Design Considerations and Guidelines

for Implementing Federated Learning in Smart Manufacturing Applications

2022-03-17

Authors:

Sourabh Bharti
Munster Technological University, Cork, Ireland
sourabh.bharti@mtu.ie

Alan McGibney
Munster Technological University, Cork, Ireland
mailto:alan.mcgibney@mtu.ie

Tristan O'Gorman
IBM
Cork, Ireland
trisogor@ie.ibm.com

CONTENTS

FIGURES

TABLES

1 OVERVIEW

1.1 INTRODUCTION

The digitization of the manufacturing sector has enabled the availability of large volumes of data, this allows industry to embrace machine learning (ML) algorithms to identify and analyze patterns in the data. These patterns can be leveraged to enable decision support for activities including process and production planning, asset performance management and energy efficient manufacturing practices etc. The accuracy and robustness of ML algorithms depend heavily upon the amount, variety, veracity of training data.

To this end, a data-sharing ecosystem is encouraged where organizations participate and contribute their data as a strategic resource for the benefit of all. The incentive for participating in such an ecosystem is the value generated by access to a larger pool of data for model training. However, due to industrial competition and data privacy concerns, organizations are reluctant to share potentially commercially sensitive data, thus, the datasets remain in silos.

Federated learning (FedL) offers a potential solution to address the conflict between data protection and participation in a data sharing ecosystem. FedL enables organizations to collaboratively train robust AI models, without the need to directly share sensitive data with each other. Despite several contributions in domains such as natural language processing and healthcare, multiple barriers exist that are hindering the uptake of FedL in the manufacturing industry.

A key challenge is the complexity associated with designing and deploying a FedL solution. It requires consideration of many constraints such as the application type, FedL client configuration, global model training orchestration, choice of encryption mechanisms to secure models and incentive mechanisms. Currently, there is a lack of a clear methodology that allows practitioners and industry stakeholders to identify and evaluate the potential of using a FedL approach for their specific use case scenarios.

While the literature explores each of the design constraints independently, there is a need to consolidate these into a common framework to support the development of FedL solutions for smart manufacturing.

1.2 PURPOSE

The purpose of this paper is to translate practical experience of designing, building and deploying FedL solutions as well as an analysis of current literature into guidance that:

- provides a perspective on the business models that enable and foster a collaborative ecosystem to increase participation in data sharing arrangements that are critical to the development of FedL solutions.

- identifies the key technological enablers (i.e., digitization) needed to support the application of FedL solutions.
- proposes a set of characteristics that align use case scenarios with the benefits of a FedL approach in a smart manufacturing context.
- offers guidelines for FedL implementation and deployment in the context of use case scenarios in smart manufacturing.

1.3 SCOPE

The scope of this paper is to explore the design requirements and implementation strategies for the use of federated learning within the manufacturing domain. Specifically, this will provide insight into approaches for creating a collaborative ecosystem where organizations can mutually benefit from robust machine learning models. This requires making high-quality data sets that are typically beyond the reach of any single organization accessible to all participants in a secure and privacy preserving manner.

1.4 STRUCTURE

The document is organized as follows:

- Chapter 1 – Introduction
- Chapter 2 – Motivation
- Chapter 2 – Design considerations for FedL enabled collaborative ecosystem
- Chapter 3 – Value of collaborative-ecosystem: potential business models
- Chapter 4 – Guidelines for implementing FedL in smart manufacturing
- Chapter 5 – Use case implementation
- Chapter 6 – Conclusion

1.5 AUDIENCE

Manufacturing industry practitioners and applied researchers working towards realizing a collaborative ecosystem using federated learning.

1.6 USE

To design the solution architecture(s) of federated learning enabled manufacturing use cases.

1.7 TERMS AND DEFINITIONS

The following terms and definitions that are key to understanding this document are:

- ML – Machine Learning
- FedL – As per Kairouz, 2021, "*Federated Learning is a machine learning setting where many clients (e.g., mobile devices or whole organizations) collaboratively train a model*

under the orchestration of a central server (e.g. service provider), while keeping the training data decentralized."'

- OEM – Original Equipment Manufacturer

2 MOTIVATION

Due to growing competition and data privacy concerns many organizations are reluctant to share their data with each other or on cloud infrastructures (for data pooling); and thus, are deprived access to the variety and veracity of having data gathered from multiple sources to train ML models (Mohr, 2021). This also hinders the potential to unlock value from unused datasets. FedL enables organizations to mutually benefit from each other's data by collaboratively training robust ML models without having to share their raw data (Kairouz, 2021).

A FedL setup typically consists of several iterative phases to support model training which is initialized by FedL clients downloading a common model from a trusted centralized server. Clients proceed to train the model using data collected locally. Once the model is trained, the client shares only the updated model parameters with the trusted server. This is followed by the aggregation of the received model updates (from all clients) to create an updated global model that can be downloaded by the clients for the next iteration of training. The process terminates once the clients reach a consensus regarding the optimally of the global model. This approach allows the model to be exposed to a significantly larger pool of data that would be impossible for a single organization to possess alone.

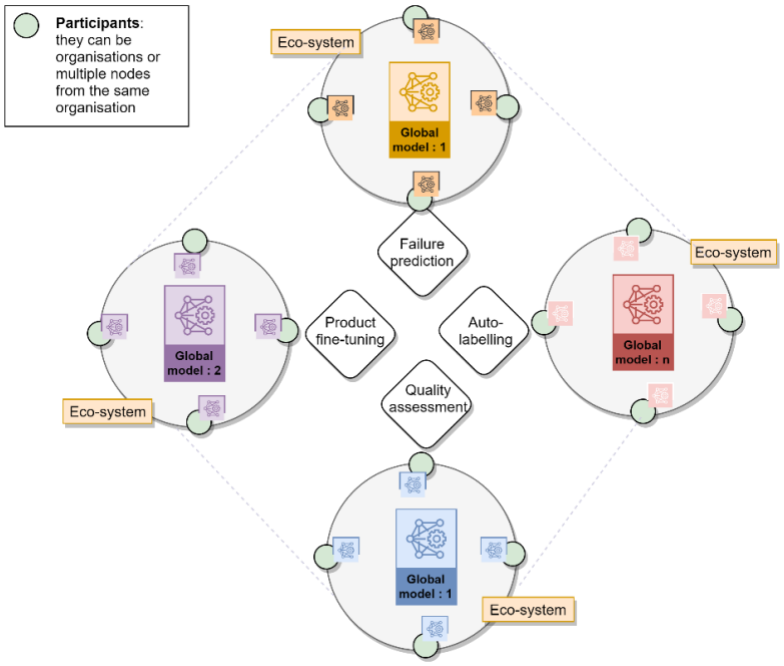

Figure 2-1: Ecosystems and Smart-Manufacturing Use Cases for Participants

The first step towards implementing FedL in smart manufacturing scenarios is to identify an appropriate use case for FedL, i.e., *where collaborative model training and/or sharing can*

improve the performance of the predictive model owned by a standalone client. Subsequently, this will bring potential organisations together to form an ecosystem in which, data can act as a strategic resource for the benefit of all. Figure 2-1 presents an example network ecosystem aligned with some example scenarios that benefit from FedL applicable in the context of smart manufacturing:

2.1 PRODUCT FAILURE PREDICTION

Large scale organizations can use on-premises computational nodes to train predictive models supporting product failure prediction using a variety of data gathered from manufacturing sites, thus forming localized data silos. On the contrary, small-scale organizations can often lack access to large training datasets and as a result develop less robust models. FedL can help in such scenarios where similar equipment operating on multiple sites of different organizations form a virtual network and share their failure patterns with each other to reach a consensus about the failure prediction model.

2.2 AUTO-LABELLING

Prediction models deployed by an organisation require an initial supervised training procedure with the labelled data. Usually, labelling is performed manually by subject matter experts. To save time, auto-labelling using transfer learning is also feasible. Auto-labelling for un-seen data can benefit from FedL as a new participatory agent (organisation) can connect with the FedL ecosystem and gain access to a robust model trained with a variety of patterns observed by other agents over a longer period.

2.3 PRODUCT OPTIMIZATION BY ORIGINAL EQUIPMENT MANUFACTURERS (OEM)

OEM provides equipment to multiple organisations and has the capacity to monitor its performance over time. This data can be utilised to optimise/fine-tune the performance of the product for their clients. Additionally, it can provide an opportunity to minimise downtime, pre-empt procurement delays by automating the procurement process of parts/equipment based on remaining useful lifetime of equipment or its components. However, OEM is often unable to gather the data from multiple organisations due to privacy and trust issues. FedL can be leveraged in such scenarios where an OEM distributes a set of global functions to the organisations along with the equipment. The functions take gathered equipment data as input and return the computed values to the OEM.

2.4 PRODUCT QUALITY ASSESSMENT

On-time product quality assessment is an important step towards cost savings and zero-defect manufacturing. An initial model can be deployed on factory floors to classify damaged products. However, it is observed (Mohr, 2020) that different damages in the same product can be

observed by different manufacturers. FedL can enable clients to exchange the information about the observed damage with each other so that a more robust model can be built.

To maximize the opportunities of a FedL solution in the context as described above it is important that practitioners utilize a formal approach to capture all characteristics of the use case and influence design choices. The motivation of this paper is to present design considerations and guidelines that are derived based on experience of successful deployments of FedL in industrial settings (Bharti, 2021). In particular, it has been demonstrated that FedL offers significant benefits in use cases such as remaining use-full lifetime (RUL) prediction and vision-based quality inspection. The lessons learned from real-world implementation offer an opportunity to provide insight and decision support for managers, developers and system integrators through a consolidation of options, constraints and design requirements that are typically faced in designing FedL solutions.

2.5 Design Considerations for FedL Enabled Smart-Manufacturing Ecosystem

As shown in Figure 2-2, there are five key considerations that should be taken into account to successfully design a typical FedL ecosystem within the context of smart manufacturing. There are existing data-sharing ecosystems that leverage similar design principles such as the methodology proposed in “Redesigning Trust: Blockchain Deployment Toolkit, 2020” from the World Economic Forum (WEF). This aims to support organizations in deploying blockchain based solutions with the key design considerations. The following derives a set of design questions specific to a FedL solution, that can be utilized by a practitioner to solicit ecosystem requirements, these will be explored further in subsequent sections.

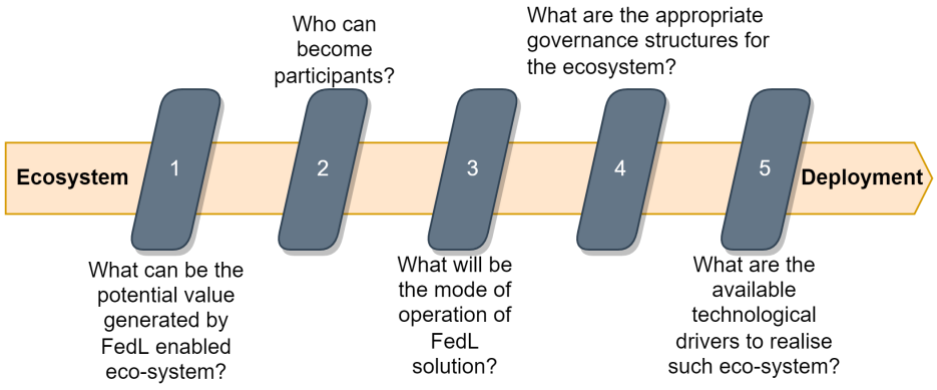

Figure 2-2: Workflow of a FedL Ecosystem

Step 1: Identify the potential value of participating in a FedL ecosystem

- What are the business challenges that can be solved by implementing a FedL ecosystem?
- What are the set of values that can be generated by the ecosystem which can be garnered by participatory organizations?

- What is the economic viability of the ecosystem (IEEE Std 3652.1-2020, 2021)? For example, what are the incentive provisions for clients to participate in such ecosystem? What is the return on investment (ROI) of participation?
- What are the potential risks (e.g. data protection, cyber security) and barriers (e.g. economic, regulatory) for participation in a FedL ecosystem?

Step 2: Potential participants

- Who can be the targeted participants based on the problem at hand and the proposed business model?
- What is the potential value that can be generated by each participant that cannot be achieved individually?
- What is the minimum number of participants required to sustain such ecosystem?

Step 3: FedL mode of operation

- What are the typical requirements (i.e., quick response time, high accuracy etc.) of the problem addressed by the FedL solution?
- Do participating organizations have suitable resources (computational and communication) to meet these requirements?

Step 4: Governance structures

- Who will orchestrate the global model training in each of the FedL mode of operation?
- What are the provisions to ensure resource (computational and communication) allocation for model training and update among participatory clients?
- What provisions are to be taken to ensure the security and quality of the shared model updates?
- What provisions are to be taken to prevent the centralized orchestrator to be compromised?

Step 5: Available technological drivers

- What are the available technological solutions to realize the FedL ecosystem?
- How prepared are participatory organizations for an initial development of such ecosystem?

Answering design questions pertaining to each step is important to gather the solution requirements and influences the methodology taken for implementing the technological solutions required. This may also include sensitizing potential participants about the business model and value that can be generated from participating in such ecosystem and preparing them fully before adapting the FedL solution.

3 VALUE OF COLLABORATIVE-ECOSYSTEM: POTENTIAL BUSINESS MODELS

The selection of the right use case is driven by a detailed assessment of potential collaborative business models which are shared among the targeted organizations for consideration and consent of participation. Betti (2020) defined a five-part framework to help organizations:

1. Understand the business challenges.
2. Develop applications to help overcome the business challenges.
3. Determine the viable applications.
4. Identify suitable collaborators for each application.
5. Define the data-sharing relationship.

For FedL, collaborations can be one of two models, sharing or allocation (Man, 2019). In the sharing model, collaborators can work together on with a pre-agreed split to capture value and realize greater scale or network effects that the pooling of data can provide, that is, the value of the data and information gained from it will increase when it is aggregated and shared. In the allocation model, value is created by optimally allocating risks between collaborators thereby allocating greater profit to the partner managing more risk (or more contributing data) in a FedL alliance. For example, as (Yang, 2019) has identified, FedL could make equitable rules for profits allocation with the help of consensus mechanism from Blockchain technologies.

At present, the primary approach to FedL alliances is via data sharing platforms provided by independent 3rd parties, who are not participants in the alliance. Such platforms may be attractive to organizations who cannot build their own platforms for data sharing or who want to outsource the challenges of managing data transactions, matching, licensing, and transfer of data (Richter, 2019). These platforms also have the benefit of solving the problem of trust, not only with having the platform being provided by a neutral 3rd party, but also through features such as certification, secure access control and digital watermarking (Richter, 2019). An example of this is the FedAI Ecosystem developed by WeBank, which offers an IoT driven data sharing platform, with applications in the areas of vehicle insurance, financial lending, and anti-money laundering.

Another scenario that utilizes the FedAI ecosystem is to create and online visual object detection platform to support collaborative object detection and minimize excessive data transfer to central cloud servers from cameras deployed in the field. Another example is IBM's Maximo visual inspection that has presented such a federated system for vision-based defect detection in manufacturing products (Bharti, 2021). Such Industrial IoT use cases can prove to be crucial in organization's growth in a competitive environment where if the defected products are delivered to customers this could not only hamper the organization's reputation and future business prospects but also result in significant losses in terms of raw material wastage and cost of recycling.

Another interesting example of this approach applied in the health industry is the MELLODY project, a consortium of 17 European partners is funded by EU Innovative Medicines Initiative (IMI) as a public–private partnership. The MELLODY alliance uses the Owkin Connect platform to manage the FedL activities that makes sure that the data (e.g. generated from personal medical devices, patient records, etc.) never leaves the data owner's system and only non-sensitive models are exchanged. To provide full traceability of the operations, the platform is based on a private Blockchain, with de-centralized control.

Future FedL alliances may follow this approach as it outsources the technology considerations to third parties, which eliminates the cost of building the technology. This also introduces an independent third party for whom there is no competitive advantage to be gained by having access to the shared data. The technology itself may solve some of the traditional challenges with data sharing alliances.

Key to applying FedL as an enabler of collaborative business models are examples of successful projects, products, and enterprises that can point to this technology as being the key differentiator. The business community will look at how successful projects like the MELLODY and FedAI ecosystem meet these challenges, before deciding to embrace FedL at scale. Data-driven and collaborative smart-manufacturing requires the addition of a new set of skills, there remains significant knowledge gaps between machine learning specialists and domain experts. New technologies such as FedL will act as fundamental enabler to deliver new modes of interaction that underpin existing and future workflows and there is a need to ensure employee support to maximize the potential of these technologies. Collaborative digital platforms support the involvement of stakeholders in the process creating a community of practice (Fit4FoF, 2018) for the design, definition and implementation of new programs and are required to provide remote, interactive upskilling and operate in tandem with workers across full process areas.

4 GUIDELINES FOR FEDL IMPLEMENTATION IN SMART MANUFACTURING

Once the business benefit of a FedL approach has been identified the next step is to define the implementation approach for the given use case scenario. To support this a decision model (Figure ***4-1***) is proposed that captures the pathway for implementing FedL solutions along with the current available technological drivers to realise the FedL ecosystem (i.e., choosing the correct building blocks of a FedL solution). The following outlines the main pathways for the proposed decision model.

4.1 CHOICE OF PARTICIPATORY CLIENT

Organizations may participate in the FedL process by using siloed data (cross-silo) on their geo-distributed data centers or on distributed IoT edge devices (cross-device). This is enabled by finalizing the design considerations (Section 3) for participating in the ecosystem. When the data across multiple clients share the same features but belong to different sample IDs, it is known as *horizontal FedL* and is mainly used to increase the variety and velocity of the input data.

On the other hand, in *vertical FedL*, data across multiple clients overlap on sample IDs but not on features. This is often used when clients participate to share the missing/intermediate features of the dataset. An example of cross-silo-vertical FedL in smart-manufacturing is monitoring the health degradation of an industrial bearing. This can be done by capturing images in a time-series and/or by recording parameters such as vibration, temperature etc.

Cross-device FedL is the most used paradigm of FedL where clients (>10,000) are un-reliable, state-less, and usually participate in a *horizontal FedL* of a light-weight predictive model. An example of cross-device FedL for smart-manufacturing is the collaboration of similar assets across multiple sites of organizations where agile edge analytics is important. The connected edge devices form network clusters i.e., all edge devices in a single cluster possess IID data about the same type of asset.

Digital twins can be the potential technology drivers for cross-device FedL as the concept provides a basis to make the transition from standalone, relatively unintelligent systems to a network of "intelligent" objects on the internet, facilitating and fueling the development of new value-added services enabled by access to data/model extracted from distributed physical assets.

Resource-constrained IoT edge devices are used as FedL clients if the global model is lightweight in terms of the number of model parameters and training data size. To validate this, the authors conducted experiments on single board computers (Raspberry pi) with 2GB RAM. A lightweight artificial neural network (ANN) with 2-hidden layers was used as a global model to be trained among three Raspberry pi devices and a centralized orchestrator (Laptop machine). The experiment results showed that almost 60% of the memory was occupied during the model training process with >85% CPU utilization throughout the experiment. This shows that such resource constrained IoT edge devices will not be able to train a more complex ML model such as deep learning, for which the edge/cloud servers are the appropriate choices.

On the other hand, the clients involved in cross-silo FedL are limited (1-100), reliable, state-full and equipped with enough computational resources. Interoperability and complex data ownership structures remain a challenge in the implementation of both cross-device and cross-silo FedL, as such the use of standards is encouraged. In the case of manufacturing standards such as OPC UA and Asset Administration Shell (AAS) are emerging as key approaches to support information modelling that can be common across all participants in the FedL network. IEEE standard 3652.1™-2020 defines an architectural framework and requirements for the application of FedL. It sets out to act as a guide to promote the use of distributed data sources and FedL without violating regulations or ethical considerations.

4.2 Choice of FedL Mode of Operation

In cross-silo FedL, any of the organisations can be elected as a trusted orchestrator and can manage the global model training task. However, when competing organisations are involved, electing a single orchestrator is often un-desirable considering the cost of manufacturing data

leakage. Thus, a fully decentralised or hybrid control is preferred where interaction is peer-to-peer among organisations. However, a decentralised consensus over the global model may take more time as compared to centralised aggregation. Moreover, a trustworthy centralised authority may still be in the charge to decide upon the global model architecture, initial hyper-parameters setting and debugging.

Implementing fully decentralised orchestration in cross-device FedL involves a huge number of un-reliable clients inching towards the consensus which may result in increased response time. However, managing a huge number of clients by a single orchestrator is also not feasible in terms of monitoring the edge resource information for appropriate client selection. To this end, a local edge server can act as an interface to the global server. The global model can be downloaded beforehand at the local edge servers also gathering the model parameter updates from clients and relaying these to the global server. This hybrid orchestration can unlock the potential of real-time data analytics in true sense by minimizing the prediction delay and maximizing the reliability of clients.

Such hybrid orchestration is realized and tested in one of our previous contributions (Bharti, 2021) in this area. The experiment results showed that utilising edge servers as interaction points for edge devices prevents client failure and minimizes the model convergence time. On the other hand, the model accuracy and convergence time suffers if all the clients are to directly interact with the global server.

As Big Data and deep learning (Kotsiopoulos, 2021) algorithms are key technological enablers for implementing cross-silo FedL, cross-device FedL on the other hand is mainly driven by technologies such as edge computing and 5G to support real-time data analytics.

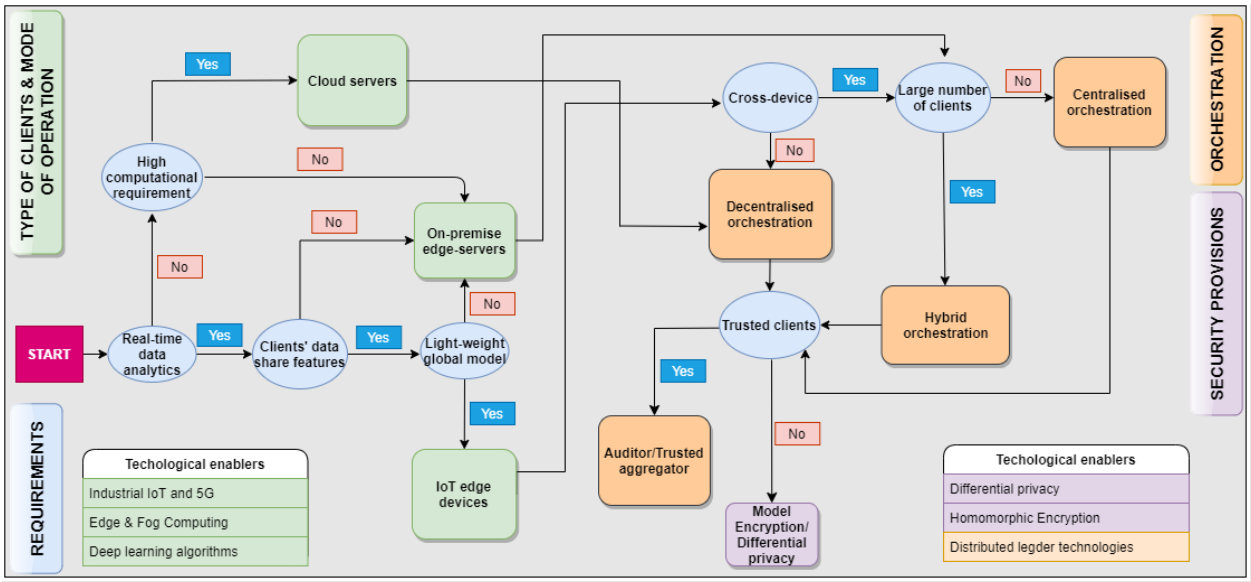

Figure 4-1: FedL Implementation Decision Model

4.3 Global Model Training Orchestration and Security Provisions

As the original data can be re-constructed from shared model updates (Kairouz, 2021) there is a requirement to further protect sensitive manufacturing raw data against the attackers. Traditionally, trusted aggregation mechanisms are employed to deal with this issue. However, it becomes a major challenge in cross-silo FedL across multiple manufacturing organizations.

Due to the preferred fully distributed control in such scenarios, a consensus relating to the global model can be achieved by utilizing Distributed Ledger Technologies - DLTs (Isaja, 2018) such as Blockchain. A corner stone to any FedL process is data integrity and DLT provides intrinsic properties that can ensure data integrity along a value chain and as such provides a single source of truth that can be used to build reliable models and analysis.

Type of Clients			Orchestration			Mode of operation		Technologies
IoT Edge devices	Edge servers	Cloud servers	Centralized	Hybrid	De-centralized	Cross-device	Cross-silo	
✓			✓			✓		5G, IIoT, Edge Computing, Light-weight machine learning, Digital twin
✓				✓		✓		
✓			✓				✓	
✓				✓			✓	
	✓		✓			✓		Deep learning, Differential privacy, Homomorphic encryption
	✓			✓		✓		
	✓			✓			✓	
		✓			✓		✓	Big data, Deep learning, Homomorphic encryption, Distributed ledger technologies

Table 4-1: Possible Configurations for FedL Implementation and Required Technological Drivers

Many platforms are quickly emerging that provide a marketplace for data and model sharing (Open Application Network, dataspace, IoTA). These marketplaces leverage Blockchain to provide a mechanism for producers and consumers of datasets that can leverage vast quantities of data in a collaborative, fair and transparent manner. DLT offers the opportunity to monetize data exchange thus incentivizing organizations to collaborate and share model updates.

Although research into model governance for FedL is still relatively immature, opportunities exist with Blockchain to ensure trust in the models being deployed for AI based systems, especially those shared, licensed, or purchased from third parties. In the context of FedL, if a client is compromised to enter incorrect model update (model poisoning attack (Chen, 2021)), Blockchain provides a consensus mechanism about the quality of the model update received from that client i.e., whether to accept the model update or reject.

One such example is proposed by (Zhang, 2021) which utilizes blockchain to build a data-sharing sharing platform for manufacturing organizations. The organizations are categorized into: (1) client organizations and (2) server organizations. Clients train the federated model while the servers owned the data-sharing platform and orchestrated the model aggregation process.

Another example of utilizing blockchain for decentralized model aggregation is proposed in (Zhao, 2021) where models related to home appliances were used to fine tune their performance i.e., energy consumption. Blockchain is again used as a model sharing platform where consumers can upload their specific local models to be aggregated by dedicated miners on the blockchain.

Given the advantages, Blockchain also faces some challenges which need to be addressed before utilizing it for FedL. In such de-centralized setup, clients must share their model updates with each other to reach to the consensus about the optimality of the global model. This may expose an organization's local model parameters to another competing organization.

Thus, another layer of security at client level is needed on top of DLT to protect the sovereignty of the model updates. Solutions such as differential privacy (DP) (Choudhary, 2019) and homomorphic encryption (Shreshtha, 2019) can prevent the re-construction of raw data from client model update. However, both are time consuming, and DP suffers in terms of model accuracy. They can be utilized along with Blockchain for cross-silo FedL whereas, they are not suitable for cross-device FedL where real-time data analytics is of paramount importance and thus a trusted centralized and/or hybrid orchestration should be an appropriate choice for such scenarios.

Once the appropriate client, mode of operation, security provisions and orchestration type are selected for the given smart manufacturing use case requirements, suitable technological drivers can be derived according to the possible FedL solution configuration (Table 4-1).

5 USE CASE IMPLEMENTATION

While the FedL implementation guidelines proposed in section 5 can be leveraged to support all use case examples as set out in section 2, this section presents the use of the decision model shown in Figure 4-1 in the context of product optimization to identify the implementation guidelines.

Product Optimization by Original Equipment Manufacturers (OEM)

This is a typical cross-silo FedL setting where multiple organizations in possession of equipment provided by an OEM, are willing to participate in a collaborative ecosystem to improve the performance of the equipment but stipulate that they cannot sharing raw equipment data with each other or with the OEM. As the variety and veracity of the equipment data gathered from one manufacturing site is not enough to mine complex data patterns, the objective is to gather the data around the equipment operating in different working conditions at their corresponding organizational sites and perform a collaborative model training using FedL.

- *Type of clients and mode of operation:* Data gathering associated with operational efficiency of equipment is typically collected in parallel to equipment's normal operational tasks and thus the product (equipment) fine-tuning/optimization does not demand real time data analytics. As per the decision model, the type of clients in this use case can be deployed either as on-premises edge servers if the global ML model is not computationally demanding or on cloud servers to support a computationally demanding model training task. This decision is driven by the amount of training data and frequency of model exchanges between participatory organizations.
- *Global model training orchestrator*: In such cross-organizational settings, participatory organizations may be hesitant in letting a single organisation orchestrate the global model training. As per the decision model, a decentralized orchestration is suited for this use case where a third party (chosen by participatory organizations) or the OEM itself can become the orchestrator. The OEM can also employ an interesting business model where the cost of model training orchestration can be accommodated by receiving a copy of the trained global model. It may benefit OEM to produce/supply more optimized products in the future. Current third-party (apart from OEM) solutions in this direction provide a secure model/data sharing platform to be utilized by participatory clients during model exchanges.
- *Security provisions*: The use case is a typical cross-silo FedL setting in which participatory organizations cannot afford to expose even their trained model parameters to each other or to the orchestrator. Thus, additional security provisions are needed in this use case to prevent any sensitive data/model leakage. In addition to that, a more secure data-sharing platform must be utilized to keep transactional (model exchanges) records intact during and post model training.
- *Technological enablers*: Sophisticated deep learning models and computationally rich model training platforms are key enablers for this use case. In addition to that, DLT such as Blockchain and model encryption techniques such as differential privacy and homomorphic encryption should be utilized for secure and auditable model exchanges.

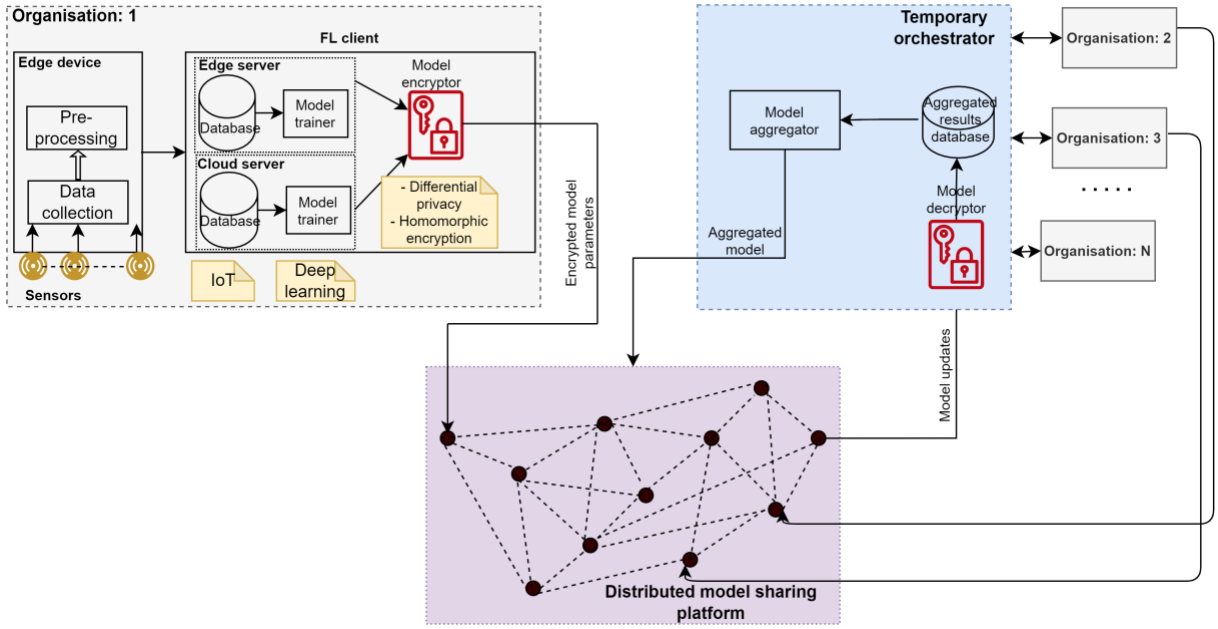

Figure 5-1: Use Case Design Architecture for Implementing FedL Solution

A high-level architecture for implementing FedL solution for this use case is presented in Figure 5-1. The key architectural components are (1) Participatory organizations, (2) Temporary orchestrator and (3) DLT platform. The interaction between clients (organizations) and aggregator(s) takes place via DLT platform where each node in the ledger represents a client. The key technological enablers along with design features are highlighted in the architecture.

6 CONCLUSION

This paper presented design considerations and guidelines for implementing FedL solutions for smart manufacturing applications. Sample use cases of FedL in manufacturing, followed by key design requirements to realise a FedL ecosystem are discussed in detail. A brief introduction to the potential business models to be leveraged by the FedL ecosystem is also presented. Finally, a decision model about various FedL architectural components (i.e., type of clients, mode of FedL operation & global model training orchestration) is proposed to support the choice of suitable FedL configuration for the given use case. Future work will include further evaluation of the effectiveness of the proposed decision model with practitioners and relevant stakeholders. Emphasis will be placed on its applicability this across multiple industrial IoT sectors and use case scenarios.

7 REFERENCES

Mohr, M., Becker, C., Moller, R., Richter, M. (2021). Towards Collaborative Predictive Maintenance Leveraging Private Cross-Company Data. *In: Reussner, R. H., Koziolek, A., & Heinrich, R. (Hrsg), INFORMATIK. Gesellschaft fur Informatik, Bonn.* (S. 427-432).

Bharti, S., McGibney, A. (2021). Privacy-aware Resource Sharing in Cross-device Federated Model Training for Collaborative Predictive Maintenance. *IEEE Access*, 9, 120367-120379

Kairouz, P., Mcmahan, B., Avent B. et al. (2021). Advances and Open Problems in Federated Learning. *Foundations and Trends in Machine Learning,* 14(1), 1-210.

Redesigning Trust: Blockchain Deployment Toolkit (2020). *https://widgets.weforum.org/blockchain-toolkit/*

IEEE Guide for Architectural Framework and Application of Federated Machine Learning (2021). *in IEEE Std 3652.1-2020,* 1-69.

Man, A., Luvison, D. (2019). Collaborative business models: Aligning and operationalizing alliances. *Business Horizons*, 62(4). 473–482.

Bharti, S., Bringing AI to your data and improve vision-based product quality inspection (2021). *https://www.ibm.com/blogs/internet-of-things/vision-based-product-quality-inspection/*

Yang, Q., Liu, Y., Chen, T., Tong, Y. (2019). Federated machine learning: concepts and applications. *ACM Transactions on Intelligent Systems and Technology,* 10(2). 1-19.

Richter, H., Slowinski, P. (2019). The data sharing economy: On the emergence of new intermediaries, *IIC - International Review of Intellectual Property and Competition Law*. 50. 4–29.

Federated AI Ecosystem: Collaborative learning and Knowledge Transfer with Data Protection (2020). *https://www.fedai.org/*

MELLODY: Machine learning ledger orchestration for drug discovery (2019). *https://www.melloddy.eu/*

Fit4FoF: Making our workforce fit for the factory of the future (2018). *https://www.fit4fof.eu/*

Kotsiopoulos, T., Sarigiannidis, P., Ioannidis, D., Tzovaras, D. (2021). Machine Learning and Deep Learning in smart manufacturing: The Smart Grid paradigm. *Computer Science Review.* 40. 100341.

Isaja, M., Soldatos, J. (2018). Distributed ledger technology for decentralization of manufacturing processes. *In 2018 IEEE Industrial Cyber-Physical Systems (ICPS)*. 696–701.

Chen, Z., Tian, P., Liao, W., Yu, W. (2021). Towards multi-party targeted model poisoning attacks against federated learning systems. *High-Confidence Computing,* 1(1). 100002.

Choudhary, O., Divanis, A., Salonidis, T., Sylla, I. et.al. (2019). *https://arxiv.org/abs/1910.02578*

Shrestha, R., Kim, S. (2019), Chapter Ten - Integration of IoT with blockchain and homomorphic encryption: Challenging issues and opportunities. *Advances in Computers.* 115. 293-331.

Betti, F., Bezamat, M., Fendr, B., Fernandez, D., et al. (2021). Share to Gain: Unlocking data Value in Manufacturing. *Technical report World Economic Forum*.

W. Zhang et al. (2021), Blockchain-Based Federated Learning for Device Failure Detection in Industrial IoT, in *IEEE Internet of Things Journal*, 8(7), pp. 5926-5937.

Y. Zhao et al. (2021), Privacy-Preserving Blockchain-Based Federated Learning for IoT Devices, in *IEEE Internet of Things Journal*, 8(3), pp. 1817-1829.

8 ACKNOWLEDGEMENTS

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement No. 847577; and a research grant from Science Foundation Ireland (SFI) under grant number 16/rc/3918 (Ireland's European Structural and Investment Funds Programmes and the European Regional Development Fund 2014-2020).

- Return to *IIC Journal of Innovation landing page* for more articles and past editions