With new data modeling and advanced analytics methods available, every industrial operation strives to become more data driven. Yet, to succeed at scale, organizations must adopt new processes for data integration and management that account for the complexity of operational data.

Optimizing Industrial Operational Data

July 2023

Written by: Jonathan Lang, Research Director, Worldwide IT/OT Convergence Strategies

Introduction

As the sophistication and capabilities of technology expand, companies of all types are seeking to become more data driven across the organization. This is particularly true in industrial operations functions in industries such as manufacturing, oil and gas, and utilities. These operations settings are highly and increasingly instrumented and connected, and the ruthless pursuit of operational efficiency and cost and waste reduction is tantamount. For these teams, data has always been integral to decision making, but the volumes, methods, and reach of its analysis continue to reach new thresholds.

In IDC's discussions with these industrial enterprises, becoming a data-driven operation has become the north star of digital transformation efforts. The user base and accessibility to operational data is expanding thanks to technologies like the Industrial Internet of Things (IIoT), edge and cloud computing, and advanced analytics and artificial intelligence (AI). The goal of this data-driven approach is to achieve two seemingly conflicting goals at once – to become leaner and more efficient while also gaining the flexibility and agility necessary to serve volatile and rapidly shifting markets. Atop the pressures of these volatile markets is an increasing social and economic pressure to reduce the nonmonetary costs of consumption through sustainability-focused business decisions. Indeed, new and existing lines of business have never demanded more visibility and access to operational performance data.

As companies pursue these endeavors, it has become abundantly clear to IT organizations tasked with architecting and standing up these new data pipelines that approaches to integrating and managing data that have been successful in back-office and enterprise-side functions are ineffective when working within operational technology (OT). Operational data and the technology landscape it is sourced from is notoriously heterogeneous, isolated, high speed, voluminous, and ungoverned. The approach many organizations have attempted to apply in OT use case pilot programs of manually cleansing and integrating data via one-off efforts does not scale and is significantly inhibiting the potential value of digital transformation initiatives.

AT A GLANCE

KEY TAKEAWAYS

» Industrial organizations are seeking to become more data driven and capitalize on new AI and analytics capabilities.

» To do this requires a stable data foundation and data operations and engineering capabilities to abstract OT data effectively.

» Companies seeking to accelerate the successful development of such a data foundation should consider tools that are industry and purpose built for the unique requirements of operations.

Weaving the Industrial Data Fabric

A confluence of technologies is enabling access to operational data — but does not offer integration or contextualization of the data. IIoT, edge computing, and cloud computing have opened the flood gates of digitizing operational assets and processes. Many organizations report having explosive growth in operational data but significant challenges in putting it to good use. Abstracting and contextualizing operational data has become the major bottleneck to scale out of virtually all value-rich use cases. According to IDC's 2022 *Worldwide IT and OT Convergence Survey* of over 1,000 IT and operations professionals, 63.1% of respondents expect operational data to grow in excess of 15% in terms of terabytes per day in just 12 months. And in IDC's 2022 *Future of Operations Survey* of a similar cohort, 53.1% of respondents stated that they have more than enough data to support their key initiatives, but that the data is trapped in silos or difficult to extract value from. This is why the majority of industrial enterprises are actively strategizing and working to develop a data abstraction and operations layer now (see in Figure 1).

FIGURE 1: *The State of Operational Data Management and Analytics Strategies*

Q *Which best describes the state of maturity of your operational data management and analytics/artificial intelligence strategy?*
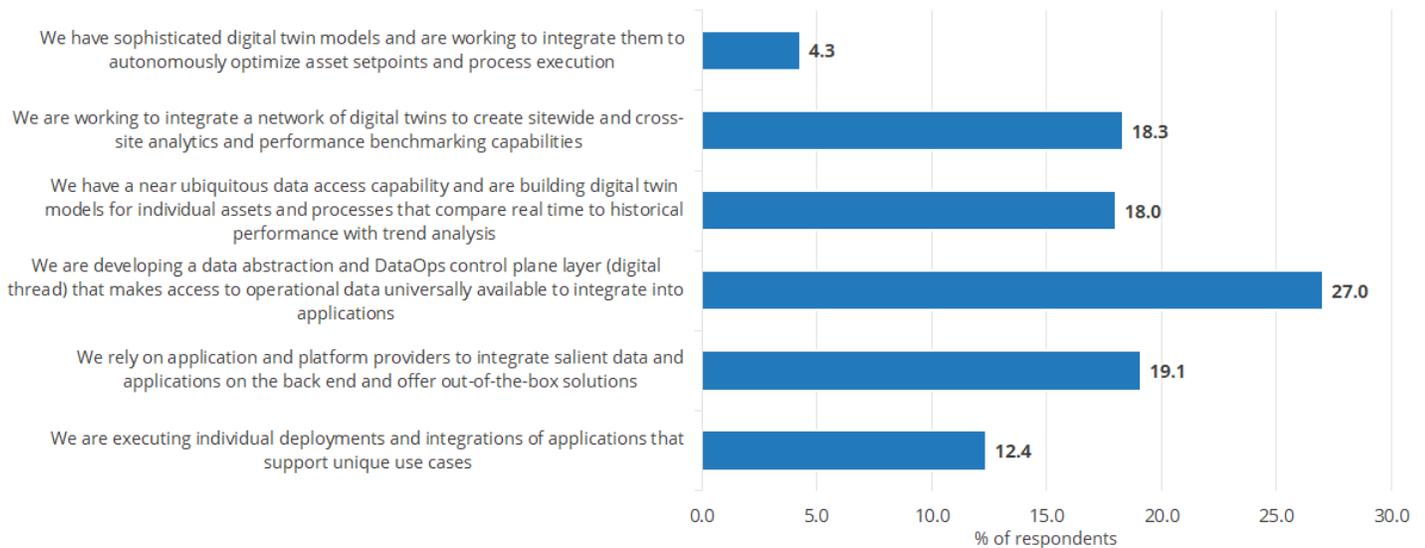
n = 1,028

Source: IDC's Worldwide IT and OT Convergence Survey, July 2022

What is observed at mature digital industrial enterprises, and what is needed by all, is a unified set of technology capabilities to enable a data abstraction layer. This abstraction layer must meet OT data where it is, regardless of data quality, legacy software and database formats, and lack of context. This market need has given rise to a new category of industrial data operations software tools that are purpose built for these unique challenges.

The data engineering efforts that take place at this data abstraction layer must be collaborative between IT and operations — but ultimately will set the conditions for scalable use of contextualized and actionable data. These data abstraction capabilities carry multiple requirements:

» Connect to diverse, high-volume structured and unstructured data utilizing a variety of protocols and data formats. Time series data stemming from data historians, IIoT platforms, and directly from dozens of industrial equipment protocols are important, but SQL databases and others also supply critical information that must be integrated.

» Abstract or virtualize the data without removing it from the source location. IDC refers to this approach as building a digital thread, which contrasts a failed approach many organizations have attempted of simply duplicating data into a data lake with the intent of untangling its meaning at a later time. In this capability, the structure of the source data must be retagged with consistent metadata attributes such that it can be analyzed effectively across multiple sources.

» Use role and access management to govern the data usage as it becomes available to new and existing roles. This data governance ensures that sensitive information is not exposed unnecessarily or edited in ways that could introduce business risk. It is also necessary to ensure that users of the data can find what is relevant to them without becoming overwhelmed.

» Build data pipelines that deliver data to a variety of destinations in support of different use cases. The key capability is to support this effort in a scalable, repeatable, templatized, and user-friendly manner. By contrast, many organizations have built custom one-off integrations to support proof-of-concept initiatives, only to realize they have built a house of cards that does not scale and requires maintenance every time there is a slight change in sources or destinations.

Many organizations may have already discovered that this independent data layer is the key to scale without building technical debt in the upkeep and customization of data integrations, but there are best practices IDC has collected from discussions with IT teams largely responsible for the technical end of these efforts:

» IT tools will not gain adoption from OT subject matter experts who hold the knowledge necessary to contextualize and upkeep the data. The environments where OT staff are asked to contribute to this new, additive responsibility to their normal day-to-day tasks must be purpose built for the industry and come from technology providers that the operations staff view as credible to be adopted sufficiently.

» Approaches must leverage industry and use case–oriented frameworks to accelerate the data ingestion process. Many companies report that the use of open-ended toolsets requires significant data science skills that they do not have in-house to use effectively.

» Approaches and tools must meet security standards and pricing/licensing models and carry change management capabilities that fit with the volumes and needs of OT data.

» Companies advise their peers to start small and get to value quickly, which in turn will ultimately accelerate investment willingness from executives. This requires subject matter expertise on the provider side to assist with rapid implementation and first use case buildout.

Benefits

The primary benefit of utilizing such an approach to operational data abstraction and management is to develop a repeatable and scalable approach to supporting OT data use case requirements that does not require massive services budgets or stacked talent pools of data scientists and engineers. However, there are additional benefits that organizations can expect:

» Traceability of data across a product, process, or site footprint for both sustainability initiatives and for satisfying the requirements of already highly regulated industries that are coming under even more regulation

» Greater collaboration between OT and IT, relieving some of the burden on either group in transformation projects and building business relationships that will extend into other areas of digital innovation and solution development

» Positioning OT data to be available for future novel analytics and use case efforts such as generative AI, next-generation automation capabilities that will be more dynamically data driven, and others yet to be revealed

Technology or Vendor Profile

Aspen Technology Inc. (AspenTech) is a global software provider that aims to help industries at the forefront of the world's dual challenge meet the increasing demand for resources from a rapidly growing population in a profitable and sustainable manner. AspenTech solutions seek to address complex environments where it is critical to optimize the asset design, operation, and maintenance life cycle (see Figure 2).

AspenTech DataWorks, a business unit of Aspen Technology, seeks to be a global leader in industrial data management solutions. The company's mission is to accelerate data-driven value creation in the asset-intensive industries through robust data software offerings. The primary objective of AspenTech DataWorks is to unlock key industrial use cases by unifying companies' industrial data architectures. From these unified environments, organizations can leverage DataWorks to aggregate and contextualize operational data across disparate sources.

FIGURE 2: *AspenTech Inmation*
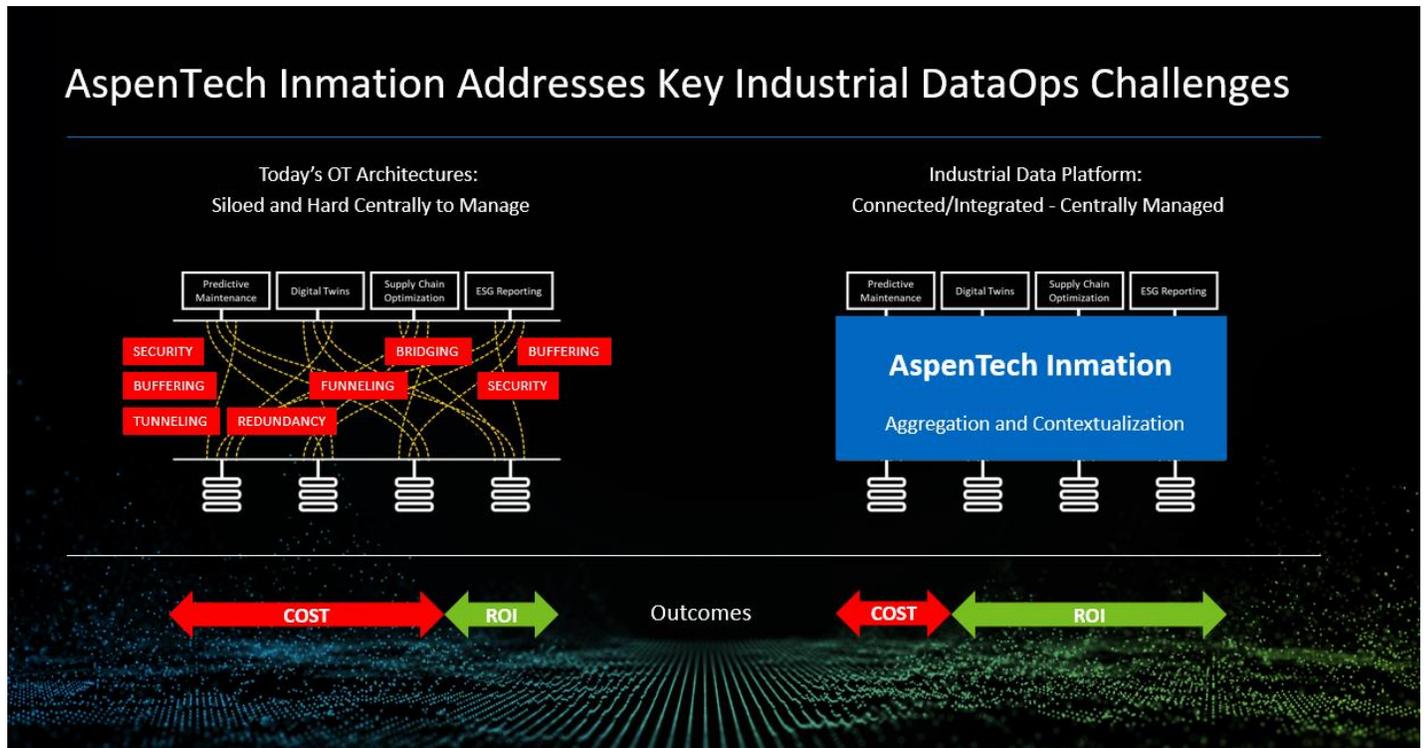
Source: Aspen Technology, Inc., 2023

Challenges

Some companies prefer these data abstraction capabilities be sourced from vendors that focus exclusively on this offering as a way to ensure they are fully agnostic to all of the systems and data in their current technology landscape. AspenTech will have to demonstrate to these companies that they are capable of supporting competitors' systems and data in addition to their own. Necessary data integrations are growing beyond time series data to include other OT and IT system data. AspenTech will have to continuously expand capabilities to support a growing data landscape in terms of both sources and destinations.

The emerging industrial data operations software market is growing explosively. Companies are being bombarded with similar data operations tools. AspenTech will have to raise awareness of its offering and educate buyers on areas of differentiation to gain meaningful traction.

Conclusion

Companies are looking to do more with what they view as underutilized and value-rich OT data. What they have found through trial and error is that point-to-point integrations do not scale and are brittle and labor intensive to maintain. With digital twins, AI, next-generation automation, and so many exciting new possibilities on the horizon, companies must have a stable data foundation in place to be able to capitalize on innovation rapidly. Developing a data abstraction layer is key to the scalable use of OT data to innovate in new use cases, and purpose-built tools offered by companies with industry and domain knowledge are critical to the success of such a technology offering.

> With digital twins, AI, next-generation automation, and so many exciting new possibilities on the horizon, companies must have a stable data foundation in place to be able to capitalize on innovation rapidly.

About the Analyst

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

Jonathan Lang is Research Director for IDC Industry Operations Insights responsible for the 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.

MESSAGE FROM THE SPONSOR

The AspenTech DataWorks Solutions

AspenTech Inmation™: Centralized data management system that allows for the ingestion, processing, augmentation and storage of diverse data, empowering your team from the plant floor to the executive suite.

Aspen InfoPlus.21®: Process Historian that helps you collect, merge, store and retrieve large volumes of time-series data and make it accessible for analysis and reporting by client users, business systems and production applications.

For more information, go to: www.aspentech.com/dataworks

IDC Custom Solutions

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