Data processing and storage in the Daya Bay Reactor Antineutrino Experiment postprint

Authors: He, Miao

Date: 2016-09-07T00:00:00+00:00

Abstract

The Daya Bay Reactor Antineutrino Experiment reported the first observation of the non-zero neutrino mixing angle 13 using the first 55 days of data. It has also provided the most precise measurement of 13 with the extended data to 621 days. Daya Bay will keep running for another 3 years or so. There is about 100 TB raw data produced per year, as well as several copies of reconstruction data with similar volume to the raw data for each copy. The raw data is transferred to Daya Bay onsite and two offsite clusters: IHEP in Beijing and LBNL in California, with a short latency. There is quasi-real-time data processing at both onsite and offsite clusters, for the purpose of data quality monitoring, detector calibration and preliminary data analyses. The physics data production took place a couple of times per year according to the physics analysis plan. This paper will introduce the data movement and storage, data processing and monitoring, and the automation of the calibration.

Full Text

Preamble

Nuclear Physics B Proceedings Supplement 00 (2015) 1-5

Data Processing and Storage in the Daya Bay Reactor Antineutrino Experiment

Miao He
Institute of High Energy Physics, Beijing
On behalf of the Daya Bay collaboration

Abstract

The Daya Bay Reactor Antineutrino Experiment reported the first observation of the non-zero neutrino mixing angle 13 using the initial 55 days of data collection. With extended data spanning 621 days, the experiment has also provided the most precise measurement of 13 to date. Daya Bay will continue operations for approximately three more years, producing about 100 TB of raw data annually, along with several copies of reconstruction data of comparable volume for each copy. Raw data is transferred to the Daya Bay onsite facility and to two offsite clusters—the Institute of High Energy Physics (IHEP) in Beijing and Lawrence Berkeley National Laboratory (LBNL) in California—with minimal latency. Quasi-real-time data processing occurs at both onsite and offsite clusters to support data quality monitoring, detector calibration, and preliminary analyses. Physics data production is performed a few times per year according to the physics analysis schedule. This paper describes the data movement and storage infrastructure, data processing and monitoring systems, and the automation of calibration procedures.

Keywords: Daya Bay, reactor, neutrino, data processing

1. Introduction

Neutrinos are elementary particles in the Standard Model that exist in three flavors: e, , and . They are produced through nuclear fusion in the sun, -decays of radioactive elements within the Earth, remnants of the Big Bang, supernova explosions, and interactions between cosmic rays and Earth' s atmosphere. In addition to these natural sources, neutrinos can be artificially generated in nuclear reactors and particle accelerators. Neutrino flavor states are superpositions of three mass eigenstates (1, 2, and 3), and one flavor can transform into another due to quantum interference among these mass eigenstates during propagation. This phenomenon is known as neutrino oscillation or neutrino mixing, characterized by mixing angles 12, 23, and 13, with oscillation frequencies determined by the differences in squared neutrino masses, Δm^2ji. For reactor-based experiments, 13 can be extracted from the survival probability of electron antineutrinos at distances of 1-2 km from the reactors:

$$P(\nu e \to \nu e) \approx 1 - \sin^2 2\theta 13 \sin^2(1.27 \Delta m^2 31 L/E)$$

where E is the e energy in MeV and L is the baseline distance in meters between the neutrino source and detector.

2. The Daya Bay Experiment

The Daya Bay experiment was designed to achieve the most precise measurement of 13 among existing and near-future experiments, with a sensitivity to $\sin^2 2$ 13 < 0.01 at 90% confidence level [1]. Situated near the Daya Bay, Ling Ao, and Ling Ao-II nuclear power plants in southern China, approximately 45 km from Shenzhen city, the experiment comprises six functionally identical reactor cores grouped into three pairs. Three underground experimental halls (EHs) are connected via horizontal tunnels. Two antineutrino detectors (ADs) are located in the Daya Bay near hall (EH1), two in the Ling Ao and Ling Ao-II near hall (EH2), and four near the oscillation maximum in the far hall (EH3). Each AD contains 20 tons of gadolinium-doped liquid scintillator (Gd-LS) as the target, surrounded by 20 tons of undoped liquid scintillator (LS) to detect gamma rays that escape the target. Electron antineutrinos interact with protons via the inverse -decay (IBD) reaction in Gd-LS, producing a positron and a neutron. The prompt signal from positron ionization and the delayed signal from neutron capture on Gd, separated by approximately 30 s, provide a distinctive time coincidence signature. The muon veto system in each hall consists of an inner water shield, an outer water shield, and a resistive plate chamber (RPC). A detailed description of the Daya Bay experiment can be found in Ref. [2].

Daya Bay began data acquisition in late 2011 with six ADs, with the final two ADs installed in summer 2012. The data acquisition software (DAQ) [3] supports multiple partitions, enabling simultaneous data taking in all three experimental halls and outputting multiple data streams. A typical physics run in each hall lasts 48 hours, interspersed with pedestal runs and electronics diagnostic runs. The typical trigger rates are 1.3 kHz in EH1, 1.0 kHz in EH2, and 0.6 kHz in EH3. Approximately 320 raw data files are produced daily, each about 1 GB in size. In 2013, the data taking efficiency exceeded 97%, with physics data taking time exceeding 95%. Based on the first 55 days of data, Daya Bay reported the first observation of the non-zero neutrino mixing angle 13 with 5.2 standard deviations [4]. The precision of sin²2 13 has been continuously improved with increased statistics, with the latest result based on 621 days of data achieving a precision of 5.6% [5].

3. Data Processing and Storage in Daya Bay

Figure 2 [Figure 2: see original paper] presents an overview of the Daya Bay data processing workflow. Raw data is first transferred to the onsite farm, then to the Institute of High Energy Physics (IHEP) in Beijing and Lawrence Berkeley National Laboratory (LBNL) in California for central storage and processing. Metadata from the online database is also transferred to the offline database. A Performance Quality Monitoring (PQM) system [6] runs onsite using fast reconstruction algorithms to monitor physics performance with a latency of approximately 40 minutes. Keep-up data processing begins as soon as data arrives at IHEP or LBNL, employing full reconstruction with the latest calibration constants. Detector monitoring plots generated during this process are published through an Offline Data Monitoring (ODM) system with a latency of about 3 hours. Extracted data quality information is stored in a dedicated database for long-term monitoring.

3.1. Data Movement and Storage

Data movement is controlled by a set of Java applications, with the primary application called SPADE. Once a raw data file is closed and recorded in the online database by the DAQ, it is automatically copied to an offline file server at Daya Bay onsite. Upon completion of the transfer, SPADE updates the file status in both the online and offline databases to TRANSFERRED, allowing the file to be deleted from online disk. The file is then transferred sequentially to the two computing centers, IHEP and LBNL, with a maximum latency of 20 minutes. SPADE also creates an XML file containing raw data descriptions for file cataloging. Multiple tools monitor network traffic at the Daya Bay site and track the transfer rates of raw data files and data volume from Daya Bay to IHEP, and from IHEP to LBNL.

Raw data is stored at the Daya Bay onsite facility for one month and archived at IHEP and LBNL, with one disk copy and two tape copies at each site. Although the infrastructure of these two clusters differs, their computing and storage capabilities are comparable. Each cluster provides nearly 1 PB of disk space and approximately 800 CPU cores, including some shared resources. Additional computing resources have been planned to accommodate increasing data volumes. By mid-2014, accumulated raw data reached 320 TB.

3.2. Offline Software

The Daya Bay offline software, known as NuWa, is built on the Gaudi [7] framework and provides full functionality for simulation, reconstruction, and physics analysis. NuWa utilizes Gaudi’ s event data service as the data manager. Raw data and other offline data objects can be accessed from the Transient Event Store (TES). The prompt-delayed coincidence analysis requires time-window lookups, which are implemented through a specialized Archive Event Store (AES). All data objects in both TES and AES can be written to or read from ROOT [8] files via various Gaudi converters. To improve analysis efficiency, Daya Bay developed an alternative Lightweight Analysis Framework (LAF) that is compatible with NuWa data objects but offers higher I/O performance through simpler data conversion, lazy loading implementation, and a flexible cycling mechanism. LAF enables both backward and forward event access through a data buffer, which also facilitates data exchange among multiple analysis modules. The NuWa and LAF packages are available to collaborators through the Subversion (SVN) code management system [9].

The NuWa auto-build system was implemented using the Bitten [10] plugin for Trac [11], an enhanced wiki and issue tracking system for software development projects. Multiple servers at Daya Bay onsite, IHEP, LBNL, and other institutes function as Bitten slaves, automatically building NuWa whenever code is updated.

3.3. Database

The offline database comprises an onsite offline database, a central master database, and multiple local slave databases, as illustrated in Figure 3 [Figure 3: see original paper]. During data taking, information such as raw data descriptions and detector status is automatically extracted from online databases to the onsite offline database via a scraper. The onsite offline database is then synchronized to the central master database at IHEP, which also contains calculated reactor neutrino flux based on reactor information provided by the power company. Multiple slave databases replicate the central database at different institutes. Offline calibration constants produced by calibration experts are first inserted into a temporary table of a local slave database, then dumped into a formatted text file after validation. This text file is archived in the Subversion system. The database manager is responsible for inserting the calibration constants into the central database using the text file with a revision number specified by the calibration expert. This standard operating procedure ensures that calibration constant updates are highly controlled, carefully documented, and easily reproducible.

3.4. Calibration and Reconstruction

Various sources are used to calibrate the detector response. Dedicated calibration runs using LEDs and radioactive sources are taken weekly. Additional calibration samples, such as PMT dark noise and neutrons produced by cosmic rays, are selected from normal physics runs. Two categories of calibration are employed. The file-by-file track automatically accumulates calibration samples from physics runs and generates calibration constants. The run-by-run track uses a script to search all raw data from calibration runs and submits jobs to compute calibration constants. Reconstruction algorithms read calibration constants from the local slave database through a Database Interface (DBI). A timestamp called the rollback date is set to select the latest records inserted into the offline database before that time, enabling version control.

3.5. Onsite Data Processing and Monitoring

The PQM system running at the Daya Bay onsite facility provides quasi-real-time data processing and monitoring during data acquisition. Sixteen dedicated CPU cores are allocated to PQM, with 40 additional cores shared with users. A Python control script runs in background mode on a dedicated server, querying the onsite offline database at fixed intervals (10 seconds). When a new record is found and the query indicates that the corresponding raw data file is already on the file server, a processing job for the new file is submitted to the Portable Batch System (PBS). The PQM job uses a recently manually tagged NuWa version to reconstruct data with the latest calibration constants from the offline database, runs analysis algorithms, and fills results into user-defined histograms. All histograms are written to a ROOT file during the finalization step, which is then merged with the accumulated ROOT file for the same run. Subsequently, a C++ histogram printer processes the merged ROOT file to generate selected figures. The control script checks job completion at 10-second intervals by detecting an empty text file created at the job’ s conclusion. For completed jobs, the control script transfers the corresponding figures to PQM disk for web display. When the run ends and all associated raw data files have been processed, the control script saves the accumulated ROOT file on the PQM disk for permanent storage. The PQM data flow is shown in Figure 4 [Figure 4: see original paper]. The reconstruction and analysis modules in PQM are configurable; the default configuration requires approximately 30 minutes per file.

3.6. Offsite Data Processing and Monitoring

As soon as a raw data file arrives at IHEP or LBNL, an application called ingest in the data movement system submits a job using the full NuWa reconstruction with the latest calibration constants—this is known as keep-up (KUP) data processing. Additional analysis modules run to enable high-level data monitoring. Output histograms categorized by detector and channel are archived in ROOT files and rendered as figures. A multifunctional webpage has been developed using the Django framework, called the Offline Data Monitor (ODM). ODM' s primary function is to display plots produced by KUP, providing additional tools to assist data monitoring and analysis, such as comparison to reference plots, display of detector configuration and status, and interfaces to online and offline databases. Two example plots from ODM are shown in Figure 5 [Figure 5: see original paper], displaying the energy spectra of prompt and delayed signals from inverse -decay selected using AES during KUP. KUP jobs have the highest priority in the job system, typically occupying 30 CPU cores. The ODM latency is dominated by reconstruction and analysis algorithms.

Physics production uses validated and frozen calibration constants and reconstruction algorithms. It includes event tagging and filtering, providing both full data samples and various reduced samples to improve analysis efficiency. The reconstructed data volume is approximately 1.2 times the raw data volume, with only the latest version kept on disk while previous versions used for publications are archived on tape. Physics production occurs once or twice per year according to physics analysis requirements, with each production requiring about one month of processing time. A special production strategy was employed for the first two publications [4, 12] to rapidly obtain physics results. The offline software was fixed at the beginning, calibration constants were generated after weekly calibration data taking, followed by physics data production for the past week. Thanks to this weekly production strategy, Daya Bay completed the analysis and reported the first physics result only 20 days after the first stage of data acquisition.

3.7. Data Quality

During keep-up data processing, information related to data quality is extracted and stored in a dedicated database, which is automatically checked by an application for data quality evaluation. Additionally, online shifters record any data issues in the database. A data quality manager is responsible for tagging each raw data file based on both automatic and manual check results, providing a preliminary good file list. After physics data production, the final good file list used for physics publications is compiled based on feedback from analyzers. A web interface for the data quality database has been implemented for long-term monitoring of data quality, as shown in Figure 6 [Figure 6: see original paper], which displays a strip chart of the antineutrino candidate rate in one AD in EH1.

4. Summary

Daya Bay is a reactor antineutrino oscillation experiment that provided the first and most precise measurement of the neutrino mixing angle 13. The offline system enables rapid and stable data transfer, processing, and quality monitoring. Daya Bay will continue data taking until the end of 2017. Computing resources and offline software will be upgraded to accommodate data processing and storage requirements.

References

[1] X. Guo, et al., A precision measurement of the neutrino mixing angle 13 using reactor antineutrinos at Daya-Bay, (2007) arXiv:hep-ex/0701029.

[2] F. An, et al., A side-by-side comparison of Daya Bay antineutrino detectors, Nucl.Instrum.Meth. A685 (2012) 78–97. arXiv:1202.6181, doi:10.1016/j.nima.2012.05.030.

[3] F. Li, X.-L. Ji, X.-N. Li, K. Zhu, DAQ architecture design of Daya Bay reactor neutrino experiment, IEEE Trans.Nucl.Sci. 58 (2011) 1723–1727. doi:10.1109/TNS.2011.2158657.

[4] F. An, et al., Observation of electron-antineutrino disappearance at Daya Bay, Phys.Rev.Lett. 108 (2012) 171803. arXiv:1203.1669, doi:10.1103/PhysRevLett.108.171803.

[5] C. Zhang, Recent Results From The Daya Bay Experiment. URL http://neutrino2014.bu.edu/

[6] Y.-B. Liu, M. He, B.-J. Liu, M. Wang, Q.-M. Ma, et al., Onsite data processing and monitoring for the Daya Bay Experiment, Chin.Phys. C38 (2014) 086001. arXiv:1406.2104, doi:10.1088/1674-1137/38/8/086001.

[7] [link]. URL http://cern.ch/gaudi/

[8] [link]. URL http://root.cern.ch/drupal/

[9] [link]. URL http://subversion.apache.org/

[10] [link]. URL http://bitten.edgewall.org/

[11] [link]. URL http://trac.edgewall.org/

[12] F. An, et al., Improved Measurement of Electron Antineutrino Disappearance at Daya Bay, Chin.Phys. C37 (2013) 011001. arXiv:1210.6327, doi:10.1088/1674-1137/37/1/011001.