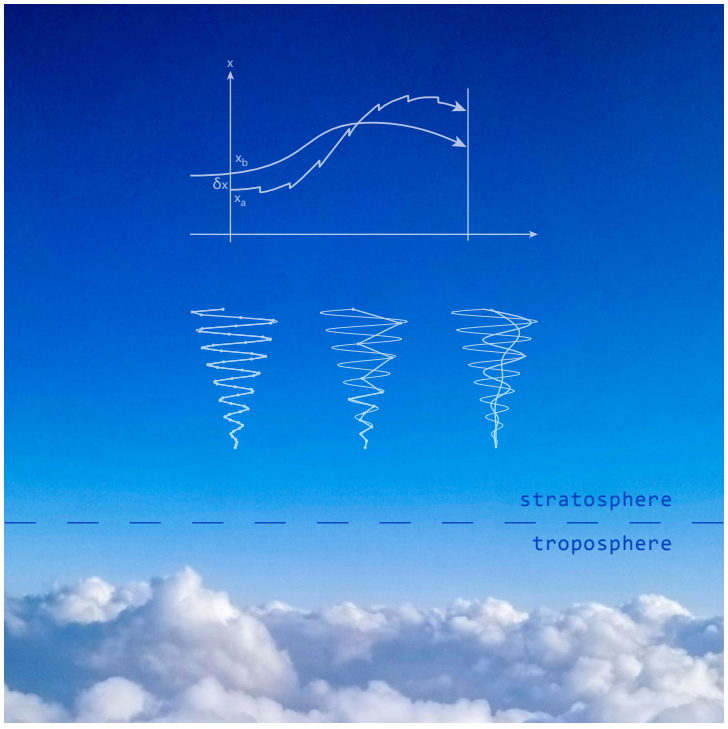
ECMWF Feature article

from Newsletter Number 163 – Spring 2020

COMPUTING

HPC2020 – ECMWF's new High-Performance Computing Facility



Cover photo: batuhanztrk / iStock / Getty Images Plus

www.ecmwf.int/en/about/media-centre/media-resources

doi: 10.21957/371iwlk49p

This article appeared in the Computing section of ECMWF Newsletter No. 163 – Spring 2020, pp. 32–38

HPC2020 – ECMWF's new High-Performance Computing Facility

Mike Hawkins, Isabella Weger

ECMWF's High-Performance Computing Facility (HPCF) is at the core of its operational and research activities and is upgraded on average every four or five years. As part of the HPC2020 project, ECMWF has recently concluded a contract for its new system, made up of four Atos Sequana XH2000 clusters that will deliver about five times the performance of the current system, made up of two Cray XC40 clusters.

The HCP2020 project

Replacing ECMWF's HPCF is a multi-year effort. The HPC2020 project started as early as 2017 with the development and approval of a business case, followed by an international procurement which was concluded at the end of 2019. The implementation phase of HPC2020 started in early 2020 (Figure 1).

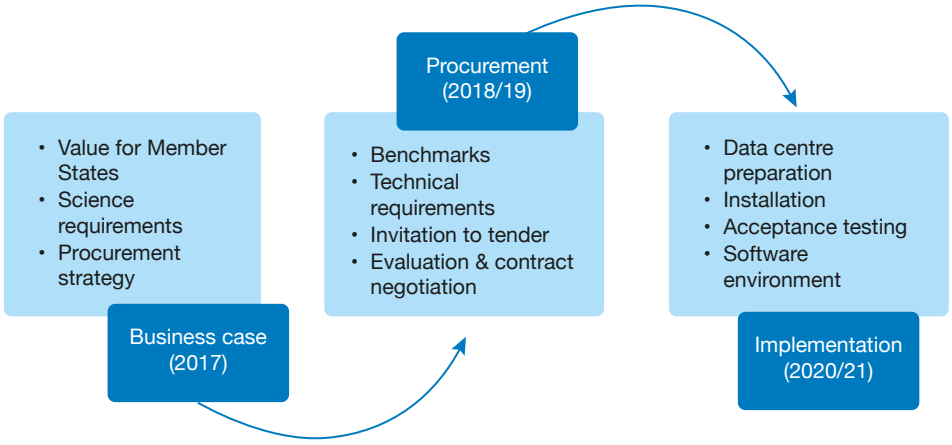


Figure 1 The HPC2020 project has moved on from the business case and procurement phases to the implementation phase.

Business case and procurement process

The goals set out in ECMWF's Strategy for the period 2016–2025 include making skilful ensemble predictions of high-impact weather up to two weeks ahead and predicting large-scale patterns and regime transitions up to four weeks ahead and global-scale anomalies up to a year ahead. These ambitious goals will only be achievable with the appropriate high-performance computing capability.

In December 2017, ECMWF's Council reviewed and approved the HPC2020 business case for an increased investment in computational resources. Recognising the crucial importance of HPC resources for the successful delivery of the Strategy, Council approved an increase of about 75% in the HPC budget, which covers the cost of the HPC service contract and a contribution to the costs of the system's electricity consumption for the envisaged four-year service period.

ECMWF's procurement approach was to maximise the performance of its main applications within the available budget envelope. Hence, the performance benchmarks were based on running two distinct workflows relevant to ECMWF: a capability benchmark which simulates the workflow of the time-critical forecasts, including product generation, at potential future resolutions; and a capacity benchmark based on the typical research workflow, to gauge the system throughput for the anticipated research load. To approximate a realistic workflow, these benchmarks were composed of inter-dependent tasks, including output and reading from persistent storage of full operational datasets.

doi: 10.21957/371iwlk49p

An invitation to tender for the provision of a high-performance computing facility was published in November 2018 and closed in March 2019. Following a review of the procurement outcome by Member State Committees, Council, in December 2019, authorised ECMWF's Director-General to sign a contract with the successful tenderer, Atos UK Ltd. The new HPCF will be provided under a four-year service agreement and will deliver a performance increase of about five over the current system, based on the time-critical capability and capacity benchmarks. It will be installed in ECMWF's new data centre in Bologna, Italy, and will run in parallel with the existing Cray HPCF until September 2021, when the new Atos system will take over the provision of the operational service for a period of four years until autumn 2025.

A new data centre for a new supercomputer

ECMWF's existing data centre in Shinfield Park, Reading, UK, which has served its purpose well for more than 40 years, will no longer be suitable to host future generations of supercomputers. Therefore, in June 2017 ECMWF Member States approved the proposal by the Italian Government and the Emilia Romagna Region to host ECMWF's new data centre in Bologna, Italy. The new data centre is being built on the site of the new Tecnopolo di Bologna campus, where the unused buildings and grounds of a former tobacco factory are being redeveloped (Figure 2).

Figure 2 ECMWF's new data centre (artist's impression) at the Tecnopolo di Bologna, Italy. (*Image: gmp von Gerkan, Marg & Partner*)

The HPC2020 project is part of the wider Bologna data centre programme (BOND), which also includes the delivery of the new data centre. The time schedules for HPC2020 and the delivery of the data centre are intimately connected. They have been designed to enable the new HPCF to become operational before the contract for ECMWF's current HPCF expires, allowing the necessary flexibility to adapt the schedules as both projects develop.

With ECMWF's entire data centre operations moving to Bologna, the migration comprises much more than just HPC applications, such as ECMWF's Integrated Forecasting System (IFS). The migration project is managing the scientific application migration and the move of the hundreds of petabytes of data held in the Data Handling System.

The new HPC2020 facility

The specification of the new system is based on ECMWF's long experience with HPC and the new requirements. Here we present some key concepts, the system configuration, and the software environment.

General concepts

At a high level, there are some important concepts in the design:

- **HPC facility.** The project is to provide a complete HPC facility, and not just a new supercomputer. The requirements include the 24x7 hardware and software support, a full-time application analyst, and customisation of the data centre to support the machines.
- **Multiple clusters.** The ECMWF HPC workload is dominated by a high throughput of short-running jobs: 90% of the resources of the system are consumed by jobs that require fewer than 8,000 processor cores. This characteristic, combined with the flexibility of the ecFlow workload management, eliminates the need to have all resources in one large cluster. For many generations of HPC systems, ECMWF has therefore had a system with two self-sufficient clusters, allowing the operational service to be run even if one cluster fails or is shut down for maintenance. Splitting the compute resources also reduces contention on shared components, such as the job scheduler and network, making for a more reliable and manageable system. For HPC2020, the number of clusters

doi: 10.21957/371iwlk49p

is increased to four to further improve resilience. For instance, this makes it possible to upgrade one of the clusters to the latest software levels as a test, while still maintaining a resilient system for operations.

- **Separation of research and operational file systems.** The operational suites run to very tight schedules. Suites are carefully set up and tested during the transfer of research to operations (R2O) to meet the production schedule, and good I/O performance is a critical part of this. To avoid the possibility of a predictable operational job competing with an I/O-intensive job from another source, there are dedicated and separate file systems for operational and research work.
- **Multiple file systems.** Like having multiple compute clusters, having multiple file systems improves resilience and maintainability and limits contention for resources at scale. These considerations have been especially important for storage sub-systems. This is partly due to the design of the Lustre file system used, which has a single metadata server per file system, and partly due to disks being mechanical devices that can fail. All file systems are connected to all four clusters. This has the great benefit that a job can be scheduled to any cluster, but it does introduce a common point of failure for the entire system, as potentially a faulty file system could affect all clusters.
- **General purpose and interactive login nodes (GPIL).** The HPC system has always run different types of workload, mainly parallel jobs that run on multiple nodes, but also jobs that only need one node or even one core. Dedicating an entire 128-core node to a job that only requires one processor core would clearly be a waste of resources, so work of this type is typically allocated to dedicated nodes where several jobs can efficiently share the node. The ECMWF Linux clusters lxc, lxop and ecgate have, in the past, provided other locations to run this workload as well. With significantly increasing data volumes, it becomes increasingly undesirable to move large amounts of data to a different platform. Consequently, and because of a large overlap of applications, all of the resources for this work are being included in the new HPC system. In addition, because of the data volumes, we expect more interactive data analysis, visualisation and software development on the system. These activities will also run on a set of dedicated GPIL nodes.
- **Time critical storage hierarchy.** Solid-state disks (SSD) have become commonplace since ECMWF procured the last HPC system. They have better access times and lower latency. They are thus a valuable means of achieving high I/O performance in a small amount of storage space, especially when accessing small files. Unfortunately, however, they are still more expensive than traditional disk-based storage at the same capacity. The new HPCF therefore has a hierarchical storage design with two pools of SSD storage in addition to traditional disk storage pools. Each SSD pool is designed to hold data generated by the operational forecast suites for a couple of days. After this time, the suite moves the data to storage pools with higher capacity, but lower performance.
- **Home file systems.** In addition to the high-performance parallel file systems, there is also a need for general storage space. In the current HPCF, the 'home' and 'perm' file systems on the HPCF are not visible from outside the system. In the new HPCF, the home and perm spaces will be common between the HPCF and other systems.

Atos Sequana XH2000 system configuration

The main system from Atos is made up of four self-sufficient clusters, also called 'complexes' (Table 1). Each cluster is connected to all the high-performance storage. There are two type of nodes that run user workloads: 'compute nodes' for parallel jobs, and 'GPIL nodes' for general purpose and interactive workloads. Other nodes have special functions, such as managing the system, running the scheduler and connecting to the storage. See Figure 3 for a schematic representation of a single cluster.

 doi: 10.21957/371iwlk49p

	Cray XC40	Atos Sequana XH2000
Clusters	2	4
Processor type	Intel Broadwell	AMD Epyc Rome
Cores	18 cores/socket, 36 cores/node	64 cores/socket, 128 cores/node
Base frequency	2.10 GHz	2.25 GHz (compute) 2.5 GHz (GPIL)
Memory/node	128 GiB (compute)	256 GiB (compute) 512 GiB (GPIL)
Total number of compute nodes	7,020	7,488
General purpose 'GPIL' nodes	208	448
Total memory	0.9 PiB	2.19 PiB
Total number of cores	260,208	1,038,848
Water-cooled racks	40	80
Air-cooled racks	0	10

Table 1 The Cray XC40 system and the new Atos Sequana XH2000 system.

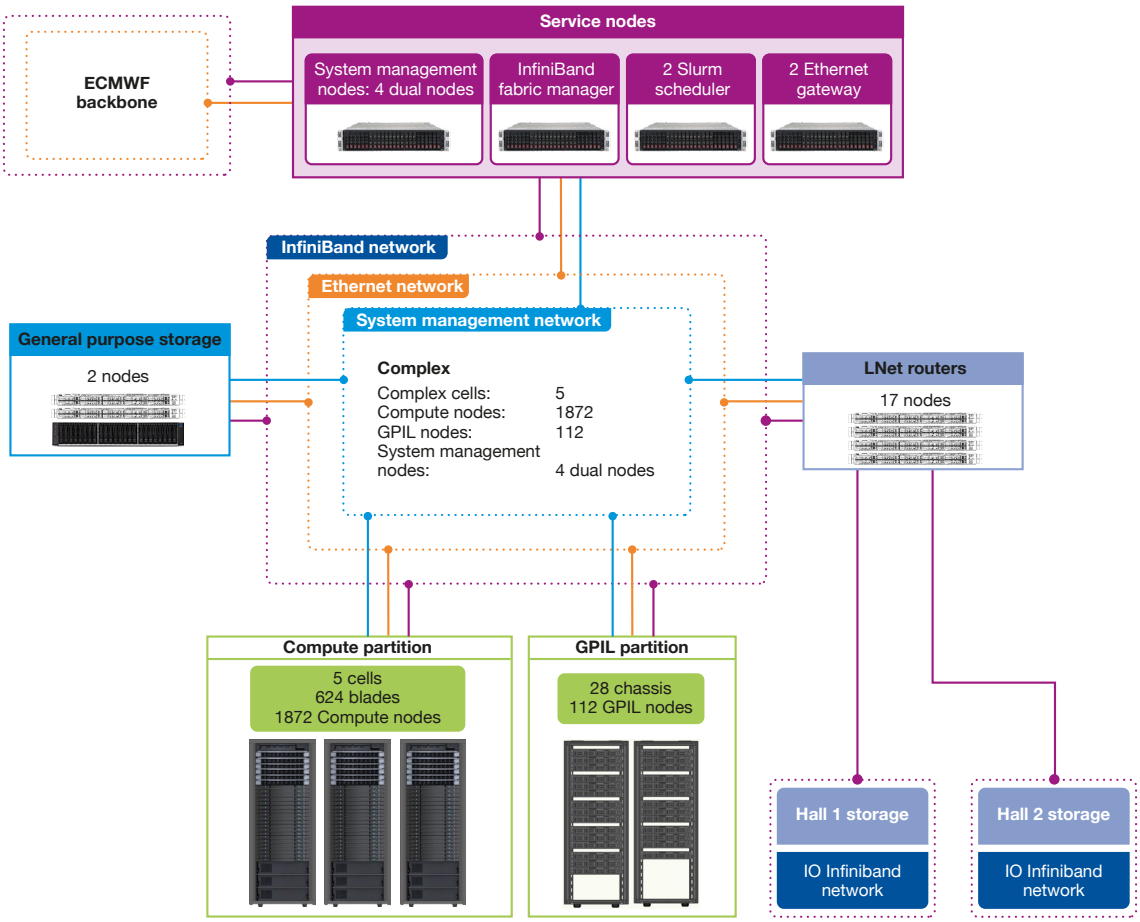


Figure 3 Overview of an HPC cluster. There are four of these in the system.

The 7,488 compute nodes form the bulk of the system and are located in Bull Sequana XH2000 high-density racks. Each rack has 32 blades (Figure 4), with three dual socket nodes per blade, and uses Direct Liquid Cooling to extract the heat from the processors and memory to a liquid-cooling loop in the rack. This cooling method allows the compute nodes to be densely packed and the number of racks to be minimised. A heat exchanger at the bottom of the rack connects to the building's water cooling system. The cooling system in the rack allows water to come in at up to 40°C. This provides plenty of opportunity for cooling using just the outside air, without any need for energy-consuming chillers. This lowers the amount of electricity required by the system and improves overall efficiency.

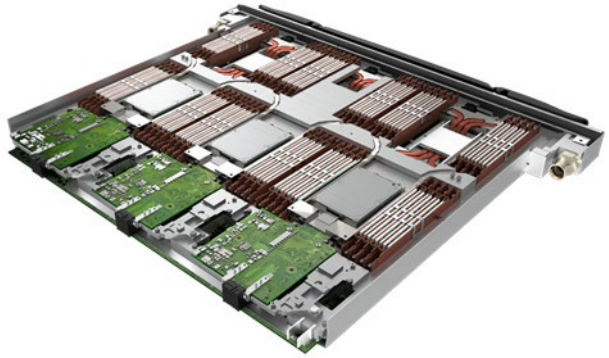


Figure 4 Sequana XH2000 AMD compute blade with three nodes.

The second type of node in the system is the GPIL nodes, which run at a slightly higher frequency and have more memory per node than the compute nodes. The different node type allows Atos to include a 1 TB solid-state disk in each node for local high-performance storage. The racks with GPIL nodes are less densely populated than the compute node racks and can therefore use a simpler cooling infrastructure. Fans remove the heat from the GPIL nodes and blow the hot air out through the water-cooled radiator 'rear-door' exchanger. While this is a less efficient cooling method compared to direct liquid cooling, and while it cannot handle the same kind of heat load or cooling water temperature, the use of standard servers facilitates maintenance and is less costly.

ECMWF's storage requirements have always been a large part of the requirements for any system. The storage of the new system, as outlined in Table 2, uses the Lustre parallel file system and is provided by DataDirect Networks (DDN) EXAScaler appliances. The use of the Lustre file system makes the solution quite like the storage on the current system. However, the newer Lustre version includes features such as 'Data on Metadata' to improve performance. Storage arrays and Lustre, which are primarily designed for handling large files, can be quite slow when handling lots of small files. The new metadata functionality allows small files to be stored on the controllers, rather than on the main disk storage arrays, which will lead to significant performance improvements.

Storage			
	Time critical – short term	Time critical – working	Research
IO Fabric racks	1 (shared) per hall, 2 in total		
Storage racks	3 per hall		9 per hall
Total racks (air cooled)	13 per hall, 26 in total		
Storage	DDN ES200NV	DDN ES7990	DDN ES7990
Usable storage	1.4 PB	12 PB	77 PB
Bandwidth	614 GB/s	224 GB/s	1,570 GB/s

Table 2 High-performance storage.

doi: 10.21957/371iwlk49p

A key differentiating factor between a high-performance computer and a lot of individual computers in a rack is the network that connects the nodes together and allows them to efficiently exchange information. The Atos system uses a state-of-the-art 'High Data Rate' (HDR) InfiniBand network produced by Mellanox. The HDR technology boosts application performance by keeping latencies (the time for a message to go from one node to another) down to less than a microsecond and enabling each cluster to have a bisection bandwidth of more than 300 terabits per second (corresponding to the simultaneous streaming of 37.5 million HD movies).

The compute nodes in a cluster are grouped into 'Cells' of four Sequana racks. Each cell has 'leaf' and 'spine' switches. Each compute node is connected to a leaf switch, and each leaf switch is connected to every spine switch in the cell, so that all the 384 nodes in a cell are connected in a non-blocking 'fat-tree' network. Each of the spine switches has a connection to the corresponding spine in every other cell, yielding a 'full-bandwidth Dragonfly+' topology. The GPIL nodes are connected to the high-performance interconnect so that they have full access to the high-performance parallel file system.

As well as the networking for the compute and storage traffic, network connections to the rest of ECMWF are needed. Gateway routers connect from the high-performance interconnect to the four independent networks in the new ECMWF data centre network, enabling direct access from other machines to nodes in the HPC system.

Software environment

The 'Atos Bull Supercomputing Suite' is the Atos software suite for HPC environments. It provides a standard environment based on a recent Linux distribution (RedHat Linux RHEL 7) supported by Mellanox InfiniBand drivers, the Slurm scheduler from SchedMD, the Lustre parallel file system from DDN, and Intel compilers. In addition, PGI and AMD compiler suites and development tools will be available.

For in-depth profiling and debugging, the ARM Forge product can be used. This software package includes a parallel debugger (DDT) and a performance analysis tool (MAP). The Lightweight Profiler (Atos LWP) is supplied as well. It provides global and per-process statistics. LWP comes with the Bull binding checker to allow users to confirm that the process binding they are using is as expected. This is an essential function to obtain better CPU performance with large core count, multi-core processors and hyperthreading.

Containerised software can be run through Slurm using standard commands. Atos provides a complete framework based on the Singularity software package as well as a Slurm plugin for submission and accounting, and a tool to help container creation by users.

Component	Description
Operating system	Red Hat Enterprise Linux
Main compiler suite	Intel Parallel Studio XE Cluster Edition
Secondary compiler suites	• PGI compilers and development tools • AMD AOCC compilers and development tools
Profiler / debug tool	ARM Forge Professional
Batch Scheduler	Slurm

Table 3 Main software components of the new HPCF.

doi: 10.21957/371iwlk49p

Implementation plan

Unlike in previous HPC migrations, where the project has been responsible for delivering the system and migrating the applications, this time the work has been split in two. This reflects the fact that the migration is a much bigger project than usual since it involves additional clusters and gateway systems and the new interactive environment. The HPC project provides the systems and a migration project in the BOND programme is responsible for the application migration. Both projects are expected to be complete by September 2021 (Figure 5).

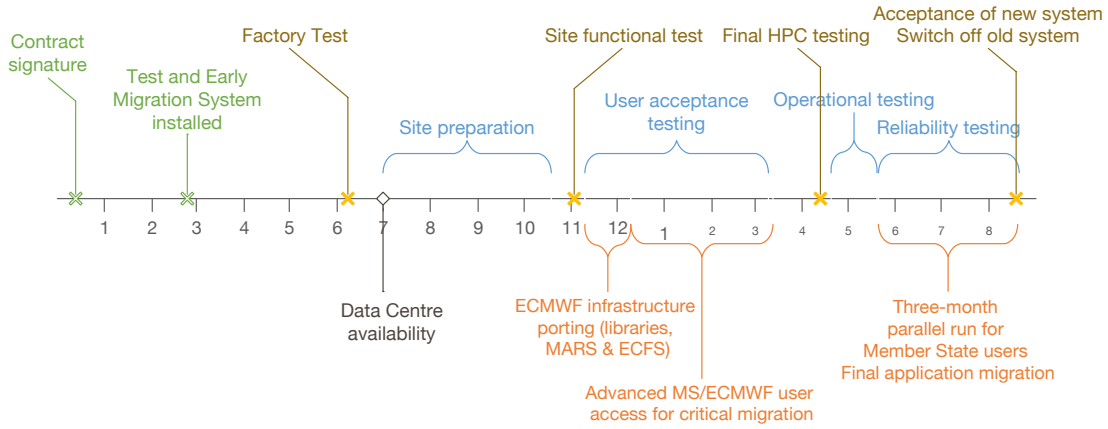


Figure 5 Implementation phase timeline. Timings may shift depending on possible impacts of the current COVID-19 pandemic on the BOND programme.

As part of the HPC service contract, Atos will supply a full-time application support analyst, based at Shinfield Park, to help users and developers port code and make the best use of the system. A full training programme is being developed and will be available to users and developers during the migration.

Test and Early Migration System

To enable porting and testing work to start as quickly as possible, a temporary 60-node 'Test and Early Migration System' (TEMS) has been installed in the Shinfield Park data centre. The system, which has been available to users since mid-March 2020, will make it possible to port and test all ECMWF libraries, utilities and applications, though with limits on the size and number of jobs that can be run.

Once the first main components of the new HPCF have been installed in Bologna, the Reading system will be dismantled and shipped to Italy, where its components will be reused in one of the later clusters. The configuration of the initial TEMS system is shown in Table 4. In this system, unlike in the main system, compute nodes and GPIL nodes have exactly the same specification. As the GPIL nodes are air-cooled, the TEMS could be installed in ECMWF's existing data centre without the need for expensive work to install water-cooling infrastructure.

Component	Quantity	Description
Compute node	40	2 x AMD 'Rome', 64 cores/socket, 512 GiB, 1 TB SSD
GPIL node	20	2 x AMD 'Rome', 64 cores/socket, 512 GiB, 1 TB SSD
General purpose storage	1	NetApp E2800 storage device with 64 TB
High performance storage	1	DDN ES7990 EDR appliance with 450 TB
High performance networking		Mellanox InfiniBand HDR 200 Gb/s

Table 4 Test and Early Migration System as installed in Reading.

doi: 10.21957/371iwlk49p

Since the system configuration is relatively small, access is limited to ECMWF application migration teams and Member State users by invitation only. Interested Member States users are advised to contact User Support via the Service Desk.

A smaller test system will be installed in Bologna to provide, over the duration of the contract, a platform for testing new software and procedures.

Acceptance process

The acceptance process for an HPCF is complex. It comprises tests to confirm that the functionality and performance of the system meet ECMWF's expectations and the vendor's contractual commitments, as well as reliability testing. Learning from the experience of previous HPC migrations, the implementation timeline and acceptance test procedure allow for more time for setup and the migration of applications to the new system.

The test steps are:

- **Factory Test** – A substantial portion of the main system is built in the vendor factory. ECMWF runs a 5-day functional test to verify that the system meets the contracted performance and functionality. This provides an opportunity for ECMWF to identify any issues early in the process. During testing in the factory, the vendor will have plenty of people on hand to see and resolve any issues that might be found. If the system passes this test, it can be shipped to the Bologna data centre.
- **Site Functional Test** – After a substantial part of the system has been installed in Italy, a functional test is run to confirm that it still works after shipping, to conduct any tests that could not be carried out in the factory, and to ascertain whether any important bugs identified in the Factory Test have been resolved. This test determines if the system is ready for user access.
- **User Acceptance Test** – This test is the first of the reliability tests. To pass it, Atos must demonstrate that the system meets specific availability targets. During this test phase, Atos will gradually build the complete final system, going from one cluster to all four, and it will use this time to resolve any outstanding issues. The first user access to the system will be allowed during this period, and the BOND project will use this time to migrate the applications. User access is expected to be possible by the end of 2020 or early in 2021.
- **Operational Test** – The operational test comprises two parts: the final functional tests, which will verify that the system fully confirms to the committed performance levels and functional specifications; and a 30-day operational reliability test, in which Atos has to demonstrate that the system fully meets the availability and reliability requirements. During this test period, the BOND migration project is expected to run the time-critical operational suite in a real-time test to prove that we can meet the forecast delivery schedule.
- **Reliability Test** – The last test is the final verification of the system's reliability and availability. During this period, ECMWF will disseminate test data to external users so that they can test their workflows and prepare for the transfer of the operational workload to the new HPCF. ECMWF and Member State users will port the remaining research workload during this period.

Outlook

From 2021 onwards, the new Atos Sequana HPCF will support ECMWF's operational and research activities for the following four years.

In previous HPC procurements, computing requirements were driven by a resolution upgrade of the high-resolution forecast (HRES) early in the contract period, followed by other forecast improvements later in the period. This led to a two-phase implementation approach: first an initial installation with sufficient performance to implement the HRES resolution upgrade, followed by a mid-term upgrade.

The Strategy 2016–2025 called for a different approach as it puts the focus on a significant increase in the resolution of ensemble forecasts (ENS). This requires a significant upgrade in computational performance, both in terms of capability and capacity, from the start, in one big step rather than two smaller steps. The upgrade of the ensemble forecast horizontal resolution from currently 18 km to 9 km (or in any case to the order of 10 km, the exact resolution depending on the testing that will take place) is expected to be implemented shortly after the new HPCF has become operational. Hence, the HPC2020 contract is a four-year service contract with no contracted mid-term upgrade.

doi: 10.21957/371iwlk49p

However, the HPC2020 contract includes options for ECMWF to enhance its HPC resources during the term of the contract. ECMWF could potentially secure additional funding to run new services, including funding of additional HPC requirements, under agreements to be concluded in the future. This may require enhancement of the HPC facility and a significant increase in capacity. In addition, in order to keep the configuration of the system in line with changing requirements, potential smaller enhancements or adaptations of the system are envisaged. These could include enhancements of the computational performance by adding further compute nodes or storage infrastructure or the introduction of other hardware, such as general purpose GPUs, to support the continued development of ECMWF's applications using state-of-the-art HPC and AI technologies.

 doi: 10.21957/371iwlk49p