



> APAC



HUAWEI

Towards the better Datacenter

Authored by DCD

www.datacenterdynamics.com

> Contents

Analysis: The future of artificial intelligence in datacentres	4
Main findings & analyses	6
The evolution of operations and management	8
<i>Figure 1: Digitalisation Spectrum</i>	8
<i>Figure 2: The ‘intelligentization’ of data center Operations & Management</i>	8
The development of AI to optimise cooling systems and improve energy efficiency	9
<i>Figure 4: Power consumption of major data centre components at PUE of 1.8</i>	10
<i>Figure 5: Averaged proportion of power consumption of cooling within data centre by PUE</i>	10
<i>Figure 6: Impact of machine control on PUE in Google data centre</i>	11
<i>Figure 7: Impact of machine control on PUE improvement across Google data centres</i>	11
<i>Figure 8: The complexity of data centre temperature control</i>	11
<i>Figure 9: ASHRAE optimal ‘envelope’ for dry bulb temperature against relative humidity</i>	11
iCooling in the Langfang datacentre	13
<i>Figure 10: Representation of layers within the neural network</i>	13
AI as a Means of Delivering Sustainability	14
Using AI to improve resilience and reduce risk	14
AI and Security	15
Developing datacentre strategies towards AI	17
The current state of AI/machine learning deployment	18
<i>Figure 11: Deployment of technologies across Asia Pacific data centres (January 2019)</i>	18
<i>Figure 12: Key drivers for current and intended adoption of AI within APAC data centres</i>	19
<i>Figure 13: Interest in AI conference content by market: DCD conference registrations 2018-19</i>	19
<i>Figure 14: Interest in AI conference content by sector: DCD conference registrations 2018-19</i>	19



> Foreword

Artificial Intelligence is emerging quickly as a key tool in the operation and management of datacentres. As datacentres continue to become more business-critical, more technologically complex and networked, and more reliant on empirical/data-driven rather than 'hunch' decision making so human capabilities are proving increasingly challenged in managing the datacentre environment. The complexity of management increases further as datacentres become absorbed into portfolios of digital infrastructure comprised of different storage processing and distribution environments. The complexity of management increases further as datacentres become absorbed into portfolios of digital infrastructure comprised of different storage, processing and distribution environments. Almost 20% of 600 datacentre owners and operators in a 2019 DCD survey indicate that they are currently deploying AI in their datacentres and a similar proportion have this as a future priority.

Cooling is very much a target for AI deployments. In most datacentres, cooling and heat removal is the single largest consumer of energy after the IT equipment and it is a major source of risk as too little cooling may lead to thermal events and the shut-down of IT equipment. This source of risk means that excessive or poorly targeted cooling represents a major source of waste since cooling represents an average of 40% of energy costs in a datacentre and a 2017 DCD study in South East Asia indicated that power costs could be halved if PUE was improved from 1.80 (close to the global average) down to 1.30. While a seismic improvement of this scale is unlikely, even more moderate improvements will make a difference to operational costs.

Yet the capability of making significant inroads into energy efficiency via cooling is limited using traditional means and purely human analysis and decision making. Cooling does not follow the linear and measurable paths of power or connectivity and is less subject to human operator control than IT utilisation. As densities grow and temperatures increase so cooling and heat removal will spawn equipment to deal with issues of environmental maintenance. Even when restricted to zones within a datacentre cooling is dynamic in variation. This level of difficulty makes it a key target for AI innovation since the complexities of the data collection, analysis and decision making associated with cooling means that a system capable of big data analytics, independent learning in the context of managing cooling in an energy efficient manner and the speed of decision making necessary to avoid a thermal event. This paper looks at the Huawei iCooling system and its AI foundation and its deployment in the Langfang Datacentre as illustrative of the principles of utilising AI in the quest for better energy efficiency.

Cooling/heat removal and energy efficiency are not the only areas of a datacentre which deploy AI to improve performance and this Paper looks also at its role in sustainability, risk management and security protocols. As its role in datacentre operations and management becomes more holistic, AI will move away from the current situation where a lot of the adoption is for single task 'narrow' application based on being embedded into equipment and solution sets towards a broader and 'deeper' application across the datacentre. ●



This White Paper has been prepared by Nick Parfitt, Senior Global Analyst at the DCD Group.



> Analysis: The future of artificial intelligence in datacentres

The deployment of AI in datacentres will continue to increase although, in common with other manifestations of digital transformation it will be an evolution (“developing from the ground up” to quote DCD editorial) rather than a revolution. There are a number of factors that will determine the speed and penetration of AI within datacentres;

1. The business case that AI makes for datacentre investors. Some of the major savings on power and operational efficiency have been noted in very large and hyperscale facilities. As AI matures as a set of technological options and as pricing levels reflect that level of maturity so the business case for smaller MTDC and enterprise facilities will gain further traction.
2. The number of datacentres that will be able to deploy AI to a level where it will effectively be able to represent a return. There seems little doubt that it seems to fit with the hyperscale operating model although not all hyperscale facilities in recent DCD research appear to have yet introduced it. Its penetration is likely to be lower among MTDCs and lower still among enterprise. While the case is made by some analysts that AI can be deployed at reasonable cost in older facilities, the question of whether the improved operation of dated or legacy environments may not simply throw into sharper relief their limited profile. However, the industry may reach the stage where the older and smaller facilities for which AI is not a realistic or affordable upgrade may not be facilities that are viable to operate anyway. ►

3. ► The extent to which AI replaces human skills or supplants them, and the extent to which it moves the industry towards 'lights out' facilities. The skills trend in datacentres over the past 5 years has been from working inside the datacentre to working across data infrastructure and interpreting business requirements, and there are still major local shortages of skills in both these areas of activity. It is possible that without major change in how the datacentre industry sources and upskills its staff, that AI may keep the industry moving rather than representing a threat to employment
 4. The deployment of AI will affect the operation of the datacentre on critical 'soft' issues such as risk management, performance, data and privacy practices, quality standards, compliance and accreditation. Some of these factors become far less clear-cut, partly because the sheer volume of data and analytics makes the tasks of overseeing the processes more complex although effectively at the end of the day there will be AI systems required to oversee other AI systems and check them for compliance. New cloud-based and AI-driven services such as datacentre management as a service (DMaaS), take huge amounts of data from datacentres and apply the shared learnings to individual situations and customers. But as with Edge computing, this creates a new legal conundrum as to who owns the data, particularly as the latest data legislation such as GDPR shifts the ownership principle from the company or individual harvesting or collecting the information to the company or individual providing it. The major case that prevented Facebook from transferring data on an EU system out of the EU indicates the extent to which the balance has changed.
 5. The degree of self-learning and automation associated with AI may make visibility more difficult. Therefore the deployment of AI and deep learning systems are most successful when the output is defined as part of the deployment process rather than after. This needs to view how machines are reaching decisions.
 6. The most major trend in AI adoption in datacentres will be away from single task/purpose applications towards deeper learning and overall management systems. This is not actually a new situation – currently datacentre operators can be frustrated by needing to deal with solutions and systems from multiple suppliers and this may be continued as suppliers embed AI into their products or systems. It is likely that to overcome this situation that an over-arching AI system to which all others are accountable. These examples can be as 'task-driven' AI in the context of the datacentre. Examples of more 'general' AI applications have emerged more recently such as the latest version of Huawei's PostgreSQL-based GaussDB claimed by the company to be the world's first AI-native database. Specifically, it adds a host of AI-based capabilities such as the ability to self-tune, automatic fault diagnosis, and some self-healing capabilities. AI acts as a performance improver in relation to GaussDB. In terms of self-tuning skills, tests indicate an AI-tuned GaussDB configuration performing up to 50 percent better than manually tuned configuration as well as configuration produced by popular machine learning-based OtterTune tool for both OLAP and OLTP databases. GaussDB supports multiple deployment scenarios, including deployment on a standalone server, as part of a private cloud, or on Huawei's public cloud. On Huawei Cloud, GaussDB is used to power various data warehousing services for its cloud customers.
 7. Technological advances in AI will include the further development of natural language processing to support 'conversational' AI. The future will see also advances in 'self-healing' datacentres. Data throughput within the datacentre will need to deal with the added requirements of the AI system and this will require increases bandwidth and speed specifications.
 8. Any future deployment of AI and machine data will need to accommodate the major emerging trend towards more distributed and 'Edge' computing. This model requires processing capacity at the point at or close to where the data is collected and a scalable source of shared information and learnings at the 'core'. The curation, management and storage of the huge amount of data that will be generated in terms of deciding what needs to be transferred to the core and what can be discarded can only be effective through a self-learning and auto-programming system. Both are requirements that can only really be fulfilled by AI with a key machine learning application. The rolling out of Edge computing will add further urgency to the deployment of AI since Edge deals with the complexity of data collection, analytics and intelligent decision-making which AI is already executing inside enterprise and service datacentres and across networks. The volume and complexity of data that Edge will generate means that legacy statistical systems will not be capable of the analytics required, and that a system that can develop an adaptable learning capacity will be the only means of running the systems.
 9. Operations and Maintenance: AI networks can also help improve equipment maintenance. Scheduling maintenance according to manufacturer guidelines is effective but costly. Many experienced datacentre engineers will tell stories of how they could tell that a piece of equipment was faulty just by the way it sounded or smelled. What if a trained AI network could predict those failures long before they were detectable to engineers? Major data centre operators are beginning to apply these type of deep learning networks today to monitor different types of machines.
- Just as it has proven difficult to provide a viable figure for the volume of data that new advances in technology will bring, so it is difficult to project the level of increase in future investments. The OECD at December 2018 indicated considerable growth in private investment from 2017 onwards and Gartner's 2019 CIO Survey indicates a growth rate of 270% in implementation since 2015 and 37% from 2018 to 2019. Gartner estimates that the enterprise AI market will be worth US\$ 6.14 billion by 2022. ●

> Main findings & analyses

What is Artificial Intelligence?

Artificial intelligence [AI] is broadly agreed to represent the principle of intelligence as demonstrated by machines in contrast to the 'natural' intelligence displayed by humans and animals. In demonstrating intelligence, the machine or device is able to perceive or sense its surroundings and adapt or take actions that maximize its chance of successfully achieving its goals or tasks. A machine or device which has an AI capability has the capacity to learn independently (hence the term "machine learning") and to increase its intelligence on that basis. AI is data collection, analytics and processing driven evolved to the extent that learning can occur independently and as part of a continuous process. The principles of AI are that it makes it possible for machines to learn from experience, to adjust to new inputs and perform tasks in a way that mimics humans. There is not one system or one capability that defines AI – it is more of a spectrum. There are intelligent machines that are not AI-enabled but which respond automatically to stimuli, for example, automated or temperature-sensitive alarm systems in buildings. While AI enables new forms of automation, there are more basic forms of automation that have existed for as long as machines.

Some commentators divide AI into 'narrow' (or 'weak' AI) which is limited in scope and usually focused on a single task based on the simulation of human intelligence and 'artificial general' (or 'strong' intelligence (AGI) which has the capability of intelligence that can be applied to solving any problem. The latter is a closer replication of human capabilities. It is based on machine learning which feeds data to a computer and uses statistical techniques to help it "learn" how to get progressively better at a task, without having been specifically programmed for that

task, eliminating the need for millions of lines of written code. Machine learning can consist of both supervised learning (using labelled data sets) and unsupervised learning (using unlabelled data sets).

While 'artificial intelligence' and 'machine learning' are used in a similar context, they are different things. 'Artificial intelligence' represents a capability whereby machines can process data in a manner taken as 'thought' and act upon the knowledge to conduct a task according to unique requirements. 'Machine learning' is a key process in developing artificial intelligence whereby data is provided to the machine once the capacity for AI has been established and the machine can learn and enhance its knowledge base from that process. In reference to this learning process a further attribute of 'depth' is expressed. Machine learning, as an enabler of AI, empowers businesses by giving their computer systems the ability to "learn." Through algorithms that allow systems to recognize patterns and automatically build analytical models, computer systems in datacentres can now be augmented with the ability to make critical decisions with less human intervention.

These analyses reflect the development from narrow into more general intelligence, increases in AI learning capacity and the directions in which AI is moving (IQ to EQ?). It is clear that AI is NOT a 'black and white' deployment whether in a datacentre or anywhere else; it is not like installing a new cooling system or even new DCIM software. The uptake of AI can be visualised more as progress along a spectrum and this is true particularly because the technology itself is still a work in progress, still evolving and therefore still to move into uncharted territory. ●

The Drivers towards AI in the datacentre

Working datacentres are rarely environments where technology is adopted for technology's sake. They are usually mission critical facilities which are expensive to operate but potentially far more expensive if for whatever reason they fail. Therefore, the deployment of technology, the management and operation of the datacentre is of primary importance to ensure availability. One of the acknowledged consequences of digital transformation (the techno-trend of which AI and machine learning can be considered major contributors) is that technology and datacentres are required to integrate more closely with business needs and this can be seen in the changing drivers. AI needs to deliver not just on datacentre management and operational requirements but on business requirements as well.

The drivers towards AI in the datacentre vary between different types of datacentre, and three main types are referenced in this Paper.

Enterprise datacentres exist to service the needs of one organisation whether private or public. These may be on-prem or collocated and managed by company staff or outsourced facility management. This category includes also datacentres run by IT and managed service companies to offer their services, although once the services offered move into the facility sphere, they would be included usually under MTDCs.

These account for over 95% of all datacentres by number but this includes multitudes of smaller, older ("legacy") datacentres. One of the major decisions faced by a large number of companies ►



► which run their own datacentres is whether they should continue to do so or whether it would make better sense to outsource or to deploy cloud. If the decision is made to continue with the on-prem datacentre then AI may find a role in delivering the extended life, operational performance and efficiency to justify that decision.

The key concerns of those running enterprise datacentres from DCD research in 2019 reflect the balance between the needs to plan and manage risk and costs (of both operations and upgrade) and to protect also against cyber-threats:

- Reducing operational costs
- Cyber-threats
- Conducting capacity planning
- The costs of upgrading and expanding datacentres, and
- Risk management strategies.

Multi-tenant datacentres [MTDCs] are service datacentres mostly run to service the IT needs of external clients on a commercial basis, although some larger enterprise facilities may operate on an MTDC basis of charging different client business units within a company. These facilities share raised-floor space, power and cooling between tenants. The portfolio of such facilities may include a variety of standalone datacentres, server rooms and disaster recovery sites. They may be built as a campus, as a dedicated purpose-built facility or, in a building shared with other commercial or industrial activities. However, the MTDC world is generally populated by smaller facilities and smaller companies than the wholesale and hyperscale sectors. MTDCs are largely unable to compete with the flexibility and scalability of cloud or with its basis of charging based on what is consumed, and datacentres based on offering 'retail' colocation have suffered accordingly. However, the relationship between different infrastructure options is dynamic and the response among all types and sizes of MTDC has been to move from a facility that is operated or leased on the basis of racks, power, connectivity and security to one that offers access to services either through in-house ecosystems or through various modes of cross-connectivity. Not all MTDCs have the facility profile to do this effectively, but colocation providers are able to offer a business case based on the argument that it is more cost effective to access a range of IT and cloud services via an outsourced facility that is able to access and provide these rather than to upgrade an on-prem. datacentre at considerable cost.

A majority of MTDCs researched offer disaster recovery services, access to ISPs, dedicated hosting, application hosting and SaaS. This indicates a fundamental transition from a business model based on infrastructure rental, to one based on IT resource rental based on interconnectivity and then to a model offering IT resource as application services based on cloud infrastructure. Different MTDCs are at different points along this spectrum. Some have had no need to move away from the original colocation model while others will have moved through the spectrum towards a service model.

From DCD research, the key concerns of the MTDC sector in 2019 include:

- Reducing operational costs
- Conducting capacity planning
- Meeting the needs of different customers
- The shortage of suitably trained and skilled staff
- Understanding new technologies
- The costs of upgrading and expanding datacentres, and
- Increasing rack densities.

The drivers of planning and cost management are balanced here against the need to meet the needs of different customers and a need to understand how new technologies, including AI may help in this.

Cloud datacentres are those operated by cloud providers for their own commercial purposes. However, major cloud providers and the wholesale colocation providers that build to provide for cloud, have developed a new class of datacentre – 'hyperscale'.

For hyperscale, the business model is the performance of the datacentre itself. The key values of hyperscale include scalability, resilience, connectivity, efficiency, replication, predictability. These are datacentre attributes that can be and are measured. They are based directly on customer demand as these are also key customer expectations of a service that is there when you want or need it, that is responsive, instantaneous and predictable. Efficiency is a consideration of operation and profitability although there is growing interest in energy consumption among Governments, media and the public and some of the protagonists in this space wear their PUE as a badge of pride.

The key requirements of the cloud and hyperscale providers from DCD research in 2019 include cost management and planning, particularly in the context of the increased demand for the cloud services they provide:

- Reducing operational costs
- Conducting capacity planning
- The costs of upgrading and expanding datacentres
- Meeting increased demands for cloud services
- Meeting compliance and accreditation requirements.

These findings from different types of datacentre indicates common drivers across all of them particularly in regard to costs of operation and upgrade, capacity planning and risk management. While all datacentres are based on a similar purpose to house IT, their business functions, technological profiles and means of delivery vary considerably. As with any other input into the datacentre – power, cooling, monitoring, for example – the deployment will be individualised according to the environment and the requirement. For enterprise, the drivers are based on efficiency and business justification; for MTDC it is the need to fit differing client needs under a standardized business model and for hyperscale it is dealing with the sheer scale and technological requirements of growth. Therefore, differing contexts mean that there is no "one size fits all" for thinking about AI in the datacentre. ●

> The evolution of operations and management

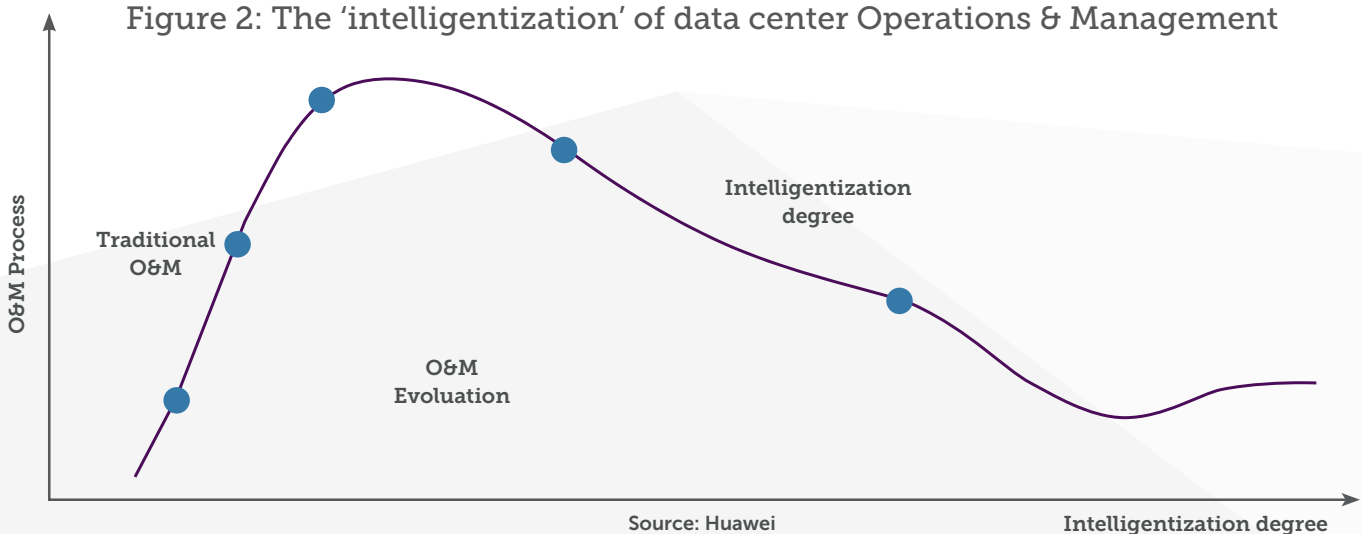
Since the principle of AI is the replication of human understanding, any of the operational and management functions that people undertake in the operation of a datacentre may, in time, be replaced or supplemented by AI. As such, AI is seen as an answer to the skills shortage although now, AI appears currently more to be replacing outdated or inferior systems than it does to be replacing human labour. In its Paper on the evolution of Digital Transformation, DCD identified a spectrum of progression that companies were taking in order to transform towards a digitalised world. The spectrum moves from companies that have no strategy in place, through an ad-hoc stage where technologies are adopted but without a clear path forward through phases where technology adoption, business objectives and processes are linked more closely together. Note that the evolutionary process described here can be applied to specific technologies and business requirements or on an integrated whole of business basis.

Figure 1: Digitalisation spectrum



Source: DCD Research/Highways & Cul-de-Sacs (2019)

Figure 2: The 'intelligentization' of data center Operations & Management



A similar process exists for the running of datacentres and Huawei have developed a model to chart this evolution of datacentre Operation & Management [O&M] from 'conventional' O&M to 'smart' O&M. This process is not linear – progress over time towards smart O&M is indicated by the x-axis while the y-axis indicates the comprehensiveness and complexity of the O&M process.

In the intelligent O&M phase towards the right of the chart, digital and intelligent means are used to solidify and simplify processes.

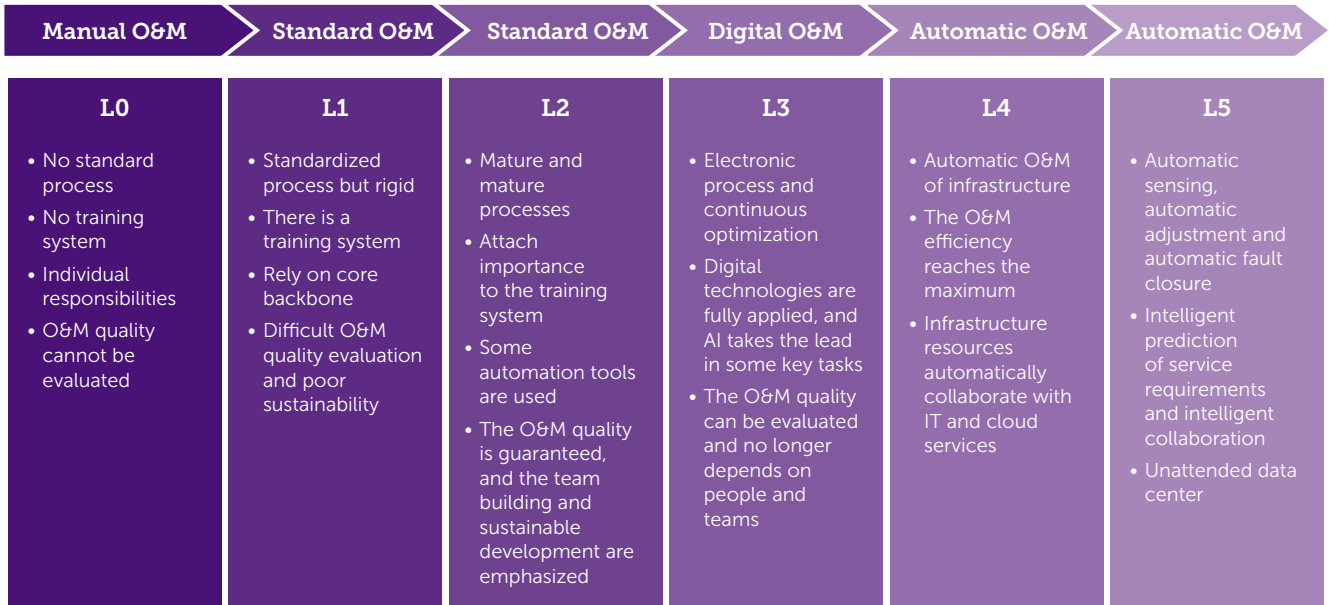
The steep path at the beginning of the process is due the lack of intelligent means and dependence on the experience and expertise of the O&M team. Sustainable management

will depend therefore on the continuous improvement of the training system. O&M efficiency decreases to some extent along with the continuous improvement of the process. However, as the O&M team gets more familiar with the process, O&M efficiency will be restored.

The trend shown here needs to work with possible constraints which arise from the habits of the conventional O&M phase. These constraints may include an overdependence on particular O&M teams or individuals that may impede the establishment of the "intelligentization" process and the development of shared experience and accumulated knowledge; the rigid use of processes which will eventually cause the O&M team to lose patience with the overall process if those don't reflect the actual O&M situations that the team meets. ▶

► The profile of the 6 stages in the evolution of conventional O&M towards smart O&M in the Huawei model are indicated below:

Figure 3 : The evolution of O&M



Source: Huawei

The evolution moves through:

- the increasing standardisation and maturity of the O&M process
- the move from human responsibility towards digital, automated and then AI options and eventually a 'lights out' datacentre

- The capability of evaluating the quality of O&M delivery

So, as part of the processes described above, in which situations have datacentres deployed artificial intelligence and to meet which needs? ●

> The Development of AI to optimise cooling systems and improve energy efficiency

Use is made of AI in datacentres to meet the key challenges of high operating costs, improved energy efficiency, capacity planning (being able to work out how the datacentre needs to be resourced to meet demand), the costs of upgrade and the development of risk management strategies.

The more general deployment of AI occurs across the tasks associated with datacentre facility management, usually in conjunction with software defined infrastructure, DCIM or equivalent software, cloud deployment and application development. More targeted cooling and improved energy efficiency is the major target of these initiatives.

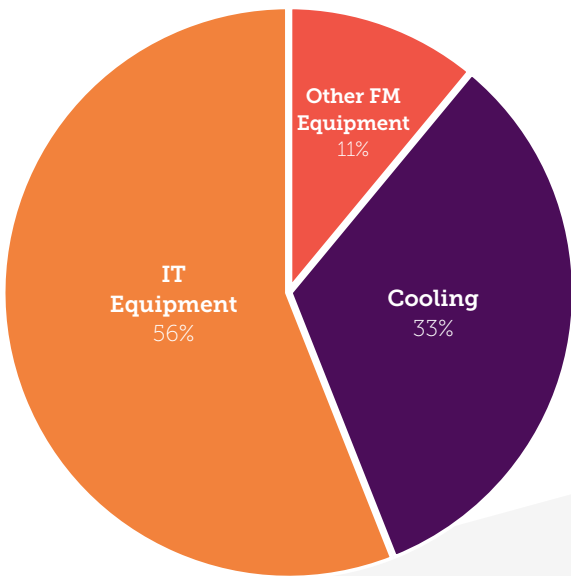
A primary driver for this approach is that cooling in most datacentres the major user of power after IT and secondly that power is the single largest operational expense, accounting for a stated 35% to 40% of operational costs based on DCD Research conducted over the period 2015 to 2019.

In the datacentre energy consumption structure, IT device power consumption and cooling power consumption are large users. According to the heat dissipation requirements of datacentre devices, the power consumed by IT devices is converted into heat and needs to be balanced by cooling capacity, the ambient temperature of the meets the requirements of the equipment. When the energy consumption of IT equipment is limited, the ►

► energy consumption of cooling equipment can be reduced by optimizing the cooling system.

The PUE of a datacentre is a comprehensive evaluation indicator. At a PUE of 1.8, 56% of power into the datacentre will be consumed by the IT equipment, one-third by cooling and the remainder by other facility equipment.

Figure 4: Power consumption of major data centre components at PUE of 1.8

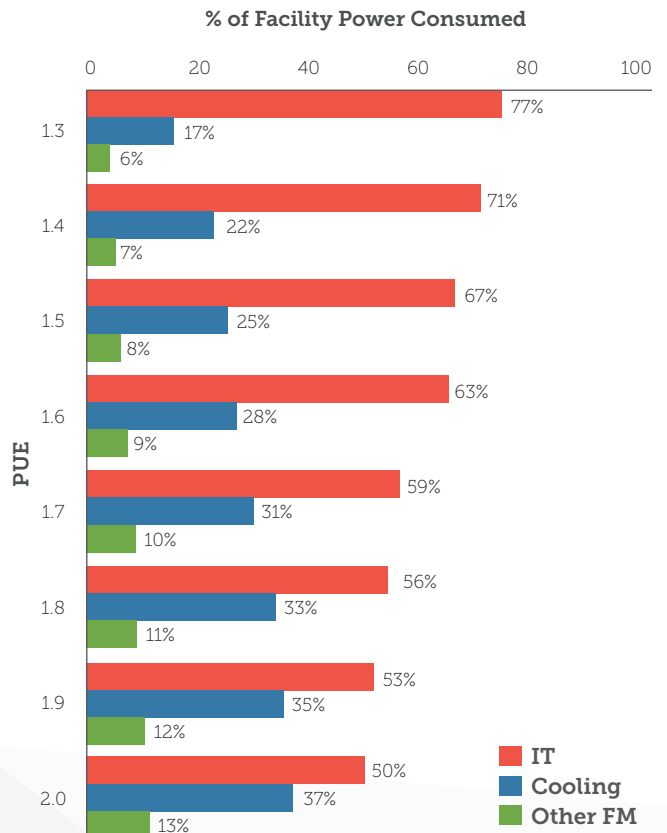


Source: DCD Research 2016-18

Due to the association between cooling and device heat dissipation, device configuration, equipment room environment, and atmospheric conditions, after the O&M reaches a certain maturity level, the manpower or expert experience cannot meet the requirements for further reducing energy consumption, for example, a small increase in the temperature of the cold aisle will cause many changes in the cooling system. For example, the power consumption of the cooling system, cooling tower, heat exchanger, and water pump will increase or decrease, and the power consumption is not linear. The result may be that the temperature of the cold aisle increases and the total power consumption increases. The interaction between refrigeration and electrical systems and various complex feedback loops make it difficult to accurately derive the efficiency of datacentres by using traditional engineering formulas.

DCD uses information collected from DCD global, EU and APAC surveys and this suggests that for a datacentre which has an overall operational cost of \$1,000,000, the amount spent on power for cooling can be halved between a PUE of 1.8 to a PUE of 1.3. It should be noted that these calculations are indicative only and do not make any assumptions of other efficiency measures such as improved utilisation of IT equipment.

Figure 5: Averaged proportion of power consumption of cooling within data centre by PUE



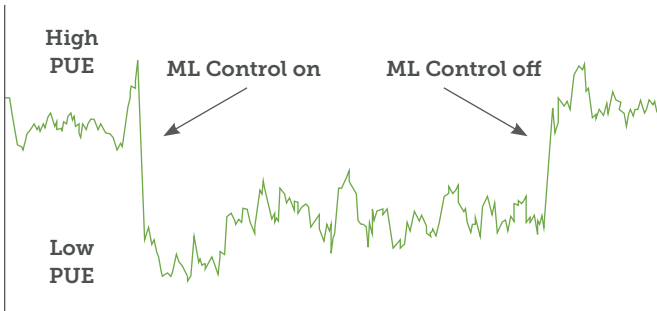
Source: DCD Research 2016-18

Some of the earliest deployments of AI for the management of cooling have come from Google which has over the past decade developed algorithms to improve the accuracy and relevance of its search functions. Within its datacentres, it has used DeepMind AI to cut its datacentre energy bills by putting the AI system in charge of power use in parts of its datacentres. This has led to a reduced requirement of power for cooling which is in most datacentres, the largest non-IT consumer of power. To achieve this, the neural networks control around 120 variables and take data from sensors located across the server racks.

The system is defined as 'self-improving' – the analysis of data allows further sensors to be deployed to improve the accuracy of the intelligence. Google claimed to have achieved a PUE of 1.12 across its Datacentres by 2014, an average that it has maintained and improved in subsequent years and it has published the charts following to illustrate the impact of the intelligent system it has developed on its PUE metric.

The first chart indicates the impact on PUE of turning the machine learning control 'on' and 'off'; the second the broad reduction in the PUE scores achieved by Google's datacentres from 2008 and 2014. ►

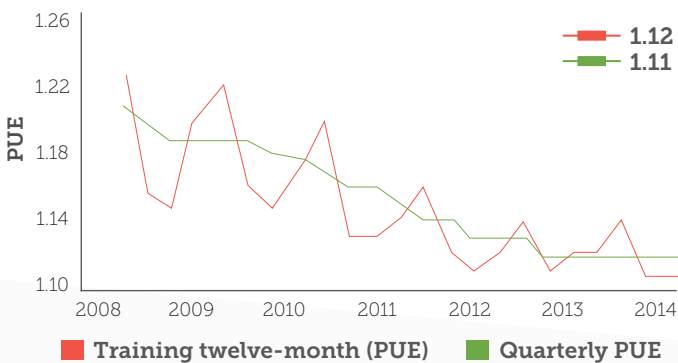
Figure 6: Impact of machine control on PUE in Google data centre



Source: Google

<https://deepmind.com/blog/article/deepmind-ai-reduces-google-data-centre-cooling-bill-40>

Figure 7: Impact of machine control on PUE improvement across Google data centres



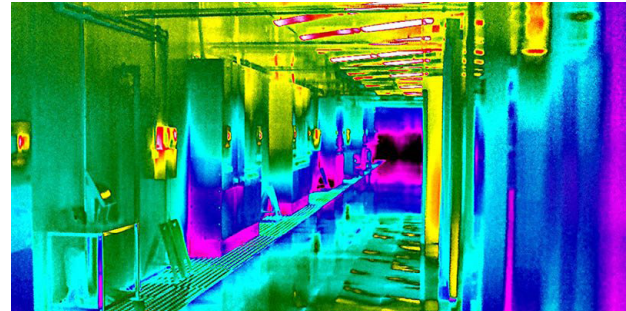
Source: Google

<https://deepmind.com/blog/article/deepmind-ai-reduces-google-data-centre-cooling-bill-40>

► Google's commitment to reducing cooling energy use has continued through the development of 'DeepMind' AI which it is claimed has reduced Google's datacentre cooling bill by 40%. Assuming the PUE of Google's datacentres to be consistent with 2014, cooling would represent around 12% of its datacentre energy consumption, and therefore a 40% saving on cooling would mean a 5% cooling across their datacentres as a whole.

The importance of cooling efficiency to improving datacentre energy efficiency is rarely questioned in the datacentre industry. The challenge is how to achieve this. The efficiency of cooling is based on a complex set of measures in comparison to other types of facility equipment. It needs to ensure the correct temperature across volumetric space (whether that is defined in terms of rooms, racks, rows or single processing units). The classic illustration of temperature in a datacentre - the ever-moving sets of colours where red or purple indicates the higher temperatures and higher risk of 'thermal events'.

Figure 8: The complexity of data centre temperature control

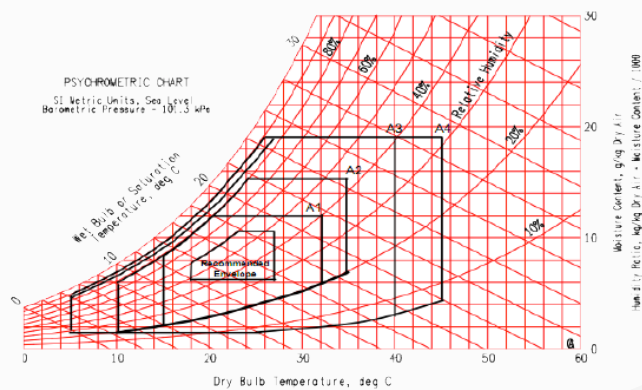


Source: Infrared Datacentre Survey; ired.co.uk

The uncertainty around how best to deal with cooling can be illustrated by the continuing reluctance of datacentres to run their IT equipment at the higher temperatures which are permitted by the manufacturers of that equipment and which are recommended also by ASHRAE [American Society of Heating, Refrigerating and Air-Conditioning Engineers]. ASHRAE has since the publication in 2004 of "Thermal Guidelines for Data Processing Environments" made recommendations as to the environmental operating conditions for datacentres. These recommendations have been updated three times since. With each succeeding edition, the facility environmental envelope ranges have broadened in response to the increased environmental ruggedness of newer generations of IT hardware. These broader ranges have allowed the facility operators the opportunity to improve their cooling energy efficiency. The industry reliance on these ASHRAE guidelines has allowed datacentre facility managers to consider increasing the operating temperatures and adjusting the humidity ranges to save energy, while considering any effects on IT equipment reliability.

The optimal area for dry bulb temperature against relative humidity is termed the 'envelope'. ►

Figure 9: ASHRAE optimal 'envelope' for dry bulb temperature against relative humidity



Source: ASHRAE

For further information visit www.ashrae.org

► Research conducted in South East Asia indicates that a majority of datacentres kept their IT equipment below the minimum recommended ASHRAE temperature. The key reason for this is the continuing concern about thermal events which might raise the temperature of the IT equipment to a level which would threaten their operation.

Wastage from power loss across the distribution chain can be measured at points along a linear system, and wastage from inefficient IT utilisation can be measured through the metrics collected on servers and processing work each achieves. Compared to these more linear datacentre systems, cooling is inherently less 'intelligent'. It is usually based on different types of equipment linked together to cool and to remove heat rather than a single unitary system. As cooling is one of the key inputs into the whole datacentre environment it is also more difficult to manage using traditional manual approaches. The classic datacentre power management system is based on the availability of back-up power sources and the 'N' classification system reflects the level of redundancy of power. Cooling is also included in these classification systems.

This process of management needs to pay attention to the two key requirements. The first is the optimization of the working status and energy consumption of the equipment. This means ensuring that each component of the cooling system runs in an efficient range according to the natural curve of the equipment. This will prevent 'thermal events'.

The second is the optimization of the system formed by the equipment working in combination. In cooling, different sets of equipment work together and the principle is to identify the optimum combination of components in the refrigeration system. In a typical cooling system, there are dozens of controllable variables on CRAC, fan, pump, chiller, cooling tower, and other hardware across the system. Different configurations across the cooling system may provide a similar cooling capacity, however, the energy consumption of different configurations might be quite different. For example, how often are the cooling towers and the cooling pumps used when a 1000KW cooling capacity is required? Which combinations and weightings of equipment deployment are more energy-efficient for different requirements? Can a situation be achieved where power provision is reduced to certain devices while increasing the overall cooling capacity across the system?

Optimization is difficult to measure and control using manual intervention since the dynamics of IT load and environment are too complex to track and even the most experienced engineer is not competent to precisely control all the variables simultaneously and instantly and as human error is still the primary cause of datacentre outage, traditional management methods continue to be followed as no one wants to take the risk of downtime.

While maintaining the capability of cooling to meet the environmental requirements of the datacentre AND do so more efficiently has been a priority for datacentres, the technology that will enable this to be achieved seamlessly has only now started to be more generally available based around the evolution of AI as a

tool for managing the datacentre.

The principle of the use of AI to increase cooling efficiency is based on a process whereby the algorithm obtains data, models, algorithm framework, management components, intelligent services, and application integration. Data is acquired via data collection, data processing, and data storage. Statistical modeling and the basic algorithm mainly resolves the sampling and cleaning of big data, obtains the optimal operator, and supports the quick solution of the model. The algorithm framework then selects the machine learning algorithm based on the logical association, performs pattern matching, and finds the applicability of the algorithm.

Considering the complexity of the datacentre cooling system, the system data needs to be obtained for the electrical system, cooling system, and environment parameters to find the system feature values and use the feature values to organize the DNN network. In a datacentre, there are many systems related to cooling. The data to be provided to the algorithm that informs the AI will come from a number of sources including big data generated by the BMS, PMS, and control system.

To track this kind of complex system, an effective algorithm needs to be found which has the necessary level of data assimilation, learning and application to optimise overall performance. Big data and AI have become looked at more closely as a means for enabling energy efficiency optimization.

The most recent incarnation of AI uses historical data to train a neural network, output the predicted PUE, and analyse the relationship between the PUE and data generated from particular components within the datacentre. The machine learning algorithm can be used to find the relationship between different devices and the parameters of different systems. A mathematical model is then established by using a large amount of sensor data in order to understand the relationship between operational parameters to find the optimal configuration to achieve the required outcome.

Using its analysis of this set of correlations, it is then able to guide the datacentre in performing the appropriate optimization control tasks according to the prevalent climate and load conditions, in order to achieve energy saving targets.

What is new about this form of AI? A neural network is a type of machine learning algorithm, which simulates the cognitive behavior of interactions between neurons in the human brain. To improve the cooling efficiency of the datacentre, a neural network is used. The neural network has an input layer, an output layer, and multiple hidden layers, and an input characteristic vector which is transformed to an output layer by means of implicit layer transformation, and a classification result which is produced at an output layer. Multi-layer perceptions and analytics can get rid of the constraints and bias of early discrete transmission functions as well as enabling continuous functions such as sigmoid or tanh to simulate the response of neurons to excitation and use the reverse propagation BP algorithm for training. ►

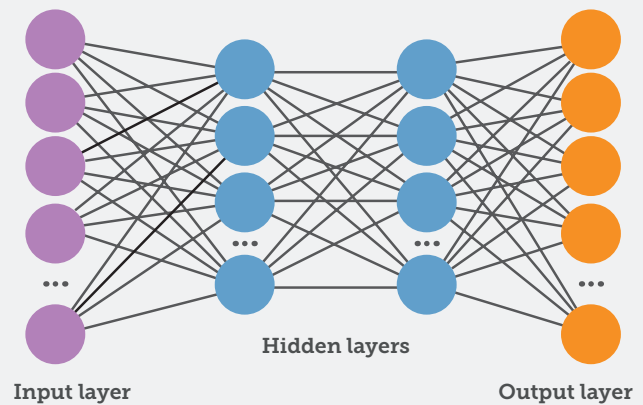


► The paradigm of developing AI for use in relation to cooling may be seen in Huawei’s development of the iCooling system. This is based on the wholistic approach in order to further optimize the relationship between the system working status and energy consumption by considering the optimization of two layers.

The AI provides training and inference platforms. The inference platform (bearer model) can be deployed independently from the management system to the physical machine, cloud (if a cloud deployment mode is used), or the management system. The training platform is deployed independently.

The management component manages the life cycle of a model based on the model evolution, such as releasing a new model and rolling back the model and supports the evolution of models after learning. The intelligent service is used to preset the inference model, recommend and predict the model, and make decisions on adjustable parameters. In the actual running process, the parameter group that can be adjusted can be obtained through the decision system. The integration service is used to visualize AI services, adapt scenarios of different cooling systems, and visualize data process analysis and control effects. ●

Figure 10: Representation of layers within the neural network



Source: r-bloggers.com

> iCooling in the Langfang datacentre

Huawei have deployed the AI-based iCooling technology at the Langfang Cloud Datacentre. Langfang is Huawei’s major enterprise datacentre in Northern China. 4,500 racks with a planned capacity of 36MW have been rolled out in the first phase of the project. The facility consumes over US\$ 20 million worth of electricity annually at its original PUE of 1.42.

According to the datacentre’s next 10-year plan, its capacity will grow 10-fold, and it will be capable of hosting 1 million server units. The considerable amount of energy being consumed and that will be consumed through expansion imposes a considerable burden on the datacentre’s operating costs. Any technology which can effectively reduce the energy consumption and thereby reduce also operating costs is therefore of considerable interest.

At Langfang, many of the conventional approaches to improve datacentre energy efficiency have already been taken. These include the management of in-row air conditions, high-efficiency power supply and distribution, and cooling based on aisle heat containment. Indirect evaporate free-cooling (AHU) will be deployed soon.

What Huawei faces at Langfang is by no means unusual. As the datacentre becomes the key enabler of digital transformation for enterprises, so their high energy consumption has drawn extensive attention not only from datacentre owners, but also

from concerned authorities and from environmental groups. In China as in other nations across the world, regulations have been introduced by local authorities in metropolitan areas such as Beijing and Shanghai which mean datacentres need to conform to legislated criteria in terms of their energy efficiency and for their power consumption. The Huawei Langfang Cloud Datacentre has therefore been designed to comply with such regulations.

The balance between meeting increasing service requirements while limiting energy consumption is the balance that datacentre owners need to achieve. For the Huawei Langfang Cloud Datacentre, this means not only cutting power costs which will increase by millions and possibly tens of millions of US dollars annually, but also achieving a better PUE that will enable the facility to host more IT infrastructure while still using a pre-determined amount of power.

But Huawei determined that the core of the solution to meeting these issues was to look again at the traditional measures for managing cooling. While these may have proven effective in the past, they were not sufficient to deliver the necessary step-changes in cooling management to improve energy efficiency. The logical approach to reducing the energy consumption of the cooling systems is to establish how the cooling system should correlate to the IT load dynamics and to control it on that basis. To address Langfang’s energy consumption challenges ►

▶ particularly in the context of cooling the datacentre, an AI-based system control – iCooling – was therefore introduced. The capability of AI to optimise datacentre energy efficiency is based on the quantity and quality of historical data used to inform and build the predictive model. At Huawei Langfang Cloud Datacentre, data is collected from over 700 points and 21 variables have been identified as those parameters which correlate most strongly with energy efficiency as measured by PUE. The statistical relationship has been modeled using the deep neural network algorithm, and the model charts the extent to which each of the variables contributes to the overall PUE score to an accuracy and at a speed beyond human capability even that of the most experienced datacentre engineer.

The 'trained' predictive model is also verified by comparing the predicted PUE to the actual running PUE, if the deviation is under ± 0.005 , the prediction is regarded as accurate. Once the accuracy of the predictive model achieves a certain level, the model is applied for energy efficiency optimization.

iCooling was first deployed on 1,500 racks at the Huawei Langfang Cloud Datacentre in May 2018. The yearly average PUE

in the pilot zone (1,500 racks) has been optimized from 1.42 to 1.304, representing an 8.17% saving in electricity consumption. At the current stage, the accuracy of PUE prediction has been modelled to 99.5% at a tolerance of 0.005. This represents the accepted level of variation between the predicted PUE and the actual PUE.

Running on the Huawei DCIM system, the inference platform collects the real-time operational data from datacentre and identifies the best cooling control strategy for the different condition (IT load, environment, and etc.). The control order for cooling system is sent to the Building Management System (BMS) to actuate the cooling system. Huawei works closely with the top BMS vendors to ensure the compatibility, so that the iCooling solution is able to be widely applied to most datacentres. Currently, iCooling is being delivered to datacentre operators, for instance, the China Unicom Zhongyuan Data Base, and China Mobile Xiamen. iCooling is also potentially adaptable to smart building energy management and Huawei is developing the 'Smart Campus' project which monitors and optimizes electricity usage in industrial parks, universities, and campuses. ●

> AI as a means of delivering sustainability

More recently and as the focus of datacentre best practice energy use has moved from energy efficiency within the datacentre to how sustainable is the source of energy, Google's AI subsidiary DeepMind has developed a machine learning algorithm to predict the productivity of wind farms up to 36 hours in advance. The system is currently used across 700MW of wind power capacity the company uses to run its datacentres and offices in the US, and it allows hourly delivery commitments

to the power grid to be committed a full day in advance. The company claims that it has boosted the value of [its] wind energy by roughly 20 percent. Other 'cloud giants' are using AI for similar purposes to Google. Conscious of sustainability issues, Microsoft has launched a new data-driven circular cloud initiative using the Internet of Things, blockchain and AI to monitor performance and streamline the reuse, resale and recycling of datacentre assets, including servers. ●

> Using AI to improve resilience and reduce risk

Not all the focus of AI is on energy efficiency or on sustainability. One key application of AI applications is based on their ability to substantially decrease the risk of downtime. Currently, downtime is one of the costliest occurrences, not only for datacentre operators, but for their clients as well. DCD research on downtime indicates an average cost of failure at just over USD 100,000 and a number of high profile downtime incidents in 2018/19 (Microsoft, British Airways, Amazon, Google, O2,

CenturyLink) indicate that the costs do not take into account damage to brand or reputation, staff morale, losses to stock values, costs of remedy and repair. AI as a driver of new security protocols can represent also a step toward maintaining availability.

AI software is also in use to assist skilled technicians and engineers in individually monitoring high numbers of generators. ▶

► This added monitoring is used preventively as well as during the critical generator monitoring in the event of a potential utility outage. Without AI, maintaining this level of careful auditory vigilance would be a near impossibility where there are large numbers of generators to monitor while utility is restored. The software in use provides an additional layer of automated surveillance to serve as an extension of the facility's technical support team. The sensors track variables to sound patterns to potentially identify and solve an issue or even predict failures before they have the chance to cause downtime. The AI's more minute sensory experience, along with its application of predictive modelling, allows the team to have eyes and ears in more places at once. While human senses may not be able to pick up the small noises that could indicate issues, the software can detect them and predict the need for preventative maintenance or immediate attention using its learned algorithms.

These types of ground-breaking deployments represent the beginning of a new wave of AI implementations and practical uses. While widespread and thorough management of datacentre operations through Artificial Intelligence may still be some time away, current implementations like those that support uptime are already proving beneficial for datacentres and their clients. With more reliable and easily monitored operations, datacentre users can rest easier knowing that their facilities can follow stringent compliance mandates for uptime and efficiency.

In one case study, Artificial Intelligence (AI) based Predictive Analytics platform is implemented for improved uptime of technology services. The system monitors 79 metrics every 30 seconds, which amount to 400GB per day. Log data reaches as many as 100 million [items]. Remote system monitoring occurs every 5 seconds using ping, SSH, and URL. It was also productized in the private cloud datacentre for diverse services, including multimedia surveillance, artificial intelligence and GPS navigation. For GPU servers, hardware-level data are also collected for change management using IPMI and DMidecode. For datacentre facility management, a 3-dimensional visualization (3DV) is used to monitor power consumption and temperature/humidity in the server rooms. 3DV interactively shows the racks and servers by rotating, panning, and zooming in/out. It also uses heatmap to show the temperature distribution.

In another user case, AI-based analytics are being deployed for network function virtualization. Ultimately, the system will automate datacentre tasks based on the results from machine learning analysis. It can execute remote tasks on hundreds or thousands of systems using automation tools. This automation will decrease downtime and human errors, so that only a small number of operators are needed to handle large-scale systems in datacentre. ●

> AI and security

As a data-driven system, AI will bring with it a new range of threat vectors and sources of risk. It will change the cybersecurity landscape. Previously, the major source of threat to a datacentre were human – accounting for 70% of all unplanned interruption according to Gartner. Now the threat will be data-driven, not just through malicious cyber-attacks but also through the corruption and degradation inherent in very large volumes of data. It will be impossible for humans to stay up to date on all the information required to protect the AI-operated datacentre. Machine learning and deep learning applications can help datacentres adapt to changing security requirements faster.

From a legacy situation where security is based on the concept of the perimeter and the key objective of dealing with threats by restricting access and creating impenetrable walls. But with a constant flux of users and data, the approach of restricting access will not be enough to ensure security. The more dynamic approach of AI-based systems can help datacentres be more secure without imposing stringent rules on their users since the requirements for security need to be balanced against those of access. In a data-driven world the focus is on predicting and meeting threats as they occur and using data

analytics to identify what is a threat, the level of threat and how it should be dealt with.

The deployment of AI as a means of delivering security is increasing. This is a response to the huge growth in threats – whether malware strains, phishing, identity falsification, viruses or other forms. The traditional approaches based on firewalls and the belief that perimeter security is the be-all and end-all of the security process are no longer adequate. AI provides the level of analytic detail and speed needed to keep security up to the level of data complexity as many of the threats. Threat-centric security models based on AI go beyond just preventing attacks to detecting potential attacks before they strike. Even where appropriate threat detection technologies are in place, the volume of alerts can be too great to be dealt with without AI-derived intelligence and analysis.

Increasingly, sophisticated algorithm-based techniques are used, not just to identify security threats but to diagnose the wider principle of 'data health'. A data pattern can be considered as the mathematical expression of specific behavior. Such behavior can either correspond to a newly discovered knowledge or something learned in the past. The ability to ►

► recognise behaviors in data has tremendous implications in detecting pre-defined network incidents, such as cell congestions, cell outage, or sleeping cells.

There are two main aspects to threat intelligence: technology or machine intelligence, and human intelligence. While machine intelligence can mine and analyse enormous amounts of data in real-time, many in the industry believe that it is not enough, and that human input is needed to refine the findings.

Threat intelligence seeks to detect anomalies, by establishing a baseline of normal behavior so that abnormalities can be detected through the use of user behavior and user analytics. Threat intelligence also looks to identify "indicators of compromise". These are the tools, techniques and procedures used by attackers from the artifacts left behind in an attack. From this intelligence, countermeasures can be implemented to stave off future attacks.

Network Anomaly Detection is the action of finding behaviors in network traffic which do not conform to expected patterns. These nonconforming behaviors may indicate threat impact, equipment performance degradation, security and intrusion detection or attack blockers when the anomalies are detected in the network in the early stages. Security measures need therefore the ability to proactively detect network anomalies and detect unknown network behaviors without using any signatures, labeled traffic, or training.

Root Cause Isolation (RCI) is the process of identifying the source of anomalies (potentially problems) in a system using only data observation. Many OSS systems and NOCs suffer from a common problem: when the network fails to function correctly, it is often difficult to determine which part is the source of the problem.

In current network systems, by monitoring certain thresholds, warnings and alarms are triggered to indicate an incident in the network. The domain experts then manually investigate the reports starting with the most critical ones. The process of resolving network incidents requires significant comprehensive knowledge about the network architecture, its elements, and their capabilities. It is more effective to use a patented unsupervised approach to measuring distances between any individual entity (such as a cell, or a KPI) and behaviors either previously identified and labeled as abnormal, or automatically learnt as a deviation from the normal expected behavior. This method avoids long post-mortem investigation times.

The fundamental challenge is that the symptoms of a failure often manifest as end-to-end failures in the operation of the system, without causing obvious failures in the system components; noticing that something has gone wrong does not necessarily provide information about where to look to fix it.

In general, cell outage takes place due to multiple reasons, such as hardware or software failures (including misconfigurations

or software bugs) or even environmental changes. Usually, the detection of a malfunctioning cell is performed through the analysis of alarms, KPIs, or in many cases, multiple customer complaints. Generally, cell outage is classified into three types: degraded, crippled, or catatonic.

Due to its complexity and ambiguous behaviour, the most difficult to detect is a degraded cell. A degraded cell can carry network traffic, but not as much as a correctly functioning cell. A crippled cell is characterised by severely degraded performance, but still provides a service to a few users. A cell which experiences complete inoperability is referred to as a catatonic cell. A sleeping cell is a cell degraded type, which is invisible for network operators through traditional alarms. This peculiarity makes a sleeping cell problem a very challenging task.

Root-Cause analysis involves an automatic investigation of problem KPIs and diagnosis regarding failure reasons through the automation of the diagnosis process by creating models per cell, KPI and area to identify the component leading the anomaly.

Not every indicator of compromise turns out to be an attack, and a challenge for threat intelligence is to reduce the number of false positives to a manageable level and to those that really warrant investigation.

In the Huawei O&M evolution model, risks are attributed mainly to manual input as part of the conventional O&M approach. In addition to monitoring system discovery and expert identification, digital O&M automatically identifies risks during O&M activities and triggers risk management. For example, PMI items are directly generated for non-compliance items found during electronic PMI (rules can be defined in the PMI template). This approach in which the O&M security depends on the DCIM system prevails over the conventional approach in which the O&M security depends on the O&M team's expertise and a sense of accountability. AI is particularly gaining traction in terms of device fault prediction and 'proactive' maintenance. Effective sample data and human experience can be used to quickly train a fault prediction model with high accuracy. Device faults can be predicted and routine PMI and maintenance become more targeted. Ever-improving prediction accuracy may eventually eliminate the need of routine manual O&M. ●



> Developing datacentre strategies towards AI

The first consideration in the strategy for AI development is the capability of the datacentre itself. Even the best intentions regarding digitalising the datacentre will come to nothing if the datacentre does not have the capacity to house and process the AI input. DCD research indicates that those datacentres in the Asia Pacific which have started to introduce an AI capability are also dealing with digital transformation data for other companies. Therefore, the two prongs of digital transformation in the datacentre - as a source of business and as a means of operation - tend to occur in combination.

The second consideration will be the infrastructural basis of development, and this may incorporate existing infrastructure or require new deployments. As the datacentre becomes more digitalised in terms of its infrastructure stack - through equipment that is coordinated and managed through operational systems, through the software defined systems adopted to reduce the costs of infrastructure, improve efficiency and to act as the most effective access into cloud systems, so its infrastructure takes on a new shape. Software-defined infrastructure (SDI) becomes more of a standard. As technical computing infrastructure entirely under the control of software with no operator or human intervention, SDI operates independent of any hardware-specific dependencies and is programmatically extensible and therefore amenable to input from AI systems.

The major resource pools in a datacentre - compute, storage and network - can all be delivered under software control, leading to the situation whereby all resources can be available on demand, in a software-defined datacentre (SDDC). In MTDCs and some enterprise datacentres, intelligence tends to be built software-defined infrastructure in combination with DCIM [Datacentre Infrastructure Management]. AI can be added onto DCIM, giving it intelligence to handle levels of information that would overwhelm humans.

This may represent a process of optimising DCIM which has been missing in that software's somewhat chequered history. While DCIM has been implemented widely in datacentres, instrumenting the IT and facilities hardware, its deployment has been slowed by perceptions of expense, complexity of operation, problems of integration with previous monitoring systems and with facility decision making. In situations where it has needed to deal with a considerable volume of data as datacentres have grown in all metrical aspects, it produces, and the complexity of the actions required to run a datacentre. This is very much the domain of artificial intelligence, and its value as an adjunct to DCIM lies in the capacity to identify happening that require attention through data pattern recognition and to use intelligence and learning to formulate the correct response. Intelligent DCIM systems will

proactively help during disaster recovery and help datacentres comply with industry standards and other regulations.

The process of developing an AI strategy may include the following elements:

- Matching the capabilities of the datacentre with the requirements that AI will make of it. Also, match the requirements against your skills profile in terms of data science, engineering, DevOps, machine learning and AI specialists, and against skills sets that will need to be brought in from outside.
- Defining the objectives at both the facility and wider business level for the deployment of AI. While AI as it advances in learning may take a more proactive role in datacentre management, so the datacentre is itself a facility of a company and subject to its corporate objectives. What are the opportunities or challenges that AI is being deployed against?
- Testing both above criteria via a 'proof of concept' project with agreed objectives and risk standards to base deployment decisions. Use testing also to validate the analytics and output models, and how these will create the basis for decisions
- Defining the lifecycles for both AI application software and product development. This will include planning, systems analysis and design, development, testing, implementation and maintenance. Similarly, understand the product development lifecycle (PDLC) for AI. This includes requirements, design, manufacturing for hardware and development for software, testing, distribution, use and maintenance, and disposal.

For enterprise, the AI strategy needs to be put into the context of the overall ROI of running the datacentre as against outsourcing or cloud. Therefore, applications to reduce OPEX and improve efficiencies will be core and the use of AI to monitor and allocate across the different components of hybrid IT as this is the growth path for the enterprise datacentre.

Hyperscale's need is to ensure the facility meets the needs of the business model which delivers on efficiency, scalability and connectivity.

The major applications of transformative technologies in datacentres to date have been in operations and management but the technology now has the potential to take this further. Some of the business decisions that an MTDC might consider AI for (beyond that of better intelligence about datacentre operations) might include the evolution into service hubs which are demand-driven, offering flexibility and scalability. The intelligent MTDC cannot operate on a 'one size fits all' customer proposition, therefore one key output will be to evolve a standard of intelligent management that allows the facility to customise the service offering based on customer needs. This process has ►

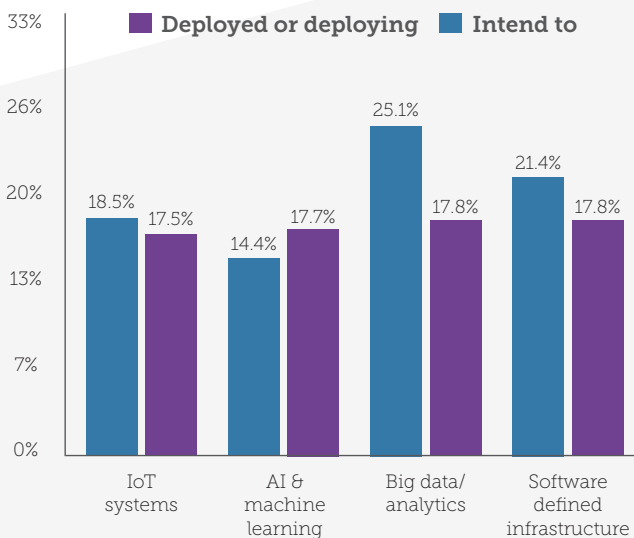
► started already in terms of different service profiles and more open systems but AI can be used to get this to a stage where it is genuinely on-demand and may be based on a level of demand that the client themselves do not have to specify; it is adapted automatically. For this sector, this represents the convergence of IT and OT systems.

As part of this process, AI can be used to make the models used for charging and specifying SLAs more sophisticated and more based on 'real time' and activity closer to a 'pay for what you use' model. One key benefit that this represents for the MTDC is negates a key benefit of using cloud and it may be necessary if cloud in different variants becomes one of the offerings. ●

> The current state of AI/machine learning deployment

Across the Asia Pacific a survey of more than 600 owners and operators of datacentres indicates that AI and machine learning have been deployed by lower proportions than IoT systems, analytics or software defined infrastructure but that intention to deploy among those who have not yet started is at very similar levels.

Figure 11: Deployment of technologies across Asia Pacific data centres (January 2019)



Source: DCD Research 2019

The introduction of digitalised data gathering, analysis and application technologies occurs within a narrow band of companies which account for 15% to 25% of the overall sample. Fairly much the same set of companies are deploying across these four technologies and will continue to do so.

The profile between different datacentre types is fairly even, although big data/analytics deployment is higher among end users (29%) and datacentres run by IT and managed service companies are slightly further advanced in terms of AI and machine learning (25%) and software defined infrastructure (27%).

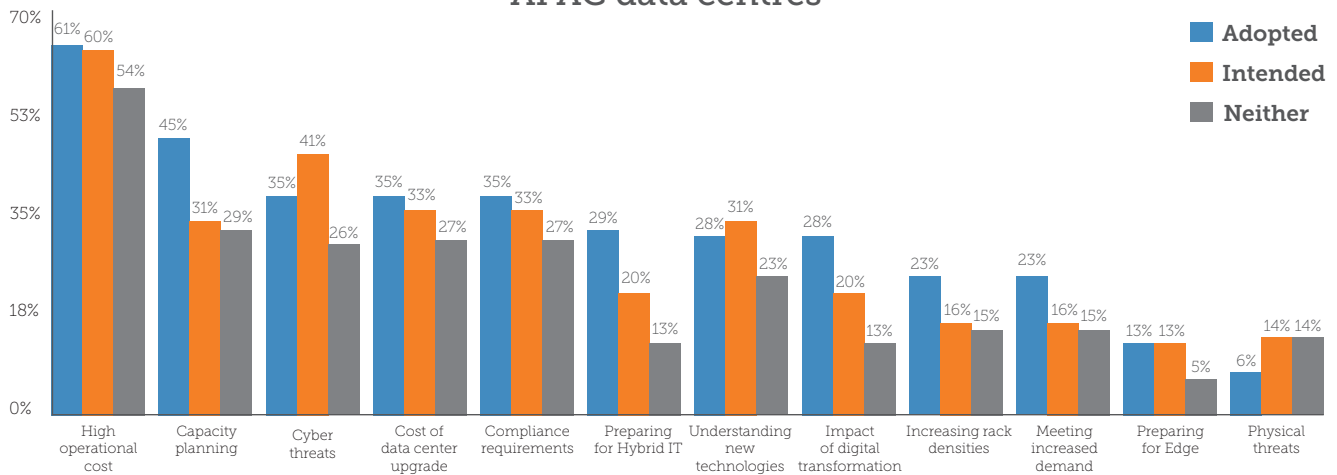
Companies with larger data infrastructure portfolios are more likely to continue to invest in future in three of these four technologies:

- Software defined infrastructure where the average footprint of those companies which will continue to invest in SDI is 950m², significantly larger than the footprint of those not looking to invest in the technology in the future at an average of 347m²
- AI/machine learning where the average footprint of those companies which will continue to invest in AI is 824m², significantly larger than the footprint of those not looking to invest in the technology in the future at an average of 384m²
- Big data/analytics where the average footprint of those companies which will continue to invest in big data and analytics is 836m², significantly larger than the footprint of those not looking to invest in the technology in the future at an average of 377m².

The geographic profile of companies which have, or which will deploy these technologies tends towards the markets of North East Asia and China, and for software defined infrastructure towards Australia and Singapore as well. Future uptake switches towards Singapore and the markets of South East Asia.

Those companies which have started to adopt AI are particularly concerned by capacity planning, preparing for hybrid IT, the impacts of digital transformation and increasing power densities. Those who are intending to deploy AI are close to those who have deployed on most other issues. Those who have neither deployed nor intend to, generally have lower concern on most issues as they are operating less technologically advanced, older facilities. ►

Figure 12: Key drivers for current and intended adoption of AI within APAC data centres

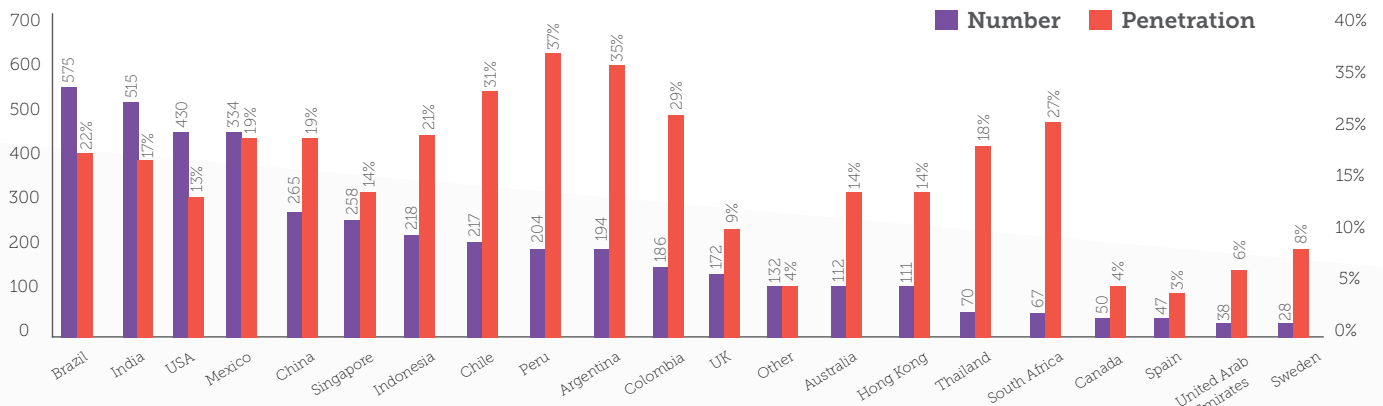


Source: DCD Research 2019

► A more global sample is represented by analysis of DCD registrations for the calendar year 2018 and their interest in AI as a content theme. 11% of registrations in markets covered by DCD across the world in 2018 show interest in AI and machine learning. The key markets in terms of numbers include a high

representation from Brazil, India, the USA, Mexico, China and Singapore. The high numbers of recently established markets are a reflection of DCD's greater coverage of the Asia Pacific and Latin America although the London and Stockholm conferences draw in delegates from elsewhere in Europe.

Figure 13: Interest in AI conference content by market: DCD conference registrations 2018

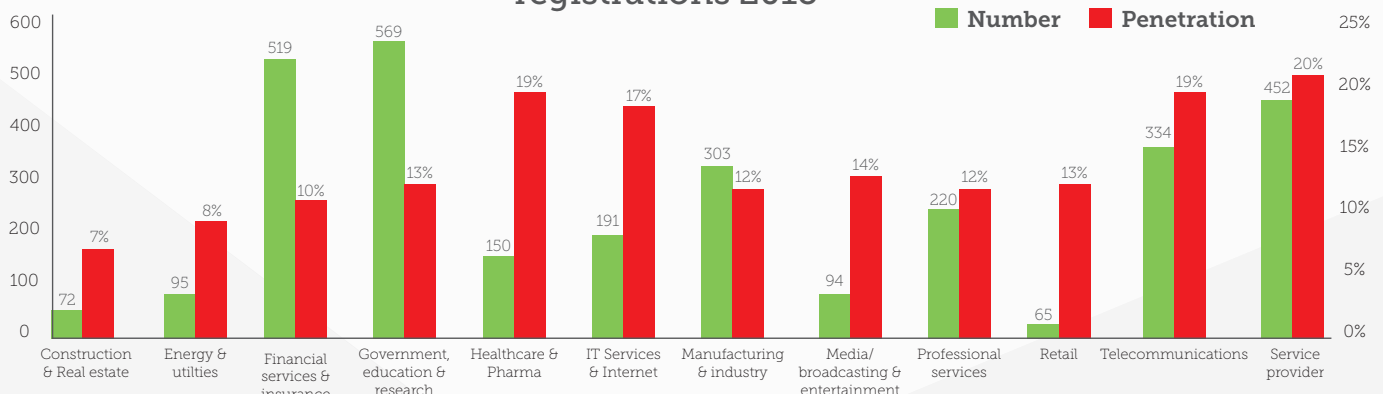


Source: DCD Registration 2018-9

The incidence of interest is highest in Latin American markets, South Africa and Indonesia. The greatest numbers across sectors include Government, the

financial sector, service datacentres and telecommunications datacentres. These latter two sectors, together with healthcare and IT services indicate the greatest level of interest. ●

Figure 14: Interest in AI conference content by sector: DCD conference registrations 2018



Source: DCD Registration 2018-9