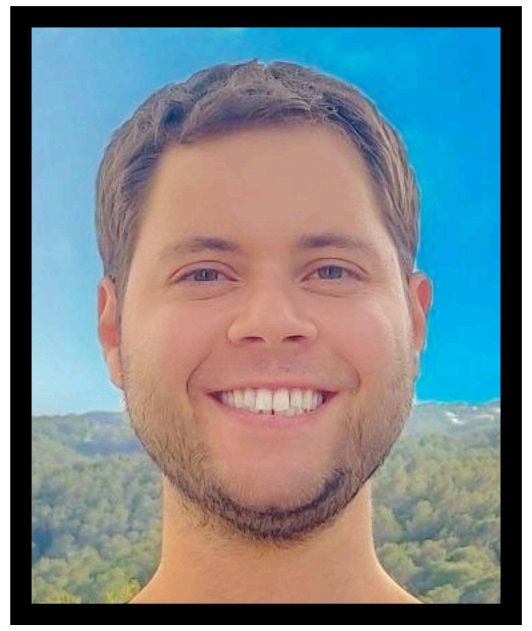
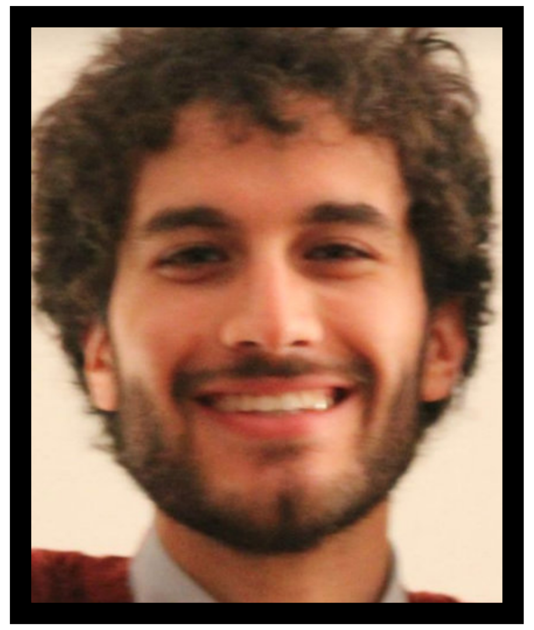
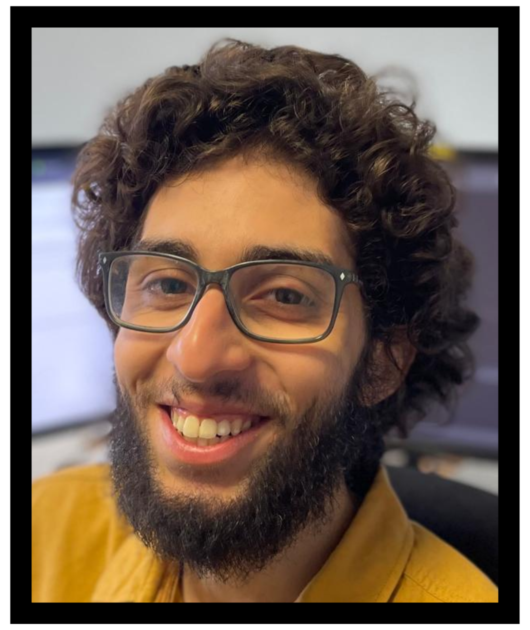
FL under intermittent and correlated client availability

Angelo Rodio

Francescomaria Faticanti

Othmane Marfoq

Giovanni Neglia

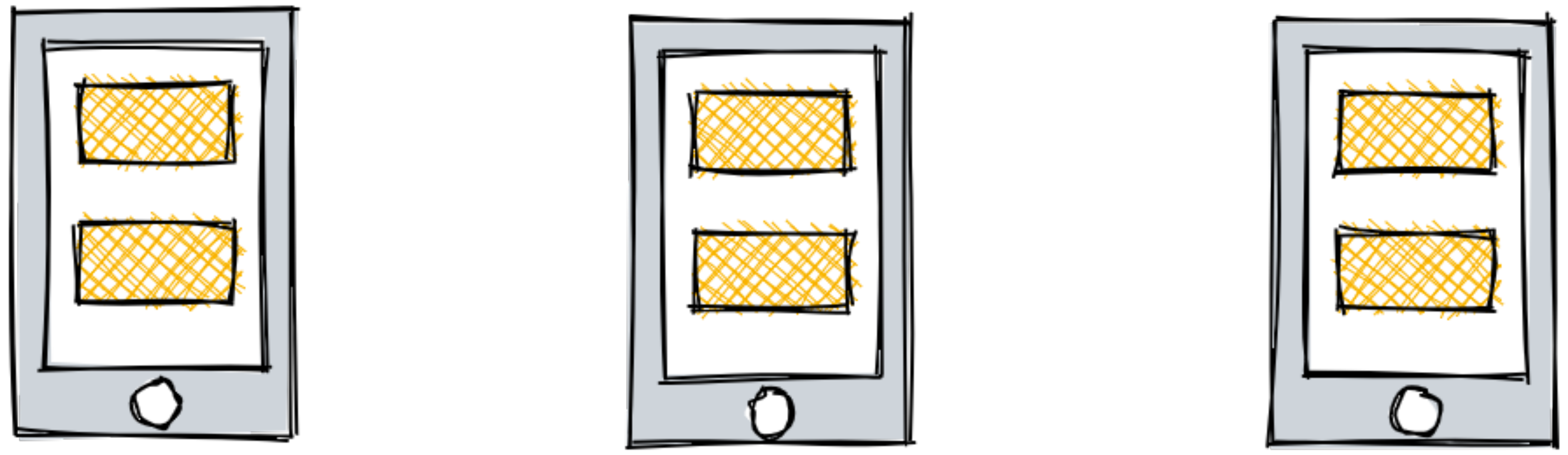
Emilio Leonardi

Inria

UNIVERSITÉ CÔTE D'AZUR

3iA Côte d'Azur
Interdisciplinary Institute for Artificial Intelligence

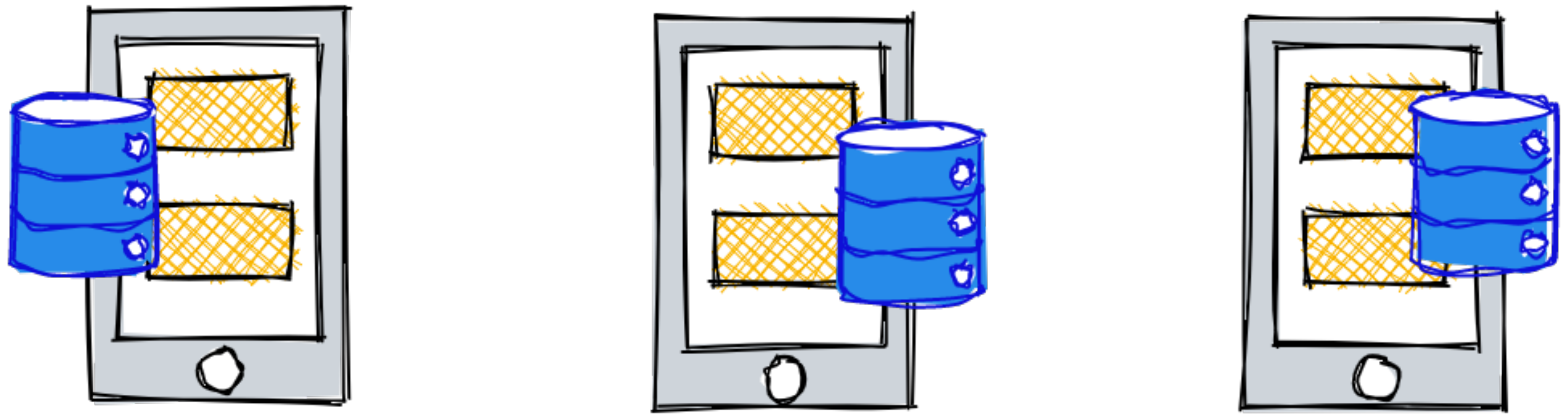
Federated Learning



Clients $k = 1, \ldots, K$

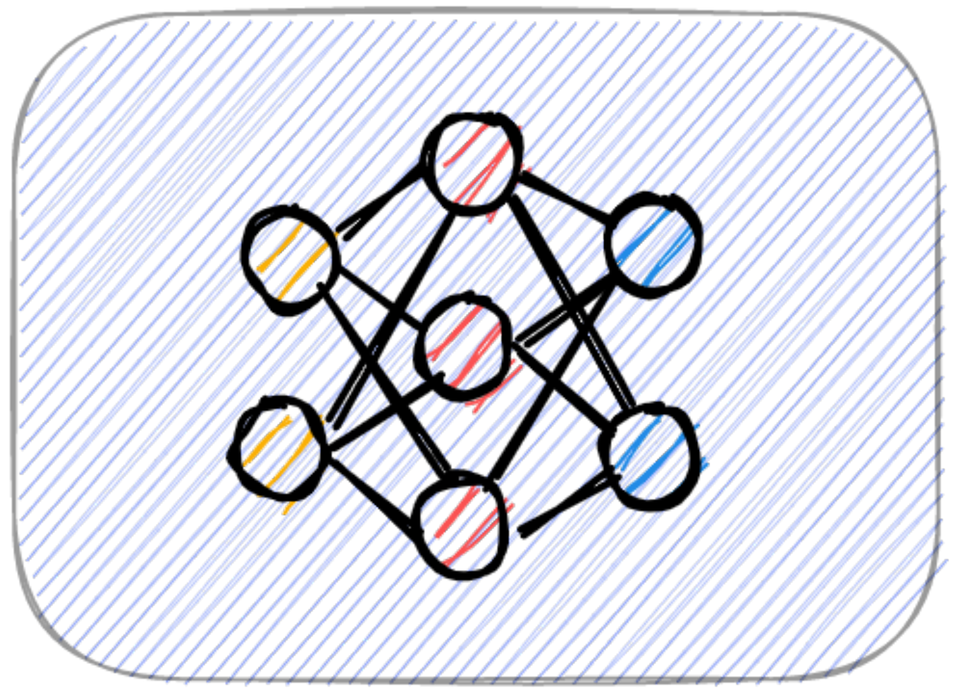
Federated Learning

Dataset $D_k = \{\xi_{kl}\}_{l=1}^{n_k}$



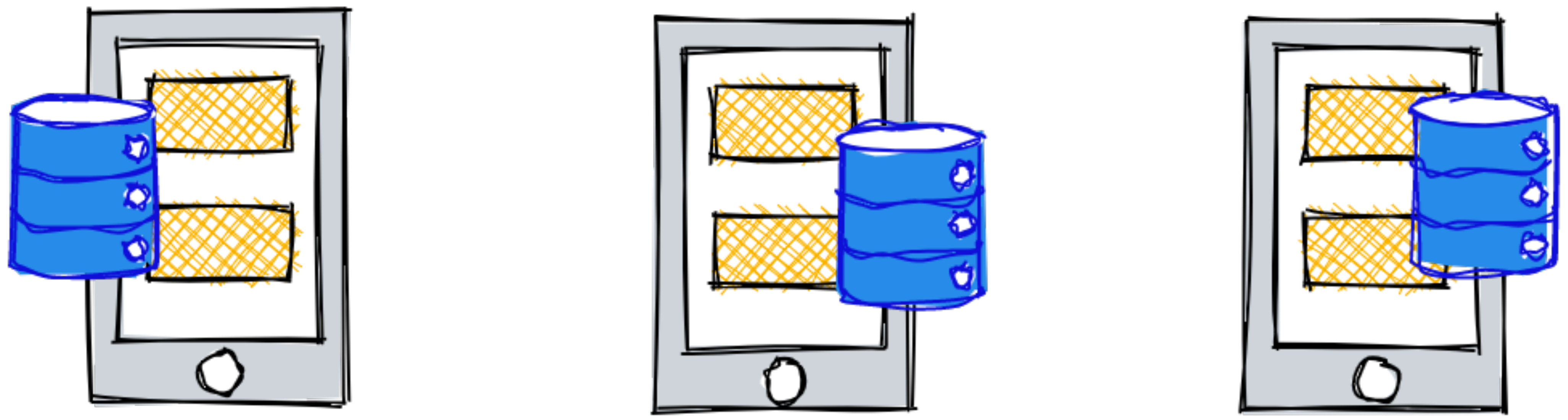
Clients $k = 1, \ldots, K$

Federated Learning



Global model
$\boldsymbol{w} \in \mathbb{R}^d$

Dataset $D_k = \{\xi_{kl}\}_{l=1}^{n_k}$

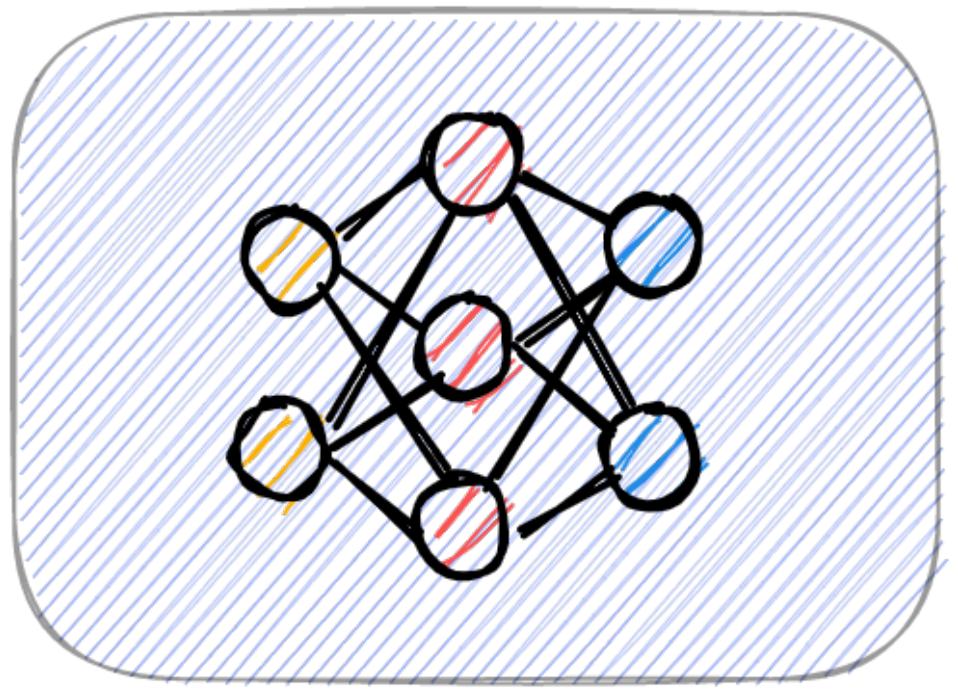


Clients $k = 1, \dots, K$

Federated Learning

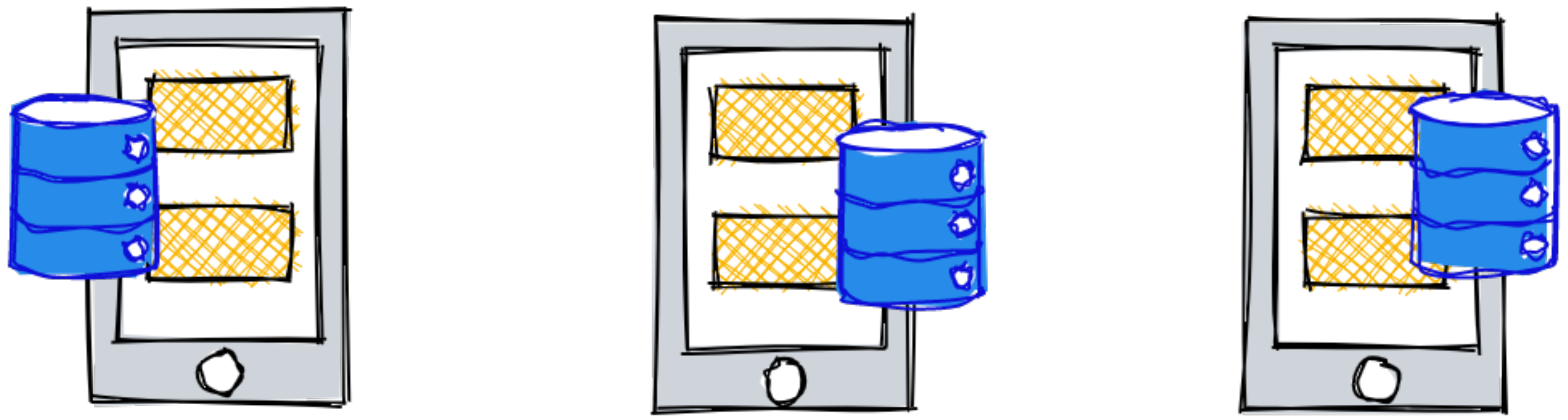
Solve the optimization problem

$$\min_{\boldsymbol{w}} F(\boldsymbol{w}) = \sum_{k=1}^{K} \alpha_k \, F_k(\boldsymbol{w})$$



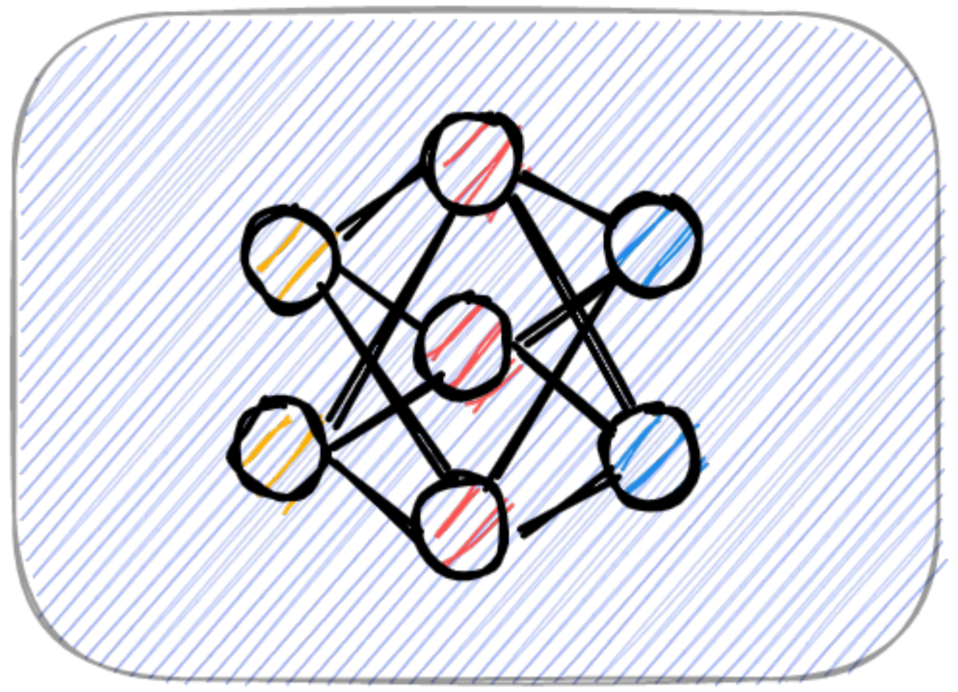
where

$$F_k(\boldsymbol{w}) = \frac{1}{n_k} \sum_{l=1}^{n_k} \ell(\boldsymbol{w}, \xi_{kl})$$



Federated Learning

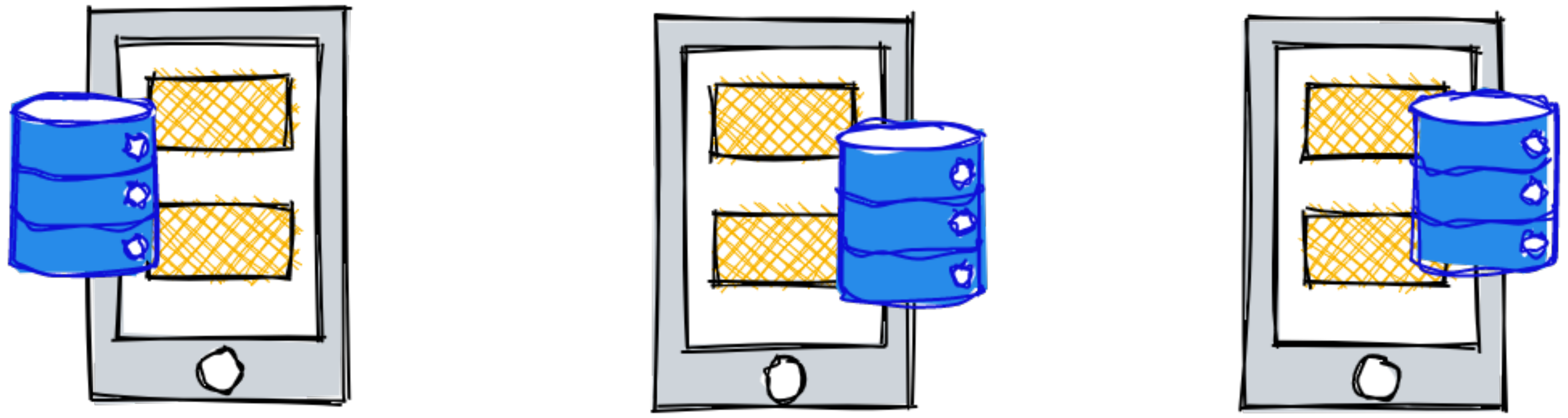
Solve the optimization problem



$$\min_{\boldsymbol{w}} F(\boldsymbol{w}) = \sum_{k=1}^{K} \alpha_k \, F_k(\boldsymbol{w})$$

$\boldsymbol{\alpha}$: importance weights

where

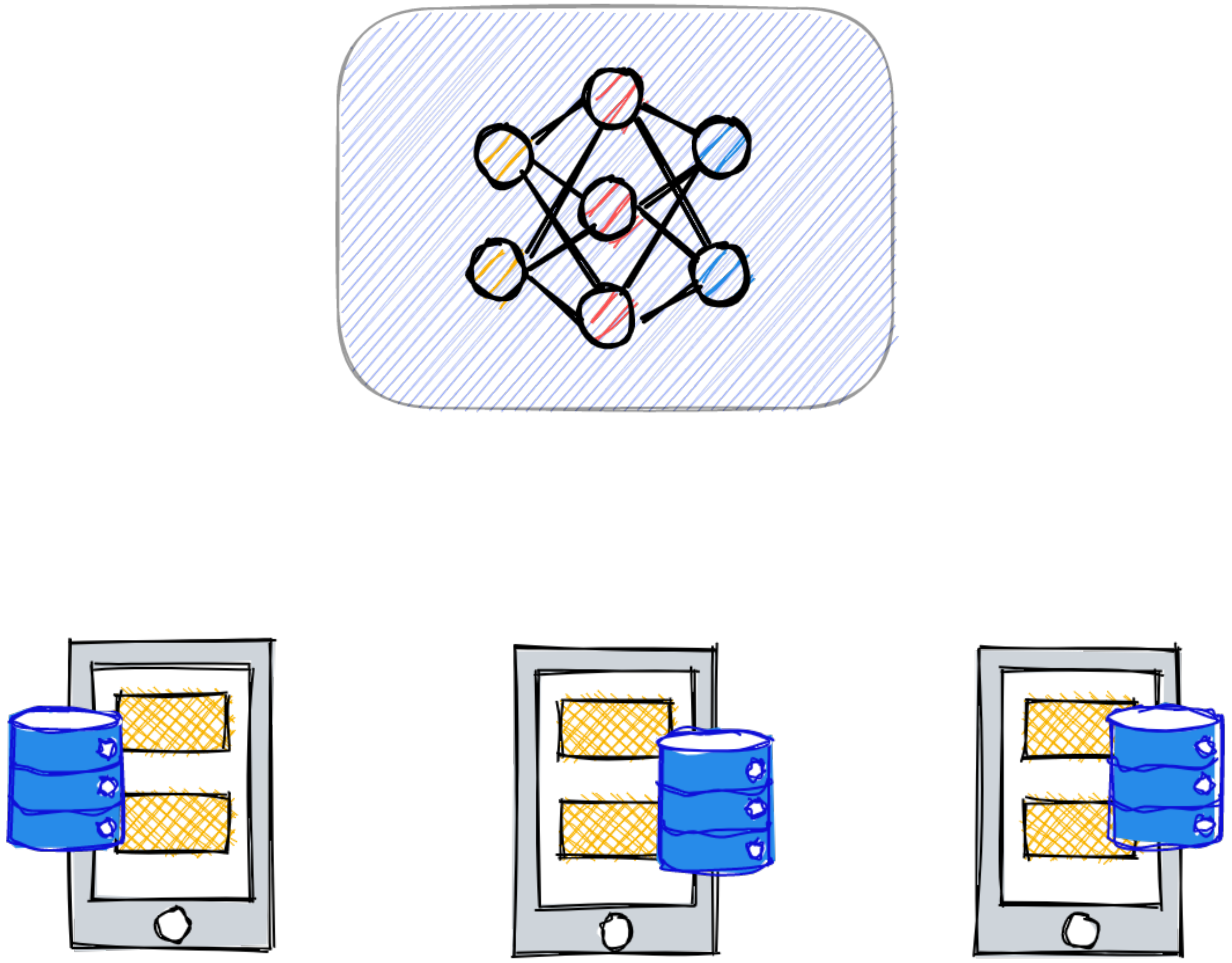


$$F_k(\boldsymbol{w}) = \frac{1}{n_k} \sum_{l=1}^{n_k} \ell(\boldsymbol{w}, \xi_{kl})$$

Federated Averaging (FedAvg)

A_t: set of active clients at time t

for $t \in \{0, \ldots, T-1\}$ **do**:

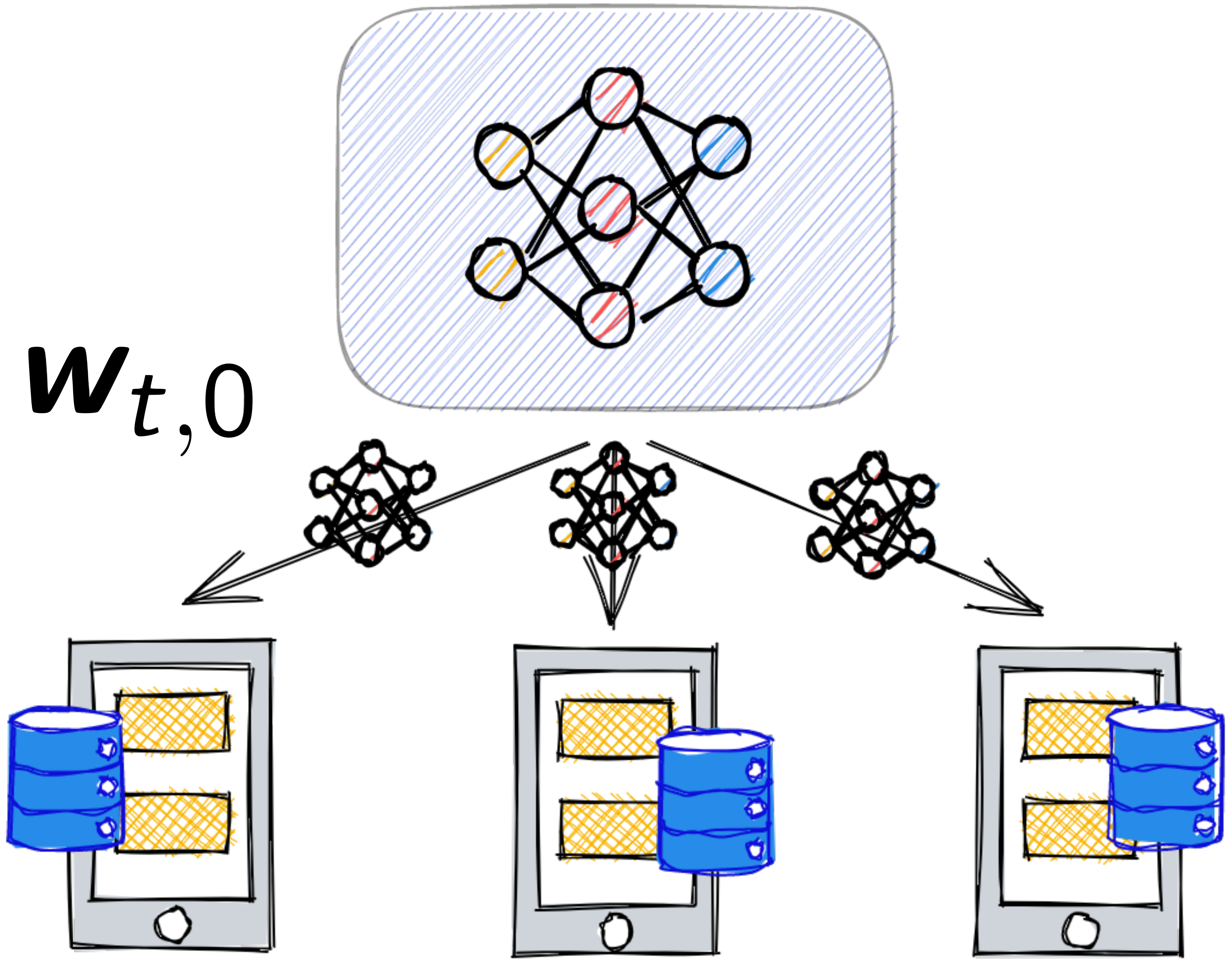


Federated Averaging (FedAvg)

A_t: set of active clients at time t

for $t \in \{0, \ldots, T-1\}$ **do**:

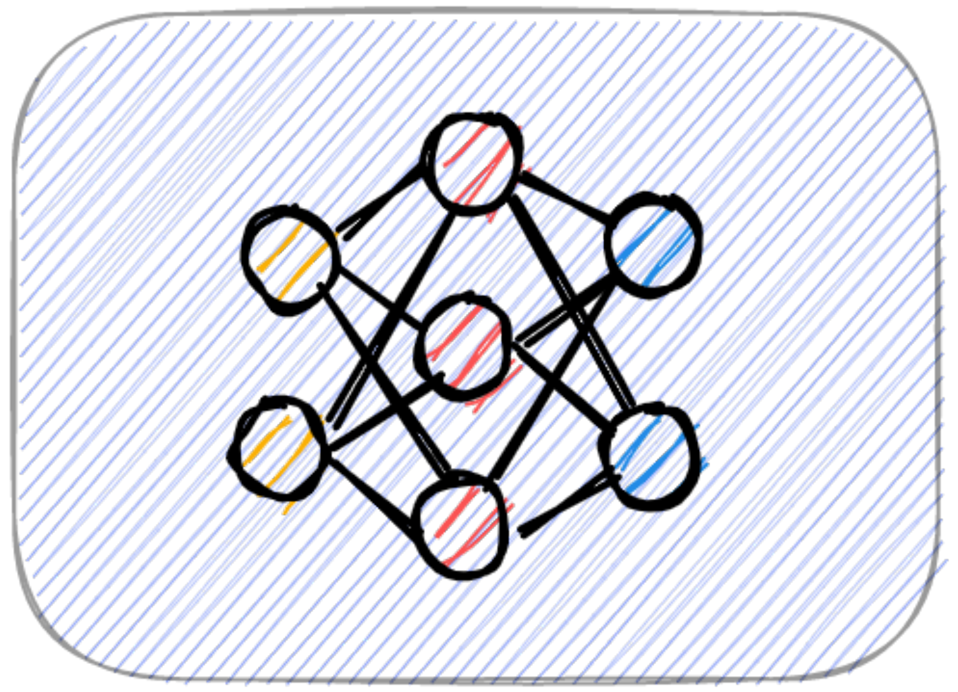
(1)



Federated Averaging (FedAvg)

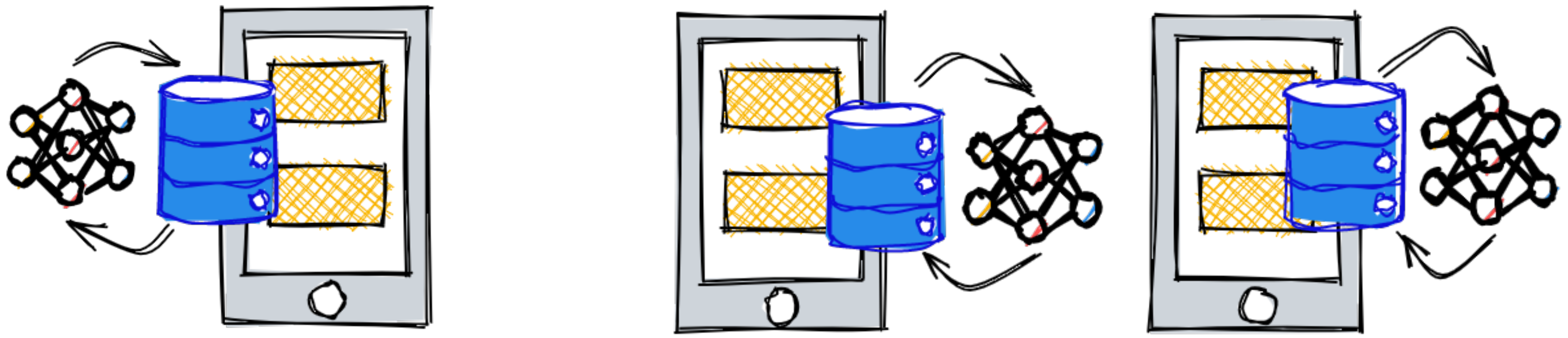
A_t: set of active clients at time t

for $t \in \{0, \ldots, T-1\}$ **do**:



(2)

$\boldsymbol{w}_{t,E}^{k}$

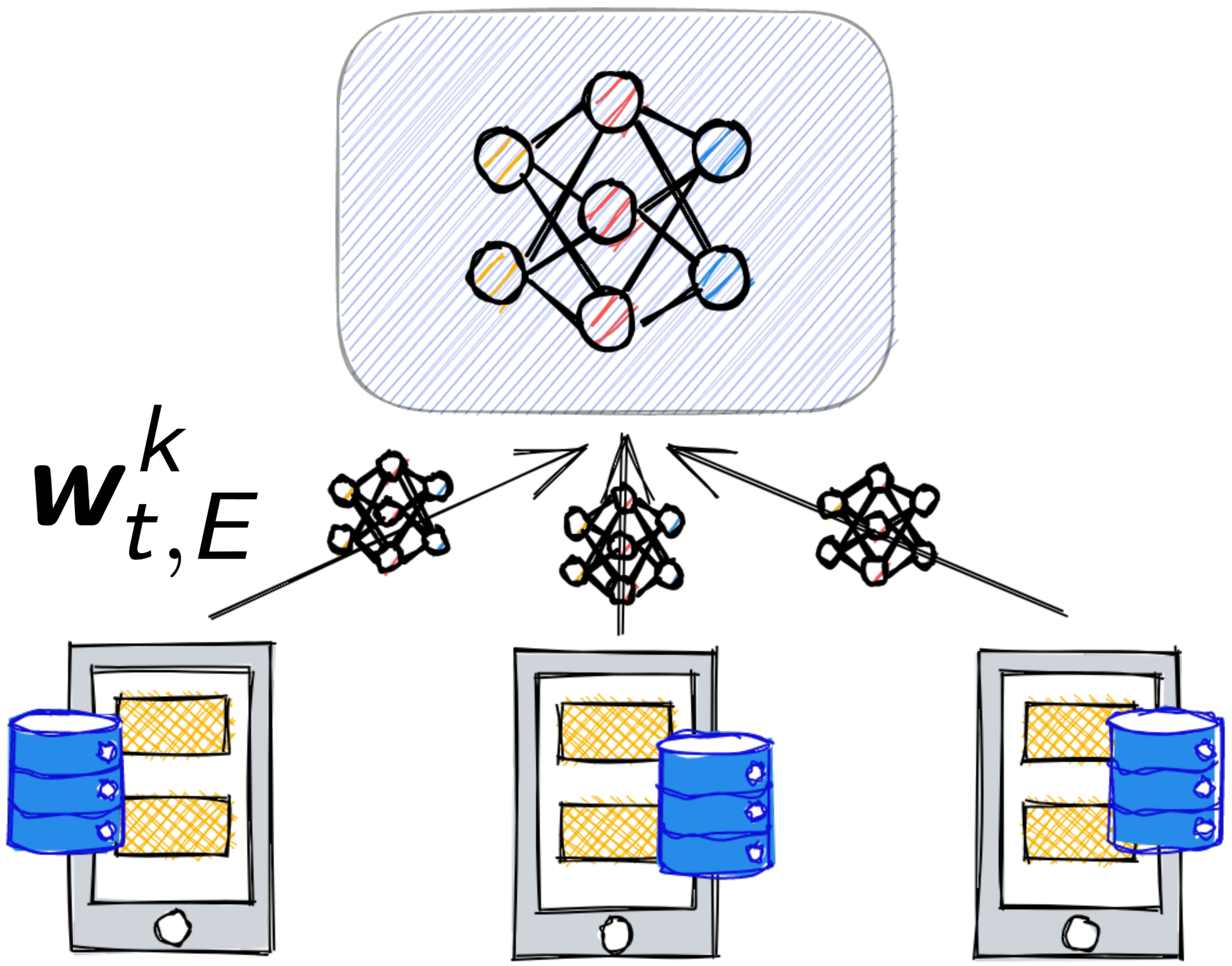


Federated Averaging (FedAvg)

A_t: set of active clients at time t

for $t \in \{0, \dots, T-1\}$ **do**:

(3)

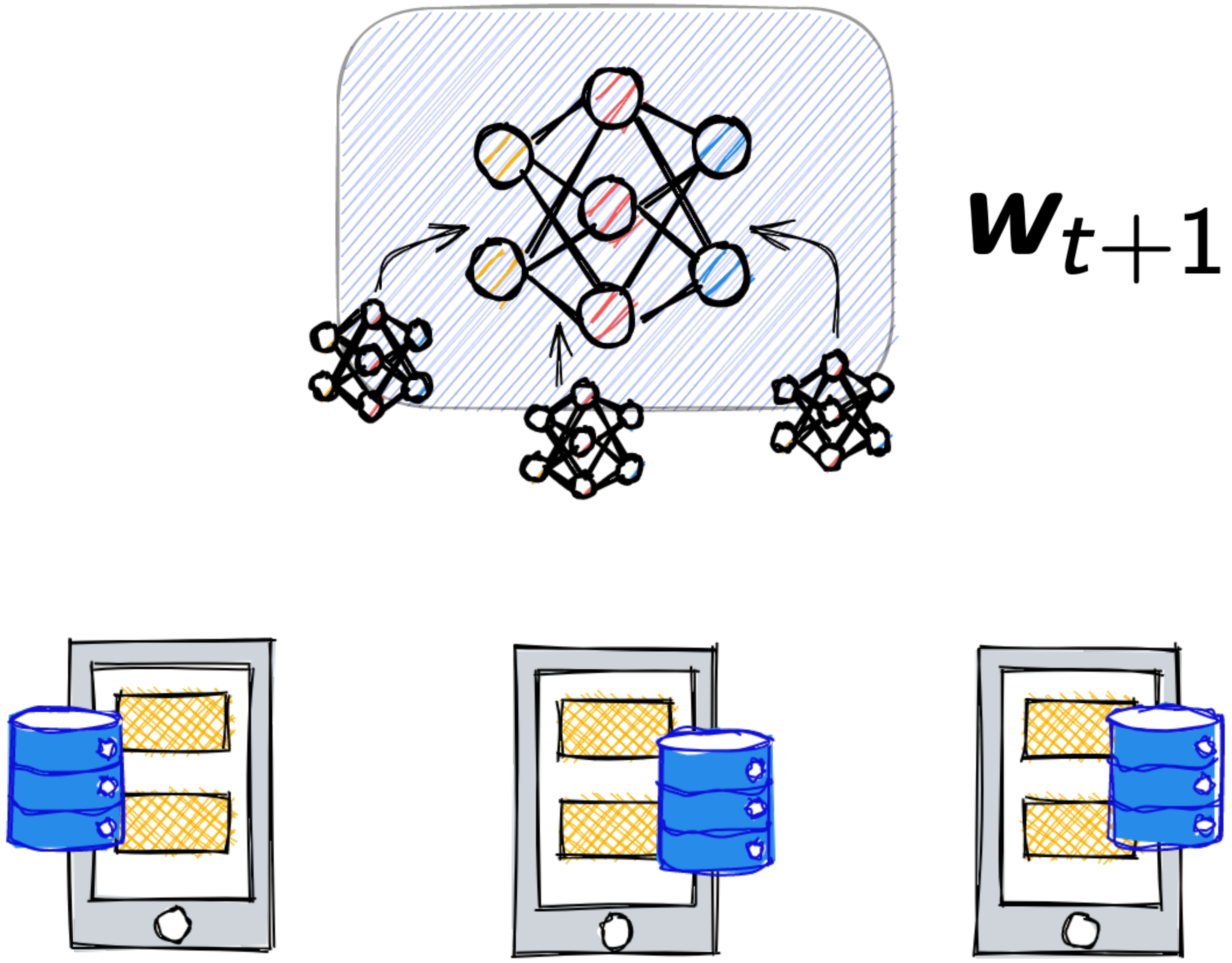


Federated Averaging (FedAvg)

A_t: set of active clients at time t

for $t \in \{0, \ldots, T-1\}$ **do**:

(4)



$\boldsymbol{w}_{t+1,0}$

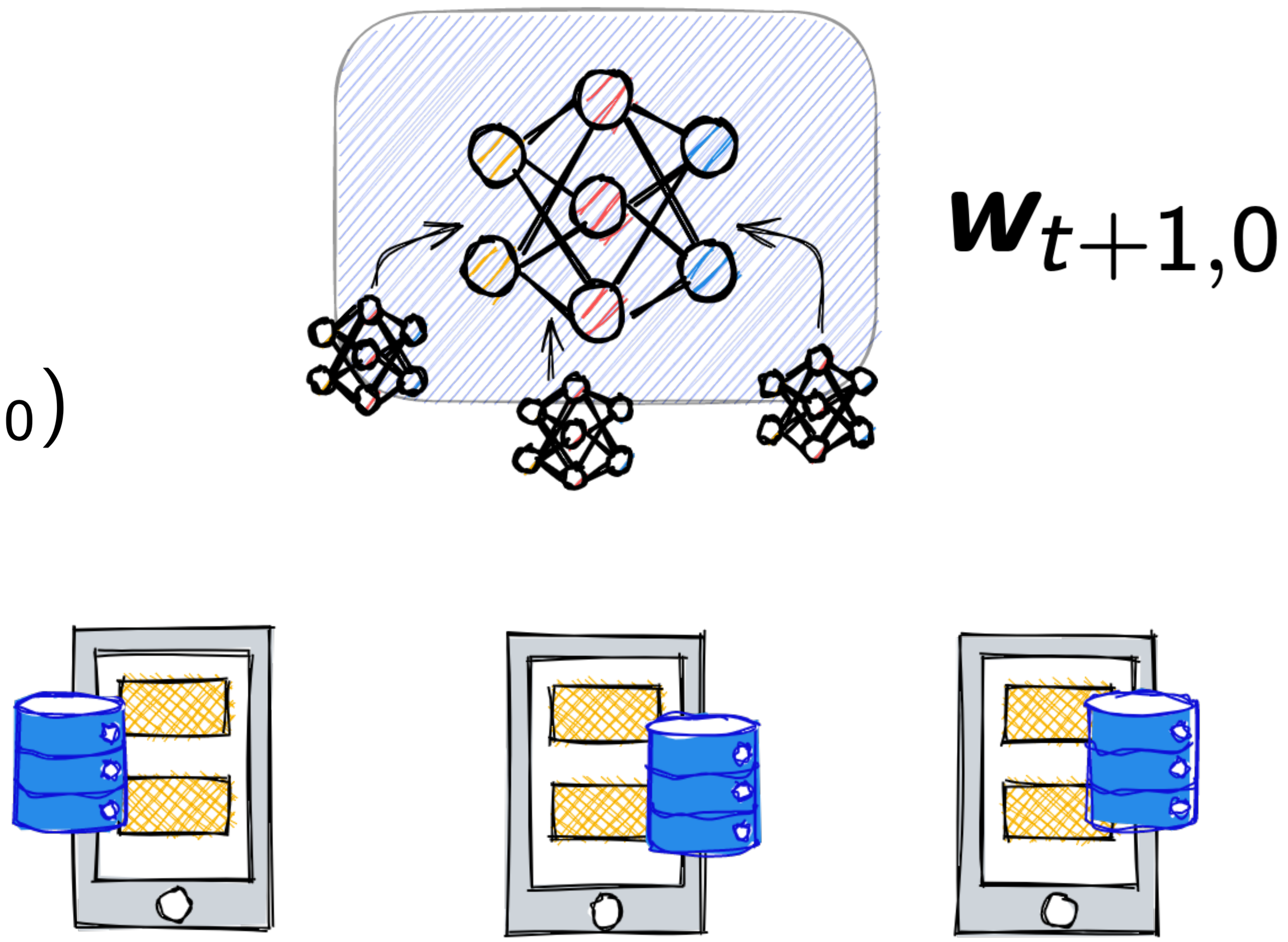
Federated Averaging (FedAvg)

A_t: set of active clients at time t

for $t \in \{0, \dots, T-1\}$ **do**:

(4) $\boldsymbol{w}_{t+1,0} = \boldsymbol{w}_{t,0} + \sum_{k \in A_t} q_k (\boldsymbol{w}^k_{t,E} - \boldsymbol{w}_{t,0})$

$\boldsymbol{q}$: aggregation weights



Intermittent and Correlated Client Availability

- **Intermittent:**

  clients are not always active

- **Correlated:**

  the activity of a client is correlated over time

  the activity is correlated across clients

Intermittent and Correlated Client Availability

- **Intermittent:**

  clients are not always active

- **Correlated:**

  the activity of a client is correlated over time

  the activity is correlated across clients

Contributions

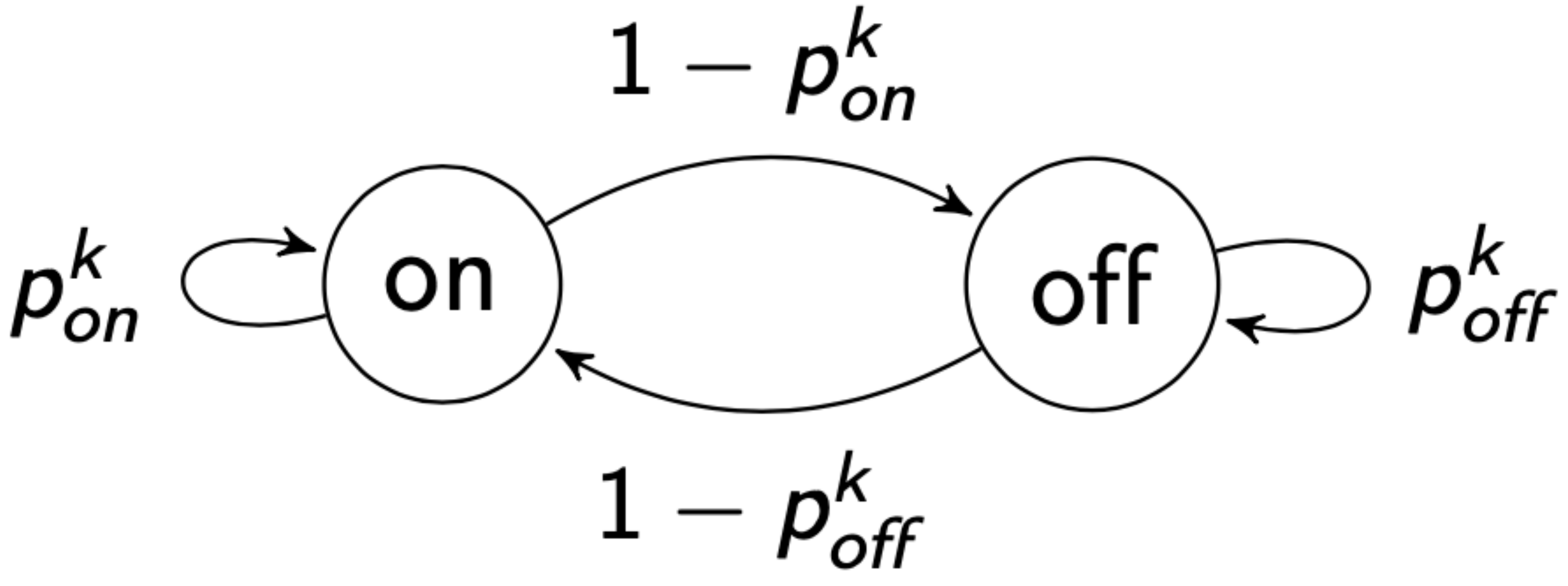
1. Understanding the effects on FL training
2. Client sampling and aggregation strategies

Main assumption

- A_t: set of active clients at time t
- Clients' activities follow a discrete-time Markov chain $(A_t)_{t \geq 0}$ with transition matrix $\boldsymbol{P}$ and stationary distribution $\boldsymbol{\pi}$

Main assumption

- A_t: set of active clients at time t
- Clients' activities follow a discrete-time Markov chain $(A_t)_{t\geq 0}$ with transition matrix $\boldsymbol{P}$ and stationary distribution $\boldsymbol{\pi}$



$$\boldsymbol{P} = \bigotimes_{k=1}^{K} \boldsymbol{P_k} \qquad \boldsymbol{\pi} = \bigotimes_{k=1}^{K} \boldsymbol{\pi_k}$$

$$\lambda(\boldsymbol{P}) = \max_{k \in [K]} \lambda(\boldsymbol{P_k})$$

Contributions

1. Understanding the effects on FL training

2. Client sampling and aggregation strategies

The effect of Intermittent Availability

Under intermittent availability $\boldsymbol{\pi}$, FedAvg converges to a biased objective $F_B(\boldsymbol{w})$

$$F_B(\boldsymbol{w}) := \sum_{k=1}^{K} p_k \, F_k(\boldsymbol{w}), \; p_k = \frac{\pi_k q_k}{\langle \boldsymbol{\pi}, \boldsymbol{q} \rangle} \qquad \neq \qquad F(\boldsymbol{w}) := \sum_{k=1}^{K} \alpha_k \, F_k(\boldsymbol{w})$$

$\boldsymbol{p}$: biased importance

$\boldsymbol{\alpha}$: true importance

The effect of Intermittent Availability

Under intermittent availability $\boldsymbol{\pi}$, FedAvg converges to a biased objective $F_B(\boldsymbol{w})$

$$F_B(\boldsymbol{w}) := \sum_{k=1}^{K} p_k \, F_k(\boldsymbol{w}), \; p_k = \frac{\pi_k q_k}{\langle \boldsymbol{\pi}, \boldsymbol{q} \rangle} \qquad \neq \qquad F(\boldsymbol{w}) := \sum_{k=1}^{K} \alpha_k \, F_k(\boldsymbol{w})$$

$\boldsymbol{p}$: biased importance

$\boldsymbol{\alpha}$: true importance

Unbiased aggregation strategy: $\boldsymbol{q} \propto \dfrac{\boldsymbol{\alpha}}{\boldsymbol{\pi}}$

The effect of Correlated Availability

$$\mathbb{E}[F_B(\bar{\boldsymbol{w}}_{T,0}) - F_B^*] \leq \mathcal{O}\left(\frac{1}{\sqrt{T}} \cdot \frac{1}{\ln(1/\lambda(\boldsymbol{P}))}\right)$$

where T is the total communication rounds and $\lambda(\boldsymbol{P})$ quantifies the correlation

The effect of Correlated Availability

$$\mathbb{E}[F_B(\bar{\boldsymbol{w}}_{T,0}) - F_B^*] \leq \mathcal{O}\left(\frac{1}{\sqrt{T}} \cdot \frac{1}{\ln(1/\lambda(\boldsymbol{P}))}\right)$$

where T is the total communication rounds and $\lambda(\boldsymbol{P})$ quantifies the correlation

Correlation slows down the convergence

The effect of Intermittent + Correlated Availability

$$\epsilon(\boldsymbol{q}) := F(\boldsymbol{w}_{T,0}) - F^* \;\leq\; \underbrace{\mathcal{O}\left(F_B(\boldsymbol{w}_{T,0}) - F_B^*\right)}_{:= \epsilon_{\text{opt}}(\boldsymbol{q})} \;+\; \underbrace{\mathcal{O}\left(d_{TV}^2(\boldsymbol{\alpha}, \boldsymbol{p})\Gamma\right)}_{:= \epsilon_{\text{bias}}(\boldsymbol{q})}$$

where $p_k = \frac{\pi_k q_k}{\langle \boldsymbol{\pi}, \boldsymbol{q} \rangle}$, $d_{TV}(\boldsymbol{\alpha}, \boldsymbol{p}) = \frac{1}{2} \sum_{k=1}^{K} |\alpha_k - p_k|$, and $\Gamma = \max_{k \in [K]} \{F_k(\boldsymbol{w}_B^*) - F_k^*\}$

The effect of Intermittent + Correlated Availability

$$\epsilon(\boldsymbol{q}) := F(\boldsymbol{w}_{T,0}) - F^* \leq \underbrace{\mathcal{O}\left(F_B(\boldsymbol{w}_{T,0}) - F_B^*\right)}_{:= \epsilon_{\text{opt}}(\boldsymbol{q})} + \underbrace{\mathcal{O}\left(d_{TV}^2(\boldsymbol{\alpha}, \boldsymbol{p})\Gamma\right)}_{:= \epsilon_{\text{bias}}(\boldsymbol{q})}$$

where $p_k = \frac{\pi_k q_k}{\langle \boldsymbol{\pi}, \boldsymbol{q} \rangle}$, $d_{TV}(\boldsymbol{\alpha}, \boldsymbol{p}) = \frac{1}{2}\sum_{k=1}^{K} |\alpha_k - p_k|$, and $\Gamma = \max_{k \in [K]}\{F_k(\boldsymbol{w}_B^*) - F_k^*\}$

- The convergence of the true objective $F(\boldsymbol{w})$ depends on $\boldsymbol{q}$

The effect of Intermittent + Correlated Availability

$$\epsilon(\boldsymbol{q}) := F(\boldsymbol{w}_{T,0}) - F^* \leq \underbrace{\mathcal{O}\left(F_B(\boldsymbol{w}_{T,0}) - F_B^*\right)}_{:= \epsilon_{\text{opt}}(\boldsymbol{q})} + \underbrace{\mathcal{O}\left(d_{TV}^2(\boldsymbol{\alpha}, \boldsymbol{p})\Gamma\right)}_{:= \epsilon_{\text{bias}}(\boldsymbol{q})}$$

where $p_k = \frac{\pi_k q_k}{\langle \boldsymbol{\pi}, \boldsymbol{q} \rangle}$, $d_{TV}(\boldsymbol{\alpha}, \boldsymbol{p}) = \frac{1}{2}\sum_{k=1}^{K} |\alpha_k - p_k|$, and $\Gamma = \max_{k \in [K]} \{F_k(\boldsymbol{w}_B^*) - F_k^*\}$

- The convergence of the true objective $F(\boldsymbol{w})$ depends on $\boldsymbol{q}$
- Propose a client aggregation strategy that minimizes $\epsilon(\boldsymbol{q})$

Contributions

1. Understanding the effects on FL training
2. Client sampling and aggregation strategies

Client sampling and aggregation strategies

$$\begin{aligned} \underset{\boldsymbol{q}}{\text{minimize}} \quad & \epsilon_{\mathsf{opt}}(\boldsymbol{q}) + \epsilon_{\mathsf{bias}}(\boldsymbol{q}) \\ \text{subject to} \quad & \boldsymbol{q} \geq 0, \; \|\boldsymbol{q}\|_1 = Q \end{aligned}$$

Client sampling and aggregation strategies

$$\begin{aligned} \underset{\boldsymbol{q}}{\text{minimize}} \quad & \epsilon_{\mathsf{opt}}(\boldsymbol{q}) + \epsilon_{\mathsf{bias}}(\boldsymbol{q}) \\ \text{subject to} \quad & \boldsymbol{q} \geq 0, \;\; \|\boldsymbol{q}\|_1 = Q \end{aligned}$$

- The solution of this optimization problem suggests:

Guidelines

Client sampling and aggregation strategies

$$\underset{\boldsymbol{q}}{\text{minimize}} \quad \epsilon_{\mathsf{opt}}(\boldsymbol{q}) + \epsilon_{\mathsf{bias}}(\boldsymbol{q})$$

$$\text{subject to} \quad \boldsymbol{q} \geq 0, \ \ \|\boldsymbol{q}\|_1 = Q$$

- The solution of this optimization problem suggests:

Guidelines

A. Some clients can be excluded from training, i.e., receive $q_k^* = 0$

Client sampling and aggregation strategies

$$\begin{aligned} \underset{\boldsymbol{q}}{\text{minimize}} \quad & \epsilon_{\text{opt}}(\boldsymbol{q}) + \epsilon_{\text{bias}}(\boldsymbol{q}) \\ \text{subject to} \quad & \boldsymbol{q} \geq 0, \ \|\boldsymbol{q}\|_1 = Q \end{aligned}$$

- The solution of this optimization problem suggests:

Guidelines

A. Some clients can be excluded from training, i.e., receive $q_k^* = 0$

B. Exclude clients with low availability π_k and large correlation $\lambda(\boldsymbol{P_k})$

Client sampling and aggregation strategies

$$\begin{aligned} \underset{\boldsymbol{q}}{\text{minimize}} \quad & \epsilon_{\mathsf{opt}}(\boldsymbol{q}) + \epsilon_{\mathsf{bias}}(\boldsymbol{q}) \\ \text{subject to} \quad & \boldsymbol{q} \geq 0, \ \ \|\boldsymbol{q}\|_1 = Q \end{aligned}$$

- The solution of this optimization problem suggests:

Guidelines

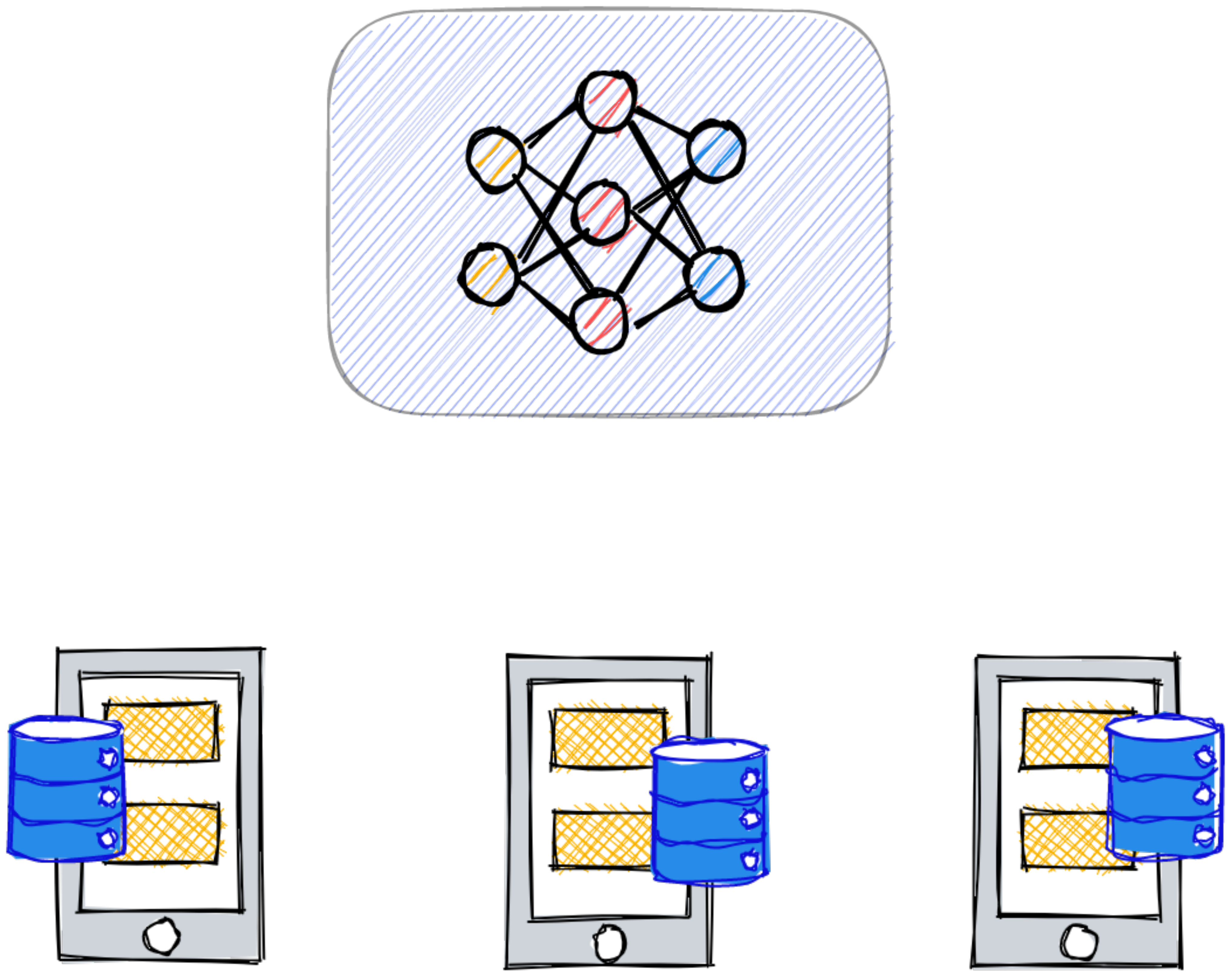
A. Some clients can be excluded from training, i.e., receive $q_k^* = 0$

B. Exclude clients with <u>low availability</u> π_k and <u>large correlation</u> $\lambda(\boldsymbol{P_k})$

C. Assign aggregations $q_k = \alpha_k / \pi_k$ to the included clients

Correlation-Aware FL (CA-Fed)

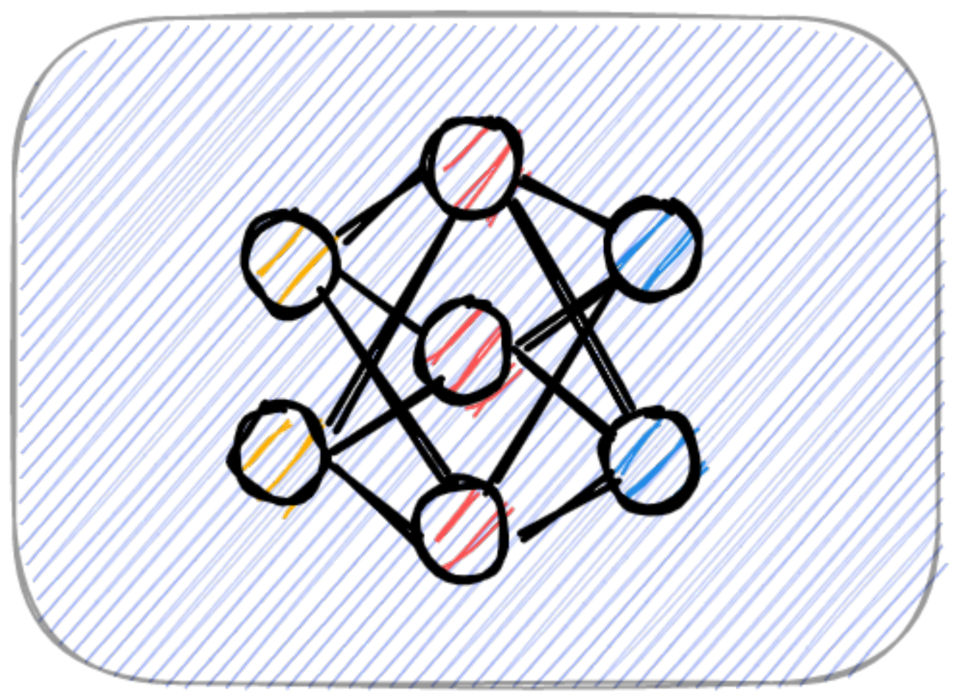
(1) The server

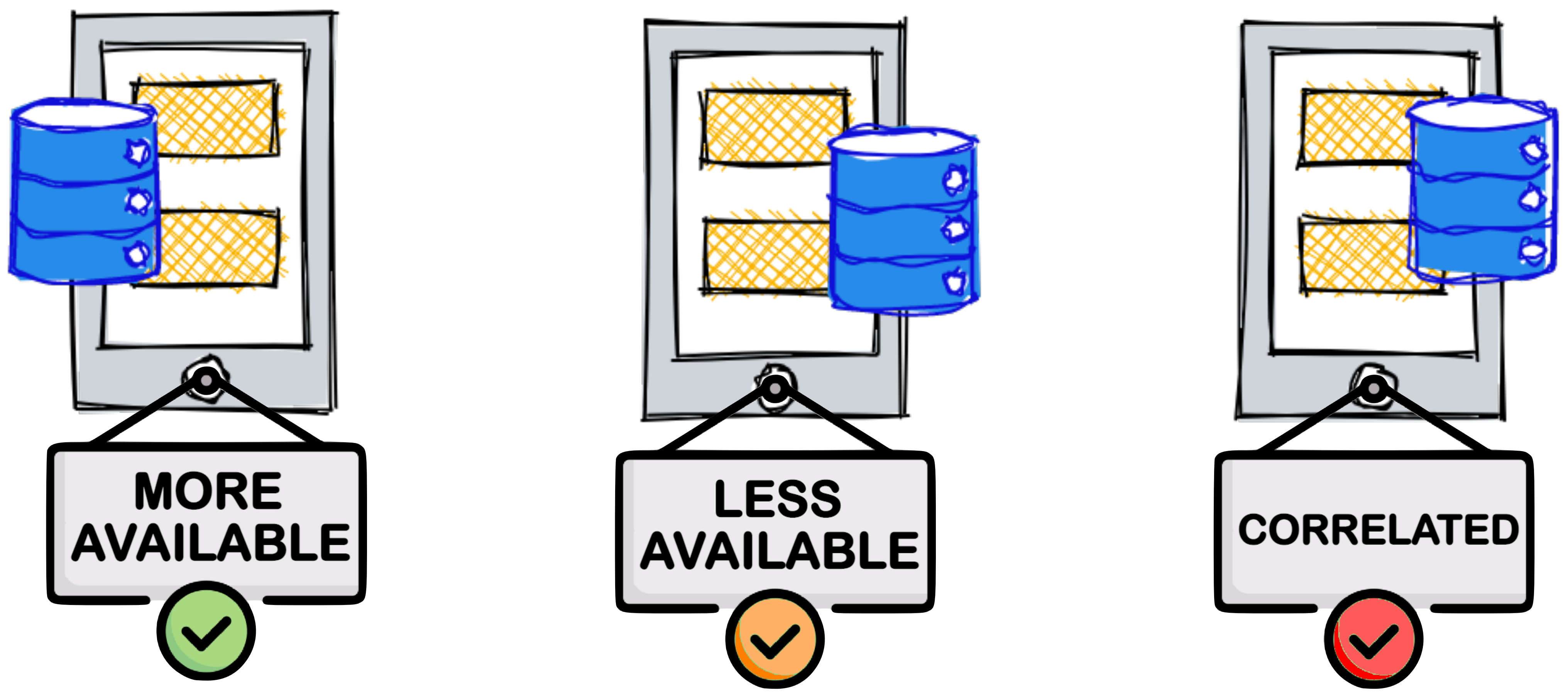


Correlation-Aware FL (CA-Fed)

(1) The server

- Estimate $\hat{\pi}^{(t)}$, $\hat{\lambda}^{(t)}$

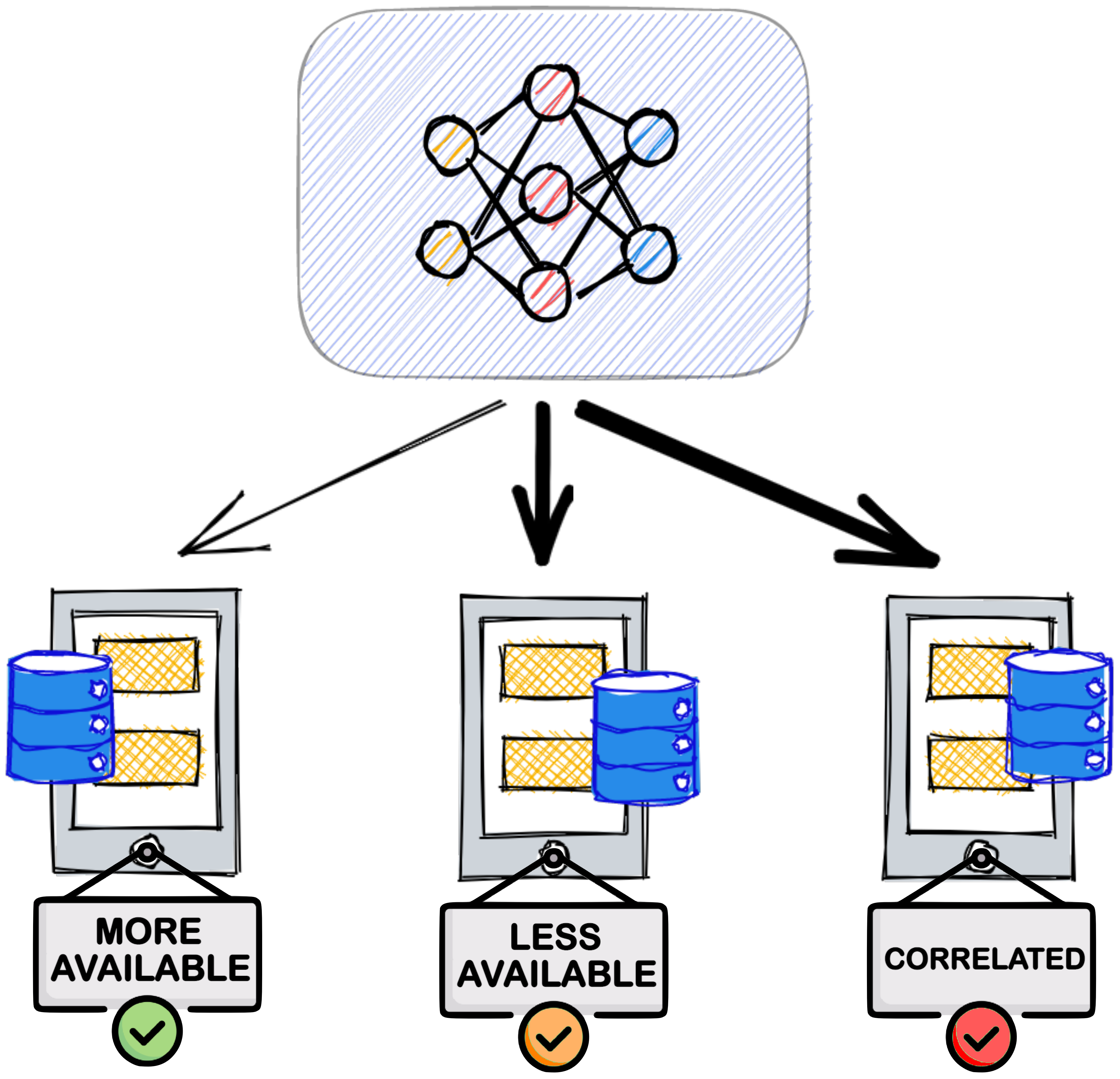




Correlation-Aware FL (CA-Fed)

(1) The server

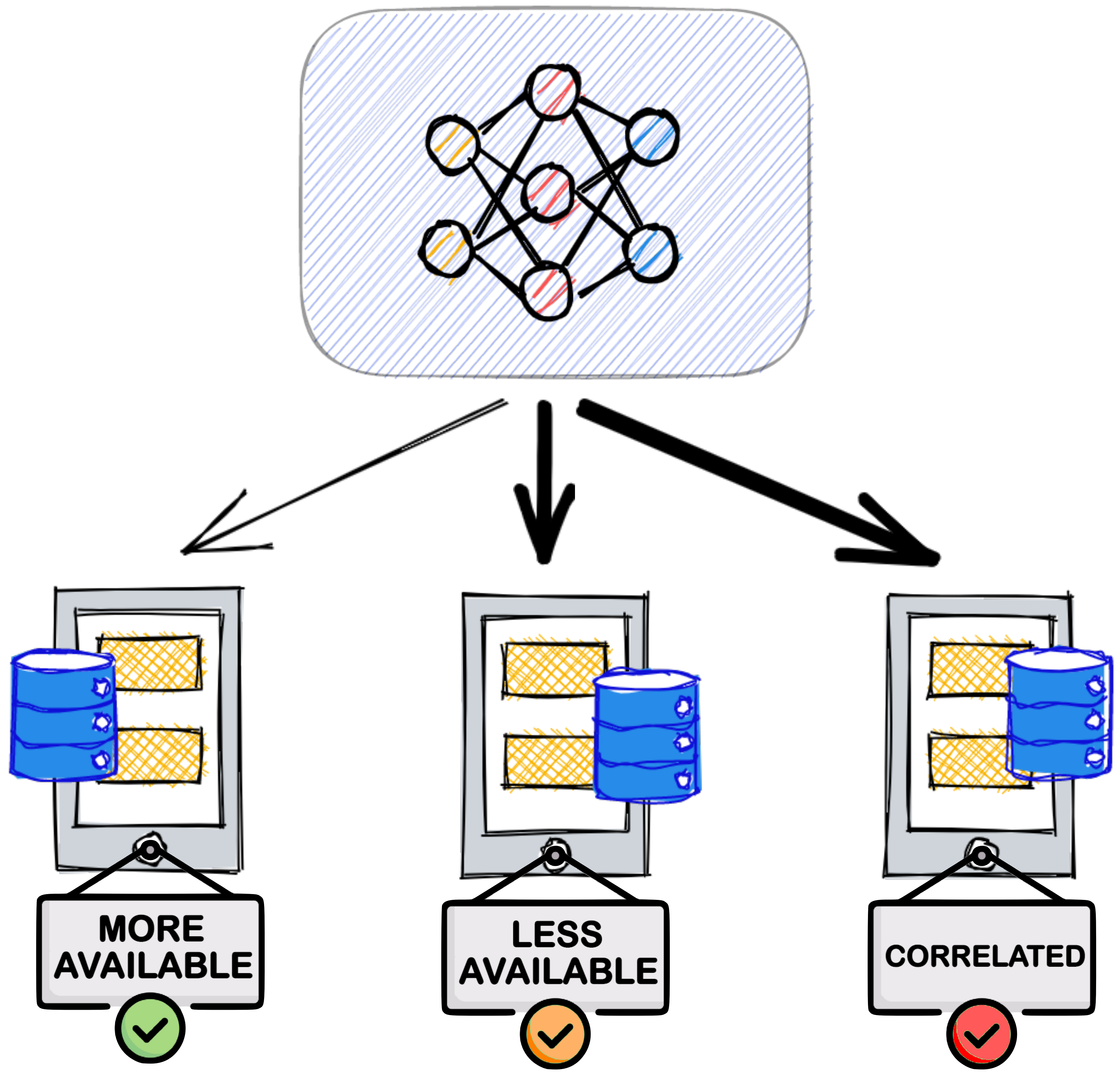
- Estimate $\hat{\boldsymbol{\pi}}^{(t)}, \hat{\boldsymbol{\lambda}}^{(t)}$
- Initialize $\boldsymbol{q}^{(t)} = \frac{\boldsymbol{\alpha}}{\hat{\boldsymbol{\pi}}^{(t)}}$



Correlation-Aware FL (CA-Fed)

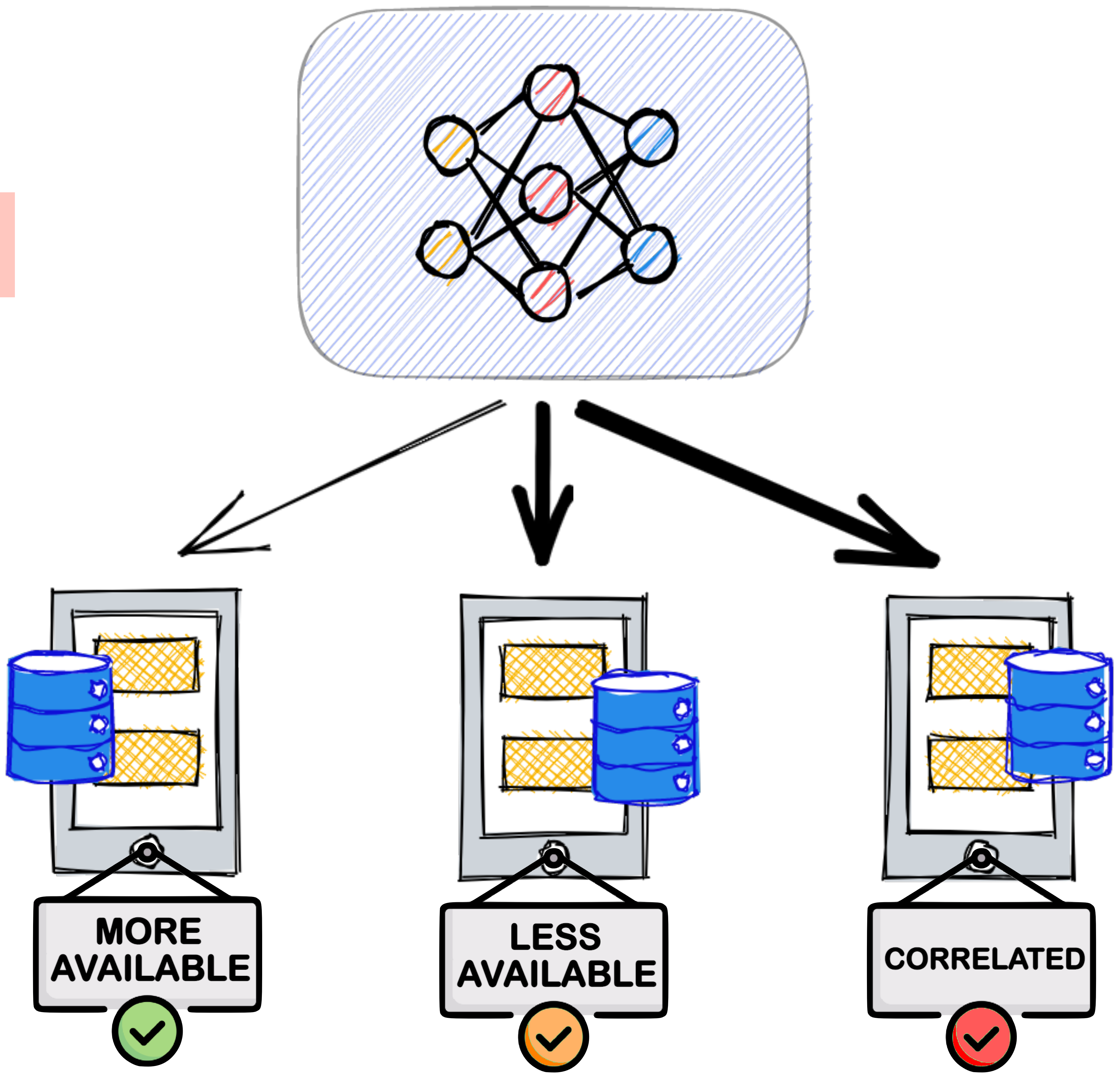
(1) The server

- Estimate $\hat{\pi}^{(t)}, \hat{\boldsymbol{\lambda}}^{(t)}$
- Initialize $\boldsymbol{q}^{(t)} = \frac{\boldsymbol{\alpha}}{\hat{\pi}^{(t)}}$
- Compute $\hat{\epsilon}^{(t)}(\boldsymbol{q})$



Correlation-Aware FL (CA-Fed)

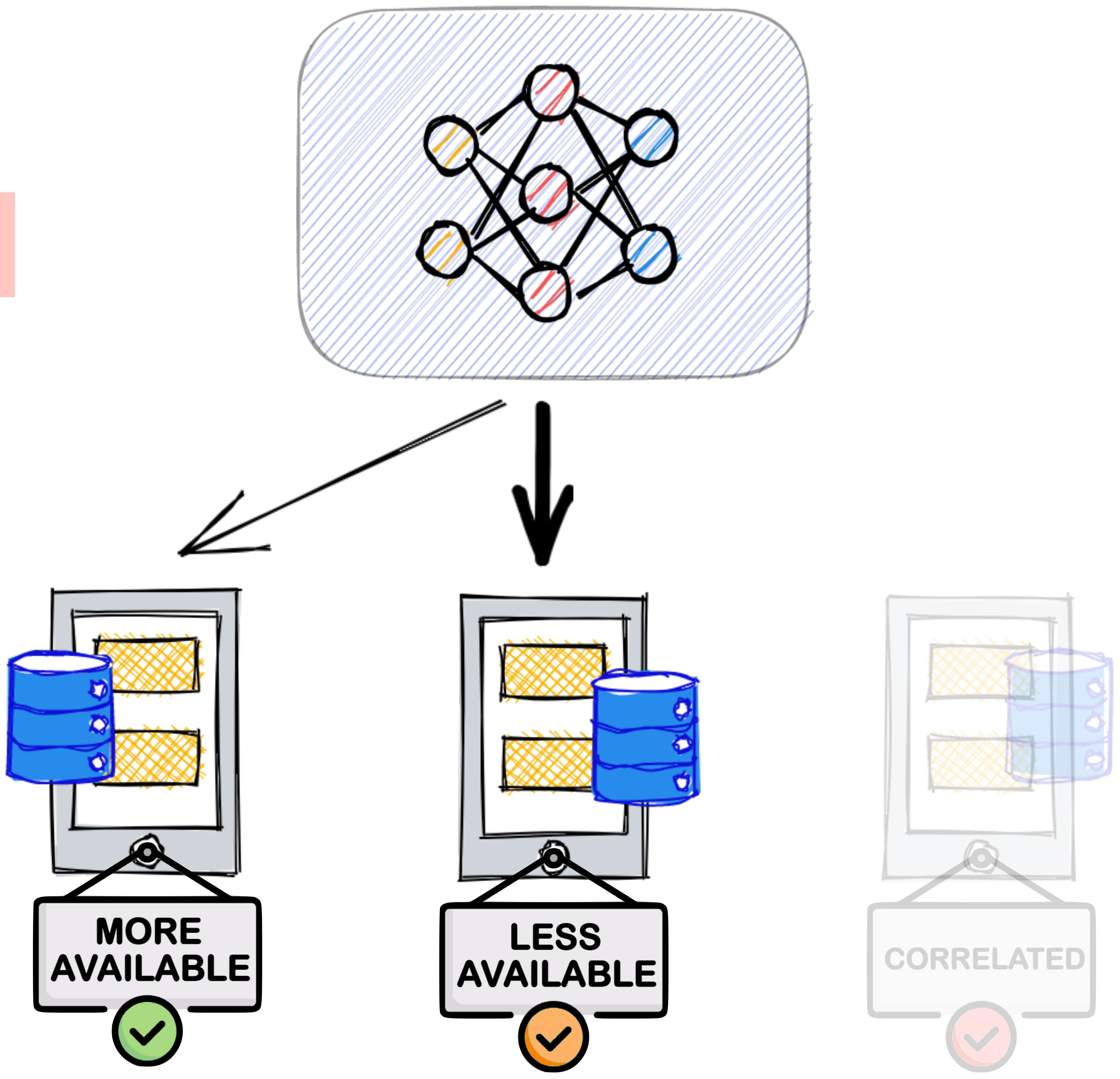
(2) For k in [K], starting from the "Less Available, Correlated" clients:



Correlation-Aware FL (CA-Fed)

(2) For k in [K], starting from the "Less Available, Correlated" clients:

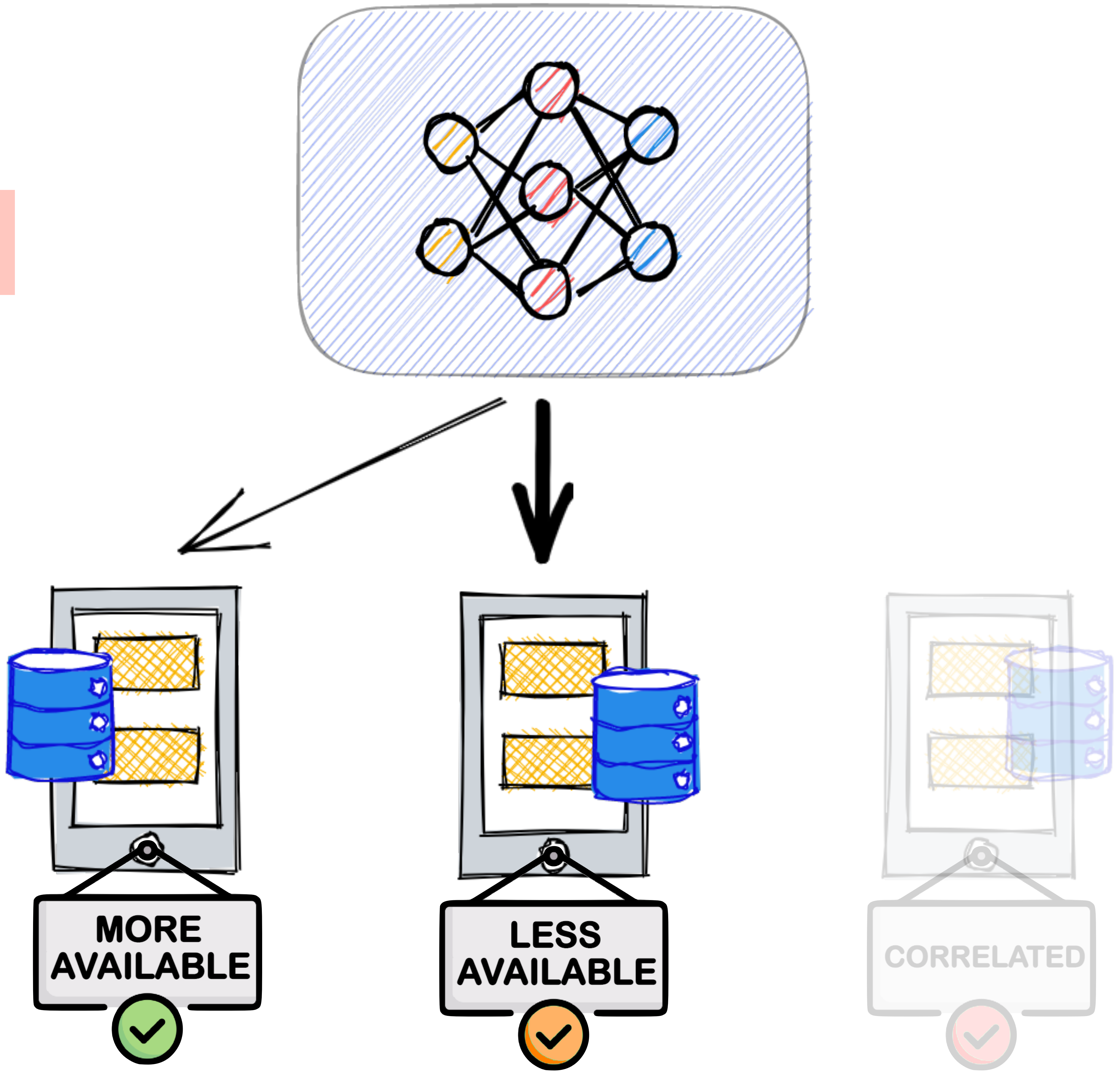
- Try $q_k = 0$



Correlation-Aware FL (CA-Fed)

(2) For k in [K], starting from the "Less Available, Correlated" clients:

- Try $q_k = 0$
- Compute $\hat{\epsilon}^{\text{new}}$

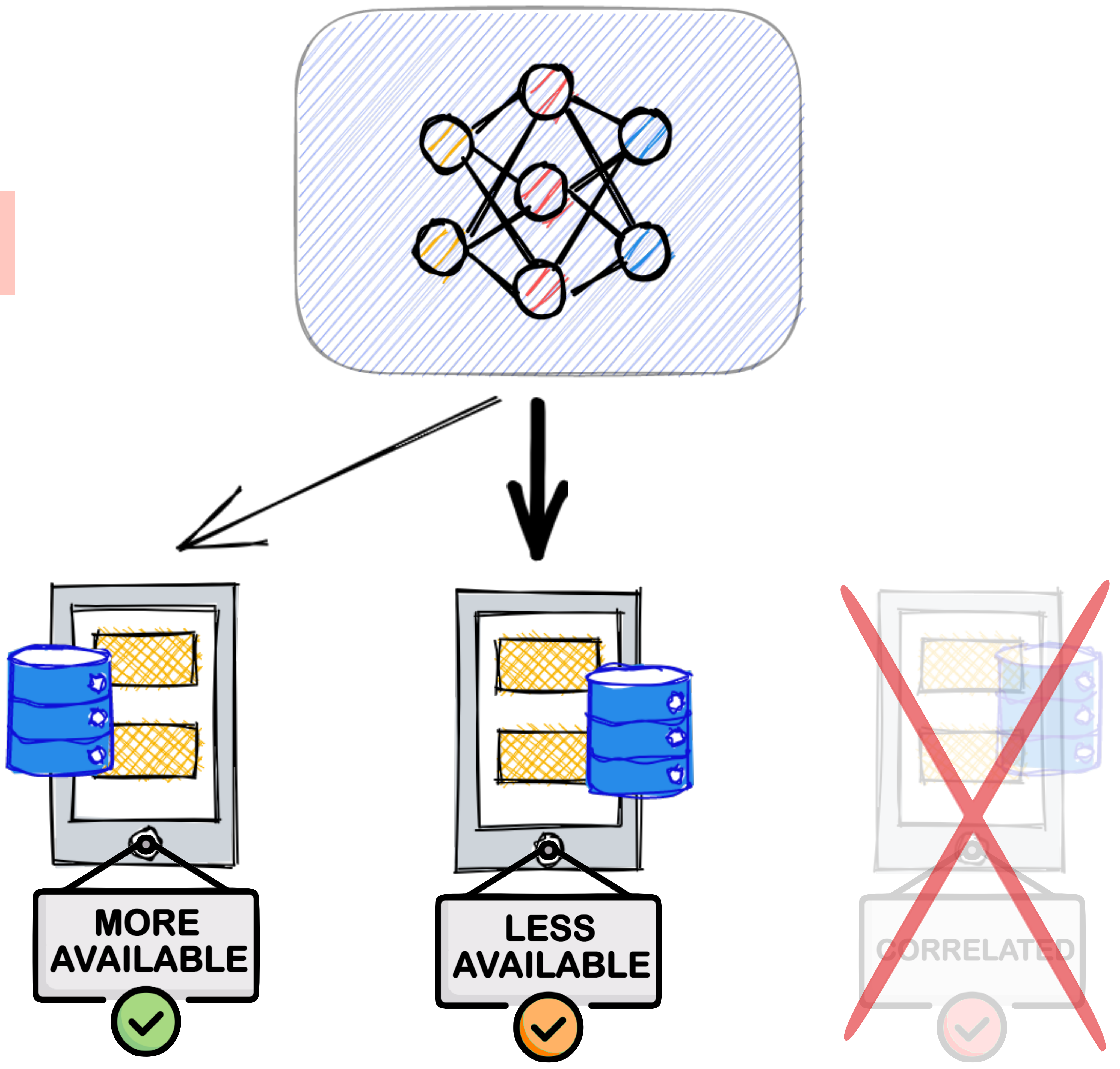


Correlation-Aware FL (CA-Fed)

(2) For k in [K], starting from the "Less Available, Correlated" clients:

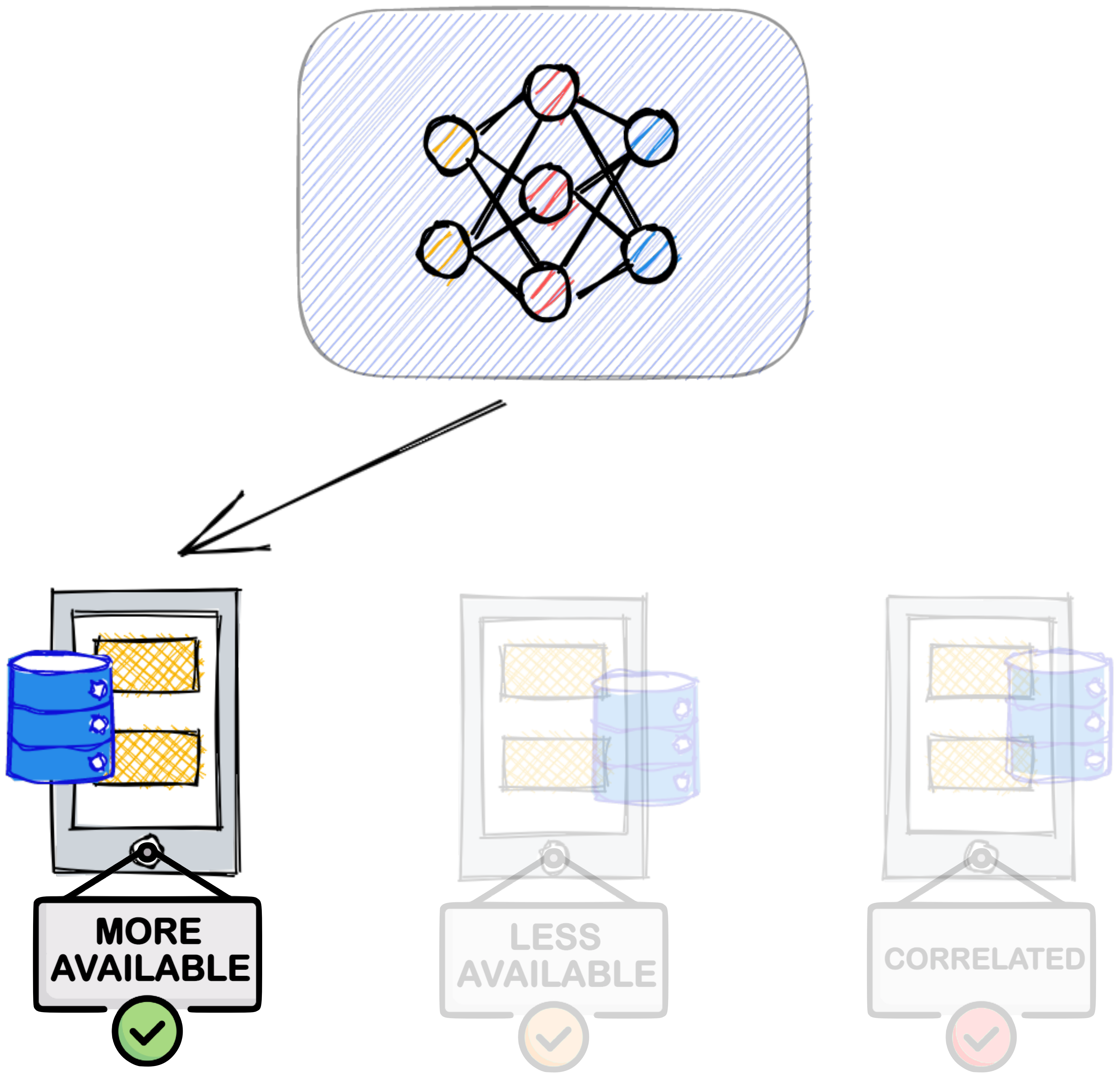
- Try $q_k = 0$
- Compute $\hat{\epsilon}^{\text{ new}}$
- If $\hat{\epsilon}^{\text{ old}} - \hat{\epsilon}^{\text{ new}} \geq 0$:

  Exclude client k



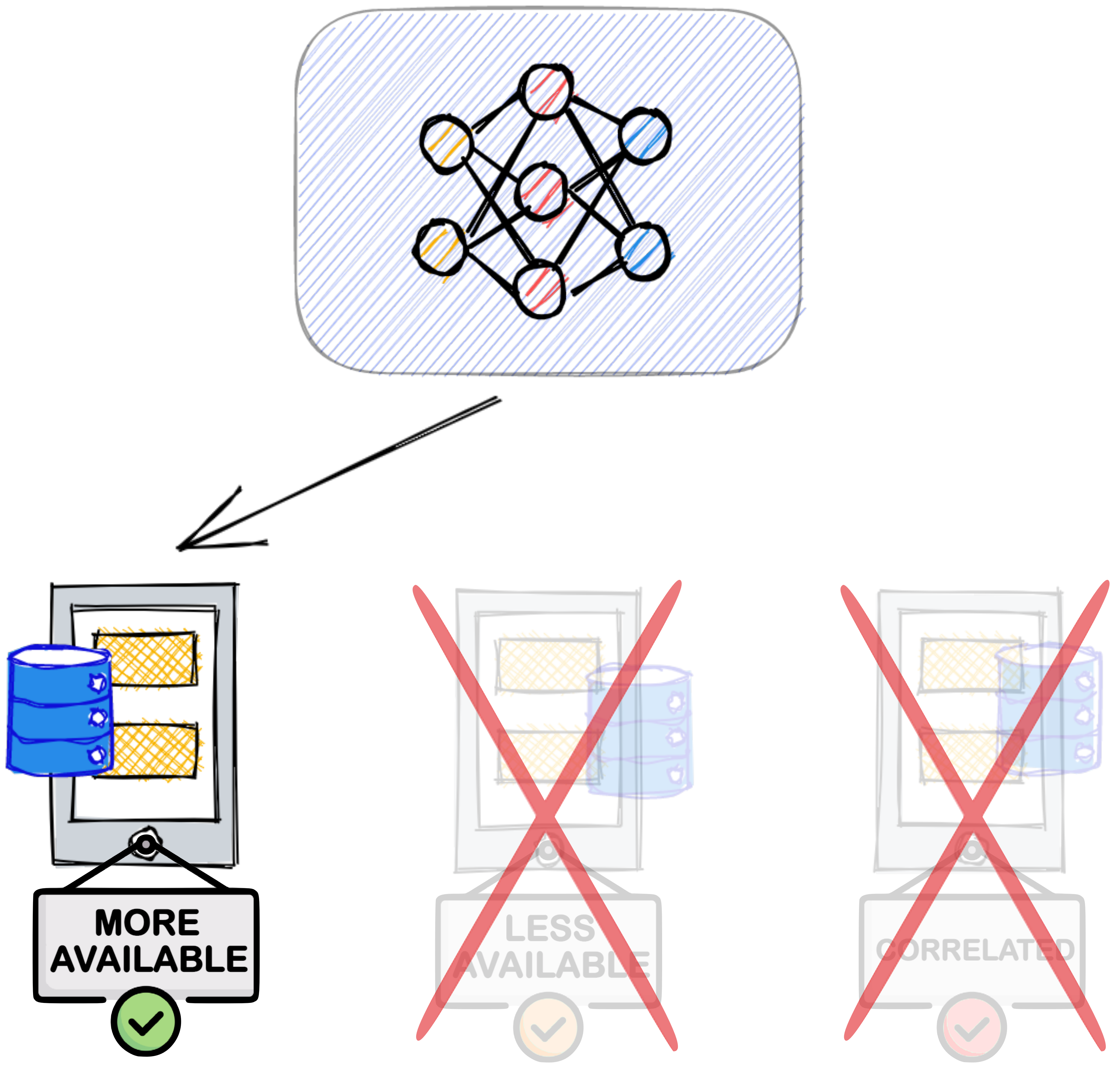
Correlation-Aware FL (CA-Fed)

(3) Only the clients $\{k \in A_t; q_k^{(t)} > 0\}$ train



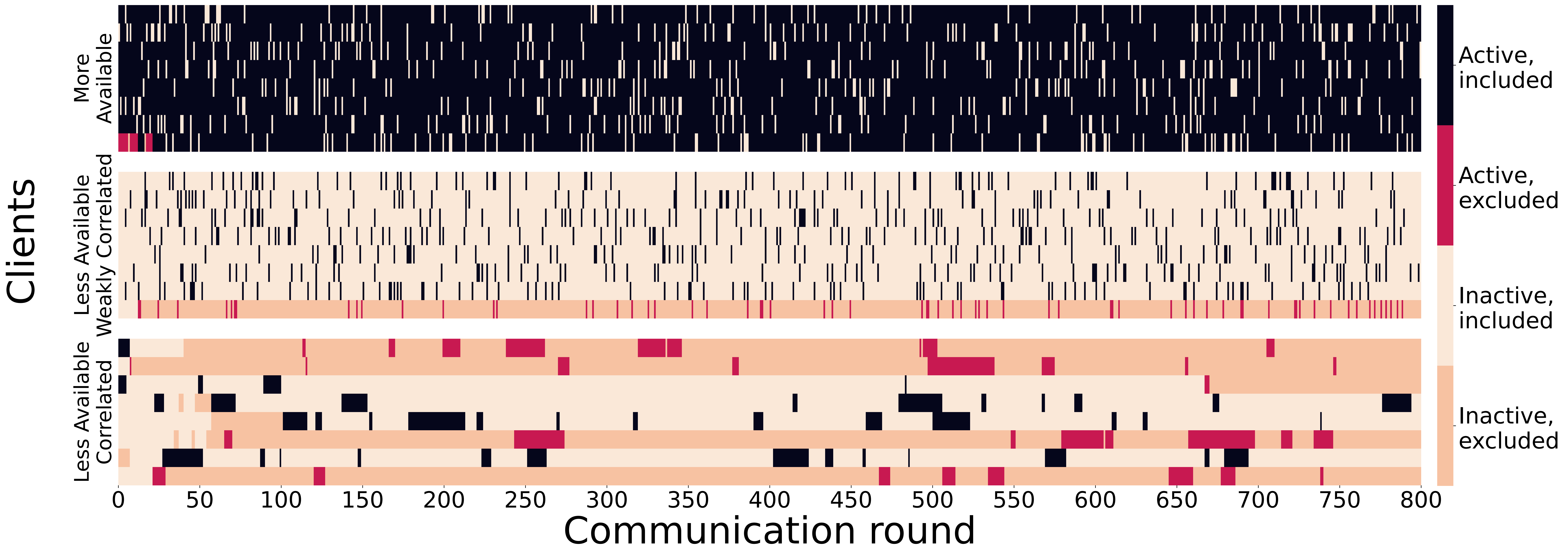
Correlation-Aware FL (CA-Fed)

(3) Only the clients $\{k \in A_t; q_k^{(t)} > 0\}$ train



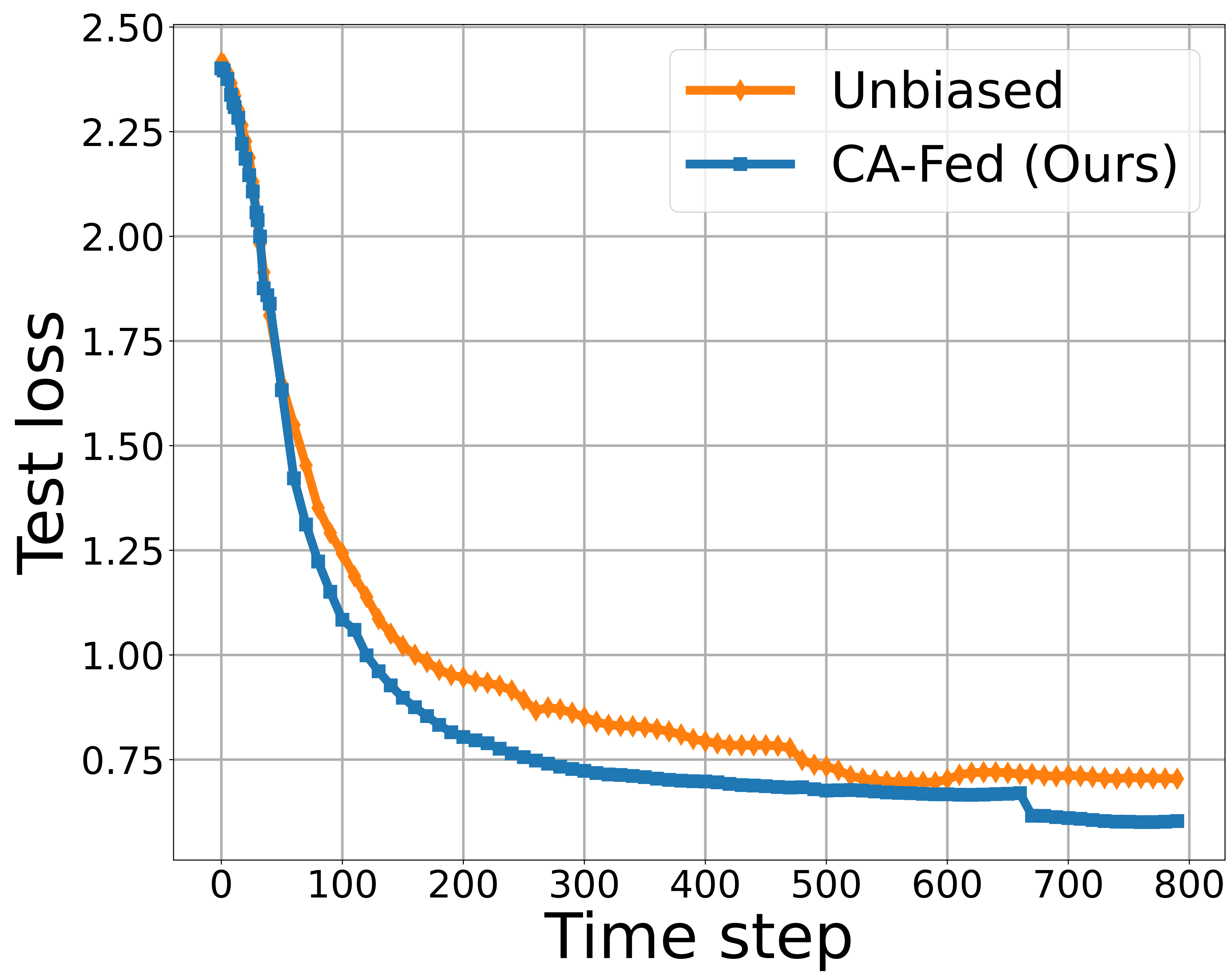
Experimental setting

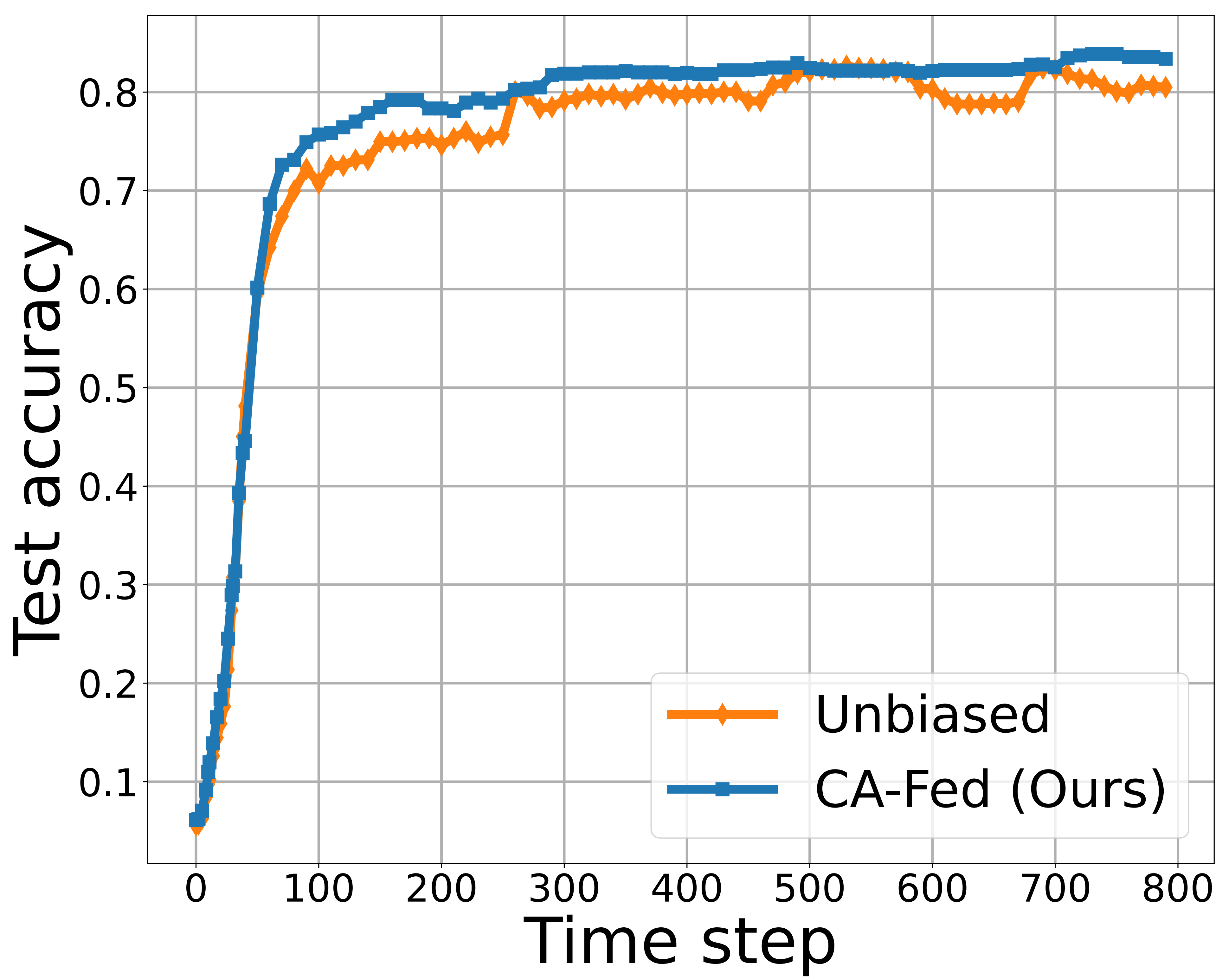
- Population of K=24 clients, divided in:



Experimental results

- We compare CA-Fed with the Unbiased baseline





CA-Fed excludes clients from training without performance drop

Conclusions

- Introducing a correlation process in the modeling of FL population
- First convergence analysis under intermittent and correlated client availability
- Adaptively excluding less available and correlated clients can be effective
- Excluding clients also reduces the overall training cost

Conclusions

- Introducing a correlation process in the modeling of FL population
- First convergence analysis under intermittent and correlated client availability
- Adaptively excluding less available and correlated clients can be effective
- Excluding clients also reduces the overall training cost

Thank you for your attention!