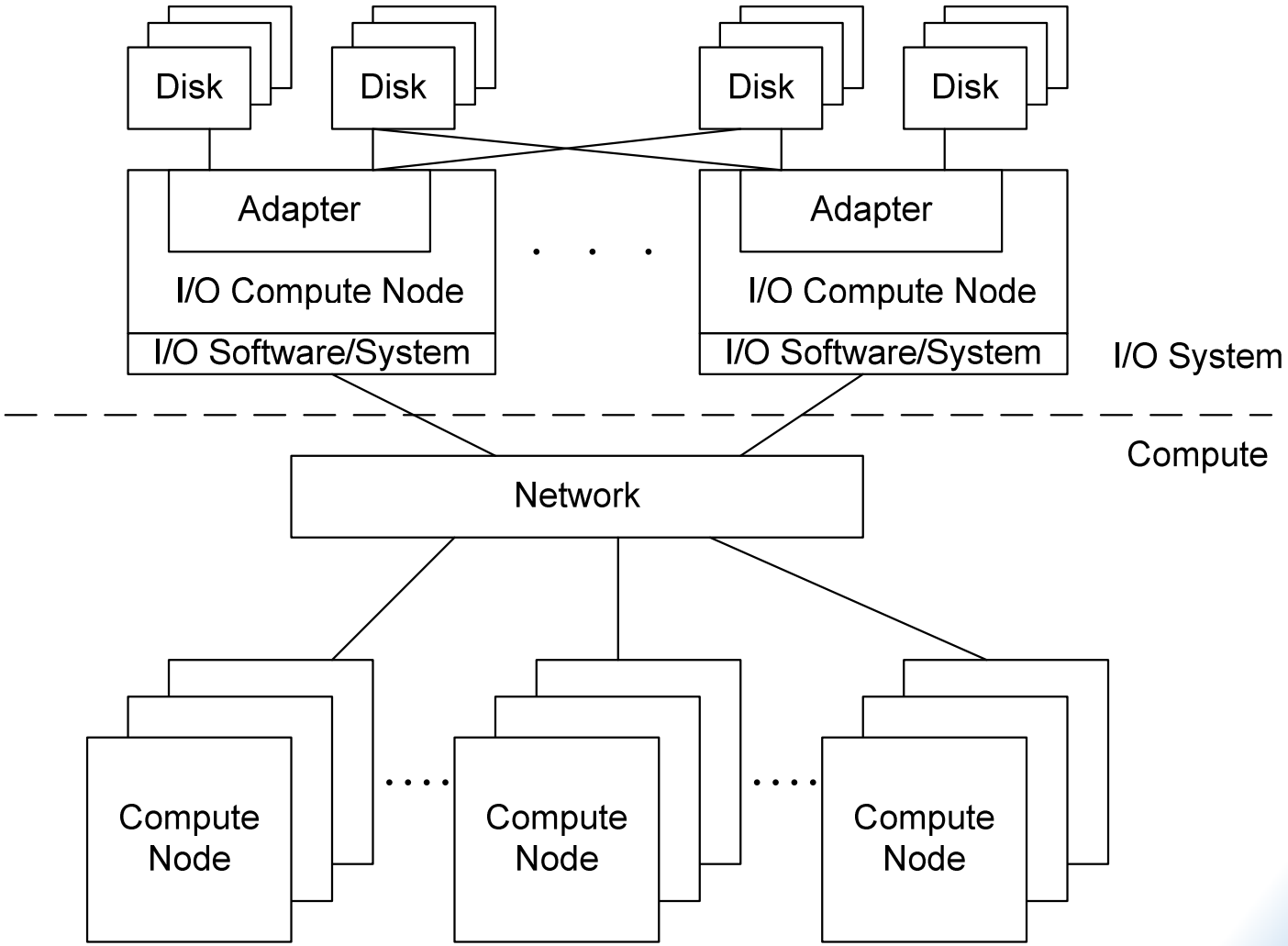
Application and Systemwide Performance

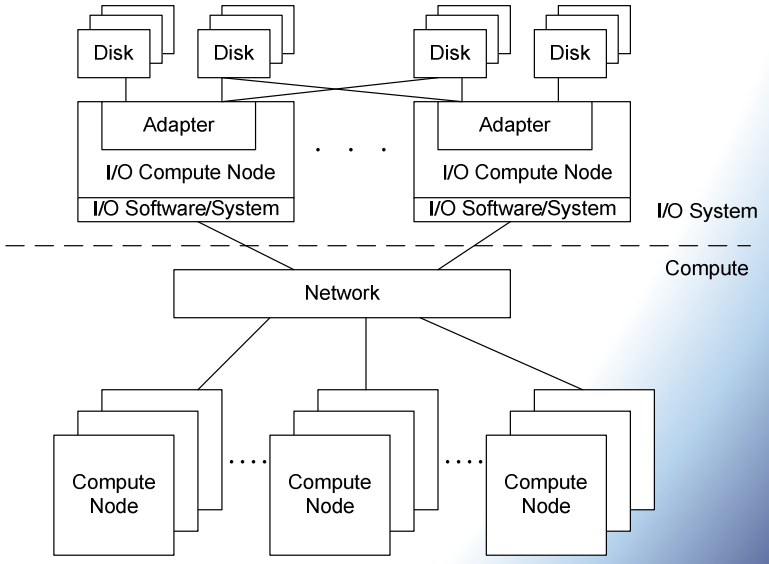
Adrian Jackson
adrianj@epcc.ed.ac.uk
@adrianjhpc

Different aspects of I/O system



Shared resource

- I/O systems commissioned for peak performance on high through put
  - Generally done uncontended
- Each part of the system is shared
  - Communication network
  - I/O nodes
    - Memory
    - Disks
  - I/O metadata server



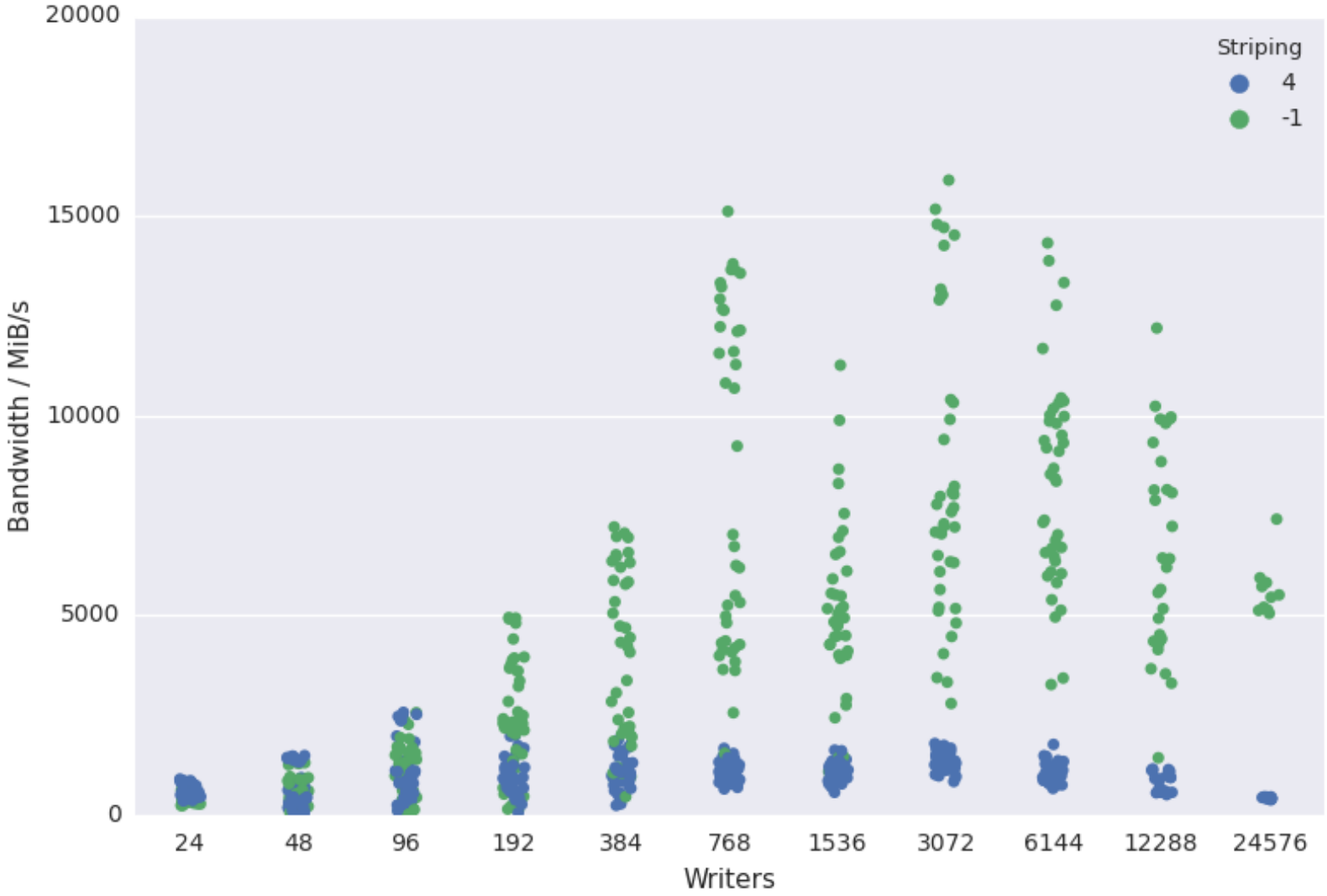
Performance dependent on operation

- Opening files, creating files, writing data, reading data
- Performance complicated by caching at multiple levels
  - Linux kernel, filesystem client, filesystem servers,…
  - Don’t care for real applications
- Scale
  - Number of processes used changes performance profiles

Different types of I/O

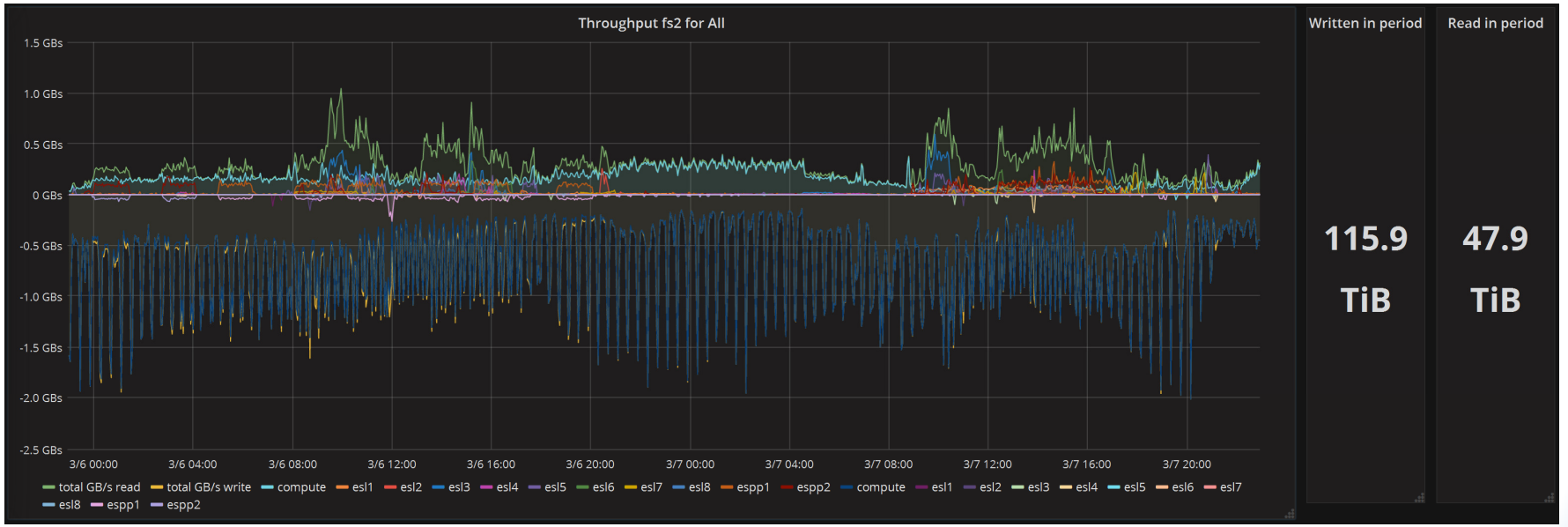
- Compute nodes
- Login nodes
- Pre/post-processing nodes
- Into or out of the system

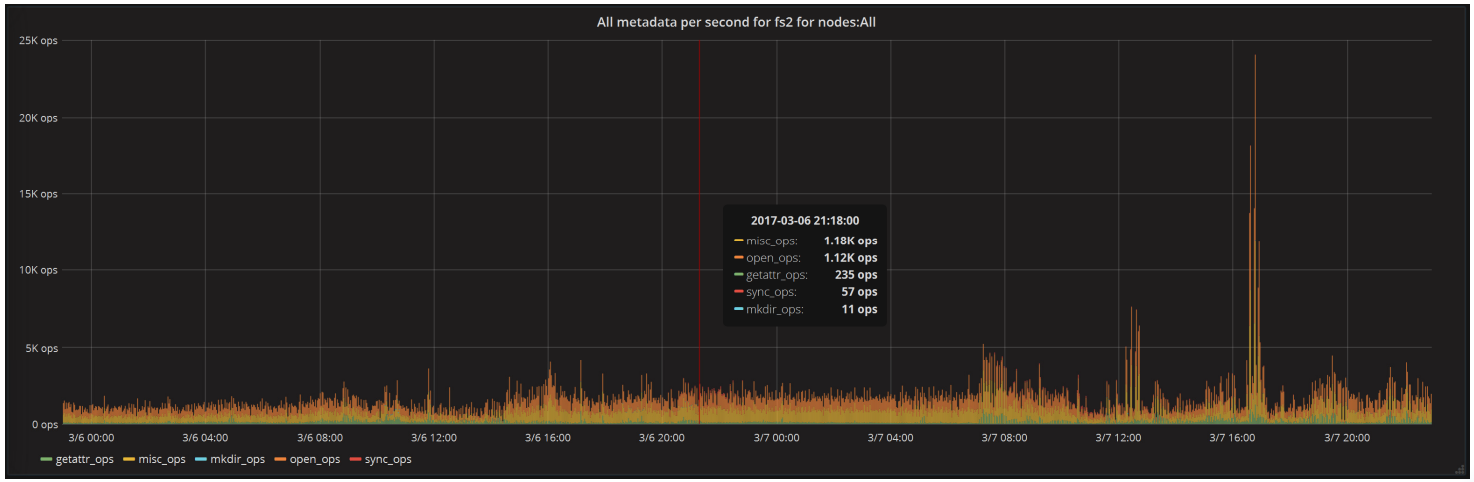
ARCHER I/O Performance



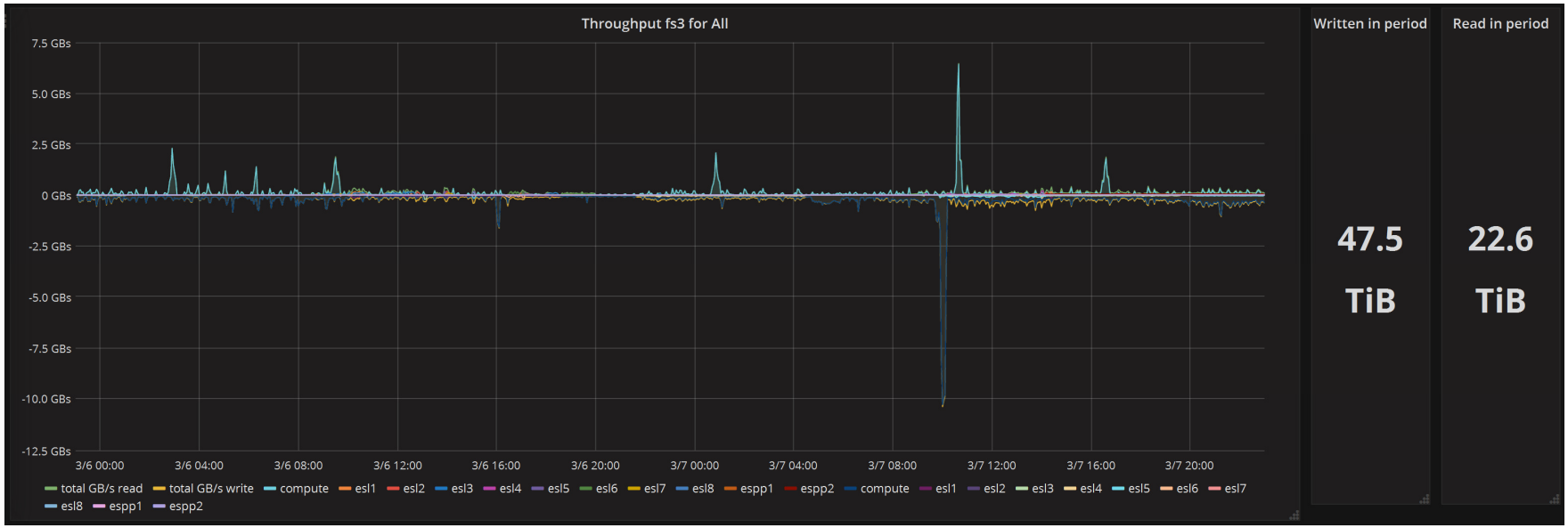
- https://www.archer.ac.uk/documentation/white-papers/parallelIO-benchmarking/ARCHER-Parallel-IO-1.0.pdf

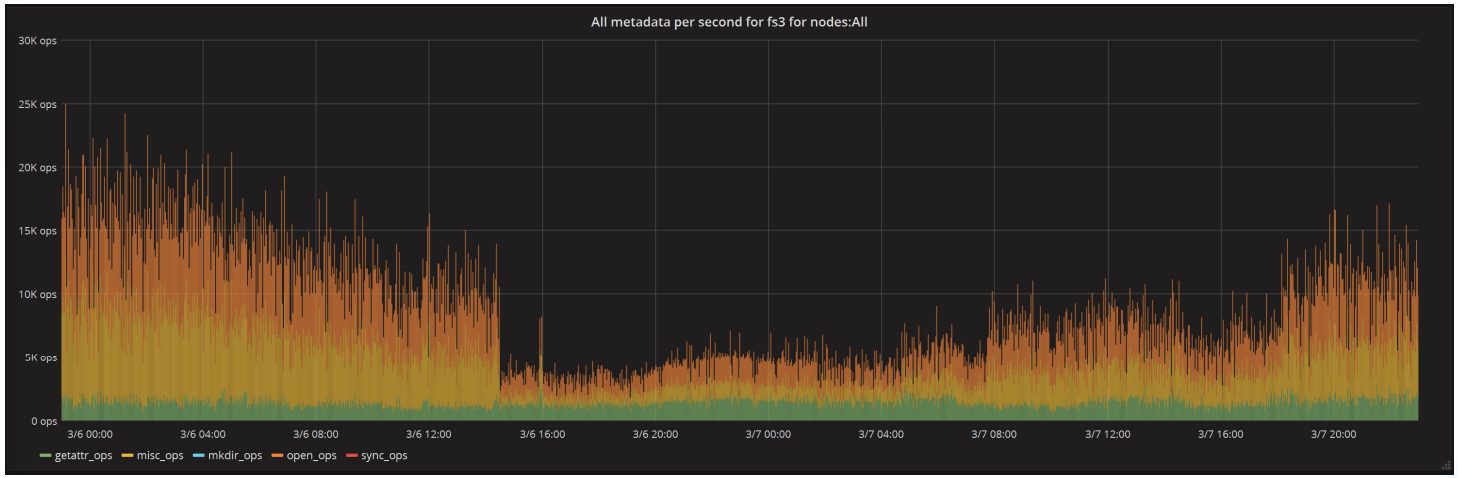
ARCHER workload



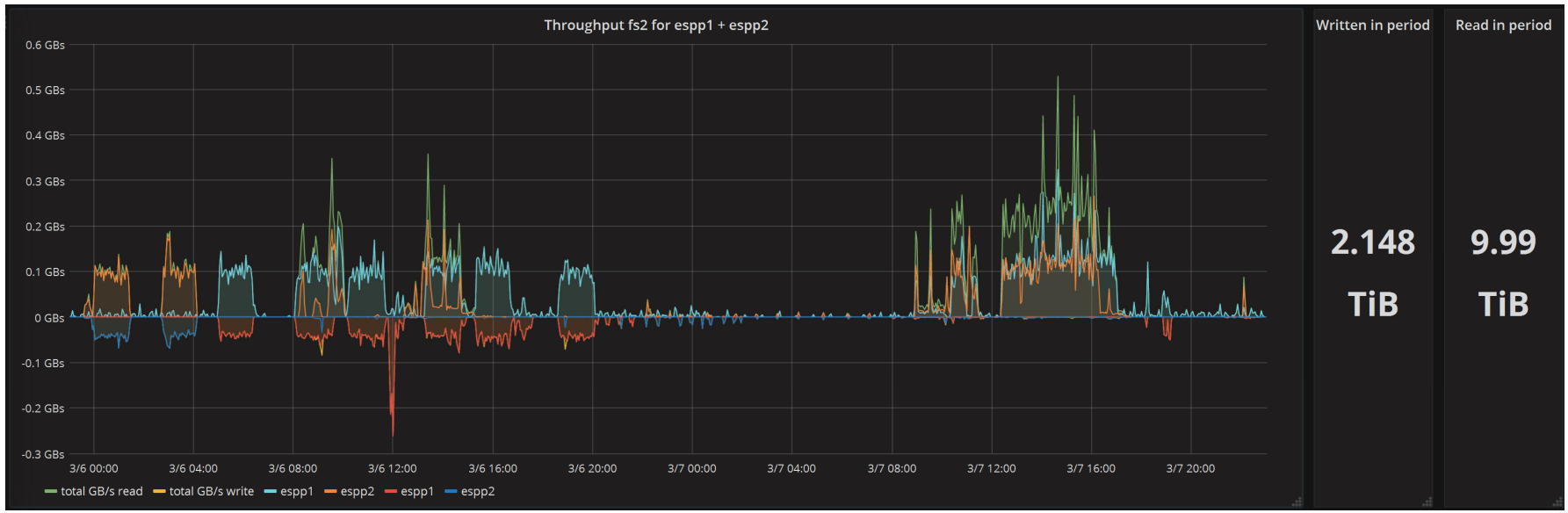


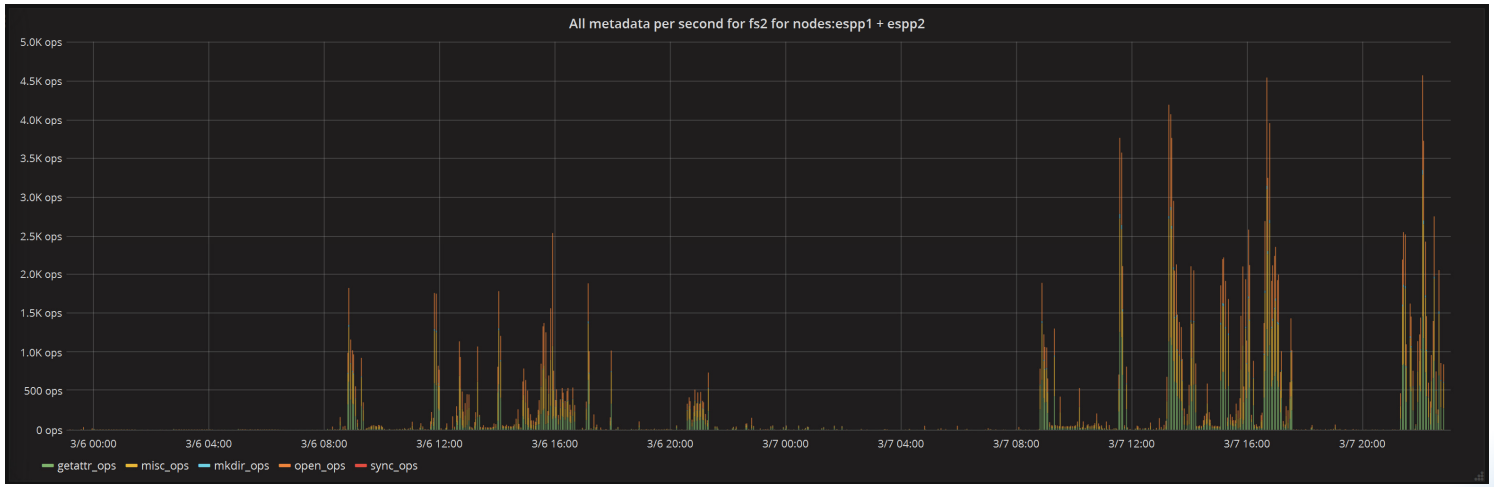
ARCHER workload



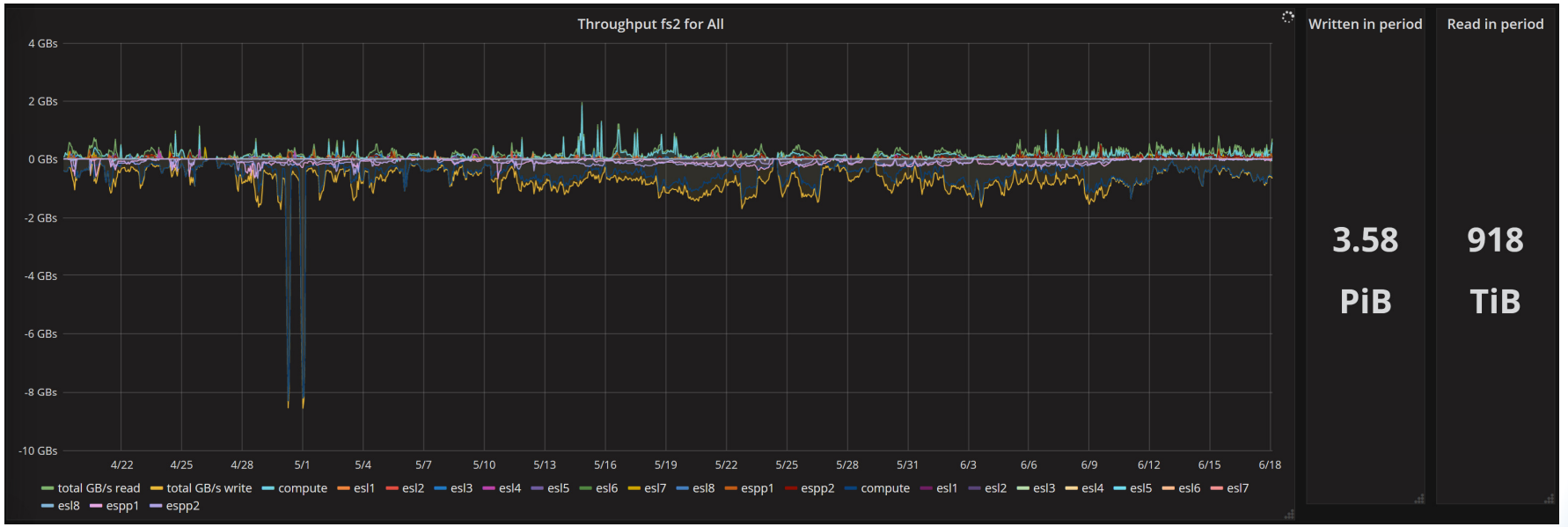


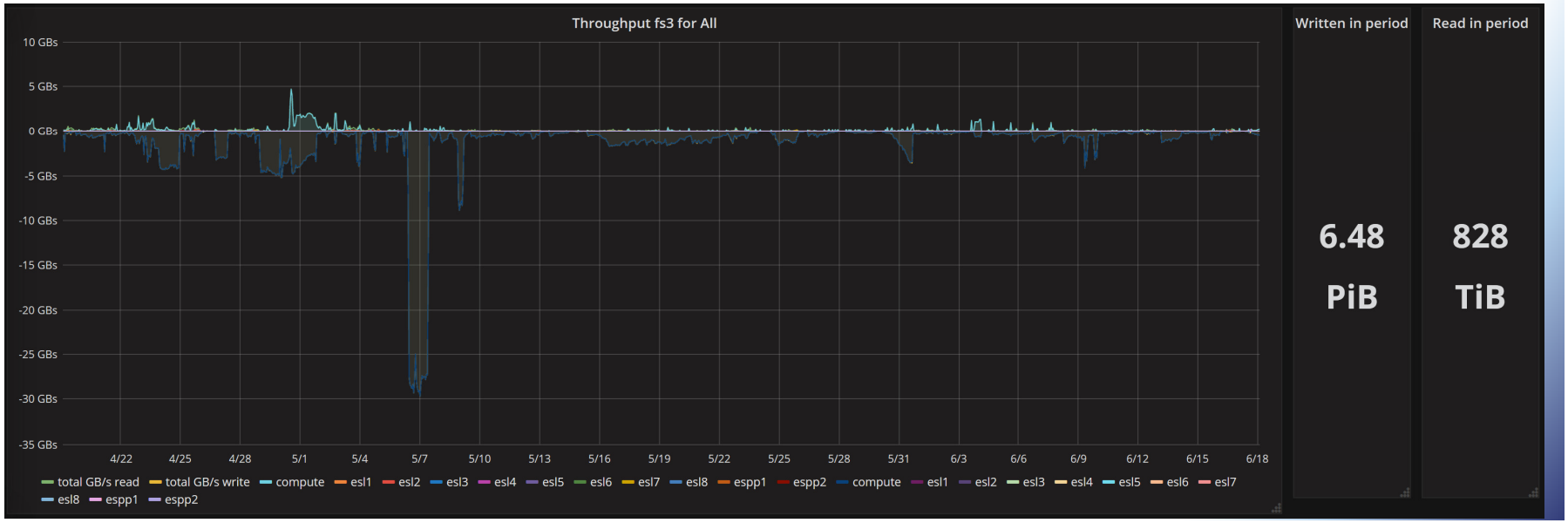
ARCHER workload



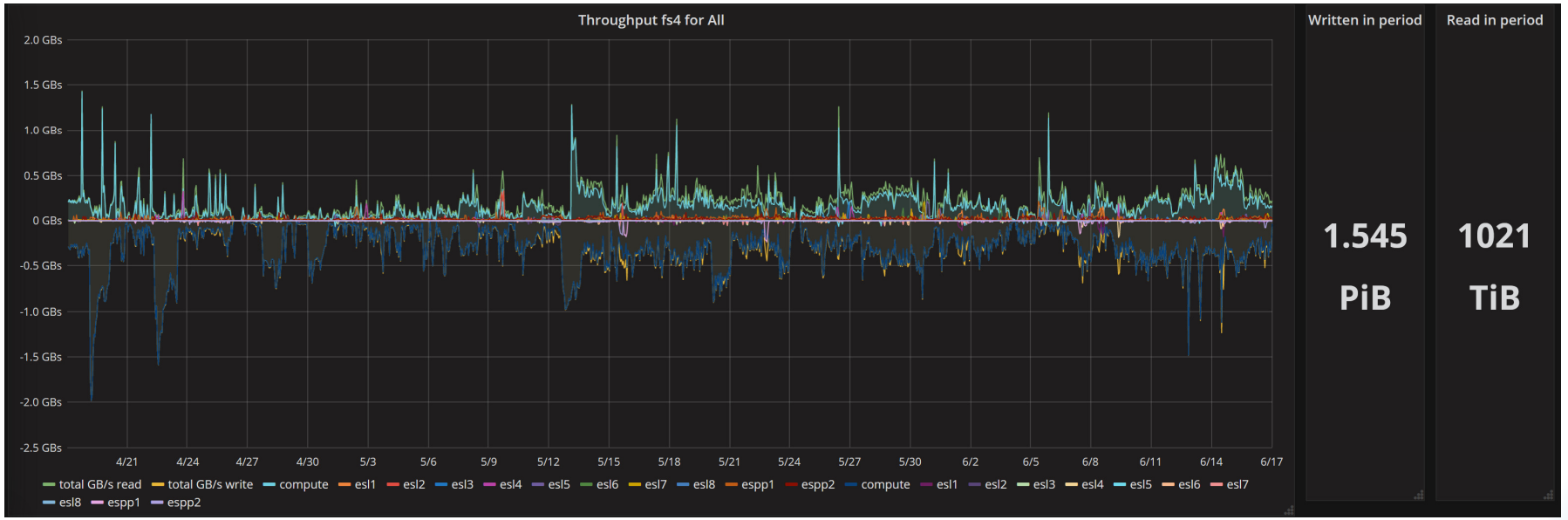


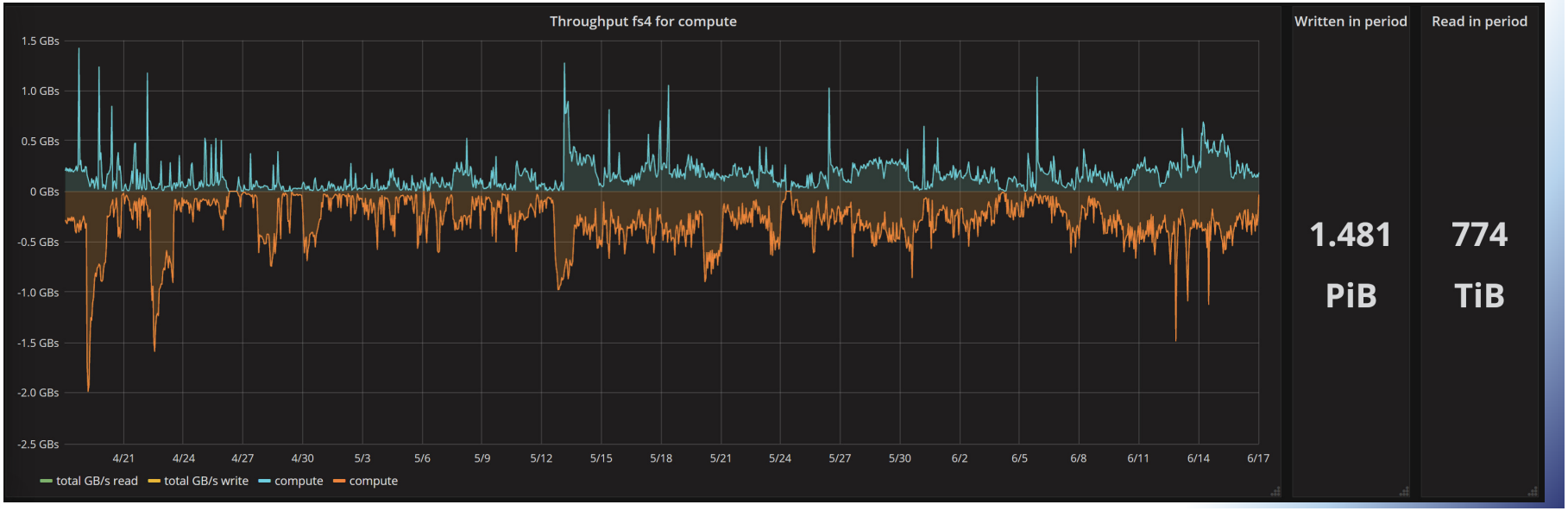
ARCHER


Throughput fs2 for All
4 GBs
2 GBs
0 GBs
-2 GBs
-4 GBs
-6 GBs
-8 GBs
-10 GBs
4/22 4/25 4/28 5/1 5/4 5/7 5/10 5/13 5/16 5/19 5/22 5/25 5/28 5/31 6/3 6/6 6/9 6/12 6/15 6/18
total GB/s read
total GB/s write
compute
esl1
esl2
esl3
esl4
esl5
esl6
esl7
esl8
espp1
espp2
Written in period
3.58
PiB
Read in period
918
TiB

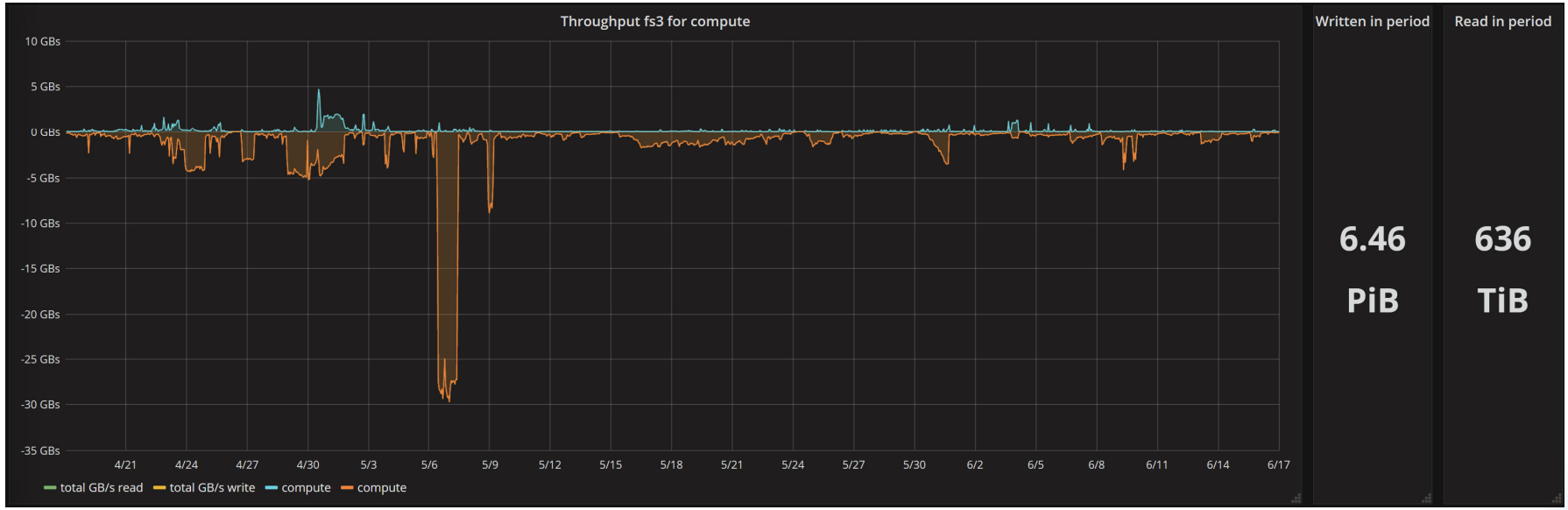

Throughput fs3 for All
10 GBs
5 GBs
0 GBs
-5 GBs
-10 GBs
-15 GBs
-20 GBs
-25 GBs
-30 GBs
-35 GBs
4/22 4/25 4/28 5/1 5/4 5/7 5/10 5/13 5/16 5/19 5/22 5/25 5/28 5/31 6/3 6/6 6/9 6/12 6/15 6/18
total GB/s read
total GB/s write
compute
esl1
esl2
esl3
esl4
esl5
esl6
esl7
esl8
espp1
espp2
Written in period
6.48
PiB
Read in period
828
TiB

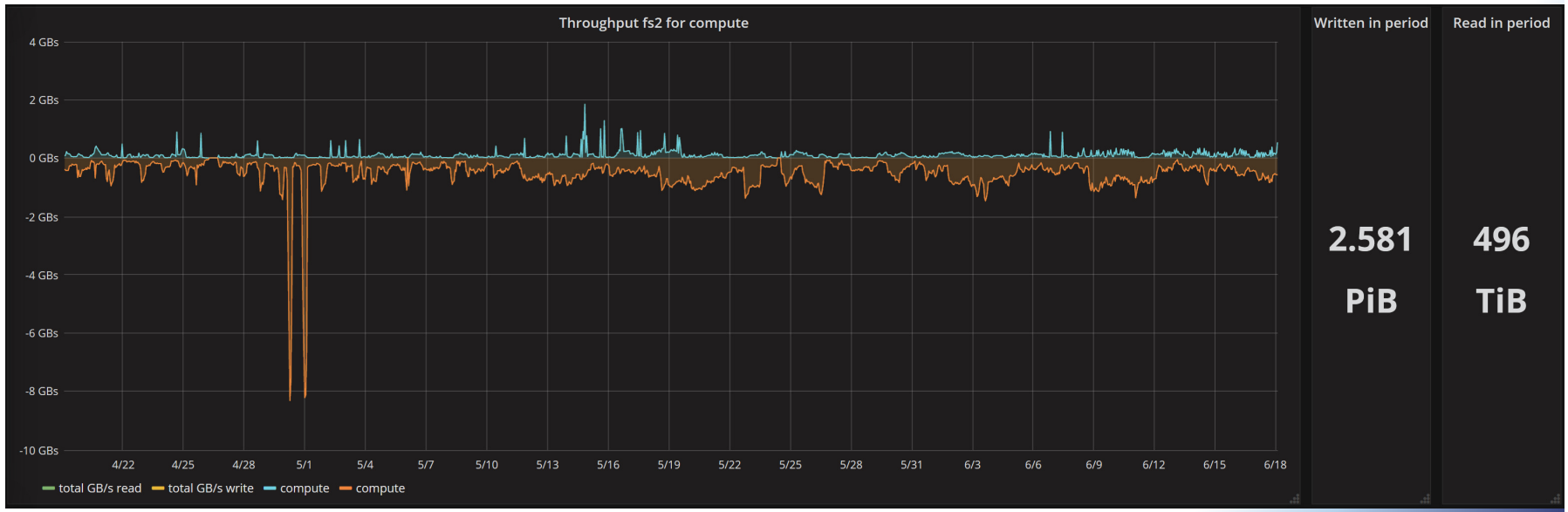
ARCHER


Throughput fs4 for All
2.0 GBs
1.5 GBs
1.0 GBs
0.5 GBs
0 GBs
-0.5 GBs
-1.0 GBs
-1.5 GBs
-2.0 GBs
-2.5 GBs
4/21 4/24 4/27 4/30 5/3 5/6 5/9 5/12 5/15 5/18 5/21 5/24 5/27 5/30 6/2 6/5 6/8 6/11 6/14 6/17
total GB/s read
total GB/s write
compute
esl1
esl2
esl3
esl4
esl5
esl6
esl7
esl8
espp1
espp2
compute
esl1
esl2
esl3
esl4
esl5
esl6
esl7
esl8
espp1
espp2
Written in period
1.545
PiB
Read in period
1021
TiB

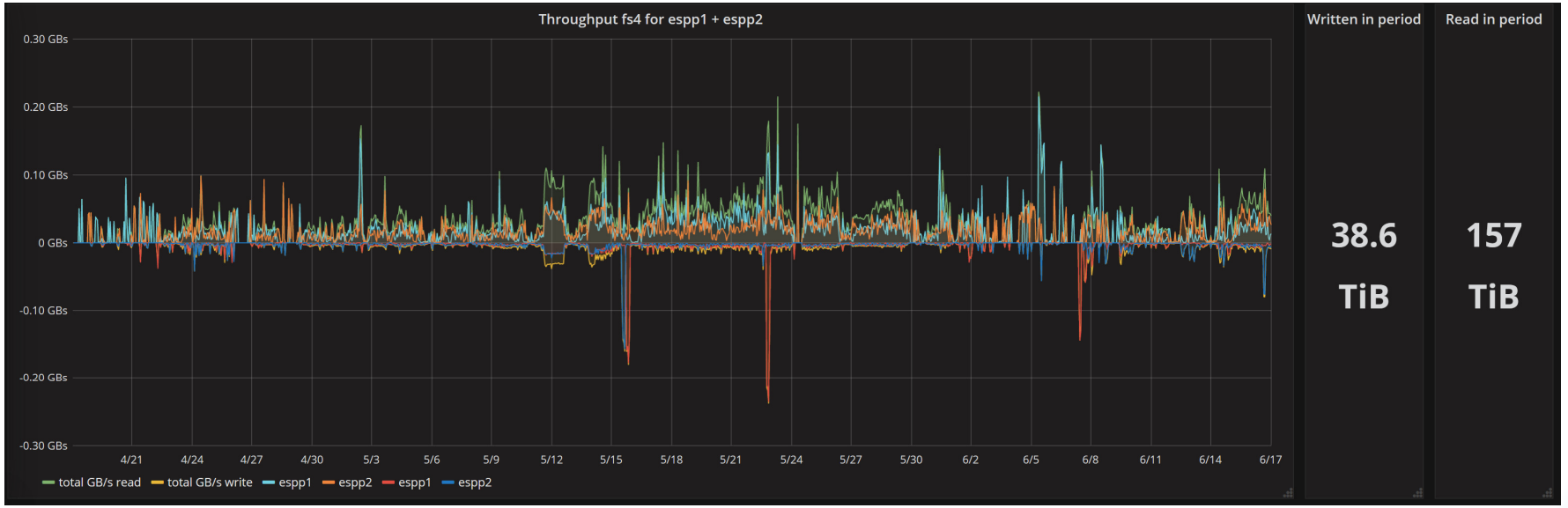

Throughput fs4 for compute
1.5 GBs
1.0 GBs
0.5 GBs
0 GBs
-0.5 GBs
-1.0 GBs
-1.5 GBs
-2.0 GBs
-2.5 GBs
4/21 4/24 4/27 4/30 5/3 5/6 5/9 5/12 5/15 5/18 5/21 5/24 5/27 5/30 6/2 6/5 6/8 6/11 6/14 6/17
total GB/s read
total GB/s write
compute
compute
Written in period
1.481
PiB
Read in period
774
TiB

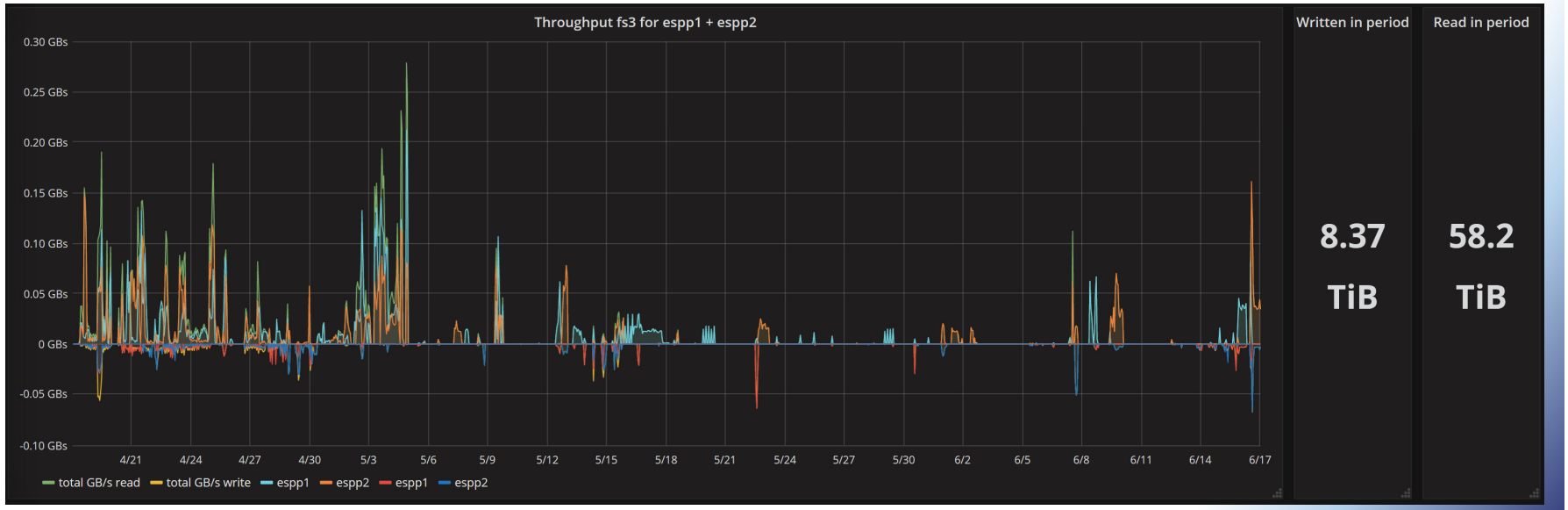
ARCHER Compute



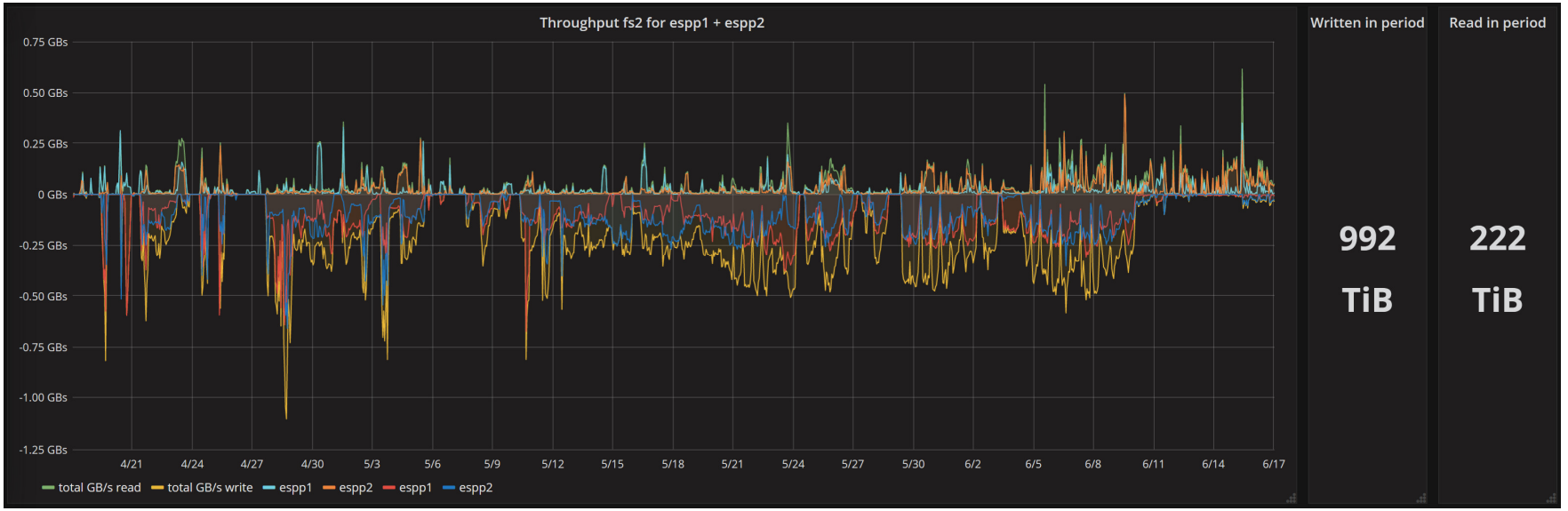


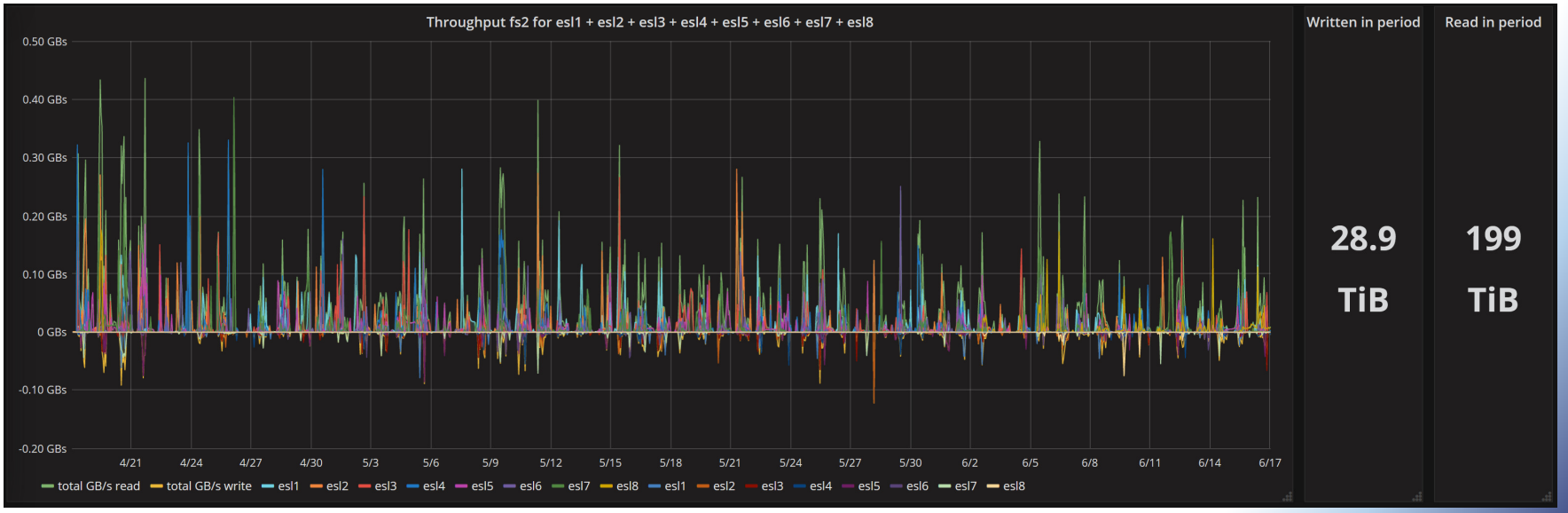
ARCHER post processing



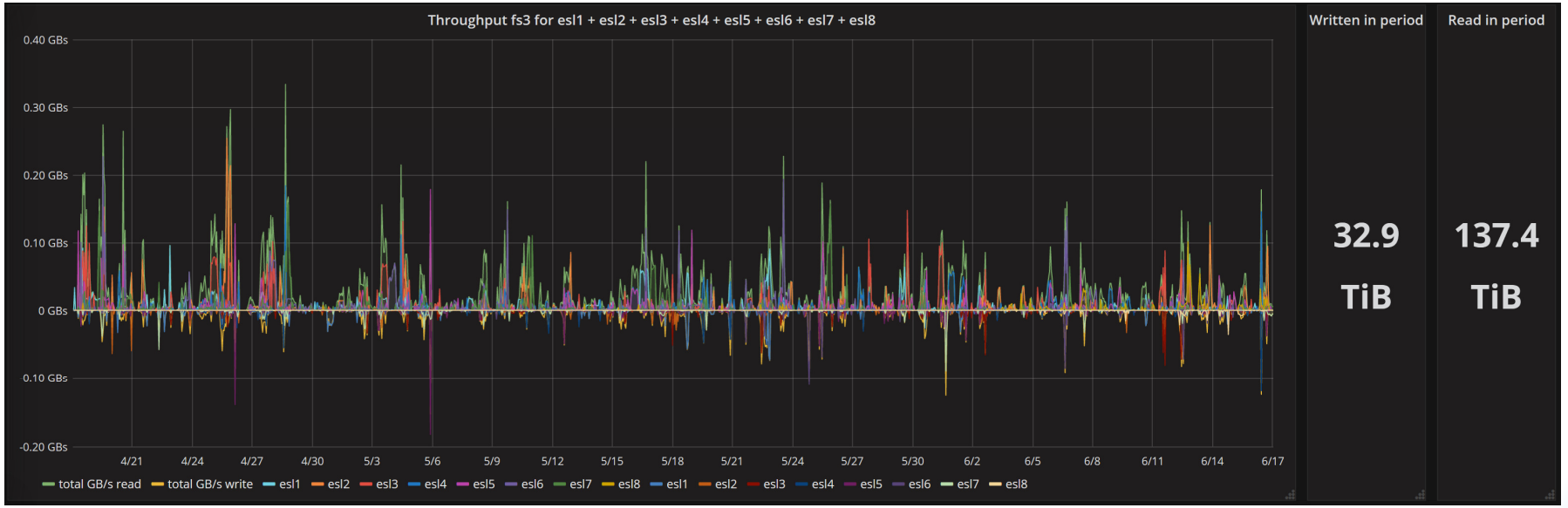


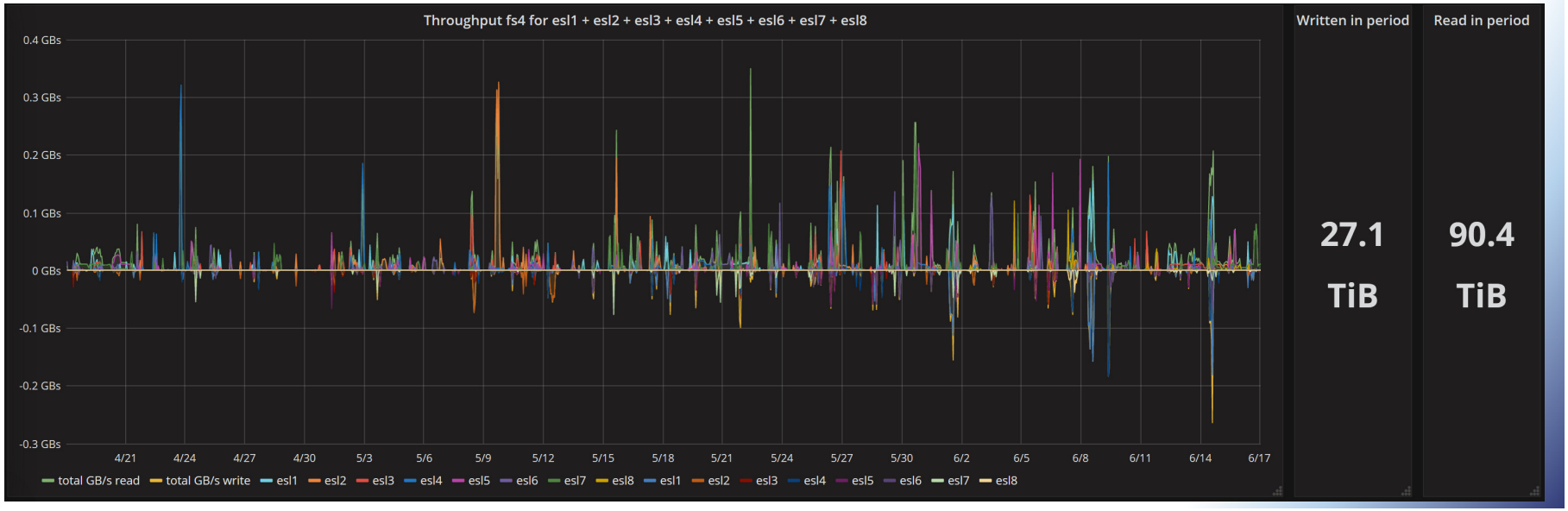
ARCHER post processing/login



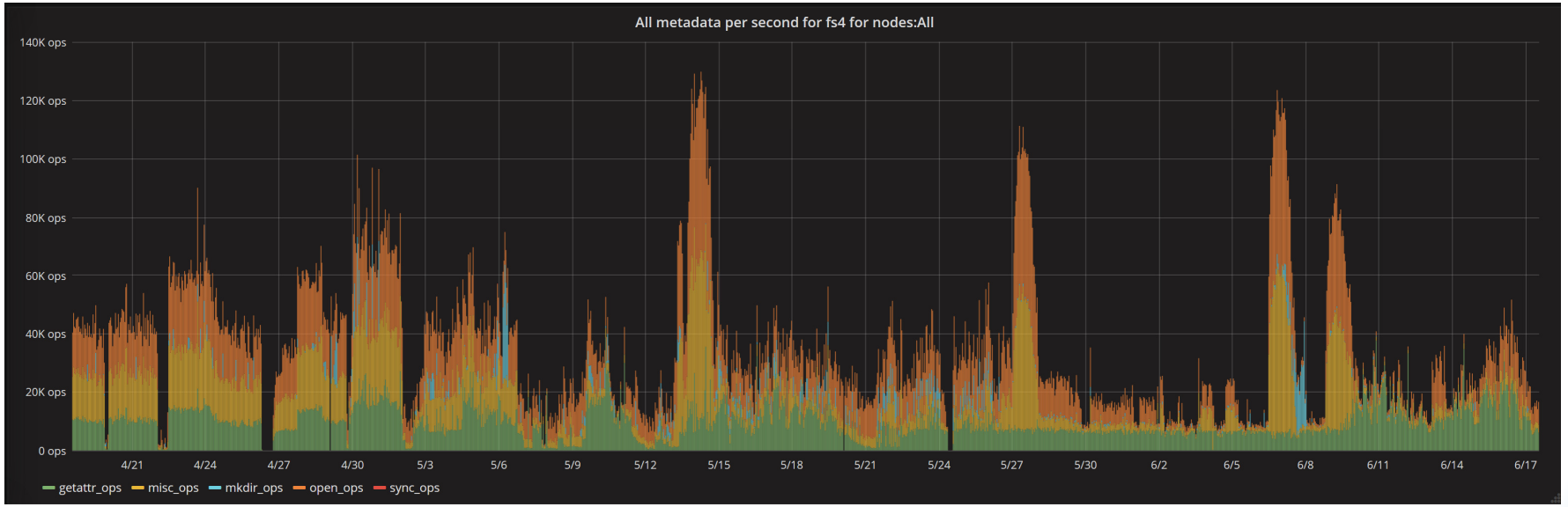


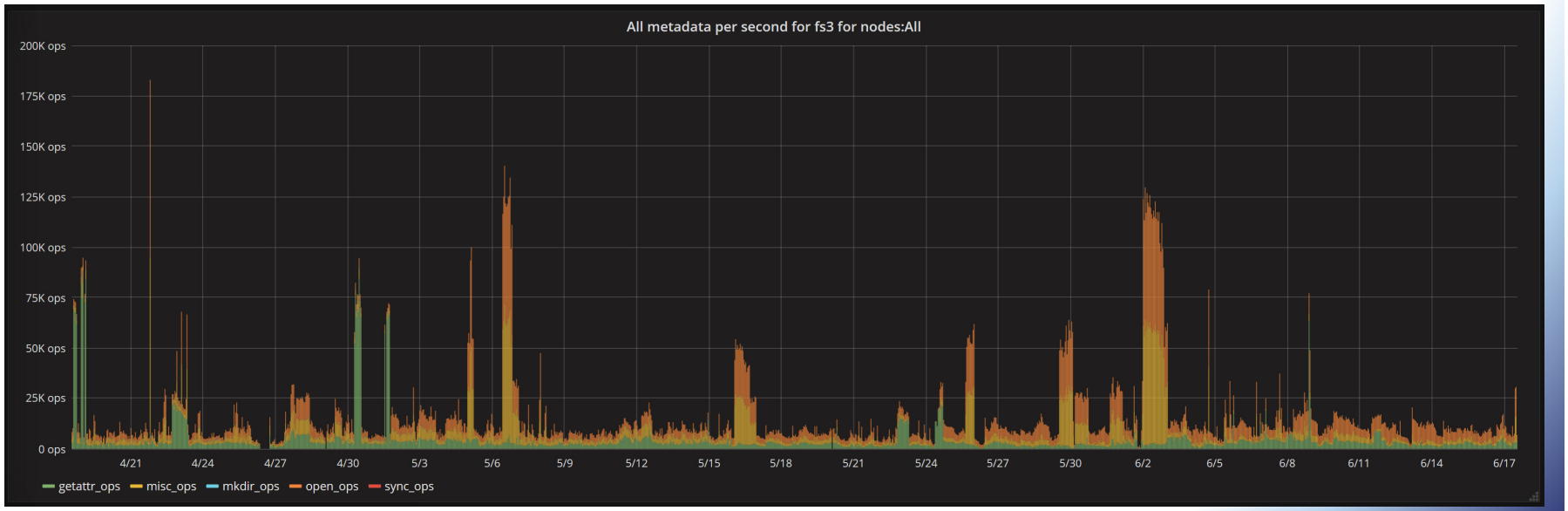
ARCHER login



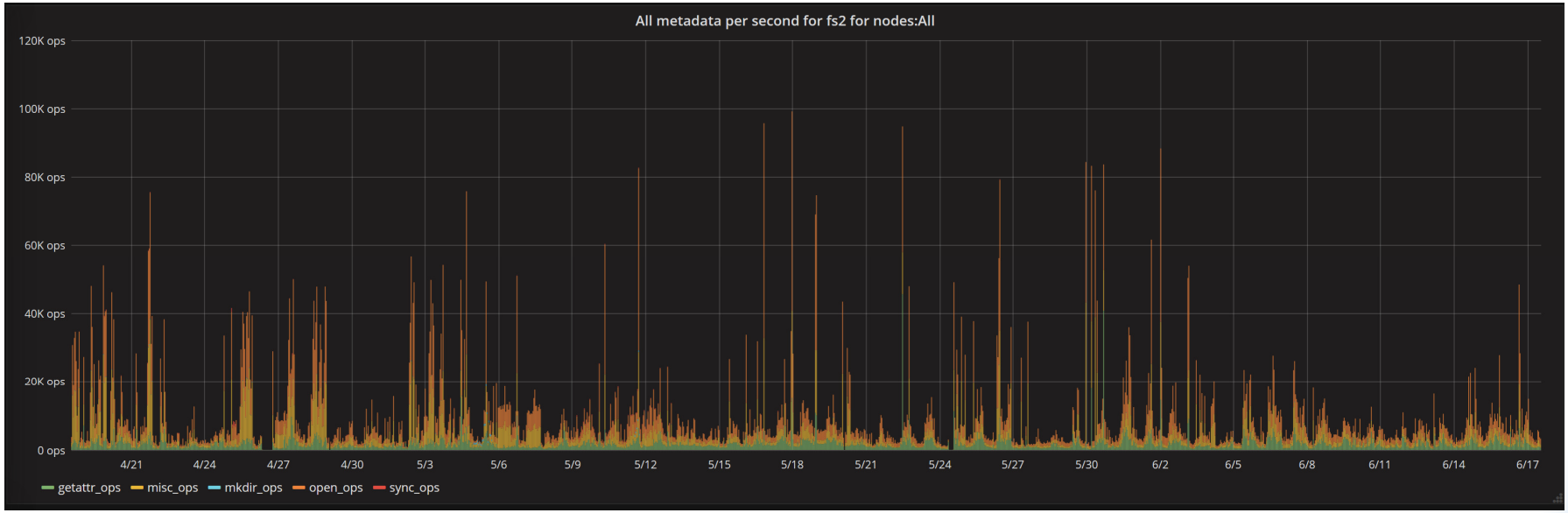


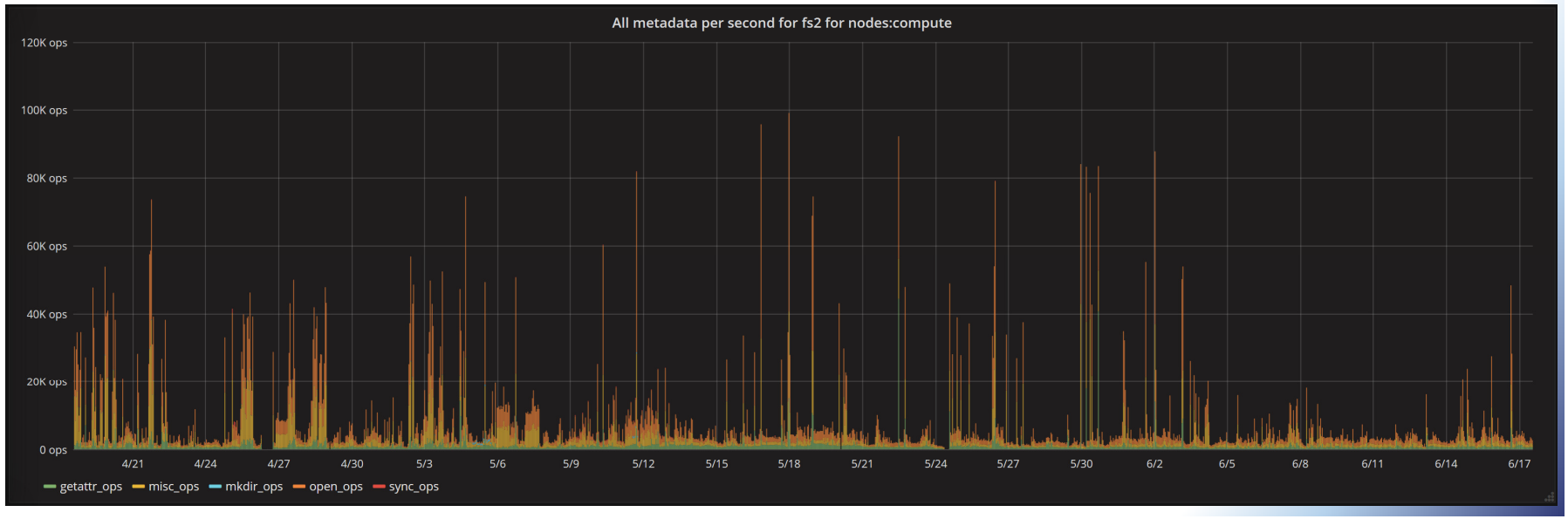
ARCHER metadata operations



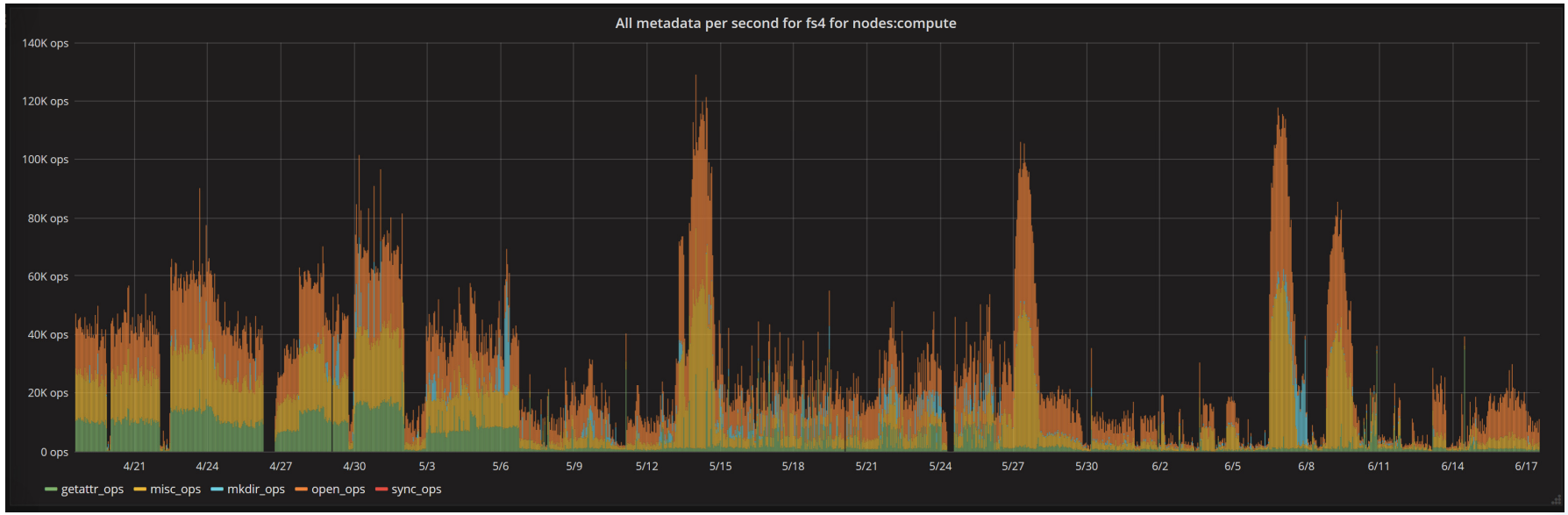


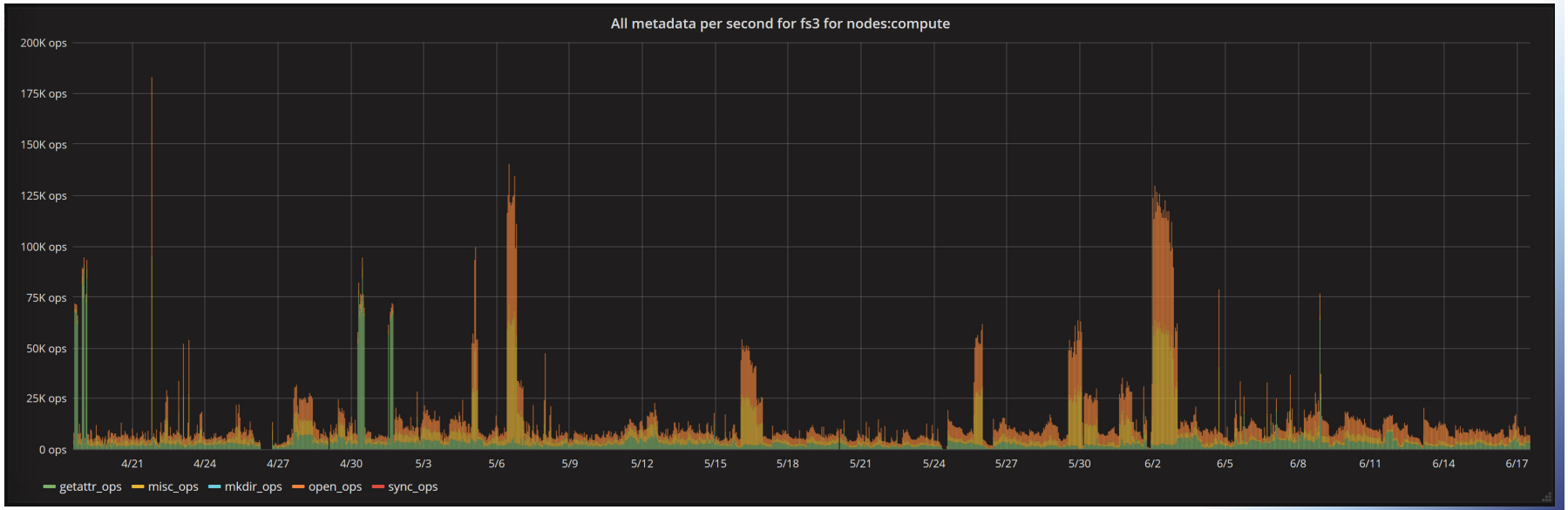
ARCHER metadata operations



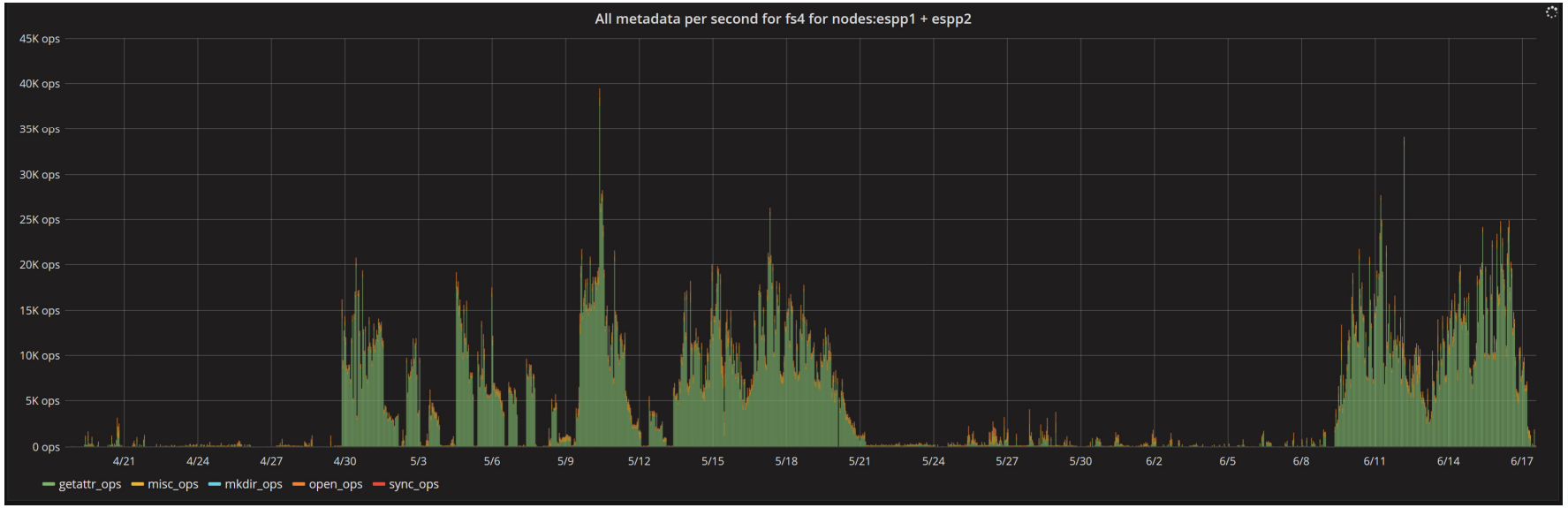


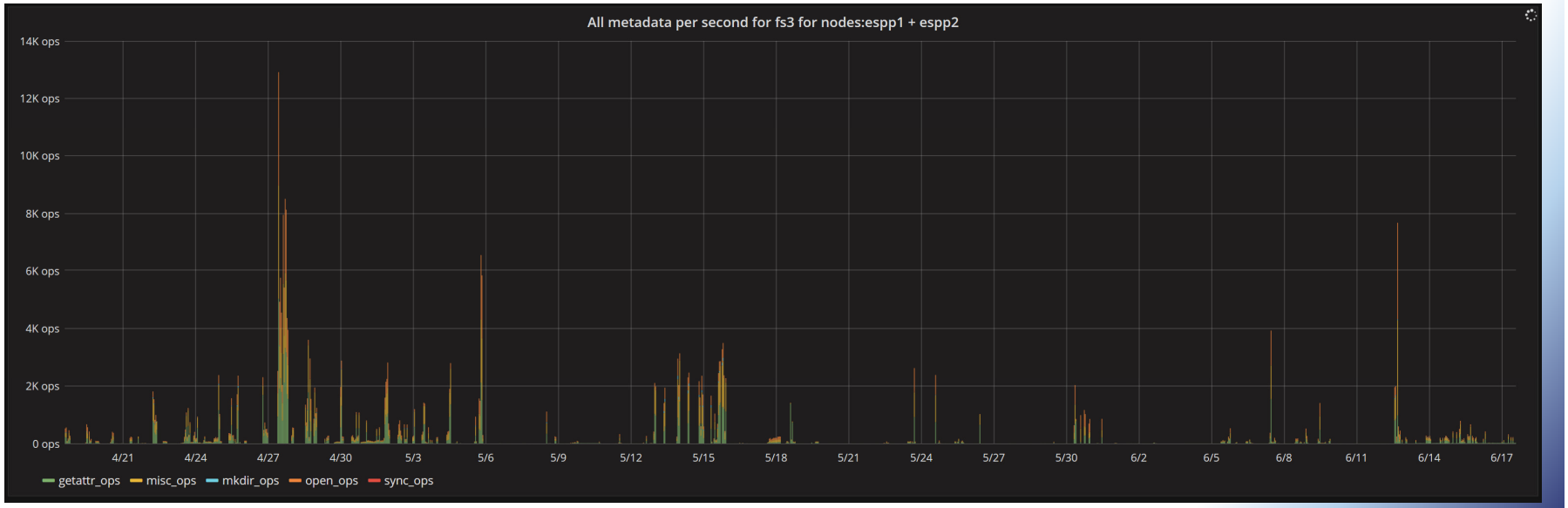
ARCHER compute metadata



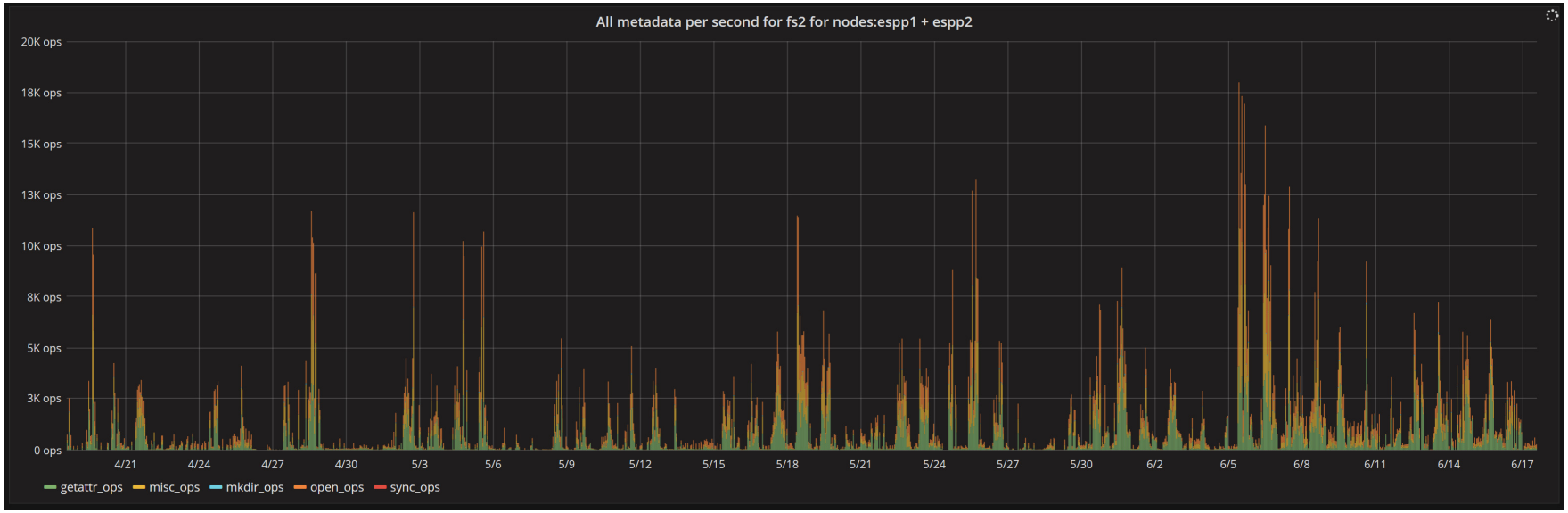


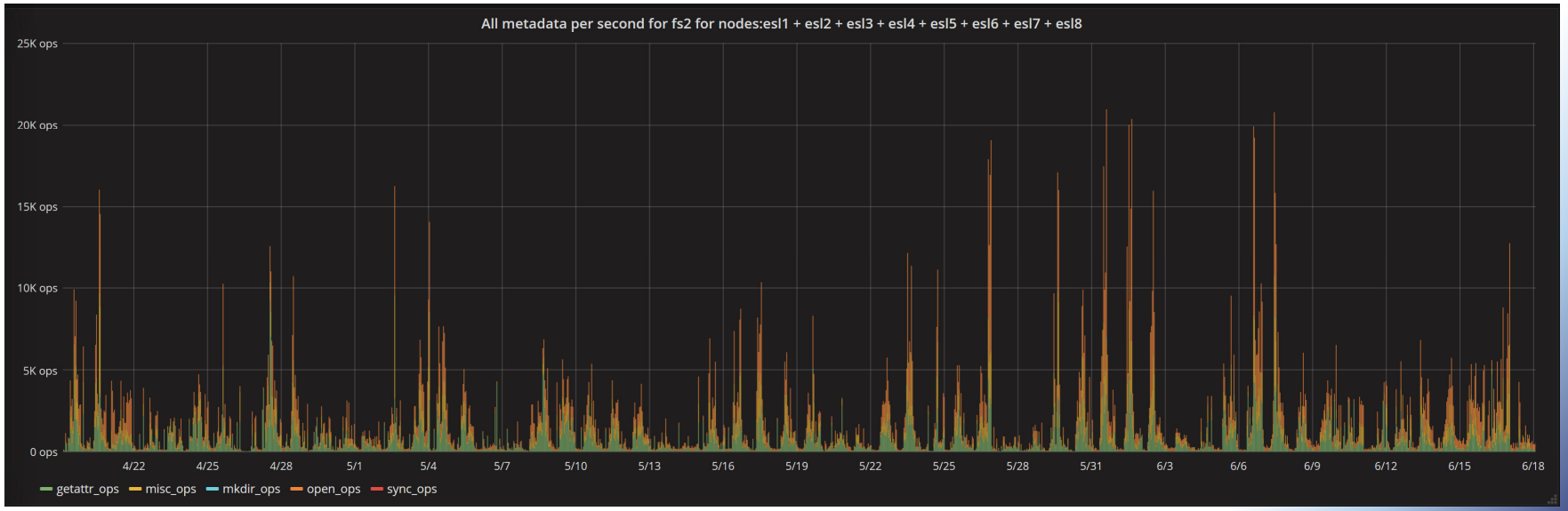
ARCHER post processing metadata



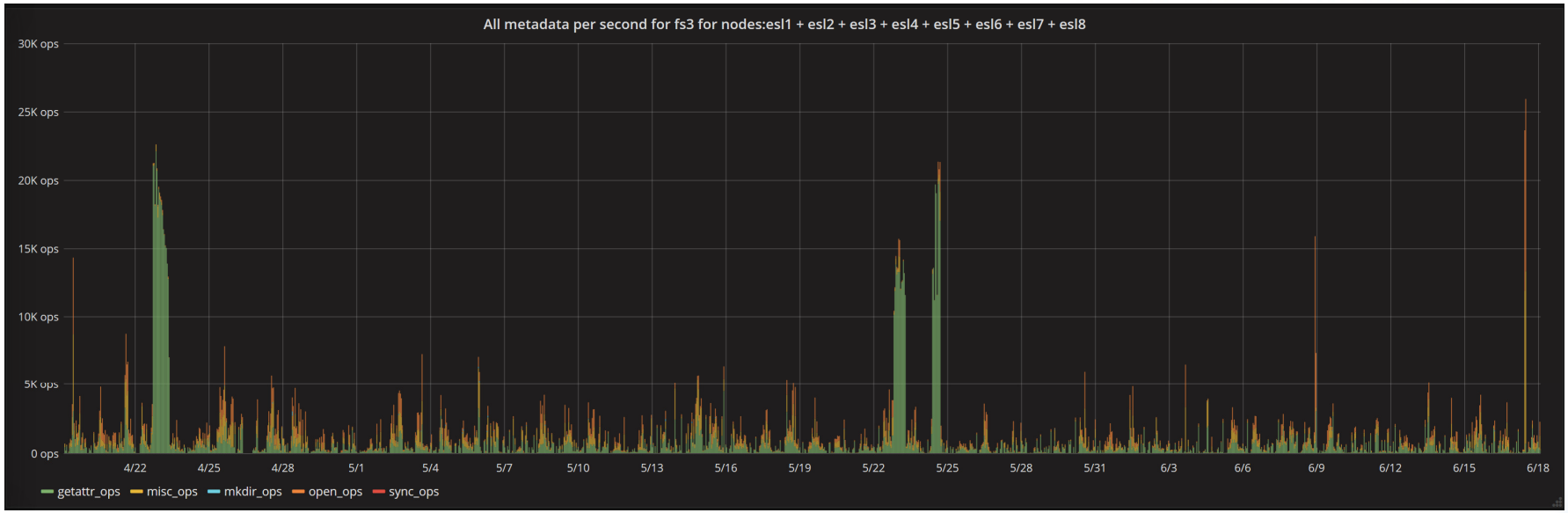


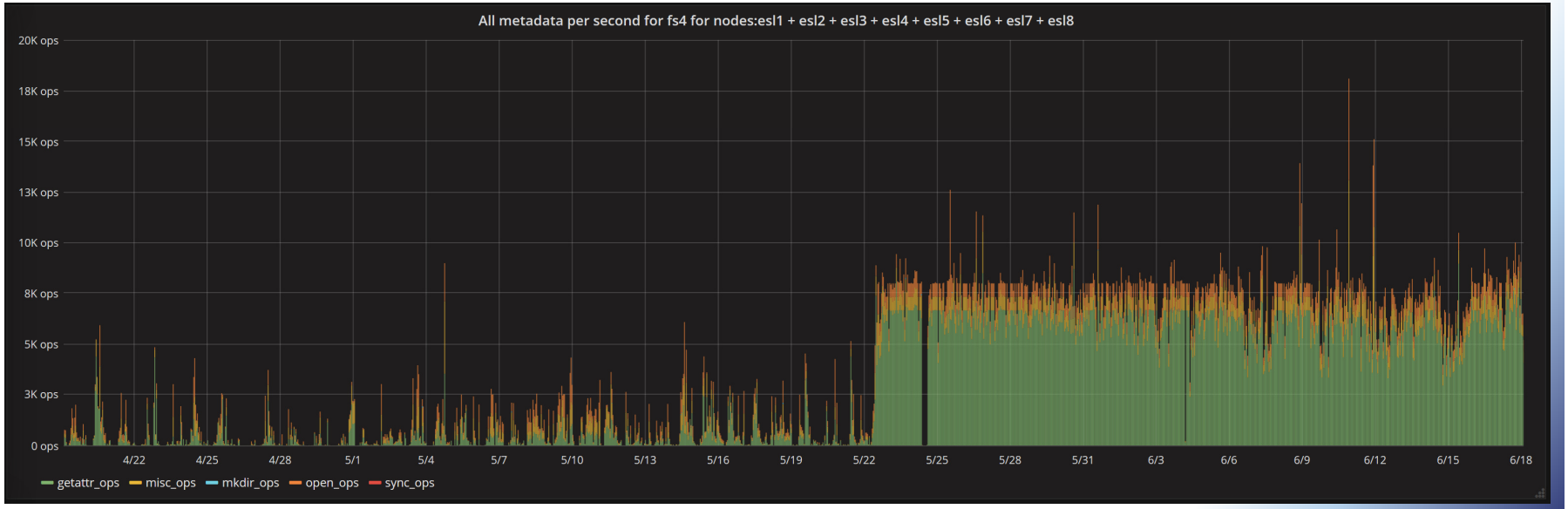
ARCHER post processing metadata





ARCHER login metadata





Summary

- Clearly there's a lot of variability in the performance/load
- Different application types have different profiles
- Depending on your I/O functionality you'll hit different bottlenecks
  - Metadata
  - Throughput
  - Contention on disks
  - …..

Exercise

- Two different exercises
  - Analyse
    - “Pen and paper”
      - Looking at benchmark performance and I/O logs to try and understand the variation/performance levels
  - Benchio
    - Run I/O benchmark that does different types of I/O:
      - Serial
      - MPI-I/O
      - NetCDF
      - HDF5
    - Run at different core counts, see what performance is achieved and what variations