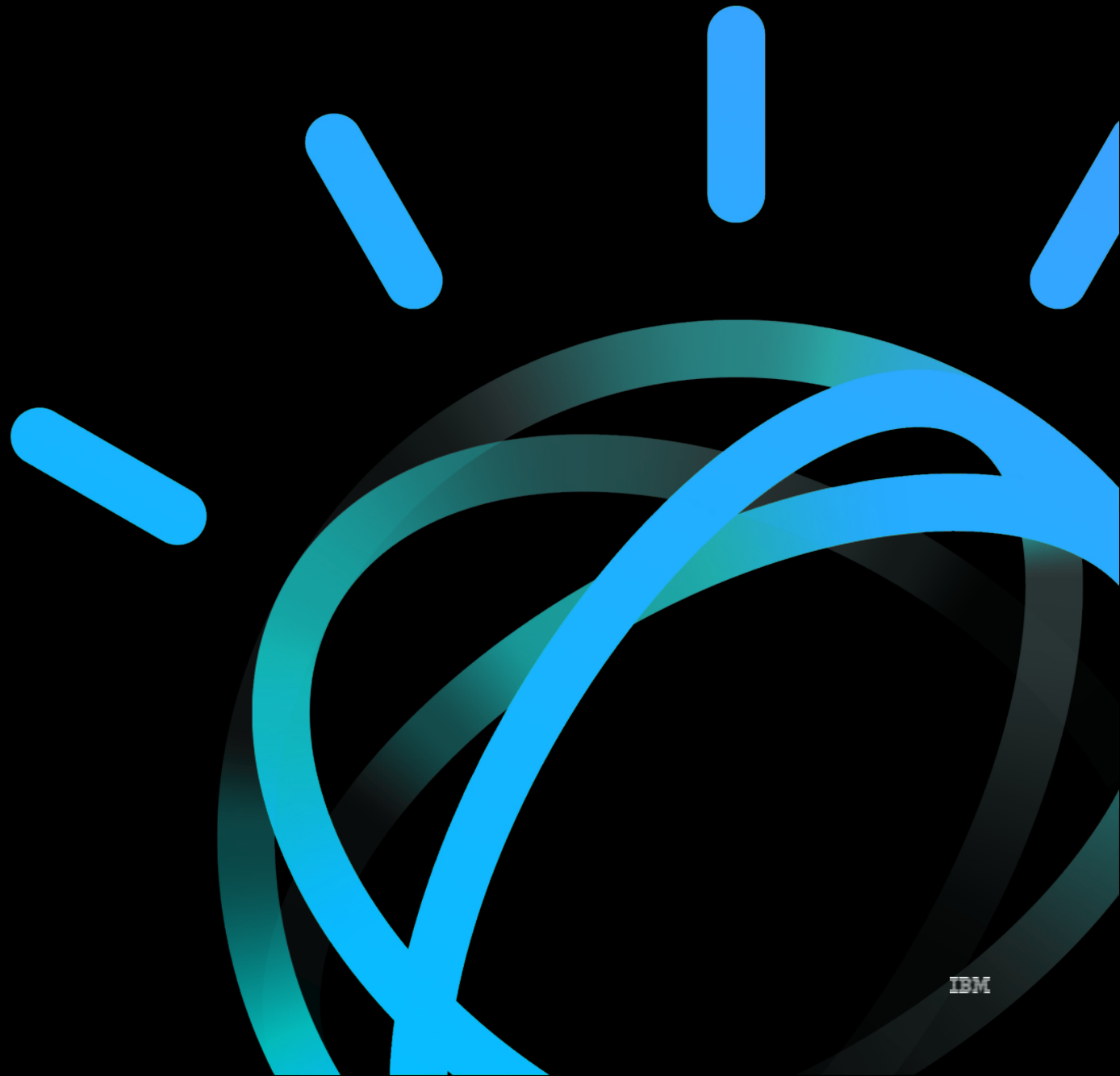
IBM Data Science & AI

Simplify, Scale, and Speed your Enterprise AI initiatives

University of Wisconsin
Data Science Bazaar

Ryan Kather
02/17/2021



IBM

Please note

IBM’s statements regarding its plans, directions, and intent are subject to change or withdrawal without notice and at IBM’s sole discretion.

Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision.

The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract.

The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user’s job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.

Agenda

AI Ladder

Cloud Pak for Data Overview

AutoAI

Components

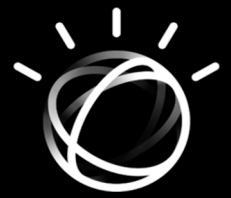
Bias Mitigation

Federated Learning

Hands On Lab

The AI Ladder

A prescriptive approach for data and AI



INFUSE - Operationalize AI throughout the business

ANALYZE - Build and scale AI with trust and transparency

ORGANIZE - Create a business-ready analytics foundation

COLLECT - Make data simple and accessible

MODERNIZE
Make your data ready for an AI and hybrid cloud world

Talent & Skills

One Platform, Any Cloud

IBM Cloud Pak for Data

Simplifies, unifies and automates the delivery of data and AI to the business

Foundational Services

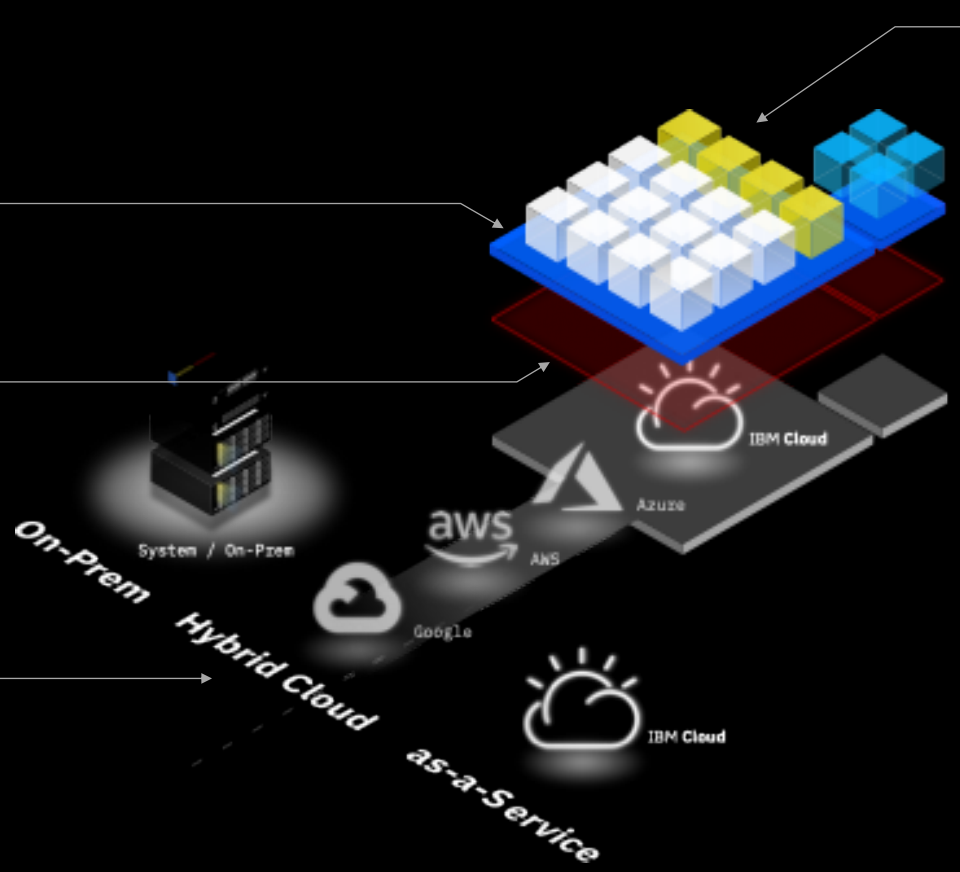
Essential to Cloud Pak for Data, Foundational Services provide a command-line interface, an administration interface, a services catalog, a central list of connections, and the central user experience.

Red Hat® OpenShift®

A streamlined multicloud container platform foundation that can run anywhere. OpenShift license dedicated to run Cloud Pak for Data.

A truly hybrid multicloud world

Don't let lock-in or major IT decisions affect how your Data and AI teams operate. Your work and data should be accessible in any hybrid multicloud strategy.



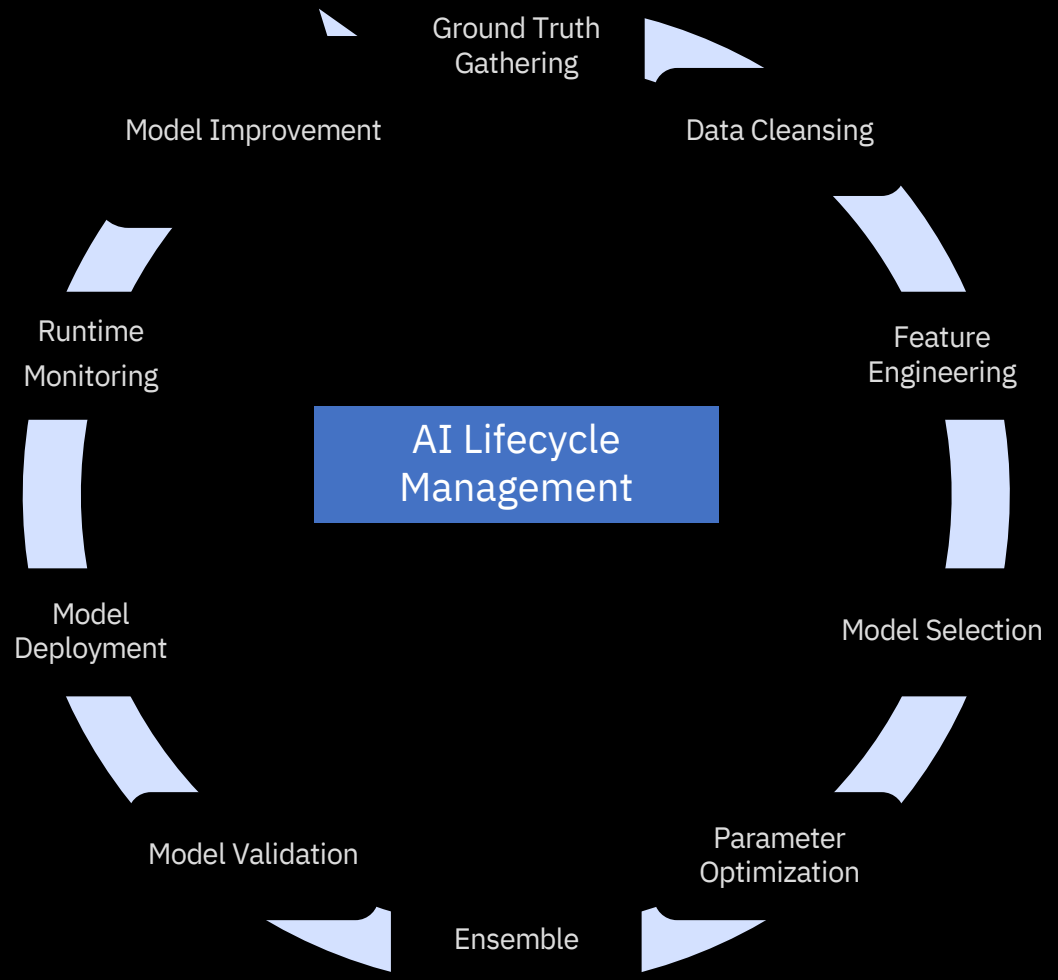
Base Services

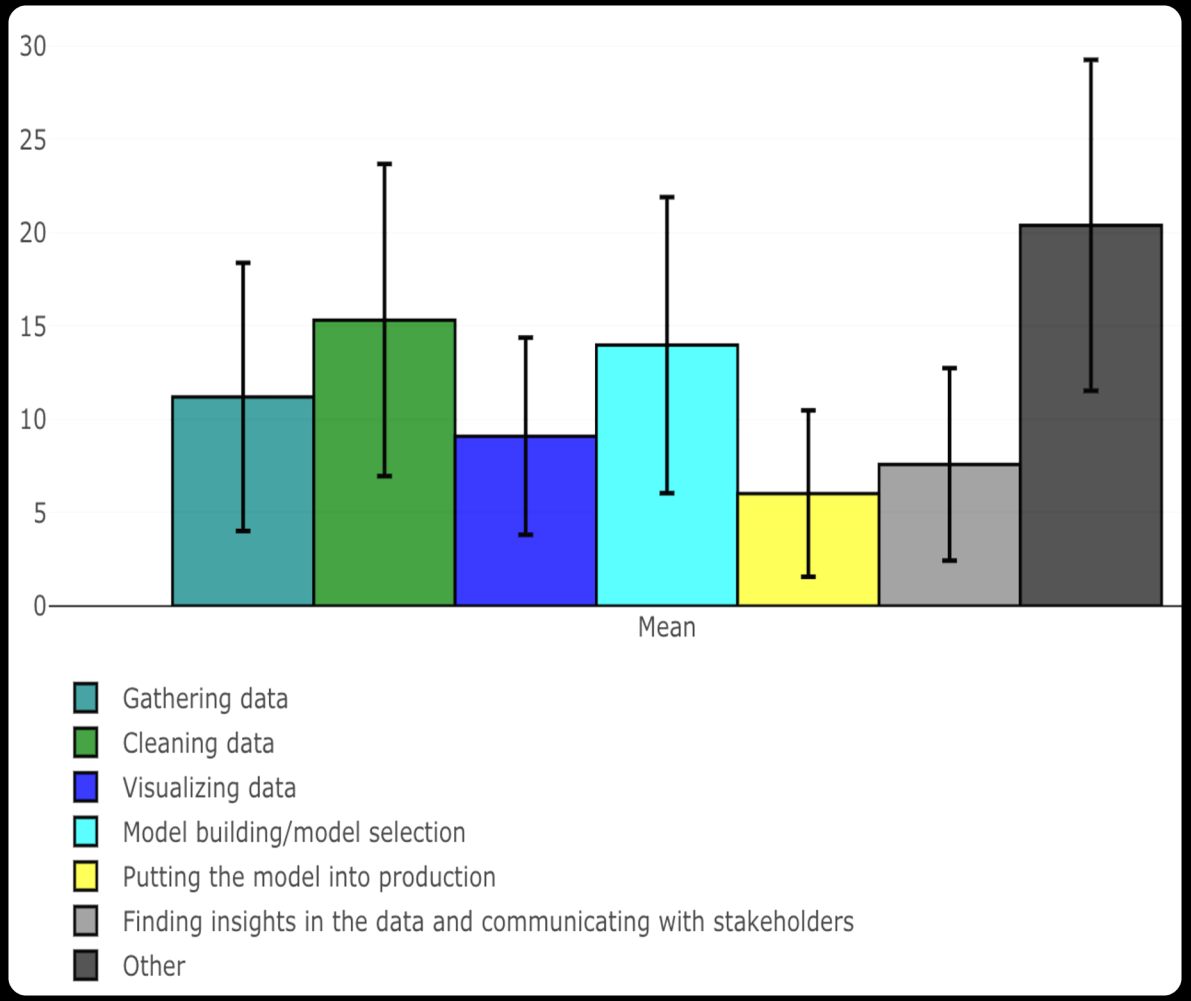
Core to Cloud Pak for Data, these Base services fastrack organizations on their **end-to-end data and AI journey** with comprehensive continuum of capabilities.

Extended Services

Beyond its Base Services, Cloud Pak for Data has a growing **ecosystem of Open Source, Partner, and IBM Extended Services** to expand the breadth of capabilities for teams.

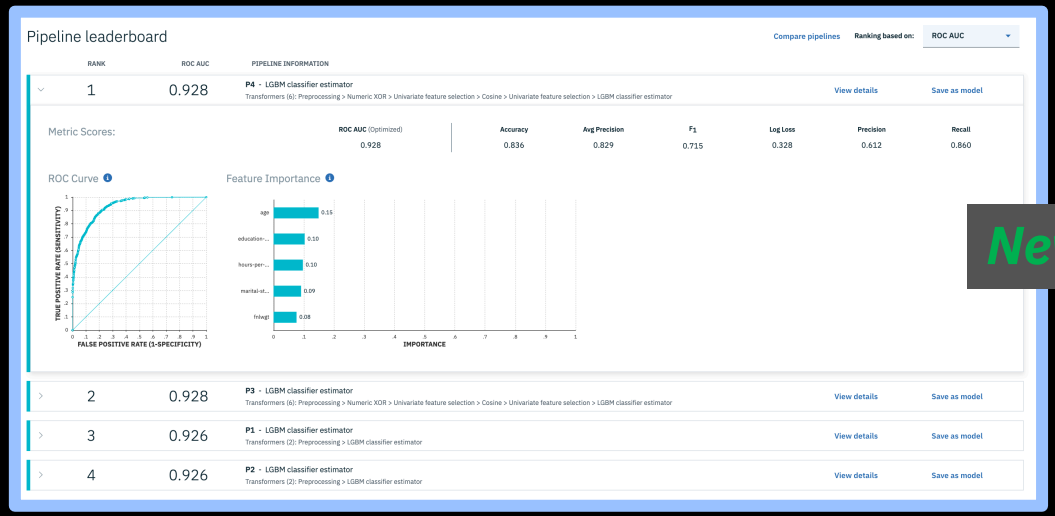
Case for AI Automation: AI Workflow's Bigger & More Complex



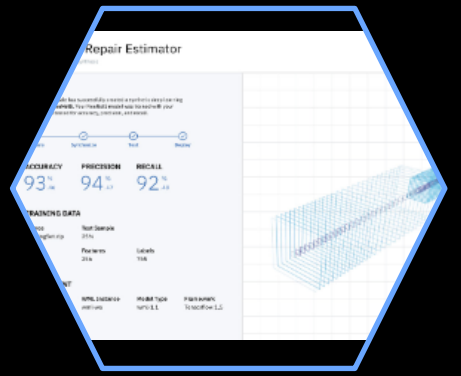


Source: https://www.kaggle.com/paultimothymooney/2018-kaggle-machine-learning-data-science-survey

IBM’s Strategy for Automation of AI Development



New



Transfer Learning

- Small data and compute requirements – leverage Watson base model
- Unstructured data (image/text)
- Featured in Watson Services, available through Watson Studio

AutoAI Experiments | Pipeline optimization

- Automation from data prep to model selection and tuning
- Structured data (csv)
- **New! AutoAI** GA as of May 2019

Neural Network Search

- Automate the generation and training of a custom deep learning model
- Unstructured data (image/text)
- **NeuNetS** as a feature of Watson Studio, available in Open Beta

Build AI with Watson and deploy anywhere with Cloud Pak for Data

AI lifecycle automation
Data and AI governance

Watson Knowledge Catalog

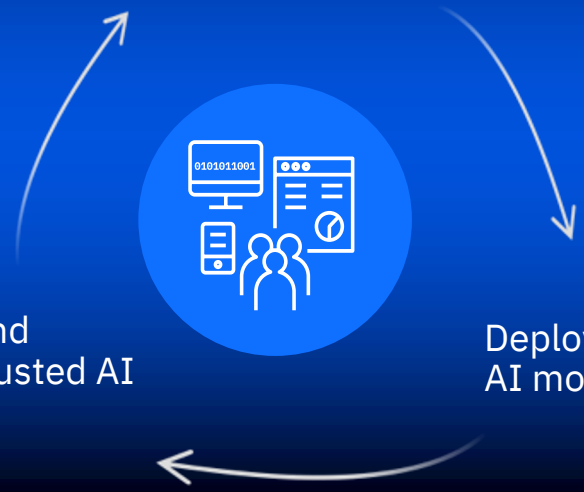
Organize

Catalog and prepare data

Watson Studio

Build, run and manage

Build and train AI models



Manage and operate trusted AI

Deploy and run AI models

Available visually or programmatically

Watson Studio Premium

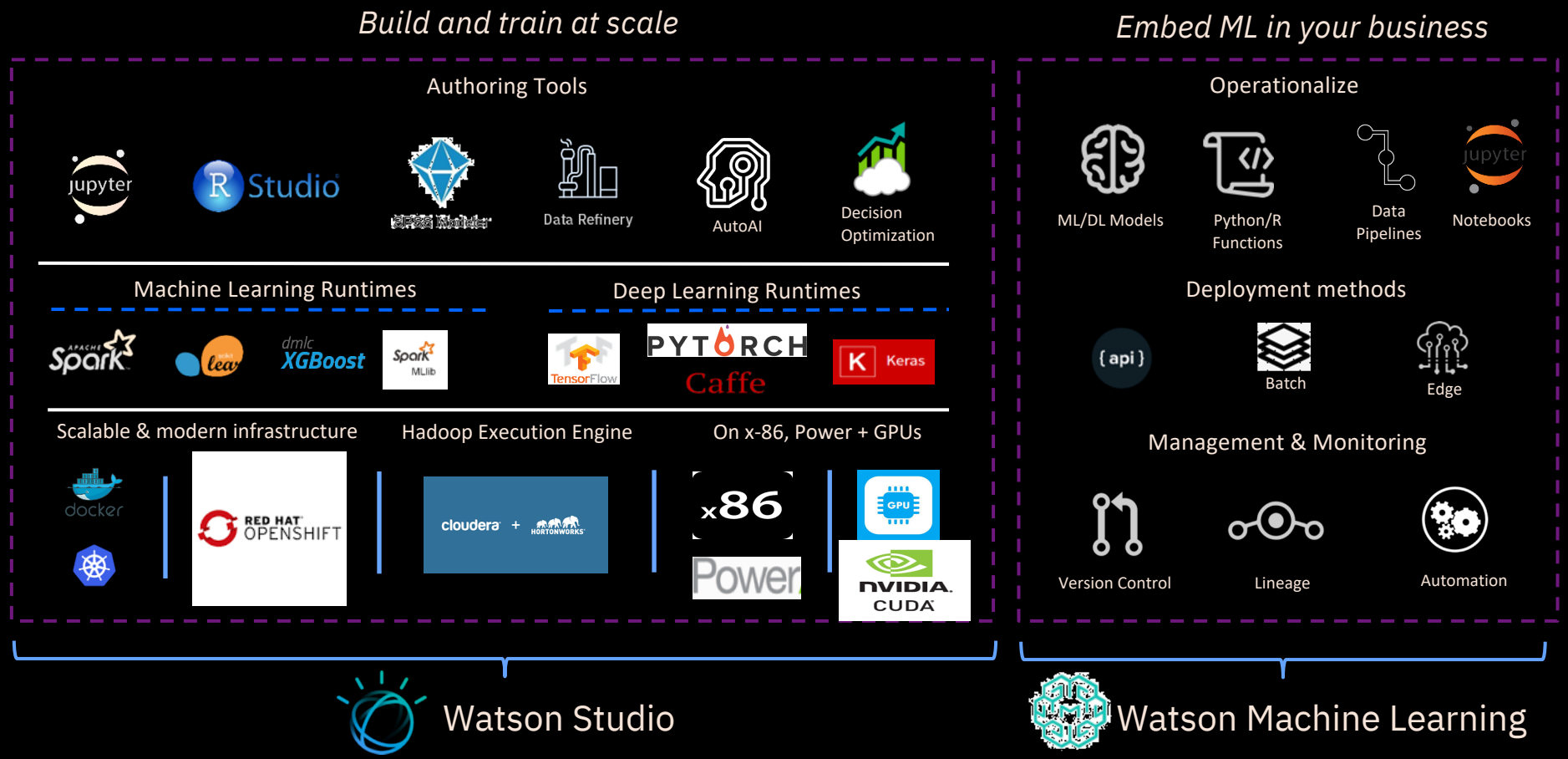
Optimize

Optimize and simplify decisions

Discovery | Prediction | Outcomes

Cloud Pak for Data

Watson Studio and Watson Machine Learning inject AI firepower into your business



IBM SOFTLAYER · On-prem · amazon web services · Azure

Mix and Match your deployment

- ✓ Cloud – IBM Cloud, Azure, AWS
- ✓ On Premise / Private Data center
- ✓ Desktop

IBM Watson Studio

Enterprise Data Science platform that helps your team work together to build models to make better data driven decisions for your business

Analyze any data, no matter where it lives

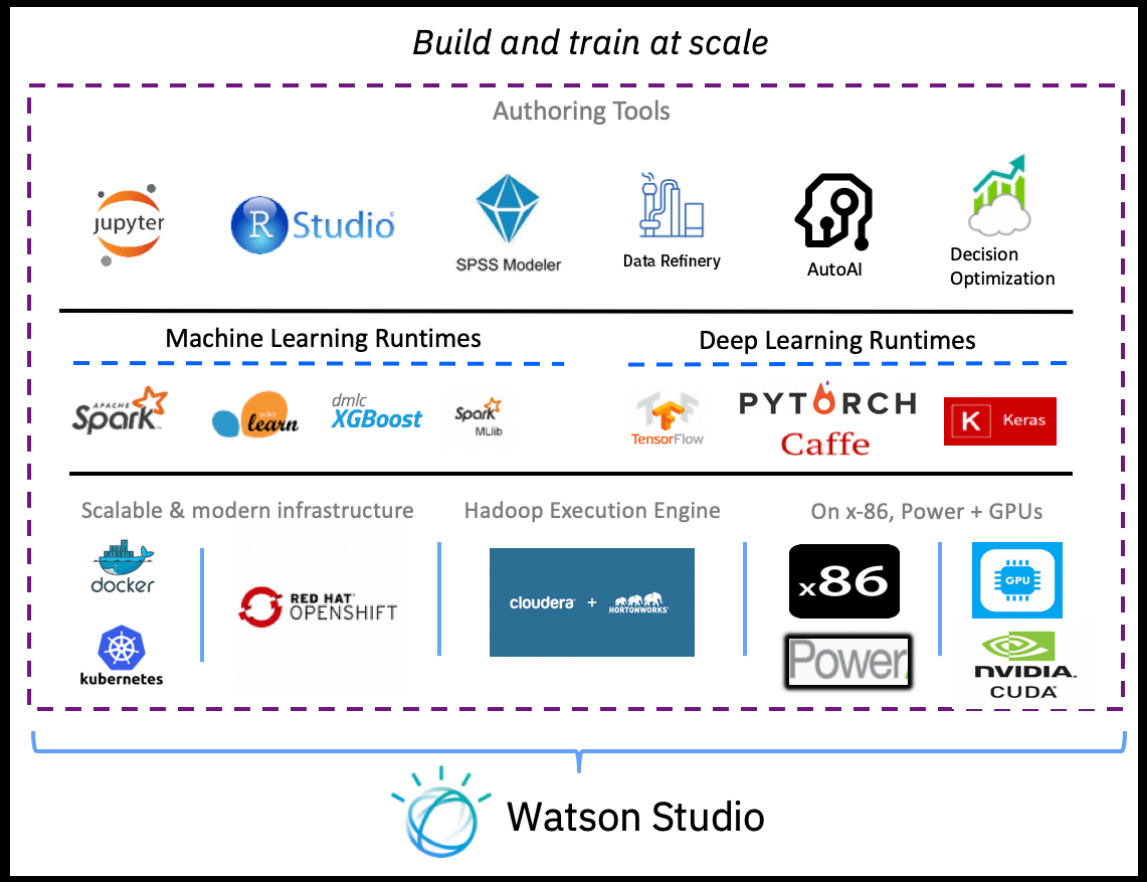
Connect to and analyze your data without moving a single byte through dozens of connectors and multiple deployment options

Empower your entire organization with notebooks, visual productivity, and automation tools

Leverage your entire organization with a variety of tools in a single integrated platform

One platform to rule them all from discovery to production

Analyze data, build predictive models, and seamlessly integrate Watson Machine Learning to deploy at scale



IBM Watson Machine Learning

Embed Machine Learning and Deep Learning in your Business

Deploy and Manage Models

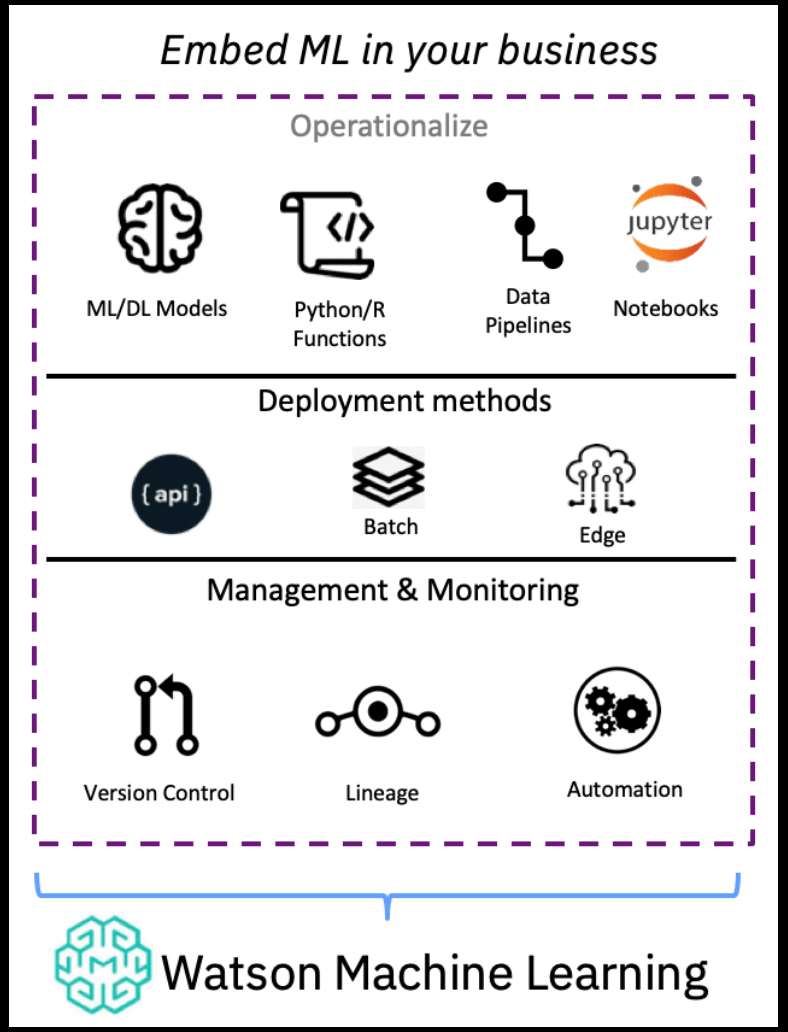
Move models to production, in an easy, secure, and compliant way

Intelligent Model Operations

Embed intelligent training services, with feedback loops that constantly learn from new data, regardless where it resides

Accelerate Compute Intensive Workloads

Distribute your deep learning training and Hadoop/Spark workloads with multi-tenant job scheduling



Do you trust your AI?

BlackRock shelves unexplainable AI liquidity models

Risk USA: Neural nets beat other models in tests, but results could not be explained

YouTube sued for using AI to racially profile content creators

They claim YouTube's algorithms discriminate against black users

Uber hit with lawsuit to reveal how its algorithm works

Sections

The Washington Post

Democracy Dies in Darkness

Get 1 year for $29

Apple Card algorithm sparks gender bias allegations against Goldman Sachs

RETAIL OCTOBER 10, 2018 / 4:04 PM / UPDATED 2 YEARS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

EFF to HUD: Algorithms Are No Excuse for Discrimination

BY JAMIE WILLIAMS, SAIRA HUSSAIN, AND JEREMY GILLULA | SEPTEMBER 26, 2019

Watson OpenScale

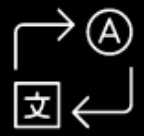
Validate and monitor AI models, deployed anywhere, to help comply with regulations, address internal safeguards, and mitigate business risk

Monitoring for compliance and safeguards

Mitigate biased model behavior

Explain model decisions

Validate and control risk



Ensure that models are resilient to changing situations

Detect drift during runtime

Generate specific model retraining inputs

Align model performance with business outcomes

Correlate model metrics and business KPIs

Actionable metrics and alerts

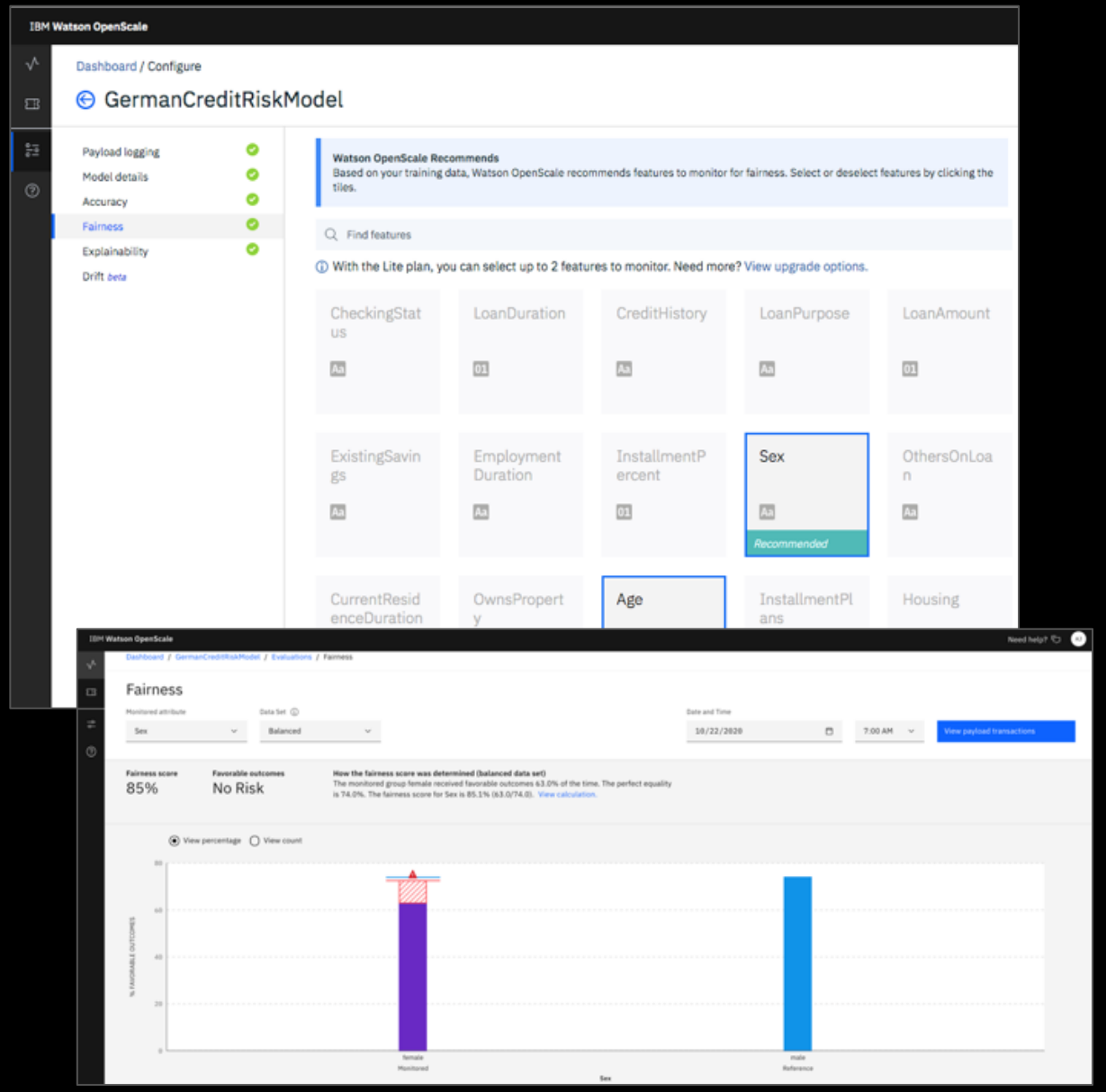
Bias Detection

OpenScale enables enterprises to enforce fairness in their model's outcome by analyzing transactions in production and finding biased behavior by the model

It pinpoints the source of bias and actively mitigates the biases found in production environment

Value:

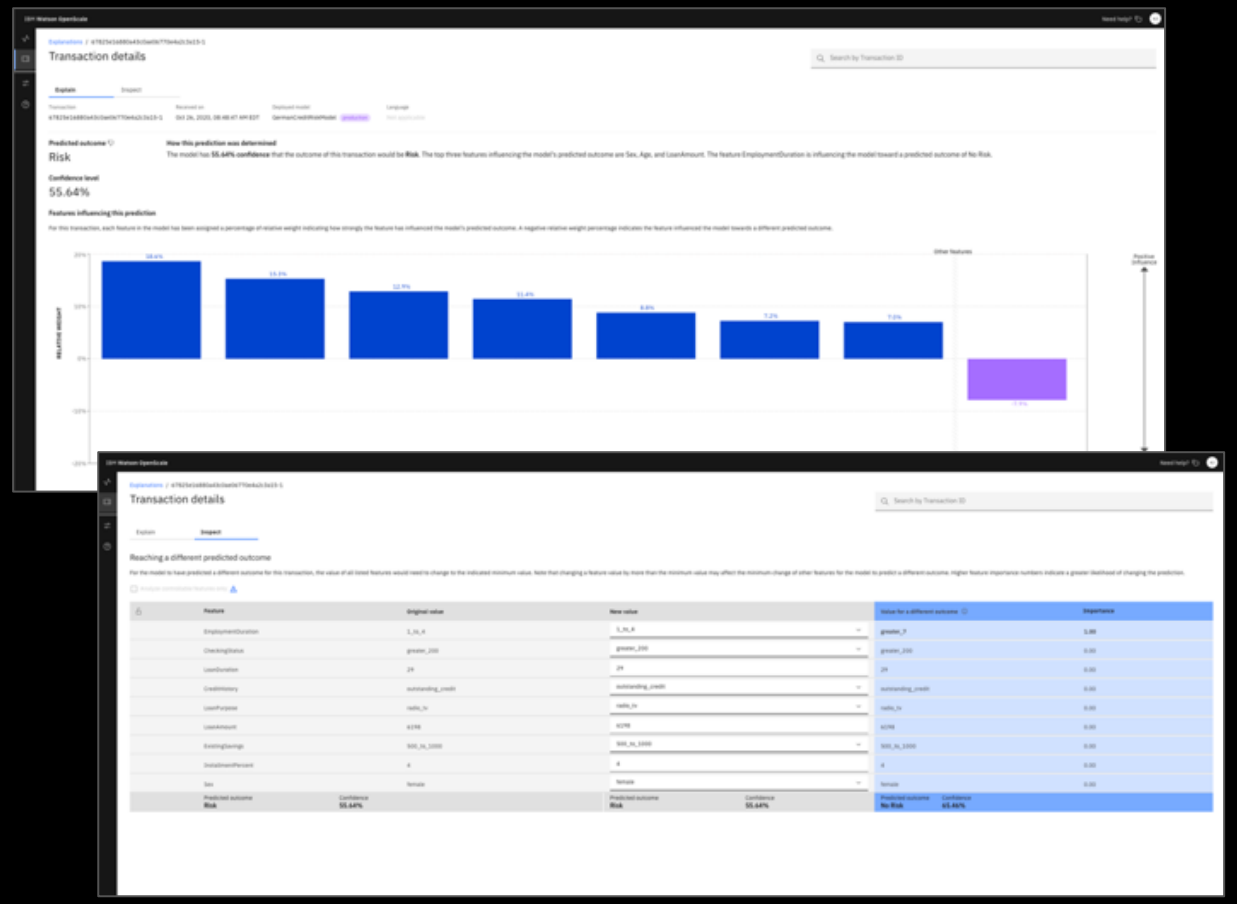
- Automatically recommend common protected attributes to monitor during production
- Detect biases in runtime in order to catch impacts on business applications and compliance requirements without time consuming, manual data analysis
- Metrics and data to help data scientists further troubleshoot issues in data sets or models
- Mitigate biases in runtime in order to enforce regulatory or enterprise fairness guardrails in real time



Explainability

OpenScale records every individual transaction and drills down into its working to explain how the model makes decisions

It provides a simple explanation that is user friendly and interactive



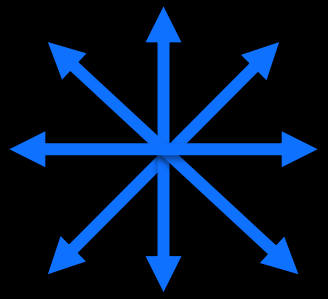
Value:

- Explain individual transaction level decisions made by the model in run time, including details about most important attributes and their values in order to assist in compliance and customer care situations
- Analyze individual transactions in a what-if manner in order to understand how model behavior will change in different business situations

Trusted AI Lifecycle through Open Source

Pillars of trust, woven into the lifecycle of an AI application

Did anyone tamper with it?



ROBUSTNESS

Adversarial Robustness 360

↳ (ART)

github.com/IBM/adversarial-robustness-toolbox

art-demo.mybluemix.net

Is it fair?

FAIRNESS

AI Fairness 360

↳ (AIF360)

github.com/IBM/AIF360

aif360.mybluemix.net

Is it easy to understand?

EXPLAINABILITY

AI Explainability 360

↳ (AIX360)

- github.com/IBM/AIX360

aix360.mybluemix.net

Using Watson OpenScale & Toolkits Together

Watson OpenScale

Online Post Processing
(Intercept & Fix Features)

AI Fairness 360

In-Processing
(Intercept & Fix Algorithm)

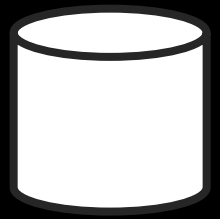
AI Fairness 360

Pre-Processing
(Intercept & Fix Training Data)

Runtime

Algorithm Design

Data Selection

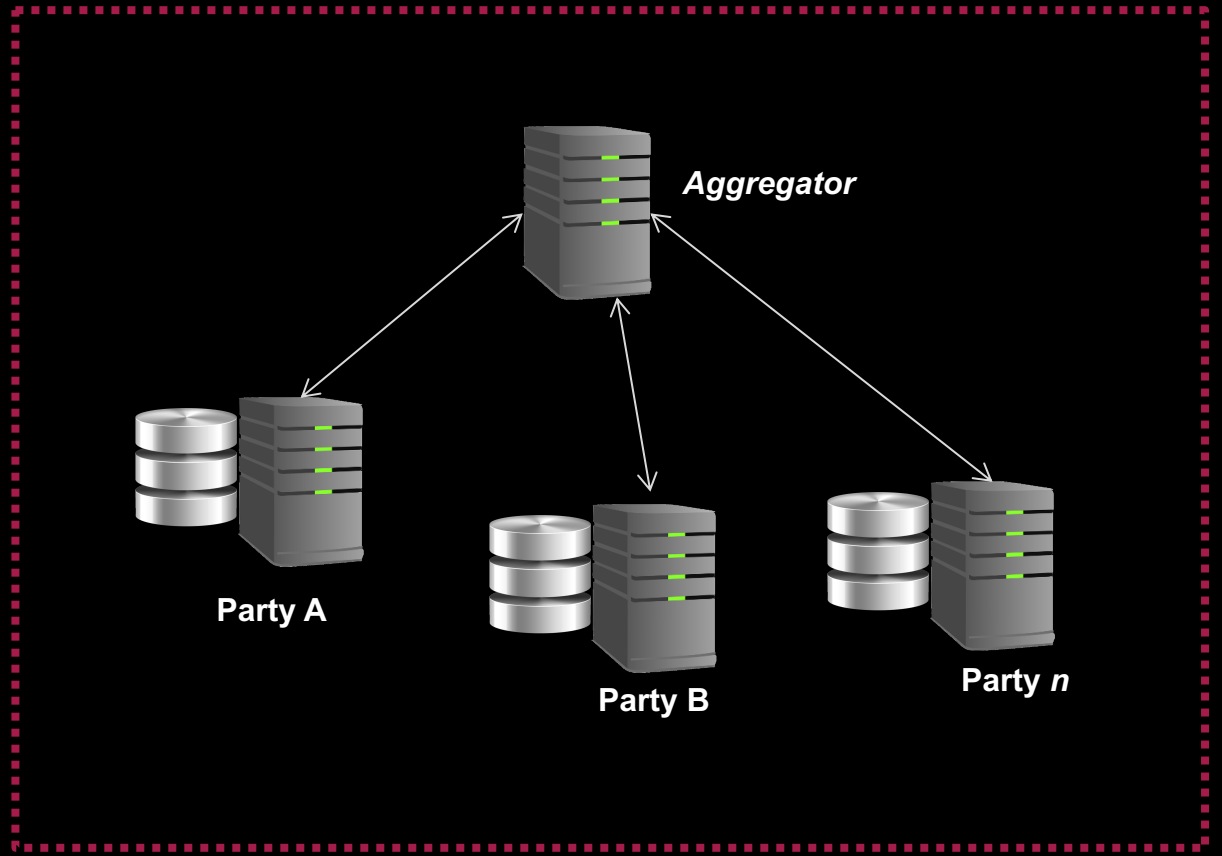


Federated Learning in 3 Steps

1. Start with a ML model

2. Configure *n* number of Parties to distribute model along with retraining ability

3. Parties collaboratively train a ML model by returning information to Aggregator while keeping training data to themselves



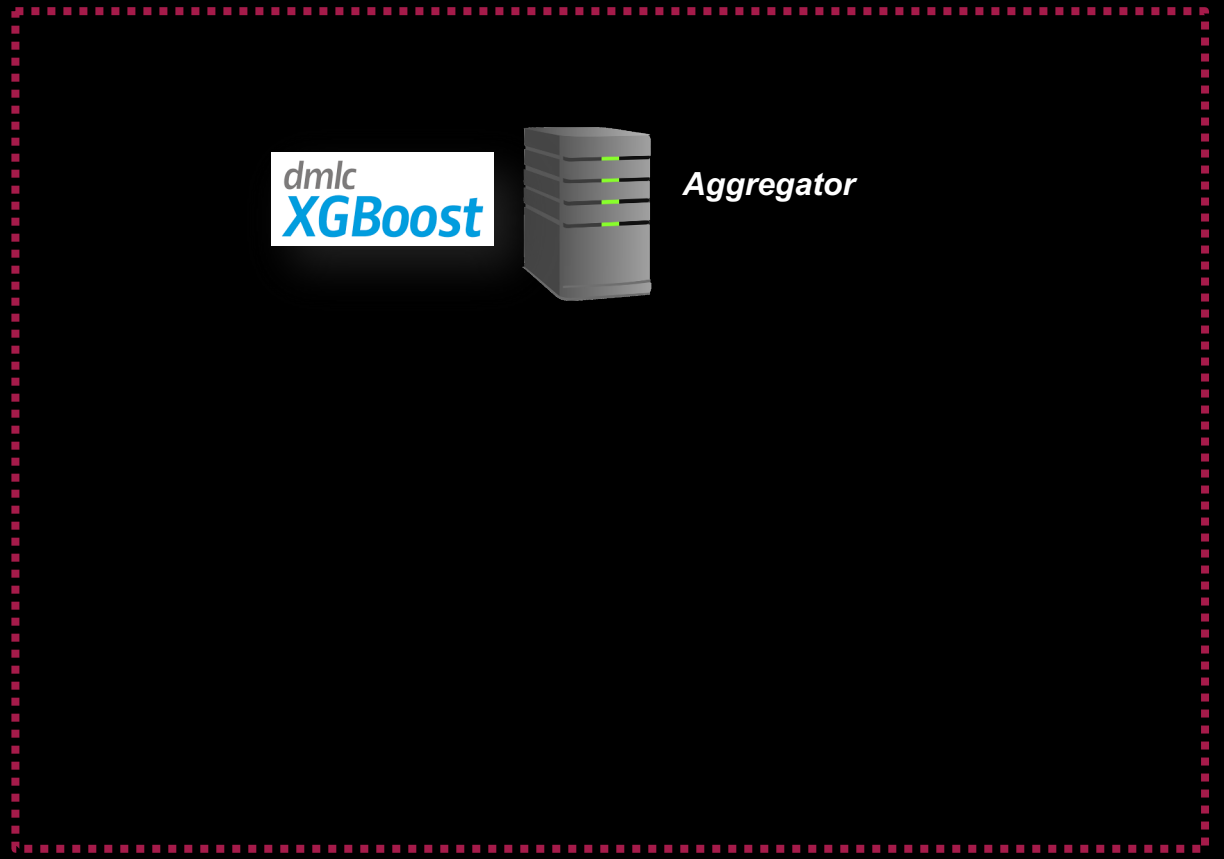
IBM

Federated Learning in 3 Steps

1. Start with a ML model

2. Configure *n* number of Parties to distribute model along with retraining ability

3. Parties collaboratively train a ML model by returning information to Aggregator while keeping training data to themselves

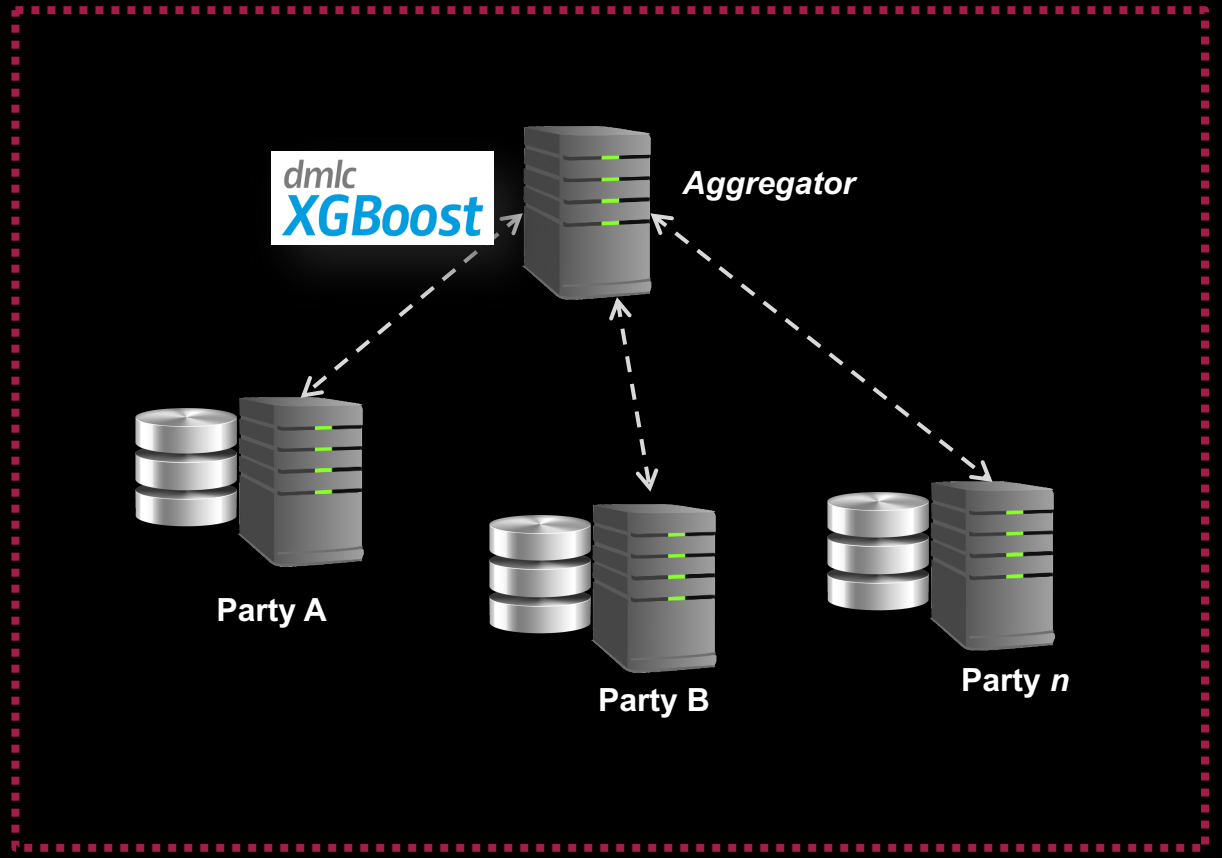


Federated Learning in 3 Steps

1. Start with a ML model

2. Configure *n* number of Parties to distribute model along with retraining ability

3. Parties collaboratively train a ML model by returning information to Aggregator while keeping training data to themselves



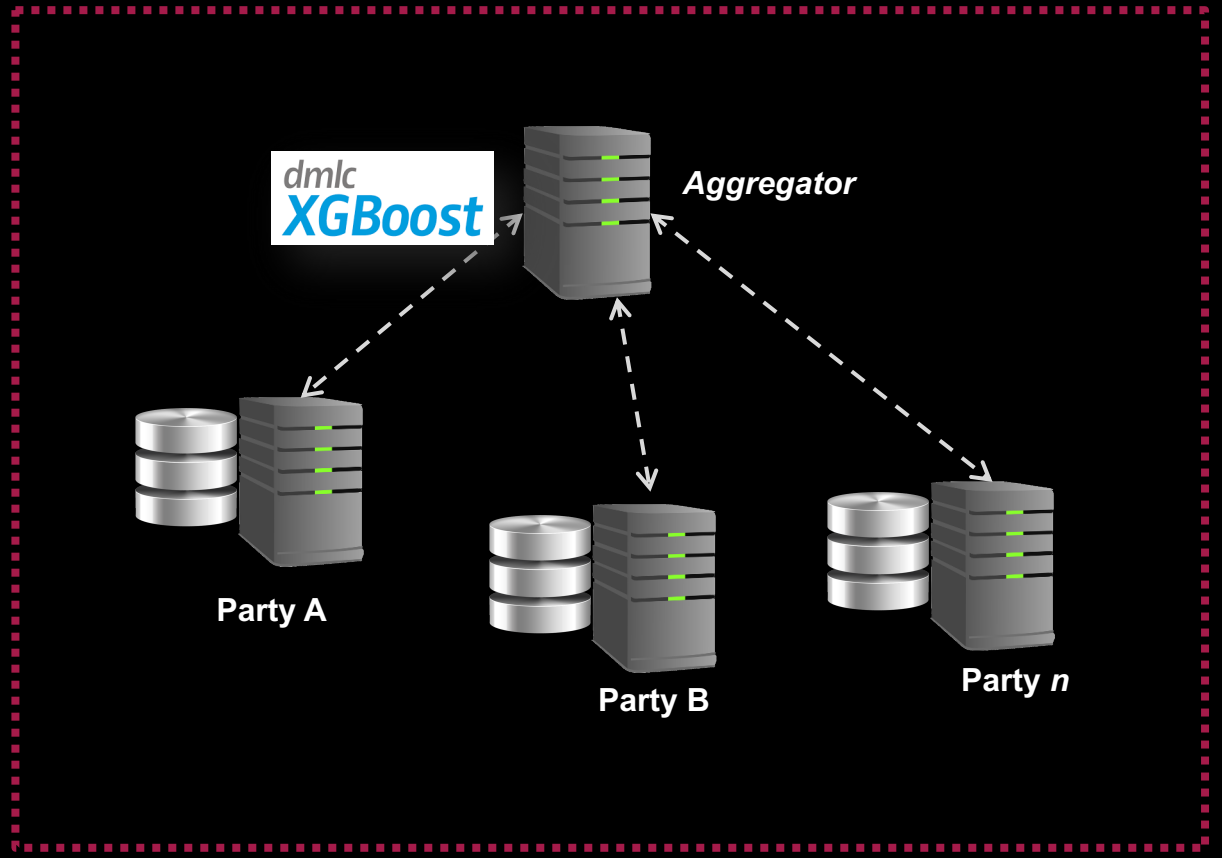
IBM

Federated Learning in 3 Steps

1. Start with a ML model

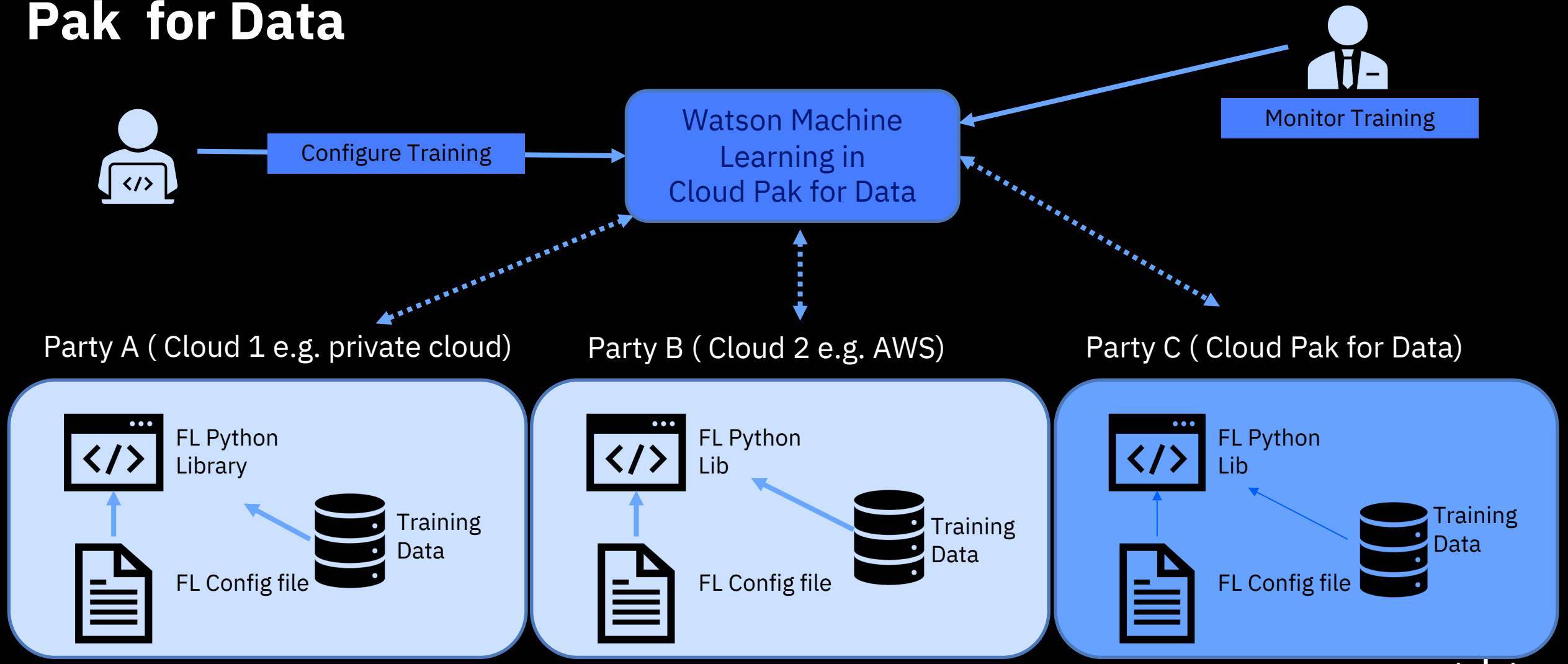
2. Configure *n* number of Parties to distribute model along with retraining ability

3. Parties collaboratively train a ML model by returning information to Aggregator while keeping training data to themselves



IBM

Federated Learning in Watson Machine Learning for Cloud Pak for Data



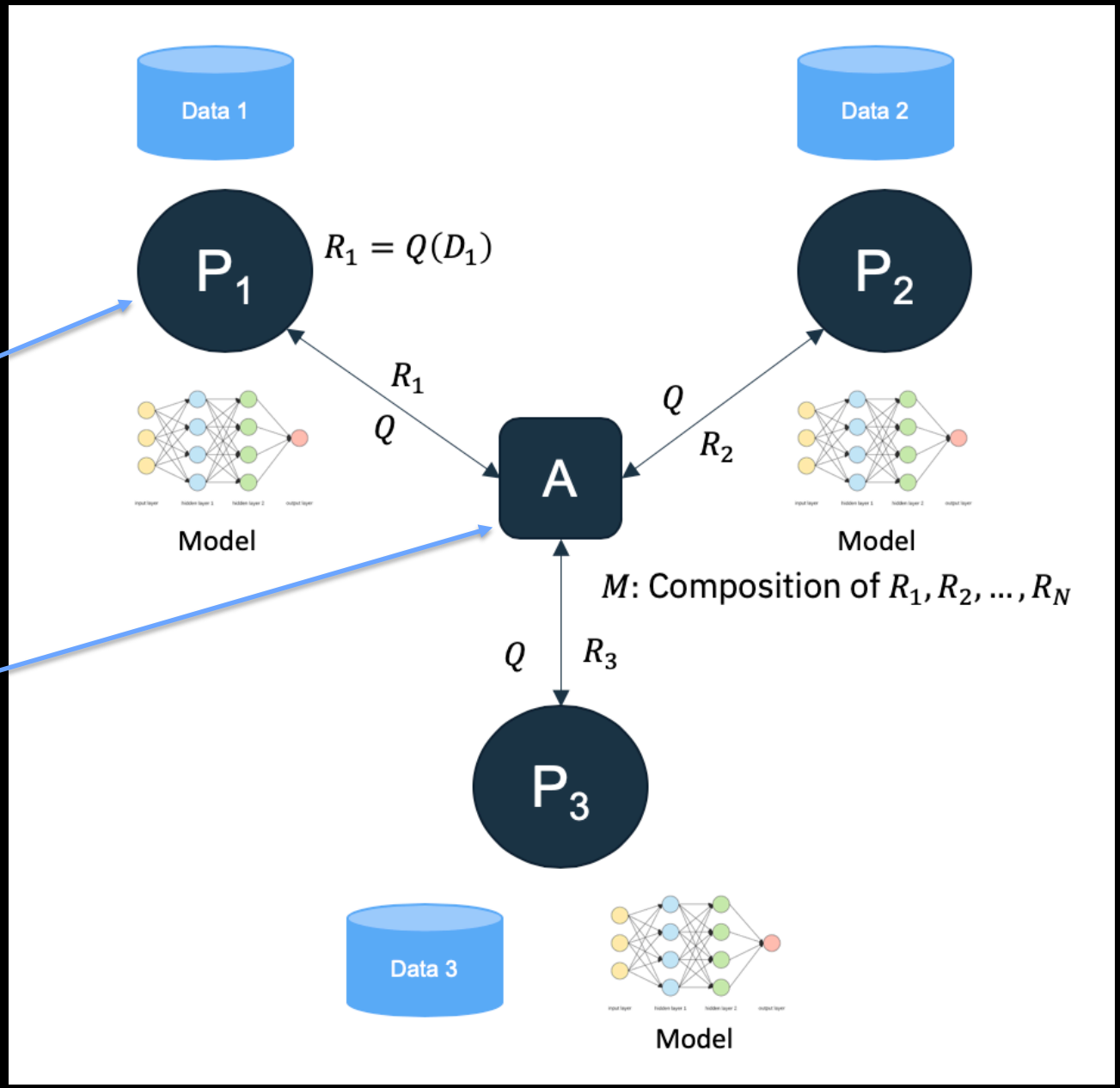
Appendix

Getting back to this …

How do you do this?

Local Training

Fusion



Different Learning Paradigms in Federated Learning

Choosing *how* and *where* the *learning occurs*.

Every ML algorithm falls into one of the three following categories defined by:

1. Model Structure

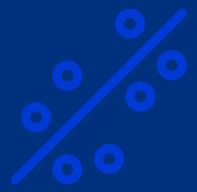
2. Learning Process (i.e. Loss)

3. Location of Model Updates

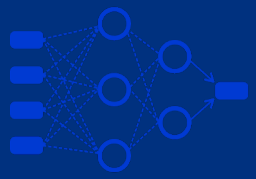
Static Model

- Predefined fixed model architecture.
- Learning (i.e. model weight updates) occurs at party side.
- Fusion of the models occurs at the aggregator.

Examples:



Linear Models



Neural Networks

Hybrid Model

- Fixed loss structure, with dynamic architecture.
- Learning process occurs in both aggregator and party side.
- Parties perform loss-based computation (i.e. compute gradients, Hessians, etc).

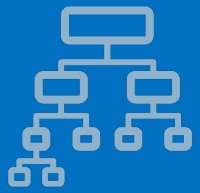
Examples:

XGBoost

Dynamic Model

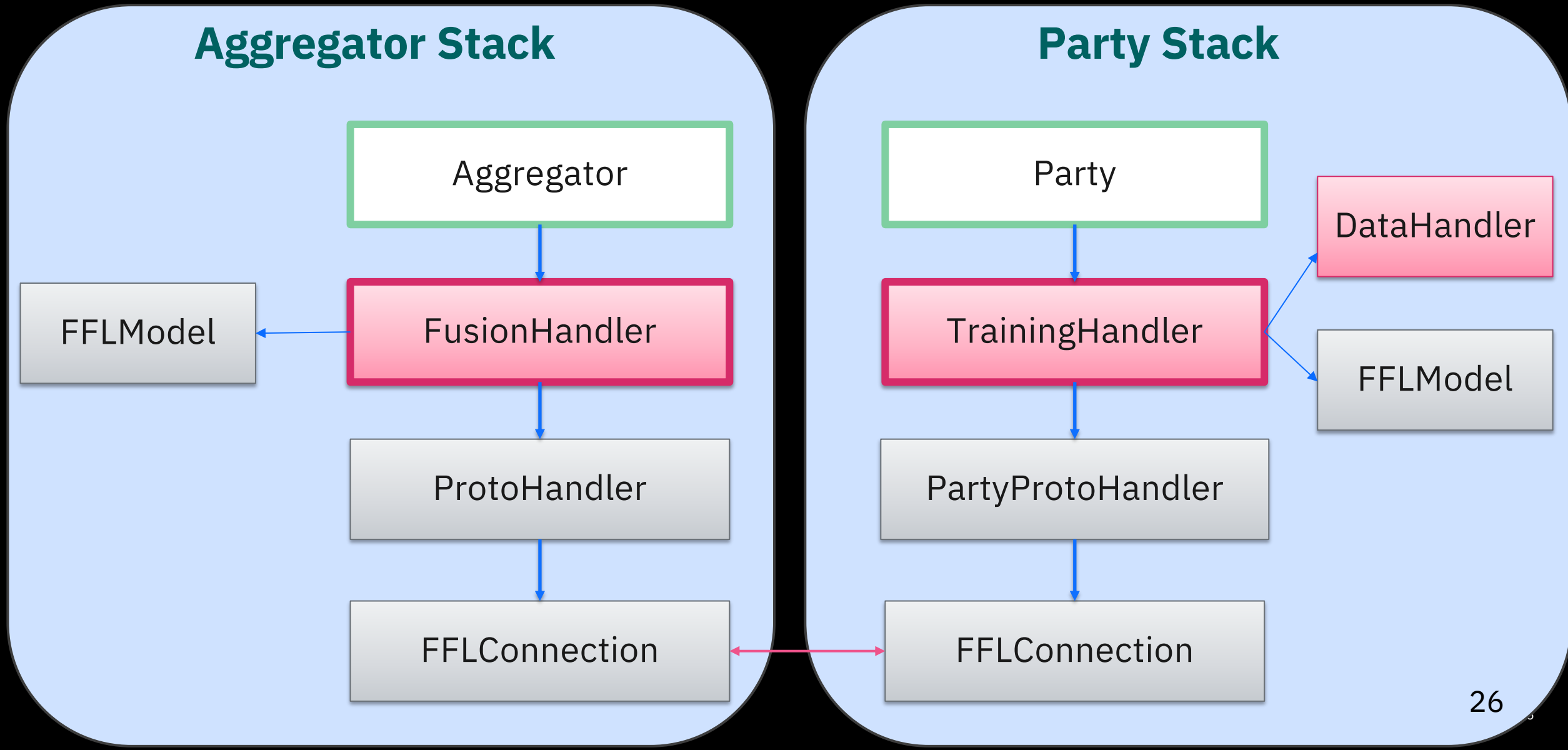
- Predefined dynamic model architecture.
- Learning (i.e. tree growth) occurs at aggregator.
- Parties perform very simple queries (i.e. get simple count values).

Examples:



Decision Trees

A look under the hood



Advanced Fusion Algorithms

Fusion algorithms combine neural networks from different parties

State of the art: model averaging

- Works for all network topologies
- Not very performant

Approach: fusion as a matching problem

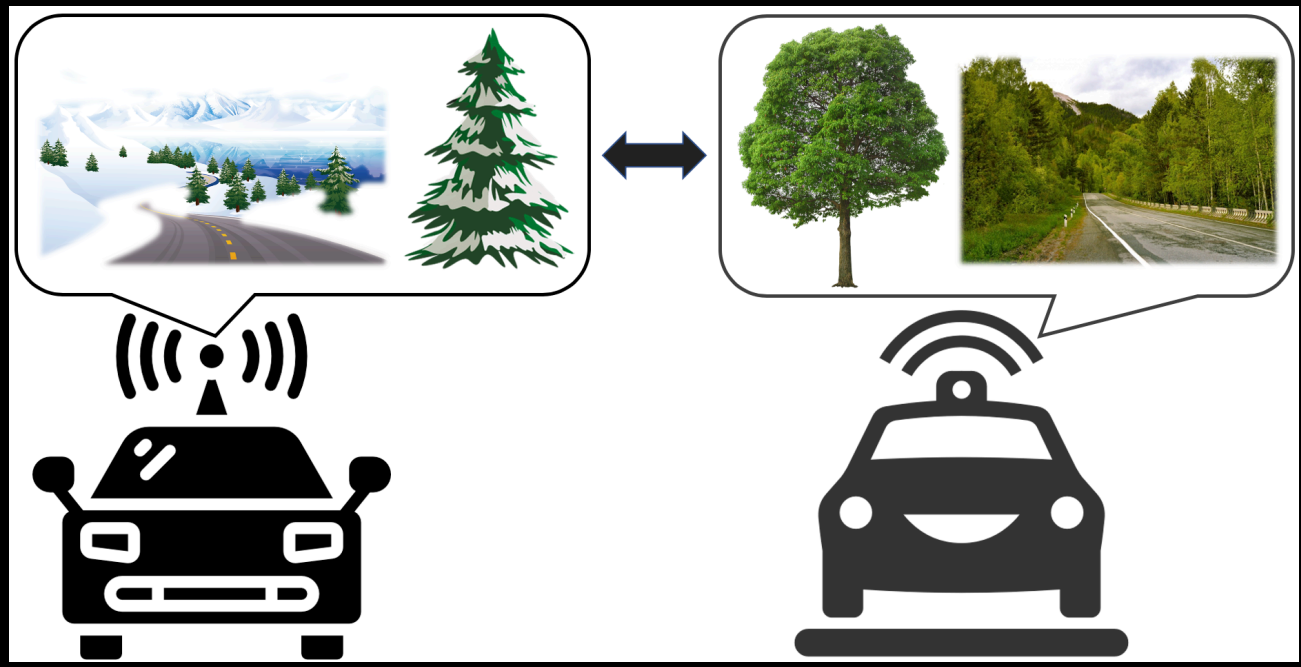
Probabilistic Federated Neural Matching (**PFNM**)

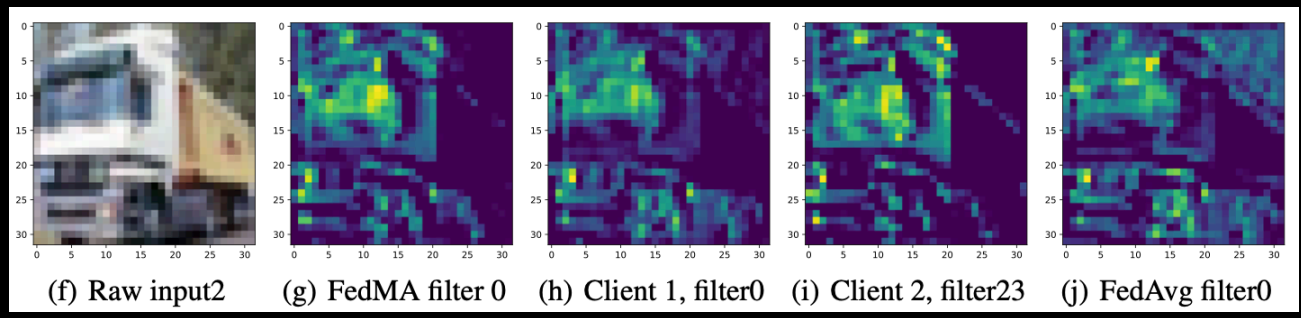
Effective algorithms leverage neural network structure for training efficiency

One-time fusion

Also for unsupervised training (SPAHM)!

Joint work: MIT-IBM Lab, U Michigan, U Wisconsin @ Madison, IBM Research





H. Wang, M. Yurochkin, Y. Sun, D. Papailiopoulos, and Y. Khazaeni. Federated learning with matched averaging. ICLR 2020
M. Yurochkin, M. Agarwal, S. Ghosh, K. Greenewald, N. Hoang, and Y. Khazaeni. Bayesian nonparametric federated learning of neural networks. ICML 2019

Heterogeneity in Federated Learning Performance

Federated learning taps into data sources in different Clouds, data centers or app silos

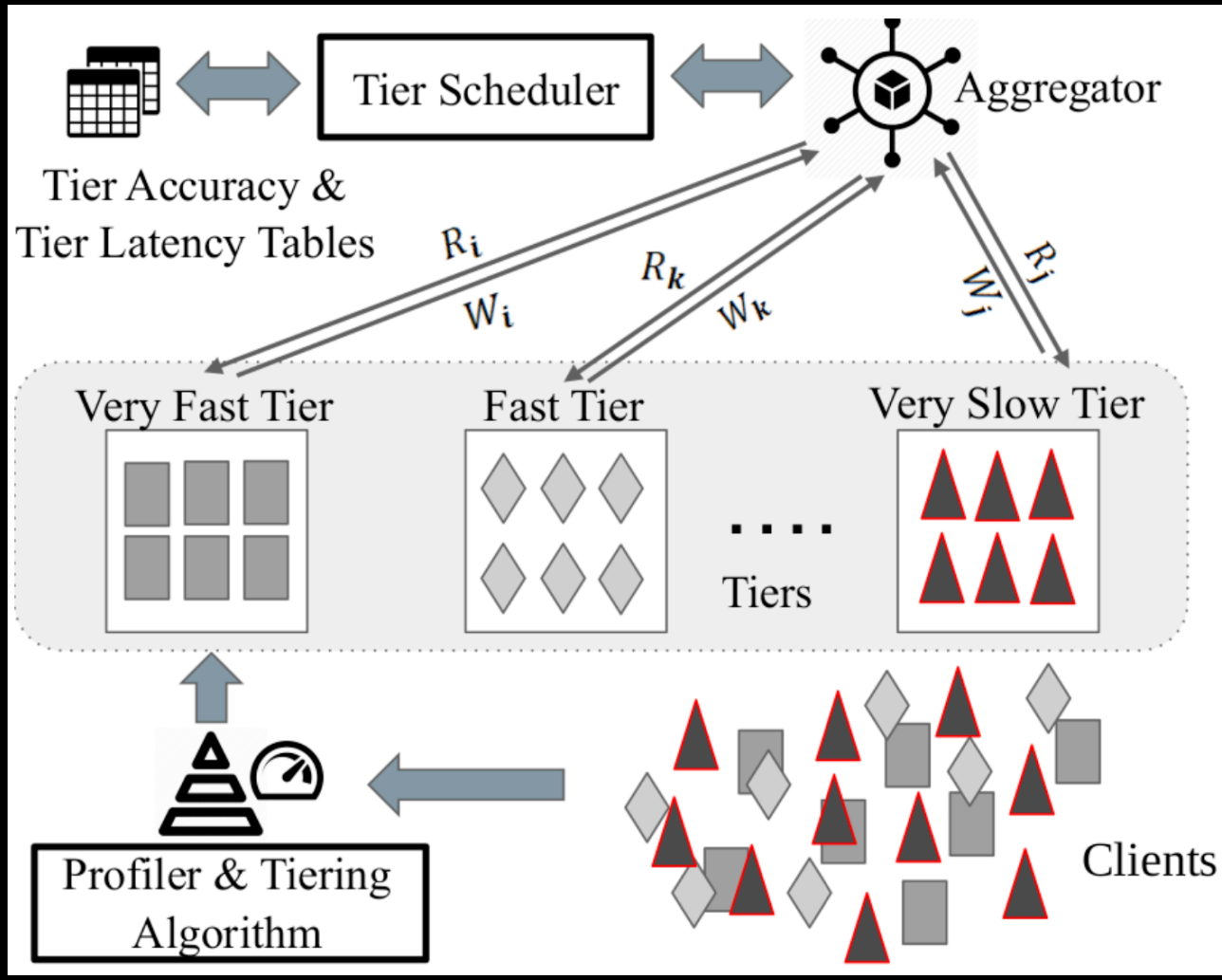
Parties are different in

- Data quantity
- Data distribution
- Training capability

That's a good thing!!

TiFL: Profiling and tiering manages heterogeneity

Joint work: George Mason, U Nevada@Reno, Georgia Tech, IBM Research



TiFL: A Tier-based Federated Learning System: Z. Chai, A. Ali, S. Zawad, S. Truex, A. Anwar, N. Baracaldo, Y. Zhou , H. Ludwig , F. Yan , Y. Cheng, accepted at ACM HPDC - HotCRP 2020.

Effective Privacy in Federated Learning

Privacy regulation:

HIPAA Health Insurance Portability & Accountability Act

General Data Protection Regulation

CALIFORNIA CONSUMER PRIVACY ACT

Problem: Training data might be inferred from DNN model

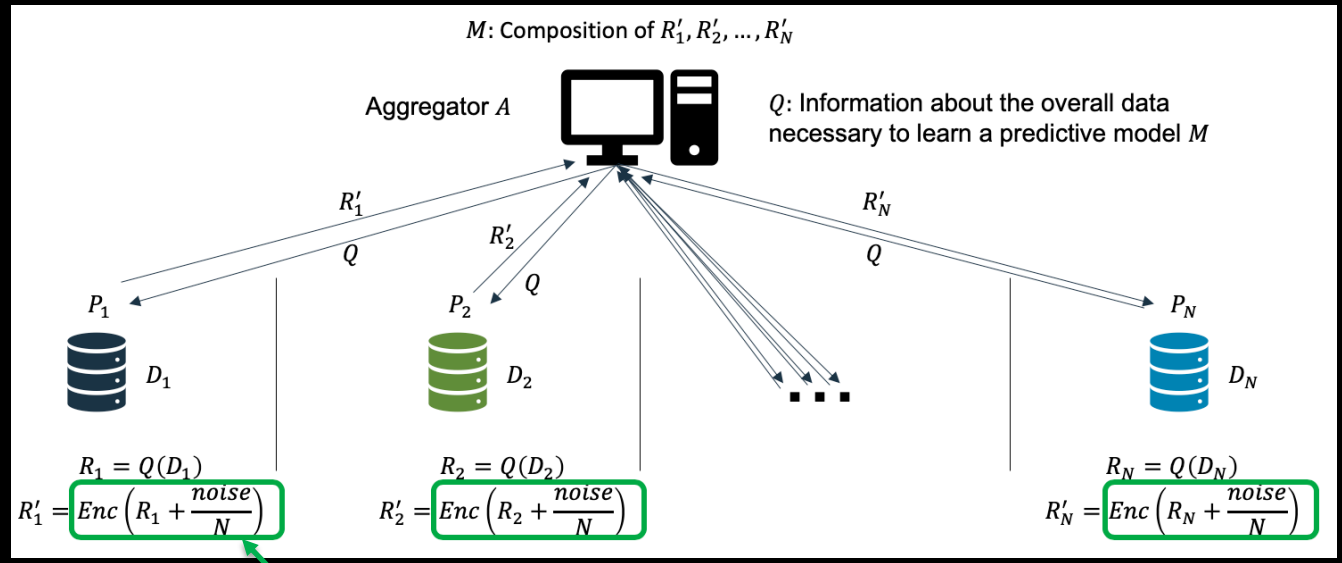
Simple solution: adding (differentially private) noise

Penalty: reduced model performance

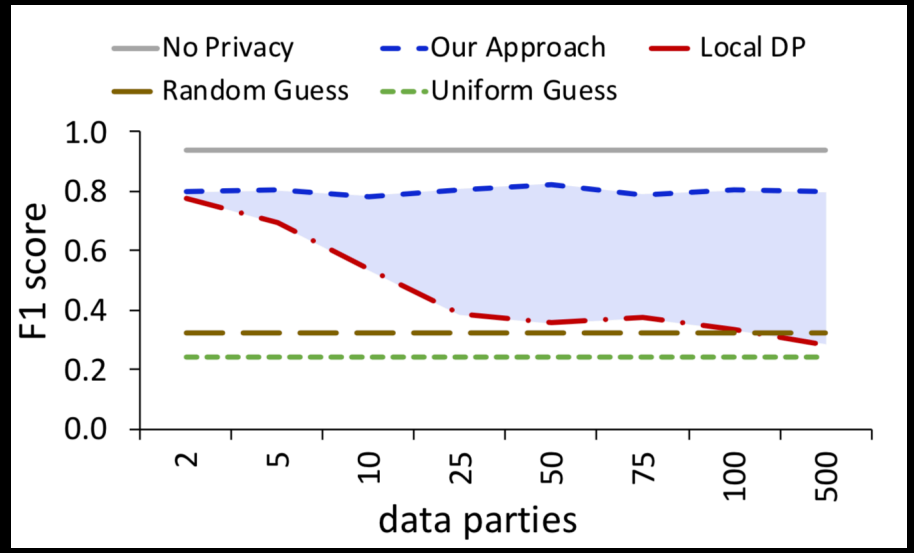
Better solutions: Hybrid One

Differential Privacy + Secure Multi-party Computation for private and performant FL

Based on Threshold Pallier Scheme



Reduces total noise from noise*N to noise

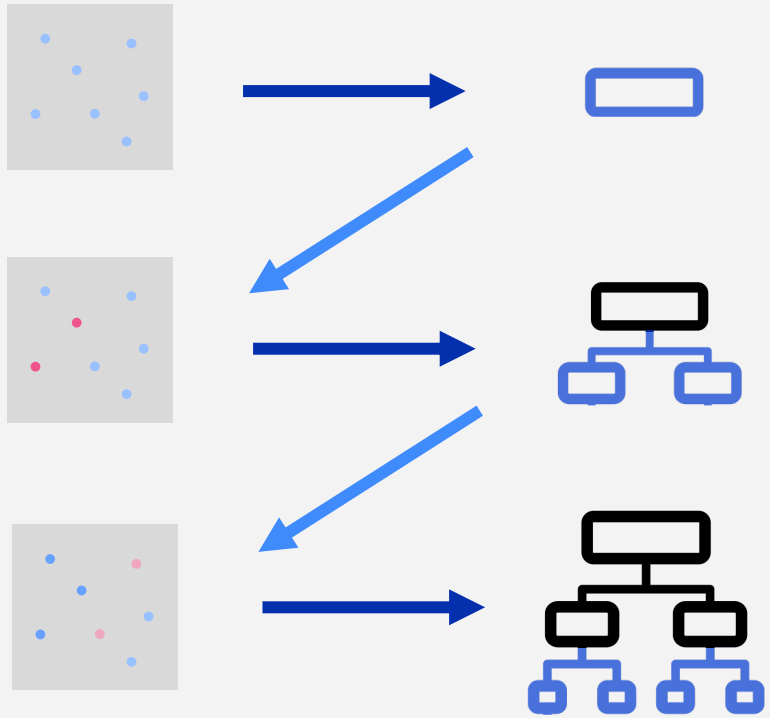


S. Truex, N. Baracaldo, A. Anwar, T. Steinke, H. Ludwig, R. Zhang, Y. Zhou: A hybrid approach to privacy-preserving federated learning, ACM AISec 2019 – Best Paper Award

Extreme Gradient Boosting (XGBoost)

- XGBoost is a decision tree-based ensemble method which utilizes a gradient-boosting based approach for optimizing against the loss function.
- Gradient Boosting methods have demonstrated state-of-the-art performance in various supervised tasks.
- Highly utilized in various settings such as classification, regression, and ranking based problems.
- Recently popular among the Kaggle community for its use in various machine learning competitions.

XGBoost kaggle



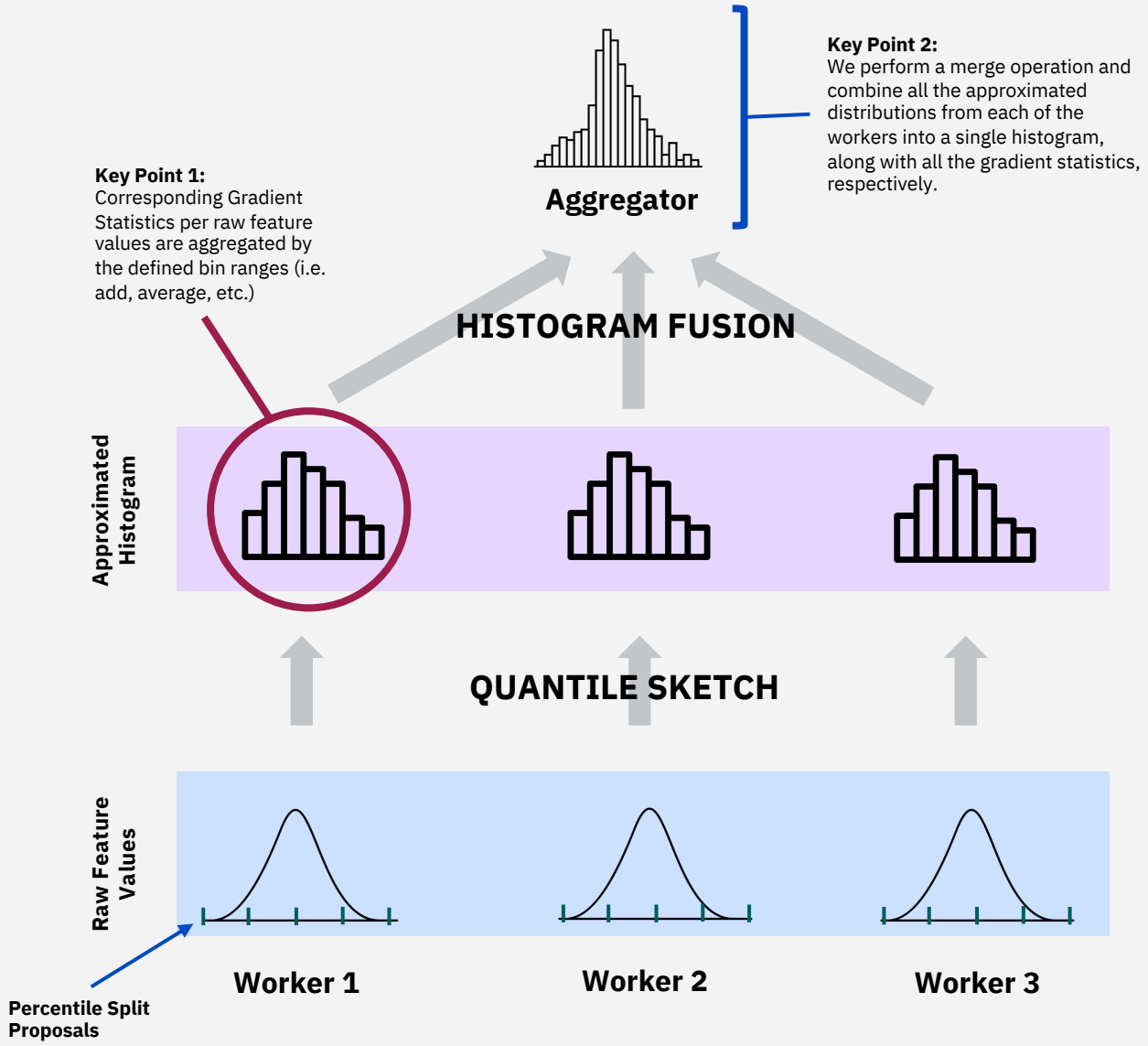
Fusion Strategy: Federated Quantile Sketch

We exploit XGBoost's *approximated histogram-based method* to introduce a **federated approach** for Gradient Boosting in FFL.

Federated Quantile Sketch is a novel fusion method which entails the following two key steps:

1. **Histogram Approximation Policy**
   Intelligently choosing the right parameters (i.e. histogram bin size) to strike a balance between **protecting privacy** while **retaining high model performance** for building our surrogate data histograms.

2. **Aggregated Histogram Fusion**
   Depending on the learning task at hand (i.e. classification vs regression), we must choose the right method for how to combine our surrogate data histograms at the aggregator.



XGBoost in Federated Learning

REPEAT PROCES UNTIL **MAX ROUNDS**. CHOOSE BEST MODEL FROM ALL ROUNDS

Aggregator (A)

STEP 1: Initialize Aggregator Model $f_0^{(A)}$ and obtain data set size from each party.

STEP 2: Send current model and epsilon parameter from Aggregator.

$f^{(A)}, \epsilon$

STEP 3: Using given base aggregator model ($f^{(A)}$), compute predictions on local participant's dataset. Compute the gradient and hessian statistics of the prediction.

Perform quantile sketch on feature values using ϵ parameter. Aggregate gradient statistics with respect to the binned split value proposals.

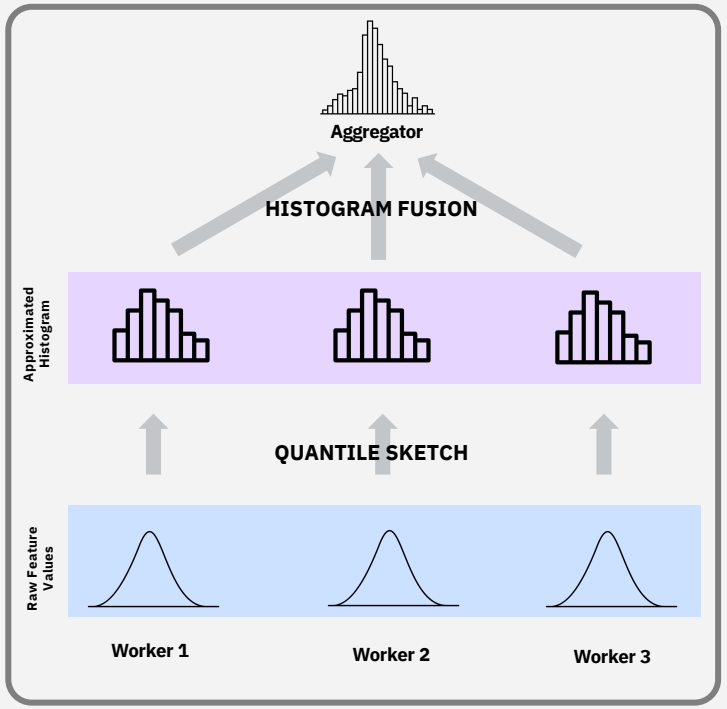
STEP 4: Send back corresponding (approximate) split value, gradient and hessian statistics from each candidate.

X, G, H

STEP 5: Merge aggregated values as single distribution and merge gradient statistics accordingly.

STEP 6: Send back split proposal, gradient, and hessian statistics. Perform split finding and build model recursively.

Aggregated Split Finding Process



P_1 D_1 | P_2 D_2 | ... | P_N D_N

X, G, H

Participants (P)

Federated Reinforcement Learning

Reinforcement learning:

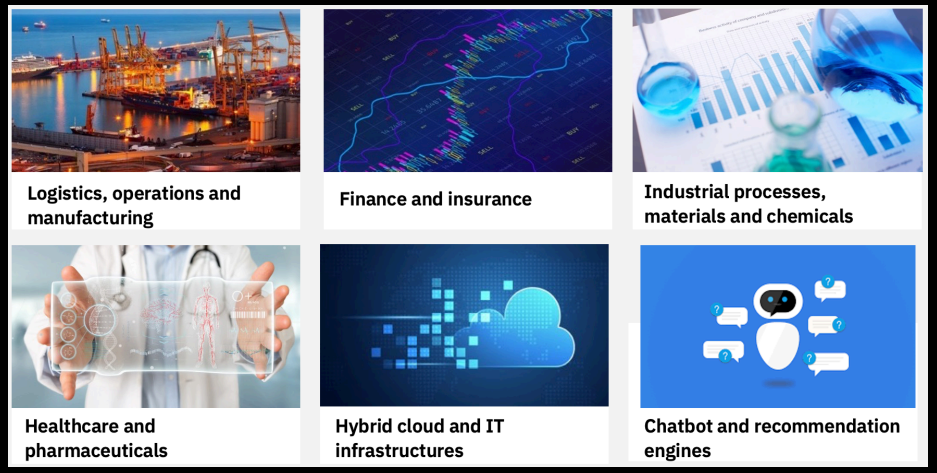
Learning from decision outcomes

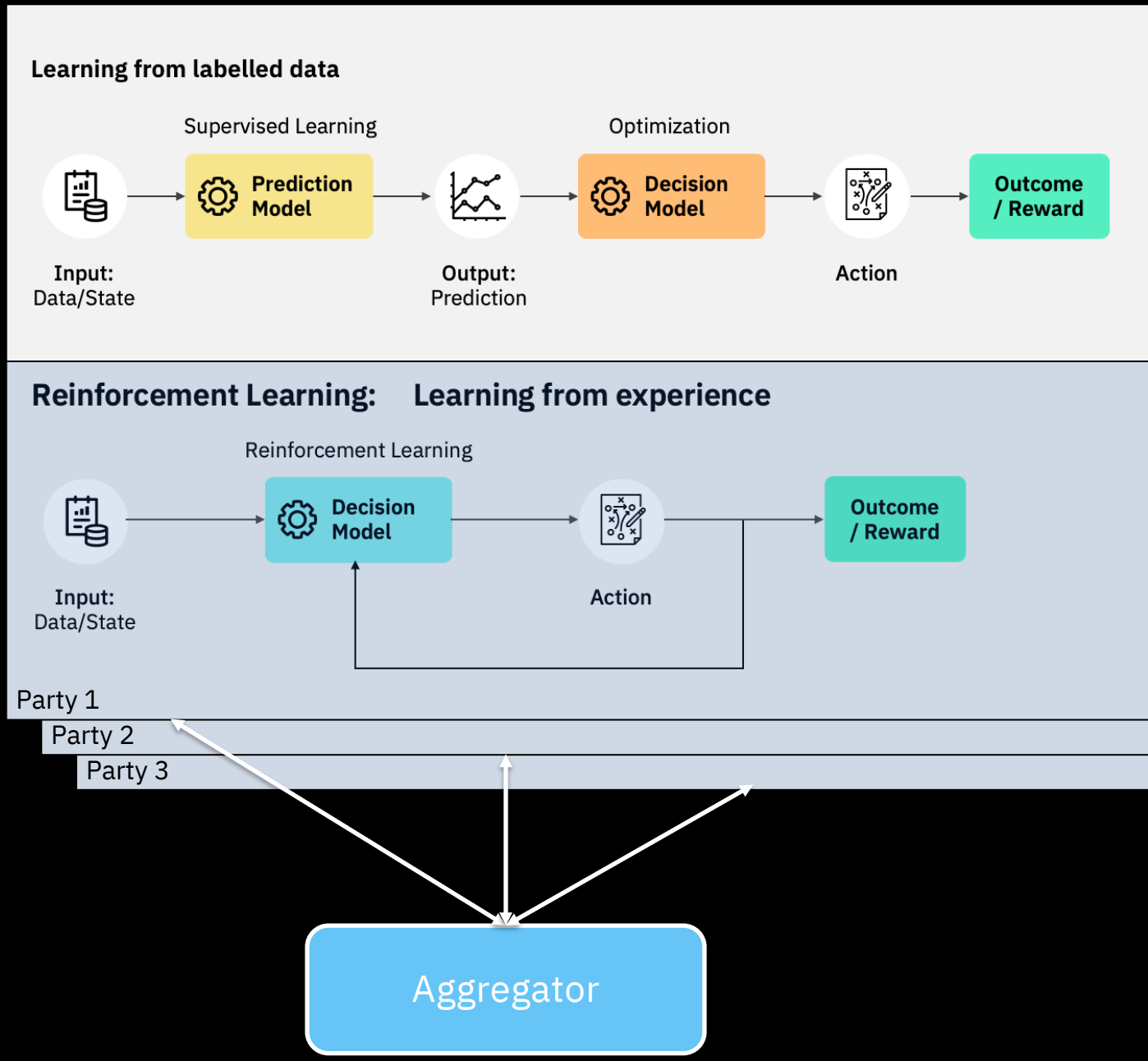
Approach:

- Decision model trained in parties with local data
- Models merged in aggregator

Uses RLLib or custom RL library

Application domains:





S. Ghosh, S. Laguna, S. Hong Lim, L. Wynter, H. Poonawala: A Deep Ensemble Multi-Agent Reinforcement Learning Approach for Air Traffic Control; arxiv