

Ellexus is the I/O profiling company. Our customers manage large compute clusters and peta-bytes of shared storage and are experts in good I/O patterns, but often their users have expertise in other areas and do not know how to optimise their applications for the IT infrastructure.

Unless users of distributed compute and storage have access to I/O profiling tools they can easily overload the file-system with bad I/O patterns, wasting a huge amount of engineering time.

Make every engineer an I/O expert

Breeze Healthcheck generates a simple report that all compute users can understand and use to optimise the way they access shared storage. It needs no configuration or training so users can get into the habit of checking their workflows before running them at scale.

Breeze Healthcheck is built on top of Breeze HPC, a detailed I/O profiling and dependency analysis tool. Breeze Healthcheck gives you a list of all the files used by your application as well as a simple report that looks for certain bad I/O patterns.

Example checks included in the Breeze Healthcheck I/O report:

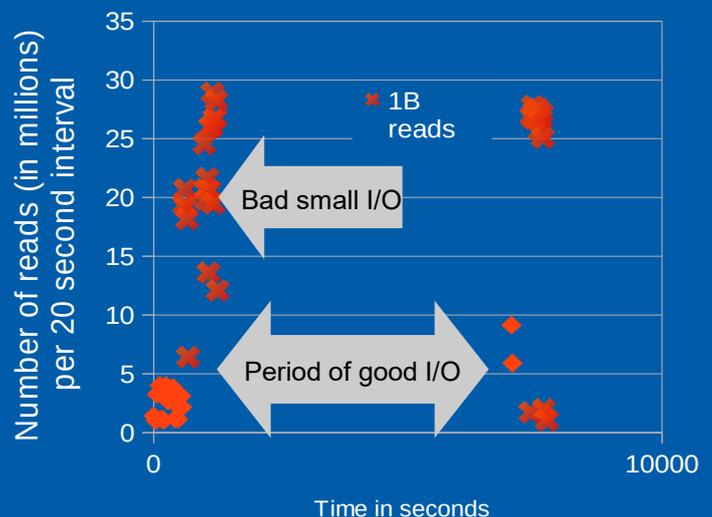
- Files or programs used in someone else's home directory and other hard coded paths that shouldn't be there.
- Lots of small temporary files saved on shared storage or not deleted.
- Programs that make very small reads and writes or very large reads and writes
- Programs that stat() or open() lots of files without using them.

Is my program slow or is the file system?

The Breeze Healthcheck report contains information about how much time is spent carrying out I/O so you can see how the performance of the file system and network is affecting your application.

Case study: Profiling genome pipelines at the Sanger Institute

Ellexus profiled one of the public genome pipelines at the Sanger Institute to look for I/O patterns that can harm the performance of shared storage. The pipeline had been optimised in some areas, but our profiling showed that there was still potential for it to be improved.



At the start there were up to a million 1 byte reads per second. These small reads harm the computational performance and create sub-optimal I/O patterns on the shared storage. This soon settles down to a longer period of good streaming I/O, but it would be worth optimising the early small reads.