Deliver a better virtual desktop experience with Dell Technologies APEX Private Cloud using NVIDIA GPUs

A VDI cluster with NVIDIA T4 GPUs delivered more frames per second and reduced end-user latency vs. the same cluster without GPUs

Selecting on-premises infrastructure with APEX Cloud Services to support your virtual desktop infrastructure (VDI) can provide IT flexibility—but the configuration you choose contributes to the type of experience desktop users will have.

In testing with APEX Private Cloud, we found that using the NVIDIA® GPU-enabled configuration provided a better overall virtual desktop experience while running the NVIDIA nVector Knowledge Worker workload compared to the same solution without GPUs. The NVIDIA GPU-powered APEX solution offered 21.4 milliseconds lower end-user latency and had a higher frame rate than the solution with CPUs alone.

To help your remote workforce better connect with a more responsive and smoother experience, consider deploying an APEX solution with NVIDIA GPUs to host your virtual desktops.

Keep latency in an acceptable range
21.4 milliseconds lower end-user latency (12.5% decrease)

Smoother user experience
8 more frames per second (57.3% increase)
How we tested

To see the impact that NVIDIA GPUs have on virtual desktop performance, we compared two nearly identical solutions on APEX Private Cloud; the only difference was one used NVIDIA GPUs, and the other used CPU alone for image processing and video rendering.

For testing, we used an APEX Private Cloud cluster consisting of four Dell VxRail™ V570F servers, each with:

- Two Intel® Xeon® Gold 6248R processors
- 768 GB of DDR4 RAM
- VMware® ESXi™ 7.0.2 hypervisor
- 56 TB of shared all-flash VMware vSAN™ storage

This four-node VxRail cluster hosted a total of 192 virtual desktops for testing.

For load generation and client hosting, we used a three-node cluster of Dell PowerEdge™ R730 servers, each with:

- Two Intel Xeon E5-2697 v3 processors
- 512 GB of DDR4 2,133MHz RAM
- VMware ESXi 7.0.2 hypervisor
- 4 TB of shared all-flash vSAN storage

We used VMware Horizon® version 2103 as the VDI solution to create and deploy a desktop pool of 192 Windows 10 VMs with 4 vCPUs and 8 GB of RAM each. For VMs configured to use GPUs, we used the Grid T4-1B vGPU profile.

We used PowerShell scripting to create a desktop pool of 192 matching VMs, prepare the environment for testing, and execute nVector test runs. We monitored the workload logs while the VMs were under test and used a PowerShell script to parse and analyze test results.

To learn more about our testing, read the science behind the report.

About the NVIDIA nVector benchmarking tool

The NVIDIA nVector benchmark quantifies the effect that virtual GPUs have on VDI performance. nVector collects a series of performance metrics to report performance. We used the following metrics for our testing:

- End-user latency (in milliseconds): Measures lag at the endpoint device when interacting with a remote application. The automated workflow captures wait times for common tasks, such as building slides and advancing through slides in the popular Microsoft PowerPoint app.

- Frame rate (frames per second): Measures the rate at which frames appear on an end user’s screen. In general, the higher the frame rate, the smoother the user experience is.

Meeting the challenges of virtual desktop performance with NVIDIA GPUs

The challenges of supporting remote workers with virtual desktop infrastructure center around providing the compute resources to deliver a smooth, local-like experience. When users must access their critical applications across a network, underpowered solutions can hamper user experience.

Latency—or the time that end users wait for an application to respond—can frustrate and distract users, making it difficult for them to stay on task. According to VMware, when protocol latency exceeds 150ms, the server goes into a warning state. In our tests with 192 VMs, the solution with GPUs performed under this threshold, while the solution without GPUs exceeded it. As Figure 1 shows, the NVIDIA GPU-enabled configuration reduced end-user latency by an average of 12.5 percent (21.4 milliseconds) running the Knowledge Worker workload.

Frame rate is another important measure of how well a VDI workload is performing. Lower frame rates can mean that end users have a choppy experience as they access the files and applications on their virtual desktops. This can again lead to frustration and distraction, lowering overall productivity. On the NVIDIA nVector benchmark, using the NVIDIA GPU-enabled configuration delivered a higher frame rate than the same servers without GPUs, averaging 57.3 percent more frames per second (see Figure 2). These results indicate that virtual desktop users could see a smoother overall experience with less lag with the NVIDIA GPU-powered solution.

Figure 1: End-user latency (in milliseconds) for both APEX Private Cloud configurations, reported by the NVIDIA nVector benchmark. Lower numbers are better. Source: Principled Technologies.

Figure 2: Frame rate (in FPS) for both APEX Private Cloud configurations, reported by the NVIDIA nVector benchmark. Higher numbers are better. Source: Principled Technologies.
Conclusion

Workers accessing their critical apps via remote desktops need a responsive experience that empowers them rather than hurts their productivity. For those using APEX Private Cloud solutions, choosing NVIDIA GPUs can help. NVIDIA nVector Knowledge Worker workload testing results showed that the NVIDIA GPU-enabled APEX Cloud configuration shortened wait times for end users by 21.4 milliseconds (a savings of 12.5 percent) and a 57.3 percent higher frame rate for a smoother video experience than the same solution without GPUs.

If your organization wants to provide a smoother virtual desktop experience for your employees, consider NVIDIA GPUs for your APEX Private Cloud solution.

---
