Contribution ID: **464**   Type: **oral presentation**

Large Scale Management of Physicist's Personal Analysis Data

Tuesday 14 April 2015 15:15 (15 minutes)

The ability of modern HEP experiments to acquire and process unprecedented amounts of data and simulation have led to an explosion in the volume of information that individual scientists deal with on a daily basis. This explosion has resulted in a need for individuals to generate and keep large "personal analysis" data sets which represent the skimmed portions of official data collections pertaining to their specific analysis. These personal analysis and simulation sets represent a significant reduction in size compared to the original data, but they can still be many terabytes or tens of terabytes in size and consist of tens of thousands of files. When this personal data is aggregated across the many physicists in a single analysis group or experiment it can represent data volumes on par with or exceeding the official "production" samples which require special data handling techniques and storage systems to deal with effectively.

In this paper we explore the toolsets, analysis models and changes to the Fermilab computing infrastructure which have been developed and deployed by the NOvA experiment to allow experimenters to effectively manage their personal analysis data and other data that falls outside of the typically centrally managed production chains. In particular we describe the models and tools that are being used to allow NOvA to leverage Fermilab storage resources that are sufficient to meet their analysis needs, without imposing management burdens of specific quotas on users or groups of users, without relying on traditional central disk facilities and without having to constantly police individuals users usage. We discuss the storage mechanisms and the caching algorithms that are being used as well as the toolkits that have been developed to allow the users to easily operate with terascale+ datasets.

Primary author: Dr NORMAN, Andrew (Fermilab)

Co-authors: MENGEL, Marc (Fermilab); Dr TAMSETT, Matthew (University of Sussex); Dr GROUP, Robert (University of Virginia); Dr ILLINGWORTH, Robert (Fermilab)

Presenter: Dr NORMAN, Andrew (Fermilab)

Session Classification: Track 5 Session

Track Classification: Track5: Computing activities and Computing models