

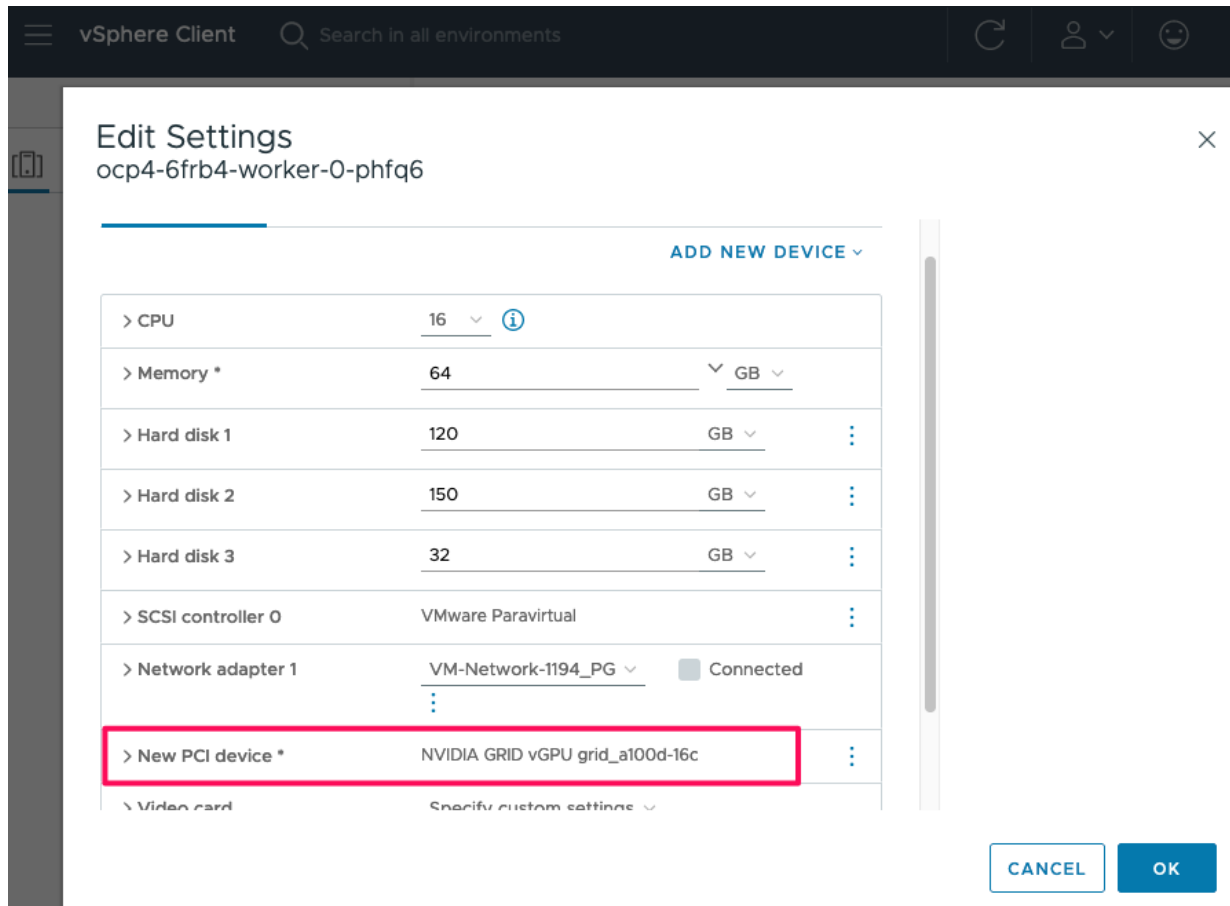
08_CVD: Add vGPUs to Red Hat OCP worker node VMs

Add vGPUs to Red Hat OCP worker nodes

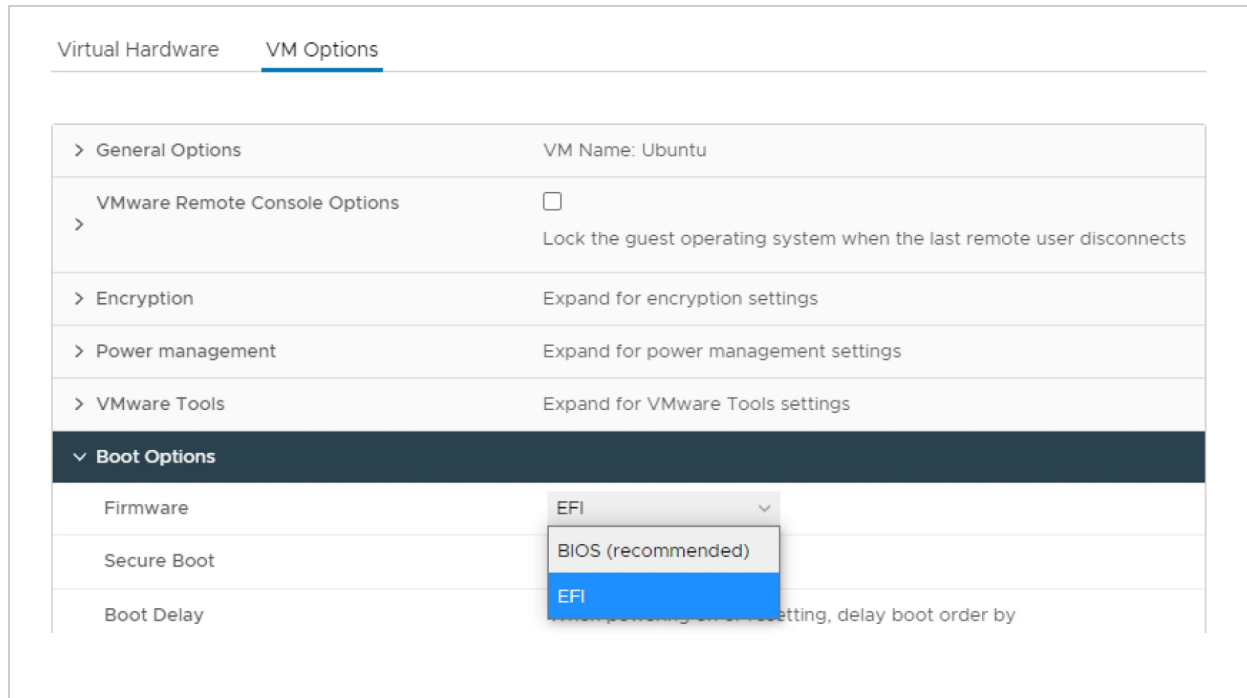
To add vGPUs to OCP worker nodes, the VMs need to be powered off.

Deployment Steps:

1. From VMware vCenter, select a worker node VM to add vGPU to and gracefully shut it down.
2. Right-click on the virtual machine and select **Edit Settings** from the menu.
3. Select the **Virtual Hardware** tab, from the **ADD NEW DEVICE** drop-down menu on the right, select **New PCI device**. From the list of available vGPU profiles, select the one that best meets your needs. On any given GPU, the first vGPU profile applied to a VM will determine the vGPU size/profile for the remaining vGPUs available, up to the maximum GPU frame buffer capacity (for e.g. 80G for A100-80G) - all vGPUs must be of the same size/profile. Multiple vGPUs can be assigned to a given VM or shared among multiple VMs.



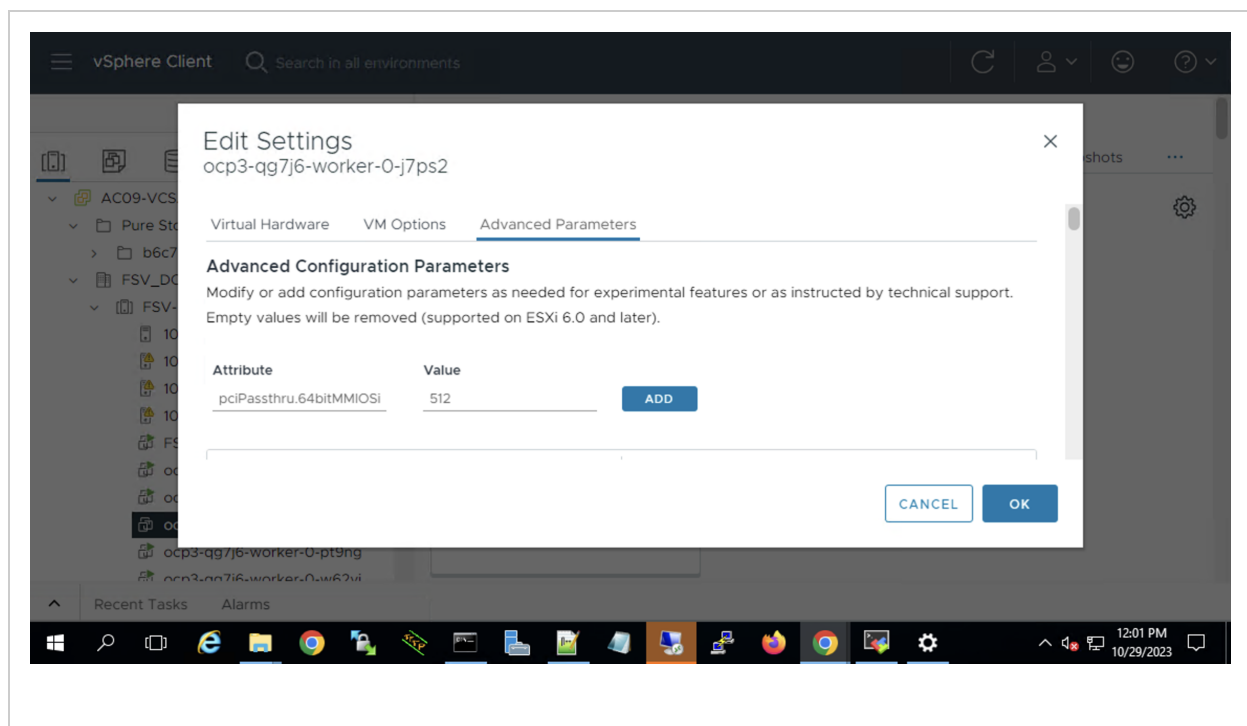
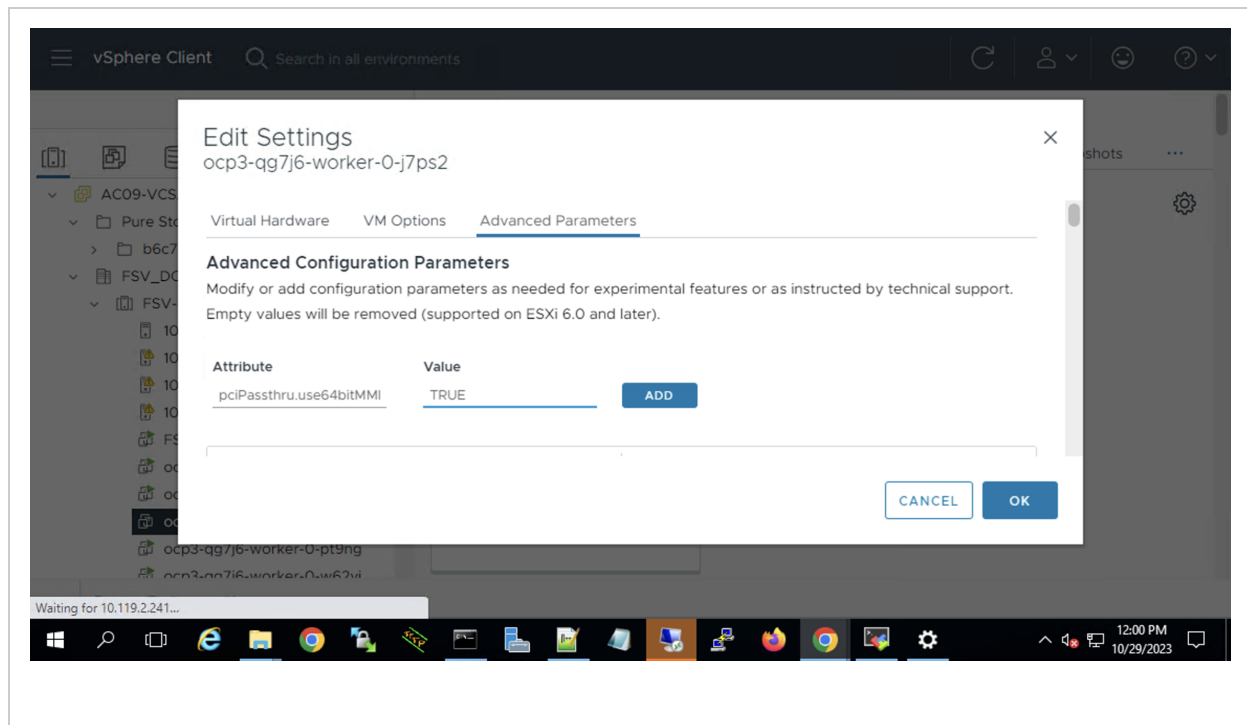
3. Select the **VM Options** tab, expand **Boot Options**, change the **Firmware** from **BIOS** to **EFI**. If **Secure Boot** (below **Firmware**) is enabled, it must also be disabled.

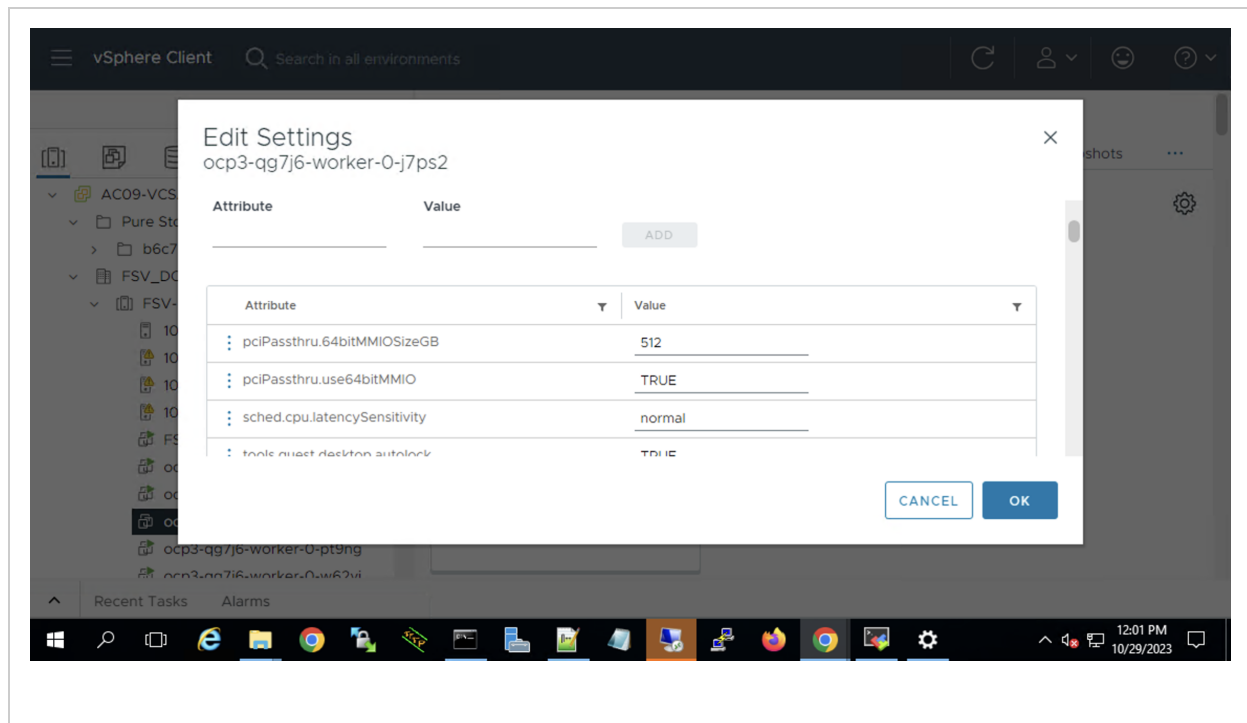


4. Adjust the Memory Mapped I/O (MMIO) settings for the VM under **Advanced Parameters**. The size will depend on the NVIDIA GPU model (see NVIDIA documentation for your GPU model):

- **pciPassthru.use64bitMMIO = TRUE**
- **pciPassthru.64bitMMIOSizeGB = 512**

Select the **Advanced Parameters** tab, add the following **Attributes** and **Values**.





5. Power the Virtual Machine backup.
 6. Repeat for remaining worker nodes.
 7. Wait for all worker nodes to be in **Ready Status** from the Red Hat OpenShift console.
-