tecnalia
MEMBER OF BASQUE RESEARCH & TECHNOLOGY ALLIANCE

DISRUPTIV

Plataforma Tecnológica Española de Tecnologías Disruptivas

Con financiación de:

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN

Convocatoria 2020 Plataformas Tecnológicas y de Innovación (PTR2020-001159)

Secretaría técnica a cargo de:

APTE

# Federated Machine Learning

- Maria Aróstegi

Haga clic aquí para unirse a la reunión

**14/12/2022**

# Perfil

Maria **Arostegi**, acabó su licenciatura en **Matemáticas** (especialidad en Teoría de Números) en 1995 en la EHU-UPV (Universidad del País Vasco), después de cursar su último año en Milán (Italia), en L'Universtità degli Studi di Milano.

En 1999 tuvo una beca de la diputación para trabajar 1 año en Iowa State University con **Carolina Cruz–Neira**, en temas de **VR**. Después de su Vuelta trabajó en varios proyectos relacionados con herramientas de modelización 3D, interacción gráfica, computación y simulación en escenarios virtuales.

Especializándose finalmente en el desarrollo matemático de las ecuaciones 3D de trasferencia de calor en procesos **siderúrgicos**, y los factores de forma de las mismas.

En 2017, cambió su campo de investigación para centrarse en algoritmos matemáticos de **analítica de datos**, Machine Learning, y tecnologías Big data (como Hadoop, Spark, etc..), cursando para ello un **master** online en "**Big data, Business Intelligence, y Data Science**". Actualmente se encuentra desarrollando un doctorado en: **Incremental Explainable ML.**

## Ultimas publicaciones:

- "**A heuristic approach to the multicriteria design of IaaS cloud infrastructures for Big Data applications**". *Maria Arostegi, Ana-Isabel Torre-Bastida, Nekane Bilbao, Javier Del Ser.* January 2018. Expert Systems DOI 10.1111/exsy.12259

- "**Concept Tracking and Adaptation for Drifting Data Streams under Extreme Verification Latency**". *Maria Arostegi Ana-Isabel Torre-Bastida Jesús López Lobo, Nekane Bilbao, Javier Del Ser.* September 2018 Studies in Computational Intelligence DOI: 10.1007/978-3-319-99626-4_2 In book: Intelligent Distributed Computing XII

- " **SLAYER: A Semi-supervised Learning Approach for Drifting Data Streams under Extreme Verification Latency**", *Maria Arostegi, Jesús Lopez Lobo and Javier del Ser*. Proceedings of the "Workshop Interactive Adaptive Learning Bilbao September 2021. IAL2021. https://www.activeml.net/ial2021/pdf/ialatecml2021.pdf

- " **Machine Learning based Soft Sensing Tool for the Prediction of Leaf Wetness Duration in Precision Agriculture**", *Maria Arostegi, Diana Manjarrés, Sonia Bilbao and Javier del Ser*. Proceedings of the "16th International Conference on Soft Computing Models in Industrial and Environmental Applications", Bilbao September 2021. SOCO21. Publication Springer Nature https://doi.org/10.1007/978-3-030-87869-6_50

- **"Edge intelligence secure frameworks: Current state and future challenges"** *Esther Villar, Maria Arostegi, Ana-Isabel Torre-Bastida, Cristina Regueiro,Juan Lopez de Armentia* August 2022. UNDER REVIEW Computers & Security

# Índice

Con financiación de:
Convocatoria 2020 Plataformas Tecnológicas y de Innovación (PTR2020-001159)

Secretaría técnica a cargo de:
APTE

# *Why do we need federated learning?*

# Data is born at the edge

Billions of phones & IoT devices constantly generate data
Data enables better products and smarter models

# Can data live at the edge?

Data processing is moving on device:

- Improved latency
- Works offline
- Better battery life
- Privacy advantages

E.g., on-device inference for mobile keyboards and cameras.


# Can data live at the edge?

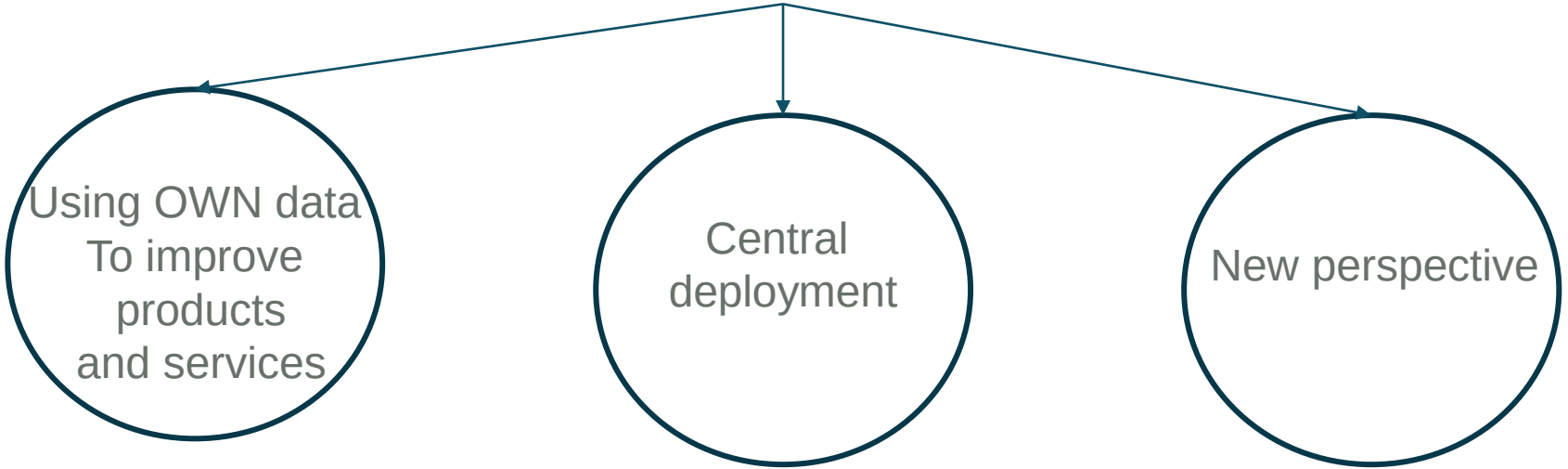

Data processing is moving on device:

- Improved latency
- Works offline
- Better battery life
- Privacy advantages

E.g., on-device inference for mobile keyboards and cameras.

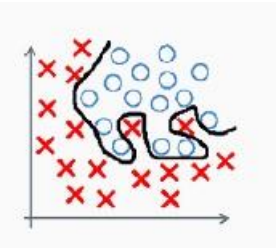
But when **computing** any **learning** process?

# **Three** options



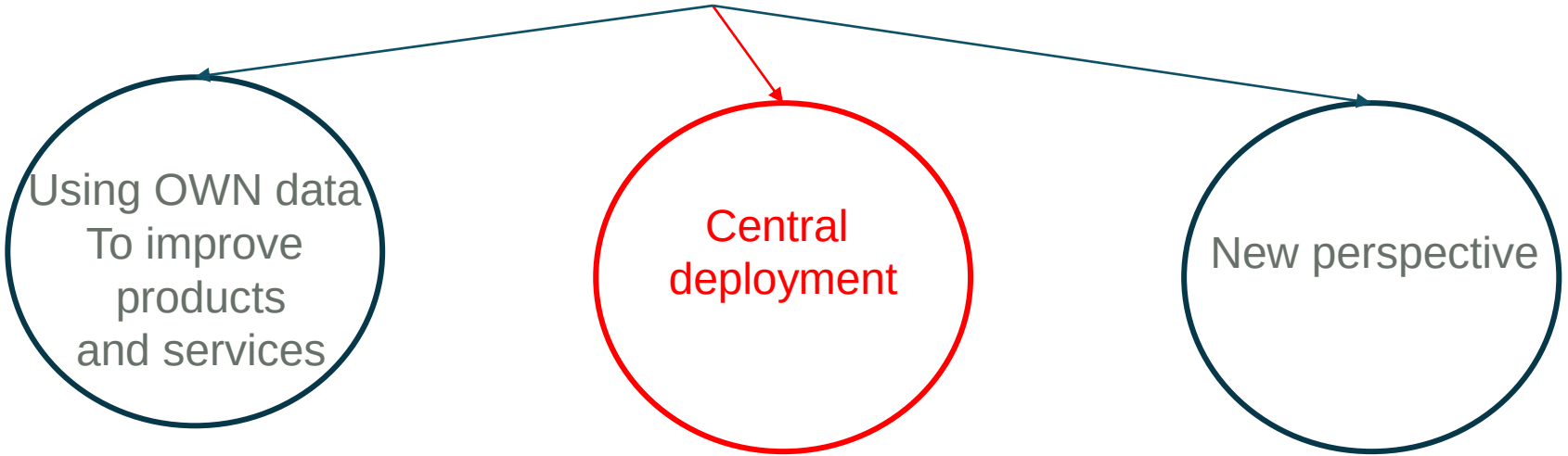
Using OWN data

HOW ABOUT **EACH PARTY LEARNING** ON ITS OWN?



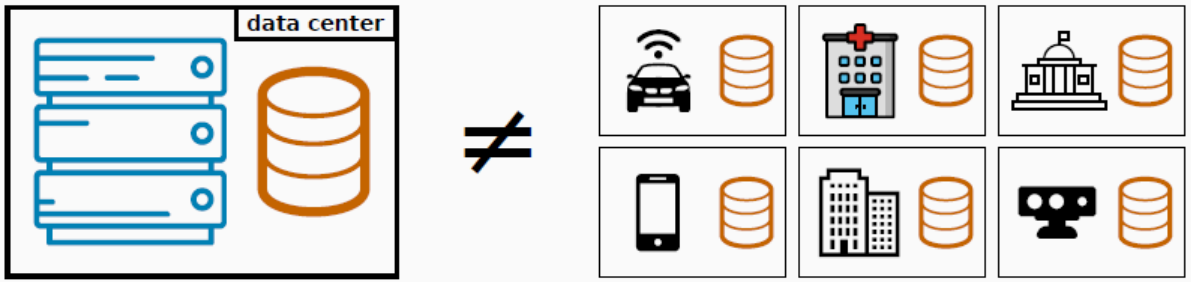
1. The local dataset may be too small

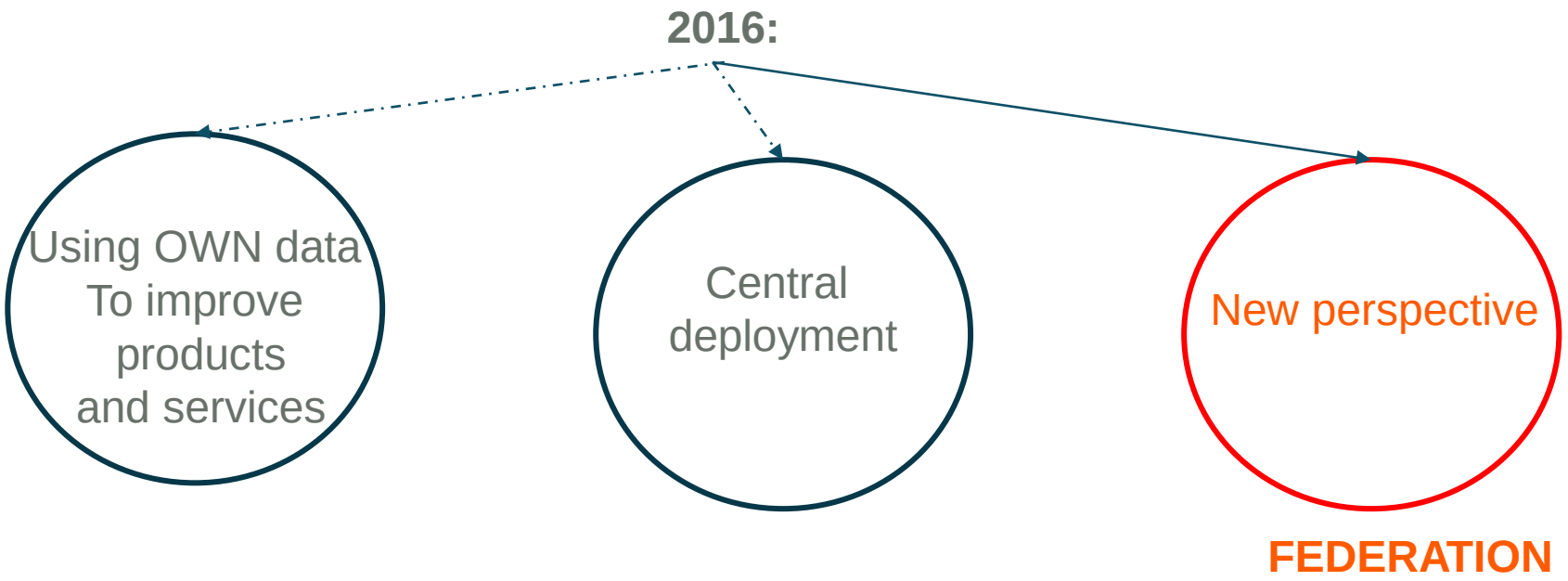
2. The local dataset may be biased

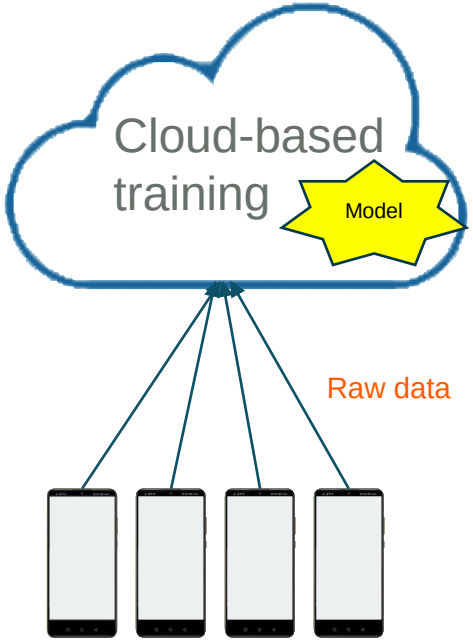

Using OWN data
To improve
products
and services
Central
deployment
New perspective

# FROM **CENTRALIZED** TO DECENTRALIZED DATA

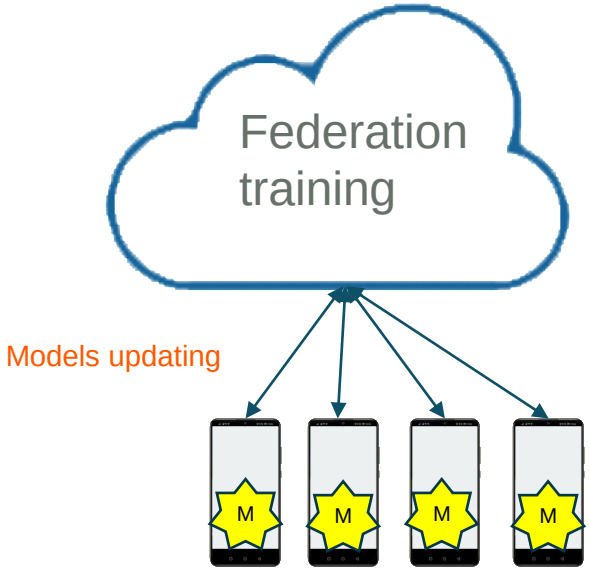
- The standard setting in Machine Learning (ML) considers a centralized dataset processed in a tightly integrated system
- But in the real world data is often decentralized across many parties




2016:
Using OWN data To improve products and services
Central deployment
New perspective
FEDERATION


Cloud-based training
Model
Raw data

# WHY?


Federation training
Models updating
M
M
M
M

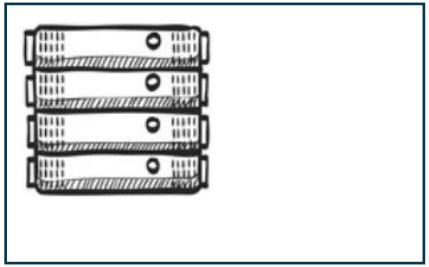
DISRUPTIVE Plataforma Tecnológica Española de Tecnologías Disruptivas
Con financiación de:
Convocatoria 2020 Plataformas Tecnológicas y de Innovación (PTR2020-001159)
Secretaría técnica a cargo de: APTE

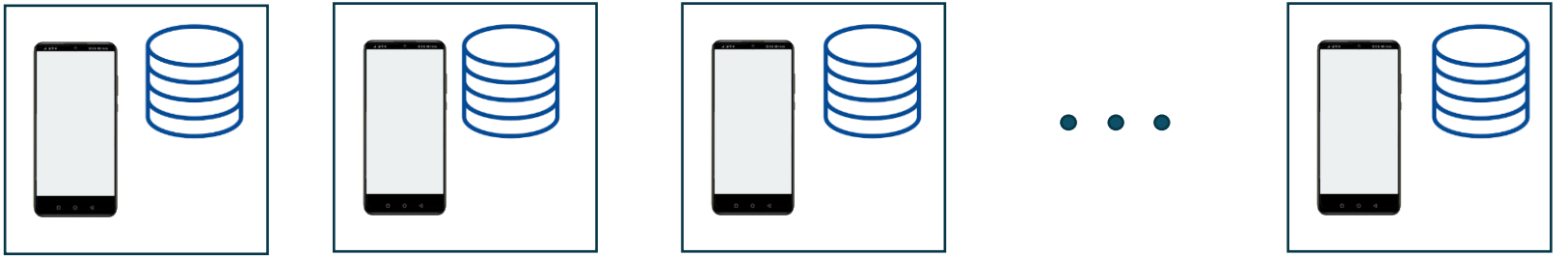
# 02

# *What is federated learning?*

# 1- Federated Training

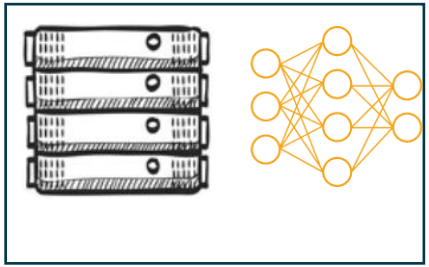
Federated Learning (FL) aims to collaboratively **train** a ML model while keeping the data decentralized

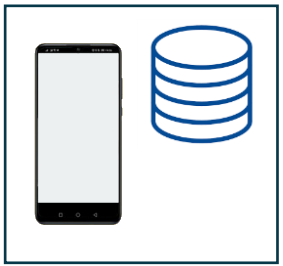


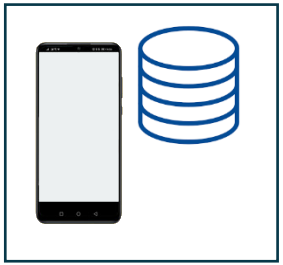


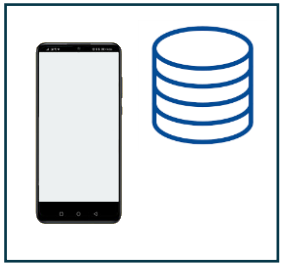
# 1- Federated Training

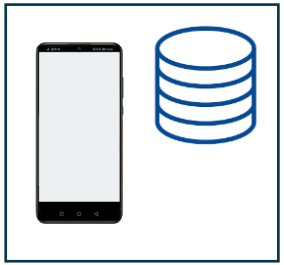
Initialize model →



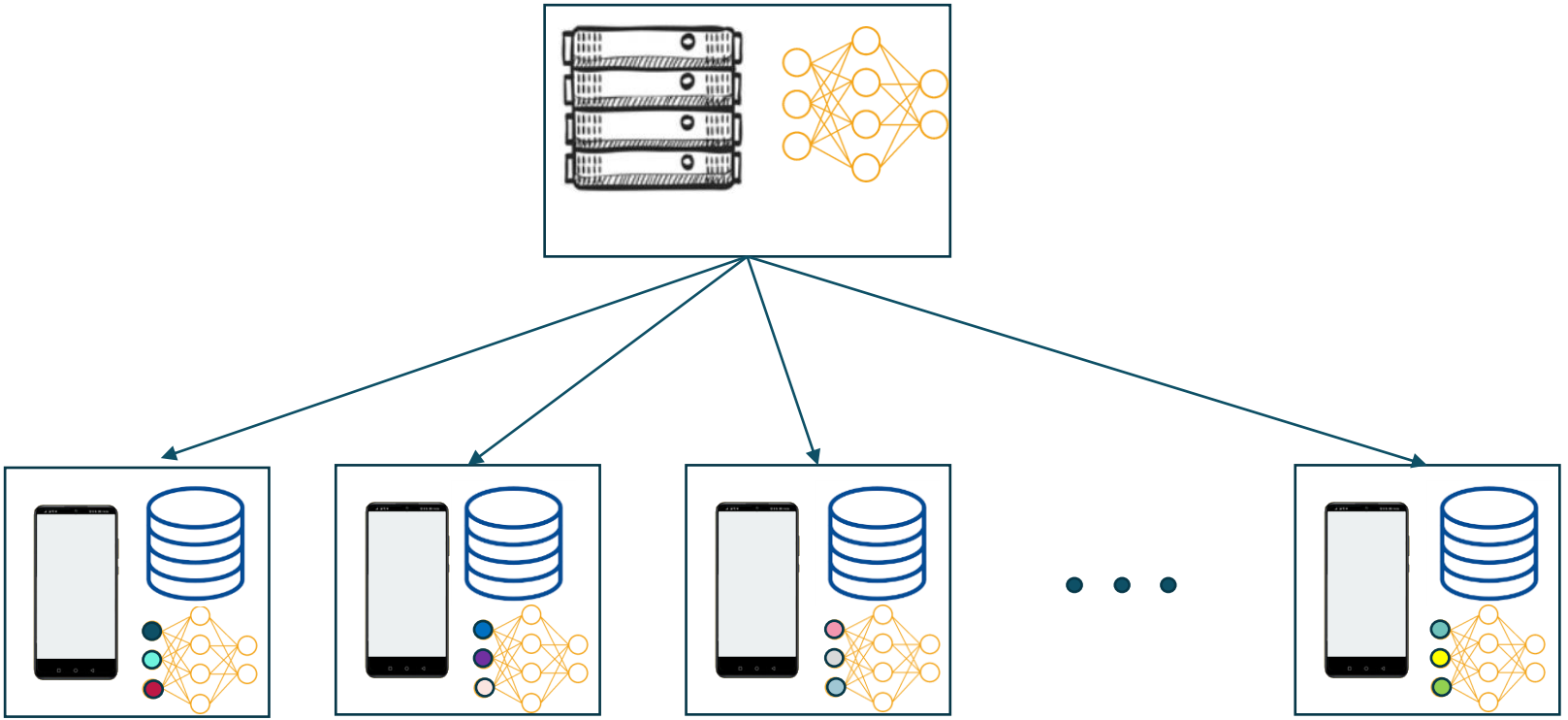








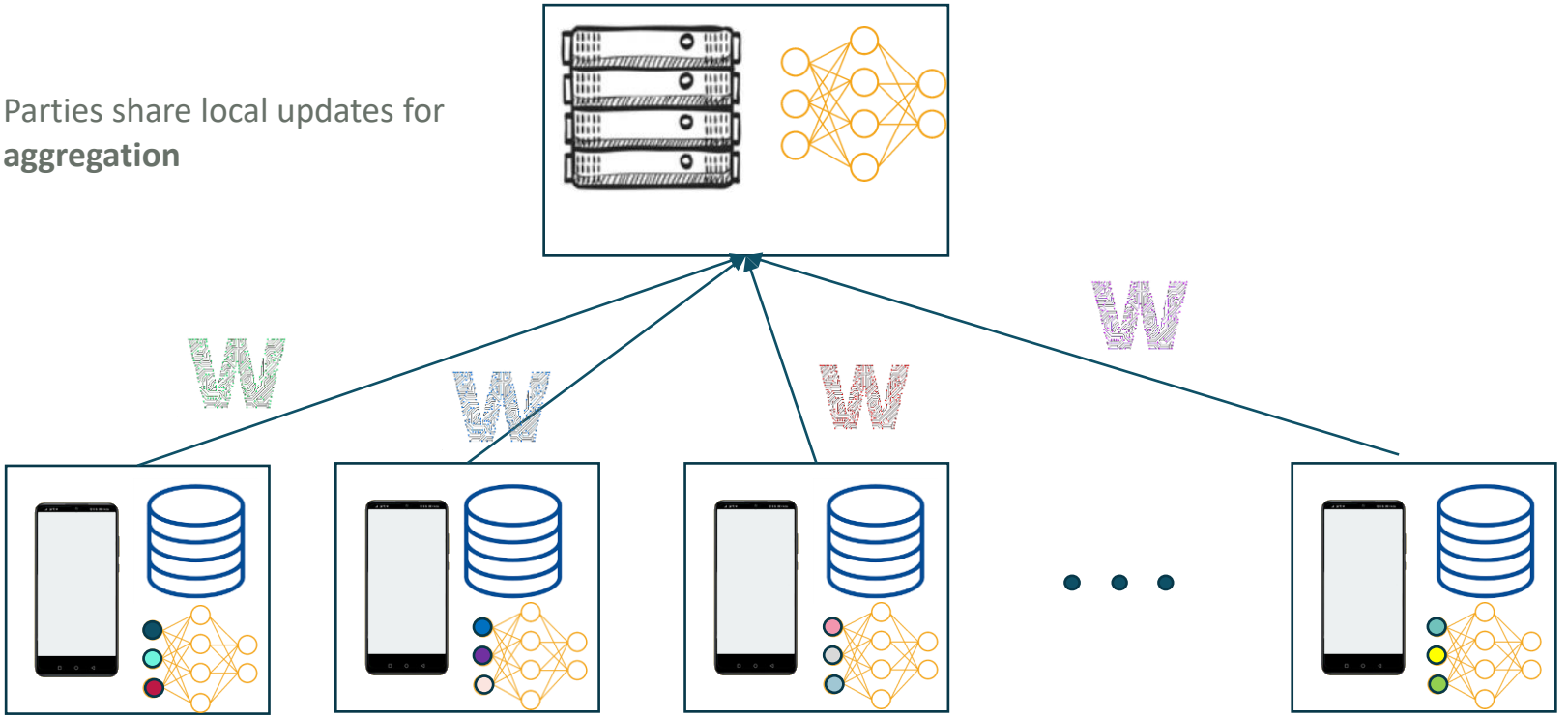
# 1- Federated Training



Each party makes an update using its **local** dataset: **FEDERATED TRAINING**

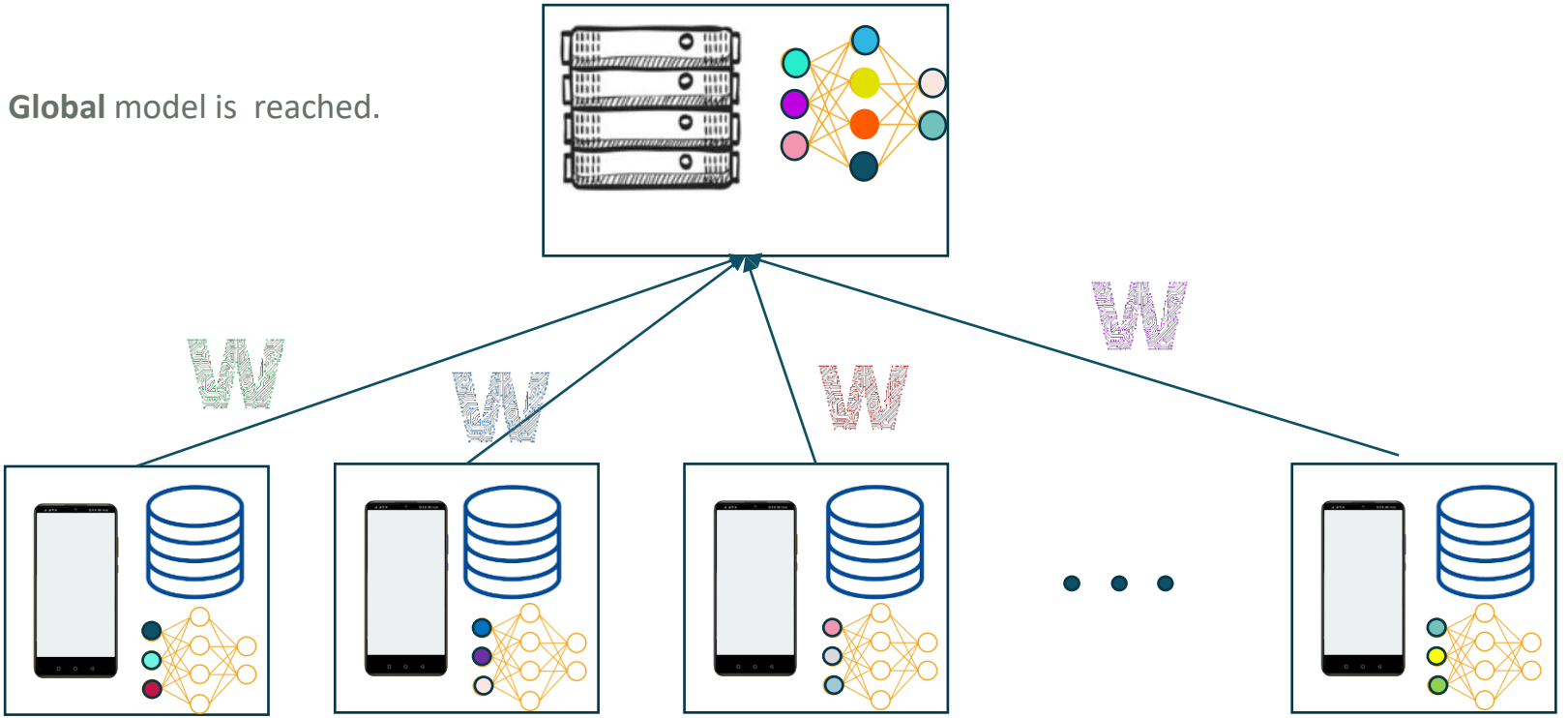
# 1- Federated Training

Parties share local updates for **aggregation**



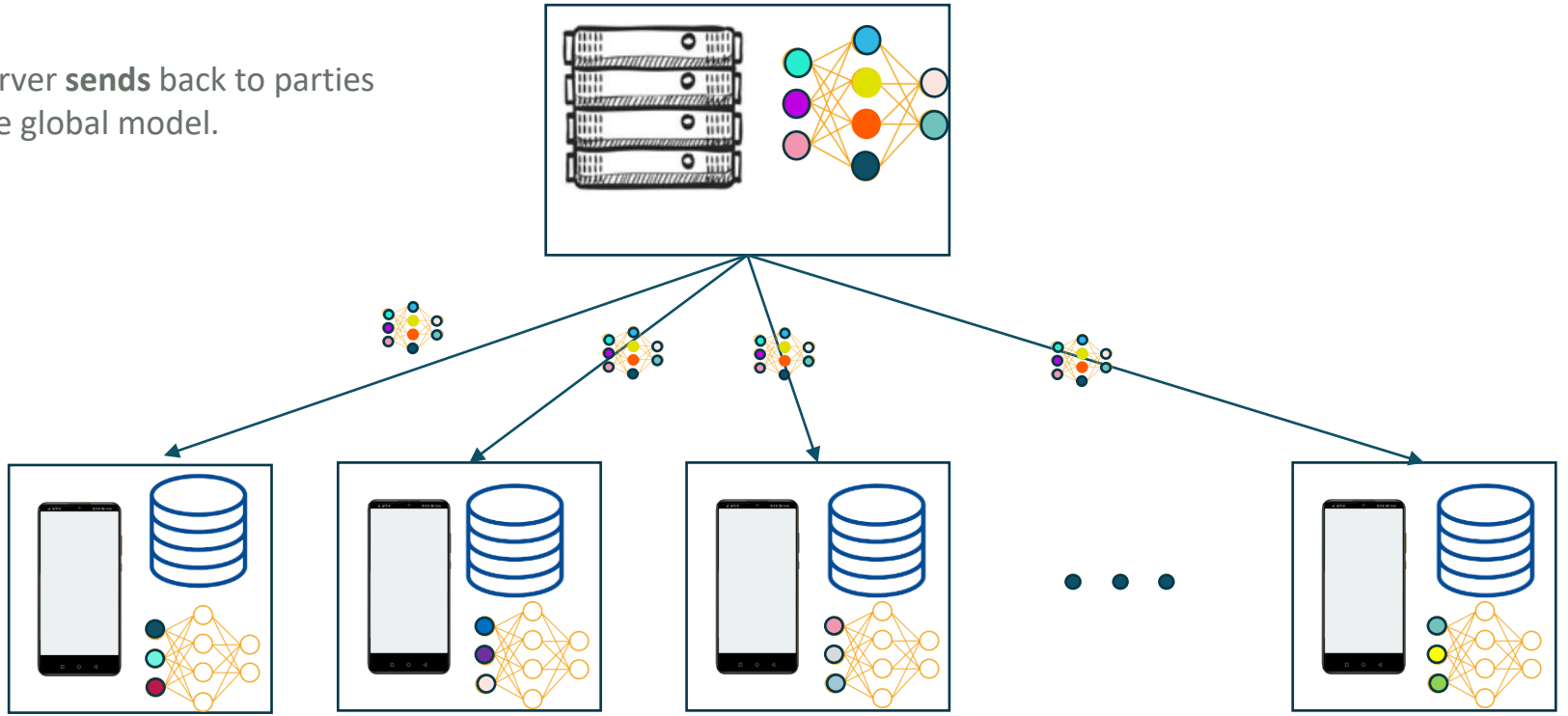
# 1- Federated Training

**Global** model is reached.



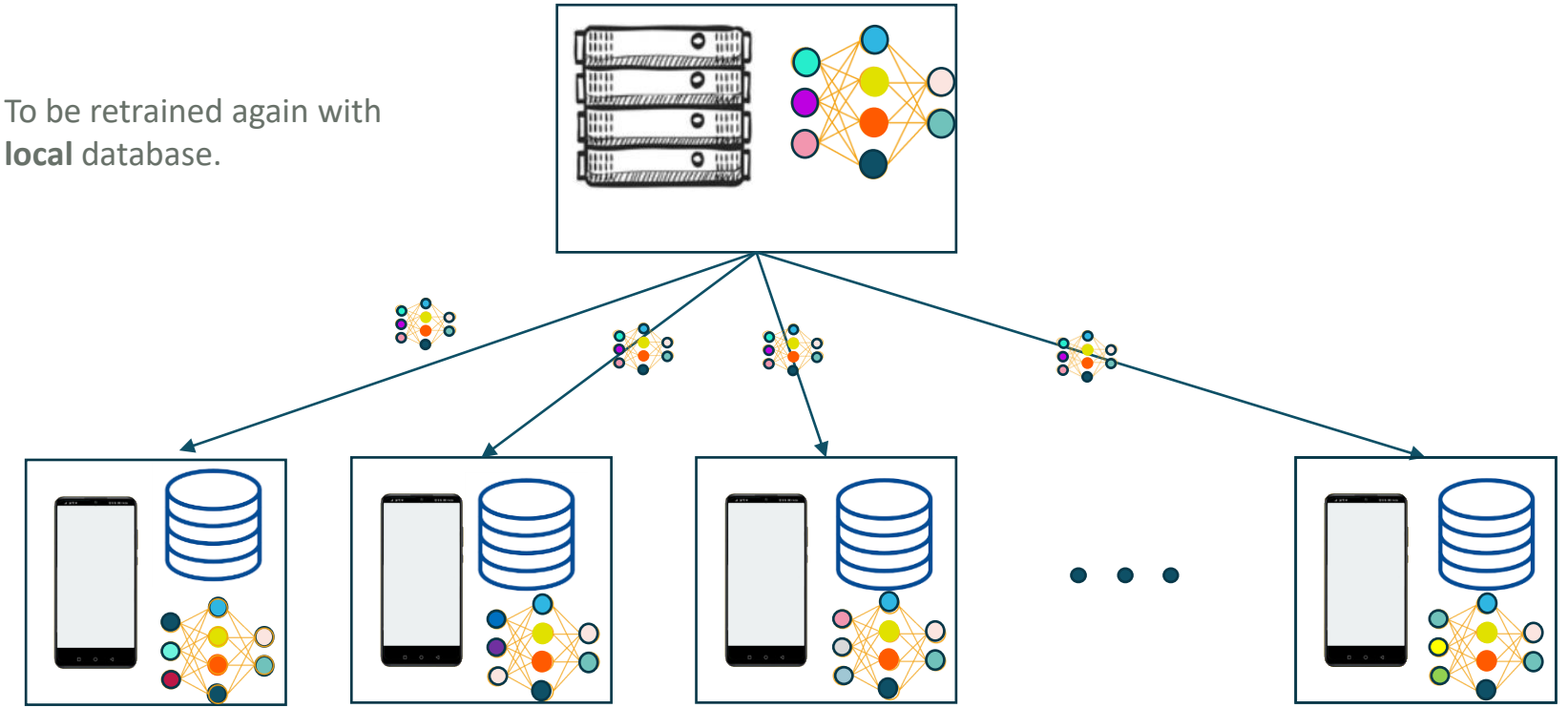
# 1- Federated Training

Server **sends** back to parties the global model.



# 1- Federated Training

To be retrained again with **local** database.



Con financiación de:
Convocatoria 2020 Plataformas Tecnológicas y de Innovación (PTR2020-001159)

Secretaría técnica a cargo de:
APTE

# 03

# *Use cases*

# Projects & Frameworks

## Terminet

Provide a set of innovative mechanisms and tools for moving AI to the edge by using cutting-edge ML technologies, avoiding data collection and offering decentralized analytics, privacy by design and data protection.

TERMINET

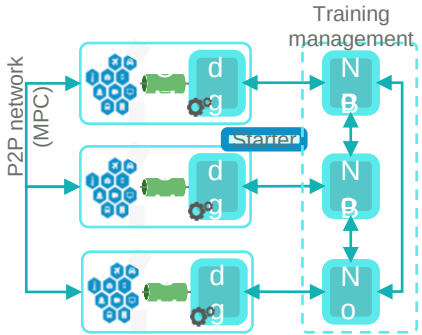
## Shinefleet

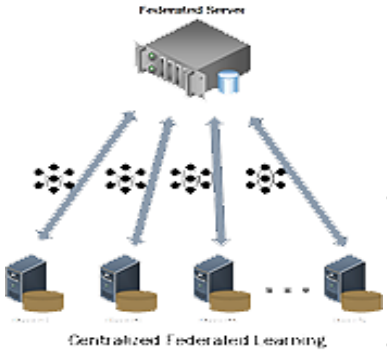
- Para el desarrollo de ***soluciones tecnológicas de hidrógeno*** para la movilidad inteligente y sostenible de flotas pesadas.
- Durante los próximos años, los participantes abordarán los principales ámbitos de investigación relacionados con ***la generación de hidrógeno de manera descentralizada***, para su uso en el transporte de mercancías en camión.

SHINE-fleet

## Decentralized

**Federated Training (Offline)**
➢FL Orchestration : Blockchain clients
➢FL Model Aggregation : MPC
➢FL Local Training :all the edges (clients)



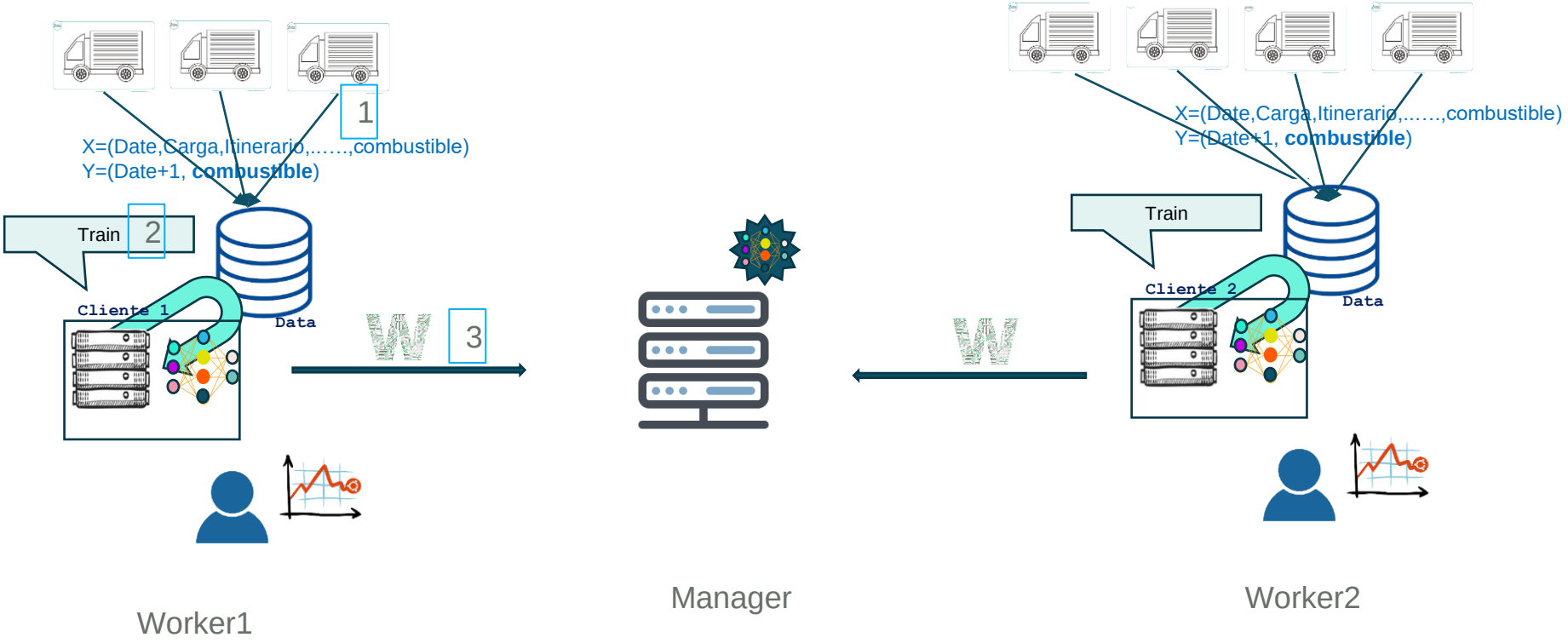




## Centralized

- Baseline for the analysis to explore codes to **speed up distributed machine learning training.**
- **Framework de instanciacion.**

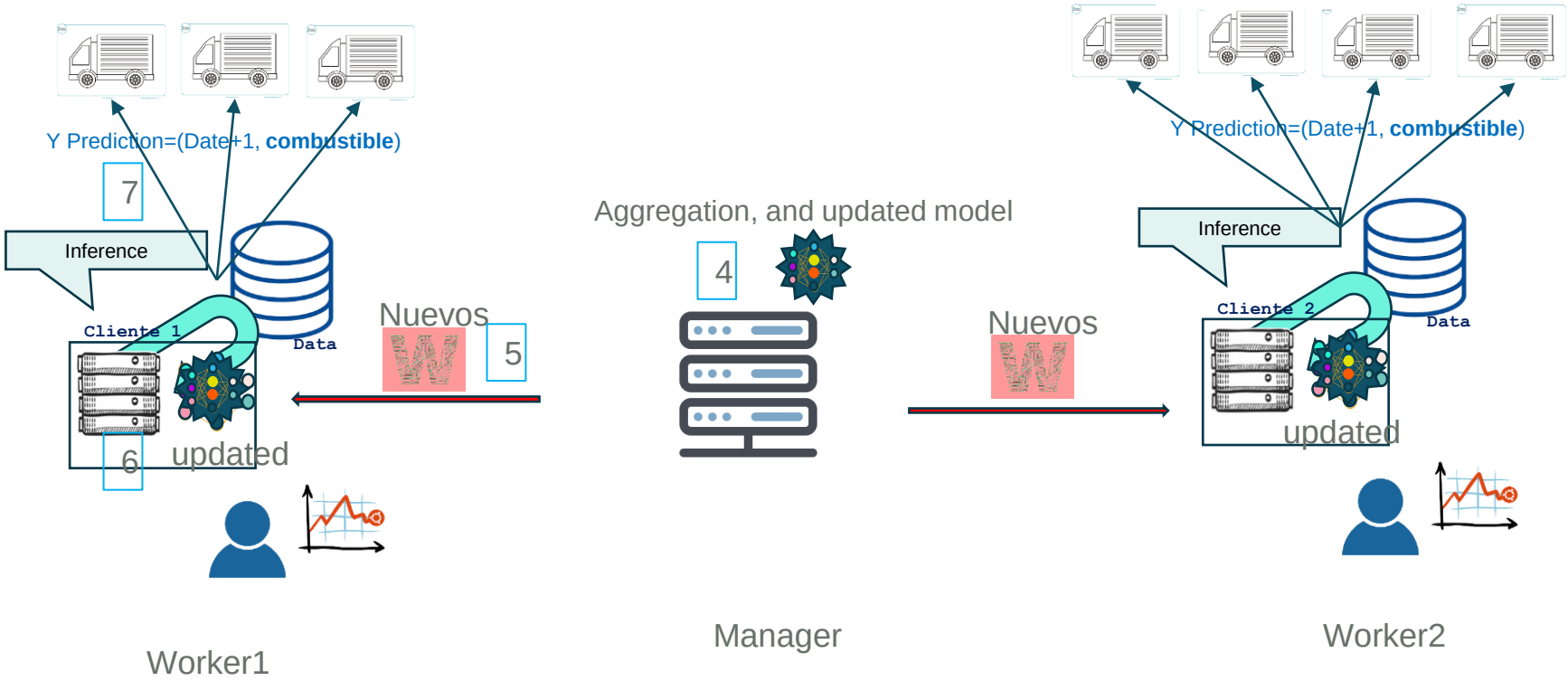
# CASOS de USO SHINE-fleet

## Flujo de FML



# CASOS de USO SHINE-fleet

## Flujo de FML



# CASOS de USO Terminet

"Pathway of Personalized Healthcare"

**Pathology**

Measurements

Reports

Patient Doctor

**Radiology**

**Radiation oncology**

**Surgery**

Measurements

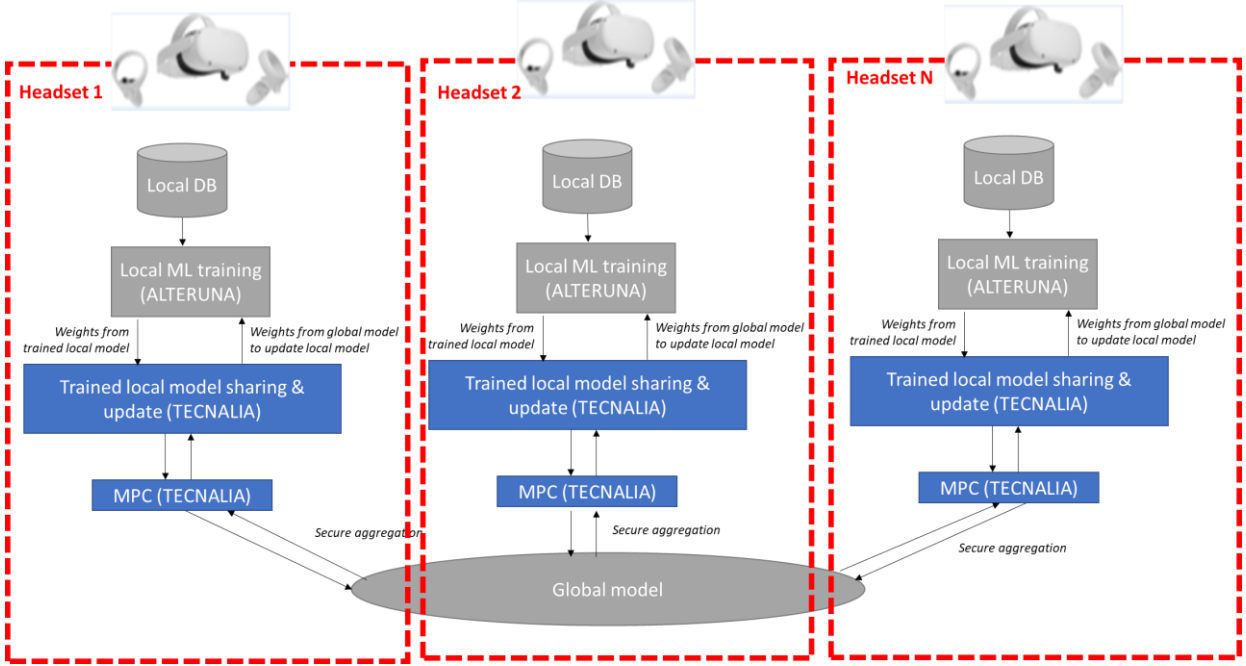
Reports

Using Healthentia apps at the edge,
Healthentia platform centrally


GemelliGenerator

# CASOS de USO Terminet

"Mixed Reality and ML Supported Maintenance and Fault Prediction of IoT-based Critical Infrastructure" addresses the area of maintenance, and lastly"



# *Questions??*

## *THANKS!!!!*