



DataOps

Fundamental for Industrial Transformation

By John Harrington

I believe the missing link to industrial transformation is a codeless solution that can aggregate, standardize, and contextualize industrial data from sensors, controls, and other industrial automation assets and systems for use by business users throughout the company and its supply chain. The missing link is an industrial data operations (DataOps) solution.

DataOps is orchestrating people, processes, and technology to securely provide trusted, understandable, and ready-to-use data to all who require it using a scalable and repeatable approach. In an industrial environment, DataOps enables collaboration among data stakeholders throughout the organization, including control and automation professionals, process and manufacturing operations, product designers, quality assurance, maintenance, data scientists, and business analysts.

Industrial DataOps solutions are an efficient way to leverage the value of industrial digital

information by efficiently performing data conditioning while providing a secure data flow to the various consuming applications running at the edge, in on-premise data centers, and in the cloud. These solutions aim to increase velocity, reliability, and quality of data analytics. As you might imagine, the need for industrial DataOps is becoming more important, as there is an increasingly large amount of data available from industrial sensors and controllers.

Although DataOps began as a set of best practices, it has now matured to become a new and independent approach to data analytics. Ultimately, industrial DataOps is a tool for manufacturers to establish and maintain their critical data infrastructure in order to achieve digitalization.

Effects of Industry 4.0

The manufacturing industry is going through a step change so significant that it is being referred to as the fourth industrial revolution.

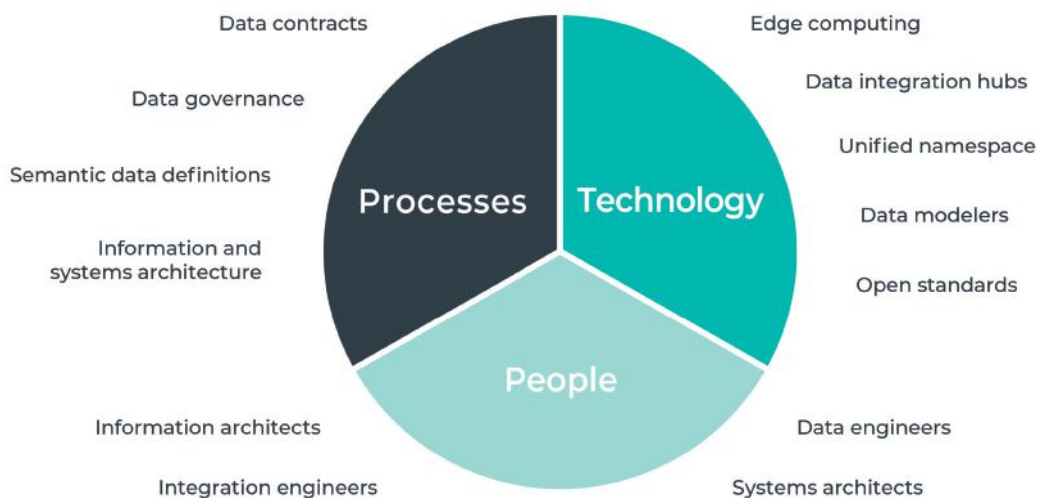
**FAST FORWARD**

- Industrial DataOps orchestrates people, processes, and technology and is an essential component for Industry 4.0 success.
- Contextualizing and standardizing data is required to efficiently leverage industrial data at scale.
- Industrial data security is critical, and the ability to use secure protocols, separate the data by consumer, and control the flow of data all contribute to securing the network, systems, and factory floor equipment.

Digitalization requires data aggregation, standardization, and contextualization at scale.

The first industrial revolution spanned the 1700s and 1800s with the adoption of external power through windmills and water wheels. The second industrial revolution included the electrification of the factory and use of motors to drive machinery in the late 1800s and early 1900s, and the third industrial revolution was defined by the control automation for those motors, which started in the mid-1900s and continues today.

The fourth wave is the cyber-physical revolution with real-time feedback between machinery control, sensors, historical data, business systems, and prescriptive analytical systems. With each industrial revolution, there have been major shifts in the processes used, products created, and dominant industrial companies. Achieving the benefits of industrial utilization and Industry 4.0 requires actionable data.



DataOps (data operations) is the orchestration of people, processes, and technology to securely deliver trusted, ready-to-use data to all who require it.

Drowning in unusable data

Industry 4.0, digital transformation, and smart manufacturing are about using disparate information to drive automated decisions from machinery to the cloud and putting more information in the hands of business decision makers when and where they need it. Companies make this transformation by adopting multiple technologies, including cloud computing, Industrial Internet of Things (IIoT) platforms, advanced analytics, augmented and virtual visualization, mobile platforms, miniature and inexpensive sensors, and networking.

Unfortunately, as these solutions have been adopted and connected, data contextualization and accessibility have been more time consuming, labor intensive, and lacking in purpose-built solutions than expected. Companies that were early in adopting Industry 4.0 technologies thought they could just “hook up” their industrial data to the analytics or visualization applications through APIs and rapidly make use of this data. They found the data was inconsistent across machinery, and data streams had no context to explain what the stream was, where it was from, what the expected tolerances were, or what the unit of measure was. The data was correlated to the controls equipment—not the way business users

think of the business. The volumes of data were immense, and most uses of the data did not require the resolution produced by the equipment.

Today, data is needed near the machinery, in on-site data centers, and in some cases, within cloud-based systems. To solve these data architecture challenges and address the need for data contextualization and standardization, a new category of software solutions is emerging. They may be the key to helping companies adopt Industry 4.0. This category is known as DataOps, or more specifically as industrial DataOps when built specifically for industrial data.

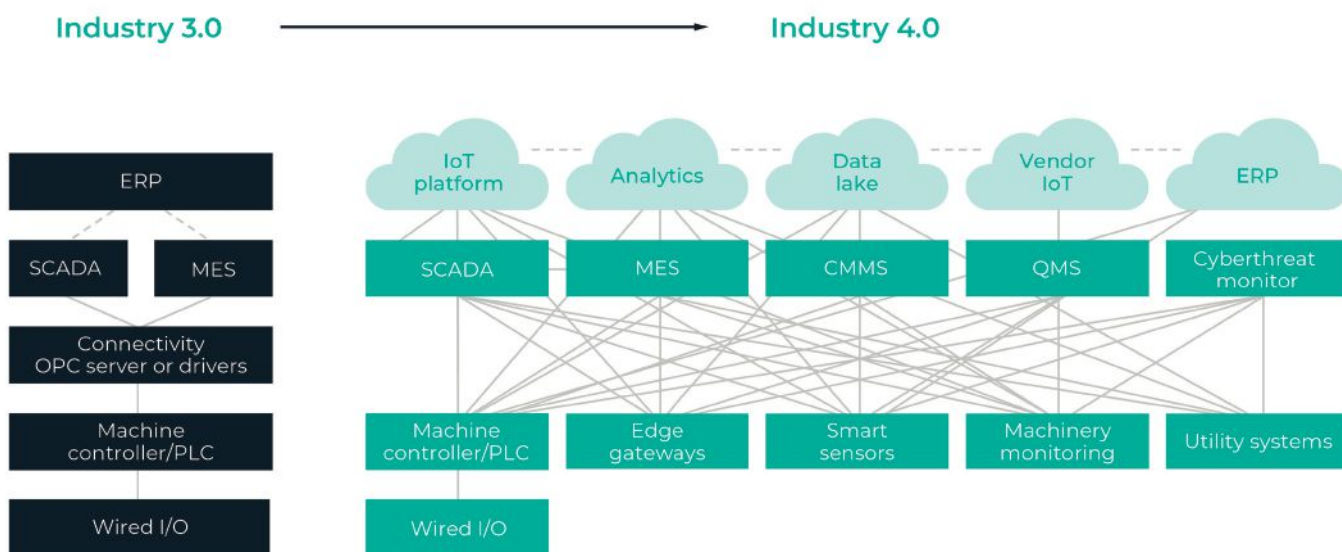
An introduction to industrial DataOps

Before Industry 4.0, industrial data architecture had evolved over many years into a layered approach defined in the Purdue Model or ISA-95. In this model, data flowed from sensors to automation controllers to supervisory control and data acquisition (SCADA) or human-machine interfaces to manufacturing execution systems (MESs) and finally to enterprise resource planning (ERP). The volumes of data dramatically decreased as they move up the stack. Data resolution was also reduced to the point that many companies updated ERP and even MES manually and did not even

connected them to factory machinery.

At the connection between each layer, a number of communication protocols for data were developed; however, most were proprietary, allowing the company developing them to connect to its own hardware or software. These protocols were unique to connecting one layer to the next and were not reused at different layer boundaries. OPC is one of the industry-adopted open protocols that was developed to move information into the SCADA layer from the device layer. Between the machine controller and software layer, OPC servers were developed to translate between proprietary controller protocols and OPC.

Processing data through layers of systems worked for many years primarily because the amount of data being moved up the stack was limited, and much of the data used by the next-level system was generated in the previous system. However, with the advent of Industry 4.0, digital transformation, and smart manufacturing, this was not the case. For example, sensor data at level 0 is not needed for process control, so it is not available in the programmable logic controller (PLC) (level 1) or the SCADA system (level 2)—but the level 0 sensor data may be needed by a level 3 maintenance system. Pushing excess unused data through systems (in this example,



Before Industry 4.0, industrial data architecture had evolved over many years into a layered approach defined in the Purdue Model or ISA-95. However, Industry 4.0 and new target applications like IoT platforms, data lakes, and machine learning applications, have made industrial data architectures exponentially more complex.

level 1 and level 2) that do not require that information can slow down and complicate processing, reduce security, and increase data vulnerability.

Industrial DataOps solutions

Digital transformation is about leveraging data to drive the business. In a manufacturing company, this means extending factory floor operations data from the traditional operations environments to business users throughout the company. These users do not have the same understanding of the manufacturing controls system but require rich data to do their jobs—and do them better. Receiving a data point feed with the name F8:4 and a value of 52.2 does not tell a maintenance engineer that this is a temperature value, that it is sensing the temperature of hydraulic oil on a stamping machine, that the unit of measure is Celsius, and that if the temperature goes over 180 degrees, it will cause premature failure of machine components.

The challenges of optimizing and controlling data flows coupled with the need to contextualize and standardize industrial data have led to the development of a new DataOps organizational function and category of software solutions to support this function. DataOps for industrial environments is different from DataOps for business transaction systems. Why? Because data in industrial environments is very inconsistent across machinery, lacks context, and is correlated to the controls equipment—not assets, processes, and products. It is used from the edge to the cloud—with data security critical at every exchange.

Five essential components

Based on my time speaking with manufacturers and analyzing available solutions, there are five essential tasks of an industrial DataOps solution that are required to achieve value.

1 Standardize, normalize, and contextualize data

Industrial data was created to control motors, valves, conveyors, machinery, and other such equipment. This data typically comes from PLCs, machine

controllers, remote terminal units (RTUs), or smart sensors. It is not uncommon for a factory to have hundreds of PLCs and machine controllers, with machinery and controllers often purchased at different times and from different vendors. As a factory grows, its needs change, and the products evolve. The data points available on the controllers vary from one controller to the next, and very few companies can enforce any consistency.

The data points on controllers were designed for efficiency of communications and use by industrial software solutions. They generally do not include any contextualization, standardization, or documentation of the data packets. This “contextualization” is often stored in an MES, asset management system, or other database system—or may just be known by the operations technology (OT) team. The context must be efficiently merged with real-time data to make the data usable. Furthermore, to get the full value from analytics, data needs to be analyzed across machinery, processes, and products. To handle the scale of hundreds of machines and controllers—and tens of thousands of data points—a set of standard models must be established within the industrial DataOps solution. The models correlate the data by machinery, process, and products and present it to the consuming applications.

2 Connect to industrial and IT systems

Industrial devices and systems and information technology (IT) systems natively communicate in different ways. Industrial devices and systems use many proprietary protocols, though support for OPC UA and other open protocols is increasing. IT systems use their own protocols to communicate, with extensive usage of APIs and bespoke integrations. IT systems communicating with edge devices have begun to use MQTT. MQTT provides a highly flexible pub/sub methodology to minimize the cybersecurity exposure and secure encrypted communications with little overhead. MQTT has been extended with the Sparkplug specifica-

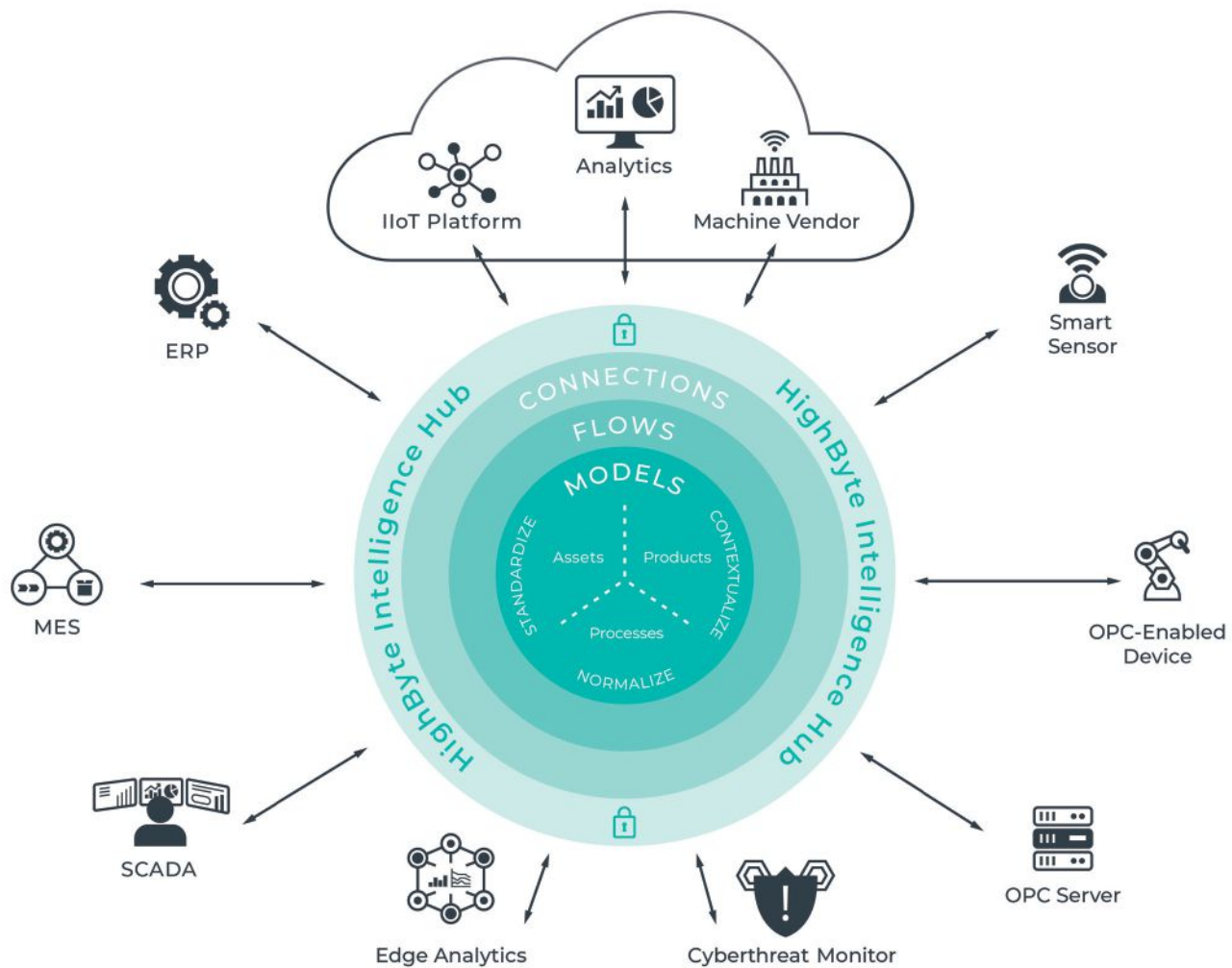
tion to make it more useful in industrial situations and make the integrations easier. On the IT side, many systems are integrated through RESTful APIs and through direct database integrations. An industrial DataOps solution must be able to integrate seamlessly with devices and data sources at the OT layer by using industry standards and providing value to business applications that conform to today's IT best practices.

3 Manage the flow of information

Information flows must be in a managed system where they can be identified, enabled, disabled, and modified. Identifying the impact of machinery changes is critical to making sure good data is being stored and connections are established when change happens. From a security perspective, it is essential to know what data is moving from system to system and to be able to turn it off. Many outside vendors now want machine data to provide enhanced service. The team operating the machine will want to be able to control what data is flowing and at what frequency or set of conditions it is moving. The operations team will also want to be able to disable the data flow if the vendor no longer needs it. Therefore, managing the flow of information is an essential component of an industrial DataOps solution.

4 Provide needed scale and security

Industrial data is different from typical transaction data stored in most IT systems. Industrial data comes from hundreds or thousands of different devices. This data must be captured, contextualized, and delivered at a resolution that is unique for each use case to satisfy the analytics or visualization needs. Industrial data is typically used milliseconds to seconds after it is created. As such, batch processing ETL (extract, transform, and load) solutions built for transactional data do not work well for industrial data. Industrial data must be curated or contextualized close to the machinery and before being stored. Industrial data often holds the intel-



HighByte Intelligence Hub is an example of an industrial DataOps solution that addresses the needs of industrial environments where data must be aggregated from industrial automation and leveraged by business users throughout the company and its supply chain. This architecture satisfies the needs of manufacturers embracing Industry 4.0.

lectual knowledge of a manufacturing plant. This data must be secured and discretely delivered to the applications that need it from the industrial DataOps solution.

5 Live at the edge

Machinery comes in many shapes and sizes and runs in many different environments. Depending on the analytic or visualization application, the data may be processed close to the machinery, in an on-premise data center, or in the cloud. The industrial DataOps solution must run close to the device and feed the applications the required data at the frequency or condition specified. However, the solution must also be able to share models across the fac-

tory and company, allowing for data standardization and normalization.

By using an industrial DataOps solution to define standard models and establish and manage integrations, the operations team can provide data to the systems and business users who are requesting it in an efficient and managed way. This application abstraction approach allows the operations team to effectively substitute changes in the factory, add new applications, and react to changes in business relationships with outside vendors.

By moving data contextualization and access within an industrial DataOps solution, the operations team can own and manage data access, accelerate analytics and visualization projects, and

maintain factory flexibility to change or add new machinery over time. ■

ABOUT THE AUTHOR



John Harrington (john.harrington@highbyte.com) is the cofounder and chief business officer of HighByte. Harrington is passionate about delivering technol-

ogy that improves productivity and safety in manufacturing and industrial environments. He has spent his 25-year career both delivering software to manufacturers and working for manufacturers in operations roles. Harrington has an MBA from Babson College and a BS in mechanical engineering from Worcester Polytechnic Institute.

Reproduced with permission of copyright owner. Further reproduction prohibited without permission.