SAVONIA

ENGINEERING AND TECHNOLOGY

BIG DATA AND ITS OPPORTUNITIES FOR THE ENGINEERING COMPANIES

RESEARCH PUBLICATION OF DIGIBOOST PROJECT

WRITER **Karapetyan Karina**

Big Data and its opportunities for the engineering companies

Research publication of DigiBoost project

Karapetyan Karina

Savonia University of Applied Sciences
P.O. BOX 6 (Microkatu 1 D)
70201 KUOPIO
tel. +358 17 255 5023
fax. +38 17 255 5043
e-mail: julkaisut@savonia.fi
www.savonia.fi/julkaisut

1st edition

ISBN: 978-952-203-210-2 (PDF)
ISSN-L: 2242-7015
ISSN: 2242-7015
Savonia University of Applied Sciences, Series D4/7/2015

Publisher Savonia University of Applied Sciences
Layout manager Tapio Aalto

CONTENTS

1 INTRODUCTION

1.1 Targets and methods

The target of the publication was to answer on two questions:

- What can the Big Data bring as the perspective to the engineering companies?
- How can it affect the future?

During the work on the publication the following methods were used:

- Literature review and case study;
- Gathering data from the publications and books, data sheets, the Internet and other sources, as well as analyzing and organizing these data;
- Interview in the Case Company;

1.2 DigiBoost research

The research was a part of the *DigiBoost - The new possibilities of industrial service business project (2015 - 2016)*, coordinated by Savonia University of Applied Sciences. The project concentrates on exploring the new possibilities of industrial service business development in close cooperation with the local engineering and energy industries. As there are multiple new business opportunities and targets related to the *Internet of Things* and *digitalization*, also the phenomenon of *Big Data* must be explored as a part of the research activities.

Big Data and Internet of Things have close relationship: IoT is one of the major reasons of Big Data explosion. The technologies provided by the Internet of Things help to listen to data, gain its value and provide the efficient results from it.

2 BIG DATA IN GENERAL

2.1 Definition

The phrase "Big Data" has dozens of various definitions offered by a wide range of organizations. In the past, the term "Big Data" was utilized for the vast amount of available information which was collected in the digital world. Today, Big Data is recognized as the rising power of the 21st century. Big Data is not just a buzzword; it is a disruptor in the technology channel.

In the book "*Big Data: A Business and Legal Guide*", James Kalyvas and Michael Overly explain Big Data as:

"Big Data is a process to deliver decision-making insights. The process uses people and technology to quickly analyze large amounts of data of different types (traditional table structured data and unstructured data, such as pictures, video, email, transaction data, and social media interactions) from a variety of sources to produce a stream of actionable knowledge." [1]

Figure 1. Big Data Symbol [2]

This definition is practical and descriptive at the same time. Big Data represents itself a process that outcomes into information which contributes decision making. Big Data cannot be measured only by the amount of the storage space or power that it consumes, it should be measured by the impact it brings into every sphere of our lives. Big Data allows us to gain significant meaning from the data at a lowest cost which was historically possible without buying expensive devices for capturing and collecting information.

One of the main qualities of Big Data is an ability to transform. The purchase and utilization of complex software or expensive techniques for collecting data for making usage of it are not needed anymore. Big Data with a combination of cloud and mobile worlds changes the game for businesses of all sizes and types.

2.2 Objectives

1. **Cost Reduction** – Big Data brings cost reductions and significant improvements to carrying out the assigned tasks in the specified time by implementing information technologies. Big Data gives organizations and companies an opportunity to reach a variety of goals by utilizing its technologies and concepts. The selected objectives have influence not only on the outcome and financial benefits, but also on the process.

2. **Time Reduction** – Big Data technologies are utilized for optimization of working processes which leads to the reduction of time to implement the assigned tasks.

3. **Developing New Big Data-Based Offerings** – Developing a wide range of product offerings and featuring by using Big Data and data scientists.

4. **Supporting Inside Business Decisions** – Big Data is applied as a structured data source for internal decision-making of companies.

2.3 Timeline

The definition of the big data has been modified through years since the new technologies and applications were introduced. During the last 20 years, the Big Data has fundamentally modified due to the way the data is utilized. The sequence of the events that had an influence on the Big Data era is presented in the Table 1.

Table 1. Big Data Timeline.

Year	Event
1991	- The Internet (World Wide Web) is introduced - The Hypertext Transfer Protocol (HTTP) is a standard for sharing information in WWW
1995	- The release of Java Platform, Java becomes the second most popular programming language after C - Global Positioning System (GPS) is utilized fully
1998	- The creation of the open-source relational database named NoSQL by Carlo Strozzi - The foundation of Google by Larry Page and Sergey Brin
1999	*Kevin Ashton coins the term "Internet of Things"*
2001	Wikipedia is introduced to the world.
2002	The Bluetooth v. 1.1 is implemented by the Institute of Electrical and Electronics Engineers or IEEE.
2003	➢The research of IDC and EMC shows that the amount of data produced in 2003 exceeds the amount of data produced in the whole human history heretofore; ➢The launch of LinkedIn.
2004	➢The foundation of Facebook by Mark Zuckerberg and his teammates.
2005	➢The Apache Hadoop project is implemented by Mike Caferella and Doug Cutting. The National Science Foundation or NFS is recommended to create professions such as high-quality data scientists to maintain the growing amount of digital information.
2007	The release of iPhone.
2008	*The amount of the different devices connected to the Internet outreaches the world's population.*
2011	- IBM's Watson computer analyzes four terabytes of data just in few seconds to conquer two human players. - UnQL is implemented for NoSQL databases. - IPv4 address space has been utilized.
2012	- The Big Data Research and Development Initiative begins, "Core Techniques and Technologies for Advancing Big Data Science and Engineering" article is published by NSF.
2013	The Big Data becomes part of our lives, since our smartphones, tablets, Wi-Fi originate data at astonishing rates. People start to access large volumes of public data and use it creatively.

2.4 Characteristics: The 3 V's model

In 2001, Doug Laney introduced the 3 V's model of Big Data in Meta-Group publication, marking its three main components: the excessive volume of data, the wide variety of different types of data and the velocity at which data is proceeded. The 3 V's mode notably identifies the traditional ways applied for gaining, storing, utilizing and analyzing data.

Figure 2 shows the expansion on the 3 V's dimensions. Combined together, the new ways of dealing with the data are necessary which leads to creating of a new industry.

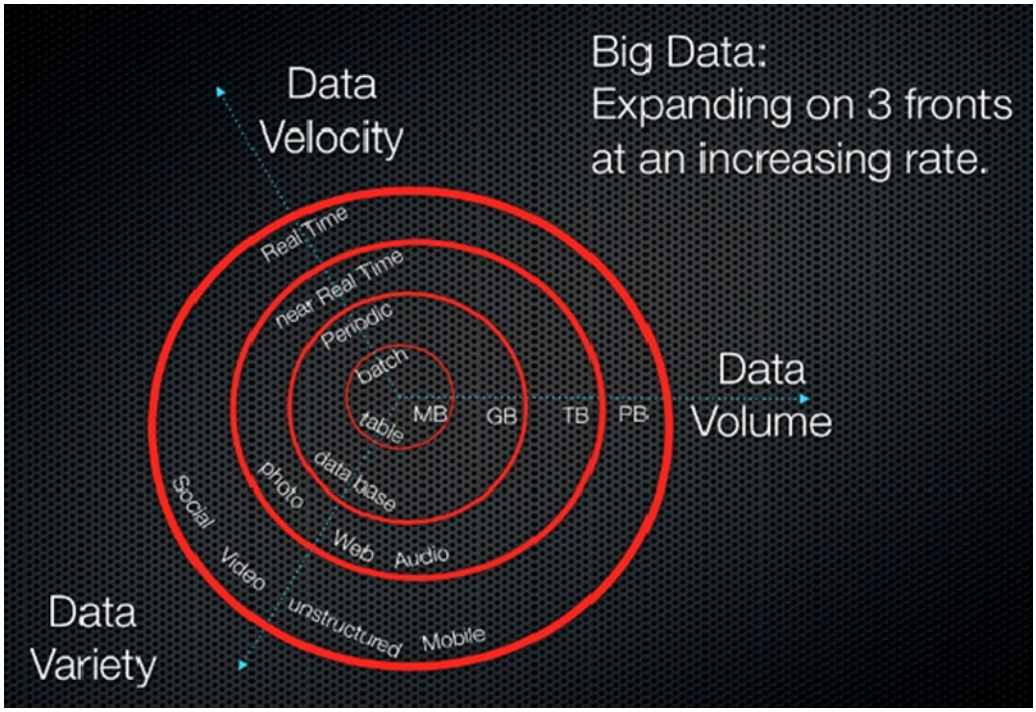

Figure 2. The visualization of Big Data [3]

2.4.1 Volume

The volume of data provided to the enterprises constantly increases since 2004. For instance, the total amount of data collected from the whole Internet in 2004 was one petabyte. One petabyte is tantamount to the collection of all television content for 100 years. In 2011 the

global quantity of information stored electronically reached one zettabyte (which is equal to one million petabytes or collection of high-definition videos with the duration of 36 million years). In 2015, the volume of the Big Data will achieve 7.9 zettabytes (7.9 million petabytes). As the predictions for the future, in 2019 the Big Data is supposed to reach one yottabyte which is an equivalent of one thousand zettabytes.

The size of the datasets which are utilized for Big Data outreached their capabilities of traditional data tools for simple maintenance of data. The visualization of the volume that Big Data takes is presented on the Figure 3.

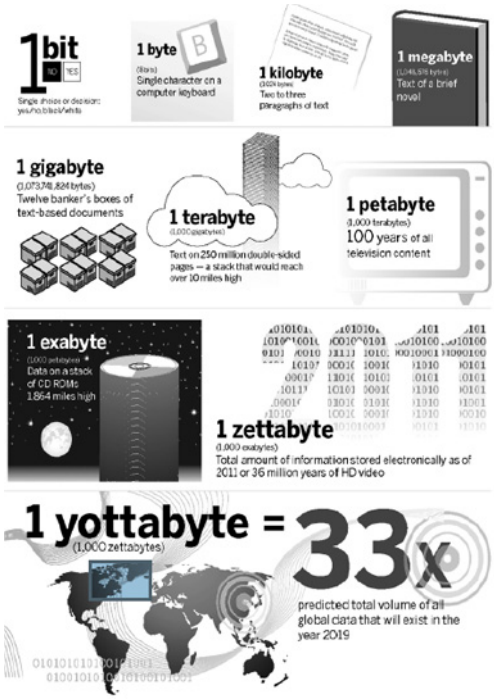

Figure 3. The visualization of Big Data [4]

2.4.2 Velocity

Velocity represents itself the speed at which the data must be kept and analyzed. The data can be described as fleetingly increasing quantity of unstructured data from ever-growing amount of sources streams across the Internet.

2.4.3 Variety

Big Data maintains the data analytics by extending the diversity of useful data to analyze. It unites the value of the data which is kept in the traditional structured databases with the significant value of novel data which is accessed from the unstructured data. The term "unstructured data" is implied to constantly growing amount of data which is not ordered. The unstructured data contains the content created by the user from the social nets, images, videos, sensor data, government

[4] Dean, Big Data, Data Mining and Machine Learning, Value Creation for Business Leaders and Practitioners, 2014

data, web files, researches, etc. Nowadays, approximately 95% of whole data existing on the Earth is considered to be unstructured. These data sources can make business intelligence highly essential.

2.4.4 A fourth V: Veracity

The amount of Big Data is increasing right along; thereafter our knowledge about Big Data deepens. IBM has applied a fourth V to the 3 V's model – veracity. Veracity is applied to the indeterminacy of data such as noise, infraction or offset which leads to deficiency and incompleteness of data.

2.4.5 The 5 V's of Big Data: Validity

Nowadays, the previous four dimensions of Big Data exceed the limits with what the traditional analytics system can handle efficiently and timely. Thus, the fifth V is introduced to the Big Data's model – validation.

Validity of Big Data means the accuracy and correctness of data for consigned usage. By utilizing data mining and analytics, the significant business information can be gained from structured, unstructured, data flow and data warehouses. The given business intelligence can be applied for developing program planning, tracking sales, following the marketing activities, measuring the efficiency between the channels, etc. Data strategy provides business the opportunity to have a better analysis of this data with the purpose of progressive profitable growth.

2.5 Applications around the world

Big Data is a top notch of composition of software development, market growth, demand of the business client and potential consumer.

Social media companies

Several popular consumer companies globally increased their usage of Big Data. For instance, Facebook utilizes Big Data for tracking us-

er's behavior in the social net. Moreover, the company offers your new friend recommendations by finding out whom else you might know. The more friends you make, the more likely you are to keep up your Facebook account. Having more friends means you can see more content, share more photos and videos and post bigger amount of status updates.

Another great example is business networking website LinkedIn which utilizes Big Data for matching job seekers with their potential employers. Headhunters can select and contact the company's hiring manager by simple search. Both of these companies became public in the last few years – Facebook stock started to be available on NASDAQ, LinkedIn on NYSE. Whilst Facebook, LinkedIn and Google are consumer companies on the surface, at the core they belong to really large Big Data companies.

In fine, venture capitalist of Silicon Valley begun to find Big Data companies like never happened before. Nowadays Big Data has several startups with which help Silicon Valley hopes to take control over Wall Street in the next few years.

Google

If there is any technology company in the whole world that demands and epitomizes Big Data so much, it can be only the search engine giant Google Inc. According to the official records of Google, the company maintains at least one trillion of queries per year.

By storing the information from the search results, Google gets information about individual's search behavior. This gives Google an opportunity to optimize advertising and monetize web traffic as no any company in the world could do. To maintain the data, Google's engineers came up with the innovative Big Data technologies.

Financial Services

The providers of financial services utilize big data analytics to develop their analysis of clients to determine customer's edibility for capital, credit, insurance and so on.

Healthcare

Healthcare services work with electronic health records of their patients which they receive from various sources such as treatments, demographics and imagery. Pharmaceutical companies create their own Big Data solutions for monitoring drug effectiveness and provide fast and more efficacious drug development.

Telecommunications and utilities

Telecommunications and utilities use Big Data in order to explore user behaviors and demands for implementing better electrical grid. Moreover, they process environmental sensor data to figure out the lacks of infrastructure and offer more productive risk management intelligence.

Airlines and trucking companies

Airlines and trucking companies utilize Big Data for monitoring fuel consumption and traffic across the regions in real time to lower costs and become more efficient.

2.6 Worldwide Big Data Revenue by Vendor in 2013

When it comes to the information technologies, the most leading and perspective companies invest as much as they can into the information as much as in technology. Getting the value from the data results into a business which makes profit not only from the technologies, but from the information (Table 3).

Table 3. 2013 Worldwide Big Data Revenue by Vendor ($US millions) [5]

Vendor	Big Data Revenue, $US millions	Total Revenue, $US millions	Big Data Revenue as % of Total Revenue, %	% Big Data Hardware Revenue, %	% Big Data Software Revenue, %	% Big Data Services Revenue, %
IBM	1368	99751	1	31	27	42
HP	869	114100	1	42	14	44
DELL	652	54550	1	85	0	15
SAP	545	22900	2	0	76	24
Oracle	491	37552	1	28	37	0,36
SAS Institute	480	3020	16	0	68	32
Cisco Systems	295	50200	1	72	12	16
Microsoft	280	83200	0	0	63	37
Amazon	275	70	1	0	0	100
Hitachi	260	89,999	1	0	0	100
Google	175	59767	1	0	0	100
Intel	165	52708	1	66	21	13
Red Hat	109	1437	8	0	78	22
VMware	80	5,207	1	0	79	21
Juniper	28	4669	1	82	0	18
Total	18607	n/a	n/a	38	22	40

2.7 Big Data in Finland

Various Finnish organizations are evidenced to use Big Data as a key area in such fields as: information technologies (especially in data compilation, data mining and cloud computing), information and communication technology (also including Internet of Things) and information security:

1. ***Cloud Software Finland*** – Cloud Software Programme is created for the development of the competitiveness of Finnish software industry in the world market. It offers novel cloud business samples, public cloud software infrastructures and software organizations samples.
2. ***The national Internet of Things project of Finland*** – national strategic R&D project which includes more than 250 scientists and experts.
3. ***Data to Intelligence*** – the programme which key issues are Big Data, data storages and client-oriented service development. The aim of the project is to create and develop intelligent tools and ways for maintaining, applying and processing data of various types to implement progressive business samples and services.
4. ***Finnish Information Security Cluster*** – an organization found by most of the Finnish information security enterprises to develop their business in national and global markets.

[5] Kelly, Big Data Vendor Revenue and Market Forecast 2013- 2017, 2014

Examples of Finnish companies that use Big Data as their basis:

1. ***Center for Science Ltd.*** – government-supported company that offers IT support and resources for other companies, research institutions and universities. The organization provides high-class software and databases, calculative resources and IT maintenance. Nevertheless to say, CSC signed multi-million euro contracts with Fujitsu and Data Direct Networks for developing Big Data needs.
2. ***The product services of Fujitsu at Finland*** – the company develops Big Data storage solutions with high efficiency calculation. Moreover, Fujitsu created Data Utilization Platform Services that are implemented by cloud.
3. ***Data Direct Network***s – an experienced in Big Data field company that develops multilevel storage of Big Data for scientific computation environment.
4. Big Data Solutions O y – the company implements information analytics by utilizing Big Data technologies and Cloud Business Intelligence.
5. **Capgemini** – Finnish company that offers clients to obtain better value form the corporative and external data by utilizing anomalous behaviour detection, business Data Lake, cybersecurity, data warehouse optimization, enterprise data hub accelerator and in-memory solutions.

3 LANDSCAPE

Big Data does four operations with data: storing, processing, analyzing and visualizing. The Big Data environment starts with infrastructure by choosing appropriate tools for maintaining data. The distinctive analytics tools provide the search of insights of the data. Hereupon, the applications are launched for working with data. Finally, all of these important components create the Big Data ecosystem in total (Table 4).

Table 4. Big Data market segmentation [6]

HARDWARE	BIG DATA DISTRIBUTIONS	DATA MANAGEMENT COMPONENTS	ANALYTICS and VISUALIZATION	SERVICES
1. Storage 2. Servers 3. Networking Examples: • Dell • IBM • Cisco • HP	1. Community Hadoop distributions 2. Non-Hadoop framework 3. Enterprise Hadoop distributions Examples: • IBM • MapReduce • Microsoft	1. Data integration 2. Data quality and governance 3. NoSQL Examples: • Data Stax • Syncsort • Informatica • IBM	1. Data visualization instruments 2. Analytic development platforms 3. Business intelligence applications 4. Advanced analytics applications Examples: • SAS Institute • Datameer	1. Consulting 2. Training 3. Software maintenance 4. Hardware maintenance 5. Hosting/Cloud Examples: • Amazon • Think Big Analytics • Services connected to enterprise distributions (e.g., Cloudera)
Next Generation Data Warehouse: • MPP, columnar data warehouse appliances • In-memory analytics engines Examples: • SAP • Teradata Aster Data • Microsoft • IBM Netezza				

3.1 Applications

Figure 4 displays the landscape of the companies that have an influence on the ecosystem of Big Data. The companies represent themselves as players of the Big Data space. Applications are Big Data business, making profit out of offering analyzed Big Data to end-users. The several examples of areas where the applications are utilized are health, retail and energy.

[6] Business Analytics 3.0 "Analytics-as-a-Service", What is a "Hadoop"? Explaining Big Data to the C-Suite, 2011

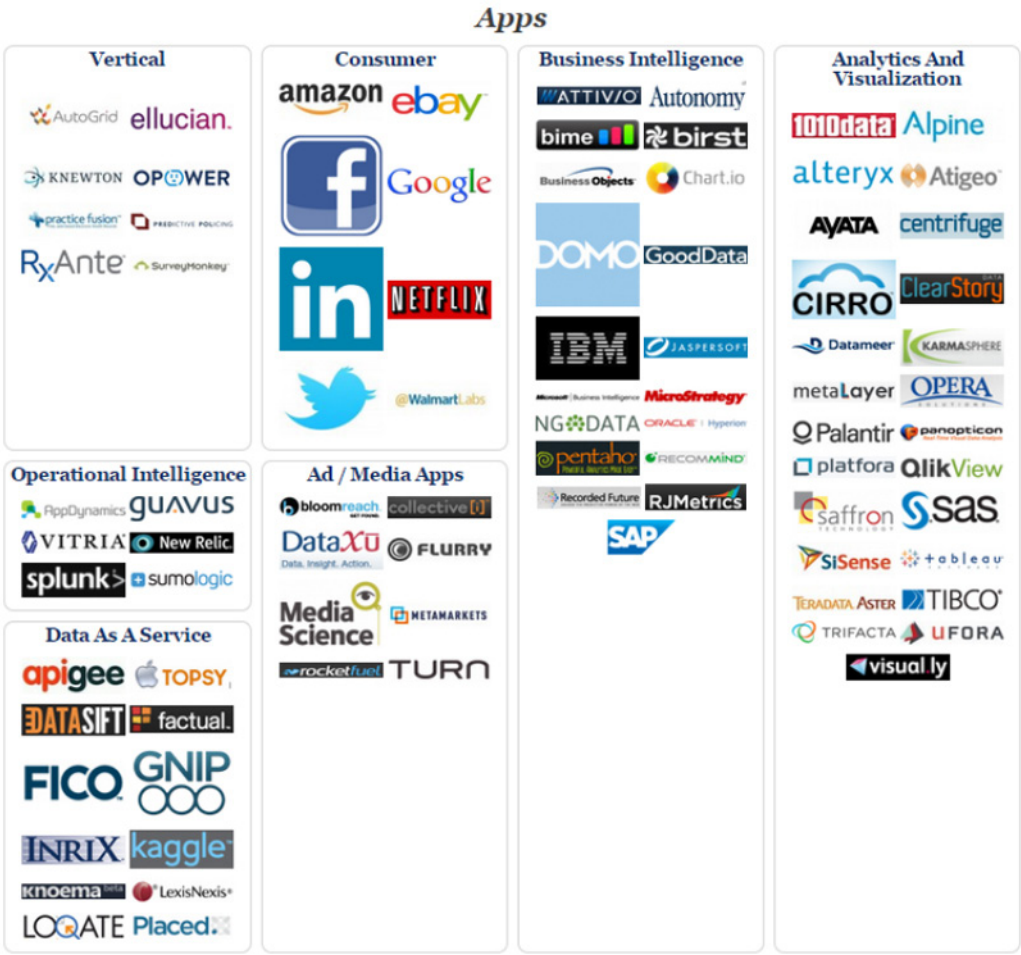

Figure 4. The landscape of the companies in the Big Data space [7]

3.2 Infrastructure

Infrastructure is the keystone of Big Data ecosystem. The main purpose of infrastructure is to store, process and analyze massive amounts of data that companies gain (Figure 5).

For a long time enterprises relied on the relational databases for maintaining structured data. However, due to the volume, velocity and variety of data, relational database cannot handle full performance which is necessary for operating with massive, complex data. Thereafter, new infrastructural technologies are launched that are capable of working with a variety of data, making it easy for systems to operate with thousands of nodes, potentially including thousands of terabytes.

[7] The Official Website of The Big Data Landscape, 2015

Figure 5. The infrastructure component of Big Data ecosystem [7]

3.3 Technologies

The examples of Bid Data Technologies are presented on Figure 6.

➢ ***Hadoop*** – an open-source framework which maintains processing, storing and making analysis of data by using Java as default. Its main principle is to break data into several parts and deliver them simultaneously, thereafter permitting processing and analyzing various parts at the same time. Hadoop is utilized for cheap commodity hardware.

- *MapReduce* – utilizes numerous analytic functions by analyzing dataflow in parallel before "decreasing" the results. "Map" stands for propagation and allocation of a query to nodes, "Reduce" – collects results and unites them into unique value.
- *YARN* – implements cluster management and timing client applications

➢ ***NoSQL*** – a broad set of various database technologies that operate with dynamic semi-structured data, making it be more comfortable for Big Data to work with.

- Apache Cassandra – an open-source database system that maintains data across numerous storage servers, providing high accessibility and low chance of failure.
- MongoDB - a cross-platform database system that operates with document-oriented information.
- Oracle NoSQL – a key-value database which supplies highly secure, accessible and reliable data storage through an adjustable suit of systems that operate as custody nodes.

➢ ***Massively Parallel Processing*** – the technologies that process massive quantity of data in parallel by using SQL as default. MPP is utilized in expensive data warehouse applications.

[7] The Official Website of The Big Data Landscape, 2015

- Berkeley Open Infrastructure for Network Computing – free intermediate system for volunteer computing and grid computing.

➢ ***Cloud*** – a wide range of products that are utilized as services and delivered through network.

- BigQuery – Google's Cloud computing product, which was specially created for maintaining Big Data.

Figure 6. The technologies of Big Data [7]

[7] The Official Website of The Big Data Landscape, 2015

4 EXPLOITING BIG DATA IN ENGINEERING COMPANIES

4.1. Small Data and local companies

Big Data is not used for all Internet of Things cases. Local companies and organizations analyse relatively small datasets for optimizing manufacturing and business processes in order to save significant amount of money.

"Small data connects people with timely, meaningful insights (derived from big data and/or "local" sources), organized and packaged – often visually – to be accessible, understandable, and actionable for everyday tasks." [8]

Small data identifies what the object is doing. However, to understand why the object is doing it, you should use Big Data.

4.2. Finnish engineering company as a case example

To provide anonymity, the company's name is not specified in the publication. The Case Company is perspective Finnish company that has a long history. The company's area of work is designing and manufacturing engineering products. The company takes the leading position in its industry in Finland and is considered to be one of the most high-end companies in its field.

The Case Company Features:

1. Manufacturing with efficient usage of material and energy
2. Production control in real time
3. 3D design and excellent calculation base
4. Customer-focused approach
5. Innovation and development
6. Minimal influence on the environment

However, the Case Company is referred to Big Data Company that utilizes Big Data technologies and Internet of Things technologies without knowing it.

[8] Small Data Group, Defining Small Data, 2013

One of the most remarkable examples of organizations that actually use Big Data Analytics technologies for developing its exploration and production is Hitachi. One of the Hitachi's areas of work is matching working field of the Case Company.

Hitachi purchased an open source of business intelligence from Pentaho for implementing Big Data and Internet of Things technologies. The price of the acquisition was between 500 and 600 million of dollars. The aim of the purchase is a part of novel global management strategy which leads to growth of the company's social innovative business in every sector of work.

"By harnessing the expertise across the Hitachi companies to deliver social innovation solutions - software, services, deep industry expertise, and the consulting needed to launch complex solutions - we can quickly accelerate our global growth. Ultimately our goal is to deliver business and social outcomes to help our customers win, while making the world healthier, safer, and more secure," [9]

- Jack Domme, the US CEO of HDS,
new chief of Hitachi America

[9] Official website of Hitachi, Hitachi Appoints Jack Domme as Chief Executive for the Americas, 2015

5 FUTURE POSSIBILITIES OF BIG DATA

5.1 New applications for the Case Company

As an example of new application for the Case Company can be usage of Prescriptive Analytics of data collected from production, process of work to forecast possible failures and determine actions to decrease the production loss to make production rate more efficient. Moreover, Big Data can offer the wider range of the production line with the same technologies.

5.1.1 3 Steps for implementing Big Data in the Case Company

Figure 7 represents the general scheme of Big Data application in the Case Company.

The process of interaction between Big Data and the Case Company:

1. Intelligent Engineering products send sensor data to Hardware of the Case Company which maintains source data into the relational databases. Moreover, the Intelligent Engineering products get customer feedback in the process of work.
2. Big Data Analytics Platform processes the machinery data from relational databases and makes analyses for maximizing production rate, minimizing amount of the required resources and lowering environmental impact. Thereafter, the Case Company has the information about the records of transactions, sensor data, customer feedback, machine behaviour, security threats, fraudulent rate, etc.
3. Next, the Case Company is able to control, manage and make developments in the infrastructure process.

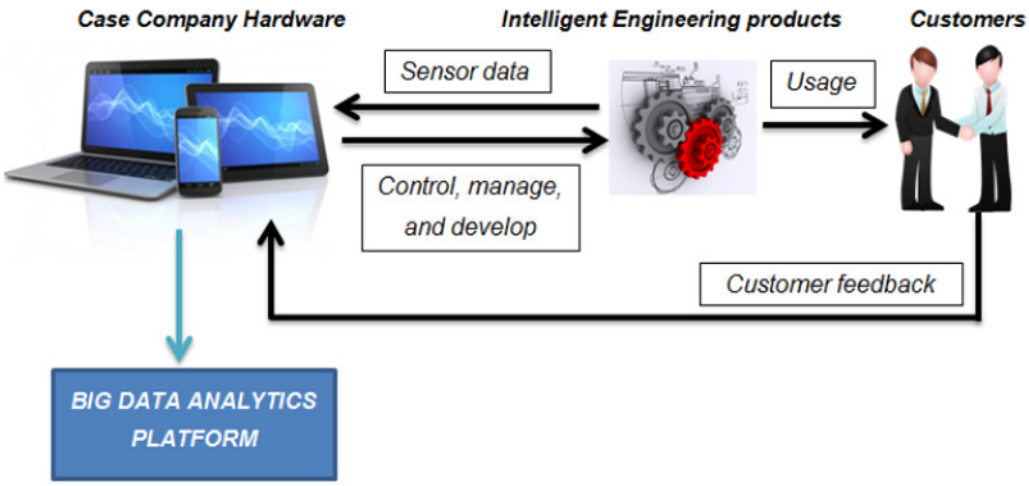

Figure 7. General Scheme of Big Data application in the Case Company

5.1.2 Technical scenario

The technical scenario of implementing Big Data in the Case Company is displayed on the Figure 8. For instance, open-source Hortonworks Sandbox which is based on Hadoop Big Data technology can be used for analysing the data.

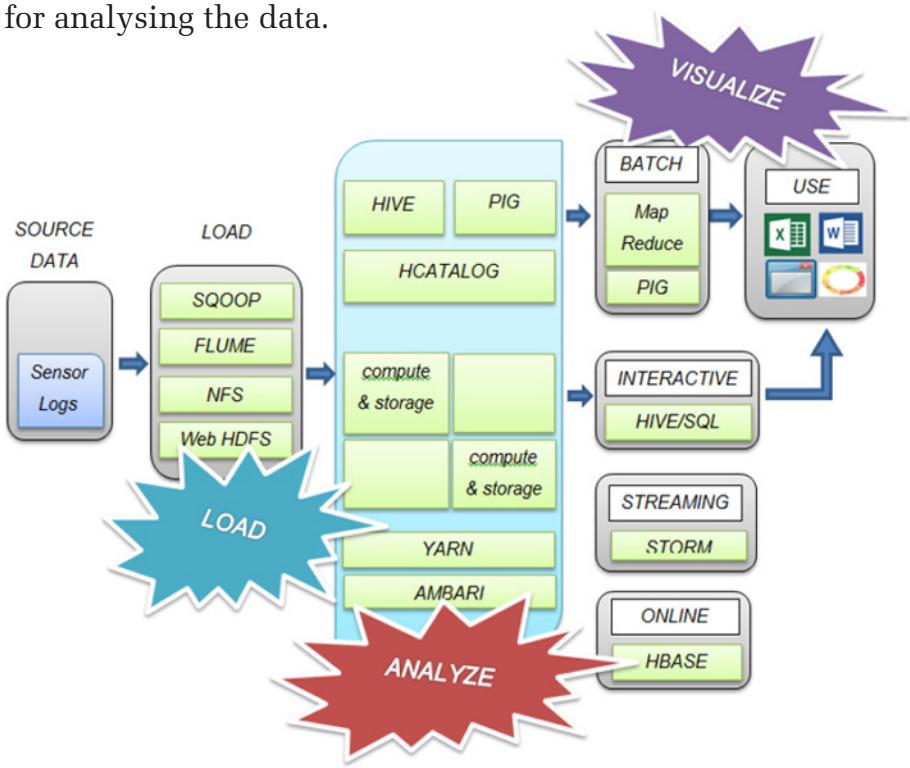

Figure 8. Technical Scenario [10]

[10] Hortonworks, How to Analyze Machine and Sensor Data, 2015

The principle of work:

1. Sensor Logs contain machinery data collected from Intelligent Engineering Products.
2. The flow of sensor data is stored in HDFS (Hadoop Distributed File System) with Flume (service which allows data flow from source into Hadoop Environment).
3. Thereafter, the structured data is forwarded to HDFS by utilizing Sqoop (a tool that transfers data between Hadoop and relational databases).
4. HCatalog (Hadoop layer which provides table data access for PIG, MapReduce, Sqoop, etc.) is responsible for building data's relational view.
5. HIVE (default standard for SQL queries) and Pig (data access engine for implementing difficult MapReduce transformations) scripts are utilized for data processing or structuring data.
6. With the help of ODBC (Open Database Connectivity) connector, the data can be transmitted into Microsoft Excel.
7. Finally, the data is visualized by utilizing Power View (the feature of SQL Server for visualizing, exploring and presenting data in various types, e.g., Excel).

Smarter decisions lead to fewer amount of resources, lower environmental influence and higher revenue. Successful and prospective companies know how to estimate resources to work with, produce and transport their production in the most effective and safest way. Big Data and Prescriptive Analytics will take a leading role in the industry in the upcoming future.

5.1.3 Investment scenario

When we talk about business scenario of Big Data technologies implementation in the Case Company, we should take into account the prices of:

- *The Software for Big Data Analytics* (e.g., free *Hortonworks HDP 2.2.4 Sandbox environment*; however, the costs for the Enterprise Support for the cluster depend on its size. Standard configuration represents itself eight nodes of Enterprise Plus at the price of approximately 30 thousand euros);

- *The official Operating System* (Windows 8 or later version) for utilizing Excel, Word and other programs for visualization of data analysis;
- *The Hardware utilized for Big Data Analytics* (the required hardware is 64bit or x64 processor);
- *Sensors in the Intelligent Engineering Products*;
- *The cost of the programmer's labour* (the specialist controls and manages Big Data Analytics Tool);

5.2 Big prospects for Big Data

Big Data is one of the leading and most efficient game-changer. By collecting information from all three primary categories of Big Data and analyzing data from various data sources and data types, Big Data Analytics helps to prescribe how to maximize the production with minimal environmental impact.

Big Data Analytics:

1. **Descriptive Analytics –** the category of Big Data Analytics which describes what has happened.
2. **Predictive Analytics** - the category of Big Data Analytics which describes what will happen.
3. **Prescriptive Analytics** - the category of Big Data Analytics which describes what will occur; when, why and how to improve the predicted future.

By gaining value from the massive amount of data, Prescriptive Analytics can discover the incredible key insights, find out possible problems and opportunities, and develop the best plan of action. Thereafter, it leads to maximizing production while minimizing cost and avoiding the environmental impact.

According to the interview of the Case Company, the following opportunities that Big Data can bring to the company for its development and settlement were revealed (Figure 9).

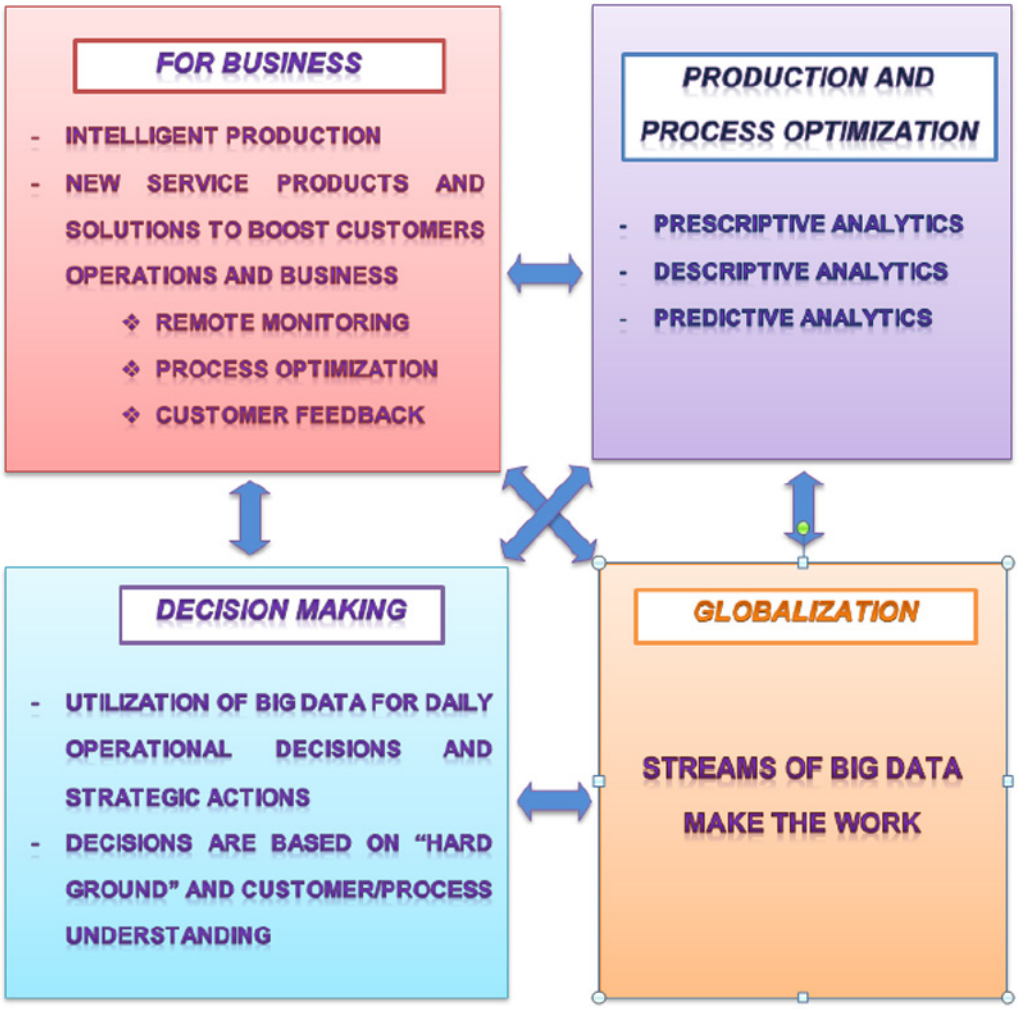

Figure 9. The possibilities of Big Data for a company

REFERENCES

[1] James Kalyvas, Michael Overly, Big Data: A Business and Legal Guide, CRC Press, (2015), pp. 1-5

[2] The Georgia Contractor, How big data is transforming the construction industry, (2015), [reference made 07.04.2015]. Available at: http://thegeorgiacontractor.com/how-big-data-is-transforming-the-construction-industry/

[3] Margaret Rouse, Essential Guide, An architect's guide: How to use big data, Big Data, (2014), [reference made 07.04.2015]. Available at: http://searchcloudcomputing.techtarget.com/definition/big-data-Big-Data

[4] Jared Dean, Big Data, Data Mining and Machine Learning, Value Creation for Business Leaders and Practitioners, Wiley, (2014), pp. 3-5

[5] Jeff Kelly, Wikibon, Big Data Vendor Revenue and Market Forecast 2013- 2017, (2014), [reference made 08.04.2015]. Available at: http://wikibon.org/wiki/v/Big_Data_Vendor_Revenue_and_Market_Forecast_2013-2017

[6] Business Analytics 3.0 "Analytics-as-a-Service", What is a "Hadoop"? Explaining Big Data to the C-Suite, (2011), [reference made 07.04.2015]. Available at: https://practicalanalytics.wordpress.com/2011/11/06/explaining-hadoop-to-management-whats-the-big-data-deal/

[7] The Official Website of The Big Data Landscape, (2015), [reference made 07.04.2015]. Available at: http://www.bigdatalandscape.com/

[8] Small Data Group, Defining Small Data, (2013), [reference made 06.05.2015]. Available at: http://smalldatagroup.com/2013/10/18/defining-small-data/

[9] Official website of Hitachi, Hitachi Appoints Jack Domme as Chief Executive for the Americas, (2015), [reference made 09.05.2015]. Available at: http://www.hitachi.ca/about/press/details/02052015.html

[10] Hortonworks, How to Analyze Machine and Sensor Data, (2015), [reference made 20.05.2015]. Available at: http://hortonworks.com/hadoop-tutorial/how-to-analyze-machine-and-sensor-data/

ADDITIONAL REFERENCES

Elorie Knilans, Avnet Advantage: The Blog, The 5 V's of Big Data, [reference made 07.04.2015]. Available at: http://www.ats.avnet.com/na/en-us/news/Pages/The-5-Vs-of-Big-Data.aspx

Kevin Normandeau, Inside Big Data, Beyound Volume, Variety and Velocity is the Issue of Big Data Veracity, [reference made 07.04.2015]. Available at: http://insidebigdata.com/2013/09/12/beyond-volume-variety-velocity-issue-big-data-veracity/

David Feinleib, Big Data Bootcamp: What Managers Need to Know to Profit from the Big Data Revolution, Apress, (2014), chapter 1, chapter 2.

Danny Sullivan, Search Engine Land, Google Still Doing At Least 1 Trillion Searches Per Year, [reference made 08.04.2015]. Available at: http://searchengineland.com/google-1-trillion-searches-per-year-212940

Eileen McNulty, Dataconomy, Understanding Big Data: Infrastructure, [reference made 11.04.2015]. Available at: http://dataconomy.com/understanding-big-data-infrastructure/

Thomas Davenport, Jill Dyché, Internal Institute for Analytics, Big Data in Big Companies, [reference made 13.04.2015]. Available at: http://www.sas.com/content/dam/SAS/en_us/doc/whitepaper2/big-data-bigcompanies-106461.pdf

JUNIPER NETWORKS, Introduction to Big Data: Infrastructure and Networking Considerations Leveraging Hadoop-Based Big Data Architectures for a Scalable, High-Performance Analytics Platform, [reference made 13.04.2015]. Available at: http://www.juniper.net/us/en/local/pdf/whitepapers/2000488-en.pdf

Big think, The Internet of Things Meets Big Data, [reference made 13.04.2015]. Available at: http://bigthink.com/think-tak/the-internet-of-things-meets-big-data

Invest in Finland, Big Data in Finland, [reference made 21.04.2015]. Available at: http://www.investinfinland.fi/industries/rd-and-innovation/big-data-in-finland/123

Datanami, Fujitsu Lets Big Data Cloud Flag Fly, [reference made 22.04.2015]. Available at:
http://www.datanami.com/2012/01/18/fujitsu_lets_big_data_cloud_flag_fly/

Datanami, Finland's Big Data Storage Leap , [reference made 22.04.2015]. Available at: http://www.datanami.com/2012/03/29/finland_s_big_data_storage_leap/

Quest born to engineer, Product Design, [reference made 20.05.2015]. Available at: http://engineering.quest-global.com/engineering-product-design.html

pixuffle.net, The Importance Of Computers In Our Dailly Lives , [reference made 20.05.2015]. Available at: http://www.pixuffle.net/the-importance-of-computers-in-our-daily-lives/

National Business Research Institute, Customer Service Surveys, [reference made 20.05.2015]. Available at: https://www.nbrii.com/products/customer-surveys/customer-service-surveys/

SAVONIA

BIG DATA AND ITS OPPORTUNITIES FOR THE ENGINEERING COMPANIES

RESEARCH PUBLICATION OF DIGIBOOST PROJECT

The target of this research was to explore the phenomenon of Big Data. The research questions were:

- How does Big Data affect the future?
- What kind of benefits and potential can it bring to the engineering companies?

The research included literature review and case study. The case study was implemented by the investigation concerning Big Data`s potential in a local engineering company.

The research was a part of the DigiBoost - The new possibilities of industrial service business project (2015 - 2016), coordinated by Savonia University of Applied Sciences.The project concentrates on exploring the new possibilities of industrial service business development in close cooperation with the local engineering and energy industries.

The DigiBoost project has been financed by the European Regional Development Fund via the Regional Council of Pohjois-Savo.

European Union
European Regional Development Fund

Leverage from the EU 2014–2020

Regional Council of Pohjois-Savo supports a successful region

DIGI-BOOST

ISBN: **978-952-203-210-2 (PDF)** | ISSN-L **2242-7015** | ISSN **2242-7015** | SERIES **D4/7/2015**

9 789522 032102