

**RESEARCHING THE POTENTIAL OF MOBILE BIG DATA
FOR BUSINESS DECISION-MAKING**
Case of Elgiganten AB - Haparanda

Hung Duc Luu

Bachelor's Thesis
Lapland University of Applied Sciences
Degree Programme in Business Information Technology
Bachelor of Business Administration

2014

School of Business and Culture
Degree Programme in Business
Information Technology

| | | | |
|----------------------------|---|------|------|
| Author | Duc Hung Luu | Year | 2014 |
| Supervisor | Vladimir Ryabov | | |
| Commissioned by | Harri Putila, Elgiganten AB - Haparanda | | |
| Title of thesis | Researching the Potential of Mobile Big Data for Business Decision-Making Case of Elgiganten AB - Haparanda | | |
| No. of pages + app. | 74 + 10 | | |

This research is focused on the area of mobile Big Data, particularly personal location data and its usage to gain insights into customers' in-store behaviours. The objective of this research is to study the practical potential of mobile Big Data in order to evaluate its impact on business decision-making in the case company. This study was commissioned by Elgiganten AB – Haparanda. The case company is a part of Elkjøp Nordic AS, the largest consumer electronics and home appliances retailer in Sweden.

The study stemmed from the need to increase the case company's understanding of the customers' behaviours. This research is practically oriented with the theoretical framework revolving around studying the mobile Big Data, sensors' tracking technologies and Apache Hadoop applications. To accomplish the objectives of this research, the business scenario using mobile Big Data to support decision making in the case company was formulated and scrutinised.

The qualitative research method was used in this single case study for the case company. Exploratory research approach was chosen due to the novelty of the research area. This research makes extensive use of both primary and secondary data. The primary data was gathered through in-depth interviews and questionnaire. The secondary data was collected from established research works conducted on mobile Big Data and its related issues, i.e. Big Data technologies, business value and ethics.

On the basis of the theoretical discussions, the business scenario, and the in-depth interviews and questionnaire, the research results indicate that mobile Big Data can provide various valuable insights into the customers' behaviours in the case company's retail store. Consequently, the business decision-making can be enhanced.

Keywords: mobile Big Data, sensor technologies, personal location data, Apache Hadoop, retail store, business value, privacy

CONTENTS

ABSTRACT

FIGURES AND TABLES

| | |
|---|----|
| 1 INTRODUCTION | 6 |
| 1.1 Background and motivation..... | 6 |
| 1.2 Description of the case company | 8 |
| 1.3 Research objectives..... | 9 |
| 1.4 Structure of the thesis | 10 |
| 2 RESEARCH SCOPE, QUESTIONS AND METHODOLOGY..... | 11 |
| 2.1 Research scope | 11 |
| 2.2 Research questions | 12 |
| 2.3 Research methodology | 13 |
| 3 ONTOLOGY OF BIG DATA IN MOBILE COMPUTING | 15 |
| 3.1 Big Data | 15 |
| 3.2 Mobile Big Data | 16 |
| 3.3 Location data | 17 |
| 3.4 Activity and location tracking sensors | 18 |
| 3.4.1 Global Positioning System (GPS) sensors | 18 |
| 3.4.2 Wi-Fi-based positioning system (WPS) sensors..... | 19 |
| 3.4.3 Accelerometer sensors..... | 20 |
| 3.4.4 Gyroscope sensors | 21 |
| 3.4.5 Other technologies | 22 |
| 3.5 Apache Hadoop | 23 |
| 3.5.1 Apache Flume and Apache Sqoop | 25 |
| 3.5.2 Hadoop Distributed File System (HDFS)..... | 26 |
| 3.5.3 Hadoop YARN..... | 28 |
| 3.5.4 Apache Hive | 30 |
| 3.5.5 Apache Pig..... | 31 |
| 3.6 Mobile Big Data visualisation | 32 |
| 3.6.1 Traffic flow map | 34 |
| 3.6.2 Route map..... | 35 |
| 3.6.3 Density map | 36 |

| | |
|---|----|
| 3.6.4 Single-point motion map..... | 37 |
| 3.6.5 Walking-speed map..... | 38 |
| 3.7 Business value of mobile Big Data..... | 39 |
| 3.8 Privacy and ethics..... | 42 |
| 4 ANALYSIS OF EMPIRICAL DATA..... | 44 |
| 4.1 Analysis of interview data | 44 |
| 4.1.1 The case company IT infrastructure | 45 |
| 4.1.2 The importance of understanding customers' in-store behaviours | 45 |
| 4.1.3 The case company policies to protect customer rights..... | 46 |
| 4.2 Analysis of questionnaire data | 47 |
| 4.2.1 Customers' demographics..... | 48 |
| 4.2.2 Customers' shopping habits and behaviours..... | 49 |
| 4.2.3 Customers' attitude toward location data | 52 |
| 4.3 Summary on the results of data collection | 53 |
| 5 BUSINESS SCENARIO USING MOBILE BIG DATA TO SUPPORT DECISION-MAKING IN THE CASE COMPANY | 55 |
| 5.1 Business scenario overview..... | 55 |
| 5.2 Activity and location sensing systems..... | 56 |
| 5.3 Total solution cost of business scenario in the case company's retail store | 58 |
| 5.4 Findings and recommendations..... | 63 |
| 6 CONCLUSIONS..... | 66 |
| REFERENCES | 69 |
| APPENDICES..... | 75 |

FIGURES AND TABLES

| | |
|--|----|
| Figure 1. Input and output information of mobile location data (Manyika et al. 2011, 92)..... | 7 |
| Figure 2. Hadoop mobile data platform..... | 25 |
| Figure 3. A client writing data to HDFS (White 2012, 71)..... | 28 |
| Figure 4. Apache Hadoop NextGenMapReduce (YARN) (The Apache Software Foundation 2014)..... | 30 |
| Figure 5. The example map of the customers' traffic flow in the store..... | 34 |
| Figure 6. The example map of the customers' routes in the store..... | 35 |
| Figure 7. The example map of the customers' distribution rate in the store..... | 36 |
| Figure 8. The example map of the customers' angular frequency..... | 37 |
| Figure 9. The example map of the customers' walking speed in the store..... | 38 |
| Figure 10. The biz-tech ecosystem reflects the complexity of today's business (Devlin 2014)..... | 39 |
| Figure 11. Respondents' domicile and gender..... | 48 |
| Figure 12. Percentage of respondents' age range..... | 49 |
| Figure 13. Store visiting frequency..... | 50 |
| Figure 14. Respondents' time spent and purposes when visiting the store..... | 50 |
| Figure 15. Respondents' habit when visiting the store..... | 51 |
| Figure 16. Respondents' perception of collecting location data..... | 52 |
| Figure 17. The Aaronia GPS – Logger (Aaronia AG 2011)..... | 56 |
| Figure 18. Components of the total solution cost..... | 58 |
| Table 1. The Big Data appliance total cost (Dijicks 2014)..... | 60 |
| Table 2. Human Actors, Roles and Annual Income in the Hadoop ecosystem | 61 |
| Table 3. The total solution cost..... | 62 |
| Table 4. The costs for individual stores and shared between stores..... | 65 |

1 INTRODUCTION

The background and motivation of this research are presented in this chapter. In addition, the brief description of the case company is discussed. Further, the research objectives are described, and the structure of the thesis is given in this chapter.

1.1 Background and motivation

Along with the rapid development of mobile technologies, mobile Big Data has become an important aspect in the retail industry. Mobile data generated from mobile devices, embedded sensors and mobile applications is proliferating exponentially. Meaningful information extracted from harnessing the immense mobile Big Data can be highly valuable for business. Retailers can use this information to sharpen the business decision-making process.

Location data is a major segment of mobile Big Data. McKinsey Global Institute analysis shows that personal location applications could create up to US \$700 billion in profit for business and consumers by 2020 (Manyika & Chui & Brown & Bughin & Dobbs & Roxburgh & Byers 2011, 92). While most research into location data focuses on its benefits for marketing and advertising purposes, little attention has been paid to its high potential in the retailing sector. Therefore, this research is undertaken to explore and study the potential of mobile Big Data for business decision-making particularly in the case company.

The research is commissioned by the store manager of the case company Elgiganten AB – Haparanda. As other brick-and-mortar retail stores, the case company is in need to understand customers' habits and behaviours in their retail store. This issue can possibly be solved by examining the location data of customers' movements in the store. Figure 1, modified from Manyika et al. (2011, 92), is a sample presentation of input/output information of the location tracking application.

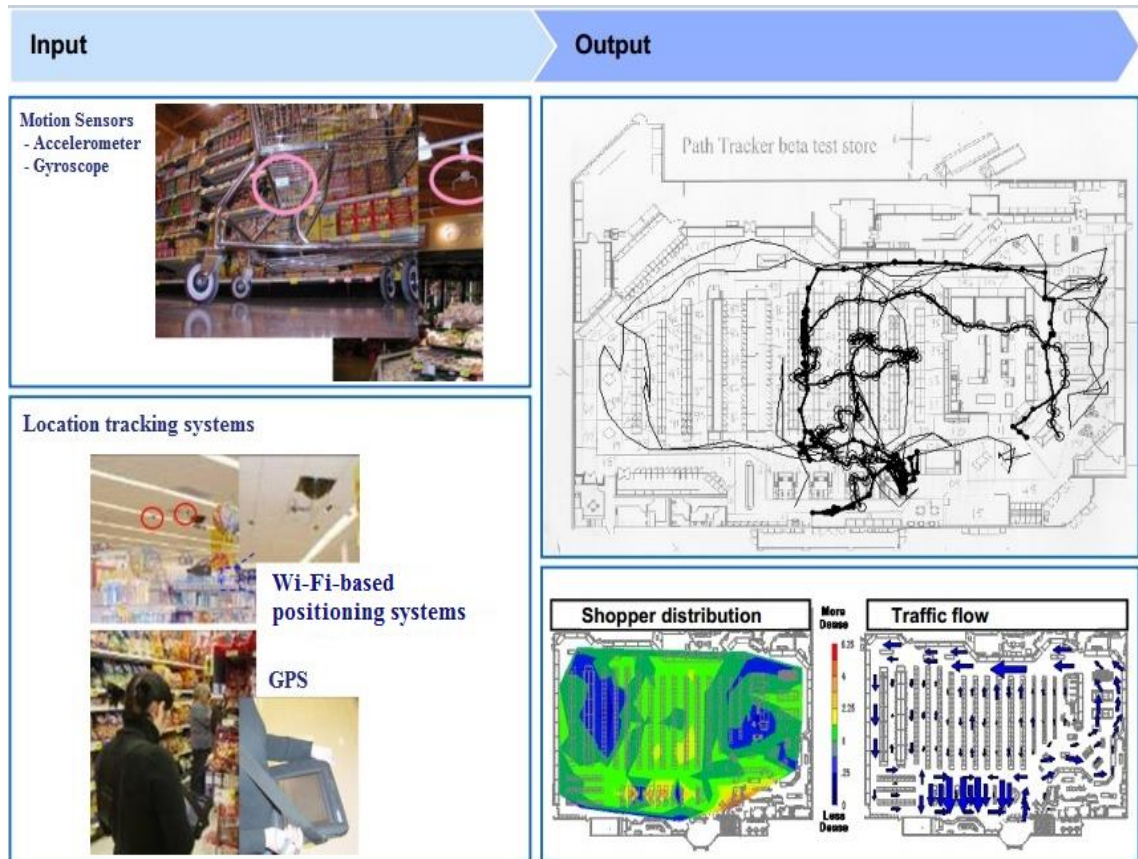


Figure 1. Input and output information of mobile location data (Manyika et al. 2011, 92)

Figure 1 illustrates the process of mobile Big Data studied in this research work. Firstly, mobile activity and location tracking devices are studied in order to evaluate its practicability and efficiency as if it were deployed in the case company store. The quality and nature of data generated from these devices is analysed. Based on the types and characteristics of this data, proper mobile Big Data tools are suggested for processing and managing the data. The collected data is analysed with Big Data analytics to provide insights into the customers' shopping patterns. The outputs of the process are the shopper in-store routes map, the traffic flow and even distribution rate between different sections in the store. By acquiring the information, the manager can enhance several business decisions from optimising the store layout to shelves' placement, sharpening stock management and improving the store services. Therefore, it will give the case company an edge over competitors in the market.

The need to foresee benefits as well as challenges before deploying a new technology is an essential prerequisite for the case company. The case company will find this research work useful in terms of recommendations and consultation on the potential of mobile Big Data to support business decision-making.

1.2 Description of the case company

This section presents the background of the case company initially. Further, the current business challenges of the case company are discussed.

Elgiganten AB is a part of Elkjøp Nordic AS and owned by British Dixons Retail PLC. The company is currently the largest consumer electronics and home appliances retailer in Sweden. Elgiganten AB has 70 retail stores located around the country with stores' floor space ranging in various sizes, from 600 m² to 6000 m². (Putila 2014a.)

Elgiganten AB aims to provide products from famous and trustworthy brands with relatively low price. The extensive range of products gives customers broader and more comfortable choices when shopping in Elgiganten retail stores. Hence, the company always aims to make every customer satisfied with great deals, great services and the industry's best customer guarantees. (Putila 2014a.)

Elgiganten AB - Haparanda (hereinafter EAH) is the case company of this research work. EAH is one of the Elgiganten AB's retail stores located in Haparanda, Sweden and classed as small-size Superstore. There are 17 full-time employees holding different positions working in the store. By having an ideal geographical location, i.e. located near Finnish border, the store has opportunity to not only serve local Swedish customers but also customers from Finland, which account for approximately 50% of the total retail store turnover. (Putila 2014a.)

EAH has around 400 customers visiting the store daily. The demographics of the customers are notably various, major aged from 14 to over 65. In addition, the store serves customers from several countries, e.g. Sweden, Finland, Russia and Norway, who allegedly have diverse shopping cultures. Besides, the case company starts to realise that making customers happy and satisfied is a crucial factor in the retailing today. (Putila 2014a.) Hence, the case company, EAH, is in need to improve the understanding of customers' shopping habits and behaviours. The knowledge can possibly help the case company to serve their customers in a smarter way.

Society for Human Resource Management (2008) points out that the salary for employees cost around 20% of total operating expenses in retail organisations. Mobile Big Data technologies are able to manage and analyse an enormous amount of data with minimum manpower required. Hence, deploying mobile Big Data can facilitate the case company to get additional customers' information at an affordable expense while still keeping human resources balanced. This research, which studies the potential of mobile Big Data, can provide a significant segment in the understanding of customers' shopping behaviours by revealing the customers' movement patterns in the case company's store.

1.3 Research objectives

The objectives of this research are to study the practical potential of mobile Big Data and to find out how to deploy mobile Big Data technologies to gain the most benefit in the case company. In order to attain the objectives, the concept of mobile Big Data is defined. The study delves into feasible mobile technologies, especially sensor technologies for tracking customers' activity and location such as global positioning system (hereinafter GPS), Wi-Fi-based positioning system (hereinafter WPS), accelerometer and gyroscope sensors. Apache Hadoop software used to manage and analyse the location data is thoroughly studied. Additionally, the research determines the business value of the extracted information from mobile Big Data and the impact of the information on the business decision-making process.

Moreover, a business scenario is analysed to offer a comprehensive view of the process of employing mobile Big Data in the case company store. The scenario allows for an investigation into challenges which can occur during the deployment. Furthermore, the scenario provides an insight in how mobile Big Data technologies can be fitted and operated effectively in the case company's store.

A set of recommendations proposed for the case company is the main practical outcome of this work. The outcome aims to pave the way for the adoption of mobile Big Data in the case company. The recommendations are made based on the nature of the store, the characteristics of the generated location data and the budget the case company is willing to invest into the new technologies. By consulting the recommendations, the case company can select the sensible mobile Big Data technologies' approaches to enhance business performance in their store.

1.4 Structure of the thesis

The thesis is divided into seven chapters. A brief introduction to the research work, the case company and the research objectives are provided in this chapter. In Chapter 2, the research scope is narrowed down and the research questions are discussed. In addition, Chapter 2 discusses and provides arguments for the research methodology. Subsequently, Chapter 3 presents the overview of mobile Big Data and different sensor technologies to support activity and location tracking, followed by an in-depth study of Apache Hadoop ecosystem. Further, mobile Big Data visualisation, business value, and privacy and ethics are discussed in Chapter 3. Chapter 4 is devoted to analysing the empirical data from the interviews and the questionnaire in order to understand the company's business and the customers' perception of mobile Big Data. In Chapter 5, the business scenario using mobile Big Data in the case company is developed and scrutinised. Finally, Chapter 6 concludes this research, discusses the results of this work and suggests directions for further research.

2 RESEARCH SCOPE, QUESTIONS AND METHODOLOGY

The scope of this research is discussed initially. Further, the research questions which are formulated to achieve the objectives of the work are discussed. Lastly, the research methodology utilised in this research is presented.

2.1 Research scope

The general aim of this research is to focus on exploring mobile Big Data, particularly personal location data and its impact on business decision-making. The research studies mobile Big Data tools and assesses its suitability for the case company store. The analysis of the empirical data is combined with the selected tools to evaluate its feasibility to apply in the store.

The sensor technologies which provide location data are studied in detail. Nonetheless, the research does not touch upon other location tracking related technologies such as RFID tags and surveillance cameras. The application of location data in retail stores is a relatively novel concept, and it concerns data from customers' in-store behaviours. Therefore, the customers' experience of mobile Big Data adoption is studied in this thesis.

A specific mobile Big Data physical installation cannot be elaborated in the case company because of time limitation. However, the researcher is willing to assist the implementation in the future if required. Besides, due to financial ramifications, the researcher is unable to obtain the equipment required for testing and experimental purposes. Hence, the practical part will primarily be based on empirical data from the case company and analysis of technologies available. Moreover, as mobile Big Data is a dynamic field with intensive revolution, the research validity might be altered in the long-term context.

2.2 Research questions

Three research questions are defined to achieve the objectives of this research work. The necessary steps to answer these questions are addressed below.

1. What are the advantages and challenges in deploying and utilising mobile Big Data technologies and tools in the case company?

Different mobile technologies for location and activity tracking are studied in this research. The technologies include GPS, WPS, accelerometer and gyroscope sensors. Apache Hadoop, the Big Data open-source software, is examined to evaluate the reliability and stability if utilised in the case company. A critical analysis is made from the study of mobile Big Data technologies and the empirical data to find out the most reasonable way of their deployment in the case company.

2. How can the use of mobile Big Data in the case company contribute to the business decision-making process?

This research analyses the business scenario in order to understand the underlying value of mobile Big Data and the impact of mobile Big Data on the case company's retail store. The findings from the business scenario can be used to facilitate the process of applying mobile Big Data in the case company. Moreover, this research question aims to discover any limitations which can occur during the implementation.

3. How do ethical concerns impact the use of mobile Big Data, particularly personal location data?

The research question aims to understand customers' point of view on personal location data. Since customers play a significant role in generating location data, it is important that they acknowledge the importance of mobile Big Data to improve their shopping experience. The questionnaire is utilised to investigate this issue from the customers' perspective.

2.3 Research methodology

The qualitative research method is used for both theoretical and practical parts of the research. The qualitative research provides a great advantage in “conducting formative evaluations, ones that are intended to help improve existing practice rather than to simply assess the value of the program or the product being evaluated” (Scriven 1991 as cited by Maxwell 1996, 21). This research aims to study the potential of mobile Big Data to improve business decision-making. The analyses of various literature and empirical data are utilised to collect the required information for the research work. Therefore, the qualitative research method is well suited.

This thesis involves exploratory research based on the literature. The exploratory research scientifically denotes extensive, purposeful and systematic data collection which aims to expand discovery based on description and direct understanding of a social area (Stebbins 2001 as cited by Given 2008, 327). The exploratory method is relevant since the area of this research revolves around exploring the potential of mobile Big Data in the case company. In addition, the research main outcome is to generate a set of recommendations matched the purpose of exploratory research.

The case study approach is applied to this research work. As stated by Yin (2009, 4), the researcher can “retain the holistic and meaningful characteristics of real-life events such as” organisational processes and the maturation of industries. The case study approach is technically defined as an empirical study which uses to investigate thoroughly a contemporary phenomenon within its realistic context when the boundaries between phenomenon and context are not clearly evident (Yin 2009, 18). In this research work, the case study approach facilitates the researcher to examine the feasibility as well as evaluate the business value of mobile Big Data application in the case company’s retail store. Moreover, the research is conducted as a single-case study for the case company since it allows critically analysing the precise nature of the phenomenon (Yin 2009, 47).

The research utilises both primary and secondary sources. The primary sources of data are derived from interviews and surveys. Relevant books, reports and articles are used as the secondary sources.

In-depth interviews are conducted with the case company store manager. This approach is chosen since the interview is one of the most crucial sources of the case study information. The in-depth interviews with the store manager allow the researcher to ask not only about the facts of the store and its related IT issues but also the opinions about this research for further enquiry. The in-depth interviews take place over an extended period of time. Adequate evidence for supporting answering the research questions can be found from interpreting the interviews. In addition, the store manager can suggest other members of staff to interview in a specific matter, as well as other sources of information. (Yin 2009, 106-107.)

The questionnaire is carried out in order to collect customers' opinions on the use of mobile Big Data. The questionnaire technique is chosen for data collection since it can reach a wide range of potential-customer respondents and receive answers in a short period of time. The questionnaire survey in this research consists of multiple choices, yes/no questions and scale questions along with an open ended question.

The requirements for choosing secondary sources are that it should be relevant and authoritative. Established books and reports from reputable authors are primarily used to conduct the theoretical part of this research. Some of the sources used in this research are acquired from Lapland University of Applied Sciences' e-resources and other established scientific electronic libraries. Literature review is utilised to collect relevant data for supporting the research throughout the entire thesis work.

3 ONTOLOGY OF BIG DATA IN MOBILE COMPUTING

This chapter presents the overview of Big Data in mobile computing, the business value as well as the moral aspects of mobile Big Data. In section 1, a brief description of Big Data is given by addressing Big Data's origin and its characteristics. Mobile Big Data and location data are thoroughly explained in section 2 and section 3, followed by the in-depth study of sensor technologies which can be utilised in the case company in section 4. Apache Hadoop and Big Data tools which are suitable for managing mobile Big Data are presented in section 5. Mobile Big Data visualisation is described in section 6. Section 7 focuses on business value derived from mobile Big Data from organisational and individual perspectives. Lastly, section 8 discusses the privacy and ethics concerning Big Data in mobile computing.

3.1 Big Data

Data represents the lowest raw format of information. In the computer science, data is often illustrated as rows and columns of organised values which depict some entities and its attributes (Dumbill 2012 as cited by O'Reilly Media 2012, 3). Data has always been a vital part of every enterprise no matter of its size. It is pointed out that if data is managed and leveraged properly, it can offer enormous business value for companies. (Mohanty & Jagadeesh & Srivatsa 2013, 1.)

Big Data can be described as massive volumes of disparate data in varying degrees of complexity and ambiguity. Big Data is generated at different velocities which go beyond the ability of typical data processing tools to manage and analyse. In other words, Big Data refers to data sets which have grown substantially in size, become difficult to handle and increasingly hard to extract value from it. Moreover, the concept has evolved to embrace not only the size of the data sets but also the process in leveraging the data. (Ohlhost 2012, 13.)

“Big Data is not a stand-alone technology but a combination of old and new technologies” which allows companies to gain insights into unique patterns of data (Hurwitz & Nugent & Helper & Kaufman 2013, 15). Decision-making process usually utilises lots of human processing and analytic refinement. Big Data technologies offer the automated data processing capability which is scalable, flexible and error-free (Krishnan 2013, 5). A survey of more than 600 business executives by Capgemini (2012 as cited by Akerkar 2014, 9) shows that Big Data is already utilised to support decision makers 58 % of the time and 29 % of the time for decision automation. In addition, the survey demonstrates that over the next three years, Big Data utilisation can enhance companies’ performance by 41 %.

3.2 Mobile Big Data

Mobile Big Data has been emerging noticeably as an essential part in the development of Big Data. Nevertheless, mobile Big Data is still a relatively novel concept, and the exact meaning is often nebulous, as the same way Big Data is often defined. The main factor which differentiates mobile Big Data from Big Data is allegedly the sources of data. Input data of mobile Big Data is primarily from mobile computing such as mobile machine logs, sensors and mobile applications.

Mobile Big Data inherits all the characteristics from Big Data which are volume, variety, velocity and veracity. Mobile Big Data can be seen as enormous amounts of mobile data generated which come with high velocity and in many forms, e.g. structured, semi-structure and unstructured. Due to the nature of mobile data which is inconsistent and latency, mobile Big Data can be uncertain and might lead to statistical errors. Hence, the cleanliness of the collected mobile data is necessary for extracting reliable information.

3.3 Location data

Location data is an important segment of mobile Big Data. The amount of available information contained the personal location data has been increasing substantially in recent years. Mobile technologies such as GPS sensors can instantaneously locate any supported devices with the accuracy within a few meters radius. Therefore, the location data has been using to create a new wave of innovative business models which are particularly suited for the case company's retail store. This revolution of the location data will keep evolving over the next decade. Along with this is the improvement and expanding in the quality, volume and accessibility of the location data. (Manyika et al. 2014, 85.)

The tracking sensors are the primary sources of the location data volume since it has to update the location data frequently. However, the size of total data from the location data is fairly small because the data necessary to capture a fixed location point is only a few bytes. Nevertheless, this small amount of the location data generated implies that this data type can possibly offer higher value per byte generated in comparison with other types of data such as machine logs or video which can go up to gigabytes. (Manyika et al. 2014, 87.) Thus, the case company aiming to use the location data can reduce the cost of storage, which is one of main concerns when deploying mobile Big Data technologies, while still being able to make the most use of it.

Using the location data can support the case company to broaden the understanding of the customers' shopping behaviours in the store. Accumulative information on walking traffic density and speed can provide detailed insights into where customers decelerate or accelerate in response to special offers and new products in the store. These patterns can be linked with data on purchased products, customers buying history and demographics. The granular information allows the store manager to get deeper perception in the customers' references. The business decisions from advertising strategy to merchandising can be enhanced. (Manyika et al. 2014, 92.)

3.4 Activity and location tracking sensors

Sensors can be defined as devices which convert a physical measure from the real world and transform it into a digital signal. Data acquisition systems process and translate the raw signal to data logs as the output. Apache Hadoop software is deployed to extract value information from this output data. In essence, sensors are continually on and capturing data in an efficient way. The fundamental principles of sensors consist of sensitivity, range, accuracy and precision. The sensitivity shows how much sensors' output alters to the changes of a measured quantity. The range illustrates the quantity scope in which sensors are able to measure. The accuracy describes the correctness of the result in compare to the actual or true value. Lastly, the precision depicts the different level when one quantity is repeatedly measured. Besides, sensors have to be calibrated with prerequisite reference standards before being utilised. (Chen & Janz & Zhu & Brychta 2012, 2.)

Activity and location tracking sensors which include GPS, WPS, accelerometer and gyroscope sensors and other potential sensor technologies are discussed in the next sub-sections. The sensors are widely used to measure human physical activities, e.g. location, motion and velocity. The specific objective of the sensor-derived data in this research is to capture the comprehensive portrait of the customers' in-store activities and behaviours.

3.4.1 Global Positioning System (GPS) sensors

GPS sensors work as receivers which determine its position on the Earth's surface. The receivers calculate the different time of receiving the radio signals sent from several GPS satellites to locate the accurate position. GPS sensors deliver certain position data in the form of latitude and longitude. The accuracy of GPS location data is within a radius of 10 meters with the use of Differential GPS. However, the location data can be accurate close to 2 centimetres if equipped special tools near the sensors. In addition, GPS sensors constantly

synchronise with the satellites' atomic clock. Hence, the sensors can provide precise timing up to a nanosecond, which is highly valuable when analysing the customers' real-time location. (Theiss & Yen & Ku 2004, 90-91.)

GPS sensors capture the detail in-store location and walking routes of customers. The location data is analysed with Apache Hadoop software to identify customers' interested routes. Lastly, the location information is matched and visualised with the case company's store floor-plan to provide insights into the customers' movements.

It can be seen that GPS offers a passive, low cost and objective way to monitor actual customers' movement patterns through the store (Duncan & Badland & Mummery 2008, 550). However, there are some practical issues with GPS sensors that need to take into consideration. The GPS signal is degraded, and the accuracy is significantly decreased in the in-door environment. Besides, sophisticated GPS sensors, which provide highly accurate position by using several GPS frequencies, can be costly. Hence, the case company should consider the benefits as well as the limitations when deploying this technology.

3.4.2 Wi-Fi-based Positioning System (WPS) sensors

WPS is one of the indoor positioning systems which is capable of locating and tracking customers' movements. WPS sensors work by analysing the Wi-Fi signals emitted by the customers' Wi-Fi-enabled devices such as smartphones or tablets. When the Wi-Fi is turned on, the devices continuously emit Wi-Fi signals searching for nearby Wi-Fi access points. The Wi-Fi signals are triangulated with WPS sensors located around the stores, and the intensity of these signals are measured to estimate the customers' position to within a few meters. It can be seen that WPS locates the customers' position in the similar way GPS works. However, WPS advantage over GPS is that it can function effectively in the in-door environment. (Clifford and Hardy 2013.)

In essence, each Wi-Fi signal being emitted does not just send the signal strength, but also a unique MAC address of each device. The unique MAC addresses from the customers' devices are collected by WPS sensors. The new MAC addresses are compared with the collected MAC addresses from the database to recognise repeat customers from new customers. With the support of Big Data analytics, it allows the case company to build up behaviour information of the return customers and get further insights into how different the new customers behave in comparison with the return ones. Additionally, the unique MAC addresses can help to understand how frequently customers visit the store as well as how many customers do not enter but just walk past the store. (Clifford and Hardy 2013; Euclid 2014a.)

Furthermore, WPS sensors can calculate the duration the customers spend in the store. For instance, if several customers enter and leave the store within 5 minutes, it might indicate that there are not enough interesting products on the shelves or the products, that the customers look for, are not available. This insight can be used by the case company to adjust the products on display. (Fu 2013 as cited by Martin 2013.)

WPS sensors can be deployed in the case company's retail store in conjunction with GPS sensors to provide ubiquitous and highly accurate location data. WPS is an efficient, low-cost way to aggregate the customers' data since it uses the customers' smartphones themselves to collect the customers' location. It should be noticed that the customers do not need to connect to any specific Wi-Fi networks or install a mobile application. WPS is able to work productively as long as the smartphones have Wi-Fi turned on and in the coverage of WPS sensors. Thereby, the case company can measure and collect the customers' data without interrupting their shopping experience.

3.4.3 Accelerometer sensors

Accelerometer sensors are capable of measuring motion data such as acceleration, vibration (periodic acceleration) and tilt (static acceleration). Micro-

electro-mechanical systems (hereinafter MEMS) accelerometers are the standard motion-detecting sensors since it provides multiple axes, wide dynamic ranges, good sensitivity, low power consumption and minuscule size. (Chen et al. 2012, 5.) Digital output of accelerometer sensors is a variable frequency square wave. In order to be understood and used for further process, the output waves have to be read through the data acquisition system and translated into a scaled acceleration data. The data is passed to Apache Hadoop software for analysis. In addition, the acceleration data itself cannot illustrate an absolute location measurement. It needs to be combined with other location data such as GPS data in respect of time in order to provide meaningful value.

Since accelerometer sensors calculate the intensity of movements, it is applicable to use in the case company's store to provide the changes in motion as well as counting steps of the customers. For instance, the sensors can measure directly the acceleration of the customers' movements. Hence, the information where the customers speed up or slow down because of interesting products or offers in a particular time is presented. Moreover, when customers interacts with demo products, accelerometer sensors can measure the vibration level and combine with time duration the customers spent on these products to assess the degree of their interest in it.

3.4.4 Gyroscope sensors

Gyroscope sensors, also known as angular velocity sensors, are used to detect the rate of the angular rotational velocity accurately even in free space. Angular velocity is defined as the change in rotational angle per unit of time (Seiko Epson 2014). MEMS gyroscopes are widely used in mobile applications due to its simplicity, reliability and low-power consumption. MEMS gyroscopes apply vibrating mechanical elements to sense the rotation based on the transmission of electricity amongst two vibration modes of an object caused by the Coriolis force. (McGrath & Scanail 2014, 50-52.)

The main usage of the gyroscope sensors in this research is to measure the angular velocity and changes in direction produced by the sensor wearers' movement. The angle changed is collected for analyses. For instance, when customers stand in front of the television section, the gyroscope sensors can sense the angular velocity of the customers if they turn to left or right. The information can be collated with the actual store layout, the location and the time data to understand which television brands or products are getting attention from the customers.

As pointed out by GSMArena (2014), accelerometer sensors can measure the directional movement and acceleration of customers but are unable to resolve the lateral orientation during this movement accurately without gyroscope sensor data to fill in. Gyroscope sensors effectively augment the information supplied by the accelerometer sensors by tracking the rotation and adding the dimension data. Therefore, in order to gain an output which is both meaningful and responsive, the data of the accelerometer sensors and the gyroscope sensors should be combined together in the same time.

3.4.5 Other technologies

Besides the major activity and location tracking sensors that studied above, the case company may consider other potential positioning and motion detecting sensor technologies for future reference. In this section, pressure sensors, magnetic sensors and its applicability for in-door use are presented.

Pressure sensors are used to measure the customers' gaits and positions accurately based on their footsteps. Furthermore, the biometric information collected can be used to classify the customers' uniqueness and determine their directions in the store. There is a recent project which deploys pressure sensors called "Smart Floor". In the project, a floor is outfitted with Smart tiles, which consists of pressure measuring sensors. As the customers walk over the tiles, data generated "in response to the weight and inertia of a body" is recorded. The data is utilised to identify the specific customers by comparing the uniqueness of footstep gesture profiles. Additionally, by evaluating time and

force of footsteps impacted on the tiles, the customers' moods can be revealed. For instance, a very high amplitude pressure such as stomp footstep may indicate a bad-temper customer while a quick, prompt footstep may show a hurried one. The advantage of this technology is that the customers do not need to carry any devices which explicitly identify them to the system. However, in order to cover the major floor area such as the case company's store, a considerable amount of pressure sensors would be required. Hence, the case company should take into account the trade-off between the value information this technology provides and its initial investment. (Orr & Abowd 2000; Middleton & Buss & Bazin & Nixon 2005.)

Magnetic sensors can pinpoint the customers' location based on measuring the anomalies in the magnetic fields. The magnetic fields are emerged from both Earth's magnetic fields as well as human sources such as building steel, concrete structures, and electronic appliances. Since each building has its unique spatial magnetic fields, the magnetic differences can be mapped and combined with physical floor plans. As the customers walk around a store, the magnetic sensors can sense the magnetic field and measure the anomalies around it. The information is pushed and analysed with the map database. The reading from the sensors and the map created by taking magnetic fields data from the store are utilised to estimate the customers' position. The accuracy of the results can range from 0.1 meter to 2 meters. The application of magnetic sensors are still being researched and developed since the magnetic fields are unstable, fluctuated and can be affected by many factors. (Belezina 2012; Haverinen & Kempainen 2009.)

3.5 Apache Hadoop

Apache Hadoop is an open source project to manage Big Data in scalable and distributed computing. Apache Hadoop is not just a single stand-alone software but a whole ecosystem of projects which work together to provide a common framework. This framework allows the distributed computing to process large data sets efficiently across clusters of computers by using simple programming

models. Furthermore, Apache Hadoop can transform commodity hardware into a coherent service which is capable of storing reliably petabytes of data. One of the key attributes of Apache Hadoop is redundancy. Since data is held on multiple data nodes, Apache Hadoop can automatically replicate the data if it is corrupted without human intervention. Besides, Apache Hadoop is highly dynamic; users can always get full access to the system with high bandwidth where all machines are available to process the data. Apache Hadoop supports real-time data; nevertheless, it primarily concentrates on batch processing, i.e. when a bulk of data is transferred to the clusters, Apache Hadoop processes the data and sends back the result when completed. (The Apache Software Foundation 2014a.)

By deploying Apache Hadoop software, the case company can get a powerful Big Data management system that provide huge distributed computations without having to pay top-notch systems or special hardware with redundant reliability since Apache Hadoop is already responsible for this reliability. Figure 2 depicts the Hadoop Mobile Data Platform which can be applied for the case company. The platform is recommended based on the characteristics of activity and location tracking sensors which generate mainly semi-structured and structured data. In addition, the Hadoop applications used in the platform are selected with the aims of processing data in an efficient manner by avoiding unnecessary steps and optimising the required output for visualisation.

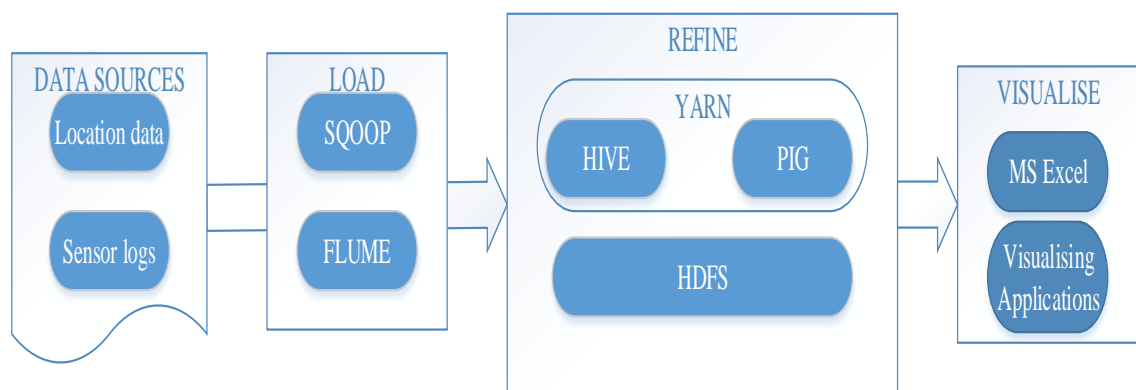


Figure 2. Hadoop mobile data platform

As illustrated above, the platform divides data processing into three main stages: Load, Refine and Visualise. First, the data sources included the

generated location data and sensor logs are transferring via Apache Flume and Apache Sqoop into Hadoop Distributed File System (hereinafter HDFS). HDFS and Hadoop YARN (hereinafter YARN) are the principal components of the platform. HDFS is capable of fault-tolerant, scalable storage of mobile Big Data while YARN provides the resource management and the framework to run Apache Hive and Apache Pig applications (Hortonworks 2014a). Apache Pig and Hive run on top of YARN to explore, structure the mobile data and turn it into useful and concrete information for the last stage. Visualised applications are utilised in the last stage to transform the processed information into the store maps which expose the aggregate customers' movements and activity in the store. The Hadoop applications are discussed thoroughly in the following sub-sections according to the flow of data in the platform: (1) Apache Flume and Apache Sqoop, (2) Hadoop Distributed File System, (3) YARN, (4) Apache Hive and (5) Apache Pig.

3.5.1 Apache Flume and Apache Sqoop

Apache Flume and Apache Sqoop are used to load mobile data into HDFS and related system, e.g. Apache Hive. The specific features and usage of each tool are described in the following paragraphs.

Apache Flume is a distributed and reliable system for collecting, aggregating, and moving efficiently large amounts of data from various sources to a centralised data storage. The system is utilised in this platform to deliver log data, manipulate and drop it into HDFS. (Zikopoulos & Eaton & Deroos & Deutsch & Lapis 2012, 74-75.) Apache Flume contains five major components which are event, source, sink, channel and agent. The Flume events are a singular unit of data flow which allows communication between different components and usually have a few bytes payload. The Flume sources exploit the events provided to it by any external data sources such as sensor data server. The external sources send the events to the target Flume sources in a recognisable format. The Flume channels assess the durability of the events and provide a link for the events between a source and a sink. In this research,

the Flume sinks perform as the final flow destination in which the data flow landed. A Fume agent is any physical Java virtual machine which can host the events flow from an external source to the next destination. A single agent is able to run numbers of sources, sinks and channels between it. Apache Flume ensures the reliability of the delivered events by using a transactional approach. (Atlassian Confluence 2012.)

Apache Sqoop is a tool developed for conveniently transferring bulk data between the structured data storage and Apache Hadoop. It can be used to import tables from the data storage by extracting rows from the tables and write the records to HDFS. Conversely, Apache Sqoop can export the processed data back to the data storage for later use. Sqoop connector is an extension framework which optimises connectivity to any external storage systems supported bulk data transfer. Sqoop connectors can work with several popular relational databases and data warehouse systems such as MySQL, PostgreSQL and Oracle. (Kimball as cited by White 2012, 527-529.)

Mobile data generated from the activity and location tracking sensors is primarily location data and sensor logs. It is suitable to use Apache Sqoop to import the structured data of GPS sensors which contain latitudes and longitudes into HDFS. While the sensor logs from WPS sensors, accelerometers and gyroscopes can flow through Apache Flume for cleansing and aggregating to HDFS for further process. The further process can be implemented with high-level Apache applications such as Apache Hive and Apache Pig.

3.5.2 Hadoop Distributed File System (HDFS)

HDFS is a distributed file system which provides elastic, versatile and clustered way to store and manage very large files with streaming data access patterns in the Hadoop mobile Big Data platform. Unlike conventional distributed file systems which require highly reliable and specialised hardware, HDFS is developed to work on the clusters of low-cost, commodity hardware. Since node

failure is likely to occur in the clusters due to the commodity hardware, HDFS is designed to handle such failure and keep operating without interruption. In HDFS, data sets are written once and read many times thereafter. Hence, Big Data analyses which involve a considerable proportion of the data sets can be performed quickly with low-latency. (White 2012, 43-44.)

HDFS deploys a master-slave architecture. A typical HDFS cluster includes a single NameNode and multiple DataNodes. Blocks are small pieces divided from large files by HDFS. The blocks are stored on DataNodes and replicated across different DataNodes for fault tolerance. The NameNode keeps track of where these blocks are located on DataNodes to make up the complete files. Moreover, the NameNode is responsible for controlling privilege to manage the files such as read, write, execution and delete of the data blocks on the DataNodes. The DataNodes store and retrieve the blocks according to the requests from the NameNode. Besides, the DataNodes periodically report back the lists of the storing blocks to the NameNode. In addition, heartbeat messages are frequently sent from the DataNodes to detect and ensure connectivity with the NameNode. (Hurwitz et al. 2013, 112-114.) A visual representation of the process when a client interacts with HDFS for storing a file to HDFS is presented in figure 3.

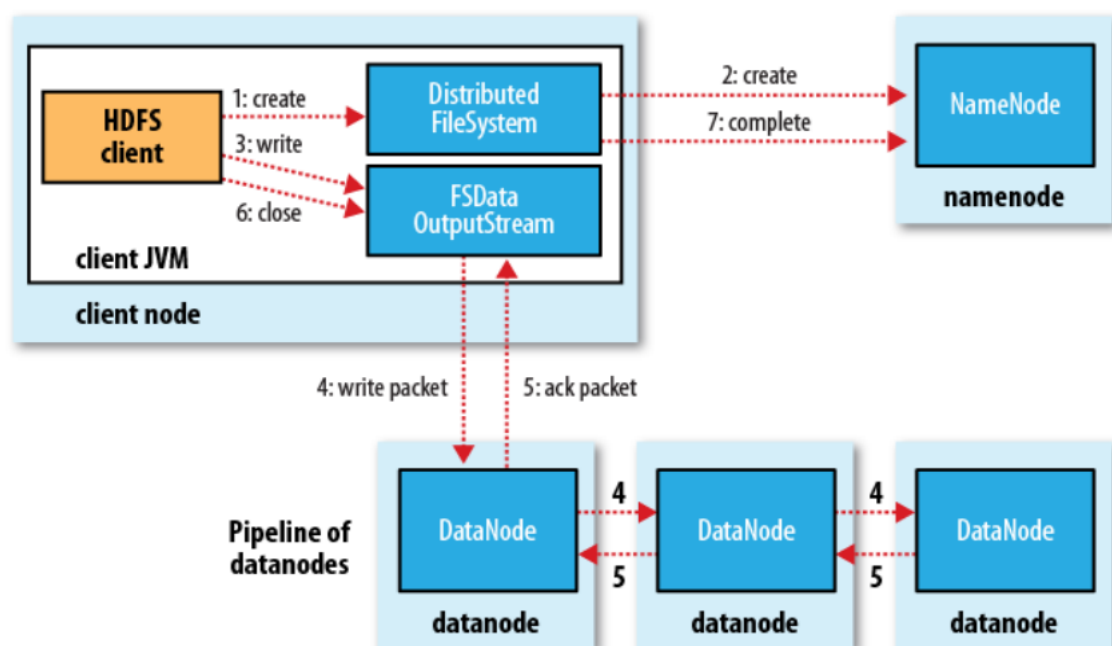


Figure 3. A client writing data to HDFS (White 2012, 71)

Figure 3 displays the main sequence of events when writing data to HDFS. Firstly, a client creates a file and sends to distributed file system. The file system calls the NameNode to create a new file without any associated blocks. The NameNode executes several tests to verify this file does not already exist and validate the client's permissions to create the file. A record of the new file is made by the NameNode if these tests pass. The client can write data to the DataNodes via the HDFS output stream. The data is divided into the blocks and stored, replicated in the DataNodes. The client closes the stream when finished writing data. The NameNode gets the block allocations of the file from HDFS output stream and close the process. (White 2012, 70-71.)

It can be seen that HDFS provides an elastic and reliable data storage in the Hadoop mobile Big Data platform. The data sets are stored and managed efficiently in the HDFS clusters. Along with the proliferation of the location data in the case company, HDFS can be easily expanded with commodity hardware to offer an efficient way to store the increasing data.

3.5.3 Hadoop YARN

YARN can be seen as the data operating system of the Hadoop mobile data platform. YARN consists of a global ResourceManager (RM) and per-application ApplicationMaster (AM) which are responsible for resource management and job scheduling/monitoring respectively. The RM acts as a transcendent authority which mediates resources amongst all the applications in the system. The RM comprises of two major components: Scheduler and ApplicationsManager. The Scheduler is strictly used to allocate resources to the running applications according to its capacities and queues. Nonetheless, the Scheduler is not responsible for monitoring or tracking the applications' status. The ApplicationsManager is utilised to assess job-submissions and negotiate the first container for executing the applications. The Node Manager (NM) is the per-node slave framework machine which is responsible for containers and monitoring its resource usage, i.e. CPU, ram and disk. The RM and the NM constitute a data-computation system which is capable of managing

applications in a distributed environment. The per-application AM is a framework specific library which is used to negotiate appropriate resources from the RM and work with the NMs to monitor and track status of the jobs. (The Apache Software Foundation 2014b.) Figure 4 illustrates the architectural view of YARN.

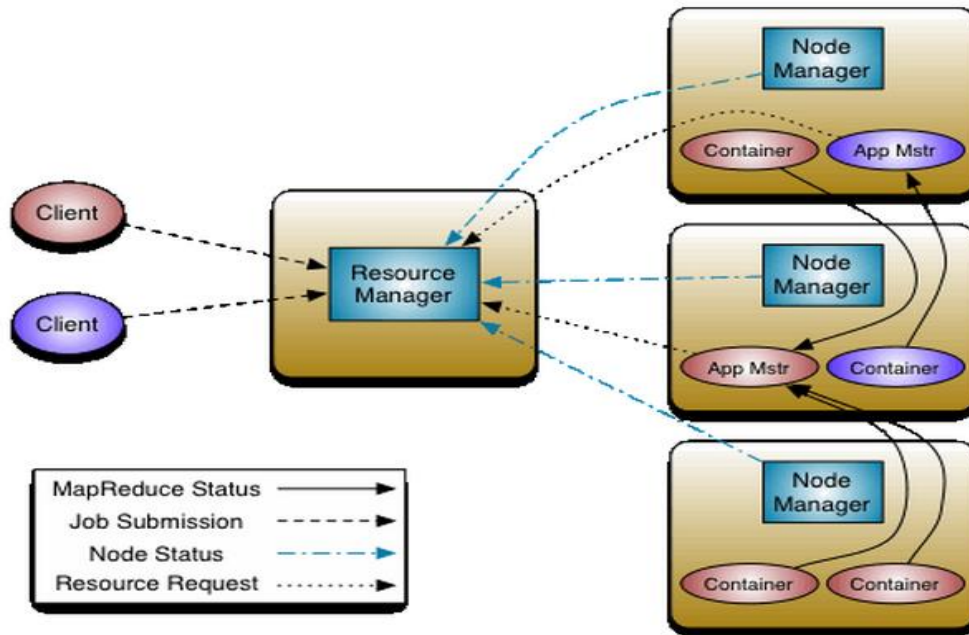


Figure 4. Apache Hadoop NextGenMapReduce (YARN) (The Apache Software Foundation 2014)

Figure 4 represents YARN architecture and its workflow. As shown in figure 4 above, an application is submitted by passing a container-launch context for the AM to the RM. The RM starts and registers the AM. The AM has to communicate routinely with RM to confirm its alive status and ask for the requirements. When the RM allocates containers, the AM constructs the container-launch context to launch the containers on the distributed NMs. In addition, the AM is solely responsible for monitoring the status of the work done in the containers. The AM will unregister from the RM when its work is accomplished and completely exit from the RM. (Valilapalli & Murthy & Douglas & Agarwal & Konar & Evans & Graves & Lowe & Shah & Seth & Saha & Curina & O'Malley & Radia & Reed & Baldeschieler 2013, 8.)

In the Hadoop mobile data platform, YARN provides a greater scalability to manage the resources. The data can be interacted simultaneously in multiple

ways without accessing each cluster for different data sets and applications. Moreover, YARN can perform as a pluggable architecture for operating the high-level Apache applications in the platform such as Apache Hive and Apache Pig. As the research keeps evolving, other applications which require streaming, searching or in-memory analytics such as Apache Solr, Apache Spark can be directly added and run with YARN at the foundation (Hortonworks 2014a).

3.5.4 Apache Hive

Apache Hive provides a framework for data warehousing on top of the Apache Hadoop. Apache Hive is a general-purpose, scalable data processing application which can run queries on the huge volumes of mobile data stored in HDFS. However, it should be taken into account that Apache Hive is not designed for quick, real-time responses. Apache Hive is most suitable for deeper data mining and sophisticated analytics of mobile Big Data. (Hurwitz et al. 2013, 124.)

Apache Hive organises data into tables like relational database management system. Both primitive data types, e.g. numeric, string, timestamp and complex data types, e.g. arrays and maps are supported by Apache Hive. The Apache Hive metadata, which contains the detailed descriptions of the Hive schema such as column types, key data and table statistics, is stored in the meta-stores. The stored data and the Hive metadata are logically combined to create Hive tables. (White 2012, 426-429.) Besides, the meta-stores can sync the data catalogue with other metadata services in the Hadoop platform (Hurwitz et al. 2013, 125).

HiveQL is an SQL-like language used in Apache Hive. HiveQL supports various SQL features such as aggregating, joining and indexing. As an open source project, HiveQL is enriched over time by added more features from the community to fulfil new requirements. HCatalog is an essential component of Apache Hive. It is built on top of the Apache Hive meta-stores and acts as a

table and storage management layer. HCatalog provides ability to read and write data quickly on the grid with different data processing applications included Apache Pig. (Atlassian Confluence 2014.)

3.5.5 Apache Pig

Apache Pig is utilised as a scripting language for processing and analysing large mobile data sets stored in HDFS. It runs as a client-side application. Apache Pig is capable of performing long series of data operations. The data is loaded from HDFS for manipulating. In the next step, the data is run through a set of transformations. Lastly, the data is dumped to the screen, and the results are stored in a file for further visualisation. (Zikopoulos et al. 2012, 65.) Thus, Apache Pig is well suited for extract-transform-load data pipelines and researching on raw mobile Big Data. Furthermore, Apache Pig is designed to be extensible. All processing parts are customised and altered by user-defined functions to meet the particular processing requirements. Nonetheless, like Apache Hive, Apache Pig is optimised for batch processing of data. It is designed to explore the whole data sets or at least the huge amount of data in the large data sets. (Hortonworks 2014b.)

Apache Pig data flow programming language – Pig Latin is developed with very rich syntax to handle any kind of data. With Pig Latin, complex tasks which consist of numerous interrelated data transformations can be simplified and encoded alongside data flow sequences. Pig programs are easy to understand, write and maintain. It can process terabytes of data simply by compiling few lines of Pig Latin from the console. Besides, Pig programs can perform a sample run on a subset of input data for testing before fully operated on the data sets. There are two execution environments to run Pig Latin programs which are local mode and MapReduce mode. In this coherent Hadoop mobile data platform, Apache Pig is performed in MapReduce mode in which all scripts are run on a Hadoop cluster. (White 2012, 367-368.)

Apache Pig and Apache Hive are suitable for analysing and exploring mobile Big Data, particularly location data and sensor logs from HDFS. Both applications provide the ability to manage and extract valuable insights from the large mobile data sets. The information derived from the critical analyses plays an important role for the platform's final stage which is visualising. Additionally, the metadata associated with the information and the accurate output from Apache Pig and Apache Hive will intensify the interpretable information. Insightful visualisation is conducted to represent concisely the information in order to provide a thorough understanding of the customers' in-store behaviours.

3.6 Mobile Big Data visualisation

Mobile Big Data provides tremendous potential value for the case company. It can help the case company to reveal many insights into the customers' behaviours that might seem impossible to collect before, such as how customers' walk in the store or how each individual reacts when seeing a special offer or a new product. However, the extracted data from mobile Big Data alone is sometimes complex and confusing to comprehend. Furthermore, it is practically awkward to "examine a million rows of data and try to make sense of it" (Yuk & Diamond 2014, 8). This is when data visualisations come in. The use of data visualisations is very essential to expose the thorough meaning of data. Visualised information can provide tremendous business value by allowing the case company to make faster and more comprehensive decisions.

Data visualisations aim to make large amounts of mobile Big Data approachable so that the case company is able to draw insights from it. In essence, good visualised information has to have certain traits which are useful, desirable and usable. The useful trait means that the case company can use the visualised information on a regular basis as well as make relevant decisions by viewing all the required data in one place. The visualised information needs to be not only easy to use but also pleasurable and comfortable to use. Lastly, it

has to provide high usability for the end-user, i.e. the case company can accomplish their tasks quickly and easily. (Yuk & Diamond 2014, 8-9.)

In this research work, the visualised information follows these traits and is primarily constructed in the forms of store maps. The example data used to visualise in this research is the analysed data from the activity and location tracking sensors. The raw data is assumed to be collected from the total of 2000 customers visited the store with different demographics and behaviours in the duration of one week. Additionally, the 2000 customers are supposed to move and spend their differently in the store.

The store maps are created based on the example data derived from the Hadoop analytics. The actual case company's store floor plan is simplified and used as the foundation for the store maps. Microsoft Visio is utilised as the primary tool to build the store maps. Five types of the store maps are proposed in this research: (1) Traffic flow map, (2) Route map, (3) Density map, (4) Walking-speed map, and (5) Single-point motion map. The maps are illustrated in the following sub-sections and the detailed information that the case company can get from of each map is discussed.

3.6.1 Traffic flow map

The traffic flow map below shows the example of the aggregate customers' movement directions in the case company's store.



Figure 5. The example map of the customers' traffic flow in the store

By analysing data from GPS, WPS, and gyroscope sensors, the major and minor walking flows and directions are presented in the map in figure 5. The case company can get the insights into the customers' needs as well as their interests by reading this map.

3.6.2 Route map

The map in figure 6 depicts the example of the accumulative customers' walking routes in the store.

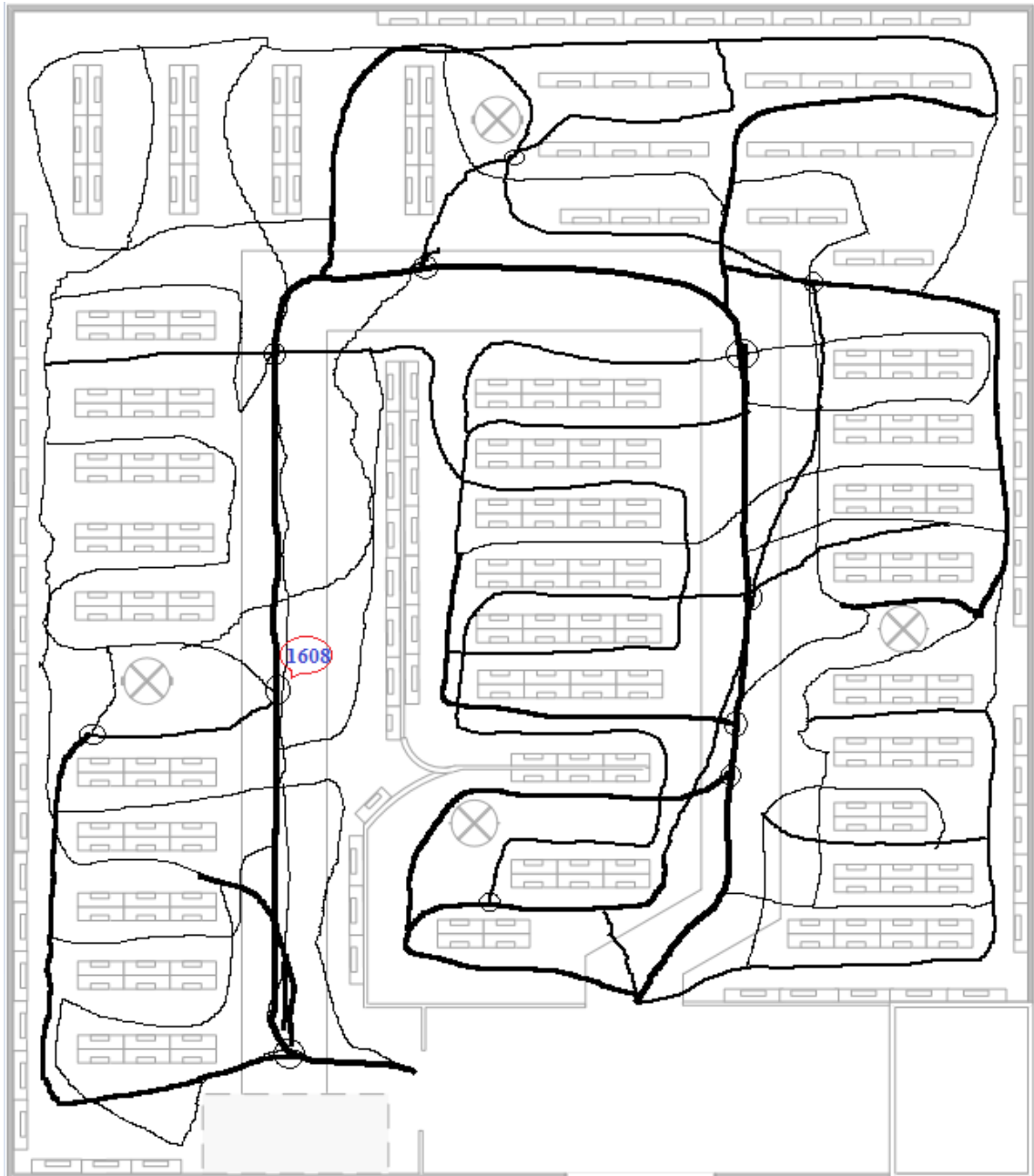


Figure 6. The example map of the customers' routes in the store

The quantity of customers walked in the same routes is represented based on the boldness of these lines in the map. Besides, the map can provide specific numbers of the customers in particular points in these lines. There are 1608

customers shown as an example for the point in the map above. Convergent points are symbolized as small circles in the map. By reading this map, the case company can comprehend the routes that customers prefer to go as well as the convergent place where most of the customers walk past.

3.6.3 Density map

Figure 7 below illustrates the example of the customers' density in the store.

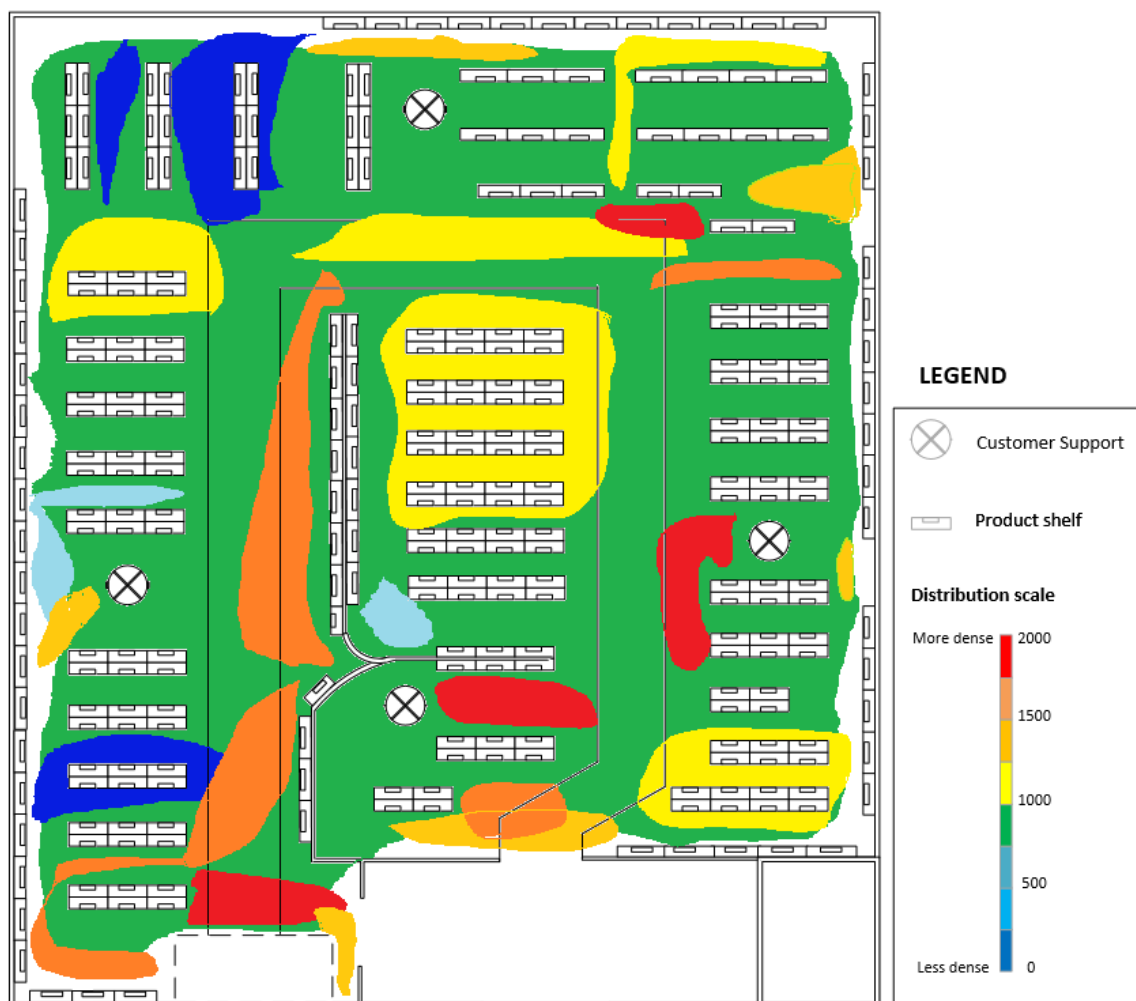


Figure 7. The example map of customers' distribution rate in the store

The map in figure 7 is developed with respect to the spatial and temporal data to highlight the customers' distribution rate in the store. The map shows different distribution rate of different areas in which the customers walked past.

Hence, it can help the case company to recognise the customers' walking habits in different sections and departments in the store.

3.6.4 Single-point motion map

Figure 8 indicates the synthesised motion information of the customers in particular places in the store.

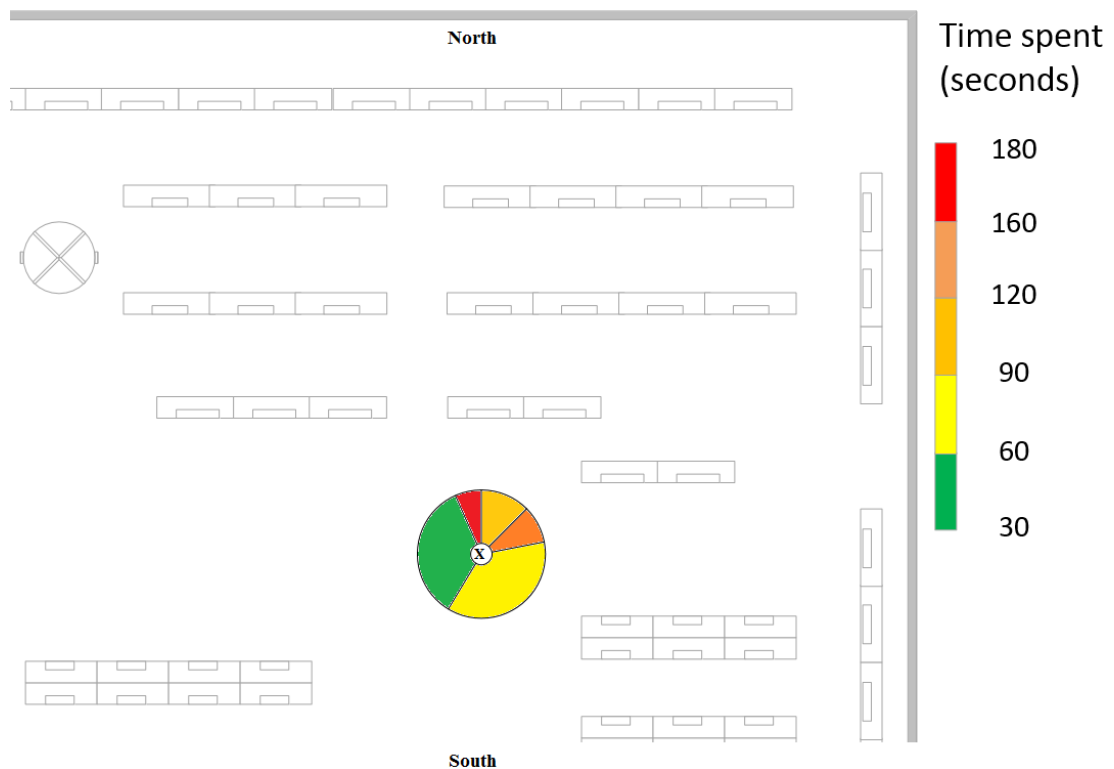


Figure 8. The example map of customers' angular frequency

This map above can provide the observed range of the customers and average time they spent in different angles based on the generated motion data from the accelerometer and gyroscope sensors. Hence, it can help the case company to understand which products are engaged more attention from the customers in compared with other products. For instance, the point X in figure 8 shows that the customers spent more time to look at products in the north (approximately 160 seconds) than they spent to view over the products in the south-east area.

3.6.5 Walking-speed map

Figure 9 below depicts the example of the customers' average walking speed in different departments in the store.

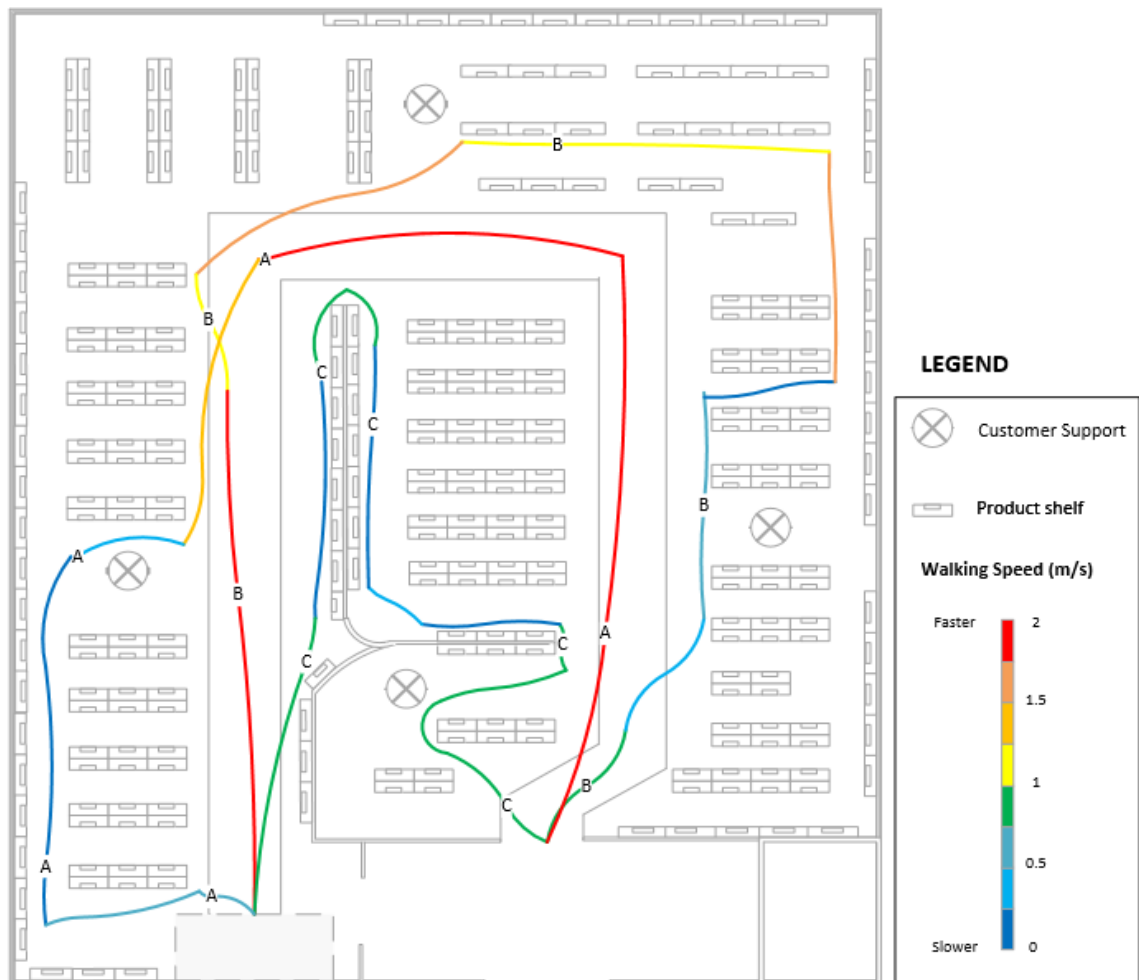


Figure 9. The example map of customers' walking speed in the store

As shown above, the map provides three example routes with different speed in each part. For instance, the route A shows the average customers' speed in different store departments. The customers in route A walked quite fast after entering the store. However, they were slow down when they moved to the left side of the store. The speed significantly decreased after they entered the department in the bottom left of the store. Besides, the case company can hover the pointer over the routes in the map to have the particular customers' speeds as well as acceleration displayed. This map can reveal the customers references and how different groups of the customers move in the store.

Furthermore, it facilitates the case company to understand what the customers are looking for and how their motions change when they see new products or special offers.

3.7 Business value of mobile Big Data

This section identifies the business value of mobile Big Data and presents the ways in which Big Data technologies can help to conceive strategic impact and provide business value for the case company. In order to understand the important of mobile Big Data in making business value, the fundamental correlation of business value and information technology (hereinafter IT) as a whole must be comprehended.

According to Carr (2003), “Today, no one would dispute that information technology has become the backbone of commerce. It underpins the operations of individual companies, ties together far-flung supply chains, and, increasingly, links businesses to the customers they serve. Hardly a dollar or a euro changes hands anymore without the aid of computer systems.” However, the discussion if IT has become a commodity or accessible investments for all companies needs to be arguably questioned, and the influence on business value requires further evaluation. Figure 10, adopted from Devlin (2014), demonstrates the correlation of the business and the IT.

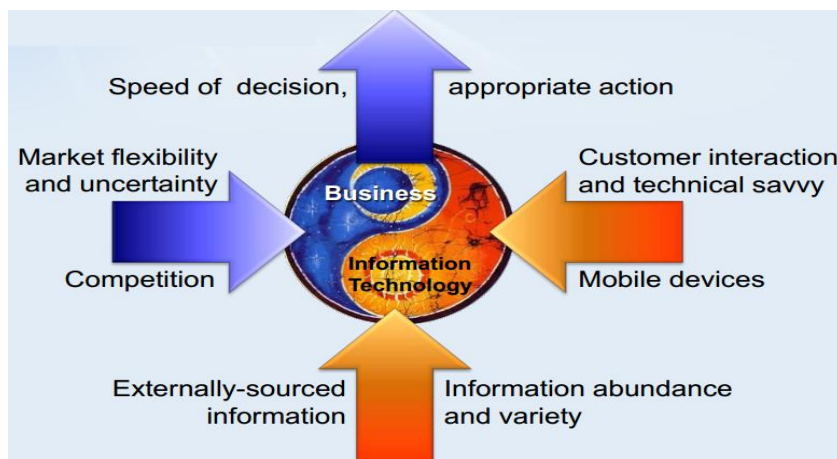


Figure 10. The biz-tech ecosystem reflects the complexity of today’s business (Devlin 2014)

As shown in figure 10, there are three main inputs and an output. On the one hand, there is externally-sourced information. The information is in abundance and variety forms. The novel external information and the mobile technologies such as mobile sensors, smartphones are generating a tremendous amount of data, leading to entirely new way of customer interaction as well as new level of technical savvy amongst customers who are interacting with businesses. On the other hand, the marketplace today is extremely flexible and uncertain at a high level of competition. All of these inputs drive companies to look for a new way of bringing business and IT together. IT and business should be integrated much more closely than ever before. Hence, it can accelerate, enable faster speed for decision-making and appropriate action. The case company should always take into account the integration of IT and business in order to clearly identify the correct strategic plans, i.e. the long-term strategies and short-term implications. To sum up, the biz-tech ecosystem is the environment in which business and IT actually have to work dependent to one another. IT drives business and business drives IT in a tied loop. (Devlin 2014.)

Mobile Big Data can be treated as a trigger for the process innovation and closely connected to the case company business strategies. However, mobile Big Data cannot create value independently but have to “be a part of business value creating process with other information systems and the company factors operating in a synergic manner” (Wade & Hulland 2004 as cited by Kohli & Grover 2008, 26). The factors are linked to the IT-based system which includes business processes, relationship assets, culture and policies as well as IT people such as managers, data analysts and developers (Kohli & Grover 2008, 26).

Mobile Big Data enables a mechanism to gather continuous, expansive and veracious data about the customers’ locations and motions in the store. The information extracted from mobile Big Data plays an essential role in filtering and sense-making capabilities. However, the information does not explain the reason customers are doing this way or another. It basically captures and presents the situation collected from the sensors’ data. The case company can have good business capabilities, yet to make an exceptional difference in

operational excellence and competitive response in the realm of Big Data, the information capabilities to effectively manage and use information are critical. These capabilities refer to information orientation which consists of the ability to administer information via sensing, organising and processing information. (Ryabov 2014; Kohli & Grover 2008, 32.)

For the case company the priority of which is operational excellence, the business value from mobile Big Data is measured in terms of key performance indicators at the process level (Bloch & Hoyos-Gomez 2008). For instance, mobile Big Data is seen as valuable when the information derived from this technology can help the company to overcome the challenge to understand the customers' in-store behaviours. Based on the acquired knowledge, the decision-making process can be quickly administered. The product layout in the store can be optimised. The logistical sector can predict the products' demands accurately in order to offer attractive sales promotions. Consequently, the logistics management can be enhanced, and the sales can be improved along with the increasing rate of customers' satisfaction. In addition, with mobile Big Data and its automation, the human resources can be reduced while the company is still able to get the valuable information of customers.

Businesses and customers can be seen as the final arbitrators of the business value creation. The intangible value derived from mobile Big Data is increasingly essential. Today, the business value contains not only the direct economic benefits such as return-on-investment or market share, but also the indirect, intangible value such as agility, flexibility and understandings to adapt for vigorous customer trends. (Kohli & Grover 2008, 33.)

Mobile Big Data can enhance the economic and the strategic dimension of the case company. Nonetheless, it is important to notice that the business value of mobile Big Data cannot be realised instantaneously, but it requires a perceptible period of time. Mobile Big Data does not explain why customers behave in different ways in the store. It basically shows how customers' behaviours are in the store and presents the collected data as the aggregate visualised information. Therefore, the success of getting business value from mobile Big

Data partially relies on the information capabilities of the company, not Big Data technologies alone.

3.8 Privacy and ethics

While mobile Big Data offers highly valuable information which would be impossible to obtain before, it also comes with new challenges concerning privacy and ethics of individuals as well as companies. Mobile Big Data itself has no country, obeys no law and can travel across any borders. It can be stored in cloud networks and physically located in any places in the world. The information extracted from Big Data can be retrieved and copied in seconds from anywhere. There are diverse perspectives about mobile Big Data, particularly personal data, i.e. data generated in the course of person's activities. One might feel it disturbing, whereas others just see it as the price for living in the digital age. (Craig & Ludloff 2011, 20.)

Oxford University Press (2014) defines privacy as "the state of being alone, not watched or disturbed by other people and being free from the attention of the public". Nonetheless, the concept of privacy has become more complex in the digital world. The right to remain relatively anonymous in society at any given time and places has certainly altered with the proliferation of the Internet and social networks. Today, most of the human activities and behaviours are digitally captured and analysed for various purposes. (Craig & Ludloff 2011, 15.) For instance, customers' shopping behaviours can be thoroughly analysed based on their movements in the stores in conjunction with their Facebook profiles, online-shopping habits and even recent statuses posted in Twitter about recently purchased products. The digital age has allegedly blurred the boundaries between privacy and publicity. It follows that privacy cannot be a mere discussion of what is right or wrong but the perception in assessment of any situations that one shows.

The private and ethical impact is highly context dependent (Davis & Patterson 2012, 14). People with different cultures, society and politics can have different

attitudes towards privacy and ethics. For example, some customers advocate the use of mobile Big Data to improve the shopping experience while few perceive that this usage would infringe upon their private life. However, there is nothing morally wrong about collecting and using personal data, e.g. customers' location data as long as the customers are aware of it.

While Big Data as a whole and personal location data in particular are ethically neutral, the use of it is not. Ethical enquiry into finding and maintaining a balance between the benefits of using mobile Big Data technologies and the detriments of individuals' privacy is imperative. The case company, in order to succeed in the highly competitive marketplace, has to explicitly assess the ethical impacts of the data collected from customers. Thus, the risks of diminishing the quality of the company's relationships with customers can be minimised. (Davis & Patterson 2012, 13-14.)

As the pervasiveness of mobile Big Data increases, complex privacy and ethical dilemmas can arise. The case company needs to be proactive in dealing with anticipated privacy and ethical issues. New policies and procedures can be established to protect the customers' privacy such as notifying before collecting data or anonymised the data prior to analytics. By doing so, the company can exploit the valuable information generated from mobile Big Data while keeping in harmony with customers.

4 ANALYSIS OF EMPIRICAL DATA

The empirical part is based on the qualitative research method utilising the approach of the in-depth interviews and the questionnaire survey. The analyses of the data collected are presented in this chapter. This chapter is divided into two sections; the first section aims to discover the case company related information while the second section focuses on analysing customers' data and their opinions revolved around mobile Big Data.

4.1 Analysis of interview data

In-depth qualitative interviews were utilised in this research as it allows the researcher to gain specific information and genuinely explore the interviewee's point of view and perspectives about the research topic. The interview procedure was semi-structured in order to keep the open-ended structure for the interviews and give the interviewee freedom to express opinions in detail. Two in-depth interviews with the store manager were conducted and recorded in this research. The transcripts of the interviews are displayed in Appendix 1 and Appendix 2 respectively. The store manager was chosen as the sole interviewee since he has an extensive and broad knowledge of IT infrastructure as well as business processes in the case company. In addition, the store manager is the main person who is responsible for making business decisions and development plans in the store. In order to provide highly reliable data, the interview transcripts are interpreted and analysed in conjunction with the real system observation, the case company source and data collected from the Internet. Furthermore, several informal discussions and e-mail enquiries were carried out with the store manager and other members of staff in the case company to obtain the necessary information of the analysis. The following sub-sections discuss the main themes of the interviews.

4.1.1 The case company IT infrastructure

The case company Elgiganten AB – Haparanda has its own enterprise resource planning (hereinafter ERP) system. The system is deployed in all the Elgiganten

AB stores in Nordic countries. MS-DOS provides the environment for the system to work in. In essence, the company ERP system is business management software. It is simple, but powerful and effective software that fulfils all the business requirements for the case company. (Putila 2014c.)

The software includes two main modules which are financial module and logistics module. The logistics module is the principal part of the case company's ERP system; it provides management information of the warehouse, and forecasts sales to optimise the store inventory. The sales and distribution features in the logistics module are responsible for managing sales enquiries, sales ordering as well as billing processes. Besides, the logistics module supports the online shopping feature. This feature can receive online orders and print out the orders. The financial module administers the finance and investment projects in the store. These modules are capable of working independently, yet always connecting and synchronising information with each other and ERP system as a whole.

4.1.2 The importance of understanding customers' in-store behaviours

The case company highly values the understanding of customers' behaviour in the store. In other words, it is one of the most important factors that the company wants to possess now. Happy customers are the key factor for the sustainable development of the company. The best way to satisfy customers is to understand what they want and what they need. The case company is intensively working on this issue by training staff communication skills such as warmly greeting every customer when they enter the store to create a friendly and pleasant shopping atmosphere. In addition, the members of staff have to be good listeners. They have to carefully hear and pay attention to what customers' needs in order to suggest some sensible recommendations according to their requirements. By comprehending the customers' shopping habits and behaviours, the staff can be progressively proactive when serving the customers.

The store manager finds it important to understand the customers' shopping behaviours. The case company has made heavy investments in order to achieve this understanding. However, by the time this research is written, the case company has as far as customers' satisfactory assessment system called "Happy-or-Not", which is used to receive feedback concerning customers' satisfaction after shopping in the store. In addition, promotional e-mail subscriptions and cookies are being used to facilitate the company's ability to understand the customers' habits in the online shopping environment.

The case company specifically attends to understanding the customers' in-store behaviours, and mobile Big Data is capable of providing the information. Therefore, the company is willing to invest in the mobile Big Data technologies if it is effective and can help the company to enhance the customers' understanding. (Putila 2014b.)

4.1.3 Company policies to protect customer rights

The case company currently operates several policies to protect customers when shopping in the store. For instance, the company provides voluntary filling form when customers buy products. It means that customers can fill in their personal information for later customer service and in case customers do not want to share their personal information, filling in the form is not required. Moreover, free opt-in and opt-out policies are applied when customers subscribe to online advertising emails as well as printed advertising newspapers. (Putila 2014b.)

According to the store manager, collecting and using personal location data, particularly customers' location data would be legal and accepted as long as customers do give consent to it. Besides, the store manager demonstrates that all customers' information is securely stored and always used with the permission of customers and only for the company purposes. (Putila 2014b.)

4.2 Analysis of questionnaire data

Questionnaire technique is used in the form of online questionnaire survey to collect customers' opinion on mobile Big Data. The reason for doing the online survey is because it is an efficiency way to reach the high amount of the different respondents in a short time period. Furthermore, the respondents can have flexible time to complete the questionnaire, i.e. the responders can do anytime without constraints of time and space. The survey targeted respondents are the case company's customers who live in the Haparanda – Tornio area and recently visited the store. Since most respondents speak English as a second language, the survey questions use easy to understand vocabulary and simple, short sentence structures. The online survey, created by SurveyMonkey tool, includes a total of 10 questions. There are 6 multiple choice questions, 2 yes/no questions, a scale question and an open-ended question, all are displayed in a single webpage. Hence, the respondents can see all the questions at one time and do not need to click many times to move to other pages in order to complete the survey.

The questionnaire was discussed and approved by the store manager. The questionnaire version which was used to collect data from the respondents is shown in Appendix 3 in this thesis. The link connected to the online survey web page was posted on the case company social network page and Haparanda – Tornio region social network page. The survey was available for the duration of one week. By the end of the survey, there are 86 responds received in total. The SurveyMonkey tool are utilised to analyse and visualise the statistics result.

The survey was divided into three categories. The first category is to determine the customers' demographics. The second category is to collect the customers' shopping habits and behaviours in the case company store. The third category is to gather the customers' attitude towards location data. Each category is discussed in each section below.

4.2.1 Customers' demographics

Understanding the customers' demographics is crucial for the case company to realise the market segmentation. Based on the cultures, genders and age groups, the case company can identify the market groups and implement proper selling and marketing strategies to target and fulfil the different range of customers' requirements. The respondents' domicile and gender received from the questionnaire survey is illustrated in the pie charts below.

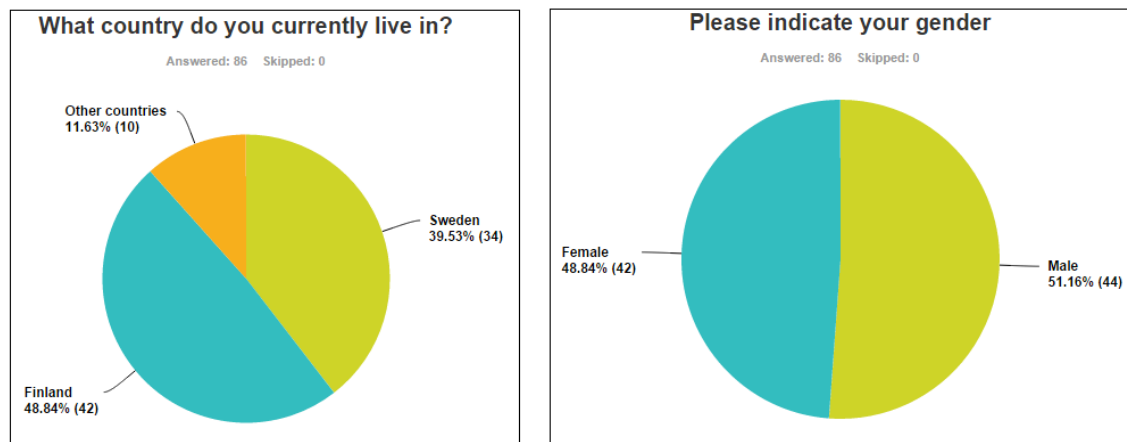


Figure 11. Respondents' domicile and gender

It can be seen in the pie charts that the number of male respondents was almost the same as the female respondents with 51.1 % over 48.9 %. However, there were some differences in the results when asking the respondents about their current place of resident. The respondents from Finland accounted for nearly half of the total responds while respondents from Sweden came in second place with around 40 %. The rest respondents (11.6 %) informed that they currently live in other countries. Figure 12 represents another aspect of the customers' demographics which is the age.

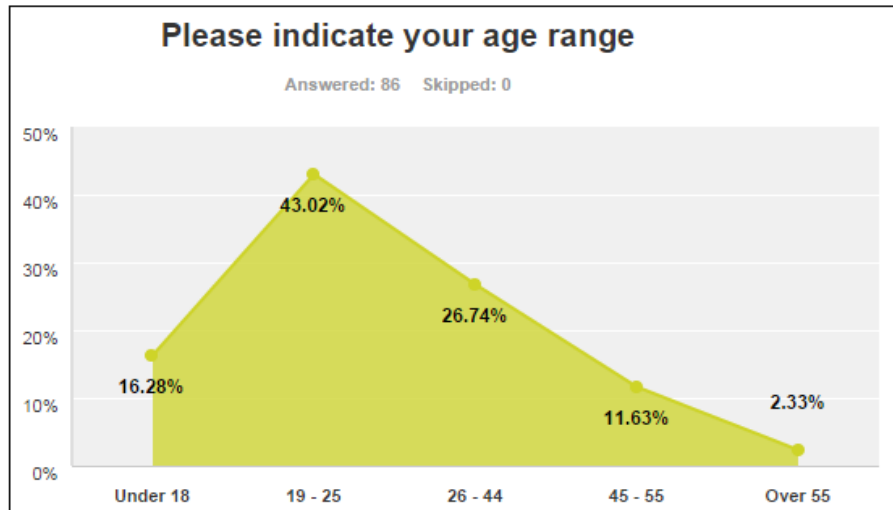


Figure 12. Percentage of respondents' age range

As shown in the area graph, the majority of the respondents was in the ages of 19 – 25 which takes up to 43 %, followed by the age group of 26 – 44 with 26.7 %. The reason for the high percent of the respondents in the age from 19 to 44 could be that it is the young working age group which has greater interest in IT devices and independent incomes in comparison with other age ranges. The age group of under 18 accounted for 16.3 % of the total respondents. The number of the respondents started to decrease after the age of 44 with 11.6 % of respondents in the age group of 45 – 55 and only 2.3 % of respondents over 55 years old. In the context of these demographics results, the case company should pay additional attention for the customers with age range from 19 to 44 when developing marketing strategies and sales promotions.

4.2.2 Customers' shopping habits and behaviours

It is necessary to understand customers' shopping habits and behaviours before applying new technologies in the case company store. There are 5 questions used in the questionnaire to enquire the respondents about their shopping habits and behaviours. The results are illustrated in the figure 13, figure 14 and figure 15 below.

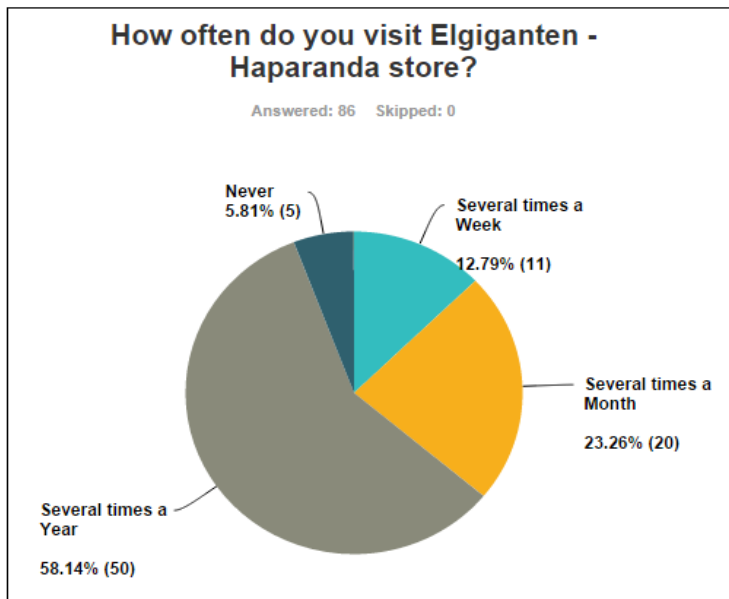


Figure 13. Store visiting frequency

The pie chart in figure 13 represents the respondents' visiting frequency in Elgiganten - Haparanda store. A significant number of the respondents (58 %) indicated that they visit the store several times a year. Approximately a quarter of the respondents visit the store few times within a month while just over one in ten visit weekly. It is interesting that around 6 % of the respondents have never visited the store.

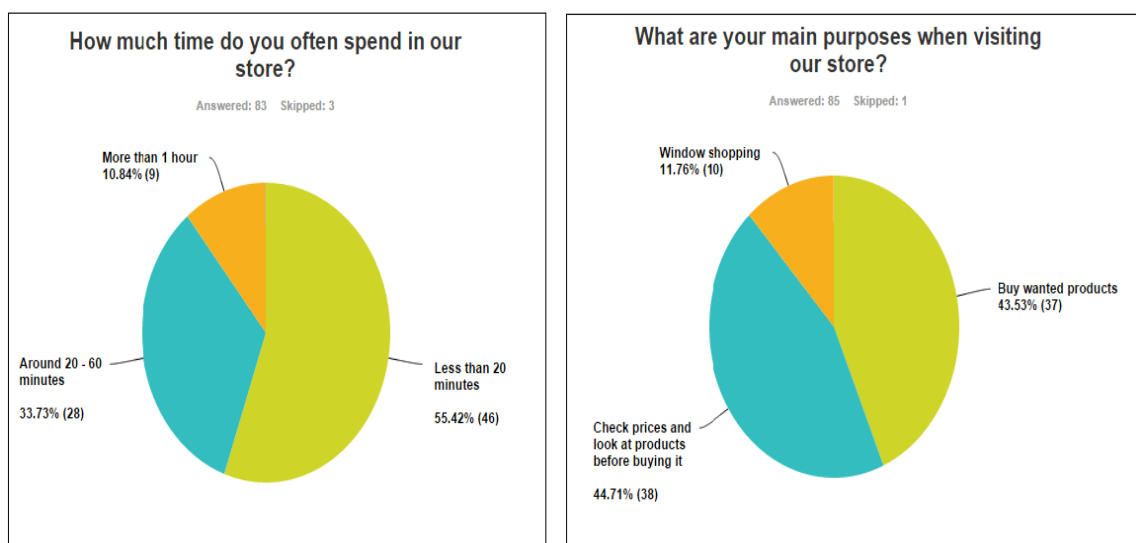


Figure 14. Respondents' time spent and purposes when visiting the store

The pie charts in figure 14 above depict the response results given by the respondents when enquired about the time they often spend in the store and the purposes of visiting the store. It is shown more than half of the respondents (55 %) generally stop by the store for less than 20 minutes. Conversely, just under 11 % of the respondents indicate that they often spend more than an hour when visiting the store, whereas approximately 34 % of the respondents usually stay in the store for around 20 to 60 minutes. Additionally, buying wanted products accounted for 43.5 % of the respondents' main purposes when visiting the store, and this figure was very similar for the "checking prices and looking at products before buying it" purpose, at 44.7 %. Another reason for visiting the store is window shopping given by less than 12 % of the total respondents.

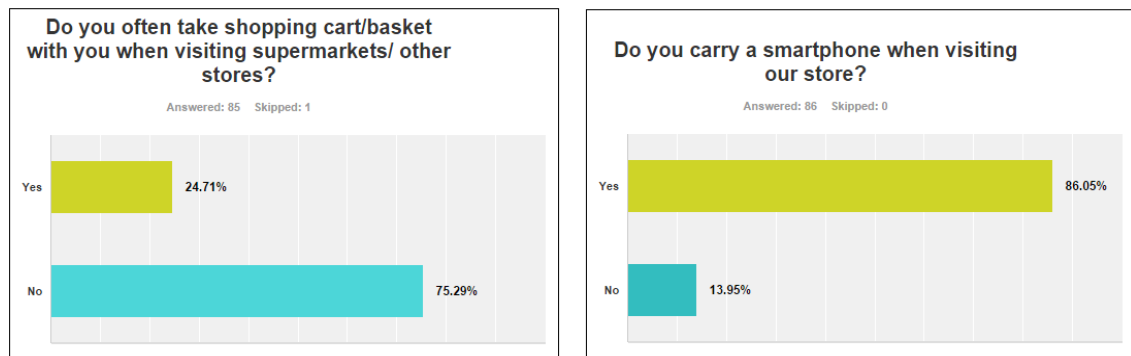


Figure 15. Respondents' habits when visiting the store

In order to assess the feasibility to embed the tracking devices in the case company store, the respondents were asked about their shopping habits, particularly whether they take baskets or carry smartphones when doing shopping. 25 % of the respondents pointed out that they often carry shopping baskets; by contrast, 75 % do not use it. In addition, the vast proportion of the respondents (86 %) mentioned that they do carry a smartphone, when only 14 % said no.

To sum up, most of the respondents usually visit the store few times a year, and the optimum time spending in the store is less than 20 minutes. In addition, the main purposes when visiting the store indicated by the respondents are to buy intended products or check products' prices and have a physical look at it.

Further, a quarter of the respondents often take shopping baskets, and the majority of the respondents carry their smartphones when visiting the store.

4.2.3 Customers' attitude towards location data

The ninth question in the questionnaire asked the respondents to rate their perception of collecting location data anonymously in the store in the five-level scale. There were two cases in which the respondents had to judge. The first case was that the respondents are notified in advance and the second case, the respondents are not. The results were illustrated in figure 16.

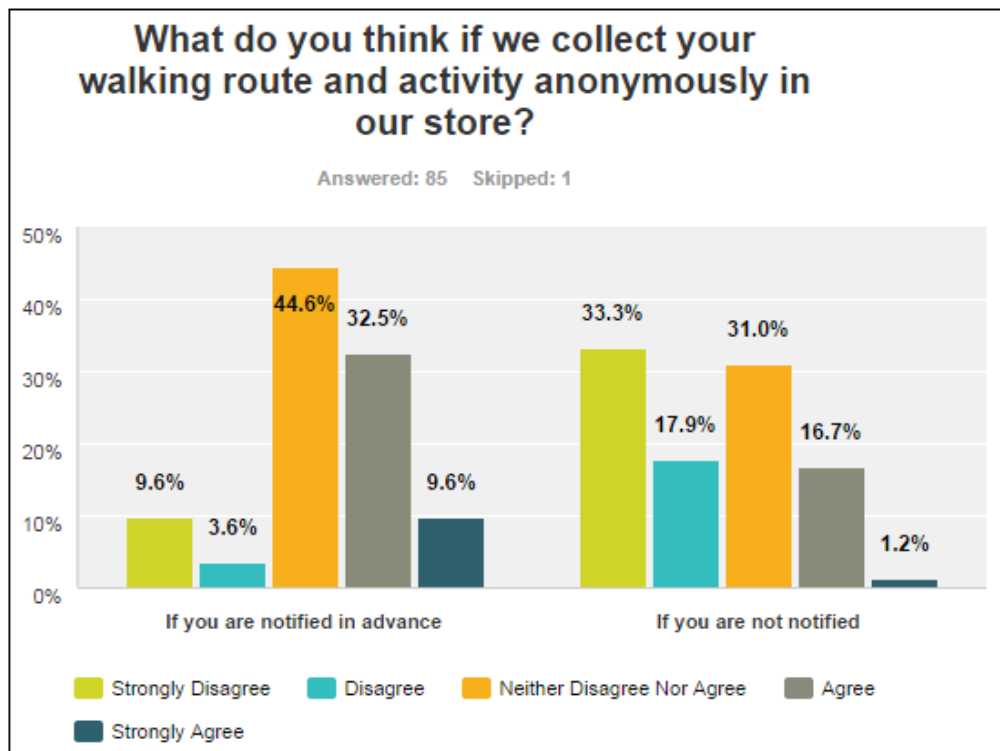


Figure 16. Respondents' perception of collecting location data

As shown in the vertical bar diagram above, the response results were noticeably varied in the two different cases. In the case when the respondents are notified in advance, there were more positive sentiments than negative ones. Approximately 42 % of the respondents indicated that they strongly agreed or agreed, followed by 44.6 % of the respondents who remained neutral about sharing their location data. Only around 13% showed disapproval of the

collection. In contrast, when the respondents are not notified in advance, the disapproval significantly increased with 33.3 % of the respondents strong disagreed and 17.9 % disagreed. A third of the respondents remained neutral while less than a half of the respondents agreed in this case in comparison with 32.5 % of the respondents agreed when they are notified. Therefore, the case company has to take notice of the importance in notifying the customers before collecting their location data in the store. Additionally, it helps the case company's customers to be aware of their right to privacy when shopping in the store.

4.3 Summary on the results of data collection

The results of the in-depth interviews and the questionnaire survey help to reveal many important aspects in the case company store as well as its customers. It can be seen that the case company does not have any infrastructure related to Big Data. Thus, the mobile Big Data platform can be installed from the ground up in the store. Moreover, the case company and the store manager in particular have a high awareness about the importance of utilising mobile Big Data to understand the customers' habits and behaviours. This awareness can provide a positive impact on deploying the mobile Big Data in the store.

Besides, the survey conducted shows that the case company store is serving a diversity of the customers living in the Tornio – Haparanda area. Most of the customers are willing to share their location data in the store for the purpose of improving the case company's customer service as long as they are aware of it. The customers carrying shopping baskets and smartphones can create favourable conditions to apply sensor devices and facilitate data collection in the store.

Furthermore, from the last open-ended question in the survey, the majority of the respondents' comments exposed optimistic views about the applicability and customer value in which location data can contribute for the customers'

shopping experience in the case company's store. Besides, one respondent gave a comment about the store layout which is worth to contemplate. This respondent's comment is as follows: "The walking route through the store feels less freedom as the walking route concept of IKEA possessed. You cannot go straight to cashier, as an example you need to buy some small products from cashier only, instead you have to walk through the whole store, since there is no short cut in the middle due to the shelves. Of course, on the marketing point of view it is a simple solution to put up shelves this certain way to engage customers to view other products as well. Anyway customers could be provided with a small shortcut in the middle. In this matter, the amount of sold products would not be that much less, but the overall customer experience could rise."

5 BUSINESS SCENARIO USING MOBILE BIG DATA TO SUPPORT DECISION-MAKING IN THE CASE COMPANY

In the previous chapters, studies into the background of mobile Big Data and empirical data have given insights and suggestions of sensor technologies and mobile Big Data applications for the case company. This chapter presents the business scenario which uses mobile Big Data to support decision-making in the case company. The business scenario helps to address the technical requirements, business value and prospective results when deploying mobile Big Data in the case company. Furthermore, based on the business scenario, the findings are discussed and the recommendations are proposed for the case company.

5.1 Business scenario overview

Mobile Big Data is a sophisticated system, and it needs a detail plan and preparation before applying in the real situation. Thus, the business scenario is developed and scrutinised. The scenario revolves around the human and resources needs for deploying mobile Big Data technologies in the case company store. Moreover, some technical difficulties and business problems can be revealed by scrutinising the scenario.

The business scenario utilises the case company's retail store as the experimental environment. The objectives of the business scenario are to present the comprehensive requirements and estimate the cost to deploy mobile Big Data technologies in the store. In order to achieve the objective, the activity and location sensing systems with different features are proposed. Furthermore, the total solution cost of the business scenario is studied, and the deploying process is described in-depth. It should be noticed that all the costs in this scenario use the euro as the sole currency. Any costs which calculate in United States dollar are exchanged to euro with the exchange rate of US \$1 equals to €0.78 (XE.com 2014).

The business value of the mobile Big Data is always important. Nonetheless, the investments in mobile Big Data technologies are only made when the expected business value is noticeably greater than the expected cost. Therefore, cost analysis and business value derived from the scenario are assessed.

5.2 Activity and location sensing systems

This section discusses the sensing systems which can use to track customers' activities and locations in the store. The systems are proposed based on the sensor technologies studied in the previous chapter. The cost and strengths as well as weaknesses when applying in the store are discussed.

The first system which has high feasibility to deploy in the case company store is the Aaronia GPS logger. The logger integrates GPS sensor, three-axis accelerometer, three-axis gyroscope sensor and compass sensor into an independent, stand-alone data logger device. The logger can provide high update frequency of "35 complete logs with all sensor data per second". The maximum rate the Aaronia GPS logger produced is around 50 megabytes per hour. The sensor data logs can be stored in micro-SD card or streaming via USB interface. There are many well-known organisations which are currently using the Aaronia GPS logger such as Deutsche Bahn, NATO and IBM Deutschland. (Aaronia AG 2011.) The actual image of the logger is shown in the figure 17 below.



Figure 17. The Aaronia GPS – Logger (Aaronia AG 2011)

As illustrated above, the logger is very mobile with three dimensions are 106x45x22mm (Aaronia AG 2011). It makes the logger be easily embedded in the shopping carts and baskets in the store. Furthermore, the logger is capable of capturing detail customers' motions and movements in the store with its high update frequency. The sensor data logs from the logger can be transferred to HDFS at the end of the day or every other day relies on the case company needs. The price for single unit of the logger is €250 (Aaronia AG 2011). For the case company which usually serves around 40 customers at a time, 40 units will be efficient. Hence, the initial cost for the will be $40 * €250 = €10,000$. Since the loggers do not require other major costs to operate and come with a 10-year warranty, the investment cost will be decreased in the long-term period.

The second system that the case company should consider is WPS. Euclid is one of the major WPS providers. For the case company store which has an area of 900 square meters, it will require 12 WPS sensors to cover the whole area and provide optimal customers' position tracking results. The sensors support plug-and-play and can be installed within two minutes using the current internet and power line in the store. There is no initial fee to obtain the sensors. However, Euclid charges monthly per-store fee for support and services. The proposed fee is €450 per month for the system applied in the case company store. (Euclid 2014b.) The data collected from the WPS sensors can be transferred to HDFS for further analysis.

Pressure measurement system is another solution to gather the location data. Smart Floor is presented as a new and cost-effective pressure measurement system to provide the customers' position data. According to the Gator Tech Smart House, the total cost for the Smart Floor system as deployed in the 32 square meters are approximately €3,200. The cost includes the installation and material expenses of the raised floor. (Kaddoura & King & Helal 2005.) Hence, it is assumed that the case company store applies the Smart Floor. In order to reduce the cost, only the main walking routes which equal to around 600 square meters are covered with the Smart Floor. The initial investment cost of this system for the case company will be $(€3,200 / 32) * 600 = €60,000$.

The suitable systems can be selected based on the case company needs. Nevertheless, the case company should carefully scrutinise the initial costs, the return on investment and the long-term costs of each system. It is important to clearly understand what each system can do in comparison with its total costs in order to gain optimal benefits. Besides, the systems can be combined and work together as a hybrid system. One system can compensate the insufficient features and take advantage of the strengths of others.

5.3 Total solution cost of business scenario in the case company's retail store

The business scenario presents the requirements of implementing the mobile Big Data technologies in the case company. The total solution cost for the business scenario is estimated and proposed for the case company. The estimate consists of all the system and hardware costs as well as the software related costs for creating the business solution. The solution includes the cost of developing and testing the management, analytic and visualised application that utilise the sensor data for the intended purposes. Figure 18 below depicts the main cost components of the total solution cost in the scenario.

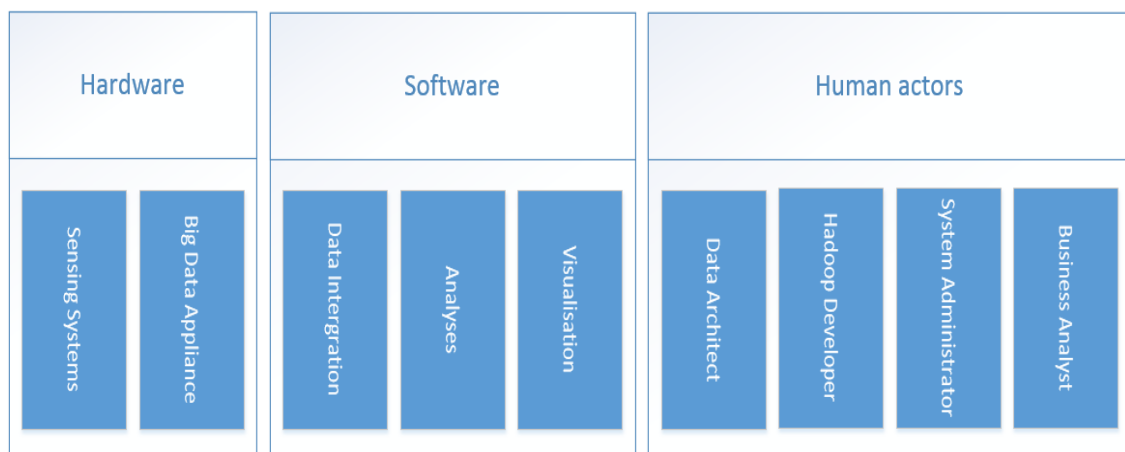


Figure 18. Components of the total solution cost

As shown above, there are three major components that contribute to the total solution cost. These components are hardware, software and human actors. Each component and its factors are studied in-depth in this section. Since mobile Big Data technologies are the long-term investment, the business

scenario calculates the solution cost in the period of three years. Additionally, it is assumed in the scenario that the hardware requirements are supplied as demanded and there is no hardware price fluctuation. Besides, the software applications are supposed to work as required and human actors do their job as usual.

In this scenario, all the components revolve around Apache Hadoop platform. Apache Hadoop is an open-source software framework for managing and processing large-scale of data sets from the activity and location sensing systems. Apache Hadoop has emerged in the retail industry as a low-cost, alternative system to the conventional data warehouse platform. It is shown that the acquisition cost of a Hadoop cluster is under €800 per terabyte of data, much lower than the average price for a data warehouse platform. (Winter & Gilbert & Davis 2013, 3-4.)

There are two key factors in the hardware component of mobile Big Data. These are the sensing systems and the Big Data appliance. The sensing systems are discussed in the previous section and the systems proposed prices are in the range of €10,000 to €60,000 depended on the selection. The Big Data appliance includes a Hadoop cluster and switches. The total cost of the Hadoop cluster comprises of production ready hardware and software. For the case company, it is assumed that the Hadoop cluster runs as a production system with hot-swappable disks and redundant computer components.

Oracle proposes a cost-effective plan to adopt the Big Data appliance named Oracle Big Data Appliance. According to Oracle, it is an optimised system built for production clusters with “pre-installed and pre-configured with Cloudera CDH and all options included”. The price of the whole system is approximately €414,000. The case company gets the following system: “Big Data Appliance Hardware (comes with Automatic Service Request upon component failures), Cloudera CDH and Cloudera Manager, All Cloudera options as well as Accumulo and Spark, Oracle Linux and the Oracle JDK, Oracle Distribution of R, Oracle NoSQL Database, Oracle Big Data Appliance Enterprise Manager Plug-In.” The Oracle Big Data Appliance hardware consists of a full rack of the

Big Data cluster with 18 Sun X4-2L servers, 288 cores, 864 terabytes disk and 3 Infinitive Band switches which allow the case company to expand the cluster in the future without requiring extra switches. (Dijicks 2014.)

Moreover, Oracle provides premier support for the system with €50,000 per year. The one-time install and configuration on site cost around €11,000 more. The configuration includes optimising and tuning the system to the Apache Hadoop, the operating system and network settings. Besides, secure authorisation methods such as Kerberos and Apache Sentry are configured. (Dijicks 2014.) The total cost for the Big Data appliance in the period of three years is presented in table 1 below.

Table 1. The Big Data appliance total cost (Dijicks 2014)

| | Year 1 | Year 2 | Year 3 | Total in 3 years |
|---------------------------------|----------|---------|---------|------------------|
| Big Data Appliance Cost | €414,000 | | | |
| Annual Support Cost | €50,000 | €50,000 | €50,000 | |
| On-site Install (approximately) | €11,000 | | | |
| Total | €465,000 | €50,000 | €50,000 | €565,000 |

As shown in table 1, the total cost of the Big Data appliance using Oracle system is €565,000. The initial investment for the system is high; however, the total average cost will gradually decreased along with the time using the system. For 3-year period, the cost per terabyte of data using this system is $€565,000 / 864 = €654$.

There are no software license fees or subscriptions for using Apache Hadoop. Hence, after acquired the Big Data cluster from Oracle, the case company can start to install Apache Flume, Apache Sqoop, HDFS, YARN with Apache Hive and Apache Pig without any additional payment. However, specialised skills will be required to configure and implement the cluster for the required applications. Besides, it is important to note that the Big Data enterprise manager software, Linux subscriptions and Big Data appliance plug-in are already included in the total price of the Oracle Big Data Appliance package.

The second component which contributes to the total solution cost in this business scenario is software operation. It includes the cost to develop and execute 3 main factors which are data integration, mobile Big Data analyses and visualised application. The cost of data integration is the cost of developing a process to purge the generated sensor data, rearrange it as needed before importing it into the Hadoop ecosystem. The mobile Big Data analyses generate the cost of developing procedural programs in Apache Hive and Apache Pig to perform complex data analyses. Visualised application accounts for the cost of developing flexible application which utilise the analysed data and visualise it as the store maps with analytical information. The real value of mobile Big Data is delivered to the case company through this software operation. (Winter et al. 2013, 19-20.)

Since Apache Hadoop is a free, open-source software library, the major software expenditure in this scenario is the labour cost of developing, managing and operating the applications and program in the Hadoop ecosystem. Thus, the human actors are addressed as the third component in the total solution cost. There are three key human resources in the human actors required for the successful ongoing use of mobile Big Data technologies, particularly Apache Hadoop. Table 2 presents the human actors, their roles and annual income in the Hadoop ecosystem. The average income data used in table 2 is referenced from PayScale Inc. (2014).

Table 2. Human Actors, Roles and Annual Income in the Hadoop ecosystem

| Human Actor | Roles | Annual Cost |
|----------------------|---|-------------|
| Data Architect | Outlines the plan and architecture of the Hadoop deployment; works with Apache Flume and Apache Sqoop to integrate data into HDFS | €83,000 |
| System Administrator | Manages and monitors the Hadoop cluster; integrate Hadoop with other applications | €47,500 |
| Hadoop Developer | Builds mobile Big Data applications in Hadoop such as visualised application and Hive queries | €61,500 |

As shown in table 2, three new positions are required to work and be responsible for different tasks in the Big Data ecosystem in this business

scenario. The Hadoop developer position can be a one-year contract in order to develop the mobile Big Data applications while the data architect and system administrator will be the long-term contract to operate the Big Data ecosystem. The total human actor cost in the first year from the three positions is €83,000 + €47,500 + €61,500 = €192,000. From the second year, there will be two positions needed which are data architect and system administrator; and the total human actor cost will be €83,000 + €47,500 = €130,500 per year. In addition, the store manager can hold concurrently as a business analyst who is responsible for identifying the value within visualised data. Further, the store manager can work with business intelligence and the case company's data warehouse.

To sum up, the total solution cost is the combination of three major components: hardware, software and human actors. The total solution cost is calculated over the period of three years. Table 3 below displays the total solution cost of this business scenario.

Table 3. The total solution cost

| Components | | Year 1 | Year 2 | Year 3 | Total in 3 years |
|--------------|--------------------|---------------------|----------|----------|-------------------------|
| Hardware | Sensing systems | €10,000 - €60,000 | | | |
| | Big Data Appliance | €465,000 | €50,000 | €50,000 | |
| Software | | 0 | | | |
| Human actors | | €192,000 | €130,500 | €130,500 | |
| Total | | €667,000 - €717,000 | €180,500 | €180,500 | €1,027,000 - €1,077,000 |

As shown above, the total solution cost for three years is from €1,027,000 to €1,077,000. While the initial investment cost to equip hardware is high in the first year, the following years show the solution cost to operate the system is substantially decreased. Hence, the longer the Big Data system is used, the lower the yearly average cost it will be. Nonetheless, the total solution cost in this business scenario does not take into account unexpected costs that might occur during the operation such as the power outage, data corruption or rising inflation.

5.4 Findings and recommendations

The business scenario highlights several important guidelines for deploying mobile Big Data technologies. The business scenario aims to help the case company to develop and manage cost-effective mobile Big Data technologies while still be able to minimise or even eliminate the risk of making any costly mistakes in the process. Besides, the business scenario shows that Big Data can help to analyse the complex mobile sensor data as an affordable cost, i.e. €654 per terabyte. It is much more cost-effective compared with traditional data warehouse.

The key sources of value of mobile Big Data come through data visualisation and analytics. Effective visualisation and analytics play a critical role in optimising the business value. Nevertheless, the business value of mobile Big Data cannot be delivered instantly. The analytics need to be tailored to meet specific business needs, and the visualised results require the proper information capabilities in order to extract value from mobile Big Data.

Moreover, selecting appropriate mobile Big Data platform is important. The right choice can reduce business risk and time to value. The Hadoop mobile data platform and the Apache Hadoop framework discussed in this research can work as the fundamental base for stable operation and development of the mobile Big Data applications. As the case company substantially grows in size, the platform has to be easily expanded to fulfil the higher performance.

It can be seen that the initial investment costs of mobile Big Data in the business scenario is the major expense. Nevertheless, the case company should not underestimate the ongoing costs of system administration, maintaining and developing complex analyses. Each of these costs can pile up over time. Hence, the case company has to carefully plan the long-term investments in operating mobile Big Data as well as assess the return on investment in developing visualisation and analytics in response to fast changes in business requirements.

The case company, as a part of large chain stores, can take full advantage of the ability to handle the enormous amount of Big Data by connecting and analysing the mobile Big Data with other stores. It will provide a wider, comprehensive view of understanding customers' behaviours in multiple locations. Furthermore, deploying mobile Big Data in the whole chain stores can help to reduce the cost and improve performance of processing data since now not only the case company but also other stores can get access and process the collected mobile Big Data within the Elgiganten AB. Consequently, the interesting comparative analytics from the different stores can be accumulated and compared. The results from one store might be applied for others.

By elevating the mobile Big Data deployment to all the chain stores in the Elgiganten AB, each store can adopt mobile Big Data with significantly lower initial investment costs. The sensing systems will be required to be equipped for each individual store. However, the Big Data appliance can be shared and used from all the stores as a Big Data appliance centre. The data collected from sensing systems in each store can be stored in the cloud storage of the Elgiganten AB. Mobile Big Data applications and analytics running on the Big Data appliance centre can analyse the data directly from the cloud. The analysed results will be sent back to the requested stores for further examination.

The store managers in each store in the Elgiganten AB can work as a business analyst to examine the results. Nonetheless, the data architect, the Hadoop developer and the system administrator can work in the single Big Data appliance centre. Thus, the labour costs for operating Big Data appliance centre can be shared between all the chain stores within the group. The human resources can be optimised and the labour costs are reduced while the stores can get the full advantage of the mobile Big Data. Table 4 presents the cost for the individual stores and the costs that can be shared between the stores when adopting mobile Big Data in all the chain stores in the Elgiganten AB.

Table 4. The costs for individual stores and shared between stores

| Components | | Individual stores | | | | Shared between stores | | | |
|--------------|--------------------|-------------------|--------|--------|-------------------|-----------------------|----------|----------|------------------|
| | | Year 1 | Year 2 | Year 3 | Total in 3 years | Year 1 | Year 2 | Year 3 | Total in 3 years |
| Hardware | Sensing systems | €10,000 - €60,000 | €0 | €0 | | | | | |
| | Big Data Appliance | | | | | €465,000 | €50,000 | €50,000 | |
| Software | | | | | | €0 | | | |
| Human actors | | | | | | €192,000 | €130,500 | €130,500 | |
| Total | | €10,000 - €60,000 | €0 | €0 | €10,000 - €60,000 | €657,000 | €180,500 | €180,500 | €1,018,000 |

As shown in table 4, when the Big Data appliance centre is shared between the stores, the costs become more reasonable for the individual stores. There is sensing systems cost that each store has to independently invest in. However, the costs of Big Data appliance, software and human actors are shared between all the 70 stores in the Elgiganten AB. As the result, each store will have to invest only €10,000 to €60,000 for the sensing systems plus the shared cost $€1,018,000 / 70 = €14,500$ for operating the Big Data appliance in the period of three years. The total solution cost for each store for three years will be from $€10,000 + €14,500 = €24,500$ to $€60,000 + €14,500 = €74,500$.

6 CONCLUSIONS

Today, with the advance in technology, mobile sensors can track and determine customers' motions and in-door positions in detail and with high accuracy. The generated data from these sensors can be examined to understand the customers' movements and activities. However, most sensor data is unstructured and semi-structured in which the conventional databases and business intelligent tools are unable to capture and make sense of. This is where mobile Big Data comes in. Mobile Big Data offers the ability to effectively manage and analyse the enormous and complex amount of the sensor data. The information derived from mobile Big Data can provide many valuable insights for the case company. By adjusting to data-driven decision-making, the case company can enhance the business in ways that were not previously achievable.

Researching the potential of mobile Big Data is the focus of this thesis. Different sensor technologies and Apache Hadoop tools were studied and its applicability was evaluated as if it were deployed in the case company. Business value of mobile Big Data was identified and it is shown that mobile Big Data can enhance both the economic value and customer value for the case company. Nonetheless, the business value requires an appreciable amount of time to be realised.

The results of the in-depth interviews and the questionnaire made it possible to answer the research questions as well as provide a solid foundation to facilitate the adoption of mobile Big Data in the case company. It is shown that the case company has a solid IT infrastructure with the coherent policies to protect the customer rights. Furthermore, customers as the most crucial factor for the thriving of the case company have positive, open-minded perceptions on the collection of the location data and its usage to improve their shopping experience as long as they are conscious about it.

The findings of the study indicate that mobile Big Data if properly exploited can have the positive impacts on the business decision-making in the case

company. The visualised information from mobile Big Data provides to a greater extent to the understanding of the customers' behaviours and habits in the store. By delving into the visualised store maps, the case company can grasp many valuable insights which could not be able to acquire before.

To conclude, mobile Big Data is a relatively novel technology for handling and analysing large amounts of different types of data, e.g. sensor logs, location data and motion data. Mobile Big Data is still in its early stages, yet it is considerably getting attention from researcher and developer community. Thus, many more up-and-coming applications are being developed and improved to take full advantage of mobile Big Data. This research addressed some major feasible potential of mobile Big Data for the case company. Mobile Big Data can be notably beneficial for the business decision-making of the case company. However, the technical challenges associated with mobile Big Data and sensor technologies need to be taken into account due to their novelty and complexity. Therefore, deploying mobile Big Data in the case company requires a scrupulous plan to avoid any issues that might occur and the members of staff need to have fundamental knowledge and understanding of mobile Big Data to operate the system in the long-term context.

On the basis of this research, the business scenario was scrutinised to demonstrate the requirements of mobile Big Data technologies and the total solution cost for the case company. However, the business scenario is not enough to deliver a comprehensive deployment plan. Therefore, further research will be necessary to evaluate the reality when the mobile Big Data is physically deployed in the case company's store. The projected return on investment of the deployment needs to be assessed. Moreover, it is suggested to conduct research on developing stable and reliable devices which can aggregate and operate the motion sensors to generate and transfer the activity and location data automatically to the HDFS. Additionally, it would be beneficial to explore and study in-depth the Hadoop ecosystem which particularly aims for sensor logs and position data by developing, coding, testing and implementing the program based on the concrete platform proposed in this research.

Further, it is important to point out that mobile Big Data does not only appear to be useful by providing the insightful information in the retailing sector. Mobile Big Data technologies can be applied for other sectors. For instance, personal location data can be deployed in the airports to track the passengers' movements. By understanding the passengers' movements and habits, the airports' managers can optimise the routes to the gates to reduce the transferring time for the passengers. Moreover, activity and location data can help to detect suspicious activities and abnormal movements in the airports. Consequently, the security, the safety and the travelling experience of the passengers can be enhanced. Besides, the applications of mobile Big Data can be extended to be used in other different environments such as amusement parks, shopping malls, hospitals and restaurants.

REFERENCES

- Aaronia AG 2011. Aaronia GPS Logger with 6 Sensors. GPS, 3D Gyro, 3D Tilt, Digital Compass, Height Sensor & Accelerometer in one device. Retrieved 8 October, 2014.
<http://www.aaronia.com/Datasheets/Spectrum_Analyzer/Aaronia-Multi-Data-GPS-Logger.pdf>
- Akerkar, Rajendra 2014. Big Data Computing. Boca Raton, London, New York: CRC Press.
- Atlassian Confluence 2012. Apache Flume / Home. Flume NG. Retrieved 14 September, 2014.
<<https://cwiki.apache.org/confluence/display/FLUME/Flume+NG>>
- Atlassian Confluence 2014. Apache Hive / Home. Retrieved 16 September, 2014.
<<https://cwiki.apache.org/confluence/display/Hive/Home>>
- Belezina, Jan 2012. Magnetic field anomalies used to pinpoint your location in a building. Retrieved 16 September, 2014.
<<http://www.gizmag.com/magnetic-anomaly-indoor-positioning/23253/>>
- Bloch, Michael & Hoyos-Gomez, Andres 2009. How CIOs should think about business value. McKinsey on Business Technology. McKinsey & Company. Retrieved 27 September, 2014.
<http://www.mckinsey.com/insights/business_technology/how_cios_should_think_about_business_value>
- Carr, Nicholas G. 2003. IT Doesn't Matter. Harvard Business Review. Retrieved 24 September, 2014.
<<http://hbr.org/2003/05/it-doesnt-matter/ar/1>>
- Chen, Kong & Janz, Kathleen & Zhu, Weimo & Brychta, Robert 2012. Redefining the roles of sensors in objective physical activity monitoring. Med Sci Sports Exerc. Volume 44. Retrieved 6 September, 2014.
<<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245644/pdf/nihms298772.pdf>>
- Clifford, Stephanie & Hardy, Quentin 2013. Attention, Shoppers: Store Is Tracking Your Cell. The New York Times. Retrieved 26 August, 2014.

- <http://www.nytimes.com/2013/07/15/business/attention-shopper-stores-are-tracking-your-cell.html?pagewanted=all&_r=1&>
- Craig, Terence & Ludloff, Mary E. 2011. Privacy and Big Data. Sebastopol: O'Reilly Media, Inc.
- Davis, Kord & Patterson, Doug 2012. Ethics of Big Data. Sebastopol: O'Reilly Media, Inc.
- Devlin, Barry 2014. The Marriage of BI and Big Data - Business unIntelligence. ACM Webinar. Retrieved 17 June, 2014.
- <<http://event.on24.com/eventRegistration/EventLobbyServlet?target=lobby.jsp&eventid=789901&sessionid=1&key=ADE41CA936251C00FC24EA0C9C5FF746&eventuserid=99533025>>
- Dijicks, Jean-Pierre 2014. The Data Warehouse Insider. Updated: Price Comparison for Big Data Appliance and Hadoop. Oracle Corporation. Retrieved 12 October, 2014.
- <https://blogs.oracle.com/datawarehousing/entry/updated_price_comparison_for_big>
- Duncan, Mitch & Badland, Hannah & Mummery, W.Kerry 2008. Applying GPS to enhance understanding of transport-related physical activity. Journal of Science and Medicine in Sport. Elsevier, Volume 12, Issue 5, 549-556. Retrieved 7 September, 2014.
- <<http://www.sciencedirect.com/science/article/pii/S1440244008002107>>
- Euclid 2014a. How Euclid location analytics works. Retrieved 22 September, 2014.
- <<http://euclidanalytics.com/products/technology/>>
- Euclid 2014b. Telephone discussion, sales representative, Euclid Inc. Conducted on 10 October, 2014.
- Given, Lisa M. 2008. The SAGE Encyclopedia of QUALITATIVE RESEARCH METHODS. Volume 1 & 2. USA: Sage Publications Ltd.
- GSMarena 2014. Glossary: Sensors. Retrieved 8 September, 2014.
- <<http://www.gsmarena.com/glossary.php3?term=sensors>>
- Haverinen, Janne & Kemppainen, Anssi 2009. Global indoor self-localization based on the ambient magnetic field. Elsevier, Volume 57, Issue 10, 1028–1035. Retrieved 7 September, 2014.
- <<http://www.sciencedirect.com/science/article/pii/S0921889009001092>>

- Hortonworks 2014a. Hadoop YARN: The Architectural Center of Enterprise Hadoop. Retrieved 10 September, 2014.
<<http://hortonworks.com/hadoop/yarn/>>
- Hortonworks 2014b. Apache Pig. A scripting platform for processing and analyzing large data sets. Retrieved 22 September, 2014.
<<http://hortonworks.com/hadoop/pig/>>
- Hurwitz, Judith & Nugent, Alan & Helper, Fern & Kaufman, Marcia 2013. Big Data For Dummies. New Jersey: John Wiley & Sons, Inc.
- Kaddourah, Youssef & King, Jeff & Helal, Abdelsalam 2005. Cost-Precision Tradeoffs in Unencumbered Floor-Based Indoor Location Tracking. The third International Conference on Smart Homes and Health Telematics (ICOST). Retrieved 8 October, 2014.
<<http://www.harris.cise.ufl.edu/projects/publications/Youssef-ICOST-2005-Final.pdf>>
- Kohli, Rajiv & Grover, Varun 2008. Business Value of IT: An Essay on Expanding Research Directions to keep up with the Times. Journal of the Association for Information Systems. Volume 9, Issue 1, Article 1, 23-39. Retrieved 7 September, 2014.
<<http://aisel.aisnet.org/jais/vol9/iss1/1/>>
- Krishnan, Krish 2013. Data warehousing in the age of big data. USA: Morgan Kaufmann.
- Manyika, James & Chui, Michael & Brown, Brad & Bughin, Jacques & Dobbs, Richard & Roxburgh, Charles & Byers, Angela H. 2011. Big data: The next frontier or innovation, competition, and productivity. McKinsey Global Institute. Retrieved 15 April, 2014.
<http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation>
- Martin, Angela 2013. Nordstrom Using Smart Phones To Track Customers Movements. CBS Local Media. . Retrieved 15 September, 2014.
<<http://dfw.cbslocal.com/2013/05/07/nordstrom-using-smart-phones-to-track-customers-movements/>>
- Maxwell, Joseph 1996. Qualitative research design: An interactive approach. Applied Social Research Methods Series. Volume 41. London, New Delhi: Sage Publications Ltd.

- McGrath, Michael & Scanaill, Clíodhna 2014. *Sensor Technologies: Healthcare, Wellness, and Environmental Applications*. New York: Apress Media.
- Middleton, Lee & Buss, Alex & Bazin, Alex & Nixon, Mark 2005. A floor sensor system for gait recognition. *Automatic Identification Advanced Technologies. Fourth IEEE Workshop*, 171-176. Retrieved 20 September, 2014.
<http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=1544421&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D1544421>
- Mohanty, Soumendra & Jagadeesh, Madhu & Srivatsa, Harsha 2013. *Big Data Imperatives: Enterprise Big Data Warehouse, BI Implementations and Analytics*. New York: Apress Media.
- O'Reilly Media 2012. *Big Data Now: Current Perspectives from O'Reilly Media. Second Edition*. USA: O'Reilly Media, Inc.
- Orr, Robert J. & Abowd, Gregory D. 2000. *The Smart Floor: A Mechanism for Natural User Identification and Tracking*. Georgia Institute of Technology. Retrieved 24 September, 2014.
<<https://smartech.gatech.edu/bitstream/handle/1853/3321/00-02.pdf>>
- Oxford University Press 2014. *Privacy*. Retrieved 24 September, 2014.
<<http://www.oxforddictionaries.com/definition/english/privacy>>
- Ohlhorst, Frank 2012. *Big Data Analytics: Turning Big Data into Big Money*. USA: John Wiley & Sons, Inc.
- PayScale Inc. 2014. *Salary Data & Career Research Center (Finland)*. Retrieved 4 October, 2014.
<<http://www.payscale.com/research/Fl/Country=Finland/Salary>>
- Putila, Harri 2014a. Email. Store manager, Elgiganten AB – Haparanda. General information on Haparanda store for the research. Sent to email address: [duchung.luu\(at\)edu.lapinamk.fi](mailto:duchung.luu(at)edu.lapinamk.fi). Conducted on 22 August, 2014.
- Putila, Harri 2014b. Interview of Mr. Harri Putila, Store Manager of Elgiganten AB – Haparanda. Conducted on 16 September, 2014.
- Putila, Harri 2014c. Interview of Mr. Harri Putila, Store Manager of Elgiganten AB – Haparanda. Conducted on 23 September, 2014.
- Ryabov, Vladimir 2014. Discussion. Principle Lecturer, Lapland University of Applied Sciences, Tornio. Conducted on September 17, 2014.

Seiko Epson 2014. Gyro sensors – How they work and what’s ahead. Retrieved 8 September, 2014.

<http://www5.epsondevice.com/en/sensing_device/gyroportal/about.html>

Society for Human Resource Management 2008. Salaries as a Percentage of Operating Expense. Retrieved 20 April, 2014.

<<http://www.shrm.org/research/articles/articles/pages/metricofthemothsalarieraspercentageofoperatingexpense.aspx>>

The Apache Software Foundation 2014a. Welcome to Apache™ Hadoop®!. Retrieved 8 September, 2014.

<<http://hadoop.apache.org/#What+Is+Apache+Hadoop%3F>>

The Apache Software Foundation 2014b. Apache Hadoop NextGenMapReduce (YARN). Retrieved 18 September, 2014.

<<http://hadoop.apache.org/docs/r2.3.0/hadoop-yarn/hadoop-yarn-site/YARN.html>>

Theiss, Adam & Yen, David & Ku, Cheng-Yuan 2004. Global Positioning Systems: an analysis of applications, current development and future implementations. Computer Standards and Interfaces. Elsevier, Volume 27, Issue 2, 89–100. Retrieved 7 September, 2014.

<<http://www.sciencedirect.com/science/article/pii/S0920548904000649>>

Vavilapalli, Vinod & Murthy, Arun C. & Douglas, Chris & Agarwal, Sharad & Konar, Mahadev & Evans, Robert & Graves, Thomas & Lowe, Jason & Shah, Hitesh & Seth, Siddharth & Saha, Bikas & Curino, Carlo & O’Malley, Owen & Radia, Sanjay & Reed, Benjamin & Baldeschwieler, Eric 2013. Apache Hadoop YARN: yet another resource negotiator. Proceedings of the 4th Annual Symposium on Cloud Computing (SOCC '13). Retrieved 17 September, 2014.

<<http://doi.acm.org/10.1145/2523616.2523633>>

White, Tom 2012. Hadoop: The Definite Guide, Third Edition. Beijing, Cambridge, Farnham, Köln, Sebastopol, Tokyo: O’Reilly Media, Inc.

Winter, Richard & Gilbert, Rick & Davis, Judith R. 2013. Big Data. What Does It Really Cost?. Winter Corporation. Retrieved 30 September, 2014.

<<http://www.wintercorp.com/login-tcod-report/>>

XE.com 2014. XE Currency Converter. Retrieved 15 October, 2014.

<<http://www.xe.com/currencyconverter/convert/?Amount=1&From=USD&To=EUR>>

Yin, Robert K. 2009. Case Study Research: Design and Methods. Fourth Edition. UK, India, Singapore, USA: Sage Publications, Inc.

Yuk, Mico & Diamond, Stephanie 2014. Data Visualization For Dummies. Hoboken, New Jersey: John Wiley & Sons, Inc.

Zikopoulos, Paul & Eaton, Chris & Deroos, Dirk & Deutsch, Tom & Lapis, George 2012. Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. Second Edition. USA: McGraw-Hill Osborne media.

APPENDICES

Appendix 1

1(5)

INTERVIEW TRANSCRIPT WITH THE STORE MANAGER OF ELGIGANTEN
AB – HAPARANDA, CONDUCTED ON 17 SEPTEMBER, 2014.

Interviewer: Could you describe the current IT infrastructure in our store?

Interviewee: We have five different departments in our store. When you come to our shop, you can see the first department that has mobile phones, GPS devices and cameras. After that you get to the computers, printers and other accessories. Then you come to photo and video department, you can see all kinds of TVs and audio-visual products. About IT, we have a lot of instruments to do the ranking of everything in the store. In our sale program, you can look for our sales and how much stuff do we have, and customers buying products rate, how many products did they buy per hour. The sale program also helps our logistic to check how many products are in the store and how many products have to order more from main warehouse. We also have promotional program to look for forecast and shopping trend for 3 weeks ahead. Today, our main way to advertise in the whole co-operation, Elgiganten in Sweden, is advertising newspapers. These newspapers will come every Monday and usually last until Sunday.

Interviewer: How important do you think to understand customers' in-store behaviours?

Interviewee: I think this is the most important thing. Because in our company Elgiganten, if we went back like 5 years ago, it was not so important. The most importance was to sell and you did not have to listen to customers. You just sent the advertising newspapers with good prices and customers came and then you just sold and got money. You did not have to care if it was the right things for customers or not. The most importance was to get the volume. But today, we try to educate every staff that it is crucial to greet customers. When you see customers you have to say "Hello" in the first place and ask if they need

any help, then you have to listen to customers. The customers say to you what he/she needs you do not say anything. You just listen and make good suggestions according to their needs. We have morning meetings before the shop opens and today, I asked my staff if they say hello to every customers and all said yes. However, I used mystery shoppers and got the result yesterday. It is shown that not every staff says "Hello" to customers and I have to take it into account. Because if the customers do not come here, we do not have business. So yes, it is very important to know how customers behave and want in the store.

Interviewer: If you can get insights into customers' behaviour in the store, how will the understanding help to improve business decision making?

Interviewee: Yes. It can help many things in our decision making. For instance, not in this business but when I worked in IKEA, when you see the customers' in-store routes map, you can get the idea of why several customers walked like that. But then when you follow the customers' routes, you can look at what kind of products that you have there. Maybe you have a wrong product there and you have an opportunity to put the better one like the one you want to sell to customers without customers knowing about that. And sometimes you maybe have like another store selling well this particular laptop but your store does not. Then you have to look where is this laptop placed in our store and if you know where customers moved, maybe you have had placed it in wrong place. So you should put this laptop in the front to get more attention.

Interviewer: Do you know Big Data, particularly personal location data and its applications to track customers' movements in the store?

Interviewee: No, not really. We have not used it. However, we have a machine; we use to measure how happy customers are when they leave our store. The machine "Happy-or-Not" has 4 buttons: deep green for very happy, green for quite good, orange for okay and red for upset, customers can push one of the buttons. The data will then be collected every morning and compared with other

stores every week. The lowest stores will get support and supervise from the headquarters.

Interviewer: What are the challenges do you think new technologies, e.g. mobile Big Data might pose to our store?

Interviewee: I think the challenges can be installing the system in the store and training our staff to use it. We do not have many full-time staff. The challenge can happen when you have to maintain it or when the system is down. It happened with our “Happy-or-Not” machine.

Interviewer: What is the budget range the company want to invest on mobile Big Data?

Interviewee: I think our parent company makes a lot of investment in order to know customers better. So we bought things like “Happy-or-Not”. We also ask for email subscription acceptance from every customer. So that you can send the new products and promotional offer electronically. You do not need this kind of advertising newspapers maybe in the future and that much cheaper. And there with online offer, you can also see the behaviour of customers, like okay he is interested in computer section so he does not need advertise for like coffee machines or TVs. So the company makes a lot of investment in that kind of technology.

Interviewer: So it means that you are willing to invest in this new technology if it is effective and help you to know customers better?

Interviewee: Yes, we do.

Interviewer: What do you think about ethical concerns when using customers’ location data?

Interviewee: You mean like collecting customers data?

Interviewer: Yes, like when you are using customers’ personal location data in our store, what are your views on this?

Interviewee: I think it is okay. If the customers, you know when the customers take the email acceptance, we ask the customers if it is okay to collect this email and we will not use this email for any other purposes. It is only for like the receipts come electronically to his/her email. And maybe you have to come back, you have some problem with your product, you always have the receipt of this product in your email. Because every time you leave your email after purchase products, the receipts will automatically go to your email. So I think as long as we keep the promise that we do not use the customer data for any other purposes and as long as you tell this to the customers and he/she has decision by themselves. I think it is okay and we also say that if you do not like the email subscription, you can come and say to us that “No, no we do not want it at all” or you can unsubscribe by yourself in your email.

Interviewer: Does our company have any policies to protect customers' privacy?

Interviewee: Yes, as I already said we do not give our customers' data. Like in Sweden, we have all kind of companies that want to have customers' data and we do not give to any others, it is only for our company and this is also to protect the customers. Because you can have like telephone numbers, addresses and other things that they do not want anyone to know. And then we have customers who do not want to leave any kind of information about themselves and that is also okay. You do not need to provide a name or any information about yourself when shopping with us if you do not want to. It is a free choice.

Interviewer: So are you interested in what customers' attitude towards personal location data and how it impacts the business? Since I plan to do a questionnaire and conduct a customer survey about this issue.

Interviewee: Yes, I am very interested in knowing what customers think about it. Because I think in there you can find a lot of things to do better.

Interviewer: Do you want to provide more information for this interview? It can be anything related to mobile Big Data, customers' behaviours or moral aspect.

Interviewee: Yes, like I said to you that today, this company thinks that the most important people are the customers. Maybe like 4, 5 years ago, it was not on the top. You know at that time the main goal was to just be number 1 in Sweden. But now, we still want to be in number 1 but we also want to have the happiest customers. Because, we think if you have a happy, satisfied customer, then he/she will always come back to your store. We want to build a good reputation. Additionally, our education always starts with those like "Say hello to the customers", "Listen to the customers" and then you always "Say goodbye to them".

INTERVIEW TRANSCRIPT WITH THE STORE MANAGER OF ELGIGANTEN AB – HAPARANDA, CONDUCTED ON 23 SEPTEMBER, 2014. *(Confidential)*

Interviewer: Could you tell me more detail about the sale system in our store?

Interviewee: Yes, it is the same system in all the Nordic countries. We have it in Elgiganten in Sweden, Gigantti in Finland and Elkjøp in Norway and Denmark. So the sale system is the program built with another logistic program. They work together.

Interviewer: So there are two separate programs?

Interviewee: Yes, you have one for logistics and one for sale. But they are working together. The logistic program is the heart. Then you have the sale program connected to that logistic one.

Interviewer: Could you describe the structure of two programs and how does it work?

Interviewee: The system we have is very simple. It looks like 1970s' system. It is based on DOS. And for every product you have unique codes. When you write it down to the logistic system, you get the store information, how many does it have and the prices. So it is quite simple and old system but it is working well. I can show it to you.

Interviewer: Yes, please.

Interviewee: The system is called "Sallo" in Swedish. You have several functions in here like ranking for one day, or longer time. Now we go to the sale program in our store. You enter the product code in the system. You see the price, product information and available in the stock and holding numbers. For more detail, logistic programs can check if new products have come to our store. And I can see the ordinary price and campaign price. There are also all the product physical details and orders.

Interviewer: So how does the system work when customers order some goods online?

Interviewee: Okay. We have something called "Collective store". You as customers can have several ways to make remote orders. They can call service centre and they can make orders from there. They can make products delivered directly to home from our Elgiganten main warehouse or they can call the service centre and come to pick it up in our store if this product is available. Then you also order it online through our Swedish website. In our system, the printer will print all the orders every night. So when we come here in the morning. Okay. Luu has ordered this one so we take the product and make it ready for you. Then you come and pick it up.

Interviewer: So the sale system is also responsible for online orders?

Interviewee: Yes. And here you can see when you check a product in our website. There will be a list of numbers of available products in all the stores. We also have post package service. When customers come to our store but we do not currently have this product. But you can pay for this; we will send it to you by post when it is available. So you have many ways to buy products.

Interviewer: What is your prospect of the future marketplace?

Interviewee: I think there will be no problems for the next 10 years. For the old generation like me, I am quite familiar with this online system because I work with it but I am not so interested in computers and something like that. But for the young generation like you, you get used to Internet and find things. Maybe in next 10 years, I think most people will do shopping online and pick the item here or by post.

Interviewer: What do you think would be the challenge to understand customers' behaviour?

Interviewee: I think the biggest issue is to educate the staff the way customers behave today and in couple of years ahead. So the whole Elgiganten have to go with the further development on how customers buy and interact with products

and we have been working with it. You know when we are collecting emails, it is an easy and efficient way to contact customers and you do not need printed advertising newspapers. Because we, elder people, come to store. But you, youngsters, maybe just come to the store to check at it and then okay, it looks like that. I will come home and order it online.

Survey on collecting Personal Location Data in Elgiganten retail stores**The Elgiganten AB – Haparanda**

We are conducting a research on the potential of mobile Big Data, particularly Personal Location Data for business decision making.

Filling in this survey will significantly help the store to improve our customer service and your shopping experience.
It will take approx. 3-5 minutes to complete the survey.

Thank you so much for participating in our survey. Wish you have a nice day :)

1. What country do you currently live in?

- Sweden
- Finland
- Other countries

2. Please indicate your gender

- Male
- Female

3. Please indicate your age range

- Under 18
- 19 - 25
- 26 - 44
- 45 - 55
- Over 55

4. How often do you visit Elgiganten - Haparanda store?

- Several times a Week
- Several times a Month
- Several times a Year
- Never

5. How much time do you often spend in our store?

- Less than 20 minutes
- Around 20 - 60 minutes
- More than 1 hour

6. What are your main purposes when visiting our store?

- Buy wanted products
- Check prices and look at products before buying it
- Window shopping

7. Do you often take shopping cart/basket with you when visiting supermarkets/ other stores?

- Yes
- No

8. Do you carry a smartphone when visiting our store?

- Yes
- No

9. What do you think if we collect your walking route and activity anonymously in our store?

| | Strongly Disagree | Disagree | Neither Disagree Nor Agree | Agree | Strongly Agree |
|--------------------------------|-----------------------|-----------------------|----------------------------|-----------------------|-----------------------|
| If you are notified in advance | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| If you are not notified | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

10. If you have a suggestion or recommendation on using your movements in our store. Please provide it below.



Send