There was a time when brick-and-mortar structures were constructed to store highly valuable assets such as gold and silver, especially by banks. With the evolution of modern banking systems and the establishment of central banks in most countries—such as the Federal Reserve in the United States—these precious metals came to be stored in highly secure facilities like Fort Knox.

However, with the dawn of the 21st century, a new and equally valuable asset has emerged: data. In today's digital economy, data is often referred to as "the new gold." Much like gold, it requires secure storage, but unlike bullion reserves data requires rapid real-time processing. As a result, data centres have proliferated across the globe, serving as modern vaults that store and process trillions of bytes of information. But unlike gold storage, operating a data centre is not only vastly more expensive but also immensely complex.

This essay explores the key challenges associated with data centres—primarily focusing on the financial and energy requirements—and highlights the nations leading the charge in this high-stakes sector.

The Enormous Costs of Setting up a Data Centre

To begin with, the scale and cost of building a large data centre are staggering. On average, such a facility requires over 100,000 square metres (approximately 1.1 million square feet) of floor space to house hundreds of servers and millions of kilometres of cabling. For context, the Empire State Building has a floor area of about 254,000 square metres, making a large data centre nearly half its size.

The cost of acquiring land and constructing a suitable infrastructure can be astronomical. Only corporations with deep financial resources can afford to invest in such ventures. Typically, the capital investment is around $12 million per megawatt (MW) of capacity. A standard large-scale data centre may consume 100 MW or more. Importantly, a significant portion of these costs—such as air conditioning and building infrastructure—is incurred upfront during the deployment of the initial few megawatts. These figures exclude the cost of IT infrastructure, including servers, computers, and high-speed cabling, which can vary greatly depending on the technology used.

According to a McKinsey report, by 2030, global capital expenditure on data centres will reach $6.7 trillion. Of this, $5.2 trillion is expected to be allocated to AI-related computing infrastructure, while the remaining $1.5 trillion will support non-AI applications. To put this into perspective, only the GDPs of the United States and China currently exceed this estimated amount.

Powering the Digital Vaults: Energy Consumption

The second major challenge is the enormous energy demand of data centres. These facilities are significant consumers of electricity, and their energy requirements have been increasing rapidly each year.

According to Roundy & Kirvan on TechTarget, although traditional corporate data centres are growing at a slower rate than 25 years ago—partly due to the shift to cloud computing—they still remain major energy consumers. Meanwhile, hyper scale cloud data centres continue to expand their energy usage. Fortunately, some of this growth is being offset by investments in energy-efficient technologies and greener systems.

However, the rise of artificial intelligence has introduced new, energy-intensive computing demands. The U.S. Department of Energy's 2024 report highlights a dramatic rise in data centre energy consumption—from 60 terawatt-hours (TWh) in 2014 to 176 TWh in 2023, accounting for 4.4% of the country's total electricity use. This surge is primarily due to the growing adoption of accelerated servers designed for AI.

Projections suggest that by 2028, data centres in the U.S. could require 74 to 132 gigawatts (GW) of electricity, which would represent 6.7% to 12% of the nation's total power consumption. The compound annual growth rate (CAGR) for this energy demand is estimated between 13% and 27% for the period 2023–2028.

Globally, the International Energy Agency (IEA) reports that data centres currently consume approximately 500 TWh of electricity. By 2035, this figure is expected to rise to 1,300 TWh, with about 550 TWh supplied by renewable energy sources.

The Sustainability Challenge

This brings us to a critical issue: sustainability. The long-term success of data centres hinges on their ability to transition to clean and renewable energy sources.

The United States leads in this regard. According to the IEA, U.S. data centres currently consume 225 TWh, of which 60% comes from gas-fired plants, and the remainder from nuclear and renewable sources. Meanwhile, China, the second-largest operator of data centres, consumes 130 TWh—but only 23% of this is derived from renewable sources. The rest comes from coal-fired power plants, raising serious environmental concerns.

For data centres to be truly sustainable, their full energy demand must eventually be met through clean energy. While progress is being made, there is still a long way to go—especially for countries heavily reliant on fossil fuels.

North-South Divide

The proliferation of data centres has created a stark divide between rich and poor countries. The high costs associated with building and maintaining data centres, coupled with their massive energy requirements, have made them inaccessible to many developing nations. As a result, data centres have become a luxury reserved for the wealthy, exacerbating the North-South divide in the digital age.

This disparity has significant implications for poorer countries, which are forced to rely on external data centres often located in more affluent nations. This dependence can lead to a loss of control over data, compromising sovereignty and potentially

enabling exploitation by developed countries. The lack of infrastructure for renewable energy sources in poorer countries further compounds the issue, as they are unable to support data centres with sustainable energy.

The inability of poorer countries to develop their own data centres using renewable energy sources creates a power imbalance, where they are beholden to external entities for data storage and processing. This can lead to exploitation, as developed countries may leverage their control over data centres to dictate terms, extract concessions, or even manipulate data for their own interests. The digital divide between rich and poor nations is thus perpetuated, with far-reaching consequences for economic development, sovereignty, and global equity.

Countrywise Number of Data Centres as on March 2024			
US	5381	Russia	251
Germany	521	Japan	219
UK	514	Mexico	170
China	449	Italy	168
Canada	336	Brazil	163
France	315	India	152
Australia	307	Poland	144
Netherlands	297	Rest of the World	2334

Data Credit: Statista