SUSTAINABLE CYBERINFRASTRUCTURES

WHITE PAPER

Mark A. Asch
University of Picardy,
Amiens, FRANCE.
EXDCI2 Project
mark.asch@u-picardie.fr

Thierry Bidot
Neovia-Innovation
Paris, FRANCE
EXDCI2 Project
thierry.bidot@neovia-innovation.eu

François Bodin
University of Rennes
Rennes, FRANCE
EXDCI2 Project
francois.bodin@irisa.fr

Maike Gilliot
ETP4HPC/TERATEC
Bruyères-le-Chatel, FRANCE
EXDCI2 Project
maike.gilliot@teratec.fr

Laurent Morin
INRIA
Rennes, FRANCE
laurent.morin@inria.fr

Recent hardware and software advances have motivated the development of a transcontinuum digital infrastructures concept to account for the convergence of data and compute capabilities *across* the digital continuum, which goes from instruments and sensors at the edge, through wireless and network channels, to HPC and cloud facilities et the center. This concept is not in a straight line from the past efforts and a paradigm change is needed: we will have to design systems encompassing hundreds of billions of cores distributed over scientific instruments, IoT, supercomputers and Cloud systems through Lan, Wlan and 5G networks.

Pushed by massive deployments of compute and storage capabilities at the edge, we require new system design to accommodate the ecosystem change to be expected in the coming decades (environmental and technological) and horizontally integrate the different actors. The new demands and challenges that combine data and compute, distributed across the continuum, and the maintenance and resource efficiencies, are pushing for drastically increased *software and hardware sustainability*. Furthermore, the need to provide high-level cybersecurity is profoundly changing the game. Efficiency and resilience will have to reach levels never achieved so far, while taking into account the intrinsic distributed and heterogeneous nature of the continuum. In addition, the question of dealing with such high volumes of data needs to be faced, and quality versus quantity will become unavoidable. These considerations will spread over all components. Long-lifetime hardware devices will have to be reconfigurable, modular, and self-aware in order to be operational over extended periods. Algorithm efficiency will need to be drastically pushed up (e.g. more efficient AI). Management and deployment of large-scale application workflows will have to be adapted, or invented. Network protocols will have to offer better control over the data logistics, etc.

It is widely recognized that AI will play a central role in these extreme-scale, continuum infrastructures. This will occur at three levels: (1) AI for Digital Infrastructure, (2) Digital Infrastructure for AI, and (3) AI for Science, Industry and Societal Challenges. The first addresses how AI can pilot and monitor the continuum and in so doing provide solutions to the points listed in the previous paragraph. The second treats the question of re-designing the cyberinfrastructure to efficiently deal with data analysis and machine learning, which means tuning of data access, I/O, and low precision arithmetic. The last deals with the ever-increasing needs to exploit AI techniques for extreme-scale, combining Data and Compute through the interpretation and coupling of computing results, measurements and observations (eg. Digital Twins in extreme earth modelling, combining climate models with satellite data and on-ground sensors).

The overall objective is to target high TRL solutions (7 and more), based on horizontal synergies between all the concerned digital infrastructure technologies: HPC, Big Data, Machine Learning, IoT, 5G, cybersecurity, processor technology and robotics. All of these components of the digital infrastructure will together be able to address the huge societal challenges and sustainable development goals by mobilizing their amazing potential all the way across the continuum.

References

[1] Asch, Moore, et al. Big data and extreme-scale computing: Pathways to Convergence-Toward a shaping strategy for a future software and data ecosystem for scientific inquiry. *Int. J. HPC and Applications*, Vol. 3, No. 4, pages 435–479. 2018.

[2] ETP4HPC. *Strategic Research Agenda For High-Performance Computing in Europe, January 2020 - European HPC Research Priorities 2021 – 2024.* Edited and published by ETP4HPC (European Technology Platform for High Performance Computing) 2020. `https://www.etp4hpc.eu/sra.html`