



# Challenges of Identifying and Utilizing Big Data Analytics in a Resource-Constrained Environment: In the Case of Ethiopia

Tigabu Dagne Akal, Tibebe Beshah, Stefan Sackmann, Solomon Negash

## ► To cite this version:

Tigabu Dagne Akal, Tibebe Beshah, Stefan Sackmann, Solomon Negash. Challenges of Identifying and Utilizing Big Data Analytics in a Resource-Constrained Environment: In the Case of Ethiopia. International Working Conference on Transfer and Diffusion of IT (TDIT), Jun 2019, Accra, Ghana. pp.234-254, 10.1007/978-3-030-20671-0\_17 . hal-02294717

**HAL Id: hal-02294717**

**<https://inria.hal.science/hal-02294717>**

Submitted on 23 Sep 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Challenges of identifying and utilizing Big Data Analytics in a resource-constrained environment: in the case of Ethiopia

Tigabu D.Akal<sup>1</sup>, Tibebe Beshah<sup>2</sup>, Stefan Sackmann<sup>3</sup>, Solomon Negash<sup>4</sup>,

<sup>1</sup>Addis Ababa University, Addis Ababa, Ethiopia, [tigabu.dagne@aau.edu.et](mailto:tigabu.dagne@aau.edu.et)

<sup>2</sup>Addis Ababa University, School of Information Science, Addis Ababa, Ethiopia, [tibebe.beshah@aau.edu.et](mailto:tibebe.beshah@aau.edu.et)

<sup>3</sup>Martin Luther University of Halle-Wittenberg, Institute of Information Science, Germany, [Stefan.sackmann@wiwi.uni-halle.de](mailto:Stefan.sackmann@wiwi.uni-halle.de)

<sup>4</sup>Kennesaw State University, Kennesaw, USA, [snegash@kennesaw.edu](mailto:snegash@kennesaw.edu)

## Abstract

Big data analytics (BDA) is the process of capturing and storing huge volume of data which has different formats and generated in high Velocity. It also refers to the process of analyzing big data for the purpose of decision making, strategic planning and policy formulation. Some of the applications of BDA include market segmentation, sales forecasting, weather forecasting, payment fraud detection, crop diseases detection, e-commerce analysis and users purchasing recommendation and others. The application of BDA is not only left for economically developed regions. It is also important for resource-constrained environments. In this study, challenges of identifying and utilizing big data analytics in the resource-constrained environment in the case of Ethiopia have been explored using some case. The case studies considered potential industries that can generate big data in Ethiopia. Ethiopian Telecommunication Corporation, Agricultural Transformation Agency, Payment systems like Hello Cash and Ethiopian Educational Networks (EthERNet) were considered as a case study. In the study, a qualitative grounded approach has been applied. Data was collected using a semi-structured interview approach. As data analysis result and discussion indicated that even if the selected potential industries have been generated big data they are not using it fully for the purpose of decision making. Potential challenges were identified in the identifying and utilizing of BDA in a resource-constrained environment. Some of these areas: lack of BDA awareness, data integration challenge, lack of skilled experts in the area, lack of data correctness and completeness, lack of standardized data registry, lack of leadership and management skill, issue of data privacy and infrastructure challenges including a huge volume of storage device constraint. Based on the identified challenges of BDA implementations in this study and possible application areas of BDA in those industries, a conceptual framework of the study were formulated.

## 1. Introduction

Big Data Analytics (BDA) is using large, diverse, and dynamic sets of user and machine-generated data as well as applying new methods of analytics to generate some interesting and prevailing knowledge [5, 20]. Data Analytics (DA) is the application of Business intelligence & analytics technologies that are based on data mining and statistical analysis [5, 38].

The application of big data (BD) and BDA in making an organizational data-driven decision has created a centre of attention over the past a few years. Some of the major areas of BD and analytics application areas are identified. Service providing divisions such as banking and finance, e-government and politics, smart health and wellbeing, telecommunication, information technology companies, security and public safety, science, technology and electronic-commerce are quickly adopted BD [5].

The application of BDA is not left only for high-income countries. It can be also highly applicable in low-income countries. But in low-income countries like Ethiopia, there is an issue of infrastructure development and know-how (understanding) of organizations on the application of BD, which are both low and hindering BDA availabilities in the organizations. The main motivation of this research is to identify and utilize the opportunities of BD and BDA application in the context of low-income countries like Ethiopia where there are different challenges like infrastructure. Through the research, the following research question were addressed: what are the challenges of identifying and utilizing BDAs in a resource-constrained environment? In this research, potential industries like Ethio-telecom, Agricultural transformation agency, payment systems (like hello cash) and education institutions were considered as a case study. The researcher has selected these cases studies due to their hugeness in producing data with different varieties and at a high-speed rate. In the meantime, these organizations are leading in the country on having a huge number of customers. The other reason behind considering these organizations is due to BDA analytics is flourishing or applying a lot in high-income countries in those industries which are addressed in the related work section.

Although big data has lots of benefits, there are some challenges related to storage and scalability issue of which “storage systems are not capable enough to store data” [1, 7, 12, 17]. Some scholars were addressed challenges related to lack of expert in the domain area, privacy and security challenge [12, 29, 39], data error [10] and challenge on representation of heterogeneous data [5, 17]. Other challenges are lie in data collection, storing, searching, analysis, sharing and visualization [1].

## **2. Related Works**

As addressed by some scholars BD and BDA have been used to address huge volume of datasets (from terabytes to exabytes) and complex datasets (captured from large-scale enterprise systems, online social graphs, mobile devices, internet-of-things and open data /public data) that they need highly developed storage management and apply systematic use of analytical applications and data representation technologies [3,5].

BDA is magnificently being used in different industries, such as telecom, environmental studies banking, education, insurance, social media user behavior identification [8, 26]. As addressed by [16] government industries use BDA to improve their capability to give services for the citizens by addressing different national challenges related to healthcare, job creation, economy, terrorism and natural disasters. For example, telecom companies have long had access to extensive bits of data with a large base of their subscribers connecting daily to their network and services

([28]. As addressed by [4] telecommunication industries are now capturing more and more data volume that is consumers are making more calls and connecting more and more to connect to the internet that is benefitting from a larger variety of sources as well as from higher velocity in data generation. Having these huge volumes of data helps companies to classify their customers' behaviors and usage patterns. Countries like China use big data to conduct predictive analytics and enhance their businesses in the banking industry by analyzing customers' behaviours through analytical modelling methodologies and techniques [25].

The other BDA application area is the agricultural sector. For the sake of addressing the current agricultural production challenges, there is a need for applying modern information technologies that help to monitor the physical environment continuously [13]. As addressed by Kamilaris and his colleagues applying modern agricultural digital technologies can produce huge quantities of data. Big data analytics technologies are applicable to the agricultural sector. For example, Hadoop and cloud-based analytics were applied in the crop identification with comparisons to prices in different seasons [22]. Hence, "big data analysis" is the term used to describe a new generation of practices [14, 24] designed so that farmers and related organizations can extract economic value from very large volumes of a wide variety of data by enabling high-velocity capture, discovery, and/or analysis [18, 27]. The application of BDA enable farmers, agricultural transformation agencies, and researchers companies to extract value-driven information from the huge collection that will enable agricultural productivity and environmental sustainability [13]. Data generated from the agricultural sector "big data" demands large investments in resource utilization including resources like a skilled expert, infrastructures for data capture, store and process [21]. The big data analysis needs to "*operate almost in real-time for some applications (e.g. weather forecasting, monitoring for crops' pests and animals' diseases)*" [11].

BDA application lies in the banking industries. The Intelligent Customer Analytics for Recognition and Exploration (iCARE) model coming with IBM software platforms has been addressed to analyze banking. As addressed by [25] the applications of the iCARE framework have been confirmed in a real case study of a bank in Southeast China. As indicated in the case study, iCARE helps generate insights for active customers based on their transaction behaviour, using close to 20 terabytes of data. Payment systems are generating huge datasets from customers in different services. But there are challenges in the application of BDA in the banking and payment systems. The challenge is related to the organizational level of handling BD for the application of BDAs. For example, smartphone users in the United Kingdom (UK) tend to do 220 tasks in every day and use their phone 1500 times per week versus only a few calls and short messages in the recent past [4]. There are also web-based applications which are encountering BD regularly, like "*recent hot spots social computing (including social network analysis, online communities, recommender systems, reputation systems, and prediction markets), Internet text and documents, Internet search indexing*" [5].

Higher education institutions need to capture and record students' academic record (tutorial data, registration data, courses data, assessment data, social activities, reading behaves, internet access inside in the university...), students behaving activities in the universities portal and social media [19]. But most of the higher education is facing

challenges in the application of BDA such as lack appropriate information infrastructure, data collection tools, automated software systems, skilled experts for effective and efficient data collection, data preparation tools for data cleansing, analysis and visualization of data [12]. As addressed by [19] some sources of higher education data are learning management systems, social media communication data, registration data, assessment data, student information data, employees' data, graduate data and others. *"Big Data can influence higher education practice, from enhancing students experience to improved academic programming, to more effective evidence-based decision making, and to strategic response to changing global trends"* [9]. But there are many implementation challenges of BDA in higher education. Some of these are data integration challenge, challenge associated with the quality of data collection and reporting and users' acceptance on the development of new processes and changing the management approach [9].

In the implementation of BDA, there are issues related to the uneven distribution of resources between the urban and rural regions that impact the development of BDA in different countries [30]. As discussed by [30], two major issues related to difficulty in the expansion of information infrastructure for the healthcare industry have been mentioned for low-income countries. There are also challenges related to data quality development of BDAs. Researches for data quality have been in the 1990s, and numerous researchers have been diverse explanations of the quality of data and distribution approaches of quality measurements [31]. The total Data Quality Management group of the Massachusetts Institute of Technology University directed by Professor Wang has done deep research in the data quality area [30]. As addressed by Wang and Strong "data quality" means "fitness of use" and argued the quality of data highly related to data from customers. As addressed by [30, 31] for uniformity data use in the healthcare industry, data should be of high quality and available so that different stakeholders are confidence to provide data-driven informed decisions making.

In addition to data quality and infrastructure challenges, organizational readiness and understanding what BD is recognized as one of the most steps prior to implementation and an important prerequisite to the achievement of BDA in relations of acceptance rate. Readiness assessment, as a comprehensive measure in order to provide a proper image of existing conditions and the preparedness of healthcare organization to change, is also a way to identify the potential cause of failure in innovation such as organizational resistance [32].

The remaining section of this paper addressed Research design in section three, data analysis in section four, discussion of the analysis result in section five and the concluding remark of the study in section six.

### **3. Research Design**

**3.1. Research Methodology:** Qualitative research approach is employed to investigate the problem mentioned above. As a result preliminary inputs (concepts) to a potential theory has been developed from collected data and analyzed from different management and other employees of the company using the semi-structured interview approach. The grounded theory was undertaken in this research. Because the nature of the research seeks to understand and explains social phenomenon or process within the context of a given application and it demands to

approve the validity of the theory from the point of those employees of the company who participated within the context during the study [14]. A summary of guidelines for information system researchers applied in the research of BDA, which includes research question, data collection, data analysis and result interpretation [20]. As discussed by Muller and his colleagues BDA researches are comfortable with data instead of theory and position either predictive or explanatory research methodology. In this research project, explanatory research methodology is selected because the researchers require the development of BDA concepts or generalizable properties associated with the challenges of identifying and utilizing BDA in a resource-constrained environment.

**3.2. Research Paradigm:** The interpretive research paradigm applied in terms of data collection and data analysis in this study. Because interpretive research paradigm can help Information system (IS) researchers to recognize human thinking and action in social and organizational contexts; it has the potential to construct thoughtful insights into information systems happenings [33]. A new conceptual framework derived from data that collected and analyzed from different stakeholders using the semi-structured interview approach. The grounded theory undertaken in this research. Because the nature of the research sought to understand and explains social phenomenon or process within the context of a given applications and it demanded to approve the validity of the justifications from the point of those stakeholders who participated within the context during the study [15].

**3.3. Data Collection and analysis approach:** The research instrument consisted of different sets of semi-structured interview guides and codes for thematic analysis. For data collection, semi-structured interviews which area open-ended and process oriented applied [34]. Primary data comprised of semi-structured interviews obtained from respondents of Ethiopian Telecommunication Corporation (ETC), Agricultural Transformation Agency (ATA), Payment Systems (Hello Cash), Ethiopian Educational Networks (EthERNet) and Independent Researcher.

During data collection, a method of purposive sampling was used to identify the sample group that have been participated in the interview. It is purpose because the interview participants were identified based on their knowledge, relationship and expertise on BDA. Table 1 below shows the different categories who were sampled, including the number of interviewees.

Table 1: Interview Sample per Category

Stakeholder Category	Semi-Structured Interviews
Independent Researcher	1
ETC Customer Operation Manager	1
Hello Cash Data Analyst	1
ATA Data Analyst	1
EthERNet Big Data Researcher	1

There were different types of interview guides depending on the participant's role. The face to face interview range between 12 minutes to 19 minutes. The collected data was analyzed thematically following from [35, 37].

#### 4. Analysis result and findings

This section presented the process and results of data analysis in order to identify patterns to come up with conceptual framework. The seven principles of interpretive research method suggest by [33], selected principles applied in this research with evidence. Since the type of data that collected is context-rich qualitative data and analysis also was qualitative (e.g., coding or content analysis). The following steps are adapted from [35, 37].

##### 4.1. Developing the Code Manual

The coding manual a data management approach that helps to organize portions of similar or related text and to come with a new understanding from the evolving patterns [37]. This was designed prior to analysis and was based on the research question. As demonstrated in table 2, the code manual is a data management tool that assists in organizing portions of similar or related text and deriving new insights from the emerging patterns [37]. It consists of the name, definition of what the theme concerns, and description of how to know when the theme exists. Researchers know to code in the qualitative analysis of interview data is also focus on something which is repeated in several phases, something that the interviewee explicitly states that it is important, something which is similar in the previously published reports and something which reminds a theory or concept [36, 37].

Table 2: Code Manual

No.	Code	Definition	Description
1	Big data	Data in different formats including text, audio, video, log files: both structured and unstructured. Its volume is huge and increasing with high velocity and different varieties.	Availability of big data in the organizations
2	Big data analytics	Process of data collection, storing and analyzing huge datasets (big data) in order to discover patterns and important information for researching and decision making.	Availability of big data analytics tools
3	Big data analytics and Telecommunication	This indicates that the impact of big data analytics in the telecommunication industry for customer relationship management, sales forecasting, marketing, fraud detection, security, intrusion detection and others.	Decision makers, policy makers and researchers applying big data analytics for the telecommunication sectors
4	Big data analytics and Agriculture	This indicates that the impact of big data analytics in the agricultural industry for weather forecasting, crop diseases detection, productivity forecasting, accurate crop predictions and policy making.	Decision makers, policy makers and researchers applying big data analytics for the Agricultural sectors
5	Big data analytics and Payment systems	This indicates that the impact of big data analytics in the payment systems for customers segmentation and	Decision makers, policy makers and researchers applying big data analytics for the Payment sectors

identifying customers product buying behaviors.

6	Big data analytics and Education	This indicates that the impact of big data analytics in the education sector for forecasting students behaviors, security of institutions prediction,	Decision makers, policy makers and researchers applying big data analytics for the Education sectors
7	Big data analytics and Lack of awareness	This is indicates that lack of awareness is one of the challenge on the application of big data analytics in organizations.	Policy makers, decision makers, top managements, data collectors and users have lack of awareness on the application of big data analytics.
8	Big data analytics and leadership	This indicates that lack of leadership skill is one of the challenges on the application of big data analytics in the organizations.	Existence of Lack of leadership skill in the organization can hinder in the big data analytics development.
9	Big data and Lack of cheaper data collection tools	This indicates that lack of cheaper data collections tools for collecting different data: images, text, sensor data, logs and many other.	Availability of cheaper data collection tools
10	Big data analytics and lack of qualified experts in the domain	This indicates lack of skilled expert has a negative impact on big data analytics application	Availability of big data analytics experts
11	Big data analytics and infrastructure challenge	This indicates that lack of stabled infrastructure hinders the application of big data analytics in an organizations	Availability of information communication technology infrastructure
12	Big data analytics and lack of modern data management	This indicates that Modern Management is required for successful application of big data analytics	Availability of Modern Management approaches'. Interviewees were asked whether modern management approach has been implied or not
13	Completeness and Correctness of data	This indicates that attributes of data.	Participants perception about capturing and storing data by focusing on correctness of data they used
14	Data privacy Concern	Data privacy is one of the major concerns in big data when users are generating data and shared it.	Data privacy need an attention for the implementation of big data analytics as addressed by the interviewees.
15	Audio, video, texts, CSV files, image and others.	This indicates data formats in the big data	Interviewees were dealt the different data formats that are generating by the users and machines in the organization
16	Big data analytics and Lack of standard in data registry	This indicates that compatibility/consistence of data formats	No standard in the data production



### Testing the reliability of codes

In this phase the applicability of the codes were tested as showed in table 3. Using **Atlas.ti**, how the codes was applied to the selected transcripts interview were checked.

Table 3: Reliability of Codes

Code	Data from Transcripts
Big data analytics and Lack of awareness	<p>The most important or first rank challenge from my experience is having awareness on the usage of big data.</p> <p>There is no awareness about the value of big data in our organizations. But there huge datasets already existed. In every day more than a Tera bytes of data are captured and recorded. When I look other countries' telecom sector they are using big data for different applications /business innovations in the industry.</p> <p>I think the fundamental challenge is related to existence awareness. I am considering awareness as the first challenge.</p>
Big data analytics and leadership	<p>There is also lack of leadership and management capabilities of people. Leaders and managers of organizations' used to think that there is no enough data for decision-making. So our country should work on knowledge capabilities of mangers and leaders of organizations.</p> <p>It is possible to bring traditional management approach in to the big data management but management of compartments, access policy management, data management and other issues management need skilled person so that the design need lots of time.</p>
Big data analytics and Lack of standard in data registry	<p>The big issue is due to lack of standard in data registry. There is no standard in the data production. For example, in our organization regions, zones and woredas are producing data. But there is huge problem in compatibility/consistence and data formats. There are many missing values or incomplete data when they are capturing.</p> <p>Every organization should have its own standard in data handling and collection. Data with its own time stamps and temporal data must be enforced. The other thing the data collection approach should digital / online mechanism</p>
Big data analytics and lack of qualified experts in the domain	<p>Resources availability like skilled professionals in the area will be faced as a challenge in the modern data management or big data analytics implementation.</p> <p>The 2<sup>nd</sup> challenge is related to knowledge gap. There are huge skill gap on the knowhow of big data applications and tools.</p> <p>As we discussed previously, currently our organization is processing data through Excel. It is challenging to process big data using Excel. This indicates that shortage of skilled manpower is challenging for big data analytics application in our organization. I think this is the same challenge in other organizations in our country. So lack of skilled manpower and big data production standardization are the main challenge</p>

in our organization might be the same throughout the country.

Our country should work on knowledge capabilities of managers and leaders of organizations. In the meantime, the philosophy of big data analytics should come to our country. Because big data analytics can be applied a lot in the area of agriculture, traffic management, health sectors, payment systems and other organizations that can generate huge volume datasets from sensors and social media.

Completeness and  
Correctness of  
data

The second challenge is having very low quality datasets. For example when customers are purchasing products or applying for some services, they should be registered in a given form with correct data sets. Some lefts empty, some fills with unrelated issues with answer to be filled for a specific questions, some filled text data for answer that required to be filled with numeric data. The form includes information related to customer information and other issues. The second challenge also raised because of lack of awareness on the value of data. That means the huge datasets that captured from some customers might be a garbage. If awareness is coming in to the board the data quality issue will be solved.

The image quality that used to capture by data collectors is very low and challenging to identify potential information of customers. Do to image quality too much information of customers' are not registering properly. Therefore, there is information inconsistency.

Big data analytics  
and infrastructure  
challenge

The third challenge is related to infrastructure especially storage capabilities. Our organization used to store maximum of 6 months for most of the users is due lack of having huge storage devices.

Well, big data starts with cheaper data collection mechanisms. When you look at these things, they are currently in their infancy in our country. Although it getting better, low data collection will always hinder us from using big data analytic system.

Even if we can collect data, storage is another bigger issue. Building a reliable, ever increasing, secured data center is not an easy task.

Infrastructure is challenging. The government should work on the designing of central infrastructure that can be used by many institutions at the same time. Building the infrastructure is very expensive. Imagine every company is working to have its own sophisticated data center how much it costs.

Big data requires huge storage infrastructure. The volume in the 3V's shows that the volume of the data is immensely large when compared to the traditional database systems. It's often in exabytes because you store the raw uncompressed and unprocessed data.

Audio, video, texts, CSV files, image and others.

Audio, text, SMS, data log files and others. All data from customers are used to capture and store for some durations. Usually from 3 to 6 months. The limit of 3 to 6 months for most users is due to lack of storage devices.

They are audio, video, texts, CSV files, image and others.

Well the core part of big data is that the 3 Vs (volume, variety and Velocity) and the Variety term implies that big data often doesn't have a single format. It can be composition of images, text, sensor data, logs and many other.

Big data analytics and organization

In our organization big data analytics can be used for marketing, network intrusion detection analysis, fraud detection and analysis, operation and maintenance, sales forecasting. For example in our organization there are more than 10, 000 telecom sites and usually it is difficult to trace which site has a problem and what kinds of problem will be happen at each site. Our organization has more than 65 million mobile customers, more than 17 million internet and data customers, over 1.2 million fixed line customers.

Currently our organization has working on a big data warehouse project which is not yet completed.

---

### Summarizing data and identifying initial themes

In this phase the interview transcripts were summarized. The researcher went through all the process of reading notes taken during interview, listening the recorded file and summarizing dataset as showed in table 4.

Table 4: Code Summary

Research Questions	Summary of Responses
Challenges of identifying and utilizing BDAs in a resource constrained Environment:	Lack of awareness on the usage of big data
	Lack of leadership and management capabilities of people in the big data analytics.
	Lack of Modern data management approach.
	Lack of standard in data registry by the organizations.
	Resources availability like skilled professionals in the area of big data analytics.
	Lack cheaper data collection tools both hardware and automated software availability.
	Data quality is very low to run big data analytics
	Infrastructure challenge especially the network infrastructure and storage capabilities.
	Big data is available in some organizations like telecom industry, payment systems and agricultural industry.

---

Big data analytics can be applicable for marketing, network intrusion detection analysis, fraud detection and analysis, operation and maintenance, sales forecasting, weather forecasting, strategy planning and policy analysis, project monitoring and evaluation, customer behavior analysis.

### Applying template of Codes and additional Coding

Using the illustration methods [37], the researcher used the codes from the code manual to the interview transcripts with the sake of identifying meaningful portions of the text as showed in table 5.

Table 5: Sample Codes from the initial code manual

Code	Description of Code	Matching text
Big data analytics and Lack of awareness	Policy makers, decision makers, top managements, data collectors and users have lack of awareness on the application of big data analytics.	“There is no awareness about the value of big data in our organizations. But there huge datasets already existed. In every day more than a Tera bytes of data are captured and recorded. When I look other countries’ telecom sector they are using big data for different applications /business innovations in the industry”
Big data analytics and infrastructure challenge	Availability of information communication technology infrastructure	“Infrastructure is challenging. The government should work on the designing of central infrastructure that can be used by many institutions at the same time. Building the infrastructure is very expensive. Imagine every company is working to have its own sophisticated data center how much it costs”
Completeness and Correctness of data	Participants perception about capturing and storing data by focusing on correctness of data they used	“The image quality that used to capture by data collectors is very low and challenging to identify potential information of customers. Do to image quality too much information of customers’ are not registering properly. Therefore, there is information inconsistency”

Based on table 2 to 5, the Groundedness and density of each considered using Atals-ti qualitative analysis tool. Table 6 below shows the Groundedness and density. Groundedness refers to total number of quotations linked to a given code. Density is the proportion of documents that link to a given code

Table 6: Code Groundedness and Density

No.	Code	Groundedness	Density
1	Big data	15	15
2	Big data analytics	12	11
3	Big data analytics and Telecommunication	5	4
4	Big data analytics and Agriculture	5	4
5	Big data analytics and Payment systems	6	5

6	Big data analytics and Education	7	3
7	Big data analytics and Lack of awareness	15	5
8	Big data analytics and leadership	5	1
9	Big data and Lack of cheaper data collection tools	7	1
10	Big data analytics and lack of qualified experts in the domain	15	6
11	Big data analytics and infrastructure challenge	17	13
12	Big data analytics and lack of modern data management	3	1
13	Completeness and Correctness of data	17	12
14	Data privacy Concern	1	1
15	Audio, video, texts, CSV files, image and others.	12	5
16	Big data analytics and Lack of standard in data registry	3	1

#### 4.2. Connecting the codes and identifying themes

Connecting codes is the process of discovering themes and patterns in the data [37]. Based on the research question, the main task of this phase is connecting the codes and identifying themes which is shown in below table 7, 8 and 9.

##### Definition of big data

Table 7 shows that the definition of big data from different viewpoints.

Table 7: Definition of Big data by different organizations and researcher

Group	Definition of big data
Education policy makers	It is a huge datasets in different formats. It is different from data which generated from traditional databases
Telecom Customer Operation manger	Huge datasets generated by the customers, machines with high speed and different formats
Independent Researcher	The core part of big data is that the 3 Vs (volume, variety and Velocity) and the Variety term implies that big data often doesn't have a single format. It can be composition of images, text, sensor data, logs and many other.
Data Analyst from Agricultural Transformation agency	Data in the form of audio, video, texts, CSV files, image and others and can be collected from different sources frequently. It can be also generated by people from projects
Data Analyst from Payment system	Data that that is generated by many users frequently with huge volume and can be generated repeatedly by the same customer.

##### Benefits of big data analytics

Table 8: Big data application in different organizations

Group	Benefits of big data analytics
Education Institution	It is important for decision making in the educational transformation. The data generated every day from higher educational institutions can help for decision making

	in the area of intrusion detection, students behavior on social network access, students performance and teachers performance. The data generated by the students from the web portal of each higher education can enable decision maker to give fast and reliable decision at any time.
Telecommunication industry	It is important for marketing, sales forecasting, customer behavior identification or customer segmentation, improve customer experiences, forecast network capacity and demand faster, fraud detection, for real time customer insight and
Researchers	It is important for theoretical contributions in different disciplines.
Agricultural Transformation agency	For weather forecasting, for crop diseases identification, environmental prediction, for project monitoring and evaluation, for country level strategic planning and policy design. It is based on the organization accumulated and daily generated big data.
Payment Systems	It is important to classify or cluster customers based on their purchasing behaviors on different application or services. It is also important for payment fraud detection and managing risks and for customer transaction activity tasks.

### Challenges of big data analytics implementation by different organizations

Table 9: Challenges of big data implementation by different organizations

Group	Challenges of big data analytics implementation
Education Institution	<ul style="list-style-type: none"> <li>- Some students provide incomplete data</li> <li>- There is a challenge on awareness</li> <li>- Lack of skilled manpower</li> </ul>
Telecommunication industry	<ul style="list-style-type: none"> <li>- There is a challenge on awareness</li> <li>- Lack of integrated systems that handle all rounded information for a given customer.</li> <li>- Infrastructure including storage device challenge</li> <li>- Data privacy challenge</li> <li>- Security is a sensitive issue to innovate the field in the organizations</li> </ul>
Researchers	<ul style="list-style-type: none"> <li>- Lack of Modern Management on big data analytics</li> <li>- Lack of cheaper hardware devices for data collection</li> <li>- There is huge gap in software system automation. And difficult to think about big data before software automation.</li> <li>- Difficult to access data from organizations to make researches in the field.</li> <li>- Lack of qualified experts in the field.</li> </ul>
Agricultural Transformation agency	<ul style="list-style-type: none"> <li>- Infrastructure challenge</li> <li>- Lack of standardized data registry</li> <li>- Low image quality generated by agricultural extension workers which is difficult for analysis and decision making</li> </ul>

- Lack of qualified experts in the field.

#### Payment Systems

- Customer provide incomplete data that is issue of data quality
- Having poor infrastructure
- Lack of modern Management on the big data analytics

### 4.3. Corroborating and legitimating coded themes

This is the final stage for the further identification of themes to be clustered from the coded text in to 5 clustered themes as showed in table 10.

Table 10: Corroborating and legitimating coded themes

First Ordered Theme	Clustered Theme	Second Order Theme
Big data analytics for decision making in the educational transformation	Theme 1: Lack of Resources availability is a challenge for big data analytics implementation	Accessibility of big data
Big data analytics for marketing, sales forecasting and customer behavior identification	Theme 2: Data correctness and completeness is a challenge for big data analytics implementation.	Positive impacts of big data analytics
Big data analytics for payment fraud detection		
Big data analytics for weather forecasting and environmental prediction	Theme 3: Lack of awareness on the application of big data analytics is a challenge for big data analytics implementation.	Challenges of identifying and utilizing big data analytics
Big data analytics to manage risks and for customer transaction activity tasks.		
Data incompleteness is challenge for big data analytics	Theme 4: Lack of leadership and management skill is challenge for big data analytics implementation	
Low image quality is challenge for big data analytics implementation		
Lack of qualified experts is challenge for big data analytics implementation	Theme 5: Big data analytics is important for decision making, marketing, sales forecasting and weather forecasting.	
Lack of cheaper data collection tool is challenge for big data analytics implementation	Theme 6: Big data analytics for customer behavior identification, customer transaction activity tasks and payment fraud detection.	
Lack of awareness on the application of big data analytics is challenge for big data analytics implementation		
Lack of leadership and management skill is challenge for big data analytics implementation		

### 4.4. Challenges of identifying and utilizing BDAs interrelationship

The following figure shows that the interrelationship of challenges of identifying and utilizing big data analytics in a resource constrained environment. It indicates the challenges of implementation and application of big data analytics interrelationship.

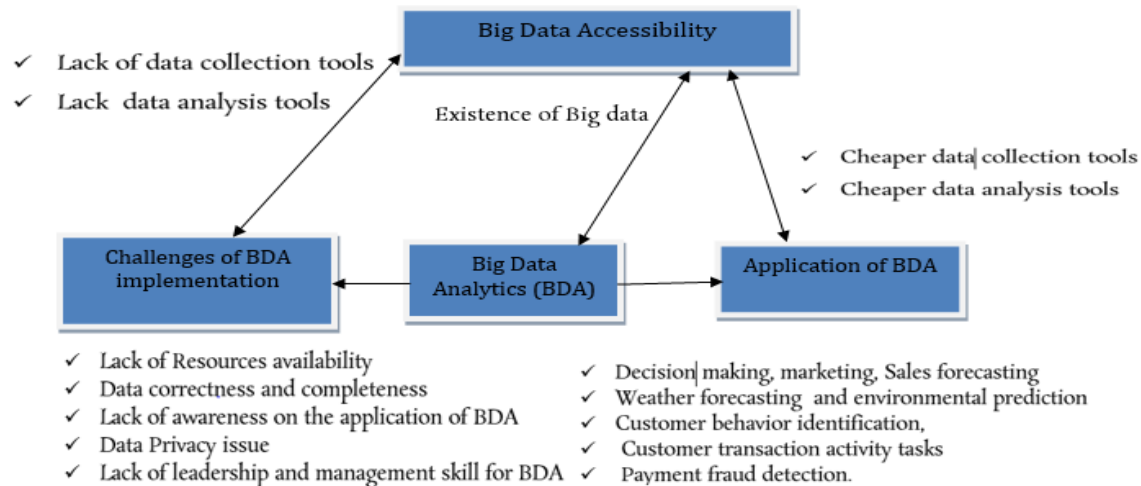


Figure 1: Utilizing BDAs interrelationship on a resource constrained environment

As shown in figure 1 above shows an illustration of the dependencies and feedback loop that exists between the application of big data analytics and challenges of big data analytics implementation in the resources constrained environments in case of Ethiopia.

## 5. Discussions

The analysis of the semi-structured interview result indicated that all the four participated organizations are producing a huge volume of data, at alarming velocity and different varieties. The independent researcher who is participated in the interviewing process indicated that there are many industries in Ethiopia which are producing big data. However, all the organizations which are participated in the study do not have proper data capturing and handling procedures.

For example in the telecom industry, the interviewee addressed the following issue: “All data from customers are used to capture and store for some duration. Usually from 3 to 6 months. After that, it can be deleted automatically. For some selected users it can be captured and stored for an unlimited period of time. The limit of 3 to 6 months for most users is due to lack of storage devices and not using the datasets for other purposes in most scenarios. Even we look only 994 (call support to help all customers) and 980 (call support to help VIP customers- enterprise companies). For example, 994 customer support service has more than 4,000 employees. One employee can give support for 250 to 260 customers. Averagely more than 1 million customers used to call for asking help in every day. Different kinds of request”. This indicated that even if they are generated a huge volume of data from customers they are not storing and applying BDA due to lack of infrastructure.

If we look at the payment system which is participated in this interview, its customers are requested many services. The organization has approximately one million customers. As the interviewee indicated that “Customers can pay their electricity and water bill. By the way hello cash includes other major businesses services like hello sera, hello



*doctor and other customer support packages are available. Hello Cash also interconnected with the metro-taxi systems of Addis Ababa. That is a customer can pay money from their Hello Cash account and can get taxi services. They have also deal with Selam Bus Transport Company so that customers can buy tickets from Selam Bus using their Hello Cash account. Customers can also buy mobile cards/ charge money their mobile at anytime from anywhere from their Hello Cash account”.*

The agricultural transformation agency (ATA) is also collecting and storing a huge volume of datasets. As the interviewee addressed the organization is collecting data for decision making, policy and strategic planning document formulation. The sources of data sets are: *“Our main source is CSA (Central Statistics Agency), World Trade Organization and data generated by the team inside the organization through interview or field visit. Internal teams are producing data during project implementations from farmers and agricultural extensions”.*

Even if organizations are generating a huge volume of data with different varieties they are not applying big data analytics. The following are the challenges that were addressed by the interviewees. Most of the challenges are related to lack of awareness, lack of skilled experts, data incompleteness, lack of stabled infrastructure, lack of leadership and management. For example, the interviewee from telecom industry addressed that “The most important or first rank challenge from my experience is having awareness on the usage of big data. There is no awareness of the value of big data in our organizations. But their huge datasets already existed. In every day more than a Terabytes of data are captured and recorded. When I look at other countries’ telecom sector they are using big data for different applications /business innovations in the industry. The second challenge is having very low-quality datasets. For example, when customers are purchasing products or applying for some services, they should be registered in a given form with correct data sets. Some lefts empty, some fills with unrelated issues with an answer to be filled for specific questions, some filled text data for an answer that required to be filled with numeric data. The form includes information related to customer information and other issues. The second challenge also raised because of lack of awareness on the value of data. That means the huge datasets that captured from some customers might be garbage. If awareness is coming into the board the data quality issue will be solved. The third challenge is related to infrastructure especially storage capabilities. Our organization used to store a maximum of 6 months for most of the users is due to the lack of having huge storage devices. There is also a lack of leadership and management capabilities of people. Leaders and managers of organizations’ used to think that there is not enough data for decision-making. So our country should work on the knowledge capabilities of managers and leaders of organizations”.

The interview result from the educational institution addressed the challenges: “I think the fundamental challenge is related to existence awareness. I am considering awareness as the first challenge. The 2nd challenge is related to the knowledge gap. There is a huge skill gap in the know-how of big data applications and tools. The 3rd challenge is related to the preference of commercial big data tools which are too expensive, not focusing on open source technologies. Another challenge is Infrastructure is challenging”.

The other challenge addressed by the agricultural transformation agency is related to lack of qualified expert and lack of data processing tool. The interviewee addressed that: “Currently our organization is processing data through

Excel. It is challenging to process big data using Excel. This indicates that the shortage of skilled manpower is challenging for big data analytics application in our organization. I think this is the same challenge in other organizations in our country. So lack of skilled manpower and big data production standardization are the main challenge in our organization might be the same throughout the country. The other big issue is due to the lack of standard in data registry. There is no standard in data production. For example, in our organization regions, zones and wordas are producing data. But there is a huge problem in compatibility/consistency and data formats. There are many missing values or incomplete data when they are capturing”.

Interviewees from the different organization have provided different recommendations that have to put for the big data analytics implementation in their organization and in the country level. For example, the interviewee from the telecom industry addressed that: “our country should work on knowledge capabilities of managers and leaders of organizations. In the meantime, the philosophy of big data analytics should come to our country. Because big data analytics can be applied a lot in the area of agriculture, traffic management, health sectors, payment systems and other organizations that can generate huge volume datasets from sensors and social media”.

The interviewee from the educational institution was addressed what their organization or the country should work on the big data analytics in the future: “The government should work on the designing of central infrastructure that can be used by many institutions at the same time. Building infrastructure is very expensive. Imagine every company is working to have its own sophisticated data centre how much it costs”. Some of the challenges were identified in this study were also addressed as a challenge by other scholars in the field. Challenges related to data correctness and data privacy [1, 2, 12]. Some of challenges like having lack of awareness of the application of BDA, lack of resource availability, lack of leadership and management skill were identified as major challenge in this research.

## **6. Conclusion**

In conclusion, big data analytics has numerous application areas in the telecom industry, payment systems, agricultural transformation agency and educational provider companies. The study indicated that all these organizations are generating a huge volume of data with different varieties and speed velocity. However, none of the organizations is applied not big data analytics for decision making, policy formulation and for future strategic planning. Rather they are using manual data analysis and simple office package software for data filtering in some situations. The challenges behind not applying big data analytics by these organizations were identified in the study. The major challenges were identified in the study. These are:

- Organizations have not their own standard in data handling and collection,
- Lack of awareness in the application of big data analytics,
- Lack of management and leadership in the area of big data analytics,

- Lack of qualified experts in the area of big data analytics,
- Lack of data correctness and completeness,
- Lack of stable infrastructure availability: there are no cheap data collection tools and software automation is still behind the current digitalization era.

Even if there are many existing challenges organizations should have come with their own integrated data storage system, cheap data collection tools. In the meantime, the government should give strong emphasis on the application of big data analytics at each organization which are generating a huge volume of data, with different varieties and speedy velocity. The government should also give emphasis on the skilled experts' development for big data analytics. Higher education should give emphasis to the hottest and highly applicable field, big data analytics. New curriculums and should be enforced in the area of big data analytics. Future research should give emphasis on the design and development of cheaper data collection and analysis tools which are more customer services oriented model.

## 7. Reference:

1. Assunção, D., Calheiros, N., Bianchi, S., Netto, S., Buyya, R.: Big Data computing and clouds. Trends and future directions, *Journal of Parallel and Distributed Computing* 79(80), 3-15 (2015).
2. Ahrens, J., Hendrickson, B., Long, G., Miller, Ross, R., and Williams, D.: Data-intensive science in the US DOE: case studies and future challenges. *Computing in Science and Engineering* 13 (6), 14–23 (2011).
3. Baesens, B., Bapna, R., Marsden, J.R., Vanthienen, J., and Zhao, J.L.: Transformational Issues of Big data and Analytics in Networked Business. *MIS Quarterly* 40 (4), 807- 818 (2016).
4. Bughin, J.: Reaping the benefits of big data in telecom. *Journal of Big Data* 3(16), 1- 17(2016).
5. Chen, H., Chiang, R., & Storey, V.: Business intelligence and analytics: From big data to big impact. *MIS Quarterly* 36(4), 1165–1188 (2012).
6. Chen, M., Mao, S., Liu, Y.: Big data: A survey. *Mobile Networks and Applications* 19(2), 171-209(2014)
7. Chen, P. and Zhang, C.: Data-intensive applications, challenges, techniques and technologies. A survey on Big Data. *Journal of Information Sciences*, (2014).
8. Cooper, J., Jones, C., Kahn, E. and Arbuckle, P.: Big data in life cycle assessment. *J. Ind. Ecol* 17, 796–799 (2013).
9. Daniel, B.: Big Data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology* (2014).
10. Fisher, D., DeLine, R., Czerwinski, M., Drucker, S.: Interactions with big data analytics. *Interactions* 19 (3), 50-59(2012)
11. Hashem, I. Yagoob, Anuar, Mokhtar and Gani: The rise of “big data” on cloud computing: review and open research issues. *Information System* 47, 98–115 (2015).
12. Kaisler, S., Armour, F., Espinosa, J. A., & Money, W.: Big Data: Issues and Challenges Moving Forward (2013).

13. Kamilaris, A., Kartakoullis, A., and Prenafeta-Boldú, F.: A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture* 143, 23–37(2017).
14. Kempenaar, C.; Lokhorst, C.; Bleumer, E.J.B.; Veerkamp, R.F.; Been, Th.; Evert, F.K. van; Boogaardt, M.J.; Ge, L.; Wolfert, J.; Verdouw, C.N.; Bekkum, Michael van; Feldbrugge, L.; Verhoosel, Jack P.C.; Waaij, B.D.; Persie, M. van; Noorbergen, H.: *Big Data Analysis for Smart Farming*. Wageningen University & Research 655(2016).
15. Kempster, S. & Parry, K. (2014), “Critical realism and grounded theory” in P. Edwards, J. O’Mahoney & S. Vincent (Eds.) *Studying Organizations Using Critical Realism: A Practical Guide* (pp. 86-108), Oxford, United Kingdom: Oxford University Press
16. Kim, H., Trimi, S., Chung, J.: Big-data applications in the government sector. *Communication. ACM* 57 (3), 78–85(2014).
17. Li, H., Lu, X.: Challenges and trends of big data analytics. In: *Ninth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC)*, Guangzhou China, 566-567(2014)
18. Lokers, Knapen, Janssen, Randen and Jansen: Analysis of Big Data technologies for use in agro-environmental science. *Environ. Model. Software* 84, 494–504(2016).
19. Minimol A.: An Efficient Way of Applying Big Data Analytics in Higher Education Sector for Performance Evaluation. *International Journal of Computer Applications* 180 (23), 2018.
20. Müller, O., Junglas, I., Brocke, J., and Debortoli, S.: big data analytics for information systems research: challenges, promises and guidelines. *European Journal of Information System* 25(4), 1-14(2016).
21. Nandyala, C., Kim, K.: Big and Meta data management for U-agriculture mobile services. *Int. J. Software Eng. Appl. (IJSEIA)* 10, 257–270(2016).
22. Ravisankar, k., Sidhardha, k., and Prabadevi, B.: Analysis of Agricultural Data Using Big Data Analytics. *Journal of Chemical and Pharmaceutical Sciences* 10, 2017.
23. Seddon, J., and Currie, W.: A model for unpacking big data analytics in high-frequency trading. *Journal of Business research* 70, 300–307(2017).
24. Sonka, S. (2016) “Big data: fueling the next evolution of agricultural innovation”, *J. Innovation Manage.* Vol. 4 (1), pp. 114–136 (2016).
25. Sun, N., Morris, J., Xu, J., Zhu, M. and Xiek.: iCARE: A framework for big data-based banking customer analytics. IBM, 2014.
26. W. Bates, D., Saria, S., Ohno-Machado, L., Shah, A. and Escobar, G.: Big Data in Health Care: Using Analytics to Identify and Manage High-Risk and High-Cost Patients. *Health Affairs* 33(7), 1123–1131(2014).
27. Waga, D., Rabah, K.: Environmental conditions’ big data management and cloud computing analytics for sustainable agriculture. *World J. Comput. Appl. Technology* 2 (3), 73–81(2014).
28. Yongjun H., Ming F., Shengyong D., and Yongbing F.: Big Data Development Strategy for Telecom Operators” *Telecommunication Science* 3(2), 2013.
29. Youssef, G., Mouhcine, G., and Hussein, T.: Big Data Analytics: Security and Privacy Challenges. *IEEE Symposium on Computers and Communication (ISCC)*, 2016
30. Brra, J., Hanseth, O., and Heywood, A.: Developing Health Information Systems in Developing Countries: The Flexible Standards Strategy. *MIS Quarterly* 31(2): 381-402(2007)
31. Wang, R. and Strong, D.: What Data Quality Means to Data Consumers. *Journal of Management Information Systems* 12 (4), 5-33(1996).

32. Ghazisaeidi, M., Safdari, R., Torabi, M., Mahboobeh, M., Jebraeil, F., and Azadeh, G.: Development of Performance Dashboards in Healthcare Sector: Key Practical Issues. *Journal of Academy of Medical Science of Bosnia and Herzegovina* 23(5), 317-321(2015).
33. Klein, H., and Myers, M.: Set of principles for conducting and evaluating interpretive field studies in Information systems. *MIS Quarterly*, 1999
34. Walsham, G.: Interpretive case studies in IS research: nature and method. *European Journal of Information System* 4(2), 74-81 (1995).
35. Boyatzis, R.: Transforming qualitative information: thematic analysis and code development. Thousand Oaks, CA, US: Sage Publications, Inc.1998
36. Fereday and Muir-Cochrane: Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development. *International Journal of Qualitative Methods*
37. Crabtree, B., and Miller, W.: A template approach to text analysis: Developing and using codebooks. *Doing qualitative research* Newbury Park, CA: Sage., 163-177 (1999)
38. Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data—evolution, challenges and research agenda. *International Journal of Information Management*, 48, 63-71.
39. Dwivedi, Y. K., Janssen, M., Slade, E. L., Rana, N. P., Weerakkody, V., Millard, J., ... & Snijders, D. (2017). Driving innovation through big open linked data (BOLD): Exploring antecedents using interpretive structural modelling. *Information Systems Frontiers*, 19(2), 197-212.