ISC High Performance

REINVENTING HPC

MAY 12 – 16, 2024 | HAMBURG, GERMANY

HPC towards the compute continuum: Introduction

Dirk Pleiter (KTH and OEHI)

Intended Learning Outcomes

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

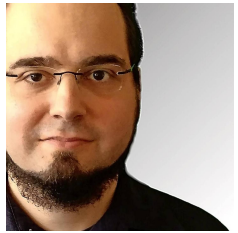
Agenda

14:00 - 14:25	Compute continuum and 1-2 use cases
14:25 - 14:45	Introduction to key technologies I: Containers and container orchestration systems
14:45 - 15:30	Hands-on I: Container building and deployment in homogeneous environments
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 - 16:30	*Coffee break*
16:30 - 16:45	Introduction to key technologies III: Prometheus
16:45 - 17:45	Hands-on III: Workflow implementation based on prepared components
17:45 - 18:00	Wrap-up

Tutorial Team

HLRS
High-Performance Computing Center Stuttgart

nag®



GOETHE UNIVERSITÄT FRANKFURT AM MAIN

KTH VETENSKAP OCH KONST

UNIVERSITÄT GÖTTINGEN

GWDG
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

HUAWEI

DECICE

OEHI

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.)

- **Edge computing**

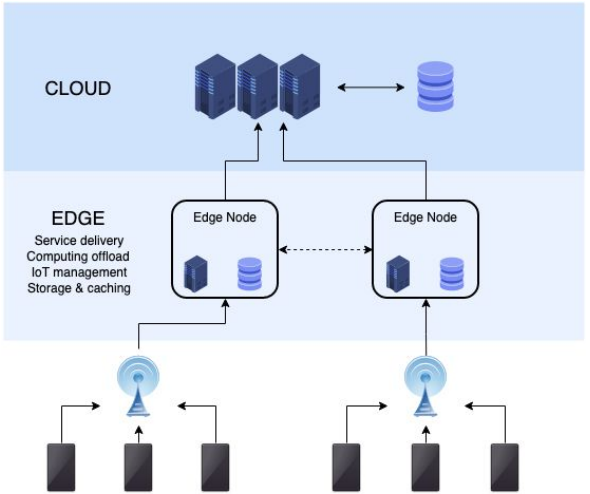
[Weisong Shi et al., 2016]

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

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

[Wikimedia Commons, 2019]



Potential Benefits from Extensions towards the Edge

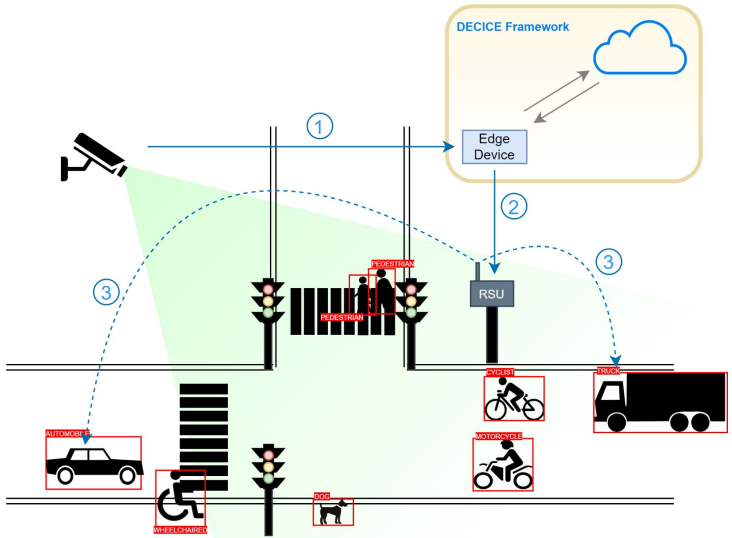
[G. Klas, 2017]

- 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

- 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
  - Central detection of service degradation
  - Upcoming needs for suitable billing mechanisms
- Resource allocation and service scheduling
- Trust and security

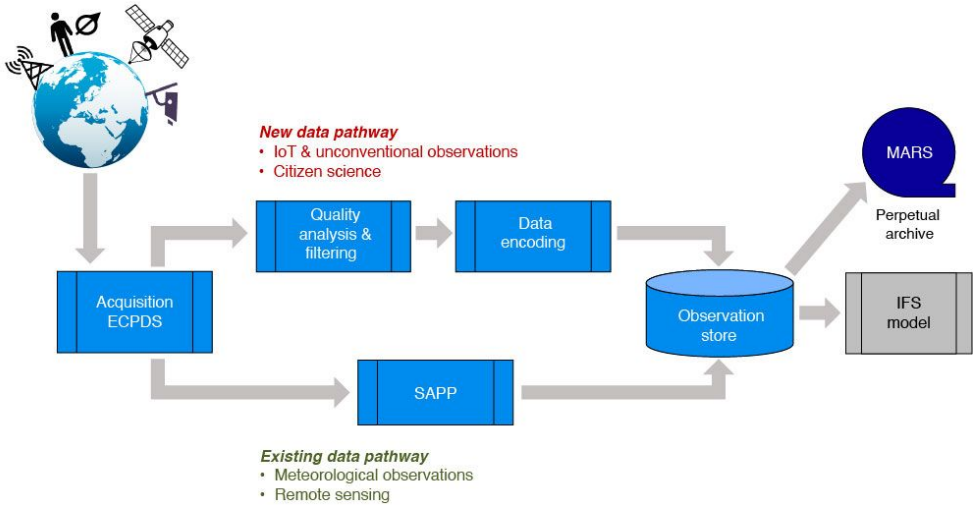
Example Use Case I: Intelligent Transportation Systems



[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



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

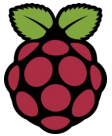
- Goal: Integrate novel observations into weather prediction workflows
- Need for compute continuum
  - Novel observations = input from IoT devices
  - Workflows comprise model simulations on HPC
- Challenges
  - Low reliability of data streams and need for filtering of noisy data at the edge
  - Heterogeneous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

Technologies Selected for this Tutorial

- Containers
  - This technology is widely used for running software applications in isolated user spaces
  - It can be used for HPC, Cloud, and the Edge
- Kubernetes
  - The most popular container orchestration system, i.e. a system for deploying containers
- KubeEdge
  - A system for extending Kubernetes to the Edge
- Prometheus
  - A toolkit for monitoring digital infrastructures

Hans-on Sessions

- Hands-on I: Container building and deployment in homogeneous environments
  - Practice how to build containers to deploy a simple service based on an MQTT message broker
  - Use Kubernetes to deploy the service in a remote Cloud instance
- Hands-on II: Container building and deployment in heterogeneous environments
  - Use Kubernetes and KubeEdge to deploy containers on edge devices (Raspberry Pis) here in the class room
- Hands-on III: Workflow implementation based on prepared components
  - Realise a more complex workflow from Cloud to Edge



Training Material and Slides

https://github.com/haicgu/HPC_compute_continuum

Temporary short link: https://shorturl.at/lnR07

Questions?