

Modern infrastructure management: accelerating productivity through machine learning by, Peter Burris

November 7th, 2017

This is a Wikibon Voice of the Community Report, sponsored by ExtraHop Networks Inc. Voice of the Community posts are identified paid posts that appear on all pages of SiliconANGLE.com, supporting editorial efforts. Premise. Information technology infrastructure data can provide valuable intelligence for reducing business overhead, improving back-office and customer-facing processes and ensuring compliance with [...]



© 2017 Wikibon Research | Page 1

This is a Wikibon Voice of the Community Report, sponsored by ExtraHop Networks Inc. Voice of the Community posts are identified paid posts that appear on all pages of SiliconANGLE.com, supporting editorial efforts.

Premise. Information technology infrastructure data can provide valuable intelligence for reducing business overhead, improving back-office and customer-facing processes and ensuring compliance with key mandates. However, automating the distillation and delivery of those insights grows increasingly difficult as the scale and complexity of the infrastructure and its operational data skyrockets. Enterprises are swamped with ever greater volumes, varieties and velocities of operational data streaming in from public, private and hybrid clouds. In many organizations, this never-ending data avalanche threatens to overwhelm IT operational management staff. Machine learning is a promising tool for assisting IT infrastructure managers in deriving data-driven insights in support of root cause analysis, workload management, performance monitoring and other key operational tasks.

IT infrastructure processes almost every interaction, transaction and communication in our lives, and much of that infrastructure has already shifted to the cloud.

Cloud IT infrastructure has become the integration fabric for every type of application. For IT operations personnel, sifting through all of that cloud-sourced data — relevant to services, applications, networking, storage, devices and other IT infrastructure-can be a daunting challenge. They must monitor and correlate data at every level of the infrastructure stack — including hardware, networks, middleware and applications—and use it to manage service levels in real time. Just as important, they must leverage that data for anomaly detection, situation response, remote diagnostics, root cause analysis, capacity management and many other applications.

How can cloud IT infrastructure managers harness infrastructure data to extract the insights needed to do their jobs at every level of necessary detail? More to the point, how can they do this in the face of growing "data friction." This refers to obstacles that prevent infrastructure managers from extracting the operational insights they need to do their jobs effectively.

Data friction stems from the complexities that administrators encounter in the processes of collecting, aggregating, analyzing, and otherwise using this data to monitor, maintain and improve service levels. And it may be partly exacerbated by enterprise investments in underpowered analytics tools.

Analytics can be a key for reducing data friction and powering agile data-driven decision support in many operational scenarios. With infrastructure wire data, the key task-focused insights may remain beyond infrastructure managers' reach if they merely try to view them using simple queries, pre-existing reports and dashboards, and other standard analytic views. For deeper insights into more complex infrastructure issues, managers may need to call on more sophisticated real-time, historical and predictive analytics.

To explore these issues, Wikibon convened a Voice of the Community CrowdChat to discuss data friction and the techniques — such as machine learning – for addressing it proactively. For an hour in September 2017, 58 industry experts who are part of the Wikibon community met online to discuss this topic. The conversation comprised 226 collaborative observations that generated close to 1 million customer impressions on social-media channels.

What follows are key findings that were surfaced in the CrowdChat:

- **Infrastructure management data is growing rapidly**: In the CrowdChat, participants cited annual growth rates in infrastructure data volumes of between 200 to 500 percent as typical.
- Data growth challenges infrastructure managers to find golden needles in unfathomably massive haystacks. In the CrowdChat, participants repeatedly commented on the fact that infrastructure data's expanding scale, scope, and speed make it difficult to drill through it all for task-specific operational intelligence.
- IT managers are well aware of the promise of data-driven infrastructure intelligence: In the CrowdChat, participants agreed that there's strategic decision-support potential locked up in infrastructure data, but the requisite analytic and data management tools and analytics are largely lacking from their operations.
- IT managers struggle against constraints that prevent them from evolving their operational analytics: In the CrowdChat, participants generally agreed that IT professional would generally embrace better infrastructure data analytics, but are often hamstrung by constraints in staffing, skills, tooling, budget and other areas that deter or discourage them from taking the necessary next step.
- IT managers have limited experience using sophisticated decision-support tools: As discussed in the CrowdChat, data-driven decision making tends to be limited to a senior cadre of business analysts in most organizations, though acknowledged widely as an aspiration. Infrastructure data can be fearsomely hard for users to find, aggregate and analyze, though machine learning can be useful in distilling relevant patterns and correlations.
- IT managers have varying confidence on the quality of their infrastructure data: As discussed in the CrowdChat, IT managers, like users in general, have varying degrees of confidence in the correctness, consistency, freshness and other quality metrics surrounding their infrastructure data. Confidence in the data hence in any data-derived infrastructure insights depends on the extent to which the manager has visibility into the data's provenance, handling and lineage.
- Machine learning, as a tool for surfacing IT infrastructure data insights, is embryonic: As noted in the CrowdChat, IT infrastructure managers' operational analytics have barely begun to incorporate machine learning for incident monitoring, anomaly detection, root cause analysis, predictive maintenance and other critical data-driven inferences. One key reason for this is that commercial IT operations management tools have only recently begun to incorporate machine learning. Another factor is that enterprises that have invested in IT operations management tools may lag considerably in upgrading or migrating to solutions that benefit from machine-learning-driven analytics.

What is a Crowdchat?

CrowdChat is a community engagement tool used by Wikibon to research innovation. A Crowdchat brings together – online – experts in a domain to discuss complex technology, social, and business issues. Wikibon posts questions to these experts, which catalyzes a bloom of conversational interactions about the subject. Wikibon analysts then combine these interactions with other research sources to develop the findings that we publish in a Voice of the Community research paper.

Wikibon will now drill down into the free-form comments and poll results that were produced in this



CrowdChat, revealing the Voice of the Community on particular issues and opportunities.

The Voice of the Community

... on growth in the volume of infrastructure data

Infrastructure management data is growing rapidly. In the CrowdChat, participants cited annual growth rates in infrastructure data volumes of between 100 and 500 percent as typical in many organizations. Much of this growth stems from the expanding scale and complexity of the infrastructure from which the data is being sourced and in which it's being consumed. Another key factor is the fact that more applications have data and analytics at their heart. Yet another growth factor is the expansion of edge computing, especially the Internet of Things, which is a non-stop source of data sensing, capture, and inference.

As regards the growing scale of infrastructure data, noteworthy comments from the community included:

- "doubling annually... volume is growing related to increased variety of metrics being captured... I see scale-out app architectures and associated service policies driving the need for increased measurements"
- "growth is as fast as enterprises can capture the data. It's not the 40 percent CAGR IDC and EMC keep citing — 40 percent is just Moore's Law. ...One way to think about growth: number of entities in the application and infrastructure landscape (ever more fine-grained) emitting ever more telemetry per entity"
- "Retail supply chain is driving enormous data sprawl."
- "for our customers it's massive growth... !... More anecdotal, but for many customers it's 2-5x annually.
 ...IT Ops is becoming a data-driven practice, hence growth. Need drives growth!... More anecdotal, but for many customers it's 2-5x annually"
- "Management data is slower than data overall, but not by much. It's massive growth. Easily 40 percent+ a year.....I just think there's a naturally offset correlation. You deal with more bytes in/out as things like HD vid, audio streams, higher graphic content etc. become the norm. But those things don't automatically drive more mgmt data."

Figure 1 presents a poll of CrowdChat participants on the subject of infrastructure data growth:





Figure 1: CrowdChat Poll: How rapidly is your enterprise infrastructure management data growing in volume?

... on the ease of extracting value from infrastructure data

Data growth challenges infrastructure managers to find golden needles in unfathomably massive haystacks.

In the CrowdChat, participants repeatedly commented on the fact that infrastructure data's expanding scale, scope, and speed make it difficult to drill through it all for task-specific operational intelligence. In addition, the ever-expanding range of data-driven decision support scenarios challenge infrastructure managers to implement analytics that are agile enough to address future query, reporting, dashboarding and other decision-support use cases.

As regards the ease of distilling actionable insights from infrastructure data, noteworthy comments from the community included:

- "Generally speaking? Harder. Rapidly. Exponentially. With the right platforms/tools? Definitely still possible and perhaps "easy". But with tools of yesteryear? Bordering on impossible soon."
- "Depends on the use case environment and again, industry segment I think. Analytics are becoming very broad based and flexible good and bad. Target rich, accuracy poor in many cases."
- "Expectations for value are increasing, so once you engage in it, you find it to be harder because you begin to discover what data is valuable and what isn't and what sources you actually need"
- "...and extracting value depends on use case much of the time, ie meaning: incident response or outage: need data *now*. forensics / capacity planning / etc... historical is needed"

Figure 2 presents a poll of CrowdChat participants on the subject of ease of extracting relevant and actionable insights from infrastructure data:

Is it getting harder or easier to extract value from your infrastructure data? (one response only)



Figure 2: CrowdChat Poll: Is it getting harder or easier to extract value from your infrastructure data?

.... on whether infrastructure data has become a strategic asset

IT managers are well aware of the promise of data-driven infrastructure intelligence.

In the CrowdChat, participants agreed that there's strategic decision-support potential locked up in infrastructure data, but the requisite analytic and data management tools and analytics are largely lacking from their operations. A key consideration is the extent to which organizations can create assets — such as machine learning and predictive analytics models — that are derived from infrastructure data and can be reused flexibly to address future decision support and automation requirements for IT operational management.

As regards infrastructure data's strategic potential, noteworthy comments from the community included:

• "Have to strike the balance – level investment in tooling and resources with derived (and derivable) value.... Winners will be the ones most able to commoditize pattern data quickly and repeatably."

Figure 3 presents a poll of CrowdChat participants on the subject of the strategic business value of infrastructure data:

Is your infrastructure data becoming more of a strategic asset or overhead burden? (one response only)



Figure 3: CrowdChat Poll: Is your infrastructure data becoming more of a strategic asset or overhead burden?

... on whether the IT organization is a facilitator in gaining value from infrastructure data

IT managers struggle against constraints that prevent them from evolving their operational analytics.

In the CrowdChat, participants generally agreed that IT professionals would generally embrace better infrastructure data analytics, but are often hamstrung by constraints in staffing, skills, tooling, budget and other areas that deter or discourage them from taking the necessary next step. Though IT professionals take pride in their data-driven ingenuity, they also generally operate as a cost center that makes the best of whatever resources (albeit limited) are at their disposal.

Figure 4 presents poll of CrowdChat participants on the subject of IT managers' commitment to maximizing the operational business value of infrastructure data:







Figure 4: CrowdChat Poll: Is your IT organization a facilitate or obstacle to gaining value from your infrastructure data?

... on the extent to which infrastructure data-driven insights permeate their organizations

IT managers have limited experience using sophisticated decision-support tools.

As discussed in the CrowdChat, data-driven decision making tends to be limited to a senior cadre of business analysts in most organizations, though acknowledged widely as an aspiration. Infrastructure data can be fearsomely hard for users to find, aggregate, and analyze, though machine learning can be useful in distilling relevant patterns and correlations. Of course, this situation parallels the meager adoption of operational business intelligence tools by knowledge workers in general, which consistently is limited to 5 to 10 percent of organizational staffs.

As regards IT management groups' use of infrastructure data-driven insights, noteworthy comments from the community included:

- "you have to know how to use and what you are doing. data itself is kind of useless unless you know what you are doing and can use the right tools"
- "Agree. But there are also cases where data that's useless today, and tomorrow, and the next day becomes invaluable 10 years from now. See: The CDC."
- "This is interesting, and 100 percent true. Data's value changes over time depending on the use case.
 Too few take this into consideration, let's hope that changes :).... From an ops perspective as an ex-Ops person myself, a very simple value model I use is (velocity/friction)
- the number of users that can put the data into action. Not academically rigorous but folks find it useful."
- "most applications today are vomiting data state, status, condition in a vacuum. not actionable need

AI to synthesize and deliver accelerated insight"

... on the ease of finding infrastructure data and associated analytics

IT managers have limited experience using sophisticated decision-support tools.

As discussed in the CrowdChat, data-driven decision making tends to be limited to a senior cadre of business analysts in most organizations, though acknowledged widely as an aspiration. Infrastructure data can be fearsomely hard for users to find, aggregate and analyze, though machine learning can be useful in distilling relevant patterns and correlations. Bear in mind that the same can be said for big data in general, especially as it grows in volume, velocity and variety.

As regards the ease of discovering relevant infrastructure data and analytics, noteworthy comments from the community included:

- "big issue is not the data, its which answers u r looking for... start w a goal (e.g. oper efficiency, security, downtime ..) define the data u need from there"
- "In other words, how much friction is there for people trying to find the data and analytics: Do they need to know a query language? Can they easily search? How long do they have to wait?.... There is a "half-life" to data value and it differs based on the analytics use case. With incident response and downtime, the half-life of data value is shorter, but for long-term capacity planning the half-life is longer."
- "I'll parse this along the boundaries of data sources: Machine data (logs, snmp, etc.) Agent data, synthetic data, and wire data. Most are quite easy to get at, it's harder to wrangle it into value."
- "No common reference model for what is the data, making it difficult to know what to find. Machine learning can really help there.... But as IT resources become more core to business behavior, a more diverse user community will demand simple ways to find, collaborate on, and act on answers."
- "Varies *wildly* with tooling and instrumentation. For some I've spoken with it's hopeless in their current deployment. For others it's seconds to visibility and insight. It's all about surfacing insights rapidly, automatically."

Figure 5 presents a poll of CrowdChat participants on the subject of easy of finding infrastructure data and associated analytics:



Figure 5: CrowdChat Poll: How easy is it for user to find the infrastructure data & *associated analytics they need?*

... on the veracity of infrastructure data

IT managers have varying confidence on the quality of their infrastructure data.

As discussed in the CrowdChat, IT managers, like users in general, have varying degrees of confidence in the correctness, consistency, freshness, and other quality metrics surrounding their infrastructure data. Confidence in the data — hence in any data-derived infrastructure insights — depends on the extent to which the manager has visibility into the data's provenance, handling and lineage. When it comes to "wire data" — in other words, raw data that streams direct from the applications and infrastructure nodes being managed — confidence tends to be high, due to the fact that it is perceived to be trustworthy (albeit raw and therefore difficult to interpret without further analytics).

In this regard, noteworthy comments from the community included:

What does the Wikibon community say about CSPs and innovation?

• "Very high confidence in highly stylized data. Modest confidence in correlated data. Limited confidence in data that ties to business activity."

Figure 6 presents a poll of CrowdChat participants on the subject of confidence in the veracity of infrastructure data:





Figure 6: CrowdChat Poll: How much confidence do users have in the veracity of infrastructure data?

... on using machine learning to unlock insights from infrastructure data

Machine learning, as a tool for surfacing IT infrastructure data insights, is embryonic.

As noted in the CrowdChat, IT infrastructure managers' operational analytics have barely begun to incorporate machine learning for incident monitoring, anomaly detection, root cause analysis, predictive maintenance and other critical data-driven inferences. One key reason for this is that commercial IT operations management tools have only recently begun to incorporate machine learning. Another factor is that enterprises that have invested in IT operations management tools may lag considerably in upgrading or migrating to solutions that benefit from machine-learning-driven analytics.

In this regard, noteworthy comments from the community included:

• "Some form of data reduction/analytics via machine learning makes the analytics supporting the operational decision making process more tractable."

Figure 7 presents a poll of CrowdChat participants on the subject of whether ML has made it easier to unlock insights in infrastructure data:





Figure 7: CrowdChat Poll: Has machine learning made it easier to unlock insights from your infrastructure?

Summary

Data friction stems from the complexities that administrators encounter in the processes of data ingestion (e.g., discovery, collection, transformation), governance (e.g., aggregation, cleansing, version control) and consumption (e.g, delivery, analysis, visualization, contextualization, anomaly detection, alerting). Increasingly, Wikibon sees infrastructure managers applying the most sophisticated machine learning, deep learning, and artificial intelligence tools for operational decision support. These technologies can help to elucidate the most noteworthy correlations, anomalies and other insights using statistics-based approaches such as artificial neural networks, support vector machines, and cluster analysis.

Action Item. Wikibon recommends that IT managers use the following multifaceted approach for deriving actionable insights from their infrastructure data:

• Aggregate infrastructure data into an IT operations data lake. Having a single repository or registry of infrastructure data can be the foundation for IT operations simplification, speeding query, reporting, dashboarding, predictive analysis, machine learning, and other analytical tools for harvesting actionable insights. As you consolidate data into this lake, make sure that it's been matched, merged, cleansed, and augmented so as to enable development of high-confidence, relevant decision-support tools for IT administrators. Also make sure that the data lake enables fine-grained lineage analysis into the sources of the data and into how it has been transformed, cleansed, and otherwise processed.

- **Provide self-service advanced analytics tools for IT administrators**. If IT management professionals can easily query, explore, and build customized analytics to leverage infrastructure data in the data lake, this can accelerate data-driven operations tasks and enable greater administrative productivity. Just as important, boosting IT operations productivity requires a comprehensive range of descriptive, exploratory, streaming, predictive, and prescriptive analytics have been pre-built for a full range of tasks. At the very least, these tools must accelerate and automate more of the monitoring, alerting, anomaly detection, root cause analysis and closed-loop remediation tasks that form the core of IT operations management.
- Leverage machine learning to sift through IT infrastructure data patterns. Drilling for insights in growing volume of IT infrastructure data demands increased reliance on machine learning tools for finding the relevant correlations, outliers, and anomalies in and across myriad data sets, both in real-time, with respect to historical trends, and with a keen eye on predictive likelihoods. In addition to relying on IT operations management tools with prebuilt machine learning tools for particular tasks, organizations should also bring more data scientists into their organizations to build and supervise machine learning models that are trained and tuned to their specific requirements.

Peter manages Wikibon and directs research at SiliconANGLE Media. His own research focuses on digital business transformation, data value, and cloud optimization. He has over 30 years of experience as an IT practitioner, analyst, research leader, and executive in technology and business. Prior to joining Wikibon, Peter ran B2B CMO and CIO research teams at Forrester, was co-Research Director at META Group (now Gartner) where he drafted the seminal Adaptive Infrastructure Strategies program and has worked as an IT and marketing leader at large, global enterprises. He is a well-regarded speaker on technology infrastructure trends and digital business transformation. He graduated from Yale University and now lives in Silicon Valley, CA.



Peter Burris

@plburris peter@siliconangle.com



© 2017 Wikibon Research | Page 14