Recent Evolution of the Offline Computing Model of the NOvA Experiment

Alec Habig
University of Minnesota Duluth, Duluth MN, USA
E-mail: ahabig@umn.edu

Craig Group
University of Virginia, Charlottesville VA, USA
E-mail: group@virginia.edu

A. Norman
Fermi National Accelerator Laboratory, Batavia IL, USA
E-mail: anorman@fnal.gov

Abstract. The NOvA experiment at Fermilab is a long-baseline neutrino experiment designed to study $\nu_e$ appearance in a $\nu_\mu$ beam. Over the last few years there has been intense work to streamline the computing infrastructure in preparation for data, which started to flow in from the far detector in Fall 2013. Major accomplishments for this effort include migration to the use of off-site resources through the use of the Open Science Grid and upgrading the file-handling framework from simple disk storage to a tiered system using a comprehensive data management and delivery system to find and access files on either disk or tape storage. NOvA has already produced more than 6.5 million files and more than 1 PB of raw data and Monte Carlo simulation files which are managed under this model. The current system has demonstrated sustained rates of up to 1 TB/hour of file transfer by the data handling system. NOvA pioneered the use of new tools and this paved the way for their use by other Intensity Frontier experiments at Fermilab. Most importantly, the new framework places the experiment’s infrastructure on a firm foundation, and is ready to produce the files needed for first physics.

1. Introduction

There are many open and exciting questions regarding neutrinos [1]. At the top of the list, the fundamental question of the mass ordering of the neutrino mass eigenstates is unknown. That is, if we consider the mass state that is predominantly composed of the electron neutrino, we do not yet know if this is the lightest mass state (the normal case), or if it is actually heavier than the state with the smallest component of electron neutrino (the inverted case). Another important question regards the CP-violating phase in the neutrino mass mixing matrix: we don’t yet know if this phase is non zero, and if it is non zero we don’t know if it is large enough to play an important role in the development of the matter/antimatter asymmetry of the universe. Recently, due to precise measurements of $\theta_{13}$ [2, 3], $\theta_{23}$ (the first of the three measured mixing
angles) has become the least precisely known mixing angle. Moving into potential new physics, perhaps there are more than three generations of neutrino species. The existence of a fourth generation (which would have to be a “sterile” neutrino due to constraints from the width of the $Z^0$ boson [4–6]) is well motivated and would also be a major discovery. Ongoing and future neutrino experiments will attempt to shed light on all of these interesting unknowns.

The NOvA experiment [7] hopes to address many of the questions mentioned above. NOvA recently finished commissioning the experiment which consists of a near detector at Fermilab and a 14 kiloton far detector located 810 km away in Ash River, MN and slightly off axis relative to the NuMI beam line. The detectors are composed of alternating horizontal and vertical planes of PVC cells filled with liquid scintillator and outfitted with wavelength shifting fiber readout by avalanche photo diodes. The far detector started collecting data in Fall 2013 and the commissioning period was completed in Fall 2014. First results are expected from NOvA in summer 2015.

NOvA follows a traditional HEP computing strategy where event-driven analysis access centrally-stored data and Monte Carlo simulated events by distributing data to parallel jobs running on grid computing clusters [8]. Like many other experiments, NOvA also relies heavily on ROOT [9] and GEANT4 [10]. NOvA uses an analysis framework shared by several other experiments at Fermilab called art [11–13]. An important goal of art, and an important feature for NOvA, is seamless integration with the data-handling tools available at Fermilab. NOvA further utilizes many other centralized resources and services. At Fermilab, FIFE (Fabric for Frontier Experiments) manages and provides access to shared resources and tools [14,15]. FIFE provides access and support for a comprehensive set of services at Fermilab that NOvA benefits from: DAQ and controls; job submission [16], grid and cloud resources; scientific data storage, access, and management, scientific frameworks and software; physics and detector simulation; databases; and scientific collaborative tools. NOvA’s use of many of these services will be discussed below. NOvA and other experiments at Fermilab benefit from this strategy to share resources and services [17].

The demand from NOvA for computing resources is large. Last year NOvA used about 15 million CPU hours and has already written more than 2 petabytes of files to tape. This total represents more than 5 million files, and raw data alone produces about 5,000 new files per day. In comparison, by the end of the experiment the total NOvA file collection could be comparable in size to everything from the Tevatron experiments (perhaps as large as 20 PB).

The NOvA offline computing model has undergone a major paradigm shift in preparation for file production for physics analysis. Describing these changes and recent successes of the effort will be the topic of this paper.

2. Transition of the NOvA Computing Paradigm

Three years ago NOvA used a large networked disk (BlueArc) for all storage and ran all batch jobs locally on cores with direct mounts available to the BlueArc disk. As the computational needs of the experiment increased it became clear that this system was not scalable.

One major change that NOvA made to improve the situation was to transition to a data handling system based on SAM [18–21]. SAM stands for Sequential data Access via Metadata and is a system that was successfully used by the CDF and D0 experiments at the Tevatron. SAM largely consists of a database serving as a per-file metadata catalog. For example, metadata may include file name, file size, number of events, run information, luminosity, details of Monte Carlo simulation, etc. Based on this metadata, queries and dataset definitions can be conducted in a flexible and robust way. SAM coordinates and manages the data movement to jobs using the desired file transfer protocol [22] (gridftp, dccp, SRM, …). SAM also interfaces with cache management, currently based on dCache as a front end to the tape system, and SAM tracks file consumption and job success providing tools for easy job recovery. The NOvA production group
modified all of their workflows to use SAM for all of the data handling needs. Input datasets are defined using SAM metadata and output files are always registered in SAM.

Another major transition for NOvA was to add the ability to run batch jobs on off-site resources. NOvA now has the capability to distribute its code using CernVM-FS [23] and the SAM-based data handling system functions well off-site. Specifically, NOvA is now capable of running on the Open Science Grid (OSG) [24] and the Amazon Cloud [25].

3. Summary of NOvA’s Computing Infrastructure

A SAM-based data handling system would not succeed without a well-planned underlying infrastructure. At NOvA typical users do most of their interactive work on virtual machines. These machines have access to the NOvA software and are available for login use [26]. They also have access to 335 TB of BlueArc disk (NFS-based NAS). The BlueArc disk is useful project space and provides storage for short-term or small data sets. For larger efforts users submit jobs to Fermigrid, a shared computing resource at Fermilab with approximately 15,000 cores. The current NOvA quota is 1300 cores but opportunistic slots are also available: at times NOvA receives up to 5000 cores on Fermigrid. SAM is used to handle data for batch jobs. There is a large shared cache disk (about 4 PB of dCache) that serves as a front end to the tape system. The tape library serves as permanent data storage. Several PostgreSQL databases are required for online and offline operations and these can be accessed via http servers for ease of off-site usage.

4. NOvA File Production Steps

The file production steps of the NOvA experiment are shown in Fig. 1. The detector data flow starts on the top left while the simulated data [27] starts on the top right. At each production step, a new data set is produced. These steps are conducted for several different data types: near and far detector; cosmic rays; neutrino beams with varying horn current; etc. After the reconstruction step [28] more advanced particle identification steps are performed (PID and
Library Event Matching, LEM [29, 30]) before the small analysis ntuples (“CAF” [31]) files are produced. One small caveat: the LEM process is more memory intensive than the other steps, so is done off-site at Caltech where the grid nodes have more available memory. Note that analysis can be done either with CAF files in ROOT, or with the output of the reconstruction, PID, or LEM jobs in the art framework.

5. Resource Requirements
In fall of 2013 NOvA conducted a workshop in preparation for the first data analysis to understand resource requirements and streamline file production tools. A production run on NOvA consists of reprocessing all of the data and producing the full set of MC needed for an oscillation analysis, and it is anticipated that this will need to be done about twice per year. In total a production run requires about 250 TB of tape storage and about 10,000 CPU days of processing. So, if the goal is to be able to conduct a production run in a reasonably short period of time (∼10 days), it must be possible to pass about 1 TB per hour through the file transfer system and one must have access to at least 1000 CPUs dedicated to the production effort.

6. Recent File Production Efforts
The current CPU quota for NOvA is 1300 cores on Fermigrid and usage has actually averaged over 2000 cores for the last three-month period (see Fig 2 for CPU usage over the last 12 months), with opportunistic use of shared resources making up the difference. In addition to that, production jobs are also run on OSG. Thus, while NOvA’s CPU needs have increased significantly over the last year, so far CPU has not been a major limiting factor. A recent grant has been obtained with the goal to demonstrate production scale processing on the Amazon cloud [25]. Using commercial cloud resources to satisfy burst computing needs will take some of the pressure off of local CPU resources.

![Figure 2. NOvA’s usage on Fermigrid over the last 12 months. Usage has increased significantly in preparation for the first physics results, but resources have been available to meet that need.](image-url)

With thousands of batch jobs running on grid resources, output files must be passed through the file transfer system (FTS) at a reasonable rate or this step will become the processing bottleneck. The file transfer system at NOvA currently consists of three servers that monitor drop...
boxes, copy files to dCache locations (then to tape), and register their locations in SAM. Before files can be accessed by an analysis or the next processing step, they must be registered in SAM. The combined FTS servers have demonstrated throughput of greater than 1 TB per hour and sustained throughput over several hours of greater than 0.5 TB per hour.

Recently, a new suite of tools for validating all steps of the production effort was created [32]. For every new software release, the validation tool checks all file production steps with a set of automated scripts and publishes the results to a web page. Failures are reported and metrics for each production step and the resulting output data can be compared with past validation checks. For example, file output size, job size, and memory usage are recorded. These metrics are also very useful for predicting the resource needs of any new request to the production group. This new tools has streamlined predicting resource needs.

Two recent production runs have demonstrated the successful transition to SAM. In spring of 2014 we had the first production run that included a substantial far detector data set and SAM was fully integrated into all file production steps. There were challenges with the new system, but in the end all steps of production ran in time to produce files for the “Neutrino 2014” conference. Currently ongoing is the winter/spring 2015 effort to produce files to be used for the first oscillation analyses from NOvA. With the pressure of first physics came many new requests including keep-up processing, new calibration requests, systematic samples, etc. Overall, the SAM paradigm has functioned well in the current intense production efforts.

7. Summary and Outlook
The increasing computational requirements of the NOvA experiment demanded a scalable data handling system and a flexible set of CPU resources. Over the last two years, NOvA migrated from a simple central disk storage model to a tiered model composed of disk, cache disk, and tape where SAM is used to track file locations. SAM provides tools for a flexible and scalable data handling system where file metadata can be used to define datasets for processing. In addition, NOvA added CPU resources by demonstrating the ability to run jobs on the Open Science Grid and is working to incorporate the Amazon Cloud as a flexible resource for times of increased CPU needs. The NOvA computing infrastructure is now on a firm foundation that should successfully provide the files required for physics analysis for years to come.

Acknowledgments
Support for the presenter was provided by the National Science Foundation, NSF RUI grant #1306944. The authors acknowledge that support for this research was carried out by the Fermilab scientific and technical staff. Fermilab is Operated by Fermi Research Alliance, LLC under Contract No. De-AC02-07CH11359 with the United States Department of Energy. The University of Virginia particle physics group is supported by DE-SC0007838 also from the Department of Energy.

References
[1] A. de Gouvea et al. [Intensity Frontier Neutrino Working Group], arXiv:1310.4340.
[2] F. P. An et al. [Daya Bay Collaboration], Phys. Rev. Lett. 108, 171803 (2012).
[3] J. K. Ahn et al. [RENO Collaboration], Phys. Rev. Lett. 108, 191802 (2012).
[4] P. Abreu et al. [DELPHI Collaboration], Nucl. Phys. B 367, 511 (1991).
[5] B. Adeva et al. [L3 Collaboration], Z. Phys. C 51, 179 (1991).
[6] G. Alexander et al. [OPAL Collaboration], Z. Phys. C 52, 175 (1991).
[7] D. S. Ayres et al. [NOvA Collaboration], FERMILAB-DESIGN-2007-01.
[8] B. Rebel, M. C. Sanchez and S. Wölfers, arXiv:1310.6964.
[9] R. Brun and F. Rademakers, Nucl. Instrum. Meth. A 389, 81 (1997).
[10] R. Brun, F. Carminati and S. Giani, CERN-W5013, CERN-W-5013.
[11] C. Green and J. Kowalkowski and M. Paterno and M. Fischler and L. Garren and others, J. Phys. Conf. Ser.,396,022020 (2012).
[12] https://web.fnal.gov/project/ArtDoc/Pages/home.aspx

[13] T. Arvanitis and A. Lyon, J. Phys. Conf. Ser. 513, 022023 (2014).

[14] P. Mhashilkar, Advances in Distributed High Throughput Computing for the Fabric for Frontier Experiments Project at Fermilab, in these proceedings.

[15] M. Kirby, J. Phys. Conf. Ser. 513, 032049 (2014).

[16] D. Box, J. Phys. Conf. Ser. 513, 032010 (2014).

[17] C. Group, S. Fuess, O. Gutsche, M. Kirby, R. Kutschke, A. Lyon, A. Norman, G. Perdue, and E. Sexton-Kennedy, Computing at the Intensity Frontier, in these proceedings.

[18] R. A. Illingworth, J. Phys. Conf. Ser. 513, 032045 (2014).

[19] A. L. Lyon, R. A. Illingworth, M. Mengel and A. J. Norman, J. Phys. Conf. Ser. 396, 032069 (2012).

[20] A. Aurisano, C. Backhouse, G. S. Davies, R. Illingworth, N. Mayer, M. Mengel, A. Norman, D. Rocco, and J. Zirnstein, Data Handling with SAM and ART at the NOvA Experiment, in these proceedings.

[21] A. Norman, Large Scale Management of Physicist’s Personal Analysis Data, in these proceedings.

[22] A. L. Lyon and M. W. Mengel, J. Phys. Conf. Ser. 513, 032068 (2014).

[23] P. Buncic, C. Aguado Sanchez, J. Blomer, L. Franco, A. Harutyunian, P. Mato and Y. Yao, J. Phys. Conf. Ser. 219, 042003 (2010).

[24] R. Pordes, B. Kramer, D. Olson, M. Livny, A. Roy, P. Avery, K. Blackburn and T. Wenaus et al., J. Phys. Conf. Ser. 78, 012057 (2007).

[25] P. Mhashilkar, Large Scale Monte Carlo Simulation of neutrino interactions using the Open Science Grid and Commercial Clouds, in these proceedings.

[26] G. Davies, J. Davies, B. Rebel, K. Sachdev, and J. Zirnstein, Software Management for the NOvA Experiment, in these proceedings.

[27] A. Aurisano, The NOvA Simulation Chain, in these proceedings.

[28] M. Baird, J. Bian, M. Messier, E. Niner, D. Rocco, K. Sachdev, Event Reconstruction Techniques at NOvA, in these proceedings.

[29] C. Backhouse and R. B. Patterson, Nucl. Instrum. Meth. A 778, 31 (2015).

[30] C. Backhouse, R. Patterson, and D. Rocco, The Library Event Matching classifier for electron neutrino events at NOvA, in these proceedings.

[31] C. Backhouse, and D. Rocco, A Data Summary File Structure and Analysis Tools for Neutrino Oscillation Analysis at the NOvA Experiment, in these proceedings.

[32] M. Tamsett, Software Framework Testing at the Intensity Frontier, in these proceedings.