21st International Conference on Computing in High Energy and Nuclear Physics **CHEP2015** Okinawa Japan: April 13 – 17, 2015

Contribution ID: **214**   Type: **poster presentation**

Data Handling with SAM and ART at the NOvA Experiment

During operations, NOvA produces between 5,000 and 7,000 raw files per day with peaks in excess of 12,000. These files must be processed in several stages to produce fully calibrated and reconstructed analysis files. In addition, many simulated neutrino interactions must be produced and processed through the same stages as data. To accommodate the large volume of data and Monte Carlo, production must be possible both on the Fermilab grid and on off-site farms, such as the ones accessible through the Open Science Grid.

To handle the challenge of cataloging these files and to facilitate their off-line processing, we have adopted the SAM system developed at Fermilab. SAM indexes files according to metadata, keeps track of each file's physical locations, provides dataset management facilities, and facilitates data transfer to off-site grids.

To integrate SAM with the Fermilab's ART software framework and the NOvA production workflow, we have developed methods to embed metadata into our configuration files, ART files, and standalone ROOT files. A module in the ART framework propagates the embedded information from configuration files into ART files, and from input ART files to output ART files, allowing us to maintain a complete processing history within our files. Embedding metadata in configuration files also allows configuration files indexed in SAM to be used as inputs to Monte Carlo production jobs. Further, SAM keeps track of the input files used to create each output file. Parentage information enables the construction of self-draining datasets which have become the primary production paradigm used at NOvA. We will present an overview of SAM at NOvA and how it has transformed the file production framework used by the experiment.

Primary author: AURISANO, Adam (University of Cincinnati)

Co-authors: NORMAN, Andrew (Fermilab); BACKHOUSE, Christopher; ROCCO, Dominick (University of Minnesota); DAVIES, Gavin (Iowa State University); ZIRNSTEIN, Jan (University of Minnesota); MENGEL, Marc (Fermilab); MAYER, Nathan (Tufts University); Dr ILLINGWORTH, Robert (Fermilab)

Presenter: AURISANO, Adam (University of Cincinnati)

Track Classification: Track3: Data store and access