IDC TECHNOLOGY SPOTLIGHT

INDUSTRIAL DATAOPS: IOT OPTIMIZED DATA EMPOWERING ENGINEERS TO OPERATIONALIZE DATA AT SCALE

Authors:

Gaurav Verma
Lorenzo Veronesi
Jean-François Segalotto

March 2021

Sponsored by COGNITE

IDC #EUR147278921

Industrial DataOps: IoT Optimized Data Empowering Engineers to Operationalize Data at Scale

Industrial Data Challenges Make Business Decisions and App Creation Challenging

Intelligent use of the wealth of data that operational technology (OT) systems produce is central to industrial organizations' efforts to improve operational excellence.

OT data is the raw material that enables organizations to build more efficient and resilient operations and improve employee productivity and customer satisfaction. Industrial organizations, however, struggle to generate value from their increasingly connected operations — with IDC research showing that only one in four organizations analyzes and extracts value from data to a significant extent.

Data dispersion and lack of appropriate tools and processes are the biggest obstacles (source: *IDC Enterprise Data Retention Survey 2019*), resulting in data workers spending almost 90% of their time searching, preparing, and governing data (*IDC Data Integration and Integrity End-User Survey 2019*). Fear of missing data value often led organizations to prioritize data centralization over data organization. In turn, this led to poorly thought-out "data swamps" that only perpetuate the issue of dark and uncontextualized data. Companies that adopted machine learning (ML) to develop predictive algorithms quickly realized how critical it is to have trusted data quality and that historical data is not always trustworthy. Many organizations are unable to address the requirements needed to achieve the data governance required to support data-driven innovation.

AT A GLANCE

KEY TAKEAWAYS

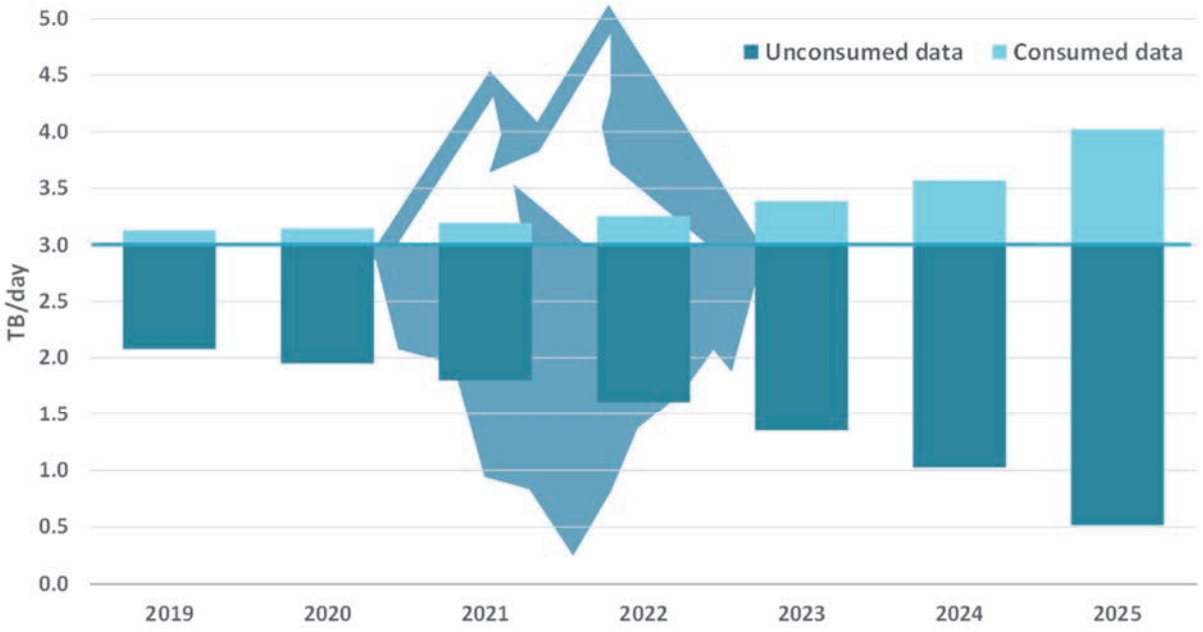
- **Industrial data challenges:** Only one in four organizations extracts value from data to a significant extent. Data dispersion and the lack of tools and processes to connect, contextualize, and govern data stand in the way of truly digital operations.
- **DataOps opportunity:** Industrial DataOps promises to improve time to value, quality, predictability, and scale of the operational data analytics life cycle. It's also a steppingstone to a new way of managing data within the wider organization, enabling it to cope with growing data diversity and serve a growing population of data users.
- **From theory to practice with industrial DataOps technology:** Establishing DataOps is a gradual process that requires greater cohesion among data stakeholders. Many companies struggle to overcome organizational inertia, and deploying the right platform — feature-rich, intuitive, and easily scalable — can become a catalyst for change.

The reality is that as operational assets become more complex, connected, and intelligent — and provide more real-time information — the complexity of enabling data-driven decision making to plan, operate, and maintain them increases. To put this in perspective, organizations across manufacturing, oil and gas (O&G), utilities, and mining expect their daily operational data throughput to grow 16% in the next 12 months (*IDC IT-OT Convergence Survey 2020*). IDC has measured the data generated daily by operations across these organizations' silos and has modeled the future expansion of data and its use across industrial sectors. Even accounting for

the growing digitization of operations, IDC predicts that only about 30% of this data will be adequately utilized in 2025 (see Figure 1).

FIGURE 1

Data Generation and Consumption in a $250 Million Industrial Operation, 2019–2025



Source: IDC, 2020

Modern Industrial DataOps

IDC defines DataOps as a methodology for industrializing data management and the data analytics value chain. It applies automation, agile methods, and DevOps practices to the data life cycle, improving time to value, quality, predictability, and scale of data analytics.

"By 2023, 60% of organizations will have begun implementing DataOps programs to reduce the number of data and analytics errors by 80%, increasing trust in analytic outcomes and efficiency of Gen-D workers."

(IDC FutureScape: Worldwide Data and Analytics 2021 Predictions)

Industrial DataOps platforms help data workers deploy automated workflows to extract, ingest, and integrate data from industrial data sources, including legacy operations equipment and technology. They offer a workbench for data quality, transformation, and enrichment, in addition to intelligent tools that apply industry knowledge, hierarchies, and interdependencies to contextualize and model data. This contextualized data is then made available through specific application services for humans, machines, and systems to leverage.

Direct and Indirect Benefits of DataOps

Efficient Data Management

DataOps maximizes the productive time of data workers with automated data provisioning, management tools, and analytic workspaces to work with and use data safely and independently within specified governance boundaries. DataOps is augmented with AI-based automation for various aspects of data management (e.g., metadata management, unstructured data management, data integration) — enabling data workers to spend more time developing use cases that deliver value.

Improved Data Accessibility and Contextualization

The current use of operational data is limited by dispersion across silos, spotty integration, and accessibility for centralized applications. Even where data sources are connected, data often lacks context due to limited documentation at the data's origin or information loss due to inconsistent structure or tagging.

By improving data accessibility, DataOps brings a paradigm shift in how the organization accesses business-critical information, improving decision-making quality and lowering risk while lowering the barriers to (and skills for) data innovation.

DataOps technology uses AI to enable rapid ingestion and contextualization of large amounts of data. A semantic model is created where domain context can be added to data types and sources by adding a contextualization process. This enables users to understand connections in terms of lineage, interrelation, and asset hierarchy.

Converging advanced AI tools and models and data platforms is critical to rapidly analyze complex data streams from disparate sources. This does not only reduce the lead time but also the total cost of data operations, including data testing, preparation, and quality management, as well as any data source modifications.

Rapid Development and Scaling of Industrial Use Cases

DataOps aims to shorten the time to value of data by making proofs of concept (PoCs) quicker and cheaper to design and offer tools to operationalize and scale. Industrial DataOps platforms provide models and frameworks to better analyze operational assets or processes. Anomaly detection in asset performance, equipment failure prediction, or root cause analysis are a few examples. These models are the foundation for the operationally focused analytics tools that differentiate them from more general business intelligence tools.

IT-OT-Engineering Technology (ET) Convergence

Traditional data silos impede the extraction of value from data. Companies can make the most industrial DataOps to begin the integration process that spans asset and data life cycles across ET, OT, and IT. The resulting converged data will support resilient decision making across the organization and unlock the potential of full-fledged digital twin applications.

Enterprise Data Governance as a By-Product

DataOps enables companies to set and enforce the basic principles for managing data. If implemented successfully, DataOps provides consistency and ROI in technology, processes, and organizational structures to improve operations data quality, integration and accessibility, and stewardship. DataOps platforms also enhance data security, privacy, and compliance with tracking, auditing, masking, and sanitation tools.

Moving from Theory to Practice

DataOps requires organizations to take some critical steps to start the journey on the right foot. For starters, it requires stronger cohesion among data stakeholders. Data science and IT must collaborate well beyond data access and resource allocation, while businesses should be involved in data projects well beyond the typical demand and validation stages.

Organizational divides add to a company's inability to access data at scale to make asset analytics pilots too long or too expensive to operationalize. Bridging these organizational and operational gaps is a balancing act that requires focus and leadership. Deploying the right tools — feature-rich, intuitive, and easily scalable — can be a catalyst for change.

Considering Cognite

Cognite is an international industrial DataOps technology developer. Its core software product is Cognite Data Fusion (CDF), a full-stack DataOps platform focusing on unifying and communicating data across IT, OT, and ET disciplines and the entire operational asset life cycle. Built fully on public-cloud-managed services, CDF can be consumed as SaaS or embedded as middleware into larger in-house data platforms as an API PaaS.

Cognite aims to support asset-intensive industries in developing and scaling mission-critical use cases across safety, efficiency, and sustainability. The company brings value to its clients by making data available, useful, and valuable.

Making Data Available

CDF offers a comprehensive portfolio of on-premises to cloud data extractors for systems and endpoints, enabling companies to free data from operational silos and make it available to users faster. It provides connectors for most standard industrial data sources, from generic extractors (e.g., ODBC, OPC-UA) to ET/OT/IT data sources (e.g., AVEVA PDMS, OsiSoft PI, OpenText Documentum, IBM Maximo) to control system interfaces and application connectors (e.g., Tableau, Spark, Emerson AMS). This enables users to ingest the complete continuum of asset data types, from CMMS to GIS, 3D model CAD, and point cloud data. CDF also comes with a customizable extractor framework to integrate other third-party or custom IT and OT systems leveraging open standards for data transfer.

Making Data Useful

Once the data is ingested, CDF cleanses and groups it in sets based on sources, and runs quality and completeness checks with use-case-specific quality rules. Data workers can leverage

Cognite's flexible industrial knowledge graph to model data across IT and OT processes and use the platform's AI-based contextualization service to link and infer relations between data. Cognite offers open source software development kits in several languages along with open source GUI elements.

Data Contextualization

Contextualization is the core of the CDF DataOps platform, establishing meaningful relationships between data sources and types to help data users find and use relevant data from assets across the operations. Contextualization is adding meaning through relationships and creating active meta data.

A petroleum engineer, for instance, would understand the sensor data streaming from an electric submersible pump of an oil-well site, but a data scientist might not. CDF's contextualization service links the pump identity from the asset hierarchy to its sensor data and related work orders and relates it to the asset's 3D model. Similarly, in the steel industry, a data scientist might not grasp the complexity of predictive quality and steel-grade monitoring and predictive models without a solid knowledge of the underlying chemicals and physics laws. Given more context with a 3D model or knowledge graph, however, a data scientist would be able to visualize the operational context to develop models and data applications — in this case, for anomaly detection.

Enabling Industrial Hybrid MLOps

Cognite offers a combination of data-driven statistical and physics-driven process modeling and simulation. While each approach has its pros and cons, an ML model based on a hybrid of the two will often provide the best results. Cognite empowers developers with workflows compatible with third-party AI tools and other necessary tools to develop, train, and manage hybrid ML models. This enables developers to quickly operationalize use-case-specific data subsets efficiently and at the desired scale.

Making Data Valuable

Enabling CitizenOps

CDF enables data users with low-code/no-code application development and model life-cycle management tools. CDF helps democratize DataOps and facilitate a "CitizenOps" model where non-professional data users can perform data management tasks and develop advanced analytics independently within specified governance boundaries. CitizenOps helps store process knowledge and maintain technical continuity so that new engineers can quickly understand, manage, and enrich existing models.

Going Beyond PoC

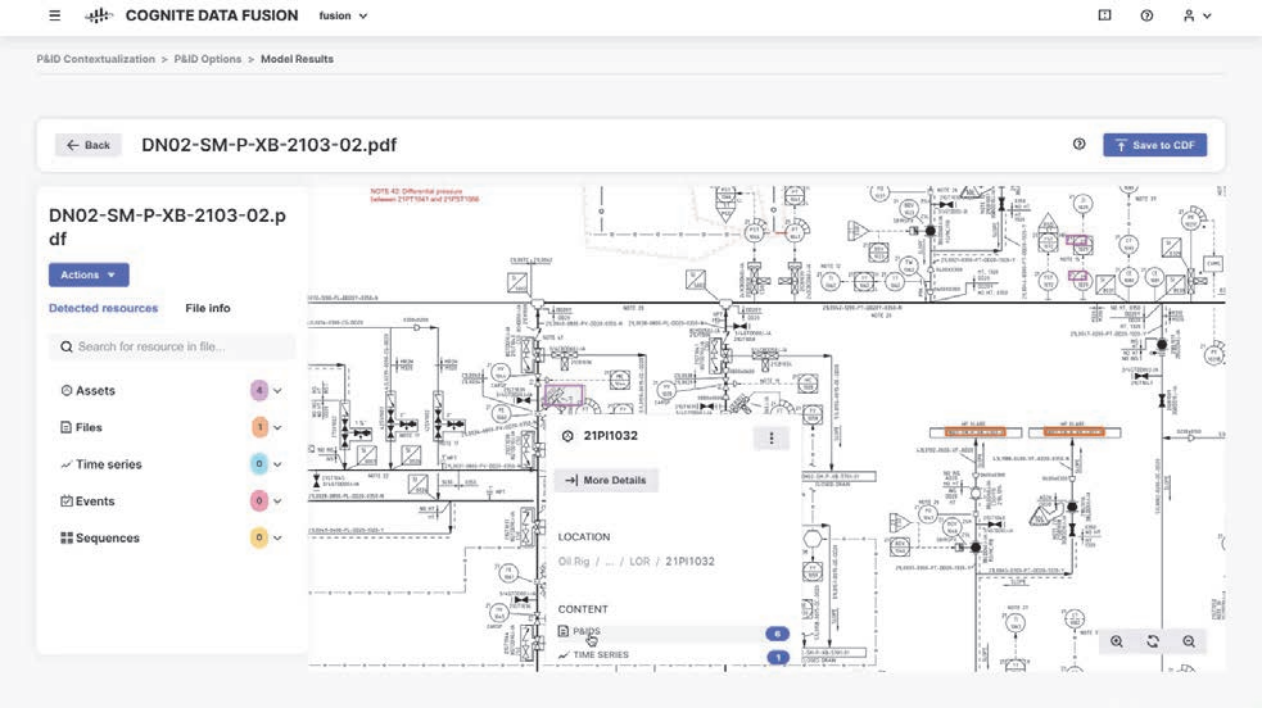
Too often, digital operation initiatives get trapped in "PoC purgatory," where scaling pilots takes too long or is too expensive. What holds them back are the IT/OT and OT/data science divides and the inability to produce and access contextualized quality data at scale.

Cognite's CDF helps bridge those divides on the path to use-case operationalization by connecting data users with disparate operational data sources. CDF provides an ML library of standard industrial use cases that help developers save time collecting data and developing and training their models. Data scientists can leverage this library and use it with component-level data. Once a use case is created, and the outcomes are satisfactory for one component of the plant, the contextualization and ambient state of asset data enables it to be scaled to plant or fleet level.

Access to Off-the-Shelf Business Applications

Cognite also offers prepackaged proprietary solutions to make industrial operations more data-driven. With a portfolio of hybrid AI algorithms for standard operational workflows common across asset-intensive industries, CDF enables data engineers to build use cases faster. Some examples of common use cases include maintenance workflow optimization, engineering scenario analysis, digitization of asset process and instrumentation diagrams (P&IDs) to make them interactive and sharable, and 3D digital twin models to support asset management.

FIGURE 2
P&ID Contextualization in Cognite Data Fusion



Source: Cognite, 2021

Meeting the Challenges

Competition has increased over the past four or five years, and the number of data software and service providers serving industrial verticals is growing. While not all vendors offer full-stack DataOps platforms, many have specialized in critically required functionalities, from asset-specific analytics to root cause analysis and advanced statistical process control. This makes

capabilities mapping increasingly complex for industrial end users in an area where vendor selection must be done carefully to avoid potential lock-in.

This adds to the flourishing digital offerings of equipment manufacturers and industrial automation and engineering software companies. Choosing between the established industrial ecosystem (with its extensive asset and system footprint and captive customers) and specialists' data solutions is increasingly challenging for buyers, especially in industries such as O&G, utilities, and manufacturing, where companies tend to work with vendors with a strong track record.

What to Consider When Adopting Industrial DataOps

Asset-intensive organizations should look to industrial DataOps to unleash ET, OT, and IT data's full potential and transform their traditional operating model.

When starting on this journey, companies should:

1. Think of AI as a critical tool for both fact-driven decision making and efficient management of the data supporting it. Bypassing human "midstream" data handling is critical.
2. "Data liberation" is critical to maximizing value from DataOps. Maximizing your data extraction capabilities will make it easier to plug DataOps into your existing IT and OT architecture, limiting the need to invest in additional systems integration and OT data sources.
3. Develop an IT/OT governance model with data governance at its core. This will dictate how new data is connected and integrated into the overall data architecture. It will also help serve a growing population of data and analytics business users.
4. Prioritize data organization over centralization. Start driving the connection and mapping of all relevant data sources with a clear list of target use cases in mind. As part of the governance model, all new data sources must have a connection, tagging, sharing, and integration plan.
5. Note that not all DataOps platform vendors have the same capabilities. Domain expertise and industry track record should drive selection criteria.

About the Analysts

Gaurav Verma, Research Manager, IDC Energy Insights

Gaurav Verma is a research manager for IDC Energy Insights Europe, focused on the O&G industry. He advises both vendors and end-user clients on how digital technologies can further innovate and transform the existing operational methodology in the O&G industry. His main areas of focus include digital transformation, IT/OT convergence strategies, business advisory, and industry trend analysis for the O&G industry.

Lorenzo Veronesi, Research Manager, IDC Manufacturing Insights

Lorenzo Veronesi is a research manager for IDC Manufacturing Insights EMEA. Veronesi supports all IDC MI research services for EMEA by analyzing IT opportunities in multiple manufacturing industry subverticals and writing quarterly industry updates and standardized reports based on annual surveys on manufacturing end users. He is often involved in consulting projects in EMEA for end users, IT vendors, and public authorities. Veronesi also actively contributes to the Worldwide Operations Technology research services and related consulting activities.

Jean-François Segalotto, Associate Research Director, IDC Energy Insights

Jean-François Segalotto is associate research director at IDC and head of IDC Energy Insights Europe. He leads region- and domain-focused research and consulting, providing full coverage of the utility technology stack and industry value chain. He is also responsible for the European Utility IT and Operational Technologies Strategies research program, covering a wide range of themes, from IT-OT convergence, smart grids, and emobility to meter-to-cash solutions and smart customer operations, smart home solutions, and smart city initiatives.

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

IDC UK

5th Floor, Ealing Cross,
85 Uxbridge Road
London
W5 5TH, United Kingdom
44.208.987.7100
Twitter: @IDC
idc-community.com
www.idc.com

Global Headquarters

5 Speen Street Framingham, MA
01701 USA
P.508.872.8200
F.508.935.4015
www.idc.com

Copyright and Restrictions