Federated ML application at edge R4 Architecture Document

Blueprint overview/Introduction

The AI Edge is an Akraino approved blueprint family and part of Akraino Edge Stack, which intends to provide an open source MEC platform combined with AI capacities at the Edge, and could be used for safety, security, and surveillance. The MEC platform, which named ote-stack, targets on shielding the heterogeneous characteristics through underlying hardware virtualization and providing an unified access for cloud edge, mobile edge and private edge. In addition, the AI Edge utilizes the cluster management and intelligent scheduling of multi-tier clusters to enable low-latency, high-reliability and cost-optimal computing support for running AI applications at the edge. At the same time, it makes device-edge-cloud collaborative computing possible.

This blueprint mainly focuses on building an edge federated ML platform to implement federated ML algorithms on servers in the edge.

Use Case

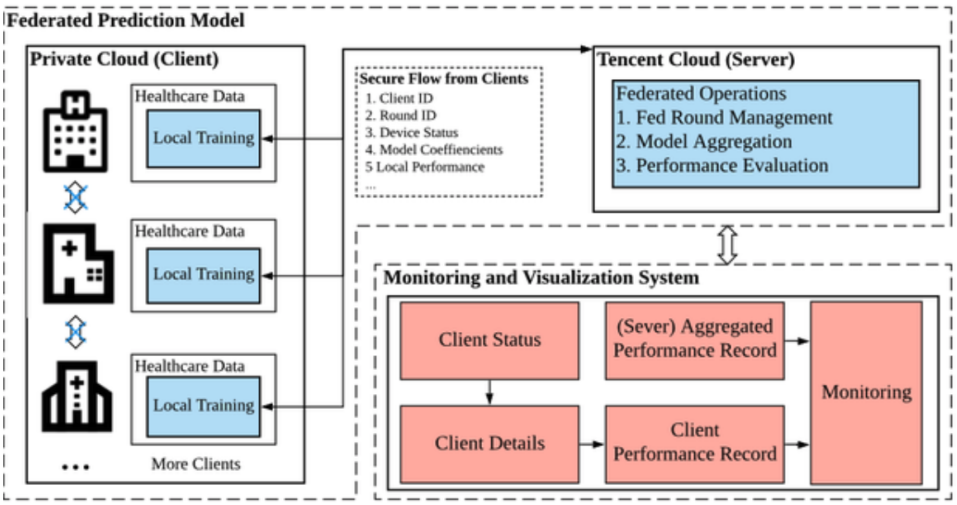
<use case 1: Federated Learning in Privacy Protection>

blocked URL

Facial recognition, voice assistant, and many other AI application require continuous training and optimization to get better performance. The traditional way is uploading the data into a centralized data server to finetune the machine learning model. But since these data may contain some privacy Infomation like facial image and voice data, it is dangerous for users to allow the producer to gather them. And since the enacting of GDPR, gathering these privacy data may cause the service provider faces some legal issues.

Federated Learning provides a solution to this dilemma. Rather than training the model in a centralized server, federated learning can allow the model to be trained on the edge devices where the data generated and only gather the gradient and model output, which can not be used to deduce the input data. In this way, the model can be updated safely and legally, users' privacy is protected and providers can update their models without any concerns.

<use case 2:Federated Learning in Data Gathering>



Deep learning is a promising way in aiding medical treatment and diagnosis, but an effective model requires quite large amounts of different data to conver ge. And the lack of sufficient training data is the biggest obstacle to developing AI models for medical applications.

Federated learning provides a safe tunnel across different medical institutions like hospitals and clinics, so that they can use train their own AI model with the data from multi-sources and without worry about the law issues and privacy problem with the help of Federated learning.

Where on the Edge

Business Drivers

The AI Edge will provide a cluster management for different logical MEC edge clusters. Through the standard api interface, the third clusters can join the management of AI Edge easily, so as to schedule deployment of an AI application to a specific edge node with the unified access. The benefits are: Lower cost on manage multiple edge clusters and more computing power of edge devices can be utilized ; Less load and latencies on network and more safely since the application is running locally; Edge cluster autonomy.

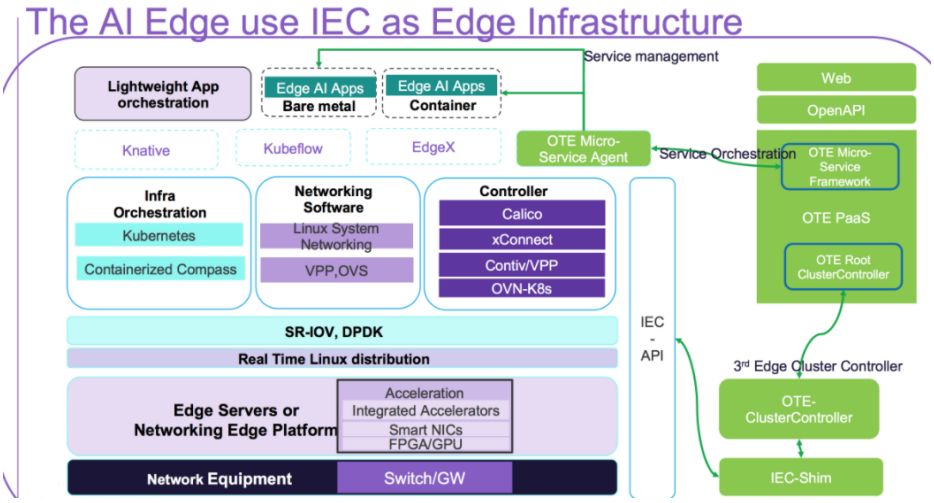
Overall Architecture

The AI Edge blueprint architecture consists of a cluster control manager with web platform at the cloud and multiple edge clusters. The number of clusters can be theoretically unlimited which can effectively solve the management and scheduling problems of large-scale mobile edge clusters in 5G era. For development environment we have tested with one IEC clusters with 3 nodes.

The cluster control manager, which consists of ote-web, openapi and lightweight cluster-controller, manages orchestration and life cycle of applications on the edge cluster and the hierarchical structure of clusters. While the ote-web and openapi provides access to the AI Edge, the cluster-controller provides core capabilities support for network connection, metadata synchronization and message transmit between cloud and edge and establishes the routing path for all edges. The edge, can be a kubernetes cluster, a k3s cluster or other private cluster, will be deployed a cluster-controller and a cluster-shim as to receive and process messages from upper cluster. Due to the autonomy of edge cluster, the network/workspace infrastructure and data volumes are managed by itself. Therefore, the deployed AI applications can still run normally when disconnection from the cloud occured.

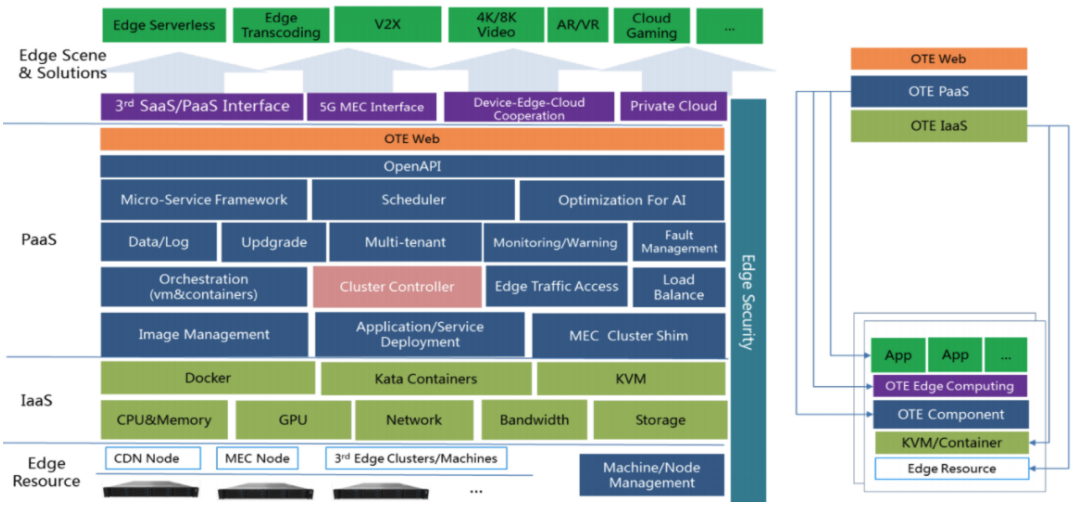
Many cloud native monitoring applications are used to collect container/node resource usage and running log, like prometheus, elasticsearch.

The below image shows the overall architecture for using IEC as edge infrastructure in AI Edge.



Platform Architecture

The detailed platform architecture of AI Edge blueprint is shown in the below diagram.



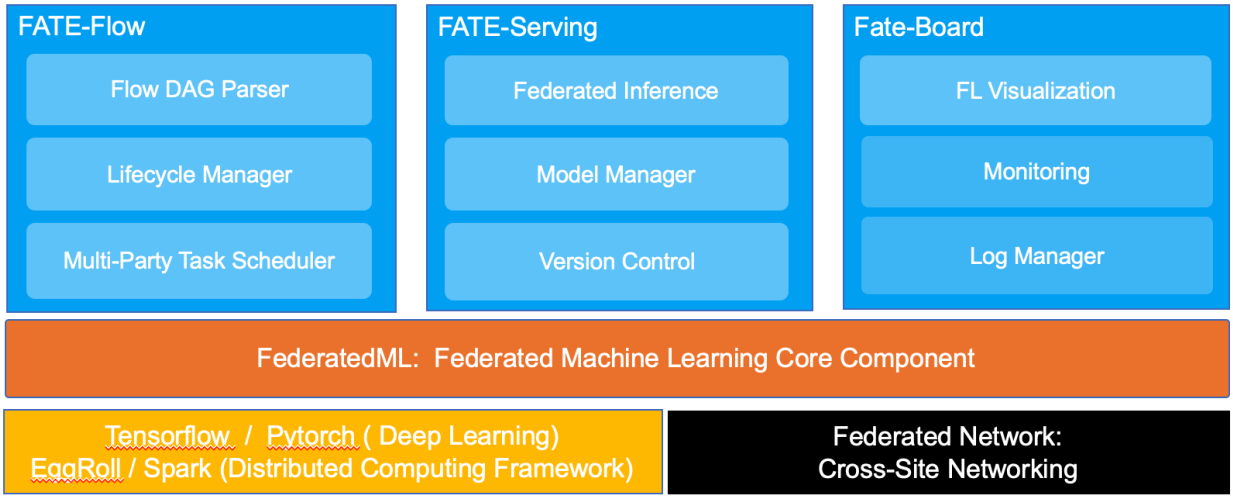
In the current release the components released are

- cluster-controller and controller-manager
- k8s-cluster-shim

Other components, such as openapi, ote-web, are currently released as docker images and will be open source in the future.

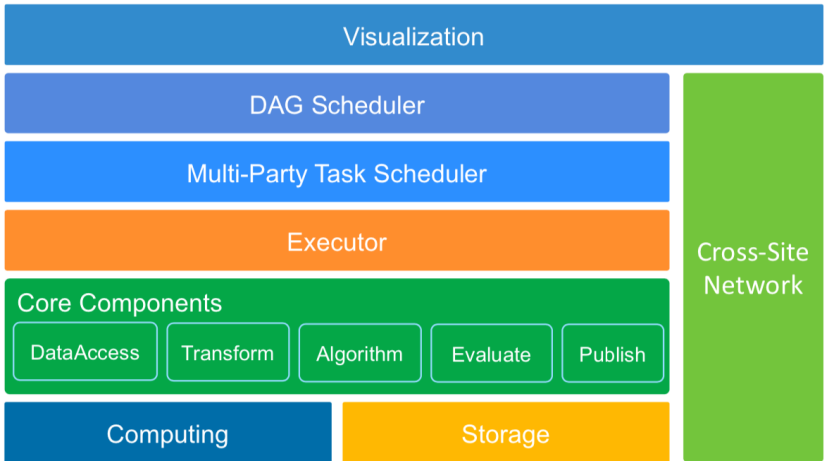
Software Platform Architecture

The below image shows the software architecture for this release.



Function Perspective

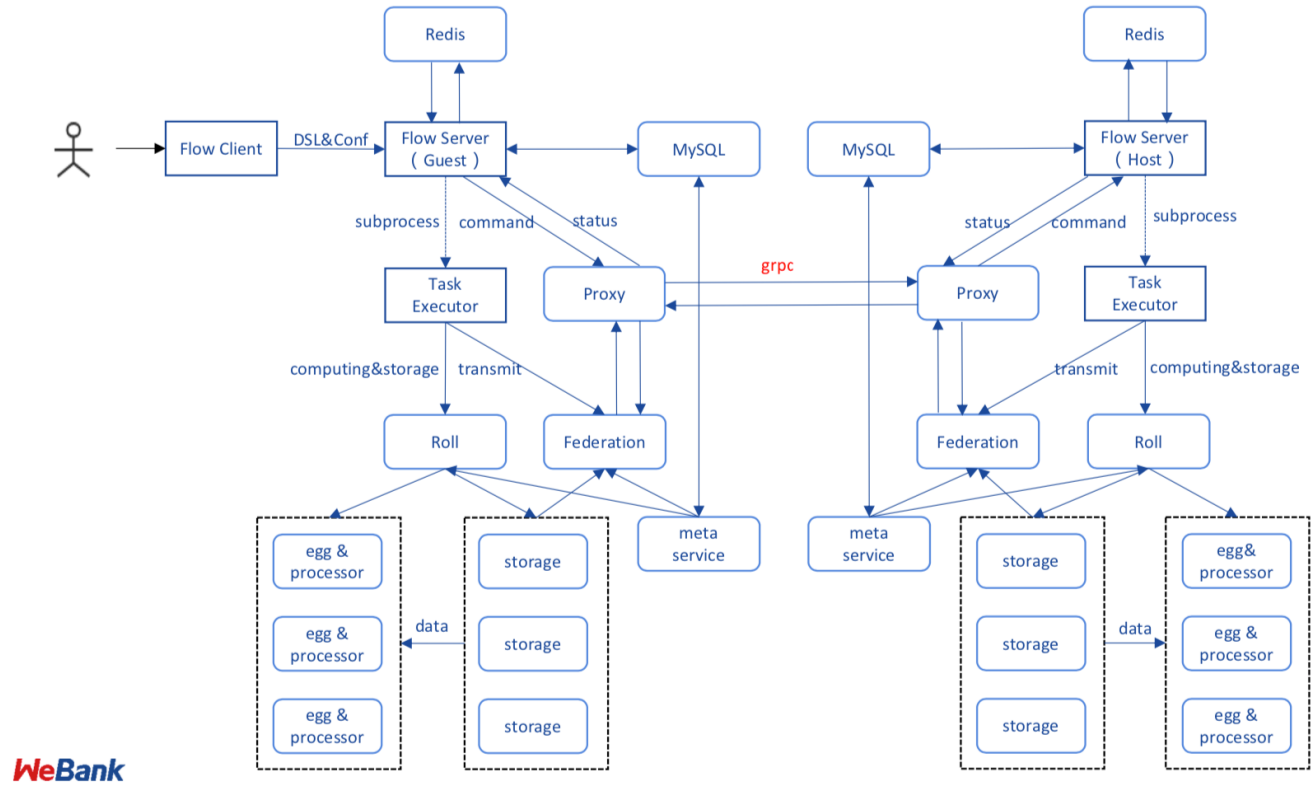
Introduction to Eggroll. What is EggRoll?



Keywords：

1. DAG
2. Component-based Pipeline
3. Multi-Party Scheduling
4. Cross-Site Transmission
5. Distributed Computing
6. Distributed Storage
7. Visualization

Basic Architecture



Data flow between and within devices.

APIs

Federated ML application at edge R4 API Document

Hardware and Software Management

Software Management: Gerrit Repo

Licensing

- Apache 2.0 license