ABC Research Alert
Vol 7, Number 3/ 2019
Published Online: 20 December, 2019
https://abc.us.org/ojs/index.php/abcra

Advanced Metering Infrastructure Data: Overviews for the Big Data Framework

Md. Mostafijur Rahman[1], Mahesh Babu Pasupuleti[2], Harshini Priya Adusumalli[3]

[1]Lecturer, Business Administration, First Capital University of Bangladesh (FCUB), Chuadanga, **BANGLADESH**

[2]Data Analyst, Department of IT, iMinds Technology Systems, Inc., 1145 Bower Hill Rd, PA, **USA**

[3]Software Developer, iMinds Technology systems, Inc., 1145 Bower Hill Rd, PA, **USA**

Abstract

The Advanced Metering Infrastructure (AMI) statistics provide real-time information about power use as well as social, demographic, and economic aspects within a community. This study proposes a Data Analytics/Big Data architecture for leveraging AMI data in Smart City applications. The framework has three main components. First, the architectural view positions AMI within the SGAM. Second, the methodological view describes the DIKW hierarchy and NIST Big Data interoperability model's translation of raw data into knowledge. Finally, human expertise and talents to analyze the results and translate knowledge into wisdom are a connecting aspect between the two approaches. A binding element that supports optimal and efficient decision-making is added to our new perspective. We made a case study to demonstrate our framework. A load forecasting application is implemented in Retail Electricity Provider (REP). Some of the REP's marketplaces have a MAPE of less than 5%. The scenario also highlights the binding element's effect on new development options and as a feedback mechanism for more forceful decision making.

Keywords

Big Data, Advanced Metering Infrastructure, AMI Data, Data Analytics, Smart Meter

Introduction

In order for Smart Cities to be successful, they must make extensive use of information-based technology. As a result, big data and data analytics have evolved into reliable tools that aid in the creation of applications for the various parties involved in them (Achar, 2018b). A key player is Smart Grids, which enable data collecting in order to construct a more evolved and efficient electrical network. Smart Grids are becoming increasingly popular. Adoption of Advanced Metering Infrastructure (AMI) is aimed at encouraging the technologies that are available to quantify and measure the flow of energy across the distribution network (Adusumalli, 2016a).

It is this infrastructure that not only assists the utility in providing information, but also allows the consumer to participate as a stakeholder in the energy value chain. AMI data represent a source of real-time information not only on electricity consumption but also on potential population behaviors, such as concentrations of people, population migration, demographic trends, and economic changes in various sectors of the population, among other things. AMI data are collected from a variety of sources, including satellites (Borlase, 2017).

As indicated in the most recent analysis issued by Berg Insight on smart meter markets, the three markets that are leading the way in smart energy meter installation are Asia–Pacific (including Japan), Europe, and the United States. According to studies, the Asia–Pacific market (which includes China, Japan, South Korea, India, Australia, and New Zealand) would have 975 million smart meters, worth USD 142.8 million, by 2024, while the European market will have roughly 223 million smart meters. The amount of information accessible is staggering when you consider that a smart meter may record data at minute intervals.

To be more specific, the imminent arrival of massive amounts of information necessitates the development of tools in the fields of Big Data and Data Analytics in order to explore and evaluate this information (Adusumalli, 2016b). In this context, a number of writers have worked on a variety of applications that attempt to extract value from the raw data collected by smart meters or AMI assets. We may name a few of the most typical topics that involve advanced analytic approaches and large amounts of data in AMI data (Adusumalli & Pasupuleti, 2017).

AMI's data processing tools, as well as their interaction with new technologies, are discussed. This group includes work on platforms for storing and running analytic applications using a variety of programs, including Hadoop, MATLAB, MADlib, System C, Hive, and Spark Streaming, as well as other related work (Shyam et al., 2015; Achar, 2018a). Performance evaluation platforms, as well as other emerging technologies like cloud computing, real-time data processing, and fog computing, have also been investigated by some of the writers as well. Authors who are dedicated to establishing methodological ways for processing AMI data might also be included in this category (Daki et al., 2017).

NIST Big Data Reference Architecture

The majority of the time, reference designs serve as a template for generating solutions in an orderly manner in a certain field, and they can be utilized for comparison and alignment (Pasupuleti & Amin, 2018). As a result of pulling together similar characteristics identified in diverse documented case studies from throughout the world, NIST has determined on a particular architecture (Adusumalli, 2018). Aspects of the reference architecture depicted in Figure 1 include broad concerns about Data Analytics, including its implications and requirements, and other general considerations.

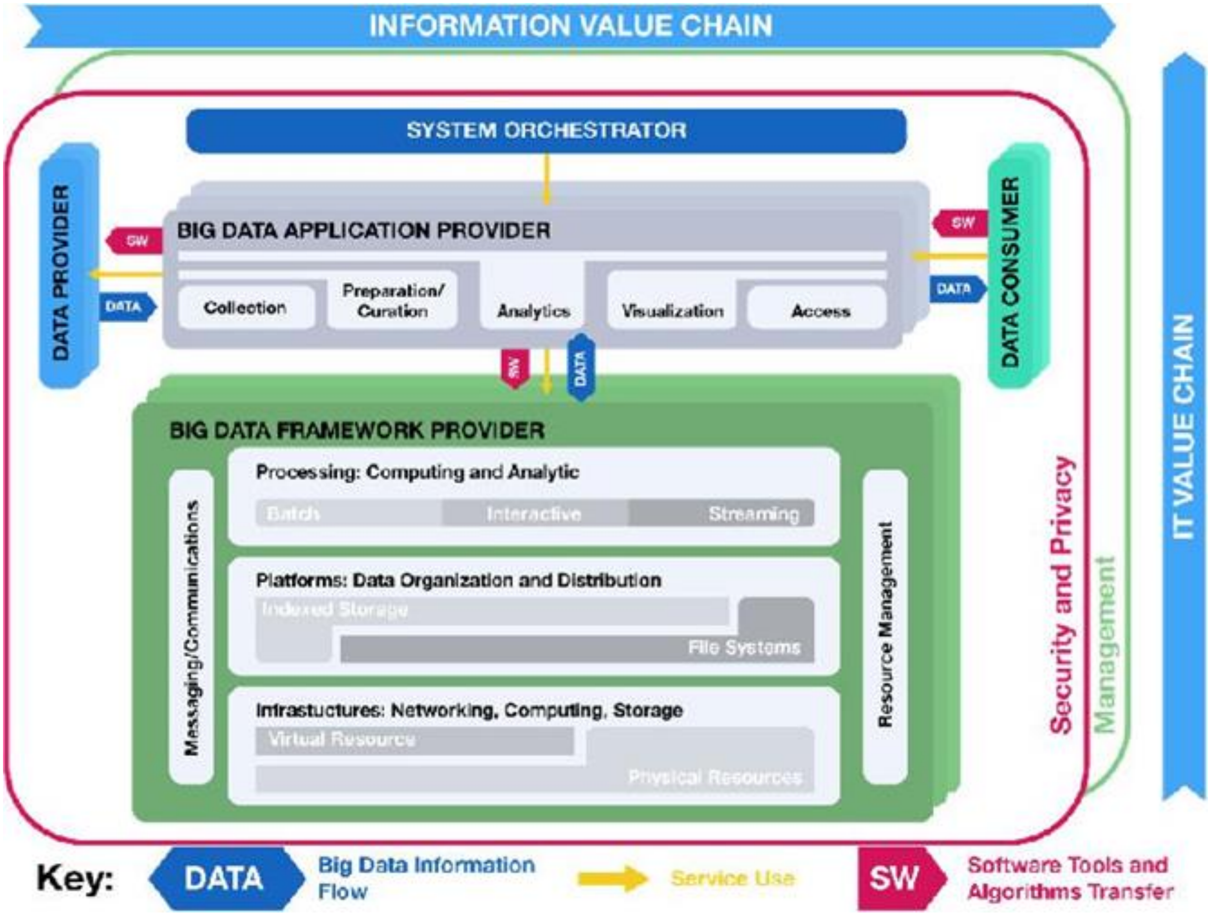

Figure 1: The National Institute of Standards and Technology's Big Data reference architecture.

The following key roles are proposed by the reference architecture.

- In a vertical operational system, the System Orchestrator defines and integrates the data application activities that are necessary into the operational vertical system. It identifies the overarching requirements for business ownership, governance, data science, and system design, as well as their respective implementations.
- Creating a computing framework that allows certain transformation applications to be executed while maintaining the privacy and integrity of data is what a Big Data Framework Provider is all about. The following resources or services are used by the Data Application Provider: Infrastructure framework (networking, computation, storage, and environmental), Data platform (physical storage, file systems, and logical storage), and Processing. Infrastructure framework: (software support for applications). This stage is the most sensitive to the nature of the data to be processed because it requires the most information. When compared to Data Analytics applications, a Big Data Analytics application may necessitate the use of a more advanced computer platform.
- Data Provider: a person or organization who adds fresh data (usually raw data) or information sources to the Big Data system, either online or offline. Also included are data permanence (hosting), data scrubbing (to eliminate PII–personally identifiable information), metadata (for history and repurposing), and policy for others' access to data, as well as querying without uploading any data (sometimes).

- Users or other systems that make use of the outcomes of the analysis are referred to as data consumers.
- Big Data Application Provider, search and retrieval, download, local analysis, reporting, and visualization of large amounts of data.
- Data Application Provider. This component manages the life cycle of data to ensure that security and privacy requirements are met. As part of this process, System Orchestrator-defined requirements are developed, as well as mechanisms for data capture and preparation, analytics (discovery for finding value in large volumes of data), visualization (exploratory, explicatory, or explanatory), and access to results from the data system. It is important to emphasize that this life cycle is relevant, in general, to both Big Data Analytic and Data Analytics applications, as previously stated.

Big Data Analysis varies from traditional data analysis in that it takes into account the volume, velocity, and variety features of the data under consideration throughout the analysis process (Pasupuleti & Adusumalli, 2018). The term is used in this context when we take the AMI data from the function layer (data from smart meters) in the SGAM and combine it with the requirements of the application. During this phase of the life cycle, information is gathered, prepared, analyzed, visualized, and made accessible.

Additional to this, the NIST reference model describes the Application Provider life cycle as having five broad stages:

- Collection: This stage is responsible for establishing a connection with the Data Provider and extracting the necessary information. Such information may be obtained from a variety of sources. This step might be referred to as the "extraction" component of the ETL (Extraction, Transformation, and Load) cycle because it involves the extraction of data.
- Preparation: During this step, we do all of the tasks necessary to make the data useful and available for analysis. • Analysis: During this stage, we conduct all of the essential analyses. Tasks such as data validation, cleaning, outlier elimination, and normalization are included in this category. A section of the ETL cycle known as "transformation" corresponds to this step.
- Analytical Techniques and Algorithms: During this stage, we implement all of the techniques and algorithms that are required to achieve the analysis goal defined by the application. It encompasses a variety of algorithms as well as statistics and machine learning methodologies. This stage is as sophisticated as the analysis requirement specifies it will be.
- It is at this point that we organize and deliver to our Data Consumer the pieces that have been derived from our analytics stage. Simple reports or even interactive applications for the end-user might be used as part of the visualization process.
- Accessibility: This stage is strongly tied to the visualization stage in terms of functionality. It is in charge of ensuring that the appropriate access is granted to the appropriate user. There are several approaches that can be used, including web services based on access roles or any technique that allows each user to get the results they require based on their role in the program (e.g., manager, operator, or supervisor).

Overall, these five processes constitute the translation of data into information that can be used by end users to their advantage. Another set of authors, on the other hand, has proposed an extended cycle (Bere et al., 2014), which aims to "organize the activities and tasks involved with the acquisition and processing of data," as well as "improve the efficiency of data acquisition, processing, and re-purposing," in order to "improve the efficiency of data acquisition, processing, and re-purposing."

In any instance, whether using the five phases suggested by NIST or the nine steps of the extended cycle presented by (Erl et al., 2016), the goal of this stage is to turn the data into useful outputs through the application of data analytic techniques and algorithms to achieve a specific purpose. Following that, the NIST reference model is introduced as a new component of the architecture suggested in this paper in Figure 2.

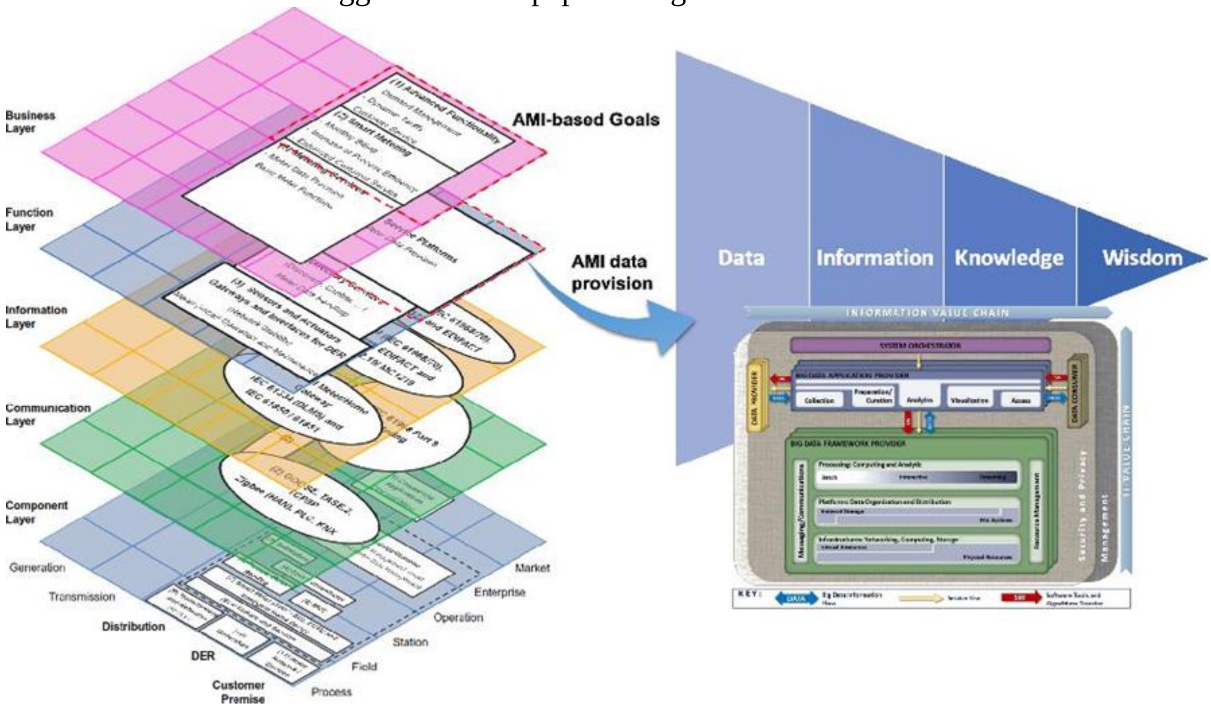

Figure 2: SGAM serves as the data provider architecture for the DIKW hierarchy, while the NIST model serves as the data transformation methodology for the hierarchy.

The architecture, business, and function layers of the AMI deployment over SGAM are depicted in Figure 2 as a relationship between the architecture, business, and function levels. As a result, the reference framework proposed by NIST serves as a mechanism for converting AMI raw data into knowledge, allowing for the advancement of the hierarchy proposed in the DIKW methodology for recognizing atypical consumption and identifying atypical consumption patterns.

An exploratory case focused on Big Data and Data Analytics was used to validate this framework in a prior iteration of the framework (Pasupuleti, 2016b). In that particular study example, we concentrated on the examination of electricity consumption data collected from smart meters in London between 2011 and 2014. By combining several data sources, we were able to gather more than 670 million data warehouse records, which were then processed and examined further. Load forecasting, customer clustering, and modeling were among the Big Data Analytics developments implemented in the prior research case, with one but more on Data Analytics due to the large number of records available (Adusumalli, 2017a). Currently, the REP is active in 30 marketplaces that are affiliated with two companies that are members of the same REP. Each market corresponds to a city in which one of the REP's enterprises has a presence, and vice versa. For the sake of this case study, Company 1 and Company 5 were designated as the names of the two companies involved.

Currently, the statistical department of the REP anticipates the total amount of energy consumed by its clients as a collective (Pasupuleti, 2015). However, when this department disaggregated the predictions by each market (city), they discovered that the Mean Absolute Percentage Error (MAPE) was 38 percent. With this case study, we will be implementing a prototype project to increase the demand forecasting of the REP's consumers, who will be grouped by market, namely by city and time intervals of 12 hours, in order to better serve them (a.m. and p.m. intervals). Due to the fact that it is an early pilot model, the implementation time for this case study was limited to three weeks.

SGAM Business and Function Layers

Similarly to the SGAM given in Figure 1, the problem outlined before defines the purpose of the study case, which is the function of the business layer: an application to improve the load forecasting of the company by market or city, which is the role of the business layer. Following the definition of the case study's goals, the function layer is responsible for ensuring that the platforms' functionality is maintained so that we may achieve the goals established in the business layer (Adusumalli, 2017b).

According to Uslar et al. (2013), this layer, in the context of AMI, includes service platforms that are responsible for the production of meter data for various roles such as distribution network operator, DER operator, and customer. It is in this sense that the REP supplied us with access to the information provided by two of its service platforms. The first platform (smart meter data platform) is responsible for delivering data connected to the metering infrastructure system, which includes data from smart meters deployed at each customer location.

The second platform (the customer data platform) is used to present information about the client. The goal described in the case study relates to the SGAM business layer (Pasupuleti, 2016a). In the same way, the data provider platforms are connected to the SGAM function layer. It is feasible to proceed to the data transformation step, as seen in Figure 1, after identifying the components of the business and function levels.

Transformation of Data into Information and Knowledge

To transform raw data into knowledge, we begin with the NIST reference framework, which is received from data platforms in the SGAM function layer, according to the DIKW model (Rowley, 2017). (Liu et al., 2017).

- System Orchestrator (also known as a system orchestrator). This function was carried out by the authors and the REP development team.
- Data Provider is a company that provides data. As specified in the SGAM function layer, the service platforms mentioned therein fulfill the job of the data provider. This function serves as a link between the architectural component (SGAM) and the methodological component (DIKW + NIST model) of the framework provided in this study.
- Data Consumption. This responsibility was taken on by the development department and the CEO of the REP. They are the ones who will be interacting with the results of the application.

- Data Application Service Provider (DASP). We designed and created all procedures associated with the Data Analytic life-cycle, which included activities such as data collection, preparation, curation, analytic, visualization, and access to results. This function is responsible for ensuring the transformation of raw data into knowledge and for overseeing the data life cycle, which was previously discussed.

Data Collection and Preparation

The conversion of data into information is the first transformation that takes place (Stoyanov and Kakanakov, 2017). This was accomplished through the completion of the first two tasks of the Data Application Provider, which are depicted in Figure 1: data gathering and data curation/preparation. This was accomplished by developing an Extraction-Transformation-Load (ETL) stage and writing several functions that were then used in conjunction with other functions from the Pandas Python package.

The Pandas library capabilities allowed us to connect to the two data systems mentioned above, which provided us with historical energy use data as well as information about our customers' preferences. The smart meter data platform produced one file for each month of historical data collected by the smart meter system (Fadziso et al., 2018). Active and reactive energy measurements were taken hour by hour, and we had access to a total of 1,949,337 records with these measures. For the purposes of the case study, only active energy use was taken into consideration. The information was organized into 56 fields, which included the customer ID, the date, the market (city), the REP Company, and active and reactive energy measures for each hour. The customer data platform offered information on 3534 customers and the features that characterized each of them. The information was organized into eleven fields, which included the customer ID, economic activity, contract-related information, electrical installation information, and market (city) for each client in turn.

As part of the development of a data warehouse, we used the Pandas library to create transformation functions that filtered corrupt or incomplete records for both customer records and consumption records (Pasupuleti, 2017). According to the REP aim, we only examined border (total) measurements and categorized consumptions by markets, rather than by consumers, in our analysis. The raw data from the smart meter data platform was supplied in a matrix-like data frame that contained the 56 fields listed above. We did, however, create 30 time-series data frames with only five fields: the REP's company ID, the market (City), the timestamp (in one-hour intervals), the type of day (business day, weekend, or holiday), and the amount of active energy spent at each time interval (in kilowatt hours). Later, we put the data frames in a warehouse with only the information needed to continue with the analysis that was requested.

Conclusion

The growth of the electrical grid in Smart Grids paves the path for the adoption of new infrastructures, such as AMI, in the future. Its deployment makes a large volume of data available, which rises at the same rate as the number of AMI project implementations. In order to take use of the availability of data from Smart Meters (AMI data), tools and platforms for processing, analysis, and utilization in fields of research such as Big Data and Data Analytics are required. In this regard, a number of authors have offered research methodologies as well as application scenarios. Applications include AMI data processing tools and their connection with developing technologies, load profile identification, load forecasting, demand response systems, as well as loss detection and prevention.

The implementation of a case study based on real data provided us with the opportunity to validate the framework. It is demonstrated in the case study that all of its components are critical in achieving outcomes that are beneficial to the functioning of a company in the electricity sector, in this case, a REP, at the worldwide level. The smart meter and customer data platforms served as data sources for the SGAM function layer, which was implemented as a function layer. The DIKW hierarchy received its data from the same source as it received its input. For the transformation of raw data into knowledge, we followed the NIST methodology: first, we constructed an ETL and a data warehouse; later, we used XGBoost to perform forecasting. In order to provide results to end-users, we used Tableau dashboards. Given the short development time, the pilot project conducted in this case study was able to lower the forecasting MAPE from 38 percent to 8.9 percent, a significant reduction.

Further research into the implementation of this framework in other Smart City scenarios, taking advantage of the abundant data available from multiple platforms, such as energy efficiency, smart mobility, smart metering (water and gas), and intelligent billing (among others), is being considered. This could result in resource optimization as well as a change in the operating dynamics of the entire Smart City concept. When it comes to Smart Cities, the overall goal should be the most optimal and efficient functioning of a system of systems, where the availability of data in real time has become a differentiating element in the marketplace.

It is also possible to explore incorporating various artificial intelligence approaches into the data transformation process in order to widen the scope of data transformation. We were able to accomplish this by incorporating other sources of information, such as data from social media networks, into our framework, allowing users of smart cities to have a stronger voice in the design of the city. Natural language processing (NLP) and emotional analysis are two methodologies that could be used to deal with this type of data, respectively. There is still opportunity for additional validations that include remote processing platforms or case studies that are more focused on data privacy and security, even if the study leaves the possibility of integrating alternative types of data processing in the paper up for discussion.

References

Achar, S. (2018a). Data Privacy-Preservation: A Method of Machine Learning. ABC Journal of Advanced Research, 7(2), 123-129. https://doi.org/10.18034/abcjar.v7i2.654

Achar, S. (2018b). Security of Accounting Data in Cloud Computing: A Conceptual Review. Asian Accounting and Auditing Advancement, 9(1), 60–72. https://4ajournal.com/article/view/70

Adusumalli, H. P. (2016a). Digitization in Production: A Timely Opportunity. *Engineering International*, *4*(2), 73-78. https://doi.org/10.18034/ei.v4i2.595

Adusumalli, H. P. (2016b). How Big Data is Driving Digital Transformation?. *ABC Journal of Advanced Research*, *5*(2), 131-138. https://doi.org/10.18034/abcjar.v5i2.616

Adusumalli, H. P. (2017a). Mobile Application Development through Design-based Investigation. *International Journal of Reciprocal Symmetry and Physical Sciences*, *4*, 14–19. Retrieved from https://upright.pub/index.php/ijrsps/article/view/58

Adusumalli, H. P. (2017b). Software Application Development to Backing the Legitimacy of Digital Annals: Use of the Diplomatic Archives. *ABC Journal of Advanced Research*, *6*(2), 121-126. https://doi.org/10.18034/abcjar.v6i2.618

Adusumalli, H. P. (2018). Digitization in Agriculture: A Timely Challenge for Ecological Perspectives. *Asia Pacific Journal of Energy and Environment*, *5*(2), 97-102. https://doi.org/10.18034/apjee.v5i2.619

Adusumalli, H. P., & Pasupuleti, M. B. (2017). Applications and Practices of Big Data for Development. *Asian Business Review*, *7*(3), 111-116. https://doi.org/10.18034/abr.v7i3.597

Bere¸ S. A., Genge, B., Kiss, I. (2014). A Brief Survey on Smart Grid Data Analysis in the Cloud. In Proceedings of the 8th International Conference on Interdisciplinary in Engineering—INTER-ENG, Tirgu Mures, Romania, 9–10 October 2014, 858–865.

Borlase, S. (2017). Smart Grids: Advanced Technologies and Solutions; CRC Press: Boca Raton, FL, USA.

Daki, H., El Hannani, A., Aqqal, A., Haidine, A., Dahbi, A. (2017). Big Data management in smart grid: concepts, requirements and implementation. *J. Big Data*, *4*, 1–19.

Erl, T., Khattak, W., Buhler, P. (2016). Big Data Fundamentals: Concepts, Drivers & Techniques, 1st ed.; Pearson: Upper Saddle River, NJ, USA.

Fadziso, T., Adusumalli, H. P., & Pasupuleti, M. B. (2018). Cloud of Things and Interworking IoT Platform: Strategy and Execution Overviews. *Asian Journal of Applied Science and Engineering*, *7*, 85–92. Retrieved from https://upright.pub/index.php/ajase/article/view/63

Liu, X.; Golab, L.; Golab, W.; Ilyas, I.F.; Jin, S. (2017). Smart Meter Data Analytics. *ACM Trans. Database Syst*, *42*, 1–39.

Pasupuleti, M. B. (2015). Problems from the Past, Problems from the Future, and Data Science Solutions. *ABC Journal of Advanced Research*, *4*(2), 153-160. https://doi.org/10.18034/abcjar.v4i2.614

Pasupuleti, M. B. (2016a). Data Scientist Careers: Applied Orientation for the Beginners. *Global Disclosure of Economics and Business*, *5*(2), 125-132. https://doi.org/10.18034/gdeb.v5i2.617

Pasupuleti, M. B. (2016b). The Use of Big Data Analytics in Medical Applications. *Malaysian Journal of Medical and Biological Research*, *3*(2), 111-116. https://doi.org/10.18034/mjmbr.v3i2.615

Pasupuleti, M. B. (2017). AMI Data for Decision Makers and the Use of Data Analytics Approach. *Asia Pacific Journal of Energy and Environment*, *4*(2), 65-70. https://doi.org/10.18034/apjee.v4i2.623

Pasupuleti, M. B., & Adusumalli, H. P. (2018). Digital Transformation of the High-Technology Manufacturing: An Overview of Main Blockades. *American Journal of Trade and Policy*, *5*(3), 139-142. https://doi.org/10.18034/ajtp.v5i3.599

Pasupuleti, M. B., & Amin, R. (2018). Word Embedding with ConvNet-Bi Directional LSTM Techniques: A Review of Related Literature. *International Journal of Reciprocal Symmetry and Physical Sciences*, *5*, 9–13. Retrieved from https://upright.pub/index.php/ijrsps/article/view/64

Rowley, J. (2017). The wisdom hierarchy: Representations of the DIKW hierarchy. J. Inf. Sci. 2007, 33, 163–180.

Shyam, R., Bharathi Ganesh, H. B., Kumar, S. S., Poornach, P., Soman, K. P. (2015). Apache Spark a Big Data Analytics Platform for Smart Grid. *Procedia Technol.*, *21*, 171–178.

Stoyanov, S., and Kakanakov, N. (2017). Big data analytics in electricity distribution systems. In Proceedings of the 2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 22–26 May 2017, 205–208.

Uslar, M., Specht, M., Dänekas, C., Trefke, J., Rohjans, S., González, J. M., Rosinger, C., Bleiker, R. (2013) Standardization in Smart Grids—Introduction to IT-Related Methodologies, Architectures and Standards; Power Systems; Springer Science & Business Media: Berlin/Heidelberg, Germany, p. 300.

--0--