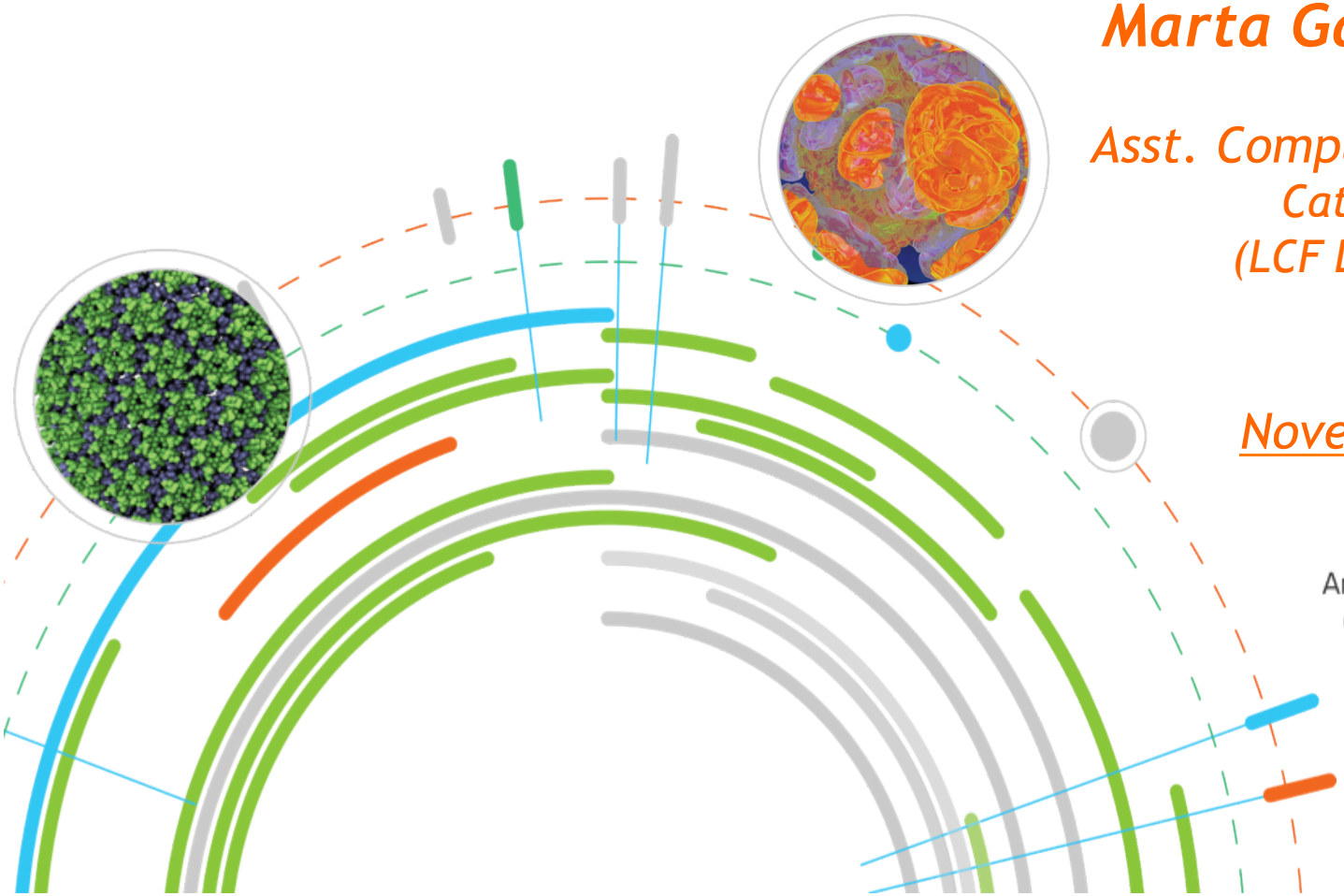
VERIFI Workshop:
Hands-on Session (step-by-step)

Marta García Martínez

Asst. Computational Scientist
Catalyst Team
(LCF Division, ANL)

November 13th, 2014



Argonne Leadership Computing Facility

Argonne NATIONAL LABORATORY

Outline

- **Part I:**
  - Hands-on session (step-by-step)
    - Objective
    - Overview of Mira (IBM Blue Gene/Q)
    - Mira job submission
- **Part II**
  - GCI concept, science behind it
- **Part III**
  - Visualization of piston engine simulations with ParaView
- **Part IV**
  - Scaling up CONVERGE™: ALCF code improvements

Objective:

Run CONVERGE™ CFD software at scale on Mira

Section:

Overview of Mira (IBM Blue Gene/Q)

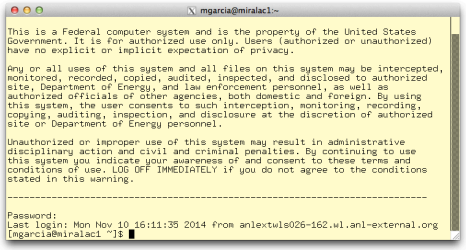
What do we need to achieve this goal?

user

1)

machine
(Wi-Fi)

2)



Terminal / X11 / Putty

3)

CRYPTOCard or token

ALCF

1)

The supercomputer

http://www.alcf.anl.gov/mira

2) **The program:**

- Director's Discretionary (DD)
  VERIFI_Workshop
- Exp. Date: 2015-01-01
- Allocation: 2,000,000 Core-Hours

http://www.alcf.anl.gov/getting-started/how-get-allocation

3) **The ALCF account:**

[@miralac1 ~]$ echo $USER

http://www.alcf.anl.gov/getting-started/new-user-guide

June 2014 | TOP500 Supercomputer Sites

RANK	SITE	SYSTEM	CORES	RMAX (TFLOP/S)	RPEAK (TFLOP/S)	POWER (KW)
1	National Super Computer Center in Guangzhou China	**Tianhe-2 (MilkyWay-2)** - TH-IVB-FEP Cluster, Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P NUDT	3,120,000	33,862.7	54,902.4	17,808
2	DOE/SC/Oak Ridge National Laboratory United States	**Titan** - Cray XK7 , Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x Cray Inc.	560,640	17,590.0	27,112.5	8,209
3	DOE/NNSA/LLNL United States	**Sequoia** - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom IBM	1,572,864	17,173.2	20,132.7	7,890
4	RIKEN Advanced Institute for Computational Science (AICS) Japan					
5	DOE/SC/Argonne National Laboratory United States	**Mira** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM	786,432	8,586.6	10,066.3	3,945

10 Petaflops =
10 quadrillion FLoating-point Operations Per Second =
10,000,000,000,000,000 FLoating-point Operations Per Second

Top #5

http://www.top500.org

ALCF Resources

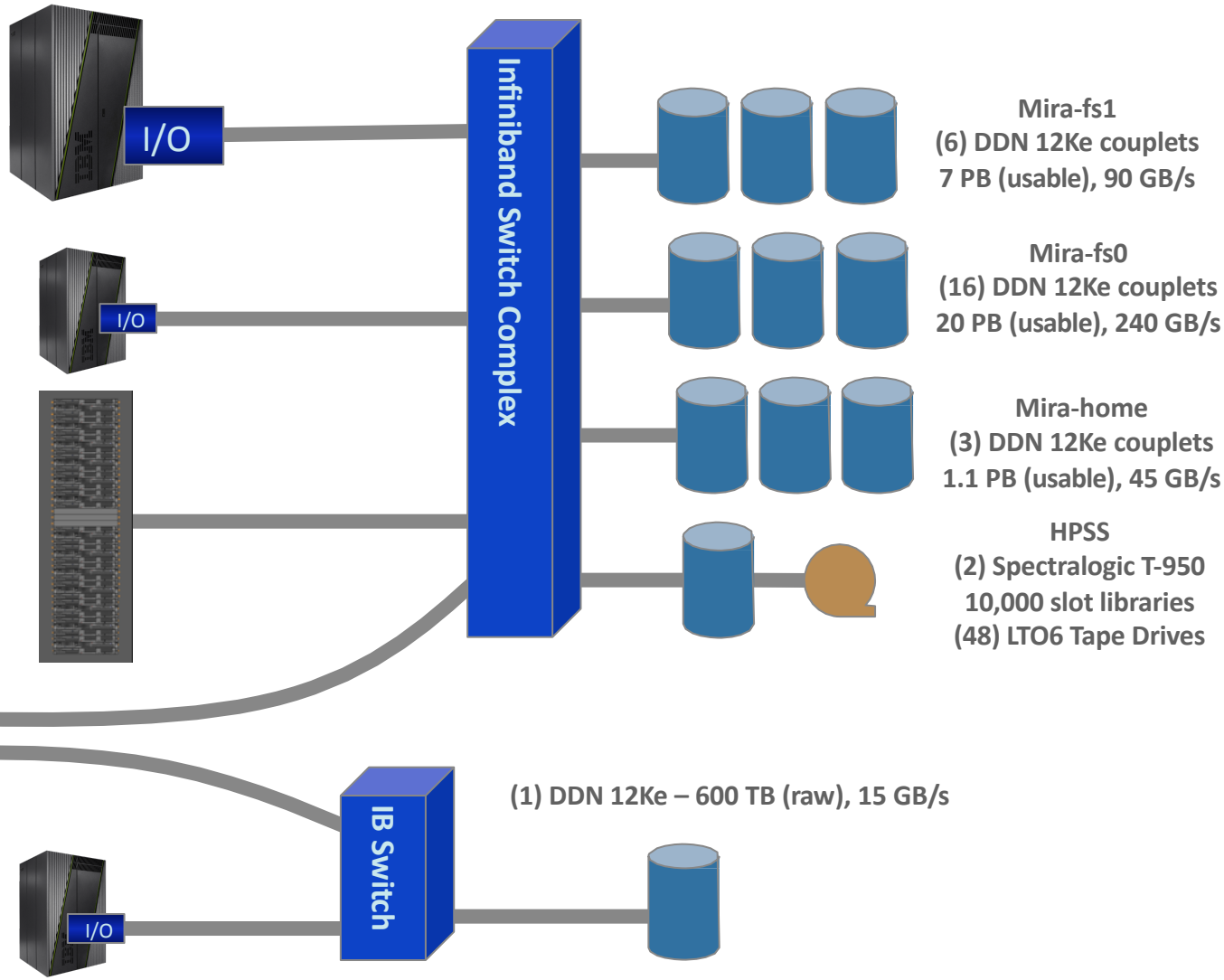
Mira
48 racks/768K cores
768 TB RAM
10 PF

Cetus (Dev)
4 rack/64K cores
64 TB RAM
838 TF

Tukey (Viz)
96 nodes/1536 cores
192 NVIDIA GPUs
6.1TB / 1.1TB GPU
RAM 220 TF

Networks – 100Gb
(via ESnet, internet2
UltraScienceNet,)

Vesta (Dev)
2 racks/32K cores
32TB RAM
419 TF

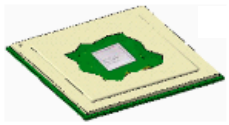


Argonne Leadership Computing Facility

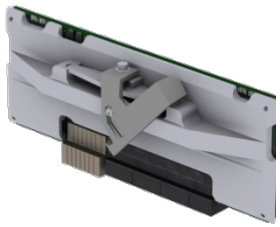
Blue Gene/Q

Chip
16+2 cores

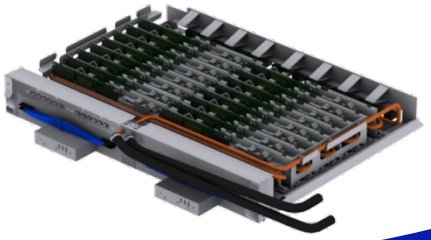
Module
Single chip



Compute card
One single chip module
16 GB DDR3 Memory
Heat Spreader for H_2O Cooling



Node board
32 compute cards,
optical Modules,
link chips; 5D Torus

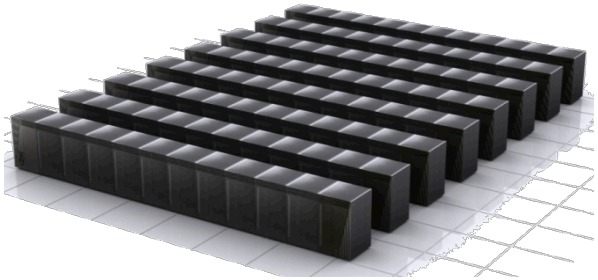


I/O drawer
8 I/O cards w/16 GB
8 PCIe Gen2 x8 slots
3D I/O Torus

Midplane
16 node boards

Rack
1 or 2 midplanes
0, 1, 2, or 4 I/O drawers

Multi-rack system
Mira: 48 racks, 10 PF/s



Argonne Leadership Computing Facility

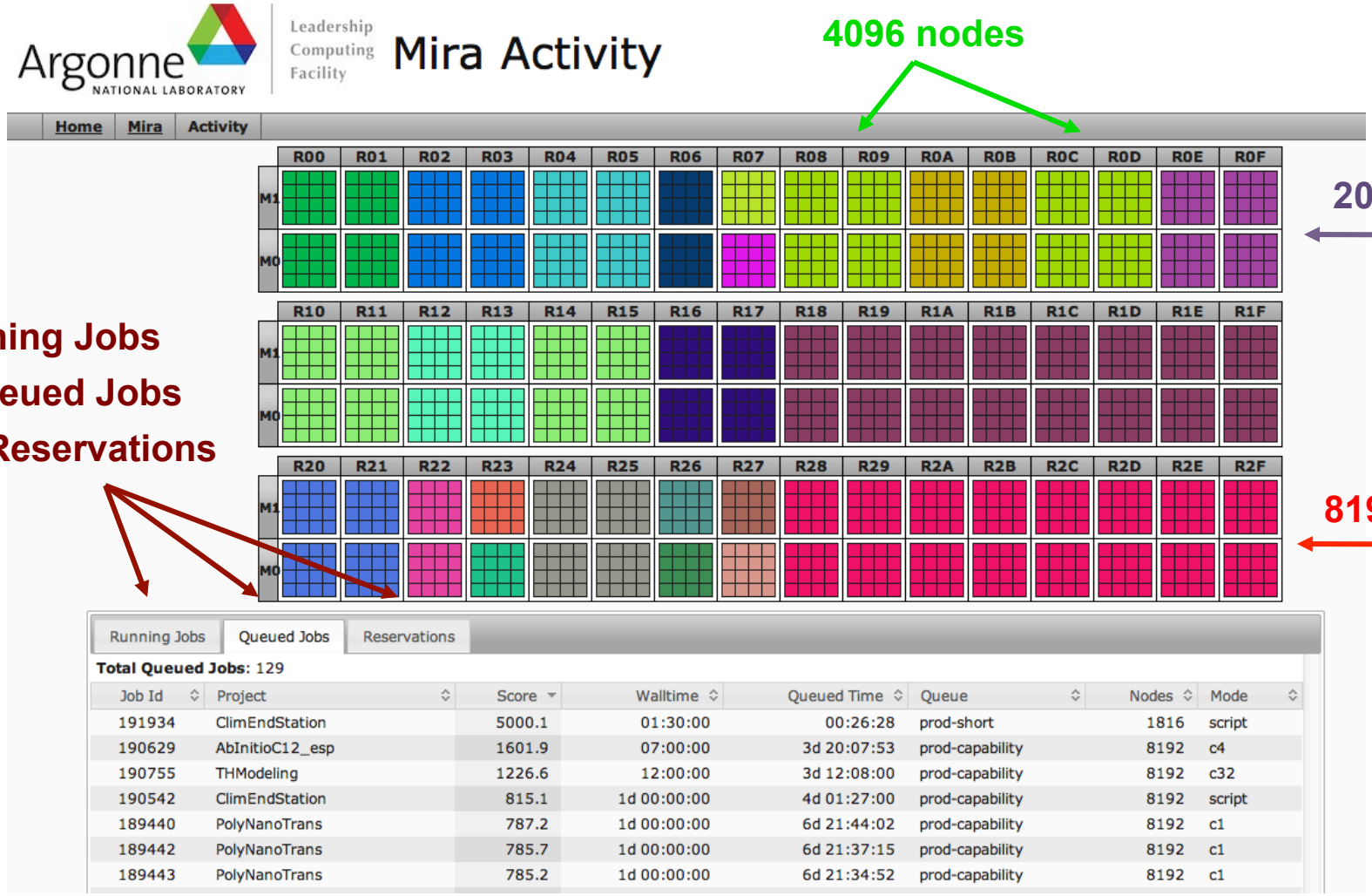
Mira: Photo Gallery

Panoramic shot of Mira

IBM Blue Gene/Q

Mira and her cables

Machine status web page



4096 nodes

2048 nodes

8192 nodes

http://status.alcf.anl.gov/mira/activity (beta, a.k.a. The Gronkulator)

Keywords

Project vs Allocation:

Projects and allocations at the ALCF are different. **A particular project might have multiple allocations of time**. For example, a discretionary project that has been approved for more than 3 times will have 3 allocations (2 are probably expired) **but just one project**. Projects will not expire, allocations will. If allocations are expired, jobs will not be able to run.

Project name: VERIFI_Workshop **Allocation: 2,000,000 Core-Hours**

ALCF account:

All computing carried out on the ALCF systems is associated with a user "account." This account is used to log onto the login servers and run jobs on the resources. **Only one account per user**.

Understanding BG/Q blocks:

It is important to keep in mind that **CPU time is charged based on partition size, not job size**. If the number of nodes required by a job is not a power of 2, the project must round up to the nearest power of 2 in order to determine the correct allocation to request (with a minimum size of 32, 128, or 512 nodes, depending on the system). If the job does not fit entirely in a partition, the unused nodes remain idle and are not available for use by anyone else; this is why we always charge for the entire partition.

For example, a 6000-node job must run on a 8192-node partition, which means the project will be charged for (8192 nodes x 16 cores) = 131,072 CPU hours. Depending on how many threads the application uses, that is anywhere from 6000 to 131,072 MPI processes.

Keywords

Mira

512 nodes = minimum partition size on Mira

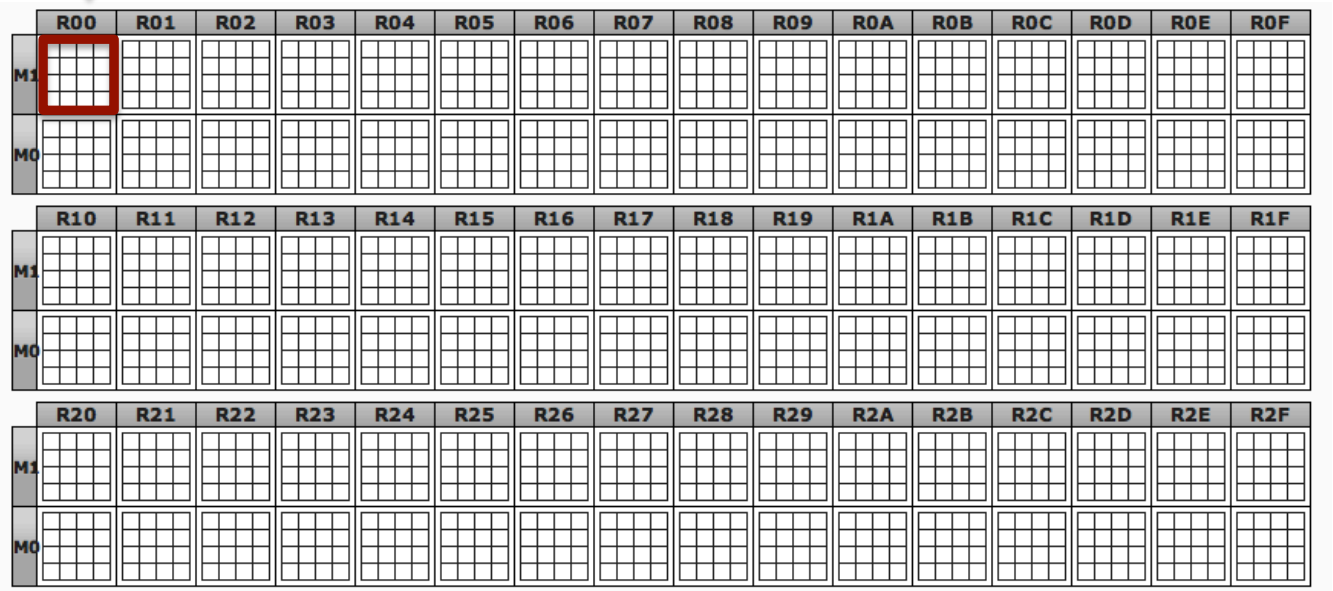
16 racks

+

16 racks

+

16 racks



Row 0

Row 1

Row 2

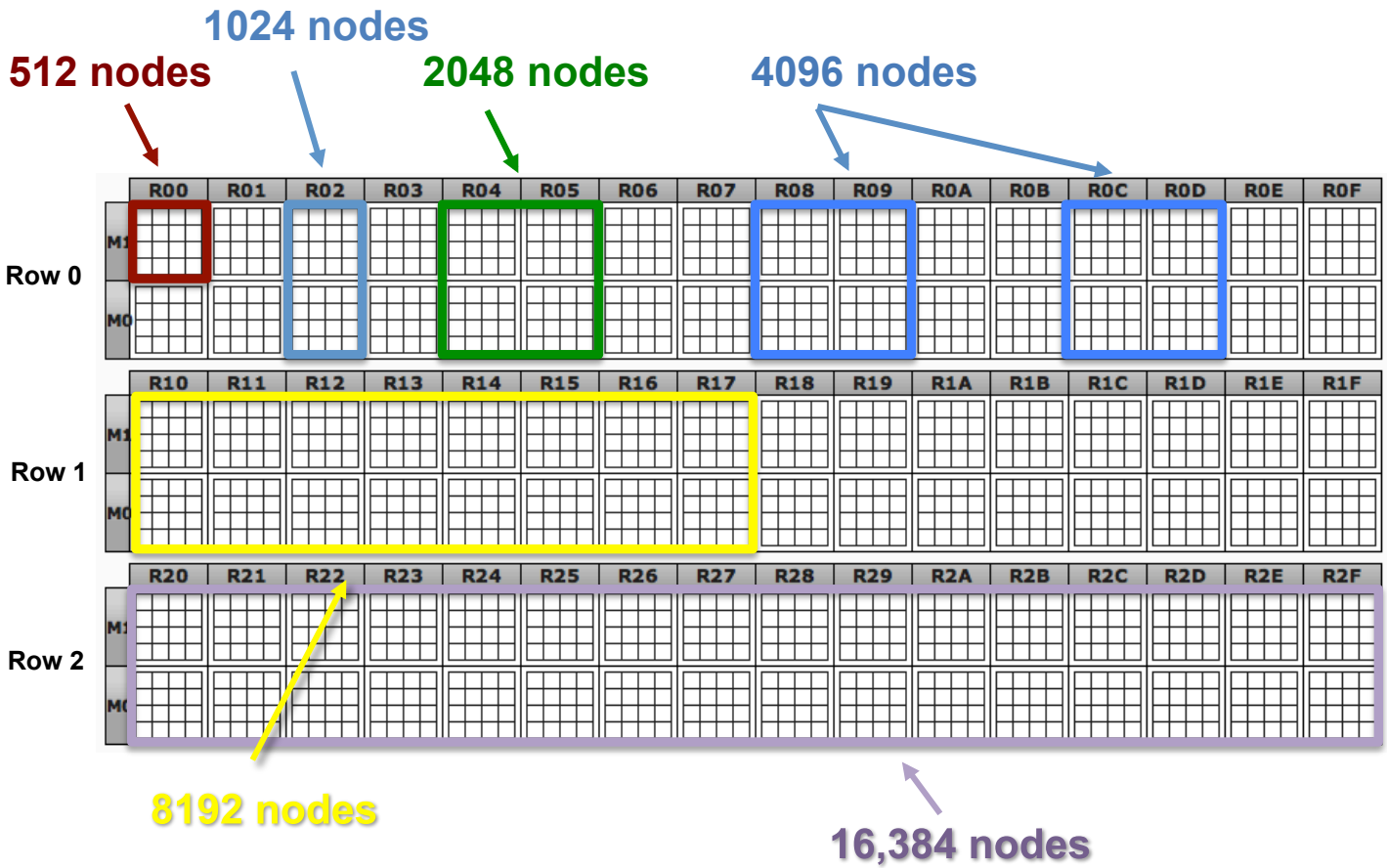
Rack
1 or 2 midplanes
0, 1, 2, or 4 I/O drawers

1 rack = 1024 nodes = 16,384 cores = 65,536 processes

Midplane
16 node boards

1 midplane = 512 nodes = 8192 cores = 32,768 processes

Mira multiple rack partitions ("blocks")

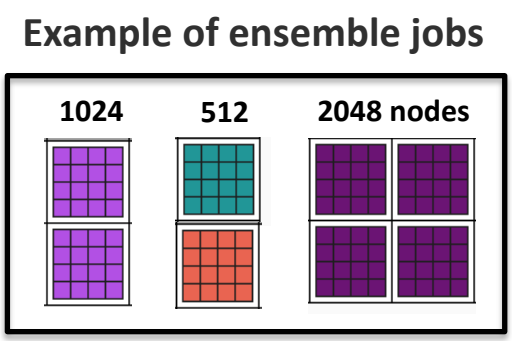


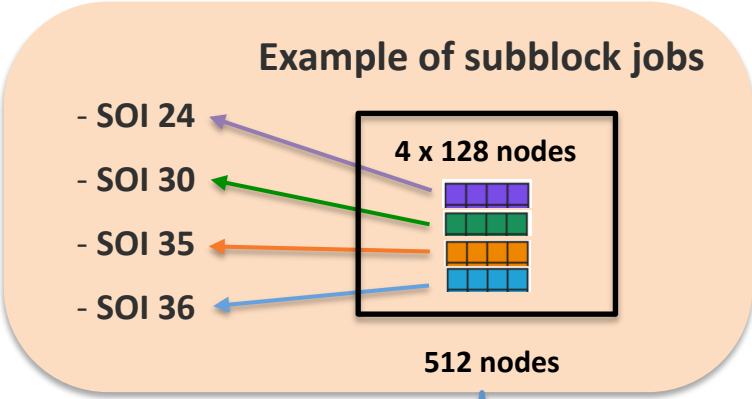
The number of large block sizes possible is:

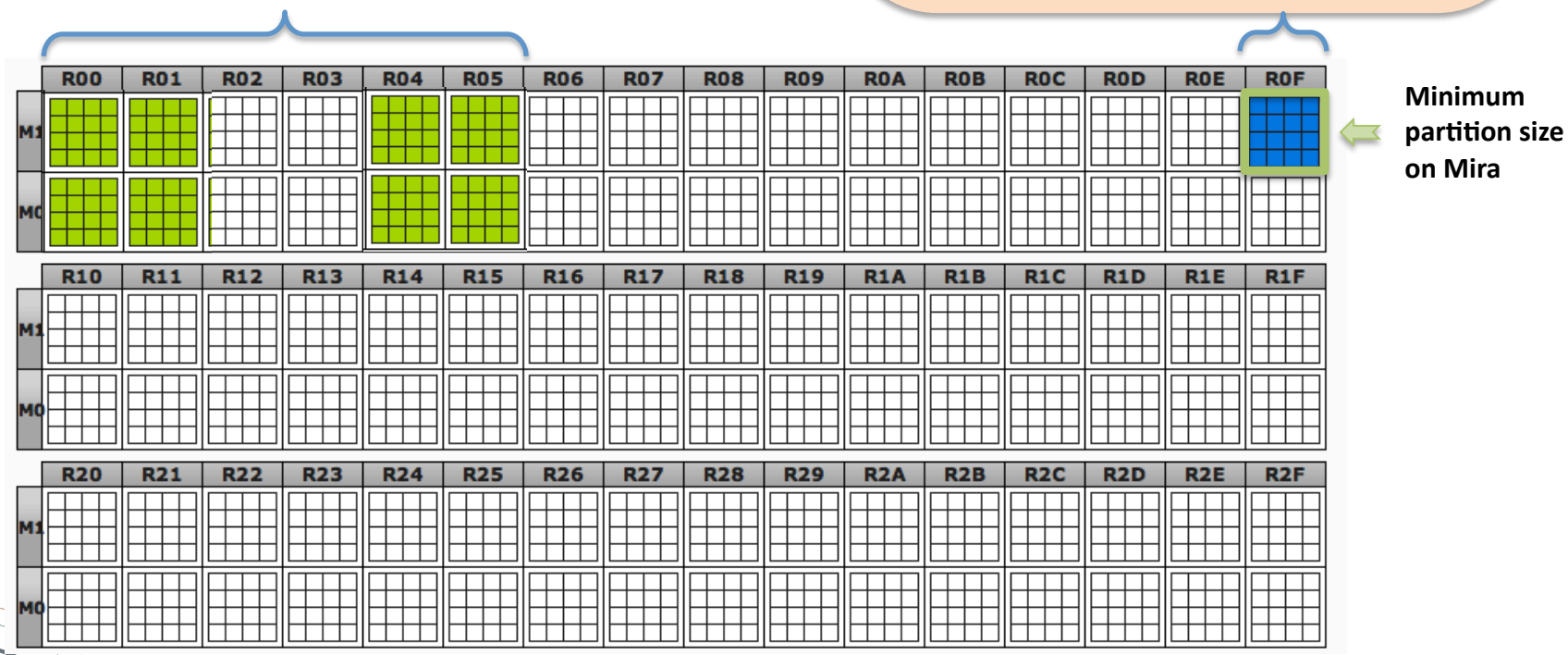
# of nodes	# of blocks
49152	1
32768	3
24576	2
16384	9
12288	12
8192	6
4096	12
2048	24
1024	64
512	96

partlist will show you if a large free block is busy due to a wiring dependency

Example of ensemble and subblock jobs







For jobs with the same characteristics: **higher job size = faster score increase**

Mira file system at ALCF

Name	Accessible from	Type	Path	Backed Up to HPSS	*Daily Snapshots	Uses
mira-home	Mira Cetus Tukey	GPFS	/home or /gpfs/mira-home	Yes	Yes	General use
mira-fs0 mira-fs1	Mira Cetus Tukey	GPFS	/projects	No	No	Intensive job output, large files

* Daily snapshots are stored for 1 week on-disk in /gpfs/mira-home/.snapshots/. These snapshots do NOT persist in the event of disk failure.

http://www.alcf.anl.gov/user-guides/bgq-file-systems

The command **myquota** will show you the quota in your home directory.

The command **myprojectsquota** will show you the current space used and quotas on any projects of which you are member of.

Section:

Mira job submission

Getting help during Hands-On session

Contact us:

ALCF staff:

- Avanthi “Lalitha” Mantrala (User Services Outreach, USO)
- Marta García Martínez (Catalyst Team, CAT)
- Kevin Harms (Performance Team, PE)
- Joe Insley (Visualization, VIZ)

E-mail: **support@alcf.anl.gov**

!! Use keyword in the title: [VERIFI] …

Argonne Leadership Computing Facility

Hands-on session

Problems: *with token?* *1st time?* *No access to the project directory?* *Other?*

- **Log into Mira**:

  > ssh *username*@mira.alcf.anl.gov

  This will lead you to your home directory (~)

- Go to the **VERIFI_Workshop project directory:**

  > cd /projects/VERIFI_Workshop

  > ls

```
Hands-On
    |----- SRC
    |          |----- converge
    |          |----- scripts
    |----- RUN
    |          |----- run_everything.sh
    |          |----- submit.sh
    |          |----- SOI
    |                     |----- 24
    |                     |----- 30
    |                     |----- 35
    |                     |----- 36
```

Hands-on session

- Create your own directory, copy the files and **get ready to submit**:

  > mkdir $USER

  > cp -r /projects/VERIFI_Workshop/Hands-On $USER

  > cd $USER/Hands-On/RUN

  > ./submit.sh

```
mgarcia@miralac1:/projects/VERIFI_Workshop
[mgarcia@miralac1 VERIFI_Workshop]$ more Hands-On/RUN/submit.sh
#!/bin/bash -x
qsub -A VERIFI_Workshop -q R.VERIFI_Workshop -n 512 -t 03:00:00 --disable_preboot --mode script ./run_everything.sh
[mgarcia@miralac1 VERIFI_Workshop]$
```

- Check if your job is in the queue and its status:

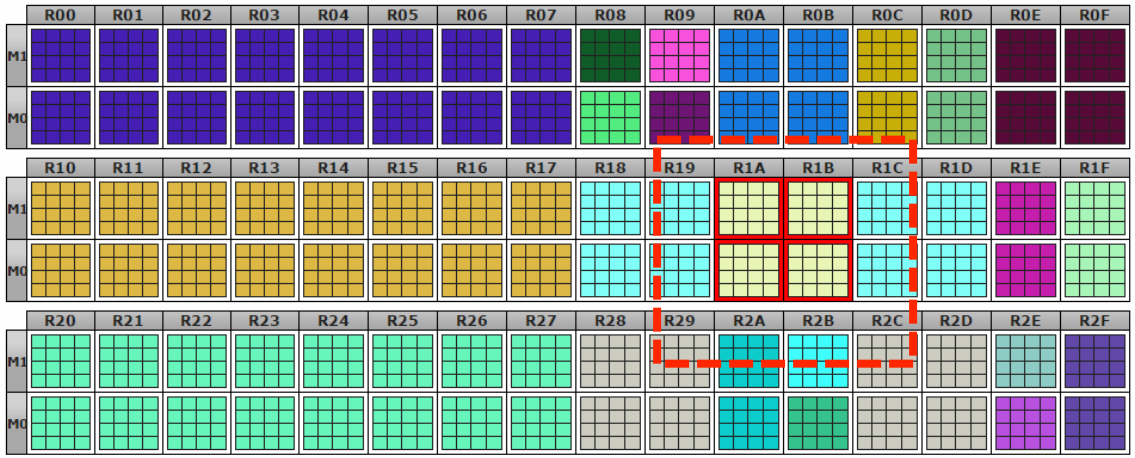
  > qstat -u $USER

Where is your job running on Mira?

Step 2:
Go to http://status.alcf.anl.gov/mira/activity

Argonne NATIONAL LABORATORY | Leadership Computing Facility | Mira Activity

Home | Mira | Activity



Step 1:
Find your Job ID by typing in terminal:
> qstat -u $USER

Step 3:
Hover mouse over your Job ID to highlight where your job is running on Mira

Running Jobs | Queued Jobs | Reservations

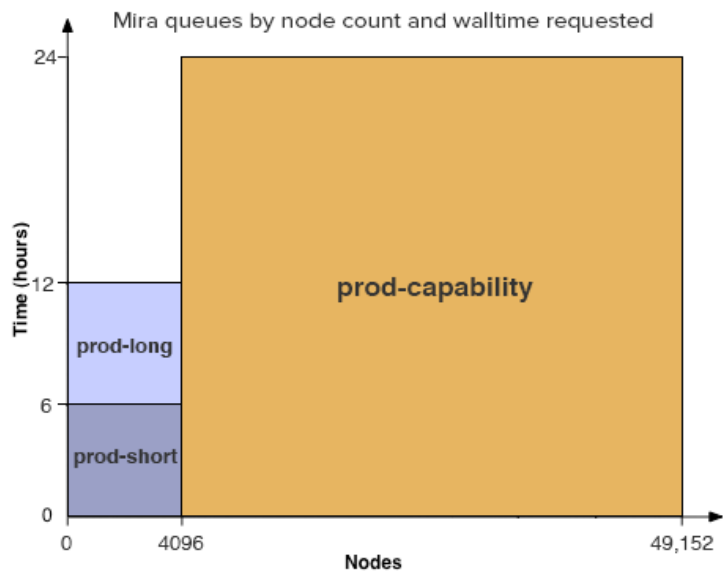
Total Running Jobs: 22

Job Id	Project	Run Time	Walltime	Location	Queue	Nodes	Mode
357918	HybridPhoto	23:18:12	1d 00:00:00	MIR-04000-37FF1-8192	prod-capability	8192	c32
348811	LiquidWater	17:25:30	1d 00:00:00	MIR-08000-3BFF1-8192	prod-capability	8192	script
354269	WindNoise	09:05:27	12:00:00	MIR-40040-73371-512	prod-long	256	script
357203	EddySim_CLF	06:58:17	12:00:00	MIR-40080-733B1-512	prod-long	512	c16
359368	HeWPlasma	05:42:10	12:00:00	MIR-40800-73BF1-2048	prod-long	2048	script
360123	DNSGeo	05:38:43	06:00:00	MIR-00000-33FF1-8192	prod-capability	8192	c32
358873	LiquidWater	05:35:55	06:00:00	MIR-44000-77FF1-4096	prod-short	4096	script
358271	HEP-ANL	05:35:24	05:50:00	MIR-44C00-77FF1-2048	prod-short	2048	c2
361312	HighLit	04:43:39	06:00:00	MIR-44800-77BF1-1-1024	prod-short	1024	script

Mira job scheduling

- Restrictions in queues
  - **prod-long:** restricted to the row 0.
  - **prod-short**, **prod-capability:** can run in the full machine

http://www.alcf.anl.gov/user-guides/job-scheduling-policy-bgq-systems



User Queued	Underlying Queue	Nodes	Wall-clock Time (hours)	Max. Running per User	Max. Queued per User
prod	prod-short	512 - 4096	0 - ≤6	5	20
	prod-long	512 - 4096	>6 - 12	5	20
	prod-capability	4097 - 49152	0 - 24	5	20
	backfill (*)	512 - 49152	0 - 6	5	20
prod-1024-torus	prod-1024-torus	1024	0 - 12	5	16
prod-32768-torus	prod-32768-torus	32768	0 - 24	1	20

(*) This queue is automatically selected based on the scheduling policy.

Getting help after the VERIFI Workshop

Online resources (24/7):

- ALCF web pages:
  - http://www.alcf.anl.gov/user-guides
  - https://accounts.alcf.anl.gov
- Mira/Cetus/Vesta/Tukey status:
  - http://status.alcf.anl.gov/{mira,cetus,vesta,tukey}/activity

Contact us:

E-mail: **support@alcf.anl.gov**

ALCF Help Desk: **Hours**: Monday – Friday, 9 a.m. – 5 p.m. (Central time)
Phone: **630-252-3111** or **866-508-9181** (toll-free, US only)

ALCF staff

News from ALCF:

E-mail via: ALCF Weekly Updates,
ALCF newsletters (monthly)
{*mira,cetus,vesta,tukey*}*-notify* lists

Argonne Leadership Computing Facility

Argonne Leadership Computing Facility
VERIFI Workshop Hands-on Session

Thank you for your attention