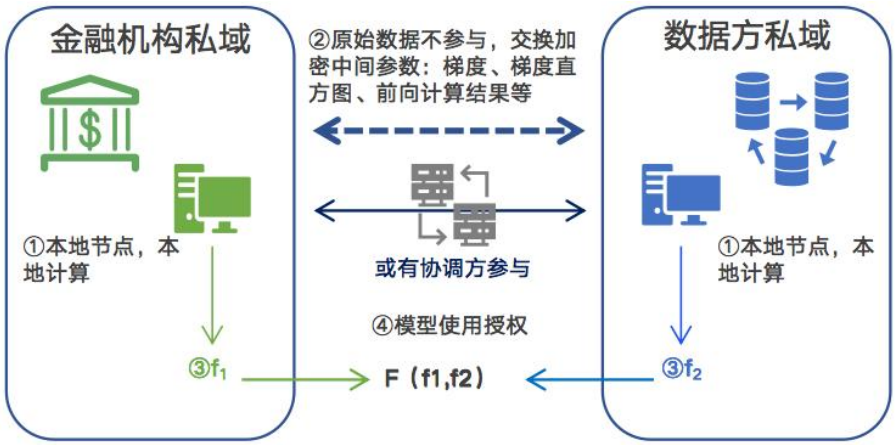
联邦学习
Federated Learning

联邦学习，又名联邦机器学习、联合学习、联盟学习，被认识是解决数据安全与人工智能两难问题的一个重要技术。联邦学习是一种隐私保护的分布式机器学习框架，各个参与方在保证各自原始数据不出域的前提下，通过交互中间数据，协作完成某项机器学习任务.

联邦学习能有效帮助多个机构在满足用户隐私保护、数据安全和政府法规的要求下，进行数据使用和机器学习建模。联邦学习作为分布式的机器学习范式,可以有效解决数据孤岛问题,让参与方在不交换原始数据的基础上联合建模,能从技术上打破数据孤岛,实现 AI 协作。



联邦学习主要分为三种类型：横向联邦学习、纵向联邦学习、联邦迁移学习：

(1) 横向联邦学习：在两个数据集的用户特征重叠较多而用户重叠较少的情况下，把数据集按照横向(即用户维度)切分，并取出双方用户特征相同而用户不完全相同的那部分数据进行训练。

(2) 纵向联邦学习：在两个数据集的用户重叠较多，而用户特征重叠较少的情况下，把数据集按照纵向（即特征维度）切分，并取出双方用户相同而用户特征不完全相同的那部分数据进行训练。

(3) 联邦迁移学习：在两个数据集的用户与用户特征重叠都较少的情况下，不对数据进行切分，而可以利用迁移学习来克服样本和标签不足的情况。

联邦学习在如下两种情况可以很好解决企业的人工智能应用难题：一是涉及到保护数据隐私和核心价值的场景，因为联邦学习的整个学习训练过程，没有传输任何原始数据；二是多方数据补充的场景，这可能存在单方样品数量不够充分或单方数据维度不够丰富的情况。例如在金融风控场景中，银行希望引入外部数据源做特征补充来建立联合模型。基于用户授权，联邦学习技术可以在保证数据安全不出库的同时，整合不同机构间对用户行为特征不同维度的捕捉，以用户为基础，形成对个人的较为全面的描述。对比传统模型方式，模型可以学到更多用户信息，从而提升模型效果，促进业务发展，实现降本增效。