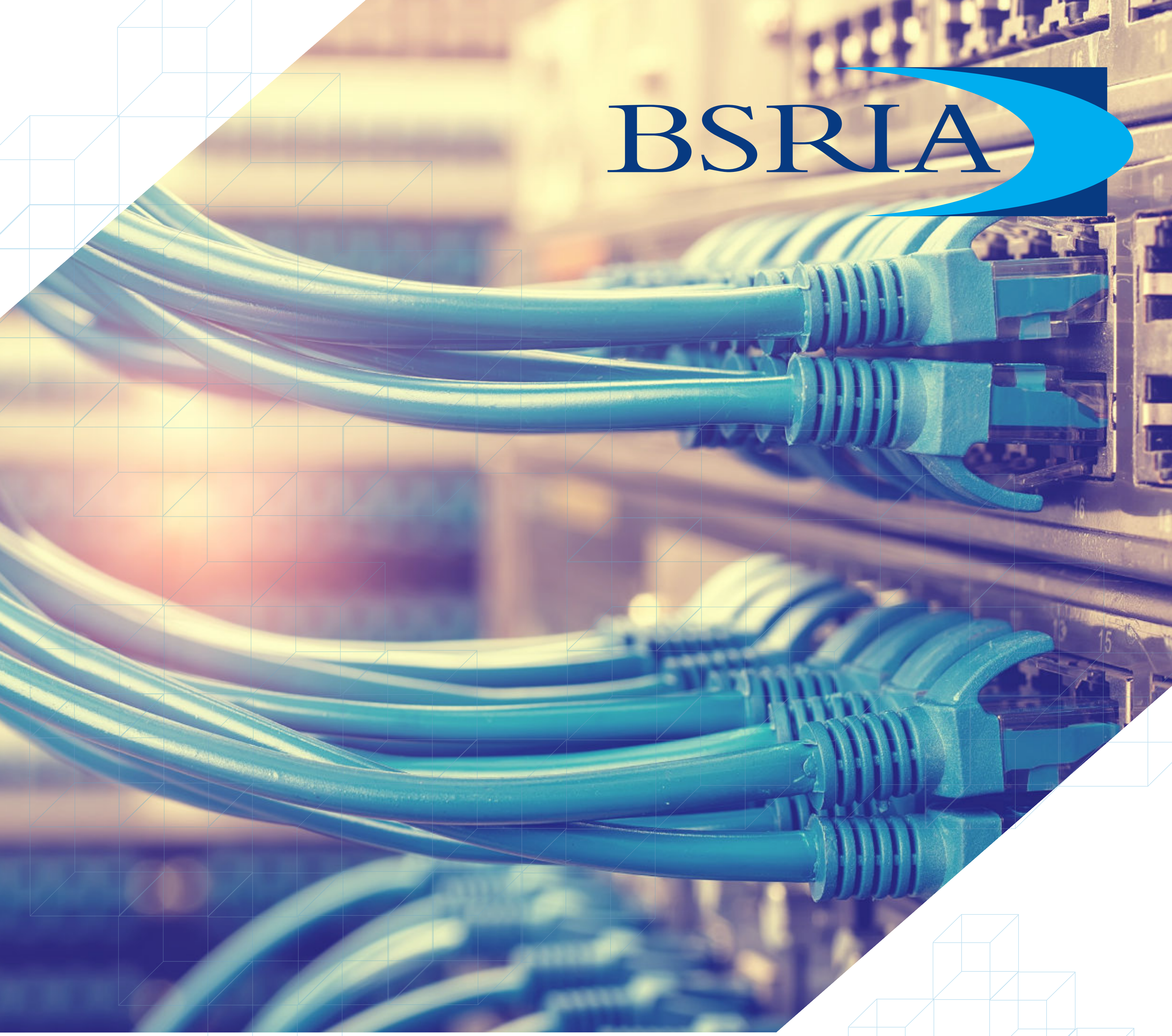


A BSRIA White Paper

DATA CENTRES - FUTURE GROWTH AND INNOVATION

Krystyna Dawson / Henry Lawson

WP 12/2020

DATA CENTRES-
FUTURE GROWTH AND INNOVATION

Krystyna Dawson and Henry Lawson

Date: MARCH 2020

Issued by: BSRIA Limited
Old Bracknell Lane West, Bracknell,
Berkshire RG12 7AH UK
T: +44 (0)1344 465600 | F: +44 (0)1344 465626
E: bsria@bsria.co.uk | W: www.bsria.com/uk/

ISBN: 978-0-86022-777-9

International Copper Association
Copper Alliance

This white paper draws on a BSRIA research project sponsored by the International Copper Association (ICA). The project explores how, thanks to its unique properties, copper facilitates the efficient generation and delivery of electricity, making it an integral part in the development of innovative technologies.

DISCLAIMER

DATA CENTRES-
FUTURE GROWTH AND INNOVATION

We are currently living through a data revolution that will change our lives for ever. It has been calculated that a serviceable recording of all human speech since our ancestors first started talking would require 42 zettabytes of storage. By 2025 the amount of data actually stored is forecast to be four times this. This data, while in a sense virtual, has to “exist” somewhere.

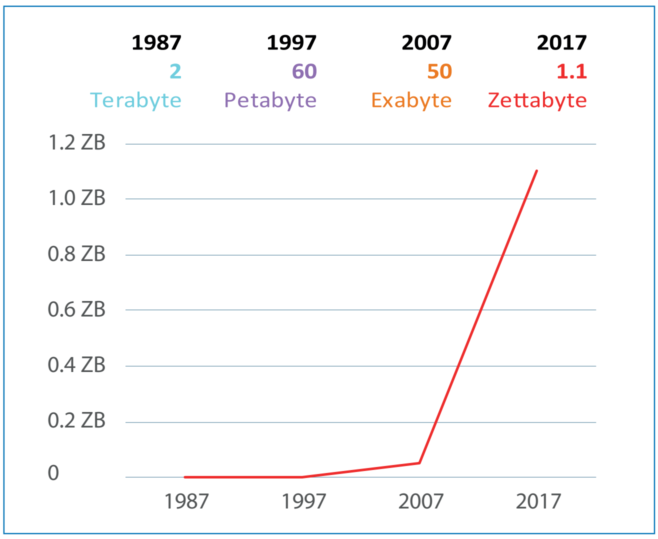
Accordingly, data centres, where data is stored and processed en masse, are now critical infrastructure. They provide part of the platform needed to support our modern way of life, our needs and our expectations that increasingly rely on processing ever-growing amounts of data. An overview of the Internet World Stats shows that over the past 30 years the growth in data use has been massive, with the last decade seeing a particularly spectacular boom.

Technological progress has opened whole arrays of new service possibilities that have found huge social and political resonance and enabled this spectacular development. The data centre industry is actually an aggregation of several industries, driven by tech giants such as Facebook, Amazon, Google and Microsoft.

Services such as Google search and Instagram uploads are delivered using huge storage and processing power from data centres that are very complex buildings,

1 kilobyte (1 kB)	= 1000 bytes	= 10^{3} bytes
1 megabyte (1 MB)	= 1,000,000 bytes	= 10^{6} bytes
1 gigabyte (1 GB)	= 1,000,000,000 bytes	= 10^{9} bytes
1 terabyte (1 TB)	= 1,000,000,000,000 bytes	= 10^{12} bytes
1 petabyte (1 PB)	= 1,000,000,000,000,000 bytes	= 10^{15} bytes
1 exabyte (1 EB)	= 1,000,000,000,000,000,000 bytes	= 10^{18} bytes
1 zettabyte (1 ZB)	= 1,000,000,000,000,000,000,000 bytes	= 10^{21} bytes

Growth of internet use



Source: BSRIA, based on Internet World Stats (https://internetworldstats.com)

designed and built to operators' specifications and demands. They include ICT infrastructure, which embraces IT equipment and communication systems and enables data centres to process, transfer and analyse enormous volumes of data. Power and cooling equipment provide a key supporting infrastructure that ensures smooth and frictionless operation of the data centres. Companies identifying key locations suitable for data centres are also part of this broader industry, as geographical and environmental conditions are of high importance in determining both the cost and the efficient operation of these facilities.

At the bottom of this huge wave of 'progress' building up to satisfy customers' demands there are governments and lobbyists who are concerned about data centres' energy consumption as well as the growing power and influence of the data centre providers. The interests of all these stakeholders, although sometime contradictory at first sight, are based above all on the potential that data processing offers to all, driving constant progress.

The appetites and interests of consumers, business and political bodies will further drive the dynamic growth in data use and with it the development of the data centre industry. Alongside the growth, the industry will also undergo a significant qualitative change.

DRIVERS AND GOALS UNLOCKING THE CHANGES

BSRIA recent research on behalf of the International Copper Association has identified some key drivers for change:

Progression of the internet of things (IoT), growing AI capabilities and the roll out of 5G networks are among the main technology-related drivers that are leading to near-exponential growth in the amount of data to be processed.

At the same time, environmental concerns, related to the high levels of energy that data centres consume, are growing. This is resulting in legislative changes that emphasise the focus on energy efficiency. This is affecting the way in which data centres are set up.

Finally, customers, who increasingly expect seamless delivery of streaming services and real-time applications to multiple devices in any location, are driving a race for improved service delivery. All of this is happening under a regime of tight cost control.

Responding to these drivers, new trends have emerged in the data centre industry that impact the scale, design and structure of data centres:

- a growth of hyperscale data centres accompanied by growth in the edge side of data processing. A key principle here is that the data is processed at the location where it can be done most efficiently.
- an integration of data centres' infrastructure systems
- an ever-stronger shift towards the use of renewable energy and towards the installation of environmentally friendly products and systems;
- a strong push for standardisation of products and systems

These main trends paint a picture of clear goals that the industry is striving to achieve. These include:

- Environmental goals such as reduction of carbon footprint, often driven by legislative pressure
- Economic goals, aiming to achieve simpler, energy efficient data centres that deliver reductions in capital and operational expenditure
- Customer-oriented goals that pursue improved delivery of services.

DATA CENTRES ARE VERY COMPLEX AND ENERGY HUNGRY

Today, most data centres are physical locations (rooms or buildings) holding critical computing and telecommunications resources for companies. These resources include servers, storage systems, databases, access networks and software that runs the applications – for business or leisure.

Data centres are expected to run effectively, delivering services without disruption. Computing and telecommunications functions require specific supporting systems, with those delivering energy and cooling holding a prominent place within the infrastructure and often requiring some level of redundancy of systems to ensure backup in case of failure.

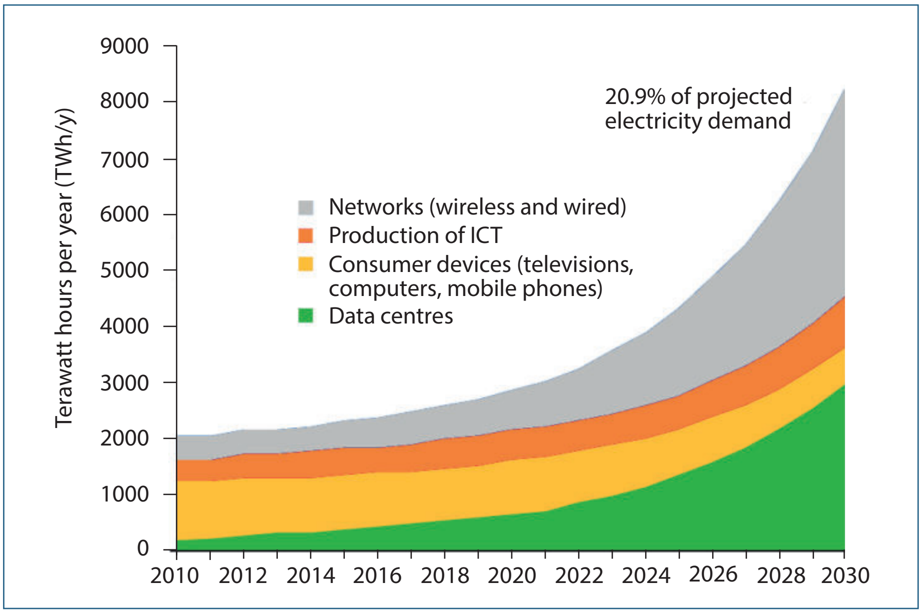
It takes large volumes of many different equipment types to keep a data centre running, including IT, electrical, HVAC, building automation and controls (BACS), security and fire detection products. It also takes a substantial amount of energy for all of the data centre's products and systems to deliver the required performance.

IoT, 5G and integration of electric vehicles will drive communication and ICT progress which is forecast to cause a strong increase in demand for electricity within the next decade; global electricity demand from data centres alone is expected to reach well over 2000 TWh annually by 2030. Such extensive energy use is of concern as data centres are, as of today, a source of growing carbon footprint and the pressure to mitigate their carbon impact in the future is enormous.

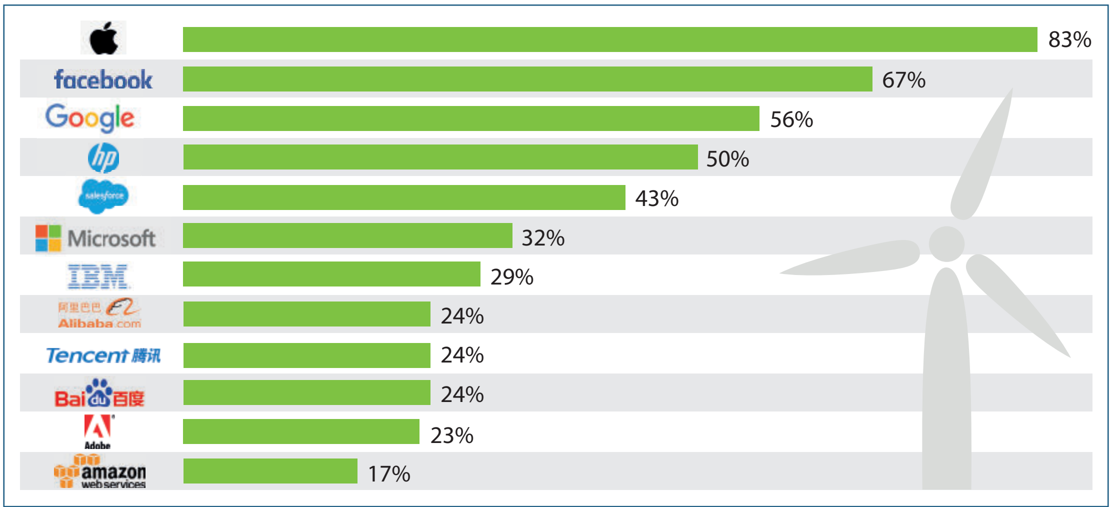
Tech companies involved in the ICT industry are responding by making themselves greener. The largest data centre players are amongst the greenest if judged by the proportion of renewable energy in their power supply.

The "greening" process is also happening in the data centre segment but the focus and scale of changes vary. The developing importance of cloud data centres and hyperscale data centres, which account for an increasing share of electricity used by data centres, is the most important factor driving this change.

Projected global electricity demand (Terawatts)



Source: Nature



The Greenest Tech Companies - Proportion of renewable energy in their power supply 2017

Source: Greenpeace

TYPES OF DATA CENTRES

Data centres can be categorised in multiple ways. For the purposes of this research, they have been categorised by size, as follows:

Hyperscale data centres occupy a large space and usually comprise a set of buildings that have their own power station just to provide power to this particular site. Each set of buildings has a separate function, providing power, cooling and IT as well as important administrative and security functions. In hyperscale data centres the setup is usually duplicated to provide an N+1 level of redundancy where N describes the primary structure and the +1 one duplication of it. If one system fails, then it is possible to easily switch to a full backup to replicate the power, cooling and processing or storage of the service.

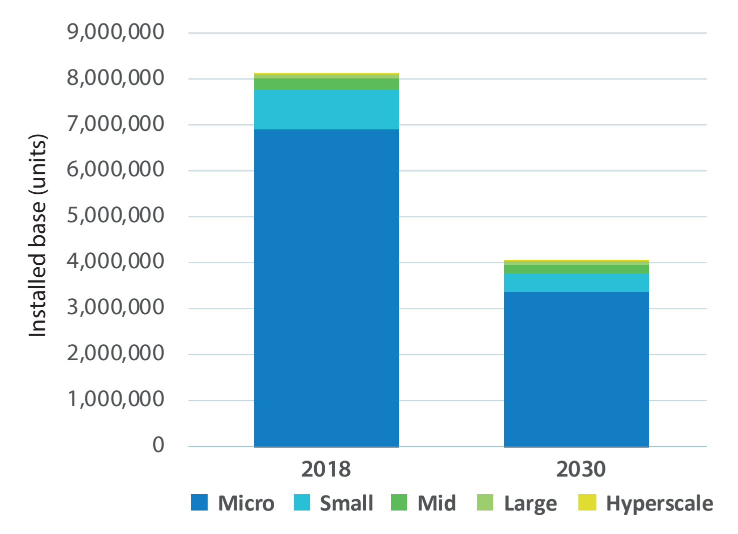
Large/midsize data centres are usually contained in one building, having their own power supply and an N+1 level of redundancy for power and cooling.

Small/micro data centres are located in a dedicated space within a building, rely on the main power supply to the building and do not have full redundancy set up.

Data centres are built using the same techniques, materials and equipment as typical office buildings; however, they differ in their robustness, scale and operational focus on their infrastructure systems.

Currently, the global structure of data centres is characterised by a strong dominance of small and micro data centres that together account for some 95% of the estimated number of just above 8 million data centres globally. Hyperscale and large data centres are a niche segment in terms of the constructed number, however, their importance in terms of data processing capacity is vital and over time it will keep growing.

Installed base of data centres by type, 2018 - 2030



Source: BSRIA research

Today, the number of hyperscale and large data centres is counted in thousands and will remain within this range even though over the next decade their number is expected to double. In the meantime, the number of small/micro data centres contained within buildings, counted in several millions today, will fall dramatically as the processing power of data will shift to the cloud and in the same time edge computing will continue to gain ground.

Edge Computing

Essentially, edge computing involves shifting computing power from the large-scale 'central' processing of traditional ICT (cloud and network services) to smaller, more focused delivery points for networks services – edge data centres. As 5G also needs a higher density of access points and lots of local processing, this move fits neatly with the required growth.

EDGE COMPUTING

IoT growth and 5G progression will provide the basis for growing user expectations towards the performance of services on smart phones and other devices.

Streaming TV services and real-time applications place huge demands on networks and on data centres, and these are just the start of a new wave of applications that will demand high speed, with seamless services delivered to a number of different devices. The next generation ICT platform is being developed to support that. Autonomous vehicles, blockchain, smarter devices, smart homes and smart cities all mean more services, more processing and more data.

The data centre industry is responding to these changes by moving towards large and hyperscale data centres on one hand and deploying edge services on the other. Together with 5G, edge represents a fundamental change to the way we think about computing and data management. It is seen as a powerful enabler to the explosive growth of IoT.

A huge amount of activity across the data centre industry in the next few years will be focused heavily on innovation on the edge. Large numbers of locally-based, self-contained data centres will be needed.

These new edge data centres will help decentralise IT infrastructure, enabling data to be processed at source, with no time delays (low latency) in data processing, supporting real-time applications for autonomous vehicle systems, industrial robots and numerous other applications.

Today, applications for 5G might be limited in scope, but as this area of technology develops and the number of IoT devices grows, new applications will be developed and rolled out. Responding to a growing trend, data centre managers are already shifting services locally, adding small edge data centres to their network.

The developing edge data centres are, however, not of the same setup and structure as current micro or small data centres as they are not necessarily located in buildings.

Edge has existed for a long time, but previously it was mostly about gaining or allowing access to services that were delivered through mainly fixed Wi-Fi or mobile networks to the growing number of IT-based devices.

Services were managed and run through large banks of servers, where the data was also stored and processed; the so called “core”. There was a traditional alliance

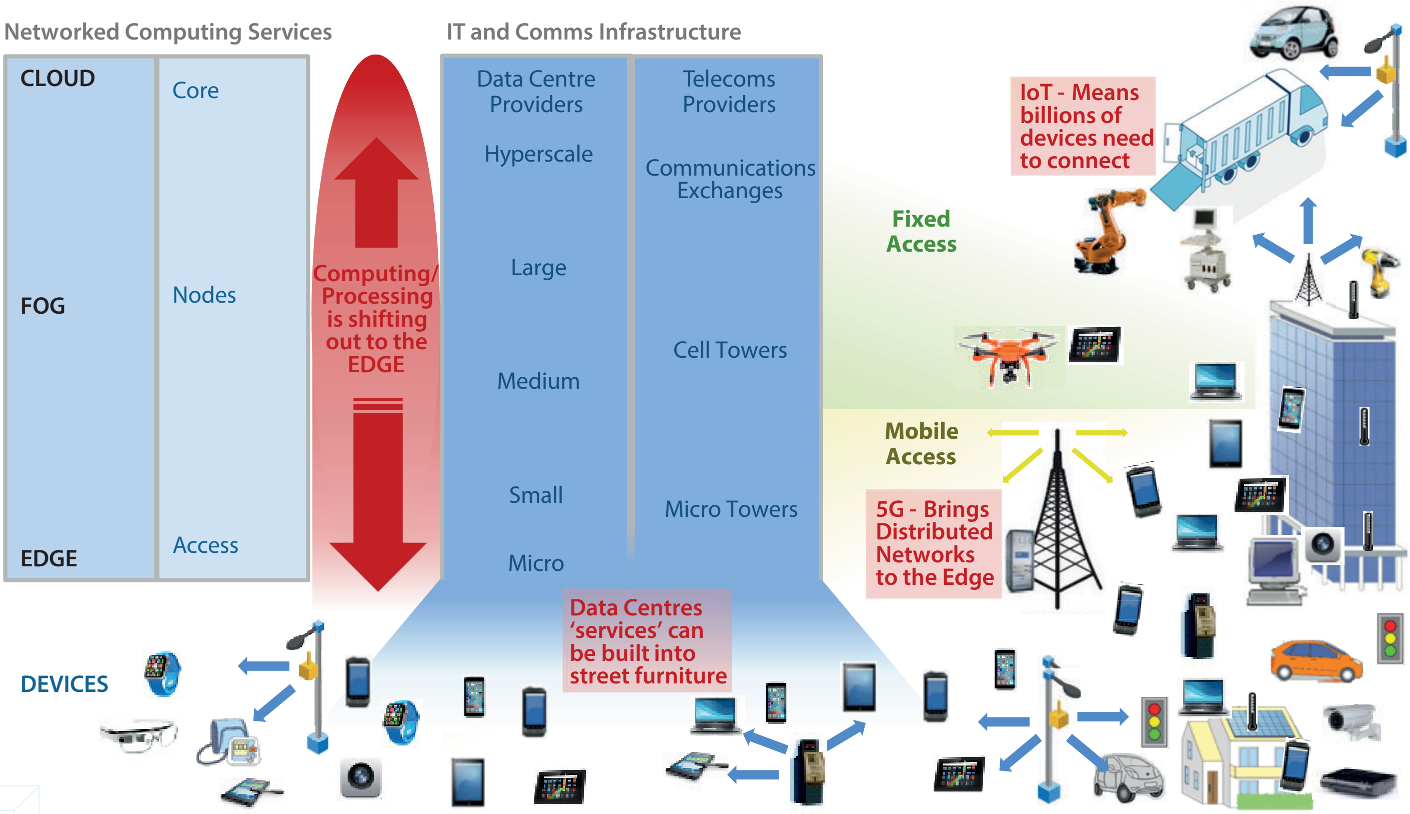
between providers of network equipment, network service providers and app developers that manage and run the system. Today, the number of devices and services related to edge is growing exponentially.

For data centres, this means that the old architecture, driven by central processing and storage, becomes unable to deliver what is required by the Internet of Things and the growing expectation of instant access to rich content and hugely sophisticated apps.

It is not just that there will be a million or billion more devices wanting to access services, but they will need to have access anytime, wherever they are, and not just where they can be plugged in. The new applications need to use local processing to operate near-real-time, with very low latency or time lag on the network calls being made to and from applications and data.

This means that a degree of processing as well as network access must be provided close to the user, at the edge of the network. Consequently, more data-centre-like services are being set up on streets and in buildings to deliver over 5G and Wi-Fi networks. Data processing points are starting to appear with much smaller footprints and where integration is key, rather than data centres. The developing data centre points will become smaller. They will be more like mobile or modular structures that can fit anywhere, for example in a building or on a street corner.

Technology development and legislative, economic and performance pressures are behind changes that have been detected in materials as well as in the type of products and systems that are deployed in currently-built and refurbished data centres worldwide.



Source: BSRIA

> The old architecture, driven by central processing and storage, becomes unable to deliver what is required by the Internet of Things and the growing expectation of instant access to rich content and hugely sophisticated apps.

The most significant changes are to be found in three type of systems that are vital for the smooth functioning of data centres:

- The ICT system includes IT connectivity, communications, apps and storage
- The power system, providing electric power to the ICT
- The environmental system that is responsible for controlling air temperature, humidity and air flow.

Traditionally these systems have had minimal interaction, despite each infrastructure system playing a vital role and generating masses of valuable data. Technology is now enabling these systems to communicate with each other, increasing efficiency and decreasing the risk of outages and downtime. Data centres are typically expected to run at 99.7% or 99.9% uptime service levels and the multiple systems must operate without interruption.

Such high performance requires high quality products and materials but, even then, a high level of utilisation translates into a shorter than the average product lifecycle. The economic & technological lifespan of the ICT infrastructure in a data centre is about three to four years. The economic lifespan of the critical power and cooling infrastructure is between 10 and 15 years. As equipment approaches the end of its life, the likelihood of failure increases, so owners carry out preventative replacement. The rapid step ups in improved functionality resulting from technology advances are also a factor driving replacement.

As IT cycles are driven by both life stage and economics, their performance and value are constantly under review. Moreover, over time, the increase in the amount of equipment in the room is also likely to increase its thermal load. After three to five years, the original cooling equipment often cannot meet the environmental requirements of the room. In consequence, evaluation of replacements for the data centre's systems usually starts from the 4th to the 6th year, and system updates will be completed in the 8th to the 15th year.

Several consumer-expectation-bound factors drive data centre performance up, but data centre managers still need to respond to pressures to lower costs and to reduce energy consumption and the overall carbon footprint. Several steps can be taken to improve performance. The ultimate choices made nevertheless require careful holistic thinking as often gains in one area may result in losses in others.

POWER CONSUMPTION

As users demand faster processing and more advanced services (such as AI, real-time analytics and faster graphics), more high-performing servers are needed with more processing/storage power in smaller and smaller units i.e. a high server density. Quick progress in both results in higher power consumption. Management and reduction of power consumption are therefore a central focus and drive change in the data centre market.

One key area where energy is often wasted lies in the over-capacity of power supply provided to the servers and to the sub-systems. Understanding of power requirements for servers at various times is key to delivering it more efficiently where and when it is needed.

It is recommended that data centre managers regularly audit their assets. In larger data centres there may be some old servers and apps that don't do any useful work and waste energy. The industry often calls these "zombie servers".

Implementation of data centre infrastructure management (DCIM) enables streamlining of data centre operations, driving innovation across ICT, power and cooling systems. It relies on real-time sensing, analysis and management of conditions both physically (in the data centre environment), and virtually (within the ICT setup i.e. when certain processing tasks are needed).

The industry is also increasingly using AI principles to plan, understand, and react to the micro-environment that is being created in modern data centres.

> Management and reduction of power consumption are therefore a central focus and drive change in the data centre market.

ENERGY EFFICIENCY

In the design and configuration of power systems, a trade-off is often needed between redundancy level and energy efficiency. High redundancy reduces operational risk but has an impact on the overall energy use as well as on capital investment. The industry trend is towards optimising redundancy while improving energy performance with upgrades of equipment.

Transformerless UPSs (uninterruptable power supplies) are an example of the latter; Increases in efficiency and improvements in design and performance can be achieved by the adoption of transformerless UPS, made possible by the introduction of insulated gate bipolar transistors (IGBTs). Transformerless UPSs have a smaller footprint, achieve higher efficiency and can be installed in a modular way.

The power distribution system is increasingly delivered using busbars that replace power distribution units (PDUs). The introduction of busbars into hyperscale data centres has been a trend in electrical design for some time and has now started trickling down to large colocation and enterprise data centres. Busbars deliver ease of design and installation, flexibility and improved footprint.

Colocation data centres are those where space and equipment is available for rental to customers. **Enterprise data centres** are those that are owned and operated by the organisation that uses the data centre.

ENVIRONMENTAL SYSTEM TRENDS

Environmental systems ensure, among other things, effective cooling in the space where data centre equipment is working as it generates large amounts of heat. Cooling can be provided by an air-based system where cooled refrigerant is pumped into a device that cools air to then pump it around server racks. Another way makes use of direct fresh air evaporative cooling, where air is taken from outside, cooled and pumped into the IT space. Hot air is then extracted from the building.

Cooling systems can consume up to 40% of the overall DC energy demand and are therefore costly. A lot of effort is directed to minimising power consumption in this area by promoting natural or free cooling where possible. Both are rapidly becoming standard in many large and hyperscale data centres if climatic conditions and air quality level are favourable.

Free cooling can be used in almost all areas north of 28 degrees latitude. It delivers a reduction in power usage, however, extra equipment is needed to maintain humidity levels and often for air filtration/purification. Water usage is higher in such systems; hence access to water needs to be assessed before deciding on the use of such systems. The advantages of free cooling and also its restrictions are driving locations of newly-constructed data centres into cooler climates, where mechanical cooling is only needed in summer months, or not at all.

Optimisation of airflows supports effective cooling in the IT space. Airflow can be measured and optimised to suit server cooling patterns, based on data from sensors within the halls. Providers are increasingly looking for cooling systems that consume less water. For most data centres, water usage is not as costly as energy consumption, but in certain geographies water usage can become a significant expense while also raising environmental concerns.

Control of air (temperature, quality, humidity) and water (quality and usage) are key for data centre operational and energy performance. A lot has been achieved on the energy efficiency front but, despite all efforts, data centres are bound to increase their energy consumption over the next decade.

RENEWABLE POWER SOURCES

Back in 2009, Google estimated that a typical search uses as much energy as would be used for illuminating a 60 watt light bulb for 17 seconds (https://googleblog.blogspot.com/2009/01/powering-google-search.html)

To mitigate the environmental impact, the focus on PUE (power usage effectiveness) in data centres is growing and consequently the importance of renewable energy sources within the data centre market is growing. Medium and small data centres often depend on electricity from the local utility, and here the growth in the share of electricity production accounted for by renewables varies greatly depending on the world region. However, large and hyperscale data centres, which often have their own power generation facilities, increasingly rely on local renewable solar or wind farms for power generation.

Low environmental impact and improved cost effectiveness of renewable energy sources are the main drivers and it is worth mentioning that the attitude adopted in assessing investments has changed significantly over the last decade. Nowadays it is much easier to achieve stability of energy cost with renewable production.

Legislative initiatives also support the drive towards energy efficiency of data centres. In Europe, the EU Code of Conduct for Data Centres was introduced in 2008, driving the trend towards increased power usage effectiveness (PUE). It points towards the ideal value of 1 for this indicator and with the global average being currently around 2.

Heat recovery and reuse. Waste heat generated by the climate control activities of data centres can be used in other applications. This technology is not new but started to make inroads to the data centre setup a few years ago as it needs careful planning and convenient locations to allow other systems to benefit.

Some good examples of innovative energy efficiency designs can be found at https://ec.europa.eu/jrc/en/news/eu-code-conduct-data-centres-10-years-improved-energy-efficiency

POLITICAL FACTORS

Concerns over cybersecurity, data security and the political and economic value of data are becoming significant factors in determining the attitudes of governments, corporations and individuals as to where data centres are geographically located and what controls are placed over the access to any data which is stored or processed within them.

Organisations may not trust certain governments or associated bodies not to interfere with data located in their territories or to safeguard it against third parties. It is harder, for example, to enforce rules, such as "the right to be forgotten", or the requirement to remove data content deemed unacceptable where the data is held in a jurisdiction over which one has little or no control.

Most notably, organisations within the European Union frequently prefer or even require that data be stored and processed within the EU with controls over who can access the data elsewhere. This has encouraged the growth in data centres in various parts of the EU and could in principle result in politically-driven growth of data centres in other countries or regions of the world as well.

DATA AND BUILDINGS

While data is also crucial to optimising and monitoring buildings' performance, it is important to bear in mind that its value is delivered through substantial energy usage. Large tech companies have already engaged heavily not only in the use of renewable energy but also in delivering data centres that are close to the ideal PUE value of 1 while keeping resilience and delivering high quality consumer service as the example of Google's St. Ghislain data centre in Belgium demonstrates.

Nevertheless, the trend of reviewing a data centre's structural setup needs to trickle down to the small/micro data centres that are located on companies' premises. Large numbers of existing data centres are either in the process of replacing several parts of the infrastructure or due to start.

While focusing on improved delivery of services and on controlling cost, it is important to maintain a holistic approach to energy efficiency in data centres. Renewable energy generation, heat recovery and re-use, energy-efficient cooling and water savings are all important factors that together allow optimum results to be achieved. If a data centre is part of a building it is crucial to connect with building services operators to deliver benefits for the entire building. Knowledge and communication are crucial in this process.

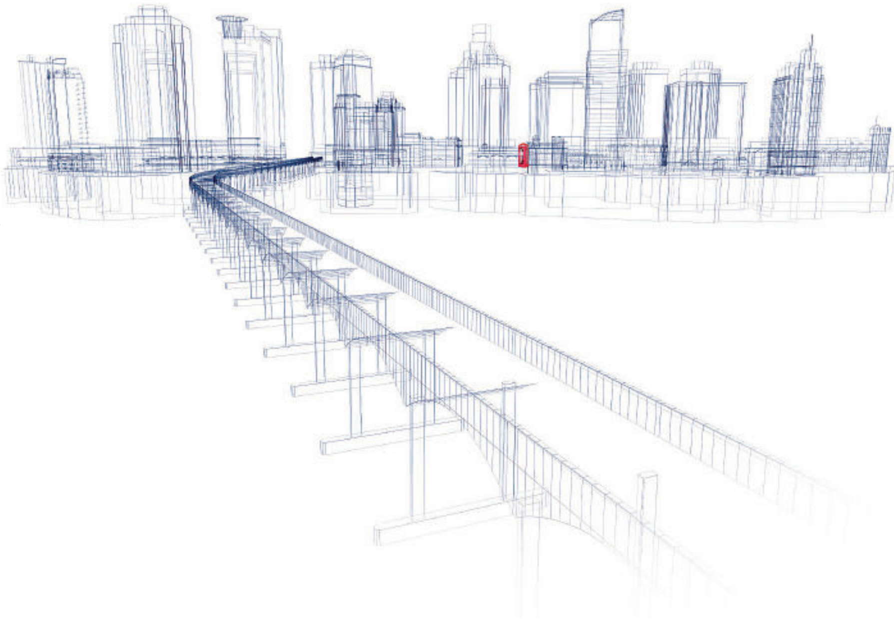
BSRIA

Improving the built environment

Delivering knowledge

Measuring compliance and performance

International



Engineering services: Test, compliance and consultancy support services for clients, designers, constructors, manufacturers and facilities managers

Instrument solutions: Hire, sales and calibration of instruments

Market intelligence: Off the shelf reports, bespoke studies and management consultancy for global markets

Information and membership: Bookshop, library, training, networks and events

Head Office: Old Bracknell Lane West, Bracknell, Berkshire RG12 7AH UK
T: 0800 254 5700 (UK) or +44 (0) 1344 465 600 | **F:** +44 (0) 1344 465 626
E: bsria@bsria.co.uk | **W:** www.bsria.com/uk/