Edge Computing & OnPrem Infra
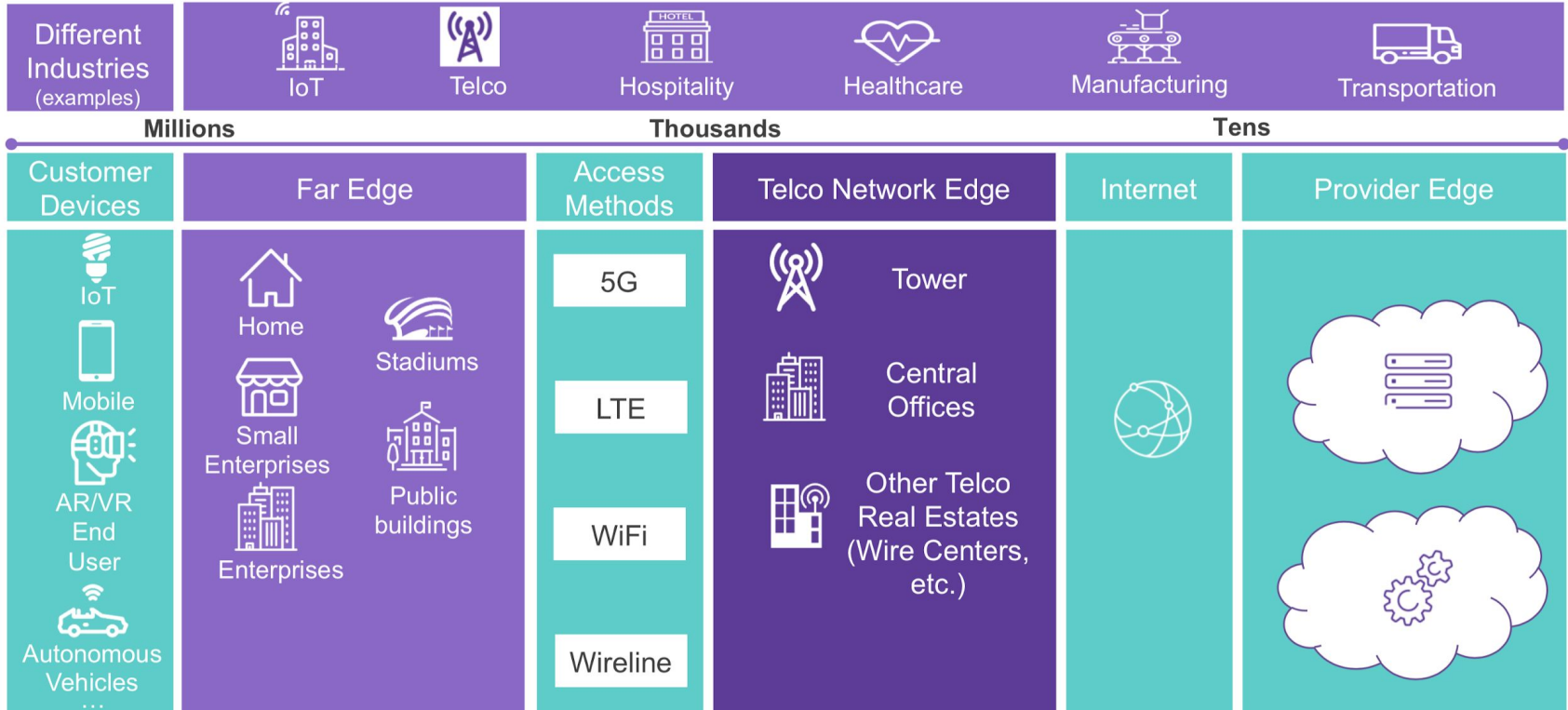
Feng Ding @ RAS

Content

- **Edge Computing Concepts**

- **KubeEdge (Architecture, Device Management, Service Discovery)**

- **Workflow Management Systems**

- **Use-cases & Outlook**

Edge Computing Concepts

Different Industries (examples)	IoT	Telco	Hospitality	Healthcare	Manufacturing	Transportation

Millions — Thousands — Tens

Customer Devices	Far Edge	Access Methods	Telco Network Edge	Internet	Provider Edge
IoT, Mobile, AR/VR End User, Autonomous Vehicles …	Home, Stadiums, Small Enterprises, Public buildings, Enterprises	5G, LTE, WiFi, Wireline	Tower, Central Offices, Other Telco Real Estates (Wire Centers, etc.)		

Edge Computing Concepts

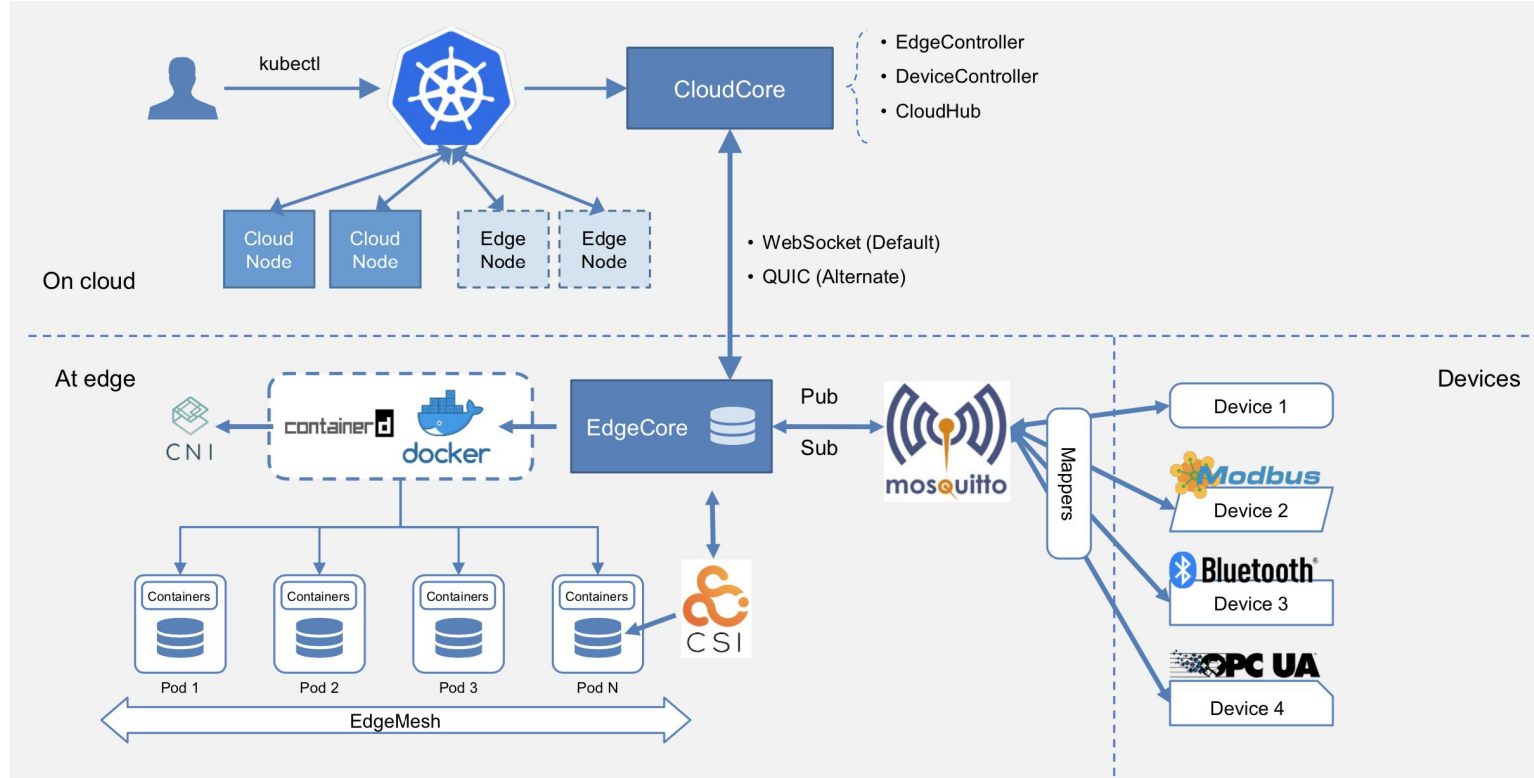| Low Latency		Bandwidth Optimization
	Edge Computing	
Autonomy		Privacy

Software Hierarchy

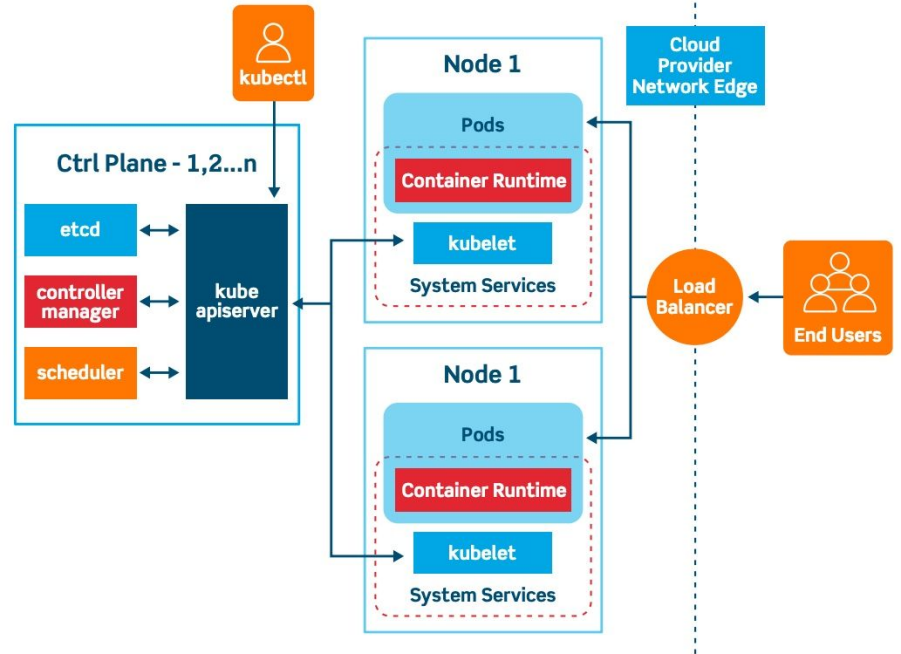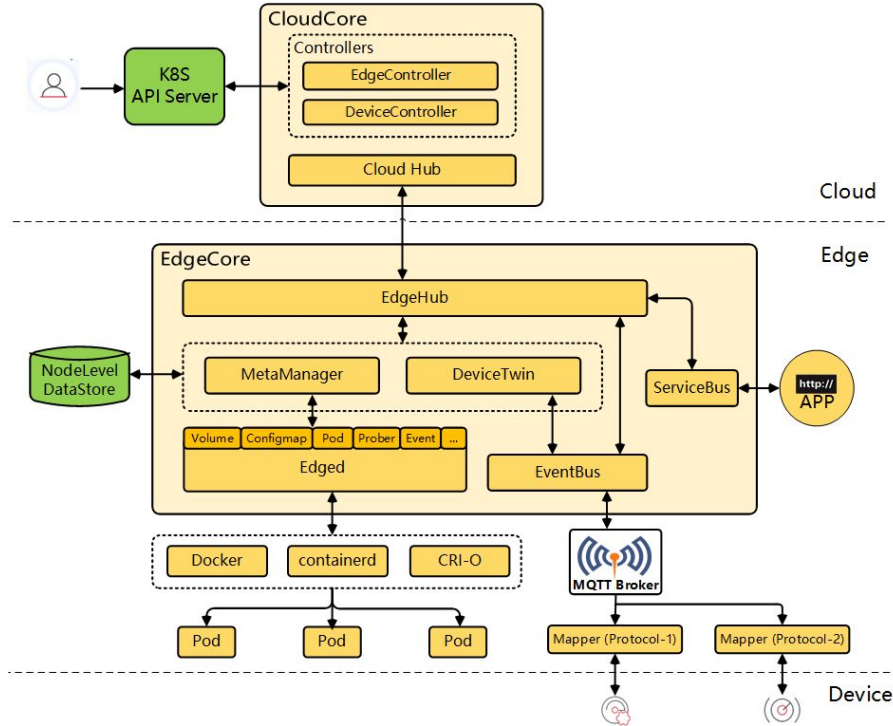Applications & Use-cases

Workflow Management Systems (WMSs)

Computing Infrastructures (K8s, Kubeedge)

Physical Resources & Networking

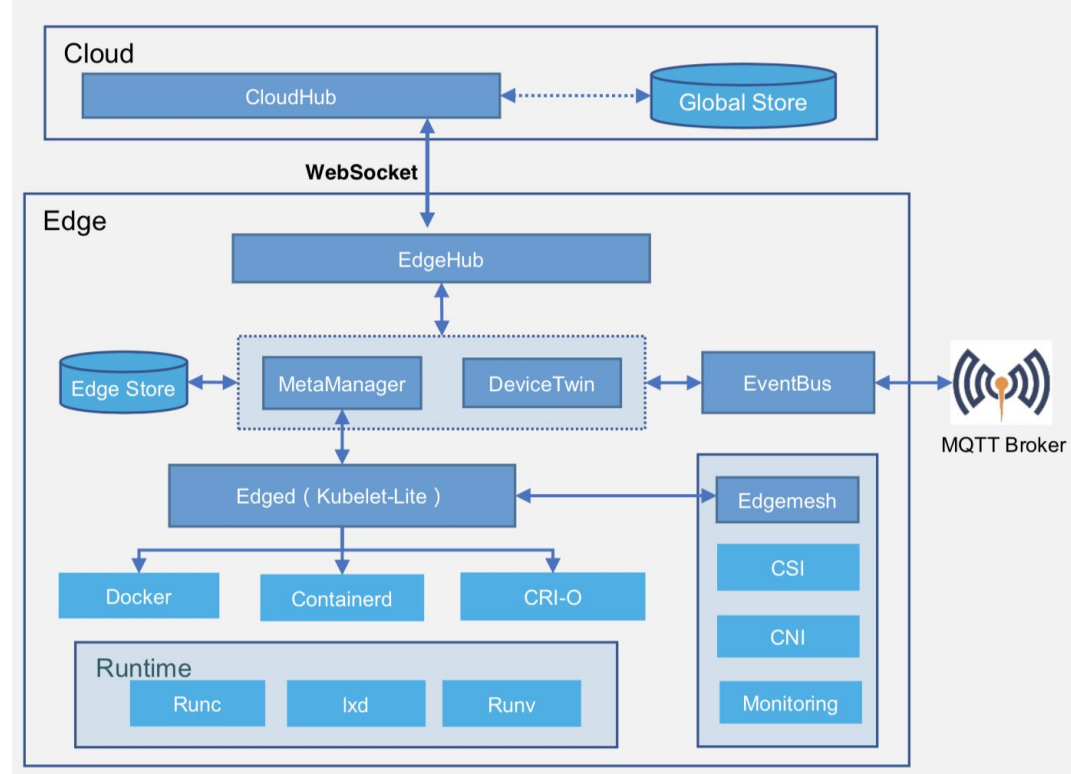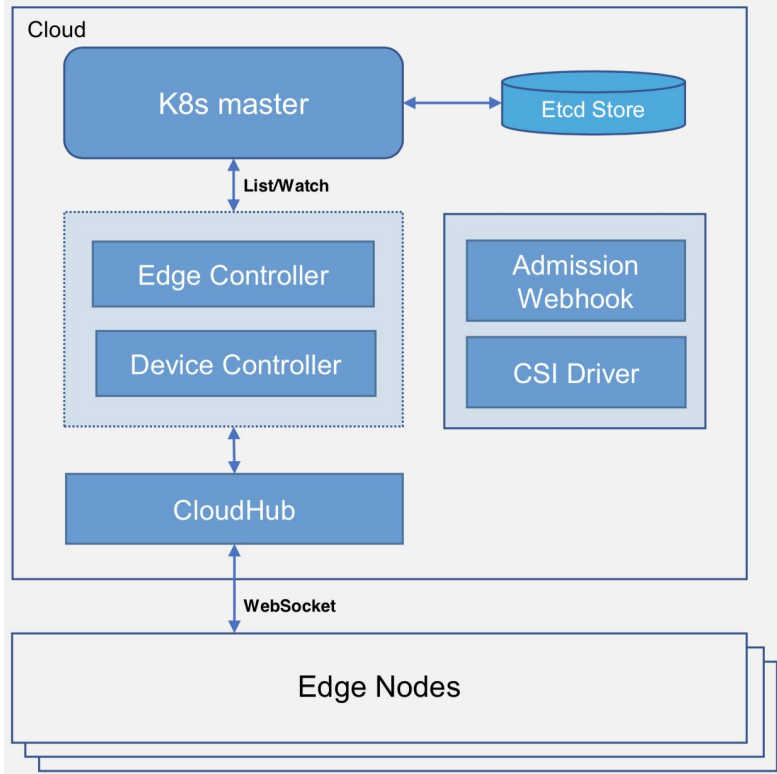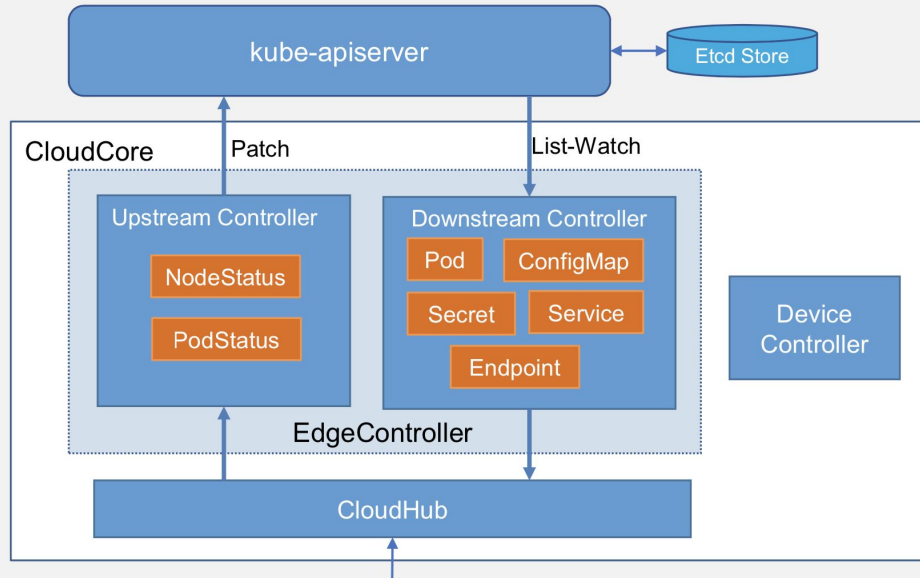# KubeEdge Architecture

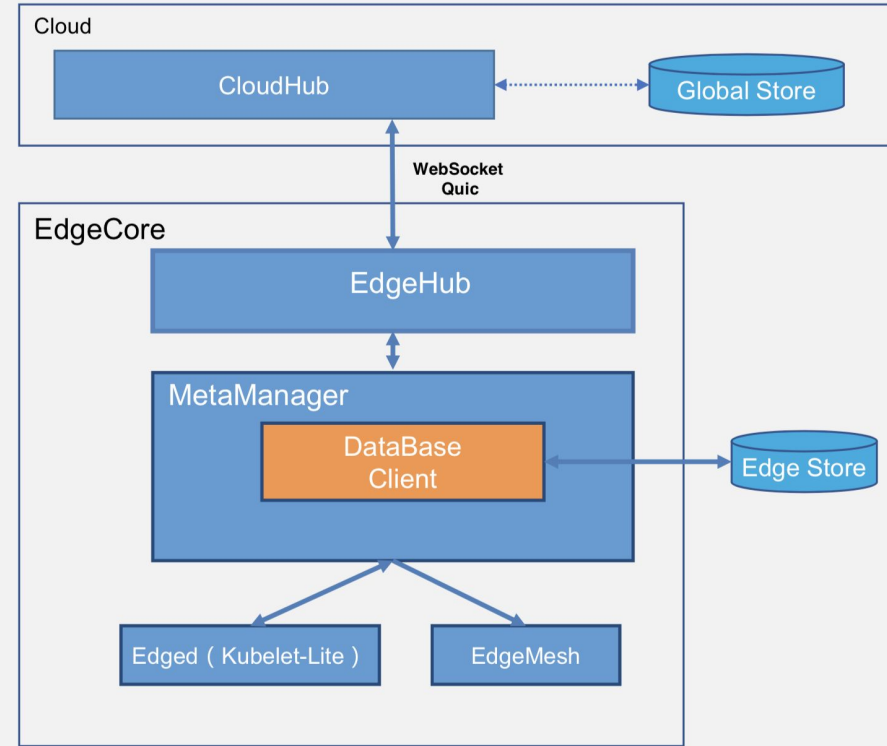# KubeEdge vs K8s
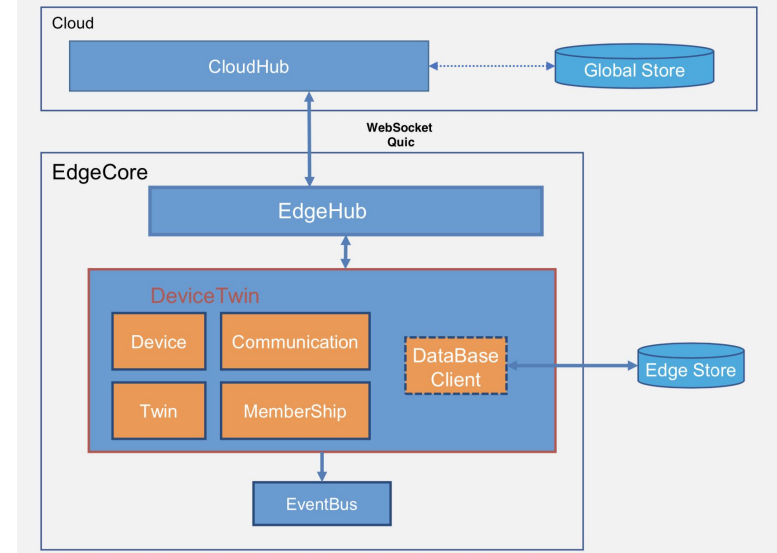
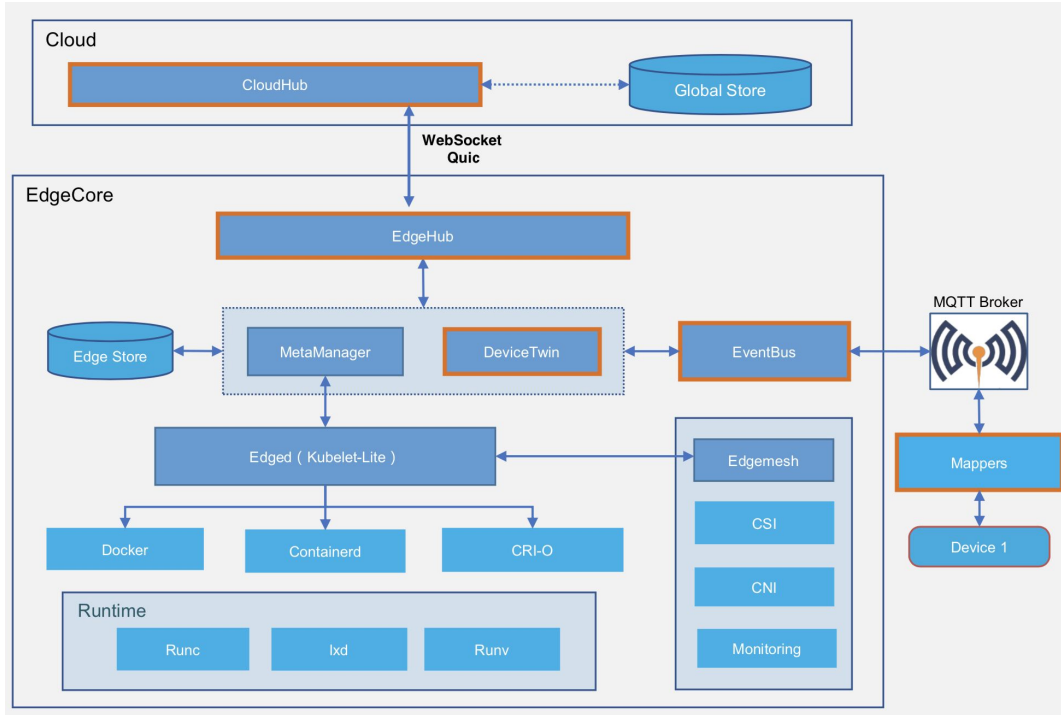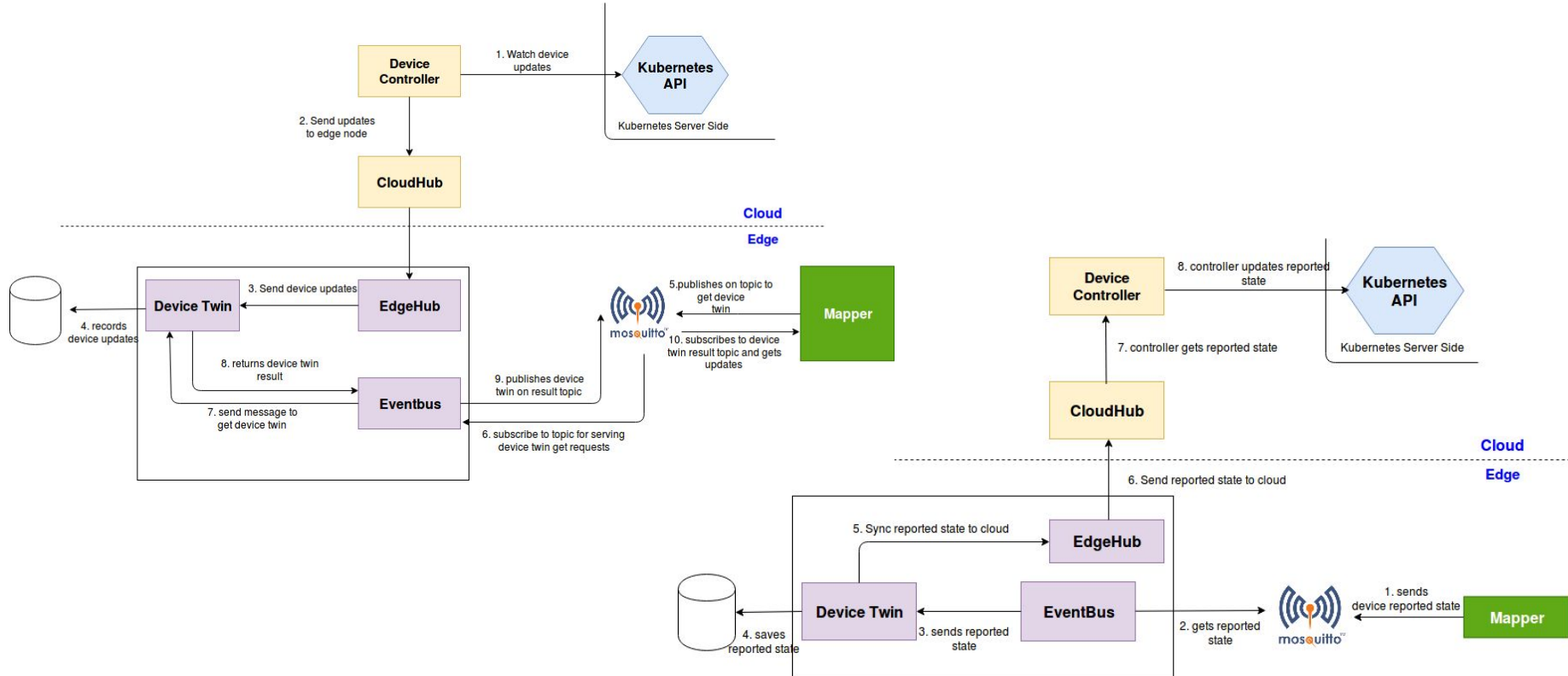# KubeEdge - Key Components

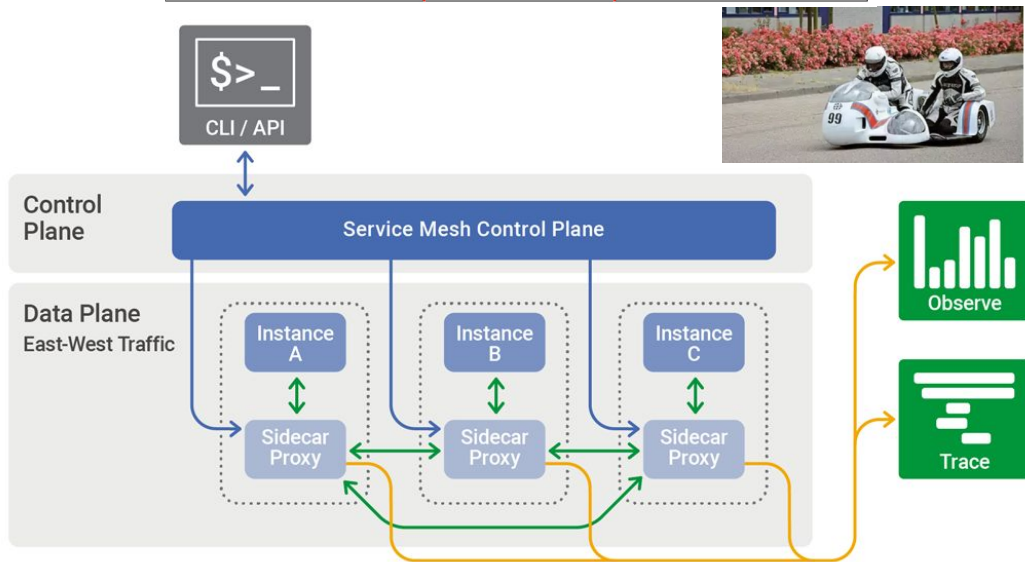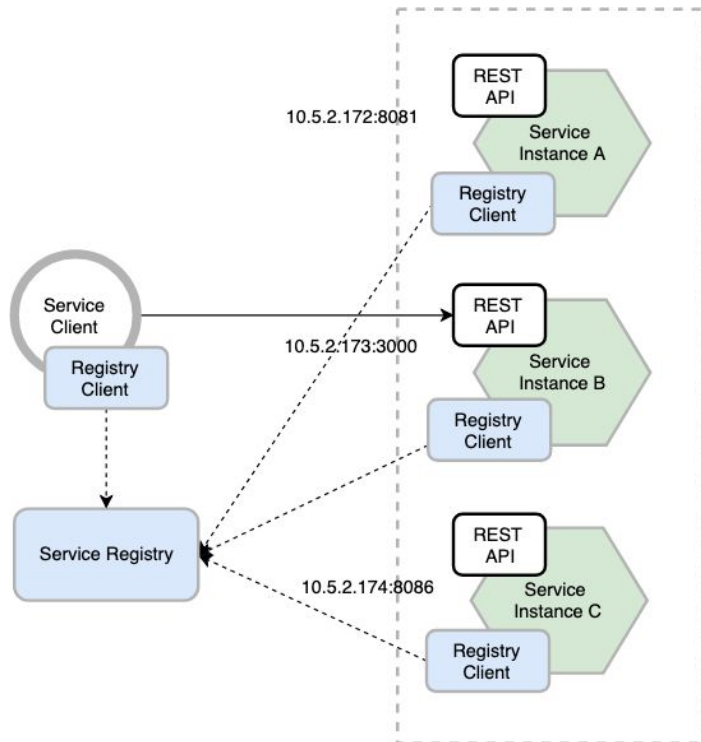# KubeEdge - Local Autonomy

Device Management - Device Twin

Device Management - Status Sync

Service Discovery

REST API
Service Instance A

10.5.2.172:8081

Registry Client

REST API
Service Instance B

10.5.2.173:3000

Registry Client

REST API
Service Instance C

10.5.2.174:8086

Registry Client

Service Client

Registry Client

Service Registry

apiserver

Client   kube-proxy

clusterIP
(Virtual Server)

Node

Backend Pod 1   Backend Pod 2   Backend Pod 3

(Real Server)

$>_
CLI / API

Control Plane

Service Mesh Control Plane

Data Plane
East-West Traffic

Instance A   Instance B   Instance C

Sidecar Proxy   Sidecar Proxy   Sidecar Proxy

Observe
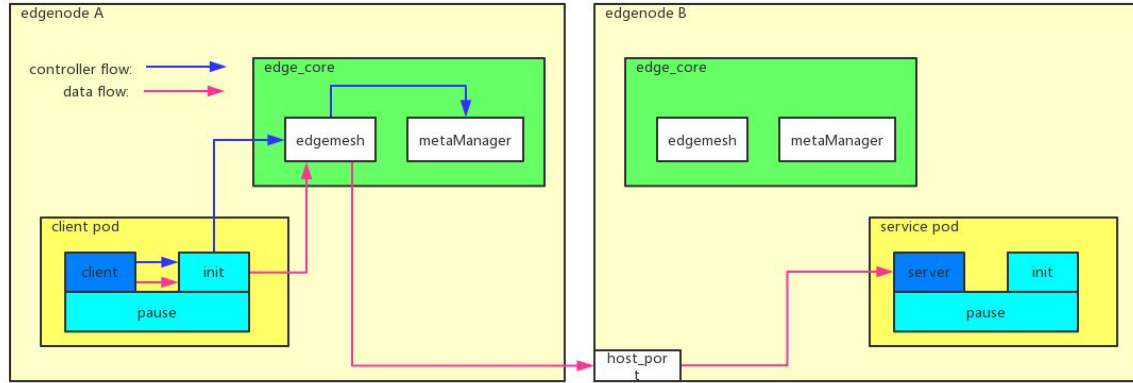
Trace

Service Discovery - EdgeMesh
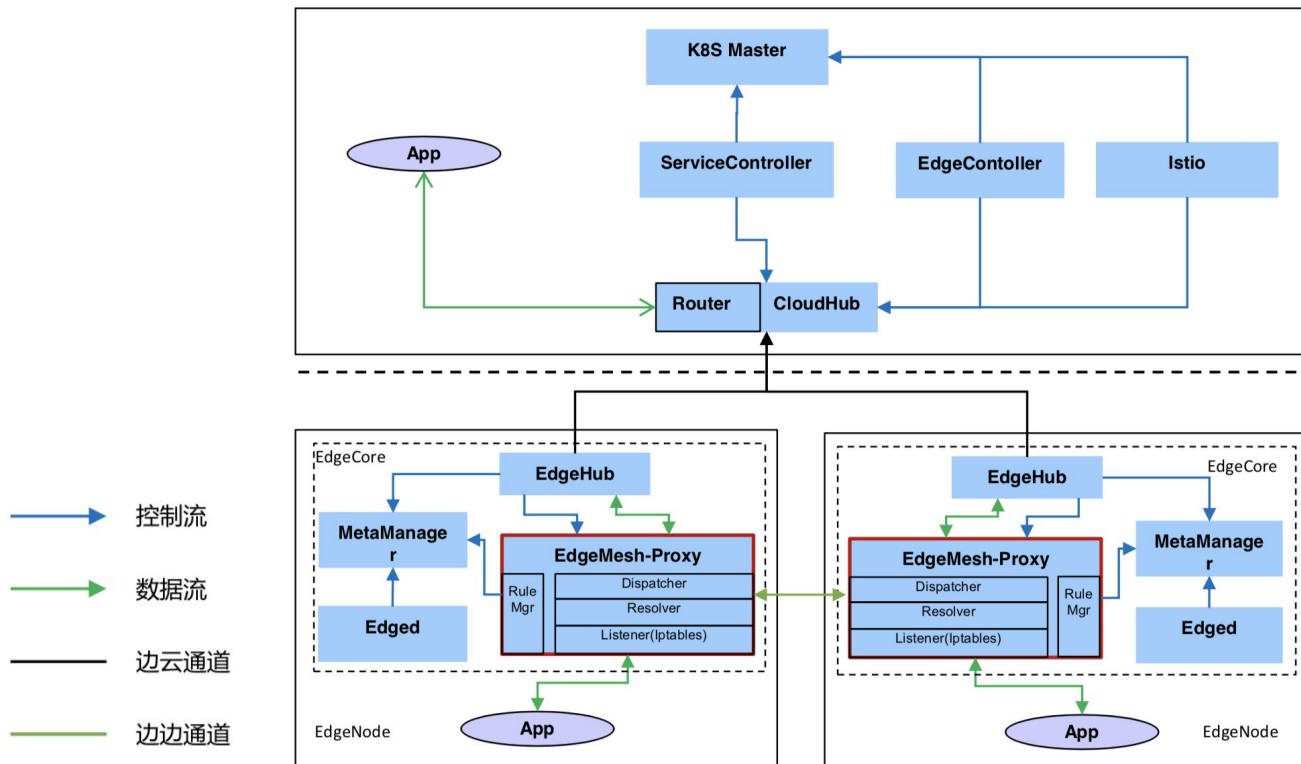
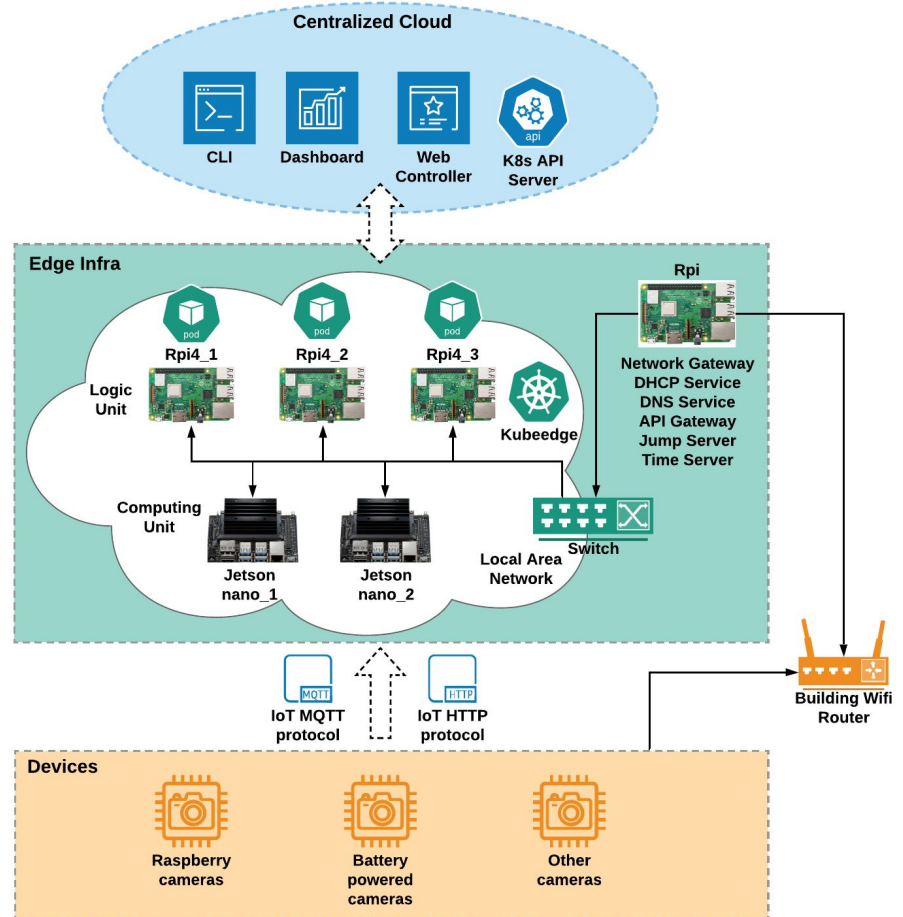# Service Discovery - Differences with ServiceMesh

- **EdgeMesh**

- **ServiceMesh**

Use cases

- **Smart Building/Park:**
  Face Detecting
  Pedestrian Flow Detecting
- **CDN:**
  Video Rendering
- **Industrial Internet:**
  Model Pull/Model Inference
- **Smart Car**

Data Science Hierarchy

THE DATA SCIENCE
HIERARCHY OF NEEDS

AI, Deep Learning

LEARN/OPTIMIZE — A/B Testing, Experimentation, Simple ML Algorithms

LEARN/OPTIMIZE — Analytics, Metrics, Segments, Aggregates, Features, Training Data

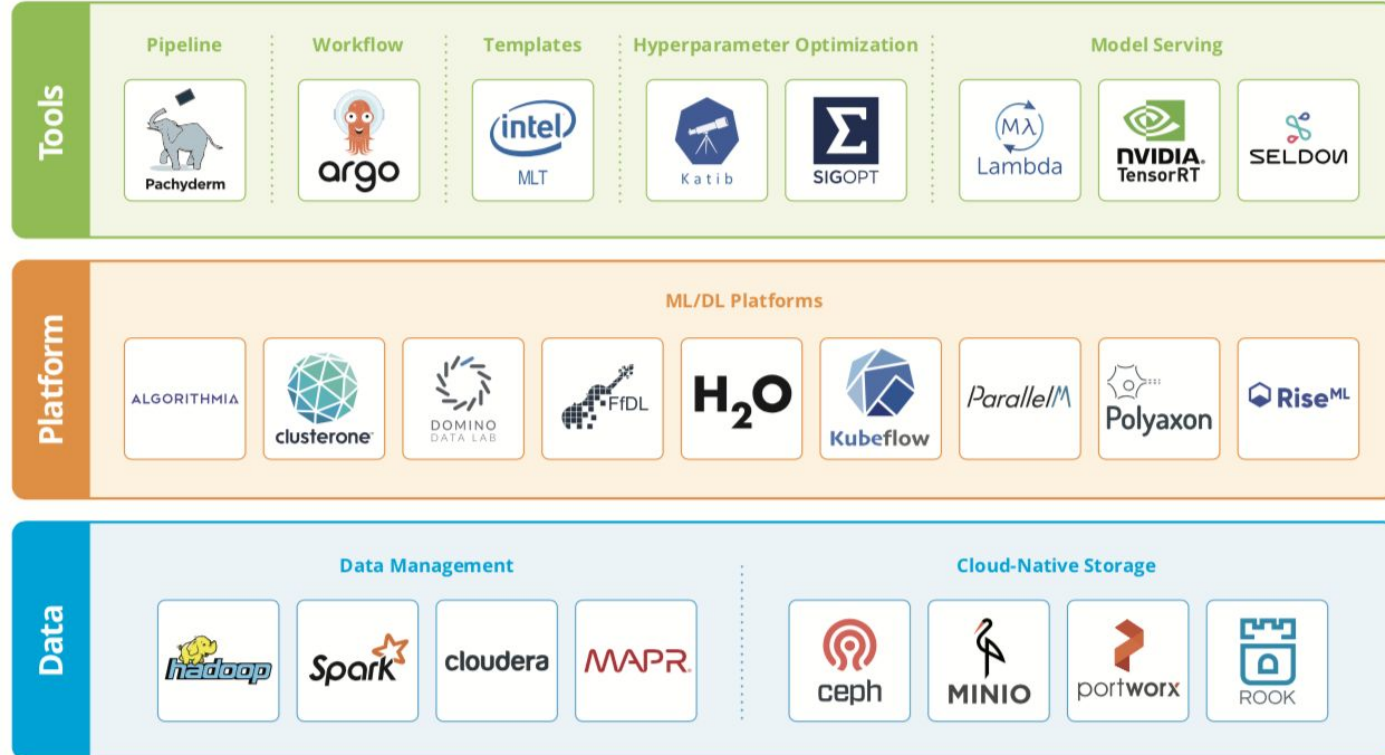EXPLORE/TRANSFORM — Cleaning, Anomaly Detection, PREP

MOVE/STORE — Reliable Data Flow, Infrastructure, Pipelines, ETL, Structured andUnstructured Data Storage

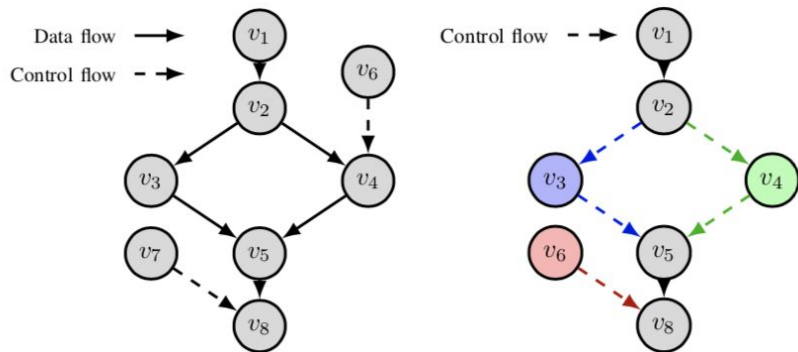COLLECT — Instrumentation, Logging, Sensors, External Data, User Generated Content

ML/AI Platform
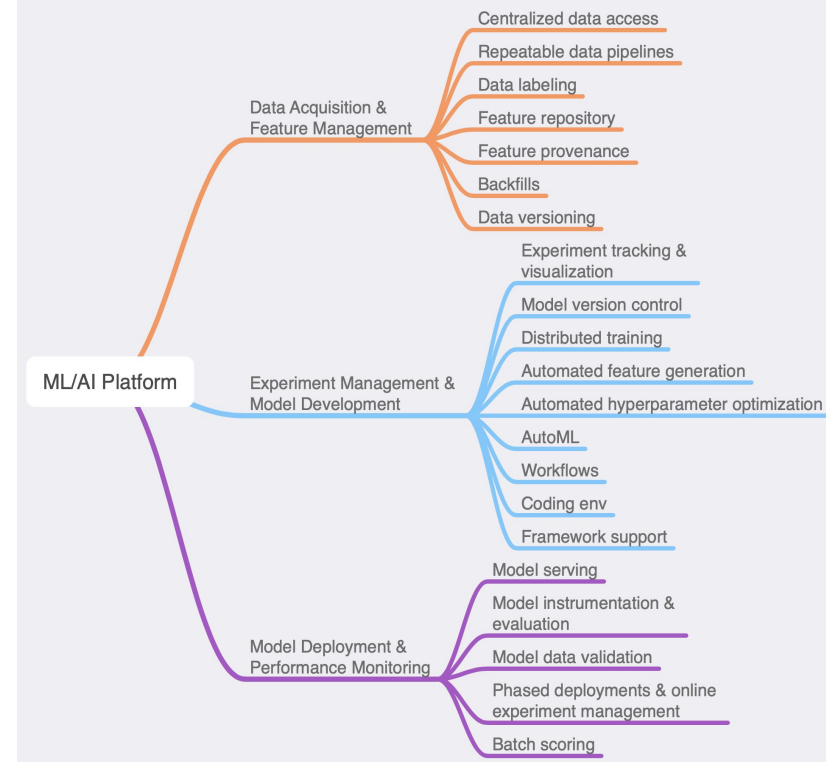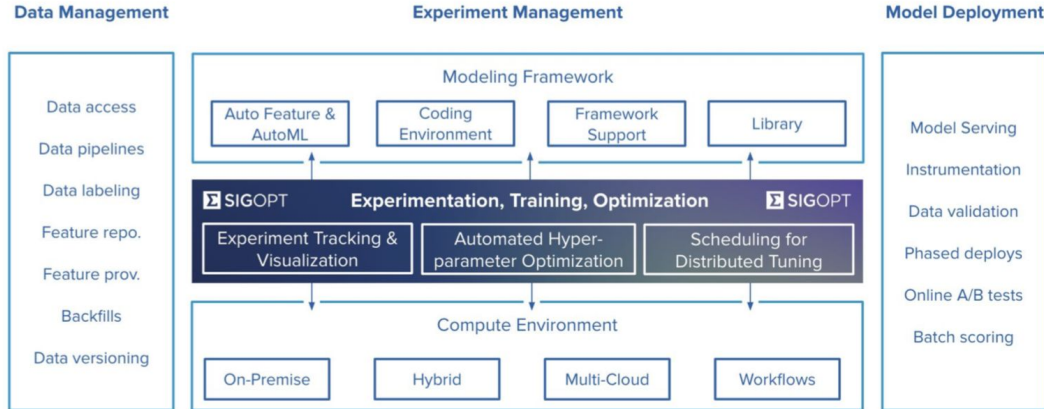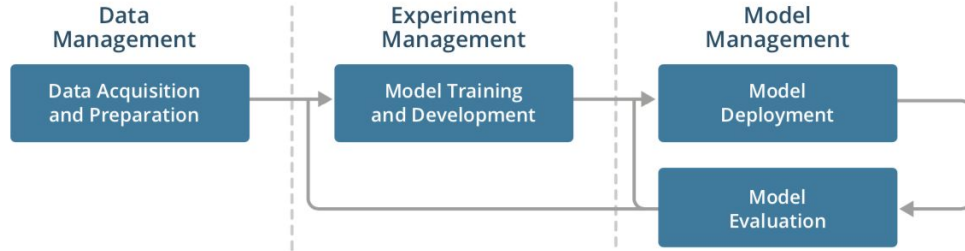
Kubernetes ML & DL Landscape
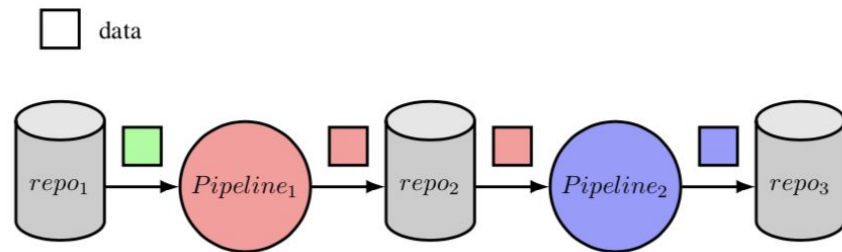
Workflow Management System

Task-oriented WMS

- **DAG:**
  node->computational task
  edge->task dependencies
- **Scheduling strategies, data management**
- **Solutions: Pegasus, Makeflow, Airflow**

Data-oriented WMS

- **DAG:**
  node->data repository
  edge->computational tasks
- **Better data provenance & versioning**
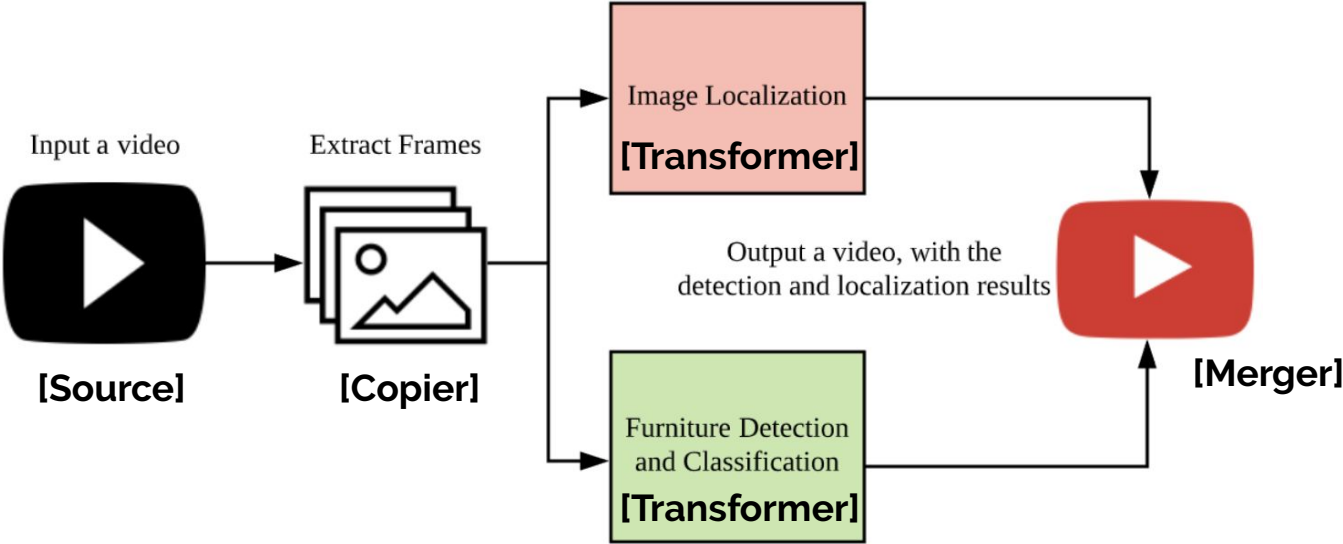- **Reproducible data pipelines**
- **Solutions: Pachyderm**

Use-Cases

People Counting

Images Collector [Source]	→	Model Serving [Transformer]	→	Result Collector [Sink]

Visual Localization

Input a video
[Source]

Extract Frames
[Copier]

Image Localization
[Transformer]

Furniture Detection and Classification
[Transformer]

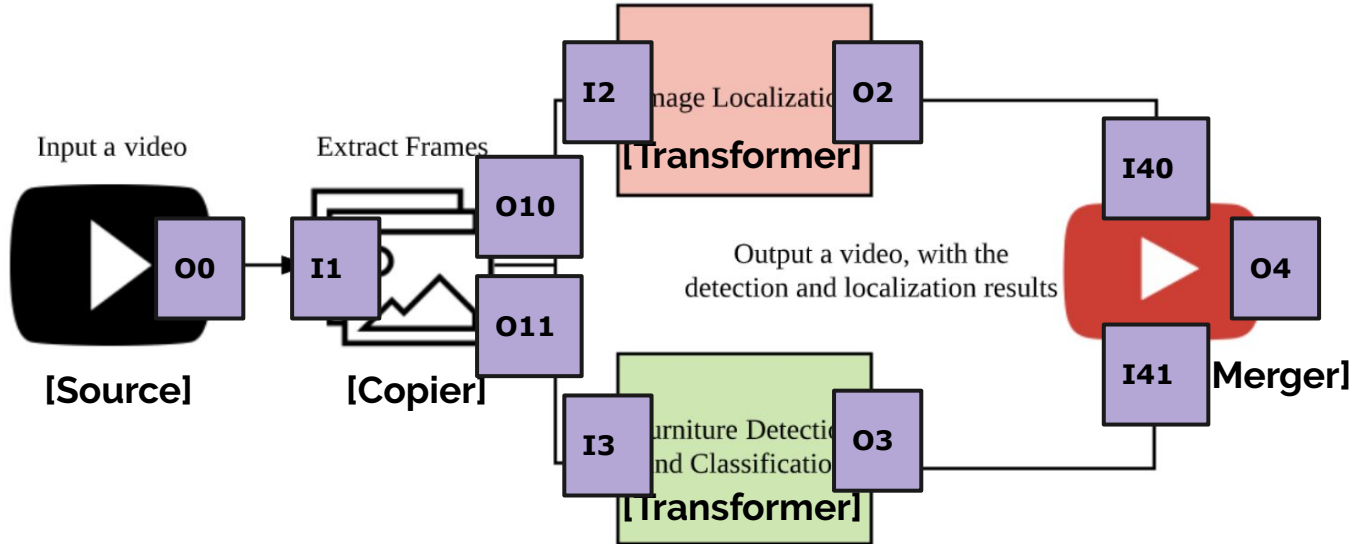Output a video, with the detection and localization results
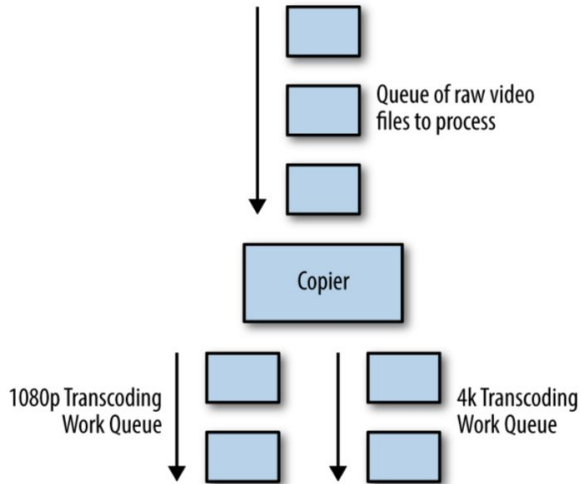
[Merger]
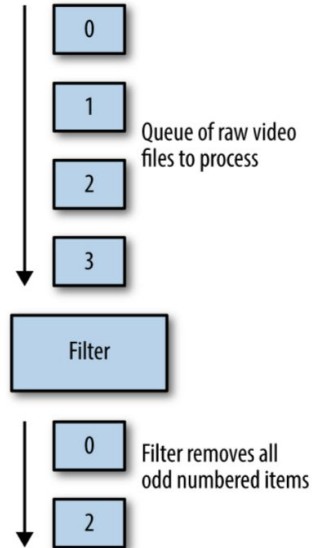
Use-Cases

People Counting
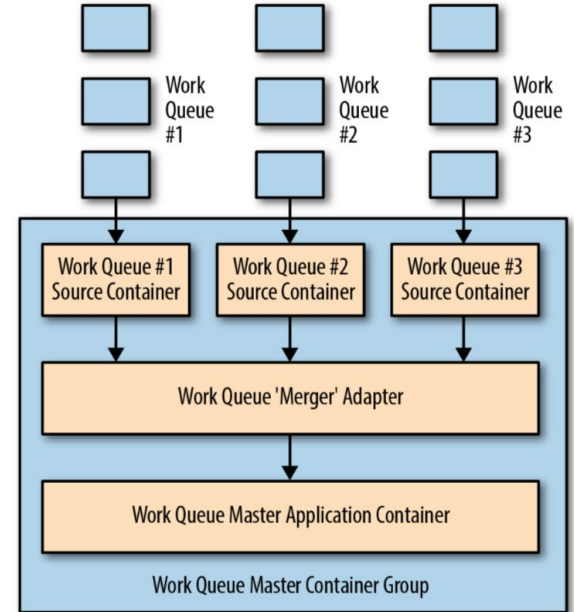

Visual Localization
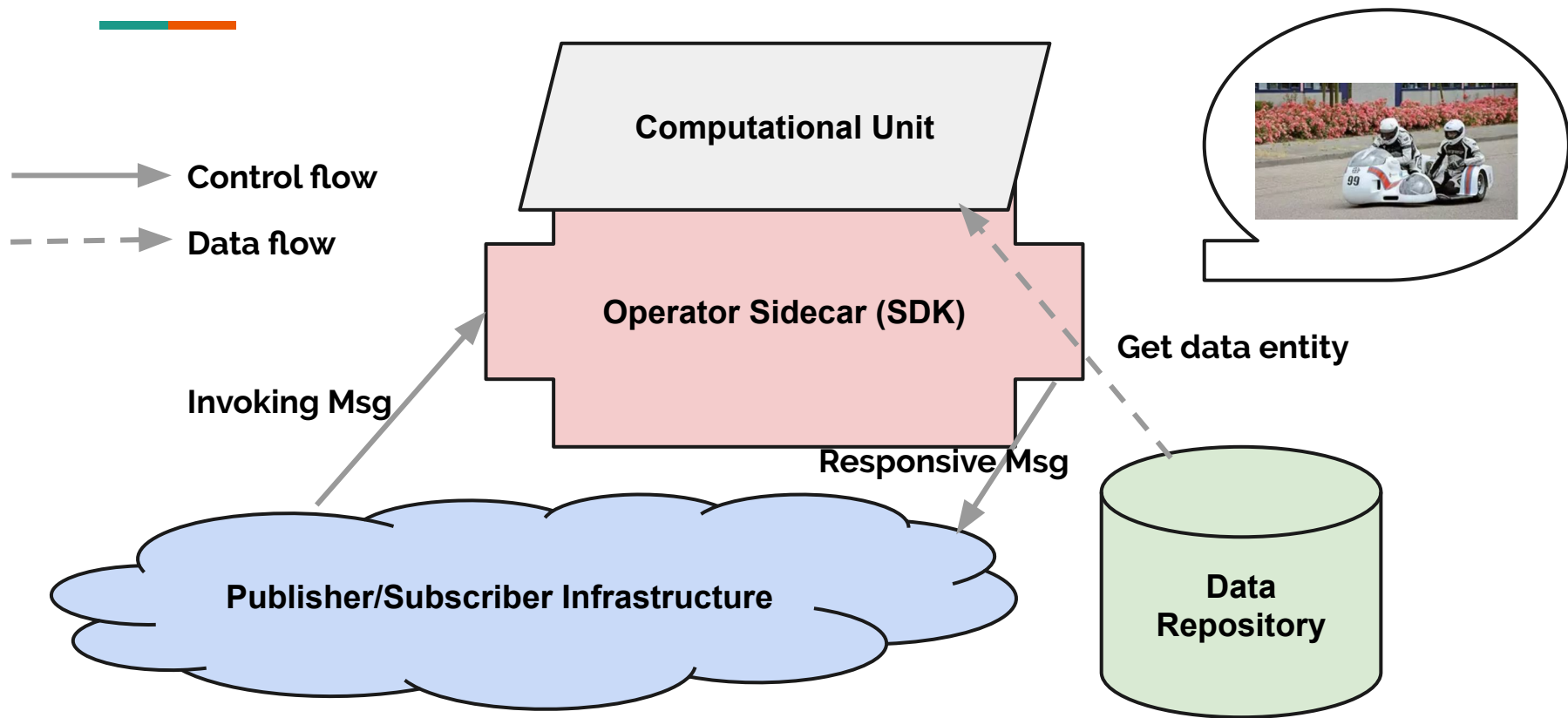
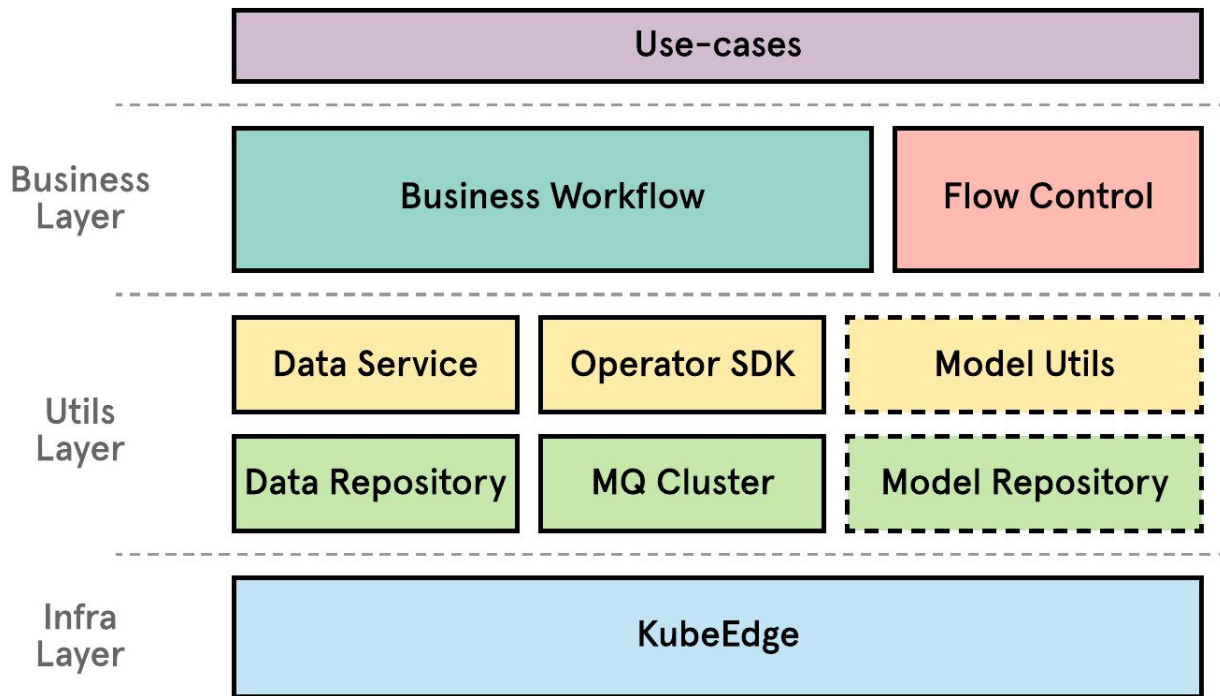# Workflow Operator

- Copier Pattern

- Filter Pattern

- Merger Pattern

Operator Sidecar

Operator Sidecar

References

- **KubeEdge Official Document**

- **Mitchell, Ryan, et al. "Exploration of Workflow Management Systems Emerging Features from Users Perspectives."**

- **Twiml, The Definitive Guide to Machine Learning Platform**

- **Twiml, Kubernetes for Machine Learning, Deep Learning & AI**