The CAPE project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101189899.

CAPE

European Open **C**ompute **A**rchitecture for **P**owerful **E**dge

The CAPE project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101189899.

CAPE

European Open
Compute **A**rchitecture for **P**owerful **E**dge

Project Overview
DSD25

Martin Kaiser, Bielefeld University
Jens Hagemeyer, Lennart Tigges, Christian Klarhorst, Björn Voß, Fred Buining, Bola Fakhoury, János Lazányi, René Griessl, Muhammad Shahzad, Yiannis Georgiou, Salim Mimouni, Pedro Velho, Michael Mercier, Eva Trungel, Julian Gajewski, Stefan Krupop, Micha Vor Dem Berge, Deepak M. Mathew, Skipis Dimitrios, Arnidis Iordanis, Orestis Vantzos, David Georgantas, Gautier Rouaze, Christoph Bühler, Guido Salvaneschi, Brandon Lewis and Angela Hauber

CAPE - European Open **C**ompute **A**rchitecture for **P**owerful **E**dge

- EU-Call: „Open Source for Cloud/Edge to support European Digital Autonomy"
- Start: 12/2024
- Consortium
  - 7x Industry (6x SME)
  - 4x Academic
  - 1x Standardization organization
- Coordinator: Bielefeld University
- Technical Coordinator: hiro

* associated participants

Big Picture of CAPE

Use Cases

Smart Grid

Deep Learning for Edge

Satellite payload data processing

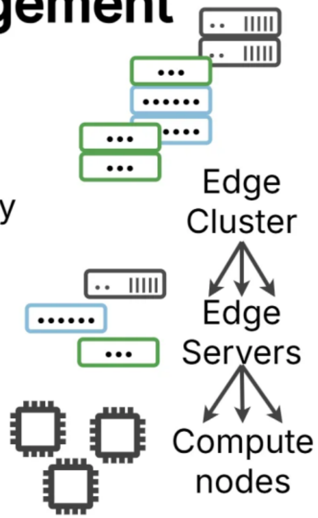
Standardization

Open-Edge-Cloud Infrastructure-Alliance

COM HPC

Open Edge Server Management

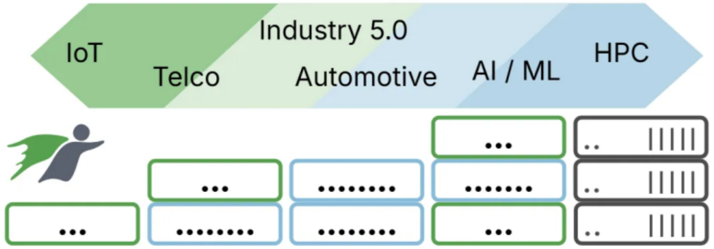
- Web-based, out-of-band configuration
- Run-time topology reconfiguration
- Live power monitoring
- Support for heterogeneous computing nodes



Software services

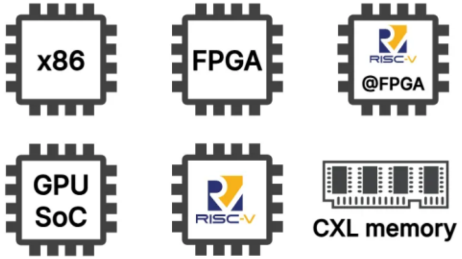
- Infrastructure from Code (IfC)
- AI-assisted orchestration and service deployment
- Resource clustering: dynamic node allocation & composition

Edge-Cloud Continuum



Hardware

Modular, heterogeneous compute nodes



Edge Server Platforms

eHPS (up to 3 nodes)
embedded **H**igh-**P**erformance **S**erver

EMDC (up to 7 nodes)
Embedded **M**icro**D**ata**C**enter

CXL switching for intra- and inter-server communication

Environment Integration

- Wall mount & 19" rack
- Passive cooling
- Waste heat recovery

Big Picture of CAPE

Use Cases
- Smart Grid
- Deep Learning for Edge
- Satellite payload data processing

Standardization
Open-Edge-Cloud Infrastructure-Alliance

COM HPC

Open Edge Server Management
- Web-based, out-of-band configuration
- Run-time topology reconfiguration
- Live power monitoring
- Support for heterogeneous computing nodes

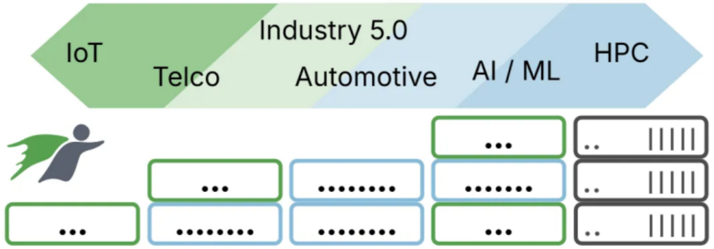
Edge Cluster

Edge Servers

Compute nodes

Software services
- Infrastructure from Code (IfC)
- AI-assisted orchestration and service deployment
- Resource clustering: dynamic node allocation & composition

Edge-Cloud Continuum



Hardware

Modular, heterogeneous compute nodes
x86

FPGA

RISC-V @FPGA

GPU SoC

RISC-V

CXL memory

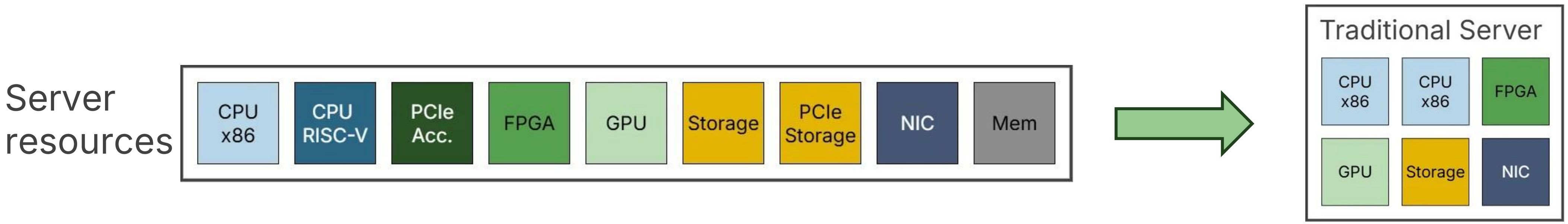
Edge Server Platforms
eHPS (up to 3 nodes)
embedded **H**igh-**P**erformance **S**erver

EMDC (up to 7 nodes)
Embedded **M**icro**D**ata**C**enter

CXL switching for intra- and inter-server communication

Environment Integration
- Wall mount & 19" rack
- Passive cooling
- Waste heat recovery

CAPE's Vision for Composible Server Infrastructure



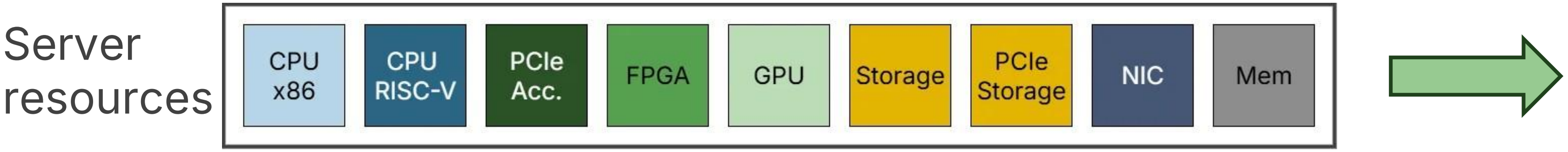
Goals

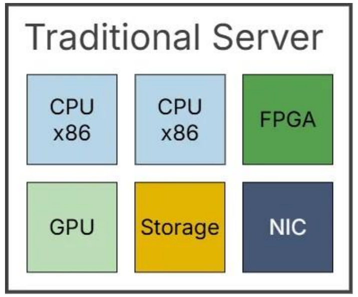
- „Application-driven infrastructure“
- Flexible and **dynamic resource allocation** in Edge-Cloud servers
- Target edge and scale-out to cloud
- Focus on standardization

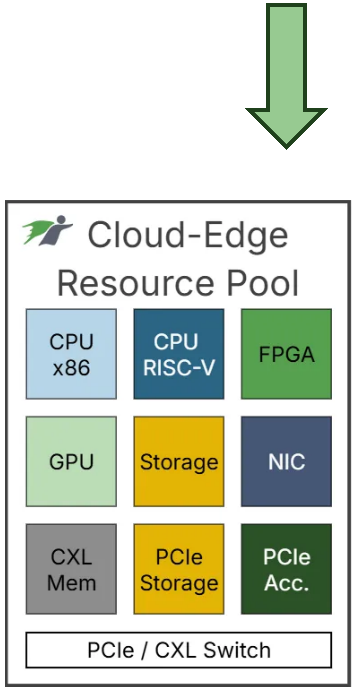
→ Use **modular compute node approach** with well-established form factors

→ Run-time **adaptable communication infrastructure** with CXL

CAPE's Vision for Composible Server Infrastructure







Goals

- „Application-driven infrastructure“
- Flexible and **dynamic resource allocation** in Edge-Cloud servers
- Target edge and scale-out to cloud
- Focus on standardization

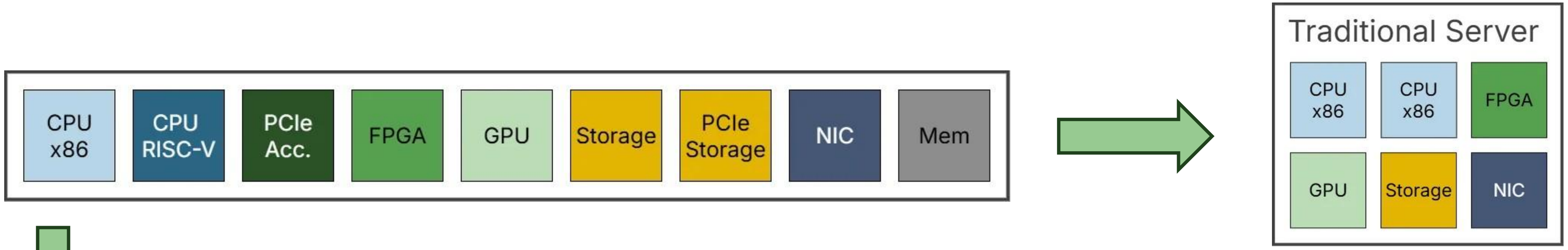
→ Use **modular compute node approach** with well-established form factors

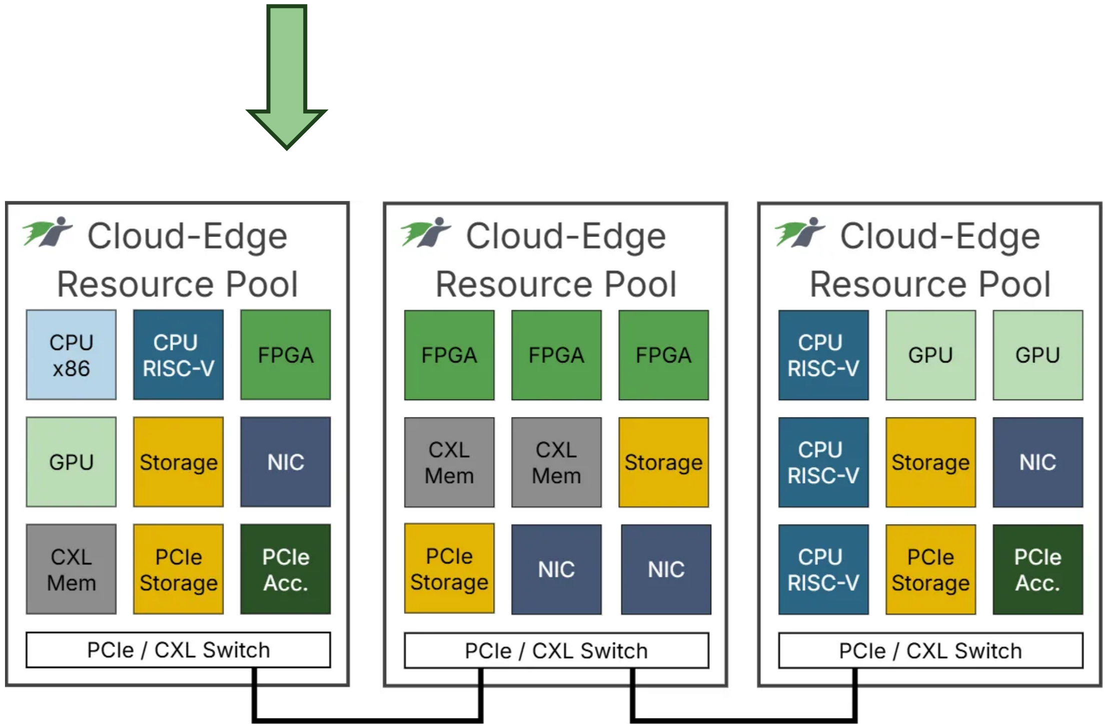
→ Run-time **adaptable communication infrastructure** with CXL

CXL (Compute Express Link) is based on the PCIe transport layer and provides for cache-coherency and memory pooling

CAPE's Vision for Composible Server Infrastructure

Server resources





Resource Pools → Composable Infrastructure

Goals

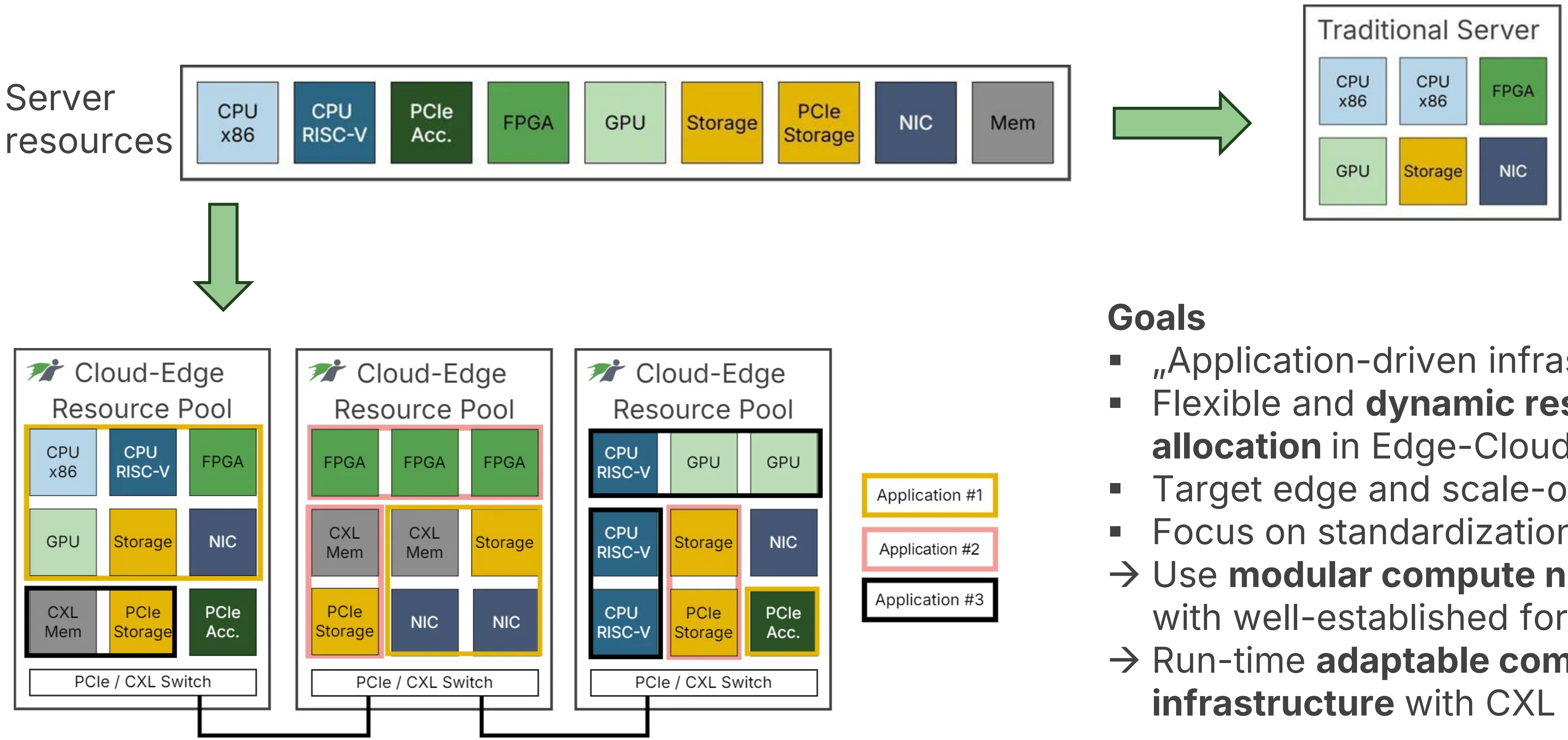
- „Application-driven infrastructure“
- Flexible and **dynamic resource allocation** in Edge-Cloud servers
- Target edge and scale-out to cloud
- Focus on standardization

→ Use **modular compute node approach** with well-established form factors

→ Run-time **adaptable communication infrastructure** with CXL

CXL (Compute Express Link) is based on the PCIe transport layer and provides for cache-coherency and memory pooling

CAPE's Vision for Composible Server Infrastructure



Goals

- „Application-driven infrastructure"
- Flexible and **dynamic resource allocation** in Edge-Cloud servers
- Target edge and scale-out to cloud
- Focus on standardization

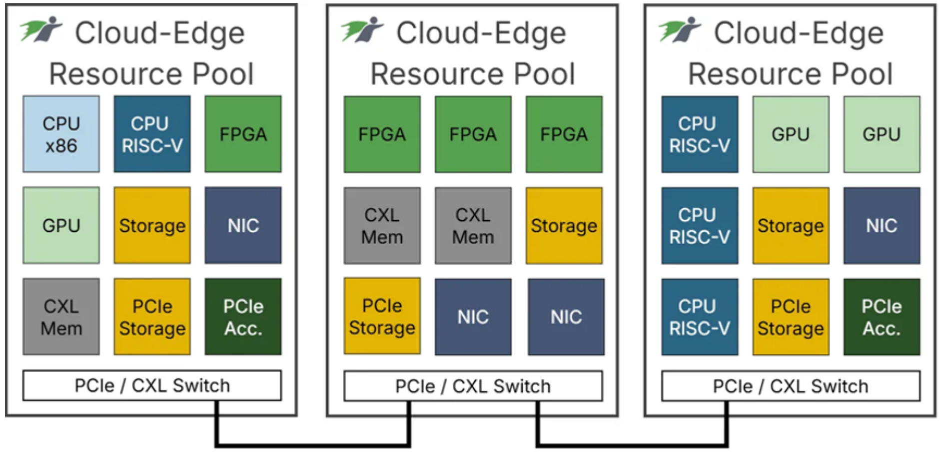
→ Use **modular compute node approach** with well-established form factors

→ Run-time **adaptable communication infrastructure** with CXL

Resource Pools → Composable Infrastructure

CXL (Compute Express Link) is based on the PCIe transport layer and provides for cache-coherency and memory pooling

Hardware Platforms

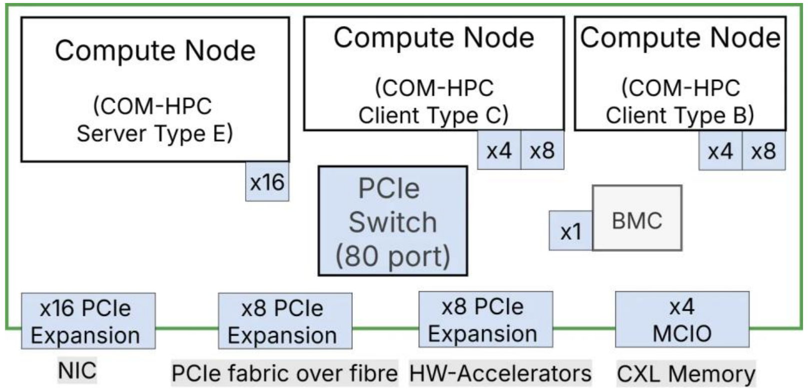


CAPE's fully
Composable Infrastructure

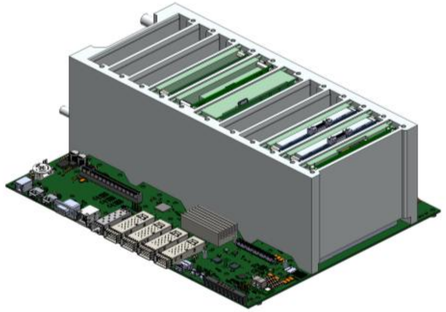
Two
implementations

Heterogeneous
edge servers

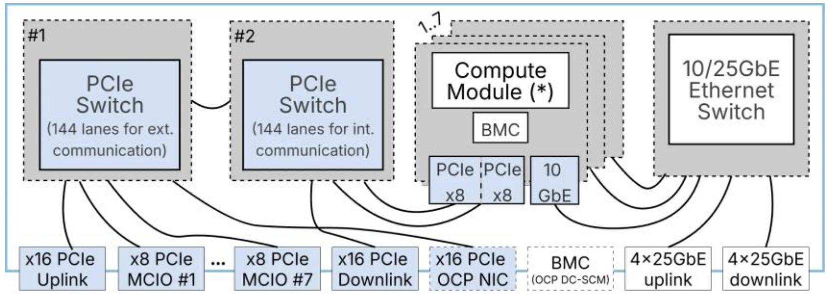
eHPS: embedded High Performance Server



Passive cooling &
building integration



EMDC: Embedded Micro Data Center



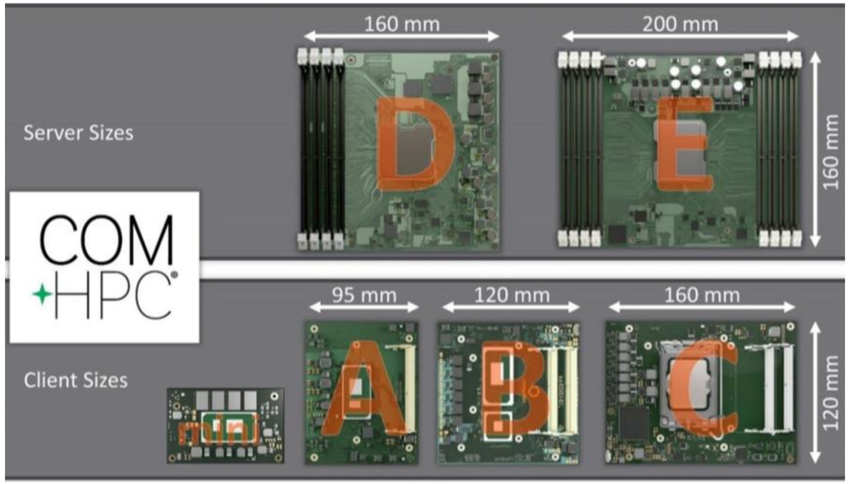
(*) Supported compute modules: COM-HPC Server Type E, Jetson AGX

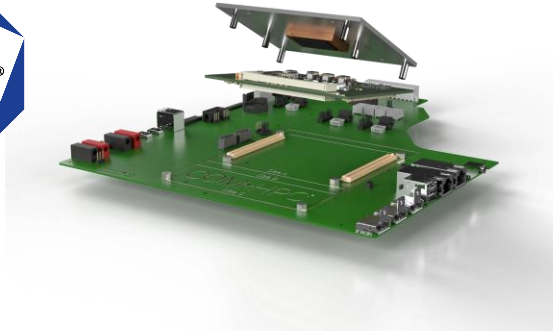
----- exchangeable module   interconnect by cable

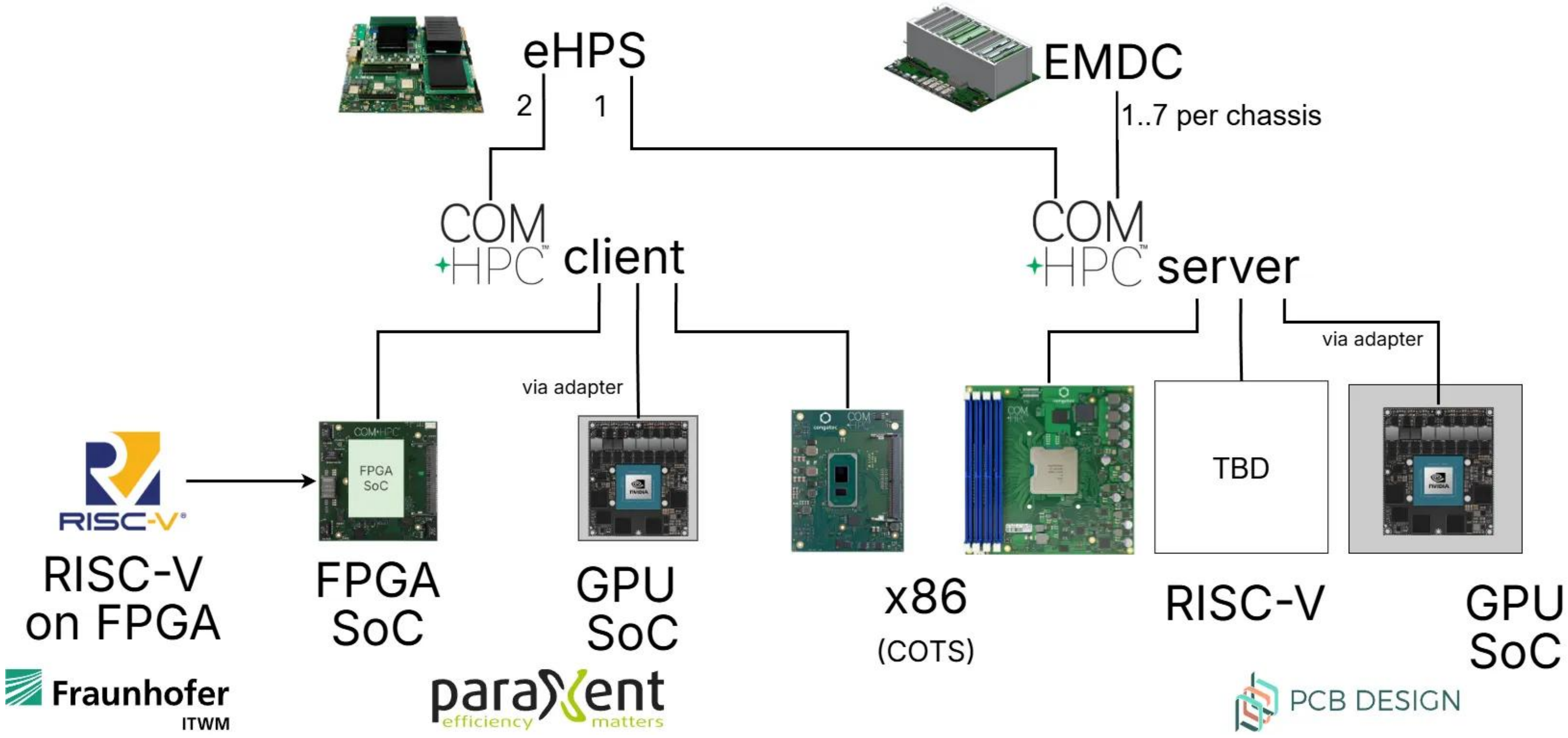
UNIVERSITÄT BIELEFELD · CHRISTMANN INFORMATIONSTECHNIK + MEDIEN · paraXent efficiency matters · PCB DESIGN · EPFL · Cadterv

Modular compute nodes

- Based on well-established form factors: COM-HPC Client and Server
- Focus on heterogeneous computing







COM-HPC is a computer-on-module form factor standard for high performance edge applications

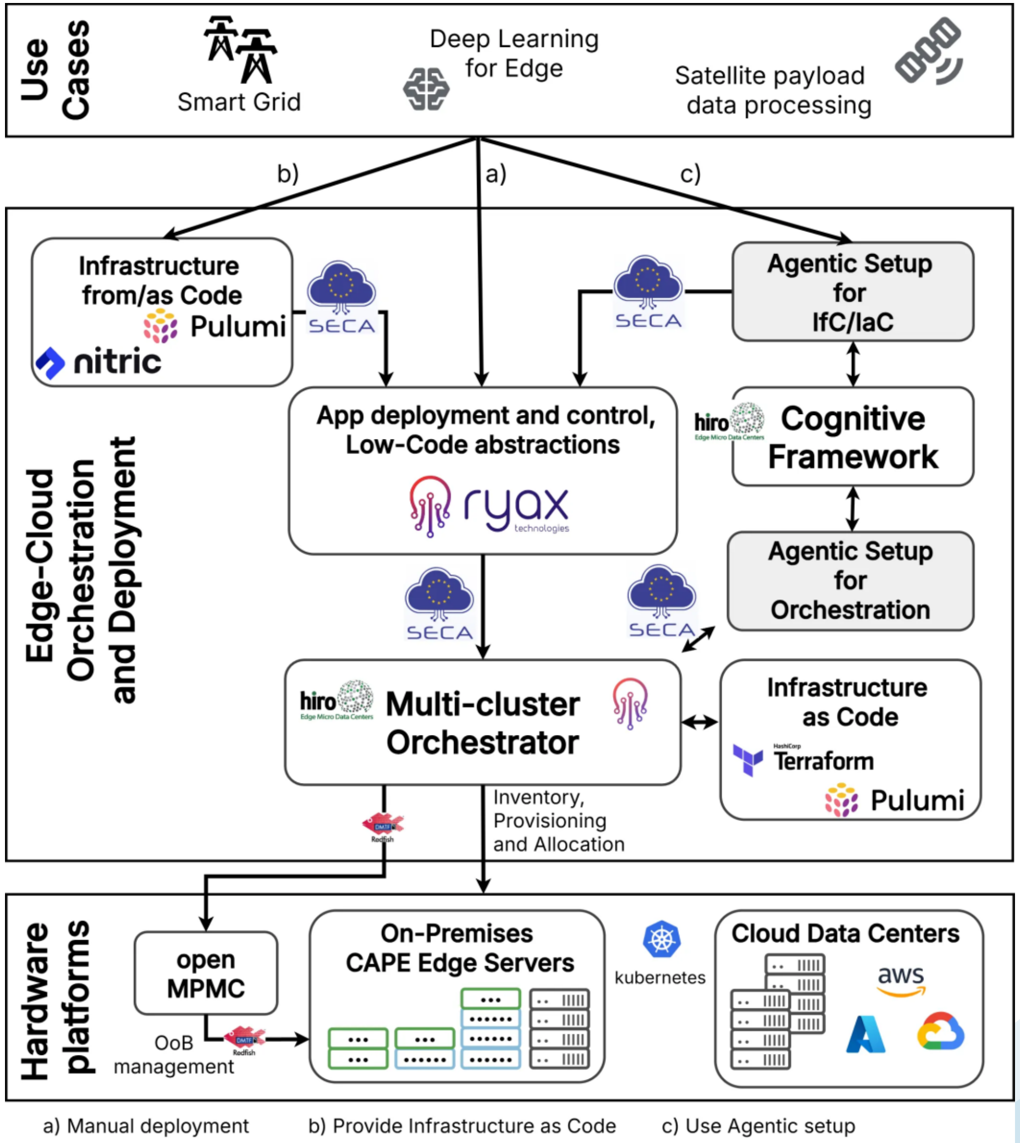
Software Architecture

Features

- Transparent deployment across edge-cloud continuum
- Heterogeneous computing resources
- Unified programming and deployment model
- Fully open source stack including (Sovereign European Cloud API)

Components / tools

- Infrastructure from/as Code (IfC/IaC): Automated Software Testing
- Ryax: Deployment, resource selection and data center management
- Orchestrator: manages inventory, configures network infrastructure and compute nodes
- OpenMPMC (Multi-Platform Management Controller): low-level server management

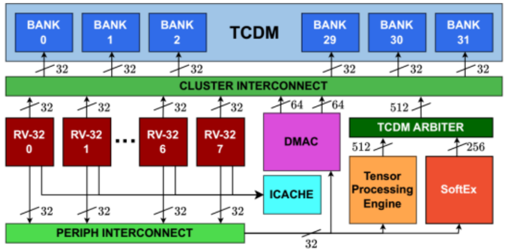


Use Cases

Deep Learning for Edge

Fraunhofer ITWM

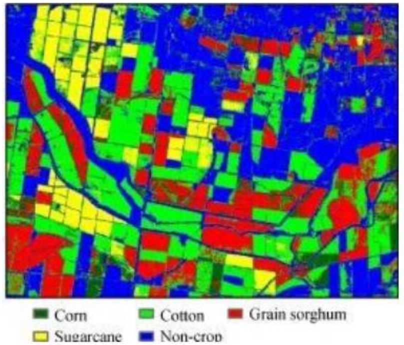
- Surveillance on RISC-V / FPGA with Acc.
- Optimize deep learning algorithms
- **Low-latency & high throughput**



Satellite Payload Data Processing

PCB DESIGN

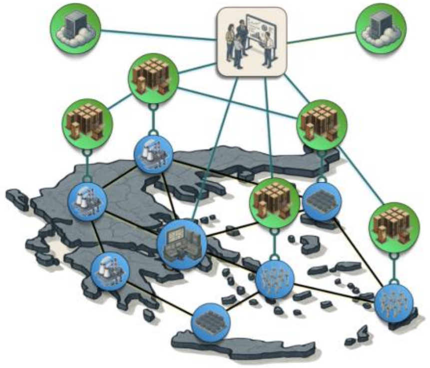
- Image (pre-)processing
- Data clustering & fusion
- **High throughput & energy efficiency**



Smart Grid

ipto

- Real-time anomaly detection
- Energy forecasting & market prediction
- **Robustness & real-time capabilities**

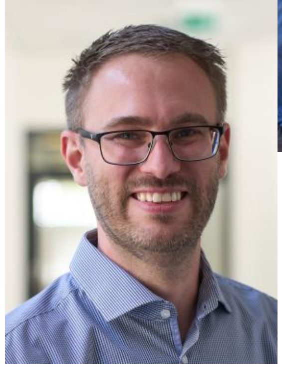


CAPE

European Open **C**ompute **A**rchitecture for **P**owerful **E**dge

CAPE LinkedIn

CAPE website



Martin Kaiser
Bielefeld University
Germany

mkaiser@techfak.uni-bielefeld.de

UNIVERSITÄT BIELEFELD

hiro Edge Micro Data Centers

PCB DESIGN

paraXent efficiency matters

Cadtern

ryax technologies

CHRISTMANN INFORMATIONSTECHNIK + MEDIEN

ipto

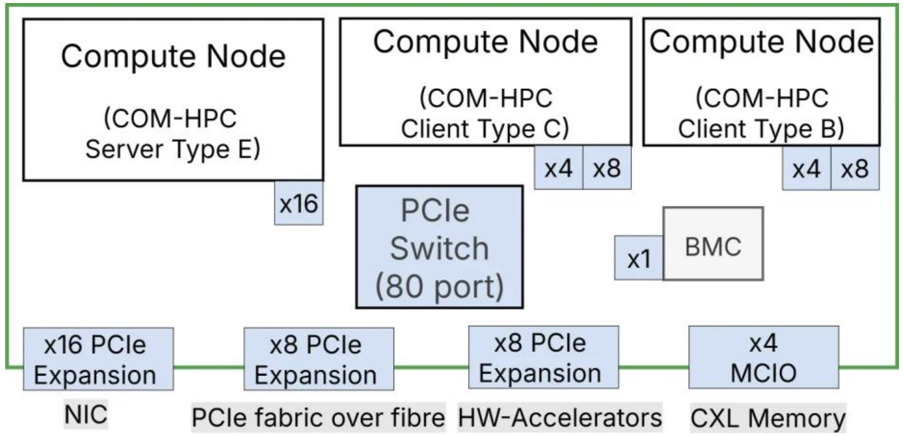
Universität St.Gallen

Fraunhofer ITWM

EPFL

PICMG

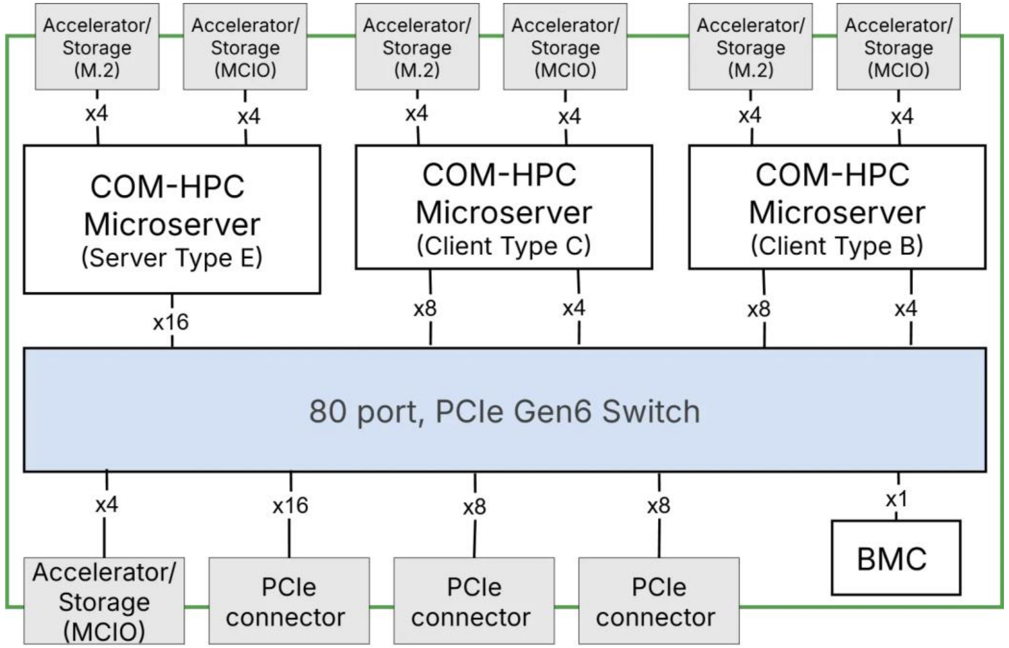
eHPS: embedded High Performance Server



System Architecture

E-ATX
Mainboard

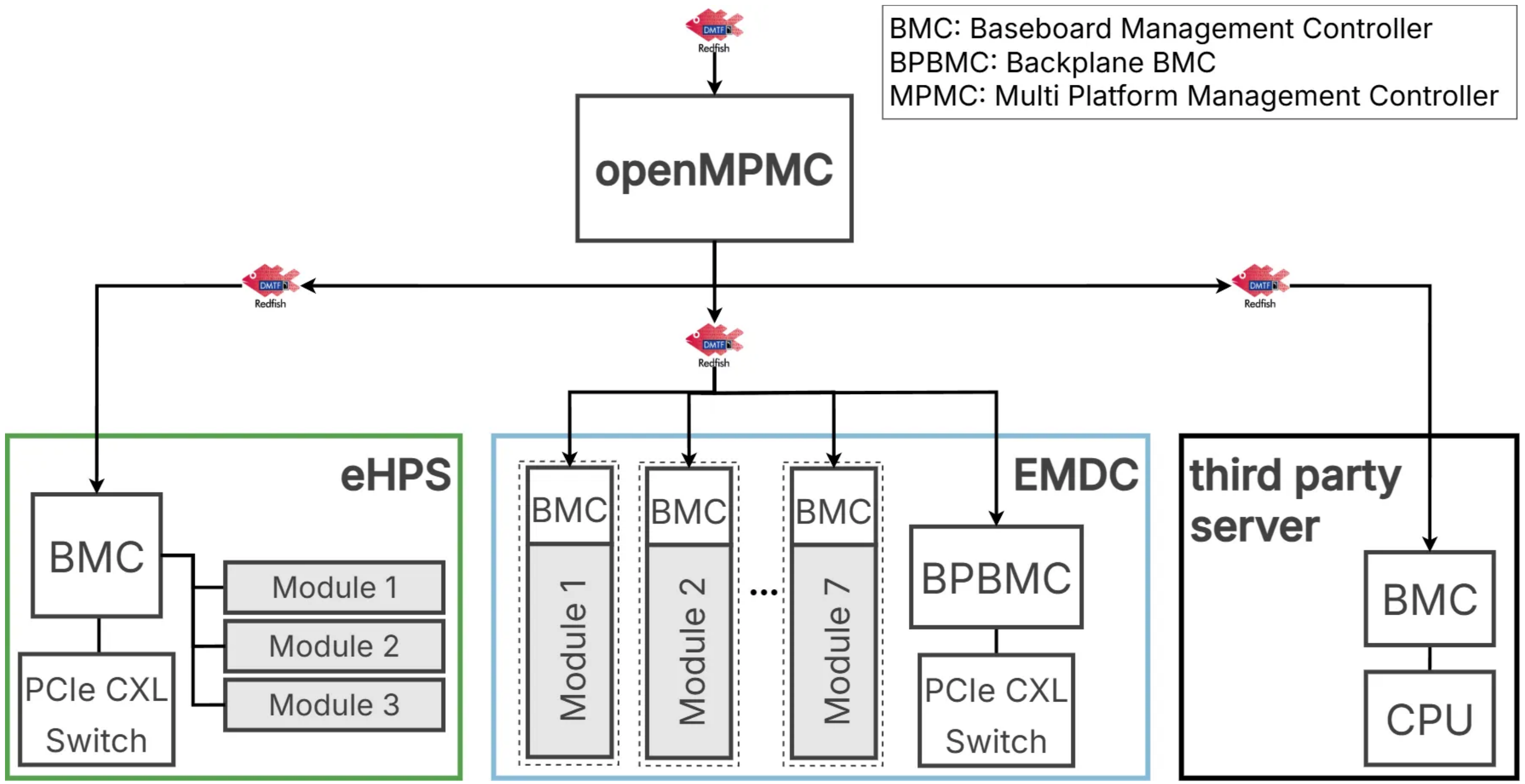
3 HU 19" housing



PCIe based communication infrastructure

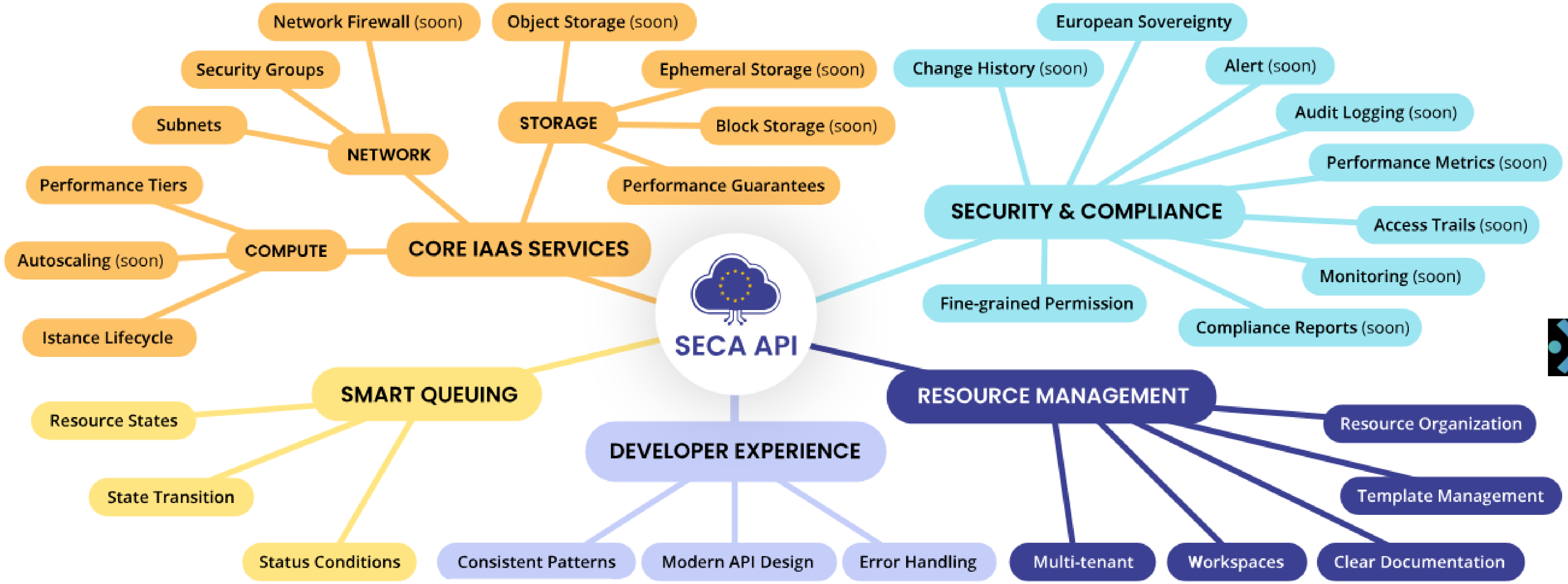
Board Management

- Redfish as open API for low-level control
- openMPMC: Multi-Platform Management Controller
- Two approaches:
  - eHPS: 1 BMC → 3 nodes
  - EMDC: 1 BMC → 1 node + 1 central BMC
- openBMC vs. custom linux



SECA (**S**overeign **E**uropean **C**loud **API**)

- open industry standard www.secapi.cloud
- API specification for Cloud Infrastructure Management



aruba.it
IONOS
dynamo