Argonne NATIONAL LABORATORY

ARGONNE LEADERSHIP COMPUTING FACILITY

U.S. DEPARTMENT OF ENERGY

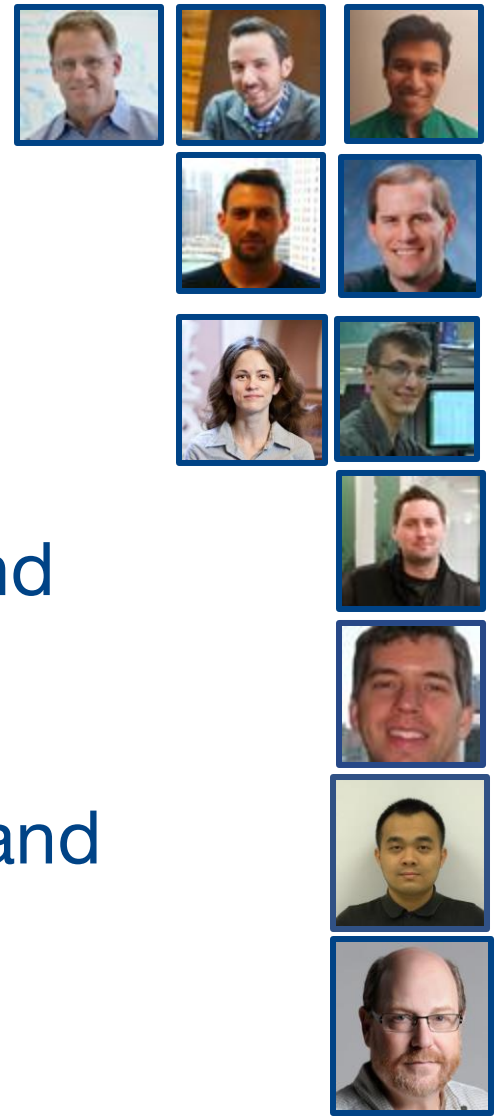
DLHub

Data and Learning Hub for Science

Ben Blaiszik (bblaiszik@anl.gov), Ryan Chard, Logan Ward,
Kyle Chard, Zhuozhao Li, Anna Woodard, Yadu Babuji,
Steve Tuecke, Mike Franklin, Ian Foster

Funding: 2018 Argonne Advanced Computing LDRD

Data and Learning Hub for Science (DLHub)



- Collect, publish, categorize models from many disciplines (materials science, physics, chemistry, genomics, etc.)
- Serve models via API to simplify sharing, consumption, and access
- Enable new science through reuse, real-time integration, and synthesis of existing models

Funding: 2018 Argonne Adv. Computing LDRD

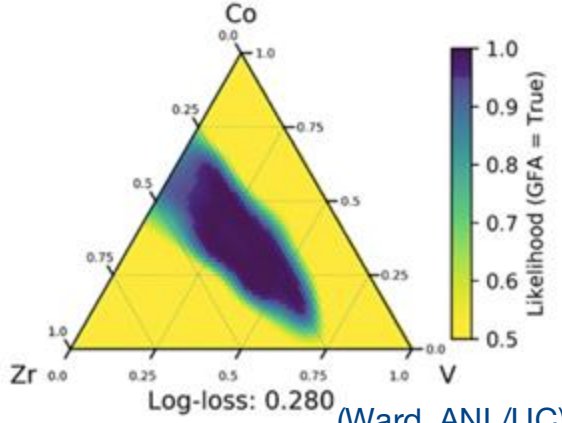
Argonne NATIONAL LABORATORY

ARGONNE LEADERSHIP COMPUTING FACILITY PETREL parsl

Select DLHub Use Cases

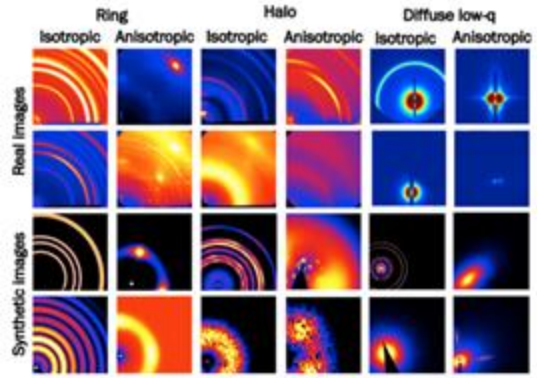
Model-driven Experimentation and Data Tagging

- Metallic glass discovery [active learning]



(Ward, ANL/UC)

- XRD beamline image tagging

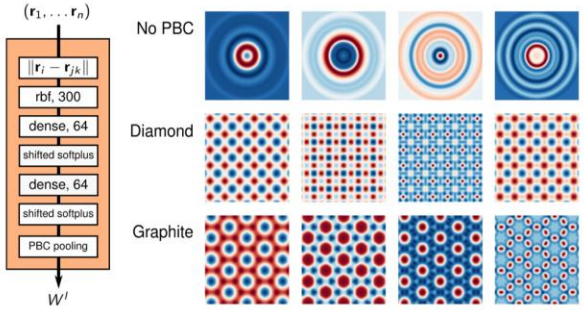


	probability	tags
1	0.999780	Beam Off Image
6	0.745707	Halo
13	0.364863	Strong scattering
7	0.082574	High background
12	0.027826	Ring
8	0.012137	Higher orders
15	0.008106	Weak scattering
9	0.001964	Linear beamstop
16	0.001744	Wedge beamstop
2	0.000982	Circular Beamstop

(Yager, BNL)

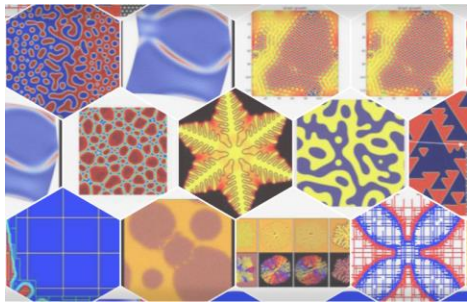
Community Model Benchmarking

- Crystal structure



(Ward, ANL/UC)

- NIST PFHub



(Wheeler, Warren, Heinonen NIST/UC/Argonne/NU)

Automated Model Retraining with New Data

- Models linked to dynamic data sources

(Center for Hierarchical Materials Design NIST/UC/Argonne/NU)

Funding: 2018 Argonne Adv. Computing LDRD

Argonne NATIONAL LABORATORY

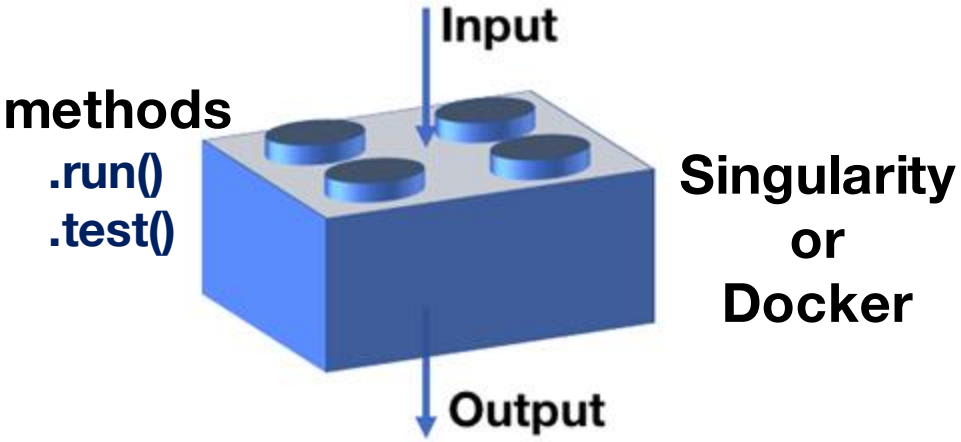
ARGONNE LEADERSHIP COMPUTING FACILITY

PETREL

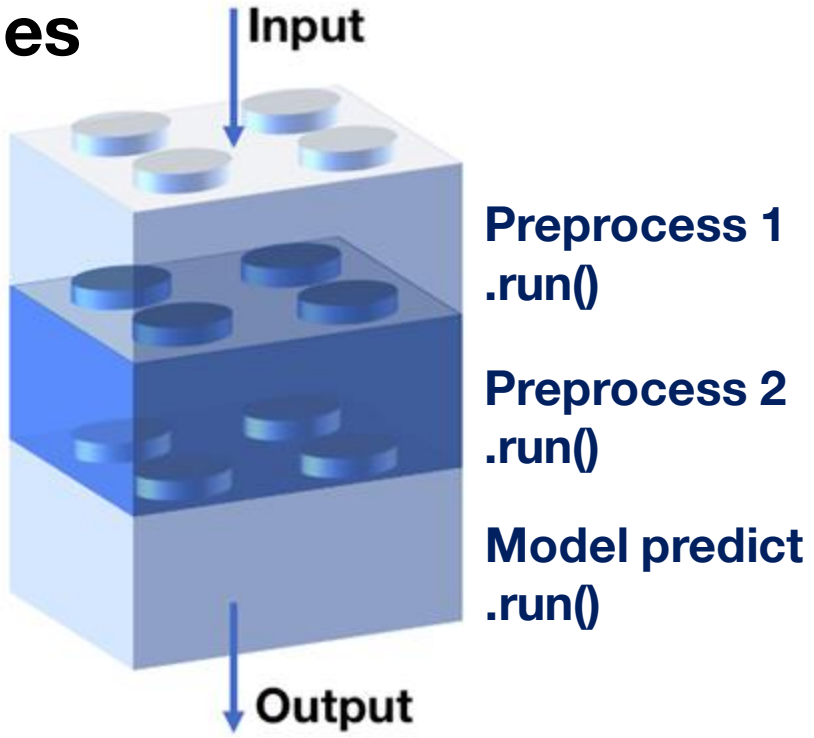
parsl

DLHub Servables and Pipelines

Servables



Pipelines



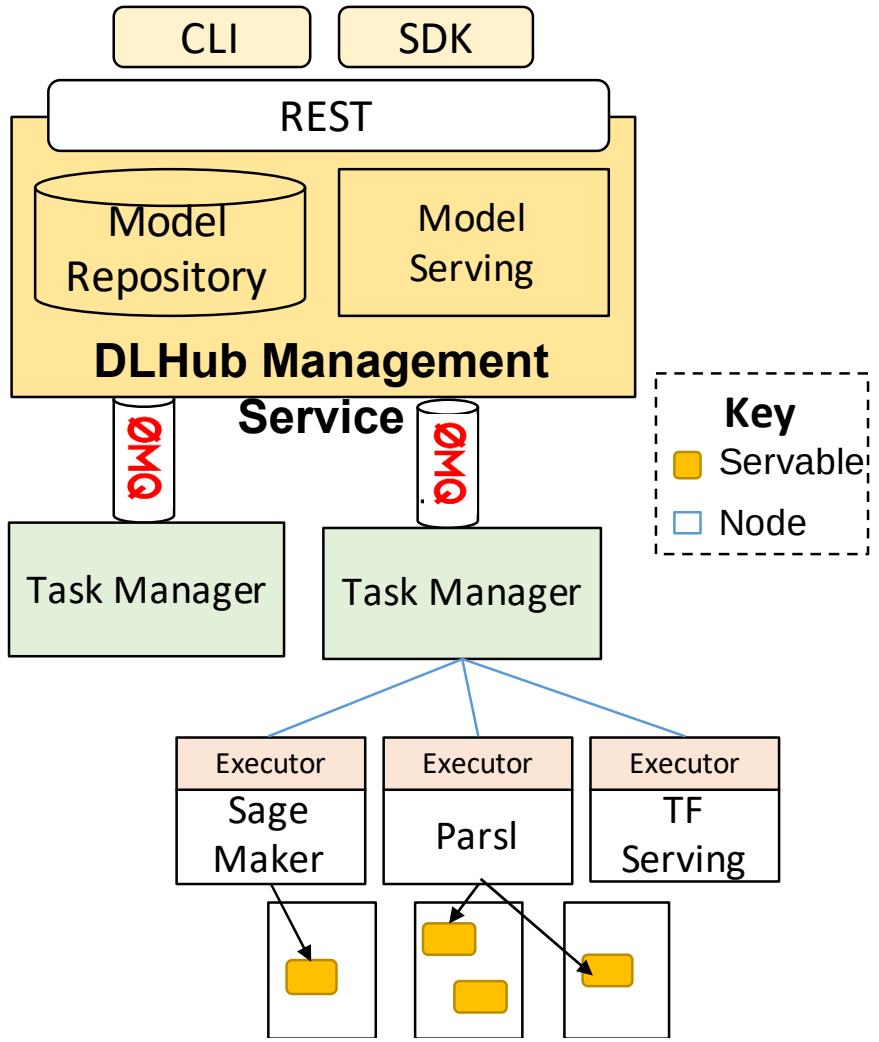
Funding: 2018 Argonne Adv. Computing LDRD

Argonne NATIONAL LABORATORY

ARGONNE LEADERSHIP COMPUTING FACILITY PETREL parsl

DLHub Architecture

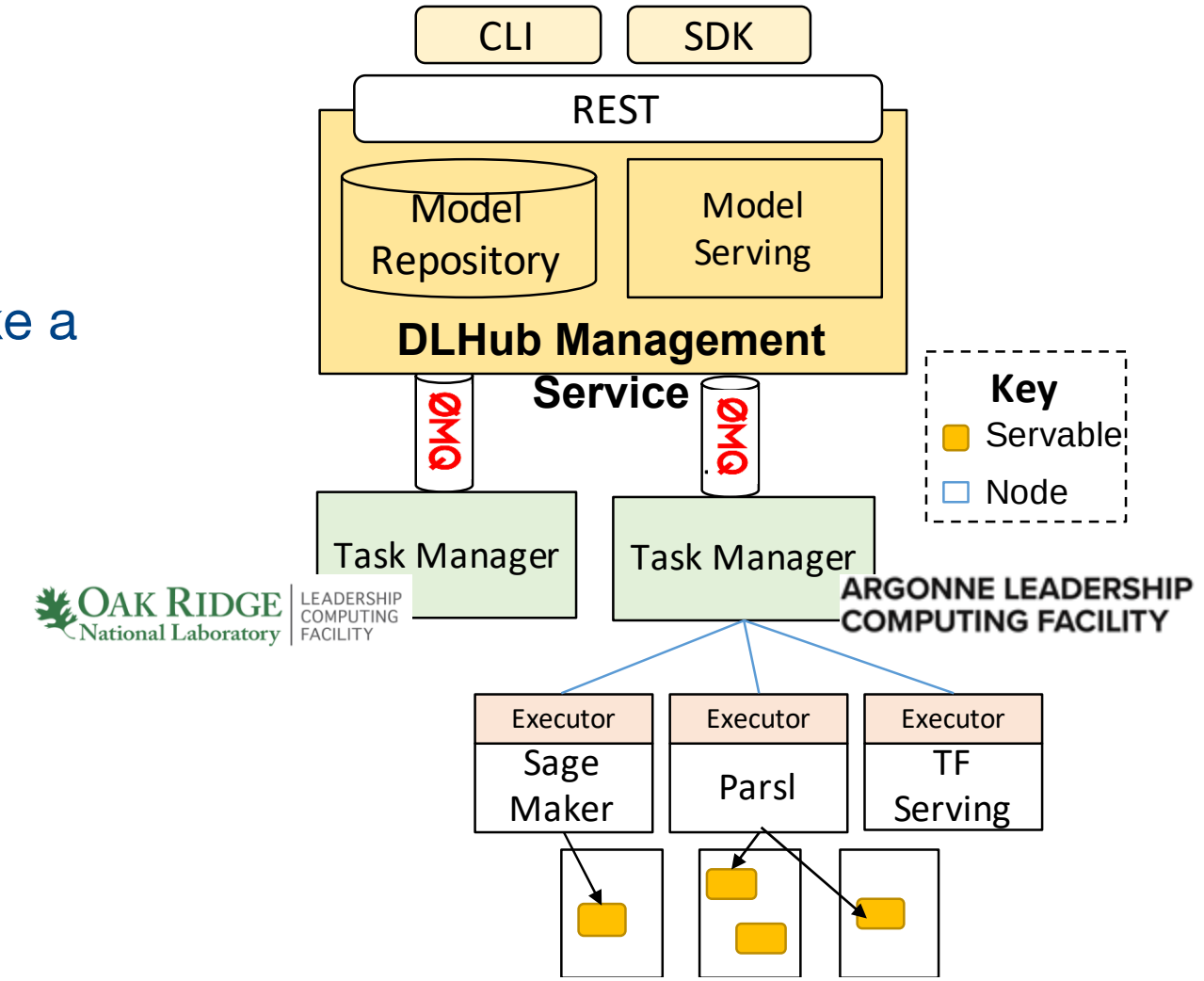
- REST API with Python SDK (available) / CLI (delivery in Nov. 2018)
  - Support model markup, data staging, registration, and invocation
- Model Repository
  - Container registry
  - Advanced search functions
  - Identifier minting capabilities



https://github.com/DLHub-Argonne

DLHub Architecture

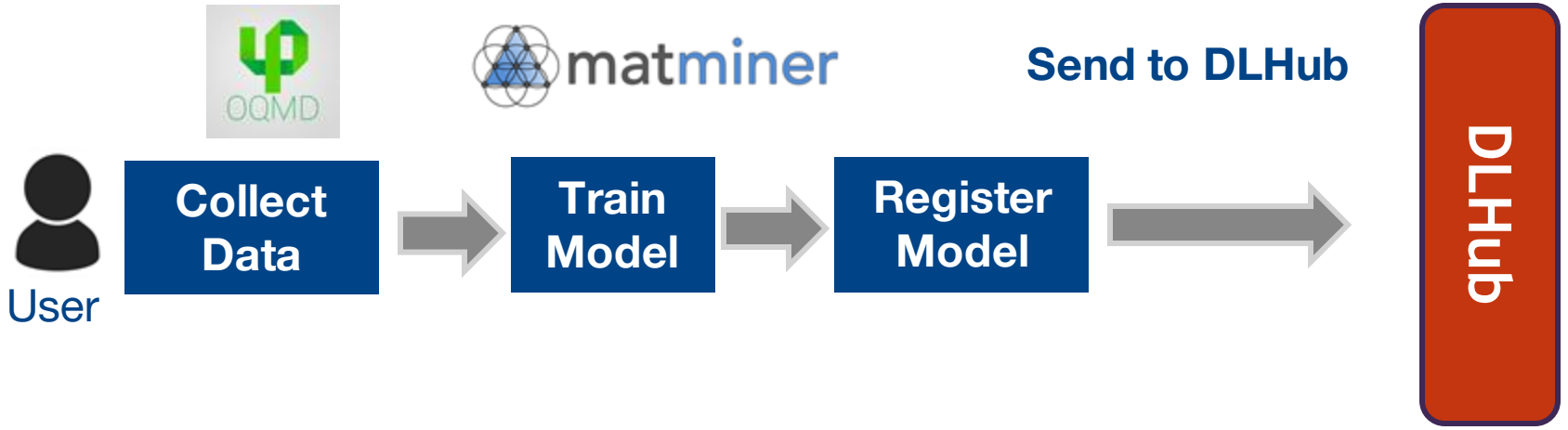
- Task Managers (TM) to support execution on various compute resources
- Executors chosen by TM to invoke a given servable'
- Caching at TM
- Data staging with Globus
- Batch submissions
- Scalability through deployment of model replicas



DLHub Model Registration and Publication

- **Register** model metadata, weights, and files to improve discoverability and reusability
- **Containerize** model to enhance interoperability
- **Identify** model with a permanent identifier (e.g., DOI, minid)
- **Version** model and data pre/post processing steps
- **Deploy** model with simplified interfaces for users
- **Control** access to model metadata and usage

- (future) **Automate** retraining and testing when new data are available



Funding: 2018 Argonne Adv. Computing LDRD

Marking up a Model - Python SDK

Existing Model

User Mark Up with SDK

SDK Extracts Metadata for Known Model Types

Send to DLHub (via Globus or HTTPS)

DLHub Containerization

Populate Search Index / Mint Identifiers

```python
from dlhub_toolbox.models.servables.keras import KerasModel
import pickle as pkl
import json

# Describe the keras model
model = KerasModel('model.hd5', list(map(str, range(10))))

# Describe the model
model.set_title("MNIST Digit Classifier")
model.set_name("mnist_tiny_example")
model.set_domains(["general","digit recognition"])

# Add link to paper describing the dataset
model.add_related_identifier("10.1109/CVPR.2007.383157", "DOI",
                             "IsDescribedBy")

model.set_authors(["Lecunn, Yann", "Cortes, Corinna"])

# Describe the outputs in more detail
model.output['description'] = 'Probabilities of being 0-9'
model.input['description'] = 'Image of a digit'
```

Python SDK - Automated Metadata Generation

Citation Metadata

```
"datacite": {
    "creators": [{
            "givenName": "Yann",
            "familyName": "Lecunn",
            "affiliations": []
        },
        {
            "givenName": "Corinna",
            "familyName": "Cortes",
            "affiliations": []
        }
    ],
    "titles": [{
        "title": "MNIST Digit Classifier"
    }],
    "publisher": "DLHub",
    "publicationYear": "2018",
    "relatedIdentifiers": [{
        "relatedIdentifier": "10.1109/CVPR.2007.383157",
        "relatedIdentifierType": "DOI",
        "relationType": "IsDescribedBy"
    }],
    "identifier": {
        "identifier": "10.YET/UNASSIGNED",
        "identifierType": "DOI"
    },
    "resourceType": {
        "resourceTypeGeneral": "InteractiveResource"
    }
},
```

DLHub Metadata

```
"dlhub": {
    "version": "0.1",
    "domains": [
        "general",
        "digit recognition"
    ],
    "visible_to": [
        "public"
    ],
    "id": null,
    "name": "mnist_tiny_example",
    "files": {
        "other": [],
        "model": "model.hd5"
    }
},
```

Access Control

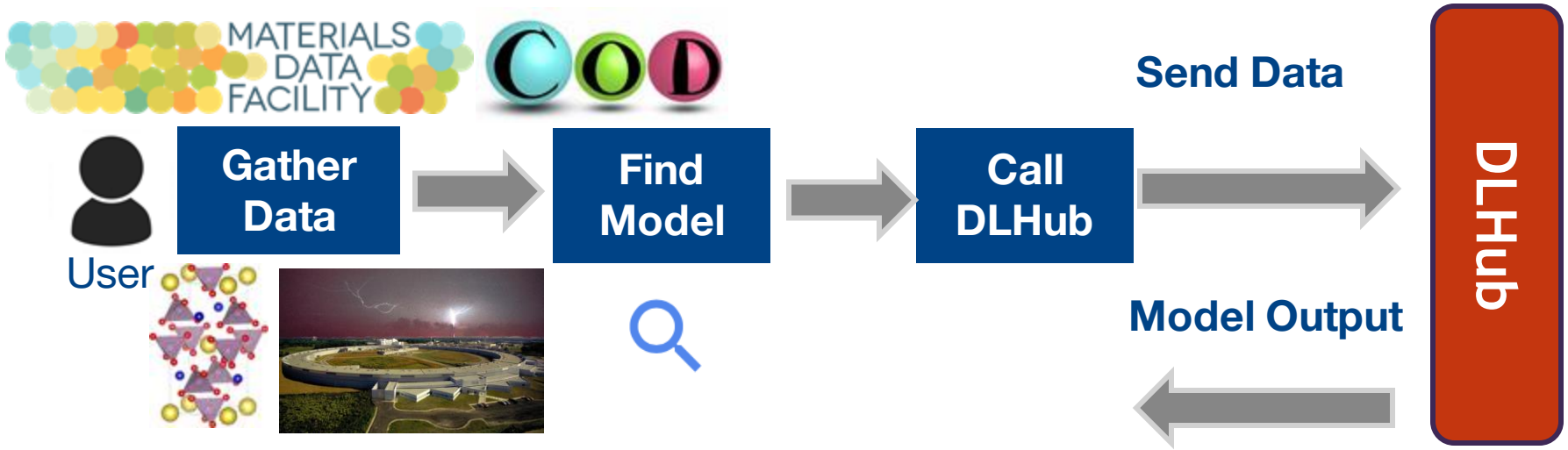
- Public
- Globus users
- Globus groups

Servable Metadata

```
"servable": {
    "methods": {
        "run": {
            "input": {
                "type": "ndarray",
                "description": "Image of a digit",
                "shape": [null,28,28,1]
            },
            "output": {
                "type": "ndarray",
                "description": "Probabilities of being 0-9",
                "shape": [null,10]
            },
            "parameters": {},
            "method_details": {
                "method_name": "predict",
                "classes": ["0","1","2","3","4",
                    "5","6","7","8","9"]
            }
        }
    },
    "shim": "keras.KerasServable",
    "language": "python",
    "dependencies": {
        "python": {
            "keras": "2.2.4",
            "h5py": "2.8.0"
        }
    },
    "type": "Keras Model",
    "model_type": "Deep NN"
```

DLHub Model Discovery and Usage

- Find curated and tested models
- Use models through simple interfaces



Funding: 2018 Argonne Adv. Computing LDRD

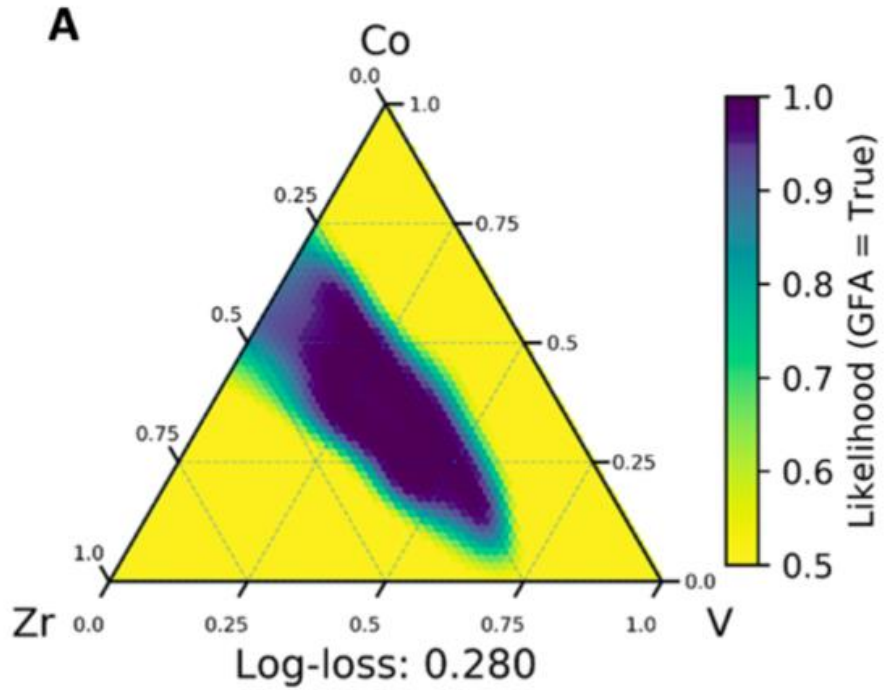
Argonne NATIONAL LABORATORY

Predicting Glass-forming Ability

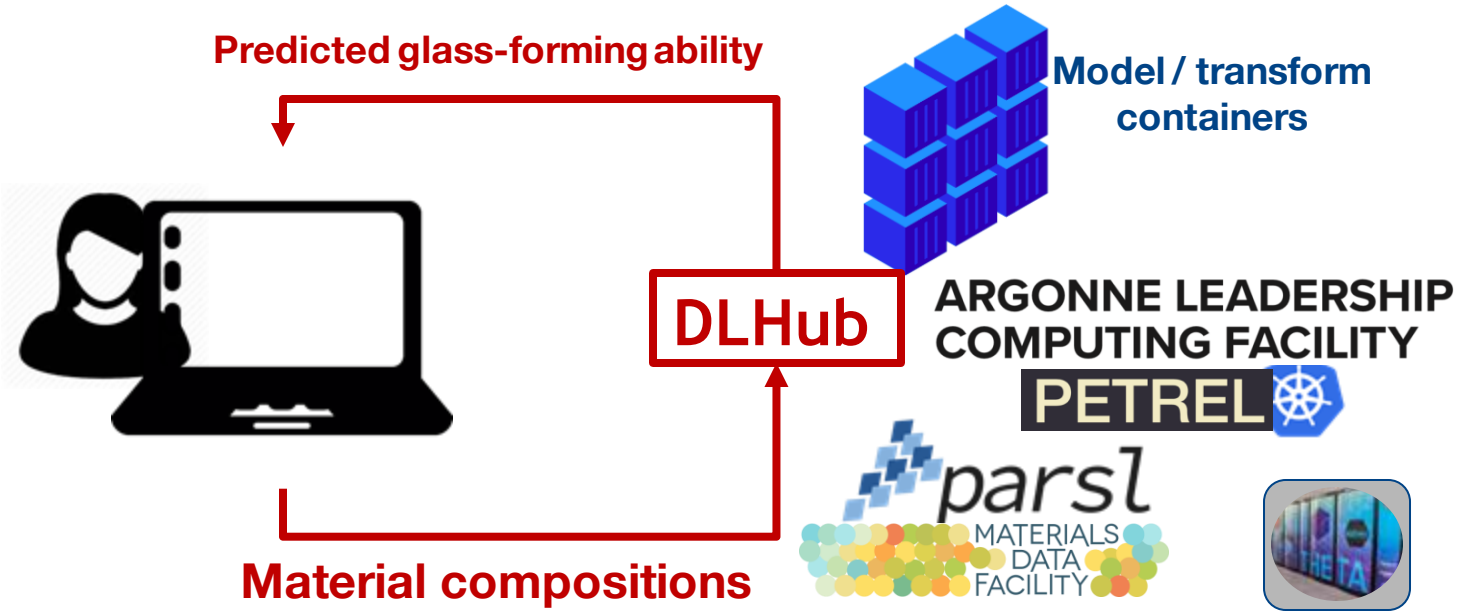
Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments

Fang Ren[1,*], Logan Ward[2,3,*], Travis Williams[4], Kevin J. Laws[5], Christopher Wolverton[2], Jason Hattrick-Simpers[6] and Apurva Mehta[1,†]

10.1126/sciadv.aaq1566



- Where are the model and trained weights?
- How do I run the model on my data?
- How can I retrain the model on new data?
- How can I build on this work?



Funding: 2018 Argonne Adv. Computing LDRD

Argonne NATIONAL LABORATORY

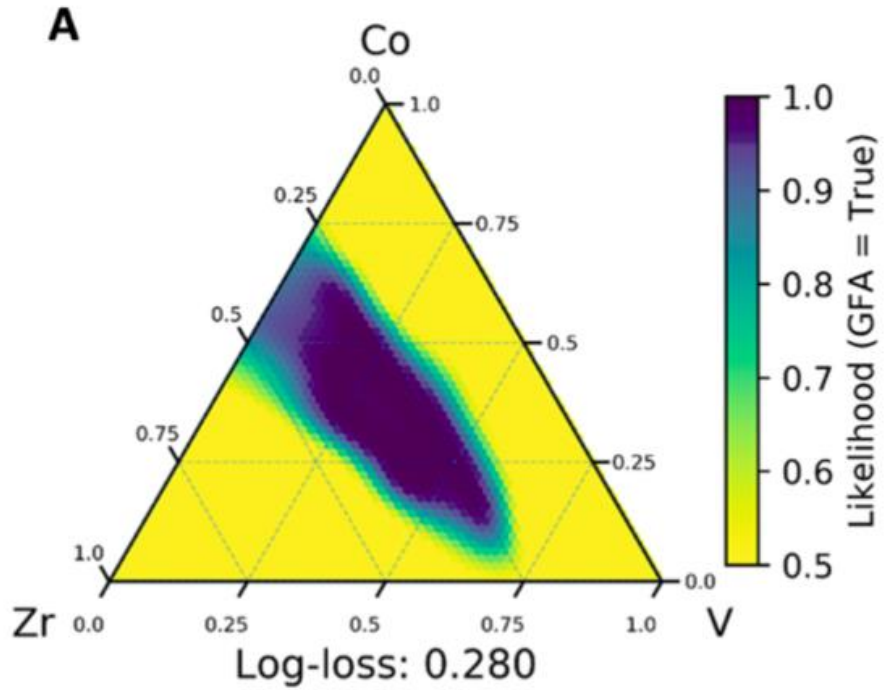
ARGONNE LEADERSHIP COMPUTING FACILITY PETREL parsl

Predicting Glass-forming Ability

Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments

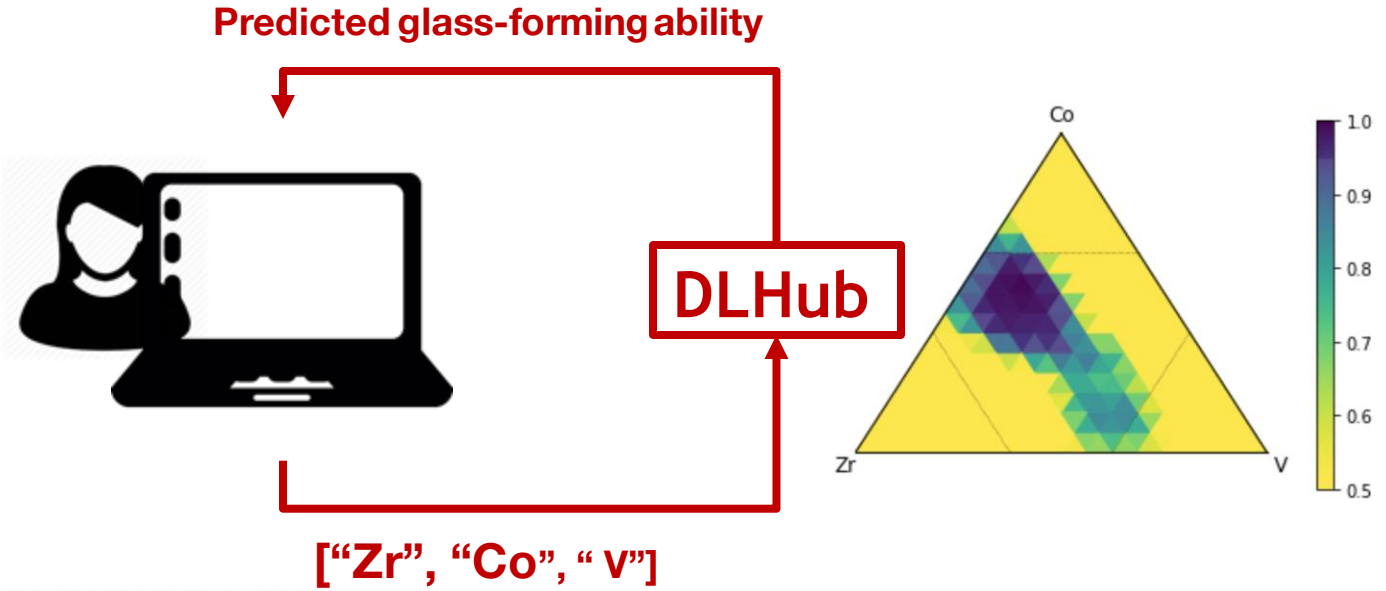
Fang Ren[1,*], Logan Ward[2,3,*], Travis Williams[4], Kevin J. Laws[5], Christopher Wolverton[2], Jason Hattrick-Simpers[6] and Apurva Mehta[1,†]

10.1126/sciadv.aaq1566




```
servable_name = "metallic_glass"
servable_id = dl.get_id_by_name(servable_name)
elems = ["V","Co","Zr"]

res = dl.run(servable_id, {"data":elems})
```



Funding: 2018 Argonne Adv. Computing LDRD

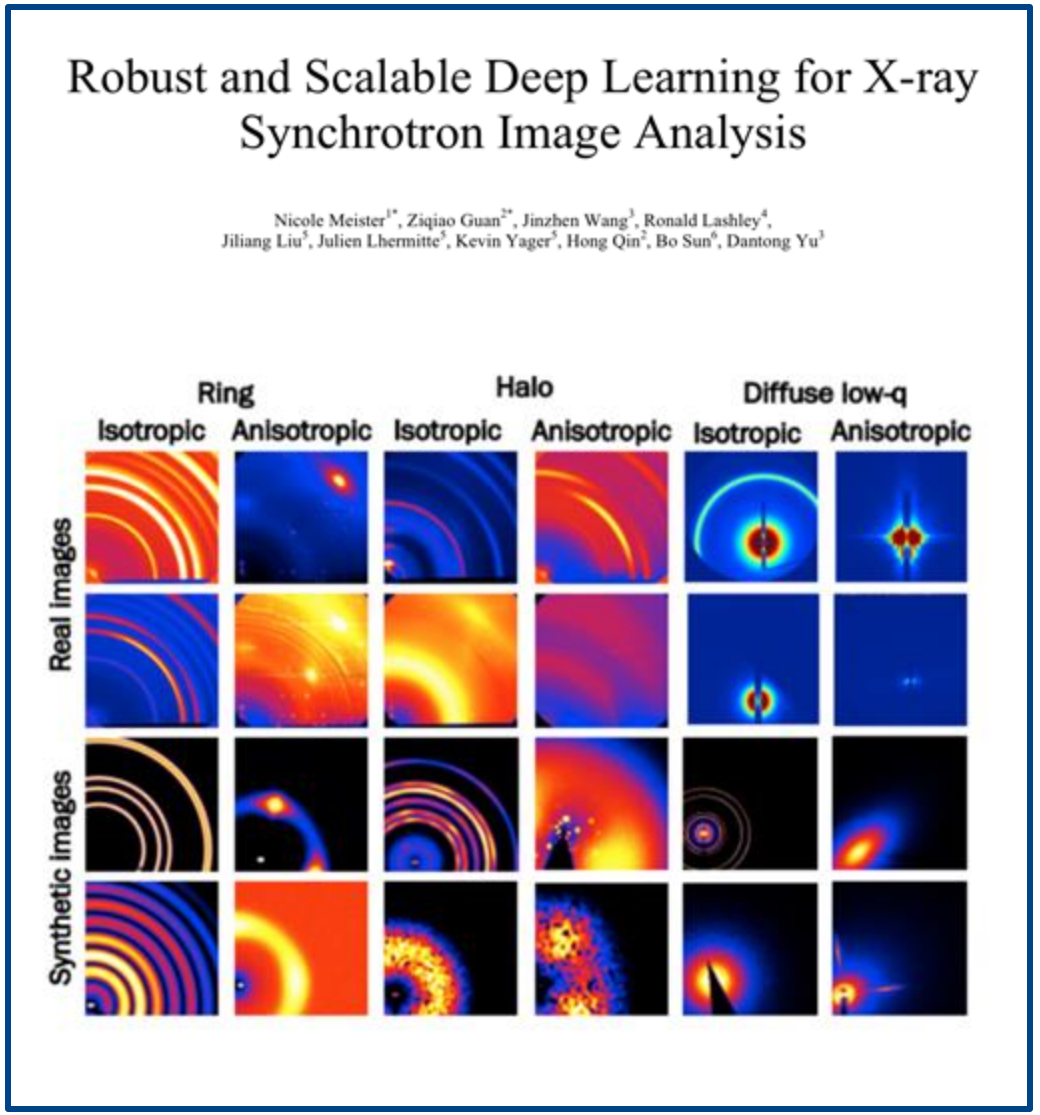
Argonne NATIONAL LABORATORY

ARGONNE LEADERSHIP COMPUTING FACILITY

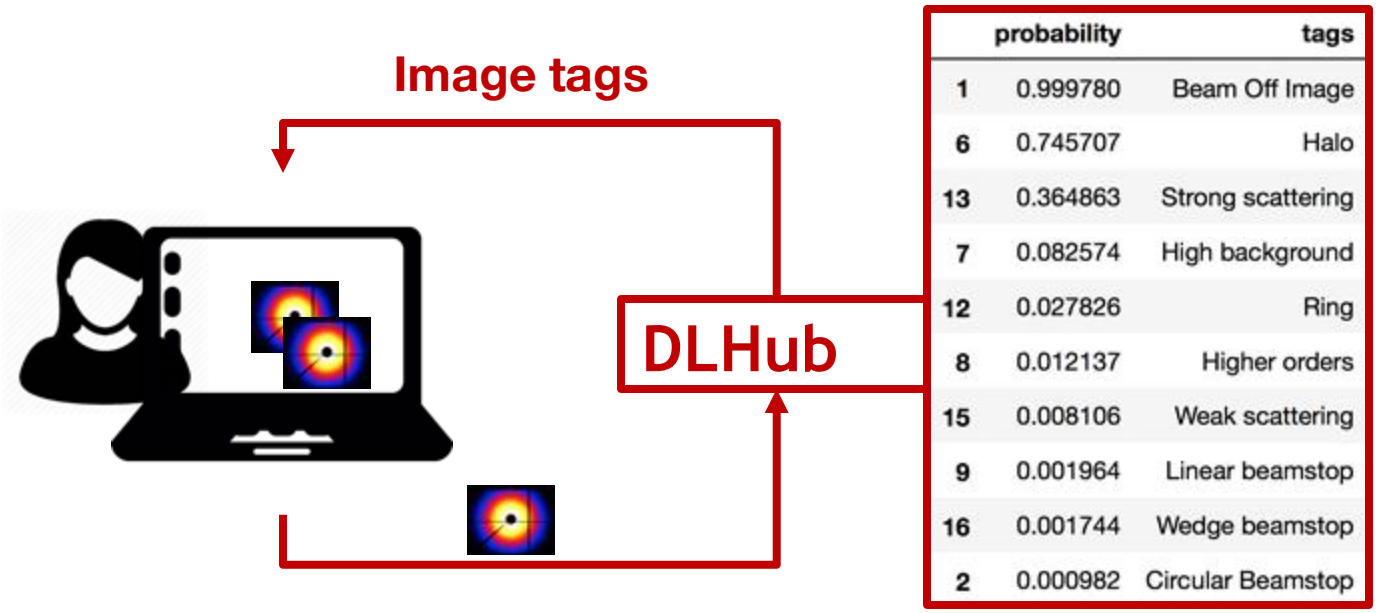
PETREL

parsl

Analyzing Beamline Images



- Stage data into containers via Globus HTTPS
- Pass valid token and data location



	probability	tags
1	0.999780	Beam Off Image
6	0.745707	Halo
13	0.364863	Strong scattering
7	0.082574	High background
12	0.027826	Ring
8	0.012137	Higher orders
15	0.008106	Weak scattering
9	0.001964	Linear beamstop
16	0.001744	Wedge beamstop
2	0.000982	Circular Beamstop

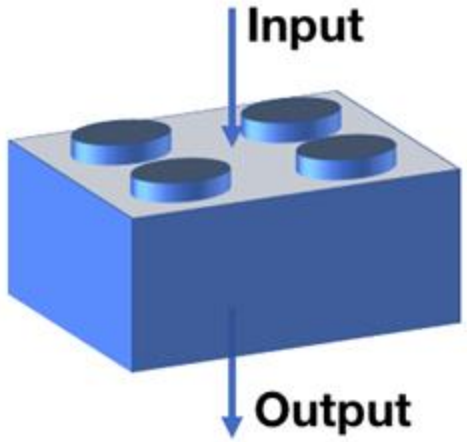
Funding: 2018 Argonne Adv. Computing LDRD

Argonne NATIONAL LABORATORY

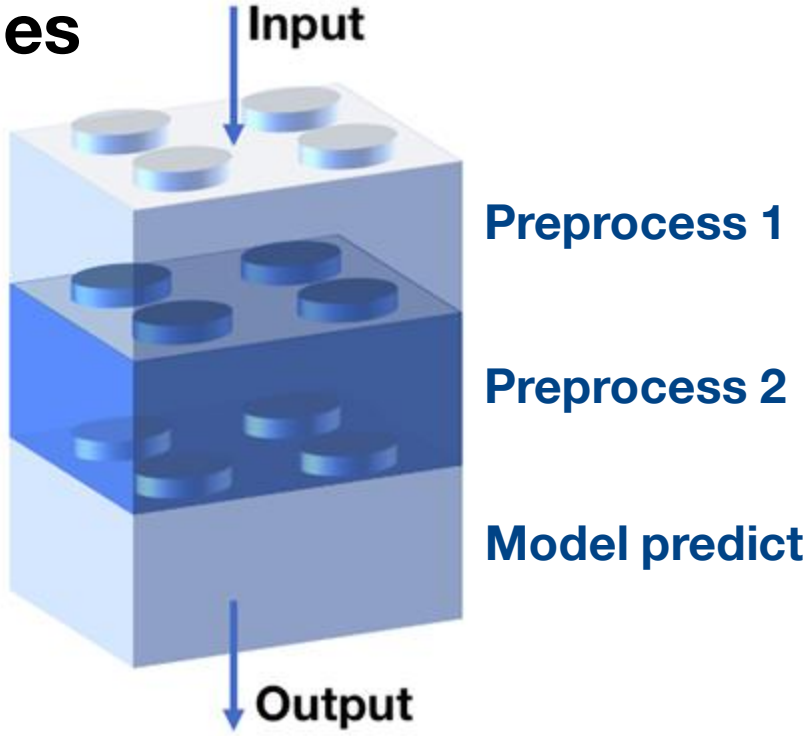
ARGONNE LEADERSHIP COMPUTING FACILITY PETREL parsl

Data and Learning Hub (DLHub): Pipelines

Servables



Pipelines



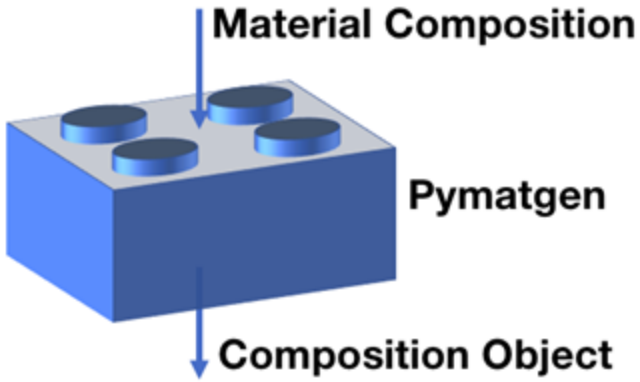
Funding: 2018 Argonne Adv. Computing LDRD

Argonne NATIONAL LABORATORY

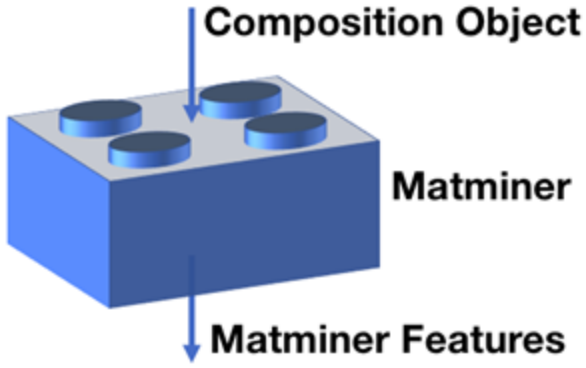
ARGONNE LEADERSHIP COMPUTING FACILITY PETREL parsl

Pipelines: Predicting Formation Enthalpy

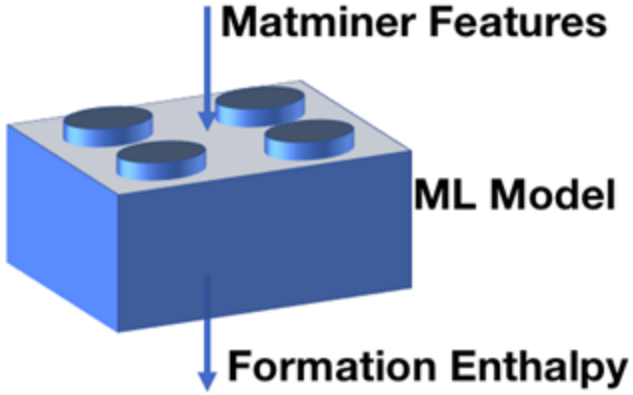
Step 1



Step 2



Step 3

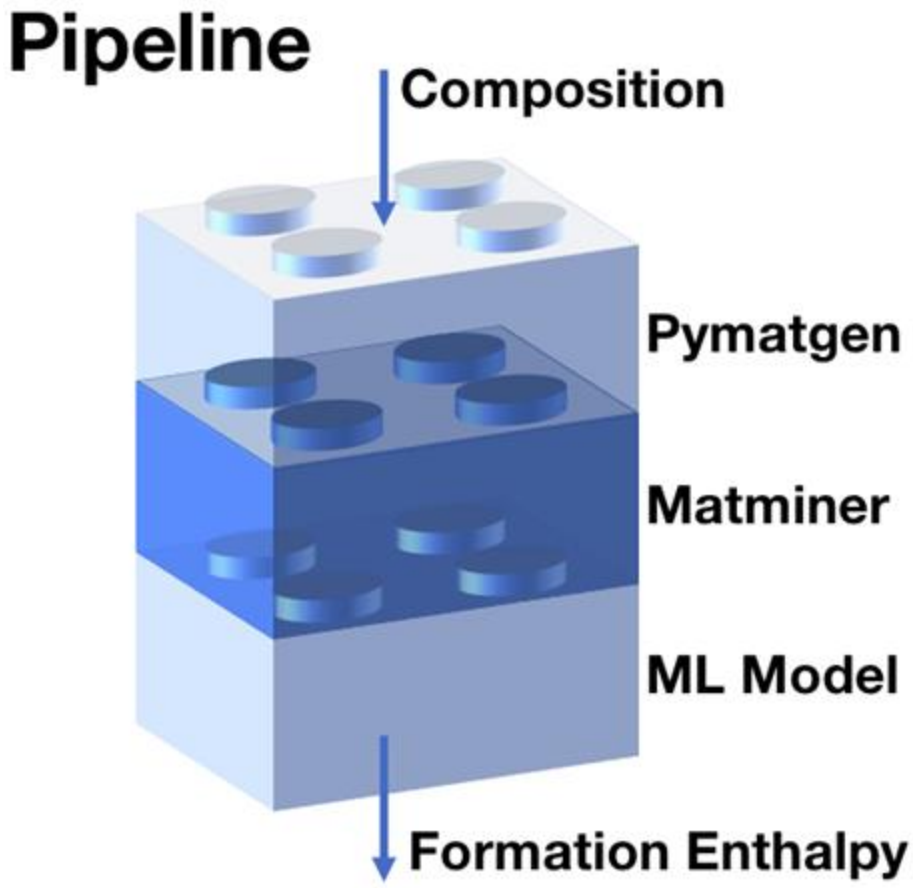


Funding: 2018 Argonne Adv. Computing LDRD

Argonne NATIONAL LABORATORY

ARGONNE LEADERSHIP COMPUTING FACILITY PETREL parsl

Predicting Formation Enthalpy



Funding: 2018 Argonne Adv. Computing LDRD

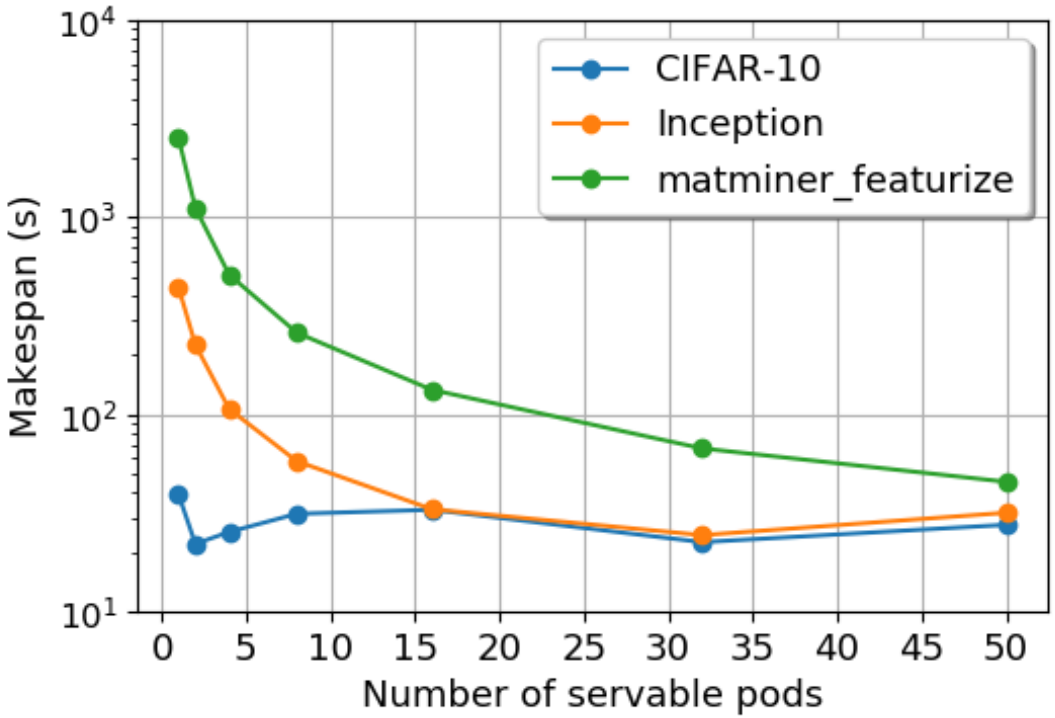
Argonne NATIONAL LABORATORY

ARGONNE LEADERSHIP COMPUTING FACILITY PETREL parsl

DLHub Performance

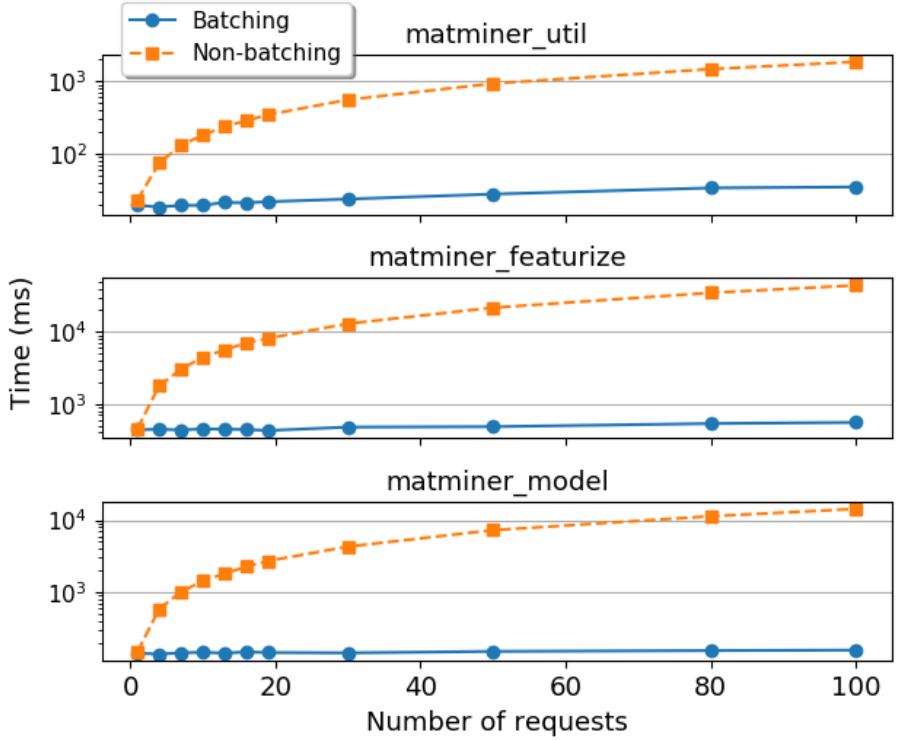
DLHub Performance

Scale Testing



The time required for the Inception, CIFAR10, and Matminer-featurize models to process 5000 inferences with varying numbers of replicas.

Batching



Servable invocation time, with and without batching.

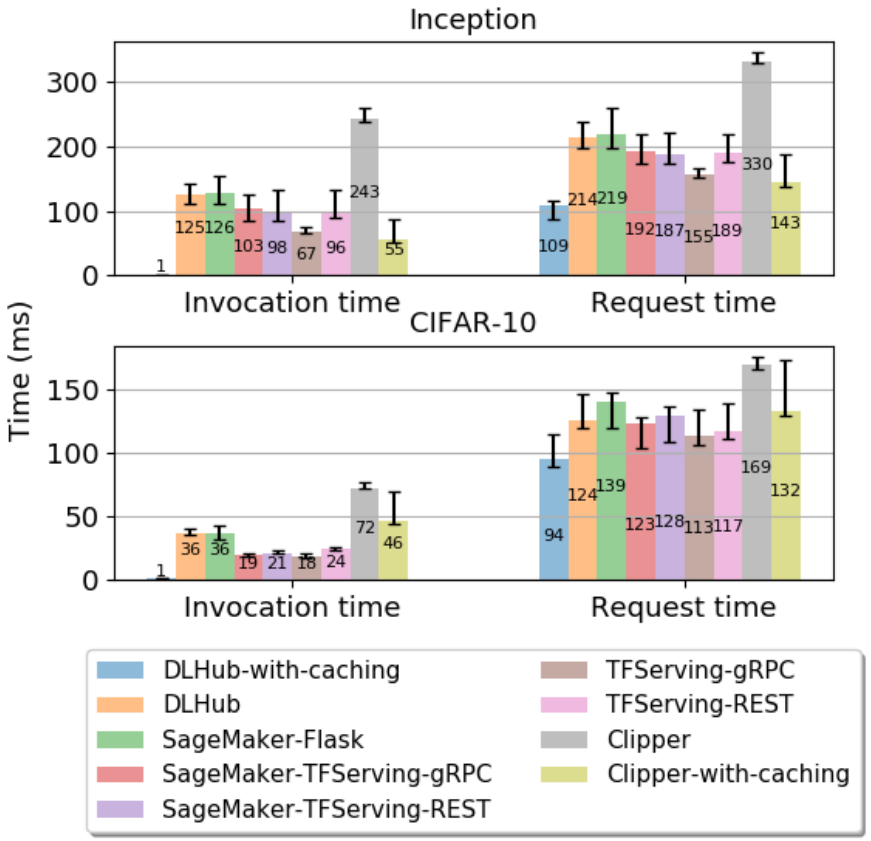
Funding: 2018 Argonne Adv. Computing LDRD

Argonne NATIONAL LABORATORY

ARGONNE LEADERSHIP COMPUTING FACILITY PETREL parsl

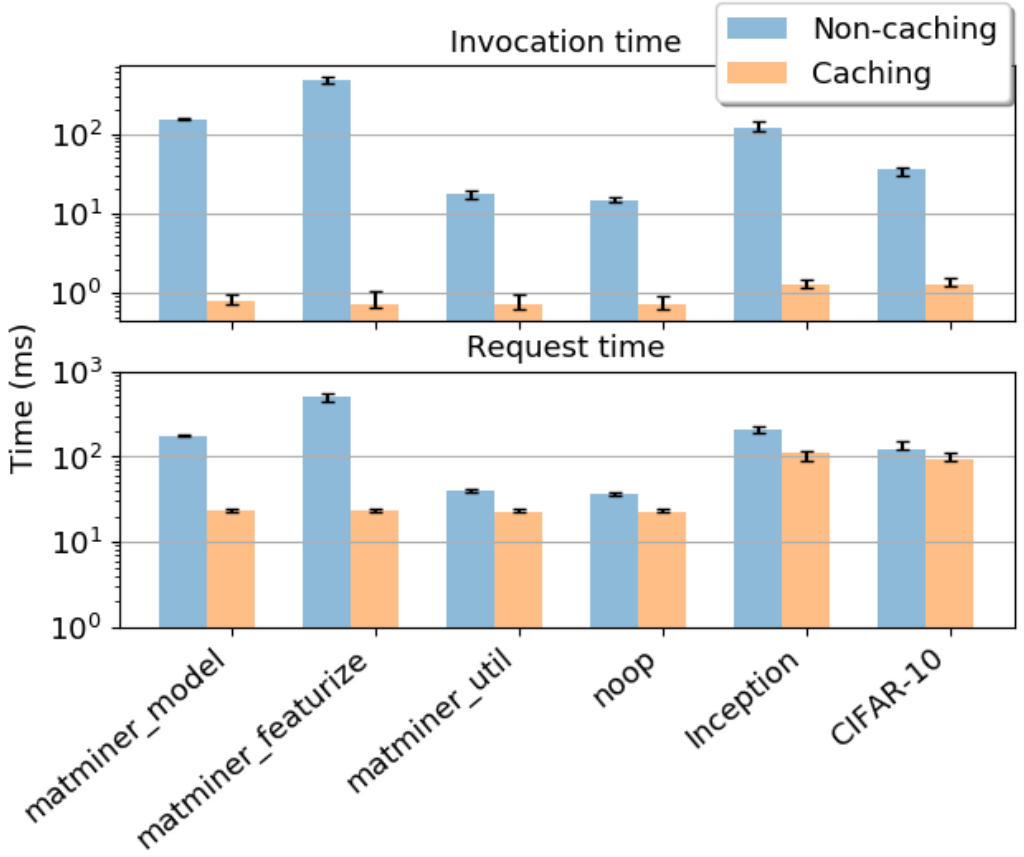
DLHub Performance

Serving General Models



Performance of different serving systems on the Inception and CIFAR-10 problems.

Caching



Performance impact of caching in DLHub. Bars and error bars show median and 5th/95th percentiles

Funding: 2018 Argonne Adv. Computing LDRD

Argonne NATIONAL LABORATORY

ARGONNE LEADERSHIP COMPUTING FACILITY

PETREL

parsl

DLHub Summary

Model deposit and discovery

- Developed a model schema to promote discovery
- Implemented advanced search and filtering
- Built ingest flow: models are dynamically staged, packaged, dockerized, published, and indexed

Model serving

- Deployed capabilities for users to run inference with SDK and CLI
- Automated testing of containers
- Implemented caching and batching

Support for multiple execution sites

- PetrelKube: Parsl, TF serving, Sagemaker
- Other: AWS, OSG

Authentication

- Protected model metadta and inference with GlobusAuth
- Secured data staging

Monitoring and statistics

- Request, invocation, data staging

Future work

- Dynamic scaling by load
- Build Web UI to create pipelines and invoke models
- Cache at the servable level within pipelines
- Couple DLHub to data sources (MDF, etc.)
- Integrate with ML frontend tools (DeepForge), optimization tools (DeepHyper), and more
- Create interface for training and retraining of models

Funding: 2018 Argonne Adv. Computing LDRD

Argonne NATIONAL LABORATORY

ARGONNE LEADERSHIP COMPUTING FACILITY PETREL parsl

Thanks to our sponsors!

ARGONNE LEADERSHIP COMPUTING FACILITY

ALCF DF

PETREL

Data Management and Sharing Pilot

Parsl

NSF

Globus

NIST

U.S. DEPARTMENT OF ENERGY

NATIONAL INSTITUTES OF HEALTH

NSF

MATERIALS DATA FACILITY

IMaD

NSF

NIST

CHiMaD

Center for Hierarchical Materials Design

DLHub

U.S. DEPARTMENT OF ENERGY

Argonne LDRD

U.S. DEPARTMENT OF ENERGY

Argonne NATIONAL LABORATORY

THE UNIVERSITY OF CHICAGO