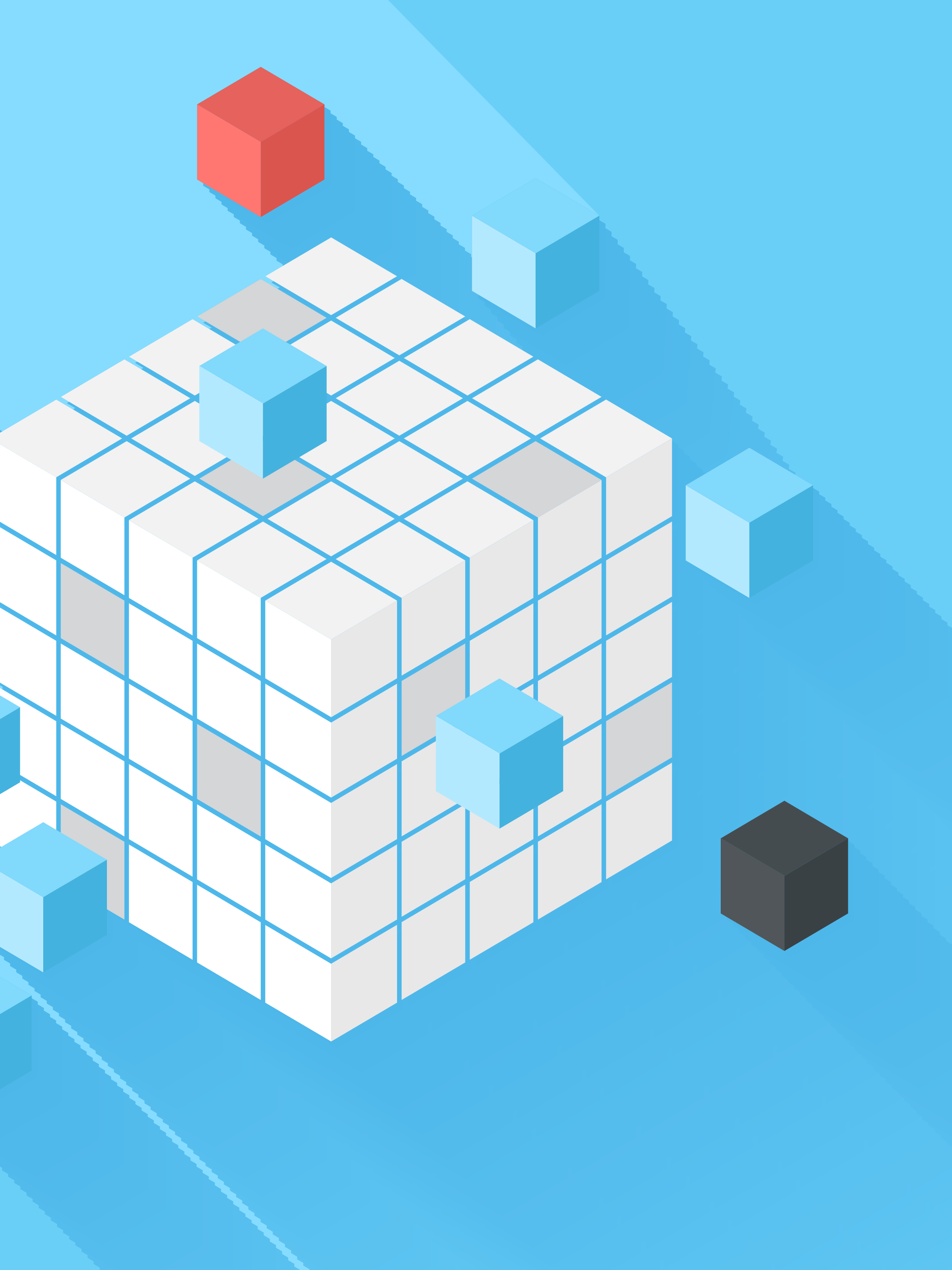
MANUFACTURING LEADERSHIP JOURNAL

Defining the Infrastructure for Big Data

Realizing the potential of data and analytics requires that manufacturers build a new global technology infrastructure, underpinned by a clear strategy, the adoption of an experimental mindset, and an investment in new skill sets.

By Steve Dertien



INFORMATION TECHNOLOGY IS REVOLUTIONIZING PRODUCTS, and, in turn, revolutionizing how manufacturers use IT. Once composed solely of mechanical and electrical parts, products have become complex systems that combine hardware, sensors, microprocessors, software, and various forms of connectivity. These "smart, connected products" and the data they generate have unleashed a new era of innovation and opportunity.[1]

Steve Dertien is Senior Vice President, Office of the CTO, at PTC. He is responsible for driving the company's technology strategy for PTC's Enterprise Products and maintaining vision alignment with PTC's customers. Previously, he served as a Technical Vice President with the PTC Enterprise Deployment Center. PTC is a member of the Manufacturing Leadership Council.

Smart, connected products, emerging across all manufacturing sectors, enable a new set of capabilities that can be grouped into four categories: monitor, control, optimize, and automate. The ability to monitor—via sensors and other data sources—exposes the product's condition, external environment, and operation. In effect, smart, connected products now have a digital voice and can exchange data with the manufacturer throughout the longest period of its lifecycle—the "use phase."

This wealth of new data, which serves as the foundation for all other product capabilities, will bring about a new standard for managing the product and service lifecycle and the customer relationship. Just think about how valuable it would be for manufacturers to stay connected to the products they develop and service every day. Instead of asking customers about product performance, manufacturers would gather design and quality insight from the product directly.

Manufacturers could provide more efficient service by knowing something was about to break instead of waiting for customers to tell them it's broken. Imagine the business growth manufacturers could drive if they knew how their product was being used, and were then able to deliver relevant and timely value-added services throughout the life of the product. On the furthest extreme, manufacturers might even provide their products entirely as a service, given the visibility and predictability provided by a smart, connected product.

The path to this value is through the aggregation and analysis of the following types of data:

- **External Data:** Third-party data from customers, partners, suppliers; the broader ecosystem, such as weather, commodity and energy prices; geomapping; and from news and social media sources, all of which informs product capabilities and analysis;
- **Enterprise Data:** Enterprise systems such as ERP, CRM, and PLM that provide data about customer preferences, sales and service history, and product details like engineering designs, warranty allowances, spare parts and inventory, suppliers, and costs;
- **Smart, Connected Product Data:** Data from product sensors, which provides insights about the product's condition such as location, temperature, and component or part failure, and data from the product's operation, including utilization, usage time and rate, and log files;

Smart, connected products change the Big Data equation for manufacturers. With product usage, condition, and performance data streaming from smart, connected products, data analysis—including enterprise and external data—can access a 360-degree view of the product and provide a richer perspective on customer needs and inform decision-making across the organization.

However, this opportunity is not without costs. Big Data is generally described as data sets whose large volume, wide variety, and high velocity make them impractical to process and analyze with traditional database technologies and related tools in a cost- or time-effective way. In order to extract the real value and promise of Big Data, practitioners must embrace an exploratory and experimental mindset regarding data and analytics.

This unique approach along with the global scale and distribution of smart, connected products changes both the infrastructure and technology choices manufacturers will be required to support. Many may be compelled to leverage cloud based infrastructures for the scalability, flexibility, and global reach and scale these providers enable.

For organizations with lightly-connected products—with low-bandwidth-modem or 2G-cellular connections, or remote devices that are intermittently connected—the challenges can be more profound. While one approach may require transitioning products to a higher bandwidth solution to leverage this new technology infrastructure, another approach may require adapting analytics capabilities directly onto the product. Adding onboard analytics will create a different set of challenges and considerations for modeling and implementing analytics within the limited hardware and software capabilities of the product.

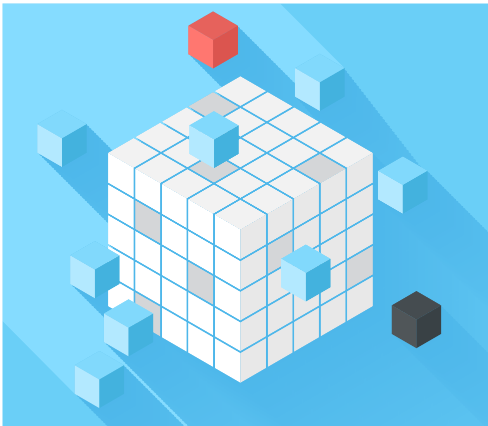
Manufacturing the Infrastructure

Building and supporting a new global technology infrastructure and capabilities for data requires substantial investment and a range of new skills—such as IoT applications development, data analytics, and Big Data management—that are rarely found in manufacturing companies.

So why invest? There are four categories of analytics that create business value opportunities for manufacturers across the enterprise. The complexity of the analytics capabilities required and magnitude of the business value opportunity increases across these four categories of analytics:

- **Descriptive:** Analysis of historical data that provides simplistic insight into past activities and performance to understand previous behavior and/or outcomes;
- **Diagnostic:** Utilization of historical data to identify a product failure pattern and determine the failure's root cause. Once the root cause is understood and diagnosed, the resolution can also be identified through mapping to knowledge management tools;
- **Predictive:** Use of modeling, data mining, and machine learning to analyze both real-time and historical data to predict and anticipate future events based on patterns found in the data;
- **Prescriptive:** Once a future event (outcome)

For organizations with lightly-connected products, the challenges with Big Data can be more profound than with other types of products.



MANUFACTURING LEADERSHIP JOURNAL

is predicted, a suggested next step and/or decision is identified, evaluated, and can be automatically enabled.

At the core of this technology infrastructure is the rules and analytics engine, which supplies the rules, business logic, and Big Data analytics capabilities that create value from data. Each of the three layers is a building block, creating some redundancy across layers. The infrastructure encapsulates the capabilities required for manufacturers to enable most Big Data use cases. (See Figure 1 below.)

The rules engine, not represented in Figure 1, serves as the business filter to express the factual conditions and regulations by which decisions should be made within the analytics engine. For example, an engine that emits an over-temperature alarm five times within an hour should execute a service diagnostic analysis, whereas five engines with the same condition in a one-month period should invoke a quality-diagnostic analysis.

In contrast, an engine that produces signals indicating low coolant levels and a rapid increase in ambient air, exhaust gas, and coolant temperatures with increasing cylinder pressure levels could indicate a predictive or even a prescriptive failure condition as imminent. The process by which analytics data are sequenced and organized can be used to guide a desired outcome or indicate when human intervention is required. Minimally, this alarm or signal could trigger a service event for a technician, or more substantially be used to alter and change the device behavior to avoid damage or defer service to a more opportune time.

1. Backend Infrastructure required for data acquisition: The bottom layer of the technology stack is responsible for the collection and storage of data from the multiple sources described above, including data governance and storage, data collection and integration, and SQL/noSQL data management. There are three types of data this infrastructure will need to manage:

- **Structured Data:** Data that is stored in a defined, fixed field within a specific record, usually including restrictions on types of data that can be inputted;
- **Semi-Structured Data:** Structured data that is not organized in a specific rational model,

Big Data Analytics Technology Infrastructure

Figure 1

or may not be organized in a recognizable structure;

- **Unstructured Data:** Data that does not have a predefined data model. This data is typically text-heavy and more complicated to analyze.

Managing the variety of data for analytics will pose technical challenges. While some organizations may rely on existing business intelligence solutions, others are looking at more analytics frameworks and approaches. Because of the volume and variety of this data, and the discovery-natured approach to creating value from Big Data, some firms are establishing "data lakes" as the source for their Big Data infrastructure.

A data lake is a data-acquisition approach that stores and holds raw data in its native format until it is needed and ignores how or why data is used, governed, defined, and secured by the organization. Information management leaders should understand the gaps in this concept—such as semantics, governance, and security—and take the necessary precautions.[2]

2. Analytics capabilities required to marshal and analyze the data: The middle layer is responsible for organizing and preparing data for analysis, and analyzing the data sets via processes and algorithms. There are many varieties of analytics capabilities, and depending on the specific use case, any one or combination of these capabilities may be required. Using an airplane manufacturer or airline operator use case helps simplify these technical concepts and highlights the unique and compounding nature of these analytics capabilities:

- **Enterprise Search:** Identifies what parts and components make up each plane, and the source of those components;
- **Log Analytics:** Identifies the log and fault diagnostics generated by the software and systems on the plane;
- **Natural Language Processing:** Interprets the comments and posts on social media regarding the plane based on flight information;
- **Statistical Analysis:** Identifies that 20 percent of a specific component that has broken on other planes in similar conditions;
- **Modeling, Simulation, and Forecasting:** Identifies how that specific component has previously yielded and correlates signals that can predict the likelihood when it will break
- **Optimization Engine:** Identifies opportunities to prevent the failure or improve the efficiency of repair when a failure occurs;
- **Machine Learning:** Implements a defined algorithmic model used to predict future conditions of interest based on prior or current information.

While not all of these analytics capabilities will be required for each use case, a wide variety of analytics capabilities and solutions are required to transform data into actionable insights.

3. Line-of-Business applications deliver the right information to the right user: Users across business functions need specific information to unlock business value, based on their role and purpose, and those needs will change over time. Within any business system are numerous pre-existing analytics capabilities, but until now many of these function with only human-captured data. The data and intelligence from smart, connected products can be combined to improve existing business systems and processes and enable entirely new processes across all functions in the enterprise.

- **Finance** can leverage product usage data to institute usage-based billing or performance-based contracts, and better manage the financial risk associated with warranty contracts by gaining visibility into product failure forecasts based on real-usage data.
- **R&D** can improve product quality by assessing failure modes based on actual product use cases, and can leverage product usage data to identify often- and rarely-used features to better invest development resources for future product revisions.

Managing the many varieties of data for analytics poses technical challenges, leading some firms to establish "data lakes" as the source of their Big Data infrastructure.

- **Operations** can connect disparate silos of manufacturing information into unified, real-time visibility across heterogeneous manufacturing systems, people, and assets to make faster and better decisions and improve operational performance.
- **Sales** can analyze product, asset, and fleet history for contract compliance and optimization, deferred and approaching service needs, product cross-sell and upsell, and consumable or replacement part opportunities.
- **Service** can automatically trigger service events based on real-time performance data and determine the best service response based on SLA entitlements and resource availability. It can also interact in real time with connected products to perform remote service activities including configuration adjustments and software updates to avoid downtime and eliminate need for on-site service calls.

In a smart, connected world, companies face new strategic choices related to data and the infrastructure necessary to leverage that data, choices that require tough tradeoffs.

Companies looking to rapidly and economically build and maintain the applications that enable non-technical users across business functions to create value from data will leverage an IoT application platform to meet the escalating demand. Conventional application development tools and approaches are not designed to work with the unique and evolving requirements of IoT applications and Big Data. An IoT application platform will streamline the development process and integrate the systems and people, as well as the data, in order to make application development and maintenance as efficient as possible.

Strategic Choices Unlock Business Value

The path to competitive advantage ultimately rests on strategy, and in a smart, connected world companies face new strategic choices related to data and the required technology infrastructure. Each choice involves tradeoffs, and each must reflect a company's unique circumstances.

A key question companies must address is what data must they capture, secure, and analyze to maximize the value of their offerings?

To determine which types of data provide sufficient value relative to cost, the firm must consider questions such as: How does each type of data create tangible value for functionality or efficiency in the value chain? How often does the data need to be collected to optimize its usefulness across business functions, and how long should it be retained?

Since data is increasingly fundamental to value creation and competitive advantage, some manufacturers choose to capture everything in a more-is-more strategy. Companies, however, must consider some key costs and risks for each type of data collected, analyzed, and stored beyond the infrastructure already discussed in this article.

First, the more data, the more complex and uncertain this process. Big Data analytics is highly process- and device-dependent and there is no "one-size-fits-all" solution. Second, there are hard costs from the additional embedded sensors, processors, and data transmission fees, and also significant security and privacy risks from amassing product data as product data clouds become new targets for hackers.

Third, many companies find the ultimate limiting factor is not technology, but the cultural and organizational change required to transform business processes based on new data insights.

Instead, a company should identify the data that reinforces its competitive positioning and creates unique value for its customers. A strategy focused on delivering optimized product or service performance must capture and analyze "immediate value" data in real time to reduce product downtime. This is especially important for complex, expensive products for which downtime is costly, such as mining or medical equipment, wind turbines, or jet engines. A strategy focused on establishing a product system or enabling a system of systems must capture extensive data across the ecosystem and multiple products.

Another key question companies face is whether they should develop a full set of smart, connected product capabilities and infrastructure internally or outsource to vendors and partners.

Developing and maintaining the new global technology infrastructure and capabilities requires significant investments that have not been typically present in manufacturing companies. Many of the capabilities required are scarce and in high demand.

A company must choose which layers of technology to develop and maintain in-house and which to outsource to suppliers and partners. In utilizing outside partners, it must decide whether to pursue custom development of tailored solutions or license off-the-shelf, best-of-breed solutions at each level. Our research suggests that the most successful companies choose a judicious combination of both.

As we have seen in previous waves of technology, early in the market we see vertically integrated solutions, where a single vendor is providing the entire technology stack. Over time we see specialization, just as Intel has specialized in microprocessors and Oracle in databases. New firms that specialize in components of the infrastructure are already emerging, and their technology investments are amortized over many thousands of customers.

A simple change in an analytics approach or business use case can result in a multifold increase in costs and time. Early movers that choose in-house development can overestimate their ability to stay ahead, potentially resulting in a slowing down of their development timeline. A better strategy is to focus on areas where, because of knowledge of the customer, product or processes, an advantage can be created over specialized technology players.

Conclusion

The technology and capabilities enabling Big Data analytics and machine learning are amongst the hottest growing technology areas fueling new business opportunities and innovation. By capturing and analyzing data during each and every stage of the product and service lifecycle, manufacturers can access the information needed to create competitive advantage, but this will require new skills, infrastructure, and cultural norms. The winners in this smart, connected world will be those who understand how to capture, analyze, and capitalize on these new streams of data. Those who don't risk placing their competitive advantage at risk.

1. "How Smart, Connected Products are Transforming Competition," Harvard Business Review, November 2014

2 "The Data Lake Fallacy: All Water and Little Substance," Gartner, July 2014

Those who understand how to capture, analyze, and capitalize on these new streams of data will be the winners in this new world.