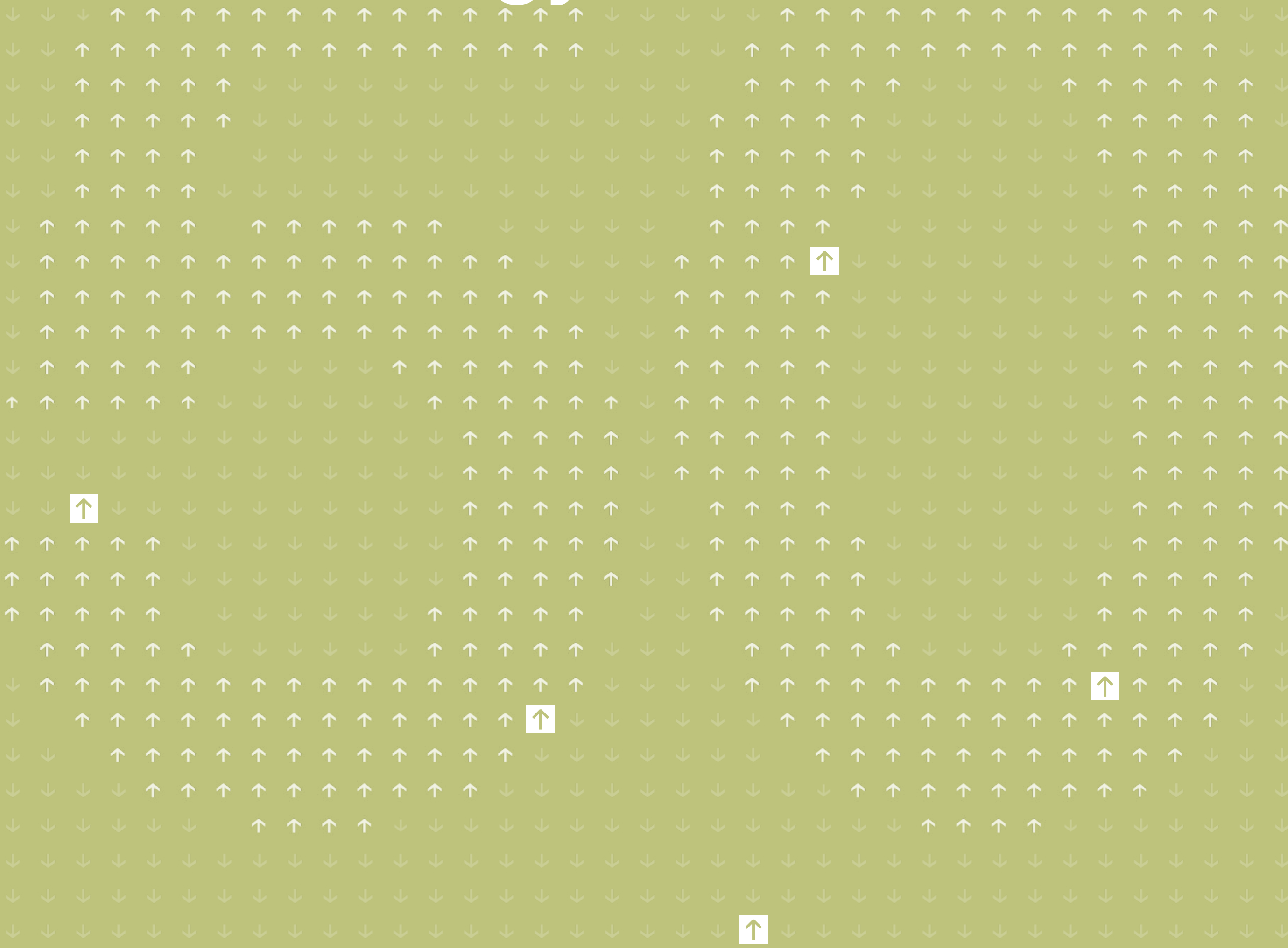
The critical role of high-performance computing in medium-range weather forecasting: half a century of technology innovation



ADVANCING WEATHER SCIENCE
THROUGH COLLABORATION

ECMWF | 50

The critical role of high-performance computing in medium-range weather forecasting: half a century of technology innovation

ABSTRACT ↘

Over the past five decades, high-performance computing (HPC) and data storage technologies have performed a pivotal role in advancing the operational time-critical capabilities and scientific research at ECMWF and among its Member States. HPC systems have enabled researchers to develop and run increasingly complex numerical weather prediction simulations, improving the model fidelity and increasing forecast lead times. Concurrently, data storage innovations have been vital to support the vast and expanding datasets, preserving half a century of meteorological data essential for weather and climate research. This article highlights the Centre's infrastructure milestones from the early supercomputers to today's HPC and storage solutions which underpin real-time prediction and multi-decadal climate analysis. Together, these services have enabled major advances in forecasting precision and scope, allowing the meteorological community to meet the urgent demands of modern weather-related challenges, including forecast reliability, emergency management planning for weather-dependent sectors, and insights into global climate change.

We describe the evolution of ECMWF's core infrastructure into a highly resilient environment necessary to deliver time-critical services. The article highlights the roles of HPC, networks, data centre design, and storage in delivering ECMWF's mission to provide reliable, cutting-edge forecasts and support future generations of weather prediction systems. With data volumes and computational demands continuing to grow, ongoing technological innovation remains an essential component to advance medium-range weather forecasting.

Reflecting on the impact of technology over the decades, the continued scientific advancements have been achieved through the growth and diversity of the communities' skills in exploiting these new capabilities. It has been a pleasure to record the massive progress that has been achieved over the last 50 years. As readers we hope you enjoy the article as much as we did in researching the history of our services, including some of the changes in fashion and designs!

INTRODUCTION →

For decades, high-performance computing (HPC) and storage technologies have been at the heart of weather forecasting services, enabling meteorologists to model complex atmospheric dynamics with ever-increasing accuracy. From the early days of numerical weather prediction (NWP) to today's exascale generation supercomputers, advancements in computational capabilities and data management have enabled significant improvements in forecast precision and reliability. Accurate medium-range weather forecasting (forecasts ranging from 3 to 15 days ahead) is essential for disaster preparedness, agriculture, transportation, and energy management. The ability to predict weather patterns with precision relies on high-performance computing, which enable meteorologists to process vast amounts of atmospheric data and run sophisticated NWP models. Storage technology advances have allowed the Centre to keep pace with processing higher and higher volumes of data from those sophisticated NWP model runs.

In the 1970s, weather prediction used early mainframe computers, which, whilst state of the art at the time, had limited computational resources, particularly with respect to processing and memory capabilities. Due to computational constraints, the model grid resolutions were relatively coarse and involved simplified physical parametrizations, leading to errors in simulating small-scale atmospheric phenomena such as localised storms. As HPC systems evolved, so did the ability to run sophisticated models that simulate atmospheric behaviour at finer spatial and temporal resolutions. This allowed numerical weather prediction centres to increase model complexity, incorporate more observational data, and extend the range of reliable predictions.

Alongside computational power, storage technologies have played a critical role in handling the exponential growth of meteorological data. The rise of satellite observations and ensemble forecasting has generated vast datasets requiring advanced, affordable high-speed storage, efficient retrieval systems, and scalable data infrastructures. Without these advancements, processing and analysing the massive volumes of atmospheric data necessary for accurate forecasting would not be possible.

25% of ECMWF's HPC capacity is allocated to Member States, providing national meteorological services with advanced computing and storage access to support research, model development, pre-operational testing, and operational enhancement activities. These resources also support Member States in exploring the use of AI-driven tools alongside traditional physics-based models.

This article describes the transition from ECMWF's early supercomputers to the latest-generation platforms and it explores the important role of the data centre and network technologies required to host and connect these systems in a highly resilient environment to operate complex time-critical services. It is also and above all a testament to the constant support and collaboration between ECMWF, its Member and Co-operating States, and the broader European Meteorological Infrastructure (EMI).

FROM PUNCH CARDS TO PETAFLOPS: 50 YEARS OF PERFORMANCE TRENDS →

ECMWF Newsletter No 3 (ECMWF, 1975) proudly reported that:

"Usually, 200 jobs a week or more are now being run on the CDC 6600 computer at John Scott House, using more than 20 hours central processor time (of the 40 hours available). On the U.K. Meteorological Office complex, a total of over 200 jobs were run in November (144 on the 360/195, 66 on the 370/158), using over 5 hours of 360/195 central processor time".

In the 50 years since that was written, the world has seen an almost unimaginable increase in the amount of computer power available; the ECMWF HPC service has increased in computational power by almost 30 million times, and 200 jobs are now run every 20 seconds rather than every week.

Figure 1: Control Data Corporation (CDC) 6600 computer hosted by CDC at John Scott House in Bracknell, UK, close to ECMWF's temporary accommodation.

ORIGINS OF ECMWF COMPUTING: FROM CDC TO CRAY

The beginnings

In December 1975, ECMWF had no dedicated data centre and rented time on a Control Data Corporation (CDC) 6600 computer to run 200 jobs a week (see Figure 1). With 40 hours of central processing unit (CPU) time available each week, the CDC supported initial model development but lacked the power necessary for operational forecasting. Whilst the CDC was itself a significant advancement in computing technology, being a factor of three times faster than the previous fastest system in the world, running a 10-day forecast would have taken 12 days!

1970s: the Cray-1 revolution

The Cray-1 was released in 1976 and was famously built in a C-shaped cabinet surrounded by a ring of benches that hid the power supplies and cooling systems (see Figure 2). The C-shape allowed the length of the myriad of cables connecting the individual module boards to be short and of similar lengths so that electrical signals arrived at the right time. Initially, ECMWF used the very first of these systems, serial number one, that had previously been installed at the Los Alamos National Laboratory in the United States. As this preceded the ECMWF headquarters, it was hosted at the Rutherford Laboratory, located in Didcot not far from Reading, UK. From 1978, ECMWF had its own permanent system, serial number nine, installed at the new headquarters at Shinfield Park in Reading.

The Cray-1 vector processing machine dramatically improved computational capabilities: the then groundbreaking 160 MFLOPS (million floating-point operations per second) reduced the forecast runtime from a theoretical 12 days to a practical 5 hours. For comparison, the Cray-1A computational capabilities were less than a tenth of the performance of today's smartwatch.

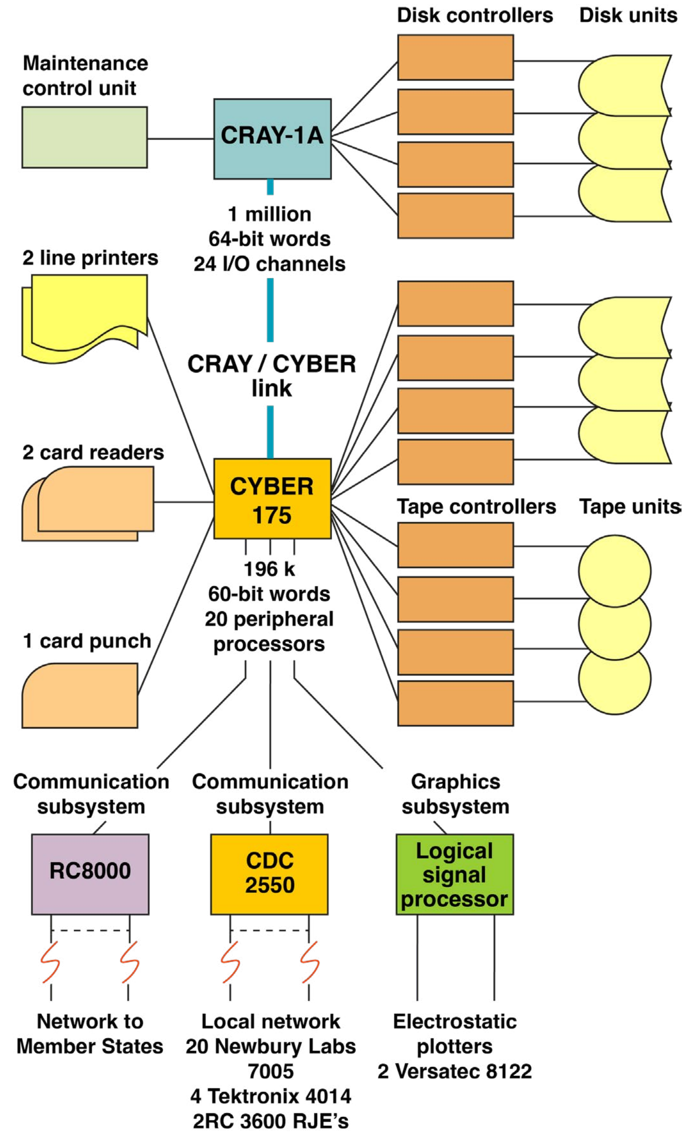
Figure 2: The Cray-1A in the ECMWF data centre in Reading (Shinfield Park).

Access to these early supercomputer systems was difficult; the Cray was a dedicated processing unit and required input and output to be fed through a front-end system (Figure 3). Programs had to be prepared and submitted on punch cards (Figure 4). Each punch card encoded a single line of code or data using a series of holes punched in predefined positions, with entire programs involving stacks of these cards. Users would prepare their job decks – often hundreds or thousands of cards long – and physically submit them to the computing centre's operations team. The Cray system relied on batch processing workflows where input from punch cards was read sequentially, compiled, and then executed by the system. This process was time-consuming and inflexible, and required meticulous card handling and error checking, as a single misplaced or incorrectly punched card could crash a job or produce faulty results. Despite these limitations, punch cards remained in use during the early years of high-performance computing, until the arrival of terminals, which gradually started being introduced at the Centre from 1979.

1980s: transition to terminals

Punch cards continued to be used into the early 1980s. However, by 1979 ECMWF had 20 alphanumeric visual display units and 4 graphical units in use; several terminals were placed in staff members' offices. Terminals proved to be very popular, so 34 more were purchased the following year, and by 1982 most staff had a terminal in their office, marking a significant productivity leap. These early systems established ECMWF as a pioneer in supercomputing for numerical weather prediction.

Figure 3: The Cray-1A and front-end systems, reproduced from ECMWF Technical Newsletter No. 1 (ECMWF, 1979). The main components were the CRAY-1A "number cruncher", a CDC CYBER 175 "front-end" and a Regnecentralen RC8000 telecommunications system, all linked via high-speed channels.



SCALING UP: CRAY TO FUJITSU AND THE PARALLEL COMPUTING SHIFT

1970s–1990s: the vector era expands

The Cray-1 system achieved its groundbreaking performance by having a small number of processors implementing vector processing where one instruction could perform calculations on a set of data, with hardware dedicated to specific instructions, e.g. addition and multiplication required separate hardware. The vector programming model carried ECMWF supercomputing through the next three decades, scaling its computing capacity with the Cray X-MP, Y-MP, and finally the C90 generations. These systems expanded vector computing, retaining a shared-memory architecture design.

1990s: Fujitsu Vector Parallel Processing (VPP) – a paradigm shift

In 1996, ECMWF transitioned to the Fujitsu VPP700 system – the first distributed-memory vector-parallel machine, providing more than 600 times the performance of the original Cray-1A. The VPP700 had 39 processing elements for computing, 6 for input/output (I/O), and 1 running the batch system and interactive work. Each

Figure 4: An ECMWF analyst using a punch card preparation system.

processing element had its own 2 gigabyte local memory, requiring applications to explicitly manage data exchange between the processors. The challenge to overcome for many scientific problems is that the calculation being worked on by one processor affects the calculation being performed on other processors, i.e. the weather in one location affects the weather in locations around it. So, there must be a very fast method of communicating information between tasks. This challenge demanded a fundamental redesign of forecasting software and introduced parallelism as a foundation of ECMWF computing models.

Networking challenges

This distributed-memory architecture required significant changes to the forecast system and a fast network to connect the processors. Without shared memory, applications had to send results to all the other parts of the simulation that might need that information. The configuration of the network design became an increasingly major element for software developers and system architects to consider, to manage costs and to optimise performance of the service. Supporting inter-process communication prompted the use of crossbar switches and network topology-aware coding strategies. Future HPC systems would move to more scalable network designs, including hypercubes and tree configurations, before eventually lighter configurations such as the dragonfly topology were introduced.

For the Fujitsu VPP700 system, the networking was a simple "non-blocking crossbar switch" and it was used for operational forecasts from September 1996.

THE MASSIVELY PARALLEL ERA: IBM TO CRAY XC30

Seymour Cray, the founder of Cray Research, was a strong advocate for the vector programming model of high-performance computing and is often quoted as saying: *"If you were ploughing a field, which would you rather use: two strong oxen or 1024 chickens?"* A statement of its time, that was soon made obsolete by the evolution of microprocessor technology. This saw the chickens evolve from the simple and specialised integrated circuit chips with a few thousand transistors in the 1970s to the current mass market devices with over 40 billion transistors. The increasing number of transistors on a chip became known as Moore's law, named after Gordon Moore, one of the co-founders of Intel, who noted in a 1965 article that circuit density was doubling roughly every year and was likely to remain that way for quite some time (Moore, 1965). The rapid development of commercial microprocessors in the 80s and 90s influenced the next phase of high-performance computing, the massively parallel systems.

2000s: IBM Power Systems

ECMWF's first massively parallel system arrived in 2003 with IBM Power 4. With around 1,400 processors based on IBM p690 servers, it vastly outperformed its predecessors. Each server had 32 processors logically partitioned into four 8-way nodes, each node with 8 gigabytes (GB) of memory, with a high-memory 12-node subset containing 32 GB per cluster.

The IBM proprietary Colony Switches temporarily provided the high-speed interconnect. However, as this was older technology, it was a short-lived system that was replaced a year later by two IBM p690+ clusters. Each 70-server cluster was connected by a pSeries "Federation" switch, which was four times faster, providing 1,700 MB/s per link bandwidth performance.

High Performance Parallel Interface (HiPPi) networks were introduced to link the HPCs with the storage clusters and with the IRIX-based SGI general-purpose systems as Ethernet could only provide 100 megabits per second (Mbps) links at the time. HiPPi could do 800 Mbps, a major achievement in High Performance Networks (HPN), which did not survive for long as Ethernet advancements quickly took over the market.

The general-purpose systems provided, amongst other things, in-house developed connectivity tools for sharing the HPC resources with Member States, enabling them to access and, almost seamlessly, submit remote jobs from their home site over the 9 Mbps Internet connection. These tools introduced cryptography into ECMWF for the first time. As the systems became more open to remote (full) access users throughout Europe, the need for security increased, access lists on the routers were no longer enough, and the market was in its infancy with regards to dedicated firewall devices. Stateless Private Internet eXchange (PIX) firewalls with very basic packet-filtering were amongst the few commercial devices available. These were not enough to keep the HPC secure, and stateful firewalls were only offered by small high-risk startup companies with very high price tags. ECMWF therefore decided to build its own firewalls based on the open-source Firewall Toolkit (FWTK) which already in the nineties offered stateful application-level security functionality. These initial firewalls were built on existing HP RISC redundant servers running HP-UX in tandem with an at the time leading-edge product that provided automatic failover in case of hardware failures (ServiceGuard). These high availability home-built firewalls were in use until the market matured and ECMWF migrated to dedicated Nokia-Checkpoint devices.

The ECMWF relationship with IBM continued across a rapid evolution through the Power5, Power6 and Power7 generations, scaling to 46,000 processors by 2013 – each processor 30 times faster than the entire original Cray-1, resulting in the Power7 system delivering around 1.4 million times the performance of the original Cray-1.

The idea to create an HPC compute grid by interconnecting the European National Supercomputing Sites materialised in 2002 with the DEISA (Distributed European Infrastructure for Supercomputing Applications) project. This project interconnected ECMWF with sites such as MPI, CINECA, CSC, LRZ, JRZ, BSC, HLRS, EPCC, CNRS, and SARA[1] with features such as cross-mounted shared filesystems (based on a multi-cluster General Parallel File System – GPFS) and a Common Software Stack Environment. DEISA later evolved into PRACE, and PRACE evolved into the current EuroHPC Joint Undertaking, which provides considerable HPC resources for scientists across Europe.

1 MPI – Max-Planck-Institute; CINECA – Italian national supercomputing consortium; CSC – IT Centre for Science; LRZ – Leibniz Supercomputing Centre; JRZ – Jülich Supercomputing Centre; BSC – Barcelona Supercomputing Center; HLRS – High-Performance Computing Center Stuttgart; EPCC – Edinburgh Parallel Computing Centre; CNRS – French National Centre for Scientific Research; SARA – Stichting Academisch Rekencentrum Amsterdam.

2010s: Cray XC30 and the dragonfly network

From the late 1990s, the Intel x86 architecture found in desktop computing systems rose to dominance. At their peak in the middle of the 2010s, Intel CPUs accounted for almost 90% of global processor sales, and the economies of scale had made it almost inevitable that the IBM successor systems would be based on Intel technology. In 2014 ECMWF installed two Intel x86-based Cray XC30 systems. Each system had around 3,500 nodes, containing two Intel Ivy Bridge 12 processor chips with 64 gigabytes memory per node, providing a total of 168,000 processors across the entire service.

The number of processors per node sharing a common memory made it important that the first level of parallel programming exploited this feature. However, the large number of nodes put heavy requirements on the high-speed interconnect fabric. Cray developed its own proprietary interconnect, known as Aries. The Aries interconnect used a dragonfly topology to limit the number of connections, particularly the longer cables, as whilst the Cray-1 was contained in one small cabinet, each Cray XC30 cluster was housed in 19 cabinets with the complete system weighing in at 93,000 kilograms. The intra-cabinet and inter-cabinet dragonfly interconnect topology hierarchy added another level of operational complexity, as applications performed better if they could keep communications within a cabinet.

Broadwell upgrade

The modular design of the Cray system allowed a mid-term upgrade in 2016 to the next generation of Intel processors. Cray replaced all the processor boards in the system with Intel Broadwell chip boards. These had 18 processor cores, and consequently took the total processor count to 260,000 without changing the overall physical footprint of the system.

NOW AND INTO THE EXASCALE FUTURE: BOLOGNA, GPUs AND AI INTEGRATION

2020s: Atos BullSequana XH2000 in Italy

Fully operational since 2022, ECMWF's Atos BullSequana XH2000 system is the latest HPC service at ECMWF, the first hosted in the purpose-built Bologna data centre (see the section 'Designing data centres: from Reading to Bologna – a chronology of key milestones') and the first to introduce four operational clusters or "complexes" (Figure 5). Hosting over 1 million AMD EPYC Rome processor cores, using NVIDIA-Mellanox InfiniBand interconnect, the system delivers more than 26 petaflops of peak performance – equivalent to 167 million times the CDC 6600. The system weighs more than 200 metric tons, requiring 40 water-cooled cabinets for the main compute nodes, and an additional 40 racks of associated servers and storage. The entire system consumes around 40 million kilowatt hours of electricity per year, equivalent to ca. 15,000 UK homes. The solution is based on AMD Rome chips, each with 64 processors, giving more than a million processor cores in the complete system. Each processor delivers 27 times the performance of the first Cray-1, resulting in an increase of 28 million times the performance of the original Cray-1 (see Table 1).

The system includes two generations of NVIDIA graphics processing unit (GPU) hardware. GPUs were originally developed for manipulating arrays of data for displays. They use massive parallelism to perform relatively simple operations on vast quantities of data as it streams though the system. Exploiting the processing power of GPUs has been difficult, especially for complex models such as ECMWF's Integrated Forecasting System (IFS), but their capability has underpinned the surge in AI and data-driven models. ECMWF's interest in this technology has grown considerably over the life of the current system. Starting from having only two GPUs in the 2019 specification, we now have 128 NVIDIA A100 GPUs and 120 Grace-Hopper chips.

Figure 5: Two of the four clusters in the Atos BullSequana XH2000 system in ECMWF's data centre in Bologna, Italy.

Table 1: Comparison of ECMWF's Cray-1A and Atos BullSequana XH2000 systems.

	Cray-1A	Atos BullSequana XH2000
Year installed	1978	2022
Architecture	Vector processor	Massively parallel cluster
Number of cores	1	1,015,808
Clock speed (MHz)	80	2,250 (2.2 GHz)
Peak perf. per core (MFLOPS)	160	36,000
Peak perf. (MFLOPS)	160	26,687,760,000
Sustained performance (MFLOPS)	50	1,401,003,800
Memory (MiB)	8	2,146,304
Disk space (GB)	2.5	10,084,400,000

Towards heterogeneous HPC

The next phase of high-performance computing at ECMWF will increasingly be driven by data-driven machine learning. Data-driven models have recently surged in skill, but have hardware and software requirements that are very different from traditional HPCs. A hybrid future in which AI workloads are performed alongside the current, traditional HPC jobs will lead to significantly more complex HPC systems, blending CPUs, GPUs, and specialist accelerators, supported by tiered storage and intelligent scheduling systems.

LOCAL- AND WIDE-AREA NETWORKING →

In the late 1970s, external telecommunications for ECMWF's computing infrastructure were basic, as the organisation focused on development of its high-performance computing facility and expertise in numerical weather prediction.

With the growth of modelling capabilities and increasing collaboration with Member States and other meteorological centres, networking infrastructure became a strategic priority early on. Local-area networks within the data centre, and wide-area networks connecting ECMWF to its global partners, have needed to upgrade in turn to match these demands.

When ECMWF produced its first operational medium-range forecast in 1979, it was receiving input observation data in raw GTS form, copied to magnetic tape by the UK Meteorological Office in Bracknell and brought across by car to the Centre. By the late 1980s, connections to facilitate data exchange with national meteorological and hydrological services (NMHSs) were in place using X.25 packet-switching technology, the standard for wide-area networks at the time.

These early networks served two critical functions: collecting observational data from weather stations across Europe and beyond, and distributing forecast products to Member States. Data transmission speeds were extremely modest by modern standards, operating at 9.6 to 64 kilobits per second but, typically for ECMWF, represented cutting-edge technology for the era.

The 1990s marked a period of digital expansion, coinciding with the rise of the Internet and more sophisticated networking standards. ECMWF began to transition to TCP/IP-based networking (Transmission Control Protocol/Internet Protocol), enabling more dynamic and reliable connections. During this period, the Centre developed its Meteorological Archival and Retrieval System (MARS) (Raoult, 1997), which necessitated faster and more scalable data access for research data transfers alongside operational outputs. Collaborations with organisations like the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) also led to increasing volumes of satellite data being acquired and processed within the data centre.

The end of the 1990s witnessed an important development in ECMWF's networking infrastructure with the establishment of the Regional Meteorological Data Communication Network (RMDCN) in collaboration with the World Meteorological Organization (WMO). This service provides a managed, high-availability, private communication infrastructure between ECMWF and the NMHSs of its Member and Co-operating States, and WMO weather centres globally.

Before the RMDCN, the NMHSs exchanged observations using individual dedicated leased lines between each pair of organisations that required it. This resulted in hundreds of international leased lines being deployed at a high cost for the Member States. ECMWF also had at least one international leased line per destination for dissemination purposes. The RMDCN introduced the concept of a single managed network to replace all these individual links, providing a more reliable service and huge financial savings for Member States. A market survey showed that there was no single telecommunications contractor with a presence in all the required countries. The idea then came to build on top of the infrastructure already deployed by civil aviation for interconnecting European airports. This resulted in the contract being awarded to SITA (Société Internationale de Télécommunications Aéronautiques) with the different local public telecommunications operators (PTTs) providing the last mile connectivity between the NMHS home sites and their closest airport.

During its first few years of operation, the RMDCN frame-relay technologies required some very static configuration at the telecomms level for each network path. This evolved in 2006 to a fully meshed MPLS IP-VPN based network – the backbone technology of the Internet as we know it today. The RMDCN became a truly any-to-any solution, allowing any site to communicate directly with another site without involving ECMWF. The very high level of reliability and security assurance provided with the RMDCN has kept the service relevant, delivering the most essential observations and products to over 50 organisations globally – far beyond ECMWF's own remit (see Figure 6).

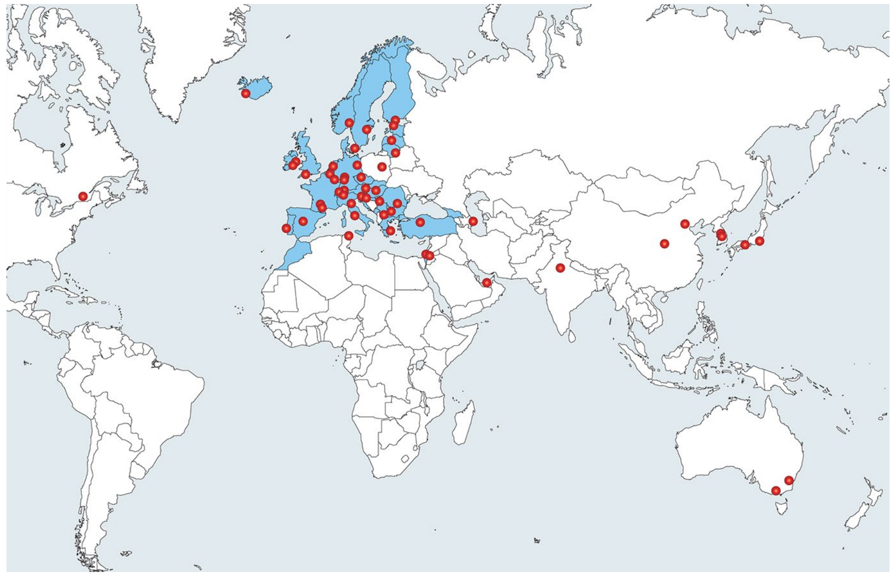


Figure 6: The Regional Meteorological Data Communication Network (RMDCN) connections at meteorological organisations around the world. The blue colour indicates ECMWF Member and Co-operating States.

As forecast models grew in complexity and output resolution, the need for faster and more scalable data centre networking became paramount. The HPC facility replacement cycles provided a natural cadence at which to also upgrade, ensuring the facilities remained free of bottlenecks and able to make best use of HPC resources. In the 2010s, ECMWF implemented 10, 40, and later 100 Gbps internal networks within its data centre to support HPC and storage systems, with 400 Gbps now on the horizon.

ECMWF has long been at the forefront of computer networking to support its high-performance computing and forecast delivery. In 2022, ECMWF began transitioning part of its operations to a new data centre in Bologna, Italy, featuring state-of-the-art networking and computing capabilities.

WMO Members now consider the Internet is sufficiently robust and mature to support safety-critical applications – so long as those applications are built in a fault-tolerant way. Many NMHSs around the world already use the Internet to support all or most of their time-critical and safety-related data exchange. ECMWF's own Internet connections now exceed 100 Gbps in aggregate.

Data exchange has come a long way in the decades since the late 1970s, from using magnetic tape and physical delivery, to today's time-critical product dissemination and popular Internet-based web services. Challenges for the future are not only in performance and reliability, but increasingly security. With a growing number of publicly accessible web services, ECMWF, like most organisations on the Internet, must face the growing information and cyber security demands of our current age.

DESIGNING DATA CENTRES: FROM READING TO BOLOGNA – A CHRONOLOGY OF KEY MILESTONES →

ECMWF has maintained world-leading supercomputing facilities for decades. The evolution of its data centre infrastructure – from the early days at Shinfield Park in Reading to the advanced systems now operational in Bologna – illustrates a continuous trajectory of innovation, resiliency and efficiency in support of meteorological research and operations.

1970s–1980s: foundations at Shinfield Park – early design and infrastructure

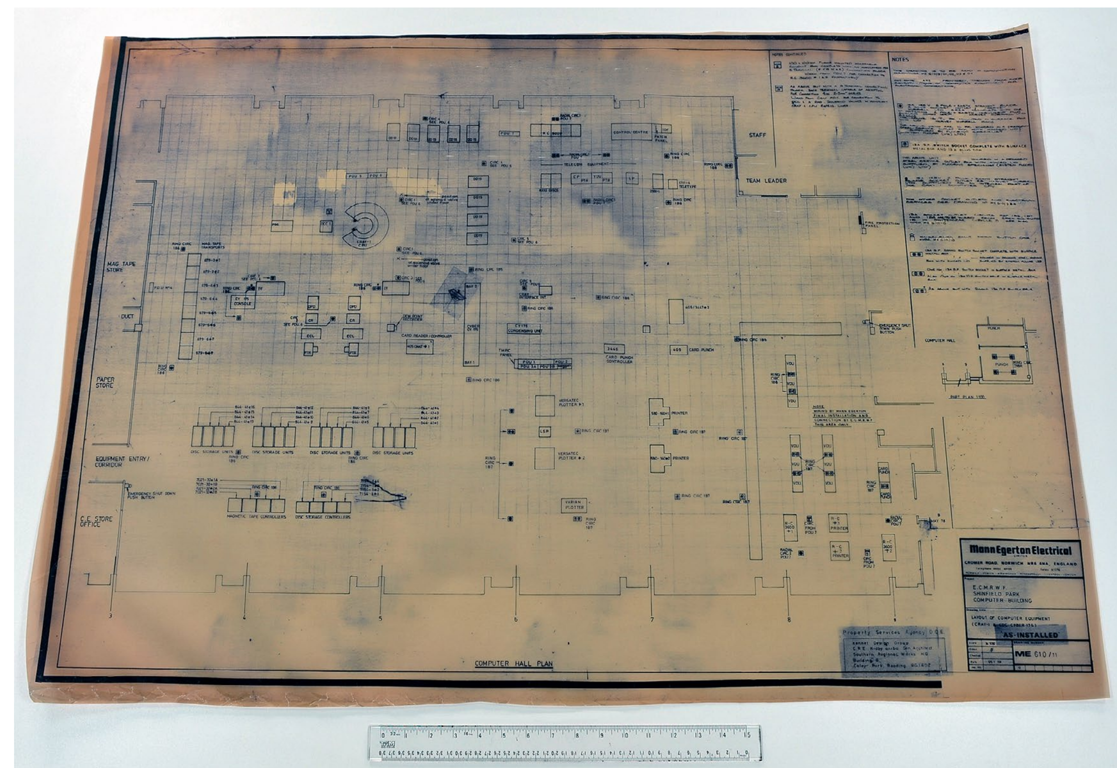
When the initial designs for Shinfield Park were developed, the site was provisioned with a 2000 kVA electrical supply. Original engineering drawings were created manually, using pens on drafting film, reflecting the standards of the time (see Figure 7).

The installation of the Cray-1A supercomputer posed unique challenges, as it was incompatible with the UK's standard voltage and frequency. To support its operation, an innovative engineering solution implemented a bespoke power delivery system, providing:

- 120 V three-phase at 400 Hz
- 120 V three-phase at 60 Hz and 50 Hz
- Standard UK voltage of 415/240 V for ancillary systems

Although not considered power-hungry by modern standards, the Cray-1A required 288 kW of power. Cooling was provided via three cooling towers at the rear of the site, supplemented by four chillers that also serviced the adjacent conference facilities. Environmental control was maintained by computer room air handling (CRAH) units positioned along the north and south walls which regulated the computer hall's internal climate.

Figure 7: The original layout of the computer hall with the Cray-1, dated October 1979.



1992: introduction of the energy centre

To boost power reliability, the energy centre was constructed and two diesel rotary uninterruptible power supply (DRUPS) systems—"KS1" and "KS2"—were installed, each delivering 1,000 kVA. These systems shared a 2000 kVA transformer and provided electrical continuity for mission-critical operations.

1996: transition to Fujitsu

The Centre transitioned from Cray to vector-based Fujitsu systems in 1996, requiring further electrical reconfiguration. These new machines specified different power requirements, necessitating transformer installations to adjust the supply voltage. Additional CRAH units were introduced to manage the increased thermal load in the computer hall, due also to the move away from water cooling to air cooling.

2002: beginning of the IBM era

In 2002, IBM systems were introduced, operating at 415 V/50 Hz, with expanded cooling demands prompting investment by the Centre in additional chiller capacity.

2004: enhanced fire protection systems

Responding to insurance mandates, multiple layers of fire protection were implemented:

- A water mist (FogTec) system was installed in offices and corridors surrounding the computer building, offering high-pressure mist-based suppression.
- Within the computer halls, a Very Early Smoke Detection Apparatus (VESDA) was installed. This system used pipes in ceiling voids to draw air samples into a detection chamber, where lasers could identify microscopic smoke particles – providing ultra-sensitive early fire detection.
- A gas-based fire suppression system using Inergen was installed in both the room and floor voids. On activation, gas was released through nozzles while wall vents opened to reduce oxygen levels, effectively extinguishing combustion. In the basement, diverter valves and 132 bottles of Inergen (stored at 300 bar) ensured system readiness.

Additionally, two more DRUPS units (KS3 and KS4) were installed in the standby house, increasing redundant power capacity to 4,800 kW.

2006–2010: computer hall expansion and upgrades

An extension to the computer hall was completed in 2006, adding 500 m^2 of data centre floor space. One new cooling loop was installed to support eight additional CRAH units. At the beginning of 2008, the existing high-voltage switchboard was replaced, by adding a temporary board outside and installing the new board in the old one's place with all the services maintained on temporary cables. New HV transformers and cables were added to increase the site's mains capacity and in 2010, the legacy DRUPS units KS1 and KS2 were decommissioned and replaced by KS5 and KS6, bringing the total uninterruptible power supply capacity of the site to 5,600 kW.

2008: return to water cooling with Power6 (High Density Computing)

The installation of the first IBM Power6 cluster marked a return to water-cooled computing. Each Power6 rack featured a redundant water-to-water heat exchanger and dedicated pumps circulating IBM's proprietary "Blue" cooling water directly to the chips.

To support this infrastructure specification, the following upgrades were performed:

- Three new chillers were installed on the far side of the building.
- As there was no direct access, a 3-metre-square tunnel was bored beneath the existing buildings to link to the new extension.
- Independent piping and pumping systems were installed, creating a completely isolated cooling loop for the Power6 systems.

2013: sustainability – free cooling implementation

The dry air cooler system was commissioned in 2013 to provide energy-efficient "free cooling". This allowed heated water from the IBM Power7 HPC systems to be cooled by ambient air through large radiator fans, dramatically reducing the reliance on mechanical chillers. As long as the return water temperature was at least 2°C above the ambient air, up to 2 MW of cooling could be achieved without using chillers.

As external temperatures approached the water temperature, fan speeds increased and three-way valves began to divert cooler water from the chiller loop to maintain temperature stability. This system recouped the £1 million construction cost within just 18 months of operation.

During this period, Chiller 4 was also upgraded – its screw compressor was replaced with a high-efficiency turbo core unit, adding an extra 350 kW of cooling capacity.

2014: water quality and cooling standards

In 2014, stricter water quality standards introduced by Cray necessitated a redesign of the cooling systems. Until then, a common water loop had been used for both computing and facility cooling. To meet the new requirements:

- Heat exchangers and dedicated pumps were installed to separate the loops.
- A full dynamic flush of the 70,000-litre system was conducted without interrupting operations, after water samples had revealed high microbial contamination, caused by the existing anti-corrosion chemicals feeding microbial growth. The system was then treated with a new passivation chemical.
- After flushing, biocide treatments were introduced, and a continuous monitoring programme was established to ensure proper water chemistry.

2020s: Bologna data centre

Purpose built to support the Centre's expanding operational and research workloads, the Bologna data centre combines leading energy efficiency, modular design, and climate-conscious infrastructure. The advanced multi-modal cooling construction strategies included:

- Groundwater wells
- Adiabatic cooling systems
- Backup mechanical chillers

At the heart of its efficiency is the multi-modal cooling system that adapts dynamically to ambient conditions. It leverages groundwater from deep wells and adiabatic cooling – where water is misted into air streams to enhance the cooling without mechanical chillers. This significantly reduces energy consumption and carbon footprint, especially during cooler months. When necessary, high-efficiency backup chillers maintain performance during peak load or adverse climate conditions.

Power reliability is ensured by a 10 MW DRUPS system, providing seamless transitions during outages without relying on traditional battery banks. Fire safety mirrors proven systems at ECMWF's Shinfield Park site, using inert gas suppression for zero-damage protection of high-value equipment.

In both architecture and operation, the Bologna data centre reflects the Centre's long-term commitment to environmental responsibility, operational resilience, and scalable computing in support of critical global weather and climate research.

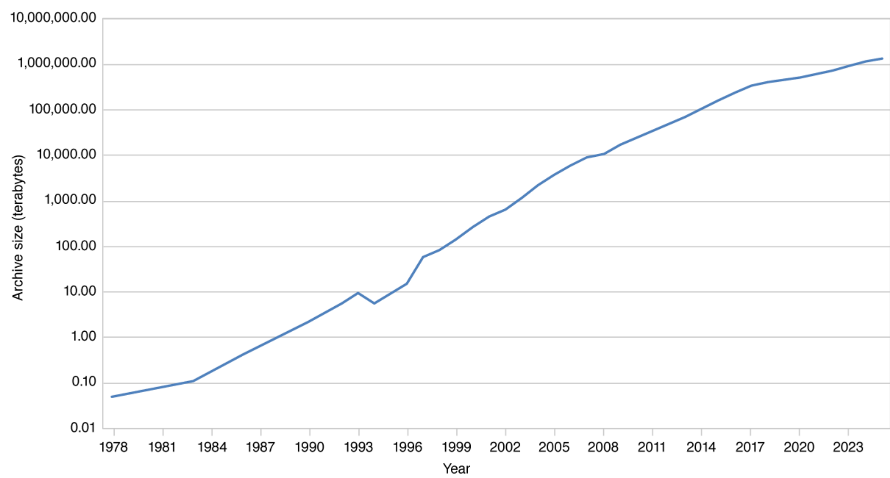
The evolution from manually drawn schematics and air-cooled Cray systems to water-cooled supercomputers and sustainable cooling strategies showcases ECMWF's dedication to staying at the technological forefront. With the Bologna data centre, ECMWF is well positioned to meet future computational and environmental challenges in the decades ahead.

CURATING 50 YEARS OF DATA: A TIMELINE OF STORAGE SERVICES →

1970s–1990s: the early evolution of ECMWF data storage

Numerical weather forecasting requires data, and consequently collecting and storing meteorological data is one of the objectives of ECMWF that was set out in the Convention establishing ECMWF. The Centre has seen transformative growth in data storage capabilities since its inception. Starting with just 50 gigabytes of storage in 1978, the archive had reached 1,300 petabytes by the end of June 2025 – a staggering 27-million-fold increase (see Figure 8).

Figure 8: Growth of data in the ECMWF archive 1978–2025 (logarithmic scale). In 1978, the archive held 50 gigabytes of data. At the end of June 2025, it held 1,300 petabytes, an increase of 27 million times.

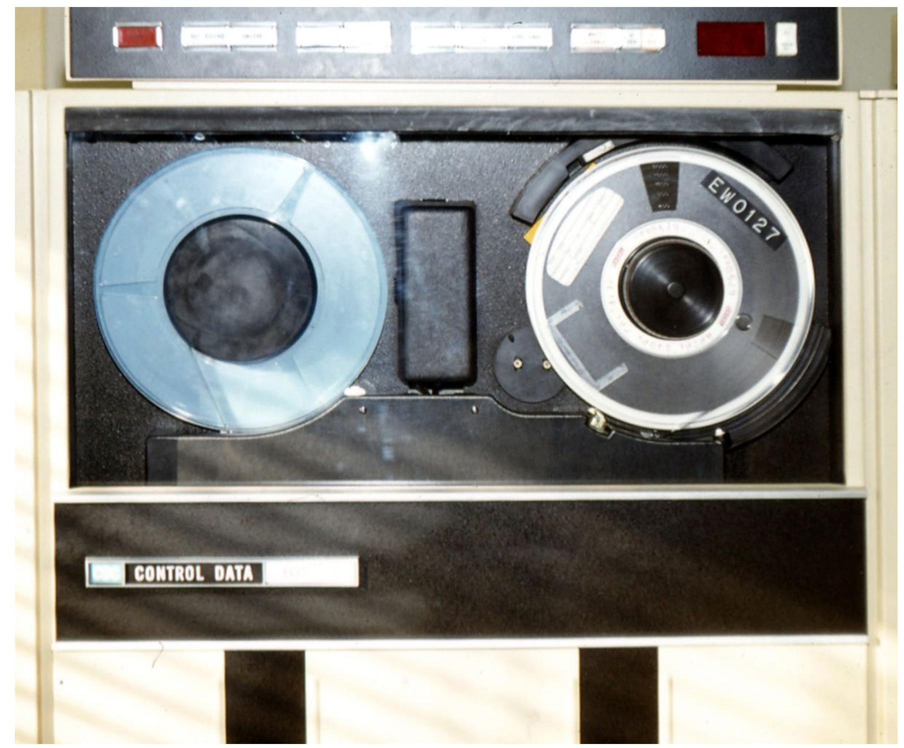


This exponential growth has paralleled advancements in high-performance computing (HPC), being symbiotically connected: HPC generates massive datasets, and the storage infrastructure preserves selected data for later retrieval by scientists. HPC systems are periodically replaced every five to seven years. Data, on the other hand, are perpetual, so the storing and protection are expensive and complex. Data need to be migrated when a new technology becomes available to improve capacities, performance and reliability and to control costs.

While disk and solid-state storage systems have provided the fast access to data, the bulk storage has always been magnetic tape. The challenge of tape is that while it provides excellent bulk storage capacity, it needs to be loaded on to a drive and accessed by passing the tape over the read and write heads. This process can be fast, but it is strictly linear, starting at the beginning and winding to the end of the used section of tape in order to write new data, or seeking the portion of the tape that contains the data of interest. The Scientific Advisory Committee in 1978 noted: "It is thought that the Centre will be able to cope for 3 or 4 years after the beginning of the operational phase by using high density magnetic tapes (and staging these to disk when necessary to assemble particular required files of data), taking into account the compression methods that will be used and the cuts in the archiving of processed data. However, such an organisation of the Centre's data bank, based on magnetic tapes in the way suggested, evidently cannot be a long-term policy." In the event, this model of data staged to disk from tape is exactly what ECMWF has continued to use to manage the growth and complexity.

The Centre initially employed Control Data 669-2 open-reel tape drives (see Figure 9). With 1,600 characters per inch (25 mm), each reel held around 40 MB across 2,400 feet (730 m) of tape, which required manual mounting, severely restricting scalability. By 1982, the system had grown to circa 1,000 gigabytes of data using approximately 15,000 tapes with 10 magnetic tape drives. However, the service provision had become cumbersome and human-intensive to support. Racks of tapes are shown in Figure 10.

Figure 9: A Control Data tape drive.



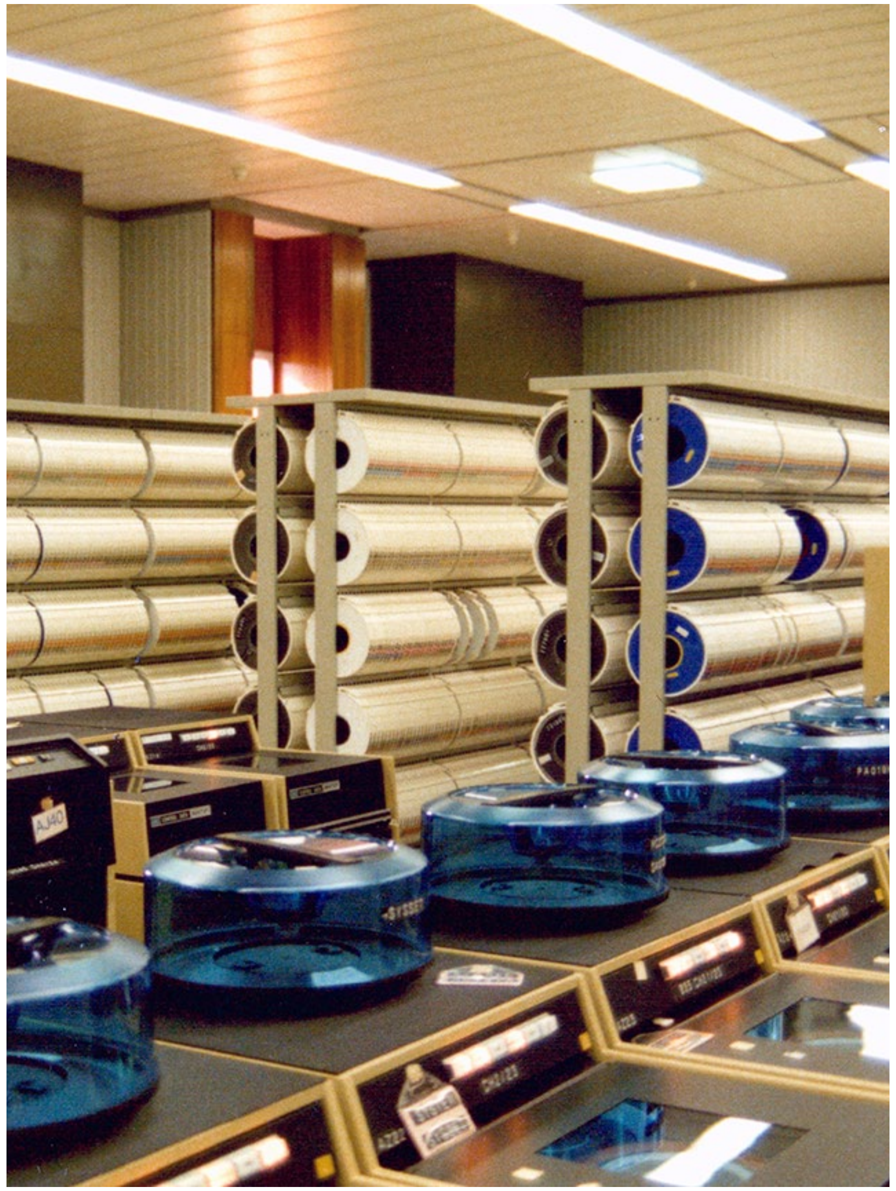


Figure 10: Racked tapes (background) and disk drives (foreground) in the Reading data centre. The blue cases on top of the disk drives were the outside packages for the removable pack, which was put into the drive unit below.

An IBM 3850 Mass Storage System (MSS) was installed at the Centre during 1983 and 1984. The Mass Storage System had a very unusual format. A group of cylindrical small cartridges, 50 mm wide and 100 mm long, each holding a 20 m spool of tape were stored in a hexagonal storage array (see Figure 11). Cartridges were moved by motorized accessor arms (see Figure 12). Each cartridge could store around 50 MB of data. Data was transferred to and from a host through IBM 3350 disk drives. This virtualised the MSS to the host, staging and de-staging data to cartridges in the background. It was the first example of “nearline” storage and pre-dated the Virtualised Tape Libraries (VTLs) that were to follow.

Around 1984, IBM introduced a family of single-reel cartridge tapes. The much smaller data cartridges (102 mm × 127 mm × 25 mm) allowed far more automation. The tape drive now threaded the tape out of the cartridge and over the read/write heads of the drive. ECMWF procured an 8-drive cartridge tape system in 1987, and by 1990 this system had 10 drives and 21,000 cartridge tapes.

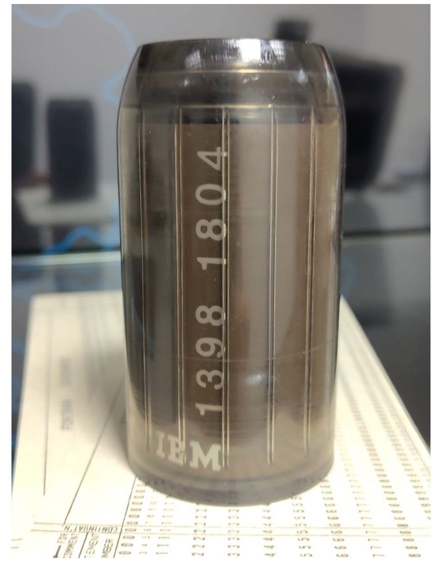


Figure 11: IBM 3850 Cartridge – the first use of helical scan reading and writing technology for magnetic tape.

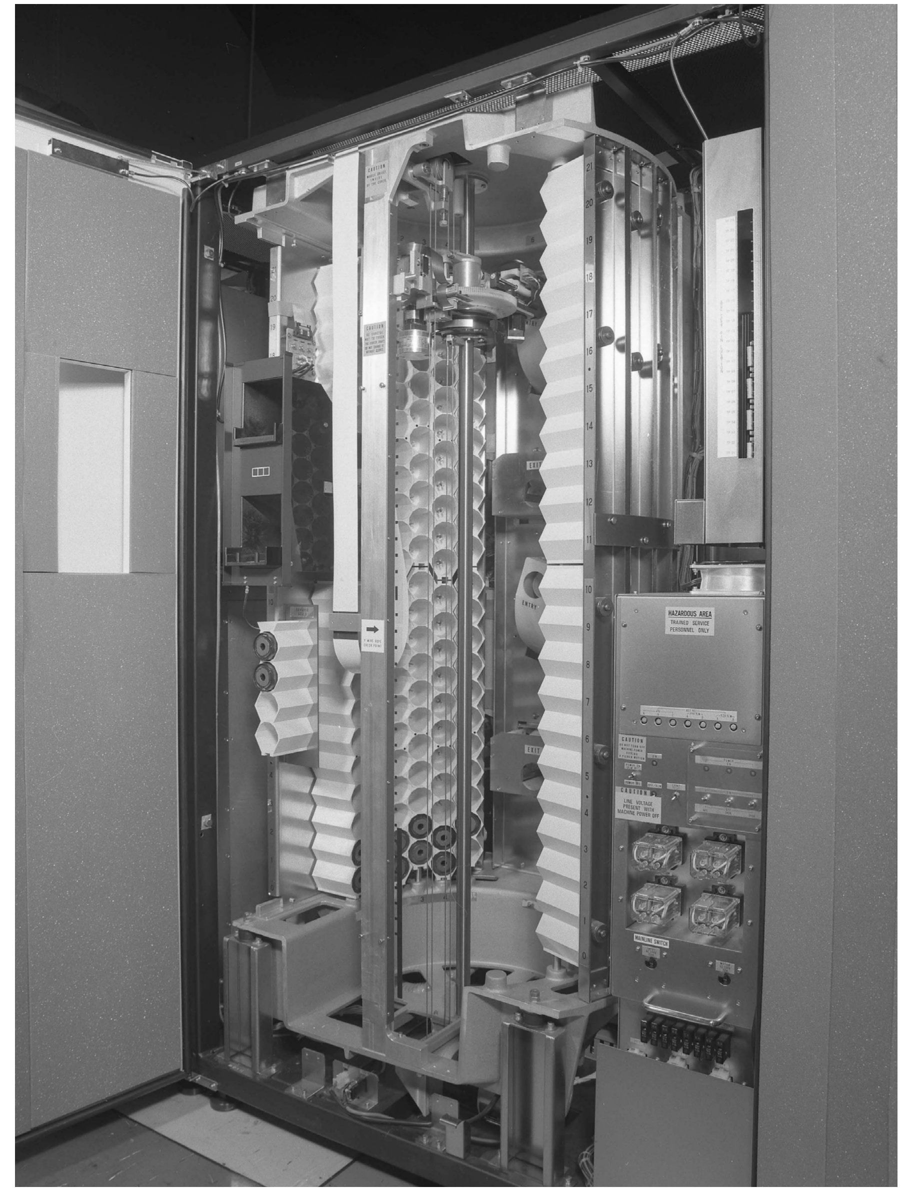


Figure 12: IBM 3850 Mass Storage System (MSS). © UKRI Science and Technology Facilities Council, available from https://www.chilton-computing.org.uk/

While the cartridge tape significantly increased the capacity of the archive and made handling the tapes much easier, it was not until 1991 that an automated tape library system was procured.

1980s–2020s: creation and development of the Data Handling System

From the early days of the Centre, it was recognised that it was strategically important to address the challenge of the ever-increasing amount of meteorological data. A Data Handling System (DHS) was created in three phases, to be the foundation of the service. A tender in 1982 resulted in the deployment of a Common File System (CFS) developed by Los Alamos National Laboratory (LANL), implemented on an IBM 4341 mainframe supported by IBM disk and tape.

As storage demands outgrew CFS scaling capabilities in the early 1990s, IBM’s ADSM (later Tivoli Storage Manager, TSM) replaced it.

Later, in a similar manner, it became clear that the TSM would not scale well into the future, and in 2002 a competition was run to find a replacement. HPSS (High Performance Storage System, a collaboration between IBM and a number of US Department of Energy laboratories) was chosen as the successor. Migration to HPSS started in earnest in 2003 with MARS and then the ECMWF File Storage (ECFS) using it in production in 2005. It provided a much better platform to scale out, where I/O movers could be added without affecting the performance of other areas of the storage. Some of the original structural decisions made in 2003 still hold today. The Centre will hold a new tender process later in 2025. However, where before it was clear a replacement was necessary, this time HPSS is still a viable option, scaling to the projected capacities into the future.

A key part of managing ECMWF's storage involves layering the services that users see on top of the automation systems, decoupling the complexity of managing the actual data storage from the use of the data. By using this design, the system is independent from the underlying hardware and software. The data can be physically re-organised without any impact on the system, allowing complex data management to be undertaken.

The Meteorological Archival and Retrieval System (MARS) was created in the early 1980s. It was developed by ECMWF to enable users to retrieve meteorological data via simple "pseudo-meteorological language" queries. It automatically identifies whether data resides on disk or tape, supporting seamless migrations and system upgrades. From managing 70 MB/day in 1985, MARS scaled to over 550 TB/day by 2025.

ECFILE and later ECFS was introduced to provide a file system interface to the users, again abstracting the details of the storage implementation from the user.

In 2022 the DHS was moved en masse to the new data centre in Bologna. Some new components – disks, Storage Area Networks (SAN) tape libraries and servers – were built ahead of time. Most of the existing equipment and tape media were packed up into flight cases and moved by road to Italy. This was done in several stages, but mostly within a concentrated four-week period. During this time, the UK-based disk systems, servers and tape libraries were rebuilt in the two data storage halls at the Bologna data centre. This was done without losing the operational MARS and ECFS services. Both had been primed with data to continue operations using a pure disk-only environment. Within three days HPSS was back and able to cope with the write workloads from the HPC. Over the following month, the remaining tape and disk data from the UK were reintroduced to HPSS to make the entire archive available again.

1990s–2020s: tape evolution, and automation

By 1991, ECMWF had introduced a phased deployment of automated StorageTek 4400 libraries – robotic systems managing up to 6,000 tape cartridges per silo (see Figure 13). The DHS progressively grew to contain five silos. Over time, the tape drives evolved from IBM 3480 (200 MB tapes) through various IBM release versions to the TS1130 drives in 2007, maintaining form factor compatibility but increasing tape density to 1 TB – providing a 5,000-fold increase in capacity in the same physical space.

In 2009, SL8500 libraries from Sun Microsystems (later Oracle) replaced these older StorageTek libraries, scaling capacity to 10,000 tapes. More importantly, the design allowed for up to 64 tape drives in each library, a necessary feature to support the particularly heavy access patterns of ECMWF. Between 2010 and 2015, tape capacity grew from 1 TB to 5 TB to 8.5 TB.

Figure 13: StorageTek 4400 nearline tape libraries (later rebranded to Powderhorn libraries).

Oracle's withdrawal from the tape market required ECMWF to pivot to IBM TS4500 and Spectra Logic Tfinity libraries (see Figure 14). ECMWF now operates 14 tape libraries with 740 enterprise-class tape drives and 80 LTO drives. These libraries manage nearly 1.275 exabytes across primary and secondary storage, showcasing both immense scale and complexity.

The Disaster Recovery System (DRS) also evolved, initially using Sony AIT tapes in a Grau AML/J library. This later moved to LTO drives within IBM TS3500 and TS4500 libraries. These systems serve as secondary storage for critical data, adding resilience and redundancy to ECMWF's storage architecture.

Figure 14: IBM tape libraries in Bologna. Photographer: Stefano Marzoli

2000s–2020s: disk storage, migrations, and looking ahead

The evolution of disk storage mirrored changes in tape systems. Early Serial Storage Architecture (SSA)-based RAID arrays gave way to Fibre Channel (FC) Storage Area Networks (SANs) attached systems. The first fibre-channel-based systems included IBM FAStT and DS series arrays, which later transitioned to V7000 and XIV systems. These newer models offered better performance and reliability.

From the mid-2010s until 2022, ECMWF deployed DDN SFA7700X and SFA14000 systems. After resolving a few initial implementation challenges these contributed to the delivery of a strong reliable service, only being retired when operations migrated to Bologna. The transition of infrastructure to Bologna enabled consolidation of the SAN resources, reducing complexity and improving manageability.

Migrations between storage systems – from CFS to ADSM/TSM and now to HPSS (High Performance Storage System) – have been monumental tasks. Each migration took over a year, requiring dual system operation and modifications to MARS and ECFS interfaces. A paradigm shift occurred in 2017 when HPSS moved from AIX to Linux, and the workload was distributed across over 100 x86-based Linux servers, compared to just 16 AIX servers previously.

MARS and ECFS have presented consistent interfaces for decades, while the underlying systems have continually evolved to ensure usability, performance, reliability, and scalability. Technologies like Ceph for distributed disk caching are being integrated, and future innovations, especially novel storage media, are being tracked to ensure the long-term future of the data archive.

WHAT WILL THE FUTURE BRING? →

The design of the next-generation HPC facilities must support an increasingly diverse mix of applications, from the traditional physics-based simulations and data assimilation to machine learning data-driven workflows. At the core of these systems will be heterogeneous compute architectures, combining general-purpose CPUs with high-throughput accelerators such as GPUs or AI-specific processors (e.g. Tensor processing units (TPUs), Graphcore Intelligence Processing Units (IPUs)). Partitions will be optimised for different workloads using flexible resource allocation to support simulation, data assimilation and machine learning inference and training tasks. These technology advancements will enable more accurate forecasts, global-to-local modelling, and faster scientific discovery, ultimately improving the ability to predict extreme weather events and long-term climate change.

By 2027, exascale computing will be widely available, enabling weather model ensembles to run at finer resolutions (4 km to 1 km scales) and supporting large-scale ensemble simulations to quantify uncertainty. AI accelerators will be deeply integrated into research computing platforms, enabling co-execution of machine learning models and traditional numerical solvers. These hybrid modelling systems will be critical for tasks such as emulating direct observation predictions, improving data assimilation cycles for near-real-time forecasting, training of ML models, and accelerating model runtimes.

To accommodate increasing demands for flexibility and collaboration, next-generation HPC systems will support cloud-native services and federated platforms. Containerised applications, workflow orchestration, and elastic scaling will allow ECMWF researchers and Member States to deploy and share models across national HPC centres and research clouds. This will foster a more collaborative and interoperable ecosystem, enhancing Europe's capacity to respond to climate and weather-related challenges with agility. These advances will also make it easier for users from diverse backgrounds – including policy, industry, and civil protection – to access and benefit from advanced weather forecasting tools and insights.

CONCLUSION

Over the past 50 years, the transformation of high-performance computing and data storage has revolutionised medium-range weather forecasting, enabling meteorologists to deliver more accurate, timely, and reliable predictions. From the early days of mainframe computing and magnetic tape storage to today's exascale, modular, cloud-enabled, AI-driven supercomputing environments, technological advancements have continuously pushed the boundaries of weather prediction.

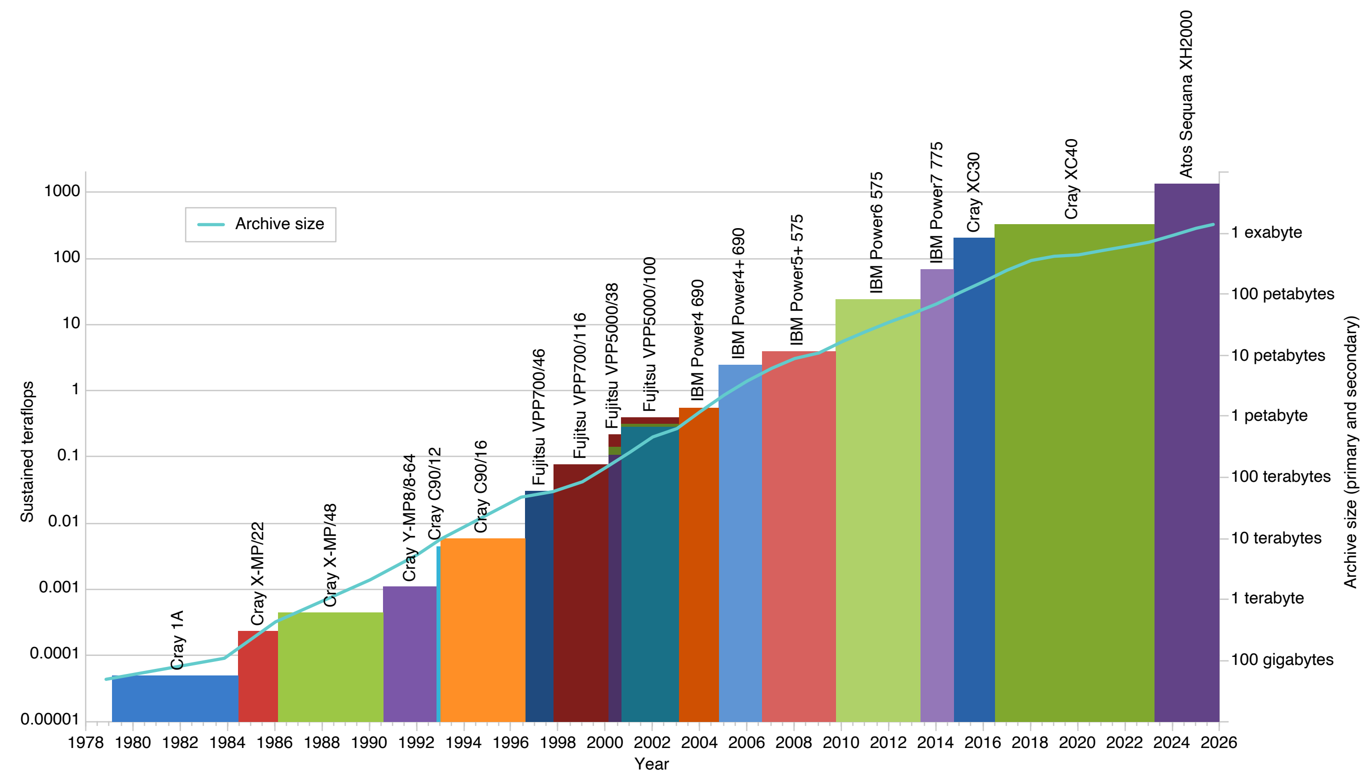


Figure 15: Evolution of ECMWF sustained HPC performance over the years with the data archive trajectory.

The shift from vector-based supercomputers to massively parallel architectures in the 1990s and 2000s allowed for higher-resolution models and the introduction of ensemble forecasting, which significantly improved uncertainty estimation. The ECMWF HPC service has increased in computational power by almost 30 million times from the original Cray-1 service in the 1970s (see Figure 15). Meanwhile, advances in distributed file systems and exascale storage technologies have enabled the handling of vast meteorological datasets, critical for data assimilation and model verification. For example, MARS has scaled from managing 70 MB/day in 1985 to over 550 TB/day in 2025.

These developments required substantial enhancements to the data centre environment to ensure the resilience and capacity to host these increasingly complex services – from mainframes that required specialised cooling through to tackling challenges related to creating carbon-neutral, sustainable data centres using renewable energy and latest technology advancements.

During the last decade, computing has witnessed major disruptive change: the emergence of commercial cloud computing has introduced flexibility and scalability, and graphical processing units have accelerated the growth of artificial intelligence and machine learning platforms. Such changes are transforming weather forecasting. AI-driven models can generate forecasts at lower computational cost, introducing the possibility of hybrid forecasting techniques which integrate traditional physics-based models with data-driven approaches. Service platforms now need to facilitate collaborative research and real-time model updates, provide cost-efficient resource allocation, and support meteorological institutions to adapt to growing data demands.

In conclusion, in just five decades, ECMWF has grown from renting time on a shared machine to operating one of Europe's powerful, time-critical operations supercomputing facilities. As data-driven science gains prominence, the Centre's next frontier lies in harmonising traditional physical modelling with machine learning applications, while working in a federated manner across Member and Co-operating States and partner institutions to share resources, expertise, and infrastructure in a truly collaborative European research ecosystem.

CONTRIBUTORS

ECMWF would like to thank the contributors to this paper:

Martin Palkovič, Christine Kitchen, Michael Hawkins, Stephen Richards, Andrew Gundry, Oliver Gorwits, Ricardo Correa.

REFERENCES →

ECMWF, 1975: *Newsletter* **No. 3**.

ECMWF, 1979: *Technical Newsletter* **No. 1**. https://www.ecmwf.int/en/elibrary/81039-technical-newsletter-no-1-february-1979

Moore, G. E., 1965: Cramming more components onto integrated circuits. *Electronics*, Volume 38, **Number 8**.

Raoult, B., 1997: Architecture of the New MARS Server, *Sixth Workshop on Meteorological Operational Systems*, 17–21 November 1997. https://www.ecmwf.int/en/elibrary/76107-architecture-new-mars-server

Find out more about ECMWF at 50

ADVANCING WEATHER SCIENCE
THROUGH COLLABORATION

ECMWF | 50