CIB World Building Congress 2019
Hong Kong SAR, China
17 – 21 June 2019

Data Centers as the backbone of smart cities: a study of their cost and benefits

Patrick T.I. Lam

Professor, Dept of Building & Real Estate, The Hong Kong Polytechnic University

(email: bsplam@polyu.edu.hk)

Daniel Lai

Former Government Chief Information Officer, HKSAR Government

(email: dsclai@gmail.com)

C.K. Leung

Assistant Professor (retired), Dept of Electronic and Information Engineering, The Hong Kong Polytechnic University

(email: enchikin@connect.polyu.hk)

Wenjing Yang

PhD Student, Dept of Building & Real Estate, The Hong Kong Polytechnic University

(email: w.j.yang@connect.polyu.hk)

Abstract

As smart cities flourish amidst rapid urbanization and ICT development, the demand for building more and more data centers is rising. Major corporations, establishments and governments need this mission critical facility to enable countless electronic transactions to take place any minute of the day. Their functional importance ranges from health, transport, payment, etc., all the way down to entertainment activities. Some enterprises own them, whilst others use data center services on a co-location basis, in which case data centers are regarded as an investment asset.

Data centers are very energy and sometimes water intensive in their operation, and their construction is usually fast tracked costing much to build, not to mention the high-value equipment contents housed therein. Their site locations need careful selection due to stability and security concerns. Hence, their life-cycle cost needs to be considered against their benefits. As an essential business continuity tool, the return on investment is a complex consideration, but certainly the potential loss caused by any disruption would be a huge amount, if at all monetizable.

This paper examines the issues and considerations of data center procurement from the cost and benefit dimensions, with an aim to illustrate the approaches for maximizing the net benefits and remain "green".

Keywords: Data centers; procurement; life cycle cost; benefits; externalities

1. Introduction

The advent of Information and Communication Technologies (ICT) has recently created an ever-increasing demand on data capture, processing, storage and transfer. Real-time commercial transactions (including online shopping and payment), the wide spread use of social media and entertainment services all add to the traffic and data storage volume. Previously, these transactions took place in teleco rooms or server rooms of individual organizations hosting these activities. However, the need for capacity expansion and more stringent security has multiplied many times, especially when smart cities sprang up in the midst, proliferating the use of Internet of Things (IoT), artificial intelligence (AI), Virtual Reality (VR), Augmented Reality (AR), multi-media, and Geographic Information System (GIS), etc. The electronic processing now takes place in centralized built facilities called the Data Centers. They have been dubbed as being "the heart of the digital economy" (Telecity Group, 2011)

Being capital-intensive and mission-critical facilities, data centers also incur substantial operational expenses. They should be located in suitable sites for connectivity and security reasons. Due to the rapid technological development, a data center life cycle spans from seven to twenty years (Sperling, 2010; Judge, 2017). Within this period, equipment replacement is highly possible if not inevitable. Hence, it is a costly investment, especially if self-owned by an enterprise. Environmentally speaking, data centers are not eco-friendly by their very nature, since they tend to consume lots of power for their equipment loads and there is a need to cool them. They are also estimated to be responsible for some 2 per cent of the world's carbon dioxide emissions (Hewlett Packard, 2012).

Based on the monetary and environmental costs, data center procurement is an important strategic decision for an entity needing IT services and the government as a facilitator. The cost and benefits should be weighed against each other in a systematic manner, in the light of resources deployment, potential pitfalls and the intrinsic value of the services to be obtained. Return on Investments (ROI) and Payback Period should be considered. The specific objectives of this study are to lay out a framework of cost and benefit analysis (as an aid to maximize the net benefit) for enterprises in a smart city context to consider owning their own data centers, or adopting other approaches such as co-location or managed services, whilst concerning about the effects of externalities. The method used in this study is based on a triangulation of literature review findings, supplemented by a working knowledge of data centers.

2. Benefits of Owning a Data Center

Many commercial and city management transactions (including health, transport and logistics, security, utilities, public finance, etc.) are being carried out online. When enterprises grow in size and operation, they need a dedicated enclosure with a controlled environment as a centralized location to install computing resources and telecommunication networks for processing, storing and transmitting their data. The computing resources include servers, storage and access devices, communications equipment (routers and switches), databases and software applications. Supporting systems are needed to enable the computing resources to function properly and these include power, air-conditioning, fire suppression, facility management and security installations. Depending on the level of availability requirements, redundant equipment and systems are in place, especially for back-up power provisions including uninterrupted power system (e.g., batteries), standby generators, chillers running at partial loads, etc. just in case a problem happens so that a seamless change-over can be effected within a certain time limit.

Having all the above as an in-house or corporate data center can bring the following advantages:

- Better control of computing resources and data to suit the operational needs;
- Better security and governance with full data autonomy (especially in applications with high privacy requirements, e.g., for banking, health and national security purposes);
- Direct relationship with own IT units for quick response;
- Maintaining intellectual properties of self-developed applications;
- Utilizing in-house customized expertise, also enabling its development;
- Establishing and maintaining the tiering level appropriate to meet the organization's need

3. Cost of Owning a Data Center

The total life cycle Cost of Ownership (TCO) includes the following (adapted from Patel and Shah, 2005):

$$\text{Cost}_{\text{TCO}} = \text{Cost}_{\substack{\text{building} \\ \text{\& land}}} + \text{Cost}_{\substack{\text{power hardware} \\ \text{\& back-up}}} + \text{Cost}_{\substack{\text{power air-con} \\ \text{\& back-up}}} + \text{Cost}_{\substack{\text{operation} \\ \text{\& maintenance *}}} + \text{Cost}_{\substack{\text{conversion} \\ \text{and demolition}}}$$

(*including facilities management, systems management and security related costs)

The availability of land for data center development depends on the infrastructure already in place (e.g., transport means, internet service carrier network coverage, water supply for cooling, power supply, submarine cable landing nearby, etc.), risk of natural disasters (e.g., flooding and earthquake), proximity to air-path (a negative factor) and governmental policy conducive to data center development or otherwise). Land bidding may jack up land cost if the site is well sought after by similar establishments. If existing buildings need to be converted into data center use, the conversion cost may be high due to various constraints, especially when building code is inflexibly applied.

Although data centers are functional buildings with no high aesthetic requirements, the construction cost are usually multiples of conventional buildings due to the high structural performance and equipment redundancy required. If adaptability is built in for scalability (e.g., allowing for "white space") and foreseeable changes, the construction cost may mount. Security and fire suppression installations may add to the cost due to stringent insurance requirements.

Apart from the installation cost of power supply equipment and backup (included in the building cost), the cost of power hardware depends on the capacity utilization factor (i.e., the utilized capacity to the installed maximum capacity, both measured in watts) and the power density (W/m^2). A lower utilization would inevitably under-recover the capital invested. A data center operated at a higher power density than designed would increase maintenance cost, whereas the opposite would result in the space not being used fully.

Power is consumed by the cooling equipment load, which is directly proportional to the power consumption and heat dissipation of the servers and other computer devices (the design cooling load factor). A data center operated at lower cooling capacity would also under-recover the capital invested. There is an increasing trend that each rack requires significantly more electricity power, i.e., higher power density).

The overall cost of operation includes the costs of facilities management, systems management, technical personnel, depreciation of IT equipment, software and licensing costs, as well as security-related costs. Since the requirements on data centers may change over their life time, especially in the light of fast developing ICT pace, the life cycle cost of data centers should include conversion and eventual demolition costs when the functional capability is surpassed by new disruptive technology.

The intangible cost of data centers include the additional carbon dioxide emissions, directly or indirectly arising from the high power consumption. In the event the backup power needs to operate, with on-site generators running, the emission may increase. Hence, a long-term solution is to foster the use of renewable energy (with the concomitant problems of instability and storage), which some data centers (e.g. Autodesk, Google and Microsoft, based on their offset purchases of solar and wind power) are doing but wide spread use of renewable energy has yet to be seen. Further discussions on such externalities are made in a later section.

4. Benefits of Outsourcing Data Center Services

Due to the high capital and operating costs in owning a data center, many corporates find it advantageous to outsource their data processing or computer hosting needs. They may provide their own servers and software applications, and rent the space, power and cooling of a co-location provider. This may be done on a retail or wholesale basis, meaning that space is either rented on a piecemeal fashion, usually by racks, or the whole (or part of) data center is just serving a corporate client exclusively. As such, data centers become an investment asset for the owners. According to a recent survey by a real estate consultancy CBRE (Sverdlik, 2018), an 8 percent rise in data center spending was expected in the year by the respondents. Amongst the respondents, only 5 percent of investors expected returns of 10 per cent and below. More than 50 percent anticipated between 10 to 20 per cent return, whilst 40 per cent of respondents has a higher return expectation in excess of 20 per cent. Some of the data center investors adopt a “pay as you grow” attitude, investing by increments as the client base grows.

For the users, colocation provides the following benefits:-

- Avoid high capital spending in a non-core business;
- Avoid investing in infrastructure technology (e.g., security) which will become obsolete fast;
- Able to shop for the best solution, bandwidth, power and minimum latency ;
- Achieve scalability as operation expands;
- Almost immediately deployable, with no need to wait for own land acquisition, data center construction, facilities installation and commissioning.
- Avoid having to employ experts on data center facilities management, security, and system operations

Possible disadvantages may include the lack of bespoke services in a multi-tenant situation and the response time for needed services (which may nevertheless be reduced by a carefully drawn up contract).

For clients expecting a higher level of services, the Managed Services type of outsourcing would be more suitable, since the data center provides application services in addition to equipment and infrastructure.

5. Costs of Outsourcing

Co-location services are usually paid for on a dollar per kW per month basis (with possibility of escalation in the long-term) or on racks usage basis. Depending on the agreement, charges may arise from the use of Internet routers or bridges, fiber and a local loop, which depends on the bandwidth and redundancy required. The level of occupancy of the data center may be factored in the overall rent payable. Server equipment is a necessary expenditure for clients who seek to co-locate, for which the space provider may charge for non-recurring installation. Clients may also be responsible for their equipment maintenance. IT and technical personnel is needed for regular monitoring and management of services and carrying out patching of operating systems or software applications as necessary, unless a managed service option is taken. It is also prudent to allow for "remote hand" service (chargeable on an hourly basis) to be provided by the data center. Caging up the servers or installing additional security facilities to enhance security may be arranged at the client's cost. Despite data centers running and storing the data, backing up important data outside the data centers or at the client's own premises is still advisable, which should be budgeted for.

6. Externalities of Data Centers

Apart from the abovementioned direct costs, data centers are both "loved" and "hated" by different stakeholders. Investors of data centers pour money into them as a burgeoning asset class (McAllister and Loizou, 2009). Governments like them since they attract multinational ICT giants to push up their economies. IT-savvy professionals welcome the job opportunities arising from them. Experts are not only required for the direct operations of data centers and the computer systems installed, but they are also needed to support and maintain the technologies, applications, databases, and systems software running at the data centers. Server and storage device manufacturers like their bulk purchases. IoT inventors and telecom businesses rely on them for their data exchange and processing capabilities. Smart cities need them for increasing efficiency and convenience for citizens. Construction contractors welcome the projects since they cost around US$20 billion a year to build around the world (Pearce, 2018). Utility companies like these big consumers for their electricity output. It is exactly this last attribute which make some people wary of them.

Environmentalists reckon that data centers currently consume more than 2 to 3 per cent of the world's electricity and account for approximately the same amount of carbon dioxide emission (also around 2 per cent) as the airline industry (Bawden, 2016; Pearce, 2018). Big data proliferation and the ubiquitous use of smart gadgets (including phones, tablets and health monitoring wearable devices) have created a massive growth in data, with 40,000 Exabytes running across worldwide networks predicted by 2020 (Paulley, 2017). Online booking, purchases and payment are a commonplace of people's lives nowadays. Social media and entertainments have fueled much of this data flow, as messages, video streaming, movie downloads are invoked every minutes in different corners of the world, including trains, ferries and even classrooms (with MOOCs[8] and SPOCs[9] getting more and more popular as an educational means). Cisco reckons that videos will occupy 82 per cent of the global Internet traffic by 2021 (Pearce, 2018). Moore's law has it that transistors in an integrated circuit are doubled every two years. In the US, the number of servers was reported to grow by 10 per cent per annum ten years ago (McKinsey, 2008), but this rate has certainly been surpassed by now. As the densely packed servers in data centers get smaller, they become much hotter. A typical rack is rated at 5 to 10kW (Brown et al, 2014) and can reach 20kW or more recently (West, 2018) in power consumption. Data centers consume double the amount of energy in every four years, and energy use is expected to triple in the coming 10 years (Harris, 2018). Currently, a data center typically consumes 40 to 400MW of power, with the higher end one called "hyperscale" (Stoller, 2018).

Since the majority of data centers are still powered mainly by generating plants burning fossil fuel, the increasing carbon footprint is also growing by staggering rates. For instance, Google's carbon footprint was around 1.8 million tonnes of CO_2 equivalent in 2013, which came mostly from its data centers (Vaughan, 2015), but note that it has used carbon offset in an attempt to reduce global emission. Vaughan (2015) also reported that using Gmail for a year emits approximately 1.2kg of CO_2 per user through data centers' power consumption.

Apart from the power consumed by servers, the need to have them cooled by chillers also adds to electricity use. In this connection, water consumption is also huge, as a cooling media. Fitzgerald (2015) likened the 800 data centers in California at that time to have used water which would fill 158,000 Olympic-sized swimming pools in a year. With water shortage there, water is a precious natural resource.

Less obvious externalities of data centers may be visualized by their locational requirements. They need to be sited in safe places, both in terms of avoidance of natural disasters (especially earthquakes and flooding) and human disturbance (for security sake), away from flight paths and preferably on flat land, as well as requiring political stability, stable and reliable power supply, data protection, personal data privacy, availability of dual power distribution paths, etc. These requirements seem to compete with human settlement needs. In a congested and hilly city such as Hong Kong, the need for housing is so strong that most data centers are built near to landfills instead. Smartly, the landfill gas is piped for small scale renewable production, which also mitigates explosion risk. Fortunately, in the same city, some use is made of deserted factory premises, which are converted into low-tier data centers.

7. Enhancing Benefits and Reducing Costs

Data centers are becoming essential infrastructure in a smart city. To enhance their benefits, they must be fully utilized, not only in terms of the space occupied, but also of capacities. It was reported that IT managers often purchased excess number of devices in data centers to ensure that capacity is adequate to meet the most extreme usage patterns. McKinsey (2008) estimated that daily server utilization was at most 5 to 10 per cent, wasting energy and capital. More recent figure updated the estimate to 15 per cent, but another estimate put the percentage of servers 'at comatose" at 30 per cent (Kepes, 2015).

[8] Massive Open Online Course (MOOC)
[9] Small Private Online Course (SPOC)

It has long been recommended that managers need to adopt a more aggressive attitude in the use of existing servers to make them work at higher utilization levels (McKinsey, 2008). One common technique is virtualization, which means sharing capacity by finding unused portions of servers to operate applications. To achieve this, end-to-end accountability needs to be established to avoid fragmented self-interest of different units in an organization. When doing IT equipment budgets, the life cycle costs of running applications should be included in procurement decisions across all business units.

Improving energy efficiency should be made a constant endeavor of data centers. On average, the Power Usage Effectiveness (PUE, which means the total power consumed by a data center divided by its IT load) is 2.5, whereas with best practice, this may be brought down to 1.6 or even lower. Apart from cooling equipment efficiency, managers of data centers have to be mindful of possible hot air short circuiting and cool air leakage. This can be achieved by proper containment using blanking plates and doors. A typically effective arrangement is to adopt the hot aisle and cold aisle strategy in the siting of server racks.

To save on water use, chiller-less cooling may be considered. With this type of technology still based on cooling tower, an economizer utilizes outside air, so that water is not evaporated all the time. Recycled water should be used wherever possible. Where the availability of water is a constraint, use of the direct expansion system for cooling rather than a chilled water system may be a necessity, although there may be a sacrifice in energy efficiency, but saving is achievable in capital cost at the same time.

With the advent of "edge computing" (data centers being near to the source of data), smaller data centers may be "installed", making more use of modular units. There is an upcoming trend to deploy servers and racks in container type enclosures with built-in power supply and air-conditioning systems. This measure enhances flexibility and adaptation, especially when computing needs change over time. Focus should now be put on the appropriate sizing of data center form factor (Miller, 2018). The cost of building massive data centers may be avoidable with careful capacity planning and a rigorous justification process. Allocating data center cost to individual business units may be a useful tactic, rather than offering it as a central provision.

A more drastic approach to reducing cooling needs is to choose data center sites in colder regions, so that outside natural air or stream/seawater may be used as cooling media. Examples are found in Scandinavian countries such as Finland, Norway and Sweden and also in Inner Mongolia Better arrangement should also be devised to make use of the waste heat of data centers for local community use in such cold regions.

In other cases, the use of redundant generating capacity available in data centers to serve the local power needs may be a way to alleviate the peak load on a city's power plant. This arrangement needs to be facilitated by an intelligent grid of a smart city to enable load balancing from multiple feed-in sources (Paulley, 2017).

As an alternative to conventional power, an increased use of renewable energy should be fostered for data centers, both by policy and technological measures. For the former, financial incentives such as Feed-in-Tariff and subsidies may be effective in the long term, whereas cap-and-trade or carbon offsetting may be used in the short to medium term. For the latter, energy storage at reasonable cost should be made the target of making renewable energy as a stable means of power supply to data centers.

Although the manpower deployment at a data center is not ordinarily high, their expertise (IT and facility management) is not abundantly available in the labor market. Hence, on a per head basis, they are increasingly expensive. It is expected that the use of Data Center Infrastructure Management system (DCIM) will be increasingly important to reduce the manpower needs further. There is also the tendency for data centers to draw more expertise and professionals to support the systems and applications running at the premises.

Finally, it must not be forgotten that data centers, being mission critical facilities, must be built and managed to provide uninterrupted availability. Downtime would be very costly to data centers, in terms of the possible compensation arising from legitimate claims of victims, not to mention the loss of reputation and business opportunities. An upfront operation and maintenance regime, coupled with continuous training and drills on attending staff, are necessary. Breach of security would be an important cost event too, as any hacking and data loss would entail substantial recovery expenses.

8. Conclusion

Data centers are becoming a common facet of life in smart cities. Besides being an essential infrastructure, they have formed a valuable asset class. Investment in data centers needs to be justified by their benefits exceeding the total cost of ownership. An alternative is to use co-located service, but even then the full cost must be budgeted for. As ICT flourishes in smart cities, we need to make data centers as "green" as possible, since they are competing for our scarce resources (especially water and power), and engender carbon emission at the generating plant. New developments in data center technology (e.g., edge computing and internal environmental control system, such as liquid cooling) may change the landscape of data center business. This paper highlights the benefit and cost (including externalities) considerations of data center procurement, which should be of useful reference value to facility managers and corporates which are involved in smart city development. Despite the functional orientations of data centers, a holistic analysis of the financial and societal implications is necessary.

Acknowledgement

The work described in this paper was fully supported by a grant from the General Research Fund of the Research Grants Council of the Hong Kong SAR Government (Project no. PolyU15233116).

References

Bawden, T. (2016) Global warming: data centers to consume three times as much energy in next decade, experts warn, *Independent*, 23 Jan.

Brown, K., Torell, W., and Avelar, V. (2014) Choosing the Optimal Data Center Power Density, White Paper 146, Schneider Electric

Harris, J. (2018) Our phones and gadgets are now endangering the planet, *The Guardian*, 17 Jul

Hewlett Packard (2012) Sustainability, Available at: www.hpl.hp.com/research/sustainability.html; Accessed on 8 Sept 2017

Judge, P. (2017) Data center life: going to extremes, *DCD Magazine*, Issue 22, June/July; Available at: www. Datacenterdynamics.com; Accessed on 7 Sept 2017

McAllister, P. and Loizou, P. (2009) The appraisal of data centres: deconstructing the cash flow, *Journal of Property Investment & Finance*, 27(1), 65-80

McKinsey (2008) *Data centers: how to cut carbon emissions and costs*, McKinsey & Company

Miller, R. (2018) Smart cities are coming, and they'll need data centers, available at: https://datacenterfrontier.com; Accessed 9 Sept 2018

Paulley, J. (2017) What do data centres need to provide to enable smart cities? *Lexology*, Available at: https://ww.lexology.com; Accessed 20 Sept 2018

Patel, C.D and Shah, A.J. (2005) *Cost model for planning, development and operation of a data center*, Hewlett-Packard Development Company

Pearce, F. (2018) Energy hogs: can world's huge data centers be made more efficient? *Yale Environment 360*, Available at https://e360.yale.edu; Accessed 10 Sept 2018

Sperling, E (2010). Next-Generation Data Centers, Forbes, March 15, available at: www.forbes.com. Accessed 8 Sept, 2017

Stoller, B. (2018) Equinix Says Its New Wholesale Data Center Business is Much Bigger Than Previously Thought, Available at: https://www.datacenterknowledge.com; Accessed on 10 Nov 2018

Sverdlik, Y. (2018) Survey: Investors Pouring Money in Data Centers Expect High Returns, available at: https://www.datacenterknowledge.com ; accessed on 18 Aug 2018

Telecity Group (2011) Annual Report and Accounts 2011, available at: www.telecitygroup.com; accessed on 10 Oct 2014

Vaughan, A. (2015) How viral cat videos are warming the planet, *The Guardian*, 25 Sept.

West, B. (2018) Data Center Power Density Trends – Where It's Been, Where It's At, and Where It's Going, 30 Aug., Available at: https://www.datacenters.com/news/data-center-power-density-trends-where-it-s-been-where-it-s-at-and-where-it-s-go; Accessed on 10 Dec 2018