HPC for Urgent Decision-Making

Exploiting the full range of HPC capabilities to promptly gain insights and support decisions in a dynamic data-driven environment

White Paper

Manolis Marazakis (FORTH),
Marc Duranton (CEA),
Dirk Pleiter (KTH),
Giuliano Taffoni (INAF),
Hans-Christian Hoppe (Scapos AG)

15/02/2022

etp4hpc.eu

@etp4hpc

EUROPEAN TECHNOLOGY PLATFORM FOR HIGH PERFORMANCE COMPUTING

Table of Contents

Introduction

Emerging use cases from incident response planning and broad-scope European initiatives (e.g. Destination Earth [1,2], European Green Deal and Digital Package [21]) are expected to require federated, distributed infrastructures combining computing and data platforms. These will provide elasticity enabling users to build applications and integrate data for thematic specialisation and decision support, within ever shortening response time windows.

For prompt and, in particular, for urgent decision support, the conventional usage modes of HPC centres is not adequate: these rely on relatively long-term arrangements for time-scheduled exclusive use of HPC resources, and enforce well-established yet time-consuming policies for granting access. In urgent decision support scenarios, managers or members of incident response teams must initiate processing and control the resources required based on their real-time judgement on how a complex situation evolves over time. This circle of clients is distinct from the regular users of HPC centres, and they must interact with HPC workflows on-demand and in real-time, while engaging significant HPC and data processing resources in or across HPC centres.

This white paper considers the technical implications of supporting urgent decisions through establishing flexible usage modes for computing, analytics and AI/ML-based applications using HPC and large, dynamic assets.

The target decision support use cases will involve ensembles of jobs, data-staging to support workflows, and interactions with services/facilities external to HPC systems/centres. Our analysis identifies the need for flexible and interactive access to HPC resources, particularly in the context of dynamic workflows processing large datasets. This poses several technical and organisational challenges: short-notice secure access to HPC and data resources, dynamic resource allocation and scheduling, coordination of resource managers, support for data-intensive workflow (including data staging on node-local storage), preemption of already running workloads and interactive steering of simulations. Federation of services and resources across multiple sites will help to increase availability, provide elasticity for time-varying resource needs and enable leverage of data locality.

Key insights

- The need to give prompt access to users and entities for time-sensitive decision support in response to unexpected events does not fit the common access/use policies of HPC centres well. Moreover, such users rely on HPC workflows that are part of data-driven processes and might not have a long-term and direct connections to supercomputing centres.
- Use cases for insight extraction and decision support in short time-scales require scheduling of resources in reaction to urgent demands, and is complicated by high dynamicity and diversity in the units of work to be managed; these include ensembles of jobs, data-staging to support workflows, as well as interactions with services or facilities external to the HPC systems/centres.
- Model-driven simulation and real-time interaction with decision makers requires interactive-style access to HPC resources, in the context of dynamic workflows processing large datasets. This leads to technical and organisational challenges. Pragmatically, we could consider maintaining both classical HPC clusters for batch-style use of computational and storage resources, together with clusters configured for dynamic workflows and interactive usage modes. For urgent decision making (such as evacuation and rescue operations planning coordination), the HPC resource requirements can be quite substantial; therefore, a static partitioning might turn out to be ineffective.
- Specific technical challenges arise in several areas: short-notice secure access to federated HPC and Cloud resources, dynamic resource allocation and scheduling (including large-scale use of accelerators), coordination of resource managers (including elastic resource allocation, unified authentication and authorisation, and usage accounting/monitoring), support for data-intensive workflows plus data staging on node-local storage, and increased interactivity (including near real-time steering of simulations and dynamic preemption of already running jobs as needed to free-up resources).
- The current state-of-the-art at supercomputer centres fails to adequately support the objective of providing prompt decision support.

Key recommendations

- Advance the state-of-the-art in short-notice secure access to federated HPC resources, system software and management to address challenges in resource allocation and scheduling, data-intensive workflow support (including data staging on node-local storage), and increased interactivity (including near real-time computational steering and preemption of jobs).
- Introduce enhancements and extensions to the relevant HPC programming models and APIs which will support expressing and managing complex and dynamic units of work, including ensembles of jobs, data-staging to support workflows, and interactions with services/facilities external to the HPC system.
- Initiate research and innovation actions, in a timeframe of 3-4 years, towards enabling interactive-style usage of HPC resources for near real-time workflows in support of insight extraction and decision making.
- Use Digital Twins to mirror various facets of complex events or systems, and develop predictive models to predetermine which line of action would be the best to take when a similar situation is evolving in real-time. The Digital Twins would run HPC simulations without the need for fast, interactive turnaround (i.e. offline operation).
- Continuously evaluate and improve reference system architectures and interoperability standards to support the effective operation of HPC environments for urgent decision making.

This white paper considers the implications of flexible and convergent usage modes for data assets and HPC infrastructure to support modelling/simulation software, as well as analytics and AI/ML-based applications. Emerging use cases from broad-scope initiatives (e.g. Destination Earth, European Green Deal and Digital Package) are expected to be built on top of federated platforms with some degree of resource elasticity and supporting the combined use of data and infrastructure, where users can build applications and integrate their data for thematic specialisation for the purposes of decision support in short timescales, including urgent response to evolving complex situations.

Figure 1 illustrates an indicative flow for decision support workflows, relying on model-based simulation and data analysis and expected to be triggered either as part of regular proactive preparation for incident response or as an urgent reactive response to unexpected events. In this white paper, we focus on decision-making support based on model execution and data processing in latency-sensitive response to evolving situations. We assume that model development and the evolution of the corresponding codebase are carried out offline, although with interactions with the data and code assets on the decision support environment. The assessment of model quality and applicability of results is also assumed to be carried out offline. Likewise, the decision support interfaces per-se (including data processing for exploration and visualisation) are outside the scope of this white paper. We focus on the requirements regarding the access to HPC and data processing resources, particularly HPC platform access, data access and workload management in latency-sensitive use cases.

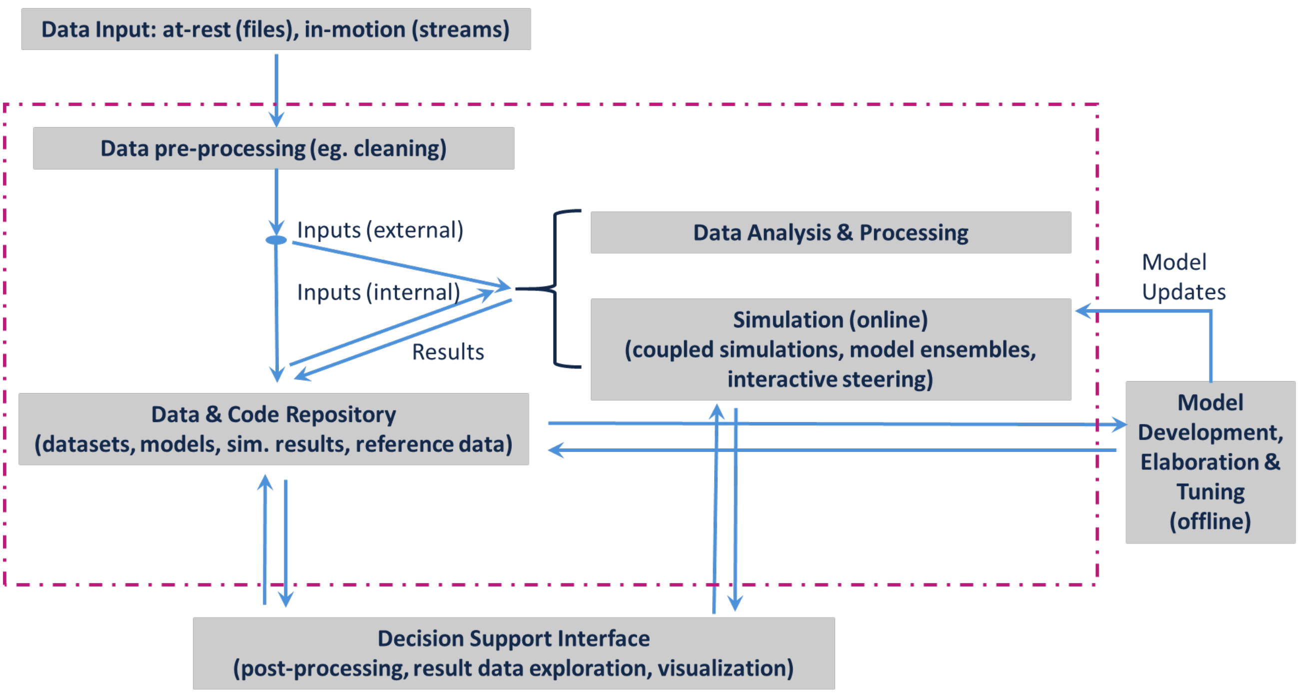


Figure 1: Decision support flow based on model-based simulation and data analysis. This white paper focuses on model execution and data processing in latency-sensitive response to evolving situations. Model development and the evolution of the corresponding codebase are assumed to be carried out offline, although with interactions with the data and code assets on the decision support environment.

We mention in passing that this type of decision support flow would also facilitate data-driven workflows in several scientific disciplines (e.g. Earth Observation, Astrophysics, Biomedicine), with a strong focus on exploratory data processing [19]. However, such scientific workflows usually lack the urgency of having to respond within hard deadlines, with unforeseeable start times and with the involvement of non-registered decision makers.

1. Context and Use Cases

This section considers requirements coming from a set of use cases that illustrate the expectations in terms of urgent computing from existing technologies [15]. The overall objective of the Destination Earth Initiative (DestinE) is to develop a "Digital Twin of the Earth" [1] - i.e. a digital modelling platform to "visualise, monitor and forecast natural and human activity on the planet in support of sustainable development". Accessible and interoperable data, combined with digital infrastructure and AI solutions, facilitate evidence-based decisions and expand the capacity to understand and tackle environmental challenges.

In the DestinE initiative [1], use cases that require urgent reactions are related to **disasters like earthquakes, tsunamis, forest fires, storms**. However other important events must be monitored such as floods, radiological incidents, epidemic outbreaks, but also astrophysical events ("space weather") as particle eruptions (flares) emanating from the Sun and impacting the Earth's magnetosphere with potentially dire disruptions of satellite operations [9] or telecommunications including the internet.

Such extreme events require a rapid response, to prevent or at least minimise human injuries or loss of life, and damage to infrastructure or property. In the worst case, the response would be in the form of data-driven management of disaster relief. The decision makers (persons or organisations) will require accurate information as quickly as possible to choose the best course of action. As clearly stated in the DestinE announcement [1], there are different classes of stakeholders: civil protection agencies, scientists and expert users, public authorities and humanitarian aid organisations. As a further example from the DestinE set of use cases, extreme weather constitutes "the second most likely global risk to the economy" and failure of climate change mitigation and adaptation is seen as "the highest global risk in terms of impact on the economy". Thus, DestinE use case #1 - "Accelerating weather-related disaster risk management" focuses on the urgent need to "improve disaster risk management with timely information on exposure of human settlements and fine scale hazard and risk information accelerating weather-related disaster risk management".

At the same time, there exists a large number of use cases that do not require strictly "urgent" reactions but rather either **rapid analysis or "medium-term" predictable responses** [9,10,11] for ever increasing (spatial and temporal) resolution. As an example, for Climate Modelling scientists, public authorities and private sector entities continuously simulate advanced and well-validated high-resolution Earth System Models (ESMs), which are the primary tools for making future projections of global climate change. An ESM is a coupled climate model that also explicitly models the movement of carbon through the entire earth system, with consideration of physical, chemical and biological processes [6]. They are linking such projected changes to allowable carbon emissions commensurate with staying below a given warming target. However, they also guide short-term and mid-term actions affecting agriculture, forest environment, and day water management, from experts, policy officers, local and global authorities. A particularly acute challenge is the sheer volume of the data resulting from weather and climate simulations with a high spatial resolution [7].

Advances in technology over the past years have opened up **many new opportunities in aiding urgent decision-makers**. In the Big Data domain, a huge amount of data is collected from numerous sources, like satellites, instruments, IoT sensors, and even social media sources such as Twitter [12]. There is a tremendous wealth of information in all this data if it could be mined, aggregated, processed, analysed, visualised and compared with numerical simulations [13]. The use of data (both observed and simulated) is crucial in a prompt decision-making environment. In particular, there are substantial benefits in facilitating the creation and access to data that are Findable, Accessible, Interoperable and Reusable without or with minimal human intervention (FAIR principles [3]), giving computational systems the ability to find, access, process, and reuse data. Moreover, following the Digital Twin approach, forecasting models could be continuously calibrated and validated.

Data comes from primary sources (e.g. physical measurement streams) and from processing pipelines (e.g. output of filtering steps or models) that may include extensive simulations, or ensembles of interrelated simulations. Incident response scenarios introduce additional constraints due to the use (and generation) of sensitive datasets, governed by stricter access and usage policies. Data annotation and metadata standards as well as established data curation processes are central to ensure that findings meet high standards of credibility and reproducibility. Metadata provides context and provenance to raw data and methods; therefore, they are essential to both discovery and validation. A good example is the meteorology field, which uses data and metadata formats standardised by the World Meteorological Organisation (WMO).

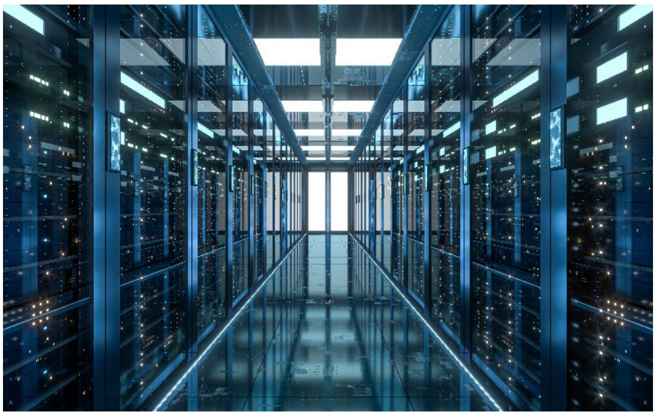
With the rapidly increasing power of HPC machines and the advent of Exascale computing, critical processes can be modelled and simulated much faster than in the past, making it possible to work on short- and mid-term predictions and analyses together with long-term studies. Examples are **complex natural phenomena** such climate change, ocean environment and sea level rising. The efficient and effective exploitation of large-scale computing capabilities is fundamental to support prompt data-driven decision making, as it is to obtain new scientific insights. A combined use of post-processing and visualisation serves to reduce, analyse and explore large volumes of data resulting from simulations. Modelling and data processing codes must be able to efficiently use heterogeneous systems that combine general-purpose processors with accelerators, high-speed networks and multiple classes of memory/storage devices, that represent the state-of-the-art and future trends in supercomputing resources (for an example, see the directions set by the EuroHPC Joint Undertaking [20]).

The opportunity of using cutting edge computing systems is not just a way to reduce the time-to-solution, but the sole viable approach to process datasets of the size and complexity expected in the type of use cases considered [12,15]. The observed data can be used as input for a numerical simulation that must then be post-processed and delivered to the decision makers or domain experts so they can take appropriate actions in near real-time. Moreover, they must be compared with theoretical simulations to enable domain experts to understand nuanced predictions, as well as shape experiments more efficiently (see Figure 1). The time scale for the analysis and simulations to be completed depends on the specific use case. However, the ability to intervene and modify pre-determined, batch processing schedules to secure the resources needed ad hoc is essential for generating actionable results within the tight time windows for events whose occurrence cannot be predicted a-priori.

Using cutting edge computing systems is not just a way to reduce the time-to-solution, but the sole viable approach to process datasets of the size and complexity expected

Data exploration and visualisation are further crucial aspects of the decision-making workflow, relying on the automated analysis and determination of possible actions (plus their consequences). Data exploration is the essential initial step in data analysis, where experts and policymakers explore a large data set in an unstructured way to uncover initial patterns, characteristics, and points of interest. Data exploration can use a combination of manual methods and automated tools to generate artefacts for decision support, such as data visualisations, charts, and initial reports. Data visualisation assists experts and policymakers in reaching a clear understanding of the information, by giving it visual context through maps or graphs. It is possible to identify different scenarios where data exploration is implemented for urgent decision making - for example: automatic decision based on pre-defined or AI driven approaches [17,18] where visualisation is used as a means for post-mortem analysis of the outcomes resulting from human-driven decision making where exploration is necessary to confine available data and focus the decisions on relevant information. As in AI models, there is the concern of explainability [22], i.e. ensuring that humans can adequately understand and oversee algorithmic processes, and establishing ways for explainable decision-making technologies to be used in combination with domain knowledge from the application areas. Finally, HPC is part of an overarching framework of organisational procedures in support of urgent decision making; therefore, HPC resource acquisition and usage protocols should be regularly reviewed and rehearsed to ensure effective response to emergencies. However, human-driven analysis and exploration steps, and protocols for orchestrating emergency response, are outside the scope of this white paper.

2. Impact on technologies and requirements



The most widely adopted and reliable method for supporting prompt decision making is the use of dedicated resources as in the case of the Japanese earthquake early warning system [14] or operational weather forecasting organisations (such as the European Centre for Medium-Range Weather Forecasts, Meteo France, UK Met Office, DWD). This is a cost-efficient way to ensure the required response times, in particular for the use cases where continuous computing is required, for example, weather forecasting or ocean and climate modelling. However, setting up dedicated resources for each domain-specific urgent computation, constantly ready and waiting to run forecasts for disasters which may never come, is economically unviable, in particular when computations take place rarely and require a huge amount of resources.

Additionally, having a fixed set of HPC resources for disaster response could also cause problems should multiple disasters occur simultaneously, as there would be no capacity to scale to larger computing resources if required. For most cases, it is therefore prudent to use time on existing international, national or regional HPC systems, or to combine dedicated resources with existing ones into a flexible and scalable environment. Rapidly engaging such resources poses significant challenges:

- Preempting already running, lower priority jobs to free required resources (rapid resource availability for urgent computing)
- Provisioning the complete workflow needed to react to an urgent (real-time or rapid analysis) event
- Performing data acquisition, execute simulations on HPC machines, run data analytics and presenting the results to decision-makers (e.g. public authorities, scientists, domain experts)
- Adapting the set of resources used and the workflow as the situation evolves in unpredictable ways.

Existing HPC infrastructures are not commonly used for urgent computing, since rapidly mobilising them faces a few (but crucial) technological challenges. For example, the usual approach of reserving resources well in advance does not meet the requirements of urgent computing. Preemptive scheduling is an effective alternative to swiftly obtain resources on HPC infrastructures when an urgent need arises unexpectedly. This requires a change in the mode of operation, and in addition the use of novel technologies: for example, checkpointing any running applications that are preempted in a non-disrupting and efficient way becomes mandatory. Furthermore, challenges related to pre-emption in the context of high-performance interconnects and accelerated architectures need to be solved and resource occupation by preempted jobs has to be minimised.

3. Challenges in Advancing HPC for Urgent Decision-Making

Figure 2 illustrates a high-level view of the architecture of an HPC-based system designed for urgent decision support. The main capabilities as expected are simulation and data management, but with additional features arising from the need to utilise distributed resources and data assets. The user interface can be part of a web portal for HPC resources.

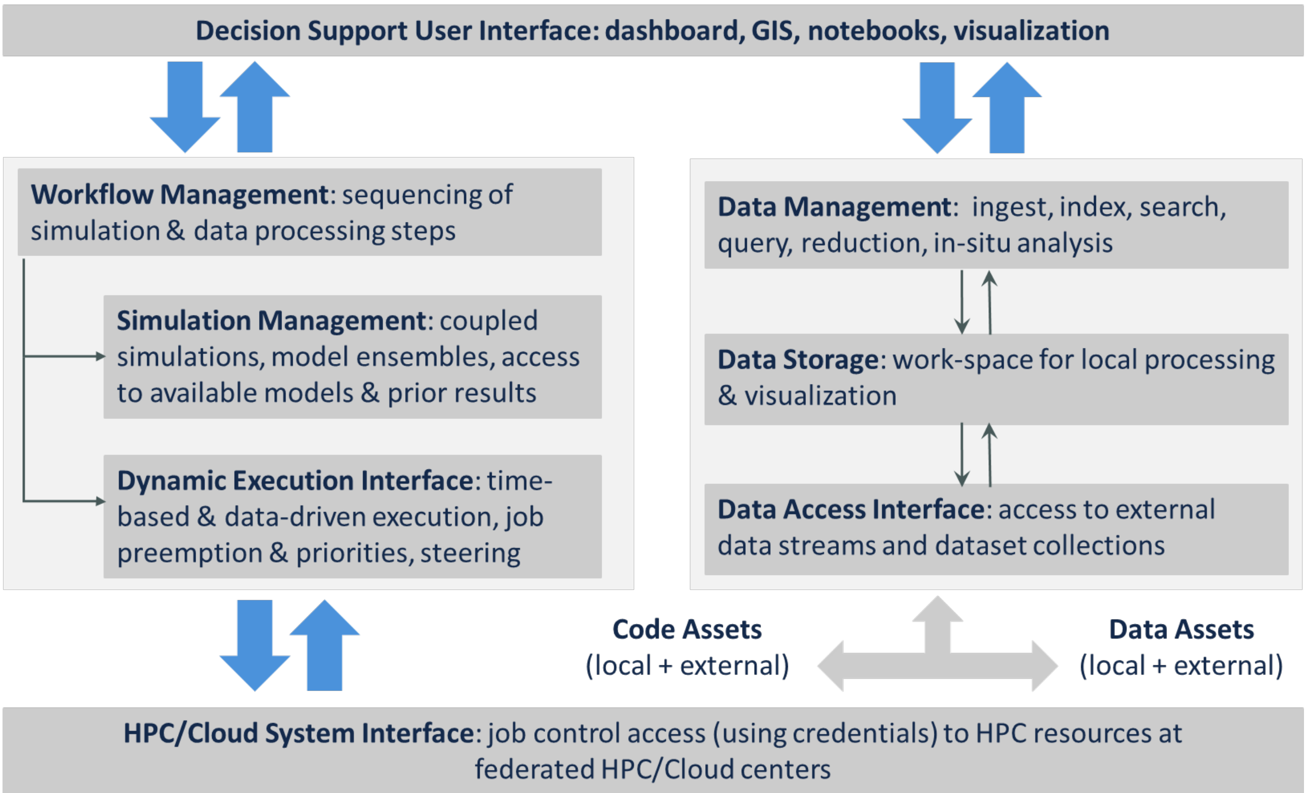


Figure 2: Architectural overview of an HPC-based urgent decision support environment.

The following list summarises ideas and challenges to consider in advancing the current state of the art for HPC-based urgent decision support:

- HPC systems tend to be optimised for job throughput so there could be very long (hours or even days) wait times in the batch queue until the resources for urgent computing become available. It is therefore necessary to **improve the management of job priorities and of job preemption, in the context of coarse-grain units of work**.
- Priority management and preemption are necessary for prompt response to incoming workloads associated with decision support workflows. Moreover, appropriate system software and application level **checkpointing and restart procedures** must be implemented to enable any preempted jobs to restart without losing significant work. Extensive usage of accelerators in HPC workflows is essential for prompt response; however, it exacerbates the complexity of state capture and restoration (in particular, in the context of checkpointing protocols).
- HPC systems are usually unique environments with specific interconnects and accelerators, as well as software sets customised to particular domains and use cases. Running an optimised code requires machine-specific porting. Deployment can be expedited with the use of (hardware platform-specific) **container technologies** that facilitate handling of software dependencies and portability.
- CPU and storage resources are unlikely to be co-located in urgent computing use cases. Input data must be accessible from the HPC systems inside supercomputing centres machines for processing, and at the same time results need to

be stored and made accessible to decision makers. **Advanced data management and caching services** should be implemented, able to interoperate with modern data storage and management architectures such as data lakes, and support interaction with large-scale data repositories of both raw and transformed data. A key problem is the staging of required data onto high-performance parallel filesystems or node-local storage.

- Opening up HPC centres for utilising their resources on short-notice and in the context of dynamic workflows will exacerbate cybersecurity challenges due to the increased attack surface. Protection and validation of system and data integrity become particularly acute challenges. **Data integrity validation**, including reliable tracking of provenance, becomes a particularly acute challenge, throughout the entire lifecycle of large-scale data sets. Executing workflows involving diverse computing and data resources entails adding data transfers and storage solutions to the decision support workflow, which can make it more difficult to ensure that data have not been corrupted in transit or at rest. When data integrity is not preserved, computations can fail and result in increased computational cost due to reruns, or worse, results can be corrupted in a manner not apparent to data analysts, and produce invalid results with severe consequences to the real-world impact of the decision-making process.
- **Simplified access to HPC resources** and the results of the urgent computing analysis is of course mandatory: the ability to obtain context-relevant information quickly and easily allows decision makers to act and respond to findings swiftly. Science platforms [23] are a promising modern approach to interact with data, resources and software, and could provide a useful starting point. APIs and common standards should be defined to develop community/service specific platforms, including standards to facilitate interaction with different workload managers.
- **Federated HPC environments** involving supercomputing centres all over Europe can be instrumental to optimise resource usage for different urgent actions. This usage mode of HPC resources requires the following:
    - Ability to pick the proper supercomputer architectures and systems according to the code and workflow to be executed
    - Identification of cross-centre policies (rather than a large set of disparate HPC centre-specific policies) to allow the use of resources that are rare and extremely expensive, including authorisation, authentication and accounting policies
    - Definition of policies for storage access and data persistence.

The idea of a Federated HPC environment opens a new set of challenges, including unified authentication and authorisation infrastructure (AAI) for user management, elastic and resilient resource management, and the definition of an innovative resource information system able to monitor supercomputer health status and load, and eventually predict resource status in advance.

A common challenge present in a wide range of use cases [1, 11, 12] is the increasing dynamicity and diversity in the definition of what units of work have to be managed, including ensembles of jobs, data-staging to support workflows, and interactions with services/facilities external to the HPC system. Our analysis of salient features of these use cases identifies the need for advances towards **more interactive access to HPC resources**, particularly in the context of dynamic workflows over large datasets. Several challenges, both technical and organisational, have to be considered. A pragmatic approach is to maintain both classical HPC clusters for batch-style use of computational and storage resources, together with clusters more tuned for dynamic workflows and interactive usage modes. HPC centres supporting prompt decision making could provision (i) one set of resources for use by predictive simulations not directly involved in the decision making in near real-time, (ii) a specialised set of resources for interactive-style use by decision makers (e.g. visualisation and data exploration) , and (iii) one set of dedicated or at least preemptable resources for actual processing within tight time constraints. Nevertheless, we should keep in mind that the HPC resource requirements for urgent decision support can be quite substantial; therefore, a static partitioning might turn out to be ineffective, necessitating a set of technical improvements along the lines suggested in this white paper.

Major challenges arise in the following areas: dynamic resource allocation and scheduling, coordination of resource managers, short-notice secure access to federated HPC and Cloud resources, data-intensive workflow support (including data staging on node-local storage), increased interactivity (including preemption and simulation steering). Additionally, cybersecurity concerns need to be considered.

Conclusions

The current state-of-the-art does not adequately cover the use of HPC environments for prompt decision support. In particular, applying HPC in a wide range of use cases for gaining insight and supporting decisions in short time scales causes increased dynamicity and diversity in the units of work which have be managed. This includes ensembles of jobs, data-staging to support workflows, and interactions with services/facilities external to the HPC systems or centres. Model-based simulation, the real-time adaptation of interlinked computational models of evolving complex processes and distributed data-driven scientific collaborations require advances towards more interactive access to HPC resources and support for preempting running jobs. Challenges in the technical and organisational area have to be addressed, all in the context of dynamic workflows over large datasets.

Specific technical challenges arise in the following areas: dynamic resource allocation and scheduling, preemption techniques with transparent and efficient job checkpointing and restart, coordination of resource managers, short-notice secure access to federated HPC resources, data-intensive workflow support (including data staging on node-local storage), interactive style and near real-time use (including preemption and simulation steering). We have outlined a set of short- to medium-term recommendations for R&D actions which would enable the effective use of HPC resources for extracting insights and supporting decisions in short time scales.

References

[1] Destination Earth (DestinE) initiative. https://ec.europa.eu/digital-single-market/en/destination-earth-destine

[2] Destination Earth: Use Cases Analysis, JRC Technical Report JRC122456, 2020. https://publications.jrc.ec.europa.eu/repository/handle/JRC122456

[3] Wilkinson MD, Dumontier M, Aalbersberg IJ, et al.. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016 Mar 15;3:160018. doi: 10.1038/sdata.2016.18. Erratum in: Sci Data. 2019 Mar 19;6(1):6. PMID: 26978244; PMCID: PMC4792175.

[4] N. Brown, R. Nash, G. Gibb, B. Prodan, M. Kontak, V. Olshevsky, and W. Der Chien, "The role of interactive supercomputing in using HPC for urgent decision making", in Proceedings of the International Conference on High Performance Computing. Springer, 2019, pp. 528–540.

[5] G. Gibb, R. Nash, N. Brown and B. Prodan, "The Technologies Required for Fusing HPC and Real-Time Data to Support Urgent Computing", in Proceedings of the 2019 IEEE/ACM Workshop on HPC for Urgent Decision Making (UrgentHPC), 2019, pp. 24-34.

[6] Earth System Modeling Framework : https://earthsystemmodeling.org/

[7] T. C. Schulthess, P. Bauer, N. Wedi, O. Fuhrer, T. Hoefler and C. Schär, "Reflecting on the Goal and Baseline for Exascale Computing: A Roadmap Based on Weather and Climate Simulations," in Computing in Science & Engineering, vol. 21, no. 1, pp. 30-41, 1 Jan.-Feb. 2019, doi: 10.1109/MCSE.2018.2888788.

[8] Baker, D.N., Erickson, P.J., Fennell, J.F. et al. Space Weather Effects in the Earth's Radiation Belts. Space Sci Rev 214, 17 (2018). https://doi.org/10.1007/s11214-017-0452-10

[9] R. Kube et al., "Near real-time analysis of big fusion data on HPC systems," 2020 IEEE/ACM HPC for Urgent Decision Making (UrgentHPC), 2020, pp. 55-63, doi: 10.1109/UrgentHPC51945.2020.00012.

[10] A. Kremin, S. Bailey, J. Guy, T. Kisner and K. Zhang, "Rapid Processing of Astronomical Data for the Dark Energy Spectroscopic Instrument," 2020 IEEE/ACM HPC for Urgent Decision Making (UrgentHPC), 2020, pp. 1-9, doi: 10.1109/UrgentHPC51945.2020.00006.

[11] Jiang, M., Bu, C., Zeng, J. et al. Applications and challenges of high performance computing in genomics. CCF Trans. HPC (2021). https://doi.org/10.1007/s42514-021-00081-w

[12] CISCO 2020, Global Network Trends Report, Tech. rep., CISCO. URL https://www.cisco.com/c/dam/m/en_us/solutions/enterprise-networks/networking-report/files/GLBL-ENG_NB-06_0_NA_RPT_PDF_MOFU-no-NetworkingTrendsReport-NB_rpten018612_5.pdf

[13] Asch M, Moore T, Badia R, 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. The International Journal of High Performance Computing Applications. 2018;32(4):435-479. doi:10.1177/1094342018778123

[14] E.Yamasaki, 2012, What We Can Learn From Japan's Early Earthquake Warning System, Momentum: Volume 1: Issue 1, Article 2.

[15] F. Løvholt, S. Lorito, J. Macias, M. Volpe, J. Selva and S. Gibbons, "Urgent Tsunami Computing," 2019 IEEE/ACM HPC for Urgent Decision Making (UrgentHPC), 2019, pp. 45-50, doi: 10.1109/UrgentHPC49580.2019.00011.

[16] Siew Hoon Leong, Dieter Kranzlmüller, "Towards a General Definition of Urgent Computing," Procedia Computer Science, Volume 51, 2015, https://doi.org/10.1016/j.procs.2015.05.402.

[17] Tzachor, A., Whittlestone, J., Sundaram, L. et al. Artificial intelligence in a crisis needs ethics with urgency. Nat Mach Intell 2, 365–366 (2020). https://doi.org/10.1038/s42256-020-0195-0

[18] Chen, N., Liu, W., Bai, R. et al. Application of computational intelligence technologies in emergency management: a literature review. Artif Intell Rev 52, 2131–2168 (2019). https://doi.org/10.1007/s10462-017-9589-8

[19] D. Elia, S. Fiore and G. Aloisio, "Towards HPC and Big Data Analytics Convergence: Design and Experimental Evaluation of a HPDA Framework for eScience at Scale," in IEEE Access, vol. 9, pp. 73307-73326, 2021. https://doi.org/10.1109/ACCESS.2021.3079139

[20] European High Performance Computing Joint Undertaking (EuroHPC JU). https://eurohpc-ju.europa.eu

[21] A European Green Deal. https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_en

[22] R. Roscher, B. Bohn, M. F. Duarte and J. Garcke, "Explainable Machine Learning for Scientific Insights and Discoveries," in IEEE Access, vol. 8, pp. 42200-42216, 2020, doi: 10.1109/ACCESS.2020.2976199.

[23] Strategic Research and Innovation Agenda of the European Open Science Cloud (EOSC), Feb. 2021. https://www.eosc.eu/sites/default/files/EOSC-SRIA-V1.0_15Feb2021.pdf

Authors:

Manolis Marazakis is a Principal Staff Research Scientist at the Foundation for Research and Technology - Hellas (FORTH - Greece).

Marc Duranton is a member of the Research and Technology Department of the Commissariat à l'énergie atomique et aux énergies alternatives (CEA - France).

Dirk Pleiter is a Professor of High Performance Computing at the Kungliga Tekniska högskolan Royal Institute of Technology and Director of the PDC Center for High Performance Computing (KTH - Sweden).

Giuliano Taffoni is a Development Scientist at the Astronomical Observatory of Trieste, Istituto Nazionale di Astrofisica (INAF - Italy).

Hans-Christian Hoppe is working as a senior project lead with Scapos AG (Germany).

Manolis Marazakis has received support from the European Union's EuroHPC research and innovation programme under grant agreement 955606 (DEEP-SEA, part of the EuroHPC-01-2019 call - project website: https://www.deep-projects.eu). National contributions from the involved state members (including the Greek General Secretariat for Research and Innovation) match the EuroHPC funding. Dirk Pleiter has received support from the European Union's H2020 research and innovation programme under grant agreement 823767 (PRACE-6IP, part of the (H2020-INFRAEDI-2018-1 call - project website: https://prace-ri.eu/about/ip-projects). Giuliano Taffoni has received support from the European Union's H2020 research and innovation programme under grant agreement 824064 (ESCAPE, part of the H2020-INFRAEOSC-2018-2 call - project website: https://projectescape.eu).

The authors would like to thank Maria S. Perez (Professor at Universidad Politécnica de Madrid, Spain) for her insightful critique on earlier drafts of this whitepaper.

Cite as: M. Marazakis et al., « HPC for Urgent Decision-Making », ETP4HPC White Paper, 2022, doi: 10.5281/zenodo.6107362.

DOI: 10.5281/zenodo.6107362