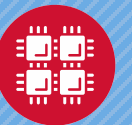


Pitzer Information Session

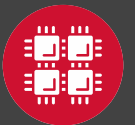
SUG

October 4, 2018



Agenda

- Pitzer hardware overview and timeline
- Software environment
- Scheduling policies and features
- Early user program
- Oakley decommissioning
- Discussion/Q&A





Pitzer Hardware Overview

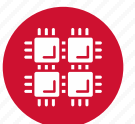
Doug Johnson, Chief Systems Architect and Manager HPC
Systems Group

Pitzer Hardware Details

- Goals
 - Complement existing systems
 - Replace Oakley with a petaflop class system
 - Keep environment consistent with existing systems
- Timeline
 - System delivered August 15, 2018
 - Full production November 2018
 - Oakley decommissioning December 2018

- Hardware at a glance

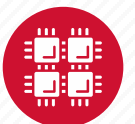
	Pitzer
Cost	\$3.35 million
Theoretical Perf.	~1300 TF
Nodes	260
CPU Cores	10560
RAM	~ 70.6 TB
GPUs	64 NVIDIA Volta V100



Pitzer Detailed Specifications

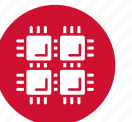
	Standard Compute Node	GPU Compute Node	Huge Memory Compute Node
Number of nodes	224	32	4
CPUs per node/Cores per node	2/40	2/40	4/80
Processor,	Intel Xeon Gold 6148	Intel Xeon Gold 6148	Intel Xeon Gold 6148
Memory (GB)	192	384	3072
GPUs	0	2 NVIDIA V100s, 16GB per GPU	0
High Speed Interconnect	Mellanox IB EDR 100Gb ConnectX-5	Mellanox IB EDR 100Gb Socket Direct ConnectX-5	Mellanox IB EDR 100Gb ConnectX-5
Internal Disk	1TB hard drive	1TB hard drive	1TB hard drive
Cooling	Liquid direct to chip	Liquid direct to chip	Air

Four login nodes, similar to standard compute node but with 382GB memory

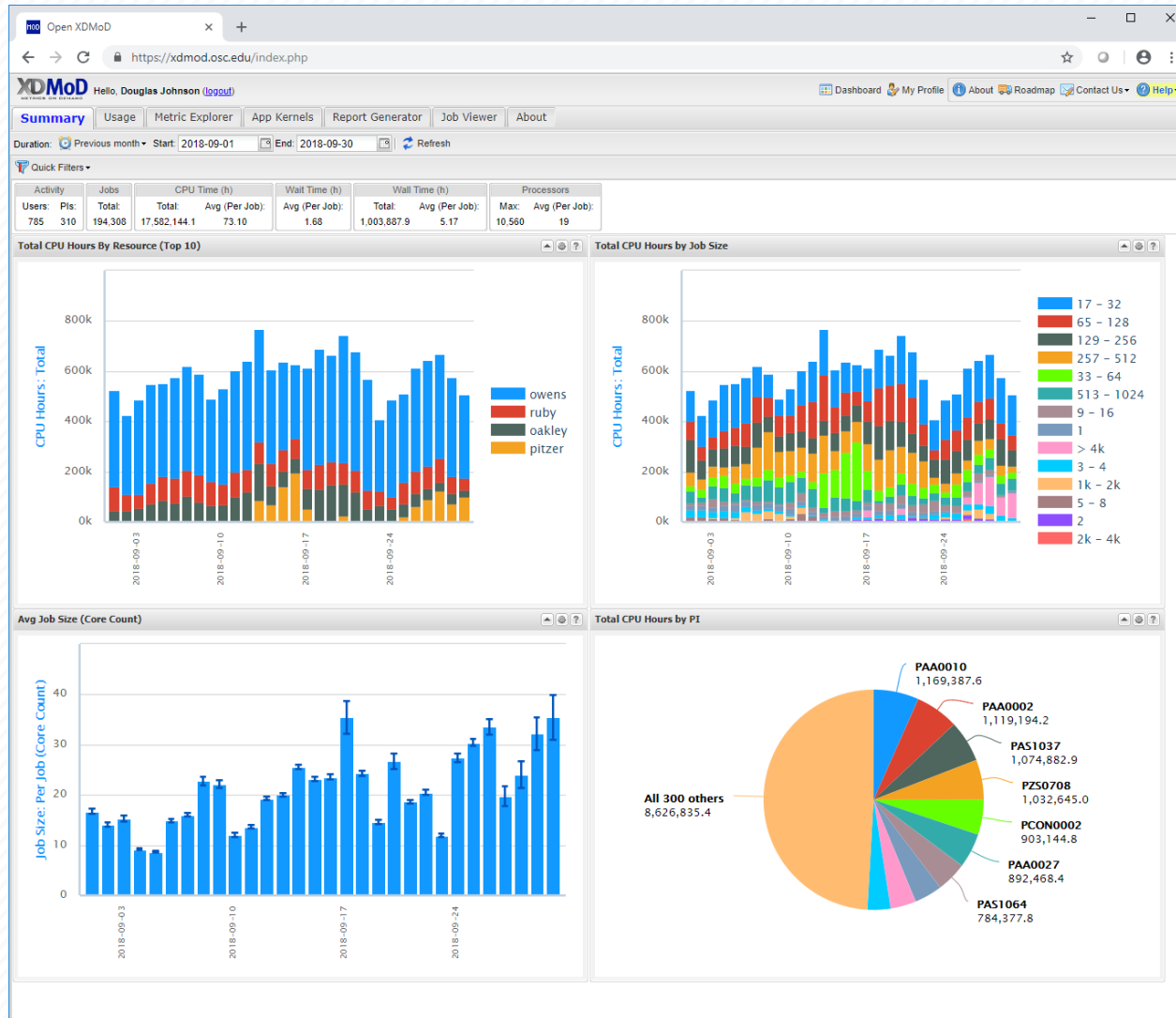


Pitzer Key Hardware Differences

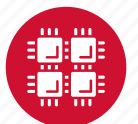
- Latest generation InfiniBand
 - ConnectX-5 cards, Switch-IB 2 switches
 - Same throughput as Owens, 100Gb/s
 - 33% higher message rate relative to Owens IB
 - Tag matching (asynchronous/one-sided optimization)
 - Full support for GDRCopy on GPU nodes
 - SHARP collective optimizations
 - SHIELD/adaptive routing
- Warm water, direct to the chip cooling
 - More consistent performance
 - More efficient, no refrigeration cycle



Pitzer Environment Differences



- Open XDMoD and SUPReMM
 - New job-level statistics
 - <https://open.xdmod.org>
- TACC XALT
 - Better software and library usage statistics
- DDN IME
 - Fast SSD cache for /fs/scratch
 - ~100GB/s
- More in the subsequent slides





Software environment on Pitzer

Heechang Na, Scientific Applications Group

- Overview
- Containers on Pitzer and Owens

Overview

- **Module system:**

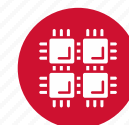
- Default modules: module load modules/au2018
- Additional software: module load <package>
 - Module load fftw3/3.3.8
- module list, module spider, module avail

- **Compilers:**

- Intel 17.0.7 and 18.0.3: -xHost
- gnu 4.8.5, 7.3.0, and 8.1.0: -march=native
- pgi 18.4: -fast

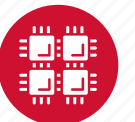
- **MPI:**

- mvapich2 2.3 for all three compilers
- Intelmpi 2017.4 and 2018.3
- openmpi 1.10.7 and 3.1.0-hpcx for intel and gnu compiler



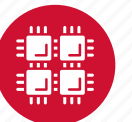
Overview

- **OpenMP:**
 - Intel, gnu, and pgi comilers support OpenMP directives
- **GPU:**
 - 64 Nvidia V100 GPUs available (2 GPUs per a node)
 - cuda 9.0.176 and 9.2.88
- **Performance and debug tools:**
 - ARM DDT, MAP and Performance Reports
 - Intel Vtune and Intel Advisor
- Pitzer environment is very similar to Owens!



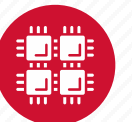
Containers on Pitzer and Owens


- **Singularity on Pitzer and Owens:**
 - module load singularity
- **Obtaining container images:**
 - **Using Hubs:** Docker Hub and Singularity Hub
 - Search for an image file from the Hub web-pages, then use “pull” sub-command to download
 - Docker: `singularity pull docker://gcc:7.2.0`
 - More images available, but possible compatibility issues with Singularity
 - Singularity: `singularity pull shub://vsoch/hello-world`
 - Better compatibility, but fewer images
 - **Other method:** copy from other computer, create from the scratch
 - Create from your local machine with root privileges
 - Create from Singularity Hub using a recipe
 - **Discussing with OSC:** `oschelp@osc.edu`



Containers on Pitzer and Owens

- **Running a container:**
 - **“shell”, “run”, and “exec” sub-command**
 - ex) singularity shell vsoch-hello-world-master.img
 - More examples in https://www.osc.edu/resources/available_software/software_list/singularity
 - **File system access**
 - Most of OSC file systems are available automatically for containers
 - Home directory
 - Working directory
 - /fs/project
 - /fs/scratch
 - /tmp
 - PFSDIR and TMPDIR (See note in https://www.osc.edu/resources/getting_started/howto)
 - **GPU usage**
 - If you have a GPU-enabled container, use the singularity flag “--nv” while running on a GPU node.



A photograph of a server room. In the foreground, there are several rows of server racks with perforated metal doors. Some racks have green indicator lights. To the left, a large black cabinet is open, revealing internal components like pipes and wiring. The floor is made of grey tiles with a pattern of small holes. A black bag is on the floor in the lower left corner. The text "Pitzer Scheduling Policies, Early Users, and Oakley" is overlaid in large blue font on the center of the image.

Pitzer Scheduling Policies, Early Users, and Oakley

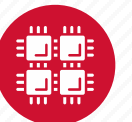
Summer Wang, HPC Client Services

Comparisons of scheduling policies

	Pitzer	Owens	Ruby	Oakley
Access	All	All	Restricted	All
Node sharing	Allow	Allow	Not allowed	Allow
Memory limit per core	TBD (Available: 4.6 GB/core)	4 GB/core	n/a	4 GB/core
GPU per node	2 gpus	1 gpu	1 gpu (Compute) 2 gpus (Debug)	2 gpus
Condo reservations	Yes	Yes	Yes	Yes
Walltime limit	Will be very similar to Oakley	See our webpages for the current policy		
Job limit	TBD due to account consolidation	See our webpages for the current policy		

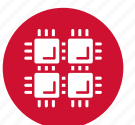
New options: Preemption and Hyperthreading

- In the development and under testing
- Preemption
 - Running jobs (preemtees requesting the new QOS of **preemptible**) can be interrupted by other jobs that are waiting on the same resources (preemptors).
 - The preemptible jobs will be **killed** to allow the preemptors to start. Preemption will be done with a **minimum runtime**.
 - Preemptible jobs will be charged a lower rate to balance the additional complexity to manage this type of workload.
- Hyperthreading
 - Some applications can benefit from hyperthreading, particularly those which are integer- or branching-bound



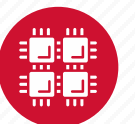
Early user program

- Application deadline: Oct 5 (Tomorrow!)
- Early user program starts on Oct 19 (tentative)
- See our webpage for more information.



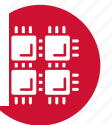
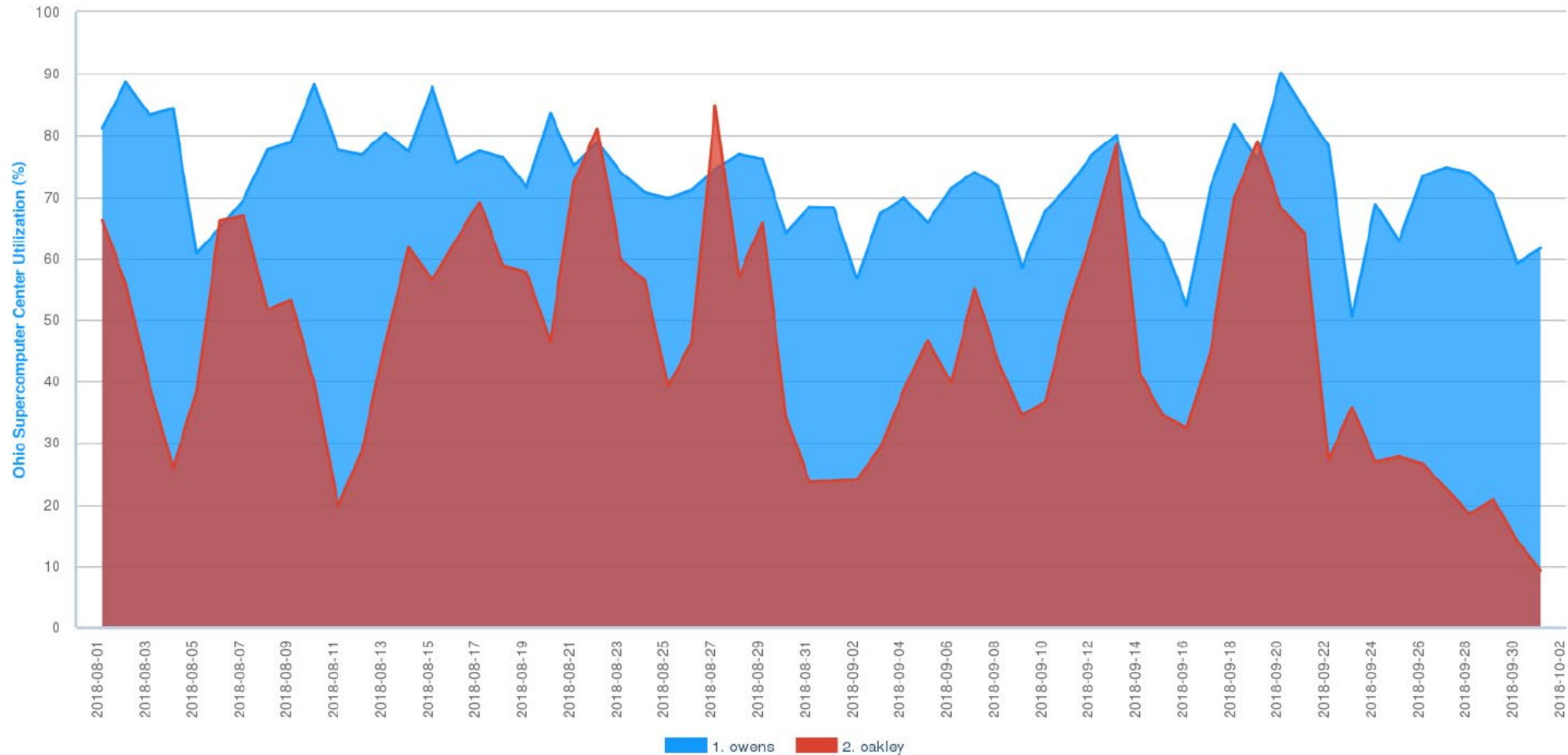
Oakley decommissioning

- By the end of 2018, after full deployment of Pitzer
- If you use Oakley only, please start to test your job on Owens
 - Hardware
 - Software
 - Scheduling impacts
 - Accounting
- It is recommended to run jobs on Oakley if possible
 - Utilization on Oakley has been low
 - This comes out to be a 50% discount with 1 RU allowing for 20 CPU hours on Oakley while 1 RU on Owens will be the standard of 10 CPU hours
 - Contact **OSC Help** if your jobs are queued due to batch limit



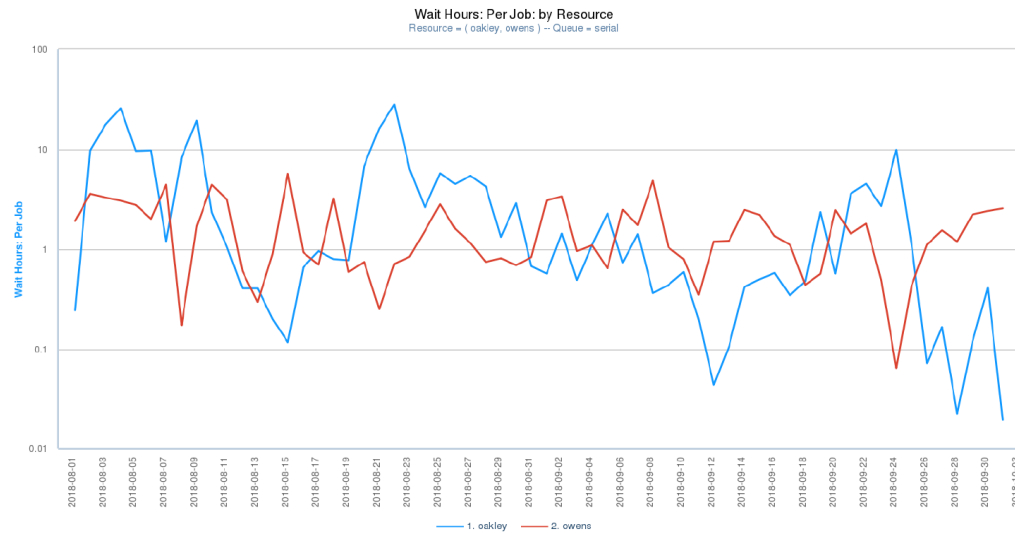
Resource utilization (Oakley and Owens)

Ohio Supercomputer Center Utilization (%): by Resource
Resource = (oakley, owens)

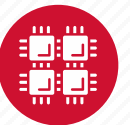
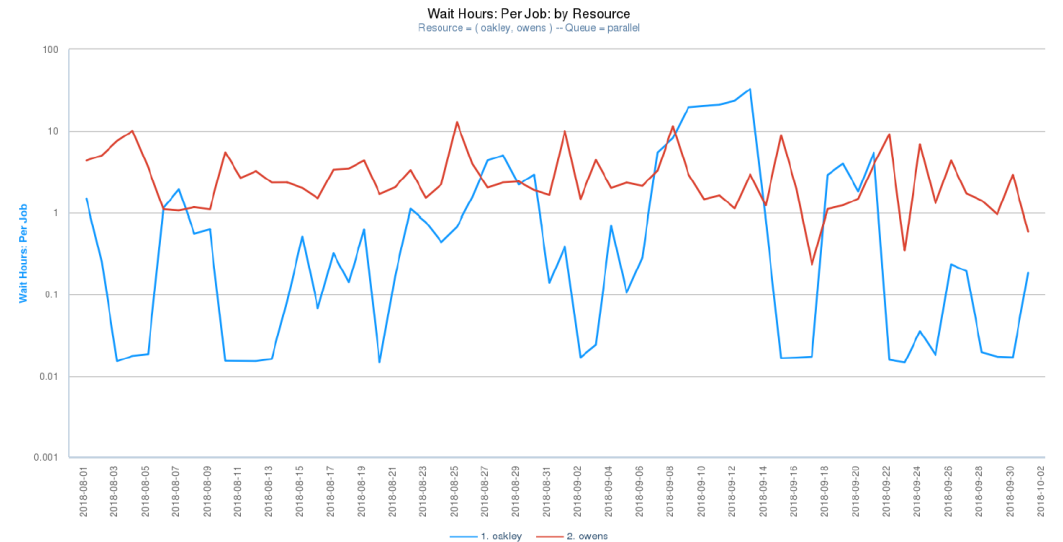


Average wait hours (Oakley and Owens)

Serial



Parallel



Discussion / Q&A

