White Paper

DataOps: The Missing Link in Your Industrial Data Architecture

Second Edition | John Harrington, HighByte Chief Product Officer

Key Takeaways

- ✓ Industrial DataOps is an essential component for success with Industry 4.0, Digital Transformation, and Smart Manufacturing use cases.
- ✓ Data contextualization and standardization are required to efficiently leverage industrial data at scale.
- ✓ Industrial data security is critical. The ability to leverage secure protocols, separate the data by consumer, and control the delivery or flow of data all contribute to securing the systems, network, and factory floor equipment.

Abstract

DataOps is a new category of software solutions that address the data architecture needs of industrial companies as they adopt Industry 4.0, Digital Transformation, and Smart Manufacturing. DataOps solutions perform data contextualization and standardization and provide secure data flow to the various consuming applications running at the Edge, in on-premises data centers, or in the Cloud.

This white paper will help you understand the role of DataOps in achieving the vision of Industry 4.0 at scale, key problems solved by Industrial DataOps, and five required components of an effective Industrial DataOps solution.

The Problem: Drowning in Unusable Data

Industry 4.0, Digital Transformation, and Smart Manufacturing aim to leverage disparate data to drive automated decisions from machinery to the Cloud and put more information in the hands of business decision makers when and where they need it.

The mainstream leap to Industry 4.0 is being fueled by the adoption of many technologies, including Cloud computing, 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 accessibility and contextualization have proven to be more time consuming and labor intensive than expected.

Companies who 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. However, 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 for the way business users think of the business. The volumes of data were immense, yet most uses did not require the data at this high resolution.

> As new technology has been adopted and connected, data accessibility and contextualization have proven to be more time consuming and labor intensive than expected.

Today, data is needed near the machinery, in on-premises data centers, and in many 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 that may be the key to helping companies adopt Industry 4.0. This category is known as DataOps, or more specifically as Industrial DataOps when the solution has been purpose-built for industrial data.

Industry 4.0

The manufacturing industry is going through a dramatic step-change in their business, so significant that it is being referred to as the Fourth Industrial Revolution.

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.

The third industrial revolution was defined by the control automation for those motors, which started in the mid-1900s and continues through today.

The fourth wave is the cyber-physical revolution offering 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.

An Introduction to Industrial DataOps

Prior to 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 SCADA/HMI to MES and finally to ERP. The volumes of data dramatically reduced as they moved up the stack. Data resolution was also reduced to the point that—in many companies—ERP and even MES systems were updated manually and were not even connected to factory machinery. At the connection barrier between each layer, a number of communication protocols for data were developed, but most were proprietary, only allowing the company developing them to connect to their own hardware or software. These protocols were unique to connecting one layer to the next and were not reused at different layers. 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. With the advent of Industry 4.0, Digital Transformation, and Smart Manufacturing, this is no longer the case. For example, sensor data at level 0 is not needed for process control, so it is not available in the 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, level 1 and level 2) that do not require that information can slow down and complicate processes, reduce security, and increase data vulnerability.

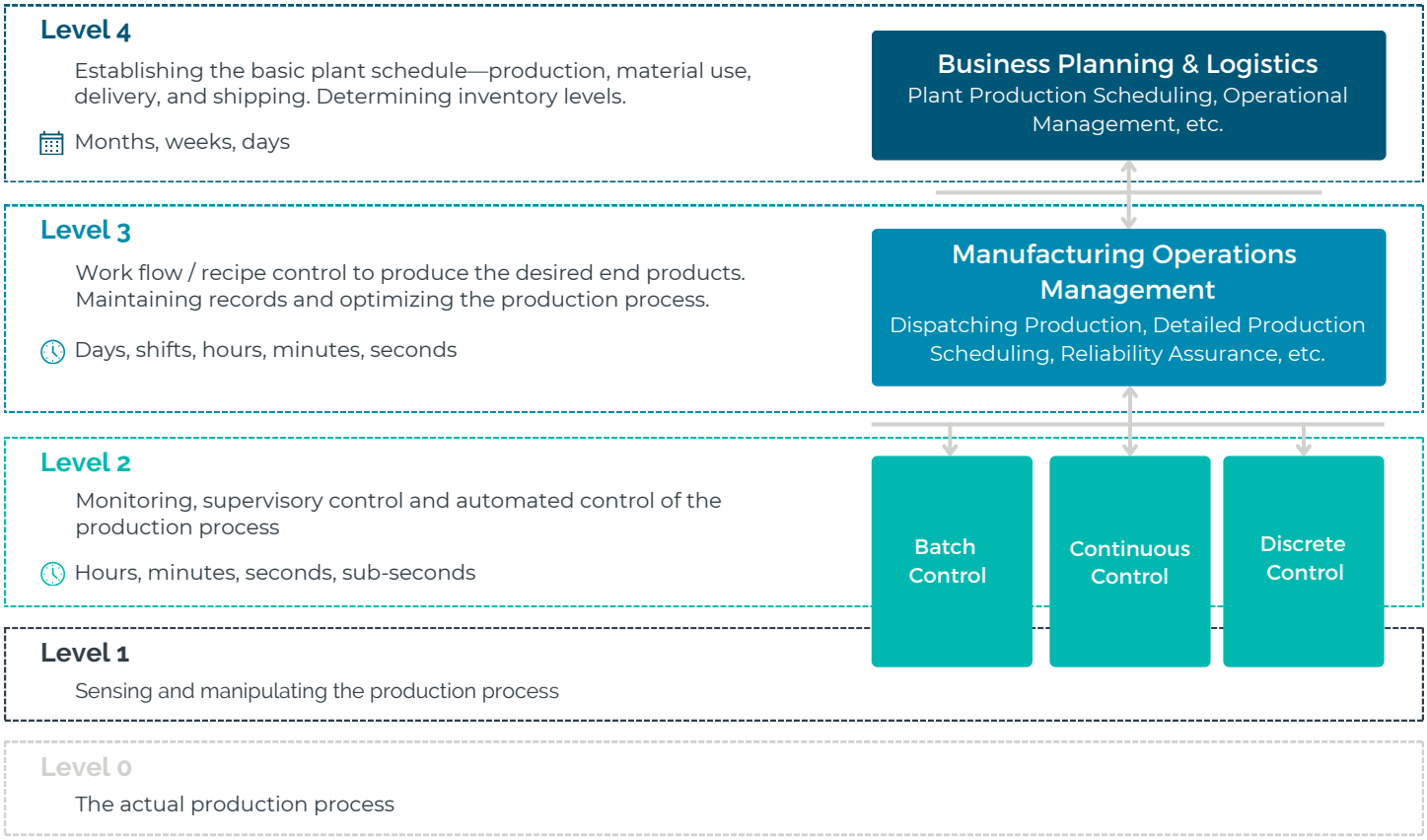

Figure 1: Example of functional hierarchy as defined by the ISA-95 specification.

Digital Transformation is about leveraging data to drive the business. In a manufacturing company, this means extending factory floor operations data outside of the traditional operations environments to business users throughout the company. These new business users do not have the same understanding of the manufacturing controls system, but they require rich data to do their job—and do it better. Receiving a data point feed with the name F8:4 and a value of 52.2 does not tell a maintenance engineer that 52.2 is a temperature value, that the sensor is measuring 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 industrial data flows coupled with the need to contextualize and standardize industrial data have led to the development of the new DataOps organizational function and software solution category to support this function. DataOps for industrial environments is different than 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.

Key Problems Solved by Industrial DataOps

- ✓ Data inconsistencies across machinery
- ✓ Lack of data contextualization and normalization
- ✓ Difficulty preparing and correlating manufacturing data for Quality and Maintenance users to analyze product or asset information, respectively
- ✓ Accessibility constraints from Operations to IT, from Edge to Cloud, and from the business to third parties
- ✓ Difficulty managing and securing data flows
- ✓ Strain on operations personnel caused by time-consuming manual contextualization and point-to-point integrations to IT systems

Required Components of an Industrial DataOps Solution

Based on my time speaking with manufacturers and analyzing available solutions on the market, there are five essential capabilities an Industrial DataOps solution needs 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. The machinery and controllers are often purchased at different times and from different vendors. As factories grow, their needs change and their products evolve. The data points available on the controllers vary from one controller to the next with very few companies able to enforce any consistency. Data points on controllers are 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. 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 DataOps solution. The models then 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 IT systems natively communicate in different ways. Industrial devices and systems use many proprietary protocols, though support for OPC UA and other open protocols are trending upward. 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 leverage MQTT. MQTT provides a highly flexible, with little overhead, pub/sub methodology to minimize the cybersecurity exposure and secure encrypted communications. A DataOps solution must be able to integrate seamlessly with devices and data sources at the operations layer by leveraging industry standards, while providing value to business applications that conform to today's IT best practices.

3 **Manage the flow of information.**
It is critical that information flows are in a managed system where they can be identified, enabled, disabled, and modified. Identifying the impact of machinery changes is critical to make sure good data is being stored and connections are established when change happens. From a security perspective, it is critical to know what data is moving from system to system and to be able to turn it off. Many outside vendors now desire machine data to provide enhanced service. The operations team, who is operating the machine, wants to be able to control what data is flowing and the frequency or set of conditions at which it is moving. They will also want to be able to disable the data flow if the vendor no longer needs it.

4 **Provide the scale and security features required for industrial use.**
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 analytics or visualization needs. Industrial data is typically used milliseconds to seconds after it is created, and batch processing ETL (Extract, Transform, and Load) solutions built for stored transactional data cannot provide the quick data access required for industrial data use. Industrial data must be curated or contextualized close to the machinery prior to being stored. Industrial data often holds the intellectual knowledge of a manufacturing plant so it must be secured and discretely delivered to the applications that need it from the DataOps solution.

5 **Live at the Edge.**
Machinery comes in many shapes and sizes, operating in many different environments. Depending on the analytic or visualization application, the data may be processed close to the machinery, in an on-premises 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, but it must also be able to share models across the factory and company, allowing for data standardization, normalization, and governance.

HighByte Intelligence Hub

HighByte Intelligence Hub is an Industrial DataOps solution purpose-built to model and manage machine, transactional, and time series data at the Edge.

The software application enables manufacturers to securely connect, merge, model, and flow valuable industrial data to and from IT systems without writing or maintaining code. It is the only solution on the market that combines contextualized and standardized data models with connections to industrial and IT systems, manages the flow of information, is scalable and secure, and has been developed with an Edge-native approach.

With the Intelligence Hub, users can speed system integration time, rapidly leverage contextualized data for analytics, AI, and ML applications, and govern data standards across the enterprise. The Intelligence Hub provides the following capabilities:

Codeless integration. Collect and publish data over open standards and native connections—eliminating the need for custom-coded integrations. Easily configure and manage multiple connections and their respective inputs and outputs within the script-free interface. Merge data from multiple systems into a complex modeled payload.

Data transformations. The hub includes a transformation engine based on JavaScript notation that allows data to be standardized and normalized for comparison and application mismatches. The transformation engine enables users to perform calculations at the Edge to improve its usability and reduce the volume of data transmissions.

Data modeling. Represent machines, products, processes, and systems with key intelligence information suited to your needs. Contextualize process data by documenting with metadata, standardizing data attributes, and normalizing units of measure. Models are easily reused within a single hub and can be easily shared across hubs.

Connection flows. Create data flows for raw data or modeled information between connections at any frequency or condition. Manage data flows within the hub to identify, enable, or disable the flow of information to applications.

Security. Exchange data using the built-in security of connection protocols. By identifying outputs by connection, administrators can implement higher-level management and security than typical pub/sub broker architectures and open, unmanaged API access.

Edge deployment. HighByte Intelligence Hub can run on your choice of light-weight hardware platforms including single board computers, industrial switches, IoT gateways, and industrial data servers. You can also deploy the Intelligence Hub as an individual software installation or Docker image, which allows you to rapidly deploy and upgrade system software components.

Conclusion

By leveraging 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 to 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 HighByte

HighByte is an industrial software company founded in 2018 with headquarters in Portland, Maine USA. The company builds solutions that address the data architecture and integration challenges created by Industry 4.0. HighByte Intelligence Hub, the company's award-winning Industrial DataOps software, provides modeled, ready-to-use data to the Cloud using a codeless interface to speed integration time and accelerate analytics. The Intelligence Hub has been deployed in more than a dozen countries by some of the world's most innovative companies spanning a wide range of vertical markets, including food and beverage, health sciences, pulp and paper, industrial products, consumer goods, and energy. Learn more at https://highbyte.com.

About the Author

John Harrington is the Chief Product Officer of HighByte, focused on product management, customer and partner success, and company strategy. His areas of responsibility include market research, customer use cases, product priorities, go-to-market, and financial planning.

John is passionate about delivering technology 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. This experience has given him a unique perspective on how suppliers and end users each play an integral role in implementing new technology solutions.

John has a Master of Business Administration from Babson College and a Bachelor of Science in Mechanical Engineering from Worcester Polytechnic Institute.

HighByte