Cerebral Approach to Track Malfunction of Assets in a Smart Power Grid Information System

B T Sayed\textsuperscript{1,a}, A Dekdouk\textsuperscript{2,b}, Y Ashraf\textsuperscript{3,c} and S Chaudhry\textsuperscript{4,d}

\textsuperscript{1}Department of Computer Science, Dhofar University, PO Box 2509, P. Code 211, Salalah, Sultanate of Oman
\textsuperscript{2}Department of Computer Science, Dhofar University, PO Box 2509, P. Code 211, Salalah, Sultanate of Oman
\textsuperscript{3}Distribution Planning Department, Dhofar Power Company S.A.O.C, PO Box 2609, P. Code 211, Salalah, Sultanate of Oman
\textsuperscript{4}Department of Computer Science, Dhofar University, PO Box 2509, P. Code 211, Salalah, Sultanate of Oman

Corresponding e-mail address: b_sayed@du.edu.om

Abstract. Failure of assets at downstream level of Smart Power Grid being barely addressed, subsequently hoarding lots of superfluous human hours, depletion of valuable time and resources, inefficient and untimely rectification of problems leading to concerns of inefficient operations and hefty loss to the power grid organization in all its aspects. There is a need to address such failures in advance, through the use of prevailing information technological inclusion, implicating efficient operations benefiting the entire sector. With Knowledge Management Systems evolving at its growing needs to address critical issues and providing feasible solutions to the digitization of information already at high demand, we propose an analytical approach merged with the concepts of big data, to track historical data of assets and resources, to identify the possibilities of failure of assets in smart power grid infrastructure. The proposed system would identify the possible failure based on the past records of an assets performance and failures, not limited to, but also based on the expected life span of the assets. The issues will be archived for further sharing while in apposite needs and its solutions would be appropriately applied throughout the power business and its subsidiary. The data would be efficiently stored and archived in the form of digitized formats such as audio, video, images, comments, documents, feedback and other visually recognizable and multimedia formats. Various technological related electronics and gadget equipment’s such as tablets, mobile phones, laptops and other hand held devices could be easily utilized to support the work force teams towards an agile operation.

1. Introduction & Statement of the Problem
With the expanding list in its never ending process, much mandatory, the assets are the kernel of any power utility company [1]. Some of the power generation companies are able to develop systems to manage their assets and resources at the upstream level of their operations, where the highest valued assets are addressed. Whether the chart is upstream or downstream, any power utility company should stay ahead of the processes, simultaneously ensuring smooth operations and low cost reins [1]. Simultaneously the smart power grid hubs require real time sensing mechanisms towards collaborative management of workers [2]. With the smart grid technology in its boom and almost with the
components are not given much priority in
s including their more complex conservative
ructural costs, and also to
mproving on operations. The main
5 and
70x763
IOP Conf. Series: Earth and Environmental Science 582 (2020) 012004
doi:10.1088/1755-1315/582/1/012004

Due to the increa
an extended
smart grid intelligent smart devices, significantly impacting increased infrast
in spite of its high input costs i.e. the increased investment portfolios, through technology budded latest
leading to reduced risk factors to an
proposed system would ensure the increase in its efficiency, decreased cost factor, timely notifications
in order to
appropriate to efficiently track and manage the data too. Looking at the asset management con
smart grid data exceptionally on its upsurge, it is important that the technological aspects be addressed
these good practices with certifications are mand
International Organization for Standardization (ISO) 55000 requirements following best industry
levels [3]. Few of the enterprises have applied the Publically Available Specification (PAS) 5
requirements [3]. Particularly the downstream asset management usually involves the efficiency of
humans and workforce teams to identify and rectify the malfunction, which in obvious cases could lead
to lots of manual intervention, and could lead to mismanagement of time, simultaneously could also lead
to wastage of resources, thereby leading to inefficient operations producing direct or indirect hefty loss
to the power generation companies in a long run.
The proposed system would support the power generation companies to strategize their operations, take
informative decisions and provide efficient and effective support to the running processes. With the
understanding that new technologies in a smart grid system could enhance the operations to an extreme
extent [6], our proposal to include knowledge management perceptive and its aspects holds a value of
appeal. Based on the recurring experience of past asset failures the proposed knowledge management
system would help identify the possible malfunction, its time frame, thereby providing an alert as well
attention to the predicted areas. The process will simultaneously and positively impact the capability of
individuals working in the project team thereby efficiently improving on operations. The main
challenge involved is the training of the workforce teams with the information technological aspect of
the proposed system, but the effort of it would be too minor in comparison to the benefits occurring from
the whole system in a long run. The proposed assets maintenance takes an evolutionary approach of life
cycles of the asset. With the value factor into consideration, based on the performance of an asset, with
the records in place, the system would propose the possibility of failure of that particular asset and
indicate towards a possible maintenance or replacement in order to ensure efficient reliable operation.
Various factors affecting the asset such as ageing, life span period, periodic maintenance tests, periodic
service alerts, components failure, components repairs, voltage fluctuation aspects, depreciation values,
asset costs, performance factors, risk analysis and other areas would be focused to determine the
probability of future malfunction of the assets. The new type of assets life span is shorter in comparison
to the traditional ones, their operating setups differ and their operating procedures too vary at different
levels [3]. Few of the enterprises have applied the Publically Available Specification (PAS) 55 and
International Organization for Standardization (ISO) 55000 requirements following best industry
practices, but this happens only at the upstream level where the cost of assets is tremendously high and
these good practices with certifications are mandatory to further emerge and stay competitive. With the
smart grid data exceptionally on its upsurge, it is important that the technological aspects be addressed
appropriately to efficiently track and manage the data too. Looking at the asset management convolution,
in order to effectively manage the assets, through the viability of knowledge management structure, our
proposed system would ensure the increase in its efficiency, decreased cost factor, timely notifications
leading to reduced risk factors to any unexpected malfunction of assets. To ensure persistent operations,
in spite of its high input costs i.e. the increased investment portfolios, through technology budded latest
smart grid intelligent smart devices, significantly impacting increased infrastructural costs, and also to
an extended common factor of extending the benefits to the end users but with nominal and affordable
costs, asset tracking and management could play a vital role to provide the right solution at the right
time.
Due to the increasing volume of unbounded grid of data occurred from various intelligent devices such
as smart meters and sensors it is vital to integrate the techniques of big data with certain agile tool to
manage and handle the data related to complex smart grid infrastructure [7]. Using the tools and techniques of big data, our aim is to store historical data related to assets, archive details of past asset failures, archive the solutions and remedies to the failures of assets, analyze the trend, generate reports and drive conclusions on the future failure of an asset, constructed on the past performance and records. The purpose of using big data lead due to the fact that the amount of data generated, at downstream level, in a smart grid is tremendous and it could be equally difficult to manage this voluminous quantity of data without the insertion of big data specifically available for such major causes, in hand in today’s era. In addition to the mismanagement of customer data, the new data generated in a smart grid technology is also not appropriately addressed or utilized which leads to critical barrier to the success of the system [7]. Thus our proposed system would help analyze the historical data and forecast the possible failure in assets that could then be rectified in advance avoiding any sort of disaster and its complicated recovery processes. Our idea is to convert data to valid information, knowledge and then develop appropriate action plans to visualize it.

The paper is written with specific concentration to critically address the downstream level asset management in a smart grid power technology, to effectually reinforce human expertise and unswerving resources, through proficient usage of technology, to ultimately benefit the community as well the power industry. Aligned with the mission of the conference, the research is aimed to embed the concept of autonomous systems which could effectually benefit the associated entities through various paradigm of smart grid and energy.

2. Big Data and Historical Data

Though there have been various technologies and methodologies to store and analyze data, the crucial mode of handling the data has been focused using the methods of relational databases and data warehousing, which has been implemented for years and is also being currently used [8] [9]. These traditional mode of data storage and archival systems has been mainly designed to work out with structured data, set of data that has been arranged in pre-defined structured format ready to be processed and analyzed as desired. Followed by analysis and design of a system, structured data requires sufficient efforts in clearly defining the fields, records, data types, relationships between data, schema validated data, a standard interface to extract data, applications to process the data, and lots others before it could be developed and implemented [8]. This process could consume stretched period deleteriously affecting the business to analyze the desired information from the stored data as by the time the data is evaluated the trend would have been over. Simultaneously, with the fact that the data would be on rising peaks on a daily basis, to handle the ever increasing large set of data consistently, also negatively affects the efficiency of the system, at a certain stage causing crash to the whole system. Adverse effects of such traditional systems are that it could cause the technical experts to waste most of their valuable time and efforts in just managing the data in contrary to significantly spend their efforts towards visualizing and analyzing the data. In addition to manage the data, a true data system should enable the experts to spend equal amount of their time and efforts to visualize and analyze the derived output and further develop on an effectively desirable system. Unstructured or semi structured data, represents a state of data which is completely or partially in its amorphous form and would have been received possibly from unidentified or unknown resources and this type of data need to be appropriately acknowledged before its convention to an organized form. In this scenario, most importantly, the challenge to address the new form of evolving semi structured or unstructured data is not inevitable. It’s evident that the unstructured form of data alike contains that vital information as seen through structured data [10]. This set of unstructured data could mean a lot to any organization. In view of a smart power grid structure, a set of unstructured or semi structured data plays a vital role where the experts pass over the related messages and information through using various platforms. From the perspective of business intelligence, these sets of inevitable unstructured data need to be carefully analyzed, archived and need to be used in applications for further amplification as and when needed. Here comes the advantage of framework of big data, where not only the cost of continuously growing structured data which is being a huge concern for any organization, simultaneously to address the unstructured data or semi structured data being inevitable becoming a challenge on the amount of never ending storage space requirement [8]. With apparent known advantages of big data framework over the traditional data such as schema-on-read, local storage,
commodity hardware, simplicity and correlation the applicability of this transformational technology gain from every angle of success, big data is the best feasible solution to such grounds [8]. In the arena of business intelligence, in today’s digital age, big data is already being applied through various applications that affect our day today experiences and life [11]. On the other hand, the concept of big data evolves where the capability of data management using commonly used software tools ends its capability to capture, manage and process data efficiently [12].

Hadoop distributed file system is capable to handle and store voluminous sets of information and also tackle the issue of unexpected infrastructure catastrophe without affecting the data. It works on clusters made up of inexpensive processing hardware and in case of failures the operations are sustained without any interruption with automated shift of hardware available within clusters. Another main advantage is its extremely secured and safe architecture that safeguards the entire data against any external unauthorized attacks [12]. In our project, the framework of big data is critically incorporated, as it is obvious to mention that the smart grid technology through smart devices generates hefty volumes of data that need to be proficiently accumulated, stored, archived, distributed and addressed. Our proposed knowledge management systems architecture stipulates such assimilation rates and provides value to the whole system.

Historical data is a data that stores organisations history in the form of a well arranged structured, semi-structured or unstructured data and is vital to be used for extrapolative analysis [13]. It could be effectively archived where examples of such semi-structured or unstructured historical data could be email communications, messages, meeting minutes, other form of communications held on formal or informal basis within an enterprise. In our proposed system the role of historical data is vital and the form of data that would be stored in historical database is the patterns of semi structured and unstructured data which then will be converted into structured data for further processing. These structured sets of historical data will equally provide the detection of error faults in the smart grid downstream components. As seen in the research there are platforms to crop, appraise, assemble, manage and present high quality information, with extreme rapid search and retrieval processes, through the use of archived historical data, using knowledge management aspects [14].

3. Data Repository

The limitations of traditional data management and approaches could be easily tackled with the concept of data lake. The concept of data lake is created to handle voluminous big data through easily affordable and economical processing hardware [15].

The data lake provides a feasible solution towards the challenges of traditional data management processes through the benefits of effectually handling voluminous data through enhanced suppleness, reduced overheads and amplified efficacy, concurrently becoming a superlative platform to much needed visualization process [15].

As proposed in our system, the concept of data repositories in the form of data refinery and data lake will help successfully execute the entire process whereby the data would have refined and archived in data lake for further processing [16]. The concept of a data lake is fetching continuing popularity as a most feasible solution to the challenges occurred from big data. Data lakes are also designed with the concept of archiving all the raw data irrespective of its known scope or usage [17]. A data lake would define all the semi-structured, unstructured and structured data that would be stored systematically and could be used for further diagnostic processes [16]. The advantages of data lake are that it enables faster insights, in the sense that it allows users to extract data from an unstructured or semi structured form for immediate application [16]. Similarly, the concept of data refinery would enable the unstructured data to be converted into a usable format of data [16]. Our system therefore uses the concept of data refinery and data lake to sort out the voluminous inflow of data.

With the advantages of big data technology to efficiently handle all types of data whether it be structured or semi structured or unstructured, simultaneously the advantage of historical data, being served with the history of archived data, the proposed system through the concepts of data engineering prepares, evaluates and visualizes the vital data for analytical and diagnostic purposes.
4. Proposed System
With reference to Fig. 1 titled “Upper Level Architecture of Big Data Incorporated with Smart Power Systems Layer”, the workforce teams could communicate and send in the messages and communications related to the assets into the Assets Data Management Systems (ADMS). The data engineer would be responsible for all the technical aspects of systems implementation, data management and data storage.

The proposed system is targeted to provide simultaneous updates to the concerned professional team at work, for its rectification. The data engineer will monitor all the messages and communications send in to the system while modeling, visualization and producing the statistics and analysis will be the responsibility of the data scientist.

Though the data engineer will be responsible for the functioning of the whole data management system, data scientist will ensure that the data flows well based on defined standards to and from the workforce teams. This process will accurately address the assets statistics and possible failure based on specified criteria. Existing technical workforce experts could be trained to act as data engineers and data scientists.

The data engineer with the support of workforce teams would create asset groups to segregate the asset entities with varying attributes. This should be done keeping into consideration the varying asset features and its distinctiveness to facilitate ADMS storage, visualization and analysis. Each component need to be segregated as every components nature varies in terms of its features and distinctiveness. The life span of each component as defined by the manufacturer should be accurately recorded in the ADMS. Data engineer will be responsible to feed all details of the assets, into the ADMS archives, such as component identification, installation date, date of first and thereafter fault concurrences, fault error, severity of the fault, performance and reliability factors, brand category (this will help in identifying the reliability of manufacturer brands and specific faults) and all other relevant details of the faulty components. Thus, historical data of each component need to be carefully evaluated. With accurately fed historical data of millions or billions of such components inclusive of IEDs, when stored and archived, the ADMS, through data scientist would be able to analyze, visualize and forecast the assets on risk, extending warning with recommendations to appropriate and feasible remedies as well corrective actions.

This would include possible asset failures, predictive time frame of an asset failure, and conclusive recommendation on brand reliabilities and commendations on every single asset component. On a long run, the benefits to the smart grid company would be tremendous in terms of costs, as almost every asset is being tracked for its reliability, supporting legitimate decision making process to assets changeover or preservation. Another huge advantage would be that drastic power failures could be avoided helping effectual improvement in operations and grid management.

Traditional database systems based on data warehousing technologies cannot specifically handle such type of numerous data that as most of the raw data is been lost and just a few percentages of data are being reached to the processing level [18]. Moreover, it is quite complicated to process the received data and time taken to process the received data and evaluate is too much thereby the data being not much useful or outdated [18]. Even some mediocre technologies could not handle capture complete amount of raw data except the data that the Information Technology (IT) department analyst would have captured and included. The exquisiteness of big data is that it could handle all types of data whether it is structured or unstructured and is capable enough to handle them efficiently and effectually irrespective of its volume [19].
Figure 1. Upper Level Architecture of Big Data Incorporated with Smart Power Systems Layer
The benefits of proposed system results in addressing, archiving and utilizing the data of assets, very appropriately to handle and treat the fundamental problem of assets failure in a smart power grid technology, to avoid any unexpected disaster and predict any upcoming calamity, evolved due to startling faults. The system could also be used as an asset failure prediction system, which means the power grid companies would be aware of any possible threat or failure of a smart grid assets or its components, much before the occurrence of real tropical incident.

As seen in Fig. 2 titled “Assets Data Management System (ADMS)”, we thus implement the framework of big data to capture the maximum amount of raw data to be converted into information and knowledge before extracting it as a valuable collaborating item useful to share and forecast the feedbacks, outcomes and results.

The historical data, both structured and unstructured, will play a core vital role towards storage of all relevant data that would work as a base or foundation to the whole process. Evident that the big data cluster or its ecology with its enormous capability of parallel computing enables data to be efficiently distributed among servers with saving a mammoth amount of processing time [20]. Though the personnel’s need to be specifically equipped with big data knowledge and skills, due to large sets of data flowing through the system, the concept of building big data support to such projects is vital. As depicted in the Fig. 2, our system is divided into three main aspects of data manipulation starting from data engineering through data preparation till data analytics contributing a data lake.

The concept of data refinery process is to eliminate incongruent data within a collective milieu to understand and produce data in an operational format [21]. Data refinery basically derives patterns from unstructured data transforms it to a meaningful data that could be used for further analytical purposes. We propose a data lexicon to be created to convert unstructured data to structured data via a viable transformation process of data refinery.

The data lexicon is a process where the data produced for analytical purpose will be further evaluated, filtered and processed to be further stored as structured data. Our system uses Optimized Row Columnar (ORC) formats as it could easily and efficiently handle storage of all those huge trillions of sets of data and allow capering over immaterial sectors of data. With the importance to maneuver large sets of vast sets of data simultaneously leading to extremely sophisticated data storage, ORC file formats structure over Record Columnar (RC) file formats have the advantage to eliminate the large complex or manually maintained indices [22]. The basic structure of ORC formats is that the data rows collections are stored in one file under large block sizes and row data in turn is stored in columnar format [22]. Thus, all the unstructured data will first be accumulated in a file and then be executed for processing. In case of Hadoop, the system will convert the unstructured data to structured data using the four tier coding in the form of dataset, map code, reducer code and driver code [23]. The data will be stored and accumulated in the data set using columns, which would indicate keywords and the keywords related information. A map code using the class unstructured would be written to set as an output to the reducer code. The map code would write the contents in context ready to be inputted to the reducer code. The reducer code will help convert the texts to string format and the resulting data will be inputted to the driver class. The driver class is used to declare and set the classes to be used as input and output to be written in the Hadoop system. The structured data will be then used for visualization purposes by the respective team.

The proposal focuses on downstream level of the smart power grid system, where all the assets listed in downstream level is being addressed, communicated, archived with its historical assets data, forecasted between the teams indicating possible failures based on its performance history and life span, thereby enabling manage the assets. The teams using Assets Data Management System (ADMS), would always be ready with alerts to drop out any unfortunate possible upcoming assets problems or disasters.
5. **Four Tier Coding**
A sample of four tier coding to convert unstructured data into structured data is forecasted as follows:

5.1 **Dataset**
- PowerTransformer transient surge voltage, arcing ground if neutral point is isolated.
- FeederPillar fuse blown due to short circuit.
- PowerTransformer power frequency over voltage, sudden disconnection of large load.
- HighVoltageOHL hot-washing performed on dirty insulators.
- HV Cables partial discharged and cable fault.
- PowerTransformer transient surge voltage, atmospheric lightening impulse.
- FeederPillar relay tripped relay due to short circuit.
- LVCables short circuit downstream between two phases.
High Voltage OHL bridging faults occurred due to natural calamities.

5.2 Map code
public class Unstructured{
    public static class Mapping extends Mapper<LongWritable, Text, Text, Text> {
        List<String> col1 = new ArrayList<String>();
        List <String> col2 = new ArrayList<String>();
        public void map(LongWritable key, Text value, Context context) throws IOException, InterruptedException {
            String line = value.toString();
            if(line != null && line.length() > 0)
            if(line.contains("PowerTransformer")||line.contains("FeederPillar")||line.contains("HighVoltageOHL")||line.contains("HVCables")||line.contains("LVCables")){
                String [] parts = line.split("t");
                col1.add(parts[0]);col2.add(parts[1]);
            }
        }
        protected void cleanup(Context context) throws IOException, InterruptedException {
            for(int i = 0; i < col0.size() && i < col1.size(); i++)
            {context.write(new Text(col0.get(i)),new Text(col1.get(i)));
            }
        }
    }
}

5.3 Reducer code
public static class Reducing extends Reducer<Text, Text, Text, Text> {
    public void reduce(Text key, Iterable<Text> value, Context context) throws IOException, InterruptedException {
        try{
            String names=new String();
            for(Text val : value){names+=val.toString()+"n";}
            context.write(key,new Text(names));
        }catch(Exception e){e.printStackTrace();}
    }
}

5.4 Driver code
public static void main(String[] args) throws Exception {
    Configuration conf= new Configuration();
    Job job = new Job(conf,"UNSTUCTURED");
    job.setJarByClass(Unstructured.class);
    job.setMapperClass(Mapping.class);
    job.setReducerClass(Reducing.class);
    job.setMapOutputKeyClass(Text.class);
    job.setMapOutputValueClass(Text.class);
    job.setOutputKeyClass(NullWritable.class);
    job.setOutputValueClass(Text.class);
    job.setInputFormatClass(TextInputFormat.class);
    job.setOutputFormatClass(TextOutputFormat.class);
    Path outputPath = new Path(args[1]);
    FileInputFormat.addInputPath(job, new Path(args[0]));
    FileOutputFormat.setOutputPath(job,outputPath);
    outputPath.getFileSystem(conf).delete(outputPath);
    System.exit(job.waitForCompletion(true) ? 0 : 1);
}

6. Visualization Tool
This huge excavation of data could only be utilized and applied as an incredible treasure only if the data flow could be appropriately tracked with the help of analytical algorithms, simultaneously being visualized effectually and proficiently through a visualization tool [24]. Consecutively, since graphical representations and illustrations enable speedy visualization of data and objects in human’s mind in comparison to the textual version of the same, consecutively in order to enable effective and efficient decision making processes, it is vital that data be represented and visualized using appropriate valid tool, to empower easy understanding and effectual support towards decision making processes [25]. Thus in
order to efficiently manage and tackle the enormous overflow of data, visualization process would play a core and critical protagonist. Our research utilizes the CanvasJS as a prospective visualization tool, to effectively supports the use of Tablets, iPhone, Android devices, Windows platform and Mac operating systems [26]. As seen in Fig. 3 titled “A Sample Visualization Tool using CanvasJS”, our system requires the field workforce teams to use roaming handheld devices to commendably communicate with the data center.

![Chart to Visualize the Smart Grid Assets Fault Analysis](https://canvasjs.com/assets/script/canvasjs.min.js)

Figure 3. A Sample Visualization Tool using CanvasJS

We thus recommend such visualization tool as a core requirement to effectively address the issues.

```html
<html><head><script type="text/javascript">
window.onload = function () {
    var chart = new CanvasJS.Chart("chartContainer", {
        theme: "light1", // "light2", "dark1", "dark2"
        animationEnabled: false,
        title: {text: "Chart to Visualize the Smart Grid Assets Fault Analysis"},
        data: [{type: "column",
            dataPoints: [
                { label: "Power Transformer", y: 10 },
                { label: "Feeder Pillar", y: 15 },
                { label: "High Voltage OHL", y: 25 },
                { label: "HV Cables", y: 30 },
                { label: "LV Cables", y: 28 }]})
    chart.render();
</script><div id="chartContainer" style="height: 370px; width: 100%;"></div>
</html>
```

7. Conclusion
The research is carried out in discussions with experts in the field whereby it was realized that the downstream level assets failure, in a smart power grid technology, is a serious concern, with minimal attention paid to it. After a careful assessment to address the issue, it was concluded that an information system to keep track of the assets failure and its behavior could help illuminate the issue to an extreme magnitude. It is vital to address the downstream level components faults in order to save on costs and apply on efficient operations. Simultaneously, it is also important to note that addressing the downstream level components is not easy without an asset management system. The proposed system encoded with knowledge based paradigm as demonstrated will be efficacious in enhancing operations and cutting irrelevant costs. In addition, the inconveniences caused to the power distribution company as well the consumers; due to deficiency of such system will be eliminated. The research extensively
applies the use of big data technology to store the huge set of transacted data which is inclusive of structured as well unstructured data. With articulated understanding that the power management companies lack sufficient technological and innovative implementations, we propose to provide a feasible and effective solution to address these faults using the designed ADMS unified with a knowledge based system paradigm.

8. References
[1] PWC. (July 2018). Assets and Operations: Strategic, integrated assets and operations - a “must”. [Online]. Available: https://www.pwc.com/us/en/industries/power-utilities/assets-operations.html
[2] Xiwei W, Qi D, Jinglin G, Hongbin G. RFID Application of Smart Grid for Asset Management. International Journal of Antennas and Propagation, 2013; Vol 2013, Article ID 264671, 6 pages. http://dx.doi.org/10.1155/2013/264671. Available: https://www.hindawi.com/journals/ijap/2013/264671/
[3] Scott S, David G. (July 2018). Smart Grid Asset Management-an Integrated Approach to Maximize Asset Utilization, Powergrid International, [Online]. Available: http://www.elp.com/articles/powergrid_international/print/volume-19/issue-6/features/smart-grid-asset-management-an-integrated-approach-to-maximize-asset-utilization.html
[4] Panda, Mrutyunjaya. Intelligent data analysis for sustainable smart grids using hybrid classification by genetic algorithm based discretization. Intelligent Decision Technologies, 2017;11(2):137-151.
[5] Komninos N, Phillipou E, Pitsillides A. Survey in Smart Grid and Smart Home Security: Issues, Challenges and Countermeasures. IEEE Communications Surveys & Tutorials, 2014;16(4):1933-1954.
[6] Raja ML, Mohd WM. Technologies Used in Smart Grid to Implement Power Distribution System. Indonesian Journal of Electrical Engineering, 2015;16(2):232-237.
[7] Daki H, ElHannani A, Aqal A. et al.. Big Data management in smart grid: concepts, requirements and implementation. Journal of Big Data, 2017;4(13)
[8] George T, Rommel G, Steve J, Charles K, Justin M. (July 2018). Understanding the Big Data World. [Online]. Available: http://www.pearsonitcertification.com/articles/article.aspx?p=2427073&seqNum=2
[9] Shams K, Farishta M.. Data warehousing: toward knowledge management. Top Health Inf Manage, 2001;21(3):24-32.
[10] Christine T. (April 2018). Structured vs. Unstructured Data, [Online]. Available: https://www.datamation.com/big-data/structured-vs-unstructured-data.html
[11] Drew R. (April 2018). Semi-Structured Data. [Online]. Available: https://www.datamation.com/big-data/semi-structured-data.html
[12] Rajan K, Sarika G. Data Refinery with Big Data Aspects. International Journal of Information and Computation Technology, 2013;3(7):655-662.
[13] Maryanne G. (August 2018). How Applications of Big Data Drive Industries. Available: https://www.simplilearn.com/big-data-applications-in-industries-article
[14] Peter NP, Rob B, Thomas C, Kevin F, Pieter F, Peter T, Harvey W. Dacura: A new solution to data harvesting and knowledge extraction for the historical sciences. Historical Methods: A Journal of Quantitative and Interdisciplinary History, 2018;51(3).
[15] Krause DD. Data Lakes and Data Visualization: An Innovative Approach to Address the Challenges of Access to Health Care in Mississippi. Online Journal of Public Health Informatics, 2015;7(3):e225.
[16] Chris C. (August 2018). Top five differences between data lakes and data warehouses. [Online]. Available: https://www.blue-granite.com/blog/bid/402596/top-five-differences-between-data-lakes-and-data-warehouses
[17] Wiktionary. (2015). Definition of Data Lake. [Online]. Available: http://en.wiktionary.org/wiki/data_lake
[18] John S, (April 2018). Platfora’s End-to-End Big Data Analytics Architecture. [Online]. Available: https://www.platfora.com/blog-post/platforas-end-end-big-data-analytics-architecture/

[19] Sivaraman R, Srinivasan S, Chandrasekeran R. Distilling Big Data: Refining Quality Information in the Era of Yottabytes. The Scientific World Journal. 2015;453597:9p.

[20] Frank L. (August 2018). Big Data Technology. [Online]. Available:https://datajobs.com/what-is-hadoop-and-nosql

[21] Carmen R. (August 2018). What’s New in Data Refinery?. [Online]. Available: https://medium.com/ibm-watson/whats-new-in-data-refinery-71ee098a89f8, 20 March 2017

[22] Christian P. (August 2018). Big Data Science and Cloud Computing. [Online]. Available: https://www.semantikoz.com/blog/orc-intelligent-big-data-file-format-hadoop-hive/

[23] Prateek. (September 2018). Converting Unstructured to Structured Data Using Hadoop. [Online]. Available: https://acadgild.com/blog/converting-unstructured-to-structured-data-using-hadoop

[24] Enrico GC, Antonio MR. Big Data Visualization Tools: A Survey - The New Paradigms, Methodologies and Tools for Large Data Sets Visualization. Presented at: 6th International Conference on Data Science, Technology and Applications. Madrid. Spain.

[25] Evgeniy YG, Vasiliy VG. Analytical Review of Data Visualization Methods in Application to Big Data. Journal of Electrical and Computer Engineering. 2013;969458:7p.

[26] Chatterjee S. (September 2018). 10 Big Data Visualization Tools Everyone in the Industry Should Be Using. [Online]. Available: https://www.promptcloud.com/blog/6-big-data-visualization-tools-for-you

Acknowledgements

This research is a partial component of a project grant titled “Towards a Future Smart Dhofar Power Grid”, Grant reference: ORG/DU/El/14/020” and supported by the Research Council of Oman. Hence we extend our sincere thanks and appreciation to the Research Council of Oman for facilitating us to successfully execute the research component.