Darshan: Improving I/O performance for scientific applications

Application:

- **Darshan** collects concise I/O access pattern information from large-scale applications

Goal:

- *Users*: improve the performance of critical scientific applications
- *Administrators*: gain insight into storage system deployments and usage
- *Researchers*: guide future research directions HPC I/O

Requirement:

- Negligible impact on production applications
  - Transparency at leadership scale
  - Enable automatically for all users
- Rapid feedback on the behavior of applications
  - User-friendly analysis tools for users to inspect I/O performance
  - Capture information from multiple levels of the I/O software stack
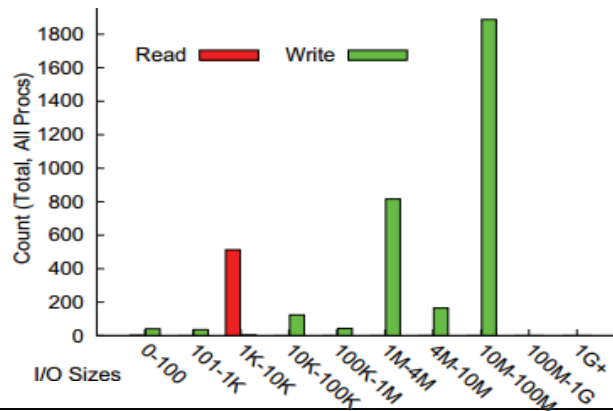
Histogram of I/O access sizes in a FLASH plot file. This data illustrated the presence of unexpected small write operations. The file layout was optimized accordingly to reduce I/O time by one third [1].

Results

- Darshan is deployed and automatically enabled for all users at ALCF and NERSC facilities.
- Darshan has been used to successfully tune a wide variety of scientific applications.
- The example shown here is from the FLASH application. FLASH can be used to solve a wide range of astrophysical, fluid dynamics, and plasma physics problems. Improvements in FLASH's I/O performance help to enhance scientific productivity.

[1] Latham et. al. A case study for scientific I/O: improving the FLASH astrophysics code. In Computational Science & Discovery, 2012