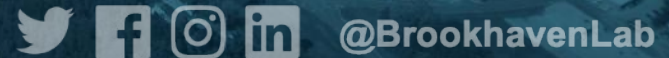




REDWOOD Track 1: High Throughput Workflow (HT-WF) Activity Summary

Tadashi Maeno (BNL)
On behalf of Track 1 Team



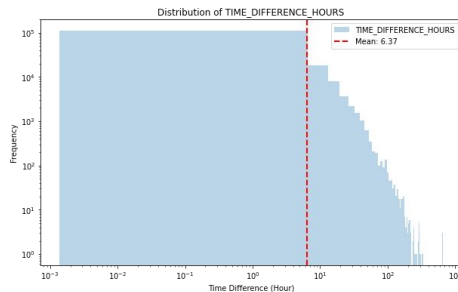
15 Jan 2025
REDWOOD General Meeting

Deliverables

- Grid system modeling → To be reported in T4
 - Paul, Raees, Fred
- Workload scheduling and data placement optimization
 - Shengyu
- Resource requirement prediction for scout jobs
 - Tasnuva
- Error classification
 - Tatiana + ATLAS
- Anomaly detection and auto retry action creation
 - To be assigned
- System Metrics
 - ATLAS

Resource Requirement Prediction for Scout Jobs

- **Scout jobs**
 - **Failure rate ~26%**
 - **~6.37 hours scouting time per Task**
- With ML pipeline
 - All 4 resource requirements for each workload upfront, in less than 60 sec
- The prediction will determine resource requirements for each workload upfront, an crucial building block in dynamically optimizing resource usage and enhancing system resilience



Work In Progress

- ✓ Automated model prediction pipeline package
- Deployment in
 - Testing environment [Jan 20-24]
 - Production [Jan 27-31]
- Model Evaluation: (Online Data)
 - Impact on Total Task completion time
 - Overall and individual model error on a weekly basis
- Paper draft (in progress)

Error Classification

➤ Goals

- Distinguish errors, e.g. temporary vs fatal, user-own vs system-own, etc
- Split catch-all errors to be more fine-grained
- Automate classification based on error patterns, characteristics, etc

➤ Steps

- New error table in the database ✓
- Logic in PanDA ✓
 - E.g. Not to penalize users in case of system-own errors
 - Git repo for data collector [[link](#)]
- Interface for error management ←
- Automation

Step 1: Error table

Jammel+Fernando

ID	Source	Code	Message	Description	...	Status
1	Pilot	12	Site error	...		Confirmed
			...			
x	Pilot	135	Site error x	...		Suggested

Step 3: Interface for error management

Jammel+Tatiana

ATLAS PanDA Dash Tasks Jobs Errors Users Sites Harvesters My PanDA
ATLAS PanDA monitor home

Confirmed errors

Source	Code	Message	Description	...
Pilot	12	Batch crash	...	

Promote selected rows

Suggested errors

Source	Code	Message	Description	...

Suggest new rule from job

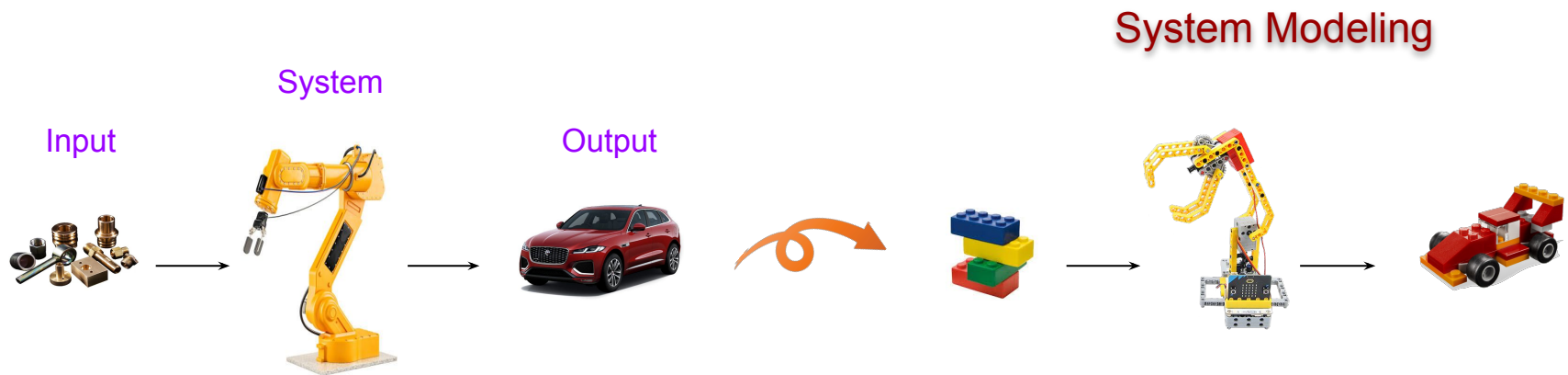
(This opens a dialog to choose a PanDAID and it will propose to create a rule from the error)

System Metrics

- One of the most crucial ATLAS milestones
 - Joint effort between ATLAS and REDWOOD
- To quantify the system-wide effects resulting from any changes
 - Lack of intrinsic or obvious metrics reflecting system performance due to the system complexity
- Status (details)
 - Defined key metrics to evaluate system performance
 - Averaged composite queuing time
 - Task active time
 - Implemented metrics collector in Dec 2024
- Next steps
 - To identify explanatory variables (metrics potentially affecting the key metrics)
 - To be enhanced to system metrics modeling

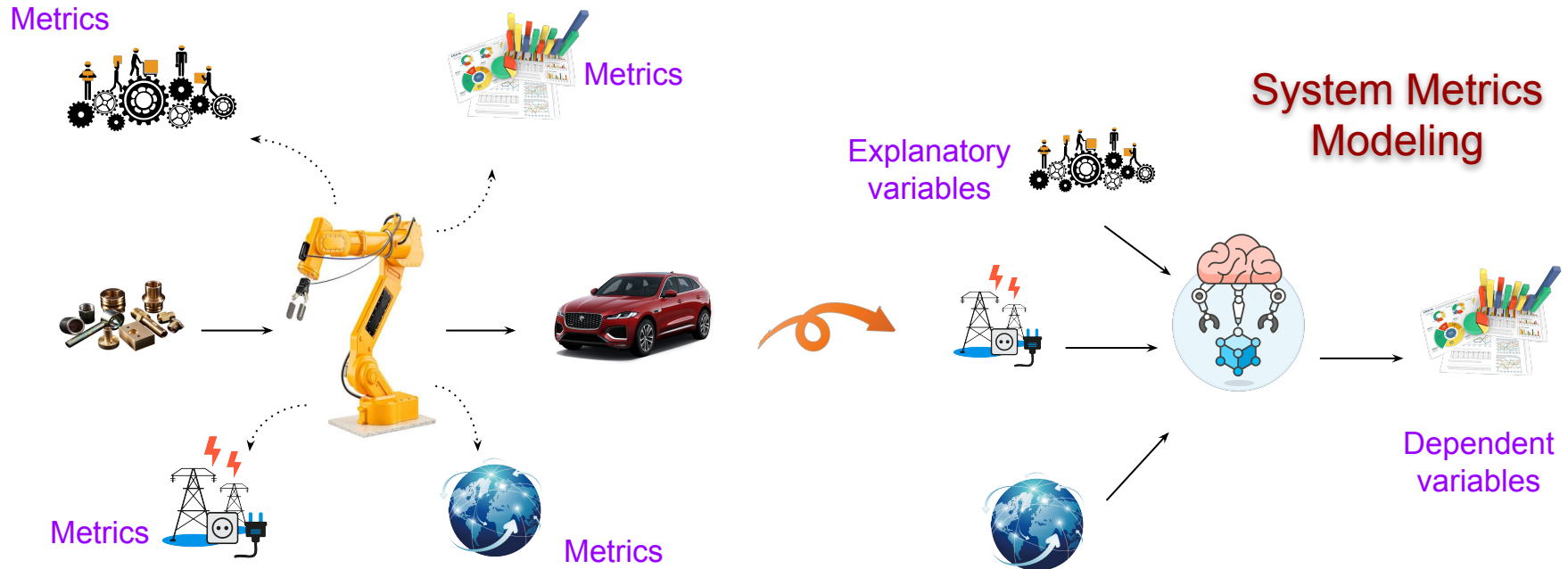
System Modeling vs System Metrics Modeling 1/2

- System modeling to build a model consist of logical representations directly mapped to system entities and processes
 - Tasks, jobs, nodes, batch queues, sites, storeages, network topologies, etc
 - Global and local fairshare, matchmaking and weight calculation in the brokerage, data motions, scheduling in batch systems, etc



System Modeling vs System Metrics Modeling 2/2

- System metrics modeling to build a model that explains relationship among various metrics
 - Measuring system behaviour through task, job, and system metrics
 - Categorizing metrics into dependent variables (outcomes = key metrics) and explanatory variables (causes)
 - Omitting description of physical system details



Monthly T1 Meeting

- It was proposed to establish monthly T1 meeting
 - The scope of general REDWOOD meeting is too broad
 - Weekly T4 meetings are focused on T4-specific
- Suggested scheduling for the first week of each month using the same time slot as the general meeting
 - Wed at 12:30 PM ET for 1 hour
 - TBC with CMU folks regarding their availability
- Presentation from one or two persons in rotation in every meeting, replacing the round-table of short reports
 - Each T1 person will give a summary report every 2~3 month