Accelerating Business Value in Industrial IoT: Smart Data for Faster Time to Value

Sujeet Chand
Rockwell Automation

Industrial systems have data. Lots of it. It is in devices, controllers, historians, databases, industrial computers, and in closed proprietary systems. Data is continually streaming from a large number of sources. The promise of the Connected Enterprise is to transform this data into actionable insights to increase productivity and create new business value such as faster time-to-market, operational productivity, asset performance, and enterprise risk management. It is easy to amass Big Data in industrial applications. The time to derive value and the cost for storing and processing Big Data can be significant, and companies are discovering that unstructured Big Data is difficult to work with. For industrial IoT applications, we recommend a different approach to processing data for business value. Rather than collecting and storing every bit of raw data, we recommend that we start with the identification of the potential new business value, and leverage domain experts' knowledge of causality in industrial systems to contextualize and model the data that drives the business outcome. We can then match the appropriate analytics and data processing solutions to dramatically reduce time-to-value and the amount of Big Data. This approach leverages the continuum of computing resources available at the Edge and to the Cloud for data processing and analytics, thus extracting value in both the OT and IT domains.

1. Introduction: Business Value from Data in Industrial IoT

A popular school of thought is to gather every bit of data from everywhere and store it in a data lake or a database and then utilize AI, machine learning and data science to extract actionable insights. This approach is illustrated in Figure 1. An upside of this approach is that we may discover new insights from correlations in large datasets such as OT data and supply chain data in an enterprise. The downside can be too much data and too little insight. Data volumes can quickly get into petabytes (a petabyte is 10^{15} bytes or 1000 Terabytes). For instance, an oil and gas company is likely to have a large amount of data available from devices such as compressors that generate 500GB of data per day, and data volumes can easily exceed a petabyte in one year. To get an idea of how large a petabyte of data is, imagine downloading and playing one petabyte of MP3 encoded songs on an MP3 player – it will take about 2,000 years to play all songs. Also, there is a significant cost in storing petabytes of data.

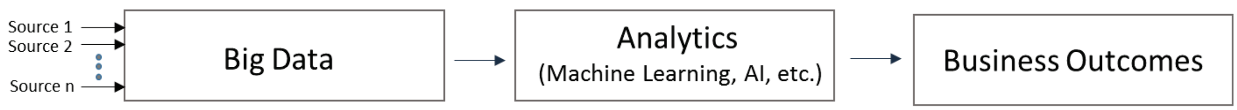

Figure 1. Industrial IoT: Common Approach to Extracting Business Value from Data

The Big Data analytics of Figure 1 is great for identifying correlations between data; however, it still takes humans to ultimately determine whether there's truly a relationship between two things and, most importantly, why such a relationship might exist. For example, research in correlations with Big Data at Harvard University showed a strong inverse correlation between the rate at which honey bees are disappearing in the U.S. to the number of Ph.D. degrees in Computer Science that are awarded in the U.S. This would mean that we need fewer computer science Ph.D. degrees to be awarded to increase the number of honey bees in the U.S.!

Results from industrial IoT pilots highlight the need for data aggregation and modeling in the OT layer to prepare the data for analytics. In industrial IoT applications, most of the systems are designed by people, and industrial processes such as pumping, packaging, sorting, filling, distillation, and brewing follow patterns of behavior that are known to designers and operators. We refer to such people who have knowledge of the machines and processes as domain experts. For many industrial applications, these experts can identify causalities in data.

We recommend changing the sequence of steps that are shown in Figure 1 to the ones in Figure 2 for industrial systems. By identifying the business value first, we can leverage the knowledge of domain experts to select the most likely data that drives the business outcomes and then match the appropriate data processing and analytics (machine learning, AI, rules, statistical processing, etc.) to gain insights faster than sifting through mountains of uncorrelated data. *The role of domain experts in identifying and contextualizing data in industrial systems for further processing through analytics is as important as the role of data scientists in industrial IoT.*

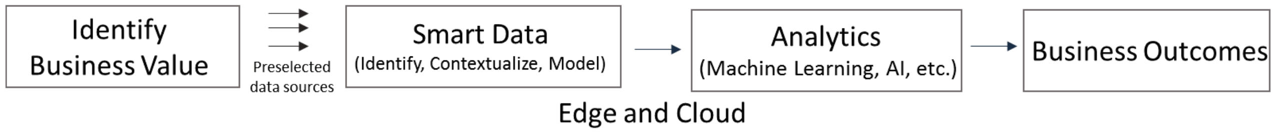

Figure 2. Industrial IoT: Approach for Faster time to Value from Data and Analytics

The key differences between Smart Data of Figure 2 and Big Data of Figure 1 are, (1) Smart Data is selected to match business outcomes (often with the help of domain experts), (2) Smart Data is contextualized, and (3) Smart Data is modeled and may include object-oriented models as defined in the next section. This approach is well suited to industrial systems that follow the laws of physics. Also, the analysis and actions that relate to real-time operation of production machines can be implemented at the Edge and acted upon at the machine level without having to wait for a higher level system to determine the necessary actions.

For stochastic systems with limited human insight into data correlations and causality, the Big Data approach of Figure 1 works best. We note that companies can use a combination of these two approaches based on the type of application and the business insights that they are after. Our experience is that subject matter experts usually know what data is relevant for industrial production problems. The challenge many times is to accurately sense parameter changes, create appropriate correlations between the input data and output effects, and develop a closed-loop control process.

2. Transforming Raw Data into Smart Data

How easy is it to form a sentence that delivers meaningful information from the letters, AACIIIMNNOPPTTUV? If these same letters were grouped into words “cavitation,” “pump,” and “in”, can we form a sentence? If we can have access to words, why do we want to deal with letters? To further complicate matters, there may be a lot more letters than the ones needed to form the sentence, “pump in cavitation.”

When processing the raw data from automation systems at the Edge or in the OT environment, we can organize and group raw data into “words” or Smart Data. Industrial systems generate data continuously over time, and this data is called time-series data. For instance, a drive will produce a continuous stream of data on motor speed. There are important trends in individual streams of time series data as well as aggregated/combined data. The data becomes more insightful and valuable when combined with time stamps and the operating state of the device or machine. Such organized and aggregated data is referred to as contextualized data. Examples of contextualized data include parameters of a device such as voltage, current, and power factor, linked with the operating state of the device such as running and faulted. Contextualized data can be created in Rockwell Automation's Logix controllers with special instructions (Add-on Instructions) and custom data structures (User Defined Types), to perform computations and create names and structure for the contextualized data. Such data is also referred to as Smart Tags. This results in discoverable information at the source that can be persisted as part of an application, and reflects the knowledge and insights of domain experts.

The controller recognizes these information data types and can be programmed to run real-time analytics with the contextualized data. *We refer to the combination of real-time control with real-time information processing as integrated control and information*.

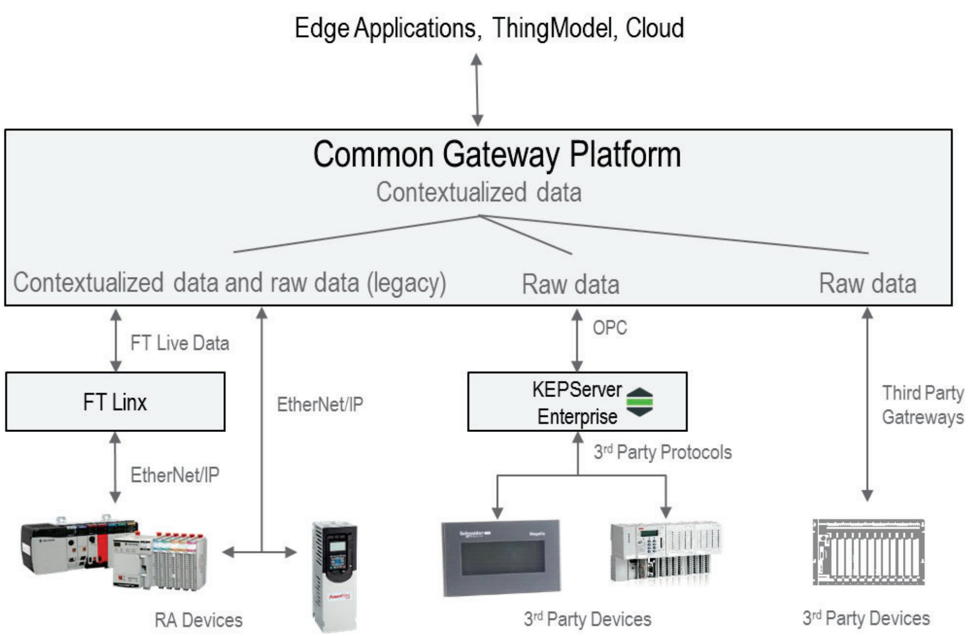

Figure 3: Transforming raw data into contextualized data at the Edge

There are many gateways available to access data from industrial assets. Gateways collect the data and egress the data as raw data, or, wrapped in an industry standard "wrapper" such as OPC as shown in

Figure 3. This streaming of raw data over time creates unstructured big data if it is not contextualized and modeled close to the source. Figure 3 illustrates the role of Rockwell Automation's Common Gateway Platform (CGP) to contextualize raw data from multiple sources. A software editor in CGP allows the naming, combination / aggregation of the data, and computation with the data to produce contextualized data.

The CGP transforms raw data from sensors, devices, machines and other sources into Smart Data for the following purposes.

(1) Create a common representation of the data. For instance, the data for Overall Equipment Effectiveness (OEE) which measures how well an asset is functioning compared to its full potential, can be in different formats from machines produced by different Original Equipment Manufacturers (OEMs.) The CGP allows us to "normalize" this data into common representations and/or add metadata to the data collected from machines. An example of normalized OEE data is the data produced by Rockwell Automation's RAPID™ templates. These templates contain timestamps and consistent reporting of quality, performance, duration, fault messages, and times in states (such as running and stopped) of machines.

(2) Organize and aggregate the data. Most industrial systems are modular and hierarchical, for example, Batch processing systems utilize a standard hierarchy of enterprise-site-area-cell-unit-equipment. Data representation and models can reflect this hierarchy. Each area in a plant can have many machines and devices. Data from devices that are connected to a machine can be aggregated at the machine level to produce contextualized data with meaning, such as machine uptime, throughput, energy consumption, and quality. Further aggregation of machine data yields cell-level and site-level data. Each level of aggregation produces important insights. These insights become possible due to the organization of the data by different levels of a plant's hierarchy. The CGP can name, process and annotate data by the system's hierarchy. This function is driven by what business outcome this data is going to serve.

(3) Time synchronization of the data. Data that is related to each other and drives a business outcome needs to be collected synchronously at the source and time-stamped to establish relevancy. For example, if our objective is to optimize the energy consumed by a mixer that is mixing cookie dough, we want to synchronize the start and stop times of the mixer with the power readings from the power meter to obtain relevant and accurate data for processing. Further analytics may require the association of the amount of dough, sugar, water, temperature, batch type, and batch sizes to

be correlated with energy, and all of this data can be time synchronized and structured in CGP for data processing and analytics.

(4) Process raw data. The CGP can create the OEE data from machines that do not provide such data, such as calculating the run time by subtracting the time that the machine was not running from planned production time. Another example of combining or aggregating data is combining voltage and current data from a Drive to provide power consumption data. The power calculation from all Drives (and other devices) can be standardized to a common representation since devices from different vendors may report current and voltage in different formats (e.g., volts, kilovolts).

(5) The Gateway can also be programmed to generate alarms and events that are triggered by limits on individual or combined data values. These notification data types are named in the Gateway.

(6) Automatic discovery of data. Since the Gateway normalizes and models the data, it can present data in a consistent model that can be automatically discovered by software applications designed to consume this data. This greatly simplifies acquisition and analytics with relevant data.

A well configured CGP and Logix controllers with structured data at the source (Smart Tags), can significantly reduce Big Data by transforming raw data into Smart Data, and eliminating the streaming and storing of all time-series data. It allows us to predefine and preformat data for consumption by models such as the "Thing Models" (higher level object-oriented models described in the next section) and analytics software. Also, the CGP integrated with containerized software and micro-services can add significant new value at the Edge in further modeling, persisting, and analyzing the data for business value.

Companies working with big data have realized that up to 80% of a data scientist's time can go into the unproductive task of data cleansing for analytics. *The CGP and Smart Data eliminate the need for data cleansing by transforming unstructured data into contextualized and structured data that is ready for analytics, dash-boarding, and mash-ups.*

We note that there are industrial assets and applications such as pipelines and remote pumping stations where the best solution is to stream data directly into the cloud due to the scarcity of local computing resources. Also, there may be legacy assets for which no models or domain knowledge is available. Such applications are well suited for machine learning with Big Data (Figure 1), and numerous software applications and tools are available to implement solutions.

3. From Smart Data to Data Models

Industrial plants have many common assets such as pumps, boilers, and compressors. Many such assets exist in a plant, and there are likely to be many types and models of these assets. To illustrate the value of data models, let's consider pumps, a common asset that is deployed widely in plants and fields. There are two basic types of pumps: dynamic and positive displacement. Dynamic pumps can be either centrifugal or special effect. Centrifugal pumps can be axial flow, mixed flow, and peripheral. A common type of centrifugal pump is an electrical submersible pump. A partial pump hierarchy is shown in Figure 4.

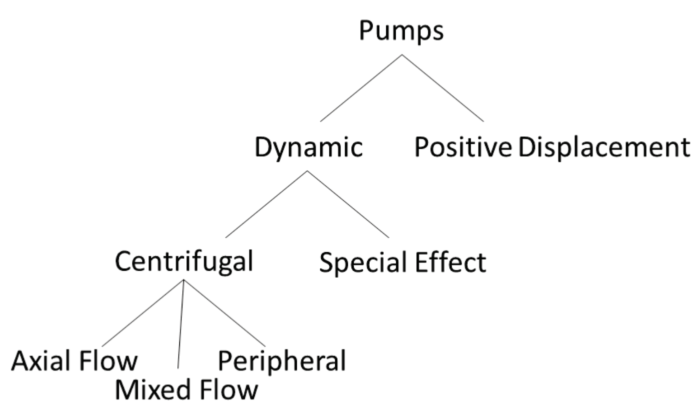

Figure 4. Types of Pumps

For each type of pump, we can identify the pump characteristics, attributes and relevant Smart Data such as pressure, flow rate, liquid volume, energy loss, and pump efficiency. Also, by utilizing a parent-child hierarchy as shown in Figure 4 and object oriented models, the data associated with each pump type is automatically "inherited" by the children in the hierarchy. Using an oil field example, there can be thousands of pumps generating data in a large oil field. The use of hierarchical models to identify the type of pump and the instance of the pump helps us organize the data for analytics, mash-ups, and dashboards. For instance, we can easily compare the efficiency of axial flow pumps to mixed flow pumps. Further, we have the option of selecting specific pumps for data analytics. By connecting Smart

Data to such object-oriented models, we now have relevant data from controllers and devices mapped to assets for rapid transformation into business value.

PTC's ThingWorx provides a world-class object-oriented modeling capability with the ability to quickly create dashboards and mash-ups through web-based applications that require minimal to no coding. Thingworx can be used to manage devices, implement response logic, and integrate third-party applications. *The combination of Smart Data and Thingworx significantly reduces Big Data, and provides the fastest way to get to business value from data.* Rockwell Automation's FactoryTalk Innovation Suite has capabilities for data collection, data orchestration, machine learning and data visualization.

4. Computing at the Edge and in the Cloud

The OT environment or the Edge has devices, machines, lines and industrial servers with significant processing capability. Transforming raw data into Smart Data and software applications for real-time analytics are ideally suited for the Edge, and enterprise level analytics are best suited for IT/cloud environment. We define real-time analytics as occurring in the OT domain (machine real-time) and resulting in *real-time actions* such as a change to a set point in a controller. Analytics in the IT/cloud level typically produce higher-level insights and dashboards. We refer to such analytics as non-realtime analytics. We note that non-realtime analytics are sometimes referred to as *human real-time analytics* (as opposed to machine real-time). For manufacturing plants with significant computing resources in the OT environment, a majority of data processing and value extraction is likely to occur in the OT environment. For field-based assets with limited compute resources, the majority of data processing is likely to occur in the cloud.

For simplicity, we will refer to three levels of data processing as shown in Figure 5: device, system, and enterprise. At the device level, simple limit checking can provide useful insights into a device's operation. At the system level, we can derive insights such as tension in a paper web that is likely to result in a web break in a few cycles. Such operational insights when implemented in controllers and industrial computers, enable real-time actions. Real-time analytics result from processing real-time data in the OT environment. At the enterprise level, selected contextualized data, or Smart Data, from the OT environment can be analyzed and combined with data from the other parts of the enterprise to develop data mash ups or dashboards that deliver actionable insights. This approach of locating data processing and analytics functions in a scalable manner is well suited to industrial IoT applications with varying time domains of processing (milliseconds, seconds, minutes, hours, etc.), asset locations (centralized, remote), and system relationships (autonomous, in-line, buffered, batched.) Analytics and AI/machine learning at each of the three levels can optimize processes and operations in industrial plants to deliver more productivity.

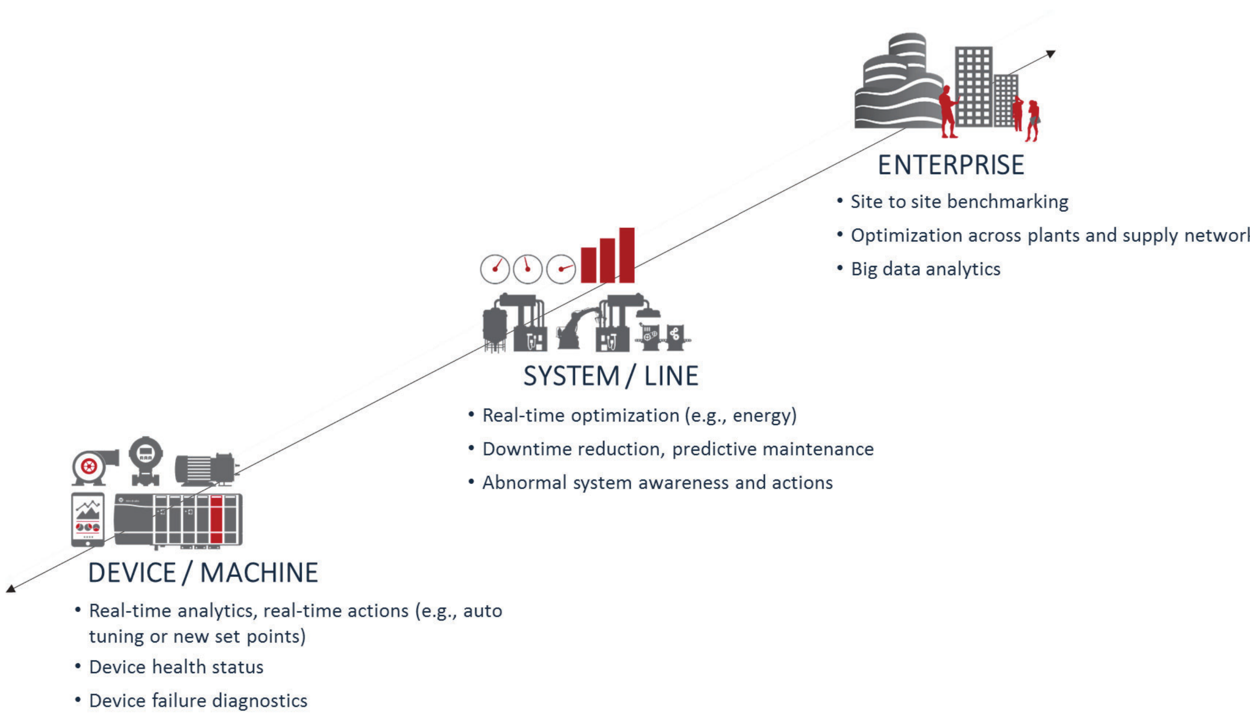

Figure 5. Scalable Processing of Data for Value

We note that cloud computing at the Enterprise level is a game changer. It allows us to leverage elastic computing and data storage for manipulating large amounts of data, running complex computations, and generating important insights by processing data from across the enterprise. Let's consider the use of cloud computing in smart automobiles today. A typical automobile has several hundred microprocessors where a significant amount of real-time data is processed for insights. These insights become available to the driver as real-time alerts, such as the vehicle needs an oil change or the engine is overheating. An unlikely scenario for the foreseeable future is to eliminate all of the on-board processing in a car, and utilize the cloud for all computing including processing safety features in the car such as airbags deployment. Cloud computing in automotive applications is well suited for higher level analytics and insights such as discovering issues in a specific vehicle model. This analogy extends to manufacturing also. Some proponents of cloud computing are contemplating that all data processing, control and safety will be moving to the cloud. We don't believe this is a likely scenario in the near future due to latency / security concerns. Cloud and "scalable computing" available in smart devices, controllers and industrial servers, provide a continuum of compute capabilities to pragmatically process data at each "layer" and deliver actionable insights that increase the plant's productivity.

5. Application of AI and Machine Learning

AI and Machine Learning hold great promise for extracting actionable insights from data. Application of machine learning at the controller / device level can help discover patterns in data to predict failures, and identify correlations between data. Modern machine learning technology allows us to discover patterns in data that can provide insights, such as informing us of potential failures. As a plant's equipment ages, discovering new correlations between data can provide valuable new insights. For instance, bearing wear is likely to change the vibration of a pump. With machine learning, we can correlate the vibration signatures to bearing failures.

The starting point for application of machine learning and AI should be the identification of the business problem. For instance, if our best opportunity for productivity improvement is to reduce the downtime of Line 5, we should focus on the Smart Data, data models, and the relevant AI/Machine Learning based analytics to predict potential downtime and implement mitigating actions.

Figure 6 illustrates the selection of Smart Data and analytics based on the business value. On the horizontal axis, we highlight four types of analytics: descriptive, diagnostic, predictive and prescriptive. At each of the three levels, device, system, and enterprise, the type of analytics to answer the questions in Fig. 6 can be different. For example, in the leftmost column of Fig. 6, if we need to know whether a pump is running okay, a comparison of the pump's real-time Smart Data with the pump's operating characteristics provides a descriptive answer such as, "the pump is running okay." AI and machine learning are not required for this type of information. However, if the question is can you predict when the pump will fail, we will likely use machine learning/AI to analyze real-time data to predict potential failure of the pump.

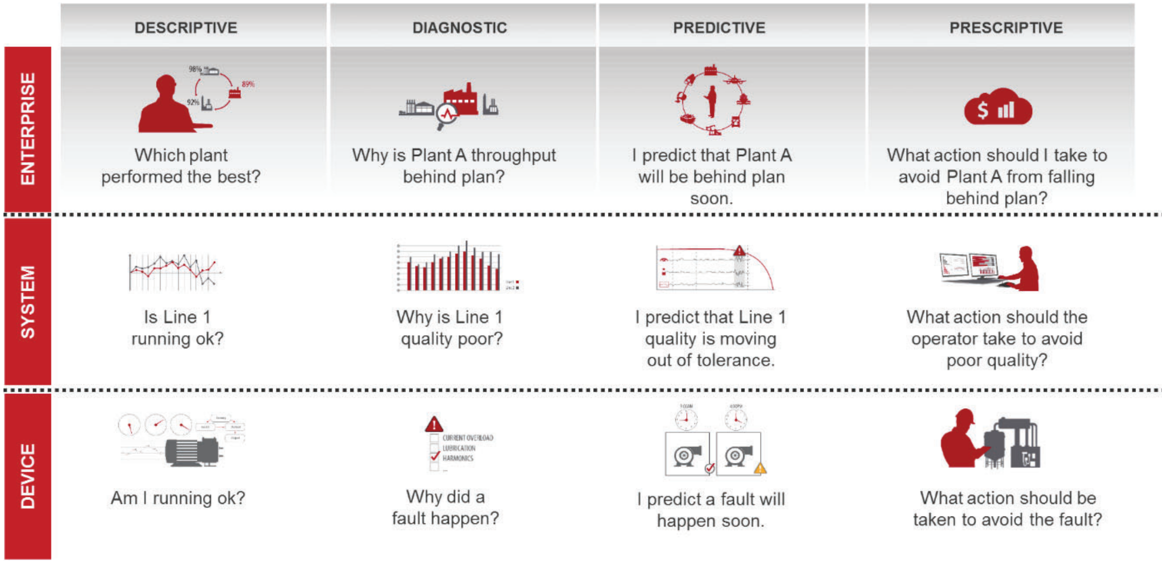

Figure 6. Business Problem Drives the Selection of Smart Data and type of Analytics

A review of the questions at each "level" in Figure 6 illustrates that it is not practical or cost effective to expect a *single* machine learning / AI / digital twin solution running in the cloud to provide answers and mitigating actions to all of the questions. The right approach is to develop solutions at the Device, System, and Enterprise levels utilizing Smart Data and data models that are matched to the question or business outcome. Solutions may include simulation, emulation, and AI/machine learning, and will deliver either real-time analytics or non-realtime analytics.

6. Summary

Big Data in industrial IoT applications will increasingly get replaced with Smart Data. Smart Data and models (Thing Models) provide a powerful framework to derive business value quickly and cost-effectively. Domain experts help identify and develop the Smart Data and Thing Models. Analytics, simulation and emulation software are selected to match the application and business outcomes. Also, industrial IoT solutions will leverage the scalable computing available at the device, system, and enterprise levels to implement solutions that deliver business value such as descriptive, diagnostic, predictive and prescriptive analytics at the Edge and in the cloud. There will be applications in Industrial IoT where human domain knowledge is unavailable, and such applications are ideally suited for the Big Data approach. Companies may use a combination of Smart Data and scalable analytics as well as Big Data for business outcomes.