Fermi National Accelerator Laboratory
LDRD Project Data Sheet - FY16

Project ID: FNAL-LDRD-2016-032
Project title: Implement open source HEP NoSQL database
Principal investigator: Jin Chang / Oliver Gutsche

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

As an alternative to current file-based data storage now employed by most High Energy Physics (HEP) experiments, NoSQL databases (such as used by Google, Facebook, and others) offer an alternative that has the potential to speed the overall computational time required to handle large data sets.

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

In order to advance HEP experimental work, new capabilities in handling large data sets are required. Thus, this LDRD project has the potential to open up new capabilities in how experiments are designed and optimized to produce new science results. There is also a DOE Office of Science emphasis on developing advanced computing including handling of big data. If successful, this project will help establish Fermilab as a contributing leader in this expanding area.

Previous year's accomplishments: (as applicable) A framework was developed for the NoSQL analysis tool. An innovation to use a striped data representation (i.e. no longer file-based, but database-based) has been developed. The implementation is cloud-friendly and fits within a client/worker/DB architecture. A demo cluster is built and the CMS Dark Matter Search data has been converted and uploaded.

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

The implementation will be moved forward with user-community prototypes to have feedback and incorporate further improvements on the implementation. This will take place before the development is finalized with both a batch ingestion tool and a local mode (laptop-mode) tool.

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

Prior year(s) costs	FY16	FY17	FY18	Total
N/A	46,206	130,272	175,000	351,478

Project Start Data: 1/01/2016 Total Approved Project funds: $ 395,000