acquiredinsights
Bringing Artificial Intelligence to Life.

January 16, 2019

Using AI/ML and Predictive Analytics to Optimize Data Centers

Tools Will Accelerate Higher Performance and Lower Costs

One hundred per cent uptime is critical and remains the goal of responsible data center management for providing exemplary customer experience, retention, and growth; it is also expensive for Data Center owners, managers, and clients.

Data Center managers know that the rate of change of hardware, software and service delivery models are getting faster every day.

Customized AI/ML solutions that can adapt to evolving and scalable technology and service delivery models is essential to driving down the operational costs for Data Centers.
This saves Data Center energy costs and reduces the carbon footprint through quick implementation and reliable performance 24x7.

While traditional data center management—presiding over a company's own servers, storage, networks and other infrastructure—isn't dead, it's becoming far less integral to the day-to-day job of IT, experts say.

By 2021, 94% of workloads and compute instances will be processed by cloud data centers; 6% will be processed by traditional data centers, according to the Cisco Cloud Global Index. And according to some estimates, more than 80% of company cloud strategies are now multi-cloud.

Close observers of the IT infrastructure market say that the top IT infrastructure trends in 2019 will continue to move data center operations outside the four walls of the enterprise. And that will have cascading effects on data center management in 2019.
(https://www.cisco.com/c/en/us/solutions/data-center/gartner-2019-top-infrastructure-operations-technology-trends.html)

These observations are reinforced when examining large global Enterprise Resource Planning (ERP) providers, for example Oracle and SAP, where they are rapidly and decisively moving their traditional on-site client installs to cloud-based services.

Key Performance Indicators (**KPI**) associated with IT and Data Center Management are many and include but are not limited to:

1. Account create success
2. Account termination success
3. Active directory performance index
4. Alert-to-ticket ratio
5. Average data center availability
6. Call center PBX availability
7. Customer connection effectiveness
8. Data center capacity consumed
9. Email client availability
10. Exchange server availability
11. Incidents from change
12. Internet proxy performance
13. Network availability - High availability sites
14. Network availability - Standard sites
15. Network manageability index
16. No problem found/duplicate tickets
17. Percentage of branch office backup success
18. Percentage of circuits exceeding target utilization
19. Percentage of IT managed servers patched at deadline
20. Percentage of production servers meeting software configuration standards
21. Percentage of security update restarts within maintenance window
22. Percentage successful remote access server (RAS) connections
23. Phone answer service level
24. Priority 1 and priority 2 network incidents meeting SLA
25. Product adoption status and compliance
26. Restore success rate
27. Server growth rate
28. Server manageability index
29. Service desk client satisfaction - Percentage dissatisfied
30. Service desk tier 1 resolution rate
31. Service desk time to escalate
32. Service desk time to resolve
33. Storage utility service availability
34. Storage utility utilization
35. Virtual machine provisioning interval
36. Virtual server utility availability
37. Web server availability

These KPIs are complicated by the fact that the global reliance on data, the complexity of technology implementations in the data center, and increasingly complicated client requirements are also advancing exponentially.

To date there has been some adoption of AI/ML technology in data centers, primarily for the purpose of power optimization, but now Acquired Insights Inc. is also introducing AI/ML for the purpose of reducing downtime, optimizing unit and system performance, predictive maintenance, and detecting issues before they cause major operational disruption.

Traditional operational support has consisted of a team of engineers, developers, and technicians certified on specific hardware and software; that skills base is also typically augmented by various networked sensors and support staff to monitor the power, back-up, and cooling infrastructure on an hourly basis.

While electronic sensors are required, the role that humans play by providing monitoring and oversight within a data center is indispensable and, for the near and foreseeable future, will remain so. The reason for humans being so indispensable is the fact that engaged employees offer an element of curiosity and contribution to the well-being of the Data Center and their employer. When a Data Center is in the middle of a crisis, it may be the product of many issues arising at the same time.

Acquired Insights has an answer for using AI/ML to identify, manage, and mitigate many of the traditional technical and operational issues that affect Data Center performance, and it's accomplished through a mobile app, the human user interface to AI/ML, called ***KaZam!*** Over a relative short period of time, humans are used to capture role-based observations and commit them to corporate memory through ***KaZam!***

Once committed to ***KaZam!***, Machine Learning is then used to learn what constitutes normal behavior and what is a failure. Artificial Intelligence is then used to glean insights, predict failures, and then push those notifications to the right roles in the data center – often before employees know they have a need for that knowledge – transitioning their historical role from reactive to proactive.

With the right tools such as the ***KaZam!*** App, a human can log multiple issues, observations, a prescribed remedy, and acquired learnings with an easy to use user interface to the monitoring system.

With the KaZam! App, a human can log multiple issues, observations, a prescribed remedy, and acquired learnings with an easy to use user interface to the monitoring system.

IBM's published research[1] identifies that 90% of data that currently exists in the world has been created in the last two years, and 90% of ***that*** data is unstructured or semi-structured data. Since businesses generally use structured data only, it means they are typically only using ~20% of available data for decision support. By contrast and comparison, if an airplane used only 20% of its engine power, it would never get off the ground.

[1] https://www.mediapost.com/publications/article/291358/90-of-todays-data-created-in-two-years.html

So a tool that can capture and tag unstructured and semi-structured data (i.e., voice dictation/memo, email, text message, photos, videos, hand written notes) from Engineers and Technicians in real time, for example, and upload it into an Enterprise Memory Management System (EMMS) integrated with AI/ML, predictive analytics and structured data sets that are already in place to monitor Data Center hardware and software, provide compelling value to Management, employees and customers. ***KaZam!*** can not only capture unstructured and semi-structured data, it can also retrieve information and recommend the next best course of action.

The ability for an engineer or technician to walk the floor, hear an unusual sound, use their hearing to identify which piece of equipment is not performing optimally, decide on a course of action that preserves the integrity of the operation, while isolating the failing equipment so can be adequately serviced, is a typical, everyday experience. But how do we capture the technician's observations and solution, codify it, submit it to a knowledge repository, combine it with historical operational data, and make everyone in that data center (or other data centers in different locations) as competent at solving that same problem, in real time? Acquired Insights' ***KaZam!*** App to the rescue!

It is disconcerting to know that some of the most advanced technology, the technology that resides in Data Centers, is monitored, remedied, and optimized primarily by human beings who physically walk the floor. Without the ability to identify problems, log solutions, and codify lessons learned in a role-based approach, and without the ability to push that knowledge to people in the same or a similar role - and often in advance of them knowing they actually need that knowledge - the costs for managing a Data Center will always be inflated. Add to that, the older-generation facilities that are not as well-instrumented as modern data centers, which can present a potential problem in creating the data repository. Additionally, facilities contain such a diverse set of equipment, it can be difficult to create a dataset that's clean enough for training the AI model. As well, not all companies whose mission-critical business infrastructure occupies these facilities will be comfortable piping their operational data into a centralized repository, for reasons such as security, compliance, and competition. With so much unstructured and semi-structured data, Acquired Insights' AI models excel in the data center environment.

Data center operators that "don't employ AI/ML to power behavioral-based security through automation, especially in the response and remediation of attacks, will leave themselves vulnerable by not keeping pace with the evolution of threat technology"

- Migo Kedem, Director of Product Management at SentinelOne, a security firm based in Mountain View, California.

What is even more shocking is the risk to Data Center operations, customer satisfaction and retention, business continuity, performance and profitability, especially given how important the human contribution is to crisis management and resolution effectiveness and efficiency.

The Human Factor

With 80+ years of data collection experience across 160 countries, and over 35 million respondents in Gallup's employee engagement database, several very scary statistic have emerged:

- A staggering **87%** of employees worldwide are not engaged at work.
- Managers account for at least **70%** of the variance in employee engagement scores.
- **51%** of Employees are actively looking for a new job or watching for new job openings.
- Companies with highly engaged workforces outperform their peers by **147%** in earnings per share.

Could poor employee engagement and/or turnover significantly compromise Enterprise Memory Management to the point that generally accepted service level agreements and KPIs are also compromised?

The American Society for Training and Development's (ASTD) research examining Turnover Related Costs (TRC) identified:

- **75%** of the demand for new employees is to simply replace workers who have left the company.
- The Average Cost of a Bad Hire (including hiring costs + total compensation + cost of maintaining the employee + disruption costs + severance + mistakes, failures, and missed business opportunities) is equal to **$840,000.** (Based on a second level Manager making $62,000 annually and terminated after 2.5 years).
- The Return On Investment for Bad Hires that leave voluntarily or involuntarily after 12 months can be defined as The Value of Contribution MINUS Total Costs (Investment[2]) X 100 = **-298%.**

  [2] *Investment is defined as Cost of Hiring + Compensation + Cost of Maintaining an Employee.*
- This gets compounded even further because good employees may have to work harder to make up for what a bad hire is not doing, putting Data Centers at risk of losing good employees.

With the research and statistics quoted above, despite rigorous maintenance, process, and escalation procedures, the best-run data centers are still ***at risk of downtime***, and that risk is often directly related to the level of engagement and/or turnover of management and employees - where experiential knowledge acquired by employees, directly relevant to data center operations, is discontinuous and at risk of being lost forever.

Asking the Right Questions to Assist in Designing a Solution

When Data Center outages occur they trigger a crisis situation, launching entire teams into a troubleshooting processes that require an auditable examination of dynamic parts to diagnose the problem to quickly and correctly bring systems back online. At Acquired Insights we think long and hard about solving these problems, but before we do, we asked ourselves a series of insightful questions, a small subset of which include:

- What if there was a way to use AI/ML and predictive analytics to accurately forecast, with a high degree of probability, when, where, and why outages would likely occur?
- What if it could be done in a way that it isolated it to the individual machine level?
- What if the solution could be implemented in a way that included multiple data sources, and multiple types of data sources, both now and in the future?
- What if the solution could scale as needed?
- What if employees were able to capture and contribute their observations to an EMMS that enabled others in the same role in the same building, or around the world, to become just as competent at forecasting and resolving the same issue in real, or near real time?
- What if it was done in a way that empowered employees to be proactive in their role of walking the floor?
- What if original knowledge contributed by an employee to this EMMS, at any time in the life of the enterprise, through Blockchain, could be attributed to an individual employee, and based on the commercial value of their contribution, that employee would receive some form of additional compensation, e.g., commission, royalties, Class B shares, etc., to create and sustain high employee engagement?

According to IDC, the average hourly cost of a data center infrastructure failure is US $100,000 per hour - and in the case of a critical application failure, that number rises to $500,000 to $1 million per hour.

For providers, a significant incident can undermine the perception of the provider as a key business partner. It can result in being perceived as a barrier to productivity and profitability that undermines a client company's ability to compete effectively. With a growing percentage of data and operations moving to the cloud, the threat to the client and the provider brand is accentuated. It is also dangerous for the provider's own corporate health because significant resources during these "all hands on deck" critical times are allocated to reacting instead of building proactive value initiatives into the business; business continuity can be put at risk quickly and a full recovery can be a multiple of the time the actual crisis consumed.

The Case for Making AI/ML a "Must" in Modern Data Centers

"Machine learning can help scale human expertise in analysis of network state and behavior", Gaurav Banga, founder and CEO of the security firm Balbix, said. It can also help evaluate whether the security controls in place are adequate and properly calibrated to defend against current threats.

In 2014, Google turned to AI to meet the challenge by building a box that tracked 19 variables including: IT load, weather, water pumps, and heating exchangers and charged the machine with calculating maximum efficiency for its server farms.

Those algorithms – since trained with billions of data points – now support Google's operations teams internationally in setting the electrical and mechanical plants for optimal performance; its Power Usage Effectiveness is consistently predicted with 99.6 per cent accuracy.

Pursuing AI/ML to Optimize Data Center Uptime

Today, the broader AI evolution continues more widely. In Data Centers, threat detection, ML- aided forensics, and Response Automation are leading areas of interest for applied AI/ML, but so much more could be done, done faster, and time *IS* money.

Beyond instances such as Google's efforts to improve energy efficiency, data centers aren't traditionally synonymous with innovation. However, the enormous potential that underpins artificial intelligence is giving birth to a new generation of AI/ML solution pioneers - like Acquired Insights - who are devising machine learning systems to tackle what is arguably the biggest challenge facing today's modern data centers: downtime, and how to reduce its risk.