DECICE

Compute Continuum Tutorial

Dirk Pleiter (KTH)

Vienna, November 2023

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Intended Learning Outcomes

DECICE

1. Describe the compute continuum concept and selected application areas
2. Analyse challenges of a compute continuum architectures and its components
3. Implement a simple workflow using containers deployed on cloud and edge nodes

Agenda

13:30 - 13:55	Compute continuum and 1-2 use cases
13:55 - 14:15	Introduction to key technologies I: Containers and container orchestration systems
14:15 - 15:00	Hands-on I: Container building and deployment in homogeneous environments
15:00 - 15:30	*Break*
15:30 - 15:35	Introduction to key technologies II: KubeEdge
15:35 - 16:00	Hands-on II: Container building and deployment in heterogeneous environments
16:00 - 16:10	Introduction to key technologies III: Prometheus
16:10 - 16:45	Hands-on III: Workflow implementation based on prepared components
16:45 - 17:00	Wrap-up

Compute Continuum: Definitions and Terminology

- Definition of **Compute Continuum** used here: [S. Moreschini et al., 2022]

  A continuum of resources available from the Edge until Cloud/HPC data centres

  - Resources = compute, storage, network
- **Cloud computing** [NIST SP 800-145, 2011]
  - NIST: “Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.”
  - Essential characteristics
    - On-demand self-service
    - Broad network access
    - Resource pooling = centralised data centres
    - Rapid elasticity

Compute Continuum: Definitions and Terminology (cont.)

DECICE

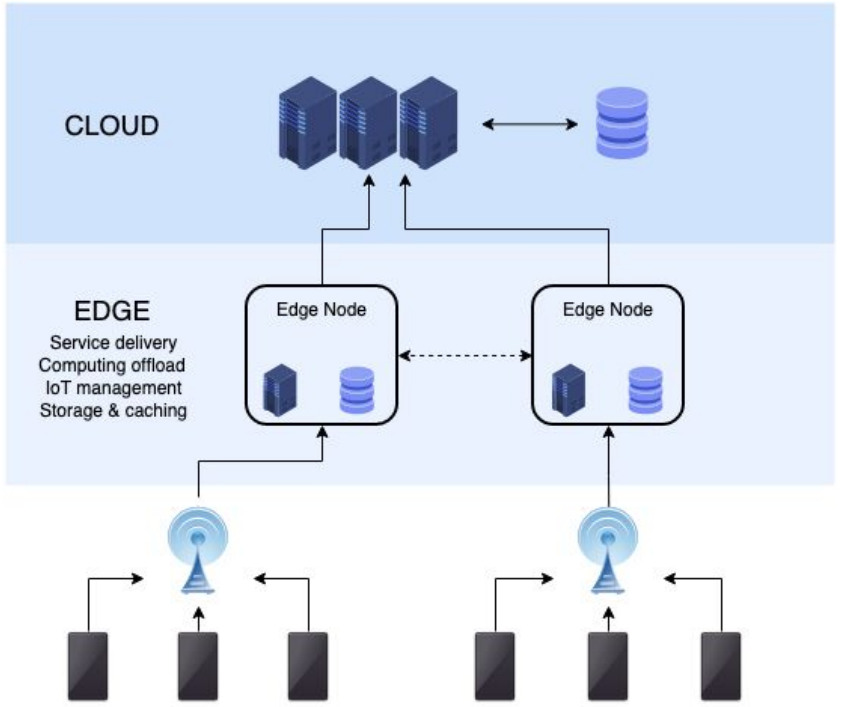
- **Edge computing**

  [Weisong Shi et al., 2016]

  A set of resources at the edge of the network with downstream connection to IoT resources and upstream connection to Cloud computing resources

  - Key role: Gateways to connect (a potentially large number of very simple) IoT devices to Cloud services

[Wikimedia Commons, 2019]



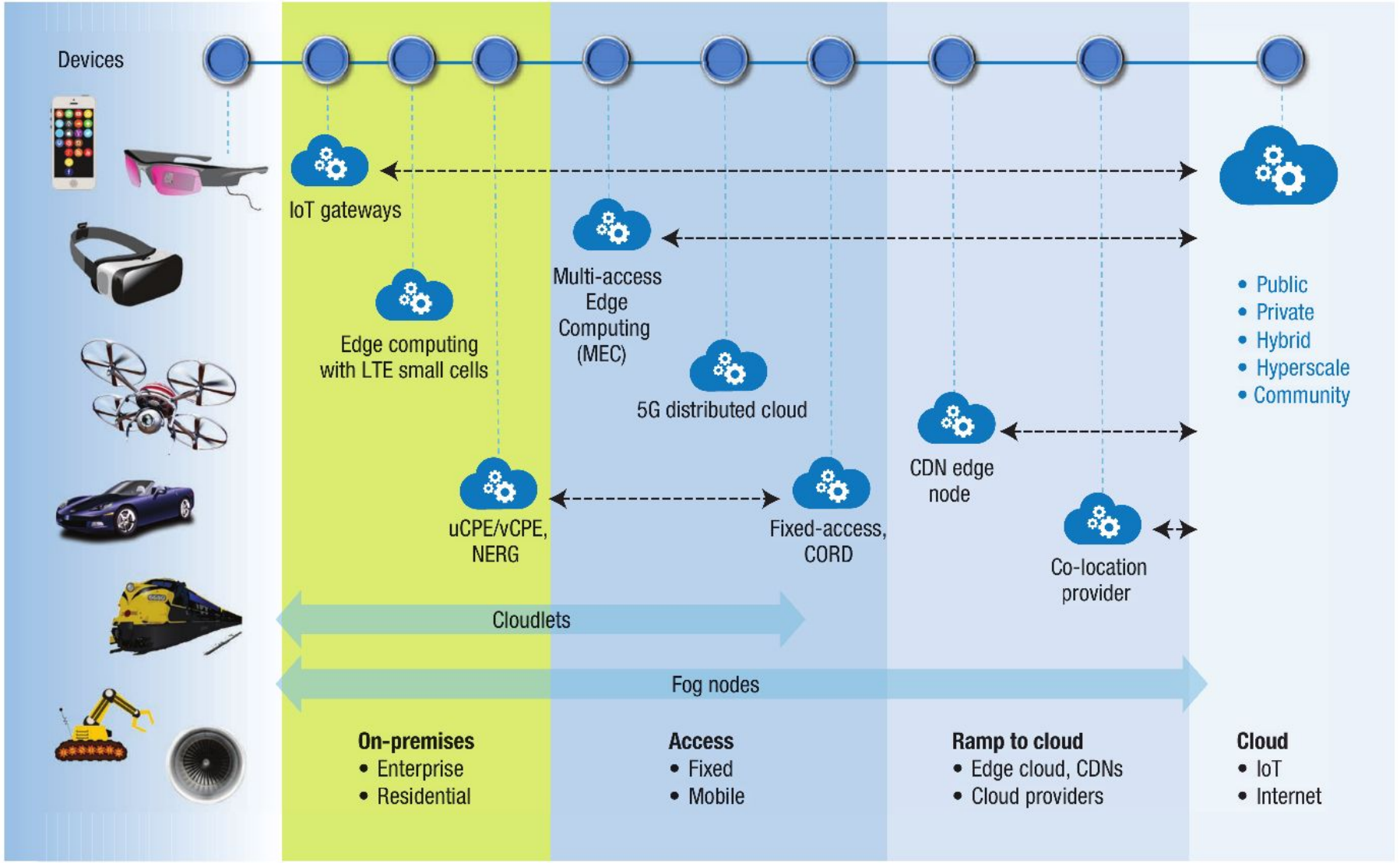
- **Fog computing**: Here considered as a synonym for Edge computing

Compute Continuum Overview

Here: A more telecommunications industry-centric view

DECICE

[G. Klas, 2017]



Potential Benefits from Extensions towards the Edge

[G. Klas, 2017]

- Low latencies and faster response from applications running at the edge
  - Virtual reality applications require round-trip-times O(10 ms)
- Reduction of data transport requirements
  - Data filtering and aggregation by edge services
- Improved data protection by keeping data locally, i.e. at the edge
  - Forward only aggregated and/or anonymised data, federated learning
- More predictable quality of service
  - Elastic Cloud services versus dedicated edge devices
- Facilitate integration of IoT devices

Selected Challenges

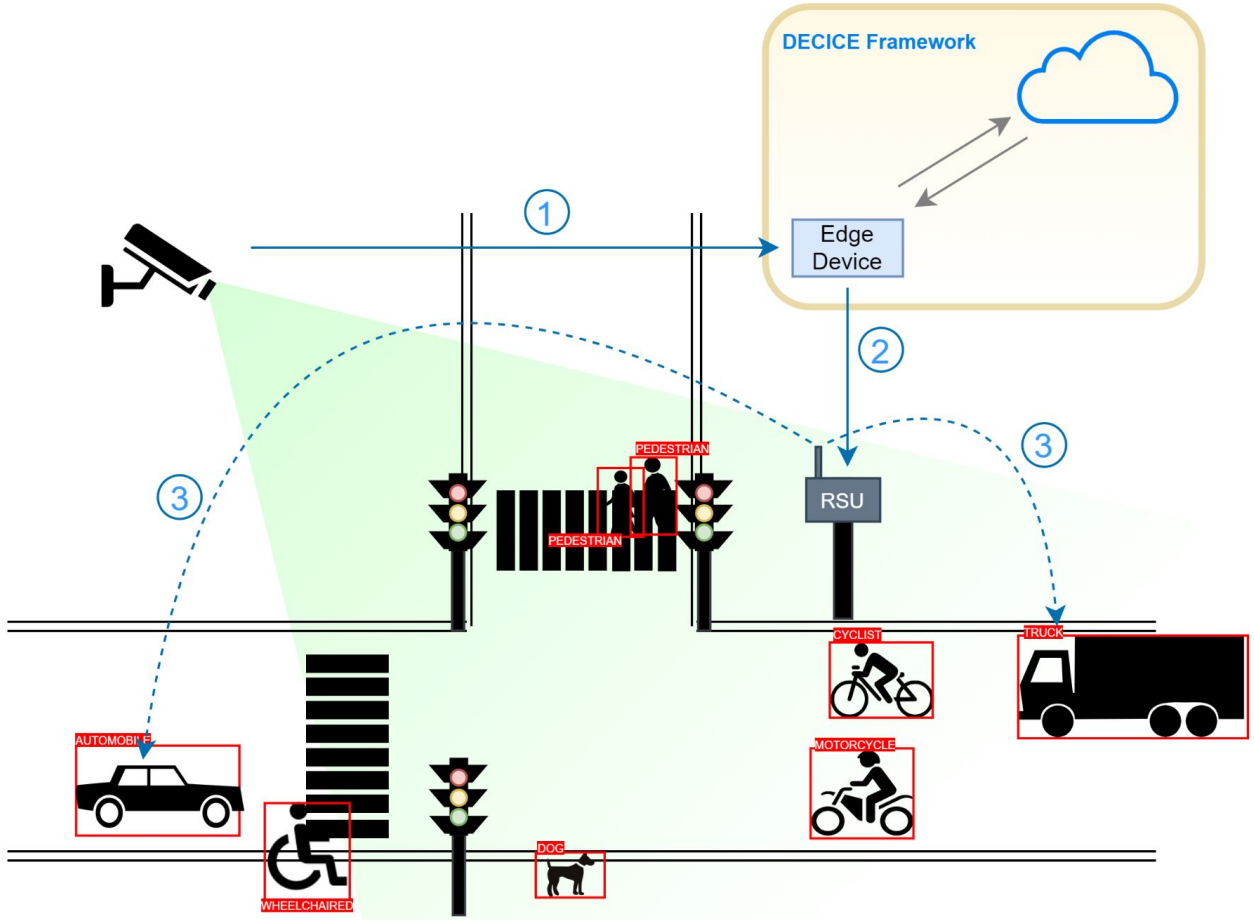
DECICE

- Edge devices are typically resource constrained
  - Power envelope of O(10 W)
- Heterogeneity of edge devices
  - CPUs with different ISA (Arm, RISC-V, …), various compute accelerators
- Potentially unstable or shielded network connectivity
  - Edge devices are often connected to a private network without public IP address
- Decentralised infrastructure monitoring
  - Service degradation
  - Upcoming needs for suitable billing mechanisms
- Resource allocation and service scheduling
- Trust and security

Example Use Case I: Intelligent Transportation Systems

[M. Soyturk, 2023]

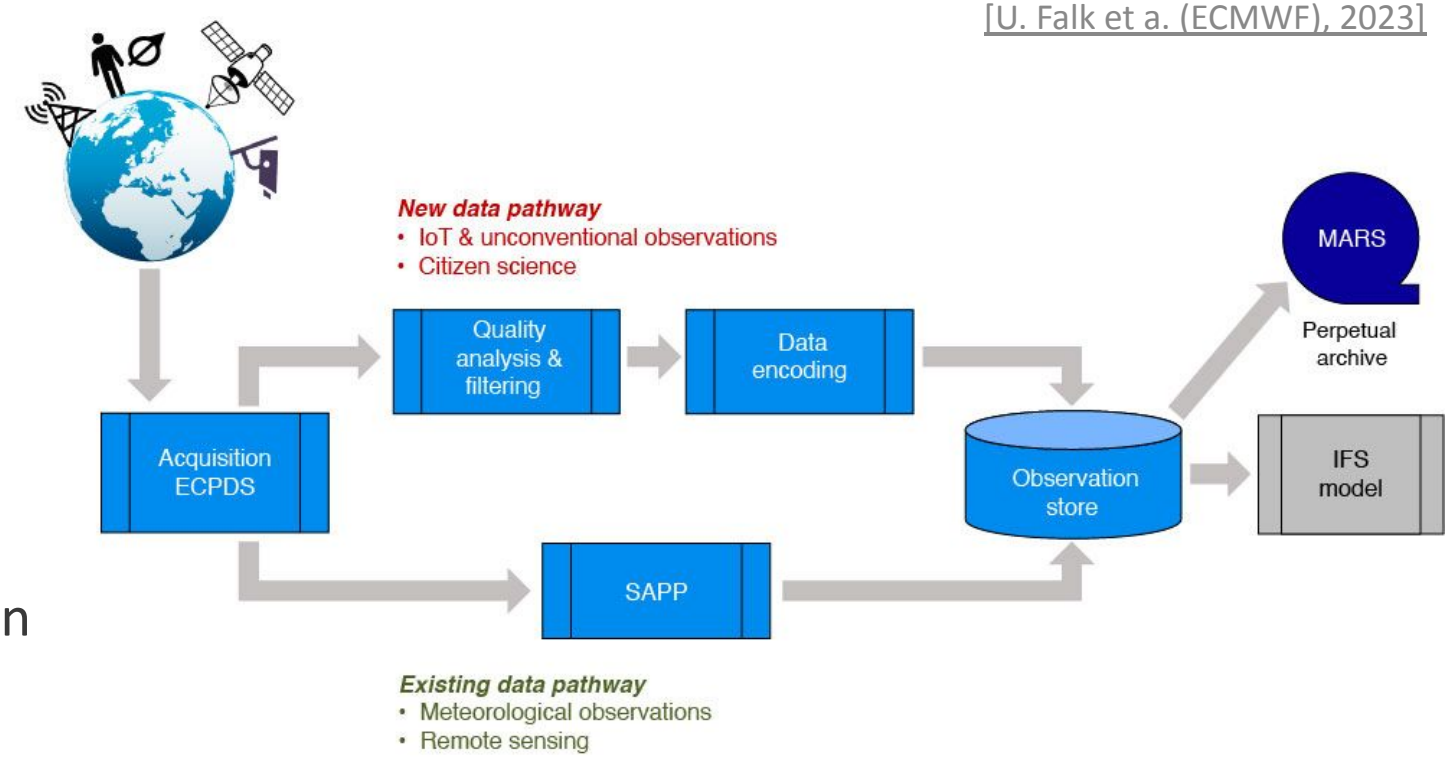
- Goal: Realise intelligent intersections to protect Vulnerable Road Users (VRU)
  - Cameras and edge devices detect VRU
  - Road-Site Units (RSU) connect to vehicles to collect vehicle sensor data and provide safety information
- Need for compute continuum
  - Local edge devices ensure low-latency responses
  - Cloud resources are used to continuously train models



Example Use Case II: Weather Prediction

- Goal: Integrate novel observations into weather prediction workflows
- Need for compute continuum:
  - Novel observations = input from IoT devices
  - Workflows comprise model simulations on HPC
- Challenges
  - Low reliability of data streams and need for filtering of noisy data at the edge
  - Heterogeneous data integration

[U. Falk et a. (ECMWF), 2023]



Questions?