

DECICE Final Project Webinar

📅 12.11.2025

Host: SYNYO GmbH



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.

Agenda



Time	Topics
10:00 – 10:05	Welcome & Introduction
10:05 – 10:20	Project Overview & Key Achievements
10:20 – 10:35	DECICE Solutions & Technological Innovations - Overview
10:35 – 11:10	DECICE Use Cases: Deep Dive
11:10 – 11:25	Q&A and Interactive Discussion
11:25 – 11:30	Closing Remarks

HOUSEKEEPING RULES



The session will be **entirely recorded** and published on the DECICE Project website.



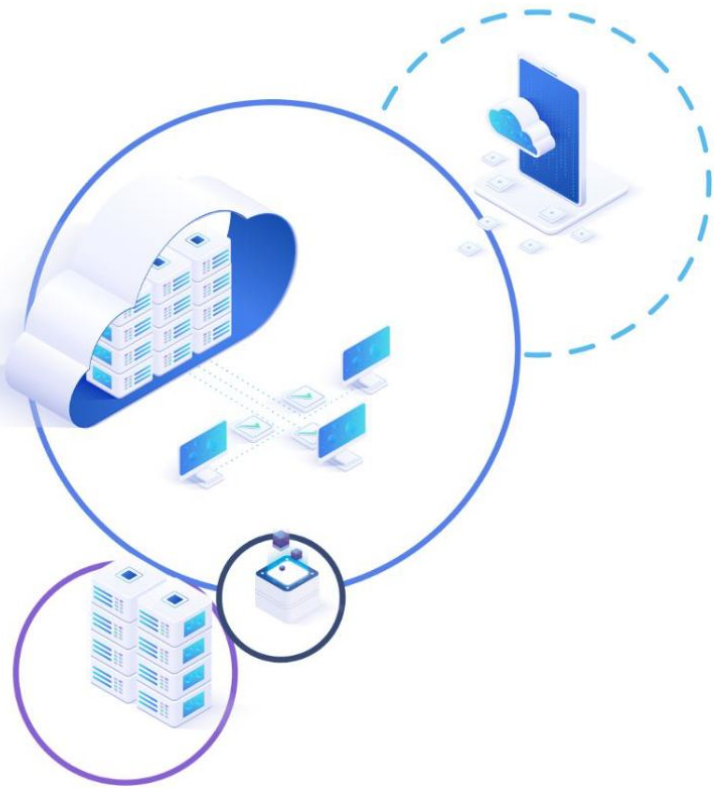
All participants except speakers and moderators will be **muted by default**.



Feel free to **post your questions** in the **chat**.



If you would like to **speak**, **raise your hand** and wait for the moderator to give you the floor.



DECICE

Device-Edge-Cloud Intelligent
Collaboration Framework

Felix Stein
The University of Göttingen

Overview

Key Objectives

Domains of Intervention

Open-source cloud management framework

HPC, Cloud Infrastructure, Edge Computing

AI scheduler for dynamic job scheduling

AI, Sustainable Computing, Energy Efficiency

API for advanced control of all resources

Networking, HPC, Cloud, Edge

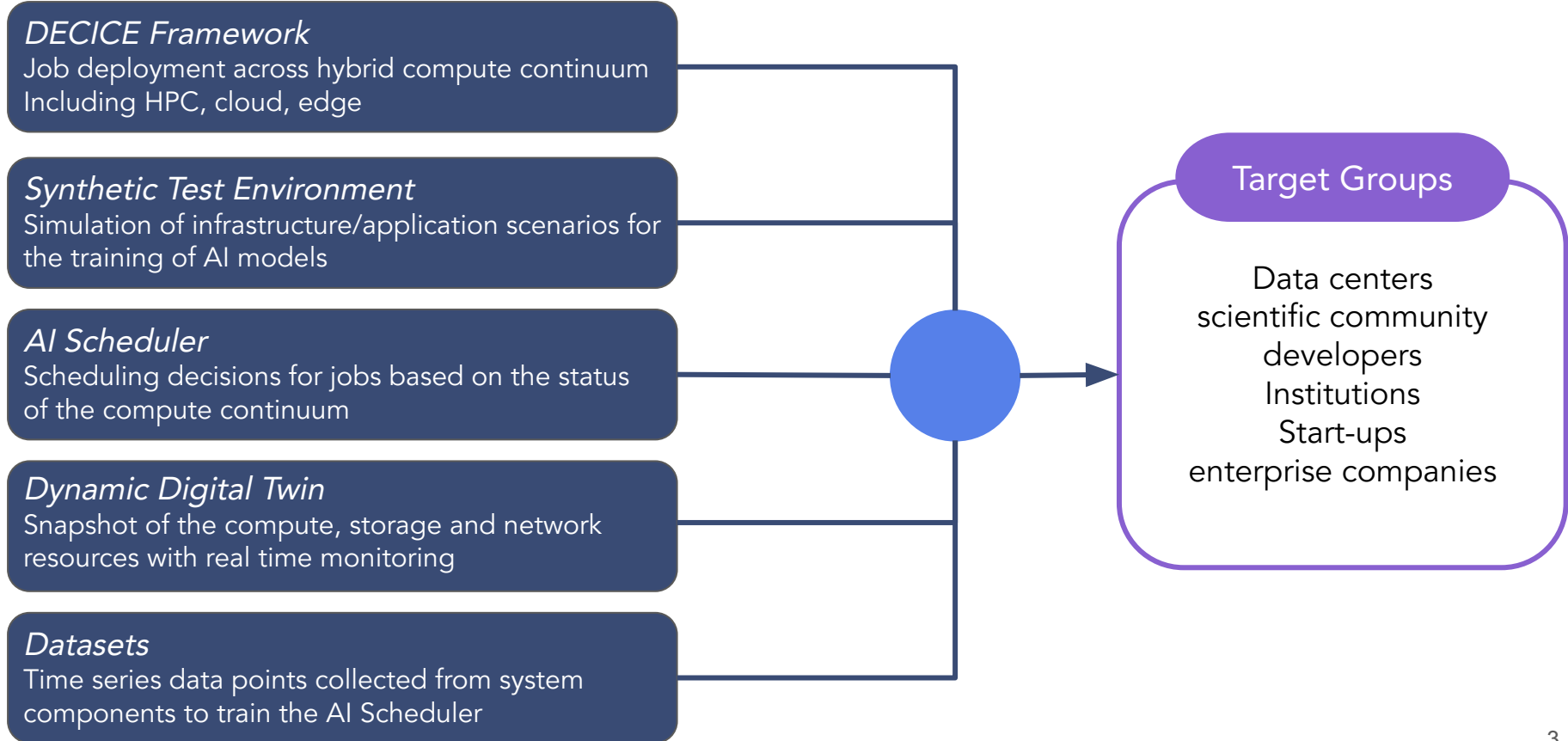
Digital twin representing the entire system

AI, System Modelling, Monitoring

Secure, compliant service deployment solutions

Cybersecurity & Compliance

Key Exploitable Results

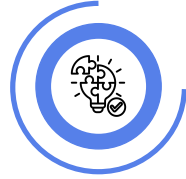


Challenges



Problem

1. Real data for training the AI Scheduler
2. Job scheduling on HPC clusters via the framework



Solution

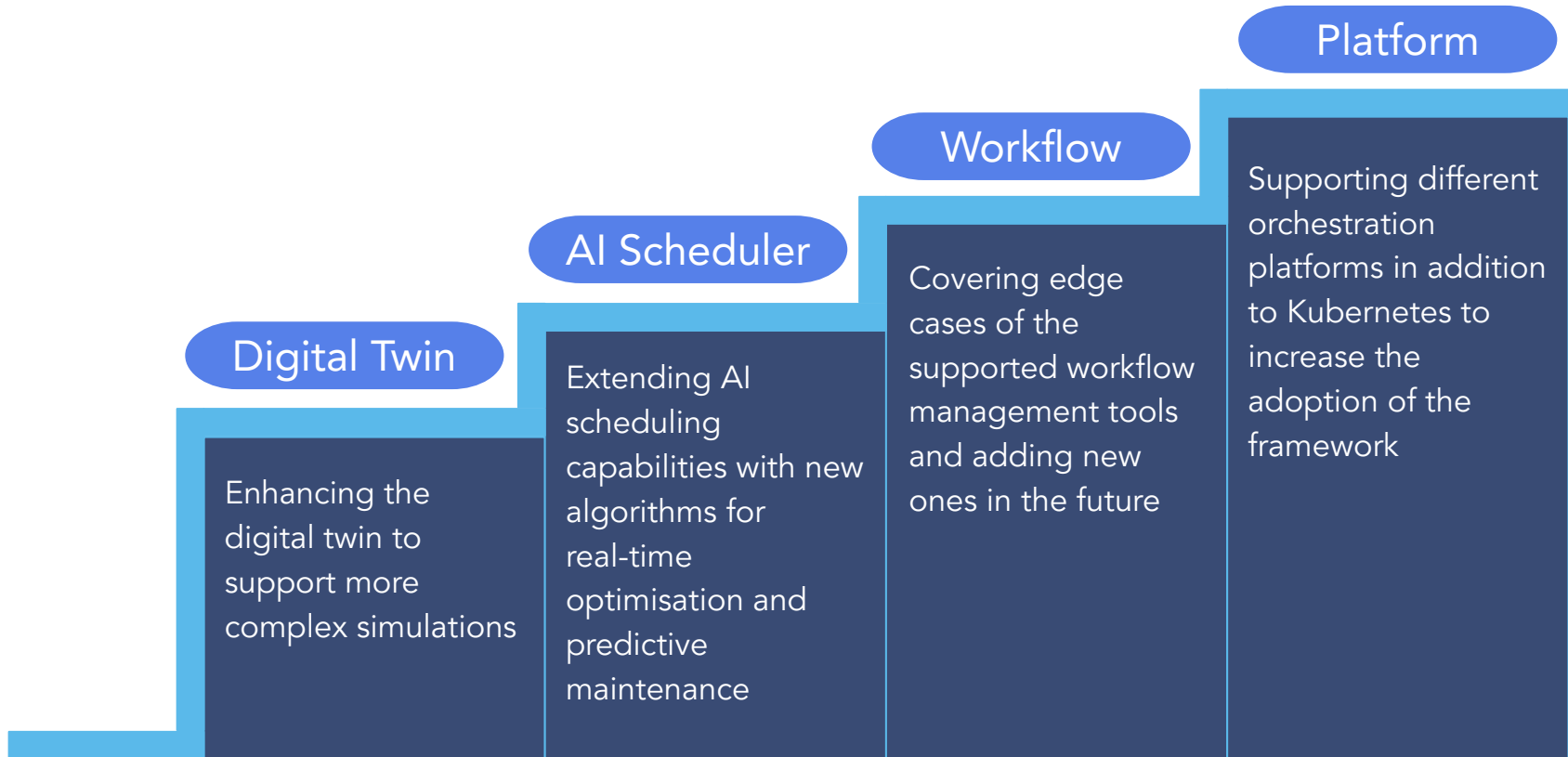
1. Collecting time series data from Prometheus and using open source datasets
2. Utilizing Slurm APIs to communicate with HPC



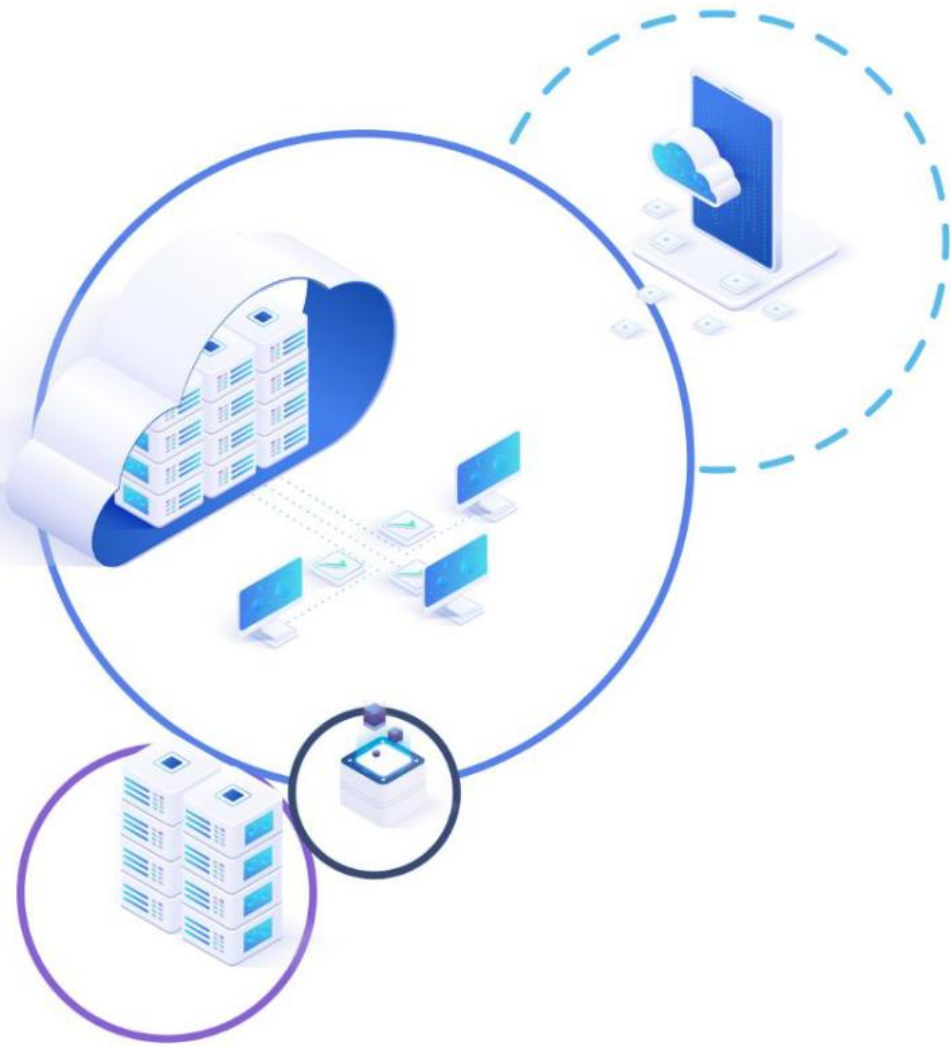
Outcome

1. Baseline data registry to train AI models
2. More control over HPC jobs and accounting compared to other open source alternatives

Way Forward



QUESTIONS



DECICE 





**Device-Edge-Cloud Intelligent
Collaboration Framework**

Mirac Aydin
University of Göttingen, Germany





Components

1. Virtual Training Environment
2. Digital Twin
3. AI Scheduler
4. Slurm Integration

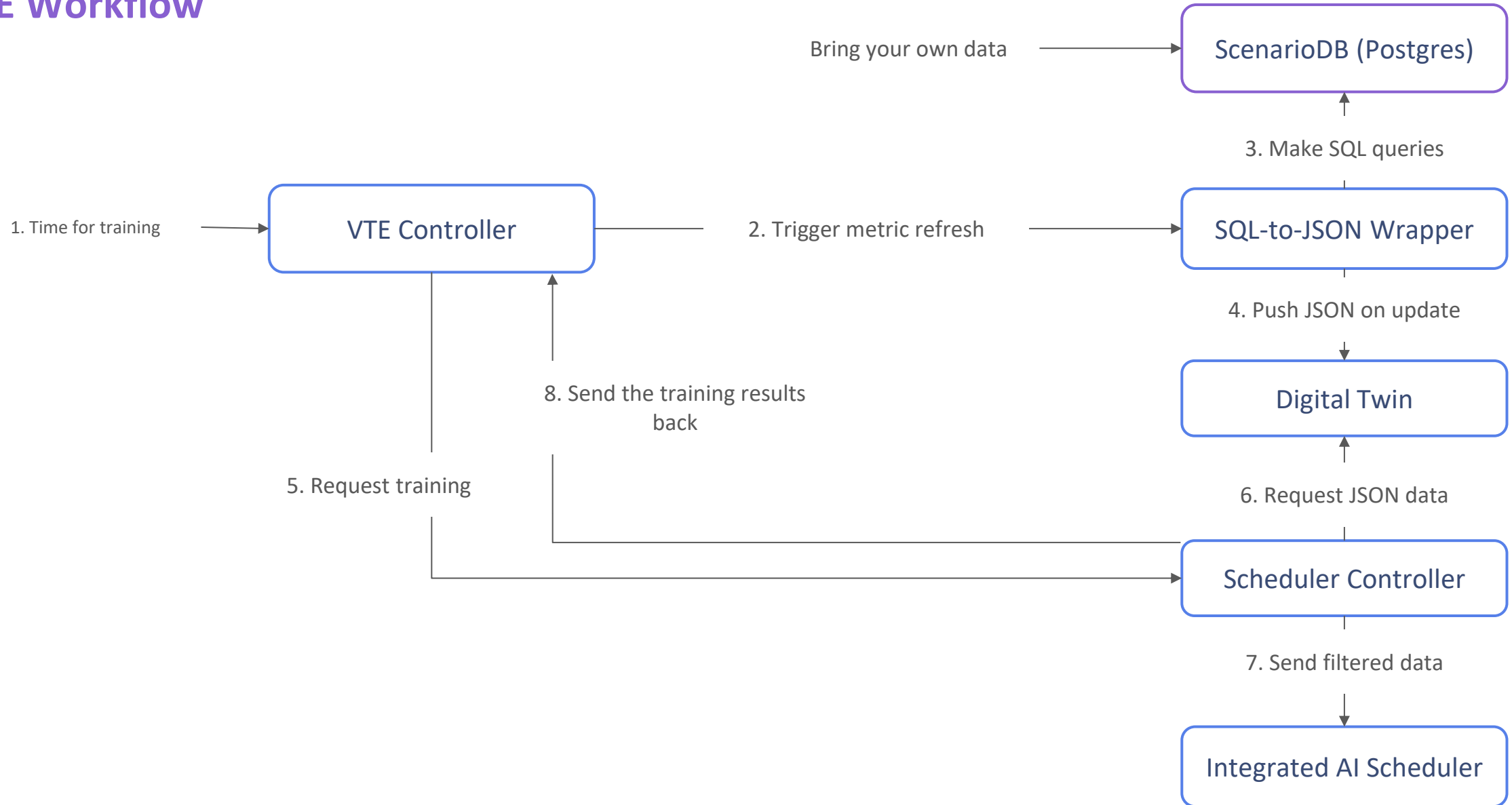
Virtual Training Environment

-  Synthetic test suite replicating system scenarios for evaluation and training
-  Simulates active and pending processes to mimic realistic conditions
-  Integrates with a digital twin for data injection in controlled experiments
-  Enables AI scheduler training in a safe, non-production environment

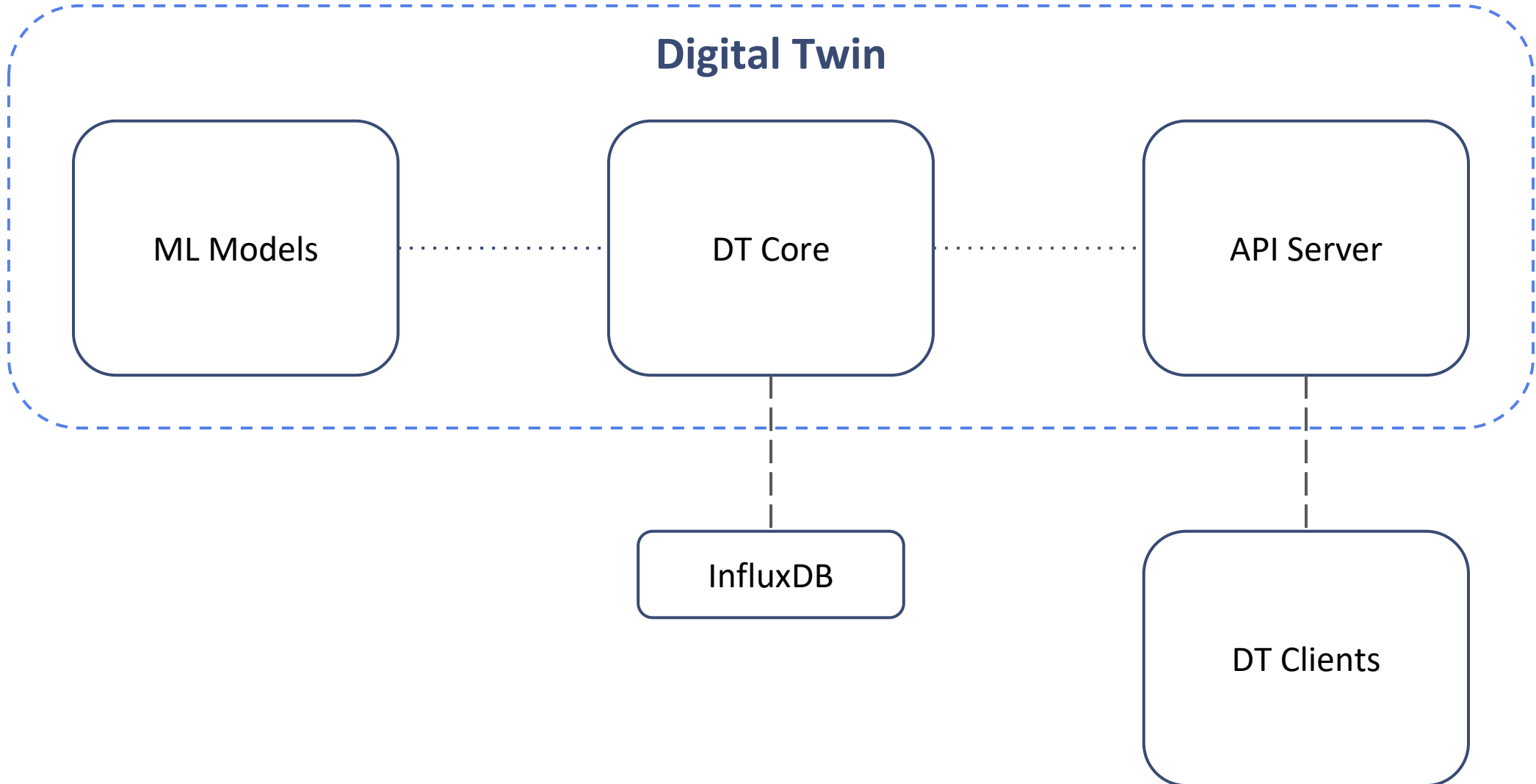
Advantages of VTE

-  Trains AI scheduler faster than real time, speeding up development
-  Tests with predefined scenarios for reliable performance assessment
-  Creates an early pre-trained model before production deployment
-  Improves the scheduler using real production data after initial training





VTE Workflow



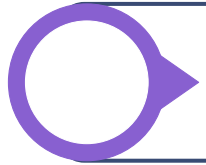
Digital Twin



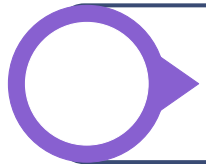
Advantages of Digital Twin

-  **Real-time monitoring:** Tracks node, job, and network status continuously
-  **AI-driven insights:** Uses anomaly detection and energy prediction
-  **Integrated MLOps:** Built on Kubeflow for streamlined training & deployment
-  **Sustainability:** Predicts carbon intensity to support environmental impact

Integrated AI Scheduler



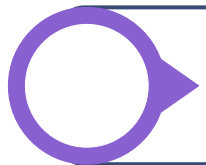
AI-driven scheduler for Kubernetes-based HPC and cloud environments



Learns from real-time and historical data using RNN, TCN







A fuzzy access controller to manage resource access across nodes

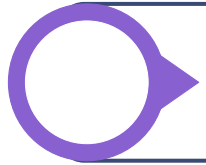


Interfaces with a Digital Twin to apply system models and data for decisions

Advantages of AI Scheduler

-  Maps jobs to nodes using Digital Twin data for optimal workload placement
-  Minimizes power consumption through energy-aware job scheduling
-  Dynamically adjusts job assignments in real time for efficiency & throughput
-  Learns from past outcomes to improve decisions & prevent resource conflicts

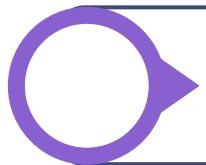
Slurm Integration



Submits HPC jobs directly via Slurm REST APIs



Supports detailed parameter definition like Slurm batch scripts

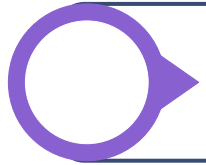


Provides a single, unified API endpoint for communication



Offers fine-grained accounting of user/group core hour usage

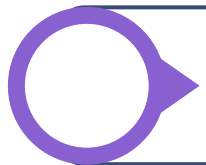
Advantages of Slurm Integration



Uses native Slurm APIs instead of a shared service account



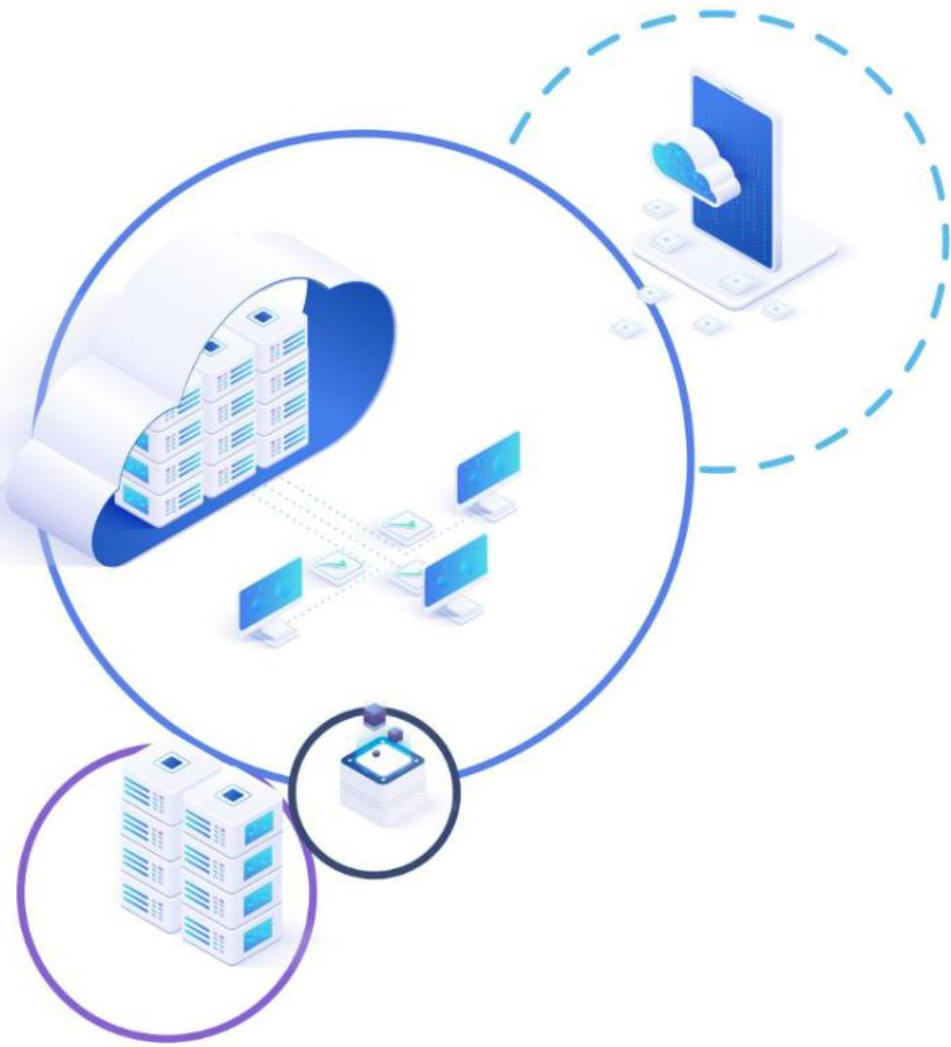
Enables secure job submission on behalf of other users



Improves system security with UNIX socket implementation



Simplifies job submission using JWT-based authentication



DECICE 

Use Case 1
Intelligent Intersection with VRU Detection

Marmara University, BigTRI
Turkey

Use Case 1: Intelligent Intersection with VRU Detection

Marmara University (UC Lead), BigTRI

Use Case Description:

- Detection and classification of the **Vulnerable Road Users (VRUs)** via image processing
 - Capture frames through **stationary roadside cameras**
 - Object detection models are employed to evaluate the spatiotemporal variations of road users
- **Identify potential risks** that may cause harm to VRUs
- **Notify nearby vehicles** with through **RSUs** (Roadside Units), using **V2X communications**

Use Case Objective:

- Demonstrate DECICE project outputs in improving Intelligent Transportation Systems, in terms of:
 - Service availability
 - Maintaining time-criticality
 - Reliable operation
 - Reduced energy consumption

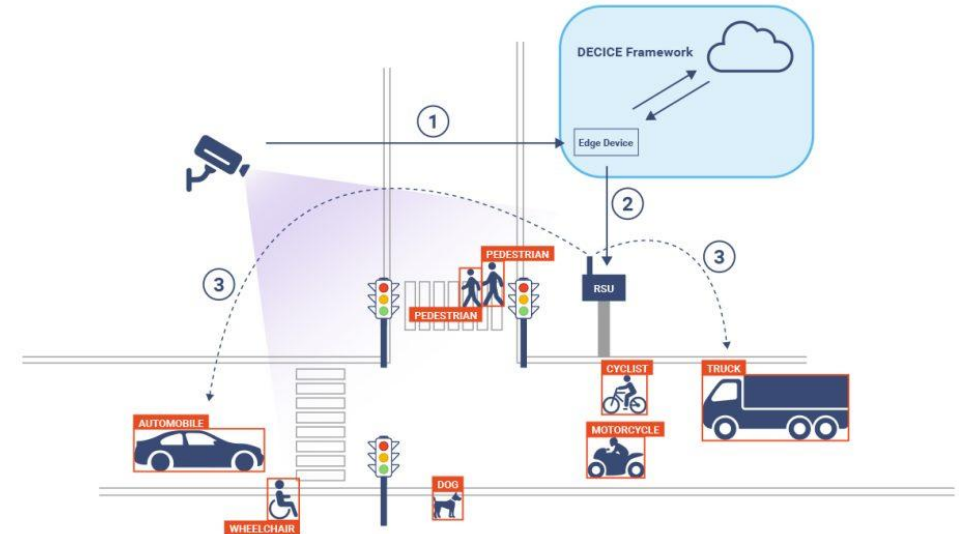


Illustration of the use case setup showing VRUs and Vehicles

Use Case 1: Intelligent Intersection with VRU Detection

Marmara University (UC Lead), BigTRI

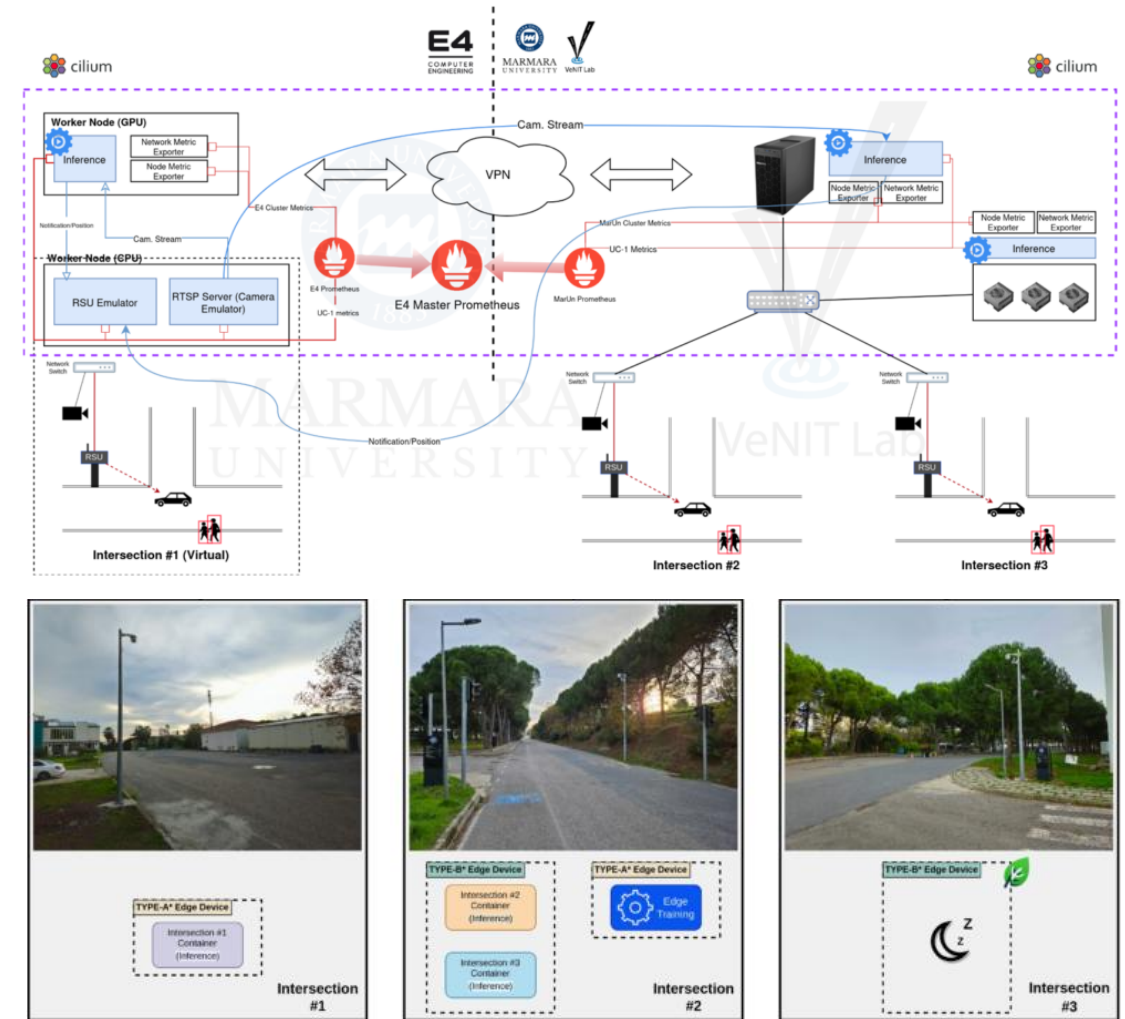


Constraints based on ETSI and 3GPP standards:

- Overall delay from camera to vehicle should be **under 300ms**
- Transferring the data from the node running the inference to the vehicle should be in **less than 100ms**
- Object detection model (inference) should be able to process **at least 10 frames per second ($\geq 10\text{Hz}$)**

Approach:

- **Multi-intersection** setup in **VeST-Ave**
 - Use of shared edge computing resources among the intersections
- **Multi-cluster** deployment by integrating E4 and Marmara University clusters
- Effective **scheduling** of workloads and **horizontal scaling** to stay within the constraints
 - Optimizing energy consumption through bin packing
 - Ensuring reliability and availability of the system
- **Monitoring network metrics in real time** between nodes and IoT devices



Use Case 1: Intelligent Intersection with VRU Detection

Marmara University (UC Lead), BigTRI



Achievements:

Scheduling of the workloads, leveraging the Digital Twin and real-time network/node metric monitoring

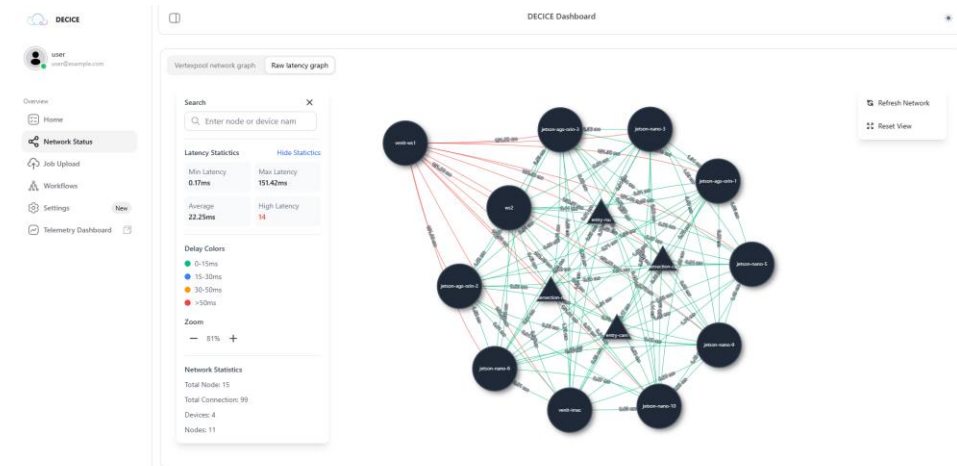
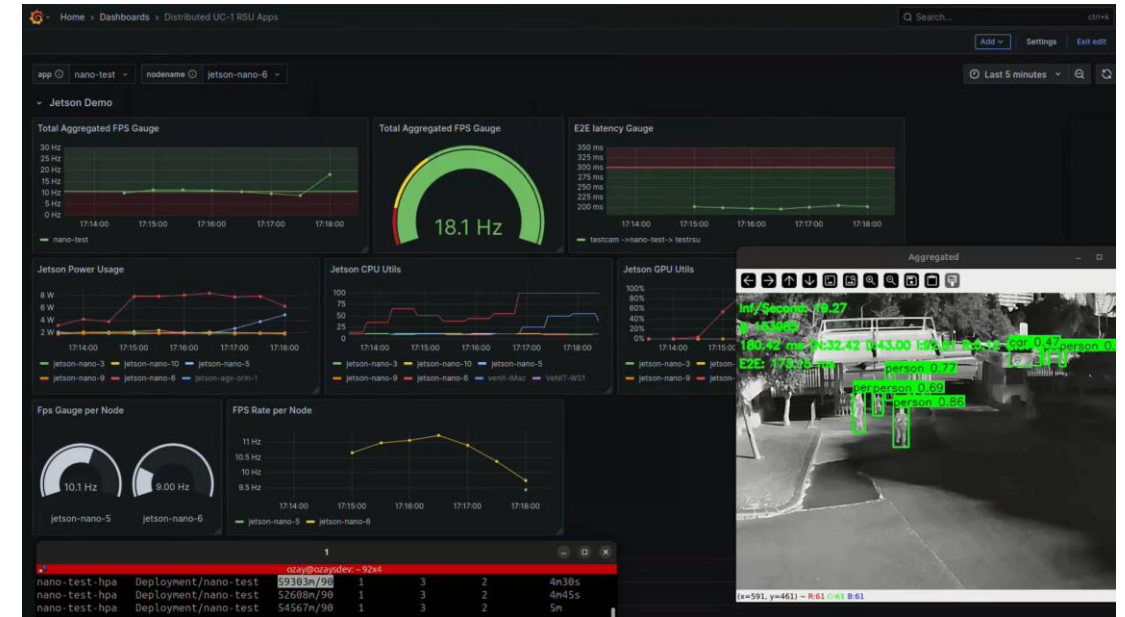
- Ensure safety - by operating below critical levels in terms of latency
 - Serve up-to-date data to nearby vehicles by placing workloads in relevant nodes
- Leverage bin packing to reduce energy consumption

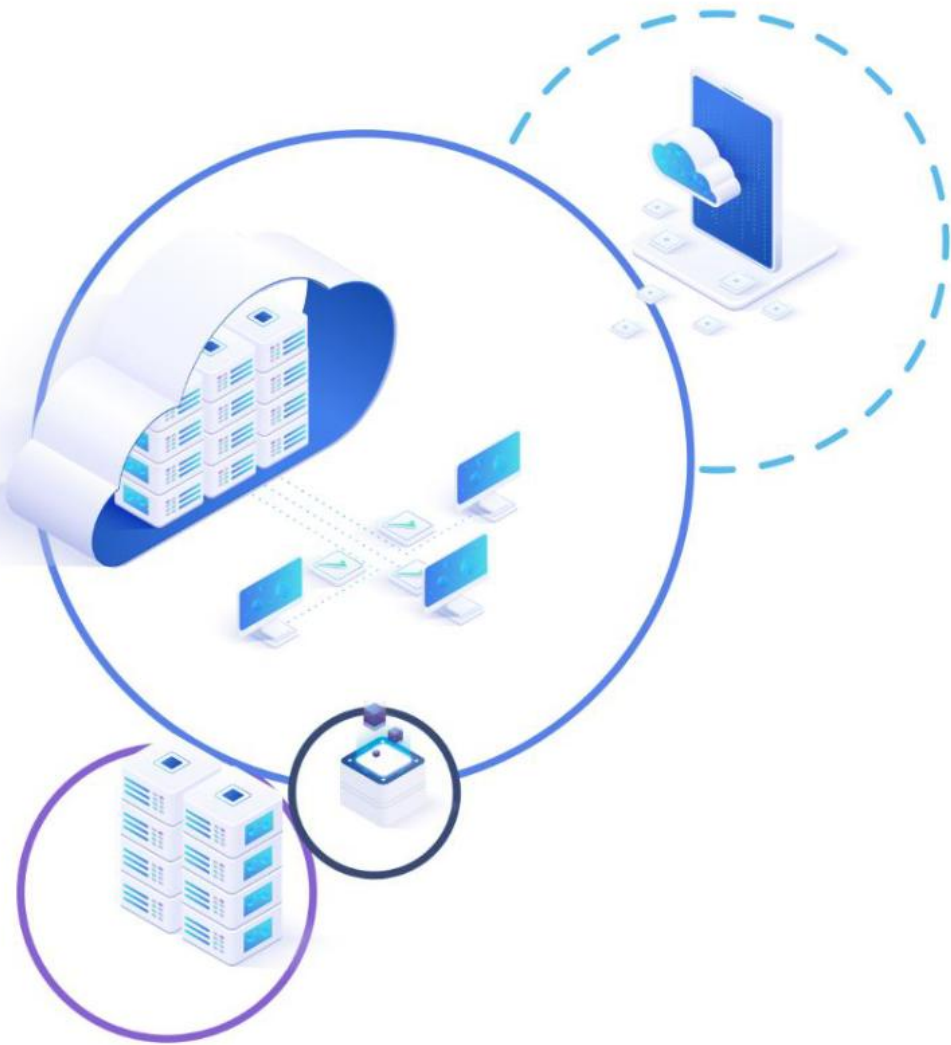
Adaptive replacement and utilization of the workloads, and horizontal scaling:

- Keep latency under safe limits even under adverse conditions
 - Rescheduling based on continuously monitored network and device metrics (through Digital Twin)
- Use of cloud resources when the model deployed on the edge node cannot provide sufficient accuracy (joint inference using SEDNA)
- Horizontal scaling of the workloads based on processing capability

Reduce energy consumption of the ITS infrastructure:

- Smart utilization of performant versus energy-efficient nodes under real-time constraints
- **5-10% decrease** in energy consumption compared to the baseline





DECICE 

Use Case 2
MRI Scans

University of Göttingen
Germany

Use Case 2 - MRI Scans

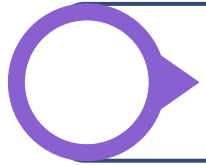
Definition



Providing fast image analysis and preventing data storage on unregistered devices.

- Reliable processing
- Fast analysis results (< 10 minutes)

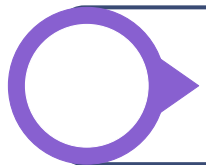
Analysis



FastSurfer for MRI analysis



Segmentation pipeline for accurate segmentation and volumetric calculation of the whole brain and selected substructures



Surface reconstruction pipeline for cortical surfaces, mapping cortical labels



Segmentation analysis can be completed under desired time limit

Use Case 2 - MRI Scans

Results on Kubernetes Cluster

GPU + CPU / Segmentation + Surface Reconstruction

1 thread + 1GPU	8 threads + 1GPU	16 threads + 1GPU	32 threads + 1GPU
43 Min	29 Min	29 Min	29 Min

CPU-Only / Segmentation + Surface Reconstruction

1 thread	8 threads	16 threads	32 threads
90 Min	-	-	33 Min

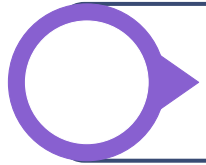
GPU + CPU / Only Segmentation

1 thread + 1GPU	8 threads + 1GPU	16 threads + 1GPU	32 threads + 1GPU
3 Min 26s	83s	74s	71s

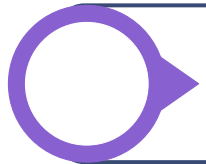
CPU-Only / Only Segmentation

1 thread	8 threads	16 threads	32 threads
29 Min	8 Min	6 Min 5s	5 Min 17s

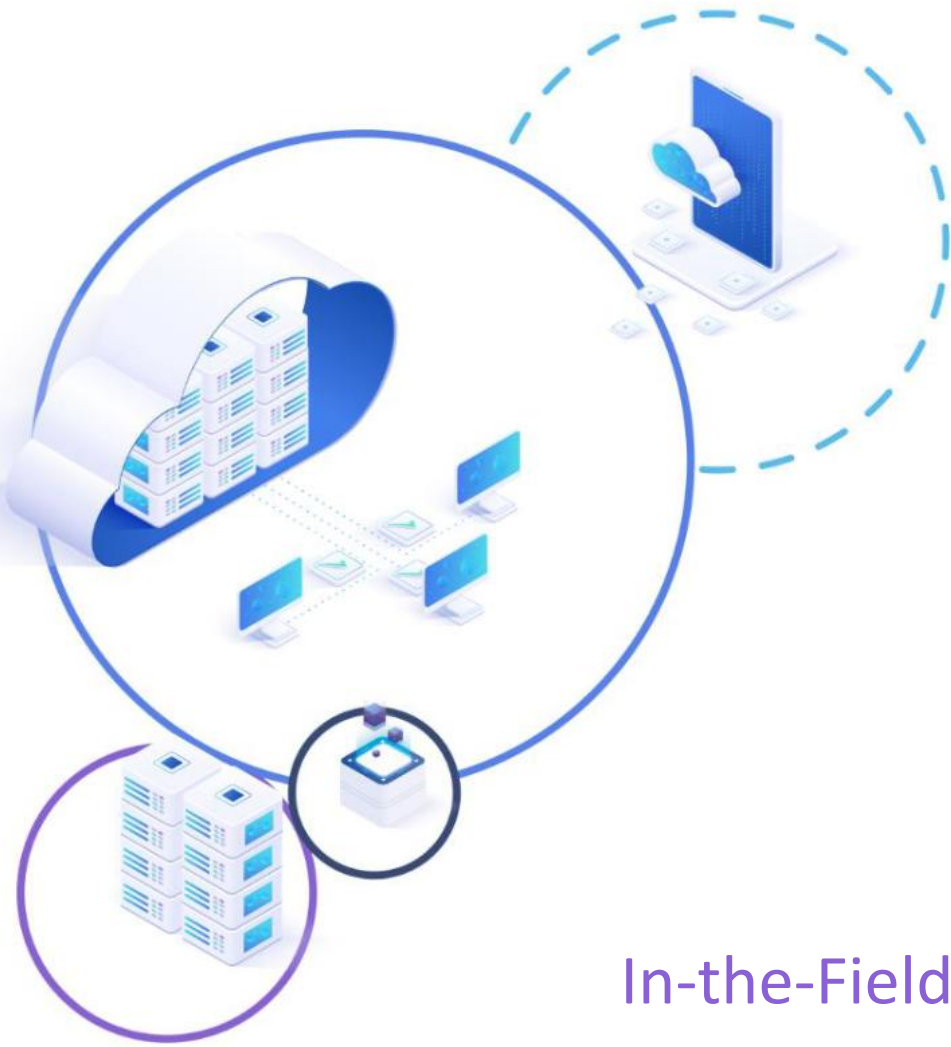
Future Work



Use case scheduling with Integrated AI scheduler



Integration with the real systems



DECICE 

Use Case 3
In-the-Field Intelligence Supporting Emergency Response

Alma Mater Studiorum - Università Di Bologna
Italy

Use Case 3 - In-the-Field Intelligence Supporting Emergency Response

Use Case Overview



Use case definition

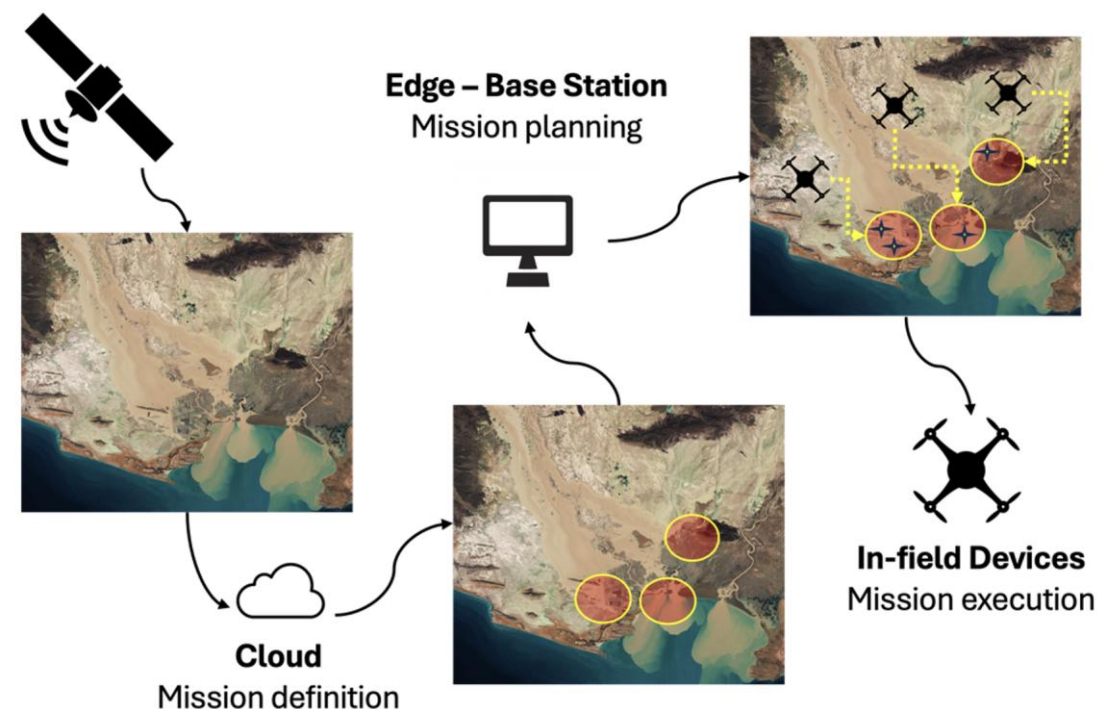
Help emergency response operators with data incoming from the field

- Leverage a priori knowledge by **processing satellite images** to identify meaningful areas and create an **exploration plan**
- Gather **data from the field** with autonomous **quadrotors**
- **Onboard processing** gathered data to extract information (e.g., search and rescue)

Use case objective

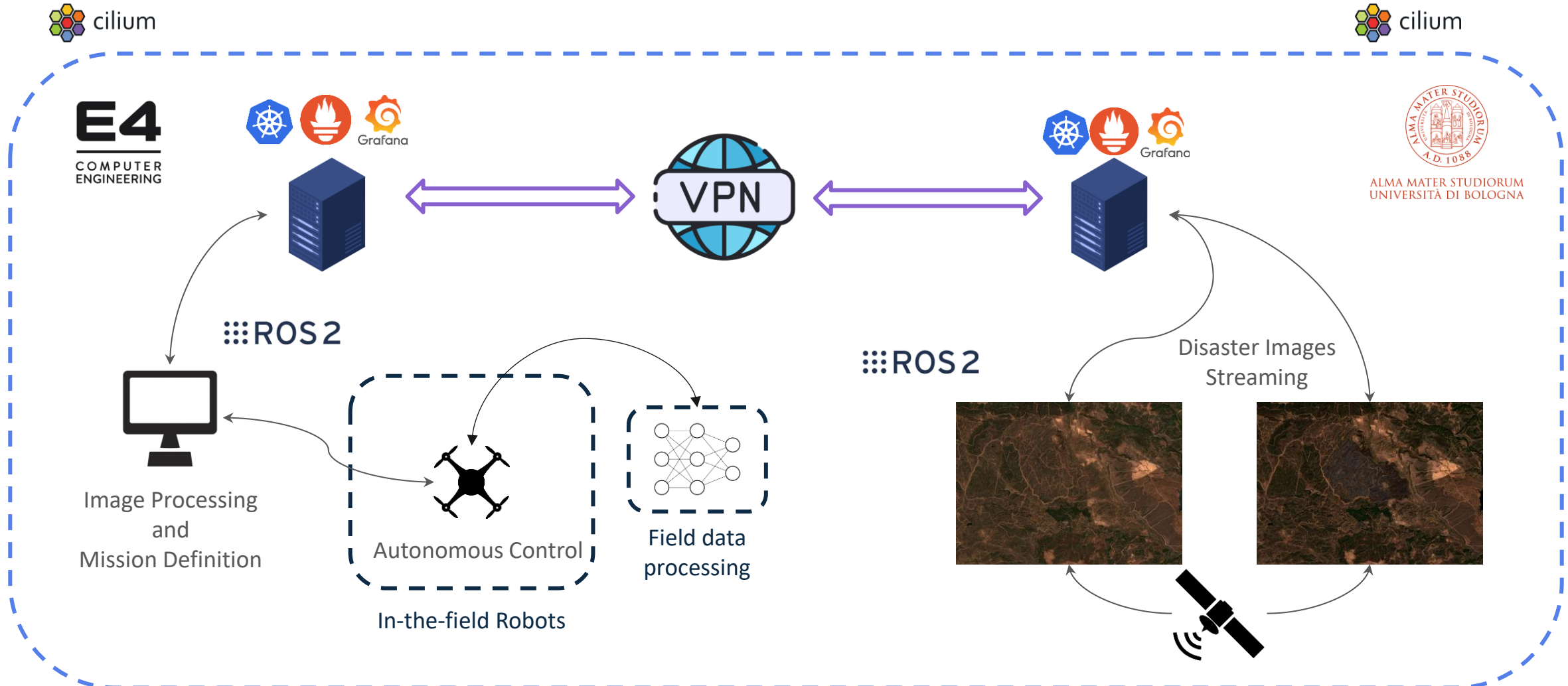
Demonstrate that DECICE can improve emergency response systems by:

- Keeping the image processing and field data tasks latency compatible with the application
- While enabling the computing continuum to
 - Save energy on the edge platforms (quadrotors)
 - Be robust to edge platforms loss



Use Case 3 - In-the-Field Intelligence Supporting Emergency Response

Use Case Architecture



Use Case 3 - In-the-Field Intelligence Supporting Emergency Response

Use Case Achievements on Kubernetes Cluster



Successfully deployed the Use Case on Kubernetes cluster:

- Ensure that the delay remains bounded to satisfy operational constraints

- Satellite image processing (700 samples)

- KPI: < 1 minute

Mean (ms)	Min (ms)	Max (ms)	Std (ms)
1054	97	1175	155

- Field data processing (15 samples)

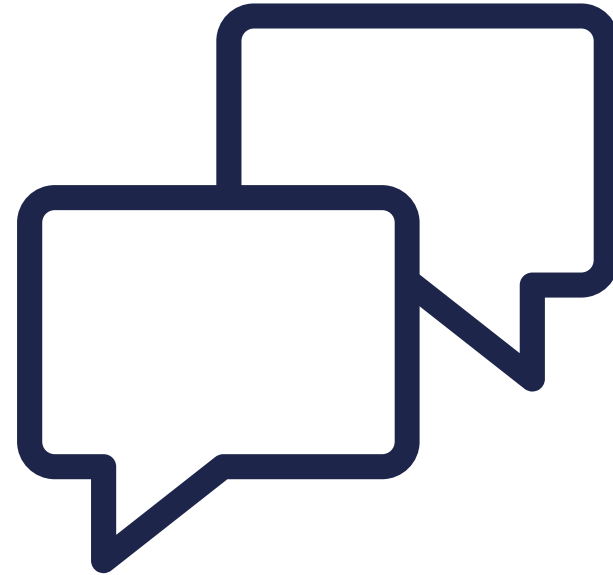
- KPI: < 10 seconds

Mean (ms)	Min (ms)	Max (ms)	Std (ms)
493	465	606	33

- Edge replacement/substitution

Mean (s)	Min (s)	Max (s)	Std (s)
341.37	332.83	349.74	5.91

Measurements of energy KPI and comparison with the DECICE AI scheduler are ongoing.



Discussion

DECICE

[Videos](#) [Publications](#) [Events](#) [News](#)



DECICE Project

Together we will shape the future.

About DECICE

DECICE aims to develop an **AI-based, open and portable cloud management framework** for automatic and adaptive optimization and deployment of applications in a federated infrastructure, including computing from the very large (e.g., HPC systems) to the very small (e.g., IoT sensors connected on the edge).

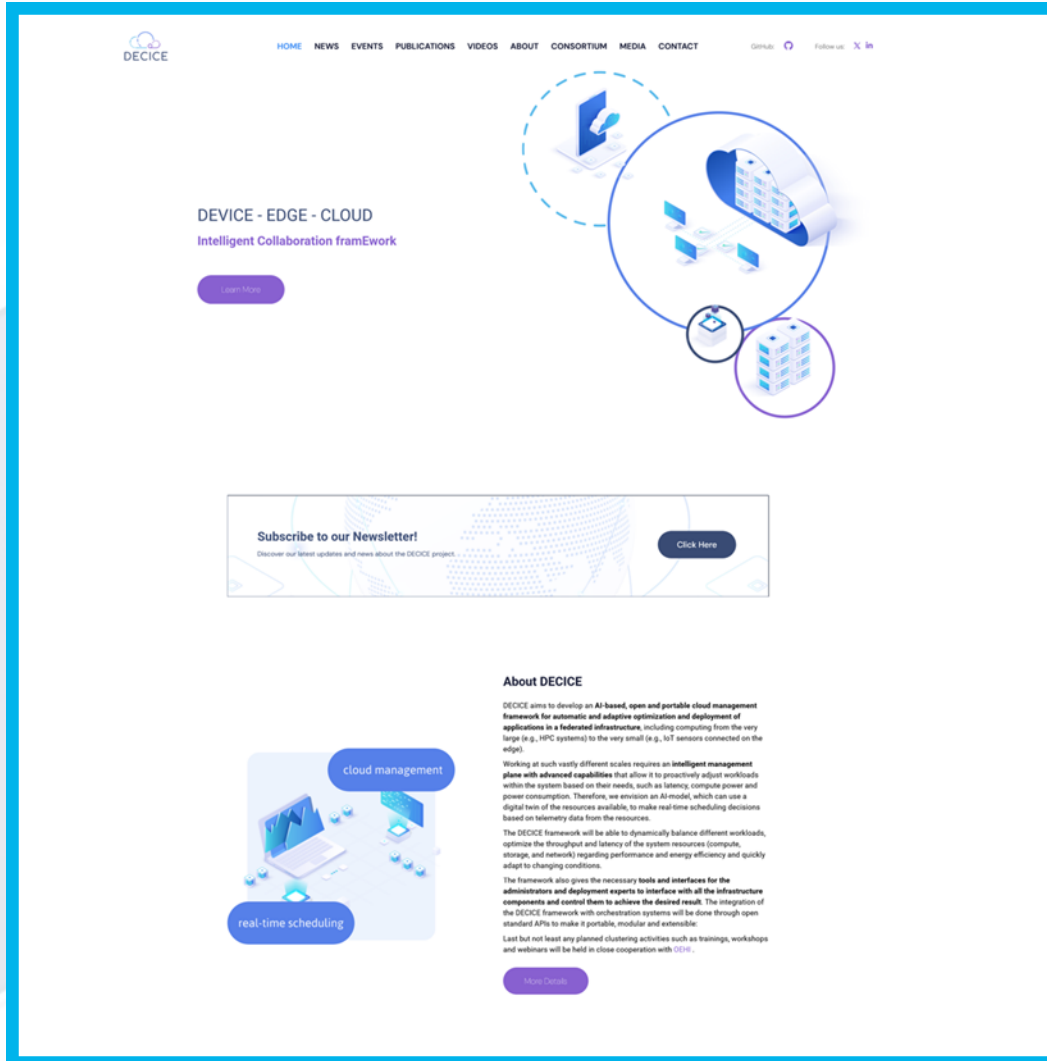
[READ MORE](#)



Newsletter

DISSEMINATION

Project website



Project Website

DISSEMINATION

Social media



LinkedIn



X

