Towards Integrated Hardware/Software Ecosystems for the Edge-Cloud-HPC Continuum

Gabriel Antoniu, Inria

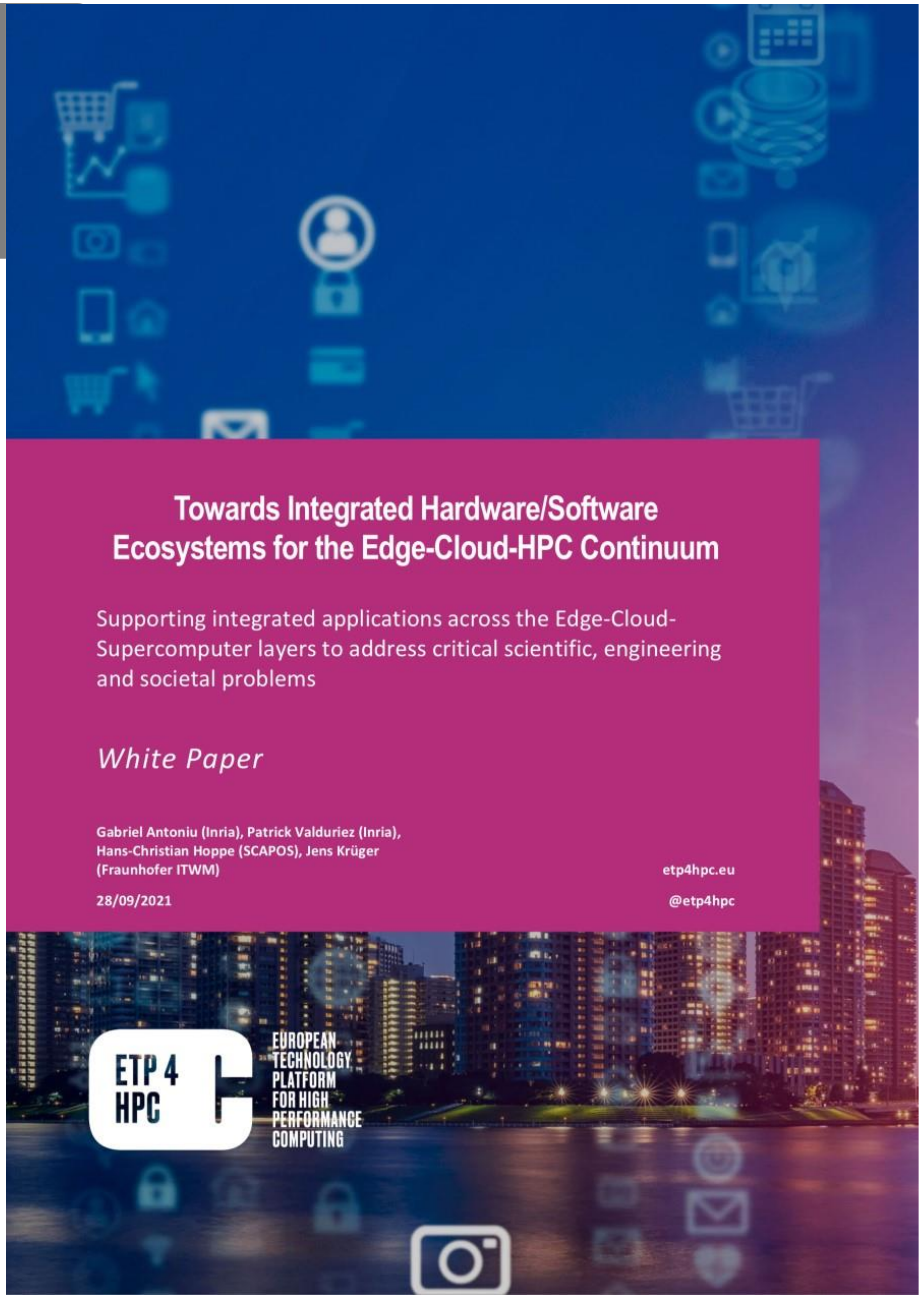
Hans-Christian Hoppe, scapos AG

26 April 2022

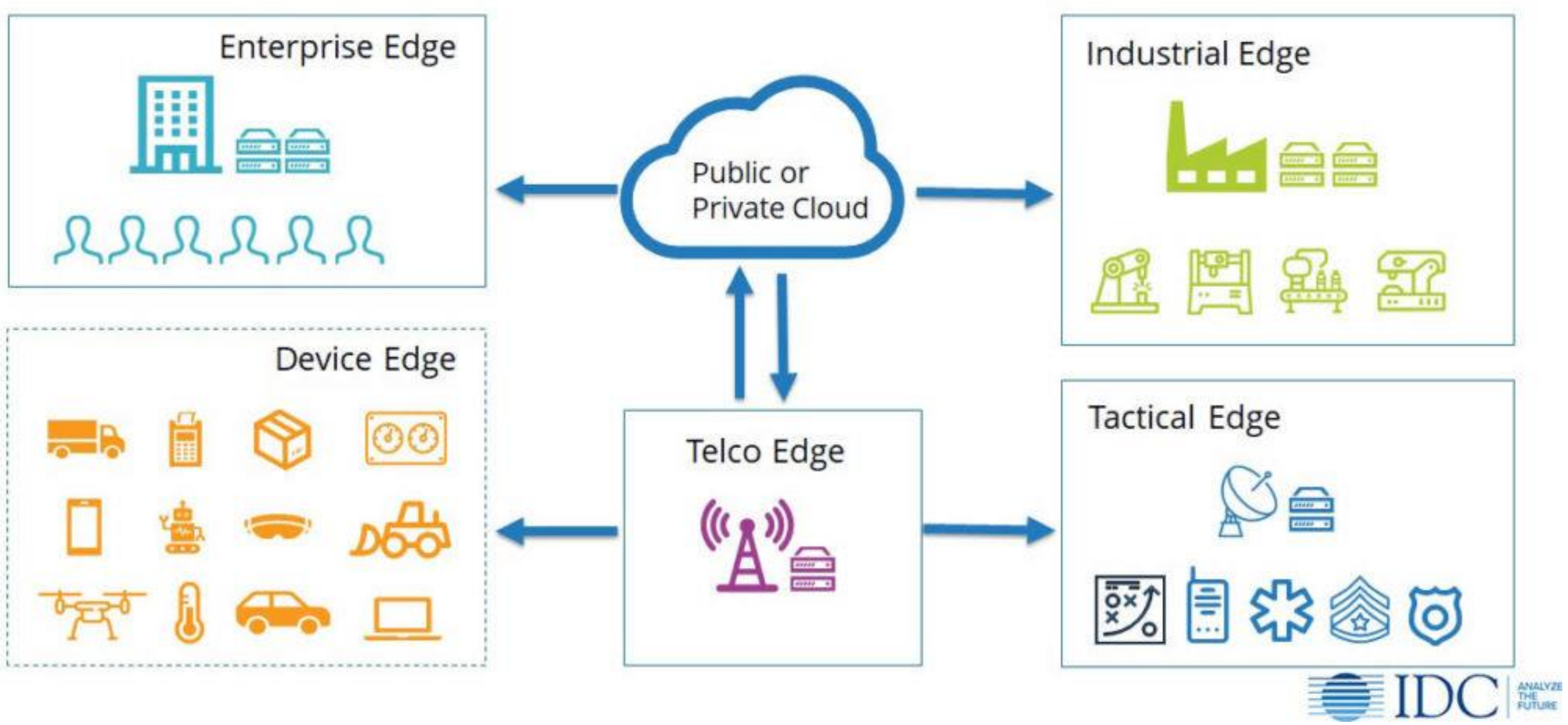
CASTIEL-ETP4HPC workshop

An ETP4HPC White Paper on Edge-Cloud-HPC Continuum Challenges

https://www.etp4hpc.eu/white-papers.html#continuum



Context: IT Investments Shift to the Edge



IDC predictions:

- *In 2022, Enterprise and service provider spending on Edge computing will reach $40 billion in 2022 in Europe, and increase with a five year annual growth rate of 16.4%*
- *By 2023, over 50% of new Enterprise IT infrastructure deployed will be at the Edge rather than corporate datacenters*
- *By 2024, the number of apps at the Edge will increase 800% (compared to 2020)*

Sources: IDC Press release on “IDC Forecasts Double-Digit Growth for European Edge Investments”, January 18, 2022
IDC FutureScape: Worldwide IT Industry 2020 Predictions
IDC blog Edge Computing: Not All Edges are Created Equal (https://blogs.idc.com/2020/06/01/edge-computing-not-all-edges-are-created-equal/)

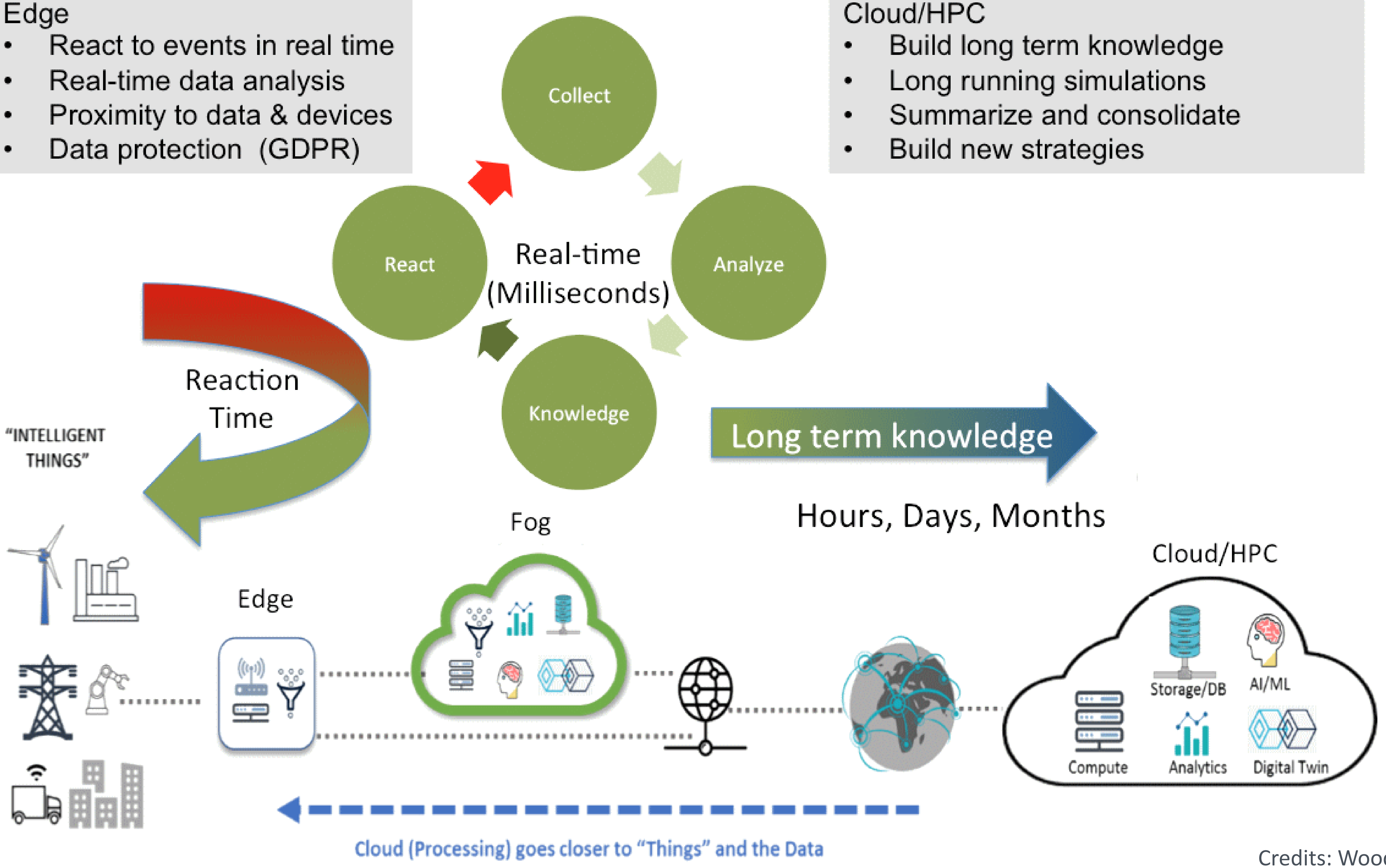
Is This a Directional Evolution?

Supercomputer → Cloud → Edge

The Digital Continuum: a Rather Circular View for Dynamic, Continuous Workflows

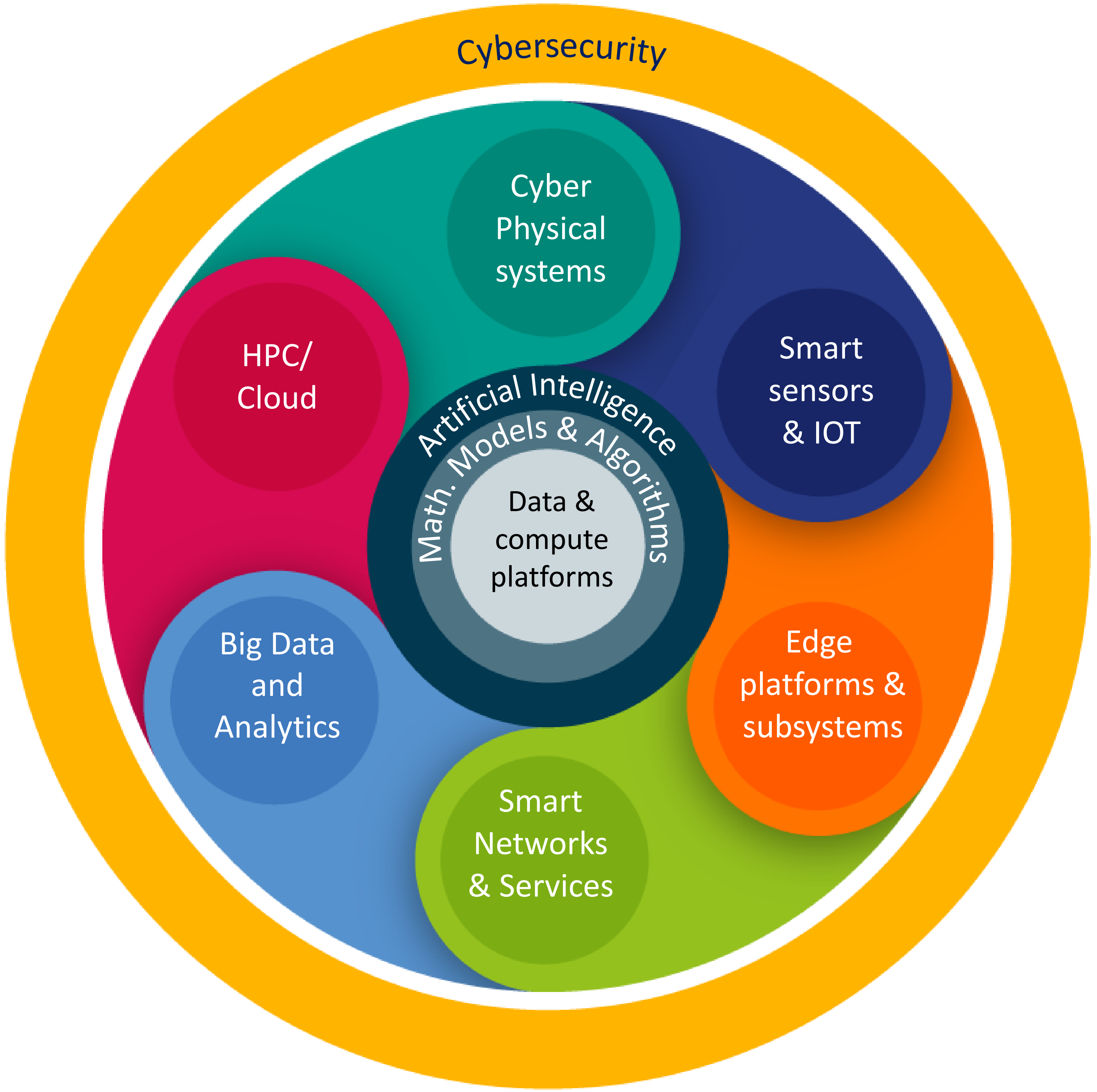
Edge
- React to events in real time
- Real-time data analysis
- Proximity to data & devices
- Data protection (GDPR)

Cloud/HPC
- Build long term knowledge
- Long running simulations
- Summarize and consolidate
- Build new strategies



Credits: Woodside Capital

The Digital Continuum: Cross-area Challenges



Original version courtesy HiPEAC

A continuous dynamic workflow

Between
Smart Sensors and IOT devices
and
HPC / cloud centers
passing through
Edge platforms & subsystems
as well as
Smart Networks and Services
executing
Simulation & Modelling, Big Data Analytics and ML*
based on
Math. Methods & Algorithms incl. MSODE**
pervasively augmented by
Artificial Intelligence
protected and secured by
Cybersecurity
back to
Cyber-Physical Systems,
all based on
Data and compute platforms (hw and sw)

* ML: Machine Learning
** MSODE: Modelling, Simulation and Optimization in Data-rich Environment

TC I
TransContinuum Initiative

Destination Earth Initiative

Goal: create and operate high precision digital models (digital twins) of the Earth to monitor and predict environmental change and human impact

Approach:

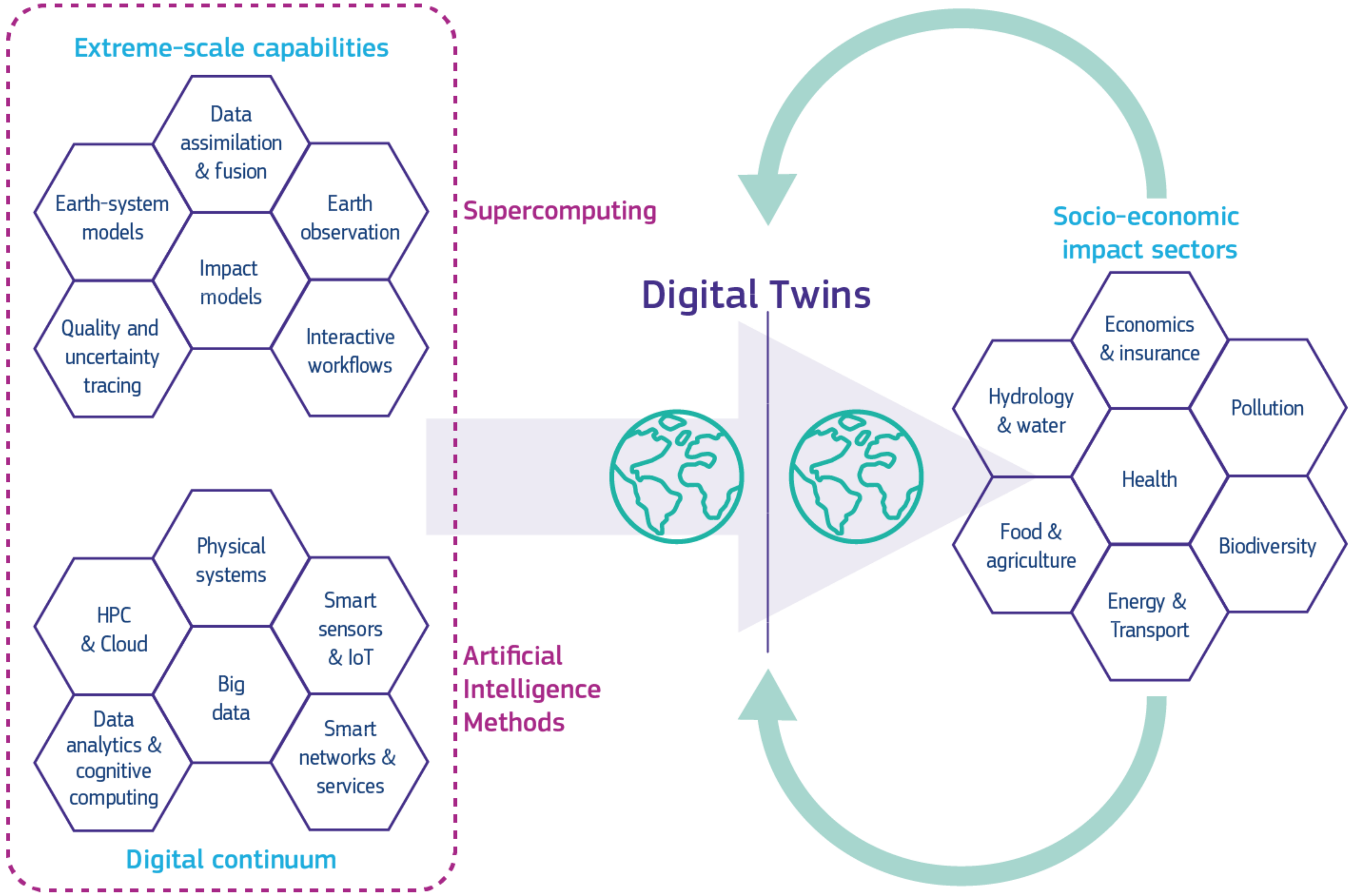
- Simulate atmosphere, oceans/rivers and cryosphere at very high resolution
- Feed in data from a huge number & variety of sensors
- Monitor & model bio- and anthroposphere

Key aspects:

- Anticipate environmental disasters and resultant socio-economic crises to save lives and avoid large economic downturns
- Enable the development and testing of scenarios for ever more sustainable development
- Near-real time analysis of climate and plant observation to correct simulation scenarios
- Integrate simulations with ML/AI-based analysis/forecasting

Need:

- Huge requirements for HPC compute and data capability
- Assimilation of diverse sensors and data streams (satellites through mobile phones)
- Disaster avoidance/recovery needs Urgent Computing capabilities



Source: https://digital-strategy.ec.europa.eu/en/library/destination-earth

TC I
TransContinuum Initiative

A Use Case in Precision Agriculture

Goal: accurately model the greenhouse costs and its production potential

Approach:

- combine models of plant and climate development with actual monitored data processed through machine learning algorithms

Key aspects:

- Near-real time data assimilation of climate and plant observation to correct the simulations scenarios
- Digital twin of the plants executed in the Cloud
- On model drift detection, automatic corrections can require dynamic resource allocations on the Edge or in the Cloud

Need:

- Deployment and the execution of a distributed, cross platform workflow from the edge to the cloud,
- Unpredictable data dynamics triggers on-demand, computationally intensive simulations in the Cloud and related processing at the Edge

A Use Case in Smart Buildings & Cities

Goal: Optimize the energy consumption of buildings

Approach:

- Combine historical operation data with real-time sensor data using ML to predict high-resolution local weather forecasts
- Control building lighting and heating/cooling/ventilation accordingly
- Perform predictive maintenance actions

Key aspects:

- Real-time Edge pre-processing of local sensor data
- Cloud-based analytics
- Further computation-bound simulations may need HPC facilities
- The input data can vary in frequency, relevance, amount

Need:

- Deployment and the execution of a distributed, cross platform workflow from the Edge to Cloud/HPC
- Dynamic reallocations of tasks and resources across the continuum

TCI
TransContinuum Initiative

Where Are We Now?

An archipelago of disconnected solutions

- Separate software stacks (HPC, Big Data on Cloud, Edge Analytics, AI) optimised for different goals, with different infrastructure requirements

What is Difficult?

- Deploy and orchestrate a combination of consistent interoperable components across the full continuum
- The different compute, storage and communication systems of a complex CC installation belong to different owners
  - Authentication
  - Interoperability
  - Heterogeneity
- How to ensure security across the continuum?
- Flexible and efficient operation of CC infrastructures

TC I
TransContinuum Initiative

Building Integrated Software Ecosystems for the Continuum: Challenges

Application-level challenges

- Traditional physics-based simulations (HPC) need to smoothly cooperate with data-driven, learning-based analytics and prediction engines (Cloud)
- Hybrid workflows: programming models, composability

Storing and processing data across the continuum

- **Coping with Extreme Volume:** support the access and processing of “cold”, historical data and “hot”, real-time data + (virtually infinite) simulated data
- **Coping with Extreme Velocity:** unified data processing (in situ/in transit, stream-based) in a common software ecosystem
- **Coping with Extreme Variety:** unified data storage abstractions to enable distributed processing and analytics across the continuum
  - Interoperable data formats
  - "Semantic interoperability” through shared ontologies
  - Storage interfaces should match the needs

Building Integrated Software Ecosystems for the Continuum: Challenges

Managing computation across the continuum

- Dynamic scheduling and orchestration of workflows which evolve at runtime, to optimize performance and energy
- Heterogeneity: wide variety of processors, accelerators, storage devices and systems, and communication systems
- Support seamless deployment and migration of workflow components
- Definition and automatic derivation of performance models

Managing dynamic workflows with ad-hoc load variation

- React to certain events, depending on data contents or on interactive requests
- Dynamically adapt the mapping of the workflow onto the infrastructure
- In some applications (e.g., disasters) parts of the infrastructure suddenly become unavailable
- Requires efficient coupling between Cloud-oriented dynamic orchestrators and traditional batch-based resource management systems, as a step towards more integrated software approaches to dynamic resource management across the continuum

Building Integrated Software Ecosystems for the Continuum: Challenges

AI-related challenges

- New heterogeneity of use cases and hardware
- The deep learning software stacks must be supported (python dependencies handling, containerisation)
- Ad-hoc training and inference runs with tight timing constraints must be supported (urgent and interactive computing)

Cybersecurity challenges

- Federated authentication, authorisation and accounting, monitoring, resource allocations, encryption, user insulation, container certification, etc.
- GDPR-related constraints: HPC centres must provide all tools necessary to address regulatory requirements.

Cooperation challenges

- Interaction of multiple expert communities (HPC, Big Data, AI, cybersecurity, IoT, 5G, etc.).
- Establishing commonly agreed, shared goals and priorities
- Need a common vocabulary and common roadmaps
- This is precisely the core motivation underlying the TransContinuum Initiative (TCI)!

Conclusions

- Edge computing is a major technological and economic trend
- Use cases combining HPC simulation, analytics and AI are emerging
  - They require Edge, Cloud and HPC
  - Adopt and evolve the “Digital Twin” approach
  - Targeting scientific, societal and business benefits
- An integrated SW ecosystem spanning across Edge, Cloud and HPC systems is key for sustained success – and it is evolving
  - Programming environment(s) to develop applications combining simulation, AI/ML and data analytics
  - Data storage, transfer, processing, assimilation across the continuum
  - Manage complex computations on large distributed, heterogeneous infrastructures
  - Efficient handling of dynamic, distributed workflows
  - Cybersecurity mechanisms to protect infrastructure & data
  - Requires cooperation across multiple areas

TCI

TransContinuum Initiative

Thank you !