Cloud Computing

Authors
Dawa L.A. Ometto (Utrecht University), **Giuseppe Gianquitto** (SURF)

1. Growing investments in the European cloud infrastructure & sovereignty
2. More AI-driven hardware in data centres to manage workloads in the cloud
3. Bringing computing resources closer to the ‘edge’
4. Growing need for automated cloud and edge resource management
5. Towards cloud native services and application design for research

Introduction

At the core of the digital infrastructure are two distinct, yet complementary, distributed computing frameworks. On one hand, cloud computing – the core focus of this chapter - provides vast, on-demand access to scalable processing power and data storage via the network, through private and public cloud providers. Cloud computing allows organisations and individuals to pay on demand instead of investing in their own costly infrastructure. On the other hand, edge computing pushes computing functions closer to where local data is generated – at the network's edge. Edge computers enable real-time responsiveness to the local data, reduce network traffic, maintain functionality during disconnections from central systems, and ensure that sensitive data can remain on the local device. Edge computing has developed rapidly due to the explosive growth of the IoT-devices (Internet of Things) loaded with large amounts and large varieties of data-producing sensors in e.g. cars, healthcare devices, public infrastructure, and manufacturing machines. Once viewed as separate approaches, these two frameworks are increasingly converging into a unified framework known as the computing continuum. This continuum seamlessly connects the smallest sensor in a peripheral device to a supercomputer in a central data centre, facilitating a variety of data processing capabilities.

The primary catalyst driving the computing continuum is the relentless advance of artificial intelligence (AI). For optimal operation of today's AI systems and the demands, there are two principal requirements. First, it needs massive, centralised high-performance computing

(HPC) resources for AI model training and large-scale and complex data analysis. Second, to ensure low latency during operation, decentralised processing is necessary to run these AI models, or their compact versions, in real-world applications, including, for example, synchronising with the cloud for further big data analysis.

The current cloud computing landscape is significantly influenced by strategic global competition. Hyperscalers (large technology companies that provide cloud computing services on a massive scale) are at the forefront of this dynamic. Hyperscalers such as Amazon Web Services (AWS), Google Cloud, Oracle Cloud Infrastructure (OCI), and Microsoft Azure deliver powerful services and have highly dominant market positions in regions like the US and Europe.

Countermeasures in Europe are evolving to strengthen Europe's digital sovereignty in the cloud sector. In 2025, the European Commission proposes the Cloud and AI Development Act, intending to at least triple the EU's data centre capacity within the next 5 to 7 years. These kinds of (policy) developments will lead to expanded European cloud infrastructures – both private and public – to meet the needs of EU businesses and public administrations who favour sovereign cloud services including European data residency. However, European-level collaboration is paramount for this ambition whereby initiatives like Gaia-X and European Open Science Cloud (EOSC) offer lessons learned.

Contributors
Alex Chiumento (University of Twente),
Javier Ferreira Gonzalez (Saxion University of Applied Sciences)

TREND #1

Growing investments in the European cloud infrastructure & sovereignty

Public Values

Autonomy	Freedom of choice | Independence | Privacy
Justice	Transparency | Sustainability | Accountability
Humanity	Safety

Maturity WATCH PLAN **ACT**

Drivers

Geopolitics & (digital) sovereignty; Compliance & regulation; Cybersecurity & trust; Automation & AI; Digital transformation; Global trade & tariffs

GenAI has influenced the cloud market considerably: half of the increase in global (public) cloud service revenues since the launch of ChatGPT (late 2022) is driven by GenAI. In Europe specifically, GenAI-specific services such as GPU-as-a-Service (GPUaaS) and GenAI Platform-as-a-Service (PaaS) are demonstrating exceptionally high growth rates over the last few years. The main beneficiaries of this European market growth have been the US companies Amazon, Microsoft, and Google, which together now account for 70% of the regional market in terms of revenues. Europe is experiencing a strong and ongoing push for digital sovereignty, aiming to reduce reliance on US-based cloud providers and ensure greater control over sensitive data. This is the reason why major global cloud providers are making significant European investments to address digital sovereignty concerns. As a result, European cloud providers' market share remains relatively constant at around 15% of the European cloud market despite the European ambitions and investments.

SIGNALS

Cloud integration in academia

Hyperscalers, like AWS and Microsoft, launch dedicated EU sovereign cloud offerings to address data residency and regulatory requirements. European Open Science Cloud (EOSC) and Gaia-X provide federated environments aligned with EU values.

AWS Unveils Independent European Governance and Operations for European Sovereign Cloud (infoq.com)

Announcing comprehensive sovereign solutions empowering European organisations (blogs.microsoft.com)

GÉANT Innovation Programme funds a Digital Research Environment (DRE) initiative to create integrated, cloud-based research platforms (connect.geant.org)

European Open Science Cloud (EOSC) develops a federated, trusted environment enabling researchers across Europe to share, access, and reuse data and services for open science (research-and-innovation.ec.europa.eu)

Size of the cloud market

Cloud market jumped to $330 billion in 2024 – GenAI is now driving half of the growth (srgresearch.com)

European cloud providers' local market share now holds steady at 15% (srgresearch.com)

The future of cloud in 2029: the journey from technology to business necessity (gartner.com)

2024 enterprise trends: cloud meets AI (redhat.com)

Cloud market share 2024 - AWS, Azure, GCP growth fueled by AI (holori.com)

Global cloud infrastructure market share 2024 (statista.com)

Amazon and Microsoft Stay Ahead in Global Cloud Market (statista.com)

AWS Remains $330bn Cloud Market Leader, Driven by AI Growth (technologymagazine.com)

Cloud Market Jumped to $330 billion in 2024 (srgresearch.com)

European cloud providers grow but lose market share to US titans (techrepublic.com)

Europe's cloud customers eyeing exit from US hyperscalers (theregister.com)

SIGNALS

European cloud infrastructure expansion

The development of a national Gaia-X testbed in the Netherlands continues, expanding the ecosystem of trusted, sovereign cloud options

Test environment for cloud services: Structura-X (tno.nl)

European cloud computing

- **Platforms** (european-alternatives.eu)
- **Policies** (digital-strategy.ec.europa.eu)
- **The cloud sovereignty nexus: How the European Union seeks to reverse strategic dependencies** (onlinelibrary.wiley.com)

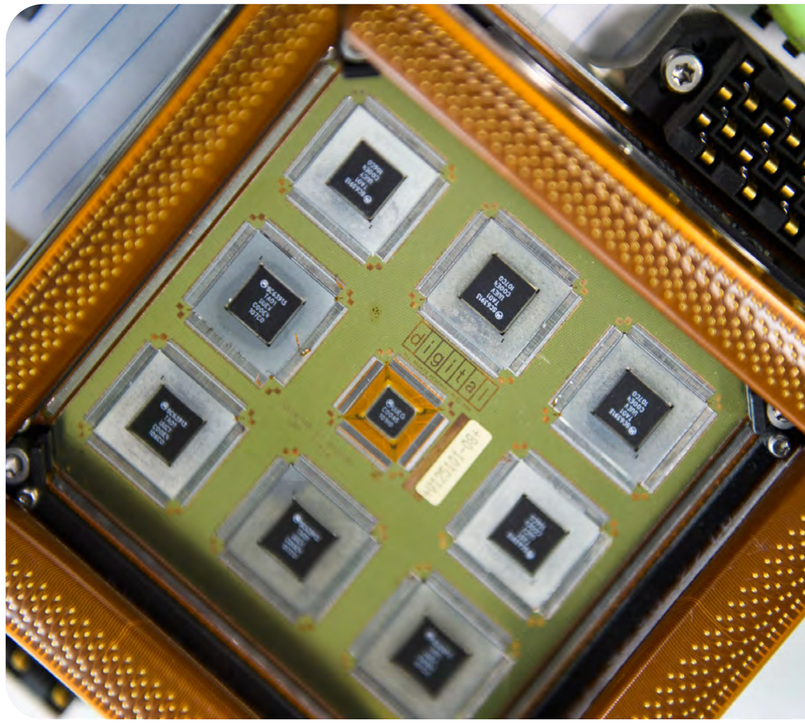
Digital sovereignty in Europe: navigating the challenges of the digital era (pppescp.com)

Designing the European processor (sipearl.com)

> ***"By 2030, 75% of European businesses should use cloud-edge technologies for their activities."***
>
> **– European Commission**

Next Euro HPC chip coming next year will be in 2026 EU exascale system (hpcwire.com)

SiPearl raises €90 million in initial closing of Series A to launch Rhea (eib.org)



IMPACT

Education

- More institutions are adopting cloud-first policies, offering students much-needed exposure to and experience with real-world infrastructure.
- A diverse training portfolio is required for students in cloud computing and high-performance computing (HPC) to meet the rising demand for scalable AI infrastructure and navigate European digital sovereignty initiatives.

Research

- Access to powerful cloud resources on demand will accelerate the timeframe towards key research discoveries and outcomes.
- Researchers require enhanced support and training to maximise their use of cloud resources through initiatives. The hyperscalers provide tailored programmes and expert assistance, for example, AWS's Open Data and Azure for Research.

Operations

- Data centres are being upgraded and federated clouds integrated to meet increasing compute demands.
- Technical support is essential to ensure that facilities on institutional campuses operate without interruption, providing reliable maintenance and troubleshooting for high-performance computing (HPC) and cloud infrastructure.

TREND #2

More AI-driven hardware in data centres to manage workloads in the cloud

Public Values

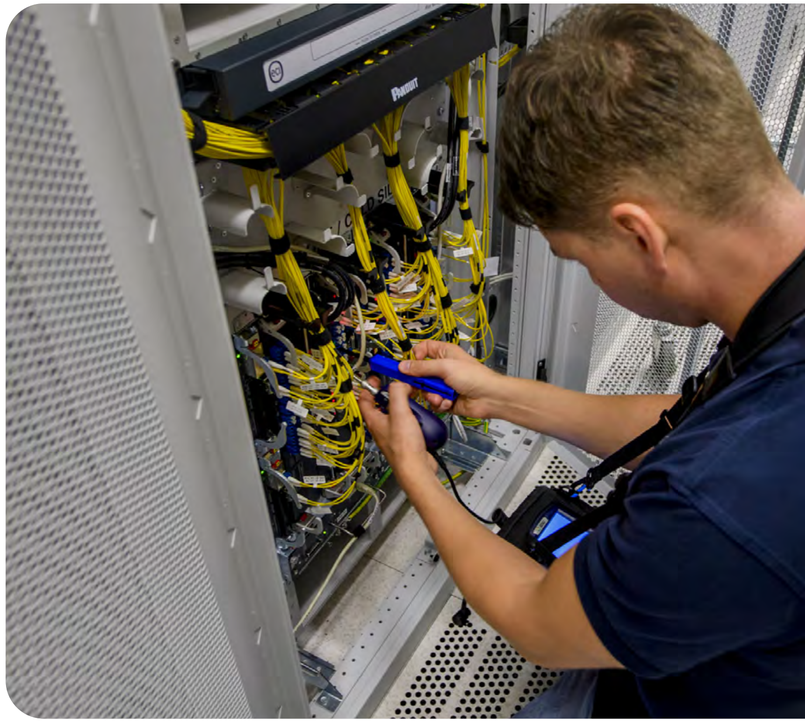
Autonomy	Independence
Justice	Sustainability | Equity | Integrity | Accountability
Humanity	Safety

Maturity WATCH PLAN **ACT**

Drivers

Automation & AI; Engineering advances & computation; Digital transformation; Connectivity & interaction; Energy supply & demand; Service-oriented & value-based economies

To support intensive workloads that are anticipated for cloud computing, cloud providers and heavy users like research institutions are investing heavily in specialised hardware. This goes beyond standard CPUs (with AMD outselling Intel in the data centre segment recently) to include powerful accelerators such as Graphical Processing Units (GPUs), which are the workhorses for AI training and whereby Nvidia's data centre GPU market share is 98%. Hyperscalers are developing specific hardware like AI chips to deliver high-performance, cost-efficient cloud services. This hardware evolution requires additional innovations in infrastructure to manage the immense power density.



SIGNALS

GPU and supercomputer upgrades

LUMI supercomputer in Finland delivers up to 380 petaflops using green energy (csc.fi)

Nvidia's new Spectrum-XGS aims to turn multiple data centers into 'one gigantic GPU' (datacenterdynamics.com)

> ***"NVIDIA's Blackwell Ultra GPUs, along with custom accelerators from Google and Amazon, drove the data center IT component market to 44% year-over-year growth in 2Q 2025."***
>
> **– Dell'Oro Group report**

Custom AI hardware and efficiency

AWS Trainium, Google TPUs, and Microsoft FPGA-based services offer cost-effective alternatives (cloudexpat.com)

AI chips are driving cloud innovation through better performance-per-dollar metrics (amd.com)

Energy-aware operation of HPC systems in Germany (frontiersin.org)

Harnessing data centre heat into reusable energy: a sustainability game changer (hpcwire.com)

IMPACT

Education

Students need to gain exposure to advanced computing environments and AI-capable hardware. This can be achieved through dedicated compulsory and elective modules/ courses, as well as project work, where the utilisation of new computing paradigms is a requirement.

Research

Thanks to the integration of advanced AI model training and cloud computing services, researchers will have faster and more feasible access to computing resources that are applicable across several disciplines.

Operations

At the research institute and university level, shared infrastructure serves to reduce costs. In addition, the deployment of liquid cooling systems for computational infrastructure will improve energy efficiency and help with the sustainability of these computational resources.

TREND #3

Bringing computing resources closer to the ‘edge’

Public Values

Autonomy	Independence | Privacy | Freedom of choice
Justice	Accountability | Integrity | Inclusion
Humanity	Safety

Maturity WATCH **PLAN** ACT

Drivers

Connectivity & interaction; Engineering advances & computation; Cybersecurity & trust; Compliance & regulation; Digital transformation

The rise of AI, IoT (Internet of Things), and future 6G networks is pushing computation away from a purely centralised model. This architectural shift involves moving data processing to the network edge to reduce latency, conserve bandwidth, and protect privacy. In practice, this means AI models need to run on-site on specialised hardware for real-time edge computing tasks such as field robotics (where a high degree of autonomy is required) or patient monitoring.

Concurrently, society is increasingly relying on and expecting complex, multimodal workflows that combine text, images, and sensor data. These workflows demand new, distributed architectures that seamlessly integrate edge devices for data collection and processing, cloud platforms for large-scale storage, and HPC backends for intensive AI model training and advanced data analysis.

SIGNALS

Edge for low latency AI

Implementing NVIDIA Jetson in field robotics and smart medical devices (industrialautomation.nl)

The FNS programme (6G Future Network Services) has developed AI-native 6G network to enable dynamic edge-cloud orchestration (tudelft.nl)

EC's next generation Internet of Things (digital-strategy.ec.europa.eu)

"Cloud-edge integration allows organisations like enterprises and research institutes to enhance their IT infrastructures with low-latency processing and real-time decision-making capabilities."

– EU Commission/Technopolis

Multimodal AI pipelines

AI4EU has fostered distributed AI services and federated learning platforms (ai4europe.eu)

Edge AI is being used for on-device inference, with cloud/HPC backends then employed for AI model training and coordination (sciencedirect.com)

Empowering edge intelligence: a comprehensive survey on on-device AI models (dl.acm.org)

Study on the economic potential of far edge computing in the future smart Internet of Things (op.europa.eu)

The global edge computing market size was estimated at USD 23.65 billion in 2024 and is expected to reach USD 327.79 billion in 2033, growing at a CAGR of 33.0% from 2025 to 2033 (grandviewresearch.com)

Hype Cycle for Edge Computing, 2024 (gartner.com)

The European cloud, edge and IoT continuum preliminary market analysis (zenodo.org)

IMPACT

Education

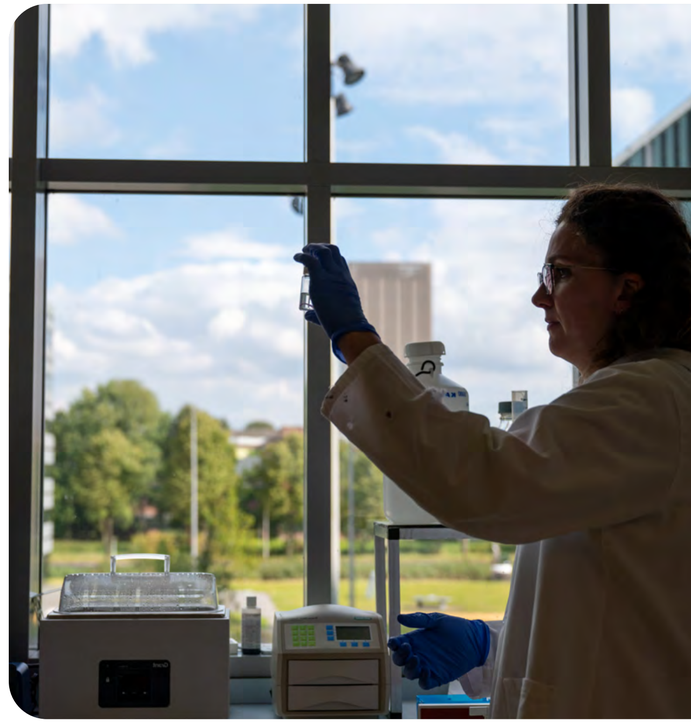
Students and staff alike need to be exposed to edge AI and network-driven architectures. This can be achieved through revision of existing modules/courses or the creation of focused content.

Research

The architectural shift will enable new experimental approaches in robotics, medicine, and environmental monitoring.

Operations

Architecture shifts are set to drive investment in IoT, network infrastructure, and local compute.



TREND #4

Growing need for automated cloud and edge resource management

Public Values

Autonomy	Privacy | Independence
Justice	Sustainability | Accountability | Equity | Integrity | Transparency
Humanity	Safety

Maturity WATCH **PLAN** ACT

Drivers

Automation & AI; Engineering advances & computation; Digital transformation; Service-oriented & value-based economies; Digital transformation; Energy supply & demand

The proliferation of AI is creating new paradigms in cloud resource management, as AI-driven automation (AIOps) becomes essential. Cloud providers are now embedding AI assistants into their platforms to automate complex tasks, such as resource provisioning - selecting and deploying computing resources for optimal application performance - and performance tuning.

In addition, the massive energy footprint of AI workloads is accelerating the push towards green computing, with a focus on sustainable data centres and intelligent workload scheduling. To manage the unpredictable costs, institutions are adopting FinOps (Financial Operations), which is the practice of financial governance that unites IT, finance, and research faculties to optimise cloud spending and maximise value.

SIGNALS

Cloud automation (AIOps)

AWS and Azure have introduced AI assistants for infrastructure usage and monitoring:

AWS AI assistant for accelerating software development and leveraging companies' internal data (aws.amazon.com)

Operate everywhere with Azure AI management and security (techcommunity.microsoft.com)

> ***"FinOps is a framework and cultural practice which maximises the business value of cloud and technology, enables timely data-driven decision making, and creates financial accountability through collaboration between engineering, finance, and business teams."***
>
> **– FinOps Foundation**

FinOps practices are being adopted to curb cloud budget overruns and track usage

FinOps Foundation (finops.org)

Cloud Strategy at the EU: Cloud computing (digital-strategy.ec.europa.eu)

Embracing cloud governance within the context of FinOps (ijfmr.com)

IMPACT

Education

As students engage with AI-driven cloud and DevOps tools, they will need focused skills in AIOps, FinOps, and sustainable computing to learn how to manage automation, costs, and environmental impact effectively.

Research

The rise of AI-driven cloud management and DevOps practices enables researchers to accelerate experimentation and scale workloads efficiently, but also demands greater focus on cost management, sustainability, and cross-disciplinary collaboration.

Operations

- These new paradigms and practices are helping institutions to manage IT complexity and to meet sustainability goals.
- Although the integration of AI-driven cloud tools and DevOps practices enhances efficiency and scalability, new strategies for financial governance, sustainability, and cross-departmental coordination may need to be considered.

TREND #5

Towards cloud native services and application design for research

Public Values

Autonomy	Freedom of choice | Independence
Justice	Sustainability | Integrity, Transparency | Accountability
Humanity	

Maturity WATCH PLAN ACT

Drivers

Automation & AI; Connectivity & interaction; Digital transformation; Service-oriented & value-based economies; Value of knowledge & skills

The adoption of cloud technologies is transforming how research software is developed and used. Research IT is shifting towards cloud-native development, where applications are built as scalable microservices and consumed via application programming interfaces (APIs).

This approach, which includes containerising workflows with tools like Docker and orchestrating them with Kubernetes, is making academic software more robust and reproducible.

Practices like Infrastructure as Code (IaC) and Continuous Integration/Continuous Delivery (CI/CD) pipelines are becoming the standard, thus allowing scientists to share the exact configurations used for different experiments and support reproducibility. This modernisation not only accelerates scientific discovery but also improves the longevity and reusability of research software, aligning it with industry-grade practices.

SIGNALS

Cloud-native design and tools

Kubernetes, Terraform, and GitLab CI/CD are being used in academic and research projects:

CERN: processing petabytes of data more efficiently with kubernetes (kubernetes.io)

A grid site reimagined: building a fully cloud-native ATLAS Tier 2 on kubernetes (inspirehep.net)

The use of infrastructure as code in regulated companies (ispe.org)

> ***"Infrastructure as Code provides several advantages toward reproducibility in computer science. ... IaC allows researchers to define and control their infrastructure accurately in a format that can be easily stored, versioned, and shared, making it easy to reproduce experiments and obtain the same results at each execution."***
>
> **- Frontiers in Computer Science**

IMPACT

Education

Cloud technologies in research software enhance education by providing students access to robust, reproducible tools that promote collaboration and innovation; however, students can maximise the potential by gaining hands-on skills in DevOps and cloud-native development.

Research

The adoption of cloud-native development supporting FAIR software practices profoundly enhances the efficiency and reproducibility of research, enabling scientists to scale experiments and accelerate discoveries and share methodologies more effectively.

Operations

The transition to cloud technologies and microservices in research software enhances cross-disciplinary collaboration and optimises IT service delivery, while also streamlining resource allocation and reducing software deployment time. However, a critical shortage of skills is hampering adoption.

More info about Cloud Computing? Visit surf.nl

SURF Utrecht
Hoog Overborch Office
Building (Hoog Catharijne)
Moreelsepark 48
3511 EP Utrecht
+31 88 787 30 00

SURF Amsterdam
Science Park 140
1098 XG Amsterdam
+31 88 787 30 00

futuring@surf.nl
www.surf.nl/en