RF Code™

White Paper

IT Starts with Information:

How Monitoring, Capacity Planning and Predictive Analysis Improve Data Center Efficiency and Facilitate Business Growth

Executive Overview

Data centers are the engines of commerce. They enable every aspect of modern civilization, from social connections to global markets. But they also use a lot of energy – enough to power entire cities. And much of that energy is wasted.

Data centers are also expensive, both to build and to run. A company cannot afford to let a multi-million dollar asset operate at anything less than optimal efficiency. This paper explains how real-time monitoring, capacity planning and predictive analysis technologies help data center operators improve agility and efficiency in their data centers and ensure higher performance at a lower cost.

Introduction

We live in an always-on, always-connected digital world thanks to data centers. While the media tends to focus on large cloud facilities run by household names like Amazon, Facebook and Google, they represent only a small percentage of the data centers that form the backbone of today's global economy. Yet for something so ubiquitous, data centers are strangely invisible and poorly understood.

Simply put, a data center is a place that contains IT equipment and the infrastructure that supports it – mostly power and cooling equipment. Although some data centers fill large warehouses, many occupy just a few hundred square feet in a building that is otherwise used for office space.

Unfortunately, data centers use a lot of energy – in some cases, up to 200 times the electricity of office space.[1] Most enterprise data centers spend millions of dollars on electricity each year. That's real money in any economy, much less in today's highly competitive global landscape. Add in rising energy costs throughout much of the industrialized world and the increased potential for climate change-related regulation,[2] and one would think most data center operators would consider improved energy efficiency their highest priority.

But do they?

No. In a recent survey conducted by the Data Center Users' Group, a collective of over 2,000 data center, IT and facility managers, energy efficiency was ranked fourth in priority by survey respondents.[3] The top concerns? Adequate monitoring and data center management capabilities, availability, and technology changes and change management.

The common denominator of these concerns is ensuring availability. In an always-on world, a data center cannot go offline. A seemingly minor change or miscalculation and an overloaded circuit trips. A computer room air conditioner (CRAC) unit malfunctions and unmonitored servers overheat. An uninterruptible power supply fails and the facility goes dark.

These are concerns that keep people up at night, because data center outages are expensive. In a 2013 survey conducted by the Ponemon Institute, respondents reported an average cost of $690,204 per incident.[4] And that number includes only readily quantifiable impacts – business disruption, lost revenue, decreased productivity, equipment repair and the like. The cost to a business's reputation is harder to measure, but it lasts longer and affects the bottom line far more than the expense of the actual event.

So how to ensure availability? Most operators use a multi-pronged approach: incorporating redundant systems in their design, applying best practices for operational and maintenance procedures, and using innovative and integrative technologies such as data center infrastructure management (DCIM) systems to improve reliability.

DCIM technologies monitor power and environmental conditions within the data center, build and maintain asset databases, facilitate capacity planning and assist with change management. In

1 *Best Practices for Data Centers: Lessons Learned from Benchmarking 22 Data Centers. S. Greenberg et al. 2006. ACEEE Summer Study on Energy Efficiency in Buildings. http://datacenters.lbl.gov/sites/all/files/aceee-datacenters.pdf*

2 *Climate Change 2014 Synthesis Report – Approved Summary for Policymakers. Intergovernmental Panel on Climate Change. 1 November 2014. http://www.ipcc.ch/pdf/assessment-report/ar5/syr/SYR_AR5_SPMcorr2.pdf*

3 *Fall 2014 Data Center Users' Group™ Survey Results. October 2014. http://www.emersonnetworkpower.com/documentation/en-US/Brands/Liebert/Documents/White%20Papers/DCUG%20Survey%20Results%20Fall%202014%20-%20Conference%20Version.pdf*

4 *2013 Cost of Data Center Outages. http://www.ponemon.org/local/upload/file/2013%20Cost%20of%20Data%20Center%20Outages%20FINAL%2012.pdf*

Uptime Institute's 2013 Data Center Industry Survey of global data center operators, respondents identified the following as the top three drivers for DCIM adoption: better management of data center capacity, identifying problems that could threaten availability, and better visibility and manageability of assets and status.[5]

This paper will focus on how real-time monitoring, capacity planning and predictive analysis technologies help data center operators improve agility and efficiency in their facilities and ensure higher performance at a lower cost.

Real-time Monitoring Improves Availability and Lowers OpEx

Cooling usually consumes the most power in a data center (that is, excluding the actual computing work). Electronic equipment puts out a lot of heat and overheated devices are more likely to fail. That's why, traditionally, data centers have been kept at temperatures more like a refrigerator than a place of business.

There are many techniques and technologies data center operators can employ to save energy in their facilities. Arguably the strategies that offer the greatest potential savings in cooling costs (e.g., free cooling, chiller-free data centers) stem from recent guidance from the American Society of Heating, Refrigeration and Air Conditioning Engineers (ASHRAE) that expands the recommended and allowable temperature and humidity ranges for data center equipment operation.[6] Today, data centers that used to operate at 55-65˚F can run at 80 or even 90˚F, and with less stringent humidity limits.

The impact of these changes on energy use can be significant: a data center can save 2-5% in energy costs for every 1˚F increase in server inlet air temperature.[7]

5 2013 Uptime Institute Data Center Industry Survey. http://uptimeinstitute.com/2013-survey-results
6 Thermal Guidelines for Data Processing Environments (2008, 2011, 2012). http://ashrae.org/resources--publications/bookstore/datacom-series#thermalguidelines
7 U.S. EPA Energy Star - Server Inlet Temperature and Humidity Adjustments. http://www.energystar.gov/index.cfm?c=power_mgt.datacenter_efficiency_inlet_temp%20

Consider the example of telecommunications giant CenturyLink, which currently operates 55 data centers worldwide. In 2011 CenturyLink had an annual electricity bill of over $80 million. Starting in a pilot facility, CenturyLink implemented a monitoring program that enabled its engineering team to safely raise supply air temperatures without compromising availability. Based on the results of that pilot study, CenturyLink identified $2.9 million in potential annual savings across its portfolio.[8]

Given ASHRAE's new guidance and the substantial savings these changes offer, one might expect that every data center operator would make these seemingly straightforward adjustments. Yet many don't; over three-quarters of the respondents to the 2013 Uptime Institute survey reported that their average server supply air temperature was 65-75˚F – far cooler than ASHRAE recommends.[9]

8 RF Code Delivers Millions of Dollars in Annual Power & Cooling Savings for CenturyLink. RF Code Case Study, 2014. http://resources.rfcode.com/rf-code-delivers-millions-of-dollars-in-annual-power-cooling-savings-for-centurylink
9 2013 Uptime Institute Data Center Industry Survey. http://uptimeinstitute.com/2013-survey-results

Power distribution and backup equipment also contributes to energy waste in the data center, due to conversion losses, poorly designed power chains and inefficient power supplies and cables. As with cooling, there are many strategies data center owners can employ to improve power efficiency, but the most obvious are on the compute side. Because most data centers provision for peak load – loads that may occur only a few days per year – low server utilization is the status quo in the industry and presents substantial opportunity for increased efficiency. Experts estimate that server utilization in most data centers is only 12-18%, and some feel up to 20% of the servers deployed in data centers are plugged in 24/7/365 but performing little to no work.[10] These "comatose" servers still draw almost the same amount of power when idle as they do when processing at full capacity. Additionally, every watt of electricity wasted at the device level has a cascade effect, as still more energy is needed to power the physical infrastructure that supports the device.

Increasing the density of the IT load per rack through consolidation and virtualization can offer substantial savings in not only equipment but also electricity and space – an important consideration if the data center is located where energy supply is constrained or electricity and real estate prices are high, as in most urban areas.

Seventy percent of the respondents to the Data Center Users' Group's recent survey reported that their average density is currently 2-8 kW/rack. But almost the same percentage predict that in two years their average density will increase to 4-16 kW/rack.[11] High density means concentrated thermal output and modified power requirements. The only way to maintain continuous availability in high density deployments is real-time monitoring and granular control of the physical infrastructure.

One of the most promising innovations in energy efficiency is power proportional computing, or matching power supply to compute demand. Yet few data center operators use available dynamic provisioning technologies or the power capping features already installed on their servers.

Why are data center operators leaving these potentially game-changing savings on the table?

Risk. Without real-time monitoring and management, raising inlet air temperature increases the risk of equipment failure. Without a detailed understanding of the relationship between compute demand and power dynamics in the data center, power capping increases the risk that the processing capacity won't be available when required.

Availability trumps savings. Availability trumps all.

Businesses don't have to choose between availability and savings – they can have both. Real-time monitoring gives businesses the information they need to manage risk, improve efficiency and decrease costs.

In an intelligent data center thousands of sensors throughout the facility collect information on temperature, humidity, air pressure, power use, fan speeds, CPU utilization, and much more – all in real time. This information is aggregated, normalized and reported in ways that allow the operator to understand and adjust controls in response to current conditions.

Consider this scenario: a technician makes an error in replacing a section of the raised floor – installs a solid tile where a perforated one was needed, or reverses a directional panel so the cool air is blowing across the aisle rather than toward the server inlets. This seemingly minor mistake can have major repercussions in terms of air flow, temperature and air pressure. The only way to catch a small issue before it becomes a large problem is through real-time monitoring.

10 NRDC Issue Paper, Data Center Efficiency Assessment: Scaling up Energy Efficiency across the Data Center Industry: Evaluating Key Drivers and Barriers. August 2014. http://www.nrdc.org/energy/files/data-center-efficiency-assessment-IP.pdf
11 Fall 2014 Data Center Users' Group™ Survey Results. October 2014. http://www.emersonnetworkpower.com/documentation/en-US/Brands/Liebert/Documents/White%20Papers/DCUG%20Survey%20Results%20Fall%202014%20-%20Conference%20Version.pdf

Photo: Connie Zhou for Google

Monitoring has benefits beyond disaster avoidance. Cloud, co-location and hosting providers can use the data collected to document their compliance with service level agreements (SLAs).[12] Monitoring data can be integrated into the facility's building management system (BMS), allowing operators to further automate and optimize control of the physical environment. Visibility at a macro and micro level improves client confidence, streamlines decision making and increases data center availability, productivity and energy efficiency.

Capacity Planning Improves ROI and Increases Productivity

Asset management is a key feature of data center capacity planning. For many organizations, asset management means recording the name and location of every piece of IT equipment in the data center in a spreadsheet or diagram. This is a labor-intensive, expensive and error-prone way to track valuable assets.

A data center is a dynamic environment. Equipment is moved every day; devices are taken offline for maintenance, and new equipment is deployed. Data centers that track assets manually are attempting to solve a modern problem using methodology that dates back to ancient Mesopotamia.[13]

In a data center that employs a static, manually maintained asset management system, staff must physically walk around the data center to conduct inventory audits. If a device is missing from its last recorded location or if information about the device is incomplete or conflicts with existing records, staff must investigate. Discrepancies must be reconciled and lost equipment replaced. In this data center, productivity is compromised, morale is low and costs are high. Trying to make capacity decisions using incomplete, inaccurate data is like working in the dark – figuratively and, eventually, literally, because this is an outage waiting to happen.

Now consider a data center with a real-time asset management system. In this data center the user knows the exact location of every piece of equipment in the facility. He can drill down to specifications, maintenance and warranty histories for every device. He may even be able to see the device on a floor

12 *RF Code Delivers Millions of Dollars in Annual Power & Cooling Savings for CenturyLink. RF Code Case Study, 2014. http://resources.rfcode.com/rf-code-delivers-millions-of-dollars-in-annual-power-cooling-savings-for-centurylink*

13 *Inventory Management History, Part One, by A. Dolinsky. Almyta Systems. http://www.almyta.com/Inventory_Management_History_1.asp*

plan and in context, with power paths, network connections and dependencies clearly mapped. This user fully understands the current position and status of every piece of equipment in the data center. He can correlate real-time monitoring data with asset management information to detect stranded capacity (for example, power is available in a given area but cooling is at its limit). He can model "what if?" scenarios for new configurations and predict what would occur should a certain piece of equipment fail. This is a data center that is consistently available to perform today and is ready to meet the demands of tomorrow.

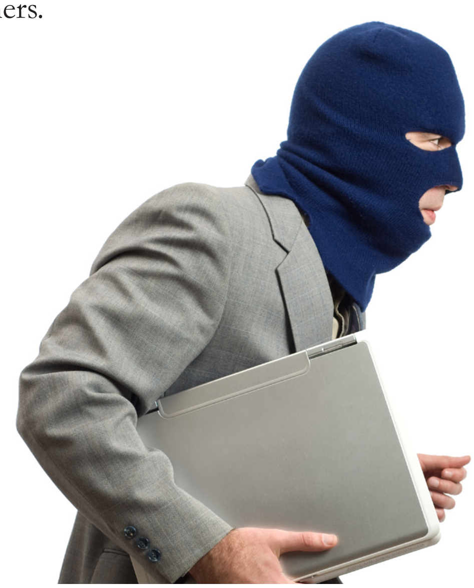
On a tactical level, real-time asset management means less time wasted in inventory reconciliation, fewer penalties associated with late lease returns, and smaller equipment replacement budgets. Warranty and depreciation information is readily available, audits are streamlined and change management is simplified. On a more strategic level, asset management systems facilitate capacity planning.

But if most data centers currently have a large amount of underutilized compute capacity, is capacity planning *really* an important concern for many data centers?

Absolutely. In an October 2014 survey of members of the Data Center Users' Group, 63% of the respondents indicated that they would run out of capacity in the next 2-5 years.[14]

Building new capacity is expensive: a data center can cost $5-10 million per MW.[15] Co-location is an option, but prices currently range from $1,000-$2,600 per rack each month, and that doesn't include electricity, bandwidth, staff, and migration costs.[16]

A company that buys new compute capacity simply because it is unable to identify that sufficient capacity already exists is making two expensive mistakes: it is wasting money to buy capacity it does not need, and it is taking funds from other initiatives that could drive business growth. There can be no doubt: the benefits of comprehensive, real-time asset lifecycle management extend beyond the walls of the data center to influence the financial stability of the company itself – and potentially, that of its customers.



While companies are acutely aware of the dangers of data loss due to hackers, a significant number of IT security breaches are the result of inadequate asset management policies.

When we think about IT security we usually think about theft – hackers breaking into a site and stealing information. The reality is that some of the most damaging IT security breaches arise from a far more mundane – and completely avoidable – cause: companies simply losing the devices on which sensitive information is stored. The Health Insurance Portability and Accountability Act (HIPAA) Security Rules require that covered entities institute meaningful policies to safeguard any hardware or portable devices on which the personal information of clients is stored.[17] Similar rules govern the Payment Card Industry (PCI).

The fines associated with asset management lapses are often in the millions of dollars and, when associated class action lawsuits are taken into consideration, can reach the billions.[18] With this kind of money

14 Fall 2014 Data Center Users' Group™ Survey Results. October 2014. http://www.emersonnetworkpower.com/documentation/en-US/Brands/Liebert/Documents/White%20Papers/DCUG%20Survey%20Results%20Fall%202014%20-%20Conference%20Version.pdf

15 2013 Uptime Institute Data Center Industry Survey. http://uptimeinstitute.com/2013-survey-results

16 Colocation Pricing Trends. Data Center Talk, January 5, 2012. http://www.datacentertalk.com/2012/01/colocation-pricing-trends/

17 U.S. Department of Health and Human Services Health Information Privacy: The Security Rule. http://www.hhs.gov/ocr/privacy/hipaa/administrative/securityrule/index.html

18 Big Data, Bigger Security Risks: How Data Centers Can Track, Manage, and Secure Data with Dedicated Asset Tracking Networks. RF Code White Paper, 2013. http://resources.rfcode.com/whitepaper-big-data,-bigger-security-risks-how-data-centers-can-track-

on the line, it's clear that a robust, real-time asset management system is not just a nicety – it's a necessity.

Predictive Analysis Facilitates Business Growth

Real-time monitoring, asset management and capacity planning offer substantial benefits in themselves, but the true potential of these technologies lies in the insights that can be gleaned through detailed analysis of the data collected. Combining environmental measurements with IT information like CPU utilization, server power use and fan speeds gives a clearer recognition of the relationships between compute demand and the physical infrastructure and enables more responsive, integrated control of the data center as a whole. Operators are able to use sophisticated efficiency technologies like dynamic power provisioning and load shifting with confidence. Using machine learning technologies, the complex relationships between millions of pieces of data collected over time can be analyzed to identify connections humans cannot grasp. Rules can be derived and policies integrated into management software to create more autonomous, optimized operation,[19] decreasing the risk of human error and improving productivity.

In other words, predictive analytics moves data center operations from a reactive to a proactive mode.

Perhaps your company has more than one data center. If you've implemented real-time capacity planning and monitoring capabilities across the enterprise, predictive analysis of the data collected could facilitate shifting workloads between facilities on an intermittent basis. The ability to dynamically shift loads between facilities with confidence not only defers capital expenditures in upgrades or new facilities but also can open the door to other efficiency initiatives (e.g., onsite renewable power generation, use of "follow the moon" strategies, or participation in your utility's demand-response program). It is also a good defense against transient threats to availability (e.g., an impending storm that may affect power supplies – an increasing risk in this era of climate change).

The longitudinal data gathered through monitoring can be analyzed to identify trends as well as predict issues: if CRAC units have a characteristic pattern of declining performance before they fail, the data center operator knows exactly when to replace a CRAC unit that's following the same pattern. Applications may have a particular cooling and power use signature under one set of circumstances and another when conditions change. Predictive analytics allows operators to understand the relationships among environmental conditions, work patterns and compute activity over time, allowing them to derive rules and develop models that assist in decision-making.

Another benefit of detailed, longitudinal data collection and analysis is that it enables more accurate measurement of the true cost of providing a service. If a business can identify precisely what application is running on a given server at a given time, track the power and cooling used to support that compute load, and incorporate data on the other resources required (e.g., staff, bandwidth, additional equipment), it can calculate precisely what to charge a specific business unit or client for that service.

This ability is nothing short of transformative in terms of business planning. Workloads can be prioritized and costs allocated based on demand and availability. Organizations can develop industry-specific productivity metrics that give users visibility into the efficiency of their infrastructure and allow them to fully understand and exploit the capacity of the data center to further business growth.

One such example is eBay's Digital Service Efficiency (DSE) dashboard. eBay's DSE quantifies data center effectiveness in the context of business key performance indicators to help the company track its progress in four target areas: performance, cost, environmental impact, and revenue.[20] Performance metrics such as these highlight the importance of data center efficiency in meeting business goals.

manage-and-secure-data-with-dedicated-asset-tracking-networks

19 *Better Data Centers Through Machine Learning. Google, May 28, 2014. http://googleblog.blogspot.com/2014/05/better-data-centers-through-machine.html*

20 *Designing an Efficiency Metric. http://tech.ebay.com/blog-post/designing-efficiency-metric*

Summary

The data center is the keystone of modern business. When it functions as intended, it sustains and strengthens the business. When it fails, the entire business is compromised.

In our always-on world, customers expect a data center to be continuously available and companies invest millions of dollars each year to meet those expectations. Unfortunately, inefficiencies mean that data center availability often comes at an unnecessarily high price, both in terms of capital and operational costs and environmental impact. To succeed in today's competitive business climate, data center operators must implement techniques and technologies that enable them to maintain continuous availability while optimizing efficiency. The secret to achieving both goals is real-time monitoring and management of the data center environment.

Monitoring technologies give the data center operator visibility into the environmental conditions throughout the facility and allow him to identify and address an issue before it becomes a problem. Real-time asset management systems ensure the operator has comprehensive information on every device in the data center, simplifying inventory management and facilitating capacity planning. Predictive analysis of the data collected enables more integrated, autonomous operation of the data center and informs decision making throughout the organization. A company armed with this combination of information and technology is able to fully understand and exploit the capacity of its data center to further business growth.

About RF Code

RF Code is the world's fastest growing, leading provider of distributed IT environmental monitoring and asset management solutions. Its patented tracking and sensor technologies are deployed by many of the Fortune 250 and help manage the global data centers of some of the largest IT service providers. RF Code is an essential component of the asset management, risk and compliance assurance and automated control systems in healthcare, IT services, industrial supply chains and natural resources/oil & gas industries. RF Code is a privately held company with investors including yet2Ventures and Intel Capital. The company is headquartered in Austin, TX, with offices and partners in the UK, EMEA, Australia, Asia and South America.

Large enterprises, such as HP, CME Group, GE, Dell and Bank of America, have already begun experiencing substantial ROI through the deployment of RF Code's automated, accurate real time monitoring systems and are continuing to see positive results through expanded deployments. In addition, global market leading DCIM suppliers, including IBM, CA Technologies, and iTRACS CommScope have eagerly integrated RF Code's real time environmental and asset monitoring technology into the foundation of their systems. The unmatched accuracy of the data RF Code delivers enables these DCIM suppliers to gain a competitive edge in the market surpassing the value of any other systems currently available.

RF Code™
9229 Waterford Centre Blvd. • Suite 500
Austin, TX 78758
Tel: 512.439.2200 • Fax: 512.439.2199
sales@rfcode.com • http://www.rfcode.com