Fermi National Accelerator Laboratory
LDRD Project Data Sheet - FY16

Project ID: FNAL-LDRD-2016-010
Project title: Preparing HEP reconstruction and analysis software for exascale-era computing
Principal investigator: Marc Paterno

Project description: (short description and explanation of cutting edge, high-risk, high-potential science or engineering)

The project is to produce a prototype software system suitable for moving high energy physics (HEP) experiment event data through multiple processing stages in an exascale-class computing facility. There are two critical components to be demonstrated: a) high-performance I/O to a parallel filesystem and b) communication of event data through node interconnects rather than through the filesystem. Simulated and experimental HEP data will be used for the I/O implementation studies, the data store methods for a high-core-count, low-memory per core system, and studying the scaling performance.

Tie to Mission: (explain the project's relevance or anticipated benefits to Fermilab's and DOE's missions)

High energy physics experimentation requires advanced computing capabilities and often breakthroughs in experimentation are the result of advances in computing. Current HEP projects take advantage of "high throughput computing" (HTC) which utilize an architecture not compatible with new "high performance computing" (HPC) machines that will make available factors of 10s and 100s more processing power. This project will perform R&D on the architectures required for HEP data to make use of HPC machines.

Previous year's accomplishments: (as applicable) Investigating I/O with the HDF5 library, contact has been made with the HDF group to implement the storage of complex experimental data. Partly done is developing an efficient means of storing the natural HEP hierarchical structure. Just started is the evaluation of parallel-processing of data from the new format.

Work proposed for current fiscal year and anticipated / desired results:

There are 3 major non-tabular data products to move to HDF5 format before timing studies can begin. There is a version of LArIAT processing to be implemented in python/numpy+MPI. A C++ version of MPI processing will follow. Publication of details will be a project deliverable that also describes performance studies.

Project funding profile: (costs, budgets, projected budgets, and total)

Prior year(s) costs	FY16	FY17	FY18	Total
N/A	189,465	190,974	170,535	550,974

Project Start Data: 1/01/2016 Total Approved Project funds: $ 627,400