



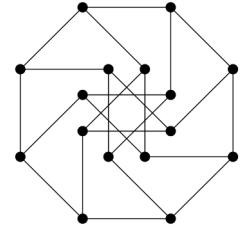
云原生联邦学习平台、实践与应用

彭麟 (Layne Peng)

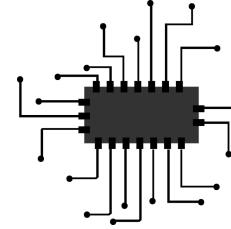
资深研究员

VMware – CTO办公室 – 云原生实验室

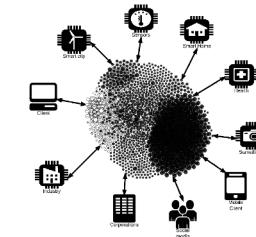
人工智能的三大要素



算法

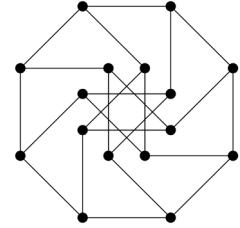


算力



数据

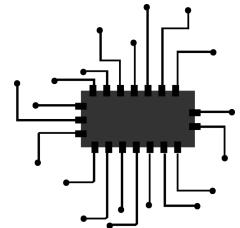
数据的现状并不理想



算法



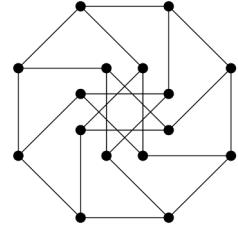
数据



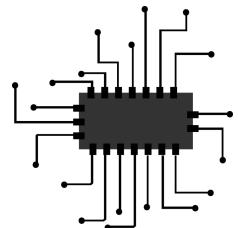
算力

数据孤岛
数据分布不均
缺少标注数据

数据的现状并不理想



算法



算力



数据

数据孤岛
数据分布不均
缺少标注数据

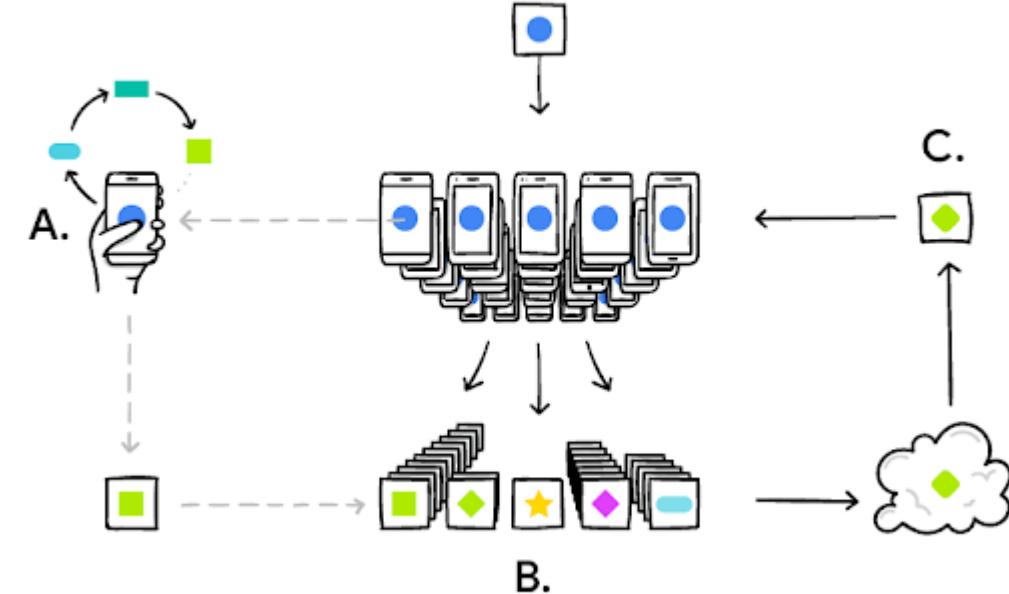
- 制造数据: GAN
- 利用公有(public)和开放(open)数据: 迁移学习
- 私有数据方合作一起训练: 联邦学习 (Federated learning)

联邦学习概念出现



数据

- 数据孤岛
- 数据分布不均
- 缺少标注数据



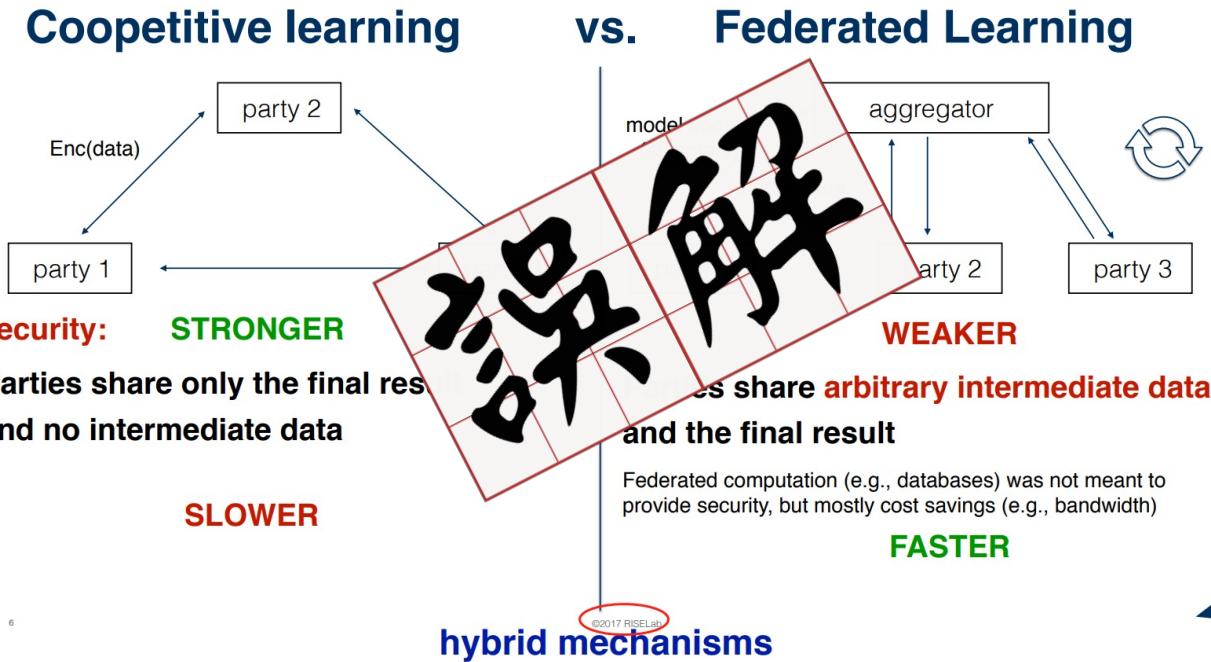
(Source: Federated Learning: Collaborative Machine Learning without Centralized Training Data, Google AI Blog, 2017)

联邦学习的误解：无隐私保护



数据

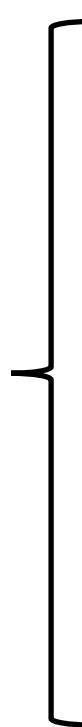
- 数据孤岛
- 数据分布不均
- 缺少标注数据



(Source: Secure Collaborative Learning, 2017)



数据



数据孤岛

数据分布不均

缺少标注数据

隐私法律法规

数据安全

◦ ◦ ◦



- 联邦学习 (Federated learning) => (安全&保护隐私的) 联邦学习

联邦学习的定义



数据

数据孤岛

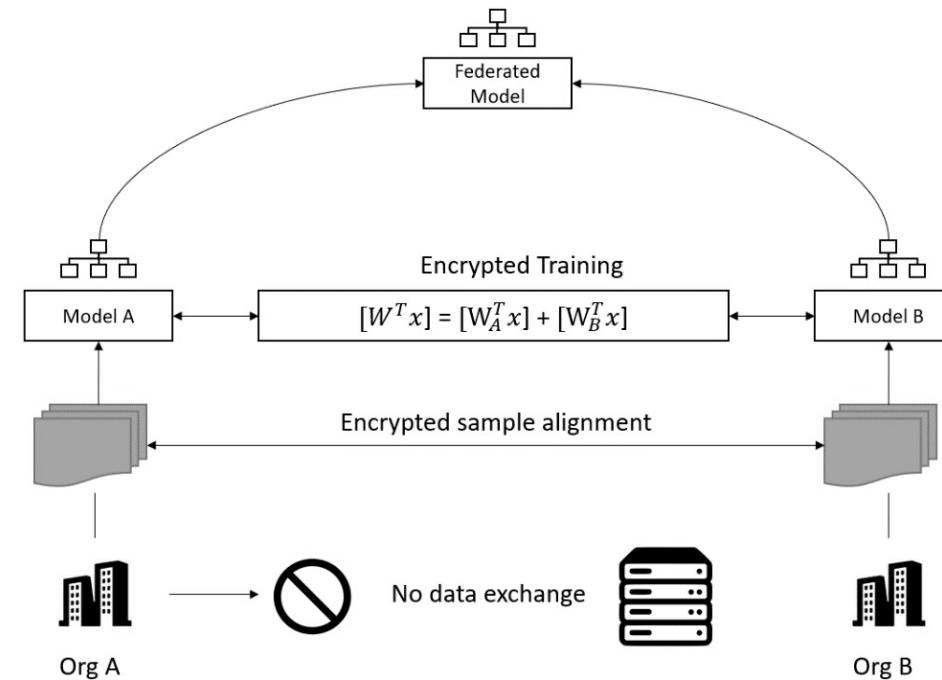
数据分布不均

缺少标注数据

隐私法律法规

数据安全

。 。 。

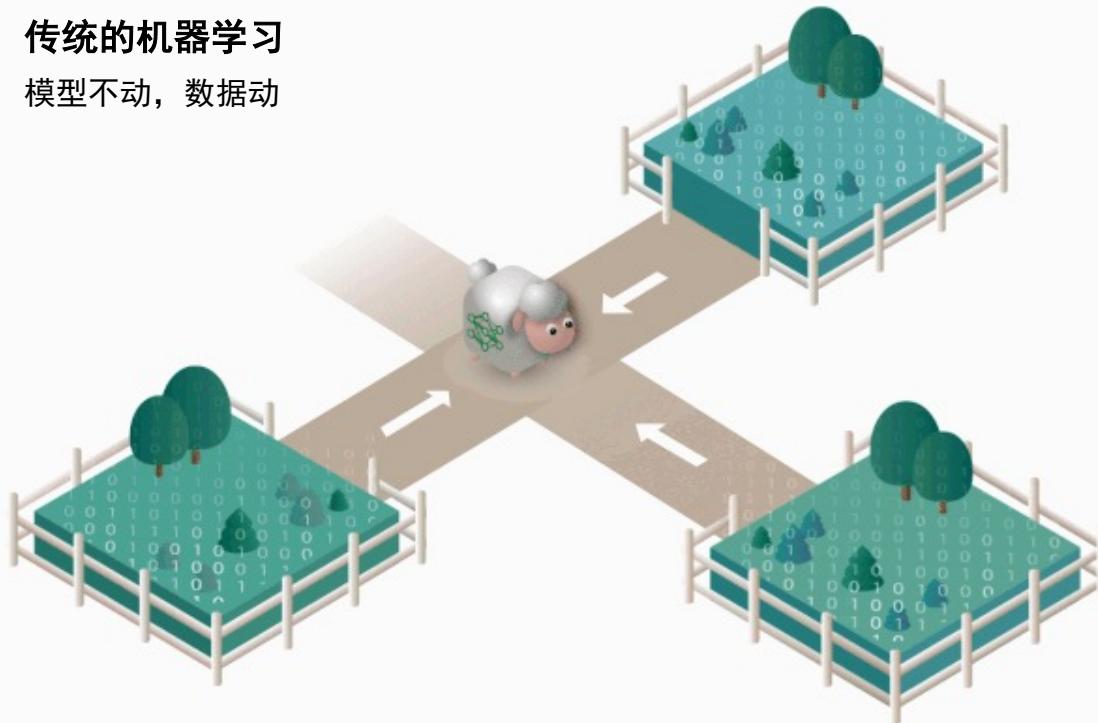


- 两个或更多的（子）组织共同训练模型
- 组织间无数据交换
- 加密模型在多方安全计算框架下共同训练：
 - 同态加密
 - 共享密钥
 - 不经意传输
 - ...

联邦学习与传统的机器学习

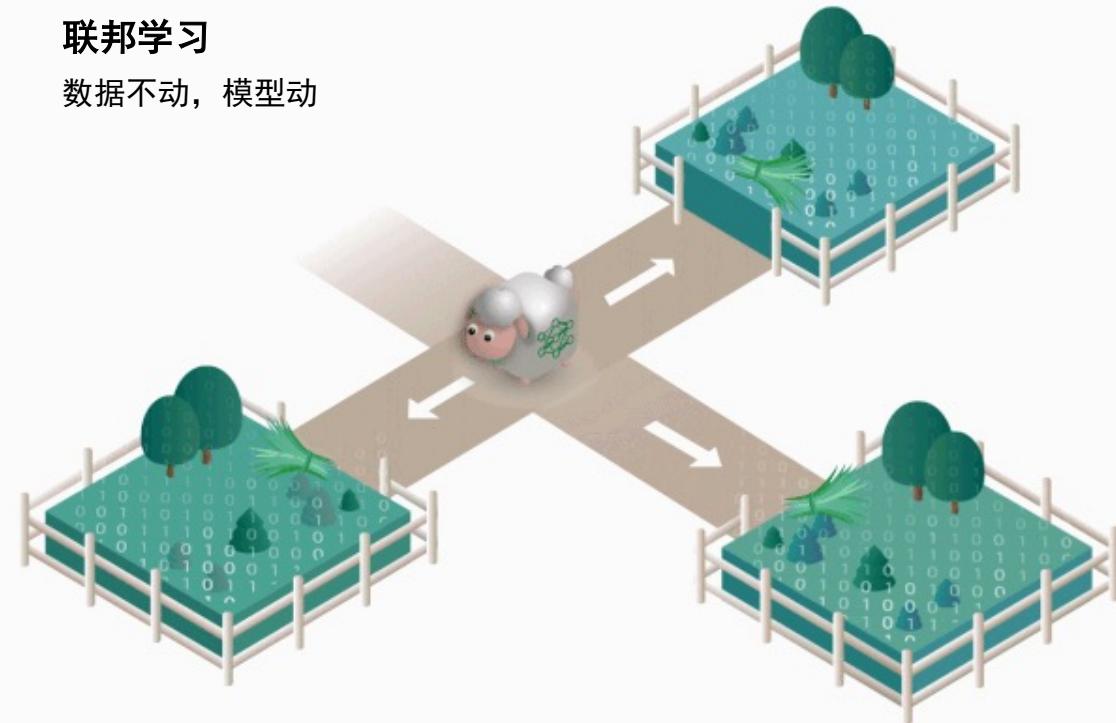
传统的机器学习

模型不动，数据动



联邦学习

数据不动，模型动



(Source: Federated Learning (Synthesis Lectures on Artificial Intelligence and Machine Learning) , Qiang yang , et al.)

数据不动模型动，数据可用不可见

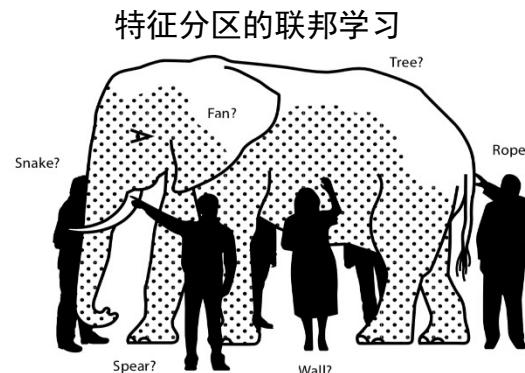
联邦学习的分类

数据孤岛情况 1: 样例分散在不同的组织, 单个组织样例不足以支持优质训练。。



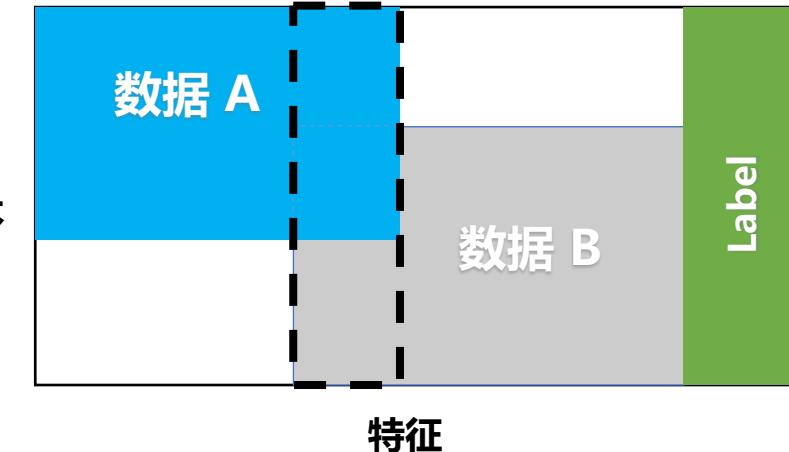
(Source: <https://ai.googleblog.com/2017/04/federated-learning-collaborative.html>)

数据孤岛情况 2: 样本数据的特征分散在不同组织, 单个组织有样本片面的理解, 造成训练结果偏差。。

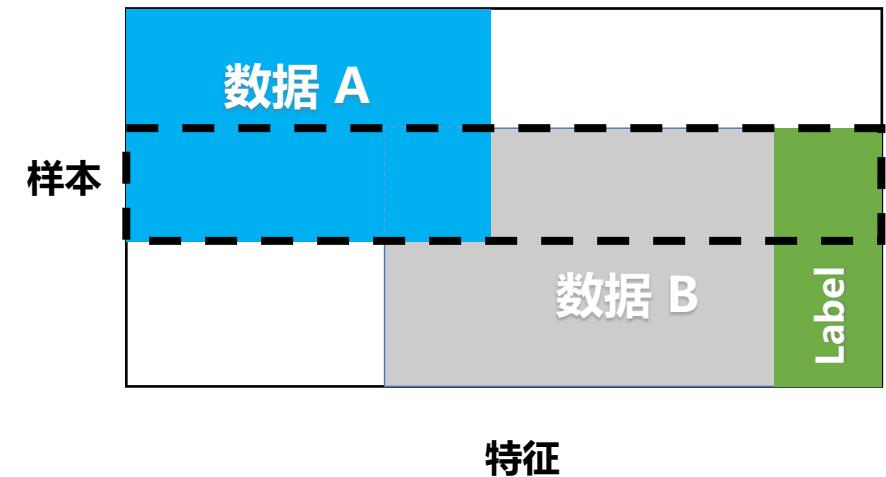


(Source: 中国寓言, 盲人摸象)

横向联邦学习/同构联邦学习

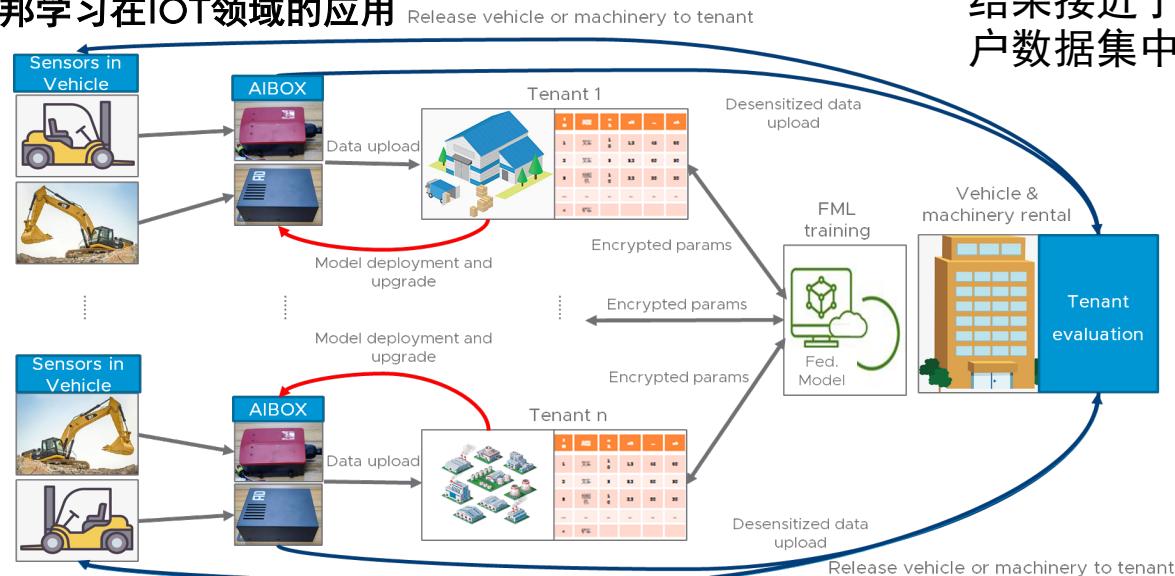


纵向联邦学习/异构联邦学习



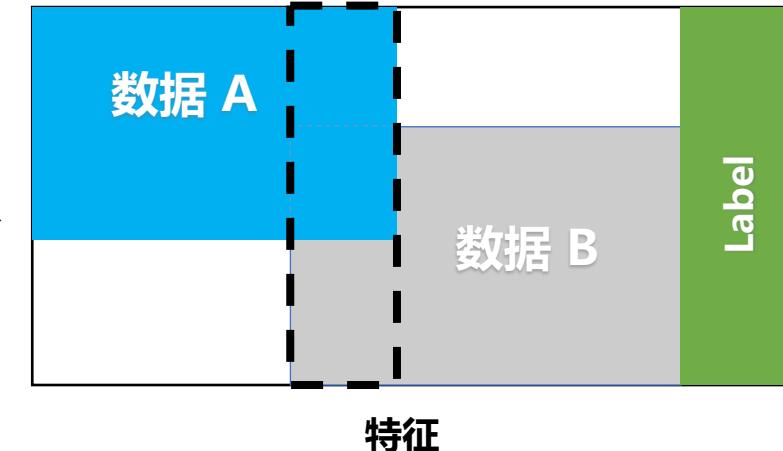
联邦学习的分类

横向联邦学习在IOT领域的应用



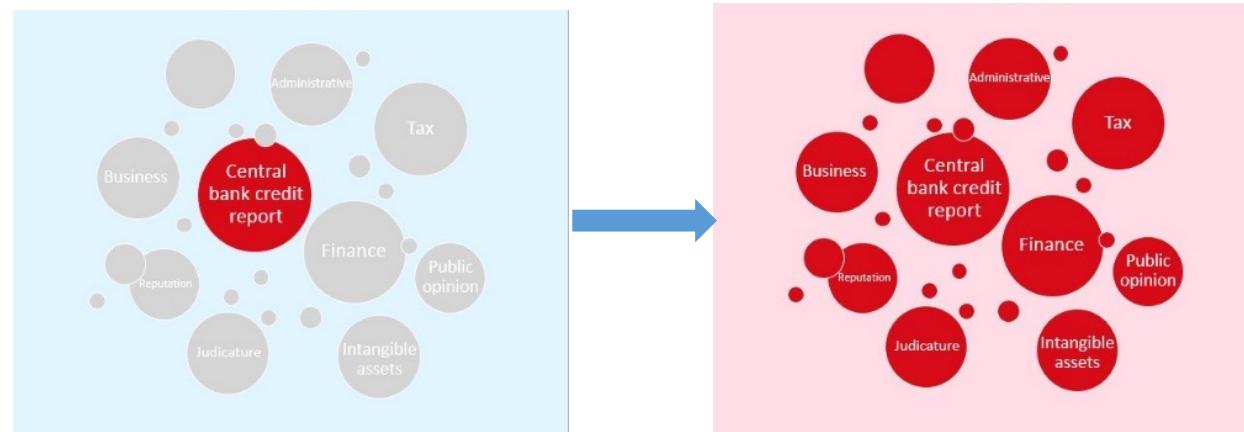
结果接近于把所有租户数据集中训练

横向联邦学习/同构联邦学习

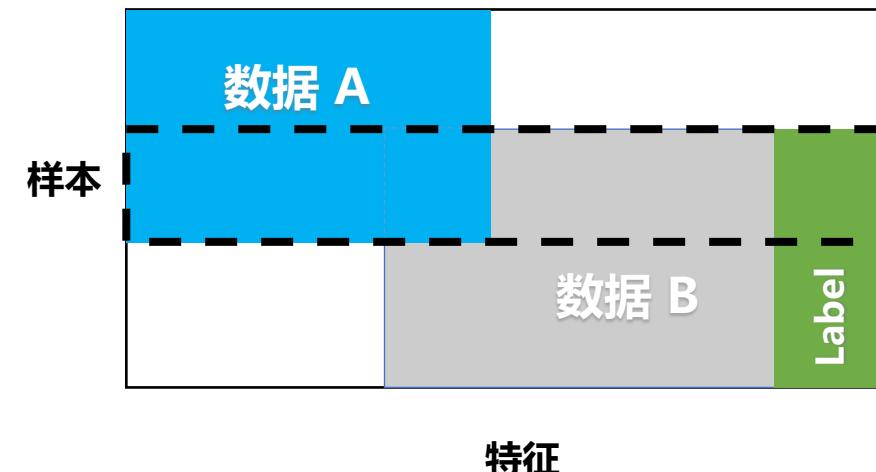


小微企业信用风险管理

多元数据来源组合获得更准确的用户画像



纵向联邦学习/异构联邦学习



FATE: Federated AI Technology Enabler



FATE: Federated AI Technology Enabler



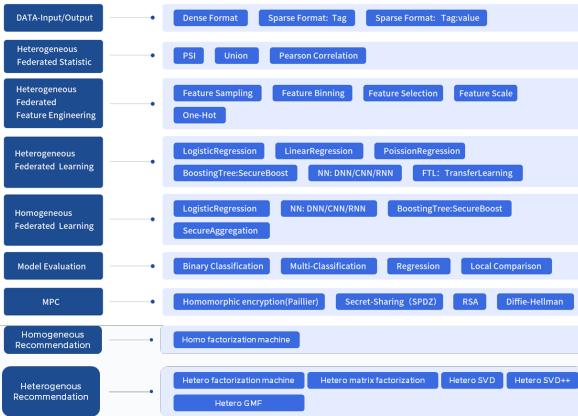
FATE是开箱即用的工业级联邦学习平台：

1. 内置典型的联邦学算法；
2. 可视化建模界面；
3. DAG工作流引擎；
4. 支持多种多方计算安全协议：同态加密、共享密钥，etc.
5. 支持审计等功能，满足银监等保要求；
6. 分布式计算、存储、传输引擎；
7. 支持异构加速器。



A permissive license whose main conditions require preservation of copyright and license notices.
Contributors provide an express grant of patent rights. Licensed works, modifications, and larger works
may be distributed under different terms and without source code.

人工智能x 云原生分论坛



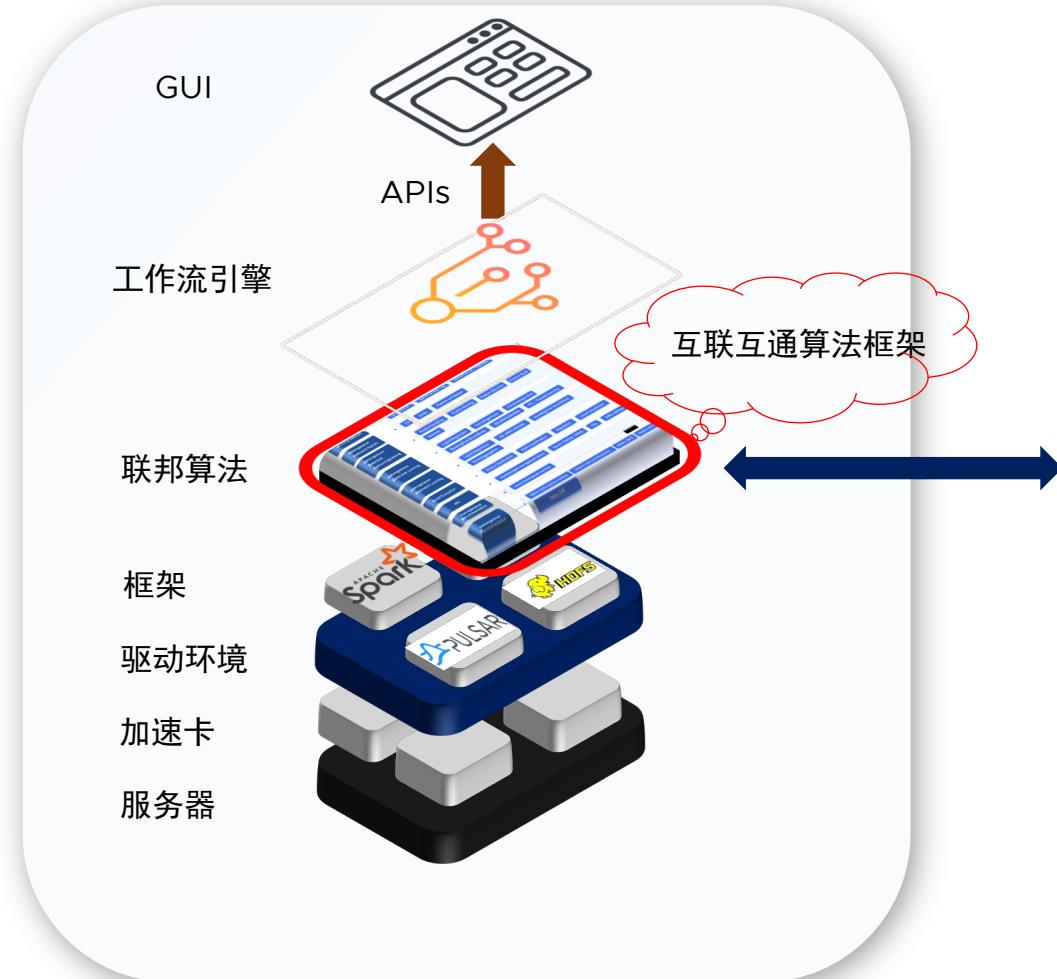
1. 开箱即用的算法；
2. 联邦学习算法开发框架：
 - a) 底层工具
 - b) 通信协议引擎
 - c) 工作流引擎
 - d) 互联互通协议
 - e) 算法编译器



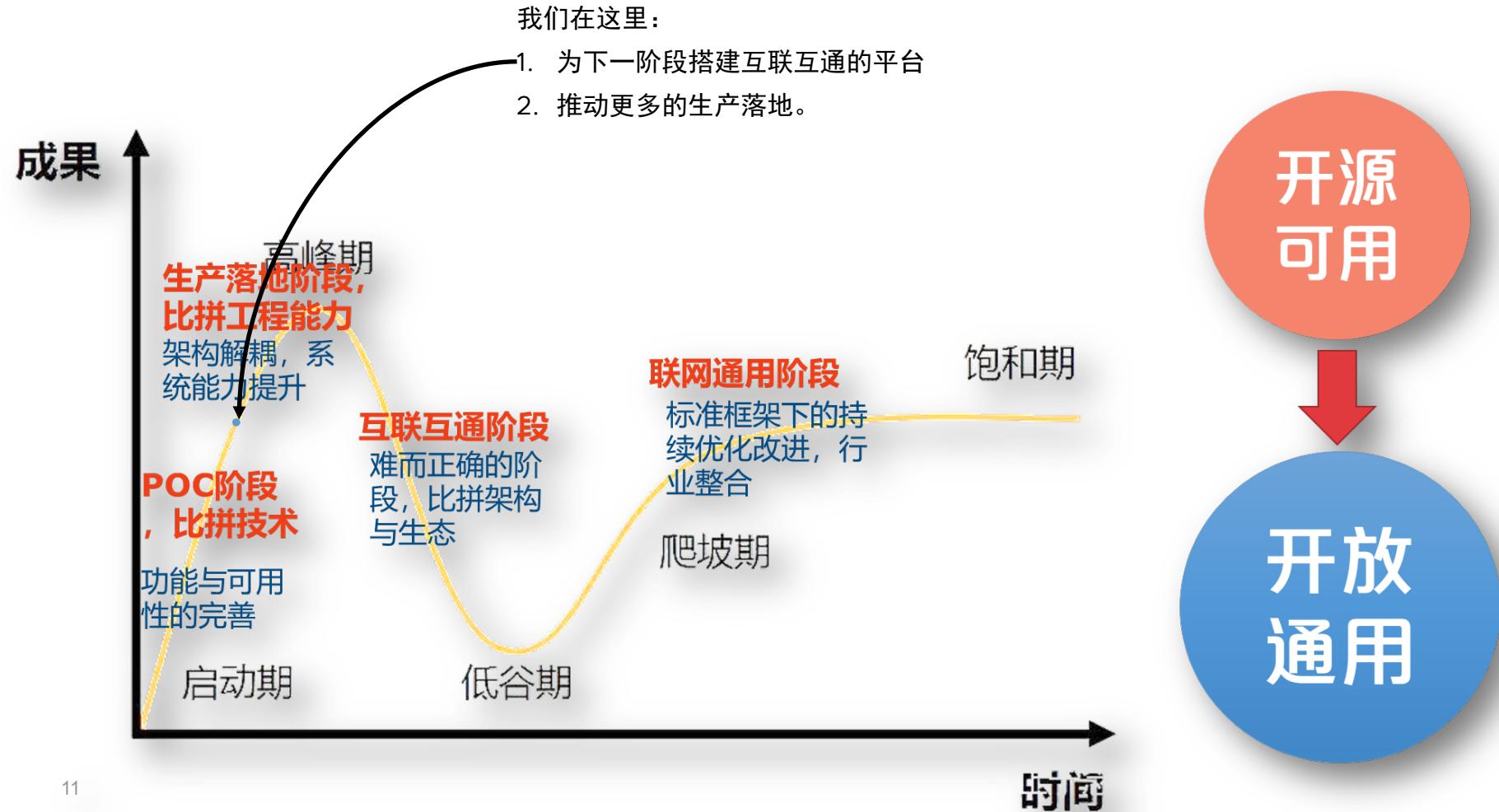
1. 重用已有算力：支持开源计算、传输、存储框架
 - a) Spark
 - b) Pulsar/RabbitMQ
 - c) HDFS
 - d) Hive
 - e) ...
2. 异构加速器：
 - a) GPU
 - b) FPGA
 - c) ARM

FATE: Federated AI Technology Enabler v1.7.0

FATE v1.7.0是一个联邦学习的生态系统（FedAI）



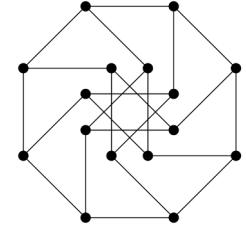
Source: [破解不同技术平台交互阻碍，「富数科技」和「微众银行」实现异构联邦学习平台互通](#)



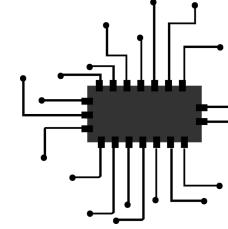
11

Source:企业级联邦学习平台建设的探索与思考，中国银联金融科技研究院，周雍恺

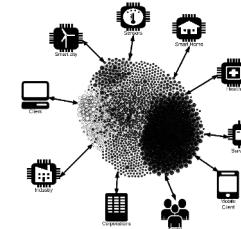
人工智能第四要素



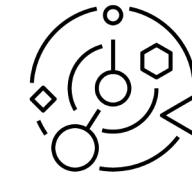
算法



算力

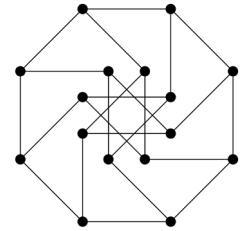


数据

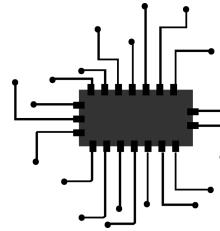


运维

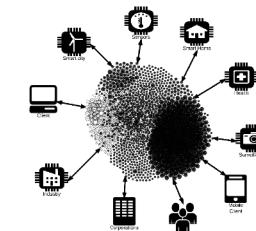
人工智能第四要素



算法



算力



数据



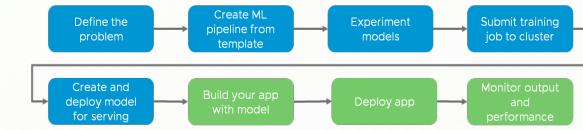
可插拔



云原生联邦学习



可扩展



全生命周期管理



安全

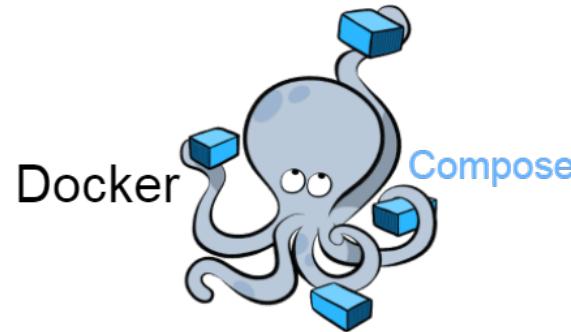


管理

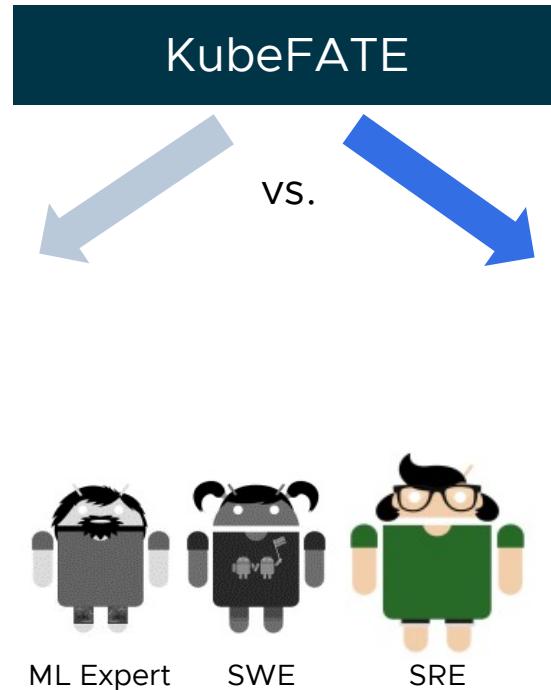


高可用

KubeFATE: 云原生联邦学习平台

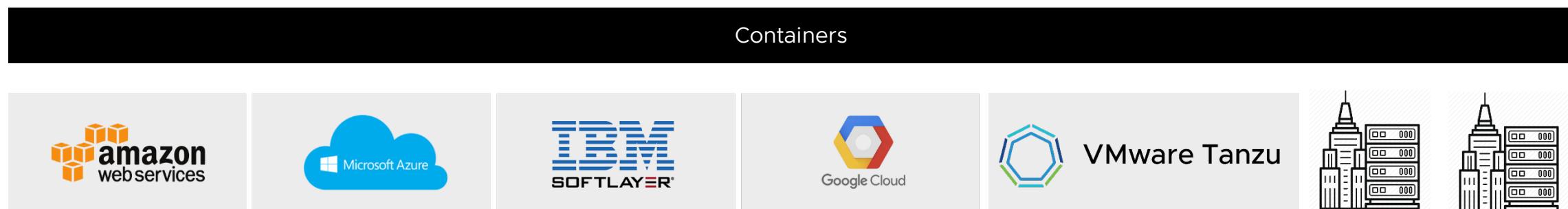


1. 测试、体验多方FATE集群；
2. 上手简单。

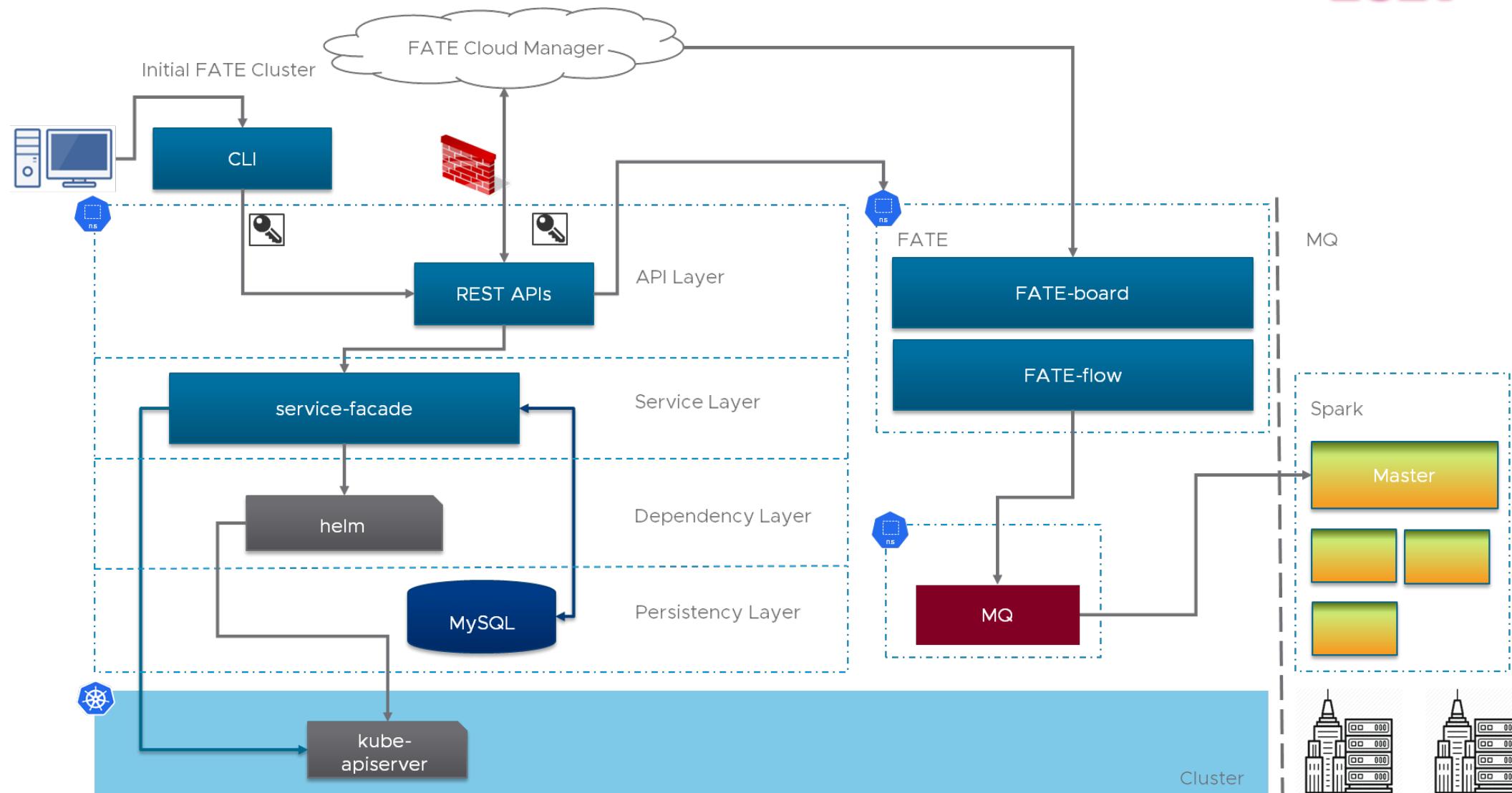


kubernetes

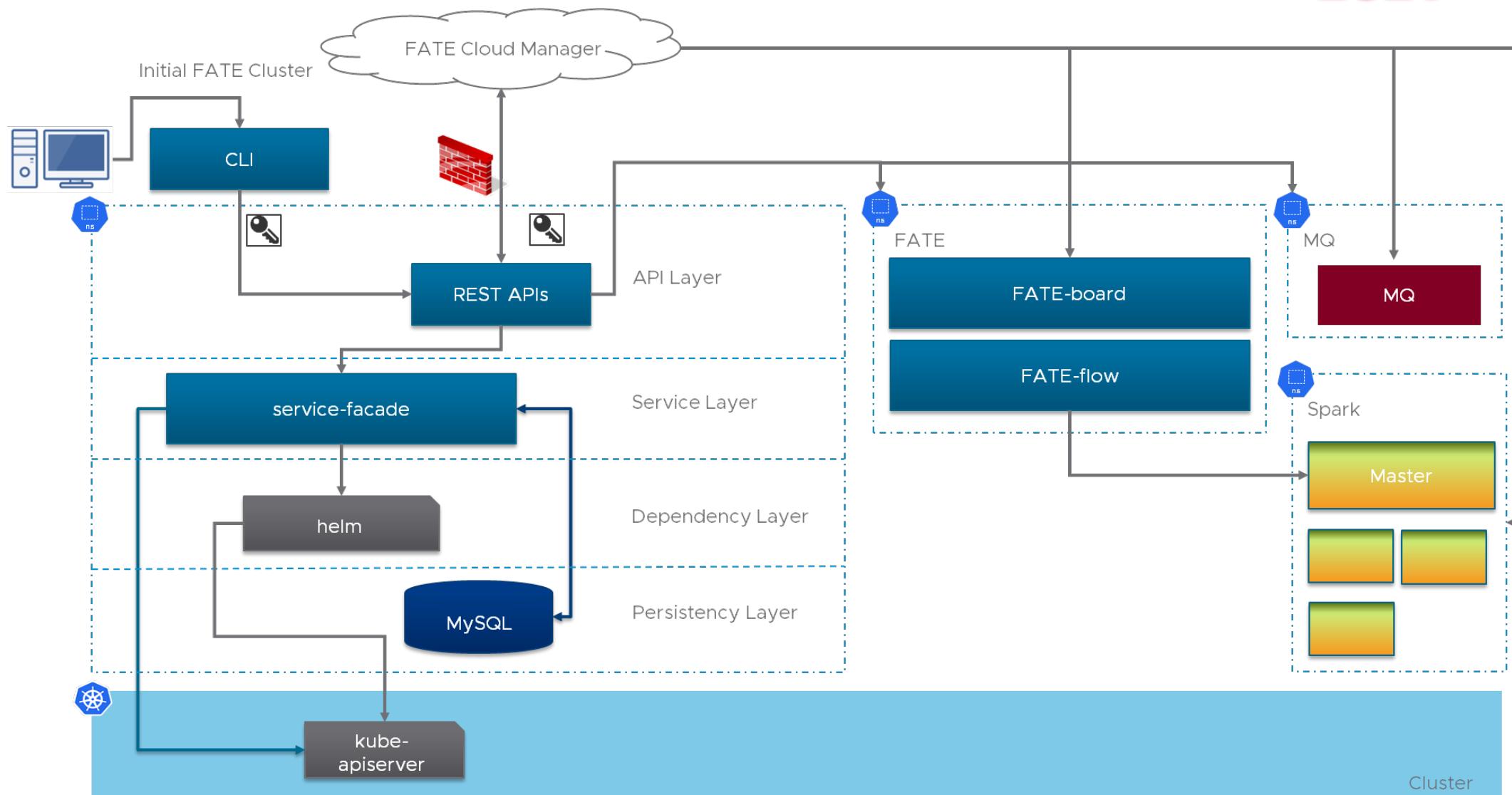
1. 面向生产环境：
 - 1) 支持多个FATE环境及集群；
 - 2) 声明式扩展能力；
 - 3) 升级，迁移；
 - 4) 日志及监控功能
2. 强大的定制功能



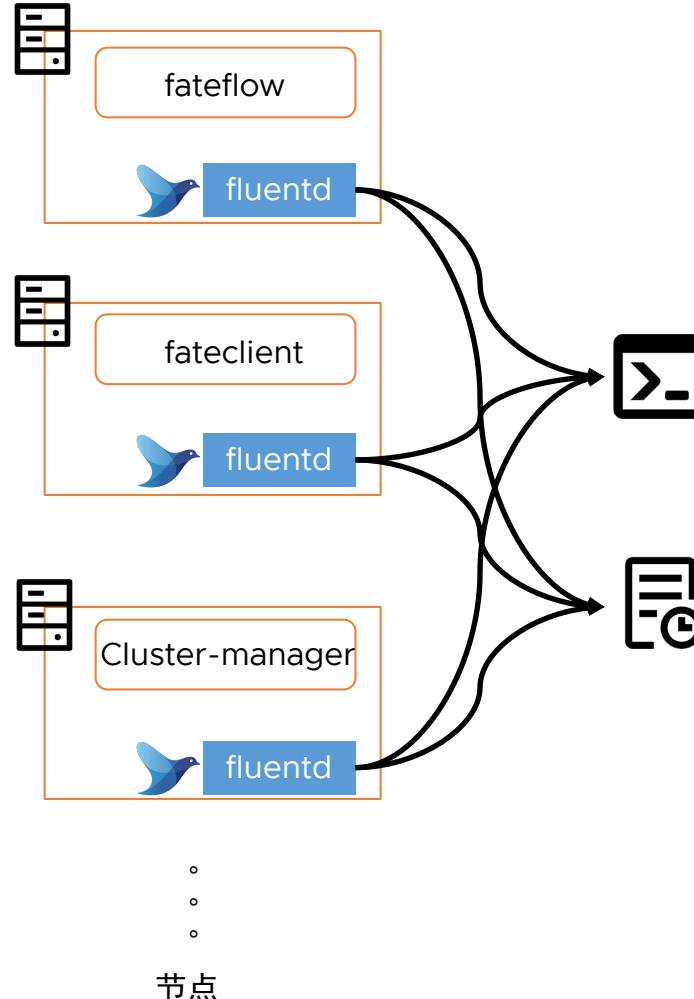
KubeFATE: 架构、模块



KubeFATE: 定制化部署



KubeFATE: 云原生管理功能 – e.g., 日志汇集



```
kubefate cluster logs [options] <cluster_ID> [modules_name]
```

[options] 是命令的选项

<cluster_ID> 是指定FATE集群的ID（必选）

[modules_name] FATE对应的模块组件

持续监控集群所有组件的日志

```
kubefate cluster logs -f 8b980f0b-b139-40b2-a94d-d5aebd14d913
```

监控python组件的日志

```
kubefate cluster logs -f 8b980f0b-b139-40b2-a94d-d5aebd14d913 python
```

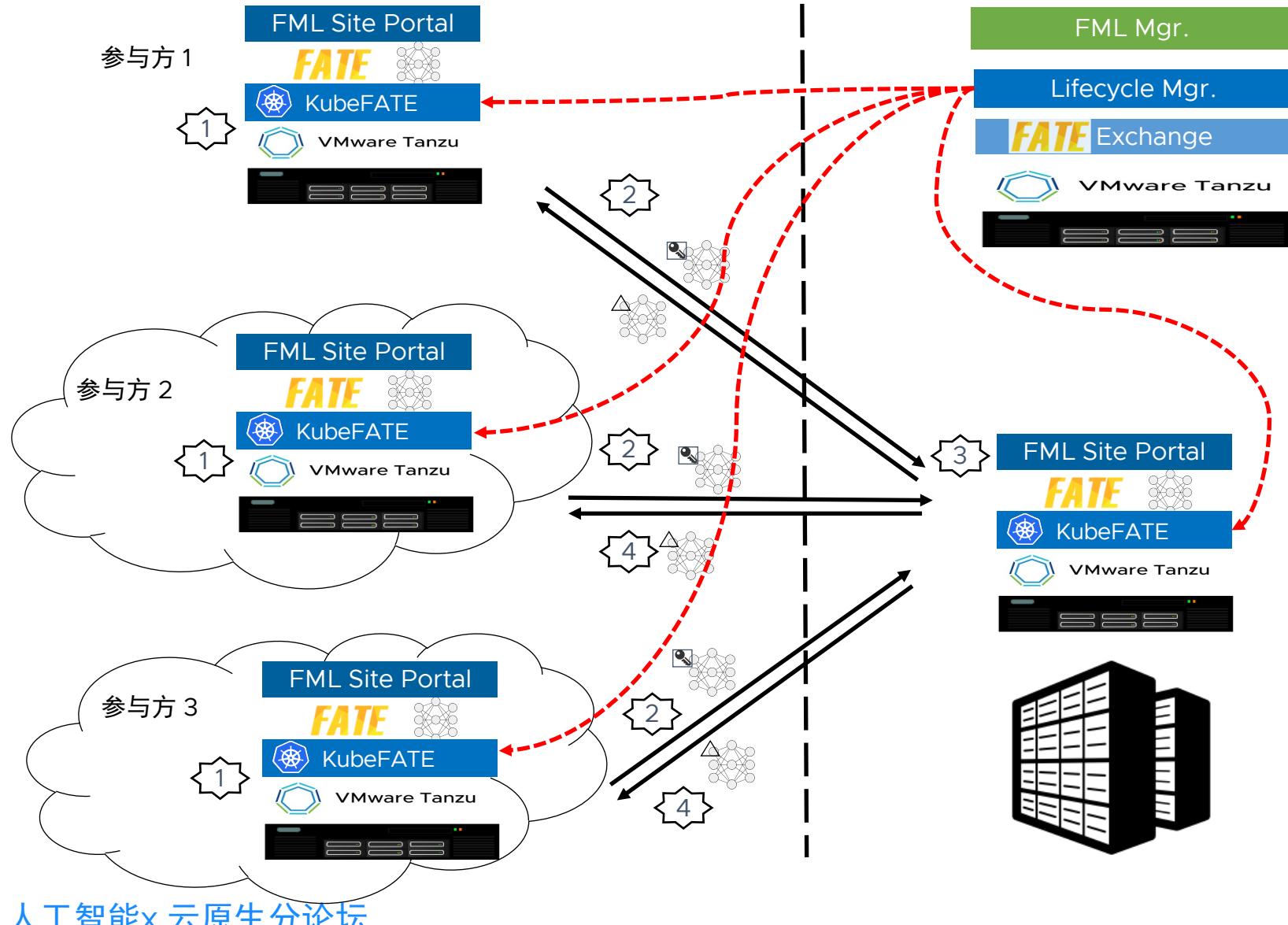
监控错误日志

```
kubefate cluster logs 8b980f0b-b139-40b2-a94d-d5aebd14d913 | grep ERROR
```

查看单个任务日志

```
kubefate cluster logs b4db45a6-e9b5-4350-8be3-511ea72c76cf | grep <Job_ID>
```

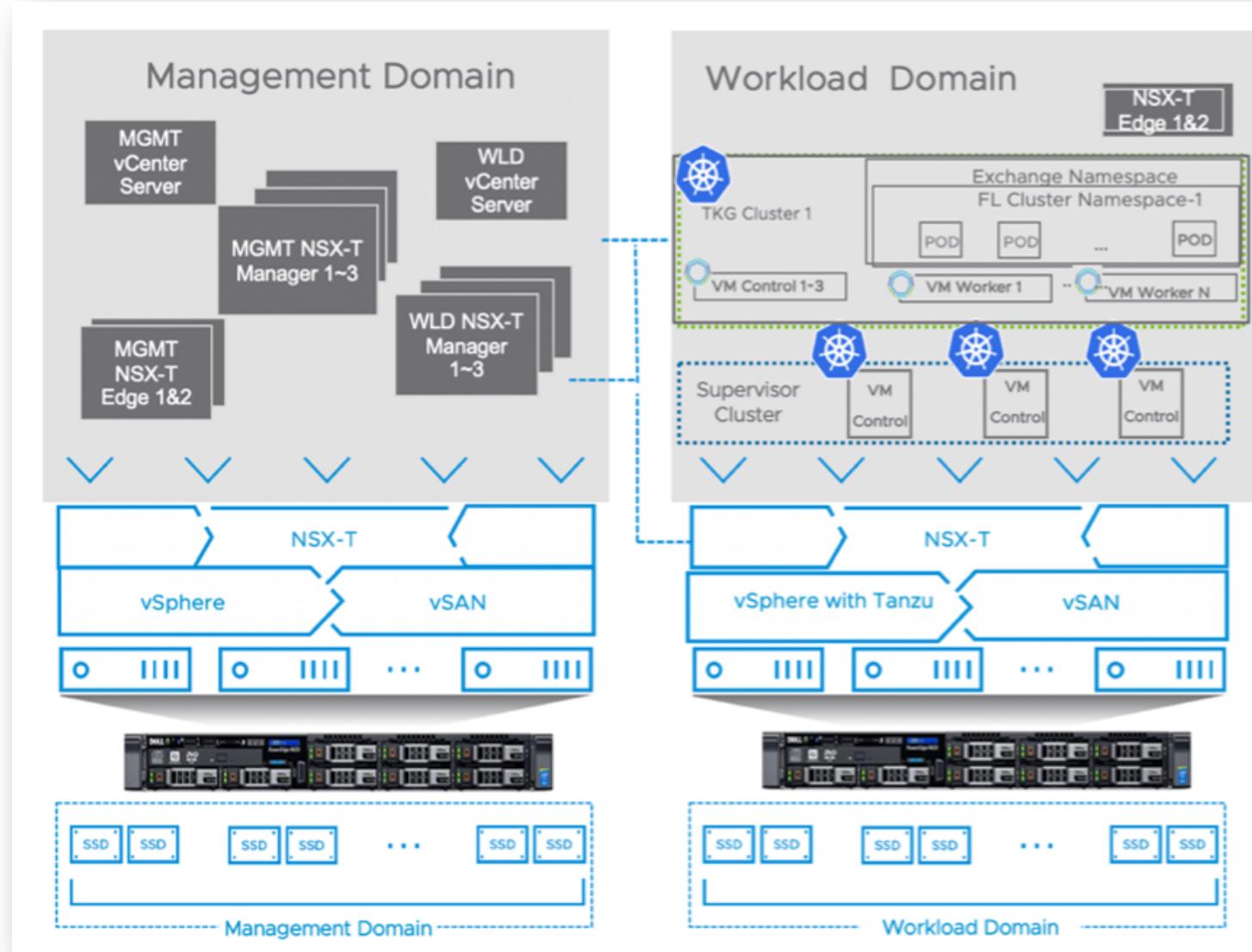
KubeFATE: FATE+VCF企业级方案



KubeFATE: FATE+VCF企业级方案

原文: KubeFATE on VMware Cloud Foundation with VMware Tanzu (<https://core.vmware.com/resource/kubefate-vmware-cloud-foundation-vmware-tanzu>)

- 技术概览
- 解决方案配置
- 解决方案验证
 - 性能测试
 - 错误验证
- 配置要求建议
- 用例



总结

1. 联邦学习是解决小数据、数据孤岛的可行方案，核心是“数据不动模型动”；
2. FATE是面向生产，有诸多成功案例的开源、开放联邦学习平台；
3. 联邦学习的复杂性提出了运维的需求，为此我们提出云原生联邦学习的概念，并开源：
KubeFATE: <https://github.com/FederatedAI/KubeFATE>
4. VMware作为FATE的主要贡献者之一，结合本身产品栈，推出FATE+VCF的企业级方案。



VMware中国研发中心



回复“kubefate”加入KubeFATE交流群



FATE联邦学习技术交流群



Thank You