Addis Ababa University
አዲስ፡አበባ፡ዩኒቨርሲቲ
SEEK WISDOM, ELEVATE YOUR INTELLECT AND SERVE HUMANITY !

**Addis Ababa University**

**School of Information Science**

# MATURITY ASSESSMENT OF BIG DATA ANALYTICS IN THE ETHIOPIAN HEALTH SECTOR

**By Hilina Terefe**

A THESIS SUBMITTED TO THE SCHOOL OF INFORMATION SCIENCE AND SYSTEMS

FOR A PARTIAL FULFILMENT OF THE MASTER OF SCIENCE DEGREE IN INFORMATION SCIENCE AND SYSTEM (INFORMATION SYSTEMS TRACK)

**December 2021**

**Addis Ababa University**

**School of Information Science**

Hilina Terefe

Advisor: Melkamu Beyene (PhD)

This is to certify that the thesis prepared by Hilina Terefe Mekete, titled *Maturity Assessment of Big Data Analytics in the Ethiopian health sector* and submitted in partial fulfillment of the requirements for the Degree of Master of Science in Information Systems compiles with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by Examining Committee:

| *Name* | | *Signature* | *Date* |
|---|---|---|---|
| *Advisor:* | Melkamu Beyene (PhD) | | |
| *Examiner:* | Getachew H/Mariam (PHD) | | |
| *Examiner:* | Million Meshesha (PHD) | | |

# ABSTRACT

*It has been experienced and showed in different researches that big data applications have assisted in reducing the cost of healthcare. In order to understand this an appropriate maturity assessment model need to be deployed to analyze the current scenario. In this research work the following research questions will be answered: What are the key pillars considered to select Big Data Maturity Model (BDMM)? Which Big Data Maturity Model is appropriate to evaluate Ethiopian health sector experience in Big Data Analytics? What is the big data maturity level of Ethiopian health sector scenario? In general the objective of this research was maturity assessment of big data analytics in the Ethiopian health sector. For this purpose, a conceptual tool known as the big data maturity model has been selected using defined criteria that encompasses model structure completeness, model development and evaluation quality, ease of application and big data value creation to best fit the Ethiopian health sector. The researcher applied deductive–exploratory research purpose with hermeneutics philosophical approach to select the best fit maturity model called TDWI (The Data Warehouse Institute) Big data maturity model. The researchers applied qualitative research to study the status of big data analytics application based on the five pillars i.e., Organization, Infrastructure, Data Management, Analytics and Governance as a reference to define the maturity level. For this research Ministry of Innovation and Technology (MINT) and Ministry of Health (MOH) are the main sources of the data. The population is selected using Purposive Sampling Technique. In this research work data is being collected directly from the identified and selected sample population. Correspondingly, interview is used as a direct data collection method. The analysis result showed that big data analytics application in the Ethiopian health sector is in its Pre- Adoption stage. One of the characteristics of an organization in a pre-adoption stage is the readiness to start doing on its responsibilities regarding big data analytics in which the health sector in Ethiopia also shows similar characteristics. In this research work there is a limited time and capacity to reach all main stakeholders within the health sector then the researcher is limited to contact the main and leading stakeholders within the health sector.*



**Key Words:** Big Data Analytics, Big Data Maturity Model, TDWI BDMM, Health Sector

## Acknowledgment

I am over helmed in all humbleness and gratefulness to acknowledge my depth to all those who have helped me to put these ideas, well above the level of simplicity and into something concrete.

I would like to express my special thanks of gratitude to my advisor Dr. Melkamu Beyene. I owe my deepest gratitude to Ato Gebrie Alamenie Technology and Innovation Institute, Computing and Analytics Directorate director for his encouragement, feedback, and continuous support.

I would like to acknowledge the Ministry of Innovation and Technology, Technology and Innovation Institute, and Ministry of Health to help me get all the necessary data.

I would like to thank my husband and parents who helped me a lot in guiding me from time to time in making this project, despite their busy schedules, they gave me different ideas in making this project valuable.  Thank You.

Contents

## List of Acronyms

| | |
|---|---|
| **BDMM** | Big Data Maturity Model |
| **BI** | Business Intelligence |
| **CAD** | Computing and Analytics Directorate |
| **EFDA** | Ethiopian Food and Drug Administration |
| **EPHI** | Ethiopian Public Health Institute |
| **GDP** | Growth Domestic Product |
| **GoE** | Government of Ethiopia |
| **MINT** | Ministry of Innovation and Technology |
| **MOH** | Ministry of Health |
| **POC** | Proof of Concept |
| **PSA** | Pharmaceutical Supply Agency |
| **TDWI** | The Data Warehousing Institute |
| **TechIN** | Technology and Innovation Institute |
| **USAID** | United Sates Aid |

# Chapter 1: Introduction

## 1.1.Background

Today, organizations are accumulating increasing amounts of big data. This tremendous amount of collected data could be customer, supplier, or operational data. Companies realized that this data and advanced analysis could provide an important strategic and competitive advantage (Manyika, 2011; Halper & Krishnan, 2013). However, they are accumulating and collecting data more than the level of the capacity they can manage. To manage this more organized and advanced infrastructure, data management, analytics, governance, and organizational process should be introduced (Halper & Krishnan 2013). In general, these initiatives collectively referred to as Big Data.

Big data is an all-inclusive approach to managing, processing, and analyzing huge volumes of data for providing insights, so that we can have a better advantage when making decisions (Wamba, Akter, Edwards, Chopin, Gnanzou, et al., 2015).

Internationally, the application and concept of big data is well known in different areas. The purpose of this work is to study the current applications, challenges and in general maturity of Big Data in Ethiopian's health care. Big Data in the health sector is characterized by the following features: heterogeneity, timeliness, ownership, and privacy.

Big Data in the health care sector has its own features, such as heterogeneity, incompleteness, timeliness, longevity, privacy, and ownership. These characteristics within the health sector bring a lot of challenges for the following areas: data storage, mining, and sharing to advance the sector.

Within the health sector the application of big data analytics could bring improved treatment with lower cost for the patient. Moreover government, research institutions and hospitals could also be advantageous from the application of big data analytics in the health sector.

According to history the healthcare sector is such that it generates huge volumes of data, which is propelled by keeping meticulous patient-health records (Raghupathi, 2014). In the past, records were stored as hard copies. Now, records are digitized, leading to large volumes of data being generated. In the era of big data, the velocity of generating data is very high such that the U.S healthcare industry alone attained 150 exabytes as in 2011, which could soon reach yottabytes (Dembosky, 2012). Today, approximately 30% of the world's data volume is being generated by the health care industry. By 2025, the compound annual growth rate of data for health care will reach 36%. (Capital Markets Episode 1, the health care data explosion) Although application of big data in Africa is still in its infancy, recent reports have shown that success has been achieved using digital surveillance to track epidemics (Nsoesie, Kluberg, Mekaru, Majumder, Khan, Hay, Brownstein, et al., 2015).

This research work attempts to study the maturity of the sector by using a selected big data maturity model as a reference. Regarding with big data management in the health care sector, Ministry of Health (MOH) and Technology and Innovation Institute under (TechIN) Ministry of Innovation and Technology (MINT) plays the greatest role.

The Ministry of Health shall have Powers and Duties to formulate the country's health sector development program, follow up and evaluate the implementation of same, support the expansion of health services coverage, follow up and coordinate the implementation of health programs financed by foreign assistance and loans, direct, coordinate and follow up implementation of the country's health information system, devise and follow up the implementation of strategies for the prevention of epidemic and communicable diseases, follow up and coordinate the implementation of national nutrition strategies, take preventive measures in the events of emergency situations that threaten public health, and coordinate measures to be taken by other bodies, ensure adequate supply and proper utilization of essential drugs and medical equipment in the country, prepare the country's health services coverage map, provide support for the expansion of health infrastructure, supervise the administration of federal hospitals, collaborate with the appropriate bodies in providing quality and relevant health professionals' trainings within the country, provide appropriate support to promote research activities intended to provide solutions for the country's health problems and for improving health service delivery, expand health education through various appropriate means, ensure the proper execution of food, medicine and health care regulatory functions (FMoH Page). In the other side TecIN is responsible for developing the innovation and technology research infrastructures, organizing and linking innovation and technology researches scattered across different, sectors to the economic and social significance, supporting manufacturers and service providers to make them internationally competitive through providing innovative and technological solutions that comes out from research and development activities through in-house problem-solving technology development and prototyping, collecting, organizing, and analyzing data that may be used by a variety of high-tech technological innovations, researching and distributing information by adding value for environmental and technological development, facilitating, implementing and improving research and development of innovations and technology in identifying sectoral development actors of the country through cutting-edge technology (Technology and Innovation Institute Organizational Overview).

The two ministers primarily work on the big data analytics in the health sector. Both the organizations have their own dedicated department working on the sector. In such time of pandemic disease like COVID 19 is spreading the benefit of big data analytics is priceless in controlling its damage and to act. Currently Ethiopian Public Health Institute (EPHI) is dedicated to managing the big data related to the pandemic disease. In addition, FMOH is working with USAID to deploy different technological solutions, funds to support the big data analytics specifically during the COVID pandemic could be mentioned. In general MINT take the initiation and establish the computing and analytics directorate to dedicatedly work on big data analytics.

In order to get advantage of the sector to the fullest and understand its maturity in the health care there should be a defined and clear framework. A selected big data maturity model assist this work to identify the maturity of big data analytics in the health sector in Ethiopia.

## 1.2 Motivation

Big data analytics have a positive impact in so many areas like in combating crime, business execution, finance, Global Positioning System (GPS), commerce, travel, urban informatics, meteorology, genomics, complex physics simulations, biology, environmental research, and health care (Chen, Mao, Liu, et al., 2014). Health care data are one of the driving forces of Big Data. As the speed of data generation is increased the volume of data is also increased in a similar manner. For example, The Human Genome Project completed in 2003 showed that, one single genome in human DNA occupies 100–150 gigabytes (Marx, 2013; O'Driscoll, Daugelaite, & Sleator, 2013). In terms of data size, Big Data in health care exceeded 150 exabytes after 2011 (Wang, Kung, Ting, Byrd, et al., 2015), and a study showed that data size in health care is estimated to be around 40 ZB in 2020, about 50 times the 2009 figure of 0.8 ZB (O et al., 2013).

There is huge and complex big data within the health care and medical sector. Considering this, applying traditional software and hardware will make the analysis so difficult (Kankanhalli, Hahn, Tan and Gao, et al., 2016), (Raghupathi W & Raghupathi V. 2014). Big data analytics includes integration of heterogeneous data, data quality control, analysis, modeling, interpretation, and validation (Wu, Cheng, Kaddi, Venugopalan, Hoffman, Wang, et al., 2017). The application of advanced big data analytics provides knowledge from the existing big amount of data which helps for strategic decision.

Specifically, the application of big data analytics in the health sector enables analysis of large datasets from thousands of patients' data, identifying cluster and correlation between datasets as well as developing models using data mining techniques.

In this research the health sector is being considered for the case study to examine maturity of big data analytics in Ethiopia. Big data analytics is contributing a lot for the health sector worldwide, in comparing this with our country experience the researcher is motivated to assess the maturity of the sector and its challenges in the health sector in Ethiopia. The world is entering the new era of the fourth industrial revolution, Ethiopia should not be a follower but an active player, unlike our previous experiences.

The researcher is motivated to contribute to the health care service enhancement through the understanding of big data analytics application and by identifying the sector's bottleneck to the application of analytics in healthcare.

## 1.3. Statement of the Problem

Although businesses are the biggest player in the use and application of big data, healthcare sector has been active in the application of big data in recent times. In many cases, big data applications have assisted in reducing the cost of healthcare (Raghupathi W & Raghupathi V. 2014), helped in the diagnosis and in the prescription of suitable treatment pathways to patients, ensured healthcare systems monitoring and augmented accountability.

Traditionally, doctors use their experience and judgment to diagnose and to make clinical decisions. However, in this era of big data, decisions made by health professionals are edging towards evidence-based medicine (Costa, 2009). Hence, patient information and data generated by other health professionals especially laboratories and radiology are transmitted to the physician in a timely manner. This further increases data that are available to make better-informed decisions for patient's care.

The application of big data analytics need assessment to understand its maturity. Big data applications have assisted in reducing the cost of healthcare (Kamadjeu, Tapang, Moluh, et al., 2005), helped in the diagnosis and in the prescription of suitable treatment pathways to patients, ensured healthcare systems monitoring and augmented accountability. In order to get the benefits of big data analytics within the health sector an initial work of understanding the existing scenario is crucial. This might be done using different tools and techniques.

In this work an assessment is done using a conceptual framework to address the following main research questions:

- What are the key pillars considered to select Big Data Maturity Model (BDMM)?
- Which big data maturity model is appropriate to evaluate Ethiopian health sector experience in big data analytics?
- What is the bid data maturity level of Ethiopian health sector?

## 1.4. Objective of the Research

### 1.4.1. General Objective

The general objective of this research is to identify the maturity level of big data analytics at Ethiopian health sector.

### 1.4.2. Specific Objectives

In order to achieve the general objective, the following specific objectives are identified:

- Prepare criteria as a pillar to select appropriate BDMM
- Assess appropriate big data maturity model for a reference that best fit Ethiopian health sector using the designed criteria
- Discuss the Ethiopian health sector characteristics according to the selected model's sub perspectives
- Identify challenges and factors that hinder the application of big data analytics in the health sector in Ethiopia

## 1.5. Scope and Limitation of the study

The scope of this research is to assess the application of big data analytics maturity in Ethiopia. For the maturity assessment a model has been selected that best fit for Ethiopian health sector by considering designed criteria. Following that the sector is being studied and analyzed according to the model's content and perspective.

As of there is a limited time and capacity to reach all main stakeholders within the health sector the researcher is limited to contact the main and leading stakeholders within the health sector. In addition, as the researcher only able to get highly experienced personnel within the sector to get better information the conclusion is totally dependent on the respondent's view.

## 1.6. Significance of the Research

The main benefit of this research is to study the readiness/ maturity of big data analytics application in the health sector in Ethiopia using a best fit big data maturity model. After learning the state of the sector application in Ethiopia the researcher will be able to define the current scenario. In line with this the research will list drawbacks in the current big data analytics application in the health sector which will be useful for the future as an input to design a better framework to implement big data analytics.

As big data analytics has an impetus effect on the growth of the health sector, studying the current state and identifying factors that hinder its development is a very crucial point. In order to develop the sector, it is very important to identify the sector's maturity. Then this research will assist the growth of the sector by understanding the maturity state and identifying drawbacks for designing better solutions.

## 1.7. Organization of the Thesis

The remaining part of the thesis is organized as follows. The second chapter deals with the literature review, related work, and conceptual framework. The literature review section tries to give a brief understanding of big data analytics, its application in the health sector and automation and the health sector in Ethiopia. In the same chapter, under the related work section, recent and related works for the topic has been discussed with critical review. Finally, the second chapter deals about big data maturity models for the health sector and identified the most compatible maturity model for Big Data Analytics application maturity assessment for the health sector in Ethiopia according to quality, ease of use, extensiveness, and business intelligence integration. The third chapter discussed the methodology for the whole research work on data collection method, its source and sample population and analytics techniques. Following the above chapter, the next chapter, the fourth chapter, discussed about Ethiopia’s health sector and big data analytics application according to the selected model. Finally, the fifth chapter discuss the research work result and findings. Lastly, the sixth chapter concludes the thesis by indicating conclusion and recommendations.

# Chapter 2: Literature Review and Related Works

### 2.1.1. Big Data Analytics

The concept of "big data" is not new; however, the way it is defined is constantly changing. A data mostly acquiring an advanced software and hardware technologies for effective and efficient visualization, analysis, and storage, this could be mentioned as one of the definitions for big data. The health sector generate data rapidly, which has original behavior of how the three Vs of data basically determined like: variety, velocity (speed of generation of data) and volume. Such data is dispersed throughout different health sector systems, health sector researchers, health insurers, government entities, and so forth. Furthermore, each of these data repositories is siloed and inherently incapable of providing a platform for global data transparency. To add to the three Vs, the veracity of healthcare data is also critical for its meaningful use towards developing translational research (Big Data Analytics in Health Care, Hindawi Publishing Corporation)

According to 15econ et al., big data is an all-inclusive approach to managing, processing, and analyzing huge volumes of data for providing insights, so that we can have a better advantage when making decisions.

Big data analytics is a way of using advanced and technological analysis techniques over huge data, unique and distinct data which involves unorganized, semi-organized and organized data from various sources and different size. The former and traditional databases did not possess the capacity for storing, processing and managing big data sates whose size or nature is further away if its capability. The most common big data characteristics are the following: high volume, high velocity, or high variety. Some of the areas contributes for the data complexity and new sources of data could be the internet of things (IOT), social, artificial intelligence (AI) and mobile. The regular source of big data is: networks, log files, audio/video, sensor devices, web and social media, transactional applications. Most of the data is produced at a huge scale and in real time.

Big data and big data analytics create the environment for researchers, business users and analysts to build well and speedy decisions using data that was previously unusable or inaccessible (and couldn't manage by traditional technique). Advanced analytical techniques such as text analytics, predictive analytics, data mining, statistics, machine learning and natural language processing could apply in business to gain new insights or perspective from prior unexploited data sources independently or together with the current big data (Ashwin, Raghuram, Reza, Soroushmehr, Fatemeh, Daniel, Kayvan, et al., 2006). One such under-developed regions of the world is Africa, and technically, the continent is adjudged to be an unhealthy continent due to prevalence of epidemics and inadequate infrastructure to sustain healthcare. While there is apparent inadequate infrastructure to sustain good healthcare in many parts of Africa, there is evolution of use of mobile technologies such as mobile phones. For instance, users of mobile phones subscription in Africa as of February (2016) have reached about 700 million. Such huge mobile phones subscription also means huge generation of big data. Despite slow growth in the application of big data in Africa, recent evidence indicates the usefulness of mobile phone data and social media such as Facebook in tracking and eradicating epidemics, though with some challenges. Due to lack of awareness of big data usefulness within Africa it is important to report its roles and potentials (Ashrafi, Lori, Jean, et al., 2014).

### 2.1.2. Big Data Analytics and the Health sector

A review article by Hindawi Publishing Corporation (Ashwin, Raghuram, Reza, Fatemeh, Daniel, Kayvan, et al., 2006) discussed three areas that the concept of big data analytics currently being applied. The first is image processing. Medical images are an important source of data frequently used for diagnosis, therapy assessment and planning. Computed tomography (CT), magnetic resonance imaging (MRI), X-ray, molecular imaging, ultrasound, photoacoustic imaging, fluoroscopy, positron emission tomography-computed tomography (PET-CT), and mammography are some of the examples of imaging techniques that are well established within clinical settings. Medical image data can range anywhere from a few megabytes for a single study (e.g., histology images) to hundreds of megabytes per study (e.g., thin-slice CT studies comprising up to 2500+ scans per study. It also demands fast and accurate algorithms if any decision assisting automation were to be performed using the data.

The encoded concept is Signal Processing. Like medical images, medical signals also pose volume and velocity obstacles especially during continuous, high-resolution acquisition and storage from a multitude of monitors connected to each patient. However, in addition to the data size issues, physiological signals also pose complexity of a spatiotemporal nature. Analysis of physiological signals is often more meaningful when presented along with situational context awareness which needs to be embedded into the development of continuous monitoring and predictive systems to ensure its effectiveness and robustness.

The last but not the least concept is Genomics. The cost to sequence the human genome (encompassing 30,000 to 35,000 genes) is rapidly decreasing with the development of high-throughput sequencing technology. With implications for current public health policies and delivery of care, analyzing genome-scale data for developing actionable recommendations in a timely manner is a significant challenge to the field of computational biology. Cost and time to deliver recommendations are crucial in a clinical setting.

Initiatives tackling this complex problem include tracking of 100,000 subjects over 20 to 30 years using the predictive, preventive, participatory, and personalized health, referred to as P4, medicine paradigm as well as an integrative personal omics profile. The P4 initiative is using a system approach for (i) analyzing genome-scale datasets to determine disease states, (ii) moving towards blood based diagnostic tools for continuous monitoring of a subject, (iii) exploring new approaches to drug target discovery, developing tools to deal with big data challenges of capturing, validating, storing, mining, integrating, and finally (iv) modeling data for each individual (Ashwin, Raghuram, Reza, Fatemeh, Daniel, Kayvan, et al., 2006). The integrative personal omics profile (iPOP) combines physiological monitoring and multiple high-throughput methods for genome sequencing to generate a detailed health and disease states of a subject. Ultimately, realizing actionable recommendations at the clinical level remains a grand challenge for this field. Utilizing such high-density data for exploration, discovery, and clinical translation demands novel big data approaches and analytics.

### 2.1.3. Automation and the Health Sector in Ethiopia

Big data is having a positive impact in almost every sphere of life, such as in military intelligence, space science, aviation, banking, and health. Big data is a growing force in healthcare. Even though healthcare systems in the developed world are recording some breakthroughs due to the application of big data, it is important to research the impact of big data in developing regions of the world, such as Africa.

In healthcare, the rate of generation of electronic health records in the present dispensation is high and such data are diverse in datatype and complexity to the extent that they cannot be processed with traditional software/hardware or with conventional data management tools and methods. The totality of this diverse data that is generated make up the "big data" in healthcare. Such data include, clinical data, doctor's comments and prescriptions, medical imaging, laboratory test results, pharmaceutical records, health insurance reports, administrative reports, patient historical data, posts on social media platforms such as, Facebook and Twitter, blogs, and other information that are not directly connected to patients, such as, news in health magazines and relevant publications in medical journals (Raghupathi W & Raghupathi V. 2014).

Just like in the business world, availability of big data in the healthcare sector, presents an opportunity to discover relationships, patterns, and trends within the data. Information derived from such sources has the capability of improving healthcare and lowering cost of health care. Therefore, what big data analytics can do in health care is, to utilize the benefits of large volumes of diverse data to obtain insights so that healthcare professionals and administrators alike, can make life-changing decisions (Raghupathi W & Raghupathi V. 2014).

The African continent has the highest mortality rate the world, and as at today, it is the only region of the world where deaths from communicable diseases exceeds the deaths from chronic diseases (Deaton & Tortora, 2016). For instance, the continent population is disabled with epidemic of diseases such as Malaria, HIV/AIDS, Polio, and in recent times the Ebola virus (Aikins & Marks, 2007). In (2005), about 18.8% of the 58.03 million global deaths are from Africa, and about 64% of these deaths were as a result of diseases (Kirigia & Barry, 2008).

The Government of Ethiopia (GOE) is working to strengthen the healthcare system to align it with the Millennium Development Goals. Ethiopia has a large, predominantly rural and subsistence agriculture population with poor access to safe water, housing, sanitation, food, and health service. The government has made significant investments in the public health sector that have led to improvements in health outcomes. Nevertheless, communicable diseases like HIV/AIDS, TB, malaria, respiratory infection, and diarrhea remain a serious challenge in Ethiopia. High fertility rates, and low contraceptive prevalence continue to drive a rapidly increasing population in Ethiopia. With a growing middle class, the GOE is facing an increase in non-infectious diseases such as cancer, diabetes, heart diseases, hepatitis B&C and high blood pressure. Mental health and eye problems are also becoming major issues in Ethiopia.

Under the second Growth and Transformation Plan (GTP II) and Health System Transformation Plan, the Ministry of Health (MOH) is implementing changes to various aspects of the healthcare system. For better management, the government has increasingly decentralized management of its public health system to the Regional Health Bureau levels. The MOH is also committed to reforming agencies such as Ethiopian Food and Drug Administration (EFDA) and the Pharmaceutical Supply Agency (PSA).

### 2.2.1 Prospects of Big Data Analytics in Africa Health Care System

The purpose of this section is to review a related work to understand the sector and how other research is contributing on this specific topic. Accordingly, the following summarized article has the purpose of reporting the prospects of big data in African healthcare. It is believed that this review will provide an update on the application of big data in the health sector in Africa continent. Healthcare systems in Africa are, in relative terms, behind the rest of the world. Platforms and technologies used to amass big data such as the Internet and mobile phones are already in use in Africa, thereby making big data applications to be emerging. One such under-developed regions of the world is Africa, and technically, the continent is adjudged to be an unhealthy continent due to prevalence of epidemics and inadequate infrastructure to sustain healthcare. While there is apparent inadequate infrastructure to sustain good healthcare in many parts of Africa, there is evolution of use of mobile technologies such as mobile phones.

### 2.2.2. Health Care Informatics and Big Data Analytics in Health Care

At the root of quality healthcare delivery is healthcare informatics. Records have proved that health informatics which encompasses how information technology is used to enhance healthcare (Carroll et al., 2003), has led to improvement in the healthcare system. Efficient healthcare delivery is information-driven, just like other business enterprises. Physicians and other Healthcare personnel need adequate and the right kind of data to diagnose diseases, prescribe drugs and provide appropriate counsel to patients. In like manner, healthcare managers rely on quality information to make right decisions (Akindele, Dharini, Ami, Oluyemi, et al., 2018).

In many countries with efficient healthcare systems, the backbone of healthcare delivery is a healthcare information system based on robust health informatics. Health informatics ensures management of data that health personnel and institutions need to perform their duties resourcefully. Such healthcare information systems enable communication, synchronization of information, and coordination of action among many professionals within healthcare (Akindele, Dharini, Ami, Oluyemi, et al., 2018).

Adequately like in the business world, availability of big data in the healthcare sector, presents an opportunity to discover relationships, patterns, and trends within the data. Information derived from such sources has the capability of improving healthcare and lowering cost of health care. Therefore, what big data analytics can do in healthcare is, to utilize the benefits of large volumes of diverse data to obtain insights so that healthcare professionals and administrators alike, can make life-changing decisions. When there is synthesis and analysis of big data in healthcare, there is revelation of associations, patterns and trends leading to insights that providers and stakeholders in healthcare can utilize in lowering the cost of healthcare, making better diagnosis and finding and using better treatments of diseases, monitoring epidemics, and preventing new epidemics (Akindele, Dharini, Ami, Oluyemi, et al., 2018).

Business Intelligence (BI) is a tool, a technology, a process, a methodology and an architecture used to collect data and store and analyze data, in order to provide deeper insights that will help in better decision making. BI is proven to have the capability to operationalize repository data such as medical records so that evidence-based medical practice is possible, thereby improving healthcare delivery. Due to the dynamics of our society in terms of changes in laws, regulations, and the amount of data generated, healthcare organizations are employing business intelligence (BI) solutions to utilize data for accurate decision-making in order to help improve services provided to patients, reduce administrative and service costs, and safeguard the prospect of healthcare industry.

### 2.2.3. Big Data Application Through the use of Digital Surveillance in the Health Care System in Africa

Although application of big data in Africa is still in its infancy, recent reports have shown that success has been achieved using digital surveillance to track epidemics (Akindele, Dharini, Ami, Oluyemi, et al., 2018). Even though there are ethical issues to contend with in the use of big data in Africa, an outbreak of an epidemic, sometimes many epidemics become a national crisis as well as a global crisis. In order to save live and settle the political issue the government took action as prevention. In combatting infectious diseases, early detection and surveillance is critical. A digital surveillance that has been used successfully used globally and in Africa is HealthMap.

HealthMap is an open-source automated electronic information system for presenting data on disease outbreaks based on geography, time, and infectious disease carriers. Health organizations such as WHO rely on systems such as a HealthMap to detect epidemics. For instance, the dreaded Ebola outbreak in West Africa first came to global knowledge on 14th March, (2014), after HealthMap retrieved it from a French news website that reported a deadly fever in Guinea. Based on that report, Sierra Leone which shares border with Guinea officially declared on 23rd of March, that Ebola might have spread into Sierra Leone. Then on the 8th of August (2014), WHO declared Ebola as an outbreak of public health emergency of international concern. In addition, through HealthMap surveillance, new cases of Polio were detected in a war-torn north-western Nigeria. The strength of HealthMap as a disease location detection tool is based in its capability to pool huge, diverse, and unstructured Internet data, and extract useful information for users of public health. The system collects alerts daily, categorizes them based on the disease and the location, saves them in a database, and then visualizes them to the user.

### 2.3. Conceptual Framework: Big Data Maturity Models for a Reference

Big data analytics is a discipline that assists to investigate methodically separate data or generally to manage informational indexes that are too huge or complex to be managed by customary information preparing application programs.

The maturity of Big Data can be defined as the evolution of an organization to integrate, manage, and leverage all relevant data sources from internal and external. It is a way of creating an innovative ecosystem, delivering insightful business value, and enabling impactful transformations. In a clear manner big data analytics is neither having advanced technology in place to manage big and vast data nor simply using social media to analyze buzz about specific brand. It is a journey that involves building an ecosystem that includes data management, technologies, governance, analytics, and organizational components.

Big Data Maturity Models (BDMM) are the artifacts or systematic tool used to measure Big Data maturity. Big Data maturity models help organizations to create structure around their Big Data capabilities and to identify and analyze where to start. Then organizations are enabled to communicate their big data vision to the entire system through the tools that are provided from the models. The state of a company's big data capability, phase of maturity or the effort required to complete their current stage and to progress to the next stage could be measured and monitored by BDMM's methodology. Furthermore, BDMM manage and measure the speed of both the progress and adoption of big data programs in the organization.

For an organization that is considering or in the process of implementing a big data project, maturity model for big data is useful for several reasons (Halper and Krishnan, 2014). Some of the reasons are discussed below:

- BDMM assists to create structure around a big data program and determine where to start.
- BDMM helps to define and identify the organization's goals and vision around the program aligning with across the entire organization's communication process.
- Finally, the model has different methodologies that create the environment to monitor and measure the status and state of the program and the effort needed to complete the current stage aligning with identifying steps to move to the next stage of maturity.

The main goals and benefits of BDMMs are (Halper and Krishnan, 2014):

1. To enable an organization to identify specific focus areas in big data in reference to identified organizational area or pillars
2. It enables to design guiding development milestones
3. To avoid obstacles or blocking factors in establishing and building big data capabilities

To assess the maturity level of big data analytics in the health sector in Ethiopia and to identify challenges the researcher will use a defined criterion to evaluate the selected 8 big data maturity assessment models for the future work. In this chapter the researcher will evaluate 8 selected models and use the selected model as a reference to assess the maturity level of big data application in the health sector in Ethiopia.

### 2.3.2. The Concept of Maturity and Maturity Model

In general, “maturity” can be defined as “the state of being complete, perfect or ready” (Simpson & Weiner, 1989) or literally as “ripeness” (Fraser et al. 2002). Then maturity implies that “an evolutionary progress in the demonstration of a specific ability or in the accomplishment of a target from an initial to a desired or normally occurring end stage” (Mettler et al, 2010). Evolutionary progress shows that the subject or process may pass through a step-by-step state to reach on a final maturity goal. The starting or bottom state stands for an initial step/state. That means an organization having little capabilities in a specific domain, while the top/ highest stage represents a conception of total maturity (Becker et al. 2009). Maturity is normally measured as capabilities or “abilities to fulfill specified tasks and goals” (Wendler, 2012). Independent from specific domains, maturity models refer to “manifold classes of entities” (Pöppelbuß & Röglinger, 2011). Mettler (2009) argues that maturation is reflected in a one-dimensional manner and that maturing subjects can be divided into three key groups:

1) Process Maturity: it extends a specific process is explicitly defined, managed, measured, controlled and effective
2) Object Maturity: it extends a particular object reaches a predefined level of complexity or sophistication
3) People Capability: it extends the workforce is able to enable knowledge creation and enhance proficiency

In addition, Kohlegger also presented a classification system which is like the above one for maturing entities. This classification system makes a distinction between object, person, and social system maturity. According to Mettler’s research a question is raised about the artifact types in which maturity is associated with or whether maturity is associated with methods, modes or theories. Then his conclusion states that “maturity positioned themselves in between methods and models, addressing both the identification of problems as well as guidelines on how to solve these problems”. A maturity model conceptually represents “phases of increasing quantitative or qualitative capability changes of a maturing element in order to assess its advances with respect to defined focus areas” (Kohlegger et al. 2009).

In other words, maturity models are an instrument to rate organizational capabilities of maturing elements, and to select appropriate actions to take the elements to a higher level of maturity (De Bruin et al. 2005; Becker et al. 2009; Kohlegger et al. 2009; Mettler 2009). “The fundamental underlying assumption of maturity models is that a higher level of maturity will result in higher performance” (Boughzala & de Vreede 2012), as well as improved predictability, control, and effectiveness (Vezzetti et al. 2014).

The benefits of maturity models have a variety of dimensions or versatile. However, the intension of maturity models can be presented as the following:

1) **Descriptive Models:** assess the current firm maturity through qualitative positioning of the firm in various stages or phases. The model does not provide any recommendations as to how a firm would improve their big data maturity.
2) **Comparative Models:** by conducting a survey containing qualitative and quantitative information, this model targets to benchmark an organization in relation to its industry equivalent or peer.
3) **Prescriptive Models:** most of prescriptive big data maturity models follow a similar modus operandi in that first the current situation is assessed and followed by phases plotting the path towards increased big data maturity.

It can be learned that all maturity models shared a common characteristic of defining elements. These basic elements of maturity models are step by step levels, s description for each step, a summary or generic description of the properties of each step/level as a whole, a number of dimensions, a number of activities or element for each dimension and a description of each activity or element as it might be performed at each step/ level of maturity (Fraser, 2002). Nevertheless, a number of certain distinctions can be made between different maturity models based on a list of different factors. A top-down or bottom-up approach can be mentioned as one of the factors in constructing maturity models. In a top-down approach first a definition is written followed by assessment items development to fit the definitions. In the other hand in bottom-up approach first items are developed later to reflect the developed items by definitions written based on the developed assessment items (Bruinet, 2005). In addition, according to Fraser et al.

differentiate maturity models by grouping or categorizing them according to their structure type and composition.

1. Maturity grids typically contain “text descriptions for each activity at each maturity level and are of moderate complexity, requiring at most a few pages of text.”
2. Likert-like questionnaires can be considered to be a simple form of maturity models. The questions are simply statements of well-known practices, which organizations use to score their relative performance on a scale from 1 to n. (Hybrid models combine the questionnaire approach with more detailed definitions of maturity.)
3. CMM like models have a more formal and complex approach, where a number of key practices and goals are specified to reach an acceptable level of sophistication.

### 2.3.3. Big Data Maturity Models

According to Halper and Krishnan (2013) the evolution and development to earn very high returns when operating at high-capacity utilization of facility, integrate, collaborate, and manage all relevant data sources from internal and external can be determined as big data maturity model. According to Bilbao (Bilbao, 2014) define Big Data maturity as a way and approach for organizational progress assessment and to awareness of relevant initiatives. In general, Big Data maturity includes building an ecosystem hat involves governance, organizational components, technologies, analytics and data management (Halper & Krishnan 2013). Big data maturity can be measure using maturity models as an artefact. A big data program structure and a determined starting point can be created by using a maturity models for big data (Halper & Krishnan 2013). It is important for an organization to define aims around the big data program and communicate their vision across the entire organization. Furthermore, big data maturity models are advantageous for using methodologies that monitor and measure the state of the program, the effort needed to complete the existing maturity level and the steps to advance to the next stage. In addition, these models manage and measure the speed of both the implementation and progress of big data program within organization. By providing a capability assessment that focuses on specific big data crucial areas in an organization, big data maturity models aim to assist guide development milestones and to avoid factors that hinder development.

According to Krishnan (2014) suggestion the previously mentioned key areas are the subcomponents of the "people, processes, technology" triangle, namely alignment, architecture, data, data governance, delivery, development, measurement, program governance, scope, skills, sponsorship, statistical model technology, value, and visualization. The big data maturity levels/ stages show the different ways in which data can be used with in a specific organization (Darwich, 2014). Furthermore, El Darwiche's underlying assumption is that a high big data maturity stage aligns with the increase of top line revenues and reduction of operational expenses. Nevertheless, being on the highest level of big data maturity often involves major investment over many decades. Per to Radcliffe's argument big data maturity models are the crucial tools for setting the direction and monitoring the development and progress of big data program.

#### 2.3.4. Big Data Maturity Models Evaluation

Within the context of maturity models the terms evaluation and assessment is different. Evaluation is understood as the process of measuring the effectiveness of the maturity model, while assessment is understood as the act of using maturity models as a tool to measure organizational capabilities. In the sense of identifying the right improvement proposals, maturity model's effectiveness is very important. By conducting a comparative evaluation, one can ensure maturity model's effectiveness (Helgesson, 2012). In this research, normative comparison is being used as a comparison. Normative comparison's goal is to define and explain invariance of the models and give a result of the best model among all alternatives being studied.

Within the world wide web there are a list of maturity models available. But it needs a study how many of those models evaluate and measure maturity in the domain of big data. According to the researcher's analysis big data maturity model's bibliographic data bases search result only empty outputs. This data shows that big data maturity models have been developed commercially and publicly, not academically. Then the search for available maturity models has been made by using Google search engine. During the search the following combination has been used:

- "Maturity Model in the big data area"
- "Big Data Maturity Model"
- "big data"

- “maturity model”

During search using the above key words there are some more methodological and practical screening criteria are considered. Like: only results from the last ten years is considered, only documents written by English is considered, relevancy and available documents has been considered. Then the search result reduced to a manageable list. After removing duplicates, eight different maturity models were finally selected which are developed independently from each other.

The selected eight models are measured depending on the following main criterias:

- Model structure completness (completeness, consistency)
- The model development and evaluation quality (Stability, trustworthiness)
- Ease of application (comprehensibility, ease of use)
- Big Data value creation (performance,relevancy, actuality)

The above categories reperesents an aggregated criteria group comprised of different decision attributes. Model Structure completeness, the first criteria group, focuses on how extensively the model was built. Especial attributes that can be featured as metric values like: business dimensions, number of maturity levels are mostly involved decision attributes in the first crieria group. The second criteria group is the model development and evaluation quality which addresses the key problematic findings from a systematic literature review. This criteria aims to evaluate how well the maturity model development process was documnted, which standardized methods and best practices were used for development, and how well the model was tested. Thirdly, ease of application, ensures the supporting tools presented to the end user when implementing the model in reality. Quality of guideliness for maturity assessment is ensured through and focus area of evaluation in addition to the improvement recommendations. The last but not the least criteria is big data value creation, evaluate the specific maturity model based on how well the model applies to the big data domain by focusing on business value creation.

| *Completeness of the model structure* | *Quality of model development and evaluation* | *Ease of application* | *Big Data value creation* |
|---|---|---|---|
| Purpose of use | Design approach | Assessment instrument | Maturing subject |
| Industry considerations | Design process | Assessment method | Domain focus of the model |
| Composition | Subject of evaluation | Application support | Target domain |
| Maturity levels | Total number of respondents that have tested the model | Practicality of evidence | Target audience |
| Business dimensions | Reliablity | Visualization | Domain capabilities and focus areas |
| Innovation | Evolution | | |
| Total number of questions for business dimensions | | | |

Table 2.1. The aggregated criteria and their decision attributes

Finally, the eight models are identified and discussed in order of discovery below:

| *Decision attribute* | *Halper* | *Betteridge* | *IDC* | *Infotech* | *Radcliffe* | *El-Darwiche* | *van Veenstra* | *Knowledgent* |
|---|---|---|---|---|---|---|---|---|
| *Name* | Big Data Maturity Model & Assessment Tool | Big Data & Analytics Maturity Model | CSC Big Data Maturity Tool | Big Data Maturity Assessment Tool | Big Data Maturity Model | Big Data Maturity Framework | Maturity Model for Big Data Developments | Big Data Maturity Assessment |
| *Primary source* | Halper & Krishnan (2013) | Betteridge & Nott (2014) | IDC (2013) | Infotech (2013) | Radcliffe (2014) | El-Darwiche et al. (2014) | van Veenstra et al. (2013) | Knowledgent (2014) |
| *Origin* | Educational (TDWI) | Business (IBM) | Business (IDC) | Business (Info-Tech) | Business (Radcliff Advisory Services) | Business (Strategy&) | Business (TNO) | Business (Knowledgent group inc) |
| *Year of publication* | 2013 | 2014 | 2013 | 2013 | 2014 | 2014 | 2013 | 2014 |
| *Maturing subject* | Process, Object, People | Process, Object, People | Process, Object, People | Process, Object, People | Process, Object, People | Process, Object, People | Process, Object | Process, Object |
| *Domain focus of the model* | Domain-specific | Domain-specific | Domain-specific | Domain-specific | Domain-specific | Domain-specific | Domain-specific | Domain-specific |
| *Industry considerations* | General focus on industries | No focus on industries | General focus on industries | No focus on industries | No focus on industries | No focus on industries | No focus on industries | No focus on industries |
| *Purpose of use* | Comparative | Descriptive | Comparative | Prescriptive | Prescriptive | Prescriptive | Prescriptive | Descriptive |
| *Target domain* | Big Data | Big Data & Analytics | Big Data & Analytics | Big Data | Big Data | Big Data | Big Data | Big Data |
| *Target audience* | Management & IT | Management | Management & IT | Management | Management | Management | Management | Management & IT |
| *Innovation* | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| *Design approach* | N/A | N/A | N/A | N/A | N/A | Top-down | N/A | N/A |
| *Composition* | Likert-like | Grid | Likert-like | Likert-like | Grid | Grid | Grid | Likert-like |
| *Design process* | Iterative | N/A | N/A | N/A | N/A | Iterative | N/A | N/A |
| *Maturity levels* | 6 | 5 | 5 | 4 | 6 | 4 | 4 | 5 |

Table 2.2. Summary ofThe Eight Big Data Maturity Models aligning to their decision attribute's characterstics Based on the above table Halper's TDWI BDMM model is the most suitable model for the researcher's assessment.

TDWI provides individuals and teams with a comprehensive portfolio of business and technical education and research to acquire the knowledge and skills they need, when and where they need them. The in-depth, best-practices-based information TDWI offers can be quickly applied to develop world-class talent across organization's business and IT functions to enhance analytical, data-driven decision making and performance.

TDWI advances the art and science of realizing business value from data by providing an objective forum where industry experts, solution providers, and practitioners can explore and enhance data competencies, practices, and technologies.

In worldwide most data warehousing professional and business intelligence would gain advice and research from TDWI research. The main TDWI research focus areas are exclusive to Business Intelligence and/ or Data Warehouse issues. It teams up with industry thought practitioners and leaders to deliver both deep and broad understanding of the technical and business challenges through the implementation and use of data warehousing and business intelligence solutions.

TDWI BDMM differ from the other model by consisting of a significant body of knowledge. It helps to understand companies' maturity level both through quantitative and qualitative approach.

### 2.3.5. TDWI Big Data Maturity Model

TDWI (The Data Warehousing Institute) Research provides research and advice for business intelligence and data warehousing professionals worldwide. TDWI research create the environment for commentary, in-depth research reports, topical conferences, and inquiry service as well as strategic planning services to vendor organization and user.

These days many organizations collect and receive a massive amount of data without analyzing their data management capacity. TDWI big data maturity model and assessment tool is designed per to organization's requests to learn and understand how their big data implementation compare to those of their peers to provide best-class insight and support. The assessment evaluates maturity according to a standard of a big data program in an objective way across different dimensions that are crucial to deriving value from big data. Based on our experience, data becomes useful when it can be analyzed, so big data analytics is a primary focus of the model.

The model can help the researcher to understand the big data journey of the health sector in Ethiopia. It provides framework with clearly identified pillars to learn the existing and the future big data deployment and its challenge.

## 2.4. Related Work

### 2.4.1. The Maturity Measurement of Big Data Adoption in Manufacturing Companies using the TDWI Maturity Model

Data has an important role in having more organized information and advance decision-making process. Companies who can process and use data in a format of large volume, variations, speed, and high data rate can take more advantage and benefits. Nevertheless, the application of big data analytics is still not very popular in Indonesia. The telecommunication, Banking and government sector can be mentioned as the leading big data analytics implementers in Indonesia. Discussion about the challenges faced in implementing Big Data and its relation to development policy in Indonesia is still small. In fact, in the future public policy will be shaped by Big Data and its application in various aspects of public life such as in the education, health and public services sectors, companies, and others.

The aim of this article is to measure the level of maturity in the implementation of big data technology in manufacturing company. This will assist the sector in having good and sufficient information for better decision.

The study uses qualitative method to analyze research data using TDWI maturity model tool. Interview technique is used to regain respondent data where interview question and preparation guidelines made by considering the five dimensions.

The research has been conducted in four steps. Defined as follows:

- Defining Variable and indicators

- Defining respondent in-depth interview: All respondents come from the ICT department with a criterion of selection is their position and work period. Furthermore, the TDWI model is chosen with consideration that it is easier to understand and accommodate big data implementation from nascent to the mature/ visionary level.
- Analyzed levels of maturity on TDWI big data maturity model
- Analyzed the TDWI Big data Maturity Assessment Criteria

After having an intensive interview and retrieved data from the respondents it is mapped on the assessment score that refers to each dimension in the TDWI maturity model. The score table for each dimension is shown in the following figure.

THE SCORES FOR EACH DIMENSION

| Score Per Dimension | Level |
|---|---|
| <15 | Nascent |
| 16-25 | Pre-Adoption |
| 26-35 | Early Adoption |
| 36-45 | Corporate Adoption |
| 46-49 | Mature |
| 50 | Visionary |

Figure 2.1. Score for each TDWI dimension

The measurement was done by using TDWI Big Data Maturity Model. Based on the results of data collection in the field, it explained and analyzed the implementation of Big Data technology in XYZ Company that has initiated and utilized Big Data in its business processes. TDWI Big Data Maturity Model analysis framework is used to determine the level of maturity of the application of Big Data technology in the company. Maturity status can be characterized by indicators on five dimensions including: organization, infrastructure, data management, analytics, and governance, so that it can be concluded whether the application of Big Data technology in the company is in the nascent, pre-adoption, early adoption level, corporate adoption, or mature/visionary.

Accordingly, the following scores has been gained after the intensive interview with the respondents within the ICT department at the manufacturing companies. The following figure shows that the score after the interview assessment completed.

TOTAL SCORE ASSESSMENT IN THE PT. XYZ COMPANIES

| Dimension | Score | Level |
|---|---|---|
| Organization | 33 | Early Adoption |
| Infrastructure | 38 | Corporate Adoption |
| Data Management | 36 | Corporate Adoption |
| Analytic | 36 | Corporate Adoption |
| Governance | 34 | Early Adoption |

From the identification process carried out above it can be concluded that the application of big data in the manufacturing companies is at the level of corporate adoption for the dimension of infrastructure, data management and analytics. However, organization and governance dimensions are still in the early adoption stage. These companies' maturity level is characterized by integrated big data supporting infrastructure, where end-users have been involved in the system and benefited. To be able to reach the mature/ visionary category this research work suggested improvement areas or like adding cloud infrastructure, improving aspects of data security, and enhancing the work culture and awareness of employees and stakeholders to utilize the big data technology that has been built.

To be fully beneficiary of the big data analytics application within a sector understanding and learning the current situation of the company is very crucial. The above reviewed work is related to this research work in its initiatives to start the work and the methodologies and tools, which is TDWI maturity model, it used to assess the maturity level within the manufacturing sector. However, as it is limited and focused on the manufacturing sector it is different from this work. Both the reviewed work and this research work uses qualitative method to conduct the research.

# Chapter 3: Methodology

## 3.1. General Approach

By using hermeneutics as a philosophical approach, this research work is mainly defined as deductive-exploratory. Interpretation and description of academic research papers are used as benchmarking criteria and a model and decisions are used for maturity assessment. This research uses concepts in different academic papers as a theoretical foundation for the research which resulted in a deductive approach.

This research work applied the research onion's concept to conduct the research. The "Research Onion" concept is proposed by Saunders defined and explained that the outer layers represent the context and boundaries within which the data collection and analysis technique will be selected.

The research onion which is used in this research as a reference is showed as follows:

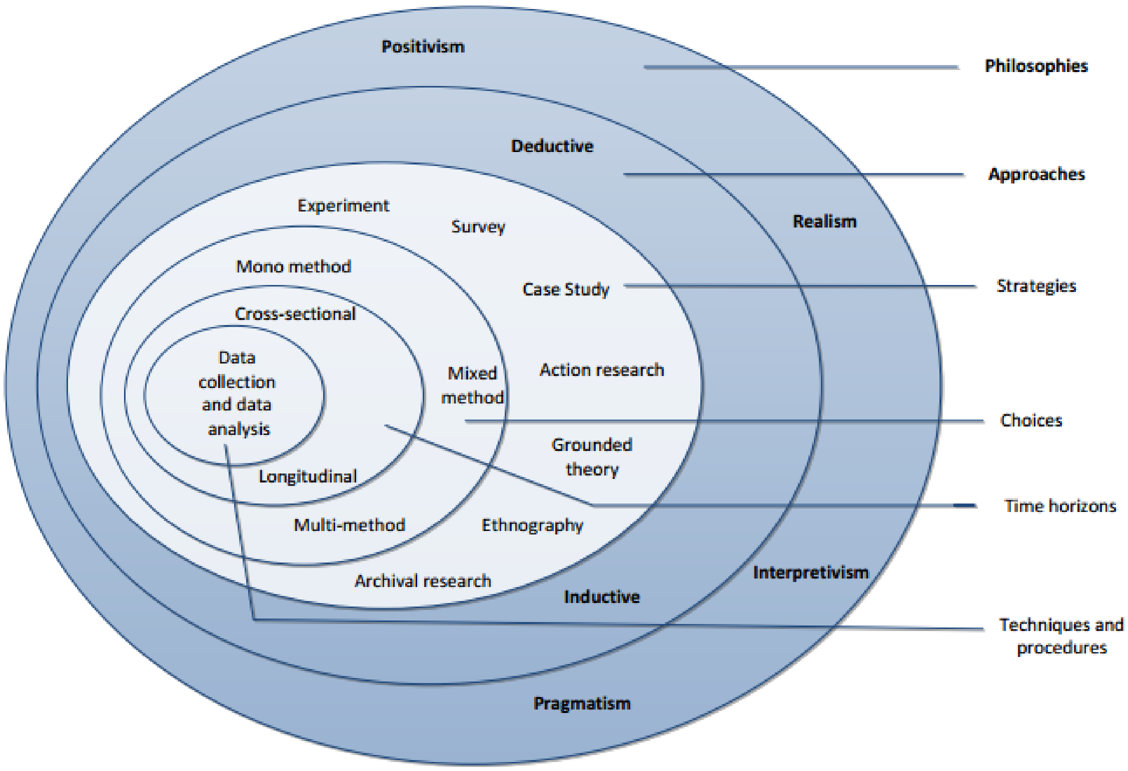


Figure 3.1. The Research Onion (Saunders, Lewis, Thornhill, et al., 20176)

## 3.2. The Research Strategy

In this research work qualitative method is being used as it is intended to explore and provide a detailed description and new insight.

The data for this research is not collected using observation or experimentation which result to the lack of empirical evidence. In this research work the researcher used conceptual approach as the qualitative research strategy, which in return guides to choose the appropriate data collection and analysis technique.

## 3.3. Data Source, Population and Sampling

For this research Ministry of Innovation and Technology (MINT) and Ministry of Health (MOH) are the main sources of the data. The two ministries work in collaboration to enhance the health sector through the application of big data analytics.

The population is selected using Purposive Sampling Technique. As the research is considering only two stakeholders and the respondents are from ICT department within this organizations.

The researcher chooses purposive sampling to simply generalize the outcome depending on the respondents' feedback. Accordingly, the experts who are directly participated in this research are selected through the purposive sampling as they have direct experience and exposure with data analytics area specifically within the health sector

Purposive sampling is a commonly used sampling strategy in that it is designed to provide information-rich cases for in-depth study. This is because participants are those who have the required status or experience or are known to possess special knowledge to provide the information researchers seek.

## 3.4. Data Collection Methods and Approach

Data collection and analysis techniques should be first identified to gather and get the right data and information. In this research work systematic literature review is being used.

The process of data collection is directly related to sampling and is best viewed as complementary to it. In this research work data is being collected directly from the identified and selected sample population. Correspondingly, interview is used as a direct data collection method. The interview questions are both structured and unstructured.

Data from document and previous work review and analysis is also considered as indirect data collection method.

## 3.5. Data Analysis Technique

For this research purpose content analysis and narrative analysis is being used. The researcher used content analysis to study and understand the sector from different documents and responses from interviewees.

For further understanding of the sector on how to different actors work in collaboration, the researcher uses narrative analysis on the information gathered from different experts who are working in the primary actors in the sector.

During content analysis technique being used for this research the following procedure has been followed.

1. The research question being designed and presented clearly
2. Each research question has its own content to be analyzed with sub classified questions associated with the interview question

3. Tag interview answers and discussions with its corresponding content according to the selected big data maturity model
4. Clearly understand collected data according to the model and relate the discussion with the appropriate meaning per the model implication
5. Analyze the implications and draw conclusion

For narrative analysis a common procedures has been followed. The procedures are discussed as follows:

- Organize and prepare the data
- Common understanding or general sense of the information
- Categorize the data
- Identify categories according to the selected big data maturity model (BDMM)
- Interpret the data

# Chapter 4: Conceptual Framework

## 4.1. Context of the TDWI Big Data Maturity Model (BDMM)

Big data is a combination of old and new technologies not a single technology, which assists companies to gain actionable insight while effectively managing data storage and load problems. Big data analytics involves and requires the ability to collect, manage and analyze potentially huge volume of disparate data, with the appropriate speed, within the right time frame, while resulting right time analysis and activity to the end user (Halper and Krishnan, 2014).

For an organization which is in the process or considering of a big data project, a maturity model for big data is useful for a list of reasons. At the beginning it assists to create structure around a big data program and determine where to start. After defining a clear starting point, it also aids to define and identify the organization's goal around the program and creates a process to communicate that vision across the entire organization (Halper & Krishnan, 2014).

Maturity models provides a method and methodology to monitor and measure the state of a program and the effort needed to complete the current stage, in addition to identifying steps to move to the next stage of maturity (Ashwin Belle, et al., 2015). It serves as a kind of odometer to manage and measure the speed of big data program progress and adoption within the company for the program. The TDWI big data maturity model comprised of five stages called Nascent, Pre-adoption, Early Adoption, Corporate Adoption, and Visionary/ Mature. The more organizations step up via these stages, they should acquire more value and benefits from their investments.

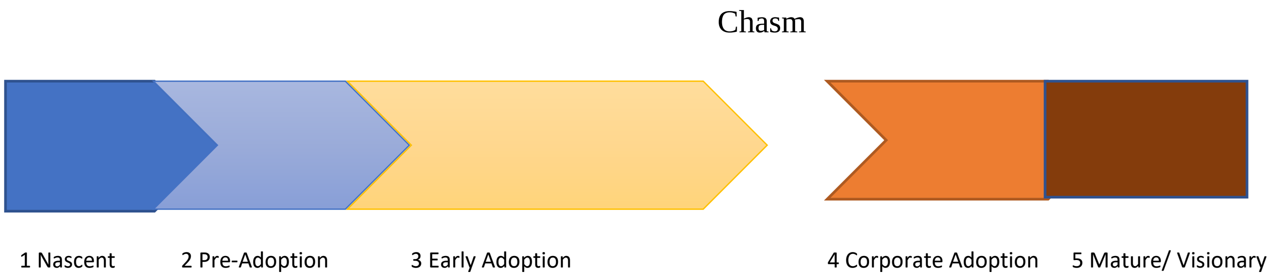


Figure 4.3. Stage of Maturity in the TDWI Big Data Maturity Model (Ashwin Belle, et al., 2015)

The maturity of big data analytics and big data program is measured in an objective way across various dimensions using TDWI big data maturity model assessment tool. These dimensions are key to gain value from big data analytics, called Infrastructure, Analytics, Data Management, Governance and Organization (Halper & Krishnan, 2014).

**Stage One: Nascent**

A pre-big data environment is represented by nascent stage. In this stage, most companies have insufficient or low awareness of big data and its appropriate value across much of the business. There is no real executive support for the program or effort, although there are pockets of people spread throughout the company who may be interested in the potential value of big data (Halper & Krishnan, 2014). Key characteristics of the nascent organization includes:

**Organization:** at this stage the organization has started the concept of analytics and furthermore it may have a data warehouse, for example, but it has not yet started to explore advanced big data journey or analytics.

**Infrastructure:** big data management requires an in depth understanding of the new infrastructure to support critical big data components like data volume, data type, data security and user, processing requirements, information storage, and access to information. An organization in nascent stage often misunderstand the need to differentiate between data and infrastructure, which leads to short lived success in their big data journey.

**Data Management:** a nascent organization will have also collected data as files with variety of formats but lacking some important points like: storage structures that are minimally defined and naming standards. A nascent internet-based business may have the same problem in terms of collecting data without real data management strategy in place or idea of what to do with the data. As a matter of fact, as no one knows what to do with the given data at this phase, data that could be useful may be discarded. Furthermore, the organization typically doesn't have much in the way of an end-state data architecture. Its data life cycle management strategy and data strategy are not strong and are often more about immediate results and silos of information than an overall plan.

**Analytics:** A nascent organization may not or may use of advanced analytics. Nevertheless, the nascent organization is often forcing the boundaries and limits what it can do analytically. To put it another way, in nascent organization analytics is happening in silos and pockets. Usually, if an organization is working only with structured data, then it is utilizing advanced analytics like predictive analytics.

**Governance**: The nascent organization has bought into the concept of analytics, and it may have a data warehouse, for example, however it has not started to explore advanced analytics or begun its big data journey. In a way it means that its governance strategy is more IT centric than IT and business centric.

**Stage Two: Pre-Adoption**

This level shows that the organization is starting to do on its responsibilities regarding big data analytics. Staff may be reading up on the related topics and attending conferences. The organization may have invested in some new technology like Hadoop to assist the big data program. However, the effort to implement big data analytics is usually departmental in scope, the organization aware of big data analytics will be implemented in the near term (Halper and Krishnan, 2014). Key characteristics of this stage includes:

**Organization:** in this stage, the organization is just beginning its big data journey. The mindset is generally around experimentation. By trying to determine the top business problems to solve, the team charged with exploring big data may have also brought in some business partners to help with this. The team realize and learn that identifying the right business problem is crucial and critical for success. The organization typically implement analytics as part of its decision-making process in various departments, however the organization is not necessarily analytics driven. At this stage some companies experience skepticism around their big data and analytics program.

**Infrastructure:** during this stage of maturity the company is exposed for trying out new technologies to assist big data analytics program like Hadoop or some other big data technologies as part of the experimentation phase leading to a proof of concept. The maintenance, installation and configuration are key points that are defined during this stage. In addition, there is a level of compliance to an enterprise standard, though, the class of infrastructure may not be close to production ready or may implement a basic use case.

**Data Management:** An organization in the preadoption stage of maturity may have started to identify and collect some big data sources (this is mostly from internal source) as part of the experimentation process. Nevertheless, the collected files lack uniformity like minimal naming standards, minimally defined storage structure, and different formats. In addition, there is a lack of data auditability and lineage, and data life cycle management.

**Analytics:** An organization in the preadoption stage has explored and started advanced analytics. There may be a group of individuals, business analysts or statisticians, who have part of the overall organization’s voice, who adept predictive modeling. This shows that they understand the value of advanced analytics.

**Governance:** Organizations at this stage might have a steering committee managing and monitoring the big data program from a governance perspective. This committee has representatives who are dedicated for providing reports on progress and later on compliance. Nevertheless, most do not.

**Stage Three: Early Adoption**

The early adoption stage is mainly characterized by one or two proofs of concept (POCs) which become more production ready and established. Organizations tend to spend a long time in this stage, usually because of it is hard to cross the chasm that leads to corporate wide adoption of big data and big data analytics (Halper & Krishnan, 2014). Key characteristics of this stage includes:

**Organization:** During this stage there is at least one executive sponsor involved in general. Nevertheless, it is also at this stage when more executives might start to become more interested in the big data program as organizations show some wins and preference in the POCs. The more an organization gets excited about the prospects of big data, more people and professionals start to come on board. In return, this means that a formal team is formed to initiate to strategize and plan for a wider big data scope. This might lead to politics may start to kick in, if exceptionally the goal of the big data project is cost containment rather than competitive advantage.

Infrastructure: At the early adoption stage there are various kinds of big data technologies in place. This might involve a Hadoop cluster or an appliance. Moreover NoSQL databases might also be in place. Nevertheless, the ecosystem or the notion of a unified architecture is not yet spread, and these technologies are not operationalized. During this stage a common typical infrastructure is a tier 2 production class cluster which is maintained and installed in the organization's data center or even in the cloud. The configuration, maintenance and installation are defined per to the organization's standard. If the NoSQL database is implemented, its purpose is not part of the operational infrastructure and generally offline. Hadoop might be used in some companies where cost saving is paramount, to take advantage of growing data volumes without expanding the data warehouse. Though this is a valid hybrid approach, it may lack well management, and the data may not be used yet for meaningful analysis.

**Data Management:** During the early adoption maturity stage organizations will have data collected as files of potentially with division or enterprise standards for naming, storage management and different formats. There may be metadata attributed at the division level and a defined end state data architecture, whether it is in a semantic layer, in a database layer or in the platform architecture, from a data strategy perspective. Except there is a typical purpose to throw data out, organization is not doing so. Nonetheless, a solid companywide big data management strategy is specifically not in place. If an organization is aware of issues relate to security and data quality, it enables to move at a fast pace to implement its big data and big data analytics initiatives.

**Analytics:** An organization in the early adoption stage might be implementing predictive or even descriptive analytics in its projects. During this stage, the specific problem that the company is trying to deal with determine which kind of analytics tool to be used. Though, the data vary among organizations in this stage, organizations are yet using one kind of data.

**Governance:** In this stage department may be running some sort of predictive model on large volumes of structured data that is stored in an appliance. However, it is yet not moved on to other format of data and departmental in scope. Even if departments try to implement different kinds of data, it is not in an assimilated or integrated manner. This means there is no integration across departments in such organization.

**Transition: The Chasm**

The Chasm is simply a transition period from Early Adoption stage to Corporate Adoption Stage. There is basically a series of barriers that organizations need to overcome as they try to move from early adoption to corporate adoption. This is usually why organizations spend a large amount of time in this stage. In addition, there is obvious challenges of getting the appropriate skill set like: advanced analytics skills, Hadoop skills and political issues may not be present in the organization (Halper & Krishnan, 2014). For a given organization to successfully cross the chasm phase, it will need to handle the following challenges:

**Funding:** Most of the time visionary executive IT champions are the reasons to drive or bootstrap many early big data projects. Indeed, it is extremely important to establish wins with these projects to stable future funding. Business involvement is critical here as IT funding is usually not enough. Mostly there is a lack of buy-in and business funding so, some organizations are not able to get POC phase or past the prototyping. There must be a connecting bridge between the business in big data initiatives and IT. For the sake of having business value with tangible business outcomes big data projects need to involve business and its important concepts.

**Data Management and Governance:** a strong data governance and management plan needs to be in place in order to move forward to sharing of data and corporate adoption. To have a common knowledge and understanding through a company, a formal format of data management and government is very crucial.

**Architecture:** Choose the best approach according to the big data analytics application stage.

**Skill Sets:** Developing the skill sets for new technologies like NoSQL or Hadoop is the main hindering factor for big data projects moving pass the chasm. Most of the time companies hire professional for this as per their affording capacity. But the new staff fall to understand the business well. There can also be a disconnect between the traditional data warehouse team and the Hadoop team. It is very crucial to get these two different teams together.

**Cultural and Politics Issue:** Big data analytics often becoming pervasive throughout the organization just because of cultural issues and political issues, typically when the motivation for big data is more about cost containment than competitive advantage. With the lack of road map and solid plan or business case in place with executive sponsorship (that staunchly supports the plan) or a determined set of product managers, this may stop the project from progress.

Governance: In order to move to the corporate adoption stage, it requires to establish a big data governance team. Some organizations assumes that they are mature only by having a big Hadoop cluster and corporate sponsorship in place.

In the big data journey crossing the chasm stage can be a longtime taking exercise unless the organization has both bottom-up and top-down strategies in place. Otherwise, it needs to compel by a strong business reason to make big data happen. In order to cross the chasm stage organizations, need to ensure that the right data architecture, governance, organizational structures, and data life cycle management security strategies are in place.

**Stage Four: Corporate Adoption**

In any organization's big data journey corporate adoption is the major crossover phase. Corporate adoption phase highly increases end user's involvement through assisting them to gain insights and transform how they do business. For example: they may change how decisions are made by operationalizing big data analytics in the organization. Mostly there are some certain gaps in an organization that are repeatedly addressed by organizations which are trying to reach at the corporate adoption maturity stage like: infrastructure, organization, data management, governance, and analytics issues (Halper & Krishnan, 2014).

**Organization:** Usually during this stage, companies end up realizing that analytics is a competitive differentiator. Innovation in data analysis and data is a core value, and a predominate analytics culture. With the data infrastructure support, the business strategy is generally bottom-up and top-down.

**Infrastructure:** The basic company infrastructure during this stage of maturity is a tier 1 production class cluster. It is maintained and installed in the data center, which might involve the cloud. The recovery or disaster recovery and backup procedure is complying with the architecture and infrastructure of the big data ecosystem. A unified architecture that takes an ecosystem approach is also in place.

**Data Management:** An organization can manage and make use of its data by becoming a more mature company. During this stage of maturity organizations can make use of many forms of data. A strong data governance policy can be well managed through a collaborative data sharing activity. The organization perceives value from all components of the data infrastructure and data siloing is being minimized during this stage.

**Analytics:** An organization might collect a lot of data without proper use of it. By analyzing new data coming into an organization quickly and made part of the logical infrastructure, it is the sign of maturity. In this phase, analytics highly supports the organization. In addition, a center of excellence (COE) is typically existing in a company during this stage in order to serve different parts of the organization. The COE involves the data science team, which might even train other groups in the process of using analytics in different forms. It is generally also having the goal to evangelizing big data and its members might also be looking for shared groups in analytic needs to determine what might be reusable in other parts of the company. During this point analytics might be functionalized as part of a business process. In other sense it means that business process might be automated or integrated with analytics. For instance, using other data about customers together with machine-generated data to determine a net best offer.

**Governance:** A truly mature organization aware of and understands that big data can be a liability waiting to happen. This organization is interested with answering WH questions. During this phase of maturity companies will have program governance in operation. The program is guided by PMO and a steering committee that oversees the program from a company perspective. A similar effort will also be in operation with a well-defined data management and strategy and a steering committee overseeing the progress of data in data governance.

**Stage Five: Visionary/ Mature**

In terms of big data and bid data analytics, only a few organizations are considered as visionary or mature currently. During this stage, organizations are executing big data programs as well-oiled machine using data governance strategies and a highly tuned infrastructure with well-established program. The program is implemented as a planned and budgeted initiatives from the company perspective (Halper & Krishnan, 2014). In the Visionary or mature stage, there is energy and excitement around big data big data analytics.

**Organization:** The visionary or mature organization demonstrate different characteristics. First, an executive has bought into big data and big data analytics view it as critical, de facto standard for how to do business. During this phase the mindset is creative, and analytics is seen as a competitive weapon. Secondly, analytics isn't simply used to drive insight or strategy, instead, the business is always looking for opportunities to use analytics in new ways. As a matter of fact, a visionary organization is characterized by its continuous determination of new ways to use and create value from analytics. Finally, collaboration and integration are a big part of the culture.

**Infrastructure and data management:** For a big data maturity the way complexity is managed is a key point. The visionary or mature company has implemented a coherent analytics infrastructure that is fully operational and can be used in the mission-critical aspects of the business. The infrastructure part involves the ability to integrate new sources of data for analytics, whether they are external or internal to the company. The infrastructure leverages enterprise NoSQL databases and Hadoop and there is disaster recovery, security, backup and recovery, proactive infrastructure monitoring and performance management in place even if the public cloud is used. The program is given the right amount of staffing and skills and treated as mission critical. Data is shard across the organization.

**Analytics:** Mature or visionary organization is continuously working to develop analytics. Specifically, this kind of organization usually use all kinds of data in addition to real time data. It uses this data to incorporate into business processes and for its decision making. The mature or visionary organization connects the dots or gups between existing assets and new data. The business then can bring new products to market and enhance product offering. The data, analytics and infrastructure are completely intertwined.

There are teams and COEs in this organization who are working to deliver exciting and new forms of analytics. A few visionary companies establish an innovation team consisting of IT and business that innovates on the technologies, take them in to production and brings them back to the business.

### 4.2. TDWI BDMM: Evaluating Benchmark score

The dimensions of the TDWI big data maturity model are formed from the benchmark survey that has approximately 50questions across the five categories (Halper & Krishnan, 2014). The following discuss the contents of the questions under each five categories:

- **Organization:** the questions under this category could be:
    - How far the key points like organizational leadership, organizational strategy, organizational culture, and fund raise assist the company's BD analytics program success?
    - What is the value given by an organization for analytics?
- **Infrastructure:** the questions under this category could be:
    - How coherent and advanced is the architecture in support of a big data initiative?
    - To what range does the infrastructure support all parts of the potential users and company?
    - To what extent the big data management approach is effective?
    - What are the key technologies implemented to support a big data initiative? And how these technologies are integrated into the existing environment?
- **Data Management:** the questions under this category could be:
    - How extensive are the volume, velocity and variety of data used for big data analytics?

- How does the organization manage its big data with the assistance of analytics? (This involves data processing and quality as well as data storage and integration issues)

➢ **Analytics:** the questions under this category could be:

- How advanced big data analytics the company is using? (This involves how the analytics is delivered and kinds of analytics utilized in the company. In addition, it involves the skills required to make analytics in place)

➢ **Governance:** the questions under this category could be:

- How the organization's data governance strategy is consistent to assist its big data and big data analytics program?

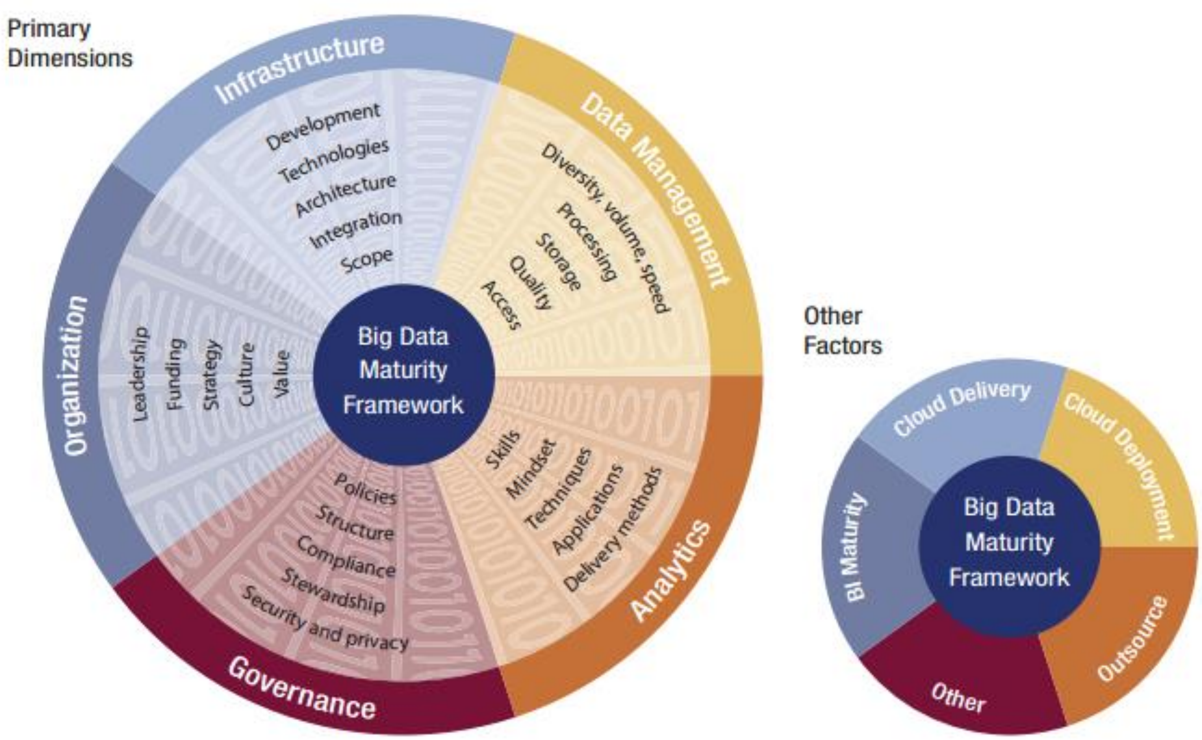


Figure 4.4. Big Data Maturity Assessment (Halper & Krishnan, 2014)

### 4.3. TDWI BDMM: In case of Health Sector in Ethiopia

The African continent has the highest mortality rate the world, and as at today, it is the only region of the world where deaths from communicable diseases exceeds the deaths from chronic diseases (Deaton, Angus, Robert, et al., 2016). The Government of Ethiopia (GOE) is working to strengthen the healthcare system to align it with the Millennium Development Goals. Ethiopia has a large, predominantly rural and subsistence agriculture population with poor access to safe water, housing, sanitation, food, and health service. The government has made significant investments in the public health sector that have led to improvements in health outcomes. Nevertheless, communicable diseases like HIV/AIDS, TB, malaria, respiratory infection, and diarrhea remain a serious challenge in Ethiopia. High fertility rates, and low contraceptive prevalence continue to drive a rapidly increasing population in Ethiopia.

From this different investment on the health sector, the investment for enhancing the application of big data application could be mentioned. In order to understand the maturity level of the sector TDWI maturity model is chosen and applied. The model has different stages and dimensions for better and detailed understand of big data maturity stage.

The Ministry of Innovation and Technology (formerly known as the Ministry of Science and Technology (MoST), Ministry of Communication & Information Technology, Ministry of Science and Technology) is an agency of the government of Ethiopia. It was established as a commission in December 1975 by directive No.62/1975. The Technology and Innovation Institute “TechIN” is an institute (reporting to MINT) of the Government of Ethiopia mandated by law with wide range of objectives to develop problem solving technologies in the social and economic sector of identified fields through researches and dissemination. This institute has a dedicated directorate called computing and analytics directorate to manage big data analytics. The health sector is considered as one of the sectors the director is working on. EPHI under MOH is also working to develop the health sector through big data analytics application.

The two ministries experience in applying big data analytics is studied according to five perspectives which are: organization, Infrastructure, Data Management, Analytics and Governance. This analysis will help to understand in which stage Ethiopian big data analytics maturity lay on. The detail characteristics of the sector according to the pillars will be discussed on following sections.

# Chapter 5: Result and Findings

## 5.1. Overview

Although maturity models were initially used in software related applications, healthcare organizations have also been using them as organizational management tools. Normally, the result of using these models is different levels/stages of maturity and activities/checklist items that they encompass. After deep evaluation of different big data maturity models this research adopted the TDWI BDMM.

The TDWI Big Data Maturity Model applied to understand the maturity level of the sector in Ethiopia based on the intensive interview and the sector study the maturity level of the sector is identified.

## 5.2. TDWI BDMM as a Reference in Ethiopian Health Sector

In order to learn big data analytics application maturity in the health sector in Ethiopia it is important to design a conceptual framework using big data maturity model as a reference. For this purpose, TDWI Big Data Maturity Model is used as a reference. Based on this model Big Data Maturity in the Ethiopia's health sector also showed one of these stages characteristics. To elaborate this in detail the characteristics of big data is discussed in detail as follows.

The Ethiopian Government is working to enhance the health sector through the application of advanced technologies. Big Data Analytics could be mentioned as one of these examples. To assist the application of this concept Ministry of Innovation and Technology (MINT) is working in collaboration with Federal Ministry of Health of Ethiopia (FMOH) to initiate the Big Data Analytics application in the health sector. The first workshop was launched after the establishment of Technology and Innovation Institute under the ministry. The workshop was titled as "The Foundation of Big Data and Computing" with about 80 participants from concerned organizations. This and similar workshops and different research enable the ministry to understand the benefit of big data analytics specially in the health sector.

**Organization:** The Ministry of Health shall have Powers and Duties to formulate the country's health sector development program, follow up and evaluate the implementation of same, support the expansion of health services coverage, follow up and coordinate the implementation of health programs financed by foreign assistance and loans, direct, coordinate and follow up implementation of the country's health information system, devise and follow up the implementation of strategies for the prevention of epidemic and communicable diseases, follow up and coordinate the implementation of national nutrition strategies, take preventive measures in the events of emergency situations that threaten public health, and coordinate measures to be taken by other bodies, ensure adequate supply and proper utilization of essential drugs and medical equipment in the country, prepare the country's health services coverage map, provide support for the expansion of health infrastructure, supervise the administration of federal hospitals, collaborate with the appropriate bodies in providing quality and relevant health professionals' trainings within the country, provide appropriate support to promote research activities intended to provide solutions for the country's health problems and for improving health service delivery, expand health education through various appropriate means, ensure the proper execution of food, medicine and health care regulatory functions (FMOH Page).

Among this managing health related data is one of its mandates. To enhance the ministry's service, it applies big data analytics in different department and institution within the ministry. The Ethiopian Public Health Institute is currently working with USAID to manage COVID 19 data which helps to understand the pandemic spread, which area and society is vulnerable and related issues.

There is a dedicated directorate for big data analytics under the Technology and Innovation Institute called Computing and Analytics Directorate (CAD) in MINT. This directorate is working on the big data analytics and computing sector in different sectors like: Health, Investment, Transportation and so on. The directorate conducted a "National Public awareness Survey" to understand the public understanding about the big data and analytics concept.

The concept of big data analytics is also practicing in the health sector by collaboration of EPHI and CAD. The MINT is giving technical assistance for MOH. This enables MOH to apply big data analytics in its decision-making process.

**Infrastructure:** The Computing and Analytics directorate is collaborating with different government offices to experience big data analytics and get advantage of it. For this purpose, the directorate is using different technologies like Python to apply big data analytics. EPHI is also working with this directorate and with the aid of USAID specifically on big data analytics and its application. The dedicated responsibility of EPHI for COVID 19 could be mentioned as the institutes work through big data analytics application.

**Data Management:** Both MOH and CAD is working to collect health related data from different governmental organizations and concerned body. The data could be in different format and form different system. However, as there is no digital data in most of the government organization the data collections are bottleneck for both MOH and CAD. This problem leads to no data auditability and lineage and less defined data life cycle management.

**Analytics:** The minister done analysis on the collected data depending on output preference. This enables CAD to align its technical or analytical skill with specific sector. This shows that the sector expert working with CAD through the analysis. CAD is working somehow advanced analysis on health-related data either by itself or in collaboration with MINT and EPHI.

**Governance:** MOH, CAD and EPHI has their own department which have the responsibility of managing and oversee the application of big data analytics program. Under the Innovation and Technology Institute CAD is responsible for big data analytics application in different sectors through the country. The health sector could be mentioned out of these sectors the directorate is working.

## 5.3. Finding and Results

### 5.3.1. Finding

By using the TDWI BDMM the big data application in the health sector is discussed in the above section. The five perspectives (Organization, Infrastructure, Data Management, Analytics and Governance) are used in studying and understanding the sector. Through the study the maturity of the big data application in the health sector is mostly show the characteristics of **Pre-Adoption stage.**

The health sector starting to enhance its services through the application of big data analytics. The workshop hosted by CAD could be mentioned as one of the initiations made by the different sector players to practice big data analytics.

The EPHI is investing a lot on different technologies including big data analytics technology to assist its application. EPHI is currently highly engaged in COVID 19 related data management duties which will add an experience for big data analytics.

However, most of the duties and responsibilities are departmental in scope which need to grow into institutionalize.

### 5.3.2. Finding/ Factors

According to the analysis the big data analytics application in Ethiopian health sector is in its Pre- Adoption stage. The main aim of this research is to identify the maturity of the application of big data analytics in the health sector. The following bulletin will discuss the characteristics of big data analytics application according to the five perspectives.

**Organization:** In the perspective of organization there are some factors that hinder the application of big data analytics or block to upgrade to the next maturity stage. There is no clear leadership and strategy on how to manage the big data. The currently existing funding needs to be permanent in order to ensure the deployment of big data analytics.

The awareness in the organization level is well initiated but the culture of the health sector should be well communicated across the organization. This will help to create a common understanding across the different actors.

**Infrastructure:** There are different technologies to assist the application of big data analytics in the health sector which is managed by different organizations. MOH and MINT should work more to integrate their system to develop a strong architecture across the country. Currently systems are not integrated which leads to hard data management scenario. The Scope of each ministry regarding with big data analytics application is vague.

**Data Management:** Volume, Velocity and Variety are all vital for health care big data analytics but there are more V words to think about like Veracity, validity, Viability, Volatility, Vulnerability, visualization, and Value. Big data analytics in medicine and healthcare covers integration and analysis of large amount of complex heterogeneous data such as various – omics data (genomics, epigenomics, transcriptomics, proteomics, metabolomics, interactomics, pharmacogenomics, diseasomics), biomedical data and electronic health records data.

In Ethiopian health sector there is big data indeed but there are some missing characteristics according the V big data characteristics. The current existing health related data is only exhibiting the characteristics of velocity and variety. The storage of this big data should also enhance. The Ministry of Innovation and Technology is currently working to enhance the storage capacity of the data which will help to advance data processing capacity. The data from the health sector is somehow accessible but it need a standard format for better accessibility. As most of governmental organizations fell to digitize their data, the data accessibility is vulnerable for difficulty.

**Analytics:** There are a lot of skilled experts in both ministry but there is a lack of expert to interpret even the existing analyzed data. This leads to miss the benefit from big data analytics. There should also be a clear roadmap on how to deliver the outputs to the concerned organization.

There should be an advanced analytical technique in order to develop and enhance the maturity of the big data analytics.

**Governance:** The first and crucial factor under this section is the lack of policy to govern the big data management and big data analytics application. For example: MOH is currently willing to work with MINT on big data analytics only with limited duties. MOH only need technical assistance from MINT. This is because of the lack of policy on how to manage data security and privacy. As known health care data is very sensitive and need security.

The other factor is there is no clear structure on how to govern big data analytics application through different programs.

## 5.4. Discussion of Result

In the above section the characteristics of big data analytics application in Ethiopian health sector has been discussed according to the five perspectives. The five perspectives are: Organization, Infrastructure, Data Management, Analytics and Governance. Through the study the maturity of the big data application in the health sector is mostly show the characteristics of **Pre-Adoption stage.**

The Pre-Adoption stage shows that the organization is starting to do on its responsibilities regarding big data analytics. Staff may be reading up on the related topics and attending conferences. The organization may have invested in some new technology like Hadoop to assist the big data program. However, the effort to implement big data analytics is usually departmental in scope, the organization aware of big data analytics will be implemented in the near term.

Accordingly, the awareness in the organization level is also well initiated in the Ethiopian health sector but the culture of the health sector should be well communicated across the organizations. Regarding infrastructure to support the generating big data, it is inefficient.

The Ministry of Innovation and Technology is currently working to enhance the storage capacity of the data which will help to advance data processing capacity. The data from the health sector is somehow accessible but it need a standard format for better accessibility.

On the subject of skill, there are a lot of skilled experts in both ministry but there is a lack of expert to interpret even the existing analyzed data. This leads to miss the benefit from big data analytics. The other major factor is the lack of policy to govern the big data management and big data analytics application.

In general the above discussed characteristics of big data analytics application in the Ethiopian health sector shows similar characteristics as of pre-adoption stage under the TDWI big data maturity model (BDMM). In which it makes it fall under the Pre-Adoption stage.

# Chapter 6: Conclusion and Recommendation

## 6.1. Conclusion

The healthcare industry is one of the intensive and sensitive organization which is generating a massive amount of big data with the various format. In order to manage and provides meaningful insight from these data, it needs serious attention to perform analytics using the right techniques and modern tools to enhance the quality of service and reduce its cost. For an organization that is considering or in the way of deploying a big data project, big data maturity model is important for a list of reasons. BD maturity can be taking up as a developing organization to handle, desegregate and earn very high returns when operating at high-capacity utilization of a facility from all pertinent data and information sources. Which implies enabling dynamic transformation and changing, deliver forward-looking ecosystem and delivering insightful business value. Which implies that BD maturity focus is not only concerned with advanced technology deployed to manage high volume of data but also the above issues.

To assess the maturity level of big data analytics in the health sector in Ethiopia and to identify challenges the Comparative Model specifically the TDWI Big Data Maturity Model is used as a reference. TDWI (The Data Warehousing Institute) research provides advice for data warehousing and business intelligence professionals and provides different research materials. The main focus areas of TDWI research are data warehouse and business intelligence issues with a collaborative work with industry thought leaders and practitioners for an output of broad and deep awareness and learning of the technical and business challenges or hindering factors around the operation and deployment of DW/BI solutions.

The Data Warehouse Institute BDMM evaluate the maturity stage of the big data analytics maturity in the health sector in Ethiopia according to five perspectives namely: Organization, Infrastructure, Data Management, Analytics and Governance. Through the analysis the researcher identified that the big data analytics maturity is in the Pre-Adoption stage.

After a deep analysis of the sector there are factors identified that hinder the application of big data analytics in the health care. Lack of clear leadership and strategy on how to manage big data, the lack of permanent fund allocation and common understanding across organization are the factors identified under the organization perspective. Lack of system integration and Vague scope through big data application are important points identified under the infrastructure point of view. Lack of data management strategy, lack of policy to handle data confidentiality and undigitized data with lack of standard is the factor under data management issue. Low Business Intelligence (BI) maturity and lack of advanced analytical skill are the factors that need improvement under the Analytics perspective. The last but not the least factor under is lack policy to govern big data application and analytics under the governance point of view.

## 6.2. Recommendation

The main contribution of this research work is to understand the maturity of big data analytics in the health sector in Ethiopia and provide recommendation to upgrade the sector's application to the fullest.

A developing country like Ethiopia is facing various problems in the health care because of lacking advanced systems to enhance the sector. Identifying problem in each sector is the beginning of the work to develop the system. The application of big data analytics has a lot of advantages for the health sector through improving its services.

In the above section the researcher finds out factors according to predefined pillars that hinder the application of big data analytics specially in the health sector through the assessment. In this section a recommendation will be developed to assist the sector's growth in the health care. The following are the recommendation that a related future work might consider:

- Design Clear Leadership strategy to manage big data analytics both in MINT and MOH.
- The concerned body to oversee big data analytics application in health care should secure permanent funding to support the sector's sustainability.
- There should be a common understanding across organization and department to contribute for the same common goal.
- All health-related systems should be integrated for the best data management process.
- Government organizations should invest on digitization to have a standard data.
- The scope of organizations (specifically MOH and MINT) should be clear in managing big data and analytics.
- Data confidentiality governance should get solution

- MOH and MINT and all concerned organization should work on Business Intelligence (BI) maturity. After having analyzed data the sector specific expert should know how to handle and use it for further decision making.
- The technologies for big data analytics should be advanced. For example: Hadoop Cluster
- A policy must be designed to govern big data management and analytics in the health care.

## References

[1] Wamba, S. F., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. International Journal of Production

[2] Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. Health information science and systems, https://doi.org/10.1186/2047-2501-2-3

[3] Dembosky, A. (2012). Big Data: Data prescription for better healthcare. Financial Times.

[4] Nsoesie et al (2015), Nsoesie, E. O., Kluberg, S. A., Mekaru, S. R., Majumder, M. S., Khan, K., Hay, S. I., & Brownstein, J. S. (2015). New digital technologies for the surveillance of infectious diseases at mass gathering events. Clinical Microbiology and Infection, 21(2), 134-140. https://doi.org/10.1016/j.cmi.(2014).12.017

[5] Kamadjeu, R., Tapang, E., & Moluh, R. (2005). Designing and implementing an electronic health record system in primary care practice in sub-Saharan Africa: a case study from Cameroon. Journal of Innovation in Health Informatics.

[6] Costa, F F. (2009). Enomics, epigenomics and personalized medicine: A bright future for drug development? Bio forum Eur.

[7] Ashwin Belle, Raghuram Thiagarajan, S. M. Reza Soroushmehr, Fatemeh Navidi, Daniel A. Beard, and Kayvan Najarian. Review Article: Big Data Analytics in Health Care. Hindawi Publishing Corporation.

[8] Ashrafi, N., Lori, K., & Jean-Pierre, K. (2014). The impact of business intelligence on healthcare delivery in the USA. Interdisciplinary Journal of Information, Knowledge, and Management.

[9] Mettler, T., & Vimarlund, V. (2009). Understanding business intelligence in the context of healthcare. Health informatics journal,

[10] Bonney, W. (2016). Applicability of Business Intelligence in Electronic Health Record. Social and Behavioral Sciences.

[11] Deaton, Angus S, and Robert Tortora. (2016). "People In Sub-Saharan Africa Rate Their Health And Health Care Among The Lowest In The World." Health Affairs

[12] https://www.facebook.com/EthiopiaFMoH/

[13] Technology and Innovation Institute Organizational Overview

[14] D. Yaramala, "HEALTH CARE DATA ANALYTICS USING HADOOP (thesis)," San Diego State University.

[15] Jan Willem Middleburg (2019), How to conduct A Big Data Maturity Assessment, white paper.

[16] Fern Halper and Krish Krishnan (2013-2014), TDWI Big Data Maturity Model Guide: Interpreting your assessment score, TDWI Research

[17] https://www.ibm.com/analytics/hadoop/big-data-analytics

[18] Akindele Akinnagbe, K.Dharini Ami Peiris & Oluyemi Akinloye (2018). Prospects of Big Data Analytics in Africa Health Care System. Global Journal of Health Science.

[19] Nahid Golafshani (2003). Understanding Reliability and Validity in Qualitative Research. The Qualitative Report.

[20] Abreham Kassahun (2018). A FRAMEWORK FOR PREDICTIVE ANALYSIS OF MALARIA DISPERSION. Addis Ababa University.

[21] Liang Hong, Mengqi Luo, Ruixue Wang, Peixin Lu, Wei Lu, Long Lu (2018). Big Data in Health Care: Applications and Challenges. Data and Information Management.

[22] Iwan Kurniawan Subagja, Nurhadifah Amaliyah, Ulpen Hiermy, Budi Tri Rahardjo, E. Laxmi Lydia, K. Shankar, Phong Thanh Nguyen (2019). Evaluation of Big Data Analytics in Medical Science. International Journal of Engineering and Advanced Technology (IJEAT).

[23] Mudassir Khan (2018). Big Data Analytics Evaluation. International Journal of Engineering Research in Computer Science and Engineering (IJERCSE).

[24] Sabyasachi Dash, Sushil Kumar Shakyawar, Mohit Sharma, and Sandeep Kaushik (2019). Big data in healthcare: management, analysis and future prospects. Journal of Big Data.

[25] Phaphama Kangelani and Tiko Iyamu. A Model for Evaluating Big Data Analytics Tools for Organization Purposes. Cape Peninsula University of Technology, Cape Town, South Africa.

[26] Kankanhalli A, Hahn J, Tan S, Gao G. Big data and analytics in healthcare: introduction to the special section. Inform Syst Front. 2016;18.

[27] Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inform Sci Syst. 2014.

[28] Wu PY, Cheng CW, Kaddi CD, Venugopalan J, Hoffman R, Wang MD. Omic and Electronic Health Record Big Data Analytics for Precision Medicine. IEEE Trans Biomed Eng. 2017.

[29] https://www.questionpro.com/blog/quantitative-research/

[30] https://www.scribbr.com/methodology/qualitative-research/

[31] Fitri Retrialisca, Umi Chotijah. The Maturity Measurement of Big Data Adoption in Manufacturing Companies Using the TDWI Maturity Model. Journal of Information Systems Engineering and Business Intelligence. Vol.6, No.1, April 2020

## Annex I

**Interview Questions**

The objective of this research is to assess the application of big data analytics in the health sector in Ethiopia. For this purpose, a prior discussion made with the primary actors related to the topic. Ministry of Health and the Ministry of Innovation and Technology is the primary actors that play their role in the health sector and in technology, respectively. The following questions are designed by reference of five crucial perspectives and will be highly assisting to understand the sector.

Thank you so much in advance for your time.

**Organization**

1) What is the Leadership structure for data management and big data analytics?
2) Is there any funding in the health sector to aid the big data analytics or any ICT enabled services?
3) Is there any strategy to assist the ICT sector, big data management, and application for enhancing the health sector?
4) How do you explain the Ministry of health working culture in relation to capable the sector through the deployment of advanced systems?
5) What is the Ministry's Vision, Mission, and Value?

**Infrastructure**

1) How do you explain the development and change of the health sector especially after the deployment of different ICT enabled services?
2) What are the currently existing systems to assist the health sector services?
3) Is there any available big data technology like Hadoop cluster, NoSQL database, or any deployed in the ministry to assist the big data analytics?
4) What type of architecture/ecosystem currently existing in the big data management structure?
5) Is any of the systems integrated with other related systems or how are the different systems operate?
6) What is the scope of the big data analytics application?

**Data Management**

1) Do you think there is big data in the health sector?
2) How do you explain this data's characteristics like variety, volume, velocity, and soon?
3) After the appropriate data is collected or received from a different system, how it is processed for further use?
4) What is the storage capacity for big data?
5) How do you explain the quality of the data?
6) Is the data or information being accessible in the minister or for other organizations?

**Analytics**

1) Are there any skilled experts for big data analytics?

2) In most of the time analyzed data put on the shelf in different organizations. How do you explain the health ministry’s experience? To what extent the organization gets to benefit from analyzed data?

3) What are the big data analytics techniques applied in the health sector?

4) After the ministry set data analysis techniques, how is the application seems?

5) How is the analyzed data delivered to all concerned bodies and how it participates in different experts?

**Governance**

1) Is there any policy, roadmap, or guideline for big data management and analytics? And how does the ministry govern it?

2) What is the Ministry of Health structure in managing big data?

3) Is there any steering committee to oversee the whole program of big data analytics and its application?

4) How do you define the stewardship structure dedicated to the big data analytics in health-related issues?

5) Is the minister concerned with answering questions such as, “Whose data, was it? Whose data is it? Where is it going? How long will it last?”

## Annex II

TECH-IN²
Technology and Innovation Institute

# Summary of Workshop on Big Data Analytics and Advanced Computing

## Given on October 15, 2019

This is a summarized feedback collected from the participants which are 26 in numbers. All the participants did not give the feedback.

1. The summary of the questions from agreement and disagreement perspective

| Questions | Strongly Agree | Agree | Disagree | Strongly Disagree |
|---|---|---|---|---|
| The workshop met my expectations. | 9 | 12 | 4 | 1 |
| This workshop increased my knowledge level in big data analytics and computing. | 9 | 15 | 2 | |
| The information and/or presentations were relevant and useful. | 12 | 14 | | |
| The presenters provide adequate time for questions and answers. | 20 | 6 | | |
| The length of the workshop was appropriate. | 9 | 11 | 5 | 1 |
| **Summary** | **59** | **58** | **11** | **2** |

2. Knowledge of the participants about Bid data before the workshop

| Have Knowledge | Intermediate Knowledge | No Knowledge |
|---|---|---|
| 14 | 8 | 4 |

3. Participants experience on Big data Projects

| Participated | Not Participated |
|---|---|
| 11 | 15 |

4. Prioritized sectors selected by participants placed **in order**
   i. Agriculture
   ii. Health
   iii. Transport
   iv. Education
   v. Satellite data, Climate modeling and weather forecasting

vi. Telecommunication
vii. Business, Finance and Investment
viii. Crime detection
ix. Government and public sector

5. **Best strategies** given by participants to enhance and implement big data analysis technology in Ethiopia.
   - ✓ Cooperation and work together among government organizations, universities, researchers and private sectors
   - ✓ There should be strong awareness creation and advocating tasks for higher government officials and targeted sectors on data importance
   - ✓ Capacity building on big data analysis, HPC
   - ✓ Building data storage infrastructure
   - ✓ Preparing data protection and access **policy**
   - ✓ Digitalizing and automating government offices to collect the data and facilitate services
   - ✓ Doing assessment on business and industry requirement so that it is possible to build the desired system
   - ✓ We have to be **IOT** user to collect data in advance
   - ✓ TECHIN should start the service and attract stakeholders
   - ✓ Use available and existing data as starting point

6. Given general perception given about the workshop
   - ✓ It was good and an organized workshop
   - ✓ It nice for beginners in data analysis task
   - ✓ It was an eye opening for researchers on big data and data science
   - ✓ TECHIN should keep on organizing another workshop in advance
   - ✓ The time given for the workshop was very short
   - ✓ It would have been better if there are practical and advanced techniques on HPC and actual big data analysis processes.
   - ✓ Hands on workshop should be included
   - ✓ Technologies and methods were not well discussed
   - ✓ Camera men disturbed the seminar