Blueprint: The AI Edge

Baidu
Table of Contents

- General Blueprint Introduction
- Blueprint Use Cases
- Network Architecture
- MEC Architecture/MEC Software Architecture
- Physical topology
- The AI Edge use ICN&IEC as Edge Infrastructure
- Future Plan
General Blueprint Introduction

- The AI Edge Blueprint focuses on establishing an open source MEC platform combined with AI capacities at the Edge, which could be used for safety, security, and surveillance.
- The first release of the AI Edge Blueprint is targeted at Q4, 2019.
- The AI Edge Blueprint is sponsored by Baidu, Arm, and Intel.
- Contact: Hechun Zhang (zhanghechun@baidu.com)
- Refer to: https://wiki.akraino.org/display/AK/The+AI+Edge+Blueprint+Family
## Blueprint Use cases

<table>
<thead>
<tr>
<th>Use case</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Security Monitoring</td>
<td>Conduct smoke detection on densely populated places such as industrial parks and community properties to quickly detect whether there is a fire, reduce the damage caused by fire, and improve the safety of the park.</td>
</tr>
<tr>
<td>Classroom concentration analysis</td>
<td>Conduct a full evaluation of the overall class and the concentration of individual students, help teachers and school authorities to fully understand the teaching situation, according to the concentration data of each course, conduct targeted class knowledge test and strengthen.</td>
</tr>
<tr>
<td>Factory safety production</td>
<td>By identifying whether employees wear helmets, safety gloves, etc., monitor and evaluate the work of workers in the workshop to help companies monitor the safety production status in a comprehensive and timely manner. Statistical results and analytical data can be used as a reference for strengthening the level of safety production management.</td>
</tr>
<tr>
<td>Kitchen hygiene monitoring</td>
<td>Monitor staff's smoking and cell phone behavior in the kitchen of the catering industry, ensuring the safety and hygiene of the food production process.</td>
</tr>
</tbody>
</table>
Network Architecture

Terminal
- IP Camera
- Camera
- Smart Terminal

Data Transmission

Edge site
- Video Processing Hub
- Result Analysis Function
- Data Collection
- Computing
- Video Analysis Model
- Data Sync
- Configuration Sync

OTE Stack
- CDN Server
- MEC Server
- Enterprise Server
- GPU
- CPE
- General Server

Cloud (public/private/…)
- Data Transmission
  - Data/Picture/Streaming
- Knowledge downlink
  - Function/Model
MEC Software Architecture

Edge Serverless | Edge Transcoding | V2X | 4K/8K Video | AR/VR | Cloud Gaming | …

3rd SaaS/PaaS Interface | 5G MEC Interface | Device-Edge-Cloud Cooperation | Private Cloud

MEC Software

Edge Security

OTE Web
OTE PaaS
OTE IaaS

App
OTE Edge Computing
OTE Component
KVM/Container
Edge Resource

OTE Web
OTE PaaS
OTE IaaS

PaaS

Data/Log
Upgrade
Orchestration (vm&containers)
Image Management

Services

Scheduler
Multi-tenant
Application/Service Deployment

Monitoring/Warning
Fault Management

Edge Traffic Access
Load Balance

KVM

Machine/Node Management

CPU&Memory
GPU
Network
Bandwidth
Storage

3rd Edge Clusters/Machines
Physical Topology

Cluster adaptation layer, responsible for accessing different types of clusters

1. Provide multiple runtime nodes to support the generation of various IaaS resources
2. Provide high performance, container network components that support network policies

Cluster Controller maintains tree topology
1. selector, cluster label screening;
2. Operator & scheduler, sub-cluster task scheduling assignment;
3. edge-tunnel & edge-handler, connect parent node and process command
4. Router, maintain routing table of all descendant clusters
5. cloud-handler & cloud-tunnel, command conversion and subcluster connection

Cluster: root

Cluster Controller

- cluster-shim
- cluster-selector
- cluster-scheduler
- cluster-operator
- edge-tunnel
- edge-handler
- router
- cloud-handler
- cloud-tunnel

K8S API-Server

Node

Cluster: child1

cluster-shim

K8S cluster

Cluster Controller

Cluster 1

Cluster 2

Cluster N

Cluster: child2

cluster-shim

K3S cluster

Cluster Controller

Cluster 1

Cluster 2

Cluster N

Cluster: child2

cluster-shim

3rd cluster

Cluster Controller

Cluster 1

Cluster 2

Cluster N

Cluster adaptation layer, responsible for accessing different types of clusters
The AI Edge use IEC as Edge Infrastructure

- **Lightweight App orchestration**
  - Knative
  - Kubernetes
  - Containerized Compass

- **Infra Orchestration**
  - Kubernetes

- **Networking Software**
  - Linux System Networking
  - VPP, OVS

- **Network Equipment**
  - Switch/GW

- **Controller**
  - Calico
  - xConnect
  - Contiv/VPP
  - OVN-K8s

- **Edge AI Apps**
  - Bare metal
  - Container

- **Edge Servers or Networking Edge Platform**
  - Acceleration
    - Integrated Accelerators
    - Smart NICs
    - FPGA/GPU

- **Service management**

- **Service Orchestration**
  - Web
  - OpenAPI
  - OTE Micro-Service Framework
  - OTE PaaS
  - OTE Root ClusterController

- **Adapter for 3rd Edge Cluster**
  - OTE-ClusterController
  - IEC-Shim

- **3rd Edge Cluster Controller**

- **IEC - API**
The AI Edge use ICN as Edge Infrastructure

**Infra Orchestration**
- Kubernetes
- Containerized Compass

**Networking Software**
- Linux System Networking
- VPP, OVS

**Controller**
- Calico
- xConnect
- Contiv/VPP
- OVN-K8s

**Accelerators**
- Integrated Accelerators
  - Smart NICs
  - FPGA/GPU

**Edge Servers or Networking Edge Platform**
- Switch/GW

**Lightweight App orchestration**
- Knative
- Kubeflow
- EdgeX

**Service Orchestration**
- Web
- OpenAPI

**OTE Root ClusterController**
- OTE Micro-Service Framework
- OTE PaaS

**3rd Edge Cluster Controller**
- Adapter for 3rd Edge Cluster

**SR-IOV, DPDK**
- Real Time Linux distribution

**ICN - API**
Future Plan

- The first demo of the AI Edge Blueprint is targeted at Oct 31st, 2019.
- The first Akraino version will be released in Q4, 2019.
# The AI Edge Blueprint Criteria

<table>
<thead>
<tr>
<th>Case Attributes</th>
<th>Description</th>
<th>Informational</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>The AI Edge</td>
<td></td>
</tr>
<tr>
<td>Blueprint Family - Proposed Name</td>
<td>New</td>
<td></td>
</tr>
<tr>
<td>Use Case</td>
<td>Safety, Security, and Surveillance</td>
<td></td>
</tr>
<tr>
<td>Blueprint proposed Name</td>
<td>The AI Edge</td>
<td></td>
</tr>
<tr>
<td>Initial POD Cost (capex)</td>
<td>Leverage Unicycle POD - less than $150k</td>
<td></td>
</tr>
<tr>
<td>Scale &amp; Type</td>
<td>Up to 4 servers, x86/ARM server or deep edge class</td>
<td></td>
</tr>
<tr>
<td></td>
<td>With nVIDIA Tesla P4/T4 GPUs</td>
<td></td>
</tr>
<tr>
<td>Applications</td>
<td>1. Small scale IoT data aggregation and machine learning platform</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. Small scale deep learning models training for video data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Model training for Acumos Apps</td>
<td></td>
</tr>
<tr>
<td>Power Restrictions</td>
<td>Less than 10Kw</td>
<td></td>
</tr>
<tr>
<td>Infrastructure orchestration</td>
<td>Docker 1.13.1 or above</td>
<td></td>
</tr>
<tr>
<td></td>
<td>K8s 1.12.5 or above- Container Orchestration</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OS – CentOS 7.0 or above</td>
<td></td>
</tr>
<tr>
<td>PaaS</td>
<td>OTE Stack</td>
<td></td>
</tr>
<tr>
<td>SDN</td>
<td>Calico container networking, or OVS-DPDK</td>
<td></td>
</tr>
<tr>
<td>Workload Type</td>
<td>Containers</td>
<td></td>
</tr>
<tr>
<td>Additional Details</td>
<td>Acumos and EdgeX interfaces will be connected in this case</td>
<td></td>
</tr>
</tbody>
</table>
The Linux Foundation

1 Letterman Drive
Building D, Suite D4700
San Francisco CA 94129
Phone/Fax: +1 415 7239709
www.linuxfoundation.org

General Inquiries
info@linuxfoundation.org

Membership
membership@linuxfoundation.org

Corporate Training
training@linuxfoundation.org

Event Sponsorship
sponsorships@linuxfoundation.org
Legal Notices

The Linux Foundation, The Linux Foundation logos, and other marks that may be used herein are owned by The Linux Foundation or its affiliated entities, and are subject to The Linux Foundation’s Trademark Usage Policy at https://www.linuxfoundation.org/trademark-usage, as may be modified from time to time.

Linux is a registered trademark of Linus Torvalds. Please see the Linux Mark Institute’s trademark usage page at https://lmi.linuxfoundation.org for details regarding use of this trademark.

Some marks that may be used herein are owned by projects operating as separately incorporated entities managed by The Linux Foundation, and have their own trademarks, policies and usage guidelines.

TWITTER, TWEET, RETWEET and the Twitter logo are trademarks of Twitter, Inc. or its affiliates.

Facebook and the “F” logo are trademarks of Facebook or its affiliates.

LinkedIn, the LinkedIn logo, the IN logo and InMail are registered trademarks or trademarks of LinkedIn Corporation and its affiliates in the United States and/or other countries.

YouTube and the YouTube icon are trademarks of YouTube or its affiliates.

All other trademarks are the property of their respective owners. Use of such marks herein does not represent affiliation with or authorization, sponsorship or approval by such owners unless otherwise expressly specified.

The Linux Foundation is subject to other policies, including without limitation its Privacy Policy at https://www.linuxfoundation.org/privacy and its Antitrust Policy at https://www.linuxfoundation.org/antitrust-policy, each as may be modified from time to time. More information about The Linux Foundation’s policies is available at https://www.linuxfoundation.org.

Please email legal@linuxfoundation.org with any questions about The Linux Foundation’s policies or the notices set forth on this slide.