Data Management in the ICARUS Collaboration

Yun-Tse Tsai

September 14, 2018

1 Definition

The data management system is expected to faithfully transfer the data collected by the data acquisition system to permanent storage, to catalog the data, and to provide collaborators with the access to the data. The team responsible for data management shall develop a system automatically transferring data during the operation of the experiment, and ensuring the integrity of the data in the permanent storage. In addition, the team shall establish a system to track the data, such as metadata in the SAM database, and make the data accessible to the collaboration. It is critical to perform those function at the rate of data production and the availability of resources.

To achieve the goals, coordination between the data management team and the scientific computing division is required. The team may utilize an existing, generic tool kit and customize to meet the specific requirement of the ICARUS experiment. It is also essential to coordinate with the data acquisition group, and possibly the trigger, beam interface groups, to obtain all the relevant data to be archived.

2 Deliverables

The deliverables of the data management task are listed below.

- A system reliable transferring the data acquired real-time to permanent storage and cataloging the files.
  - The system should be capable of transferring the data with a bandwidth in ecessive of the DAQ rate by 33%.
  - Throughput of the network has to be identified and stress tested.
  - The data files in permanent storage have to be cataloged. For example, a appropriate format of metadata should be defined.
  - Monitors of the resource availability and back pressure have to be implemented.
  - An automatic alerting system of back pressure should be established.

- A mechanism integrating the neutrino beam data obtained from the IFBeam database with the data collected at the detector. (Coordinate with the beam interface group.)
- The procedure of maintaining and backing up the databases recording information in real-time, such as EPIC database for slow control, run configuration database, and event database. (Coordinate with the DAQ, slow control group, and the Scientific Linux Administration Management team in Scientific Computing Division.)
- A system providing reconstructed data and accessible by analyzers. (Coordinate with the offline production team.)
- (Open to discussion) A shifter-friendly online run catalog.
  The database may collect the essential configuration of the detector, such as the cathode voltage, the gain and shaping time of the TPC electronics, and provide some plots for shifters to check. Such a catalog is deployed in Minerνa and MicroBooNE.
- (Open to discussion) A data quality monitor. (Coordinate with the DAQ and offline software groups.)

Notes

- Most of the development and maintenance efforts of data management can be done off-site.

Acknowledgement

I thank Angela Fave and Michael Kirby (MicroBooNE data management convener) for their comments on this document.