



Hewlett Packard
Enterprise



Framework for tracking metadata, lineage and model provenance in hybrid simulation-AI HPC exascale workflows

Martin Foltin, Andrew Shao, Rishabh Sharma, Shreyas Kulkarni, Annmary Justine Koomthanam, Aalap Tripathy, Cong Xu, Suparna Bhattacharya, Paolo Faraboschi: AI Research Lab, Hewlett Packard Labs
Wenqian Dong: Oregon State University
Brian Sammuli: General Atomics

May 8, 2025

Outline

Problem Statement

SmartSim

Common Metadata Framework (CMF)

Integration of CMF and SmartSim

Continual Learning Workflow

Example Use Case

Summary and Next Steps



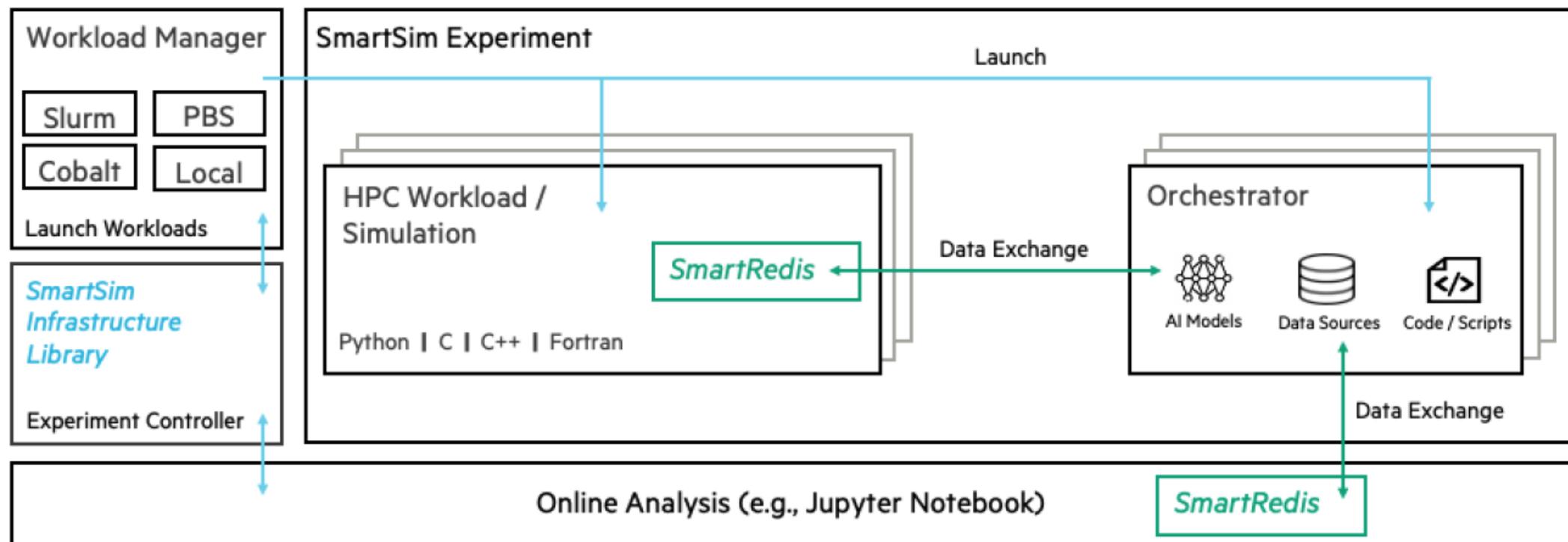
Problem Statement

- AI has the potential to significantly accelerate HPC simulations by providing simulation surrogate models (AI in the loop) and by automating simulation steering (AI over the loop)
- AI models in HPC simulations require periodic re-training to better align with input data distributions (i.e., correct for out-of-distribution data outliers, drifts, etc.)
- Tracking of the AI model provenance helps improve model quality
 - Enables a-posteriori analysis of model performance changes with new datasets, improving data sampling and model selection
 - Supports unwinding to previous iterations when performance degrades
 - Helps algorithms for catastrophic forgetting avoidance
 - Provides better understanding of the conditions where models perform well for future reuse
- AI model provenance tracking in hybrid AI-HPC workflows makes it difficult to use existing frameworks
 - AI-HPC workflows involve multiple components producing metadata at different volumes and velocities
 - Metadata and lineage from AI model inference may need to be recorded from every MPI rank at low overhead
 - Metadata and lineage tracking from simulation inner loop may require Fortran/C++ APIs to integrate to established codes
 - Metadata tracking from AI inference, simulation inner loop, and the workflow manager need to be combined in one framework



SmartSim

- Minimum changes required in Fortran/C/C++ simulation code: compilation with lightweight SmartRedis client
- AI model inference using familiar Pytorch / Tensorflow AI frameworks in SmartSim Orchestrator
- Online analysis (e.g., simulation results visualization) also possible by connecting to SmartSim Orchestrator

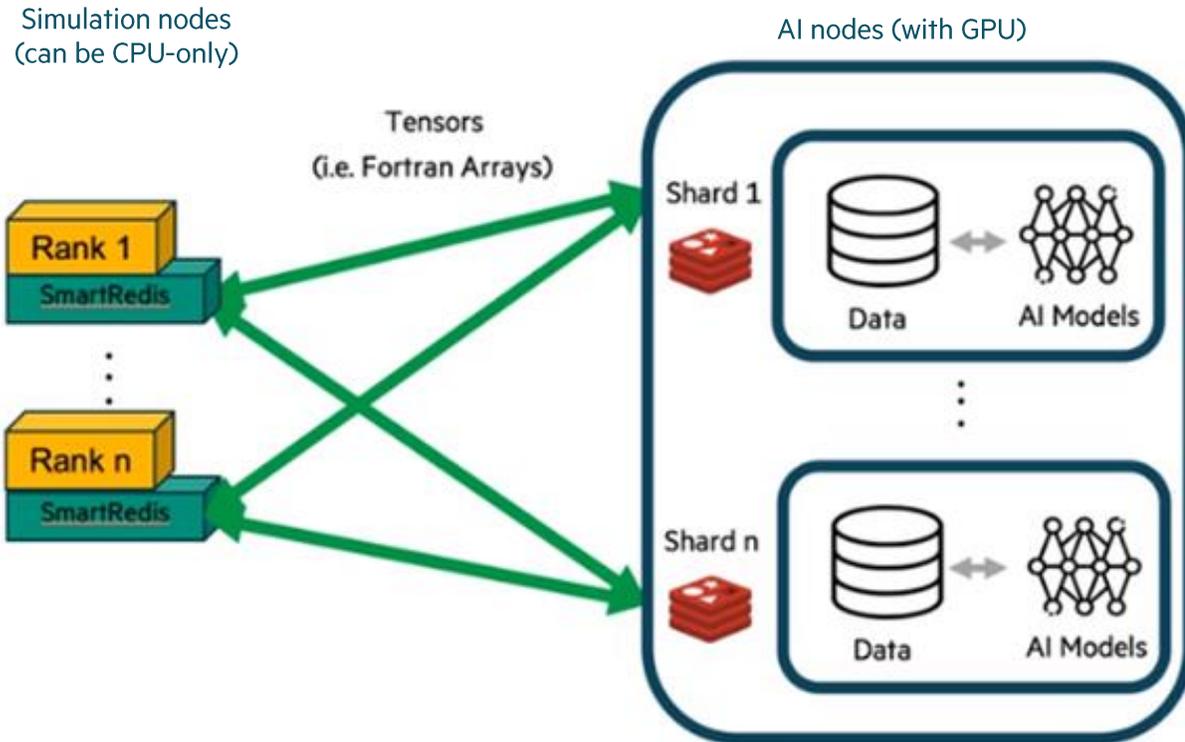


<https://github.com/CrayLabs/SmartSim>

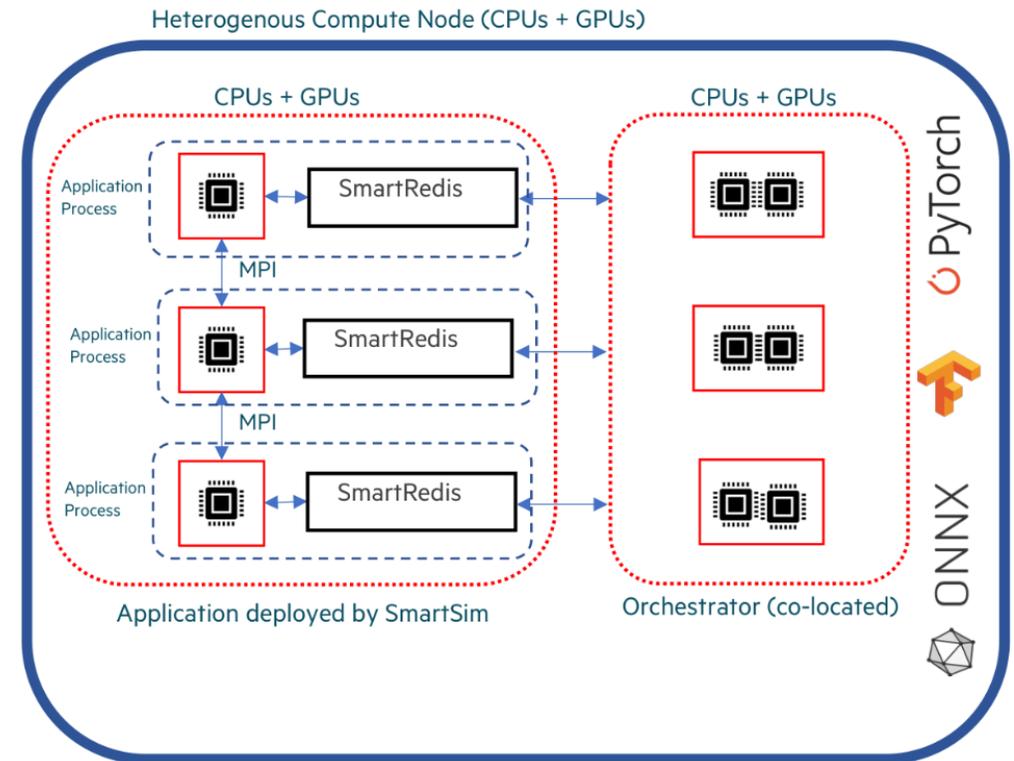
SmartSim Scalability

- Standalone deployment for high throughput: Orchestrator off simulation nodes, distributed among multiple AI nodes
- Co-located deployment for lowest latency: Orchestrators collocated with clients - simulation and AI on the same node

Standalone Deployment



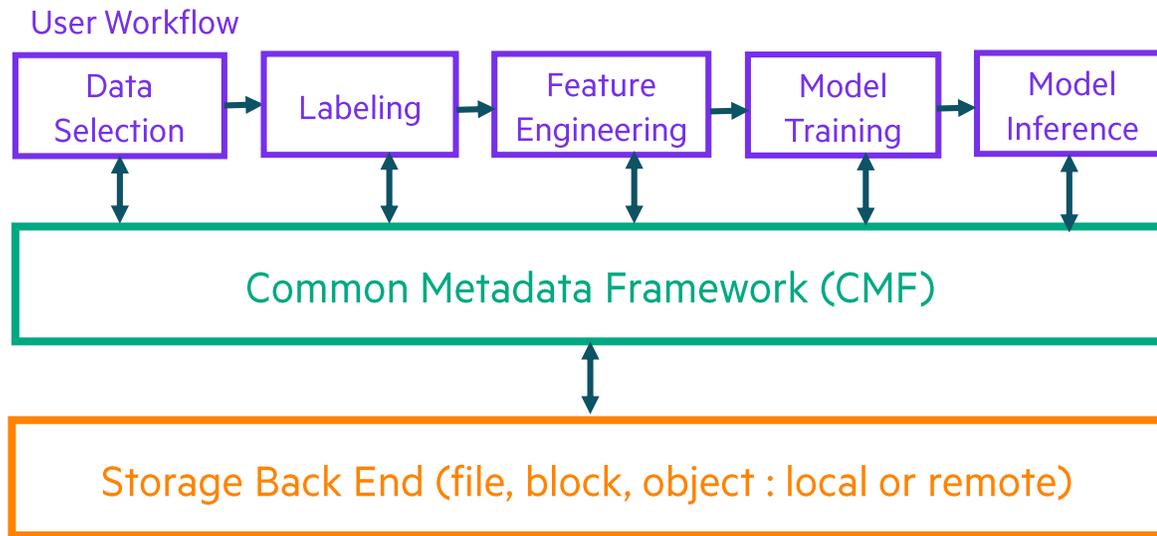
Co-located Deployment



Common Metadata Framework (CMF)

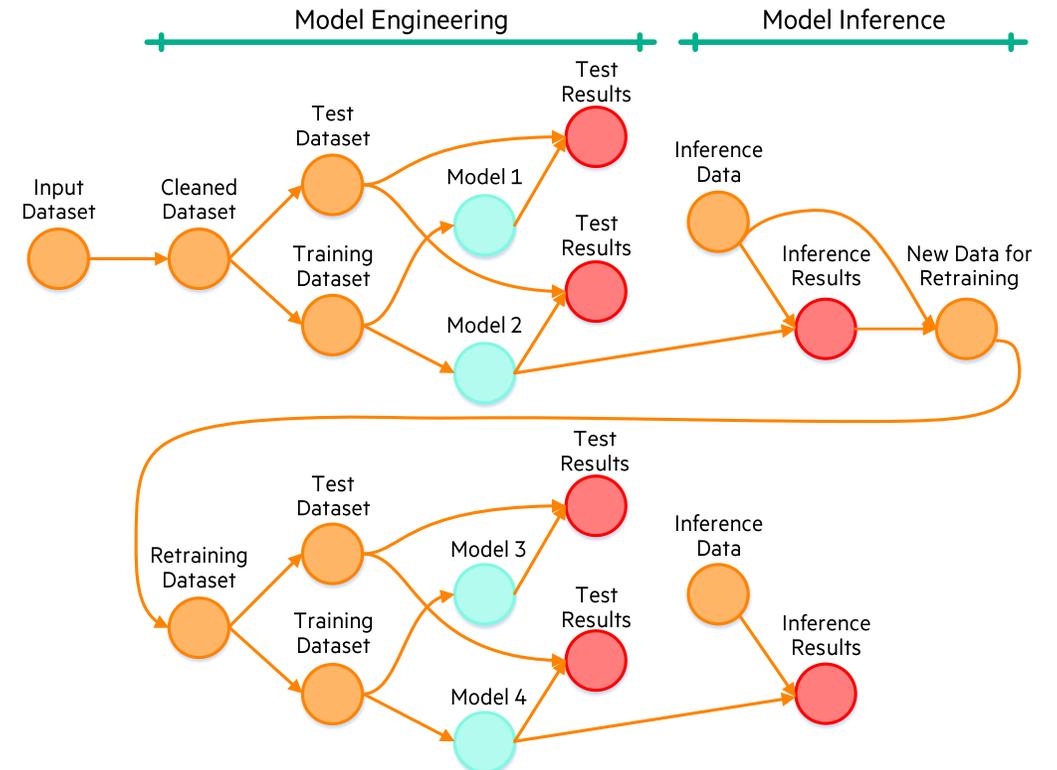
- A software layer between user workflow and storage back-end that records metadata, lineage and provenance
- Enables concurrent and scalable management of code, metadata, data and AI model lifecycle with Git-like simplicity
- Supports distributed workflows and tracks all workflows in one place for visibility and reuse

CMF in Software Stack (AI Model Training Example)



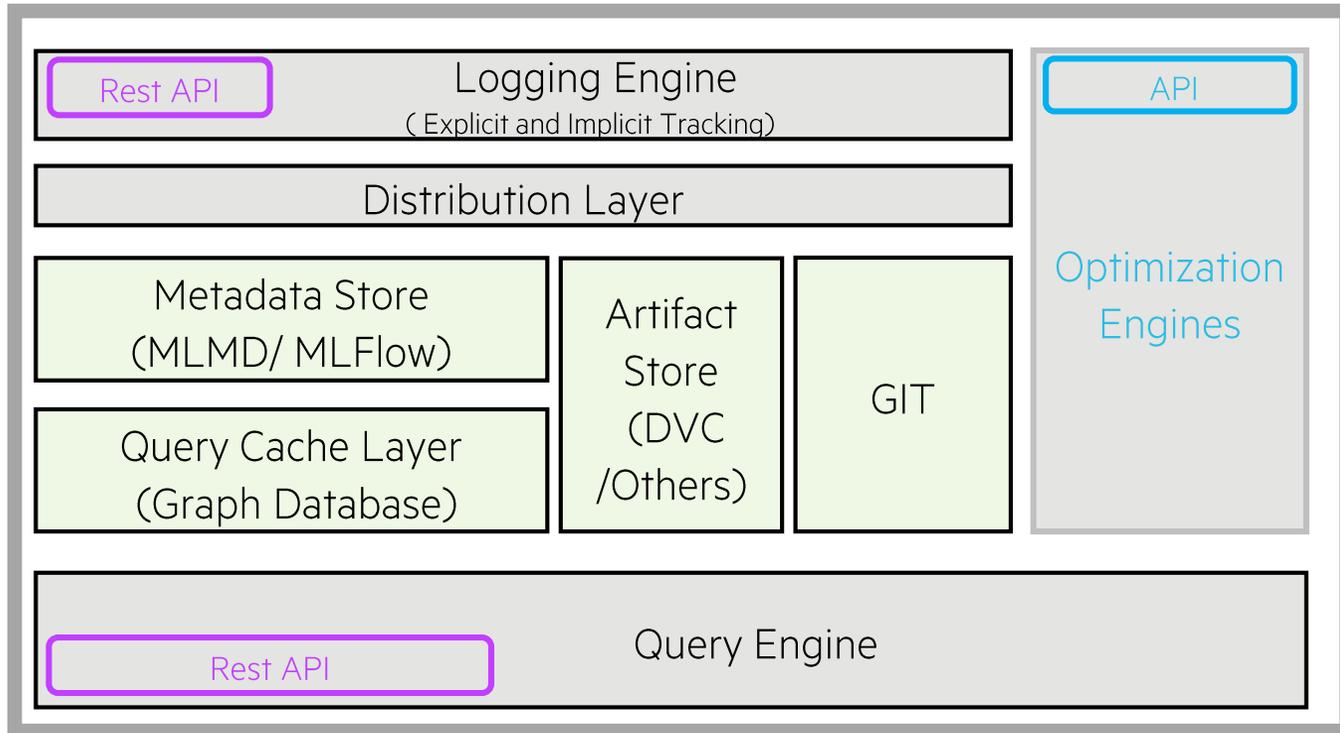
<https://github.com/hewlettPackard/cmf>

CMF Lineage Example (AI model Continual Learning)



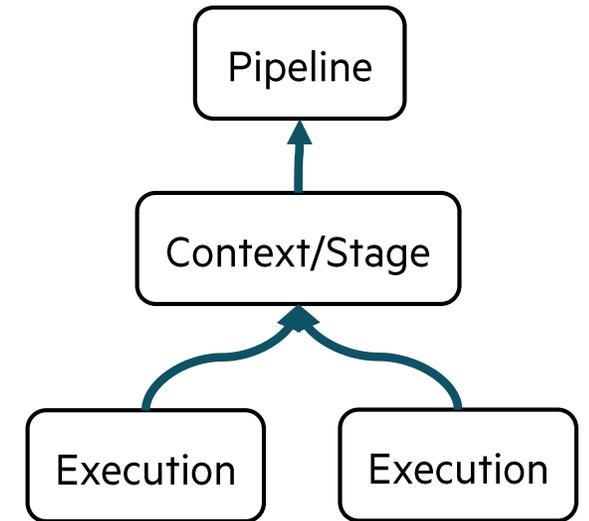
CMF Architecture

COMMON METADATA FRAMEWORK



- Common Metadata Framework
- Existing Frameworks
- Text Under Development

WORKFLOW ABSTRACTION



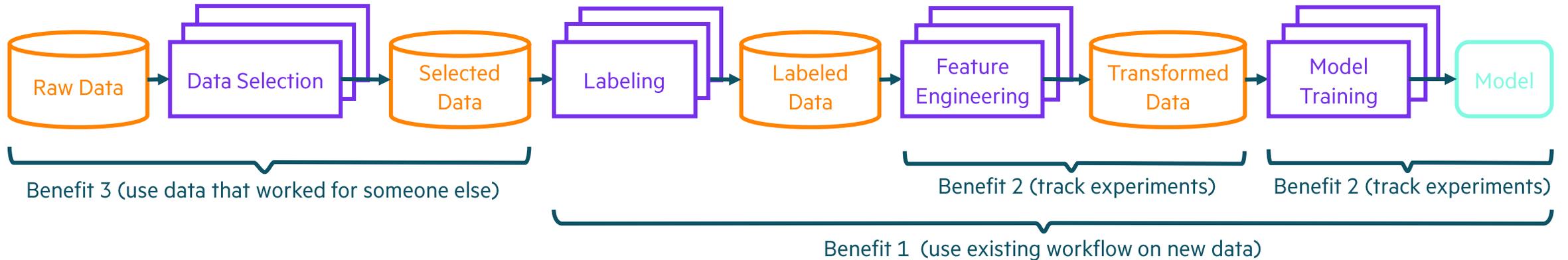
CMF library
(log, query, optimize)

CMF commands
(clone, push/pull)

CMF service

Input and Output Artifacts,
Parameters, Metrics, recorded for
each element

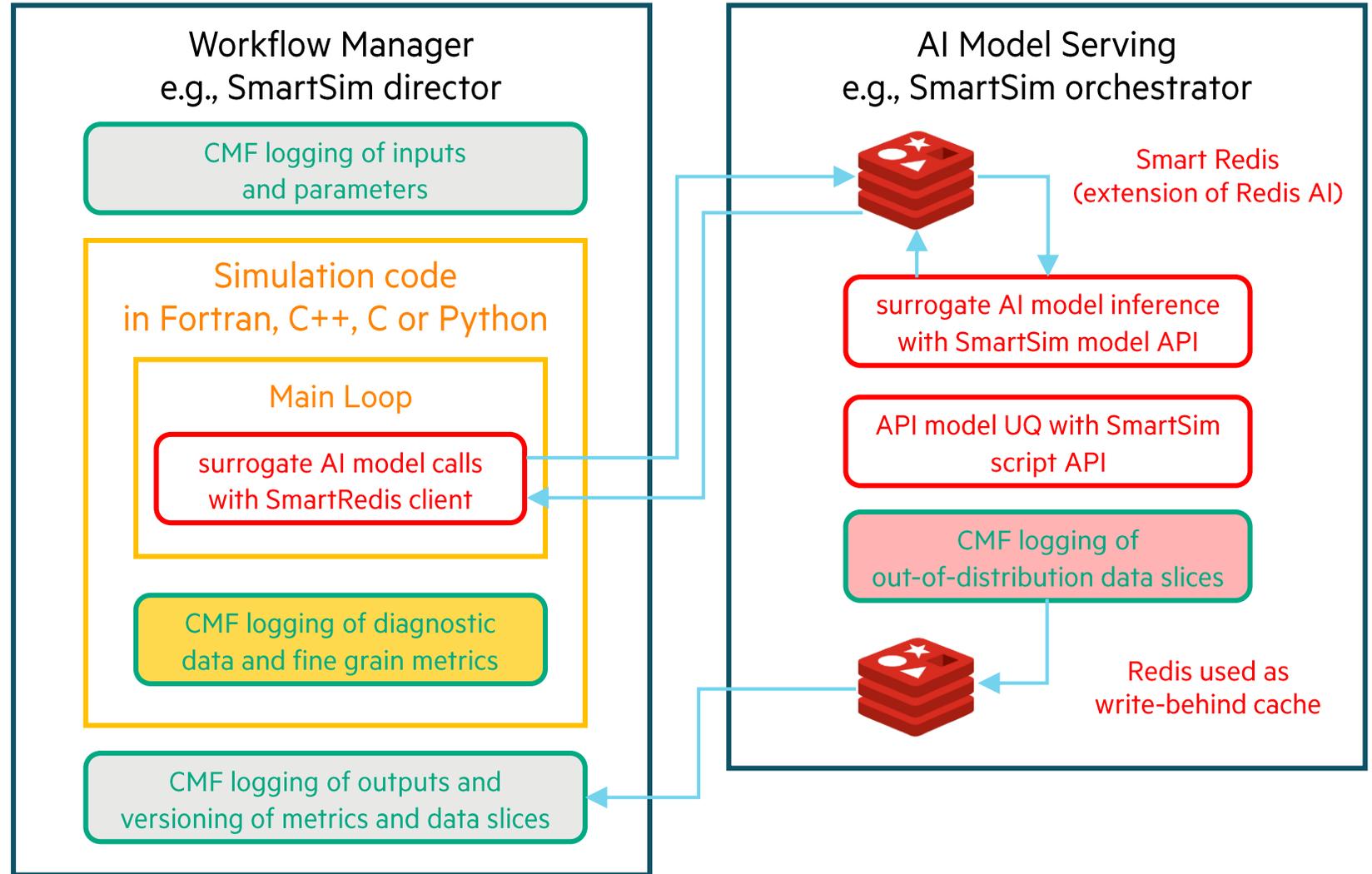
Example CMF Benefits in AI Workflows



1. Facilitates workflow reuse and collaboration
2. Simplifies management of AI experiments
3. Accelerates data discovery
4. Manages model lifecycle (as shown on previous slide)
5. Enables model certifiability

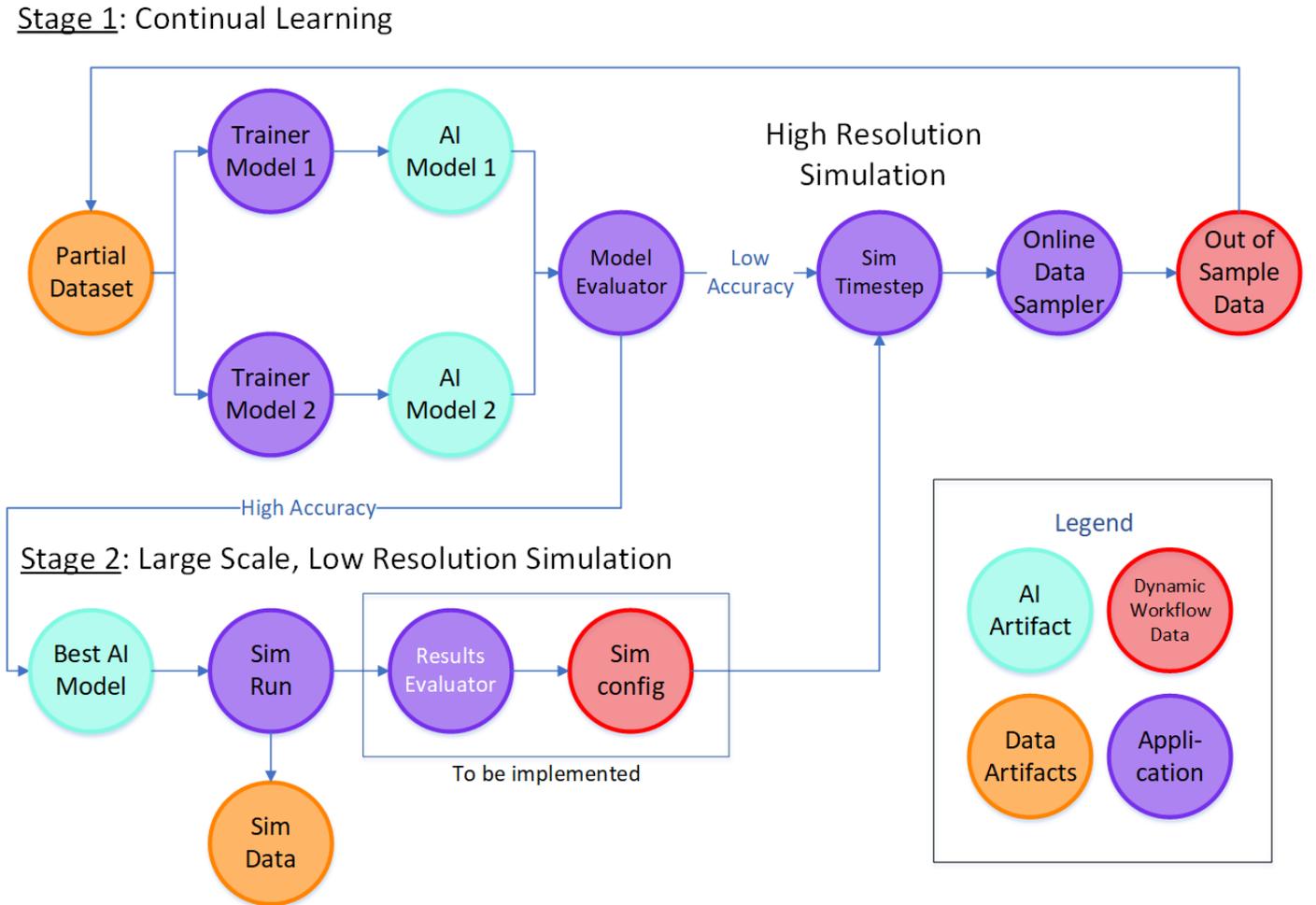
Integration of CMF with SmartSim

- Fortran/C++ APIs for logging of diagnostic data from simulation inner loop (CMF log_metrics)
- Blue-print for CMF tracking from SmartSim Orchestrator (e.g., out-of-distribution data)
- Write-behind caching of data logged in SmartSim Orchestrator
- Integration and versioning of metadata & artifacts tracked from:
 - Workflow Manager
 - Simulation Inner Loop
 - SmartSim Orchestrator



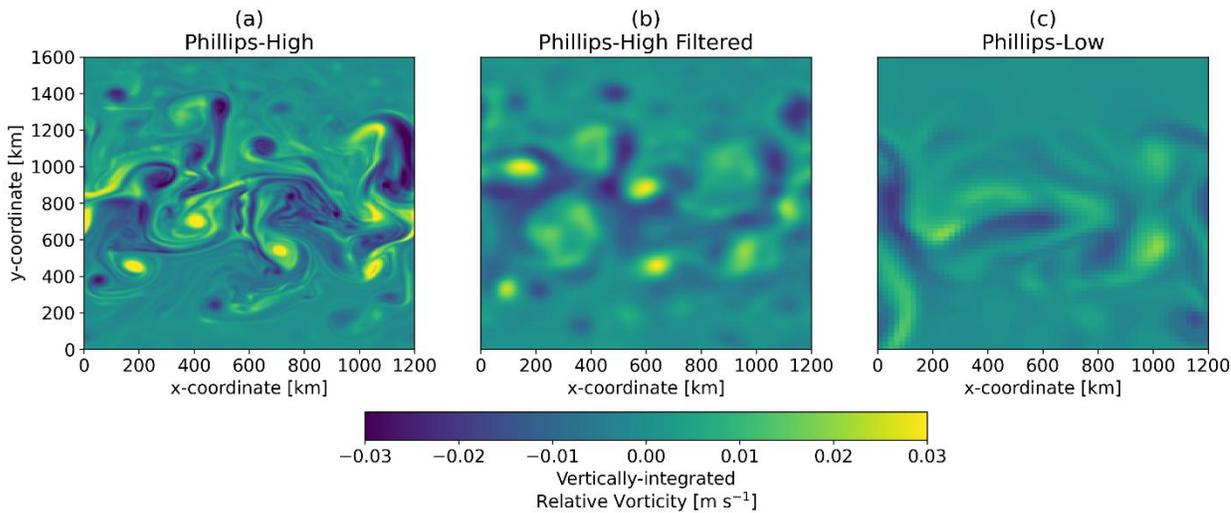
Acceleration of CFD Modeling Example

- AI model of eddy quantities (turbulence) accelerates CFD simulation $\geq 10x$
- Model periodically retrained for out-of-distribution data
- Data and trained model metrics tracked by HPE Common Metadata Framework (CMF)
- Workflow orchestration and coordination by HPE SmartSim
 - Large scale, coarse resolution simulation with AI inference
 - Small scale, fine resolution simulations providing training data
 - AI model training. The best from multitude of models selected based on CMF metrics

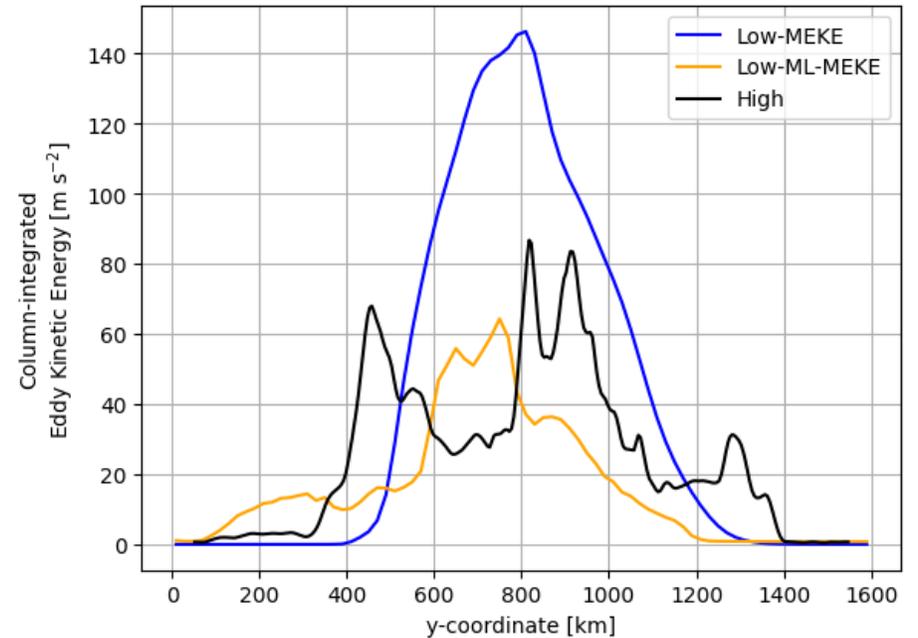


CFD Modeling Results

Similar scales of resolved features in (c) Phillips-Low coarse-resolution simulation with AI surrogate and in (b) Phillips-High Filtered high-resolution training set suggest that the coarse simulation is capable of resolving turbulence



Eddy Kinetic Energy (EKE) y-profile of Low-ML-EKE coarse-resolution simulation with AI surrogate corresponds better to High resolution ground truth than the Low-MEKE simulation with physics parametrization



Summary and Next Steps

- Developed a framework enabling model provenance tracking (incl. training datasets and model performance metrics) across different components of hybrid simulation-AI workflow
 - Workflow Manager
 - Simulation Inner Loop
 - AI Inference Orchestrator (e.g., SmartSim Orchestrator)
- Demonstrated benefits of this framework in enabling continual learning of AI surrogate model for eddy quantities (turbulence) in CFD simulations
 - Model performance metrics guide selection of training datasets and of an optimal model
- Working on applying this framework in other science areas
 - Magnetic Fusion Kinetic Equilibrium simulations
 - Molecular Dynamics simulations of nanoparticle structures for catalysis



Acknowledgements

- Foltin, Shao, Sharma, Kulkarni, Koomthanam, Tripathy, Dong, Xu, Battacharya, and Faraboschi all gratefully acknowledge the support of internal funding by HPE
- Some of the work by Foltin, Sharma, Kulkarni, Koomthanam, Tripathy and Sammuli were supported under Department of Energy award DE-SC0024409 and DE-SC0024426



Thank you

martin.foltin@hpe.com

