

GE Digital

Unlocking Business Value Through Industrial Data Management



Authors

Main author: Shefali Patel

Contributing authors: Jeremiah Stone, Scott Duhaime and Venkat Eswara

Contents

- 03 The Opportunity and Potential of Industrial Big Data
- 05 Industry Mega Trends and Challenges for Big Business
- 08 GE's Approach: Maximizing Value from Machines and Enterprise Data
- 12 Rethinking Data Management—Control to Cloud
- 14 Customer Case Studies
- 16 Conclusions



The Opportunity and Potential of Industrial Big Data

The industrial network of connected machines with data analytics is creating a new frontier of competitiveness for information-centric modern organizations. We are citizens of a data-driven century, in the early stages of a digital industrial revolution. Orchestration of digital machines will lead to unprecedented change in the way companies perform, compete in their industries, and unlock new value from their data and business models.

Technological progress in information systems, broadband, communications and artificial intelligence is generating enormous wealth across industrial and consumer sectors. Advancements in cloud

and in-memory computing, distributed databases and data science, along with proliferation of sensing and controls, have given rise to "big data"—fundamentally changing the ways in which individuals, organizations and machines interact with their intelligent environments.

Convergence of these physical and digital improvements has paved a way for the Industrial Internet, driving efficiency gains across the production value chain.

At the heart of this revolution is an invaluable resource: data that can be converted rapidly into insights, leading to smarter business decisions and increased automation.

Based on research from the Center of Digital Business at MIT's Sloan School of Management, data-driven decision-makers (DDD) demonstrate 4% higher productivity, 6% greater profitability and 50% higher market value from IT, according to Eric Brynolfsson. Data-driven decision-making also improves other performance measures such as a) return on assets, b) return on equity, c) asset utilization (output per total assets) and d) market value (market-to-book ratio).

"Information is the oil of the 21st century, and analytics is the combustion engine."

— Peter Sondergaard,
SVP, Gartner Research

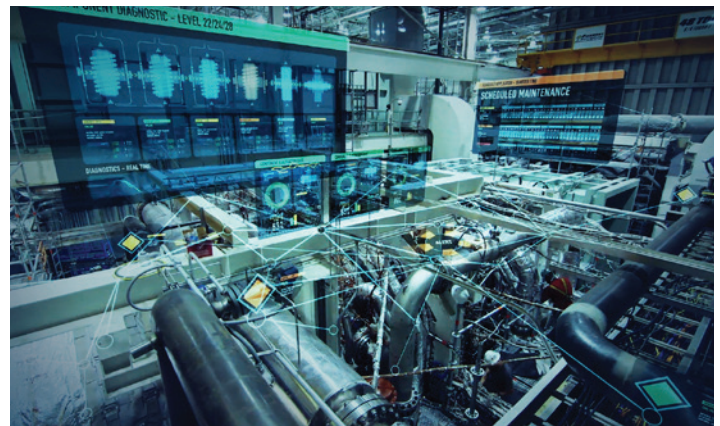
With lower cost of sensors, we can now measure many things that we could not before, and therefore control our technology more precisely and less expensively than ever before. Utilization of Internet technologies has made it easier to access data. Data can help us make better predictions and take smarter actions. We can be collectively objective, rather than individually subjective. We can do so in areas where we formerly acted based on intuition and assumption rather than by data and analysis. As big data

tools proliferate, they are changing how we behave, automate machines and reap benefits of digitization.¹

However, despite the promises of big data, companies struggle to exploit its value. Why? Abundant data by itself solves nothing. Its sheer volume and variety exceeds human capacity to configure it efficiently. Inherent challenges tied to evolution and integration of information and operation technology make it **difficult to glean intelligence from unorganized data, compromising digital literacy.**

At GE, we are developing a better way—a better way to manage industrial big data that triggers insights. We are in the early stages of a long journey of discovery and invention, taking a longer-term view to strategic data management and its technologies that translate to business advantage.

Abundant data by itself solves nothing.



What if invisible insights into your business became visible? What if terabytes of data pulsing through your operations were captured and stored securely in the field and in the cloud so that it could be accessed in real time? Imagine an enterprise world with a single source of truth—a panoramic view of your entire fleet and operation with the ability to zoom into powerful telescopic detail in a cost-effective way, and a workforce focused on resolving issues, innovating and collaborating across silos.

This paper is about Industrial Data Management: its challenges and opportunities, with a central focus on foundational technologies that lay the groundwork to win in the future of the Industrial Internet.

Industry Mega-trends and Challenges for Big Business

Before we discuss industrial big data solutions, let us review some of the mega-trends that are impacting the state of the market.

They create both challenges and opportunities for progress, while shaping our choices for technology investments. These major market trends fall into three different areas, all relevant to digital transformation:

The rise of big data and the Internet of Things

With the cost of computing, bandwidth and sensors decreasing multifold in recent years, there has been an explosion of embedded devices that can communicate with one another and churn out volumes of data. big data can be defined as high-volume, high-velocity and /or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision-making and process automation.²

The physical world is being digitized. Smart objects linked through wireless networks that carry information are forming a system called the Internet of Things (IoT).

The basis of IoT, the real-time dynamic analysis of data, is challenging business models built on static and rigid information architectures. We need fluid business models based on agile software platforms to catch this technology wave and maximize its value before it diminishes.

Disruption of industries

From domains as diverse as entertainment, retail and transportation, digital innovations have disrupted industries all around us. Missing the digital beat is falling prey to disruption. Disruption happens slowly, **as centralized incumbents get displaced by nimble companies that have commercialized cheaper, more convenient and widely available digital technologies to underserved consumers.** It creates new markets.

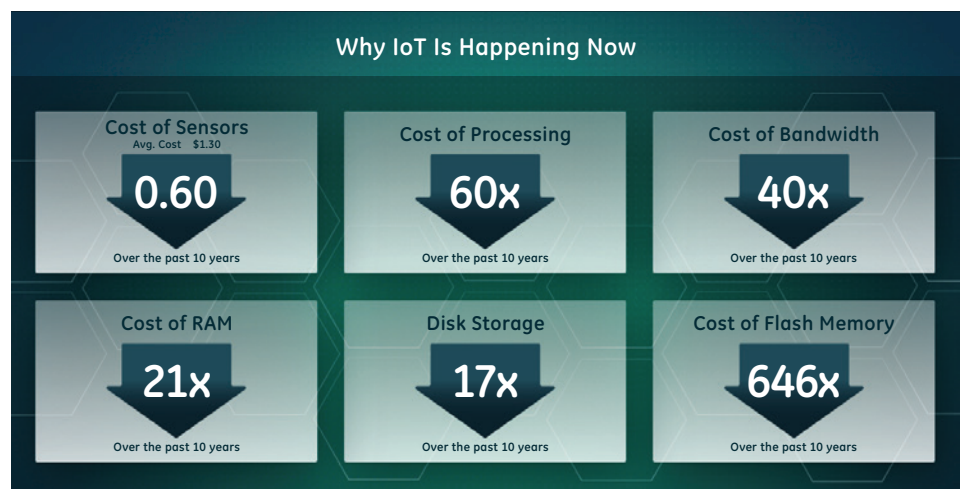
At GE, we have disrupted ourselves by infusing an ethos of entrepreneurship, reinventing our business model, leading the practice of the Industrial Internet for our customers, and creating GE Digital.

The future of work

The nature of how work gets done today is also evolving with demographic, globalization and technology changes. The composition of today's workforce is ever so diverse, with digitally savvy millennials working side by side with more experienced knowledge workers.

As IoT technologies further penetrate the workplace, they will create new ways of organizing work and dispersing knowledge across distributed teams.

With virtually every device in the workplace emitting data, it will take new ways to manage data-driven organizations and get the most of out of our human capital.



Source: Goldman Sachs Global Investment Research, John C. McCallum Research, TCG Advisors

Industry and Customer Challenges

Taken together, these trends are carrying us forward. They are also creating new challenges for business leaders. We have grouped them into technology, business and organizational challenges.

Technology challenges

Islands of disparate data

Keeping up with a flood of information is difficult. Most companies struggle with data deluge driven by lower cost storage, sensing and communications technologies. But a few have figured out how to exploit their data. big data that is neither **structured nor contextualized is strenuous to cost-effectively store and analyze in its entirety** through traditional computing approaches.

What causes data islands?

Data islands are created as a byproduct of operational and project moment-in-time decisions not made in the context of a larger data strategy. Layering of legacy systems conjoined with newer technologies and lack of governance for data systems also results in **data-related islanding of internal departments and work groups. It can lead to limited purview and inhibit collaboration.** Data gets siloed, whether it is enterprise data, equipment data inside an organization, or data across different organizations.

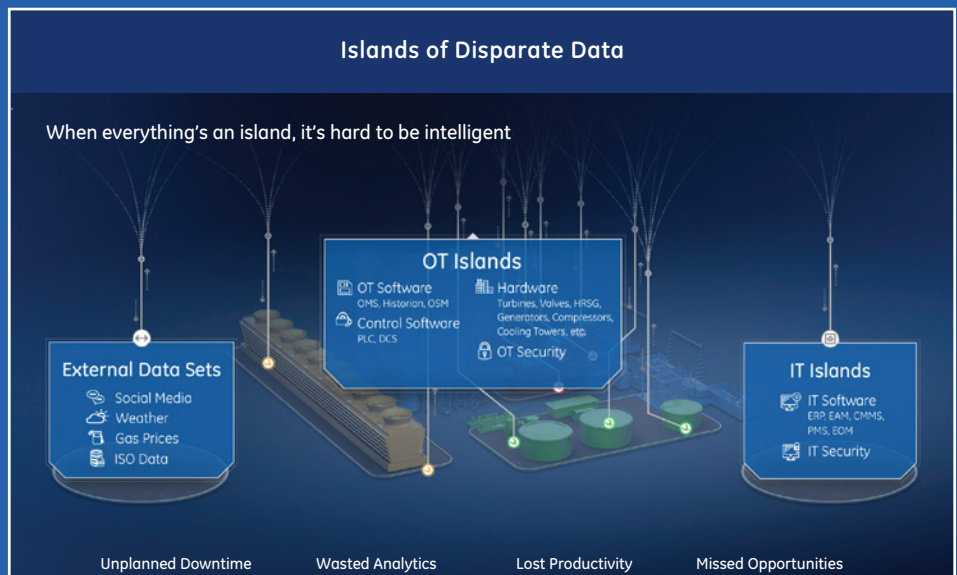
This fragmentation makes data discovery difficult and presents complex technical and organizational

challenges. When the data is scattered throughout plant and the enterprise, integrating and analyzing it manually becomes resource-intensive and tedious. **This has an opportunity cost. By the time data is organized, its value may have been lost and the personnel too fatigued to derive any insights.**

To extract meaning and value from data, new systems are required to handle the challenges posed by the volume, velocity and variety of these big data sets.

Technology integration

The management of industrial technology has traditionally been split between two separate fields: information technology (IT) and operations technology (OT). IT worked from the top down, deploying and maintaining data-driven infrastructure largely to the management side of business. OT built from the ground up, starting with machinery, equipment and assets, and moving up to monitoring and industrial control systems. With smarter machines, big data and the Industrial Internet, the worlds of IT and OT are converging. Traditional enterprise data management, such as ERP or CRM, is being dwarfed



by operations data due to sheer volume and variety. But most of this data is still in the dark. **IT and OT, developed separately with independent systems**

architectures, need to come together and find common ground to develop a new infrastructure.³

Business challenges

New sources of revenue and profitability

With a highly volatile market environment and costs of maintaining aging infrastructure, companies are continually challenged to sustain their profitability by finding new sources of revenue. **Manufacturers are seeking ways to lower capital expenditures, and they need a single source of the truth** to help them make the right decisions for improved performance, while mitigating risk from unexpected incidents.

There is a need to link analytical systems to operational systems. Today, most business analytics do not support any connection back to the originating systems of the data. Analytics are on an island, as well, inhibiting the ability to take action in a reliable and effective way due to the onus on the individual to connect the worlds in their brain and connect the systems and workflows via their own initiative

Asset-level visibility

Improved capacity utilization is one of the great benefits of state-of-the-art information systems. To achieve production targets, operators need

to be able to monitor assets in real time and ensure all assets (across all plants) are performing at an optimal level. They need increased visibility and better insights that can be acted upon. This enables them to detect anomalies and fix issues before they occur, yielding no unplanned downtime.

Asset Performance Management and operations optimization software can provide operators with answers on what equipment is most important, how it should be maintained and how unexpected failures can be avoided.

Organizational challenges

Aging workforce and knowledge capture

Aging of the workforce is impacting a number of industries. Retirement of experienced workers is expected to create a skills gap. While younger generations of workers will bring new skills, **it is crucial that the knowledge and experience accumulated by more senior workers is captured and made accessible to the new workforce.** Inability to institutionalize this knowledge can be detrimental to the apprenticeship

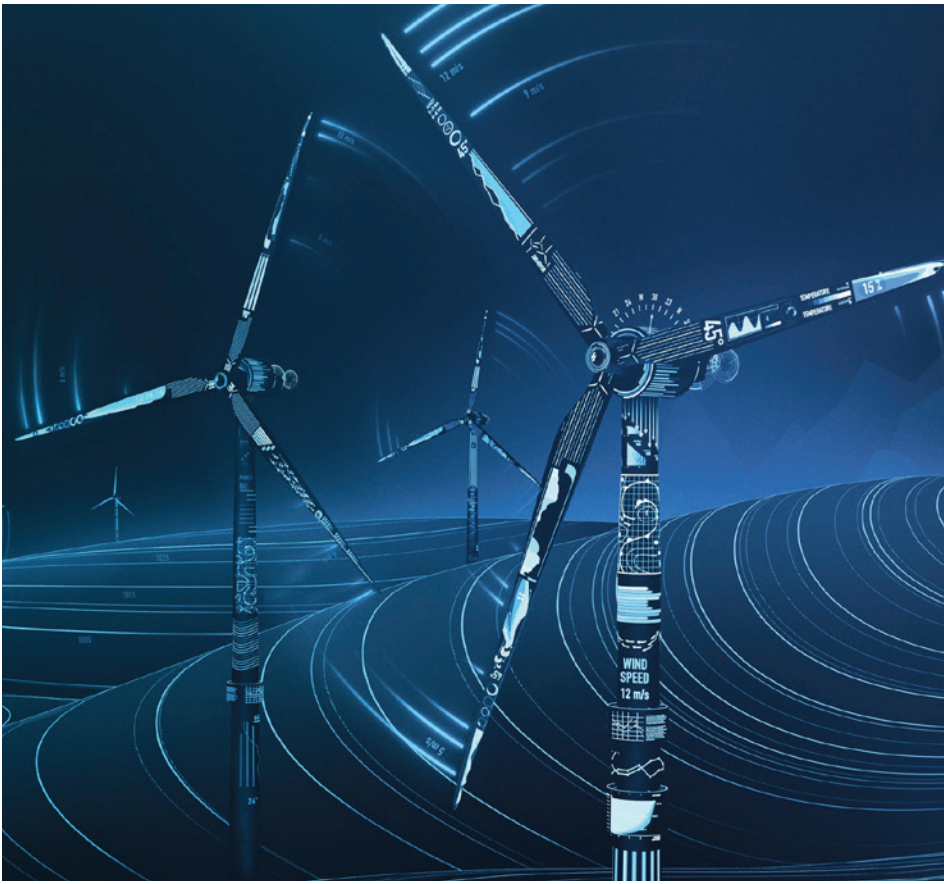
of younger employees. Preparing for this impending change by using digital technologies can ease the transition.³

Cyber security

As billions of assets get smarter, network and store information on the cloud, they will be exposed to digital privacy risks. Just in the last year, there have been several cases of data breaches causing significant damage to all parties involved (companies and their supply chains, as well as consumers). Cyber

attacks pose a range of threats—from personal devices to corporate IT systems—making individuals and institutions vulnerable to financial and physical harm.

There is growing awareness and paranoia among stakeholders, and an urgency to mitigate these risks. Vendors are deploying solutions to prevent these cyber events and protect against digital crime. As we invest in digital technologies, cyber security capability must be part of the selection criteria.



GE's Approach: Maximizing Value from Machines and Enterprise Data

Industrial Internet technologies can turn the challenges discussed earlier into opportunities for improved productivity.

Many industrial companies have already started their digital journeys towards Industrial Internet maturity. Technologies including enterprise data management and predictive analytics that we have been deploying for our customers are now seeing double-digit performance gains across the following sectors: power generation, oil and gas, transportation and health care.

Reaching maturity involves five stages with corresponding technology components that allow an enterprise to connect, monitor, analyze, predict and optimize their assets and operations.

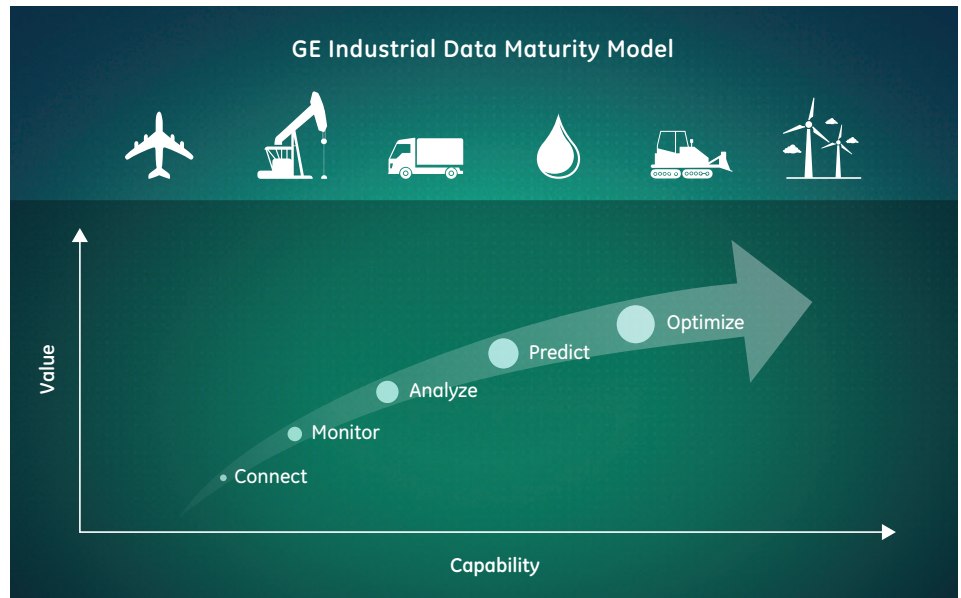
This can be accomplished on the premises and in the cloud, depending on the current configuration of your data architecture.

5 stages of asset and operations maturity: connect, monitor, analyze, predict and optimize

The first step of the Industrial Internet maturity model is to connect all critical assets across the enterprise. This is not a trivial task as we are referring to hundreds of discrete instrumented components with different communications, networking protocols and underlying data formats, all generating terabytes of data. **It requires a highly scalable and fault-tolerant software system that enables cost-effective data storage, data visualization and analytics. This foundation of the Industrial Internet starts with data management.**

Historically, it was possible for a single machine or a handful of machines to receive and store data. However, the growth in sheer volume of devices and sensors coupled with the desire to perform rapid data mining on larger amounts of historical data (requiring that they be kept in memory or on a disk) necessitated the development of systems with new technical approaches. **Enterprise Historian, designed and built by GE, addresses this challenge in a cost-effective way.**⁴

To illustrate this point, let us first think through how industrial data becomes industrial big data. As evidenced by figure 1, a single machine in the manufacturing of baby care products, which generates 152K data samples per second, provides 4 trillion samples a year. Added to this are other streams of data such as geolocation (GIS), alarms, and maintenance and shift logs. big data systems that support handling of increasingly fast and large volumes of heterogeneous data can replace the



traditional archive-and-ignore model of data management—and at a lower cost. They make it possible for structured and contextualized **data to be online and available anytime for analysis and mining. They not only give data scientists access but also a visual representation of what is happening to an asset, making it easy to spot unusual conditions.**⁴

Data historians are special-purpose database applications. They are designed to **efficiently store and analyze large quantities of time-series data with ultra-fast read and write performance.** They are capable of storing up to millions of data points per second, capturing streams of data in real time from sensors located across a manufacturing facility, power plant or other such sensor-rich

environments. **While very efficient, most data historians in the marketplace are single-server solutions, limited by the memory and disk capacity of a single machine.**⁴

Our next-generation historian efficiently stores and performs historical analysis on hundreds of terabytes of time-series data. It is seamlessly scalable in how much data it can store, with the flexibility to address:

- Consistent data collection, storage, federation and data modeling across asset, site, plant and enterprise
- On-premise execution with the ability to manage data at rest in the cloud
- Fast data modeling with data discovery models to process a variety of data sets (structured, semi-structured and unstructured)

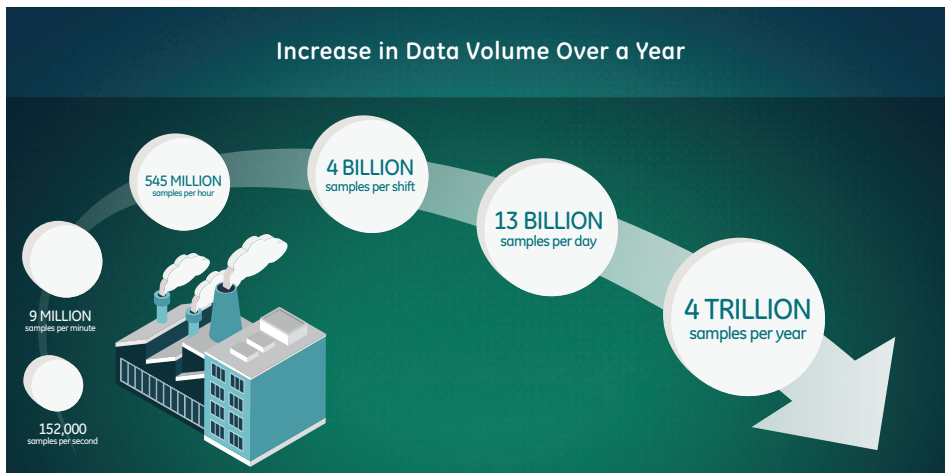


Figure 1

Next, let us review underlying technical capabilities and the business outcomes of data historians.

Data collection

Data historians collect time-series data in real time from distributed, disparately structured sources. With GE's Historian, the first value customers have realized is through simplifying data collection and aggregation functions. A mix of redundant, fault-tolerant collection options that can be paired with sites at different levels of automation, coupled with the ability to embed pre-handling and calculation functions enable a more flexible way of maintaining data flows to the central operational data store (ODS) with lower demand on IT and engineering resources.

Data storage

Our customers are realizing value in the way GE's Historian stores data. With no compression at all, it offers much higher disk space efficiency than a relational database (RDB). When using a 1% dead band compression, it delivers even greater efficiency for enhanced performance and reduced maintenance. Efficient data storage and compression enable high performance and minimize maintenance.

Why is this important? Consider the big data example of the baby care line that collects 152K samples per second. Uncompressed, this data could reach 40 TBs annually. With standard GE Data Historian compression, this is easily reduced to 4 TBs. At a cost of \$5,000 per terabyte, your storage budget for this one machine shrinks from \$200,000 to \$20,000. In addition, once the data is in

Historian, it cannot be edited or changed (except during a re-run of data collection). It is secured in a highly trusted enterprise data store.

Data services

A central enterprise data store built on near-real-time data opens up new opportunities to use current data in support of a range of systems and procedures. A key part of Historian's enabling technology is a **services-orientated architecture**. It provides a set of data models that can be adapted to reflect common types of assets (and related events, personnel and materials) deployed across various operations. Source data elements from existing automation and systems can then be mapped to the items defined in the model. Users commonly take advantage of this modeled Historian data to:

- Feed maintenance systems with asset usage data
- Deliver comparative views of asset and process health across like assets or installations, regardless of differences in underlying automation and systems
- Use data and calculations from Historian as the basis for triggering a range of corrective or keep-running actions, delivered through electronic work instruction systems like Workflow, or even as outputs back out to existing systems (SCADA being most common)

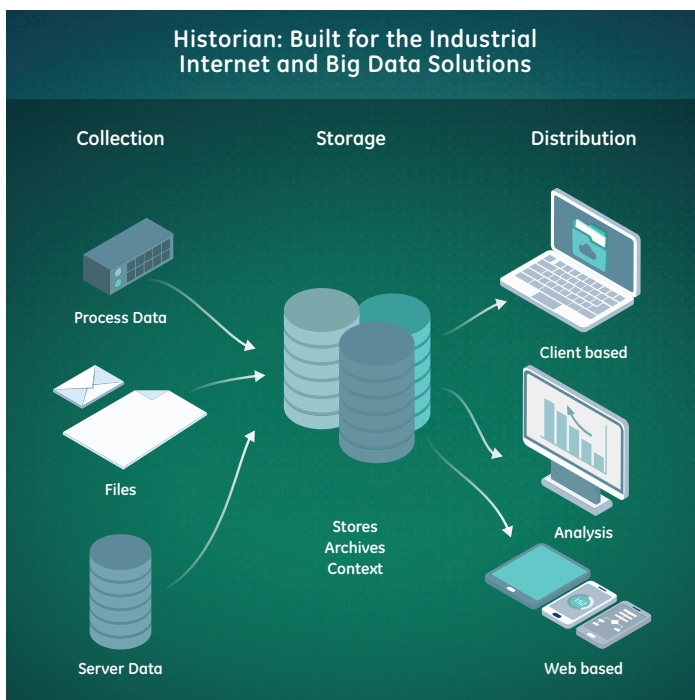


Figure 2

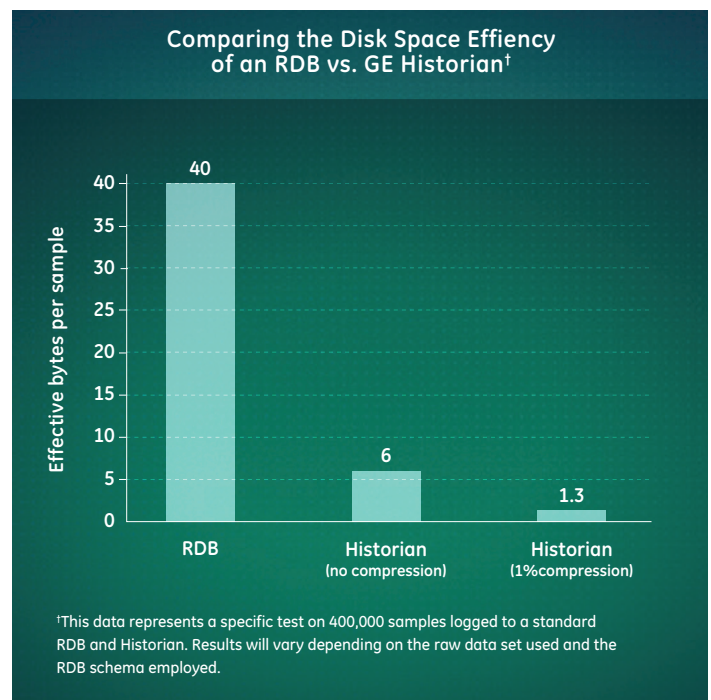


Figure 3

Data access: On-premise to cloud

Another two considerations for IT leaders are: a) the cost of moving large data files across the network and b) long-term data storage requirements, specifically determining where the final resting place of data should be.

GE Historian solves for both these problems.

Per our previous discussion on data compression, consider moving uncompressed archives of 10 GB or more across various networks, say from a control network, through a firewall, into a DMZ (demilitarized zone) or a business network, then again across another firewall boundary, into an external facing network or public Internet for cloud storage. Then consider the 10x compression of GE Historian and the impact to your network of transferring 1 GB of archive files.

Beyond efficiency, **GE Data Historian has a number of native ways to move the data.** From one Historian to another, one can use a server-to-server collector for streaming data. **Moving from premise-based Historian to cloud-based time-series open source, Historian offers a server-to-cloud collector.** To move larger files, one could use the Historian HD ingestion service to move the tag configuration and archive data onto a Hadoop Distributed File System (HDFS) data lake on-premise or in the cloud.

This also allows IT managers creative options for long-term storage. For instance, one customer, who keeps 30 days of historical data at the plant level, aggregates all plant data onto a central Historian temporarily, and then moves all this data to an HDFS cluster for their “data at rest” strategy. This approach has financial benefits. A terabyte that costs \$5,000 to store in a traditional manner may cost \$1,000 in an HDFS cluster. With the server-to-cloud collector capability, the costs of highly compressed GE Data Historian file drops from \$20,000 to \$5,000 in HDFS.

Time to value: From calendar time to watch time

Within the Enterprise Historian, the time-series data is stored in an HDFS. **The near-linear scalability of Hadoop allows our Enterprise Historian platform to scale out as the volume of data grows over time.** Up to 20+ years of industrial big data generated from the installed base of equipment can be stored online and mined on demand, replacing months of manual effort, to explore much smaller data sets.

Industrial Data Management is the foundation for the Industrial Internet and goes beyond historian time-series data. To effectively and efficiently deliver next-generation analytics and applications, Industrial Data Management uses time-series data generated from machines and equipment in conjunction with structured data such as enterprise resource planning (ERP), customer relationship management (CRM), geolocation, semi-structured data such as machine logs and digital inspection data, and unstructured data such as content (images, manuals, video, etc.). This is critical to bringing a complete contextual and situational analysis for the assets. In addition, it becomes the basis to empower personas beyond process engineers, such as data scientists and enterprise operations, to unlock the value using advanced data management and analytic capabilities.

Rethinking Data Management

Data value chain from control to cloud.

We are learning a lot along with our customers as we redefine GE as a data-driven business. Beyond just assets, our customers are looking at how to manage their data across the enterprise and run analytical queries that provide immediate tangible insights to operations and improve business metrics. To achieve better business outcomes, digital industrial companies like ours are rethinking how they manage the data in the field in order to maximize the value of the analysis in the cloud.⁵

Figure 4 below depicts how data management capabilities can traverse the enterprise—single-site, on-premise installations to multi-site, cloud-based solutions. The cloud-based solutions go beyond semi-structured time-series data to include structured transactional system data, and unstructured web and machine data, merging these data streams into a single view. Evolving the use and capability of data historians to be cloud-based, is central to rethinking data management.⁵

Critical to this journey is to bring a wide array of data together and relate it into a common structure such as a data lake in Predix™, our cloud platform for the Industrial Internet, as shown in figure 4. Business-level analyses based on statistical models can then be performed with Enterprise Historian solutions as the foundational data layer,⁵ as shown in figure 5.

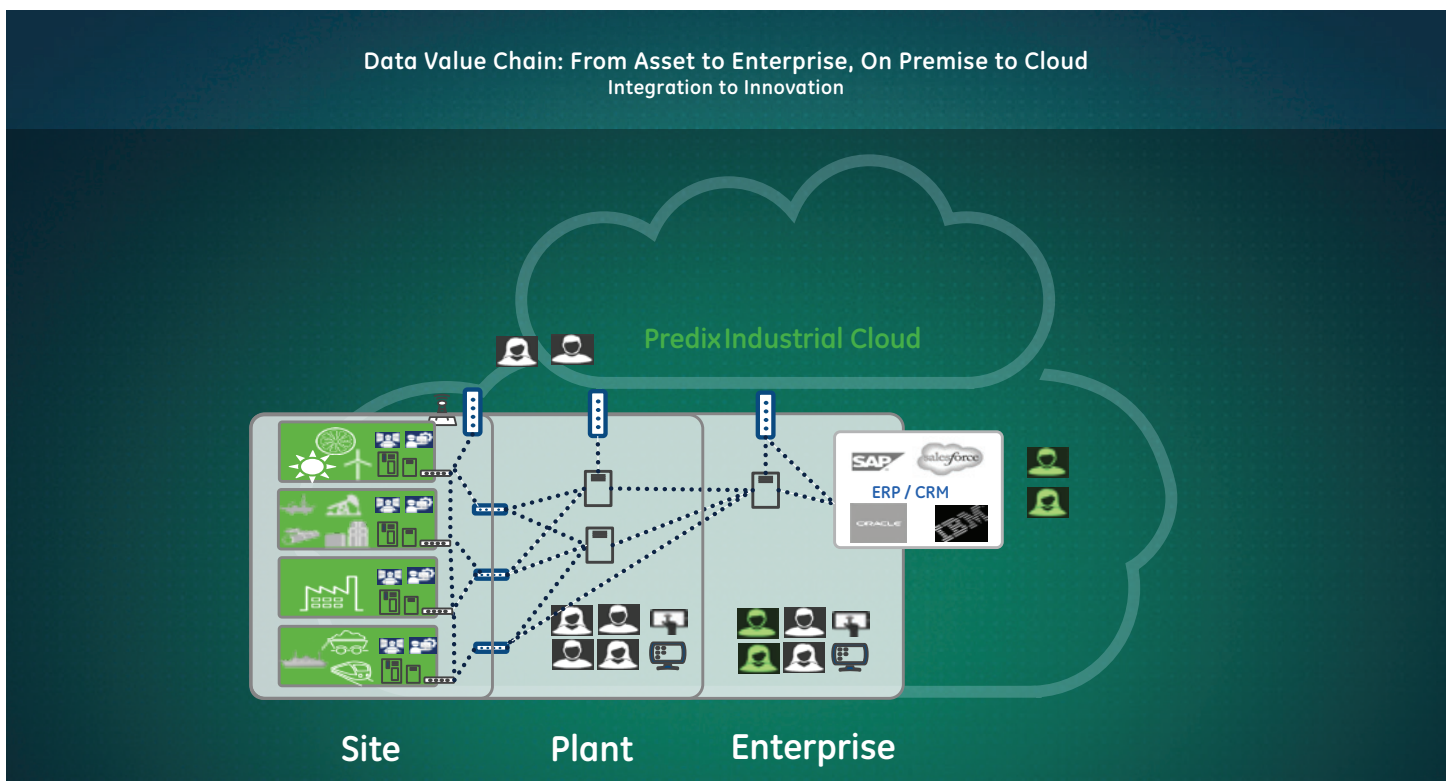


Figure 4

This data set can be looked at as a flow horizontally across the enterprise and into a cloud-based environment, where that data can be managed in much more cost-effective ways. A key dimension in control to cloud is integration with systems in the control environment at the asset layer, with operational systems and with enterprise systems. This is part of the reason why the horizontal flow is important, as well—a single source of data across layers of responsibility, rather than competing sets of authoritative data.

Industrial companies desire to have a common data management capability. At GE Digital, we are finding that our historical capabilities in control systems and data management and processing capabilities in the field are complemented very well with our capabilities in the cloud. We are starting to design this data flow holistically with Predix.

Outcomes from the data value chain approach — from the control system to the cloud — include:

- A dramatic **decrease in cost** by managing operational data more effectively through a decrease in cost of IT infrastructure and operations
- An **increase in speed of deployment**, speed of development of new analytical queries and applications. A unified, systemic approach that results in repeatable speed and time to value
- A **decrease in complexity** with focus on optimizing at an infrastructure level capability within the field

Companies that can leverage this data value chain horizontally across their businesses, from their machines to the cloud, will reach new levels of efficiency and business performance. There are other possibilities starting to emerge, as well. As we start to facilitate a greater flow of data and the right data into the cloud, the quality of our analyses improve along

with our ability to ask “what if” questions. We are able to model our assets and drive **predictive analytics using our SmartSignal technology**, most notably in the area of early anomaly detection in asset performance. We are able to see deviations from normal or desired asset behavior long before they become visible to standard operational systems. Continued maturity expands into operations optimization and business optimization capabilities.⁵

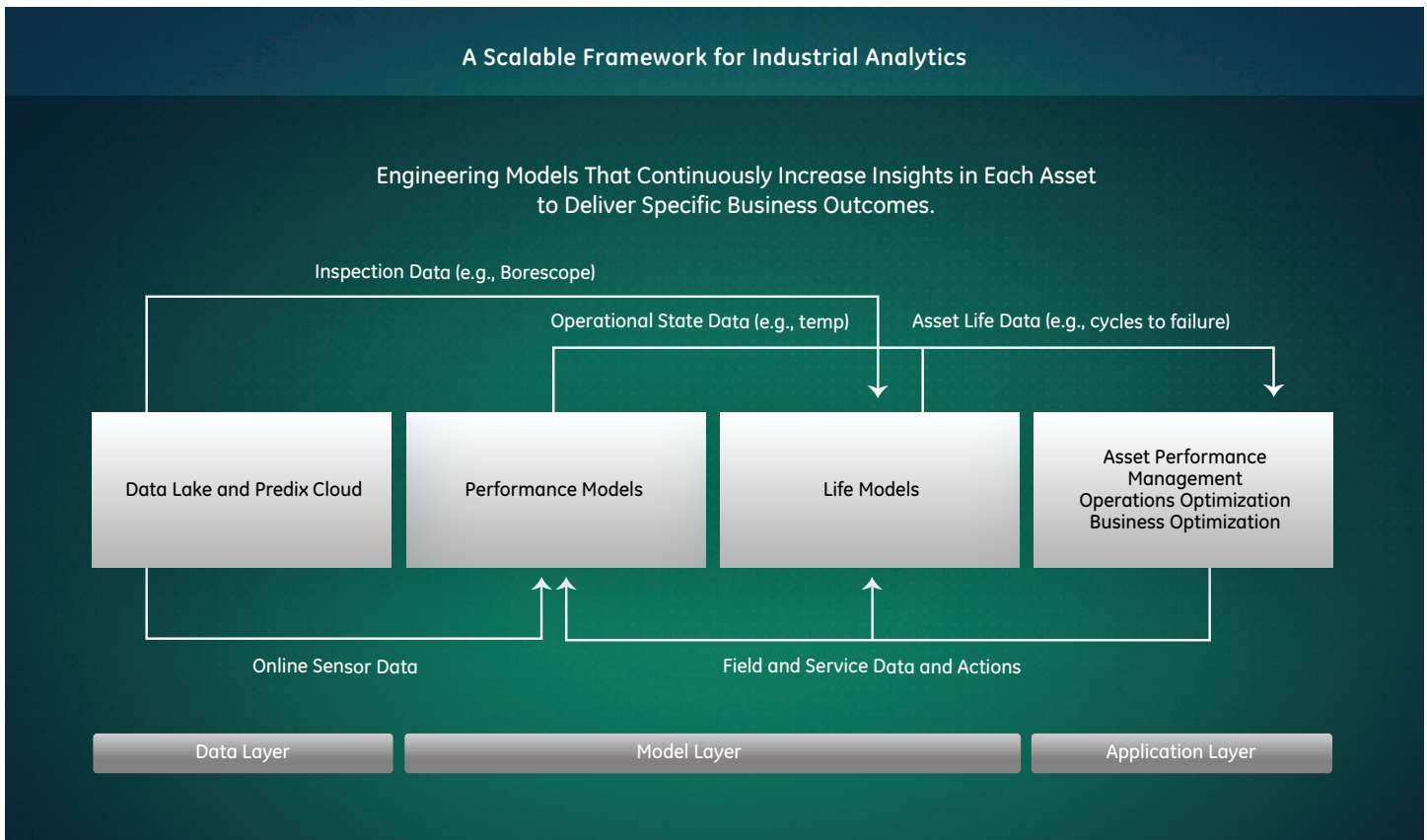


Figure 5

Customer Case Studies

At GE's Remote Monitoring and Diagnostics (RM&D) Center in Atlanta, we are leveraging our Data Historian capabilities to collect data generated from 1,600 gas-fired turbines around the globe. The total output of these turbines can support the annual energy needs of over 60 million homes. The data feeds a central GE Historian cluster. A team of more than 20 M&D engineers analyzes this data to assist customers in enhancing their asset reliability and performance 24/7/365. This data is then fed into a big data Historian running in a Hadoop file cluster, where this team runs complex analytics across 100 million fleet operating hours.⁴

Business outcomes of GE Historian at our RM&D Center include:

Increased productivity

We have reduced decision times and minimized effort spent managing data. This time to find cross-fleet patterns and creating test rules, which used to be in weeks or months, has reduced to minutes or hours.

Reduced costs

Independent software development costs have been reduced by \$3 million.

Decreased infrastructure costs scaling well over \$1+ million include:

Data collection:	From batch to real time
Storage size:	10x reduction
Database cost:	4x reduction
App server cost:	4x reduction
Data retrieval:	10x improvement
Software dev:	Internal to COTS (commercial off-the-shelf)

Higher-quality analytics

Data-driven decisions are correct and optimal. We can now use larger data sets to create rules, thus reducing rule errors on unseen issues.⁴

Increased customer satisfaction

Our customers are experiencing our greater speed and consistency in addition to savings through higher-quality analytics, data-driven analysis and decisions. The questions posed and the answers available to this team via the GE Historian infrastructure have generated customer savings estimated at more than \$100 million.

Elevated team effectiveness

Digitized knowledge reduces training needs and increases staff flexibility. Experienced engineers with valuable knowledge can now be more productive, and new

engineers with less historical knowledge are more effective, improving productivity by nearly \$9 million.

Enabled growth

Digitized knowledge is allowed to focus more on creation of new digital products and services.

Subjecting machinery data

Subjecting machinery data to analysis and data mining operations has yielded significant amounts of productivity for GE and business benefits to GE's customers through better management of their equipment (avoiding unplanned downtimes). **To date, the team has led 15 patents on the Enterprise Historian system and adjacent technologies.**⁴



Figure 6

Case studies below demonstrate how we are partnering with customers in different industries

One of the world's largest independent oil and natural gas exploration and production companies

was challenged with disparate islands of information. They had grown rapidly through acquisition, which resulted in no central data source or a way to collect and report this data. They chose GE Historian as the central data store, and they now collect information from over 20,000 wells. They are able to generate daily operating reports from the information received into this single trusted data store. They worked with GE to create custom collectors from different SCADA systems. They are also automating the modeling required to add a new asset to the Historian database in watch time, not calendar time, saving hundreds of man-hours in labor per year.

A global manufacturer of consumer packaged goods

captures machine data in over 150 line-level GE Historians for local analysis. This data is first aggregated into a Plant Historian, which generates manufacturing metrics like operating efficiency, downtime and waste. This data is then fed into a big data Historian running on Cloudera to run fleet-level queries, taking pressure off the production system and reducing the individual storage requirements at the line and plant levels, where the customer keeps only the last 30 days of data. This customer expects to save millions annually in reduced storage and infrastructure costs.

A major US municipality with 43 political subdivisions

serving more than a million customers is leveraging GE Historian to support their wet weather management and prevent sewer overflow. With this solution, they will meet EPA regulations for water quality and gain visibility to 120 remote pump and lift stations, providing an energy management solution for seven major wastewater treatment plants. With a five-year capital budget of about \$908 million, a 1% decrease in operating costs will save them nearly \$2 million per year and avoid government fines for non-compliance.

Oil and Gas Exploration and Production



Consumer Goods Manufacturing



Waste-water Management



Conclusion

Industrial companies have begun an exciting digital journey. At the heart of this transformation is the power of data analytics to unlock new sources of value. However, the challenges of big data, threat of digital disruption and changing workforce dynamics are real. **In order to exploit the fast-moving technology wave of the Industrial Internet, companies need to think strategically and holistically about the foundational elements of their data architecture, starting with Industrial Data Management.**

GE has invested in a software solutions portfolio to provide our customers the building blocks of achieving Industrial Internet maturity. On-ramping with cost-effective data management technologies that can aggregate, store, analyze and visualize terabytes of data pulsing through assets and systems across the enterprise is a critical foundational step. They give business leaders and operators a single source of truth to improve asset-level visibility, cross-operation performance, knowledge capture and employee collaboration. Having these capabilities in the field and in the cloud sets up the enterprise to extract value from insights that would have otherwise remained hidden within islands of dark and disparate data.

We have deployed data management, predictive analytics and advanced control systems that are yielding operational improvements and increased productivity for our customers. As a first mover of the Industrial Internet, **our mission is to continue to lead the market with our technologies, our domain experience and an unparalleled ecosystem of partners.**

References

- ¹ McAfee, Brynjolfsson; *big data: The Management Revolution*, Harvard Business Review, 2012
- ² Gartner definition of big data. <http://www.gartner.com/it-glossary/big-data>
- ³ Annunziata, Bell, Buch, Patel, Sanyal; *Powering the Future: Leading the Digital Transformation of Energy*, 2015
- ⁴ Puig, Lu, Interrante, Pool, Aggour, Botros; *Enterprise Historian for Efficient Storage and Analysis of Industrial Big Data*, 2013
- ⁵ <http://www.geautomation.com/blog/rethinking-data-management-and-cloud-horizontal-data-value-chain-part-1>
- ⁶ <http://blog.insresearch.com/will-the-datahistorian-die-in-a-wave-of-iiot-disruption>

For more information visit www.ge.com/digital



Imagination at work

GE Digital
2623 Camino Ramon
San Ramon, CA 94583
www.ge.com/digital