LOIS ID: 11184

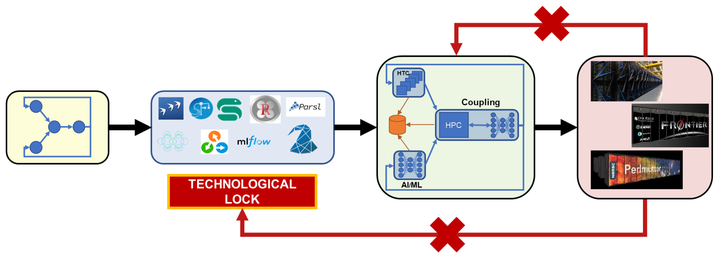
OAK RIDGE National Laboratory

# Laboratory Directed Research and Development

## Accelerating the Design of Scientific Workflows with Simulation-Based Rapid Prototyping

**Program: Strategic Hire**
**PI: Fred Suter (suterf@ornl.gov)**

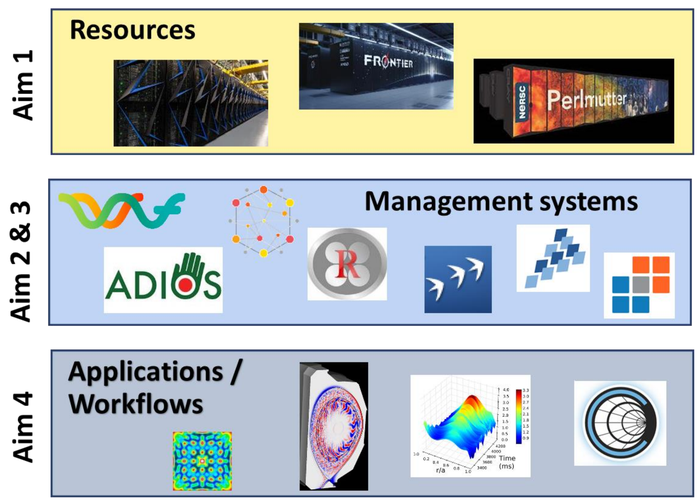
### Traditional Workflow Design




1. **Design abstract** workflow to solve a scientific problem
   - Identify components and dependencies
2. **Select ONE** workflow management system **among MANY**(1)
   - Empiric choice (recommendation, brief study, careful investigation, …)
   - **Technological lock**
3. **Implement concrete** workflow
4. **Deploy and test** on target infrastructure(s)
   - Time- and resource-consuming process
5. ~~Change the workflow management~~

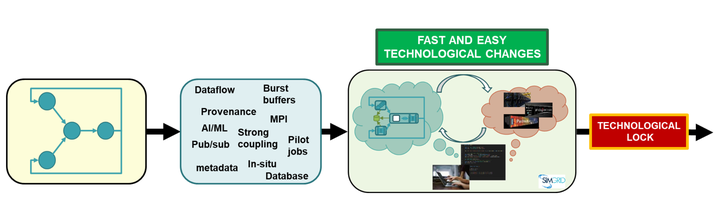
(1) https://s.apache.org/existing-workflow-systems

### Project Overview




**Develop a comprehensive simulation-based framework**
**to help domain scientists prototyping their workflows**
**and provide them with the most efficient design and implementation**

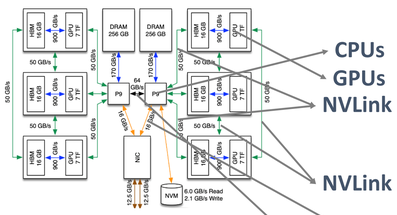
### Simulation-based Rapid Prototyping

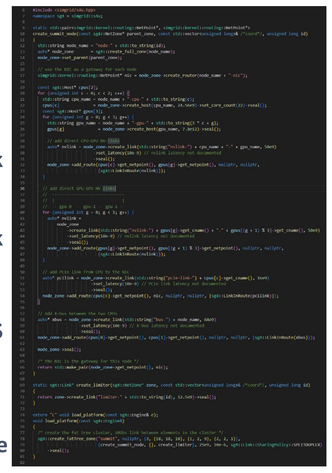



1. **Design abstract** workflow to solve a scientific problem
   - Identify components and dependencies
2. **Identify patterns** and adapted candidate strategies
   - User engagement
   - Pencil and papers prototyping
3. **Implement** workflow **simulator**
   - **Test** on simulated infrastructure(s)
   - Fast **iterative refinement** process
   - Converge on **concrete** workflow
4. Select the **most adapted** workflow management system(s)
5. **Deploy** on target infrastructure(s)

### Modeling Leadership Class Supercomputers and Beyond

**Summit in less than 80 lines**



Leverage SimGrid's **programmatic resource description** API

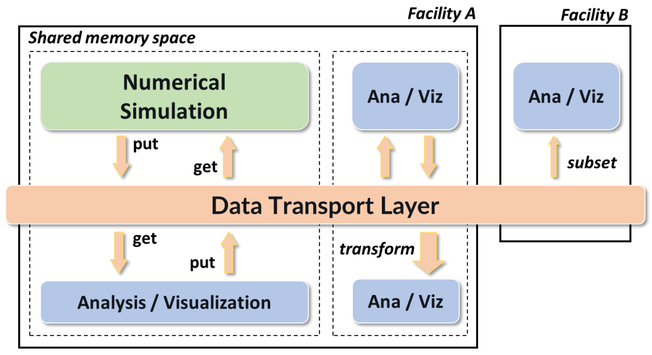
**Calibrate** from logs and benchmark results

**Full Fat Tree**

Include into **Edge-to-HPC** computing continuum scenarios

SIMGRID
https://simgrid.org

### Versatile Data Transport Layer




- Expose a simple **Publish/ Subscribe** abstraction to applications and services
- **Decouple** actual data movement and storage

- Enable **testing and evaluation** of several options
  - File-based
  - In-situ data streaming
  - In-memory
  - No-cost

### Exposing Workflow Patterns and Management System Features

**Workflow Patterns**

1. Strong Code Coupling and Analytics
2. Ensemble Contributing to a Common Data Set
3. AI/ML-based Steering
4. Edge-to-HPC Multi-Stage Analysis
5. Digital twins

F. Suter, R. Ferreira da Silva, A. Gainaru, S. Klasky.
**Driving Next-Generation Workflows from the Data Plane**
*19th IEEE International eScience Conference*, 2023

**Workflow Management Systems Features**

Focus on tools supported by the **ECP ExaWorks** project

- **Radical-Pilot** ➞ Ensembles and Pilot jobs
- **Parsl** ➞ Dataflow-based workflows
- **Swift/T** ➞ Many-task workflows
- **EFFIS** ➞ Strong and loose coupling

ExaWorks ECP

Research sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U.S. Department of Energy.

U.S. DEPARTMENT OF ENERGY