

**March 2015**

# IoT and Big Data Combine Forces

The Internet of Things produces big data that can drive improvements in manufacturing and light industrial applications

Author: Advantech

E-mail: [eainfo@advantech.com](mailto:eainfo@advantech.com)

**March 2015**

The Internet of Things (IoT) and big data are two hot topics with respect to commercial, industrial, and other applications. The term “IoT” was coined in 1999 and refers to the world of devices connected to the Internet, which is the method by which much of big data is collected, concentrated and curated. Big data additionally refers to the analysis of this information to produce useful results.

A primary driving force behind the IoT and big data has been the collection and analysis of data concerning consumer behaviors in order to find out what people buy and why. An example of this is the loyalty cards that shoppers use at grocery stores and other retail outlets. Using these cards, retailers and their suppliers can ascertain what customers buy which products, and can then use this information to increase sales and profits.

Commercial off-the-shelf (COTS) technologies, such as the PC and various Microsoft operating systems, have long migrated to industrial automation applications in the manufacturing sector. This is also the case with the IoT and big data, as the COTS technologies that underpin the implementation of these concepts are migrating from the commercial sector to industrial automation.

This white paper will first show why manufacturers should integrate the IoT and big data into their industrial automation systems, and then it will reveal how this can be done using products and systems available today. The white paper will conclude with a look at the future of the IoT and big data, illustrating how these concepts will help create the factory of the future.

#### Why Implement the IoT and Big Data?

The commercial and governmental sectors are showing how the IoT and big data can be used to improve operations, but how can these concepts be profitability extended into manufacturing and other related industrial sectors?

It helps to first define where this data is coming from, figure 1 shows some of main vertical markets where big data is collected. These markets include but aren't limited to agriculture, electricity, forestry, water treatment and virtually every type of manufacturing facility.

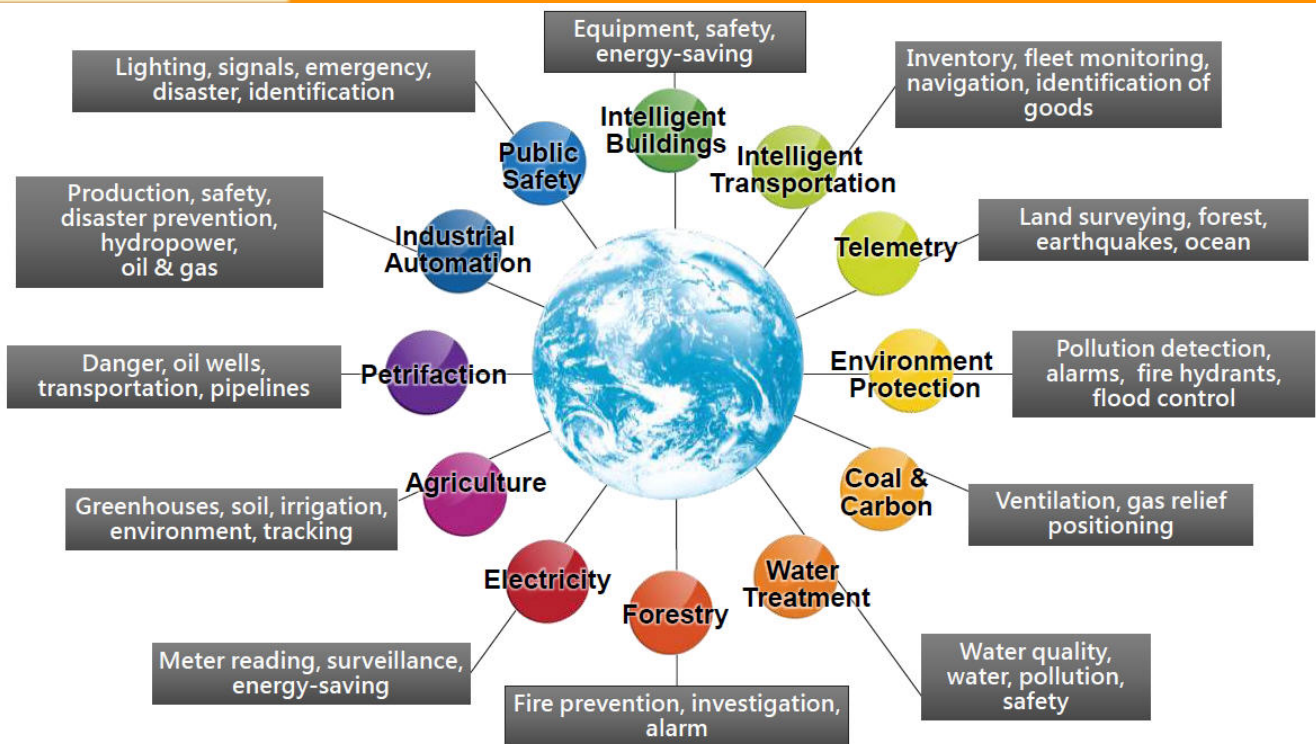


Figure 1 - Big data is produced across a wide array of vertical markets, where it can be used to improve operations.

As Sanat Joshi explains in ISA’s *InTech*, “Big data is the new norm for enterprise analytics and it is pervasive across many industries.” He says that data becomes big “when the volume, velocity, and/or variety of the data exceed the abilities of your current IT systems to ingest, store, analyze, or otherwise process it<sup>1</sup>.” Big data typically conveys an individual perspective of perceived magnitudes of information greater than expected, and may express a conundrum of comprehension and utilization.

Data input originally came from limited human-based measurements, observations and manual data entry—but most data today is automatically generated by sensors, with the trend of field sensors becoming more numerous and smarter with more reporting capabilities continuing to grow.

Once this rapidly expanding stream of data is collected, it then needs to be concentrated and curated, terms that will be explained in detail later in this white paper. This curated data must then be visualized and analyzed in order to improve operations, a process that will also be expanded upon.

Table 1 lists how manufacturing and other related vertical industry operations can be improved by the intelligent implementation of the IoT and big data. These operational improvements will result in better products, increased throughput, less downtime, and lower costs. Figure 2 lists some of the specific areas of manufacturing that will benefit including but not limited to production, process control, and packing, and test and inspection.

Table 1: Why Implement the IoT and Big Data?

1. Predictive health monitoring
2. Less downtime
3. Lower reject rates
4. Improved quality
5. Higher throughput

6. Improved safety
7. More efficient use of labor
8. Enable mass customization



Figure 2- The IoT and big data can be used to improve many aspects of manufacturing including production, process control, packaging, and test and inspection.

Industrial packaged equipment is typically automated with embedded controllers or programmable logic controllers (PLCs), and both types of controllers are commonly available with Ethernet connectivity (Figure 3).



Figure 3 - Embedded Controllers provide real-time automation for industrial processes and machines as well as act as data concentrators

Even individual sensors, analyzers, radio frequency identification (RFID) hardware, vision systems, and other devices are becoming “smarter” and are offering network connectivity. Thus, more data than ever is available from the factory floor and process plants. On top of this, automation components have more memory, processing power, and standard functionality on-board.

This combination of information and capabilities can be harnessed by a top-level predictive health monitoring system. Motor run times and start counts, valve actuation counts, and other tracking metrics can be configured to alert the user when equipment requires service. More detailed diagnostics such as temperature and vibration monitoring can trigger a useful warning of imminent trouble. In fact, many devices now include powerful diagnostic information on-board, just waiting to be tapped by end users.

For manufacturing industries, uptime is crucial. Enhanced diagnostic capabilities directly lead to reduced downtime as problems can be identified quickly or even before they fully emerge. With a wide enough range of system information, predictive analysis methods can potentially alert the user to trouble before it even happens. As Kevin Ashton puts it in *RFID Journal*, "If we had computers that knew everything there was to know about things - using data they gathered without any help from us - we would be able to track and count everything, and greatly reduce waste, loss, and cost<sup>2</sup>."

In fact, for less severe problems it is entirely possible for systems to proactively and automatically adjust themselves or take corrective actions, all without any required human interaction. Such a system would obviously result in lower reject rates, improved product quality and higher throughput.

Even though the IoT and big data may just seem like autonomous functions once they are implemented, their positive impact on operations personnel can be substantial. The ever expanding availability of sensed information means there is a reduced need for operators to venture out into the field, especially vital in hazardous environments such as refineries. Additional data from IoT devices can also lead to improved process control that makes operations safer.

Self-diagnosing systems mean that maintenance technicians can spend less time troubleshooting dead ends and more time performing hands-on repairs. Predictive health monitoring allows repair events to be scheduled in a coordinated manner that minimizes production impacts, and reduces the human toll associated with events such as third-shift emergency repairs.

Combining the IoT and big data can even facilitate new capabilities. For instance, systems configured with enough intelligence can offer "mass customization", which uses computer-aided manufacturing methods to enable individually customized products produced with efficiencies and costs approaching mass production. Customers can thus have access to a new variety of customized products at a less than a full-custom price, while manufacturers are able to tap new markets and realize increased profits.

While speed is important in harvesting big data from the IoT, it's important to point out that determinism is the real goal. Determinism implies that each action can reliably predict another action. The more information we have, the more deterministic we can become since we can correlate more inputs to derive more accurate output predictions.

More accurate predictions will in turn improve prognostics and health management and other predictive technologies. This ever-improving automated sense/predict cycle leads to increased productivity due to reduced manual data collection and interaction. Better quality products are made closer to specifications and tolerances, with lower reject rates and higher throughput.

But all these benefits can't be realized until data is first collected, which starts at the sensor level.

### Collection

It's obvious how designers are incorporating sensors into countless devices used throughout our everyday lives. Smartphones react to user inputs by tracking the smallest of gestures, automobiles constantly monitor vehicle motions to achieve stability control, and buildings vary environmental controls to provide both comfort and energy savings.

In some ways, the vast availability of sensors and their produced data has provided kindling for the IoT and big data fire, while the automated method of querying these sensors, largely over the Internet, has sparked the fire to life.

When looking from a system-wide perspective, the sensing function can be considered to exist at the lowest level, or out on the "edge". Table 2 offers a list of common sensor types, and shows how the IoT sees the world.

Table 2: Types of Sensors

1. Temperature
2. Pressure
3. Level
4. Flow
5. Density
6. Proximity
7. Position, angle, displacement, distance, speed, acceleration
8. Vision
9. Vibration
10. Current, voltage and other power parameters
11. pH and other analytical values
12. Moisture, humidity and other weather conditions

The classic four measurements for production processes are flow, level, pressure, and temperature. Additionally, chemical, pharmaceutical, water treatment, and food industry processes will rely on analytical measurements of pH, dissolved oxygen and many other specific variables. Environmental and facility control systems will look for temperature, moisture, humidity and other weather conditions.

Extending into the mechanics of many types of equipment and machinery, there are various sensors for measuring position, angle, displacement, distance and proximity. Speed and acceleration sensors monitor the change of those same sensed values over time, while vibration sensors monitor the frequency and/or amplitude of motion.

For electrically driven equipment, sensing devices report the instantaneous current and voltage values and other related power parameters. Even more esoteric is the use of vision sensing systems to track produced parts, verify quality, or simply read tags. Vision systems in particular can generate huge amounts of image data.

Sensors generate numerous types of signals, but these various signals must all be transduced from the raw measurement into a useful signal format for transmission. A classic method of transmitting a single analog value is to use a transducer that scales the value into a 4 to 20 mA signal wired to a control system analog input module.



More advanced “smart” transducers can convert the sensed signal into a standard digital communications protocol, Ethernet-based in many cases. Signal converters are also available that can effectively upgrade legacy type signals and equipment by transforming proprietary or obsolete serial signals to modern open network protocols.

“Smart” transducers or sensors are preferred for several reasons. Not only can the primary sensor signal be monitored, but very often there are additional process values available that prove useful. For example, most smart pressure, flow and level instruments include temperature measurement, very useful for signal compensation and general health monitoring. Smart sensors and instruments usually also collect and transmit diagnostic information including device status, alarms and events.

Finally, sensors and transducers networked via serial or Ethernet links can provide simplified installation, particularly for wireless variants. There are many paths to collect data, and this data must then be concentrated.

### Concentration

Data typically journeys from networked sensors to some form of data storage or concentration device. This device or devices can exist in a number of locations. Classically, the data would be deposited in large PC server or mainframe systems in a control room or mission control center. Table 3 identifies some other data concentration options.

Table 3: Types of Data Concentrators

1. Data loggers
2. Embedded controllers
3. Embedded PCs
4. PLCs
5. Multi-variable instruments and transducers
6. Smart sensors

Older data loggers, many still in use today, consist of round or scrolling paper charts with ink pens tracing information, similar to an analog earthquake seismograph. Today’s data loggers are fully digital and offer numerous improvements.

Solid state data loggers can accept large quantities of input channels, capturing values at defined sampling intervals with the capacity to store tens of thousands of data values in volatile or non-volatile memory. This type of device makes it practical and economical to extend data logging to remote locations. These data loggers may be network-connectable, or may have removable memory cards or USB connectivity for archiving the stored data to a higher-level system.

Embedded controllers are often used as a dedicated means to automate equipment. Although primarily tasked with cost-effective machine control, these controllers often have sufficient memory and processing capabilities to also perform data logging. Similarly, embedded PCs have more than enough power to perform automaton and data logging tasks, and often are configured with rotating or solid state hard drives that can provide significant data storage capacity.

PLCs are used in many of the same applications as embedded controllers. Historically, a PLC’s strong suit has been reliable, flexible, and high-speed control of machinery. They have also been a common gateway for providing raw data to higher level data concentrators.

Recent years have found PLC processing power and memory capabilities increasing, and in fact the most capable PLCs are now often called process automation controllers. Whatever the moniker, the latest PLCs offer impressive non-volatile storage capabilities, and some even offer historian modules that plug into the PLC chassis. For many cases, this makes the PLC itself a viable option for concentrating data.

And looking back for a moment at sensing devices, the most capable of these often have on-board data logging. Often these devices can log to memory cards or USB memory, or they can be networked into a higher level data concentrator.

Just as there are myriad options for collecting data, there are more options than ever for concentrating the abundance of this information. Data concentration can be scaled to meet the need, ranging from many smaller field devices all the way up to large centralized or distributed PC-based systems. It's becoming more straightforward to reduce the amount of manual human intervention required to obtain useful data, which in turn drives up the quantity and quality of data available for curation, and eventually interpretation.

### Curation

Once data has been harvested from sensors, and concentrated onto some type of storage system, it becomes important to organize the information in a manner that helps users make sense of it. The activity of "curating" this data typically involves the use of database software. Curating is usually accomplished at the PC server level, since database management is an advanced software function that requires equally capable hardware.

Some of the most well-known commercially available business databases include Oracle Database, Microsoft SQL Server, Microsoft Access, SAP Sybase, and IBM DB2. These software packages are used throughout industry for maintaining all sorts of data, and they offer a platform for querying and analyzing information.

However, not all data is exactly the same. In the business and commercial world, data often consists of customer information, financial transactions, product stock numbers, stocking quantities, and shipping information. This transactional data requires very secure data storage, with relatively slow rates of data collection. Interruptions to data access and other functions aren't critical.

When we look at the manufacturing world, we find that a large amount of the information obtained from sensors consists of time-series information where each sample is the value of a sensor with a timestamp. This real-time data is sampled repeatedly. Process data may also include alarm values and discrete process events. Manufacturing data must be stored and accessed in real-time, and uptime is critical.

With so much information potentially available, it's critical to capture what is most important. *Control Engineering* finds that "Lost in the big data flood are indicators that can help plants understand the dynamically changing risks and avoid some of the \$10 billion losses the U.S. chemical and petrochemical industry experience annually (due to unexpected shutdowns)<sup>3</sup>." Without informed analysis, full value is not received from the effort and expense spent in collecting, concentrating and curating data.



There is a class of operational or industrial databases or historians that have developed specifically to address the needs of manufacturing processes. These software packages are optimized to address tag-based data (such as might be obtained from IoT devices), and to log large amounts of time-series data at high speed. Another characteristic of process historians is that users can choose the desired sample rate and precision. Careful configuration of these attributes can allow the historian to compress the data storage size, and reduced database sizes improve query response times.

Curated data is commonly available within an organization over the facility or company network, which is often referred to as an “intranet” or a “private cloud”. The Ethernet is considered a “public cloud”, but there are secure methods to link multiple facilities or data sources using it.

Easy access to historical database information and the ability to retrieve this data in a variety of useful ways (such as raw value, average value over a given period, and a graph of values over a given period), powers higher level analysis of big data.

### Visualization and Analysis

To a great extent the data collection, concentration, and curation activities occur silently and unseen after initial configuration. These activities must be in place to act as a foundation for visualizing and analyzing information. However, simply presenting users with large tables of values from a historical database will not help most people to understand the data, since most people are more visually oriented towards graphical representations of data such as charts, graphs and other symbols.

Data scientist Nate Silver, interviewed by Jon Gertner in *Fast Company*, relates that “The flood of data means more *noise* (i.e., useless information) but not necessarily more *signal* (i.e., truth).” He goes on to state that “People blame the data, when they should be asking better questions<sup>4</sup>.” The right data visualization tools can help users understand data and ask the right questions, and to quickly interpret and act upon the answers.

Fortunately, more options than ever are available to display quantities of information and help users make sense of it (Table 4). At the field level, many analyzers and data loggers offer local displays or come with associated operator interfaces. These devices often allow users to view trends of data, alarms and events.

#### Table 4: Types of Visualization and Analysis Tools

1. Smart phones
2. Tablets
3. Operator Interface Terminals
4. PC-based HMIs
5. Data loggers
6. Transducer and analyzer displays

For data coming from an embedded controller or a PLC, a local operator interface terminal (OIT), which is a dedicated device as opposed to a typical PC, is a good choice for displaying data and alarms for the related equipment or area (Figure 4).



Figure 4 - A local operator interface terminal is often the best choice for displaying and analyzing data in a defined location or area.

If the controller/OIT is provided by an equipment manufacturer or system integrator, there is a good chance that useful local indications, data logging/trending, and alarming capabilities will be built-in and thus readily available without any extra effort required by the end user.

For many applications, the point where visualization and analytical performance really gains traction is when software on a PC platform taps into historical data. The software historians mentioned in the “Curation” section offer their own visualization and reporting tools, and most PC-based HMI packages include these and other data analysis features. Operators can trend various data points over chosen time ranges, and effectively perform their own investigation of the available data (Figure 5).

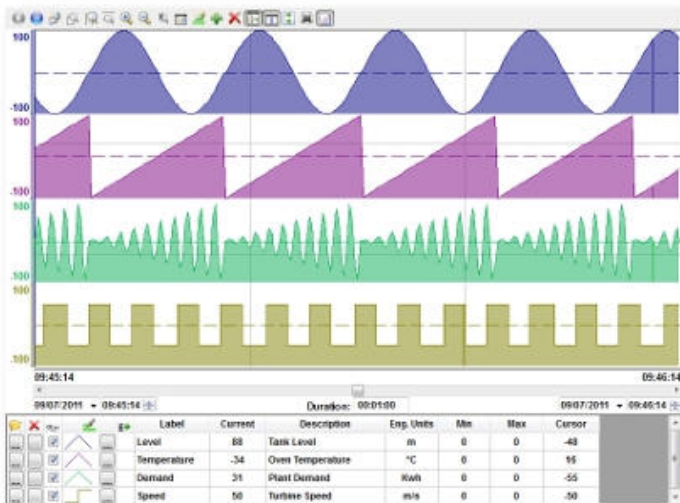


Figure 5 - PC-based HMI packages and operator interface terminals typically include data analysis features such as trending of various data points, over chosen time ranges.

These products can be configured to offer desired reports and readouts, or can be dynamically operated by users on the fly. Some modules allow users to develop advanced equations using process data in order to solve a problem or optimize an operation. This empowers researchers, scientists, and manufacturing personnel to optimize their operations.

Data analytics, which in manufacturing applications is defined as using statistics to monitor an industrial process, has many aspects. In a *ControlGlobal.com* article, Greg McMillan and Stan Weiner describe just a few effective methods of data analysis such as principle component analysis (PCA) and partial least squares (PLS). “PCA is used for early fault detection and employs two statistics to explain if the process is within control limits.” Additionally, “PLS is used to predict the end-of-batch quality<sup>5</sup>.”

Another development in this arena is the proliferation of web-based visualization products. This allows “big data” to be harnessed by any user with a web browser and the proper credentials. Products such as Advantech’s WebAccess offer rich graphics and animation, a library of objects with connectivity to a variety of data sources, and reporting features (Figure 6). Browser access to these types of web servers ensures simplified deployment throughout an organization.

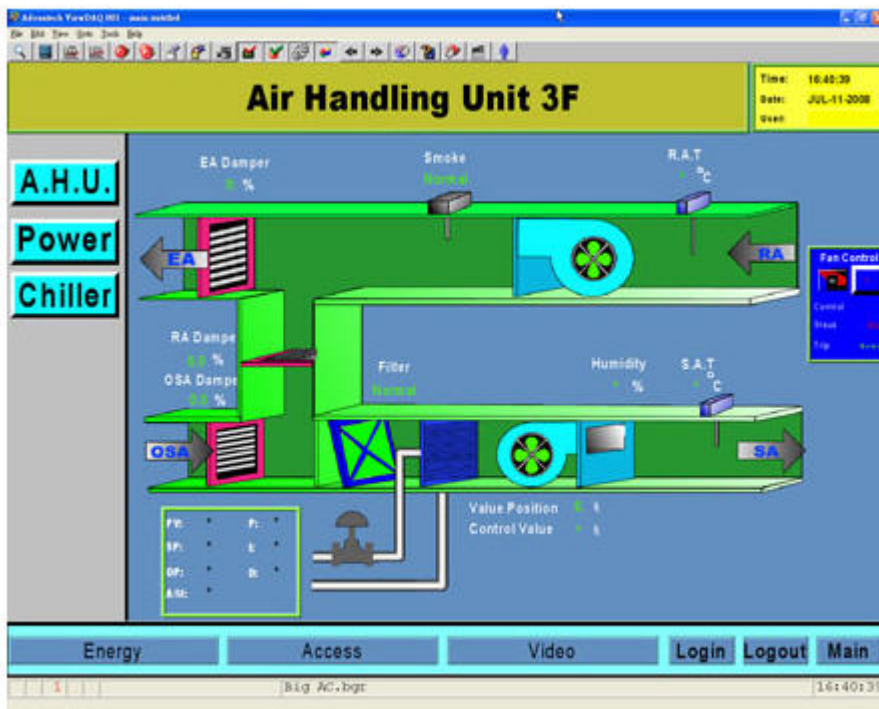


Figure 6 - Web-based visualization products allow big data to be viewed via any web browser, offering rich graphics and animation that help users interpret data into actionable information.

Some companies, such as Tableau Software, are taking this trend a step further with desktop and web-accessible software that promises drag-and-drop ease of performing data analysis. This effectively opens up the field of data analytics to everyday users, allowing them to “tell stories with data”. This type of innovation will enable a wider audience to exploit the IoT and big data productively.

And no discussion of visualization would be complete without touching on the expanding role that smartphones and tablets play in connecting to automation, historian, and database software. How fast

are mobile devices impacting the world? According to Joe Feeley in *ControlDesign.com*, “By the end of this year alone, the number of mobile-connected devices will exceed the earth's population<sup>6</sup>.”

Most visualization software products offer mobile modules that enable portable devices to connect to the host system over the Internet or an intranet. Mobile applications offer much of the same functionality as their big brother PC-based applications, but in a format appropriate to the device size. In some cases, users may find that their preferred operator interface experience is to use their own portable device wherever they go instead of booting up their laptop or seeking out a public PC.

A popular method of summarizing the results of collection, concentration, and curation is to create a simplified visual “dashboard”. Just as in a car, a graphical dashboard shows essential operating parameters of the system and warns of critical problems. There is much more information beneath the surface, but the dashboard distills the key elements to show the most important facts about what is happening, allowing users to quickly interpret conditions and make smart decisions.

### Conclusion

The IoT and big data are empowering extensive changes in the manufacturing and automation world. Writing for *Automation.com*, Bill Lydon indicates that European sources are calling this “industry 4.0”, indicating “that industry is entering a fourth industrial revolution” of higher technical integration. The internet of things is creating a bridge between the virtual and the real world<sup>7</sup>.”

Typical automation system designs have previously incorporated a top-down approach, with relatively limited control actions to maintain specific process variables. As the IoT and big data allows sensed information to become more rich and readily available, and as the capability has increased to capture and analyze that information, the ability to control processes becomes far more powerful and far-reaching.

For example, a traditional synchronous manufacturing process consists of all items moving down the production line and being handled identically. However, consider an alternate scenario where the item being produced boasts an IoT smart tag with memory, embedded with the requirements of that item.

This would enable asynchronous manufacturing, where the part/subassembly/batch carries its own information and drives the manufacturing process. In fact, German government agencies are working to formalize an object structure to facilitate exactly this type of “smart” manufacturing, and the IoT and big data are two of the most important tools for enabling this bottom-up interactive manufacturing approach.

IoT and big data are reworking the relationship of machines-to-machines as well as people-to-machines. Many new hardware and software technologies have been developed to bring field sensor information from the very edge of the process, to collect it in a distributed or centralized manner, and to curate it through databases and historians. Each of these data harvesting tasks is becoming more automated, which removes the elements of delay and error associated with manual readings and data entry.

Improving and automating data collection, concentration and curation enables end users to take full advantage of visualization and analysis software to make their operations more efficient. As key components of these activities, the IoT and big data are forces ready to be harnessed with today's products and systems to improve the efficiency, safety and profitability of manufacturers worldwide.

## References:

1. Big Data; Sanat Joshi; *ISA InTech*:  
<http://www.isa.org/InTechTemplate.cfm?template=/ContentManagement/ContentDisplay.cfm&ContentID=93360>
2. That 'Internet of Things' Thing; Kevin Ashton; *RFID Journal*:  
<http://www.rfidjournal.com/articles/view?4986>
3. Process Risk Assessment Uses Big Data; Ankur Pariyani, PhD; Ulku G. Oktem, PhD; Deborah L. Grubbe, PE; *Control Engineering*:  
<http://www.controleng.com/single-article/process-risk-assessment-uses-big-data/632b3ce8d25102b9ab558b3833cc5885.html>
4. 100 Most Creative People In Business 2013, Nate Silver; *Fast Company*:  
<http://www.fastcompany.com/3009258/most-creative-people-2013/1-nate-silver>
5. Drowning in Data, Starving for Information; Greg McMillan and Stan Weiner; *Control*:  
<http://www.controlglobal.com/articles/2010/AutomationData1003.html>
6. 'Internet of Things' Becomes Internet of Everything; Joe Feeley; *Control Design*:  
<http://www.controldesign.com/articles/2013/feeley-gigabytes-how-quaint.html>
7. Industry 4.0 Gaining Momentum; Bill Lydon; *Automation.com*:  
<http://www.automation.com/automation-news/article/industry-40-gaining-momentum>