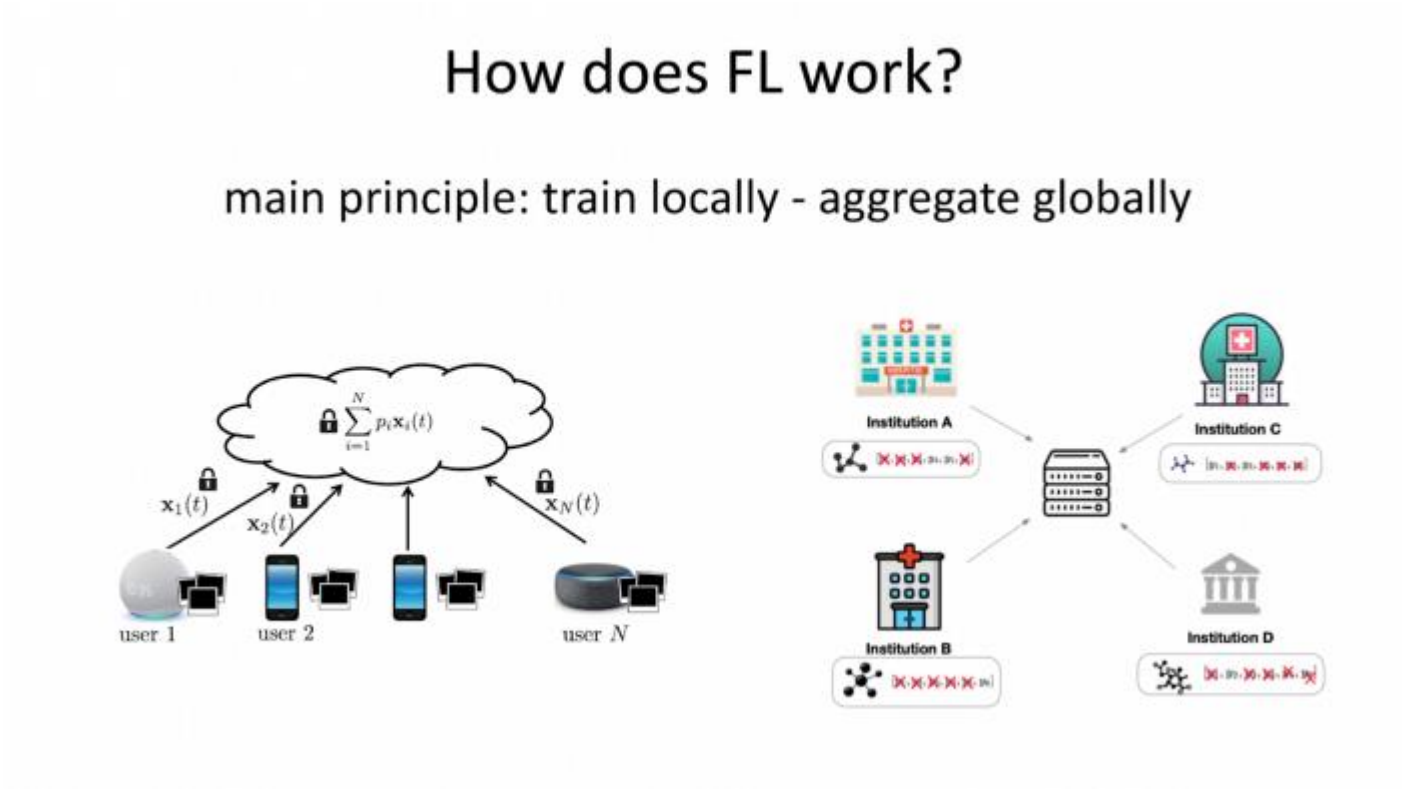
FedML

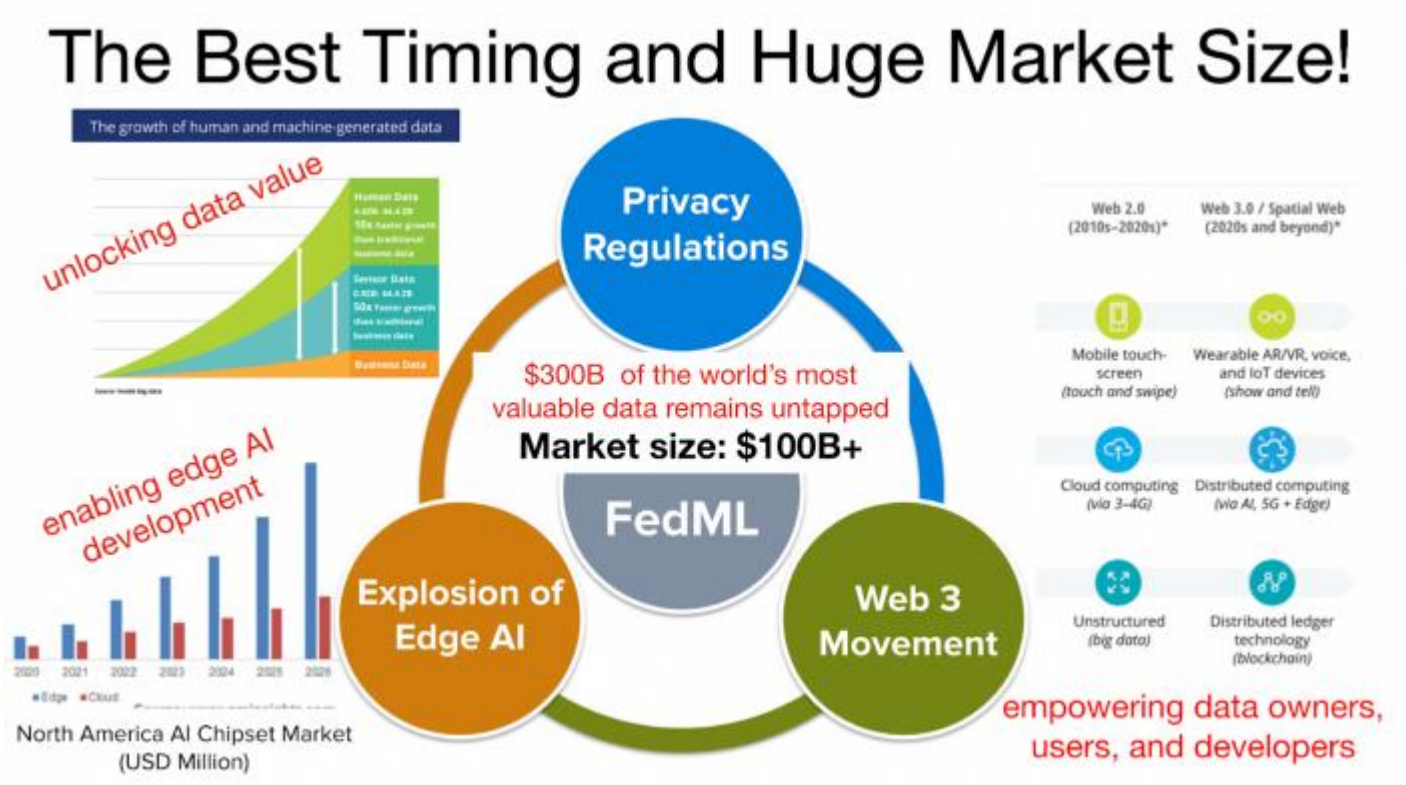
FedML Product Overview, https://medium.com/@FedML/fedml-ai-platform-releases-the-worlds-federated-learning-open-platform-on-public-cloud-with-an-8024e68a70b6



FedML

FedML Product Overview, https://medium.com/@FedML/fedml-ai-platform-releases-the-worlds-federated-learning-open-platform-on-public-cloud-with-an-8024e68a70b6

In this era, federated learning technology is critical because it is at the historical intersection of the three major technological hotspots: privacy computing, edge AI, and Web3/Blockchain.



FedML

FedML Product Overview, https://medium.com/@FedML/fedml-ai-platform-releases-the-worlds-federated-learning-open-platform-on-public-cloud-with-an-8024e68a70b6

What never before have we considered is distributed AI algorithms on a large number of nodes with scattered small data while simultaneously taking into account security, system efficiency, and model accuracy.
In addition, many concepts in blockchain, including data decentralization computing, ownership verification, traceability, incentive mechanism, and trusted security, all coincide with the idea of federated learning to make the data value flow safely.

FedML

FedML Product Overview, https://medium.com/@FedML/fedml-ai-platform-releases-the-worlds-federated-learning-open-platform-on-public-cloud-with-an-8024e68a70b6

In academia, the most popular open-source framework is undoubtedly FedML (https://github.com/FedML-AI), which is widely used around the world (see https://fedml.ai/use-cases/). (there are also some other frameworks, but FedML is the one cited most: FedML-196; Flower-96; PySyft-32; FedScale-20; FATE-14).

Today, the FedML team has further upgraded these academic achievements into an industrialized platform. Its mission is to build open and collaborative AI anywhere at any scale. In other words, FedML supports both federated learning for data silos and distributed training for acceleration with MLOps and Open Source support, covering cutting-edge academia research and industrial grade use cases.

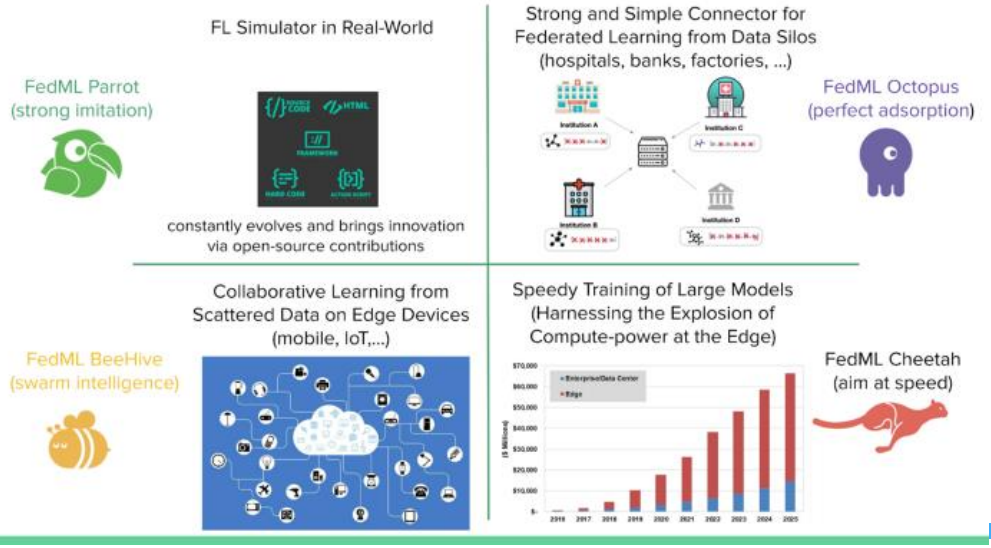
Different from the closed-source commercial platforms of most companies or the restricted mode of "applying for a trial," FedML focuses on building a public MLOps (ML Operations) platform, which is open to global users for free.

FedML

FedML Product Overview, https://medium.com/@FedML/fedml-ai-platform-releases-the-worlds-federated-learning-open-platform-on-public-cloud-with-an-8024e68a70b6

At its current stages, FedML provides the following services:

- FedML Parrot — Simulating federated learning in the real world.
- FedML Octopus — Cross-silo Federated Learning for cross-organization/account training, including Python-based edge SDK.
- FedML Beehive — Cross-device Federated Learning for Smartphones and IoTs, including edge SDK for Android/iOS and embedded Linux.
- FedML MLOps: FedML's machine learning operation pipeline for AI running anywhere at any scale.
- Model Serving: providing a better user experience for edge AI.



FedML

FedML Product Overview, https://medium.com/@FedML/fedml-ai-platform-releases-the-worlds-federated-learning-open-platform-on-public-cloud-with-an-8024e68a70b6

FEDML MLOPS USER GUIDE, https://open.fedml.ai/octopus/userGuides/index

Write Once, Run Anywhere: Seamlessly Migrate Your Local Development to the Real-world Edge-cloud Deployment

Video Tutorial: https://www.youtube.com/embed/Xgm0XEaMlVQ

FEDERATED LEARNING ON ANDROID SMARTPHONES,
https://open.fedml.ai/beehive/userGuides/index

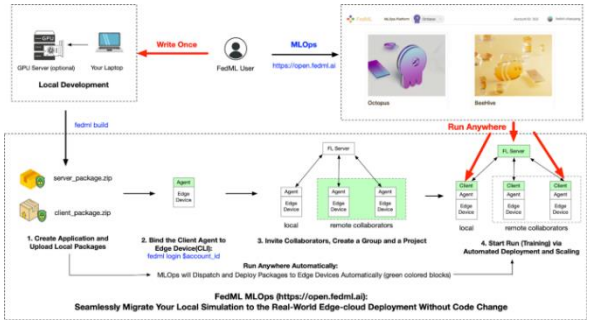
FedML

FedML Product Overview, https://medium.com/@FedML/fedml-ai-platform-releases-the-worlds-federated-learning-open-platform-on-public-cloud-with-an-8024e68a70b6

3.1 Seamless migration between simulated experiment and real deployment, zero code modification

FedML can help users seamlessly migrate the code of experimental simulation (POC) to the actual system (Production) to carry out experiments under real private data and distributed training systems of edge devices.

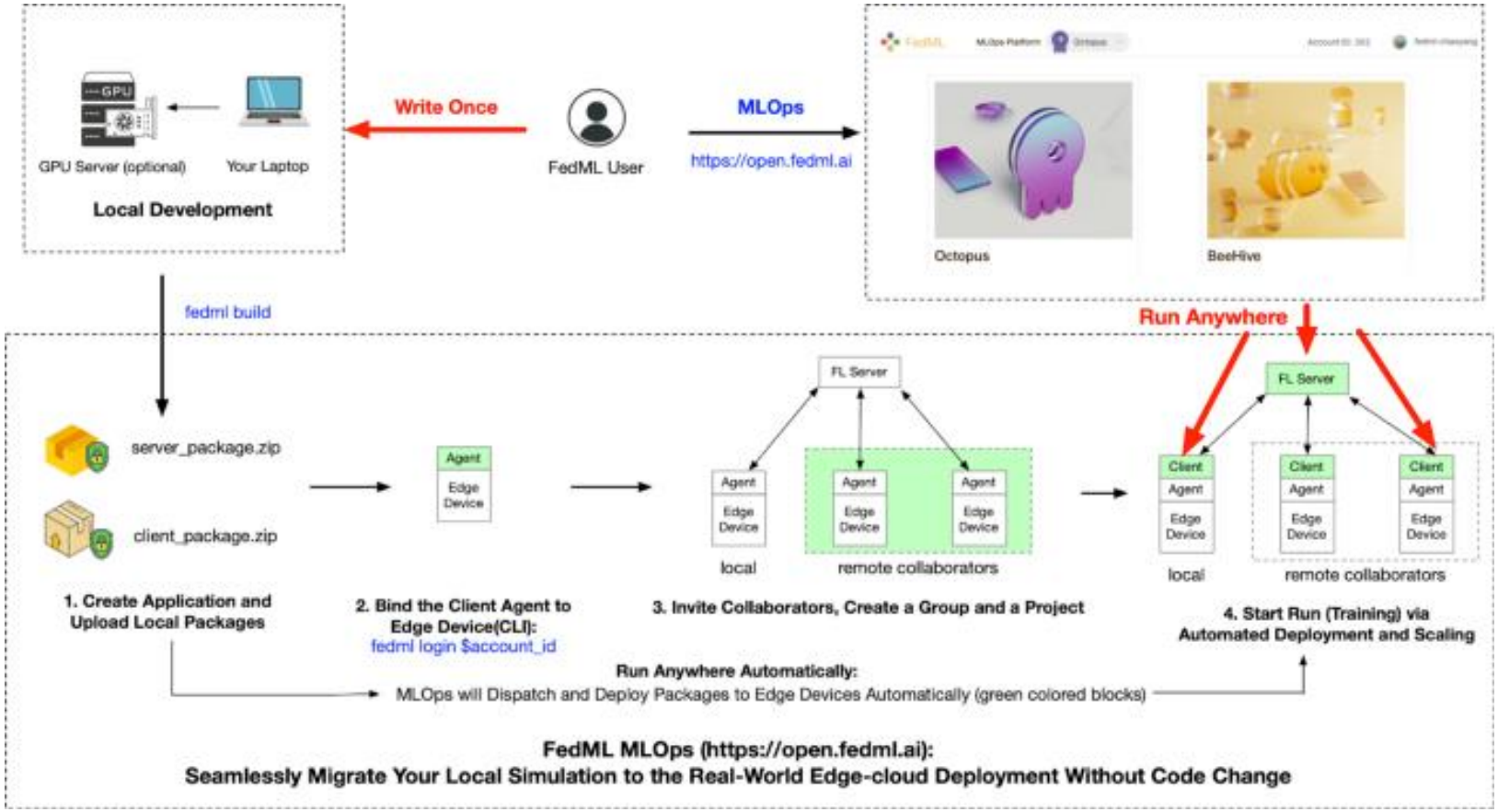
- After it is verified that federated learning can produce modeling benefits on specific applications, users can use FedML MLOps to upgrade the simulation into production without modifying the code.
- The simulated source code can be deployed directly to edge devices with real data. As shown in the upper left of Figure 1, the user first completes local development and debugging in "Local Development" and then generates installation packages.
- Finally, through simple commands and UI interactions, these installation packages and scripts can be distributed to any private device (shown in red on the right of Figure 1).



FedML

FedML Product Overview, https://medium.com/@FedML/fedml-ai-platform-releases-the-worlds-federated-learning-open-platform-on-public-cloud-with-an-8024e68a70b6

3.1 Seamless migration between simulated experiment and real deployment, zero code modification



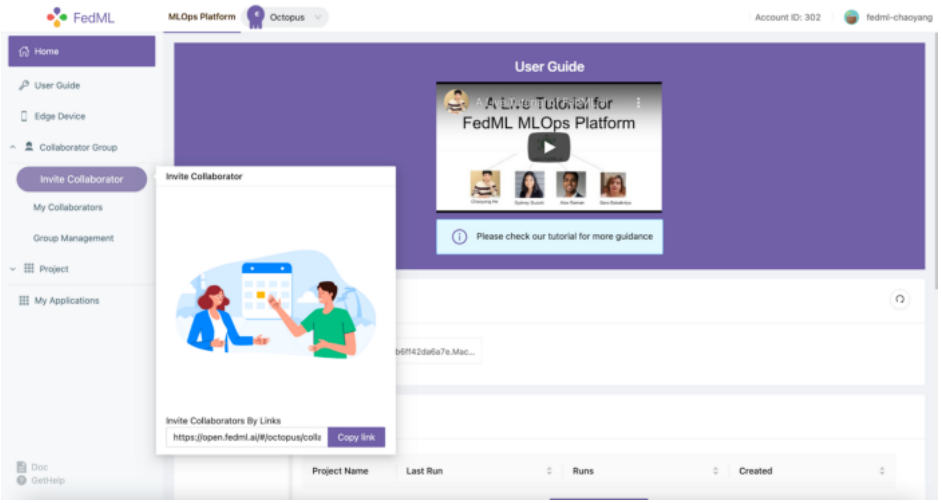
Federated Learning

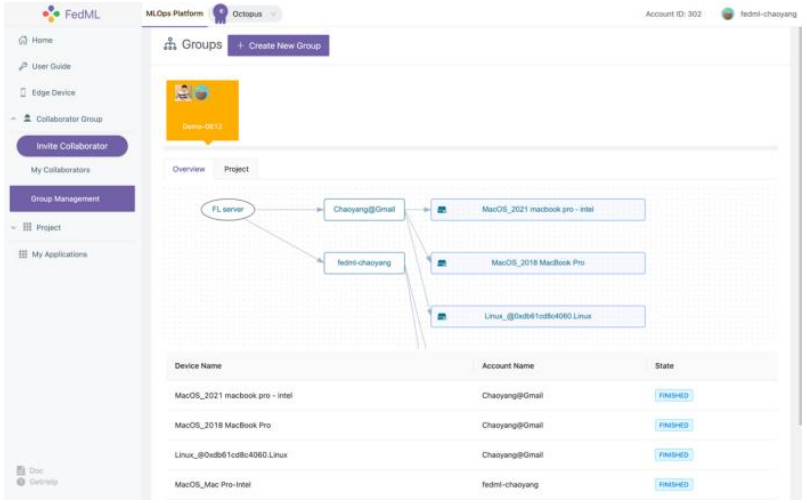
FedML Product Overview, https://medium.com/@FedML/fedml-ai-platform-releases-the-worlds-federated-learning-open-platform-on-public-cloud-with-an-8024e68a70b6

3.2 One-line command to complete the edge deployment

3.3 Support simplified collaboration anywhere: multinational, cross-city, multi-tenant

FedML collaboration has become extremely simple. Just as you can collaborate on documents with friends, you can easily create a federated learning group by sending an invitation link. There are no geographical, national, or city restrictions anywhere in the world.



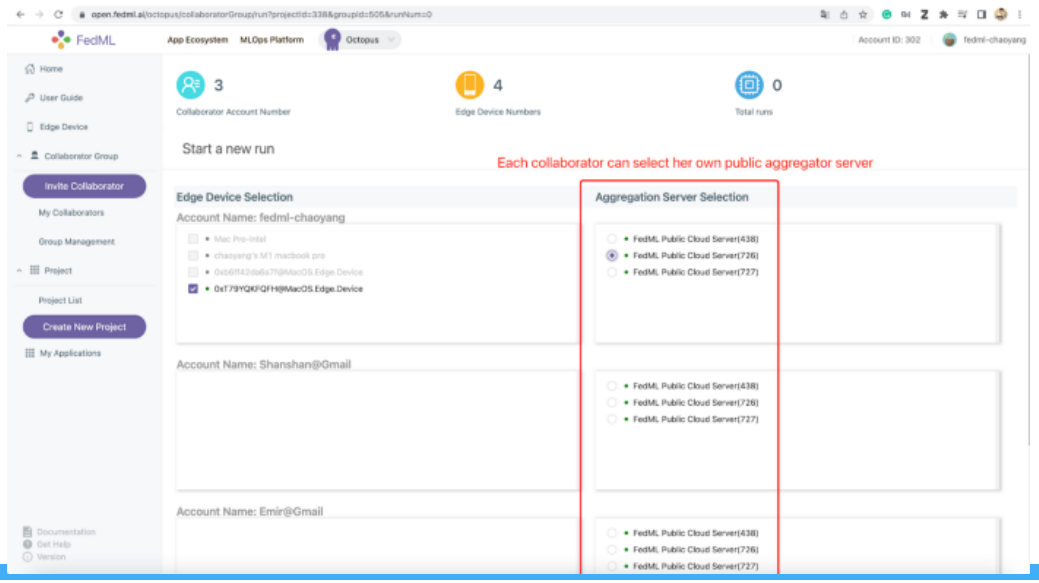


FedML

FedML Product Overview, https://medium.com/@FedML/fedml-ai-platform-releases-the-worlds-federated-learning-open-platform-on-public-cloud-with-an-8024e68a70b6

3.4 Provide free public cloud aggregation server and private cloud deployment with Docker

To reduce the difficulty of federated training, FedML's open platform provides a public cloud aggregation server for everyone. Users can arbitrarily select free service nodes offered on the public cloud when initiating training.
FedML also considers more secure and strict deployment requirements for platform users. For this reason, FedML platform has also developed a private aggregation server library that can be deployed freely. It still only needs a one-line command, that is, the "fedml login -s" command to run a secure docker environment at any self-hosted server.



FedML

FedML Product Overview, https://medium.com/@FedML/fedml-ai-platform-releases-the-worlds-federated-learning-open-platform-on-public-cloud-with-an-8024e68a70b6

3.5 Experiment monitoring and analysis capabilities tailored for distributed training

In addition to the features mentioned above of lowering the user access threshold, the FedML platform also helps AI application modeling by providing experimental tracking, management, visualization, and analysis capabilities. Key capabilities currently supported include:

1. Edge device training status tracking.

2. Custom metrics reporting, such as the accuracy in common classification tasks in deep learning, the error rate of regression tasks, and even the running time of the system state, memory usage, and GPU utilization.

3. Profiling flow and edge device system performance. It can help users to view the execution performance of different subtasks on each edge device, which is convenient for analyzing the bottleneck of the system.

4. Distributed logging. This is a capability that the current general-purpose machine learning platform does not have, and it is convenient to track and analyze the anomalies that occur on each device in a real-time manner.

5. The experimental report allows users to compare multiple experimental results.

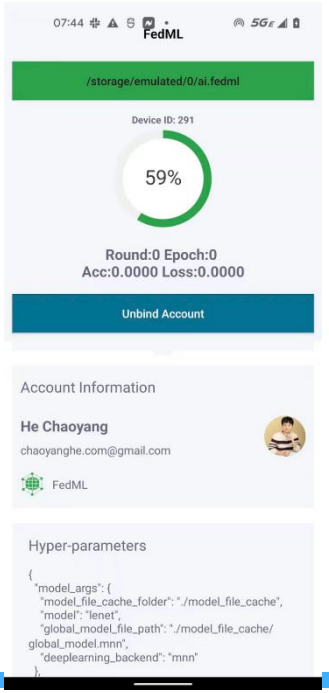
FedML

FedML Product Overview, https://medium.com/@FedML/fedml-ai-platform-releases-the-worlds-federated-learning-open-platform-on-public-cloud-with-an-8024e68a70b6

3.6 Unified cross-platform design, supporting smartphones and IoT devices

FedML also recently released the Android platform for mobile devices; details can be found at the following links:

FedML Android Platform: https://github.com/FedML-AI/FedML/tree/master/android
FedML IoT Platform: https://github.com/FedML-AI/FedML/tree/master/iot



FedML

FedML Product Overview, https://medium.com/@FedML/fedml-ai-platform-releases-the-worlds-federated-learning-open-platform-on-public-cloud-with-an-8024e68a70b6

4. Three-in-one Strategy for Open Collaboration: Open Source, Open Platform, Collaborative Application Ecosystem

In addition to operating open source communities, FedML promotes open collaboration and open source research and development from multiple product perspectives.

- Besides the open source library (https://github.com/FedML-AI),
- and open platform (https://open.fedml.ai),

FedML has also developed the collaborative App Ecosystem (users can visit https://open.fedml.ai and find "App Ecosystem" on the left top).

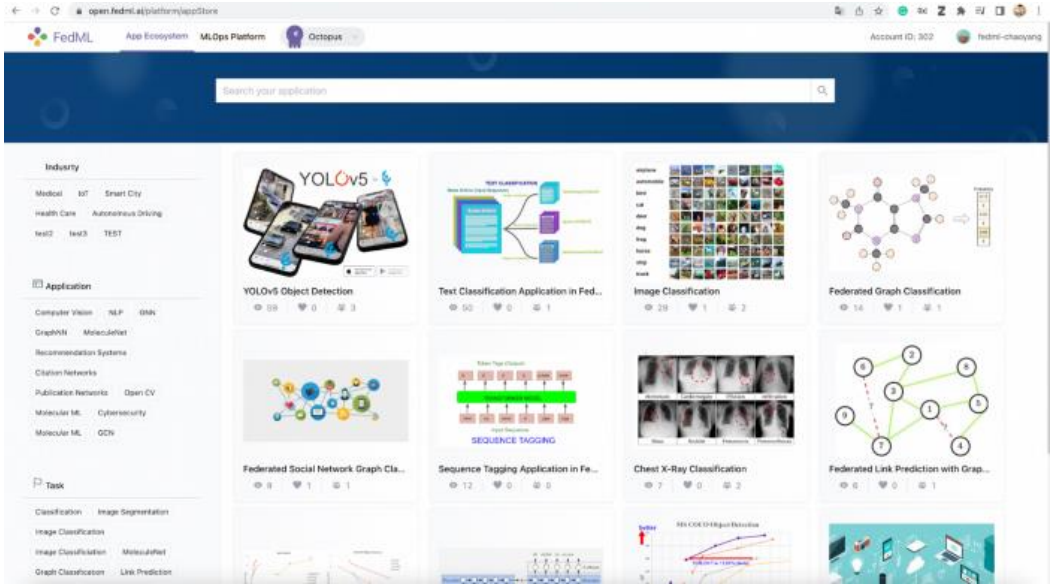
- The App Ecosystem and the platform cooperate with each other to continuously enrich the application ecosystem. The first version has completed the open collaboration of over 20 applications.
- Users can contribute and share the application. Each application includes all the FedML-based source code of an AI application, including model definitions, training scripts, and configuration files.
- At present, the App Ecosystem covers mainstream AI application scenarios such as computer vision, natural language processing, graph data mining, and the Internet of Things.

FedML

FedML Product Overview, https://medium.com/@FedML/fedml-ai-platform-releases-the-worlds-federated-learning-open-platform-on-public-cloud-with-an-8024e68a70b6

4. Three-in-one Strategy for Open Collaboration: Open Source, Open Platform, Collaborative Application Ecosystem

If the open platform reduces the difficulty of actual building deployment of the federated learning system to the lowest level, then the App Ecosystem is used to lower the AI application R&D threshold for practitioners: A company needs not to hire high-cost machine learning teams but rather needs only one engineer who can do "one-click import" on the basis of community results and use the application directly without intensive development circles.



FedML

FedML Product Overview, https://medium.com/@FedML/fedml-ai-platform-releases-the-worlds-federated-learning-open-platform-on-public-cloud-with-an-8024e68a70b6

5. Simple and flexible APIs, boosting innovation in algorithm and system optimization

- First, from the application point of view, FedML does its best to shield all code details and complex configurations of distributed training. Data scientists and engineers at the application level, such as computer vision, natural language processing, and data mining, only need to write the model, data, and trainer in the same way as a stand-alone program and then pass it to the FedMLRunner object to complete all the processes. This greatly reduces the bar for application developers to perform federated learning.
- Secondly, the FedML team believes that the design of the API should conform to the current technology development trend and should not assume that today's technology is the final solution; rather, it should be iterated as it progresses. We can see that the algorithm innovation of the open source community is still very active, and many more user-valued algorithms continue to be innovated every month. It is based on this background that FedML considers making custom APIs flexible enough to empower algorithm innovation. To this end, FedML abstracts the core trainer and aggregator and provides users with two abstract objects, FedML.core.ClientTrainer and FedML.core.ServerAggregator, which only need to inherit the interfaces of these two abstract objects and pass them to FedMLRunner. Such customization provides machine learning developers with maximum flexibility. Users can define arbitrary model structures, optimizers, loss functions, etc. These customizations can also be seamlessly connected with the open source community, open platform, and application ecology mentioned above with the help of FedMLRunner, which completely solves the long lag problem from innovative algorithms to commercialization.

FedML

FedML Product Overview, https://medium.com/@FedML/fedml-ai-platform-releases-the-worlds-federated-learning-open-platform-on-public-cloud-with-an-8024e68a70b6

5. Simple and flexible APIs, boosting innovation in algorithm and system optimization

- Finally, FedML believes that although FL is a comprehensive technology that combines security, system efficiency, and model accuracy, the first priority is still "ML-oriented Research and Development". For example, security and system optimization are definitely important, but it is not a good product design for ML users if they have a huge learning burden on security and system design — this would eventually cause the core users to abandon the product. Therefore, in terms of architecture, FedML considers that security, privacy, and system optimization should all serve ML. The details of these auxiliary modules are hidden throughout layered design, and it is ultimately through this that the best ML experience is achieved. This responsibility is accomplished through FedML Flow.

FedML

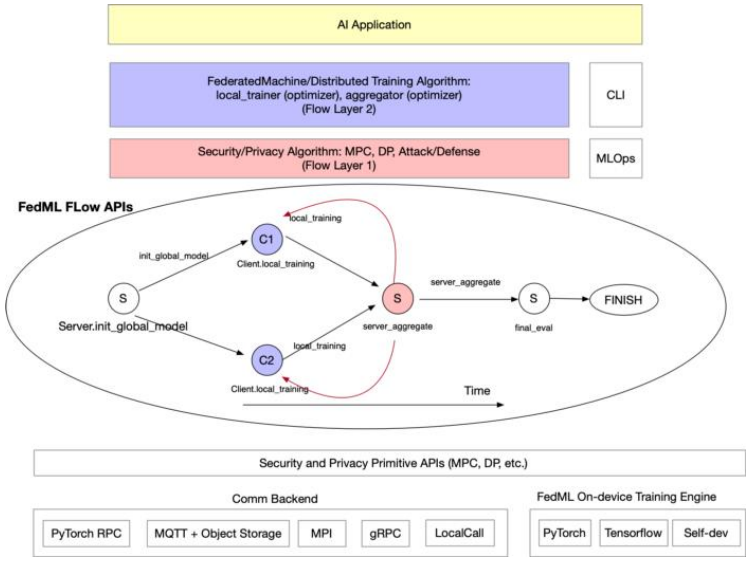
FedML Product Overview, https://medium.com/@FedML/fedml-ai-platform-releases-the-worlds-federated-learning-open-platform-on-public-cloud-with-an-8024e68a70b6

5. Simple and flexible APIs, boosting innovation in algorithm and system optimization

Specifically, as shown in the figure, FedML regards distributed computing processes such as complex security protocols and distributed training as a directed acyclic graph (DAG) flow computing process, making the writing of complex protocols similar to stand-alone programs.

- Based on this idea, the security protocol Flow Layer 1
- and the machine learning algorithm process Flow Layer 2

can be easily separated so that security engineers and machine learning engineers can perform their duties without having to master multiple technologies with the same mindset.



Federated Learning

FedML Product Overview, https://medium.com/@FedML/fedml-ai-platform-releases-the-worlds-federated-learning-open-platform-on-public-cloud-with-an-8024e68a70b6

5. Simple and flexible APIs, boosting innovation in algorithm and system optimization

For a more intuitive understanding of FedML Flow, the following example demonstrates the process of implementing the FedAvg algorithm and adding multiple distributed tasks through FedML Flow.

- First, each distributed node can be regarded as an abstract FedMLExecutor, which is carried in an independent process and is responsible for executing a specific task. This task can be training or some protocol messages, thus maintaining a high degree of flexibility and abstraction.

- Flow is a framework that helps to transfer the behavior of these distributed Executors, and it can arrange the order of task execution and message passing between tasks. Specific to a FedAvg algorithm, we can define a Client Executor and a Server Executor object and use their custom functions as tasks in the flow.

- Through the Flow API, users can freely combine the execution processes of these Executors. The following code shows the entire process of model initialization, multiple rounds of training, and finally, distributed evaluation of the model.

- The most important thing is that is that this programming example only happens on the personal computer of FedML users and does not require any distributed system development skills

FedML

FedML Product Overview, https://medium.com/@FedML/fedml-ai-platform-releases-the-worlds-federated-learning-open-platform-on-public-cloud-with-an-8024e68a70b6

```
    if args.rank == 0:
        executor = Server(args)
        executor.init(device, dataset, model)
    else:
        executor = Client(args)
        executor.init(device, dataset, model)

    fedml_alg_flow = FedMLAlgorithmFlow(args, executor)
    fedml_alg_flow.add_flow("init_global_model", Server.init_global_model)
    fedml_alg_flow.add_flow("handle_init", Client.handle_init_global_model)
    for round_idx in range(args.comm_round):
        fedml_alg_flow.add_flow("local_training", Client.local_training)
        fedml_alg_flow.add_flow("server_aggregate", Server.server_aggregate)
    fedml_alg_flow.add_flow("final_eval", Server.final_eval)
    fedml_alg_flow.build()

    fedml_runner = FedMLRunner(args, device, dataset, model, algorithm_flow=fedml_alg_flow)
    fedml_runner.run()

class Client(FedMLExecutor):
    def local_training(self):
    def handle_init_global_model(self):

class Server(FedMLExecutor):
    def init_global_model(self):
    def server_aggregate(self):
    def final_eval(self):
```

FedML

FedML Product Overview, https://medium.com/@FedML/fedml-ai-platform-releases-the-worlds-federated-learning-open-platform-on-public-cloud-with-an-8024e68a70b6

6. Release Cross-silo FL dataset with Okwin, allowing research face real scenarios

Different from the synthetic or hypothetical data combed by papers such as LEAF and FedScale (note: only natural ID segmentation does not represent the authenticity of products and business scenarios), the datasets released by FedML and Okwin are taken from real federated learning scenarios.

As shown in the figure below, the current version mainly focuses on medical scenarios, containing 7 naturally partitioned medical datasets covering multiple tasks, models, and data modalities, each with baseline training codes for everyone to conduct research and development. All these results are published on the FedML open platform; see https://open.fedml.ai for details (click "App Ecosystem" after logging in).

Dataset	Fed-Camelyon16	Fed-LIDC-IDRI	Fed-IXI	Fed-TCGA-BRCA	Fed-KITS2019	Fed-ISIC2019	Fed-Heart-Disease
Input (x)	Slides	CT-scans	T1WI	Patient info.	CT-scans	Dermoscopy	Patient info.
Preprocessing	Matter extraction + tiling	Patch Sampling	Registration	None	Patch Sampling	Various image transform	Removing missing data
Task type	binary classificatior	3D segmentation	3D segmentation	survival	3D segmentation	multi-class classificatio	binary classification
Prediction (y)	Tumor on slide	Lung Nodule Mask	Brain mask	Risk of death	Kidney and tumor masks	Melanoma class	Heart disease
Center extraction	Hospital	Scanner Manufacturer	Hospital	Group of Hospitals	Group of Hospitals	Hospital	Hospital
Thumbnails							
Original paper	Litjens *et al.* 2018	Armato *et al.* 2011	Perez *et al.* 2021	Liu *et al.* 2018	Heller *et al.* 2019	Tschandl *et al.* 2018 / Codella *et al.* 2017 / Combalia *et al.* 2019	Janosi *et al.* 1988
# clients	2	5	3	5	6	5	4
# examples	399	1,018	566	1, 088	96	23, 247	740
# examples per center	239, 150	670, 205, 69, 74	311, 181, 74	311, 196, 206, 162 51	12, 14, 12, 12, 16, 30	12413, 3954, 3363, 225 819, 439	303, 261, 46, 130
Model	DeepMIL [63]	Vnet [97, 100]	3D U-net [23]	Cox Model [31]	nnU-Net [66]	efficientnet [117] + linear layer	Logistic Regression
Metric	AUC	DICE	DICE	C-index	DICE	Balanced Accuracy	Accuracy
Size	50G (850G total)	115G	444M	115K	54G	9G	40K
Image resolution	0.5 μm / pixel	~1.0 × 1.0 × 1.0 mm / voxel	~ 1.0 × 1.0 × 1.0 mm / voxel	NA	~1.0 × 1.0 × 1.0 mm / voxel	~0.02 mm / pixel	NA
Input dimension	10, 000 x 2048	128 x 128 x 128	48 x 60 x 48	39	64 x 192 x 192	200 x 200 x 3	13

FedML

FedML Product Overview, https://medium.com/@FedML/fedml-ai-platform-releases-the-worlds-federated-learning-open-platform-on-public-cloud-with-an-8024e68a70b6

7. Published 50+ top scientific papers, covering key challenges such as security, efficiency, weak supervision, and fairness

All papers are summarized at https://doc.fedml.ai/resources/papers.html

8. Academia Sponsorship

For more details, please visit: https://fedml.ai/academia-sponsorship/

9. FedML Team

AIFactory

https://aifactory.space/

디지털 헬스케어 서비스를 위한 연합학습 현황 및 과제

이노피아테크 이광기 (kwangkeelee@gmail.com)

Federated Learning

Federated learning is a machine learning setting where multiple entities (clients) collaborate in solving a machine learning problem, under the coordination of a central server or service provider. Each client's raw data is stored locally and not exchanged or transferred; instead, focused updates intended for immediate aggregation are used to achieve the learning objective.

Advances and Open Problems in Federated Learning,
https://arxiv.org/abs/1912.04977

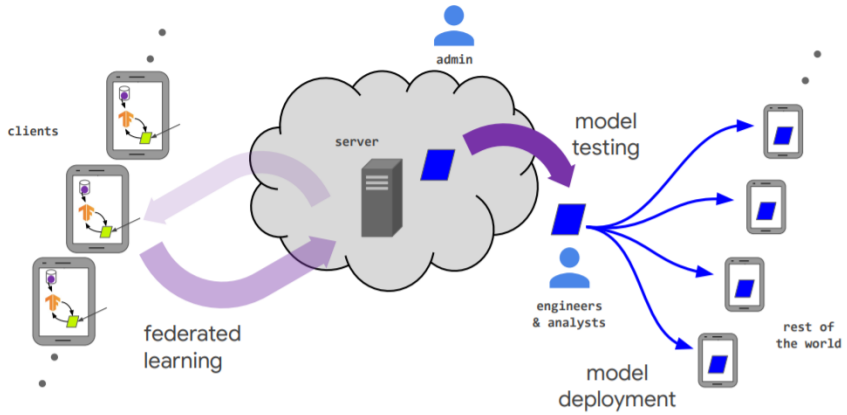
연합학습은 중앙 서버 또는 서비스 제공자의 관리 하에, 다수의 클라이언트/디바이스가 기계학습 문제를 해결하기 위해 협력하는 기술

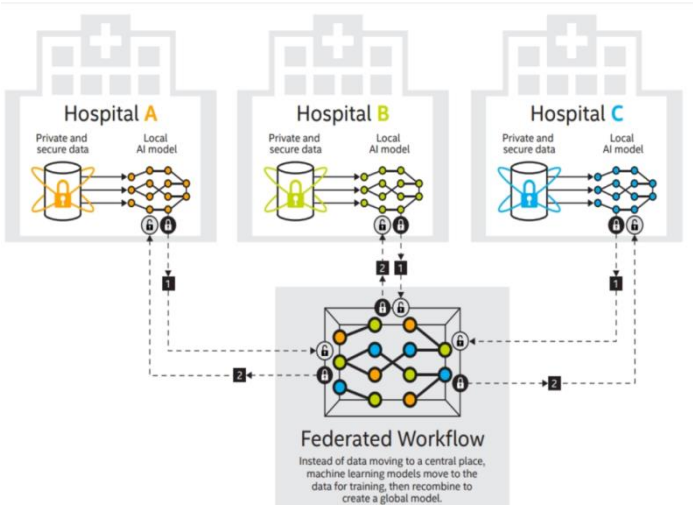
- 각 클라이언트/디바이스는 보유한/생산한 원시 데이터를 교환 또는 (중앙으로) 전송하지 않고, 로컬모델 학습에만 사용함으로써, 데이터 생산자의 프라이버시 보호
- 각 클라이언트/디바이스에서의 학습 결과는 (중앙의) 글로벌 모델 학습에 반영/기여. 'A fed B'학습의 성능은 'A+B'성능에 근사
- 데이터 생산자의 프라이버시 보호, 통신 오버헤드 감소

Federated Learning

▶ 개인 정보의 노출/침해 없이, 데이터를 확보/활용할 수 있는 연합학습 기술

- 인공지능 모델을 학습하기 위해서는 많은 양의 데이터가 필요하지만, 데이터 프라이버시 정책 등으로 인하여 (개인)데이터 수집/활용에 제약
- 기존에는 중앙 서버에 모든 데이터를 수집 후 학습하는 과정이 일반적으로, 프라이버시 침해 위험이 존재. 이를 개선하기 위해 각 디바이스에서 로컬 모델을 학습하고 이를 동기화하는 연합학습 기술 필요성 대두
- 연합학습 기술은 사용자 로컬 데이터에 직접 접근하지 않으면서 모든 사용자들의 정보를 반영한 글로벌 모델을 학습하여 이용할 수 있음





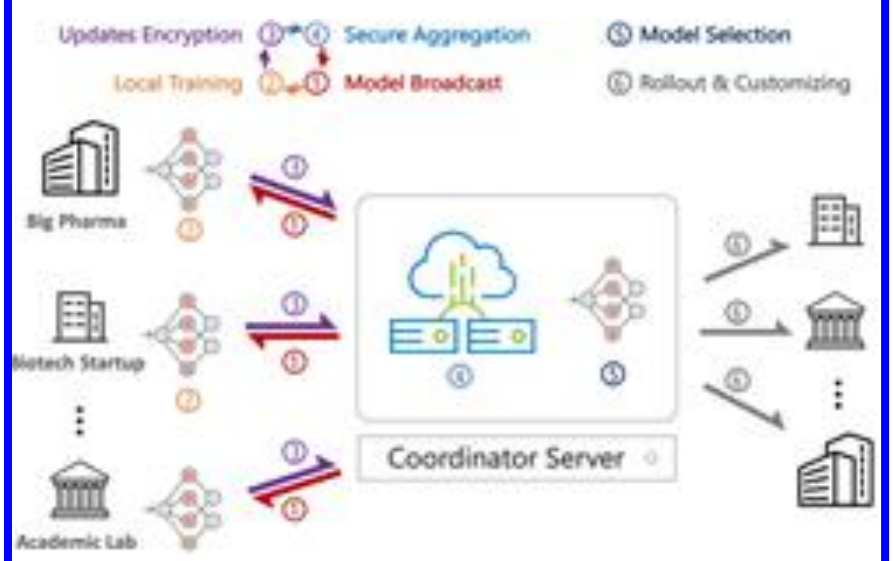
연합학습 개요

- 연합학습은, 로컬 데이터 샘플을 보유하는 다수의 분산 에지 장치 또는 서버들이 원시 데이터를 교환/공유하지 않고 기계학습 문제를 해결하기 위해 협력하는 기술
- 각 로컬노드(클라이언트/디바이스)는 생산한/보유한 원시 데이터를 로컬모델 학습에만 사용함으로써, **데이터 생산자/제공자의 프라이버시를 보호하고, 데이터 소유/활용의 파편화 문제를 해결**
- 모든 로컬 데이터 세트가 하나의 서버에 업로드/공유 되는 전통적인 중앙집중식 기계학습 방식 혹은 로컬 데이터 샘플이 동일하게 분포 (identically distributed) 된다고 가정하는 전통적인 분산접근 방식과는 대비됨
- 연합학습은 데이터 소유/관리/활용의 파편화 문제를 해결하기 위한 사일로-교차(Cross-silo) 연합학습, 디바이스/서비스 사용자 데이터를 활용하기 위한 디바이스-교차(Cross-device) 연합학습으로 특징과 이슈를 구분

	분산학습 (Datacenter distributed learning)	**사일로-교차 연합학습 (Cross-silo federated learning)**	**디바이스-교차 연합학습 (Cross-device federated learning)**
환경	단일 크러스터 혹은 데이터센터가 대규모 데이터로 학습	서로 다른 기관(의료 혹은 금융) 혹은 지리적으로 분산되어 있는 데이터센터들이, 각자의 사일로 데이터를 학습	**클라이언트는 많은 수의 모바일 혹은 IoT 디바이스**
데이터 분산	데이터는 중앙에 저장되며, 클라이언트들은 데이터에 제한 없이 접근, 혼합	**데이터는 로컬에서 생성, 분산되어 있음. 각 클라이언트는 자신의 데이터를 저장하며 다른 클라이언트의 데이터를 읽을 수 없음. 데이터는 iid (independently or identically distributed) 하지 않음**	
오케스트레이션	중앙에서 데이터 관리와 학습을 관장	**중앙 오케스트레이션 서버/서비스 주도로 학습을 관장하지만, 원시 데이터에는 접근하지 않음**	
데이터 가용성	모든 클라이언트가 항상 가용		**일정 시간에, 일부 클라이언트만 가용**
분산 규모	1 - 1000 클라이언트	2 - 100 클라이언트	**10^{10} 까지 대규모**
주요 병목	Computation (연산량 및 연산속도)	연산 및 통신	**일반적으로 통신이 주된 병목**

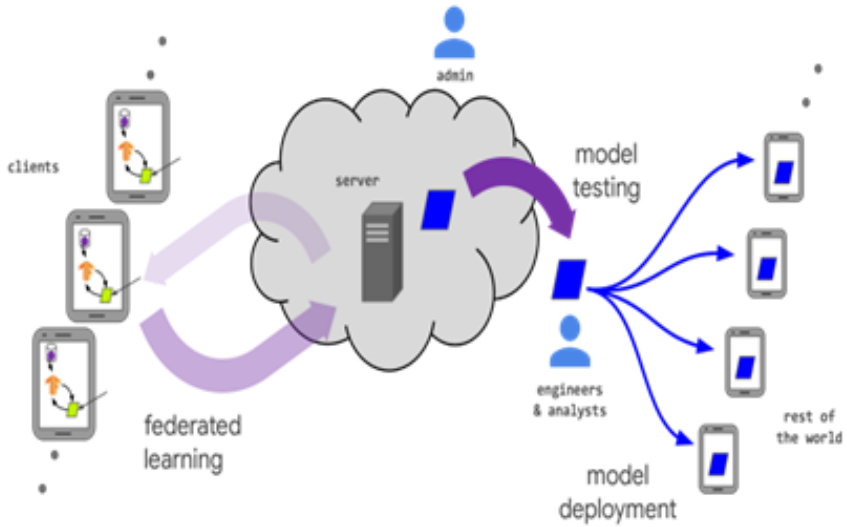
Advances and Open Problems in Federated Learning, https://arxiv.org/abs/1912.04977

연합학습 개요 : Cross-silo vs. Cross-device



사일로-교차 연합학습 (Cross-silo FL) :

- 서로 다른 기관 (의료 혹은 금융) 혹은 지리적으로 분산되어 있는 데이터센터들이, 각자의 사일로 데이터를 학습 : 2 - 100 clients
- 데이터/통계적 이질성, 디바이스/시스템적 이질성 문제 小
- 모든 클라이언트가 항상 가용



디바이스-교차 연합학습 (Cross-device FL) :

- 사용자의 개인 디바이스 (휴대폰, IoT) 가 개인 데이터를 학습 : Massive # of clients
- 데이터/통계적 이질성, 디바이스/시스템적 이질성 문제 大
- 일정 시간에 일부 클라이언트만 가용하고, straggler effect 대응 필요

* **통계적 이질성**: 다수의 다양한 사용자/디바이스, 동적 환경 및 시공간으로부터 수집된 데이터는 독립동일분포(iid: independent identically distributed) 조건을 만족하지 못하고 비균일/불균형의 특성을 지님

** **시스템적 이질성**: 연합학습에 참여/기여하는 디바이스의 성능과 기능 및 네트워크 환경이 다양하고, 디바이스의 추가, 변동이 지속적으로 발생

인공지능 기술청사진 2030 2차년도 보고서,
https://www.iitp.kr/kr/1/knowledge/openReference/view.it?ArticleIdx=5248&count=true

연합학습 개요

Applications of cross-device federating learning

What makes a good application?

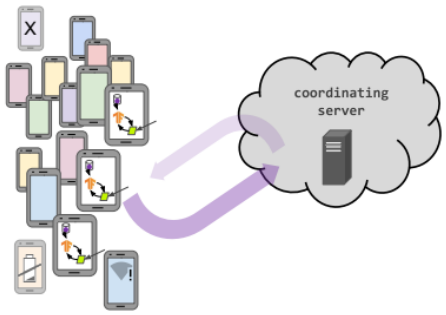
- On-device data is more relevant than server-side proxy data
- On-device data is privacy sensitive or large
- Labels can be inferred naturally from user interaction

Example applications

- Language modeling for mobile keyboards and voice recognition
- Image classification for predicting which photos people will share
- ...

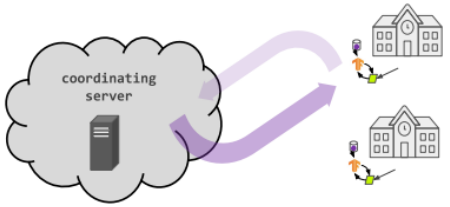
Cross-device federated learning

millions of intermittently available client devices



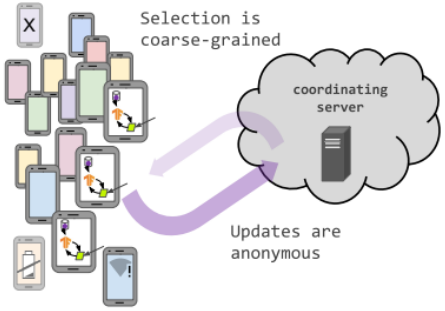
Cross-silo federated learning

small number of clients (institutions, data silos), high availability



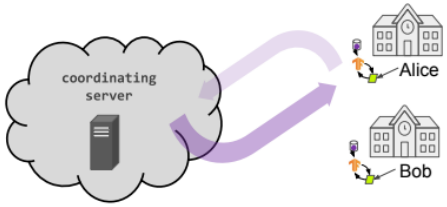
Cross-device federated learning

clients cannot be indexed directly (i.e., no use of client identifiers)



Cross-silo federated learning

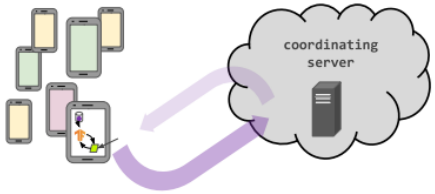
each client has an identity or name that allows the system to access it specifically



Cross-device federated learning

Server can only access a (possibly biased) random sample of clients on each round.

Large population => most clients only participate once.

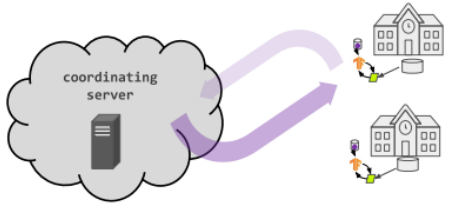


round 2
(completely new set of devices participate)

Cross-silo federated learning

Most clients participate in every round.

Clients can run algorithms that maintain local state across rounds.

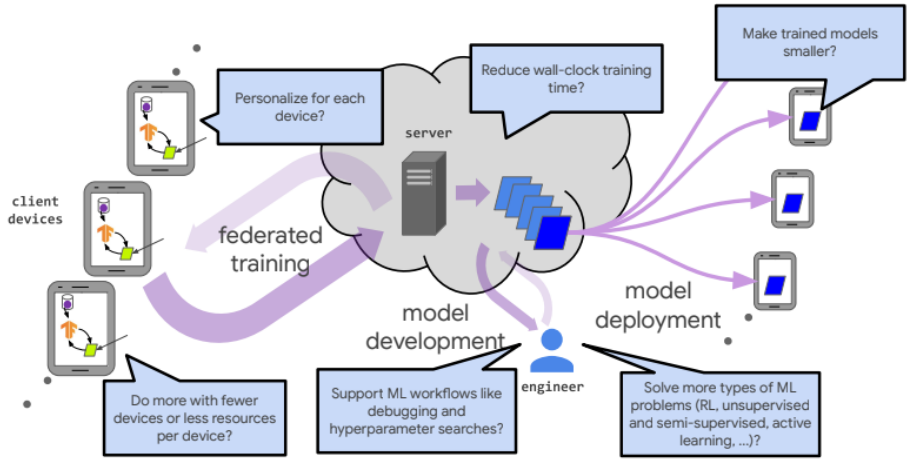


round 2
(same clients)

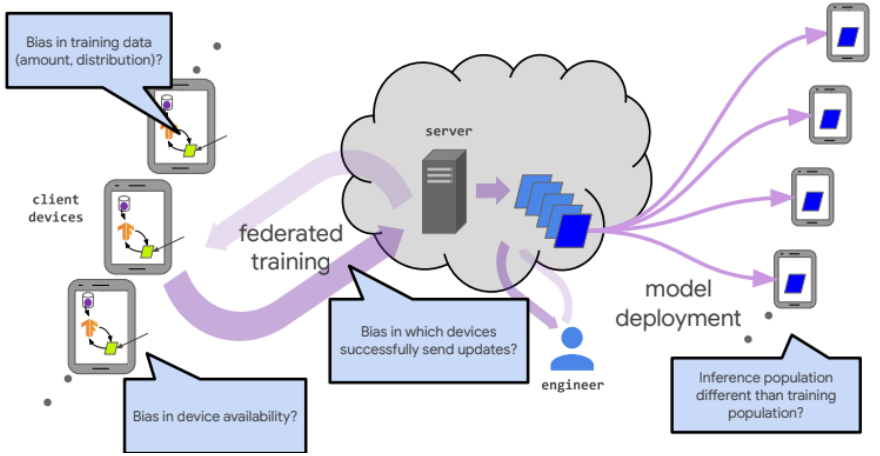
Federated Learning Tutorial@NeurIPS 2020, https://sites.google.com/view/fl-tutorial/

연합학습 개요

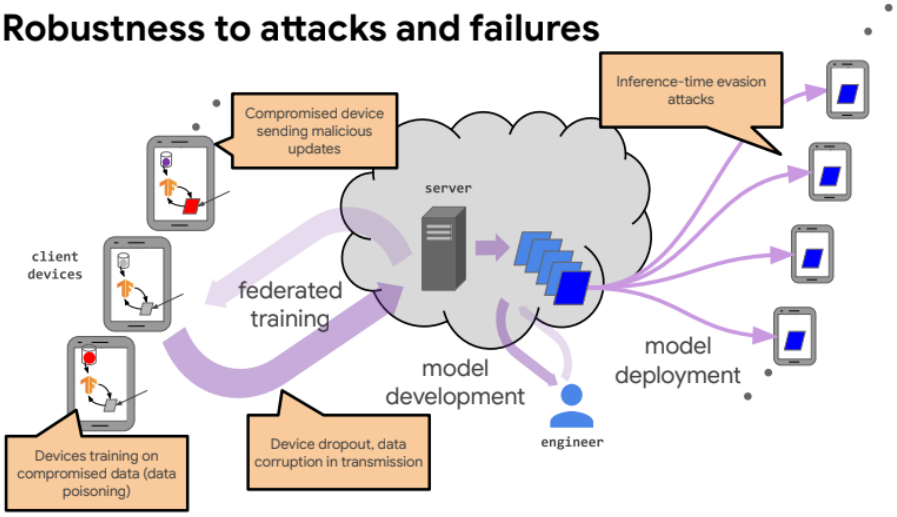
Improving efficiency and effectiveness



Ensuring fairness and addressing sources of bias



Robustness to attacks and failures



Advances and Open Problems in Federated Learning

Peter Kairouz[7*] H. Brendan McMahan[7*] Brendan Avent[21] Aurélien Bellet[9]
Mehdi Bennis[16] Arjun Nitin Bhagoji[13] Keith Bonawitz[7] Zachary Charles[7]
Graham Cormode[23] Rachel Cummings[6] Rafael G.L. D'Oliveira[14]
Salim El Rouayheb[14] David Evans[22] Josh Gardner[24] Zachary Garrett[7]
Adrià Gascón[7] Badih Ghazi[7] Phillip B. Gibbons[9] Marco Gruteser[7,14]
Zaid Harchaoui[24] Chaoyang He[21] Lie He [4] Zhouyuan Huo [20]
Ben Hutchinson[7] Justin Hsu[25] Martin Jaggi[4] Tara Javidi[17] Gauri Joshi[2]
Mikhail Khodak[2] Jakub Konečný[7] Aleksandra Korolova[23] Farinaz Koushanfar[17]
Sanmi Koyejo[7,16] Tancrède Lepoint[7] Yang Liu[12] Prateek Mittal[13]
Mehryar Mohri[7] Richard Nock[1] Ayfer Özgür[15] Rasmus Pagh[7,10]
Mariana Raykova[7] Hang Qi[7] Daniel Ramage[7] Ramesh Raskar[11]
Dawn Song[16] Weikang Song[7] Sebastian U. Stich[4] Ziteng Sun[3]
Ananda Theertha Suresh[7] Florian Tramèr[15] Praneeth Vepakomma[11] Jianyu Wang[2]
Li Xiong[5] Zheng Xu[7] Qiang Yang[8] Felix X. Yu[7] Han Yu[12] Sen Zhao[7]

[1]Australian National University, [2]Carnegie Mellon University, [3]Cornell University, [4]École Polytechnique Fédérale de Lausanne, [5]Emory University, [6]Georgia Institute of Technology, [7]Google Research, [8]Hong Kong University of Science and Technology, [9]INRIA, [10]IT University of Copenhagen, [11]Massachusetts Institute of Technology, [12]Nanyang Technological University, [13]Princeton University, [14]Rutgers University, [15]Stanford University, [16]University of California Berkeley, [17]University of California San Diego, [18]University of Illinois Urbana-Champaign, [19]University of Oulu, [20]University of Pittsburgh, [21]University of Southern California, [22]University of Virginia, [23]University of Warwick, [24]University of Washington, [25]University of Wisconsin-Madison

Abstract

Federated learning (FL) 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. FL embodies the principles of focused data collection and minimization, and can mitigate many of the systemic privacy risks and costs resulting from traditional, centralized machine learning and data science approaches. Motivated by the explosive growth in FL research, this paper discusses recent advances and presents an extensive collection of open problems and challenges.

Advances and Open Problems in FL

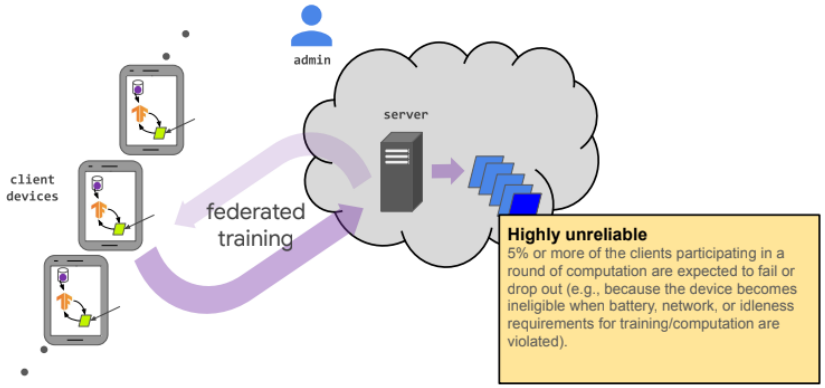
58 authors from 25 top institutions

arxiv.org/abs/1912.04977

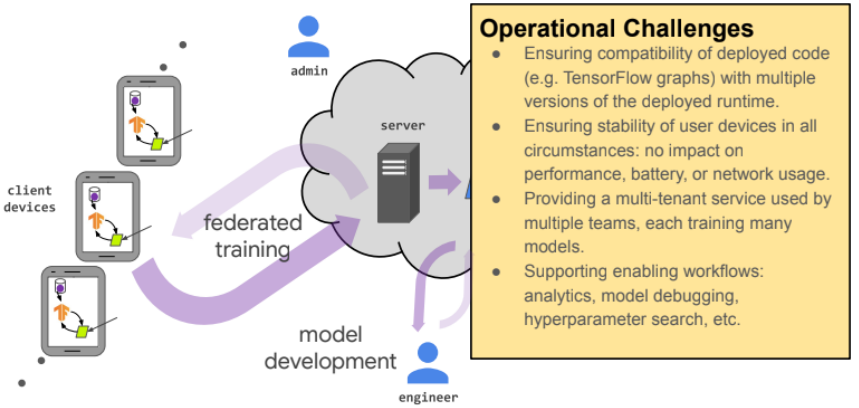
Federated Learning Tutorial@NeurIPS 2020, https://sites.google.com/view/fl-tutorial/

연합학습 개요

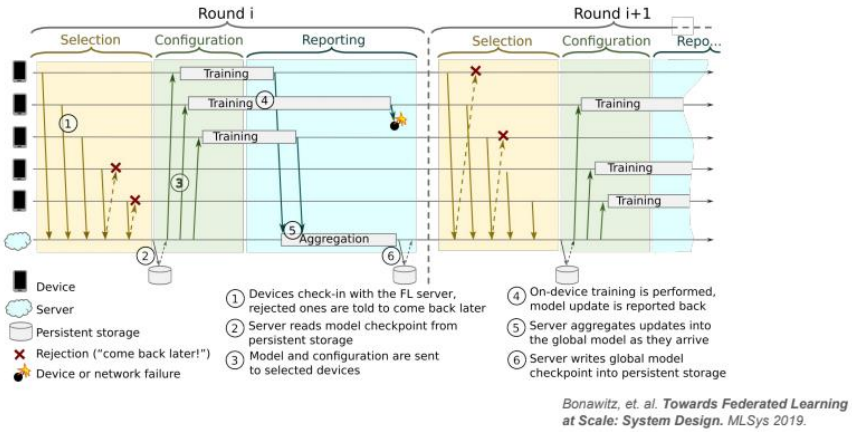
System challenges in cross-device FL



System challenges in cross-device FL

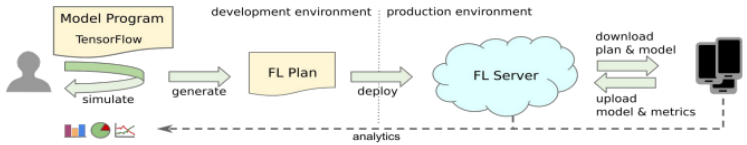


An example cross-device federated learning protocol



Bonawitz, et. al. ***Towards Federated Learning at Scale: System Design.*** *MLSys 2019.*

Developer workflows in federated learning



- Model developers depend on the production system for experimentation
  - They only have access to proxy data but not to the real data
  - Develop in Python, then push the result automatically to production and get metrics back
- Experimentation must never affect the user experience on devices
  - Training has no visible effect to the user -- inference models are manually pushed
  - Device architecture ensures that device health is not affected

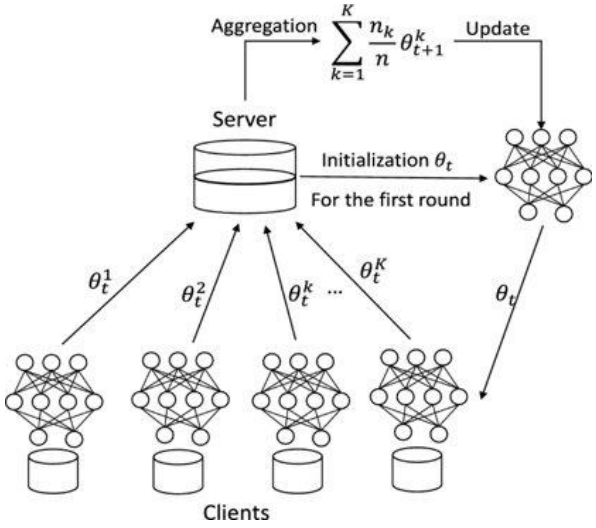
Federated Learning Tutorial@NeurIPS 2020, https://sites.google.com/view/fl-tutorial/

Open FL Platforms

Open FL Platform (Active)	Affiliation
Flower: A Friendly Federated Learning Framework, https://flower.dev/	U of Cambridge, Samsung
FedScale: A scalable and extensible federated learning engine and benchmark, http://fedscale.ai/	U of Michigan
FedML: https://fedml.ai/	USC & FedML Inc.
FLSim: https://github.com/facebookresearch/FLSim	Meta/Facebook Research

Open FL Platform (Etc)	Affiliation
TensorFlow Federated (TFF), https://www.tensorflow.org/federated?hl=ko	Google
PySyft, https://github.com/OpenMined/PySyft	OpenMined
FedAI, https://www.fedai.org/	WeBank
Open Federated Learning (OpenFL) - An Open-Source Framework For Federated Learning, https://github.com/intel/openfl	Intel
IBM Federated Learning, https://ibmfl.mybluemix.net/	IBM Research.
OpenFed, https://github.com/FederalLab/OpenFed	--
FL_PyTorch, https://github.com/burlachenkok/flpytorch	--

Vanilla Federated Learning



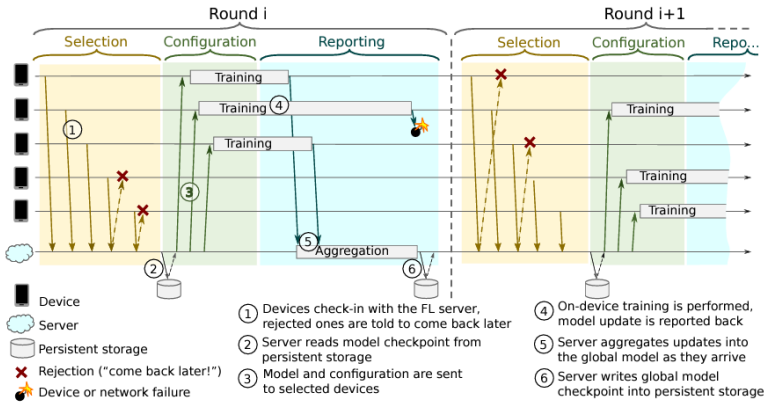


Figure 1: Federated Learning Protocol

Algorithm 1 `FederatedAveraging`. The K clients are indexed by k; B is the local minibatch size, E is the number of local epochs, and η is the learning rate.

Server executes:

initialize w_0

for each round $t = 1, 2, \dots$ **do**

$m \leftarrow \max(C \cdot K, 1)$

$S_t \leftarrow$ (random set of m clients)

for each client $k \in S_t$ **in parallel do**

$w_{t+1}^k \leftarrow \text{ClientUpdate}(k, w_t)$

$w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$

ClientUpdate(k, w): *// Run on client k*

$\mathcal{B} \leftarrow$ (split $\mathcal{P}_k$ into batches of size B)

for each local epoch i from 1 to E **do**

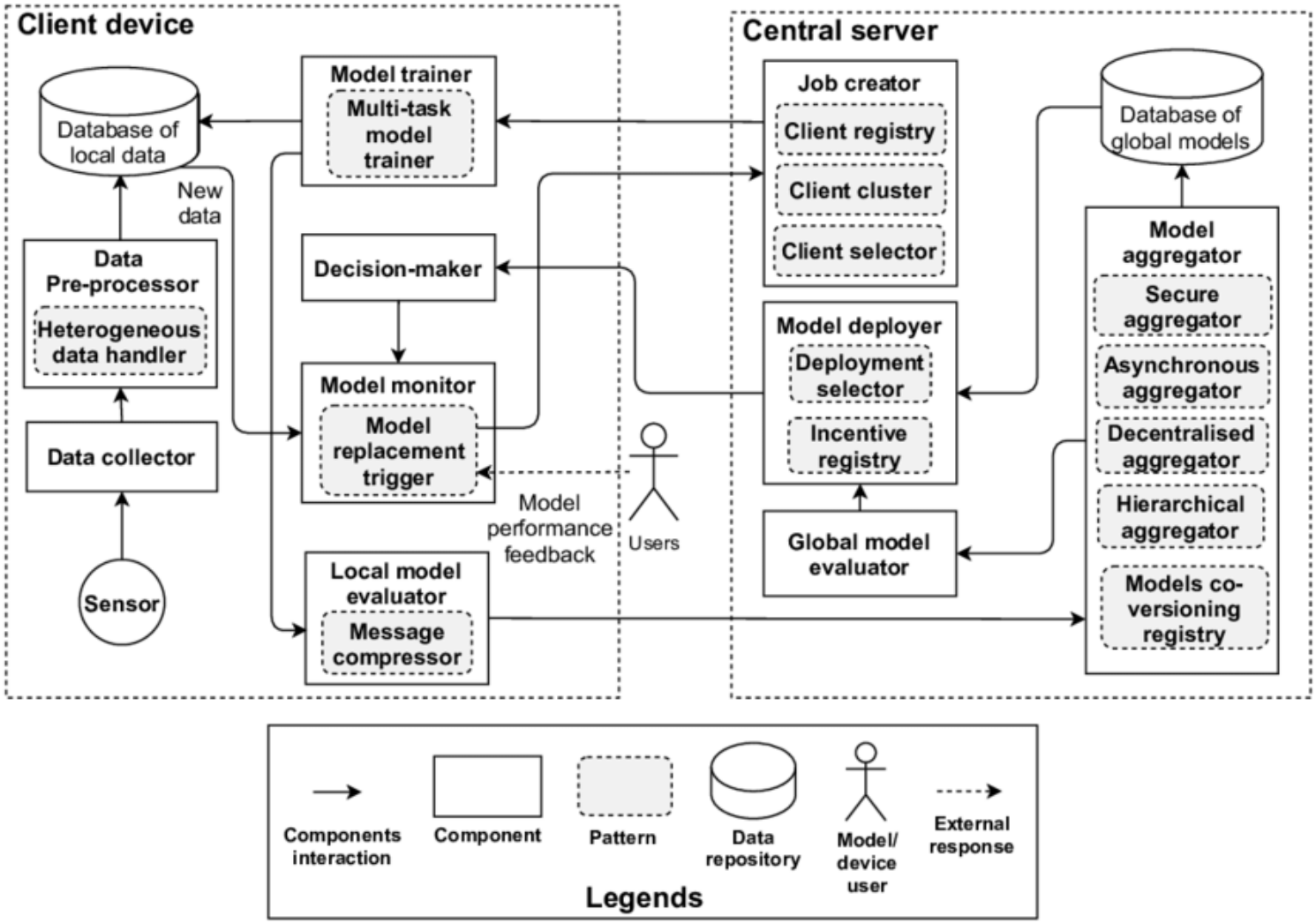
for batch $b \in \mathcal{B}$ **do**

$w \leftarrow w - \eta \nabla \ell(w; b)$

return w to server

FLRA: A Reference Architecture for Federated Learning Systems

FLRA: A Reference Architecture for Federated Learning Systems, https://arxiv.org/abs/2106.11570



Personalization for FL

*** 연합학습은 일반적으로 모든 디바이스 및 사용자에 공통으로 적용되는 글로벌모델을 학습하는 것을 목표로 하고 있으나, 동적인 디바이스 환경의 데이터 이질성 및 디바이스 이질성으로 인하여 **모든 디바이스에서 잘 동작하는 하나의 모델을 학습하기 어려우며, 개별 디바이스 및 사용자 관점에서 최적의 성능이 보장되지 않음.** 동적인 디바이스 환경에서 각 사용자 및 디바이스의 특징과 애플리케이션 요구사항을 최적 반영하기 위해서는, 글로벌 모델 뿐 만 아니라 **개인화·로컬 모델(locally adapted personalized model)의 성능을 최적화**할 수 있는 연합학습 기술 필요

Personalization 방식	특징
Adding User Context	▪ user clustering where similar clients are grouped together and a separate model is trained for each group.
Transfer Learning	▪ some or all parameters of a trained global model are re-learned on local data. ▪ To avoid the problem of catastrophic forgetting [21] [22], care must be taken to not retrain the model for too long on local data. A variant technique freezes the base layers of the global model and retrains only the top layers on local data. Transfer learning is also known as fine-tuning, and it integrates well into the typical federated learning lifecycle.
Multi-task Learning	▪ multiple related tasks are solved simultaneously allowing the model to exploit commonalities and differences across the tasks by learning them jointly
Meta-Learning	▪ **MAML builds an internal representation generally suitable for multiple tasks, so that fine tuning the top layers for a new task can produce good results. MAML proceeds in two connected stages: meta-training and meta-testing.** ➢ **Meta-training builds the global model on multiple tasks, and** ➢ **meta-testing adapts the global model individually for separate tasks.**
Knowledge Distillation	▪ extracting the knowledge of a large teacher network into a smaller student network by having the student mimic the teacher.
Base + Personalization Layers	▪ the base layers are trained centrally by Federated Averaging, and the top layers (also called personalization layers) are trained locally with a variant of gradient descent
Mixture of Global and Local Models	▪ Instead of learning a single global model, each device learns a mixture of the global model and its own local model.

Survey of Personalization Techniques for Federated Learning, https://arxiv.org/abs/2003.08673

Blockchain-based trustworthy federated learning architecture

Towards Trustworthy AI: Blockchain-based Architecture Design for Accountability and Fairness of Federated Learning Systems, https://ieeexplore.ieee.org/abstract/document/9686048

https://github.com/Kwangkee/FL/blob/main/FL%40CSIRO.md#towards-trustworthy-ai

- However, federated learning systems struggle to achieve and embody responsible AI principles. In particular, federated learning systems face accountability and fairness challenges due to multi-stakeholder involvement and heterogeneity in client data distribution. To enhance the accountability and fairness of federated learning systems, we present a blockchain-based trustworthy federated learning architecture.
- We designed the architecture based on a reference architecture for federated learning system named FLRA [6]. https://arxiv.org/abs/2106.11570

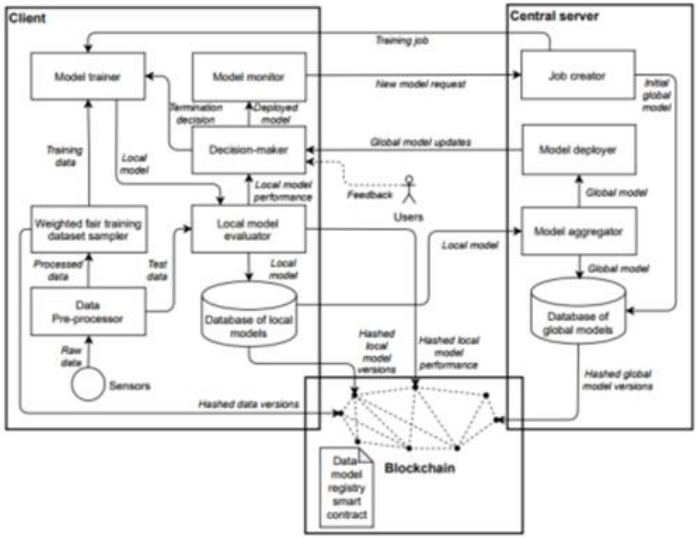


Fig. 1: Blockchain-based Trustworthy Federated Learning Architecture

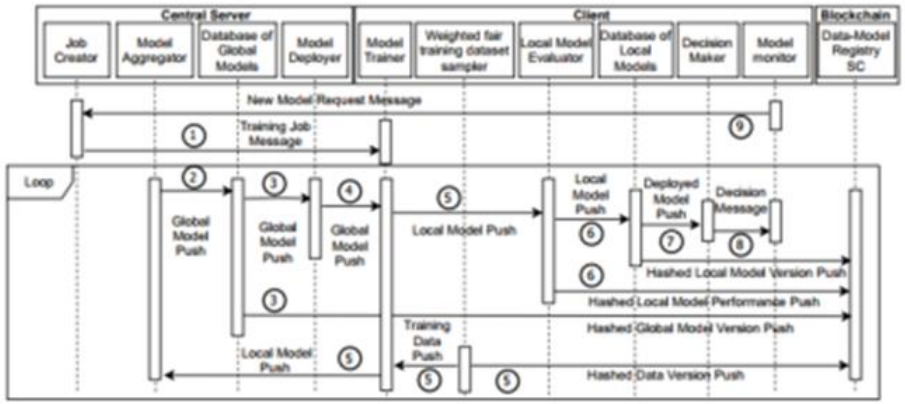
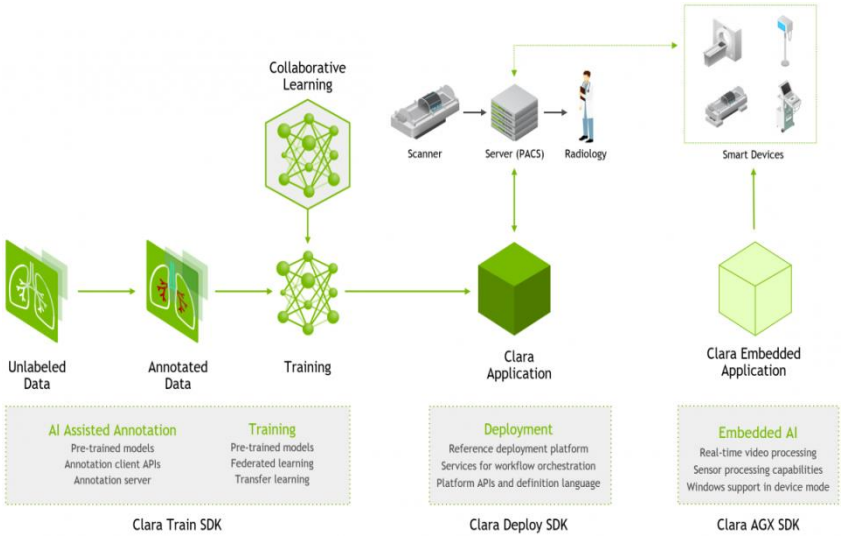


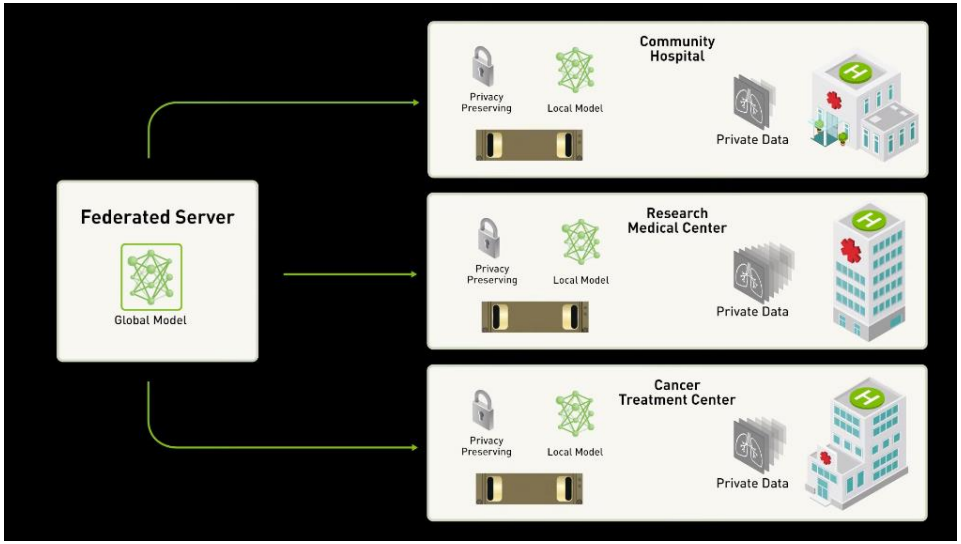
Fig. 2: Sequence Diagram of Blockchain-based Trustworthy Federated Learning Process

Federated Learning powered by NVIDIA Clara

- An Application Framework Optimized for Healthcare and Life Sciences Developers, https://developer.nvidia.com/clara
- Federated Learning powered by NVIDIA Clara, https://developer.nvidia.com/blog/federated-learning-clara/

- Transforming AI Healthcare with Federated Learning, https://news.developer.nvidia.com/transforming-ai-healthcare-with-federated-learning/
- https://www.nature.com/articles/s41746-020-00323-1



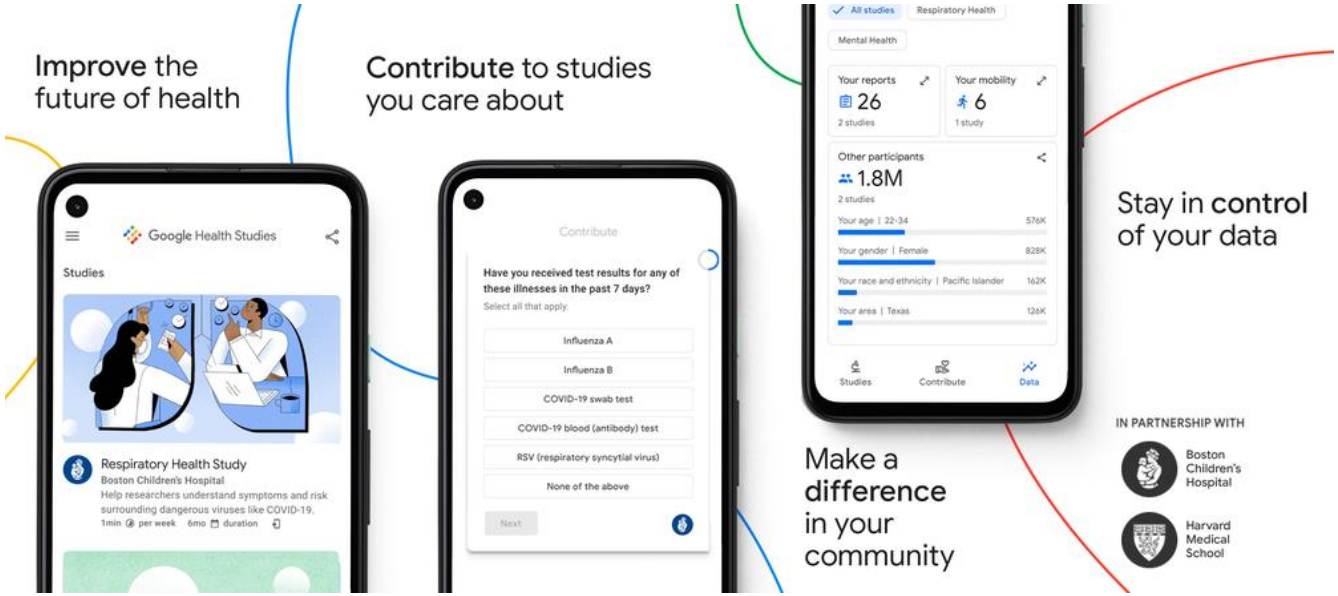


Google Health Studies : 연합학습 적용

Blog: https://blog.google/technology/health/google-health-studies-app/
App: https://play.google.com/store/apps/details?id=com.google.android.apps.health.research.studies

The Google Health Studies app is now available in the Google Play Store, and we're inviting people to download the app to join this initial study. We look forward to partnering with health researchers and to making it possible for more people to participate in these important studies.

... this first study utilizes federated learning and analytics—a privacy technology that keeps a person's data stored on the device, while allowing researchers to discover aggregate insights based on encrypted, combined updates from many devices. This means researchers in this study can examine trends to understand the link between mobility (such as the number of daily trips a person makes outside the home) and the spread of COVID-19, This same approach powers typing predictions on Gboard, without Google seeing what individuals type.

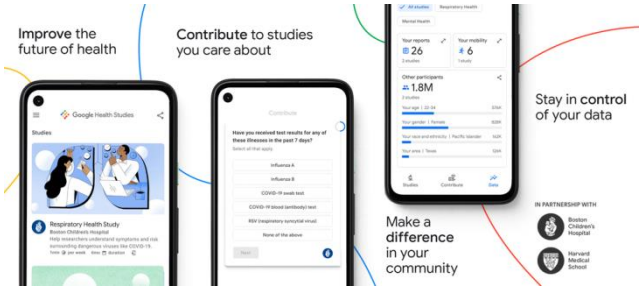


Google Health Studies : 연합학습 적용

Blog: https://blog.google/technology/health/google-health-studies-app/
App: https://play.google.com/store/apps/details?id=com.google.android.apps.health.research.studies

Google Health Studies lets you securely contribute to health research studies with leading institutions, right from your phone. Volunteer for studies that matter to you and represent your community.

Simply download the app and enroll in a study.



Help researchers make advancements in medicine and healthcare:
- Self-report symptoms and other data
- Volunteer for multiple studies in one app
- Track your information with digital health reports
- Learn research findings from the studies you participate in

Help scientists better understand respiratory diseases.
The first study available is a respiratory health study conducted by Boston Children's Hospital and Harvard Medical School. If you participate in this study, you'll provide data to help researchers understand how demographics, health history, behavior, and mobility patterns contribute to the spread of respiratory illnesses. Upcoming studies will research mental health and diabetes.

You're in control of your data: In the respiratory health study, your personal information is kept on your device. Researchers only see aggregated study data combined from all participants. This allows researchers to collect the information needed to advance the study without seeing individual details.

Your input matters: Google Health Studies aims to create opportunities for more people to participate in health research. By contributing, you'll represent your community and start improving the future of health for everyone.