research

CHECKLIST REPORT

2018

# Six Steps for Generating Business Value from Operational IoT Data

## How IT and operations can harness the potential of machine data and analytics for improved operational efficiency and business performance

By David Stodder

Sponsored by:

OSIsoft.

tdwi
Transforming Data With Intelligence™

DECEMBER 2018

TDWI CHECKLIST REPORT

# Six Steps for Generating Business Value from Operational IoT Data

How IT and operations can harness the potential of machine data and analytics for improved operational efficiency and business performance

By David Stodder

tdwi

**Transforming Data With Intelligence™**

555 S. Renton Village Place, Ste. 700
Renton, WA 98057-3295

**T** 425.277.9126
**F** 425.687.2842
**E** info@tdwi.org

**tdwi.org**

## TABLE OF CONTENTS

## FOREWORD

In its research, TDWI regularly finds that most organizations view improving operational efficiency and effectiveness as their top priority. As the pace of digital transformation and instrumentation quickens, organizations need operational intelligence—that is, the ability to collect, integrate, and analyze a variety of data, including streaming and event data, in time to support smarter operational decisions and actions. If managers in operations can gain data insights that enable them to reduce waste, anticipate maintenance, allocate resources better, and cut unnecessary steps to streamline processes, they can deliver significant top- and bottom-line business benefits. The more IT can enable operations to use data effectively, the more IT will share credit for realizing these benefits.

There's no shortage of data in operations. The question is how to gain value from it. The Internet of Things (IoT)—the growing network of connected physical objects that are embedded with sensors, software, and other technologies—is becoming a major contributor to the tsunami of operational data. TDWI research anticipates that IoT data collection will grow by 200 percent in the next few years; other research firms say that the number of interconnected devices, from parts in a tractor, truck, or washing machine to personal mobile devices, could hit 50 billion as soon as 2020.

IoT will drive new levels of competition based on the quality of operational intelligence. With greater instrumentation producing important data, functions in organizations that may not have previously focused on how to use data to increase their operational excellence will need to consider solutions and practices that enable them to ingest and analyze IoT and other operational data. These functions include IT, which today must monitor and manage a growing array of assets. Others could include field sales, marketing, customer support, product and service development, and risk management.

This TDWI Checklist focuses on how organizations can develop strategies for improving operational intelligence, particularly by realizing value from IoT data. Although collecting this data is important, organizations obviously want to do more than that. They need to prepare it well and integrate it with other data so they can analyze and visualize trends and patterns for predictive insights. They also need to enable real-time awareness of events that require immediate response if conditions warrant.

Organizations need to consider what data management and analytics strategies are best for realizing the potential of operational data such as that generated by IoT activity. Additionally, as IoT grows, organizations need to recalibrate governance and security for appropriate protection—topics that this Checklist will also address.

## NUMBER ONE

CONSOLIDATE IOT AND OTHER OPERATIONAL DATA TO SHORTEN PATH TO VALUE

To tap the value of IoT data, it is important for organizations to develop a consolidation strategy. The goal should be to establish an operational core of IoT data that can be accessed and used in real time. Organizations also need a strategy for enabling integrated views of IoT and other operational data so decision makers can understand it completely.

IoT devices typically use different networks, protocols, and technical standards for communication and interoperability than those for regular operational data. This adds complexity to the problem of ingesting and consolidating the data that IoT devices generate. Some IoT systems use dedicated networks calibrated to how much data they transmit, how frequently, and how much power they require. Thus, as an organization creates its strategy for consolidation, it is important to tap IT networking expertise to ensure a good flow of data for operational reporting, management, and analytics.

Operational data originates from many sources; IoT merely increases the number, often considerably. An oil and gas company, for example, might have IoT sensors on an array of drilling, transportation, refining, production, and other equipment from which it wants to collect data for monitoring and analysis. Most likely, the firm is already managing operational data from traditional seismic and other types of monitors as well as from control systems, asset management, supply chain management, and business applications. Traditionally, organizations have ingested operational data from sources into applications and data management systems by developing numerous point-to-point custom programs. Unfortunately, the result is often "spaghetti code" that is difficult to maintain.

The strategy of some organizations is to create a data lake to centralize the collection of operational data, including that from IoT sources. Data lakes built on Apache Hadoop clusters can offer horizontal scalability to accommodate big data volumes. Data lakes can appear to be a solution to consolidating data because it is easy to ingest raw, detailed data into them without first manipulating it to conform to a predefined schema and data model. However, ultimately, the data has to be restructured and contextualized whenever it is read from the data lake. This can increase latency and preparation overhead.

In addition, although a data lake may make it easier to centralize data collection, it does not solve how to get the data *out* to meet reporting, analytics, or other downstream requirements. Organizations have to resort to more custom coding to organize and extract data, particularly for the operational time-series analysis that is critical for improving operations. This can add

further complexity to existing spaghetti code problems and extend project timelines.

Most operational data is time-series data; when collecting this data, it is important to preserve both the source of the data as well as the chronological order of its arrival. Organizations will want to monitor and analyze how operational data changes over time. IoT sensors can enable organizations to gain discrete measurements of changes in systems such as the rate at which objects carrying or monitored by sensors are moving through factory processes and behavioral or quality changes over time. An engineer, for example, could learn of unexpected behavior in a pump as recorded by changes in IoT sensor data. Discovering the rate of performance degradation for a pump that has not been maintained for over a year can help a company move to condition-based rather fixed schedule maintenance and help engineers decide to take action and prevent a catastrophic failure.

Data management systems that can reduce manual work and reliance on point-to-point spaghetti code to collect and organize operational data can enable an organization to focus greater attention on time-series analysis and resulting insights. Organizations should evaluate data management systems for how well they collect and store the time-series context of the data and support time-series monitoring, reporting, and analysis. If organizations are collecting IoT and other production data in a data lake, they will need additional programs to understand changes over time and recognize potential faults or anomalies in the data because no data is ever perfect.

## NUMBER TWO

REDUCE DATA PREPARATION DELAYS AND INCONSISTENCIES

The sooner users and their applications have access to data ingested from operational systems for visualization and analysis of events, trends, and patterns, the sooner organizations can take appropriate action. Unfortunately, data preparation—that is, all the steps that must occur to make raw data ingested from one or multiple sources usable for the requirements of users and applications—can be notoriously slow. Organizations need to examine where they can apply technologies and practices to reduce the time it takes to collect and consolidate the data, improve its quality and completeness, standardize how it is defined and structured, and enrich and transform the data as necessary.

IoT and operational data present preparation challenges. Data collection is the one of the first steps in the data preparation process. As noted in the previous section, procedures for collecting and analyzing this data are often either fully manual or involve custom, case-by-case programming. Plant engineers, for example, frequently collect plant data manually from numerous sources, in some cases where it is difficult to collect due to remote locations or hazardous operating environments. As a result, it may only be collected once a day, which can leave parts of the plant, including vital equipment, unmonitored for long stretches of time. Deployment of IoT sensors can be helpful in such situations if the organization can automate data collection and prepare the data as needed for downstream use.

However, TDWI finds that much manual data preparation, including cleansing and completing the data, is done on spreadsheets. This can present consistency and quality problems for organizations because there is typically little standardization and documentation of macros, formulas, and processes undertaken to prepare the data. In addition, some attributes of operational data are hard to capture and interpret using spreadsheets. Consider, for example, the difficulty of defining discrete events occurring in continuous streams of IoT data. Yet, it is critical during data collection to define points in time and reflect changes in the state of plants, equipment, or other objects between points in time.

Manual data preparation through use of spreadsheets can also introduce inconsistencies in how parts in hierarchies and those in nested processes are handled. If these are not correctly represented in the data, equations for uses such as tracking temperatures, speed, and other changes could be incorrect. In addition, even sophisticated spreadsheet users often struggle to account for, prepare data for, and enable analysis of all the attributes that describe a complex asset such as a pump or a machine. If errors and inconsistencies become embedded in the data, it can take weeks or months for organizations to uncover them, which can have many negative repercussions, not the least of which is a lack of confidence in the data for reporting and analytics.

It is important for organizations to invest in solutions that can handle data preparation requirements, especially as they begin to work with fast, continuous, and high-volume streams of IoT data. Such solutions reduce bottlenecks and streamline the flow of data from ingestion through preparation so users and applications have properly enriched and transformed data that can support a variety of use cases.

IT can play an important role in stewarding users to apply common data preparation processes and best practices for producing trusted, quality data. However, IT must work closely with subject matter experts in operations who are closest to the data and know best how it should be collected and prepared to preserve its meaning and value.

## NUMBER THREE

DEVELOP ANALYTICS TO UNCOVER INSIGHTS THAT SOLVE PROBLEMS

Analytics helps organizations solve problems, reduce costs, and innovate by improving processes to better serve customers and business partners. As analytics advance into data science and artificial intelligence (AI), projects mix elements of statistics, programming, computer science, operations research, and data visualization, typically with the objective of developing models and hypotheses to learn from the data. However, with all the attention today on advanced analytics and big data science, organizations should not overlook the potential of analytics to address the "low-hanging fruit"—that is, less spectacular and more easily achievable solutions to everyday problems.

Maintenance offers a prime example. Many studies show that unplanned downtime due to either poor or unnecessary maintenance is a significant factor driving higher costs in many industries. Analytics can help organizations establish smarter maintenance, potentially saving costs that are incurred either through unneeded procedures or by not bringing attention to conditions that require immediate action to prevent damage to components or avoid a larger systems failure. Organizations frequently base maintenance procedures on the failure-rate history of a component. The history might indicate that a pump, for example, will fail every 10 years. Without analytics, preventive maintenance procedures based strictly on the failure-rate history could either force unneeded maintenance or cause personnel to overlook a case where a pump needs maintenance sooner.

Analytics combined with data visualization and alerts can help organizations study patterns in data from IoT devices and other operational sources to discover faults sooner, potentially in real time. Then, organizations can calibrate preventive maintenance based on actual data rather than on only historical records that could be inaccurate or not based on a complete view of the variables affecting maintenance needs. Analytics can help organizations spot simple changes in procedures, such as more frequent monitoring, that could lead to more efficient maintenance.

The inclusion of IoT data streams gives organizations important data for running predictive models focused on determining what conditions are most influential in causing problems in systems, equipment, and processes. Organizations could perform time-series analysis to examine a sequence of events recorded by sensors to predict what is likely to occur within specific time intervals.

Blending analysis of different types of data is critical. With access to data gathered from perhaps thousands of IoT sensors in a consolidated data management system, users can blend into their analysis geolocation data to see if placement of resources on a factory floor and variables such as temperature and humidity based on location could be affecting asset performance. Logistics and fulfillment operations could analyze blended data to better understand and respond to delays in shipments. In these ways, analytics can help operations managers develop practical solutions.

## NUMBER FOUR

USE SELF-SERVICE VISUALIZATION TO MAKE DATA AND ANALYTICS ACTIONABLE

Collecting, preparing, and analyzing data are critical steps in enabling executives, operations managers, and frontline personnel to make intelligent, data-driven decisions and take action with greater confidence. However, data visualization is what makes insights actionable. Organizations need to ensure that users can develop, share, and apply visualizations to interact with data quickly and correctly.

Visualizations such as dashboards are increasingly common as the users' point of interaction with data and analytics. Some dashboards are simple; they are primarily for data consumption through charts, gauges, and other visual objects that show status of key performance indicators (KPIs) and additional metrics. In-depth dashboards and graphical user interfaces (GUIs) offer greater selections of visual objects to fit a fuller range of data consumption and analytics needs. For example, a user might select a colorful heat map to show contrasts between data values to make them easier and faster to understand than if they were in a spreadsheet, table, or text comment.

Operational personnel need flexibility in their dashboards so that as new sources of data or new analytics requirements arise, personnel can adapt. For many organizations, the need for flexibility has driven growth in self-service visualization and analytics. Users do not want to be dependent on IT for every change to visualizations or for new data they need to incorporate into their analysis. With drag-and-drop GUIs and wider selections of visualizations, modern self-service visualization and analytics solutions offer advantages over spreadsheets, desktop databases, and simple reporting applications. Spreadsheets in particular can require numerous calculations that only expert users can develop and run to produce the visual reports and dashboards that self-service tools can increasingly create automatically.

The trend toward cloud services is enabling self-service visualization and analytics solutions to implement standard APIs that make it easier to pull in a wider range of data, not just what is stored in on-premises servers. Cloud-based solutions can also tap cloud computing's elastic scalability to enable organizations to use only

the computing power they need to respond when data volumes and analytical complexity rise or recede.

Three areas demand attention to ensure that visual data and analytics are actionable for operations:

- **Performance metrics.** KPIs and other metrics are commonly used for defining and communicating objectives and giving operational managers the means to measure performance against thresholds. Visualizations can enable managers to easily compare actual performance against forecasts. However, TDWI research finds that users get frustrated if they are unable to drill down to deeper levels of source data, including IoT sensor data, to understand KPI results and consider what actions to take. Organizations should evaluate technologies and improve practices to ensure that users can integrate analytics into performance management and interact with data through visualizations to more fully understand KPIs and metrics.
- **Prescriptive insights.** Analytics is evolving from purely descriptive analytics about what has happened to advanced predictive and prescriptive analytics. Predictive analytics focuses on developing models for evaluating patterns in data and providing insight into what will happen. Prescriptive analytics is about building recommendations for the best course of action based on predictive and descriptive insights. Organizations should evolve visualizations toward making prescriptive insights part of actionable information; then, operations personnel could be made aware of situations and receive response recommendations.
- **Real-time data views and analytics.** IoT data is most valuable if it can be viewed in, at, or near real time. Visualizations must become capable of handling large streams of real-time updates, which users might want to view as part of time-series analysis visualizations. Using real-time data views, organizations can set up alerts to draw attention to critical situations discovered in analysis of IoT data. Organizations should, however, avoid "alert fatigue"—that is, having too many alerts that either distract personnel or cause them to take none of them seriously. Organizations should differentiate between situations that demand only passive monitoring of conditions over time and scenarios where real-time alerts are warranted.

## NUMBER FIVE

ENCOURAGE OT-IT COLLABORATION TO ACHIEVE DATA-DRIVEN ENTERPRISE GOALS

IoT, the rise of the Industrial Internet, and the growing instrumentation of many aspects of business environments are all part of a larger objective held by many organizations to become data-driven in decisions and actions. Technology solutions are vital to achieving data-driven ambitions; but an organizational culture is also required that supports the collection and use of data, the development of analytics insights, and the operationalizing of insights into business processes and actions. To become more data-driven by using IoT and operational data effectively, organizations need their operational technology (OT) and IT functions to improve communication and collaboration.

OT focuses on technology development and management for manufacturing, processing, and other industrial control systems. Thus, the development of the Industrial Internet is happening primarily in the OT environment. In most organizations, OT has grown separately from IT; OT has focused on systems that run key business processes and amass data records of time-sensitive, process-oriented activities. IT, on the other hand, has traditionally focused on managing the business. IT manages primarily enterprise business systems, business applications, networks, security, and data management designed to record the state of the business.

IoT development brings OT and IT responsibilities closer together for two reasons. First, because getting value out of IoT requires new strategies for networks, systems, security, and data management, all traditionally part of IT's domain. Second, OT and IT have a shared interest in the data; it can be valuable for both running operations and managing the business.

IoT data can provide new perspectives on business management issues such as productivity and efficiency, inventory and resource allocation, and consumer market behavior. IoT data can offer various functions in organizations a more detailed understanding of how to optimize processes, ensure quality, and improve asset health. For these reasons, OT and IT will need to collaborate to ensure that the collection, preparation, and management of data generated by IoT and other operational systems meet the needs of both operational and business functions.

Here are two ways in which OT and IT can work together to meet the larger objectives of data-driven organizations:

- **Empower subject matter expert (SME) decision makers.** Most IT organizations do not have the knowledge to supply OT decision makers with all the data, visualizations, views, and analytics they need. Operations teams frequently apply domain expertise in thermodynamics, physics, the

development of empirical models, and other disciplines as they work with IoT and other operational data. Similarly, business decision makers apply their own domain expertise. SMEs on both sides can provide critical knowledge for defining metadata, master data, and important attributes and variables for analytics. Self-service technologies and cloud services can help SMEs do more on their own and enable IT to focus on facilitation, provisioning trusted data, and the management of infrastructure resources.

- **Share leadership through a center of excellence (CoE).** The emerging Industrial Internet and IoT are stretching existing practices and technologies beyond traditional boundaries. Organizations need to address change management. In recent years, when firms have embarked on new technology or business initiatives, CoEs have proven to be helpful as internal committees that provide leadership, guide funding, identify training requirements, and share best practices. A CoE focused on IoT and the Industrial Internet can facilitate adaptation of systems and culture so the organization as a whole can successfully manage change. OT and IT should be part of the CoE and share in its leadership.

### NUMBER SIX

BALANCE GOVERNANCE AND SECURITY TO REALIZE VALUE FROM DATA

TDWI finds that governance and security are top-of-mind challenges for most organizations as they seek to democratize data access and incorporate new data types and sources into their environments. The advent of the Industrial Internet and expansion of IoT in operations only heightens the importance of creating comprehensive governance and security strategies that are aligned with changes to the organization's data environment.

As they deploy IoT in operations, organizations need to think through how to secure IoT devices themselves as well as the networks that connect them because these could be hacked and misused. IT will need to update network security procedures if IoT networks are to be connected to the organization's other networks. IT will need to ensure that security is up to date and that the software or firmware on the devices themselves is properly updated.

Regarding the data, however, once it is consolidated into centralized data management, security procedures become somewhat clearer. Organizations can set up perimeter security, database security, and security of the data in motion as it is ingested, collected, moved, and analyzed. Organizations need to pay special attention to securing vulnerabilities, such as when data passes from sources to the consolidated data management system over networks.

Governance intersects with data security due to regulations and internal policies about protecting personally identifiable information (PII) and other sensitive data. Governance rules and policies need to clarify how personnel are to adhere to regulations such as HIPAA, the European Union's GDPR, and many other data protection regulations within the U.S., Canada, China, and other regions.

In many organizations, the purview of governance has been broadening to include data quality and stewardship. Data stewards are typically data-savvy SMEs or IT personnel; their focus is on identifying how the organization can improve data assets through better quality, consistency, and completeness. Stewards can be mentors and guide other users to the best data for their visualizations and analytics. Stewards can also be part of governance by being knowledgeable about rules and policies and helping users adhere to them.

A key objective should be to balance enforcement of governance and security priorities with users' need to view, report on, and analyze the data. Unnecessarily strict governance and security will dampen the organization's ability to realize value from the data. Organizations should address finding the correct balance in CoE (or governance committee) meetings. The CoE should also head up evaluation of data management technologies and cloud services that can support the organization's requirements for security, governance, and stewardship of IoT and other operational data.

## A FINAL WORD

This Checklist explained six steps for generating value from operational IoT data. Organizations have the potential to use this data to gain new levels of understanding into how they can manage operations and run OT more efficiently and effectively. At the same time, these new data sources could provide new business management perspectives about what factors most impact the value of operations.

Organizations need to examine their data management strategy to ensure they have the right technology and/or cloud services to handle IoT and other operational data, especially to reap value from it as soon as possible through data visualization and analytics. Technology choices are key to supporting data-driven goals for informed decisions and actions.

## ABOUT OUR SPONSOR

www.osisoft.com

OSIsoft's PI System is one of the world's most widely used technologies for the Industrial Internet of Things. The PI System captures data from sensors, manufacturing equipment, and other devices and transforms it into rich, real-time insight that people use to reduce costs, dramatically improve overall productivity, and create new products and services.

Some of the world's largest organizations—including over 1,000 leading utilities, 90% of the largest oil and gas companies, and more than 65% of the *Fortune* 500 industrial companies—rely on the PI System for critical processes and decisions. The PI System can be found inside wind farms, advanced manufacturing facilities, and control rooms everywhere. OSIsoft collaborates with hardware and software vendors such as Rockwell Automation, SAP, Microsoft, Dell, HPE, and ESRI on products and technologies for bridging the IT-OT gap. Worldwide, the PI System manages over 2 billion sensor-based data streams.

## ABOUT THE AUTHOR

**David Stodder** is senior director of TDWI Research for business intelligence. He focuses on providing research-based insight and best practices for organizations implementing BI, analytics, performance management, data discovery, data visualization, and related technologies and methods. He is the author of TDWI Best Practices Reports and Checklist Reports on AI for BI, cloud analytics, data visualization, customer analytics, big data analytics, design thinking, and data governance. He has chaired TDWI conferences on visual BI and analytics, business agility, and analytics. Stodder has provided thought leadership on BI, information management, and IT management for over two decades. He is an industry analyst, having served as vice president and research director with Ventana Research, and he was the founding chief editor of *Intelligent Enterprise* and *Database Programming & Design*. You can reach him by email (dstodder@tdwi.org), on Twitter (@dbstodder), and on LinkedIn (linkedin.com/in/davidstodder)

## ABOUT TDWI RESEARCH

TDWI Research provides research and advice for BI professionals worldwide. TDWI Research focuses exclusively on analytics and data management issues and teams up with industry practitioners to deliver both broad and deep understanding of the business and technical issues surrounding the deployment of business intelligence and data management solutions. TDWI Research offers reports, commentary, and inquiry services via a worldwide membership program and provides custom research, benchmarking, and strategic planning services to user and vendor organizations.

## ABOUT TDWI CHECKLIST REPORTS

TDWI Checklist Reports provide an overview of success factors for a specific project in business intelligence, data warehousing, analytics, or a related data management discipline. Companies may use this overview to get organized before beginning a project or to identify goals and areas of improvement for current projects.