Science domains in NeSI cluster job submissions - first results

Gene Soudlenkov
g.soudlenkov@auckland.ac.nz

Outline

NeSI platforms

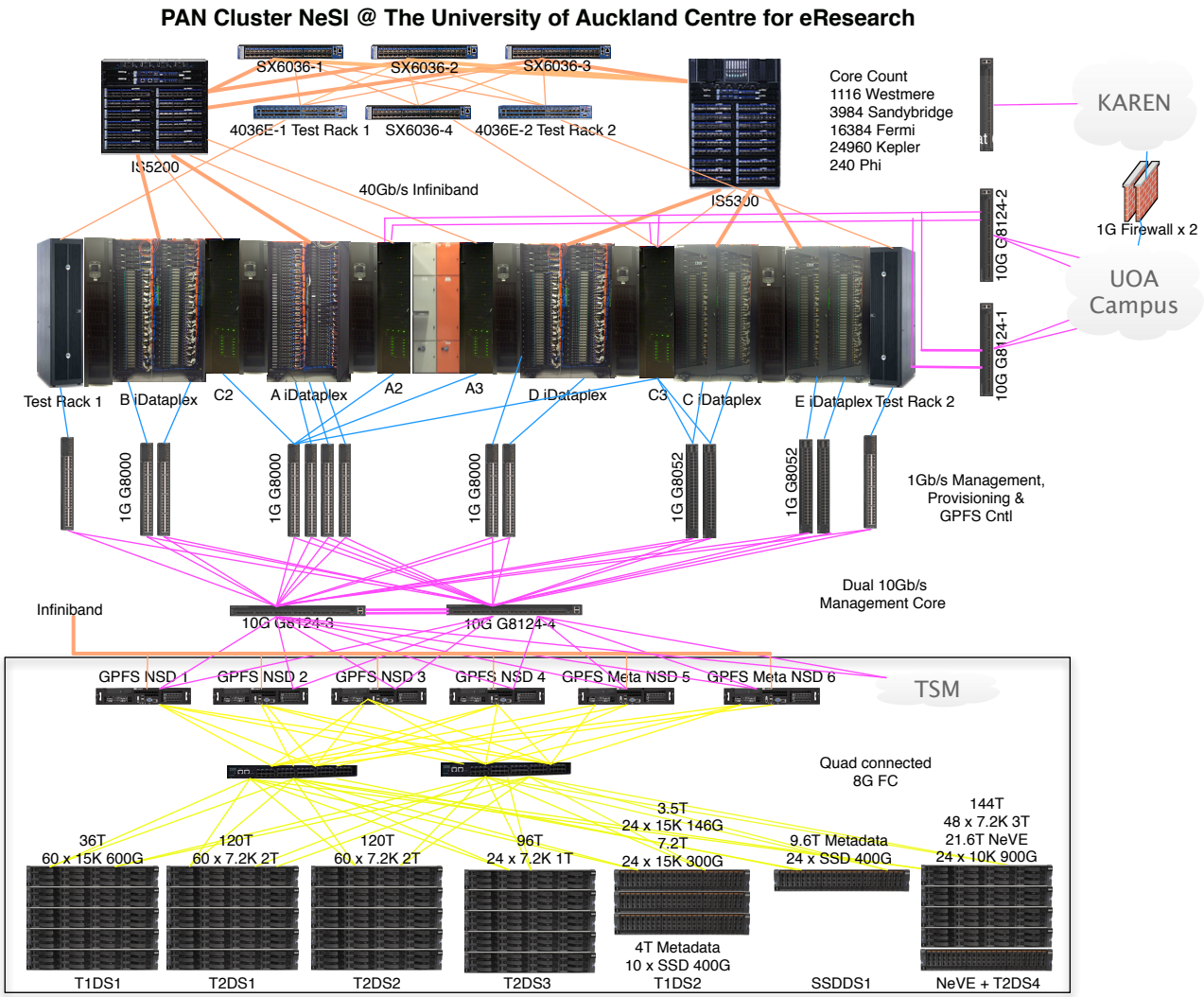
NeSI platforms available for the researchers:

- **BlueFern**: BlueGene/P {8192 cores, 8TB total memory}
- **Fitzroy**: P575/POWER6 {3456 cores, 8TB total memory}
- **BlueFern**: P755/POWER7 {416 cores, 1.5TB total memory}
- **Pan**: x86 cluster {5000 cores, 16TB total memory}

NeSI platforms

- over **700** registered users
- **700** registered projects
- Universities traditionally have varied job patterns
- Audit databases did not cater for meta information
- Central projects database was not wired to carry quantitive jobs information
- University of Auckland Pan cluster was used for carrying out database test

Pan cluster


PAN Cluster NeSI @ The University of Auckland Centre for eResearch
SX6036-1
SX6036-2
SX6036-3
4036E-1 Test Rack 1
SX6036-4
4036E-2 Test Rack 2
IS5200
IS5300
Core Count
1116 Westmere
3984 Sandybridge
16384 Fermi
24960 Kepler
240 Phi
KAREN
1G Firewall x 2
UOA Campus
40Gb/s Infiniband
10G G8124-2
10G G8124-1
Test Rack 1
B iDataplex
C2
A iDataplex
A2
A3
D iDataplex
C3
C iDataplex
E iDataplex
Test Rack 2
1G G8000
1G G8052
1Gb/s Management, Provisioning & GPFS Cntl
Infiniband
10G G8124-3
10G G8124-4
Dual 10Gb/s Management Core
GPFS NSD 1
GPFS NSD 2
GPFS NSD 3
GPFS NSD 4
GPFS Meta NSD 5
GPFS Meta NSD 6
TSM
Quad connected 8G FC
36T
60 x 15K 600G
120T
60 x 7.2K 2T
120T
60 x 7.2K 2T
96T
24 x 7.2K 1T
3.5T
24 x 15K 146G
7.2T
24 x 15K 300G
4T Metadata
10 x SSD 400G
9.6T Metadata
24 x SSD 400G
144T
48 x 7.2K 3T
21.6T NeVE
24 x 10K 900G
T1DS1
T2DS1
T2DS2
T2DS3
T1DS2
SSDDS1
NeVE + T2DS4

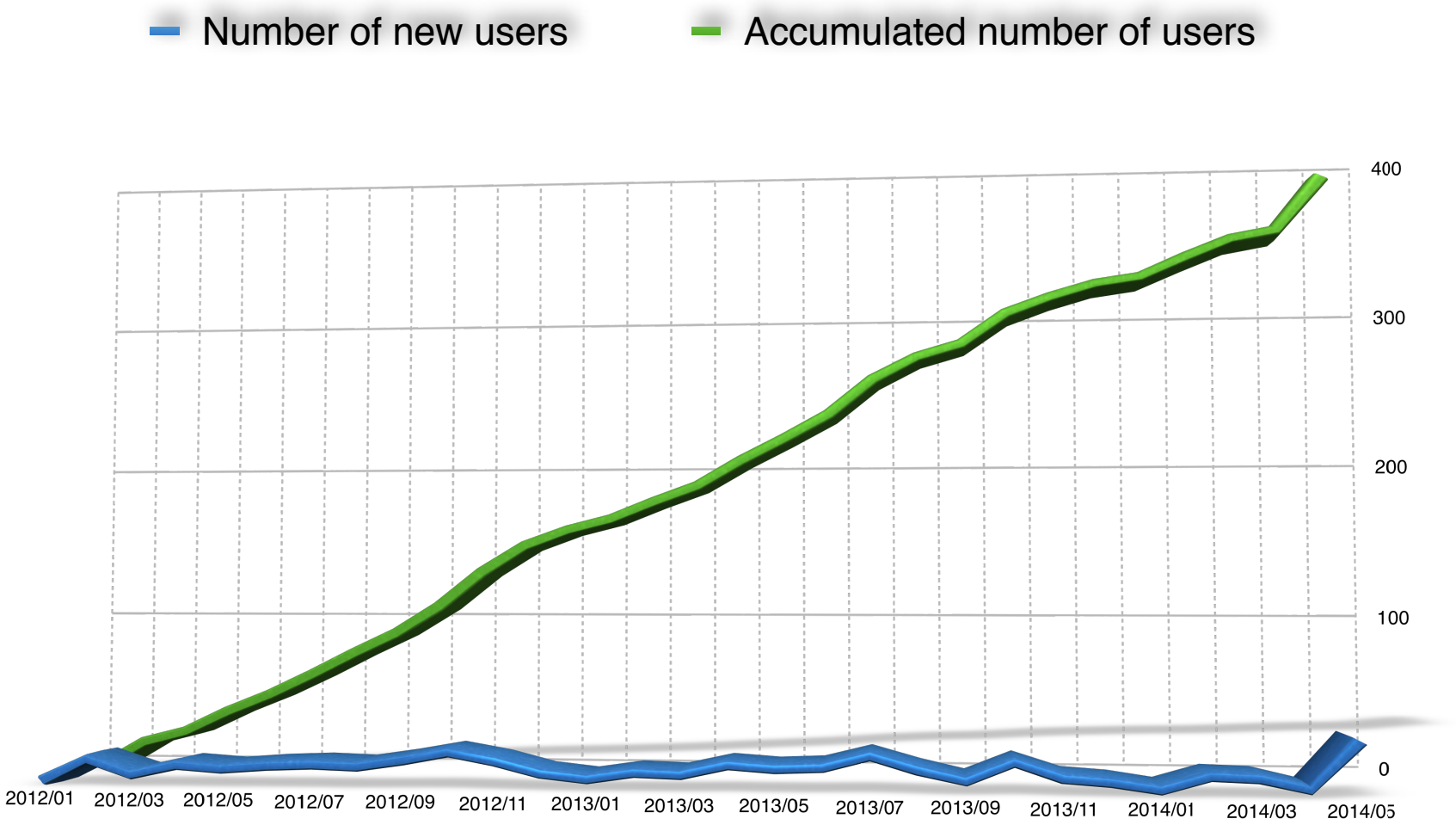
Pan cluster

- Conceived in 2010 and launched in January, 2011
- **5,000** x86 cores, **40,000** GPU cores
- **400TB** shared file system, **GPFS**®
- **InfiniBand**®interconnect
- **LoadLeveler**®batch job scheduler, being phased out in favour of **Slurm**®

Cluster operations

- over **400** registered users
- **300** users submitting jobs on regular basis
- **350** registered projects
- Average daily job throughput: **6000** jobs
- Average daily compute time: **120,000** core hours

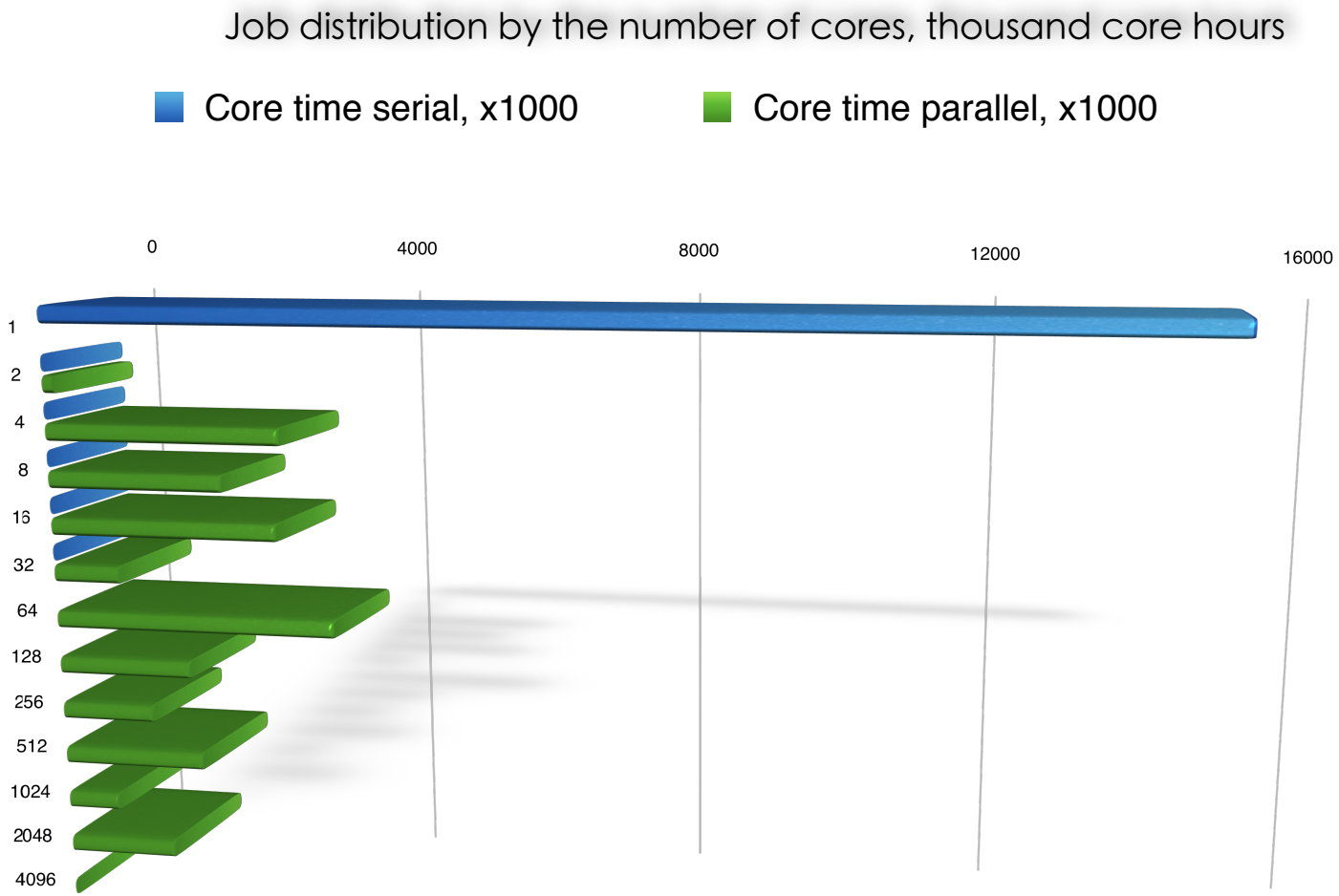
Users intake



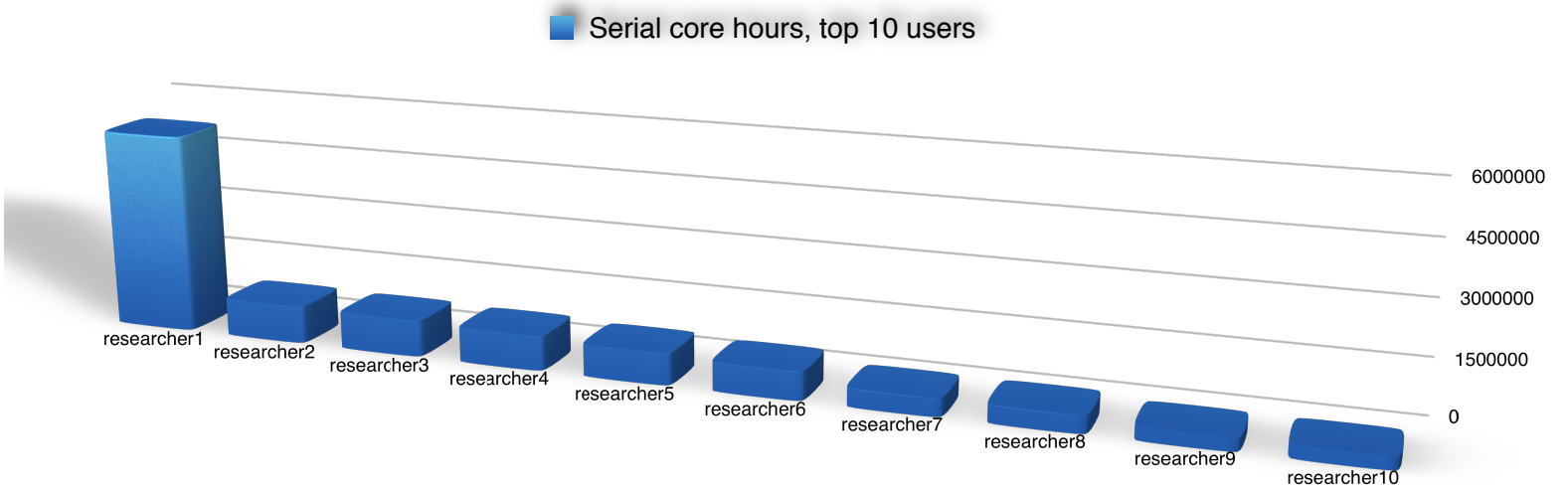
Users intake

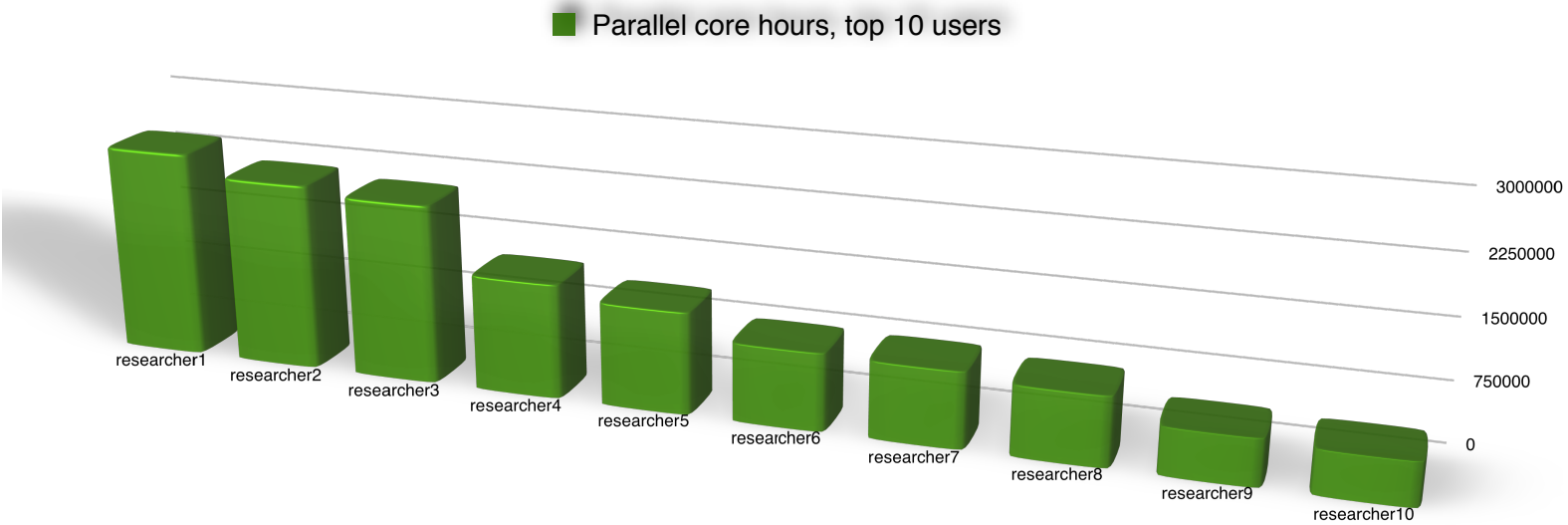
- Users are counted from the time they submit their first job
- Users are associated with affiliation metadata - we can follow institutional use
- Each user participates in one or more projects
- We did **not** know what science domains their work related to

Jobs structure



Jobs structure

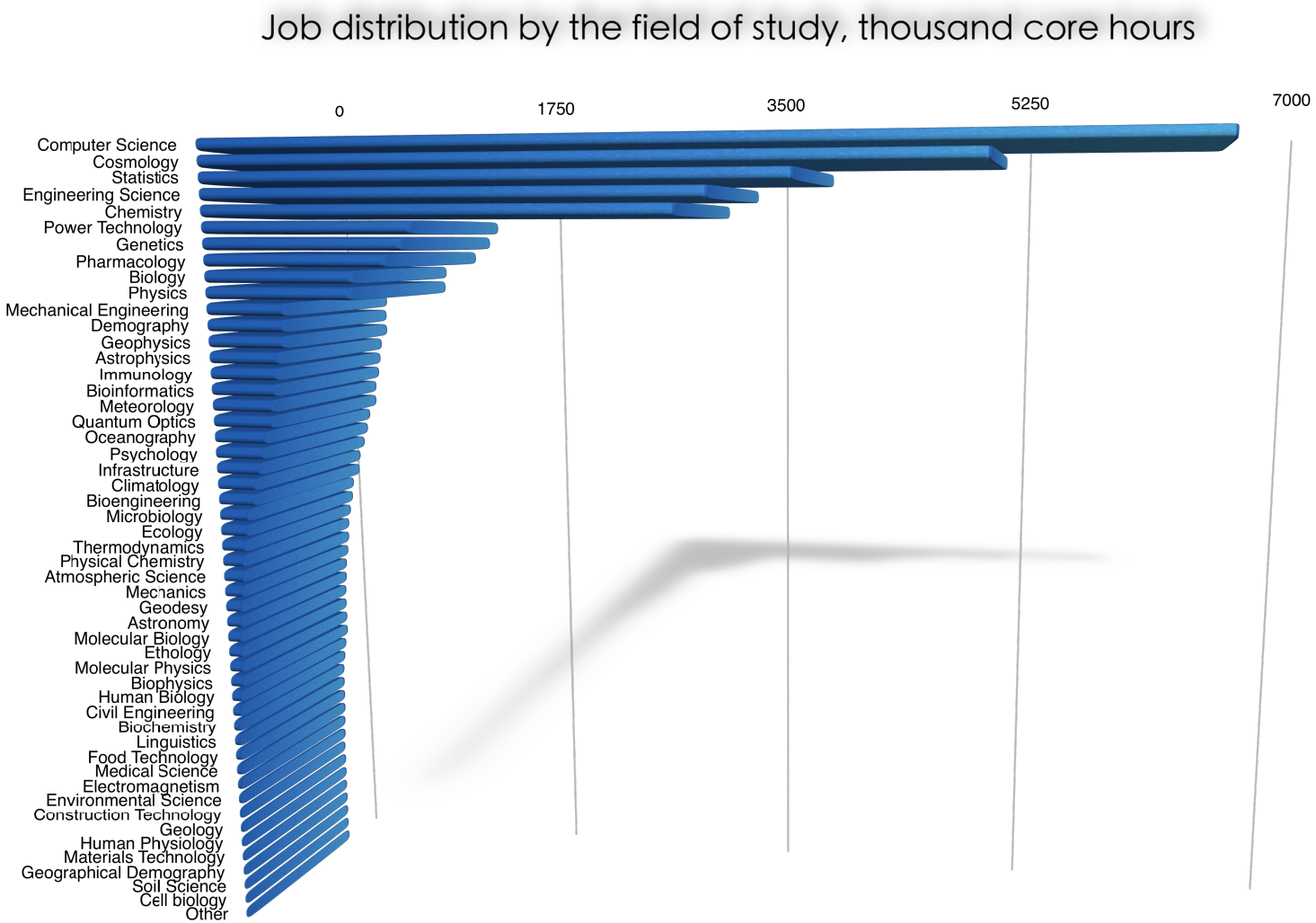




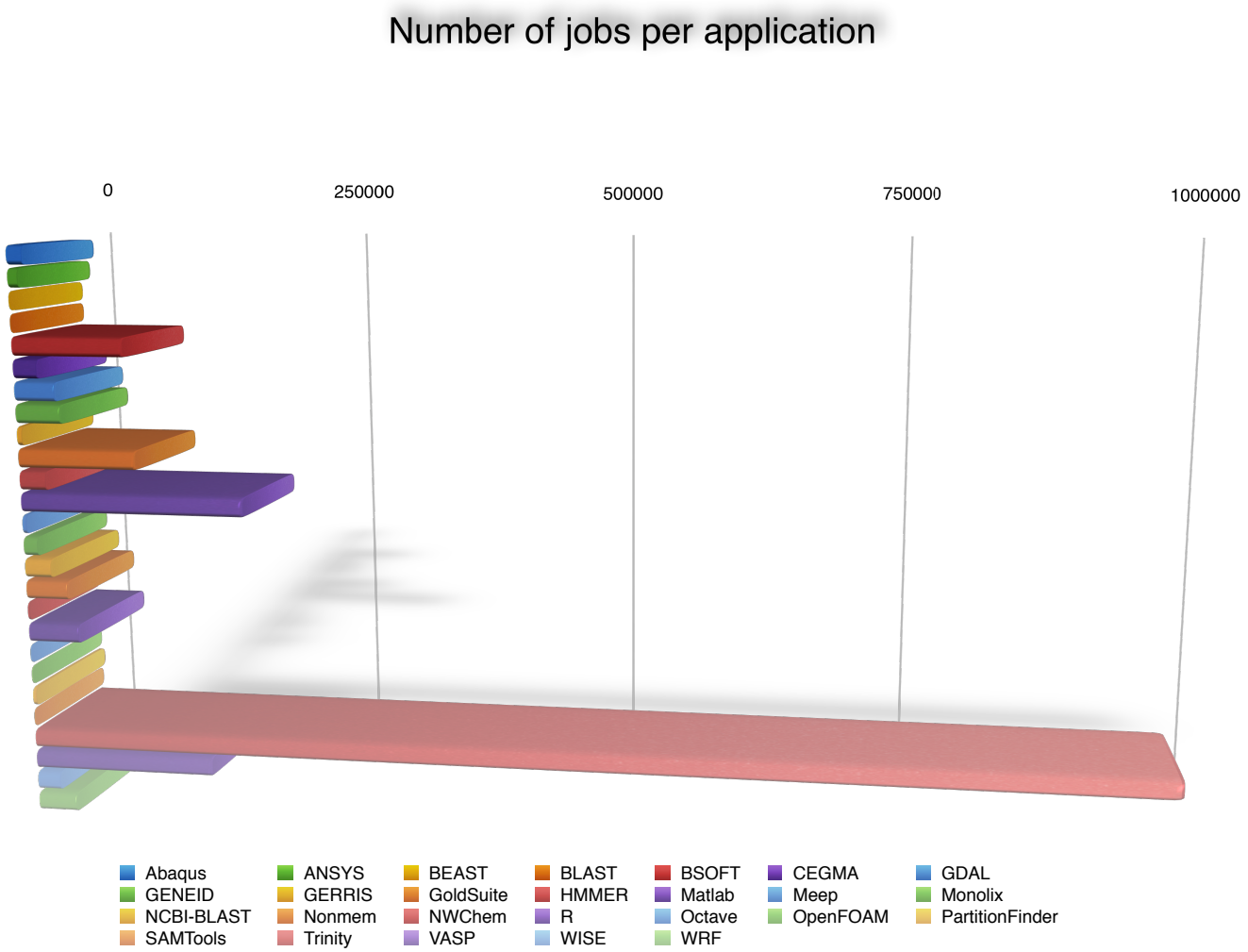
Jobs structure

- The majority of the jobs were serial jobs
- There are users capable of utilising multiple nodes in a single job
- We knew who were our largest users
- We did **not** know what science domains their work related to

Science domains spread



Applications spread



Conclusion

So, why do we need this science domain information? There is a multitude of reasons and many of them have siginificant practical benefit:

- We can better understand what sciences we enable
- We can prioritize job scheduling based on the science needs
- We can predict applications usage and optimize packages placement
- We can plan for the future with regard to computational resources and support component

Questions & Answers