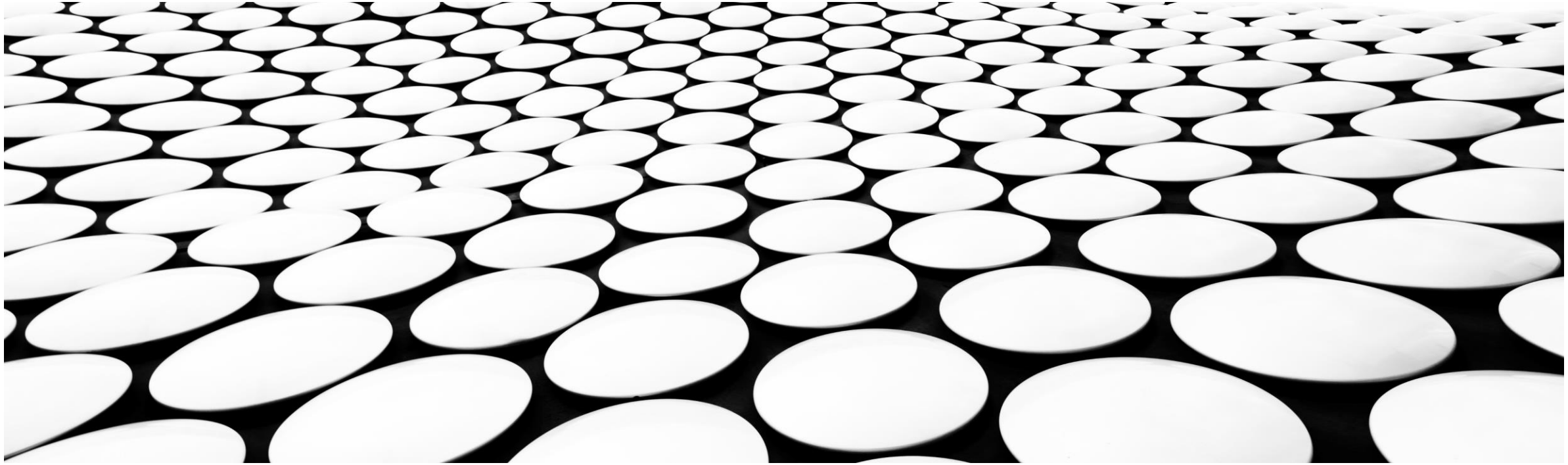
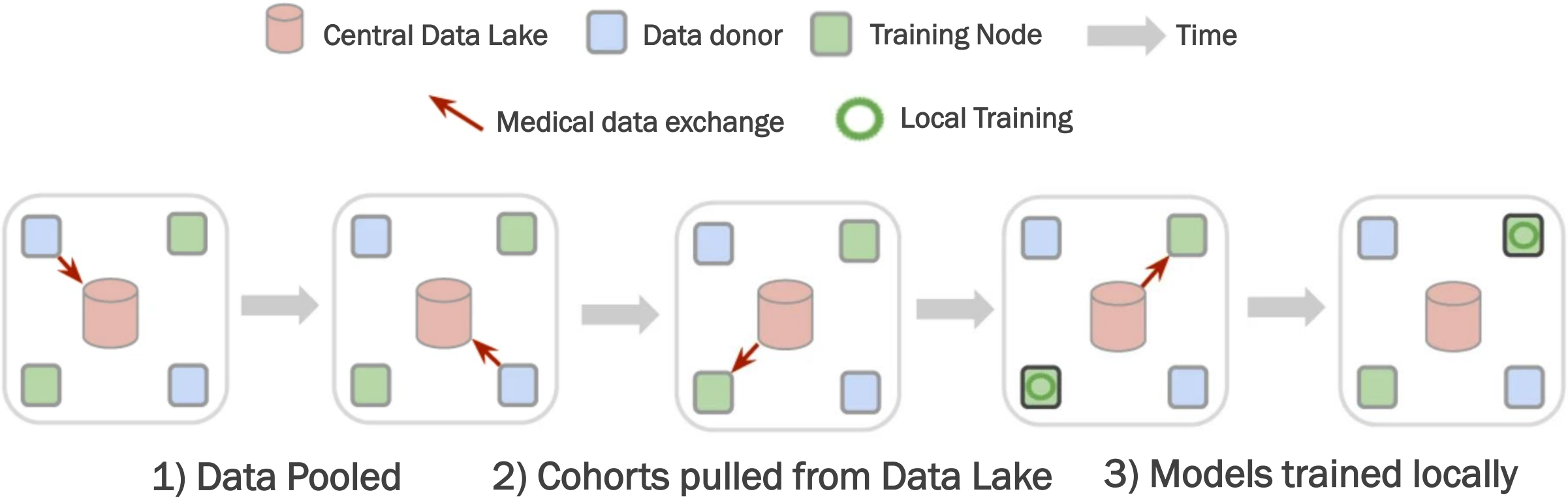


FEDERATED LEARNING APPLICATION IN CANCER PATIENT DATA ANALYSIS

IVAN KNEŽEVIĆ - SEEIIST

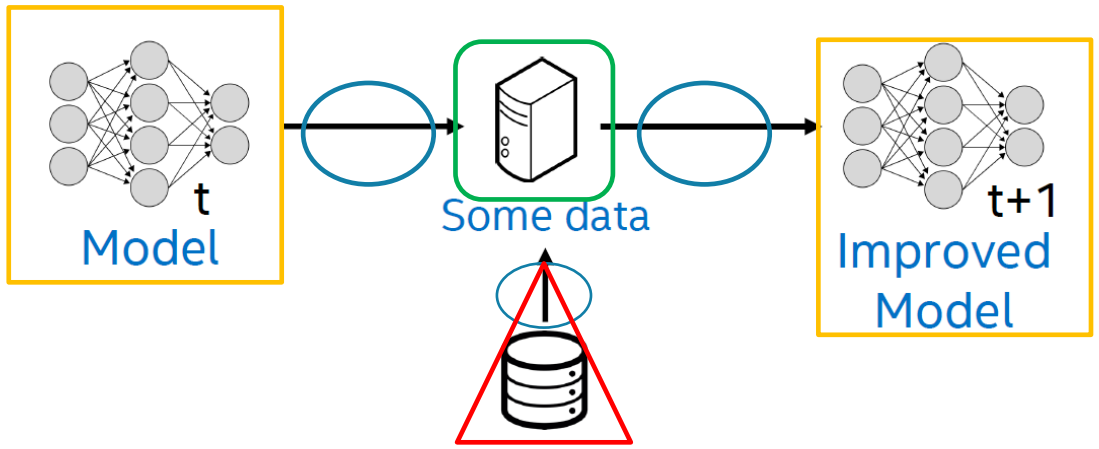


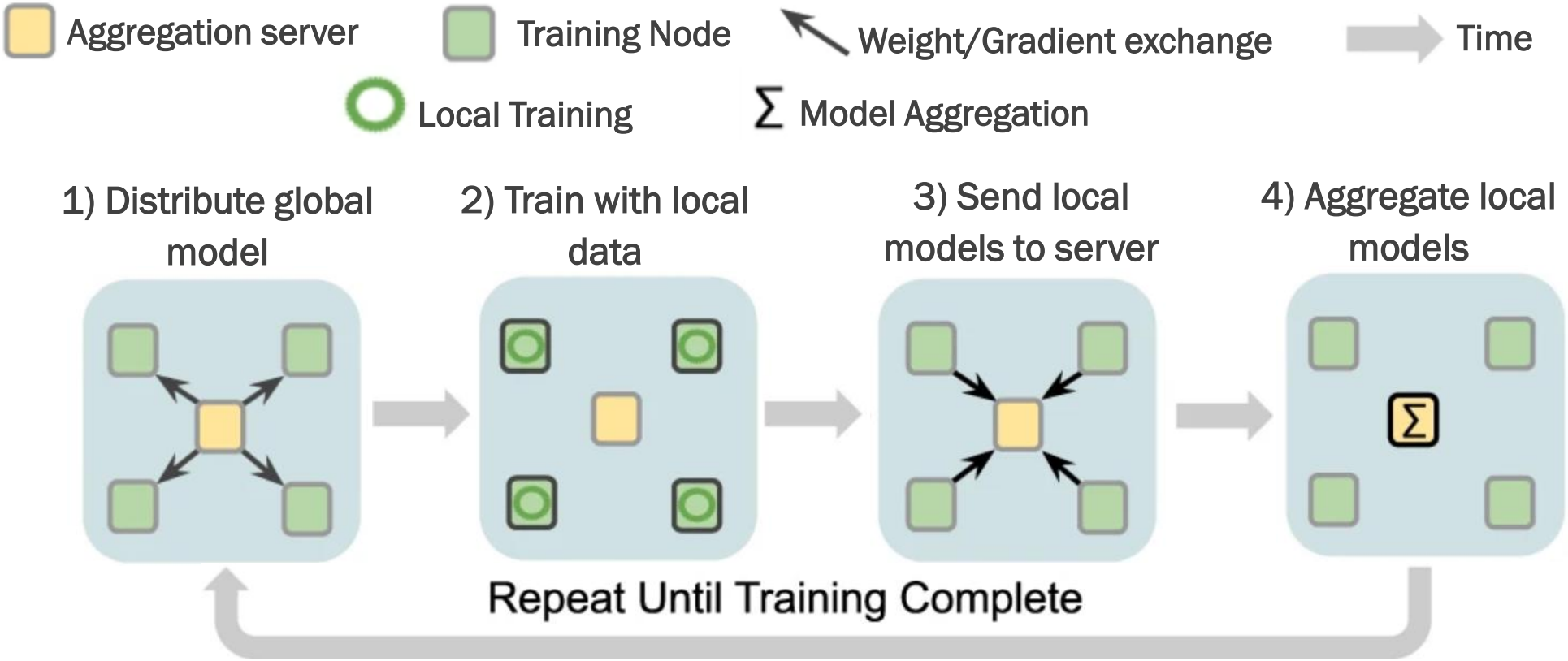
CENTRALIZED MACHINE LEARNING WORKFLOW

CENTRALIZED MACHINE LEARNING PROBLEMS

HUGE PROBLEM FOR CANCER PATIENT DATA ANALYSIS

- **Connectivity** – data must be transmitted over a stable connection
- **Architecture** – requires high computational power central server and big-data system support
- **Dataset limit** – eventually hit the limit of dataset provoking decrease in information value
- **Privacy** – sensitive operational data must remain on site



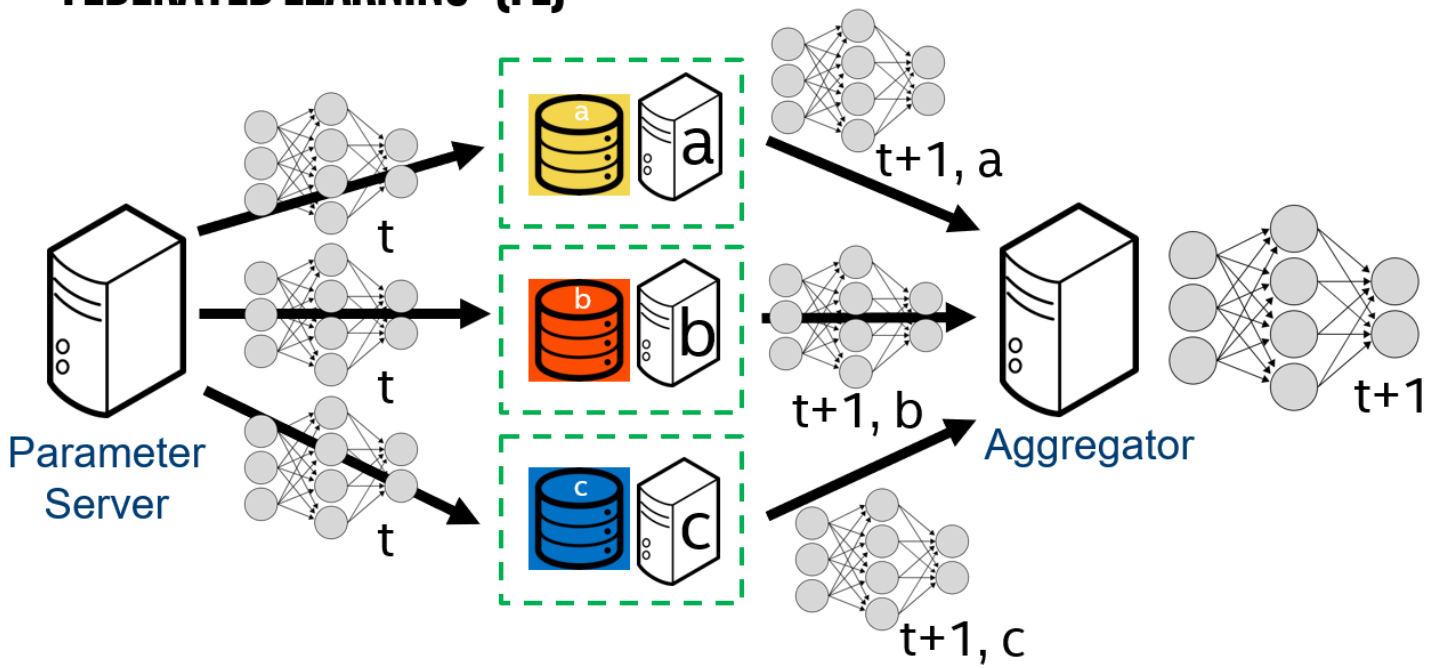


FEDERATED LEARNING WORKFLOW

FEDERATED LEARNING

- PARAMETER SERVER
- COLLABORATORS
- AGGREGATOR

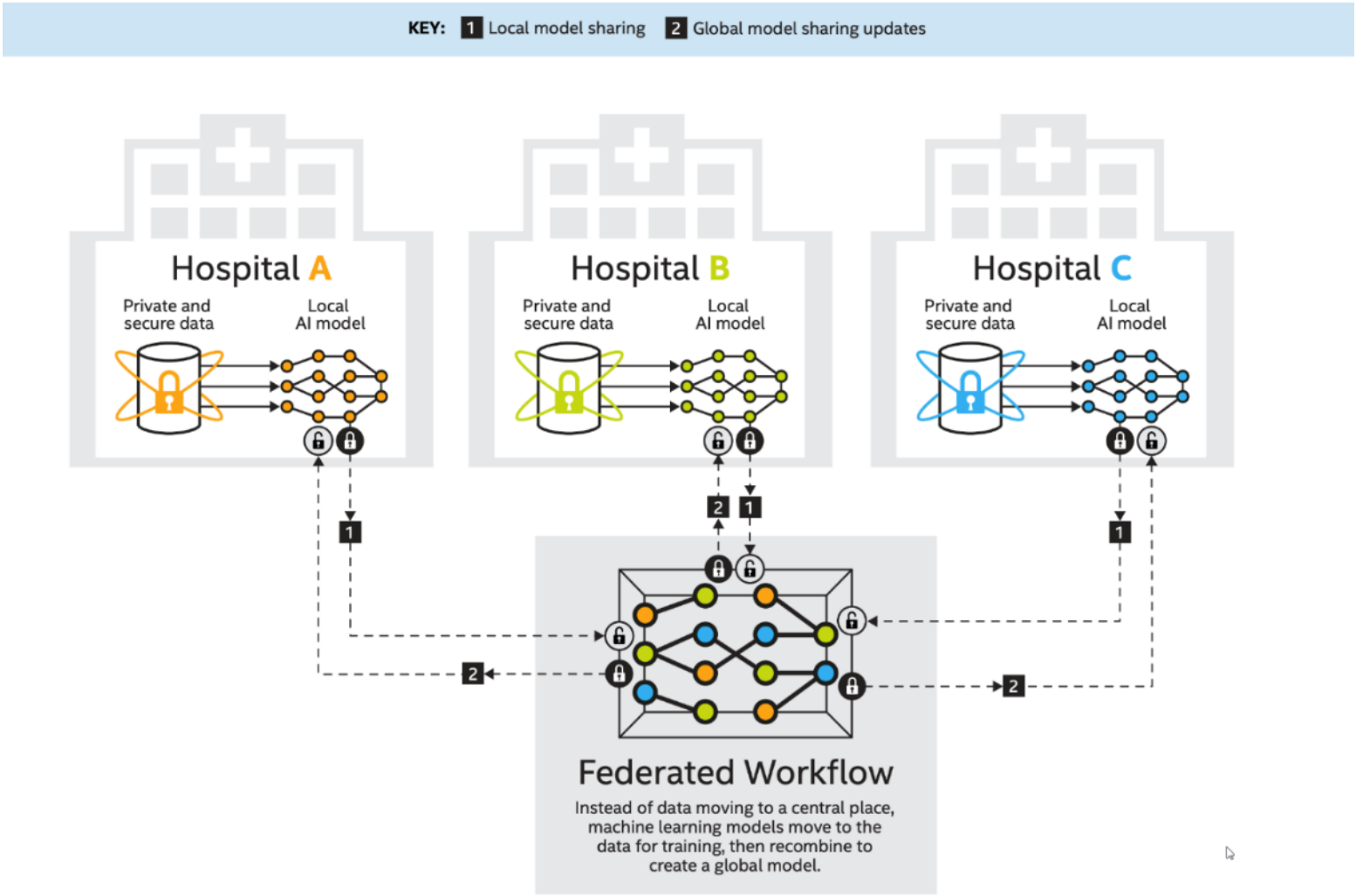
"Instead bringing data to the model, we take model out to the data" ,unknown Google engineer ☺



Machine learning technique that trains an algorithm across multiple decentralized edge devices or servers holding local data samples, without exchanging them.

FEDERATED LEARNING

HEALTH CARE ENVIRONMENT



FEDERATED LEARNING

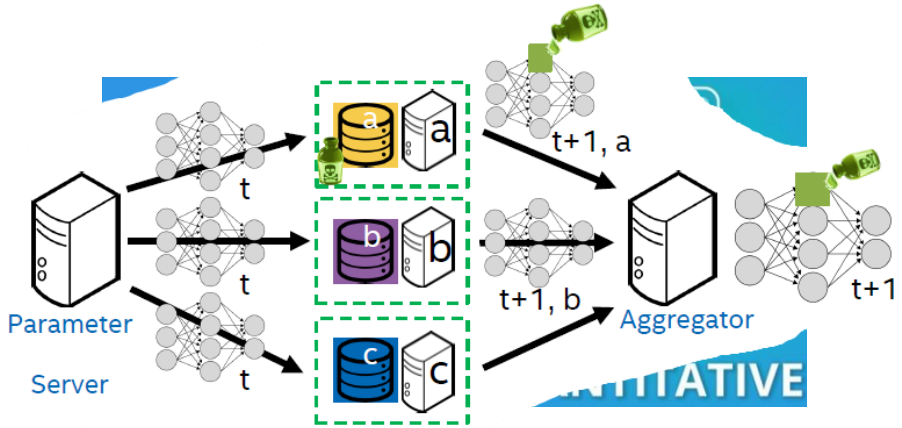
CHALLENGES AND CONSIDERATIONS

Data heterogeneity - qualitative vs. quantitative

- Variety of modalities, dimensionality and characteristics
- Acquisition differences
- Global optimal solution may not be optimal for an individual local participant

Privacy and security - privacy vs. performance

- "Data poisoning"
- Extraction attacks – recover training data from model



FEDERATED LEARNING

IVAN.KNEZEVIC@CERN.CH

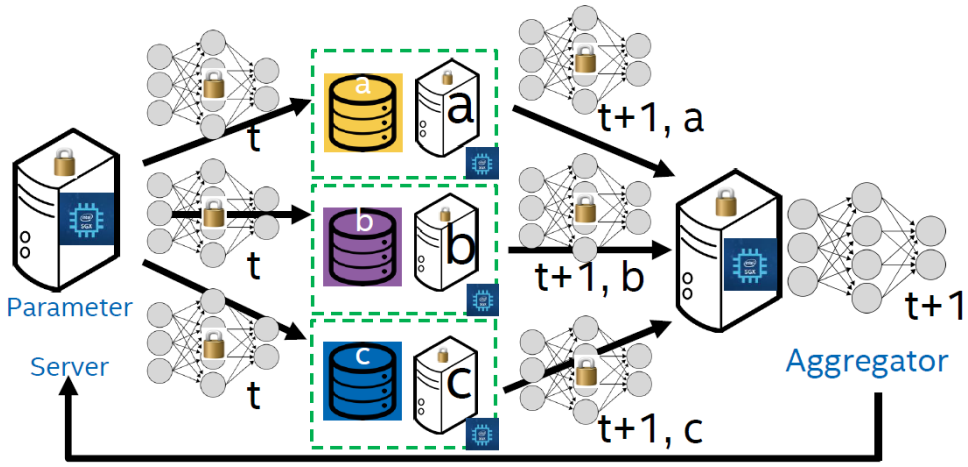
PROPOSED SOLUTIONS

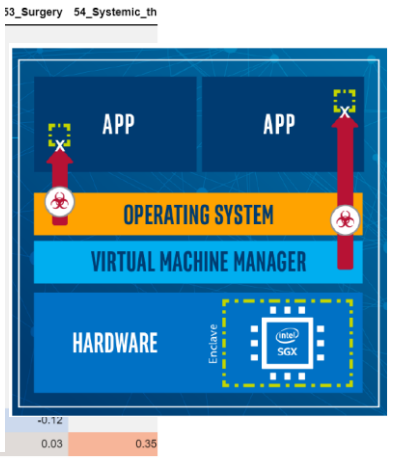
Data heterogeneity

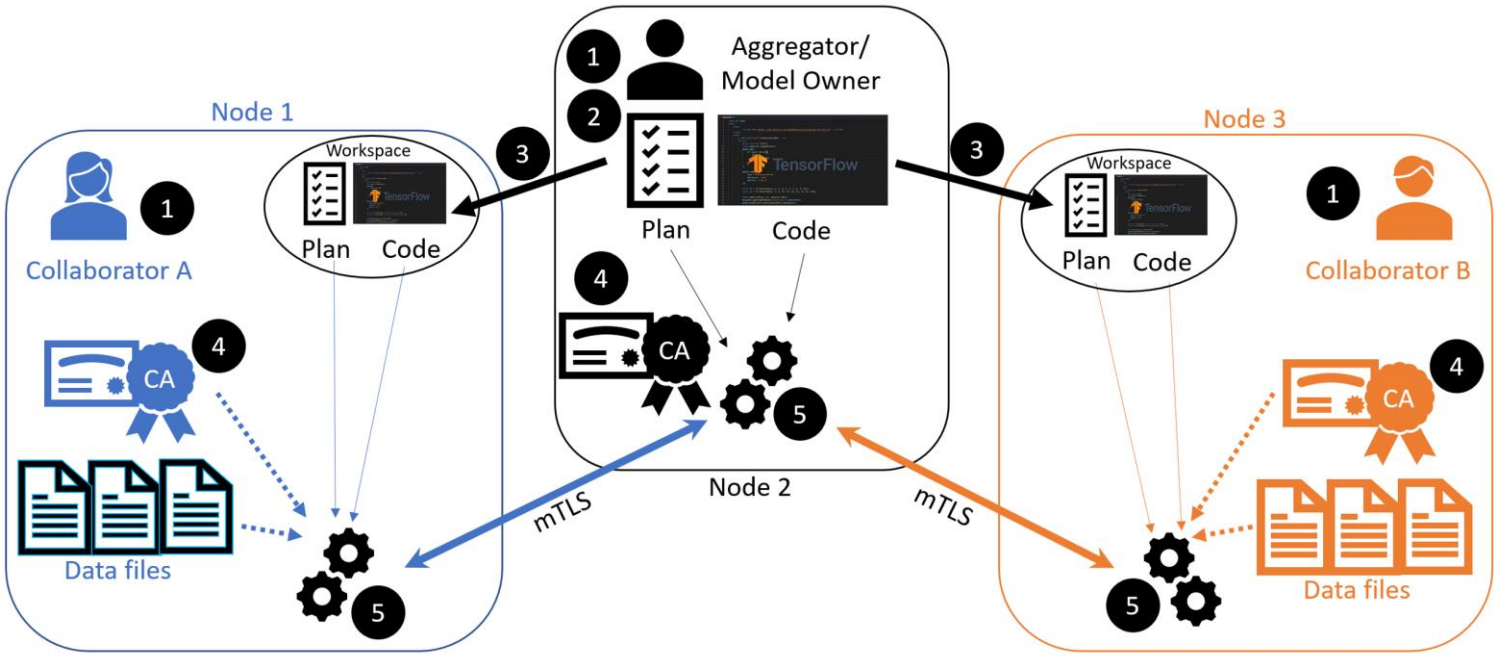
- Proposed usage of homogenous ENCR dataset
- Corelation matrix analysis for quantitative data
- Contingency table analysis for qualitative data
- Analysis of different ML algorithms

Privacy and security

- Homomorphic encryption or PKI
- IntelFL framework integration with SGX technology







1. Install OpenFL in a Python environment on all machines in the federation
2. Create FL workspace on aggregator machine
3. Move workspace to the other machines in the federation
4. Make sure everyone has their own valid PKI certificate
5. Start the nodes

RUNNING THE FEDERATED USING INTEL OPENFL

FEDERATED LEARNING

ALGORITHM

Require: num_federated_rounds T

1: **procedure** AGGREGATING

2: Initialise global model: $\mathrm{W}^{(0)}$

3: **for** $t \leftarrow 1 \cdots T$ **do**

4: **for** *client* $k \leftarrow 1 \cdots K$ **do** ▷ *Run in parallel*

5: Send $W^{(t-1)}$ to client k

6: Receive model updates and number of local training iterations $(\Delta \mathrm{W}^{(t-1)}, Nk)$ from client's local training with $\mathrm{L_k}(\mathrm{Xk}; \mathrm{W}^{(t-1)})$

7: **end for**

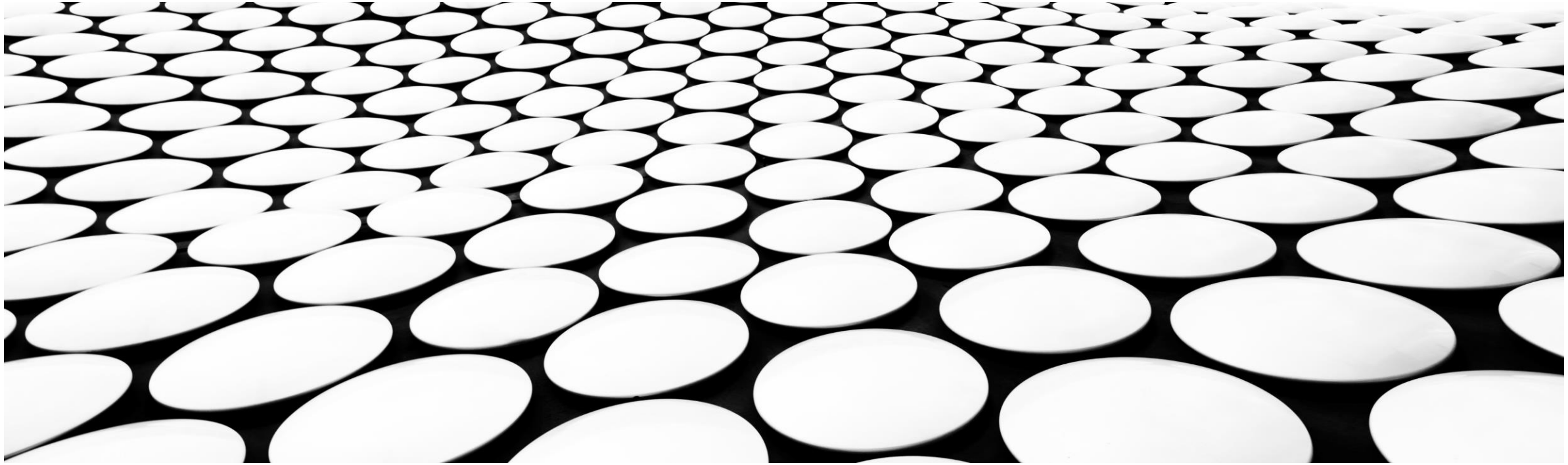
8: $\mathrm{W}^{(t)} \leftarrow W^{(t-1)} + \frac{1}{\sum_k N_k} \sum_k (Nk \cdot W_k^{(t-1)})$

9: **end for**

10: **return** $\mathrm{W}^{(t)}$

11: **end procedure**

FL algorithm via Hub & Spoke (Centralised topology) with `FedAvg` aggregation.



THANK YOU FOR YOUR ATTENTION

QUESTIONS?