SNIA CSTI | CLOUD STORAGE TECHNOLOGIES

Why Distributed Edge Data is the Future of AI

Live Webinar

October 3, 2023

10:00 am PT / 1:00 pm ET

Today's Presenters

Erin Farr
Vice Chair SNIA Cloud Storage Technologies Initiative
Storage CTO Office, IBM

Rita Wouhaybi
Senior AI Principal Engineer
Office of the CTO, Network & Edge Group
Intel

Heiko Ludwig
Senior Manager
AI Platforms
IBM Research

The SNIA Community

200
Corporations, universities, startups, and individuals

2,500
Active contributing members

50,000
Worldwide IT end users and professionals

SNIA CSTI | CLOUD STORAGE TECHNOLOGIES

What We Do

Educate vendors and users on cloud storage, data services and orchestration

Support & promote business models and architectures: OpenStack, Software Defined Storage, Kubernetes, Object Storage

Understand Hyperscaler requirements
Incorporate them into standards and programs

Collaborate with other industry associations

SNIA Legal Notice

- The material contained in this presentation is copyrighted by SNIA unless otherwise noted.
- Member companies and individual members may use this material in presentations and literature under the following conditions:
  - Any slide or slides used must be reproduced in their entirety without modification
  - SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of SNIA.
- Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.
- The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

  NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.

Agenda

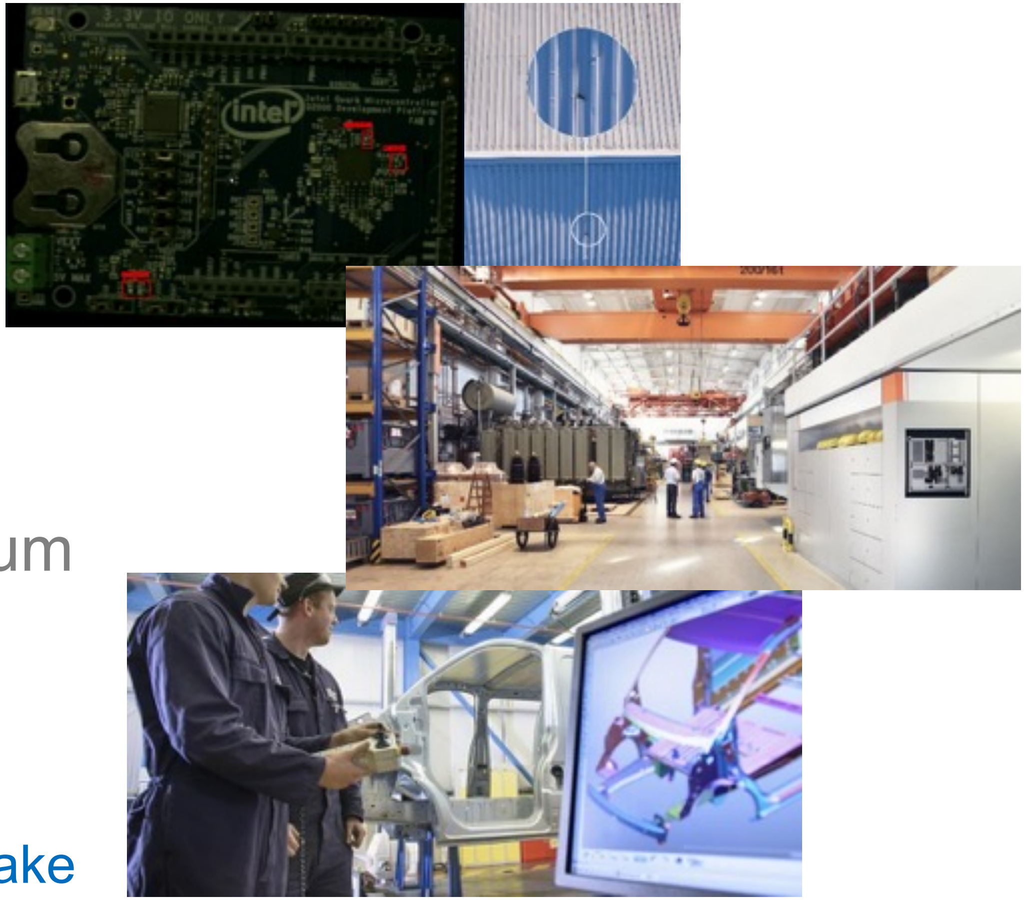
- Edge AI – Use Cases and Lifecycle
- Edge Data – What is New/Different?
- Federated Learning at the Edge
- Dealing with Not Having Shared Storage
- Privacy for Edge AI

Edge AI

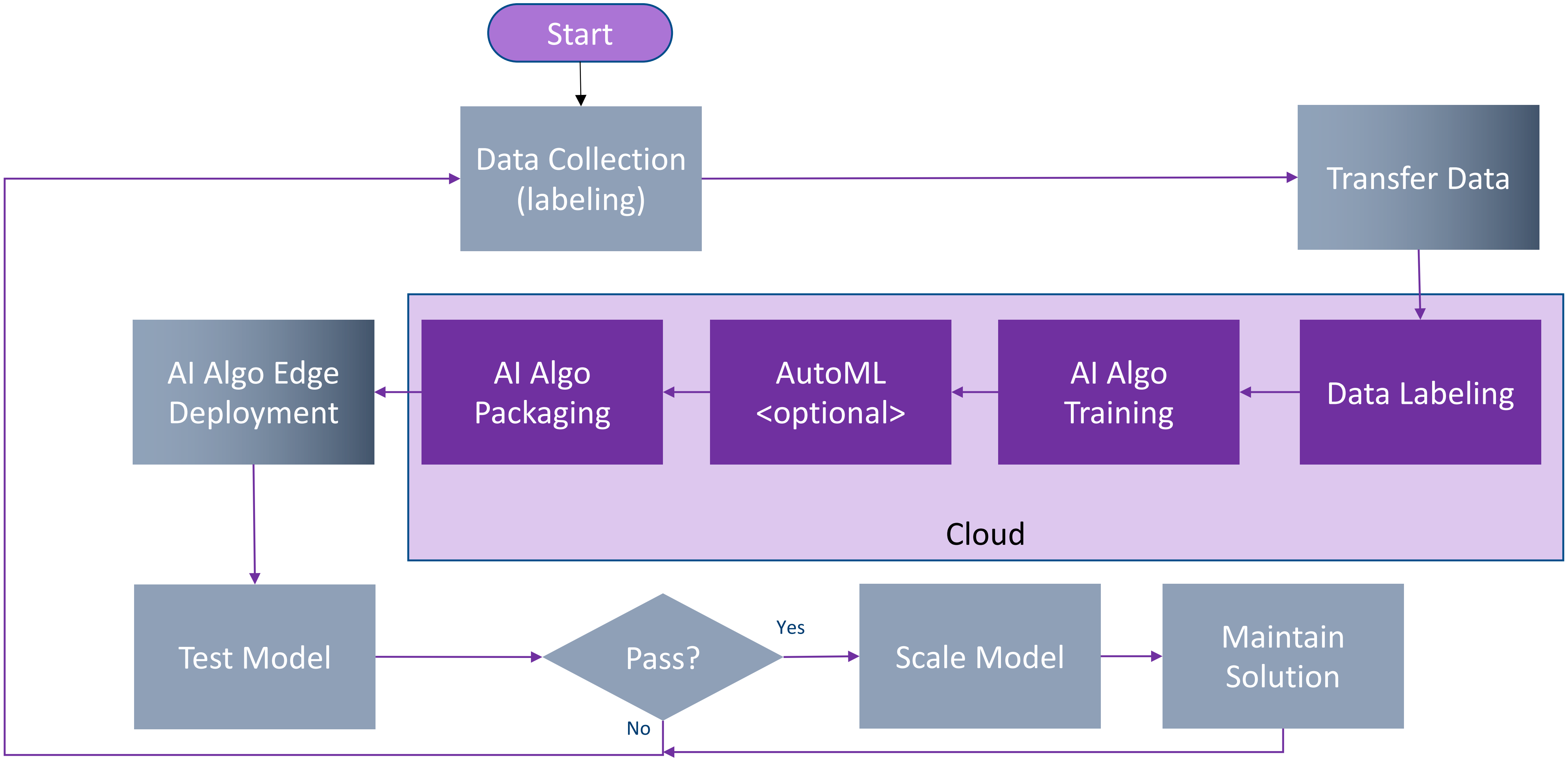
Use Cases and Lifecycle

Why AI at the Edge?

- Why?
  - Latency
  - Privacy
  - Bandwidth
- Challenges
  - Shift from inference only to the full continuum
  - Edge applications =
    - Data pipeline (ingest, infer, publish, store, visualize)
    - Business logic (send an alert, stop a process, take an action)



Phases of AI at the Edge



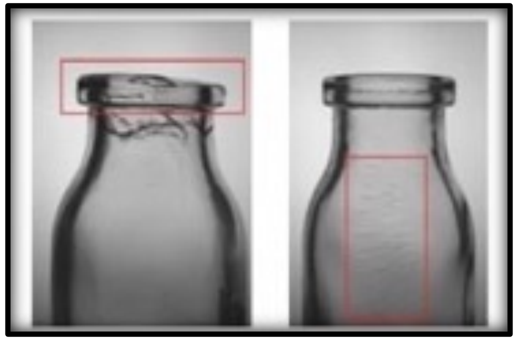
SNIA. CSTI | CLOUD STORAGE TECHNOLOGIES

AI Edge Use Cases

How can I better
PREDICT AND
Reduce
Downtime?

How can I optimize
operation
for higher throughput?

How can I
meet rising requirement on
product quality?



How can I leverage
AI for better business
outcomes?

Edge Data

What is New/Different?

Data

Edge data

- Highly distributed
- Crosses boundaries (OT)
- Heterogenous compute
- Dynamic

Cloud data

- Aggregated
- One policy (IT)
- Elastic abstracted compute
- Static (highly available)

Edge Data

What is needed?

- Elastic compute
- Data abstractions (multi-modal support, availability, reliability, redundancy, …)
- Ease of testing and experimentation

How to get there?

- Advances in AI are proving that making sense of data is very valuable (LLMs)
- Rethinking the persona (democratizing AI)
- AI (and data) as a tool

Federated Learning at the Edge

Machine learning without sharing data?

What is Federated Learning?

- Multiple parties
- Train a machine learning model
- Collaboratively
- Without sharing training data

Neonatal Care

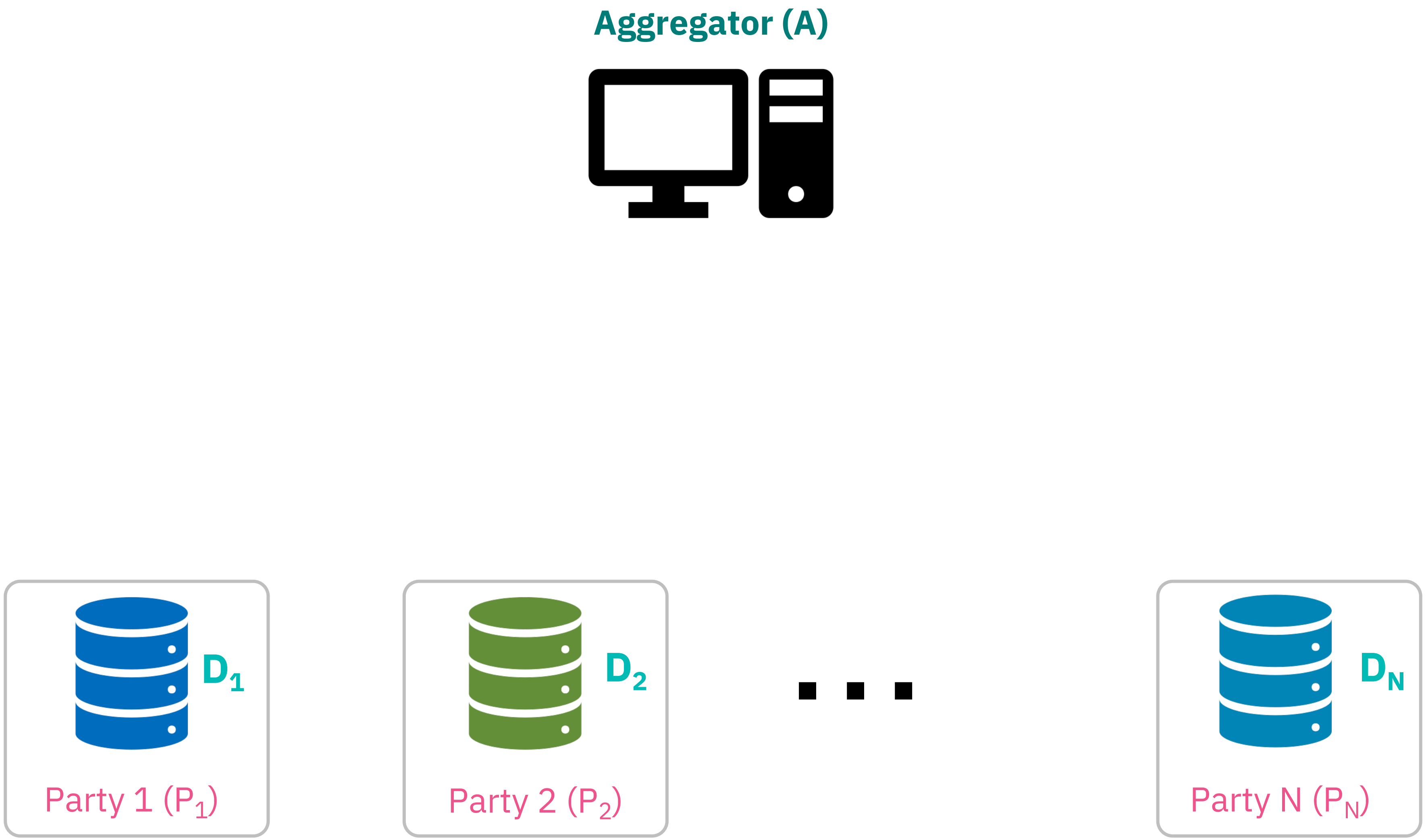
- Innocens predicts sepsis incidents
- More data from more hospitals
- Federated gradient boosted trees

SNIA. CSTI | CLOUD STORAGE TECHNOLOGIES

Environmental Monitoring

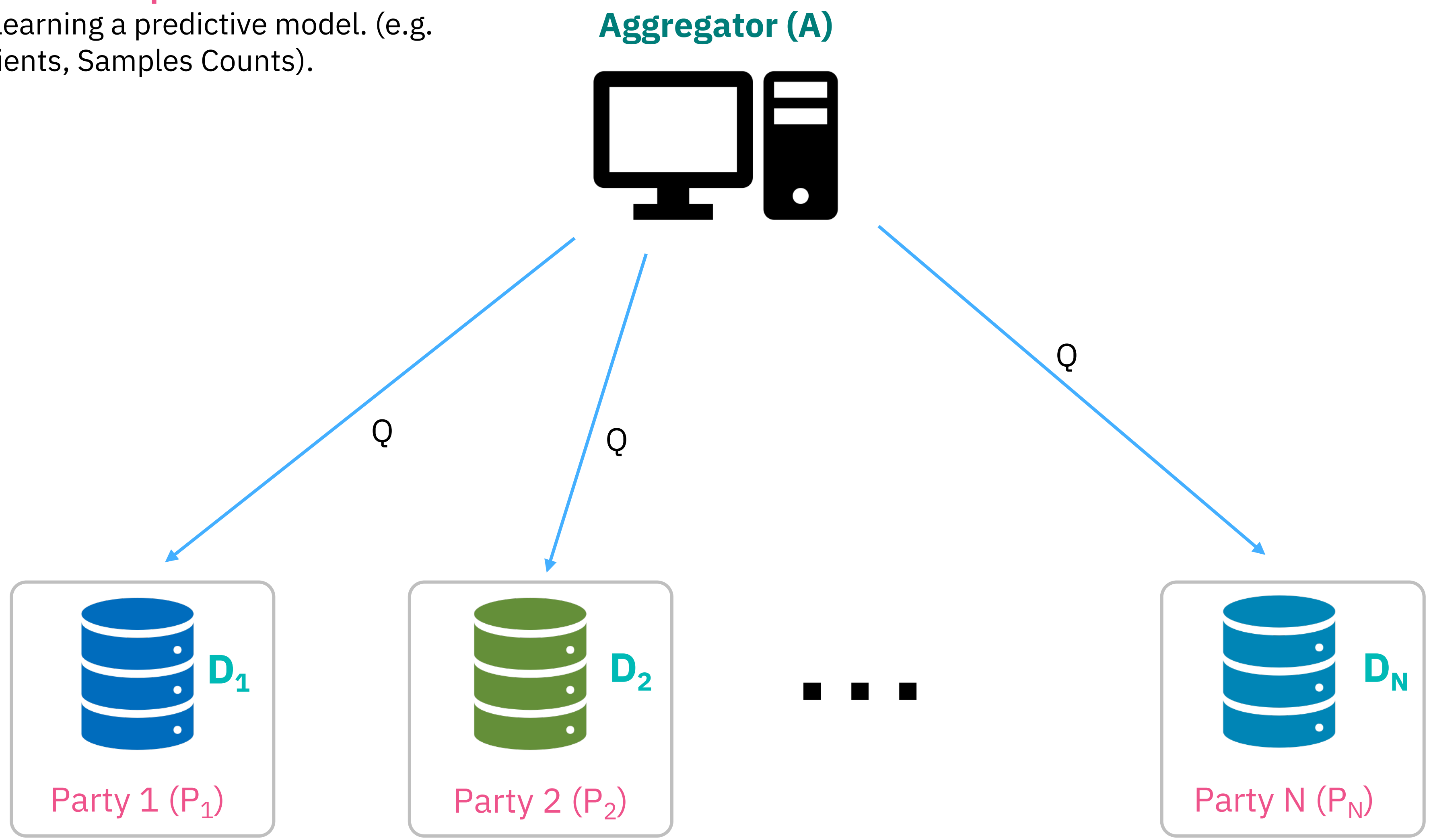
- Visual inspection
- Oil fields are typically shared
- Pipeline of DNNs

Basic Federated Learning



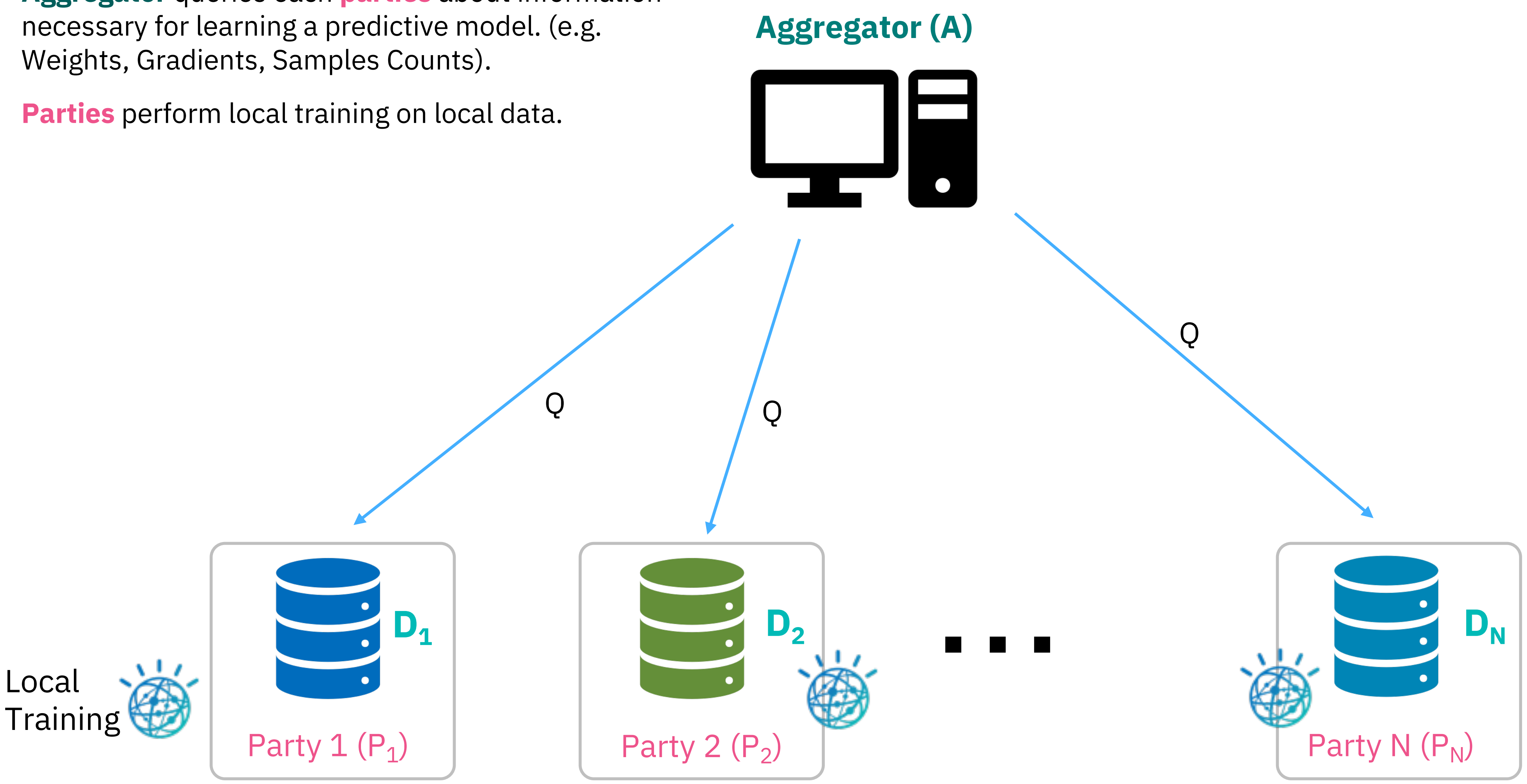
Basic Federated Learning

1. **Aggregator** queries each **parties** about information necessary for learning a predictive model. (e.g. Weights, Gradients, Samples Counts).



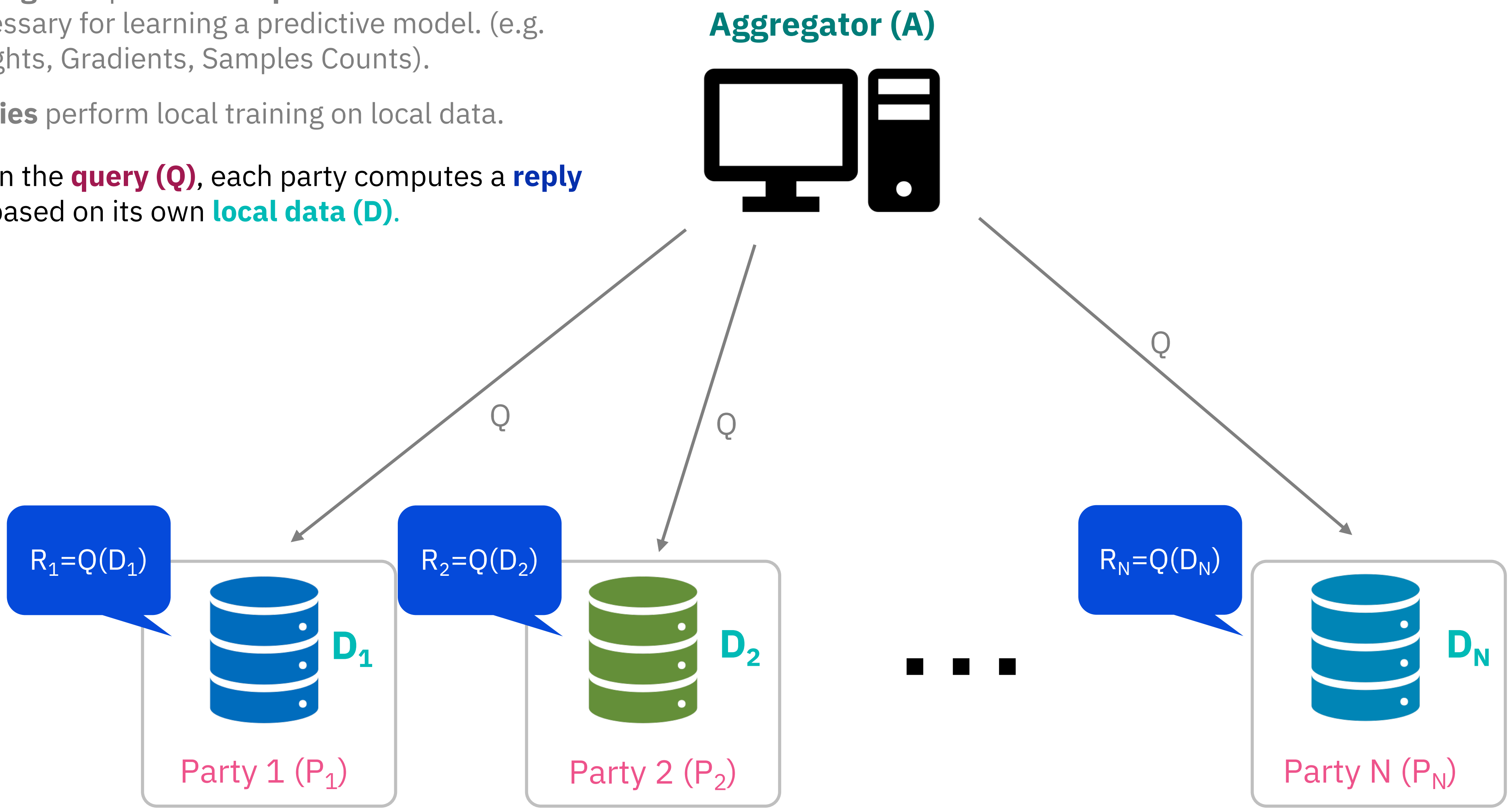
Basic Federated Learning

1. **Aggregator** queries each **parties** about information necessary for learning a predictive model. (e.g. Weights, Gradients, Samples Counts).
2. **Parties** perform local training on local data.



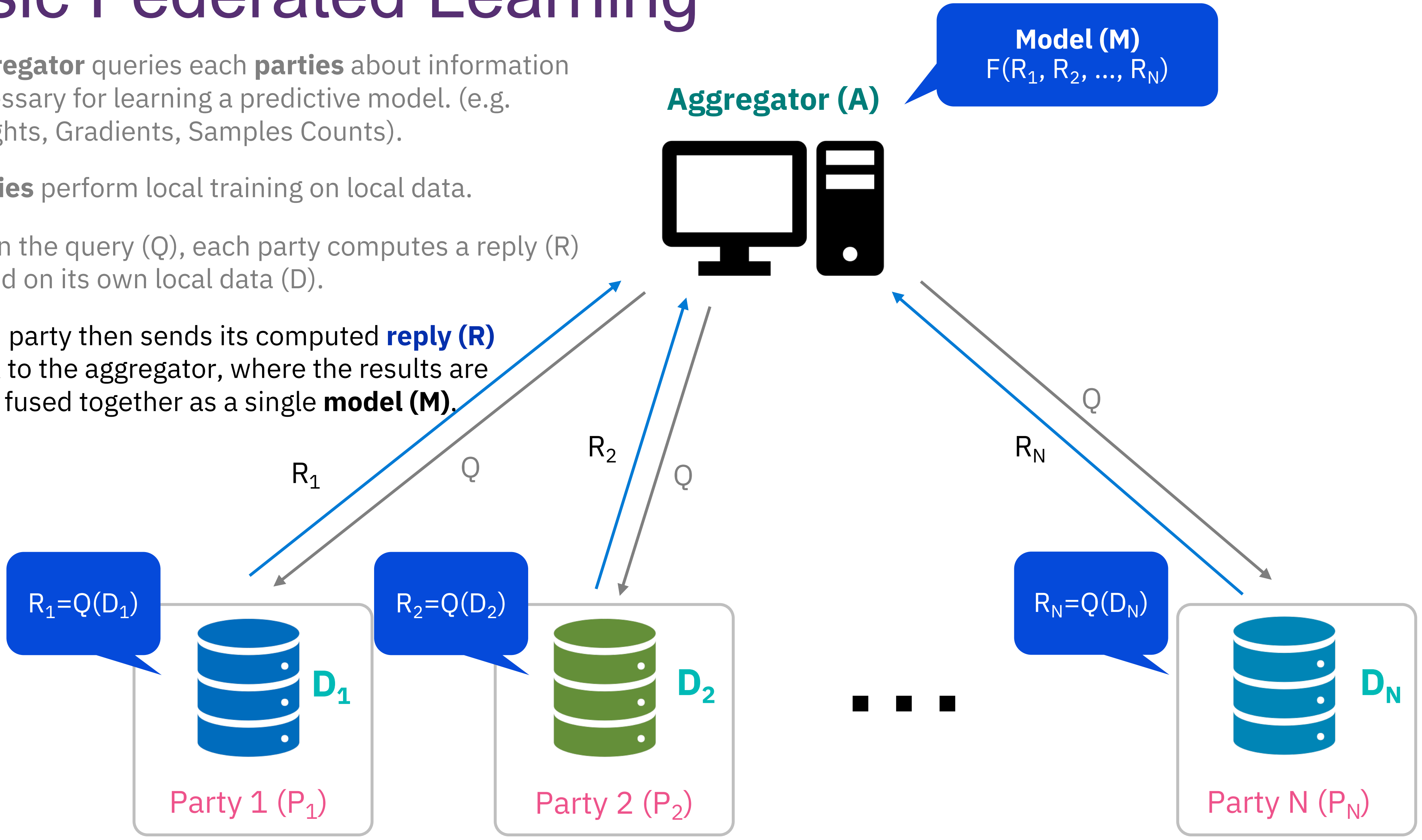
Basic Federated Learning

1. **Aggregator** queries each **parties** about information necessary for learning a predictive model. (e.g. Weights, Gradients, Samples Counts).
2. **Parties** perform local training on local data.
3. Given the **query (Q)**, each party computes a **reply (R)** based on its own **local data (D)**.



Basic Federated Learning

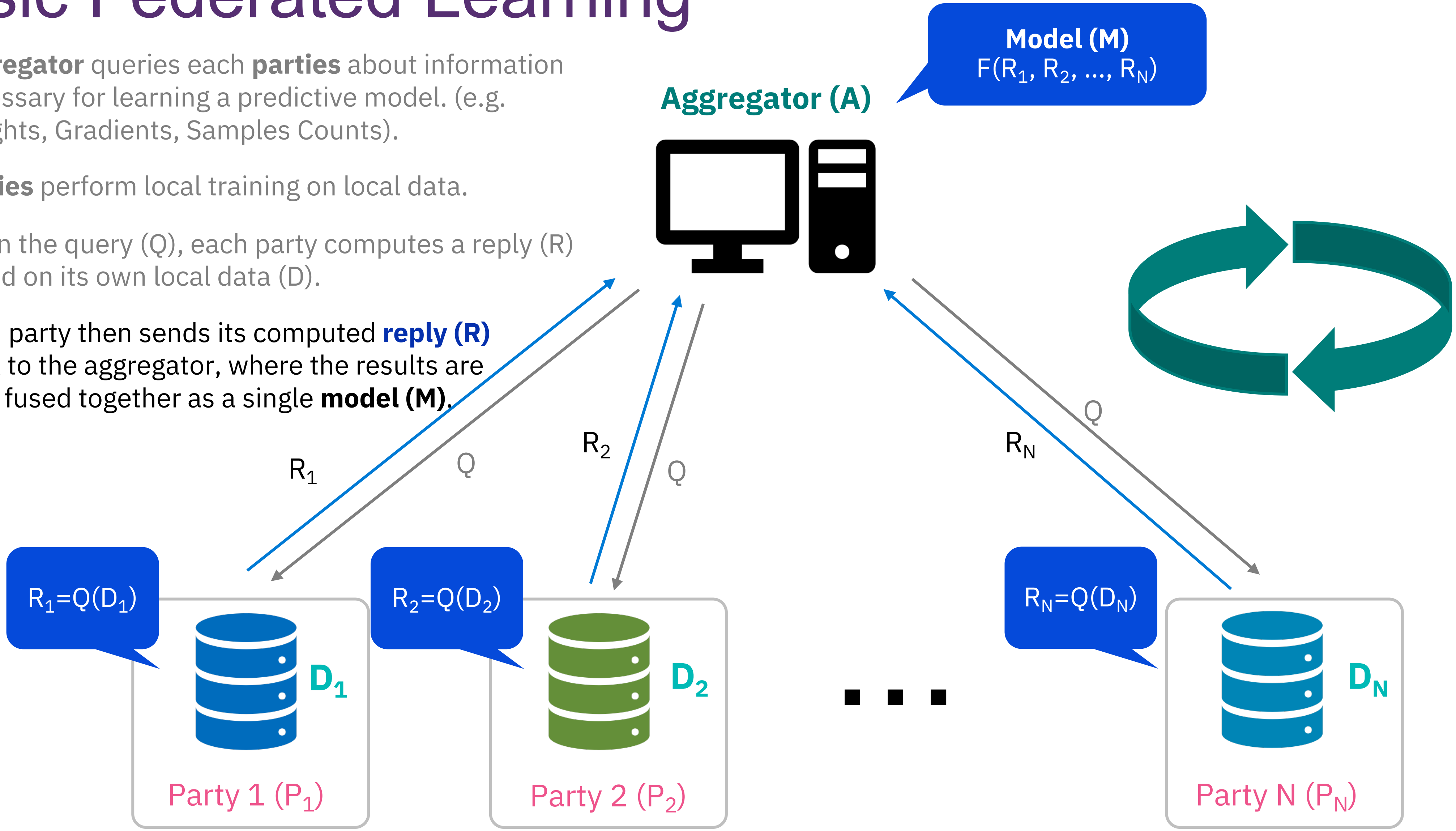
1. **Aggregator** queries each **parties** about information necessary for learning a predictive model. (e.g. Weights, Gradients, Samples Counts).
2. **Parties** perform local training on local data.
3. Given the query (Q), each party computes a reply (R) based on its own local data (D).
4. Each party then sends its computed **reply (R)** back to the aggregator, where the results are then fused together as a single **model (M)**.



SNIA. CSTI | CLOUD STORAGE TECHNOLOGIES

Basic Federated Learning

1. **Aggregator** queries each **parties** about information necessary for learning a predictive model. (e.g. Weights, Gradients, Samples Counts).
2. **Parties** perform local training on local data.
3. Given the query (Q), each party computes a reply (R) based on its own local data (D).
4. Each party then sends its computed **reply (R)** back to the aggregator, where the results are then fused together as a single **model (M)**.

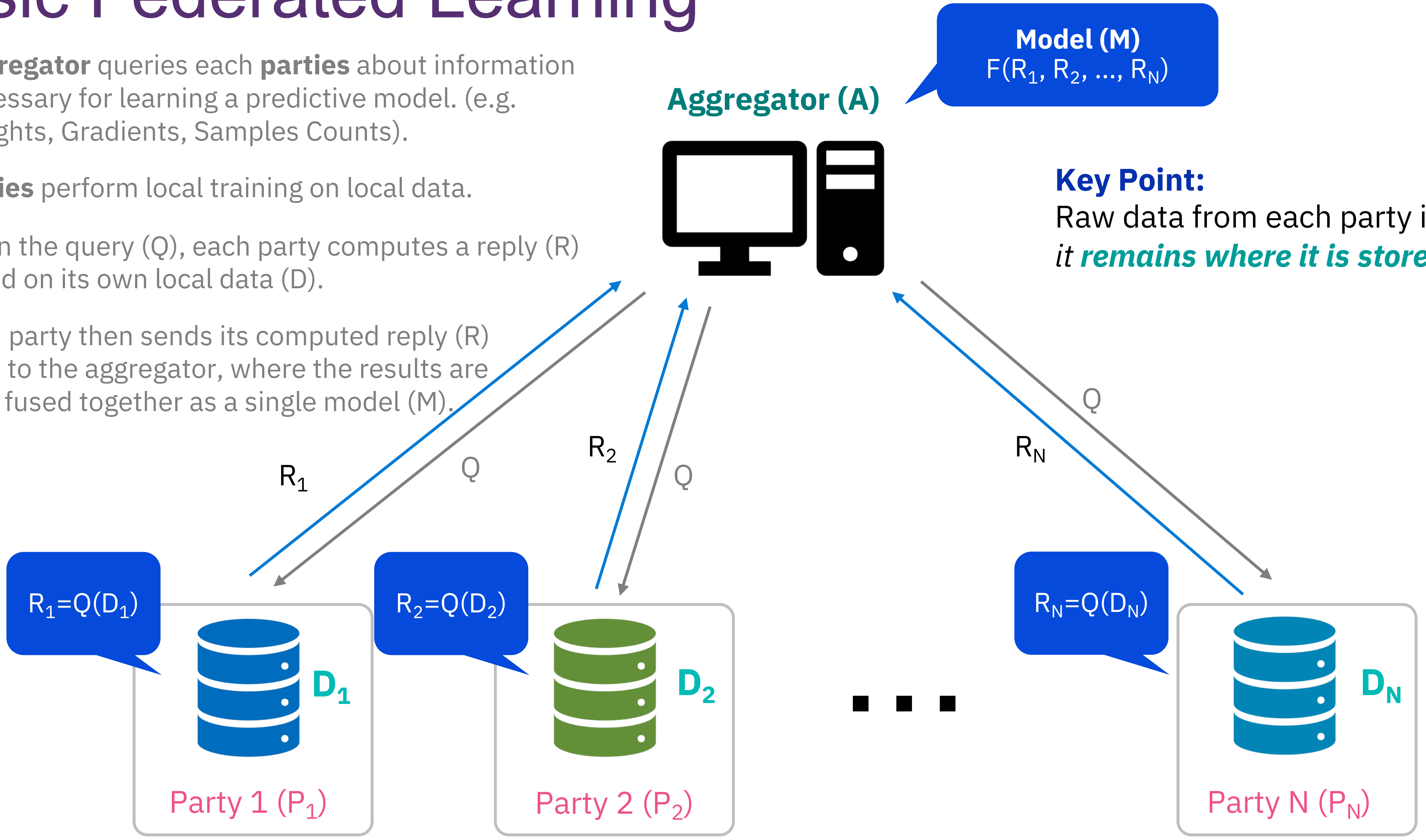


 |

Basic Federated Learning

1. **Aggregator** queries each **parties** about information necessary for learning a predictive model. (e.g. Weights, Gradients, Samples Counts).
2. **Parties** perform local training on local data.
3. Given the query (Q), each party computes a reply (R) based on its own local data (D).
4. Each party then sends its computed reply (R) back to the aggregator, where the results are then fused together as a single model (M).

Key Point:
Raw data from each party is **never shared**, *it* ***remains where it is stored***.

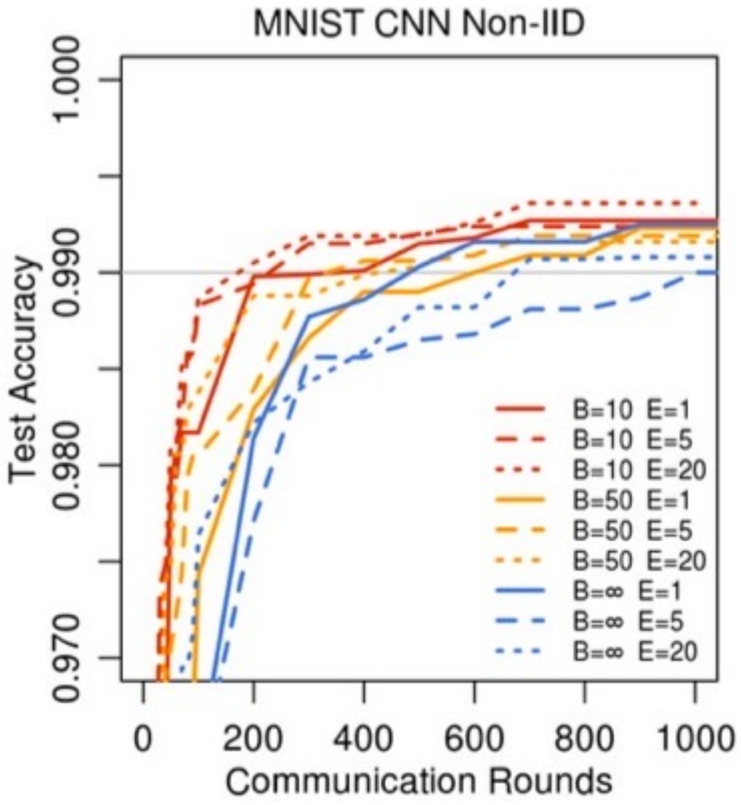


 |

Dealing with Not Having Shared Storage

Hyperparameter Tuning for Federated Learning

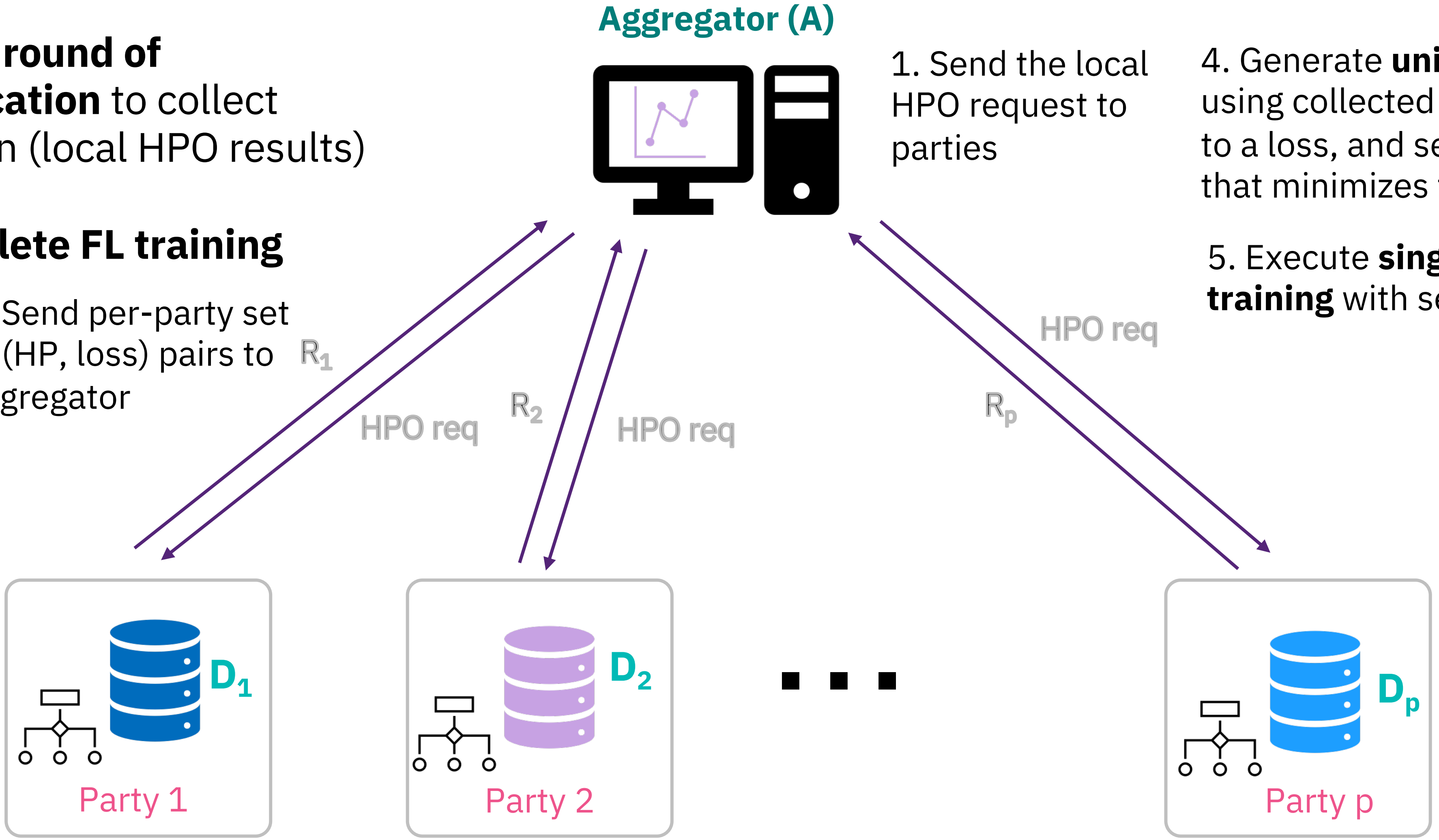
- Hyperparameters are important for model performance.
- HPO is typically automated or assisted.
- Federated Learning makes
  - HPO harder due to lack of data access
  - Introduces new HPs



Ref: McMahan, Brendan, et al. "Communication-efficient learning of deep networks from decentralized data." *Artificial Intelligence and Statistics*. PMLR, 2017

Single-shot HPO Framework: FLoRA

- **One extra round of communication** to collect information (local HPO results)
- **One complete FL training**



1. Send the local HPO request to parties

2. Run **independent local HPOs** to generate set of (HP, loss) pairs

3. Send per-party set of (HP, loss) pairs to aggregator

4. Generate **unified loss surface** using collected info to map any HP to a loss, and select a **single HP** that minimizes this loss surface

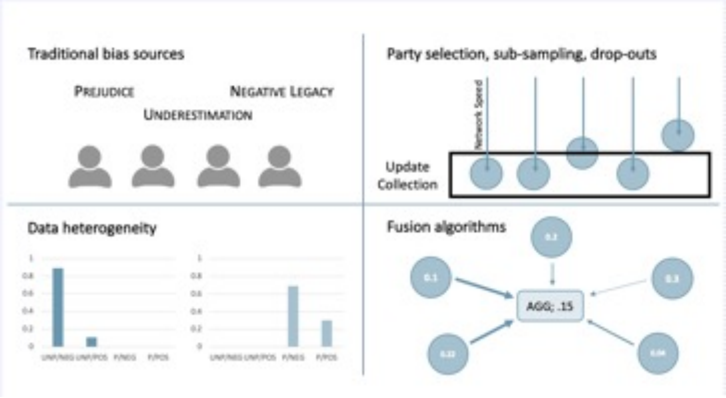
5. Execute **single federated training** with selected HP

Ref.: Y Zhou, P Ram, T Salonidis, N Baracaldo, H Samulowitz, H Ludwig. "Single-shot general hyper-parameter optimization for federated learning". *In International Conference on Learning Representations, 2023*

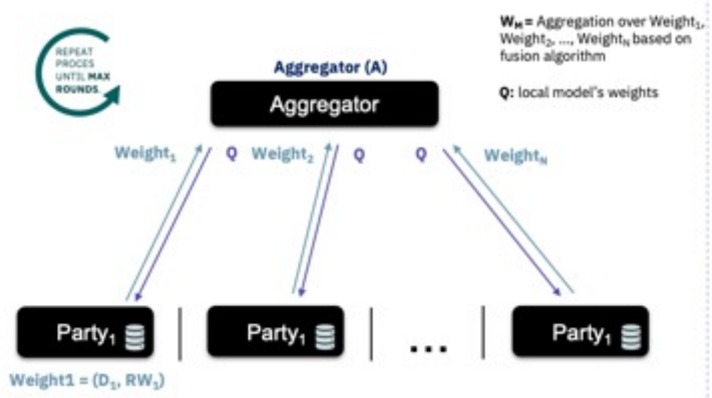
Fairness and Bias Mitigation in Federated Learning

- **Problem:**
  - Traditional methods to analyze and mitigate statistical bias in ML require full training dataset
  - Data locations introduce new sources of bias
- **Solution:**
  - Pre-processing: Local re-weighing
  - In-processing: Federated prejudice removal

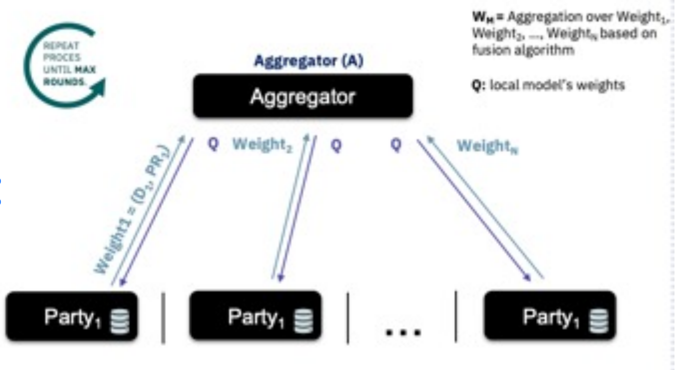
New sources of bias:



Local re-weighing:



Federated prejudice removal:

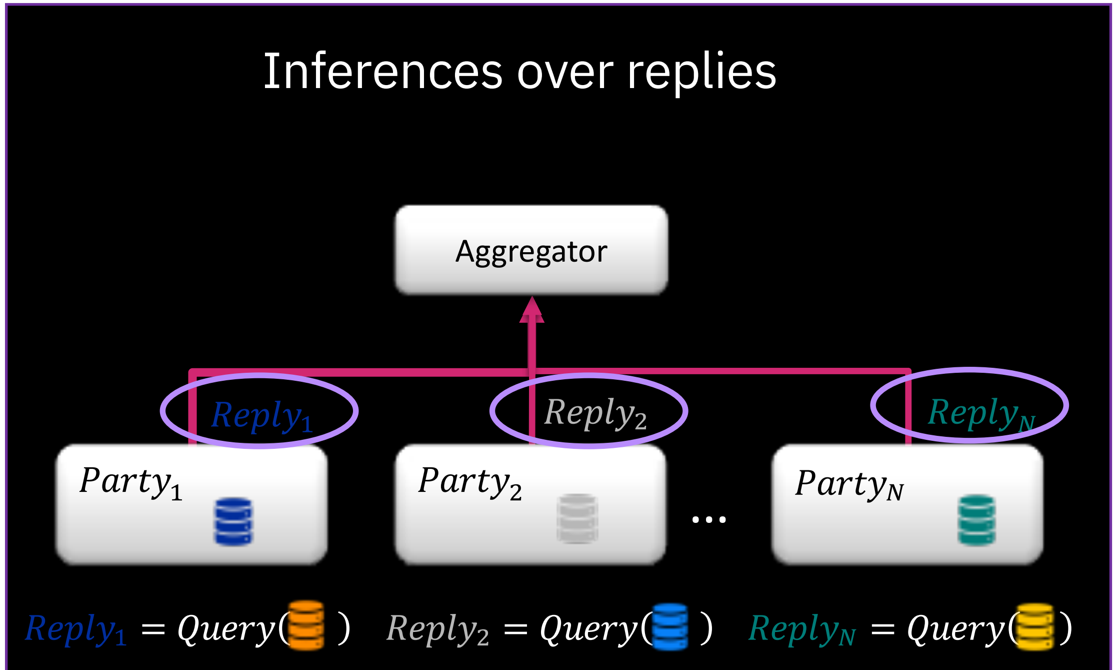


$$-L(D;\Theta) + \eta R(D,\Theta) + \frac{\lambda}{2}||\Theta||_2^2$$

Privacy for Edge AI

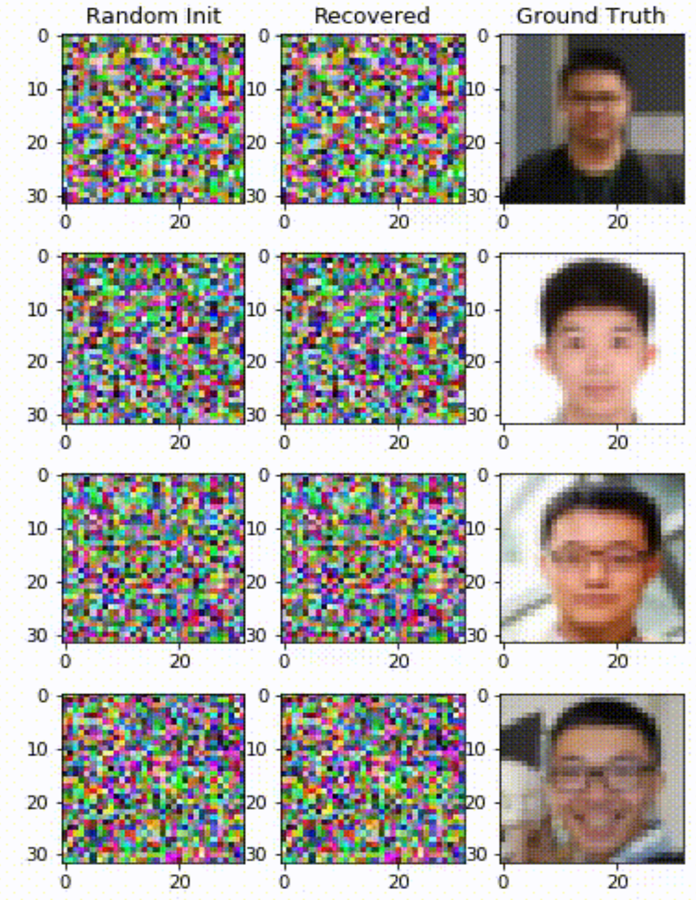
Inference Threats

- In untrusted environments adversaries may try to infer information by analyzing other parties replies



[1] Le Trieu Phong et. al. 2018. Privacy-Preserving Deep Learning via Additively Homomorphic Encryption
[2] IG: Inverting Gradients (NeurIPS 2020),
[3] DLG: Deep Leakage from Gradients (NeurIPS 2019),
[4] iDLG: Improved Deep Leakage from Gradients
[5] Melis et. al 2019 S&P Exploiting Unintended Feature Leakage in Collaborative Learning :

Taken from **Deep Leakage From Gradients**
https://github.com/mit-han-lab/dlg



Examples
1. Inference based on gradients exchanged [1-4]
2. Gradient of a bag of words: non-zero means the data has a word
3. Properties e.g., was someone wearing glasses [5]

Privacy Techniques

Differential privacy,

Fully Homomorphic,

Functional encryption with TPA,

Functional encryption using decentralized MIFE,

Threshold Pailler,

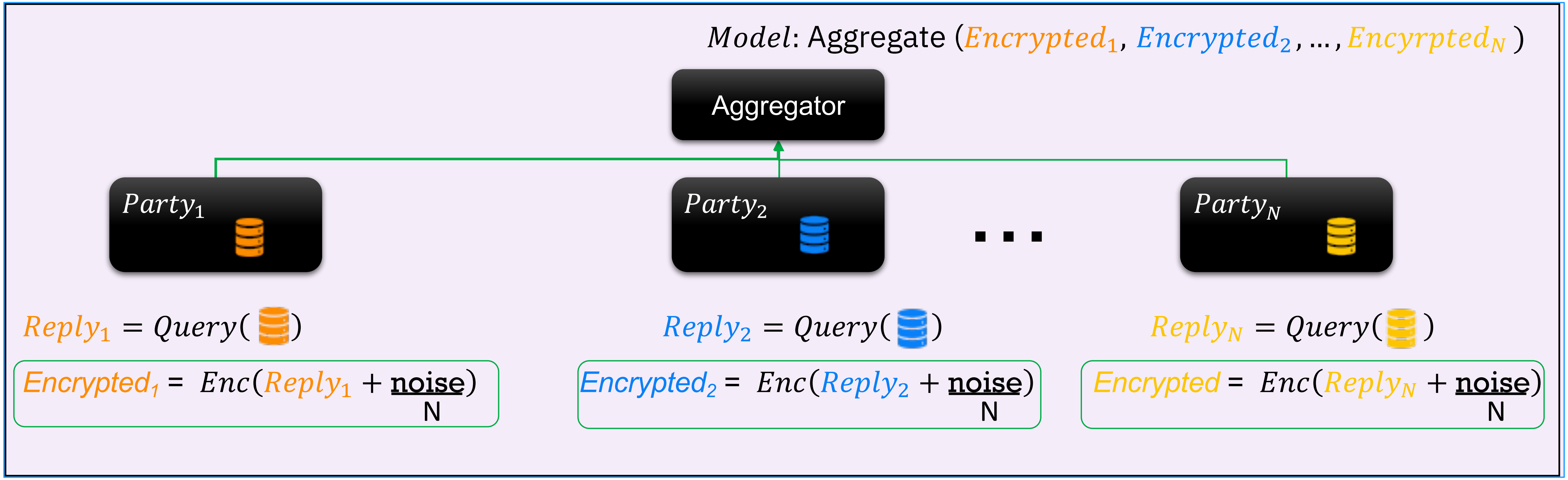
Pailler,

Secret Sharing and

Trusted Execution Environments

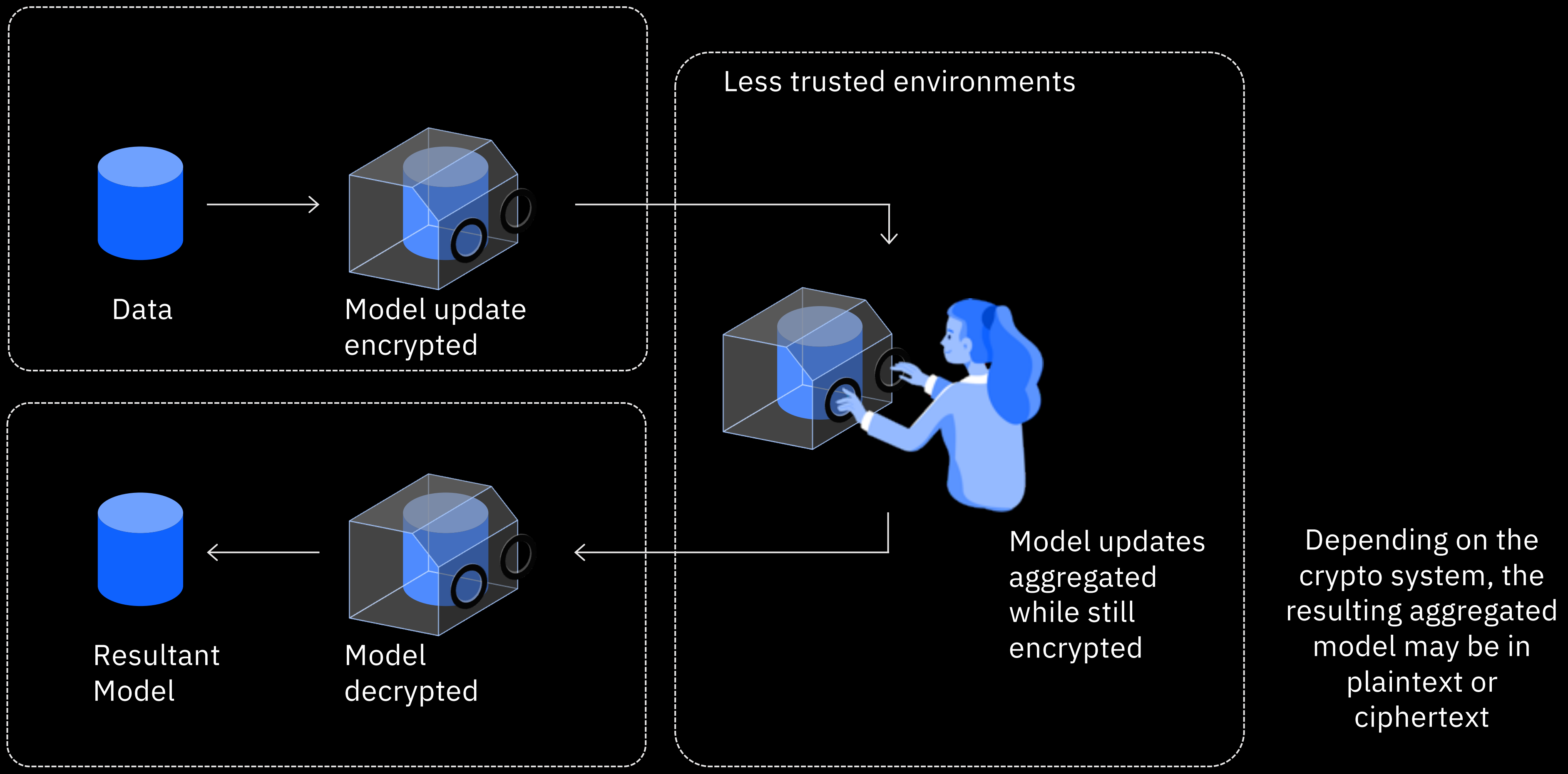
SNIA. CSTI | CLOUD STORAGE TECHNOLOGIES

Combine multi-party computation and differential privacy to achieve higher model accuracy through reduced noise

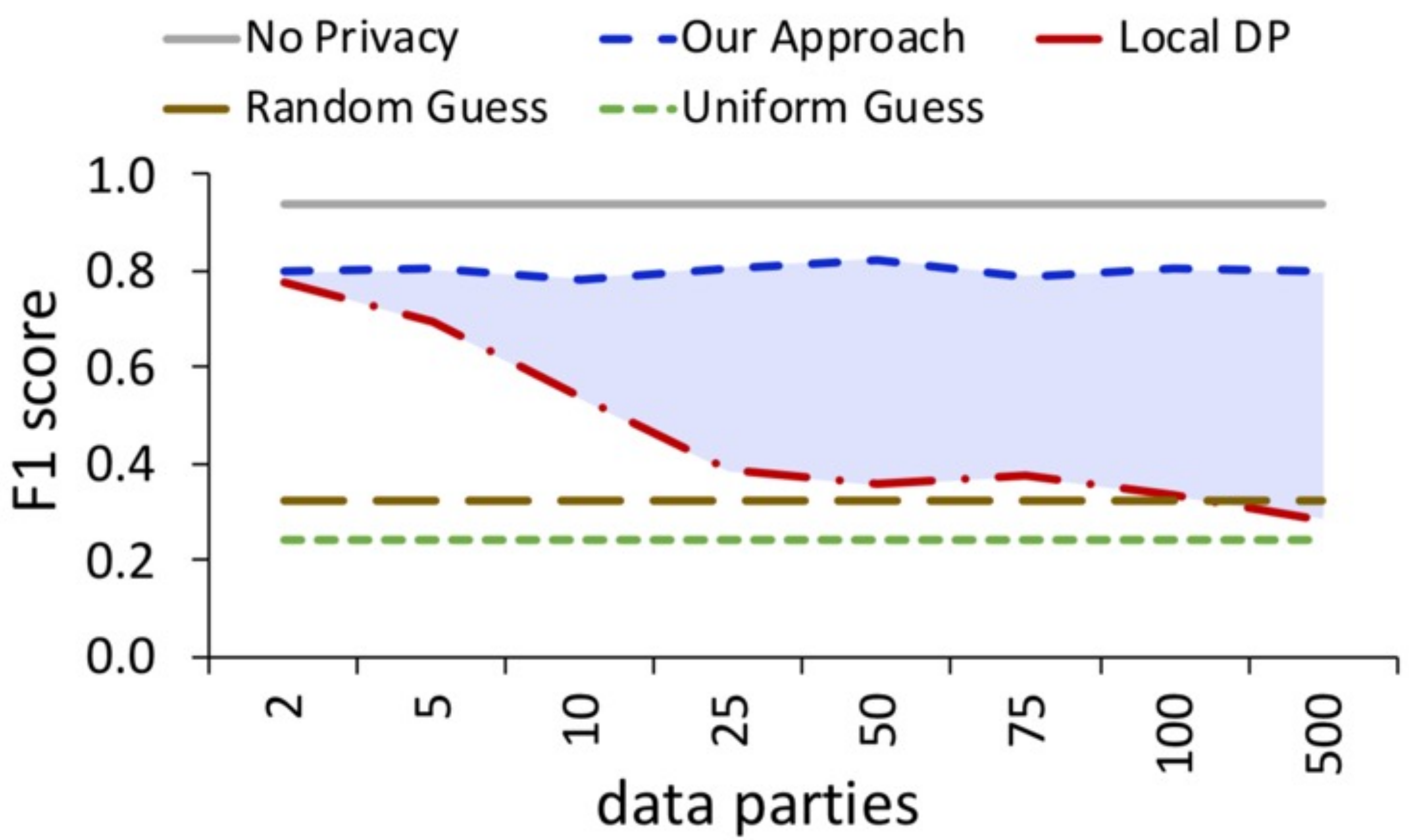


Multi-Party Computation

- Provides privacy of the inputs



Combination Approach Outperforms Baselines



- The more parties the data is divided amongst, the more noise the *Local Differential Privacy* approach requires, resulting in poor accuracy
- In contrast, our approach produces stable accuracy
- The baseline converges with the "random guess" line at around 100 parties

Results shown for decision tree, Nursery dataset from UCI, $\epsilon = 0.5$, 10 parties

A Hybrid Approach to Privacy-Preserving Federated Learning Truex et al
(Best paper award)

Summary

Summary

- AI is changing edge compute forever; edge is also changing AI
- Data at the edge is complex, noisy, and distributed
- Solving the data challenges is the key to democratizing AI

Join us in innovating in edge AI!

Q&A

Thanks for Viewing this Webinar

- Please rate this presentation and provide us with feedback
- This webinar and a copy of the slides are available at the SNIA Educational Library https://www.snia.org/educational-library
- A Q&A from this webinar will be posted to the SNIA Cloud blog: www.sniacloud.com/
- Follow us on X (formerly) Twitter @SNIACloud

Thank You!

SNIA. CSTI | CLOUD STORAGE TECHNOLOGIES