IDC ANALYZE THE FUTURE

IDC TECHNOLOGY SPOTLIGHT

Sponsored by: Seeq

Operational data is at the core of both continuous improvement and data-intensive transformations. There are unique challenges and opportunities for innovation represented by synthesis and analysis of time series and operational data.

Enabling Engineers with Data Science

June 2020

Written by: Kevin Prouty, Group Vice President, IDC Energy Insights and IDC Manufacturing Insights; Jonathan Lang, Research Manager, Worldwide IT/OT Convergence Strategies, IDC; and Reid Paquin, Research Director, Manufacturing IT Priorities and Strategies, IDC Manufacturing Insights

Introduction

Engineers and subject matter experts within operations settings are tasked with driving operational excellence to improve quality, safety, and throughput in production operations. The new technologies proliferating as part of Industry 4.0 initiatives are raising and expanding performance expectations. Rapid, data-driven insights are becoming critical to balance resiliency and agility with efficiency. As an IT professional, supporting these efforts is imperative as operations leaders grapple to get the most out of investments in technologies that generate and use large volumes of data.

AT A GLANCE

KEY TAKEAWAYS

- Many popular BI applications are well suited to visualization but not analysis of time series data.
- Operations personnel require self-service analytics applications to realize full value from time series and telemetry data.
- To accelerate the transition to a market-driven operation, enterprises must cleanse and merge operational and enterprise data in a scalable way.

Like the data itself, the insights data offers are transient in nature, streaming in real time as the operation executes its core functions. The insights hidden within data in motion form the foundation for the kind of resilient decisions that operations staff need to meet rising business expectations. Operational data is at the core of not just continuous improvement but also data-intensive transformations underway in all aspects of manufacturing. The ability to leverage both historical and streaming operational data is needed to gain insights required for the diagnostic, predictive, and prescriptive capabilities that are necessary in a digital world. The speed and scale of these data sets require a unique blend of synthesis and analysis of high volumes of time series and operational data, contextualized into action plans by subject matter experts within operations.

Many enterprise-side business intelligence (BI) applications are well suited for synthesizing IT-centric data from large and diverse static data sets. They are also good visualization applications. But they are not well suited to manufacturing analytics tasks because of the unique nature of time series data and the changing needs of operations to be agile and market driven. Many BI applications are not built for engineers who are immersed in the operational aspects of production assets and processes. They are built for analysts who are immersed in data and reporting. The time is now for operations workers to be empowered with self-service applications that deliver on their unique needs and help rapidly extract insights from existing and new time series data emerging from Industry 4.0 initiatives. Additionally, IDC predicts that by 2021, 80% of all industrial companies will have merged operational data streams with enterprise data streams to

support broader and more rapid operational innovation.[1] The first step to achieving these higher-level enterprise insights is to get your data house in order or make the right information available in the right format; operational time series data is a critical component of these efforts.

The Challenge with Time Series Data

What is time series data? In manufacturing, time series data is a collection of measurements captured chronologically from the physical world via sensors. It is the fuel that drives strategic use cases possible because of the industrial Internet of Things (IIoT). For process-centric, asset-intensive environments, unlocking value from operational data has historically been a challenge. Volumes are immense and heterogeneous, and most importantly, operational data lacks context. Complicating matters, data from operational systems is now exploding in volume, largely driven by the increase in time series data or sensor telemetry stemming from the rise in IoT, edge computing, and connected assets (see Table 1).

TABLE 1: ***Data Generated in Operations by Industry, 2019–2025 (TB/Day)***

	2019	2020	2021	2022	2023	2024	2025
Semiconductor	1.7	2.2	2.9	4.0	5.6	8.1	13.5
Medical devices	2.0	2.3	2.6	3.1	3.8	4.6	6.2
Automotive OEM	1.9	2.1	2.5	3.0	3.6	4.4	5.8
Electric assembly	1.3	1.5	1.8	2.2	2.7	3.5	5.0
Petrochemical	1.3	1.5	1.8	2.2	2.7	3.5	5.0
Chemicals	1.3	1.5	1.8	2.2	2.7	3.5	5.0
Power	1.1	1.3	1.6	2.0	2.5	3.2	4.5
Paper and textiles	1.1	1.3	1.5	1.7	2.1	2.6	3.4
Oil and gas	1.0	1.1	1.3	1.6	1.9	2.3	3.1
Automotive suppliers	1.2	1.3	1.5	1.7	1.9	2.3	2.8
Aerospace	0.9	1.0	1.2	1.4	1.7	2.1	2.7
Metals	1.0	1.1	1.2	1.4	1.6	1.9	2.3
Pharmaceuticals	1.3	1.4	1.5	1.6	1.8	2.0	2.3
Heavy industry	0.8	0.8	0.9	1.1	1.2	1.4	1.8
Food and beverage	0.8	0.9	1.0	1.1	1.2	1.3	1.5

[1] *IDC Futurescape: Worldwide IT/OT Convergence 2020 Predictions,* IDC #US45597519, October 2019

TABLE 1: ***Data Generated in Operations by Industry, 2019–2025 (TB/Day)***

	2019	2020	2021	2022	2023	2024	2025
CPG	0.7	0.8	0.9	0.9	1.0	1.1	1.3
Gas	0.5	0.6	0.6	0.7	0.7	0.8	1.0
Water	0.3	0.3	0.4	0.4	0.4	0.5	0.6
Light industrial	0.3	0.3	0.4	0.4	0.4	0.5	0.6

Source: IDC, 2019

To make sure that one aspect of the accelerating data management issue is called out plain and clear, we'll mention process data historians. These are databases and systems built to connect to DCS, SCADA, and HMI applications that support high-frequency data collection, highly efficient data storage and compression, and rapid data retrieval and are accessible by engineering staff. These data historians are developed by many companies, including industrial automation vendors and some independent software companies. While historians have been a critical data store for over 30 years, manufacturing companies and their engineers continue to install more historians, and there is a renewed focus on getting more value out of the data in those databases.

Traditional BI applications have typically been limited in how they access and manipulate the data in those historians. The reason process data historians exist is to collect time series data.

This operational time series data is only as valuable as the ability to interpret it in context and turn around decisions quickly. This is nothing new — this analysis has been used to the same end for 30 years. However, the availability of data science innovation in self-service analytics for engineers and subject matter experts, the volumes of IIoT data, and the increased pressure to deliver data-driven insights are all new. Virtually all major Industry 4.0 use cases rely on some form of data-driven decision making including the following:

- **Descriptive.** What is happening or has happened? Industry 4.0 is creating pressure for operations to move from retrospective failure and performance analysis to near-real-time monitoring and the ability to report on past-happenings on demand. In addition to fulfilling regulatory requirements, the business expects new key performance indicators (KPIs) and performance dashboards to be available on demand.
- **Diagnostic.** Why did it happen? Understanding the root cause of performance or nonperformance is the key to driving improvements, whether in planning, production, or quality. Institutional knowledge and context are key, and these investigations rely on the ability of engineers to manipulate and contextualize operational performance data in ways that are not yet — and may never be — codified into algorithms suitable for automated intelligence. This human expertise is what separates operational laggards from leaders. A good reason to undertake a diagnostic investigation is to understand why a production deadline was missed.

- **Predictive.** What will happen in the future? To reduce unplanned downtime, improve productivity, and ultimately become a market-driven operation requires foresight into a variety of future scenarios. Operations staff need to understand what challenges or bottlenecks will arise so that they can deploy resources efficiently and proactively and plan for change that is being pushed down from higher levels within the enterprise at increasing rates. The greatest example of this is predicting asset failure to shift from unplanned to planned downtime maintenance models.
- **Prescriptive.** What decision will produce the greatest outcomes? Confident modeling of future scenarios is the holy grail of the market-driven operation. The need for operational agility and speed may result in the development of sophisticated digital twins of both assets and processes in the future, but operations must meet rising reliability and agility expectations today. Prescriptive analysis is necessary to optimize the use of resources or assets, for example.

Furthermore (beyond the plant, in an IT or corporate or business context), time series data is difficult to reconcile and interpret in the broader business (ERP systems, HR systems, market strategy) context necessary for these use cases to succeed and scale. Leveraging time series data to augment and drive deeper context and meaning for traditional BI applications is the path to innovating toward real business goals.

For 30 years, spreadsheets have proven to be the only game in town when it comes to synthesizing time series data to drive the types of insights mentioned previously. Much to the chagrin of operations engineers, business analysts, and subject matter experts, the unique requirements to interpret time series data have gone unmet for decades, and manual development of spreadsheet macros remains the only choice, until recently, across the majority of manufacturers today. It is evident that these methods are risky and unscalable, particularly when placed in the context of the rise in time series types of data and the need for more resilient decision making.

Time series data is particularly difficult to synthesize and interpret early in its life cycle. Traditional spreadsheets remain the default tool to prepare time series data for analysis or use with enterprise BI applications for a number of reasons:

- Transforming data into insights requires subject matter expertise to contextualize data for a given question. The value of those insights is dependent on timeliness. Therefore, engineers utilize spreadsheets because they are one of the only tools that satisfy both of these requirements.
- Operational data exists across a variety of applications. To gain the full set of data related to a particular event or moment in time, one must correlate data across multiple databases and applications such as MES and historians. Data formats across databases may be highly inconsistent, requiring manual adjustment in order to compare data in one system to that of another.
- Units of measurement are inconsistent. Normalizing for different time zones, different units of measure, and so forth is a manual process. Interpolation of time series data is risky without the necessary context as to what measure of data is being analyzed.
- Inconsistencies in collection rates and data tagging across like assets in different settings or even within the same setting means human judgment is often required to gain "like for like" understanding of multiple streams of data required for an analysis.

- Manual processes persist in addressing the sheer volume of data and intermittent network connectivity in some manufacturing settings, so noisy signals or gaps in data must be manually identified and cleansed to avoid skewing analysis.
- Useful reference frames for time series data can vary significantly. Setting boundaries or isolating trends within fixed time periods is a critical capability for analyzing operational data. Flexibility is required to adjust these parameters even within a single analysis.
- Outside perspective is often needed. In addition to multiple data sources for the time series data, oftentimes an operations analysis or report will require outside context from one-off reports or from applications such as batch and MES to enable subject matter experts to put context to the signals they are interpreting.

Indeed, engineers have their work cut out for them and are often and increasingly required to analyze this data rapidly. In reality, the hours engineers spend in spreadsheets is detrimental to productivity because it requires even the most skilled engineers to spend too much time. The keys to alleviating the current reliance on spreadsheets are experience, institutional knowledge, and self-build sheets that have limited potential for reuse and require an individual worker's insight to utilize effectively. It is high time to codify and standardize the way workers engage with this data if enterprise expectations of improved throughput, quality, and other plant metrics are to be met.

Outside the operations setting, business analysts require cleansed, contextualized, and dimensionalized data to deliver business intelligence on operations data, which requires days or weeks to report on common inquiries such as comparing the energy demands to produce similar products. While BI applications such as Power BI and Spotfire have cut cycles from the visualization and presentation of structured enterprise data from ERP and other systems, their sweet spot is in data visualization across relational data sets. Time series data has become the bottleneck to resilient decision making even at the enterprise level.

Time series data has become the bottleneck to resilient decision making even at the enterprise level.

At the same time, this fully contextualized, higher order analysis, which combines enterprise, customer, and operations data, is the key to unlocking the tremendous opportunities that Industry 4.0 promises. Engineers and industrial executives are being asked to be more resilient, agile, and strategic in their decision making. The industry at large is moving toward being market driven. When enterprise and operations executives can remain in near-real-time alignment to make data-driven decisions, they can predict and prescribe operational changes that enable them to meet customer demand more quickly and efficiently, with confidence. This represents a potent dynamic to achieve. To enable this dynamic requires IT to empower the organization with the data, communication, and infrastructure in support of decision making that is continuous, consistent, and accessible to every role and segment of the organization.

Benefits

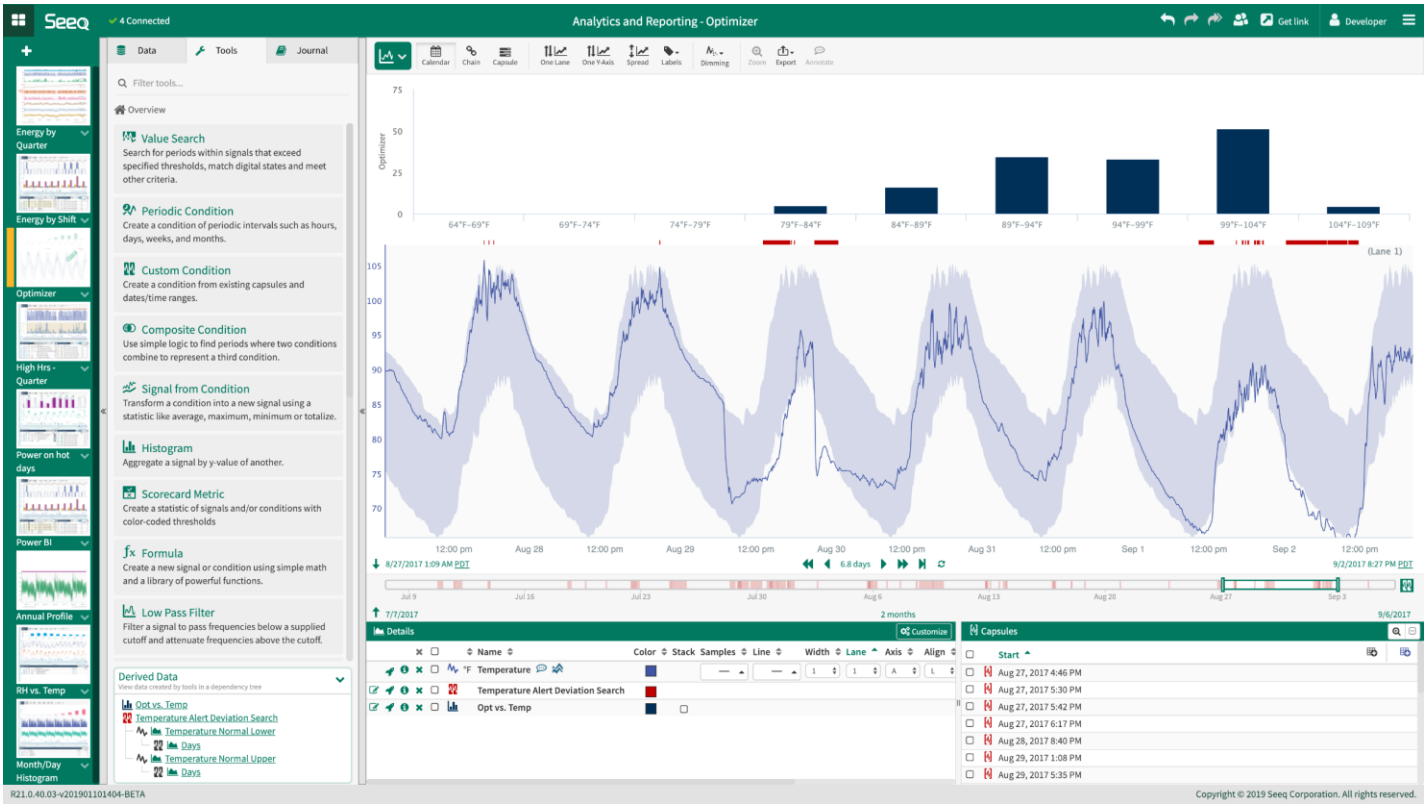
Self-service applications for engineers that are purpose built for connecting to, cleansing, and analyzing time series data are emerging and are alleviating the bottleneck that data imposes, enhancing data value and driving benefits for the workforce. Such applications:

- Enable IIoT initiatives to scale beyond a few assets by shortening the cycle of achieving value. Models can be reused across like assets, a common pain point in scaling IIoT projects and achieving cross-site visibility into performance.
- Support flexibility for engineers to delve into novel data analysis across multiple data sets quickly and reliably, without demanding IT resources. With the rate of change occurring in a market-driven operation, those IT resource demands will otherwise be significant.
- Encourage the development of skills to analyze time series data. Without the development of these skills, which are in short supply, the lack of time series analysis capabilities is likely to become more acute because experienced engineers are retiring from of the workforce. Data scientists are in high demand and difficult to recruit into operations roles. By standardizing the data investigation process and providing self-service applications, organizations can enable every engineer to achieve the level of data science competency required to meet today's demands.
- Free up time for engineers and business analysts to focus on more strategic analysis and process improvements rather than the tactical work of data cleansing and formatting.
- Create consistency in the way data is interpreted and reported across multiple site locations, enabling the sharing and reuse of models and best practices as well as enabling cross-site visibility.
- Enrich traditional BI applications for the enterprise more easily with time series data and vice versa, providing consistency, context, and depth of insight to engineers and business leaders alike.
- Maintain and manage analytical and artificial intelligence (AI) models as well as track simple assumptions made in modeling and decision making.

Considering Seeq

Seeq is a provider of advanced analytics applications for the asset-intensive process manufacturing, energy, and utilities industries. Seeq's offerings focus on creating value from both historical time series data and streaming data generated by increasingly digitized assets. To enable engineers to find insights in this data, Seeq aims to make the disparate and heterogeneous data streams available to subject matter experts who can ask the right questions to act on the data. This data is accessed from process historians, SQL databases, IIoT clouds, data lakes, and other databases. Seeq's goal is to provide an environment that helps customers answer questions across the spectrum of operational scenarios from a descriptive and diagnostic understanding of the past through real-time monitoring to predictive and prescriptive scenarios that optimize future outcomes (see Figure 1).

FIGURE 1: ***Seeq Is a Self-Service Advanced Analytics Application for Process Manufacturing Data***



Source: Seeq, 2020

To enable users to find insights, Seeq delivers unified and self-service applications for analysis across various data silos — with features for data cleansing, performing calculations within time-based parameters (what Seeq calls "capsules"), and enabling contextualization from other operations systems such as MES and EAM. Seeq browser-based application access varies by audience and is enabled by Seeq Server, which can be installed on premises or in the cloud. The company's modules are geared for the following constituents:

- For engineers and their peers, Seeq Workbench enables self-service data modeling and analysis capabilities to answer questions about asset and operational performance. By allowing engineers to directly manipulate the parameters of analysis to account for novel scenarios, Seeq aims to empower engineers to take action quickly and optimize operational processes proactively from the point of activity. This self-service approach also aims to alleviate demand for IT to continuously update and manage analytics models.
- For engineering leads and operations managers, Seeq Organizer unifies multiple analyses to provide trending dashboards and scorecards that give an overview of operational performance in real time. Seeq Organizer also provides a single source of truth for employees across the organization to leverage insights created in Seeq. The intent is to increase operational reporting as well as collaboration with capabilities to annotate and share feedback.

- For data-savvy engineers and data scientists, Seeq Data Lab offers additional flexibility to leverage Python libraries and integrate with a broader variety of applications. The Jupyter Notebooks–based front end allows data scientists to participate in industrial analytics by publishing custom algorithms for use by engineering staff, increasing the collaboration with operations staff to develop new and enhanced models.

At the core of the Seeq Server is the Seeq Calculation Engine, which utilizes domain-specific principles developed especially for time series data. By accessing data at the source, identifying salient time periods for like comparison, and matching the time "capsules" across heterogeneous data stores, the Seeq Calculation Engine sets itself up to be capable of analyzing large data sets at scale.

Additionally, the preparation and analysis of time series data enable the data to be more useful and accessible to business analysts who require it to be integrated in reporting conducted in traditional BI applications. The usability and accessibility open a broader audience, including business executives, to unlock the potential value of time series data. Analysis also enables insight to be discovered when this data is contextualized against enterprise and operations applications.

Challenges

While the concept of role-specific data and analytical workbenches is appealing, there are a number challenges and inhibitors to realizing the full value of this approach:

- Cultural issues of defaulting to spreadsheets are probably the biggest challenge. Every engineer and technical staffer quickly becomes an expert in the nuances of manipulating data and deriving insights from spreadsheets.
- The next biggest challenge is the broader data management issue. Isolated projects with a limited number of engineers, assets, and data sources tend to give a false impression of the effort needed to get the datasphere of an organization ready for full system production.
- A less commonly discussed issue, yet a major pain point for all analytical and AI projects, is paralysis through pilots — that is, designing a pilot in a narrow silo and failing to take into consideration how the pilot represents the broader enterprise world.
- The coordination of IT and operations is often overlooked: data and expertise from the engineers and algorithms from the data scientists. Operations' technical capabilities tend to focus on process optimization, while IT tends to focus on enterprise reporting and infrastructure. Building the governance model for technical interaction between IT and operations is critical.

All of these challenges are capable of stopping a project, but they aren't insurmountable. They require careful planning and an intimate understanding of each of the organizational silos involved and their motivations to invest in these projects.

Conclusion

IDC believes time series data not only is critical to the stability and reliability of ongoing operations but also is the key to unlocking transformational use cases in operations settings. To be able to manage and extract this value, enterprises require advanced analytics applications that can analyze time series data in the broader context of both plant data such as quality, safety, and yield and business data such as pricing and orders. Core guidance to enterprises looking to scale

and succeed in their transformation is to enable access to data by those who have expertise and experience. While IIoT platforms can aggregate raw signals and BI applications can access and present business data, it is necessary to bridge the technology gap that exists between the two to realize the full promise of IIoT.

About the Analysts

Kevin Prouty, *Group Vice President, IDC Energy Insights and IDC Manufacturing Insights*

Kevin Prouty is Group Vice President for IDC Energy Insights and IDC Manufacturing Insights. He is responsible for managing a group of analysts that provide research-based advisory and consulting services that will enable energy executives in oil and gas and utilities to maximize the business value of their technology investments and minimize technology risk through accurate planning. Kevin's research specialties are utilities, manufacturing, enterprise applications, and product innovation.

Jonathan Lang, *Research Manager, Worldwide IT/OT Convergence Strategies*

Jonathan Lang is Research Manager for IDC Manufacturing Insights responsible for the Worldwide IT/OT Convergence Strategies practice. Mr. Lang's research focuses on digital transformation strategies in environments where operations technologies are deployed, including manufacturing, utilities, oil and gas, and healthcare provider settings. As IT capabilities redefine and extend the core value drivers of operations technologies, Mr. Lang's research examines strategies, road maps, and governance models to drive this convergence and manage the new data and processes it requires.

Reid Paquin, *Research Director, Manufacturing IT Priorities and Strategies*

Reid Paquin is Research Director for IDC Manufacturing Insights responsible for the IT Priorities and Strategies (ITP&S) practice. Mr. Paquin's core research coverage includes IT investments made across the manufacturing industry and manufacturers' progress with digital transformation. Based on his background covering the manufacturing space, Mr. Paquin's research also includes an emphasis on the technology enablers that help manufacturing executives make better-informed operational decisions.

MESSAGE FROM THE SPONSOR

Seeq is an advanced analytics solution for process manufacturing organizations to rapidly make data-based decisions that improve production and business outcomes. Supported data sources include historians, IIoT platforms, and cloud services as well as contextual data in manufacturing and business systems. With diagnostic, monitoring, and predictive analytics, Seeq enables organizations to derive more value from the data they collect. Typical users of Seeq include process engineers for analytics, plant executives for dashboards and KPIs, managers for reporting, and data scientists developing machine learning algorithms and accessing production data sets.

Oil & gas, pharmaceutical, specialty chemical, utilities, renewable energy and numerous other vertical industries rely on Seeq to improve production outcomes, including yield, margins, quality, and safety. Headquartered in Seattle, Washington, Seeq was founded in 2013 and is a privately held virtual company with employees and partners in the United States, Asia, Canada, Europe, and South America.

IDC Custom Solutions

The content in this paper was adapted from existing IDC research published on www.idc.com.

IDC Research, Inc.
5 Speen Street
Framingham, MA 01701, USA
T 508.872.8200
F 508.935.4015
Twitter @IDC
idc-insights-community.com
www.idc.com