International Journal of Advance Research in Science and Engineering
Volume No. 10, Issue No. 06, June 2021
www.ijarse.com
IJARSE
ISSN 2319 - 8354

AN EXTENDED TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (ETF-IDF) BASED DATA WAREHOUSE FRAMEWORK FOR DECISION MAKING

Usman Dahiru Haruna

Department of Information Technology,

ModibboAdama University of Technology, Yola, Nigeria.

&

Prof. Aminu A. Ahmad

Department of Management and Information Technology,

Abubakar Tafawa Balewa University Bauchi, Nigeria.

Abstract

*Organization are experiencing an exponential increase in data without right technology to handle the growth. though some organization uses the right mechanism for the storage known as data warehouse. but the data is increasing in developing nation organization with the exception of few of them with the right technologies known as data warehouse and data mining. this paper review, analyze data warehouse frameworks and proposed an extended term frequency-inverse document frequency (ETF-IDF) based data warehouse framework that will ensure quality data in to the warehouse for inform decision making. The paper recommends*full adoption and implementation of this framework as an integrated enterprise software for evaluation and success quantifiable matrix visibility and also further improvement should be done on the framework to accommodate other component of linguistics that serve as issue that reduce the quality of data in data warehouse.

Key word: *Data warehouse (DW), Extraction, Transformation & loading (ETL), DW Frameworks*

1.0 Introduction

The importance of data warehouse (DW) cannot be over stated in view of the exponential increase in size of data generated within and outside organizations for the purpose of providing matrix for making projections and decision making. For examples the activities and revenue of major e-commerce businesses such as Alibaba, Amazon, and Zalando indicate that the total online retail revenues in China, Japan, South Korea, India, and Australia will nearly double from US$733 billion in 2015 to US$1.4 trillion in 2020 [1].

So also, with the increased in social media data alongside Internet of Things (IoT), we witness an increasing explosion of stored and circulating data on the web, which produced a huge flow of data in various formats. These data can be structured, unstructured or semi-structured which at the end make complex for the current DW frameworks to handle. In the same vein there will be a significant increase in data that will be generated within and outside organization which will require an efficient data warehouse to handle the exponential growth. Current database tools and technologies cannot handle the load and analytic process of data into meaningful information. Organizations already collect vast amounts of data which has been growing significantly. These datasets are rich and growing and needs tools to produce insight from the records which is expected to support the decision-making process [2]. Therefore, it's against these background that this research seeks to analyze the available DW frameworks so as to explore the activities involve in integrating silos data sources that may be containing structure and unstructured data set and the activities involve in ETL process for better and qualitative data in a data warehouse to guide in making informed decision at both strategic and tactical level management decision levels.

1.1 Statement of Problem

Traditional data warehouses are unable to meet the growing needs of the modern enterprise to integrate and analyze a wide variety of data generated by social, mobile and sensor sources for making projections and decision making [3] Even though ETL is the main process in traditional data warehouse technology but cannot handle unstructured data [2]. Several attempts were made to build data warehouse framework to analyze such growing unstructured data to make informed decision. [21] in trying to improve data warehouse performance, he integrates Query Cache method to enhance the performance of ETL processing and minimize response time, in the same vein data coalition rule which applies in the proposed data warehouse to detect errors, dirty data and faults that could exist therein.

[14] in an attempt to create better DW mechanism for businesses, realized current Extract-Transform-Load (ETL) tools are not suitable because they do not consider semantic issues in the integration process. ETL tools did not support processing semantic data in order to create a semantic DW, which lead to the need for building a new DW that will integrate internal and external data published in different formats which requires semantic integration. A programmable framework was developed, named Semantic ETL (SETL) that facilitates users to build a semantic data warehouse but the framework fail to set up basic semantic ETL operators that can be combined to perform any semantic ETL operations to explore the Link between the internal and external sources because the framework did not consider issues like duplication and context of utterance.

[24] in their ability to bring high data quality to the data warehouse from both internal and external sources using the ETL process identified two Data quality approaches in the ETL process: process centered approaches and data centered approaches within the data centered approaches three component of linguistics have been highlighted: syntax, semantics and pragmatic. Where in [14] framework consider semantics component of linguistics. In an effort to improve the existing data quality in DW frameworks this paper proposed an improved data warehouse framework

that will integrate both syntactic and semantic linguistic component to ensure correct and quality data in to the warehouse for making inform decision.

1.1 Aim and Objectives

The main aim of this study is to design an improve data warehouse framework for inform decision making. To achieve these, the paper breaks the aim in to the following specific objectives as follows:

1. Perform a systematic evaluation of existing data warehouses for the proposed framework.
2. Identify suitable ETL syntactic mechanism to ensure data quality in the proposed framework
3. Propose an improve data warehouse framework by integrating the ETL mechanism identified

2.0 Data Warehouse Concept

DW is an integrated and historized collection of data generally used to make strategic decisions by means of Online Analytical Processing (OLAP) techniques. It integrates data that come from independent heterogeneous operational data-sources and create a single view about the organization [3] It's important to know that database computing has shifted its focus from operational to decisional concerns. This change in computing focus has become the impetus for the development of data warehousing technologies.

Similarly, DW is viewed as the combination of concepts and technologies that facilitate organizations to manage and maintain historical data obtained from operational and transactional applications, it also helps knowledge workers (executives, managers, analysts) to make quicker and more informed decisions [2]. DW is created within an organization as a separate data store whose primary purpose is data analysis for the support of management's decision-making processes [4] Often, the same fact can have both operational and analytical purposes. DW data is subject oriented, non- volatile, time variant and integrated data and it consists of collection of data for decision support system which has one-level, two level and three level architecture [5].
[6] is of the view that it is a kind of management technique that collect business data from different stations of the enterprise network, so that it can provide efficient data analysis to decision makers. There are some architectural requirements which would govern development of architecture, some of them are: identifying potential users, defining security requirements, skill requirements etc. from the conceptual views above, we can deduce that. Data warehouse referred or entails the same meaning by the different authors. Because the central aim attached to all the views is to have huge information that will be analyze and aid in sound decision making.

2.1 Data warehouse architecture

[2] introduce data warehouse architecture which could substitute traditional data warehouse in educational system to reduce difficulties associated with traditional data analysis which will have potential of enriching the education system with new learning ways, and making decision making by policy makers more effective and efficient. In the same vein [7] are of the view that Data Warehouse architecture provide integrated environment from the database which does not produce any data but can store data from various databases. It consists of 4-layers. Database, Data

International Journal of Advance Research in Science and Engineering
Volume No. 10, Issue No. 06, June 2021
www.ijarse.com
IJARSE
ISSN 2319 - 8354

Staging, Data presentation and various data mining tools. they further said, Database is nothing but a source for the data which consists operational source systems and from other sources like Oracle, SQL. Data staging area is a Data Warehouse environment which includes ETL which is an Extraction, Transformation, and Loading (ETL) of data. Data Presentation layer includes OLAP and a Data Warehouse. While [5] are of the view that, Data Warehouse consists of tool like query based from the data base, tools for reporting and analysis. The general data warehouse design which include bottom tier, front and back room as can be seen in fig. 1 below

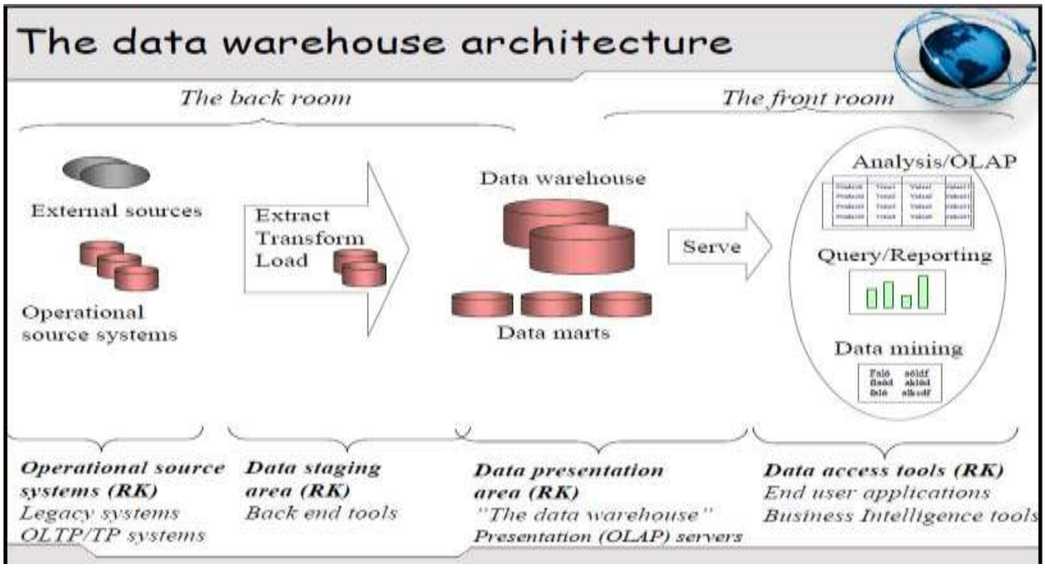

Figure 1: data Warehouse Architecture

Source: [5]

From the architectural view in relation to data warehouse, we can clearly see that a Data Warehouse consists of 4-layers. Database, Data Staging, Data presentation and various data mining tools as viewed by other scholars while [5] present the same layers in an encapsulated concept namely: Bottom Tier; Back Room; The Front Room. And also, backup their opinion with a design as can be in seen in figure 1 which clearly depict the actual data warehouse architecture.

2.2Data warehouse failure and success factors

[9] in trying to assess the failure factors of warehouse productivity in the Malaysian logistics service sector, by applying a fuzzy analytic hierarchy process method identified labor productivity, warehouse utilization and inventory space utilization. In the same vein [10] identified four dimensions in order to cope with potential failures: operational, organizational, decision-making, and cooperative dimensions. In other hand, many researchers identified their own lists of success factors. For example, easy to use and speedy information retrieval while [11] presented three (3) data warehouse success factors as clearly defined business needs/benefits; top management support; and user involvement/participation, in the same vein,[12] identified eight comprehensive data warehouse

success factors which virtually include all what the first two authors mentioned, they include: Organization Culture, Technical Tools, Management Support, User Involvement, Quality of Data Sources, Self-efficacy, Knowledge Sharing, Clear Objective, Scope and Goals and this research tend to be in line with them because they are comprehensive and seem complete

2.3Extraction, Transformation and Loading (ETL)

[13] is of the view that ETL are set of processes for getting data from systems into a data warehouse while [14]is of the view that Extract-Transform-Load (ETL) is the backbone process of a DW, and the ETL design and deployment takes up to 80% of the time in DW projects. ETL processes physically integrate data from multiple, heterogeneous sources in a data warehouse. ETL tools are pieces of software responsible for the extraction of data from several sources, cleansing, customization and insertion of the data into a data warehouse. The extraction phase largely deals with the technical heterogeneity of the different sources and imports relevant data into a staging area. The transformation phase is the heart of an ETL process whereby all the data is brought into a common data model and schema using mapping technology, standardized consolidated to a single representation. In the load phase, the integrated, consolidated, and cleaned data from the staging area is loaded into the appropriate databases/data marts of the warehouse [15].[16] as oppose to [15] looked at ETL as a process, not a physical implementation. ETL systems differ from data warehouse to another data warehouse and even sandwiched between department information marts within a data warehouse, ETL process is not a one-time event, because ETL is an integral, ongoing and recurring part of a data warehouse. they further assert that ETL system contains four discrete functional fundamentals: Extraction, transformation, loading and meta data.

Hence, ETL removal constituent is in charge for extracting data commencing the supply system. During removal, data may be removed from the resource system or a copy made duplicate and the original data retain in the source system as perceived by [16] So also, extraction phase largely deals with the technical heterogeneity of the different sources and imports relevant data into a staging area waiting to be transform as perceived by [15] So also, transformation phase is the heart of an ETL process whereby all the data that's extracted by the previous process will be brought into a common data model and schema using mapping technology, standardized consolidated to a single representation. [15] In other perspective transformation elementis responsible for data validation, data accuracy,data type conversion, and business rule application [16]

[16]looked at Loading component as one responsible for putting transformed data into the data warehouse. The integrated, consolidated, and cleaned data from the staging area is loaded into the appropriate databases/data marts of the warehouse [15]. The ETL meta data functional element is responsible for maintaining information about the movement and transformation of data and the operation of the data warehouse. It also documents the data mappings used during the transformations. Meta data logging provides possibilities for automated administration, trend prediction, and code reuse [16]From the foreseeing, [15] presented a view about ETL concept but [16] did not stop at the level of first author, they further extend their discussion with a particular concept meta data which keeps log

of what happens within the ETL process. But they are all of the view that, ETL is an indispensable component of data warehouse, because data warehouse cannot stand as it is with the processes imbedded inside ETL activity.

2.4Reviews on data warehouse frameworks

[17] affirmed the volume of generated and stored data from social media has been increasing knowing that analyzing and understanding this kind of data can offer relevant information in different contexts and can assist researchers and companies in the decision-making process. However, the data are scattered in a large volume, with different sources and formats. He then developed a social media data integration model based on a data warehouse to reduce the computational costs related to data analysis, and techniques to discover useful knowledge. The results showed that the proposed data warehouse improves the quality of data mining algorithms, while being able to reduce the execution time by avoiding redundant data in to the DW.

[2] introduce big data in to data warehouse for academic purpose knowing ETL is the main process in traditional data warehouse technology but cannot handle unstructured data. Hence, the need for flexible ETL process which can handle several data quality issues, e.g. duplicate data, inconsistency and garbage data. Where they combined Hadoop and RDBMS as data ingestion/staging tool, but also as data management and data presentation platform though it improves the decision variable by reducing duplicated data, inconsistency data, and garbage data but there is need for OLAP Navigator and MDX Query Editor to create powerful report.

[3] in a similar research, observed that Traditional data warehouses are unable to meet the growing needs of the modern enterprise to integrate and analyze a wide variety of data stored in NoSQL databases, data warehouse schema from document-oriented databases was designed which used DBLP" and the" LINKEDIN". But the framework cannot aggregate-oriented databases such as: Column Oriented Database and Key/Value Database. Unlike [18] that construct a data warehouse accessing framework for campus power usages, Internet of Things (IoT) technology, was used to analyze and real-time monitor the power consumption data of buildings or equipment handled by Hadoop ecosystem. And Hive as a data warehouse, Spark as a data ETL tool, and Impala as a big data search engine as a multi-layer architecture. Yet cannot handle semi structured and unstructured data and there is need to append more hosts and monitor the condition of each host to enhance the robustness of system.

[14] in a similar work in order to create better means of decisions making for business realize Current Extract-Transform-Load (ETL) tools are not suitable for such effort because they do not consider semantic issues in the integration processing. the tools neither support processing semantic data nor create a Data Warehouse (DW) that requires semantic integration. a programmable framework, named Semantic ETL (SETL) that facilitates users to build a DW. SETL uses T-Box as an underlying schema to integrate heterogeneous data sources was developed by (Rudra et al., 2017). SETL was builds on Semantic standards. Which supports semantic data sources in addition to traditional data sources, semantic integration, and creating or publishing a semantic DW in terms of a knowledge base. A comprehensive experimental evaluation comparing SETL to a solution made with traditional tools on a concrete use case, shows that SETL provides better programmer productivity, knowledge base quality, and

performance. A well-defined set of basic semantic ETL operators that can be combined to perform any semantic ETL operations Besides, developing techniques to explore the Link between the internal and external sources should be considered.

[19] similarly, build computational framework for the management of BD using binary classification, by subjecting what customers are saying about the firm through reviews in social media content, with the best results being found for binary classification. But the models can not consider unstructured data and other rating options. because the cleaning was not done properly because SQL cannot remove noise like (special characters, blanks, nulls) for predictions. [20] as opposed to Jose et al. build framework for querying structured and unstructured data in medical field using Hadoop approach which provide unlimited query support and efficiency in processing unstructured data. The framework is efficient that it utilizes the distributed computing power of Hadoop clusters. The study showed that efficiency is gained by separating the query into two phases (structured and unstructured). But it fails to extend processing time by handling the variety of the data in different stages.

[21] recognized that ETL is the lead procedure to fetch all the data in the form of homogenous and standard environment. They incorporate “Query Cache method” to enhance the performance of ETL processing and minimize response time and data coalition rule to detect errors, dirty data and faults that could exist therein, but the framework has some limitation of not being able to identify duplicity error (example DOB and date of birth) and date time format.

[22] in the same vein to improve virtual data warehouse using Efficient Data Access (EDA) offers several advantages such as real-time analytic reports and reduced maintenance among others. The evaluation of the framework revealed that the use of virtual query processing suffers from certain drawbacks including extra load on the OLTP system and substantial increase in the overall query response time and need for implementing VDW environment in the distributed systems. But [23] looked at it from a different dimension by proposing a new network-based model to uniformly represent and handle structured, semi- structured and unstructured sources of a data warehouse. However, they are generally incapable of managing unstructured data and are not lightweight and flexible enough to be used in the improve data warehouse context.

3.1 Decision making

[25]One of the most important functions of a manager is to take decisions in the organization. Success or failure of an organization mainly depends upon the quality of decision that the managers take at all levels. The quality of managers’ decisions is the Yardstick of their effectiveness and value to the organization though depends on the quality of data provided to him that will serve as a matrix for the decision. while [26]looked at decision making as the process of responding to a problem by searching for and selecting a solution or course of action that will create the most value for organizational stakeholders or process of identifying and solving problems.

International Journal of Advance Research in Science and Engineering
Volume No. 10, Issue No. 06, June 2021
www.ijarse.com
IJARSE
ISSN 2319 - 8354

3.2 Data Warehouse and Decision Making

In this modern era of globalization and high level of integration, new techniques, models and tools supporting the concept of Business Intelligence (BI) by providing easy and quick access to information and knowledge, real time scenarios, data visualizations and dashboards, summary reports, and other analytical tools for data, text and web mining have changed the way business make decision. Because data and information are treated as an important asset in an organization to impart quality and summarized knowledge for valuable, quick and effective decisions. Data warehouse give organizations access and privilege to access the data but do not guarantee the integrity of the data and adequacy of response time. But it tries to solve such problems by providing technology which enables the user or decision maker to process the huge amount of data in a short amount of time. through extracting the knowledge in a real time and its help the user in decision making. Data warehousing alongside data mining provide the right foundation for building decision support and executive information system tools which help to measure the progressing speed of organization toward its goal, provide a technology that enables the user or decision maker in the corporate sector/government to process the huge amount of data and make decisions which are useful for whole organization [27].

DW is a change agent in business trends these days. but small and big organizations are collecting and using data from various sources through DW concept to identify their own business trends, understand their strengths and weaknesses of their competitor improve their progressing speed towards the goal and expand their business. Furthermore, it forms an integrated environment where data from disparate systems is brought together and presented in a consistent manner to make key decisions. In order to support these decisions, quality of data in warehouse must be reliable. If there are issues in data loaded in warehouse, business users lose trust and the information from the warehouse becomes unreliable [28] but [29] are of the view that the best decisions are made when all the relevant data available is taken into consideration. The best possible source for that data is a well-designed data warehouse. it provides leverage for management in an organization to make effective decision at all three managerial levels (strategic, tactical and operational [30]

3.3 Result and Discussion

[31] are of the view that the previous two methods (soft TF-IDF and TF-IDF) concentrate on two different causes of record duplication, namely typographical error and varying word order. However, in case both types of error occur this research adopts the method that extend TF-IDF method by [32] to address two common situations in duplicate detection: sparsity due to missing entries and large numbers of duplicates.

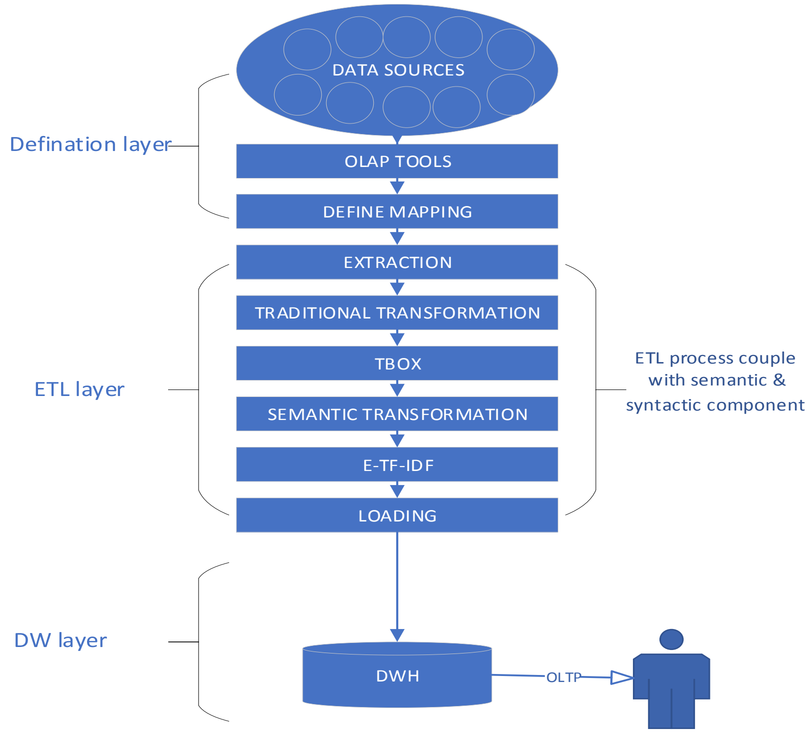

Figure 2. Proposed Syntactic-Semantic ETL DW framework for decision making

The definition layer integrates different components required to define the schema of an SDW, to define the different sources that feed data in the SDW, and to define the mappings between the sources and the target. The data in an SDW are semantically connected with other internal and/or external data. Mappings define the relationship between the elements of a source and the corresponding elements in the target T-Box. the ETL component handle the extraction of the data from discrete data sources in to single schema. Traditional Transformation includes operations known from traditional ETL tools, such as cleansing the data and formatting the source data according to the target schema. Then Semantic Transformation includes operations to create RDF triples according to the semantics of the target T-Box from the data output by the Traditional Transformation component. And lastly the E-TF-IDF includes removing duplicate data, recalculating data, normalizing data, renaming attributes, checking integrity constraints, refining data, unique identifier generation, creating new attributes based on existing attributes, null value and de fault value handling, noisy data filtering, sorting data, and grouping/summarizing data. After such processes the data will be load in to the data warehouse to be subjected in to any data miming algorithms depending on the reason for the mining by either data scientist or data analyst.

4. 0 Conclusion

Building a data warehouse that will integrate internal and external data published in different formats requires semantic integration. Here, we have proposed framework, named Syntactic-Semantic ETL (SETL) that facilitates users to build a DW. SSETL uses T-Box as an underlying schema to integrate heterogeneous data sources and E-TF-IDF to complete the data cleaning process to have a super clean data to make inform decision that will help organization in achieving their organizational objectives and goal and also survive within the todays competitive market alongside knowledge/technology driven economy.

Thus, it can be concluded that Data Warehouse is important for the success of an organization as they combined different data sources into one unified source and gives fruitful dashboards and reports thereby provide variable insights for better decision making, revenue enhancement, cost reduction, summary reports on business operations, better analysis of market positioning, segmentation and competitive intelligence. They deliver outputs at fast speed, lower cost and higher quality than the competitors thereby enabling competitive advantage to the organizations.

5.0 Recommendation

In an effort to improve the existing data quality in DW frameworks this paper recommends full adoption and implementation of this framework as an integrated enterprise software. for evaluation purpose and success quantifiable matrix visibility at the same time further improvement should be done on the framework to accommodate context of utterance (pragmatic) as a component of linguistics that serve as issue that hinder the quality of data within data warehouse.

References

[1] L. Hao, Y. Xuan and T. Xiang, A synchronized production-warehouse management solution for reengineering the online-offline integrated order fulfillment. Journal of T*ransportation Research Part E (122)*, 2019, 211–230.

[2] W. Leo and K. Yulia, Data warehouse with big data technology for higher education.
4th Information Systems International Conference Bali, Indonesia, Procedia Computer Science, *124*, 2017, 93–99.

[3] B. Senda, N. Ahlem and G. Faiez, Design a data warehouse schema from document-oriented database. *23rd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, Procedia Computer Science 159*,2019, 221–230.

[4] S. Sivaranjani, S. Sivakumari and M. Aasha, Crime prediction and forecasting in tamilnadu using clustering approaches. *IEEE International Conference on Emerging Technological Trends, 978*(1), 2016, 13-19.

[5] M. Achudhan and J. Prem, Mathematical modeling and control of an electrically heated catalyst. *International Journal of Applied Engineering Research*, 9(23),2014, 13-23.

[6] M. Noor, H. Siti, M. Rosmayati, A. Masita and S. Muhammad, A review on a classification framework for supporting decision making in crime prevention. *Journal of Artificial Intelligence 8(1)*, 2015, 17-34.

[7] S. Inderpal, Study of data warehouse architecture. *International Journal of Engineering Sciences & Research Technology, 2(7*), 2016, 1675-1680.

[8] T. Seema and J. Snehal, The study on data warehouse design and usage. *International Journal of Advance Research in Computer Science and Management Studies*, 4(5), 2016, 21-23.

[9] H. Nur, S. Noorul and F. Syed, Empirical evidence on failure factors of warehouse productivity in Malaysian logistic service sector, *The Asian Journal of Shipping and Logistics, 34(2)*, 2018, 151-160.

[10] Z. Manel Z. Sahbi and G. Faiez, Specification of the data warehouse for the decision-making dimension of the Bid Process Information System. *Procedia Computer Science 159*, 2019 1190–1197.

[11] D. Shobha, Key success factors for data warehouse implementation: analysis. *Oriental Journal of Computer Science & Technology,7*(*2)*, 2014, 231-236.

[12] P. Thiti and R. Siriluck, Data warehouse implementation success factors and the impact of leadership and personality on the relationship between success factors. *Journal of Business and Economics, 4(10)*,2013, 948-956.

[13] A. Ardhian, Extract transform load (ETL) process in distributed database academic data warehouse. *Journal on Computer Science and Information Technologies. 4*(2), 2019, 64-71.

[14] P. Rudra, H. Katja, B. Torben and R. Oscar, SETL: A programmable semantic extract-transform-load framework for semantic data warehouses. *Information Systems 68*, 2017, 17–43.

[15] M. Christopher, M. Rahab and N. Njeri, A case for judicial data warehousing and data mining in Kenya. *American Journal of Computing Research Repository, 4(1)*, 2016, 7-14.

[16] H. Gajareand S. Rangdale, ETL Data conversion: extraction, transformation and loading data conversion. *International Journal of Engineering and Computer Science, 6(10)*, 2017 22545-22550.

[17] C. Carlos, M. Martha, A. Erwin, M. Manuel and L. Elizabeth, Multidimensional analysis model for a document warehouse that includes textual measures. *Decision Support Systems 72*, 2015, 44–59.

[18] H. Chih, C. Fuu, T. Chao and C. Sheng, on construction of a big data warehouse accessing platform for campus power usages. *Journal of Parallel and Distributed Computing 133*, 2019, 40–50.

[19] J. Jose, G. Israel, L. Jose and R. Belen, Towards a big data framework for analyzing social media content. *International Journal of Information Management. 44*, 2019, 1-12.

[20] I. Sarmad and R. Mohammad, Unstructured medical image query using big data -an epilepsy case study. *Journal of Biomedical Informatics 59*, 2016, 218–226.

[21] P. Tiwari, A. Mishra, V. Kumar and B. Terfa, Improved performance of data warehouse. *International Conference on Inventive Communication and Computational Technologies* 2017.

[22] A. Fakhri, A. Awais, I. Muhammad, A. Mafawez, R. Mujeeb and J. Bilal, Efficient data access and performance improvement model for virtual data Warehouse. *Sustainable Cities and Society. 35*, 2017, 232–240.

[23] P. Giudice, L. Musarella. G. Sofoand D. Ursino, An approach to extracting complex knowledge patterns among concepts belonging to structured, semi-structured and unstructured sources in a data lake. *Journal of Information Sciences,478*, 2019, 606–626.

[24] M. Souibgui, F. Atigui, S. Zammali, S. Cherfi and Y. Ben, Data quality in ETL process: *A preliminary study. Procedia Computer Science. 159*, 2019, 676-687.

[25] N. O. James and Edwin A. Effective decision-making and organizational goal achievement in a depressed economy. *International Journal of research and Development Studies. 8(1)*, 2017.

[26] A. L. Mark and D. S. David. Organizational decision-making models: comparing and contrasting to the stinson wellness model. *European Journal of Management*, *14(3*), 2018, 13-28.

[27] P. Monika, S. Sukhdev and S. O. Sukhwinder. Impact of data warehousing and data mining in Decision making. *International Journal of Computer Science and Information Technologies. 4(6)*, 2016, 995-999.

[28] G. Rahul. Maintaining data quality in data warehouse. *International Journal of Scientific and Research Publications. 6(12)*, 2016, 283-285.

[29] I. Nwakanma, O. Egbivwie, U. Azike, A. Nwaobilo, C. Ibeji and N. Ohia. The role of data warehousing concept for improved organizations performance and decision making. *International Journal of Computer Science and Mobile Computing. 3(10)*, 2018, 451-455.

[30] V. Palak**, S.** Saurabh and S. Swapna. Business intelligence: escalation of data warehousing and data mining for effective decision making. *International Journal of Advanced Science and Technology*, *29(5)*, 2020, 1377-1388.

[31] S. Sadiq, and M. Indulska. Open data: Quality over quantity. *International Journal of Information Management. 37*(2), 2017, 150-154.

[32] V. G. Yves, H. Blake, M. Anna, M. Daniel, V. Ryan and L. B. Andrea. Unsupervised record matching with noisy and incomplete data. *International Journal of Data Science and Analytics IJDSA,6(2)*, 2018, 109–129.