



GPU considerations

NetApp Solutions

Dorian Henderson, Suresh Thoppay
April 16, 2021

Table of Contents

GPU considerations 1

GPU considerations

GPUs are typically used for graphic visualization (rendering) by performing repetitive arithmetic calculations. This repetitive compute capability is often used for AI and deep learning use cases.

For graphic intensive applications, Microsoft Azure offers the NV series based on the NVIDIA Tesla M60 card with one to four GPUs per VM. Each NVIDIA Tesla M60 card includes two Maxwell-based GPUs, each with 8GB of GDDR5 memory for a total of 16GB.



An NVIDIA license is included with the NV series.

Graphics Card

Sensors

Advanced

Validation



Name NVIDIA Tesla M60

Lookup

GPU GM204

Revision

FF

Technology 28 nm

Die Size

398 mm²

Release Date Aug 30, 2015

Transistors

5200M

BIOS Version

84.04.85.00.03

☐ UEFI

Subvendor NVIDIA

Device ID

10DE 13F2 - 10DE 115E

ROPs/TMUs 64 / 128

Bus Interface

PCI



Shaders 2048 Unified

DirectX Support

12 (12_1)

Pixel Fillrate 75.4 GPixel/s

Texture Fillrate

150.8 GTexel/s

Memory Type GDDR5 (Hynix)

Bus Width

256 bit

Memory Size 8192 MB

Bandwidth

160.4 GB/s

Driver Version

27.21.14.5257 (NVIDIA 452.57) / 2016

Driver Date Oct 22, 2020

Digital Signature

WHQL

GPU Clock 557 MHz

Memory

1253 MHz

Boost

1178 MHz

Default Clock 557 MHz

Memory

1253 MHz

Boost

1178 MHz

NVIDIA SLI

Disabled

Computing

☒ OpenCL☐ CUDA☒ DirectCompute☒ DirectML

Technologies

☒ Vulkan☐ Ray Tracing☐ PhysX☒ OpenGL 4.6

NVIDIA Tesla M60



Close

With NetApp HCI, the H615C GPU contains three NVIDIA Tesla T4 cards. Each NVIDIA Tesla T4 card has a Turing-based GPU with 16GB of GDDR6 memory. When used in a VMware vSphere environment, virtual machines are able to share the GPU, with each VM having dedicated frame buffer memory. Ray tracing is available with the GPUs on the NetApp HCI H615C to produce realistic images including light reflections. Please note that you need to have an NVIDIA license server with a license for GPU features.

Graphics Card

Sensors

Advanced

Validation



Name NVIDIA GRID T4-8Q

Lookup

GPU TU104

Revision

A1

Technology 12 nm

Die Size

545 mm²

Release Date Sep 13, 2018

Transistors

13600M

BIOS Version

0.00.00.00.00

☐ UEFI

Subvendor NVIDIA

Device ID

10DE 1EB8 - 10DE 130F

ROPs/TMUs 8 / 160

Bus Interface

PCI



Shaders 2560 Unified

DirectX Support

12 (12_2)

Pixel Fillrate 4.7 GPixel/s

Texture Fillrate

93.6 GTexel/s

Memory Type

GDDR6

Bus Width

256 bit

Memory Size

8192 MB

Bandwidth

Unknown

Driver Version

27.21.14.5257 (NVIDIA 452.57) / 2016

Driver Date

Oct 22, 2020

Digital Signature

WHQL

GPU Clock

585 MHz

Memory

0 MHz

Shader

N/A

Default Clock

585 MHz

Memory

0 MHz

Shader

N/A

NVIDIA SLI

Disabled

Computing

☒ OpenCL☒ CUDA☒ DirectCompute☐ DirectML

Technologies

☒ Vulkan☒ Ray Tracing☐ PhysX☒ OpenGL 4.6

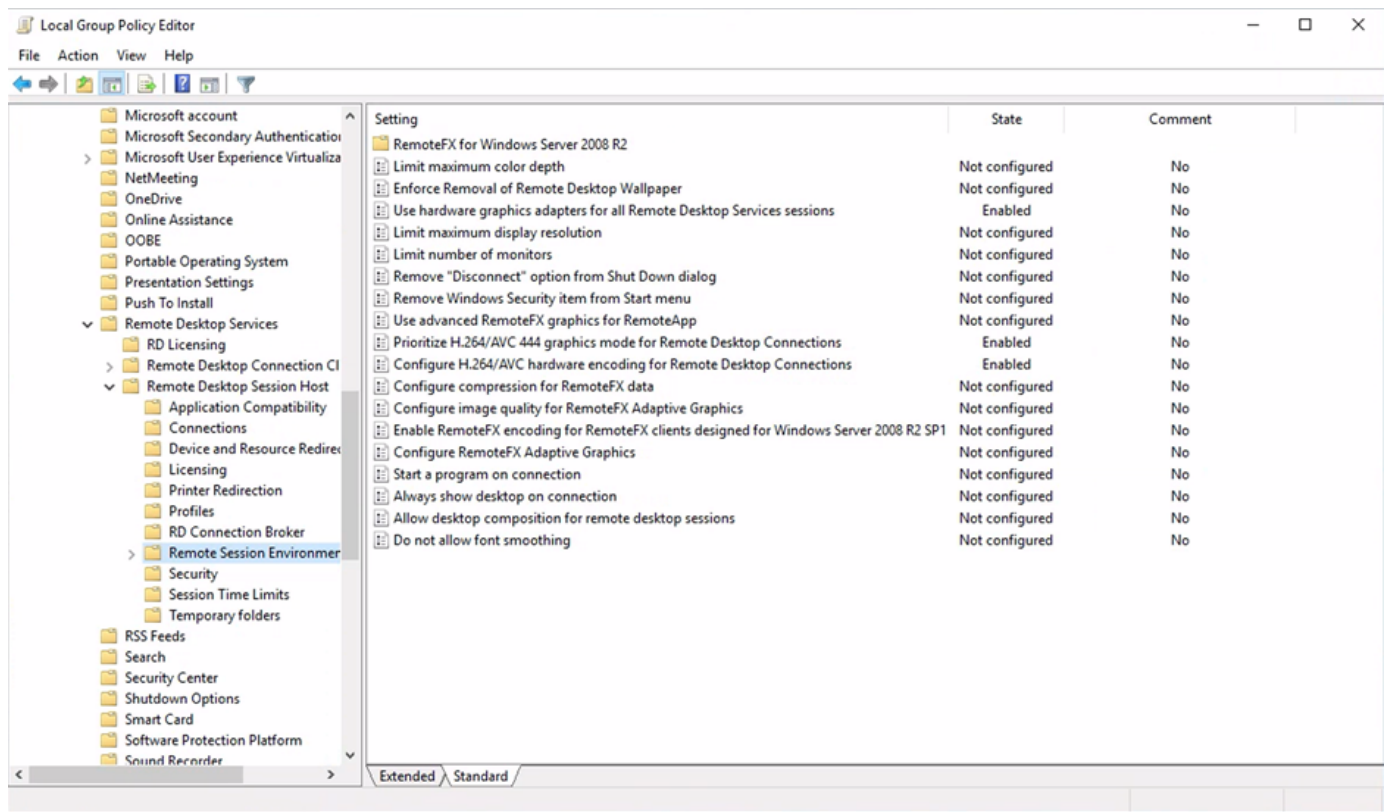
NVIDIA GRID T4-8Q



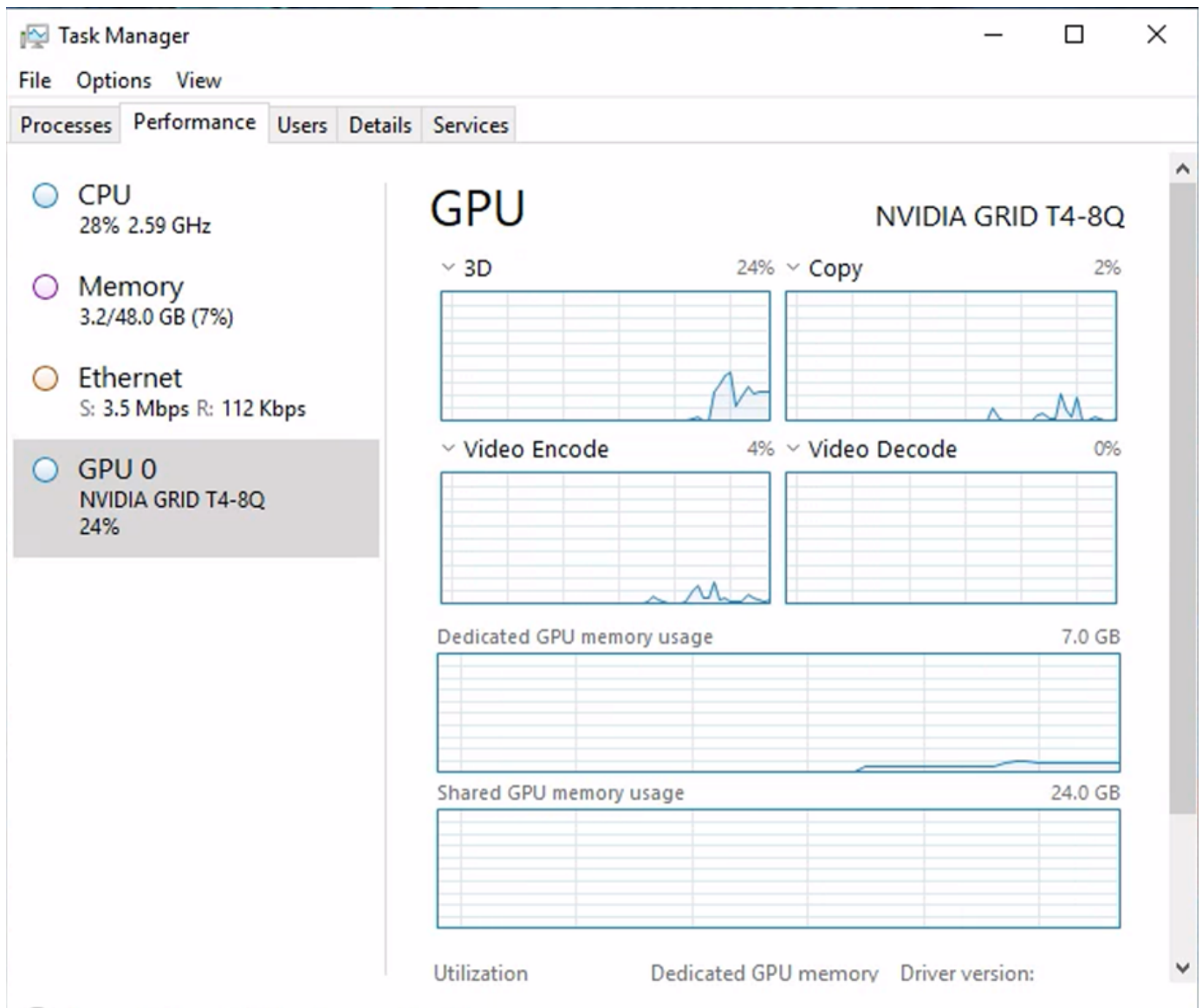
Close



To use the GPU, you must install the appropriate driver, which can be downloaded from the NVIDIA license portal. In an Azure environment, the NVIDIA driver is available as GPU driver extension. Next, the group policies in the following screenshot must be updated to use GPU hardware for remote desktop service sessions. You should prioritize H.264 graphics mode and enable encoder functionality.



Validate GPU performance monitoring with Task Manager or by using the nvidia-smi CLI when running WebGL samples. Make sure that GPU, memory, and encoder resources are being consumed.



To make sure that the virtual machine is deployed to the NetApp HCI H615C with Virtual Desktop Service, define a site with the vCenter cluster resource that has H615C hosts. The VM template must have the required vGPU profile attached.

For shared multi-session environments, consider allocating multiple homogenous vGPU profiles. However, for high end professional graphics application, it is better to have each VM dedicated to a user to keep VMs isolated.

The GPU processor can be controlled by a QoS policy, and each vGPU profile can have dedicated frame buffers. However, the encoder and decoder are shared for each card. The placement of a vGPU profile on a GPU card is controlled by the vSphere host GPU assignment policy, which can emphasize performance (spread VMs) or consolidation (group VMs).

Next: [Solutions for industry.](#)

Copyright Information

Copyright © 2021 NetApp, Inc. All rights reserved. Printed in the U.S. No part of this document covered by copyright may be reproduced in any form or by any means-graphic, electronic, or mechanical, including photocopying, recording, taping, or storage in an electronic retrieval system-without prior written permission of the copyright owner.

Software derived from copyrighted NetApp material is subject to the following license and disclaimer:

THIS SOFTWARE IS PROVIDED BY NETAPP "AS IS" AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, WHICH ARE HEREBY DISCLAIMED. IN NO EVENT SHALL NETAPP BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

NetApp reserves the right to change any products described herein at any time, and without notice. NetApp assumes no responsibility or liability arising from the use of products described herein, except as expressly agreed to in writing by NetApp. The use or purchase of this product does not convey a license under any patent rights, trademark rights, or any other intellectual property rights of NetApp.

The product described in this manual may be protected by one or more U.S. patents, foreign patents, or pending applications.

RESTRICTED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (c)(1)(ii) of the Rights in Technical Data and Computer Software clause at DFARS 252.277-7103 (October 1988) and FAR 52-227-19 (June 1987).

Trademark Information

NETAPP, the NETAPP logo, and the marks listed at <http://www.netapp.com/TM> are trademarks of NetApp, Inc. Other company and product names may be trademarks of their respective owners.