HighByte

CFE Media and Technology™

Improve industrial data integration with ETL software

Extract, Transform, Load (ETL) software can help improve data gathering for Operations Technology (OT) applications, but there are major challenges with data integration that companies need to overcome.

Author: John Harrington, HighByte Co-Founder & Chief Business Officer

This article first appeared in its original form in the June 2019 issue of Control Engineering, a CFE Media publication.

Introduction

Most people are familiar with Industrie 4.0, Smart Manufacturing, and the Industrial Internet of Things (IIoT). These terms are used to describe the tremendous changes in operations technology brought on by a surge in underlying technologies including cloud, big data, smart sensors, single board solid state computers, wireless networks, analytics, application development platforms, and mobile devices.

Some of these technologies are not new, but recent price drops and improved ease of use have increased their usage. These technologies are being combined with traditional Operations Technology (OT) like control systems and Manufacturing Execution Systems (MES) to improve operations and business functions of industrial companies by providing more data—and tools to leverage that data.

Many of these technologies were first developed for Information Technology (IT) departments to interact with other business disciplines like marketing, sales, logistics, and finance. Given the vast amount of data in manufacturing and the need to improve operations, these tools are being evaluated and adopted by IT. However, operations teams looking to leverage industrial data face unique challenges around data integration, which have increased the effort required to deploy such systems.

The IT industry solved its data integration challenges by creating Extract, Transform, Load (ETL) software, which integrates business systems into analytics systems. These solutions are designed to extract data from other systems and databases like Customer Relationship Management (CRM) and Enterprise Resource Planning (ERP), combine this data in an intermediate data store, and transform the data by cleaning, aligning and normalizing it. The data is then loaded into the final data store to be used by analytics, trending and search tools.

So why can't ETL solutions be used by operations to prepare industrial data? Simply put, industrial data coming off the controls system in a factory has different challenges than transaction data from business systems. Let's look at these challenges in more depth.

Extract

Operational data is not all stored in a database as transactions waiting to be extracted. Rather, it is available in real time from Programmable Logic Controllers (PLCs), machine controllers, Supervisory Control and Data Acquisition (SCADA) systems, and/or time series databases throughout the factory. Instead of extracting data from a handful of large databases, data must be collected from hundreds of devices and systems.

Transaction processing systems store complete records for each transaction, but in factories, process data is not captured as "transactions." A high-volume discrete manufacturer cannot store the complete data set for each component that comes off the line. A batch manufacturer often needs to store more than a single value per

batch. Industrial data must be collected at a high rate to catch any anomalies and then stored at different rates based on the use case. This makes extraction much more complex (see Figure 1).

Transactional Data vs Industrial Data

FIGURE 1

	TRANSACTIONAL	INDUSTRIAL
Source	Databases	PLCs, SCADA, Historians
Number of sources	1-5 large databases or logs	Hundreds of devices and systems and thousands of data points
Data resolution	Transaction with multiple data points	Unique data points stored at rate of change
Data state	Data at rest in database	Streaming data
Data transformation execution	Batch	Real-time
Data format	Standardized	Unique sets of data points defined per machine
Missing data source	Corrupted data in database	Network outage, connection lost
Inconsistencies	Mistaken entry	Inconsistent definitions
Location of consuming applications	Cloud	Edge, data center, and cloud

Figure 1: Comparison of transactional data extraction to the more sophisticated requirements of industrial data extraction across a number of factors.

Transform

Data transformation on operational data requires more of a conditioning than a transformation.

Operational data storage often happens periodically—every second, minute or hour. The stored data may be an actual value like the quantity produced, or it could be statistical calculations of the raw data like the average, minimum, and maximum temperature values checked every second, but recorded every hour.

Data points on PLCs generally have an address or name and a value. However, these data points only provide a process or controls-centric view of the data. There are no descriptions, units of measure, operating ranges, or other descriptive information.

This creates challenges as industrial data is used outside of the controls environment for machine maintenance, process optimization, quality, and traceability. In these cases, the data must be analyzed and aligned by machine for machine maintenance, by process for process optimizations, and by product for quality and traceability. The required data often is available, but must be correlated and sometimes transposed into a usable format.

Typical factories also have machinery from many different vendors and equipment that has been purchased over a 10 to 30-year timespan. This variety in machinery results in a wide variety of available data. Some data points may have different names while others may have different units of measure or different measurements entirely. For analytics, trending, or any sort of data analysis to be possible, the data points must be standardized, normalized, and in some cases, calculated based on component measures.

Finally, analytics data generally is not as critical as controls data, so companies have started to use lower-cost sensors to collect data for non-critical analysis. However, these sensors can fail or drift so having redundant sensors with external data validation is important to ensure good data is being stored.

Load

With the introduction of these new technologies, the number of business users who want access to high resolution, automated data feeds from operations has increased. They use unique systems to analyze and make use of the data and have differing requirements. These business users vary by company but often include manufacturing operations, maintenance, quality and value engineering. Machine vendors also have started to sell service contracts with requirements for real-time data collection.

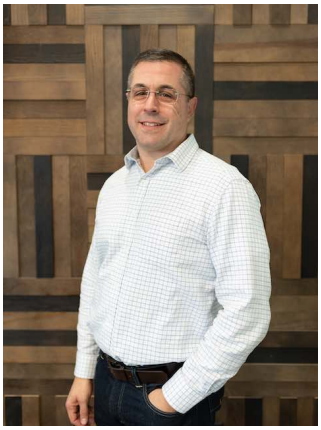
Managing the delivery of data is important. There are security risks as well as significant costs associated with storing incorrect, corrupt or useless data.

Industrial data extraction and transformation must happen close to the production machinery. This allows the data to be used by local edge analytics and sent to on-premises data centers or the cloud based on which is more efficient.

Realizing Data's Value

The need to extract, transform and load operational data is as great as—if not greater than—the need for ETL in a typical IT business system integration. Yet industrial ETL has unique and sophisticated requirements. This demands a rethink of data architecture and the creation of new industrial data infrastructure solutions. These new industrial data infrastructure solutions must simplify and streamline data integrations for industrial companies to achieve the value expected from Industry 4.0, Smart Manufacturing, and the IIoT.

About the Author



John Harrington is the Co-Founder & Chief Business Officer of HighByte, focused on defining the company's business and product strategy. 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 in Business Administration from Babson College and a Bachelor of Science in Mechanical Engineering from Worcester Polytechnic Institute.

About HighByte

HighByte is an industrial software development company in Portland, Maine building solutions that address the data architecture and security challenges created by Industry 4.0. We believe contextualized and standardized data is essential for Industry 4.0 to reach broad adoption. That's why we've launched HighByte Intelligence Hub—enabling manufacturers to securely connect, model, and flow valuable industrial data throughout their extended enterprise without writing or maintaining code. HighByte Intelligence Hub is the first DataOps solution purpose-built to meet the unique requirements of industrial assets, products, processes, and systems at the Edge. Learn more and request a free trial at https://highbyte.com.