Argonne NATIONAL LABORATORY

# Multitenancy at the ALCF

MAGIC Meeting, Feb 1st, 2023

William (Bill) E. Allcock, ALCF Director of Operations/Manager Advanced Integration Group

# To date, ALCF has only done whole node scheduling

- You can call that space shared multitenancy, node level multitenancy, system multitenancy, or no multitenancy as you choose.
- Why?:
  - —User Isolation:
    - Security: We can't get an acceptable (to us) level of user isolation at an acceptable level of performance.
    - Fault/Failure: It is too easy for one user to impact the other users on the node if they have a misbehaving application.
  - —Performance: Everyone complains about this, but it is even more true for the leadership computing facilities. More about that on the next slide.
  - —The architectures we ran (notice the past tense): "Thin" nodes, but lots of them on highly scalable networks.

# A Small Digression on Leadership Computing

- How are we different?  Our fundamental mission is to run the computations that simply can't be run anywhere else. To enable science that couldn't otherwise be done because it either couldn't be computed, or couldn't be computed in a reasonable amount of time.  I have often said that our tagline should be: "If you **CAN** run somewhere else, you **SHOULD** run somewhere else" because we turn away a significant percentage of the applicants for time on our systems.  We want the biggest, baddest computational science on the planet with the greatest impact. We regularly run jobs that consume ½, ¾ or the entire machine in a single job.
- What does that have to do with multitenancy?  Everyone complains about the performance hit that comes with multitenancy.  The users we tend to have are generally scraping for every last FLOP they can get and are as close to the bare metal as they can be, so the idea of sharing a node is the antithesis of what they want.  They would kick everyone else off the system so they could have the network to themselves if they could get away with it ☺.
- AI is changing this, however.  The science researchers using AI are usually users of standard packages, not developing the AI algorithms, so they are not pushing the limits of the nodes as they might have in the past.

# The times, they are a-changin’

- Architectures:
  - —Our first three systems were Blue Genes:  Thin, low power consumption nodes, on a highly scalable network.  Sharing those nodes really made no sense.  There wasn’t enough there to share and the system support wasn’t there to do it. Though even on Mira, it had a quad FPU and very few projects took advantage of it.  These are all retired.
  - —Theta: A bit of a “tweener”.  It has fairly weak x86 cores, but each node has 64 of them.  Maybe there are enough resources there to share, but again, with the proprietary network and stack, it is difficult.  Theta is currently scheduled to retire 12/31/2023.
  - —Polaris (our newest system): This is a ”fat” node. A single 32 core Milan CPU and four Nvidia A100 GPUs.  Users are going to have a difficult time taking full advantage of the computational resources on the node.  To avoid wasting computational resources we are investigating options for how to share nodes.
  - —Aurora (being installed): Even fatter. 2x  Intel Xeon® CPU Max (the CPU formerly known as Sapphire Rapids with HBM) CPUs and 6x Intel® Data Center GPU Max (the GPU formerly known at Ponte Veccio).
  - —Networks: Polaris and Aurora are also running Slingshot which is less proprietary and more Ethernet like, which gives us options.

# The times, they are a-changin’

- Security
  - —Confidential computing is becoming a hot topic and vendors are adding HW support for user isolation.  I am by no means an expert in this field, but it is an area we are keenly interested in and investigating.  Some examples:
    - The Intel Xeon® CPU Max , which will be in Aurora, has full memory encryption and Trusted Domain Extensions (TDX) though the latter is more for cloud computing and virtual machines.
    - The Nvidia Hopper H100 GPU has a number of hardware isolation features to support confidential computing.  This also helps with fault isolation.
    - Smart NICs: Using a smart NIC as a “trusted” piece of HW that the user never gets on and assuming the node is utterly untrusted
- Performance
  - — As encryption becomes the norm and is built directly into the data path, the impact on performance drops.
    - 20%-30% or more was not uncommon for encryption in software
    - Intel claims 2% overhead for their Full Memory Encryption
    - Someday, hopefully in the not too distant future, end-to-end encryption with appropriate hand-off between subsystems will be the norm and you won’t be able to move data unencrypted because there will be no benefit to doing so.

# How are we going to experiment with multitenancy?

- Where projects explicitly allow / request it
  - —We are working closely with the Advanced Photon Source (APS) at Argonne on on-demand computing, so that ALCF can process the data as it comes off the beam line and get them results in time to adjust the next data acquisition. While they are working on porting codes to GPUs, that takes time and some codes are just inherently serial, so they want to mix CPU heavy and GPU heavy/CPU light jobs on the same node.
- "Virtual nodes"
  - —This is effectively cgroups, but it is a PBS (which is the scheduler we run and contribute to) concept. For instance on Polaris, our Milan 32 core CPU is made up of (4) 8 core chiplets and the topology is such that each of the GPUs hangs off of one of the chiplets, so we can very nicely carve the system up into four single GPU "vnodes" that PBS will recognize and schedule as if it were a normal node. All the risks of a user breaking out of the cgroups isolation are still there though.
- Greed ☺
  - —For projects that are out of time and are constrained to run opportunistically in backfill, we are considering adding a flag they can add saying they are willing to accept the risk of sharing a node in order to increase their ability to get extra hours through backfill.
    - Though, since most of our users specify all 32 cores in their request the opportunities here may be limited.
    - On the other hand, it might be incentive for them to be more accurate about their resource needs…

Argonne NATIONAL LABORATORY

# Questions?

This research used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357.

*"Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Networking and Information Technology Research and Development Program."*

The Networking and Information Technology Research and Development (NITRD) Program

**Mailing Address:** NCO/NITRD, 2415 Eisenhower Avenue, Alexandria, VA 22314

**Physical Address:** 490 L'Enfant Plaza SW, Suite 8001, Washington, DC 20024, USA Tel: 202-459-9674, Fax: 202-459-9673, Email: nco@nitrd.gov, Website: https://www.nitrd.gov