rescale

Energy Reality: Soaring Costs, Scarcity, and Sustainability

High-Performance Computing and real-world energy sustainability challenges

White Paper

Introduction

Geopolitical turmoil, dramatically higher energy costs, and renewed emphasis on sustainability are throwing every aspect of business operations into sharp relief. For companies running compute-intensive workloads such as simulations, the energy demands of data centres are under increasing scrutiny. Can high-performance computing contribute to a greener world?

Energy Reality: Soaring Costs, Scarcity, and Sustainability

The war in Ukraine shattered the European – and possibly the global – geopolitical landscape. Alongside the violent catastrophe of war, energy costs soared.

Yet even in a future where supplies may be volatile in price, availability, and volume, the demand for energy will, of course, continue. In this new world, what steps can consumers and businesses take to minimise power usage while maximising efficiency and outcomes?

Facing the new realities, the term 'energy sobriety' is emerging in France, in some ways similar to the earlier German "Energiewende," reflecting the desire to think more deeply about consumption from all aspects, emphasising best use of a precious resource.

Global data centres, particularly HPC data centres, are naturally coming under increasing scrutiny. Engineers now not only look for the most-efficient way to run workloads, but also at the efficiency and sustainability of the data centre itself. For HPC tasks, a new energy equation is taking shape, influenced by a complex mix of compute performance, cost, sustainability, and price.

To tackle these challenges, new tools are emerging that enable HPC users to assign workload to the platforms best-suited to individual tasks. These tools will help to meet sustainability targets even as users reduce costs, improve throughput, and cut energy consumption.

As Europe and the world adjusts to the new realities of war, HPC will also respond. Even though the commercial imperatives will continue to apply, green data centres and sustainable computing will now play a central role as we reach out for energy sobriety.

Capacity Squeeze and Price Hikes

War has revealed the weakness of relying on a dominant gas and oil supplier, and Western European countries have embarked on an unprecedented scramble to realign energy strategies. For example, in 2021 Germany relied on Russia for 49% of its gas, yet wants to end all gas imports from Russia by mid-2024. For neighbour Austria the position is even more extreme, relying on Russia for 86% of its gas.

As the prices of primary fuels such as gas and coal have soared, so too has the cost of electricity. Europe generates around 19% of its electricity from gas and 15% from coal. Gas prices – and thus electricity prices – tripled in the six months February-August 2022, and ten-fold since January 2021.

Yet for many years, activists and politicians have been driving a sustainability revolution, arguing for urgent transition away from fossil fuels to wind, hydro, and solar. Caught up in the push for reduced environmental damage, nuclear power ambitions have been dramatically curtailed, and in Germany have been abandoned completely.

However, even though nuclear power accounts for 26% of electricity generation in Europe, it is being scaled back at precisely the moment when other capacity is being squeezed, either for environmental or price reasons.

Regardless of the political situation, the environmental reasons for turning away from nuclear power remain unchanged. Extinction Rebellion (UK), Letze Generation (Germany), and Dernière Rénovation (France) will continue to argue that ruination of the planet and the demise of humanity are still inevitable.

The result of these twin forces is that the energy supply side – fossil fuels and nuclear – of power generation will not change, and high energy prices are likely to be the 'new normal.'

In the same way that the COVID-19 pandemic brought forward many changes that were already waiting in the wings – remote working, video conferences, cloud collaboration – the war has accelerated changes that had similarly been poised to occur: reduced consumption, increased efficiency, and the shift to renewables.

Data centres are not immune. Variously estimated to account for between 1% to 1.5% of global electricity consumption, there is a renewed drive toward efficiency.

And energy-intensive High-Performance Computing is under particular pressure.

Balance of Power

In 2011, Chancellor Angela Merkel declared that Germany would close its last nuclear power station by the end of 2022. The announcement followed the tsunami in Japan that devastated the Fukushima nuclear power station. By contrast, on Monday 26 September 2022, Vice-Chancellor and Economy Minister Robert Habeck partially reversed the policy, and stated that at least two of the country's nuclear-powered stations would continue until April 2023. Mr Habeck is – or was – a prominent environmental campaigner, and a former chairman of the Alliance'90/The Greens group in Schleswig-Holstein, the northernmost German state.

By comparison, France operates 56 nuclear power reactors (second only to the US, with 92, and ahead of China at 54). President Emanuel Macron announced on 10 February 2022, two weeks before Russia invaded Ukraine, that France should build six new reactors by 2028, and a further eight by 2050, with the first station commissioned by 2035. France is also committed to multiplying solar energy capacity tenfold, and building 50 offshore wind farms.

Nuclear power takes significant time and investment. For example, the UK government approved the new Hinkley Point C power station in 2008. After design, the build is expected to take around ten years to complete, finally generating 3.3 MW by 2027.

Friends of the Earth, founded in 1971, has consistently opposed nuclear power, presenting cogent arguments around waste disposal, safety, the environment, and – more recently – comparative cost. Though renewables do not provide baseline capacity, the investment cost and time-to-market are much lower: the Hornsea 1 & 2 wind farms on the UK coast produce around 2.5 GW at full capacity, and took four years to build.

Continuing the no-fossil-fuels agenda, governments of all persuasions continue to support the 2015 Paris Accord and resolutions adopted by COP26 in 2021. Seventeen weeks before the Russians invaded Ukraine, the UK published its Net Zero Strategy in October 2019, setting out policies and proposals for decarbonising all sectors of the UK economy to meet net zero by 2050. There are no immediate signs that any of these commitments will change.

Market turmoil calls for a profound re-examination of energy use. High-Performance Computing is not exempt.

Energy-Efficient Simulation

> High Performance Computing data centres are expected to grow at a CAGR of 8.9% from 2022-2032, reaching $30.1 billion.

Simulations are the most-common workloads for High-Performance Computing data centres. Simulation growth is driven at least in part because it is so much more cost-effective than crashing cars, flying aeroplanes, and stress-testing bridges. For example, the largest and fastest wind tunnel at the NASA Glenn Research Center consumes up to 200 MW when at full power. While simulations cannot replicate every single aspect of real-world interactions, even the very largest individual server – such as an IBM Z mainframe – consume less than 5 kW, three orders of magnitude less than the wind tunnel.

Naturally, simulation also offers massively enhanced product design, with the ability to test, adjust, re-test in thousands of iterations. For example, in the period 2005-2017, air travel fuel consumption per passenger per kilometre reduced by 24%, thanks to improved aerodynamics, propulsion, and materials. Every tiny design change was flown and tested in virtual wind tunnels for thousands of hours before committing to the time, cost, risk, and environmental expense of a real-world flight.

At the same time, computing efficiency has skyrocketed. Over ten years 2010-2020, performance measured by Floating Point Operations Per Second per Watt (MFLOPS/W) has increased by a factor of up to 2,000.

But performance per watt alone misses many critical sustainability issues, such as the processor utilization, lifecycle cost of manufacture and disposal, and the energy costs of data centre climate control. Being able to combine these factors to create a unified view is becoming significant, not least because sustainability is now a reporting requirement for public companies in many jurisdictions.

The International Sustainability Standards Board (ISSB) is a new standard-setting board created in November 2021 by the IFRS, alongside the International Accounting Standards Board (IASB), which defines global accounting principles. IASB standards have been largely adopted worldwide, and it seems likely that governments and corporations will adopt ISSB standards, too. For example, Japan has members on the ISSB board, and in September 2022 the International Cooperation Forum and Meeting of African Ministers of Finance, Economy and the Environment in Egypt expressed support for the ISSB initiative.

The new equation for HPC sustainability therefore includes balancing the commercial imperatives of cost and time, compute efficiency, power consumption, lifecycle impacts, and regulatory reporting.

Cloud Means Clear Skies Ahead

The meteoric rise in cloud computing has transformed HPC possibilities. On-premises systems of fixed capacity have had their day, replaced by on-demand resources from Amazon, Google, Microsoft and many more.

Previously it was important to keep on-premises systems fully utilised to ensure greatest return on investment, and the consequence was workload queuing. Presuming that simulations deliver competitive advantage; delays have a direct impact on commercial performance.

The flexibility offered by cloud computing has solved the utilisation question. Systems can be created, sized, and configured for specific task, and fully utilised. In the data centre, the nature of shared workload means that the systems are collectively run at high utilisation rates, which is also in the cloud services provider's commercial interest.

Different workloads may run optimally on specific platforms, too, and cloud computing creates the ability to direct jobs to the most appropriate service. Rather than using the all-things-to-all-people on-premises system, users can pick and choose the processor, operating system, bandwidth, storage and more, designed to optimise performance and accelerate results delivery.

In that process it may be more energy-efficient to select a system configuration that consumes greater absolute power for a shorter period, or perhaps find a low-power solution that runs for longer. Either way, assessing the total energy consumed is a key metric when calculating, for example, the carbon cost of a simulation task.

For organisations with very large HPC workloads, it will also be important to include total data centre power consumption and lifecycle calculations, and report on sustainability as an integrated metric. Data centres located where ambient temperatures allow power-hungry air-conditioning systems to be replaced by natural cooling – see boxouts – show impressive results, which will also affect sustainability outcomes.

Further, data centres that consume less total power will act as an important protection against possible future energy shortages and price increases; security of supply may be partly dictated by type of supply, such as renewables.

Advanced cloud HPC management solutions – see boxout on Rescale – offer unified support for both compute orchestration and sustainability analysis. Only by adopting an integrated approach to both computing and sustainability will enterprises be able to maximise both commercial advantage and environmental responsibility.

In recent years, consumer-facing industries such as fashion, food, autos, and aviation have experienced considerable green pressure to improve sustainability. For HPC, the same pressures are now building, including new standards, potential legislation, the energy crunch, and environmental reporting.

The competitive and commercial balance is now swinging fast towards a new, sustainable, equilibrium. Compelled and accelerated by geopolitics, and enabled by dramatic rises in computing power and orchestration capabilities, cloud offers the right path for next-generation HPC.

RESPONSIBLE COMPUTE
YOUR GREEN PARTNER IN THE NORTH

Iceland and Borealis

Borealis Data Center (BDC) builds and operates sustainable data centres using renewable green energy. Located in Iceland, Borealis offers year-round natural cooling, between North America and Europe, two of the world's key markets.

In August 2022, BDC and IT service provider Origo announced a new joint-venture company, Responsible Compute, which will deliver sustainable HPC solutions in partnership with Rescale.

Responsible Compute will allow Rescale to provide customers with unprecedented access to carbon neutral computing allowing them to run massive compute workloads, powered by hydro and geothermal energy, in a responsible and sustainable way. Responsible Compute will deliver carbon-free solutions along with market-leading emissions reporting and world-class performance.

Microsoft

Microsoft and Project Natick

Project Natick is a Microsoft research project to investigate the potential benefits that a standard, manufacturable, deployable undersea datacentre could provide to cloud users all over the world.

The Natick Phase 1 vessel was operated on the seafloor approximately one kilometre off the Pacific coast of the United States from August to November of 2015. The Phase 2 vessel was deployed at the European Marine Energy Centre located in the Orkney Islands, UK, in June of 2018.

The project reflects Microsoft's search for cloud datacentre solutions that are less resource-intensive, and that offer rapid provisioning and lower costs. Future research will explore powering a Natick datacentre by a co-located ocean-based green power system, such as offshore wind or tide, with no grid connection.

Google Cloud

Google and Carbon Emissions

Google Cloud Computing (GCP) offers a series of energy data reporting tools, including metrics for the percentage of carbon-free electricity, grid carbon intensity, and net operational greenhouse gas emissions (read more here).

GCP users can access carbon footprint reporting tools for their cloud account. For example, the dashboard contains four key reports; gross monthly carbon emissions, and by project, product and region, alongside high-level summaries for yearly and past-month footprint. Google aims to run its data centres on carbon-free energy, 24/7, by 2030.

aws

AWS and Customer Carbon Footprint

Amazon Web Services (AWS) enables users to track, measure, review and forecast carbon emissions. The summary provides carbon savings when compared with an equivalent on-premises workload, both by infrastructure and by energy source and geography.

Based on data from 451 Research, AWS estimates that its cloud services can reduce carbon footprint of a specified workload by 88%, rising to 96% when AWS is powered with 100% renewable energy (read more details here). AWS expects to reach the renewables target by 2025.

Rescale Platform

Rescale builds cloud software and services that enable organisations of every size to deliver engineering and scientific breakthroughs that enrich humanity. Rescale offers high-performance computing and the orchestration and reporting layers that fully empower engineers while ensuring very high levels of security and control for IT managers.

With solutions and services for a wide range of industries, with uses cases from Applied AI/ML to Workflow Automation, the Rescale platform enables the benefits of cloud HPC for every workload type. Rescale enables automation and optimisation of end-to-end IT and R&D workflows, with outcome-based SLAs and full-stack support.

From the foundation through enterprise and government-specific editions, the platform delivers optimal cloud HPC management, including AI-supported configuration and full sustainability management.