

# Integrating Prediction, Provenance, and Optimization into High Energy Workflows

*Wednesday, October 12, 2016 12:30 PM (15 minutes)*

Motivated by the complex workflows within Belle II, we propose an approach for efficient execution of workflows on distributed resources that integrates provenance, performance modeling, and optimization-based scheduling. The key components of this framework include modeling and simulation methods to quantitatively predict workflow component behavior; optimized decision making such as choosing an optimal subset of resources to meet demand, assignment of tasks to resources, and placement of data to minimize data movement; prototypical testbeds for workflow execution on distributed resources; and provenance methods for collecting appropriate performance data.

The Belle II experiments deal with massive amounts of data. Designed to probe the interactions of the fundamental constituents of our universe, the Belle II experiments will generate 25 petabytes of raw data per year. During the course of the experiments, the necessary storage is expected to reach over 350 petabytes. Data is generated by the Belle II detector, Monte Carlo simulations, and user analysis. The detector's experimental data is processed and re-processed through a complex set of operations, which are followed by analysis in a collaborative manner. Users, data, storage and compute resources are geographically distributed across the world creating a complex data intensive workflow.

Belle II workflows necessitate decision making at several levels. We therefore present a hierarchical framework for data driven decision making. Given an estimated demand for compute and storage resources for a period of time, the first (top) level of decision making involves identifying an optimal (sub)set of resources that can meet the predicted demand. We use the analogy of unit commitment problem in electric power grids to solve this problem. Once a cost-efficient set of resources are chosen, the next step is to assign individual tasks from the workflow to specific resources. For Belle II, we consider the situation of Monte Carlo campaigns that involves a set of independent tasks (bag-of-tasks) that need to be assigned on distributed resources.

In order to support accurate and efficient scheduling, predictive performance modeling is employed to rapidly quantify expected task performance across available hardware platforms. The goals of this performance modeling work are to gain insight into the relationship between workload parameters, system characteristics, and performance metrics of interest (e.g., task throughput or scheduling latency); to characterize observed performance; and to guide future and runtime optimizations (including task/module scheduling). Of particular interest, these quantitative and predictive models provide the cost estimates to the higher-level task scheduling algorithms, allowing the scheduler to make informed decisions concerning the optimal resources to utilize for task execution.

## Tertiary Keyword (Optional)

Distributed workload management

## Secondary Keyword (Optional)

Data processing workflows and frameworks/pipelines

## Primary Keyword (Mandatory)

High performance computing

**Primary author:** Dr KERBYSON, Darren J. (Pacific Northwest National Laboratory)

**Co-authors:** Dr STEPHAN, Eric (Pacific Northwest National Laboratory); Dr YIN, Jian (Pacific Northwest National Laboratory); Dr BARKER, Kevin (Pacific Northwest National Laboratory); Dr HALAPPANAVAR, Mahantesh

(Pacific Northwest National Laboratory); SCHRAM, Malachi; Dr TALLENT, Nathan (Pacific Northwest National Laboratory); Dr FRIESE, Ryan (Pacific Northwest National Laboratory)

**Presenter:** SCHRAM, Malachi

**Session Classification:** Track 4: Data Handling

**Track Classification:** Track 4: Data Handling