BIG DATA : MERGING DIVERSE PRODUCTION FILES IN MANUFACTURE OF ELECTRONIC COMPONENTS AND BOARDS INDUSTRY.

LEOWANIE LEOW SUI QIN

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

BIG DATA : MERGING DIVERSE PRODUCTION FILES IN MANUFACTURE OF ELECTRONIC COMPONENTS AND BOARDS INDUSTRY.

LEOWANIE LEOW SUI QIN

The thesis is submitted in partial fulfillment of the requirements for the award of Bachelor of Technology Management (Technology Innovation)

Faculty of Technology Management & Technopreneurship
Universiti Teknikal Malaysia Melaka

JUNE 2016

‘I/ We, hereby declared that I/We had read through this thesis and in my/our opinion that this thesis is adequate in terms of scope and quality which fulfill the requirements for the award of Bachelor of Technology Management (Technology Innovation)

SIGNATURE : _____

NAME OF SUPERVISOR : DR. FAM SOO FEN

DATE : _____

SIGNATURE : _____

NAME OF PANEL : PUAN EDNA BINTI BUYONG

DATE : _____

DECLARATION

"I admit that this report is the result of my own, except certain explanations and passages where every of it is cited with sources clearly."

SIGNATURE : _____

NAME : LEOWANIE LEOW SUI QIN

DATE : _____

DEDICATION

For my beloved family and friends

ACKNOWLEDGEMENTS

I would like to express my greatest gratitude to the individuals who have helped and supported me throughout my Undergraduate Project Programme (PSM) entitled Big Data : Merging Diverse Production Files in Manufacture of Electronic Components and Boards Industry.

I am very grateful and thankful to my supervisor, Dr. Fam Soo Fen for her continuous support and guidelines from precise and constructive advices with ongoing guidance and encouragement for my final year project. I am very grateful and truly appreciate Dr. Fam for allocating her precious time and infinite coaching for which I have gained abundance of new knowledge and unforgettable life-time experience in learning a new software, R in the process of completing my research project. Dr. Fam's guidance and coaching throughout the journey of my PSM has certainly improved my critical thinking and analytical skills to empower me to become a more resourceful, innovative and creative person.

A significant thank you goes to my panel, Puan Edna binti Buyong for her helpful and positive comments and guidance for my final year report. A very grateful and thankful note for the compassionate advice on improving my final year report.

I would also like to wish a very sincere thank you to my parents for their continuous support who had inspired and encouraged me to complete this report. Their moral support to is my motivation to complete this report fully with my dedication. I am also appreciative towards my peers for their fruitful advices and infinite sharing of our journey to complete our individual projects.

ABSTRACT

This research highlights the challenges of Big Data phenomenon in the merging process of multiple production files in the manufacturing of electronics industry. Datasets produced in the production floor is vast in volume, high in velocity and comes in variety due to the data are obtained from different machines. The Big Data revolution has challenge the merging process of diverse files in the production floor as merging the datasets is important for further analysis of the data to gain insights. The characteristics and review of Big Data 5V is illustrated in this research. This research describes the software R, a statistical package as an alternative to merge diverse data files and then proceeds to illustrate the use of R by the steps of import data from Microsoft Excel and also merging the diverse Big Data files in the production floor by using R. In addition, a survey is carried out with respondents from the manufacturing of electronics components and boards industry as participants to study the awareness and acceptance of software R as an alternative for merging Big Data production files in the electronic components and boards manufacturing industry, in Penang, Malaysia. This research explores the software R and employs R as an important role for alternative in merging diverse Big Data files the manufacturing production floor prior to the Big Data phenomenon. As a conclusion, R plays an important role as an alternative for merging diverse Big Data files in the Big Data revolution in the manufacturing production floor of electronic components and boards manufacturing.

Keywords : Big Data, R Statistical Programme, Electronics Manufacturing, Merging diverse files

ABSTRAK

Kajian ini menggariskan cabaran fenomena Big Data dalam proses penggabungan berbilang fail pengeluaran dalam pembuatan industri elektronik. Kumpulan data yang dihasilkan di lantai pengeluaran adalah luas dalam jumlah, tinggi halaju dan datang dalam pelbagai disebabkan oleh data yang diperolehi dari mesin yang berbeza. Revolusi Big Data mempunyai mencabar proses penggabungan yang pelbagai fail di lantai pengeluaran sebagai menggabungkan set data adalah penting untuk analisis selanjutnya data untuk mendapatkan pandangan. Ciri-ciri dan kajian Big Data 5V juga digambarkan dalam kajian ini. Kajian ini menerangkan R perisian, pakej statistik sebagai alternatif kepada bergabung fail data yang pelbagai dan kemudiannya untuk menggambarkan penggunaan R oleh langkah-langkah data import dari Microsoft Excel dan juga menggabungkan pelbagai fail Big Data di lantai pengeluaran dengan menggunakan R. Di samping itu, kajian dijalankan dengan responden daripada pembuatan komponen industri elektronik dan papan sebagai peserta untuk mengkaji kesedaran dan penerimaan perisian R sebagai alternatif untuk menggabungkan fail pengeluaran Big Data dalam industri pembuatan komponen elektronik dan papan, di Pulau Pinang, Malaysia. Penyelidikan ini juga meneroka perisian R dan menggaji R sebagai peranan yang penting untuk alternatif dalam penggabungan pelbagai Data Big fail di pengeluaran pembuatan dalam fenomena Big Data. Kesimpulannya, R memainkan peranan yang penting sebagai alternatif untuk menggabungkan pelbagai fail Big Data dalam revolusi Big Data di lantai pengeluaran industri pembuatan komponen elektronik dan papan.

Kata Kunci : Big Data, Program Statistik R, Industri Pembuatan Elektronik, Pengabungan pelbagai fail

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATION

ABBREVIATION	MEANING
ANZ	Australia-New Zealand Bank
Big Data 5V	Big Data Volume, Variety, Velocity, Veracity, Value
FIZ	Free-Trade Industrial Zone
GDP	Gross Domestic Product
IDC	International Data Corporation
IBM SPSS	IBM Statistical Product and Service Solution
R&D	Research and Development
ROI	Return On Investment
SAS	Statistical Analysis System
α	Confidence Level

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

Towards the 21st century today, the rapid development of technology has contributed to the growth of information age and the era of Big Data. The exponential increase in the amount of data generated by the technology savvy individuals and organizations has overloaded the server network and systems contributing to the Big Data hype. Big Data is an axiom to portray large volumes of high velocity, dense and variable data that needs innovative procedures and technologies to facilitate the capturing, loading, sharing, managing, and studying of the information. Similarly, (Emani, et al., 2015) defines big data as information and datasets whose size is beyond the capability of typical database software tools to capture, store, manage and analyze.

Big Data has been captivating its form in organizations, as it is ordinary for any corporation to generate and traffic data in capacity exceeding seven exabytes yearly due to its operations. One exabyte of data is comparable to more than 4,000 times the information stowed in the United States of America Library of Congress (Nedelcu, 2013). In the competitive manufacturing environment of electronic components and boards, vast amount of rich multiple datasets are gained in the production floor. This is due to the increase of records measurements from different physical surroundings and processes to support the manufacturing facilities (O'Donovan, et al., 2015).

According to (Wilschut, et al., 2015) the multiple datasets gained in an average electronics manufacturing production floor exceeds twenty gigabytes daily. The multiple datasets collected by employees frequently require further analysis to gain context and insights of the data as a whole for better understanding. (Lee, et al., 2013) discuss the present manufacturing procedures involve better-quality analysis of diverse multiple datasets from machines and processes. In addition, the electronics manufacturing data includes diverse levels of characters and can be categorized by multiple initials (Moyne, et al., 2015).

(Alexandros and Jagadish, 2012) explains the value of datasets increases when it can be interrelated with other data, thus merging datasets is a main initiator of creating a new value. The merging of multiple files of large datasets into single file collected by the employees in the production floor remains a challenge due to the nature of Big Data. This research studied the merging of diverse multiple datasets in the Big Data environment to support the next-generation infrastructure and technologies for electronics manufacturing.

1.2 Problem Statement

In an average corporation, massive amount of datasets is needed by top management to propose the best decision making and forecasting for production floor projections. This is due to the fact that vast data is needed for further analysis may come from different departments, employees and machines. Manufacturing procedure involves greater analysis of innumerable data from different machines and processes in the production floor (Lee, et al., 2013). The critical challenge prior to further analysis of immense amount of data is to merge the diverse source of multiple vast datasets in a single worksheet. In the context of Big Data 5V, merging multiple datasets files exceeding gigabytes in volume pose a challenge as the large capacity of the datasets may cause the running software to experience lag, a visible pause between the action of the user and the reaction of the server in the running application.

(Wilschut, et al., 2015) proposed that physically manually skimming and categorizing the data is reflected as too labour exhausting and has prohibited the use of the Big Data in daily decision making by the top management in production floor. The common software used to merge datasets by the manufacturing industry is Microsoft Excel as it is user friendly and easy to use for trainees in data analysis. However, the fast moving pace of operations in the production floor requires advance software to sustain the capacity of the growing Big Data 5V phenomenon. According to (Tsuda, et al., 2015) there is a rising necessity for application to process enormous volumes of Big Data instantaneously by machine-to-machine configurations in the manufacturing field. An alternative method of merging Big Data is needed to enhance the performance of the existing software and the growing data capacity generated in the production floor.

This research was to identify an alternative method of merging Big Data multiple dataset files to solve the limitation of existing software in the fast pace development of manufacturing of electronic and electronic components and boards.

The research questions are as below:

- What are the current approaches in merging of diverse Big Data files?
- What are the obstacles that challenge merging of diverse Big Data files?
- What is the innovative approach strategy to overcome the challenge merging of diverse Big Data files?

1.3 Research Objectives

The era of Big Data phenomenon has drastically changed the manufacturing of electronics operations due to the vast amount of multiple datasets collected in the production floor daily for further analysis to gain insights and value. In addition, the multiple datasets is also important for top management as a decision-making support. However, the merging of the multiple datasets that exceeds twenty gigabytes daily

into a single file collected by employees in the production floor remains a challenge due to the Big Data nature. This research helps to bridge the fissure between merging the multiple datasets into a single file and Big Data nature to support the future of manufacturing infrastructure and systems.

The objectives of this research are:

- To explore the current approaches in merging of diverse Big Data files.
- To determine the obstacles that challenge in merging of diverse Big Data files.
- To identify the innovative approach strategy to overcome the challenges merging of diverse Big Data files.

1.4 Scope, Limitation and Assumptions of the Study

This research was focused to assess the current practices and obstacles in merging multiple Big Data files in the field of manufacture of electronic and electronic components and boards in Malaysia. The limitation of this research includes the context of merging multiple datasets in size of Big Data only. This research was also limited to studying the fissure of merging multiple datasets into a single file to ease data analysis in the production floor. This research was only solely focused to Microsoft Excel users to understand the challenge of merging multiple Big Data files.

There are two key assumptions of this study. First, this research assumed that the respondents answer the questionnaire accurately with honesty. Second, this research focused on respondents' behavior without taking the account of environmental and economical factors.

1.5 Importance of the Study

This study examined the current practices and obstacles in merging multiple Big Data files in manufacturing of electronics components and boards industry. This research pursued to understand the obstacles in merging of multiple Big Data files with the existing software available in the 21st century today. Besides that, this research also focused to recommend innovative methods to merge multiple Big Data files into a single file to enable employees to perform more resourceful and effective data analysis.

Furthermore, this research was significant to merging multiple Big Data files into a single file and the Big Data environment towards a more proficient workload in the production floor and to aid in effective data analysis.

1.6 Summary

The rapid growth of technology and information age has radically changed the business environment including the manufacturing field as datasets are being generated throughout the organization's daily operations leading to the Big Data phenomenon. The merging massive datasets produced and gathered in the production floor during the manufacturing operations remains a challenge. This is due to the available traditional software to process the Big Data has several bugs due to the Big Data nature. This research explored the nature of the Big Data and its challenges in the production floor of manufacturing of electronics components and boards. In addition, this research investigates the current practices of merging multiple Big Data files into a single spreadsheet file and the obstacles that restricts the competency of merging multiple Big Data files due to the limitation of the available software. Furthermore, this research enabled technical learning of the innovative suggestions to overcome merging of diverse Big Data files into a single spreadsheet file. This research offered depth understanding and new knowledge to alternative software available other than Microsoft Excel to merge Big Data diverse files into a single spreadsheet file.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The chapter discusses the definition and nature of Big Data. The Big Data and its environment are explored in details for further depth understanding. The literature review defines all concepts of merging multiple Big Data files in manufacturing of electronics components and boards was supported by academic articles from international journals. A theoretical framework is illustrated at the end of this chapter for the study of this research.

2.2 Definition and Nature of Big Data

The age of Big Data has approached the fast moving pace business world today. In the fast moving pace of development in economics and technologies of the twentieth century, infinite changes has brought to our lives, education, career, and just about everything in the world by the Internet and the way we work with information. The Big Data definition by (Bello-orgaz, et al., 2015) describes Big Data as high-volume, high-velocity and high-variety information resources that mandate cost-effective, inventive forms of information processing for enriched decision making. Simultaneously, (Krishnan, 2013) defines Big Data as datasets whose size is beyond the capability of typical database software tools to capture, store, manage and analyze.

Datasets that are so gigantic and complicated that traditional data processing systems are not able to execute analysis, transfer and visualization efficiently are denoted to as Big Data (Stich, et al., 2015). As datasets continues to grow, technologies to support the data application has to be further improve to accommodate the data generation by the manufacturing industry. Big Data has changed the information systems technologies to be more dynamic and enhanced to a higher technology to constitute the massive unlimited datasets generated from the production floor of electronics manufacturing industry.

Since the beginning of the computer era, the manufacturing sector was the first intensive consumer of datasets by implementing information technology to initiate quality enhancement and efficiency of the productions. Present manufacturing procedures need greater analysis of diverse datasets from machines and processes (Lee, et al., 2013). Merging of the datasets is important for analysis due to the nature of the Big Data and its environment (Nedelcu, 2013). However, manually merging and processing manufacturing the diverse datasets from hundred machines is labour exhaustive and time consuming in the production floor (Wilschut, et al., 2015). In addition, the merging of diverse datasets is challenging due to the Big Data nature and environment (Tsuda, et al., 2015).

2.3 Characteristics of Big Data Environment

Characteristics of Big Data began with three V's, which are volume, variety, and velocity (Chen and Storey, 2012). The three V's of Big Data characteristics, Volume, Variety and Velocity were originally presented by Gartner and Brynjolfsson from year 2012 to year 2013 (Fosso, et al., 2015). However, up to date, the three V's of Big Data has grew to five V's that adds the Value and Veracity characteristics into Big Data. The 5V Big Data environment has a significant influence towards the information age. With the rapid evolution and infinite datasets generation by human and machine, Big Data is here to stay. Thus, understanding the 5V environment of Big Data is critical for manufacturers and business organizations to exploit and benefit from Big Data.

Volume denotes the amount of scale of data size. (Chen and Storey, 2012) describes Big Data Volume is expressed to be in multiple terabytes and petabytes. The massive volume amount of data is so huge that traditional information systems may be dysfunctional to be able to store and analyze the data. According to (Fosso, et al., 2015) the volume attribute of Big Data is classified as the great volume of data that either consume vast storage or entail of large number of files. Data volume measures the amount of data available to an organization to access it.

Variety aspect of Big Data is described by the type of structured, unstructured or semi-structured data. Unstructured records are data of text, images, audio and video that is not simply not easy to be assessed by machineries (Chen and Storey, 2012). (Arun, 2014) categorized Big Data variety is the degree of the lushness of the data exemplification such as text, images video, and audio. In addition, from an analytic viewpoint, it is perhaps the major complication to successfully using sizeable volumes of data. However, (Emani, et al., 2015) describes Big Data variety as these data do not have a static configuration and hardly exist in a flawlessly systematic form and prepared for processing.

The third V is the Velocity of the Big Data. Velocity is the speed of data that can be delivered to the user or machine instantly when extracting data from a particular source. Velocity also incorporates the streaming of rapid data into mass storage for latter batch managing (Emani, et al., 2015) The degree of which data are produced and the rate at which it should be examined and acted upon as described by (Chen and Storey, 2012) is the velocity characteristics of Big Data. (Sharma, et al., 2014) explained that data is being gathered at a very fast speed.

The fourth V, named Value, was presented by International Data Corporation (IDC), Oracle and Forrester back in year 2012 (Fosso, et al., 2015). The fourth characteristic of Big Data comes into picture due to how does the data serve as the value when the data is uncountable, having a massive amount of volume. For any research, or any forecasting methods it is often said that the more data, the better, but how the data contributes to becoming a value.

The fifth and final V for Big Data characteristics is the Veracity proposed by White in year 2012 (Fosso, et al., 2015). Data veracity is much associated with the value of the Big Data as per discussed earlier. May it be massive amount of data, collected data used for forecasting, customer's data, any data that could be found, the trustfulness of the data remains on the person whom will analyze and investigates the data. Datasets collected are mainly used for decision making in the organization, thus veracity is dependent on the decision maker. Furthermore, (Fosso, et al., 2015) describes veracity as the characteristic of impulsiveness of some data obliges analysis of Big Data to achieve trustworthy forecast.

2.4 Approaches in Merging Diverse Big Data Files

According to (Bolón-Canedo, et al., 2015) datasets are being amassed at an extremely unprecedented speed and consistently processed precipitously for top management decision making. Moreover, datasets are no longer a result of the industry, but has become a vital connection of all aspects (Jin, et al., 2015). The current approach in merging diverse and multiple Big Data files in the production floor of electronics manufacturing industry is done manually through the copy and paste query in Microsoft Excel. As of datasets are retrieved from different machines and operating software in the production floor, merging the numerous datasets for further data analysis explores the capacity and capability of Microsoft Excel. Microsoft Excel serves as a platform to be used by every employee for their daily work task as it is user-friendly and has features for quick data analysis (Ravichandran and Kumar, 2015).

Another technique used to merge Big Data diverse files into a single spreadsheet file is through the command by macros in Microsoft Excel (Alexandros and Jagadish, 2012). Macro is sequence of commands and guidelines that can be categorized together as a specific command to complete a repetitive task automatically without any manual input. Employees are able to benefit from the macro function of Microsoft Excel to further shorten the time of merging the multiple datasets for immediate data analysis.