



Introducing RC's new NVIDIA Grace Hopper Superchip



Research Computing
UNIVERSITY OF COLORADO BOULDER

Introducing RC's new NVIDIA Grace Hopper Superchip

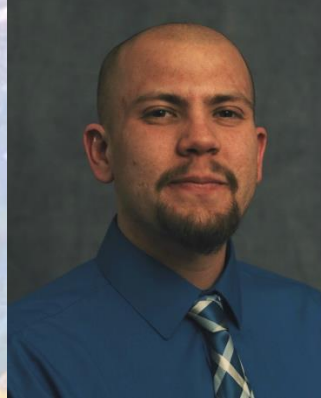
Instructor: Brandon Reyes

- Research Computing
- Website: www.rc.colorado.edu
- Documentation: <https://curc.readthedocs.io>
- Helpdesk: rc-help@colorado.edu
- Survey: <http://tinyurl.com/curc-survey18>

Meet the User Support Team



Layla
Freeborn



Brandon
Reyes



Andy
Monaghan



Michael
Schneider



John
Reiland



Dylan
Gottlieb



Mohal
Khandelwal



Ragan
Lee

Slides

https://github.com/ResearchComputing/introducing_rc_gh200_quick_byte



Session Overview

- GH200 architecture overview
 - Hardware specs
- What kind of workflows will benefit from the GH200 architecture?
- RC's approach to software management
- Beta testing phase
- How can I run on the GH200s?

GH200 architecture overview

The Grace Hopper Superchip (GH200) is a newer chip provided by NVIDIA that allows the Grace CPU and Hopper GPU to concurrently and transparently access both the CPU and GPU memory

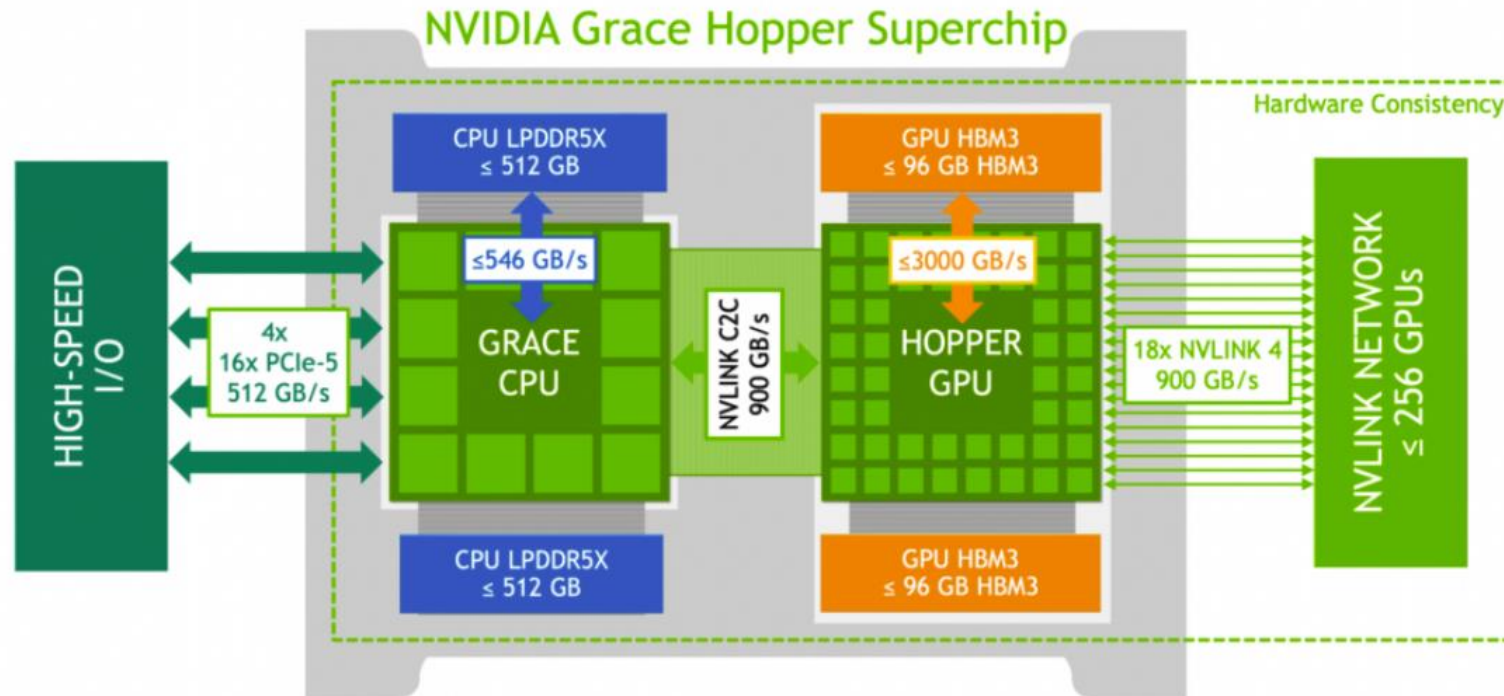


Image provided by <https://resources.nvidia.com/en-us-grace-cpu/nvidia-grace-hopper>

- Highlight
 - ARM based
 - Grace CPU has 72 cores and roughly 480 GB of RAM
 - Hopper GPU is an H100 and has roughly 100 GB of VRAM
 - CPU and GPU are connected via NVIDIA NVLink-C2C
 - Memory coherent, low latency, and high-bandwidth (900 GB/s)
 - FAST I/O speed (512 GB/s)
 - Roughly 1.7 TB of usable SSD on the node
 - We have not enabled linking of multiple nodes

- NVLink-C2C Allows for seamless memory management

Thank you!

Survey and feedback

<http://tinyurl.com/curc-survey18>



Slides

https://github.com/ResearchComputing/introducing_rc_gh200_quick_byte

