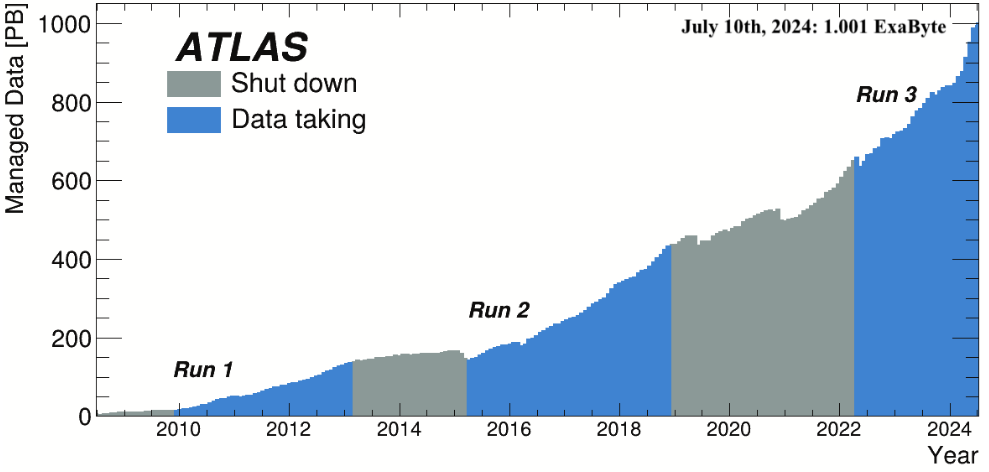
REDWOOD : Resilient Federated Workflows in a Heterogeneous Computing Environment

Lead PI: Alexei Klimentov (BNL)

Co-PIs: BNL: Adolfy Hoisie, Tadashi Maeno, Shinjae Yoo
ORNL: Scott Klasky
SLAC: Wei Yang
UMass Amherst: Verena Ingrid Martinez Outschoorn
Carnegie Mellon University: Yiming Yang
University of Pittsburgh: Joseph Francis Boudreau

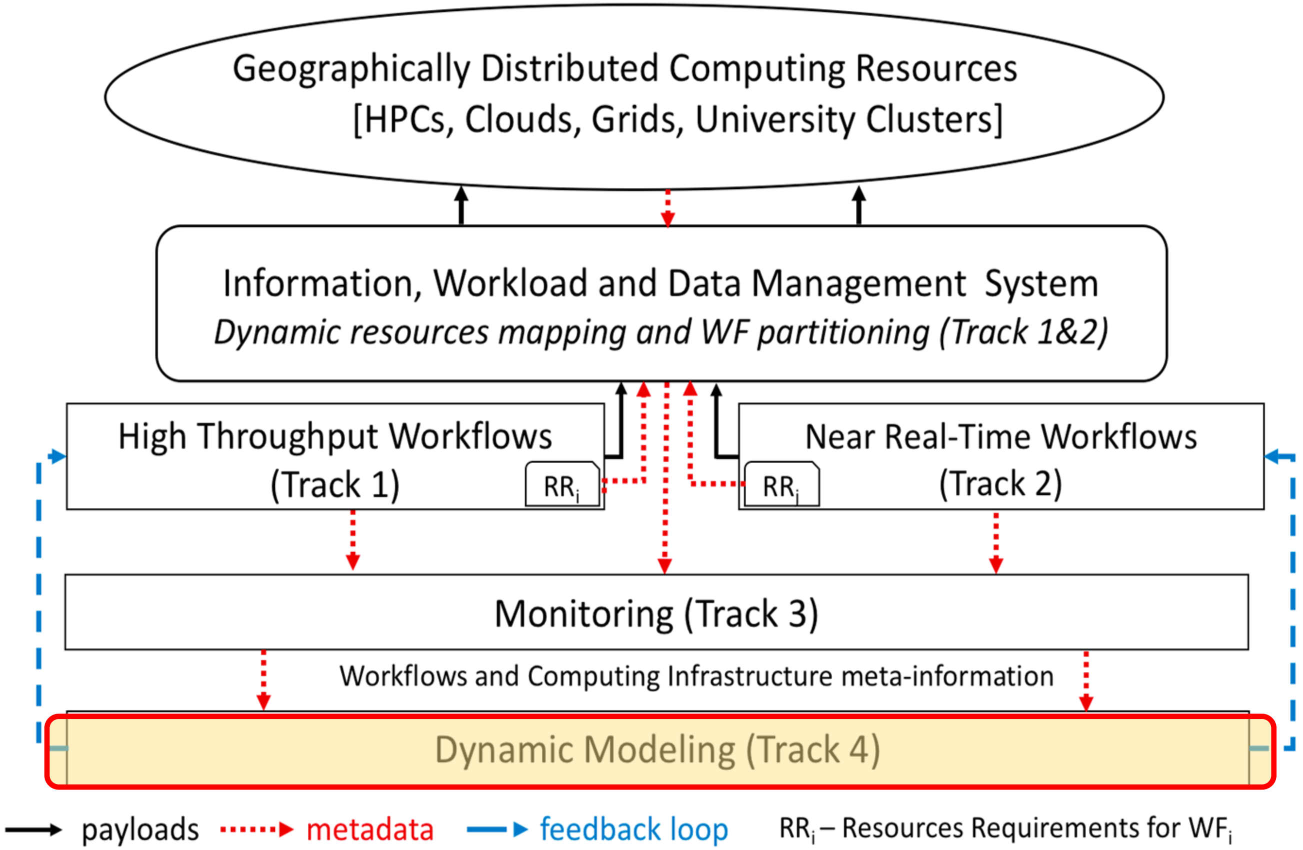
ASCR PM monthly update, Sep 24, 2024 | Presenter: David Park (BNL) on behalf of REDWOOD collaboration

Project Goals and Organization



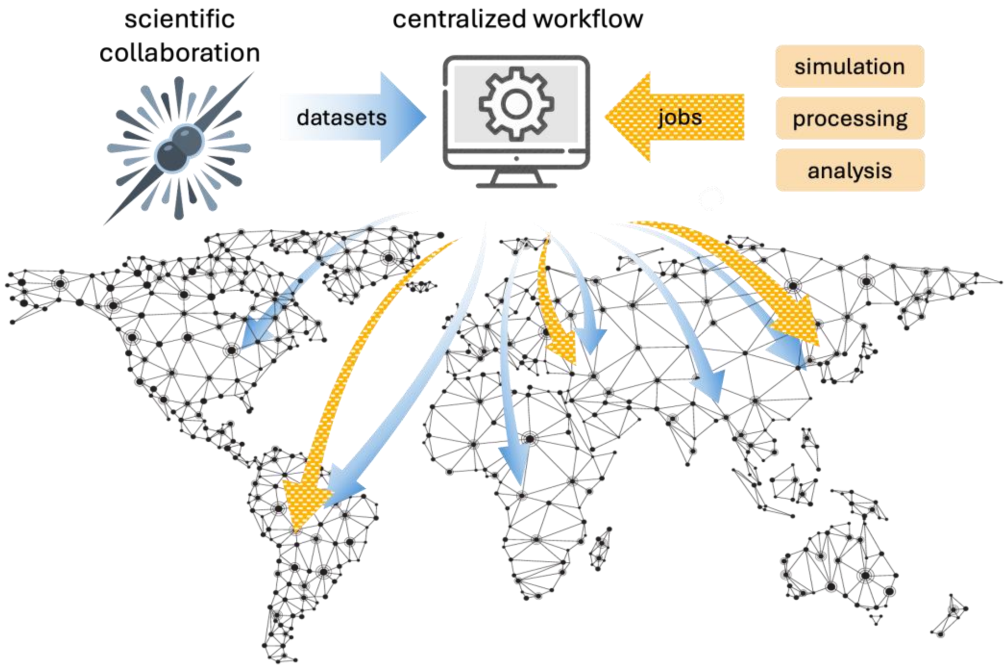
- Motivation: Extreme large data volumes and increasingly complex computation workflows in many scientific domains

- Goal: Optimal data placement and workload scheduling enhancing the resilience, throughput, and resource utilization.

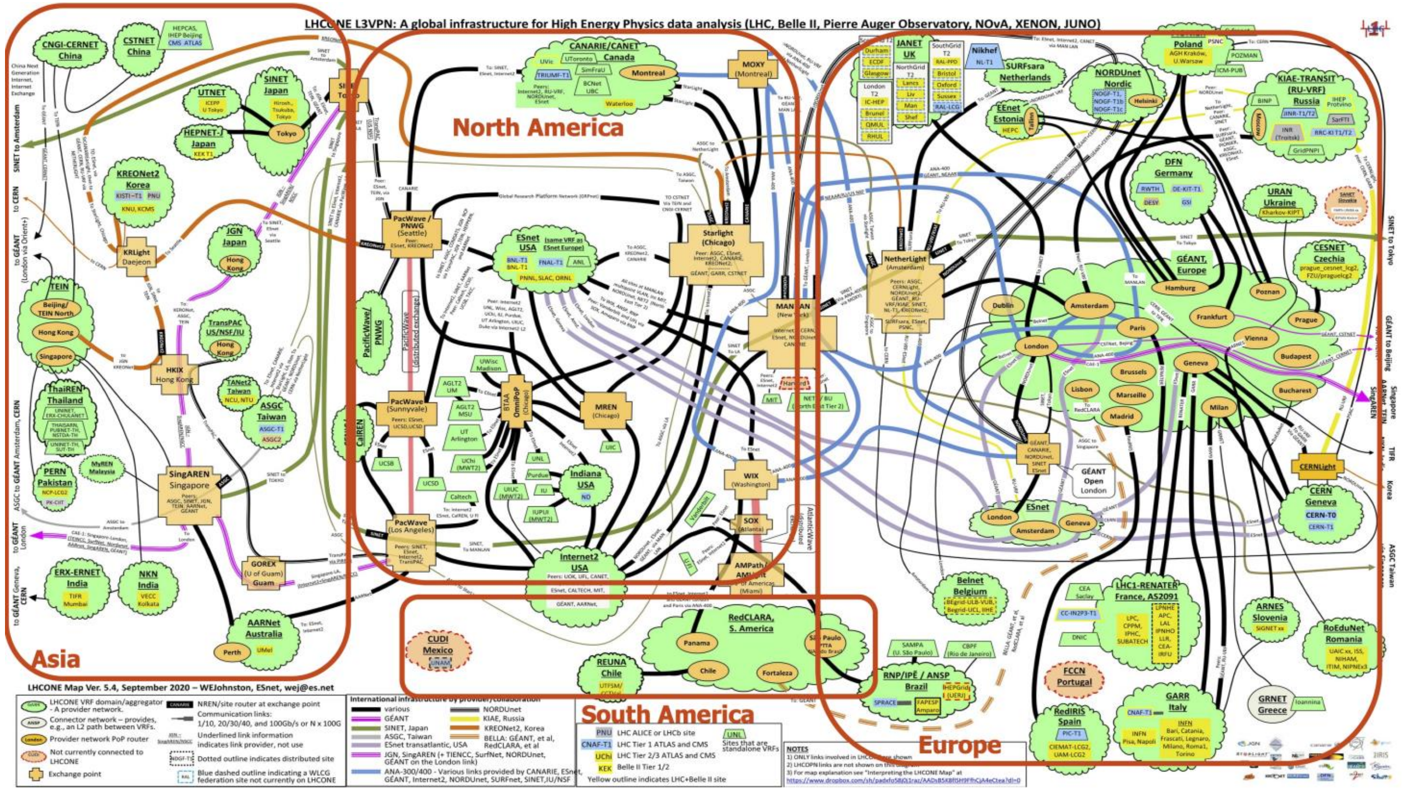


Current Landscape of Distributed Computing

- Jobs and datasets are distributed and computed across 150 sites in 40 countries on all continents but Antarctica

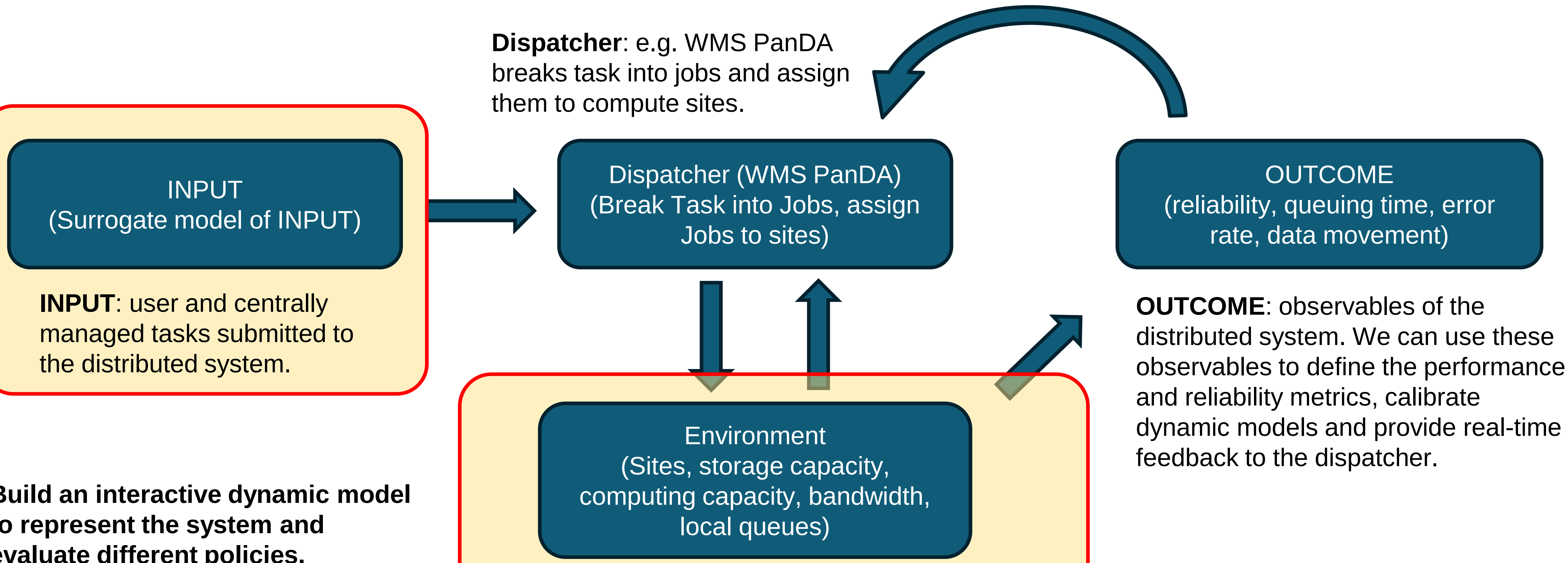


Centralized workflow manages jobs and datasets



World Nuclear and Particle Physics Research Network

Four Interacting Components of the Dynamic Model



Dispatcher: e.g. WMS PanDA breaks task into jobs and assign them to compute sites.

INPUT: user and centrally managed tasks submitted to the distributed system.

OUTCOME: observables of the distributed system. We can use these observables to define the performance and reliability metrics, calibrate dynamic models and provide real-time feedback to the dispatcher.

Environment: different sites have different computing capabilities, storage, data partition, network bandwidth, local jobs, and failure rate, etc.

Build an interactive dynamic model to represent the system and evaluate different policies.

Brookhaven National Laboratory

Four Interacting Components of the Dynamic Model

Real-time Feedback

Dispatcher: e.g. WMS PanDA breaks task into jobs and assign them to compute sites.

INPUT
(Surrogate model of INPUT)

INPUT: user and centrally managed tasks submitted to the distributed system.

Dispatcher (WMS PanDA)
(Break Task into Jobs, assign Jobs to sites)

OUTCOME
(reliability, queuing time, error rate, data movement)

OUTCOME: observables of the distributed system. We can use these observables to define the performance and reliability metrics, calibrate dynamic models and provide real-time feedback to the dispatcher.

Environment
(Sites, storage capacity, computing capacity, bandwidth, local queues)

Environment: different sites have different computing capabilities, storage, data partition, network bandwidth, local jobs, and failure rate, etc.

Build an interactive dynamic model to represent the system and evaluate different policies.

Brookhaven National Laboratory

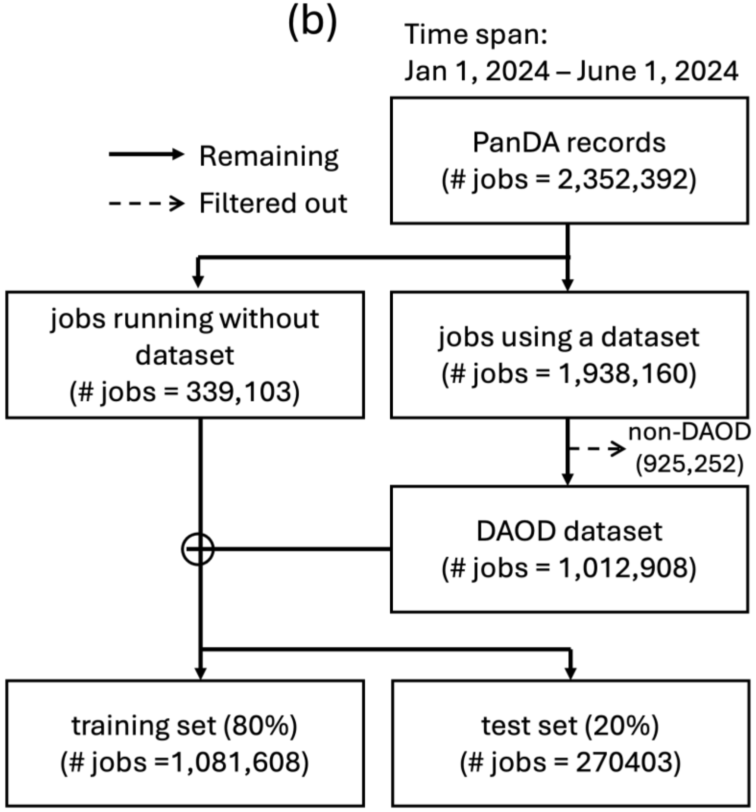
Workload Management System (WMS PanDA) Records

- Preprocessing pipeline (b) and preprocessed data samples (a)

(a)

	creation time	computing site	DAOD dataset features					status	workload
			project	prod step	data type	nfiles	size		
type	N	C	C	C	C	N	N	C	N
# unique	N/A	83	14	4	54	N/A	N/A	4	N/A
samples	2024-03-24 21:09:26	ANALY_BNL_VP	data16_ 13TeV	deriv	PHYS	10.0	1.86e+10	finished	620760.0
	2024-02-18 23:37:50	SWT2_CPB	mc21_1 3p6TeV	deriv	PHYS	3.0	1.66e+10	finished	303960.0
	2024-04-22 08:57:48	CERN	mc21_1 3p6TeV	deriv	PHYS	1.0	3.49e+09	failed	3300.0
	2024-03-24 17:48:13	BNL	mc20_1 3TeV	deriv	EGAM1	8.0	5.22e+10	finished	7010880.0
	2024-01-07 09:39:54	ANALY_ARNES_ DIRECT	data18_ 13TeV	deriv	PHYS	1.0	2.59e+09	finished	45000.0

(b)



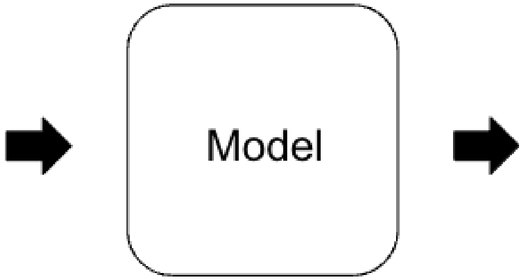
D. Park et al., "AI Surrogate Model for Distributed Computing Workloads", to appear at sc24 AI4Science workshop

Generative Models for Tabular Data

Number of data – Train: 1,343,792 (60%) / validation: 447,931 (20%) / test: 447,931 (20%)

creationdate	computingsite	workload	jobstatus
2024-03-11 08:43:26	TRIUMF	244150.0	finished
2024-02-12 06:51:24	AGLT2	0.0	closed
2024-02-11 11:42:23	BNL	351720.0	finished
2024-03-17 22:52:56	TOKYO	5460.0	failed
2024-01-21 18:17:05	ANALY_ARNES_DIRECT	1173400.0	finished
2024-05-05 20:15:07	SWT2_CPB	263880.0	finished
2024-02-05 08:44:23	praguelcg2	122220.0	finished
2024-05-27 08:21:09	FZK-LCG2	185640.0	failed
2024-03-24 15:59:45	UKI-NORTHGRID-MAN-HEP	436920.0	finished
2024-04-29 03:11:47	INFN-LECCE	182300.0	finished

Samples of training data



creationdate	computingsite	workload	jobstatus
1.710744e+09	IN2P3-LAPP	4.775945e+04	finished
1.710744e+09	TRIUMF	1.661405e+04	finished
1.711332e+09	CERN	2.614423e+03	finished
1.714942e+09	SWT2_CPB	6.659398e+03	finished
1.713719e+09	TRIUMF	1.020332e+05	finished
...	...	...	...
1.713725e+09	NSC	8.748761e+05	finished
1.714943e+09	SWT2_CPB	3.329313e+06	finished
1.708938e+09	SWT2_CPB	1.212568e+03	finished
1.708937e+09	CERN-T0	0.000000e+00	closed
1.714940e+09	BNL	4.665673e+03	failed

synthetic data

Baselines: tabular generative models

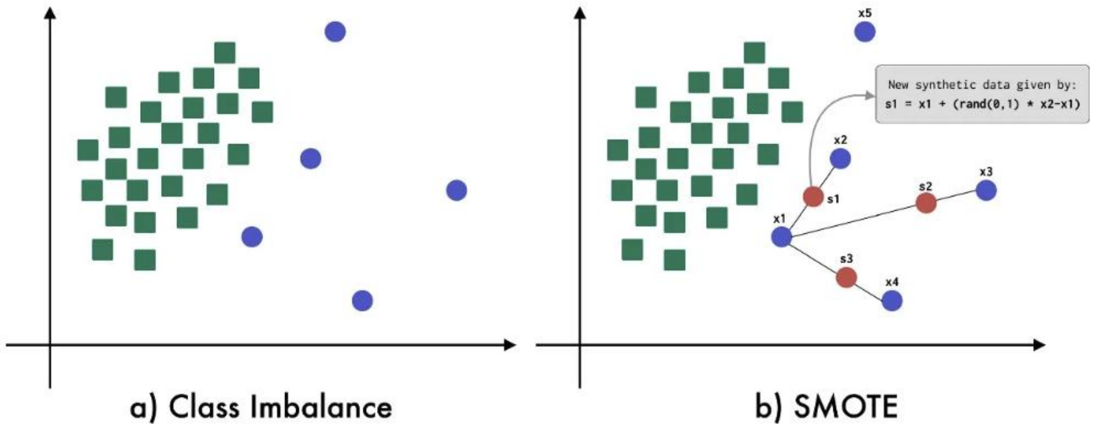
SMOTE: Non-DL algorithm working based on nearest neighbor.

TVAE: Variational autoencoder as backbone

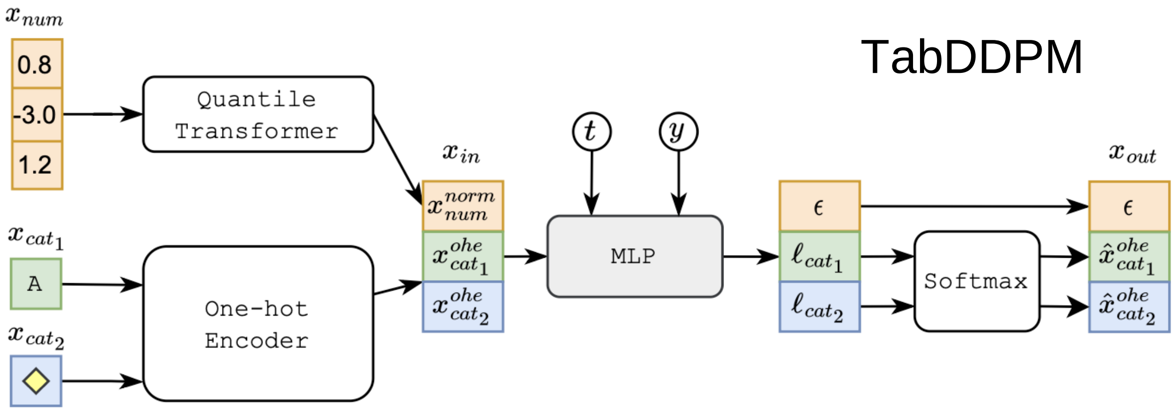
CTABGAN+: best tabular model with generative adversarial networks

TabDDPM: Diffusion model backbone

SMOTE



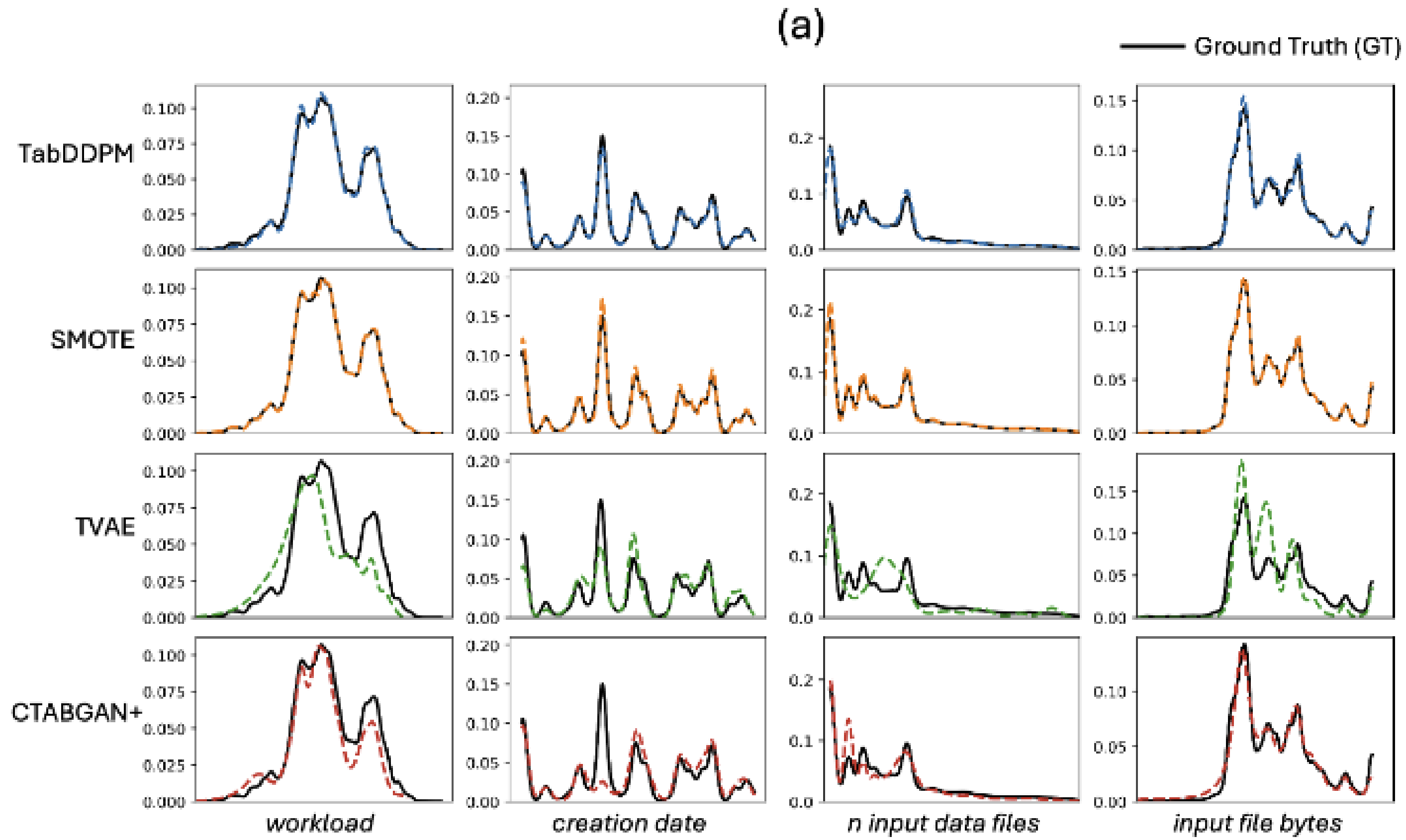
TabDDPM

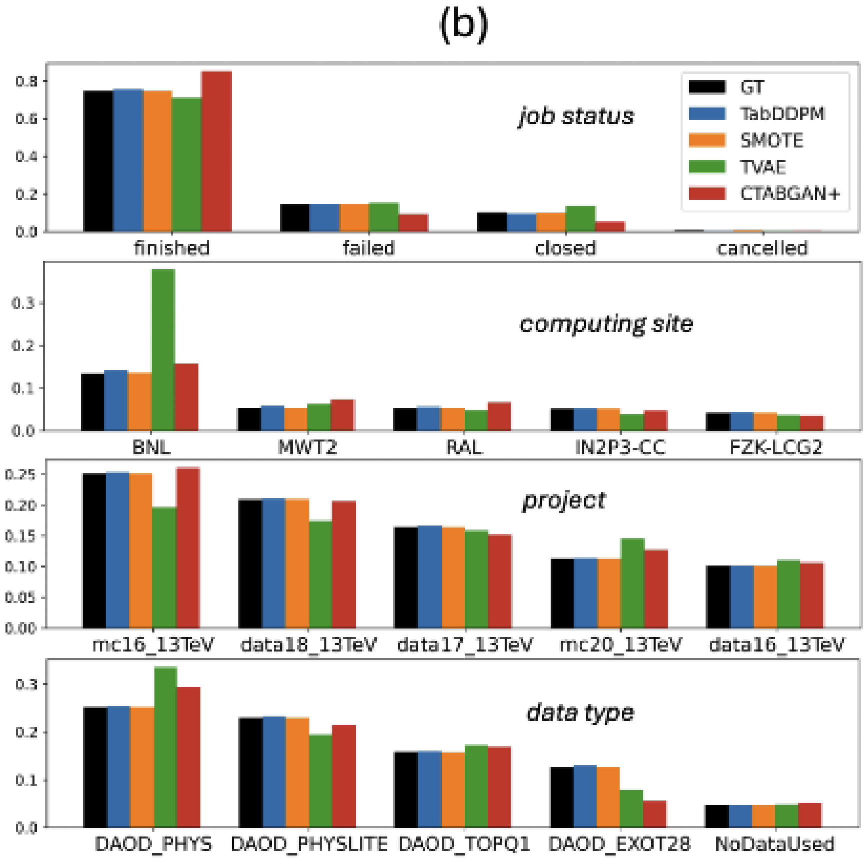


Brookhaven National Laboratory

Measuring Generative Performances: Results

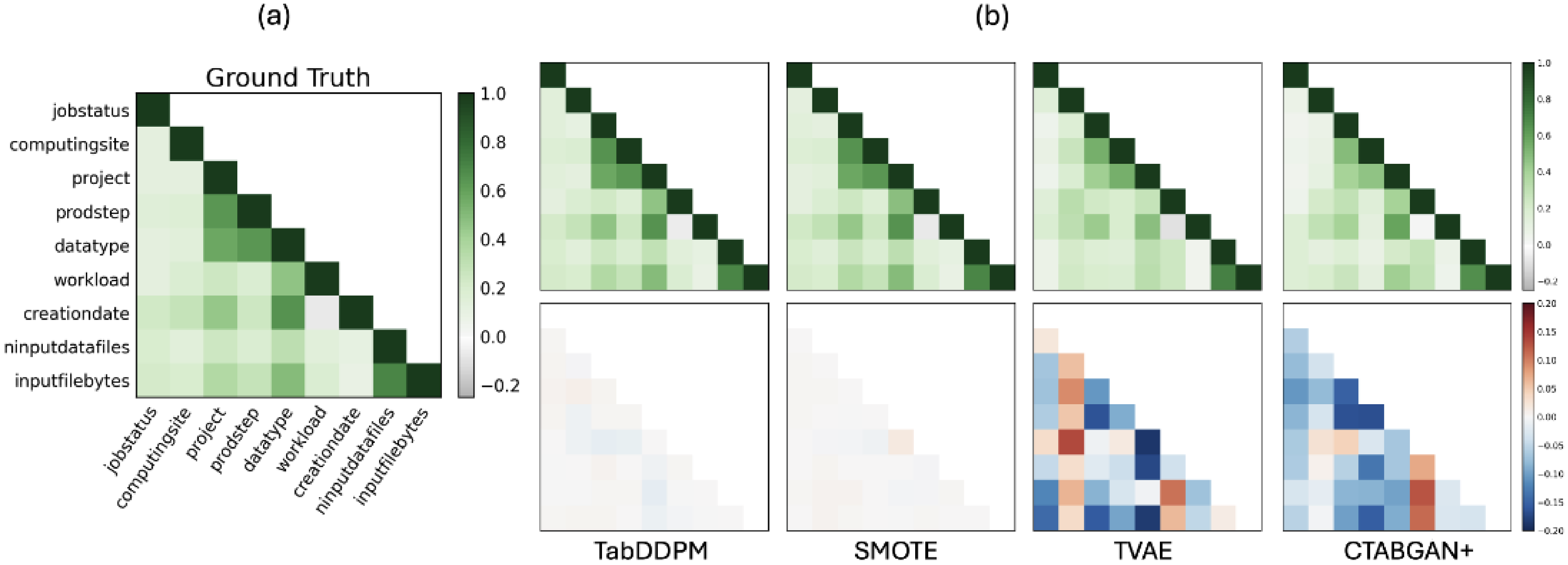
(1) Per-feature evaluation





Measuring Generative Performances: Results

(2) Correlations between feature pairs



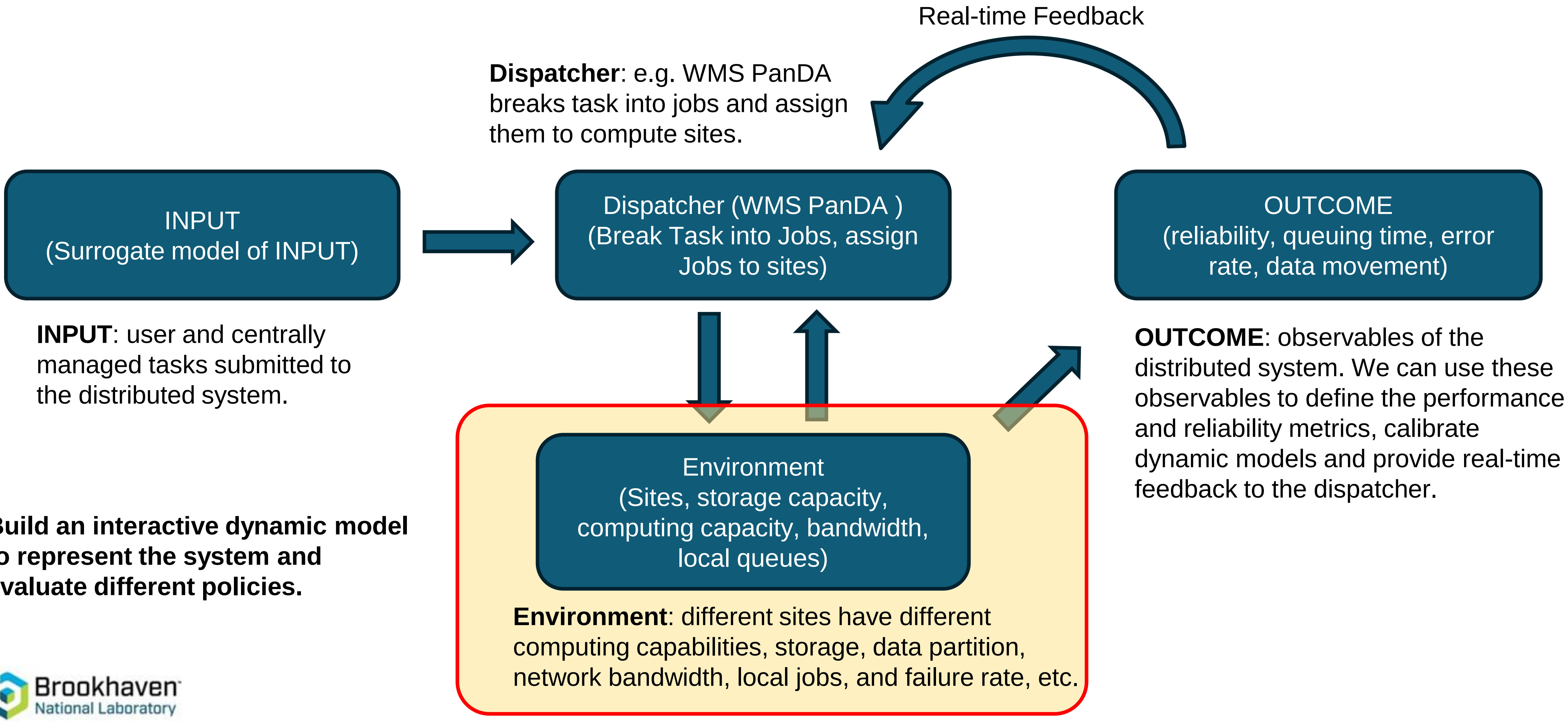
Measuring Generative Performances: Results

(3) Minimizing privacy risk: distance to closest record (DCR)

TABLE I
PERFORMANCE COMPARISONS ON SURROGATE MODELS.

Model	WD ↓	JSD ↓	diff-CORR ↓	DCR ↑	diff-MLEF ↓
TVAE	0.961	0.806	0.653	**0.143**	5.875
CTABGAN+	1.0	0.820	0.658	0.105	10.464
SMOTE	**0.871**	**0.799**	**0.011**	0.001	**0.058**
TabDDPM	0.874	**0.799**	0.036	0.025	0.826

Four Interacting Components of the Dynamic Model



Build an interactive dynamic model to represent the system and evaluate different policies.

Brookhaven National Laboratory

Simulating Distributed Computing Environment

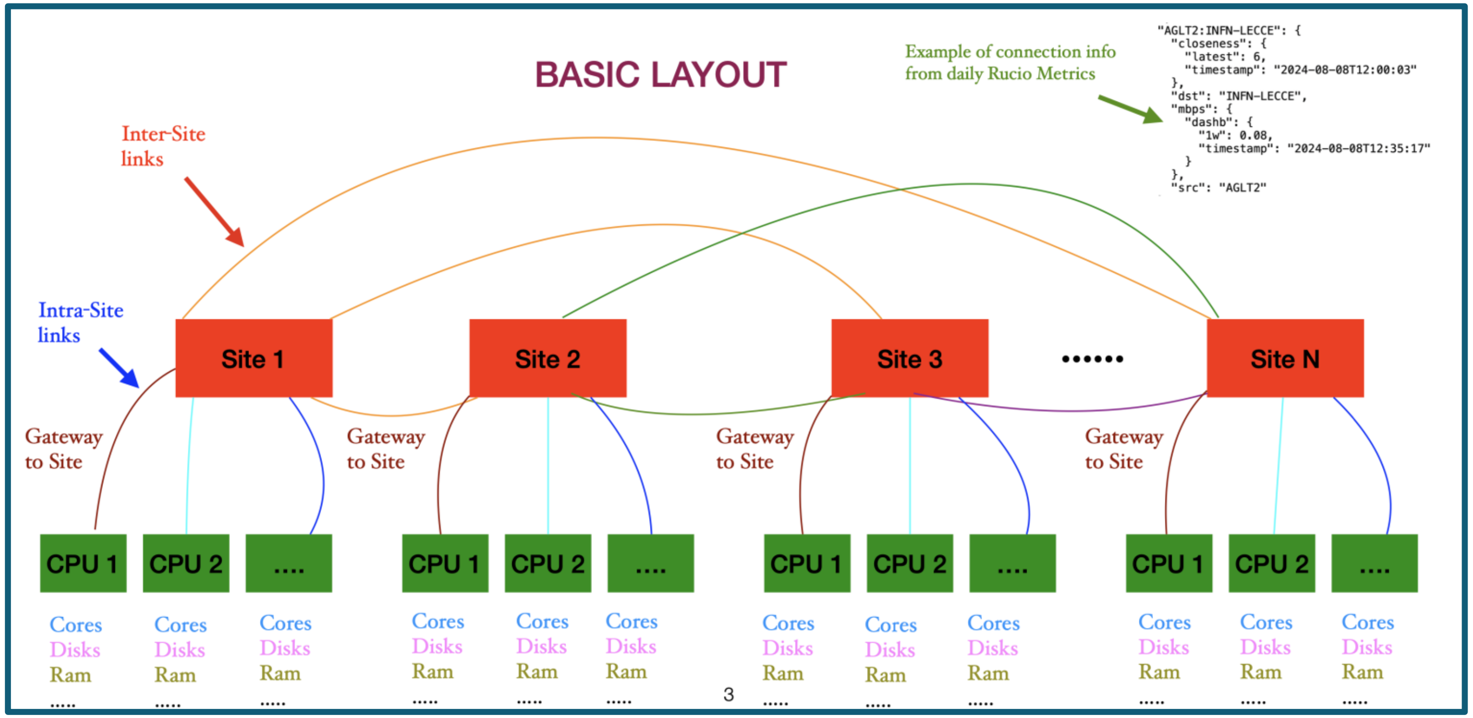
SIMGRID A framework for developing simulators of distributed applications targeting distributed platforms, which can in turn be used to prototype, evaluate and compare relevant platform configurations, system designs, and algorithmic approaches.

WRENCH An open-source framework designed to make it easy for users to develop accurate and scalable simulators of distributed computing applications, systems, and platforms. WRENCH is built op top of SimGrid.

DCSim Simulator for the simulation of high energy physics workloads on distributed computing systems with clusters, worker nodes, storages and caches. Essentially this was developed for simulating CMS grid system.

Simulating Distributed Computing Environment

- **SimGrid** is implemented for distributed computing environment.



  - **Working together**: Raees Khan (Pitt), Sairam Sri Vatsavai (BNL), Joseph Boudreau (Pitt), Paul Nilsson (BNL), Frederic Suter (ORNL)

Simulating Distributed Computing Environment

- Work-in-progress
  - Working on SimGrid, WRENCH and DCSim in parallel, and comparison studies are underway.
  - Integration of more **realistic statistics** for real job records and **surrogate models** into the simulation.
  - **Discover and address the shortcomings of SimGrid simulation** in, but not limited to, monitoring, dataset movements, delays of datasets and computing resources, computing sizes, initial dataset placement, etc.

Summary

- Delivered the surrogate model for 150-day WMS PanDA records: publication to appear at SC24 AI4Science workshop.
- The surrogate model successfully learns the joint distribution of WMS PanDA table as well as the time dynamics.
- Simulation modeling effort for distributed computing environment is underway based on three frameworks: SimGrid, WRENCH, and DCSim.
- Reflecting real data inputs from the surrogate model for SimGrid is the next step for speeding up event-based simulation.

Brookhaven National Laboratory