 VMs vs 1-Noble, 2 x Intel Xeon Platinum 8180 Processor on Wolf Pass SKX with 768 GB of Total Memory on VMware ESXi6.0 U3 GA using Guest OS RHELL 6 64bit. Data Source: Request Number: 2563. Benchmark: server virtualization consolidation. Score: 1580 @ 90 VMs. Higher is better.

About the Data

To conduct this analysis, the TSO Logic platform algorithmically analyzed a subset of our 100,000-instance repository of on-premise OS instances across companies evaluating cloud. The platform created a fine-grained statistical model of each organization's compute resources to determine the most cost-effective place to run each workload. Ingesting millions of data points from the current environments—including age, generation and configuration of all hardware, the OSs they're running, and each instance's utilization—it algorithmically profiled compute patterns. It then used machine learning influenced algorithms and pattern matching to determine the best fit for each workload from thousands of potential cloud options. Using up-to-date, validated information from Intel and AWS, the platform normalized and compared processing capabilities and costs across AWS instance types using various generations of Intel Xeon processors.

Cost reduction scenarios described are intended as examples of how a given Intel- based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.

About Intel

Intel (NASDAQ: INTC), a leader in the semiconductor industry, is shaping the data-centric future with computing and communications technology that is the foundation of the world's innovations. The company's engineering expertise is helping address the world's greatest challenges as well as helping secure, power and connect billions of devices and the infrastructure of the smart, connected world – from the cloud to the network to the edge and everything in between.

To learn more, contact us at info@intel.com

About TSO Logic

TSO Logic helps organizations automate discovery and TCO analysis of their compute environments to fuel successful cloud migration and enterprise transformation. We create fine-grained statistical models of all on-premise resources to help you automatically identify the most cost-effective place to run each workload. Using validated information from Intel and AWS, we normalize and compare processing capabilities between various generations of Intel processors and the myriad of options in the AWS cloud.

With the automated, algorithmic analysis of the TSO Logic platform, you can keep up with constantly changing cloud catalogs and ensure that you are always running each workload in the optimal place, at the lowest cost.

To learn more, contact us at info@tsologic.com