

Solving big data problems with a little data approach

By Derek Thomas



ABOUT THE AUTHOR

Derek Thomas is the vice president of marketing and strategy for Emerson's machine automation solutions business, responsible for global sales, strategy, marketing and product management. He has an MBA from Washington University in St. Louis and a BS in mechanical engineering from Purdue University.

Nearly every industrial facility has an opportunity to create value from collected and stored big data by implementing Industrial Internet of Things (IIoT) and other operational improvement initiatives. In the process industries, this data often resides in centralized control systems and historians, while in discrete part manufacturing, the data is more likely to be dispersed across the plant or trapped within machines. But no matter where the data is collected or stored, the best approach to creating value is often to start small, with a "little data" approach.

McKinsey & Company gives some insight into the scale of data analysis opportunities. "Most data generated by existing IIoT sensors is ignored. In the oil-drilling industry, an early adopter, we found that only 1 percent of the data from the 30,000 sensors on a typical oil rig is used, and even this small fraction of data is not used for optimization, prediction, and data-driven decision making."

Many big data projects fail because a "boil the ocean" approach is pursued, whereby substantial time and capital are committed upfront in the hope of analyzing all the stored and incoming data to derive insights. These types of approaches usually begin with discussions of what technologies should be used, particularly the cloud and other IT-related infrastructures, and often end with frustration and unsatisfactory results, even after months or even years of effort.

A better and more practical way is to start at the machine or production-line level by defining specific problems that are solvable with better use of little data. Focusing the field of view to a specific, defined asset reduces complexity and simplifies the search for a solution. This simplification is crucial, because the people implementing a little data project should be the personnel most familiar with operations.

Another advantage of the little data approach is it quickly yields tangible improvements by empowering users to find, solve, improve, and move on to the next opportunity. This creates positive momentum within a company, and as experience and comfort increases, it becomes easier to scale efforts to larger data sets using the lessons learned.

For these and other reasons, a little data approach is the most practical path for IIoT projects at many industrial plants, but it requires different technologies than the IT-centric methodology used

in many big data projects. The plant personnel most familiar with operations are generally also very competent when it comes to real-time control systems. This is by necessity; these systems keep plants running smoothly, and adjustments to these systems are often required to improve operations. Unfortunately, most real-time controllers do not have the required capability to analyze data produced by field devices to generate insight, a requirement for little data projects.

This type of edge processing has traditionally required a separate industrial computing device and software solution to store and process data. Integrating these elements with the existing controller and network was often problematic due to: the complexity of setting up, programming, and managing in two different environments; synchronization; lag/latency; and other issues. A little data project using two separate devices could thus become quite complex and unreliable, slowing implementation and driving up costs.

A modern class of edge controller addresses these issues by combining two functions into a single device. The first function is real-time control, much like what was done by a traditional programmable logic controller. The second set of functions is performed by a computing platform with a processor capable of data storage, analysis, and a wide range of other tasks—similar to what could be done with an industrial PC. Because both functions are performed in one device, there is no additional effort required to integrate two components—data is simply passed between the two functional areas safely and securely.

Once the edge controller stores and processes the data already being collected for real-time control, results are readily transmitted to enterprise platforms, such as manufacturing execution systems, enterprise resource planning, maintenance management, and other analytics systems—both on premises and cloud-based—through the typical industrial or Ethernet protocols. These higher-level platforms thus have the information required to improve operations.

Big data projects seem to call for big and complex solutions, but a large-scale approach often fails due to high costs and excessive implementation time. A better way is to begin with targeted efforts for analyzing little data to create insights, creating value and building momentum. ■