arm

Arm Research

SMARTER

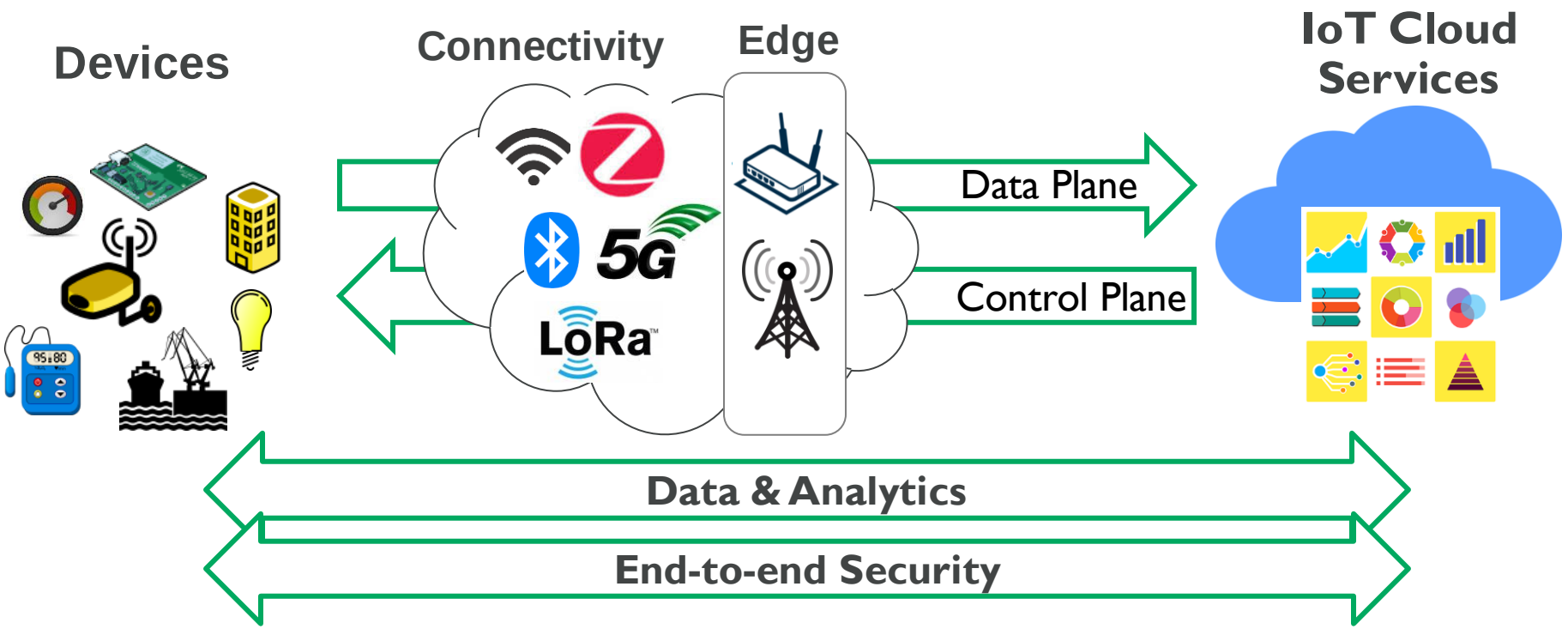
Experiences with Cloud Native on the Edge for Smarter Cities

Eric Van Hensbergen, Alex Ferreira, Chris Adeniyi-Jones, Edmund Grimely-Evans, Josh Minor, Mark Nutter, Luis E. Peña, Vasileios Laganakos, Jon Hermes

Computing on the Edge with Kubernetes Conference – October 2020

Motivation: A trillion devices from edge to cloud

Vision: To enable an intelligent and secure world with a trillion connected devices



Significant R&D investment to build industry-leading IoT systems and cloud services

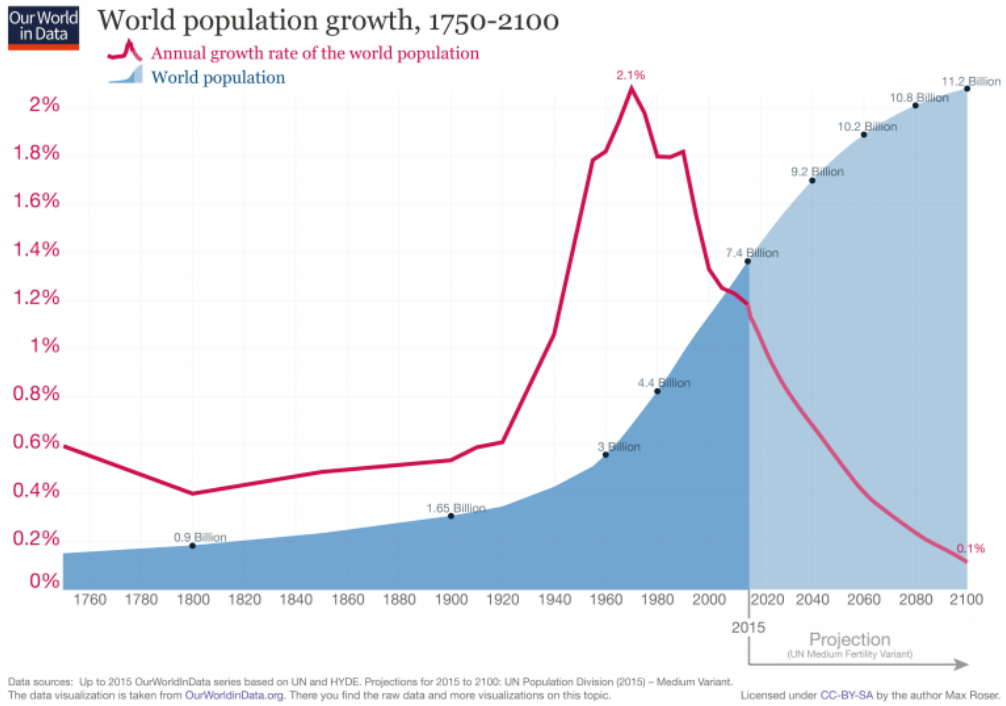
THE GLOBAL GOALS
For Sustainable Development

https://globalgoals.org
https://arm.com/company/sustainability/realizing-the-global-goals

World Population Trends

1750-2015 and projections until 2100

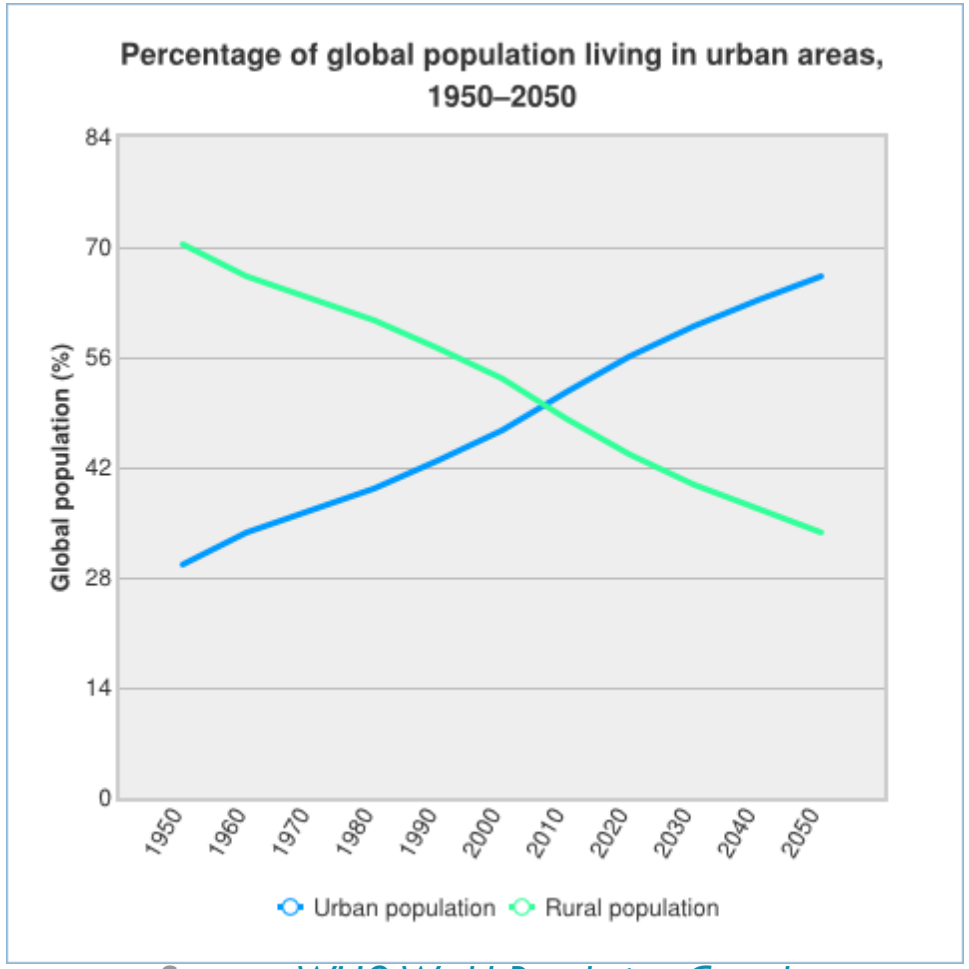
- Before 1800, world population growth rate < **1%**
- In 20th century, annual growth rate peaked at **2.1%** in 1962
- Since peaking, annual growth rate has gone down, with expected **0.1%** growth by 2100



Source: *Our World in Data*

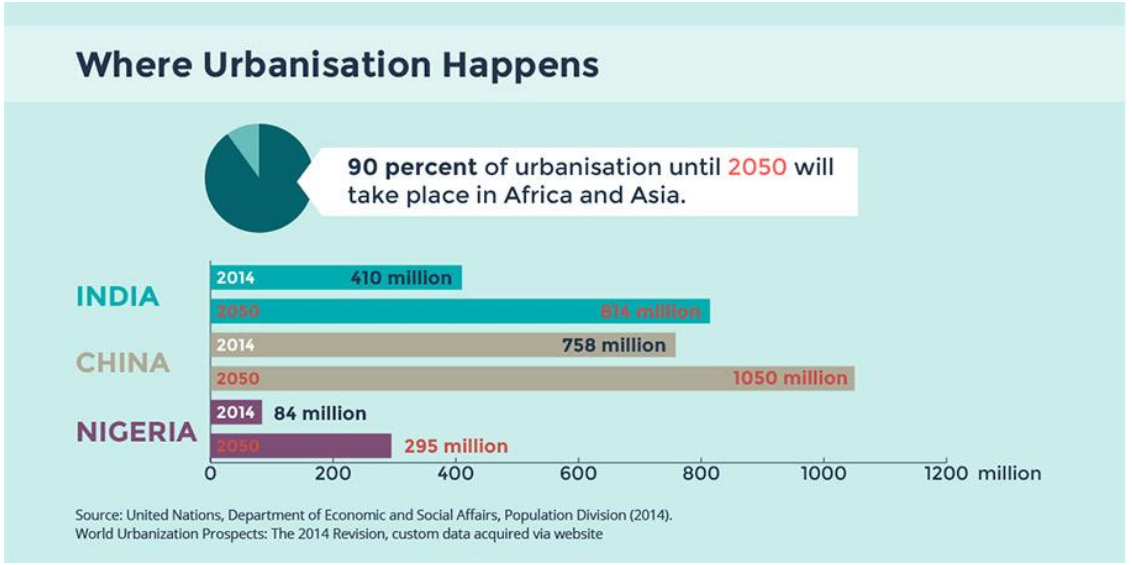
Urban Populations Trends

By 2030, urban areas are projected to house **60%** of people globally



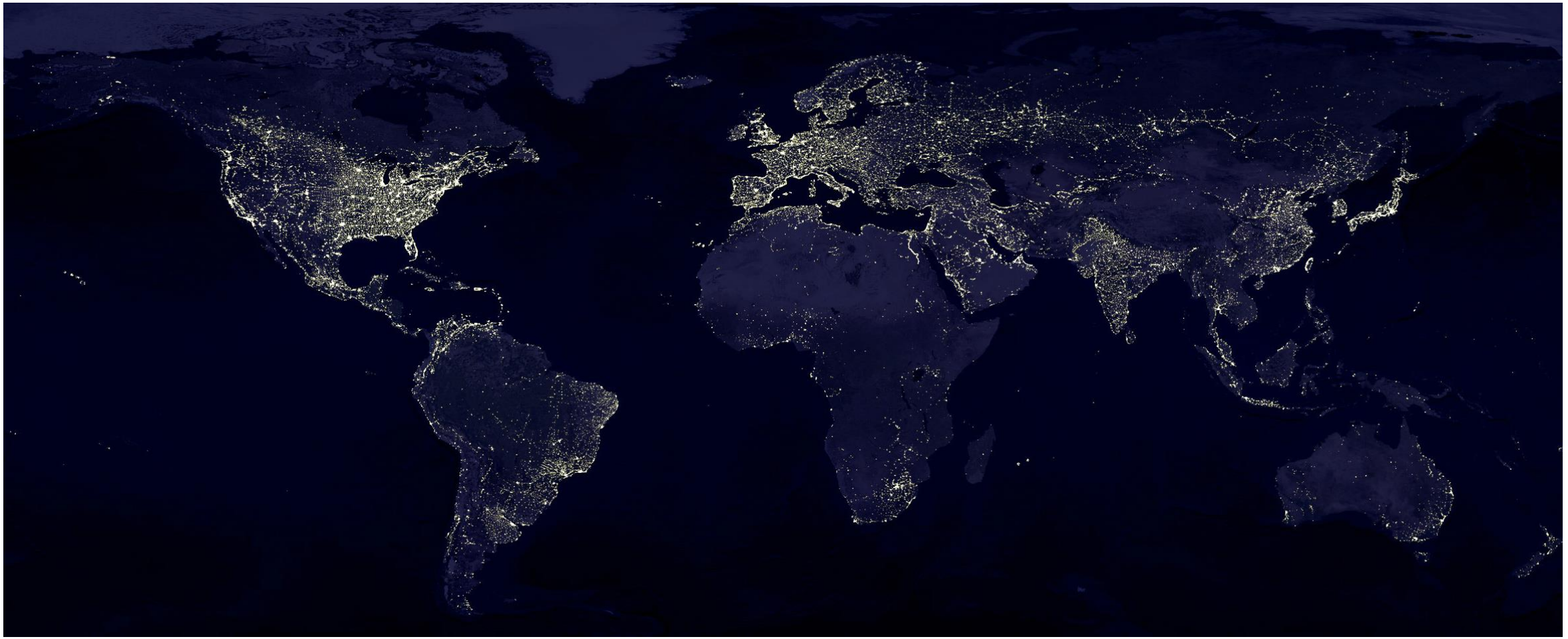
Source: *WHO World Population Growth*

- The world's urban population is growing fastest in Asia and Africa



Energy

Cities already consume **75%** of global energy resources and account for **80%** of emissions.



Source: *NASA Earth at Night*

Water

Developed World

- Infrastructure in expected to last +50 years
  - Up to 20% of clean water lost due to leakage annually ($27B)

Developing World

- Resources are scarce
  - Cape Town has recently run out of fresh drinking water
- Infrastructure in developing world often does not exist
  - WHO: Up to 4,000 people die every day to water-borne disease

Source: *Northeast Group LLC*

Smart City Objectives

"The effective integration of physical, digital and human systems in the built environment to deliver **sustainable, prosperous and inclusive future** for its citizens" – *British Standards Institute*

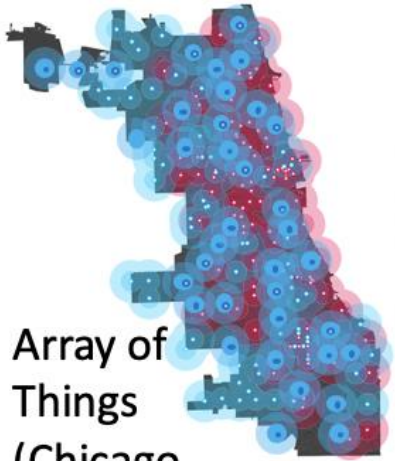
Smart Cities could deliver …

If cities across the globe today were to universally adopt, and deploy, smart city technology and services, what would the benefits be for citizens?"

Source: *Juniper Research*

AI@Edge

Building an open infrastructure for Edge-to-Cloud AI
for society and infrastructure: health, quality of life, environmental protection, safety



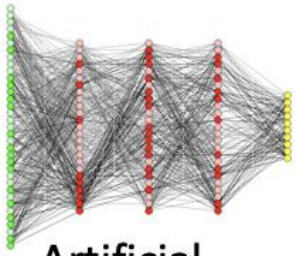
Array of Things (Chicago, .uk, .tw, etc.)

Air Quality

Camera, Mic, LIDAR, etc.

Edge Computing

Google TPU



Artificial Intelligence

AI & HPC/Cloud

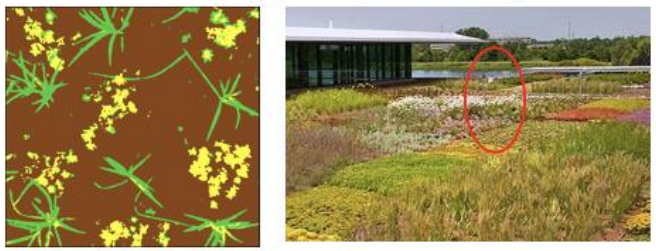
Waggle Platform

Missing:

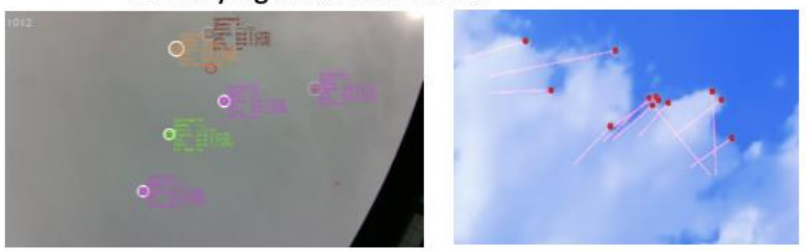
- The open programming framework for Edge-to-HPC/cloud
- A public-good global testbed to build AI to improve society
- National education curriculum for sensors and an AI future

50 frames to detect flooding and segment image

Understanding plant species and growth



Identifying what is in the air



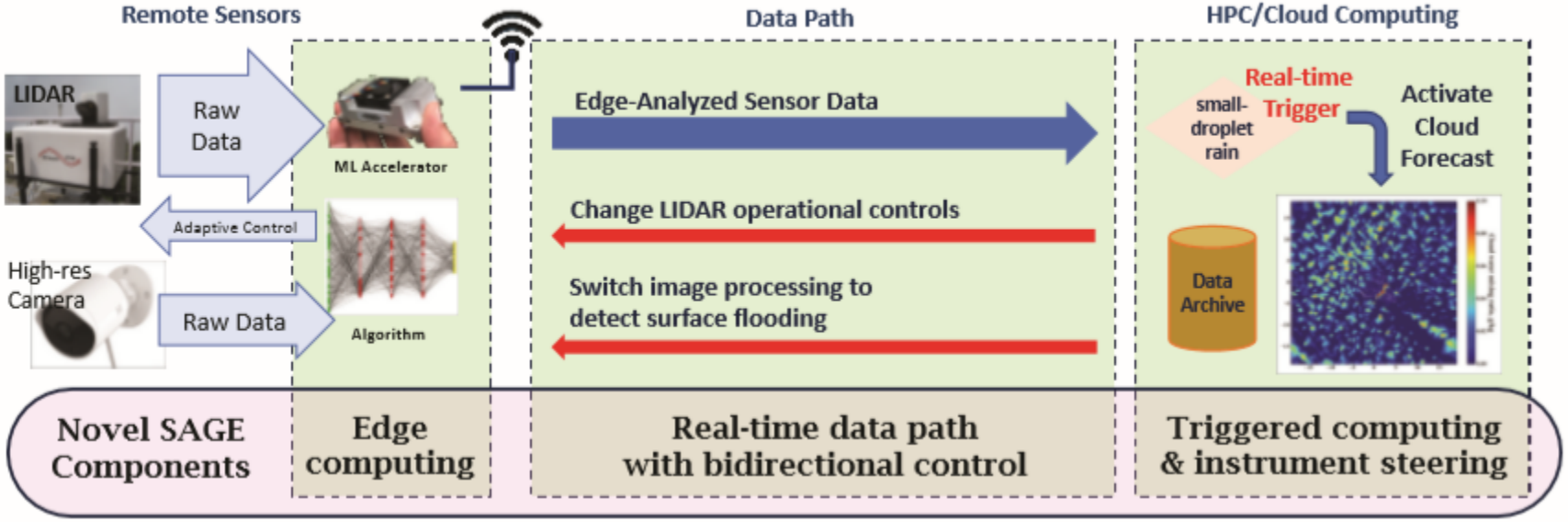
the power grid

Pete Beckman, Charlie Catlett, Nicola Ferrier, Rajesh Sankaran: Argonne National Laboratory

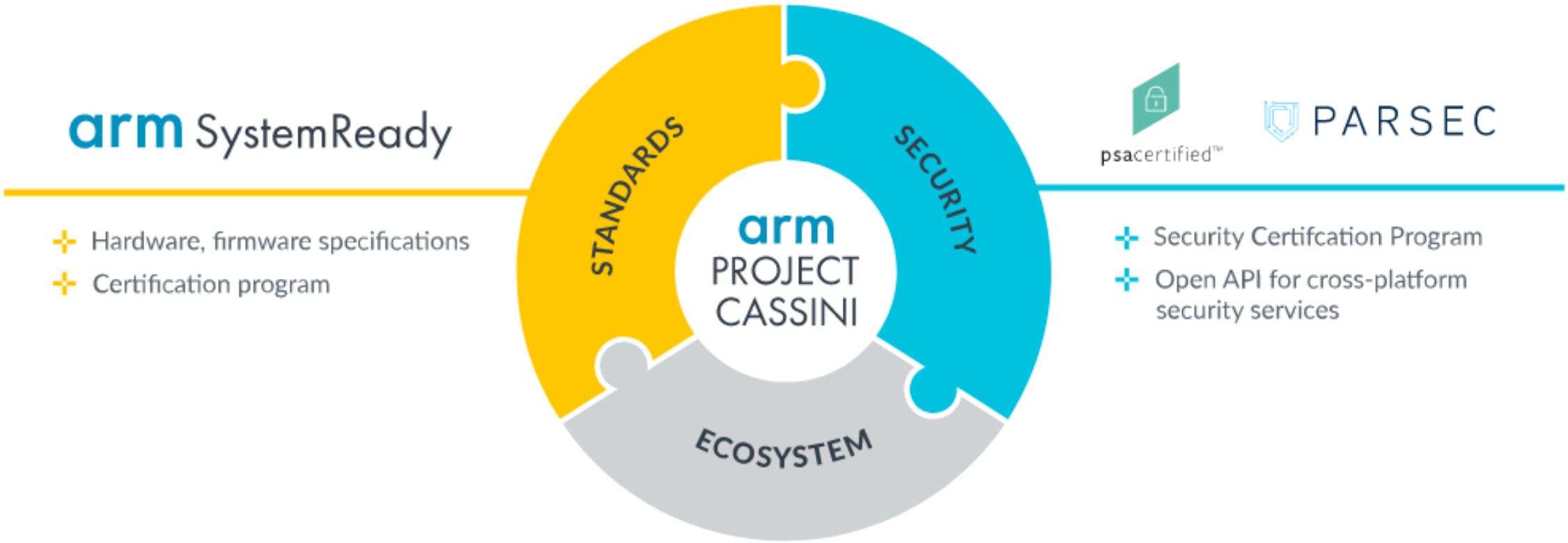
NSF SAGE

http://sagecontinuum.org/

Base Type	Common Sensors (all nodes)
Air Quality	PM2.5, EPA criteria pollutants
Environment	Meteorology, Light (ambient, IR, UV), Sound (10 octaves), Vibration, Magnetic field
Imaging	HD sky, Ground cameras
Extensions	**Capabilities & Features**
HD Air Quality	Airquality, Meteorology, 3D wind, Rain sensor, IR camera
Seismic	Meteorology, Seismic, Vibration, HD cameras w/ 360° View
Sky	Meteorology, Wide angle HD sky cameras, Lightning sensor
Wildfire	Meteorology, Anemometer, Air-quality, HD cameras for 360° view
Urban AoT	Airquality, Meteorology, Sky and street cameras, Microphone
Wildlife	Meteorology, Microphones, HD cameras for 360° view
Doppler Lidar	Range- and time-resolved measure-ments for line-of-sight air velocity, attenuated aerosol backscatter
μ-Pulse Lidar	Altitude of clouds, detection of atmospheric aerosols
μ-Rain Radar	Measurement of rain rate, liquid water content and drop size distribution, near ground to 3000 m
Wind Profiler	Radar wind and precip profiles through the boundary layer
ETCE	Eddy covarariance turbulence exchange between soil, vegetation, and atmosphere

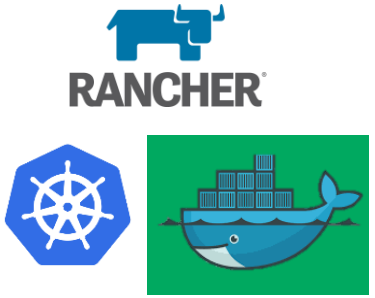


arm

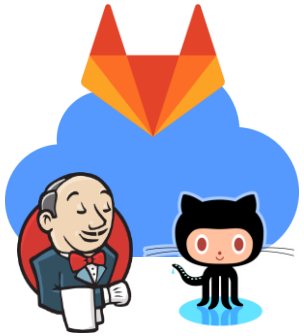


https://arm.com/project-cassini

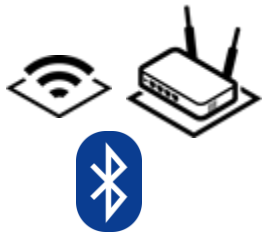
Cloud Native for the Edge



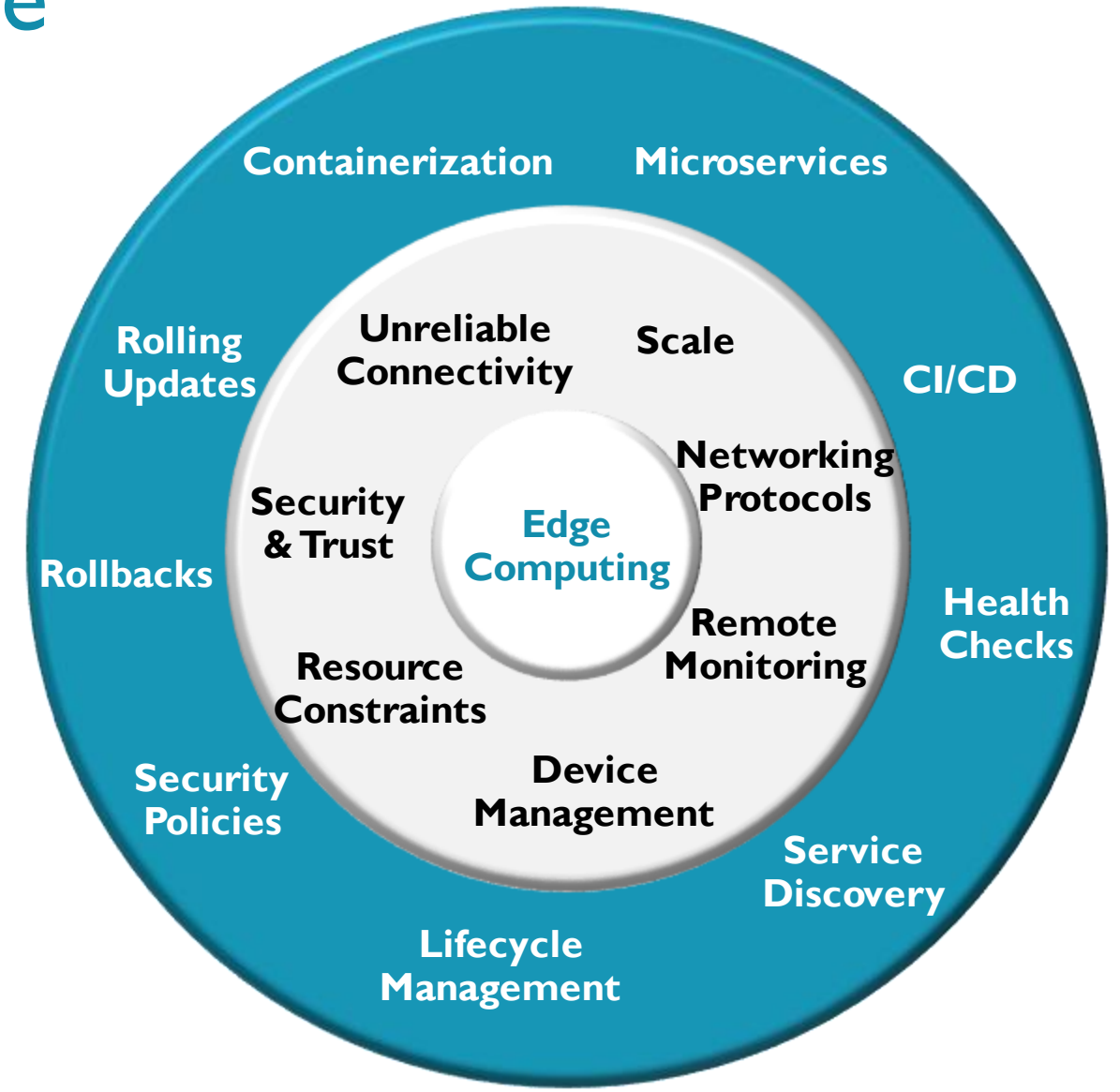
Base development environment on existing cloud-native methods, ecosystem, APIs, and tool chains.



Bring server-style modern application management, continuous integration and deployment to the edge.



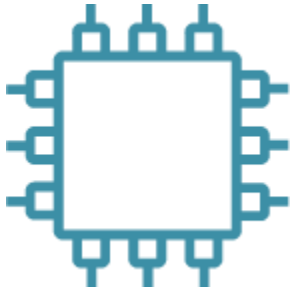
Extend tools & technology where necessary to optimize for IoT and Edge and target wide-scale versus cluster deployment.



arm

What's different on the Edge?

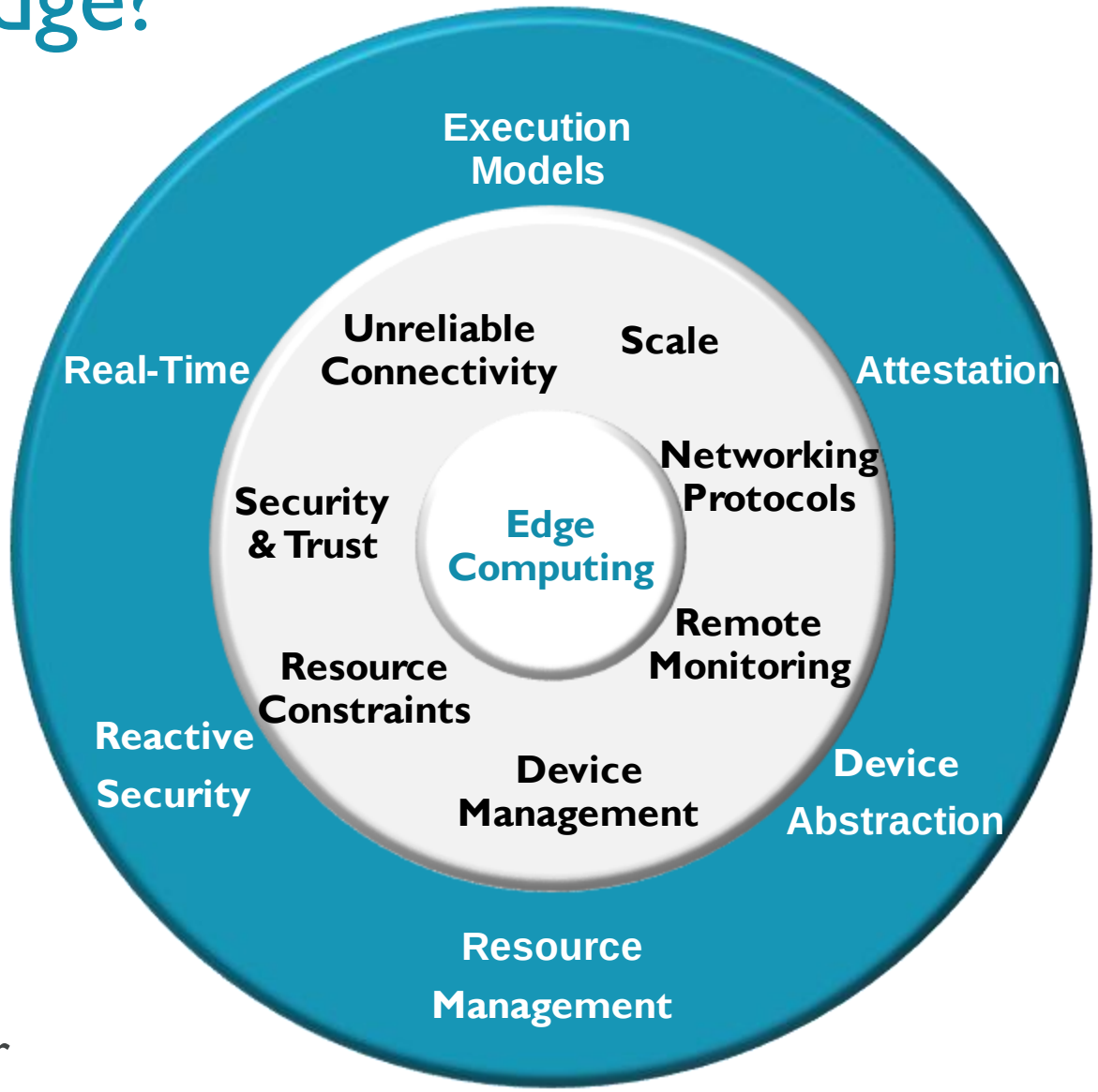
Different attack vectors require a more complete security model and potentially different approaches to securing multi-tenancy



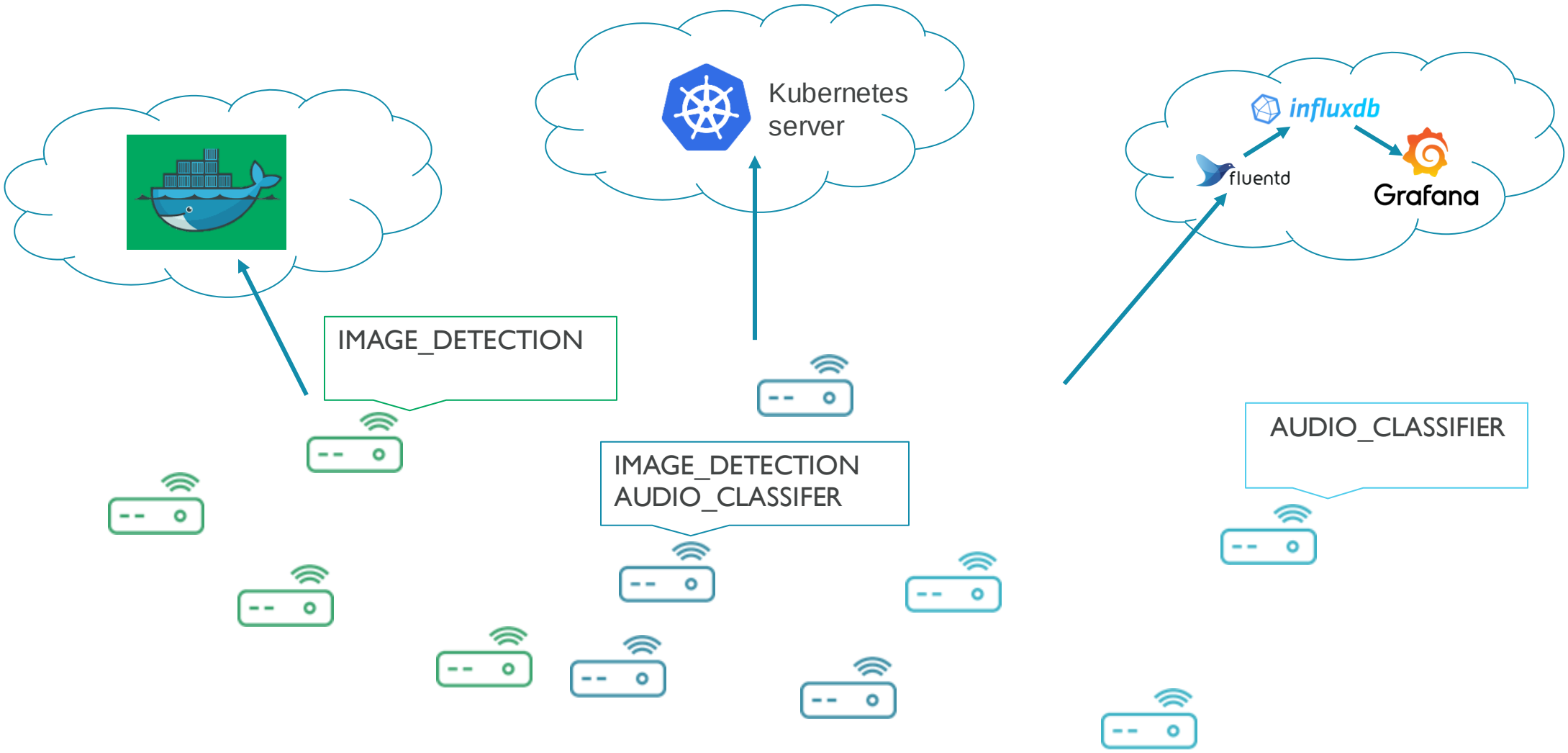
More constrained node environments, limited virtualization for devices, with no ability to scale-out to other nodes in cluster

Different connectivity model, potentially unreliable, constrained metered bandwidth, and local-first service model

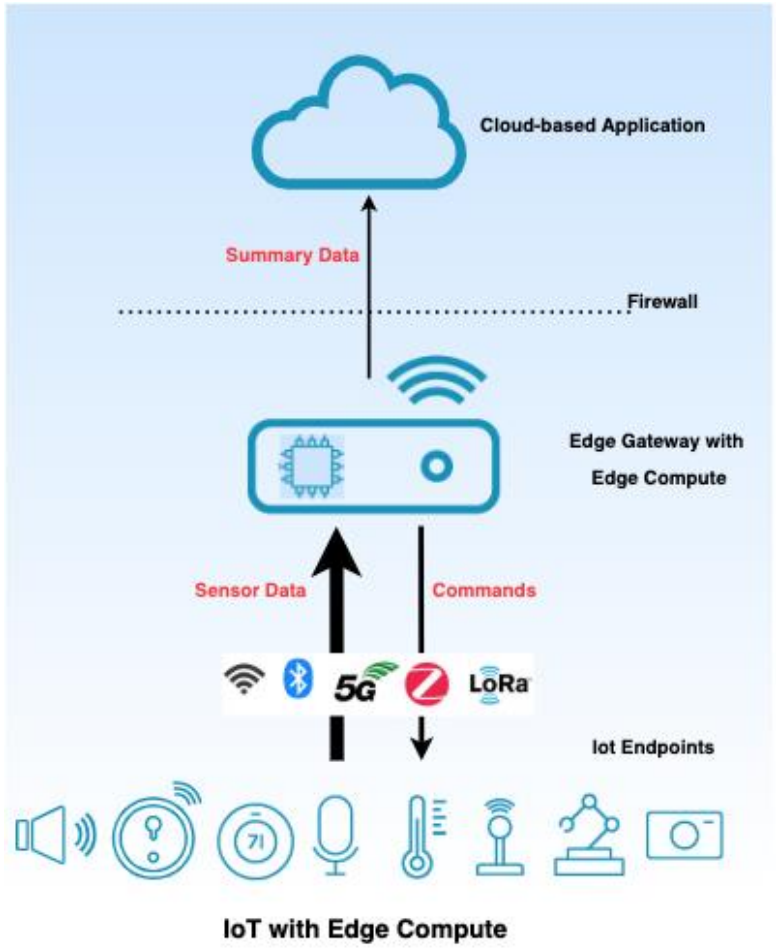
Physical sensors which must be allocated, multiplexed, secured, and managed – something that just doesn't happen in a cloud data center



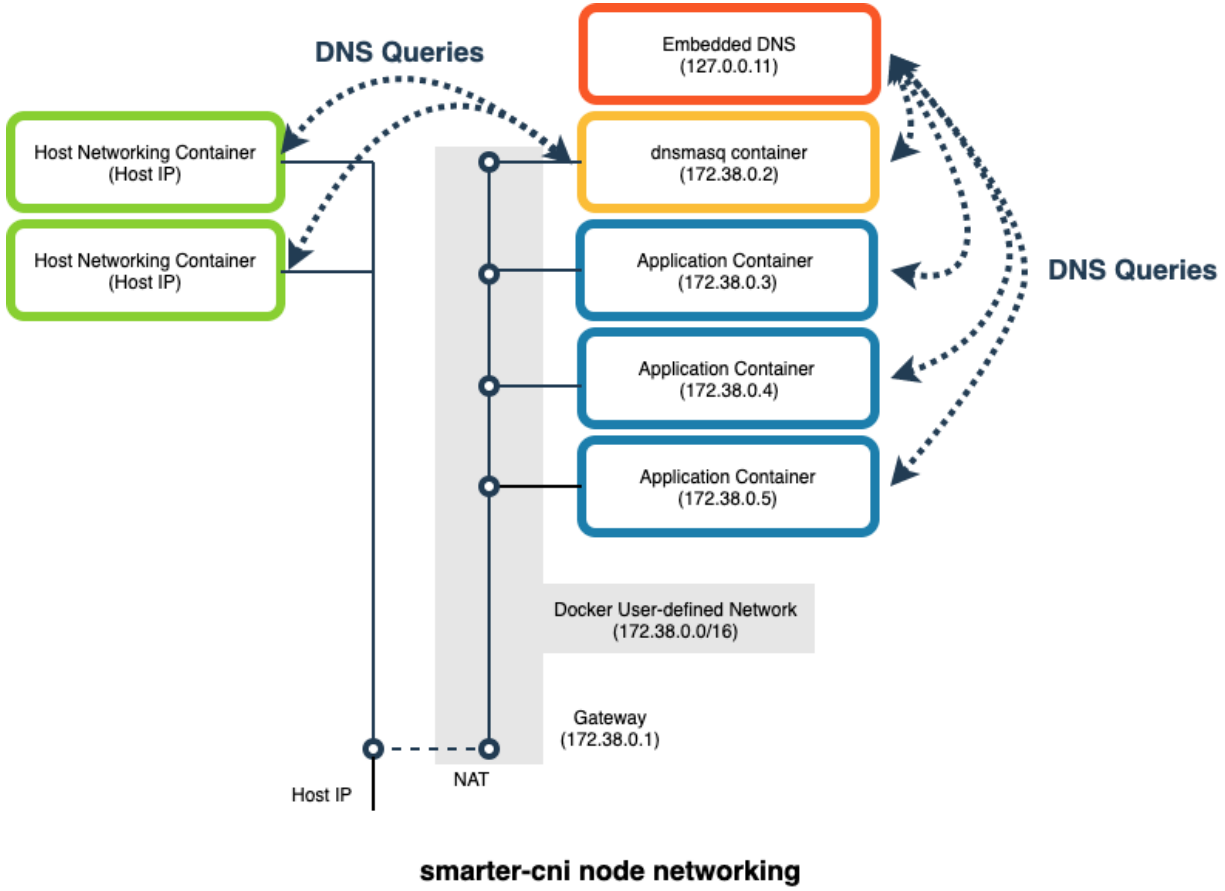
Kubernetes – provisioning and orchestration



Edge Networking



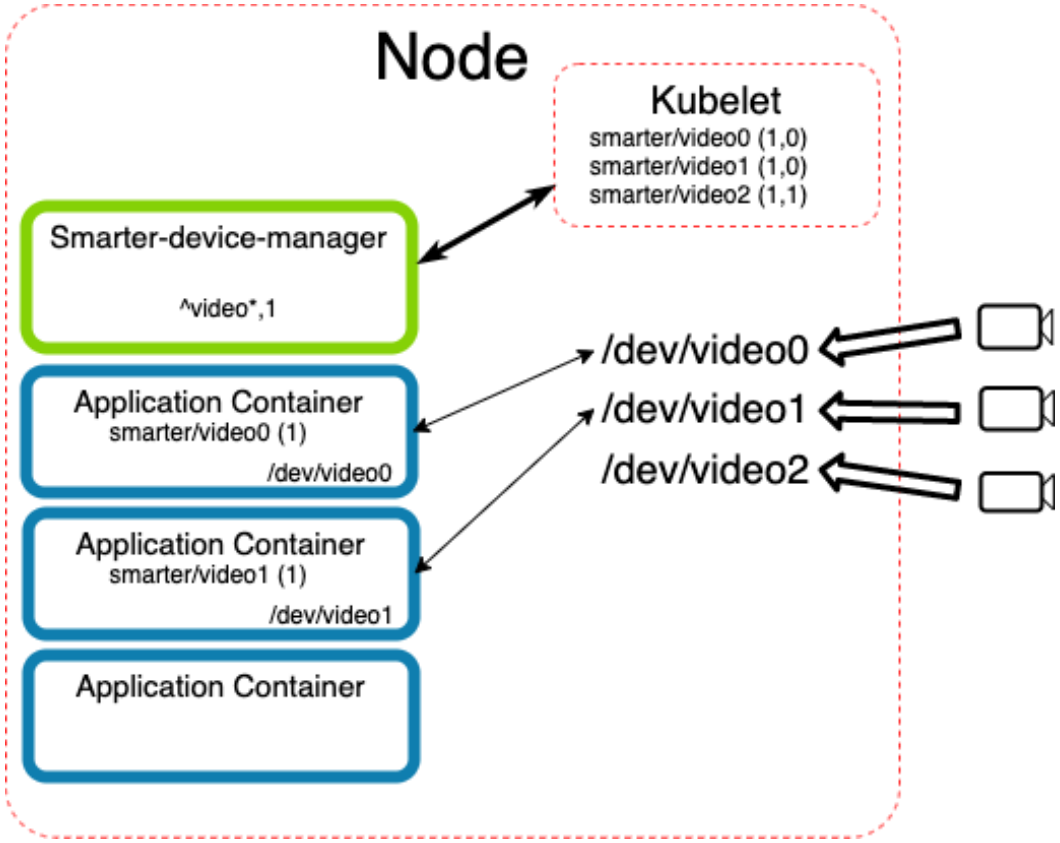
IoT with Edge Compute



smarter-cni node networking

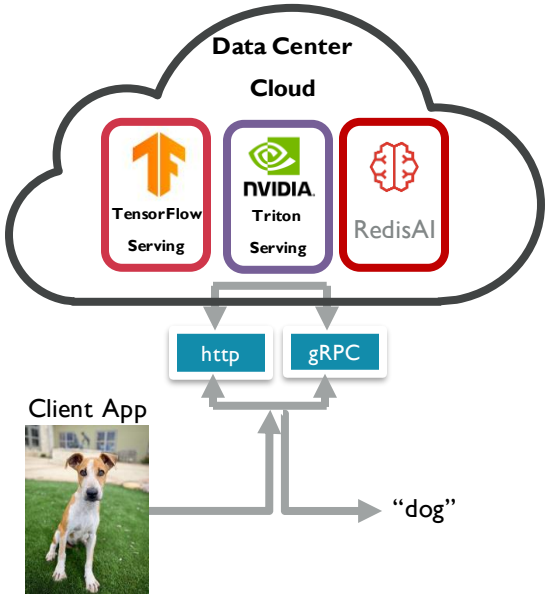
Device Management

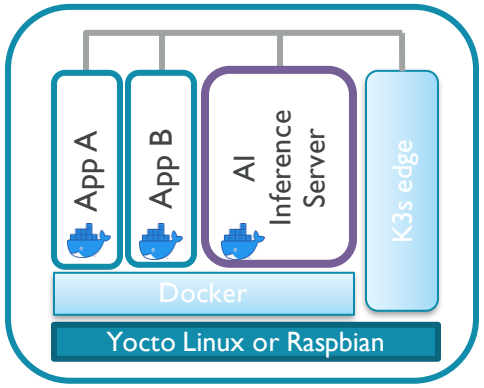
```
apiVersion: apps/v1
kind: DaemonSet
metadata:
  name: pulseaudio
  labels:
    k8s-app: pulseaudio
spec:
  selector:
    matchLabels:
      name: pulseaudio
  template:
    metadata:
      labels:
        name: pulseaudio
    spec:
      nodeSelector:
        pulse: "enabled"
      hostname: pulse
      containers:
      - name: pulseaudio
        imagePullPolicy: IfNotPresent
        image: registry.gitlab.com/arm-research/smarter/edge-workloads/pulseaudio:v1.0.0
        resources:
          limits:
            cpu: 350m
            memory: 100Mi
            smarter-devices/snd: 1
          requests:
            cpu: 250m
            memory: 100Mi
            smarter-devices/snd: 1
      terminationGracePeriodSeconds: 15
```



AI Inference Serving

Adopt and extend cloud-native best practices to edge

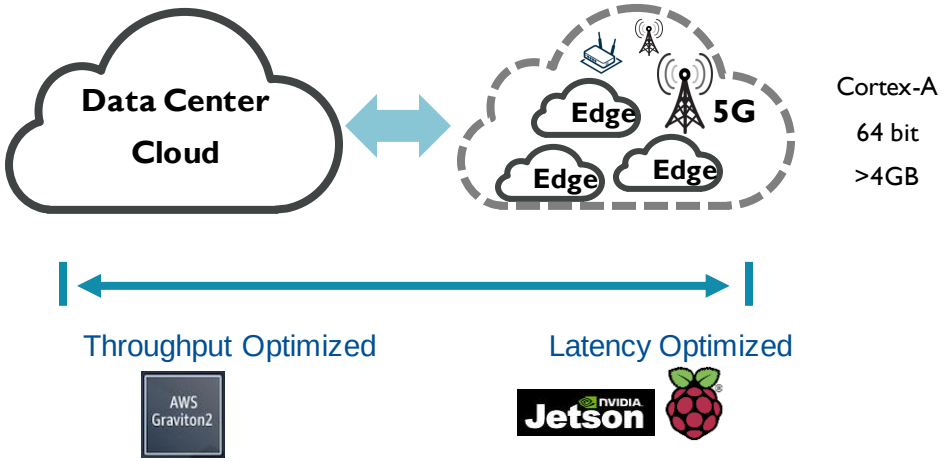




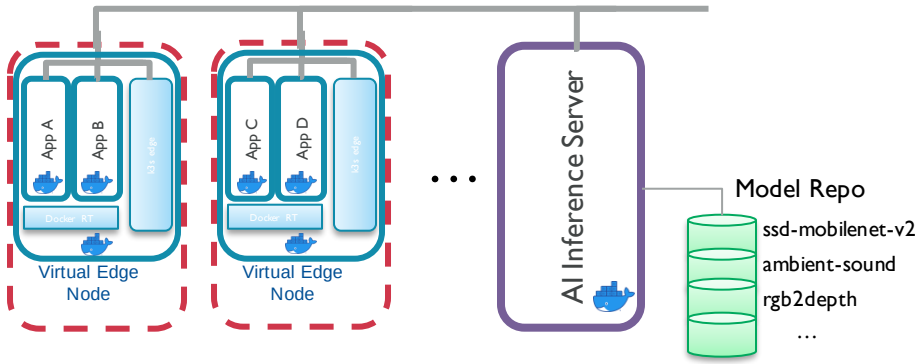
Physical Edge Node

Outcomes

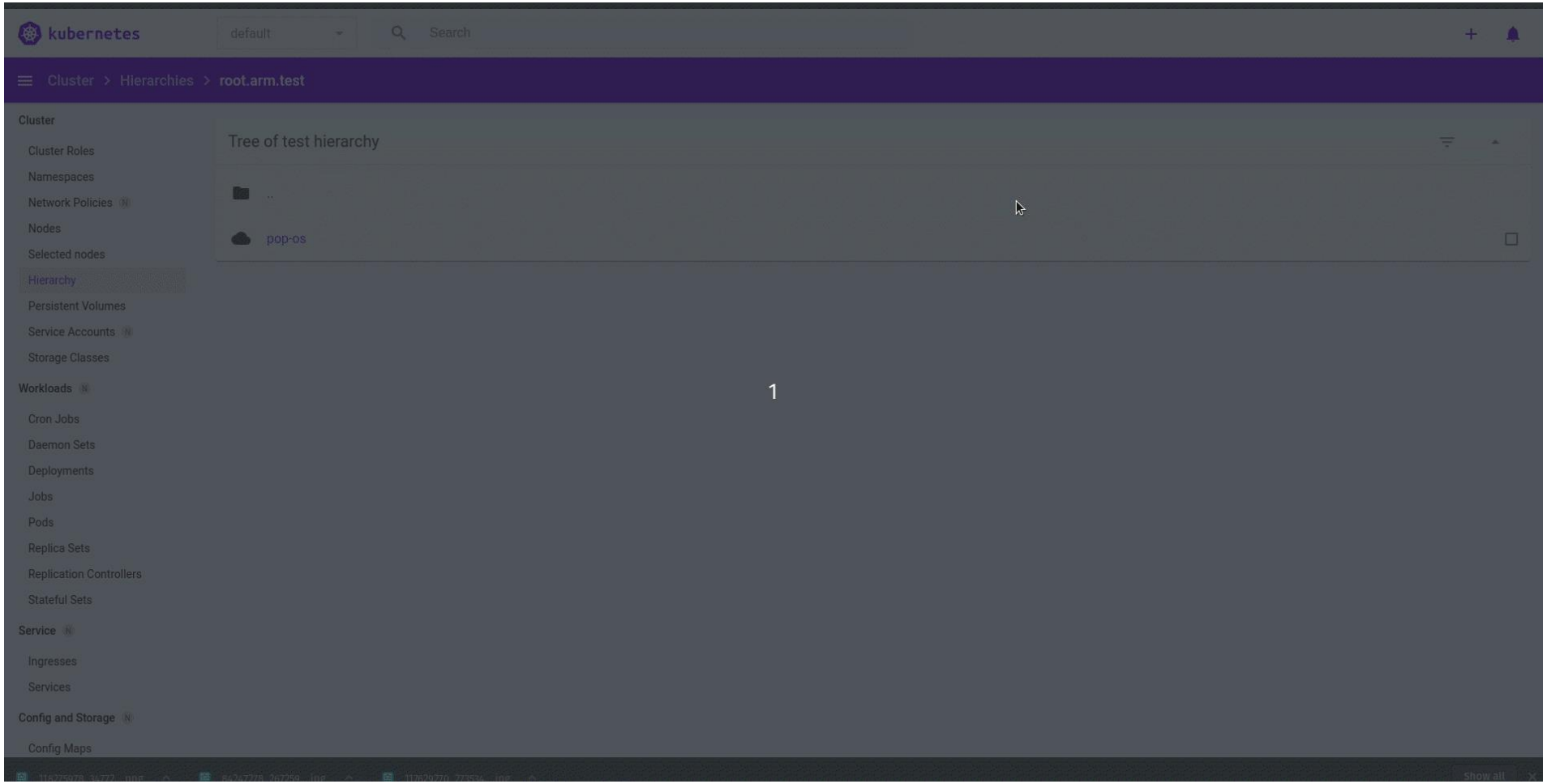
1. Flexibly move AI Inference between cloud and edge



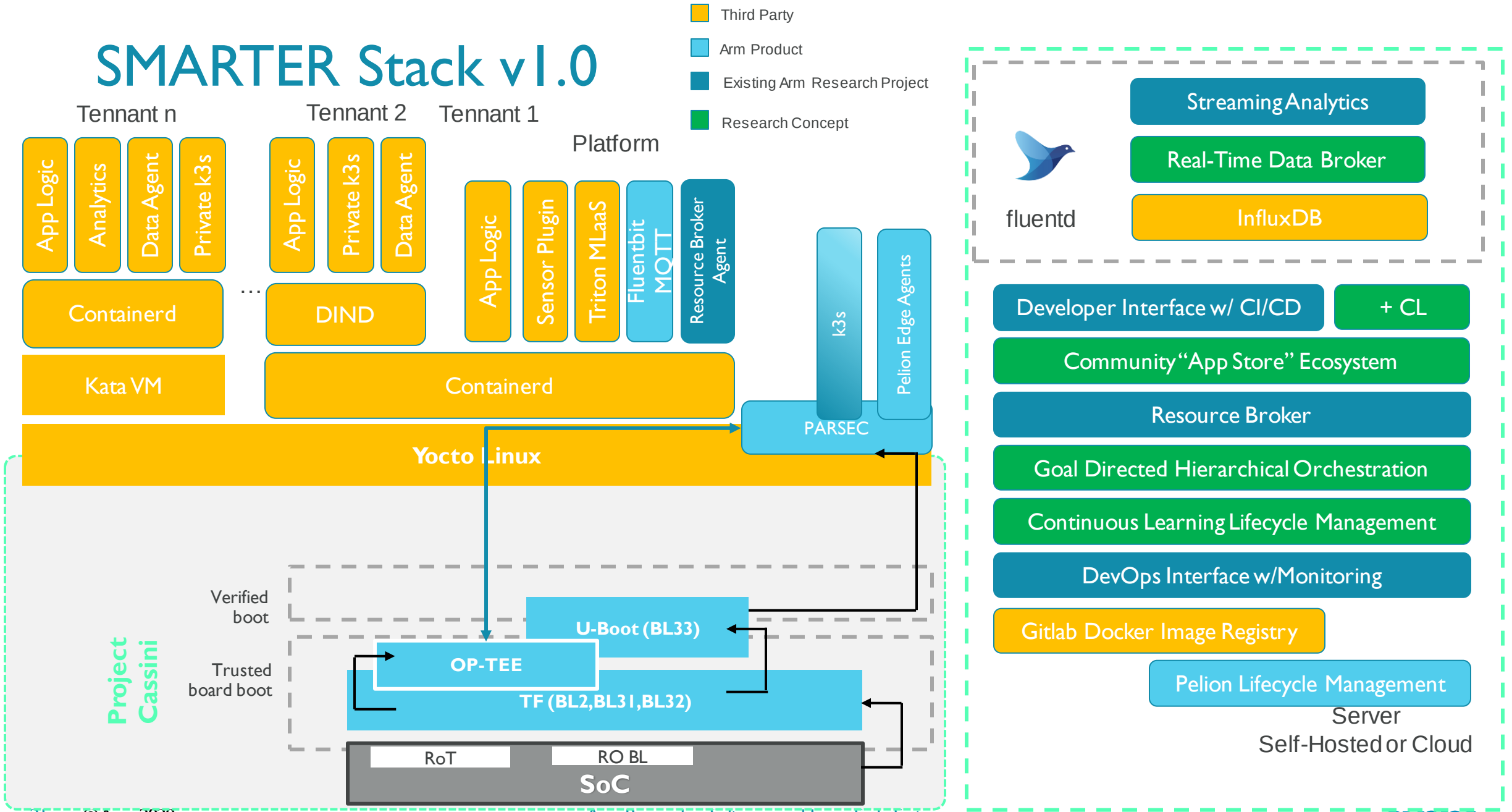
2. Efficiently share accelerators among edge applications



Extend dashboard for geographic and hierarchy



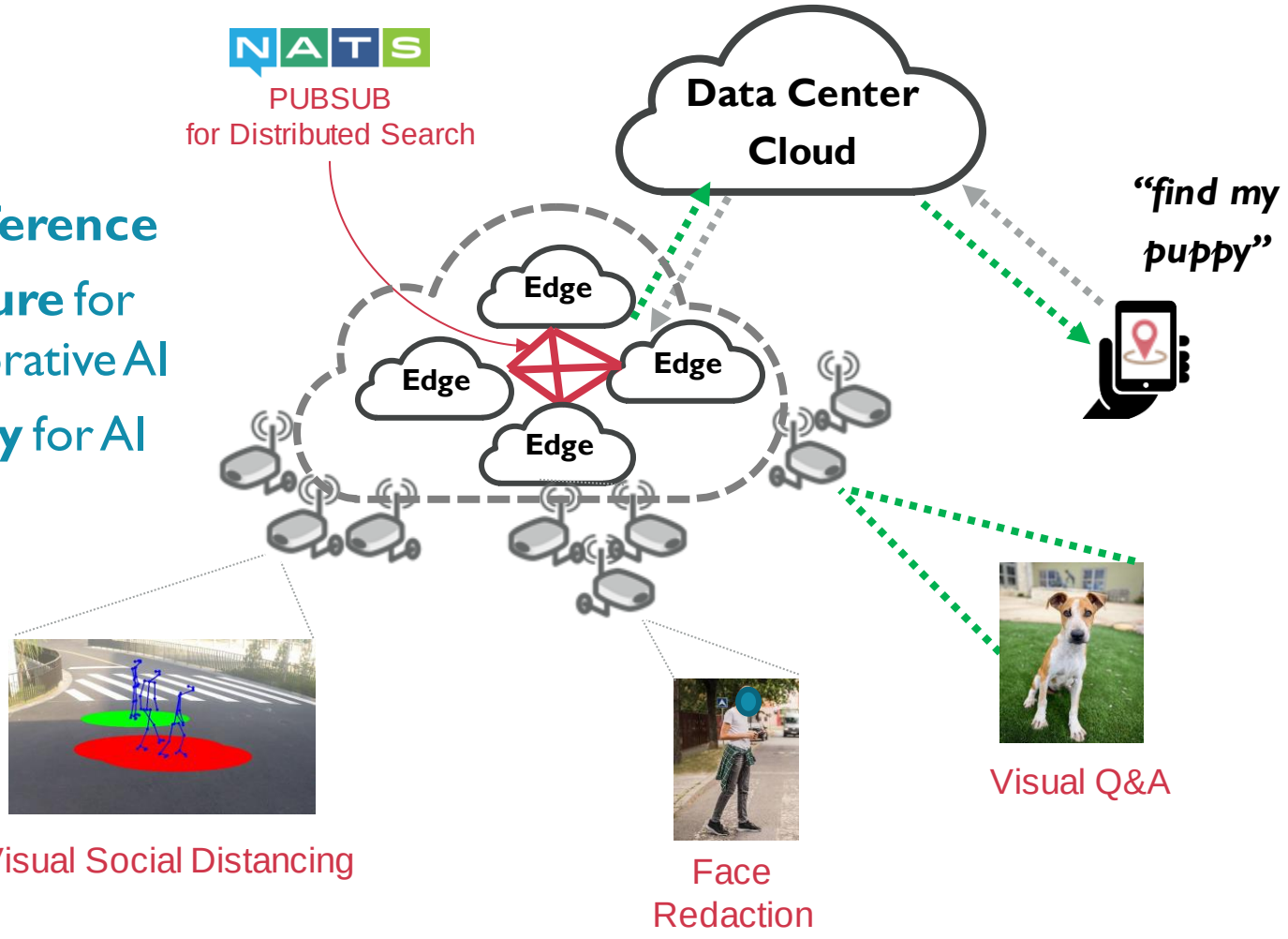
SMARTER Stack v1.0



Upcoming Work

Distributed Arm Compute for AI Inference

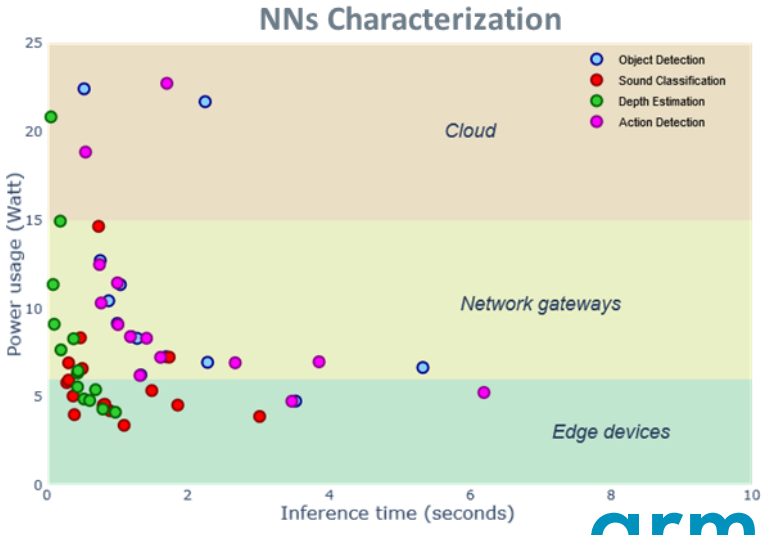
- Efficient **streaming AI inference**
- **Distributed Infrastructure** for Secure, Respectful, Collaborative AI
- **Hardware Multi Tenancy** for AI accelerators



Long-term goals

- Realize this Distributed Compute System, that makes use of the available Arm IP, from the Edge to the Cloud.
  - Hardware-Software co-design.
  - Software implementation on existing hardware is the 1st step.

- Improve Arm IP and Arm Tooling.
  - Model Characterization – help partners identify the right Arm IP.
  - Make it easy to develop for Arm – high-performing Apps.
  - Increased focus on serverless / function as a service.
  - Optimize communication inside and outside node.
  - Expand intersection of project from A-class to also integrating R-class and M-class systems.



For additional information….

- http://getsmarter.io
  - Example project contains cloud and client-side instructions to deploy pedestrian/car detector and audio classifier

- Array of Things and Waggle Platform (inspiration for base architecture and applications)

- Arm Project Cassini
- Arm Pelion Edge

SMARTER is an open-source reference, please comment, contribute, extend!

arm