UF AI Initiative and HiPerGator 3.0 & AI

January 14, 2021

UF | Information Technology

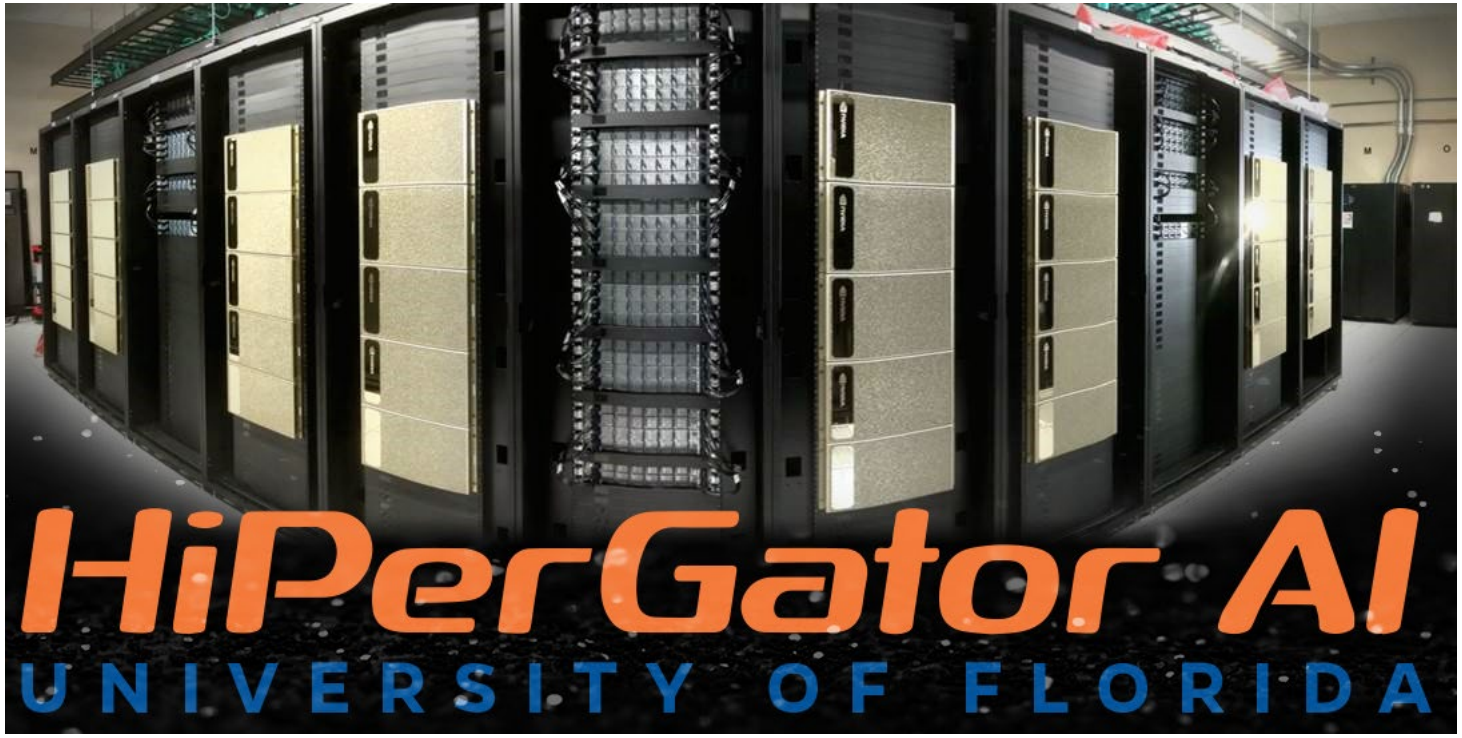
Rise to Five

UF AI University - timeline

- April 29, 2020 – Provost's AI initiative
- May 14, 2020 – Jensen Huang announces A100
- June 25, 2020 – two DGX A100 nodes arrive at UF
- July 21, 2020 – UF and NVIDIA announce partnership
- Nov-Dec 2020 – HiPerGator 3.0 and NVIDIA SuperPOD delivery
- Jan 2021 – HiPerGator 3.0 in production
- Jan 2021 – HiPerGator AI system validation and early user access

HiPerGator 3.0

- HiPerGator evolution
  - HiPerGator 1.0 – 2013 – 16,000 AMD cores – 4 GB RAM/core
  - HiPerGator 2.0 – 2015 – 30,000 Intel cores – 4 GB RAM/core
- HiPerGator 3.0
  - Dec 2019 – 608 new Nvidia RTX 2080ti and RTX 6000 GPUs
  - July 2020 – 4 PetaByte new “blue” fast storage
  - Jan 2021 – 30,720 AMD EPYC “Rome” 2.0 GHz cores – 8 GB RAM/core
  - ~May 2021 – 9600 AMD EPYC “Milan” cores – 8 GB RAM/core
  - Total core count 70,320 cores
    - retire 16,000 cores of HiPerGator 1.0
  - Double precision Linpack (HPL) ~1 Petaflops = 1 M x 1 B ops/sec



- All access to HiPerGator AI goes through HiPerGator
- Open data login nodes
  - Interactive shell work and batch job work
- Restricted data login through ResVault server
  - All work in VMs running on secure VM hosts
  - Interactive work and batch job work in (clusters of) secure VMs

HiPerGator AI
UNIVERSITY OF FLORIDA

NVIDIA SuperPOD

- 140 Nvidia DGX A100 nodes
- 17,920 AMD 7742 2.25 GHz “Rome” cores w. 8 GB RAM per core
- 1,120 Nvidia “Ampere” A100 GPUs
- 4 PetaByte all-flash DDN A3I AI400 storage
- 250 InfiniBand and Ethernet Mellanox switches
- Double precision Linpack (HPL) 13.75 Petaflops = 13.75 M x 1 B ops/sec
- AI floating point operations 0.7 Exaflops = 0.7 B x 1 B ops/sec

How to get started?

- For education use in courses
  - Contact UFIT Research Computing to set up an allocation for a semester
  - RC staff provides training in class
  - TAs respond to student problems, RC staff handles system problems
- For research use:
  - Principal investigator investments
    - For faculty and collaborators
  - College or department investments
    - For all faculty in the unit
- Visit https://www.rc.ufl.edu/access/purchase-request/

How does it work?

- Allocations for research, called “investments”
  - Configure a virtual “cluster” with
    - A number CPU cores with 8 GB RAM/core
    - A number of TB of storage
      - Blue → high performance, for running jobs
      - Orange → good performance, for keeping data accessible
      - Red → super performance for HiPerGator AI scratch
    - A number of GPU cards
  - Duration
    - multiple of 6 months for “service/lease”, or 5 year for “hardware”
  - See https://www.rc.ufl.edu/services/rates/
  - Purchase form https://www.rc.ufl.edu/access/purchase-request/

A flexible path to start…

- Consider the recommended option to get started:
- Buy a shared allocation, “investment,” for
  - College
  - Department
  - Institute
- This gives flexibility to provide faculty and their collaborators access
  - learn,
  - explore,
  - experiment, and
  - Develop courses
- Then funded projects can buy dedicated investments as needed

Training resources

- Take user training basic and any advanced training on use of HiPerGator
  - https://help.rc.ufl.edu/doc/Training
- AI specific preparation
  - NVIDIA has a lot of great material at the Deep Learning Institute (DLI)
  - https://www.nvidia.com/en-us/deep-learning-ai/education/
  - Get DLI Ambassador certified to teach the training materials
  - https://developer.nvidia.com/dli/amb_program_benefits
  - UFIT Research Computing Ying Zhang has DLI Ambassador certificate

Information: news, events, training, support, consulting, …

- UF: the AI University
  https://ai.ufl.edu
- UFIT Research Computing infrastructure
  https://www.rc.ufl.edu