# Overview of Federated Learning

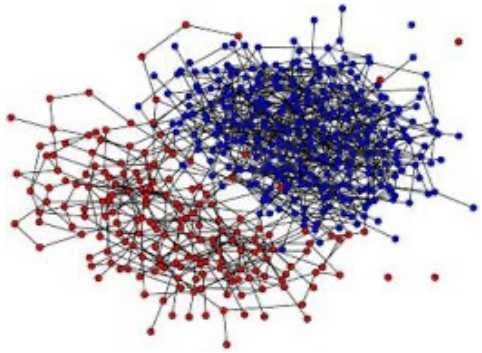
Dr. Lotfi ben Othmane
University of North Texas

# Administrative

- Quiz 2 is graded
- The grades will be released this week
  - Assignment 1
  - Project phase 2
- Coming:
  - Quiz 3 – 3/19
  - Assignment 2 – 3/23
  - Project presentation – 4/2

# Machine Learning Training

## Data/Input



## Training algorithm

Iterative training

1. Select initial *k* data points
   $m_1^1\ m_2^1 \ldots \ldots . m_k^1$
2. For each round *t* assign each data point to the nearest cluster *I*

$$C_i^t = \{x_p : \| x_p - m_i^t \|^2 \leq \| x_p - m_j^t \|^2\ \forall\ 1 \leq j \leq k\}$$

3. Recompute the means of the clusters

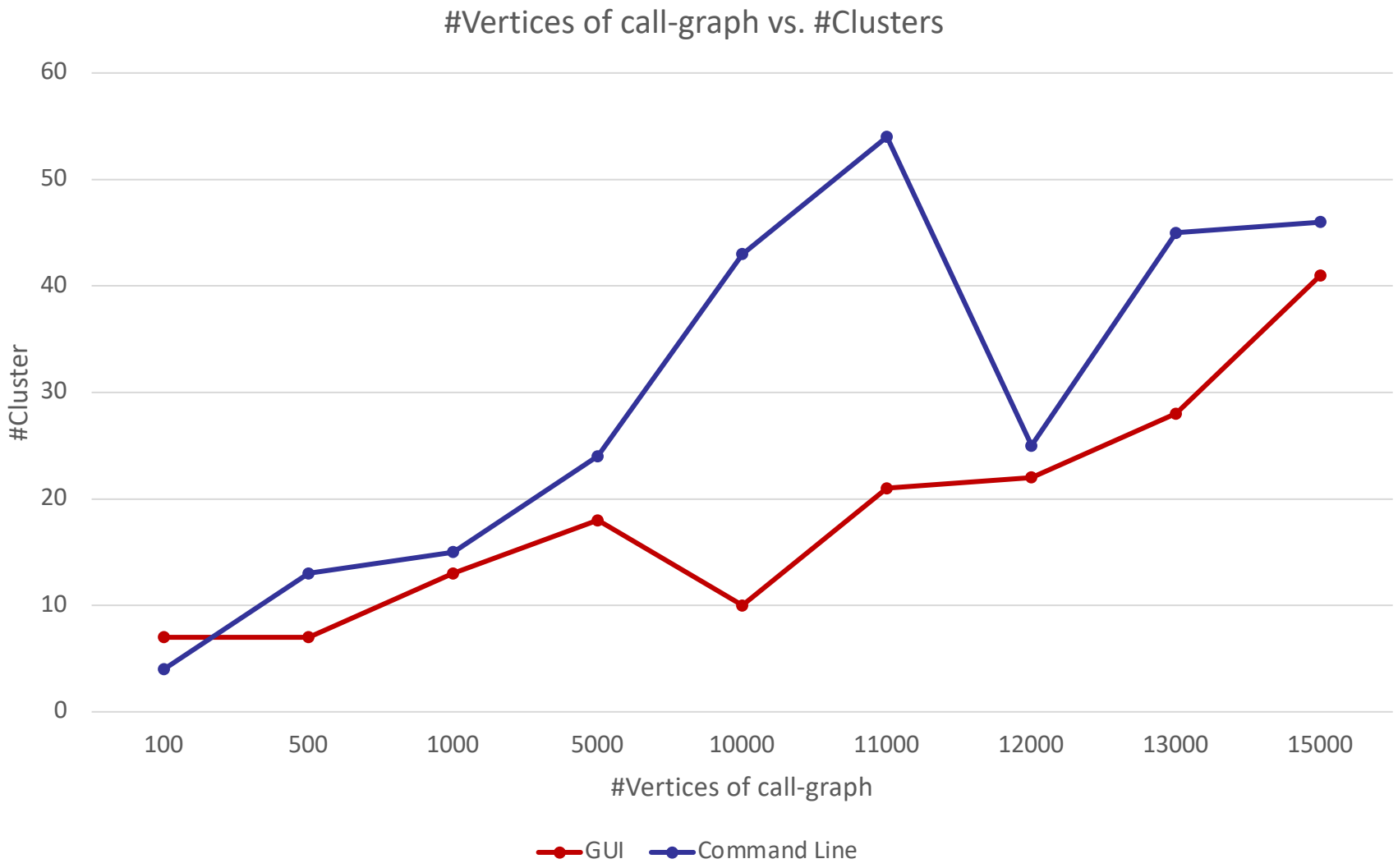
$$m_i^{t+1} = \frac{1}{|C_i^t|} \sum_{x_j \epsilon C_i^t} x_j$$

4. Stop when the assignment no longer change

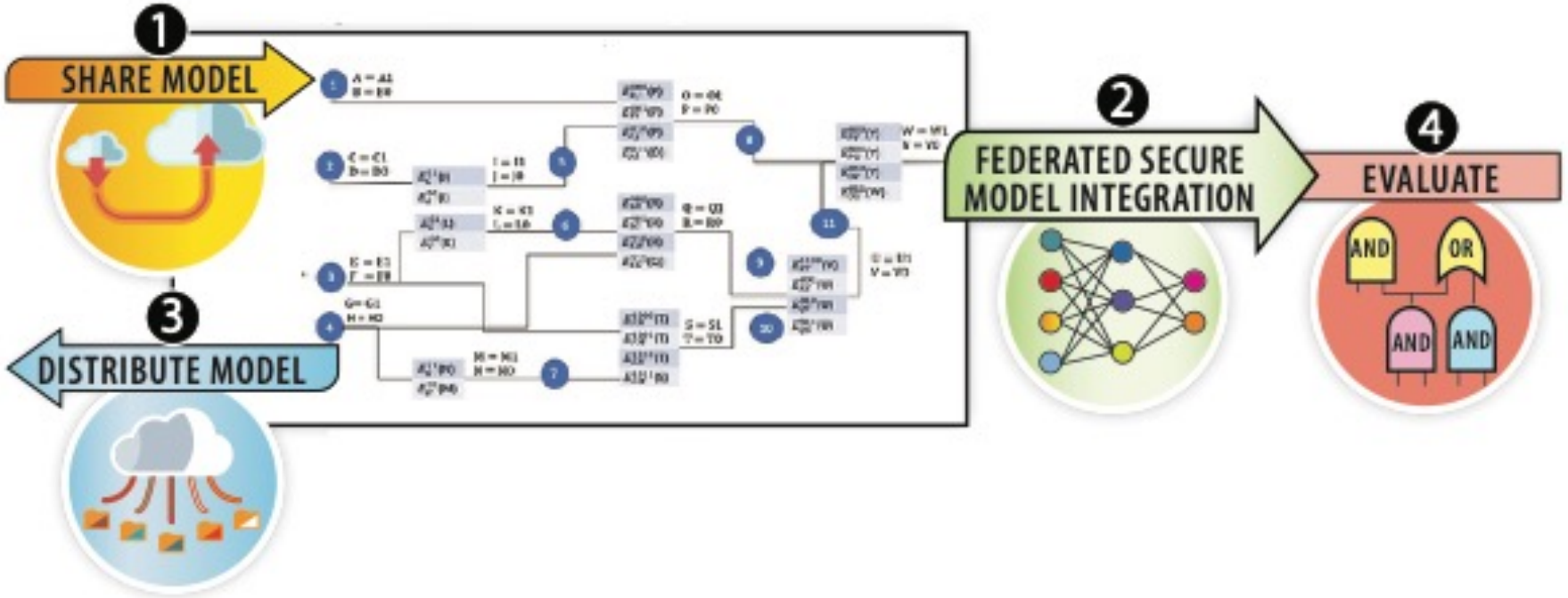
## Quantity to minimize

$$\sum_{i=1}^{k} \sum_{x \epsilon S_i} \| x - y_i \|^2$$

# Challenge 1: Large Size of the Data



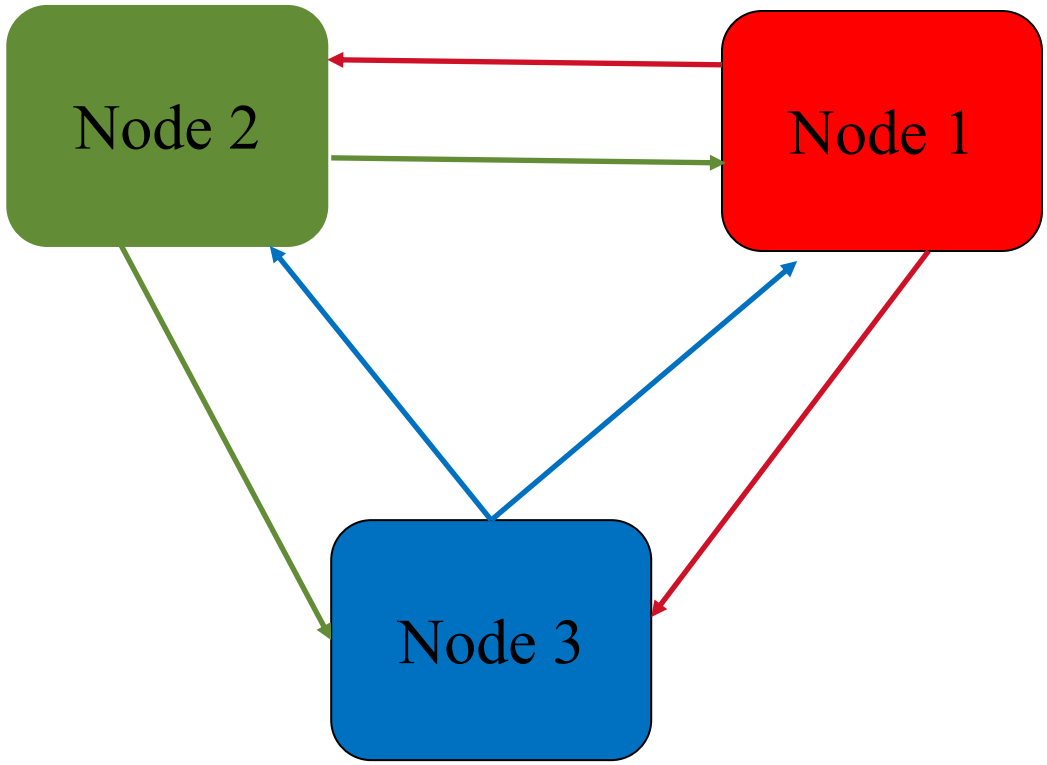
# Challenge 2: Collaboration Using Private Data




Ownership, Privacy, and locality of the data

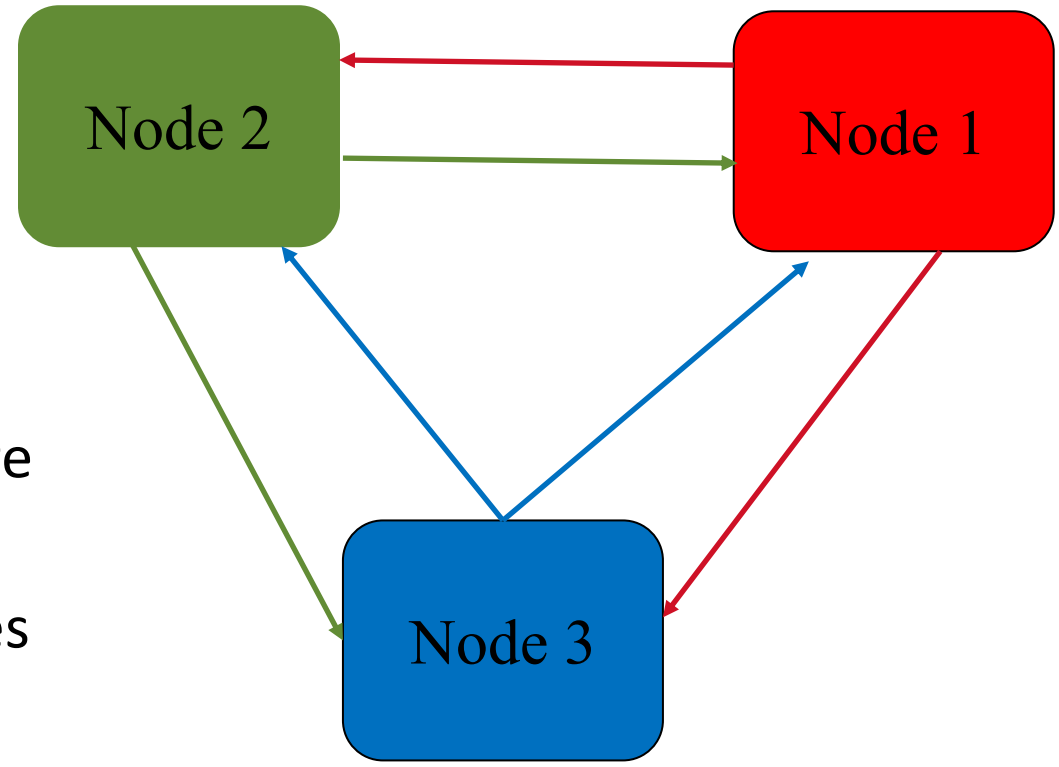
# Federated Learning

Federated Learning is a distributed machine learning approach that enables training a decentralized data residing in a distributed system.



# Federated Learning

- Collaboration in the learning architectures varies




Node tasks:

1. Train model on its data
2. Distribute its model share
3. Aggregate the model shares of the other nodes
4. Go to 1

The protocol must terminate

# Federated Learning vs Regular ML

Algorithm

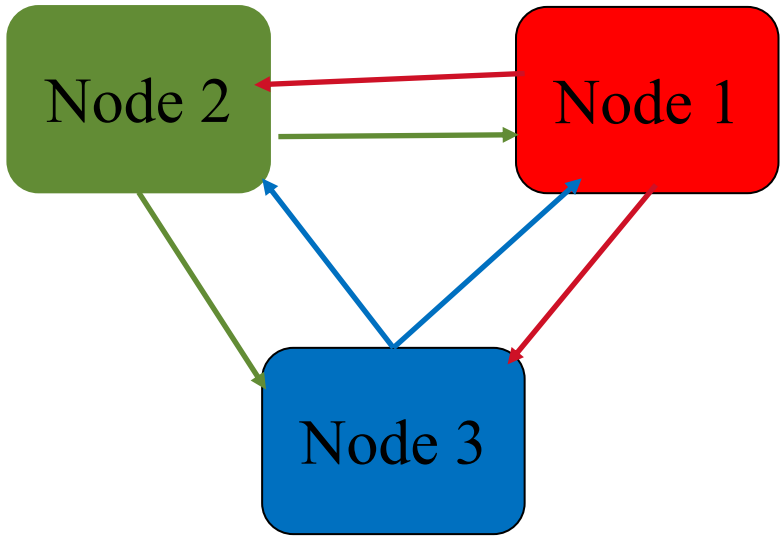
1. Select initial *k* data points $m_1^1\ m_2^1 \dots \dots . m_k^1$
2. For each round *t* assign each data point to the nearest cluster *I*

$$C_i^t = \{x_p : \| x_p - m_i^t \|^2 \leq \| x_p - m_j^t \|^2\ \forall\ 1 \leq j \leq k\}$$

3. Recompute the means of the clusters

$$m_i^{t+1} = \frac{1}{|C_i^t|} \sum_{x_j \epsilon C_i^t} x_j$$

4. Stop when the assignment no longer change




Computation

Communication + Computation

# Federated Learning vs Regular ML

Phases of federated learning (training, aggregation, and distribution) depend on the ML algorithm.

K-means Algorithm

1. Select initial *k* data points $m_1^1\ m_2^1 \ldots \ldots . m_k^1$
2. For each round *t* assign each data point to the nearest cluster *I*

$$C_i^t = \{x_p : \| x_p - m_i^t \|^2 \leq \| x_p - m_j^t \|^2 \ \forall\ 1 \leq j \leq k\}$$

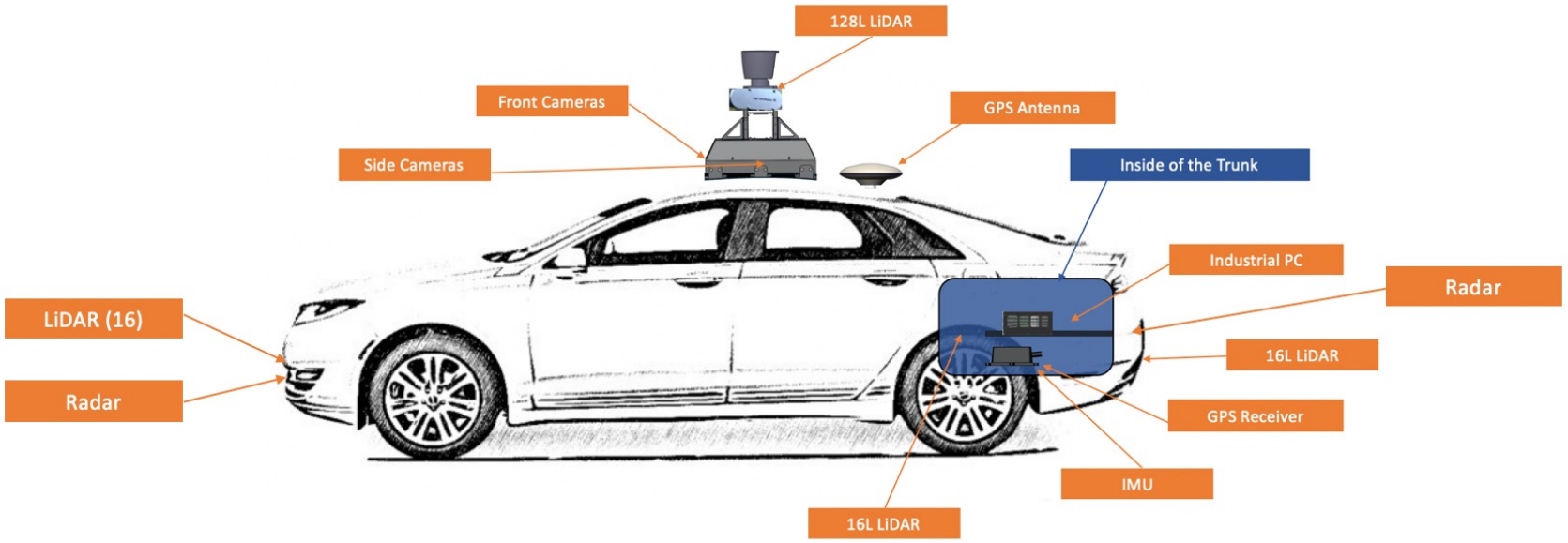
3. Recompute the means of the clusters

$$m_i^{t+1} = \frac{1}{|C_i^t|} \sum_{x_j \epsilon C_i^t} x_j$$

4. Stop when the assignment no longer change

# Uses of Federated Learning

Should we use federated learning for object identification?

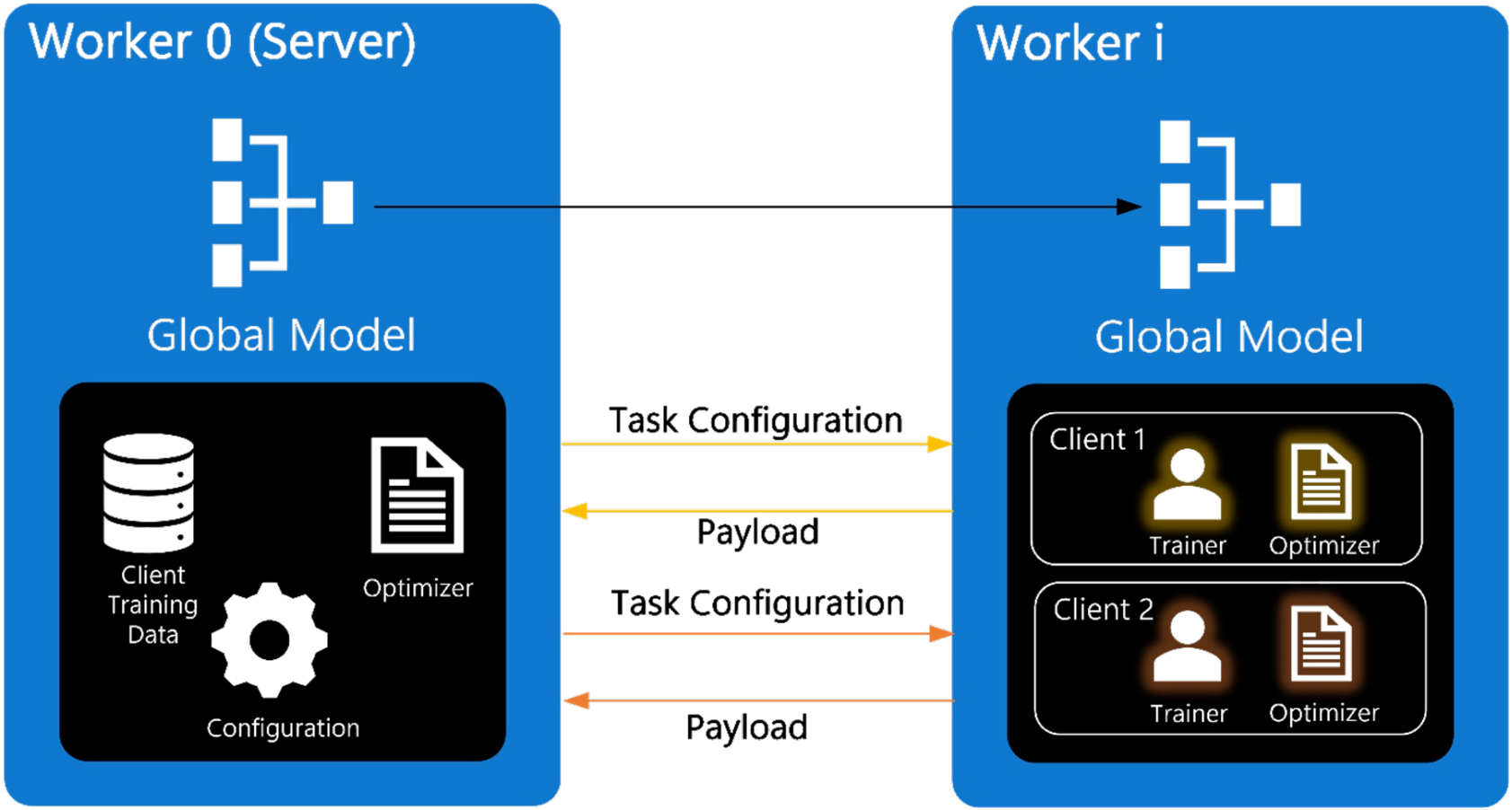


# Uses of Federated Learning

Give two examples where federated learning is

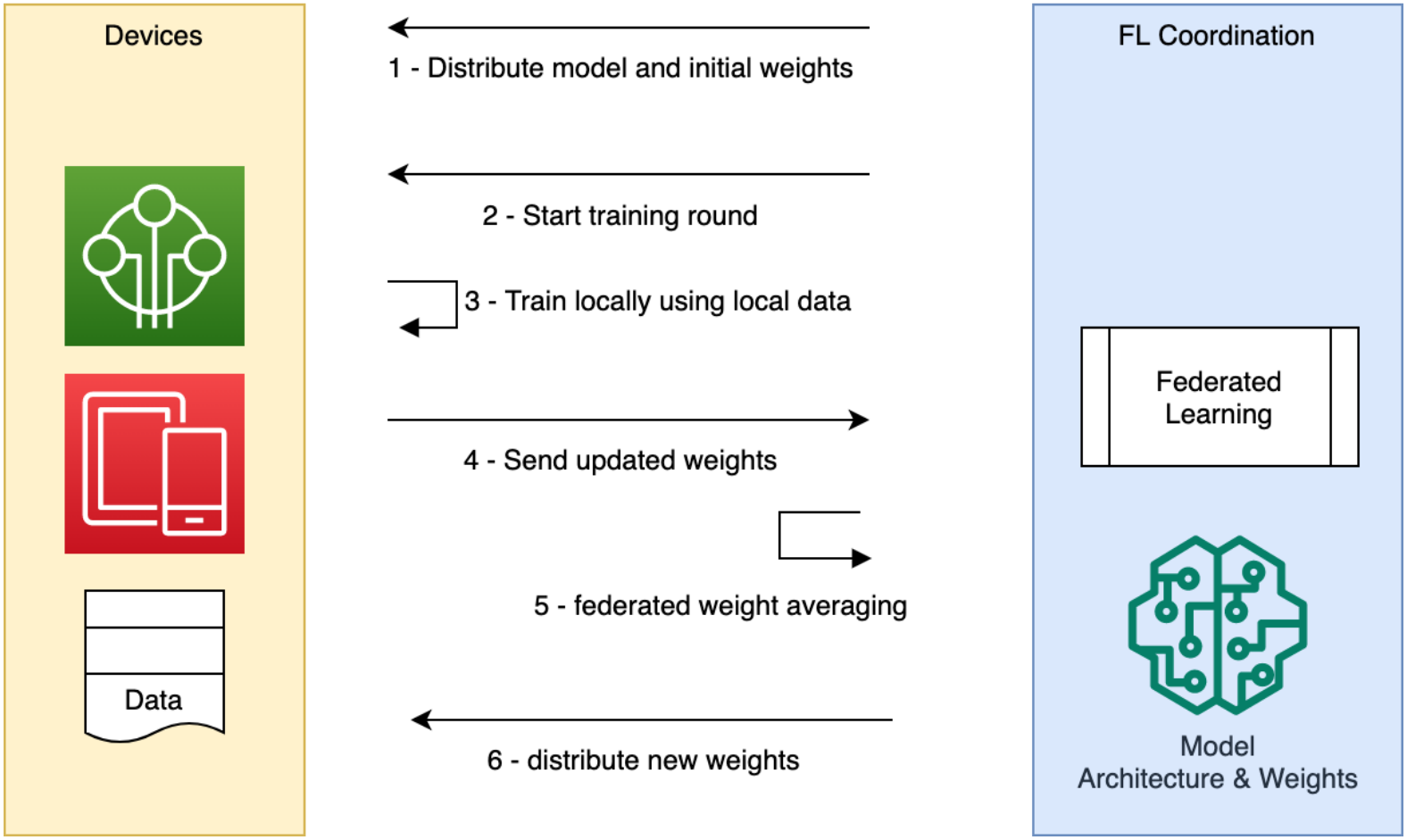
1. Not an option
2. Could be preferred

# Federated Learning Platform - Microsoft




https://www.microsoft.com/en-us/research/blog/flute-a-scalable-federated-learning-simulation-platform/

# Federated Learning Platform - Amazon




https://aws.amazon.com/blogs/architecture/applying-federated-learning-for-ml-at-the-edge/

# Uses of Federal Learning

Federal learning suits the following scenarios?

1. Build ML model from sensitive data
2. Build ML model using data controlled by a set of partners
3. The dataset is not large
4. Tolerate accuracy losses of 10%
5. We need to label the data

Which of the scenarios is correct?

Thank you

Any Question?