HPC Workflows a' Comin'

CJ Newburn, HPC Lead for NV Compute SW

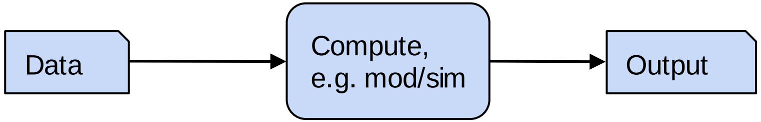
P3HPC Forum, Sep 2, 2020

- A new era: workloads → workflows
- HPC x AI
- HPC at the Edge
- Towards a metaflow and common infrastructure
- Where, what, and how
- The 3 Ps

An invitation to collaborate

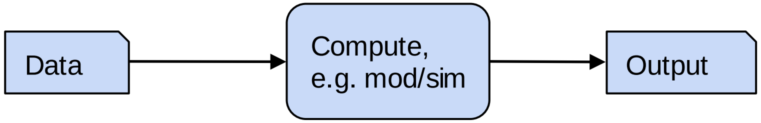
Workloads: First there was simple HPC

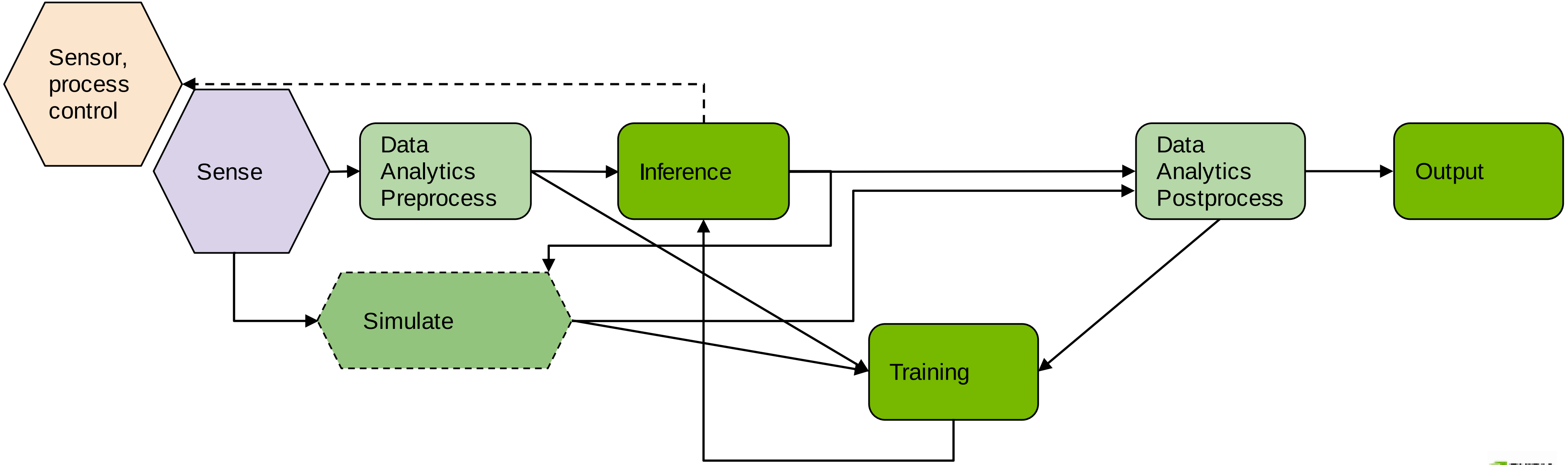
One and done



Workflows: Then there were mashups - light, accurate, realtime

Complex inter-relationships of stages vs. one and done





Common interests for workflows

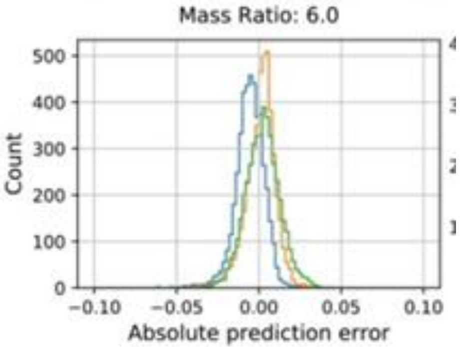
Relatively-new characteristics

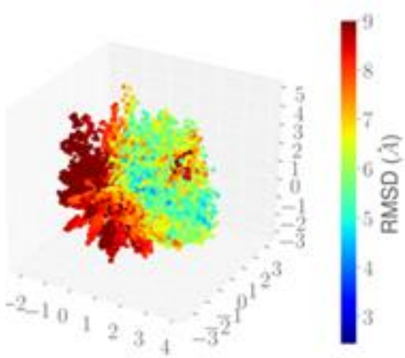
- HPC practitioners are being presented with a new set work work~~load~~flows
  - **Mashup** of HPC, data analytics, deep learning, visualization
  - Data + **metadata**
  - Common data **representation** and **interoperability** are key
  - Moving toward higher-level **abstractions**, reusable infrastructure
  - **Projection** of numerical data into consumable insight (vis) with automated action
  - **Orchestration**, management, containers
  - **Monitoring** for bottlenecks and continuous improvement
- Performance, portability, and productivity all have a part to play

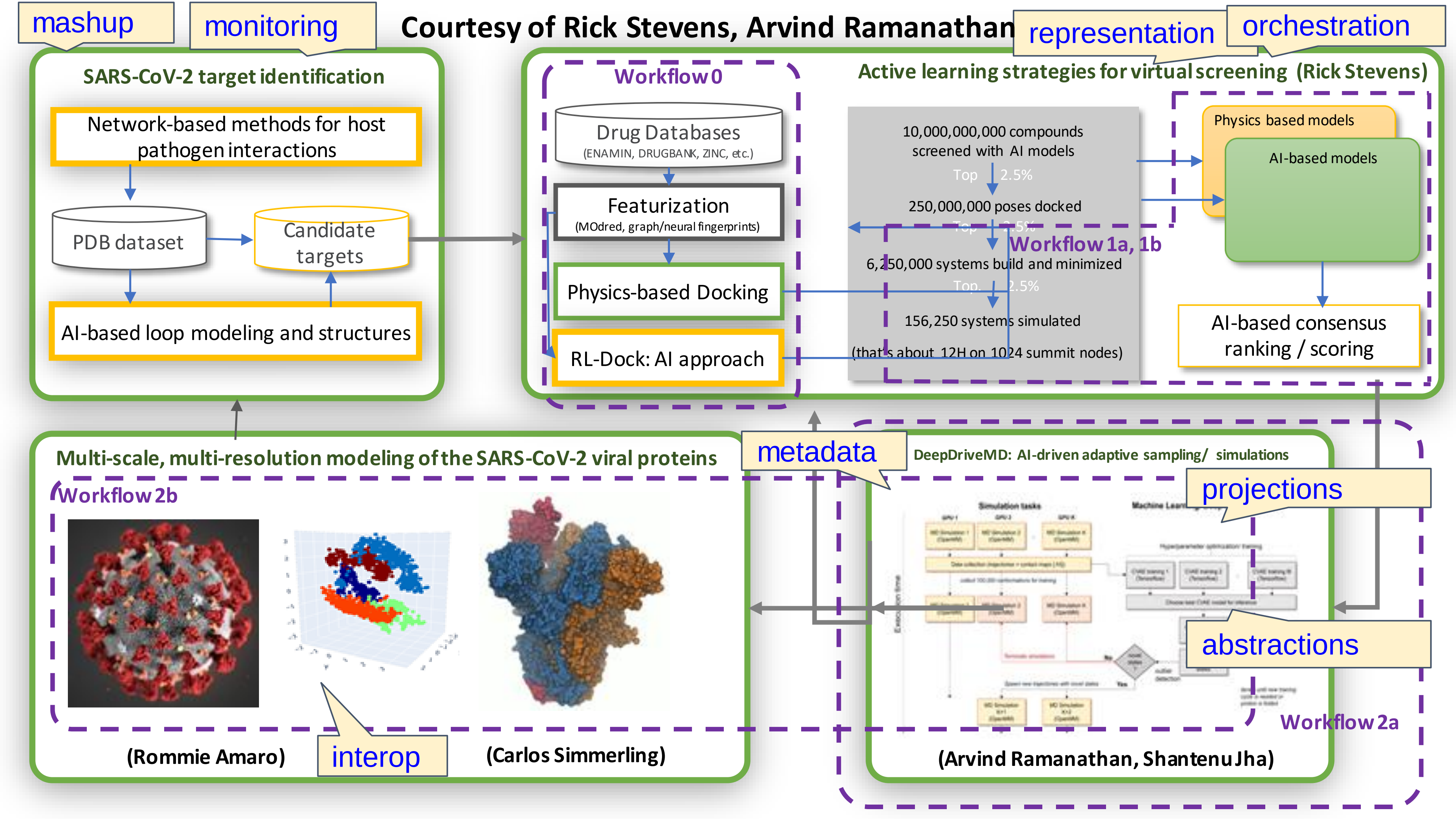
Fully-integrated AI-assisted HPC simulation

Enables new science in areas like CoViD-19

- DeepDriveMD: protein folding, docking
- ATOM: molecular design tools
- ProtTrans: computational biology ~ natural language processing (link)
- Physics-inspired DL to characterize black hole mergers (link)
- ECMWF: Weather modeling



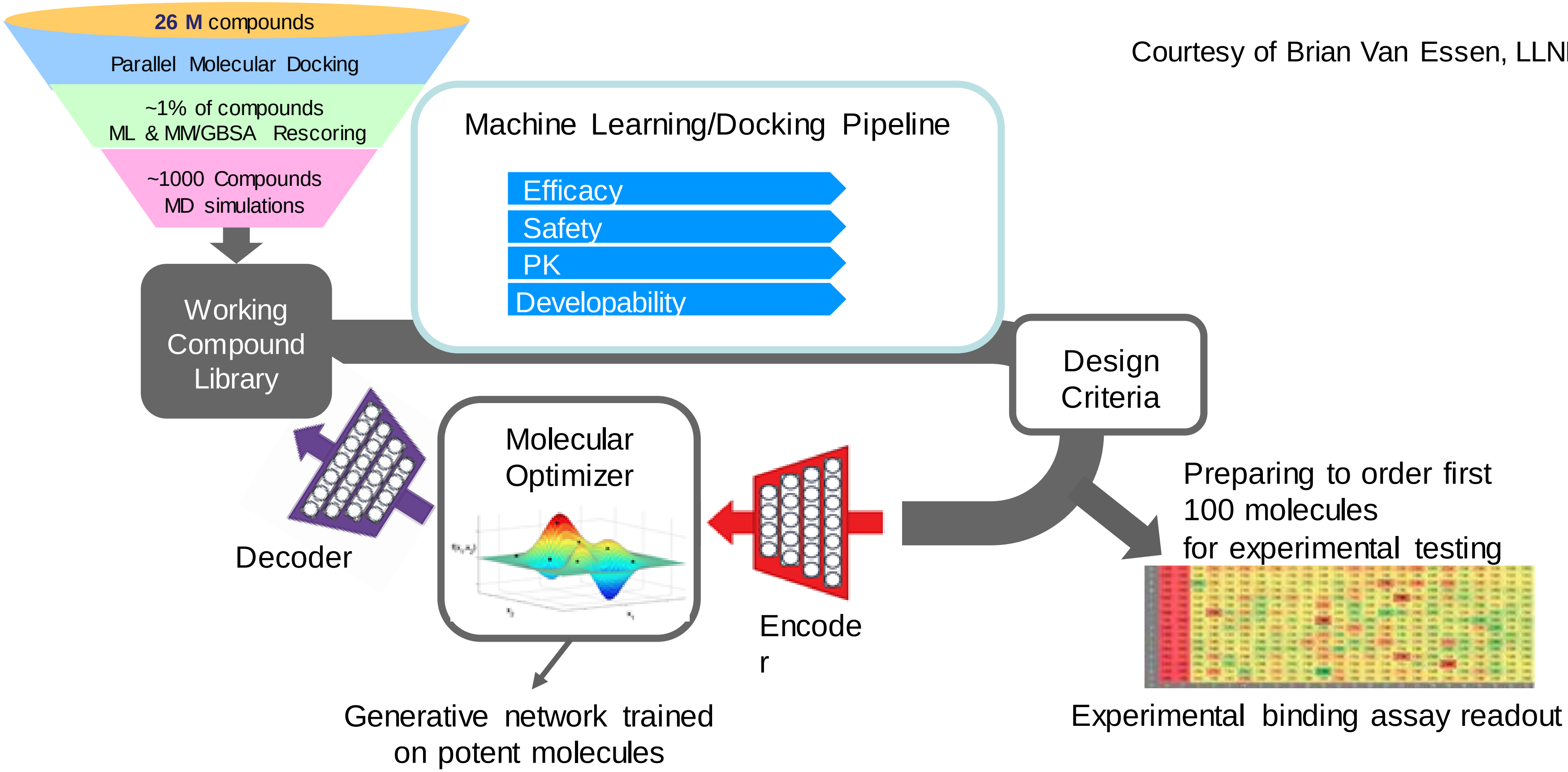




Courtesy of Rick Stevens, Arvind Ramanathan
mashup
monitoring
representation
orchestration
SARS-CoV-2 target identification
Network-based methods for host pathogen interactions
PDB dataset
Candidate targets
AI-based loop modeling and structures
Workflow 0
Drug Databases
(ENAMIN, DRUGBANK, ZINC, etc.)
Featurization
(MOdred, graph/neural fingerprints)
Physics-based Docking
RL-Dock: AI approach
Active learning strategies for virtual screening (Rick Stevens)
10,000,000,000 compounds screened with AI models
Top 2.5%
250,000,000 poses docked
Top 2.5%
Workflow 1a, 1b
6,250,000 systems build and minimized
Top 2.5%
156,250 systems simulated
(that's about 12H on 1024 summit nodes)
Physics based models
AI-based models
AI-based consensus ranking / scoring
Multi-scale, multi-resolution modeling of the SARS-CoV-2 viral proteins
metadata
DeepDriveMD: AI-driven adaptive sampling/ simulations
Workflow 2b
projections
Simulation tasks
abstractions
Workflow 2a
(Rommie Amaro)
interop
(Carlos Simmerling)
(Arvind Ramanathan, Shantenu Jha)

ATOM design tools will be applied to propose new improved molecular structures

Courtesy of Brian Van Essen, LLNL

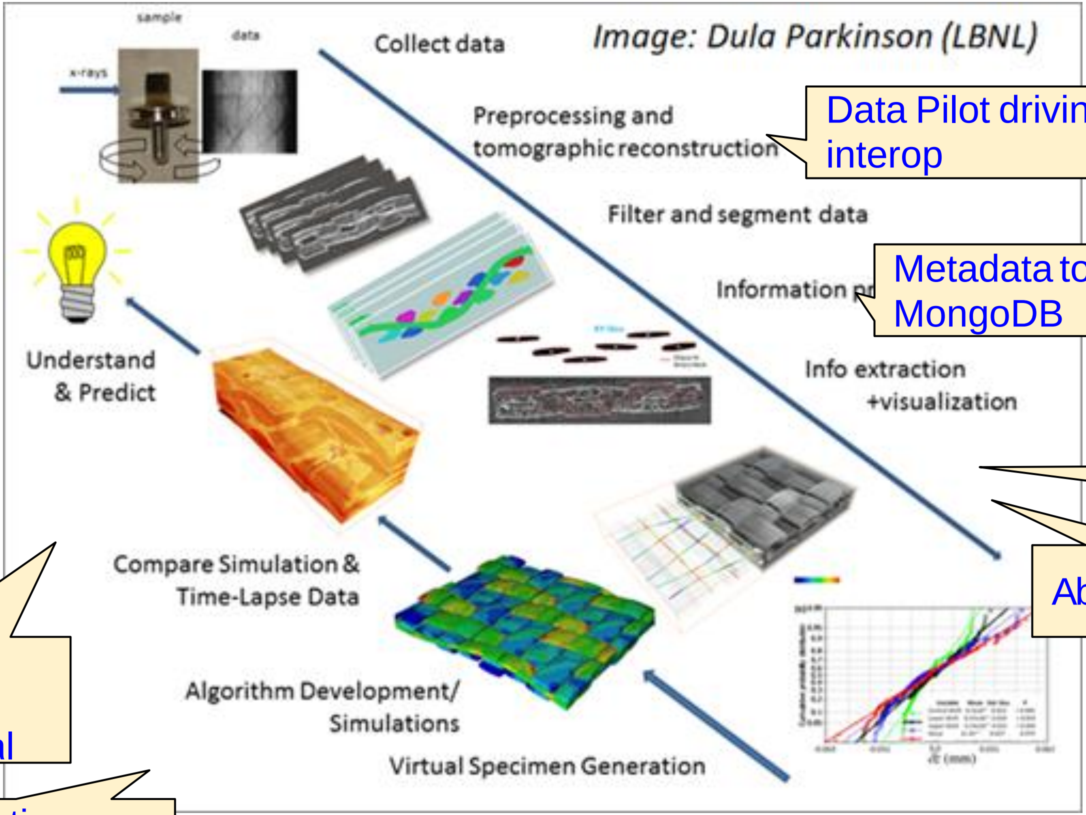


HPC at the Edge usage models

HPC at the Edge: realtime compute at edge → lower bandwidth, higher quality data back to DC

- Light sources at LBNL, BNL, ANL, SLAC
- ORNL: CNMS microscopy, FedSci, MANTiD, Manufacturing Design Facility
- VA Tech: Microtomography
- Health: NVIDIA Clara

Courtesy of Hari Krishnan and Alex Hexemer, LBNL.ALS



Data Pilot driving interop

mashup

Metadata to MongoDB

Projections: realtime vis for quality verification

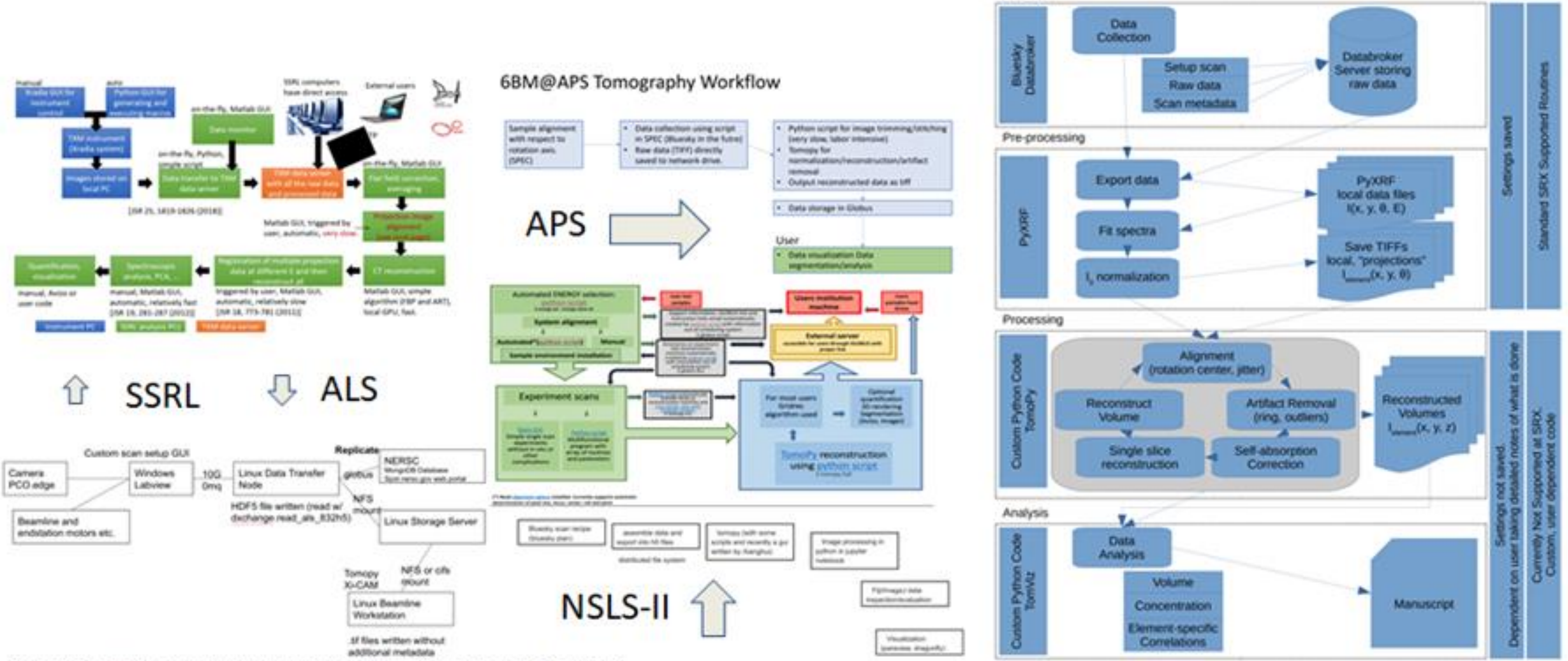
Abstractions

Orchestration, monitoring:
ad hoc → central

Representations:
MsgPack → Arrow?

Data pipeline typical of several beamlines at the ALS: imaging, scattering, micro-diffraction, hybrid.

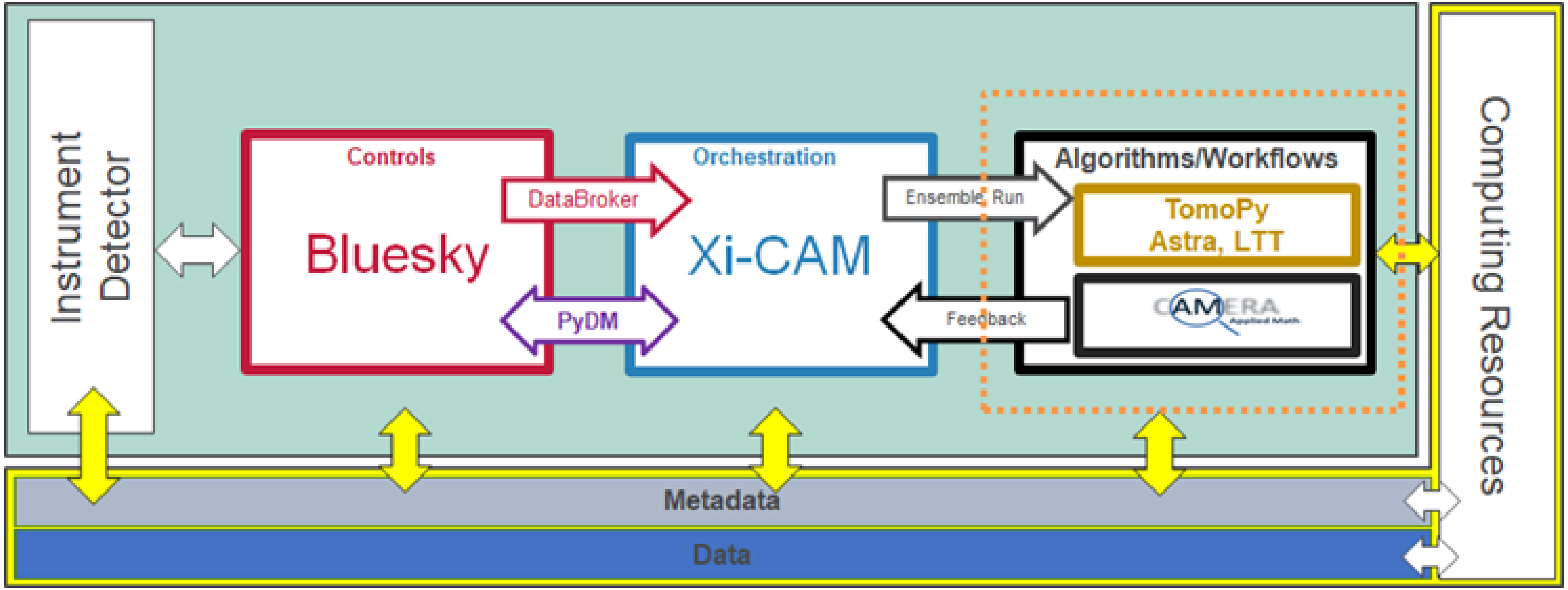
Multiple light sources do similar operations



Credit: Data Pilot Tomography Breakout Report (All credit goes to respective authors)

Toward shared DoE infrastructure

Build once, use everywhere



Data collection

Data sanitization

Noise reduction, crop padding, normalization

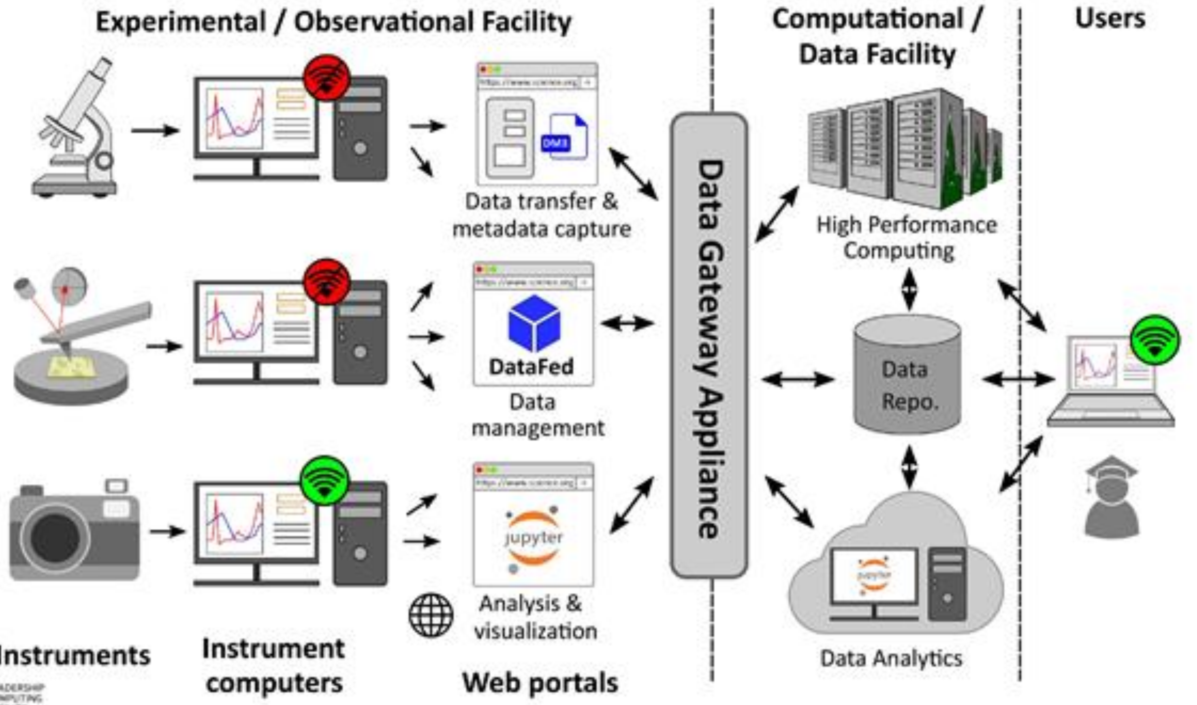
Visualization

Reconstruction segmentation/registration

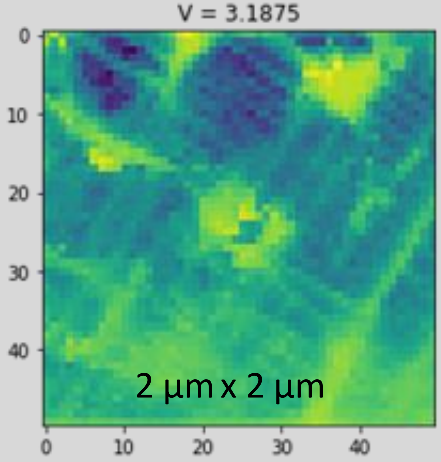
Toward commonality and integration on the data plane

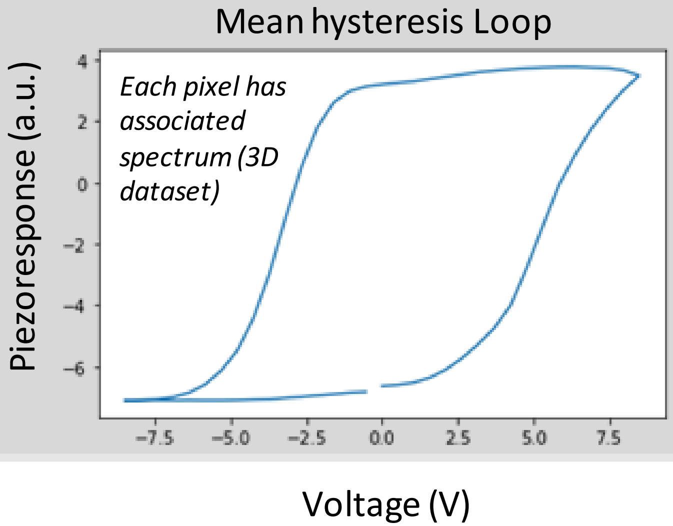
A Systemic Approach to Facilitating Reproducibility via Federated, End-to-End Data Management
Dale Stansberry, Suhas Somnath, Gregory Shutt, and Mallikarjun Shankar @ ORNL, Advanced Data and Workflows Group

- Inadequate metadata for AI
- Containers
- Orchestration/security
- Higher-level APIs
- Edge-to-exascale workflow (control, interactive steering, design of experiment)
- Data lifecycle management



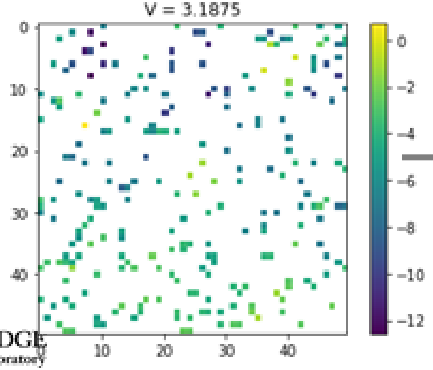
Courtesy of Rama Vasudevan, ORNL.CNMS





- Each measurement takes ~4 seconds
- We need to optimize sampling in the (x,y) space
- Utilize active learning

Sample 10% pixels initially



Dimensionality Reduction

mashup

RED = Edge Compute

interop

Gaussian Process to predict unobserved pixels

Representation: [pyUSID]

Criterion

E.g. max uncertainty
Max area
Etc.

Select Next pixels; Measure

Dimensionality Reduction

Oak Ridge National Laboratory

FedSci: SW Framework for Federated Science Instruments

T. Naughton, S. Hitefield, L. Sorrillo, N. Rao, J. Kohl, W. Elwasif, J-C. Bilheux, H. Bilheux,
S. Boehm, J. Kincl, S. Sen and N. Imam @ ORNL

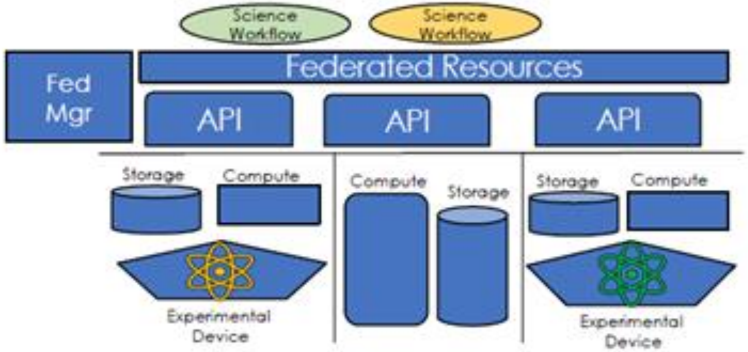
Perf: device → HPC resources

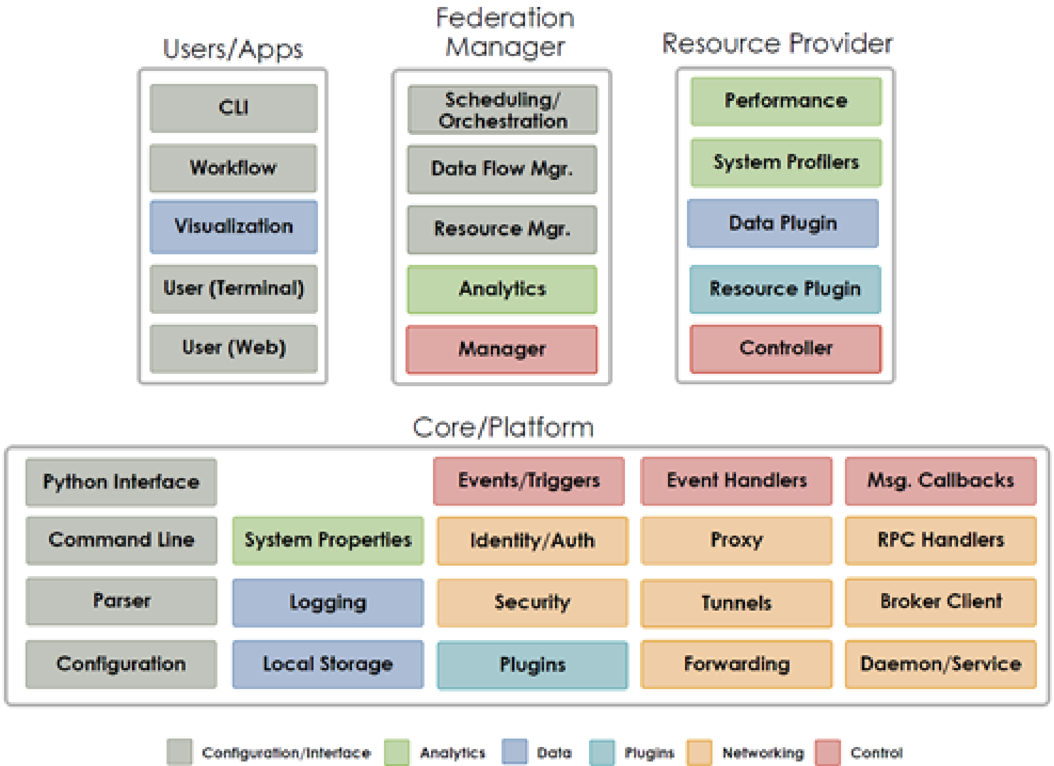
Portability: OSS toolkit, many sites

Productivity: time to discovery

Abstraction: plugins, Python APIs

Mashup, sched/orch, interop, vis



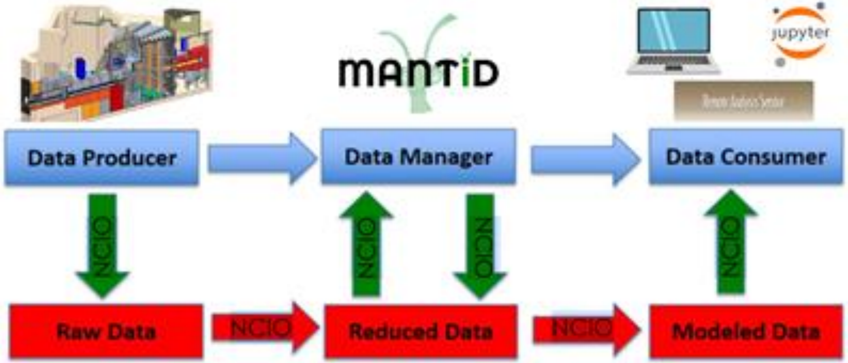


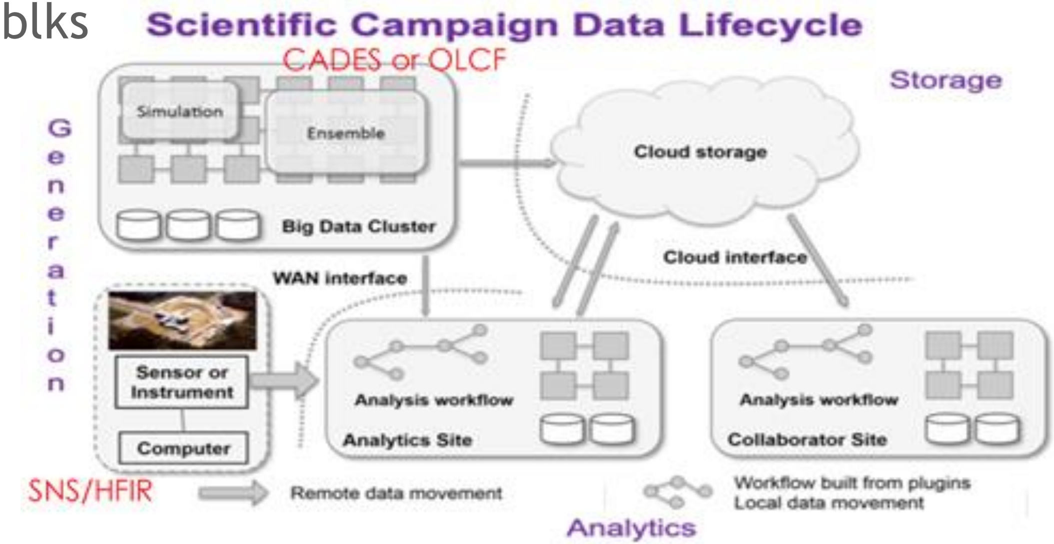
MANTiD @ ORNL

"Performance improvements on SNS and HFIR instrument data reduction workflows using Mantid"

William F Godoy, Pete Peterson, Steven Hahn, John Hetrick, Jay Billings @ ORNL

Perf: focused on IO
Mashup of HPC, analytics, DL, vis
Common data rep, interop
Higher-level abstractions, building blks
Monitoring



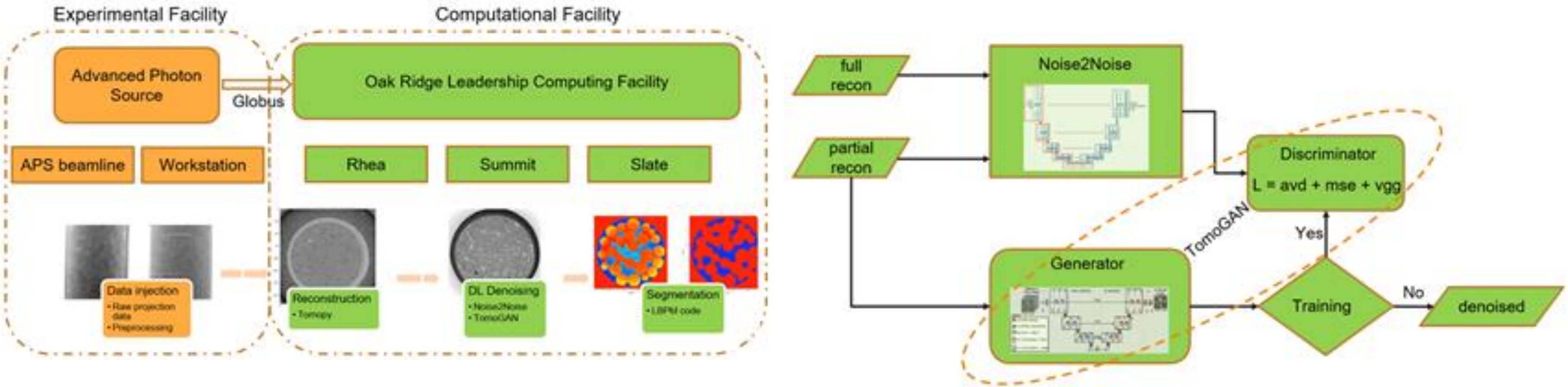


Micro-tomography @ VA Tech

Toward Real-Time Analysis of Synchrotron Micro-Tomography Data: Accelerating Experimental Workflows with AI and HPC: J McClure, J Yin, R Armstrong, K Maheshwari, S Wilkinson, L Vlcek, Y Wang, M Berrill, M Rivers

Perf: data bottleneck

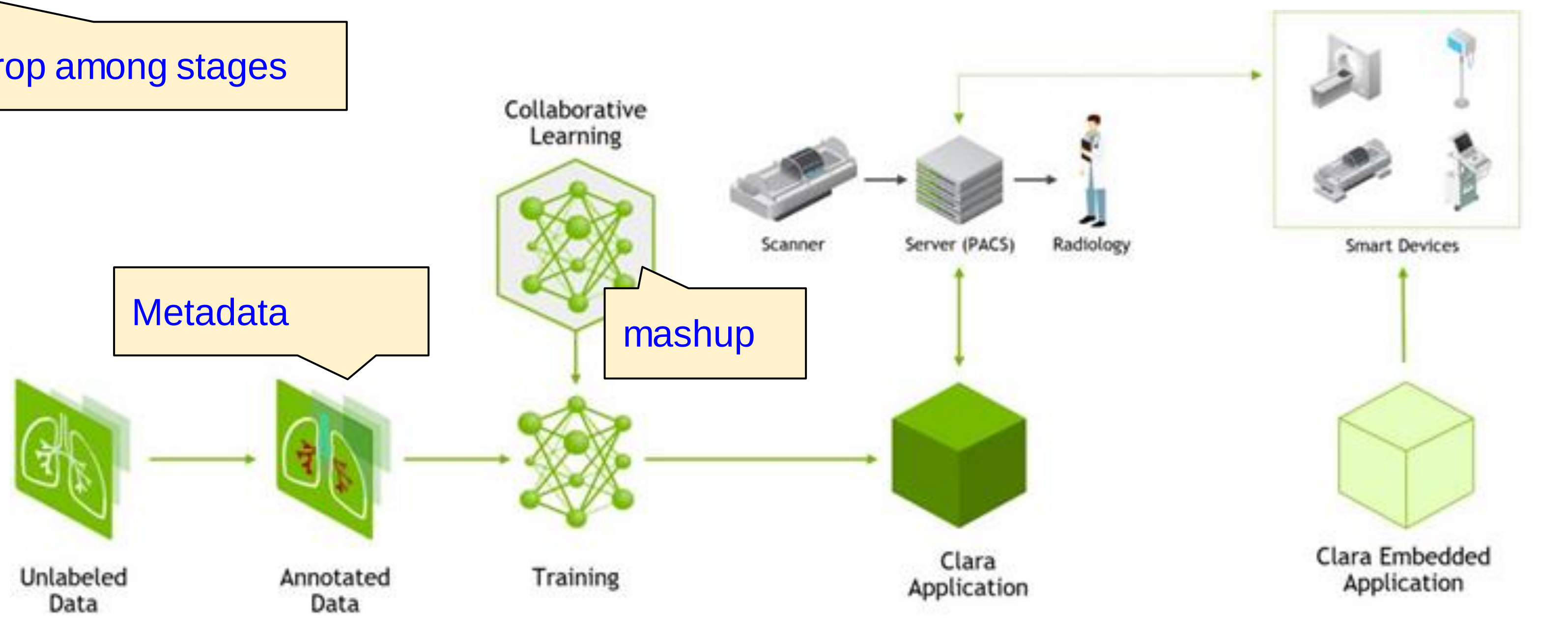
Mashup, vis, metadata



NVIDIA Clara for Healthcare

Orchestration, monitoring: EGX

Interop among stages



Metadata

mashup

DICOM Adaptor (DICOM to/from MHD)
AI Assisted Annotation (AIAA)
Data Converter (DICOM to NIfTI)

Pre-Trained Models
Transfer Learning Toolkit (TLT)
Federated Learning

Inference (TensorRT)

abstractions

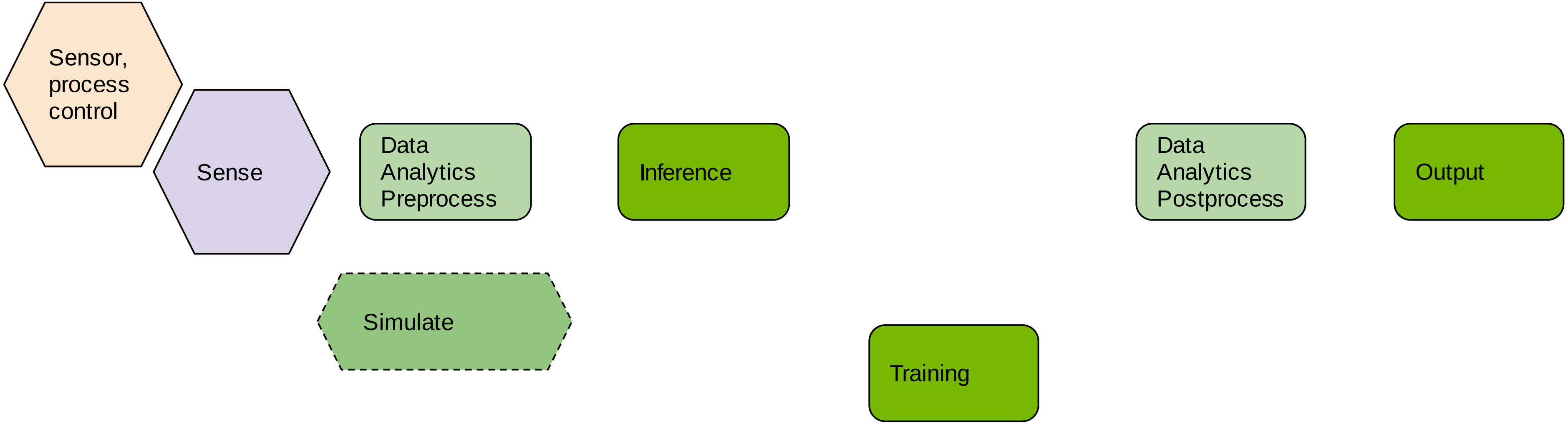
Visualization

projections

New workloads → workflows for a new era

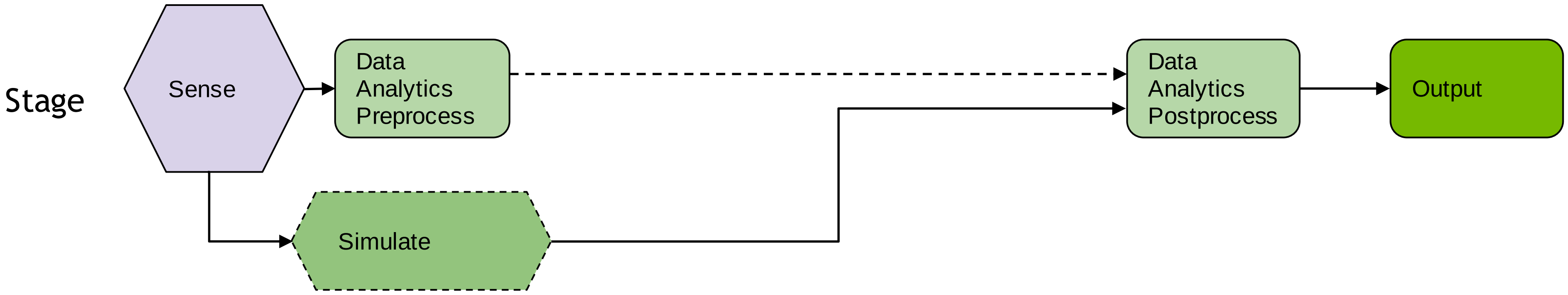
Mash-up of HPC, DL, DA, Vis in accelerated systems vs. just CPUs

Beginning to explore rich interaction among stages

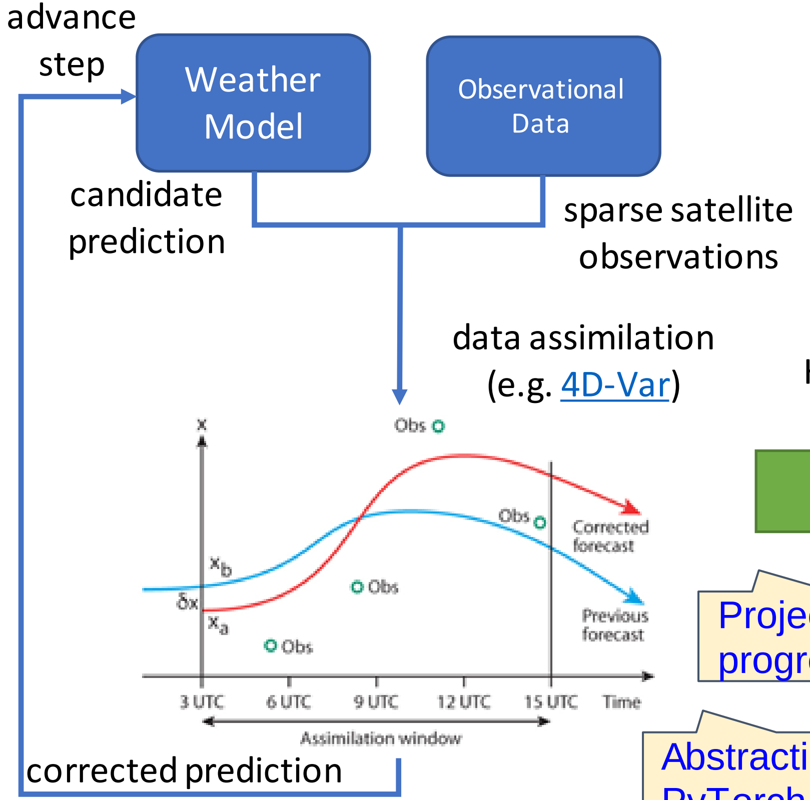


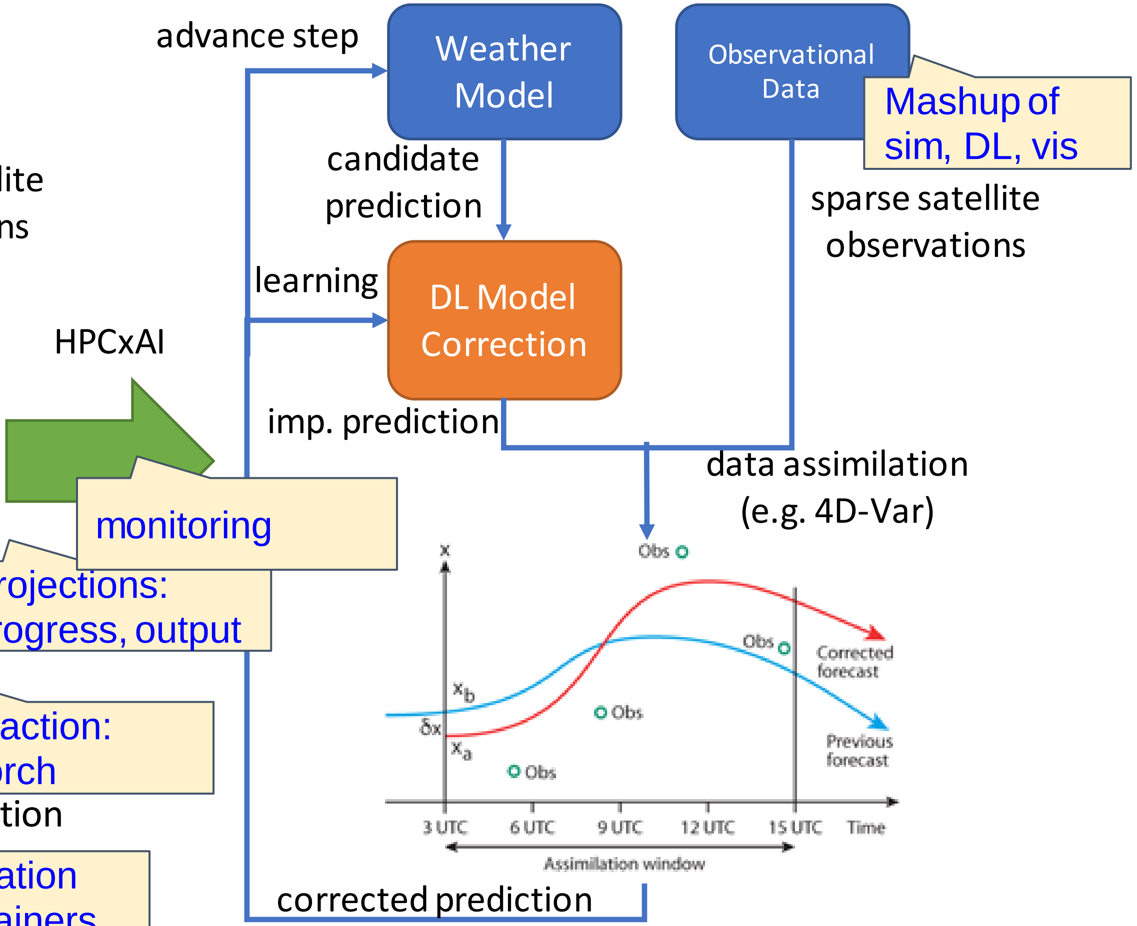
Interaction among stages: Sensors+HPC+DA+Vis

Combine sensor data with simulation, analyze, visualize



Courtesy Thorsten Kurth. Exploratory effort with ECMWF

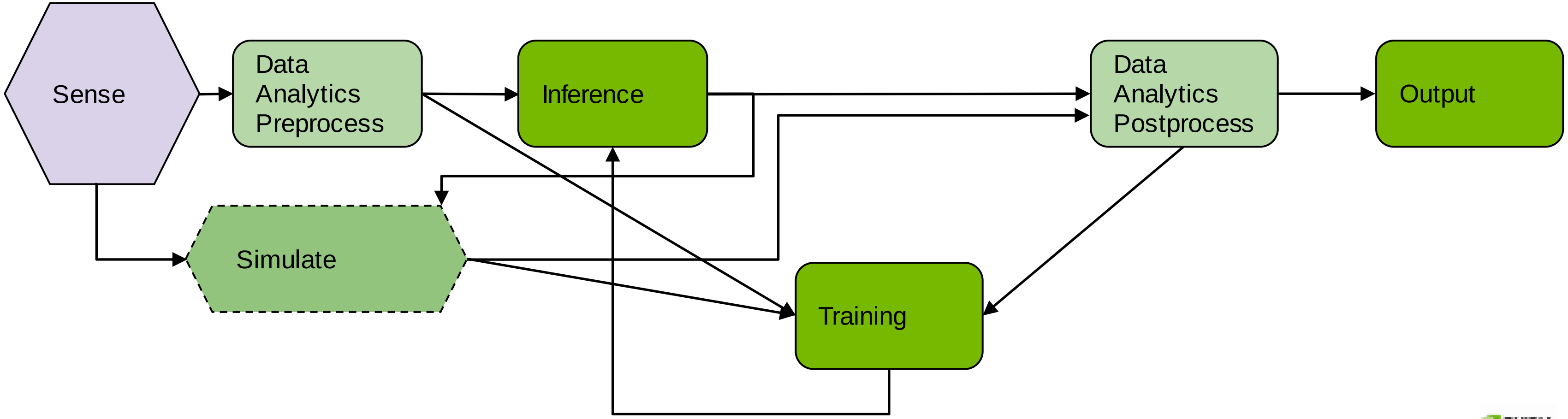




- requires tight integration of AI and simulation
- monitoring training/sim cycle for potential instabilities

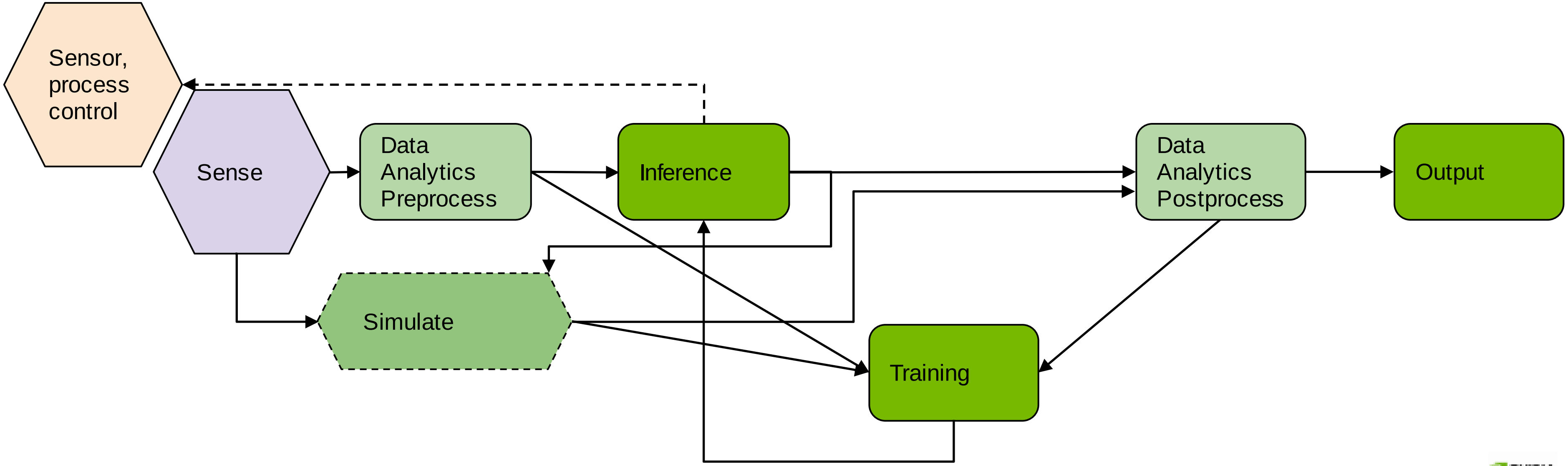
Interaction among stages: Sensors+HPC+DL/ML+DA+Vis

Augment model training with sensor data that corrects simulation



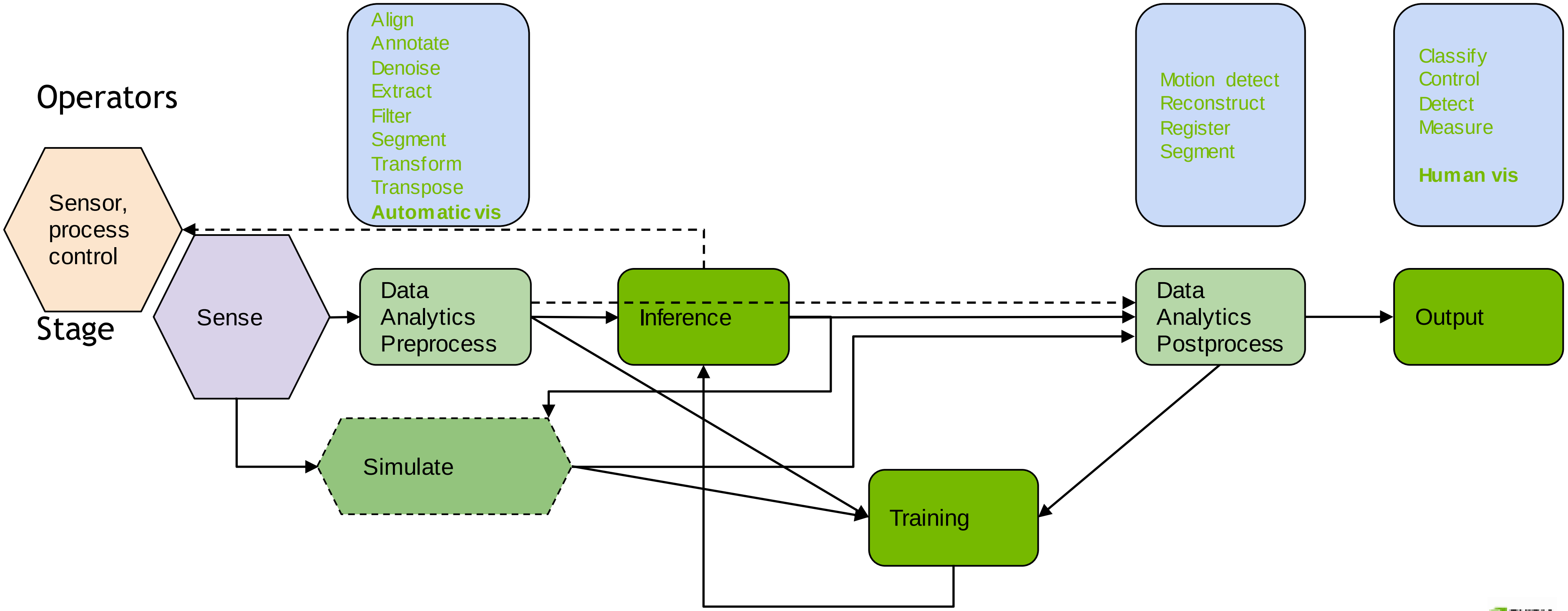
Interaction among stages: Sensors+HPC+DL/ML+DA+Vis+Ctrl

Use of model to shorten feedback loop: correct setup, increase sampling of what's unexpected



Common building blocks in various stages

Build up perf-tuned, portable/HLL building blocks to enhance productivity



Where should the action be?

Compute, storage, learning/inference

- Near the edge
  - Create higher-quality data to send to datacenter at lower bandwidth
  - Data may have locality at the edge, e.g. industrial control
  - Local learning, local activities related to federated learning
  - Inference, and an increasing fraction of training
- Fungible - either DC or edge (portable)
  - Computing platform is common between sensors and base station/datacenter
- In the base station or datacenter (if security allows)
  - Very-large-scale strong scaling, not power constrained
  - Persistent, longitudinal data
  - Complex learning, in service of transfer learning
  - Human-experimental interaction
  - Hub for federated learning

Converting numerical data into insightful projection

Lossy but discriminant compression/distillation of data that enables targeted action

- Trends
  - Sensor fusion of multi-modal sensor input, e.g. pressure, vibration, IR, temperature
  - Make data more accessible to humans or to AI networks
    - Augmented input: Fake info to support perception, e.g. color, lens flare
    - Simulation results comparisons become "observables"
- Challenges
  - Interoperability between data representations, tool diversity
  - Life-like cinema comes to science: Omniverse, Paraview, Houdini; auto-driving
- Where can you "*in situ*" it?
  - Edge - non-human agent is automated to reduce latency, effort
  - Datacenter - refined human perception: insight, education, marketing/Hollywood

Orchestration, monitoring, and resilience

- Workload: run ‘til it’s done
- Workflow: tune during the run
- Needs
  - Observability/monitoring, (health/effectiveness) analysis, vis
  - During the run, from cloud adaptively resched, re-orchestrate
  - Recover from transient and permanent failures, disconnection
  - Easy, secure, automated update
  - K8s device plugins, GPU operators
- Reference: EGX
  - “Orchestration for the Edge” at ISC20.HPCW

Performance, portability, productivity

Monitoring and continuous improvement of all of these

- Performance
  - Data movement, access, management - MagnumIO for multi-GPU, multi-node
  - Spanning nodes, processes, containers - best practices
  - Representations - numpy/cupy arrays, Arrow, USD, pyUSID, MsgPack, DICOM, NIfTI
- Portability
  - CPU and GPU versions of building blocks
  - Relocatable work: near the edge or back in the data center
- Productivity
  - Building blocks and higher-level abstractions, C++ and Python

Call to action

An invitation to collaborate as we begin a new era

- Workflows vs. workloads: mashups, interoperability, data representation, metadata, orchestration/containers/monitoring, abstractions, projections/vis

- Performance: Data movement, access, management, representation
- Portability: CPU/GPU, shift work/data/vis to where it makes most sense
- Productivity: increasing communal pool of building blocks on common platform

- Share workflows, help find commonality, build platform-level infrastructure

NVIDIA