

A Little Data Approach to Solving Big Data Problems

By Derek Thomas

Vice President of Sales & Marketing • Machine Automation Solutions at Emerson

Creating value from big data is top of mind for many executives across a wide variety of industries. These executives perceive the potential to substantially improve operations by creating insights into the terabytes of existing data stored in their organizations' databases, with more data flooding in every second. Ironically, the best way to approach this problem can often be to start with little data problems and insights before progressing to big data solutions.

With big data opportunities come challenges, and perhaps the greatest is the sheer volume of data. There is often a temptation to tackle the issue all at once, with mega-scale projects ambitiously gathering all the data from various sources into a data lake, either on premise, in the cloud, or a hybrid of the two. Once stored, this data must be cleansed to remove anomalies, and this is an ongoing process as more data floods. Another issue arises because big data often comes with significant dilution of insights as data naturally starts to normalize and hide intricacies of events.

During this process, it's easy for a company to focus on the infrastructure and technologies instead of the problem, which can lead to discouragement due to the complexity, cost, and scale of implementation. These and other issues can cause companies to lose sight of the goal, which is creating an acceptable return on investment by realizing operational improvements.

Because big data and the tasks required to create value from it presents a formidable challenge, a better answer may be to consider a solution involving little data. For someone new to IoT, greatly reducing the field of view to a specific, defined source and need reduces the complexity and simplifies the search for a solution. This approach yields real tangible improvements and creates positive momentum by allowing users to find, solve, improve, and move on quickly. As experience and comfort increases, it becomes easier to scale to larger data pools because there is a baseline of insights to build upon.

One of the main technologies for creating value from little data is edge computing,

whereby data produced by field devices is analyzed by a field-located controller to generate insight. This information can be supplied to the right personnel, close to the source for appropriate action.

In the past, this type of edge processing would have required the addition of a separate industrial computing device and software solution to store and process the data. This new device would then have had to be integrated with the existing controller and network, often a problematic step due to the complexity of setting up and programming in two different environments, synchronization, lag/latency, and other issues.

But today, an edge controller can be used to perform both functions. The first is real-time control, much like a traditional programmable logic controller (PLC), and the second is data storage and analysis. Because both functions are performed in one device, there is no additional effort required to integrate two components.

Once the edge controller stores and processes the data already being collected for real-time control, results can be transmitted to enterprise platforms such as MES, ERP, maintenance management, and other systems—both on premises and cloud-based—through the use of embedded connectivity capabilities. These higher-level platforms thus have better information to use in various ways to

improve operations.

Higher-level connectivity is simplified because an edge controller not only stores and analyzes data, but also supports embedded IT-compatible connectivity and security to communicate over standard networking technologies and protocols. Ethernet is of course the network of choice for most higher-level communications and is well supported by a typical edge controller, but this is not enough as various industrial Ethernet protocols must be supported as well.

Common IT protocols must also be supported, as well as industrial protocols, such as PROFINET or WirelessHART. Cloud-based communications often require their own lightweight protocols, such as OPC UA, MQTT or AMQP, so these must be supported as well.

This article has introduced three new but related concepts for deriving value from big data: little data, edge processing, and embedded connectivity. Each of these concepts will be explored in detail in three subsequent columns, providing a roadmap for new applications and implementations to create value from the data available in industrial automation systems.