

# 21st International Conference on Computing in High Energy and Nuclear Physics (CHEP2015)



Contribution ID: 171

Type: oral presentation

## ATLAS I/O Performance Optimization in As-Deployed Environments

*Monday, April 13, 2015 6:00 PM (15 minutes)*

I/O is a fundamental determinant in the overall performance of physics analysis and other data-intensive scientific computing. It is, further, crucial to effective resource delivery by the facilities and infrastructure that support data-intensive science. To understand I/O performance, clean measurements in controlled environments are essential, but effective optimization requires as well an understanding of the complicated realities of as-deployed environments. These include a spectrum of local and wide-area data delivery and resilience models, heterogeneous storage systems, matches and mismatches between data organization and access patterns, multi-user considerations that may help or hinder individual job performance, and more.

The ATLAS experiment has organized an interdisciplinary working group of I/O, persistence, analysis framework, distributed infrastructure, site deployment, and external experts to understand and improve I/O performance in preparation for Run 2 of the Large Hadron Collider. The adoption of a new analysis data model for Run 2 has afforded the collaboration a unique opportunity to incorporate instrumentation and monitoring from the outset.

This paper describes a program of instrumentation, monitoring, measurement, and data collection both in cleanroom and grid environments, and discusses how such information is propagated and employed.

The paper further explores how these findings inform decision-making on many fronts, including persistent data organization, caching, best practices, framework interactions with underlying service layers, and settings at many levels, from sites to application code. Related developments that increase robustness and resilience in the presence of faults by improving communication between frameworks and underlying infrastructure layers are also discussed.

**Primary author:** Dr MALON, David (Argonne National Laboratory (US))

**Presenter:** MAIER, Thomas (Ludwig-Maximilians-Univ. Muenchen (DE))

**Session Classification:** Track 3 Session

**Track Classification:** Track3: Data store and access