

Full Length Research Paper

User-Centred Scalable Big Data Visualizer for Power Consumption Data in the Electrical Secondary Distribution Network

¹Lucina Lawi and ^{2*}Ellen A. Kalinga

¹University of Dar es Salaam Computing Centre, University of Dar es Salaam, Tanzania

²Department of Computer Science and Engineering, University of Dar es Salaam, Tanzania

*Corresponding author: ellenakchale@gmail.com

ABSTRACT

Establishment of Smart Grids for electrical power has been practised worldwide for the purpose of bringing reliability, security, and efficient management of electrical power networks for enhancing quality service to the society. Apart from the potential aim, smart grid has been a challenge to developing countries, including Tanzania from cost and technology point of view. Due to the use of many smart equipment involved in smart grids like Advanced Metering Infrastructure (AMI) equipped with smart meters and sensors, handling and managing big data has been a challenge. Among the challenges is the issue of visualizing the Big Data due to big volume generated with high velocity. This paper is developing a user-centered scalable big data visualizer for the electrical secondary distribution network by making use of design process model by Akanmu et al. (2017) and design activity framework by McKenna et al. (2014). The approach involves three phases: pre-development, development and post-development phase. The paper reviews several approaches in visualization and demonstrates effective big data visualization. The paper managed to visualize households' units purchased against power consumed as well as balancing visualization of transformer phases.

Keywords: Smart Grid, Big Data Visualization, Visualization Models, Tanzania.

INTRODUCTION

Electrical Power Network comprises of generation, transmission and distribution networks. Generation is concerned with the creation of electrical power from using different types of power sources like hydropower plants while transmission sends power to potential areas for use. The transmitted

power is stepped down ready to be distributed to end feeders/ users. Due to the complexity of the whole electrical power network from generation to distribution, its management require introduction of intelligence in the network using Information and Communication Technology (ICT). Ibwe *et al.* (2018) state that the Smart Grid introduces intelligence and

smartness in the electrical power system that significantly increases the efficiency and effectiveness in management of the entire electrical power system. In developing countries like Tanzania, electrical networks, especially secondary distribution electrical network is highly complex.

The effort done by the public electrical utility company in Tanzania, Tanzania Electric Supply Company Limited (TANESCO) to address this complexity was to develop smartness in the generation and transmission part of the electrical network, i.e. primary network. Smartness in the secondary distribution network is in the pilot stage, where TANESCO established the Distribution Control Centre (DCC) located at Kinondoni North for Dar es Salaam city in January 2014. TANESCO established the Supervisory Control and Data Acquisition (SCADA) system, integrated with Distribution Management System (DMS), and Geographic Information System (GIS) in order to monitor and control the medium voltage (33 and 11 kV) distribution network safely and securely.

Among the challenges facing Smart Grids in extending it to secondary electrical power distribution network is the issue of visualizing the Big Data due to big volume generated from smart meters installed at each consumer, sensors and actuators in the network. The volume of data generated increases since the network and the customer base is increasing rapidly with development. According to Daki *et al.* (2017) smart grid systems generate a large quantity of data. For example, SCADA system collects data every 2–5s, AMI system collects data every 1–15 min etc. So, utilities face great number of challenges

from strategy to performance in data management. Assuming that each smart meter can send the consumer energy usage every 15 min, so every million meters can generate 96 million readings per day instead of one meter reading a month in a conventional grid. In addition to energy management, smart grids require great data management to be able to deal with high velocity, important storage capacity and advanced data analytics requirements. Since vast amount of data is continuously being generated, there has to be mechanisms to analyze and make good use of the data generated.

Visualization transforms data or information into graphical forms to be represented on the computer display. Hence, visualization deals with both transformations and representations. Smart grid data requires complex treatment because of their nature, distribution and real time constraint because of their large volume (large amount of data), high velocity (rate at which data arrive and the time in which it must be acted upon) and variety (different types of structured and unstructured data), which are the characteristics of big data. Hence, its visualization faces a variety of challenges, which limit its effectiveness in providing visibility of data within the grid. This paper focuses on developing a scalable big data visualizer application to visualize energy consumptions at the distribution part of the electricity power grid. TANESCO being the government organization managing all issues of electricity in Tanzania from generation, transmission and distribution, need an automated mechanism which will read and record consumed Kilowatthour (KWh) units in households, while comparing with the purchased KWh units.

STUDIES IN DATA VISUALIZATION

Big Data Visualization

Olshannikova *et al.* (2015) defines big data as a large quantity of datasets, presented in variety of formats coupled with high velocity to provide timeliness requirements, which gives the ability to extract different meanings from the collected data sets. Due to the fact that big data is a complex entity, its visualization faces a great deal of challenges. The characteristics of big data require new forms of processing to enable enhanced process optimization, insight discovery and decision-making. The challenges lie in data capture, storage, analysis, sharing, searching, and visualization. Visualization can be thought of as the “front end” of big data (Wang *et al.*, 2015). Daki *et al.* (2017). Nga *et al.* (2012) state that data visualization has a great role, because it improves the assessment of smart grid. There is a great number of visualization techniques based on multivariate high dimensional visualization for 2D and even the 3D visualization. Smart grids face enormous variables that complicate data presentation such as 3D Power-map. However, scatter diagram, parallel coordinate, and Andrew curve may resolve the problem of high dimensional data. According to Ali *et al.* (2016) the most important visualization feature is interactive. Also, visualization must display relevant information.

Visualization Models

The different models used in the development of data visualizers differ in factors used. One such factor is the role played by user in the development process. Here, general visualization models and user centered visualization

models are analyzed. The general models are Ben Fry model (Fry, 2007) and the Nested model (Munzner, 2009). The user-centered design (UCD) models proposed includes those by Robinson *et al.* (2005), Roth *et al.* (2010), Koh *et al.* (2011), Meyer *et al.* (2012) and Akanmu *et al.* (2017). However, only that of Nested model from general models and a model by Akanmu *et al.* (2017) from UCD has been taken into account.

Nested Visualization Model

Nested visualization model by Munzner (2009) under general models have stages as shown in Figure 1, which include domain characterization for familiarization of oneself with domain, design of data and task abstraction for mapping of problem and data from language of specific domain to description of operations and data types, design of visual encoding and interaction for choosing visualization techniques and principles, while techniques and design of algorithms is for implementation of visual encoding and interaction design.

McKenna *et al.* (2014) improved nested model by linking both decision and design process models. They proposed the design framework linking design process and visualization design decisions as depicted in Figure 2. The four main activities in the design framework are understand, ideate, make and deploy. “Understand” activity involves the first two phases of the model; domain characterization and data/task abstraction. “Ideate” activity generates ideas for supporting first activity to understand the problem domain. “Make” turns ideas into tangible prototypes using third phase output and the forth phases; techniques

design and algorithm to test design decisions made at the technique and algorithm levels. “Deploy” activity brings a prototype into effective action in the real world settings for supporting

target users’ works and goals to a useable visualization system involving actual implementation of algorithms and chosen visualization techniques.

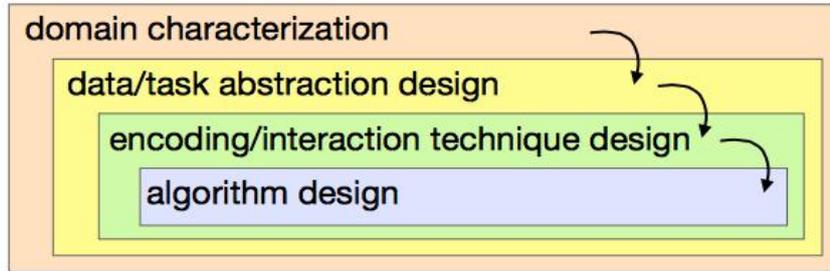


Figure 1: Stages of nested visualization model (Source: Munzner, 2009)

	<i>u</i>	<i>i</i>	<i>m</i>	<i>d</i>
domain characterization	○			
data /task abstraction	○	○		
encoding/interaction technique		○	○	
algorithm design			○	●

Where: u-understand; i-ideate; m-make; d-deploy

Figure 2: Design activity framework with nested model (McKenna *et al.*, 2014)

User Centered Design

The UCD model proposed by Akanmu *et al.* (2017) has three iterative phases; pre-development, development, and post-development as detailed in Figure 3. Pre-Development Phase is dealing with conceptualization. It has three different stages: domain analysis and dataset collection; data preparation; and users’ tasks elicitation. Development Phase is meant to produce a usable visualizer for exploration and visualization of data, and supports decision-making processes while Post-Development Phase for reviewing the technical aspect of visualizer to ensure

its workings and operations. It has deployment, testing, and Users’ experience evaluation of the visualizer.

APPROACH USED TO DEVELOP BIG DATA VISUALISATION

The research paper made use of design process model by Akanmu *et al.* (2017) and design activity framework by McKenna *et al.* (2014). The Akanmu *et al.* (2017) design process model incorporates iterative process with descriptions as one of the stages in the visualizer development. It involves three phases; pre-development, development and post-development

phase. Tables 1 and 2 show the summary of activities done at each phase and how the two models were related to one another for the development process.

The chosen model is systematic with defined set of activities which were performed in development process using Object Oriented Analysis and Design (OOAD) with Unified Modelling Language (UML). Several

artifacts like Functional Requirements, Non-Functional Requirements, Use case Diagrams, and Use Case Descriptions in the requirement analysis were applied. In the design and implementation relevant components were identified, Jupiter notebook was used as a Web server which supports Julia, Python and R programming languages was considered. The Jupiter notebook framework can work with data from databases as well as from csv files.

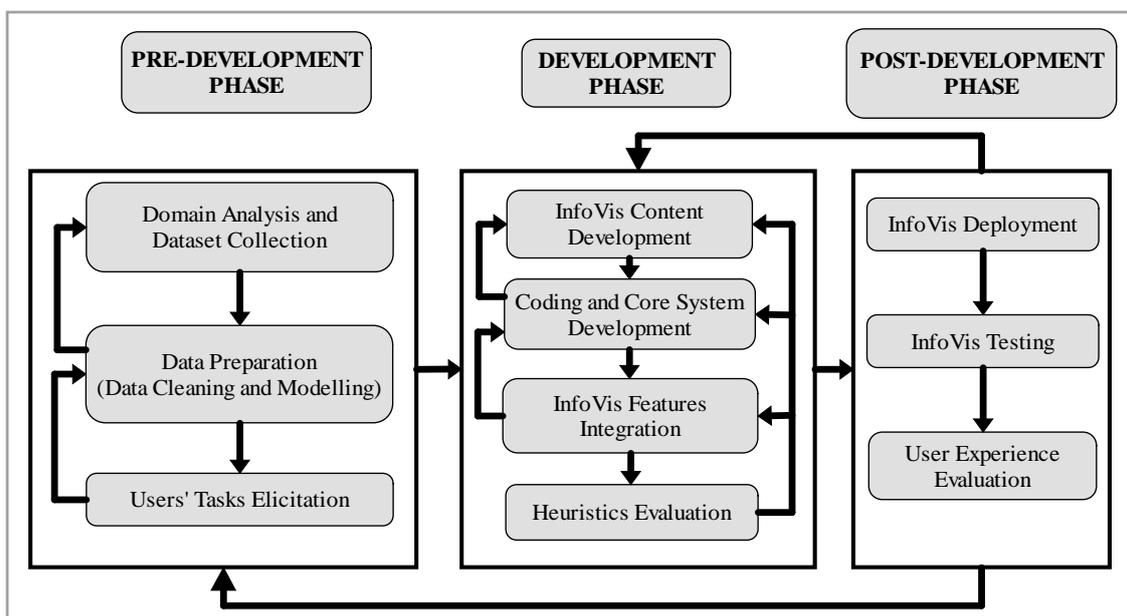


Figure 3: A process model for designing information visualization (Akanmu *et al.* (2017)

BIG DATA VISUALIZER REQUIREMENT SPECIFICATION

Requirement gathering and analysis corresponds to a hybrid user-centered design model in the pre-development phase of the system development and design activity framework adopted in this work involving domain analysis, data preparation and visualizer task elicitation.

Domain Analysis

At the domain analysis and dataset collection phase, requirement gathering was done with cooperation of the utility company staff and from literature. According to Ibwe *et al.* (2018) stakeholders of the electrical power system in Tanzania include TANESCO, the public power utility company responsible for generating, transmitting and distributing electricity and is the principle stakeholder who provided the main information which served as

inputs to this work like Energy consumption, Voltage profiles, Current load profiles, Load profiles of customers and the estimated consumption. Other stakeholders are Ministry of Energy, which is responsible for formulation and articulation of policies to create an enabling environment for stakeholders, Energy and Water Utilities Regulatory Authority (EWURA), which is responsible to regulate the water supply and sanitation services, to promote renewable energy, setting utility tariffs, licensing, and monitoring of standards related to electricity, petroleum, and natural gas and water development projects. Others are Rural Energy Agency (REA), which is responsible for handling access to modern energy services in Rural Tanzania and Tanzania Communications Regulatory Authority (TCRA), which is responsible for regulating the communications and broadcasting sectors in Tanzania. However, TANESCO is the main stakeholder and the prototype will be tested mainly by them.

Visualization needs for Substation unit include: each transformer actual consumption and transformer overload, phase overload and imbalance, current and voltage for each phase and get information to facilitate decision-making on feasibility to connect new customers.

Data Preparation and Analysis

The data generated at the secondary distribution network exhibit big data characteristics that have to be considered in design and development of the visualizer. Data collected at TANESCO Distribution Control Centre in Dar es Salaam indicated that there are around 606,726 electrical energy

customers in Dar es Salaam alone which must send data and viewed at real time, thus the data is of high velocity. Considering all customers' consumption details need to be viewed and analyzed, assuming every customer sends data every 15 minutes, hence every hour there is 36,403,560 readings, which is a huge volume. The data that need to be viewed involves both structured and unstructured data from sensors and GIS. The data being huge, visualization solution has to be scalable, handle a variety and fast changing data and allow users to be able to efficiently use the application.

The data collected from TANESCO were that of customer household and transformer. For the transformer readings there were six files of which three were for the current readings for the three phases and the other three for the voltage readings for the three phases and all of them were used. Such transformer data can be seen in Figure 4. As part of data it indicates the current from Phase A in January 2014. The collected data from six files was then cleaned and loaded into the database for merging the files. The six files of transformer data were merged into one file for visualization of a single transformer with its phases. Figure 5 shows the merged file for one transformer data for visualization. Figure 5 shows the total merged current for all phases (Phase A, Phase B and Phase C) of the same Transformer.

For the household data, the merging of the data was done so that there could be one file for each household combining consumption data as well as units bought. Figure 6 shows one such merged file for customer household data.

Table 1: Summary of activities done at each development phase

Phase Name	Stage	Action Taken/Activities to be done
<p><i>Pre-Development Phase:</i></p> <p>✓ This is considered as requirement gathering stage.</p> <p>✓ It corresponds to understand stage in the design activity framework</p>	Domain Analysis & Data collection	<ul style="list-style-type: none"> • Identify users among electrical energy stakeholders, their knowledge preference as well as their roles. • Studying user current work practices and their needs. • Determining what kind of decision making are supposed to be obtained. • Identify the kind of data to be collected for visualization.
	Data preparation & modelling	<ul style="list-style-type: none"> • Check the characteristics of the data collected. • Stating what the expected quality of data is. What data structure and dimensionality to address. • Looking how data quality will confirm with the business value. • It also ensures the data structure and dimensionality address the domain need. • Missing data values, incoherent data dimensions, and noise, are treated and modelled.
	Users' task elicitation	<ul style="list-style-type: none"> • Tasks that will achieve the earlier identified users' needs in domain analysis are conceptualized. • Identify functional and non-functional requirements from requirements gathering and from literature review. • Enumerate the core visualizer's tasks will then be done. • Evident features as well as hidden features need to be realized.
<p><i>Development Phase:</i></p> <p>✓ This stage corresponds to two main activities in the design activity framework which are ideate and make</p>	Visualizer Components Development	<ul style="list-style-type: none"> • Involves design of the visualizer based on analyzed information found in pre-development phase • Setting up design goals and objectives in terms of visualization options to be used, components as well as design principles. • Identifying components that make up the visualizer application. • Generating ideas concerning what encoding and interaction techniques are better for the power industry in the electrical secondary distribution network. • Design visual encoding and interaction techniques. • Visualizer Component Development entails the design of the prototype, starting with a low-fidelity form. • Analysis and transformation of the findings from pre-development phase will be done using draw.io.
	Coding And Core Content Development	<ul style="list-style-type: none"> • It involves the realization of the visualizer analytics dashboard. • Choosing the architecture to be used is done. • Developing the components (visual interface, data binding and visual exploratory mechanism, and database). • Developing the content delivery techniques • Choosing the programming language to be used
	Visualizer Features Integration	<ul style="list-style-type: none"> • Perfect the prototyping process, by bringing all the peculiar visualizer characteristics into reality. • A review of the technical aspects is also done to actualize prototype of high fidelity.
	Heuristic Evaluation	<ul style="list-style-type: none"> • Formal evaluation will be done to receive feedback from usability/HCI and data analysts/administrators using the prototype present. • Choice of evaluators needs to be done. • Methods of evaluation need to be identified. • Evaluators are chosen mainly from usability specialists or users with usability experience
<i>Post-Development</i>	Visualizer Deployment	<ul style="list-style-type: none"> • Visualizer system prototyped for use is deployed. • It involves interfacing the visualizer dashboard with analytical

Phase Name	Stage	Action Taken/Activities to be done
Phase		tool as well as with the real time database server that receives data from customer households and from the transformer. <ul style="list-style-type: none"> • It also involves reviewing the technical aspect of visualizer and ensures that its workings and operations are in order. • A cross-browser assessment will be done to check the compatibility of the visualizer design with the major web browsers.
	Visualizer testing	<ul style="list-style-type: none"> • Testing of the developed visualizer and debugging for errors • Dummy data will be generated and tested • Visualizer techniques overview (with details), zooming, filtering and mouse-touch interactivity, and scrolling are tested.
	Users' experience evaluation.	<ul style="list-style-type: none"> • Evaluation stage to find out how the developed visualizer performs. • Filling of evaluation questionnaire for usability test by users.

Table 2: Relation of Akanmu *et al.* (2017) UCD Model and Nested General Visualization Model

User-Centred Phase	Design Activity framework				Aim/Motivation	Outcome
	U	I	M	D		
<i>Pre-development</i> <ul style="list-style-type: none"> • Domain analysis • Dataset exploration and modelling • Visualizer Tasks elicitation 	ð				Gather, observe, and research available information to find the needs of the user.	Acquired knowledge about the target set of users, their domain-specific questions and goals, their workflows, and the types of measurements or data they have acquired
<i>Development</i> <ul style="list-style-type: none"> • Component development • Coding and core content development • Features integration • Heuristic evaluation 		ð	ð		i) To generate good ideas for supporting the understand outcomes. ii) To concretize ideas into tangible prototypes.	i) A set of ideas that are most often externalized in a variety of forms, from sketches to wireframes to even low-fidelity prototypes ii) A set of prototypes
<i>Post-development</i> <ul style="list-style-type: none"> • Deployment • Testing • User Evaluation 				ð	Bring prototype into reality in order to support the target users' work and goals.	Usable system

Visualizer System Requirements

Visualizer for the data from TANESCO (big data) has a set of defined specifications to meet its needs. They are core activities, functional and non-functional requirements. Table 3 summarizes the functional requirements. The identified core activities and their sub-activities are:

- (a) Management of Big Data
 - Capture transformer voltage and current.
 - Capture customer's consumptions
 - Read household consumption and units purchased.
 - View balancing of transformer phases
- (b) Management of Big data visualizer reporting
 - Search for household electrical data.
 - Display real time household consumption visualization.
 - View visualization for consumption data for household.
 - View visualization for comparison between consumption and units purchased for the household.
 - View visualization for phase voltages for a transformer.

- View visualization for phase current for a transformer.
 - View visualization for phases' load differences between phases in the transformer.
- (c) General Issues of the system
 - Register users
 - Login
 - Authenticate users
 - Generate statistical report
 - Connect to analytical server and database

Data from survey of various users of visualization tools show that scalability is one of key requirements for a big data visualizer. Interactivity, usability and interoperability are other non-functional requirement required for the big data visualizer. Table 5 shows the non-functional requirements for a big data visualizer.

Three actors identified are: System Administrator, Revenue Protection Engineer and Substation Engineer. These actors correspond to the Revenue Protection Unit and Substation Units respectively. Figure 7 shows a use case diagram of the Visualizer. Table 6 shows one sample use case description of the "View Graphical Representation" for the Transformer Balancing use case.

Date/Time	Value	Status
1/28/2014 5:40:00 AM	72.88	A
1/28/2014 6:00:00 AM	77.96	A
1/28/2014 6:20:00 AM	74.24	A
1/28/2014 6:40:00 AM	75.96	A
1/28/2014 7:00:00 AM	71.02	A
1/28/2014 7:20:00 AM	60.81	A
1/28/2014 7:40:00 AM	60.77	A
1/28/2014 8:00:00 AM	52.42	A
1/28/2014 8:20:00 AM	47.80	A
1/28/2014 8:40:00 AM	53.56	A
1/28/2014 9:00:00 AM	55.24	A
1/28/2014 9:20:00 AM	54.15	A
1/28/2014 9:40:00 AM	59.72	A
1/28/2014 10:00:00 AM	56.50	A
1/28/2014 10:20:00 AM	59.77	A

Figure 4: Current data from one phase of the transformer

	DATEMMDDYYYY	TIME_12H	CURR	VOLT	CUR_STATUS	VOLT_STATUS	PHASE	YEAR
0	10/10/2014	12:00 pm	227.75	717.61	NaN	NaN	A	2014
1	10/10/2014	2:00 pm	254.05	708.58	NaN	NaN	A	2014
2	10/10/2014	1:00 am	144.62	709.76	NaN	NaN	B	2014
3	10/10/2014	10:00 pm	487.68	718.18	NaN	NaN	A	2014
4	10/10/2014	3:00 am	274.81	719.05	NaN	NaN	A	2014
5	10/10/2014	4:00 pm	243.12	708.10	NaN	NaN	A	2014
6	10/10/2014	10:00 pm	362.38	719.82	NaN	NaN	C	2014
7	10/10/2014	11:00 am	215.14	711.52	NaN	NaN	C	2014

Figure 5: Sample cleaned and merged transformer data for visualization

SNO	MONTH	CUSTOMERNO	CONSUMPTION	UNITS	LOCATION	
0	NaN	JAN	10200	10643.0	1852.2	UBUNGO
1	NaN	FEB	10200	9473.0	2121.7	UBUNGO
2	NaN	MAR	10200	10180.0	580.0	UBUNGO
3	NaN	JAN	20300	8025.0	7018.0	KIMARA
4	NaN	FEB	20300	11192.0	11220.5	KIMARA
5	NaN	MAR	20300	8816.0	9513.7	KIMARA

Figure 6: Output for customer data

Table 3: Functional requirements of the big data visualizer system

Ref. No.	Requirement Description		Category
R1	Management of Big Data		
	R1.1	The system should be able to capture customers' consumption with respect to time.	Hidden
	R1.2	The system should allow to read captured customers' consumption with respect to time.	Evident
	R1.3	The system should allow comparing of phase currents for the different phases within the transformer so as to help identify balance between phases.	Hidden
	R1.4	The system should allow to view balancing status of phase currents of a transformer	Evident
	R1.5	The system should be able to locate customers through geographical information system	Hidden
	R1.6	The system should allow to read or view locations of customers	Evident
	R1.7	The system should be able to compare and analyze the relationship between units bought and actual consumption for a household within a given time.	Hidden
	R1.8	The system should allow to read customer's units bought in relation to consumption	
R2	Management of Big Data Visualizer Reporting		
	R2.1	The system should represent visualized information graphically, including: ✓ Consumption trend of each household for a given period of time. ✓ Transformer voltage and current for each phase.	Hidden
	R2.2	The system should allow viewing of: ✓ Consumption trend of each household for a given period of time. ✓ Transformer voltage and current for each phase.	Evident
	R2.3	The system should allow users to search for varieties of information based on customer meter number or transformer name.	Evident
	R2.4	The system should display real time consumption data within the given household and transformer.	Evident
	R2.5	The system should allow to view graphical statistics for power consumptions, transformer phase balancing etc	Evident
R3	General issues of the system		
	R3.1	The system should allow system users to login into the system	Evident
	R3.2	The system should authenticate users of the system	Hidden
	R3.3	The system should allow system users to logout from the system	Evident
	R3.4	The system should create connection mechanisms with other systems like GIS	Hidden

Table 5: Non-functional requirements for big data visualizer system

S/N	Attribute	Description
1.	Scalability	Big data visualizer should be flexible in manipulation of large amount of data so as to accommodate big data issues.
2.	Interactivity	Big data visualizer should allow user to interact with the application in order to get the required information.
3.	Usability	Big data visualizer should be easy to use. It should not require much training before using.
4.	Interoperability	Big data visualizer should be able to interact with other systems like database server.

Table 6: Use case description for “view graphical representation” for transformer balancing

Use Case Name	“View Graphical Representation” for Transformer Balancing
Use Case Description	This use case allows viewing of transformer balancing graphically to a span of time selected
Primary Actor	System
Secondary Actor	User
Pre-condition	Transformer Currents and voltages continuously captured and stored in the system
Post-condition	The system displays graphical visualization of a transformer balancing in-terms of Current and Voltages
Main flow	<ol style="list-style-type: none"> 1. System displays a form to select a transformer to visualize the balancing and a form to select the time span to visualize 2. User selects a transformer to visualize the transformer balancing and selects the range of time to display the trend of transformer balancing 3. System validates the entered information 4. System findings the concerning stored data from database 5. System display graphical visualization of the transformer balancing with real values
Alternative Flow	If the range is not proper, system should display an error.

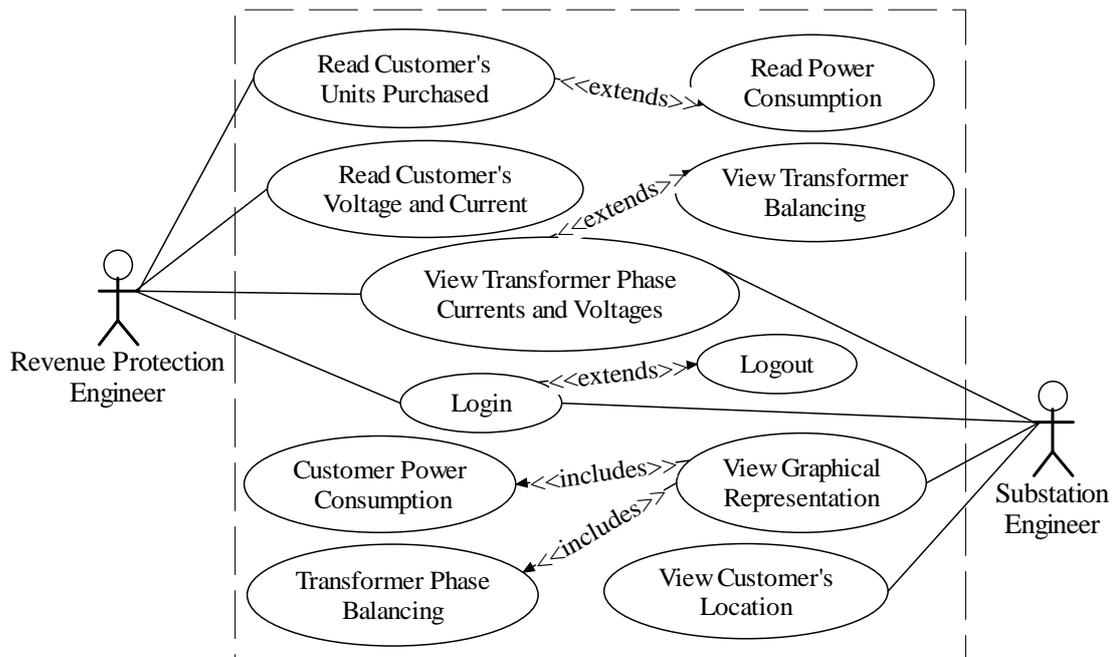


Figure 7: Big data visualizer use case diagram

DESIGN AND IMPLEMENTATION OF BIG DATA VISUALIZER SYSTEM

System design and implementation falls in both developments and post-development phases of the adopted system development approach in this work Akanmu et al. (2017) and design activity framework by McKenna et al. (2014).

Design of Big Data Visualizer

The system architecture included the webserver, user interface (Browser), database connector and database as shown in Figure 8.

- User Interface: will be through Web Browsers to view graphical interface pages.

- The web server: is making use of an open source Jupyter notebook. The Jupyter notebook supports Julia, Python and R programming languages and it can work with data from databases as well as from csv files.
- Database connector: facilitates the connection of Jupyter notebook to the database via Python sql query. Connection to get data to be used for visualization by jupyter notebook can also be done through csv files.
- Database Server: is where captured data is being temporary stored and used when visualization is called. Database that has been used is mysql database (temporarily used to hold historical data as the work mainly was on big data visualization and not to create big data database). Data can be taken from the database through database connector to create visualizations. Queries are generated according to the user and the data is extracted and visualization plotted.

The sample code as a database connector to extract data from csv file is as shown in Figure 9, while the database structure created is as shown in Figure 10.

Encoding and Visualization Technique

Different data types require different types of encoding techniques. There are categorical, ordinal, quantitative, relational and spatial data types. There are recommended visualization techniques based on the data type categories. According to Bach *et al.* (2016) some of the visualization techniques with reference

to the data types are visualization based on points, regions and lines. Visualization based on points is mainly for monthly or yearly data. It is hard to visualize data with daily intervals for five years with visualization based on points. The data being massive visualization based on regions is mostly used, choropleth map is one of the best methods for that. The kind of visualization technique chosen depend also on the objective of the data to be viewed like trend, progress or change over time line graph is used; dual axis is for four dimension data and scatter plots is for analyzing multiple data sets for similarities or differences (Bach *et al.*, 2016). Table 7 shows encoding and visualization techniques.

Implementation of Big Data Visualizer System

Python programming language has been used. Database was created with six tables: Users, CustomersHousehold, Transformers, HouseholdReadings, TransformerReadings and HouseholdPurchase. User table is for holding data for system users, CustomersHousehold table is for storing electrical customer general information. Transformer table is for storing general information about the transformer. The other two tables: transformerReadings and HouseholdReadings are for storing readings from transformers and household meters respectively. The data that the visualizer will make use of most comes from transformerReadings and HouseholdReadings Tables. Figure 11 shows the tables created into database.

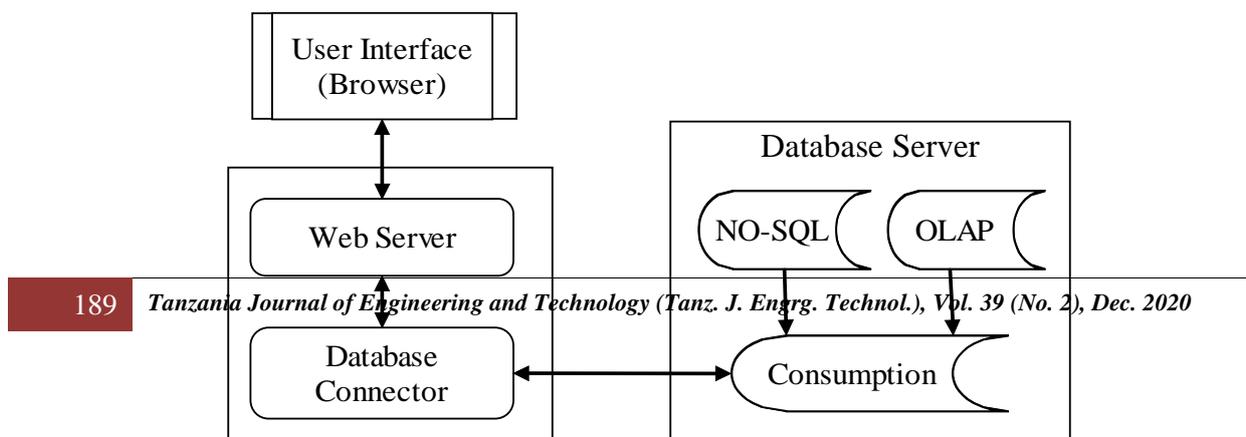


Figure 8: Big data visualizer system architecture

```

In [2]: import pandas as pd
#Reading data and saving as dataframe
Household=pd.read_CSV("C:/Users/User/Documents/Lucina/Ta
nescoData/householddata2.CSV")
#Plotting data for comparing customer consumption and units
purchased
Household.plot.bar(y=["UNITS","CONSUMPTION"],x='MONTH');
    
```

Figure 9: A Sample code extract to create visualization for household consumption data

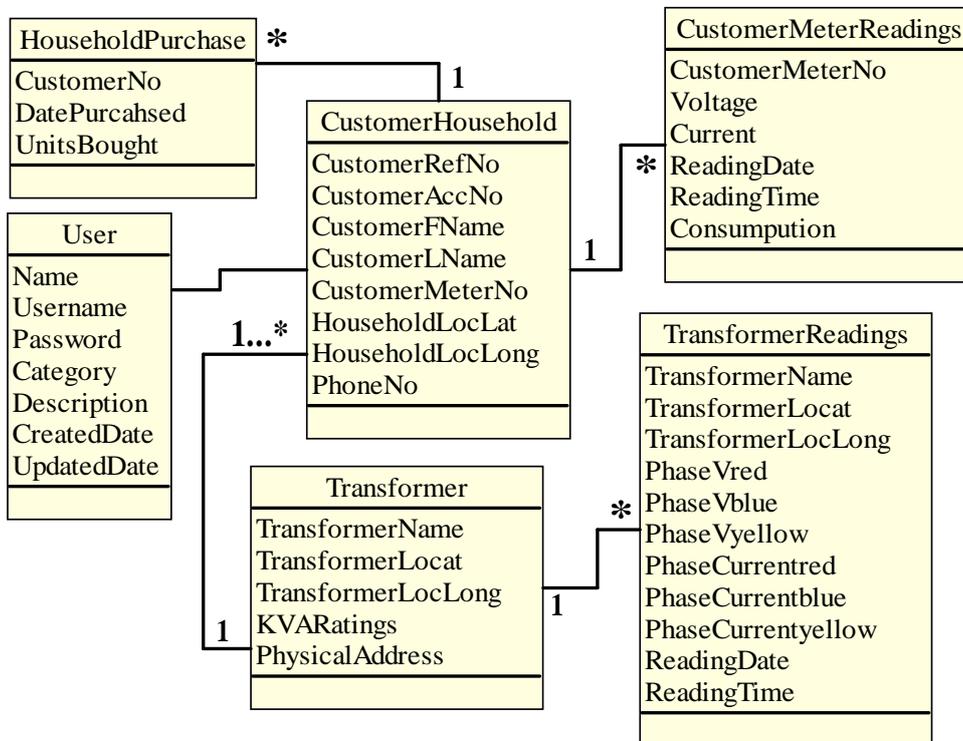


Figure 10: Visualizer system database structure

Table 7: Encoding and visualization techniques (Source: Bach *et al.*, 2016)

Type of Data to be visualized	Encoding method	Dimension
Relationship between actual consumption and units bought	Scatter plot	2 D
Household consumption over time for one customer	Line graph	2 D
Voltage and current for different phases within the transformer for five years (yearly interval)	Horizontal Bar	3 D
Voltage or current for different phases of the transformer for 5 years (20 minutes interval/monthly)	Choropleth map/line graph	Spatial-temporal
Status change and alerts (phase overload)	Animation on a map	Spatial-temporal
Details of voltage and current with geographical location for a transformer with interactivity.	Buffer, zoom and brush	Spatial-temporal

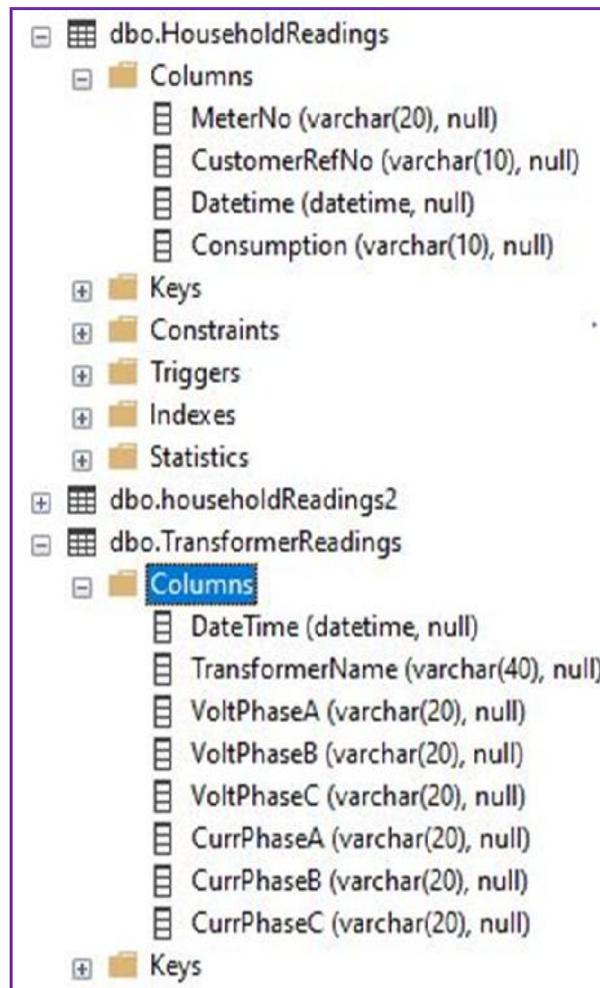


Figure 11: Big Data Visualizer in the database

Implementation of visualization system, include:

(a) Visualizer system login page and home page dashboard

The visualization system allows a user to login in Figure 12 first and thereafter be

allowed to use the system. Successful login will be provided with the dashboard as shown in Figure 13. The dashboard allows a concerned person to quickly see how many transformers and households are registered into the system. In this case

for testing purpose: 3 transformers and 3 households have been registered. The dashboard also shows the total data present in the system in its totality and their statistical consumption.

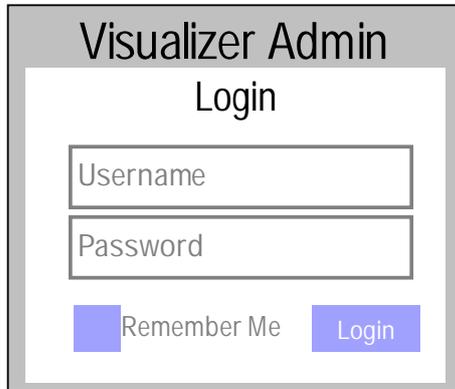


Figure 12: Visualizer system login page

(b) Visualization of household unit purchased and its consumption

When selecting to visualize the power used by a single household, a form will be displayed with an option to select a

household. Three households identified by ID: 10200, ID: 10201 and ID: 20300 were considered. As an example Figure 14 shows the purchased versus the consumption units in year 2018 of a household identified by “ID: 10200”. The exact units for a single month can be displayed by keeping a cursor on top of a waveform.

(c) Visualization of Transformers balancing

Transformer balancing can be visualized in almost four options. This is to give more clarity on understanding the trend on transformer balancing for decision making.

- ✓ For the whole year (Transformer as Kwababu for year 2016) (Figure 15).
- ✓ For a certain month within the selected year, say April 2016 (Figure 16).
- ✓ For the selected day within a month of that year, say on 17th April 2016 (Figure 17).

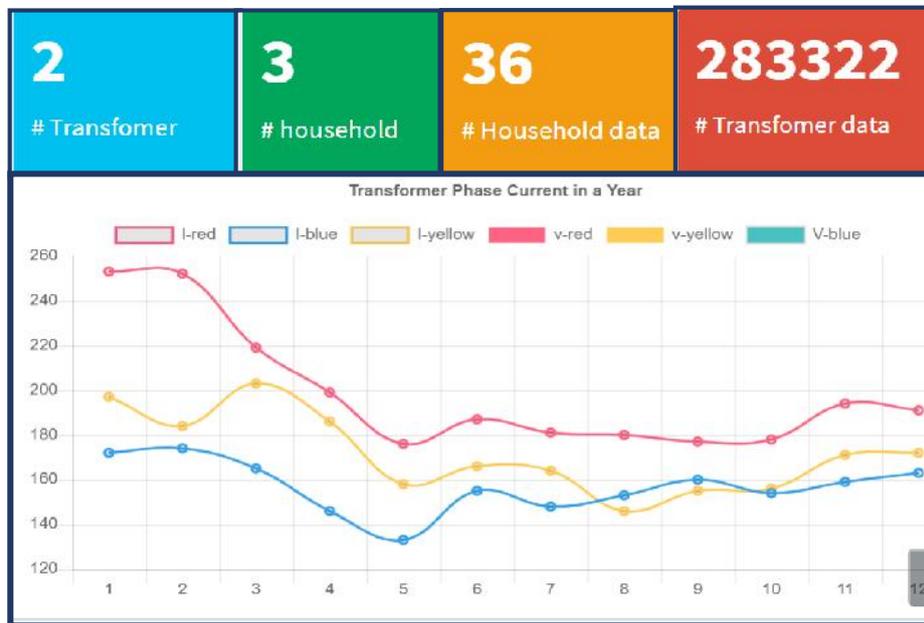


Figure 13: Visualizer system dashboard page

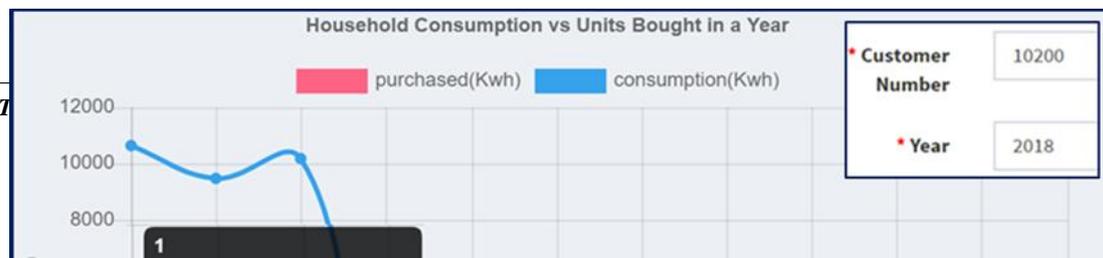


Figure 14: Household ID 10200 purchased versus the consumption units in year 2018

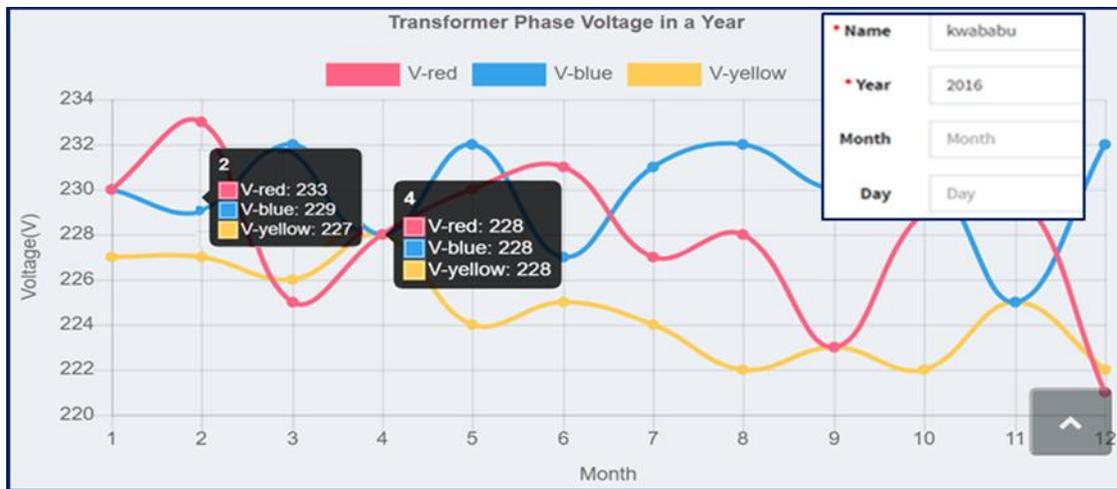


Figure 15: Visualization of “Kwababu” transformer balancing in year 2016

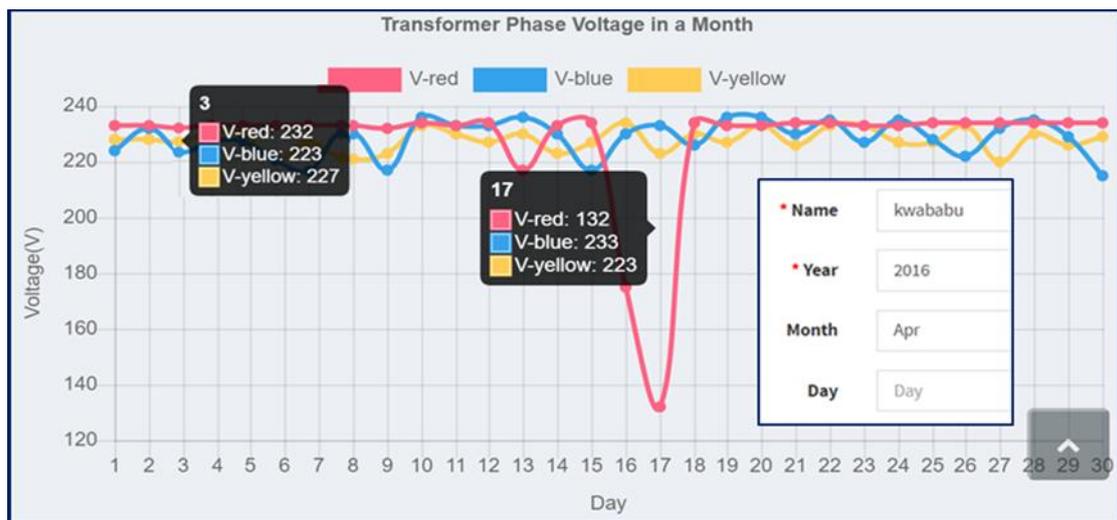


Figure 16: Visualization of “Kwababu” transformer balancing in April, 2016

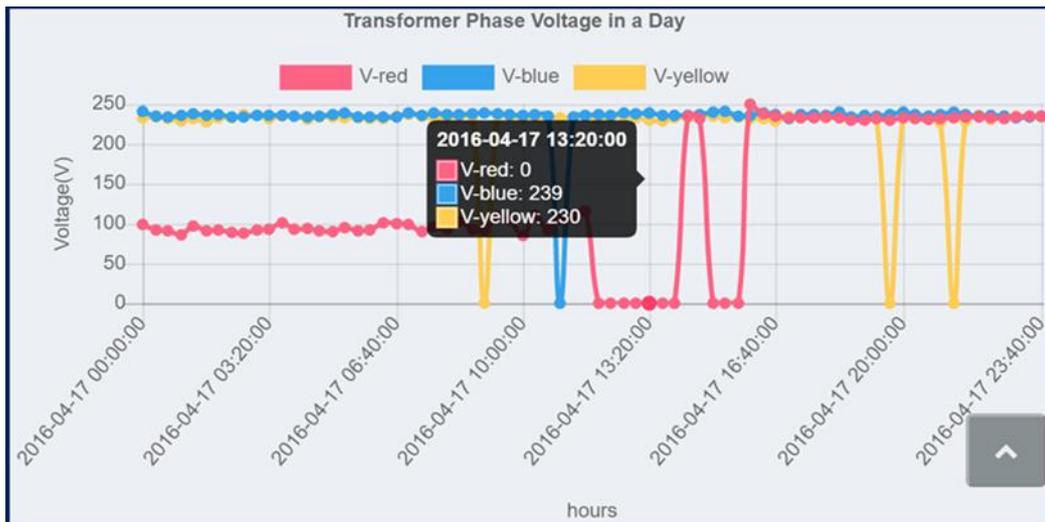


Figure 17: Visualization of “Kwababu” transformer balancing on 17th April, 2016

The research work visualized the Kwababu transformer balancing for five consecutive years from 2015 to 2019 and the waveform

could not be clearly seen as captured in Figure 18.

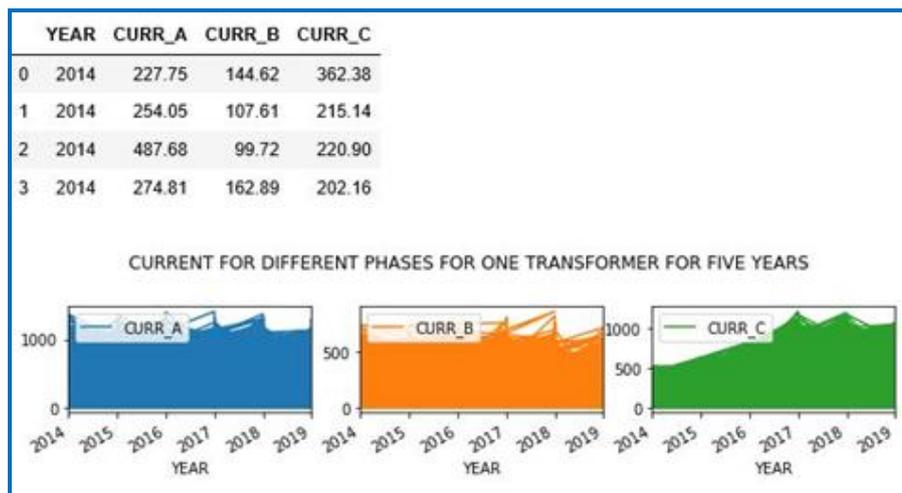


Figure 18: Transformer phases visualization graphical representations

The following is the code used to generate visualization of currents for the three phases of the transformer.

```
import matplotlib.pyplot as plt
Current11.plot(y=['CURR_A', 'CURR_B','CURR_C'],x='YEAR',title= 'CURRENT FOR DIFFERENT PHASES FOR ONE TRANSFORMER FOR FIVE YEARS', subplots=True, layout=(3,3), figsize=(10,5))
plt.show()
```

DISCUSSION

Visualization of power usage to households and transformer balancing is necessary. This paper has developed a big data visualizer system by making use of historical measured values from households for year 2018 and transformers for 5 years 2015 to 2019. Values include: purchased and consumed units in terms of kilowatthour (KWh) for households. While the measured values for transformers were phase voltages and currents. The study sampled three households and two transformers.

Looking into the household identified as 10200, visualization shows clearly the diversion of the purchased units and consumed units. The diversion of the two units is too different from January to March. In January (Month 1), the consumption was reading 10,643 Kwh while the purchased units were only 1852. Reasons for this big difference should be found, especially this time when households are using prepaid electrical meters. From Month 4 (April) to Month 12 (December) at least the purchased units were less than the consumed units.

The sampled visualization of “Kwababu” transformer balancing shows the consistent unbalancing of voltage phase throughout the year 2016, however none of the voltages were below or above the range used by TANESCO to be accepted of 207V minimum to 253V maximum. Considering only one month, in this case April, it is seen that on 17th April 2016 Red phase was less loaded (132) compared to Blue (233V) and Yellow (223V) phases.

Going further to visualizing the transformer balancing per a single day of 17th April 2016, visualization gives a room to see the measured voltage values per hours in intervals of 20 minutes. Some hours show phases to read 0V. This can be because of the power outages to phases for a certain period. Outage can be due to

normal maintenance routing by TANESCO or faults to that phase.

Slight differences in power purchased and power consumed can be due to normal losses caused by problems in electrical installation system itself. But the large difference between the two can alert the Power Utilization Company on the existence of unfair usage of power. Since there are inadequate mechanisms to monitor the dishonest power usage to households, especially in developing countries, visualization can be a starting point to identify those frauds. The main source with unfair power usage is on electricity theft, mainly caused by Electrical Meter tempering or bypassing.

Problems in power balancing to transformer phases are very critical in many developing countries including Tanzania. This is because of the nature of distribution of power to end feeders, where to a big extent only one phase to streets is provided. Therefore, to any new customer to be added to the same phase without considering whether that line or phase is overloaded or not.

CONCLUSIONS

The visualizations considered in this paper were divided into two parts, concerning household data and concerning transformer data. For household data, information that is being visualized is mainly concerning the actual consumption for each household as well as units purchased by the customer. For transformer data, the information that is being visualized is concerning the transformer phase consumptions in terms of voltage and current. Through visualizing the analytical reports on the purchased and consumed units, TANESCO can be able to judge the unlawful moves in households. Through the developed visualizer, one can visualize a single household and detect where there

is any unusual conduct of electricity. Problem on transformer balancing should start from distributing three phases to all streets, that in when balancing can easily be achieved. The paper used historical data in tables from six files for transformer data and two files for household data obtained from TANESCO. However, visualization from real environment is needed for better decision when adding households to transformers.

REFERENCES

- Ali S.M., Gupta N., Nayak G.K. and Lenka R.K. (2016). Big Data Visualization: Tools and Challenges. 2nd International Conference on Contemporary Computing and Informatics (IC3I). DOI:[10.1109/IC3I.2016.7918044](https://doi.org/10.1109/IC3I.2016.7918044)
- Akanmu S. and Jamaludin Z. (2017). A process model for designing information visualization. Computer Science, Corpus ID: 23029756
- Bach B., Shi C., Heulot N., Madhyastha T., Grabowski T. and Dragicevic P. (2016). Time curves: folding time to visualize patterns of temporal evolution in data. *IEEE Trans. Vis. Comput. Graph.* 22(1): 559–568. Available online at HAL Id: hal-01205821 <https://hal.inria.fr/hal-01205821>
- Daki H., El Hannani A., Aqqal A., Haidine A. and Dahbi A. (2017). Big Data Management in Smart Grid: Concepts, Requirements and Implementation. *Journal of Big Data* volume 4, Article number: 13. <https://doi.org/10.1186/s40537-017-0070-y>
- Fry B. (2007). The Seven Stages of Visualizing Data. *Visualizing Data*. Retrieved from <https://www.oreilly.com>.
- Ibwe K.S., Kalinga E.A., Mvungi N.H., Tenhunen H. and Taajamaa V. (2018). The Impact of Industry Participation on Challenge Based Learning. *The International Journal of Engineering Education*. 34(1): 187–200. TEMPUS Publications. Accessed from <https://www.ijee.ie/contents/c340118.html>.
- Koh L.C., Slingsby A., Dykes J. and Kam T.S. (2011). Developing and Applying a User-Centered Model for the Design and Implementation of Information Visualization Tools. 15th International Conference on Information Visualisation, 90-95.
- McKenna S., Mazur D., Agutter J. and Meyer M. (2014). Design activity framework for visualization design. *IEEE Transactions on Visualization and Computer Graphics*, 20(12): 2191–2200. <https://doi.org/10.1109/TVCG.2014.2346331>.
- Meyer M., Sedlmair M. and Munzner T. (2012). The four-level nested model revisited: blocks and guidelines. *Proceedings of the 2012 BELIV Workshop: Beyond Time and Errors - Novel Evaluation Methods for Visualization*, 11:1-11:6. <https://doi.org/10.1145/2442576.2442587>.
- Munzner T. (2009). A Nested Process Model for Visualization Design and Validation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6): 921–928. <https://doi.org/10.1109/TVCG.2009.111>
- Nga D.V., See O.H., Quang D.N., Xuen C.Y. and Chee L.L. (2012). Visualization Techniques in Smart Grid. *Smart Grid and Renewable Energy*, 3(3): 175 – 185. <https://doi.org/10.4236/sgre.2012.33025>.
- Olshannikova E., Ometov A., Koucheryavy Y. and Olsson T. (2015). Visualizing Big Data with augmented and virtual reality: challenges and research agenda. *Journal of Big Data*, 1–27. <https://doi.org/10.1186/s40537-015-0031-2>.

Robinson A., Chen J., Lengerich E., Meyer H.G. and MacEachren A. (2005). Combining Usability Techniques to Design Geovisualization Tools for Epidemiology. *Cartography and Geographic Information Science*, 32(4): 243-255. doi: [10.1559/152304005775194700](https://doi.org/10.1559/152304005775194700)

Roth R., Ross K.S., Finch B.G., Luo W. and MacEachren A. (2010). A user-

centered approach for designing and developing spatiotemporal crime analysis tools. GIScience Conference, Zurich, Switzerland.

Wang L., Wang G. and Alexander C.A. (2015). Big Data and Visualization: Methods, Challenges and Technology Progress, 1(1): 33-38. <https://doi.org/10.12691/dt-1-1-7>.