Open Source AceCon

2021 智能云边开源峰会

AI x Cloud Native x Edge Computing

人工智能 × 云原生 × 边缘计算

云原生隐私计算平台

Henry Zhang

Cloud Native Lab, OCTO China, VMware

Cloud Native Lab

- Focus on cloud native leadership and technologies
- Incubates solutions to add values to modern application platform

Ecosystem

KubeCon+CloudNativeCon

CCF Tech Frontier

China Open Source Summit

Cloud Native Industry Summit

Projects

HARBOR FATE KubeFlow PySyft

Co-innovators

腾讯云

京东

中国移动 China Mobile

滴滴 滴滴一下 美好出行

WeBank

Alibaba Cloud

CloudChef 骞云科技

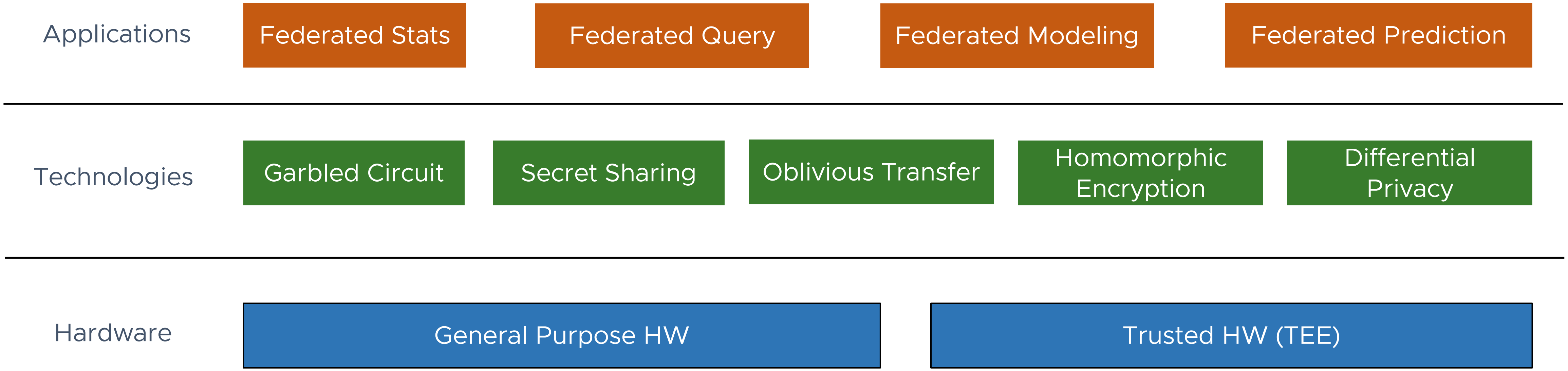
caicloud 才云

灵雀云 alauda.cn

ise2c

Privacy preserving computation

- Usable but not visible: make use of the data without leaking the data
- Three categories: MPC, federated learning, TEE
- Privacy-enhancing computation a top strategic technology trend for 2021 – Gartner*



Applications: Federated Stats | Federated Query | Federated Modeling | Federated Prediction

Technologies: Garbled Circuit | Secret Sharing | Oblivious Transfer | Homomorphic Encryption | Differential Privacy

Hardware: General Purpose HW | Trusted HW (TEE)

* https://www.gartner.com/en/newsroom/press-releases/2020-10-19-gartner-identifies-the-top-strategic-technology-trends-for-2021

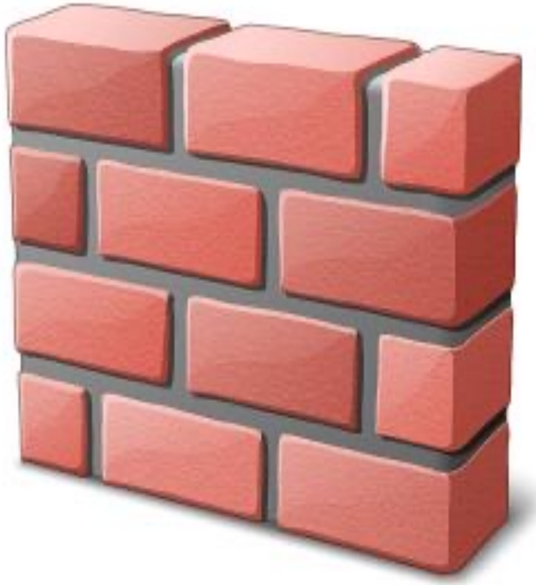
Federated Learning = AI + Privacy Preserving Computation

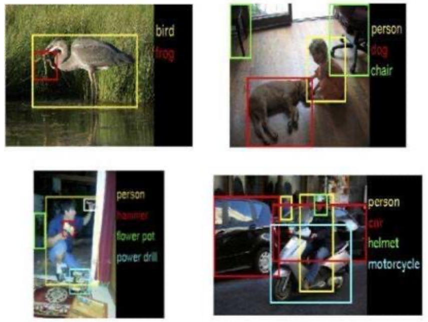
- FML is the top 9 AI trends in 2020 – CBInsights*
- With the growing interest of FML, customers need to have a way to leverage data from different organizations while preserving data privacy.
- Customers need to run FML workload across multiple cloud, either on-prem or in public cloud.

* https://www.cbinsights.com/research/report/ai-trends-2020/

Challenges to AI: small data set and fragmented data

Enterprise A





Enterprise B

- Security in data sharing
- Lack of labeled data
- Isolated datasets

Over 80% of enterprises have information silos!

Challenges to AI: Data Privacy and Confidentiality

General Data Protection Regulation (GDPR)

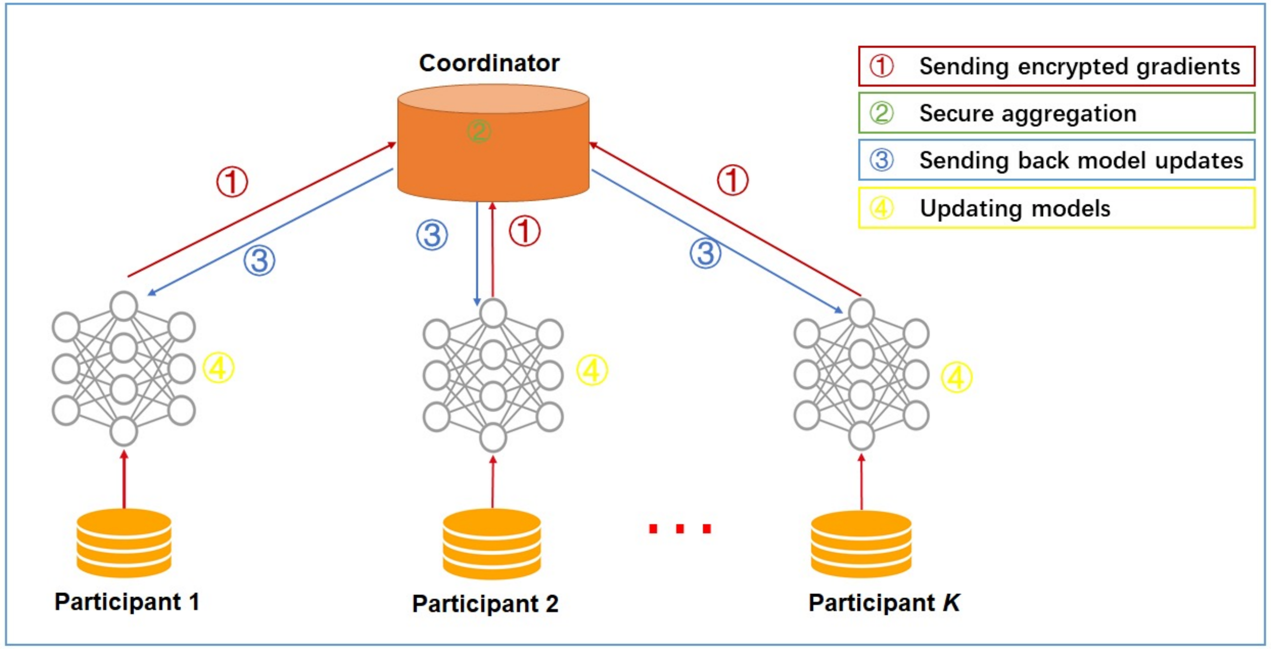
California Consumer Privacy Act (CCPA)

China’s Data Security Law

Personal Information Protection Act

Federated Machine Learning for Preserving Data Privacy

- Participants co-build an FML model
- Exchange no raw data to preserve privacy
- Model is lossless



Qiang Yang, et al., *Federated machine learning: Concept and Applications*, 2019.

Cross-Bank Federation

ISP-Bank Federation

ISP-Insurance Federation

ISP-Retailors Federation

Benefits of Federated Learning

- **Globally optimized model built from data of multiple organizations**

- **Preserve data privacy and confidentiality**
  - Homomorphic encryption
  - Secure Aggregation protocol
  - Differential Privacy (DP)

- **Communication Cost Reduction**
  - Local computation of data
  - Reducing communication rounds and data per round

FATE: Federated AI Technology Enabler

- An open source project hosted by Linux Foundation
  - 3400+ stars, 7800+ commits, 50+ contributors
- Support both horizontal and vertical federated machine learning
- Design for enterprise use cases

https://github.com/FederatedAI/FATE

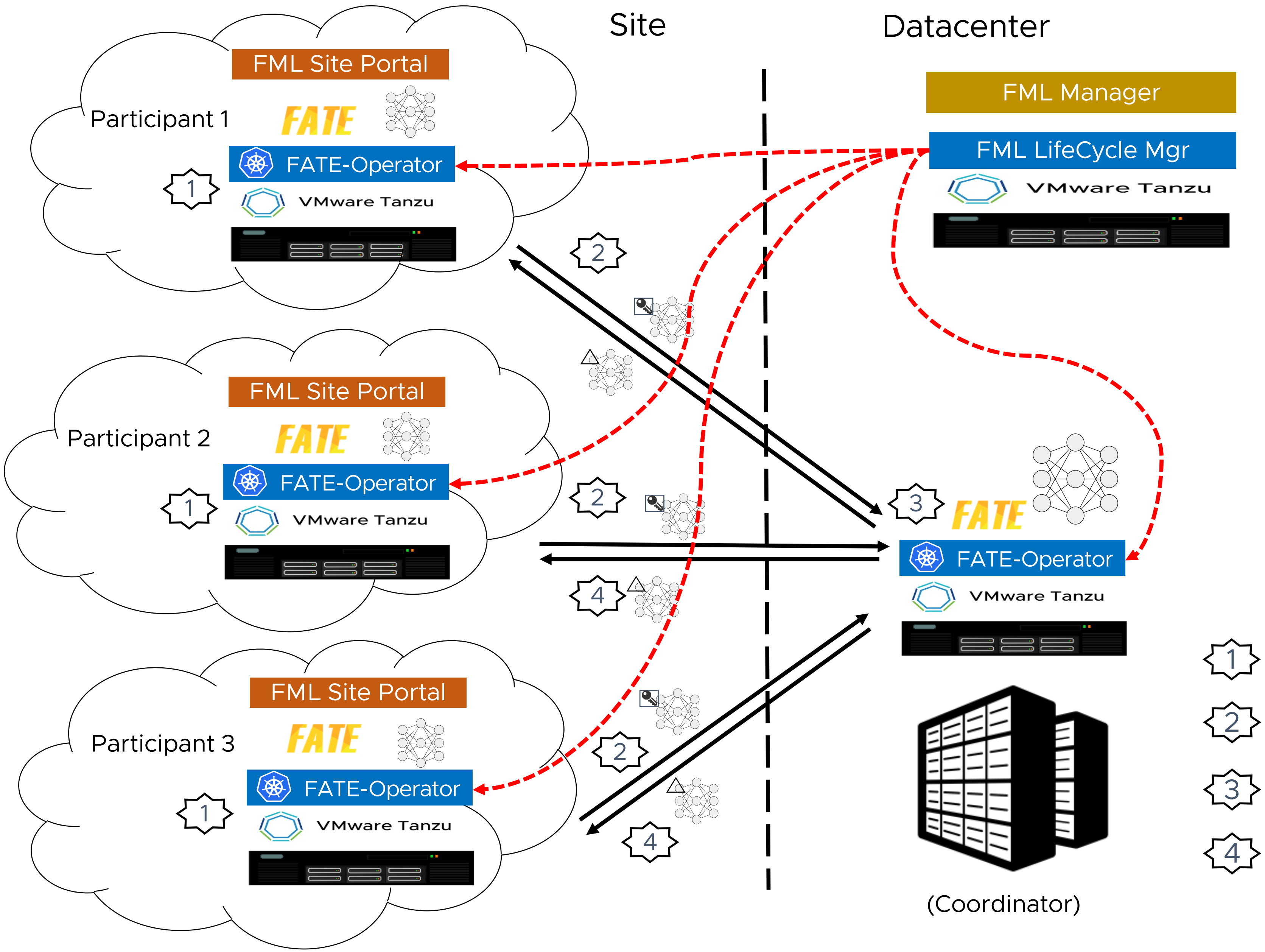
FATE

LINUX FOUNDATION

VMware’s Contribution to FATE

- TSC member of FederatedAI, under Linux Foundation
- Key contributor to OSS projects: FATE, KubeFATE
  - 7 contributors
  - 400+ commits
  - 15+ releases
- Contribution to Kubeflow project : FATE-Operator
- Active participation in FL community & evangelism:
  - Meetups
  - CCF Tech Forum
  - Academic and industry conferences

FML - Operational Model



Data mgmt, model mgmt, Authn/z, …

Lifecycle Manager：provisioning, federation setup, monitoring, etc.

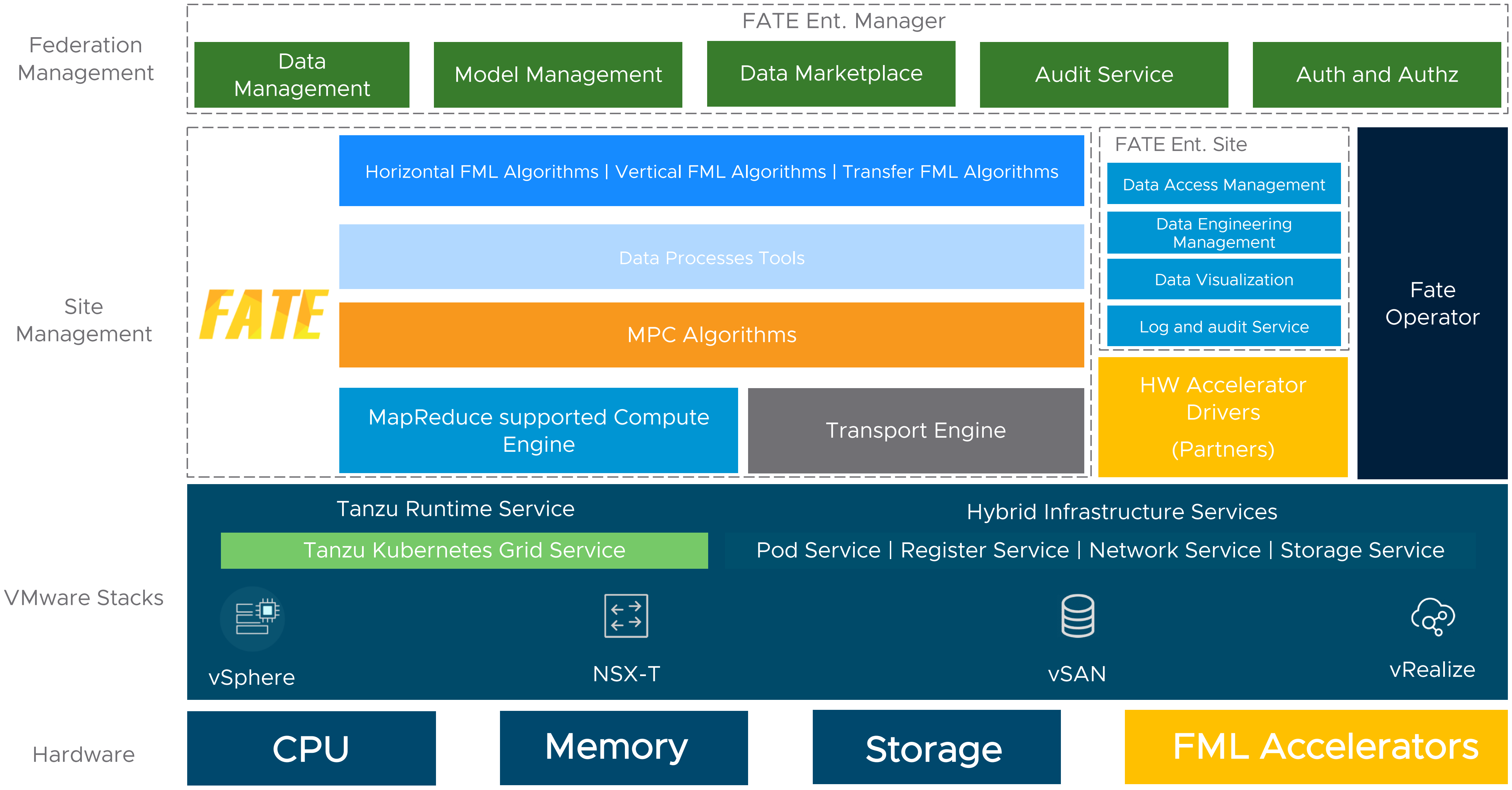
1 Train model locally

2 Send encrypted **model** to coordinator

3 Aggregate the encrypted model

4 Send back the updated encrypted model

Architecture Components



Federation Management

FATE Ent. Manager

Data Management | Model Management | Data Marketplace | Audit Service | Auth and Authz

Site Management

Horizontal FML Algorithms | Vertical FML Algorithms | Transfer FML Algorithms

Data Processes Tools

MPC Algorithms

MapReduce supported Compute Engine

Transport Engine

FATE Ent. Site

Data Access Management

Data Engineering Management

Data Visualization

Log and audit Service

HW Accelerator Drivers

(Partners)

Fate Operator

VMware Stacks

Tanzu Runtime Service

Tanzu Kubernetes Grid Service

Hybrid Infrastructure Services

Pod Service | Register Service | Network Service | Storage Service

vSphere

NSX-T

vSAN

vRealize

Hardware

CPU

Memory

Storage

FML Accelerators

Thank You