

KubeEdge Edge Service Blueprint

A blueprint family which showcases end-to-end solution for edge services with KubeEdge centered edge stack. The first release will focus on the ML inference offloading use case.

KubeEdge is a CNCF sandbox project widely adopted as a key industry reference Edge Computing architecture. With KubeEdge, application vendors can push a wide range of applications at the edge, including Mobile Edge services. It is an upstream project to Akraino community. There are other Akraino Blueprint projects with KubeEdge incorporated. This project family will focus on solutions centered around KubeEdge. Future releases may incorporate various components, e.g. hardware infrastructure etc.

Weekly Meeting: <https://zoom.us/j/91049610205>, every Tuesday 20:00 PDT.

Slack Channel: <https://fedge.slack.com/archives/C0155MP4TSB>

Use Case Details

| Attributes | Description | I n f o r m a t i o n a l |
|---|--|---|
| Type | New | |
| Industry Sector | Cloud, Enterprise, Telco | |
| Business driver | Edge computing leverages edge locations to distribute application loads among device/edge/cloud. A service layer is required to bridge infrastructure platform and applications. e.g. load distribution coordination, hardware platform agnostic, etc. KubeEdge extends native containerized application orchestration capabilities to hosts at Edge. Along with other vertical domain support such as device twin at edge, KubeEdge edge service stack is geared to offer feature rich support to applications while remain platform neutral. | |
| Business use cases | KubeEdge Edge service can be deployed at enterprise edge or as a cloud edge extension interfacing telco network. It offers support for following use cases: <ul style="list-style-type: none">• ML offloading for inference and training in image recognition for mobile phones• Automatic Speech Recognition (ASR) in operation field• Manufacture production line defect inspection• IoT gateway• Mobile Edge enabler | |
| Business Cost - Initial Build Cost Target Objective | KubeEdge is a software layer. Its managed applications can run on any kubernetes environment. Validated edge stack including hardware choices should have manageable cost suitable for edge deployment. | |
| Business Cost – Target Operational Objective | KubeEdge edge service provides service portal for operational management. It supports zero touch deployment and monitoring capabilities. | |
| Security need | KubeEdge supports application oriented security SPIFFE spec. | |
| Regulations | N/A | |
| Other restrictions | N/A | |
| Additional details | N/A | |

Blueprint Family Details

| Use Case Attributes | Description | Informational |
|--|--|---------------|
| Type | New | |
| Blueprint Family | KubeEdge Edge Service | |
| Use Case | Telco edge and enterprise edge | |
| Blueprint proposed | Central Office deployments <ul style="list-style-type: none"> • ML inference offloading Customer Premise deployments <ul style="list-style-type: none"> • ASR at operation field (future proposal) | |
| Initial POD Cost (capex) | Less than USD100K | |
| Scale | From 1 server to a rack. | |
| Applications | Any type of edge services | |
| Power Restrictions | Varies | |
| Preferred Infrastructure orchestration | OpenStack - VM orchestration Docker/K8 - Container Orchestration OS - Linux VNF Orchestration - ONAP | |
| Additional Details | N/A | |

Blueprint Details

| Case Attributes | Description | Informational |
|------------------------------|--|---------------|
| Type | New | |
| Blueprint Family | KubeEdge Edge Service | |
| Use Case | Facial emotion recognition task offloading to edge node | |
| Blueprint proposed Name | ML Inference Offloading | |
| Initial POD Cost (capex) | Less than 100KUSD | |
| Scale & Type | 1 x86 server With Nvidia Tesla P4/T4 GPUs | |
| Applications | Deep learning models(facial expression) offload from mobile device to Edge | |
| Power Restrictions | Varies | |
| Infrastructure orchestration | Docker 18.09 OS – Ubuntu18.04 Python 3.5 ~3.7 CUDA>10.1 GPU driver release 19.03 | |
| PaaS | KubeEdge, Kubernetes | |
| SDN | N/A | |
| Workload Type | Containers | |
| Additional Details | N/A | |

Committers

Yin Ding of Futurewei is the PTL (1 May 2020 - 30 April 2021).

| Committer | Committer Company | Committer Contact Info | Committer Bio | Committer Picture | Self Nominate for PTL (Y/N) |
|------------------|--------------------------|-------------------------------|----------------------|--------------------------|------------------------------------|
| Jane Shen | Futurewei | jane.shen@futurewei.com | | | |
| Yin Ding | Futurewei | yin.ding@futurewei.com | | | Y |
| Tina Tsou | ARM | tina.tsou@arm.com | | | |
| Xuan Jia | China Mobile | jiaxuan@chinamobile.com | | | |
| Jiafeng Zhu | Futurewei | jjafeng.zhu@futurewei.com | | | |
| Hanyu Ding | China Mobile | dinghanyu@chinamobile.com | | | |
| Jeff Brower | Signallogic | jbrower@signallogic.com | | | |
| Cindy Xing | Microsoft | cixing@microsoft.com | | | |
| Hao Xu | Futurewei | hxu1@futurewei.com | | | |

Contributors

| Contributor | Contributor Company | Contributor Contact Info | Contributor Bio | Contributor Picture |
|--------------------|----------------------------|---------------------------------|------------------------|----------------------------|
| May Chen | | | | |
| | | | | |
| | | | | |

[Akraino KubeEdge Edge Service Blueprint.pptx](#)