Achieve more OLTP database work[*]

38% more MySQL transactions per minute using the HammerDB TPROC-C benchmark

Execute machine learning algorithms in less time[†]

Up to 51% less time to carry out HiBench Spark workloads

Lower monthly cost to run a given workload

10% less for MySQL VM[Δ]

21% less for cluster of HiBench Spark machine learning VMs[◊]

Get greater performance on MySQL™ and Spark™ machine learning workloads by selecting Azure® Standard_HB120-64rs_v3 virtual machines based on 3rd Gen AMD EPYC™ 7V13 processors

If your organization is one of the many that are shifting critical applications to the cloud, you know that cloud service providers offer a staggering number of virtual machine options. In your quest for the best performance, an important factor to consider is the processor that powers the VMs.

We ran benchmarks on two high-performance 64-vCPU Microsoft Azure VMs:

- Standard_HB120-64rs_v3 VMs, based on AMD EPYC™ 7V13 processors
- Standard_E64ds_v4, based on 2nd Gen Intel® Xeon® Platinum 8272CL processors

We measured MySQL online transaction processing using the HammerDB TPROC-C OLTP workload on a single VM of each type. We also measured Apache Spark™ performance using two HiBench workloads on a cluster of five VMs of each type. We found that on all three workloads, the Standard_HB120-64rs_v3 VMs based on the AMD EPYC 7V13 processor outperformed the VMs based on the 2nd Gen Intel Xeon Platinum 8272CL processor, with improvements of up to 51 percent on HiBench. By selecting higher-performing VMs, you can do more work with the same number of VMs, which can help contain your cloud footprint.

In addition to their performance advantages, Pay-as-You-Go pricing for the Azure Standard_HB120-64rs_v3 VMs based on the AMD EPYC 7V13 processor is lower than that for VMs based on the Intel Xeon Platinum 8272CL processor.[1] We found that selecting Azure Standard_HB120-64rs_v3 VMs based on an AMD EPYC 7V13 processor for the setups we tested could reduce your monthly expenditure by as much as 21 percent.

One Azure Standard_HB120-64rs_v3 VM, based on AMD EPYC 7V13 processors, vs. one Azure Standard_E64ds_v4 VM, based on 2nd Gen Intel Xeon Platinum 8272CL processors, on HammerDB TPROC-C benchmark.

†Cluster of five Azure Standard_HB120-64rs_v3 VMs, based on AMD EPYC 7V13 processors, vs. cluster of five Azure Standard_E64ds_v4 VMs, based on 2nd Gen Intel Xeon Platinum 8272CL processors, on HiBench Logistic Regression and Latent Dirichlet Allocation workloads.

ΔOne Azure Standard_HB120-64rs_v3 VM, based on AMD EPYC 7V13 processors, vs. one Azure Standard_E64ds_v4 VM, based on 2nd Gen Intel Xeon Platinum 8272CL processors, running 24 hours a day for one month in East US region.

◊Cluster of five Azure Standard_HB120-64rs_v3 VMs, based on AMD EPYC 7V13 processors, vs. cluster of five Azure Standard_E64ds_v4 VM, based on 2nd Gen Intel Xeon Platinum 8272CL processors, running 24 hours a day for one month in South Central US region.

Comparing two similar Azure virtual machines

Table 1 presents some of the configuration details of the two VMs we tested. (For more complete configuration information, see the science behind the report.) Our goal was to compare virtual machines that were as similar as possible aside from the processor, but because Azure offers only prepackaged VMs, the configurations differ slightly. As the table shows, the Standard_HB120-64rs_v3 VM backed by the Intel Xeon Platinum 8272CL processor has more memory than the VM backed by the AMD EPYC 7V13 processor and the core frequency of its processor is 0.5 GHz lower. The Intel VM we tested provided the closest available comparison on CPU, vCPU count, and memory capacity. The majority of Azure VMs powered by 2nd Generation Intel Scalable processors use the Intel Xeon Platinum 8272CL processor.

Note that for the MySQL database testing, we used a single VM of each type. For the Spark testing, we used five VMs (one manager node and four worker nodes) of each type.

Table 1: VM configuration information. Note: One of the VMs we used in testing—the HiBench manager node—did not have an additional (data drive) disk. Source: Principled Technologies.

	Standard_HB120-64rs_v3	**Standard_E64ds_v4**
Platform details		
VM series and size	Standard_HB120-64rs_v3	Standard_E64ds_v4
Operating system name and version/build number	CentOS Linux release 8.4.2105 4.18.0-305.7.1.el8_4.x86_64	
Processor		
Vendor and model	AMD EPYC 7V13	Intel Xeon Platinum 8272CL
vCPU count	64	64
Core frequency (GHz)	3.1	2.6
Memory module(s)		
Total memory in system (GB)	448	504
Network		
Network bandwidth	50 Gb/s Ethernet (40 Gb/s usable)	
Local storage (OS)		
Number of drives	1	
Drive size (GB)	30	
Drive information	Premium SSD LRS	
Local storage (data drive)		
Number of drives	1	
Drive information	Premium SSD LRS	
Drive size (TB)	8	
IOPS	20,000	
Performance tier	P80	
Drive bandwidth (MB/s)	900	

About the workloads

We conducted three workloads:

- The HammerDB TPROC-C workload, which uses MySQL
- The HiBench Spark Latent Dirichlet Allocation (LDA) workload
- The HiBench Spark Logistic Regression (LR) workload

Note: For geographic diversity, we used Azure VMs located in different regions. We ran the HammerDB tests on virtual machines in the East US region and the HiBench tests on VMs in the South Central US region. (For complete details on our testing, see the science behind the report.)

About 3rd Gen AMD EPYC 7V13 processors

These 64-core processors use AMD Infinity Architecture and are part of the AMD EPYC 7003 Series. The latest offering from AMD, 3rd Gen EPYC processors offer increased I/O with up to 32MB L3 cache per core, 7nm x86 hybrid die core, and new security features such as Secure Encrypted Virtualization - Secure Nested Paging (SEV-SNP).[2] According to AMD, "On premises, in the cloud, in containers, virtual machines, or on bare metal, 3rd Gen AMD EPYC 7003 Series CPUs are the market's best performing x86 server processor, helping provide faster time to results."[3]

Learn more at https://www.amd.com/en/processors/epyc-7003-series.

More MySQL OLTP database transactions

Companies rely on OLTP database applications for online shopping, customer relationship management, order entry, and more. The higher the transaction rate your cloud-based virtual machine can deliver, the better you can support your users.

We used the HammerDB TPROC-C benchmark to measure number of MySQL transactions per minute the two Azure VMs could handle. As Figure 1 shows, the Azure Standard_HB120-64rs_v3 VM backed by the AMD EPYC 7V13 processor achieved 38.5 percent more MySQL transactions per minute than the VM backed by the Intel Xeon Platinum 8272CL processor. This advantage could allow you to carry out a fixed amount of work using fewer VMs.

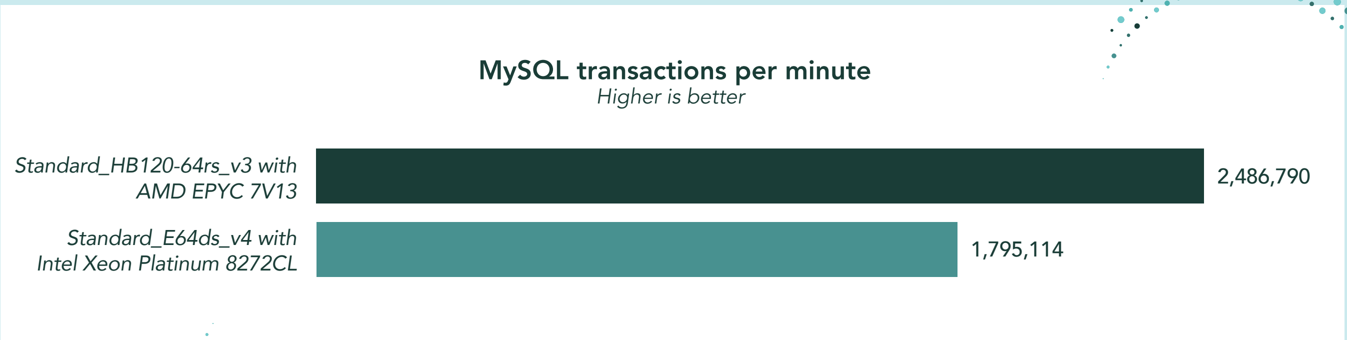

Figure 1: The number of MySQL transactions per minute the two VMs achieved on a HammerDB TPROC-C OLTP workload. Higher numbers are better. Source: Principled Technologies.

About the HammerDB TPROC-C workload

To assess each VM's MySQL database performance, we used a TPC-C-like OLTP workload from the HammerDB benchmarking suite called TPROC-C. Even though the HammerDB developers derived this workload from the TPC-C standard, it is not a full implementation of TPC specifications. Therefore, the results in this paper are not directly comparable to officially published TPC results.

To learn more about HammerDB and the TPROC-C benchmark, visit www.hammerdb.com.

Speedier completion of Spark machine learning workloads

Companies across many industries are integrating machine learning applications into their businesses. These powerful applications can be very compute-intensive, making it important to select a virtual machine platform that delivers on performance. We tested two Spark-based machine learning workloads from the HiBench benchmark suite, Latent Dirichlet Allocation and Logistic Regression. To do so, we created a five-node Hadoop® cluster (four worker nodes and one manager node) using each VM type.

Latent Dirichlet Allocation

Latent Dirichlet Allocation, or LDA, is a method of topic modeling that uses the words in a document to determine categories or topics to which the document belongs. One potential application of LDA is generating recommendations. For example, a library could use LDA to recommend books for you based on the topics of books you have previously checked out.

As Figure 2 shows, the Hadoop cluster of Standard_HB120-64rs_v3 VMs backed by the AMD EPYC 7V13 processor achieved more than twice the average throughput per worker node of the cluster backed by the Intel Xeon Platinum 8272CL processor, an improvement of 107 percent. Based on this greater rate, the cluster with the AMD EPYC 7V13 processor completed the LDA workload in less than half the time (see Figure 3). This advantage could allow organizations to analyze data more efficiently and reduce the number of VMs they need.

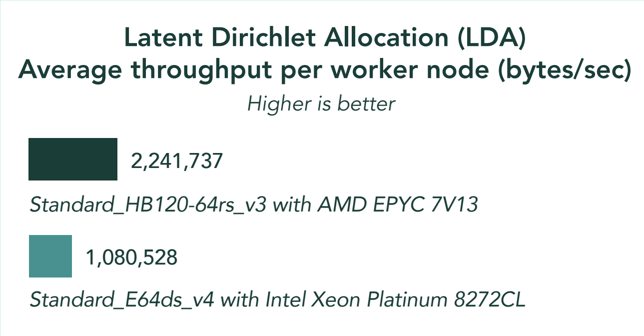

Figure 2: Throughput in bytes per second that two VM clusters achieved on a HiBench LDA workload. Higher numbers are better. Source: Principled Technologies.

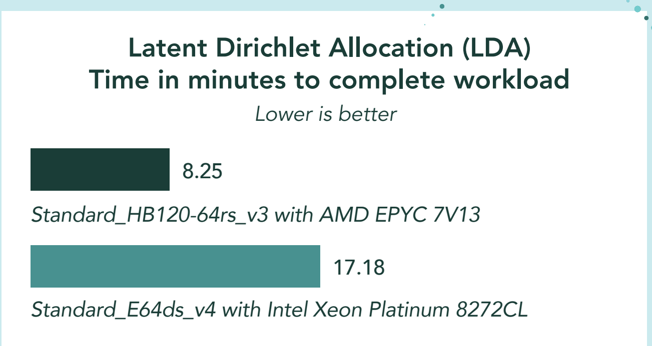

Figure 3: Time the two VM clusters took to complete a HiBench LDA workload. Lower numbers are better. Source: Principled Technologies.

Logistic Regression

Logistic Regression predicts probability by taking some number of input variables and producing a binary prediction such as on/off, 0/1, true/false, etc. An ecommerce company could use an LR algorithm to determine whether to recommend and display a certain similar product in an online shopper's cart based on the variety of other products they've viewed during their online shopping trips.

As Figure 4 shows, the cluster of Standard_HB120-64rs_v3 VMs with the AMD EPYC 7V13 processor achieved 59.0 percent greater average throughput per worker node than the cluster of VMs with the Intel Xeon Platinum 8272CL processor, which allowed it to complete the Logistic Regression workload in 37.1 percent less time (see Figure 5). As with LDA, this time savings could allow companies to execute their workloads with fewer virtual machines, leading to savings.

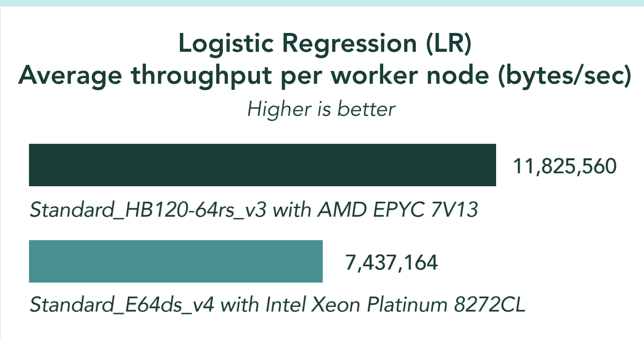

Figure 4: Throughput in bytes per second that two VM clusters achieved on a HiBench Logistic Regression workload. Higher numbers are better. Source: Principled Technologies.

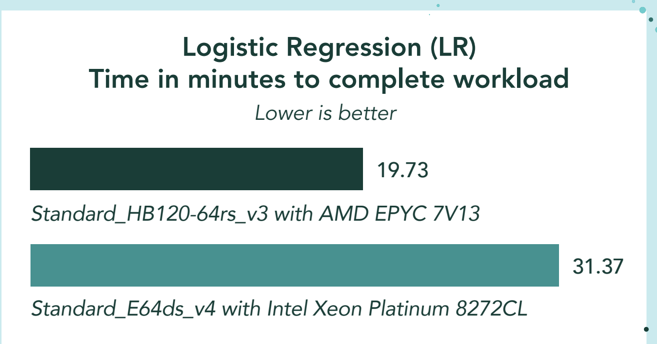

Figure 5: Time the two VM clusters took to complete on a HiBench Logistic Regression workload. Lower numbers are better. Source: Principled Technologies.

About HiBench

With Hadoop, Spark, and streaming workloads, the HiBench benchmark suite lets users measure speed, throughput, and resource utilization of big data frameworks.[4]

Learn more at https://github.com/Intel-bigdata/HiBench#readme.

Looking at pricing

When selecting public cloud VMs for your organization's critical workloads, cost is always an important consideration. We used "Pay-as-You-Go" pricing from the Azure VM pricing calculator[5] to estimate the monthly cost of the two environments we tested.

Using the Virtual Machines option from the Azure VM pricing calculator, we specified the exact configurations we used in our testing and recorded the cost estimates the calculator produced. The options to construct each virtual machine included the region in which it was run, its operating system and type, and the VM series and size. In addition to the VM's compute specifications, we added the number of disks necessary to support the application's database. Finally, we defined the bandwidth as a $0 cost data transfer type of internet egress routed via the public internet. We combined these three components—compute, storage, and network—for an estimated cost to run each application 24 hours a day for one month. (In the Azure pricing calculator, the cost for 730 hours equals the cost for 1 month. 730 hours, is one-twelfth of 8,760, the number of hours in 365 days.) We add the VM's monthly compute costs to the cost to support the necessary number of disks of the appropriate type (i.e., Premium SSD) and performance tier (i.e., P80 - 20,000 IOPs and 900MB/s).

Cost of the MySQL test environment

For the MySQL test infrastructure, we include a single VM of each type in our pricing model. Each of these VMs, on which we installed MySQL, had a single disk attached that held the database we used for testing. We exclude the VM we used as the HammerDB driver system from our pricing model because it was simulating remote users conducting transactions and not a part of the actual MySQL deployment.

As Figure 6 shows, selecting the Standard_HB120-64rs_v3 VM backed by the AMD EPYC 7V13 processor to run our MySQL environment for one month in the East US Azure region on a Pay-as-You-Go basis would cost $736 less than using the VM backed by the Intel Xeon Platinum 8272CL processor. This is a savings of 10.5 percent.

Cost of the Spark test environment

For the Spark test configuration, we built a four-worker-node Hadoop cluster with one manager node. Because the manager node is an integral part of a Hadoop infrastructure, we included its VM as part of our pricing model. Our test configuration consisted of the five VMs (one manager and four workers), each with identical hardware and software configurations. In addition, each worker node had a managed disk defined and attached to the VM. These four disks housed the Hadoop file system with the Spark dataset. We entered all the pertinent parameters into the calculator and recorded the estimated costs to run the application for one month on a Pay-as-You-Go basis.

As Figure 7 shows, selecting the Standard_HB120-64rs_v3 VM backed by the AMD EPYC 7V13 processor to run our HiBench workload for one month in the South Central US Azure region would cost $7,409 less than using the VM backed by the Intel Xeon Platinum 8272CL processor, a savings of 21.1 percent.

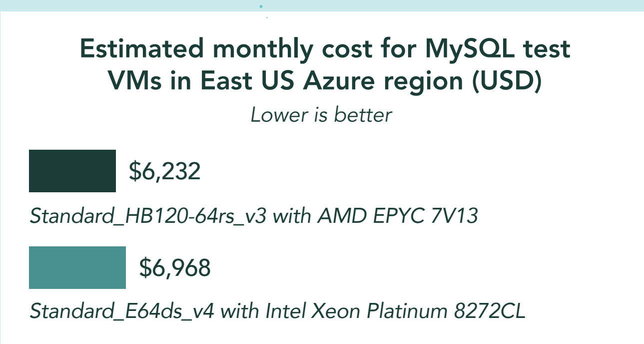

Figure 6: Estimated Pay-as-You-Go monthly costs for using the two VM types to complete our HammerDB MySQL workload. Lower numbers are better. Source: Principled Technologies.

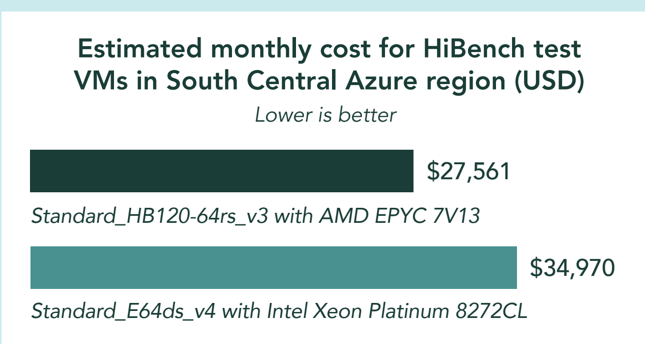

Figure 7: Estimated Pay-as-You-Go monthly costs for using the two VM clusters to complete our HiBench workload. Lower numbers are better. Source: Principled Technologies.

Conclusion

The flexibility that companies gain when moving workloads to the cloud may be high, but the monthly bill from their cloud provider can also be significant. This makes it important to select virtual machines that deliver strong performance while remaining cost-effective. In our testing, Azure Standard_HB120-64rs_v3 virtual machines powered by AMD EPYC 7V13 processors outperformed VMs powered by Intel Xeon Platinum 8272CL processors on OLTP and machine learning workloads, increasing throughput by as much as 107 percent on the HiBench test. Looking at Pay-as-You-Go pricing of Azure VMs, we also found that selecting Standard_HB120-64rs_v3 virtual machines with AMD EPYC 7V13 processors for the VM configurations we tested could lower a company's monthly expenditure on cloud compared to selecting Standard_E64ds_v4 VMs with Intel Xeon Platinum 8272CL processors.

---

1 Azure VM pricing calculator, accessed October 19, 2021, https://azure.microsoft.com/pricing/calculator/.
2 "AMD EPYC 7003 Series Processors," accessed October 19, 2021, https://www.amd.com/en/processors/epyc-7003-series.
3 "AMD EPYC 7003 Series Processors," accessed October 19, 2021, https://www.amd.com/en/processors/epyc-7003-series.
4 GitHub Repository for HiBench, accessed October 19, 2021, https://github.com/Intel-bigdata/HiBench#readme.
5 Azure VM pricing calculator, accessed October 19, 2021, https://azure.microsoft.com/pricing/calculator/.

Read the science behind this report at https://facts.pt/BCoarNA ▶

Principled Technologies® **Facts matter.®**

This project was commissioned by AMD.