Big Data Case Study – Navya Mote
(follow this structure for your report)

1) **Who** (company, agency, organization)

Who they are, what do they do?

What is their role/purpose?

"General Electric (GE) is an American multinational conglomerate corporation incorporated in New York[1] and headquartered in Boston, Massachusetts"[2]. The revenue generated in 2017 was US$122.09 billion[3] and the number of employees being 295,000 as of 2016. It was established by Thomas Edison and is the first private company in the world to own its own computer system, in the 1960s[4]

GE is a powerhouse of a corporation that is involved in almost every area of industry being appliances, aviation, capital, consumer electronics, power, digital, energy connections, global research, consulting services, industrial solutions, intelligent platforms, lighting, mining, oil&gas, healthcare and transportation to name a few.

The purpose of GE is to build products that help build the world and to make them work better.

2) **Need**

Why did they want/need to do a big data project?

They wanted to build a platform that would help their separate machines and tools connect with each other and to be able to constantly report their status. Since just one gas turbine of theirs generates about 500GB of data every day, it became a big data project.

In essence, the idea is that all the separate machines and tools which make an industry possible will be "smart" – connected, data-enabled and constantly reporting their status to each other in ways as creative as their engineers and data scientists can devise[5].

What *questions* did they want to answer?

They wanted to be able to increase efficiency by monitoring every aspect of an industrial operation and to able to tweak it for optimal performance. They also wanted to reduce the downtime caused by machinery break down . With the help of sensors, they would know exactly when to replace the worn part.

What *data sources* were available?

Their wind turbines are deployed with sensors which stream the data constantly to the cloud. As for homes, they have smart meters which record data on power consumption. Large amounts of data is also recorded from every aircraft and every aspect of ground operations.

---

[1] EDGAR Search Results , header. Retrieved December 21, 2015, https://www.sec.gov/cgi-bin/browse-edgar?action=getcompany&CIK=0000040545&owner=include&count=40

[2] Fact Sheet, About Us. GE. Retrieved August 1, 2016, https://www.ge.com/about-us/fact-sheet

[3] GE ANNOUNCES FOURTH QUARTER 2017 RESULTS, January 24, 2018 - GE (NYSE:GE) announced results today for the fourth quarter of 2017, https://www.ge.com/investor-relations/sites/default/files/ge_webcast_pressrelease_01242018.pdf

[4] Big Data Case Study Collection, Bernard Marr ,2015, Wiley, https://www.bernardmarr.com/img/bigdata-case-studybook_final.pdf

[5] Big Data Case Study Collection, Bernard Marr,2015,Wiley, https://www.bernardmarr.com/img/bigdata-case-studybook_final.pdf

GE intends to take the warning and error alert data it has been able to harvest the past five years and use it to identify patterns of behavior that lead to more precise scheduled maintenance[6]

Who *owns* the data;

what *access rights*?

what *privacy* issues?

what *quality* issues?

GE generate, capture and analyze internal data from the operation of their machines, as well as external data from a wide range of outside suppliers including meteorological, geopolitical and demographic data. This external data includes satellite imagery. Just one of their gas power station turbines generates around 500 gigabytes a day, covering everything from the environmental temperature it is operating at to its efficiency at converting burning coal into electrical energy [7]

Since the data gathered is from machines that are manufactured by GE, they do not face problems related to accessing the data. Also as GE takes the necessary precautions to keep the data secure, they do not face problems relating to privacy. Sometimes the data collected is fragmented and soiled, for which GE writes algorithms to derive useful information.

3) Challenges

What *technical* and *organizational* challenges did they face?

Why was this a "big data" problem? (what were the “Vs”?)

The technical challenges that GE faced was that the datasets could be fragmented and soiled which would make it difficult to use elsewhere in the business. Since Operational technology and Information technology systems operate separately, it would lead to duplication as the roles they play in an organization merge at times. Edge devices may not always be connected due to privacy, security, regulatory reasons but need to be functioning even when the connection is temporarily unavailable. Applications need to analyze data and send back insights in real-time due to which there might not be enough time to place the data in the cloud. Diversity in the practices used across the organization might lead to performance variability. They wanted the platform to be machine-centric, to be able to store, manage, integrate and access heterogenous data. It also needed to be able to predict, to guide the user with intuitive experiences on a device of their choice and to deliver the information securely with the help of cloud.

Their engineers had experience as either mechanical engineers or computer scientists but not enough in the area of analytics. As the demand for services capable of carrying out high speed analytics on large volumes of data increased, they needed to increase their investment in analytics staff and research.

One of the largest challenges facing GE’s system architects and engineers while implementing Industrial Internet infrastructure was scale. They had to keep in mind the vast number of industries their machines operate in and the spread of their customers worldwide. While catering to this, they had to work on centralizing the storage and

---

[6] THINK BIG DATA , Andrew Lampitt, InfoWorld, JAN 3, 2013,https://www.infoworld.com/article/2616433/big-data/general-electric-lays-out-big-plans-for-big-data.html

[7] Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results, By Bernard Marr, ©2016 Bernard Marr, http://onlinelibrary.wiley.com/doi/10.1002/9781119278825.ch19/pdf

perform analysis on the data collected, which was a huge task. They also had to be prepared for the volume of data that they would encounter with as Industrial data is growing twice as fast as any other sector. For GE, the sheer volume of data was the biggest big data problem. In the case of its aircrafts, they also provided real-time alerts which would address another big data problem which is of velocity.

Some amount of variety is also present as GE works with external data from a wide range of outside suppliers including meteorological, geopolitical and demographic data. This external data includes satellite imagery.

4) **Requirements**, resources needed

What HW/SW resources did they need to conduct the project?

People, process, technology

They needed a software platform that would be reusable, build apps quickly, be able to leverage work from elsewhere, reduce sources of error, develop and share best practices and lower risk of cost and time overruns.

Due to the growing demand from GE's Industrial customers, they needed to start looking for talent that would be able to cater to the needs of the industrial internet.

They needed a cloud model that would lower costs, be centrally managed and have a shared infrastructure in a pay as you go subscription model. Cloud would also help them scale to meet different business workloads as it can adjust capacity on demand. To be able to generate and deliver analytical insights that can be modelled across the organization. They also needed a edge and cloud architecture, that would reduce latency for mission-control and safety-critical features and adhere to SLA's and regulatory compliance.

5) What was the approximate project *schedule*? *Project plan*? *Deliverables*?

The inception of GE Software was first announced in November 2011 as a part of GE's Global Research Center. This project was headed by GE Digital CEO Ruh and was based in San Ramon, California . The project was rolled out in November 2012 and was GE's first-ever uniform product offering, under the Predictivity brand[8]. In 2012 they announced $1 billion was being invested over four years in their state-of-the-art analytics centre in San Ramon, California, in order to attract pioneering data talent to lay the software foundations of the Industrial Internet[9][10]

6) Solution

What *resources* were used?

Internal and commercial

People and technologies

[8] GE and the Industrial Internet, Karim R. Lakhani, Marco Iansiti and Kerry Herman, APRIL 2014 (REVISED MARCH 2015),https://www.hbs.edu/faculty/Pages/item.aspx?num=47272

[9] Big Data Case Study Collection, Bernard Marr,2015,Wiley,https://www.bernardmarr.com/img/bigdata-case-studybook_final.pdf

[10] How GE Is Using Big Data to Drive Business Performance, Bernard Marr, 2014, https://www.smartdatacollective.com/how-ge-using-big-data-drive-business-performance/

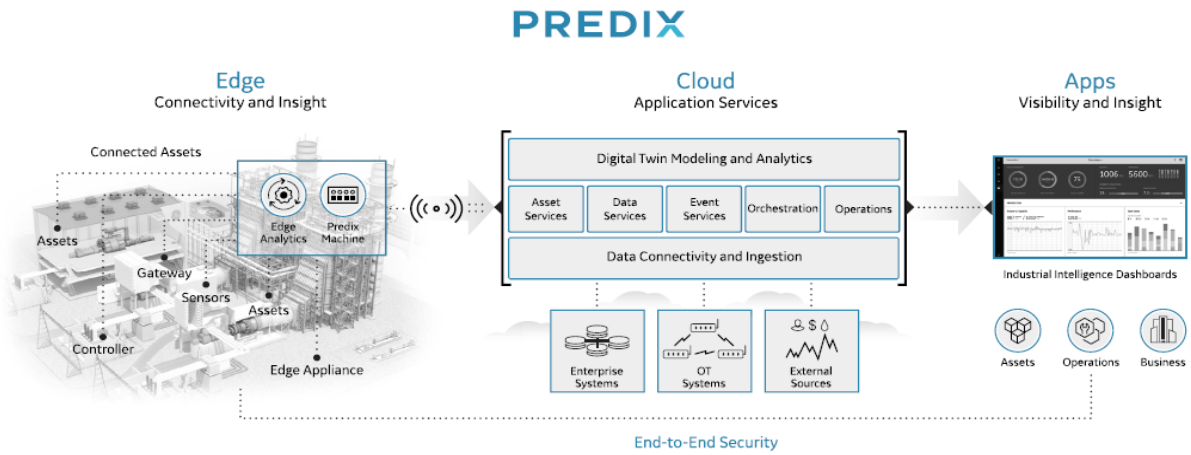
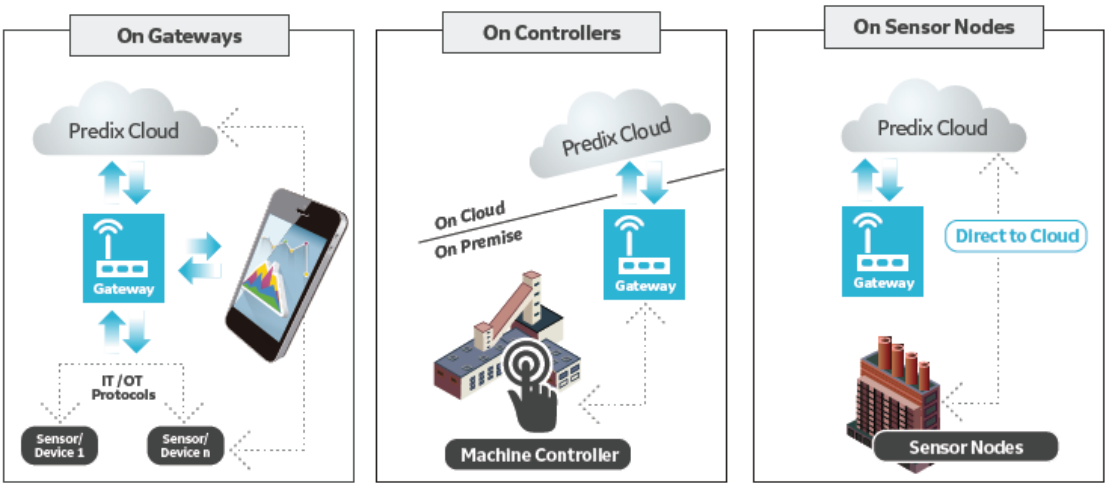


Fig 1 [11]

GE let go of their go-it-alone cloud strategy and chose to use other cloud services such as Amazon AWS and Microsoft Azure.

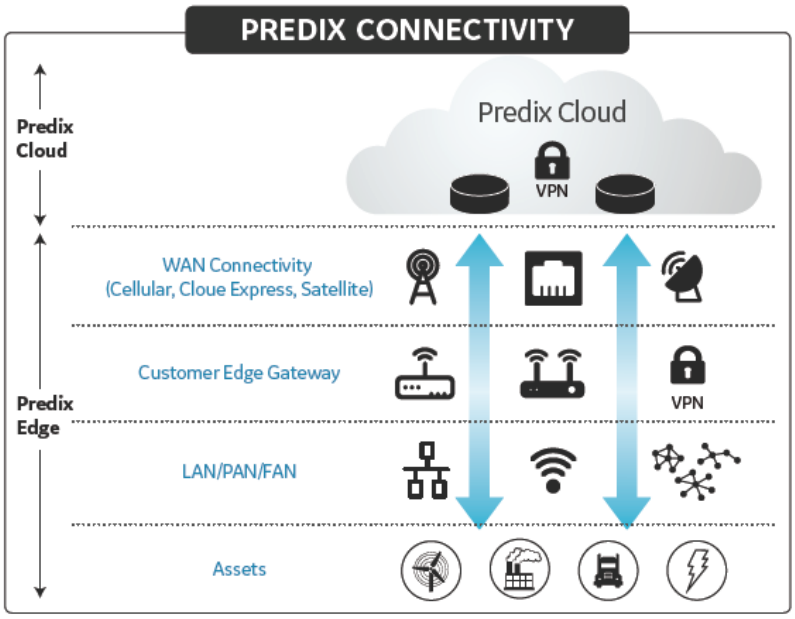
Predix Edge is an integration of Predix Machine, Predix connectivity, and Predix Edge Manager. Predix Machine is a software that develops and deploys machine apps, and can build a new generation of social and interoperable products. The hardware components required by Predix is sensors, controllers, gateways and on-premise appliances. Predix Edge Manager is used to manage the Predix Machine running on the hardware.



[11] predix.com , © 2016 General Electric Company – All rights reserved, the GE Monogram and Predix are trademarks of General Electric Company,https://www.predix.com/sites/default/files/predix-the-industrial-internet-platform-from-ge-digital-brief.pdf

Fig 2[12]

Predix connectivity that would offer seamless, secure, and reliable end to end connectivity between Predix Edge and controller devices. End-to-end route and flow management between edge and cloud without multiple carrier arrangements. Protocol-agnostic network configuration and management for M2M (Machine-to-Machine) and M2C(Machine-to-Cloud) connectivity. Centrally managed policies driving QoS and bandwidth optimization. Policy-driven data forwarding between multiple cloud and on-premises destinations. Physical connectivity globally via cellular, fixed or satellite networks through partnerships with communication service providers. Secure virtual private network (VPN) between the edge and cloud, ensuring data privacy and asset protection. Ability to manage and control the edge assets by providing remote access via VNC, RDP, SSH, and HTTP. End-to-end monitoring and notifications about the connectivity between Predix cloud and edge assets. One-stop-shop billing and reporting for all connectivity and IP services. Zero touch provisioning with a self-management portal.



A truly global industrial platform needs to consume and analyze massive volumes of data by connecting to a wide variety of machines, sensors, control systems, data sources, and devices.

Fig 3[13]

Predix Edge Manager provides a centralized view of edge devices that are running Predix Machine. It help to manage, configure, and administer the edge devices, apps and users. Another use of the Edge Manager is that it can quickly determine device condition and connectivity health while managing apps.

Predix is built on an open source Platform-as-a-Service(PaaS) known as Pivotal's Cloud Foundary. Using Predix, app developers can quickly build, test, deploy, and scale

[12] predix.com , © 2016 General Electric Company – All rights reserved, the GE Monogram and Predix are trademarks of General Electric Company,https://www.predix.com/sites/default/files/predix-the-industrial-internet-platform-from-ge-digital-brief.pdf

[13] predix.com , © 2016 General Electric Company – All rights reserved, the GE Monogram and Predix are trademarks of General Electric Company,https://www.predix.com/sites/default/files/predix-the-industrial-internet-platform-from-ge-digital-brief.pdf

application in a matter of hours and days instead of weeks or months. The programming languages used to publish analytics is Python, Java, and Matlab.

In the Bay Area of San Francisco, they were able to find startups working on analytics led technology in which they invested, acquired or partnered with. To meet the company goals of Industrial Internet, GE recruited Ruh from Cisco to lead the initiative.

7) **Results/Findings**

What *results* were achieved?

GE's Predictivity platform has been successfully implemented and allows real-time automated analysis. With Industrial Internet of Things, machines are now able to order new parts for themselves and therefore minimizing expensive downtime. GE has said that its contractors use to lose an average of $8 million per year due to unplanned downtime[14].

Their wind turbines are deployed with sensors which stream data constantly to the cloud. This data can be used to fine-tune the pitch, speed, and direction the blades are facing to capture the maximum wind possible. Therefore, the Green industries are benefitting from this technology deployed on about 22,000 wind turbines. These turbines communicate with each other with the help of sensors. And with the responses that they get from other turbines, they mimic it to become more efficient and also pool the resources in case one device on a neighboring turbine should fail.

Their data gathering extends into homes too – millions are fitted with their smart meters which record data on power consumption, which is analyzed together with weather and even social media data to predict when power cuts or shortages will occur.[15]

Was the project *successful*?

Based on the overall results, I would say that the project was successful in implementing the goals

Were there any *surprises*?

Yes, there were. They had planned to build their own cloud platform but by then Amazon had come up with AWS and Microsoft with Azure whose cloud services would cost lesser and save GE time. They also did not anticipate that converting their existing objects to be compatible with Predix would be time consuming as the existing algorithms were written in different programming languages.

8) **Critique**

How could the project have been *improved*?

Looking at what some analysts and investors have to say about GE's progress in implementation of Industrial Internet project, it's current growth rate is too slow to reach the target of $12 billion by 2020. They also say that GE decided to go digital sooner than other companies which could have costed the company more time and money to better their strategy. I think that the project could have been improved if there was better planning at the initial stage of implementation. Since they took a big step by going digital, they could have addressed one problem at a time instead of implementing the Industrial Internet across all industries. That way they could have made sure that it is working well in one industry and moved on to the next. Breaking the problem down into smaller problems would have made it easier to deliver and meet the target. I think implementing large scale is what could have taken more time and made things

---

[14] How GE Is Using Big Data to Drive Business Performance, Bernard Marr, 2014, https://www.smartdatacollective.com/how-ge-using-big-data-drive-business-performance/

[15] Big Data Case Study Collection, Bernard Marr,2015,Wiley,https://www.bernardmarr.com/img/bigdata-case-studybook_final.pdf

complicated. More time and money could have been spent on research to test the solutions before deploying it across industries.

9) Lessons learned

What *lessons* were learned from conducting the project?

Initially GE had planned to build data centers on their own cloud platform to be called as Predix Cloud. But after Amazon.com Inc and Microsoft Corp built AWS and Azure that offer cloud services, GE changed their course. As GE let go of their go-it-alone cloud strategy and chose to use other cloud services, they were four months behind schedule. When they had AWS and Azure to use and did not have to build their own cloud platforms, their engineers focused on applications, which turned out to be more profitable than the platform alone. This was the biggest lesson learned as said by GE Digital CEO Ruh. Since GE has different systems residing in their businesses, they faced challenges adapting to the Predix software. The existing algorithms for monitoring its machines were written in different coding languages, which would be time consuming transferring them to Predix. GE acquired companies like Meridium and ServiceMax , which bought more code that needed to be converted in order to work on Predix. Because of which, software installation took longer than they expected. Some of GE's executives acknowledged that Predix experienced technical issues and was running behind schedule. The code had bugs and applications lacked features required by clients.

What specific *actions* were taken as a result of the project?

They had to take a time out in May and June 2017 to be able to fix the bugs and make Predix stable. There were few organizational changes such as Predix Chief Harel Kodesh was succeeded by Patrick Franklin. It was Franklin who called the time out to fix Predix. GE Digital CEO Ruh said that as the leadership changed and evolved with the business, GE now has the right people to help Predix grow.

Value obtained

What *value* was obtained as a result of the project?

When technology is interconnected, it can improve efficiency to a large extent across all sectors. Instead of waiting for the machines to fail in order to replace its parts, a better way would be if the machines could predict themselves when a part would need replacement. And that is exactly what GE is trying to do with the help of Industrial Internet of Things. Also as machines would be more reliable with these predictions in comparison to people, it is indeed a good idea to show them how to do it. One of the ways to bring in this efficiency is by combining data with analytics.

GE's value proposition is that a 1 percent boost in, say, jet turbine productivity, can save a company billions of dollars. The company believes that the market for a platform and applications in the industrial segment could reach $225 billion by 2020[16]

Although GE have not released overall figures, they have said that their industrial customers can expect to save an average of $8 million per year from the reduction in machine downtime alone. They have also said that if their systems can improve operating

[16] Why GE is winning the war for tech talent , Clint Boulton , MAR 16, 2017,https://www.infoworld.com/article/3181890/it-management/why-ge-is-winning-the-war-for-tech-talent.html

efficiency across five of their key sectors then businesses adopting their Big Data-driven technology can expect to make combined savings of $300 billion. Analytics in general, GE CEO Jeff Immelt has claimed, could create between $10 trillion and $15 trillion value in the world economy over the next 20 years.[17]

10) Explain/define terms

Include *explanation* of any *technical terms* relevant to the project

Edge devices : An edge device is a device which provides an entry point into enterprise or service provider core networks

Latency : The delay before a transfer of data begins following an instruction for its transfer

SLA's : Service Level Agreement

Regulatory compliance : Regulatory compliance describes the goal that organizations are aware of and take steps to comply with relevant laws, policies, and regulations

Sensors : A sensor is a device that detects events or changes in the environment and sends the information to other electronics, frequently a computer processor

Controllers : A chip or expansion card that interfaces with a peripheral device

Gateways : A link between two computer programs allowing them to share information and bypass certain protocols on a host computer

On-premise : On-premises software is installed and runs on computers on the premises (in the building) of the person or organization using the software, rather than at a remote facility such as a server farm or cloud

Protocol : The set form in which data must be presented for handling by a particular computer configuration, esp. in the transmission of information between different computer systems.

Agnostic : Agnostic, in an information technology (IT) context, refers to something that is generalized so that it is interoperable among various systems

QoS : Quality of Service is the description or measurement of the overall performance of a service, such as a telephony or computer network or a cloud computing service, particularly the performance seen by the users of the network

Policy-driven data : Use policies to meet the network performance and design requirements of modern data center and cloud environments

VPN : A virtual private network (VPN) extends a private network across a public network, and helping users to send and receive data across shared or public networks as if their computing devices were directly connected to the private network

VNC : In order to remotely control another computer, we use Virtual Network Computing(VNC) which is a graphical desktop sharing system that uses the Remote Frame Buffer protocol to establish the connection

RDP : Remote Desktop Protocol (RDP) is a proprietary protocol developed by Microsoft, which provides a user with a graphical interface to connect to another computer over a network connection

SSH : In order to operate network services securely over an unsecured network, we make use of the secure shell

HTTP : Hypertext Transfer Protocol (HTTP) is an application-layer protocol for transmitting hypermedia documents, such as HTML

---

[17] Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results, By Bernard Marr, ©2016 Bernard Marr, http://onlinelibrary.wiley.com/doi/10.1002/9781119278825.ch19/pdf

Platform-as-a-Service(PaaS) : Platform as a Service (PaaS) is a category of cloud computing services that provides a platform allowing customers to develop, run, and manage applications without the complexity of building and maintaining the infrastructure typically associated with developing and launching an app
Pivotal's Cloud Foundary : Pivotal Cloud Foundry is engineered to deliver a single software platform to run an entire enterprise—capable of scaling to support hundreds of IT teams and thousands of applications

11) References

Provide *appropriate citations and references (see FAQ document)*

[1] EDGAR Search Results , header. Retrieved December 21, 2015, https://www.sec.gov/cgi-bin/browse-edgar?action=getcompany&CIK=0000040545&owner=include&count=40
[2] Fact Sheet, About Us. GE. Retrieved August 1, 2016, https://www.ge.com/about-us/fact-sheet
[3] GE ANNOUNCES FOURTH QUARTER 2017 RESULTS, January 24, 2018 - GE (NYSE:GE) announced results today for the fourth quarter of 2017, https://www.ge.com/investor-relations/sites/default/files/ge_webcast_pressrelease_01242018.pdf
[4] Big Data Case Study Collection, Bernard Marr ,2015, Wiley, https://www.bernardmarr.com/img/bigdata-case-studybook_final.pdf
[5] Big Data Case Study Collection, Bernard Marr,2015,Wiley, https://www.bernardmarr.com/img/bigdata-case-studybook_final.pdf
[6] THINK BIG DATA , Andrew Lampitt, InfoWorld, JAN 3, 2013,https://www.infoworld.com/article/2616433/big-data/general-electric-lays-out-big-plans-for-big-data.html
[7] Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results, By Bernard Marr, ©2016 Bernard Marr, http://onlinelibrary.wiley.com/doi/10.1002/9781119278825.ch19/pdf
[8] GE and the Industrial Internet, Karim R. Lakhani, Marco Iansiti and Kerry Herman, APRIL 2014 (REVISED MARCH 2015),https://www.hbs.edu/faculty/Pages/item.aspx?num=47272
[9] Big Data Case Study Collection, Bernard Marr,2015,Wiley,https://www.bernardmarr.com/img/bigdata-case-studybook_final.pdf
[10] Big Data Case Study Collection, Bernard Marr,2015,Wiley,https://www.bernardmarr.com/img/bigdata-case-studybook_final.pdf
[11] Why GE is winning the war for tech talent , Clint Boulton , MAR 16, 2017,https://www.infoworld.com/article/3181890/it-management/why-ge-is-winning-the-war-for-tech-talent.html
[12] Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results, By Bernard Marr, ©2016 Bernard Marr, http://onlinelibrary.wiley.com/doi/10.1002/9781119278825.ch19/pdf
[13] predix.com , © 2016 General Electric Company – All rights reserved, the GE Monogram and Predix are trademarks of General Electric Company,https://www.predix.com/sites/default/files/predix-the-industrial-internet-platform-from-ge-digital-brief.pdf
[14] GE shifts strategy, financial targets for digital business after missteps , Alwyn Scott , AUGUST 28, 2017 ,https://www.reuters.com/article/us-ge-digital-outlook-insight/ge-shifts-strategy-financial-targets-for-digital-business-after-missteps-idUSKCN1B80CB

[15] Behind GE's Vision For The Industrial Internet Of Things , JON GERTNER , June 2014, https://www.fastcompany.com/3031272/can-jeff-immelt-really-make-the-world-1-better

[16] General Electric,2018, https://www.ge.com/about-us/building

(highlighted items are especially important)