RTX Server Virtual Workflow

So what we're looking at here is how an NVIDIA RTX Server can host multiple desktops for a visual effects studio, or small studio. So we have three windows here which as actually three virtual machines.

The one on the right here, with the robot, is set up [for the way] a lighter would need to do their work. So a lighter, as the name implies, is doing a bunch of work to figure out where the lights need to go, how it should look, what colors. They need a lot of GPU power. So, in this case, we've got two virtual GPUs assigned into this machine - two RTX8000's in this case. And so you see as I work on this scene, I can make changes, I can change lights. The tool we're working in here is actually Autodesk Maya and the Autodesk Arnold Renderer.

Now, not everybody in a creative studio needs all that power. Maybe here we'll go to the top left which would be, say, what an animator would do. You can see, we can sit here as we move through the timeline in Maya here, we can scrub, we can change, we can adjust the animation and work. This doesn't need quite so much GPU power -- this actually just one RTX8000. Maya, when you're animating like this, actually likes a lot of GPU memory. So that full RTX8000 and the 48GB of memory helps all the frames, all the sequence, stay in memory and move smoothly.

Down to the bottom left is something a little smaller. Maybe somebody is just working on props, or concept art. They don't need a lot of performance. They don't need a lot of memory. In this case, we have half an RTX8000 going.

Now as you can see, all three of these are working at the same time because it's three different virtual machines on one RTX Server. Now the advantage to the studio that deploys this is it makes it very easy for them to reallocate the resources to whatever the artist needs. Got a client review coming up? You can turn all these things off, put all of the GPU in one virtual machine and get really fast and quick image resolution. So you can kind of move [GPUs] around based on what the task is. It's a lot easier than changing the GPU in somebody's physical workstation, to change it in a virtual workstation.

Now while this demo kind of represents what three different people would do, my honest answer is if what the artist needs is all three of these running at the same time on their desk, this is one of the easiest ways to get three machines on one desk, as well. [All the GPUs] can all come to [that one] workstation because of the RTX Server that's [out] there.

Also, we'll accelerate the batch rendering, when you turn in your job to go. Here's our queue manager, showing our batch render queue. It takes about 35 minutes per frame on the CPU. This is a dual Xeon render node - kind of a standard piece. So 30 minutes per frame.

The GPU one…. here, actually we started both of these this morning. We've completed 12 frames on our CPU node. The GPU node has already completed 150 frames. And we decided to submit it again, so we have four more going. We're running about 8 minutes per frame on the GPU. So you get that acceleration for every frame. But since we're on a system with 4 GPUs, we can render four frames at the same time. So you can see we've got four going simultaneously. So each one is faster, and we've got four going at the same time.

This is the same server. So for the studio, they can allocate between batch rendering and desktop use. Nighttime comes, the can move VMs off desktop use, put them to batch rendering. Crunch time comes, they can add more to the batch rendering. So it gives them lots of flexibility to adjust what they need to tailor how the GPUs get to their artists to maintain their production workflow.