ACES

Autopoietic Cognitive Edge-cloud Service

Introductory Project Presentation

Speaker Name

Funded by
the European Union

ACES at a glance

Project title	Autopoietic Cognitive Edge-cloud Service
Call and topic	HORIZON-CL4-2022-DATA-01 Grant Agreement Number: 101093126

Members

8 Beneficiaries
3 associated partners

Funds

€ 5 543 925.00

Duration

36 months
Start Date: 01/ 01 /2023

Motivation

ACES Motivation

Challenge: increasing need for cloud services at the edge (edge–services) caused by rapidly growing quantity and capabilities of connected and interacting edge devices exchanging vast amounts of data.

<u>**Specific challenges to cloud computing architectures at the edge**</u>, such as:

i) ability to **provide end-to-end transaction resiliency** of applications broken down **in distributions of microservices**;

ii) creating **reliability and stability of automation in cloud management** under increasing complexity

iii) **secure and timely handling** of the increasing and latency sensitive **flow (east-west) of sensitive data and applications**;

iv) need for **explainable AI and transparency of the increasing automation in edge-services** platform by operators, software developers and end-users.

<u>Solution</u>: infused autopoiesis and cognition on different levels of cloud management to empower with AI different functionalities such as: workload placement, service and resource management, data and policy management.

Specific solutions:

End-to-end transaction resiliency to fulfil the varying end-user requirements, such as stipulated by Quality of Service (QoS), Service Level Objectives (SLO) and Service Level Agreement (SLA)

Autonomy and self-regulating mechanisms that go beyond recursive self-reference and provide systems stability, locally and edge-wide

Horizontal' (east-west) flow of data and applications between sites, and the interactions between data producers and data consumers, in terms of bandwidth, energy efficiency, security, privacy, etc.

Explainability of the autonomous operations on the platform to the different stakeholders i.e. operators, application developers and end-users

Deliver low-overhead (costs, latency, energy, labour) security in stream and in real time in a zero-trust environment

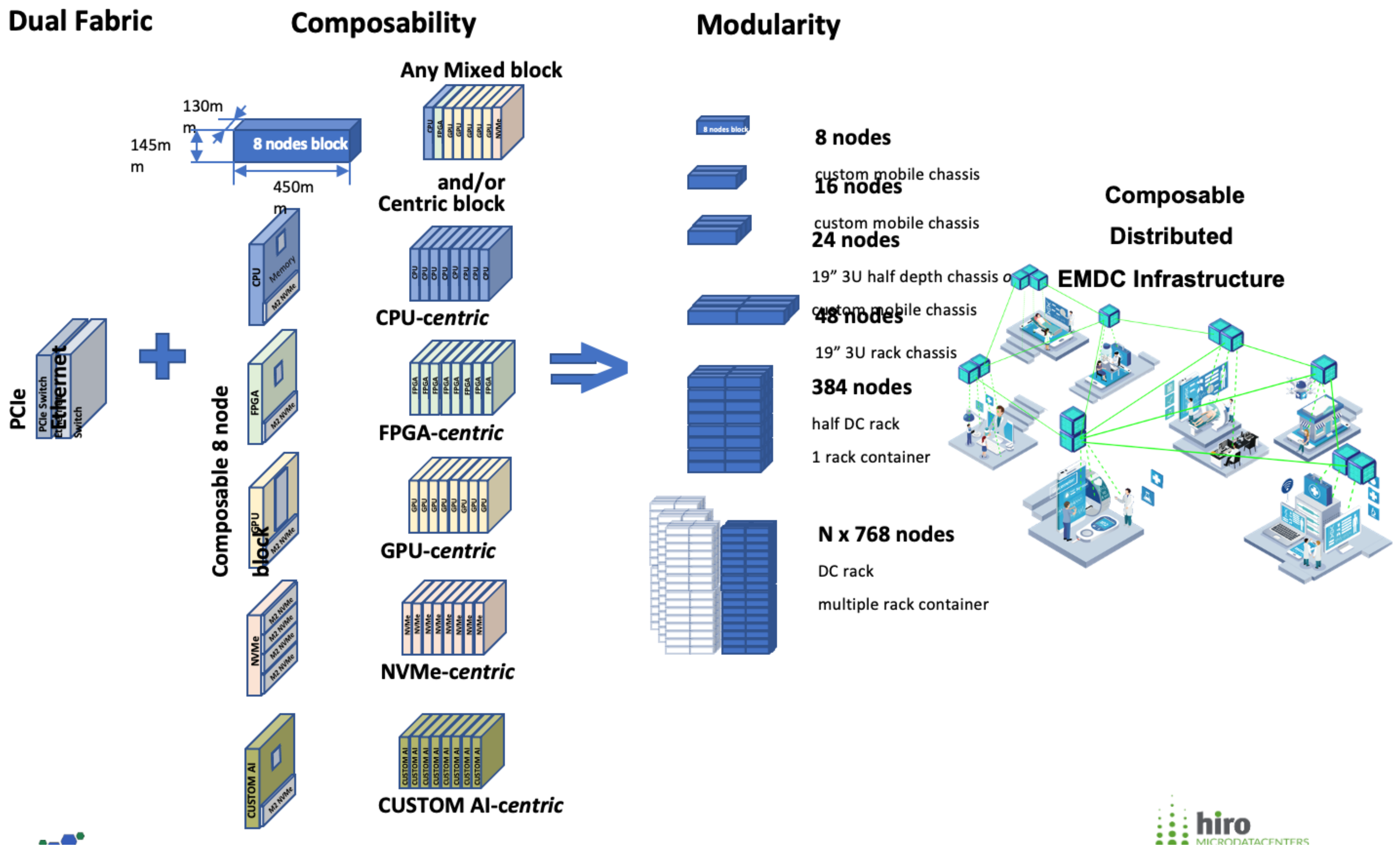
Vision

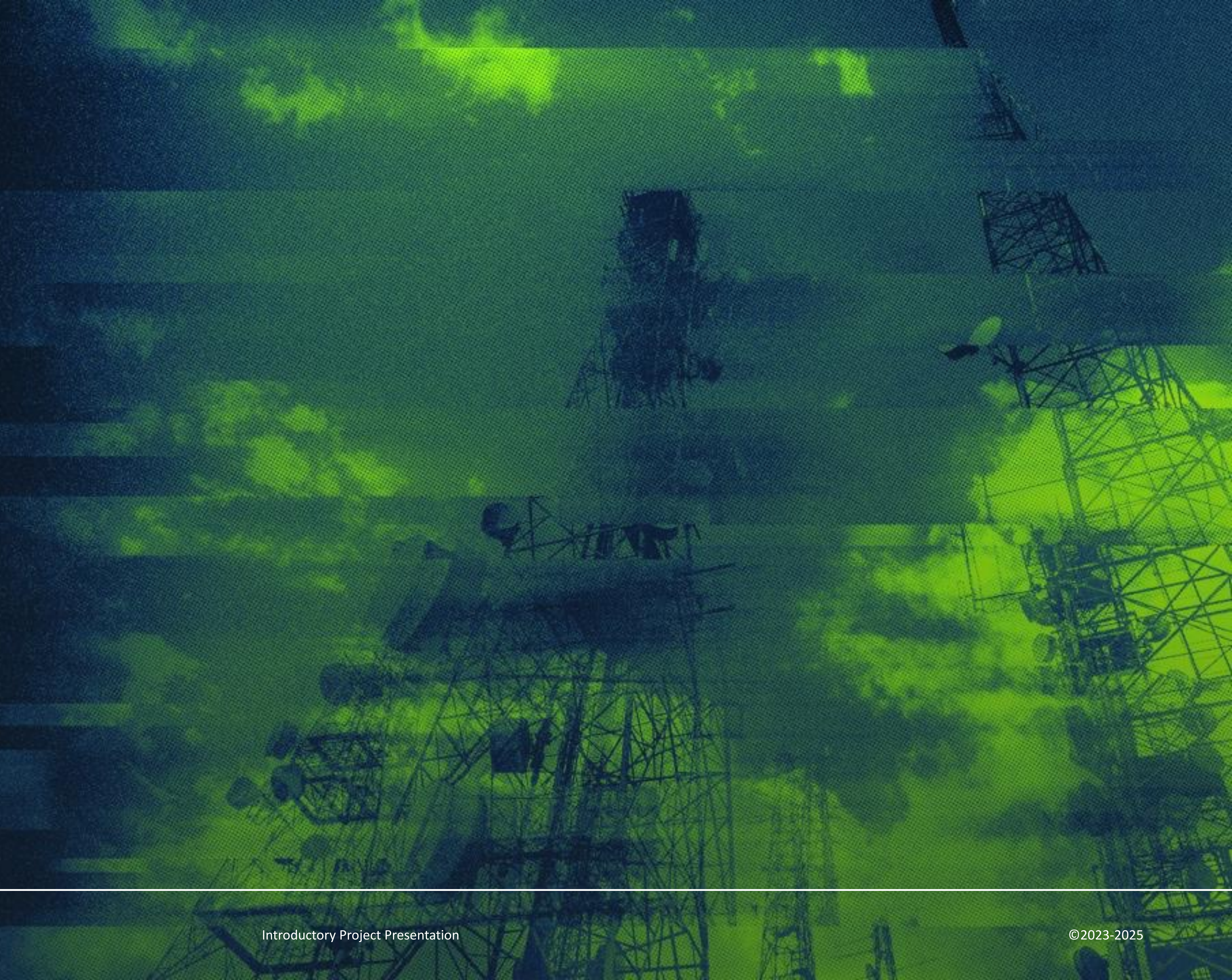
ACES Vision

ACES vision is to is to research an evolution of cloud computing, an edgeservices cloud with hierarchical intelligence, specifically Autopoiesis and cognitive behaviours, to manage and automate a compute platform, network fabric, storage resources, virtualization, and analytics to increase resilience while managing simultaneous service constraints.

- **Cognition** is defined as the ability to acquire and process information, apply knowledge, and autonomously change the inner circumstances to provide better services in response to the fluctuations in the environment. Because cognition by definition creates intelligent self-adaptation up to the same level of complexity as its analysed environment, ACES needs an additional control approach to protect the system from damage or destruction in situations where its cognitive complexity handling capabilities are exceeded.
- **Autopoiesis** means self-producing and maintaining itself, and is defined as the ability to produce more of their own organization principles than the ones produced by analysing the environment (cognition). This type of self-adaptive systems is a field of investigation, autonomic computing, that studies how systems can achieve desirable behaviours on their own, for example, through the use of AI

ACES - Customizable Edge mesh of Edge MicroDatacenters (EMDC's)





Objectives

Main Objective

ACES will provide an edge-services cloud with hierarchical intelligence, specifically autopoiesis and cognitive behaviours, to manage and automate a compute platform, network fabric, storage resources, virtualization, and analytics to increase resilience while managing simultaneous service constraints

ACES Specific Objectives

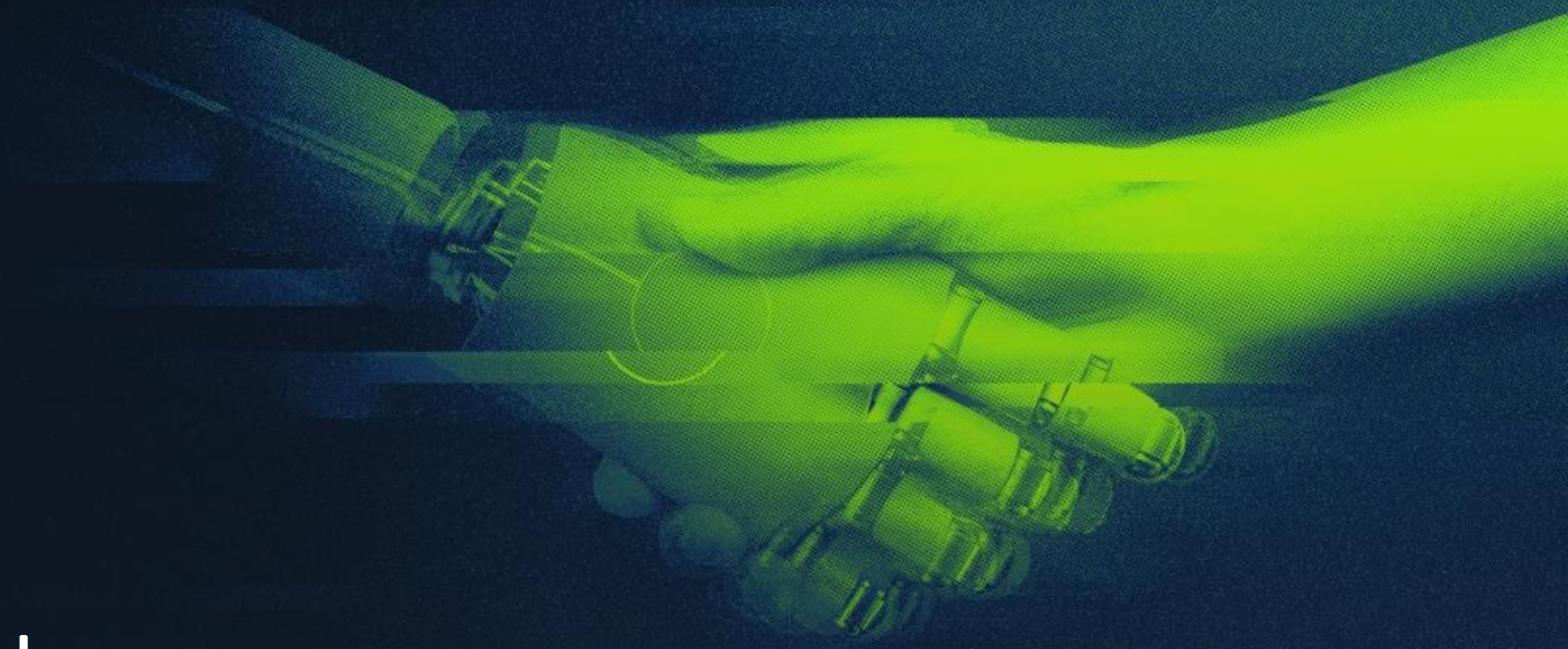
Develop an Autopoiesis cognitive edge-services cloud that enables infrastructure and intelligent operation leaders to deploy and operate an infrastructure that converges and automates the compute platform, network fabric, storage resources, virtualization, and analytics to increase resilience while managing simultaneous service constraints

Research Autopoiesis cognitive edge-services that orchestrate resource and workload management with respect to a wide range of edge-relevant requirements, such as latency, energy efficiency and security, and balancing these criteria for multi-tenancy at a single site and across multiple sites

Research autopoiesis cognitive edge-services that optimize the data management, data storage, data replication, acceleration and data movement

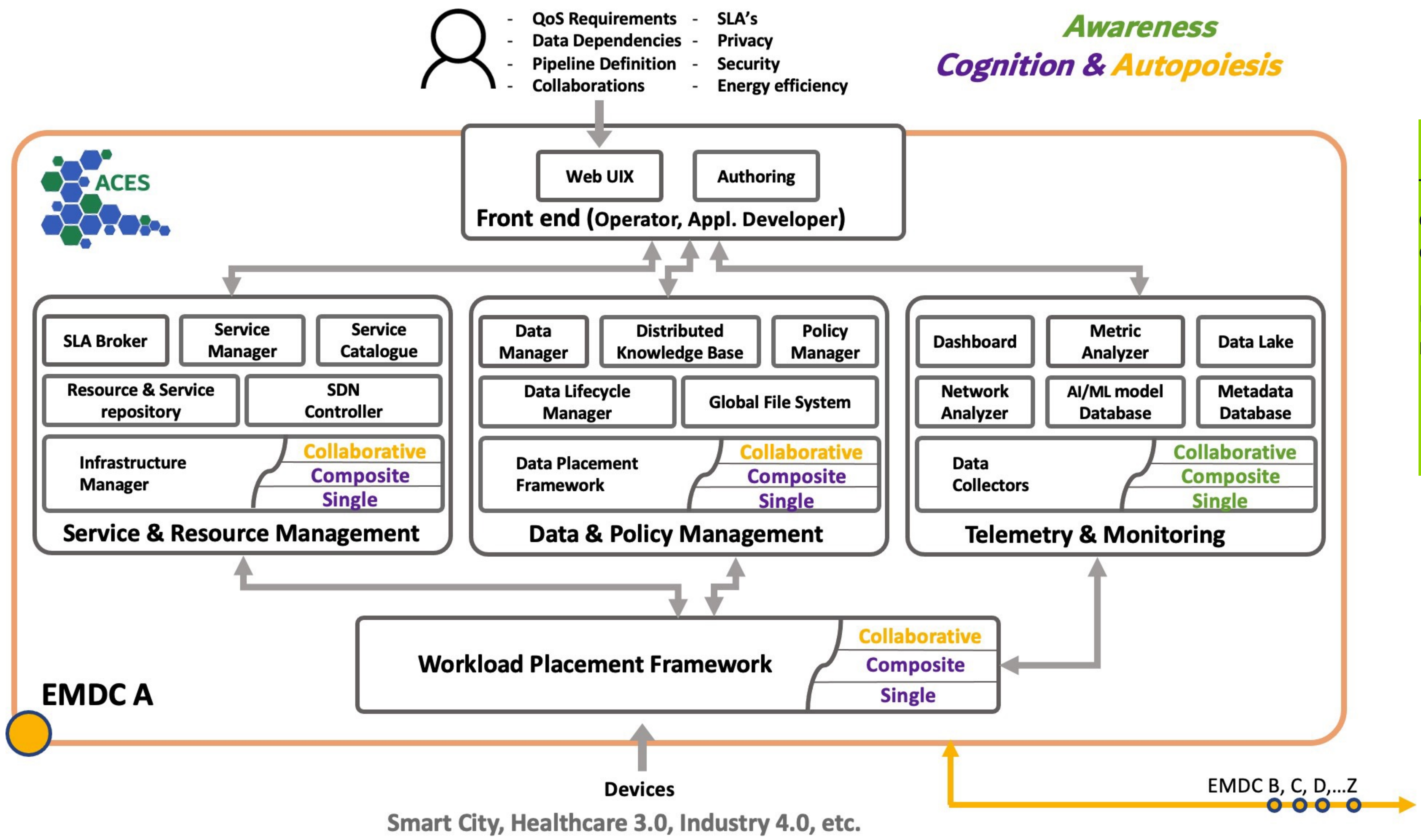
Research services (including visualisation and recommendation framework) to improve the experience of operators, software developers and developing applications and end-users

Test and demonstrate the effectiveness and generality of ACES by evaluating three real-life use cases that exhibit cognitive edge-services with various levels of awareness, autonomy, actionability within the services, the edge-services stack and with regard to the hardware



Approach

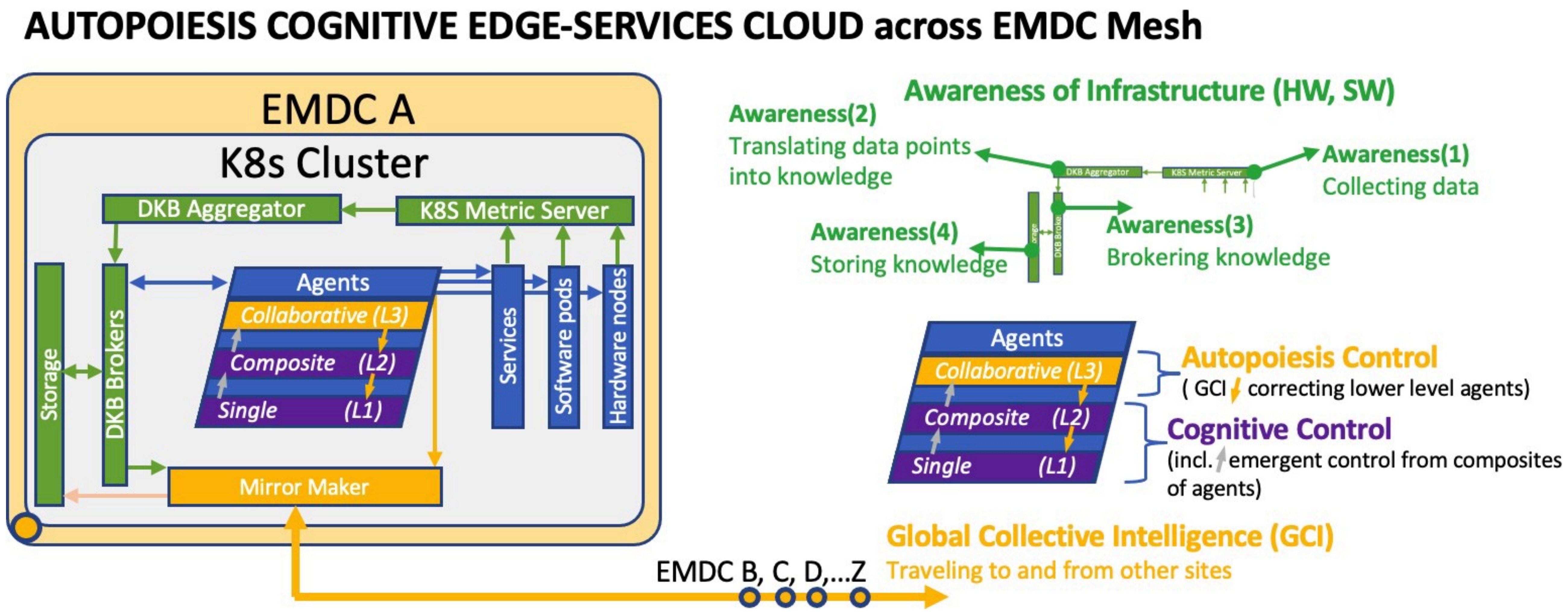
ACES conceptual architecture



The Autopoiesis Cognitive Edge-Services Cloud (ACES) project develops a distributed, opportunistic, collaborative, heterogeneous, self-managed, self-organizing edge services environment, primarily edge-to-edge (east-west) and secondly on the edge-to-cloud continuum (south-north).

Key aspects : allocating, placement and monitoring workloads on a network of edge nodes, including cloud resources, using AI/ML algorithms.

ACES approach – integrated autopoietic cognitive framework



ACES will develop a multi-agent framework with different levels of Awareness and Cognitive and Autopoiesis control executed across services.

The different levels are:

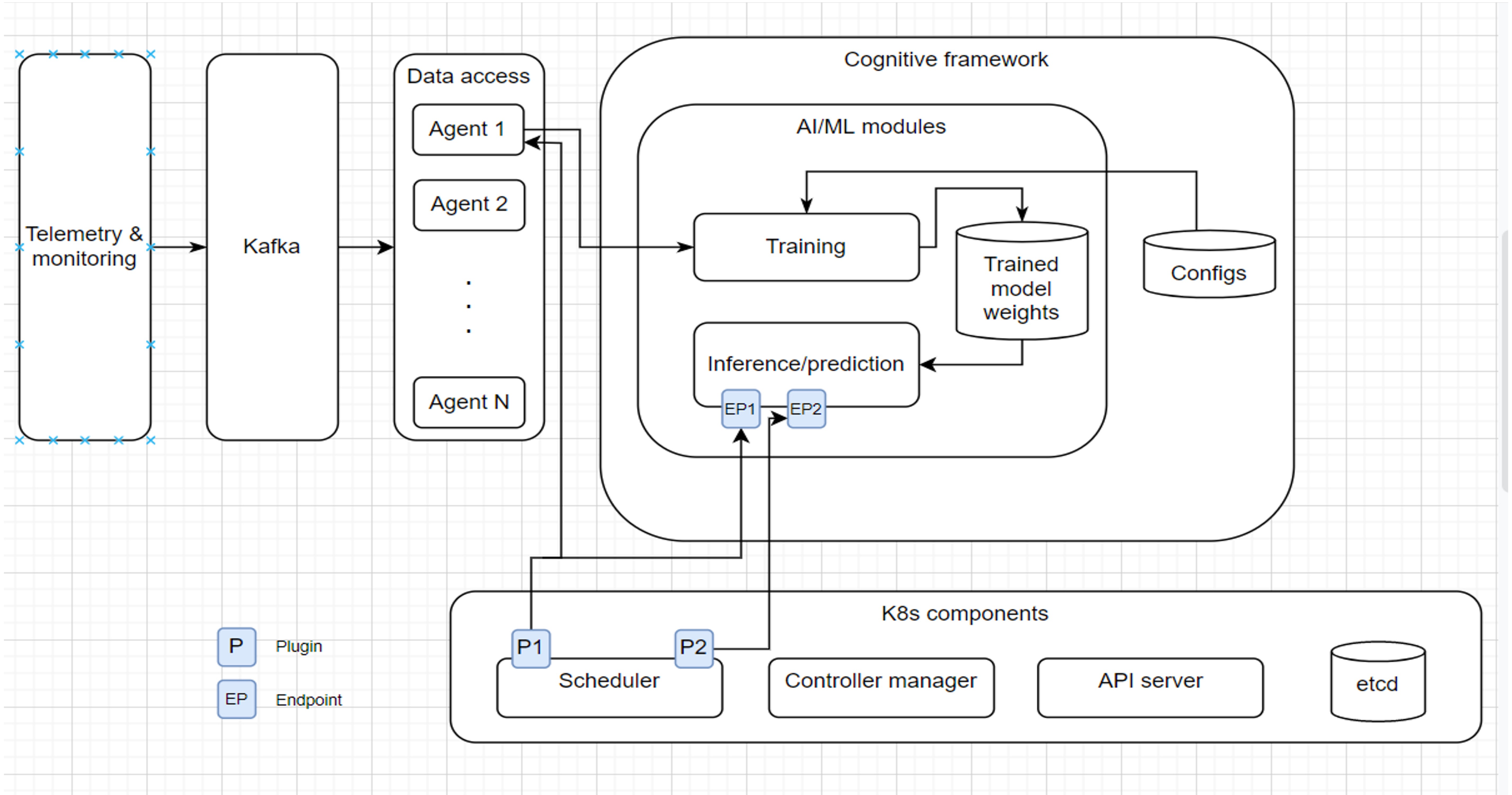
- Single (L1) lower-level knowledge and agents around single service, pod, HW node
- Composite (L2) higher level knowledge and agents managing composites of L1 agents within one EMDC
- Collaborative (L3) knowledge and agents active for collaborations between EMDC's across multiple sites

Supervised Learning will be implemented at the level of the L1 single agents in order to perform state estimation tasks in an efficient way, while geometrical compression space techniques will be used in order to explore in a smart way the solution space of the ML problem model

As the complexity of tasks and the number of agents grow, the L2 agents coordinating a composite of L1 agents become of fundamental importance

L3 agents, based on swarm intelligence are highly scalable, adaptive and robust with little computational overhead, will facilitate the collaboration between L2 agents in one site and across multiple sites

ACES approach – cognitive engine



ACES approach – distributed knowledge base/storage

Distributed knowledge base will store;

(a) Current and previous states of the system (including deployed applications, available computing resources, etc.)

(b) Desired state of the system based on functional (incl. microservices network of deployed applications) and non-functional specifications (QoS, SLOs regarding response time, latency, target resource utilisation, etc.) from user/admin/operator(s)

(c) Current, past and predicted models (incl. meta models, surrogate models) of the system, its environment generated via AI/ML and their effectiveness.

(d) Current and past execution plans devised by planner module, implemented by executor module.

(e) Actual interfaces code to enable informed self- adaptation for various operating aspects (incl. new AI/ML, scheduling and resource management algorithms)

(f) Pre-stored policies and predefined configurations. While a system administrator periodically updates policies an AI-based autonomic agent is to handle the execution automatically. The distributed knowledge base provides a centralized location of information for the local various running tasks (threads, processes) and exchange information for those tasks executed in coalition with other locations.

ACES approach – distributed knowledge base/communications

Communications:

- among large numbers of heterogeneous “things” will have to be flexible and adaptive to rapidly changing situations and local needs as well as robust with respect to time-varying topologies and network failures.
- Highly intelligent automation will be required to continually allocate and reconfigure the communication network’s resources.
- The decentralized architecture needs to leverage communication-constrained distributed consensus algorithms that
  - operate seamless in-network, offering the possibility of optimization and data processing in assets directly at the edge;
  - are parallel, by exploiting possibly onboard multicore architectures, to accommodate the need of fast (real-time) processing and optimization; and
  - are flexible–adapting smoothly to different types of network topologies (e.g., meshed or hierarchical networks)–and robust to possible failures of individual assets, network variability (e.g., time-varying topologies), and asynchronous modus operandi.

The Distributed Knowledge Base (DKB) is also stored on a HDFS backend for long-term storage and collaboration with cognitive services in nearby fog and distant cloud.

ACES approach – natively edge aware

Cloud computing

Mechanisms such as virtualization, auto-scaling, and load balancing are so fundamental to the cloud computing model that most cloud platforms provide them as native platform functions

This is possible because cloud computing provides concepts such as virtual machine specifications, SLAs, datacenter regions, etc., as platform abstractions

These abstractions in turn allow platform operators to develop complex operational mechanisms, such as auto-scaling, VM consolidation, or energy- aware machine provisioning, that work transparently to the platform user

Edge computing

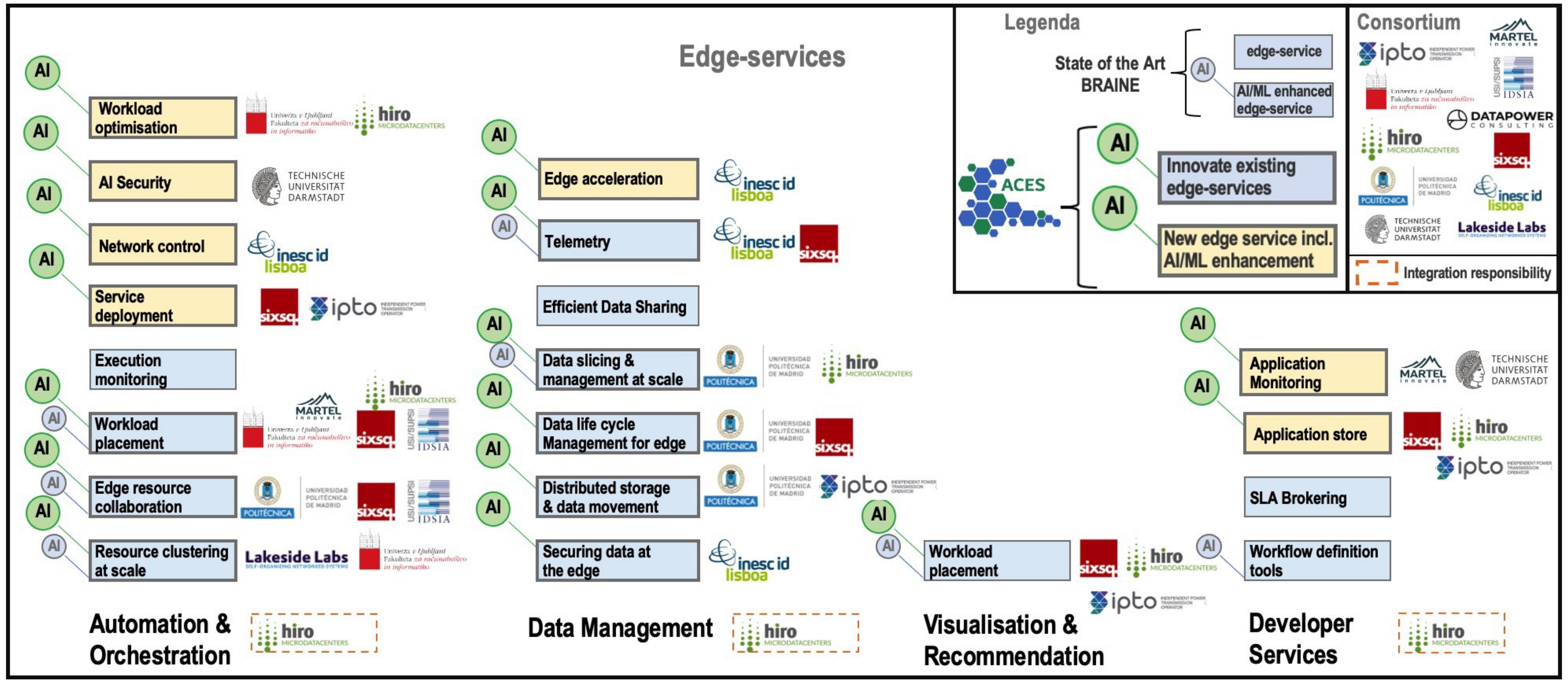
- In edge computing, concepts such as device proximity, data locality, or energy consumption play a critica role in the effectiveness of these mechanisms, and therefore have the same level of significance
- ACES elevates them to first-class citizens, allowing operators to develop reusable operational mechanism that work for a variety of edge applications, rather than something ad-hoc tailored to a specific applicatio

ACES approach – intelligence

Beyond the state of the art in intelligence:

- ACES will go beyond conventional use of ML techniques for merely operating in a reactive mode, rather will be proactive in optimizing the workloads that are deployed across multiple edge-sites
- ACES develops scalable reinforcement learning techniques for workload deployment and optimization that are highly scalable and efficient for responding when target times available to respond or recover are extremely low
- ACES rely on ML model formulations in order to obtain parsimonious and efficient problem representations, thanks to feature extraction/selection and kernel approaches
- ACES will develop coordination techniques based on swarm intelligence that work by exploiting local rules and information to create a global solution via emergence

ACES approach – edge services



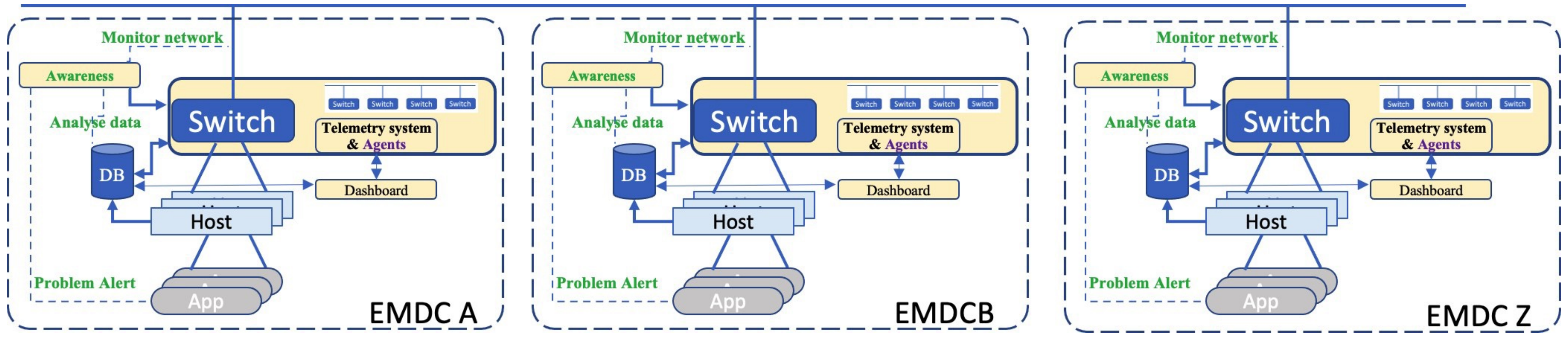
The ACES project builds on the concept and state of the art demonstrated in the Braine project (April 2020-April 2023), a cognitive framework and set of edge-services to support Big Data and AI applications, based on the requirements from 4 usecases: Smart hospitals, Smart City, Industry 4.0 and Supply chain 4.0.

The Braine concept is a solid foundation for ACES's research to push beyond state of the art in the development of the edge-services cloud stack and implement autopoiesis cognitive frameworks.

In particular ACES will: innovate and extend limited capabilities and autopoiesis cognition of existing services and develop new autopoiesis cognitive edge-services;

The cognitive edge-services will also be able to stretch across multiple geographically distributed sites.

ACES approach – network accelerated



ACES will be network accelerated, follow an SDN design augmented with ML-based control, and Network Function Virtualization (NFV).

To guarantee hundreds of Gbps or even Terabit performance with software-based, virtualized network functions, ACES will explore:

- programmable data planes,
- programmable switches (e.g., the Intel/Barefoot Tofino),
- SmartNICs (e.g., FPGA- and NPU-based, DPUs/IPUs), together with new CPU architectures (e.g., Intel Xeon D) and kernel-based mechanisms (e.g., eBPF/XDP)

enabling improvements on network visibility, telemetry, and application performance.

ACES will integrate these high-performance programmable devices to improve security (e.g., to speed up intrusion detection), network performance (e.g., by running AI/ML and cryptographic operations directly in the switches), and service performance (e.g., by offloading specific application functionality).

To address the software engineering challenge that entail supporting this myriad of target platforms, ACES will develop program synthesis techniques based on formal methods (e.g., symbolic execution) and on AI/ML models. Contrary to the state-of-the-art rule-based techniques [Zha20], our synthesis-based approach can explore the entire problem space, improving solution. For scalability we will devise domain-specific search heuristics.

ACES approach – data centricity

ACES will develop a decentralised data management service that directly supports **edge services**, as well **as AI models** that are executed across devices, while cooperating with workload management and network intelligence elements to be aware of bandwidth, resources, management and orchestration.

- data for any service to operate within an edge location will be transparently provided by this component.
- data will provide the execution context for edge services, with the added feature of proactively managing this context in order to adequately enable the service.
- location, performance and reliability requirements will be enforced by ACES
- adequately support the continuously and dynamically changing data flow between devices. i.e., the horizontal flow of data and intelligence from applications across the Edge with decentralised data operations (storage, cleaning) and training of AI.
- a data-centric approach for coherently integrating the heterogeneous systems, algorithms, applications, and models into a single unified framework, capable of aggregating data, acquired and processed by different sources, into a resource exploitable by users.
- ACES' architecture will be designed for scalability, efficiency, dependability/ trustworthiness, interoperability, adaptability and transparency regarding data management and distribution.

ACES approach – open source alignment

All outcomes of the ACES project will be open-sourced and made available to various stakeholders and edge community initiatives through connections of the consortium members, and were not available through official channels.

At the beginning of the project to support requirements and state-of-the-art identification and standardization.

Interaction with the community continues to share ACES progress, design choices and obtain feedback get feedback on the developments and demonstrations.

ACES Use Cases/1

ACES envisions three use-cases: Market place and asset distribution (UC1), Distributed process management (UC2) and IoT based asset monitoring and management (UC3).

The usecases can be found in many different Industry verticals, and place a wide range of requirement on the edge-services cloud, and allow the consortium to demonstrate different selections of services in each usecase. The Energy sector digital transformation is of such magnitude that we found all three usecases in one industry sector and entity, the Greek Transmission Operator, IPTO. The installation of an edge-services cloud infrastructure and the effort for an end-user to migrate their applications to the edge-service cloud is such that working with one usecase partner is very cost and time efficient.

ACES Use Cases/2

Market place & assets distribution

Description of Current Status and Challenges

- Through a Market Clearing algorithm (MC) Local Energy Markets optimize utilization of local energy assets, minimize energy flows, relieve transmission and distribution grids thereby reducing congestions
- As the grid becomes a more fine-grained distribution of demand and supply, the MC becomes a computational challenge given increasing complexity of bid orders from the technically restricted productions assets, bid orders allowing the demand side to set more criteria, the 15min-cycle to match demand and supply by the MC, across various demand and supply locations and entities

Advancements Brought by ACES :

- Has no single point of failure (distributed), fault tolerant
- Provides local powerful storage and processing of data (MC algorithm), for peak performance scalable to other sites
- Provides near real time, low latency, processing
- Secures local storage and edge-wide exchange (multi site) of privacy sensitive data
- Securely distributes data that are classified
- Provides a means of upscaling the computation to a neighbouring area in case of a local mismatch

Distributed Process Management:

Description of Current Status and Challenges

- Energy transmission develops into coalitions of smart decentralized management systems and smart substations, with a diverse set of monitoring devices, advanced measuring and control systems of field conditions in sub-seconds.
- Fundamental for digital transformation is the move from Central Dispatch Systems and BaU schema, to resilient distributed and adaptive grid management.

Advancements Brought by ACES :

- Power flow tool; calculates microgrid power flows based on data from sensors (CTVT, PMU, etc).
- Area Control Error tool; calculates the local Area Control Error (ACE) for the microgrid
- ML microgrid health algorithm: offers information, sensor-based measurements, health predictions, alert
- Automation and AGC; EN 61850-protocol based automatic control in the smart substations, alerts AutomatiGenerator Control (AGC) in the case of an ACE discrepancy
- Digital twin for automation: incorporates IoT data from Grid Assets and Measurement Points and creates a real time digital twin of the different interconnected microgrids for automation validation

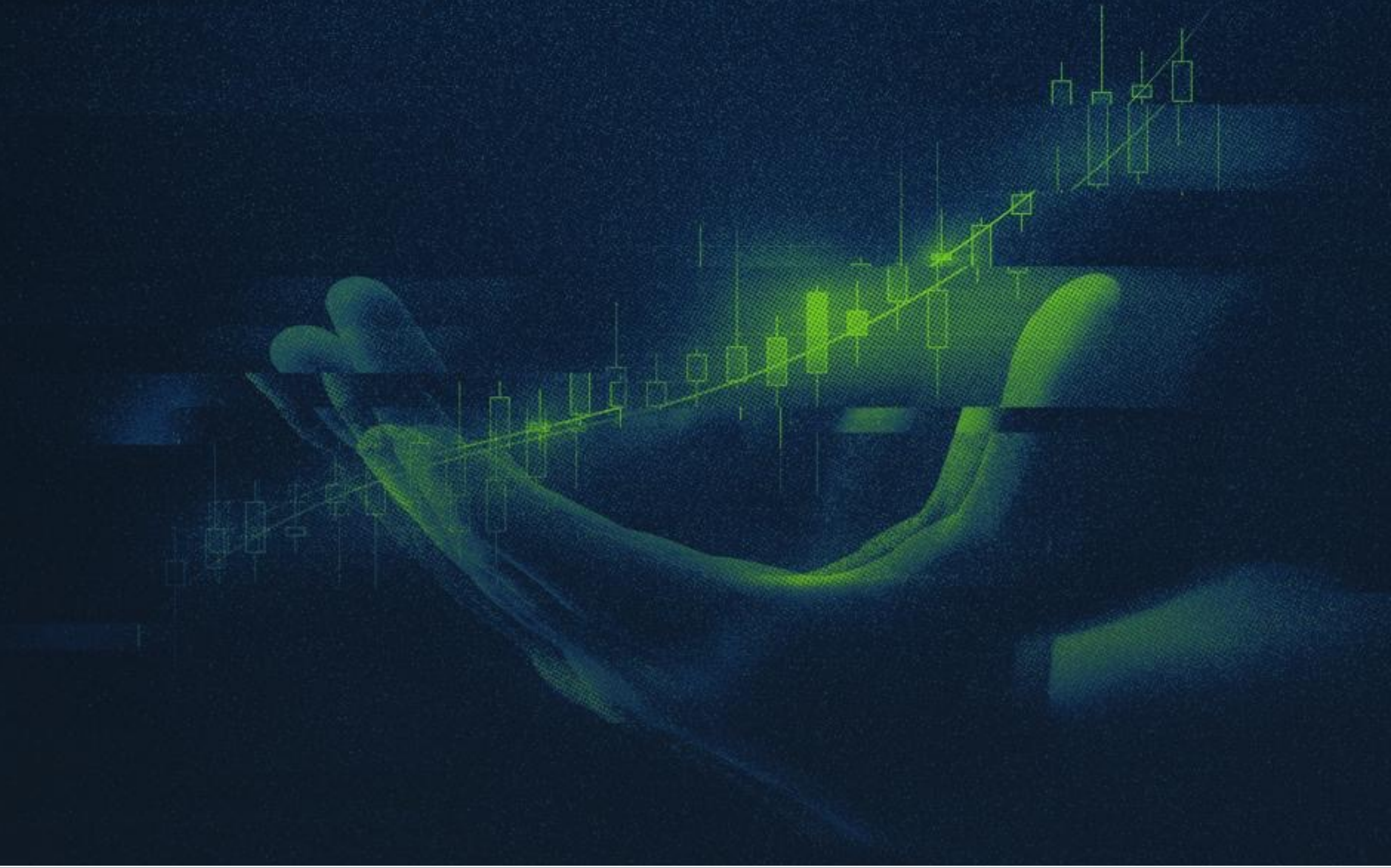
IoT based Asset Monitoring and Management

Description of Current Status and Challenges

- Network operators have periodic planning cycles for assets-maintenance
- Periodic on-site inspections can be replaced by advanced metering, sensor data and GIS systems for real time outage detection, prediction and more reliable investment planning and deferral
- Big data analytics offer network operators insights in asset loss of life and fine-grained right-sizing of the assets, risk-based asset management and reduce the total cost of ownership
- The big data sensor streams from 3 microgrids (Cyclades, Crete, Attica) will be analysed (baseline assets)

Advancements Brought by ACES :

- Providing the services to support federated learning across multiple sites for the ML health index tool
- Supporting multi-level and multi-site maintenance scheduling by using use both local and distributed data (Logistics Data, Human in the Loop)
- Optimal handling of the IoT Sensor Swarms and their produced data streams



Results

Key Outcomes

Autopoiesis cognitive edge-services cloud: intelligent highly automated full edge-services stack

Swarm technology-based methodology and implementation for orchestration of resources in the edge

Platform with awareness tools, AI/ML agents for workload placement, service and resource management, data and policy management, telemetry and monitoring

Edge-wide workload placement and optimization service and implementation of scalable ML to facilitate this

Autopoiesis agents to safeguard stability in situations of extreme load and complexity, specifically for workload placement, service and resource management, data and policy management, telemetry and monitoring

App store for classification, storage, sharing and rating of AI models used in ACES, automated deployment of ACES infrastructure AI models

Impact

Facilitate the establishment of a data infrastructure at pan-European, specifically aimed to share and re-use of data in the context of EOSC, GAIA-X, and the common European data spaces, all in various application areas.

Ensure the sharing and manipulation of data in compliance with needs of data processors and data rightsholders and other stakeholders in line with the current national and European legislation (e.g., GDPR)

Boost the efficiency of data operations from the point of view of energy footprint and increase the visibility and usage of renewable energies for fuelling the storage of data, in this way facilitating the reduction of the CO2 emissions from fossil fuel energy sources

Satisfy the growing demand for energy efficient datacentres by providing blueprints on how to ensure power efficiency and high performance

Contribute to the European cybersecurity and privacy preserving efforts by facilitating the safe and secure data handling thereby

Facilitate the fair, humane, and ethically sound data collection, processing, and manipulation, following the principles of responsible/trustworthy AI.

Consortium

Partners

hiro MICRODATACENTERS

inesc id lisboa

MARTEL innovate

TECHNISCHE UNIVERSITÄT DARMSTADT

Lakeside Labs

UNIVERSIDAD POLITÉCNICA MADRID

POLITÉCNICA

Univerza *v Ljubljani*

Scuola universitaria professionale della Svizzera italiana

SUPSI

ipto

INDEPENDENT POWER TRANSMISSION OPERATOR

sixsq.

DATAPOWER CONSULTING

ACES

Thank you!

Introductory Project Presentation

Speaker Name

Funded by
the European Union