asiaeuniversity

THE BENEFITS, CHALLENGES AND CRITICAL SUCCESS FACTORS OF BIG DATA IN PROJECT MANAGEMENT - A LITERATURE REVIEW

BY
CHOY LAI FUN
C60401130004
MICTM

Project Paper Submitted in Partial Fulfillment of the
Requirement for the
Master in ICT Management

asiaeuniversity

THE BENEFITS, CHALLENGES AND CRITICAL SUCCESS FACTORS OF BIG DATA IN PROJECT MANAGEMENT - A LITERATURE REVIEW

BY
CHOY LAI FUN
C60401130004
MICTM

Project Paper Submitted in Partial Fulfillment of the Requirement for the Master in ICT Management

\bstract

n the information era today, the emerging of modern and digital economy has spurred an npprecedented growth of data usage and traffic across the world via our daily activities. This volution of data sources had introduced the term "BIG DATA".

ienerally, big data has been emerging rapidly for the recent decades to better serve the industries nd communities with enriched insight of the data assets. The deployment of big data enables usinesses to improve their performance with enhanced strategic advantages via sets of sophisticated nd innovative data exploration and exploitation tools and techniques for optimum effectiveness and fficiency.

ipecifically in the ICT project development and management environment, big data helps the project eam to harvest valuable project-related knowledge to facilitate intimate management to ensure the uccess of the project development towards the defined project objectives for strategic project nvestment i.e. effective manipulation of big data brings strategic opportunities to the project success. Iowever, the improper deployments of big data also possess various severe risks and challenges to he project teams. Therefore, an unambiguous set of critical success factors of big data adoption in roject management is vital to well equip the project management team so that they can fully harness he power of big data for strategic project opportunities while mitigate its risks to the project nvestment, development and management.

'his study focuses on the benefits and challenges as well as the critical success factors of big data in acilitating the ICT project management. The study is conducted based on literature review from the nline sources from the similar researches and forum related on the mentioned aspects.

Acknowledgement

Foremost, I would like to express my sincere gratitude to my supervisor, Assoc. Prof. Dr. Roshayu Binti Mohamad for her continuous support in my Master study with her excellent guidance, patience, motivation and knowledge. Her guidance and immerse knowledge motivated me throughout the process of research and writing of this paper. She is always supportive to me with her encouragement and insightful feedback to lead my working towards the right direction for this paper.

Besides, I would like to thanks to my fellow programme-mates in AeU for all the discussions and motivations throughout the process of completing this paper.

Last but not least, I would like to thanks my family members for all their incredible support always.

Table of Contents

Topics **Page**

Table of Contents

Table of Contents

Topics **Page**

CHAPTER 1: INTRODUCTION

1.1 The Research Topic

In a recent study conducted by IBM, it is reported that there are about 2.5 quintillion (10^{18}) bytes of data has been created daily in the recent two years while the estimated growth of data traffic expected by the year 2020 is approximately 50 times higher than now (Duijvestijn & Stahl, n.d.). Technically, big data handles this mass volume of information sourced from heterogeneous origins via sophisticated tools at rapid pace to serve users' unique requests. This data management technology makes the values of data assets really BIG to all the industries and individuals. And, hence it introduced the term "big data."

Generally, the term "big data" is derived from the data-related phenomenon on its rapid expansion with advanced data exploration and exploitation tools. In the context of information management, the big data is well-defined as the information assets with 3 V's, namely high-volume, - velocity and –variety (Gartner, 2001). These data assets could be fully harnessed for enhanced insight and decisions using innovative and sophisticated analytic tools and techniques for unique users' data needs.

Since the year 2012, the big data has started to infuse to our digital realm with a mass involvement of businesses and society (Schroeck et al., 2012). The emerging of big data has changed our way of doing business, communication and our daily life in the data-driven environment today.

Of all the changes in the data-driven environment, the breath-taking growth of data volumes has yielded various opportunities and challenges to the businesses and community. Specifically in the field of ICT project development and management, there are plenty of data, information, reports, updates and feedback produced and communicated throughout the entire project life cycle. These data could be derived from varying project stakeholders or sourced from the project environment internally or externally.

This mass volume of data assets origin from heterogeneous sources are consolidated and managed systematically to provide enriched insights towards the project success. By adopting the big data, it facilitates the project development and management teams in effective data management for improved project development and management optimization, efficiency and effectiveness.

Same as any other industries, the big data also possess some challenges. Often these big data challenges emphasize on the data handling aspects, such as data creation, storage, sharing, analysing, searching and governance. The inappropriate use of data assets could greatly impact the performance, management and development of individual or the entire aspects of the projects.

1.2 The Problem Statement

Inevitably, the availability of enormous data assets from heterogeneous data sources had raised the interests of the businesses to further discover its hidden insights and values. With the unprecedented growth of data volumes, the conventional databases and SQL-based analysis tools are no longer adequate to analyse and spawn the increasing information demands. In addition, the advanced data exploration and exploitation aspects are served as the key strategies to all the industries at all level of management or professionals for strategic operations and optimum management. Hence, an effective data management is essential.

From the ICT project management perspective, every ICT project generates an immerse set of project-related information. The adoption of big data allows a comprehensive set of project-related information to be captured, analysed and reported for optimum project management (Mullaly, 2015) and project team success. The project-related data sources such as resource plans, budgets, schedules, issue lists, change requests, project schedules and many more are required to be systematically created, collected, gathered, quantified, qualified, planned, monitored and reported to facilitate all the project stakeholders across various key knowledge areas in ICT project management.

Commonly, the success of a project relies to the effectiveness of the project management while the effectiveness of project management is associated with the optimum use of data assets. However, the growing data volumes and the demanding user requests have made the data manipulation and management more challenging. Therefore, the big data with advanced analytic tools is superbly in-time to offer a strategic data management to the current needs.

In the project management perspective, the big data supports the project planning, execution, risk management, quality assurance and integration activities with enhanced project quality and improved projects teams' productivity with valuable project-related insights. Often, the adoption of big data helps the project management teams to discover the hidden project-related insights for project effectiveness and success (Keshavarz & Candidate, 2014).

However, how does the big data contribute or impact to the project management work? What are the critical success factors that shall be noted for big data effectiveness in supporting project management?

Despite the challenges of harnessing the full power of data for perfect project insights especially during project life cycle with plenty of uncertainties and the involvement of project stakeholders from multiple disciplines, it is significant to study the benefits and challenges as well as the critical success factors of the big data in project management for optimum project performance and management.

1.3 Objectives

This paper demonstrates the extent and the outcomes of the big data deployment in the ICT project management processes. The aim for this paper is to examine the benefits, challenges and critical success factors of big data in the field of ICT project management.

In considering the growing power and values of data analytic, data flows, computation and storage for project-related decisions, it is essential to identify the opportunities and challenges of big data in the project management so that the project success can be more committed. Besides, the project management team could essentially weigh the benefits of big data for their investment and management decisions timely to ensure the defined project objectives can be achieved within the defined project's triple constrains, namely project scope, time and costs with agreed quality.

Besides, this study also examines the critical success factors of big data in the context of project management to lead the project team to effectively deploy and align the big data application to the project success while enhancing the project team development.

1.4 Research Questions

Practically, the real values of the big data do not merely restricted to its 3 V's characteristics. When big data is deployed in the project management environment, it is important to identify its uses and benefits while determining its potential challenges to for the project management team's attention. Furthermore, the recognition of the critical success factors of big data in project management could effectively help the project team to fully utilize the big data to optimize the data management decisions and analytics for enhanced project planning and quality assurance while reducing the project risk for a guaranteed project success.

Therefore, the above mentioned perspectives lead to the presentation of this report by examining the contributions and impacts of big data as well as its critical success factors in the area of project management.

The paper is guided by the following research questions: -

1. What are the benefits and challenges of big data in supporting project management?
2. What are the critical success factors of big data in the field of project management?

1.5 Significance

This study is designed to examine the benefits and challenges of the big data deployment in the field of ICT project management to raise the awareness of the project manager and project stakeholders on its potential opportunities and risks to the success of project and team development.

In addition, it is crucial for the project manager to recognize the critical success factors for better data management in the processes of project management so that the power of big data could be fully harnessed to support the project initiation, planning, execution, monitoring and controlling with a comprehensive closing down for its completeness.

Obviously, big data is important to support not only the current project but also for the future project by providing reliable and accurate data sources with appropriate lessons learnt from the existing or past projects.

Therefore, the big data's innovative data analytic and utilization facilities are good in empowering the project team in smart decision making while enhancing the appropriateness in project communications, optimizing the project costs and resources management and allowing close monitor and control on the project risks and quality for the project.

1.6 Methodology

A literature research is conducted based on the big data deployment in the project management perspectives for a more comprehensive analysis via an interpretive model.

This research is done by literature search in three key stages to produce a comprehensive review on the deployment of big data in project management. Firstly, literature searches have been conducted by studying the adoption of "big data in project management" via various online databases such as ACM Digital Library, ScienceDirect, BigData Society, Google Scholar and other online resources for the relevant cited papers and reports.

Secondly, the relevant sources also collected by attending big data-related forum, named SAS Forum KL 2015 organized by SAS, a big data provider, dated August 2015 at Le Meridien Hotel KL.

Thirdly, the big data theories are related with some case studies for a more comprehensive review on its deployment in the field of project management.

The compiled works from the above sources were used as the result of this paper.

1.7 Dissertation Overview

Generally, this paper is organized as follows.

Chapter 1: Introduction

> This section provides the background knowledge regarding the big data, the problem statement, paper objectives, research questions, significance, research methodology and paper's organization to offer a fundamental introduction about the paper.

Chapter 2: Literature Review

The section provides a literature review on the big data technology and project management. The focus of this section emphasizes on a general context and background information about the big data technology and project management embracing the benefits and challenges of big data, as well as the critical success factors of deploying big data in various industries including the project management perspective.

Section 3: The Application of Big Data in Project Management

This section discusses the findings on the benefits and challenges of incorporating big data in the ICT project management processes and its critical success factors to be noted for the implementation of big data in ICT project management perspective by relating the big data's theories with real cases.

Section 4: Conclusion

This section summarizes and concludes the key findings of the study and highlights the limitation of the current study with the recommendations for future works suggested.

CHAPTER 2: LITERATURE REVIEW

2.1 Overview

With the advancement of technology, the concept of digitization has been accelerated and expanded to almost all aspects of our life. The rapid growth of the digital devices, storage and network have greatly spurred the creation and utilization data sources using sophisticated and innovative tools and techniques for enhanced organizational or individuals' performance and development. Obviously, the hidden values of these data sources have raised the attention and interest of business to further explore and exploit for improved strategic advantages.

According to a survey conducted by Accenture (2014), titled "Accenture Big Success with Big Data Survey", it is reported that 92% of the respondents were satisfied with the outcomes of big data while 94% agreed that the applications of big data met their requirements and expectations. The study also found that big data has been served as a vital tool to deliver in-depth analytics for more accurate and precise costs and time management, product development and optimization while facilitating smart decision makings for businesses.

2.2 The Big Data

2.2.1 The Characteristics of Big Data

2.2.1.1 The Definition of Big Data

As any other terms in the field of ICT, big data is commonly interpreted with varying definition based on the nature of the industries. Nonetheless, most of the definitions were derived from Gartner. Hence, Gartner's definition will be used as a key reference in this report.

In 2001, Gartner asserted a widely accepted definition on big data as data-related phenomena which possess the following characteristics, *"...high-volume, -velocity and –variety..."* to facilitate improved insight, analytics, decisions and process optimization.

The significance of the mentioned 3 V's are further defined as follow.

(1) Volume

The volume of big data is associated to the size of the data sources. Big data embraces an immerse set of data resources to cover the relevant aspects and dimensions of the phenomena. An adequate data size is significant to provide ample sources and supportive information to facilitate comprehensive in-depth operations and decisions enhancement.

(2) Variety

To facilitate big data operations and management, the metadata of big data are collected and derived from a pool of diversified pertinent channels via internal and external sources, through Internet of Things (IoT), web, log files, social media, etc. These data assets sourced from heterogeneous sources were recorded in varying structures regardless structured, semi-structured or unstructured forms like document-oriented, table-oriented, key-value store and graph databases in origins for further correlation and exploitation.

(3) Velocity

The third V, Velocity, defines the speed of delivering the information as per user requests. Basically, the big data processes the users' data requests at real time basis despite involving a high volume of a variety of data sources for quick decisions, updates and prompt actions.

Besides the 3 V's defined by Gartner, Schroeck et al. (2012) also added that the quality and correctness of the data assets are essential to facilitate effective data management. Hence, the fourth V was included to further explain the characteristic of big data.

(4) Veracity

Veracity is referred to the quality of metadata for utmost accuracy and reliability. The metadata used will produce the analytic results that influence further interpretations and decisions. Hence, the big data's content quality complies to the relevant standards, metrics, models and scientific algorithms is a key success factor to its deployment.

2.2.1.2 The Working of Big Data

Typically, the big data platform works closely with an immense scale of data resources assembled from heterogeneous origins pragmatically with extended repositories to facilitate enriched forensic observation and analytic on the hidden values of the data assets.

According to (Klein & Gorton, 2015), the working of a big data system can be described at a horizontal scale as the diagram illustrated in **Figure 2.1** below.

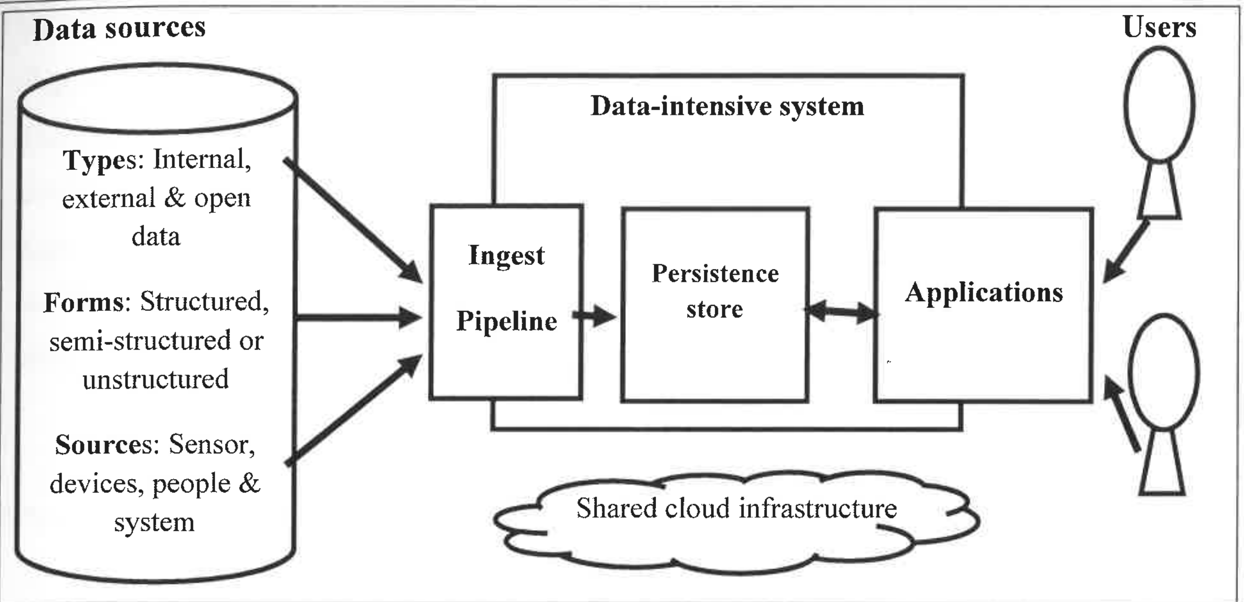
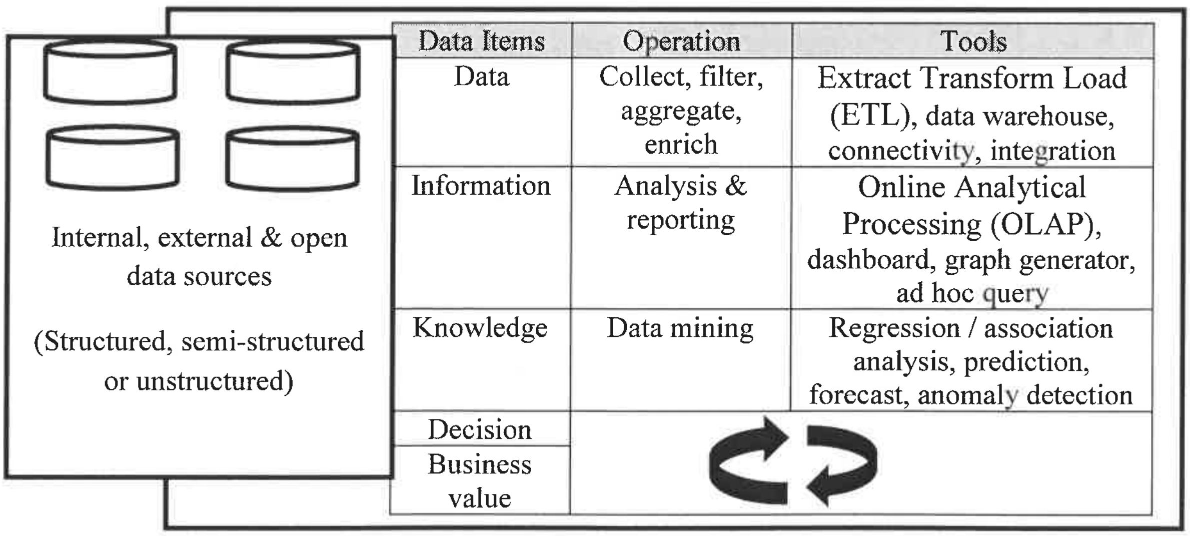


Figure 2.1: The working of a big data system (Klein & Gorton, 2015).

The **Figure 2.1** above shows that the data sourced from internal, external or open data types in the forms of structured, semi-structured or unstructured models are obtained via various sources like sensors, devises, people or system. This data is ingested into a data-intensive system in a shared cloud environment via an ingest pipeline. The data ingested is stored in a pool of centralized repositories to facilitate user requests. During the big data processing, the users' requests are sent through applications where these requests are distributed and executed simultaneously at various computation and storage nodes to speed up the performance (Kromer, 2014) for almost instant responses.



Data Items	Operation	Tools
Data	Collect, filter, aggregate, enrich	Extract Transform Load (ETL), data warehouse, connectivity, integration
Information	Analysis & reporting	Online Analytical Processing (OLAP), dashboard, graph generator, ad hoc query
Knowledge	Data mining	Regression / association analysis, prediction, forecast, anomaly detection
Decision		
Business value		

Figure 2.2: A schematic diagram to illustrate the big data environment in facilitating the data creation and process operations (Olszak & Ziemba, 2007).

To better illustrate the big data environment, Olszak & Ziemba (2007) also drafted another diagram as shown in **Figure 2.2** above with an insight overview on the big data environment. The **Figure 2.2** above illustrates the tools and techniques used in the big data operations with various data items sourced from heterogeneous origins. This data are collected, filtered, aggregated, enriched, analyzed, reported and further mined via the tools such as ETL, data warehouse, OLAP, dashboard and many other innovative and sophisticated techniques for enhanced insights and decisions.

2.2.1.3 The Key Sources of Big Data

Commonly, the data sources in big data are collected or derived from a variety of origins with adequate volume to support imposed accuracy and reliability in analytics. A study from Schroeck (2012) reported that the most significant primary data sources were obtained from the internal organization while the secondary data sources mostly sought from government, business partners, publics and so on either collaboratively or independently. Some examples of these popular data sources used in big data environment are listed as **Table 2.1** below.

Data	Internal data sets	External data sets	Open data	Others
Sample	• Business transactions • Log data • Events data • Social media / social network data • Multimedia data (text and media) • Sensors and actuator data	• Business partners: Supply-chain, customer service, strategies, trading details • Mobile data • Internet query and information • Intellectual properties and research data sets • Risks flags	• Public data • Government regulations, policies, announcement, etc.	• Ego-nets • Channels • Lifestyle • Voice • Risks flags

Table 2.1: A list of data sources for big data technology.

Despite its complexity, the big data supported by a vast volume of a heterogeneous data sources at rapid pace with enhanced veracity to perfectly support intelligence analytic for improved business opportunities and optimization (Choudhury, 2013). The 4 V's of big data helps the individuals and management to precisely plan, estimate, monitor and control the relevant phenomena with quick decisions.

2.2.1.4 The Technology and Infrastructure of Big Data

A report from McKinsey & Company (2011) suggested that some sophisticated infrastructures are mandatory in big data to ensure the efficiency and effectiveness of its operations. Some of these key infrastructures are described below in the **Table 2.2** below.

Key Infrastructure		Functions
1	Information integration	The integrated architecture includes scalable data architecture with unified management to support enriched computing, storage and connectivity. It facilitates instant data accesses and analytics with transparency in decisions and policies setting.
2	Scalable storage infrastructure	Big data is supported by a pool of storage servers and aggregated network to accommodate the rapid data processing and management.
3	Security and governance	Big data embraces a set of data security and governance models with legal, ethical, regulatory policies and the necessary measures to facilitate seamless data exploration and exploitation with improved data quality and intelligence.

Table 2.2: Some key infrastructure of Big data (McKinsey & Company, 2011).

According to McKinsey & Company (2011) and Schroeck et al. (2012), some key tools and techniques are essential to optimize the big data performance. These tools and techniques are described in the **Table 2.3** as shown below.

Key Tools and Techniques		Functions
1	A/B testing tools	A type of statistical hypothesis testing tool uses 2 variants, namely variant A (as control) and variant B (as treatment) in a controlled experiment. Often, it uses the data sourced collected from survey, offline source or other more complex phenomena for effective and accurate analysis.
2	Crowd-sourcing	Big data uses a collection of information origin from a large group of heterogeneous sources via online communities, public, devices to provide quality metadata for further analytic and management.
3	Machine learning	An artificial intelligence used to study the patterns, leaning theories and predicting the potential outcomes from the learned mode.
4	Simulation	The simulation facilitates decisions, collaboration, automated-related algorithms and rules by modeling the real environment for accuracy.
5	Visualization	Big data offers visual imagery such as diagrams, graphs, charts, images and so on for effective information presentation and communication.
6	Predictive modeling	Predictive modeling in big data uses what-if analysis and use cases to offer enriched insights of future state. It is essential for risk assessment, hypothesis testing, forecasting, strategic decisions, etc.

Table 2..3: Some key tools and techniques of Big data highlighted by McKinsey & Company (2011) and Schroeck et al. (2012).

2.2.2 The Benefits of Big Data

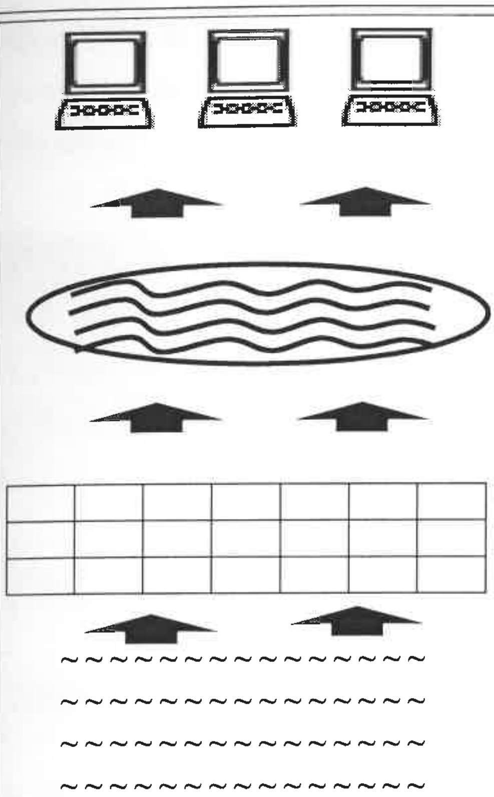
According to a survey commissioned by Glenn Finch, Steven Davison, Christian Kirschniak, Marcio Weikersheimer (2013), it is reported that there are about 63% of the respondents participated in the survey committed that big data create positive return via advanced analytic while 69% of them agreed that big data significantly impact the organizational outcomes.

Schroeck et al. (2012) added that big data does not bring any values until it is used to solve imperative business challenges. Big data enables the data ocean to be fully harnessed and further drilled down via sophisticated integration, filtering and analysis for richer insight of data sources and improved decisions for improved performance in businesses. For instance, a report on the Big Data Consumer Guide from Open Data Center Alliance (Odca, 2012) reported that big data helps to reduce the expenses of hardware, software and application developments via effective data management. Some other common benefits of big data to businesses are including the followings: -

(1) Facilitate data archiving and integration

Big data consolidates the data flooded from heterogeneous origins into centralized repositories despite varying data genres, structures and semantics for dynamic deployments (Keshavarz & Candidate, 2014). Its data integration models assemble the broad data context to essentially facilitates an extensive series of exploitation applications (Baru, Bhandarkar, Nambiar, Poess, & Rabl, 2012). For instance, big data with cloud-based tools enable the users to enjoy the ubiquity access with sophisticated analytic features at security-rich environment.

A study from Johan et al. (2014) reported that big data is integrated via an efficient data utilization schema designed by John Donovan from AT&T (as cited in Johan et al., 2014). The Donovan's data utilization schema integrates the data assets from diverse origins and structures to provide enriched data sources for further exploration and exploitation. Furthermore, the schema also optimizes the big data integration and exploration via advanced mapping, simulation through various types of APIs for users' efficiency as described in the **Figure 2.3** below.



Application Program Interface (API)
API is a software intermediary to bridge the users and big data connection and enable users to explore a large volume of data for their unique data requirement.

Data lake
Data lake offers a centralized repository to integrate massive volume of data sources from any origins at diverse structures for further analytic and exploitation.

Data grid
The data grids systematical organizes the data sources in matrix forms to facilitate the user access, alternation, transfer and exploration.

Data stream
Data stream involves data transmission process into the repositories to offer the root of data sources for data processing and flowing.

Figure 2.3: An AT&T framework of data integration and utilization model designed by Donovan (as cited in Johan et al., 2014).

Jewell et al. added that big data commonly includes searching, discovery, indexing and navigation systems for optimum data integration and management. For instance, IBM offers tools like InfoSphere with indexing capability for flexible data integration and utilization. The IBM's InfoSphere indexing helps users to accelerate the process of searching, navigating, discovery and exploration via position-based index approach while index replication helps minimize redundancy and index distribution facilitates distributed processing.

(2) Support data mining and analytic

According to a study reported by Russom (2013), it is found that about 61% of the respondents from the survey committed that big data helps to improve the data analytic on the user requirements, nature and relationships between processes, people, patterns for insight analysis, fraud detection and anomalies discovery on the root causes, customer behaviors, business trends and changes, etc. to gain advanced strategic advantages.

In view of the unique users' information needs, big data offers the analytical tools such as A/B testing, visualization, simulation, etc. for advanced data management. E.g.: Apache Hadoop offers an affordable storage for massive volume of variety data to be managed with satisfactory velocity.

Rusitschka & Ramirez (2014) added that the process of data mining and analytic are proceed according to well-recognized sequence to ensure its accuracy and effectiveness as illustrated in the **Table 2.4** below and described as followed.

Analytic processes	Tools	Outcomes
1:Descriptive	Standard algorithms & assumptions	Summary, report & dashboard
2: Diagnostic	Data mining and reasoning models	Current status / situation
3: Predictive	Data mining, machine learning, neural network, reasoning models and optimization	Future status / situation
4: Prescriptive	Real time modeling	Possible actions & outcomes with evaluation

Table 2.4: The big data analytical process (Rusitschka & Ramirez, 2014)

Step 1: Often, the data analytic process is started with a descriptive analytic to provide an overview on the data content. In the descriptive analytic, data assets are explored, aggregated, filtered, summarized or visualized with a set of standard algorithms and necessary assumptions to reflect the real conditions and situations. Its outcomes are presented in the forms of conventional reporting or innovative dashboards forms for greater visualization.

Step 2: The descriptive insights are then further analyzed using diagnostic analytics and semantic technologies with more accurate reasoning models for enhanced investigation and enriched decision support. Diagnostic analytic supports individual actions processing, tracking and diagnostic via data mining and reasoning models for accuracy to alert the users on the current situation for necessary strategies or measures instantly.

Step 3: With the detailed diagnostic insights collected, the data science practitioners dynamically cluster, test, explore and exploit the data sources using know-how analytic for prediction in a cost-effective way. The predictive analytic uses data mining and machine learning capabilities to perform correlation and probability computation from the vast historical data for enhanced reliability (Odca, 2012).

Step 4: Lastly, the prescriptive analytics suggests the possible actions based on the systematic forecast and comprehensive evaluations on the possible outcomes obtained from predictive analytic using simulation and optimization models. The recommended set of best solutions is then prescriptively analyzed for their appropriateness and potential outcomes for enhanced decision automation.

Therefore, it is clear that an effective data mining and analytic are much depending on the data integration. To facilitate an effective classification, indexing and information manipulation in an integrated environment, natural language processing, search techniques as well as various algorithms and models are necessary for greater efficiency.

(3) Aids business optimization and decision making

The large stream of data from a variety of sources in real time is challenging yet essential to provide novel insight to help the organizations to rules their businesses for the optimum customer-centric performance.

Provost & Fawcett (2013) urged that "*the best decisions shall be made based on rata analytic rather than purely on intuition*". The emerging of data democracy has empowered the decision-makers with sets of comprehensive analytic results (Labrinidis & Jagadish, 2012) and supported by outcome interpretation capabilities based on varying assumptions, parameters, data resources and evaluations for more precise decisions towards the define business or project objectives. With big data capability, businesses can empower more decision makers instead of sorely depend on a few elites in business product development and management.

2.2.3 The Challenges of Big Data

Notwithstanding the big data technology continues to emerge and evolve, challenges continue to be engendered (Keshavarz & Candidate, 2014) which can deter the new users to the big data pools. Some of these commonly challenges found are including the followings: -

(1) Quality of data

In the recent forum conducted by SAS, one of the speaker, Prof. Francisco N. de los (Reyer, 2015), an analytic advisor from Thakral One, claimed that the data quality is the utmost significant challenges to the big data users. According to Prof. Francisco, the context of data quality encompasses the accessibility, accuracy, appropriateness, ease of use, completeness, interpretability, relevant, reputation, security and value-added.

The massive data storage and processing frameworks with various tools increase the barriers in assuring the data quality which could result severe risks and governance issues in security and privacy aspects (Yew & Architect, 2015). Often, the key factors of poor big data quality are caused by inefficient work flow, disjointed data sources and inconsistent standardization (Ramanathan, n.d.). Furthermore, the assorted patterns in workflow management among various data experts always provoke separation in data management.

Besides, Labrinidis & Jagadish (2012) claimed that the nature of big data itself already nuisanced the data quality. Todays, most users at any age, professioanls or industries are demanding for real time processing instead of batch processing. Hence, the rapid acquisition and qurying rates have tighten response time limits. In addition, Koronios, Gao, & Selle (2014) also added that people, processes and technology are the key sources that drive the quality of data due to their challenges in data security, privacy and governance aspects.

(2) Data security, privacy and governance issues

The big data is complex, in its physical realm and operational methods, and for sure not to forget its data quality, security and privacy models as well as its governance frameworks (Berman, 2013). Besides, Russom (2013) also reported that the big data security, privacy and governance are the key factors that impede the success of big data management in the respondents' organizations.

In addition, the emergent of big data in cloud technology has further attracted the risks of privacy and security break (Cuzzocrea, 2014). The sensitive and confidential data content stored in distributed data repositories for ubiquity accesses via cloud nodes and vehicular networks open doors to security breaches.

Apart from that, Carter (2011) also found that the real time processing with overloading amount of alerts tail closely could be annoying and subsequently stimulate users' ignorance or simply clicked away. As the consequences, the real time anomalies cannot be responded promptly for efficiency.

Besides, the massive data sources with anonymized content create tension to future analytics in quantifying its originality or replication (Daries et al., n.d.), especially in the open environment via network. More severely, the growth of data volumes had accelerated from a single silos into a proliferate of siloed big data repositories in 21 century (Mcheick, Grant, & Mohammad, 2014). This silos integration makes big data governance more complex.

(3) The complexity of big data technology
Labrinidis & Jagadish (2012) claimed that businesses have started to embrace additional system to for varying purposes like risk assessments, operational logs and social media analytics. The integration of these separated existing systems and big data management tools could be challenging and complex.

The issues of connectivity, compatibility, accessibility and governance between the big data platform with the existing systems require double endeavours from all the parties involved especially from the programming teams in amalgamating the relevant modules for integration and providing comprehensive testing for data quality assurance and governance.

For instance, a duplicate set of data sources at different systems enable backup for contingency plan. Obviously, this is a good practice for smooth operation in data management and processing. However, the data duplication makes the operation more complex in the big data environment (Rajan, Ginkel, & Sundaresan, 2012). The use of redundant data sources with overlapping and redundant set of interdependencies caused complexity and confusion in maintenance and monitoring (Odca, 2012).

In addition, the acceleration of data growth and values had challenged the normal database, data management and processing (Koronios et al., 2014). The massive data volumes from heterogeneous data repositories and structures have gradually shift the normal database system into Hadoop with Hadoop Distributed File System (HDSF) features and MapReduce with parallel processing framework and yet the changing requirements from the user groups are keep changing with more unique and innovative specifications. Therefore, Johan et al. (2014) believed that power of data assets can be fully harnessed via the right infrastructure, right talents and organizational alignment.