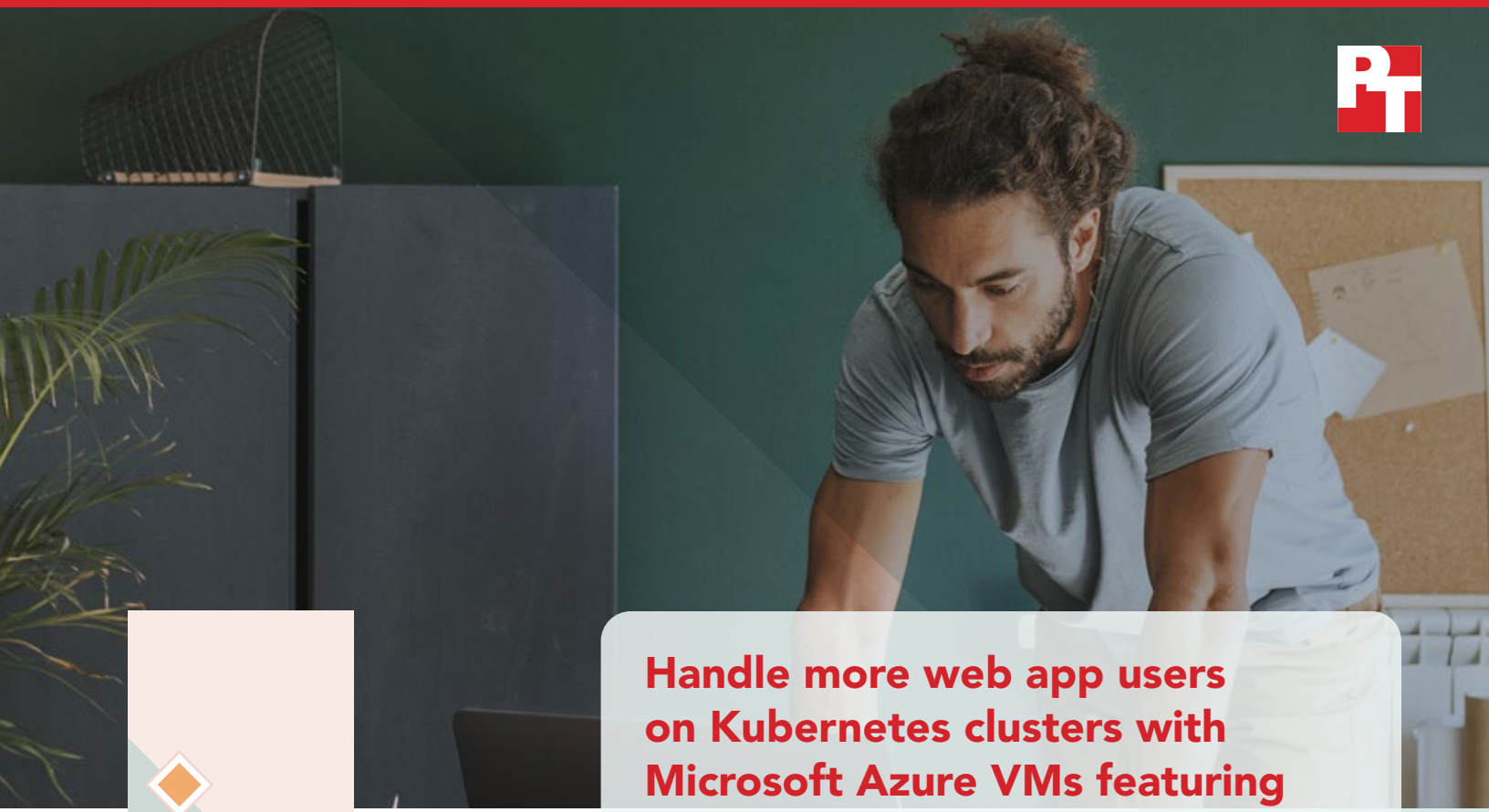


Handle more web app users on Kubernetes clusters with Microsoft Azure VMs featuring 2nd Generation Intel Xeon Scalable processors

Newer D- and E-series VMs with Cascade Lake processors vs. older VMs with processors from previous generations

Organizations that host multi-tiered web applications on cloud servers will want to use VMs capable of handling high-traffic peak times and increased user load.

At Principled Technologies, we used a multi-tiered web application (Weathervane 2.0) to assess the Kubernetes performance of various server clusters. These clusters ran VMs from four different Azure series: Newer Ds_v4 and Es_v4 series VMs featuring 2nd Generation Intel Xeon Scalable Processors (Cascade Lake), and older VMs that used Intel Xeon processors from a previous generation. For each data point we measured, we found that the clusters using newer VMs supported more Kubernetes web server users than the older VMs. In addition, there was no difference in cost between the newer and older versions of the VMs, meaning that newer Microsoft Azure VMs and Cascade Lake processors were more cost-effective.

Handle up to 1.58x the Weathervane users
with small D8s_v4 series VMs*

Handle up to 1.77x the Weathervane users
with medium E16s_v4 series VMs**

Handle up to 1.58x the Weathervane users
with large D32s_v4 series VMs*

*compared to older Ds_v3 series VMs
**compared to older Es_v3 series VMs

Our testing

Our testing involved Kubernetes and containerized applications. For those unfamiliar with these terms, here is a brief overview.

A container is a unit of software packaged up with everything required to run that software in a standalone state. Containers comprise everything an application needs to run, including binaries, libraries, dependencies, and of course, the application itself. Kubernetes is an open-source platform for deploying and managing applications that run in containerized environments.

We compared Kubernetes multi-tiered web app performance among three-node cloud server clusters. These clusters ran VMs from four different Azure series: Newer Ds_v4 and Es_v4 VMs featuring Intel Cascade Lake processors versus these VMs' older counterparts, Ds_v3 and Es_v3 VMs that featured Intel processors from a previous generation. (Note: VMs from the v3 generation are available with several CPU configurations. Please see the Science behind this report for more information.) The D-series VMs, Ds_v4 and Ds_v3, are general-purpose offerings, whereas the E-series VMs, Es_v4 and Es_v3, are memory-optimized.

We created these clusters using Azure Kubernetes Service (AKS), a tool that configures clusters to user-defined specifications without requiring hands-on setup. It is a fast method for creating functional Kubernetes clusters on Azure, but it also provides less flexibility and visibility into the underlying nodes.

We tested Kubernetes on both classifications of VM to measure the performance benefit that newer VMs would deliver for various sizing needs. We tested all VMs in the East US region. Figures 1 and 2 illustrate our cloud-based environment.

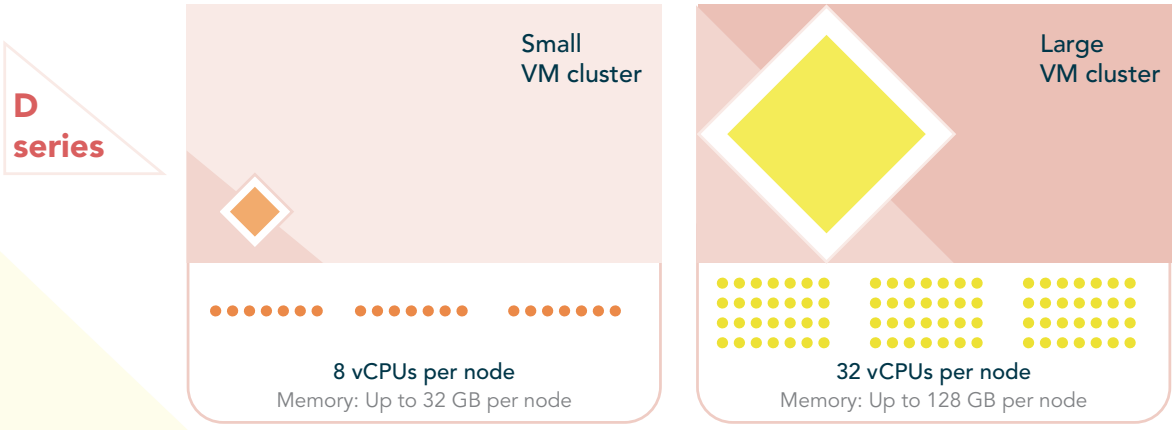


Figure 1: Key specifications for the D-series VMs. Source: Principled Technologies.

To test each Kubernetes cluster, we used the Weathervane 2.0 benchmark from VMware. According to the VMware website, Weathervane uses a "realistic multi-tier web application" called Auction, and consists of several service tiers.[1] Auction is a real-time app for hosting and managing online auctions. In the Weathervane implementation of Auction, simulated users populate the app, view photos and descriptions of the items up for sale, and compete against each other to make bids. Weathervane 2.0 measures the maximum number of simulated users (WvUsers) a cluster can support without breaking specified quality-of-service requirements.

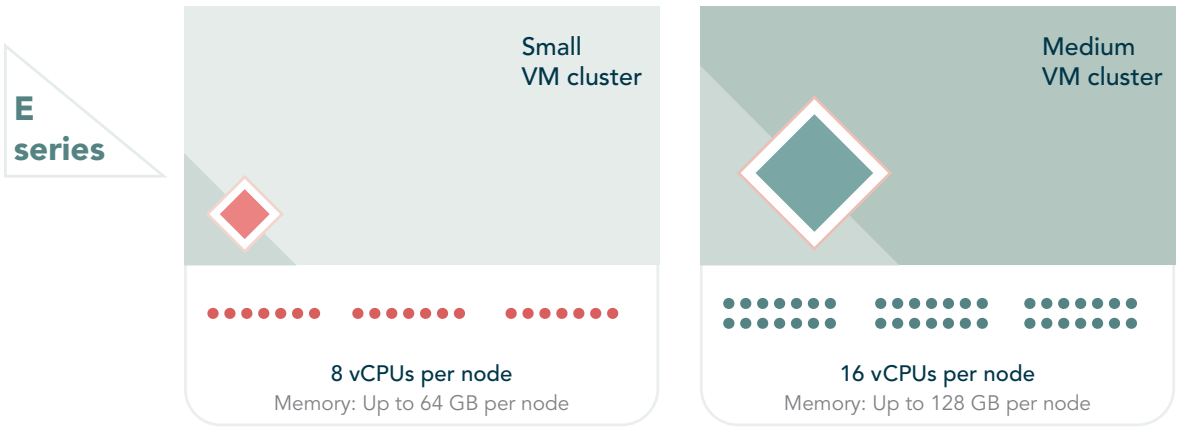


Figure 2: Key specifications for the E-series VMs. Source: Principled Technologies.

D-series VM testing results

Microsoft markets its Azure D-series VMs as being able to meet requirements for most workloads and use cases, positioning them as viable options for companies that require performance befitting a consistent user load.

Small VMs

Businesses that host smaller websites, such as for internal portals with employee resources and applications, may find that 8vCPU VMs from the Azure D-series line suit their needs nicely. Figure 3 shows that the cluster with newer D8s_v4 VMs supported 1.58 times the number of Weathervane 2.0 users as the cluster with older D8s_v3 VMs that used older-generation Intel Xeon processors.

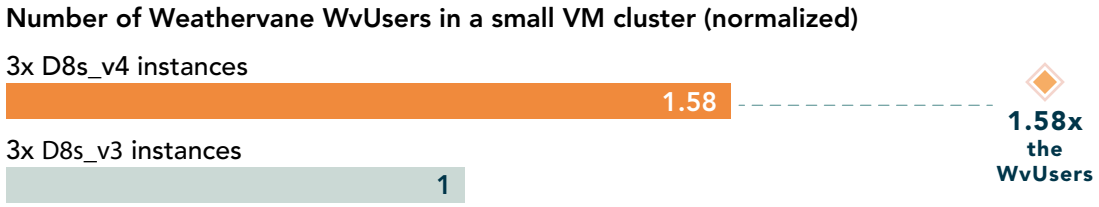


Figure 3: Relative number of Weathervane 2.0 users each cluster supported, normalized to the score achieved by the Ds_v3 instances. More users is better. Source: Principled Technologies.

Large VMs

Organizations that need to support larger public-facing websites will naturally want larger VMs to sustain them. In our tests, the cluster with D32s_v4 VMs powered by Intel Xeon Cascade Lake processors again supported 1.58 times the number of Weathervane 2.0 users compared to the cluster that used older D32s_v3 VMs.

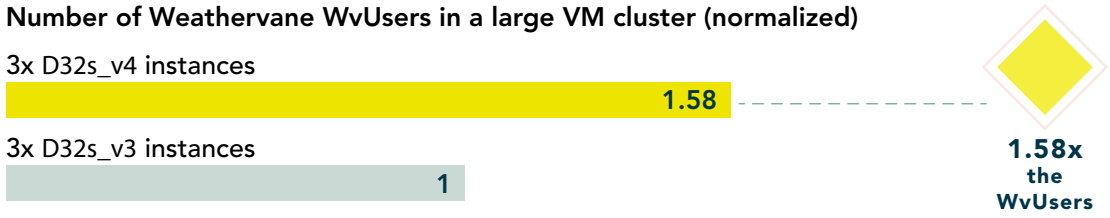


Figure 4: Relative number of Weathervane 2.0 users each cluster supported, normalized to the score achieved by the Ds_v3 instances. More users is better. Source: Principled Technologies.

Cost analysis: D-series VMs

Our hands-on testing suggests that the newer D-series VMs are more cost-effective for Kubernetes workloads than VMs from the previous version. There's no cost difference between the older and newer VMs, yet the cluster with Ds_v4 VMs delivered 1.58 times the performance of the cluster with Ds_v3 VMs.[2]

E-series VM testing results

Microsoft targets its Azure E-series VMs for memory-intensive applications such as SAP HANA, other applications based on relational databases that have large caches, and even in-memory analytics. This greater memory footprint may be useful for web applications within larger environments that support a large user base.

Small VMs

If your organization's internal website uses Kubernetes-managed applications that require more powerful and robust resources than the norm, you may want to investigate whether 8vCPU E-series VMs are right for your needs. In our tests, the cluster with E8s_v4 VMs powered by Intel Xeon Cascade Lake processors handled 1.48 times the number of Weathervane 2.0 users as the cluster with older E8s_v3 VMs that had Intel Xeon processors from a previous generation.

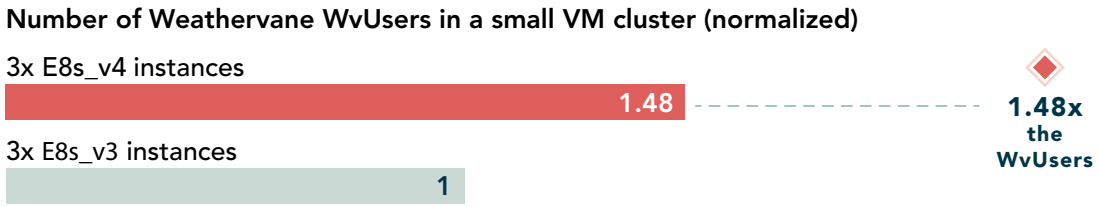


Figure 5: Relative number of Weathervane 2.0 users each cluster supported, normalized to the score achieved by the Es_v3 instances. More users is better. Source: Principled Technologies.

Medium VMs

If your organization's public-facing website runs an app that has more intense resource needs, Azure E-series VMs with 16 vCPUs may satisfy your requirements. Figure 6 shows that in our hands-on tests, the cluster with E16s_v4 VMs featuring current-generation Intel Xeon Cascade Lake processors supported 1.77 times the number of Weathervane 2.0 users as the cluster with E16s_v3 VMs and older-generation Intel processors.

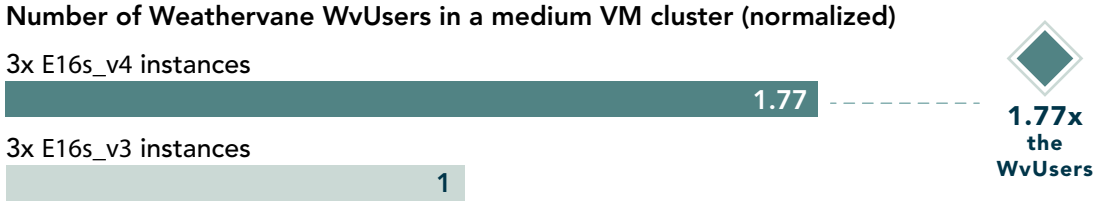
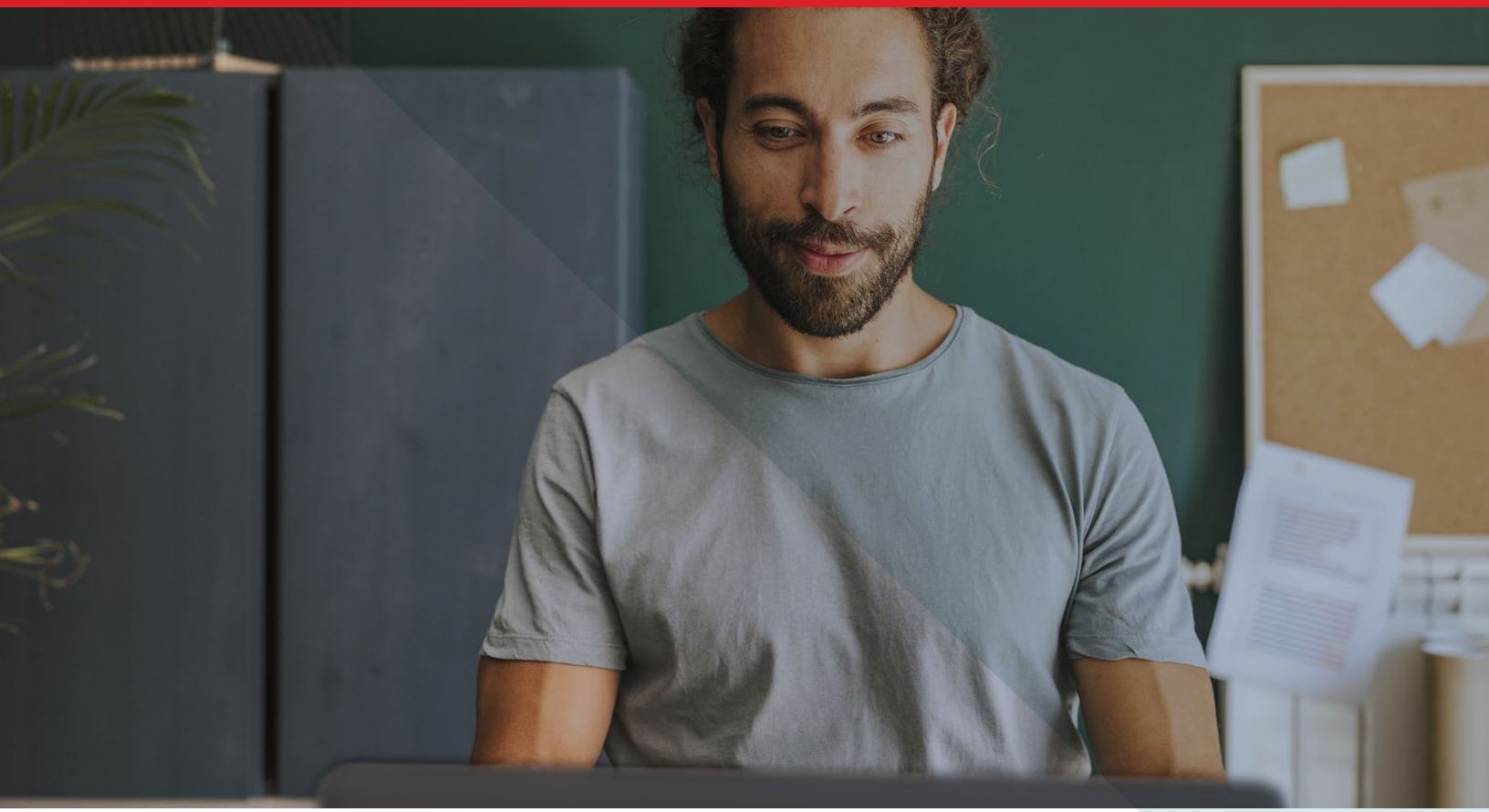


Figure 6: Relative number of Weathervane 2.0 users each cluster supported, normalized to the score achieved by the Es_v3 instances. More users is better. Source: Principled Technologies.

Cost analysis: E-series VMs

Newer VMs are once again the better investment for Kubernetes web server work, according to our hands-on tests. The latest generation of E-series VMs cost exactly the same as the previous version, but offered from 1.48 to 1.77 times the Kubernetes performance.[3]



Conclusion

When considering which cloud VMs to trust with your organization's Kubernetes-managed web server applications, it helps to understand the sort of performance you can expect from a variety of VM types.

We tested clusters that used general-purpose and memory-optimized VM series with different generations of Intel processors and at different VM sizes for a total of eight data points. In each case, we found that clusters using newer VMs featuring Intel Cascade Lake processors supported more Weathervane 2.0 users than clusters that used older VMs with Intel processors from a previous generation. Supporting more web app users can mean improved application performance and being able to support more users during peak times without breaking quality of service requirements.

---

1 Harold Rosenberg, "Weathervane 2.0: An Application-Level Performance Benchmark for Kubernetes," accessed April 5, 2021, https://blogs.vmware.com/performance/2020/02/weathervane2-kubernetes.html

2 "Windows Virtual Machine Pricing," accessed April 5, 2021, https://azure.microsoft.com/en-us/pricing/details/virtual-machines/windows/

3 "Windows Virtual Machine Pricing," accessed April 5, 2021, https://azure.microsoft.com/en-us/pricing/details/virtual-machines/windows/

Read the science behind this report at http://facts.pt/8q2N5VE ▶

Principled Technologies®

Facts matter.®

This project was commissioned by Intel.