A Framework to Standardize Site-Wide Analytics Results – perspective from collaboration between a large industrial user, an energy data mining company and a utility program provider

Ramesh Rao, Graphet Data Mining
Brody Wilson, IBM

ABSTRACT

Large industrial sites are increasingly leveraging data analytics through alerts and notifications from many different systems to help manage complex operations. This presents challenges for energy management in that a clear overarching picture of site-wide analytics results is difficult to achieve. Due to the various analytics output formats and platforms the most obvious opportunities to impact energy efficiency can be lost in the noise.

Without a reasonable method to compare all analytics outputs, prioritizing and acting on them is difficult. To empower energy efficiency and innovation in energy management, the outputs of analytics tools must be standardized across platforms. This paper illustrates the need for a common framework for analytics reporting and establishes the fundamental requirements for it.

To meet the need for data portability across systems, the commercial HVAC/Controls industry has an initiative called project-haystack while the Capital Facilities Industry has proposed an XML schema solution named cfiXML for electronic data exchange. These efforts standardize the data on which analytics are performed, but do not standardize the format of analytics results. A framework for analytics results will help leverage new technologies like Machine Learning and Artificial Intelligence while accelerating collaboration across analytics providers.

The idea for an analytics results framework emerged from a Graphet Data Mining project at IBM's Data Center an industrial site in Boulder, CO. Graphet, Xcel Energy, and IBM teamed up recently to identify energy efficiency opportunities at this site.

IBM has a mature global energy management program certified under ISO 50001. Its energy management program goes beyond traditional best practices to embrace new technology; however data mining analytics results from disparate systems still poses some challenges. We believe the initiatives underway in the HVAC and Capital facilities Industries are valuable but need to be taken further to address these challenges.

This article will examine the outputs from diverse analytics platforms at sites like IBM Boulder to make the case for a common language and framework to guide intelligent and effective actions for energy efficiency and critical operational needs.

Introduction

With the advent of the Internet of Things (IoT), where every device can communicate and vie for attention, Big Data has become a reality. Systems once considered discrete and independent are now expected to share data to inform each other's operation The term Big Data articulates the concept that data is coming from a large number of sources and is beyond the capability of a typical database to store, capture and analyze. McKinsey Global Institute estimated that nearly all sectors of the US economy had an average of 200 terabytes of stored data per company,[1] and that many sectors had more than 1 petabyte in mean stored data per company (McKinsey 2009).

Economist Herbert Simon (1971, 40-41) noted, “A wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it.”

In manufacturing or commercial facilities, site operations and maintenance personnel receive large amounts of data from four basic sources:

1. **Automation Systems (SCADA) –** Systems that control and monitor coordinated operation of multiple pieces of equipment.
2. **Equipment Optimization Systems** - Third party systems that provide an optimization “overlay” for facility operations.
3. **Data Collection, Aggregation, and Fault Detection and Diagnostics (FDD) Systems –** Systems that provide a rule based analytical engines for fault detection and diagnostics.
4. **Equipment controllers –** These are the input/output “field level” devices, often OEM provided, that directly control the equipment.

Each of these sources (systems and platforms) often has its own graphical interface and alerting capabilities ranging from audible horns with on-screen only information, to mobile/pager notifications, to sophisticated alert visualizations across systems. However, not all alerts are created equal. The validation, root cause analysis, and elimination of false positives require subject matter expertise and the ability to prioritize the urgency and importance of actions.

The idea that site personnel can efficiently, reliably and consistently operate and maintain facilities with more dashboards, alerts and alarms is an unrealistic expectation. Striving to make energy management a priority and proactively monitoring equipment for performance degradation is difficult with this flood of information. Competing business priorities such as employee morale, energy and environmental factors, production throughput, safety standards, etc. make the process of sifting through and prioritizing alerts a daunting task. Yet, the risk of a missed impactful alert can be expensive. Perhaps a better starting point is identify strategies to filter out the noise in the outputs from these various platforms so that the insight for action comes through clearly, and to the right people.

A framework that offers interoperability among multiple platforms will enable intelligent triage and analytics of alerts, alarms, and FDD results. This article presents the need for this common framework, in the real world context of a large data center site that is actively driving energy efficiency through ISO 50001 activities.

Background

The IBM facility in Boulder, CO is a 2.5 million square foot campus that includes data centers, offices and laboratories (see Fig1). The campus has been in operation since 1965 and is served by five chilled water plants, a central steam plant, and has nearly 170 air handling units.

[1] for companies with more than a 1000 employees

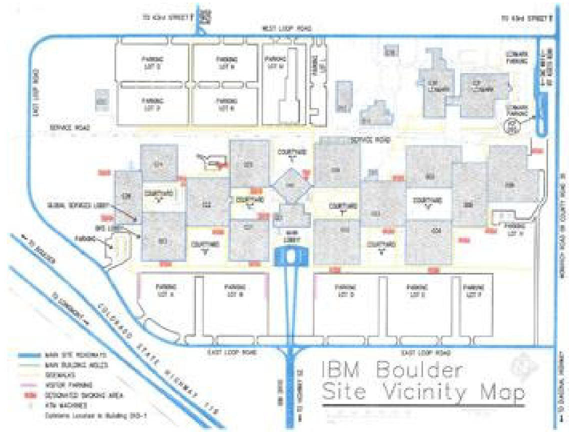

Figure 1. IBM Boulder campus

The campus utilizes a wide range of automation and controls systems, SCADA systems for sub-metering energy usage, historian databases to archive data and a comprehensive suite of analytic and support tools for data analysis. Graphet Data Mining analyzed the historical usage patterns and operating efficiencies of existing systems at the Boulder site as part of Xcel Energy's Process Efficiency Program. Xcel Energy's Process Efficiency Program is a holistic Strategic Energy Management program (SEM) to continuously improve performance by achieving energy and cost savings over the long term. SEM replaces the project centric approach that has dominated the landscape for many decades. It emphasizes equipping and enabling plant management and staff to impact energy consumption through behavioral and operational change (CEE 2014).

As a participant in Xcel Energy's PE program, IBM Boulder facility set management and technical priorities for energy. A key outcome from this engagement was an energy plan that established a set of short, mid and long term energy conservation opportunities.

These opportunities were identified by Graphet Data Mining by utilizing data and analytics in a site wide assessment. The savings estimated for IBM Boulder is 8-11% of current energy costs. IBM is in the process of implementing these recommendations by applying innovative approaches and machine learning to improve energy efficiency. A breakout of the energy usage based on historical data is shown in Figure 2.

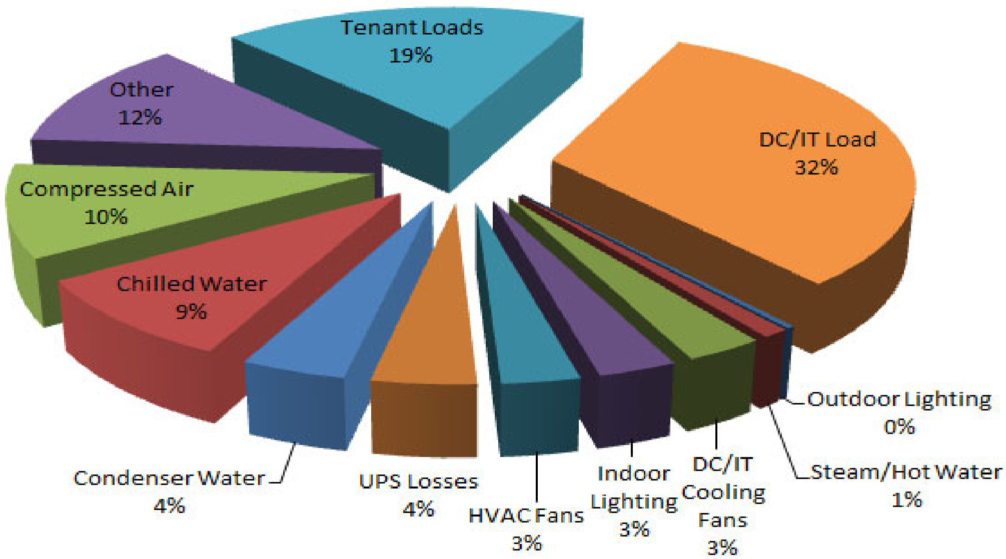

Figure 2. Breakdown of energy usage based on historical data for IBM Boulder campus

Graphet provided IBM with simple prioritized lists with simple visualizations to understand the relative importance of the areas to address and act upon. One such simple visualization is shown below:

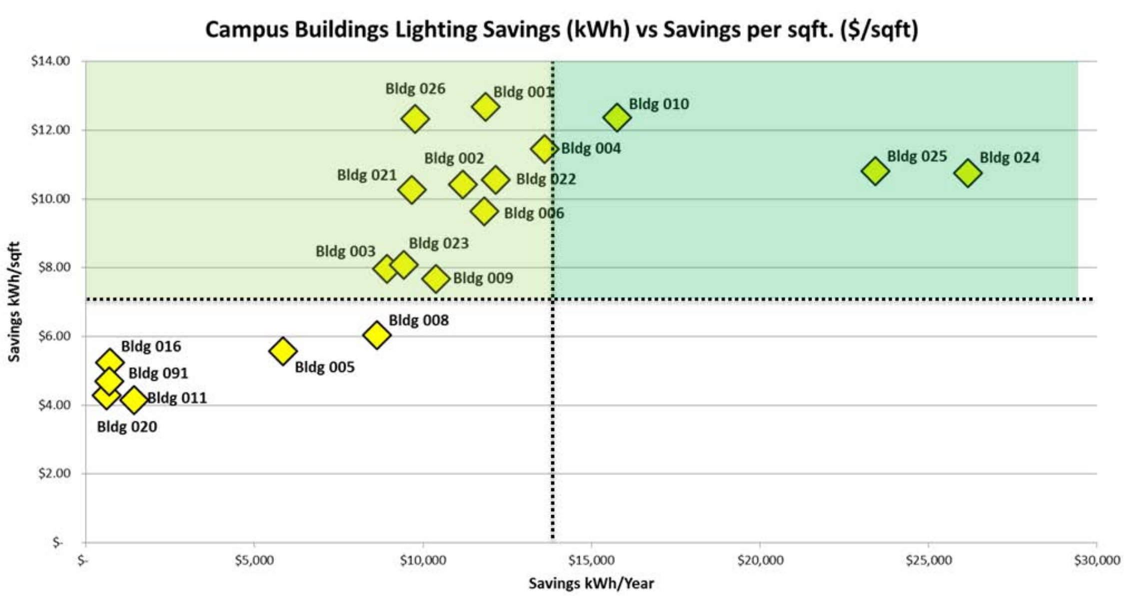

Figure 2 - Prioritized depiction of lighting savings opportunities for IBM Boulder Campus

Graphet's data mining process generated a set of easily digestible priorities for action by using the data already being collected at the IBM facility. The process of compiling, understanding and prioritizing inputs from disparate systems was labor intensive and required significant domain expertise. The shared experience and the desire to create a more sustainable environment provided the impetus for a framework that would help prioritize alerts through interoperability among the various systems and platforms.

Alerts, Alarms, Fault Detection and Diagnostics Environment

As the Internet of Things (IoT) emerges in the facilities management industry, the amount of data available and the corresponding analytics results are growing exponentially. In general the use of fault detection and diagnostics is becoming more pervasive, and moving from a single, centralized software platform, to a decentralized networked environment where alarms, alerts and diagnostics occur at every device and equipment level. For large facilities, the number of alarms, alerts, and FDD results can quickly grow beyond the capability of the, increasingly limited, facilities and energy management staff. This can often translate into an 'Analysis-Paralysis' condition rather than well-informed decision making.

In a recent study, more than 700 energy, sustainability, facility and finance professionals were surveyed about their interest and level of competency with Energy Management Systems. Of those surveyed 23% of the professionals cited interpreting data as one of their most formidable challenges. When asked about their organizations' investment in gathering and analyzing energy data, one-third of the participants said the majority of their sites have meters, EMS, or other data collection/monitoring devices installed. So, while the volume of data being collected is increasing, the challenge to interpret it in a way that drives actionable insights still remains (Ecova 2015).

In large, critical application environments significant events can have an "alarm cascade", wherein a single equipment failure triggers subsequent alarms on down-stream equipment. As system size and complexity grows the alarm cascade effect is amplified such that a single breaker trip (for example) can result in several hundred downstream alarms that might all be annunciated simultaneously. For field technicians in the "first responder" role, identifying root cause in the midst of an "alarm cascade" can become extremely difficult at a time when seconds in response matter.

The plant operations at the IBM Boulder facility are monitored and controlled by the four distinct types of software platforms defined in the previous section – details on their operation are provided here. These systems are somewhat hierarchical in their operation as listed here:

- Comprehensive Analytics Platform
- Third-Party Optimization Systems (Chilled Water Optimization, Data Center Cooling Optimization)
- Automation Systems (BAS or SCADA)
- Original Equipment Manufacturer (OEM) Controllers

Each of these systems generates alarms or notifications for operators and energy managers, many with their own unique path to the intended recipient. The ability of these systems to report alarms varies. Some OEM equipment email alarms to pre-defined email addresses, some automation systems are connected to a paging server and send out alarms via a paging system, others only display an on-screen notification.. These systems target a responsible recipient and require human intervention; they do not lend themselves to further automated data mining and analysis.

The output from these systems are themselves new data sources begging for the same integration and analysis the industry has achieved with low-level time-series data. Coordination and prioritization of alerts from control and data analytics systems is a must for effective energy management in the next decade. It would be useful to examine a few key aspects of these distinct software platforms before presenting a framework for their interoperability.

Comprehensive Analytical Software Platforms

Comprehensive analytical platforms have the capabilities to analyze data to identify issues, faults, and opportunities for improved performance and operational savings (SkyFoundry 2014). In order to represent the data in a meaningful way, especially when accessed from multiple sources, they use a tagging approach to capture the data and put the end-user in the driving seat by providing the ability to create custom rules based on hundreds of standard analytic functions. The rules, once activated, can be applied to new as well as historic data.

IBM‘s Real Estate Strategy and Operations (RESO) organization utilizes SkySpark’s analytical platform in conjunction with other IBM software tools like Tririga and Maximo. These tools provide an extraordinary opportunity to identify faulty equipment behavior. At the IBM Boulder facility, dozens of actionable alerts are generated each day. Currently the energy management/operations staff prioritizes action on these alerts based on knowledge, experience and understanding of the site.

Building Automation Systems

Building Automation Systems broadly refer to software platforms on which a network of equipment controllers are coordinated to control the building’s environment, operate the systems according to occupancy and energy demand, and alert the operator when systems function outside of the operational threshold. When facilities are controlled in this centralized manner, it creates an efficient environment for the facility staff to operate these systems and adopt sustainable practices to reduce energy spend.

While automation system providers are increasingly offering FDD as a feature, the majority of systems installed today are limited to threshold alarms with subsequent priorities and acknowledgement by the operator. Simple alarms like “*The motor was commanded to start but didn’t*”, “*temp sensor X is above Y threshold*” are the standard of BASs today. In large facilities with hundreds of motors, and thousands of sensors, they can create large volumes of alarms.

Third-party Optimization Systems

Third party companies develop software that employs proprietary methods and algorithms to analyze and optimize systems. This includes identifying energy conservation opportunities (ECOs) to meet energy saving goals of a company. These software products work at varying levels with the BAS, sometimes taking complete control of it (BuildingIQ 2015). Depending on the sophistication of the software, it can utilize real-time data, trended/historical data, and utility bills to optimize performance. These software tools provide features to view the performance of the systems usually in the form of a web-based user interface.

As the pervasiveness of products like these increases, the alerts they produce (while substantially less than from the BAS system itself) need to be integrated and considered alongside alerts being generated from other systems.

Original Equipment Manufacturers (OEMs)

Original Equipment Manufacturers may provide a basic user interface for given equipment with the capability to view alarms, adjust set-points, and set initial configurations. Increasingly, OEM's are providing FDD on-board their equipment as well. While the ease of integrating OEM controllers with building automation systems has improved dramatically, the alarms and FDD results coming from OEM controllers are not yet adequately integrated into the overall alarms & alerts framework.

The case for a Standardized Framework

The methodology for prioritization and response to alerts can vary significantly among facilities operations teams and response to alarms is usually a manual process. Urgent alarms are acknowledged and addressed immediately by operations staff. Less urgent alerts (especially FDD results) go through multiple levels of screening and prioritization. Alerts deemed as coming from bad sensors and bad data connectivity often do not require physical repairs, but nonetheless require changes to the data collections system or analytic engines generating the alert. Trend data analysis is often required to help filter true issues from false positives. Once deemed actionable, necessary repairs are entered into a work order system or Computerized Maintenance Management Systems (CMMS) to dispatch technicians for remediation. Figure 4 provides a snapshot of the alarm summary received by the Plant Manager. In this instance the facilities management team is collecting alarms from several source systems and aggregating them into a single feed.

The overall goal of any framework is twofold:

- Standardize meta-data to contain data necessary to route, prioritized analyze, and address alerts in a time efficient manner, independent of source.
- Standardize data structure/format associated with alerts to enable higher level analytics and prioritization to be accomplished with Machine Learning and Artificial Intelligence.

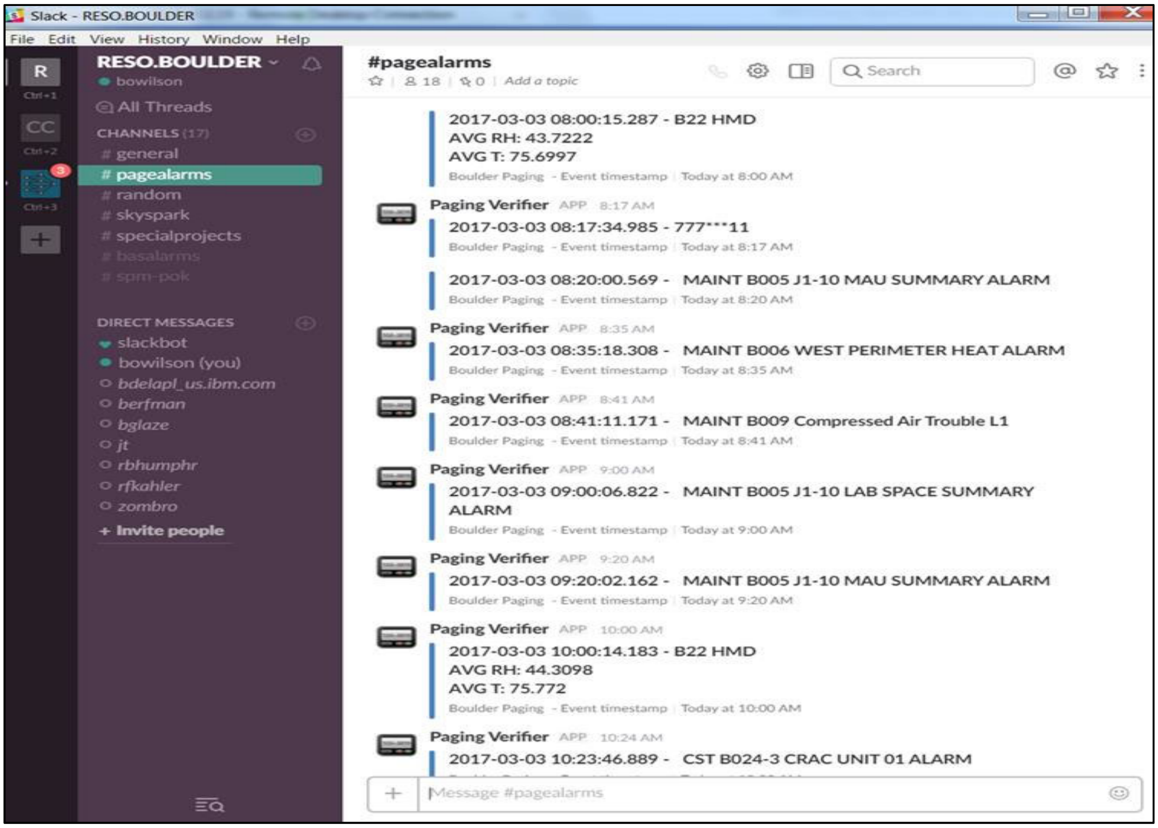

Figure 3-Typical alarm summary page received by a PM

Higher level analytics of alerts and alarms can provide significant benefit for the resiliency and efficiency of large sites. For example, the "alarm cascade" events described above make it challenging for human technicians to decipher root cause, but typically such events have a unique alarm "cluster" or "fingerprint". With some training and/or programed logic, machine learning algorithms could interpret an alarms cluster and report most likely root causes. An intelligent system might generate the work order automatically, rather than require the typical human analysis and manual work order generation. Another example might be a schema that continuously prioritizes alerts from all systems so alerts are delivered to specific personnel and each person's alert list is continuously prioritized. Other applications of alert/alarm analytics abound, but in large diverse sites, all require machine-to-machine interoperability and a common data framework.

Several efforts exist today to help independent controls and analytics tools share data seamlessly. Two such efforts are:

- Project Haystack (an open source initiative to streamline data from the IOT) and,
- cfiXML (a software independent system based on W3C schema developed by National Institute of Standards and Technology and Fiatech's Automating Equipment Information Exchange (AEX) project)

The identifiers or tags for each piece of equipment are distinct and comprehensive. These efforts aim to standardize the conceptual data models associated with operational data to speed up what is often a manual, labor intensive, data "mapping" process. To date however, extending these data models to alarms, alerts, and FDD results data has not yet been pursued.

Such efforts are but the first step toward achieving effective interoperability. For a schema of alert prioritization and routing to be accomplished, the people or machines that perform the prioritization and routing need to have detailed information about the system/equipment generating the alerts. This information could be stored in the alerts meta-data structure and carry the semantic data model of the equipment generating the alert.

While individual vendors may have sophisticated alert prioritization within their own platform, single platform/single vendor environments are increasingly rare. A cross-industry effort to standardize the modeling and taxonomy is necessary before vendor agnostic alert/alarm analytics is possible.

For example, at an industrial site, a single day's list of notifications will include alerts requiring immediate action and while other require future action, depending on the priority assigned by site personnel. For a large site, this list could include as many as 200 notifications in a day. It can be difficult for operators to identify those alerts that require immediate responses versus those that can tolerate a slower more proactive response.

Alert prioritization systems will require a high degree of customization. It is beneficial to consider the term "SMART"(**S**pecific,**M**easurable,**A**ctionable,**R**ealistic,**T**ime-based) as a mnemonic in this environment so that alert response decision-making aligns with business goals.

The MESA (Manufacturing Enterprise Solutions Association) has identified 28 metrics that are most utilized by manufacturers (Davidson 2013). These metrics can be broadly grouped into 7 classes. These are:

- Improving Responsiveness and Cycle Time
- Improving Quality
- Improving Efficiency
- Ensuring Compliance
- Reducing Maintenance
- Reducing Costs
- Increasing Flexibility or Innovation

Whether using "SMART" criteria or those offered by MESA, any alert prioritization system should aim to be an enhancement to, or extension of, the deep human knowledge and experience that is currently used to prioritize alert responses in large industrial sites. Technology now has the ability to capture our human experience and expertise for tasks like alert prioritization, but this technology will be more easily and effectively deployed if the players in the industry come around a common framework for alerts from all systems.

Requirements of Standardized Framework

The xml tag specification and/or project haystack tag definitions for identifying equipment may be too cumbersome to implement at the alert level at each piece of equipment. A more standardized mechanism for alert handling needs to be developed and implemented across the industry. At an information technology (IT) level, each device participating in the IOT, already knows how to identify itself. This self-identification capability should be extended to address the need for integration with other systems in a more meaningful way. One such idea could be to include IT's self-identification mechanisms (such as a mac or IPv6 addresses) to the specifications espoused in cfiXML or Project Haystack.

Interoperability is the ability of computer systems or software to exchange and make use of information and is the most important feature required to address alert prioritization. The standards set forth by project haystack or cfiXML are an important first step towards interoperability of systems and tools available for facilities. An increasing number of vendors are beginning to take advantage of these standards to create more seamless integration between systems. Some vendors, however, still lack APIs or use proprietary data formats that make interoperability difficult. Interoperability, not just of low-level equipment data, but of higher level alert data is required to:

- Identify cascading alarms and display in chronological order
- Combine and categorize alerts
- Provide embedded links in the alert itself to access the static/dynamic data associated
- Route the alerts to the appropriate personnel

As an example of what such a framework would look like, a system that borrows from the IT industry using machine learning to create recommendations is proposed. In the IT space, Big Data situations use Hadoop to create recommendations. These recommended systems can use opinion quintuples, meaning, and five attributes per alert that can serve as inputs for data mining, to automatically produce recommendations that can help guide site personnel actions. In the facility, this would translate to a quintuple comprising of:

- T – tag of equipment
- X– alert text about equipment
- M – measurement threshold crossed
- A – action taken
- S – severity of incident / or priority level

Table 1. Quintuple for Sample Alerts

#	T – tag of equipment	X– alert text about equipment	M – measurement threshold crossed	A – action taken	S – severity (priority)
1	cmd, cool, valve, water, equipRef, chilledWaterPlantRef, point	CHW Valve has been at 100% position for 24 hours	100% for 24 hrs	Technician investigate cause of excessive CHW use. Ineffective coil? Commanded output overridden?	6
2	condenser, pump, cmd, run, chilledWaterPlantRef, equipRef	BAS not following setpoint - Number of condenser water pumps requested = 1, number of condenser water pumps running = 2.	command vs. actual mismatch	Technician investigate why BAS not following optimization command. Point overridden? Bad BAS programming?	5
3	battery, run, status, UPS, power	UPS X in Building Y is discharging batteries	Emergency equipment running	Critical Systems Team responds within 5 minutes to ensure generators start normally, or UPS switches back to utility if available	1
4	run, status, pump, water, heating, equipRef, hotWaterPlantRef	HWP XX failed to start.	command vs. actual mismatch	In severely cold temperatures, maintenance technician responds immediately, If temps above 40F, work order is generated to investigate.	3
5	dataCenter, rack, inlet, temp, sensor, racRef	2% of rack inlet temperature sensors are greater than 77F.	2% of sensors above threshold.	Critical Systems Team investigates and determines of high temps are distributed (probably not a problem), or localized (indicating loss of cooling)	2
6	approach, temp, sensor, equipRef, chilledWaterPlantRef	Approach temperature of plate-in-frame heat exchanger has been above the "time to clean" threshold, for 48 operational hours.	approach >4F for 48 operational hours	CMMS receives alert and creates work order for heat exchanger cleaning. Work order is released after operator approval.	4

Table 1 shows how this classification can be effective. The Tag identifies the specific piece of equipment that can be cross referenced to static data such as equipment specifications. The text provides the context in which to evaluate the alert. The measurement threshold allows for comparison with operational requirements. The action provides a recommendations and insights and the severity allows for prioritization. The example above demonstrates that an innocuous message of UPS battery discharging is the highest prioritization in a data center, while the chilled water valve being 100% open needs more investigation and a proactive response.

Conclusion

This article has demonstrated the need for a comprehensive framework to lessen the reactive burden or operations managers. When a framework as proposed is established, we envision systems that would effectively prioritize analytics results, route them to the proper person, and make recommendations on initial actions. Eventually we imagine additional features such automated root cause analysis with appropriate work order generation, and consolidation of cascading alerts could be incorporated. The contours of the framework should address:

- Standardization of equipment specifications and operational characteristics
- Multiplicity of sources, that can self-identify and communicate
- SMART criteria to capture essential and relevant information for analysis
- Interoperability of analytical systems in a collaborative environment for sharing low-level equipment data as well as analytics results.
- Enhancements to alert generation to capture actions as well as the alert text itself

References

BuildingIQ. 2015. *Managed Services – The Future in Building Automation.* buildingiq.com/resources/white-papers/managed-services-wp/

CEE (Consortium for Energy Efficiency). 2014. *CEE Strategic Energy Management Minimum Elements*. Boston, MA. library.cee1.org/content/cee-strategic-energy-management-minimum-elements/

Davidson. 2013. 28 Manufacturing Metrics that Actually Matter (The Ones We Rely On). *LNS Research (blog),* October 09. blog.lnsresearch.com/blog/bid/188295/28-Manufacturing-Metrics-that-Actually-Matter-The-Ones-We-Rely-On

DOE (Department of Energy). Accessed 2017. *Improve Operations & Maintenance.* energy.gov/eere/buildings/improve-operations-maintenance

Ecova. 2015. *2016 Energy and Sustainability Predictions.* Spokane, WA. s3.amazonaws.com/uploads.ecova.com/2016/04/27220750/2016-Energy-and-Sustainability-Predictions-Report.pdf.Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A. h. Byers. 2011. *Big Data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute. www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

Simon, H. 1971. Computers, communications, and the public interest: Designing organizations for an information-rich world. Baltimore, MD: The Johns Hopkins Press.

SkyFoundry. 2014. Find What Matters™ in the Data from Equipment Systems and Smart Devices. *SkySpark*, 2014. skyfoundry.com/file/8/SkySpark-Overview-Brochure.pdf