A CONCEPTUAL FRAMEWORK FOR A MULTIDIMENSIONAL MODEL OF TALENT MANAGEMENT DATA WAREHOUSE

AZWA ABDUL AZIZ, MUHAMMAD SANI ILYASU, IBRAHIM ABDUL KARIM MAHMOUD, RAJA HASYIFAH RAJA BONGSU, JULAILY AIDA JUSOH

Faculty of Informatics and Computing Universiti Sultan Zainal Abidin, Tembila Campus, 22200 Besut, Terengganu, Malaysia
E-mail: azwaaziz@unisza.edu.my, msnailiyasu@gmail.com, rajahasyifah@unisza.edu.my, julaily@unisza.edu.my

ABSTRACT

The modern business environment and dynamic economic situation has made the workforce and business ethics more informed, versatile and sophisticated. This development forges a complex business atmosphere that compels industries to compete productively for sustainable growth. Thus, it becomes imperative for organizations to manage their talents (workforce) effectively to achieve sustenance in the modern economic climate. Regrettably, it has been challenging for HR managers to identify, select and retain competent individual’s that suits their industrial needs (talent management). Therefore, this paper proposes a conceptual framework for a Multidimensional Model of a Data Warehouse (DW) for Talent Management (TM). Further, a hybrid approach of multidimensional modeling will be adopted in developing the model of the DW. In addition, student personal information and academic performances from institute of higher learning across Malaysia will be used as the data sources. Similarly, industrial needs (job vacancies) outlined by Multimedia Development Corporation (MDEC) Malaysia will be considered as user requirement. The proposed DW will provide information that will facilitate direct mapping of candidate to industrial needs among other TM practices. Also, the DW will facilitate various TM related analytics using appropriate Business Intelligence tool.

Keywords: Data Warehouse, Multidimensional Modeling, Star Schema, Business Intelligence, Talent Management.

1. INTRODUCTION

The technologically advanced and dynamic modern economic situation has made the workforce and business ethics more informed, versatile and sophisticated. Consequently, the business environment becomes so complicated that organizations have to compete for sustainable development. Thus, it becomes imperative for organizations to develop effective human resource activities to manage their talents efficiently. Industries of all scale and sizes recognize effective talent management as a robust tool for competitive advantage [1]. TM can be simply described as a forecast for human capital demands and planning how to meet the demand [2]. TM practices consume a significant amount of organizational resources. According to Economist Intelligence Unit (2006) various stakeholders significantly shows concern to the issue of talent management beyond HR academics and professionals [3]. Unfortunately, it has been challenging for HR managers to identify and select and retain competent individual’s that suits their industrial needs (talent management) [1]. Thus, it becomes imperative to develop a centralized repository of information about industry relevant talents to facilitate effective Talent Management (TM). This is because according Collins et al. (2009) development of talent pool is among the three vital phases of TM [4].

Therefore, this paper proposes a conceptual framework for a Multidimensional Model of Talent Management Data warehouse. DW has proved to be worthy in business analyses and decision-making processes. Thanks to its features that improves identification, coordination, and utilization of information asset. Further, in DW development, user requirements often determine the information demands to be met and decide the structure of the target DW [5]. In this regard,
MDEC jobs with highest vacancy as at 2013 will be considered as the requirement. Also, academic data of graduates from various IHLs across Malaysia will be used as the data source. Moreover, the collected data will be analyzed in accordance with multidimensional modeling guidelines. This is to identify the multidimensional concepts within the data. And eventually, through step by step guide the model of DW will be built based on these concepts in consideration of the user requirement.

However, DW development involves extraction of information from multiple and usually heterogeneous sources. This phase is very vital in DW development and it determines the quality of data stored in the DW. Also, it involves a lot of transformation and cleansing of data before it is finally loaded into the DW. This operation is achieved through special data integration tools [6, 7].

Finally, this DW will serve as a consistent source of quality assured information of talents relevant to the industrial needs. This information will facilitate stakeholders’ decisions regarding TM practices. And also, it will facilitate various TM related analytics with the aid of appropriate Business Intelligence tool. However, this study is limited to recruitment aspect of TM and only considers ICT jobs and candidates.

2. DEFINITION OF DATA QUALITY

2.1 Data Warehouse (DW)

Information is considered as organizational greatest asset on which the success of the organization in gaining competitive edge depends. As a result of the advancement in information technology, storing organizational data is no longer a problem irrespective of its nature or magnitude. However, accessing such data easily and harness its value by the organization is difficult. This difficulty can be attributed to the heterogeneity in format and diversity in sources of the data [7].

However, the advent of relational databases in the early 1980’s provides enhanced access to the vital information buried deep within a data. But unfortunately, that did not suffice, because databases were model for smooth transactional processing. And they were not always optimized to provide sophisticated reporting and analytical services that will facilitate strategic decisions of the management [8].

The evolution of the concept of the practice known today as Data Warehousing dates back to late 1980s, [10] invent the term “Business Data warehousing”. The concept was borne out of the need for easy access to consolidated and quality assured organizational data that can be used for management strategic decisions [8].

According to Inmon DW can be described as a subject oriented, integrated, time-variant and non-volatile collection of data in support of management’s decision-making [9]. Inmon is considered as a father of data warehousing. Further, in 1990s Inmon lays the theoretical underpinning of the data warehousing concept. Also, he develops the DW architecture called Corporate Information Factory. This architecture has the enterprise wide view of company’s data and follows top-down approach of DW design. (That is, the DW is designed as a centralized repository for the whole enterprise. Therefore, dimensional data marts containing data needed for specific departments are derived from the enterprise DW only after the DW has been completely developed). To date Inmon remain one of the major players in data warehousing [10].

Another key player in the field of data warehousing is Ralph Kimball. In 1996 Kimball publish his first book about data warehouse titled The Data Warehouse Tool Kit. This book covers industry-oriented, practical examples for OLAP-style modeling of Data Warehouse [11]. Kimball is an advocate of multidimensional approach of building data warehouse. Contrary to Inmon, Kimball view data warehouse as a collection of departmental individual data marts, integrated together to form the data warehouse. His methodology indicates bottom up approach to building data warehouse in which autonomous data marts for each department are firstly built and later combined into a larger data warehouse [10]. Moreover, according to Kimball DW can described as a decision support tool that obtain its data from various heterogeneous sources and transform it into meaningful information for decision making [5, 10].

2.1.1 Data warehouse Architecture

DW is made of the following major component

**Data source:** these are the sources of DW input data usually operational systems of records that capture business transactions [12].

**Data staging area:** is the part of DW between Operational sources and Data presentation area where the vital Extraction, Transformation and Loading process (ETL) take place. This area does
not produce query for business presentation services and is off-limit to users [6, 13].

**Data presentation area:** this is the part where processed data is organized, stored and made available for direct query by users [6]

**Data access tools:** these are capabilities that can be provided to business users to enable them exploit the presentation area for analytical decision making. E.g. Business intelligent tools [13]

**Data Mart:** Is a subset of data warehouse usually containing data or information of a particular business line that serves as a medium through which users access the data warehouse [6].

**Meta Data:** is a data that describe other data [15]. The figure 1 shows the architecture of the proposed data warehouse model [13].

![Figure 1 Architecture Of The Proposed TM DW](image)

**2.1.2 Goal of Data Warehousing**

The fundamental goal of data warehouse is to provide a secure, adaptive and resilient repository of consistent and easily accessible organizational information that can serve as the bases for improved decision making [6].

**2.2 Multidimensional Model (MDM)**

Since late 19th century, Multidimensional database that originates from multidimensional matrix algebra has been used for manual data analysis. But, the practice known today as MDM was first introduced in [6] where the first methodology to drive data warehouse logical schema was introduced [5, 10].

MDM is a data model that structure data into facts (the subject of the analysis) and dimensions (the different perspectives to view and analyze the subject). And represent it in a star or snowflakes schema. However, the dimension, dimension level and dimension descriptors constitute what is referred to as dimensional concepts. These dimensional concepts form the n-dimensional data space in which the fact is placed, viewed and analyzed from various perspectives [5]. Further, the fact/dimension dichotomy forms the basis of multidimensionality that gave birth to the multidimensional structure known as Data Cube [6, 10]. The figure 2 illustrates a multidimensional view of sales data.

![Figure 2 Multidimensional Cube](image)

MDM is widely accepted as the basis of data warehouse design as it provides comprehensive and easy to understand visualization of data for non-expert end-users [5, 6, and 10].

MDM methodologies can be categorized within Data-driven also referred as supply-driven, Requirement-driven also known as goal or demand-driven and hybrid frameworks. Data-driven applies techniques and procedures applied on a relational schema on data sources to identify multidimensional data. While the Requirement-driven approach give more relevance to the user multidimensional requirements and later map them onto the data sources. Also, the hybrid approach combines both frameworks. This approach designs data warehouse from data sources but in consideration of the end-user requirement [5, 14, 12].

**2.2.1 Star schema**

Star schema is a DW schema that consists of one or more central fact table linked to numerous dimension tables .Moreover, the term star schema is originated from the physical model of the schema which looks like a star. It has a central fact table with links emanating to the surrounding dimension tables in a star shape as shown in figure 3 [11, 6].
However, star schema is optimized to querying big data. This capability makes it desirable for use in data warehouses and data marts to support OLAP Cubes, BI and other analytics applications. [11].

2.2.2 Fact table

Fact table is the central table in a multidimensional schema that contains measures or metrics of business process or subject of the analysis. Fact table accommodates numerical values and foreign key to the dimension tables. In addition to that, fact tables can record events at an atomic level of detail (as they are designed to a low level of granularity) [5, 10, 11]

2.2.3 Dimension tables:

Dimension table is the table that records the attributes that describe the fact data [5].

2.2.4 Benefits of a star schema

Star schema provide some functionalities that made it desirable for data warehousing, these include but not limited to [10];

i) Simpler queries
ii) Simplified logic for business reporting
iii) Improved query performance
iv) Fast aggregation
v) Populating or feeding data cube

2.3 Business Intelligence (BI)

BI is a collection of methodologies and technologies for collection, integration and analysis of enterprise information aimed at enabling knowledge workers in making informed decisions [17]. However, BI has over the years become a ubiquitous technology that is found in almost every sector of human endeavour. Thanks to the dwindling cost of managing big data and its features that accords user with more sophisticated data analysis capabilities. These enable users to better harness their data assets, deliver new functionalities and draw informed business decision [8].

BI tool consists of data warehouse, OLAP servers and so on at the back

2.4 Talent Management (TM)

TM can be simply described as a forecast of human capital demand and creating a plan on how to meet that demand [2]. Moreover, a more detailed definition of the concept was given by Collins et al [3]. According to [3] TM can be perceives organizational talent management strategy as activities that involve three vital phases. Firstly, the systematic identification of positions which differentially contributes to organization’s sustainable competitive advantage. Secondly, the development of a talent pool of high potential and high performing incumbents to fill this roles. Lastly, the development of a human resource architecture to facilitate filling these positions with competent incumbents and ensure their continued commitment to the organization [3]. Figure 1 below shows talent management cycle.

However, a substantial part of strategic Human Resource Management literature recognizes human capital as a key to gain and sustain a competitive edge [20]. For this reason, TM becomes a major priority of HR managers. In Europe TM is considered as one of the five crucial challenges of HR [21].

2.5 Related Works

A lot of researches discuss the practice and application of information technology in talent
management and human resource practices. These similar researches provide models for information systems, decision support systems and data warehouse to automate and facilitate various talent management and HR activities.

Zhang, (2009) proposes a framework that uses DW technology in University teachers’ appraisal. The DW serves as a source of information that is used to evaluate teachers’ performance in teaching methodology and scientific research. Moreover, the data warehouse mainly stores three data sets about each staff these are the basic information data sets which consist of curriculum vitae technical and professional titles, Research information data set which comprises of publications, workshops, project and work done and lastly teaching information data set which include teaching workload and ability level. However, this information is extracted and cleansed of unnecessary information from the university HR database and stored in the data warehouse to provide consistent and consolidated information to enable the relevant stakeholders to make informed decisions while evaluating staff.

Similarly, Suresh and Mahale, (2011) proposed an architecture for student performance analytics using data warehouse in E-Governance System. Moreover, a multidimensional model was used to develop the star schema the data warehouse. The data warehouse will extracts, cleanse and consolidate student academic performance data from various higher institutions. This will enable relevant stakeholders to perform various analytics on students performance, recommend and implement changes where necessary for improvement of students performance.

Mohammed and Anad, (2014) proposed data warehouse architecture for Human Resource in Iraqi. The data warehouse will consolidate data about competent graduates from universities in Iraq. This will enable governmental and non-governmental human resource organizations to easily access the information, identify and hire competent talent for their organizations.


Faiyaz et al (2009) presented an approach for ranking job applicants using decision support system. The model used a fuzzy based agent for group decision support of applicant ranking within recruitment systems to handle uncertainties and inconsistencies in the collective decisions of recruitment expert panel. However, the system will enable automation of requirement specification process and applicant matching/ranking. The system has been tested and proved worthy in residential care sector in which it produced ranking decisions that were consistent with those of human experts.

Another research conducted by Doka et al proposed a framework for Integrated Decision Support System (IDSS) based on Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) model to facilitate Human Resource Selection (HRS). The proposed model consists of TOPSIS, Fuzzy TOPSIS and Fuzzy Set Theory integrated together with aim of providing an avenue for decision makers to conduct fair HRS.

Ahmad and Abdallah, (2009) Proposed an expert system called Smart Evaluation on Job Vacancy Application System that intuitively shortlist candidates for job application. The proposed system applies Ruled-based expert system techniques to generate shortlist of candidates that satisfies the company’s requirement. The system screen series of job applications online and generate a list of candidates to be invited for the interviews.

This researches show the benefits that can be derived from using DW technology in supporting various HR activities. However, these studies have contributed immensely in facilitating and providing solution to various HR and TM practices. Alas, there are still some loopholes to be plugged. Because almost all the proposed DW models reviewed does not give much consideration to the industrial needs to be satisfied. Thus it is very likely to have a gap between industrial needs and the supplied candidates. Similarly, the most of the DSS focuses on the selection criteria that usually consider overall CGPA of a candidate rather than the individual courses performance that determines the strength of that candidate in his area of specialization.

Therefore, the propose model tend to bridge the aforementioned gap by adopting a hybrid approach.
of MDM which considers both the data sources (candidates’ information) and user requirement (industrial needs).

3. RELATED RESEARCH

The study is based on the principle of multidimensional modeling. The simplicity and understandability as stated in section 2.2.4 informed the choice of the method. Moreover, a hybrid approach of MDM will be adopted in developing the model of the DW. This is in order to better understand problem and provide a holistic solution. Therefore both data sources (student information) and user requirement (industrial needs) will be considered during business process definition. After wards Kimball approach will be adopted to identify and model the multidimensional concepts.

This implies that, Kimball’s MDM guidelines will be applied on the IHL’s data in consideration of MDEC jobs to drive the schema of the DW model. Like every other DW that conforms to Kimball’s approach, it consists of four basic steps []:

Steps I: Definition of business process: This is the first stage of DW development. It involves the choice of the subject on which the DW will be built. This is achieved through careful evaluation of the available data and in consideration of user requirement. That is the IHLs data and MDEC job position respectively. In this case, Talent is the business process. Therefore available IHLs data and MDEC job position will be considered to properly understand the type of talent needed by the industries.

Steps II: Declaration of grain: The atomicity of the data to be stored in the fact table is declared

Step III: Choice of dimension: In this stage the dimensions that apply to each fact row are chosen. The available data is thoroughly evaluated to identify and choose the multidimensional concepts that will form the DW. These are the dimensions, measures. The data will be analyzed based on two key elements. These elements are the subject that is being analysed and the evaluation criteria. Further, the analyzed subject represents a dimension and the analysis criteria represent the measure. However in this study, the corresponding dimensions are the student (talent), institution attended, the years spent, the subjects offered, and the program and the measures are number of years spent, number of subjects offered and so forth.

Step IV: Identification Facts

This is final step of the model design, a set of dimensions and their associated measures makes a fact. further, facts usually contains numerical values of performance indicators or measure. In this regard the corresponding facts are the Marks scored per subject and Grade Point Average (GPA) are the corresponding facts that forms the DW schema.

3.1 Conceptual Framework

The purpose of this framework is to understand the industrial talent needs) and identify the various sources of data to be used in developing the model. Also to clarify the phases and processes involved in developing the proposed model Figure 5 show the conceptual framework.

![Conceptual Framework](image)

As illustrated in the figure 5, the conceptual framework consists of three phases as follows;

3.1.1 Phase I: This phase involves identification of the various sources of the data needed and requirement to be considered in developing the DW
model. In this study, personal information and academic records of student across IHLs in Malaysia will serve as the data sources. On the other hand, job vacancies identified by MDEC will serve as the user requirements.

3.1.2 Phase II: This is the phase where jobs are mapped to relevant course. Recall that the DW is for TM and therefore needs to contain the data that can facilitate such activities. To achieve that, the two set of data obtained were then mapped together based on students courses related to respective jobs. That is the IHLs data and MDEC job position respectively.

Further, the mapping is between the courses and relevant jobs not job to candidates. Because each job position has some basic courses a particular talent needs to offer in school in order to have the requisite skills to perform that job and the competency of a candidate is defined by his possessed skills and the skills are achieved through the courses offered. For example, a talent needs to pass programming, system analysis and design and other relevant courses to be employed as a software developer. This mapping was later translated into relationships from which some of the dimensions were derived.

3.1.4 Phase IV: This is the modelling phase where the schema of the target DW is modelled. This is where MDM comes into play. However, Kimball’s MDM guidelines will be applied on the IHL’s data inconsideration of MDEC jobs to drive the schema of the DW model. Like every other DW that conforms to Kimball’s approach, it consists of four basic steps.

3.1.5 Phase IV: This phase is where the vital extraction transformation and loading process is built and executed an. DW obtains its data from various and often heterogeneous data sources. For these reason, the data need to be transformed into a consistent format needed to be stored in the DW. Thus the ETL becomes imperative. This phase begins with the extraction of data from the various identified data sources. The data extracted will be converted and stored in a single database.

Secondly, the data will be transformed into an appropriate format for storing in the target database. During this process, a series of transformation rules and functions are applied on the extracted data. In addition, it permits data from different sources to be joined, refined and categorized based on specific attribute. Furthermore, field and columns with similar attributes but different nomenclature are identified and converted into the format allowed by target database. For instance gender and sex represent the same attribute but has different naming. Incidence like this will be checked before the data is transferred in to the target database. Similarly unwanted data and duplicate data are removed from the data before it is migrated in to the target database. Eventually, this process cleans the data and ensures consistency of the data type and format.

Finally, after the data is cleaned and formalized, it will be loaded into the DW according to the logical schema definition. However, this is achieved through special migration and cleaning tools called Extraction, Transformation and Loading (ETL) Tool.

4. RESULTS AND DISCUSSION

DW implementation can be very crucial and costly especially when various heterogeneous systems are involved. However, it significance outweigh all the implementation bottlenecks. Successful implementation of this framework will provide a DW that will serve as a consistent source of information about industry relevant talents. Additionally, having already mapped the curses to the jobs will facilitate direct mapping of candidate to jobs. This because each job(s) identified has been mapped with the required course(s) and ultimately to any candidate that offer such course(s). Various related analytics such mapping talents to industrial needs using appropriate BI tool.

5. CONCLUSION AND FUTURE WORK

TM is considered by industries globally as an effective instrument to gain and sustain competitive advantage and growth however the environment or economic situation. However, this research presented a multidimensional model for data warehousing to integrate and consolidate data about potential candidates (talents). The model will serve as source of quality assured information that will aid HR managers in identifying suitable candidates for job recruitment. Also it will enable them to perform various TM related analytics using appropriate BI tool. In future the model will be developed and implemented.

6. ACKNOWLEDGEMENT

The author will like to extend a warm appreciation to the Kano State Government, Nigeria.
and University Sultan Zainal Abidin, Malaysia for their unflinching support to ensure success of the study.


