OAK RIDGE National Laboratory

Data management strategies for large scale workflows coupling simulations to AI

Ana Gainaru

Multicore World, 16-20 February 2026

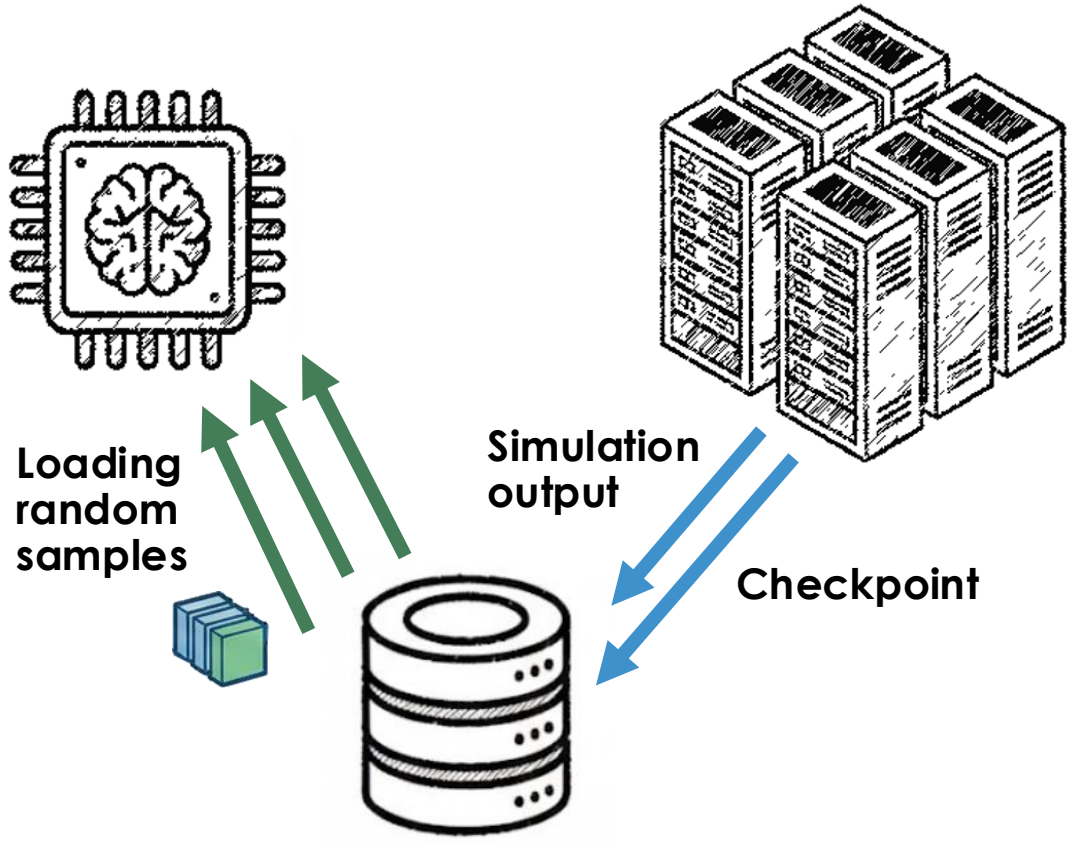
ORNL is managed by UT-Battelle LLC for the US Department of Energy

U.S. DEPARTMENT OF ENERGY

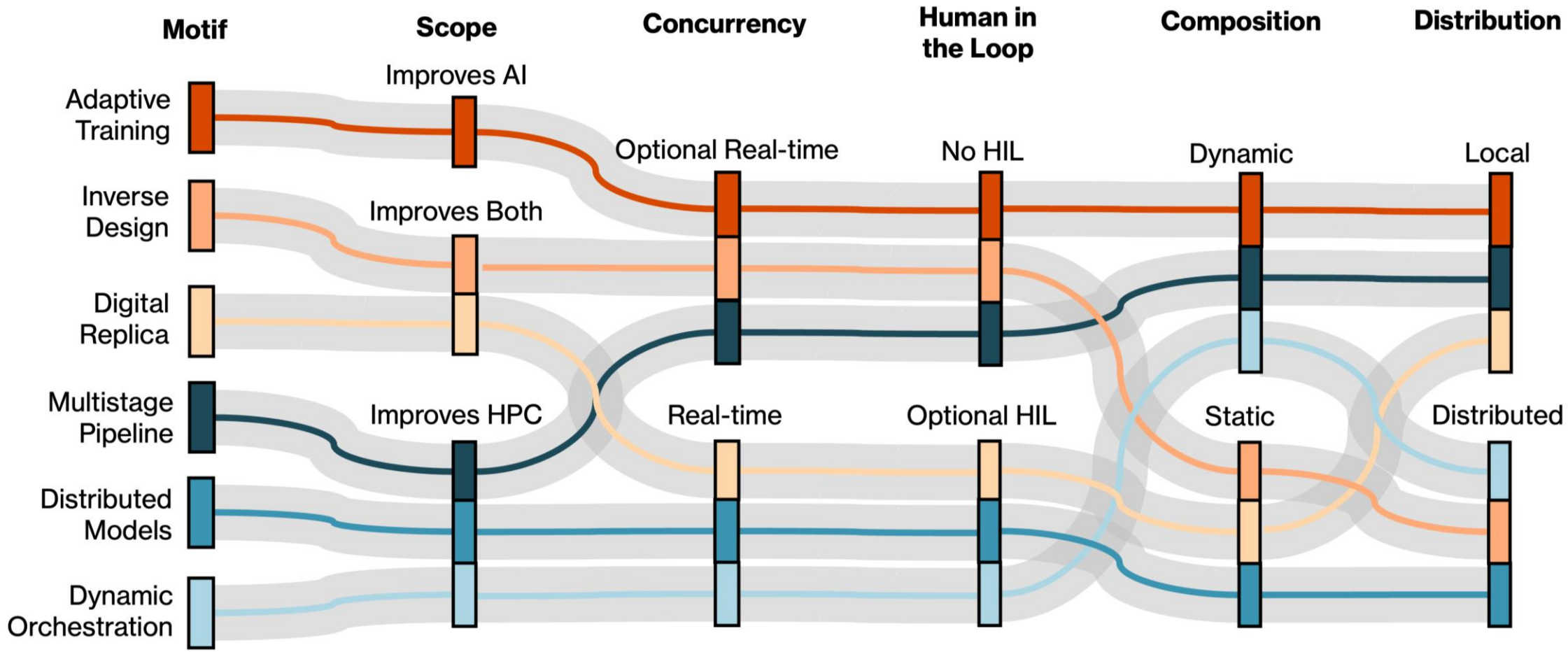
gainarua@ornl.gov

Workflows coupling simulations to AI

- **Traditional simulations**
  - Parallel I/O libraries for high throughput write
  - Feeds data to AI consumers
  - Could steer AI training based on drift detection
- **AI analysis codes**
  - Multi-threaded data loaders for high throughput read
  - Steers the HPC simulation
  - Starts ensemble simulations for training
- **Storage**
  - Store simulation data
  - Store provenance data for training
  - Store model checkpoints



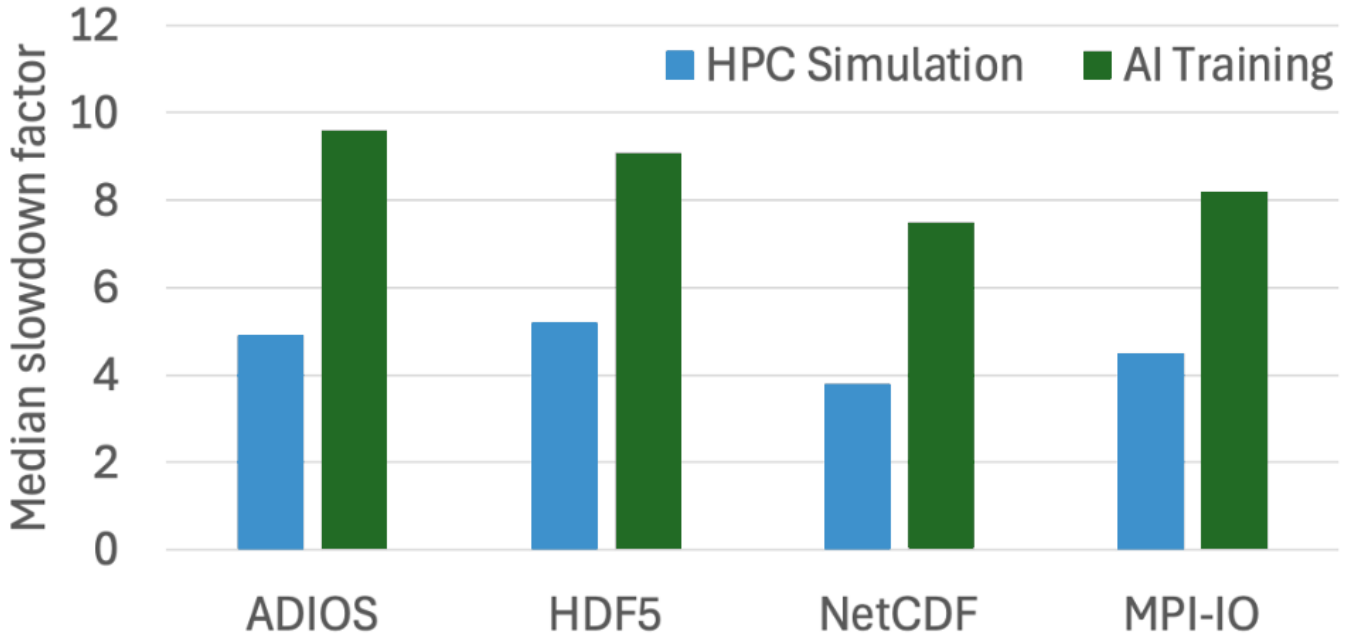
Types of workflows



OAK RIDGE National Laboratory

AI-coupled HPC Workflow Applications, Middleware and Performance, Wes Brewer, Ana Gainaru, et al, arXiv 2024

Filesystem limitations



Results for POSIX calls on the Lustre filesystem

Median slowdown factor for the I/O performance of two applications

- **Running coupled vs running sequentially**
- HPC code generating S3D combustion data (1.5TBs)
- AI code training in batches over random samples of generated data

Table of contents

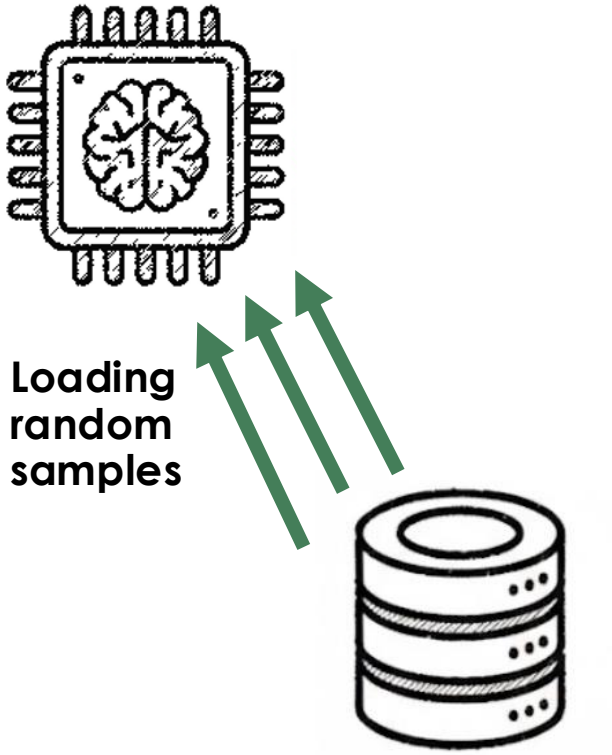
- Start from simple to complex workflows
- Data management strategies
  - Transferring data between HPC and AI
  - Finding relevant datasets for training
  - Preparing data in AI-digestible format
  - Connect to agentic frameworks

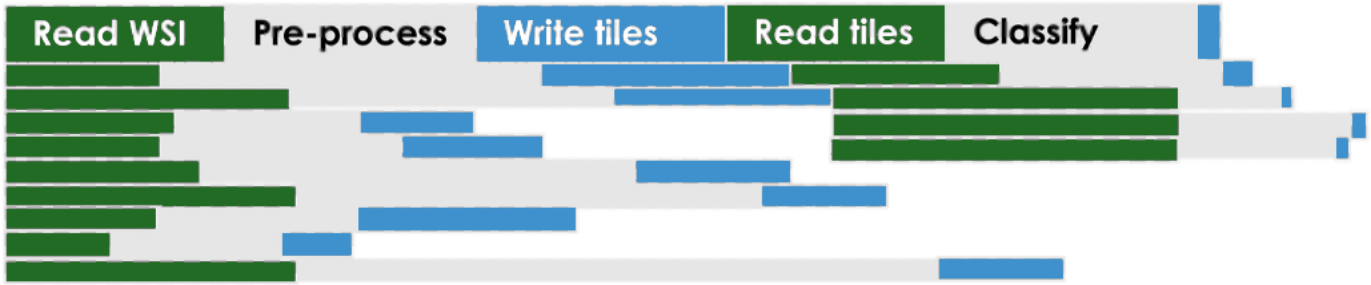
AI workflows in their simplest form

Inference for WSI segmentation

National Science Foundation

- HPC and AI are separated
  - Datasets exists on storage
    - Previous simulation runs
    - Experimental data
  - Pre-process is required
    - Mapping and normalization
    - Batch creation





OAK RIDGE National Laboratory

Understanding the Impact of Data Staging for Coupled Scientific Workflows, Gainaru et al. TPDS, 2022

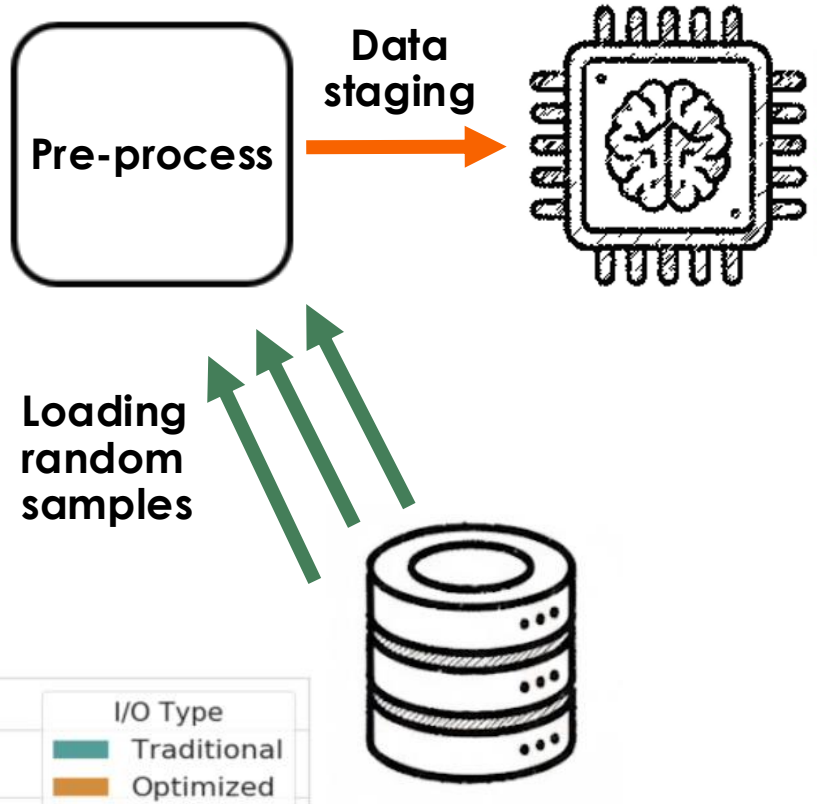
AI workflows in their simplest form

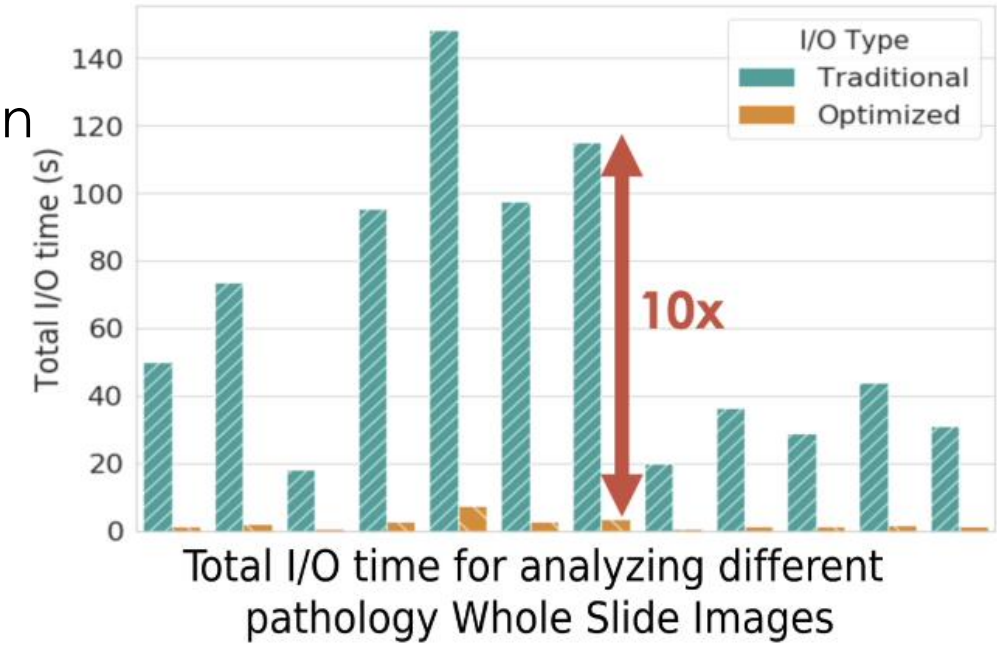
Inference for WSI segmentation

National Science Foundation

- HPC and AI are separated
  - Datasets exists on storage
    - Previous simulation runs
    - Experimental data
  - Pre-process is required
    - Mapping and normalization
    - Batch creation

Staging through memory eliminates the filesystem slowdown



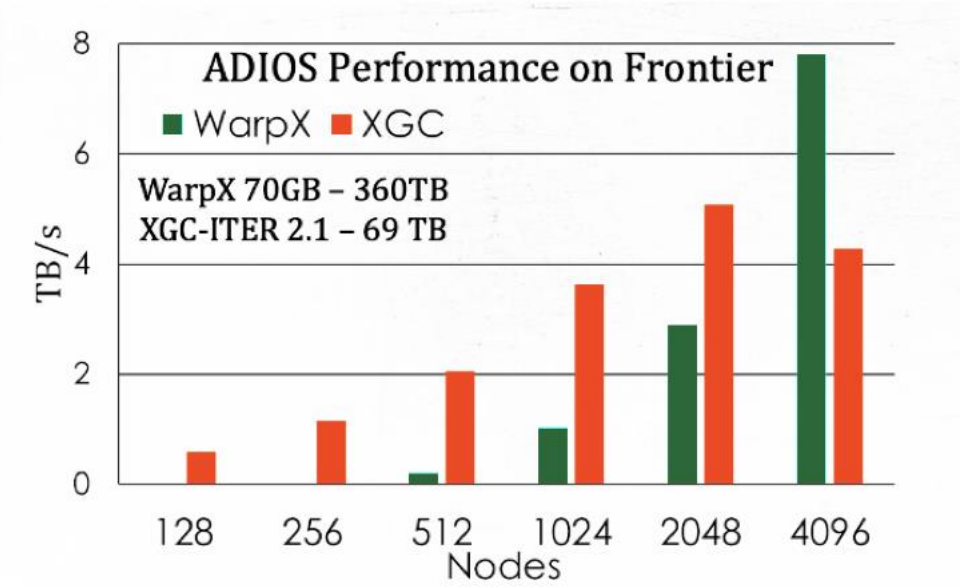


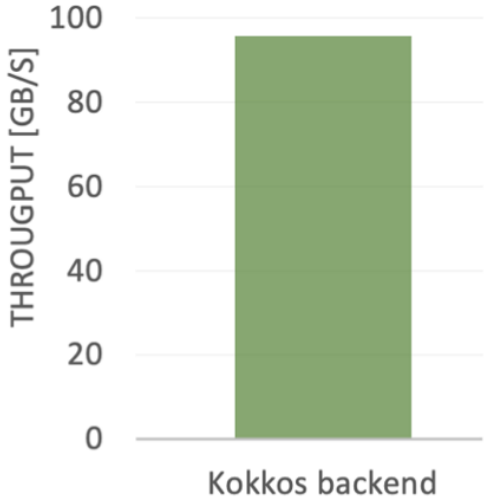
OAK RIDGE National Laboratory

Understanding the Impact of Data Staging for Coupled Scientific Workflows, Gainaru et al. TPDS, 2022

I/O Framework

- HPC parallel I/O library (ADIOS)
  - Self-described data
    - Concepts of variables and simulation steps
    - Statistics for each block of data
  - Publish/Subscribe API
  - Query engine
    - Fetch desired variables, steps and blocks
- GPU-backed to connect to AI
  - Streaming directly to pytorch tensors
  - Round robin, random access patterns



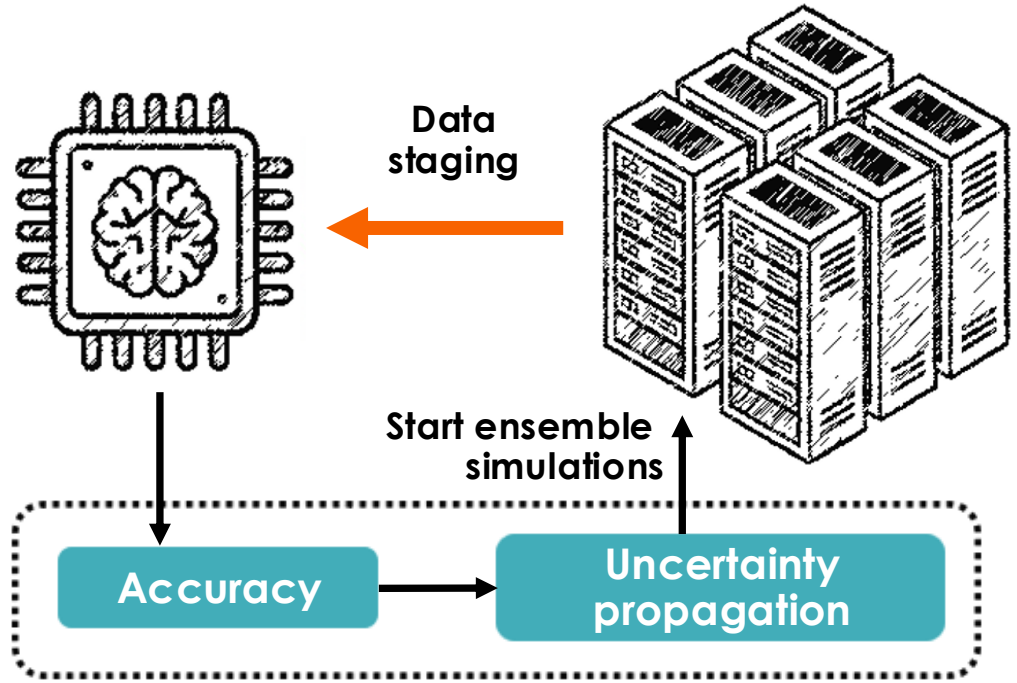


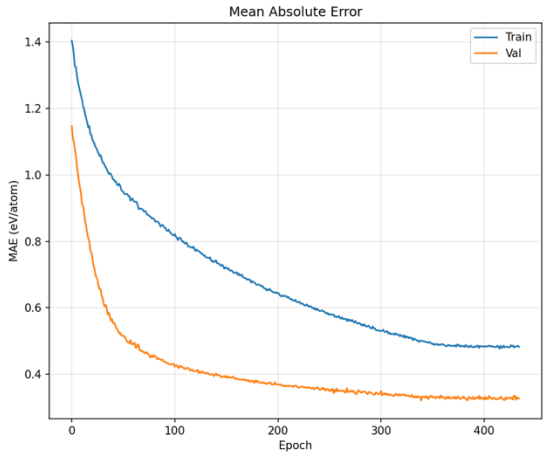
Throughput for streaming 1TB of fusion data from 900 nodes (7200 processes) to tensors in 100 nodes for random access

Active learning

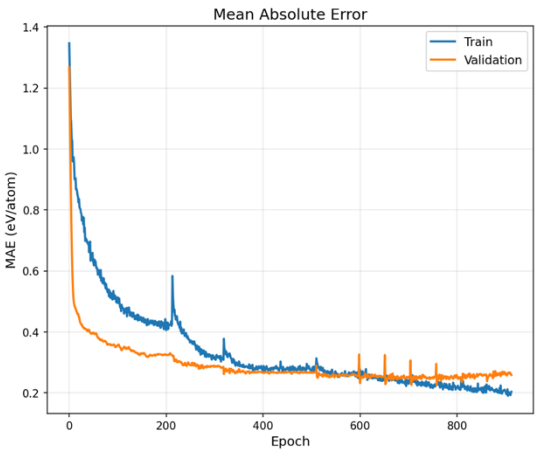
- Traditional training
  - Naive uniform sampling of the input space
  - Imbalance data coverage
    - Some areas unnecessary over-sampled while others have sparse data coverage
  - Increasing the amount of training data to eventually fill the coverage gaps
- Active learning
  - Balance exploring uncertain areas with exploited areas
  - Bayesian approach to capture uncertainty in a deep neural network

Staging through memory eliminates the filesystem slowdown





Classical Training



Active Learning

OAK RIDGE National Laboratory

Adaptive Generation of Training Data for ML Reduced Model Creation, Mark Cianciosa et al, BTSD 2022

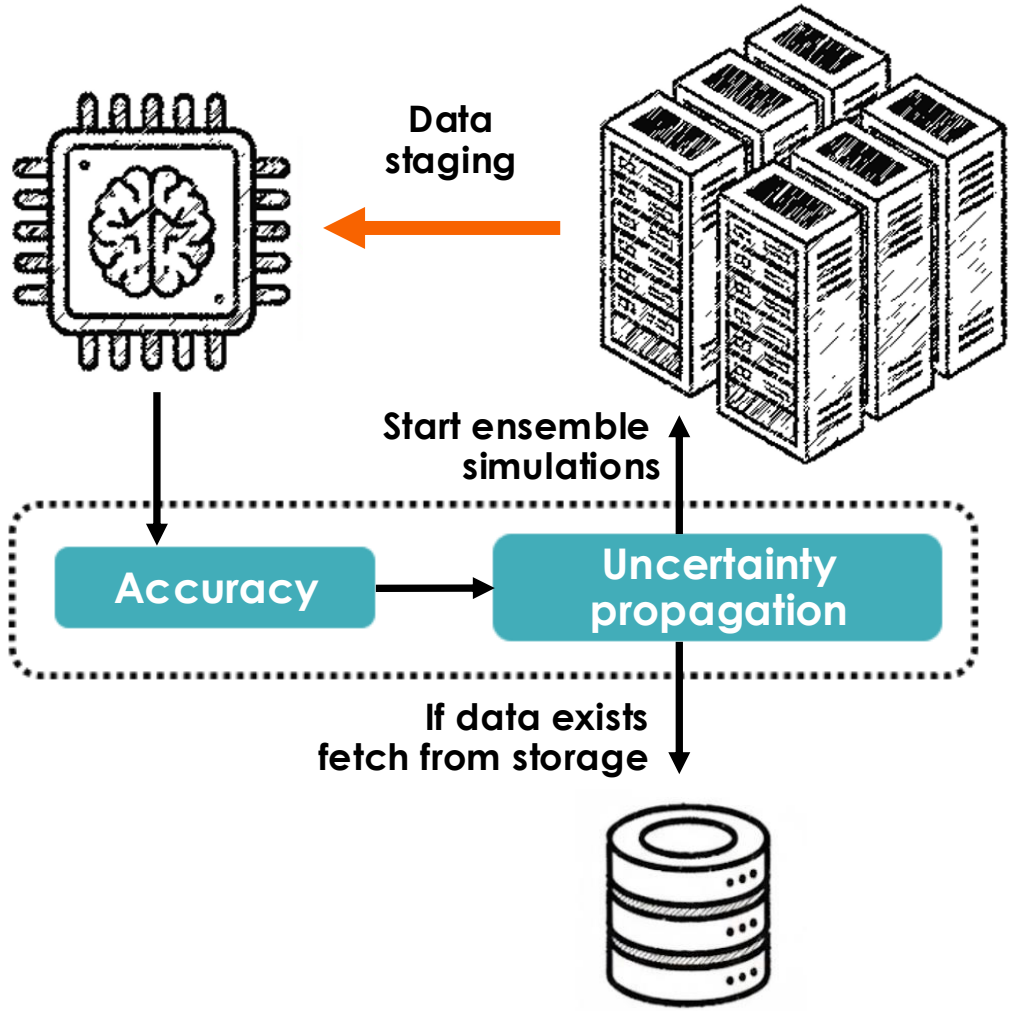
More efficient active learning

- Fetch data if it exists
  - Large scale sampling data
  - Large number of samples
  - Distributed datasets

Find relevant datasets

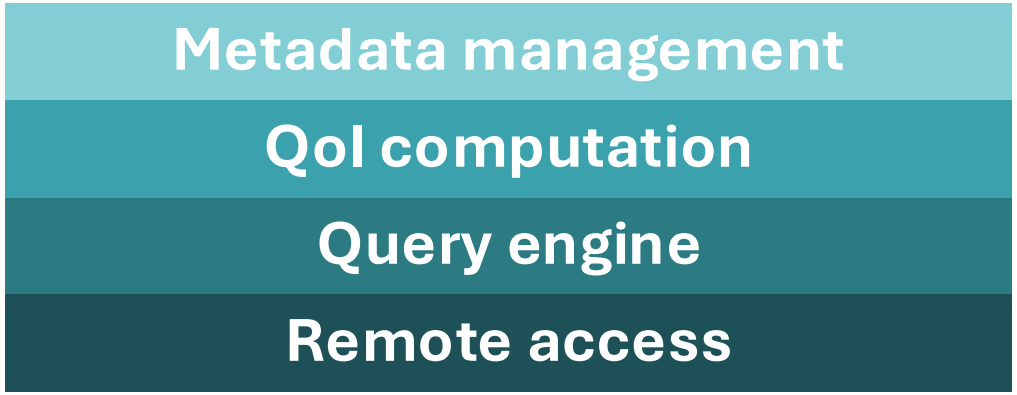
Fetch relevant data parts

Prepare the relevant data parts



Framework to manage campaigns of datasets

- How do we store data and make them known and accessible by AI and humans?
  - Assumption: Leave the data where it is (found its long-term storage) and access it on demand



Framework to manage campaigns of datasets

- How do we store data and make them known and accessible by AI and humans?
  - Assumption: Leave the data where it is (found its long-term storage) and access it on demand

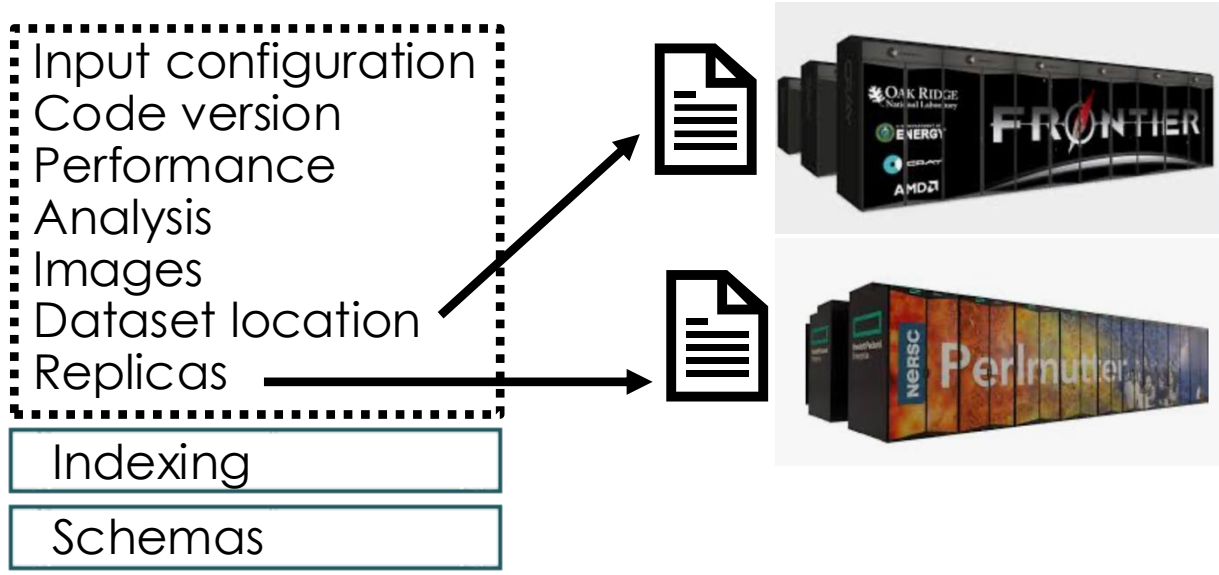
Collect and distribute metadata across multiple runs and multiple analysis

Metadata management

QoI computation

Query engine

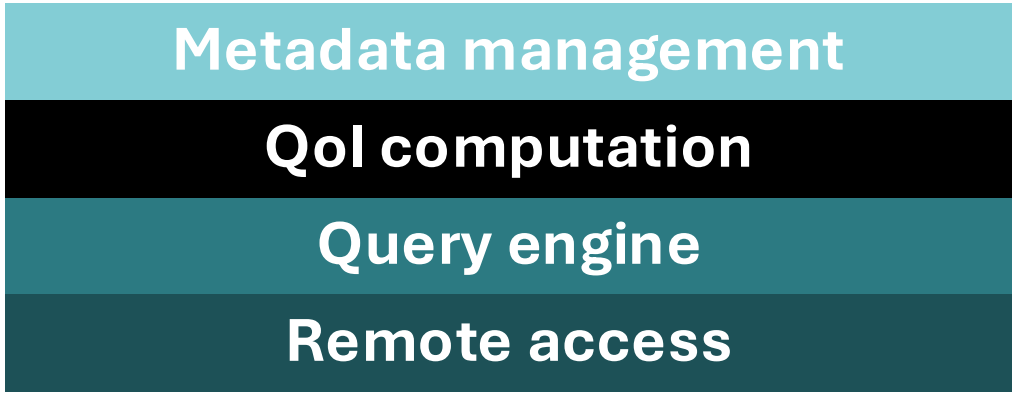
Remote access



Framework to manage campaigns of datasets

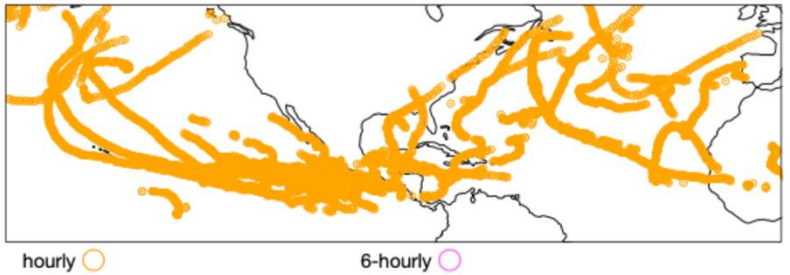
- How do we store data and make them known and accessible by AI and humans?
  - Assumption: Leave the data where it is (found its long-term storage) and access it on demand

Collect and distribute metadata across multiple runs and multiple analysis



Compute on the fly math expressions over existing datasets and store statistics

Cyclones found in 6-hourly data

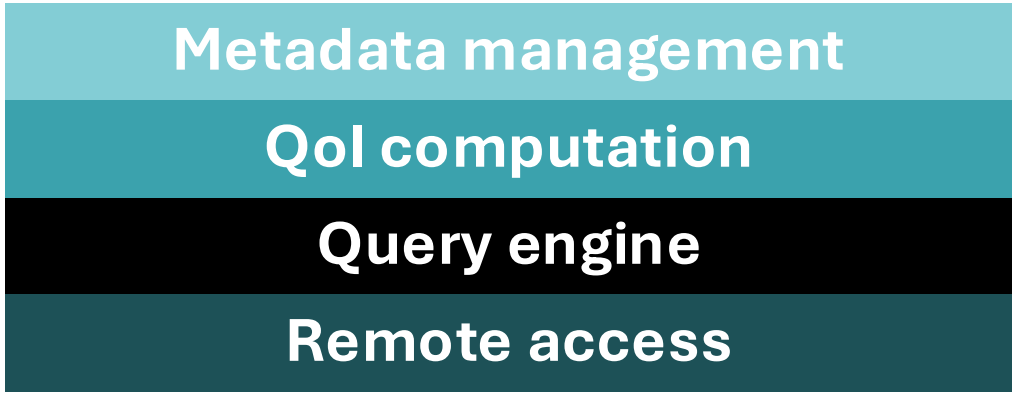


- Store derived data
  - If access of the data is frequently needed
- Store only metadata
  - For querying
- Store only the expression
  - QoI for compression

Framework to manage campaigns of datasets

- How do we store data and make them known and accessible by AI and humans?
  - Assumption: Leave the data where it is (found its long-term storage) and access it on demand

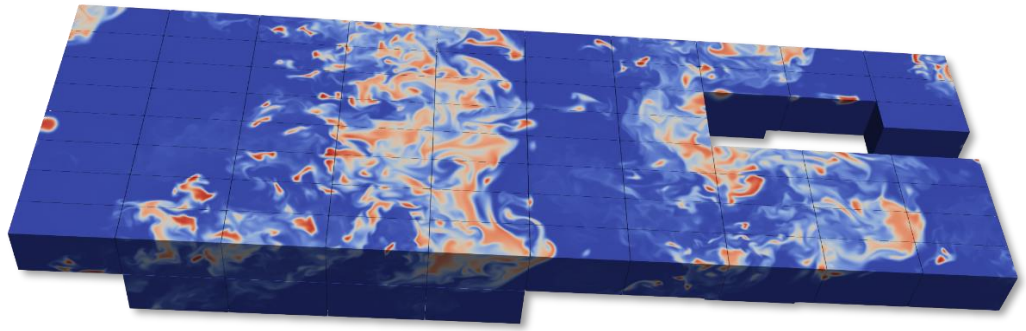
Collect and distribute metadata across multiple runs and multiple analysis



Compute on the fly math expressions over existing datasets and store statistics

Query QoI and data to restrict read to regions and objects of interest in the format needed

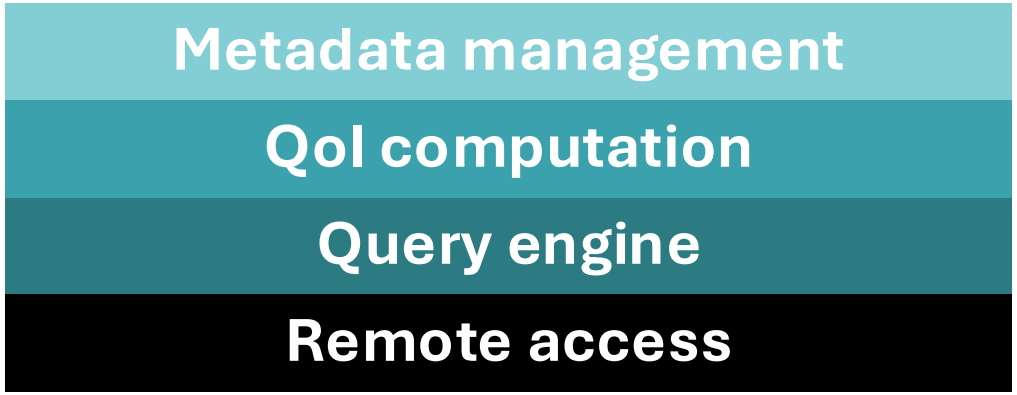
- Multi level querying engine
  - On campaign indexes
    - e.g. get all runs ran by version X of XGC using diffusion coefficients Y
  - On QoI and variables
    - E.g. get blocks where temp > 50F
  - On accuracy



Framework to manage campaigns of datasets

- How do we store data and make them known and accessible by AI and humans?
  - Assumption: Leave the data where it is (found its long-term storage) and access it on demand

Collect and distribute metadata across multiple runs and multiple analysis



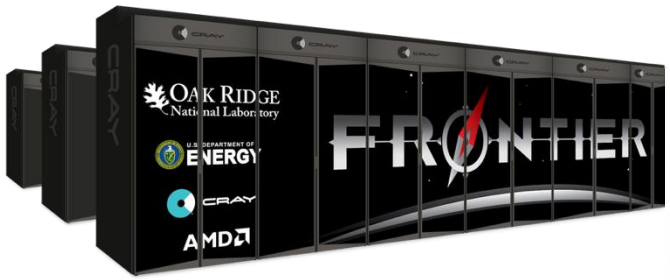
Compute on the fly math expressions over existing datasets and store statistics

Query QoI and data to restrict read to regions and objects of interest in the format needed

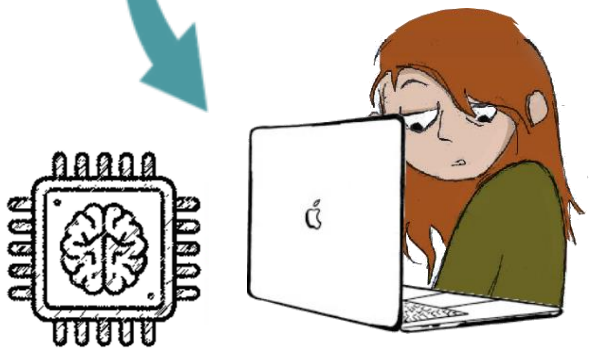
Use more advanced methods than FTP for remote access with new abstractions

- M-dimensional sub-selection of N-dimensional array
- Lossy compression to guaranteed user-specified accuracy
- Integration into existing I/O library (same code for remote and local data)
- Caching data for accelerating future reads

Demo remote access



1.5 TB of S3D data



Visualize temperature areas of intense burning

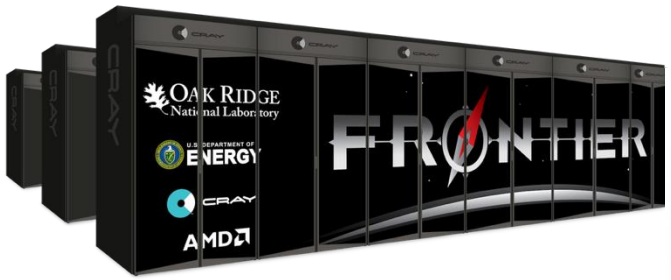
```
drwxr-sr-x 2 pnorbert csc143 99328 Feb  5 12:24 ptj.field.bp
frontier:pnorbert@login04:/lustre/orion/csc143/proj-shared/s3d/demo_single_step/run
$ du -sh ../data/ptj.field.bp/
1.5T    ../data/ptj.field.bp/
frontier:pnorbert@login04:/lustre/orion/csc143/proj-shared/s3d/demo_single_step/run
$ bpls -l ../data/ptj.field.bp/ -a
  double   derive/magnitude  {2560, 960, 3456} = 3.17757e-05 / 0.123181
  double   pressure          {2560, 960, 3456} = 14.255 / 14.3143
  double   species           {19, 2560, 960, 3456} = -3.94432e-16 / 0.787592
  string   species_names     attr   = {"H2", "NH3", "NO", "N2O", "O2", "H", "O", "OH", "HO2"
, "H2O2", "NO2", "HNO", "N", "NNH", "NH2", "NH", "H2NO", "N2"}
  double   temp              {2560, 960, 3456} = 6.16111 / 17.8856
  double   uvel              {2560, 960, 3456} = -0.105443 / 0.11861
  double   vvel              {2560, 960, 3456} = -0.0940508 / 0.0981026
  double   wvel              {2560, 960, 3456} = -0.103878 / 0.10689
frontier:pnorbert@login04:/lustre/orion/csc143/proj-shared/s3d/demo_single_step/run
$ bpls -l ../data/ptj.field.bp/ -d temp | less
frontier:pnorbert@login04:/lustre/orion/csc143/proj-shared/s3d/demo_single_step/run
$ bpls -l ../data/ptj.field.bp/ -d derive/magnitude
  double   derive/magnitude  {2560, 960, 3456} = 3.17757e-05 / 0.123181

bpls caught an exception
No data exists for this variable
```

Frontier

- 1.5TB combustion data
- Magnitude of velocity identified as QoI
  - Statistics are stored but no data

Demo remote access

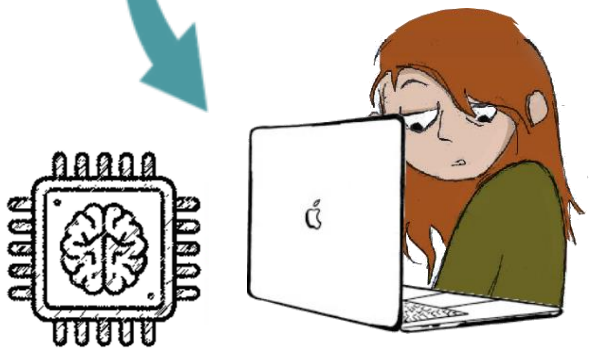


1.5 TB of S3D data

```
frontier:pnorbert@login04:/lustre/orion/csc143/proj-shared/s3d/demo_single_step/run
$ adios2_campaign_manager info  demoproject/frontier_s3d_001
info archive
ADIOS Campaign Archive, version 0.1, created on 2024-04-04 14:37:07.116121
hostname = OLCF   longhostname = frontier.olcf.ornl.gov
    dir = /lustre/orion/csc143/proj-shared/s3d/demo_single_step/run
        dataset = ../data/ptj.field.bp     created on 2024-04-03 15:25:14
frontier:pnorbert@login04:/lustre/orion/csc143/proj-shared/s3d/demo_single_step/run
$ dropbox-sync-csc143-to-cloud
```

Frontier

- Campaign file of 1.2 MB
  - Contains location of the dataset
  - Synced with laptop via Dropbox



Visualize temperature areas of intense burning

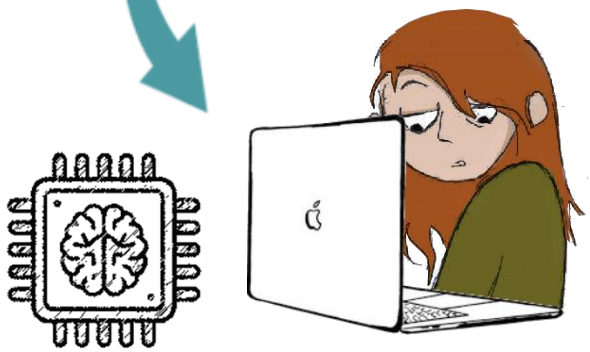
```
adios@LAP131864:~/test/demo$ adios2_campaign_manager info csc143/demoproject/frontier_s3
info archive
ADIOS Campaign Archive, version 0.1, created on 2024-04-04 14:37:07.116121
hostname = OLCF   longhostname = frontier.olcf.ornl.gov
    dir = /lustre/orion/csc143/proj-shared/s3d/demo_single_step/run
        dataset = ../data/ptj.field.bp     created on 2024-04-03 15:25:14
adios@LAP131864:~/test/demo$ bpls -la csc143/demoproject/frontier_s3d_001.aca
  double   ../data/ptj.field.bp/derive/magnitude  {2560, 960, 3456} = 3.17757e-05 / 0.12
  double   ../data/ptj.field.bp/pressure          {2560, 960, 3456} = 14.255 / 14.3143
  double   ../data/ptj.field.bp/species           {19, 2560, 960, 3456} = -3.94432e-16 /
  string   ../data/ptj.field.bp/species_names     attr   = {"H2", "NH3", "NO", "N2O", "O
"O", "OH", "HO2", "H2O", "H2O2", "NO2", "HNO", "N", "NNH", "NH2", "NH", "H2NO", "N2"}
  double   ../data/ptj.field.bp/temp              {2560, 960, 3456} = 6.16111 / 17.8856
  double   ../data/ptj.field.bp/uvel              {2560, 960, 3456} = -0.105443 / 0.1186
  double   ../data/ptj.field.bp/vvel              {2560, 960, 3456} = -0.0940508 / 0.098
  double   ../data/ptj.field.bp/wvel              {2560, 960, 3456} = -0.103878 / 0.1068
adios@LAP131864:~/test/demo$
```

Local laptop

- List of variables as all metadata as if data is local

Demo remote access

1.5 TB of S3D data



Visualize temperature areas of intense burning

```
file1 = open(queryFile, "w")
queryContent = [
    '<?xml version="1.0"?>\n',
    "<adios-query>\n",
    '  <io name="query">\n' '   <var name="' + queryVarName + '">\n
    '    <op value="AND">\n',
    '      <range  compare="LT" value="' + str(maxval) + '"/>\n',
    '      <range  compare="GT" value="' + str(minval) + '"/>\n',
    "    </op>\n",
    '  <boundingbox start="' + start3D + '" count="' + count3D + '
```

Local laptop

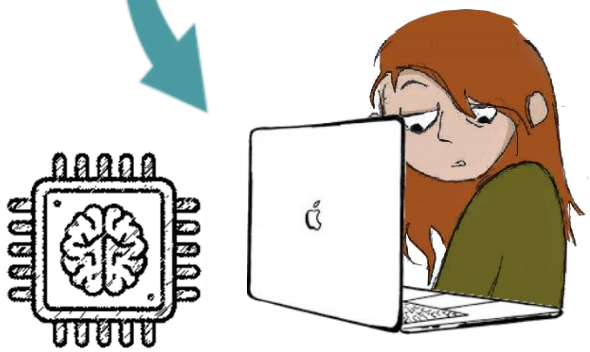
- Define query
  - In exactly the same way it would be defined for local data
  - Get temperature where magnitude < 0.7 with 0.01 accuracy

```
adios@LAP131864:~/test/demo.sc24$ time python3 plot/plot2D.py -i csc143/demop
ier_s3d_001.aca -v ../data/ptj.field.bp/temp -plane xz -o temp -e 0.1
Variable min=6.16111 max=17
GS Plot step 0 processing si                                    tation step 0
```

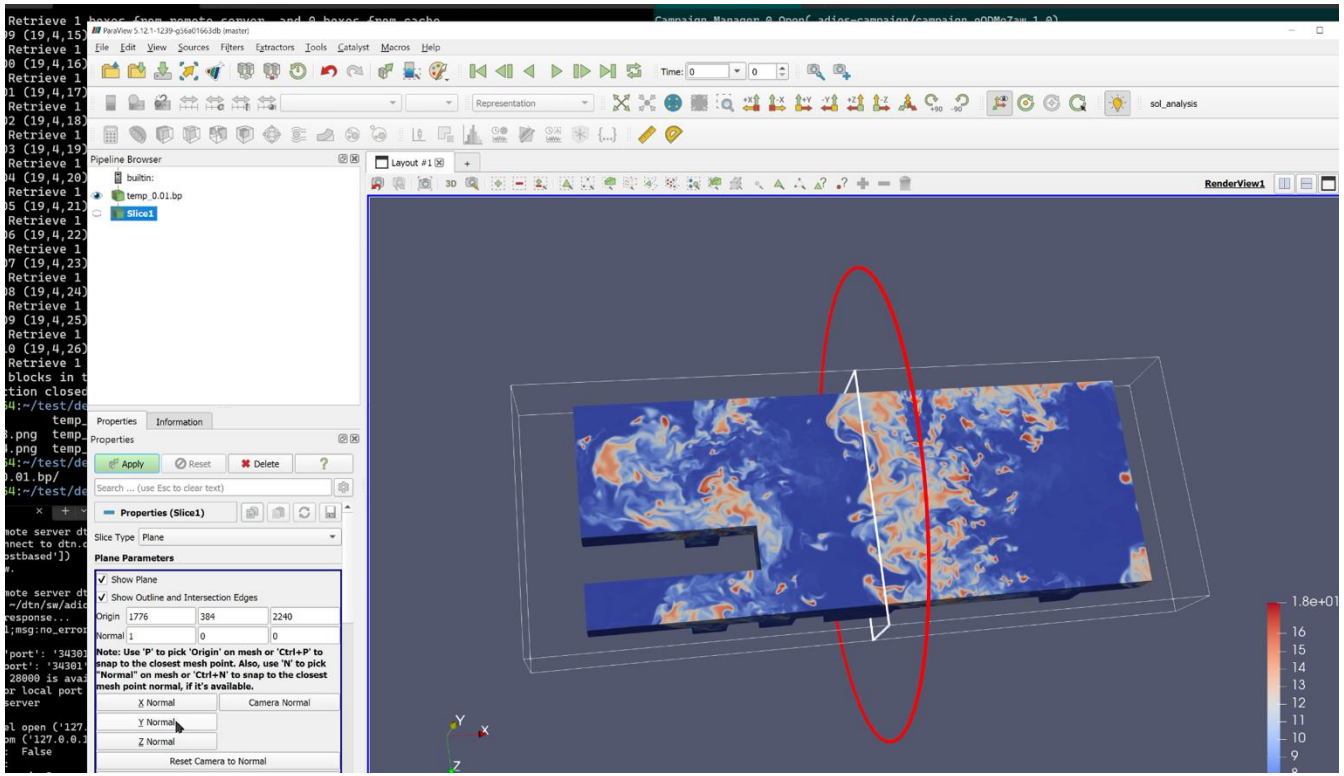
- Download blocks
  - 139 blocks
  - 13.8 seconds
  - 1.3 GB total data

Demo remote access

1.5 TB of S3D data



Visualize temperature areas of intense burning



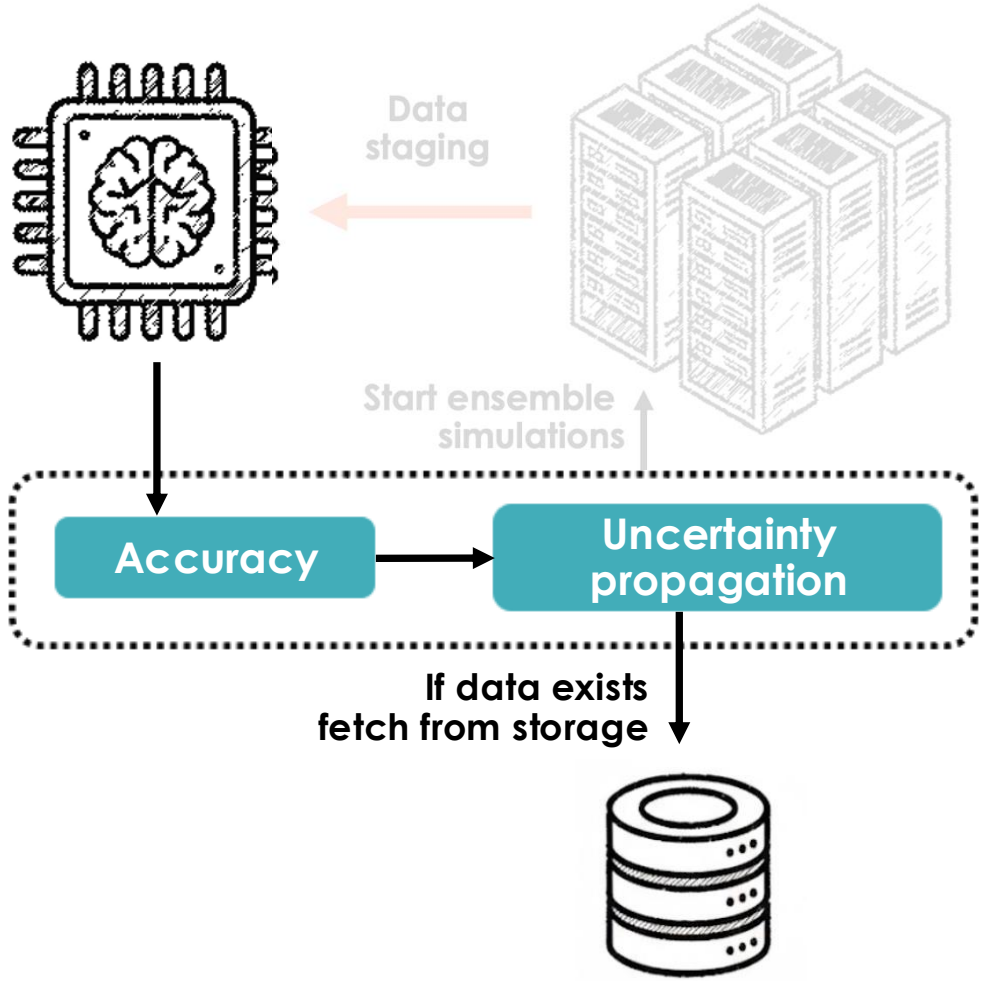
Local laptop

- Visualize data in Paraview or any other software

OAK RIDGE National Laboratory

HPC Campaigns for AI

- Single campaign
  - Campaign lp700_p3.75_d2.0.aca
  - Number of runs: 251
  - Number of variables: 12550
  - Read all variables
    - On Perlmutter directly from files: 9.2s
    - On Perlmutter from portals.nersc.gov: 32s
    - From my computer at home from portals.nersc.gov (TN): 47s
- All campaigns
  - Read all campaigns to create a 2D table for training the DivControlNN surrogate model
    - On Perlmutter: 42-83 minutes depending on file system responsivity

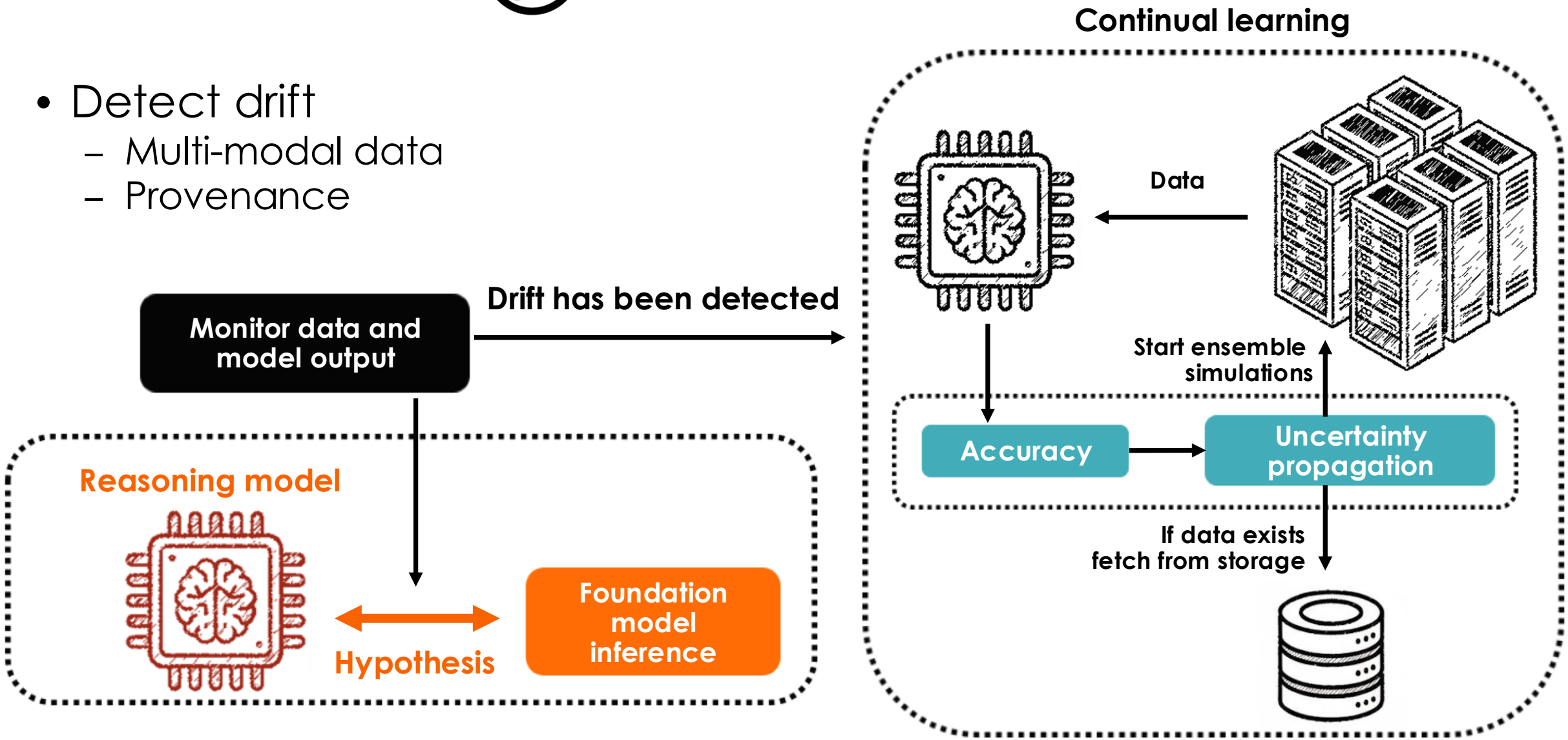


UEDGE KSTAR24 data

ModCon project

Genesis Mission

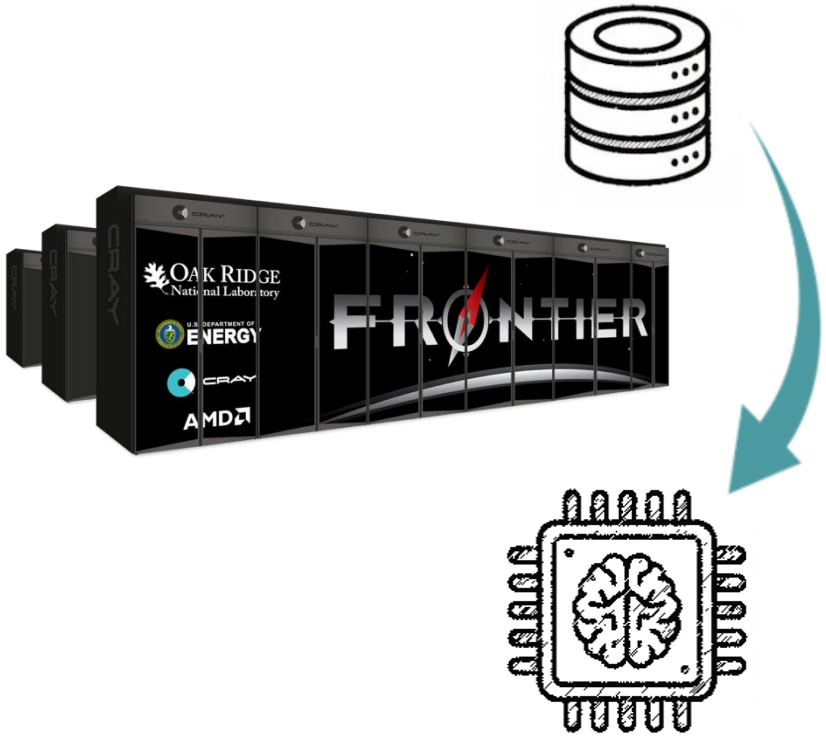
- Detect drift
  - Multi-modal data
  - Provenance



OAK RIDGE National Laboratory

Demo agentic framework

1.2 GB based on over 500 DFT simulations and 2 million downloaded entries



Reproduce the findings of a fission paper

Journal of Nuclear Materials 510 (2018) 331–336

Contents lists available at ScienceDirect

Journal of Nuclear Materials

journal homepage: www.elsevier.com/locate/jnucmat

Stability of U_5Si_4 phase in U-Si system: Crystal structure prediction and phonon properties using first-principles calculations

D.A. Lopes*, V. Kocevski, T.L. Wilson, E.E. Moore, T.M. Besmann

Nuclear Engineering Program, Department of Mechanical Engineering, University of South Carolina, Columbia, SC, 29209, USA

ARTICLE INFO

Article history:
Received 30 May 2018
Received in revised form
6 August 2018
Accepted 13 August 2018
Available online 14 August 2018

ABSTRACT

U-Si systems have recently received considerable attention due to the potential application of U_3Si_2 as a high-density fuel under an accident tolerant fuel initiative. However, the thermodynamic stability of the more recently reported adjacent U_5Si_4 phase is uncertain and could play a significant role in fuel performance. In this work, the enthalpy of formation of the phase predicted by density functional theory (DFT) using the DFT + *U* formalism is used with an evolutionary algorithm (USPEX) to evaluate stability and possible atomic structures for U_5Si_4. The structure of U-Si convex hull phases and the confirmed U_3Si_2 structure were predicted providing confidence in the reliability of the evolutionary algorithm, as well as to obtain a convex hull with comparable enthalpies of formation. Subsequently, the code was 18 and 36 atoms, predicting a 36-atom hexagonal symis the lowest-energy configuration, agreeing with that calculations using the density functional perturbation ed structure is dynamically unstable, exhibiting negative ated that generated shears are directed toward the foralogous to those occupied by carbon atoms in $U_{20}Si_{16}C_3$. phase from assessed U-Si phase equilibria is currently

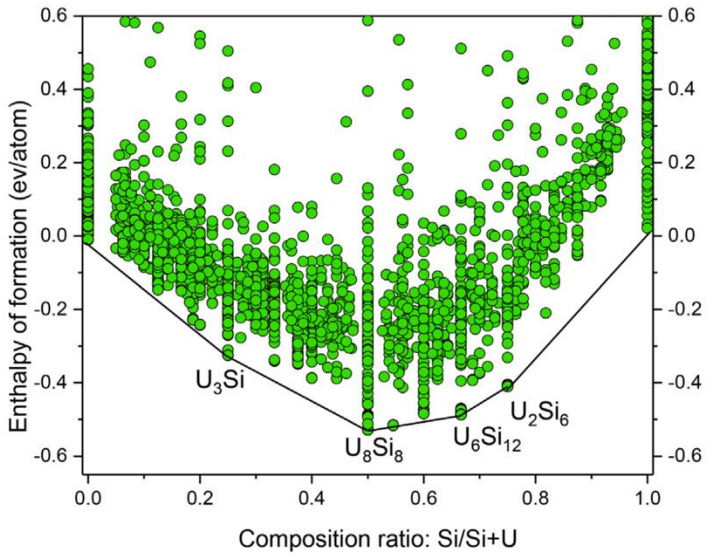


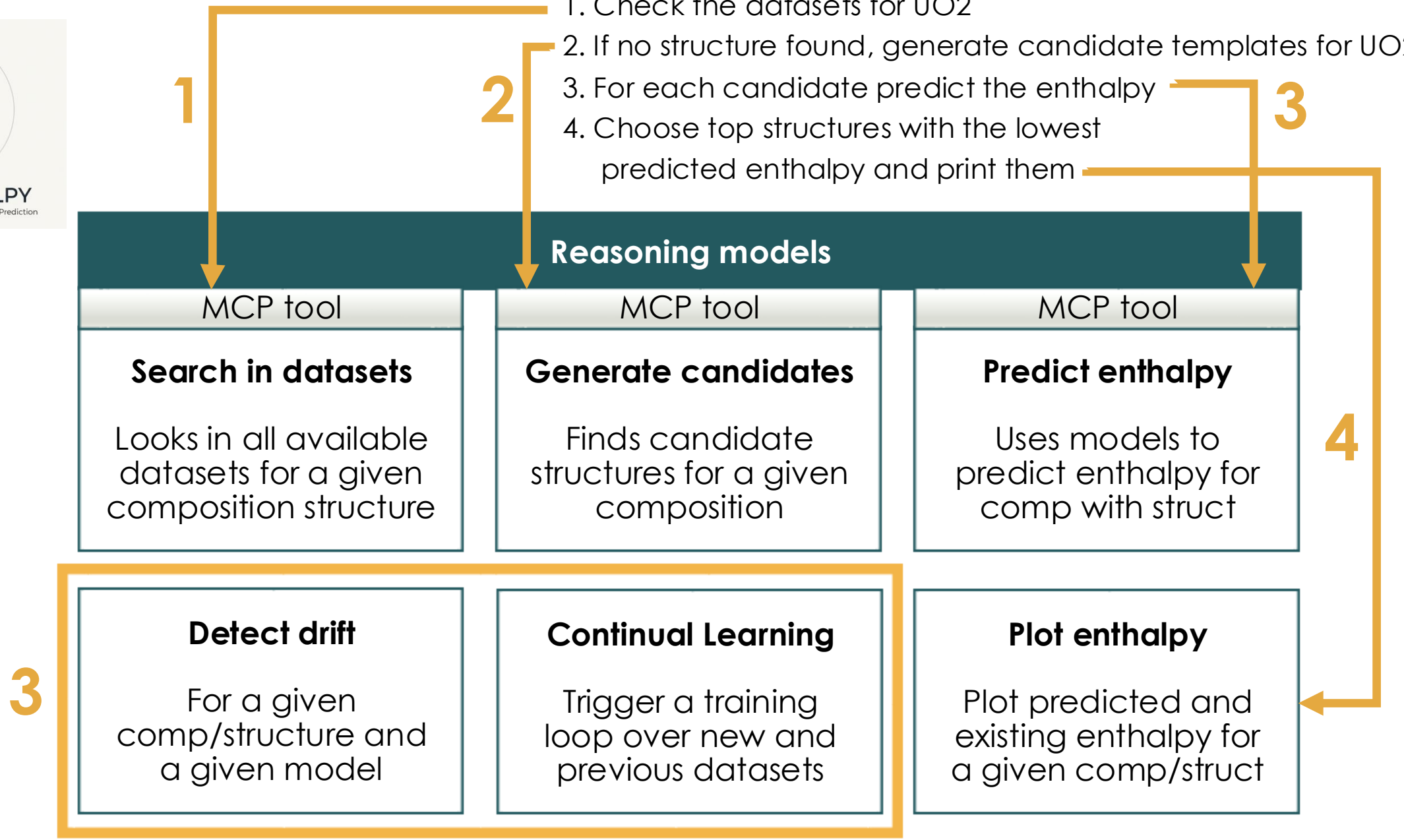
Fig. 2. Convex hull for the U-Si phase space obtained from variable-composition USPEX calculations mode in (enthalpies relative to Si-diamond and α-U, with U_{eff} = 1.5eV), for 2500 structure optimizations.

MCP tools

AERIS ENTHALPY
Reasoning Models for Material Enthalpy Prediction

> Find structure candidates for UO2 that could be stable

1. Check the datasets for UO2
2. If no structure found, generate candidate templates for UO2
3. For each candidate predict the enthalpy
4. Choose top structures with the lowest predicted enthalpy and print them



OAK RIDGE National Laboratory

Discovering new materials

- Original study
  - 6 month
  - 500+ DFT simulations ran
  - Over 10k node hours
- Reasoning model
  - One week
  - Drift detected <10 times
  - Training 1 node hour
  - Total < 1k node hours

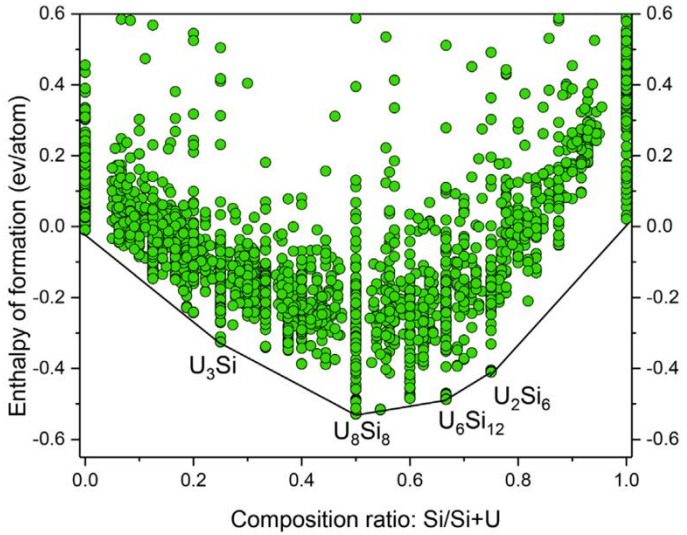
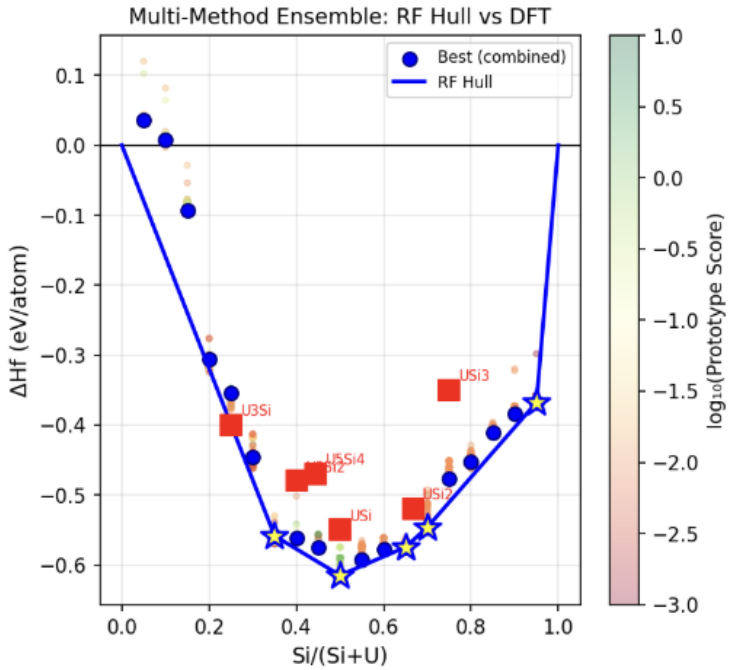
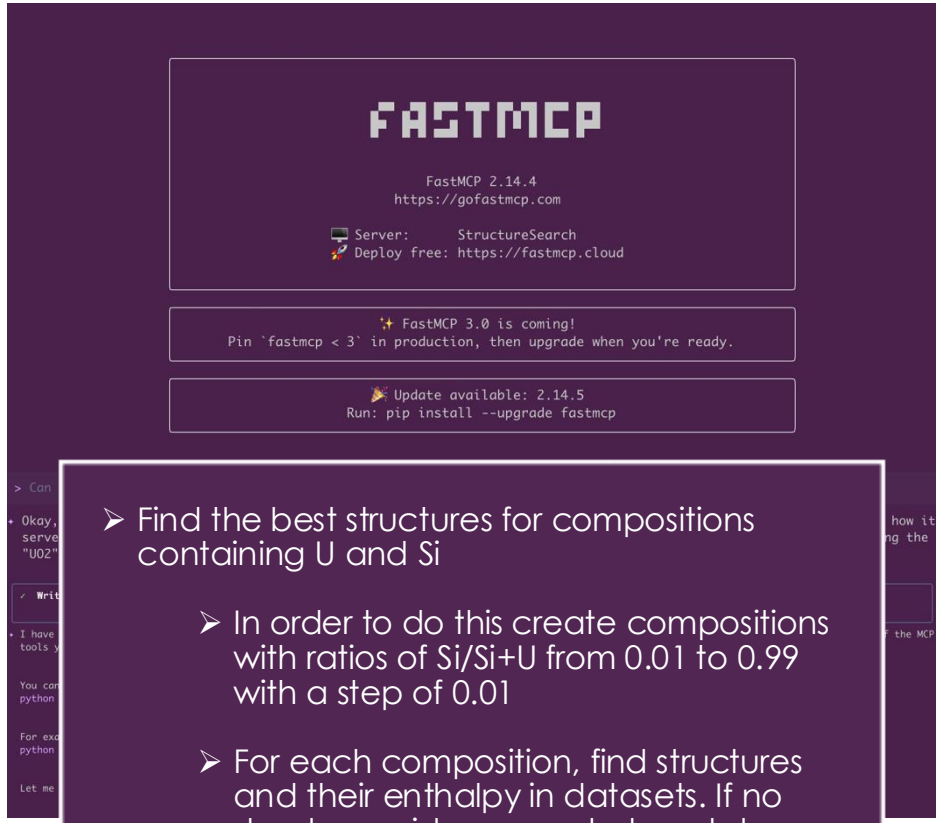


Fig. 2. Convex hull for the U-Si phase space obtained from variable-composition USPEX calculations mode in (enthalpies relative to Si-diamond and α-U, with U_{eff} = 1.5eV), for 2500 structure optimizations.





- Find the best structures for compositions containing U and Si
  - In order to do this create compositions with ratios of Si/Si+U from 0.01 to 0.99 with a step of 0.01
  - For each composition, find structures and their enthalpy in datasets. If no structure exists, generate template structures and predict the enthalpy for each template
  - Return the 100 entries with lowest enthalpy per atom
  - Use available tools to find structures, generate templates and predict enthalpy

Conclusion

Thank you !
gainarua@ornl.gov

Useful

- Co-design between CS and domain scientists
  - Provenance tracking to detect successes and fails

Usable

- Reasoning models
  - Documentation
- Metadata management
  - Schemas
- Federated access

Composable

- Plug and play solutions are needed
- MCP tools
- Dynamic behavior
  - Performance monitoring

Trustable

- Provenance tracking for post-mortem understanding
- Test and evaluation
  - Understand the drift between data and models