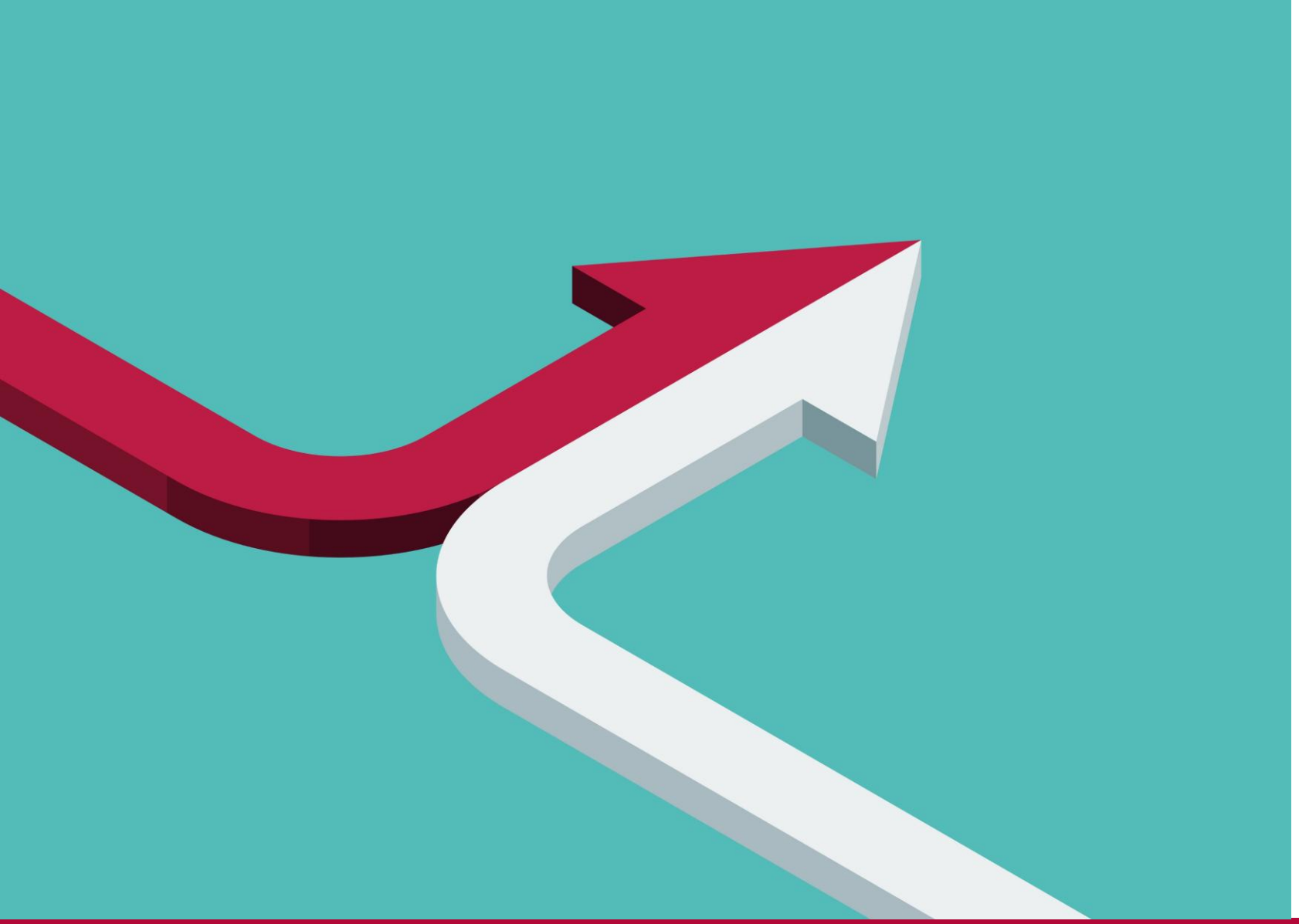


Command Lines vs. Requested Resources

How Well Do They Align?

Ben Fulton
Abhinav Thota
Scott Michael
Jefferson Davis

RESEARCH TECHNOLOGIES
INDIANA UNIVERSITY
University Information Technology Services

PERVASIVE TECHNOLOGY INSTITUTE
INDIANA UNIVERSITY

Motivation

Node capacity

	Cores/Node	RAM/Node
Big Red	4	8GB
Big Red 2	16/32	32/64GB
Big Red 200	64/128	256/512GB

Motivation

Disciplines

	IUPUI	IUB	Total
Big Red 2	68	141	159
Karst	117	164	214
Total	136	201	243

Motivation

- Job Exclusive Mode
- User Exclusive Mode
- Shared Mode

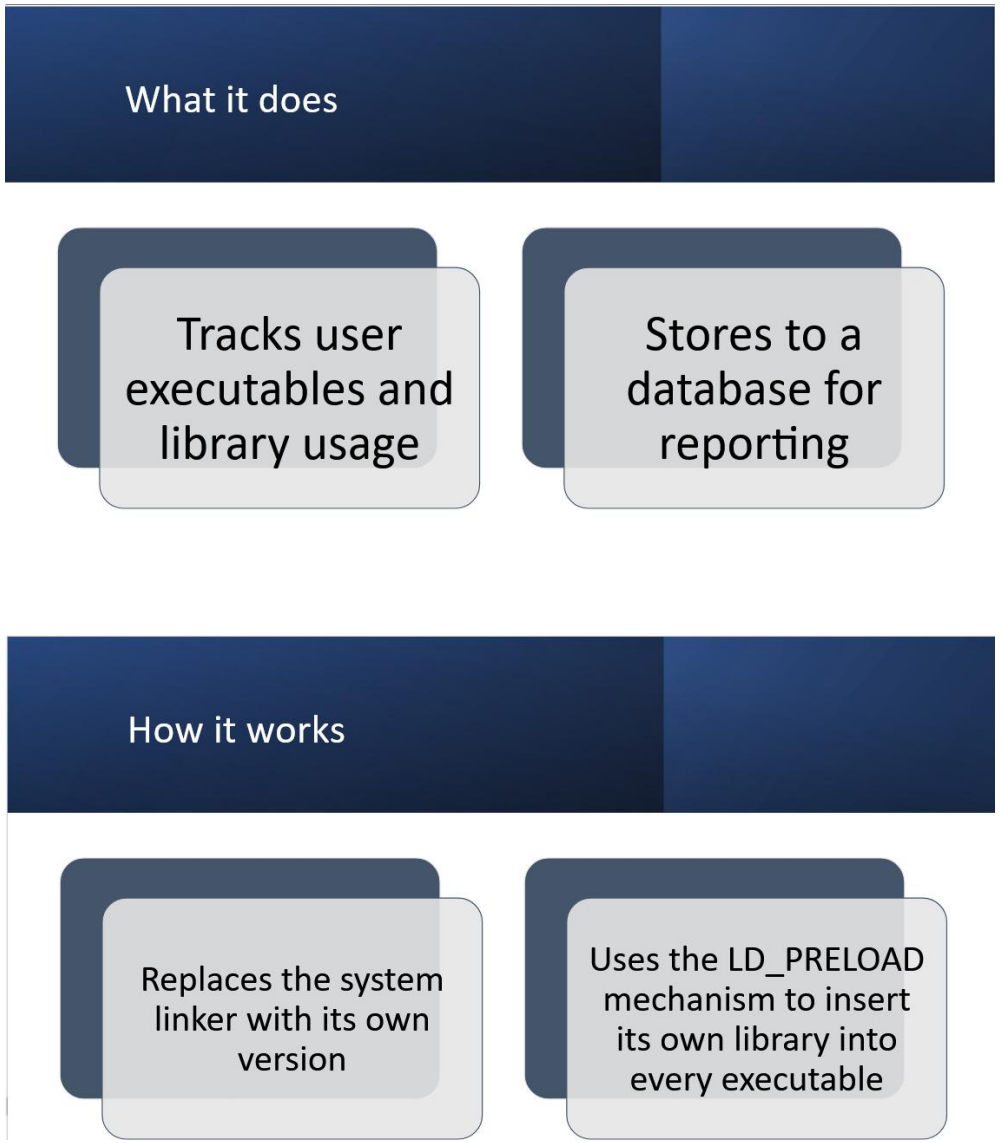
Motivation

Are users effectively using the resources they ask for?

Motivation

1. What resources are users asking for?
2. What resources are they using?

Data Collection: XALT



Data Collection: XALT

What it tells us	
Executable Path	`/geode2/home/u030/befulton/Quartz/bio/apps/bin/samtools`
Command Line	`samtools view SRR15447420_b4_wim.bam`
Slurm Job ID	`3380668`
Environment Variables	`MKLROOT=/geode2/soft/hps/rhel8/intel/22.3/mkl/2022.2.0`
Python and R Modules	`(None)`

Data Collection: XALT

Sampling	
Duration	Probability
< 5 minutes	`0.0001`
5-10 minutes	`0.01`
> 10 minutes	`1`

Application Suites	
App	Code
Rosetta	`Rosetta*`
rmpisnow	`Rosetta*`
Blastp	`NCBI-Blast*`

Data Collection: Slurm

Requested Resources

- Memory
- Tasks
- GPUs and CPUs per task

Job Scripts:

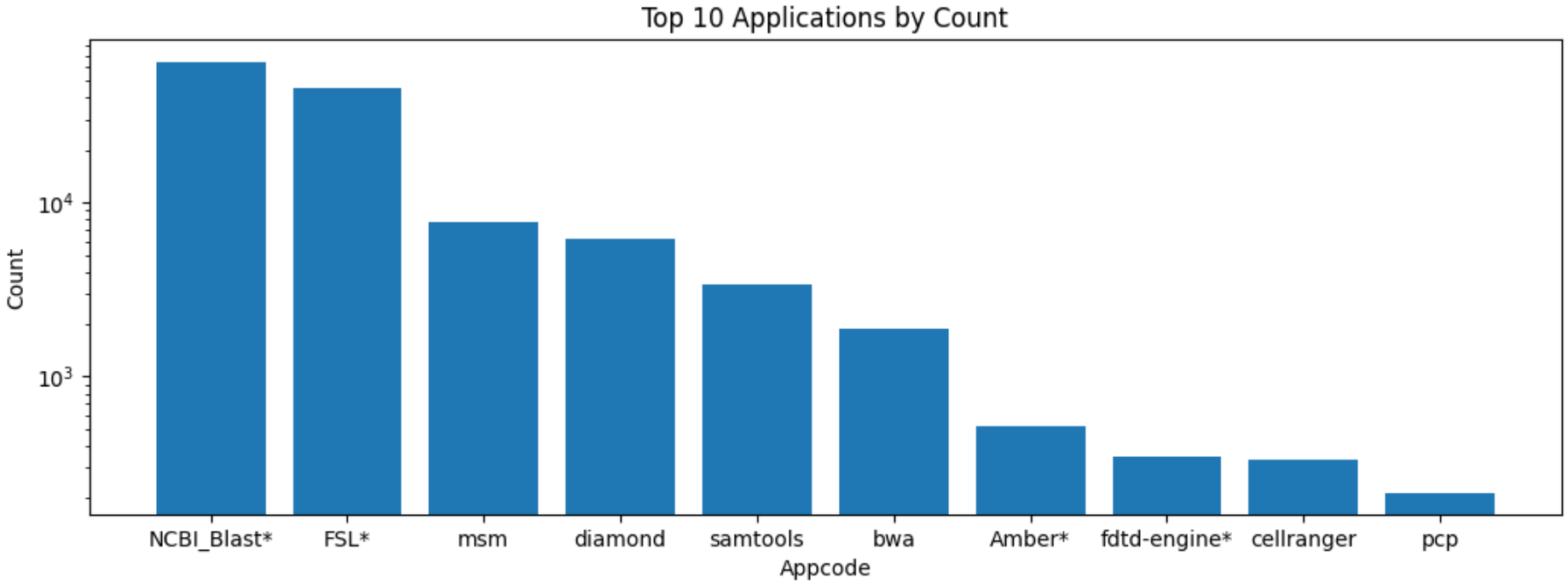
- srun calls
- OMP_NUM_THREADS settings

Collected Data Size by Cluster

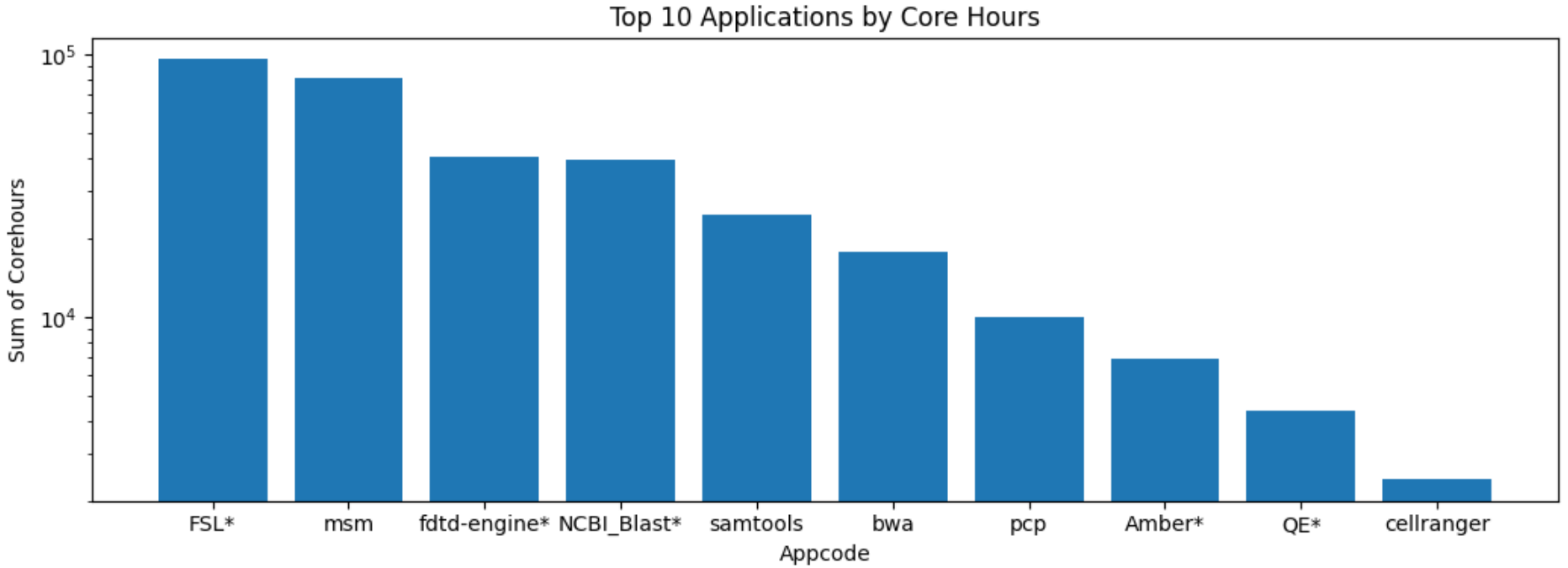
- Pulled the top 100 applications/application suites
- These included applications such as a.out
- Whittled the list down to 40 applications we thought we could get interesting data from

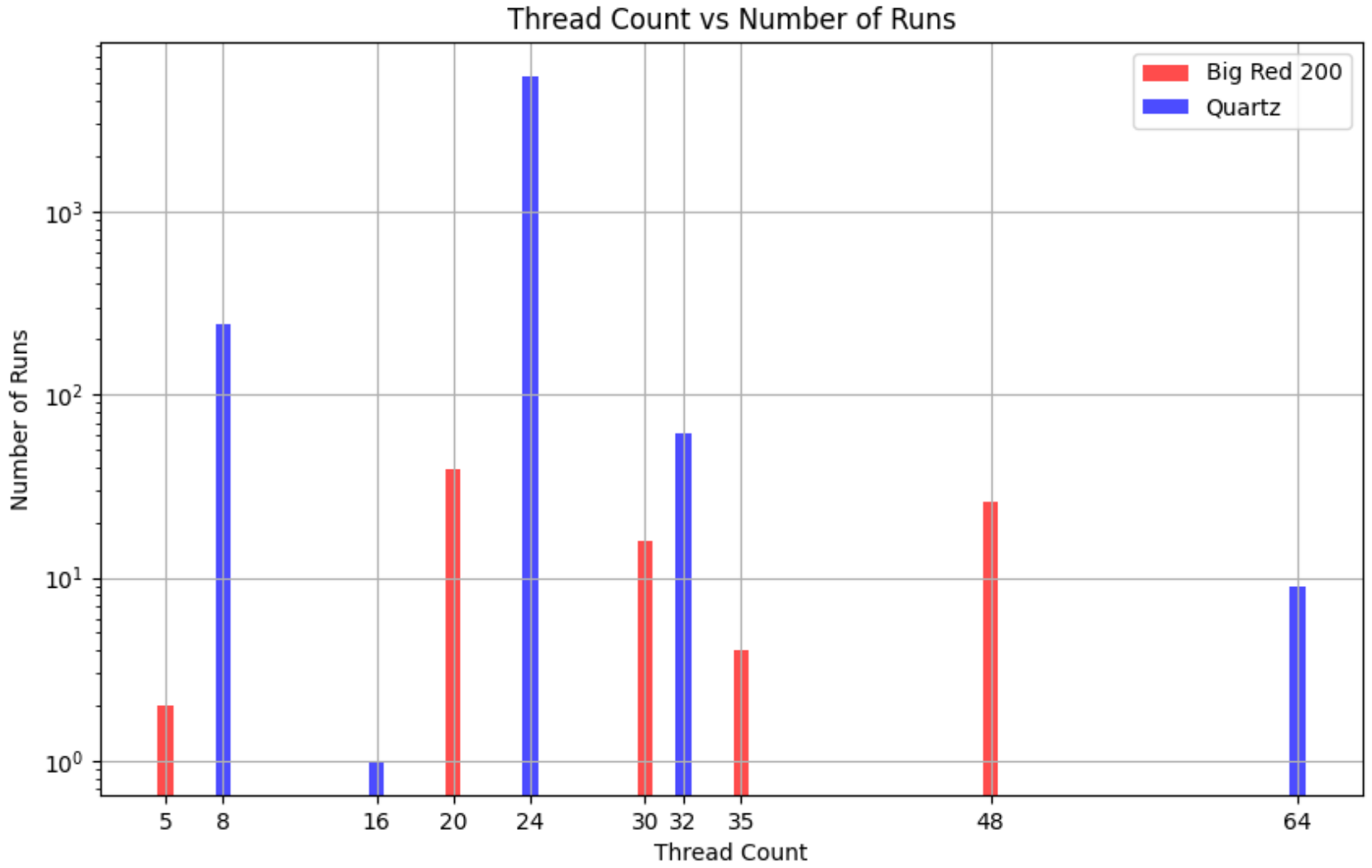
	Quartz	Big Red 200
Command Lines	1,374,246	1,072,780
Job Requests	478,886	521,179
OMP_NUM_THREADS	159,071	19,167
srun	164,245	55,556

Quartz Top Applications

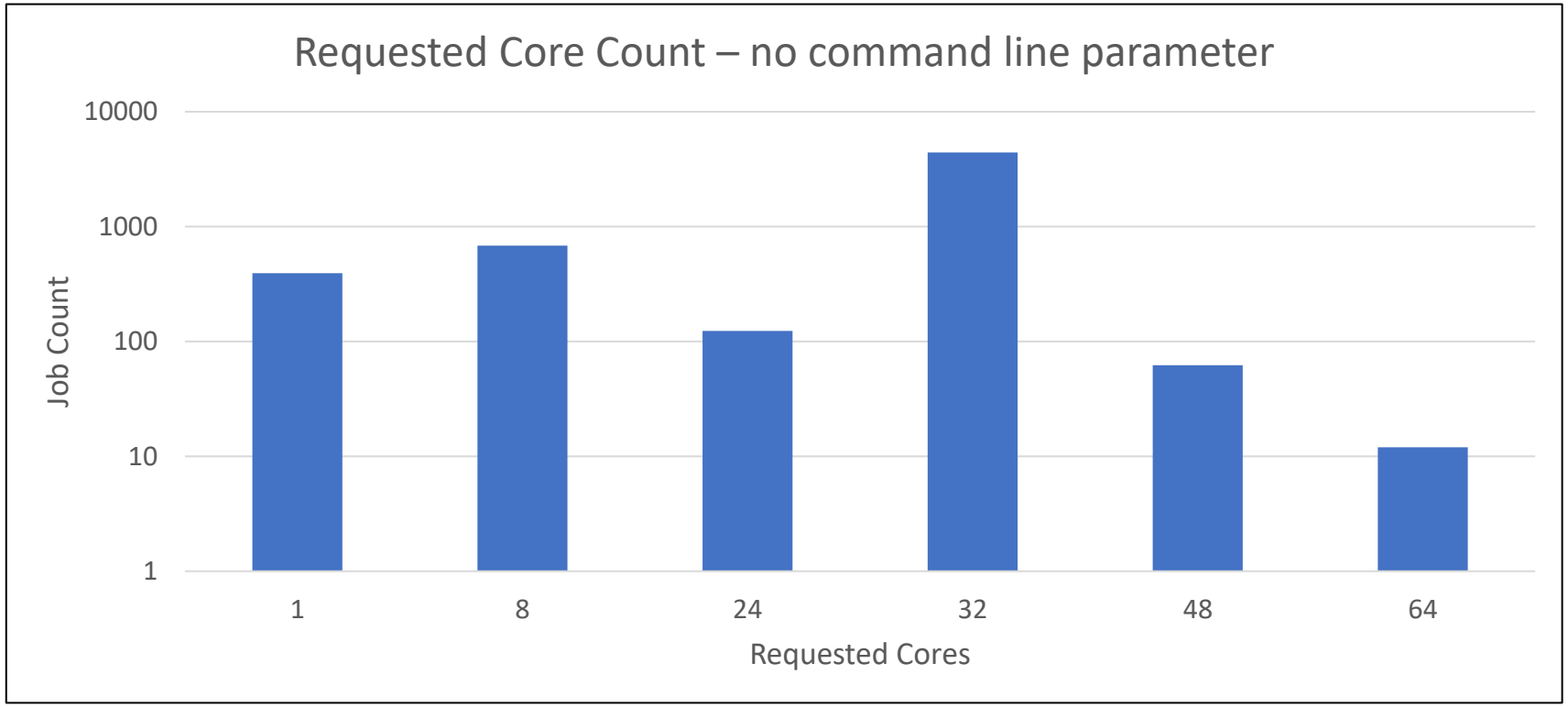


Quartz Top Applications




Thread Count vs Number of Runs
Big Red 200
Quartz
Number of Runs
10^3
10^2
10^1
10^0
5
8
16
20
24
30
32
35
48
64
Thread Count

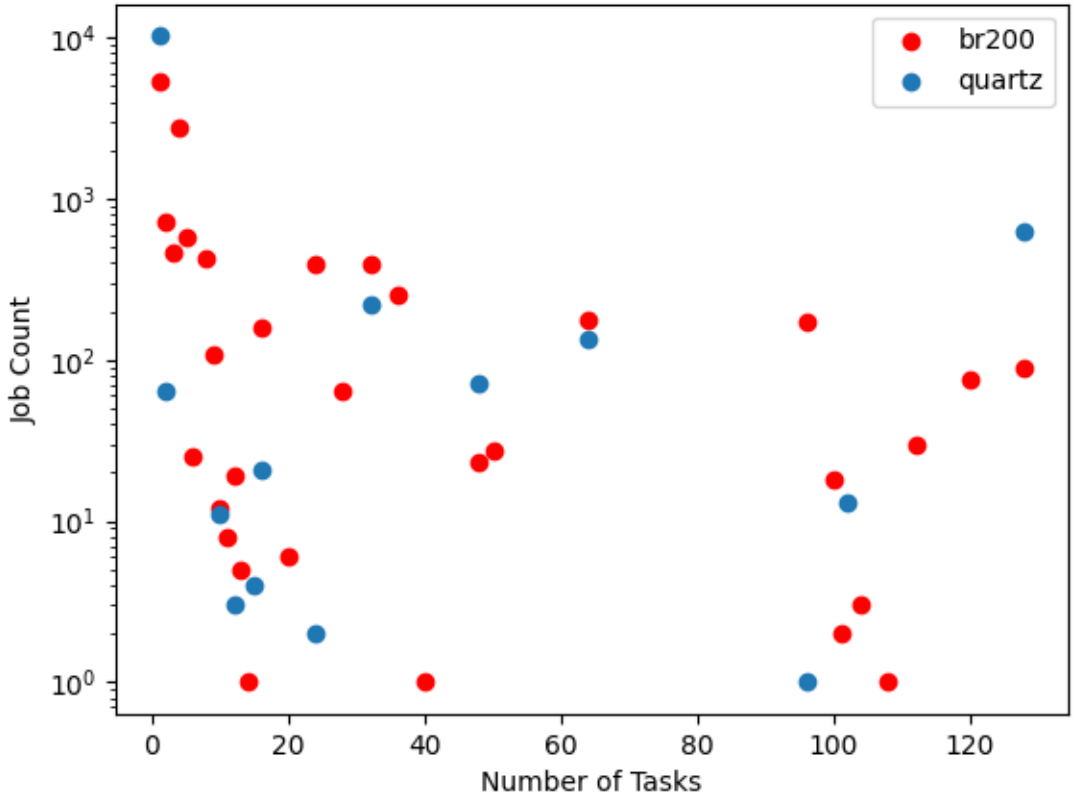
Requested Cores



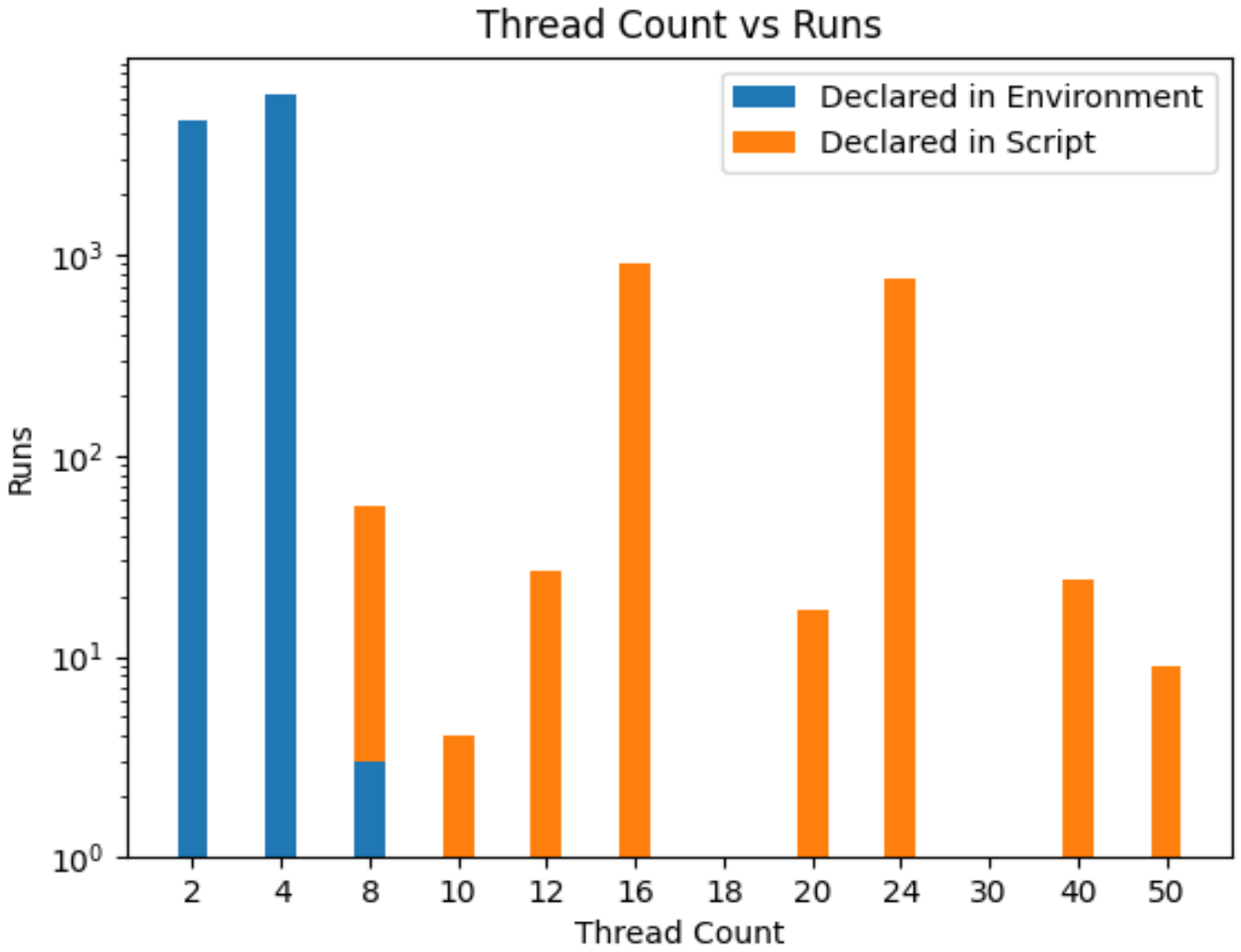
Requested Tasks

	Quartz	Big Red 200
Numeric Task Count	4,284	870
Non-Numeric Task Count	16	434
No Task Count	11,413	12,410
Matching	4,240	814

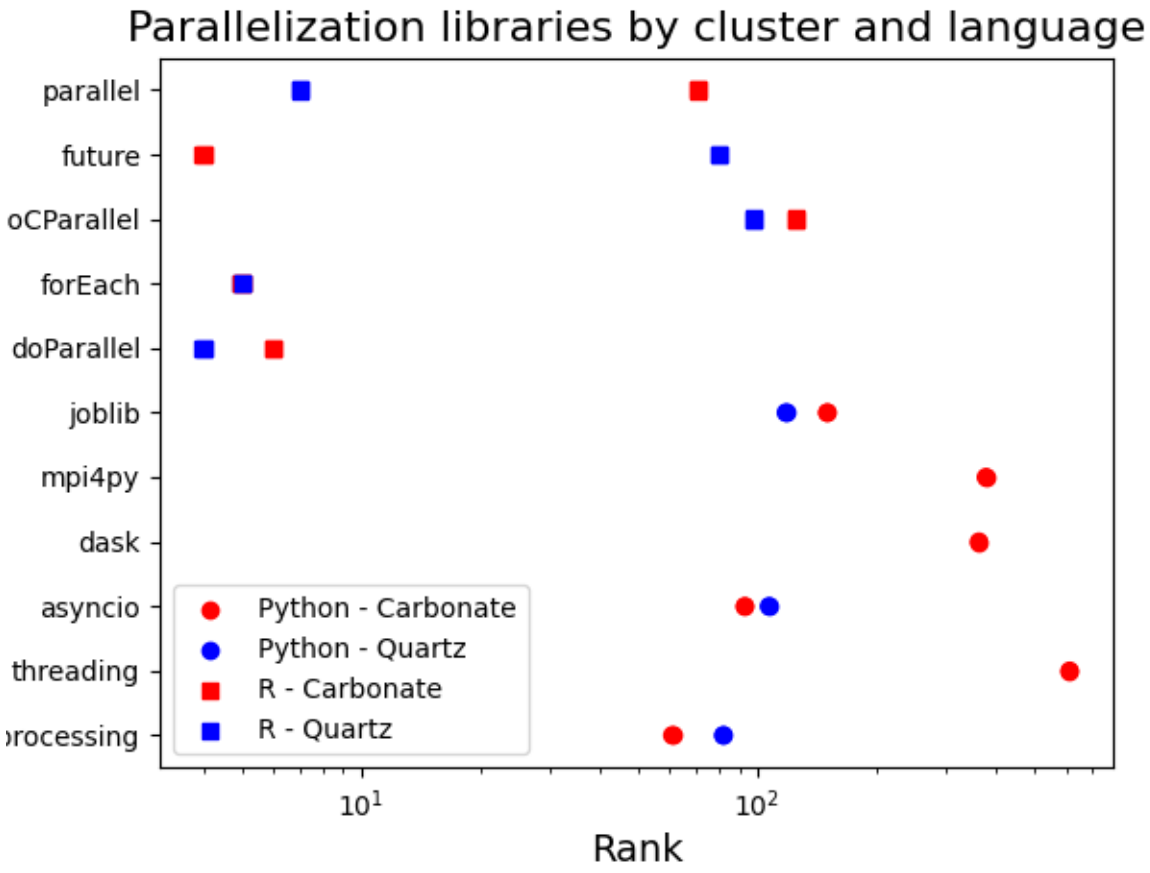
Requested tasks for jobs with no task specification in an srun call



OMP_NUM_THREADS



Scripting


Parallelization libraries by cluster and language
parallel
future
oCParallel
forEach
doParallel
joblib
mpi4py
dask
asyncio
threading
rocessing
Python - Carbonate
Python - Quartz
R - Carbonate
R - Quartz
10^1
10^2
Rank

Conclusions and Future Work

- There is a divide between “expert” users and other users
- This divide can be detected by examining commands

- Ways to improve user behavior
- Scripting
- AI/Deep Learning applications
- GPU’s
- Applying AI to this data