The Case for an Industrial Big Data Platform

Laying the Groundwork for the New Industrial Age

GE Software

It will take resources and effort,
but the Industrial Internet can transform
our industries and lives— pushing the
boundaries of minds and machines.

gesoftware.com

Executive Summary

Digital platforms can transform industries, allowing new ways for businesses to connect and co-create value. Such a platform is needed to serve the unique requirements of industrial companies. Industries such as aviation, healthcare, mining, oil and gas, power generation, and transportation represent upwards of 30% of the global economy, and touch the lives of almost everyone on the planet. These capital-intensive industries have long-lived assets such as aircraft, generators, locomotives, and turbines that are mission-critical and require considerable monitoring and service throughout their 20- to 50-year lives. A big data platform that brings new value to the wealth of data coming from these assets, their processes, and the enterprises in which they exist, will set the stage for a new wave of productivity gains and information-based services.[1]

When compared with data in other sectors (e.g., government, financial services, and retail), industrial data is different. Its creation and use are faster; safety considerations are more critical; and security environments are more restrictive. Computation requirements are also different. Industrial analytics need to be deployed on machines (sometimes in remote locations) as well as run on massive cloud-based computing environments. As a result, the integration and synchronization of data and analytics, often in real time,

are needed more than in other sectors. Industrial businesses require a big data platform optimized for these unique characteristics.

The need for a new industrial big data platform is also driven by the advent, and thus ubiquity, of new and cheaper forms of computing, storage, and sensor technology, as well as the growing complexity of industrial companies themselves. Furthermore, industrial operators, from the COO to the field technician, are more mobile than ever, and are looking for more consumer-like experiences in their workplaces, especially as the current generation retires. An integrated platform responds to these dynamics, while unearthing opportunities to connect often highly disparate operations and IT organizations.

GE is uniquely positioned to work with today's software and hardware providers to help define and deliver this platform. GE's customers are the archetypes of the industrial world, and no company understands their requirements and the software platform they need better. But this will need to be a collective effort. The new era of industrial productivity that can be unleashed by an industrial big data platform needs an ecosystem larger than GE if it is to transform not just the industrial world, but the global economy.

[1] See Industrial Internet: Pushing the Boundaries of Minds and Machines: http://www.ge.com/docs/chapters/Industrial_Internet.pdf

1

The Industrial Big Data Platform at Work: The Optimized Wind Farm

A modern 200-turbine wind farm provides an excellent way to understand the requirements of an industrial big data platform. Each turbine contains approximately 50 sensors and control loops with data sampled at different rates, depending on the types of analyses needed and the processes they inform.

The first level of analysis and interaction occurs at the edge – the wind turbine. Real-time analytics within the turbine controller use sensor data collected, and saved in an onboard data historian,[2] every 40 milliseconds to optimize the pitch of the turbine's blades, the conversion of rotational energy into electricity, and to determine whether electricity should be stored in / discharged from batteries or sent to the transmission grid.

The second level of analysis and interaction occurs at the farm. The farm controller receives more than 30 signals from each turbine at 160-millisecond intervals, and its real-time analytics ensure the right combination of turbines deliver predictable power to the utility. In addition, the farm monitoring software processes 200 tags[3] from each turbine at a one-second interval. These near real-time analytics evaluate turbine health and performance for the site operations team.

Turbine data is then transmitted at one-minute intervals from this and other wind farms to a remote monitoring center miles away. There, teams of data scientists and engineers analyze individual turbines and entire wind farms to finely tune their process and asset algorithms, and these enhanced analytics are pushed down to the machines. Simultaneously, a high-performance computer cluster is utilizing years of operating data to build predictive models that find correlations and critical issues hidden amongst the

tags of thousands of individual turbines. When millisecond granularity is needed to match signatures and patterns found in the streaming asset and process data, queries from the monitoring center can be distributed over thousands of wind turbines' data historians. The insights help operations teams prioritize maintenance, parts warehousing, logistics, and other services for the farm.

Finally, analyses that blend operational data from the farm with financial and other data from the operator's enterprise systems are delivered to the CFO's office in the form of forecasts and power production reports. Similarly, the VP of Operations can access reports on turbine capacity, while field management uses reports that detail a prioritized list of maintenance requirements. All are delivered with a modern user interface on whichever is the most appropriate device for the job.

The key to success for this wind farm lies in the ability to collect and deliver the right data, at the right velocity, and in the right quantities to a wide set of well-orchestrated analytics and provide insights at all levels in the operation. This requires a distributed computing fabric optimized for industrial big data in its many forms and in support of its many different uses. These requirements span the industrial world and are what drive the need for an industrial big data platform.

[2] A data historian is a database optimized for time-series or streaming process and asset data

[3] A tag is a raw sensor data point or calculated value with various associated properties including name, value, time stamp, data quality, data type

The key to success for this wind farm lies in the ability to collect and deliver the right data, at the right velocity, and in the right quantities to a wide set of well-orchestrated analytics...

Figure 1: The Wind Farm in the New Industrial Age

	Turbine (Machine)	Wind farm (Plant)	Power producer (Enterprise)
Analytics	Asset optimization	Operations optimization	Business optimization
Data Quantity	>100 tags	>6,000	>1,000,000 tags
Data Frequency	40 milliseconds	1 second	1 second - 10 minutes

Modern wind farms have differentiated requirements for data analysis and operations depending on the type of analysis required, and the industrial big data platform must support them all. At each level in the wind farm's operations – from the individual turbine to the grid to which it supplies power – different data types, data frequencies, analytics, and response times are needed.

The Industrial Big Data Platform at Work: The Optimized Mining Operation

While the quantity of data and the cadence of analysis are different than on a wind farm, the purpose is the same.

Another application of an industrial big data platform can be seen in mining operations. These operations can be summarized in three phases: 1) extracting ore, 2) grinding it down to size for further processing, and 3) separating and concentrating the usable components. Throughput, recovery rate of valuables, and quality of the final product depends on the ability to compensate for process disturbances and asset availability.

At first glance, mining operations are different from wind farms. While wind farms are part of a supply chain that responds to supply and demand signals in milliseconds, mining operations are tied to a supply chain that moves raw materials, concentrates, and downstream products at slower speeds. Nonetheless, there are similarities in terms of the distributed computing fabric needed to execute analytics at different points and at different velocities in the mining value chain.

The concentrating phase of a mine is typically governed by five to ten grinding circuits. Each grinding circuit has 20 to 50

sensors and controllers mounted to drives, gearboxes, pumps, and other components. In addition, each circuit has onboard compute power and storage where the first level of process optimization occurs. Seventy tags are aggregated once per second, and local analytics optimize each grinding circuit's performance. While the quantity of data and the cadence of analysis are different than on a wind farm, the purpose is the same. In both cases, a critical process – in mining, optimizing concentrator throughput – is being optimized.

The next level of analysis occurs at the mine's operating center, where thousands of tags from all concentrators arrive every 15 seconds. These tags are orchestrated with data from dozens of other mining processes and used to analyze the grinding circuit's performance and push operational adjustments back to the grinding circuits' control systems.

As in the wind example, the remote monitoring center receives data from dozens of mines across the globe, this time at 10-minute intervals. At the center, teams of data scientists and engineers utilize massive cloud computing systems to analyze years of historical data to find process or asset anomalies, predict outcomes, and further optimize operations. Finally, concentrator data are blended with data from the mining company's enterprise systems to provide key performance and operations analytics relevant to the CFO, VP of Operations, and other key stakeholders.

The Industrial Big Data Platform at Work: The Optimized Transportation Company

Additional requirements for an industrial big data platform can be seen in the transportation industry. While wind farms monitor and optimize fixed assets, assets in the transportation industry – such as aircraft and locomotives – are designed to be in motion for the majority of their operational lives. Like wind farms and mines, data is transmitted in varying cadences and quantities depending on the analytic, and where it is performed – at the machine or enterprise. Unlike wind farms and mines, however, the communications between these mobile assets and their analytical systems often take place only when the asset has arrived at a destination. This changes how the data and analytics are used.

For an aircraft engine, onboard intelligence must be able to act immediately, without human input, and with extreme degrees of reliability (99.999999999%) and confidence when a potentially significant event is detected or predicted. But, this autonomy of action must be augmented by an understanding of when an escalation is needed. In such a case, on-board analytics for mobile assets must be able to transmit the essential tags[4] via satellite communications to the operations center while in transit. This means that the industrial big data platform must support high levels of mission-criticality, redundancy, and confidence in order to conform to the real time, in-transit requirements of a mobile asset.

For the transportation industry's communications needs, the industrial big data platform has to be able to transfer large quantities of data during what may be a brief refueling or reloading stop. A modern aircraft engine, which generates hundreds of tags every ten milliseconds, can generate one terabyte of sensor data per flight, which means a fleet of hundreds of airplanes will continually create peaks and valleys in an airline's data volumes. While downloading may take place only once or twice a day, the analytics at the airline's flight operations center and an OEM's remote monitoring center must be able to flag potentially anomalous information as it is being downloaded and analyze it in real time. This allows the airline to optimize its maintenance functions, keeping the fleet as close to plan as possible.

Mobile assets operate in continually changing environments, which impact asset degradation at differing rates (e.g., flying through hot, sandy deserts degrades aircraft engine fan blades faster than flying through mild climates), and their locations have to be accounted for in operations and maintenance functions. Unlike fixed assets, these assets have to travel to a maintenance center, as opposed to having maintenance performed on site, adding further complexity to scheduling, parts availability, and other maintenance and operations factors. And significant events must be acted on in real time, while the enormous amounts of data collected by these assets means that data communications must be managed in order to optimize relatively scarce in-motion network bandwidths.

Figure 2: Unique Analytic Requirements of Industrial Companies
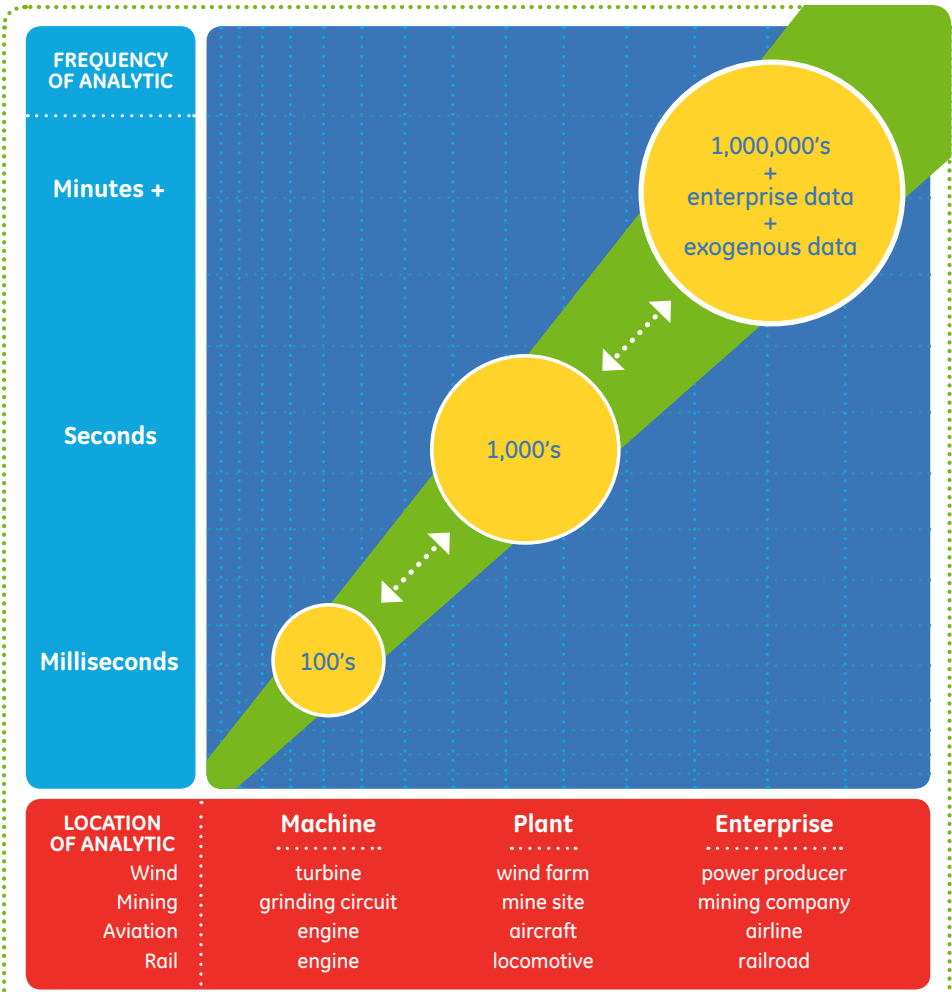
Industrial companies have multiple data and analytical requirements depending on the point of analysis. Machine-level data analysis tends to use relatively small amounts of data for very high-frequency analytics. At each successive level in the analytical hierarchy – at the plant, and the enterprise – the frequency of data collection and analysis decreases while the total quantity of data to be analyzed, and need for operations and IT to integrate, increases.

[4] In addition to typical tag properties, mobile asset tags include properties like latitude, longitude, and altitude.

Optimizing Industrial Data Management and Operations: Key Requirements

The preceding examples illustrate the unique set of big data requirements that must be addressed in order for industrial operators to achieve the many efficiency opportunities in a cost-effective, efficient manner. Ensuring that all industrial companies can reap these asset and operational benefits requires the confluence of six key capabilities. The power of the industrial big data software platform is that it brings these capabilities together in a single technology infrastructure that truly allows the whole to be greater than the sum of the parts.

The key capabilities of the platform are:

1. Data collection and aggregation:
Industrial companies are faced with a rapidly increasing mountain of data and devices that is growing in both quantity and complexity. The platform must allow the collection and aggregation of data and information from the widest possible range of industrial devices and software systems, as well as those from enterprise and web-based systems. It must integrate and normalize different data types (streaming sensor data vs. transactional enterprise data), different response times (once per ten milliseconds vs. once per day), and different business requirements (real-time process optimization vs. less real-time asset optimization), and reconcile their use at different levels of analysis. And the platform must allow companies to move beyond traditional offline data warehouses and make use of modern distributed, scalable, near-line computing environments like Hadoop that can support large-scale query and analysis on all types of data.

2. Advanced analytics at the point of need:
The requirements of analytics managed by industrial companies to optimize their assets and operations are constantly in flux as their ecosystems change. This requires what GE calls a software-defined machine – the ability for assets to be abstracted into software and to run in connected virtual environments where analytics are continually tuned to the requirements of specific devices, business processes, and individual roles. Analytic requirements may range from embedded systems that analyze data and respond in milliseconds to complex, predictive modeling analytics, deployed on massive cloud-computing infrastructures that can sort through terabytes of widely disparate data. The platform also needs to provide a library of standard analytics that allow operators, if they desire, to design, test, and rapidly deploy new analytics.

3. Cloud-agnostic, deployment independence:
Underlying these technology changes are new deployment models like cloud and mobility that are further changing the economics of computing and business. Industrial companies need a highly flexible deployment architecture that allows them to mix and match technology deployment methods – and avoid vendor lock-in – as their needs and technological options change. For companies that are bound by regulatory requirements, this may mean supporting private cloud deployments; for other companies, it may mean supporting third party public clouds from various providers. The industrial big data platform must be able to ensure that deployment flexibility – on-premise, in the cloud, or some combination – is built into the platform.

4. Extensibility and customizability:
Taking advantage of new revenue opportunities requires industrial operators to adapt quickly to changing customer requirements and dynamic competitive and regulatory environments. This requires the industrial big data platform to be highly extensible and based on standardized APIs and data models that allow it to adapt to new capabilities, new devices, new

data types, and new resources as they become available, while still preserving the capabilities that continue to impart value. The platform also has to support an ecosystem of in-house and third party developers that can enhance existing industrial solutions and innovate new ones. And the ability of the platform to leverage commonalities of the entire industrial sector must be tempered with the ability to customize the platform to meet specific company, geographical, and/or vertical requirements.

5. Orchestration: The industrial big data platform must support the orchestration of information, machine controls, analytics, and people in order to ensure that the different components of the industrial big data world interoperate effectively. This requires the ability to schedule and manage how information, analytics, machines, and people work in sync with each other. A machine-level analytic that detects and responds locally to an operational anomaly must also be able to set in motion other analyses and follow-up actions (e.g., rescheduling flights or moving spare parts) across the system. This requires the ability to tune and adapt easily as data, process, and business models change.
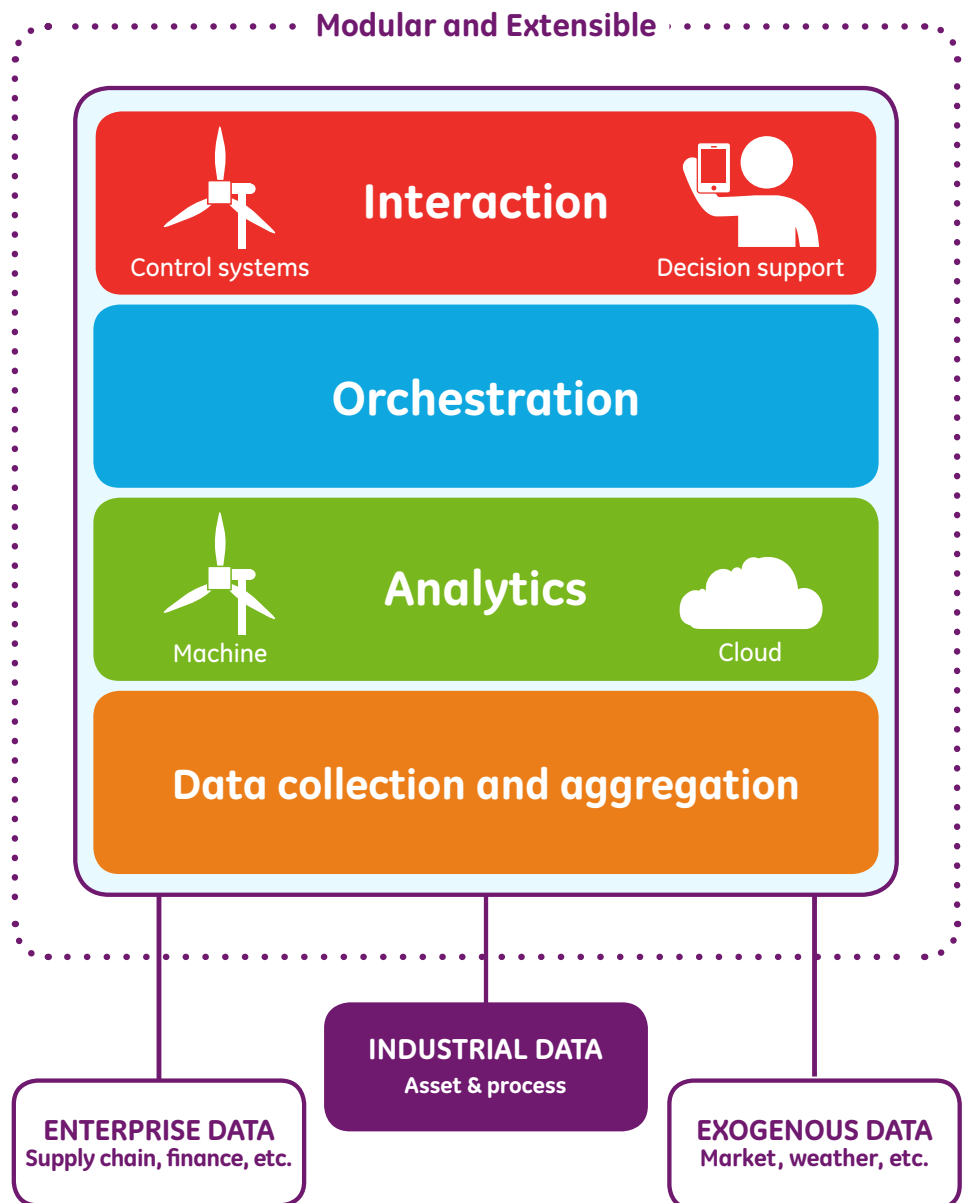
6. Modern user experience: The work force across all industries is in the midst of a generational shift. The next generation is demanding user experiences more in line with the consumer technologies they have become familiar with. This means that the industrial big data platform has to deliver the above components within the context of a modern user experience that is no longer bound to a desktop. The platform has to be as relevant to the pilot flying at 35,000 feet as it is to the mining engineer repairing equipment deep under the earth. This includes supporting a wide range of mobile devices and user interaction models, as well as ensuring that the user experience is tailored to the individual's role and requirements at a particular time and location.

The Industrial Internet is changing how data and information are used by industrial operators, and these changes are opening up new opportunities for them and their ecosystem partners. This degree of change requires a new platform. Existing data platforms, the majority of which were designed to handle traditional back office

IT requirements, are not adequate to meet the particular needs of modern industrial companies.

What is needed is a single industrial data platform that supports the unique requirements of the industrial world. With the degree of change already in motion, and with more on the way, only an industrial big data platform that combines the above capabilities will allow the vision of the modern industrial world to be realized.

Figure 3: The Architecture of the Industrial Big Data Software Platform

Modular and Extensible

Interaction
Control systems — Decision support

Orchestration

Analytics
Machine — Cloud

Data collection and aggregation

INDUSTRIAL DATA
Asset & process

ENTERPRISE DATA
Supply chain, finance, etc.

EXOGENOUS DATA
Market, weather, etc.

How to Lay the Groundwork for the New Industrial Age: Data, People, and Architecture

While GE and its partners are actively pursuing the realization of this industrial big data platform and its new capabilities, industrial companies that want to stay ahead of the curve can begin to make the promise a reality in the following ways:

Focus on your data: Begin the process of understanding and collecting the different kinds of data that are inside and outside your company. Some of these data are readily accessible; some can come from adding sensor capabilities to existing equipment; some will come from a next-generation sensor and monitoring infrastructure; and others will come

from third party information services. Understanding what these data are and how they can be used to derive new business value is a core competence of the new industrial age.

Focus on your people: Start looking at the talent requirements that will be needed to make use of these data. New roles such as data scientist or Chief Data Officer may be necessary, and business managers and executives may need training to be able to leverage the big data opportunities that await your company. The promise of the industrial big data platform cannot be realized if operations and IT continue to function as separate silos within the organization. The expertise and experience of these two groups must be brought together in order to create the new analytics and operational models that can leverage this opportunity. Sponsor workshops and contests. Get IT and operations in the same room and have them teach each other about their core processes, data, and operational requirements.

Focus on your architecture: Solely utilizing traditional data warehouses and batch analytics is no longer an option. Making use of an IT infrastructure that scales between always-on, real-time and near real-time, in-memory analysis and machine-learning, batch analytics at different locations from machine to cloud - private, public, or hybrid - is an important starting point for industrial companies looking to plan for the new industrial age. This means getting on board with technologies like Hadoop and NoSQL, among others. These and other technologies are already becoming essential components of the technology infrastructure of industrial companies, and their importance will grow over time.

Conclusion:
The Future Will be Collaborative

The vision set forth in this paper is rapidly becoming a reality across a range of industries – wind, mining, aviation, and rail included. But the process of consolidating this progress and driving the fulfillment of an industrial big data platform is one that will require the help of an ecosystem of software, hardware, and service providers – some established, and some start-ups. And central to all this will be the participation of the industrial operators – the airlines, the mining companies, and utilities, to name few – that stand the most to gain by its successful build-out. This is the mission that GE Software has signed on to deliver to the industrial world. As we move forward with our mission, we welcome your participation. Together we can ignite the next Industrial Revolution.
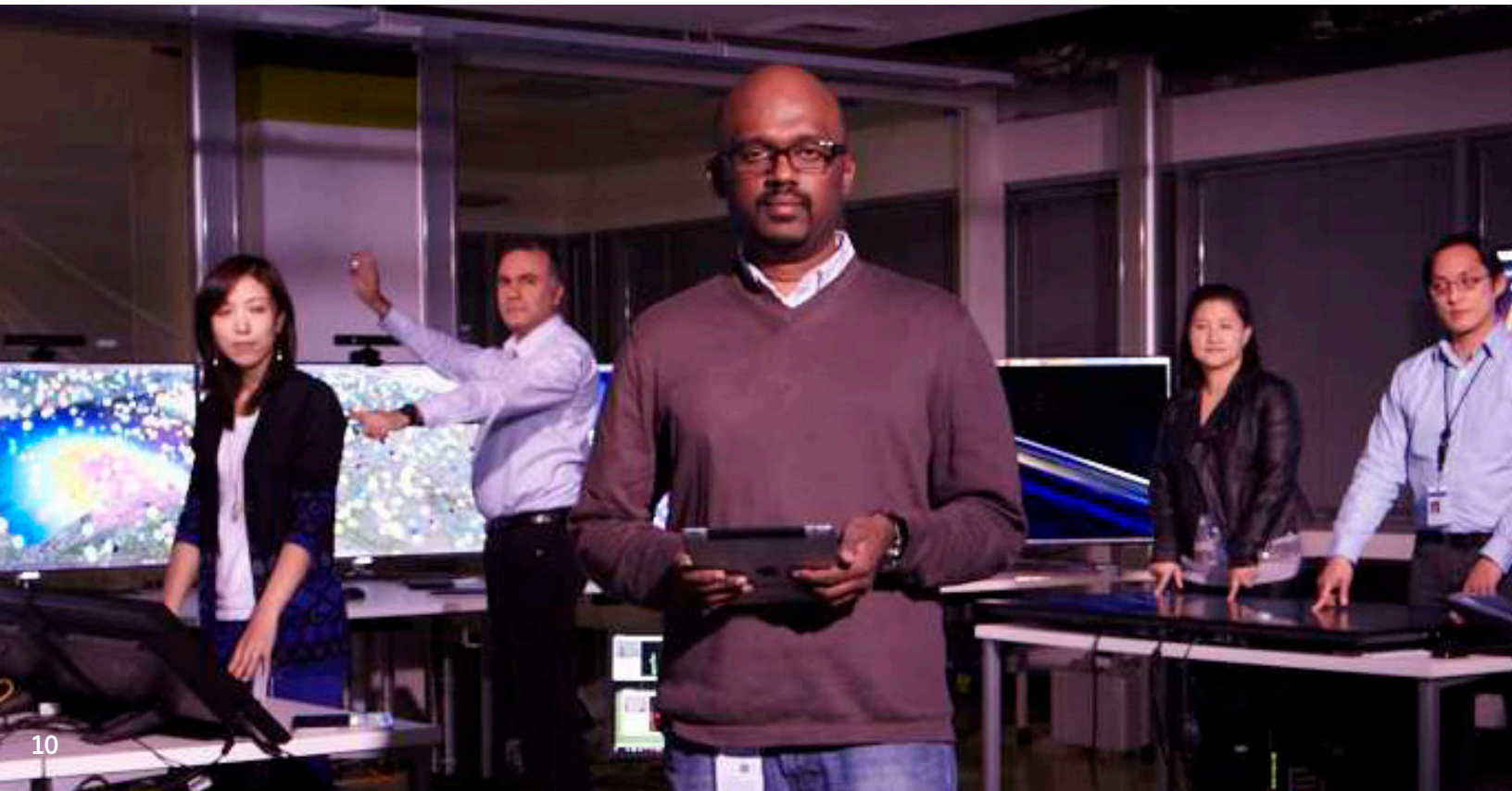
imagination at work

gesoftware.com
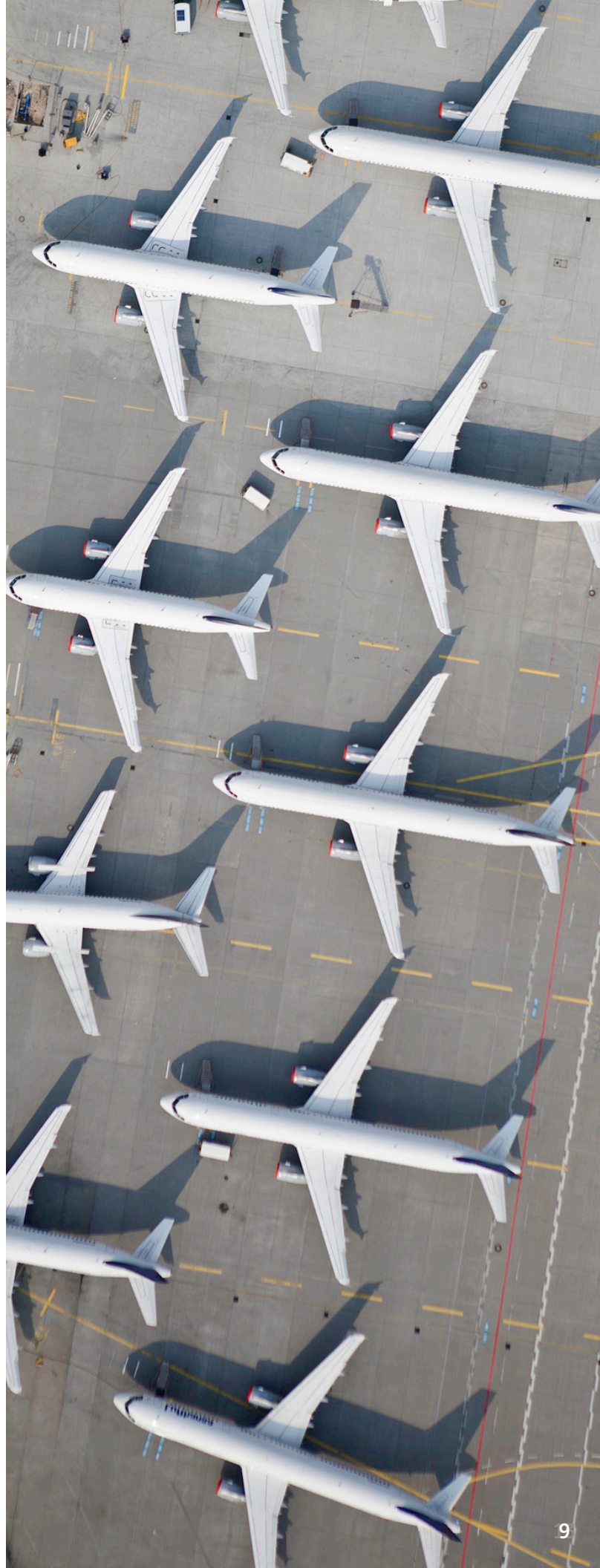
Driving the next Industrial Revolution

GE Software helps customers realize greater returns faster from what GE calls the Industrial Internet – a technology framework that connects machines, facilities and people to deliver business value and improve productivity. The Industrial Internet will drive profound productivity gains not seen since the Industrial Revolution.

The world is on the threshold of a new era of innovation and change with the rise of the Industrial Internet. It is taking place through the convergence of the global industrial system, with the power of advanced computing, analytics, low-cost sensing and new levels of connectivity permitted by the Internet. The deeper meshing of the digital world with the world of machines holds the potential to bring about profound transformation to global industry, and, in turn, to many aspects of daily life, including the way many of us do our jobs. GE is working with its partners and customers to develop the software and analytics infrastructure, tools, and services to make the Industrial Internet a reality.

This advanced and intelligent technology framework uses Big Data to predict and respond efficiently to change.

gesoftware.com

GE Software