Federated Learning in Cloud and Edge Environments: Privacy, Communication, and Application Perspectives

1st Dhvanil Patel
Institute of Technology
Nirma University
Ahmedabad, India
25MCE006@nirmauni.ac.in

2nd Pallav Shah
Institute of Technology
Nirma University
Ahmedabad, India
25MCE025@nirmauni.ac.in

Abstract—**The rapid growth of data-driven applications has brought us to immense problems in terms of privacy and communication, efficiency and scalability in conventional centralized machine learning systems. Federated Learning is emerging as A transformative paradigm that would enable collaborative model training across distributed clients without directly sharing raw data. This methodology also corresponds very well with principles of cloud computing, where computational resources and data are inherently distributed across a number of nodes and regions. Leveraging cloud infrastructures, federated learning enhances computational scalability, fault tolerance, and deployment flexibility enabling organizations to train global models efficiently while preserving User data privacy.**

This survey explores the integration of federated learning. cloud computing environments, with a focus on three major aspects: applications, communication architectures, and privacy preserving mechanisms. We analyze the role of cloud-enabled federated systems in fields including healthcare, IoT, and smart industry, highlighting how cloud orchestration supports multi-client coordination, and model aggregation. Furthermore, we have shown key strategies to optimize communication, including asynchronous aggregation, compression techniques, and edge cloud collaboration to minimizing bandwidth costs and training delays. we also reviewed Various privacy and security related aspects, including differential privacy and secure aggregation and blockchain-based trust frameworks. this paper provides a comprehensive understanding of current trends, challenges, and research opportunities in federated learning on cloud platforms. We conclude that Effective Integration of federated learning with Cloud and Edge Paradigms represents a significant step toward realizing large-scale, privacy aware, and energy-efficient intelligent systems.

Index Terms—**Federated Learning, Cloud Computing, Edge Computing, Privacy Preservation, Distributed Training**

I. Introduction

The increase in data-driven services in the cloud era has led to a paradigm shift in how organizations train and deploy their machine learning models. Traditional centralized learning approaches require aggregating large volumes of raw data from distributed sources into a one central server. While effective in homogeneous and controlled environments, such architectures raise critical concerns regarding data privacy, latency, communication overhead, and regulatory compliance [1]. To address these challenges, federated learning has emerged as a decentralized machine learning paradigm where multiple clients, for example, mobile devices, IoT sensors, or enterprise servers, collaboratively train a shared global model under the coordination of a central cloud aggregator without sharing raw data [3].

This naturally matches the core principles of cloud computing, where the computational resources are dynamically provisioned, orchestrated, and scaled across heterogeneous nodes. The integration of federated learning into both the cloud and edge computing infrastructures is a key transformative advance towards privacy-preserving and communication-efficient artificial intelligence [2]. Shifting the computation closer to the sources of data while orchestrating via the cloud allows federated learning to minimize requirements on data transfers while maintaining global model accuracy [4]. This creates a synergy that is the basis of what is now called
textitCloud-Centric Federated Learning (CCFL), a field which fuses federated optimization, network management, and cloud resource allocation [6].

However, FL deployment in cloud environments poses several system-level and theoretical challenges. Communication efficiency is one of the most important issues. iterative model update exchanges among distributed clients and cloud servers might be time-consuming and bandwidth-heavy [3]. To address this challenge, efforts have been made to reduce network overhead without sacrificing model performance by leveraging techniques such as model compression, gradient quantization, update sparsification, and asynchronous aggregation [2]. Efficiency in communication must be balanced against fairness and scalability across clients with different computational capabilities and network conditions, common in multi cloud or edge cloud environments [6].

In addition to being efficient in communication, privacy and security form fundamental pillars in any FL system, especially when sensitive or regulated data, such as healthcare records or financial transactions, are at stake. Although FL ensures that raw data stay only on local, recent studies demonstrate that model gradients can still leak sensitive information via inference or reconstruction attacks [1]. Because of this techniques for preserving privacy, such as Differential Privacy, Secure Multiparty Computation, and Homomorphic Encryption, have become important components of the federated architecture [10]. In a cloud-hosted FL system, such mechanisms are further combined with transport encryption schemes and secure aggregation pipelines for data and model parameters to be protected while in transit [5]. The delicate tradeoff between the two objectives-privacy preservation and model utility-should be maintained; excessive noise addition or cryptographic overhead reduces model generalization capability and also system performance [10].

Beyond efficiency and privacy, real-world cloud-based applications of federated learning have demonstrated significant societal and industrial impact. For instance, FL in healthcare enables the collaboration of multiple hospitals to analyze medical images and forecast diseases without violating patient data confidentiality [4]. In the domain of IoT, federated frameworks make smart devices and sensors learn collaboratively from distributed data, enhancing services such as traffic prediction, energy optimization [5]. Equally, ecosystems of autonomous vehicles leverage cloud-enabled FL to continuously improve perception models across geographically distributed fleets while respecting data sovereignty constraints [6]. Ultimately, all these application domains have important implications for cloud infrastructures regarding scalable FL orchestration, offering high performance compute, global synchronization, and compliance management [3].

From the cloud computing perspective, federated learning transforms the traditional cloud-client relationship into a more collaborative and intelligent interaction. The cloud no longer acts as a centralized data processor but assumes an active role of federation coordinator, being responsible for task scheduling, model aggregation, fault tolerance, and trust management [2]. The further development of Edge and Fog computing extends this model even further by offloading partial computation to edge nodes, reducing latency and bandwidth utilization [3]. This is an optimal hierarchical topology for FL deployments. a cloud or edge device continuously changing where the data locality and scalability are achieved simultaneously. However, such integration introduces new orchestration challenges like client selection, resource heterogeneity, and multi-tenant model management are active research frontiers [6].

In addition to the technical challenges, governance and regulatory compliance form another critical dimension of federated learning in cloud contexts. Data protection laws, such as the General Data Protection Regulation and the Health Insurance Portability and Accountability Act, impose strict controls over how user data can be collected, processed, and transferred across borders. FL inherently enables compliance by avoiding raw data sharing, but meta-data leakage, model inversion, and jurisdictional inconsistency can still pose risks [1]. Therefore, cloud providers are increasingly integrating policy-aware federated orchestration layers, ensuring that model aggregation and data processing respect legal and ethical standards [10].

Given this multifaceted landscape, this survey paper provides a review of federated learning from the cloud computing perspective, where the focus is on three major perspectives.

- **Communication efficiency:** Focusing on bandwidth optimization and the cloud–edge coordination strategies [3].
- **Privacy and security:** Discussing theoretical and applied techniques that ensure confidentiality and integrity in federated learning environments [1].
- **Application domains:** Highlighting how federated learning empowers real-world cloud-integrated systems such as healthcare, IoT, and intelligent mobility [4].

This paper is structured as follows. Section II introduces the fundamentals and taxonomy of federated learning. Section III explores cloud–edge architectures and communication-efficient algorithms. Section IV Gives a brief overview of various types of FL algorithms and emerging methods. Section V discusses privacy preservation and governance challenges in FL systems. Section V outlines open research challenges and future directions before concluding in Section VI.

Through this synthesis, we aim to provide readers-particularly those in the cloud computing and distributed systems community with clear theoretical understanding of how federated learning is transforming the architecture, efficiency, and security of intelligent cloud ecosystems [6].

II. Background and Fundamentals of Federated Learning

A. Concept of Federated Learning

Federated Learning is a distributed machine learning technique which enables the training of models across multiple devices or organisations without sharing raw data with a central server [3]. The main goal of FL is to maintain data privacy while leveraging the distributed computational and data resources across clients. Unlike traditional centralized learning, where all data is aggregated on a single server for training, FL allows participants (clients) to locally train models on their private data, and merely exchange model parameters or gradients with a coordinating server [1]. This server aggregates the updates to form a global model, a process normally referred to as *federated averaging (FedAvg).*

Mathematically, if there are N participating clients, each holding local data D_i, the global objective function can be represented as:

$$\min_{w} F(w) = \sum_{i=1}^{N} \frac{|D_i|}{|D|} f_i(w) \tag{1}$$

where $f_i(w)$ denotes the local loss function of the client i, and $|D| = \sum_i |D_i|$ is the total data size. The aggregation process ensures that the final model w reflects patterns learned collectively from all the devices while maintaining data locality and confidentiality [4].

B. Types of Federated Learning

Federated Learning supports multiple data-partitioning scenarios, allowing it to operate across heterogeneous organisations, cloud environments, and multi-tier infrastructures.

Depending on how data is distributed among participating clients, FL is generally divided into three major paradigms.

- **Horizontal Federated Learning (HFL):** In HFL, clients share the same feature space but possess different user samples. This is the most widely deployed FL setting and is commonly used in cross-device environments, where large numbers of devices contribute small, non-identically distributed local datasets. For example, multiple hospitals can collaborate using identical electronic health record schemas but with distinct patient data. Many large-scale industrial FL systems (e.g., recommendation models, mobile keyboard prediction) adopt this paradigm due to its scalability and communication efficiency. Several research directions in HFL—such as secure aggregation, differential privacy, and communication compression—are extensively surveyed in recent FL security literature [1], [10]. In cloud ecosystems, HFL is often deployed through containerized clients operating on separated organizational silos allows for privacy-preserving model training without any exchange of raw medical or behavioral data [4].
- **Vertical Federated Learning (VFL):** VFL applies when clients have the same user data but maintain different feature spaces. This scenario is prevalent in corporate and cross-institutional collaborations, such as a bank and an e-commerce platform jointly training risk assessment or fraud detection models. Since matching of user identities is necessary, VFL relies more heavily on cryptographic techniques such as Secure Multiparty Computation (SMC) and Private Set Intersection (PSI), as surveyed in [9]. These methods ensure that sensitive identifiers can be aligned without exposing proprietary attributes. VFL is particularly important in privacy-restricted cloud ecosystems where organizations cannot share features directly but still benefit from collective intelligence [10].
- **Federated Transfer Learning (FTL):** FTL is used when clients differ in both user data and feature spaces, with only very minimal or partial overlap. This configuration is suitable for scenarios where data heterogeneity is extreme, such as cross-domain recommendation systems, multi-country enterprises, or organizations with inconsistent data schemas. FTL leverages transfer learning techniques to bridge representation gaps between heterogeneous domains and mitigate distribution mismatch. In cloud–edge architectures, FTL enables lightweight edge clients with a limited number of datasets to benefit from richer feature spaces available in the cloud. Recent studies on cross-silo and corporate FL [4] highlight that FTL is useful when data variability across organizations is too high for traditional HFL or VFL to converge efficiently [4].

These configurations make FL highly adaptable for cloud-based and big data ecosystems, where data sources are geographically distributed, semantically heterogeneous, and subject to privacy regulations. By supporting multiple data-partitioning strategies, FL enables collaborative intelligence across organizations while preserving confidentiality and legal and operational constraints [3].

C. Architecture of Federated Learning Systems

Federated learning system have several distributed components that work together to enable privacy preserving and collaborative model training. While the logical structure is simple, real-world deployments—especially those spanning cloud, edge, and cross organizational silos introduce additional architectural requirements related to security, scalability. The core architecture typically includes three major components:

- **Clients (Edge Devices or Organizational Silos):** Clients are the one who hold private, locally generated, or domain-specific data. Depending on the deployment, clients may be mobile devices, IoT sensors, enterprise servers, or entire corporate databases. Each client performs local model computation without exposing raw data to external entities. Cross silo FL systems, such as those explored in [4] paper, rely on enterprise level clients (eg.banks, hospitals, or media companies) that contribute significantly larger and more structured datasets than typical cross device scenarios. Similarly, cloud–edge collaborative studies [3] shown the role of edge servers as intermediate clients that reduce communication bottlenecks and perform hierarchical aggregation.
- **Federation Server (Global Aggregator):** The central orchestrator coordinates training rounds by initializing the model, selecting the clients, receiving encrypted or privacy preserved updates, and aggregating them (commonly using FedAvg or its optimized variants). The aggregator is also responsible for establish security protocols such as secure multiparty computation, homomorphic encryption, and differential privacy, as described in [1], [10]. In advanced deployments, the aggregator may itself be distributed across cloud data centers or implemented as a serverless orchestration layer, as proposed in EneA-FL [2], which dynamically selects clients based on energy constraints and system heterogeneity.
- **Communication Layer:** This layer allows for secure, bidirectional transfer of model parameters between clients and the aggregator. In cloud-native FL systems, communication channels are often encrypted and optimized through compression or sparsification to reduce bandwidth consumption. A number of works identify communication as a significant bottleneck in distributed learning, particularly for large scale or multi organization settings [8]. These challenges are ameliorated in cloud infrastructures by the availability of high-throughput networking, regional data centers, and distributed compute fabrics. In addition, blockchain-enhanced FL frameworks [5] bake trust and auditability directly into the communications layer, recording model updates on a decentralized ledger with no requirement for a fully trusted central server.

Some cloud computing platforms such as AWS SageMaker, Google Cloud AI Platform, and Azure ML serve as a robust backbones for deploying FL models due to their scalable compute clusters, managed orchestration services, container environments, and built-in privacy controls. Recent surveys on secure FL architectures [1], [10] shown that cloud environments are now able to support hardware-based encryption, secure enclaves, key management systems, and automated monitoring system, making federated learning more reliable for large-scale and cross-institutional deployments. Additionally, cloud–edge collaborative frameworks [3] illustrate how hierarchical model aggregation distributes computation across layers, improving latency, reducing energy consumption, and enabling FL to scale seamlessly across heterogeneous big data ecosystems.

D. Federated Learning Lifecycle

The operation of a federated learning system follows a proper structured, iterative workflow that coordinates distributed clients and global servers to collaboratively train the machine learning model without exposing any private data related to client. Although the lifecycle appears linear, several stages require careful orchestration in real world cloud and cross silo deployment, particularly due to some challenges such as client heterogeneity, communication overhead, and privacy preservation the typical FL lifecycle involves the following stages:

1) **Client Selection:** At the starting of each round, the federation server determines that which subset of clients will participating in training round. Selection can depend on many different factors such as device availability, network stability, data quality, geographic distribution, or energy constraints. Serverless orchestration frameworks like EneA-FL [2] dynamically select clients based on resource budgets and device heterogeneity to minimize training latency. In cross silo environments [4], clients often represent organizations rather than only devices, which make selection dependent on corporate agreements, policies, or data readiness.
2) **Model Initialization:** The server shares an initial global model or pre-trained parameter set to all the participating clients. In cloud edge architectures [3], initialization may occur hierarchically. the cloud sends the model to edge servers, which further distribute it locally to IoT devices. This transfer based initialization is vary common in federated transfer learning (FTL), where global models must align heterogeneous feature spaces across organizations.
3) **Local Training:** Each client train the received model using it's own local dataset for a fixed number of epochs. Since data distributions are often non-IID across clients, local training can introduce "client drift," affecting convergence stability. Studies on federated learning [7] highlight that clients may train model on evolving or domain specific data, requiring techniques to maintain performance across multiple task. Local training is typically supported by cloud managed execution environments such as containerized runtimes, secure enclaves, or edge accelerators.
4) **Model Uploading:** After training, clients send back updated model to their (gradients or parameter deltas) server. To preserve privacy, updates are often encrypted or masked using secure aggregation, differential privacy, or homomorphic encryption [1], [9]. Communication efficient strategies like sparsification, quantization, or asynchronous uploading [8] are essential in reducing uploading bandwidth consumption in large scale cloud deployments.
5) **Aggregation:** The server aggregates updates from participating clients to form an improved global model. The widely adopted FedAvg algorithm computes a weighted average of model updates, but recent studies propose alternatives that address non-IID data, fairness, and communication imbalance [10]. In cloud–edge hierarchies [3], aggregation may occur in multiple layers—edge servers combine local client updates before forwarding them to the cloud, reducing communication cost and improving latency.
6) **Model Evaluation and Redistribution:** The updated global model is validated on held out datasets or evaluated through client side performance reports uploaded by the client. Once the updated parameters improves accuracy thresholds or stability conditions, the model is updated and redistributed to the clients for the next communication round. In corporate FL settings [4], evaluation may also include fairness metrics or domain specific skills. this Redistribution process continues iteratively until convergence or until further communication is no longer cost effective.

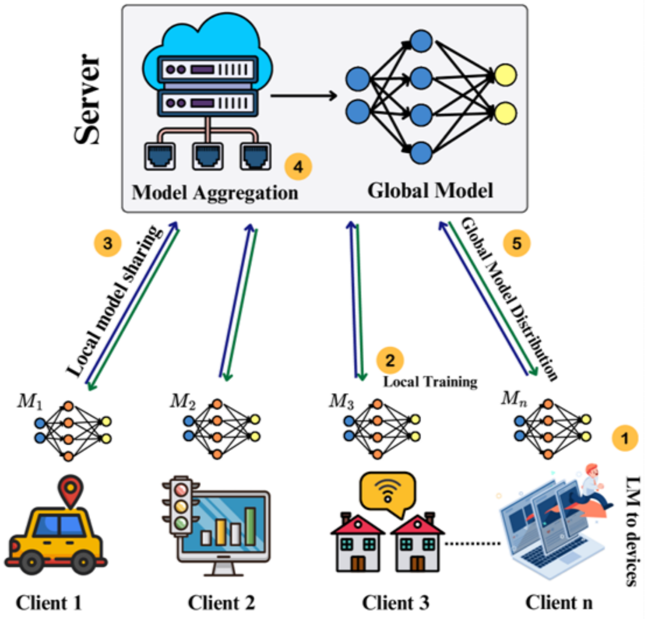

Fig. 1: federated learning lifecycle across distributed clients.

This iterative lifecycle continue running until the global model reaches satisfactory performance or until marginal improvements are outweighed by communication delays, resource consumption, or privacy utility trade-offs. Advanced FL architectures may also allow early stopping policy, partial client participation, or dynamic model personalization to furthermore optimize the lifecycle within large scale cloud and big data ecosystems [10].

E. Integration of FL with Cloud Computing

The interplay of FL and cloud computing is one of the cornerstones of modern distributed intelligence. Cloud environments provide elastic computing power, global data connectivity, and orchestration frameworks that enhance the scalability and reliability of FL deployments. Cloud based FL can leverage container orchestration like Kubernetes, data pipelines, and federated orchestration frameworks such as TensorFlow Federated or FedML.

Further, clouds serve as a trusted aggregator. This makes hybrid architecture viable, in which the computation is done on edge nodes and aggregation/model governance on cloud nodes. Integration reduces communication overhead [3] considerably and builds cross-silo collaborations over sensitive domains such as healthcare, finance, and IoT ecosystems.

III. Federated Learning in Cloud Computing

Federated Learning (FL) has emerged as a key paradigm for decentralized intelligence model, but its full potential is realized when we combined it with the scalability, reliability, and computational elasticity of cloud computing. This section explores how cloud based infrastructures enable efficient federated model training, address communication and privacy challenges, and facilitate cross domain applications such as healthcare, IoT, and finance [3].

A. Federated Learning and Cloud Synergy

The integration of the FL with cloud computing creates a multi-layered architecture where edge devices perform computation by their own while cloud servers handle model aggregation, coordination, and storage. Cloud based orchestration frameworks (e.g., AWS SageMaker Federated Learning, Google Cloud Vertex AI, and Azure FL Studio) provide APIs and secure data transfer protocols to manage thousands of clients at parallel.

In the typical setup, the cloud acts as a central node which provides efficient model distribution, monitoring, and performance tracking across multiple clients. This hybrid design enhances system scalability while maintaining data locality at the client side. The cloud edge continuum therefore forms a flexible computational pipeline [3] where real time data remains client side decentralized still collective intelligence continues to grow through iterative training cycles.

Cloud platforms also play a crucial role in resource allocation and the task scheduling. By leveraging container orchestration tools such as Kubernetes or Docker Swarm, cloud systems can dynamically assign training workloads, balance energy consumption, and ensure fault tolerance. This orchestration is vital for the large scale FL experiments [2], [4] where participants may have heterogeneous hardware configurations and unstable network connectivity.

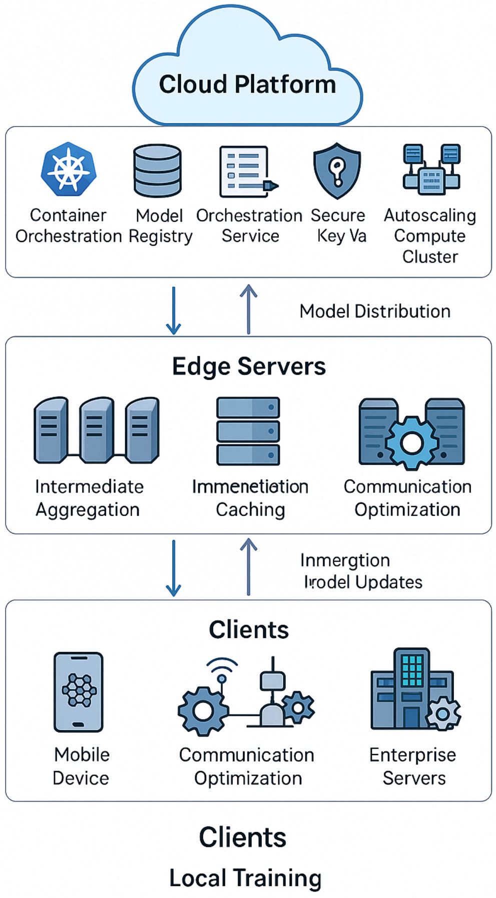

Fig. 2: Cloud-native federated learning pipeline.

B. Communication Efficiency in Federated Learning

One of the fundamental challenge of the FL is the communication bottleneck caused by frequent model updates between clients and the cloud server in very short time. Each communication round involves the transfer of model parameters or gradients, which can be vary significant for deep learning models with millions of parameters. In a cloud setting, this problem is exacerbated by the scale of participating devices or client and potential network latency [8].

To overcome this, researchers and engineers have developed several optimization strategies:

- **Model Compression and Quantization:** Techniques like weight pruning, sparsification, and gradient quantization reduce the data size transmitted to the cloud aggregator without significantly degrading model accuracy [8].
- **Client Selection and Sampling:** Instead of involving all clients in the every round, randomized or importance

based sampling ensures that only representative subset participates, thereby minimizing bandwidth usage [2].

- **Adaptive Federated Optimization Algorithms:** Algorithms like FedProx, FedNova, and SCAFFOLD improve convergence speeds and stability which reduce the number of required communication rounds [10].
- **Hierarchical and Clustered FL:** By introducing intermediate aggregators (edge servers), updates are first combined locally before sending it to the main cloud server this step significantly cutting down overall communication costs [3].

Cloud systems can further support these mechanisms by using content delivery networks (CDNs) and edge caching, ensuring low latency communication paths between clients and federators. This co-optimization of networking and computation forms a central advantage of cloud integrated FL architectures.

C. Privacy Preservation and Security in Federated Learning

While efficient communication is important for scalable federated learning, protecting privacy is still the primary motivation behind this paradigm. Since clients operate in distributed and potentially untrusted environments, cloud-enabled federated learning systems must ensure that model updates do not expose sensitive information through gradient inversion, membership inference, or reconstruction attacks. Recent surveys demonstrate that even partial gradients can leak significant private details [1], [10]. This makes strong privacy tools essential in both cross-device and cross-silo settings [4]. Several techniques are commonly integrated into modern federated learning systems:

- **Secure Aggregation Protocols:** Secure aggregation enables the server to compute the combined updates from clients without accessing individual contributions. This prevents leakage of client-specific information and is particularly important in enterprise and cross-silo FL scenarios [4]. Contemporary secure aggregation protocols rely on cryptographic masking, pairwise secret sharing, or distributed key generation, which align with secure computation strategies highlighted in recent FL security surveys [1]. Cloud-managed infrastructures further strengthen these protocols through secure key vaults and encrypted communication channels [3].
- **Differential Privacy (DP):** Differential Privacy introduces mathematically calibrated noise into gradients or model parameters, limiting the influence of any individual user or institution. DP is widely used in industrial federated learning due to its provable privacy guarantees [10]. However, excessive noise may degrade accuracy, especially under non-IID and heterogeneous data distributions. Cloud-based FL systems mitigate this by tuning privacy budgets, applying adaptive noise mechanisms, or aggregating DP-protected updates within hierarchical cloud–edge deployments [3].
- **Homomorphic Encryption (HE):** Homomorphic Encryption enables arithmetic operations directly on encrypted values, allowing servers to aggregate model updates without decryption. This is crucial when the aggregation server is semi-trusted or operated by an external cloud provider. Although computationally expensive, improvements in encryption schemes and hardware acceleration within cloud platforms reduce the overhead [1]. HE-based FL is valuable in multi-organization environments where raw gradients must remain hidden from all collaborators.
- **Secure Multi-Party Computation (SMC) and Private Set Intersection (PSI):** In vertical and cross-silo FL, organizations must securely match user identifiers without exposing proprietary features. PSI protocols, as detailed in [9], enable this matching while preserving confidentiality. SMC further allows secure joint computation of model updates without revealing intermediate values [1]. These approaches are essential in enterprise FL applications such as banking–e-commerce partnerships, cross-hospital research, and multi-institution analytics.
- **Blockchain-Integrated Federated Learning:** Blockchain introduces decentralization, transparency, and tamper resistance into FL workflows. Instead of relying on a single trusted server, model updates can be recorded as immutable ledger entries, ensuring auditability and providing trustless verification [5]. Smart contracts automate client selection, update validation, and incentive mechanisms, making blockchain-based FL appropriate for multi-stakeholder ecosystems.

In cloud environments, privacy protection is enhanced by data governance frameworks and compliance standards such as GDPR, HIPAA, and ISO/IEC 27018. Modern cloud platforms provide secure enclaves, hardware-backed key management systems, encrypted channels, and continuous compliance monitoring [3]. These capabilities ensure that FL workflows adhere to regulatory requirements and organizational privacy policies. Multiple FL security surveys emphasize that strong privacy-preserving techniques are essential not only for safeguarding sensitive data but also for fostering trust, enabling collaboration among distributed organizations, and scaling FL across heterogeneous cloud–edge ecosystems [1], [10].

TABLE I: Privacy-Preserving Techniques in Federated Learning

Technique	Advantages	Limitations
DP	Simple, strong privacy guarantees, easy to integrate.	Can lower accuracy; needs careful noise tuning.
SMC	No data exposure, suitable for cross-silo FL.	High computation and comm. cost.
HE	Operates on encrypted data, prevents leakage.	Slow and resource-intensive.
Blockchain	Adds trust, auditability, and transparency.	Latency and scalability issues.

However, privacy mechanisms often come at the cost of increased computation and communication overhead. Therefore, balancing privacy–utility trade-offs remains an active research area, particularly in multi-tenant cloud environments.

D. Attack Vectors in Federated Learning

Despite its privacy-preserving design, Federated Learning remains vulnerable to a wide range of attacks that exploit its decentralized and collaborative nature. The survey by Manzoor et al. [1] provides a comprehensive classification of attack vectors that target either the integrity of the global model or the confidentiality of client data. Understanding these threats is essential for securely deploying FL in cloud and cross-organizational environments, where system heterogeneity increases risk [3], [4].

1) Data Poisoning Attacks: In data poisoning attacks, malicious clients intentionally manipulate their local training data to negatively influence the global model. Since the server cannot inspect raw data in FL, poisoned samples can introduce biased decision boundaries or degrade performance. These attacks are particularly harmful in cross-device FL settings where the server lacks visibility into client behavior, and in cloud–edge deployments where edge nodes may aggregate updates from multiple clients [1], [3].

2) Model Poisoning Attacks: Rather than corrupting data, attackers deliberately alter their model updates before submitting them to the server. This manipulation may involve scaling gradients, injecting adversarial parameters, or steering the optimization trajectory. Manzoor et al. [1] emphasize that model poisoning can be even more effective than data poisoning because it bypasses local training and gives attackers direct control over the aggregated update.

3) Backdoor Attacks: Backdoor attacks aim to embed hidden triggers into the global model. The model performs normally on benign inputs but misbehaves when a specific trigger is present. The distributed trust model of FL makes such attacks particularly potent. A single compromised client can introduce a persistent backdoor, especially when naive averaging is used without anomaly detection [1].

4) Inference Attacks (Reconstruction and Membership Inference): Inference attacks attempt to recover sensitive client information from shared gradients or model parameters. Gradient inversion can reconstruct private samples, while membership inference tests whether specific records participated in training. According to [1], FL systems lacking robust privacy mechanisms—such as differential privacy or homomorphic encryption—are at risk of exposing confidential medical, financial, or personal data. These vulnerabilities also relate to emerging concerns around fairness and leakage trade-offs discussed in [10].

5) Sybil and Free-Rider Attacks: In Sybil attacks, a single adversary controls multiple fake clients to manipulate the aggregation process, overpowering honest users. This threat is particularly critical in cross-silo FL environments, where the number of clients is small [4]. Free-rider attacks occur when clients avoid local training but still receive the global model, benefiting without contributing meaningful updates [1].

6) Communication-Layer Attacks: Man-in-the-middle and eavesdropping attacks target insecure communication channels between clients, edge nodes, and the aggregation server. Although cloud infrastructures typically provide encrypted channels, misconfigured or resource-limited edge environments may still be vulnerable [3]. Replay attacks are another threat, where stale updates are injected to interfere with the convergence of the global model.

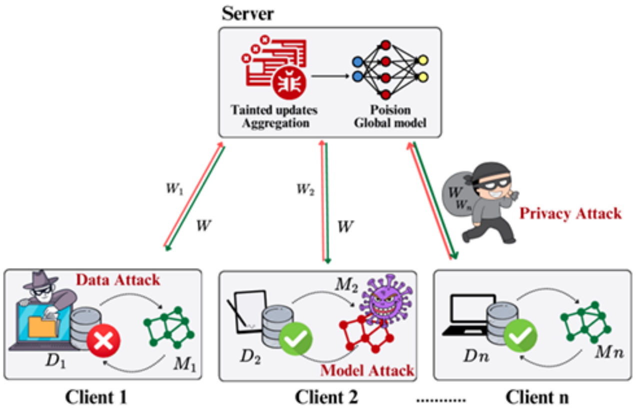

Fig. 3: Major attack types in federated learning.

Summary: These attack vectors show that while FL provides inherent privacy advantages, it is not automatically secure. Strong defenses—including secure aggregation, differential privacy, anomaly detection, blockchain-based traceability, and trusted execution environments—are necessary to maintain model integrity and client confidentiality [1], [5]. The detailed taxonomy in [1] is essential for designing resilient FL systems capable of scaling across diverse cloud–edge and cross-organizational settings.

E. Applications of Federated Learning in Cloud Environments

The combination of FL and cloud computing has accelerated adoption across diverse application domains where data sensitivity and decentralization are paramount [1], [3].

1) Healthcare: Healthcare data is often separated across hospitals, clinics, and diagnostic centers because of privacy rules. Cloud-based FL allows collaborative model training among these organizations without revealing patient data [1]. Examples include federated diagnostic imaging, electronic health record analysis, and medical text processing. For example, federated CNN models used over cloud infrastructure have shown high accuracy in detecting diseases and predicting personalized treatment [3], [10].

2) Internet of Things (IoT): In IoT systems, millions of edge devices generate large amounts of real-time data. FL combined with cloud computing creates a learning framework where edge nodes train local models and send updates to cloud servers for aggregation [3]. This setup supports smart city applications, autonomous vehicles, and predictive maintenance systems while significantly reducing the need to transmit raw data over the network. Blockchain-assisted FL approaches have also enhanced trust in IoT ecosystems [5].

3) Finance and Smart Contracts: Financial institutions can use FL to detect fraud patterns, score credit risk, or predict market trends while keeping customer data private. When deployed on secure cloud infrastructure, FL frameworks let regulated entities share model insights safely in environments that are friendly to audits [4], [10].

4) Edge–Cloud Collaborative Systems: Cloud-edge hybrid architectures are becoming more common in industrial automation and smart manufacturing. Here, federated models trained at the edge are combined in the cloud for overall improvement of production efficiency, fault detection, and logistics management [2], [3].

These applications underline the transformative role of cloud-enabled FL as a bridge between distributed intelligence and centralized orchestration [1], [3].

F. Key Observations

From the above discussion, it becomes evident that:

- Cloud computing provides the computational backbone and orchestration control essential for scaling federated systems [2], [3].
- Communication optimization and privacy-preserving mechanisms form the technical core of sustainable FL deployment [1], [10].
- Application domains such as healthcare, IoT, and finance demonstrate the real-world viability of cloud-based federated systems [1], [4], [5].

The convergence of these factors sets the stage for emerging paradigms like federated cloud services, cross-cloud collaboration, and federated multi-agent learning, which are discussed in later sections of this survey [4], [8].

IV. Overview of Federated Learning Algorithms and Emerging Methods

Federated Learning has evolved from simple distributed averaging methods to a varied range of algorithms [1]. These algorithms aim to tackle issues like differences in data, privacy concerns, and problems with communication in cloud-edge environments [3], [10]. This section gives a brief overview of popular federated learning algorithms and the new research trends affecting current FL projects [7], [8]. The goal is to help readers understand the key developments related to cloud-focused FL systems [2].

A. Foundational and Advanced FL Algorithms

The classical Federated Averaging (FedAvg) algorithm serves as the baseline for most FL implementations [1]. FedAvg allows selected clients to train a shared model locally using stochastic gradient descent (SGD). After training, clients send their updates to a central server, which then does weighted aggregation. Although FedAvg is efficient and straightforward, its performance drops significantly with non-i.i.d. data distributions, client dropout, and differing resource environments [10].

To overcome these issues, several variants of FedAvg have been proposed. FedProx adds a term that keeps local updates closer to the global model, which makes it more robust in environments with diverse devices. SCAFFOLD uses control variates to correct client drift, which helps stabilize training in highly non-i.i.d. settings. FedNova normalizes local updates to reduce biases from unequal computation among clients.

Hierarchical Federated Learning (HFL) expands aggregation to intermediate edge servers, which greatly decreases communication costs in large IoT deployments [2], [3]. Personalized Federated Learning (pFL) methods like pFedMe, FedPer, and FedBN improve client-specific adaptability. This leads to better performance in diverse and multi-domain environments [7], [8]. These methods are especially effective for cross-silo FL deployments in healthcare, finance, and enterprise cloud ecosystems [1], [4].

TABLE II: Comparison of Key Federated Learning Algorithms

Algorithm	**Core Idea**	**Advantages**	**Limitations**
FedAvg	Local SGD + global averaging	Simple, scalable, low communication cost	Poor performance under non-IID data; unstable convergence
FedProx	Adds proximal term to local training	Handles device heterogeneity; stabilizes aggregation	Higher computation; slower convergence
SCAFFOLD	Control variates correct client drift	Strong performance under non-IID data	Increased communication and storage overhead
FedNova	Normalized local update contributions	Fairness under heterogeneous computation	Sensitive to hyperparameter settings
Hierarchical FL	Multi-tier aggregation (edge + cloud)	Reduced uplink traffic; scalable for IoT	Requires reliable coordination; higher system complexity
Personalized FL	Client-specific objectives or model layers	High adaptability; improved per-client accuracy	Reduced global consistency; harder deployment
Split Learning	Model split between client and server	Low client compute; better privacy	High latency; frequent communication rounds
Robust FL (Krum, Bulyan)	Filters/weights updates to resist attacks	Protection from poisoning and malicious clients	High computation; may reduce accuracy

B. Emerging Trends in Federated Learning

Recent developments in federated learning have picked up speed due to challenges from large-scale cloud deployment, strict privacy laws, and complex IoT systems [2], [3]. Several new trends are changing the algorithmic and architectural foundations of modern FL frameworks.

Communication-efficient FL has become essential because of the high uplink traffic during iterative training [2]. Techniques like gradient sparsification, quantization, periodic aggregation, and asynchronous updates aim to reduce communication costs while maintaining accuracy. These methods are especially valuable in wireless and edge environments with limited bandwidth [3].

Privacy-preserving FL has gained attention with the growing use of differential privacy, secure multiparty computation, homomorphic encryption, and blockchain-based trust systems [1], [5], [9]. These techniques help manage privacy and utility trade-offs while meeting compliance needs in cross-silo FL deployments [10].

Continual and lifelong FL focuses on allowing models to evolve as new tasks and data distributions appear. This makes it an important approach for IoT, autonomous systems, and industrial settings [7]. Personalized FL is also increasing due to the demand for client-specific customization in collaborations between different institutions [4], [10].

Federated multi-agent learning is another emerging trend, combining reinforcement learning with FL ideas to support distributed decision-making in smart grids, transportation networks, and cooperative robotics [8].

TABLE III: Emerging Trends in Federated Learning

Trend	Drivers	Advantages	Challenges
Communication-Efficient FL	Bandwidth limits; 5G/6G; massive IoT	Faster training; reduced communication cost	Possible accuracy loss from aggressive compression
Privacy-Preserving FL	Data protection laws; cross-silo collaborations	Strong confidentiality; secure aggregation	Privacy–utility trade-offs; high computational cost
Cross-Silo FL	Healthcare, finance, enterprise systems	High accuracy; structured participation	Requires strong governance and trust mechanisms
Continual FL	Dynamic IoT and Industry 4.0 streams	Adaptive; memory-efficient	Risk of catastrophic forgetting
Blockchain-Assisted FL	Need for auditability and trust	Tamper-proof logs; strong integrity	High energy cost; scalability limits
Federated Multi-Agent Learning	Mobility, logistics, energy systems	Distributed intelligence; cooperation	High system complexity

V. Research Trends and Open Challenges in Cloud-Based Federated Learning

Despite the increasing development of cloud-based Federated Learning (FL) systems, several unresolved issues still hinder their large-scale use. These challenges come from the interaction between distributed learning dynamics, varied client environments, and the complicated management needed in cloud infrastructures [3]. This section looks at the key research areas and open challenges that shape the current state of cloud-integrated FL.

A. Communication Overhead and System Scalability

One of the main challenges in FL is the high communication cost of iterative model updates between clients and cloud servers [10]. As more devices join, the total bandwidth and synchronization demands increase. While model compression, gradient sparsification, and asynchronous updates have been suggested, these methods often lead to a drop in accuracy or instability during training, especially with non-IID data settings [1]. Additionally, network diversity, where clients have different bandwidth capabilities, makes synchronization harder.

Current research is looking into hierarchical or multi-tiered FL architectures. In these setups, edge nodes do initial aggregation before sending data to cloud coordinators [2], [3]. This approach lowers communication demands and efficiently supports millions of clients. Researchers are also exploring how to combine Software Defined Networking (SDN) and Network Function Virtualization (NFV) with FL management in the cloud. This integration could help allocate network resources dynamically, depending on client activity [2].

B. Data and Model Heterogeneity

In real-world cloud deployments, clients often have different data distributions and model capacities, which creates non-IID challenges. This variety results in biased global models that perform poorly for minority client groups [10]. To address this, researchers are investigating personalized federated learning and meta-learning strategies. These methods enable each client to keep a partially personalized model while also benefiting from shared global knowledge [7].

From the cloud perspective, this variety requires flexible resource management, load balancing, and scheduling algorithms to ensure fair participation among clients with different computing abilities [3]. Researchers are also looking into integrating serverless computing into FL pipelines. This would allow for model training in the cloud as needed without requiring ongoing infrastructure setup [2].

C. Privacy–Utility Trade-off

The main goal of FL is to protect privacy. This focus creates a trade-off between keeping data confidential and maintaining model accuracy [1], [10]. Techniques like differential privacy (DP), secure multiparty computation (SMC), and homomorphic encryption (HE) provide solid theoretical safeguards, but they also raise computation and communication costs [1], [9].

Current research aims to improve this balance with adaptive privacy budgets, client-level noise scheduling, and federated trust management frameworks that measure privacy leakage risk in real time [10]. Another emerging trend is federated auditing. Here, the cloud serves as a trusted verifier to check compliance with privacy policies like GDPR or HIPAA without accessing raw data [3]. Blockchain-based methods are also becoming popular. They use smart contracts to handle model contribution records and reward systems, promoting accountability and transparency in collaborations between institutions [5].

D. Security Threats and Robustness

FL systems are still at risk from poisoning and inference attacks. In these attacks, bad actors insert altered gradients to damage the global model or steal sensitive information from other users [1]. In a cloud environment, these risks increase because of the shared nature of virtual infrastructure. To tackle these issues, strong aggregation algorithms like Krum, Trimmed Mean, and Bulyan have been suggested. These algorithms filter out unusual updates during aggregation [10].

Additionally, zero-trust architectures and confidential computing technologies, such as Intel SGX and AMD SEV, are being used in cloud platforms. These technologies help to separate training environments and stop unauthorized access to data during model aggregation [3].

E. Interoperability and Cross-Cloud Federation

As organizations increasingly use multi-cloud strategies, making different cloud systems work together has become a significant challenge. Differences in APIs, data formats, and security standards limit smooth collaboration between cloud environments [4].

Research is shifting towards standardized FL orchestration frameworks and federated APIs. This will allow for interoperability across AWS, Azure, GCP, and private cloud infrastructures [3]. The idea of Federated Cloud Learning (FCL) is emerging. In this model, multiple cloud environments act as high-level nodes in a global federation. This setup enables collaborations between enterprises while keeping control at the cloud level [4].

F. Emerging Research Trends

Recent developments show a strong movement toward the fusion of FL with advanced cloud technologies [2], [3]:

- **Hierarchical and Split Learning:** Combining local FL with cloud-level split architectures for better memory and latency efficiency [3].
- **Edge–Cloud Co-Federation:** Seamless migration of training between edge devices and cloud servers based on resource availability [2].
- **Federated Transfer Learning (FTL):** Leveraging shared representations across clouds and clients with dissimilar data features [4].
- **AI Governance in Federated Systems:** Research on policy frameworks and ethical standards ensuring fairness, transparency, and responsible AI deployment in federated cloud environments [10].

Together, these trends reflect the evolution of FL from a distributed learning technique into a complete cloud-native ecosystem, enabling collaborative AI development under stringent data governance requirements [1].

TABLE IV: Summary of Challenges in Cloud-based Federated Learning

Challenge Area	Description	Current Research Focus
Communication Overhead	High latency and bandwidth usage in global updates	Compression, clustering, asynchronous updates
Data Heterogeneity	Non-IID data causing biased global models	Personalized and meta-FL approaches
Privacy–Utility Trade-off	Balancing data privacy with accuracy	Adaptive DP and secure aggregation
Security Risks	Malicious clients and poisoning attacks	Robust aggregation, blockchain audit
Scalability	Supporting millions of clients on clouds	Hierarchical and federated orchestration
Interoperability	Multi-cloud collaboration challenges	Federated APIs, standard orchestration layers

G. Discussion

The combination of Federated Learning (FL) and cloud computing presents both an opportunity and a challenge. Cloud infrastructure allows for the large-scale deployment and management of FL systems [2], [3]. However, it also brings complexities related to security, governance, and interoperability [1], [10]. The key to future progress is creating flexible federated frameworks that consider network conditions and privacy budgets [10]. This ensures efficient communication, secure aggregation, and fair model performance among all participants [1].

In summary, integrating Federated Learning with cloud computing has seen significant success but still has room for improvement. Addressing challenges like communication scalability, the balance between privacy and utility, and cross-cloud interoperability will shape the next phase of FL adoption [4]. New approaches such as federated orchestration, blockchain-based trust mechanisms, and policy-driven AI governance are likely to change how intelligent systems are collaboratively trained in the cloud era [2], [5], [10].

VI. Conclusion and Future Directions

Federated Learning (FL) marks a significant change in how intelligent systems are trained and used in the age of cloud computing. By decentralizing the learning process and allowing for data control, FL reduces privacy risks and supports large-scale collaboration among different organizations [1]. Cloud computing offers the necessary computational support, flexible scalability, and management infrastructure that help federated systems run efficiently at a large scale [2], [3].

In this survey, we have examined the basics of FL, its connection with cloud architectures, and the main technological factors that enable this collaboration. We discussed three key aspects: communication efficiency, privacy protection, and application development [10]. These are essential for creating lasting federated systems in the cloud. New trends like hierarchical aggregation, model compression, and blockchain auditing show how researchers are working to balance accuracy, speed, and security in real-world applications [2], [5].

Despite these advancements, several challenges have not yet been resolved. Issues such as communication delays, differing data and system types, and the balance between privacy and utility still pose problems for large-scale FL deployment [10]. Additionally, ensuring protection against harmful behaviors, achieving compatibility across different cloud platforms, and creating energy-efficient management are all areas needing more focus [4]. The combination of federated intelligence with technologies like edge computing, 5G/6G networks, and serverless cloud systems will likely shape the next generation of distributed AI ecosystems [2], [3].

Future research will likely concentrate on the following key areas:

- **Cross-Cloud Federation and Interoperability:** Developing standardized APIs and orchestration layers allows secure collaboration between different cloud providers without compromising compliance or performance [4].
- **Adaptive Privacy Preservation:** Designing dynamic privacy mechanisms can automatically balance data protection with model utility based on the context and sensitivity of the application [10].
- **Energy-Efficient Federated Learning:** Using cloud-native sustainability tools and green computing frameworks helps minimize carbon footprints during large-scale FL operations [2].
- **Autonomous Federated Orchestration:** Employing reinforcement learning and AI-driven scheduling can automate client selection, resource allocation, and model aggregation in complex cloud ecosystems [8].
- **Explainable and Responsible Federated AI:** Ensuring transparency, accountability, and fairness in federated model decisions helps build trust among collaborating organizations and users [10].

In conclusion, the fusion of Federated Learning and Cloud Computing is shaping a new era of distributed intelligence. This approach values data privacy, scalability, and global cooperation. As research continues to develop, this integration will play a crucial role in creating secure, intelligent, and fair AI systems capable of addressing large-scale societal challenges in healthcare, finance, IoT, and beyond [1], [3].

References

[1] H. U. Manzoor, A. Shabbir, A. Chen, D. Flynn, and A. Zoha, "A Survey of Security Strategies in Federated Learning: Defending Models, Data, and Privacy," *Future Internet*, vol. 16, no. 10, p. 374, 2024. :contentReference[oaicite:0]index=0

[2] A. Agiollo, P. Bellavista, M. Mendula, and A. Omicini, "EneA-FL: Energy-aware Orchestration for Serverless Federated Learning," *Future Generation Computer Systems*, vol. 154, pp. 219–234, 2024. :contentReference[oaicite:1]index=1

[3] G. Bao and P. Guo, "Federated Learning in Cloud–Edge Collaborative Architecture: Key Technologies, Applications and Challenges," *Journal of Cloud Computing: Advances, Systems and Applications*, vol. 11, no. 94, 2022.

[4] S. Kalloori and A. Srivastava, "Towards Cross-Silo Federated Learning for Corporate Organizations," *Knowledge-Based Systems*, vol. 289, p. 111501, 2024.

[5] S. Wankhede and N. Patel, "Federated Learning and Blockchain Approach for Securing IoT Data," *Computer Networks*, vol. 245, p. 110927, 2025.

[6] J. Zhang, X. Wei, and L. Chen, "Distributed Machine Learning for Next-Generation Communication Networks: Privacy, Fairness and Efficiency, and trade-offs" *Information Sciences*, vol. 671, p. 120310, 2025.

[7] H. Birashk and S. Khan, "Federated Continual Learning for Task-Incremental and Class-Incremental Problems," *Engineering Applications of Artificial Intelligence*, vol. 136, p. 109199, 2025.

[8] Y. Jing, K. Zhao, and F. Wu, "Federated Multi-Agent Reinforcement Learning:A comprehensive survey of methods, applications and challenges " *Applied Soft Computing*, vol. 154, p. 111278, 2025.

[9] Y. Meng and L. Zhang, "A Survey on Secure Multi-Party Computation Techniques Based on Private Set Intersection," *Computer Standards & Interfaces*, vol. 89, p. 103767, 2026.

[10] L. Wang, H. Li, and Y. Liu, "Linkage on security, privacy and fairness in federated learning: New balances and new perspectives" *Expert Systems with Applications*, vol. 243, p. 123174, 2025.