Camera Ready Paper

Development of Analytical DataMart and Data Pipeline for Recruitment Analytics

Ashish Chandra Jha[1], Sanjeev Kumar Jha[2], and J. B. Simha[3]

[1,2,3]REVA Academy for Corporate
Excellence, REVA University,
Bengaluru, India

ashish.ba05@reva.edu.in[1]

sanjeev.ba05@reva.edu.in[2]

jb.simha@reva.edu.in[3]

Abstract. The HR department handles all the data regarding the recruitment process while also analyzing them to select suitable candidates for the organization. The HR department interacts with the data regarding the recruitment using many tools for data analysis and interpretation. Such data should be organized properly to ensure the screening of candidates. Each system gives their own report in their own format. Data is not integrated and aggregated at proper granularity suitable for analysis. This can be mitigated by the use of Dimensional and Analytical DataMarts. The dimensional DataMarts will contain transitional data modified for analysis with dimensions and facts. These DataMarts can be used for any adhoc analysis like drill down/roll up, slice and dice, drill through, comparative analysis etc. The analytical DataMart will contain the aggregated data in one row per employee format. This will be task specific like Quality of Hire modeling, Cost of Hire, Time to Hire, demand prediction etc.

This paper has focused on creating a recruitment DataMart for Human Resource Department with the assistance of the data modeling technique. This Paper has also discussed the exploratory data analysis and model building in relation to the DataMarts for the recruitment.

Keywords: HR Analytics, Recruitment, DataMart, Analytical DataMart, Dimensional DataMart.

1. Introduction

The success and growth of any organization primarily depend on its employees. Without employees, no organization can run as they are the workforce for the organization. Various employees take up different roles in the organization and perform the duties assigned to their positions. The employees of an organization are crucial for the organization and therefore the department that majorly deals with them becomes crucial as well. This department is the Human Resource Department of an organization. The HR department is responsible for almost all aspects concerning the workforce of an organization. The importance of employees for an organization makes the HR department very important. The Human Resource Department is responsible for the effective administration of the employees of an organization to help in its overall growth. Moreover, an organization can only succeed if competent people are recruited in it. This department needs to carefully examine and select suitable persons for the various roles in the organization.

1.1. Recruitment Process
Recruitment can be defined as the overall process of choosing the right candidates for job positions at an organization based on several criteria. The recruitment process is usually handled primarily by the Human Resource department of an organization. This process might differ for every organization but there are certain general steps involved in it that most organizations seem to follow. These are Job Description Preparation, Sourcing, Screening, Selecting, Hiring, and Onboarding.

1.1.1. JD Preparation

Job description refers to the details of the job position for which the recruitment is taking place. It contains all the necessary information regarding the job such as-

- Qualifications required for the job

- Skills required to perform the necessary functions of the job

- Different responsibilities that the selected candidate has to take care of

- Suitable characteristics for the candidates for the job such as age, experience, etc.

- Salary expectations concerning the job

- Working hours

The job description is an important aspect of the recruitment process as it notifies the interested candidates with all the necessary information about the job. It is the primary source of attracting candidates for the job position.

1.1.2. Sourcing

Sourcing is the process of collection of data relevant to the position for which recruitment is being done. In other words, this process involves the collection of resumes or data of various candidates who are suitable for the concerned position. The HR department scrambles through multiple portals for data collection such as Naukri.com, LinkedIn, monsterindia.com, etc. This process does not involve the selection of the candidates, it only involves the collection of the information of the relevant candidates for further screening and selection.

1.1.3. Screening

Screening is the first assessment of the candidates with the necessary qualifications for the job position. All the resumes and data that are collected in the sourcing process are reviewed and screened by the HR department. This stage finds out those candidates who will be advancing to the next level of selection. Screening the resumes of candidates and selecting the top ones from the whole lot is the usual activity in this process.

1.1.4. Selecting

Selecting is the main stage of choosing the most suitable candidate for the job for which the recruitment is being done. Those candidates who passed the screening stage have to go through this phase of selection which might involve multiple tests. Personal tests, group discussions, aptitude tests are some of the most common tests that are used to select the most suitable candidate for the job post. The candidate who is selected also gets to negotiate his/her salary with the HR department in this phase itself. The best candidate is selected for the job and no more candidates are assessed after this selection stage.

1.1.5. Hiring

The candidate who is finally selected in the selecting stage is hired by the company. This occurs in this phase called hiring. The concerned candidate is presented with the offer letter from the organization's HR department. All the terms and conditions of the job along with other necessary information regarding the job are presented to the chosen candidate at this stage. If the organization and the chosen candidate agree on all the terms and conditions set forward, the candidate is hired into the organization.

1.1.6. Onboarding

Onboarding is the last phase in the recruitment process and by this stage, a suitable candidate is already selected and hired for the concerned job position. In this phase, the new employee is welcomed into the organization by the HR department. Onboarding is the process of introducing the new employee to the organization, its cultures, rules & regulations, his/her role, other employees, and the way of working. It is a way of familiarizing the new employee about the various aspects of the organization and the job role.

HR Analytics is defined as a HR process which uses the HR data (e.g. recruitment data) and domain knowledge to do start predicting the performance and cost involved of the people on the basis of their work. (Kailash)

2. Literate Review

According to the Gartner, human resource leader has begun analyzing the HR data at each level to improve the HR efficiency, recruitment efficiency and enhance employee experience (George S. , 2012). In the fast-

growing world with their emerging new challenges, it is very difficult to hire skilled people which will be in the organization for a decent span of time. To create such system, we need good quality of data backed by domain knowledge to provide analytical suggestion to hire a candidate. To create such analytical system, the business data should be extensively available to the hiring manager to work align to find the suitable candidate (Kimball M. R., Data Warehousing, Business Intelligence, and Dimensional Modeling Primer, 2016).

For a smooth functioning of an organization, their strategic decision should be backed by data. Data helps us to find the insights or patterns among the people or work. This data should be properly created and organized in such a way that it can be further used to perform any analytical tasks. The success of human resource department with the help of analytics can be measured as the success of the organization. HR department tries to understand the business requirement and ensure the finding of right candidate according to their skills and qualifications. To hire such a niche level people we need to create a precise DataMart and use its data for analytical view. Human resource (HR) DataMart is the base stone for building an enterprise data warehouse. The paper presents the implementation process of HR DataMart starting from implementing DataMart schema to online analytical processing (OLAP) reports (Kimball M. R., 2016).

There are quite a few DataMarts available in the market these days. Among them, quite popular are the Kim Kimball HR DataMart which is the base for many available DataMarts, Oracle HR DataMart, SAP HR DataMart.

2.1. HR DataMart by Ralph Kimball

The Kimball HR DataMart is based on the ideology of Ralph Kimball according to whom the data warehouses should be model using dimensional models such as the star schema or snowflake schema. Kimball's approach was the bottom-up approach that involved creating DataMarts first to allow quick analysis and interpretation of the data in them. These small DataMarts can later be combined to form a normal data warehouse. Kimball focused on the use of dimensional models to improve the performance of the users of the data warehouse (Oracle, Recruiting Data Mart Dimension Tables, n.d.).

In order to understand Kimball's HR DataMart, we have to look at the star schema dimension model. A star schema is a tool for dimensional modeling of data by organizing it to allow analytical operations to run on it. The star schema presents the data in the form of fact tables and dimension tables. The fact tables contain the primary data regarding the business process, in our case, the recruitment data, which are termed as 'facts'. The other tables or dimension tables contain the data associated with the primary fact tables. The dimension table contains 'dimensions' which sort of describe the data in the facts table.

The dimensional tables surround the fact table depicting the descriptive nature of the facts. The multiple dimension tables around the fact table make this model look something like a star due to which it is known as the star schema.

Kimball's HR DataMart focused on doing the harder things in the beginning while keeping the easy things for the end. Below is the Employees tracking DataMart from Kimball. I could not find anything specific to only recruitment in Kimball DataMart, which we will be disusing inferring Kimball DataMart concept (Kimball M. R., Data Warehousing, 2016). Kimball HR DataMart is shown in Figure 1.
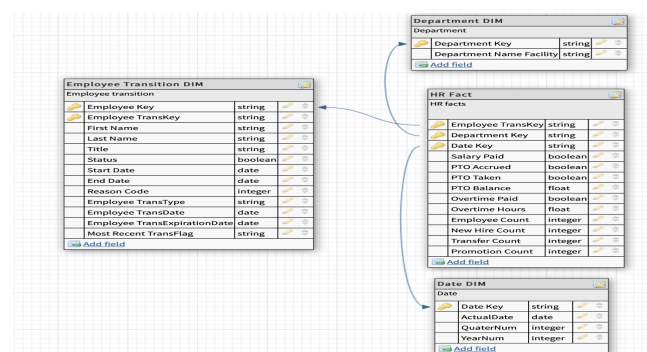

Figure 1 - Kimball HR DataMart (Ralp Kimball, 2016)

2.1.2. HR DataMart by Oracle

Then we have Oracle DataMart. Oracle Recruitment DataMart is a data storehouse of enlistment also, staffing drives, status, costs, and results. This information store contains more outlined data and all current and memorable well-known enrolment drives including open positions, orders, candidates, applications, results, and enrolment cost. Recruitment measurements can be reported and investigated by a wide assortment of credits, including specialty unit, division, work code, area, status, status reason, reference class, and subcategory, etc. The Recruiting information store furnishes staff and the board with data

expected to settle on informed choices with respect to current and future recruitment drives (Oracle, Recruiting Data Mart Dimension Tables, n.d.).

Oracle gives a total and self-administration arrangement that permits business groups to get the profound, reliable, information-driven experiences they need to settle on speedy choices. Business groups can rapidly join all vital information across various sources and organizations, including spatial and diagram, in a combined data set to drive secure cooperation around a solitary wellspring of truth given by information stores. Analysts can without much of a stretch influence self-administration information devices and implanted AI—with zero coding needed—to speed up information stacking, change, and readiness, consequently find examples and patterns, make expectations, and gain experiences dependent on information with straightforward ancestry (Prado, 2010) (Hamoud, 2020). Oracle DataMart has been shown in Figure 2.
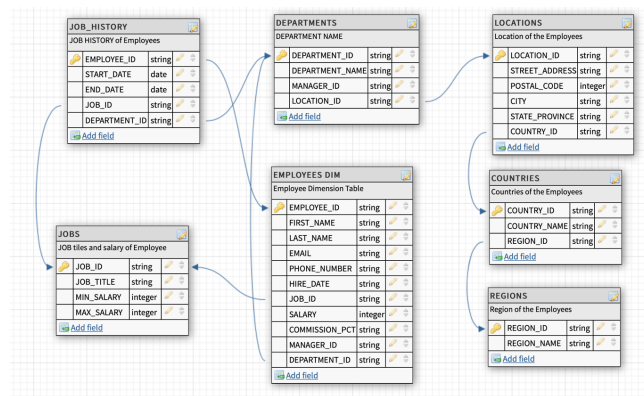

Figure 2 - Oracle HR DataMart (Oracle, Recruiting Data Mart Dimension Tables, n.d.)

3. Methodology

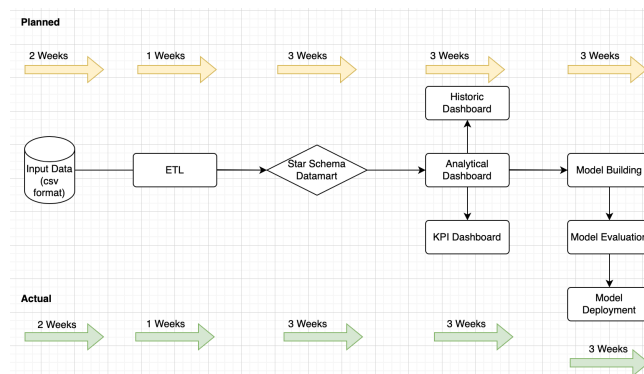The following point consists of the proposed methodology: -

Figure 3 - Pipeline

3.1. Load the Data :- At this stage, the data will be load to the system. Here we are accepting only csv file as input data for now.

3.2. ETL and finding fact and dimension tables:- Data exploration and further cleaning as per the requirement. Ingest the clean data into system. Dimension table the tables of a star schema which store the attributes that describes the object. Here dimension tables are Candidate Reference Id, Manager Id, Job Role, Job Category, Source, Salary Band, Experience etc.

Facts are the numeric dimensions of the business. They support numerical computations used to provide details regarding and investigate the business. Here fact tables are Cost of Hire, Quality of Hire, Time to Hire.

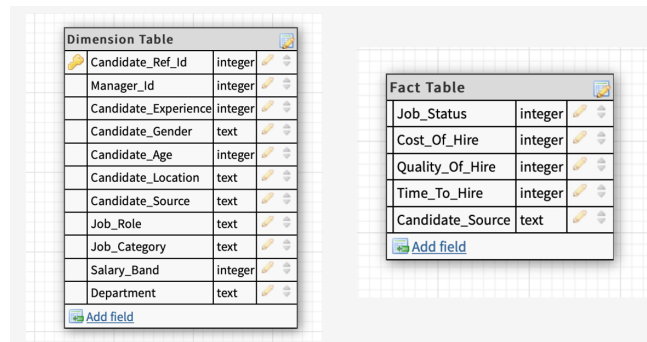Figure 4 shows the dimension and fact tables.

Figure 4 – Dimension & Fact Tables

3.3. Star Schema DataMart:- The clean data will be used to find the dimension and fact tables. Further star schema DataMart will be created.

Figure 5 shows the star schema for recruitment process. Star schema has been taken under consideration for this paper. Snowflake schema is just to understanding in detail purpose. Figure 6 shows Snowflake schema.
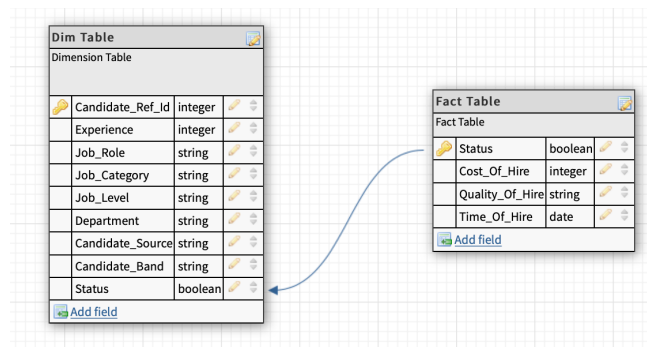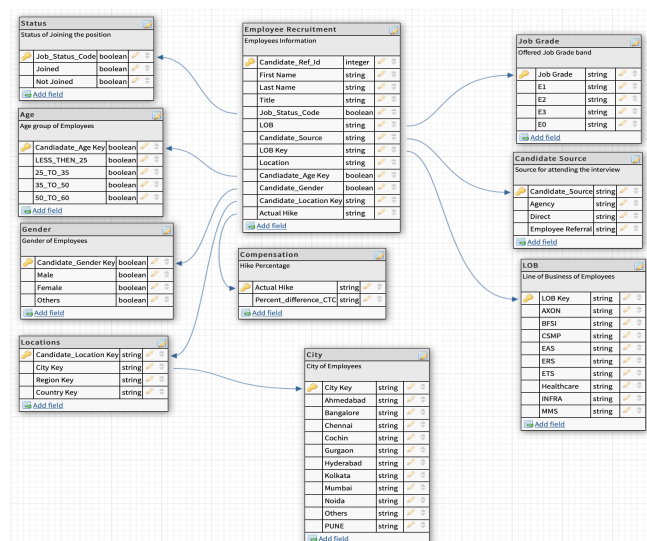

Figure 5 - Star Schema

Figure 6 - Snowflake

3.4. Creating Analytical Dashboards:- Analytical dashboard, historic dashboard and KPI dashboard will be created. Analytical DataMart helps us to create decision management platforms where data scientists can use the analytical DataMart for their analysis. Analytical DataMart Example has been shown in Figure 7.
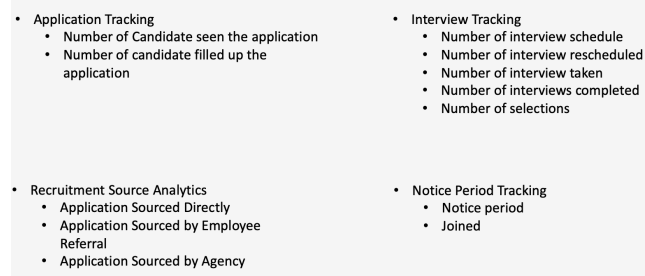

Figure 7 - Analytical DataMart

3.5. Model Building:- After data modelling by building DataMart, segmentation has been done using K-Mean Algorithm. To check the number of clusters, elbow method has been used which is shown in Figure 8.
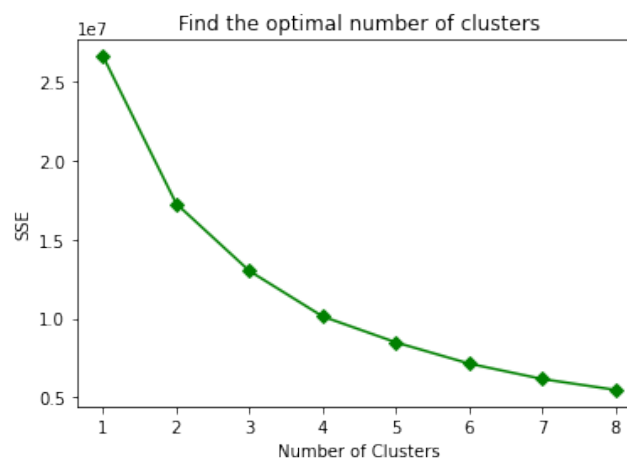

Figure 8 – Elbow Method for K-Mean Clustering

Once the data is segmented in clusters, historic dashboard has been built on data studio tool. The Historic Dashboard has been shown below in below figure. Historic Dashboard shows the historic trend of the recruitment referring to other features like cost of hire, hiring cost and hiring time. Historic Dashboard has been shown in Figure 9.
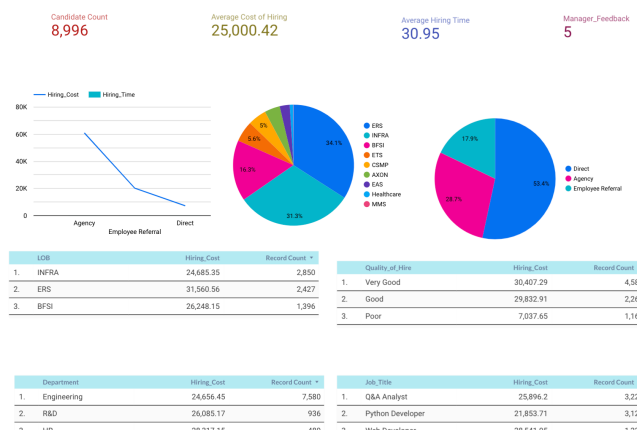

Figure 9 - Historic Dashboard

3.6. Deployment:- The DataMart has been designed on dbdesigner which generates the SQL code which can be further developed in any database tool like Microsoft SQL server management studio or oracle database or can be built on any cloud like Amazon Web Services, Google Cloud Platform, Microsoft Azure. Various dashboards like historic and predictive dashboards have been made on Datastudio which is a free tool from Google. Alternately one can use Power BI or Tableau as a substitute of Datastudio. For segmentation (modeling), K-Means algorithm has been used to build the model. The model has not been deployed as of now.

4. Results and Discussion

After A DataMart is the access layer of a data warehouse used to offer data to users in business intelligence. Small chunks of the data warehouse are often referred to as DataMart. Data warehouses typically hold enterprise-wide data, whereas DataMart typically store data unique to a department or team. The primary goal of DataMart is to offer the most relevant data for BI to the business user in the shortest time feasible. Millions of records can be stored in a DataMart, which necessitates terabytes of storage. The following are some of the benefits of employing a DataMart:

- Allows users to have access to the precise type of data they require, which improves end-user response time.

- A data warehouse that has been compacted and narrowed in scope.

- Each one is devoted to a single item or function.

- Implementing a full data warehouse is more expensive.

- Contains a lot of information.

- It is less crowded and simply contains necessary company information and data.

- Assists in the integration of all data sources

- The building and use of a DataMart result in a significant amount of data summarization.

In this paper, a DataMart has been built up referring to Kimball HR DataMart and Oracle recruitment DataMart. A star schema with facts and dimension tables has been built up to develop KPI DataMart and analytical DataMart. Further, segmentation model has been created using K-Mean algorithm to cluster the similar kind of candidates. As per elbow method, five clusters have been created to segment the candidates. Later, historic dashboard has been developed to analyze the candidate profiles in details.

5. Conclusion and Future Scope

Different human resource Datamart like Kimball HR Datamart, Oracle HR Datamart, SAP- HANA Datamart has been studied and inferred. Based on inferences from popular Datamart, industry expert experience and through research done, a new Datamart for recruitment has been created. This Datamart will be a good option to start building the recruitment Datamart. The K-Mean algorithm has been used to build a model on recruitment data. Candidates are segmented under 5 classes with different features. Historic and predictive analytical dashboard has been created on recruitment dataset.

Future work will be extending recruitment process to other human resource process like Employee Lifecycle, Employee Exit (Retention) etc. Will be creating Datamart, create dashboards and build model for the same process.

References

Kailash, M. P. (n.d.). "HR Analytics Methodical Measurement of HR Processes," International Journal of Innovative Science and Research Technology , 2020.

George, S. (2012, April 14). *Inmon or Kimball: Which approach is suitable for your data warehouse*. Retrieved from https://www.computerweekly.com/tip/Inmon-or-Kimball-Which-approach-is-suitable-for-your-data-warehouse

Kimball, M. R. (2016). Data Warehousing, Business Intelligence, and Dimensional Modeling Primer. In *Kimball The Data Warehouse Toolkit 3rd Edition* (pp. 7-16). Wiley.

Kimball, M. R. (2016). Employees Tracking for Profile . In *The Data Warehouse Toolkit: The Defi nitive Guide to Dimensional Modeling, Third Edition* (pp. 263-265). John Wiley & Sons, Inc.

Oracle. (n.d.). *Recruiting Data Mart Dimension Tables*. Retrieved from https://docs.oracle.com/cd/E41507_01/epm91pbr3/eng/epm/phcw/concept_HCM WarehouseStructure-399b81.html.

Kimball, M. R. (2016). Data Warehousing. In *Business Intelligence, and Dimensional Modeling Primer," in Kimball The Data Warehouse Toolkit 3rd Edition* (pp. 7-16). Wiley.

Ralp Kimball, M. R. (2016). Employees Tracking for Profile. In *The Data Warehouse Toolkit: The Defi nitive Guide to Dimensional Modeling, Third Edition* (pp. 263-265). John Wiley & Sons, Inc.

Oracle. (n.d.). *Recruiting Data Mart Dimension Tables*. Retrieved from https://docs.oracle.com/cd/E41507_01/epm91pbr3/eng/epm/phcw/concept_HCM WarehouseStructure-399b81.html

Prado, C. F. (2010). Using OLAP Tools for e-HRM: A case study. *International Journal of Technology and Human Interaction*.

Hamoud, M. A. (2020). Improve HR Decision-Making Based On Data Mart and OLAP. *Journal of Physics*.

Oracle. (n.d.). *Recruiting Data Mart Dimension Tables*. Retrieved from https://docs.oracle.com/cd/E41507_01/epm91pbr3/eng/epm/phcw/concept_HCM WarehouseStructure-399b81.html

Udhay Kailash, M. P. (2020). HR Analytics Methodical Measurement of HR Processes. *International Journal of Innovative Science and Research Technology* .

Alexandra. (2019, August 27). *25 Practical Tips For Building And Managing A Solid Recruitment Funnel*. Retrieved from https://harver.com/blog/recruitment-funnel/

George, S. (2012, April 14). *Inmon or Kimball: Which approach is suitable for your data warehouse*. Retrieved from https://www.computerweekly.com/tip/Inmon-or-Kimball-Which-approach-is-suitable-for-your-data-warehouse

Ralph Kimball, M. R. (2016). Data Warehousing, Business Intelligence, and Dimensional Modeling Primer. In *Kimball The Data Warehouse Toolkit 3rd Edition* (pp. 7-16). Wiley.

SAP. (2018, 1 24). *SAP HANA as Data Mart.* Retrieved from https://help.sap.com/doc/e95f6750b0fd10148ea5c6be75016694/1.0.12/en-US/SAP_HANA_Master_Guide_en.pdf

Gartner. (2019, 06). *Gartner Identifies Three Most Common AI Use Cases in HR and Recruiting*. Retrieved from https://www.gartner.com/en/newsroom/press-releases/2019-06-19-gartner-identifies-three-most-common-ai-use-cases-in-

ttttt. (n.d.). Retrieved from Gartner Identifies Three Most Common AI Use Cases in HR and Recruiting

google. (n.d.). Retrieved from www.google.com

Alysson Prado, C. F. (2010). Using OLAP Tools for e-HRM: A case study. *International Journal of Technology and Human Interaction* .

Alaa Khalaf Hamoud, M. A. (2020). Improve HR Decision-Making Based On Data Mart and OLAP. *Journal of Physics*.

Inuwa, I. (2015). Design of a Data Warehouse Model for a University Decision Support System. *Journal of Information & Knowledge Management , 5*.