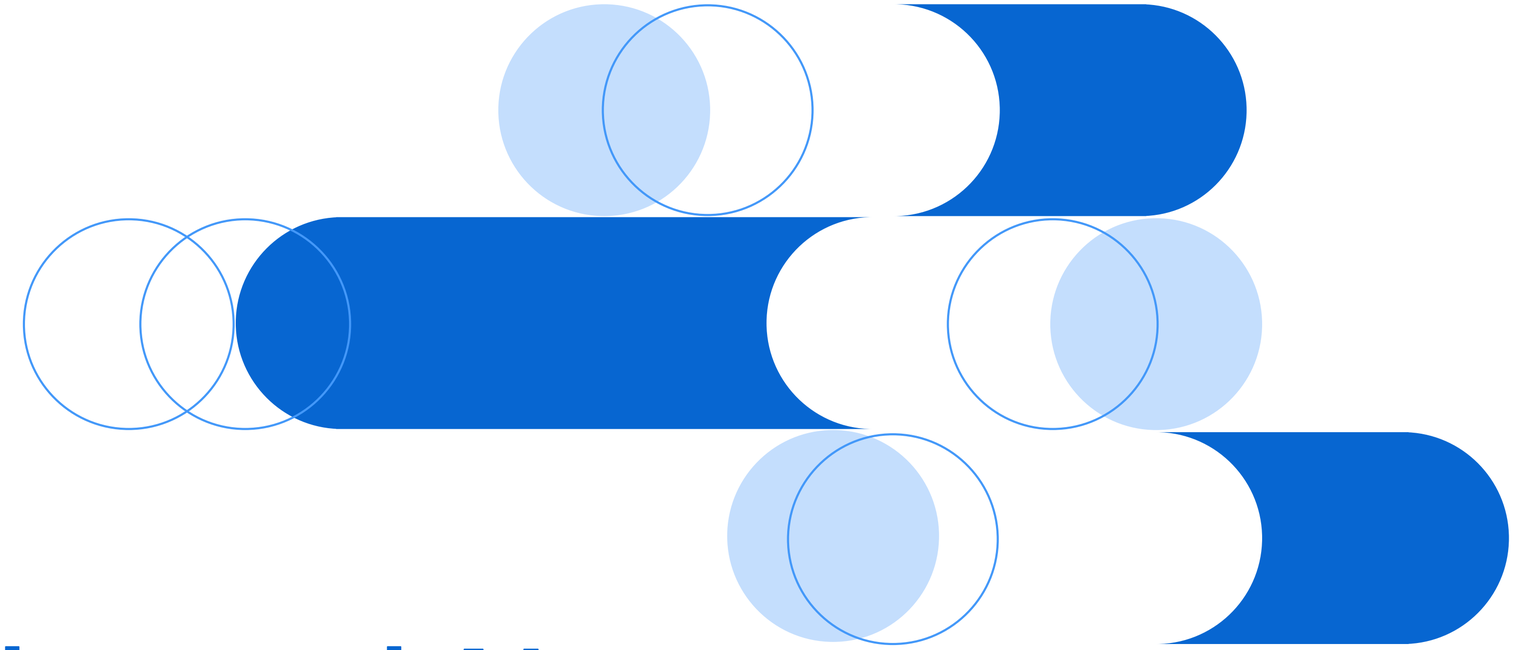


Rethinking data and AI in the cloud: How to move to a sustainable future

Contents

[1] Source: The Verge
[2] Source: UN
[3] Source: PwC
[4] Source: Guardian

Foreword

By Jerry Williams, Chief Environmental Officer, SAS

In the past two years, we've seen generative AI tools catapult the technology into the public realm – with ChatGPT becoming the fastest growing internet app of all time, attracting 100 million new users in only two months.[1]

So commonplace has the technology become that it's now difficult to remember a time when it didn't exist, and both business and political leaders alike recognise its potential to drive innovation and economic prosperity.

However, the massive computational power that AI demands is a growing environmental concern – especially as the 2030 deadline to cut global emissions by almost half draws nearer.[2]

As well as the ethical imperative, businesses are under growing pressure to prioritise ESG – not only to comply with relevant regulations but also to attract investors. Almost 80% of them say that ESG is an important factor in their investment decision-making,[3] so firms must scrutinise every part of their operations, including AI in the cloud, to secure the funding they need to survive and grow. In both the EU and US, there are signs that regulators will expect the AI industry to be transparent about its energy consumption.[4]

Demand for generative AI tools has increased the use of GPUs (graphics processing units) which are very energy intensive. This should not distract us though from the power consumption needed for the typical data and AI workloads, which use a large amount of CPUs (central processing units) to perform efficiently.

The cloud hyperscalers, Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP), are continually working to cut energy consumption in their data centres. We're also seeing new generations of CPUs from NVIDIA, AMD and Intel among others that can dramatically reduce energy requirements, even for generative AI applications.

It's not only up to hardware providers and hyperscalers to make a difference. It's also up to the individual companies that use data and AI workloads in their daily operations to optimise their processes. The cumulative impact of this will be significant.

Greater efficiency in AI model development, made possible by the availability of data and AI platforms that are optimised to run in the cloud, will also help teams to reduce unnecessary duplication and waste, and minimise energy consumption. Low and no-code platforms that boost productivity also help organisations to reduce the time spent using compute resources in the cloud, which is equally important for minimising the environmental impact.

Embedding sustainability into the culture of an organisation leads to better decision-making around everything from model development and governance, to which technology partners to use.

With insights from industry experts, this report aims to uncover the biggest sustainability challenges around AI and analytics in the cloud – and how they can be addressed.

Executive summary

This report will discuss:

The growing problem of data consumption

We explore the scale of the problem facing decision-makers due to exponential demand for data. Industry figures demonstrate how much cloud computing is contributing to carbon emissions – but we also look at what the cloud hyperscalers are doing to address these problems. Finally, we delve into the problem of cloud proliferation in organisations, resulting in inefficiencies and excessive computational processing.

Senior decision-makers in large organisations are grappling with two challenges – they need to harness data and AI on an ever-increasing scale while also urgently reducing carbon emissions.

They not only have an ethical imperative to mitigate the impact of climate change but a commercial and regulatory one too. Customers and end-consumers are increasingly scrutinising the activities of organisations, including their supply chains, and more of them are now legally required to report on sustainability.

As it stands now, many decarbonisation initiatives could be neutralised as data hungry technologies, including AI, are zealously adopted. Instead of reducing their use of these technologies, leaders want solutions that allow them to innovate in the most sustainable way possible. We spoke to three experts, working in different areas of data, to find out what steps organisations can take to make the most of data-led technologies, while also reducing their carbon footprint.

Executive summary

Expert view: Yves Mulkers, Data Strategist at 7wData

Yves Mulkers looks at some of the ways organisations can optimise their IT infrastructure and reduce energy consumption – including virtualization; using AI to manage data and refine large language models. Acknowledging that sustainability is a big challenge for decision-makers, he underlines the importance of working with partners who are continually driving technical innovations, and who share their values.

Expert view: Spiros Potamitis, Senior Data Scientist, SAS

As addressing climate change becomes more urgent, SAS' Spiros Potamitis explains how organisations could drastically reduce carbon emissions by adopting environmentally friendly cloud practices. He urges them to use monitoring tools to understand their usage, build holistic ESG dashboards, use tools to assess the carbon footprint of their data & AI workloads and partner with software and cloud vendors who are making continuous improvements to operate more sustainably.

Expert view: Luke Davies, Managing Director of Datacenters at GlobalConnect

Luke Davies takes us through some of the big innovations taking place in the design and development of data centres, including better cooling techniques – as well as the small steps operators can take to bring down energy consumption.

A plan for achieving sustainability in the cloud

We look at the five practical steps organisations can take to optimise their cloud infrastructure, and improve the energy efficiency of their operations:

Step 1: Create a sustainability culture

Step 2: Understand the environmental impact of computational workloads – and make continuous improvements

Step 3: Tackle cloud proliferation

Step 4: Take a strategic approach to model development with a low/no/yes code platform

Step 5: Choose the right partners

Big data, big problem?

The rapid and widespread adoption of AI, including data-hungry generative AI tools, in recent years has raised a number of ethical questions – one of the most important of these is how to minimise the environmental impact as models become ever more powerful.

One report suggests that ChatGPT 3 consumes the equivalent of 500ml of water for every straightforward 20-50 questions answered (and ChatGPT4 consumes even more).[5]

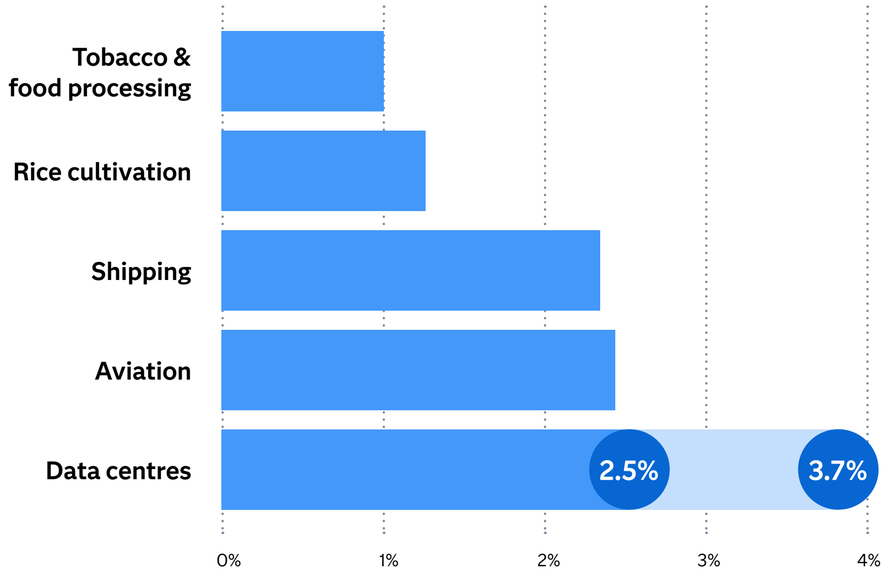
Elsewhere, data experts at Digiconomist have speculated that if Google were to power its entire search engine operations using AI, it would require as much electricity as is needed to power the Republic of Ireland.[6] Others have warned that the AI industry could use the same amount of energy as the Netherlands by 2027, unless growth slows down.[7]

Data centres are well-known to be resource-intensive, both directly in terms of water and energy consumption, and indirectly in terms of construction, equipment and land use.

Global emissions from cloud computing range from 2.5% to 3.7% of all global greenhouse gas emissions, thereby exceeding emissions from commercial flights (about 2.4%) and other existential activities that fuel our global economy.

Global cloud computing emissions exceed those from commercial aviation

Share of global CO_2 emission generated by sector/catagory



Graph reproduced using analysis from Climatiq[8]

[5] Source: Cited in Gartner
[6] Source: Mirage News
[7] Source: BBC News
[8] Source: Climatiq

Yet a slowdown in AI growth would be a step backwards, dampening benefits like efficiency, productivity, and innovation – just as organisations are starting to realise them. For a start, the ability to make sense of ESG data, through real-time monitoring and reporting using AI, ensures better reporting, compliance and sustainability strategies. Not only that but AI, and other data-driven technologies are also key to many ambitious projects aiming to address environmental, social, and humanitarian challenges.

The tech industry recognises the problem and we've seen a wave of innovations in data centre design and management in response to growing digitisation.

Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) have made good progress in improving sustainability – which means organisations migrating to the cloud from on-prem or traditional data centres reduce their environmental footprint. Microsoft, which has committed to becoming carbon negative by 2030, claims that using Azure is up to 93% more energy-efficient than using a traditional enterprise data centre.[9]

Organisations will rely on innovations in data centre design and operation to meet their own sustainability objectives, and ESG reporting requirements. Some are already scrutinising their cloud and software vendors, like any other supplier, and factoring in carbon reduction to their decision-making.

But is it enough? The IEA warns that emissions would need to drop by half by 2030 to be in with a chance of achieving net-zero admissions by 2050.[10] Growing global demand for AI could be a serious barrier to achieving this ambition. Sam Altman, CEO of OpenAI, the creator of ChatGPT, recently said that an energy breakthrough, including nuclear fusion and cheaper solar power and storage, is needed because future AI will consume even more power than was previously expected.[11]

Whether or not we'll see a once-in-a-generation shift in energy supply, of the kind that Altman posits, is unclear. And while the cloud and software vendors play a crucial role in reducing emissions, there are steps that organisations can take to optimise their own IT estate, and limit its energy requirements.

One of these is addressing cloud proliferation – a growing problem due to organisations adopting a mix of public and private cloud platforms over the years.

Our own research found that cloud and analytics sprawl is causing problems for 99% of all the large organisations we surveyed.[12]

Accuracy, efficiency and rising costs for infrastructure, storage and processing/computing power are the biggest issues they experience – but another often hidden problem is the environmental impact because of the huge amount of processing power consumed by data and AI workloads, especially where data has to be moved from one location to another. The slow creep of junk data, under-utilisation of features and storage, and duplicating workloads in different places can all increase energy requirements, which is why optimising cloud environments is critical in reducing both financial costs, and potentially, carbon emissions.

As CIOs and CTOs start to evaluate their cloud strategies and infrastructure, there are key questions they need to ask around measuring and communicating the value of the cloud, and whether they have full visibility of their workload inventory.[13] This process should also shed light on where efficiencies could be made that ultimately reduce computing power.

[9] Source: Microsoft
[10] Source: IEA
[11] Source: Reuters
[12] Source: SAS
[13] Source: CIO

Industry insight 1

Yves Mulkers, Data Strategist, 7wData

Yves Mulkers, Data Strategist at 7wData, talks to us about why data optimisation and sustainability should be a part of every digital transformation strategy, and how the right partnerships can help organisations manage data more efficiently.

What can organisations do to measure and monitor the environmental impact of their data estate (the infrastructure used to manage data[14])? Are there any key metrics they should include when making their assessment?

Energy consumption, carbon dioxide emissions, waste, and water usage are important data points to create key metrics (such as power utilisation efficiency) to include when looking to reduce your impact on the environment.

But we shouldn't consider the data in isolation because, while it might increase energy consumption, it can improve decision-making and lead to a net reduction in carbon emissions. For example, a supply chain business might use data analytics to optimise transport routes so there's an overall improvement in sustainability.

Should data optimisation – and sustainability – be part of every digital transformation strategy?

There's no doubt that data optimisation and sustainability should be a part of every digital transformational strategy because they share common goals. Optimising your data means you can make faster and better decisions, improve your data quality and work efficiently, which can also improve the sustainability of your operations.

In my experience, sustainability in data management is only now starting to become a priority. However, it's been a hot topic in the design and operation of data centres for some time, which is why moving to the cloud is a wise move because the data centres take care of many complex processes.

There will always be challenges because sustainability is complex – and it's not solely down to the technical team to solve the problems. Technology is constantly evolving and has the potential to vastly improve people's lives as well as be profitable for companies. But organisations also need to think about whether the benefits can be justified if there's a heavy impact on the environment.

Raising awareness and changing mindsets can help to bring about meaningful change. Environmentally friendly practices should also be promoted within organisational teams or stakeholder groups so sustainability is a strategic priority at the C-level from the start. Unlike servers and other hardware, data isn't tangible to many people, so we need to find ways to visualise it and make the environmental impact clear.

[14] Source: Microsoft

Is junk data a big problem in organisations, and how should they address it

A lot of organisations still have a large amount of junk data because, like data in general, it's abstract and hard to visualise compared to physical waste. Some organisations aren't even aware of the type of data they have yet it's still copied over, maintained and backed up – all of which requires computing resources and energy.

In data management, you should therefore focus on data products and tie that into the business value you're creating. There are tools that can give you a complete overview of your data, so you can determine what's of value and what has become junk.

What's your thoughts on the amount of data being wasted by organisations who are having to move it around?

There are a lot of wasted resources including time and energy, not to mention the increased costs, when moving data around inefficiently. Having worked in data integration, you see how much effort the teams spend on mapping, analysing and moving to different environments.

It's not easy to manage and process vast amounts of data which is why it becomes resource-intensive. This is why organisations need solid data management practices, optimisation, and strategic planning to reduce the movement of data and improve efficiency.

As demand for AI and analytics continues to grow, how can organisations optimise their IT infrastructure and model development to minimise energy consumption?

Many organisations now recognise they need to prioritise or target inefficiencies to mitigate environmental impact – so it's likely that we will see a technological solution leading to greater efficiency, storage and better algorithms that consume less energy within the next few years.

Companies are already optimising their CPU with more efficient and better-performing algorithms. The technological evolution is already there, so we're focusing on specific workloads, especially training algorithms to have AI workloads.

The market is following the technical evolution of the hardware too. Previously, users might have chosen a CPU that could handle any kind of workload – but now the CPU vendors are refining their products so that users can manage their workloads more efficiently.

How can technical teams work with other stakeholders within an organisation to manage data more efficiently and reduce their carbon footprint?

Data centres are focusing on their environmental impact with some operating with renewable energy or working towards carbon neutrality. Working with data centres that are already making progress in this area is a way for organisations to improve their energy efficiency.

Within an organisation, stakeholders need to work together to understand the environmental impact of projects and identify and eliminate inefficiencies when building data and AI solutions. Sometimes you need to acknowledge that a solution might not have been developed in an environmentally friendly way today – but you know what steps you can take to improve it in the future. Similarly, it's important to understand the impact on a small scale so you can see what it would be when the solution is scaled up.

How can technical teams manage conflict around sustainability and other interests?

Conflicts of interest are challenges for any organisation. Those pushing ahead with green initiatives might find they are difficult to achieve from an economic perspective. But there's always a way to solve an issue – perhaps not right away but by being mindful of sustainability and working in an ethical way you can make real progress, especially when combined with the power of analytics to address climate issues.

Are there any innovations/methods you've seen, or that are on the horizon, which could help organisations get more from their data while managing it more efficiently?

Virtualisation is accepted as a standard optimisation tool in the server world, but it's sometimes difficult to prove its value in many companies. Yet it can play an important role in reducing storage and CPU requirements, and the need for physical servers.

Using AI for managing data is another innovation – in terms of optimising resources, weeding out junk data and ensuring good data quality. When your data quality is poor, you need to rerun it until it delivers the value that you are expecting. If you can tackle that earlier in the chain, you rerun it less and can save time and improve efficiency.

Pre-trained large language models (LLMs) are energy-intensive so it's likely we'll see AI being used to refine and optimise them.

Sustainability might feel like a big challenge for organisations to tackle, which is why partnerships with vendors are so important. SAS, for instance, continually optimises its algorithms for efficiency and performance. But it's not just about the technology itself but shared values, including sustainability. This makes it easier for organisations to see how new innovations in the software will help them achieve their objectives.

About 7wData

Yves Mulkers is a leading data strategist and the founder of 7wData – a resource for digital and disruptive technologies. Yves specialises in data integration and is experienced in the technical and strategic side of data, and how it can help businesses remain competitive and optimise their processes.

"Stakeholders need to work together to understand the environmental impact of projects and identify and eliminate inefficiencies when building data and AI solutions."

Industry insight 2

Spiros Potamitis, Senior Data Scientist, SAS

We caught up with Spiros Potamitis, Senior Data Scientist at SAS, who explains how organisations can reduce the environmental impact of data and AI workloads in the cloud.

Is it fair to say that organisations haven't always considered the environmental impact of moving to the cloud and adopting technologies like AI/ML?

Organisations often think that environmental responsibility is primarily a cloud vendor obligation, but the truth is, it's a shared responsibility. While the hyperscalers are ethically bound and increasingly required by emerging global regulations to keep their environmental impact to a minimum, organisations must have accountability for their computational efficiency in the cloud and for applying responsible 'cloud citizen' practices.

Even small improvements can make a big difference especially when we consider the cumulative impact. It's worth clarifying that the cloud is more environmentally friendly compared to on-prem deployments, due to economies of scale and aggregated resources.

Optimisation is more readily achievable in larger data centres compared to on-prem environments managed by disparate teams who do not share common objectives. A growing trend is for hyperscalers to provide free emission savings calculators, so their customers can get an estimate of their carbon emissions and how much they can reduce their environmental impact when moving to the cloud.

Is that starting to change as ESG reporting requirements and pressure from customers and the public grows?

A recent Gartner report includes sustainable technology as a top 10 strategy trend for 2024. It has also predicted that by 2027, 25% of CIOs will have their compensation linked to their sustainable technology impact.[15]

The increasing number of environmental regulations is helping shape the way people think about the topic. Across the globe, we have more than 3,000 climate laws and policies which demonstrates that it's a global challenge, and we all have a part to play.

Research shows that around 80% of investors say ESG plays an important part in deciding where to invest because they see the bigger picture.[16] This view is heavily influenced by the consumer trend of prioritising sustainability in the past few years.

While environmental regulations can vary by country, organisations can be strategic about where they store their data. They can choose hosted service providers who have data centres around the world and can compare environmental benefits across regions.

[15] Source: Gartner
[16] Source: PwC

How do organisations go about quantifying their environmental impact when their cloud infrastructure might be complex and managed/owned by different teams, and delivered by external providers?

Most major cloud platforms have introduced emissions impact dashboards, where users can monitor their carbon footprint. It enables them to view, track and report their carbon emissions related to their cloud usage. While these reports or dashboards can provide reasonable emissions estimates, they often overstate reductions or savings by using their services because of broad-based assumptions about how a potential customer is managing their own environments. Estimates can also vary significantly when working with various external providers depending on calculation methodologies, access to clean energy, data centre practices and a host of other assumptions.

This makes it hard to see the bigger picture and as technology gets more complicated, it becomes more difficult to get reliable estimates. To get a more accurate figure, organisations need to develop their own ESG dashboards which includes their technology stack, and direct and indirect activities that link to their carbon emissions. This will require investment.

Are data centre operators making improvements in reducing energy consumption?

Cloud operators are investing heavily in AI enabled and higher-efficiency technologies, resource utilisation strategies, renewable energy, offsets and numerous initiatives to reduce their environmental impact. As well as improving sustainability, this can also reduce costs for customers.

Data centres also need to work to reduce their indirect emissions. For example, they can increase the lifespan of existing equipment by using smarter software that reduces stress and increases asset longevity.

As well as the environmental impact, are senior leaders increasingly concerned about the financial costs too?

Financial cost has become a top consideration, even rivalling security. Senior leaders are more concerned about cost because they are facing unexpected surprises due to the lack of cloud optimisation strategies, and appropriate tools.

Organisations are also becoming more aware of how the cloud works and cloud economics, and technology providers are offering more intelligent monitoring tools. They now have the opportunity to control their cloud usage and optimise their most resource intensive processes.

Investments in newer, more efficient hardware can dramatically reduce energy demand. There are also clever ways of using software on older machines to achieve more with less. Organisations can't afford high-cost resources when they are trying to maintain their margins, so they need to find a strict balance.

Could you explain what impact inefficiencies in analytics are having on energy requirements? How have they come about and do they often go unnoticed?

Not all algorithms are designed with efficient scalability in mind. For example, you might have a data set with 1,000 customers and your algorithm takes 10 seconds to run. This doesn't mean that running the same algorithm for one million customers will increase the processing time proportionally – it could be exponentially longer. Even if cost was not an issue, and you increased the computing power proportionally to the size of the data, you might still not get the expected results if you don't have the right technology and optimised algorithms in place.

The impact of inefficient processing can be analysed in different ways, for example, using monitoring tools such as the Green Algorithms[17] calculator which estimates carbon emissions when running a workload in the cloud. If you reduce the time needed to run analytical workloads in half, it will also have the added benefit of significantly reducing your carbon emissions.

[17] Source: Green Algorithms

Should organisations be striving towards small improvements and incremental change rather than being overwhelmed by the task, or limiting cloud adoption?

When it comes to the environment, you have to go all in. Given the urgency for actions needed to avoid the worst case scenarios of climate change, hosting data with responsible service providers who can manage resources more efficiently and cost effectively, could quickly reduce your emission's footprint – we're talking double digit percentages. However, a migration from an on-prem solution has to be assessed thoroughly and the workloads and code have to be optimised where possible, in order to bring both financial and environmental benefits to the organisation.

Something else to consider is using platforms that come with optimised algorithms that scale efficiently right out of the box. This can help you adopt an environmentally friendly approach, but also enable you to reduce your compute resources and the time spent using them.

You've previously said 'using resources in the cloud only when needed is an effective way to reduce greenhouse gas emissions'. Do organisations need to think smarter about how they use the cloud, and how can this be achieved?

Organisations should complete a thorough assessment of their cloud environment and continuously scale their resources based on their data, AI and analytics needs.

Switching on and off resources when not in use or according to the task in hand can make a big difference as the energy consumption depends more on the memory available than the actual usage. Under-utilisation of resources is an issue for organisations.

Large organisations also have data in multiple cloud platforms and move it around, which some may argue is an unnecessary use of the cloud. Having duplicate copies and your data in different locations across the world uses more power in terms of storage and computing. This is a complicated issue because of vendor lock-in concerns and some organisations may need their data on-prem for security or regulation purposes.

There's a lot to consider, and it has to be assessed on a case-by-case basis. One solution is to use a data & AI platform that allows users to access the data wherever it is stored – be that in different cloud platforms or even on-prem.

Do you have any other tips or examples of best practice you can share?

Sun and wind-based scheduling is definitely worth exploring. Organisations could schedule their data & AI workloads, in the daytime when low carbon power sources like wind and solar are plentiful.

In-database processing is another useful concept. This means that you don't have to move your data to the AI platform, process it, and then move it back. As a result, you don't need more computing resources to do the work, and you can save time because it's faster. It can get more complicated based on your database, and the way you want to process your data but there can be significant benefits in terms of efficiency.

Find out more: Exploring the Environmental Impact of AI and Analytics in the Cloud *[SAS webinar]*

Industry insight 3

Luke Davies, Managing Director of Datacenters, Globalconnect

More organisations are moving their IT infrastructure to the cloud – helping them to reduce costs, improve security, scale faster and keep their energy consumption down. Here, Luke Davies, Managing Director of Datacenters at GlobalConnect, talks us through some of the latest innovations in data centre design and operation, and how this can support sustainability strategies.

Data centre operators and their customers have been conscious of the potential for a lighter carbon footprint for some time now. It's partly for this reason that we see so many hyperscalers and other multinationals building or using data centres in the Nordics. Here we are blessed, not just with a lower energy Total Cost of Ownership (TCO) than other key European data centre locations, but also with a higher 'green quotient' (i.e. the share of total power generation from non-fossil sources).

Nordic data centre operators are not content with the natural advantages of the region. We are also working hard to increase the efficiency of our facilities. We have seen flagship efforts, like the recent naming of Google and Microsoft among the founders of a European innovation hub that aims to decarbonise data centres. Most operators are also working with an individual plan to decarbonise.

A key plank in many of these plans is district heating offtake. For example, we launched a successful pilot to reuse heat from one of our Stockholm data centres in a local district heating system.

Now we have adapted our largest data centre in Greater Copenhagen to provide free energy for 1,500 households.[18] We worked in close partnership with the local district heating association to ensure we could provide excess heat in the right way. The piping is now ready for them to couple up to. We have decided not to charge the association for the heat we provide, because we are happy to make a contribution to the local community and we see an advantage in reducing the power bills of our own customers.

Elsewhere in the data centre, we are working with a battery of liquid-based cooling technologies. Last year, we became the first co-location provider in Europe to offer submerged cooling technology to customers, which can reduce power consumption for cooling by as much as 90%.[19] We are now designing installations for customers who need direct to chip liquid cooling, whereby a heat exchanger is placed over the chip in the customer's server to cool it directly. These technologies are not just more efficient than conventional air cooling; they are necessary as workloads become more intense due to advanced analytics and artificial intelligence.

And then there are all the small, well-known but sometimes-forgotten initiatives which can reduce energy consumption – switching out old light bulbs for LEDs, consistently installing blinds when racks are not filled to capacity, using cold or hot aisle containment to reduce wastage, as well as the swapping out old, inefficient compressors, coolers and UPSes (Uninterruptible Power Supply devices) for more modern ones. Taken together, these incremental improvements can make a real difference.

[18] Source: GlobalConnect
[19] Source: GlobalConnect

Demand for AI and advanced analytics

Demand for AI and advanced analytics is growing rapidly, and while nobody can reliably put numbers on it, the total data centre capacity needed to support this growth is likely to be as significant as the boom driven by cloud migration.

The Uptime Institute has found that server racks densities are increasing, albeit slowly. Today, the average server rack density is 6 kilowatt (kW) per rack, and most operators don't have any racks above 20kW.[20] But we're going to see AI workloads that need to be operated at 40-50kW, and ultimately higher.

We're already getting requests for capacity from GPUaaS (GPU-as-a-service) companies and are talking to hyperscalers about how to support their growth plans. What's clear to us is that regions like the Nordics will play an important role in supporting the coming AI wave because of the attractiveness of the region from a power TCO perspective.

Shifting priorities

The urgency of climate change, public pressure and new regulations all mean businesses are increasingly prioritising sustainability when choosing providers.

There has been a first wave of customers for whom sustainability is really important because of their corporate governance, financing or values. As a result, they're scrutinising data centre operators to find out what proportion of their energy comes from renewables. Some are also becoming more interested in innovations like immersive cooling – but I'm looking forward to these technologies having wider appeal because that will lead to more industry-wide changes in data centre design.

Without data centres, there will be no AI – so there's a structural incentive to make them as efficient as possible, otherwise it would be financially and environmentally unsustainable. The high cost would put the brakes on innovation projects, some of which are designed to tackle the challenges around climate change, such as drought detection or natural disaster management.

"Without data centres, there will be no AI – so there's a structural incentive to make them as efficient as possible."

[20] Source: Uptime Institute

The simplest, and in my opinion most effective, way to achieve this is to locate data centres in areas where renewable energy is plentiful. Again, the Nordics benefit from cooler temperatures and a good supply of hydro-energy which will have a natural ‘greening’ effect on AI.

AI itself can improve the efficiency of data centre operations too. We use AI to optimise airflow in data halls, and recognise security threats more efficiently. When you have thousands of cameras and sensors across your data centre estate, AI could help security teams to respond quicker.

There’s a lot to come – we’re continuously looking for ways to use AI to improve energy efficiency across all the equipment we have, such as identifying areas with spare capacity. This, combined with other AI-led optimisations, could transform the sustainability of data centres – enabling organisations to push ahead with new innovations and meet their ESG objectives and obligations.

About GlobalConnect

GlobalConnect is one of the leading digital infrastructure and data communication providers in the Nordic region, driving more than half of all data traffic in and out of the Nordics. GlobalConnect delivers fiber-based broadband services to more than 800,000 private consumers and end-to-end connectivity solutions to 30,000 B2B customers via its network comprising 215,000 km fiber cable lengths across Denmark, Norway, Sweden, Germany and Finland.

Forecasting the future:
5 steps to efficient data and AI in the cloud

1 Create a sustainability culture

Sustainability is a collective effort, requiring buy-in from stakeholders at every level. By prioritising it, organisations can reap benefits that extend far beyond ESG reporting compliance. It can strengthen their reputation among customers and investors, attract climate conscious talent, and potentially reduce costs as cloud storage and hosting fees climb.[21]

To avoid accusations of greenwashing, organisations should scrutinise and optimise all areas of their operation – including ones that might not always be on the public's radar, like data and AI workloads.

2 Understand the environmental impact of computational workloads – and make continuous improvements

Organisations cannot improve their model development if they don't know how energy-intensive their computational workloads are now. Monitoring tools help you identify resource intensive processes, while the Green Algorithms calculator can give you a good idea of your carbon footprint, based on your average run times; the type, number and thermal design power (TDP) of cores; available memory; and your platform and server.

SAS recently used the Green Algorithms calculator to determine how much carbon could be reduced if an organisation were to use SAS Viya, its cloud-based data and AI platform.

Independent research has found that the SAS Viya platform is, on average, 30 times faster than commercial and open source alternatives.[22] SAS then inputted this metric into the calculator and found that, because of its speed, an organisation using Viya could cut up to 50 tons of CO_2e from their footprint over five years.[23]

To put this into perspective, a mature tree would take more than 4,500 years to absorb this much. The cumulative impact of this, multiplied across companies and industries around the world, could result in a significant reduction of carbon emissions.

[21] Source: Guardian
[22] Source: Futurum Group
[23] Source: SAS

Continuous assessment and optimisation

A continuous improvement framework, which assesses and optimises organisations' cloud usage, is key to reducing their energy requirements and environmental impact. It involves the following steps:

Assess...

1. where your data resides currently and where it needs to be moved for processing

2. current data and AI workloads' processing times

Optimise with...

3. wind & solar based scheduling in the cloud

4. autoscaling of cloud compute resources where possible

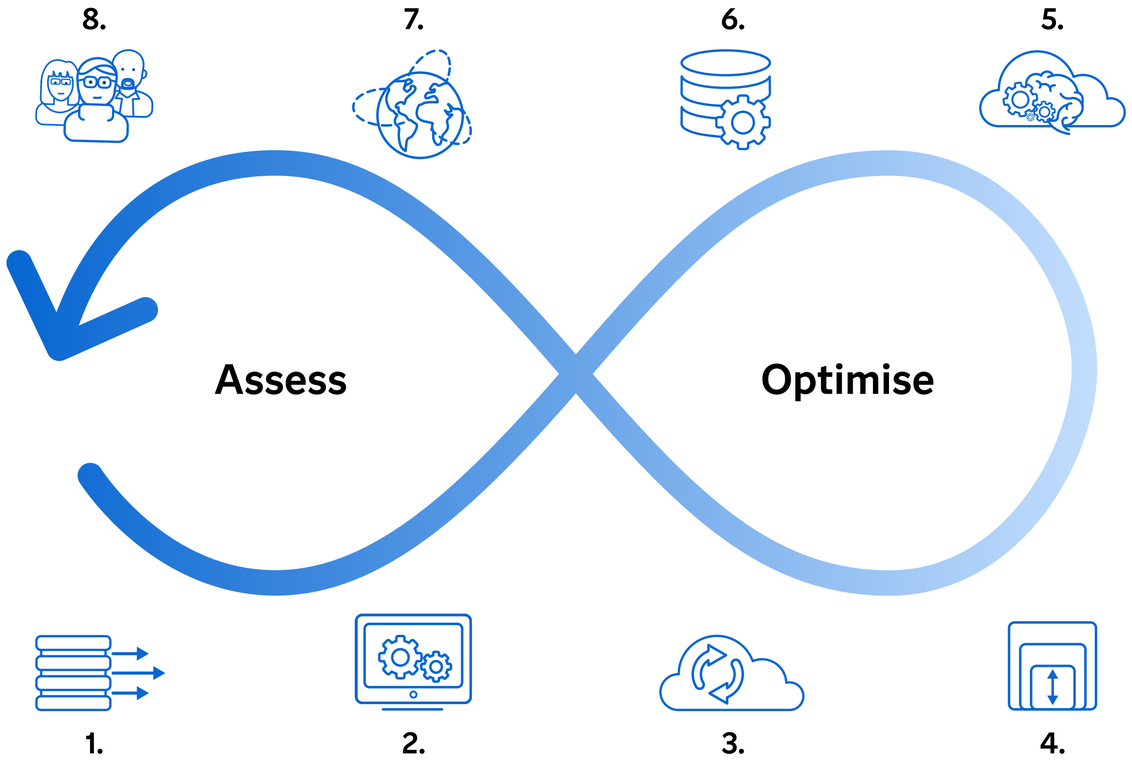
5. AI and ML optimised algorithms to run in the cloud

6. in-database processing

Back to Assess...

7. the carbon intensity of regions of data centres in the cloud of choice

8. the team's skills and training needs.



Source: SAS

3 Tackle cloud proliferation

There are sound reasons why organisations have to use multiple public and private clouds – namely, security, performance and preventing a single source of failure.

However, a poorly-managed multi-cloud environment can lead to duplicated workloads, wasted storage and longer times to access analytics, which reduce energy efficiency. With a single cloud agnostic data and AI platform, teams can connect to different clouds with ease. It is also possible to access data wherever it is located, preventing the need to move data or pool data from different sources.

Read more: A silver lining from every cloud: How to avoid the pitfalls of multi cloud and analytics platform environments

4 Take a strategic approach to model development with a low/no/yes code platform

A low/no/pro-code platform that is optimised to run in the cloud allows people in all parts of the business to reduce design time and use highly performant models to improve their decision making. As long as these tools are developed, trained and tested within an ethical and compliant framework, where sustainability is key consideration, they can maximise the value from data and AI workloads, without investing in additional CPUs or unnecessarily extending the time needed using compute resources in the cloud.

5 Choose the right partners

Organisations with a strong sustainability culture are more likely to choose technology vendors that share their values and can help them to achieve their sustainability goals. This collective pressure from businesses, and ultimately consumers, will pave the way for future innovations in data centres, hardware and software.

sas viya

How can SAS Viya help?

Find out how SAS Viya can support your cloud migration strategy and optimise your data and AI here.

sas.com/uk/viya