## 联邦学习开源平台功能分析

1. **FedML Open Platform**

**1、模拟实验与真实部署迁移**

FedML是一个工具，它可以帮助人们把在电脑或者开发服务器上做的测试转换成在真实世界中使用的程序。这个工具可以让那些在实验里测试的代码，不需要做任何改变，就能直接用在真实的设备上，处理真实的数据。

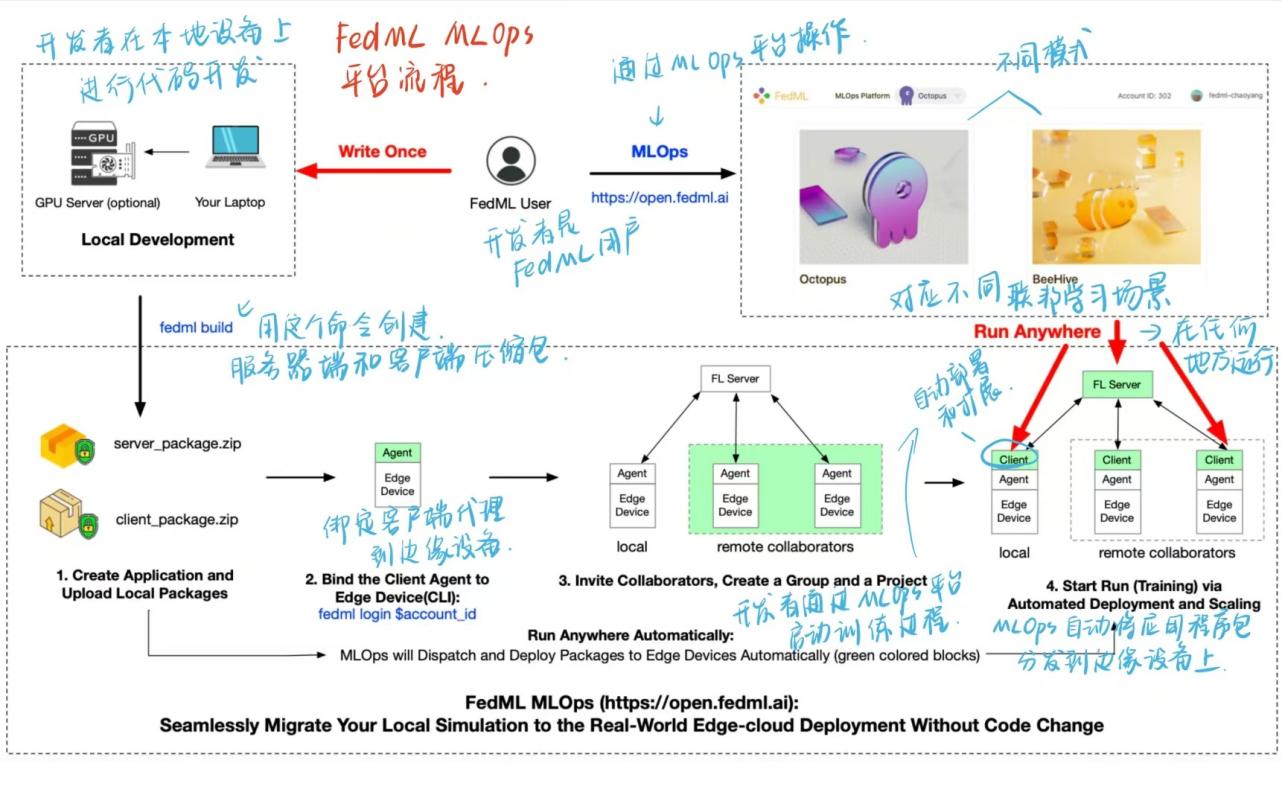
具体来说，这个过程是这样的：

• 首先，人们在自己的电脑或者开发服务器上做实验，看看他们的联邦学习是否有效。

• 一旦确认有效，他们就可以使用FedML这个工具，把实验阶段的代码变成可以在真实设备上运行的程序。

• 这个过程很简单，只需要几个命令和一些界面操作，就可以把程序安装包和脚本发送到任何他们想要的私有设备上。

简单来说，FedML起到了一个桥梁的作用，把测试实验中的代码和真实世界中的应用连接起来，使得开发者只需要编写一次代码，就可以在不同的环境个设备上运行。



**FedML MLOps工作流程**

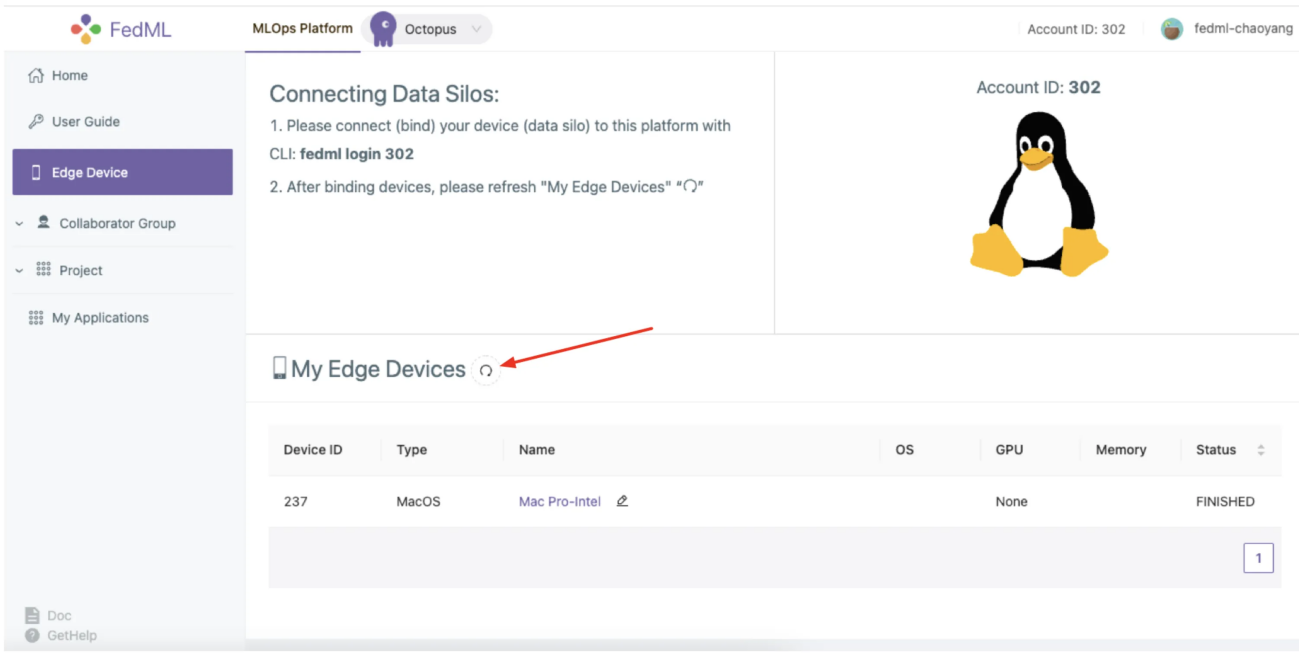
**2、边缘模型训练部署**

·用户登录平台：首先，你需要在FedML平台上创建一个账户，然后使用你的账户信息登录。

·使用“fedml login”命令：登录后，你只需要在你的电脑上输入一个简单的命令“fedml login”，这个命令会帮助你把你的边缘设备（比如手机、平板电脑或者其他智能设备）和FedML平台连接起来。

·查看边缘设备：登录并关联设备后，你可以在FedML平台的“Edge Devices”（边缘设备）页面上看到你已经绑定好的设备。这些设备就是你用来进行联邦学习训练的设备。

·提高研发效率的命令：除了“fedml login”这个命令，FedML还提供了其他一些命令，这些命令可以帮助你更快地开发和测试你的联邦学习项目，比如上传代码、配置训练参数等。



**3、更具安全性和隐私性的服务器部署**

FedML提供了两种不同的服务器部署方式：公有云和私有云（包含Docker）

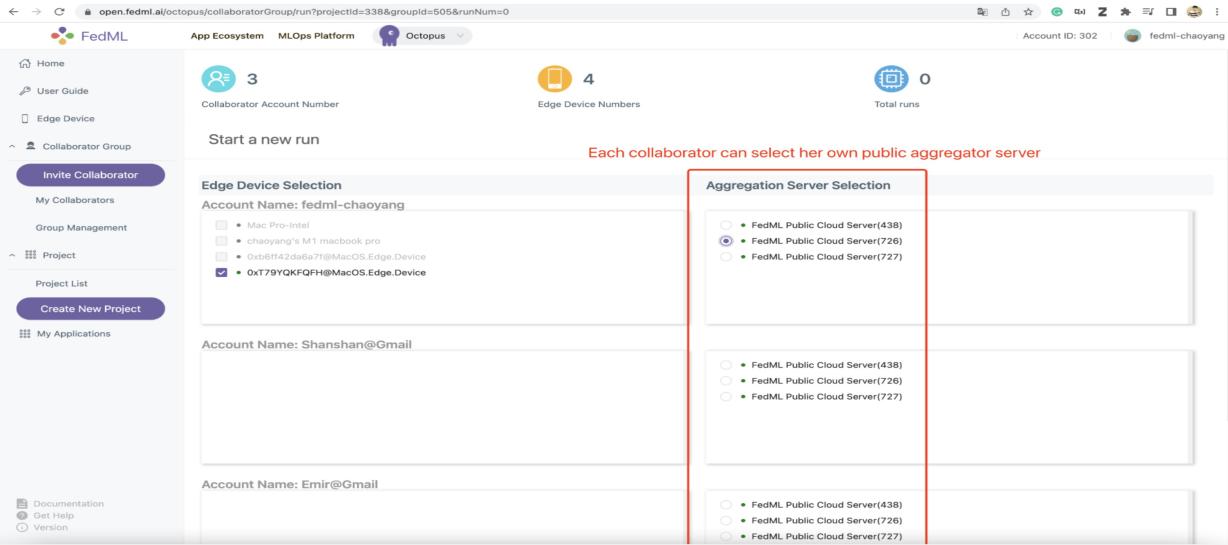
·公有云聚合服务器：

FedML提供了一个公有云服务，这个服务里有免费的服务器节点，用户可以利这些节点来训练自己的联邦学习模型。

·私有云部署：

对于那些需要更高安全性和隐私保护的用户，FedML也支持在私有云上部署聚合服务器。这意味着用户可以自己控制服务器的部署，确保数据的安全性。

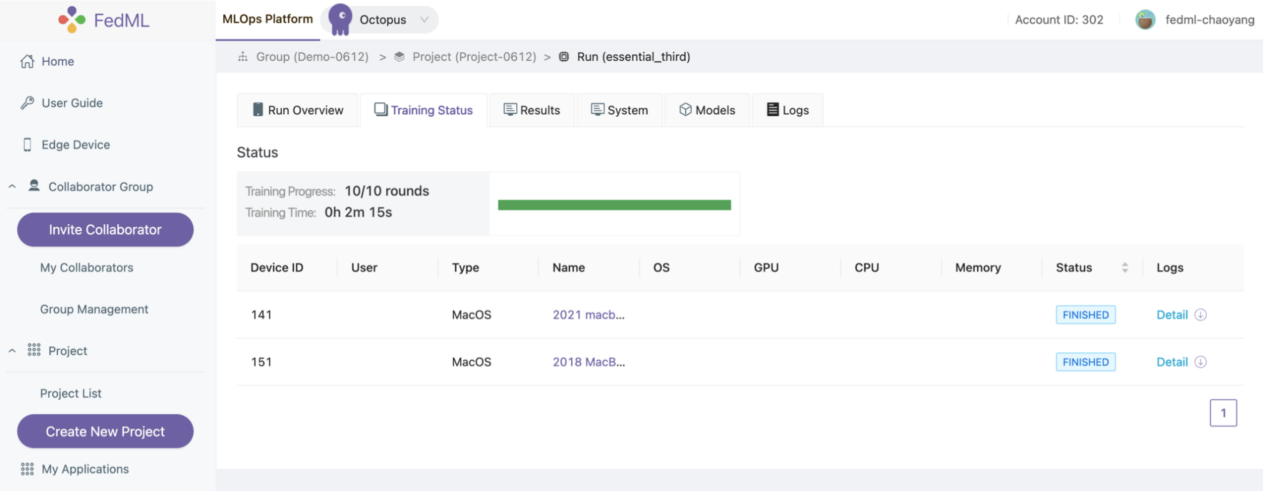
·Docker模式：

Docker是一种容器化技术，可以确保应用程序在任何环境中都能以相同的方式运行。FedML利用Docker来帮助用户部署私有聚合服务

1. **分布式训练中的追踪、管理和分析**

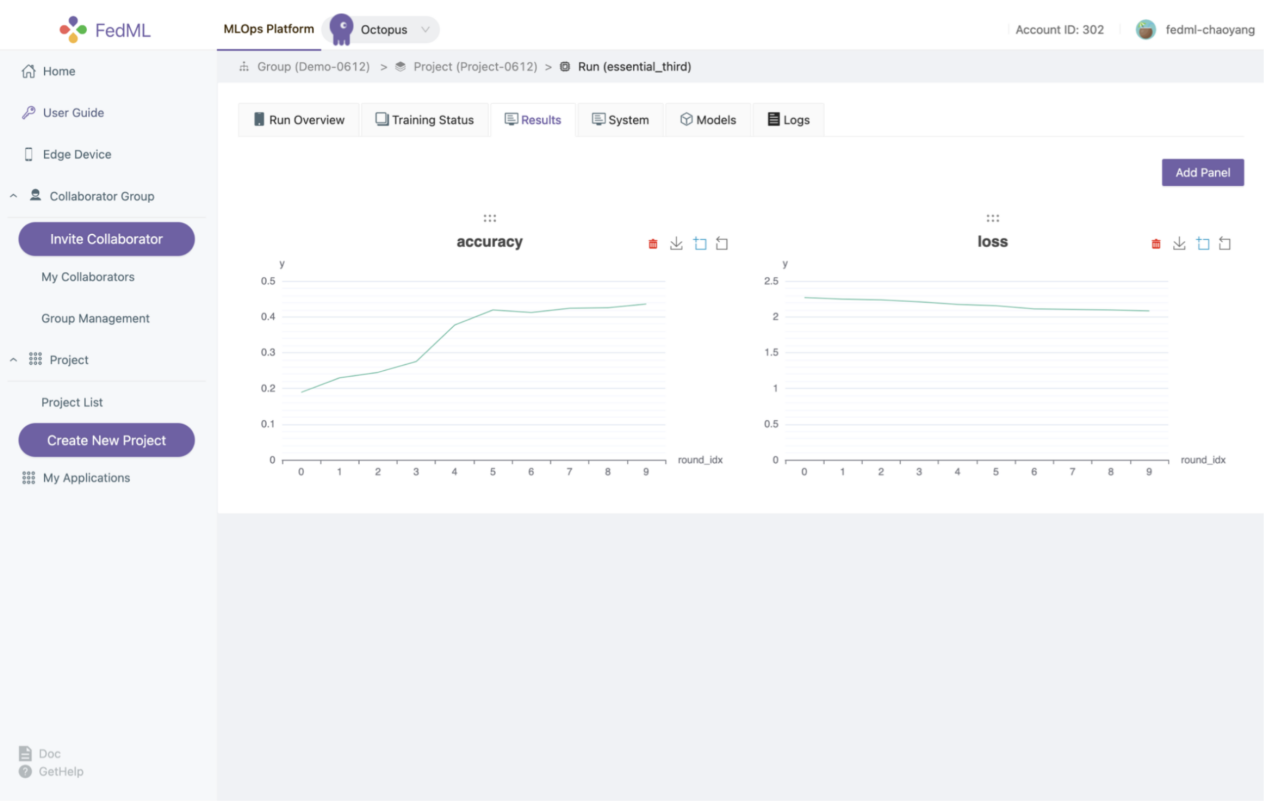
·边缘设备训练状态追踪：

就像你可以通过手机应用追踪你的快递包裹一样，FedML平台允许你实时查看每个参与训练的边缘设备（比如手机、平板电脑等）的状态，了解它们是否在正常工作，以及它们的训练进度。



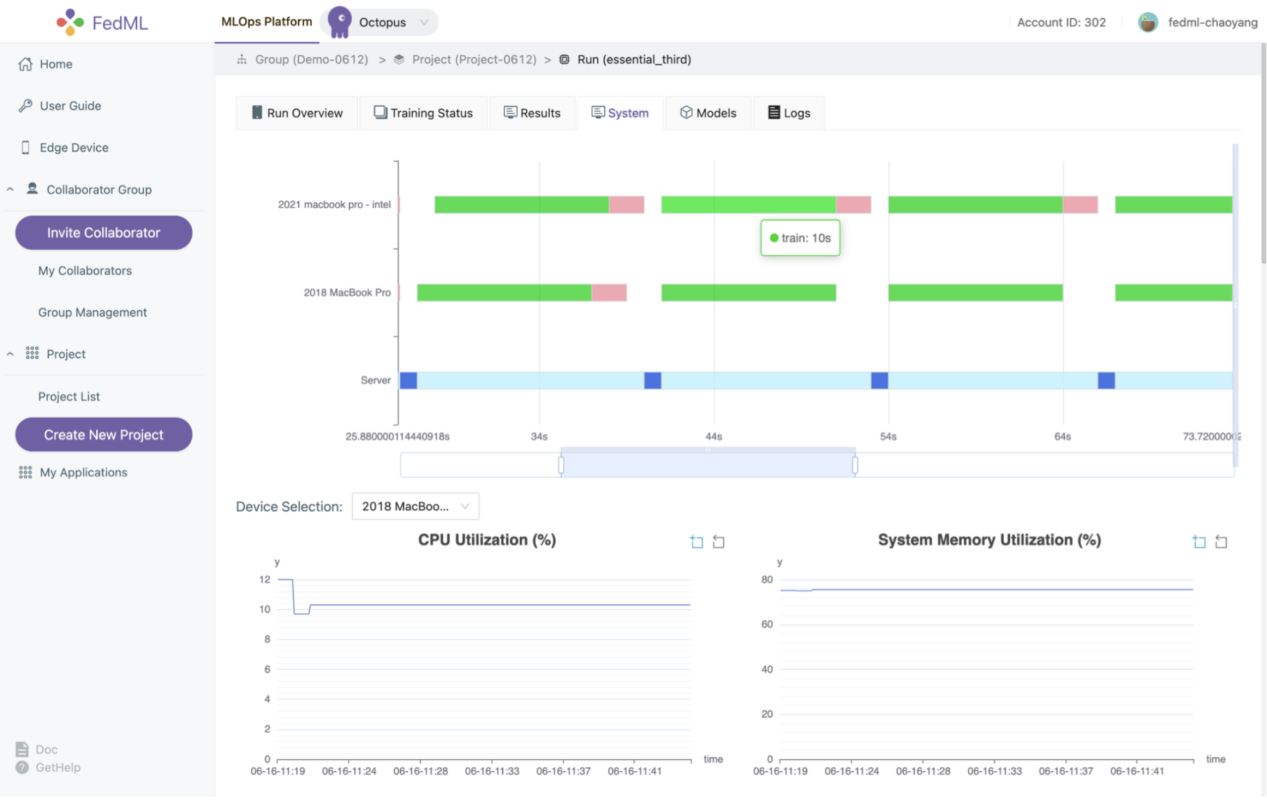
·自定义metrics上报：

在训练AI模型时，你可能会关心模型的准确率（比如分类任务）或者误差率（比如回归任务）。FedML平台允许你设置这些关键指标，并且自动上报这些指标，让你可以监控模型的性能。此外，你还可以看到系统资源的使用情况，比如CPU、内存、GPU的使用率。



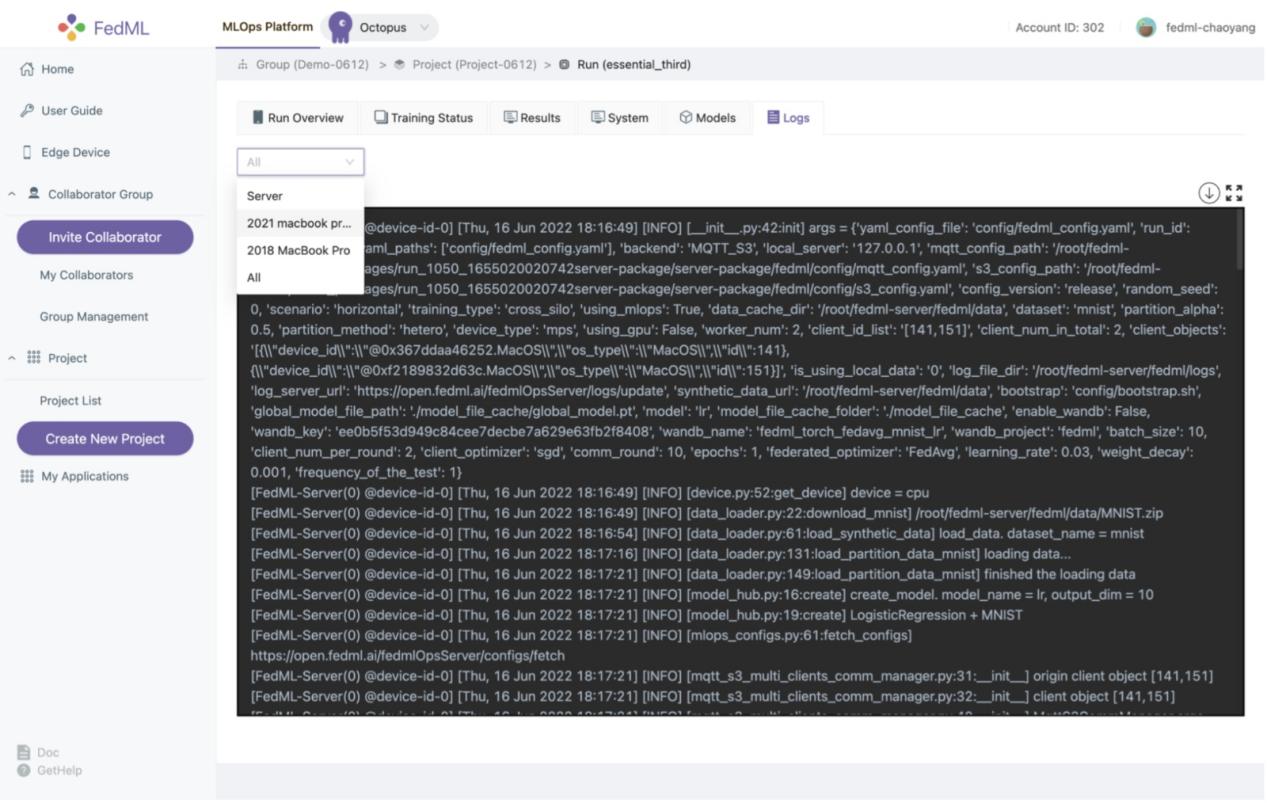
·Profiling flow与边缘设备系统性能：

这就像是给你的设备做一个体检，FedML平台可以帮你分析每个边缘设备上不同任务的执行情况，让你知道哪里可能存在问题或者瓶颈，从而进行优化。



·分布式日志：

想象一下，如果每个参与训练的设备都能记录下自己的“日记”，那么当出现问题时，你就可以通过查看这些“日记”来找出问题所在。FedML平台的分布式日志功能就是这样，它允许你实时查看每个设备上的日志，帮助分析和解决异常。



·实验报告：

有时候，可能需要比较不同训练实验的结果，看看哪个模型或者参数设置更好。FedML平台提供了实验报告功能，让你可以方便地对比不同实验的结果。

1. **极简灵活的API设计**

·简化分布式训练：

FedML通过API设计，让开发者在编写机器学习模型、数据和训练程序时，不需要关心复杂的分布式训练细节。就像写单机程序一样简单，然后把代码交给FedMLRunner这个对象，它会自动处理所有分布式训练的过程。这样大大降低了使用难度。

·灵活的API设计：

FedML认为API设计应该随着技术发展而不断迭代，不应该一成不变。他们提供了灵活的自定义API，让开发者可以根据自己的需要定义模型结构、优化器和损失函数等。这样的设计让算法创新更加灵活，并且可以和FedML的开源社区、开放平台和应用生态无缝衔接。

·核心抽象对象：

FedML提供了两个核心的抽象对象FedML.core.ClientTrainer和FedML.core.ServerAggregator，开发者只需要继承这两个对象的接口，就可以完成自定义训练和聚合过程。

·面向机器学习研发：

FedML强调，尽管安全、系统效率和模型准确性都很重要，但最重要的是让机器学习研发变得简单。他们通过分层设计，把安全、隐私和系统优化的细节隐藏起来，让机器学习开发者可以专注于机器学习本身。

·FedML Flow：

FedML Flow是一个框架，它把复杂的安全协议和分布式训练看作是有向无环图（DAG）的流式计算过程。这样，安全工程师和机器学习工程师可以各自专注于自己的领域，而不需要掌握对方的技术。

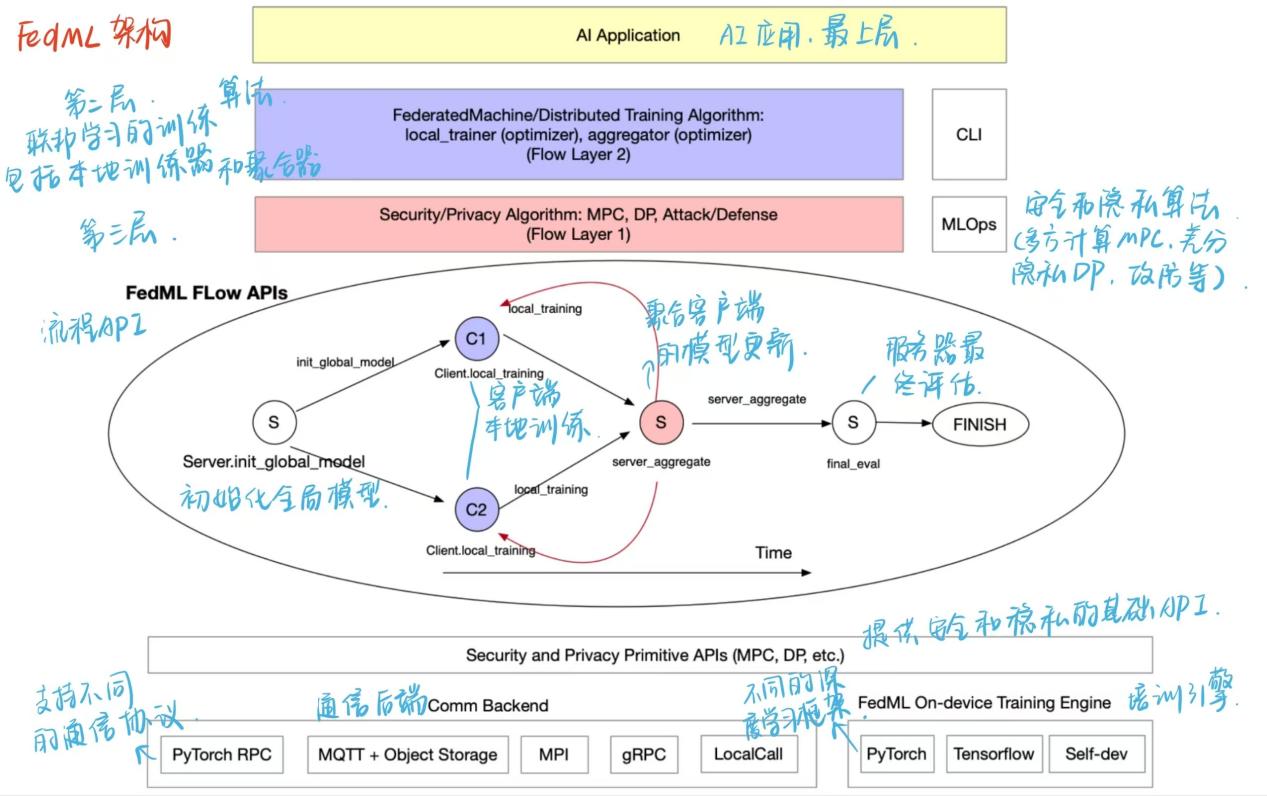
·FedMLExecutor和Flow：

FedML通过FedMLExecutor和Flow来管理和编排分布式任务。FedMLExecutor负责执行具体的任务，而Flow则负责定义这些任务的执行顺序和消息传递。开发者可以通过Flow API自由组合这些Executor，实现复杂的分布式算法，比如FedAvg。

·简化代码编写：

开发者可以在自己的电脑上编写这样的代码，而不需要掌握分布式系统的开发技能。这意味着，即使是个人开发者，也可以轻松地实现复杂的分布式机器学习算法。

总的来说，FedML通过提供简单、灵活的API和框架，让机器学习开发者可以更容易地进行算法创新和系统优化，而不需要深入了解复杂的分布式系统开发。

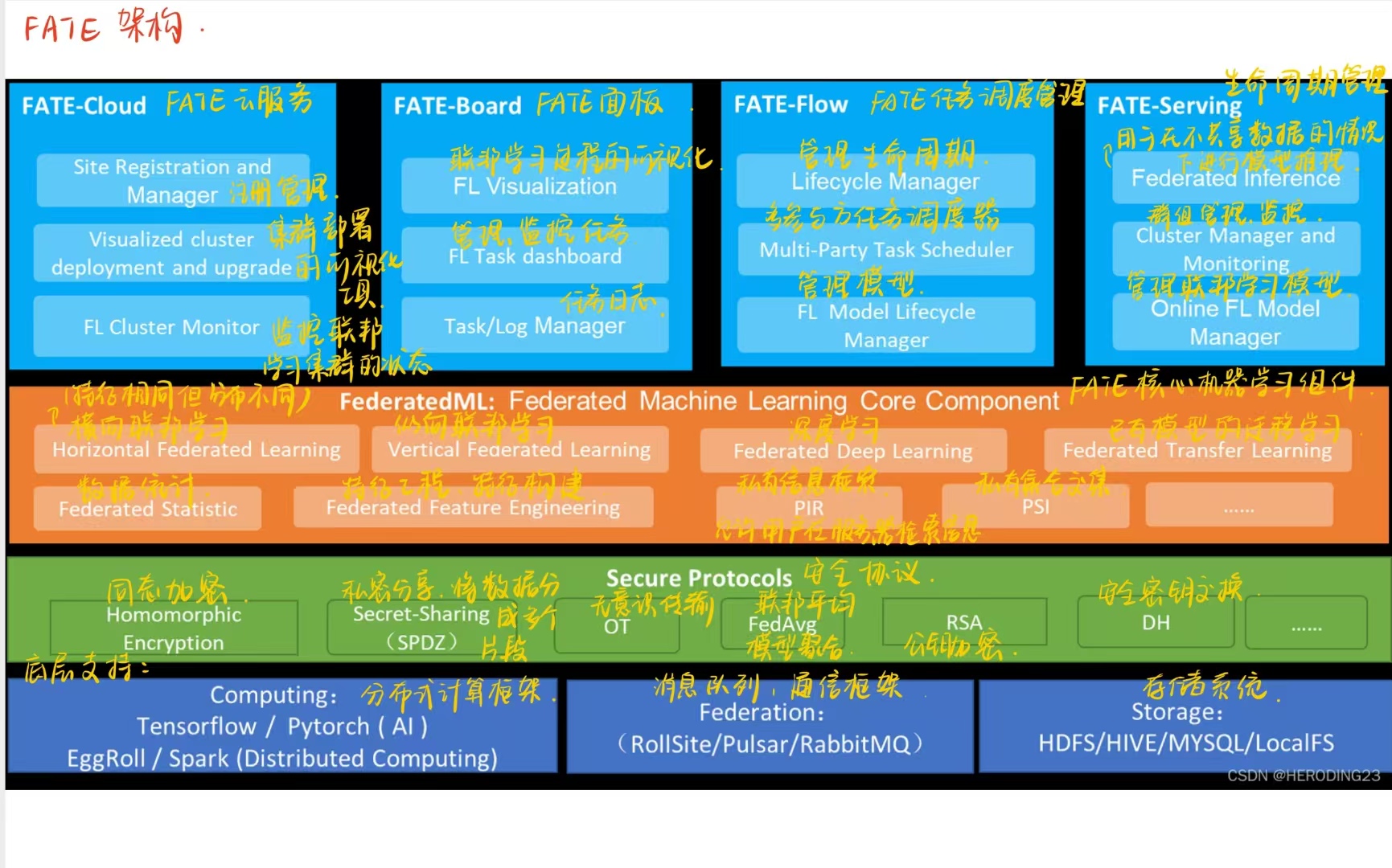


1. **FATE**

**1、FATE平台架构概述**

FATE是微众银行开发的联邦学习平台，是全球首个工业级的联邦学习开源框架，该平台为联邦学习提供了完整的生态和社区支持，为联邦学习初学者提供了很好的环境。

FATE的官方文档提供了FATE的架构图，可以看到FATE的架构自上而下分为了四层，最上面一层是FATE提供的服务，包括FATE云服务，FATE面板，FATE任务调度管理，生命周期管理等，往下一层是FATE的机器学习核心组件，包括横向联邦学习、纵向联邦学习等，再往下一层是应用在FATE中的安全协议（隐私保护算法），包括同态加密，FedAvg，RSA等，最后一层是底层框架，又可以分为平行的三层，包括计算框架（TensorFlow，Pytorch，Spark），消息队列协议，以及存储框架。

****

1. **FATE的主要功能**

·提供了一种基于数据隐私保护的分布式安全计算框架；

·为机器学习、深度学习等常用算法提供高性能的安全计算支持；

·支持同态加密、秘密共享等多种多方安全计算协议，确保数据和模型的安全；

·提供友好的跨域交互信息管理方案和开发文档，极大方便开发人员使用

1. **TensorFlow-federated**

TFF，全称TensorFlow Federated，是谷歌开源的一个联邦学习平台架构。它基于我们熟知的TensorFlow深度学习框架构建，专门用于处理分散在不同设备上的数据，比如你的安卓手机。TFF的设计初衷是为了在保护用户隐私的同时，让多个设备能够协同工作，共同训练出一个机器学习模型。TFF并不是一个独立的联邦学习生态体系，它更像是TensorFlow的一个扩展，专门针对联邦学习场景进行了优化。通过TFF，开发者可以利用TensorFlow的强大功能，同时实现联邦学习的独特需求。

TensorFlow-federated的三种不同的数据处理和存储方式：

1. 本地数据处理：

• 图中显示了一个手持设备（可能是手机或平板电脑），它在本地进行数据处理。

• 数据（绿色方块）在设备上被处理，然后结果（蓝色方块）被发送到服务器或云端。

• 这种方式强调了数据在本地处理的隐私保护，因为原始数据不需要离开用户的设备。

1. 集中式数据处理：
2. • 多个设备（手机）将数据（绿色方块）发送到一个中心服务器。

• 在服务器上，数据被集中处理，可能涉及到数据的整合和分析。

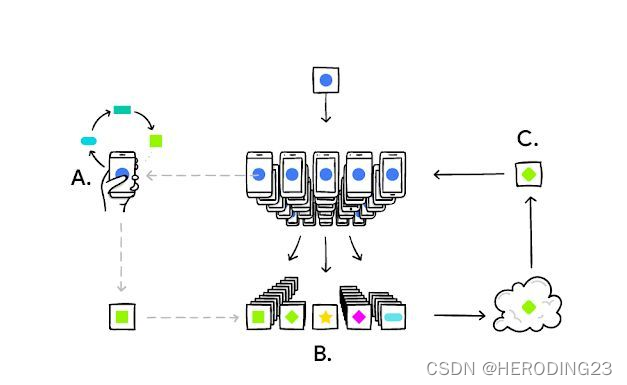
• 这种方式在传统的云计算中很常见，所有数据都被发送到云端进行处理。

1. 云端数据处理：

• 数据从设备（绿色方块）发送到云端。

• 在云端，数据被处理（可能是存储、分析或其他操作），然后结果被返回到设备。

• 这种方式强调了云计算的灵活性和可扩展性，但可能涉及到数据隐私和安全的问题，因为数据需要被传输到远程服务器。



功能方面，用户可以通过TensorFlow-federated实现如下功能：

·通过Federated Learning API与TensorFlow/Keras交互，完成分类、回归等任务：在不共享原始数据的情况下，完成数据分类和预测等任务。这在保护用户隐私的同时，还能让模型学习到有用的信息。

·通过TFF提供的Federated Core API，在强类型函数编程环境中将TensorFlow与分布式通信运算符相结合，表达新的联邦算法：TFF提供的Federated Core API，就像是给程序员提供了一套强大的工具，让他们能够在一个严格的编程环境中，将TensorFlow的计算能力与分布式通信运算符结合起来，创造出新的联邦学习算法。

· TFF使用的数据隐私保护主要是差分隐私，安全机制在TensorFlow/Privacy中实现，开发者可以自行调用安全算法：TFF使用的数据隐私保护主要是差分隐私，这是一种通过添加一些统计噪声来保护个人数据隐私的技术。安全机制在TensorFlow/Privacy中实现，这意味着开发者可以在这个框架中直接调用安全算法。

1. **总结**

总结以上几个平台，可以大致确认平台应该拥有的一些功能：

·边缘设备的管理部署、模型训练、绑定客户端代理

·服务器端的模型聚合

·客户端与服务器端交互过程的隐私安全问题