

# Integrating Prediction, Provenance, & Optimization into Large-Scale Workflows

Pacific Northwest  
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1963

**MALACHI SCHRAM**

ON BEHALF OF THE IPPD PROJECT

ILKAY ALTINTAS, KEVIN J. BARKER, DARREN J. KERBYSON,  
KERSTIN KLEESE VAN DAM, TODD ELSETHAGEN, RYAN FRIESE, MAHANTESH  
HALAPPANAVAR, BIBI RAJU, ARUN SATHANUR, ALOK SINGH, ERIC STEPHAN, NATHAN  
R. TALLENT, JIAN YIN



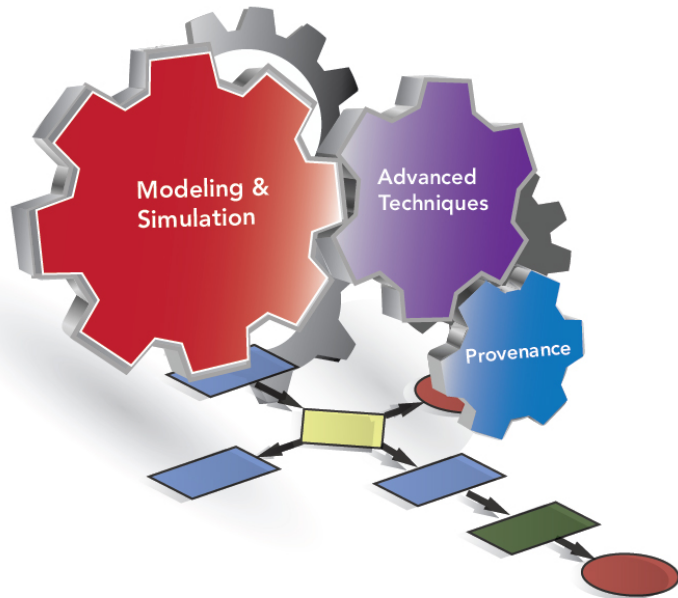
**BROOKHAVEN**  
NATIONAL LABORATORY

**Pacific Northwest National Lab**

**CHEP2016**

San Francisco, Oct 10-14<sup>th</sup>, 2016

# IPPD: Integrated Performance Prediction & Diagnosis



- ▶ Aim to provide an integrated approach to the modeling of extreme scale scientific workflows
- ▶ Brings together researchers working on modeling / simulation / empirical analysis, workflows and domain scientists
- ▶ Builds upon existing research much of which has focused to date on large-scale HPC systems and applications

**BROOKHAVEN**  
NATIONAL LABORATORY



*Explore in advance* – Design-space exploration & Sensitivity Analyses

*Optimize at run-time* – Guide execution based on dynamic behavior

# Key research areas in IPPD



## ▶ Provenance

- Capture empirical performance information from workflows enabling baseline performance to be established; to identify and help diagnose variability; to feed simulation and modeling;

## ▶ Modeling and Simulation

- Develop modeling and simulation to enable both exploration in advance of possible workflow configuration and optimizations, as well as rapid performance prediction to guide the dynamic adaptation of workflows and optimization of resource utilization.

## ▶ Advanced Techniques

- Explore novel optimization techniques for workflow optimization for both processing and data organization. Leverage unit-commitment ideas (from the smart power-grid), and transparent page management and compression (for data-movement / storage reduction)

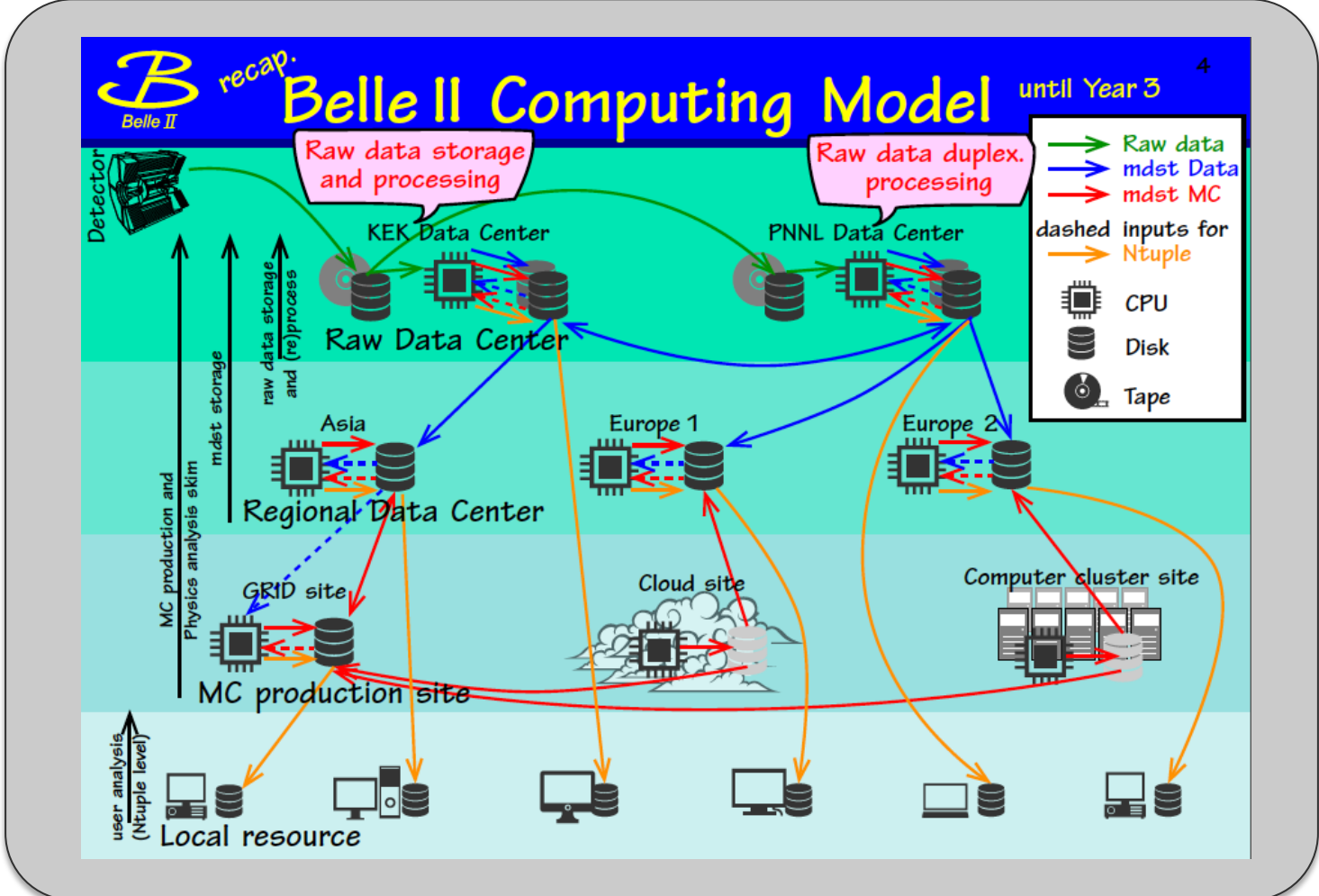
## ▶ Workflows – Belle II, ACME

- Initially explore our developments for analyzing, optimizing and considering future design of the Belle-II High Energy Physics project
- Also extending to others including metagenomics, climate.

# High Energy Physics: Belle II Analysis Workflow



- ▶ Belle II has adopted a geographically distributed computing model
- ▶ Several compute workflows



# IPPD's Capability Demonstrator: 'Enhanced' Belle II Workflow Execution

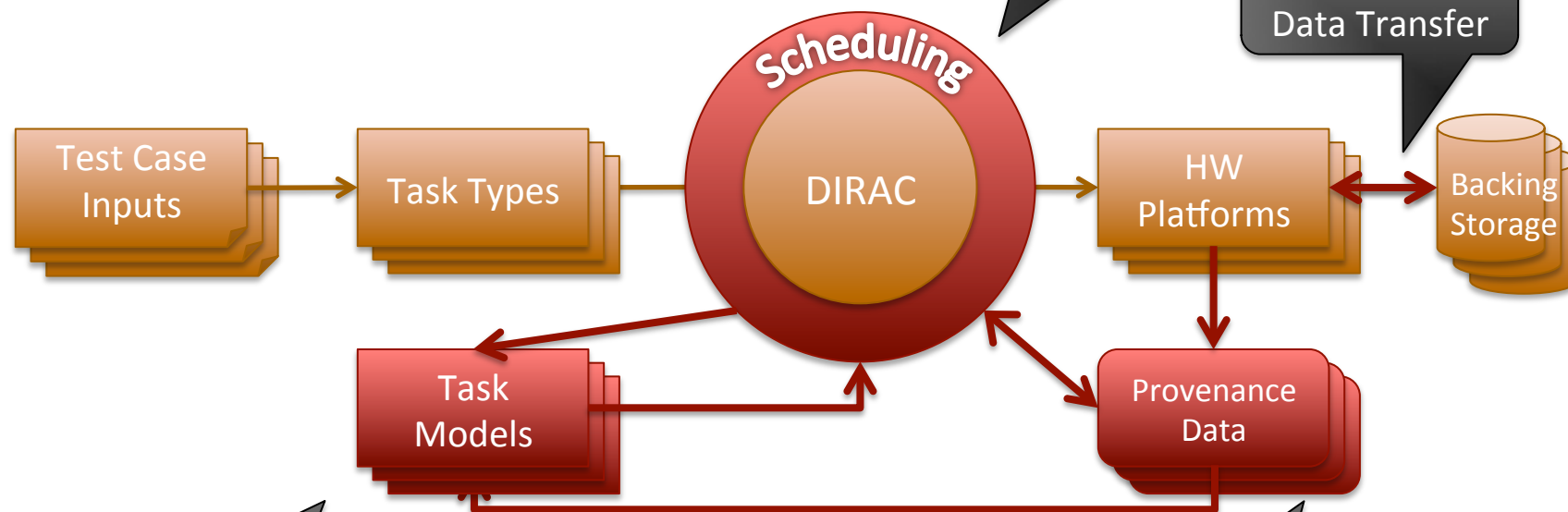


## Challenge

- Mitigate contention
- Consider power/energy

1. Scheduler avoids I/O & network contention

3. Optimized Data Transfer



2. Accurate task predictions enable good schedules

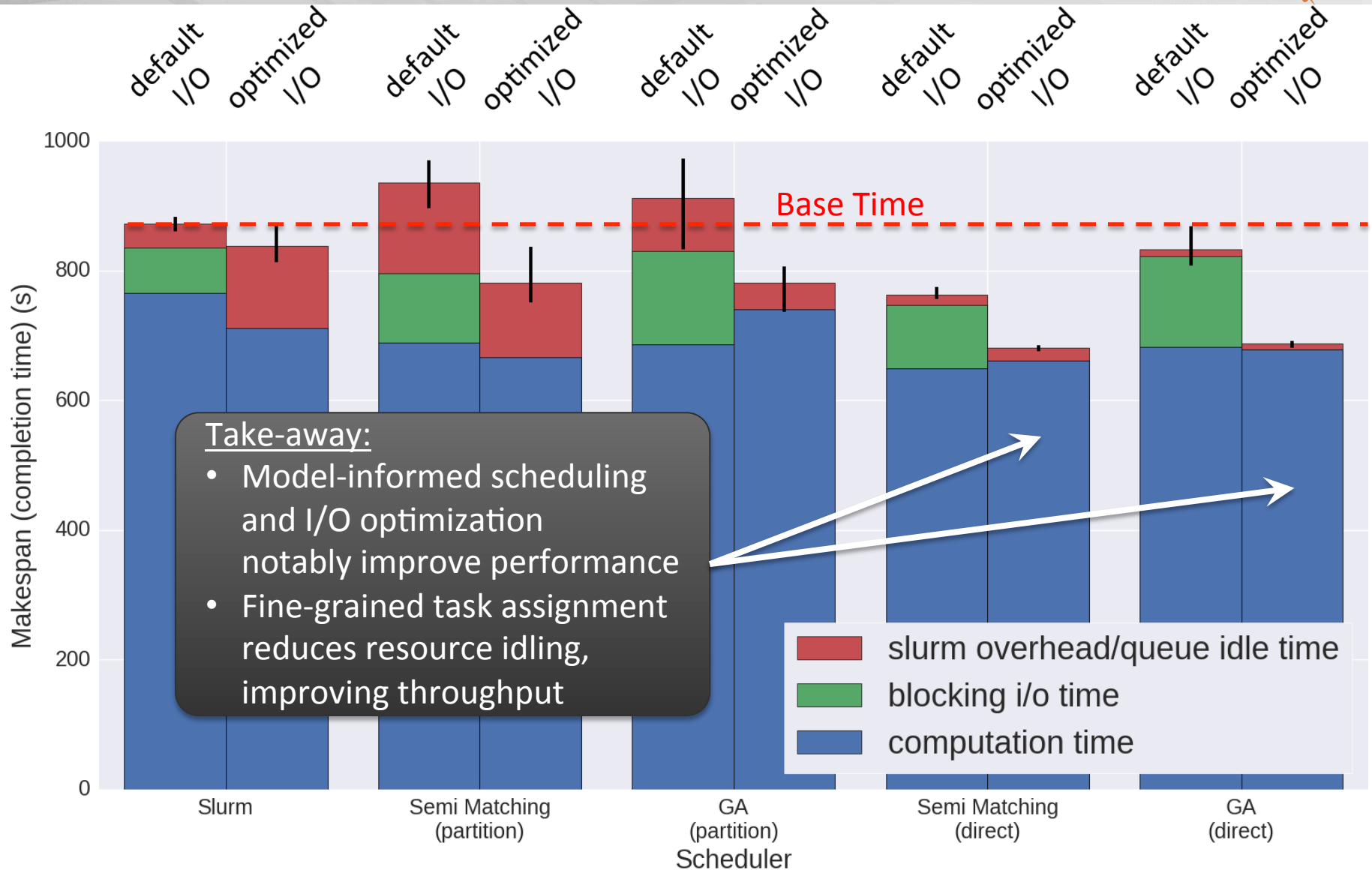
4. Provenance feeds back performance data & Model Validation

# IPPD's 'Enhanced' Belle II Workflow Execution



- ▶ Integration of: Prediction, Intelligent Scheduling, I/O optimization, and Provenance data capture
- ▶ On-line integration into extreme-scale wide-area Belle-II processing
  - 3 test-cases: ***Small – small cluster, limited task types, multiple node***
    - Medium – multi-cluster, additional task types
    - Large – predictive, represents possible future
- ▶ ***Small case*** in progress, techniques now integrated, examining impact
  - System: SEAPEARL local PNNL cluster highly instrumented
    - 4 node types (Intel Ivybridge, Haswell, AMD Fuzion)
  - Workload (based on sample Belle II tasks with imbalance)
    - Itop – 100's tasks, each with 1-100 events (work)
    - Monte Carlo – 100's tasks, each with 1-100 events (work)
  - Initial experiments being used to explore impact of IPPD techniques
    - SLURM with Scheduling (semi-matching, GA) at partition level, node level
    - I/O optimizations (with & without pre-feeding from 2 remote servers)
  - Collecting Provenance during all experiments

# Capability Demonstrator: Results



# Summary



- ▶ Integration of expertise: compute performance modelers, system software, data scientists, and domain scientists
- ▶ Integration of compute modeling with optimization and provenance for small case Belle II workflows
- ▶ Currently working on medium case distributed workflows and network/storage contention
- ▶ Looking at further workflows including ACME (climate)
- ▶ Generality of techniques will lead to increased impact
  - Looking towards using IPPD's approach with other WMS