

The FIFE Project at Fermilab: Computing for Experiments

D. Box, J. Boyd, V. Di Benedetto, P. Ding, D. Dykstra, M. Fattoruso, G. Garzoglio, K. Herner*†, T. Levshina, M. Kirby, A. Kreymer, A. Mazzacane, M. Mengel, P. Mhashilkar, V. Podstavkov, K. Retzke, N. Sharma

Fermi National Accelerator Laboratory

E-mail: kherner@fnal.gov

The FabrIc for Frontier Experiments (FIFE) project is an initiative within the Fermilab Scientific Computing Division designed to steer the computing model for non-LHC Fermilab experiments across multiple physics areas. FIFE is a collaborative effort between experimenters and computing professionals to design and develop integrated computing models for experiments of varying size, needs, and infrastructure. The major focus of the FIFE project is the development, deployment, and integration of solutions for high throughput computing, data management, database access and collaboration management within an experiment. To accomplish this goal, FIFE has developed workflows that utilize Open Science Grid compute sites along with dedicated and commercial cloud resources. The FIFE project has made significant progress integrating into experiment computing operations several services including a common job submission service, software and reference data distribution through CVMFS repositories, flexible and robust data transfer clients, and access to opportunistic resources on the Open Science Grid. The progress with current experiments and plans for expansion with additional projects will be discussed. FIFE has taken the leading role in defining the computing model for Fermilab experiments, aided in the design of experiments beyond those hosted at Fermilab, and will continue to define the future direction of high throughput computing for future physics experiments worldwide.

*38th International Conference on High Energy Physics
3-10 August 2016
Chicago, USA*

*Speaker.

†On behalf of the FIFE Project.

1. Introduction

Fermilab¹ is the world's leading laboratory for particle physics research in neutrino physics, and is a key contributor on experiments covering the full range of physics drivers in high-energy physics today. The current future precision muon and neutrino experiments have significant computing resource requirements, comparable to previous-generation collider experiments with collaborations that were several times larger. Many of these experiments are one or two orders of magnitude smaller than LHC experiments, and may lack available effort to design a completely new analysis framework and job submission system. The Fabric for Frontier Experiments (FIFE) Project [1, 2] within the Fermilab Scientific Computing Division aims to meet these requirements and challenges by working closely with experiments and service providers to develop a common, modular toolkit covering the complete range of necessary services. These services include job submission and monitoring, file delivery and cataloging, storage solutions, analysis and reconstruction frameworks, and collaboration services such as databases and document storage. Some examples of these tools include the jobsub infrastructure [3], the SAM file delivery and metadata catalog service [4], the Intensity Frontier Data Handling Client for file transfer [5], and the ART framework for reconstruction and analysis [6]. Figure 1 shows how these components can interoperate in the case of analysis or reconstruction jobs. Having these common tools that are adaptable to such a wide variety of experiments and detectors saves countless hours of otherwise duplicated experiment effort, and makes it easy to work on multiple experiments, as is common in neutrino physics.

The experiments currently using at least some of the FIFE tools cover all physics drivers in high energy physics. Taken together, these experiments are now typically running between 15,000 and 30,000 jobs per day around the world, and transferring approximately 1.8 PB per week in and out of Fermilab's public dCache system, as shown in Figure 2. While these numbers do not equal those of an LHC experiment, they are approaching the same scale, and require the same types of solutions as they continue to increase each year.

2. Addition of Open Science Grid and international computing resources

A major focus within FIFE is to increase the amount of experiment computing done opportunistically outside of Fermilab. Since the Fermilab job submission infrastructure utilizes GlideinWMS [7], it is straightforward to gain opportunistic access to Open Science Grid (OSG) sites. If experiments ensure that their code does not depend on any Fermilab-specific resources, and is available through CVMFS repositories [8], they can increase the resources available to them manyfold. The Mu2e experiment has been tremendously successful in their effort to use OSG computing, consuming over 50 million CPU hours since early 2015 at no cost to the experiment.

Nearly all experiments using the FIFE tools have collaborating institutions outside of the United States. These institutions may often have significant computing resources available to them, but integrating them into existing job submission systems can be extremely challenging. FIFE experts have made significant strides here in the past year by utilizing the GlideinWMS infrastructure,

¹Fermilab is operated by Fermi Research Alliance, LLC under Contract number DE-AC02-07CH11359 with the United States Department of Energy.

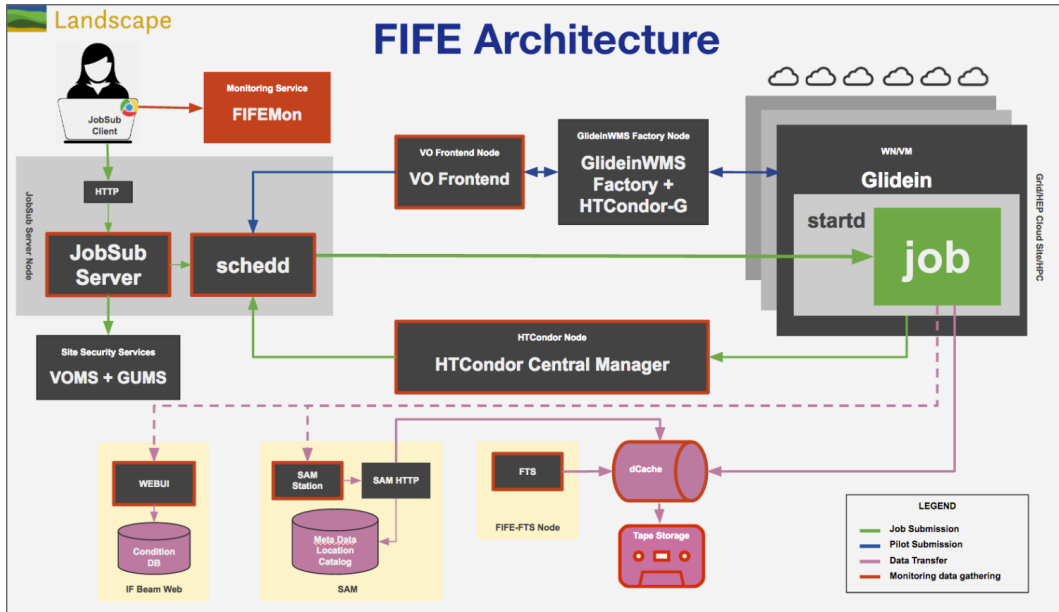


Figure 1: Diagram of the FIFE toolset architecture, include job submission and monitoring tools, data transfer and storage options, and communication with remote sites. The user interface for job submission is in the *jobsub_client* software package. This package communicates with the jobsub HA server and HTCondor schedd, which in turn communicate with a HTCondor central manager and GlideinWMS factory. The GlideinWMS factory takes care of provisioning job slots on computing resources, and each such resource communicates with the HTCondor schedd and central manager to match jobs to slots. Job output typically passes through Fermilab public dCache (shown in the bottom right), which can also act as a high-speed frontend to tape storage. File delivery and metadata cataloging are provided by SAM (with the cataloging via http), and the File Transfer Service (FTS) can automatically tag dCache and tape locations for new files in the SAM metadata catalog. Other required input information, such as beam conditions, may come in to the job via a WebUI.

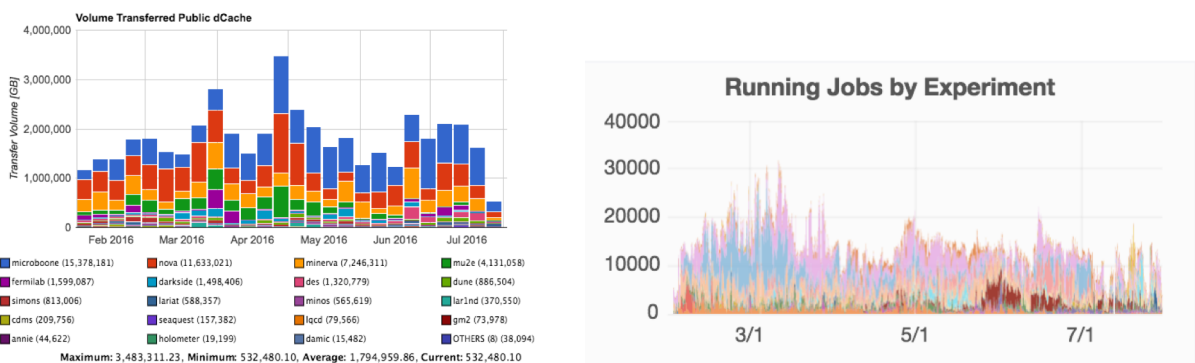


Figure 2: Left: Total amount of data transferred in and out of Fermilab dCache by experiment from February to August 2016. Each bin represents a 7-day period. The totals include raw detector data, simulation and reconstruction activity, and user analysis jobs. Right: The total number of running jobs from FIFE experiments on a daily bases from February to August 2016. Each color represents a different experiment.

POS (ICHEPP2016) 176

and following the OSG prescription for integrating new sites. The ability of GlideinWMS to interface with a variety of different local batch systems reduces the burden on local administrators. There are currently five sites in Europe that can run jobs for one or more FIFE experiments. The total integration time for new sites has steadily decreased and is now down to as little as one week in some cases.

3. Recent advances in the FIFE toolset

3.1 Monitoring tools

The complex nature of the job submission system and grid environment, coupled with experiments' demand for near 100% reliability of the system requires a robust and sophisticated set of monitoring tools. The tools should make it easy for administrators to respond to downtimes and identify inefficiencies, while also providing useful information to end users about their jobs and what they can do to improve throughput. Over the past year FIFE has completely revamped its job monitoring suite, known as FIFEMON [9], and now runs a system based on common, open source tools such as the ELK stack. Information about job status and infrastructure health feeds into FIFEMON, and Grafana is used for visualization. The code base is publicly available and there is an open dialogue between Fermilab and CERN about adapting the tools for CERN purposes. Figure 3 shows an example of some of the monitoring available to each experiment.

Based on experience with FIFEMON, similar technologies are being used in the Open Science Grid's next-generation accounting tool, GRACC [10]. This tool, intended as a long-term Gratia replacement, will provide the same information to service providers and users as Gratia does, but with an improved UI and a lower support burden.

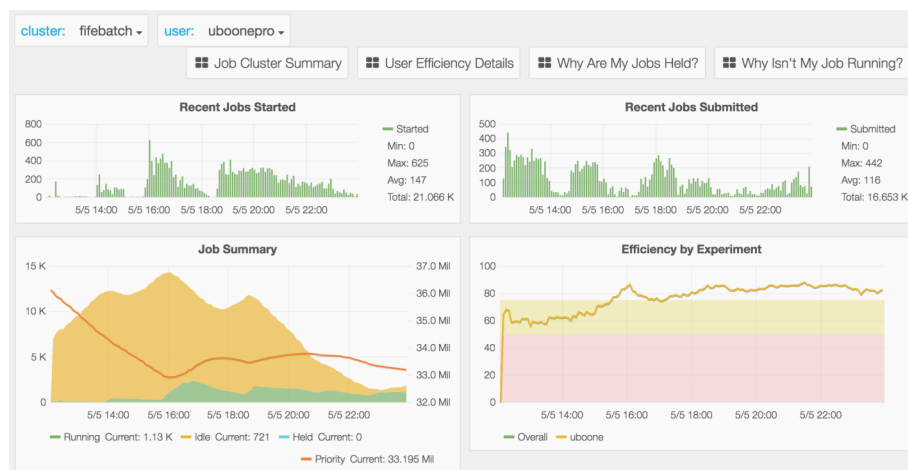


Figure 3: Screen capture of an overview page in FIFEMON. Upper left: jobs recently started. Upper right: jobs recently submitted. Lower left: overview of user jobs vs. time, categorized by job state. Lower right: total efficiency for all running jobs, calculated as CPU time divided by wall time since job start. There are also links in the upper right that go to pages explaining the reasons that any jobs have been held, and compare available job slots to your jobs' requirements to help understand why jobs are not starting.

3.2 Workflow management

The FIFE toolset now includes a full meta-workflow management system that combines job and HTCondor DAG submission, file delivery via SAM, and job monitoring tools. This service, the Production Operations Management Service (POMS), is primarily intended for experiment production teams, but it is open to user analysis as well. The production team or end user defines a high-level "campaign" structure that describes the input dataset, types of jobs to be processed, any dependencies or stages in the workflow structure, and the system will automatically set up and submit the proper job types, including any dependencies as required. POMS tracks the status of every job, and can automatically resubmit recovery jobs or DAGs with no user intervention if it detects a failure. A database stores information about every job's configuration, making it easy to resubmit certain stages of a workflow with different settings. POMS also has a mode where the user can retain full control over job submission, and use POMS only for jobs monitoring and campaign progress tracking. We expect that POMS will greatly improve experiments' productivity as they adopt it into their standard operations.

3.3 Continuous integration

FIFE is also developing a continuous integration suite based on Jenkins. The suite provides automatic building and validation of both common software packages such as ART, and any experiment-specific software that the experiments wish to include. Currently supported platforms include Scientific Linux Fermi 5 and 6, Scientific Linux 7, Ubuntu, and OSX. The common packages are typically built in all of the Scientific Linux flavors, with some packages built in Ubuntu or OSX as requested by the experiments. The experiments are free to choose any or all OS flavors when building their software. A key feature of this suite beyond simply building and validating the software build itself is the ability to run physics validation jobs. These jobs run on the grid and use a standard, well-understood set of input data. The jobs use the new release to process these input files, and generate a set of plots and tuples for comparison to reference values. If the outputs from new release differ from the reference sets, the system can automatically send alerts to the developers, saving the time and effort of by-hand comparisons.

4. Future Directions

The FIFE Project's future focus will cover three main areas: helping to shape the HEP computing model of the future, lowering access barriers to computing resources, and improving our existing services. The HEP computing model of the future will likely include increased use of High Performance Computing (HPC) resources and cloud resources, both in terms of job processing and perhaps storage, and increased use of multithreaded programs. The FIFE job submission infrastructure can now allow experiments to run on allocation-based HPC resources. As the HEP computing model incorporates more cloud-based resources, especially for meeting large, short-term computing power demand, it will be critical to ensure that users can seamlessly access these resources via familiar tools. The FIFE Project will work very closely with the Fermilab HEPCloud project, one of the leaders in the field in this emerging area [11]. While most software currently used by FIFE experiments is single-threaded, memory and speed requirements will make multi-

threading a necessity. FIFE experts will work closely with developers on the various experiments to teach them proper multithreading techniques as needed.

Lowering barriers to computing resources includes making it easier for end users, especially those not based at Fermilab, to easily access all available resources. One initiative underway now is the Distributed Computing Access via Federated Identities (DCAFI) project. The eventual goal of the project is to enable users to access Fermilab resources using their own institutional credentials, if their institution is part of the federated trust realm. This change will reduce the burden on collaborators who are not based at Fermilab; a large number of such people will work on the DUNE collaboration in the coming years.

As we have seen, FIFE offers a wide variety of tools to help experiments meet their computing needs. These tools must be constantly modernized and improved from the end user perspective. Future efforts in these areas will include scaling up our storage offerings, adding capabilities to SAM file delivery to allow for a "send the jobs to the data" model, and continued enhancements to the POMS and CI systems. FIFE's role in these enhancements is to work hand in hand with the experiments to ensure that all of their requirements are being met and that feedback makes it to the service providers.

5. Summary

The FIFE Project is a Fermilab-based initiative to steer the computing model for non-LHC experiments across all areas of high energy physics. FIFE offers a complete, modular set of tools to both experiment production teams and end users, covering everything from job submission to storage to complete workflow management. Taken together the experiments using the FIFE tools are approaching the scale of a single LHC experiment. The FIFE toolset is being constantly expanded and improved, including a completely new job monitoring system. As the HEP computing model evolves in the coming years, FIFE expects to take a major role in shaping it, and is structuring its future efforts accordingly.

References

- [1] M. Kirby, J. Phys. Conf. Ser 513 no. 3, p. 032049. IOP Publishing, 2014.
- [2] D. Box *et al.*, J. Phys. Conf. Ser. 664, no. 6, p. 062040. IOP Publishing, 2015.
- [3] D. Box, J. Phys. Conf. Ser 513, 032010 (2014).
- [4] R. A. Illingworth, J. Phys. Conf. Ser 513, 032045 (2014).
- [5] A.L. Lyon, M.W. Mengel, J. Phys. Conf. Ser. 513, 032068 (2014).
- [6] C. Green *et al.*, J. Phys. Conf. Ser. 396, 022020 (2012).
- [7] I. Sfiligoi *et al.*, 2009 WRI World Congress on Computer Science and Information Engineering (CSIE2009), vol. 02, pp. 428-432, IEEE, 2009.
- [8] J. Blomer *et al.*, J. Phys. Conf. Ser. 331, 042003 (2011).
- [9] <https://fifemon.github.io/>
- [10] <https://opensciencegrid.github.io/gracc/>
- [11] B. Holzman, HEPCloud, Proceedings of this conference, in preparation.