OAK RIDGE National Laboratory

ATLAS TIM – 01/23/25 – Stony Brook University

Simulating HEP Workloads with SimGrid

Fred Suter

Oak Ridge National Laboratory

U.S. DEPARTMENT OF ENERGY

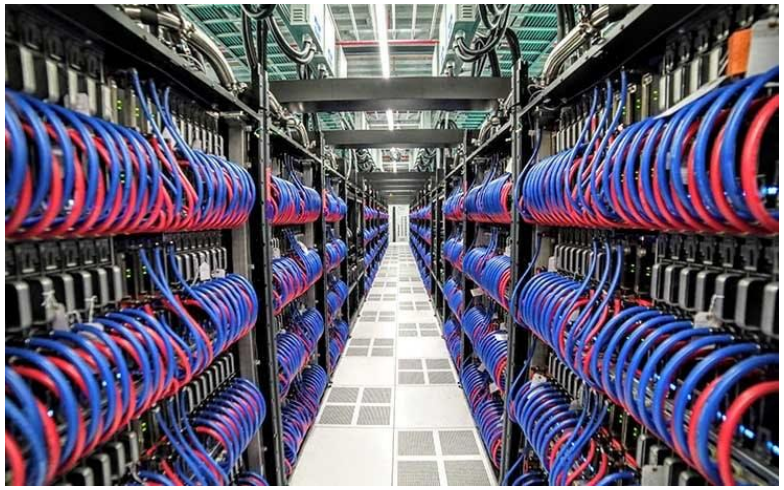
ORNL IS MANAGED BY UT-BATTELLE LLC
FOR THE US DEPARTMENT OF ENERGY

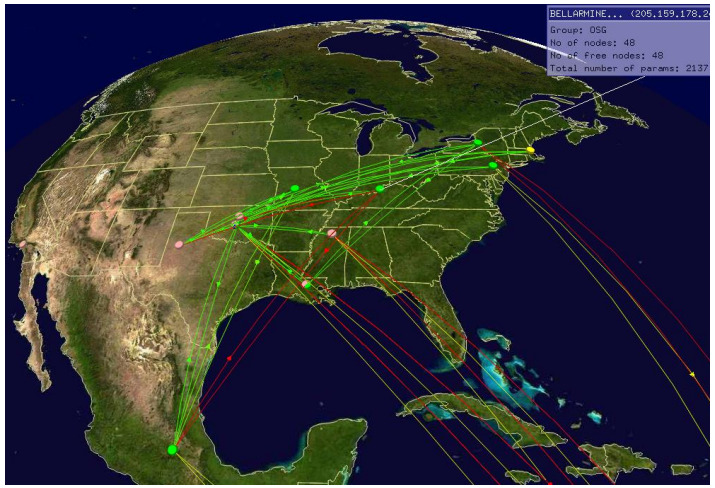
OAK RIDGE National Laboratory

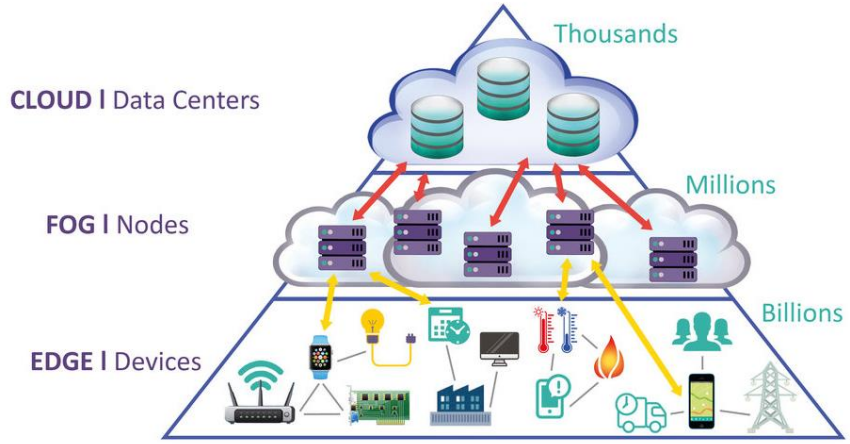
A Brief History of SimGrid

Our Scientific Objects: Distributed Systems

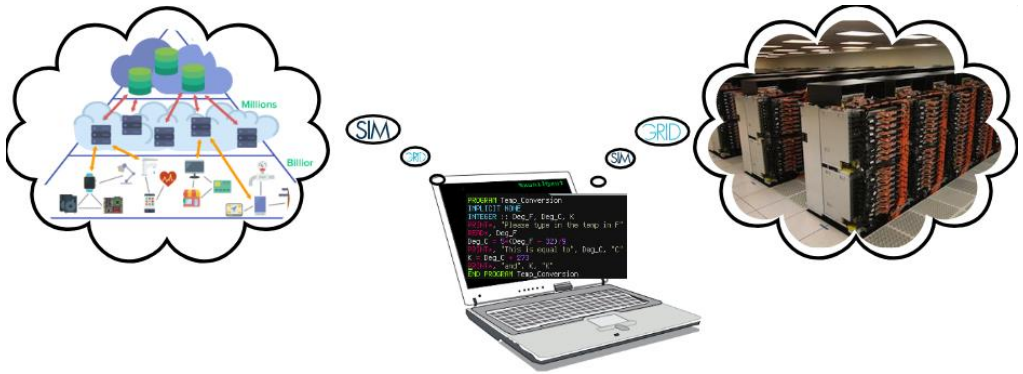
- **Clusters, supercomputers, peer-to-peer systems, grids, clouds, . . .**





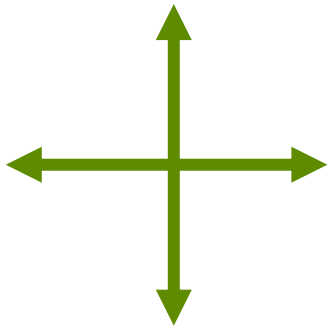


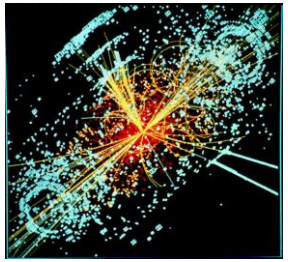
- **How to study these systems and their applications on my laptop?**



A Physics Approach of Computer Science

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^1 + i\phi^2 \\ \phi^0 + i\phi^3 \end{pmatrix}$$





Theory: Informs on what should happen

Observations: **Real** applications on **real** systems

Emulation: **Real** applications on system **models**

Simulation: Application **models** on system **models**

Simulation: **Fastest way from idea to data**

Challenges

Accuracy	Versatility
	Utilisability
Scalability	Extensivity

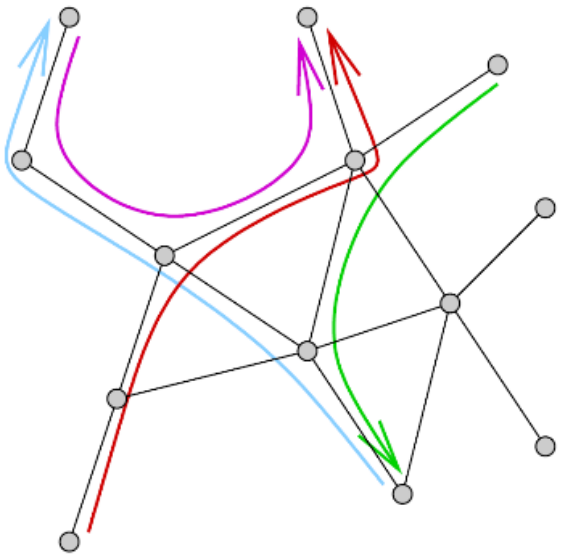
SimGrid in a Nutshell

- **Discrete Event Simulator** (sequential, but fast)
- **Base Abstractions**

Actors	Activities	Resources
Program anything you want/need	Computation, communication, I/O	CPUs, Links, Disks, …

Mailboxes / MessageQueues Rendez-vous points between actors

- **Simulation kernel main loop**
  1. Compute **share** of resource allocated to every activity (resource sharing algorithms)
  2. Compute the **earliest finishing** activity, **advance** simulated time
  3. Remove finished activity
  4. Loop back to 2
- **Flow-level models**
  - Boils down to solve a **linear max min** problem
  - Good tradeoff between **speed and accuracy**
  - Mutiple optimization techniques and specializations



SimGrid History

1998 – 2001: Factor student code (DAG scheduling)

Casanova, H. **Simgrid: a toolkit for the simulation of application scheduling**

2001– 2005: CSP and improved network models

Legrand, A., Marchal, L., Casanova H.
Scheduling distributed applications: the simgrid simulation framework

SG1 SG2 SG3 SG4

2005– 2014: Versatility, Accuracy, Scalability

Casanova, H., Legrand, A., Quinson, M.
Simgrid: A generic framework for large-scale distributed experiments
Casanova, H., Giersch, A., Legrand, A., Quinson M., Suter, F.
Versatile, scalable, and accurate simulation of distributed applications and platforms

2014– 2025: Utilisability and Extensibility

Casanova H., Giersch A. Legrand, A., Quinson M, Suter, F.
Lowering Entry Barriers to Developing Custom Simulators ofDistributed Applications and Platforms with SimGrid

OAK RIDGE National Laboratory

cnrs Inria OAK RIDGE National Laboratory

OAK RIDGE National Laboratory

SimGrid and HEP Workloads

MartinWillSim (ca. 2009)

MARTINWILLSIM

GRID Simulator

Martin Barisits Will Boyd

Supervised by Mario Lassnig and Vincent Garonne

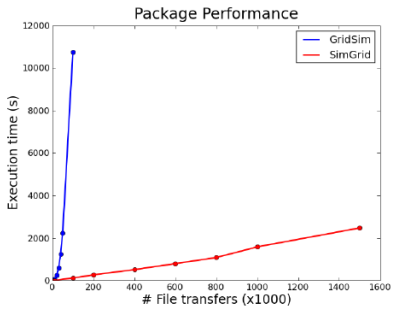
August 13, 2009

The Problem

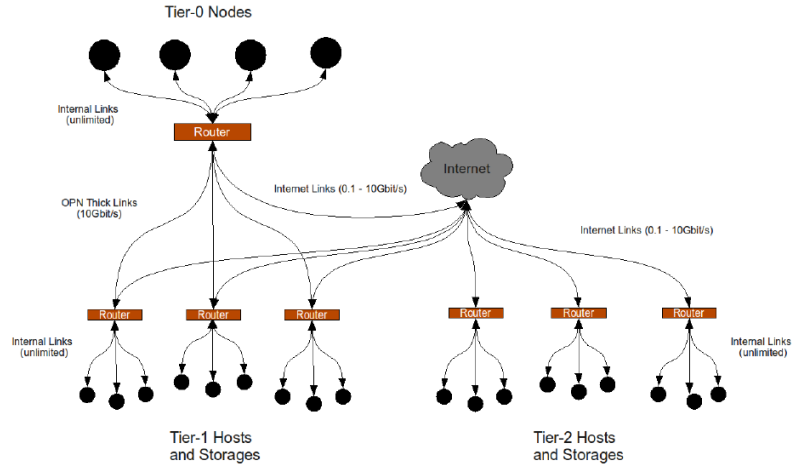
- Goal: Test data distribution strategies
- Need: Simulator
- Need: Ability to load the Simulator with the current GRID Topology
- Need: Inject the Simulator with realistic workloads
- Process Results

Package Performance

- Attempted to simulate one day on GRID (1.5 million file transfers)
- GridSim: exponential in CPU time with increasing transfers
- SimGrid: linear in CPU Time with increasing transfers



MARTINWILLSIM GRID Topology



The topology that is generated for simulation

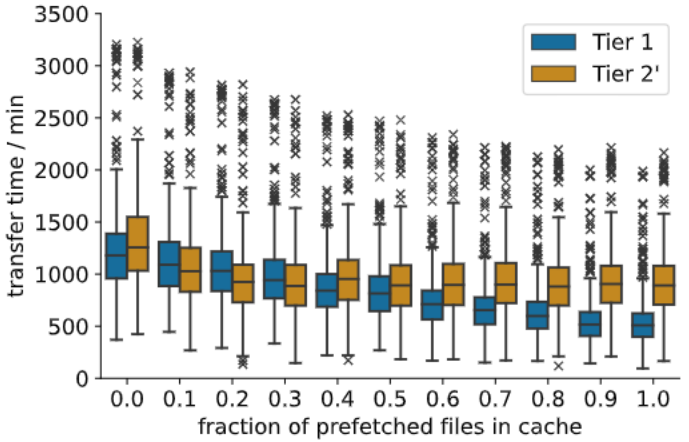
OAK RIDGE National Laboratory

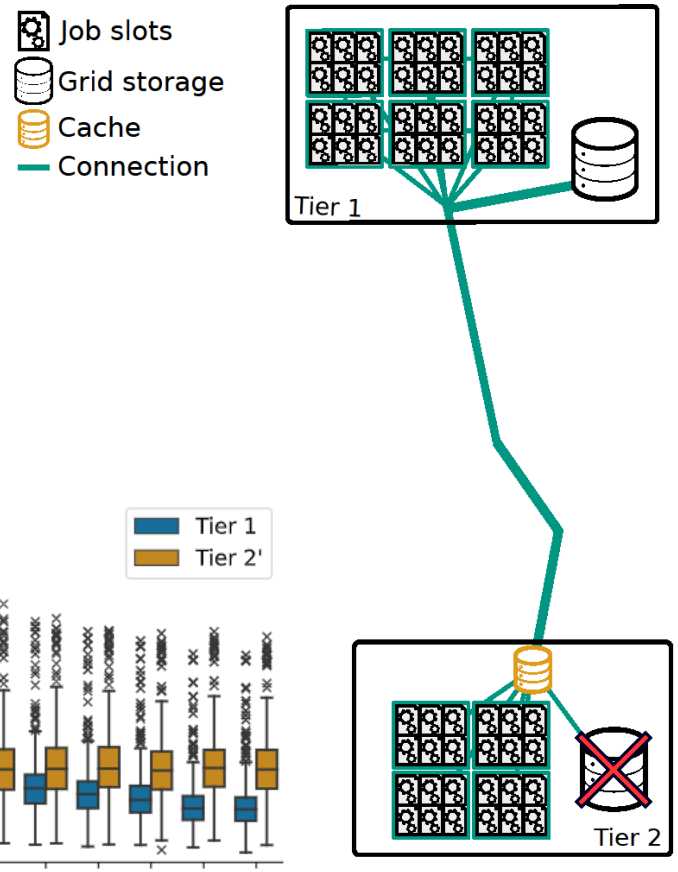
Planning a Computing and Storage Infrastructure for HEP

- Study **hypothetical** interplay of
  - A **Tier 1** center
  - A **Tier 2 with grid storage replaced by cache** (Tier 2')

Tier 1	Tier 2'
$\mathcal{O}$(40k) cores	$\mathcal{O}$(20k) cores
80 Gbps bandwidth storage	80 Gbps bandwidth cache
2 ⊗ 100 Gbps local network	40 Gbps local network
100 Gbps network between sites	

- **Prefetch** a fraction of the files from cache
  - **ρ = 0:** Every file is transferred from Tier 1
  - **ρ = 1:** Every file is fetched from the cache on Tier 2'

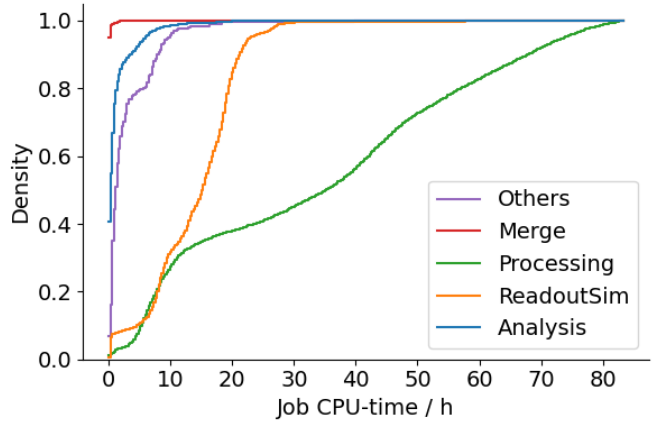


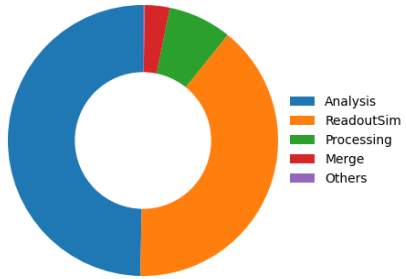


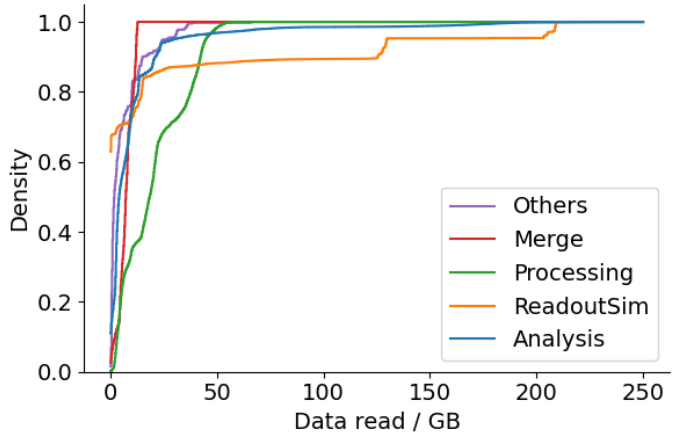
Horzela, M., Casanova, H., Giffels, M., Gottmann, A. Hofsaess, R., Quast, G., Rossi Tisbeni, S., Streit, A., Suter, F.
Modeling Distributed Computing Infrastructures for HEP Applications, EPJ Web of Conf., 295 (2024) 04032

DCSim: Implementation of (HEP) Extensions and Simulator

- **Define workloads of jobs**
  - Number of operations to execute
  - Memory
  - Size of input and output files
  - Submission time
- **Define platform (network & hosts)**
  - **properties:** #cores, CPU-speed, RAM, disk, bandwidth
  - **role:** worker, storage, data cache, scheduler, …
- **Instantiate initial deployment of files on storage systems**
- **Start the simulation!**
  - Jobs are scheduled and run
  - Input-files are streamed and cached
  - Caches evict files if necessary
  - Job dynamics are monitored

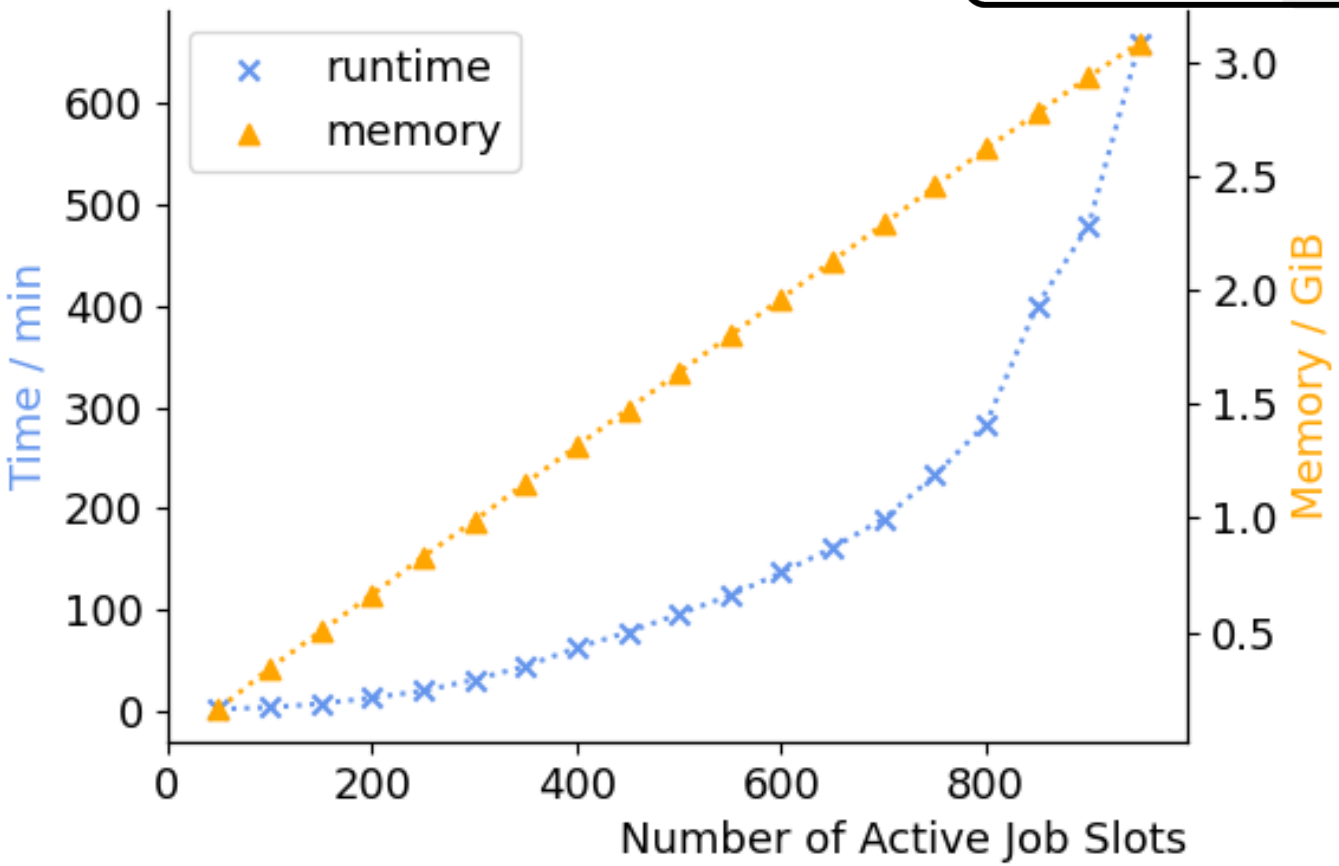






https://github.com/HEPCompSim/DCSim

DCSim: A Scalability Headache!



CyberInfrastructure Simulation

- A layer above SimGrid to make writing simulators of complex distributed computing systems easy
- Comes with built-in simulation abstractions (batch-scheduled cluster, cloud platform, storage systems, etc.)

WRENCH

- Led to new developments in SimGrid
  - Message queues
  - I/O streams
  - Compound activities
  - File System module
- And many optimizations in WRENCH
  - Memory and Speed
  - Automated calibration

OAK RIDGE National Laboratory

SimGrid and REDWOOD

History repeats itself …

The Problem

- Goal: Test data distribution strategies
- Need: Simulator
- Need: Ability to load the Simulator with the current GRID Topology
- Need: Inject the Simulator with realistic workloads
- Process Results


Questions?
Thank you for your attention

U.S. DEPARTMENT OF ENERGY

ORNL IS MANAGED BY UT-BATTELLE LLC
FOR THE US DEPARTMENT OF ENERGY