PARSL & FUNCX FEST 2022

ARGONNE PRIVACY-PRESERVING FEDERATED LEARNING WITH FUNCX

APPFL

funcX

TRUNG-HIEU HOANG

Ph.D. Candidate
Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign
Data Science and Learning Division, Argonne National Laboratory

U.S. DEPARTMENT OF ENERGY Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.

Sep 14, 2022

TEAM

Ravi Madduri

Kibaek Kim

Ryu Minseok

Hieu Hoang

Eliu Huerta

Pradeep Natarajan

Puneet Batra

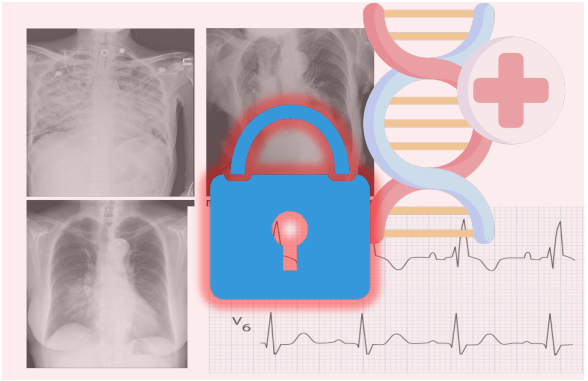
Maryellen Giger

Sam Armato

Multi-disciplinary team from ANL, LLNL, Harvard, UChicago and UIUC

KEY MOTIVATIONS & OBJECTIVES

Data shift in Machine Learning

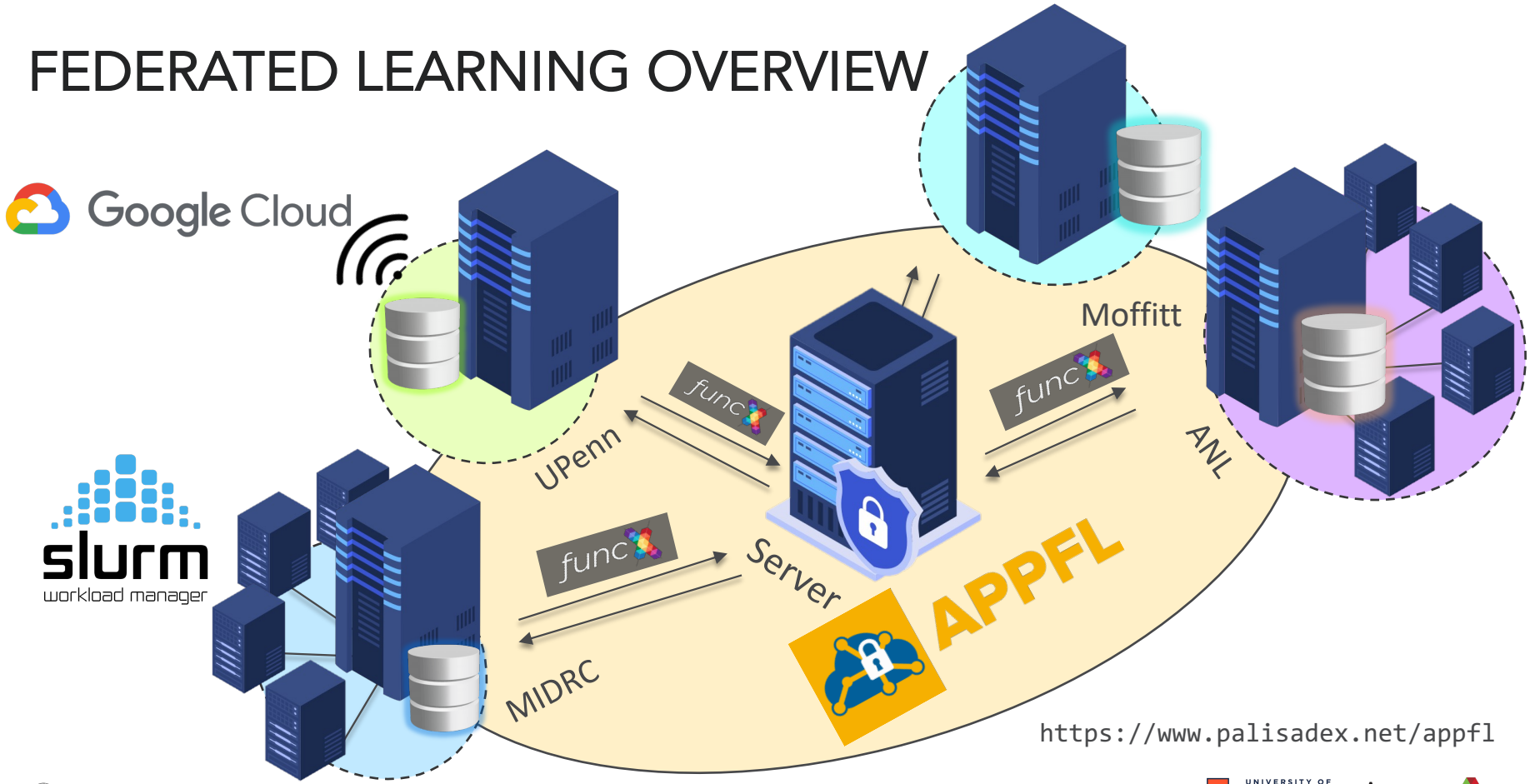


Policy concerns with Biomedicine Data

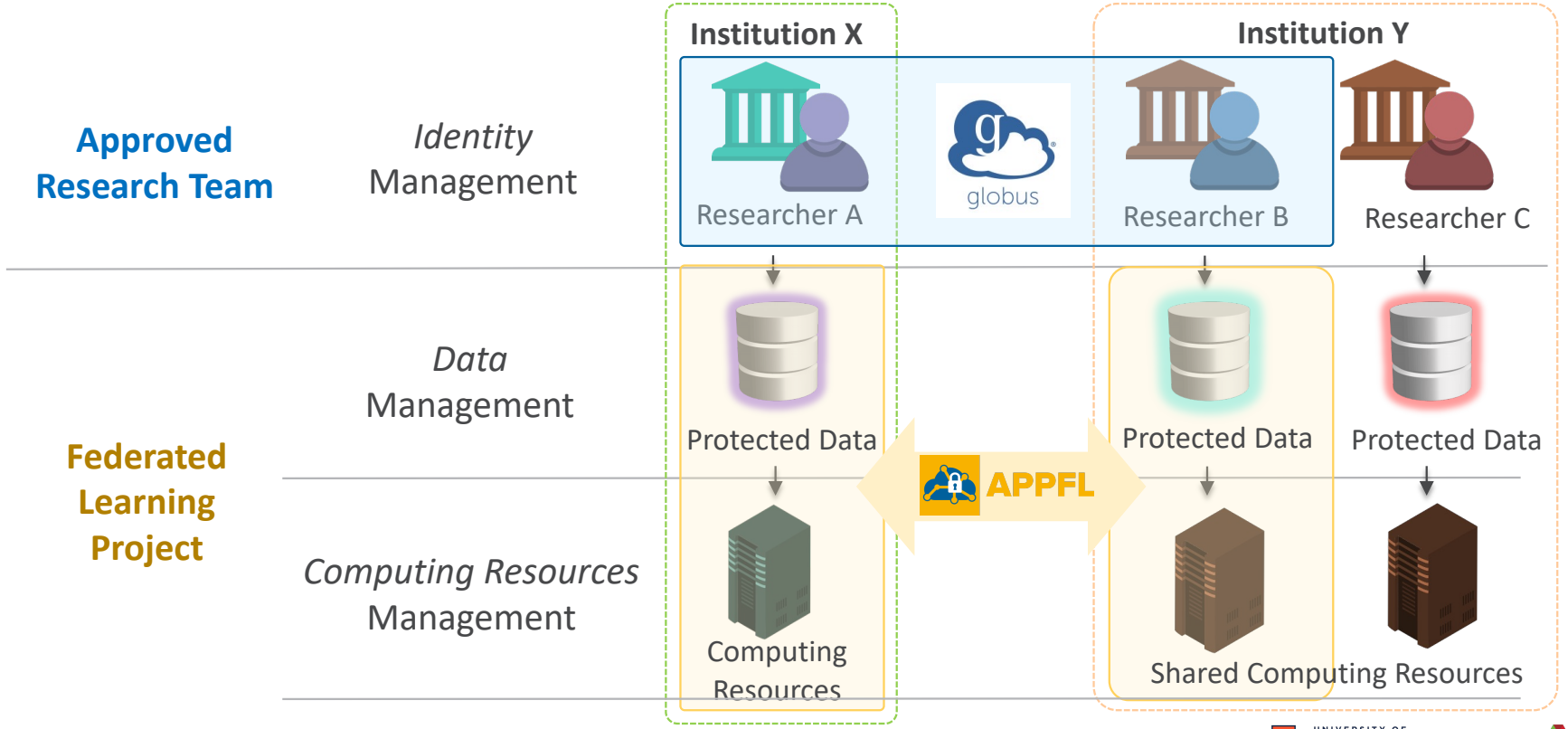
Key objectives

To develop and implement **APPFL (Argonne Privacy-Preserving Federated Learning)** framework that implements **Differential Privacy algorithms** for training **Federated Learning models** with the **biomedical datasets from multiple organizations**

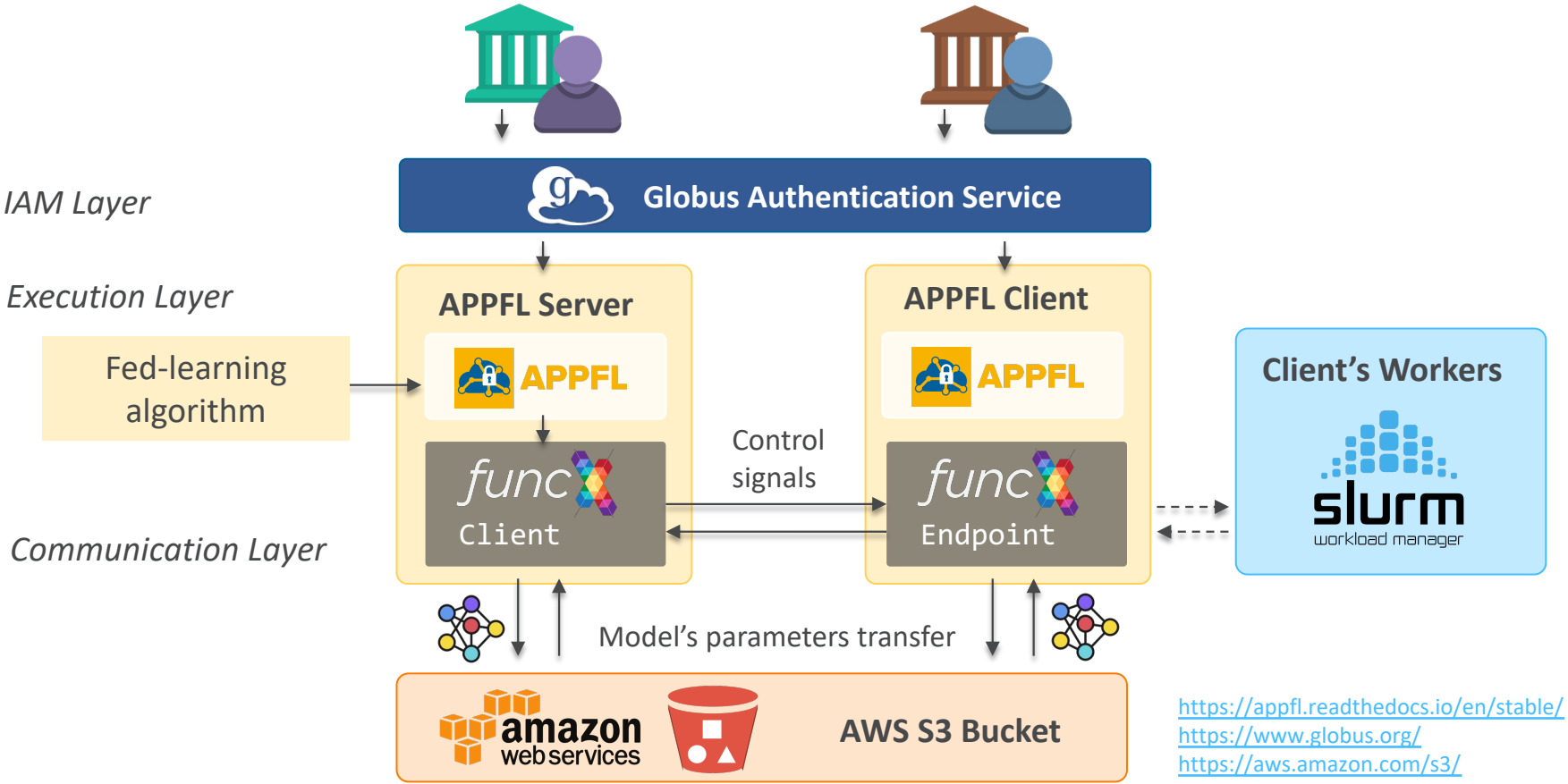
FEDERATED LEARNING OVERVIEW



FEDERATED LEARNING RESOURCES MANAGEMENT



KEY COMPONENTS



https://appfl.readthedocs.io/en/stable/
https://www.globus.org/
https://aws.amazon.com/s3/

SETTING UP AN EXPERIMENT

Users need to define the following files:

User-defined Model & Data-loader

`model.py` `dataloader.py`

Using PyTorch, as regular Python scripts

Client Configuration

`clients.yml`

Parameters, funcX-endpoint IDs and data loaders at all clients

Server Configuration

`server.yml`

Parameters, fed-learning algorithm at server

EXAMPLES – FUNCX TASK EXECUTION LOG

timestamp	task_name	client_name	status	execution_time
8/16/22 9:00 AM	client_validate_data	uiuc-cig-01-gpu-02	success	13.63
8/16/22 9:00 AM	client_validate_data	uchicago-gpu	success	39.04
8/16/22 9:01 AM	client_training	uiuc-cig-01-gpu-02	success	34.95
8/16/22 9:01 AM	client_training	uchicago-gpu	success	65.08
8/16/22 9:02 AM	client_training	uiuc-cig-01-gpu-02	success	31.73
8/16/22 9:02 AM	client_training	uchicago-gpu	success	35.21
8/16/22 9:03 AM	client_training	uiuc-cig-01-gpu-02	success	31.51
8/16/22 9:03 AM	client_training	uchicago-gpu	success	35.42
8/16/22 9:03 AM	client_training	uiuc-cig-01-gpu-02	success	28.75
8/16/22 9:03 AM	client_training	uchicago-gpu	success	35.63


```
- task_name: client_training
endpoint: uiuc-cig-01-gpu-02
start_at: '2022-08-16
09:01:22.433471'
end_at: '2022-08-16 09:01:57.384578'
events:
start_endpoint_execution: '2022-08-16
09:01:22.552091'
stop_endpoint_execution: '2022-08-16
09:01:54.578533'
timing:
load_dataset: 7.529
download_global_state: 0.626
load_global_state_to_device: 3.94
training_client_update: 19.404
epoch_1:
val_before_update_val_set: 0.726
train_one_epoch: 1.849
epoch_2:
train_one_epoch: 1.852
```

TAKEAWAYS

- **APPFL:** An **open-source** framework for **privacy-preserving federated-learning** tasks
- **funcX** helps us to efficiently perform remote task execution at clients **without thinking too much about the underlying computing infrastructure**
- **funcX authentication** via Globus provides a reliable way for **managing data/computing resources access** for teams of cross-institution researchers
- Our project facilitates collaborations on developing machine learning algorithms in many biomedicine researches. **We are actively looking for collaborations in the future!**

`https://appfl.readthedocs.io/en/stable/`

THANKS FOR YOUR ATTENTION

hthieu@illinois.edu

https://appfl.readthedocs.io/en/stable/

https://github.com/APPFL/APPFL/tree/thoang/funcx