Ricky's Afterthought:

Data centres

A.C. (Ricky) Metaxas

Life Fellow St John's College Cambridge UK
Contact E-mail: acm33@cam.ac.uk

Why do we need data centres, what is the implication for energy consumption and what form of energy should be used?

With the exponential expansion of Artificial Intelligence (AI) pervading all diverse areas one requires massive data centres to store and disseminate the information and as a consequence the power requirement is huge.

Data centres are at the heart of digital technology. They are a physical facility where businesses store many servers, computer equipment and storage data enabling cloud computing, AI and the internet itself to function.

Servers

A server is a specialized computer or software system designed to provide services, data, or resources to other computers, known as clients, over a network.

These services can range from delivering web pages and emails to storing and managing files or running applications. These machines run on a client-server model, where clients request specific services or resources, and the server fulfils these requests. In addition data centres also include many rooters and switches and other hardware equipment which enable the flow of data between servers and external networks with increasing speed which necessitates more power. In short, data centres role is to process, store, and communicate the data behind the enormous amount of information that needs to be processed daily whether it be streaming video, email, social media, online collaboration, or scientific computing.

These servers require massive amounts of electricity to function which puts a strain on the ability to provide continuous uninterrupted power.

Back up generation

Data centres operate continuously so backup systems such as diesel generators or battery storage must be in place in case of an outage from the grid. Because they function continuously they generate enormous amount of heat which must be dissipated so cooling systems are used which themselves require an increasing amount of energy to function.

Type of electricity generation

The transition from fossil fuels to electricity is at the heart of what AMPERE members have been advocating for over 40 years.

For example, use of electricity for drying of textile packages is carried out by Radio Frequency (RF) particularly after the removal of as much moisture as possible by mechanical means (mangling) or putting the product through centrifuges. Ideally, the RF power should be derived from electricity generated from nuclear or renewables such as wind or solar especially as these produce no greenhouse gases. But when considering the massive amount of power required for data centres questions arise as to whether nuclear, solar or wind generated electricity would suffice. These data centres require continuous power therefore powering these by renewables may pose a problem in that wind generation depends on atmospheric conditions and solar is not always available unless the data centre is located in a desert-like environment such as the one in Reno, Nevada.

Further, as renewable energy takes much longer to bring online than building data centres, oil and gas would still have to play a major role in bridging the gap in energy supply.

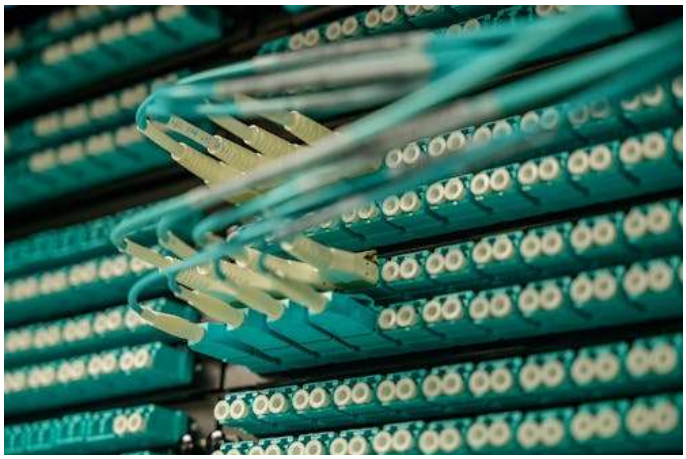
Some have advocated that as we need all the available resources for electricity generation for powering data centres, gas fired power stations alongside nuclear should continue to be used because their carbon footprint is less than that from oil or coal fired stations. Some systems use combined heat and power (CHP) systems to produce both electricity and heat from the same source. Cogeneration, as it is termed, may play a large part in powering power-hungry data centres.

Fuel cells can also be used as a power source which converts hydrogen (preferably itself generated through electrolysis powered by renewables) to produce electricity which can power a data centre (See Issue 107 July 9 2021).

Typical data centre courtesy of Pexels.

Cooling

Some AI data centres are deploying liquid cooling solutions to improve thermal management as computer workloads increase. Such an innovative method is direct-to-chip cooling which dissipates heat directly from the chip, using liquid's high thermal transfer properties allowing data centres to support higher rack densities while maximizing energy efficiency. That in itself can cause problems because the water consumption used in some data centres is extremely high, often hundreds of millions of gallons per annum which could provoke the fury of local residents once they realise that their water shortages could be blamed on nearby data centres.



An array of cables and sockets in a typical rack courtesy of Pexels.

However, air-assisted liquid cooling offers a strategic advantage for businesses aiming to harness AI and maintain a competitive edge. Combining efficient room and direct liquid cooling methods can help organizations lower energy costs, boost performance, and meet AI data centre demands. Another way to reduce the amount of water used is to deploy a closed system whereby cold water is circulated around the chips producing very high temperatures in the water which is in turn cooled by fan power behind the racks. This of course requires even more electrical power.

Large data centres

China and the USA are spearheading the data centre revolution. One of the largest is China Telecom, a large campus in Mongolia, where it is cold and allows air cooling which reduces costs, holding 50000 servers and aiming at 1million savers. In the USA the Citadel in Nevada is powered entirely by renewable energy, having a capacity of 130 MW or 55kW per rack. E-bay, Microsoft, Google, Amazon all make use of this facility for cloud, AI and e-commerce data transmission. Oracle Corporation, which has built its reputation for database management software, is heavily committed to cloud computing, and its co-founder billionaire Larry Ellison has promised to build many more data centres around the world having already 160, 6 alone in the UK. The company is spear heading its cloud-based infrastructure to serve the public sector and businesses alike.

Mark Zuckerberg remarked recently that he was going to allocate hundreds of $1b towards

building and powering data centres presumably to satisfy Meta's operations and not rely on systems that serve many companies. No doubt Google, Amazon Microsoft and e-Bay have similar aims.

One of the largest and more sustainable in Europe is in Portugal run by data centre company Start Campus' called Sines. The IT capacity is about 1.2 GW powered off the grid. Currently France's Paris-Saclay facility run by Data4 aims to increase the existing data centres from 13 to 30 to be completed in 2029 with a total power capacity of 375 MW. Powered by the very powerful Villejust electricity substation it has two underground voltage lines of 90kV each. It is interconnected to over 70 telecom operators and has access to more than 150 cloud platforms. Germany has some 490 data centres while France and Italy 320 and 209 respectively. Estimates of data centres in Canada, Australia, India and Japan are all in the region 250-300.

The UK scene

Currently there are over 400 data centres in the UK and predictions are that the number in the next 10 years is set to increase tenfold. Water reservoirs are being constructed so that there is sufficient water for cooling. The electricity demand is also set to dramatically increase which may almost certainly put up the cost for local consumers, although many suggest that any increase in electricity costs should be borne by the massive companies such as Microsoft and Meta owning these data centres. Apparently the UK is only behind the USA and Germany in the number of data centres already under construction and projected into the future. Most of these are to be built near London and surrounding counties although sites in Wales and Scotland will also have data centres.

One such system is the CWL1 Vantage data centre located near Cardiff, Wales, where two data centres, and a third under construction, are powered by renewable energy. It has a capacity of 148 MW, with a direct-private 400kV super grid connection. Many investment companies are also in need of large data centres such as the Blackstone Group wishing to build a £10bn unit in Blyth near Newcastle starting in 2031 and to be completed in 3 years. It involves massive buildings occupying some 540,000 m^2. Although this is huge one must remember that old power stations are being demolished and in their place giant data centres are springing up. If truth be told it is fair to say that it would be very difficult to predict how many data centres currently exist in the UK but one estimate suggests that 13% of data centres are in London and its surroundings.

A word of caution

A note of caution must be aired on the coexistence of the huge drive of AI in many areas, which suggests exponential acceleration of energy usage, with climate ambition towards net zero. A report from the University of Cambridge suggests that this can only be achieved if the strategies for digital and climate are aligned. I suggest that similar dilemmas exist in many countries worldwide.

Postscript

In my Afterthought article in Issue 116 on AI, I mentioned the impact that Nvidia has made worldwide, co-founded by Jensen Huang currently valued at $5tn and is the leading manufacturer of high end Graphics Processing Units. Huang and his family were in Cambridge last November to be awarded The Hawking Fellowship, an honour bestowed in memory of Professor Stephen Hawking to someone who has made a significant impact on science and particularly AI. After the ceremony he and his family were hosted at the Master's Lodge at St John's College. During his visit to the UK he also met King Charles who awarded him with the 2025 Queen Elizabeth Prize for Engineering during a ceremony at St James's Palace.